Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE
10.1029/2018GC007891

The Seagoing Scientist's Toolbox: Integrated Methods for Quality Control of Marine Geophysical Data at Sea

Michael Hamilton[1,3], Paul Wessel[1], Joaquim Luis[2], Brian Taylor[1], and Youngtak Ko[3]

[1]Department of Earth Sciences, University of Hawai'i at Mānoa, Honolulu, HI, USA, [3]Deep-Sea and Seabed Mineral Resources Research Center, Korea Institute of Ocean Science and Technology, Busan, South Korea, [2]Departamento de Ciências da Terra, do Mar e do Ambiente, Universidade do Algarve, Faro, Portugal

Key Points:
- We propose streamlined handling of underway geophysics and sonar data for seagoing scientists and the academic fleet
- The approach posits the archive file upfront which is iteratively generated and inspected as new data come in
- Incoming underway geophysics and sonar data are preprocessed and monitored in Google Earth

Correspondence to:
M. Hamilton,
mith@hawaii.edu

Citation:
Hamilton, M., Wessel, P., Luis, J., Taylor, B., & Ko, Y. (2019). The seagoing scientist's toolbox: Integrated methods for quality control of marine geophysical data at sea. *Geochemistry, Geophysics, Geosystems*, *20*, 5415–5424. https://doi.org/10.1029/2018GC007891

Received 23 AUG 2018
Accepted 15 OCT 2018
Accepted article online 01 OCT 2019
Published online 29 NOV 2019

Abstract We announce a new and integrated system for planning and executing marine geophysical surveys and for scrutinizing and visualizing incoming shipboard data. The system incorporates free software designed for use by scientists and shipboard operators and pertains to underway geophysics and multibeam sonar surveys. Regarding underway data, a crucial first step in the approach is to reduce and merge incoming center beam depth, gravity, and towed magnetic data with navigation, then reformat to the standard exchange format. We are then able to apply established quality control methods including along-track and cross-track analyses to identify error sources and to incrementally build the candidate archive file as new data are acquired. Regarding multibeam data, these are subjected to both an automated error removal scheme for quick visualization and to subsequent ping editing in detail. The candidate archive file and sonar data are automatically and periodically updated and adapted for display in Google Earth, wherein survey planning is also carried out. Data layers are also updated automatically in Google Earth, allowing scientists to focus on visual inspection and interpretation of incoming data. By visualizing underway and sonar data together with reference gravity, magnetic, and bathymetry grids in Google Earth, data familiarity is enhanced and the likelihood of noticing extreme errors increased. We hope scientists will embrace these techniques so that each data set being submitted to a data repository is vetted by the seagoing science party.

1. Introduction

The flow of geophysical data from the acquisition of a value by a measuring device until it reaches its final destination (hopefully) at a publicly accessible data repository concentrates around the scientist(s) who devises and conducts the expedition and who is responsible for data evaluation and interpretation. Seagoing scientists are intimately involved with the collection of new data and are best suited to scrutinize newly acquired data. This responsibility has long been a primary function of scientists. However, in recent times, research vessels simultaneously collect oceanographic, meteorologic, and geophysical data regardless of onboard personnel. Science parties tailored to fulfill primary objectives are often unable to vet all forms of incoming data. Evaluation of remaining data is typically being omitted, is performed by seagoing contractors, or is performed after-the-fact by data users, hydrographic offices, and other parties. So much of the seafloor has yet to be mapped (Wessel & Chandler, 2011) that even unvetted marine geophysical data are essential to many investigations. Ensuring quality is therefore of paramount importance.

Of significance to those attempting to make use of the data sets collected to date is the fact that many surveys archived at the world's largest repository for underway geophysical data, which is maintained by the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI, formerly National Geophysical Data Center), were not subjected to scientific scrutiny. Traditionally, NCEI performed little in the way of quality control and the National Science Foundation does not pay for routine processing of underway data. Typically, time and resources are directed to ensure the quality of underway data only when an onboard scientist has an interest in the data. Insufficient data visualization and quality control during acquisition and archival preparation has led to extreme errors being promulgated to the scientific community for decades. Three examples of this tendency are shown in Figure 1. For instance, Figure 1a illustrates a common operator error (i.e., loss of sonar bottom tracking) that is best dealt with while at sea and that in many cases would not be rectifiable after-the-fact. In this case, single-beam and multibeam sonars acquired data simultaneously, according to the survey's metadata header; hence, the source

