Computers & Geosciences 36 (2010) 348–354

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Tools for analyzing intersecting tracks: The x2sys package

Paul Wessel*

Department of Geology and Geophysics, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1680 East–West Road, Honolulu, HI 96822, USA

ARTICLE INFO

Article history:
Received 23 February 2009
Received in revised form
28 May 2009
Accepted 29 May 2009

Keywords:
Track intersection
Crossover errors
Leveling
Data quality improvement

ABSTRACT

I present a new set of tools for detection of intersections among tracks in 2-D Cartesian or geographic coordinates. These tools allow for evaluation of crossover errors at intersections, analysis of such crossover errors to determine appropriate linear models of systematic corrections for each track, and application of these corrections and further adjustments to data that completely eliminates crossover discrepancies from final 2-D data compilations. Unlike my older x_system tools, the new x2sys tools implement modern algorithms for detecting track intersections and are capable of reading a wide range of data file formats, including data files following the netCDF COARDS convention. The x2sys package contains several programs that address the various tasks needed to undertake a comprehensive crossover analysis and is distributed as a supplement to the Generic Mapping Tools, making them available for all computer platforms and architectures.

1. Introduction

Measuring one or more observables as a function of position or time along-tracks is a standard procedure in many scientific or engineering endeavors. Examples include underway geophysical observations collected by oceanographic vessels, remotely-sensed data collected by orbiting satellites, airborne collection campaigns, or terrestrial measurements of various types obtained along crisscrossing traverses. In many applications, the users will compile all data obtained by one or more such expeditions and use interpolation methods to resample the measured data from their arbitrary locations along track lines onto an equidistant lattice required for further analysis. It has long been known that complications are associated with this seemingly innocuous task, giving rise to what is known as crossover errors (COE) (e.g., Wessel and Watts, 1988; Smith, 1993). These are defined to be the difference between the two repeat measurements at track intersections. Such COE can result from a variety of sources. For instance, the coordinates of the track may have uncertainties and therefore the location of the track intersection may not really correspond to the correct repeat measurement point along the two tracks. Unless the phenomenon being measured is constant, one will find nonzero COE simply due to this mis-registration (e.g., Fox et al., 1992; Neumann et al., 2001). Interactive tools, such as MB-system's mbnavedit (Caress and Chayes, 1996) may be used to correct such problems. Other sources of difficulty derive from variability of the phenomena itself or how the observations are obtained, such as the introduction of aliasing. Some phenomena, such as the Earth's magnetic field, change over time and unless proper corrections are made, one may end up with COE between tracks from different epochs. Other COE may derive from improper or missing calibration of instruments, and in some cases different tracks may have used different units for the measurements. There are thus numerous examples of studies in which COE or leveling corrections have been applied (e.g., Foster et al., 1970; Swan and Young, 1978; Yarger et al., 1978; Kellogg, 1979; Prince and Forsyth, 1984; Green, 1987; Nishimura and Forsyth, 1988; Tai, 1988; Wessel and Watts, 1988; Ghosh and Hall, 1992; Minty, 1991; Klokocnik et al., 1998; Huang and Fraser, 1999; Neumann et al., 2001; Mauring and Kihle, 2006; Huang, 2008) and some software solutions have been published (Wessel, 1989; Hsu, 1995).

The approach presented here divides the many tasks that comprise a comprehensive crossover analysis into smaller steps, and each step has been encoded as a stand-alone program. The step-by-step assessment makes it easy to test intermediate results and make sure all procedures are well described and documented. While most of the tools deal with crossover calculation and extraction, some are dedicated to maintaining a database of all tracks and their content, which is useful when tracks contain more than one type of measurement and when one needs to access just the tracks that contain a particular measurement of interest. In the next sections, I present an overview of the various tools, and then demonstrate their use with a synthetic case study. In the interest of brevity and to preempt confusion arising from possible future syntax changes, I will not describe the particular program syntax here since these are best left to the documentation that comes with the package

* Tel.: +1 808 956 4778; fax: +1 808 956 5154.
E-mail address: pwessel@hawaii.edu

0098-3004/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2009.05.009

because it will be updated, as changes require. The x2sys package is distributed as an official supplement to the Generic Mapping Tools (Wessel and Smith, 1995) and benefits from the general processing capabilities of GMT. In turn, GMT uses the x2sys algorithms for hidden tasks such as determining intersections between map contours and annotation lines.

2. The x2sys family of tools

There are nine tools in the x2sys suite, each dedicated to a particular aspect of the analysis. Not all are required necessarily for a particular analysis; I will indicate their typical use as I discuss their general purposes.

