GEOLOGICAL SURVEY OF CANADA

OPEN FILE xxxx

NavCleaner – An interactive program for cleaning navigation derived from marine expedition data including SEGY sources

R. C. Courtney

2016

Natural Resources Canada Ressources naturelles Canada

Canada

GEOLOGICAL SURVEY OF CANADA

OPEN FILE xxx

NavCleaner – An interactive program for cleaning navigation derived from marine expedition data including SEGY sources

R. C. Courtney

2016

Available from

Geological Survey of Canada

601 Booth Street

Ottawa, Ontario K1A 0E8

R. C. Courtney

2016: **NavCleaner -** An interactive program for cleaning navigation derived from marine expedition data including SEGY sources, Geological Survey of Canada, Open File xxxx.

Open files are products that have not gone through the GSC formal publication process.

NavCleaner – An interactive program for cleaning navigation derived from marine expedition data including SEGY sources

Abstract

NavCleaner is a Microsoft Windows 7 program that was developed to efficiently examine and correct navigational data collected during marine expeditions of the Geological Survey of Canada. Two types of data streams are considered here: (1) navigational position data derived from GPS receivers logged on research vessels and (2) navigational data embedded in SEGY (Norris and Faichney, 2002) data collected during high resolution seismic surveys conducted during marine expeditions.

Since many high-resolution seismic systems emit an impulsive sound at a repetition rate based on two-way travel time (the time for a sound wave to travel from the sounder to the seabed and back) , the records derived from these sounders are often biased to oversample in shallow waters when compared to number of soundings made in deeper water.

This program offers a method to resample the seismic section with a prescribed constant shot distance spacing. This approach uses the Douglas-Peucker algorithm (DOUGLAS and PEUCKER, 1973) to simplify the survey vessel's track, reducing the effects of GPS and other sources of positional error that contributes to errors in the estimate of cumulative distance along track.

The program was written in C# using Microsoft Visual Studio 2013. All source code is freely available for any use. A compiled deployment version is included in this release which can be used and redistributed without restriction.

Keywords: seismic data, navigation, SEG-Y, C#, Microsoft Windows 7

Table of Contents

Introduction

The GSC has been collecting digital seismic data with embedded navigational data since the early 1990's and has used and continues to use SEGY (Norris and Faichney, 2002) as its primary format for storing its digital seismic, sounder and sidescan data. The GSC has also compiled navigation from hundreds of expeditions that are stored in databases and made available online (e.g., http://ed.gdr.nrcan.gc.ca/index_e.php).

The process of examining and correcting the positional data collected during these efforts is substantial and represents one of the more time-consuming tasks in the process of bringing these data from the field to our databases. Generally, an effort is made to quality control positional data from our expedition navigation loggers but the data in our SEGY data holdings are rarely even examined. The presence of "bad" positional data in these SEGY streams has consequences in processing and interpretation.

The GSC's high resolution inventory of digital SEGY data is considerable. But the installation of high-resolution sub-seabed sounders on Canadian Hydrographic Service vessels and their deployment during most of their multbeam bathymetric surveys over the last fifteen years, as well as similar seismic data derived from ArcticNet activities in the Canadian high latitudes, have accentuated the need to develop techniques to efficiently and effectively assess these data in order to use these data to augment our data inventory and regional seismic coverage.

Release Details

This release package contains 2 separate deployment packages:

(1) A version that can be installed in a local folder.

(2) More formal deployment that installs in C:\Program Files (x86) and puts entries in Start->Program menu and short-cut on Desktop.

Details

(1) A release that can be installed in local folder with no need for administrative permission: Unzip the file NavCleanerl.zip to a local folder. NavCleaner.exe is the executable image. Double-click on this item to proceed. Two sample file are included: 97009PGC.NAV – a sample ASCII file and (2) test3_246_1408_to_246_1909.sgy a sample SEG-Y file.

(2) A Windows 7 install package that needs administrative privileges: Unzip the NavCleanerInstaller.zip file and follow standard installation procedures. Contains the same files as (1) but packaged to created shortcuts and menus

Mainfest of Release:

Name	type	comment
NavCleaner.exe	executable	Main program
Converters.dll	dll	Convert to IBM fp to IEEE fp
DouglasPeucker.dll	dll	Line thinning library
Geographic.dll	dll	GeographicLib library
msvcp110.dll	dll	Microsoft C++ redistributable package
msvcr110.dll	dll	Microsoft C++ runtime redistributable package
NETGeographic.dll	dll	.NET wrappers for GeographicLib
Oracle.ManagedDataAccess.dll	dll	Oracle ODP.NET database operations
SEGYlib.dll	dll	SEGY library
TeeChart.dll	dll	Graphics library
97009PGC.NAV	txt	Sample ASCII nav file
test3_246_1408_to_246_1909.sgy	SEGY	Sample SEGY file
NavCleaner.docx	docx	This document
Readme.txt	txt	
NavCleanerInstall.zip	zip	Windows 7 Installer package

Interface

The interface allows the user to load ASCII navigational data or SEG-Y Data. Within NRCan, the user can also access the Expedition Database to examine preloaded data or to merge these data into SGEY files.

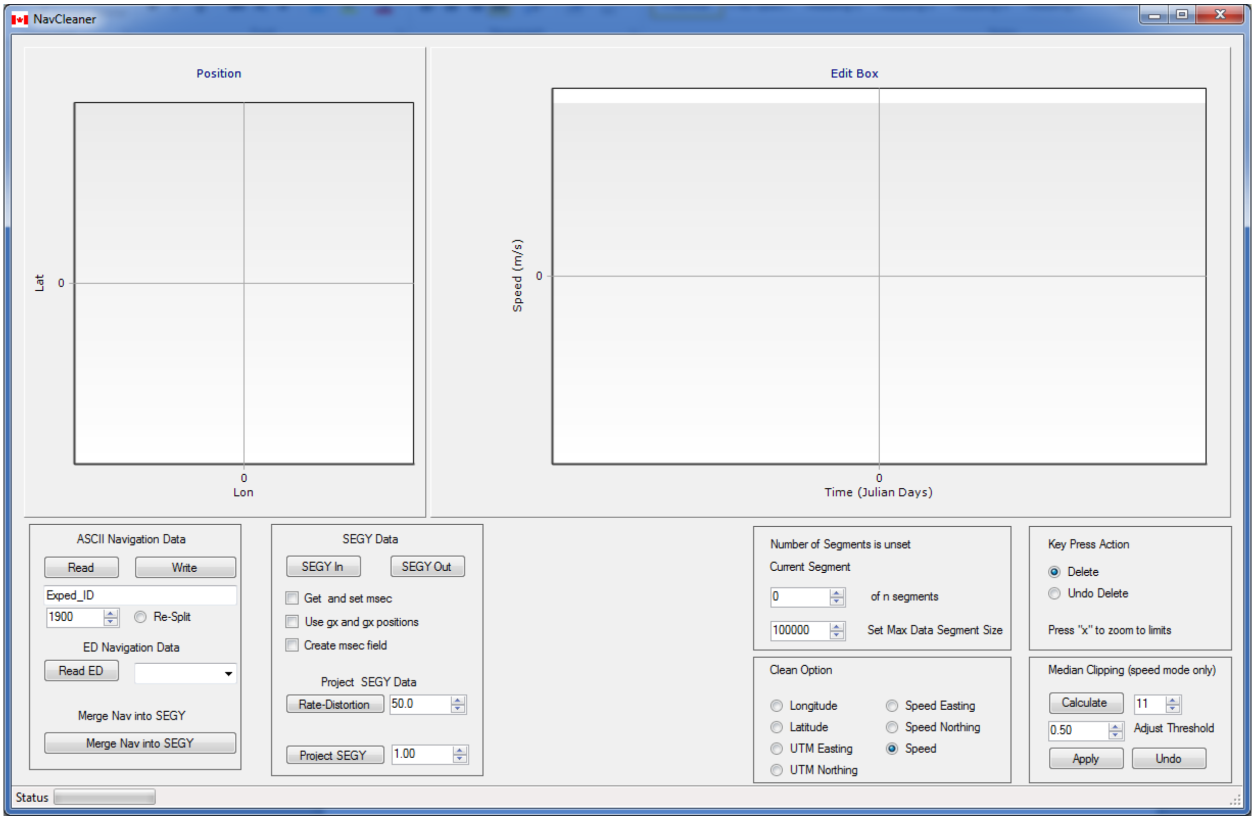

Fig. NavCleaner interface

Launch the NavCleaner interface by choosing Start > All Programs > NRCan > NavCleaner from the start menu (Fig. 1).