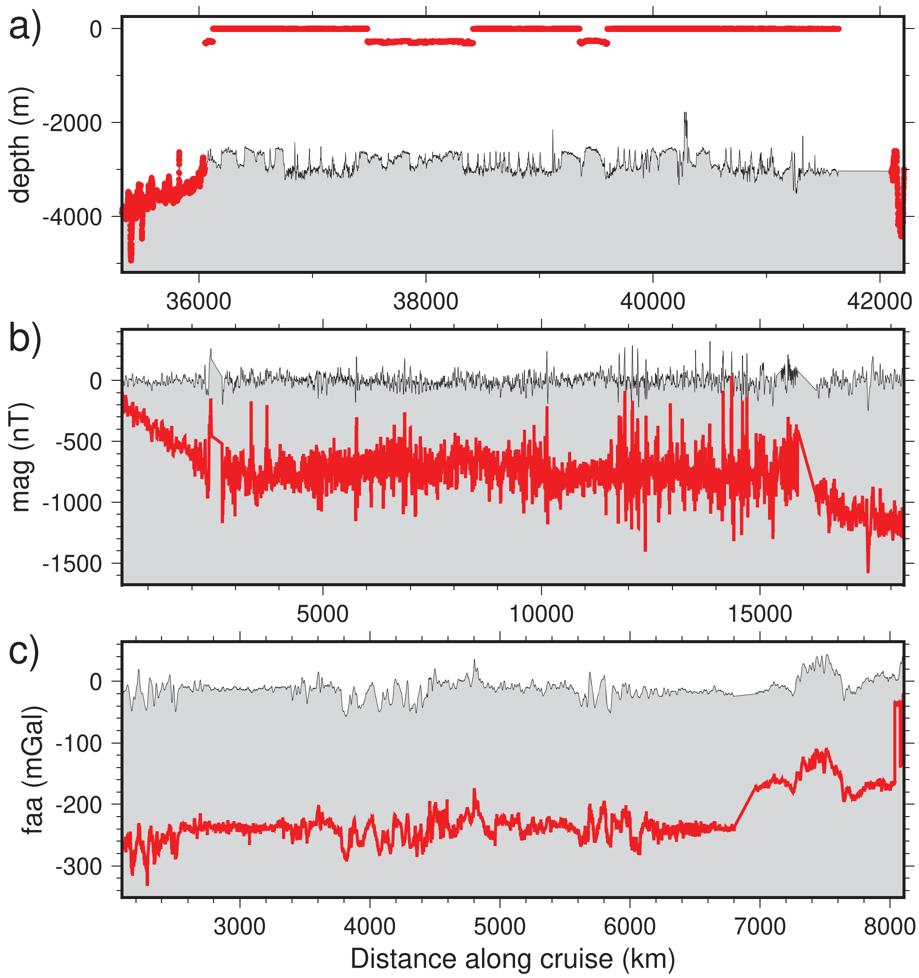

Figure 1. Examples of extreme errors in National Centers for Environmental Information's track line geophysics archive: (a) Lamont Leg C2607 (National Geophysical Data Center id 01010253) acquired single-beam and multibeam data along the northern East Pacific Rise in 1985. Despite two sonars running simultaneously, ~6,000 km of spurious depths (red) were archived. In contrast, global predicted bathymetry (Becker et al., 2009), sampled along the ship track and shown in gray, exhibits no such error. (b) Marine magnetic anomalies (red), from Lamont Leg EW9210 (01210049) to the northwestern Atlantic in 1992, have a mean value of ~ −750.4 nT, differ substantially from the global magnetic anomaly compilation of Maus et al. (2009; gray) and are correctable using the latest reference fields. (c) Free air gravity anomalies (red), acquired southeast of New Zealand by the National Science Foundation Leg NBP9604 (88010008) in 1996, show nonconstant offsets from altimetry-derived gravity (gray; Sandwell et al., 2014), requiring careful correction and inspection prior to utilization. Panel (c) adapted from Chandler and Wessel (2008).

institution could fix this by replacing erroneous values. Figure 1b shows magnetic anomalies that were possibly computed using an outdated/inaccurate magnetic reference field. Anomalies recomputed using the Definitive Geomagnetic Reference Field (not shown here) are dramatically improved from a mean value of −750.4 to 3.1 nT. According to Chandler and Wessel (2012), outdated magnetic anomalies is the most pervasive error in the NCEI track line archive, affecting at least 89% of magnetic surveys. The remaining 11% of magnetic surveys are not recomputable because these archives lack total field values. Although some were reduced using nearby observatories, ideally these surveys will be resubmitted along with total field values but they can likely be utilized, as is, given further isolation of the crustal field by each data user and for each affected survey. Not shown are a host of other data quality issues such as data reported in the wrong units (e.g., fathoms instead of meters), data inappropriately scaled by factors of 10 by archivists, unreconciled gravimeter drift and vessel motion artifacts, a systemic tendency to omit observed gravity from the majority of archive files (62% of gravity legs lack observed gravity measurements), ~7,500-m depth offsets stemming from erroneous digitization of two-way travel time echograms, inaccurate navigation especially among the

10.1029/2018GC007891

many older data sets, excessive speed due to temporary loss of satellite or radio navigation signals, navigation passing over land, and the presence of outliers and unrealistic data gradients (Chandler & Wessel, 2008, 2012; Smith, 1993; Wessel & Watts, 1988). These and a plethora of other data problems warranting our diligent supervision (e.g., hardware malfunctions, operator errors, and software crashes) continue to occur at sea.

Wessel and Watts (1988) analyzed discrepancies at gravity track intersections globally, via crossover error (COE) analysis, and were able to reduce COE standard deviations from 22.43 to 13.96 mGal by modeling systematic drift and offsets. Their study, which included 834 gravity cruises available at the time, found that instrument drift, DC shifts, poor navigation, and inadequate vessel motion corrections were the major sources of uncertainty. Smith (1993) also used COE analysis to examine the accuracy of depth soundings for 2,253 surveys and reported major error sources including extreme errors in two-way travel time digitization, inconsistent sound speed velocities, and a multitude of inappropriately scaled data sets and surveys reported in the wrong units. Quesnel et al. (2009) examined 2,411 magnetic surveys and by recomputing anomalies using comprehensive magnetic modeling and correcting for diurnal variation (which was not applied in most archived data sets), and by reducing intercruise discrepancies using COE analysis, substantially reduced magnetic field intensity COE from 179.6 to 35.9 nT. At the time of this writing, however, NCEI archives 5,249 depth, 2,965 magnetic, and 2,092 gravity tracks (Hamilton et al., 2019). COE corrections apply within the closed set of included tracks; hence, the corrections mentioned above are now outdated.

