BIOINFORMATICS REVIEW

Vol. 19 no. 1 2003
Pages 2–9

Wavelets in bioinformatics and computational biology: state of art and perspectives

Pietro Liò

Department of Zoology, University of Cambridge, Cambridge, UK and European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, UK

Received on October 17, 2002; revised on May 21, 2002; July 7, 2002; accepted on July 7, 2002

ABSTRACT

Motivation: At a recent meeting[†], the wavelet transform was depicted as a small child kicking back at its father, the Fourier transform. Wavelets are more efficient and faster than Fourier methods in capturing the essence of data. Nowadays there is a growing interest in using wavelets in the analysis of biological sequences and molecular biology-related signals.

Results: This review is intended to summarize the potential of state of the art wavelets, and in particular wavelet statistical methodology, in different areas of molecular biology: genome sequence, protein structure and microarray data analysis. I conclude by discussing the use of wavelets in modeling biological structures.

Contact: plio@hgmp.mrc.ac.uk

INTRODUCTION

The name wavelets means small waves (the sinusoids used in Fourier analysis are 'big waves'), and, in short, a wavelet is an oscillation that decays quickly. The field has progressed so rapidly that a widely accepted definition of wavelets is very general: 'Wavelets are building blocks that can quickly decorrelate data' (Sweldens, 1996).

In recent years, wavelets analysis has been applied to a large variety of biomedical signals (Aldroubi and Unser, 1996), and there is a growing interest in using wavelets in the analysis of sequence and functional genomics data. Therefore, this review is intended to give a relatively accessible introduction to wavelet analysis for bioinformaticians and computational biologists. The paper first establishes some necessary basic mathematical background and terminology relating to wavelets. After briefly mentioning the more well-established techniques of wavelets in statistics, I focus on the applications and perspectives in different molecular biology areas.

[†] XIX SMC 2001 'Wavelets in Statistics', Vico Equense, Naples, I, 2–7 April 2001.

MATHEMATICAL BACKGROUND

Wavelet transform

Let us set the scene with two biological examples. Hearing entails a transform from time to frequency accomplished by the hairy cells in the cochlear and neurons in the auditive cortex (Schreiner *et al.*, 2000). In vision, the red, green and blue cone receptors in the retina pass along signals for the visual cortex where color composition occurs (Nijhawan, 1997). Although there is no area of research in which Fourier transform (FT) has not proved useful, nature is still ahead of mathematics in transforming a signal from the time or space domain into the frequency domain (see the discussion on natural stimulus statistics in Donoho and Flesia, 2001). The wavelet transform (WT) is relatively new (early 80s) and has some similarities with the Fourier transform (FT). Wavelets differ from Fourier methods in that they allow the localization of a signal in both time and frequency. Figure 1a shows the difference between the FT, windowed FT and WT in terms of time and frequency localization. A WT of a signal typically outperforms an FT when the signal under consideration contains discontinuities and sharp spikes. In wavelet theory, a function is represented by an infinite series expansion in terms of dilated and translated version of a basic function ψ called the 'mother' wavelet (Daubechies, 1992; Chui, 1992; Mallat, 1989). The simplest example of a wavelet basis is the Haar basis (Figure 1b, left); other frequently used wavelet bases are those developed by Daubechies (1992, Figure 1b, right). Several wavelet families, with different properties (orthogonal, biorthogonal, semiorthogonal) have recently been developed (Daubechies, 1992; Chui, 1992; Mallat, 1999). The continuous wavelet transform, (CWT) for a function $f(t)$ is defined as:

$$CWT(f,a,b) = a^{-1/2}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)\mathrm{d}t$$

where a (the scale parameter) >0, b (the translation parameter) $\in \Re$. The CWT maps a one-dimensional signal to a two-dimensional time-scale joint representation. It is

