Adam M. Sykulski
Host Organisation: University College London, London, UK
Overseas Host: NorthWest Research Associates, Seattle, USA
19 June 2017

IOF Final Report

Statistical modelling and estimation for spatiotemporal data with oceanographic applications

The volume and variety of data collected from our oceans is rapidly increasing, from instruments such as surface drifters, sub-surface floats, underwater gliders, satellite altimeters, and acoustic sensors. There are now billions of observations, and the analysis of this data is key to our understanding of climate change and the environment at large.

This project was concerned with developing new statistical methods for the analysis of such datasets. We sought to build models that could exploit the volume of data to estimate new features of the ocean, at previously unresolved resolutions.

A particular focus was on data from the Global Drifter Program, a vast and unique collections of observations encompassing most of the globe's ocean surface. In total over 20,000 drifters have been deployed since 1979, resulting in over 100 million observations of measurements including surface velocity, sea surface temperature, and salinity. The challenge with such datasets are that the drifters are moving in time and space, so rather than observing fixed-point measurements, we instead observe "trajectories" where the observations from each instrument are moving across and between the oceans. This challenge required new models and new statistical methodology for their implementation.

The main contribution of this project was the delivery of a fully-integrated statistical model (from design to implementation stage) that can be used to better understand the trajectories from drifters and related instruments. Our model uses stochastic processes to describe the variability and uncertainty that is inherent to drifter trajectories, in addition to capturing swirling rotational oscillations caused by the Coriolis effect - the effect due to the Earth's rotation on its own axis - as well as from tides and local features known as eddies. Details of our model have been published to both the statistics [1] and physical oceanography [2] communities. Furthermore, the results from our global analysis have been disseminated at leading international conferences such as Ocean Sciences, the American Geophysical Union Fall meeting, the European Geophysical Union General Assembly, the Joint Statistical Meetings, and the Royal Statistical Society Annual Conference.

The implementation of our model to large-scale data required several new methodological developments in statistics and spatiotemporal data analysis. These include a computationally-efficient procedure for estimating parameters from the model [3], a method for simulating synthetic data from the model [4], and methods to check whether the assumptions of the model hold for a particular dataset, in this case local stationarity [5]. We also contributed to a new interpolated product from the Global Drifter Program dataset [6], which was then used in subsequent analyses to maximise the resolution of our output.

In addition, when certain assumptions of the model are not met, we have extended our model to data that is anisotropic (i.e. rotationally-invariant in space) [7] or is rapidly-varying and not locally stationary [8] - the latter of which being key to analysing drifter trajectories at or near the Equator, a previously poorly understood region in this context. Finally, from this research we developed an analogous modelling approach which was transferrable to other physical applications [9], namely the analysis of seismic signals. We anticipate further such synergies in the analysis of spatiotemporal data in numerous other Big Data applications.

The impact of this research is significant in several directions, and to numerous parties. First of all, better modelling of the ocean feeds directly into climate and environmental models, which is important to numerous academic and research institutions globally. In addition, our model exploits Big Data to yield output that is high-resolution in space and time. This can be used to better understand local features of our oceans, and hence our model can be used to improve modelling of oil spills or biological transport for example, which is of interest to industrial as well as governmental bodies.

The lead researcher, Dr Adam M. Sykulski, has moved to a new position at Lancaster University, and welcomes contact with regards to this project at a.sykulski@lancaster.ac.uk

References

[1] A.M. Sykulski, S.C. Olhede, J.M. Lilly and E. Danioux (2016). Lagrangian time series models for ocean surface drifter trajectories. Journal of the Royal Statistical Society (Series C): Applied Statistics, 65(1), 29–50.

[2] J.M. Lilly, A.M. Sykulski, J.J. Early and S.C. Olhede (2017). Fractional Brownian motion, the Matern process, and stochastic modeling of turbulent dispersion. Nonlinear Processes in Geophysics. Discussion paper, https://doi.org/10.5194/npg-2017-15, in revision.

[3] A.M. Sykulski, S.C. Olhede and J.M. Lilly (2016). The de-biased Whittle likelihood for second-order stationary stochastic processes. arXiv preprint: 1605.06718

[4] A.M. Sykulski and D.B. Percival (2016) Exact simulation of noncircular or improper complex-valued stationary Gaussian processes using circulant embedding. Proceedings of the 26th IEEE Workshop on Machine Learning for Signal Processing.

[5] T.E. Bartlett, A.M. Sykulski, S.C. Olhede, J.M. Lilly and J.J. Early (2015). A power variance test for nonstationarity in complex-valued signals. Proceedings of the 14th IEEE International Conference on Machine Learning and Applications, 911–916.

[6] S. Elipot, R. Lumpkin, R.C. Perez, J.M. Lilly, J.J. Early and A.M. Sykulski (2016). A global surface drifter dataset at hourly resolution. Journal of Geophysical Research - Oceans, 121(5), 2937–2966.

[7] A.M. Sykulski, S.C. Olhede, J.M. Lilly and J.J. Early (2017). Frequency-domain stochastic modeling of stationary bivariate or complex-valued signals. IEEE Transactions on Signal Processing, 65(12), 3136–3151.

[8] A.P. Guillaumin, A.M. Sykulski, S.C. Olhede, J.J. Early and J.M. Lilly (2017) Analysis of nonstationary modulated time series with applications to oceanographic flow measurements. Journal of Time Series Analysis (accepted, in press).

[9] A.M. Sykulski, S.C. Olhede and J.M. Lilly (2016). A widely linear complex autoregressive process of order one. IEEE Transactions on Signal Processing, 64(23), 6200–6210.