2.1. Initializing with x2sys_init

The tool x2sys_init is the starting point for anyone wishing to use x2sys; it initializes a set of databases that are particular to one kind of track data. These data, their associated databases, and key parameters are given a shorthand notation called an x2sys tag. The tag keeps track of common settings such as data format, whether the data are geographic or Cartesian, what units to use for speed and distance, the extent of the data domain, how to recognize data gaps along-tracks (based on exceeding time or distance criteria) and the optional binning resolution for track indices (more on this below). Running x2sys_init is a prerequisite to running any of the other x2sys programs. A key input to x2sys_init is a description of the data file format. This information is encoded in an ASCII table called a definition file that must have the extension ".def". The definition file has two sections: Header information and column information. All header information starts with the character # in the first column, immediately followed by an upper-case directive. Spaces or tabs separate directives from any optional argument. Five header directives are recognized: ASCII indicates that the data files are in ASCII format, BINARY states that the data files are native binary files, and NETCDF states that the data files are COARDS-compliant 1-D netCDF files (see http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html). SKIP takes an integer argument that is either the number of lines to skip (for ASCII files) or the number of bytes to skip (for native binary files); it is not used with netCDF files. GEO indicates that these files are geographic data sets, with periodicities in the x-coordinate (longitudes). MULTISEG means each track consists of multiple segments separated by a GMT multi-segment header; again, this directive is not used with netCDF files.

The column information consists of one line per column in the order that the columns appear in the data file. For each column you must provide the seven attributes *name type NaN NaN-proxy scale offset oformat*. Here, *name* is the name of the column variable. It is required that you use the special names lon (or x if Cartesian) and lat (or y) for the two required coordinate columns, and time when optional time data are present; other column names have no restrictions. The *type* is always **a** for ASCII representations of numbers, whereas for binary files you may choose among **c** for signed 1-byte character (−127,+128), **u** for unsigned 1-byte (0–255), **h** for signed 2-byte integers (−32768, +32767), **i** for signed 4-byte integers (−2,147,483,648, +2,147,483,647), **f** for 4-byte floating points, and **d** for 8-byte double precision floating points. For netCDF files, simply use **d** because netCDF will automatically handle type-conversions during reading. NaN should be **Y** if certain values (e.g., −9999) are to be replaced by NaN (i.e., IEEE "Not-a-Number"), and **N** otherwise; *NaN-proxy* should hold the designated replacement value. Next, *scale* is used to multiply the data after reading, while *offset* is used to add to the scaled data. Finally, *oformat* is a C-style format string used to print values from this column. If you specify – as the *oformat*, then GMT's formatting machinery will be used (i.e., you can any format you like by manipulating the settings for D_FORMAT, PLOT_DEGREE_FORMAT, PLOT_DATE_FORMAT, and PLOT_CLOCK_FORMAT).

Table 1
Example of a file format definition file (here, geoz.def).

```
# Definition file for X2SYS processing of ASCII lon,lat,z files
#
# This file applies to plain 3-column ASCII files
#
#
#ASCII          # The input file is ASCII
#SKIP 1         # The number of header records to skip
#
#name  intype  NaN-proxy?  NaN-proxy  scale  offset  oformat
lon    a       N           0          1      0       %10.5 f
lat    a       N           0          1      0       %9.5 f
z      a       N           0          1      0       %6.1 f
```

Some file formats already have definition files that come with the x2sys distribution. These include *mgd77* for plain ASCII MGD77 data files (Hittelman et al., 1977), *mgd77+* for enhanced MGD77+ netCDF files (Wessel and Chandler, 2007), *gmt* for old mgg supplement binary files (Wessel and Smith, 1991), *xy* for ASCII Cartesian (x, y) tables, *xyz* extends *xy* with one z-column, *geo* for plain ASCII (longitude, latitude) files, and *geoz* which extends *geo* with one z-column. The latter data file format will be used for the examples in this paper and is reproduced in Table 1. For any other formats you must craft your own definition file; the x2sys_init documentation provides further examples.

2.2. Maintaining a track database

If your track data have more than one observed data type or you anticipate getting additional tracks later (and at that time will need to recompute cross-overs involving these new tracks and the older tracks they cross), I recommend you maintain and use a track database. During the initialization above you can specify a binning interval, which lets the database know the geographic extent of each track at that resolution. E.g., if you choose 1° then the database will maintain indices at the 1° × 1° resolution and thus know all such bins that a track passes through. There are three x2sys programs that maintain and use the track database: (1) x2sys_binlist will read all new data tracks and create a bin index file based on the chosen bin interval, (2) x2sys_put will read the new bin index file and append the information to the binary database created by x2sys_init in the previous section, and (3) x2sys_get is used to query the database about tracks that satisfy specific criteria (e.g., geographic region, type of observables), or optionally to provide a list of track pairs that might cross based on their indices and thus should be subject to further crossover assessment.

2.3. Determining crossover errors

The program x2sys_cross is used to determine all intersections between two tracks ("external cross-overs") or within a single track ("internal cross-overs"), and will report the time, position, distance along-track, heading and speed along each track segment, and the crossover error (COE) and mean values for all observables in the track files. You may specify a list of tracks (in which case the program will examine all possible pairs for crossings) or a list of pairs directly (such as provided by x2sys_get). Unlike my original crossover detector (Wessel, 1989), the crossover locations are found by implementing the general line intersection algorithm of Sedgewick (1990). The values of the observations at the intersections are interpolated from the nearby along-track values using the specified linear, cubic, or Akima (1972) spline interpolator. The output of x2sys_cross contains COE information for each track pair that generated intersections and can be appended to a growing data file with all COE obtained to date. I refer to this file as the crossover database.

