

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/235011246>

Development and Testing of Local Automated Glider Editing Routine for Optical Data Control

Article · May 2010

CITATION

1

READS

92

4 authors, including:



Weilin Hou

United States Naval Research Laboratory

102 PUBLICATIONS 1,015 CITATIONS

SEE PROFILE



Robert Arnone

University of Southern Mississippi

277 PUBLICATIONS 5,627 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Maritime Advanced Geospatial Intelligence Craft [View project](#)



Dynamic anomaly bio-physical hotspots in Gulf of MExico [View project](#)



NRL/MR/7330--10-9247

Development and Testing of Local Automated Glider Editing Routine for Optical Data Quality Control

WEILIN HOU
DEREK BURRAGE
MICHAEL CARNES
ROBERT ARNONE

*Ocean Sciences Branch
Oceanography Division*

May 10, 2010

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 10-05-2010		2. REPORT TYPE Memorandum Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Development and Testing of Local Automated Glider Editing Routine for Optical Data Quality Control				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 0305160N	
6. AUTHOR(S) Weilin Hou, Derek Burrage, Michael Carnes, and Robert Arnone				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 73-9056-00-5	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory Oceanography Division Stennis Space Center, MS 39529-5004				8. PERFORMING ORGANIZATION REPORT NUMBER NRL/MR/7330--10-9247	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Space & Naval Warfare Systems Command 2451 Crystal Drive Arlington, VA 22245-5200				10. SPONSOR / MONITOR'S ACRONYM(S) SPAWAR	
				11. SPONSOR / MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT In this work, the automated glider optics quality control algorithms are described, followed by the testing of the performance statistics using optical data from various gliders including Slocum and Seaglidars, at different locations and involving over 7,000 profiles. Visual inspection by experts validates the performance, with errors less than half a percentage point.					
15. SUBJECT TERMS Glider Quality control Automated					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Weilin Hou
Unclassified	Unclassified	Unclassified	UL	40	19b. TELEPHONE NUMBER (include area code) (228) 688-5257

TABLE OF CONTENTS

Introduction.....	1
Automated Optics QC.....	3
1 Program Structure	3
2 Testing areas	4
3 Calibration.....	4
4 Optics Quality Control tests and flag setting at each depth.....	4
4.1 Global bounds check.....	5
4.2 Surface Depth test.....	6
4.3 Spike test.....	6
4.4 Running Standard Deviation test	6
4.5 Constant Profile test.....	7
4.6 Backscatter Buddy check.....	8
4.7 Interpolation over spikes.....	8
4.8 Convert beta to bb.....	8
4.9 Smooth optics variables	8
5 Overall optics quality flags	9
6 Error Flag Statistics.....	9
7 Validation Using Optical Variable Transects and Anomalous Profiles.....	13
Conclusions.....	19
References.....	19
Table 1. NRL/NAVO glider optical parameters for BUFR table.....	20
Table 2. Glider optical parameters in BUFR format for LAGER Optics V1.0.	21
Table 3. BUFR Quality Control (Class 63)	22
Table 4. Statistical Results under Different Testing Conditions.....	23
Table 5. Overall flag stats for all profiles tested.....	24
Appendix A.....	29

Introduction

This report gives an overview of the structure and performance of LAGER Optics v1.0 (Local Automated Glider Editing Routine: Optics, version 1), using field exercises data. LAGER Optics v1.0 is the first version of a suite of algorithms for automated data processing from optical sensors within the ocean glider quality control system developed for the Naval Oceanographic Office (NAVO) Glider Operations Center (GOC). The LAGER Optics v1.0 is an upgrade to the successful LAGER v1.0, which has been transitioned to NAVO and currently operational in NAVO GOC. The LAGER v1.0 methods and operation are fully documented in Carnes (2008) and Carnes and Hogan (2009). LAGER Optics v1.0 is a core part of LAGER v2.0, which also includes algorithms updates and conversion from previous Matlab functions to FORTRAN subroutines. The complete description of LAGER 2.0 will be found in a separate document. LAGER v1.0 was designed and developed primarily to process data from temperature and conductivity sensors (from which salinity and density are derived). It was developed to import raw data returned from ocean gliders operated by NAVO, to detect and flag erroneous observations, and to reformat data and flags into a format accepted by the NAVO Real-Time Data Handling System (RTDHS) in Binary Universal Format for the Representation of Meteorological Data (BUFR, details below), adapted for oceanographic data types. The apparently erroneous glider observations, which cannot be corrected automatically, require manual intervention. An interactive GUI-based manual utility GUI (MUG) for man-in-the-loop quality control is used to change the quality control (QC) flags and/or to edit the observations using the controls available in the GUI, operated by NAVO GOC personnel. The initial version is written in Matlab (MathWorks Inc.). A FORTRAN version of MUG, is under development, and will be delivered under LAGER v2.5. This validation test report is primarily aimed to test the automated optics QC module, which was also first developed in Matlab and subsequently rewritten in FORTRAN 77/90 to enable faster processing in an operational environment.

To answer the needs of Mine Warfare (MIW), Anti-submarine Warfare (ASW), safety of navigation, monitoring global climate change, and battlespace environment sensing, optical sensors have been fitted and tested on different types of gliders, and have proven successful in assessing environmental conditions. They provide inputs as well as validation to system performance prediction, diver visibilities, optimizing 3-dimensional ocean optical forecasting, and benefit higher level tactical decision making. The data flow can be viewed as a 5-step process, with the inclusion of optics, as shown in Figure 1.

When fully implemented, LAGER v2.0 will incorporate the Optics quality control (QC) algorithm suite described in this report. The optics algorithms will be used to process data from several optical sensor types, including instruments that measure the beam attenuation coefficient (c), the total scattering coefficient (b), the backscatter coefficient (b_b), the fluorescence returns and thus derived concentrations of chlorophyll, phycoerythrin, and colored dissolved organic matter (CDOM), downwelling irradiance (E_d), photosynthetically available radiation (PAR), and other derived optical parameters such as diver visibility (Table 1). All current optical sensors including ECO series from Wetlabs (bb2f, bb2slo, fl3, SAM, BAM, AUVb, ECO-PAR) and OCR504/7 from Satlantic are included in the current tested version, with flexibility built in to allow for expansion to meet future needs. Additionally, Table 1 outlines the optical parameter

typical range, resolution, bandwidth, and channel limits, which are used in the LAGER Optics v1.0 routines and will be discussed in later sections.

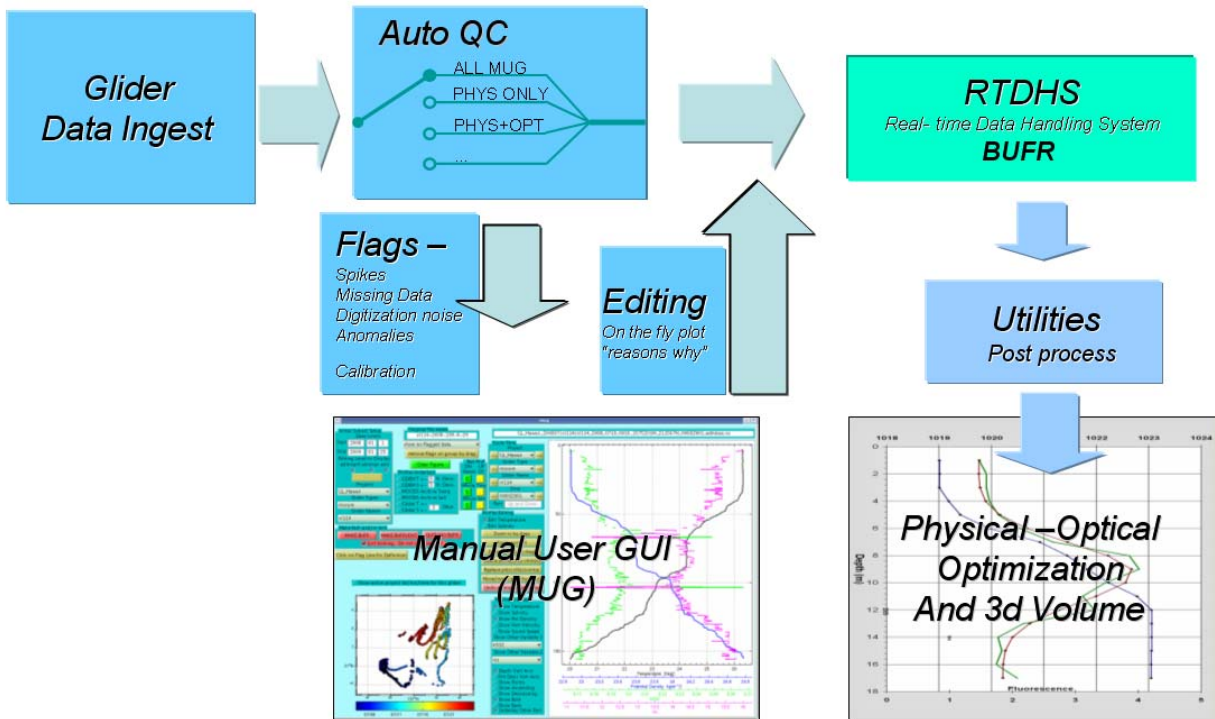


Figure 1. Data flow chart of LAGER Optics v1.0. Data ingest: reads in real-time raw data at NAVO, in converted netcdf format. Auto QC is the key step in LAGER Optics v1.0 and the main subject of this test report. It uses flags to classify different types of erroneous data and the combined quality of the flags determines whether Manual User GUI (MUG) is needed for closer inspection. The resulting data is then sent to the database in BUFR format required by RTDHS.

To meet automated data QC, downstream model requirements and archive needs, BUFR format is adapted from meteorological research to be used at NAVO GOC. All glider sensor outputs including temperature, salinity are examined and converted in BUFR data tables. Details of such tables can be found in Mangin 2009 (Ref 3). The NRL/NAVO optics table (Table 1) has been derived and incorporated into NAVO BUFR tables, with an assigned Class 60. For convenience of discussion, the optical variables corresponding to Table 1 are shown in Table 2, adopted from Ref 3 for convenience.

This report describes and validates the automated LAGER V2.0 (Optics V1.0) algorithms using backscatter, fluorescence and irradiance observations from several Slocums and Seagliders operated by NAVO and by NRL during recent exercises and research experiments. During the recent US Navy Arabian Gauntlet 2009 exercise, LAGER was used operationally by the NAVO GOC to test its performance. The validation reported here is primarily intended to demonstrate that the LAGER scheme for deciding whether to send an optics profile for manual QC, based on the automated QC results, is adequate. The scheme works well if it sends profiles to manual QC only when there are unflagged errors remaining in the file after the automated QC operations are

applied. A complementary validation of the hydrographic observation QC results has already been performed as part of the LAGER 1.0 reporting process and is described in Carnes and Hogan² (2009).