ASCII Data

Format of ASCII Navigation Data

The format of ASCII navigational data, compliant with this tool, is a form used regularly at the GSC Atlantic and GSC Pacific. The file format follows a sequence :

Expedition_id   number_of_points_in_following_segment

Coded_time   decimal_latitude   decimal_longitude

Coded_time   decimal_latitude   decimal_longitude

.

.

Coded_time   decimal_latitude   decimal_longitude

Where the coded time follows the format DDDHHMMSS where DDD is the Julian Day, HH is the hour of the day, MM the minute and SS the second of the positional fix. The header line is optional and NavCleaner will segment any navigational stream into discrete segments. The sequence of the Expedition_id and the following lines can be repeated for any number of segments. NavCleaner will calculate new segments breaks regardless of the breaks speficied in the source file

For example,

97009PGC   63845

272202513 48.656485 -123.489080

272202518 48.656503 -123.489432

272202523 48.656523 -123.489810

272202528 48.656550 -123.490218

272202533 48.656573 -123.490618

272202538 48.656600 -123.490995

272202543 48.656610 -123.491397

Loading ASCII Data

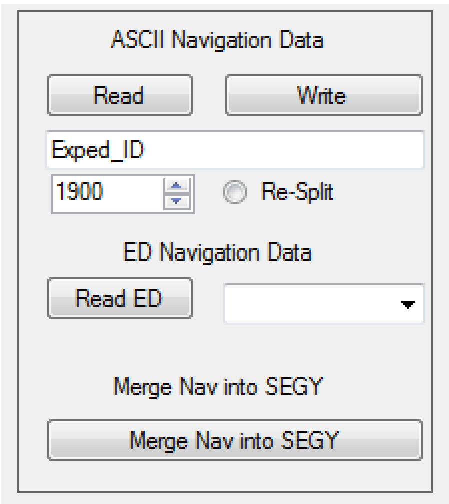

Fig.2 ASCII Navigation Data I/O panel

Before loading ASCII data into the application, first specify the year of the expedition in the numeric up/down entry box (set to 1900 by default) since this information is not included in the input ASCII navigation file. Also enter the Expedition id in the text box located immediately underneath the “Read” button.

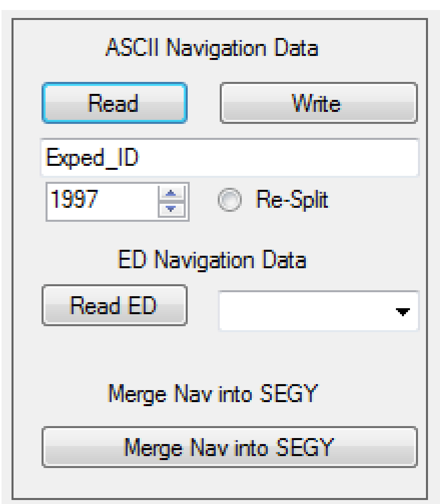

Fig. 3 Entries in the ASCII Navigation panel

Next press the”Read” button and choose the ASCII file to be considered. The positional data will be read and the navigation data are scanned and breaks in the time sequence in the navigation stream are identified. Use the “Re-Split” button to ignore any pre-existing line split in the input file, to regenerate completely new line separations. The navigational data is then displayed in the graphic plotting boxes in the upper half of the application window (Fig. 4).

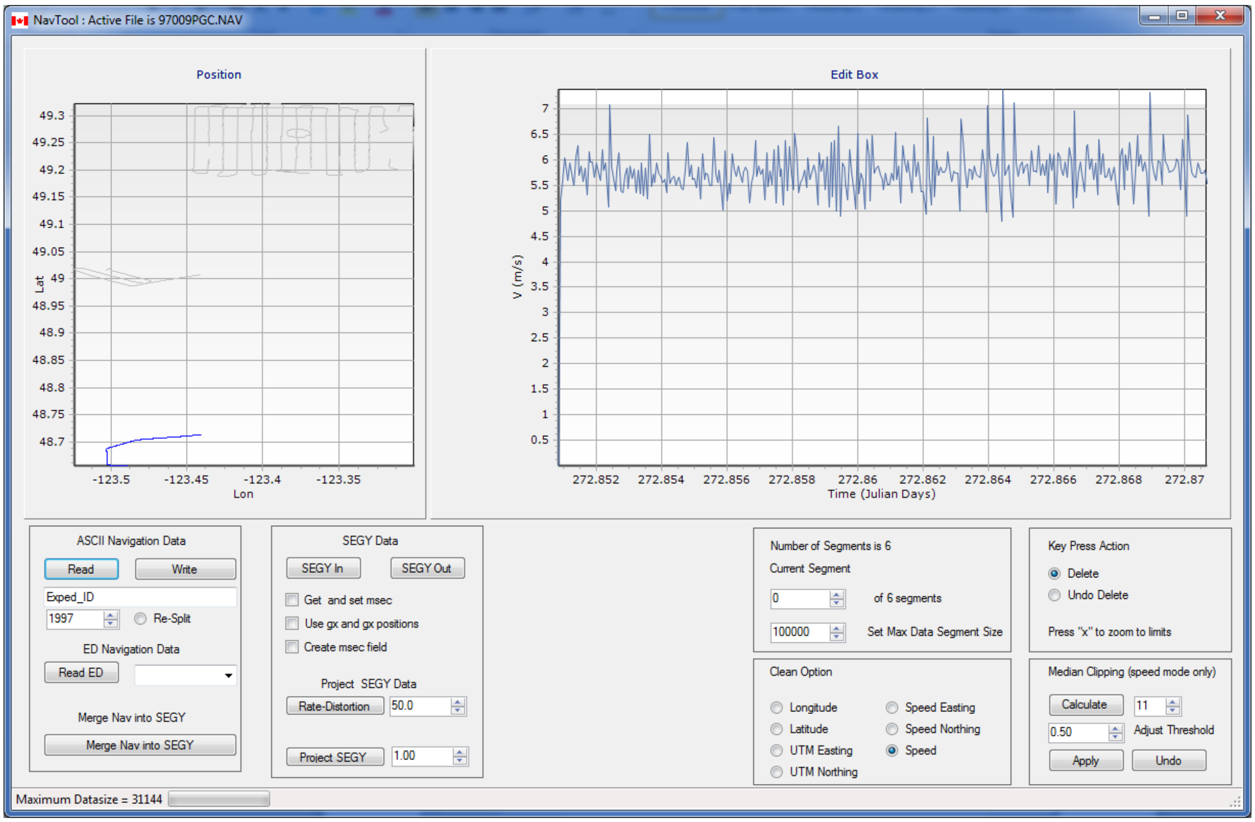

Fig. 4 ASCII Navigation data read into the system.

The left hand graphic displays all the segments in an x/y plot with the active segment highlighted in blue. The right-hand plot displays one of the attributes (longitude, latitude, x, y, speed in the x direction, speed in the y direction, and absolute speed) as a function of time displayed in units of Julian Days. The aspect ratio of the left hand plot can be adjusted by dragging the boundary between the two plots with the mouse button 1 depressed.

Note: The x and y “UTM” coordinates are calculated using a WGS84 Transverse Mercator projection with the central meridian set to the longitude of the 1st point in the navigation input series.

Editing ASCII Data

The right hand graph is used to edit points in the input data stream. As the navigation data is divided into segments, the current segment can be chosen using the numeric up/down box as shown in Fig 5.

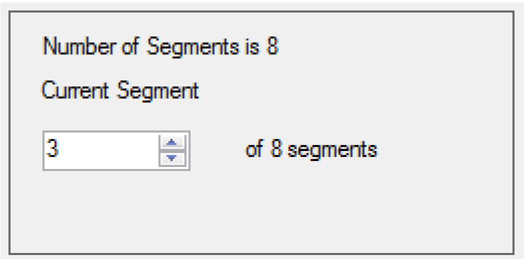

Fig.5 Segment choice panel

Since the segment number counts up from 0, the maximum segment number in Fig. 5 is 7.

Use the radio buttons in the Clean Option panel (Fig. 6) to choose the attribute to edit.

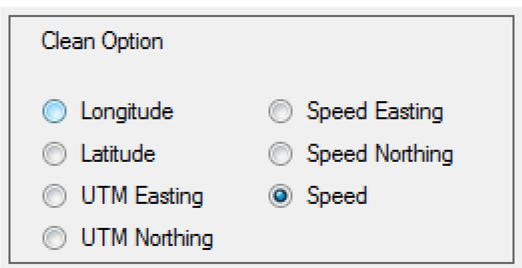

Fig. 6 Clean Option panel

By default, the “Speed” option is preselected and, probably, will be the option most usually employed in cleaning data. The speed of the research vehicles is usually constrained to the interval [0, 10 m/s] so estimates outside this range often arise from erroneous positional estimates. Some variability in the speed estimate is normal. Since the speed is calculated by the change in position between successive fixes, GPS and other sources of positional error or true short-term variation in position (running to a sea for example) give rise to variations in the speed estimate often in the order of 2-4 m/s. (Fig. 7).

