A wavelet based method for the estimation of the power spectrum from irregularly sampled data

Osmo Kaleva
Department of Mathematics
Tampere University of Technology

Heimo Ihalainen
Department of Measurement and Information Technology
Tampere University of Technology

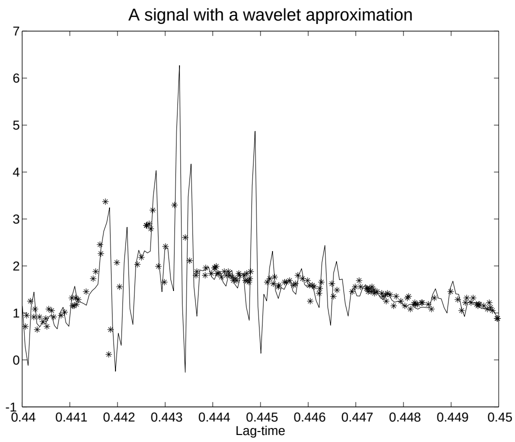
Pentti Saarenrinne
Department of Energy and Process Engineering
Tampere University of Technology

Abstract

The problem of estimating the power spectrum from irregularly sampled data appears in many technical applications. Recently it has been studied by many authors. For the current status of the problem see for instance Benedict *et.al.* (1998) and Ware (1998) and references therein.

Most techniques for spectrum estimation can be classified into one of the following groups: slotting technique and cosine transform, direct transformation and reconstruction with uniform resampling plus FFT.

In this paper we present a method of the third group above. We recontruct a signal using a wavelet approximation, resample the approximated signal uniformly and finally compute the power spectrum with the standard FFT. The figure below gives an example of an irregularly sampled data and its wavelet approximation.



A good exposition on wavelets is for example Mallat (1998).

1 Introduction

The problem of estimating the power spectrum from irregularly sampled data appears in many technical applications. Recently it has been studied by many authors. For the current status of the problem see for instance Benedict *et.al.* (1998) and Ware (1998) and references therein.

According to Benedict *et.al.* (1998) most techniques for spectrum estimation can be classified into one of the following groups: slotting technique and cosine transform, direct transformation and reconstruction with uniform resampling plus FFT. In this paper we present a method of the third group above. We recontruct a signal using a wavelet approximation, resample the approximated signal uniformly and finally compute the power spectrum with the standard FFT.

The paper is organized as follows. In section 2 we give some basic properties of wavelets and there use in approximation. Then we discuss some implementation details of the method. We use MATLAB as a computational tool. In section 4 we give the results obtained and compare them to the results obtained by some classical methods. First of these methods is a correlation based method due to van Maanen. Here the starting point is the slotting technique with local normalization. Van Maanen and Oldenziel (1998) recommend curve–fitting the locally normalized ACCF in order to remove variability in the ASDF estimates almost completely. They have developed an eight–parameter autocorrelation model, which is extremely flexible and can be analytically Fourier transformed.

For reducing variance in the slotting technique *a local normalization* has been introduced by Tummers and Passchier (1996) and van Maanen and Tummers (1996). Here an ACCF normalized by a variance estimate particular to each slot is used as the basis for the cosine transform. While this estimate has been shown to have significantly lower variance for small lag times than basic slotting technique normalized by $C_{uu}(0)$, the variance at large lag times is unchanged. This improves the spectral estimator.

Second test methods are classical approximation methods, where the signal is approximated by an evenly sampled signal. Sample and hold, linear interpolation and exponentially settled sample and hold are the most common methods used for the approximation, see for instance Adrian and Yao (1987). We will apply the sample and hold and linear interpolation method to our test data.

The final section is devoted to conclusions.

2 Mathematical background

2.1 Basic properties of wavelets

In this section we shall give some basic properties of wavelets. We begin with multiresolution analysis, which forms a mathematical foundation for studying a signal at different scales. For proofs and additional information a reader should consult Daubechies (1992), Hernandez and Weiss (1996) and Strang and Nguyen (1996).

Definition 1. A multiresolution analysis (MRA) is a sequence of closed subspaces $V_j \subset L^2(\mathbb{R}),\ j \in \mathbb{Z}$, with the following properties:

- **(i)** $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$, i.e $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset L^2(\mathbb{R})$,
- **(ii)** $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$ for all $j \in \mathbb{Z}$,
- **(iii)** $\cap_{j\in\mathbb{Z}} V_j = \{0\}$ and $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$,
- **(iv)** there exists an $\varphi \in V_0$ such that $\{\varphi(x-k) \mid k \in \mathbb{Z}\}$ is an orthonormal basis for V_0.

Given a multiresolution analysis we denote by W_j the orthogonal complement of V_j in V_{j+1}. Hence $V_{j+1} = V_j \oplus W_j$ for all $j \in \mathbb{Z}$. Using the property (iii) of an MRA we obtain the following representations:

$$V_{j+1} = \bigoplus_{k=-\infty}^{j} W_k,\ V_{j+1} = V_{j_0} + \bigoplus_{k=j_0}^{j} W_k \text{ for } j \geq j_0 \text{ and } L^2(\mathbb{R}) = \bigoplus_{k=-\infty}^{\infty} W_k.$$

In addition a repeated application of (ii) of an MRA shows that $\{2^{\frac{j}{2}}\varphi(2^j x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_j. The basic principle of wavelet analysis is that there is a function ψ such that $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0. It follows that $\{2^{\frac{j}{2}}\psi(2^j x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_j.

