Quality Control Challenges-Utilizing NetCDF files and Check Factor Data

J. Thomas (SIO), C. Olfe (SIO)

Scripps Institution of Oceanography 9500 Gilman Drive #0213 La Jolla, CA 92093-0213

1. History

Several on-going national and international efforts are addressing quality control for real-time data. As communication between our maritime communities expands, quality control standards that can be applied globally and produce reliable, accurate results are becoming critical. One method of developing and improving quality control criteria is to examine long term data sets, analyzing statistical parameters such as maximum thresholds and trends.

The Coastal Data Information Program (CDIP), based at the Scripps Institution of Oceanography, University of California, San Diego, provides an example of a long term data set of wave measurements. Since its inception in 1975, CDIP has collected more than 1 terabyte of wave data and operationally has more than 3.8 billion buoy hours of “good” spectral data collection. CDIP is funded by the California Department of Boating and Waterways (CDBW) and the United States Army Corps of Engineers (USACE). At this time CDIP has over 50 active wave buoys that are deployed throughout the coastal United States, including the South Pacific, the Caribbean and Alaska. These instruments contribute to the state and federal missions of increasing boating safety, navigation and protection of coastal structures.

2. Quality Control

While CDIP was still in its infancy, automated quality control was developed. The quality control parameters were based on very stringent criteria. In 1988, these parameters were formalized with the USACE and the National Data Buoy Center (NDBC), resulting in the publication of the “Wave Data Analysis System”. Further, in 2003, the grass-roots organization comprised of government, commercial and academic entities formed the “Quality Assurance of Real-Time Data” (QARTOD). These organizations agreed upon basic, minimum quality control parameters for waves. These parameters have served as a basis for the NOAA/USACE National Waves Plan and are now in the process of being adopted and published by the Integrated Ocean Observing System (IOOS).

With 36 years of experience, CDIP demonstrates how pre-deployment verification procedures and the automated quality control has worked well in assuring accurate, well-defined wave data. A suite of diagnostic tools assist in monitoring the data. All data have accompanying metadata and data rejections are well documented. Data which pass the quality control test become public and are posted on the CDIP website and distributed every 30 minutes to the NDBC and the National Weather Service. Data that are rejected become non-public and will not

978-0-933957-40-4 ©2013 MTS

be distributed. However, all time-series and spectral data are archived, and both public and non-public data sets are freely available to the public.

3. Expanding Quality Control

Now that CDIP has a vast archive of data, occurrence of instances outside our standard quality control are of interest. This paper portrays certain examples of data that present challenges to the standard quality control parameters. These instances affecting the instrument are caused by such events such as (but not limited to) excessive marine growth or current factors. The response of the buoy is well documented in many of these cases.

Since the early 1990's, most of CDIP's wave observations have been collected with the Datawell directional wave buoys. These buoys return the inverse of the wave ellipticity, or 'check factors' for each frequency band in the spectra they calculate. The ratio of horizontal displacements (x, y) to vertical displacements (z), these check factors are an important tool in performing QC on the buoy data. In general, the mid-frequency and high-frequency check factor values should be very close to one. The low-frequency bands often have very low energy levels, so their check factor values are more variable. When the mid- or high-frequency check factors deviate significantly from one, it indicates that there is likely an issue with the buoy. Figure 1 is an example of a deployment at Monterey Canyon, California. Although anti-fouling coatings are used on the buoys, they accumulate considerable marine growth. During this May 2011, when the growth was particularly abundant, the buoy became covered with barnacles and other organisms. The check factors reflected this condition. The plot below shows the signal before and after the buoy hull and upper mooring was cleaned.