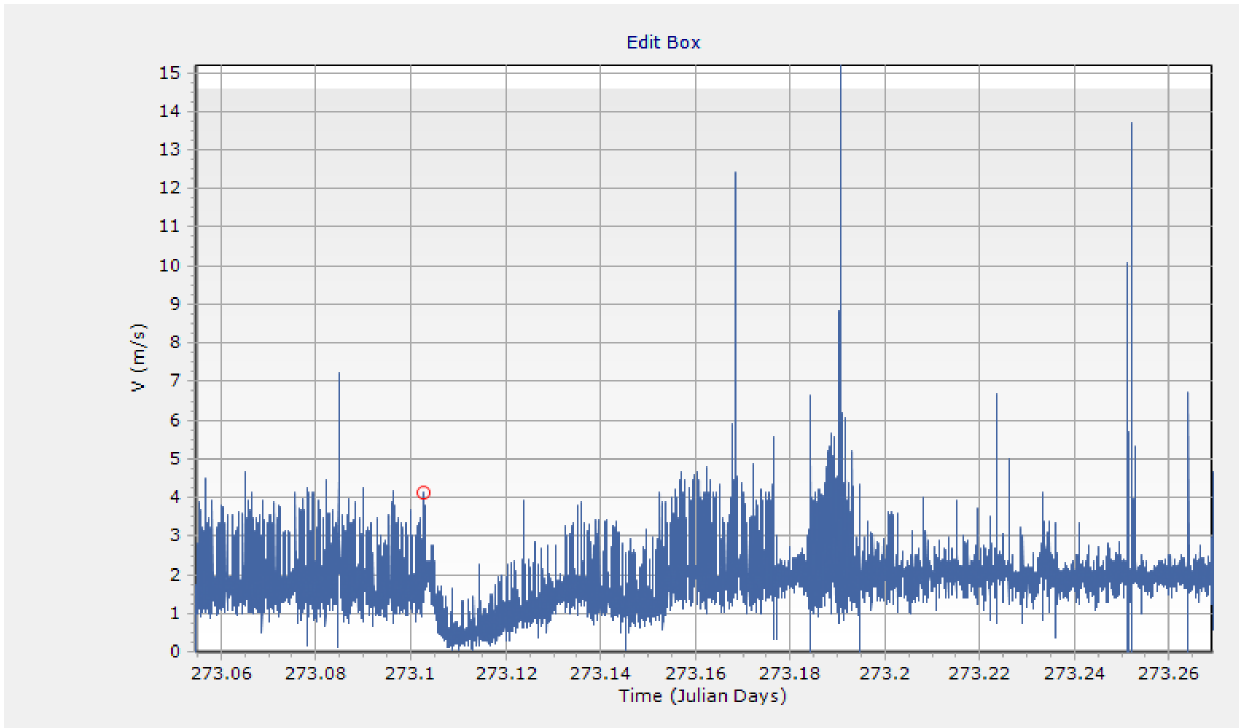

Fig. 7 Speed estimates as a function or time in segment

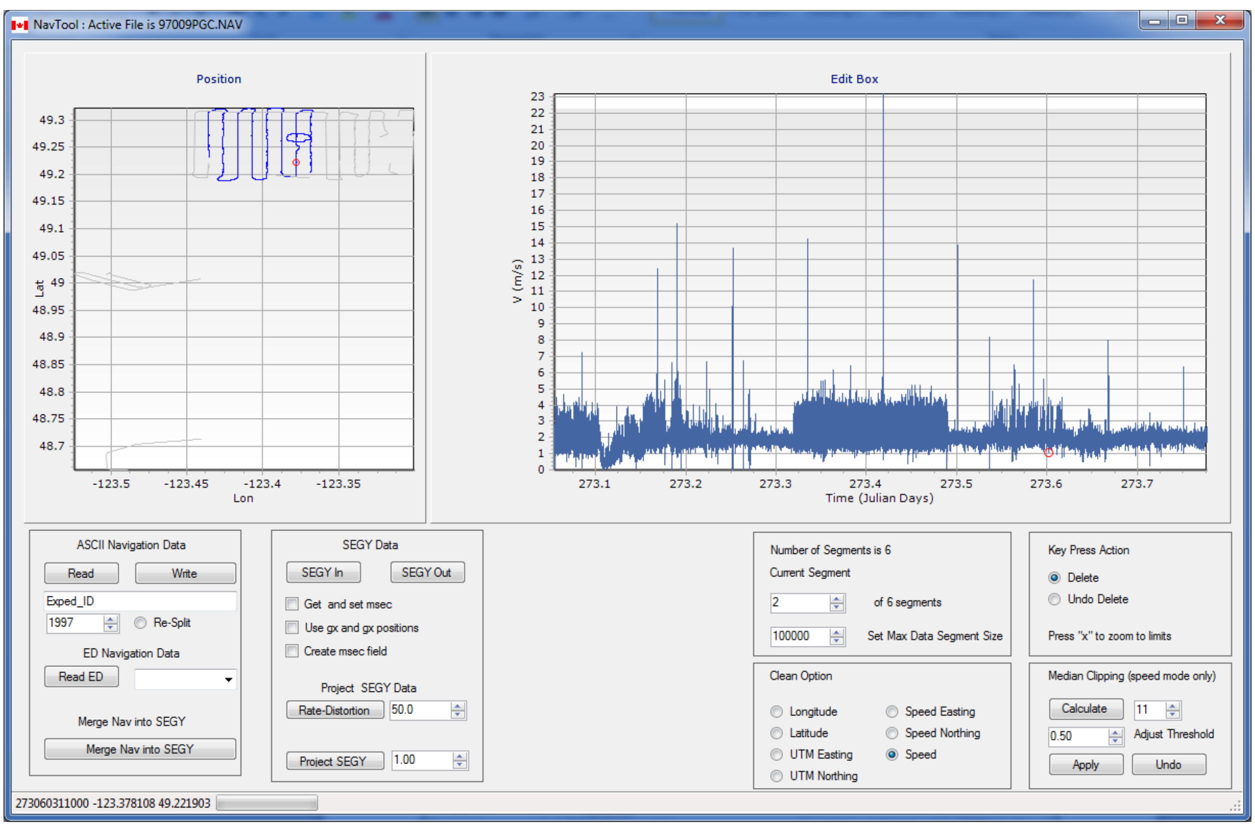

Fig. 8 Selection of a point in the edit box.

Use the mouse pointer and click on any points on the segment using the felt mouse button (Fig. 8). The selected point will be marked with a red circle in both the edit panel and the adjacent position panel. The values associated with the point are also displayed in the lower left-hand corner status

Use the mouse pointer to click on a point in the Edit Box panel. This will ensure the focus will be on that graph. The action of subsequent keyboard actions is determined by the choice of "Delete" or "Undo Delete" mode in the Key Press Action panel (Fig. 9) :

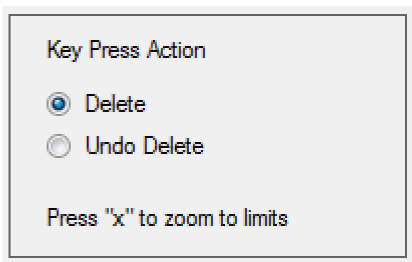

Fig.9 Key Press Action panel

This panel determines the keyboard key press action. If "Delete" is chosen, the any key press (other than "x") will delete the selected point. The space bar is a convenient choice for deleting points.