Automated Optics QC

1 Program Structure

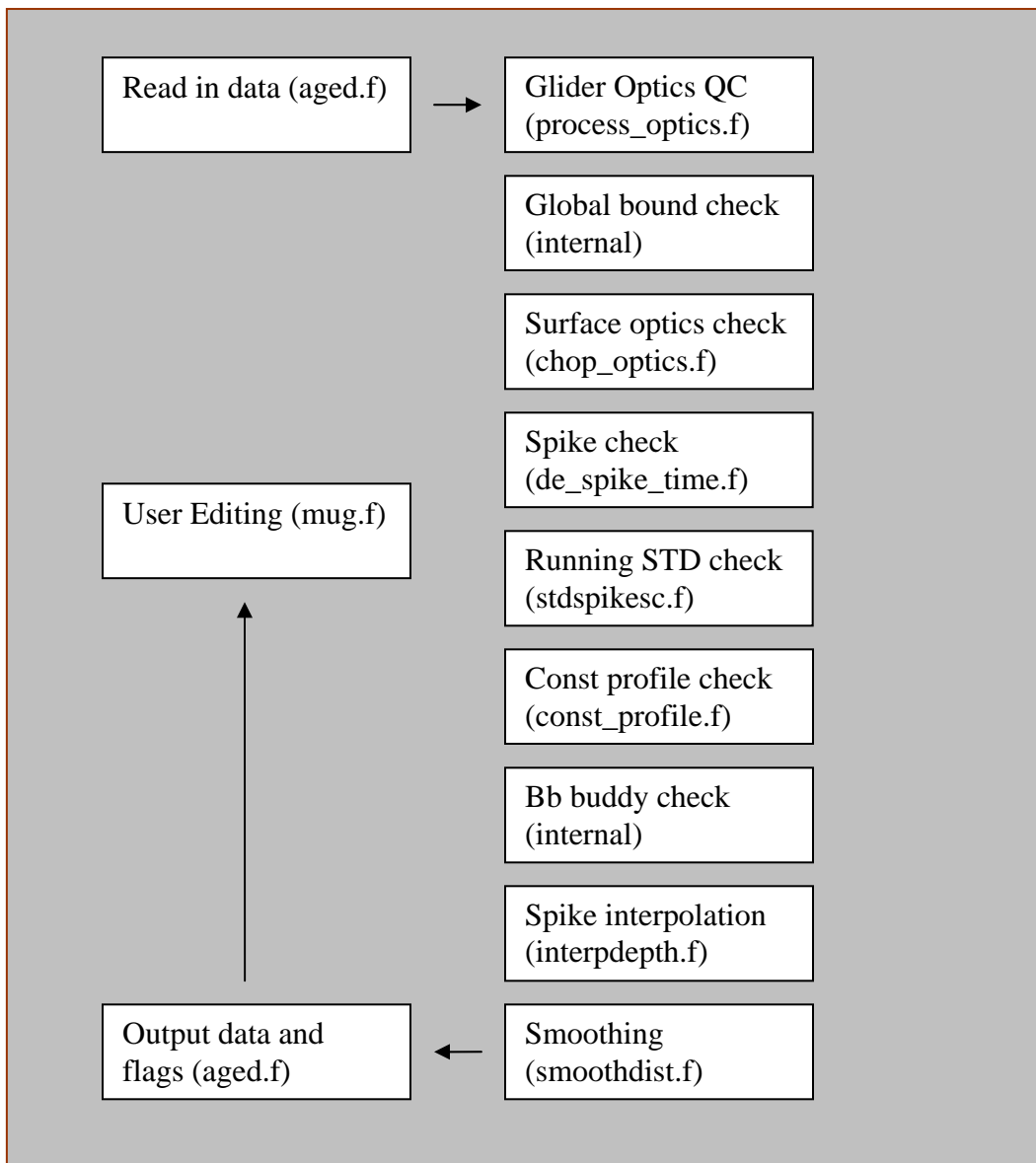


Figure 2. LAGER Optics V1.0 data quality check flow chart. Notice that manual QC in FORTRAN is currently in beta stage and will be delivered in LAGER V2.5.

2 Testing areas

Data from 8 glider deployments at different locations consisting of 7,067 profiles (Table 4), from Biospace (2008), RIMPAC 2008 exercise, Arabian Gauntlet 2009 exercise are used for statistical analysis, necessary for overall program parameter settings. Exercise Arabian Gauntlet 2009 and Exercise RIMPAC 2008 deployments are used to validate the LAGER Optics V1.0 AutoQC module. Detailed area maps can be found in subsection 7 below.

3 Calibration

Calibration of the optics variables is generally performed prior to the optics QC steps. However, in cases where the process_optics.f subroutine determines that calibrated data are already available (i.e., the sensor data was calibrated during deployment, or immediately after glider recovery), that step is skipped. The calibrate_optics.f subroutine called by process_optics.f loads calibration data from files stored in the LAGER/SETUP_FILES/calibration folder, based on the type of sensor variable (e.g., backscatter, bb, downwelling irradiance and such.), its frequency band (e.g., 470, 532, 670, 700nm, etc.) and the date of the dive (the calibration files contain calibration parameter data pertaining to a range of dates, spanning the intervals between successive calibrations). For most sensors, only a simple linear calibration formula operating on the sensor raw counts and parameterized by a scale factor and offset (typically dark counts) is needed. This contrasts with the calibration of the hydrographic data, which requires a larger number of parameters, specifying higher-order polynomials, for determining temperature and conductivity, with the additional complications that the salinity is a function of both temperature and conductivity, and there are significant response lags that must be taken into account to obtain valid profiles¹ (Carnes 2008). Future versions will apply more sophisticated approaches in calibration and correction algorithms, including but not limiting to fixes to compensate for surface fluorescence signals.

4 Optics Quality Control tests and flag setting at each depth

Quality control tests for optics variables implemented in LAGER, follow certain algorithms and conventions for detecting and flagging bad data that were developed for processing the hydrographic eg., temperature, T, and salinity, S, data. However, because of the multiplicity of optics variable types and frequency bands, a 2D variable array structure is used in contrast to the 1D vector approach used individually for T and S. The reader is referred to the LAGER 1.0 manual¹ (Carnes, 2008) for details of the hydrographic data QC implementation. Two of the hydrographic quality control tests (algorithms) implemented in LAGER are used with minor adaptations for the Optics QC (OQC). These are the Global bounds check and the Spike test. The latter defines a spike to be a single datum departing significantly from its neighboring values (single-point spike). It has the additional feature of being able to detect such a spike in the presence of a gradient (as a function of depth) of the relevant hydrographic or optical variable. Since segments of anomalous optical data may span several data points, a method to detect these cases, in the form of a running standard deviation filter, is also included in the OQC. This filter finds and flags data that depart significantly in value from a local mean computed inside a ‘running’ window (depth interval) that is moved through the entire sampled depth range of the profile.

In addition to the variable value range check, the depth of each sample point is also checked and ‘chopped’ short, if necessary, to eliminate values that appear to lie above, or too close to the surface, where processes such as wind waves and bubbles could make optical measurements such as backscattering and downwelling irradiance either invalid or difficult to interpret reliably. If this depth is shallower than a given depth (default 1 m), or is negative, the variable value is flagged accordingly. A constant profile check is also done to determine if the values for a particular optics variable are constant throughout the sampled depth range, which was proven critical in assessing real-time optical sensor performance such as those during the RIMPAC 2008 exercise. This might occur for a variety of reasons, such as instrument sampling faults, or sensor contamination, which suggest the data are invalid. Details concerning the implementation of these filters are given in the following subsections.

In the present LAGER Optics 1.0 version, all the OQC tests are independent of geographic location and time of year, and the tests are also independent of depth, except in cases where different critical test values are used in two different depth ranges. The universal character of the tests weakens their capability to detect erroneous anomalies in specific regions. However, most of the parameters used to configure the OQC algorithms are stored in information files, and may be edited to adapt them manually, if necessary, to specific regions, or seasons. A future version of LAGER will include capabilities to adapt these parameters to specific regions automatically, based on a regional climatology or historical data, or alternatively, specific regimes, such as shallow or deep ocean settings, or proximity to the coast, and relationship to other physical parameters. This is particularly important for optical data, for which the dynamic range varies widely, depending upon the prevailing oceanographic setting or conditions. In highly productive, nutrient-rich regions and/or close to estuarine freshwater sources, values of Chlorophyll-a and Colored Dissolved Organic Matter (CDOM) at the surface, or in the pycnocline, may reach highly-elevated values, and may exhibit very steep property gradients, that are not observed in deeper water.

For the hydrographic variables, several glider-specific tests have been added to the OQC to detect and flag specific known types of bad behavior exhibited by either specific brands of gliders or by all types of gliders. In most cases, the glider-specific tests are functions of the vertical velocity of the glider which is employed as a substitute for the more-difficult-to-determine total speed of the glider through the water. There are some early indications that similar provisions will need to be made for the optics data. However, insufficient data have been accumulated to clearly establish the need for such corrections, and to reliably specify a remedial strategy.

Optical error flags are initially set to zero at each depth. Tests are performed on observed temperature and salinity values prior to any tests performed on optical values. If a test is failed at a specified depth, then the corresponding flag is set to the failure flag value for that test at that depth. The sequence of tests is described next.

4.1 Global bounds check

The algorithm for this check is implemented in the FORTRAN subroutine, `process_optics.f`. $V_flag(i) = 2$ if $V(i) < Vmin$ or $V(i) > Vmax$, where $V(i)$ denotes the value of a particular optics

variable (e.g., backscatter at 660 nm, or bb660), at depth $Z(i)$, i is the depth index, and V_{min} and V_{max} denote the extreme values representing the acceptable natural range of values of that variable. The values of parameters V_{min} and V_{max} are defined in the /LAGER/SETUP_FILES optics_variables_info.dat file in the LAGER/SETUP_FILES folder (shown in Appendix A), adapted from Table 1 and 2. Note that $V_{flag}=1$ is not used.

4.2 Surface Depth test

The algorithm for this check is implemented in the FORTRAN subroutine, chop_optics.f, called by process_optics.f. The flag value is set to 7 ($V_{flag}(i) = 7$), if $depth(i) < D_{max}$, where D_{max} is the maximum chop depth, maxchopdepth, defined in the process_optics.f subroutine. This will be moved to the optics_variables_info.dat file, in a subsequent version.

4.3 Spike test

The algorithm for this check is implemented in the FORTRAN subroutine, de_spike_time called by process_optics.f. The flag value is set to 4 ($V_{flag}(i) = 4$), if $|V(i)-(V(i+1)+V(i-1))/2| - |V(i+1)-V(i-1)| > K$, where $K = K_s$ ($Z(i) < 500$ m) or $K = K_d$, ($Z(i) \geq 500$ m). K_s and K_d are the minimum deviations that could represent a spike in the depth ranges 0-500m or ≥ 500 m, respectively. Notice that the spike test utilizes the optical parameter table (Table 1), particularly the resolution of the instruments for this task. Changes can be made through values defined in LAGER/SETUP_FILES/optics_variables_info.dat (Appendix A).

4.4 Running Standard Deviation test

The algorithm for this check is implemented in the FORTRAN subroutine, stdespikesc.f called by process_optics.f. $V_{flag}(i) = 5$ if $|(V(i)-V_{mu}(i-w/2:i+w/2))| > L$ where $L = stdevfactor * V_{sd}(i-w/2:i+w/2)$, w is the width of the moving window, V_{mu} and V_{sd} are the local running mean and standard deviation, respectively, of the data falling within this window, and stdevfactor is a specified fraction or multiple of V_{sd} . If however, the absolute difference $|(V(i)-V_{mu}(i-w/2:i+w/2))| < abslimfactor$, where abslimfactor is a specified multiple of the sensor resolution, then the datum is not flagged. The rationale for this is that deviations that do not depart significantly from the running mean may be retained as they could be accounted for simply by normal instrument noise fluctuations. They will, in any case, be eliminated or reduced in subsequent smoothing operations. The values of stdevfactor and abslimfactor are specified in the optics_variables_info.dat file which defines QC parameters for each variable type.