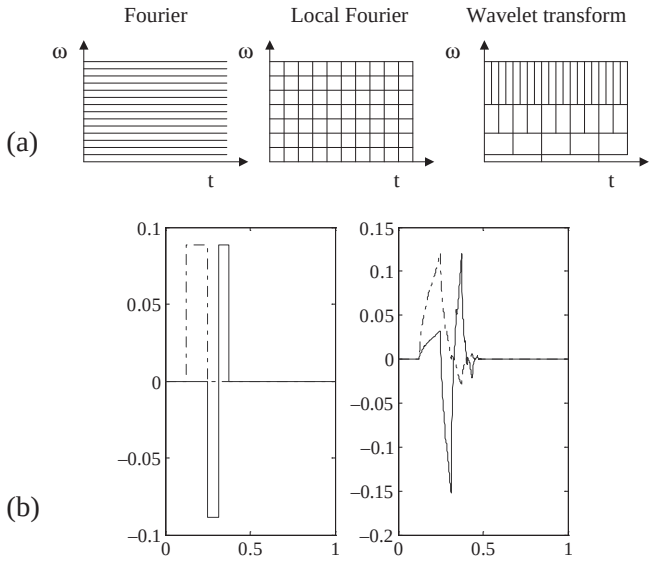

Fig. 1. According to the Heisenberg uncertainty principle, we cannot know precisely both the position in time (or space) and the frequency of a signal. With the Fourier transform, the localization of a signal that occurs in a short time interval is lost when transformed to frequency (a, left); the windowed Fourier transform usually represents a remarkable improvement (a, central). With wavelets, the time/frequency plane is patched with rectangles of different shape. At low frequency the height of the boxes are longer (poor frequency resolution) and widths are shorter (better time resolution) (a, right). Mother wavelet (line) and a related function, termed Father wavelet (dash) of Haar wavelet basis (b, left) and Daubechies with $N = 4$ (b, right). See Daubechies (1992), Chui (1992), Mallat (1989) for a comprehensive description of the relation between Father and Mother wavelets.

calculated by continuously shifting a continuously scalable function over a signal and calculating the correlation between the two. The resulting wavelet coefficients are highly redundant. Because in molecular biology and genetics we are more concerned with discretely sampled rather than continuous functions, let us briefly summarize first the main qualitative features of the discrete wavelet transform (DWT).

The discrete wavelet transform

A WT decomposes a signal into several groups (vectors) of coefficients. Different coefficient vectors contain information about characteristics of the sequence at different scales. Coefficients at coarse scales capture gross and global features of the signal while coefficients at fine scales contain local details. The DWT is an economical way to compute the WT, because it is computed only on a dyadic grid of points, where the subsampling is at a different rate for different scales. The DWT is commonly introduced using a matrix or a computational form. In matrix form we can represent the DWT (Mallat, 1989) through an orthogonal matrix

$$W = [W_1^{\mathrm{T}}, W_2^{\mathrm{T}}, \ldots, W_J^{\mathrm{T}}, V_J]^{\mathrm{T}},$$

where J is the largest level of the transform, and '$^{\mathrm{T}}$' indicates transpose. A DWT is applied to a vector X of observations as $d = WX$ and decomposes the data into sets of wavelet coefficients

$$d = [d_1^{\mathrm{T}}, d_2^{\mathrm{T}}, \ldots, d_J^{\mathrm{T}}, c_J^{\mathrm{T}}]^{\mathrm{T}}$$

with $d_j = W_j X$, $c_J = V_J X$. At scale $\tau_j = 2^{j-1}$, or level j, there are $n/2^j$ coefficients d_j, which are associated with changes in averages of the data on a scale $\tau_j \Delta t$ with Δt the time interval between consecutive observations. Each wavelet coefficient at that level tells us how much a weighted average of the data changes from a particular time period of effective length $\tau_j \Delta t$ to the next one. Scaling coefficients c_J are instead associated with averages of the data on scales $\tau_{J+1}\Delta t$ and higher, with J the largest level of the DWT. The WT is a cumulative measure of the variations in the data over regions proportional to the wavelet scales; for increasing values of j, the coefficients describe features at lower frequency ranges and larger time periods.

In the undecimated WT there are n coefficients at every scale, i.e. the number of coefficients does not decrease with the level (Mallat, 1989; Shensa, 1992). Although the transformation is not orthogonal anymore other useful features are gained. Coefficients are translation-equivariant, i.e. circularly shifting of the data is reflected in the same shifting of the coefficients. The undecimated WT is capable of handling data with arbitrary size, i.e. it does not require the sample size n to be a power of two. These two properties are particularly useful in biological sequence analysis because they allow precise correspondence between wavelet coefficients and sequence patterns. It is noteworthy that the undecimated WT is superior to the ordinary DWT in many statistical applications (Nason and Silverman, 1995; Nason *et al.*, 2000).

Multiresolution analysis

A WT leads to an additive decomposition of a signal into a series of different components describing smooth and rough features of the signal. In fact we have

$$X = W^{\mathrm{T}} d = \sum_{j=1}^{J} W_j^{\mathrm{T}} d_j + V_J^{\mathrm{T}} c_J = \sum_{j=1}^{J} D_j + C_J$$

with D_j the detail of the signal describing changes at the scale τ_j and C_J the smooth component associated with variations at scales τ_{J+1} and higher.