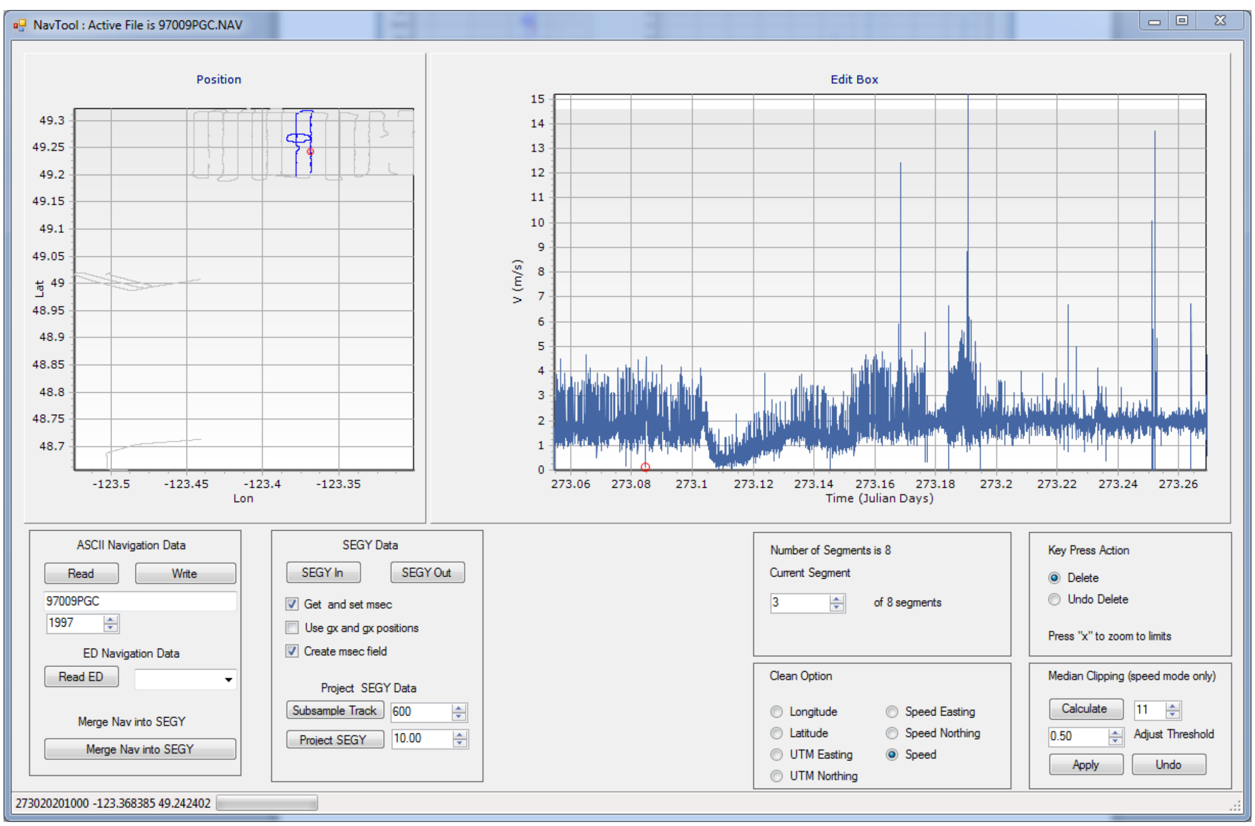

Fig. 10 Deleting a point

The selected point is deleted and is moved to a deleted points buffer. It is also displayed as a red "x" in both graphic plots. Zooming in on the deleted point and examining the nature of the point using other Clean Option attributes allows the nature of the aberrant value to be determined. In the example shown in Fig. 11, a sequence of successive values was deleted as the GPS signal appears to have jumped by around 8 meters east.

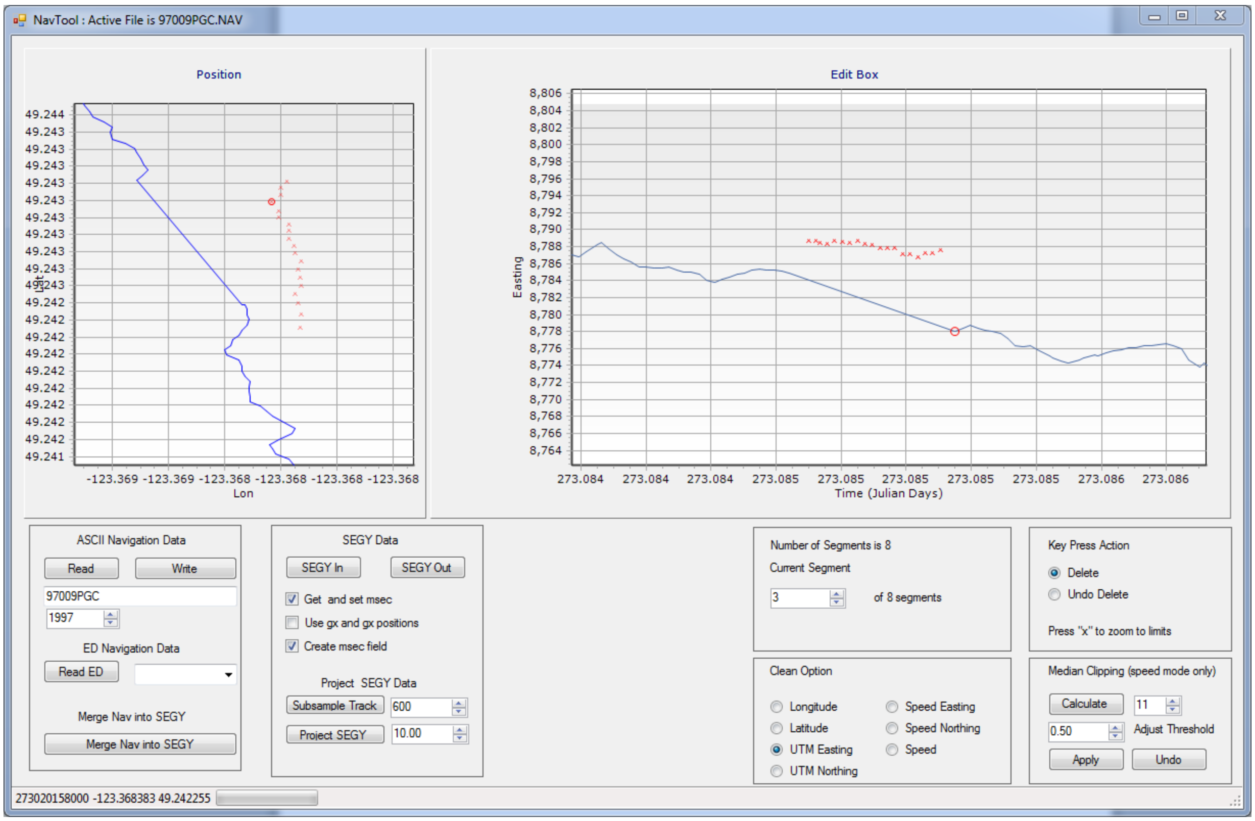

Fig. 11 Deleting a navigation jump

Note : To zoom-in in either plot, depress the left mouse button on the upper left hand corner of the area to be viewed and, keeping the left mouse button depressed, drag out a bounding rectangle to the lower right hand corner of the desired area. To zoom back to the default limits, press the key "x". The plot can be panned, by pressing the right mouse button down and dragging the point within the plot boundary.

To restore a deleted point, choose the "Undo Delete" option in the Key Press Action panel. Move to the desired deleted point (red x) to restore. The selected point should be designated with a red circle. Then press any key to restore the fix.

Automated Median Estimation of Aberrant Values

The Median Clipping panel can be used to automated data cleaning. This panel only operates when the speed option is chosen (Fig. 12).

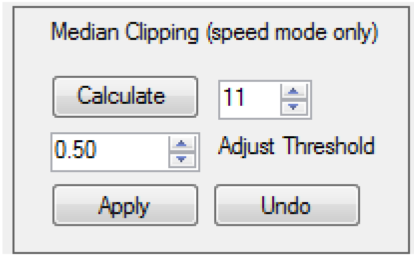

Fig. 12 – Median Clipping panel

Two entries determine the action of the median clipping filter: (1) the filter width (set to 11 points by default) set the number of neighbouring fix estimates used to calculate the median speed of the vessel and (2) the speed threshold (set to 0.5 m/s by default) sets the upper and lower bounds of acceptable vessel speeds with respect to the median estimate.