This filter is adaptive in that it detects values that are anomalous to the extent that they deviate significantly from the local mean value, with ‘significantly’ referring to a large departure, or outlier, relative to the local variability of the variable, or normal instrument noise. Because the center of the running window moves through successive depth indices, i , each datum is assessed in the context of its spanning window. It is thus possible for adjoining data values to be considered anomalous (outliers) and to be flagged accordingly.

A special provision is made in this test to ensure anomalous values are detected near the very top and bottom of the profile. This is done by using a fixed window, within which all data are tested, at the top and bottom of the profile, while a running window, centered on each datum in succession is used at intermediate depths. As a consequence the test is biased slightly deep and shallow, respectively, near the top and bottom of the profile, but this is considered a minor tradeoff considering the importance of testing all data values near these depth extremes.

4.5 Constant Profile test

The algorithm for this check is implemented in the FORTRAN subroutine, `const_profile.f`, called by `process_optics.f`. V flag(i) = 6 if $\max(V(i)) - \min(V(i)) < V_{res}$, where V_{res} is the resolution (minimum digitized sample increment) of the given sensor. The resolution is defined in the `optics_variables_info.dat` file (Appendix A). This is a critical test to detect malfunctions of newly developed optical sensors and/or deployment errors.

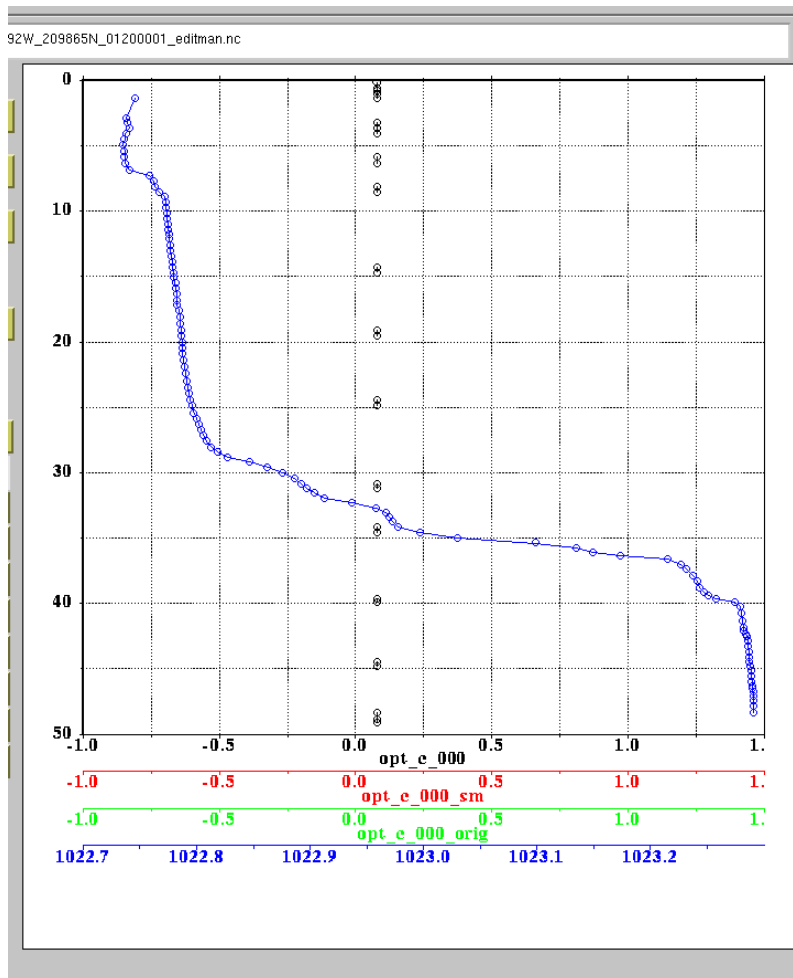


Figure 3. Example to show profiles in question, using Exercise data from RIMPAC 2008. The current constant check alerts the system and sends the profile in question to manual QC for inspection. Density profile is shown in blue.

4.6 Backscatter Buddy check

In the algorithm `process_optics.f`, there is a check to see if all backscatter bands simultaneously exhibit flag 4 (single point spike) or flag 5 (standard deviation anomaly) at the same depth level. If that is the case, the error flags are turned off. The rationale for this is that a spike or anomaly occurring in all the bands, is unlikely to be an isolated instrument or sensor sampling error, but is likely to be due to an actual observable change in water mass characteristics.

4.7 Interpolation over spikes

The algorithm `interpdepth.f` called by `process_optics.f` optionally interpolates linearly over all missing data (gaps) in the profile, as well as replacing erroneous (flagged) values with interpolated values. The interpolation is linear, with respect to depth. The option to interpolate over missing data gaps is presently turned off, to limit the risks involved in extrapolating data over large depth ranges, at points removed from actual observations.

4.8 Convert beta to bb

Although not shown in the flow chart (Fig. 2), the algorithm `beta2bb.f` is called from `process_optics.f` to convert the original and interpolated backscatter sensor data from `beta`, which is observed at a particular backscattering angle (114° currently), to `bb`, which is a measure of backscatter integrated over the back scattering angles (90 to 180) observed by the sensor. This derivation is presently done after the optics QC algorithms are applied. In a future version, consideration will be given to doing this conversion prior to QC, to avoid spurious values appearing due to errors in calibration offset.

4.9 Smooth optics variables

After the QC algorithms are applied and interpolation has been performed to correct flagged values, a smoothing operation is also applied. The operation is performed by the `smoothdist.f` subroutine, called by `process_optics.f`. The resulting smoothed variable is output separately from the unsmoothed interpolated data. The algorithm smooths by averaging the data within a moving window spanning a specified depth range. The window size, which is currently applied to all optics sensor data is set to 5 m. This parameter value was selected, based on spectral analysis of the backscatter data, which shows that it coincides with its spatial decorrelation scale. The decorrelation scale represents the ‘cutoff’ wave length where the ‘red’ spectrum of the backscatter signal levels out at a low value presumed to coincide with the instruments noise floor. In a future version of the code, this parameter will be made accessible through the `optics_variables_info.dat` file (Appendix A), so that its value can be customized, if necessary, for each individual sensor type.

5 Overall optics quality flags

As discussed previously, the following flags are used for current optics QC modules:

Flag name	assigned value
Global bounds check	2
Bb buddy check	4
Running STD check	5
Constant value check	6
Surface depth check	7

The LAGER Optics 1.0 automated QC sends a multi-sensor multiple-band optics file for further evaluation in the manual QC step if, for any of the individual profiles in the file, the number of profile observations flagged as bad is greater than a given percentage (default 15%, for the current validation test, this value is set to 10%) of the total number of observations of that profile. In future versions, additional criteria will be used to determine if profiles are bad enough to be deleted or require manual editing on the basis of the temperature and salinity or optics variable values. (See Ref 1 for details concerning the corresponding criteria for the hydrographic data). Hence, a file may be sent to manual QC because either or both of temperature or salinity require manual editing, or because at least one of the optics profiles requires manual editing. In addition, an individual T, S or optics variable profile may be flagged for deletion if the number of bad data or gaps are excessive. A `needs_manual_editing_flag` is assigned in each file to each ascending or descending profile (each value summarizes quality of the corresponding optics variable profile).

`needs_manual_editing_flag = 0` indicates that manual editing is not required.

`needs_manual_editing_flag = 1` indicates that the file containing this profile should be sent to the manual editor for further examination even if the `needs_manual_editing_flag` for the other profile (if it exists) in the file is set to 0 (doesn't need manual editing).

`needs_manual_editing_flag = 1` if:

1. If a given percentage (default 15%) of the depths have V flag > 0, but not including V_flag = 7.
2. for Slocum profiles only, at least one depth gap between consecutive observations was ≥ 10 m or two gaps ≥ 6 m.

6 Error Flag Statistics

The goals of LAGER include automatically detecting and flagging erroneous observations prior to being passed to the RTDHS, and identifying apparently erroneous glider observations that cannot be corrected automatically, and thus require manual intervention. It is thus desirable, to automatically detect and correct as many of the errors as possible, while minimizing detection failures (failure to detect bad data) and false alarms (incorrectly rejecting good data). Since manual intervention is labor intensive, we also want to minimize the number of data being sent (to MUG) for Manual QC.

There is, however, inevitably, a degree of subjective judgement involved in determining the appropriate parameter to use in the AutoQC algorithms described above, as well as in using Manual QC operations to determine, 'by eye', whether an error has occurred. This subjectivity can only be reduced by using 'expert judgement' to determine the parameters and to assess the performance of the QC operations. Hence, that is the approach we have adopted in this report. An independent assessment of physical T and S variables is possible by comparison with climatological data bases¹ (e.g., GDEM, see Carnes, 2008). However, such climatologies are not yet available for bio-optical data (backscatter, attenuation, fluorescence etc). Furthermore, the bio-optical data exhibit a wider dynamic range, appear in general to be more 'noisy', and are subject to greater environmental variability than is the case for T and S. This makes the optical variables both prone to error, while being subject to fewer 'realistic' constraints. Determining whether a particular profile requires manual intervention is thus to some degree an arbitrary process. However, as long as the 'arbitration' is carried out by an expert knowledgeable in measurement instrumentation, and experienced in acquiring and editing bio-optical data, the results can be expected to be satisfactory.

In this first version of the OQC, we use a test based on a simple bad data threshold to determine if a profile requires manual QC. This test is based on the following rationale: If the percentage of data that was detected and corrected by the AutoQC process exceeds a certain threshold (default 10% for current validation test), it is likely there are other problems or undetected errors with the data set, that may require manual correction. Similarly, in this case, a manual check is considered desirable, to confirm that the automatic detection and correction of a relatively large number of errors was successful.

In the next section, we discuss the validation of the Automatic and Manual QC operations, based on a range of data from various missions and geographical zones, using an 'expert judgement' approach. In the remainder of this section, we present and discuss statistics that show what percentage of the data files are sent to manual QC for a given bad data threshold, as applied to each profile and variable type appearing in the file. We also give statistics indicating the reasons why the various profiles were flagged as bad (based on the QC algorithms applied). These statistics are applied to data from various geographic regions and they are accumulated over all profiles and all files obtained from a specific glider within the specified region and sampling period. For practical reasons, the sampling period is defined to be a calendar month or less.

We note here that, in the terminology of LAGER, a 'file' contains all the data from a single 'dive', which in general might include several descending and ascending profiles, while a single profile generally contains data obtained from multiple sensors and wavebands (eg. bb470, bb532, fl2 cdom & etc.). While the existing seagliders typically generate files with multiple profiles, the currently available slocums generate files with only one, or a pair of (ascending and descending) profiles. In the existing version 1.0 of LAGER, however, all files, regardless of glider type, are constructed so they comprise a single profile, or at most, a pair of ascending and descending profiles. In a future version of LAGER, this constraint will be relaxed, so that multiple profiles (or profile pairs), comprising a complete 'dive' will appear in a single file. For the purpose of our statistical analysis, therefore, a file generally contains 1 or 2 profiles.

Results of the statistical analysis of the percentage number of files that are sent to Manual QC are shown in Table 4. This Manual QC percentage is based on the percentage of bad data appearing in each of the profiles, considering all the variables/wavebands in each profile appearing in the file. If one or more of the individual variable/wave band records has more than the specified percentage threshold of bad data, then the corresponding records are marked for Manual QC, and the entire file (dive) is sent to Manual QC (Mug). Under operational conditions, files for which none of the individual variable/band records exceed the threshold would bypass Manual QC and pass directly to the BUFR file creation process (see Carnes, 2008 for details). For the purpose of the validation tests discussed in the next section, however, all the files were sent to manual QC, and each erroneous profile was checked to determine if it was, indeed, flagged for Manual QC or not.

