

Scripps Institution of Oceanography Technical Report

SeisDig

Software to Digitize Scanned Analog Seismogram Images

User's Manual

Peter D. Bromirski

and

Steve Chuang

Scripps Institution of Oceanography
Integrative Oceanography Division
University of California, San Diego

June 20, 2003

This software is Copyright ©2002 by The Regents of the University of California. All Rights Reserved.

Permission to use, copy, modify, and distribute this software and its documentation for educational, research and non-profit purposes, without fee, and without a written agreement is hereby granted, provided that the above copyright notice, this paragraph and the following three paragraphs appear in all copies.

Permission for commercial use or to incorporate this software into commercial products may be obtained by contacting the Technology Transfer Office, 9500 Gilman Drive, 411 University Center, University of California, La Jolla, CA 92093-0093, (858) 534-5815, invent@ucsd.edu .

This software program and documentation are copyrighted by The Regents of the University of California. The software program and documentation are supplied "as is", without any accompanying services from The Regents. The Regents does not warrant that the operation of the program will be uninterrupted or error-free. The end-user understands that the program was developed for research purposes and is advised not to rely exclusively on the program for any reason.

IN NO EVENT SHALL THE UNIVERSITY OF CALIFORNIA BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, INCLUDING LOST PROFITS, ARISING OUT OF THE USE OF THIS SOFTWARE AND ITS DOCUMENTATION, EVEN IF THE UNIVERSITY OF CALIFORNIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE. THE UNIVERSITY OF CALIFORNIA SPECIFICALLY DISCLAIMS ANY WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE SOFTWARE PROVIDED HEREUNDER IS ON AN "AS IS" BASIS, AND THE UNIVERSITY OF CALIFORNIA HAS NO OBLIGATIONS TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS.

This software was developed with government support under Office of Naval Research Award Number: N000140110403. The government has certain rights in this software.

Development of this software was also supported under a grant from the G. Unger Vetlesen Foundation.

Contents

1	General Overview	5
2	Main Features	6
2.1	Installation and Matlab/System Requirements	7
2.2	Image File Formats	7
2.3	Required User Inputs	7
2.4	Pre-Processing	8
2.4.1	Preparing Images for Digitizing	9
2.5	Edge Lines	9
2.6	Trace Activation (Line Numbers)	10
2.7	Trace Digitizing and Filtering	10
3	Menus	12
3.1	File Menu	12
3.1.1	Importing Images and Processing Data	12
3.1.2	Saving Images and Processing Data	12
3.1.3	Exporting Digitized Data	12
3.2	Image Menu	14
3.2.1	Image Manipulation Functions	15
3.2.2	Image Filtering	15
3.2.3	Invert Image	15
3.2.4	Reloading	16
3.3	Edit Menu	16
3.3.1	Undo	16
3.3.2	Overlay Display Options	16
3.3.3	Show Opposite Image Edge	17
3.3.4	Show Reference Line	17
3.3.5	Delete Line Trace Data	17
3.3.6	Delete Line Control Points	17
3.3.7	Minute Marks	17
3.3.8	Debug/Restore	17
4	GUI Functions	18
4.1	Navigation	18
4.1.1	Shift Image	18
4.1.2	Image Magnification	18
4.1.3	Image Ratio	19
4.1.4	Auto Scroll	19
4.2	Digitizing Control/Trace Editing Functions	19
4.2.1	Control Points	19
4.2.2	Select Points	20
4.2.3	Erase Trace	20

5	Time and Amplitude Scale Factors	21
5.1	Time (Independent Variable) Scale Factor	21
5.2	Trace Amplitude (Dependent Variable) Scale Factor	22
6	Digitizing Quality Control	23
6.1	Interactive Trace Modification/Editing	23
6.1.1	Spline Fitting	23
6.2	Trace Concatenation	24
7	Digitizing Procedure	25
7.1	Seismograms	25
7.2	Generic Series: Time or Otherwise	26
8	Keyboard Shortcuts	27
9	SeisDig SAC2000 Structures	28

1 General Overview

Analog seismic data have been recorded on numerous pen and paper drum-type seismographs throughout the world over the past 100 years or so. Digitizing scanned images of resulting seismograms using the available version of digitizing software (*NxScan V. 8.0*) was extremely problematic because *NxScan* was written for Sun Microsystem workstations during the 1980's and uses the obsolete, unsupported "pixrect" primitive graphics library written for the 1980's Sun architecture and operating system. Additionally, *NxScan* was not designed for the efficient digitization of day-long seismograms for climatological studies (e.g. *Bromirski et al.* [1]). After considerable effort was spent attempting to improve and stabilize *NxScan*, incompatibilities of pixrect with current Sun architectures continued to cause repeated crashes, making the efficient digitization of month-long time series impossible. Thus, *SeisDig* was developed using Matlab, which makes *SeisDig* essentially computer and operating system independent. The hope is that *SeisDig* can be relatively easily maintained as Matlab and computer systems evolve.

Examples and illustrations in this manual were made from Galitzin seismograms from the Berkeley Seismological Laboratory (BSL) archives that were electronically "scanned" at 400 dpi (dots/inch) using an IDEAL FSS8000 full length scanner and stored as "tif" files. Each scanned seismogram image consists of a series of quasi-parallel sequential 30-minute or 60-minute traces that generally cover about a day, with two partial traces at the beginning and end of each image. One-minute timing marks usually cause small gaps at the end of each minute. In addition to high amplitude earthquake signals, the arrival of high amplitude storm waves at nearby shorelines (*Bromirski and Duennebier* [2]) causes elevated noise levels called "microseisms", both types of signals resulting in the occurrence of adjacent trace intersections, overlaps, and reduced line thickness that complicate digitization.

Storage problems over time have damaged some seismograms (particularly at edges), further complicating digitization of full seismograms. Degradation of the seismogram paper quality (especially for many pre-1950 seismograms), and the difficulty in perfectly aligning the 40 inch long seismograms in the scanner feed groove, commonly result in either slightly skewed or slightly curved trace images, typically exaggerated at the trace ends. These scanning artifacts become apparent if the horizontal:vertical image ratio is augmented, but are difficult to identify in the raw tif images. The effects of these scanning artifacts were minimized by fitting a third order polynomial to the image traces to establish a reference line for each trace to determine the digitized trace amplitudes.

Because of presumably varying voltage, Galitzin seismograph drum speed variability resulted in the Time Scale Factor (TSF) varying from about 450 to 480 pixels/min. Impact of the variable TSF is minimized by using an average pixels/min determined for the entire seismogram. When digitizing relatively short duration earthquake signals, estimates of the TSF should be made either just prior to the event or from an average of the TSF for traces before and after the event.

2 Main Features

SeisDig is a Graphical User Interface (GUI) based interactive digitizing tool. Digitizing operations are mouse selected. Because digitizing numerous entire seismogram images efficiently is critical for studying the characteristics of extreme storms and their occurrence for climate change analyses, automated trace image tracking is necessary. Efficient digitizing necessitates a minimum of user interaction.

