Big Data methodology for large-scale spatiotemporal oceanographic datasets

Adam M. Sykulski[1], Jonathan M. Lilly[2], Sofia C. Olhede[3], Jeffrey J. Early[2]

[1] Data Science Institute and the Department of Mathematics and Statistics, Lancaster University, UK
[2] NorthWest Research Associates, Seattle, USA
[3] Department of Statistical Science, University College London, UK

The volume and variety of data collected from our oceans is rapidly increasing, from instruments such as surface buoys, sub-surface floats, and underwater gliders, and many other high-technology sources such as satellite altimetry and acoustic sensing. The analysis of such data is key to our understanding of climate and the environment, particularly as the ocean is recognised as having a pivotal role in determining decadal-to-century scale climate change.

In this talk we will present some recent developments in using stochastic processes to capture the spatiotemporal dependencies inherent in such data. The aim is to construct stochastic process models that simultaneously provide a good fit to observations, but are also constructed from geophysical fluid flow principles such that estimated parameters of the model are physically informative and in meaningful units.

We will demonstrate the theoretical motivation behind using the Matérn process to model particle velocities in oceanographic fluid flow. The Matérn process is a non-Markovian fractal-type stochastic process that is receiving increasing attention in spatiotemporal statistics because of its flexibility in setting a continuous differentiability parameter, which controls the roughness or smoothness associated with a time series or random field. We will demonstrate how the Matérn process is part of a family of continuous Gaussian processes which encompasses Brownian motion and the Ornstein-Uhlenbeck process. Furthermore, we will demonstrate how the Matérn process is in fact related theoretically to fractional Brownian motion, and can be interpreted as *damped* fraction Brownian motion. When the Matérn is applied to oceanic particle velocities, then the presence of a non-zero damping parameter controls the diffusivity—the rate of particle dispersion in fluid flow—which would otherwise converge to infinity with regular (un-damped) fractional Brownian motion. This makes the Matérn a more appropriate model for analysing turbulence in fluid flow as we shall demonstrate.

To implement the Matérn process on large-scale oceanographic datasets there are a number of additional challenges. First, the physical processes of interest have spatial dependencies are significantly heterogeneous and temporal dependencies that are nonstationary. Furthermore, the acquired data sets are usually irregularly sampled in both time and space with most instruments. We shall demonstrate how locally stationary space-time modelling using the Matérn process can resolve spatial-temporal dependencies in 100 million data points acquired from the Global Drifter Program—a large collection of particle velocities collected from over 10,000 freely-drifting surface buoys over the past 38 years.