The function φ is called *a scaling function* and ψ *a basic wavelet*. In the following we use notations

$$\varphi_{j,k}(x)=2^{\frac{j}{2}}\varphi(2^j x-k) \text{ and } \psi_{j,k}(x)=2^{\frac{j}{2}}\psi(2^j x-k).$$

Now $\varphi_{j,k}\in V_j$ and $\psi_{j,k}\in W_j$. Since V_j and W_j are orthogonal and all the spaces W_j are mutually orthogonal, we have the following orthogonality relations:

$$\int_{-\infty}^{\infty}\varphi_{j,k}(x)\varphi_{j,l}(x)\,dx=\delta_{k,l},$$

$$\int_{-\infty}^{\infty}\psi_{i,k}(x)\psi_{j,l}(x)\,dx=\delta_{i,j}\delta_{k,l},$$

$$\int_{-\infty}^{\infty}\varphi_{i,k}(x)\psi_{j,l}(x)\,dx=0 \text{ for } j\geq i,$$

where $i,j,k,l\in\mathbb{Z}$ and $\delta_{i,j}$ is the Kronecker delta. This means that scaling functions are orthogonal within a scale whereas wavelets are ortogonal across scales.

It follows that if $f\in L^2(\mathbb{R})$ and $P_j f$ denotes the orthogonal projection of f onto V_j then

(1) $$P_j f(x)=\sum_{k=-\infty}^{\infty}c_{j,k}\varphi_{j,k}(x)$$

and alternatively

(2) $$P_j f(x)=\sum_{k=-\infty}^{\infty}c_{0,k}\varphi_{0,k}(x)+\sum_{l=0}^{j-1}\sum_{k=-\infty}^{\infty}d_{l,k}\psi_{l,k}(x) \text{ for } j\geq 0,$$

where

(3) $$c_{j,k}=\int_{-\infty}^{\infty}f(x)\varphi_{j,k}(x)\,dx \text{ and } d_{l,k}=\int_{-\infty}^{\infty}f(x)\psi_{l,k}(x)\,dx.$$

Kelly *et.al.* (1994) have shown that under certain assumptions on the scaling function and the basic wavelet both of these expansions converge pointwise to f almost everywhere as j tends to infinity. Furthermore analogously to FFT wavelet approximation also has Gibbs effect, see Kelly (1996).

2.2 Compactly supported wavelets

Recall that we want to reconstruct f from an irregular sample. Hence we compute $P_j f$ for some $j\geq 0$ and use the approximation $f\approx P_j f$. In consequence we have to compute numerically inner products (3). Since we only have a finite sample it is of advantage if the scaling function and basic wavelet are compactly supported, which means that they are zero outside some closed bounded interval.

For the approximation we can use either the presentation (1) or (2). However in (1) all the dilations are of the same scale. Hence the basic functions are translations of one function. It follows that it is not possible to utilize the irregularity of the original sample. Instead (2) allows us to add details of different scales into the projection, where there is data available.

So we construct a compactly supported scaling function φ and basic wavelet ψ, compute their values at a regular mesh, compute innerproducts (3) and finally form $P_j f(x)$.

Since $\varphi \in V_0 \subset V_1$ and $\varphi_{1,k}$ is an orthonormal basis for V_1 we have *the dilation equation*

$$\varphi(x) = \sqrt{2} \sum_k c_k \varphi(2x-k), \text{ where } c_k = \sqrt{2} \int_{-\infty}^{\infty} \varphi(x)\varphi(2x-k)\,dx.$$

We see that if φ is compactly supported, then only finite number of c_k's are nonzero. On the other hand each solution of the the dilation equation

$$\varphi(x) = \sqrt{2} \sum_{k=0}^{K} c_k \varphi(2x-k), \text{ where } c_0, c_K \neq 0, \tag{4}$$

is continuous, if $K \geq 3$, and compactly supported with support $[0, K]$. Daubechies and Lagarias (1991), (1992) have shown that under certain conditions on the coefficients c_k the solution of the dilation equation (4) is unique up to a scale factor and is a scaling function of an MRA. Furthermore it is known that if the solution $\varphi \in C^m$ then the length of the support of φ is at least $m+2$. It follows that it is not possible to construct wavelet bases generated by a compactly supported C^∞–function.