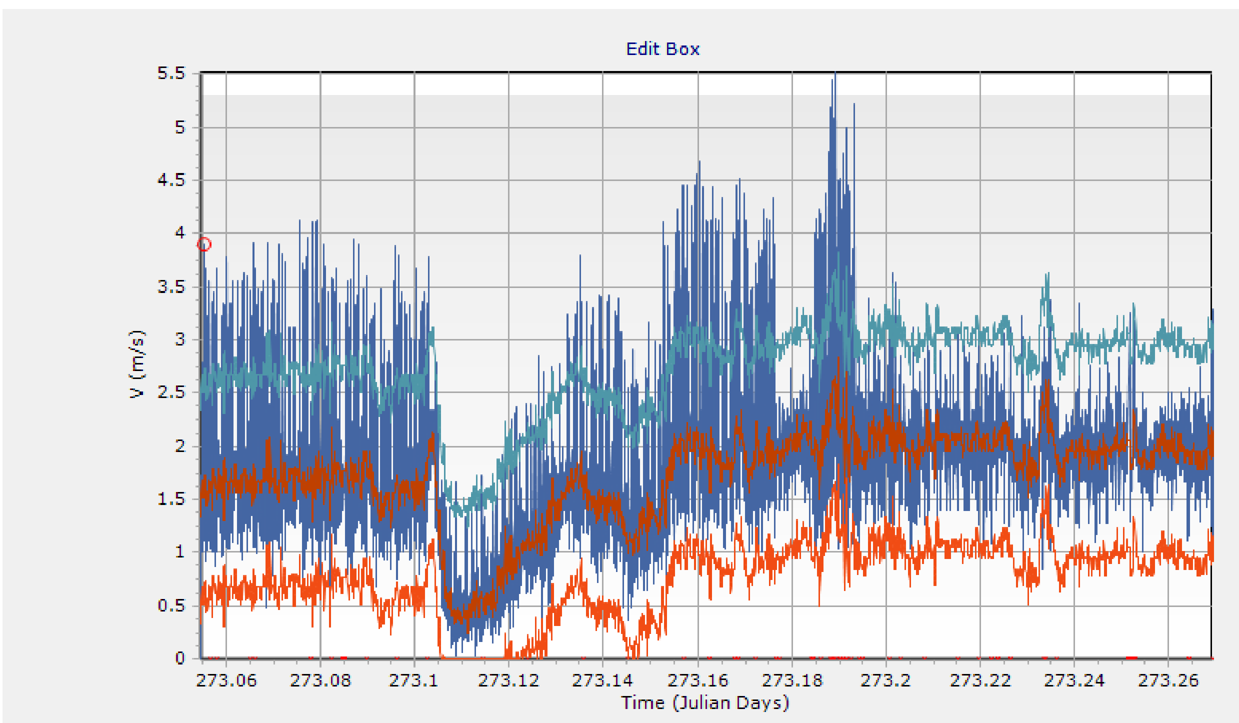

Fig. 13 Speed threshold set to 1.0

Press “Calculate” to estimate the upper and lower and median values (Fig. 13). Then press “Apply” to apply the filter (Fig. 14). “Undo” will reset the entire segment, removing all edited points. This filter can be very aggressive and remove a lot of points.

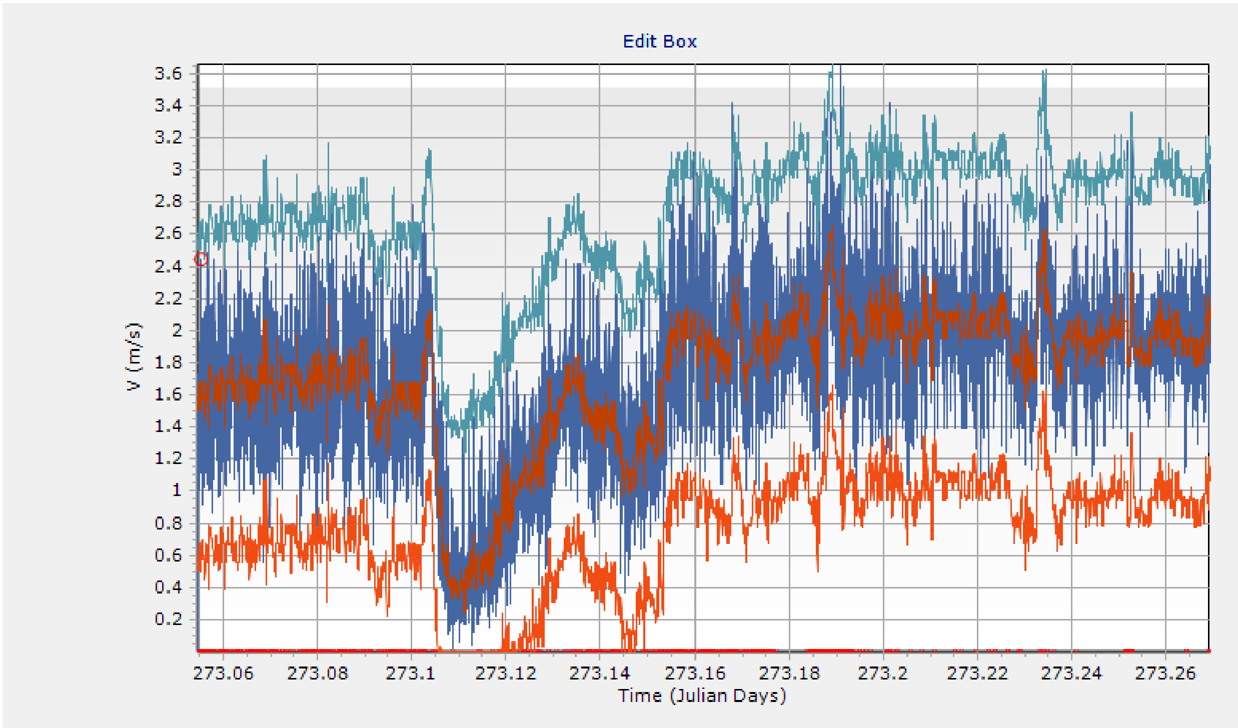

Fig. 14 Median filter applied to segment

Save Cleaned ASCII Data

The cleaned navigation file is written to a user chosen folder with the name <Expedition ID>_cln.nav, in this example, 97009PGC_cln.nav. Once the program is closed or a new data set is load, all edits and deleted points are discarded and cannot be restored.

Loading Expedition Database Navigation

For those within the NRCan / RNCan network, navigation stored in the Expedition Database (ED) can be accessed using the “Read ED” buttons in the ASCII Navigation panel (Fig. 2). The middle combination box can be used to choose the desired expedition from a list loaded at program start-up from ED. If the list of expeditions is not there, then the linkage to the database was unsuccessful.

This option may be used to load and examine navigational data that has been loaded into ED. An ASCII file containing these data can be generated using the “Write” button in the ASCII Navigation panel.

Merge Navigation into SEGY

The lower button is Fig. 2, labeled “Merge Nav into SEGY” can be used to merge the active cleaned navigation data set, whether it was loaded from an ASCII source or downloaded from ED, into a SEG-Y file. This option assumes that the each trace in the SEGY file is time-tagged and, thereby, linkable to the navigation source.

SEGY Data

Effect of GPS Noise on Cumulative Distance Estimation

SEGY formatted data is commonly used to store single channel and multichannel seismic data. At the Geological Survey of Canada, seismic, sounder and sidescan data has been recorded in this format since 1992. These data sets contain navigational data of varying degrees of quality. In the early 1990's, GPS positional information was degraded with Selective Availability (SA) and, even when differential GPS corrections were available, the reception of these corrections was often irregular and contributed to a non-Gaussian noise component to the signal which cannot be estimated through simple averaging.

Even after SA was discontinued in May 2000, the accuracy limitations of GPS positions are not insignificant. At typical vessel survey speeds of 2.5 m/s, GPS positional inaccuracy is of the same order of the distance traversed by the survey vessel during each second. In shallower waters, shot repetition rates for high resolution sounders can be as frequent as 2 shots/second so clearly the relative positions of these shots is strongly biased by GPS and other sources of positional error. Consequently, estimates of the cumulative distance along the survey track calculated using the differences between successive shot positions can overestimate the true value by up to a factor of two.

This program offers a two-fold solution to the effect : (1) a procedure to edit out obvious outliers from the embedded navigation stream and (2) a procedure to project the SEGY data onto generalized straight line segments where shot to shot spacing can fixed to with centimeter accuracy.

Read SEGY

SEGY data containing time-referenced navigation data is handled using the SEGY data panel (Fig. 15). At present, this software does not process SEGY data that is not time referenced.

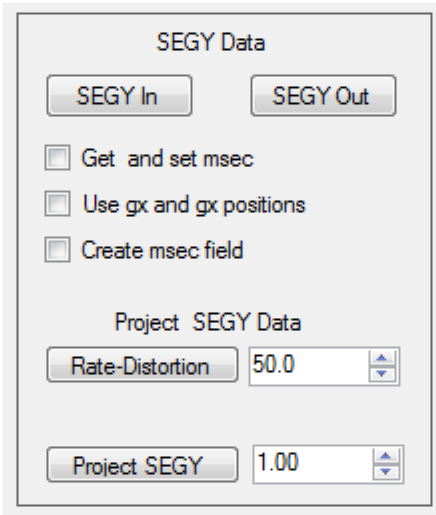

Fig. 15 SEGY Data panel

There are a couple of check boxes to consider before reading in the data.