Table 1. Software for wavelet analysis

Software	State	Source	Reference	Web address
LiftPack	Free	Ansi C	Fernandez *et al.* (1996)	http://www.cs.sc.edu/∼fernandel/liftpack
LastWave	Free	AnsiC	emmanuel.bacry@polytechnique.fr	http://www.cmap.polytechnique.fr/∼bacry/LastWave
Waili	Free	C++	Uytterhoeven *et al.* (1998)	http://www.cs.kuleuven.ac.be/∼wavelets
Wavelet Explorer	Com	Mathematica	http://www.wolfram.com	http://www.wolfram.com
Wavelab	Free	Matlab	Buckheit and Donoho (1995)	http://www.stat.stanford.edu/∼wavelab
Wavebox	Com	Matlab	Taswell (1995)	http://www.wavbox.com
Smoothing Toolbox	Free	Matlab	Antoniadis *et al.* (2001)	http://www.Imc.imag.fr/SMS/software.html
Rice Wavelet Toolbox	Free	Matlab	http://www.dsp.rice.edu/publications	http://www.dsp.rice.edu/software/RWT
Wavelet Toolbox	Com	Matlab	http://www.mathworks.com	http://www.mathworks.com
Wavelet Software	Free	Matlab	Torrance and Compo (1998)	http://paos.colorado.edu/research/wavelets
Curvelets256 Toolbox	Free	Matlab	Candes and Donoho (1999)	http://www.acm.caltech.edu/∼emmanuel
Wavethresh	Free	Splus	Nason and Silverman (1995)	http://www.stats.bris.ac.uk/∼wavethresh
S+ Wavelets Toolkit	Com	Splus	Bruce and Gao (1996)	http://www.insightful.com
Thresh	Free	Splus	Ogden and Parzen (1996)	http://lib.stat.cmu.edu/S/thresh
Wavepot	Free	Splus	Raimondo (2002)	http://www.maths.usyd.edu.au:8000/u/marcr/Wavepot

Note: List of useful software related to the wavelet techniques described in this review. Entries are: software name, availability (Com, commercial; Free, freeware), software platform, reference, web address.

Implementation of the DWT

From a computational point of view, the DWT proceeds by recursively applying two convolution functions, known as quadrature mirror filters, each producing an output stream that is half the length of the original input, until the resolution level zero is reached (Mallat, 1989). This algorithm, termed the 'pyramid' algorithm, is faster than the fast FT, being of complexities, O(n) and O($n \log_2 n$), respectively. If the quadrature mirror filters are applied J times, at each level $j = 1, \ldots, J$ the transform produces two vectors of coefficients, c_j of scaling coefficients and d_j of wavelet coefficients. The vector d_j is kept while c_j is processed through the two filters. At the last level J, both c_j and d_j are kept. A pyramid algorithm exists for 2D data also (Mallat, 1989). An inverse wavelet tranform can be defined and allows reconstruction of a signal from its wavelet decomposition. A variety of software available from the web is listed in Table 1.

WAVELET STATISTICS AND MODELING

Many researchers feel that there is a need to use more statistics in bioinformatics. Therefore, I describe briefly (and give references for) some wavelet applications in statistics that might be important in bioinformatics. For more complete coverage, see Vidakovic (1999), Hardle *et al.* (1998) and Percival and Walden (2000). Wavelet techniques are now used in many statistical areas, for example, density estimation (Johnstone *et al.*, 1992; Donoho *et al.*, 1996a), nonparametric regression (Donoho and Johnstone, 1994, 1998; Donoho *et al.*, 1996b), change-point problems (Ogden and Parzen, 1996; Wang, 1995; Raimondo, 1998), time series analysis (Morettin, 1996; Percival and Walden, 2000; Nason *et al.*, 2000) and estimation and simulation of fractional Brownian motion (Flandrin, 1992).

A recent study on measles dynamics demonstrates how wavelet phase analysis has considerable potential as tool for ecological time series and spatial analyses (Grenfell *et al.*, 2001). In bioinformatics, many algorithms for pattern recognition are based on hidden Markov models (HMM) or neural networks. It is worth mentioning that there are wavelet denoising algorithms based on HMM (see http://www-dsp.rice.edu/publications).

Dealing with noise in biological signals

In statistics, the recovery of the underlying function from a noisy signal is generally modeled using regression models; several authors have proposed wavelet estimators (Donoho and Johnstone, 1994, 1998; Nason, 1996; Amato and Vuza, 1997; Antoniadis, 1996; Antoniadis *et al.*, 2001). Consider the standard univariate regression: $y_i = f(x_i) + \epsilon_i$, where $i = 1, \ldots, n$, and ϵ_i are independent $N(0, \sigma^2)$ random variables; f is the 'true' function. We can reformulate the problem in terms of wavelet coefficients: $\hat{w}_{jk} = w_{jk} + \epsilon_{jk}$, where j is the level ($j = 0, \ldots, j - 1$), and k, the displacement ($k = 0, \ldots, 2^j - 1$). It is often reasonable to assume that only a few large coefficients contain information about the underlying function, while small coefficients can be attributed to noise. Shrinkage consists in attenuating or eliminating the smaller wavelet coefficients and reconstructing the profile using mainly the most significant wavelet coefficients and all the scaling coefficients. Several shrinkage approaches have been proposed. For example, the 'hard' threshold approach selects coefficients