It appears that if φ is as above then

$$\psi(x) = \sqrt{2} \sum_{k=0}^{K} d_k \varphi(2x-k), \text{ with } d_k = (-1)^k c_{K-k},\ k = 0, \ldots, K \tag{5}$$

is a basic wavelet for the MRA generated by φ and it has the same regularity properties as φ. In addition the values of φ and ψ are easily computed at dyadic rationals of $[0, K]$ using a *cascade algorithm*. It is easy to see that the support of $\varphi_{j,k}$ and $\psi_{j,k}$ is the interval $I_{j,k}$, where

$$I_{j,k} = [\frac{k}{2^j}, \frac{k+K}{2^j}].$$

Daubechies (1992) was first to give a family of coefficients, which generate an MRA with a compactly supported scaling function and basic wavelet. It appears that the number of nonzero coefficients must be even. Furthermore if $K = 2N-1$ then for the Daubechies wavelets we have

$$\int \varphi(x)\ dx = 1 \quad \text{and} \quad \int (x+1-N1)^j \psi(x)\ dx = 0 \text{ for all } j = 0, \ldots, N-1.$$

In other words the basic wavelet has N vanishing moments when translated to $[-N+1, N]$.

3 Implementation details

The crucial point of the method is the evaluation of the inner products (3). Here the integrands are given at an irregular mesh, so we have to use an integration method, which applies to this case. The trapezoidal rule is generalized easily also to irregular case. Alternatively we may use a generalized Simpson rule or cubic integration rule as explained in Davis and Rabinowitz (1975). The Simpson method consists of fitting an interpolating parabola to three consecutive points and integrating the average of the overlapping functions.

As seen in the following figures, the approximation depends on the integration method. Figure 1 gives an approximation with trapezoidal integration and Figure 2 with quadratic integration. The integration produces plenty of peaks, which clearly are artefacts due to the method. In order to improve the power spectrum estimate, we use a statistical method to replace these outliers. This is done as follows. First compute the first and third quartile, denoted by q_1 and q_3 respectively. The interquartile range is given by $\mathrm{iqr} = q_3 - q_1$. Now the observations below $q_1 - 1.5\,\mathrm{iqr}$ and above $q_3 + 1.5\,\mathrm{iqr}$ are identified as mild outliers. Respectively the observations below $q_1 - 3\,\mathrm{iqr}$ and above $q_3 + 3\,\mathrm{iqr}$ are extreme outliers. For further information see for example Milton and Arnold (1995).

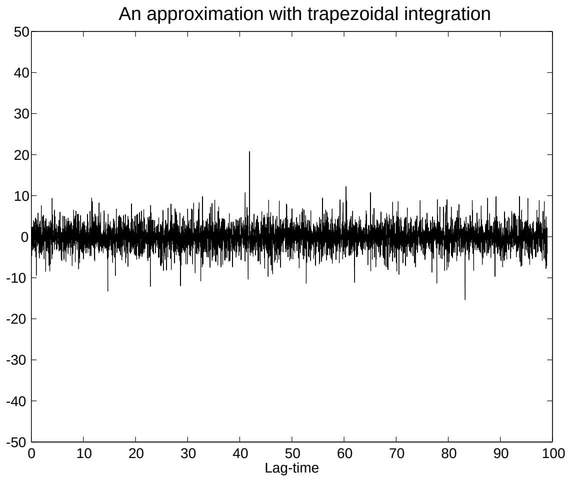


Figure 1: A wavelet approximation of a simulated data set with trapezoidal integration

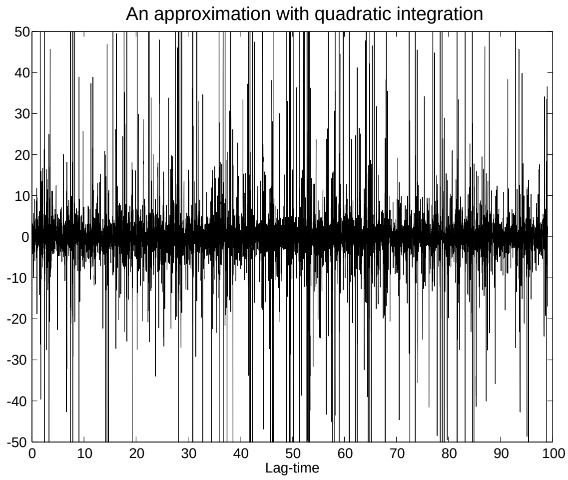


Figure 2: A wavelet approximation of a simulated data set with quadratic integration

The equalization changes the variance of the data. Hence, before computing the spectrum estimate, we renormalize the equalized data such that its variance equals to the variance of the original data.

4 Results obtained

For testing the wavelet method we use a simulated data set due to Nobach. The signal has mean velocity 0 and velocity variance equal to 1. LDA data rate is 100Hz and bias as well as noise are included. The theoretical spectrum of the signal is known.

We also use a measured signal. The probe volume was situated 10 mm after an impeller trailing edge and 5 mm from chamber wall of a model water turbine. The tip leakage and the trailing wakes of the impeller blades produce a periodic highly fluctuating signal.

The following figures give a part of a measured data with its approximations. In Figure 3 we give an interpolating cubic spline and in Figure 4 Daubechies wavelet approximation supported on [0,3]. Integration used is the quadratic one. We see that long gaps in the data are problematic with the wavelet approximation as well as traditional nonlinear interpolation.

