Probability Theory and Related Fields manuscript No.
(will be inserted by the editor)

Natalia Bochkina

Besov regularity of functions with sparse random wavelet coefficients

Received: 17 March 2006 / Revised: date

Abstract We consider a problem of determining the necessary and sufficient conditions of the membership in Besov spaces (with probability 1) of random functions defined in terms of the wavelet decomposition with random wavelet coefficients. Motivated by the settings of the Bayesian non-parameteric regression where the majority of the functions to be estimated are assumed to be regular enough so that the majority of the wavelet coefficients are zero we consider a probabilistic model for the wavelet coefficients which allows for their sparsity. We consider such probabilistic models for the coefficients of the orthogonal and the continuous wavelet transforms.

Keywords Besov spaces · random function · orthogonal wavelet transform · continuous wavelet transform · Poisson process

1 Motivation.

Consider the following white noise model

$$dY(t) = f(t)dt + \sigma_n dW(t), \tag{1}$$

where $\sigma_n^2 = \tau^2/n$, $f \in L^2[0,1]$ is unknown function of interest and W is a standard Wiener process. In practice, since instrumentally acquired data is usually discrete, it is simplified to an equally spaced regression:

$$y_i = f(i/n) + \sigma_n \varepsilon_i, \quad i = 1, \dots, n, \tag{2}$$

$\varepsilon_i \sim N(0,1)$. Brown & Low (1996) showed the asymptotic equivalence between (2) and the original non-parametric regression model (1) under mild conditions.

Address: Imperial College London, Norfolk Place, London W2 1PG, UK
E-mail: N.Bochkina@ic.ac.uk

Wavelet transforms are a commonly used tool in signal and image analysis due to fast decomposition algorithms and sparsity of the representation (e.g. Mallat [16], Chapter 26). It possesses such attractive properties as adaptivity to the regularity of the function, sparsity in the number of wavelet coefficients sufficient to represent vast majority of functions with high precision and decorrelation of coloured noise. Wavelet-based estimators of f are shown to perform well and possess good optimality properties (Donoho and Johnstone [8], [10]). Using the Besov spaces to characterise the decay of wavelet coefficients is most natural for two reasons: firstly, the Besov spaces characterise the smoothness of a function more precisely than, for instance, the Sobolev spaces, and secondly, the characterisations are expressed by simple conditions on wavelet coefficients. The Besov spaces consist of functions of different levels of regularity and include, in particular, the well-known Sobolev $W_2^s = B_{2,2}^s$ and Hölder $C^s = B_{\infty,\infty}^s$ spaces of smooth functions, and also less traditional spaces, like the space of functions of bounded variation, sandwiched between $B_{1,1}^1$ and $B_{1,\infty}^1$. For more details see Meyer [17] and Triebel [22]. Donoho and Johnstone [9] discuss the relevance of the Besov spaces for various scientific problems. The Besov spaces have been widely used in wavelet context (e.g. Donoho and Johnstone [7], [8], [9], Abramovich et al [1], [2]) due to being wide enough to contain irregular functions which can be represented in terms of wavelets. Donoho and Johnstone [9] discuss the relevance of Besov spaces for various scientific problems.

An example where these results are important is a study of optimality of Bayesian estimators over functional spaces. A Bayesian approach has been popular in statistical applications of wavelets since it allows to model sparsity of wavelet coefficients in an interpretable way to produce a variety of wavelet estimators of a function. A widely studied case of the prior distribution of wavelet coefficients modelling sparsity is a mixture of atom at zero and normal distribution where the estimators can be obtained in closed form (e.g. Clyde and George [5], Abramovich et al [1], Vidakovic [23]). Abramovich et al [1] also investigated whether such choice of prior distribution is equivalent to the prior assumption of Besov membership with probability 1 of the function of interest for the normal mixture prior. This property, for instance, can be used to specify the prior distribution if the regularity of the function to be estimated is known in advance. However it has been shown that wavelet estimators with normal prior distribution of the non-zero wavelet coefficients, although computationally simple, do not possess good optimality properties since they severely shrink large wavelet coefficients due to light tail of normal distribution (Bochkina and Sapatinas [4]). Heavy tailed prior probability models have been proposed by Vidakovic [23], Pensky [20], Johnstone and Silverman [13]. Pensky [20], whilst studying optimality of Bayesian wavelet estimators with normal and heavy tailed mixture prior distributions over the Besov spaces, provided a necessary condition for a function to belong to Besov spaces a priori with probability 1.

Here we provide a necessary and sufficient condition of the Besov spaces membership of a function with more general mixture distribution of its wavelet coefficients, namely of atom at zero and an arbitrary distribution, thus extending the Besov membership criterion to a wider class of prior distributions. We show that in the majority of the cases only the knowledge of the finite absolute moments is necessary to obtain the criterion, whereas for some combination of parameters the knowledge of the tail behaviour is necessary.

Moreover, we also show that this necessary and sufficient criterion, under stronger assumptions of the regularity of the wavelet function, holds for the mixture model adapted to the continuous wavelet transform. The continuous wavelet transform provides a greater flexibility of modelling functions of interest since it does not have dyadic restrictions on the indices of the wavelet function. Since the wavelet coefficients are related to the size of the discontinuity of a function at the corresponding frequency and location, continuous wavelet transform is particularly well suited to practical problems where the frequencies of the non-zero wavelet coefficients are of interest, and it finds its application to data in meteorology, oceanography and medicine (e.g. Ouergli [19], Polygiannakis et al [21], Mochimaru et al [18]). To model the sparsity of wavelet coefficients of the continuous wavelet transform, we use approach suggested by Abramovich, Sapatinas and Silverman [2] by modelling a set of indices of the non-zero wavelet coefficients as a two-dimensional Poisson process.

Here we adapt the general probabilistic model we consider for the orthogonal wavelet transform, a mixture of atom at zero and an arbitrary distribution, to the continuous wavelet transform and study the regularity properties of a function with such continuous wavelet coefficients, thus generalising approach of Abramovich et al [2] who considered only normal distribution for the non-zero values.

We proceed as follows. In Section 2 we give definitions of the wavelet transforms and Besov spaces. In Section 3 we describe the probabilistic (prior) model for the orthogonal wavelet coefficients and state the results on the connection of the parameters of the probabilistic model and the membership of the function in Besov spaces. In Section 4 we specify the probabilistic model for the continuous wavelet coefficients and state results about the regularity properties of the function with such wavelet coefficients. We conclude with discussion.

2 Wavelet transform and Besov spaces.

Below we give a brief introduction to wavelet transform and Besov spaces based on Vidakovic [24], Daubechies [6] and Abramovich et al [2].

2.1 Orthogonal wavelet transform

A wavelet basis is determined by a wavelet and a scaling function $\psi(x)$ and $\phi(x)$ which we assume to be of regularity r and have periodic boundary conditions on $[0,1]$. A set of functions $\{\phi(x), \psi_{jk}(x), j \in \mathbb{N}, k = 0, \dots, 2^j - 1\}$ forms an orthonormal basis of $L^2([0,1])$ where the functions $\psi_{jk}(x)$ are derived from the wavelet function by dilation and translation: $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$.

Since wavelets form an orthogonal basis of $L^2([0,1])$, a set of wavelet coefficients

$$w_{jk} = \int_0^1 f(x)\psi_{jk}(x)dx, \quad j = j_0, j_0+1, \ldots, \quad k = 0, \ldots, 2^j - 1, \qquad (3)$$

for some non-negative integer j_0 and the scaling coefficients

$$u_{j_0,k} = \int_0^1 f(x)\phi_{j_0,k}(x)dx, \quad k = 0, \ldots, 2^{j_0} - 1, \qquad (4)$$

uniquely define a function in $L^2([0,1])$. The orthogonal wavelet transforms arise naturally in a framework of the multiresolution analysis (Mallat [15]) which has important applications in functional approximations and regularity theory (Meyer [17], Walter [25]) and in signal processing (Mallat [16]).

2.2 Continuous wavelet transform.

A wavelet basis $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ with dyadic shift and scale can be extended to a set of functions with arbitrary shift and scale, i.e. $\psi_{a,b}(x) = |a|^{1/2}\psi(a(x-b))$, $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}$. This set of functions can be used to perform the continuous wavelet transform $T^{wav}: \quad L^2(\mathbb{R}) \to L^2(\mathbb{R}\setminus\{0\}\times\mathbb{R})$ in the following way:

$$T^{wav} f(a,b) = \langle f, \psi_{a,b}\rangle_{L^2},$$

where $\langle \cdot, \cdot \rangle_{L^2}$ is the scalar product in Hilbert space $L^2(\mathbb{R})$. Since a can be interpreted as a frequency we can restrict ourselves to the case $a > 0$. This corresponds to the case where both f and ψ are "analytical" signals, i.e. if $\text{supp}(\hat{f}) \subset (0,\infty)$, $\text{supp}(\hat{\psi}) \subset (0,\infty)$, where $\hat{f}$ is the Fourier transform of f, which implies $T^{wav} f(a,b) = 0$ for $a < 0$.

Now we give a short review of the continuous transform and its properties we shall apply later. See Vidakovic [24] for further details. If the *admissibility condition*:

$$C_\psi = 2\pi \int_0^\infty \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi = 2\pi \int_{-\infty}^0 \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty$$

is satisfied it is possible to recover a function from its continuous wavelet transform

$$f(x) = C_\psi^{-1} \int_0^\infty da \int_{\mathbb{R}} db T^{wav} f(a,b)\psi_{a,b}(x). \qquad (5)$$

Note that the admissibility condition implies $\hat{\psi}(0) = 0$, i.e. $\int_{\mathbb{R}} \psi(x)dx = 0$ which is one of the main properties of wavelet function.

To simplify the notation, we introduce $\lambda = (a,b)$. Taking $f = \psi_\lambda$ in equation (5) and denoting the continuous wavelet transformation of f by $F(\lambda) = T^{wav} f(\lambda)$, we have the following equation for $F(\lambda)$:

$$F(\lambda) = C_\psi^{-1} \int_{\mathbb{R}^2} F(\lambda')\langle \psi_\lambda, \psi_{\lambda'}\rangle d\lambda'.$$

Therefore $K(\lambda,\lambda') = \langle \psi_\lambda, \psi_{\lambda'} \rangle$ is a reproducing kernel for $T^{wav}f(\lambda)$. Since functions ψ_λ, $\psi_{\lambda'}$ are self-similar, the number of arguments in the kernel K can be reduced without loss of generality by introducing another kernel K_0 defined on $[0,\infty)\times(-\infty,+\infty)$ in the following way:

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

Then the original kernel K can be expressed in terms of the kernel K_0:

$$K(\lambda,\lambda') = K_0(a/a', a'(b-b')),$$

where $\lambda = (a,b)$, $\lambda' = (a',b')$. In particular case, where $\lambda' = \lambda_{jk}$, $K(\lambda,\lambda_{jk}) = K_0(a2^{-j}, 2^j b - k)$. These kernels have the following properties.

1. $K^2(\lambda,\lambda') \leqslant 1$.
2. If $[L_\psi, U_\psi]$ is the support of function ψ, then $K_0(u,v) \neq 0$ only if $L_\psi - U_\psi/u \leqslant v \leqslant U_\psi - L_\psi/u$.
3. If $f \in C^n(\mathbb{R})$ and $f^{(n)} \in C^\rho(\mathbb{R})$ with $\rho \in (0,1)$, $f^{(m)}$, $m = 0,1,\ldots,n$, are bounded and square-integrable, and ψ has n vanishing moments then

$$|\langle f, \psi_{u,v} \rangle| \leqslant C u^{-(n+\rho+1/2)} \tag{6}$$

uniformly in $u \geqslant 1$ (Daubechies [6]). Taking $f = \psi$, the following estimates hold for the kernel of compactly supported wavelets with r vanishing moments (Abramovich et al [2]):

$$|K_0(u,v)| \leqslant C u^{-(r+\rho+1/2)}, \quad u \geqslant 1;$$
$$|K_0(u,v)| \leqslant C u^{r+\rho+1/2}, \quad u \leqslant 1.$$

Here $C^\rho(\mathbb{R})$ is a Hölder space with exponent ρ which is defined in the next section.

2.3 Hölder spaces

Hölder spaces C^s are generalisations of C^n of functions with n continuous derivatives to the case of any positive real s.

Definition 1 For $s \in (0,1)$ we define

$$C^s(\mathbb{R}) = \left\{ f \in L_\infty(\mathbb{R}) : \sup_{x,h} \frac{|f(x+h) - f(x)|}{|h|^s} < \infty \right\}.$$

For $s = n + s'$, $n \in \mathbb{N}$ and $s' \in (0,1)$ we define

$$C^s(\mathbb{R}) = \left\{ f \in L_\infty(\mathbb{R}) \cap C^n(\mathbb{R}) : \frac{d^n}{dx^n} f \in C^{s'} \right\},$$

where $C^n(\mathbb{R})$ is a space of n times continuously differentiable functions on $\mathbb{R}$.

2.4 Besov spaces

Since the Besov spaces $B^s_{p,q}$ can be characterised by moduli of the wavelet coefficients of its elements for $s > r > 0$ where r is regularity of wavelet function (Donoho and Johnstone [8], [10]) we use this characterisation to define the Besov spaces. The explicit definition of Besov spaces can be found in Triebel [22].

