P3.11 ESTIMATION OF THE INTEGRAL TIME SCALE WITH TIME SERIES MODELS

Stijn de Waele[a) 1], Arjan van Dijk[b)], Piet Broersen[a)] and Peter Duynkerke[b)]
[a)]Delft University of Technology, Delft, The Netherlands.
[b)]Utrecht University, Utrecht, The Netherlands.

INTRODUCTION

Climate- and regional models solve the dynamics of the atmosphere with a resolution Δ of 10 km or more. They use parameterizations to incorporate the effects of turbulence in the atmospheric boundary layer (ABL). Fluxes, entrainment and dissipation are associated with characteristic time-scales $\mathcal{T}$. They are often constructed from characteristic length scales L for these quantities (temperature, wind velocity, humidity, chemicals) and a characteristic turbulent velocity. The spatial resolution Δ of the regional model determines where the spectra are truncated and which parts of the dynamics should be parameterized. Ratio Δ/L characterizes the importance of these subgrid-parameterizations. With increasing computational resources, the regional models are refined and the spectral truncation induced by the model resolution shifts towards the turbulent regime. Estimation of L from measurements is therefore of increasing importance for understanding the dynamics of the ABL (Nicholls and LeMone 1980). In many cases L equals the height of the boundary layer, but sometimes it does not: in Large Eddy Simulations of a convective boundary layer Jonker et al. (1999) found that, for certain scalar entrainment ratios, the growth of the L for scalars did not follow the growth of L for velocity and temperature.

In this paper we will discuss the estimation of an integral time-scale T from a given signal which via Taylor's hypothesis gives integral length L. The integral time scale is often defined in terms of an autocovariance function that is given by the time-average of shifted product of observations $x_n^* x_{n+m}^*$. We will show that application of this definition leads to results that are difficult to interpret. One problem for this definition is the influence of the subtraction of the estimated mean (Sreenivasan et al. 1978) or a trend correction (Kaimal and Finnigan 1994, p.276). Due to these difficulties, this definition is not used in practice in this pure form. Instead, an integral time scale is obtained using the truncated autocovariance function, or, equivalently, the windowed periodogram or smoothed Fast Fourier Transform (FFT). The smoothing is highly arbitrary and therefore estimation of T via FFT is a subjective business.

To eliminate this arbitrariness, a new estimator is presented. It is based on recent results in the field of time series analysis (Broersen 2002). These results have been incorporated in a Matlab Toolbox ARMASA, that can be downloaded freely (Broersen 2001). This new estimator is compared to classic methods based on shifted-products or FFT in simulations as well as on experimental data.

[1]*corresponding author address:* S. de Waele, Delft University of Technology, Lorentzweg 1, NL - 2628 CJ Delft, The Netherlands; `S.deWaele@tn.tudelft.nl`

1. DEFINITIONS

Our analysis starts at the point where a one-point measurement has been done in the atmosphere and a time-series is available for analysis. We are dealing with sampled signals and therefore all definitions given here are in discrete time. The sampling time is denoted T_s; The number of observations available is N; the total measurement time is $T_m = NT_s$. It is assumed that the signal is a stationary stochastic signal x' (turbulence) plus a deterministic additive or multiplicative trend g (e.g. daily cycle):

$$x_n = x'_n + g_n \qquad \text{or} \qquad x_n = x'_n \cdot g_n. \tag{1}$$

Only trends with a clear physical interpretation are estimated and eliminated. The autocovariance function $R(m)$ of the remaining turbulent time-series is given by:

$$R(m) = E\{x'_n x'_{n+m}\}. \tag{2}$$

The autocorrelation function $\rho(r)$ is the autocovariance function normalized by the signal variance σ^2. In the frequency domain, a stochastic signal can be described by the power spectrum $h(f)$. The definition of the power spectrum can be found in Priestley (1981). An important derived result is the Wiener-Khintchine theorem, which states that the power spectrum $h(f)$ is related to the autocovariance function $R(m)$ by the Fourier Transform.