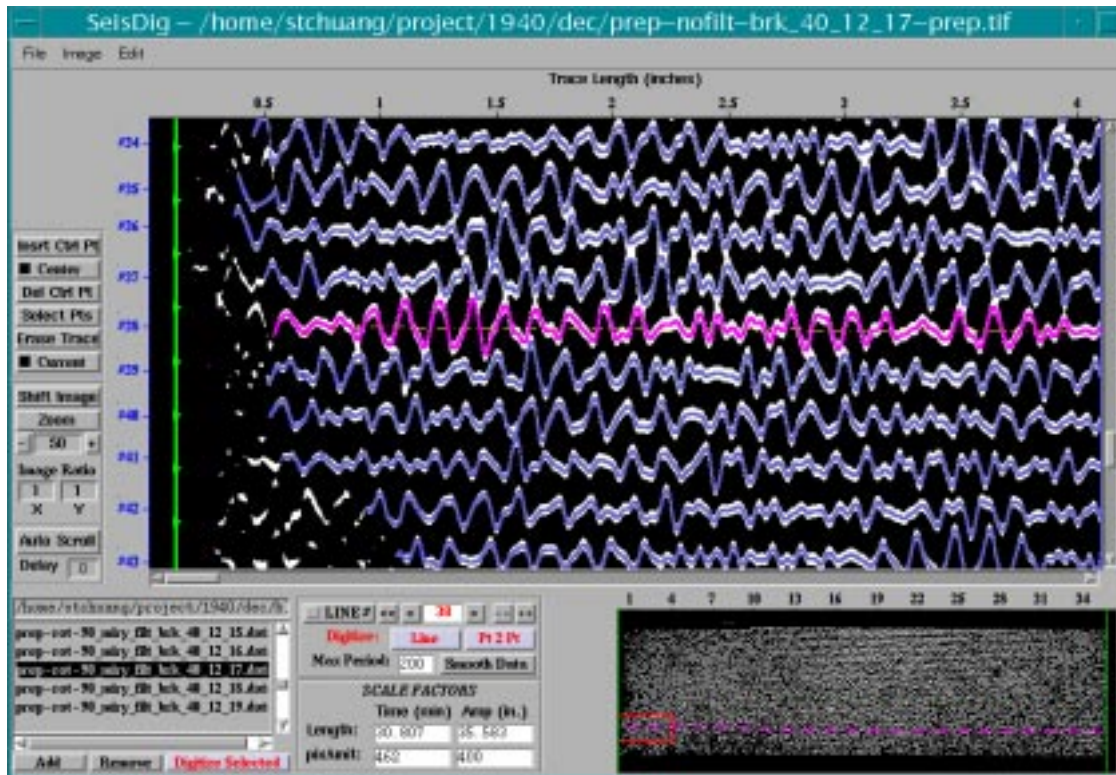


Figure 1: A portion of the seismogram for Dec. 17, 1940 viewed with *SeisDig* shows the trace variation observed in pre-1962 Galitzin seismograms, with an example of faded trace lines at the seismogram edge that sometimes occurs in older seismograms. The scanned traces (thick white lines) show examples of adjacent trace intersections during relatively high amplitude noise levels. The projected digitized traces (blue lines) successfully track past these irregularities. The magenta trace is the current (active) trace under inspection. The *Edge Line* (vertical green line) is shown on the left, its position adjusted to conserve the total elapsed time of the seismogram. The full seismogram image is shown in the locator panel at the lower right, with the portion of the image shown in the main image panel outlined by the red box and the active trace identified by the dashed magenta line. The right *Edge Line* is set at the well defined trace ends. The paper seismogram physical trace length in inches (determined using the *Amp. SCALE FACTOR* value displayed, i.e. the DPI) is shown at the top of both the main *SeisDig* and the locator panels.

Timing marks, occurring at 1-minute intervals in most Galitzin seismograms, cause small gaps in the seismogram image, and high amplitude signals result in trace intersections and overlaps (see Fig. 1). The trace width can vary within a trace as well as from trace to trace. Each of these, as well as other gaps and thin or intermittent lines, can cause incorrect trace-tracking. The necessity of setting control points at gaps and intersections was substantially eliminated with a trace-tracking algorithm that follows the center of the trace image (in the vertical cross-section) and uses a combination of slope estimates from both the preceding digitized points and the trace amplitude range to constrain the forward digitizing search. Trace tracking is further constrained to recross the baseline within a user selected time period. Incorrect handling of trace intersections and other image irregularities (commonly occurring when high amplitude signals cause repeated trace intersections or overlaps) are easily identified by visual inspection of the projected digitized trace overlaying the raw image trace. The GUI allows interactive modification of the digitized trace(s) (as well as setting control points at trace ends for pre-processing) prior to saving to an output file for cases when image quality and/or complexity warrant user interaction.

SeisDig operates on black-and-white images, 0-and-1 pixel values, respectively. Grayscale or color images are automatically converted into black-and-white.

2.1 Installation and Matlab/System Requirements

A “tar” file containing the Matlab *SeisDig* m-files and sample image files can be downloaded from <http://ccs.ucsd.edu/~peter/> and should be untared in the user's *~/matlab* directory, creating a *SeisDig* directory with a data subdirectory containing sample image files, and should be added to the user's MATLABPATH. Minimum system requirements: Matlab V.6.0 or higher with Image Processing and Signal Processing Toolboxes, and 256 Mb memory; Matlab v.6.1+ with 512+ Mb memory recommended.

2.2 Image File Formats

Acceptable input image file formats are those that Matlab function “imread” supports. These include TIF, JPEG, GIF, BMP, PNG, HDF, PCX, XWD, ICO, CUR. Images of these types can be imported directly without requiring conversion or modification.

2.3 Required User Inputs

If the image does not contain timing marks or only a section of a trace is to be digitized, the user is required to enter the time in minutes (1) between the Edge Lines (generally between identifiable minute marks), which is needed to determine the Time Scale Factor. The “dots per inch” (DPI) image scan rate (2) determines amplitude scaling (the default is 400 dpi). If the DPI is unknown, the physical length of the horizontal dimension (independent variable) of the scanned image must be entered. In the *Export (Preview)* window (see Section 3.1.3, Fig. 5), the user is required to enter (3) the absolute Start Time of the seismogram or the section of the image to be digitized, and (4) the desired digital output sampling rate, i.e. $1/\Delta$. When digitizing entire seismogram

images, the absolute End Time (5) may also be needed to ensure that the independent variable (time) is correctly conserved during trace concatenation. Additional header information prior to output, such as station name and channel identifier, are optional but can also be entered in the *Export* window.

2.4 Pre-Processing

Much of the time required for digitizing an entire seismogram image (depending on the degree of user interaction needed) is CPU and memory intensive, with digitizing a 30 min trace requiring about 1.5 min or more. For efficient digitizing of day-long seismogram images, it is useful to determine the projected digitized traces for the entire seismogram prior to inspection, modification, and saving. After loading the image file, control points must be selected at both ends of all traces to be digitized while in *Insert* mode. These points should be near the trace ends, but need not be exactly at the end of the trace image. For efficient control point selection along one side of the seismogram, activating the *Line #* check box allows automatic incrementation of the line number when adding successive line-end control points. Control points can be set at trace ends along one side of the seismogram either with individual mouse clicks at each trace end, or setting a series of end control points by holding down the left mouse button on the non-image region (black) above the initial trace end of the series and dragging the cursor to the non-image region below the last trace in the series (i.e. black-to-black). Line #'s for groups of multiply-selected control points are automatically incremented from top-to-bottom, and decremented when dragging the cursor from bottom-to-top (this makes setting control points at trace ends more efficient). Groups of incorrectly labeled trace-end control-point line numbers can be selected by toggling the *Select Points* button and holding down the right mouse and dragging the cursor. Selected groups of control points can be decremented or incremented using the -- or ++ buttons, respectively.