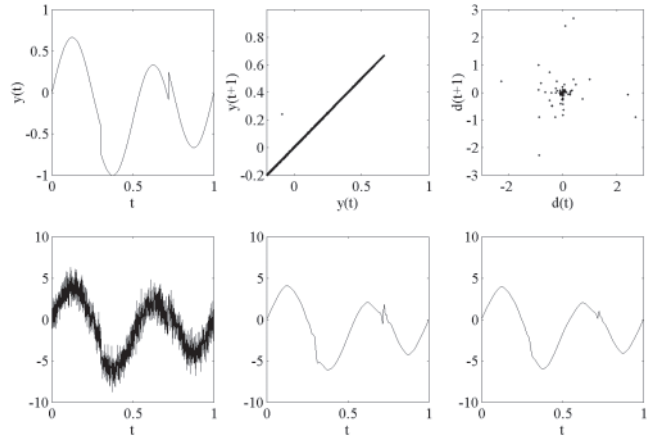

Fig. 2. Statistical properties of wavelets. Upper series: a HeaviSine function is sampled at 2048 data points uniformly spaced on $[0, 1]$ (left); then, the values of $y(t + 1)$ are plotted against $y(t)$ for the same function in time (center; correlation $= 0.99$) and the values of $d(t + 1)$ are plotted against $d(t)$ for the same function in wavelet domain using Daubechies $N = 4$ (right; correlation $= -0.14$); bottom series: a noisy HeaviSine function (left) is shrinked using 'hard' universal (center) and 'soft' universal thresholds (using Daubechies $N = 4$; right).

using a 'keep or kill' policy. Using 'soft' thresholding, if the magnitude of the wavelet coefficient is greater than (less than, respectively) the threshold, the coefficient is shrunk toward zero by an amount that depends on how large the magnitude of the coefficient is (is set to zero, respectively). Donoho and Johnstone (1994) proposed the 'universal' threshold, $\lambda_{un} = \sigma\sqrt{2\log n}$, and showed that it performs very well in both hard and soft thresholding. Thresholds can also be chosen based on the data using a hypothesis testing procedure (Ogden and Parzen, 1995, 1996; Abramovich and Benjamini, 1996; Raimondo, 2002). Data-adaptive thresholds might become very important in analyzing molecular biological data because hypothesis testing procedures can be used to test the appropriateness of various thresholds to the data under different biological assumptions (see for example Liò and Vannucci, 2000a). Finally, it is worth mentioning that several authors have proposed Bayesian thresholds and have reported interesting results (Abramovich *et al.*, 1998; Vidakovic, 1999, and references therein). Figure 2 illustrates the decorrelation (upper series) and denoising (bottom series) properties of wavelets.

Estimating variability over scales

A wavelet variance is a scale-by-scale decomposition of the variance of a signal (Percival, 1995). Replacing 'global' variability with variability over scales allows us to investigate the effects of constraints acting at different time or space scales. An estimate of the wavelet variance at a given scale is obtained by summing the squares of the wavelet coefficients and dividing by the number of them. With bivariate signals, the wavelet covariance is given by the sum at a given level, of the cross-products of coefficients with the same location (Lindsay *et al.*, 1996; Serroukh and Walden, 2000). In a similar way, wavelet cross-covariance, correlation and cross-correlation can be computed (Whitcher *et al.*, 2000; Percival and Walden, 2000). It is worth mentioning that a plot of the sum of the squares of the coefficients at each scale is termed a scalogram (Flandrin, 1988).

MOLECULAR BIOLOGY DATA AND WAVELETS

We briefly review the most interesting applications in the following areas: genome sequence analysis, protein structure investigation and gene expression data analysis.

Genome sequence analysis

Several authors showed that wavelets can be useful in detecting patterns in DNA sequences (Arneodo *et al.*, 1996, 1998; Dodin *et al.*, 2000). In particular, Audit *et al.* (2001, 2002) analyzed bending properties of sequences and provided evidence that the existence of long range correlations in the small-scale regime (10–200 bp) which are actually observed in eukaryotic genomes, in contrast to their absence in eubacterial genomes, depends on nucleosome patterns. Liò and Vannucci (2000b) showed that wavelet variance decomposition of bacterial genome sequences coded as G,C$=$ 1 and A,T $=$ 0 (or -1) can reveal the location of pathogenicity islands (see also Vannucci and Lio, 2001).