Define Besov sequence norm $b^s_{p,q}$ on the coefficients of the orthogonal wavelet transform of function f from $L^2([0,1])$ defined in Section 2.1 for $p, q \geqslant 1$, $s > 0$ and $s' = s + \frac{1}{2} - 1/p$, as follows:

$$||\mathbf{w}||_{b^s_{p,q}} = ||\mathbf{u}_{j_0}||_p + \left\{ \sum_{j=j_0}^{\infty} 2^{js'q} ||\mathbf{w}_j||_p^q \right\}^{1/q}, \quad 1 \leqslant q < \infty, \tag{7}$$

$$||\mathbf{w}||_{b^s_{p,\infty}} = ||\mathbf{u}_{j_0}||_p + \sup_{j \geqslant j_0} \left\{ 2^{js'} ||\mathbf{w}_j||_p \right\}, \quad q = \infty, \tag{8}$$

where $\mathbf{u}_{j_0} = (u_{j_0,0}, u_{j_0,1}, \dots, u_{j_0,2^{j_0}-1})$ is a vector of scaling coefficients at level j_0, vectors $\mathbf{w}_j = (w_{j,0}, w_{j,1}, \dots, w_{j,2^j-1})$ consist of wavelet coefficients at level j for $j \geqslant j_0$, and the vector $\mathbf{w} = (\mathbf{u}_{j_0}, \mathbf{w}_{j_0}, \mathbf{w}_{j_0+1}, \dots)$ is the union of these vectors, i.e. the complete set of wavelet coefficients. The key property of this norm defined on wavelet coefficients is that if that the regularity of the wavelet function r is such that $r > s > 0$, then the Besov norm of function f is equivalent to the Besov sequence norm of its orthogonal wavelet coefficients (Donoho and Johnstone [10], Theorem 2).

3 Regularity properties, orthogonal wavelet transform.

In this section we specify the probabilistic model for wavelet coefficients of the orthogonal wavelet transform and connect it with regularity properties of the function with such coefficients, expressed in terms of the Besov norm as defined in Section 2.4. Since our model is stochastic, these results hold with probability 1.

3.1 Probabilistic model for wavelet coefficients

We model the distribution of the orthogonal wavelet coefficients as a mixture of the point mass at zero and some distribution h_j:

$$w_{jk} \sim (1 - \pi_j)\delta_0 + \pi_j h_j, \tag{9}$$

where $\tau_j > 0$, $\pi_j \in [0,1]$, wavelet coefficients w_{jk} are independent, and distribution h_j has cumulative distribution function $H_j(x) = H(x/\tau_j)$ continuous at $x = 0$ for identifiability, and that $|\operatorname{supp}(H)| = \infty$. The latter assumption implies that we consider distributions defined on at least one semiline and thus do not consider distributions with both finite end points. In the context of wavelet regression, this implies that we do not introduce any bound on the wavelet coefficients of a function. Usually one considers distributions H such that $\text{supp}(H) = \mathbb{R}$ but here technically we can relax the condition to include only one of the infinities since the results of the paper depend only on the heaviest tail of the distribution. If H has a finite support, the necessary and sufficient conditions for a function to belong to a Besov space with probability 1 will be less restrictive. Note that we do not make assumptions of symmetry or continuity of the distribution H. If distribution H has finite variance then parameter τ_j is the standard deviation otherwise it can be treated as a scaling parameter. If we need further assumptions about the distribution H later we shall add them at the relevant time.

3.2 Hyperparameters

The considered probabilistic model (9) has two parameters: the variance (or squared scaling parameter) τ_j^2 of distribution H_j and the proportion of non-zero coefficients π_j both of which depend on the decomposition level j. Due to the sparsity property of wavelet transform, i.e. that in most cases a function can be well approximated by a finite number of non-zero wavelet coefficients, we choose parameters τ_j^2 and π_j such that they tend to zero as the decomposition level j tends to infinity. We shall follow the model of Abramovich et al [1] who assumed that they decrease exponentially:

$$\tau_j^2 = 2^{-\alpha j} C_1, \quad \pi_j = \min(1, 2^{-\beta j} C_2), \tag{10}$$

where C_1 and C_2 are positive constants possibly dependent on sample size n, $\alpha \geqslant 0$, $\beta \geqslant 0$. Abramovich et al [1] took $C_1 = \tilde{C}_1/n$ due to considering the corresponding probabilistic model for the wavelet coefficients d_{jk} of the discrete wavelet transform which are related to the wavelet coefficients of the orthogonal wavelet transform by $d_{jk} \approx w_{jk}/\sqrt{n}$ (Vidakovic [24]). Pensky [20] considered parametrisation where both C_1 and C_2 are dependent on n. Here we do not specify how C_1 and C_2 depend on n since it does not affect considered Besov regularity properties of the function.

Thus, as we shall see below, only two hyperparameters, α and β, affect the regularity properties of the corresponding function. To visualise the effect of α and β on the regularity of the function, Abramovich et al [1] performed a simulation study using model (9) for different values of α and β with H being normal distribution.

It follows from the probabilistic model (9) that the number of non-zero wavelet coefficients at level j has binomial distribution with parameters 2^j and π_j. Therefore the expected number of non-zero wavelet coefficients is $2^j \pi_j$ which equals $C_2 2^{(1-\beta)j}$ at finer levels of wavelet decomposition. In case $\beta > 1$ the number $\mathcal{N}$ of the non-zero wavelet coefficients at all decomposition levels is finite almost surely because its distribution is proper (this can be shown, for instance, using the characteristic function of distribution of $\mathcal{N}$). If $\beta \leqslant 1$, the number of the non-zero wavelet coefficients is infinite. In case $\beta = 1$ it has the same distribution at different levels j, whereas in case $0 \leqslant \beta < 1$ its expected value increases as $j \to \infty$. Thus we have three different cases to consider: $\beta \in [0, 1)$, $\beta = 1$, $\beta > 1$.

We connect regularity of the function with hyperparameters α and β only. In practice, constants C_1 and C_2 can be estimated using the method of moments suggested by Abramovich et al [1] or a maximum (marginal) likelihood method of Johnstone and Silverman [13].

3.3 Model with two hyperparameters.

To study the regularity properties, we make three assumptions. The first one, assumption W, is about the regularity of the wavelet transform which is necessary to apply the wavelet representation of the Besov norm (Section 2.4). The second, assumption B, describes the probabilistic model we use for wavelet coefficients of function f. And finally, the third assumption,assumption H, is about the properties of distribution H which is part of our probabilistic model in assumption B. Different assumptions have to be made for different combinations of parameters (β, p, q).

As we discussed in the previous section, in case $\beta > 1$ the overall number of the non-zero wavelet coefficients is finite with probability one. Thus, if H is a proper distribution, a function with wavelet coefficients obeying (9) is a finite linear combination of wavelet functions with probability 1 therefore it almost surely belongs to the same Besov spaces as the wavelet function, i.e. to the Besov spaces with parameters $0 < s < r$, $1 \leqslant p, q \leqslant \infty$ where r is the regularity of the wavelet function. Further we consider cases where a function has infinite number of wavelet coefficients almost surely, i.e. cases with $0 \leqslant \beta \leqslant 1$.

Assumption W: *ψ is a wavelet function of regularity r, where $0 < s < r$.*

Assumption B: *Wavelet coefficients of function f are independent and obey the probabilistic model (9):*

$$w_{jk} \sim (1 - \pi_j)\delta_0 + \pi_j h_j(x),$$

where distribution h_j has cumulative distribution function $H_j(x) = H(x/\tau_j)$ continuous at $x = 0$, $\tau_j^2 = 2^{-\alpha j} C_1$, $\pi_j = \min(1, 2^{-\beta j} C_2)$, where C_1, C_2 are positive constants, $\alpha \geqslant 0$, $0 \leqslant \beta \leqslant 1$, $\alpha + \beta > 0$. We also assume that $|\operatorname{supp}(H)| = \infty$.

Assumption H: *Suppose ξ has distribution H.*

1. $0 \leqslant \beta < 1$, $1 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$: *assume that* $E|\xi|^p < \infty$. *If* $q < \infty$, *we also assume that* $E|\xi|^q < \infty$.
2. $0 \leqslant \beta < 1$, $p = \infty$, $1 \leqslant q \leqslant \infty$: *assume that distribution of* $|\xi|$ *has tail of one of the following types:*
   (a) $1 - H(x) + H(-x) = c_l x^{-l}[1 + o(1)]$ *as* $x \to +\infty$, $l > 0$, $c_l > 0$; *if* $q < \infty$, *assume that* $l > q$;
   (b) $1 - H(x) + H(-x) = c_m e^{-(\lambda x)^m}[1 + o(1)]$ *as* $x \to +\infty$, $m > 0$, $\lambda > 0$, $c_m > 0$.
3. $\beta = 1$, $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$: *assume that* $E|\xi|^q < \infty$.
4. $\beta = 1$, $1 \leqslant p \leqslant \infty$, $q = \infty$: *assume that* $\exists \epsilon > 0$: $E \log(|\xi|) I(|\xi| > \epsilon) < \infty$.

Note that in cases 1, 3, 4 it suffices to know the finiteness of moments of distribution H, i.e. it must have a finite absolute moment of some order

Table 1 The necessary and sufficient conditions stated in Theorem 1 ($\delta_{s\alpha} = s + 1/2 - \alpha/2$).

Parameters	$0 \leqslant \beta < 1$	$\beta = 1$
$1 \leqslant p < \infty$, $1 \leqslant q < \infty$	$\delta_{s\alpha} - \beta/p < 0$	$\delta_{s\alpha} - 1/p < 0$
$1 \leqslant p < \infty$, $q = \infty$	$\delta_{s\alpha} - \beta/p \leqslant 0$	$\delta_{s\alpha} - 1/p < 0$
$p = \infty$, $1 \leqslant q < l$ or $q = \infty$, $1 - H(x) + H(-x) \sim c_l x^{-l}$	$\delta_{s\alpha} + (1-\beta)/l < 0$	$\delta_{s\alpha} < 0$
$p = \infty$, $1 \leqslant q \leqslant \infty$ $1 - H(x) + H(-x) \sim c_m e^{-(\lambda x)^m}$	$\delta_{s\alpha} < 0$	$\delta_{s\alpha} < 0$

greater or equal to 1. But in case 2 ($p = \infty$, $0 \leqslant \beta < 1$) it is necessary to know the exact type of the heaviest of the left and right tail behaviour of the distribution. This is due to the difference in the asymptotic distribution of the maximum of a large number of independent identically distributed random variables with different tail behaviour (see proof of Theorem 3 in the next section). Also, we can see that cases $0 \leqslant \beta < 1$ and $\beta = 1$ corresponding to equal and varying number of wavelet coefficients between the levels, require different assumptions about distribution H. In case $\beta = 1$, $q = \infty$ the assumption about distribution H is the weakest.

To distinguish between cases of distributions with polynomial and exponential tails, we introduce an auxiliary variable:

$$\delta_H = \begin{cases} \frac{1-\beta}{l}, & H \quad \text{has polynomial tail and } p = \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Note, that in case $p = \infty$, constant c_m may also depend on λ.

Throughout this paper, by the statement [(either) A, or B and C] we understand the logical statement $A \vee (B \wedge C)$. Now we formulate the criterion.

Theorem 1 *Consider function f and its wavelet transform under assumptions W, B and H.*

Then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\text{either} \quad s + \frac{1}{2} - \frac{\alpha}{2} - \frac{\beta}{p} + \delta_H < 0,$$

$$\text{or} \quad s + \frac{1}{2} - \frac{\alpha}{2} - \frac{\beta}{p} = 0 \quad \text{and} \quad 0 \leqslant \beta < 1,\ 1 \leqslant p < \infty,\ q = \infty.$$

Summary of the necessary and sufficient conditions is given in Table 1. For finite p, we can rewrite the necessary and sufficient condition in the following form:

$$s + \frac{1}{2} < \frac{\alpha}{2} + \frac{\beta}{p},$$

showing that hyperparameters α and β limit the possible regularity of the function, and the contribution of the β controlling the descrease of the number of the non-zero wavelet coefficients at increasing resolution levels, is scaled by parameter p. If p is infinite, the hyperparameter β is no longer important

Table 2 The necessary and sufficient condition in case $h = t_\nu$.

Parameters	The necessary and sufficient condition
$0 \leqslant \beta < 1$, $1 \leqslant p < \nu$, $1 \leqslant q < \nu$	$s + 1/2 - \alpha/2 - \beta/p < 0$
$0 \leqslant \beta < 1$, $1 \leqslant p < \nu$, $q = \infty$	$s + 1/2 - \alpha/2 - \beta/p \leqslant 0$
$0 \leqslant \beta < 1$, $p = \infty$, $1 \leqslant q < \nu$	$s + 1/2 - \alpha/2 + (1 - \beta)/\nu < 0$
$\beta = 1$, $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \nu$	$s + 1/2 - \alpha/2 - 1/p < 0$
$\beta = 1$, $1 \leqslant p \leqslant \infty$, $q = \infty$	$s + 1/2 - \alpha/2 - 1/p < 0$

for the regularity properties in the case of a distribution with quickly decreasing tail such as power exponential. However, for a distribution with a more slowly decreasing tail such as polynomial it is still essential:

$$s + \frac{1}{2} + \frac{1}{l} < \frac{\alpha}{2} + \frac{\beta}{l},$$

where the tail parameter l now scales hyperparameter β. In order to illustrate this, we consider several examples of distribution H.