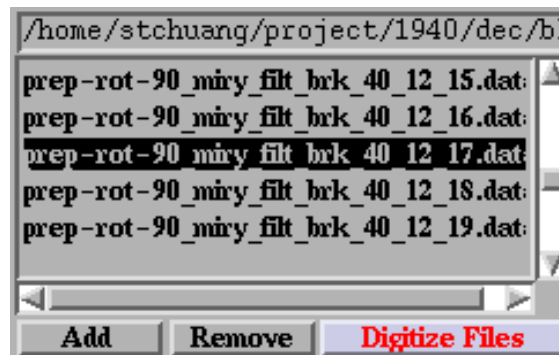


Figure 2: The pre-processing file list showing typical filenames produced by *prep_images.m*. The highlighted preped image file was digitized, with the output displayed in Fig. 1.

The default intermediate processing filenames have the same basename as the image file, although the user can enter any filename. Projected trace data and/or associated control point and Edge Line data are saved in a file having a “.data” extension (unless the user enters a different filename) by

clicking on the “*Save Data*” button. This filename automatically appears in the “*Preprocess File List*”, located in the lower left corner of Fig. 1. The *.data files are actually *.mat Matlab “save” files. The directory path that these files are stored in is shown above the *Preprocess File List*. Double-clicking on filenames in this list, either image *.tif or *.data files, will load that file. Files to be digitized can be added or removed from the *Preprocess File List* (Fig. 2) using the “*Add*” or “*Remove*” buttons, respectively, with the files to be processed highlighted after being selected from the list by mouse click. Preprocess files not in the current directory can be added to the list using their complete pathnames. The CPU intensive operations can then be performed at a convenient time by clicking on the “*Digitize Selected*” button, with computed projected traces stored in a *.data file that has “tr-” prepended to the filename. “*Load Data*” loads the selected trace tr*.data object file for inspection and correction, with the user prompted to load the associated image file for comparison with the projected traces if it is not already loaded. Inspection of the projected digitized traces can be done conveniently using the *Auto Scroll* function.

2.4.1 Preparing Images for Digitizing

Removing stray pixel “noise” from images and 90° rotations are time consuming operations. When digitizing multiple seismograms that require the same operations to prepare the image for trace end control point selection and digitizing, it may be more convenient to do filtering and rotation operations using the *prep_images.m* function that can be run independently of *SeisDig* (“help *prep_images*” in Matlab gives usage). This function performs the same operations as in the *Image* menu within *SeisDig*, creating a new tif file (with the filename prepended with “prep-” and abbreviations of all operations performed, see Fig. 2). Preped tif files can be loaded later to set control points at trace ends, with control point and image data automatically stored in a new file having the same basename as the preped tif file but with a “.data” appendage, with the new filename automatically added to the *Preprocess File List* for digitizing.

2.5 Edge Lines

Edge lines are used to delineate the horizontal (independent variable, time) trace extent. Edge lines need to be set prior to *Export*. Edge line locations can be adjusted by positioning the cursor over the Edge Line, holding down the left mouse button, and dragging the Edge Line to the desired location. Incorrect scanning can cause the trace ends of seismograms to not be aligned directly beneath each other, i.e. non-vertical, which can be identified by increasing the horizontal:vertical *Image Ratio*. To account for the slight slope of the trace ends and conserve the total elapsed time in the seismogram, the Edge Lines may be rotated by positioning the cursor over the upper end of the Edge Line, holding down the left mouse button and dragging the cursor to either side.

When digitizing full seismograms, Edge Line locations are selected by picking a balance between the quality of the image trace ends and the trace length of the paper seismogram (keeping in mind that there is a slight slope to the spiraling traces, resulting in the trace length being slightly longer than the seismograph drum circumference). If the trace length is measured, adjust the Edge Lines until the value in the Scale Factors *Amp. Length* dialog box matches the measured length. If the image quality is poor at one end (as in Fig. 1), set the Edge Line at the good end and iteratively adjust the other Edge Line to the measured trace length and/or until the Total (elapsed) Time in the

seismogram (determined in the *Export (Preview)* window, see Fig. 5 in Section 3.1.3) is adequately conserved.



Figure 3: Line activation and digitizing controls.

2.6 Trace Activation (Line Numbers)

When multiple traces are being digitized, the current (active) line number is displayed in the “*Line #*” dialog box (Fig. 3) below the main *SeisDig* image panel (Fig. 1), with the current trace under inspection identified with a magenta line (other projected digitized traces have blue lines). The location of the active trace is also identified by a magenta line in the locator panel. Line numbers are also displayed to the left of the main image panel. Line numbers are useful when editing connection sections between successive concatenated traces. A line is “activated” or made “current” (i.e. selected for control point changes and/or digitizing) by entering the desired line number in the *Line #* dialog box and striking “Enter” on the keyboard, using the associated forward > and backward < arrows for scrolling, or by clicking on the desired line number to the left of the main image panel. The << button makes the lowest numbered trace the active trace. For efficiency when sequentially setting pre-processing control points for successive trace end points at one side of the seismogram image, the line number is automatically incremented if the *Line #* check box is activated. If left and right side line numbers are inadvertently misaligned (indicated by the yellow baselines having substantial slopes because the endpoints are incorrectly selected on adjacent traces), resequencing can be accomplished by selecting groups of control points on one side of the seismogram (by holding down the right mouse button and dragging the cursor while in *Select Points* mode) and clicking on either backward -- or forward ++ buttons.

2.7 Trace Digitizing and Filtering

Individual trace digitizing is accomplished by clicking on “*Line*” (Fig. 3) to digitize the entire trace, or “*Pt 2 Pt*” to digitize the portion of the trace between selected points. The *Line* and *Pt 2 Pt* buttons (red with blue background) operate on the current active line indicated in the *Line #* dialog box. The *Pt 2 Pt* function is used during projected trace modification prior to export. The *Pt 2 Pt* end points, Beginning and End, are selected with left mouse clicks after clicking on the *Pt 2 Pt* button. To facilitate the selection of the End point in *Pt 2 Pt* mode, right mouse clicks shift the image only horizontally. Note that floating baselines are used because of potential trace curvature. The baseline location is determined from a linear least squares fit to all control points in

a line, and the trace-tracking algorithm operates relative to this baseline. Consequently, to avoid the possibility of changing acceptable previously digitized portions of a trace, the *Pt 2 Pt* operation should generally be used when modifying previously digitized data after inserting additional control points within the poorly tracked segment. The maximum time in seconds allowed between recrossings of the reference baseline is given by "*Max. Period*", and is particularly useful for high amplitude data with multiple overlaps.

The image line-thickness variability along the trace causes small jumps in the projected digitized trace unrelated to the frequency content of the underlying signal. Consequently, a three-point triangular filter is automatically applied to the digitized output. If inspection of the digitized trace warrants further filtering, additional triangular filtering of the current (active) projected digitized trace can be applied by clicking on the "*Smooth*" button, with each mouse click successively applying a three-point triangular filter to the current projected digitized trace, which can be undone if the result is unsatisfactory. Note that this is a lowpass filter that will cause a loss of energy at higher frequencies.

3 Menus

All operations necessary to digitize an entire seismogram image can be performed within the GUI, including image filtering (for “noise” reduction), image rotation and inversion, and output trace inspection and modification.

There are three main menus at the top left corner of the GUI (see Fig. 1): *File*, *Image*, and *Edit*.

3.1 File Menu

Operations under the “*File*” menu deal with loading and saving image and data files.

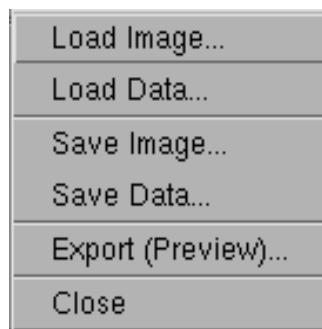


Figure 4: Operations performed under the *File* menu.

