An example of Implementing Quality Codes for Time Series Data

Monique Harvey, Tasman District Council

Since the first of the National Environmental Monitoring Standards (NEMS) was drafted in 2010, Tasman District Council (TDC) has been applying quality codes to our continuous environmental monitoring data. NEMS were developed to stem a growing divergence in monitoring techniques. Their value comes from describing what good data looks like, what processes should be carried out to enable people to achieve a certain standard of data, and to allow you to apply quality codes to your data. The latter allows you to share your data more easily, ensures that your data is used appropriately, and it also provides a tool to analyse and report on how well you perform over time.

The NEMS quality codes are aligned internationally with the approach taken by the Open Geospatial Consortium (OGC) which also uses 6 identical codes, and this in turn aligns with the new WaterML 2.0 standard for the interchange of water data between organisations. The NEMS quality codes are shown in Table 1.

Table 1: NEMS Quality Codes

NQCS Quality Value	Quality Zone	Description
100	Missing Record	Data is Missing
200	No Quality OR Non Verified	Data is any of the following: - Raw, - Of unknown quality - Not assigned a final quality code - External and not coded to match NQCS
300	Synthetic	Estimated Data from Relationships and/or Calculations and/or Limited Measured Data.
400	Poor Quality	Measured Data which has been compromised in its ability to represent the monitored parameter. Or data has undergone significant modification.
500	Fair Quality	Measured Data that does not meet operational standard / best practices at the time of acquisition. The Data is considered to be a fair representation of the monitored parameter.
600	Good Quality	Data measured using best practice that meets operational standards at time of acquisition. Data considered to be a good representation of the monitored parameter.

The codes POOR, FAIR and GOOD should not be seen as indicating the quality of the work that has been done collecting the data. In fact the majority of the data we code as QC400 *(poor)* is collected for a specific purpose which doesn't require data of high accuracy, or it may be collected in a location where we are constrained in our ability to collect data of a higher standard.

Data types regularly coded at TDC include; water level, rainfall, groundwater level, water temperature and conductivity.

All telemetered data is assigned a code of QC200 (*No Quality or Non Verified)* when it is transferred from Hydrotel to Hilltop via the HydroHill process. Manual data downloads from loggers are also assigned the QC200 code when the file is imported into Hilltop Manager.

The data remains as QC200 until it is checked, processed and archived to our data archive. It is during the final copy to the archive that a quality code of either QC400, QC500, or QC600 is assigned, depending on:

- Site setup
- Sensor and logger type/setup
- Deviation from reference measurement or verification checks.

Missing record (QC100) is processed as soon as possible after it has occurred, or synthetic data (QC300) is created and these are coded appropriately. A processing register is maintained for each site tracking data movement into the archive and quality codes used.

When we commenced this process, one NEMS standard was drafted and subsequently changed. At this stage the quality codes assigned under that previous version of NEMS have been left as they are. As an example, it is now "easier" to obtain QC600 data for rainfall – in the draft standard the acceptable deviation between primary reference gauge and the intensity gauge was 5% or less, and now it is 10% or less.

Historic Data

To allow for analysis across the entire record at a site, quality codes have been assigned at all of our open river water level recording sites (40 sites) right back to the start of the record. This has been an ongoing project, and has even seen the tape reader dusted off!

The graph in Figure 1 shows the water level record for Motueka at Woodstock, starting in 1969. Variations in colours on the graph indicate changing quality codes. It allows any unique site specific issues to be displayed in the data, over time or over a stage range. Here at Woodstock between 1998 and 2009 there was an issue with partial blocking of the tower intakes, for stage levels between 1200m and 3000m. The data in this stage range was coded QC400 to show lower reliability of data.

