

Primary author: **Scott, Nicola** (Met Office, Saughton House, Broomhouse Drive, Edinburgh, EH11 3XQ, UK), nicola.scott@metoffice.gov.uk

Abstract ID: 7P4

Automated Quality Control of Marine Observations at the UK Met Office

So much data.....So little time!

This is fast becoming a wide spread issue, with so much money being invested into new data collection methods, particularly automated systems, volumes of available data are steadily increasing. However, only a fraction of this investment is allocated to the equally important task of data management and in many cases this cost is being reduced year on year. So the question is 'How do we continue to quality control (QC) an increasing amount of data to meet user requirements with less resources'? The answer.....Automation!

This may seem quite obvious but it is essential to strike the right balance between automation and some manual intervention.

The UK Met Office has invested time during 2008/2009 in automating the majority of its marine data QC processes to increase efficiency and to standardise quality.

The majority of problems highlighted using marine QC software are due to general internal inconsistency which are mostly too difficult to correct with confidence. Whereas, out-of-range and rate-of-change problems can have a bigger impact but can be easier to correct. So the Met Office have used this theory combined with a variety of user requirements to develop a mostly automated marine QC system.

The poster presentation will show in detail the approach the Met Office have taken in automating and where it can add greater value by using a little manual effort. The poster will also quantify the efficiencies made and as a result the amount of data now able to be processed.

UK VOS data is made freely available in delayed-mode through the Global Collecting Centres (GCCs).

gcc@metoffice.gov.uk

TIME SAVINGS

By implementing this new QC system the amount of processing time saved has allowed much more data to be processed. Within the Met Office there is now one full-time employee processing delayed-mode VOS data. This consists of around **100k UK observations and 1 million observations from other countries** (received via the GCCs). Previously this work would have needed three full-time employees to process giving a **saving around 70%**.

USER REQUIREMENTS

There are various users of the Met Office's Marine Climate Data including **climate researchers, educational institutes, legal enquiries and commercial businesses**. Each user has their own specific requirements and it can be difficult to cover all requirements within one dataset with an automated QC system. However, common requirements include always having **access to original data**, only changing data if sure the new value is correct and only deleting data if certain it is erroneous and could impact on averages and extremes etc. By using a predominately flagging system the user can judge for themselves (with guidance) how to use the data.

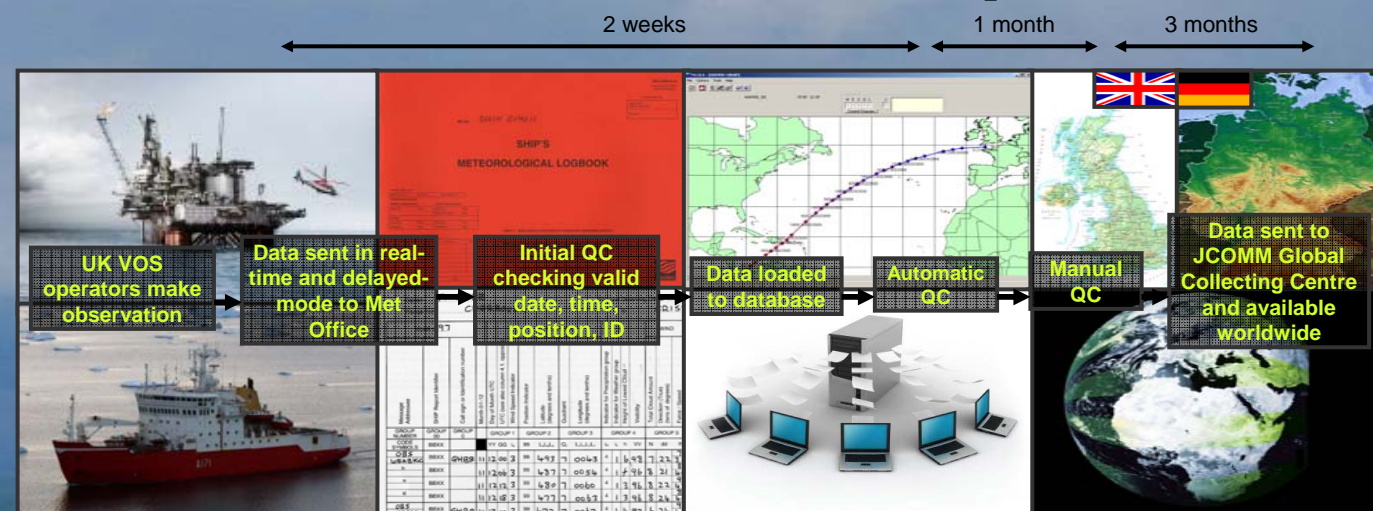
The Met Office's new Marine Data QC system strikes a balance between automation and manual QC. With over 70% of checks that highlight some less significant inconsistencies – the system automatically sets flags to convey this information – whereas more erroneous issues which could have greater impact are handled manually.

QUALITY METADATA (aka Flags)

Flagging data in a useful way is essential for conveying information on quality to the user, but it is even more important when QC is (mostly) automated. Deciding on a quality metadata structure is no easy feat, balancing the benefits of lengthy, complex and expensive to store flagging, oppose to short, simple and less expensive to store.

The Met Office has used a more simple flagging structure for each met/ocean element within an observation (see table below for details). Each element flag has a dedicated column within the Oracle database – an example of a flag might be 1006 (where QC is complete (L=6) but the value was estimated manually (E=1) and is no longer suspect) or 0141 (where only automatic QC is complete (L=1) and the elements is suspect (S=1) failing a range check (Q=4)).

Met Office Marine QC process



QC CHECKS

There are **299 quality checks** carried out on each observation loaded to the Met Office Oracle climate database. The table gives details of the amount and types of checks and (for UK data) which aspects of QC are done automatically / manually. Automatic = 213 checks, Manual = 86 checks.

Check type	Number of checks	Automatic / manual
On Land	1	Manual
Out of Range	70	Manual
Rate of Change and Speed	15	Manual
Internal Consistency	210	Automatic
Comparison with Climatology (mean ± 4 StDev)	3	Automatic

Flag	Flag name	Description	Values	Meaning
E	Estimate	Indicates if an estimate has been made to the element	0	No estimate made – value original
			1	Value estimated manually
			2	Value estimated automatically
S	Status	Indicates if element appears suspect	0	Value not suspect
			1	Value suspect
Q	Query	Indicates which type of QC check has failed	0	Original value not queried
			4	Failed range check
			5	Failed internal consistency / climatology check
L	Level	Indicates the stage to which QC has been carried out	6	Failed rate-of-change check
			0	No QC carried out (since ingestion)
			1	Automatic QC complete
			6	Automatic and Manual QC complete
			9	Level of QC unknown (old data yet to be processed)

DATA ACCESS

Once a quarter the Met Office contribute their data to the JCOMM Marine Climatological Summaries Scheme (MCSS) via the Global Collecting Centres. Data from the last quarter for all MCSS Contributing Members are then available to download from a Met Office FTP site.

Ftp site: ftp.metoffice.gov.uk
Username: marinegcc
Password: tevygu8a