Several time scales can be defined that characterize the autocorrelation function. A frequently-used characteristic time scale is the sum of the autocorrelations:

$$T^{(1)} = T_s \sum_{r=-\infty}^{+\infty} \rho(r). \tag{3}$$

This characteristic time scale is proportional to the power spectrum at frequency $f = 0$, and to the asymptotic expression for the variance of the estimated mean $\hat{\mu}$ (Priestley 1981, p.320):

$$T^{(1)} = \frac{h(0)}{\sigma^2} = \frac{T_m}{\sigma^2} var(\hat{\mu}). \tag{4}$$

Another integral time scale representing the correlation length is defined as the sum of squares of the autocorrelation function:

$$T^{(2)} = T_s \sum_{r=-\infty}^{+\infty} \rho(r)^2 \tag{5}$$

All non-zero autocorrelations give a contribution to this time scale. This integral time scale is proportional to the asymptotic expression for the variance of the estimated variance $\hat{\sigma}^2$ for normally distributed signals (Priestley 1981, p.326). One must keep in mind that T is *not* the integral timescale in the atmosphere, but the characteristic time scale in the measured time-series. With use of the mean velocity this time T will give the integral length scale L in the atmosphere.

2. ESTIMATORS

For the current analysis, it is assumed that the deterministic trend g in x consists only of a constant mean value μ. The standard estimate of the mean is denoted $\hat{\mu}$. The signal remaining after subtraction of the estimated mean is denoted $x^* = x_n - \hat{\mu}$. We will now discuss two types of estimators: The shifted-products or FFT-based estimator and the estimator based on time series models. The integral time scale $T^{(1)}$ will be used as an example in the definitions. Estimators for the integral time scale $T^{(2)}$ can be formulated in a similar fashion.

SHIFTED PRODUCTS ESTIMATORS

This class of estimators is based on the Blackman-Tuckey (BT) method for spectral estimation. It uses the frequency domain expression for the integral time scale (4). The BT estimate for the spectrum is based on the periodogram or raw FFT, given by the absolute square of the Fourier Transform of the signal x^*:

$$I(f) = T_s \left| \sum_{n=1}^{N} x^* e^{-i2\pi f n} \right|^2. \quad (6)$$

The straightforward estimate for the integral time scale is to take the periodogram at $f = 0$. However, the periodogram at $f = 0$ is exactly equal to zero, independent of the characteristics of the process at hand: $I(0) = 0$. To avoid this problem, we have to use the smoothed periodogram. This means that a weighted average is taken of neighboring periodogram estimates:

$$\hat{h}(f) = \int_{F=-1/2T_s}^{+1/2T_s} I(F) W(f-F) dF. \quad (7)$$

Another argument for smoothing is that it reduces the erratic behavior of the raw periodogram. A typical example of a weighting function is the Daniel window, where the average of a number of neighboring periodogram points is taken. The resulting Blackman-Tuckey estimate $\hat{T}_{BT}^{(1)}$ is given by the smoothed periodogram at $f = 0$:

$$\hat{T}_{BT}^{(1)} = \hat{h}(0) = \int_{F=-1/2T_s}^{+1/2T_s} I(F) W(F) dF. \quad (8)$$

The Blackman-Tuckey estimate is representative for a larger category of estimators. The BT estimate is equal to a weighted sum of estimated covariances:

$$\hat{T}_{BT}^{(1)} = \sum_{r=-r_t}^{r_t} w(r) \hat{R}_{SP}(r). \quad (9)$$

where $\hat{R}_{SP}$ is the shifted-products estimator:

$$\hat{R}_{SP}(r) = \frac{1}{N} \sum_{n=1}^{N-|r|} x_n^* x_{n+|r|}^*. \quad (10)$$

Similar to the frequency domain approach, it would seem natural to take the sum over all available covariance estimates. However, as in the frequency domain approach, this yields a value of exactly zero, independent of the characteristics of the process at hand:

$$\sum_{n=-N}^{N} \hat{R}_{SP}(r) = 0. \quad (11)$$

Therefore, this is not a useful estimate for the integral time scale. This result shows that it is important to make a clear distinction between the definition and an estimate of the integral time scale. The integral time scale is defined in terms of the autocovariance function, given by the *expectation* of a shifted product $x'_n x'_{n-m}$ (equation 2). The estimate (11) is based on an estimate of the autocovariance, given by the *time average* of $x'_n x'_{n-m}$ over the given set of observations.