(1) Get and set msec – Some SEGY files recorded at the GSC stored the millisecond field in the SGEY trace header using the time basis location (byte locations 166-167). Checking this box transcribes the millisecond value to the trace header Lag Time B position.

(2) Use gx and gy position - GSC digitizers often recorded UTM positions in the trace header source x and source y locations and put the corresponding latitude and longitude values in the group x (gx) and group y (gy) positions. Since the UTM zone is not recorded in these files, the gx and gy values can be used. The coordinates scaling of these latitude and longitude values was fixed at 1e6.

(3) Create msec field – Some SEGY files did not record the millisecond field. The program can estimate the missing fields and insert in each trace header at the Lag Time b position.

To read in a SEGY file, press the “SEGY In” button and using the popup widget to select the desired file. The program will apply the check box options and segment the navigation from the SEGY file into contiguous segments as with the ASCII navigation (Fig. 16).

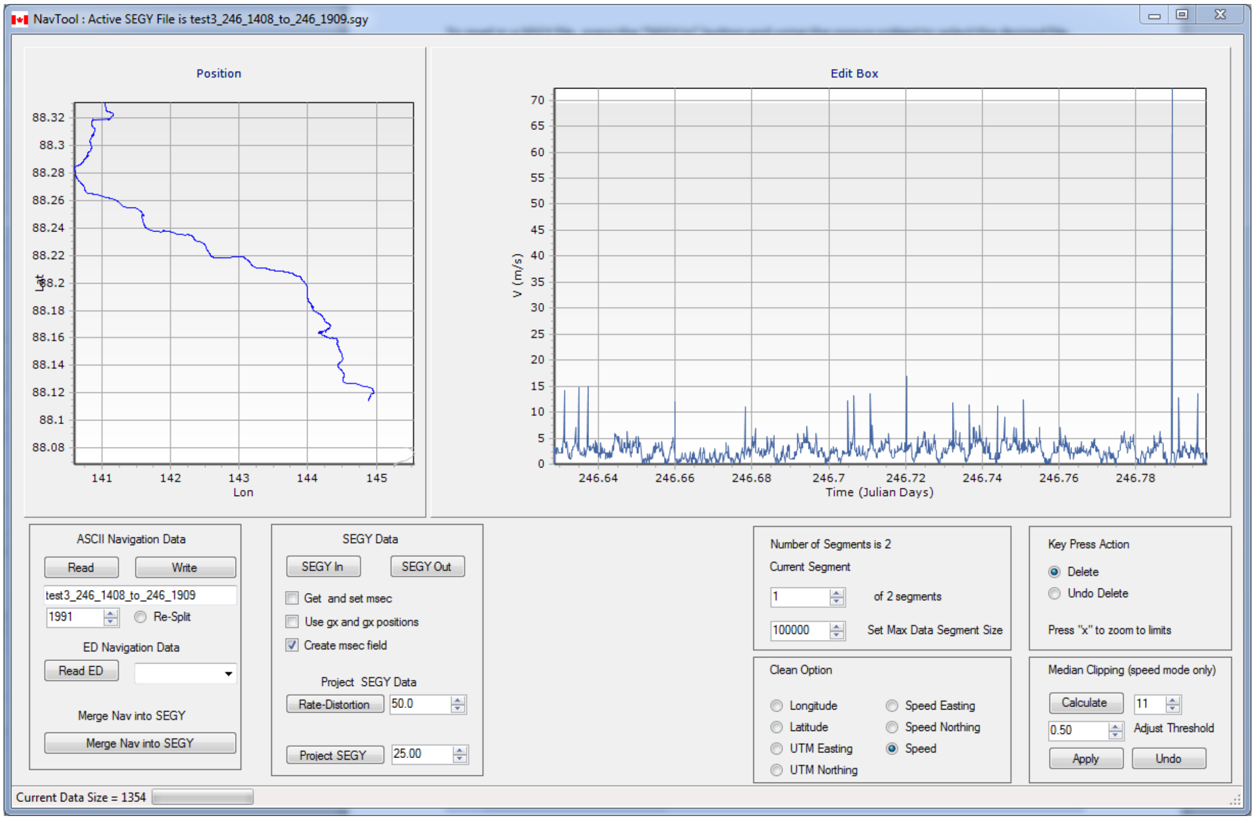

Fig. 16 SEGY navigation data

Edit SEGY Navigation Data

Use the same approach as was used with ASCII data stream to edit the navigation (Fig. 17) .

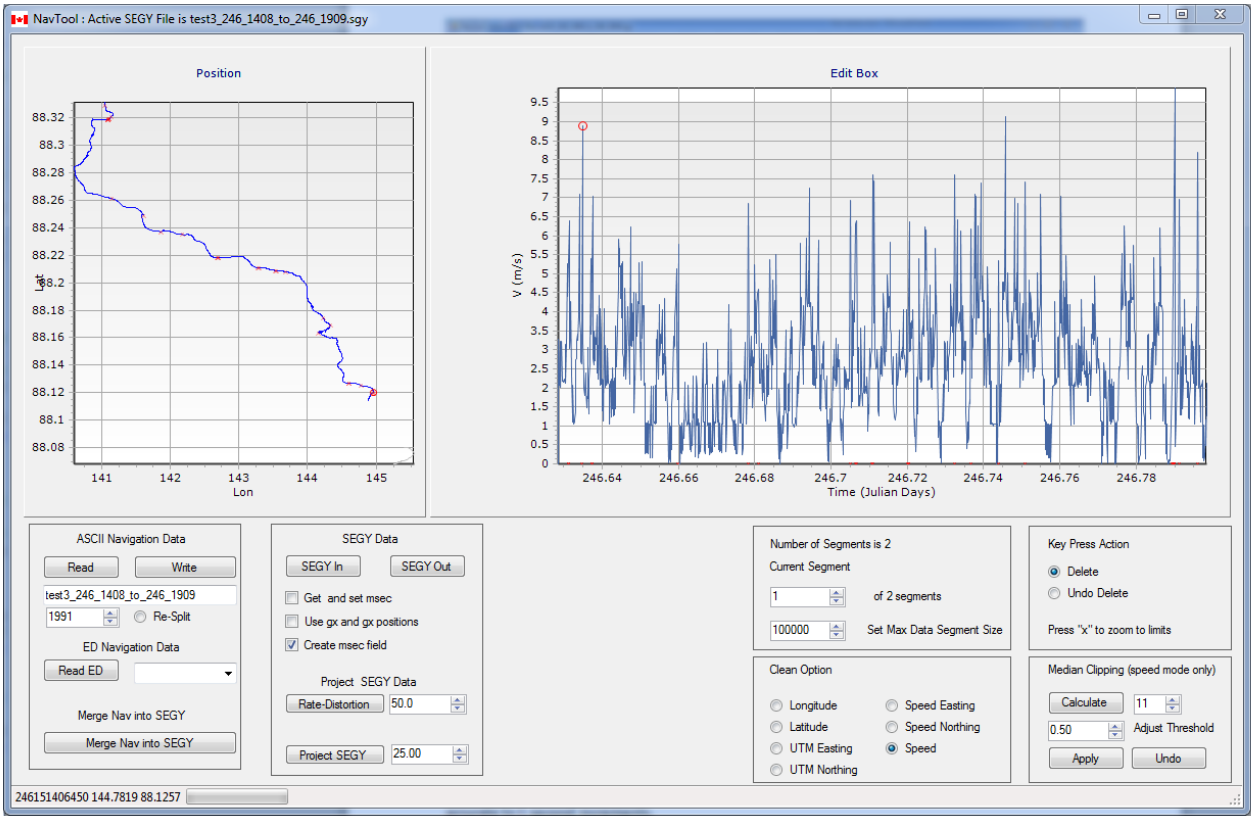

Fig. 17 Cleaned Navigation in SEGY file

Save Cleaned Navigation in SEGY

The cleaned navigation can be saved to a new SEGY file using linearly interpolated estimates for deleted values. Press “SEGY Out” to generate the SEGY file with cleaned estimates. In this example, the millisecond field has been also estimated and injected into the trace headers of the saved file. If the input file name is “TEST.sgy” then the saved file name will be “TESTcln.sgy”.

Save Cleaned Navigation from SEGY into ASCII format

The cleaned navigation can also be saved as ASCII data using the “Write” button on the ASCII Navigation Data panel. Since the ASCII format does not have a millisecond field, the time codes in this file are only accurate to 1 second increments.

Constant Shot Spacing SEGY Output

The recorded SEGY navigational track can be simplified using Douglas-Peucker reduction methods (DOUGLAS and PEUCKER, 1973) to minimize the effects of GPS and other noise in estimating the cumulative distance along track. In this method, a minimum set of straight line segments are used to approximate the track governed through the prescription of a maximum cross-track error threshold.

