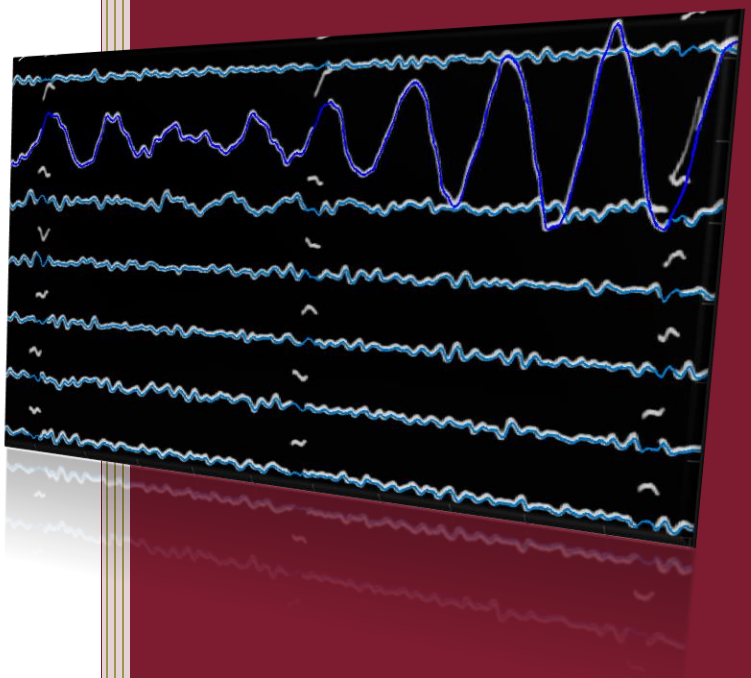
2015

DigitSeis v0.53



Petros Bogiatzis and Miaki Ishii

Harvard University

For any questions, bug reports, suggestions, comments, and requests of new features please contact with Petros Bogiatzis at *petrosbogiatzis@fas.harvard.edu.*

Introduction

DigitSeis is a digitization software that allows the conversion of digital raster images of analog seismograms to digital time series data (Figure 1). It is written in MATLAB (http://www.mathworks.com/products/matlab/) and it is uses image processing algorithms to automate whenever possible the digitization procedure. It is best to be used with MATLAB version 8.2.0.701/R2015b. The software is developed under the assumption of human supervision and input at special cases such as trace crossings. DigitSeis vectorizes the raster seismogram images, and in addition, it analyzes and corrects for various distortions and synchronizes traces by identifying and utilizing the time mark offsets. The results are saved in readily usable format in MATLAB or SAC.

The following files are also needed to be in MATLAB's path:
"*rgb2hsv_fast.m*" that can be found at
http://www.mathworks.com/matlabcentral/fileexchange/15985-fast-rgb2hsv/content/rgb2hsv_fast.m
"writesac.m" that can be found in SAC distribution, under sac/utils:
http://ds.iris.edu/ds/nodes/dmc/forms/sac/

The conversion procedure from raster image to digital time series is separated into the following stages: Preparation, Identification/Classification, Timing and Digitization. They are described in detail in sections below.

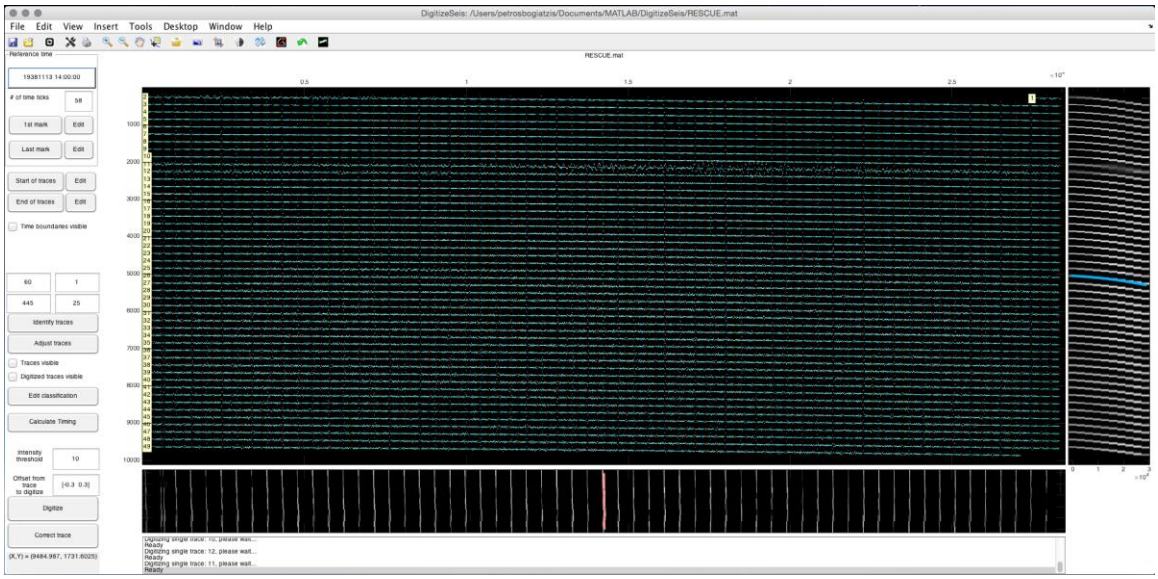


Figure 1*: Screenshot of DigitSeis working environment.*

Loading scanned seismogram

DigitSeis supports all image formats that are currently supported by MATLAB (e.g., for version 8.2.0.701/R2015b, the supported file extensions are: bmp, cur, fts, fits, gif, hdf, ico, j2c, j2k, jp2, jpf, jpx, jpg, jpeg, pbm, pcx, pgm, png, pnm, ppm, ras, tif, tiff

and xwd). The scanned image should represent the waveform with dark pixels over high intensity background (e.g., Figure 2) or the opposite way. In the first case the user should work in the complement (negative) of the image when asked during loading of the image. The user can change the image polarity after it is loaded using the  button. The input image should be either in RGB (Red, Green, Blue), HSV (Hue, Saturation, Value), or single channel (gray scale) format, and of any color depth higher than or equal to 8-bit. The image could be either in the form shown in