Protein structure investigation

Wavelets have been applied to all aspects of protein structural investigations: primary sequence evolution (Morozov *et al.*, 2000; Rzhetsky and Morozov, 2001), secondary (Liò and Vannucci, 2000a) and tertiary structure determination (Murray *et al.*, 2002; Mandell *et al.*, 1997, 1998; Hirakawa *et al.*, 1999), refinement of X-ray crystallography (Main and Wilson, 2000; Ferrer *et al.*, 1998), drug design and visualization (Carson, 1996). Mandell *et al.* (1997, 1998) showed that the phase plots of Morlet WT of the hydrophobic profiles of amino acid sequences can be related to the content of secondary structures (alpha helices and beta sheets) and can be used to classify proteins. They also investigated fluctuations of hydrophobicity along the sequence to derive information on channels, pore and receptors. Recently Murray *et al.* (2002) used the CWT to analyze the hydrophobicity and relative accessible surface area of a variety of repeating protein motifs such as TIM barrels, propellor blades, coiled coils and leucine repeats.

Microarray data analysis

Microarray technology allows us to analyze the expression patterns of hundreds of genes. The use of statistics is the key to extracting useful information from this technology. Because of this, CAMDA, critical assessment in microarray data analysis (http://bioinformatics.duke.edu/CAMDA) has been established to provide a community-wide critical assessment of different techniques used in microarray data analysis. Klevecz (2000) used wavelet decomposition and denoising techniques to analyze expression microarray data and found that the expression of most yeast genes oscillate, including both cell cycle regulated genes and ones not related to the cell cycle. The author found two major periodicities, one of ∼40 minutes and a second one of ∼80 minutes and hypothesized that part of the noise in expression microarray data may depend on the genes being expressed with oscillatory dynamics.

Microarray data analysis can also benefit from data compression. It is general practice to keep entire microarray images for reanalysis in case better statistical methods become available. Wavelet-based techniques are now the new compression standard: the lifting scheme is the basis of JPEG2000 (see http://www.jpeg.org for more details). Previous JPEG algorithms worked in terms of eight-by-eight squares; what makes wavelets better than other compression methods is their ability to adapt to the size and location of regions in the image. Jörnsten and Yu (2000) proposed a microarray image compression technique, termed 'comprestimation', with a lossless or lossy coded data structure. The authors discussed the question of optimal statistical estimation based on lossy compressed data and determined an upper bound on the minimum achievable loss of estimation efficiency due to compression. Myasnikova *et al.* (2001) used wavelets to analyze gene expression measured using tagged antibodies in a set of embryos. These authors obtained a detailed gene expression map of a morphogenetic field from fragmentary data. Recently Efron *et al.* (2001) have shown that the 'False Discovery Rate' (FDR) is a very useful inferential approach in the analysis of microarray data. The FDR is a relatively new and important idea in multiple comparisons: the FDR is the expected proportion of rejected hypotheses that are falsely rejected (Benjamini and Hochberg, 1995). When the model is sparse, FDR-like selection yields estimators with strong large sample adaptivity properties. One natural application of FDR is threshold selection in wavelet denoising. In wavelet thresholding, FDR is the proportion of wavelet coefficients erroneously included in the reconstruction among those included. This approach might lead to other applications of wavelets in microarray data analysis.

PERSPECTIVES: VERTICAL INTEGRATION OF MOLECULAR DATA AND PHYSIOLOGY

With the completion of the human and mouse genome projects the attention of bioinformaticians will move to functional genomics and to the integration of molecular and physiological data.

Modeling biological microstructures

Recently, wavelets have been used to analyze the chromatin distribution of the nucleus of breast tissue after Feulgen coloring staining (Van De Wouwer *et al.*, 2000). The analysis of the tissue and cell surface texture allows the characterization of invasive breast cancer and other cellular pathologies and represents a very active area of research: see also Li *et al.*, 2001 and the analysis of structure and morphogenic regulation organization in filamentous fungal colonies by Jones (1996).

Wavelets seem more suitable for describing scaling biological structures and signals than other mathematical transforms because wavelet basis functions can have different complex shapes, each suitable for a different class of problems. There is on going research on defining wavelets for general geometries (curves, surfaces, manifolds), over irregular sampling and for particular problems. For example, curvelets (useful for curved edges) and ridgelets (useful for straight edges) have elements distributed across a range of scales and location along with orientations (Candes and Donoho, 1999). These elements have increasing numbers of distinct directions as the scales go finer. The need for these new mathematical transforms comes in part from study of the human visual system using natural stimulus statistics. Edges are the dominant features both in human perception (providing segmentation into objects) and in mathematical settings (Donoho and Flesia, 2001). Therefore, wavelet methods (and extensions beyond wavelets) appear to be a natural way to achieve vast improvements in the quality of statistical analysis of biological structures and patterns.