Fig. 18 Navigation (green) simplified with a cross track threshold of 100 m compare to original navigation (blue).

At larger values of the cross-track threshold, a complex track can be simplified to a few straight line segments (Fig. 19)

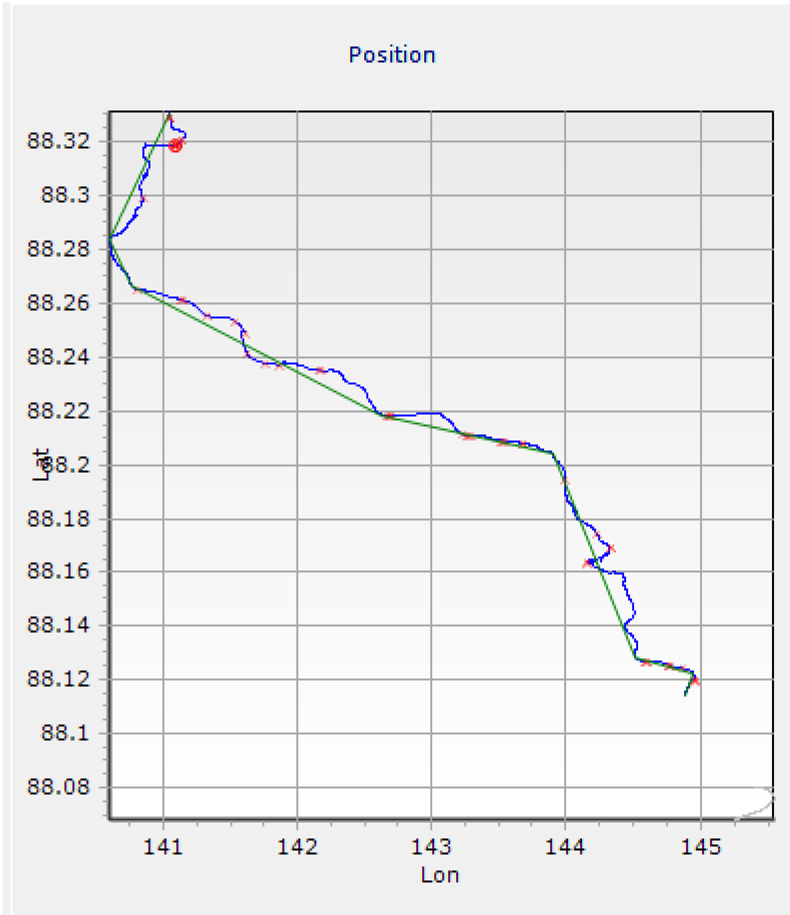

Fig. 19 Navigation (green) simplified with a cross track threshold of 750 m compared to original navigation (blue).

The input SEGY data can be mapped to these simplified tracks using a prescribed constant shot spacing where the shots on the new track are generated by using the closest trace perpendicular to point on the interpolation line. The remapped section will be not biased by shot-to-shot positional inaccuracy and would have a more representative measure of true distance along track. The degree to which the cross-track threshold can be increased to reduce line complexity depends, of course, on the cross-track variability of the geology of the seabed.

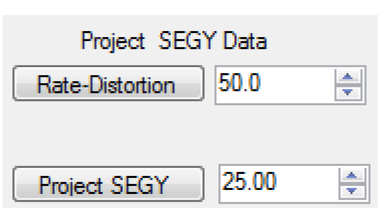

Fig. 20 Parameters for simplifying track.

When the source SEGY file is first read, the median shot separation is calculated and inserted into the lower numeric up/down box shown in Fig. 20. The cross-track threshold is set in the upper numeric up/down box shown in Fig. 20. It can be changed and the resultant track displayed using the “Subsample Track” button shown in Fig.20.

A new track-interpolated SEGY is generated through pressing “Rate-Distortion”. If the source file name is “Test.sgy” then the new SEGY file will, in this example, have the name “TEST_SC25.sgy” where the suffix “_SC25” designated speed corrected at 25 m shot spacing.

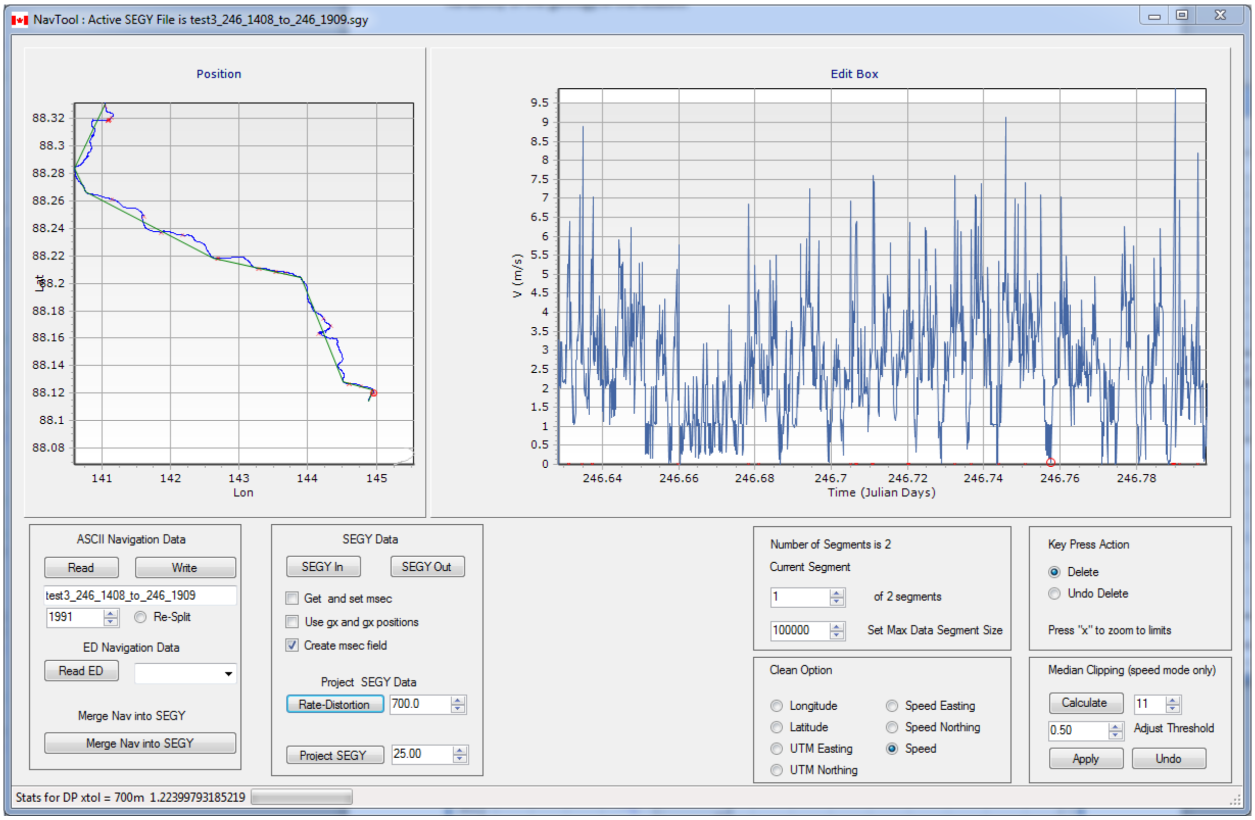

Fig. 20 Output corrected SEGY speed as a function of time

In the new file, speed varies in each straights line segment to ensure a 25 meter shot spacing. This example was derived from seismic data collected from an ice breaker where progress through the ice was variable.