Chandler and Wessel (2008) examined 4,980 marine geophysical archive files using along-track analysis, which is different from COE analysis in that errors are found internally within a single cruise by comparing shipboard data to global grids and to anomalies recomputed using the latest reference fields and gravity formulae, among many other tests. The study found extreme data errors such as data reported in the wrong units or scaled by factors of 10, navigation problems such as excessive speed and navigation-over-land, excessive truncation of precision, and untied gravity anomalies. Whereas the Chandler and Wessel (2008) study was largely automated and geared for exploratory data analysis and calibration of the along-track methods, Chandler and Wessel (2012) analyzed 5,230 archive files along track in a time-intensive, cruise-by-cruise basis and generated errata correction tables for each survey, which correct the most extreme errors. That study reported error reductions comparable to the COE studies mentioned above while leaving the majority of data untouched. The study also reported that 69% of gravity archive files lack raw gravity data and that 89% of magnetic anomalies are outdated and need to be recalculated. The remaining magnetic surveys failed to report total field data; hence, these anomalies cannot be recalculated. The along-track analysis and errata-based correction approach, although time-consuming, effectively identifies and corrects major errors without impacting the majority of measurements. Furthermore, errata tables do not lose relevance through time except in cases where a new errata table must be generated and reviewed due to a new version of an existing archive being released.

Within the American fleet, an academic consortium known as Rolling Deck to Repository (R2R) was established to rectify many data quality problems, to standardize a wide variety of incoming data types and to streamline submission to repositories (Miller et al., 2010). R2R plays an important role in the current framework where at-sea data handling is sometimes inadequate or nonexistent. A lack of geophysicists and standard data handling practices on many expeditions along with a lack of in-house data archival expertise has necessitated postcruise processing by a third party. The task that R2R faces is challenging, though, in that they are not a large group yet are required to vet and prepare vast and widely varied data from the entire American fleet. Marine research is quite costly, and close inspection is required of all incoming data; hence, we argue that more data scrutiny needs to be focused on the acquisition front. The odds of detecting and correcting extreme errors are much more in our favor at sea when data reviewers are most available.

Recently, Seabed 2030, a new and international global mapping effort, has been proposed by the General Bathymetric Chart of the Oceans group with plans to carry out systematic bathymetric mapping for all depths exceeding 200 m (Mayer et al., 2018). They estimate costs on the order of USD $3 billion and 350 ship years to complete a full coverage global bathymetric map. The Seabed 2030 project has no requirement for the collection of gravity or magnetics data. Prior to that, the Global Ocean Mapping Program (GOMaP) project was put forth by Carron et al. (2001) of the U.S. Naval Oceanographic Office and Naval Research Laboratory. The GOMaP project proposed full bathymetric and sidescan coverage at 100 m or better resolution and to concurrently acquire magnetics, gravity, subbottom profiler, and oceanographic and meteorologic data, for all depths exceeding 500 m. Although GOMaP estimated that the deep oceans could be fully mapped within ∼200 ship years, which could be carried out in a 20–30 year time frame given sufficient international support at an estimated cost of USD $8 to $16 billion, the project lacked needed support. While ambitious, these proposals indicate growing international support for a transition from focused to systematic surveying and will require data handling infrastructure as well as considerable financial backing for successful completion.

We posit that a better solution for handling underway gravity, towed magnetic, and bathymetric data is outlined herein and centers around creating the archive file first, while at sea. This step is essential because the archive file is required to be formatted in the standard data exchange format (currently MGD77T) suitable for redistribution, and because such a standard format file can be scrutinized using publicly available data quality control tools by shipboard technicians and the seagoing science party. We hope to convey that archive files should be created first under the close supervision of seagoing geophysicists and/or marine technicians, while still at sea, and subjected to wide-ranging scrutiny and constant monitoring. If processed data sets are to be passed to data repositories, rather than unprocessed data being sent from marine centers, then perhaps infrastructure (e.g., this proposal) should be put in place to facilitate this exchange and to lower the data scrutiny threshold so that quality control can be accomplished even when no geophysicists are onboard. To facilitate this important goal, and to help relieve processing redundancy caused by the promulgation of unprocessed data sets, here we describe our methodology for testing and visualizing data, while at sea, using freely available tools.

2. Methods

Many data handling procedures have been put into practice in the past. Here, we present our approach that utilizes currently available tools under active development that are distributed freely and that, if utilized, may greatly increase processing efficiency at sea. We have applied the skills developed over the course of our many days at sea and many data sets processed into a set of steps that performs these tasks in a timely fashion. If this approach is adopted by the community, after the initial period of assessing each vessel's different data structure issues, conversion into standard data exchange formats such as MGD77T (Hittleman et al., 2010), the Generic Mapping Tools (Wessel et al., 2019) mgd77 table format (herein referred to as GMT DAT), or the NetCDF compliant MGD77+ format (Wessel & Chandler, 2007), enables thorough data quality assessment and control using along-track (Chandler & Wessel, 2008, 2012) and crossover analysis. Crossover analysis is supported in that the mgd77 formats are supported by GMT's crossover analysis toolkit, x2sys (Wessel, 2010), for cases in which sufficient ship track intersections occur.

At sea, it is an iterative process by which the raw data are reformatted into the standard exchange format and a metadata header is compiled (Hamilton et al., 2019), and errors are detected along track using the GMT *mgd77sniffer* program, which outputs an errata table. When reviewing errata tables compiled from existing data sets, one typically reads the errata table with the aid of an errata review plot (e.g., Figure 2; Chandler & Wessel, 2012) and chooses which error flags to apply. However, at sea, when generating new data sets, once this table is free of systematic errors, and once the data reduction and reformatting steps are functional and any apparent errors are corrected, it is generally sufficient to let the along-track analysis step run automatically and instead examine the data visually. It is helpful, for example, to compare shipboard gravity to the latest satellite-derived gravity grid (e.g., Sandwell et al., 2014) visually in Google Earth (www.google.com/earth/). The proposed methodology is divided into archive file assembly including reduction of potential field data (Hamilton et al., 2019), along-track analysis (section 2.1), crossover analysis (section 2.2), sonar processing (section 2.3), data visualization (section 2.4), and automation at sea (section 2.5). Brief examples of commands are interspersed in the text; however, the tools are undergoing development; hence, syntax will change as the tools evolve. Documentation for the MGD77 and X2SYS tools can be found online (www.generic-mapping-tools.org), while shell scripts, source code, and documentation pertaining to this study can be found at www.soest.hawaii.edu/mgd77 website.

