Journal of Agricultural Engineering

Appendix

Soil electrical conductivity estimated by time domain reflectometry and electromagnetic induction sensors: Accounting for the different sensor observation volumes

Ali Saeed,[1] Alessandro Comegna,[1] Giovanna Dragonetti,[2] Nicola Lamaddalena,[2] Angelo Sommella,[3] Antonio Coppola[1]

Journal of Agricultural Engineering 2017; XLVIII:716_Appendix
doi: 10.4081/jae.2017.716

The FT of discrete stationary series of length M equispaced at intervals Δs $\{x_s, s=0,1,\ldots,M-1\}$ (with *s* being spatial or temporal location on the series) is defined as (Shumway, 1988):

$$X(k) = M^{-1}\sum_{s=0}^{M-1}(x_s - x_M)\exp(-2\pi i \nu_k s) \tag{A1}$$

for $k = 0,1\ldots,M-1$. In Equation A1, $X(k)$ are the Fourier coefficients, $i=\sqrt{-1}$, $\nu_k = k/M$ is the frequency (or wave number) in cycles per unit distance (or time) and x_M is the sample mean. If the series is detrended, x_s in Equation A1 is the detrended series.

The FT in Equation A1 may be written in terms of sine and cosine transform by noting that:

$$\exp(-2\pi i \nu_k s) = \cos(-2\pi \nu_k s) - i\sin(-2\pi \nu_k s) \tag{A2}$$

This way, Equation A1 becomes:

$$X(k) = X_C(k) - iX_S(k) \tag{A3}$$

The Fourier coefficients $X(k)$ are complex numbers. Most software packages (MatLab, SAS, Microsoft Excel, …) have a built-in fast Fourier transform (FFT) algorithm that considerably speeds up the computation of Equation A1, with the sine and cosine transforms available immediately from the real and imaginary parts of the computed X(k).

The real part of the FFT corresponds to the cosine series and the imaginary part corresponds to the sine. The MatLab FFT returns data that can be used to get the following coefficients:

$$\begin{aligned} a_k &= -\frac{2}{M} imag(X(k)), \quad 0 < \mathrm{k} < \frac{\mathrm{M}}{2} \\ b_k &= -\frac{2}{M} real(X(k)), \quad 0 < \mathrm{k} < \frac{\mathrm{M}}{2} \end{aligned} \tag{A4}$$

that can be used for recovering the original data signal by:

$$x(s) = a_0 + \sum_{k=0}^{(M-1)/2} \left(a_k \sin(2\pi \nu_k s) + b_k \cos(2\pi \nu_k s) \right) \tag{A5}$$

As most of the variability of the original EMI and TDR data is contained in the first (low frequency) three to six harmonics, these may be retained while removing the higher frequencies harmonics, to rebuild a smoothed data series through the Equation A5 without the noise. The cut-off frequency for smoothing the original data may be identified by looking at the power spectral density of the data series.

The periodogram, which may be written as the squared modulus of the FT:

$$P_x(\nu_k) = |X(k)|^2 = \left[X_C^2(k) + X_S^2(k) \right] = X(k)\overline{X(k)} \tag{A6}$$

where the overbar denotes complex conjugate, is approximately an unbiased estimator for the spectrum (Shumway, 1988). Each value of $P_x(\nu_k)$ has two degrees of freedom and its interpretation is generally difficult with excessive scatter of neighbouring values and occurrence of unexpected peaks. Also, the variance does not decrease to zero when the sample size tends to infinity. For this reason, it is common practice to average adjacent values of the periodogram to obtain estimates with higher degrees of freedom, and thus create a smoothed power spectrum.

The average spectral estimator in a frequency interval centered on ν_k is defined as:

$$f_x^{P,B}(\nu_k) = L^{-1} \sum_{l=-(L-1)/2}^{(L-1)/2} P\left(\nu_k + \frac{l}{M} \right) = L^{-1} \sum_{l=-(L-1)/2}^{(L-1)/2} |X(k+l)|^2 \tag{A7}$$

where L is some odd integer considerably less than M defining the averaging window. In frequency terms, the averaging window may be expressed as a bandwidth *B=L/M* (cycles per point) centered on ν_k. $f_x^{P,B}(\nu_k)$ is the periodogram-based power spectrum averaged on *B*. It is distributed approximately as a chi-squared χ^2 variable in which the degrees of freedom depend on the window width L used.

The 100(1-α) confidence interval for the smoothed spectrum can be calculated as:

$$\frac{2Lf_x^{P,B}(\nu_k)}{\chi_{2L}^2(a/2)} \leq f_x^n(\nu_k) \leq \frac{2Lf_x^{P,B}(\nu_k)}{\chi_{2L}^2(1-a/2)} \tag{A8}$$

where α is the significance level and $f_x^n(\nu_k)$ is the background noise power spectrum. The null hypothesis is $f_x^{P,B}(\nu_k) = f_x^n(\nu_k)$ *vs* $f_x^{P,B}(\nu_k) \neq f_x^n(\nu_k)$. If $f_x^n(\nu_k)$ falls within the interval in Equation A7, we fail to reject the hypothesis. If not, the estimated power spectrum at a given frequency ν_k has to be considered significantly different from that of the assumed background noise. For the case of a white noise, implying a uniform distribution of the power spectrum across frequencies, $f_x^n(\nu_k)$ can be considered as the mean of all power spectrum estimates.

Reference

Shumway R.H. 1988. Spectral analysis and filtering. In: R. Johnson and D.W. Wichern (Eds.), Applied statistical time series analysis. Prentice Hall, Englewood Cliffs, NJ, USA, pp. 47-116.