In Table 5, statistics for particular areas and calendar months (col 1) and specific gliders (col 2) are shown along with the number of files containing valid optics data (col 3). A series of tests was performed using these data sets, which counted the number of errors of any type used to flag the data appearing at each depth level, and occurring in each variable/wave band in the file. The specified percentage threshold levels of bad data (col 4) was chosen, not necessarily exhaustively, from the range 1, 2, 5, 10, 15, 20, 25, 30 %. The corresponding percentage of files in the data set that would be sent to Manual QC, given this threshold is then shown (col 5). A Manual QC (%) value of 0.00 indicates that less than 0.01 % of data would have gone to Manual QC – a negligible number). For most data sets, the percentage going to manual QC is small when the threshold is high (high tolerance for bad data), and it increases monotonically (but not necessarily gradually), as the threshold is reduced (low tolerance for bad data). In exceptional cases, however, the Manual QC (%) starts at a relatively high level and increases as the threshold is reduced. The exceptions apply exclusively to Slocums and, as will be discussed below, occur only in cases where a large number of profiles are flagged as being uniform in value (and hence highly likely to be erroneous). Ignoring these exceptions the Manual QC (%) values for a specified bad data threshold of 15 % are generally smaller than 7%, (with one exception of ~ 20%) for Slocums (slxxx), and 1% for Seagliders (sgxxx). The Seagliders appear to be generally less noisy (fewer spikes and other anomalies, see next section) and subject to errors than the slocums. However, this might partly be a consequence of sampling bias, since the Seagliders are often deployed in deeper waters (subject to less variability), further from the coast (lower values and weaker gradients), and sample over a wider depth range (tendency for most variables to be small and uniform at depth) than is the case for the Slocums.

Table 5 also gives an indication of the likelihood (percentage) of individual flag values appearing in data set, with the flag counts (expressed as a percentage of the total number of data) accumulated over all variables/bands and files. The flag types listed are 2 for data out of range, 4 for a single point spike, 5 for a running standard deviation anomaly, 6 for a constant profile and 7 for data that appears within 1 m of (or above!) the surface. In the case of flag 6, if this occurs at all (and this is the case only for slocums), it is either small (indicating a small number of constant profiles) or very large (indicating a large number of constant profiles). When large, this number tends to be exceptionally large, because only cases where the entire profile is uniform are detected (as distinct from shorter segments of a profile, that otherwise contains good data), and because, in that case, all data points in the profile are flagged as bad. During AutoQC, when it is possible to interpolate over bad (flagged) data, the number 30 is added to the error flag.

However, for the purpose of this analysis, flagged data corresponding to both interpolated and uninterpolated data were combined (i.e., flag values of 32, 34 or 35, where counted along with flag values of 2, 4 and 5, and the resulting sums were given the labels 2, 4 or 5, respectively). In the case of flag types 6 and 7, interpolation is never done, and corresponding data values are instead flagged as missing (and given special values), so summation was not necessary, in that case. Note that data with flag 7 is not necessarily erroneous (it could have been correctly measured), but is excluded because of the likely complications for optical data interpretation in the near surface (surface water wave and bubble affected near surface zone).

For Slocums, as a general rule, flag values of 2 (range check) appear in less than 0.3 % of the data, while values of 4 (single-point spike) can be as high as 5.3 %. Standard deviation anomalies are slightly less frequent by a factor of 2 or more, with maximum values of about 2.4%. As indicated above flag 6 values (uniform profiles) are usually small (less than 2.3 %), but may be exceptionally large from 13 to 67 % for certain regions and gliders. All of the slocums that show large numbers of uniform profiles are those owned either by NRL (sl077, sl079, and sl083), and deployed during BIOSPACE_200806, or by Rutgers University (sl016), deployed during Hawaii_200807 (RIMPAC 2008), and the problem might be accounted for either by operational difficulties or known problems with internal data communications that occurred in the earlier slocum gliders.

For Seaglidors, generally, the percentage of flagged values is significantly smaller than is the case for Slocums. Flag 2 (range checks) occur less than 0.1% of the time. Flag 4 (single-point spikes) occur generally less than 0.5% of the time (exception ~2.0% for sg132 in Wash_200705). Flag 5 (standard deviation anomaly check) occurs typically less than 1% of the time. Finally, as noted above, Flag 6 (uniform profiles) is never significant for Seaglidors and Flag 7 (near surface check) occurs less than 2.0 % of the time, the latter being similar to Slocums.

It should be appreciated that the statistics reported here are quite strongly dependent upon the adoption of particular values for the key parameters employed in the QC algorithms. As a general rule, these parameter values are specified in the optics_variables_info.dat file (Appendix A). However, in some exceptional cases the values are directly coded in the present version, while in a future version all of the significant parameters will be accessible through the information file. To aid interpretation of the above statistics, the principles used in setting some of these parameters are given here. A key parameter value from which others are derived is the instrument resolution, which generally accords with the smallest increment in value which can be represented in the digital data record (This shows up in the MUG plots, as a distinct quantum jump in value or locally uniform values, in weak gradient areas). It would be possible, in principle, for this value to be customized for particular instruments or sensor units (via specific instrument calibration records), but the value adopted here, and in LAGER Optics 1.0, is that which is considered typical of the relevant instrument sensor type and wavelengths. From this quantity the threshold spike size, or 'spike', utilized in the test for flag 4, is specified as a multiplicative factor of 10 larger than the resolution. The 'abslimfactor' utilized in the test for flag 5, is specified as a multiplicative factor of 3 larger than the resolution. The 'stdevfactor' which is a non-dimensional quantity determining the number of local standard deviations that define a deviation anomaly for flag 5 is also specified in the information file. These parameters and others, act together with the logic of the various QC algorithms to determined the outcome

of the particular QC checks, and directly affect the statistics reported here. The implication is that after implementation of the latest version, the QC performance can be fine-tuned, as more experience is gained and larger historical data sets become available to facilitate performance improvements.

7 Validation Using Optical Variable Transects and Anomalous Profiles

The above sections explain the LAGER Optics V1.0 automated QC algorithms and statistics under different parameter settings. The process is adjustable using the parameter table described in `optics_variables_info.dat` under `/LAGER/SETUP_FILES/` directory. It can be seen that for the 7,067 profiles tested, statistics show that about 30.5% of the profiles will be sent to GOC operators for manual inspection, when a 15% threshold is set for the optical data points in question of each individual profile. The performance of the algorithm is further validated by visual inspection of 522 profiles from Slocum glider s1113 during Arabian Gauntlet 2009 Exercise (August 14-24, 2009). The area is shown in Figure 4, with the glider track near shore from northwest to southeast, marked in red.

Continuous profiles from deployments ranging from August 18 to August 19 are used. Figure 5 shows the time series of all optical profiles from backscattering 470nm during the entire exercise. The selection process was based on the quality of the optics sensor, the system performance reliability including data collection and transmission, as well as statistical analysis based on number of profiles in questions. The intentional bias is toward more troublesome profiles, to test the robustness of the algorithm. It is natural to see more profiles being sent to manual QC queue for this reason. Also, to make sure that no bad data slip through the process, a tighter than usual criteria setting is applied, which means once 10% of the data points in a profile is in question, due to various flag settings discussed above, this profile will be sent for manual inspection, as compared to default 15% value set. Out of 522 profiles inspected visually by the authors, 305 were sent to manual QC for inspection. 217 were sent to RTDHS directly. There are 2 profiles out these pass-thru that are mistakenly marked, which translate into 0.38% error.

The first one is from the profile number s1113_0818_1151_0000401, as shown in Figure 6. The questionable behavior is apparent as the sparseness of the data should present the profile for manual inspection. This is likely due to the complication from both surface check and lack of sparseness check. This can be improved in the LAGER Optics V2.0.

The second profile in question is shown in Figure 7. This is another profile that was not flagged to be sent to manual QC. Profile is from Slocum glider s1113 on August 18, 2009 with the profile number s1113_0818_1459_0000512. The questionable behavior can be observed for data in the middle of the water column where single point spikes were flagged. Such flags are within the criterion stated in previous sections. Further investigation is necessary and a fix will be in place for V2.0.

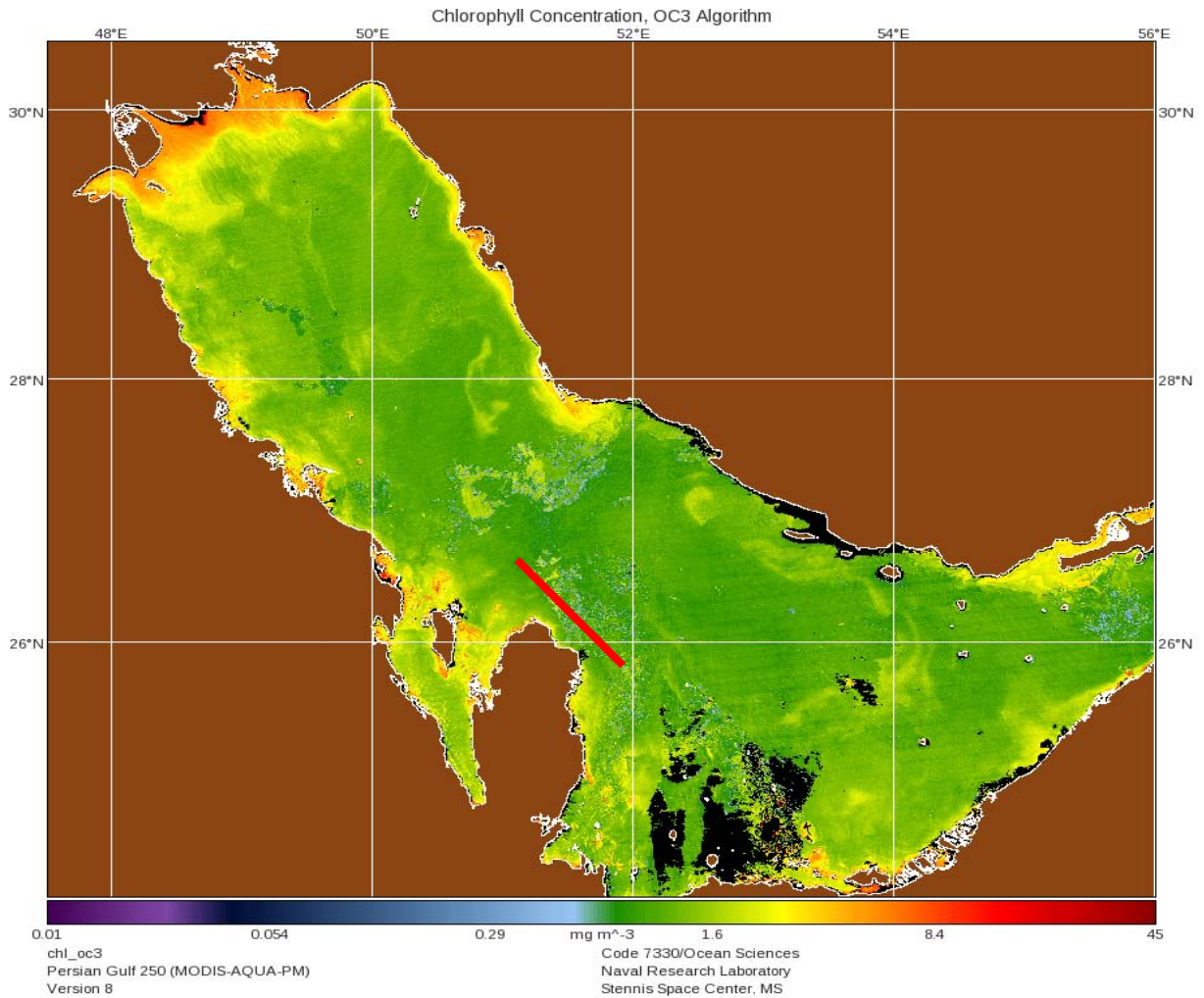


Figure 4. Arabian Gauntlet 2009 Exercise area shown using MODIS image derived chlorophyll product from NRL Ocean Sciences branch (7330). Red line depicts glider deployment tracks, from where validation test data were used.