Physiome data?

There are many examples of application of wavelets in physiology. For example, Diserbo and colleagues used wavelets to localize stationary segments in long single-channel current recordings and infer the gating channel mechanisms (Diserbo *et al.*, 2000). Alt *et al.* (1998) analyzed the courtship signals in different species of Drosophila and confirmed original findings by Kyriacou and Hall (1986) that mutations in the *per* (period) gene alter the interpulse intervals in songs. Currently, there is considerable interest in integrating quantitative biological information at different size scales, from molecules to cells, tissues, organs and organisms. A hint at the potential of wavelets in modeling the scaling behavior of biological

systems comes from the results of wavelet-based modeling of physical systems showing scaling behavior in, for example, turbulence (Katul *et al.*, 2000).

CONCLUSION

The aim of this paper was to increase the familiarity of wavelet techniques in the bioinformatics community and to provide useful references to the literature and to available software. In conclusion, I am delighted to report that the application of wavelets in molecular biology is a thriving field of research. It has two important and linked benefits: an improved ability for capturing hidden components from biological data and a better link between biological systems and the mathematics objects used to describe them.

ACKNOWLEDGEMENTS

The author is supported by an EPSRC/BBSRC Bioinformatics Initiative Grant. I would like to thank M.Vannucci and U.Amato, A.Antoniadis, N.Goldman, G.Nason for a number of comments which led to important improvements to this paper. I wish to extend an apology to the authors of the many excellent papers that could not be cited directly because of space constraints.

REFERENCES

Abramovich,F. and Benjamini,Y. (1996) Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.*, **22**, 351–361.

Abramovich,F., Sapatinas,T. and Silverman,B.W. (1998) Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc. B*, **60**, 725–749.

Aldroubi,A. and Unser,M. (eds) (1996) *Wavelets in Medicine and Biology*. CRC Press, Boca Raton, FL.

Alt,S., Ringo,J., Talyn,B., Bray,W. and Dowse,H. (1998) The period gene controls courtship song cycles in *Drosophila melanogaster*. *Anim. Behav.*, **56**, 87–97.

Amato,U. and Vuza,D.T. (1997) Wavelet approximation of a function from samples affected by noise. *Rev. Roumanie Math. Pure Appl.*, **42**, 481–493.

Antoniadis,A. (1996) Smoothing noisy data with tapered coiflets series. *Scand. J. Statist.*, **23**, 313–330.

Antoniadis,A., Bigot,J. and Sapatinas,T. (2001) Wavelet estimators in non parametric regression: a comparative simulation study. *J. Statist. Soft.*, **6**, i06.

Arneodo,A., d'Aubenton-Carafa,Y., Bacry,E., Graves,P.V., Muzy,J.F. and Thermes,C. (1996) Wavelet based fractal analysis of DNA sequences. *Physica D*, **1328**, 1–30.

Arneodo,A., d'Aubenton-Carafa,Y., Audit,B., Bacry,E., Muzy,J.F. and Thermes,C. (1998) What can we learn with wavelets about DNA sequences. *Physica A*, **249**, 439–448.

Audit,B., Thermes,C., Vaillant,C., d'Aubenton-Carafa,Y., Muzy,J.F. and Arneodo,A. (2001) Long-range correlations in genomic DNA: a signature of the nucleosomal structure. *Phys. Rev. Lett.*, **86**, 2471–2474.

Audit,B., Vaillant,C., Arneodo,A., d'Aubenton-Carafa,Y. and Thermes,C. (2002) Long-range correlations between DNA bending sites: relation to the structure and dynamics of nucleosomes. *J. Mol. Biol.*, **316**, 903–918.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B*, **57**, 289–300.

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

Buckheit,J.B. and Donoho,D.L. (1995) WaveLab and reproducible research. In Antoniadis,A. and Oppenheim,G. (eds), *Wavelets and Statistics*, Lecture Notes in Statistics, 103, Springer, New York, pp. 55–81.

Candes,E.J. and Donoho,D.L. (1999) Ridgelets: a key to higher dimension intermittency? *Phil. Trans. R. Soc. Lond. A.*, **357**, 2495–2509.

Carson,M. (1996) Wavelets and molecular structure. *J. Computer-Aided Molecular Design*, **10**, 273–283.

Chui,C.K. (1992) *An Introduction to Wavelets*. Academic Press, New York.

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

Diserbo,M., Masson,P., Gourmelon,P. and Caterini,R. (2000) Utility of the wavelet transform to analyze the stationarity of single ionic channel recordings. *J. Neurosci. Methods*, **99**, 137–141.