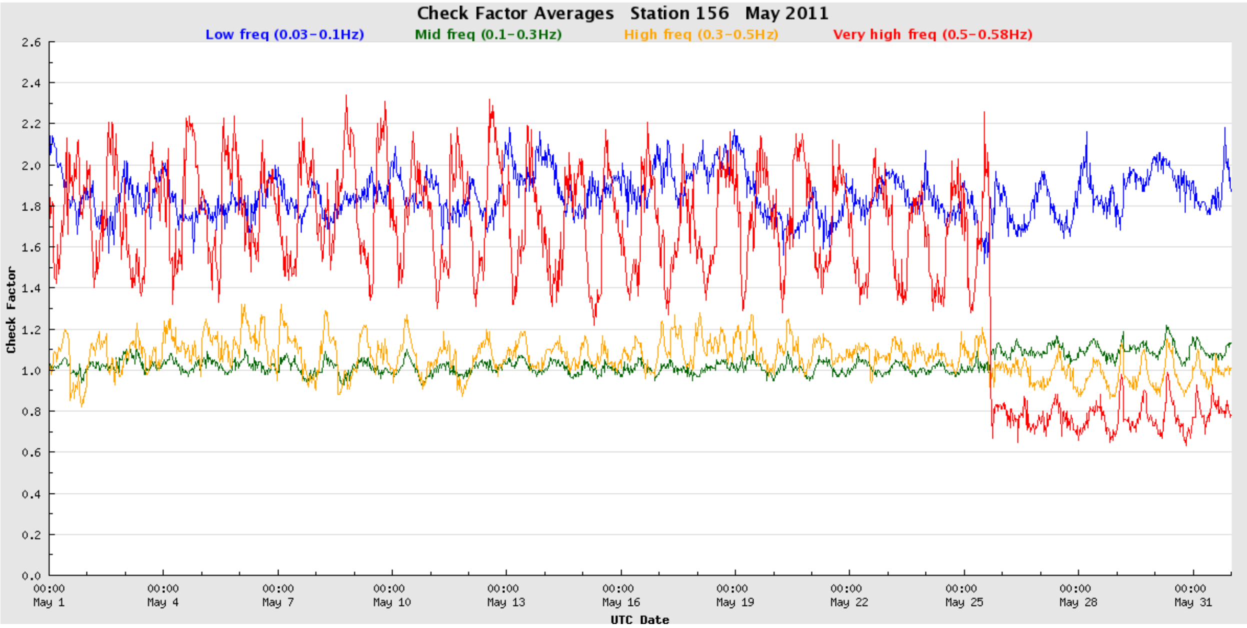

Figure 1. Monterey Canyon California

4. Utilizing Check Factors

In order to look for data anomalies, empirical data has been used to find the range of normal average check factors. See Figure 2.

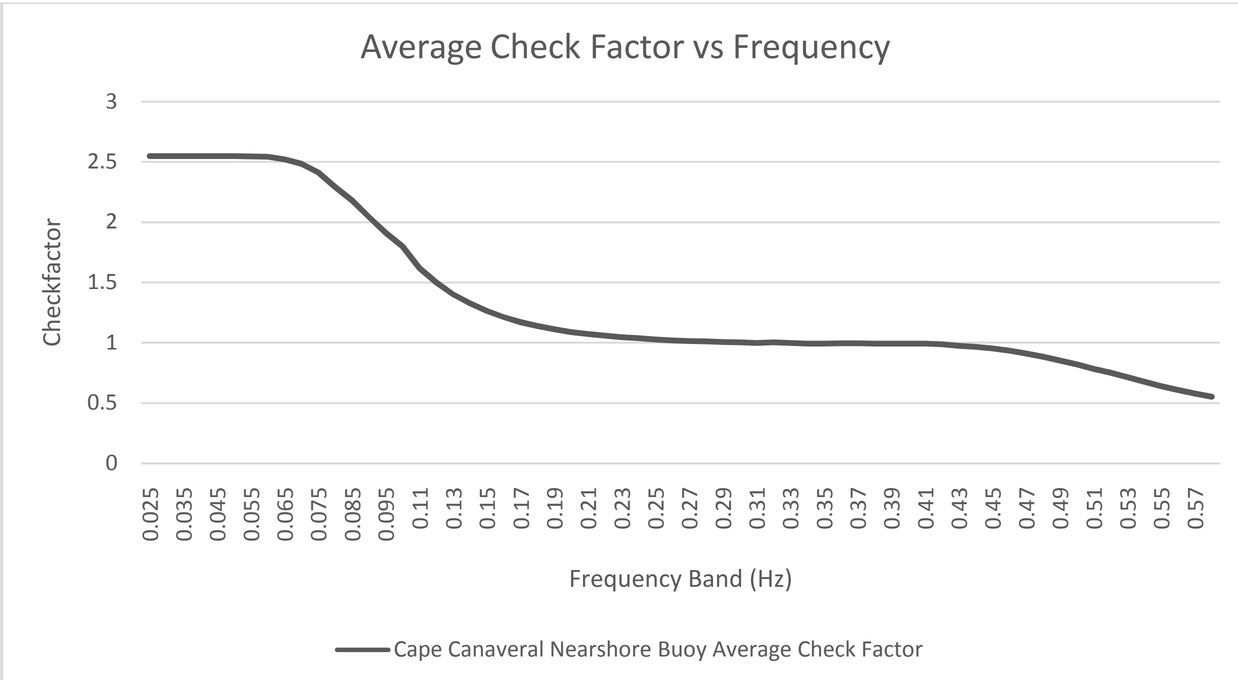

Figure 2. Normal Buoy Average Check Factor vs. Frequency

Since the low frequency energy bands typically have low energy and usually have inconsistent check factor data, the mid to high frequency bands were utilized for real time data quality control. Different data anomalies tend to affect different frequency bands, the frequency of the check factor out of range gives a clues as to the issue with the data. For example, bio-mass accumulation tends to add mass to the buoy and dampen its response to high frequency waves, the high frequency bands are particularly useful for indicating bio-mass accumulation. Figure 3 shows the Harvest California buoy with early indications of bio mass accumulation.

