Wavelet-Based Trend Detection and Estimation

Peter F. Craigmile[1] **and Donald B. Percival**[2,3].

[1]Department of Statistics, Box 354322, University of Washington, Seattle. WA 98195–4322.

[2]Applied Physics Laboratory, Box 355640, University of Washington, Seattle. WA 98195–5640.

[3]MathSoft Inc., 1700 Westlake Avenue North, Suite 500, Seattle. WA 98109–3044.

Suppose that $Y(t)$ is the value of an observable time series at time t, where t can take on a continuum of values. Scientists often speak of $Y(t)$ as consisting of two quite different unobservable parts, namely, a so-called trend $T(t)$ and a stochastic component $X(t)$ (sometimes called the noise process) such that

$$Y(t) = T(t) + X(t), \tag{1}$$

where it is assumed that the expected value of $X(t)$ is zero. There is no commonly accepted precise definition for trend, but it is usually spoken of as a nonrandom (deterministic) smooth function representing long-term movement or systematic variations in a series (for example, Priestley (1981) refers to a trend as "…a tendency to increase (or decrease) steadily over time …[or to] fluctuate in [a] periodic manner," while Kendall (1973) states that "the essential idea of trend is that it shall be smooth …"). The problem of testing for or extracting a trend in the presence of noise is thus somewhat different from the closely related problem of estimating a function or signal $S(t)$ buried in noise. While the model $Y(t) = S(t) + X(t)$ has the same form as Equation (1), in general $S(t)$ is not constrained to be smooth and thus can very well have discontinuities and/or rapid variations.

The detection and estimation of trend in the presense of stochastic noise arises in a number of important environmental applications (one example of considerable interest is whether or not there has been an upward trend in the northern hemisphere temperature measurements over the past century due to global warming). The recent advent of wavelets as a tool for time series analysis has sparked interest in handling trend via this approach. A wavelet analysis is a transformation of $Y(t)$ in which we obtain two types of coefficients, namely, wavelet coefficients and scaling coefficients (these are sometimes referred to as, respectively, mother wavelet and father wavelet coefficients). Together these coefficients are fully equivalent to the original time series because we can use them to reconstruct $Y(t)$. Wavelet coefficients are related to changes of averages over specific scales (usually taken to be powers of two), whereas scaling coefficients can be associated with averages on a specified scale. Since the scale that is associated with scaling coefficients is usually fairly large, the information that these coefficients capture agrees well with the notion of trend. The general idea behind trend analysis with wavelets is to associate the scaling coefficients with the trend $T(t)$ and the wavelet coefficients (particularly those at the smallest scales) with the noise component $X(t)$.

To carry out a wavelet analysis, let $\phi(t)$ denote a scaling function. This function satisfies a 'two scale' relationship

$$\phi(t) \equiv \sum_{k\in\mathbb{Z}} c_k \phi(2t-k)$$

and is such that the $\phi(t-k)$'s for $k \in \mathbb{Z}$ form an orthonormal set of functions; i.e., $\int \phi(t-k)\phi(t-l)\,dt$ is unity if $k = l$ and is zero otherwise (here $\mathbb{Z}$ is the set of integers). The wavelet function is then defined by the equation

$$\psi(t) \equiv \sum_{k\in\mathbb{Z}} (-1)^k c_{1-k}\, \phi(2t-k).$$

Letting $\phi_{j,k}(t) \equiv 2^{-j/2}\, \phi(2^{-j}t-k)$ and $\psi_{j,k}(t) \equiv 2^{-j/2}\, \psi(2^{-j}t-k)$, the continuous wavelet transform of a square integrable time series $Y(t)$ consists of the scaling and wavelet coefficients

$$\alpha_{J,k} = \int \phi_{J,k}(t)\, Y(t)\, dt \text{ and } \beta_{j,k} = \int \psi_{j,k}(x)\, Y(t)\, dt, \qquad j = J, J+1, \ldots, \tag{2}$$

where $k \in \mathbb{Z}$. We can reconstruct $Y(t)$ from these transform coefficients using

$$Y(t) = \sum_{k\in\mathbb{Z}} \alpha_{J,k}\, \phi_{J,k}(t) + \sum_{j=-\infty}^{J} \sum_{k\in\mathbb{Z}} \beta_{j,k}\, \psi_{j,k}(t).$$