1. **Normal distribution,** $h(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$. In this case, assumption H is satisfied for all combinations of the parameters. Therefore the necessary and sufficient condition for a function with wavelet coefficients obeying the model (9) with $H(x) = \Phi(x)$ to belong to $B^s_{p,q}$ with probability 1 is

$$\begin{aligned} &s + \frac{1}{2} - \alpha/2 - \beta/p < 0, \quad \text{or} \\ &s + \frac{1}{2} - \alpha/2 - \beta/p = 0 \quad \text{and} \quad 1 \leqslant p < \infty, \quad q = \infty, \quad 0 \leqslant \beta < 1. \end{aligned}$$

   This result coincides with the one stated in Abramovich et al [1].
2. **Laplacian (double-exponential) distribution,** $h(x) = \frac{\lambda}{2} e^{-\lambda|x|}$. Since Laplacian distribution has exponential tail and all its moments are finite, the necessary and sufficient condition is the same as for the normal distribution.
3. **T distribution with ν degrees of freedom,** $h(x) = C_\nu(1+x^2/\nu)^{-\frac{\nu+1}{2}}$, $\nu \geqslant 1$. T distribution with parameter $\nu \geqslant 1$ has a polynomial tail and finite moments of order less than ν. The necessary and sufficient condition for a function to belong to the Besov spaces in terms of its wavelet coefficients is given in Table 2. Due to a smaller number of finite absolute moments, the conditions are more restrictive than for distributions with power-exponential tail. As we shall see in Section 3.5, to obtain the necessary and sufficient condition, assumption H cannot be weakened. Hence for certain values of the parameters, for example, in case $\beta < 1$ with $\nu < p < \infty$ or $\nu < q < \infty$, functions from these Besov spaces cannot have wavelet coefficients obeying the specified probabilistic model.

3.4 Model with three hyperparameters.

As we saw in the previous section, Theorem 1 implies that functions with normal or exponential tail of the distribution of its wavelet coefficients fall in exactly the same Besov spaces. Also, parameter q of the Besov spaces is not strongly related to the hyperparameters (10) of probabilistic model. In this section we introduce the third hyperparameter γ which is related to parameter q and which separates the cases of normal and exponential distributions. We keep parameter π_j the same and introduce additional factor to the parameter τ_j:

$$\tau_j^2 = j^{\gamma} 2^{-\alpha j} C_1, \quad j > 0, \quad \tau_0^2 = C_1, \tag{11}$$

where γ takes values in $\mathbb{R}$. To extend Theorem 1 to the model for the wavelet coefficients with modified representation of parameter τ_j^2, we need to replace assumption B with assumption B'.

Assumption B': *The wavelet coefficients of function f are independent and obey the probabilistic model (9) with*

$$\tau_j^2 = j^{\gamma} 2^{-\alpha j} C_1,\ j > 0,\ \tau_0^2 = C_1,$$
$$\pi_j = \min(1, 2^{-\beta j} C_2),$$

where C_1, C_2 are positive constants, $\alpha \geqslant 0$, $0 \leqslant \beta \leqslant 1$, $\alpha + \beta > 0$ and $\gamma \in \mathbb{R}$. We also assume that $|\operatorname{supp}(H)| = \infty$.

Since the necessary and sufficient condition under assumption B' is not as succinct as under assumption B, we split the general result into four theorems according to the four cases in assumption H. Obviously, the new results will include Theorem 1 as a special case with $\gamma = 0$.

To simplify the notation, we introduce the following parameter:

$$\delta = \begin{cases} s + \frac{1}{2} - \alpha/2 - \beta/p, & p < \infty, \\ s + \frac{1}{2} - \alpha/2, & p = \infty. \end{cases}$$

Now we state the theorems.

Theorem 2 *Consider the Bayesian model (9) for the orthogonal wavelet coefficients d_{jk} of the function f under assumptions W and B' with $0 \leqslant \beta < 1$, $1 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$. We also assume that $\int_{-\infty}^{+\infty} |x|^p dH(x) < \infty$ and, if $q < \infty$, that $\int_{-\infty}^{+\infty} |x|^q dH(x) < \infty$.*

Then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\delta < 0, \quad or \quad \delta = 0 \quad and \quad \begin{cases} \gamma < -2/q, & if \quad q < \infty, \\ \gamma \leqslant 0, & if \quad q = \infty; \end{cases}$$

In the next theorem we state the result for two types of distributions: those with polynomial and those with power exponential tail, although a similar result can be obtained for other distributions.

Table 3 Criterion for γ in case $\delta = 0$ ($\delta + (1-\beta)/l = 0$ in the last row).

Parameters	$1 \leqslant q < \infty$	$q = \infty$
$\beta = 1$, $1 \leqslant p \leqslant \infty$	$\gamma < -2/q$	$\gamma < 0$
$0 \leqslant \beta < 1$, $1 \leqslant p < \infty$	$\gamma < -2/q$	$\gamma \leqslant 0$
$0 \leqslant \beta < 1$, $p = \infty$, $1 - H(x) + H(-x) \sim c_l x^{-l}$	$\gamma < -2/q$	$\gamma < -2/l$
$0 \leqslant \beta < 1$, $p = \infty$, $1 - H(x) + H(-x) \sim c_m e^{-(\lambda x)^m}$	$\gamma < -2/q - 2/m$	$\gamma \leqslant -2/m$

Theorem 3 *Consider the Bayesian model (9) for the orthogonal wavelet coefficients w_{jk} of the function f under assumptions W and B' with $0 \leqslant \beta < 1$, $p = \infty$, $1 \leqslant q \leqslant \infty$.*

1) If $1 - H(x) + H(-x) = c_m e^{-(\lambda x)^m}[1 + o(1)]$ as $x \to +\infty$ for some $m > 0$, $\lambda > 0$ and constant $c_m > 0$, then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\delta < 0, \quad or \quad \delta = 0 \quad and \quad \begin{cases} \gamma < -2/m - 2/q, & if \quad q < \infty, \\ \gamma \leqslant -2/m, & if \quad q = \infty. \end{cases}$$

2) If $1 - H(x) + H(-x) = c_l x^{-l}[1 + o(1)]$ as $x \to +\infty$ for some $l > q$ if $q < \infty$ and constant $c_l > 0$ then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\delta + \frac{1-\beta}{l} < 0, \quad or \quad \delta + \frac{1-\beta}{l} = 0 \quad and \quad \begin{cases} \gamma < -2/q, & if \quad q < \infty, \\ \gamma < -2/l, & if \quad q = \infty. \end{cases}$$

Theorem 4 *Consider the Bayesian model (9) for the orthogonal wavelet coefficients w_{jk} of a function f under assumptions W and B' with $\beta = 1$, $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$. Assume also that $\int_{-\infty}^{+\infty} |x|^q dH(x) < \infty$.*

Then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\delta < 0, \quad or \quad \delta = 0 \quad and \quad \gamma < -2/q.$$

Theorem 5 *Consider the Bayesian model (9) for the orthogonal wavelet coefficients d_{jk} of the function f under assumptions W and B' with $\beta = 1$, $1 \leqslant p \leqslant \infty$, $q = \infty$.*

If $\delta \neq 0$ we assume that $\exists\, \epsilon > 0$ $\int_{|x|>\epsilon} \log(|x|) dH(x) < \infty$;

if $\delta = 0$ and $\gamma < 0$ then we assume that $\int_{-\infty}^{+\infty} |x|^{-2/\gamma} dH(x) < \infty$.

Then, for any fixed value of u_{00}, $f \in B^s_{p,q}$ almost surely if and only if

$$\delta < 0, \quad or \quad \delta = 0 \quad and \quad \gamma < 0.$$

We can see that when $\delta < 0$ the third parameter γ can take aribitrary values, however, in the boundary case when $\delta = 0$, γ is limited to negative, and in some cases, zero values (in case $\beta < 1$, $p = \infty$ and polynomial tail read $\delta + (1-\beta)/l$ instead of δ). For the summary of the values of γ in the boundary case see Table 3.

Remark 1 In the case $\beta > 1$ the number of the non-zero wavelet coefficients is finite almost surely (Section 3.2) thus the function f belongs to the same Besov spaces as the wavelet function ψ with probability 1 which are $B^s_{p,q}$ with $0 < s < r$, $1 \leqslant p, q \leqslant \infty$.

Remark 2 In all theorems stated above, we considered wavelet basis with initial level $j_0 = 0$. Note that the results of the theorems can be easily extended to other values of j_0 by fixing the vector of scaling coefficients $\mathbf{u}_{j_0} = (u_{j_0,0}, \ldots, u_{j_0,2^{j_0}-1})$ and considering the probabilistic model (9) for the wavelet coefficients w_{jk} for $j \geqslant j_0$.

In the proofs of all theorems we consider vector z_j of normalised wavelet coefficients $z_{jk} = \tau_j^{-1} w_{jk}$, $k = 0, \ldots, 2^j - 1$. Then its absolute value $|z_{jk}|$ has the distribution function $1 - \pi_j[1 - H(x) + H(-x)]$, $x \geqslant 0$. In all proofs we use the equivalence of the Besov norm of a function and the Besov norm of its wavelet coefficients (Donoho and Johnstone [10]) discussed in Section 2.4 which takes place under assumption W of the theorems. We will denote generic positive constants by C which may take different values even within a single equation.

Proof (Theorem 2)

Let ν_p be the pth absolute moment of a distribution H. Then the expression for the mean of $||z_j||_p^p$ is:

$$E||z_j||_p^p = E\left(\sum_{k=0}^{2^j-1} |z_{jk}|^p\right) = \sum_{k=0}^{2^j-1} E|z_{jk}|^p = \pi_j \nu_p 2^j = C_2 \nu_p 2^{(1-\beta)j}.$$

Now we show that $2^{-(1-\beta)j}||z_j||_p^p$ converges to $C_2\nu_p$ almost surely as $j \to \infty$ using the first Borel-Cantelli lemma (Billingsley [3]). Given $\epsilon > 0$, Markov inequality ($P\{|X| > \epsilon\} \leqslant E|X|/\epsilon$) implies that

$$\sum_{j=0}^{\infty} P\left\{|2^{-(1-\beta)j}||z_j||_p^p - C_2\nu_p| > \epsilon\right\} \leqslant O(1)\epsilon^{-1} \sum_{j=0}^{\infty} 2^{-(1-\beta)j} < \infty,$$

which, by the first Borel-Cantelli lemma, is equivalent to

$$2^{-(1-\beta)j}||z_j||_p^p \to C_2\nu_p \quad \text{almost surely} \quad \text{as} \quad j \to \infty.$$

For a finite q the Besov norm of wavelet coefficients can be represented in terms of $||z_j||_p$:

$$\begin{aligned} ||w||_{b^s_{p,q}} &= |u_{00}| + \sum_{j=0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |z_{jk}|^p \tau_j^p\right)^{q/p} \\ &= |u_{00}| + \sum_{j=0}^{\infty} j^{\gamma q/2} 2^{jq(s'-\alpha/2+(1-\beta)/p)} [2^{-(1-\beta)j}||z_j||_p]^q, \end{aligned}$$

which is finite almost surely if and only if the series $\sum_{j=0}^{\infty} j^{\gamma q/2} 2^{jq(s'-\alpha/2+(1-\beta)/p)}$ is finite, since

$$E||z_j||_p^q \leqslant CE||z_j||_q^q = C2^j E|z_{jk}|^q < \infty$$

by statement 2 of Lemma 1 below and the assumptions of the theorem. The summand of the series is $j^{\gamma q/2} 2^{jq(s+1/2-\alpha/2-\beta/p)}$ therefore the series converges if and only if either $s+1/2-\alpha/2-\beta/p < 0$, or $s+1/2-\alpha/2-\beta/p = 0$ and $\gamma < -2/q$.

For infinite q the Besov norm of wavelet coefficients is represented in terms of $||z_j||_p$ in the following way:

$$\begin{aligned} ||w||_{b^s_{p,\infty}} &= |u_{00}| + \sup_{j\geqslant 0} \left\{ 2^{js'} \left(\sum_{k=0}^{2^j-1} |w_{jk}|^p \right)^{1/p} \right\} \\ &= |u_{00}| + \sup_{j\geqslant 0} \left\{ j^{\gamma/2} 2^{j(s+1/2-\alpha/2-\beta/p)} 2^{-j(1-\beta)} ||z_j||_p \right\}. \end{aligned}$$

The supremum is finite almost surely if and only if $j^{\gamma q/2} 2^{j(s+1/2-\alpha/2-\beta/p)}$ is finite for $j \geqslant 0$ and $j \to \infty$ which holds if and only if either $s + 1/2 - \alpha/2 - \beta/p < 0$, or $s + 1/2 - \alpha/2 - \beta/p = 0$ and $\gamma \leqslant 0$.

Now we apply the result of Donoho and Johnstone [10] that under the assumptions of the theorem the finiteness of the Besov norm of the wavelet coefficients is equivalent to the finiteness of the Besov norm of the function.

Proof (Theorem 3.)