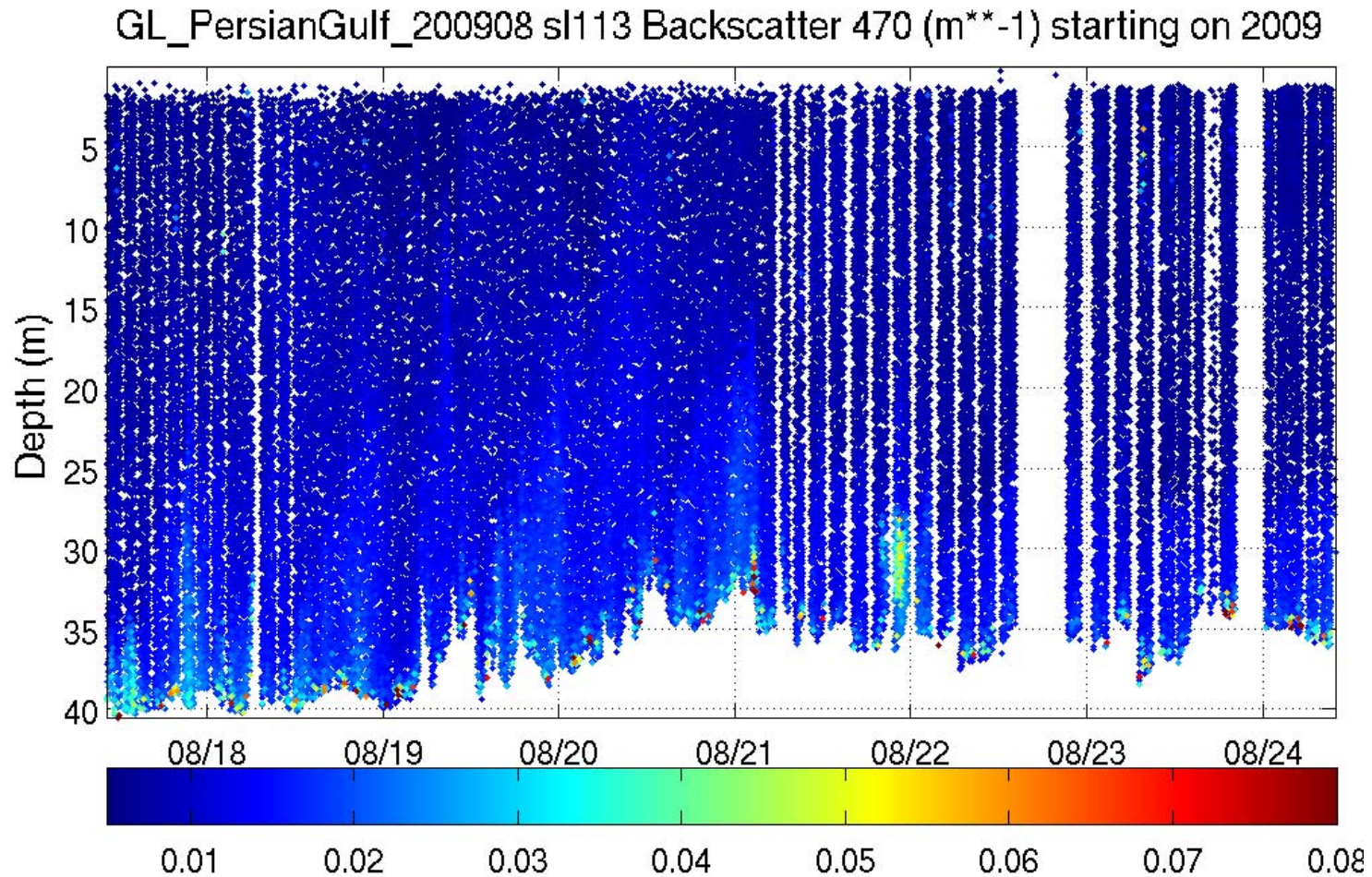


Figure 5. Time series of optical backscattering at 470nm (bb470) from slocum sl113 during Arabian Gauntlet 2009 Exercise. The validation test performed here uses data from August 18-19 period, which is a good representation of overall data structure, although it is biased towards more questionable profiles. It has less data gaps due to data transmission problems found during later stages of the exercise.

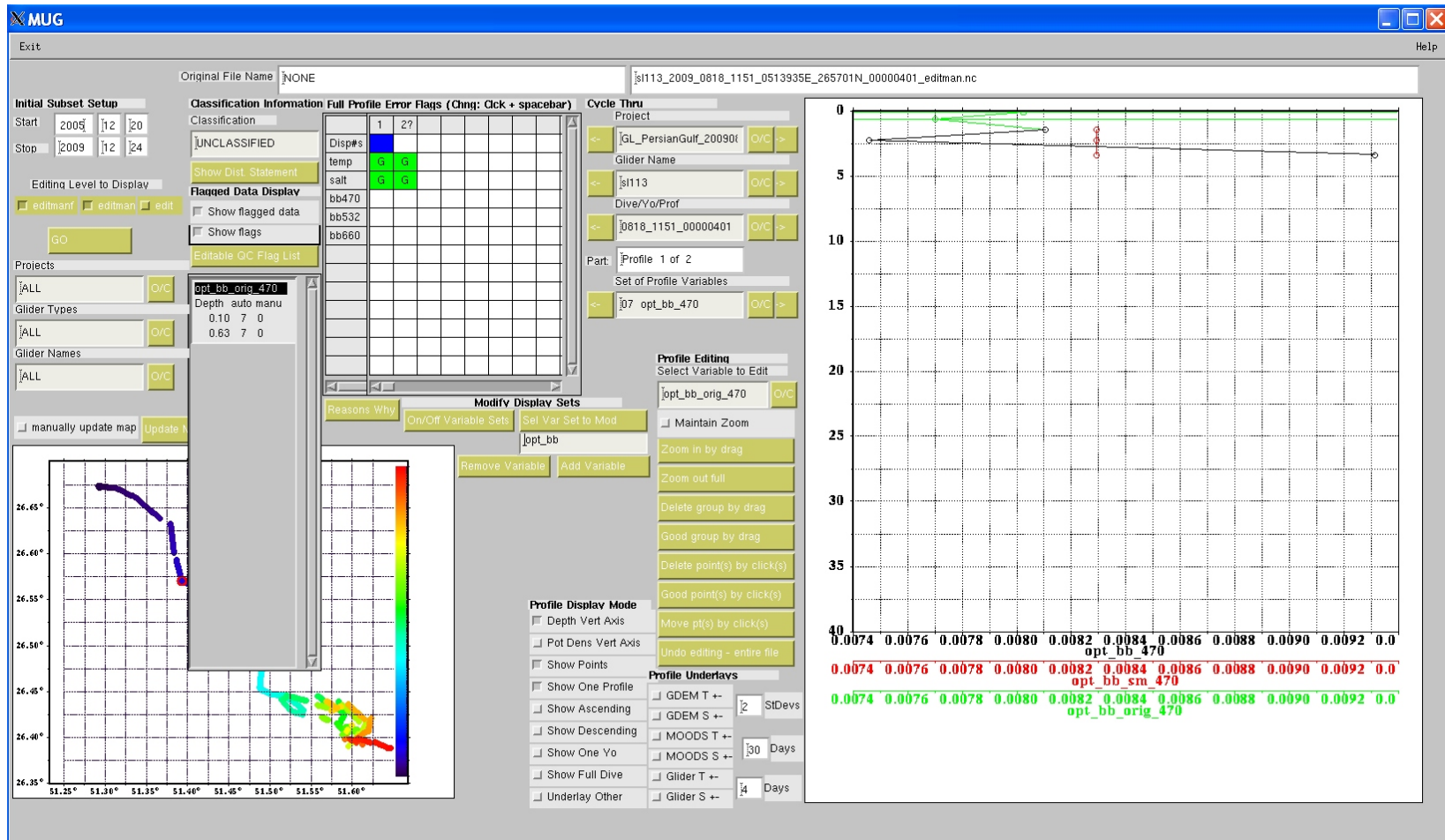


Figure 6. Sample profile that was not flagged to be sent to manual QC. Profile is from Slocum glider s1113 during Arabian Gauntlet 2009 on August 18, 2009. The profile number is s1113_0818_1151_0000401. The questionable behavior is apparent as the sparseness of the data should present the data for manual inspection. This is likely due to the complication from both surface check and lack of sparseness check. Notice individual flag values are shown both in drop-down box on left panel, as well as in the figure window on the right, marked by green lines.