Dodin,G., Vandergheynst,P., Levoir,P., Cordier,C. and Marcourt,L. (2000) Fourier and wavelet transform analysis, a tool for visualising regular patterns in DNA sequences. *J. Theor. Biol.*, **206**, 323–326.

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

Donoho,D.L. and Johnstone,I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.

Donoho,D.L. and Flesia,A.G. (2001) Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics? *Technical Report*. Stanford University, Department of Statistics, Stanford, CA.

Donoho,D.L., Johnstone,I.M., Kerkyacharan,G. and Picard,D. (1996a) Density estimation by wavelet thresholding. *Ann. Statist.*, **24**, 502–539.

Donoho,D.L., Johnstone,I.M., Kerkyacharan,G. and Picard,D. (1996b) Wavelet shrinkage: asyntopia (with discussion). *J. Roy. Statist. Soc., B*, **57**, 301–370.

Efron,B., Storey,J.D. and Tibshirani,R. (2001) Microarray, empirical bayes methods, and false discovery rates. *Technical Report*. Department of Statistics, Stanford University, see http://www-stat.stanford.edu/∼tibs/research.html.

Fernandez,G., Periaswamy,S. and Sweldens,W. (1996) LiftPack: a software package for wavelet transform using Lifting. In Aldroubi,A., Unser,M. and Laine,A.F. (eds), *Wavelet Applications in Signal and Image Processing IV*, Proc. SPIE 2825, pp. 396–408.

Ferrer,J.L., Roth,M. and Antoniadis,A. (1998) Data compression for diffraction patterns. *Acta Crystallogr D Biol. Crystallogr.*, **54**, 184–199.

Flandrin,P. (1988) Time-frequency and time-scale. *IEEE Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling, Minnesota, Minneapolis*. pp. 77–80.

Flandrin,P. (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theor.*, **38**, 910–917.

Grenfell,B.T., Bjornstad,O.N. and Kappey,J. (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature*, **414**, 716–723.

Hardle,W., Kerkyacharian,G., Picard,D. and Tsybakov,A. (1998) Wavelets, approximation, and statistical applications. *Lecture Notes in Statistics, 129*. Springer, New York.

Hirakawa,H., Muta,S. and Kuhara,S. (1999) The hydrophobic cores of proteins predicted by wavelet analysis. *Bioinformatics*, **15**, 141–148.

Jones,C. (1996) 2-D wavelet packet analysis of structural self-organization and morphogenic regulation in filamentous fungal colonies. In Stocker,R., Jelinek,H., Durnota,B. and Bossomaier,T. (eds), *Complex Systems*, Vol. 96, IOS Press, Amsterdam, pp. 12–23.

Johnson,K.F. and Simon,S.M. (2001) Call to work together on microarray data analysis. *Nature*, **411**, 885.

Johnstone,I.M., Kerkyacharan,G. and Picard,D. (1992) Estimation d'une densité de probabilité par méthode d' ondelettes. *Compt. Rend. Acad. Scie. Paris A*, **315**, 211–216.

Jörnsten,R. and Yu,B. (2000) Comprestimation: microarray images in abundance. *Conference on Information Sciences and Systems, March 15–17*. Princeton University.

Katul,G., Vidakovic,B. and Albertson,J. (2000) Estimating global and local scaling exponents in turbulent flows using wavelet transformations. *Phys. Fluids*, **13**, 241–250.

Klevecz,R.R. (2000) Dynamic architecture of the yeast cell cycle uncovered by wavelet decomposition. *Functional and Integrative Genomics*, **1**, 186–192.

Kyriacou,C.P. and Hall,J.C. (1986) Interspecific genetic control of courtship song production and reception in *Drosophila*. *Science*, **232**, 494–497.

Li,S., Hu,K., Cai,N., Su,W., Xiong,H., Lou,Z., Lin,T. and Hu,Y. (2001) Automatic analysis of image of surface structure of cell-wall-deficient EVC. *Biomed. Mater. Eng.*, **11**, 159–166.

Lindsay,R.W., Percival,D.B. and Rothrock,D.A. (1996) The discrete wavelet transform and the scale analysis of the surface properties of sea ice. *IEEE Trans. Geosci. Remote Sensing*, **34**, 771–787.

Liò,P. and Vannucci,M. (2000a) Wavelet change-point prediction of transmembrane proteins. *Bioinformatics*, **16**, 376–382.

Liò,P. and Vannucci,M. (2000b) Finding pathogenicity islands and gene transfer events in genome data. *Bioinformatics*, **16**, 932–940.

Main,P. and Wilson,J. (2000) Low-resolution phase extension using wavelet analysis. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 1324–1331.

Mallat,S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Mallat,S.G. (1999) *A Wavelet Tour of Signal Processing*. Academic Press, San Diego.