2.4. Reporting crossover statistics

The program x2sys_report can read the crossover database and report on the statistics of COE for each track. It has options that allow excluding tracks and limit what kind of output is reported. The program can also be used after crossover corrections have been determined: x2sys_report can report on the crossover statistics after you apply such corrections; this allows you to gauge improvement due to the systematic corrections.

2.5. Extracting crossover data for analysis

The crossover database contains information for COE for all the observables provided by the track data. However, when seeking to find systematic trends to correct, one must work with one type of data because such corrections tend to be data specific. Here, x2sys_list will read the entire database and extract a subset of COE based on your criteria. Typically this involves restricting the output to COE of a certain data type. Furthermore, it also means you may need to request one or more auxiliary values at the intersection points because these may be required in the analysis. Examples of such auxiliary information includes distance or time (absolute or relative to start of track, or the interval between repeat measurements), various geometric parameters (angles between tracks or their azimuths), the track IDs, x and y (or longitude, latitude for geographic data), and speed. Finally, it is common to treat internal and external COE in separate analyses. The output of x2sys_list is the data that will be examined for systematic effects.

2.6. Determining systematic corrections

Next, x2sys_solve will use the crossover information extracted with x2sys_list to solve for systematic corrections that may then be applied per track to improve overall data quality. Several systematic corrections can be solved for using a least-squares approach. As discussed in Section 2.4, only one data type can be processed at a time. There are several systematic trends that you may try to fit to the data; others may be added as new data types demand them. As this paper is published, you can choose among the following linear functions $C_p(\mathbf{x})$, where $\mathbf{x}$ are the R parameters for track p that I will fit simultaneously using a least squares approach; in each case the expected input record content (generated by x2sys_list) is given (subscripts 1 and 2 in the record descriptions refer to the two tracks):

1. $C_p(\mathbf{x})=a_1$ (a constant offset [R=1]); records must contain track ID_1, ID_2, COE.
2. $C_p(\mathbf{x})=a_1+a_2 d$ (linear drift; d is the distance along-track [R=2]); records must contain track ID_1, ID_2, d_1, d_2, COE.
3. $C_p(\mathbf{x})=a_1+a_2(t-t_0)$ (linear drift; t_0 is the start time of the track [R=2]); records must contain track ID_1, ID_2, t_1-t_0, t_2-t_0, COE.
4. $C_p(\mathbf{x})=a_1 z$ (a unit scale correction [R=1]); records must contain track ID_1, ID_2, z_1, z_2.
5. $C_p(\mathbf{x})=a_1+a_2\sin^2(y)$ (theoretical gravity datum correction [R=2]); records must contain track ID_1, ID_2, latitude y, COE.
6. $C_p(\mathbf{x})=a_1+a_2\cos(h)+a_3\cos(2h)+a_4\sin(h)+a_5\sin(2h)$ (magnetic heading correction [R=5]); records must contain track ID_1, ID_2, headings h_1, h_2, COE.

Of these, functions 1–4 are fairly generic corrections applicable to many data types, while 5 is specific to gravity anomalies (e.g., Grant and West, 1965) and 6 is unique to marine magnetic data sets (e.g., Bullard and Mason, 1961; Buchanan et al., 1996). I expect this list to grow longer as users gain experience with the package and propose additional correction types.

Based on suggestions by Menke (1984), the correction terms, $\mathbf{a}$, for each track are determined as follows. Here, T is the total number of tracks, which combine to produce N cross-overs. For each crossover k there are two tracks involved: i_k and j_k. For readability, I drop the subscripts on i and j. At the crossover the observed track parameters are $\mathbf{x}_{ik}$ and $\mathbf{x}_{jk}$ and the crossover value $\Delta_k=z_{ik}-z_{jk}$. I seek corrections, for each track p, of the form

$$C_p(\mathbf{x}) = \sum_{r=1}^{R} a_{pr} f_r(\mathbf{x}), \tag{1}$$

where a_{pr} are the R unknown coefficients and f_r are the R basis functions chosen to model the corrections. Thus, (1) can describe any correction model that is linear in its parameters, $\mathbf{a}$. This gives a total of $M=TR$ model parameters a_{pr} to solve for. When corrections are *subtracted* from each track, the misfit at the kth crossover becomes

$$\begin{aligned} e_k &= [z_{ik} - C_i(\mathbf{x}_{ik})] - [z_{jk} - C_j(\mathbf{x}_{jk})] = \sum_{r=1}^{R} [a_{jr} f_r(\mathbf{x}_{jk}) - a_{ir} f_r(\mathbf{x}_{ik})] + \Delta_k \\ &= \Delta C_k + \Delta_k. \end{aligned} \tag{2}$$

Because the various tracks may be of differing quality we allow each track to have a weight inversely proportional to the standard error in the data, i.e., $w_i=\sigma_i^{-1}$. The total weighted misfit is as usual defined to be