The Besov sequence norm in case $p = \infty$ is expressed in terms of the following random variable: $\xi_j = ||z_j||_\infty = \max_{k=0,\dots,2^j-1}(|z_{jk}|)$. The distribution function of $|z_{jk}|$ is

$$P\{|z_{jk}| < x\} = 1 - \pi_j + \pi_j H(x) - \pi_j H(-x) = 1 - \pi_j[1 - H(x) + H(-x)],$$

$x > 0$, and the distribution function of ξ_j as $j \to \infty$ for a fixed $x > 0$ is

$$\begin{aligned} F_{\xi_j}(x) &= [P\{|z_{jk}| < x\}]^{2^j} = (1 - \pi_j[1 - H(x) + H(-x)])^{2^j} \\ &= \exp\{-C_2 2^{(1-\beta)j}[1 - H(x) + H(-x)]\} \left[1 + O(1)2^{(1-2\beta)j}\right]. \end{aligned}$$

To find the asymptotic distribution of ξ_j as $j \to \infty$ it appears that we cannot apply Extreme Value Theory directly because the distribution of z_{jk} depends on j. Nevertheless the proof of the Theorem 1.6.2 (Leadbetter et al [14], p.17) stating the asymptotic distribution of the maximum under different conditions remains valid in the case $F_{|z_{jk}|}(x) = 1 - C_2 2^{-\beta j}[1 - H(x) + H(-x)]$ (for sufficiently large j) because $F_{|z_{jk}|}(x)$ depends on j monotonically.

We study the convergence of the sequence norm of the wavelet coefficients separately for distributions with power exponential and polynomial tails.

1. **The power exponential tail.** For this type of distributions $1 - H(x) + H(-x) = c_m e^{-(\lambda x)^m}[1 + o(1)]$ the asymptotic distribution of $||z_j||_\infty$ is

$$P\{||z_j||_\infty > a_j x + b_j\} \to 1 - \exp\{-e^{-x}\},$$

and the constants can be chosen in the following way (Leadbetter et al [14]):

$$a_j = [j(1-\beta)\log 2]^{\frac{1}{m}-1}(m\lambda)^{-1},$$
$$b_j = [j(1-\beta)\log 2]^{\frac{1}{m}}\lambda^{-1}.$$

If we show that for any $\epsilon > 0$,

$$\sum_{j=0}^{\infty} P\left\{\left|\frac{||z_j||_\infty}{b_j} - 1\right| > \epsilon\right\} \tag{12}$$

is finite, then according to the first Borel-Cantelli lemma, it is equivalent to $||z_j||_\infty/b_j \to 1$ almost surely. The asymptote of the summands in (12) is:

$$\begin{aligned}P\{|||z_j||_\infty/b_j - 1| > \epsilon\} &= P\{|||z_j||_\infty - b_j| > \epsilon b_j\}\\ = \exp\{-e^{\epsilon b_j/a_j}\}[1+o(1)] &+ 1 - \exp\{-e^{-\epsilon b_j/a_j}\}[1+o(1)]\\ = \exp\{-e^{\epsilon c j}\}[1+o(1)] &+ \exp\{-\epsilon c j\}[1+o(1)],\end{aligned}$$

which tends to zero as $j \to \infty$ and whose series converges. Therefore $\frac{||z_j||_\infty}{b_j} \to 1$ holds almost surely.

Hence the necessary and sufficient condition for f to belong to $B^s_{p,q}$ almost surely is the finiteness of the series $\sum_{j=1}^{\infty} 2^{js'q}\left(j^{\gamma/2}2^{-j\alpha/2}j^{1/m}\right)^q$ if $1 \leqslant q < \infty$, and in case $q = \infty$, the finiteness of $\sup_{j\geqslant 1}\{j^{\gamma/2}2^{j\delta}j^{1/m}\}$.

Therefore the necessary and sufficient condition is

$$\begin{aligned}1 \leqslant q < \infty: &\quad \delta < 0, \quad \text{or} \quad \delta = 0 \text{ and } \gamma < -2/m - 2/q,\\ q = \infty: &\quad \delta < 0, \quad \text{or} \quad \delta = 0 \text{ and } \gamma \leqslant -2/m.\end{aligned}$$

2. **Distribution with polynomial tail.** For this type of distributions $1 - H(x) + H(-x) = c_l x^{-l}[1+o(1)]$ the asymptotic distribution is

$$P\{||z_j||_\infty > b_j x\} \to 1 - \exp\{-x^{-l}\}, \quad x > 0,$$

and the constant can be chosen as $b_j = B2^{(1-\beta)j/l}$ for large j (see Leadbetter et al [14]) where $B = (c_l C_2)^{1/l}$.

Therefore the asymptotic distribution of $\zeta_j = ||z_j||_\infty/b_j$ is Type 2 of the Extreme Value Distributions. In order to show that in case $1 \leqslant q < l$ the sum $\sum_{j=1}^{\infty} 2^{js'q}\left(j^{\gamma/2}2^{-j\alpha/2}2^{(1-\beta)j/l}\zeta_j\right)^q$ converges almost surely, we apply the three series theorem together with the theorem of monotone convergence. Since ζ_j^q are independent non-negative random variables with the same asymptotic distribution and positive finite mean (due to Lemma 2 in the appendix and $q < l$), the convergence of the above sum with probability 1 is equivalent to convergence of the sum $\sum_{j=1}^{\infty} j^{q\gamma/2}2^{jq[\delta+(1-\beta)/l]}$ which converges if and only if

$$\begin{aligned}&\delta + (1-\beta)/l < 0, \quad \text{or}\\ &\delta + (1-\beta)/l = 0 \quad \text{and} \quad \gamma < -2/q.\end{aligned}$$

In case $q=\infty$ we need to show that $\sup_{j\geqslant 1}\{j^{\gamma/2}2^{j\delta}2^{(1-\beta)j/l}\zeta_j\}<\infty$ almost surely. The asymptotic distribution of $\max_{1\leqslant j\leqslant n}\{j^{\gamma/2}2^{j\delta}2^{(1-\beta)j/l}\zeta_j\}$ as $n\to\infty$ is the Extreme Value Distribution of the second type with its distribution function being:

$$\prod_{j=1}^{n}\exp\{-(j^{-\gamma/2}2^{-j\delta}2^{-(1-\beta)j/l}x)^{-l}\}=\exp\left\{-x^{-l}\sum_{j=1}^{n}j^{l\gamma/2}2^{j(l\delta+1-\beta)}\right\},$$

$x>0$. This asymptotic distribution is proper if and only if the sum $\sum_{j=1}^{n}j^{l\gamma/2}2^{j(l\delta+1-\beta)}$ converges as $n\to\infty$ i.e. if and only if

$$\begin{array}{l}\delta+(1-\beta)/l<0,\quad \text{or}\\ \delta+(1-\beta)/l=0\quad\text{and}\quad\gamma<-2/l.\end{array}$$

Since the Besov norm of a function and the Besov norm of its wavelet coefficients under assumption W are equivalent (Donoho and Johnstone [10]), the statement of the theorem is proved.

In order to prove Theorem 4 we need the following lemma.

Lemma 1 *Consider the model (9) with $\beta=1$, $1\leqslant p\leqslant\infty$. Define random variables $\zeta_{jk}\sim H$, $k=1,\dots,N_j$, where $N_j\sim Pois(C_2)$ is a Poisson random variable, $0<C_2<\infty$. Then the following statements hold:*

1. *For $\forall m>0$ and sufficiently large j, $||\zeta_j||_m\stackrel{D}{=}||z_j||_m$.*
2. *For $\forall m>0$ and sufficiently large j, $E|z_{jk}|^m<\infty\quad\Leftrightarrow\quad E||z_j||_p^m<\infty$.*

By $\xi_1\stackrel{D}{=}\xi_2$ we denote the equality of distributions of ξ_1 and ξ_2. In case $m<1$, $||z_j||_m$, strictly speaking, is not a norm but we use the same notation as in the case $m\geqslant 1$ for convenience.

Proof (Lemma 1)

1. To prove this part of the lemma, we use the argument of Abramovich et al [1]. Our model for wavelet coefficients assumes that with probability $1-\pi_j$ a wavelet coefficient is zero and with probability π_j it is non-zero, the coefficients are independent and there are 2^j of them at level j. So the number M_j of non-zero wavelet coefficients at level j is binomial with parameters 2^j and π_j. For sufficiently large j, $M_j=N_j$ almost surely where N_j is a Poisson random variable with parameter C_2. Therefore N_j has the same distribution for every j. Thus if ζ_j is a vector of N_j independent random variables with distribution H which is independent of j, norms of ζ_j and z_j are equal almost surely:

$$||z_j||_m^m=\sum_{k=0}^{2^j-1}|z_{jk}|^m=\sum_{l=1}^{M_j}|z_{jk}^{(l)}|^m\stackrel{D}{=}\sum_{l=1}^{N_j}|z_{jk}^{(l)}|^m\stackrel{D}{=}\sum_{l=1}^{N_j}|\zeta_{jl}|^m=||\zeta_j||_m^m,$$

where $z_{jk}^{(l)}$ are non-zero normalised wavelet coefficients.

2. Since $P\{N_j < \infty\} = 1$, for $N_j > 0$ we can apply Lemma 2 to norms of ζ_{jk}, implying that $||\zeta_j||_l \leqslant ||\zeta_j||_v \leqslant N_j^{\frac{1}{v}-\frac{1}{l}} ||\zeta_j||_l$. Since for $N_j = 0$ $||\zeta_j||_v = 0$ and thus the inequality is trivial, taking $l = p$, $v = m$ for $p > m$ and $l = m$, $v = p$ for $p < m$, we have that $E|\zeta_{jk}|^m < \infty \Leftrightarrow E||\zeta_j||_p^m < \infty$. According to the first statement of the lemma, for sufficiently large j, norms $||\zeta_j||_m$ and $||z_j||_m$ have the same distribution, therefore finiteness of the mth absolute moment of z_{jk} is equivalent to $E||\zeta_j||_p^m < \infty$, $1 \leqslant p < \infty$. For $m = p$ the statement is trivial.

Proof (Theorem 4.) By the equivalence of the norms given by equation (7), $f \in B_{p,q}^s$ almost surely if and only if

$$\begin{aligned} ||w||_{b_{p,q}^s} &= |u_{00}| + \sum_{j=0}^{\infty} 2^{js'q} \tau_j^q ||z_j||_p^q \\ &\stackrel{D}{=} |u_{00}| + \sum_{j=0}^{\infty} j^{\gamma q/2} 2^{jq(s'-\alpha/2)} ||\zeta_j||_p^q < \infty \end{aligned} \tag{13}$$

with probability 1, where ζ_j is a vector of N_j independent identically distributed random variables with distribution H as defined in Lemma 1, and N_j has Poisson distribution with parameter C_2. To find the condition of convergence of the norm (13) we use the property following from the monotone convergence and the three series theorems that if Z_n are independent and identically distributed non-negative random variables with strictly positive finite mean, and a_n are non-negative constants, then $\sum a_n Z_n$ is convergent almost surely if and only if $\sum a_n$ is convergent. Therefore, the finiteness of expectation $E||\zeta_j||_p^q$, according to Lemma 1, is equivalent to finiteness of $E|\zeta_{jk}|^q$ which holds under the assumptions of the theorem. Thus, condition (13) is equivalent to condition $\sum_{j=1}^{\infty} 2^{j\delta q} j^{\gamma q/2} < \infty$. This condition is satisfied if and only if either $\delta < 0$, or $\delta = 0$ and $\gamma < -2/q$. Thus, the theorem is proved.

Proof (Theorem 5.) In this case $f \in B_{p,\infty}^s$ almost surely if and only if

$$\sup_{j \geqslant 1} \{2^{js'} \tau_j ||\zeta_j||_p\} < \infty,$$

which is equivalent to $\sup_{j\geqslant 1}\{2^{j\delta} j^{\gamma/2} ||\zeta_j||_p\} < \infty$, where ζ_j is as defined in Lemma 1. Appealing to Borel-Cantelli lemmas we deduce that the former condition holds if and only if there exists a constant $c > 0$ such that

$$\sum_{j=1}^{\infty} P\left\{2^{j\delta} j^{\gamma/2} ||\zeta_j||_p \geqslant c\right\} < \infty.$$

Since the random variables $||\zeta_j||_p$ are independent and identically distributed random variables, the condition above can be rewritten as

$$\exists c > 0 : \quad \sum_{j=1}^{\infty} P\{\eta \geqslant c 2^{-j\delta} j^{-\gamma/2}\} < \infty, \tag{14}$$

where random variable η has the same distribution as $||\zeta_j||_p$. To obtain a necessary and sufficient condition for convergence, we use Lemma 3 (see appendix) considering separately two cases: $\delta = 0$ and $\delta \neq 0$. In the former case, (14) holds if and only if $\gamma < 0$ and $E\eta^{-2/\gamma} < \infty$. If $\delta = 0$ and $\gamma \geqslant 0$, for any $c > 0$

$$\sum_{j=1}^{\infty} P\{\eta \geqslant c j^{-\gamma/2}\} \geqslant \sum_{j=1}^{\infty} P\{\eta \geqslant c\} = \infty,$$

i.e. convergence does not hold for any finite constant c for $\gamma \geqslant 0$. Note that if H had a finite support, the convergence could take place for $\gamma = 0$. Since η has the same distribution as $||\zeta_j||_p$, the necessary and sufficient condition for $E||\zeta_j||_p^{-2/\gamma}$ to be finite is $\gamma < 0$ and $E|z_{jk}|^{-2/\gamma} < \infty$ (Lemma 1).