2.1. Along-Track Analysis

Along-track analysis, detailed extensively by Chandler and Wessel (2008, 2012), is a technique by which marine geophysical track line files are examined internally for various types of errors. For instance, data values should range within reasonable limits. Vessel position should not pass over land, time should not be disrupted as a vessel crosses a time zone boundary, nor should vessel speed exceed realistic limits. Additionally, certain header fields describe the data records, such as the geographical extents of the survey, the

10.1029/2018GC007891

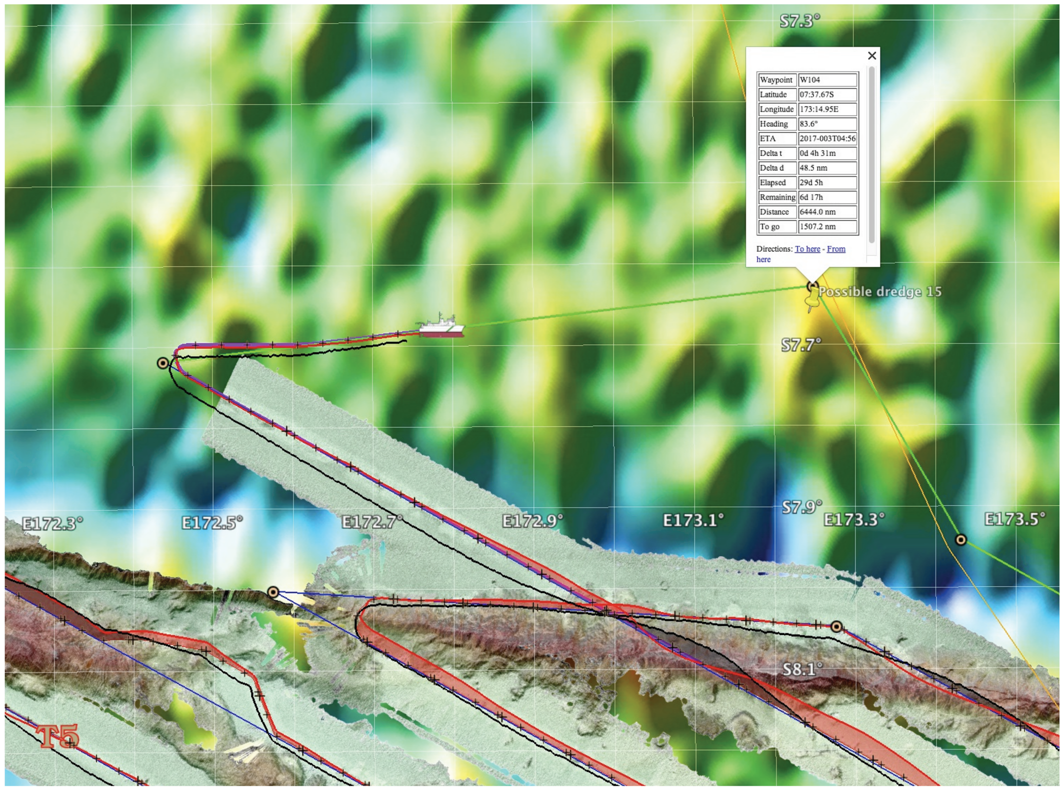

Figure 2. Monitoring full resolution geophysical data in Google Earth during R/V *Kilo Moana*'s KM1609 expedition during 2016/2017. Multibeam backscatter and bathymetry, cleaned and processed automatically using SURVEY, is overlain by reduced gravity and magnetic anomalies. These anomalies are shown as black and red along-track wiggle plots, respectively. The background image is vertical gravity gradients (Sandwell et al., 2014) and the pop-up box shows waypoint information.

10° by 10° boxes crossed, among others, and should agree. Reported gravity and magnetic anomalies also need to agree with anomalies that are recalculated internally by the program; this test can detect instrument drift and tares, outdated or erroneous magnetic anomalies, and fidelity issues among the various data columns. Shipboard depth and gravity measurements are also checked against global grids using regression analysis for checking that the shipboard data are appropriately scaled and not vertically offset. We also use an excessive offset test, which flags excessively large offsets between ship and gridded data as suspicious. Along-track analysis is an effective way to uncover problems in both new and existing data sets. With existing data sets an errata table is generated (e.g., *mgd77sniffer 08040006 -R152.1/202.85/-12.3/21.4 -De -Gdepth,topo.18.1.grd -Gfaa,grav.24.1.grd -Gnav,dist2coast.grd*) and reviewed with the aid of an errata review plot; detected errors are flagged for correction by the reviewer (Chandler & Wessel, 2012). Then, by converting the archive file to the extended MGD77+ format into which error flags are also applied (i.e., *mgd77manage cruiseid -Ae*), corrected data can be obtained via *mgd77list* (e.g., *mgd77list cruiseid -Fdat*). When building an archive file from new data, errata output should be reviewed but significant errors should be corrected by fixing the error source and the archive file should be recompiled. Then, the next generation errata table can be reviewed to ensure the error has been corrected and that no other significant errors persist. After several iterations of this, and after dealing with the various other mishaps and errors that tend to occur at sea, the archive file should be stable and the data ready for visual inspection.