3.1.1 Importing Images and Processing Data

Images are imported (loaded) by clicking on the “*File*” button at the top left hand corner and selecting “*Load Image*”, and then selecting the desired image file from the directory dialog box that pops up. Similarly, files containing control points and pre-processed projected trace estimates can be loaded by selecting “*Load Data*”. The filename of the loaded image being displayed is shown at the top of the *SeisDig* window.

3.1.2 Saving Images and Processing Data

Similar to importing images, modified images and processing data (including control points and Edge Line locations) can be saved for further editing by clicking on the “*Save Image*” and “*Save Data*” buttons, respectively. All user entered data, including control points, Edge Line locations, and projected digitized trace estimates are stored in the tr*.data files under *Save Data*. Changes to a raw image (either orientation or filtering) can be stored by clicking on the “*Save Image*” button.

3.1.3 Exporting Digitized Data

Digitized data are saved to disk by clicking on “*Export (Preview)*”. An *Export* window (Fig. 5) pops up in which the user must enter the Start Time and the sampling rate (1/Delta) of the digitized

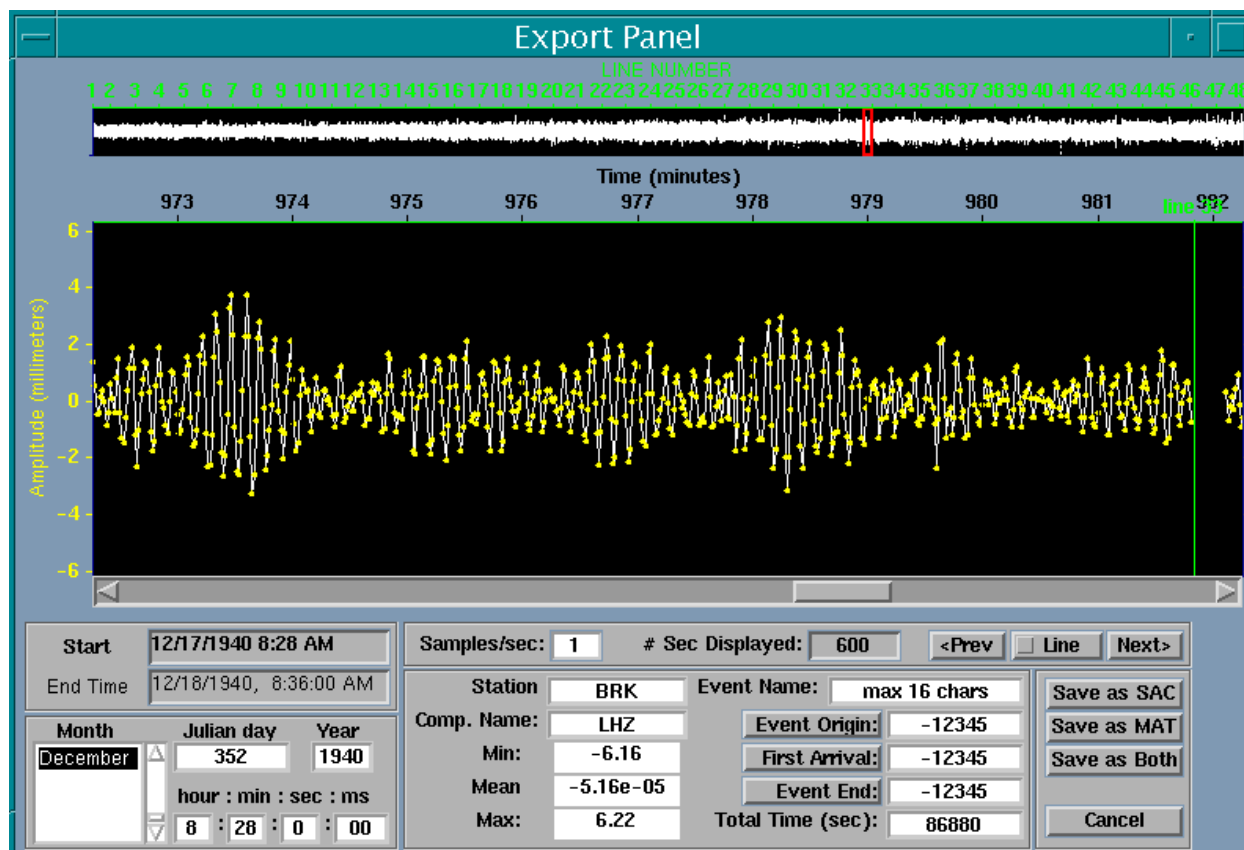


Figure 5: Layout of the “*Export (Preview)*” pop-up window. The top panel shows the concatenated output of the digitized data for the full seismogram shown in Fig. 1, containing 48 traces of approximately 30 min duration each. The portion of the digitized trace outlined by the red locator box (600 s) in the upper panel is expanded in the main panel for detailed inspection. The yellow dots indicate the sampled amplitudes to be exported. Note the semi-regular coda-like arrivals, indicative of microseisms generated by the arrival of ocean swell at nearby shorelines.

trace data, and any Seismic Analysis Code (SAC) header information available. “# *Sec. Displayed*” gives the length of the time segment displayed in the main panel in seconds. The location of the expanded trace section shown in the main panel can be selected by either clicking on the trace in the top panel, dragging the slide bar at the bottom of the main panel, or clicking on the “<Prev” or “Next>” buttons (when the associated checkbox is set to *Line*) to inspect connections between consecutive concatenated traces, with the connection points indicated by vertical green lines in the main panel (see Fig. 5). Toggling the *Line* checkbox in the *Export* window switches to *Nmax*, which allows jumping to successive trace maxima (useful for locating spikes that are generally associated with improperly digitized segments).

Header information for SAC output requires entry of the sampling rate (1/Delta, samples/s) and the Start Time of the data, which includes date information. The Start Time corresponds to the stan-

standard SAC header “Reference Time”, but differs in *SeisDig* where it must be the beginning time of the data being digitized. The inverse of the sampling rate, i.e. the sampling interval Delta, is output in the SAC headers. The locations of samples taken for a given sampling rate are determined by the number of pixels/s, which is obtained from the measured pixels/min from the seismogram image. The 400 dpi Galitzin images have about 7.7 pixels/s. Unlike *NxScan* which truncated the pixels/s estimate causing slightly shifted spectral estimates, *SeisDig* uses spline interpolation to determine the trace amplitudes at the locations specified by the sample rate entered.

To check the conservation of total elapsed time when digitizing full seismograms, the End Time needs to be entered. A pop-up window gives the Digitized Total Time (determined from the digitizing process) and the Entered Total Time (determined from the Start and End Times entered) in seconds. Typically, these differ by a few hundred seconds for a full day-long seismogram. The user has the option of accepting the Entered Total Time (generally recommended), in which case the sampling interval (i.e. the Time Scale Factor, TSF) will be adjusted accordingly, and/or slightly shifting one or both of the Edge Lines and then re-Exporting. Note that the TSF is an average estimate from all the individual lines in the seismogram, and thus has an associated variance.

Digitized data are exported by selecting either “Save as SAC” (the common SAC format) or “Save as MAT” (a Matlab *.mat file that mimics the SAC2000 Matlab output). SAC2000 can be downloaded from <http://www.llnl.gov/sac/>. The Matlab output *.mat file includes a trace vector (*SeisData*), and a header “structure” (*SACdata*) that contains necessary information found in the SAC2000 Matlab header, including a *SeisDig* struct containing digitizing parameters (see Section 9). There are often numerous data gaps in the older Galitzin seismograms (see Fig. 1). Thus, in order to keep the number of data files to a manageable level, the SAC files generated by *SeisDig* differ from standard SAC files in that missing data gaps are designated by “-12345”, and with NaNs in the Matlab *.mat output files. The output digitized trace amplitudes are in millimeters and correspond to trace amplitudes of the paper seismogram. The SAC header variable “scale” is set equal to 1.