In case $\delta \neq 0$, in order to apply Lemma 3, we divide the sum (14) into two:

$$\begin{aligned}
&\sum_{j=1}^{\infty} P\left\{\log\eta - \log c \geqslant -\delta \log 2j \left(1 + \frac{\gamma}{-2\delta\log 2}\frac{\log j}{j}\right)\right\} \\
&= \sum_{j=1}^{\tilde{j}} P\left\{\log\eta - \log c \geqslant -\delta \log 2j \left(1 + \frac{\gamma}{-2\delta\log 2}\frac{\log j}{j}\right)\right\} \\
&+ \sum_{j=\tilde{j}+1}^{\infty} P\left\{(\log\eta - \log c) I\left(\log\eta - \log c > 0\right) > -\delta\log 2j\right\},
\end{aligned}$$

which converges for some c if and only if $\delta < 0$ and $\exists c > 0$ such that $E(\log\eta - \log c) I\left(\log\eta - \log c > 0\right) < \infty$, or equivalently, $E\log\frac{\eta}{c} I\left(\eta > c\right) < \infty$. Note that the random variable under the expectation is non-negative with probability 1. Since we assumed the support of H to be infinite, for $\delta > 0$ convergence does not hold for any finite $c > 0$ since

$$\begin{aligned}
\sum_{j=\tilde{j}+1}^{\infty} P\left\{\log\frac{\eta}{c} I\left(\log\frac{\eta}{c} > 0\right) > -\delta\log 2j\right\} &\geqslant \sum_{j=\tilde{j}+1}^{\infty} P\left\{\log\frac{\eta}{c} I\left(\log\frac{\eta}{c} > 0\right) > 0\right\} \\
&= \sum_{j=\tilde{j}+1}^{\infty} P\left\{\eta > c\right\} = \infty,
\end{aligned}$$

where $\tilde{j} < \infty$ is such that for $j > \tilde{j}$, $1 + \frac{\gamma}{-2\delta\log 2}\frac{\log j}{j} > 0$.

It is easy to show that under the assumptions of the theorem this condition holds for $p = \infty$:

$$
\begin{aligned}
& E \log(||\zeta_j||_\infty) I\left(\log(||\zeta_j||_\infty) > \log c\right) \\
&= E \log(\max_{k=1\dots N_j} \{|\zeta_{jk}|\}) I\left(\max_{k=1\dots N_j} \{|\zeta_{jk}|\} > c\right) \\
&= E \max_{k=1\dots N_j} \{\log(|\zeta_{jk}|)\} I\left(\max_{k=1\dots N_j} \{|\zeta_{jk}|\} > c\right) I(N_j > 0) \\
&\leqslant E \sum_{k=1}^{N_j} \log(|\zeta_{jk}|) I\left(|\zeta_{jk}|/c > 1\right) I(N_j > 0) \\
&= E(N_j \mid N_j > 0) E \log(|\zeta_{jk}|) I\left(|\zeta_{jk}| > c\right) < \infty,
\end{aligned}
$$

since $\cap_{k=1}^{N_j} \{|\zeta_{jk}| > c\} \subset \{|\zeta_{jk}| > c\}$ for any k and $E \log(|\zeta_{jk}|) I\left(|\zeta_{jk}| > c\right) < \infty$ for some $c > 0$ is one of the assumptions of the theorem.

For $1 \leqslant p < \infty$, we can use the result for $p = \infty$ to show that, similarly,

$$
\begin{aligned}
& E\left[\log\left(\frac{||\zeta_j||_p}{c}\right)\right] I\left(\log ||\zeta_j||_p > \log c\right) \\
&= \frac{1}{p} E \log\left(\sum_{k=1}^{N_j}\left(\frac{|\zeta_{jk}|}{c}\right)^p\right) I\left(\sum_{k=1}^{N_j} |\zeta_{jk}|^p > c^p\right) \\
&\leqslant E \log\left(N_j^{1/p} \max_{k=1,\dots,N_j}\left\{\frac{|\zeta_{jk}|}{c}\right\}\right) I\left(|\zeta_{jk}|^p > c^p,\, k = 1\dots, N_j\right) I(N_j > 0) \\
&= E \log\left[\max_{k=1,\dots,N_j}\left\{\frac{|\zeta_{jk}|}{c}\right\} I\left(|\zeta_{jk}| > c\right)\right] I(N_j > 0) + \frac{1}{p} E \log N_j I(N_j > 0) \\
&< \infty
\end{aligned}
$$

under assumption of the theorem that $E \log(|\zeta_{jk}|) I\left(|\zeta_{jk}| > c\right) < \infty$ for some $c > 0$.

Therefore the $b^s_{p,q}$ norm of the wavelet coefficients is finite if and only if either $\delta < 0$ or $\delta = 0$, $\gamma < 0$. Therefore, the theorem is proved.

3.5 Sufficiency of assumption H.

For each of the theorems above we made assumptions about distribution H necessary to obtain Besov membership (with probability 1). Now we study what happens if assumption H fails.

1. Theorem 2. In the case $1 \leqslant p < \infty$, $0 \leqslant \beta < 1$ we assume the existence of the integral $\int_{-\infty}^{+\infty} |x|^p dH(x) < \infty$. We use it to prove that the strong law of large numbers holds for the sequence $\{|z_{jk}|^p\}_{k=0}^{2^j-1}$, i.e. that $2^{-j(1-\beta)}||z_j||_p^p \to c > 0$ almost surely where the existence of $E|z_{jk}|^p$ is essential. According to Feller [11] (Theorem 4, p.241), existence of the expectation is not only sufficient for the strong law of large numbers but is

also necessary otherwise for any numerical sequence $\{c_j\}$ with probability one

$$\limsup_{j\to\infty} |2^{-j(1-\beta)}||z_j||_p^p - c_j| = \infty.$$

Therefore this assumption cannot be weakened.
In case $1 \leqslant q < \infty$, we also assume that $\int_{-\infty}^{+\infty} |x|^q dH(x) < \infty$. This assumption is required for convergence of the series $\sum_{j=0}^{\infty} a_j X_j$ with $X_j = [2^{-(1-\beta)j}||z_j||_p]^q$ and $a_j = j^{\gamma q/2} 2^{jq(s'-\alpha/2+(1-\beta)/p)}$.
If the integral is infinite, for any sequence $a_j > 0$ the series $\sum_{j=0}^{\infty} a_j X_j$ has infinite expectation, since X_j are independent identically distributed random variables with infinite expectation: $EX_j = \infty$ and thus the series is infinite with probability 1. Therefore this assumption cannot be weakened.

2. Theorem 3, $p = \infty$, $0 \leqslant \beta < 1$. We consider two types of tail of distribution H: power exponential and polynomial with no additional assumptions.
3. Theorem 4. In the case $1 \leqslant q < \infty$, $\beta = 1$ we assume the existence of the integral $\int_{-\infty}^{+\infty} |x|^q dH(x) < \infty$ which cannot be weakened (see the case $q < \infty$ for Theorem 2).
4. Theorem 5, $\beta = 1$, $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$. The following assumptions are made:
   – if $\delta \neq 0$: $\exists \epsilon > 0$ such that $E \log(|\xi|) I(|\xi| > \epsilon) < \infty$;
   – if $\delta = 0$ and $\gamma < 0$: $E|\xi|^{-2/\gamma} < \infty$,

   where ξ has distribution H. In case $\delta = 0$ and $\gamma < 0$ the assumption is necessary and sufficient due to Lemma 3 given in the appendix.
   In the case $\delta \neq 0$, as can be seen from the proof, condition $E \log(|\xi|) I(|\xi| > c) < \infty$ implies that $E \log(||\zeta_j||_p) I(||\zeta_j||_p > c) < \infty$. Now we show that the reverse holds. In case $p = \infty$ we have:

$$\begin{aligned}
&E \log(||\zeta_j||_\infty) I\left(\log ||\zeta_j||_\infty > \log c\right) \\
&= E \log(\max_{k=1,\ldots,N_j} \{|\zeta_{jk}|\}) I\left(\max_{k=1,\ldots,N_j} \{|\zeta_{jk}|\} > c\right) I(N_j > 0) \\
&= E\left(\max_{k=1,\ldots,N_j} \{\log |\zeta_{jk}| I\left(|\zeta_{jk}| > c\right)\} \mid N_j > 0\right) P\{N_j > 0\} \\
&\geqslant E \log |\zeta_{j1}| I\left(|\zeta_{j1}| > c\right) P\{N_j > 0\},
\end{aligned}$$

since the expectation given $N_j = 0$ is zero and $E \max(\xi_1, \xi_2) \geqslant E\xi_1$. Therefore, the equivalence holds for $p = \infty$.
For $1 \leqslant p < \infty$, since $N_j < \infty$ almost surely, inequality $||\zeta_j||_p \geqslant ||\zeta_j||_\infty$ holds stochastically by Lemma 2 (see appendix) and thus, since $\forall c > 0$ function $\log(x) I(x > c)$ increases monotonically, inequality

$$E \log(||\zeta_j||_p) I(||\zeta_j||_p > c) \geqslant E \log(||\zeta_j||_\infty) I(||\zeta_j||_\infty > c)$$

also holds, implying that

$$E \log(||\zeta_j||_p) I(||\zeta_j||_p > c) \geqslant E \log |\zeta_{j1}| I\left(|\zeta_{j1}| > c\right) P\{N_j > 0\}.$$

Thus, if $E \log |\zeta_{jk}| I\,(|\zeta_{jk}| > c) = \infty$ for all $c > 0$, $E \log(||\zeta_j||_p) I(||\zeta_j||_p > c) = \infty$. Since ζ_j are independent, the second Borel–Cantelli lemma implies that for the considered set of parameters the Besov sequence norm is infinite, and thus the necessary and sufficient condition does not hold.

Therefore, the assumptions about distribution H cannot be weakened for Theorems 2, 4 and 5. Theorem 3 can be extended to other types of monotonic tail behaviour of distribution H.

3.6 Some particular cases.

Remark 3 1. Only in the case $p = \infty$, $0 \leqslant \beta < 1$ (Theorem 3 with polynomial tail), parameters of the tail behaviour of the distribution H are explicitly related to the condition on the parameters of the Besov spaces.

2. Each of Theorems 2, 4 and 5 states the same result for distributions with all finite absolute moments.

Introduction of the third parameter γ allows to distinguish between the regularity properties of functions whose wavelet coefficients have all finite absolute moments but different tails, for example, between distributions with normal and exponential tails which we cannot do using Theorem 1. Therefore if we want to compare the Besov spaces for functions with normal and Laplacian distributions of wavelet coefficients the only difference is in the case $p = \infty$, $\beta < 1$. The statement for these special cases is given below.

Proposition 1 *Let $0 \leqslant \beta < 1$, $p = \infty$, $1 \leqslant q \leqslant \infty$. Consider the Bayesian model (9) for the wavelet coefficients w_{jk} of the function f with assumptions W and B'.*

1. H **is normal distribution.** *For any fixed value of u_{00}, $f \in B^s_{\infty,q}$ if and only if*

$$\begin{aligned} 1 \leqslant q < \infty: &\quad \textit{either} \quad \delta < 0, \quad \textit{or} \quad \delta = 0 \quad \textit{and} \quad \gamma < -2/q - 1;\\ q = \infty: &\quad \textit{either} \quad \delta < 0, \quad \textit{or} \quad \delta = 0 \quad \textit{and} \quad \gamma \leqslant -1. \end{aligned}$$

This coincides with Theorem 2 stated in Abramovich et al [1].

2. H **is Laplacian distribution.** *For any fixed value of u_{00}, $f \in B^s_{\infty,q}$ if and only if*

$$\begin{aligned} 1 \leqslant q < \infty: &\quad \textit{either} \quad \delta < 0, \quad \textit{or} \quad \delta = 0 \quad \textit{and} \quad \gamma < -2/q - 2\\ q = \infty: &\quad \textit{either} \quad \delta < 0, \quad \textit{or} \quad \delta = 0 \quad \textit{and} \quad \gamma \leqslant -2. \end{aligned}$$

We can see that for the same values of the hyperparameters α, β, γ functions with the normal model of wavelet coefficients belong to a wider class of Besov spaces compared to the functions with Laplacian model. Therefore, if we want to span a larger set of functions (for fixed values of the hyperparameters) we need to choose a lighter tail.

4 Regularity properties, continuous wavelet transform.

In this section we extend the result stated in Section 3.3 to the continuous wavelet transform.

4.1 Probabilistic model.

We assume that wavelet and scaling functions ψ and ϕ have compact support $[0, 1]$ under periodic boundary condition and are of regularity r. We model a function f on $[0, 1]$ as a sum of high and low frequency components f_w and f_0:

$$f(x) = f_0(x) + f_w(x) = \sum_{i=1}^{M} \eta_{\lambda_i} \phi_{\lambda_i}(x) + \sum_{\lambda \in S} \omega_\lambda \psi_\lambda, \tag{15}$$

where $\psi_\lambda(x) = a^{1/2}\psi(a(x - b))$, $\lambda = (a, b)$, $a \geqslant a_0 = 2^{j_0}$, $b \in [0, 1]$, and similarly, $\phi_\lambda(x) = a^{1/2}\phi(a(x-b))$, $M < \infty$ and λ_i are such that $a_i \leqslant a_0 = 2^{j_0}$, $b_i \in [0, 1]$ where a_0 is at least twice the length of the support of functions ψ and ϕ. The coarse component of the function, f_0, is considered to be a finite linear combination of scaling functions with real-valued coefficients η_{λ_i}. For the high frequency component, we adapt the probabilistic model (9) used for the wavelet coefficients of the orthogonal wavelet transform to the continuous wavelet coefficients ω_λ and the set of their indices S.

