

$\mathcal{O}(N \log N)$ bias-corrected maximum likelihood estimation of stochastic processes with applications to a global oceanographic dataset

A.M. Sykulski*, J.M. Lilly, S.C. Olhede, J.J. Early

* *Department of Statistical Science, University College London, UK
NorthWest Research Associates, Seattle, USA
e-mail: a.sykulski@ucl.ac.uk*

Abstract: Modern Big Data methods are important for the analysis of time series. First of all, this is because of the almost ubiquitous increase in the length, frequency and number of time series samples obtained across many applications. Second, many time series exhibit rich dependency structure, such that the analysis of Big Data using many standard time series techniques (that address these complex dependencies) is unrealistic, as these methods scale poorly with increased volumes of data. For these reasons we seek computationally efficient approximate methods.

In this talk, we present a modification of Whittle Likelihood inference for the parameters of stochastic processes in time. The Whittle Likelihood estimates the parameters of a process by modelling its spectral density, and is commonly used due to its $\mathcal{O}(N \log N)$ computational speed, where N is the length of the data. This is in contrast to time domain methods, such as maximum likelihood, which typically require $\mathcal{O}(N^3)$ computation. Our modification, referred to as the De-Biased Whittle Likelihood, accounts for the estimation bias inherent to standard Whittle Likelihood for moderate sample sizes. This bias is attributed to blurring—the phenomena of leakage and aliasing when estimating the spectrum from a finite duration, discretely sampled realisation of the process. Addressing this bias is particularly important when short windows of data are used, as is desirable when a long time series exhibits nonstationarity. We prove consistency of our estimator, and demonstrate both theoretically and empirically the importance of this bias correction. The De-Biased Whittle Likelihood is also computationally faster than standard Whittle Likelihood, as aliased frequencies are naturally treated correctly.

We then estimate parameters of stochastic processes fitted to time series observed from a major oceanographic dataset produced by the Global Drifter Program. Drifters are freely-drifting and satellite-tracked ocean surface instruments, data from which is used to study ocean circulation. We study the properties of the Lagrangian trajectories mapped out by these drifters. The stochastic process model is physically motivated, and the estimated parameters summarise important features such as the diffusivity and the spectral slope. Diffusivity is the rate at which particles spread out from an initial location due to random motions, and the spectral slope is the rate of decay of the velocity power spectrum. The latter quantifies the smoothness of the trajectory, and is related to the concept of fractal dimension.

The $\mathcal{O}(N \log N)$ efficiency of the De-Biased Whittle Likelihood is essential for this application. There have been over 11,000 drifters in operation, representing over 60 million position reports in total. The modelling and estimation is performed in the frequency domain as, in addition to requiring computational efficiency, this is where the stochastic processes are most easily constructed from physical understanding. Furthermore, our frequency domain likelihood technique can be used to perform inference semi-parametrically across a selected range of frequencies, outside of which either the data is contaminated, or the model is known to be misspecified. The talk will conclude with a discussion of the physical interpretation of the observed spatial patterns in the estimated parameters.