To inform the user about the effects of simplifying the ship track, a secondary window pops up with rate-distortion information for the selected segment.:

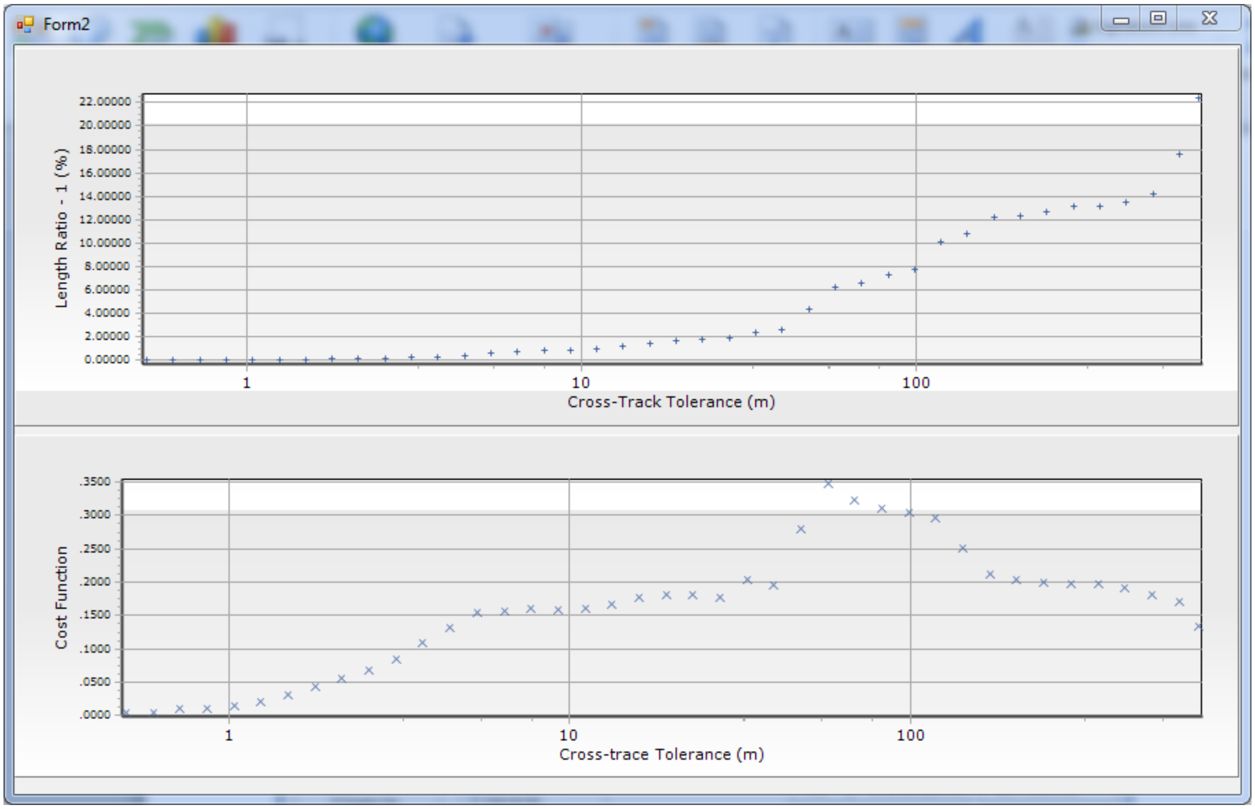

Fig. 21 Rate Distortion curves for track sub-sampling. The top graph shows the relative overall track shortening induced by the Douglas-Peucker track sub-sampling as a function of cross-track tolerance. The lower panel graph plots a cost function (the product of compression ratio times the relative track length ratio). For this example, there is little track distortion below a cross-track tolerance of 20-30 m while the cost function is maximized at cross-track values around 40 m.

For this example, a choice of a cross-track tolerance of 20 to 30 meters does not substantially shorten the total track length while a choice of 700m yields a track length 22% shorter than the recorded values. Arguments could be made that the survey vessel length could be a "natural" cross-track threshold or half the vessel length. Through this criterion on could "dial-in" the natural dynamics of the surveys vessel and its ability to move in a near- linear motion over this prescribed distance. For this example, the survey vessel was 128 m (USCGC Healy).

In ice-covered seas, as is the case for this data set, the ships's motion is highly affected by the process of breaking ice which includes multiple forward and backward manuveours. Since the seismic collection is often time based, the resulting record was highly affected by these artefacts, justifying the over-simplification of the ship track by a cross-track tolerance far in excess of the ship's length.

In Fig.22, the original SEGY (top) and the speed corrected SEGY data (bottom) are compared with the bottom axis showing the cumulative distance along track. The original SEGY shows the effect of variable shot frequency and ship speed as the horizontal axis label spacing varies. In addition, the original section's maximum cumulative distance exceeds that of the corrected section by circa 30%, attributable in part to positional inaccuracy and the reduction of the navigational path complexity.

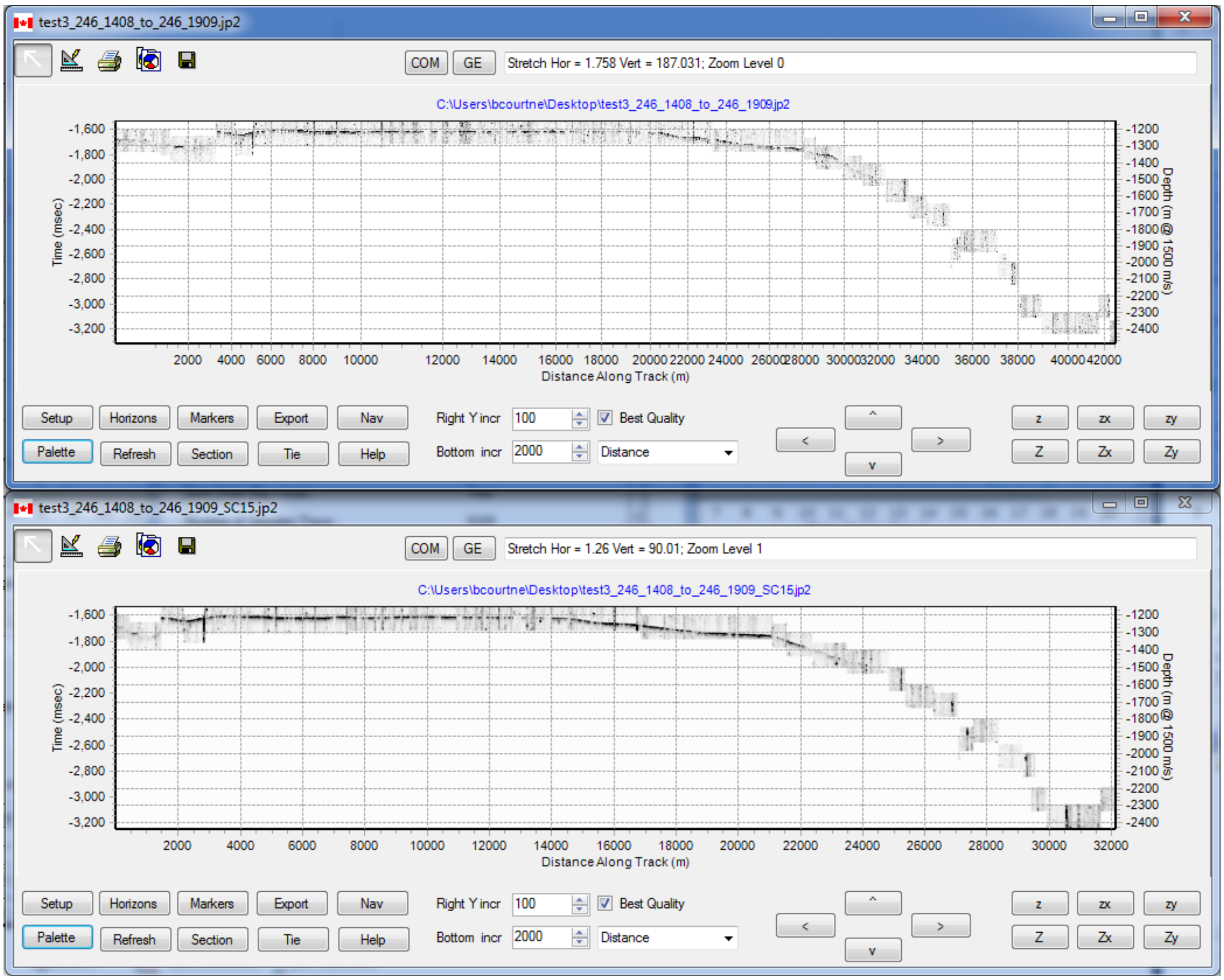

Fig. 22 Comparison between source SEGY and speed corrected SEGY for a 700 m cross-track tolerance. Note the maximum Distance Along Track in the speed-corrected strip has decreased by some 30% relative to the uncorrected original section. In addition the geometry of the shelf and slope rise has been significantly changed.

References

Anonymous , An O(nlogn) implementation of the Douglas-Peucker algorithm for line simplification: [http://dl.acm.org/citation.cfm?id=178097].

DOUGLAS, D.H., and PEUCKER, T.K., 1973, ALGORITHMS FOR THE REDUCTION OF THE NUMBER OF POINTS REQUIRED TO REPRESENT A DIGITIZED LINE OR ITS CARICATURE: Cartographica: The International Journal for Geographic Information and Geovisualization, v. 10, p. 112 122.

Norris, M., and Faichney, A., 2002, SEG Y rev 1 Data Exchange format: Technical Standards Commitee SEG (Society of Exploration Geophysicists).