The *Begin Value* and *End Value* fields in the *SeisDig* SAC header are set to the first and last data points exported. If an earthquake event is embedded within the output data, the *Event Origin*, *First Arrival*, *Event End* times may be selected by mouse clicking on points in the main *Export* panel after activating the respective button, with middle mouse button clicks causing lateral translation of the trace image centered at the cursor location. The event times are indicated by vertical magenta lines in the main *Export* panel. The values for the event times stored in the SAC header are the number of samples from the beginning of the output trace. These points can be located in absolute time from the Start Time of the data file and the Delta. Choosing event times is optional, with these values set to “-12345” in the SAC header if unspecified. Entering -12345 <cr> in any of the Event fields erases the associated magenta line.

3.2 Image Menu

The digitizing algorithm operates from screen left to right. Thus, the independent variable (time) must be increasing from left to right. Some seismograms need to be scanned lengthwise or inverted

because of physical damage to the paper during storage. Consequently, several image manipulation functions (Fig. 6) are included to align the image correctly for digitization. Aging or damage to the paper seismogram during storage can result in the inclusion of non-seismic “image noise” that needs to be filtered for efficient and accurate digitization by the trace-tracking algorithm. Most of these operations are time consuming. When digitizing multiple seismograms having the same characteristics, it may be more convenient to perform these operations external to *SeisDig* using *prep_images.m* (see Section 2.4.1).

Mirror About Y-Axis
Mirror About X-Axis
Rotate 180
Rotate +90 (slow)
Rotate -90 (slow)
Filter Image Noise (slow)
Erode Image (slow)
Dilate Image (slow)
Invert Image
Reload From Disk

Figure 6: Operations performed under the *Image* menu.

3.2.1 Image Manipulation Functions

Seismogram images can be rotated about either the screen x-axis or y-axis using “*Mirror About X-axis*” or “*Mirror About Y-axis*”, respectively. Rotation of the image about the screen z-axis can be done by clicking on the “*Rotate 180*” for 180° rotation, or “*Rotate +90*” or “*Rotate -90*” buttons for clockwise and counter-clockwise 90° directions, respectively (the $\pm 90^\circ$ rotations are much slower operations because of memory reallocation).

3.2.2 Image Filtering

“*Filter Image Noise*” removes spurious isolated pixels and offset minute marks that interfere with accurate automated digitizing, and is recommended for most images. “*Erode*” and “*Dilate*” functions effectively contract or expand, respectively, the image trace line thickness, but are not recommended for use except in special cases. Because image filtering operations cannot be reversed, these features should be used with caution. If the results of any of these functions are unsatisfactory, the image can be reloaded.

3.2.3 Invert Image

This function switches the background/foreground from white-to-black and vice versa. *SeisDig* requires a black background, i.e. pixel values set equal to zero.

3.2.4 Reloading

In the event that an image is over-filtered or in some way corrupted during digitizing operations, the current image file can be reloaded using the “*Reload From Disk*” option, which clears all objects and control points.

3.3 Edit Menu

Digitizing modifications can be facilitated by being able to add or remove previously determined overlays, as well as undo incorrect trace modifications. Such functions that have proved useful are found in the “*Edit*” menu (Fig. 7).

3.3.1 Undo

All operations are stored for possible restoration to the prior state (except *Image* menu operations). Successive operations can be reversed by clicking on “*Undo*” or striking “<ctrl> u” using the keyboard, or also by striking “u” if the main image panel has been activated with a mouse click (i.e. is the current Matlab object).

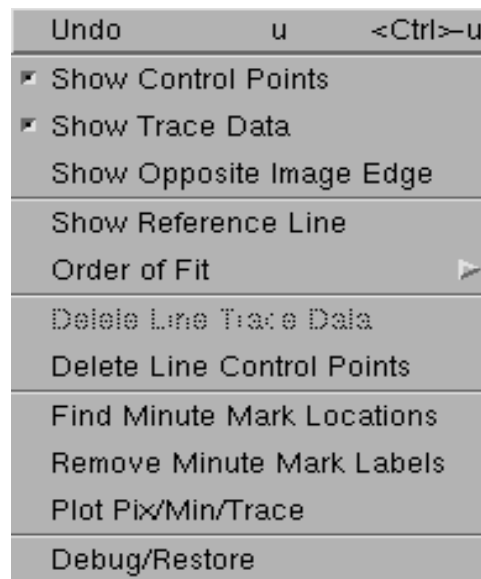


Figure 7: Operations performed under the *Edit* menu.

3.3.2 Overlay Display Options

Operations for the entire image include erasing or redisplaying all control points and/or the projected digitized traces from the viewed image by toggling the “*Show Control Points*” and “*Show Trace Data*” buttons.

3.3.3 Show Opposite Image Edge

When digitizing multiple traces, setting control points at trace ends for preprocessing and adjusting the Edge Line fit, checking trace concatenation and trace numbering, or checking the reference line fit to the image, “*Show Opposite Image Edge*” provides a convenient function to observe the traces at opposite ends of the seismogram image simultaneously.

3.3.4 Show Reference Line

The seismogram image traces can be slightly curved due to incorrect paper feed in the scanning operation. To establish a reference line for determining output digitized trace amplitudes that accounts for this slight curvature, a polynomial fit to the digitized points of orders 1, 2, or 3 can be chosen using the “*Order Fit*” option. It is recommended that full traces be digitized using a third order polynomial (the default). However, when digitizing shorter trace segments, a linear reference line may be more appropriate. “*Show Reference Line*” overlays the selected polynomial fit on all traces for visual inspection.

3.3.5 Delete Line Trace Data

If the trace-tracking algorithm gets completely “lost”, it is sometimes necessary to redigitize an entire trace. “*Delete Line Trace Data*” deletes all previously digitized data points only for the current (active) line.

3.3.6 Delete Line Control Points

“*Delete Line Control Points*” deletes all control points from the current (active) line.

3.3.7 Minute Marks

The locations of minute timing-mark (gaps) are estimated using a modification of the trace-tracking algorithm, and are numbered sequentially and overlaid on the image using the “*Find Minute Mark Locations*” function. Identifying minute marks is used to determine the Time Scale Factor, i.e. the number of pixels/min (see Section 5.1 below), needed to determine output trace sampling locations. The DPI needs to be entered (or estimated) to determine the minute mark locations. The overlaid minute mark numbers can be removed for the entire image with the “*Remove Minute Mark Labels*” function. The *Find Minute Mark Locations* function may not work well for images with trace overlaps. When digitizing full seismogram images, it can be useful to examine the variation in the pixels/min and minutes/line estimates for all the traces using the “*Plot Pix/Min/Trace*” function.

3.3.8 Debug/Restore

Occasionally, problems with either Matlab or the trace-tracking algorithm can occur. The “*Debug/Restore*” function resets the GUI buttons, and restores line labeling and the baseline locations.

4 GUI Functions

Various functions for moving around the seismogram image and correcting poorly tracked traces are activated by mouse clicking on the appropriate “button” at the left of the main image panel (Fig. 1). That function remains active until either reclicking or another button is activated. Because each of these functions use left mouse clicks, only one of these functions can be active at one time.

