GLOS QA/QC

Overview

GLOS provides a variety of data to the public through our public-facing Data Portal as well as other tools and products. Because much of this data could be used for decision support processes, GLOS is committed to providing valid, quality data to the user. For the most part, GLOS has relied on processes which put the burden of quality assurance on the data provider and ensures the levels of quality by vetting the providers and requiring through the contractual agreement that they manage and monitor quality control. Quality control tests are performed by GLOS on all real-time partner data. GLOS does not allow suspect data to be displayed in any of our products.

I. Quality Assurance Procedures

GLOS relies on contracted partners to provide the best practices for quality assurance (QA) on specific instruments. The principle investigators (PIs) are internationally recognized scientists with a great deal of expertise making in-situ and remote sensing observations in their particular field. All instruments are deployed and maintained by their respective research groups and they advise the GLOS DMAC team as to quality assurance. Equipment are calibrated, operated, and maintained in accordance with manufacturing guidelines and/or national IOOS program guidelines, when available.

An example of this type of quality assurance is that performed by LimnoTech which has contracted responsibilities for ten buoys and ten fixed sensors.

- Purchase of quality instruments
- Adherence to manufacturer guidelines for calibration and deployment
- Performance of intra-deployment cleaning and re-calibration
- Regular monitoring of reported observations and advising GLOS of suspect data for censoring
- Returning sensors post-deployment to manufacturer for cleaning and calibration

GLOS obtains near-real-time data from external partners including federal and non-federal agencies or organizations. For all external partners providing data to GLOS, GLOS DMAC screens the provider to ensure the provider is an authoritative source and is maintaining the observational equipment that collects the original source data according to published manufacturer standards and in line with scientific best practices. At this time, contracted partners report on their QA/QC procedures informally through annual in-person meetings and in consultation with the GLOS DMAC team.

Given that, measures are being put into place to a) require as part of the contract process at a minimum annual quality assurance reports – part of the 2016 contract process; and b) implement QARTOD compliance and testing by October 31, 2017.

II. Data Quality Control Procedures

Quality control procedures are described below GLOS data streams. Additional detail is shown in Table 1.

Sensor Data: Sensor data provided by contracted partners make up the bulk of our non-federal data ingestion. Raw data are generally assembled by the data provider and transferred to GLOS as near-real-time datasets. GLOS's current quality control process consists of the following steps:

- Feed/continuity monitoring: The GLOS DMAC ingest procedure tracks receipt of new data and alerts the DMAC team when data are not received on schedule so that appropriate action(s) can be taken
- Syntax check: The GLOS DMAC ingest procedure validates the structure of data submittals and rejects malformed XML; missing values are placed in the data stream where data are not available.
- Range check: The GLOS DMAC ingest procedure performs parameter-specific range checks based on limits derived from observation history and informed by subject matter experts.

In addition, if GLOS is informed by users or providers that data have issues, GLOS manually addresses the issue if possible or removes the data from our products.

GLOS plans on implementing all required QARTOD tests by 10/31/2017 for parameters with QARTOD manuals and will also identify and implement appropriate tests for non-QARTOD parameters.

Additional information is available elsewhere, including:

- Appendix A for sample programming for these validation tests;
- Table 1: Summary of Data Served
- Asset/Inventory spreadsheet (separate document)

Glider Data: GLOS does not perform QC on glider data because these data are submitted directly to the IOOS Glider DAC (Data Assembly Center) by the Provider. In addition, access to glider data is not currently provided to the public via any of our GLOS products. GLOS anticipates adding transient access to GLOS partner glider data by December 31, 2016, relying on Glider DAC QC and data services.

Satellite Products: A set of quality tests (flags) are applied by the Provider (MTRI) during image processing for each standard product to ensure only valid data are generated.

Historical and Other Data: GLOS data partners may also process and store data, then later transfer it to GLOS as part of an historical or archival process or to provide larger amounts of data to the user community. As an example, GLOS maintains a collection of manually curated datasets in netCDF format. These datasets contain various observations collected from discrete as well as continuous sources and generally are posted either in a one-off basis or updated annually. QA/QC information for manually curated datasets, where available, is referenced in the associated metadata records. Going forward,

GLOS will ensure that QA/QC information is included in or linked from the associated metadata record for all datasets.

