Extract from IOC Training Manual, Vol. III, 2000.

5. Data Documentation and Processing

5.1 Documentation and Archiving

A sea level agency should aim to not only operate gauges to its best ability, but also to provide proper documentation, data processing and archiving functions. Documentation has already been alluded to in previous sections. All tide gauge operations (equipment change notes, calibration records, maps, photographs etc.) must be documented within an overall, preferably computerised system so that the information is not lost to future analysts. The tide gauge data themselves must be checked (and if necessary corrected) for their quality and properly documented before being passed to scientists in the wider community.

5.2 PC-Based Software

The aim of data processing software should be to ensure the quality of any data which are subsequently archived and used for scientific analysis. Tide gauge data which are nowadays collected by electronic methods tend to have fewer errors than existed in the days of paper charts and inaccurate clocks. However, there is still a need to check for timings, data spikes and gaps etc. by, for example, 'buddy checking' with respect to predicted sea level time series (determined from historical tidal constants) or to data from a nearby gauge.

Many organisations have developed their own processing software to validate incoming data in varied formats and media which are specific to their requirements. However, three organisations have developed PC-based software as a contribution to GLOSS with the aim of enabling participating countries to be able to process and validate their own records. These organisations are:

- The University of Hawaii, which has produced a package led by Patrick Caldwell which is probably the most commonly used package within the GLOSS community for sea level data quality-control purposes.

- The PSMSL/Proudman Oceanographic Laboratory, which has produced a package led by Colin Bell, Ian Vassie and Philip Woodworth called TASK-2000 (Tidal Analysis Software Kit) which is based in the TIRA etc. programmes used at POL for many years.

- The Australian National Tidal Facility, which has own package derived originally from POL software.

For contact addresses and emails, see:

http://www.pol.ac.uk/psmsl/training/training.html or email psmsl@pol.ac.uk.

The manuals accompanying these packages describe the data processing and quality control procedures used.

Each of these packages runs at the present time under DOS only. We understand that each set of authors has plans to develop the packages under later versions of Windows. However, at the time of writing, they are not yet available.

5.2.1 Comments on Tidal Predictions

Each of these packages has been developed primarily for sea level data quality control. However, each also has the capability for determining tidal constants and for calculating tidal predictions. However, a user should be aware that the tidal software is not the same in each package. The Hawaii package contains tidal software constructed by Dr. Mike Foreman from Canada, whereas the TASK and NTF packages contain tidal software originally written at Bidston Observatory. 'Tidal constants' can have slightly different meanings depending on the software and should never be mixed between packages. For example, constants determined by one package should never be used to compute predictions with another one.

So far as we know, this problem is restricted to the annual harmonic (what tidalists call S_a). However, there may be other differences which we are not aware of. (See the manual which accompanies the TASK-2000 package for more information on this topic.)

5.2.2 Tidal Filters for 'Mean Sea Level'

Volume 1 of this Manual was somewhat ambiguous in its recommendations for computations of monthly mean values from hourly data, prior to submission of the monthly and annual means to the PSMSL. Volume 1 demonstrated:

- The summation of a month's high and low waters to make Mean Tide Level (MTL) for the month.
- A method by which a simple filter (the Doodson X0 filter) could be used as a 'tide killing filter' to compute so-called 'daily mean values' which could then be summed over the month to make the monthly Mean Sea Level (MSL).
- A method by which the simple arithmetic average of hourly values for a month can be used to calculate the monthly MSL.

Some comments on these options are appropriate. First, one has to realise that MTL is not the same quantity as MSL because of shallow water tidal effects and, while time series of MTL are adequate for many purposes, they should never be mixed with time series of MSL because of the systematic offsets between them (see Pugh, 1987 for explanation of the shallow water tides which cause the offsets). Of course, MTL was frequently used in former times rather than MSL, when agencies did not have access to adequate computing capacity to handle hourly values. The PSMSL always strives to keep MSL and MTL time series separate in its data bank.