The BT estimate is closely related to the Bartlett method. In this approach, a less erratic spectral estimate is obtained by dividing the data into M blocks of equal size $N_B = N/M$. From each block m, the periodogram $I_{(m)}$ is obtained. Similar to the smoothed periodogram, this results in a smoother spectral estimate. The resulting integral time scale $\hat{T}_B^{(1)}$ is the sum of the averaged periodogram estimates:

$$\hat{T}_B^{(1)} = \frac{T_s}{\hat{\sigma}^2} \frac{1}{N_B} \sum_{m=1}^{N_B} I_{(m)}(0). \quad (12)$$

The Bartlett estimate for $T^{(1)}$ is equal to the blocking estimate. The blocking estimate is based on the third interpretation of the integral time scale as the variance of the estimated mean (equation 4) (Broersen 1998). An estimate of the mean $\hat{\mu}_m$ is obtained from each block. The sample variance of these m estimates provides an estimate for the variance of the estimated mean:

$$\hat{T}_B^{(1)} = \frac{T_m}{\hat{\sigma}^2} \widehat{var}(\hat{\mu}) = \frac{T_m}{\hat{\sigma}^2} \frac{1}{M} \sum_{m=1}^{M} (\hat{\mu}_m - \hat{\mu})^2. \quad (13)$$

We will now compare the BT and the Bartlett estimators. A first remark is that they are closely related. Although not exactly equal, the BT estimate with lag window size r_t is practically equal to the Bartlett estimate for block size $N_B = r_t$:

$$\hat{T}_B^{(1)} \approx \hat{T}_{BT}^{(1)} \quad (14)$$

A preference can be expressed for the Blackman-Tuckey estimate, because the BT estimator uses all products $x_n x_{n+r}$ present in the signal to get an estimate of $R(r)$. With the Bartlett estimate, some contributions x_n and x_{n+r} are not in the same block and therefore they are not used. With the Welch method of spectral analysis, overlapping blocks are used (Stoica and Moses 1997). With this method, too, there are discarded products $x_n x_{n+r}$.

The same conclusions can be drawn for the other integral time scale, $T^{(2)}$. A minor difference is that no exact equivalence exists between the Bartlett estimate and the blocking estimate for $T^{(2)}$.

Some disadvantages are associated with the shifted-product or FFT-based estimators. Note that we can formulate these problems either in the time domain or in the frequency domain, since the BT estimate has both a time domain and a frequency domain interpretation. First, the window size or 'amount of smoothing' of the periodogram is chosen by the experimenter. This problem is found in the BT approach, but also in the Bartlett approach. Here, the amount of smoothing is given by the number of blocks. Second, the covariance estimate (10) contains a triangular bias. Finally, not all power spectra can be modelled accurately. For instance, power spectra containing both sharp peaks and flat regions are not accurately modelled. If no or little smoothing is applied, the sharp peak is modelled accurately, but the spectrum in the flat region remains erratic; if more smoothing is applied, the flat region is more accurate but the sharp peak is smoothed out.

TIME SERIES MODELS

We just showed that using all shifted-products estimates for the autocovariances to obtain an estimate the integral time scale leads to a meaningless result (equation 11). This is a consequence of the fact that the number of estimated parameters is equal to the number of observations N: $N-1$ correlations $\hat{\rho}(1),...,\hat{\rho}(N-1)$ and the mean $\hat{\mu}$. To get a practical answer, the number of parameters used to describe the data has to be reduced.