2.2. Crossover Analysis

After extreme errors are flagged and removed using along-track analysis, additional internal error reduction can be carried out by analyzing measurement discrepancies at ship track intersections. Widely used for analyzing survey tracks that intersect with other survey tracks, especially in gravity and magnetic surveys, COE analysis can greatly inform the scientist as to the magnitude and source of errors in geophysical data. We perform COE analysis using the method of Wessel (2010), which involves creation of an x2sys database particular to the project and consisting of all involved survey tracks. At sea, COE analysis would typically be internal only given the preliminary nature of the newly acquired magnetic and gravity anomalies, for example, in order to examine instrument drift. Correction against other published data sets is more meaningful after potential field anomalies are fully reduced. Because COE corrections increasingly lose their relevance as new surveys are added, they need to be updated as new surveys become available and for analyzing new shipboard data. In light of this, the x2sys toolkit was designed for creating and incrementally updating COE databases and correction tables. Because the tools support exchange formats including MGD77T, MGD77+, GMT DAT, and generic XYZ data structures, reanalyzing COE and updating correction tables should be easier within this methodology.

2.3. Sonar Processing

It is not often justifiable for a scientist to be trained to use commercial sonar processing software. Such software packages may vary from ship to ship, year to year, as do the seagoing personnel, and they typically provide a limited number of licenses. Some choose to hire contractors to do the sonar processing work. However, this quickly becomes expensive and more importantly the opportunity for increasing data familiarity and learning valuable processing skills is lost. For the processing of deep sea sonar data in academic settings, commercial software and professional sonar processors are not generally required because such data are not subject to stringent processing standards as is the case for inshore data used to chart harbors and their approaches. For the latter case, and for larger projects consisting of multiple surveys needing to be tied to a reference datum, additional rigorous processing steps are required, which will likely require the expertise of professional hydrographers.

The free MB-System toolkit (Caress & Chayes, 1995) is a fully viable solution for academic settings, which is complementary to our script-based processing methodology. The MB-System command line tools support most sonar formats, their conversion, automatic cleaning, data extraction, gridding, and plotting. The package also offers mbeditviz, a capable 3-D ping editor, which works symbiotically with the other MB-System tools. Using MB-System avoids time-consuming conversions and processing flag disconnects that can arise when using separate commercial processing software. At sea, we add MB-System conversion, automatic cleaning (using mbclean) for visualization purposes, and gridding to our iterative processing loop so that we can automatically update our sonar data display in Google Earth as new data come in. We then use the 3-D editor to remove outliers for producing archive-quality data, in the same manner as typically done with commercial processing software. Furthermore, MB-System can be used to process unlimited numbers of multibeam surveys which allows the user to create and update large data compilations. For these reasons, the MB-System toolkit is sufficient for most academic purposes.

2.4. Data Visualization

Data are often plotted in a time-intensive manner that involves finding the optimum map scale, extents, projection, color scheme, wiggle scale and orientation, etc., at each of various subregions of a survey. By switching from statically produced paper plots to automatically loading cruise data in Google Earth instead, which does not involve uploading to remote servers, geophysical data can be plotted at full resolution in Google Earth's interactive environment. Hence, one can easily zoom in or out, pan around, edit survey waypoints, compare shipboard data with other data sets, measure distances, examine sonar, gravity, and magnetic data in context, further scrutinize the incoming data, and spend a great deal more time focusing on what the data mean. Global grids of satellite altimetry-derived gravity (e.g., Sandwell et al., 2014), predicted-observed bathymetry compilations (e.g., Ryan et al., 2009; Smith & Sandwell, 1994; Weatherall et al., 2015), and global magnetic anomaly grids (e.g., Dyment et al., 2015) provide highly suitable background imagery.

A new set of shipboard preprocessing and visualization tools, known as “SURVEY” (Wessel et al., 2017), was recently prototyped aboard R/V *Kilo Moana* during leg KM1609, which examined the breakup of the Ontong Java Nui superplateau via detailed mapping and sampling of the Ellice Basin. SURVEY was designed for management and visualization of a geophysical survey within Google Earth. As with other survey management programs, a set of waypoints is required. With SURVEY, these can be created within Google Earth and edited as the cruise progresses. SURVEY tallies several useful statistics such as distance to next waypoint, distance to end of cruise, average speed, survey time remaining based on average speed, and heading. It also highlights the current survey line and plots vessel position at near real-time using incoming vessel navigation. For sonar data monitoring, SURVEY copies new sonar files to the local disk, preprocesses them into MB-System format, runs mbclean to automatically remove extreme bathymetric outliers, then interpolates