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{e_k}{\sigma_k}\right)^2 = \sum_{k=1}^{N} \omega_k^2 e_k^2, \tag{3}$$

where the weight given to the kth crossover error is determined from the track weights as

$$\omega_k^{-2} = w_i^{-2} + w_j^{-2}, \tag{4}$$

by assuming the inherent standard errors of each track is independent of other tracks. I obtain the M normal equations by setting the M partial derivatives of χ^2 to zero:

$$\frac{\partial \chi^2}{\partial a_{ps}} = 0 \Rightarrow \sum_{k=1}^{N} \omega_k^2 e_k \frac{\partial e_k}{\partial a_{ps}} = \sum_{k=1}^{N} \omega_k^2 (\Delta C_k + \Delta_k) \frac{\partial e_k}{\partial a_{ps}} = 0, p=1,T; s=1,R, \tag{5}$$

i.e.,

$$\sum_{k=1}^{N} \omega_k^2 \Delta C_k \frac{\partial \Delta C_k}{\partial a_{ps}} = -\sum_{k=1}^{N} \omega_k^2 \Delta_k \frac{\partial \Delta C_k}{\partial a_{ps}}, \quad p=1,T; s=1,R, \tag{6}$$

where

$$\frac{\partial \Delta C_k}{\partial a_{ps}} = \begin{cases} -f_s(\mathbf{x}_{ik}), & p=i, \\ +f_s(\mathbf{x}_{jk}), & p=j, \\ 0, & p \neq i,j. \end{cases} \tag{7}$$

These M equations form the square matrix equation $\mathbf{Na}=\mathbf{b}$ that can be solved for $\mathbf{a}$. I wish to populate the matrix $\mathbf{N}$ and the vector

b in a loop over M as I process the individual cross-overs Δ_k. When the loop over M is at parameter number (p,s) the program visits each crossover $k=1,N$ and obtain the track IDs i and j. I update the rows in **N** and **b** as follows: If $p=i$ then I add these values to the two sets of R columns $(i, r=1,R)$ and $(j, r=1,R)$ in **N** and to the single row entry in **b**:

$$\{\omega_k^2 f_r(\mathbf{x}_{ik}) f_s(\mathbf{x}_{ik})\} \quad \{-\omega_k^2 f_r(\mathbf{x}_{jk}) f_s(\mathbf{x}_{ik})\} \quad +\omega_k^2 \Delta_k f_s(\mathbf{x}_{ik}). \tag{8a}$$

If $p=j$ then the terms to accumulate instead become

$$\{-\omega_k^2 f_r(\mathbf{x}_{ik}) f_s(\mathbf{x}_{jk})\} \quad \{\omega_k^2 f_r(\mathbf{x}_{jk}) f_s(\mathbf{x}_{jk})\} \quad -\omega_k^2 \Delta_k f_s(\mathbf{x}_{jk}). \tag{8b}$$

Otherwise I do nothing, as per (7). If $C_p(\mathbf{x})$ contains a constant offset (as many of the functions listed above do), then the system is under-determined and I must add a final constraint that sets the sum of all such offsets to zero. This is implemented via a Lagrange multiplier, λ, which is subsequently discarded:

$$\begin{bmatrix} \mathbf{N} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}. \tag{9}$$

These correction terms, per track and data type, are written to a correction table that other programs may use to apply these systematic corrections to the tracks specified.

2.7. Obtaining corrected data

After the crossover correction table has been created, one can use the x2sys_datalist tool to read the track data and automatically have the systematic corrections applied. Typically such corrected data provide a much-improved data set for further analysis. However, as only a portion of COE can be traced to (and corrected by) systematic effects, there will be residual COE even in the corrected data. Such mismatches may lead to artifacts when trying to process the data further (e.g., in gridding). To completely eliminate COEs, I follow established procedures (e.g., Mittal, 1984; Hsu, 1995) and use splines to distribute the remaining COEs along-track between neighboring intersections according to weights inversely proportional to the remaining COE variance for each track. This step requires the use of x2sys_report to set up the adjustment files before using x2sys_datalist.

3. Example of complete crossover analysis

To illustrate the use of the x2sys package I will present a synthetic case derived from real data. To simplify the presentation I will not reproduce actual command lines here (since they are subject to possible future syntax changes) but instead discuss in general terms and show the improvements graphically. The x2sys package contains data and scripts that will reproduce all the examples in this paper. Fig. 1 shows a satellite-derived free-air gravity anomaly grid (Sandwell and Smith, 2009) and a set of artificial ship tracks chosen to give fairly complete track coverage of this region. The grid is sampled along the tracks and the resulting synthetic data tracks constitute the data set I will analyze. Fig. 2a shows the sampled anomalies along-tracks while Fig. 2b shows a gridded solution where I have used the sampled track data with a minimum-curvature spline-in-tension interpolator (Smith and Wessel, 1990), with the tension parameter set to 0.25 (an appropriate value for gravity). All major features in Fig. 1 have been successfully recovered by the sampling and gridding. However, as along-track values need to be interpolated to yield an estimate exactly at the crossover points, we find the 1515 cross-overs generated yield a standard deviation of 4.63 mGal, with a mean close to zero. I now degrade the track data by adding arbitrary offsets and linear drift to about half of the tracks (see Table 2). After gridding these modified data