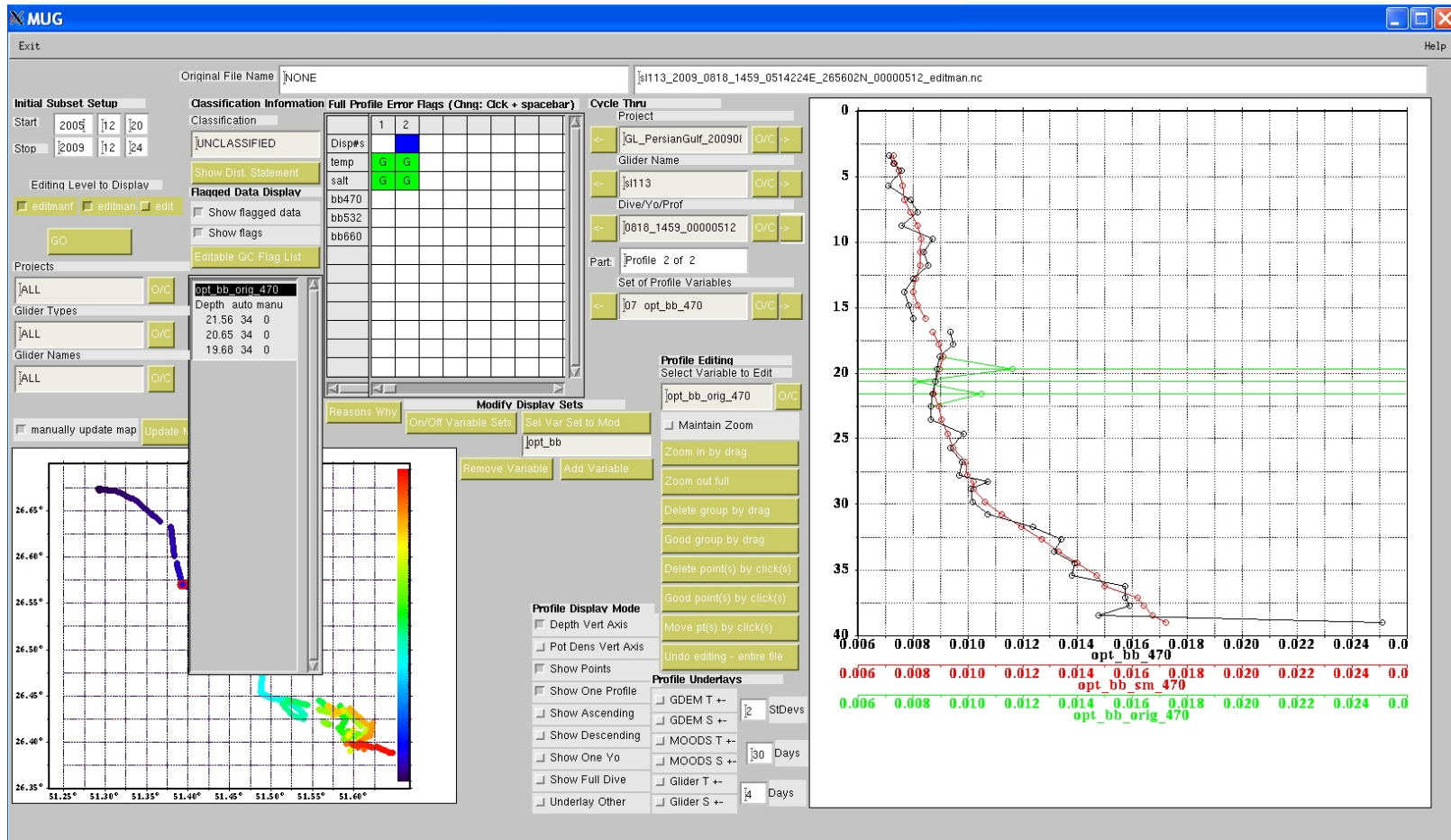


Figure 7. Sample profile that was not flagged to be sent to manual QC. Profile is from Slocum glider sl113 during Arabian Gauntlet 2009 on August 18, 2009. The profile number is sl113_0818_1459_0000512. The questionable behavior can be observed for data in the middle of the water column where single point spikes were flagged. Such flags are within the criterion stated in previous sections. Further investigation is necessary. Notice individual flag values are shown both in drop-down box on left panel, as well as in the figure window on the right, marked by green lines.

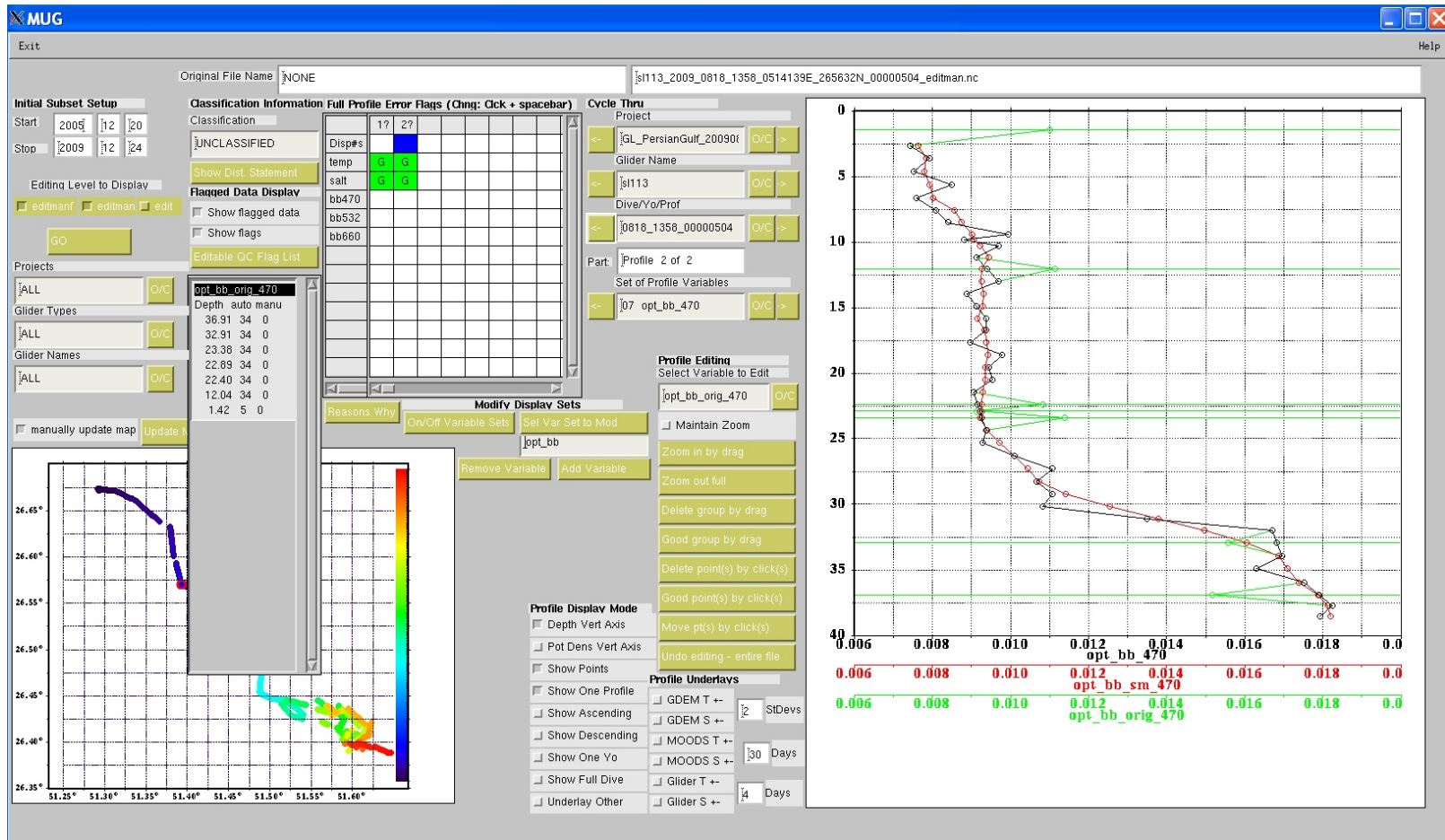


Figure 8. Sample profile that was not flagged to be sent to manual QC. Profile is from Slocum glider sl113 during Arabian Gauntlet Exercise on August 18, 2009. The profile number is sl113_0818_1358_0000504. This profile has been flagged to be manually examined, although it does not show any apparent errors, after auto QC process. This is likely due to the tight criteria currently in place for this validation testing.

Over all, more profiles were sent to manual QC than default values indicated, due to the selection of certain particular segments (more troublesome profiles). A majority of these profiles were sent due to tight spike check thresholds. One example can be found in Figure 8. This profile has been flagged to be manually examined, although it does not show any apparent errors, after auto QC process. This is likely due to the tight criteria currently in place which should be relaxed in order to lessen the burdens on the operators.

Conclusions

Through the testing of 7,067 NAVO collected glider optics data profiles from various Slocum gliders deployed over different regions and time, the results shown LAGER OPTICS V1.0 can automate QC process under various conditions. For the data we tested, results shown approximately 30.59% of the total profiles were sent to manual QC for further inspection, when the default value of 15% flag threshold is used. A tighter threshold will prompt more data to be inspected manually (eg 10% threshold would send about 40% data for manual inspection, based on current statistics of Slocum data). For validation tests, a default value of 10% of overall flags, instead of 15% was used, to ensure a tighter than usual quality control. It performs extremely well with optics data from slocum glider s1113 during Arabian Gauntlet 2009 Exercise, with under 1% (actual value 0.38%) of sampled profiles failed to be checked by the operator. It is worth noticing that this validation test was done using more questionable profile segments. It should be anticipated that the average performance should be better than shown here, both in terms of erroneous profiles going into RTDHS as well as numbers of profiles sent to MUG. It has been observed that optical data from Slocum gliders are inherently more noisy and error-prone for various reasons, including data storage and transmission, when compared to those from Seaglidern. Under similar conditions, for example, less than 1% of Seaglider data would be sent for manual inspection, even with a high, 10% threshold. (The exact figure using Table 5, would be 0.11% for a 15% threshold, and 0.81% for 10% threshold.)

References

1. Michael Carnes, LAGER Manual (Local Automated Glider Editing Routine) Version 1.0, 2008
2. Carnes, M. and P. Hogan, 2010, Validation Test Report for LAGER 1.0, NRL/MR/7330--10-9238, 68 pp.
3. Sam Mangin, U.S. Naval Oceanographic Office BUFR, December, 2009

Table 1. NRL/NAVO glider optical parameters for BUFR table.

Traditional optical variable names and units are used. Typical values listed are adapted from open ocean or clear water environments. Bandwidth parameters are from current or past instruments tested on Slocum gliders and Seagliders, with names denoted from raw data stream. The spectral information is not listed in this table and is subject to change depending on individual sensors onboard gliders. Multiple wavelengths of certain sensors (e.g. backscattering or bb) should be expected and have been implemented in the BUFR table by means of repetition (See Table 2).

Name	Description	Units	Typical value	Range	Resolution (step)	possible channel #s	Bandwidth (nm)	Instruments
bb	backscatter	m^{-1}	0.004	0~0.1	0.0001	<255	30	bb2f,bb3slo
Ed	Irradiance	wm^{-2}	10	0~500	0.1	<512	10	OCR504/7I (*)
c	beam attenuation	m^{-1}	0.2	0~100	0.01	<255	10	BAM (x)
b	Scattering	m^{-1}	0.1	0~100	0.01	<255	10	AUVb
vis	visibility	m	10	0.1~100	0.1	<255	-	AUVb,SAM, BAM
PAR	photo. avail. rad.	mEinstein. m^{-2}	100	0~2500	1	<255	-	ECO-PAR (x)
Fl-chl	[Chlorophyll-a]	$mg\ m^{-3}$	0.5	0~100	0.01	470/695	10	bb2f,fl3
Fl-phyco	phycoerythrin	ppb	0.05	0~100	0.005	540/570	10	fl3
Fl-cdom	color dissolved	ppb QS	2	0~2000	0.1	370/460	10/70	bb2f,fl3

Table 2. Glider optical parameters in BUFR format for LAGER Optics V1.0.