10.1029/2018GC007891

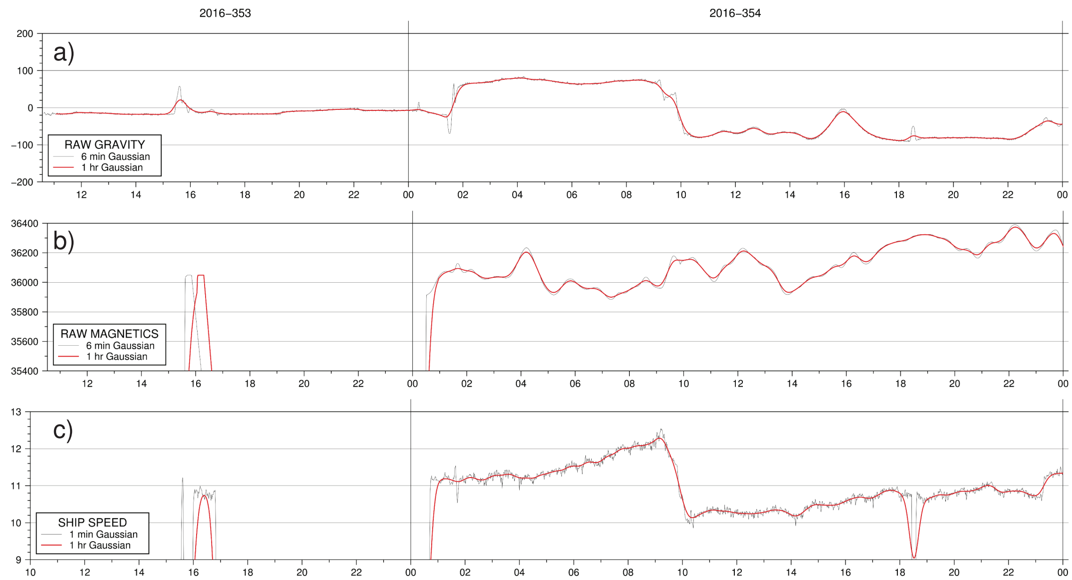

Figure 3. Plots of (a) raw gravity and (b) raw magnetics, Gaussian filtered at wavelengths of 6 and 60 min, and (c) ship speed, Gaussian filtered at wavelengths of 1 and 60 min, produced during days 353–354 of cruise KM1609.

multibeam and backscatter measurements onto grids which then get converted to PNG images wrapped in Google's Keyhole Markup Language (KML). To streamline these tasks, which typically run every 30 min as each new sonar file comes in, it may be necessary to divide a large survey into several smaller mapping zones of approximately equal area. Within the current mapping zone, SURVEY splits the zone's sonar data into $1 \times 1°$ KML files so that full resolution data can be viewed in Google Earth, as illustrated in Figure 2. Because SURVEY was developed during a single cruise in which a Kongsberg EM122/EM710 sonar combination was deployed along with a BGM-3 gravimeter and Geometrics G-882 cesium fluxgate magnetometer, refinement will be necessary in future applications, especially when deploying different instruments. Nonetheless, the interactive graphical display of survey data at full resolution, especially when situated prominently for all on board to see, clearly illuminates newly charted seafloor fabric in a manner that is useful for science and cruise administration. This capability could be a useful tool for seagoing geophysicists.

Although SURVEY is designed to pre-process data for rapid visualization in Google Earth, it also produces stand-alone plots of raw gravity, magnetics, and ship speed (Figure 3) for monitoring potential field and navigation data. In contrast, our archive and quality control methodology is designed for data scrutiny and elimination of serious errors for the preparation of a fully vetted data set in the standard archive format. Candidate archive files also need to be visually reviewed, and this is accomplished by converting gravity, magnetics, and center beam depths from the archive format into KML along-track wiggles using GMT's *gmt2kml* program. These are then loaded in Google Earth, as shown in Figure 2. SURVEY can be used without archival preparation and quality control, but then a vetted and fully reduced archive file would not be completed by survey end. Conversely, archival preparation and quality control can be performed without SURVEY, but cruise management, sonar processing, and raw data visualization would need to be done on top of that, potentially at a considerable cost. Hence, the two techniques are distinct and complementary. For visualizing point and wiggle features in Google Earth, and because KML files cannot be arbitrarily large, some decimation or, preferably, splitting into multiple segments within the KML layer, will likely be required for longer surveys. By using autorefreshing KML containers in Google Earth, the various geophysical KML files update themselves in real time, as new data are acquired, greatly contributing to automation.

It is also useful to visualize additional metadata such as the locations of sound velocity casts, dredge targets, and a backward looking navigation log reporting past ship locations, time, heading, and speed at various intervals. During the KM1609 cruise, we plotted sound velocity cast locations (e.g., the red T5 in Figure 2 marks the location where a T5-type expendable bathymetric thermograph sound velocity probe was carried out), past and potential dredge locations such as shown in Figure 2, and navigation fixes at 10-min intervals and at each sonar file's start time (the small and larger plus symbols in Figure 2, respectively).

Table 1
Algorithm for Automatic Archive File Creation and Scrutiny

```
while (true)
    Copy the latest underway geophysics files to processing workstation
    Validate raw underway data
    Filter raw potential field data
    Resample potential field data (optional)
    Merge geophysics with navigation into raw archive file data records
    Create the archive file header via mgd77header
    Reduce gravity and magnetic data to create reduced archive file
    Convert reduced archive file to ASCII MGD77T
    Scrutinize data along-track via mgd77sniffer (outputs an errata table)
    Review errata table*
        *After resolving error sources errata approval can be automated
         (e.g., sed -e 's/? Errata table ver/Y Errata table ver/g')
    To apply remaining (minor) errata corrections
        1. Convert reduced data to MGD77+ format
        2. Apply errata corrections (mgd77manage -Ae)
        3. Extract corrected data via mgd77list
    Update gravity and magnetic KML wiggle files using gmt2kml
end
```

2.5. Automation at Sea

Once the above methodology is in place and extreme errors have been resolved, automation is accomplished by compiling the steps into a script and calling it at the desired time interval or, if processing time is not constant, by executing the script within an infinite loop which will restart immediately after the previous iteration completes (e.g., Table 1). Processing time will increase throughout a survey if the process is started from scratch at each iteration; hence, it is advantageous to process new measurements only and to carefully append these reduced and merged records to the growing archive file.

Multibeam sonar files spanned 30 min each, so we set up a repeating process to execute SURVEY commands shortly after each new file finished logging. Steps included fetching and preprocessing the new sonar file, recompiling the multibeam and backscatter grids, and creating updated KML tiles. In addition to this cron daemon, another cron daemon was set up to automatically update the ship's position in Google Earth, which was also accomplished using SURVEY. Plots of ship speed and raw magnetics and gravity data were also updated at 1 minute intervals via cron jobs calling SURVEY modules. These operations are described in detail in SURVEY's documentation (Wessel et al., 2017).