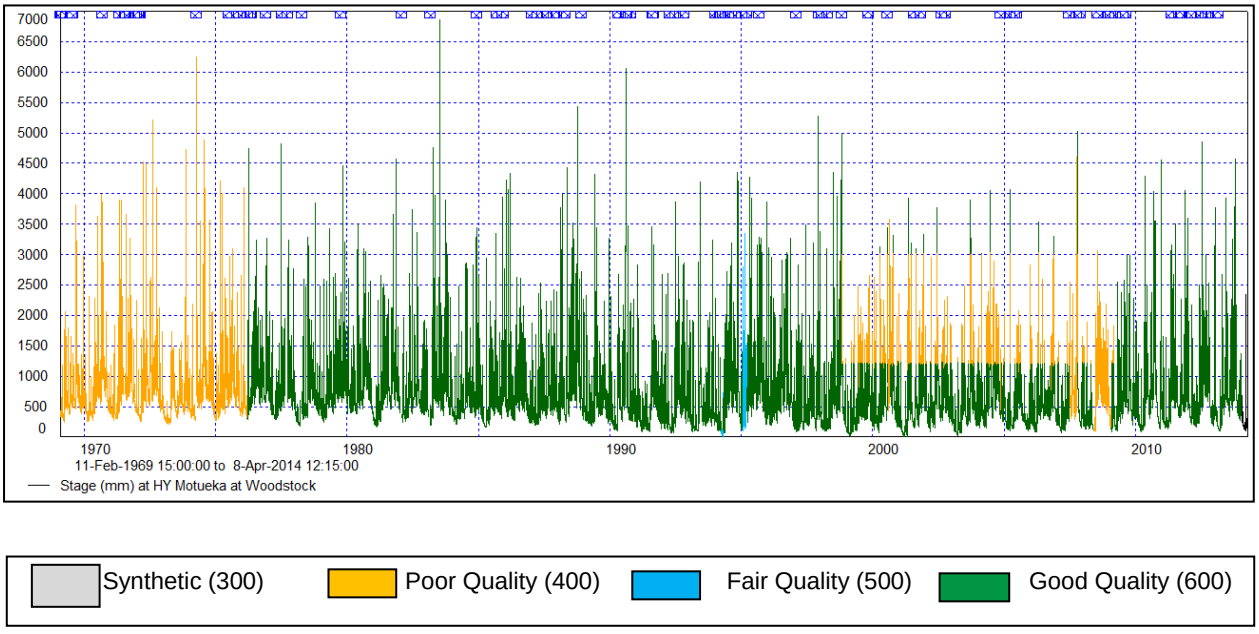

Figure 1: Graph of quality coded stage data for Motueka at Woodstock (shared NIWA site)

Assigning quality codes to historic data from open rainfall sites is now underway, and half of the 47 sites have been done so far. Historic quality codes are assigned based on the type of site setup, logging interval and sensor resolution.

Analysis of Quality Codes

Aside from colourful graphs of our data, quality coding data allows for quick analysis of data quality over time with individual or grouped sites.

Using Hilltop Hydro we are able to break down archived data by quality code, and show proportions of missing record, synthetic record and data quality. We can also show if there is any data without a quality code.

The pie chart (Figure 2) shows a summary of data quality for all water level, rainfall and groundwater data in the last quarter of 2013. A PSummary process is run on a collection of sites in our archive, with the output giving both the number of days and proportion of time for each code class.

We also create the same chart based on the 'best case quality code achievable'. That is, for each site we assign a maximum achievable quality code and then can compare how sites measured up over the quarter.

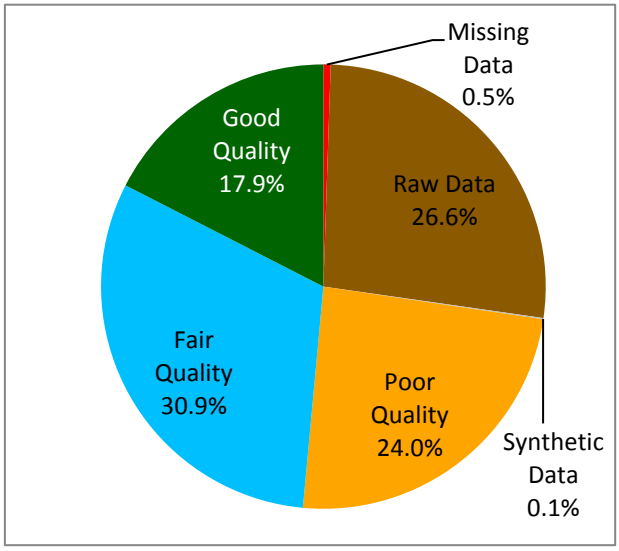

Figure 2: Summary of data quality for all water level, rainfall and groundwater data in the last quarter of 2013 at all sites