F	X	Y	Scale	Reference Value	Bit Width	Unit	Description
0	60	001	4	0	10	m-1	Backscatter
0	60	002	1	0	9	wm-2	Irradiance
0	60	003	2	0	14	m-1	Beam Attenuation
0	60	004	2	0	14	m-1	Scattering
0	60	005	1	0	10	m	Visibility
0	60	006	0	0	12	mEinstein.m-2	Photosynthetic Available Radiation
0	60	007	2	0	14	mg/m3	Chlorophyll-a
0	60	008	3	0	17	Ppb	Cyanobacteria
0	60	009	1	0	15	Ppb	Color Dissolved
0	60	010	0	0	10	Nm	Center Wavelength
0	60	011	0	0	7	Nm	Bandwidth
0	60	012	0	0	96	CCITT_IA5	Instrument

Table 3. BUFR Quality Control (Class 63)

F	X	Y	Scale	Reference Value	Bit Width	Unit	Description
0	63	001	0	0	6	Code Table	Automated QC Flags – Temperature
0	63	002	0	0	6	Code Table	Automated QC Flags – Salinity
0	63	003	0	0	4	Code Table	Manual QC Flags – Temperature
0	63	004	0	0	4	Code Table	Manual QC Flags – Salinity
0	63	005	0	0	4	Code Table	Full Profile QC Flag - Temperature
0	63	006	0	0	4	Code Table	Full Profile QC Flag - Salinity
0	63	007	0	0	6	Code Table	Automated QC Flags – Optics
0	63	008	0	0	4	Code Table	Manual QC Flags – Optics
0	63	009	0	0	4	Code Table	Full Profile QC Flag – Backscatter
0	63	010	0	0	4	Code Table	Full Profile QC Flag – Irradiance
0	63	011	0	0	4	Code Table	Full Profile QC Flag – Beam Attenuation
0	63	012	0	0	4	Code Table	Full Profile QC Flag – Scattering
0	63	013	0	0	4	Code Table	Full Profile QC Flag – Visibility
0	63	014	0	0	4	Code Table	Full Profile QC Flag – Photosynthetic Available Radiation
0	63	015	0	0	4	Code Table	Full Profile QC Flag – Chlorophyll-a
0	63	016	0	0	4	Code Table	Full Profile QC Flag – Cyanobacteria
0	63	017	0	0	4	Code Table	Full Profile QC Flag – Color Dissolved
0	63	018	0	0	4	Code Table	Full Profile QC Flag – Center Wavelength

Table 4. Statistical Results under Different Testing Conditions

Data from Slocum gliders are used to test the performance of current Glider Optics V1.0 auto QC algorithms. Notice that optical data from Slocum gliders typically suffer from more noises and data issues than those from Seagliders. This is the main reason for choosing Slocum. Detailed comparisons between optics from Slocum and Seagliders can be found from Table 5 and more details in text.

Glider	Date	Exercise	Number of Profiles	30% flagged	25% flagged	20% flagged	15% flagged	10% flagged	5% flagged
SI113	Aug 2009	Arabian Gauntlet	1038	0.289	0.624	2.70	12.1	40.4	83.8
SI005	June 2008	BioSpace	360	3.06	3.61	6.67	19.4	37.22	80.3
SI016	June 2008	BioSpace	168	0	0	0	0.595	0.595	0.595
SI027	June 2008	BioSpace	216	63.0	63.0	63.0	63.0	67.1	67.1
SI079	June 2008	BioSpace	1188	37.1	37.1	37.1	37.1	40.6	71.5
SI082	June 2008	BioSpace	1454	0.138	0.138	1.03	2.82	11.9	61.9
SI83	June 2008	BioSpace	408	60.3	60.3	60.3	60.3	60.3	80.6
SI016	July 2008	RIMPAC	2235	47.7	47.7	48.2	49.4	50.9	54.5
Total			7067	26.45	26.56	27.38	30.59	38.63	62.54

Table 5. Overall flag stats for all profiles tested

Area	Glider	NoFiles	Threshold (%)	Manual QC (%)	Individual Flag %				
					2	4	5	6	7
GL_PGulf_200908	sl113	1038	30	0.29					
GL_PGulf_200908	sl113	1038	25	0.67					
GL_PGulf_200908	sl113	1038	20	2.7					
GL_PGulf_200908	sl113	1038	15	12.14	0.273	5.28	1.451	0.013	0.26
GL_PGulf_200908	sl113	1038	10	40.37					
GL_PGulf_200908	sl113	1038	5	83.82					
GL_BIOSPACE_200806	sl005	168	30	0					
GL_BIOSPACE_200806	sl005	168	25	0					
GL_BIOSPACE_200806	sl005	168	20	0					
GL_BIOSPACE_200806	sl005	168	15	0.6	0.034	3.097	1.929	2.323	1.069
GL_BIOSPACE_200806	sl005	168	10	0.6					
GL_BIOSPACE_200806	sl005	168	5	19.64					
GL_BIOSPACE_200806	sl005	168	2	82.74					
GL_BIOSPACE_200806	sl016	360	30	3.06					
GL_BIOSPACE_200806	sl016	360	25	3.61					
GL_BIOSPACE_200806	sl016	360	20	6.67					
GL_BIOSPACE_200806	sl016	360	15	19.44	0.107	0.601	1.621	0	0.07
GL_BIOSPACE_200806	sl016	360	10	37.22					
GL_BIOSPACE_200806	sl016	360	5	80.28					
GL_BIOSPACE_200806	sl077	219	30	63.01					
GL_BIOSPACE_200806	sl077	219	25	63.01					
GL_BIOSPACE_200806	sl077	219	20	63.01					
GL_BIOSPACE_200806	sl077	219	15	63.01	0.037	1.061	0.978	24.404	0.627
GL_BIOSPACE_200806	sl077	219	10	67.12					
GL_BIOSPACE_200806	sl077	219	5	79					
GL_BIOSPACE_200806	sl079	1188	30	37.12					
GL_BIOSPACE_200806	sl079	1188	25	37.12					
GL_BIOSPACE_200806	sl079	1188	20	37.12					
GL_BIOSPACE_200806	sl079	1188	15	37.12	0.005	2.375	1.065	13.546	0.615
GL_BIOSPACE_200806	sl079	1188	10	40.57					
GL_BIOSPACE_200806	sl079	1188	5	71.55					
GL_BIOSPACE_200806	sl082	1454	30	0.14					
GL_BIOSPACE_200806	sl082	1454	25	0.14					
GL_BIOSPACE_200806	sl082	1454	20	1.03					

GL_BIOSPACE_200806	sl082	1454	15	2.82	0.011	2.154	2.387	0	0.732
GL_BIOSPACE_200806	sl082	1454	10	11.97					
GL_BIOSPACE_200806	sl082	1454	5	61.9					
GL_BIOSPACE_200806	sl083	408	30	60.29					
GL_BIOSPACE_200806	sl083	408	25	60.29					
GL_BIOSPACE_200806	sl083	408	20	60.29					
GL_BIOSPACE_200806	sl083	408	15	60.29	0.009	1.43	1.641	20.065	0.809
GL_BIOSPACE_200806	sl083	408	10	60.29					
GL_BIOSPACE_200806	sl083	408	5	80.64					
GL_BIOSPACE_200805	sl005	760	30	1.71					
GL_BIOSPACE_200805	sl005	760	25	1.97					
GL_BIOSPACE_200805	sl005	760	20	2.63					
GL_BIOSPACE_200805	sl005	760	15	6.97	0.06	3.829	2.132	0.619	0.642
GL_BIOSPACE_200805	sl005	760	10	33.95					
GL_BIOSPACE_200805	sl005	760	5	88.16					
GL_BIOSPACE_200805	sl016	828	30	0.12					
GL_BIOSPACE_200805	sl016	828	25	0.12					
GL_BIOSPACE_200805	sl016	828	20	0.72					
GL_BIOSPACE_200805	sl016	828	15	1.81	0.072	0.653	1.667	0.01	0.53
GL_BIOSPACE_200805	sl016	828	10	7.25					
GL_BIOSPACE_200805	sl016	828	5	32.97					
GL_BIOSPACE_200805	sl016	828	2	79.59					
GL_Hawaii_200807	sl016	2235	30	49.35					
GL_Hawaii_200807	sl016	2235	25	49.49					
GL_Hawaii_200807	sl016	2235	20	49.8					
GL_Hawaii_200807	sl016	2235	15	50.69	0.002	0.052	0.151	67.134	0.436
GL_Hawaii_200807	sl016	2235	10	52.04					
GL_Hawaii_200807	sl016	2235	5	56.69					
GL_Hawaii_200807	sg138	159	20	0					
GL_Hawaii_200807	sg138	159	15	0.63	0.008	0.17	0.282	0	1.236
GL_Hawaii_200807	sg138	159	10	3.77					
GL_Hawaii_200807	sg138	159	5	11.32					
GL_Hawaii_200807	sg138	159	2	30.19					
GL_Hawaii_200807	sg138	159	1	45.28					
GL_Hawaii_200807	sg137	140	20	0					
GL_Hawaii_200807	sg137	140	15	0.71	0.003	0.151	0.321	0	0.017
GL_Hawaii_200807	sg137	140	10	0.71					
GL_Hawaii_200807	sg137	140	5	1.43					
GL_Hawaii_200807	sg137	140	2	6.43					
GL_Hawaii_200807	sg137	140	1	16.43					

GL_Hawaii_200807	sg136	148	20	0					
GL_Hawaii_200807	sg136	148	15	0	0	0.068	0.284	0	0.266
GL_Hawaii_200807	sg136	148	10	0					
GL_Hawaii_200807	sg136	148	5	0.676					
GL_Hawaii_200807	sg136	148	2	2.703					
GL_Hawaii_200807	sg136	148	1	16.22					
GL_Pacom_200712	sg135	31	20	0					
GL_Pacom_200712	sg135	31	15	0	0	0.045	0.053	0	0.873
GL_Pacom_200712	sg135	31	10	0					
GL_Pacom_200712	sg135	31	5	0					
GL_Pacom_200712	sg135	31	2	7.407					
GL_Pacom_200712	sg135	31	1	7.407					
GL_Pacom_200712	sg136	31	20	0					
GL_Pacom_200712	sg136	31	15	0	0	0.039	0.093	0	0.116
GL_Pacom_200712	sg136	31	10	0					
GL_Pacom_200712	sg136	31	5	0					
GL_Pacom_200712	sg136	31	2	0					
GL_Pacom_200712	sg136	31	1	6.452					
GL_RYUK_200807	sg137	62	20	0					
GL_RYUK_200807	sg137	62	15	0	0	0.11	0.573	0	0.003
GL_RYUK_200807	sg137	62	10	1.613					
GL_RYUK_200807	sg137	62	5	1.613					
GL_RYUK_200807	sg137	62	2	4.839					
GL_RYUK_200807	sg138	27	20	0					
GL_RYUK_200807	sg138	27	15	0	0	0.093	0.375	0	0.294
GL_RYUK_200807	sg138	27	10	0					
GL_RYUK_200807	sg138	27	5	0					
GL_RYUK_200807	sg138	27	2	0					
GL_RYUK_200807	sg138	27	1	18.519					
GL_Straights_200905	sg132	94	15	0	0	0.149	0.844	0	0.135
GL_Straights_200905	sg132	94	10	0					
GL_Straights_200905	sg132	94	5	0					
GL_Straights_200905	sg132	94	2	4.255					
GL_Straights_200905	sg135	29	15	0	0	0.012	0.149	0	0.629
GL_Straights_200905	sg135	29	10	0					
GL_Straights_200905	sg135	29	5	0					
GL_Straights_200905	sg135	29	2	0					
GL_Straights_200905	sg135	29	1	0					