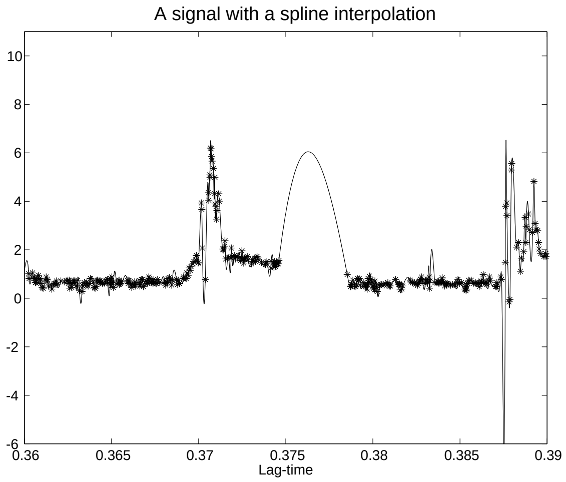


Figure 3: A part of the measured data set with cubic spline interpolation

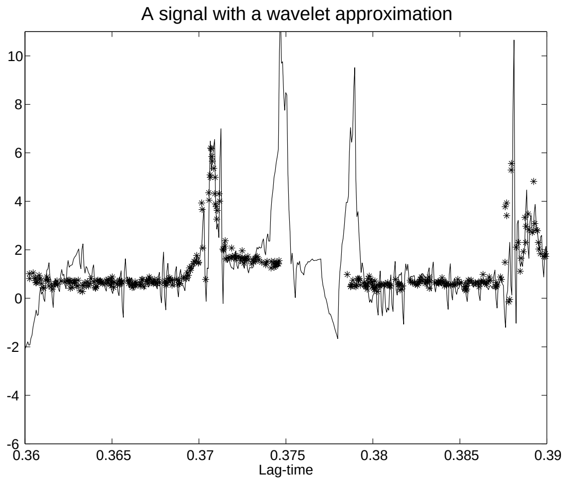


Figure 4: A part of the measured data set with Daubechies wavelet approximation

The peaks, which are due to gaps in the data and the integration method used, are clearly seen if we plot the whole data with its wavelet approximation. Figure 5 gives a Haar wavelet approximation. To demonstrate how these peaks depend also on the wavelet used in Figure 6 we plot the data with a coiflet approximation. Coiflets are wavelets, where also the scaling function has vanishing moments. They are more symmetric than Daubechies wavelets and have longer support. In general if we use a coiflet, where the scaling function has $N-1$ vanishing moments and the basic wavelet has N vanishing moments, then the length of its support is $6L-1$, where $N = 2L$. These wavelets were also developed by Ingrid Daubechies, see Daubechies (1992).

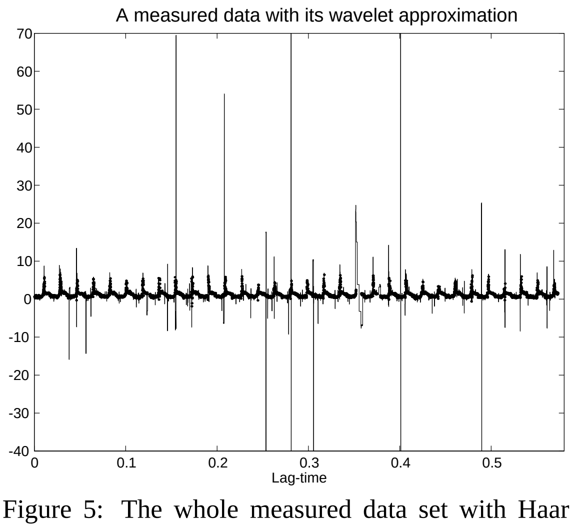


Figure 5: The whole measured data set with Haar wavelet approximation

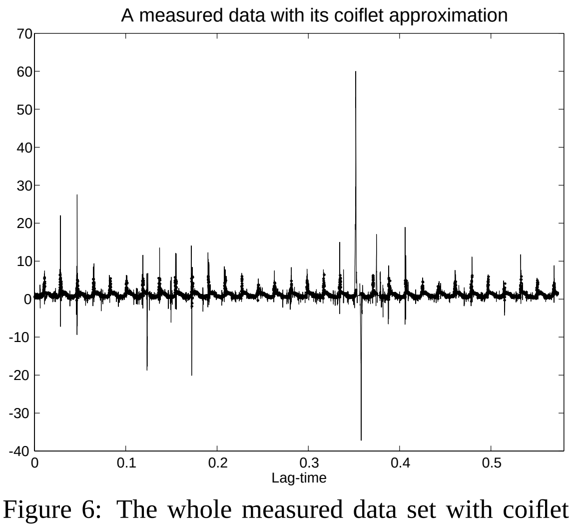


Figure 6: The whole measured data set with coiflet approximation

For the simulated data the wavelet method gives the following spectra. Figure 7 is the theoretical spectrum. Then Figure 8 gives the mean of five wavelet spectra from different realizations of the data and finally Figures 9 and 10 give a wavelet and coiflet spectra with oversampling 6. In all cases we use quadratic integration and replace extreme outliers with boundary values.