Once the sonar tiles are updated with the latest data, the autorefreshing KML container causes Google Earth to redraw the KML layer at regular intervals specified in the container. When automated, it is possible for the seagoing scientist to focus on more important tasks, such as cruise management and data scrutiny and interpretation.

3. Discussion

Although it is important that scientists remain intimately involved in data handling while at sea, it is foreseeable that these methods might be applicable by marine centers as well. For instance, raw data to archive file conversion, merging, and anomaly reduction can all be done within the ship's processing systems. Hence, scientists could then go right to errata table review, data scrutiny in Google Earth, and monitoring free-air gravity, residual magnetic anomalies, and swath imagery in real time, rather than watch the less intuitive meter counts that are typically reported at present and without waiting on data processors to provide sonar maps. Processing sonar data using MB-System and automatically plotting in Google Earth accomplishes much of what the seagoing scientist requires; hence, under this methodology it is generally not necessary for scientific purposes to learn commercial processing software or contract a sonar data processor. For University-National Oceanographic Laboratory System cruises, perhaps the completed archive file could be shared with R2R in order to minimize loss of metadata, while international operators may wish to share their file with NCEI to ensure global availability. These methods apply to geophysically oriented academic surveys but with more stringent sonar processing can be applied in nautical charting missions as well. We also encourage improvement of the proposed methodology such as by integrating further quality control tests from other researchers and handlers of marine geophysical data, including R2R. In particular, further improvement of data quality among existing data sets could be achieved through a community approach to along-track analysis and errata table review.

The existence of extreme errors in data archived to date and the potential for new errors to arise due to the ongoing lack of real time quality control both imply that error sources should be dealt with while at sea so that subsequent researchers can fully utilize marine geophysical data without needing to guess at what might have gone wrong during acquisition. We envision several scenarios for accomplishing quality control at sea. One option lies with the scientists funded to collect and interpret the data. Seagoing geophysicists possess first person exposure to the new information, including intimate knowledge of hardware malfunctions, mishaps, and causes of other data inconsistencies as advantages over others who might be tasked with data processing and archival after-the-fact. Although nongeophysicists cannot be expected to take on this role, the necessity to ensure that data are valid can not be reasoned away given the high cost of their acquisition, their importance to other sectors of science and the poor likelihood that a given study area will be revisited. Ideally this would involve geophysicists joining nongeophysical surveys but adding such a requirement to all seagoing investigators would likely be intolerable and would likely be inconsistently applied. To avoid such a burden, perhaps fleet management organizations such as University-National Oceanographic Laboratory System, National Oceanic and Atmospheric Administration, and Japan Agency for Marine-Earth Science and Technology, rather than principal investigators, could organize a pool of seagoing geophysical technicians who would participate in all expeditions lacking sufficient geophysical expertise with costs to be included in the daily rate. A third scenario is to train the current pool of marine technicians to perform data scrutiny at sea. We are often told that the main responsibility for seagoing technicians is to ensure instruments are running and data are being recorded. This is of course true, but it is equally important that the collected data not suffer from serious error, rendering the data useless. Thus, acquisition and quality control must go hand in hand and be part of the expectation for marine surveys and therefore should happen at sea. Given that day rates for marine research are already sufficiently high (e.g., ~\$30–\$40k/day and up for modern regional/global class research vessels), that it is reasonable to expect advanced data visualization and quality control systems to be included in daily costs, and that considerable pushback would likely result from tasking scientists or marine technicians with the quality control burden, it seems that the preferred scenario would be to task fleet management organizations with installing at sea quality control systems and with providing sufficient expertise on nongeophysically oriented expeditions.

Toward Seabed 2030's goal of mapping the world's deep oceans, our proposed methods can aid on the back end of such a global mapping effort, for archiving and quality control purposes. Given that the Seabed 2030 project appears to be getting underway, now is an opportune time to implement enhanced community infrastructure such as what we are proposing. For example, magnetic data acquisition is especially desirable within the framework of a global mapping effort and also to ongoing survey efforts in general, due to the importance of such data and the age and sparse coverage of existing magnetics. On the eve of this immense undertaking, it is advantageous to ensure that the essential forms of data will be acquired. For instance, in addition to compiling a complete map of seafloor bathymetry, simultaneous acquisition of all routinely collected geophysical and oceanographic data, including towed magnetics, should be feasible at little extra cost. Such an expanded scope is in line with what GOMaP proposed and would be of immense value to the scientific community.

In order to accomplish these many tasks and to fully deploy the seagoing scientist's toolbox, we recommend implementing these methods on as computationally and graphically powerful a workstation as practical running either UNIX, Linux, or Mac operating systems, wherein the tools listed above are fully supported and under development. For optimum data visualization, a large 4 K or better monitor or large-screen television is also recommended. And while it would be possible to bypass archival preparation and use the other techniques, quality control would not be possible as intended. Furthermore, the potential would be greatly increased that someone without first-hand acquisition experience and perhaps without in-depth awareness of the acquired data would be required to create the archive file. Therefore, we encourage those with geophysical expertise to create the archive file first and to fully scrutinize it while at sea.

4. Conclusions

Marine geophysical data can be converted to the standard data exchange format, reduced, analyzed along track, and plotted in Google Earth in real time, while at sea, using the proposed techniques. The methods liberate the seagoing scientist from needing to redundantly prepare maps and data plots at each area of interest. By plotting processed data automatically in Google Earth using the SURVEY package, scientists can more easily visualize data and conduct survey management. By visualizing the various acquired data sets simultaneously, extreme errors will be more noticeable, especially when compared to reference data including gridded gravity, magnetics, and bathymetry. This methodology requires only freely available software and eliminates, in most cases, the need for hiring expensive data processors. It is our hope that each data set being submitted to a data repository will be free from extreme errors.