GL_Norway_200908	sg136	72	15	0	0.001	0.133	0.506	0	0.542
GL_Norway_200908	sg136	72	10	0					
GL_Norway_200908	sg136	72	5	2.778					
GL_Norway_200908	sg136	72	2	4.167					
GL_Norway_200909	sg136	376	15	0	0.01	0.32	0.645	0	0.415
GL_Norway_200909	sg136	376	10	0					
GL_Norway_200909	sg136	376	5	0					
GL_Norway_200909	sg136	376	2	10.106					
GL_Phillipin_200901	sg131	162	15	0	0	0.049	0.623	0	0.26
GL_Phillipin_200901	sg131	162	10	0					
GL_Phillipin_200901	sg131	162	5	0					
GL_Phillipin_200901	sg131	162	2	30.864					
GL_Phillipin_200901	sg133	142	15	0	0	0.061	0.317	0	0.281
GL_Phillipin_200901	sg133	142	10	0					
GL_Phillipin_200901	sg133	142	5	0					
GL_Phillipin_200901	sg133	142	2	1.409					
GL_Phillipin_200901	sg138	111	15	0	0	0.101	0.132	0	1.515
GL_Phillipin_200901	sg138	111	10	0					
GL_Phillipin_200901	sg138	111	5	0					
GL_Phillipin_200901	sg138	111	2	3.604					
GL_ETai_200911	sg163	112	15	0	0.049	0.05	0.268	0	0
GL_ETai_200911	sg163	112	10	0					
GL_ETai_200911	sg163	112	5	0.893					
GL_ETai_200911	sg163	112	2	4.464					
GL_ETai_200911	sg159	77	15	0	0.097	0.167	0.517	0	0
GL_ETai_200911	sg159	77	10	0					
GL_ETai_200911	sg159	77	5	0					
GL_ETai_200911	sg159	77	2	0					
GL_Beaufort_200908	sg131	149	20	0					
GL_Beaufort_200908	sg131	149	15	0.671	0.09	0.535	0.983	0	0.657
GL_Beaufort_200908	sg131	149	10	2.013					
GL_Beaufort_200908	sg131	149	5	14.094					
GL_Beaufort_200909	sg131	95	20	0					
GL_Beaufort_200909	sg131	95	15	0	0.005	0.306	1.05	0	0.727
GL_Beaufort_200909	sg131	95	10	0					
GL_Beaufort_200909	sg131	95	5	4.2105					
GL_Beaufort_200909	sg131	95	2	50.526					

GL_Wash_200705	sg132	55	15	0	0.006	2.089	0.734	0	2.117
GL_Wash_200705	sg132	55	10	7.273					
GL_Wash_200705	sg132	55	5	34.546					
GL_Wash_200705	sg132	55	2	78.182					

Appendix A

optics_variables_info.dat

This file is the parameter file for most of the automated optics QC modules, as shown in Figure 2. This file resides in `SETUP_FILES` subdirectory under `LAGER`. Modifications of parameters in this file will directly affect the program performance. It should be handled with caution, and knowledge of specific optical sensors and its calibration procedures.

```
#
# File Format Description
#
# General
# All lines must start in column 1.
#
# Lines starting with asterisk (*) are comments
# Comment lines can occur anywhere within the file and are not
# required to occur anywhere within the file. For readability, a
# comment line should be placed between each block of lines defining
# a variable or variable group.
#
# Lines with a variable definition block can be in any order.
#
# If a block of lines following the line: groupsize N, where N
# is greater than 1, must have N variable, type, name, units,
# and validrange lines. In addition, if there is at least one
lines.....
#
# Description of each line type
#
# Line Type      Description
# 1. nvarinput   Number of variables input as processing group.
#                Example: groupinsize 1
#                groupsize followed by integer indicating
#                the number of variable required during processing.
#                Each new group sequence MUST begin with a groupinsize
line.
#
# 2. nvaroutput  Number of variables output from processing group.
#                Example: groupoutsize 2
#                If missing, then groupoutsize 1 assumed.
# 3. inname      input variable name
#                Example:
#                If nvarinput > 1, then each variable name in group must
#                be listed on separate lines.
#
# 4. calibrated  Variable is already calibrated.
#                Examples:
#                calibrated yes
#                calibrated no
#
# 5. type        Generic variable type, e.g., bb, ed, FL-chl
```

```

#
# 6. units          Units of calibrated variable.
#
# 7. outname       Name of calibrated variable.
#
# 8. validrangemin Valid range minimum of calibrated variable
#
# 9. validrangemax Valid range maximum of calibrated variable
#
# 9. wavelength
#
# 10. bandwidth
#
# 11. scalefactor
#
# 12. enddef       Each group sequence MUST end with an enddef line.
#
nvarinput = 1
inname = wlbb2f_blueCount
calibrated = no
instrument = bb2f
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0.0
validrangemax = 0.1
resolution = 0.0001
wavelength = 470
bandwidth = 30
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = wlbb2f_redCount
calibrated = no
instrument = bb2f
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0.0
validrangemax = 0.1
resolution = 0.0001
wavelength = 700
bandwidth = 30
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = wlbb2f_fluorCount

```

```

calibrated = no
instrument = bb2f
description = chlorophyll-a concentration from fluorescence
type = Flchl
units = mg/m**3
outname = chlorophyll-a
validrangemin = 0
validrangemax = 100
resolution = 0.01
wavelength = 470
scalefactor = 1
spike = 0.2
stdevfactor = 2.0
abslimfactor = 0.03
enddef
#
nvarinput = 1
inname = sci_bb3slo_b470_scaled
calibrated = yes
instrument = bb3slo
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0
validrangemax = 0.1
resolution = 0.0001
wavelength = 470
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = sci_bb3slo_b532_scaled
calibrated = yes
instrument = bb3slo
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0
validrangemax = 0.1
resolution = 0.0001
wavelength = 532
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = sci_bb3slo_b660_scaled
calibrated = yes
instrument = bb3slo
description = backscattering coefficient

```

```

type = bb
units = 1/m
outname = backscatter
validrangemin = 0
validrangemax = 0.1
resolution = 0.0001
wavelength = 660
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = sci_ocr504I_irrad1
calibrated = yes
instrument = ocr504I
description = downward irradiance
type = ed
units = w/m**2
outname = irradiance
validrangemin = 0
validrangemax = 500
resolution = 0.1
wavelength = 0
bandwidth = 10
scalefactor = 1.0
stdevfactor = 3.0
abslimfactor = 0.3
spike = 50
enddef
#
nvarinput = 1
inname = sci_ocr504I_irrad2
calibrated = yes
instrument = ocr504I
description = downward irradiance
type = ed
units = w/m**2
outname = irradiance
validrangemin = 0
validrangemax = 500
resolution = 0.1
wavelength = 0
bandwidth = 10
scalefactor = 1.0
stdevfactor = 3.0
abslimfactor = 0.3
spike = 50
enddef
#
nvarinput = 1
inname = sci_ocr504I_irrad3
calibrated = yes
instrument = ocr504I
description = downward irradiance
type = ed

```

```

units = w/m**2
outname = irradiance
validrangemin = 0
validrangemax = 500
resolution = 0.1
wavelength = 0
bandwidth = 10
scalefactor = 1.0
stdevfactor = 3.0
abslimfactor = 0.3
spike = 50
enddef
#
nvarinput = 1
inname = sci_ocr504I_irrad4
calibrated = yes
instrument = ocr504I
description = downward irradiance
type = ed
units = w/m**2
outname = irradiance
validrangemin = 0
validrangemax = 500
resolution = 0.1
wavelength = 0
bandwidth = 10
scalefactor = 1.0
stdevfactor = 3.0
abslimfactor = 0.3
spike = 50
enddef
#
nvarinput = 1
inname = sci_bbfl2s_bb_sig
calibrated = no
instrument = bbfl2s
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0
validrangemax = 0.0001
resolution = 0.0001
wavelength = 0
bandwidth = 30
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 0.3
enddef
#
nvarinput = 1
inname = sci_bbfl2s_chlor_sig
calibrated = no
instrument = bbfl2s
description = chlorophyll-a concentration from fluorescence
type = Flchl

```

```

units = mg/m**3
outname = chlorophyll-a
validrangemin = 0
validrangemax = 100
resolution = 0.01
wavelength = 0
bandwidth = 10
scalefactor = 1
spike = 0.1
stdevfactor = 2.0
abslimfactor = 0.03
enddef
#
nvarinput = 1
inname = sci_bbfl2s_cdom_sig
calibrated = no
instrument = bbfl2s
description = color dissolved organic matter fluorescence
type = Flcdom
units = ppb QS
outname = color dissolved
validrangemin = 0
validrangemax = 2000
resolution = 0.1
wavelength = 0
bandwidth = 10
scalefactor = 1
spike = 5.0
stdevfactor = 2.0
abslimfactor = 0.3
enddef
#
nvarinput = 1
inname = sci_bbfl2s_bb_scaled
calibrated = yes
instrument = bbfl2s
description = backscattering coefficient
type = bb
units = 1/m
outname = backscatter
validrangemin = 0
validrangemax = 0.1
resolution = 0.0001
wavelength = 661
bandwidth = 30
scalefactor = 1
spike = 1.e-3
stdevfactor = 2.0
abslimfactor = 3.e-4
enddef
#
nvarinput = 1
inname = sci_bbfl2s_chlor_scaled
calibrated = yes
instrument = bbfl2s
description = chlorophyll-a concentration from fluorescence
type = Flchl

```

```

units = mg/m**3
outname = chlorophyll-a
validrangemin = 0
validrangemax = 100
bandwidth = 30
wavelength = 0
resolution = 0.01
scalefactor = 1
spike = 0.1
stdevfactor = 2.0
abslimfactor = 0.03
enddef
#
nvarinput = 1
inname = sci_bbfl2s_cdom_scaled
calibrated = yes
instrument = bbfl2s
description = color dissolved organic matter fluorescence
type = Flcdom
units = ppb QS
outname = color dissolved
validrangemin = 0
validrangemax = 2000
bandwidth = 10
resolution = 0.1
wavelength = 0
scalefactor = 1
spike = 1.0
stdevfactor = 2.0
abslimfactor = 0.3
enddef
#
nvarinput = 2
nvaroutput = 3
inname = sci_auvb_ref
inname = sci_auvb_sig
calibrated = no
instrument = auvb
instrument = auvb
instrument = auvb
description = derived diver visibility range
description = beam attenuation coefficient
description = total scattering coefficient
type = vis
type = c
type = b
units = m
units = 1/m
units = 1/m
outname = visibility
outname = attenuation
outname = scattering
scalefactor = 1
scalefactor = 1
scalefactor = 1
wavelength = 0
wavelength = 0

```

```

wavelength = 0
resolution = 0.01
resolution = 0.01
resolution = 0.01
validrangemin = 0.1
validrangemax = 100
validrangemin = 0
validrangemax = 100
validrangemin = 0
validrangemax = 100
bandwidth = 10
bandwidth = 10
bandwidth = 10
spike = 1.0
spike = 1.0
spike = 1.0
stdevfactor = 2.0
stdevfactor = 2.0
stdevfactor = 2.0
abslimfactor = 0.03
abslimfactor = 0.03
abslimfactor = 0.03
enddef
#
nvarinput = 1
iname = sci_fl3slo_chlor_units
calibrated = yes
instrument = fl3slo
description = chlorophyll-a concentration from fluorescence
type = Flchl
units = mg/m**3
outname = chlorophyll-a
scalefactor = 1
resolution = 0.01
wavelength = 0
bandwidth = 30
validrangemin = 0
validrangemax = 100
spike = 0.1
stdevfactor = 2.0
abslimfactor = 0.03
enddef
#
nvarinput = 1
iname = sci_fl3slo_phyco_units
calibrated = yes
instrument = fl3slo
description = phycoerytherin fluorescence
type = Flphyco
units = ppb
scalefactor = 1
outname = cyanobacteria
resolution = 0.03
wavelength = 0
bandwidth = 10
validrangemin = 0
validrangemax = 2000

```



```
spike = 0.3
stdevfactor = 2.0
abslimfactor = 0.09
enddef
#
nvarinput = 1
inname = sci_fl3slo_cdom_units
calibrated = yes
instrument = fl3slo
description = color dissolved organic matter fluorescence
type = Flcdom
units = ppb QS
outname = color dissolved
scalefactor = 1
resolution = 0.09
wavelength = 0
bandwidth = 10
validrangemin = 0
validrangemax = 100
spike = 0.9
stdevfactor = 2.0
abslimfactor = 0.27
enddef
```