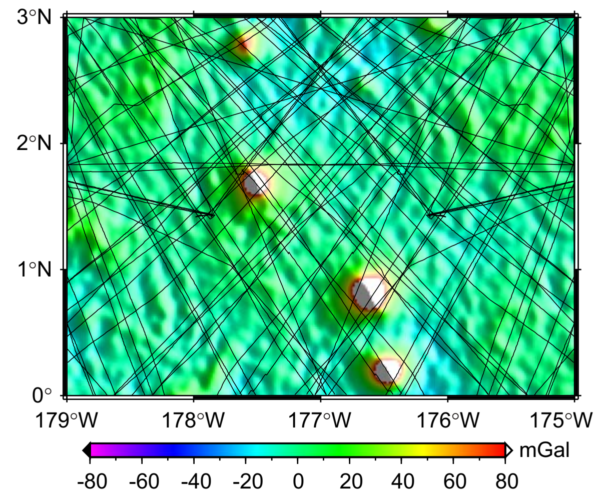

Fig. 1. Shaded-relief image of free-air gravity anomalies from satellite altimetry for a small region of Ratak seamount chain (Sandwell and Smith, 2009). Superimposed are ship tracks taken from other parts of world and shifted to overlay this region.

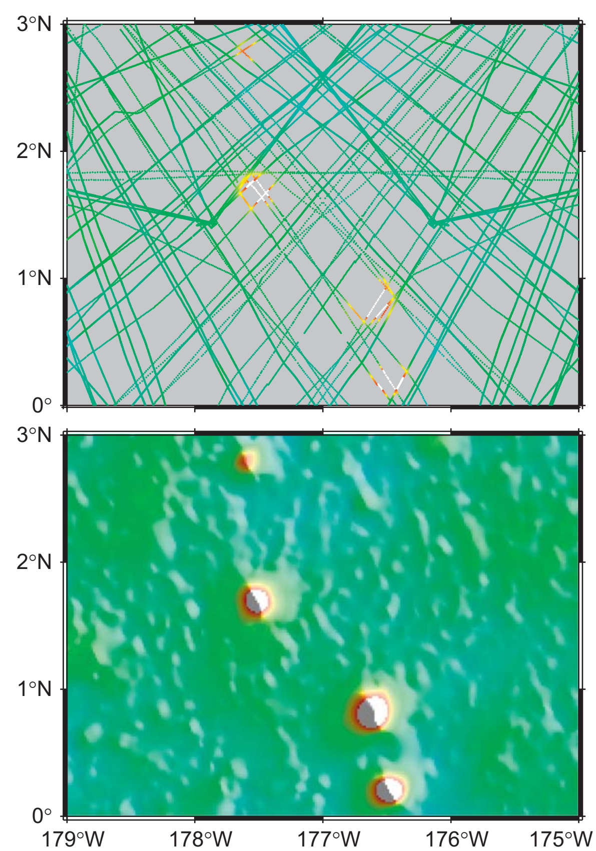

Fig. 2. (a) I use GMT's *grdtrack* program (Wessel and Smith, 1998) to sample grid along-tracks shown. (b) Using GMT's minimum-curvature spline-in-tension gridding algorithm *surface*, I obtain new grid based on sampled track data. Tension used in all gridding was 0.25 (Smith and Wessel, 1990). COE mean and standard deviation are −0.05 and 4.63 mGal, respectively.

I obtain Fig. 3, which clearly shows the unfortunate effects of gridding a data set that has not gone through a proper crossover analysis. No amount of filtering can remove these artifacts without also losing most of the short-wavelength signal of interest.

Table 2
Comparison between offsets and drifts for least squares solution and initial adjustments.

Track	a_1	a_2	offset	drift	Track	a_1	a_2	offset	drift
T_00	7.5	0.000	9.5	0.000	T_33	3.0	−0.029	–	–
T_01	0.2	−0.015	–	–	T_34	1.8	−0.016	–	–
T_02	−1.5	−0.005	–	–	T_35	21.4	−0.013	20.0	0.000
T_03	−6.5	0.024	−3.3	0.020	T_36	18.5	−0.099	–	–
T_04	−1.1	−0.003	–	–	T_37	−0.6	0.007	12.4	−0.006
T_05	3.6	−0.027	5.0	−0.020	T_38	−2.0	−0.004	–	–
T_06	−0.8	−0.005	–	–	T_39	2.5	0.010	6.6	0.000
T_07	−1.8	0.001	–	–	T_40	11.4	0.012	13.7	0.012
T_08	12.5	−0.016	13.5	0.006	T_41	3.4	0.015	5.0	0.000
T_09	−2.1	0.000	–	–	T_42	−14.1	−0.000	−12.5	0.000
T_10	3.5	−0.013	1.5	0.000	T_43	11.8	0.004	10.6	0.008
T_11	5.5	0.017	10.0	0.000	T_44	−18.8	0.015	−15.6	0.014
T_12	−3.8	0.008	–	–	T_45	−1.8	−0.005	–	–
T_13	−3.9	0.022	–	–	T_46	−0.2	0.014	0.0	0.020
T_14	−2.1	0.000	–	–	T_47	−0.7	−0.000	–	–
T_15	−2.1	−0.001	–	–	T_48	−3.9	0.015	–	–
T_16	0.6	−0.012	–	–	T_49	−1.4	−0.015	0.0	−0.020
T_17	−17.8	0.010	−15.7	0.010	T_50	1.4	−0.015	–	–
T_18	−3.4	0.003	–	–	T_51	12.6	0.019	16.3	0.000
T_19	−4.5	0.009	–	–	T_52	−7.6	−0.095	−13.6	0.000
T_20	−0.5	−0.013	–	–	T_53	−8.1	0.002	−6.0	−0.005
T_21	−1.3	−0.009	–	–	T_54	5.0	−0.000	8.9	0.000
T_22	4.5	0.010	7.7	0.000	T_55	−2.4	0.000	–	–
T_23	3.4	−0.032	4.5	−0.010	T_56	−1.6	−0.000	–	–
T_24	−1.7	−0.002	–	–	T_57	−15.6	−0.008	−14.0	0.000
T_25	5.3	0.001	9.9	0.000	T_58	−1.6	−0.010	–	–
T_26	−10.5	0.001	−7.8	0.000	T_59	2.1	0.030	4.8	0.030
T_27	12.7	0.012	14.5	0.006	T_60	0.8	0.122	–	–
T_28	1.6	−0.016	–	–	T_61	−2.1	−0.003	–	–
T_29	−2.9	0.008	–	–	T_62	−3.5	0.024	0.0	−0.006
T_30	−11.4	0.009	−8.8	0.008	T_63	−9.0	0.028	−6.7	0.000
T_31	4.6	−0.012	3.5	0.000	T_64	15.1	−0.013	13.2	0.000
T_32	0.7	−0.010	–	–	T_65	−2.1	0.037	–	–