Acronyms

MGD77	Marine Geophysical Data Exchange Format - 1977
MGD77T	Tab-delimited MGD77
GMT	Generic Mapping Tools
NCEI	National Centers for Environmental Information
KML	Keyhole Markup Language

Acknowledgments
Funding was provided in part by U.S. National Science Foundation Grants 1458964 and 1558403 and by Korea Institute of Ocean Science and Technology Grants PM59941 and PM60321. This paper greatly benefitted from critical reviews by Andrew Goodwillie (LDEO) and an anonymous reviewer and is an expanded response to a thesis defense question asked by Bruce Appelgate on the applicability of along-track analysis to real-time data acquisition. All data used are archived in the Marine Trackline Geophysical Data repository (https://doi.org/10.7289/V5CZ35DR) and Multibeam Bathymetry Database (https://doi.org/10.7289/V56T0JNC). We thank leg KM1609's marine technicians for exceptional support and for ensuring uninterrupted access to incoming data. This is SOEST Publication 10821 and KIOST Publication XXXX.

References

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., et al. (2009). Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. *Marine Geodesy*, *32*, 355–371.

Caress, D. W., & Chayes, D. N. (1995). New software for processing sidescan data from sidescan-capable multibeam sonars. In *Proceedings of the OCEANS Conference*, Challenges of Our Changing Global Environment, MTS/IEEE.

Carron, M. J., Vogt, P. R., & Jung, W.-Y. (2001). A proposed international long-term project to systematically map the world's ocean floors from beach to trench: GOMaP (Global Ocean Mapping Program). *International Hydrographic Review*, *2*(3), 49–55.

Chandler, M. T., & Wessel, P. (2008). Improving the quality of marine geophysical track line data: Along-track analysis. *Journal of Geophysical Research*, *113*, B02102. https://doi.org/10.1029/2007JB005051

Chandler, M. T., & Wessel, P. (2012). Errata-based correction of marine geophysical trackline data. *Geochemistry, Geophysics, Geosystems*, *13*, Q10004. https://doi.org/10.1029/2012GC004294

Dyment, J., Lesur, V., Hamoudi, M., Choi, Y., Thebault, E., Catalan, M., WDMAM Task Force, WDMAM Evaluators, & WDMAM Data Providers (2015). World Digital Magnetic Anomaly Map version 2.0, map available at http://www.wdmam.org.

Hamilton, M., Wessel, P., Taylor, B., & Luis, J. (2019). Producing marine geophysical archive files from raw underway data. *Computers & Geosciences*, *133*, 104321. https://doi.org/10.1016/j.cageo.2019.104321

Hittleman, A. M., Groman, R. C., Haworth, T. L., Holcombe, T. L., McHendrie, G., & Smith, S. M. (2010). *The Marine Geophysical Data Exchange Format MGD77 - MGD77T/Legacy MGD77—Key to geophysical records documentation 10 (revised)*. Boulder, Colorado, USA: National Geophysical Data Center.

Maus, S., Barckhausen, U., Berkenbosch, H., Bournas, N., Brozena, J., Childer, V., et al. (2009). EMAG2: A 2-arc-minute resolution Earth magnetic anomaly grid compiled from satellite, airborne and marine magnetic measurements. *Geochemistry, Geophysics, Geosystems*, *10*, Q08005. https://doi.org/10.1029/2009GC002471

Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., et al. (2018). The Nippon Foundation—GEBCO Seabed 2030 project: The quest to see the world's oceans completely mapped by 2030. *Geosciences*, *8*, 63. https://doi.org/10.3390/geosciences8020063

Miller, S. P., Carbotte, S. M., Stocks, K., Ferrini, V., Arko, R. A., Chandler, C. L., et al. (2010). Rolling deck to repository (R2R): Next steps in ocean exploration for data dissemination and discovery. *Eos, Transactions American Geophysical Union*, 91, OS13C–124.

Quesnel, Y., Catalan, M., & Ishihara, T. (2009). A new global marine magnetic anomaly data set. *Journal of Geophysical Research*, *114*, B04106. https://doi.org/10.1029/2008JB006144

Ryan, W., Carbotte, S., Coplan, J., O'Hara, S., Melkonian, A., Arko, R., et al. (2009). Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems*, *10*, Q03014. https://doi.org/10.1029/2008GC002332

Sandwell, D. T., Muller, R. D., Smith, W. H. F., Garcia, E., & Francis, R. (2014). New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. *Science*, *346*(6205), 65–67.

Smith, W. H. F. (1993). On the accuracy of digital bathymetric data. *Journal of Geophysical Research*, *98*(B6), 9591–9603.

Smith, W. H. F., & Sandwell, D. T. (1994). Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research*, *99*(B11), 21,803–21,824.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., et al. (2015). A new digital bathymetric model of the world's oceans. *Earth and space Science*, *2*, 1–15. https://doi.org/10.1002/2015EA000107

Wessel, P. (2010). Tools for analyzing intersecting tracks: The x2sys package. *Computers & Geosciences*, *36*, 348–354.

Wessel, P., & Chandler, M. T. (2007). The mgd77 supplement to the Generic Mapping Tools. *Computers & Geosciences*, *33*, 62–75.

Wessel, P., & Chandler, M. T. (2011). The spatial and temporal distribution of marine geophysical surveys. *Acta Geophysica*, *59*(1), 55–71.

Wessel, P., Chandler, M. T., & Wright, N. (2017). SURVEY—Tools for at-sea data processing and display via Google Earth. http://www.soest.hawaii.edu/pwessel/survey

Wessel, P.,Luis, J., Uieda, L.,Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D., (2019). The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, *20*. https://doi.org/10.1029/2019GC008515

Wessel, P., & Watts, A. B. (1988). On the accuracy of marine gravity measurements. *Journal of Geophysical Research*, *93*(B1), 393–413.