These datasets include:

- MTRI Ranger III data – vessel of opportunity data, uploaded annually. Range checking is completed for all data, and the sampling instruments are compared to independent ones before deployment. If values exceed the acceptable range, they are not used in the resulting output data sets.
- MTRI Great Lakes Optical Properties Database (GLOPGD) – geodatabase. Data submitted have passed an extensive suite of unit tests which require data to meet quality standards before inclusion.
- MTRI thermistor string data – in-situ water temperature. Range checking is completed for all data, and the string is compared to an independent instrument before deployment. If values exceed the acceptable range, they are not used in the resulting output data sets.
- GLERL hydrology – Great Lakes hydrology-related datasets developed by NOAA's Great Lakes Environmental Research Laboratory – federal data source

Federal Data: Production data services operated by federal agencies are assumed to comply with the QA/QC policies of that agency. GLOS DMAC monitors data transmission characteristics to screen for transfer errors or service interruption but does not perform quality checks on the actual data.

III. QARTOD

In order to improve QA/QC, GLOS is currently working with IOOS and other RAs to implement a comprehensive QARTOD plan that will cover all of GLOS's ingested non-federal sensor datasets. GLOS currently anticipates implementation of all required and selected recommended QARTOD tests for available sensor-collected parameters by 10/31/2017.

Non-federal, non-sensor datasets (for example, MTRI Great Lakes Optical Properties Database (GLOPGD)) are publicly exposed through GLOS only if sufficient QA/QC documentation are also available. GLOS is responsible for the gathering and assessment of QA/QC documentation for these datasets, though IOOS agreement may sometimes be sought as to the sufficiency of documentation prior to publication of the data.

IV. Summary of QC for Data Served by GLOS

Table 1. GLOS-Supported (Non-Federal Partners)

GLOS currently performs feed monitoring, syntax check, and range check tests on all ingested non-federal real-time partner data, while glider data are submitted to and QC'd by the IOOS Glider DAC. Ranges used in the range check tests are shown for real-time parameters. Additional partner QA/QC requirements will be instituted in the next partner contract cycle (starting 2016), while additional tests based on QARTOD guidance will be implemented for real-time partner data.

Platform	Variables	QC Validation Tests	Performed by
Buoys (non-federal/GLOS supported) *(NOTE: most GLOS buoy data are also submitted to the NDBC, and can be distinguished by the NDBC ID on the Asset Inventory. However, the data provided in the GLOS products for those buoys or stations come from the GLOS feed in order to avoid latency issues and to provide content that is not otherwise available via NDBC.)*	water temperature air temperature dew point relative humidity significant wave height wave period wind direction wind speed maximum wind gust barometric pressure	between 0-40 deg C between 0-50 deg C between -30 and 50 deg C between 0-100% between 0-10 m between 0-15 sec between 0-360 deg between 0-50 m/s between 0-50 m/s between 700-1200 mmHg	GLOS
Stations (Weather/Environmental)	wind direction air temperature relativity humidity	between 0-360 deg less than 0-50 deg C between 0-100%	GLOS
Stations and/or Buoys (HABs)	wind direction wind speed maximum wind gust air temperature water temperature significant wave height wave period dew point barometric pressure relative humidity turbidity chlorophyll phycocyanin (blue green algae) specific conductivity water temperature pH dissolved oxygen	between 0-360 deg between 0-50 m/s between 0-50 m/s between 0-50 deg C between 0-40 deg C between 0-10 m between 0-15 sec between -30 and 50 deg C between 700-1200 mmHg between 0-100% between 0-1000 NTU between -5 and 120 RFU between -5 and 120 RFU between 100-700 uS/CM between 0-40 deg C between 6-10 between -1 and 25 mg/L	GLOS

	dissolved oxygen saturation thermal string	between 0 and 150% between 0-40 deg C	
Gliders			Provider submits to Glider DAC
Satellite - MODIS (non-federal)	lake surface temperature lake surface chlorophyll natural color imagery color dissolved organic matter dissolved organic carbon suspended minerals	A set of quality tests (flags) are applied during image processing for each standard product to ensure only valid data are generated. MTRI remote sensing research professionals test input reflectance data to make sure negative values are not used in the water quality retrieval algorithm (Shuchman et al. 2013[1], Shuchman et al. 2006[2]). Retrieved values are then excluded for each product that do not fit into a predefined range of accepted values. Finally, a visual inspection of the retrievals is performed to determine if the images are complete enough (e.g., they are not overly patchy due to sporadic cloud/haze cover) to post.	Provider (MTRI)
Thermistor String Data	water temperature	Range checking is completed for all data, and the string is compared to an independent instrument before deployment. If values exceed the acceptable range, they are not used in the resulting output data sets.	Provider (MTRI)

[1] Shuchman, R. A., Leshkevich, G., Sayers, M. J., Johengen, T. H., Brooks, C. N., & Pozdnyakov, D. (2013). An algorithm to retrieve chlorophyll, dissolved organic carbon, and suspended minerals from Great Lakes satellite data. Journal of Great Lakes Research, 39(S1):14-33.

