GEOPHYSICAL RESEARCH LETTERS, VOL. 23, NO. 4, PAGES 407-410, FEBRUARY 15, 1996

# Filtering non-stationary geophysical data with orthogonal wavelets

Frédérique Moreau, Dominique Gibert, and Ginette Saracco

Géosciences Rennes - CNRS/INSU, Université de Rennes 1, France.

**Abstract.** A filtering method based on both orthogonal wavelet decomposition and chi-squared statistics is proposed to clean non-stationary signals embedded in a gaussian white noise. An application to a time series of thermistance data recorded in an underground quarry illustrates the interest of the technique.

## Introduction

The wavelet theory (*Grossmann and Morlet* [1984]; see *Meyer* [1990], *Daubechies* [1992], and *Holschneider* [1995] for reviews) constitutes a powerful framework to process and analyze non-stationary geophysical signals [*Foufoula-Georgiou and Kumar*, 1994]. In particular, the important problem of signal denoising has recently been addressed by means of both continuous [e.g. *Mallat and Hwang*, 1992] and orthogonal wavelet decompositions [*Donoho and Johnstone*, 1994; *Saito*, 1994]. Denoising needs to distinguish the noise from the signal and, depending on the particular models assumed for these components, distinct algorithms may be proposed. In the present study we address the particular issue of cleaning signals embedded in gaussian white noise through orthogonal wavelet decomposition. We propose a special-purpose filtering criterion based on a Chi-Square Thresholding (CST), and compare its performance to those of two general-purpose and popular threshold criteria: the Akaike's Information Criterion (AIC) [*Akaike*, 1965] and the Minimum Description Length (MDL) [*Rissanen*, 1978; *Wax and Kailath*, 1985]. Examples with synthetic tests and real geophysical data are given.

## Denoising Signals with Wavelets

An orthogonal basis reads $\left\{2^{-m/2}\psi\left(2^{-m}t-n\right)\right\}$ with $(m,n)\in\mathbb{Z}^2$ where the analyzing wavelet $\psi(x)$ is an oscillating function localized around the origin. The wavelet coefficients of a discrete signal may be efficiently computed via a pyramidal algorithm *Mallat* [1989] and provide a way to examine the information content of the original signal in the time-scale half-plane. If the input signal $s$ counts $K=2^N$ values, the first $2^{N-1}$ wavelet coefficients correspond to the finest scale available and fixed by the sampling interval, the next $2^{N-2}$ coefficients are for the immediately upper scale (i.e. twice the finest scale), and so on until the last coefficient which corresponds to the largest scale available (i.e. the length of the signal). A filtered signal is obtained by performing an inverse wavelet transform with a subset of the initial $K$ wavelet coefficients.



Paper number 96GL00155
0094-8534/96/96GL-00155$03.00

### Filtering Criteria

Since each of the $K$ coefficients may be either rejected or retained, the set $\mathcal{A}=\{s_l\}$ of the a priori possible filtered signals possesses $2^K$ elements, and a filtering criterion is necessary to decide which of the $s_l$'s is the denoised signal. Of course, the final choice depends on the problem at hand, and for the particular case of gaussian-white and zero-mean noise we propose the CST criterion whose "best" signal $s_{CST}$ verifies

$$CST\left(s_{CST}\right)\simeq p_0 \qquad (1)$$

with

$$CST\left(s_l\right)\equiv prob\left(\chi_n^2\leq\frac{\left\|s-s_l\right\|^2}{\sigma^2}\right) \qquad (2)$$

where $\chi_n^2$ is the Chi-square probability function with $n$ degrees of freedom. The variance, $\sigma^2$, of the noise is assumed a priori known, and

$$\left\|s-s_l\right\|^2\equiv\sum_{i=1}^{K}\left(s_i-s_{l,i}\right)^2. \qquad (3)$$

The probability threshold $p_0$ in (1) fixes the level of risk accepted that some noise remains in $s_{CST}$.

In order to show the reader that the choice of a particular filtering criterion is critical and strongly depends on the problem at hand, we consider two general-purpose criteria: the Akaike's Information Criterion (AIC) [*Akaike*, 1965] and the Minimum Description Length (MDL) [*Rissanen*, 1978]. These criteria are often used to choose among a collection of a priori models, like for instance ARMA models in signal processing [*Wax and Kailath*, 1985], with different complexities $k$. The MDL criterion has been used by *Saito* [1994] in the context of wavelet filtering. When applied to the gaussian case, the AIC and MDL criteria respectively retain the signals $s_{AIC}$ and $s_{MDL}$ such that

$$AIC\left(s_{AIC}\right) = \min\left[AIC\left(s_l\right)\right], \qquad (4)$$

$$MDL\left(s_{MDL}\right) = \min\left[MDL\left(s_l\right)\right], \qquad (5)$$