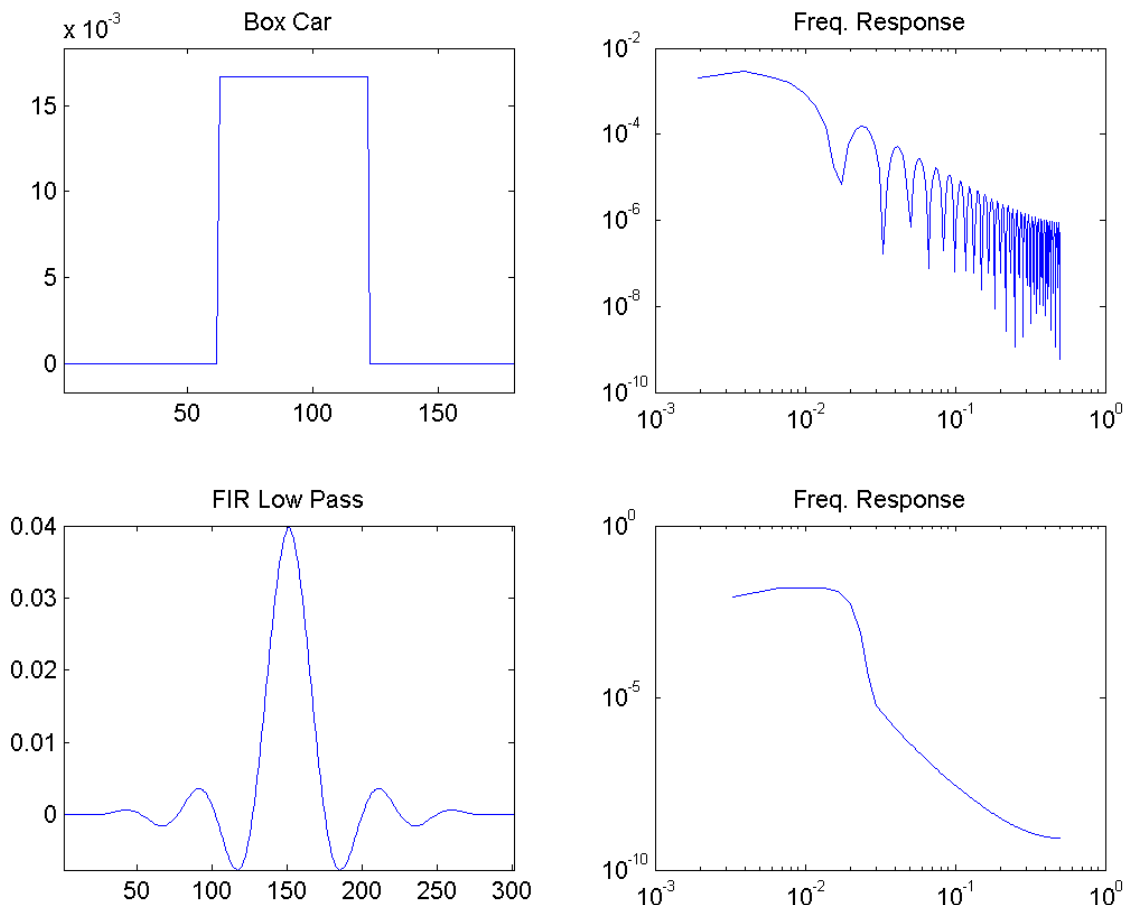


Time Averaging and Resampling

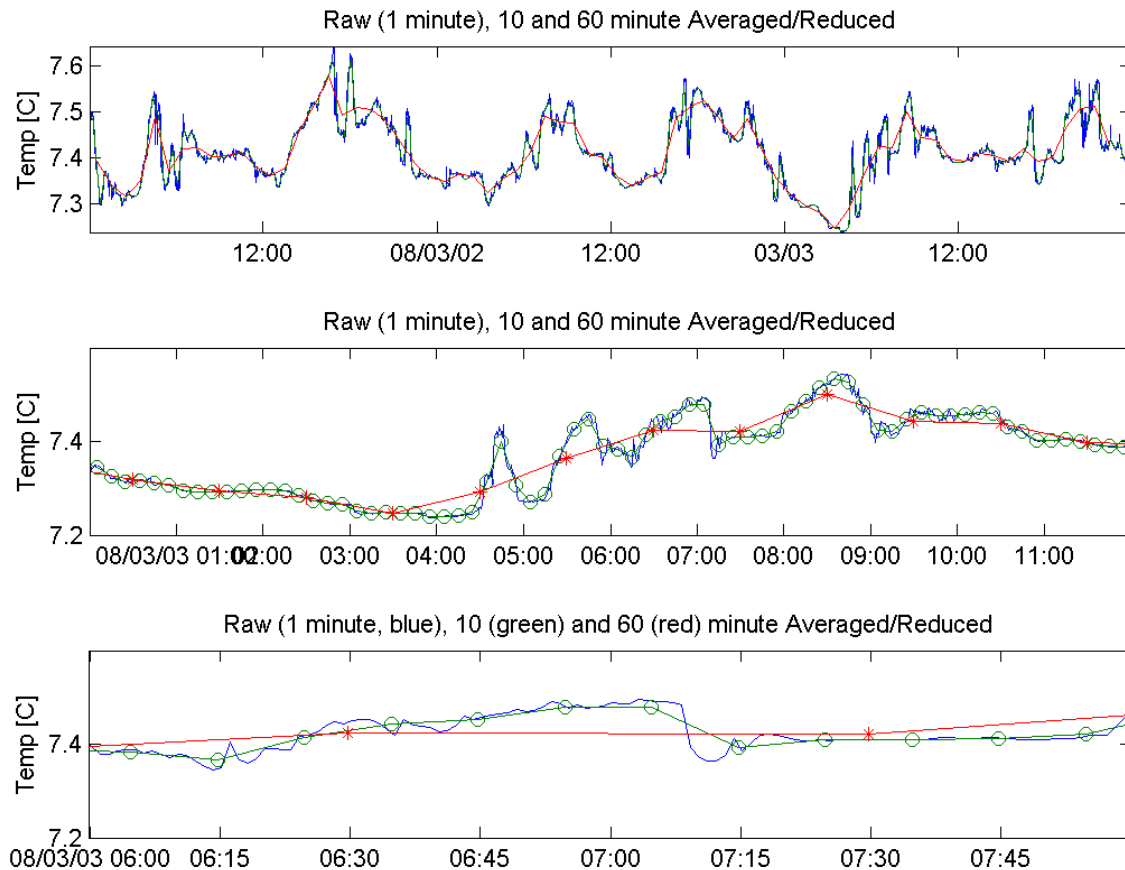
Ocean Networks Canada offers a number of data products with a reduced data density obtained by time averaging. Our sensors typically fall into one of two sampling categories, either fast at/or near 1Hz sampling (e.g. ADCP, ZAP, Nortek Vector) and slower sampling at one minute intervals (e.g. CTD, Oxygen, Pressure). Over long periods, both of these data types produce high volumes of raw data. As the length (duration) of the time series grow to many years, and become of more interest for assessments of inter-annual variability, for example, the volume of data can become unmanageable. To assist and reduce the volume of data for long-duration data requests, we offer time averaged data as a option for all of our tools: data products, services/API and viewing apps such as Plotting Utility. The reduction in data volume can be many orders of magnitude, which can greatly reduce the size of the request and the speed and ease with which one can analyze the information.

Although there are a number of techniques for computing the time average, we have adopted a simple box-car approach with a centered time stamp. In other words, for the 10 minute averaged scalar data, the reduced time series has values at 00:05, 00:15, 00:25, etc., while the time window is defined as 00:00 (inclusive) to 00:10 (exclusive), in that case. The time stamps and windows are taken as one would expect for other intervals, for example, daily is taken from midnight to midnight, weeks start Sunday at midnight, with the start of the box-car being inclusive, the end point is exclusive, so that the end point of one box-car is the start of the next (zero overlap or gap). Only "good" or "clean" data values that pass our data QA/QC procedures contribute to the time average. In general, it is suggested that if a user is interested in plotting or analyzing only a few days of data, then they should request the full un-reduced (un-averaged) data. If one is interested in processes of a week to a few months, then the 15 minute averaged data is recommended, and if one is interested in signals with long time scales, from months to multiple years, then the 60 minute averaged data is recommended. ONC plotting applications will automatically choose the averaging interval, depending on the size of screen and duration of the request, an automatic option is also the default for plots generated from Data Search. Minimum and maximum value resampling is also offered, denoted commonly as min/max and min/max+avg, the latter being very useful for plotting.

