

SOP 11: Data Quality Assignment (OC_M Creation) version 2

Summary

At the end of each survey year, a review is done to determine if recalibration data indicate that sensor drift over the season is significant. The original log sheet images are also reviewed for annotations regarding data exceptions. The raw and processed vertical profile data plots are reviewed both separately and jointly for irregular instrument readings.

The Project Leader flags questionable and unusable casts for the year using the data quality report spreadsheets. Inoperable individual sensors encountered during a cast are also noted on the sheets. The Data Manager, after validating the spreadsheet forms, posts the forms to the web and then updates appropriate columns in the cumulative database to reflect quality levels of all observations.

Detailed Steps

A. Project Leader tasks

1. Obtain a blank Data Quality Report form from the “Toolbox” link on the SEAN website.
 1. Its form is an XLSX workbook.
 2. Repeatedly copy the “template” tab until there are enough blank spreadsheets to hold every dump in the cruise year.
 3. Rename each tab with the four-digit dump number it reflects.
 4. Head the top of each form with cruise-specific information taken from the OC_H field log sheets.
 5. Record results of the analyses described below on the appropriate sheet.
 1. Use the analysis order shown below.
 2. If a cast is marked “bad” at a point in the process for any reason, no further quality review for that cast is performed.
 3. If a cast is marked “questionable,” review continues in order to determine if it should be further marked down to “bad.”
2. Review the latest certified calibration certificates.
 1. Examine the certified images on the SEAN website of the pre- and postseason calibration documents (deliverable OC_G).
 2. Following the guidance in SOP 9, note any data quality exceptions on the data quality report sheets.
 3. Be aware it is possible this may impact the entire season if a serious drift was present.
3. Review all certified field log sheet images for the year on the website (deliverable OC_H).
 1. Individual cast exceptions may be found in the COMMENTS field of the field log sheet image.
 2. Overall cruise exceptions and problems may be found in the NOTES field of the log sheet.
 3. Be sure to check the back side image of each sheet for any further quality details.
 4. Mark technically flawed casts as “bad” on the quality spreadsheet.
 1. Use the codes listed at the bottom of the sheet, which follow the standard used by the Ocean Data View system.

2. Include a brief explanatory comment for any bad cast. Each comment will be embedded in the Data Quality Comment column of the cumulative database in order to inform future researchers when they select data to work with.
3. Data Quality Comments should *not* be recorded for casts having “Good” quality.
5. Mark and comment casts as questionable if one or two sensors appear faulty for some or all of a cast.
 1. It isn’t possible to provide a quality rating at the CTD/dump/cast/individual-sensor level, so any problem has to be recorded for the cast as a whole.
6. Check appropriate boxes on the Data Quality Sheet to mark any failed or disconnected sensors for particular casts or for the dump as a whole.
 1. When the OC_M is later applied to the OC_D database, this will cause the database to have nulls in the corresponding cells, removing the wild readings typical of broken sensors.
4. Review the raw in-season profile plots for individual casts made for quality assurance purposes during the season (SOP 7).
 1. Mark and comment as “questionable” casts having one or two improbable plot lines.
 2. Mark and comment as “bad” casts displaying three or more improbable plot lines.
5. Review the processed postseason profile plots made for individual casts made for quality assurance purposes after the season (SOP 8).
 1. Mark and comment as “bad” casts for which nothing is plotted.
 1. The signal processing applications discard observations that are widely outside acceptable limits. In a seriously failed cast it is possible for all points to be deleted.
 2. Compare processed postseason plots against raw in-season plots.
 1. If strong discrepancies appear, investigate possible causes.
 2. The Project Leader’s judgment, following currently accepted practices in the field, will determine whether to mark casts as questionable or bad, based on in-season/postseason plots.
6. Submit the deliverable XLSX file via email attachment to the Data Manager for validation.
 1. Specify in the message body it is deliverable OC_M for cruise year YYYY, as defined in protocol OC-2014.1.