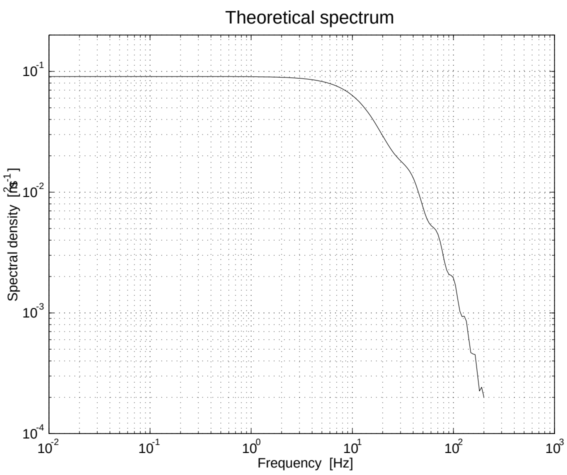


Figure 7: Theoretical spectrum of simulated data set

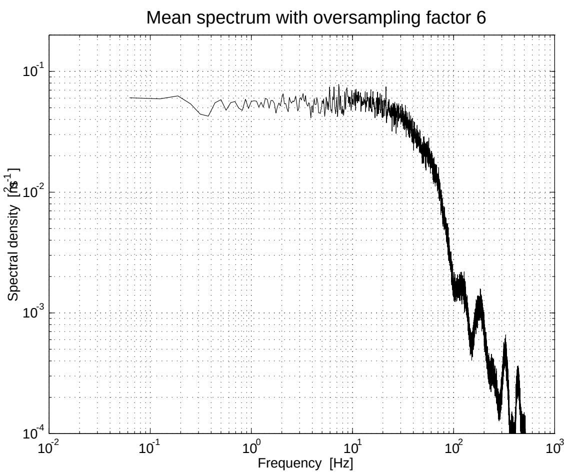


Figure 8: Mean of five wavelet spectra from 6 times oversampled simulated data sets

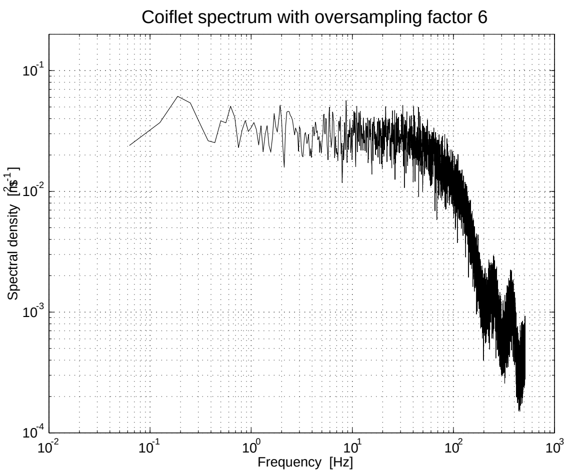


Figure 9: Coiflet spectrum from 6 times oversampled simulated data set

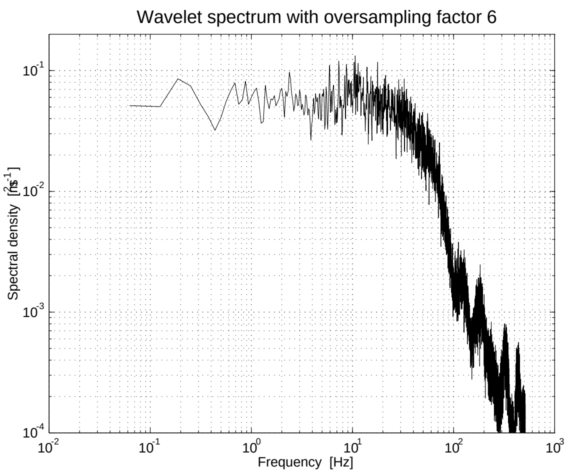


Figure 10: Wavelet spectrum from 6 times oversampled simulated data set

In Figures 11,12 and 13 we give spectra obtained by linear and cubic spline interpolation method. Then in Figure 14 there is a spectrum obtained by correlation method.