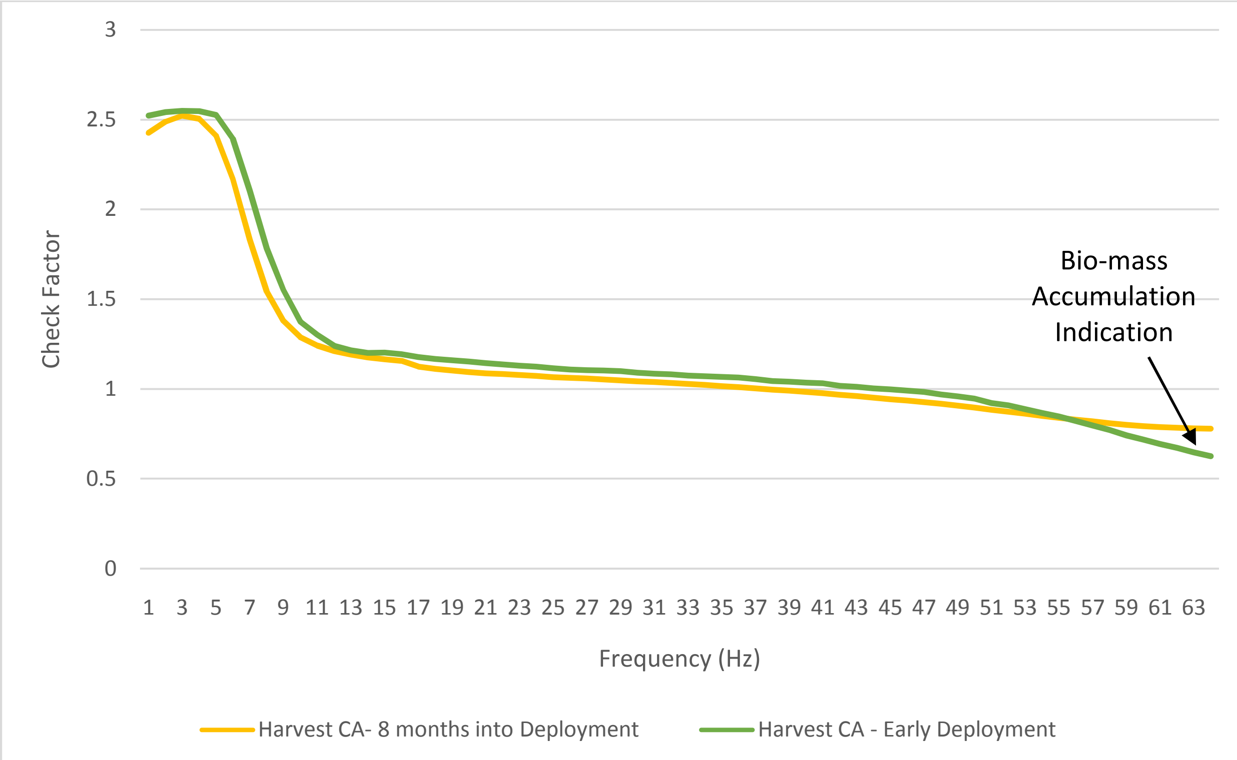

Figure 3. Early Indication of Bio-mass Accumulation

As indicated in Figure 3, the check factor values can be utilized for another form of quality control. In the upper and lower frequency ranges, the lack of energy and lower resolution from the buoy give check factor values that vary from the normal range (0.9 to 1.1). To tightly monitor check factor anomalies, the value where the computer flags the data varies with frequency. Figure 4 indicates a normally operating buoy check factor range, the Upper and Lower limit bars indicate the maximum and minimum allowable check factor value before the computer identifies an anomaly.