Defining set S of wavelet indices $\lambda = (a, b) \in [a_0, \infty) \times [0, 1]$ corresponds to selecting a set of indices where the wavelet coefficients are non-zero. Here it is modelled as a Poisson process with intensity $\mu(\lambda) = C_\mu a^{-\beta}$, $\beta \geqslant 0$, $C_\mu > 0$. The intensity of the Poisson process $\mu(\lambda)$ determines the number of elements of the process around location λ. This is an analogue of the parameter π_j we used for the orthogonal model (9) which is the proportion of non-zero wavelet coefficients at resolution level j. Similarly to the case of the orthogonal wavelet transform, we assume that the intensity, and thus the proportion of the non-zero wavelet coefficients, is independent of the shifting parameter b and decreases as the scaling parameter a increases.

Since the distribution of the wavelet coefficients ω_λ depend on the Poisson process S, we model the distribution of wavelet coefficients ω_λ conditioned on the Poisson process S in the similar way as the distribution of non-zero orthogonal wavelet coefficients. Therefore we assume that $\omega_\lambda \mid S$ are independent and have distribution H_λ:

$$\omega_\lambda | S \sim H_\lambda(x), \tag{16}$$

where $H_\lambda(x) = H(x/\tau_\lambda)$ is a distribution function, continuous at $x = 0$, $\tau_\lambda^2 = C_\tau a^{-\alpha}$, $\alpha \geqslant 0$, $C_\tau > 0$.

Hyperparameters α and β have the same interpretation as their counterparts in the model for the orthogonal wavelet coefficients in Section 3.2.

4.2 Link between orthogonal and continuous wavelet coefficients.

Since the equivalence between the Besov norm of function f and its wavelet transform is given in terms of the orthogonal wavelet $w_{jk} = \langle \psi_{jk}, f \rangle$ and scaling $u_{j_0 k} = \langle \phi_{j_0 k}, f \rangle$ coefficients, in order to prove the finiteness of the Besov norm we need to express the orthogonal wavelet coefficients in terms of known coefficients ω_λ:

$$w_{jk} = \langle \psi_{jk}, f \rangle = \sum_{\lambda \in S} \langle \psi_{jk}, \psi_\lambda \rangle \omega_\lambda + \sum_{i=1}^{M} \eta_{\lambda_i} \langle \psi_{jk}, \phi_{\lambda_i} \rangle = \sum_{\lambda \in S} K(\lambda, \lambda_{jk}) \omega_\lambda,$$

where $K(\lambda, \lambda') = \langle \psi_\lambda, \psi_{\lambda'} \rangle$ is the reproducing kernel of the wavelet transform defined in Section 2.2, and $\psi_{jk} = \psi_{\lambda_{jk}} = \psi_{(2^j, k2^{-j})}$. Note that $\langle \psi_\lambda, \phi_{\lambda_i} \rangle = 0$ for $\lambda \in S$ and considered λ_i since $\psi_\lambda \in W_{j_0} \perp V_{j_0} \ni \phi_{\lambda_i}$. If we introduce another kernel $W(\lambda, \lambda') = \langle \phi_\lambda, \psi_{\lambda'} \rangle$ we can write a similar expression for $u_{j_0 k}$:

$$u_{j_0 k} = \langle \phi_{j_0 k}, f \rangle = \sum_{\lambda \in S} W(\lambda, \lambda_{j_0 k}) \omega_\lambda + \sum_{i=1}^{M} \eta_{\lambda_i} \langle \phi_{\lambda_i}, \phi_{j_0 k} \rangle.$$

The second summand is a constant we denote by C_w. Therefore the orthogonal wavelet w_{jk} and scaling $u_{j_0 k}$ coefficients can be expressed in terms of random wavelet coefficients ω_λ in the following way:

$$\begin{aligned} w_{jk} &= \sum_{\lambda \in S} K(\lambda, \lambda_{jk}) \omega_\lambda, \\ u_{j_0 k} &= \sum_{\lambda \in S} W(\lambda, \lambda_{j_0 k}) \omega_\lambda + C_w. \end{aligned} \tag{17}$$

Properties of kernels K have been discussed in Section 2.2. Since kernel W is based on the scaling function ϕ which has the same regularity properties as the wavelet function ψ, it can be bounded in the same way as kernel K (e.g., take $f = \phi$ in (6)).

4.3 Regularity properties.

In this section we connect the hyperparameters of the model for wavelet coefficients α and β with parameters of the Besov spaces p and s. As in the case of the orthogonal wavelet transform, we split the necessary assumptions into three groups.

Assumption W: *ψ is a compact-supported wavelet function of regularity r, $\psi^{(r)} \in C^\rho$, where $\rho \in (0,1)$ is the exponent of Hölder continuity, such that $0 < s < r$ and $2(r+\rho) > 1 + \alpha$.*

We need the regularity of the wavelet function to be greater than the parameter s of the Besov space in order to be able to use the equivalence of the Besov sequence norm $b^s_{p,q}$ on wavelet coefficients and the Besov norm of the function with these coefficients. We can see that this assumption includes assumption W from Section 3.3, with addition of an extra regularity condition on the wavelet function which is due to complexity of the wavelet indices.

Assumption B: *We use the model for wavelet coefficients:*

$$\omega_\lambda | S \sim H_\lambda(x), \tag{18}$$

where S is a Poisson process on $\lambda = (a,b) \in [a_0, \infty) \times [0,1]$, $a_0 = 2^{j_0} \geqslant 2\max\{|\operatorname{supp}(\psi)|, |\operatorname{supp}(\phi)|\}$, with intensity $\mu(\lambda) = C_\mu a^{-\beta}$, $0 \leqslant \beta \leqslant 1$, and $H_\lambda(x) = H(x/\tau_\lambda)$ has finite variance $\tau_\lambda^2 = C_\tau a^{-\alpha}$, $\alpha \geqslant 0$, $\beta + \alpha > 0$. We also assume that $|\operatorname{supp}(H)| = \infty$.

Assumption H:

1. $0 \leqslant \beta < 1$, $1 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$*: assume that distribution H has the absolute moment of order p; if $q < \infty$, we also assume that distribution H has the absolute moment of order q.*
2. $0 \leqslant \beta < 1$, $p = \infty$, $1 \leqslant q \leqslant \infty$*: assume that distribution H has power exponential tail, i.e. $1 - H(x) + H(-x) = c_m e^{-(\nu x)^m}[1 + o(1)]$ as $x \to +\infty$, $m > 0$, $\nu > 0$, $c_m > 0$.*
3. $\beta = 1$, $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$*: assume that distribution H has the absolute moment of order q.*
4. $\beta = 1$, $1 \leqslant p \leqslant \infty$, $q = \infty$*: assume that exists $\epsilon > 0$ such that $\int_{|x|>\epsilon} \log(|x|) dH(x) < \infty$.*

The assumptions about distribution H are almost the same as in case of the orthogonal wavelet transform, with the additional assumption of the finite variance. Similarly to the orthogonal model, in the case $p = \infty$ and $\beta < 1$ we consider two particular cases: a distribution with the power exponential tail in the theorem because the criterion for this type of distributions coincides with the criterion for other cases, and a distribution with polynomial tail in the proposition after the theorem. Now we can give the criterion linking the model for wavelet coefficients and the parameters of the Besov spaces.

Theorem 6 *Assume that continuous wavelet transform (15) of function $f : [0,1] \to \mathbb{R}$ follows assumptions W, B, H, for $1 \leqslant p, q \leqslant \infty$. Then $f \in B^s_{p,q}$ almost surely if and only if*

$$\begin{aligned} \textit{either} \quad & s + 1/2 - \beta/p - \alpha/2 < 0, \\ \textit{or} \quad & s + 1/2 - \beta/p - \alpha/2 = 0 \quad \textit{and} \quad \beta < 1,\ q = \infty,\ 1 \leqslant p < \infty. \end{aligned}$$

Proposition 2 *In the case $p = \infty$, $1 \leqslant q \leqslant \infty$, $0 \leqslant \beta < 1$ and $1 - H(x) + H(-x) = c_l x^{-l}[1 + o(1)]$ as $x \to +\infty$, $l > 0$, $c_l > 0$, and $l > q$ if $q < \infty$, under assumptions B and W,*

$$f \in B^s_{\infty,q} \quad \textit{a.s.} \quad \Leftrightarrow \quad s + 1/2 - \alpha/2 + (1-\beta)/l < 0.$$

We can see that the necessary and sufficient criteria for a function to belong to the Besov spaces are the same for the continuous as for the orthogonal wavelet transform. Note that similarly to the case of the orthogonal wavelet transform, in case $\beta > 1$ the expected number $\int_{a_0}^{\infty} \int_0^1 a^{-\beta} da db$ of non-zero wavelet coefficients ω_λ is finite. Therefore, for $\beta > 1$, function f belongs to the same Besov space as the wavelet function, i.e. $B^s_{p,q}$ with $1 \leqslant p, q \leqslant \infty$ and $0 < s < r$.

Remark 4 It can be shown that the criterion for the Besov membership of f with probability 1 under the three parameter model with $\tau_\lambda^2 = C_\tau a^{-\alpha}(\log a)^\gamma$ is the same as for the orthogonal wavelet transform considered in Section 3.4. In this case, assumption H of Theorem 6 (and Proposition 2) needs to be extended similarly to the assumption H in case of the orthogonal wavelet transform, and the proof of Theorem 6 can be adapted to prove the criterion for the extended model.

Proof (Theorem 6 and Proposition 2) *Sufficiency*

The proof of sufficiency is different for finite and infinite parameter p therefore we consider these two cases separately.

1. **Case** $p < \infty$**.**

Denote $\delta = s + 1/2 - \alpha/2 - \beta/p$ and $s' = s + 1/2 - 1/p$. We need to prove that if $\delta < 0$ or $\delta = 0$ in case $0 \leqslant \beta < 1$, $1 \leqslant p < \infty$, $q = \infty$, then $||w||_{b^s_{p,q}} < \infty$.

Using the result of Lemma 4 given in the appendix, and Jensen's inequality, we can show that $||w||_{b^s_{p,q}} < \infty$ for $q = 1$:

$$\begin{aligned}
E||w||_{b^s_{p,1}} &= E\sum_{j=j_0}^{\infty} 2^{js'}\left(\sum_{k=0}^{2^j-1}|w_{jk}|^p\right)^{1/p} + E\left(\sum_{k=0}^{2^{j_0}-1}|u_{j_0k}|^p\right)^{1/p} \\
&\leqslant \sum_{j=j_0}^{\infty} 2^{js'}\left(\sum_{k=0}^{2^j-1}E|w_{jk}|^p\right)^{1/p} + \left(\sum_{k=0}^{2^{j_0}-1}E|u_{j_0k}|^p\right)^{1/p} \\
&\leqslant \sum_{j=j_0}^{\infty} 2^{js'}(C2^j 2^{-jp\alpha/2}2^{-j\beta})^{1/p} + C2^{j_0(1/p-\alpha/2-\beta/p)} + C2^{j_0/p}C_w \\
&= \sum_{j=j_0}^{\infty} C2^{j\delta} + C2^{j_0(\delta-s')} + C2^{j_0/p}C_w,
\end{aligned}$$

which is finite if $\delta = s + 1/2 - \alpha/2 - \beta/p < 0$. Since $B^s_{p,q} \subset B^s_{p,1}$ for all $1 < q < \infty$ (Hardle et al [12]) we conclude that if $\delta < 0$ then $f \in B^s_{p,q}$ almost surely.

In the case $q = \infty$ we need to show that the following expectation is finite:

$$\begin{aligned}
E||w||_{b^s_{p,\infty}} &= E\sup_{j\geqslant j_0} 2^{js'}\left(\sum_{k=0}^{2^j-1}|w_{jk}|^p\right)^{1/p} + E\left(\sum_{k=0}^{2^{j_0}-1}|u_{j_0k}|^p\right)^{1/p} \\
&\leqslant \sup_{j\geqslant j_0} 2^{js'}\left(\sum_{k=0}^{2^j-1}E|w_{jk}|^p\right)^{1/p} + \left(\sum_{k=0}^{2^{j_0}-1}E|u_{j_0k}|^p\right)^{1/p} \\
&\leqslant \sup_{j\geqslant j_0} C2^{j(s'+1/p-\alpha/2-\beta/p)} + C2^{j_0(-\alpha/2+(1-\beta)/p)} + C2^{j_0/p}C_w \\
&= \sup_{j\geqslant j_0} C2^{j\delta} + C2^{j_0(\delta-s')} + C2^{j_0/p}C_w,
\end{aligned}$$

which is finite if $\delta \leqslant 0$.

2. **Case $p = \infty$.**

Here we also prove the sufficiency condition of Proposition 2, i.e. we consider symmetric distributions H which have power exponential and polynomial tail. We introduce another parameter, δ_H, which is different for different types of distributions H:

$$\delta_H = \begin{cases} (1-\beta)/l & \text{for a distribution } H \text{ with tail } x^{-l},\ l > 0, \\ 0 & \text{for a distribution } H \text{ with tail } e^{-(\nu x)^m},\ m > 0. \end{cases}$$

Then the necessary conditions given in the theorem and in the proposition can be unified as $\tilde{\delta} < 0$, where $\tilde{\delta} = s' - \alpha/2 + \delta_H$. In the case of H having a power exponential tail considered in the theorem, $\tilde{\delta}$ coincides with δ defined in the previous part of the proof.