The graph in Figure 3 shows for all current water level sites back to 1960, the percent of missing and synthetic record for each year, and (hopefully) the declining trend of less of each as technology and skills improve. The red line indicates the days of data recorded each year, and although this is increasing, the percent of missing or synthetic record is not increasing. This type of analysis is made possible as quality codes have been assigned back in time.

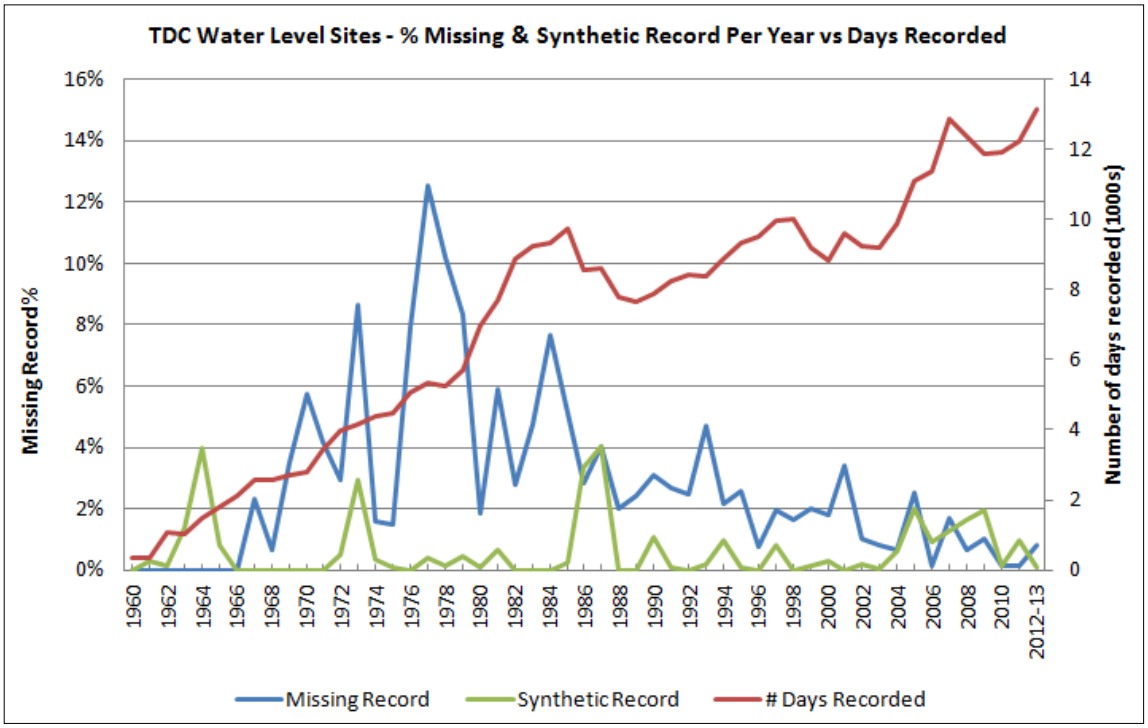

Figure 3: Percent of missing and synthetic record for each year for all current water level sites back to 1960

Implementing quality codes to both historic and current data has allowed us to show improvements in data quality over time as new technologies improve data acquisition, as well as highlighting areas where site changes/upgrades would be beneficial to improve data quality. Many of our groundwater level sites don't have the recording resolution or accuracy at present to achieve a QC500 or QC600 code, however a code of QC400 may be sufficient for the purpose it is being recorded for.

It allows for easy tracking of missing and synthetic record and has enabled us to set realistic targets for both.