4.1 Navigation

Moving within the image can be done by (1) using the scroll bars on the bottom and right hand sides of the displayed image, (2) by clicking on the desired location in the locator panel at the lower right, (3) by using the “*Shift Image*” button, or (4) centering the image at the current cursor location by right clicking. The displayed portion of the image in all cases is identified by a red box in the locator panel at the lower right, which displays the full image and identifies the current active trace with a horizontal magental line.

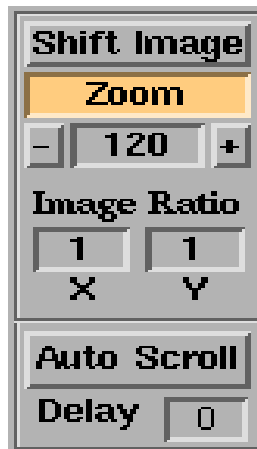


Figure 8: Image navigation controls. The *Zoom* button is activated, i.e. beige.

4.1.1 Shift Image

Activation of the *Shift Image* button allows “grabbing” and dragging the displayed image by holding down the left mouse button and moving the cursor.

4.1.2 Image Magnification

Activating the “*Zoom*” button allows image magnification (left button, zoom in) or reduction (right button, zoom out), with the zoomed image centered at the cursor location. The size and location of the red location box in the locator panel is adjusted accordingly. Zooming can also be accomplished by clicking on the “+” or “-” buttons adjacent to the zoom magnification dialog box, with zooming directed towards the upper left corner of the image. Clicking the middle mouse button centers the image at the cursor. The degree of zooming for both mouse clicks in the main image panel and

using the + or - buttons is a pre-determined percentage of the size of the initial red locator box, i.e. the size of the initial image displayed in the main image panel. The degree of zooming desired can also be manually entered in the *Zoom* dialog box, also relative to the initial main image (red locator box) size.

4.1.3 Image Ratio

Increasing the apparent relative amplitude of the dependent variable (i.e. the vertical or y-direction with an x:y *Image Ratio* of 1:50) is useful for identifying irregularities resulting from improper scanning, i.e. excessive trace curvature (the trace images should be nearly linear). Conversely, to check alignment of the trace ends, a horizontal:vertical *Image Ratio* of 100:1 facilitates the accurate selection of Edge Line location and slope.

4.1.4 Auto Scroll

After pre-process digitization of a full seismogram, cursory inspection of the projected digitized data to identify incorrectly traced segments can be performed using “*Auto Scroll*”, which steps through the entire seismogram in increments the size of the red box in the locator panel, with the viewing time in seconds between successive portions entered in the “*Delay*” dialog box. Either mouse clicking on *Auto Scroll* or anywhere in the *SeisDig* window stops scrolling.

4.2 Digitizing Control/Trace Editing Functions

Digitizing trace images requires choosing control points to constrain the digitizing algorithm, and to correct poorly tracked trace segments. Only one “trace editing” button can be activated at a time, with the activated button becoming green (see Fig. 9). Trace editing functions apply only to the line number displayed in the *Line #* dialog box, the active trace, unless the “*Current*” checkbox is toggled to “*All Lines*”. The *Current / All Lines* checkbox applies to all trace editing functions.

4.2.1 Control Points

Control points are needed at the beginning and end of each trace to identify the time-axis (independent variable) baseline and the trace extent, and within traces when the trace-tracking algorithm has difficulty correctly following a trace. Control points need not be set at the exact end of the trace (the trace-tracking algorithm will go to the end regardless). Control points can be added at multiple trace ends by holding down the left mouse button and dragging the cursor from black-to-black (non-image) locations, with a control point inserted at each crossing of a trace image (white region). Control points can be added only to the current active line indicated in the *Line #* dialog box (unless the *Line #* check box is activated for adding control points at multiple trace ends for pre-processing). Control points associated with the number in the *Line #* dialog box (active) are red, while inactive are blue. Activating the “*Insrt Ctrl Pt*” button allows adding as many control points as needed at selected cursor locations with left mouse clicks. Control points generally should be vertically centered in the image trace (the default setting). Centering can be turned off for special cases by toggling the “*Center*” check box (changes to “*Exact*”), and is particularly useful turned



Figure 9: Trace editing controls. The trace editing button *Insrt Ctrl Pts* is activated, i.e. green, with the *Center* checkbox activated causing subsequent control points inserted to be vertically centered in the *Current* trace indicated in the *Line #* dialog box.

off when adjusting the tracking at trace intersections.

Incorrectly added control points can be deleted at any time by activating the “*Del Ctrl Pt*” button and then clicking on the erroneous control points with the mouse. Individual control points can be deleted from any trace at any time with mouse clicks by toggling the *Current* checkbox (the default) to *All Lines*. Successively added control points can be sequentially deleted by striking the “u” key for “Undo” after activating the main image panel is made the active object by mouse click. All control points in a trace may be deleted with *Delete Line Control Points* in the *Edit* menu.

4.2.2 Select Points

Groups of control points can be selected for deletion or for incrementing line numbers by activating the “*Select Points*” button and then holding down the left mouse button and dragging the cursor to encompass control points on the *Current* line or on multiple lines when the *All Lines* button is active. To delete selected groups of control points, select the control points and then strike the keyboard “Delete” key. To increment or decrement, select and then click on the ++ or -- buttons, respectively.

4.2.3 Erase Trace

High amplitude signals result in trace overlaps that generally require interactive modification of the projected digitized trace. Incorrectly tracked trace data can be removed using the “*Erase Trace*” function, holding down the left mouse button and dragging the cursor to delineate projected trace data to be erased. The *Current / All Lines* check box applies to *Erase Trace* (as for other trace editing functions) so that incorrectly tracked signals can be erased without affecting previously correctly digitized adjacent traces. An entire digitized trace may be deleted with *Delete Line Trace Data* in the *Edit* menu.

5 Time and Amplitude Scale Factors

Scale factors for both the independent variable (Time for seismograms) and the dependent variable (Amplitude) (see Figures 1 and 10) need to be either entered or computed prior to exporting digitized data.

SCALE FACTORS		
	Time (min)	Amp (in.)
Length:	30.807	35.583
pix/unit:	462	400

Figure 10: Time (independent) and Amp (amplitude, dependent) variable Scale Factors, showing values typical for earlier Galitzin seismograms (see Fig. 1).

5.1 Time (Independent Variable) Scale Factor

The digitizing Time Scale Factor (TSF) is needed to determine the sampling “locations” along a projected trace to obtain the digitized output. An accurate estimate of the TSF (pixels/unit(min)) of the image is also needed to ensure conservation of elapsed time between Start and End Times of the seismogram. The seismograph drum spin rate varies over time (even hourly) for older Galitzin seismograms, giving variable trace lengths (i.e. number of minutes/trace). Thus, the TSF should be estimated for each seismogram. The TSF is determined automatically when digitizing full seismograms (that have minute mark gaps) as the mean of the pixels/min estimates (excluding outliers) for all lines to be digitized. The variation in the (pixels/min)/trace can be examined using the *Plot Pix/Min/Trace* function in the *Edit* menu, and which loads the automatically-generated associated “gaps-” file.

When digitizing short segments of seismograms, the TSF can be estimated by entering the *Length* (number of “units”) between two selected points within a trace (generally minute marks) delineated with the Edge Lines (i.e., for seismograms, the number of minutes) in the Scale Factors “*Time (min)*” dialog box. The number of minutes can be determined using the “*Find Minute Mark Locations*” function in the *Edit* menu, with the minute mark gaps numbered sequentially between control points and shown on the trace to check that the minute marks were correctly identified. These numbers can be removed using *Remove Minute Mark Labels* in the *Edit* menu.