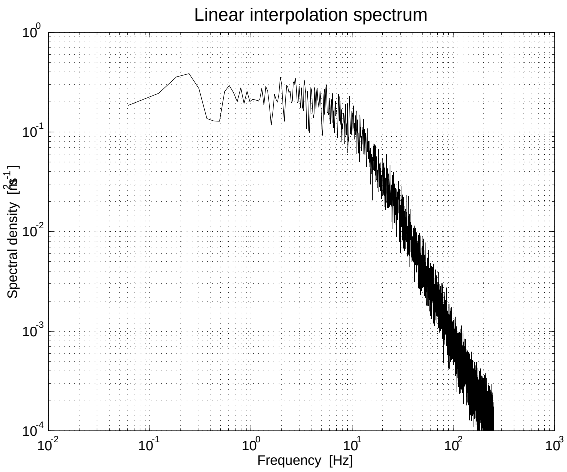


Figure 11: Linear interpolation spectrum of simulated data set

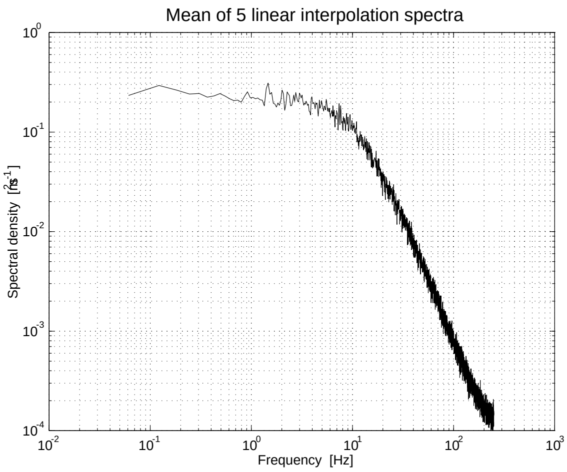


Figure 12: Mean of five linear interpolation spectra of simulated data sets

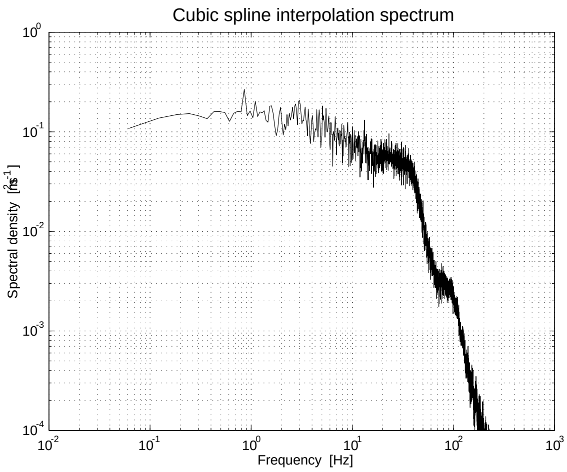


Figure 13: Cubic spline interpolation spectrum of simulated data set

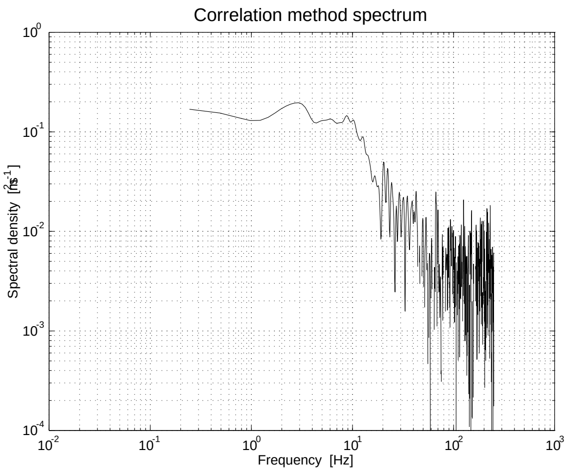


Figure 14: Correlation method spectrum of simulated data set

For the measured data we obtain the following results. In Figures 15 and 16 there are a wavelet and coiflet spectrum with oversampling factor 2. Then Figures 17 and 18 give 6 times oversampled spectrum with Haar wavelet and quadratic as well as cubic integration.