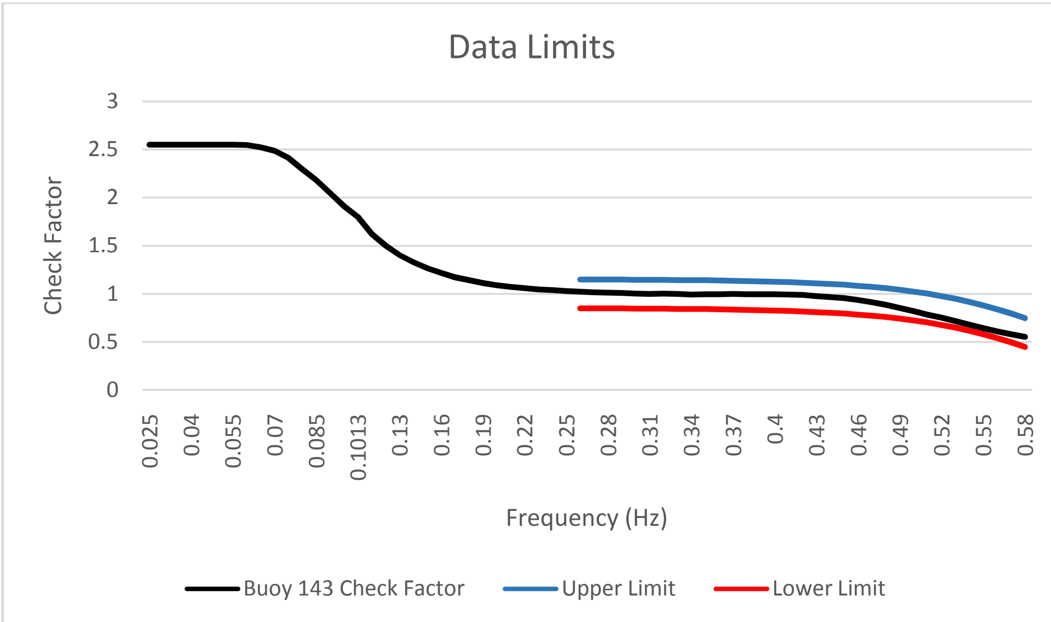

Figure 4. Data Limits

Figure 5 shows a buoy with extensive bio-mass accumulation. The high frequency check factors are out of the normal range and can be flagged accordingly.