When tick (minute) mark gaps are absent (as in Fig. 1), the image quality is poor, or multiple trace overlaps occur, the minute marks may not be identified correctly with either the *Find Minute Mark Locations* function or by visual inspection. In such cases, the TSF (pixels/min) must be entered manually. This is done by activating the *Time, pix/unit* dialog box by mouse click and entering the pixels/min value (or using the default estimate of 470), and striking the Enter key. The computed

trace length (in minutes) based on the entered TSF (and the Edge Line positions) will then appear in the *Length* dialog box.

The TSF can also be determined for the full seismogram using the the Start and End Times (annotated on the physical seismogram) entered in the *Export (Preview)* window. If Start and End Times are entered and accepted, the computed TSF will not be used. Because the trace length in older Galitzin seismograms can vary from slightly less than 30 min to about 31 min (the *Plot Pix/Min/Trace* function in the *Edit* menu also plots the number of minutes per line), the TSF estimate should be obtained from the displayed seismogram image or from a trace chronologically close.

5.2 Trace Amplitude (Dependent Variable) Scale Factor

The Amplitude Scale Factor (ASF, dependent variable) for the digitized time series corresponds to the image DPI (dots per inch). The DPI is used to convert the seismic digitized trace output amplitudes from pixels to millimeters. The DPI is generally known and should be entered in the Scale Factors “*Amp*” *pix/unit* dialog box (the default is 400 dpi).

If the DPI is unknown, the original paper seismogram physical size, i.e. the measured trace length, can also be used to determine the number of pixels/inch or DPI amplitude scaling of the image. The units of the paper seismogram physical trace length that generated the scanned image are in “inches” to conform to the DPI convention. Enter the paper seismogram trace length in the *Length* Scale Factors *Amp* dialog box (see Fig. 10). The number of pixels between the Edge Lines set at the trace ends then gives the ASF.

6 Digitizing Quality Control

The projected output trace of the digitizing algorithm should be inspected to ensure the quality of the digitized data. Because there are several factors that affect the trace image quality, initial projected digitized trace estimates often require correction. After *Pre-Processing*, entire images can be scanned for digitizing errors using *Auto Scroll*. Prior to saving to disc during *Export*, the concatenated digitized traces are displayed in a scrollable window that appears after clicking on *Export (Preview)* in the *File* menu. After correction, the digitized data must be re-*Exported*, and modified sections should be re-inspected.

6.1 Interactive Trace Modification/Editing

Incorrect image trace-tracking is easily identified by comparing the projected digitized trace with the underlying image, or by identifying inconsistent amplitudes in the *Export (Preview)* window. Poor trace tracking generally results in high amplitude “spikes”, which can be located by searching through the trace in the main *Export* panel using the scroll bar or by toggling the *Line* checkbox to *Nmax* (the “next maximum”) and examining successive maxima. Poorly tracked sections should be modified in the main *SeisDig* panel and then re-*Exported*.

Changing the trace-tracking can be accomplished by inserting control points at locations where the projected digitized trace deviates from the underlying image. Clicking on *Insrt Ctrl Pt* activates *Insert* mode. In most cases, re-digitizing modified trace sections should be done using the *Pt 2 Pt* function. This reduces digitizing time and avoids the possibility of adversely affecting previously digitized trace segments that were acceptable.

Control points should then be inserted to guide the tracking along the correct path. Groups of control points can be inserted in poorly tracked segments along a single trace (while in *Insert* mode with the *Current* checkbox active) by holding down the right mouse button and dragging the cursor horizontally from a non-image region to another non-image region (black-to-black) along a trace, with control points set at each image location (white) crossed. The *Pt 2 Pt* digitizing function can then be applied. On occasions when the image quality is poor but the trace is potentially identifiable, it is useful to load the raw, unfiltered seismogram image, overlay the projected digitized data, (determined from the filtered image) and use the raw image data as the guide for setting control points to improve trace tracking. Although generally unnecessary, sections of a digitized trace that have been poorly tracked can be deleted with “*Erase Trace*” by holding down the left mouse button and selecting the desired area. Several applications of the *Pt 2 Pt* digitizing function (with repeated additions of control points) should be used prior to resorting to *Spline Fitting*.

6.1.1 Spline Fitting

In instances of extended poor image quality but with a visually identifiable trace, holding down the left mouse button and dragging the cursor from an identified portion of the trace image to another identified image point (“white-to-white” while in *Insert* mode) identifies a section for “spline” fitting. Other pre-existing control points within the splined section become incorporated in the spline fit. Repeated clicks on identifiable trace segments within the spline section constrains the spline fit.

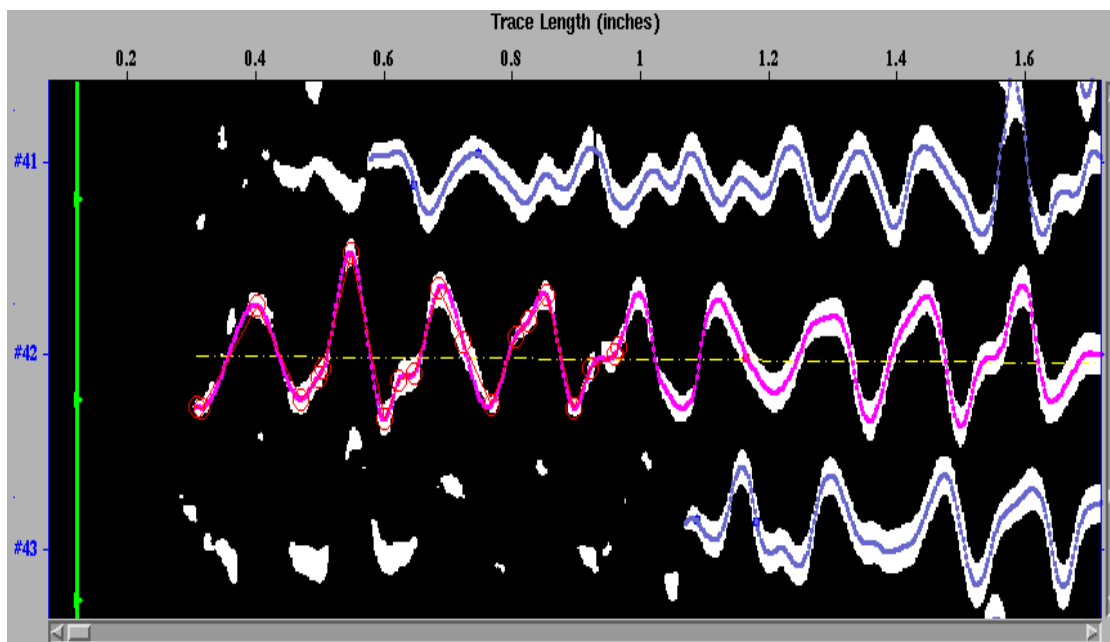


Figure 11: The main *SeisDig* panel showing an example of a portion of a trace image at the beginning of Line 42 (up to about the 0.95 inch trace location) that was spline-fitted. Red circles connected by thin red lines identify locations of spline control points. Compare with the unmodified projected trace determined by the digitizing algorithm (Fig. 1, beginning of Line 42).

Note that individual control points cannot be deleted within spline sections. If the splined section is unsatisfactory, use *Erase Trace* to delete the projected trace in the splined section, followed by *Select Pts* and striking the keyboard Delete key (see Section 4.6) to delete all control points within a splined section. Figure 11 shows a splined section at the beginning of Line 42, up to about the 0.95 in trace location. The blank space between the (green) Edge Line and the beginning of the trace results from fading of the seismogram traces at this edge, with the Edge Line set to conserve the Total Time in the seismogram.