B. Data Manager tasks

1. On receipt of the submission, assign the next formal submission number to this file.
 1. Use the “Update Submission Log” web tool on the NPS network at http://sean-dm.glba.nps.gov/0_submission_update.aspx.
 2. Complete Submission_Log details up through the Submission_Date column.
 3. Use Cruise Year as the submission unit.
2. If this is the first OC_M submission for the cruise year, update the deliverable tracking spreadsheet with the date of first submission under OC_M.
3. Save the attached file to the staging area for validation as \\sean.glba.nps.gov\data\SEAN_Data\Staging\OC\OC_M\nnnn\Quality_YYYY.xlsx, where nnnn is the submission number and YYYY is the cruise year.
4. Verify the OC_A/OC_G postseason calibration deliverables have been certified.

1. Postseason calibration provides required information for producing OC_M.
2. Be sure the date for “Postseason calibration assessment” is recorded in the OC_M column of the deliverable tracking spreadsheet. Add it if necessary.
5. Validate the XLSX file according to current criteria.
6. Record validation deliverable status in the Submission_Log using the web tool.
7. If submission fails mandatory criteria,
 1. reply with a “failure email” that includes:
 1. The submission number assigned
 2. The deliverable ID
 3. The protocol ID
 4. Documentation listing all the specific mandatory and optional criteria failed
 2. Update the submission log with error counts, date validation attempted, and status of “F” for failed.
8. If submission passes mandatory criteria, reply with a “success email” that includes:
 1. The submission number assigned
 2. The deliverable ID
 3. The protocol ID
 4. Documentation listing any specific optional criteria failed
 5. Request to certify deliverable as complete

C. Project Leader tasks

1. On receipt of a failure email:
 1. Make corrections so the deliverable meets mandatory criteria.
 2. Restart the process from the beginning to initiate a new submission with the corrected deliverable candidate.
2. On receipt of a success email, review any failed optional criteria:
 1. If these are acceptable:
 1. Reply with a “certification email” stating the deliverable is certified and may be applied to the database and disseminated.
 2. If these are unacceptable:
 1. Reply with a “withdrawal email,” stating the deliverable is withdrawn.
 2. Take remedial action to obtain a corrected deliverable.
 3. Restart the process from the beginning.

D. Data Manager tasks

1. On receipt of a withdrawal email:
 1. Mark the withdrawal in the Submission_Log’s Status column with the web tool.
 2. Terminate the process.
2. On receipt of a certification email:
 1. Verify the OC_D for the cruise year has been marked as certified. If not, notify the Project Leader that certification cannot be registered until OC_D has been certified and suspend processing until OC_D is ready.
 2. Due to the nature of these data, no sensitive information is in the deliverable.

3. Generate a PDF version of the single spreadsheet file containing all sheets in the workbook to staging directory, naming it Quality_YYYY.PDF.
4. Copy the staged certified spreadsheet and PDF files to the Development environment at ...\\AUXREP\\OC\\OC_M\\.
5. Create and test web page links to the new file.
6. Copy the revised files to the Staging/Integration environment and test they are accessible.
7. Update the staging database contents to reflect quality flags and adjustments listed on the OC_M.
 1. Use the OC_M update tool on the Data Management website at http://sean-dm.glba.nps.gov/OC_DM_update_OCM.aspx.
 2. Apply updates with care. There is no way to undo application of OC_M without recreating the underlying OC_D data.
8. Mark the certification in the Status column in Submission_Log using the web tool.
 1. Mark any previously certified submission for this unit as decertified.
9. Update the deliverable tracking spreadsheet with the date of certification for OC_M.
10. Propagate from Staging/Integration to Production environment, including revised database content.
 1. Due to infrastructure limitations, moving database content to Production is queued up but only executed about once per year.
11. Verify deliverable(s) are accessible from Production website.
12. Update the deliverable tracking spreadsheet in Development with dates for the Repository and Dissemination milestones.
13. Create a PDF from the latest tracking spreadsheet containing all years and propagate it to all websites.
14. Perform an update for both the NODC and IRMA Data Store copies of the revised OC_D data, if they have previously been run through certification.