In practice we observe a time series at a finite number of times t_n; i.e., we have

$$Y(t_n) = T(t_n) + X(t_n), \qquad n = 0, \ldots, N-1, \tag{3}$$

where we only require that $t_n < t_{n+1}$; i.e., the spacing between observations can be irregular. If we further restrict the times to be regularly spaced so that $t_n = n\Delta t + t_0$ for some sampling interval $\Delta t > 0$, we can make use of the discrete wavelet transform (DWT), which (to a certain degree of approximation) can be regarded as a discretisation of Equation (2). To be precise, let $\mathcal{W}$ denote a orthonormal DWT matrix of order $N \times N$ and $\mathbf{Y} \equiv [Y(t_0), \ldots, Y(t_{N-1})]'$. Then the wavelet coefficients are given by $\mathbf{W} = \mathcal{W}\mathbf{Y}$. We can decompose these coefficient as $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_J, \mathbf{V}_J]'$, where

- $\mathbf{W}_j$ are the $N/2^j$ wavelet coefficients, associated with changes in averages on scale 2^{j-1} and related to times spaced 2^j units apart; and
- $\mathbf{V}_J$ are the $N/2^J$ scaling coefficients, associated with averages on scale 2^J and related to times spaced 2^J units apart.

We will restrict our discussions here to two distinct cases, namely, that of uncorrelated and correlated error processes. In the uncorrelated case, since the wavelet transform is orthonormal, the resulting wavelet coefficients are uncorrelated, and, assuming that the errors have constant variance, all wavelet coefficients across different scales will have a constant variance too.

Antoniadis, Gregoire, and McKeague (1994) consider uncorrelated noise in their model, adapting the wavelet estimator so that it works like a kernel estimator. The advantage of this is that we can then formulate an estimator based on irregularly sampled data. As is common with other methods, the authors in effect use the scaling coefficients to estimate the trend. Given observations $Y(t_n)$, the estimate of $T(t)$ is given by

$$\widehat{T}(t) = \sum_{n=0}^{N-1} Y(t_n) \int_{A_n} E_J(t,s)\, ds,$$

where the A_n's are a set of intervals such that (i) their union forms a partitioning of an interval covering all the observations times t_n and (ii) $t_n \in A_n$; and

$$E_J(t,s) \equiv 2^{-J} \sum_{k \in \mathbb{Z}} \phi(2^{-J}t - k)\phi(2^{-J}s - k).$$

In their statistical analysis of the properties of the estimator, they find a bias/variance relationship dependent on the number of scales, and the wavelet function chosen. The variance of the trend estimate can be unstable in certain cases. The authors show that cross-validation is a useful tool for choosing J (one could conceivably also select the wavelet basis via cross-validation).

A more interesting situation is when we observe trend plus correlated noise. Under certain models and choice of wavelet function, the wavelet transform decorrelates the noise process and allows us to simplify the statistical analysis involved.

Brillinger introduces his method of trend assessment via wavelets in two papers (Brillinger 1994; Brillinger 1996). Suppose we observe a time series of the form given by Equation (3) with regular spacing between observations. We assume that, as we increase the sample size N, we obtain more information about the trend; i.e., as we increase N by a factor of, say, two, we decrease Δt by a half so that we sample more finely over a fixed interval of time. By assuming some regularity for the trend and by limiting how dependent and near to Gaussianity the noise process is, Brillinger obtained an estimate of trend based upon the scaling coefficients and a certain number of important large scale wavelet coefficients. The estimate thus captures the longer variations in the time series, matching our notion of trend. The author considers the asymptotic properties of such an estimator and shows that the bias is related to the level J we choose to use and that the bias decreases with sample size. The covariance structure of the estimator is related to the spectrum of the process at zero frequency and to the shape of the wavelet function used. The author also shows how his work extends to, e.g., spatial or point processes.

