Choosing a Dataset

In this course we will work in detail on time series methods, particularly Fourier and wavelet-based methods. These are extremely powerful and sophisticated methods which require study to fully appreciate and confidently apply.

An important part of this course is applying the methods to a dataset of your choice. This is because, when we apply a method to a problem we care about, we are more likely to understand it better.

However, not all datasets are a good match to the methods we will be using. Therefore some care must be taken in choosing a dataset. For this reason, I will be communicating with participants ahead of the course time to make sure that the chosen datasets will be suitable. Please understand a dataset may not be a good match for this particular course and yet nevertheless be of great scientific value.

Here is a little information about what constitutes an appropriate dataset. A time series is a one-dimensional dataset. Some examples are current meter velocity measurements, wind speed and direction measurements, temperature or pressure records, Lagrangian drifter trajectories, seismic records, and so forth. Datasets consisting of multiple or many time series (e.g. arrays measuring the same quantity at different locations) are fine and are actually encouraged. Time series output from numerical models are fine too. Sometimes, the measurement dimension is not really time—for example, alongtrack altimeter data, or 1D slices through a model; this is generally also fine. Model data in particular may be sometimes analyzed fruitfully using time series methods in multiple directions, e.g. time, x, y, or z.

A few tips for things that may not be suitable. The methods we're working with here are most useful when (i) there are a sufficient number of data points in the time series, say, on the order of 1000 or greater, (ii) the data is sampled at more or less uniform time intervals, without major gaps, (iii) the original measurements are of sufficiently high precision, and (iv) we are dealing with original, quality controlled measured data, not filtered, smoothed, OI-interpolated, or aggregated datasets; in other words, the output of a single instrument or set of instruments.

Examples of data that can be problematic are sediment core data or geochemical measurements, which can often contain a relatively small number of points with a lot of scatter and with non-uniform sampling. These types of data are best treated with different methods than those we're working with, so if you have data of this type we'll need to look at it closely prior to the course to make sure it's suitable. The same goes for datasets in which the "time" dimension is highly asymmetric, such as oceanographic CTD profiles. Finally, time series of quantities that are intrinsically one-signed—such as precipitation rate, kinetic energy, or incoming solar radiation—are also not well matched to the course methods for reasons that we will discuss.

If the time series you are most interested in at the moment is one that is not a good match for this course, I recommend finding an alternate time series that is relevant to your research and that you can work with for the purposes of understanding these methods. Sooner or later, they will prove to be of great value in your career.