[2] Shuchman, R., Korosov, A., Hatt, C., Pozdnyakov, D., Means, J., & Meadows, G. (2006). Verification and application of a bio-optical algorithm for Lake Michigan using SeaWiFS: a 7-year inter-annual analysis. Journal of Great Lakes Research, 32(2):258-279.

Ranger III Ship Mounted Sensor	various	Range checking is completed for all data, and the sampling instruments are compared to independent ones before deployment. If values exceed the acceptable range, they are not used in the resulting output data sets.	Provider (MTRI)
Great Lakes Optical Properties Geospatial Database (GLOPGD)	various	Data submitted have passed an extensive suite of unit tests which require data to meet quality standards before inclusion.	Provider (MTRI)

Table 2. FEDERAL Partners

The following providers are federal partners and no GLOS QA/QC is performed. However, we do verify whether the data is sound and will not allow suspect data into the GLOS data stream.

Provider	Description	Service type
Center for Operational Oceanographic Products and Services (CO-OPS)	buoys or stations	SOS
Coordinated Lake Levels	water levels	THREDDS
Great Lakes Environmental Research Laboratory (GLERL)	hydrologic datasets (NBS, atmospherics, precipitation, air temperature); HABs monitoring	THREDDS
Great Lakes Coastal Forecasting System (GLCFS POM)	model result datasets	THREDDS
HECWFS Huron to Erie Connecting Waterways Forecasting System FVCOM	model result datasets	THREDDS
National Data Buoy Center	buoys or stations *(NOTE: most GLOS buoy data are also submitted to the NDBC, and can be distinguished by the NDBC ID on the Asset Inventory. However, the data provided in the GLOS products for those buoys or stations come from the GLOS feed in order to avoid latency issues and to provide content that is not otherwise available via NDBC.)*	SOS
National Estuarine Research Reserve System	buoys or stations	SOS

NOAA Coastal Relief	bathymetry	THREDDS
NOAA Regional Science Consortium	buoys or stations	SOS
NWS	buoys or stations	SOS
SLRFVM Upper St. Lawrence River FVCOM	model result datasets	THREDDS
U.S.C.G. Marine Reporting Stations	buoys or stations	SOS
USGS	buoys or stations	SOS

Appendix A

Details of Current Validation Tests

The *glos_obs* package checks the validity of the observed floating point number for observation values. For example, for surface water temperature, the Java codes below make sure only a valid floating point number will be populated. Otherwise, a “Float.NaN” will be inserted to indicate an invalid record.

```
nlist=message.getElementsByTagName("wtmp1");
                                        if(nlist!=null&&nlist.getLength()>0)
                                        {
                                            try
                                            {
        val=Float.parseFloat(nlist.item(0).getTextContent());
                                                if(val!=missingVal)
                                                        o.setWaterTemperature(val);
                                                else

        o.setWaterTemperature(Float.NaN);
                                            }
                                            catch(NumberFormatException e)
                                            {
                                                o.setWaterTemperature(Float.NaN);
                                            }
                                        }
```

Before the observation record is injected into the database, *glos_obs* will conduct a range check on the observation data. The valid range is based on observation history and input from subject matter experts such as the buoy owners.

```
if(!Float.isNaN(result=ob.getWaterTemperature()))
{

if(result>0.0001&&result<50.0)
{

        Ids=getSensor(pstSelSensor,pstInsSensor,

platformId,ConfigManager.getWaterTempSensorId(),

ConfigManager.getWaterTempMeasureId(),
                                                                        "WTMP",
                                                                        alt);

        InsertResult2D(pstInsObs,platformId,Ids[0],ConfigManager.getWaterTempMeasureId(),o
b.getDate(),lon,lat,alt,result);

                                                }
                                        }
```

Similarly, *obs2nc* has Lua code to validate observation data as they are placed in NetCDF files for archival submittals.

```
    sea_surface_water_temperature={

              standard_name="sea_water_temperature",

              long_name="sea water temperature at surface",

              featureType="timeSeries",

              units="degree_Celsius",
```

```
            source="platform/45167/45167_sea_surface_water_temp",
            depth=0.0,
            keywords="EARTH SCIENCE > OCEANS > OCEAN TEMPERATURE > SEA SURFACE
TEMPERATURE",
            keywords_vocabulary="GCMD Earth Science Keywords. Version 5.3.3",
            validator=function(val)
                        if val>100 or val<-10 then
                            return false
                        else
                            return true
                        end
                end        }
```