Gilbert (1999) considers a test for the onset of trend. His test evaluates the null hypothesis that $T(t) = \mu$ for some constant μ and all t versus the alternative hypothesis that $T(t) = \mu$ only over some initial stretch of the time series, after which the trend function has a slope that is either (i) nonnegative and nondecreasing or (ii) nonpositive and nonincreasing (more precisely, the first and second derivatives of $T(t)$ are either both nonnegative or both nonpositive). One special case of the alternative hypothesis is a time series whose mean value is initially constant, after which it increases linearly. His test is based on approximating the trend in terms of the scaling function using a least squares fit and exploits the fact that $T(t)$ is quite different before and after the onset of trend under the alternative hypothesis. The author illustrates his test statistic via Monte Carlo studies and finds that it works well in the case of first order autoregressive noise whose unit lag autocorrelation ranges between -1 and about 0.6.

Craigmile, Percival, and Guttorp (2000) consider trend estimation under a specific model (although it can conceivably be extended quite generally). They consider observing a polynomial trend plus Gaussian fractionally differenced (FD) noise and apply the DWT to separate a time series into pieces that can be used to estimate both the FD process parameters and the trend. The estimation of the FD process parameters is based on an approximate maximum likelihood approach that is made possible by the fact that the DWT decorrelates FD process approximately, and can be tailored to the dependence observed in the series by assuming an autoregressive (AR) process for the wavelet coefficients within scales. The authors illustrate that the dependence between wavelet coefficients on different scales decays with increasing wavelet order (see Craigmile (2000) for further details). Once the FD process parameters have been estimated, one can test for the presence of trend by comparing the sum of squares of the observed scaling coefficients and the so-called "boundary" wavelet coefficients to what would be expected under the null hypothesis of no trend with FD noise. Additionally, one can use the observed scaling and boundary wavelet coefficients to estimate the trend. Such an estimate is similar to most authors above in that it captures the long range changes, but because of the form of the filtering operations in the DWT is more variable at the end-points. The authors show how their method can be extended to handle certain non-linear trends (with a sacrifice in bias), and they apply the approach to analysing the induced Seychelles sea surface temperature series of Charles, Hunter, and Fairbanks (1997).

As we noted above, the problem of extracting a trend $T(t)$ buried in noise is quite close to that of estimating a signal $S(t)$ in the presence of noise. Wavelets provide a very flexible tool for estimating complicated signals. Most of these estimators are variations on the idea of 'wavelet shrinkage' introduced by Donoho, Johnstone and co-workers in a series of papers (Donoho and Johnstone 1994; Donoho, Johnstone, Kerkyacharian, and Picard 1995; Donoho and Johnstone 1998). The basic difference between wavelet shrinkage and what we have discussed up to this point is that the former takes into account influential wavelet coefficients at even the very finest scales. If we are willing to expand the notion of trend to allow for abrupt changes (common in certain physical phenomena), we can apply the following wavelet-based procedures for estimating $S(t)$ as a means for estimating trend.

Johnstone and Silverman (1997) and Johnstone (1999) extend the theory of wavelet shrinkage to the correlated noise case. As mentioned above, the wavelet transform of uncorrelated noise will also be uncorrelated. Thus, given a noise process with constant variance, the resulting wavelet coefficients will also have constant variance – this often leads to a choice of a global threshold for all the wavelet coefficients in wavelet shrinkage. In the dependent case, under certain noise processes, including long memory processes such as fractional Brownian motion (fBm) or FD processes, different wavelet levels have varying statistical properties. This leads us to the concept of thresholding wavelet levels using a different threshold for each level dependent on the sample size and amount of variation present on that level. After thresholding we take the inverse wavelet transform to obtain our estimate of the signal/trend. In the above papers the theory of the uncorrelated case is extended to this situation, and it is shown that we obtain good rates of estimation for a wide class of functions. Examples of the method are given in Johnstone and Silverman (1997), along with a mention of available software.