Offset (mGal) and drift (mGal/km), if listed, reflect the values used to degrade the data. The a_1 and a_2 parameters are the least-squares solution.

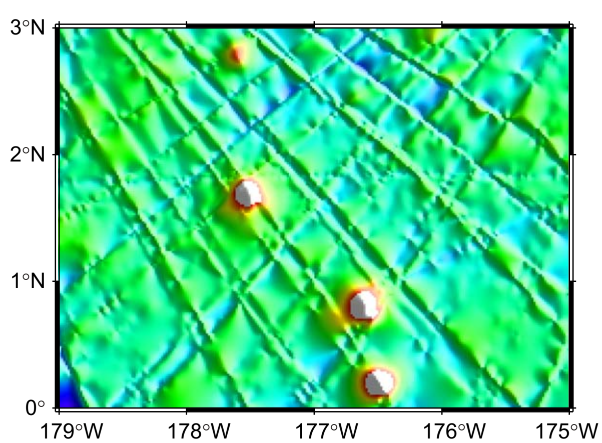

Fig. 3. Shaded-relief image as in Fig. 2b after about half of tracks had arbitrary linear trends (or just constants) added to them, changing values by 0–20 mGal. This introduces artifacts that produce COE. If data are not corrected these COE will lead to unacceptable gridded solutions. COE mean and standard deviation are now −1.03 and 11.20 mGal, respectively.

I use x2sys_init to define the particular data format and set parameters common to all the tracks used (such as how to compute distances, that the data are geographic, what the file extension is, etc.). For this simple example (a one-time correction) I do not need to interact any further with a track database (see Section 2.2). Once completed, I subject all the tracks to a complete crossover calculation. Here, x2sys_cross returns a single file with all the intersections and the estimated values of the gravity anomaly. I chose to get only external cross-overs and used an Akima (1972) spline to estimate the data values at the intersection points. The next step is to extract all the cross-overs and the auxiliary parameters needed to examine systematic trends. Deciding to determine linear trends, I use x2sys_list to extract distances along each track for each COE, the COE value, and the ID of the two tracks involved. These data are passed along to x2sys_solve, which sets up and solves the least squares problem of determining the best-fitting model parameters outlined in Section 2.5. These parameters are written to a correction table that now can be used by other x2sys programs as well as the mgd77 supplement (Wessel and Chandler, 2007). Table 2 compares the least-squares correction terms to the arbitrary changes used to degrade the data. Given that the synthetic data already had significant COEs simply due to interpolation, I did not expect to recover the exact parameters used to degrade the data. However, the most significant (i.e., largest) terms are reasonably well reproduced in the solution.

Now, I let x2sys_datalist extract all the track data and apply the least-squares corrections on the fly. Gridding the corrected data results in a much-improved image (Fig. 4a). However, I note there are linear track artifacts remaining in the solution (here highlighted by the artificial shading), most likely due to the initial data sampling in different directions on the grid. In contrast, real data would have other peculiarities that would result in similar residual mismatches. One way to further reduce the influence of these residual COE is to distribute the remaining mismatch along the tracks between the intersection points (e.g., Mittal, 1984; Hsu, 1995). I use x2sys_report with the correction table to determine spline adjustments for each track, and finally use x2sys_datalist again to give us the final corrected and adjusted track data.