A promising approach is to use statistical model inference. With statistical model inference a parametric time series model is determined automatically by the data. Recently, a reliable time series algorithm has been developed based on the Autoregressive (AR), Moving Average (MA) and the combined ARMA models (Broersen 2001). This ARMAsel algorithm is based on statistical order selection, type selection and parameter estimation. This alternative for the FFT-based estimators does not have the disadvantages mentioned in the former section (Broersen 2002). The method to estimate model type, order and parameters (ARMAsel) stops admitting more parameters when inclusion of a new parameter would not lead to *significantly* more knowledge about the signal.

With an ARMA(p,q) model, a signal x is modelled as a white noise signal ε filtered by a rational filter:

$$x_n + a_1 x_{n-1} .. + a_p x_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} .. + b_q \varepsilon_{n-q}. \quad (15)$$

Parameters $a_1,...,a_p$ are the autoregressive (AR) parameters; the number of AR parameters is AR order p. Parameters $b_1,..,b_q$ are the Moving Average (MA) parameters; q is the MA order. An AR(p) model is a model with q=0; a MA model is a model with p=0. The ARMAsel routines, with which we estimated parameters a_i and b_i, can be downloaded from `http://www.tn.tudelft.nl/mmr`.

All process characteristics can be derived from the ARMA-parameters. The power spectrum h is given by

$$h(f) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|1+\sum_{k=1}^{q} b_i e^{-j\omega k}|^2}{|1+\sum_{k=1}^{p} a_i e^{-j\omega k}|^2}. \quad (16)$$

Similarly, the autocorrelation function and the integral time scales $T^{(1)}$ and $T^{(2)}$ can be calculated directly from the parameters (Broersen 1998).

3. COMPARISON

The theoretical analysis of estimators is based on the comparison with the Cramèr-Rao lower bound for the parameter accuracy. If the true process is an ARMA-process, the lower bound is closely approximated with time series modelling (Broersen 1998). No such optimality result can be derived for the shifted-product- or FFT-based estimators.

To compare time series modelling and the shifted-product estimator, we simulated N observations of a turbulent process with a -5/3 slope in the power spectrum to mimic an inertial subrange:

$$h(f) = \gamma \frac{1}{1+\left(\frac{f}{0.01}\right)^{5/3}}, \quad (17)$$

where γ is a normalizing factor. The signal is sampled with sampling frequency $f_s = 1$. The true integral time scales are $T^{(1)} = 26$ and $T^{(2)} = 11$.

For the Blackman-Tuckey approach, a window type and window size has to be chosen. Here, we have used the Parzen window (Priestley 1981) with window size $N/10$. This corresponds to smoothing the periodogram by averaging 10 neighboring points. In terms of the Bartlett estimate, this corresponds to dividing the data into $M = 10$ blocks.

In the table the root mean square error is given for the two estimators. The integral time scale has been estimated for a varying number of observations N. The automatically selected ARMAsel model provides the most accurate result for both integral time scales. It could be argued that a better result could have been obtained with a different choice for the window size. However, a more accurate window size can only be chosen if additional information is available about the process at hand. Conversely, the result obtained with the ARMAsel algorithm has been obtained using the data only.