Second, the filter described in Volume 1 for MSL computation (the Doodson X0 filter), which is the filter in the TASK package, is not the same one as used in the Hawaii software. The former is a simple 39-point moving arithmetic filter. The latter performs a pre-whitening tidal reduction for 4 constituents (M2, S2, O1 and K1) for each month of data and then low-pass filters the residuals using a 119-point filter. The resulting 'daily means' are then summed in each case to compute the monthly mean.

One might ask, how similar (or dissimilar) are the daily means produced with the Doodson and Hawaii filters? (Obviously a 'daily mean' computed by simple averaging will be contaminated by the tidal signal). And how similar are the monthly means computed in the three cases?

A test was performed by the PSMSL in January 2000 using over 800 station-years of hourly data from a CDROM of sea level data produced by the University of Hawaii, using two years of data from each station. Each year of data was subjected to the three averaging methods. Results showed that the standard deviation of the difference between daily means produced using the Hawaii and Doodson filters was 4.7 mm. The standard deviation of differences between (monthly means determined by the Hawaii filtering to dailies then summation to monthlies) and (Doodson filtering to dailies then summation to monthlies) was 1.3 mm. Meanwhile, the standard deviation of differences between (monthly means determined by the Hawaii filtering to dailies then summation to monthlies) and (simple monthly arithmetic average) was 1.2 mm. All these tests were performed for complete months of data only as spurious differences can be obtained from filtering data with gaps.

The conclusions to be made from this from the point of view of monthly MSL computation are:

- If one only has high and low water data, then the summation to MTL described in Volume 1 is acceptable as long as the resulting monthly means are clearly documented as MTL and not MSL.

- The choice between (filtering with either filter to dailies and then averaging to monthlies) or (simple arithmetic averaging over a month) is not critical. However, the former is probably a more mathematically rigorous procedure in principle and is the one we recommend in this Manual for future research.

The conclusion to be made from the point of view of daily mean computation follows from the Hawaii filter being wider than the Doodson. Consequently, it should provide a tighter rejection of tidal bands and a better definition of daily means. However, the differences are almost always sub-centimetre and, if significantly different values are obtained with the use of the two filters, it probably means that the period under study contains considerable energy (e.g. due to a large storm surge) which will require further study with hourly values.

5.2.3 Comments on Computations of Extremes

The study of sea level extreme levels is an important subject both within the scientific (e.g. climate change) and engineering communities. For a discussion of the statistical analysis of extremes, see Pugh (1987) or Coles and Tawn (1990). However, it seems that there has never been a formal definition of how to compute the 'turning points' at each high and low tide, from which to find the extremes from a set of data (e.g. the single annual extremes from one year of data, or the 'N-largest' from the year), with which to perform the statistical analyses.

There are two common methods used to find the turning points. The first is to use (perhaps 6) hourly values of sea level spanning high (low) tide, and to interpolate them using a cubic spline in order to find the maximum (minimum) value for each tide. The highest (lowest) turning point for the year is then defined as the annual extreme. A second is to simply find the highest and lowest hourly values as observed in the data set for the year. Clearly, the first method will produce higher (lower) annual extreme maxima (minima).

The situation is more complicated if one samples at intervals of 6 or 15 minutes rather than an hour. Does one define the extremes based on the original sample of 6 minute data or from low-pass filtered hourly values? Common sense suggests that one should use the highest rate of sampling one has available in order to find the real extremes because, even if the real extreme persisted only for instant, it might be enough to result in over-topping and flooding. In addition, it is well-known that in many applications hourly sampling is insufficient to provide a full description of the sea level time evolution (e.g. in storm surge research in NW Europe).

In practice, one suspects that the resulting data set of annual extreme levels, and the results of the statistical analyses based on that data set, will be relatively insensitive to the method used to find the turning points, given the wide range of level within any distribution of annual extremes. However, for the purpose of this Manual we recommend:

- If only hourly values of sea level are available, the turning points should be defined by means of cubic spline interpolation of (perhaps 6) values spanning each high and low tide.
- If higher frequency sampling of sea level is available (e.g. 6 or 15 minutes), spline interpolation could again be used, but one suspects the turning point levels will be little different to simply noting the observed maximum (minimum) values at each tide.