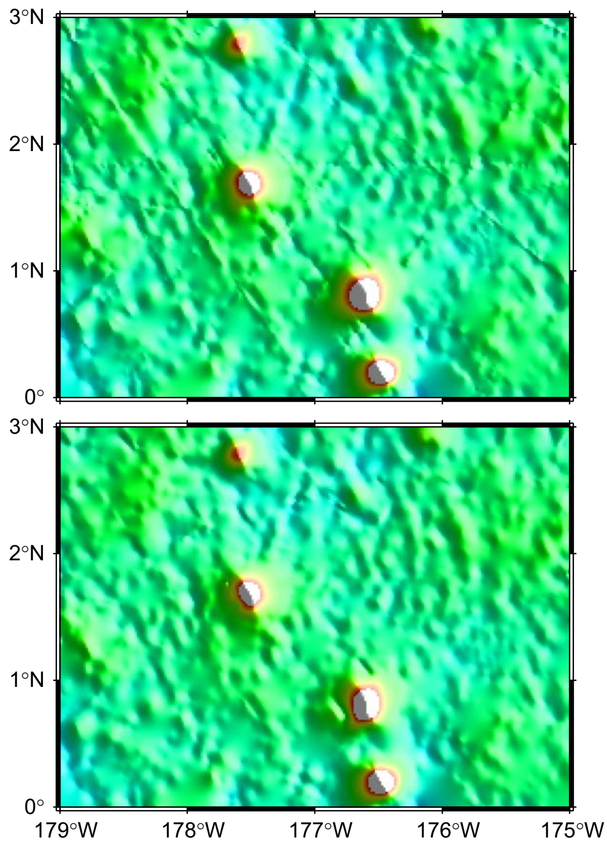

Fig. 4. (a) Using a least-squares method to determine best linear trends simultaneously for all tracks, I remove these trends and regrid corrected data. Most artifacts are gone but some remain. COE mean and standard deviation are now 0.02 and 4.21 mGal, respectively. (b) By distributing remaining COE along-tracks between intersections, I am able to almost completely remove any remaining artifacts as well as eliminate COE discrepancies.

Gridding now yields a further improved image (Fig. 4b), which only has a few minor artifacts remaining.

In this example, I could easily check the performance of the procedures since I knew exactly what the data should look like (Fig. 1). For instance, a comparison of misfit between the ground truth (Fig. 1) and the grids derived from tracks gave standard deviations of 2.2 mGal (Fig. 2b), 8.3 mGal (Fig. 3), and 3.2 mGal (Fig. 4a). The dramatic improvements offered by the crossover analysis package should give us increased confidence that the package will be very useful when processing new data as well. However, we note that the procedure of distributing residual COE along the tracks between intersections (e.g., Mittal, 1984; Hsu, 1995) may indeed result in a visually more pleasing image (Fig. 4b), but the slight degradation in standard deviation to 3.7 mGal indicates that this last improvement may come at a small price.

4. Discussion

There are numerous sources of error that can afflict along-track measurements, and many of them are not of a systematic nature that can be modeled using the linear form of (1). For instance, a tare in the marine gravity observations introduces a step-function at the time when the tare occurs. Unless that time is known *a priori* the step-function represents a nonlinear model. Even more difficult are problems that cause the recorded data to temporarily become unreliable. Examples include the effects of power outages, magnetic storms, failure of gyros, and other infrequently occurring circumstances that clearly are difficult to address with a linear model. Hopefully, such problem sections can be removed during the initial data processing step or discovered during an along-track quality control procedure (e.g., Chandler and Wessel, 2008). Alternatively, analyzing COEs versus time for one problem track can reveal such problems (e.g., Wessel and Watts, 1988) and steps may be taken (e.g., remove the bad data) before linear correction terms are determined. In general, scientists should always perform exploratory data analysis to determine unusual or problematic data sections.

As derived herein, all COE along a track contribute equally to the solution (assuming track weights are held fixed). However, this may give too much influence to large COE that may simply result from the nonlinear effect of positioning errors in areas of large data gradients (e.g., Schmitt et al., 2008). It may be worthwhile to explore more complex weighting schemes in which individual COE are weighted in inverse proportion to the local data gradients.

Acknowledgments

This work was supported by Grant OCE-0623409 from the US National Science Foundation. I thank Dan Scheirer, Neil Mitchell, Seung-Sep Kim, and Michael Chandler for helpful reviews. This is SOEST contribution no. 7828.

References

Akima, H., 1972. Interpolation and smooth curve fitting based on local procedures. Communications of the ACM (Association for Computing Machinery) 15, 914–918.

Buchanan, S.K., Scrutton, R.A., Edwards, R.A., Whitmarsh, R.B., 1996. Marine magnetic data processing in equatorial regions off Ghana. Geophysical Journal International 125, 123–131.

Bullard, E.C., Mason, R.G., 1961. The magnetic field astern of a ship. Deep-Sea Research 8, 20–27.

Caress, D.W., Chayes, D.N., 1996. Improved processing of Hydrosweep DS multibeam data on the R/V Maurice Ewing. Marine Geophysical Researches 18, 631–650.

Chandler, M.T., Wessel, P., 2008. Improving the quality of marine geophysical track line data, along-track analysis. Journal of Geophysical Research 113, B02102, doi:10.1029/2007JB005051.

Foster, M.R., Jines, W.R., Van der Weg, K., 1970. Statistical estimation of systematic errors at intersections of lines of aeromagnetic survey data. Journal of Geophysical Research 75 (8), 1507–1511.

Fox, C.G., Chadwick, W.W., Embley, R.W., 1992. Detection of changes in ridge-crest morphology using repeated multibeam sonar surveys. Journal of Geophysical Research 97, 11,149–11,162.