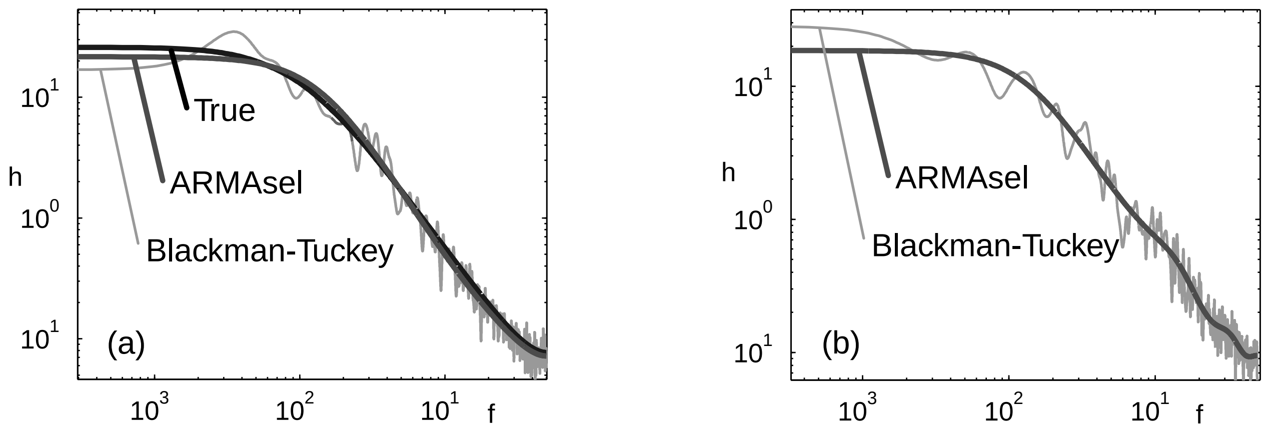

Figure 1: A comparison of the Blackman-Tuckey estimate and the ARMAsel estimate of the power spectrum from N = 5000 observations of simulated data (a) and N = 5000 observations of experimental data (b).

N	500	1000	5000	10000
BT	12.0 / 6.6	7.7 / 5.5	9.6 / 4.9	7.8 / 4.6
ARMAsel	10.7 / 3.5	7.9 / 2.4	5.1 / 1.3	4.6 / 0.9

Table 1: The root mean square error of the estimated integral time scales $T^{(1)}$ and $T^{(2)}$ for a varying number of observations N. The results are given for the Blackman-Tuckey estimate (BT) and ARMAsel time series analysis (average of 100 simulation runs)

An example for the power spectra estimated from a single set data is given in figure 1. The simulation result (figure 1a) shows that if a -5/3 slope that is present in the true spectrum it is modelled very accurately with time series modelling. The two estimators have also been applied to experimental data (figure 1b). These data are vertical velocity measurements of a stable boundary layer, obtained at the Cabauw site, The Netherlands. The behavior of the estimated power spectra is similar to that found in the simulations.

4. CONCLUSION

With ARMAsel we are capable of estimating the integral time-scale T in a series of one-point measurements without the draw-backs of classic methods. There is no subjective influence of choosing a degree of smoothing, nor does the method suffer from the bias that is known to affect FFT-based estimates for T. The integral time-scale can be used via Taylor's hypothesis to estimate the integral length scale L of turbulence in the ABL. The ARMAsel routines can be downloaded from `http://www.tn.tudelft.nl/mmr`.

Acknowledgment
The research of De Waele was supported by the Technology Foundation STW, the applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

The research of Van Dijk was supported by the EU, contract EVK2-CT-1999-00051 EUROCS (European Project on Cloud Systems in Climate Models).

5. REFERENCES

Broersen, P. M. T., 1998: Estimation of the accuracy of mean and variance of correlated data. *IEEE Transactions on Instrumentation and Measurement*, **47**, 1085–1091.

— 2001: Matlab toolbox ARMASA. Available as Download from http://www.tn.tudelft.nl/mmr.

— 2002: Automatic spectral analysis with time series models. *IEEE Transactions on Instrumentation and Measurement*, **51**, in press.

Jonker, H., P. Duynkerke, and J. Cuijpers, 1999: Mesoscale fluctuations in scalars generated by boundary layer convection. *J.Atm.Sci.*, **56**, 801–808.

Kaimal, J. and J. Finnigan, 1994: *Atmospheric Boundary Layer Flows; Their Structure and Measurements*. Oxford University Press, New York, USA.

Nicholls, S. and M. LeMone, 1980: The fair weather boundary layer in gate: The relationship of subcloud fluxes and structure to the distribution and enhancement of cumulus clouds. *J.Atm.Sci.*, **37**, 2051–2067.

Priestley, M. B., 1981: *Spectral Analysis and Time Series*. Ac. Press, London.

Sreenivasan, K. R., A. J. Chambers, and R. A. Antonia, 1978: Accuracy of moments of velocity and scalar fluctuations in the atmosperic surface layer. *Boundary-Layer Meteorology*, **14**, 341–359.

Stoica, P. and R. L. Moses, 1997: *Introduction to Spectral Analysis*. Prentice Hall, Upper Saddle River.