Again, we need to show that if $\tilde{\delta} < 0$ then $||w||_{b^s_{\infty,q}} < \infty$, i.e. that the following inequalities hold almost surely:

$$\max_{k=0\ldots 2^{j_0}-1}(|u_{j_0k}|^q) + \sum_{j=j_0}^{\infty} 2^{s'qj}\{\max_{k=0\ldots 2^j-1}(|w_{jk}|)\}^q < \infty, \quad q < \infty,$$
$$\max_{k=0\ldots 2^{j_0}-1}(|u_{j_0k}|) + \sup_{j\geqslant j_0}\{2^{js'}\max_{k=0\ldots 2^j-1}(|w_{jk}|)\} < \infty, \quad q = \infty. \quad (19)$$

For the proof we need the Markov inequality that for a positive random variable ξ and a constant $B > 0$, $P\{\xi > B\} \leqslant E\xi/B$. We consider probabilities $P\{|w_{jk}| > A\}$ and $P\{|u_{j_0k}| > B\}$ for some $A, B > 0$. According to the Markov inequality,

$$P\{|w_{jk}| > A\} = P\{|w_{jk}|^{\varkappa} > A^{\varkappa}\} \leqslant A^{-\varkappa}E|w_{jk}|^{\varkappa},$$
$$P\{|u_{j_0k}| > B\} \leqslant B^{-\varkappa}E|u_{j_0k}|^{\varkappa}$$

for any $\varkappa > 0$, and in the case of a distribution with polynomial tail there is an additional condition that $\varkappa < l$. We can use an upper bound for the expectation of the absolute moments of wavelet and scaling coefficients given in Lemma 4 (see appendix):

$$P\{|w_{jk}| > A\} \leqslant A^{-\varkappa}E|w_{jk}|^{\varkappa} \leqslant cA^{-\varkappa}2^{-j(\beta+\varkappa\alpha/2)};$$
$$P\{|u_{j_0k}| > B\} \leqslant cB^{-\varkappa}2^{-j_0(\beta+\varkappa\alpha/2)}.$$

If we set $A = 2^{\gamma j}$ and $B = 2^{\iota j_0}$ then the inequalities for the wavelet and scaling coefficients can be rewritten as

$$P\left\{\max_{k=0\ldots 2^j-1}|w_{jk}| > A\right\} \leqslant \sum_{k=0}^{2^j-1} P\{|w_{jk}| > A\} \leqslant c2^{j(-\gamma\varkappa+1-\beta-\varkappa\alpha/2)},$$
$$P\left\{\max_{k=0\ldots 2^{j_0}-1}|u_{j_0k}| > B\right\} \leqslant c2^{j_0(-\varkappa\iota+1-\beta-\varkappa\alpha/2)}.$$

Therefore $\max_{k=0,\ldots,2^j-1}\{|w_{jk}|\} \leqslant A = 2^{\gamma j}$ almost surely if

$$\gamma - (1-\beta)/\varkappa + \alpha/2 > 0. \quad (20)$$

Similarly, $\max_{k=0,\dots,2^{j_0}-1}\{|u_{j_0k}|\} \leqslant B = 2^{\iota j_0}$ almost surely if

$$\iota - (1-\beta)/\varkappa + \alpha/2 > 0. \tag{21}$$

Suppose the inequalities (20) and (21) hold. Then conditions (19) are satisfied if, almost surely,

$$\max_{k<2^{j_0}}(|u_{j_0k}|) + \sum_{j=j_0}^{\infty} 2^{s'qj}||w_j||_\infty^q \leqslant 2^{qj_0\iota} + \textstyle\sum_{j\geqslant j_0} 2^{qj(s'+\gamma)} < \infty, \quad q<\infty,$$
$$\max_{k<2^{j_0}}(|u_{j_0k}|) + \sup_{j\geqslant j_0}\{2^{js'}||w_j||_\infty\} \leqslant 2^{j_0\iota} + \sup_{j\geqslant j_0}\{2^{j(s'+\gamma)}\} < \infty, \quad q=\infty,$$

which hold if $s'+\gamma<0$. The only condition on ι is $\iota > (1-\beta)/\varkappa + \alpha/2$ for some $\varkappa$. So we shall choose $\varkappa$ later and take $\iota = (1-\beta)/\varkappa + \alpha/2 + 1$.

Combining the requirement $s'+\gamma<0$ with condition (20) we have that γ should belong to the interval $((1-\beta)/\varkappa - \alpha/2, -s')$. If there exists $\varkappa>0$ such that this interval is not empty then the necessity is proved. In case $\beta=1$, the interval is not empty if and only if $\tilde\delta = s' - \alpha/2 < 0$.

In case $\beta<1$ we consider two cases separately where distribution H has polynomial or power exponential tail.

1. For a distribution with the power exponential tail $\tilde\delta = \delta = s' - \alpha/2$ and the interval for γ is $(\delta - s' + (1-\beta)/\varkappa, -s')$ for some $\varkappa>0$. Since $\beta<1$ and $\delta<0$ this interval is not empty if $\delta + (1-\beta)/\varkappa < 0$ i.e. if $\varkappa > -(1-\beta)/\delta$. For example, we can choose $\varkappa = 2\frac{1-\beta}{-\delta}$ and $\gamma = -s' + \delta/4 \in (-s' - \delta/2, -s')$.
2. For a distribution with polynomial tail $\tilde\delta = s' - \alpha/2 + (1-\beta)/l$, and the conditions on $\varkappa$ and γ are:

$$-s' + \tilde\delta + (1-\beta)(\frac{1}{\varkappa} - \frac{1}{l}) < \gamma < -s' \quad \textit{and} \quad 0<\varkappa<l.$$

   Since $\beta<1$, the interval for γ is not empty if $\tilde\delta + (1-\beta)(\frac{1}{\varkappa} - \frac{1}{l}) < 0$ which takes place for $\varkappa$ such that

$$\frac{1}{l} < \frac{1}{\varkappa} < \frac{1}{l} + \frac{-\tilde\delta}{1-\beta}.$$

   Thus there exists a pair $(\gamma,\varkappa)$ satisfying the stated conditions, and therefore the necessity of the proposition is proved.

The necessary condition for $||w||_{b^s_{p,q}}$ to be finite holds if $\delta<0$ (or $\tilde\delta<0$ in the proposition) or $\delta=0$ in case $1\leqslant p<\infty$, $q=\infty$, therefore the sufficiency is proved.

Necessity

The idea is to reduce the problem to the known case considered in Section 3, i.e. to the model where coefficients w_{jk} are independent and whose distribution is a mixture of the point mass at zero and distribution H_j. In the considered case w_{jk} is a sum of the same independent random variables with different coefficients, therefore w_{jk} are dependent:

$$w_{jk} = \sum_{\lambda \in S} K(\lambda, \lambda_{jk}) \omega_\lambda. \tag{22}$$

To create independent wavelet coefficients, we divide the area $[a_0, \infty) \times [0, 1]$ into non-overlapping rectangles $\mathcal{I}_{jk}$ and define new random variables w'_{jk} as a sum similar to the sum defining w_{jk} but over a restricted summation area $S \cap \mathcal{I}_{jk}$ instead of S:

$$w'_{jk} = \sum_{\lambda \in S \cap \mathcal{I}_{jk}} K(\lambda, \lambda_{jk}) \omega_\lambda.$$

Note that stochastically $|w'_{jk}| \leqslant |w_{jk}|$. Now we define rectangles $\mathcal{I}_{jk}$. Since function $K_0(u, v)$ is continuous and $K_0(1, 0) = 1$ we choose c_0 in $(0, 1)$ such that $K_0(u, v) > 1/2$ for all (u, v) in $[1, 1+c_0] \times [0, c_0]$ and define the rectangles as $\mathcal{I}_{jk} = [2^j, 2^j(1 + c_0)] \times [2^{-j}k, 2^{-j}(k + c_0)]$. Note that the definition of c_0 implies that for any $j \geqslant j_0$, $k = 0, \ldots, 2^j - 1$, $K(\lambda, \lambda_{jk}) > 1/2$ for $\lambda \in \mathcal{I}_{jk}$.

Now we find analogues of the proportion of non-zero coefficients at level j π_j and the variance of a wavelet coefficient at level j τ_j used in our model in Section 3.3. The expected number of wavelet indices λ falling into the region $\mathcal{I}_{jk}$ is $c \int_{\mathcal{I}_{jk}} a^{-\beta} da db = c_1 2^{-j\beta}$, therefore the probability that there is at least one λ in $\mathcal{I}_{jk}$ is $c_2 2^{-j\beta}$ for some constant c_2.

Next we estimate the variance of w'_{jk}:

$$\mathrm{Var}(\omega_\lambda | \lambda \in S \cap \mathcal{I}_{jk}) = B a^{-\alpha} \geqslant B(1 + c_0)^{-\alpha} 2^{-j\alpha} = 4 c_3 2^{-j\alpha},$$

so $\mathrm{Var}(K(\lambda, \lambda_{jk}) \omega_\lambda | \lambda \in S \cap \mathcal{I}_{jk}) \geqslant c_3 2^{-j\alpha}$ since $K(\lambda, \lambda_{jk}) > 1/2$ for $\lambda \in \mathcal{I}_{jk}$. Therefore $\mathrm{Var}(w'_{jk}) \geqslant c_4 2^{-j\alpha}$.

We define w^0_{jk} for the same j and k as w_{jk} such that they are independent and have a mixture distribution

$$w^0_{jk} \sim \pi_j h_j(x) + (1 - \pi_j) \delta_0(x),$$

where distribution h_j has variance $\tau_j^2 = c_4 2^{-j\alpha}$ and $\pi_j = \min\{1, c_2 2^{-j\beta}\}$.

By the construction of the wavelet coefficients w'_{jk} and w^0_{jk} the following inequalities hold stochastically: $|w_{jk}| \geqslant |w'_{jk}| \geqslant |w^0_{jk}|$. Therefore using Theorem 1 which can be applied under the assumptions of the theorem and the proposition, it follows that if the original function $f \in B^s_{p,q}$, i.e that $||w^0||_{b^s_{p,q}} \leqslant ||w||_{b^s_{p,q}} < \infty$ almost surely, then $\delta < 0$. Thus the theorem and the proposition are proved.

5 Discussion

In this paper we provided necessary and sufficient conditions for the Besov membership with probability 1 of a random function which is defined as a wavelet decomposition with random wavelet coefficients, following a mixture model of the atom at zero and some arbitrary distribution, under some additional assumptions. The form of the parameters of the distribution of the wavelet coefficients was motivated by previously considered prior distribution of Bayesian wavelet estimators for non-parametric function estimation. This condition was provided for both orthogonal and continuous wavelet transforms where in case of the continuous wavelet transform the atom at the zero part of the mixture in the orthogonal case was replaced by two-dimensional Poisson process model for selecting the wavelet indices where the wavelet coefficients are not zero.

We also showed that the assumptions about the underlying distribution of the non-zero wavelet coefficients for the necessary and sufficient condition cannot be weakened. In case $p = \infty$ (and $\beta < 1$) we considered only two major cases of tail behaviour, with power exponential and polynomial tails. However, the results can be extended to distributions with other types of monotonic decay.

The results for the continuous wavelet transform can be applied to a wavelet transform on an irregular grid which corresponds to various applications and where the irregularity of the grid can be modelled as a Poisson process. An extension of this result to other models of irregularity, e.g. other random two-dimensional processes to model the selection of the non-zero wavelet coefficients, can be a topic for future research.

The obtained results can also be used to compare the effect of different tail behaviour of the wavelet coefficients on the regularity properties of the function. For instance, it follows that with probability 1 the Besov spaces contain more functions whose wavelet coefficients have distribution with normal or exponential tail than those with polynomial tail. It also follows from the assumptions of the theorems that the more moments a distribution has, the wider is the set of the corresponding Besov spaces. In case $p = \infty$ and $\beta < 1$, the necessary and sufficient condition depends on the tail of the distribution directly, i.e. the heavier the tail is, the narrower is the set of the corresponding Besov spaces as we saw on the example of t, normal and Laplacian distributions.

An important application of the results of this paper is to study a priori Besov membership of Bayesian wavelet estimators. Also, these results can be used in Bayesian regression modelling to specify the probabilistic distribution of wavelet coefficients if there is some information available about the regularity of the function of interest. If the regularity of a function is known a priori, it can be used to specify the hyperparameters of the prior model for wavelet coefficients and to choose the distribution of non-zero wavelet coefficients. The results obtained in this paper can be adapted to most prior models, and thus a necessary and sufficient condition of a priori Besov membership of a function can be compared to other properties of the Bayesian models, e.g. frequentist optimality over the Besov spaces, thus making a more precise statement compared to that in Pensky [20]. In particular, the results of this paper can be extended to other parametrisations of the proportion π_j and the variance τ_j of the non-zero wavelet coefficients, for example, they can be considered to be different functions of the levels j for fine and coarse resolution levels.

Finally, these results can be used to study regularity properties of a random function or process which can be modelled in terms of wavelet transform.

Acknowledgements The author was financially supported by Overseas Students Research Award Scheme during her PhD study at the University of Bristol. The author would like to thank her PhD advisor Bernard Silverman for fruitful discussions, support and encouragement, and Theofanis Sapatinas for helpful comments.

A Additional statements

The following lemma is about the equivalence of l^p norms in a finite-dimensional space.

Lemma 2 *For any* $0 < v < l \leqslant \infty$, $x \in \mathbb{R}^n$, $n \in \mathbb{N}$, $||x||_l \leqslant ||x||_v \leqslant n^{\frac{1}{v}-\frac{1}{l}}||x||_l$.