Ghosh, N., Hall, S.A., 1992. Crossover analysis of magnetic data over the Colombian Basin, western Caribbean Sea. In: Arora, B.R. (Ed.), Geomagnetic Studies at Low Latitudes, Memoirs of the Geological Society of India, vol. 24. Geological Society of India, Bangalore, India, pp. 93–101.

Grant, F.S., West, G.F., 1965. In: Interpretation Theory in Applied Geophysics. McGraw-Hill, New York 583 pp.

Green, A.A., 1987. Leveling airborne gamma-radiation data using between-channel correlation information. Geophysics 52, 1557–1562.

Hittelman, A.M., Groman, R.C., Haworth, R.T., Holcombe, T.L., McHendrie, G., Smith, S.M. 1977. The marine geophysical data exchange format – "MGD77". In: Key to Geophysical Records Documentation 10 (revised). National Geophysical Data Center, NOAA, Boulder, CO. http://www.ngdc.noaa.gov/mgg/dat/geodas/docs/mgd77.htm.

Hsu, S.-K., 1995. XCORR: a cross-over technique to adjust track data. Computers and Geosciences 21 (2), 259–271.

Huang, H., 2008. Airborne geophysical data leveling based on line-to-line correlations. Geophysics 73, F83–F89.

Huang, H., Fraser, D.C., 1999. Airborne resistivity data leveling. Geophysics 64, 378–385.

Kellogg, W.C., 1979. Discussion on: diurnal drift removal from aeromagnetic data using least squares, by H.L. Yarger, R.R. Robertson, R.L. Wentland (Geophysics 48, 1148–1156). Geophysics 44 (9), 1592–1593.

Klokocnik, J., Lemoine, F.G., Kostelecky, J., 1998. Reduction of crossover errors in the earth gravity model (EGM) 96. Marine Geodesy 21, 219–239.

Mauring, E., Kihle, O., 2006. Leveling aeromagnetic data using a moving differential median filter. Geophysics 71 (1), L5–L11.

Menke, W., 1984. In: Geophysical Data Analysis: Discrete Inverse Theory. Academic Press, Orlando, Florida 260 pp.
Minty, B.R.S., 1991. Simple micro-leveling for aeromagnetic data. Exploration Geophysics 22, 591–592.
Mittal, P.K., 1984. Algorithm for error adjustment of potential field data along a survey network. Geophysics 49 (4), 467–469.
Neumann, G.A., Rowlands, D.D., Lemoine, F.G., Smith, D.E., Zuber, M.T., 2001. Crossover analysis of Mars Orbiter Laser Altimeter data. Journal of Geophysical Research 106, 23,753–23,768.
Nishimura, C.E., Forsyth, D.W., 1988. Improvements in navigation using SeaBeam crossing errors. Marine Geophysical Researches 9, 333–352.
Prince, R.A., Forsyth, D.W., 1984. A simple objective method for minimizing crossover errors in marine gravity data. Geophysics 44, 1079–1083.
Sandwell, D.T., Smith, W.H.F., 2009. Global marine gravity from retracked geosat and ERS-1 altimetry: ridge segmentation versus spreading rate. Journal of Geophysical Research 114, B01411, doi:10.1029/2008JB006008.
Schmitt, T., Mitchell, N.C., Ramsay, A.T.S., 2008. Characterizing uncertainties for quantifying bathymetry change between time-separated multibeam echo-sounder surveys. Continental Shelf Research 28, 1166–1176.
Sedgewick, R., 1990. In: Algorithms in C. Addison-Wesley, Reading, MA 657 pp.
Smith, W.H.F., 1993. On the accuracy of digital bathymetric data. Journal of Geophysical Research 98 (B6), 9591–9603.
Smith, W.H.F., Wessel, P., 1990. Gridding with continuous curvature splines in tension. Geophysics 55 (3), 293–305.
Swan, D., Young, I.F., 1978. Crossover differences as an indication of survey accuracy. Marine Geophysical Researches 3, 393–400.
Tai, C.-K., 1988. Geosat crossover analysis in the tropical Pacific 1. Constrained sinusoidal crossover adjustment. Journal of Geophysical Research 93 (C9), 10,621–10,629.
Wessel, P., 1989. XOVER: a cross-over error detector for track data. Computers and Geosciences 15 (3), 333–346.
Wessel, P., Chandler, M.T., 2007. The mgd77 supplement to the Generic mapping tools. Computers and Geosciences 33 (1), 62–75.
Wessel, P., Smith, W.H.F., 1991. Free software helps map and display data. Eos Transactions, AGU (American Geophysical Union) 72 (41), 441.
Wessel, P., Smith, W.H.F., 1995. New version of the Generic Mapping Tools released. Eos Transactions, AGU (American Geophysical Union) 76 (33), 329.
Wessel, P., Smith, W.H.F., 1998. New, improved version of generic mapping tools released. Eos Transactions, AGU (American Geophysical Union) 79 (47), 579.
Wessel, P., Watts, A.B., 1988. On the accuracy of marine gravity measurements. Journal of Geophysical Research 93, 393–413.
Yarger, H.L., Robertson, R.R., Wentland, R.L., 1978. Diurnal drift removal from aeromagnetic data using least squares. Geophysics 43 (6), 1148–1156.