The box-car windowing and averaging technique is perhaps the simplest averaging method and has the advantages of being easily understood and is generally applicable. However, there are consequences users should be aware of. Shown immediately below are four plots showing the filter shape and frequency response for both a box-car (upper two panels), and a more sophisticated low-pass finite impulse response (FIR) filter averaging (lower panels). Both "filters" have an equivalent reduction to approximately 60 minutes, assuming an original sample rate of 1 minute. The frequency response of the box-car averaging is seen to have a more gentle roll-off of the higher frequencies and substantial "ringing". The FIR filter has a sharper cut-off and virtually no "ringing" (smooth reduction in the frequencies removed). Although FIR and related filter and sub-sample techniques may reduce some undesirable artifacts, they are also more complicated to both interpret and manage.



In the plots below, we show the effects of the box-car averaging on a simple temperature time series recorded at our DDL site (Delta Dynamics, Strait of Georgia, VENUS array) in March 2008. The top panel shows three days of data: the raw 1 minute samples (blue), the 10 minute averaged data (green) and the 60 minute averaged data (red). Clearly some of the larger and high-frequency oscillations in the data can already be seen to be missing in the 60 minute averages (red line). The middle panel shows a 12 hour segment, while the lower panel shows a 2 hour segment. For these shorter duration plots, the averaged data is clearly not representing many of the key variations in the signal. The 10 minute averaged data is centered on the 00:05 minute marks, while the 60 minute averaged data is centered on the 00:30 intervals. If short-term variability or signal statistics are important, we do not recommend analysis of the reduced averaged data. If one is interested in general trends, or long-term variability, the lower volume averaged data may provide a fast and informative alternative.



Tidal Aliasing

Users should also be aware of the strength and prevalence of the tidal signal in all Ocean-sourced data. As seen above, the box-car method is an acceptable anti-alias low pass filter, but it is not perfect as the high frequency roll off is not as steep or deep as a FIR filter. As such, some aliasing of signals at frequencies near the resampling frequency can occur. Further, when making plots of tide-influenced data, one should either clearly represent the tidal signal with averaging intervals below 6 hours or "average out" the tidal signal with averaging intervals of greater than 24 hours. In general, it is recommended that users avoid resampling/averaging intervals between 6 and 12 hours, inclusive. Cautious users may extend that range to include 24 hour / daily averages, although daily averages are also intuitive and useful. Plotting services and applications, such as Plotting Utility, will not offer or use these resampling/averaging intervals. Other services and products will provide users with warnings when these intervals are selected.

Mobile / Spatial Data

Data acquired from a moving platform has additional considerations compared to a stationary platform. Standard time series analysis techniques may not be appropriate. Spatial averages, binning, maps, etc, maybe better approaches than temporal resampling and standard measurement versus time plots. It also depends on the instrument, sensor and deployment type. For instance, devices may or may not be orientation (heading, pitch, roll) sensitive, such as a temperature sensor compared to a directional sonar. (Footnote: most, if not all our directional devices offer processed data that are corrected for orientation dynamically, but raw data is not corrected.). All measurements are sensitive to position (latitude, longitude, depth). In order to make our users aware of these considerations, all mobile data that is resampled and/or plotted versus time will include a warning. In some cases, resampling or plotting by time is appropriate. Our mobile data products also include any available positioning and attitude data as detailed here: [Positioning and Attitude for Mobile Devices](#), so that users may make use of the spatial context in their analysis.

Upsampling

Upsampling can occur when the averaging interval is less than the sample period of the data. In this case, the time series will become regularly spaced and filled with NaN values for intervals where there is no data. Upsampling can be useful for creating an aligned time series, which is a data set comprising of multiple sensors that all have the same time stamps, when those sensors have different sample rates. Upsampling is not restricted, but it is good practice to avoid upsampling. The [70% rule](#) applies to all resampling, meaning that if one upsamples by a significant fraction (more than about 50%) almost all the upsampled data will be NaN.

Related articles

- [Time Averaging and Resampling](#)
- [Download data using API](#)
- [Find and Download an Audio Highlight](#)
- [Use Python to download data files](#)
- [Create Hydrophone Long Term Spectral Averages](#)