Proof Hölder inequality implies that for $q = l/v > 1$ and $q < \infty$,

$$\sum_{k=1}^{n} |x_k|^v \leqslant \left(\sum_{k=1}^{n} |x_k|^{vq}\right)^{1/q} n^{1-1/q},$$
$$||x||_v \leqslant n^{1/v-1/l}||x||_l.$$

The inequality $||x||_l \leqslant ||x||_v$ follows from the nesting properties of the norms l^p.

For $l = \infty$, obvious inequalities

$$\max_{k=1,\dots,n}\{|x_k|^v\} \leqslant \sum_{k=1}^{n} |x_k|^v \leqslant n \max_{k=1,\dots,n}\{|x_k|^v\}$$

are equivalent to $||x||_\infty \leqslant ||x||_v \leqslant n^{1/v}||x||_\infty$, which completes the proof of the lemma.

Lemma 3 *If* Y *is a non-negative random variable then* $\sum_{j=0}^{\infty} P\{Y > j\} < \infty \Leftrightarrow EY < \infty$.

Proof The last statement is based on the representation of EY in terms of $P\{Y > x\}$: $EY = \int_0^\infty P\{Y > x\}dx$ (Feller [11], p.234), and thus the expectation is sandwiched between the approximations of the integral of the non-increasing function $P\{Y > x\}$ by the upper and lower Riemann sums:

$$\sum_{j=1}^{\infty} P\{Y > j\} \leqslant EY \leqslant \sum_{j=0}^{\infty} P\{Y > j\},$$

which proves the lemma.

To prove Theorem 6, we need to know the expectation of $|w_{jk}|^p$. The following lemma describes the asymptotic behaviour of the absolute moments of wavelet and scaling coefficients.

Lemma 4 *Suppose that the assumptions of Theorem 6 hold. Let w_{jk} and u_{j_0k} be the orthogonal wavelet and scaling coefficients corresponding to wavelet coefficients ω_λ. Then, as $j \to \infty$, for any $\varkappa \geqslant 1$,*

$$E|w_{jk}|^\varkappa = c2^{-j(\beta+\alpha\varkappa/2)}[1+o(1)], \qquad (23)$$
$$E|u_{j_0k}|^\varkappa = c2^{-j(\beta+\alpha\varkappa/2)}[1+o(1)].$$

Proof Lemma 4.

The wavelet and scaling coefficients depend on two types of random variables: the Poisson process S and the wavelet coefficients (17) with indices coming from this process. In order to find the moments of wavelet and scaling coefficients we start with the moments conditioned on the Poisson process.

1. The pth absolute moment of $(w_{jk}|S)$.
   Applying Jensen's inequality, we obtain an upper bound on the pth absolute moment of w_{jk} conditioned on Poisson process S:

$$E\left(|w_{jk}|^p|S\right) = E\left(\left|\sum_{\lambda\in S} K(\lambda,\lambda_{jk})\omega_\lambda\right|^p |S\right) \leqslant E\left(\sum_{\lambda\in S}|K(\lambda,\lambda_{jk})|^p|\omega_\lambda|^p|S\right)$$
$$= \sum_{\lambda\in S}|K(\lambda,\lambda_{jk})|^p E\left(|\omega_\lambda|^p|S\right) = \nu_p\sum_{\lambda\in S}|K(\lambda,\lambda_{jk})|^p a^{-\alpha p/2},$$

   where ν_p is the pth absolute moment of distribution H, $\lambda=(a,b)$. If we narrow the former sum on the following rectangles

$$\mathcal{I}^0_{jk} = [a_0,\infty)\times(2^{-j}k-1/2, 2^{-j}k+1/2), \qquad (24)$$

   it does not change since $K(\lambda,\lambda_{jk})\neq 0$ if and only if

$$b-k2^{-j}\in[-1/a, 2^{-j}]\subseteq[-2^{-j_0},2^{-j_0}]\subseteq[-1/2,1/2].$$

   Consider a map f_{jk}: $[a_0,\infty)\times[0,1]\to(0,+\infty)\times(-\infty,+\infty)$ which transforms the arguments of the kernel K to the arguments of the kernel K_0, i.e. it transforms $\lambda=(a,b)$ to $(u,v)=(2^{-j}a, 2^jb-k)$. The image of a rectangle $\mathcal{I}^0_{jk}$ under the transformation f_{jk} is the following rectangle:

$$\mathcal{I}_j := f_{jk}(\mathcal{I}^0_{jk}) = [a_02^{-j},\infty)\times[-2^{j-1},2^{j-1}].$$

   Let S'_j be a Poisson process on $\Omega=\{(u,v): u>0, -\infty<v<+\infty\}$ with intensity $2^{-j\beta}u^{-\beta}$. Applied to $S\cap\mathcal{I}^0_{jk}$, the map f_{jk} gives us a Poisson process with the same distribution as $S'_j\cap\mathcal{I}_j$.
   Now we make a substitution $(u,v)=f_{jk}(a,b)$ in the obtained sum in order to have the expectation in terms of the kernel K_0:

$$E(|w_{jk}|^p|S) \leqslant \nu_p\sum_{\lambda\in S}|K(\lambda,\lambda_{jk})|^p a^{-\alpha p/2} = \nu_p\sum_{(u,v)\in S'_j\cap\mathcal{I}_j}|K_0(u,v)|^p 2^{-j\alpha p/2}u^{-p\alpha/2}$$
$$= \nu_p 2^{-j\alpha p/2}\sum_{(u,v)\in S'_j\cap\mathcal{I}_j}|K_0(u,v)|^p u^{-p\alpha/2}.$$

2. Asymptotic behaviour of $E(|w_{jk}|^p)$.
   Denote the sum in (25) by $Z_j^{(p)} = \sum_{(u,v)\in S'_j\cap\mathcal{I}_j}|K_0(u,v)|^p u^{-p\alpha/2}$. Then, an upper bound for the expectation of a wavelet coefficient can be found as $E|w_{jk}|^p \leqslant c2^{-j\alpha p/2}EZ_j^{(p)}$.

We can use the Kernel properties (Section 2.2) to find the upper bounds for the summands of $Z_j^{(p)}$:

$$\begin{aligned} |K_0(u,v)|^p u^{-p\alpha/2} &\leqslant C u^{-p(r+\rho+1/2)-p\alpha/2} \\ &= C u^{-p(r+\rho+1/2+\alpha/2)} \quad \text{for} \quad u \geqslant 1; \\ |K_0(u,v)|^p u^{-p\alpha/2} &\leqslant C u^{p(r+\rho+1/2)-p\alpha/2} \\ &= C u^{p(r+\rho+1/2-\alpha/2)} \quad \text{for} \quad u \leqslant 1. \end{aligned} \tag{25}$$

Now, to estimate the expectation of $Z_j^{(p)}$, we apply Corollary 1 by Abramovich et al [2].

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 $\varepsilon\mu$, 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*

$$\int_\Omega \min(1, |g(x)|)\mu(dx) < \infty, \quad c_l = \int_\Omega |g(x)|^l \mu(dx) < \infty \quad \textit{for some} \quad l > 0 \tag{26}$$

Define $Y_\varepsilon = \sum_{X \in S_\varepsilon} g(X)$. Then $E|Y_\epsilon|^l = \varepsilon c_l + o(\varepsilon)$ as $\varepsilon \to 0$.

We apply the corollary with $X = (u,v)$, $\Omega = [0,\infty) \times (-\infty,+\infty)$, $d\mu(u,v) = u^{-\beta} dudv$, $\varepsilon = 2^{-\beta j}$ and $g(X) = |K_0(u,v)|^p u^{-p\alpha/2} \geqslant 0$. The assumptions of the corollary are satisfied if the integral $\int_\Omega g(x) d\mu(x)$ is finite. Using the bound (25) of $|K_0(u,v)|^p u^{-p\alpha/2}$, we can estimate the required integral:

$$\begin{aligned} \int_\Omega g(x) d\mu(x) &= \int_0^{+\infty} \int_{-\infty}^{+\infty} |K_0(u,v)|^p u^{-p\alpha/2} dudv \\ &= \int_0^1 \int_{-\infty}^{+\infty} |K_0(u,v)|^p u^{-p\alpha/2} u^{-\beta} dudv \\ &+ \int_1^{+\infty} \int_{-\infty}^{+\infty} |K_0(u,v)|^p u^{-p\alpha/2} u^{-\beta} dudv \\ &\leqslant C \int_0^1 u^{p(r+\rho+1/2-\alpha/2)-\beta}(1+1/u) du \\ &+ C \int_1^{+\infty} u^{-p(r+\rho+1/2+\alpha/2)-\beta}(1+1/u) du, \end{aligned}$$

since the Kernel property (Section 2.2) gives us that $v \in [-1/u, 1]$.
For the first integral to be finite we need to show that $p(r+\rho+1/2-\alpha/2)-\beta > 0$. One of the assumptions of the theorem is that $r+\rho+1/2-\alpha/2 > 1$. Since $p \geqslant 1 \geqslant \beta$, the inequality is satisfied and the first integral is finite. The second integral is finite due to the same assumption W that $r+\rho+1/2-\alpha/2 > 1$:

$$-p(r+\rho+1/2+\alpha/2) - \beta < -p(1+\alpha) - \beta < -p \leqslant -1.$$

Hence both assumptions of Corollary 1 are satisfied implying that

$$EZ_j^{(p)} = c2^{-\beta j} + o(2^{-\beta j}) \quad \text{as} \quad j \to \infty,$$

which gives a bound for the pth absolute moment of w_{jk}:

$$E|w_{jk}|^p = C2^{-j(\beta+\alpha p/2)}[1+o(1)]. \tag{27}$$

Since the kernels W and W_0 defined in Section 2.2 have the same properties as the properties of the kernels K and K_0 we used in the proof, the same asymptotic result holds for $E|u_{j_0 k}|^p$.

References

1. Abramovich, F., Sapatinas, T., Silverman, B.: Wavelet thresholding via a Bayesian approach. J. Roy. Statist. Soc. Ser. B **60**(4), 725–749 (1998)
2. Abramovich, F., Sapatinas, T., Silverman, B.: Stochastic expansions in an overcomplete wavelet dictionary. Probability Theory and Related Fields **117**, 133–144 (2000)
3. Billingsley, P.: Probability and measure. Wiley, New York (1995)
4. Bochkina, N., Sapatinas, T.: On the posterior median estimators of possibly sparse sequences. Annals of Institute of Statistical Mathematics **57**, 315–351 (2005)
5. Clyde, M., George, E.: Empirical Bayes estimation in wavelet nonparametric regression. In: P. Muller, B. Vidakovic (eds.) Bayesian Inference in Wavelet Based Models, pp. 309–322. Springer-Verlag, New York (1999)
6. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
7. Donoho, D., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
8. Donoho, D., Johnstone, I.: Minimax risk over l_p-balls for l_q-error. Probability Theory and Related Fields **99**, 277–303 (1994)
9. Donoho, D., Johnstone, I.: Adapting to unknown smoothness via wavelet shrinkage. J.Amer.Statist.Assoc. **90**, 1200–1224 (1995)
10. Donoho, D., Johnstone, I.: Minimax estimation via wavelet shrinkage. The Annals of Statistics **26**, 879–921 (1998)
11. Feller, W.: An Introduction to probability theory and its applications, vol. II. Wiley, New York (1966)
12. Hardle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets, Approximation and Statistical Applications. In: Lect. Notes Statist., vol. 129. Springer-Verlag, New York (1998)
13. Johnstone, I., Silverman, B.: Needles and hay in haystacks: Empirical bayes estimates of possibly sparse sequences. Ann. Statist **32**, 1594–1649 (2004)
14. Leadbetter, M., Lindgren, G., Rootzen, H.: Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York (1983)
15. Mallat, S.: A theory of multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Machine Intell. **11**(7), 674–693 (1989)
16. Mallat, S.: A wavelet tour of signal processing. Academic Press, London (1999)
17. Meyer, Y.: Wavelets and operators, *Cambridge studies in Advanced Mathematics*, vol. 37. Cambridge University Press, New York (1992)
18. Mochimaru, F., Fujimoto, Y., Ishikawa, Y.: Detecting the fetal electrocardiogram by wavelet theory-based methods. Progress in Biomedical Research **7**(3), 185–193 (2002)
19. Ouergli, A.: Hilbert transform from wavelet analysis to extract the envelope of an atmospheric mode: examples. Journal of Atmospheric and Oceanic Technology **19**(7), 1082–1086 (2002)
20. Pensky, M.: Frequentist optimality of Bayesian wavelet shrinkage rules for gaussian and non-gaussian noise. Ann. Statist **34**(2) (2006)
21. Polygiannakis, J., Preka-Papadema, P., Moussas, X.: On signal-noise decomposition of timeseries using the continuous wavelet transform: Application to sunspot index. Monthly Notice of the Royal Astronomical Society **343**(3), 725 (2003)
22. Triebel, H.: Theory of Functional Spaces, vol. 2. Birkhauser-Verlag, Basel (1990)
23. Vidakovic, B.: Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. J. Amer. Statist. Assoc. **93**, 173–179 (1998)
24. Vidakovic, B.: Statistical Modeling by Wavelets. John Wiley and Sons, New York (1999)
25. Walter, G.: Wavelets and other orthogonal systems with applications. CRC Press Inc., Boca Raton, FL (1994)