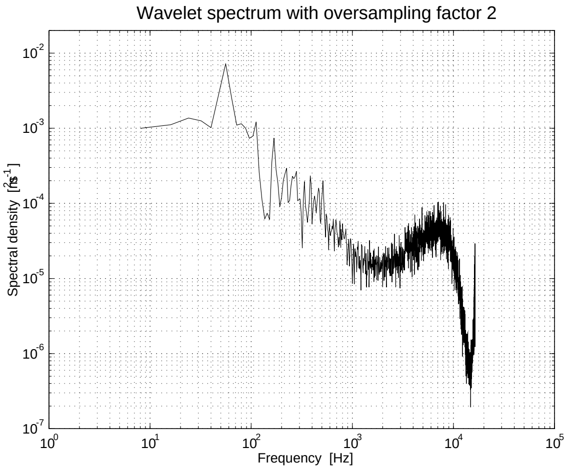


Figure 15: Wavelet spectrum from 2 times oversampled measured data set

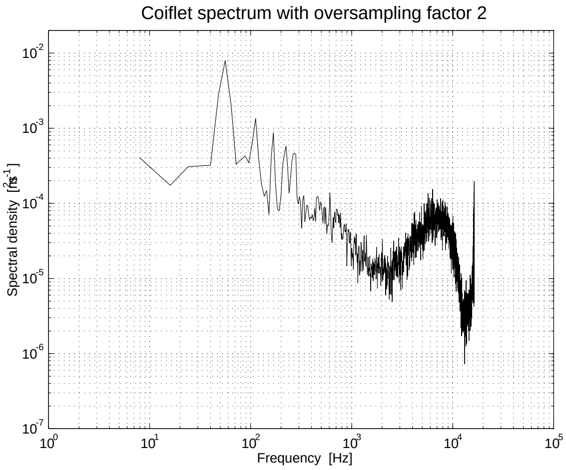


Figure 16: Coiflet spectrum from 2 times oversampled measured data set

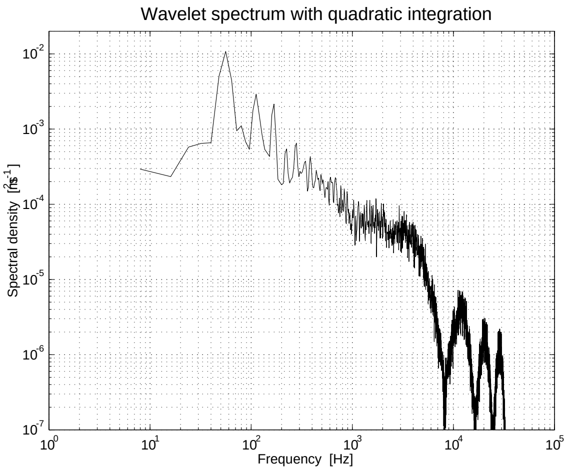


Figure 17: Wavelet spectrum from 6 times oversampled measured data set with quadratic integration

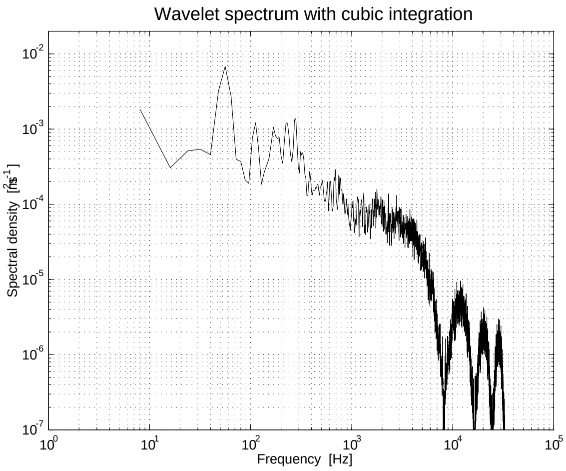


Figure 18: Wavelet spectrum from 6 times oversampled measured data set with cubic integration

Finally we give nearest neigbor, linear and cubic interpolation spectra for the measured data set.