Mandell,A.J., Selz,K.A and Shlesinger,M.F. (1997) Wavelet transformation of protein hydrophobicity sequences suggests their memberships of structural families. *Physica A*, **244**, 254–262.

Mandell,A.J., Owens,M.J., Selz,K.A., Morgan,W.N., Shlesinger,M.F. and Nemeroff,C.B. (1998) Mode matches in hydrophobic free energy eigenfunctions predict peptide-protein interactions. *Biopolymers*, **46**, 89–101.

Morettin,P.A. (1996) From Fourier to wavelet analysis of time series. In Pratt,A. (ed.), *Proceedings in Computational Statistics*. Physica, New York, pp. 111–122.

Morozov,P., Sitnikova,T., Churchill,G., Ayala,F.J. and Rzhetsky,A. (2000) A new method for characterizing replacement rate variation in molecular sequences. *Genetics*, **154**, 381–95.

Murray,K.B., Gorse,D. and Thornton,JM. (2002) Wavelet transforms for the characterization and detection of repeating motifs. *J. Mol. Biol.*, **316**, 341–363.

Myasnikova,E., Samsonova,A., Kozlov,K., Samsonova,M. and Reinitz,J. (2001) Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics*, **17**, 3–12.

Nason,G.P. (1996) Wavelet shrinkage using cross-validation. *J. Roy. Statist. Soc., B*, **58**, 463–479.

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

Nason,G.P., von Sachs,R. and Kroisandt,G. (2000) Wavelet processes and adaptive estimation of the evolutionary spectrum. *J. Roy. Statist. Soc. B*, **62**, 271–292.

Nijhawan,R. (1997) Visual decomposition of colour through motion extrapolation. *Nature*, **386**, 66–69.

Ogden,R.T. and Parzen,E. (1995) Change-point approach to data analytic wavelet thresholding. *Statist. Comput.*, **6**, 93–99.

Ogden,R.T. and Parzen,E. (1996) Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computat. Stat. Data Anal.*, **22**, 53–70.

Percival,D.B. (1995) On estimation of the wavelet variance. *Biometrika*, **82**, 619–631.

Percival,D.B. and Walden,A.T. (2000) *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge.

Raimondo,M. (1998) Minimax estimation of sharp change points. *Ann. Statist.*, **26**, 1379–1397.

Raimondo,M. (2002) Wavelet shrinkage via peaks over threshold. *Intersat*, May, 1–19.

Rzhetsky,A. and Morozov,P. (2001) Markov chain Monte Carlo computation of confidence intervals for substitution-rate variation in proteins. *Pac. Symp. Biocomput.*, **2001**, 203–214.

Schreiner,C.E., Read,H.L. and Sutter,M.L. (2000) Modular organization of frequency integration in primary auditory cortex. *Annu. Rev. Neurosci.*, **23**, 501–529.

Shensa,G. (1992) The discrete wavelet transform: wedding the à trous and Mallat algorithms. *IEEE Trans. signal process.*, **40**, 2464–2482.

Serroukh,A. and Walden,A.T. (2000) The scale analysis of bivariate non-Gaussian time series via wavelet cross-covariance. *Technical Report TR-98-02*. Imperial College, Statistics section.

Sweldens,W. (1996) Wavelets: what next? *Proc. IEEE*, **84**, 680–685.

Taswell,C. (1995) Wavbox 4: a software toolbox for wavelet transforms and adaptive wavelet packet decomposition. In Antoniadis,A. and Oppenheim,G. (eds), *Wavelets and Statistics*, Lectures Notes in Statistics, 103, Springer, New York, pp. 361–375.

Torrance,C. and Compo,G.P. (1998) A practical guide to wavelet analysis with significance and confidence testing. *Bull. Am. Meteor. Soc.*, **79**, 61–78.

Uytterhoeven,G., Van Wulpen,F., Jansen,M., Roose,D. and Bultheel,A. (1998) WAILI: a software library for image processing using integer wavelet transforms. In Hanson,K.M. (ed.), *Medical Imaging: Image Processing, Vol. 3338 of SPIE Proceedings*. pp. 1490–1501.

Vannucci,M. and Lió,P. (2001) Non-decimated wavelet analysis of biological sequences: applications to protein structure and genomics. *Sankhya B*, **63**, 218–233.

Van De Wouwer,G., Weyn,B., Scheunders,P., Jacob,W., Van Marck,E. and Van Dyck,D. (2000) Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei. *J. Microsc.*, **197**, 25–35.

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

Wang,Y. (1995) Jump and sharp cusp detection by wavelets. *Biometrika*, **82**, 385–397.

Whitcher,B., Guttorp,P. and Percival,D.B. (2000) Wavelet analysis of covariance with application to atmospheric time series. *J. Geophysical Research—Atmospheres*, **105**, 14941–14962.