Wang (1998) considers the application of wavelets to the detection of jumps and sharp cusps. Mathematically we say that the signal $S(t)$ has a sharp α-cusp at t if for $0 \leq \alpha < 1$

$$|S(t+h) - S(t)| \geq K|h|^{\alpha},$$

as $h \to 0$ for some constant $K \geq 0$. We say that it is a jump if $\alpha = 0$. Thus in practice we can consider a cusp to be an abrupt change of the level of the trend over a small time period. Wang considers the model of a unknown signal that might have some cusps plus independent and identically distributed noise. More precisely for the theory to work he assumes a finite number of cusps, and that the unknown signal is differentiable except at these cusp points. We now transform our observed time series using the wavelet transform, and across the finer wavelet scales we then search progressively for the largest wavelet coefficients on that scale (remember that wavelet coefficient are changes of averages, and thus a coefficient of large magnitude implies a large change in the original process). Large wavelet coefficients that are co-located in time across different scales provide estimates of the cusp points. The number of cusps can be estimated by the number of wavelet coefficients that exceed a given threshold (which is equal to the so-called 'universal threshold' of Donoho and Johnstone to a first order of approximation).

In conclusion, wavelets have emerged in the last decade as a useful tool for the analysis of trend. There is still much work to do in the area. Due their time/scale interpretation, wavelets can also analyse nonstationary processes, and investigation of trend in this case looks promising. Work in the area of multiple time series seems useful, as well as an assessment of the irregular sampling case for dependent noise processes. Other wavelet tools such as maximum overlap (or non-decimated) wavelet transforms and wavelet packets may also be of use in this area. We close by noting that these ideas may be extended to the realm of spatial processes. Close examination of the methods and models in this situation is advisable and needed.

References

Antoniadis, A., G. Gregoire, and I. McKeague (1994). Wavelet methods for curve estimation. *JASA 89*, 1340–1353.

Brillinger, D. (1994). Some river wavelets. *Enivronmetrics 5*, 211–220.

Brillinger, D. (1996). Some uses of cumulants in wavelet analysis. *Nonparametric Statistics 6*, 93–114.

Charles, C. D., D. E. Hunter, and R. G. Fairbanks (1997, August). Interaction between the ENSO and the Asian monsoon in a coral record of tropical climate. *Science 277*, 925–928.

Craigmile, P. F. (2000). *Wavelet Based Estimation for Trend Contaminated Long Memory Processes*. Ph. D. thesis, University of Washington, Department of Statistics. To appear.

Craigmile, P. F., D. B. Percival, and P. Guttorp (2000). Assessing non-linear trends using the discrete wavelet transform. Technical report, National Research Center for Statistics and the Environment, University of Washington (Forthcoming).

Donoho, D. L. and I. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*(3), 425–455.

Donoho, D. L. and I. M. Johnstone (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics 26*(3), 879–921.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1995). Wavelet shrinkage: Asymptopia? (with discussion). *JRSSB 57*(2), 301–369.

Gilbert, S. D. (1999). Testing for the onset of trend using wavelets. *Journal of Time Series Analysis 20*(5), 513–526.

Johnstone, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica 9*, 51–83.

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

Kendall, M. (1973). *Time Series*. London: Charles Griffin.

Priestley, M. B. (1981). *Spectral Analysis and Time Series. (Vol. 1): Univariate Series*. London: Academic Press.

Wang, Y. (1998). Change curve estimation via wavelets. *Journal of the American Statistical Association 93*, 163–172.