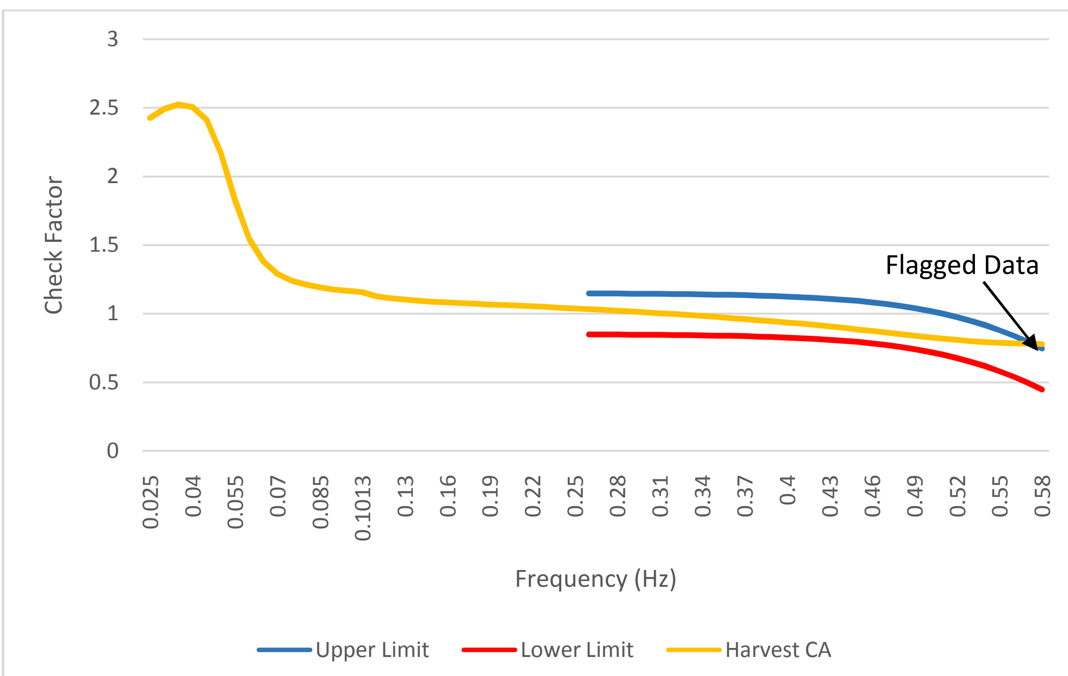

Figure 5. Individual Data Flagging

5. Using NetCDF Files

In the past, CDIP has had various forms of quality control looking at various factors of the data. However, the method of flagging the data to the user has been a simple public/non-public indication on the file. As CDIP transitions to a more interoperable NetCDF file format, the program now has the ability to flag individual data sets within the file. For example, if the high frequency data is deemed to be suspect or out of normal ranges, the individual data can be indicated as unacceptable quality, leaving the rest of the data available as reliable quality data. Furthermore, individual parts of the data can labeled with the source of the data anomaly. If the flag given suspects a particular anomaly, the file can display what the anomaly is and why the computer flagged it. The data can be flagged as good, suspect, or bad in accordance with Ocean Data Standards (UNESCO 2013 - IOC Manuals and Guides, 54, Volume 3 Version 1). Figure 5 in the previous section displays a buoy with high-frequency data flagged as 'suspect'. The rest of the data is within limits and can be deemed 'good' quality data.

Figure 6 displays a buoy at Central Cook Inlet in Alaska. This buoy was subject to heavy currents and stretched the buoys bungee to the maximum affecting the data. The figure shows the mid frequency data can be flagged as suspect, but the higher frequency data was within normal limits.

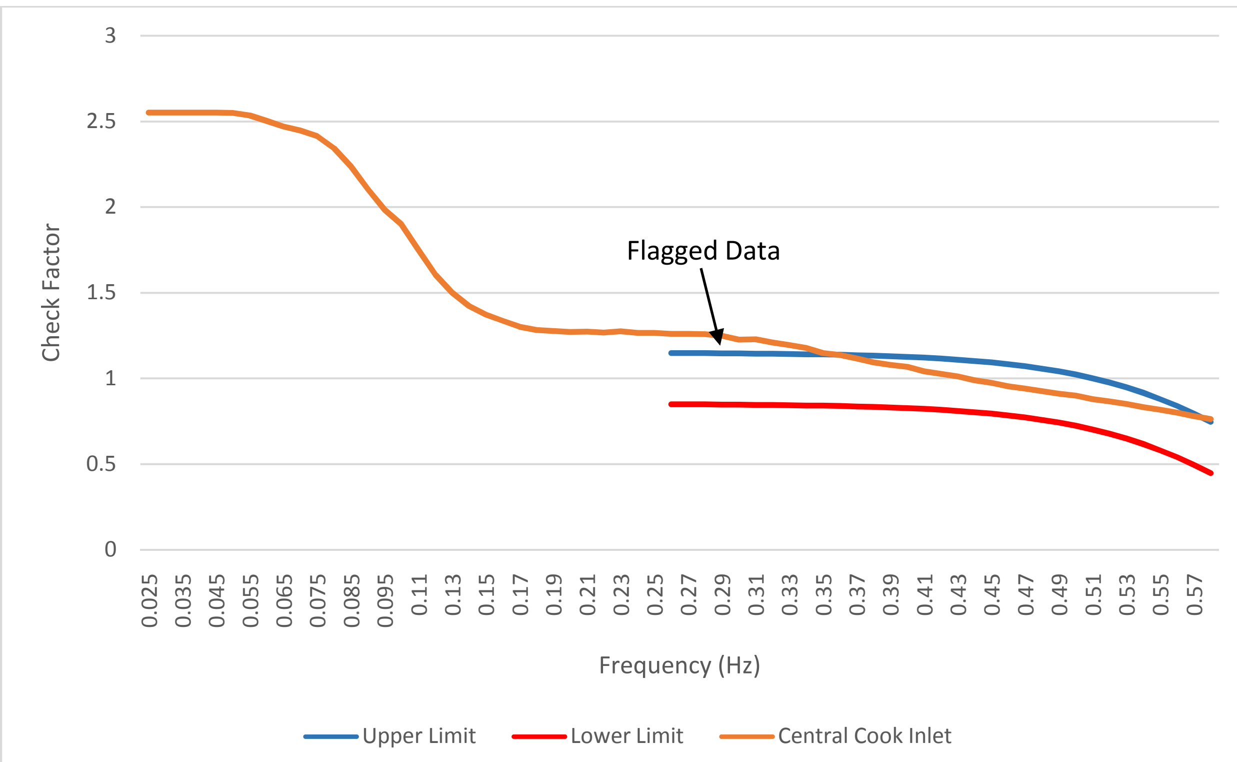

Figure 6. Cook River Inlet Anomaly

Utilizing check factors data and NetCDF files gives CDIP the ability to flag individual data sets, the end user has more flexibility with using the data and assurances that the individual data set of interest is of acceptable quality.