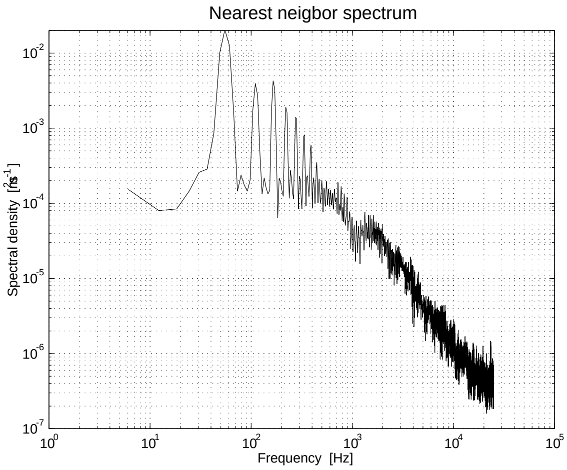


Figure 19: Nearest neigbor spectrum of measured data set

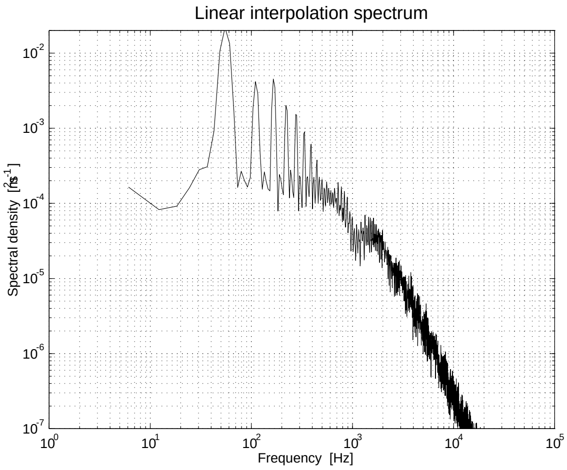


Figure 20: Linear interpolation spectrum of measured data set

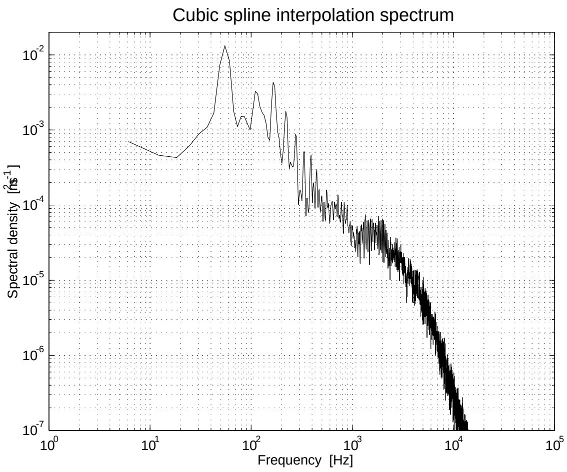


Figure 21: Cubic spline interpolation spectrum of measured data set

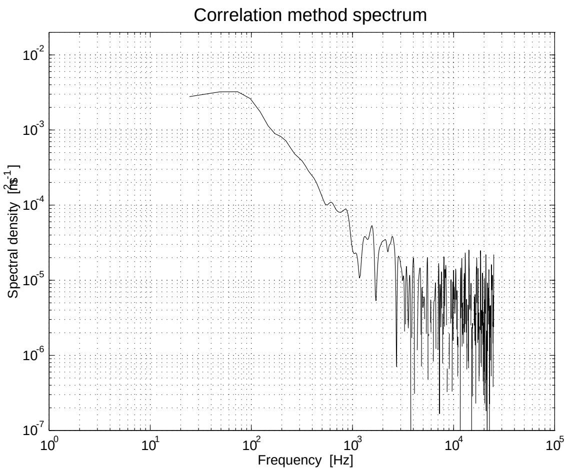


Figure 22: Correlation method spectrum of measured data set

5 Conclusions

In this paper we have presented a wavelet based method for spectrum estimation. The evidence shows that the method gives similar results as classical resampling methods. For simulated data the level of flat part of the spectrum corresponds the theoreticalspectrum. Also the decrease of the spectrum is in accordance with the theory. For measured data we obtain the expected peaks in the spectrum. It seems that wavelet spectrum is less noisy than other interpolation spectra.

In addition we have seen that wavelet method yields plenty of artificial peaks in the signal approximation, which should be removed. The spectrum also depends on the wavelet used and the oversampling factor in resampling. In conclusion we think that the wavelet method gives reasonable results and reserves further study.

References

[1] Adrian, R. J. and Yao, C. D. (1987). *Power spectra of fluid velocities measured by laser Doppler velocimetry,* Exp. in Fluids **5**, 17-28.

[2] Benedict, L. H., Nobach, H. and Tropea, C. (1998). *Benchmark tests for the estimation of power spectra from LDA signals,* Proc. Ninth Int. Symposium on Applications of Laser Techniques to Fluid Mechanics, Vol II, Lisbon, Portugal.

[3] Daubechies, Ingrid (1992). **Ten Lectures on Wavelets**, SIAM.

[4] Daubechies, Ingrid and Lagarias, Jeffrey C. (1991). *Two-scale difference equations I. Existence and global regularity of solutions,* SIAM J. Math. Anal. **22**, 1388–1410.

[5] Daubechies, Ingrid and Lagarias, Jeffrey C. (1992). *Two-scale difference equations II. Local regularity, infinite product of matrices and fractals*, SIAM J. Math. Anal. **23**, 1031–1079.

[6] Davis, P. and Rabinowitz, P. (1975). **Methods of Numerical Integration**, Academic Press.

[7] Eugenio Hernández, Eugenio and Weiss, Guido (1996). **A First Course on Wavelets**, CRC Press.

[8] Kelly, Susan E., Kon, Mark A. and Raphael, Louise A. (1994). *Pointwise convergence of wavelet expansions*, Bull. Amer. Math. Soc. **30**, 87–94.

[9] Kelly, Susan E. (1996). *Gibbs Phenomenon for Wavelets*, Appl. Comput. Harmonic Analysis **3**, 72–81.

[10] van Maanen, H. R. E. and Oldenziel, A. (1998). *Estimation of Turbulence Power Spectra from Randomly Sampled Data by Curve-Fit to the Autocorrelation Function Applied to Laser Doppler Anemometry*, Meas. Sci. Technol. **9**, 458–467.

[11] van Maanen, H. R. E. and Tulleken, H. J. F. A. (1994). *Application of Kalman Reconstruction to Laser-Doppler Anemometry Data for Estimation of Turbulent Velocity Fluctuations*, Proc. Seventh Int. Symp. on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal.

[12] van Maanen, H. R. E. and Tummers M. (1996). *Estimation of the autocorrelation function of turbulent velocity fluctuations using the slotting technique with local normalisation*, Proc. Eighth Int. Symp. on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal.

[13] Mallat, S. (1998). **A Wavelet Tour of Signal Processing**, Academic Press.

[14] Milton, J. S. and Arnold, Jesse C. (1995). **Introduction to probability and Statistics**, McGraw Hill.

[15] Nobach, Holger. *LDA Benchmark Generator II*, http://www-nt.e-technik.uni-rostock.de/˜nobach/

[16] Strang, Gilbert and Nguyen, Truong (1996). **Wavelets and Filter Banks**, Wellesley-Cambridge Press.

[17] Tummers, M. J. and Passchier, D. M. (1996). Spectral Analysis of Individual Realization LDA Data, Report LR 808, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands.

[18] Tummers, M. J. and Passchier, D. M. (1996). *Spectral Estimation Using a Variable Window and the Slotting Technique with Local Normalization*, Meas. Sci. Technol. **7**, 1541-1546.

[19] Ware, Anthony F. (1998). *Fast approximate Fourier transform for irregularly spaced data,* SIAM Rev. **40**, 838–856.