6.2 Trace Concatenation

Digitizing consecutive traces in a seismogram image requires concatenating the trace at one end of the image to the next trace at the opposite end. Storage problems over time have resulted in degraded image quality at the edges of some seismograms (see Figures 1 and 11). Minute marks and faded trace ends can result in gaps at trace connections that need to be accounted for. When digitizing full seismograms, the user sets the “Edge Line” location for the end of the image that is best defined. The other Edge Line can be iteratively adjusted to minimize the time difference between the digitized and entered Total (elapsed) Times, determined from the Start and End Times entered by the user in the *Export (Preview)* window.

The concatenated digitized traces can be examined by clicking on the *Export (Preview)* button in the *File* menu, with the connection locations identified for inspection (see Fig 5). Any needed modification at trace connections (or Edge Line adjustment) can then be performed in the *Seis-Dig* window (near the connection locations of the indicated line number in the *Export* window) and then re-Exporting. Data gaps less than 50 pixels (about 5 s) that commonly occur between concatenated traces (and at minute marks) are automatically splined across, with larger gaps designated as missing data.

7 Digitizing Procedure

7.1 Seismograms

- Import the image, rotate and mirror the image as needed so that Time (the independent variable) increases from left-to-right and top-to-bottom. Filter the image noise. Because filtering the image is CPU and memory intensive in Matlab (i.e. slow), the *Save Image* option should be used so that image filtering and rotation do not have to be repeated (be careful not to overwrite the original image unless desired). If more than one image having the same characteristics is being digitized, *prep_images.m* can be used external to *SeisDig*.
- Set control points at both end points of all lines (or the line segment) to be digitized. It is more efficient to set all the control points on one side of the seismogram image, and then the other side. When setting trace-end control points for successive lines, “activating” the *Line #* check box makes it unnecessary to change the line number in the *Line #* dialog box.

The following two operations are necessary only when digitizing multiple traces, although this procedure may be useful in estimating the arrival time of an earthquake signal if arrival time determination by inspection proves problematic or background noise estimates are needed.

- Go to the file menu and click on *Save Data*, resulting in the control points “prep-*.data” file-name being added to the *Preprocess File List*. Click on *Digitize Selected* (pre-processing takes 30-45 min/seismogram, depending on machine architecture).
- After digitizing is completed, load the control points and projected digitized trace data using *Load Data* in the *File* menu. Loading the “tr-*.data” file will automatically load the image and the projected traces in which the control points were selected.

The following are necessary for all digitized data.

- Compare the projected digitized traces with the underlying image and interactively correct poorly tracked trace segments (see Section 4.2).
- Enter the DPI. If the DPI is unknown, enter the physical length of the horizontal dimension of the scanned paper seismogram in “inches” in the *Amp, Length* dialog box.
- A least squares fit is used to determine the reference line for each trace. The linearity of each trace and the reference line third order polynomial fit can be clearly observed by changing the

horizontal:vertical image scaling ratio to 1:50. Change the order of fit if desired in the *Edit Menu*.

- Set the Edge Lines at both ends of the image. Setting the Edge Lines is necessary only when concatenating multiple traces, determining the total elapsed time in the seismogram, or when determining the Scale Factors (either the TSF or check the DPI). If timing mark gaps are absent, use the TSF determined from the entered Start and End Times in the *Export (Preview)* window or use the default pixels/min value (470, or an estimate from the closest chronologically available seismogram) in the *Time (min), pixels/unit* dialog box.
- *Export* and *Preview* to ensure digitization (and concatenation) of the traces was done properly. *Save Data* in case further adjustments are necessary at a later time. Adjust connection segments if needed by modifying the digitized data at the trace ends and recheck.
- Enter the required header information in the *Export* window, Start Time (and End Time for full seismograms), and the sampling rate. Adjust Edge Lines if needed. *Re-Export*.
- Save the digitized data in either or both SAC2000 Matlab file and standard SAC formats.

7.2 Generic Series: Time or Otherwise

Digitized data may be obtained for other images where the independent variable is increasing monotonically. Determination of the independent and dependent variable scale factors is made by entering the range of the independent variable and the physical dimension of the scanned hard copy in the *Length* dialog boxes, respectively. The remainder of the digitizing procedure follows that for seismograms.

References

- [1] Bromirski, P.D., R.E. Flick, and N. Graham (1999). Ocean wave height from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific, *J. Geophys. Res. C*, 104, 20,753-20,766.
- [2] Bromirski, P.D. and F.K. Duennebieer (2002). The near-coastal microseism spectrum: Spatial and temporal wave climate relationships, *J. Geophys. Res.*, 107(B8), 2166, doi:10.1029/2001JB000265.

8 Keyboard Shortcuts

Sometimes it can be more efficient using the keyboard to employ trace editing and digitizing functions instead of the mouse.

Notes:

1. All lower and upper case letters will have the same function unless some other functionality is explicitly stated.
2. Keypress functions only work when the figure is the active object. The keyboard will be ignored if any text fields, buttons, or sliders are selected. The exception is when using <ctrl> with a key.

[0-999] <enter> : Jump to line number (Ex: 16<enter> goes to line 16)

x / X : decrease / increase x-axis ratio

y / Y : decrease / increase y-axis ratio

<space> : Digitize Pt2Pt

> : Increment line number

< : Decrement line number

M : Reset line number to top

+ : Zoom in

- : Zoom out

I : Toggle Insert Ctrl Pt

D : Toggle Delete Ctrl Pt

L : Toggle Select Pts

E : Toggle Erase Trace

H : Toggle Shift Image

Z : Toggle Zoom Button

u, <ctrl> u : Undo

“.” : Toggle “Center” checkbox

? : Toggle “Current / All Lines” checkbox

 : Delete selected control points

Keyboard “arrows”: ↑, ↓, ←, → : Scroll image in main panel

` : Enable keyboard prompt debugging

9 *SeisDig* SAC2000 Structures

The *SeisDig* Matlab header structs do not include all those found in the Lawrence Livermore National Laboratory (LLNL) SAC2000 release. When in SAC2000, more header information can be obtained with “SAC> help mat”. A brief comparison of structs in SAC2000 and *SeisDig* is given below.

Identical to the SACdata structures in SAC2000:

struct times:

delta b e o a f
t0 t1 t2 t3 t4 t5 t6 t7 t8 t9
ko ka kt0 kt1 kt2 kt3 kt4 kt5 kt6 kt7 kt8 kt9 kf

struct station:

stla stlo stel stdp
cmpaz cmpinc
kstnm kcmpnm knetwk

struct event:

evla evlo evel evdp
nzjday nzyear nzhour nzmin nzsec nzmsec
kevnm mag imagtyp imagsrc

trcLen (integer type in SAC2000)

Structs Not included in SAC files exported from *SeisDig*:

user, descrip, evsta, llnl, response

New struct in *SeisDig* export files:

struct SeisDig:

num_pixels : total number of pixels traced
px_per_minute
samples_per_sec
px_per_delta
original_px_per_minute : value listed in SCALE FACTORS box in main *SeisDig* window
sampling_method: always “nan_spline”
drum_period : value listed in SCALE FACTORS box in the *SeisDig* window (minutes)
physical_length : value listed in SCALE FACTORS box in the *SeisDig* window (inches)
dpi : dots per inch
dpmm : dots per mm
image_filename : filename of image file digitized