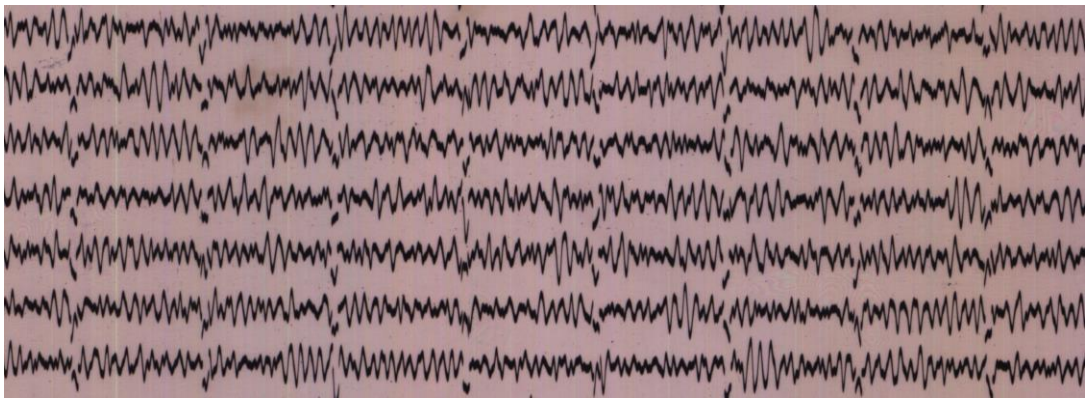


Figure 2: *Example showing an example of an input image. Waveforms are represented with darker pixels compared with the high intensity background. The opposite case is also supported.*

To load the image file, click the "*open*" tool,  , browse to the directory of the desired image, select it and click open. When open is clicked in a valid image file, the program reads the image and prepares it for digitization. More specifically, initially it converts it to an 8-bit (256 different intensity values per channel) image. This color depth provides an adequate balance between efficiency and preservation of information. Since the distinction of the signal from the background is based on the intensity rather than the actual color, therefore if the image is in RGB format it is first mapped into HSV color space and only the V channel is kept that represents the luminance information. Next, the complement of the image is computed or not depending the original form of the image, and finally high intensity values represent the signal and dark values the background. This configuration is more appropriate for the subsequent phases of the digitization as the signal is expressed with high values of intensity while the absence of signal with zero value (dark). Finally, a histogram correction is applied automatically to the image before it is shown to the screen to enhance the waveform over the background and to suppress intermediate intensity values that are associated the exposure, the long-term storage, and the scanning procedure (Figure 3). Again the user can select not to apply the histogram adjustment if the image is already in good condition.

During this procedure informative messages appear at the output message box, informing the user for the tasks that are currently performed. The user should always wait for the indication "Ready." before proceeding to any further task.

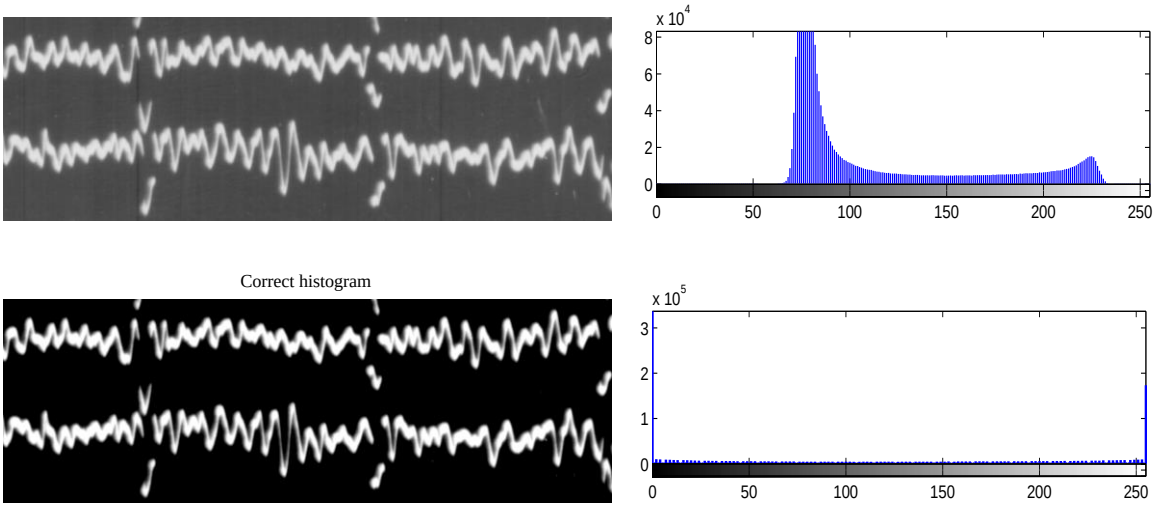


Figure 3: *Example of the automatic image enchantment during the preprocessing stage. In this particular case the histogram of image was corrected to improve the contrast and depress issues associated with the scanning procedure or the paper. The left column shows the image before (top row) and after (bottom row) the histogram correction, while the right column shows the corresponding histogram of the intensity values of the image.*

In cases where the image contains large stains, and/or the contrast varies spatially, the "*remove background*" functionality can be used by pressing the button. DigitSeis uses Gaussian filtering to produce a smoothed version of the image that captures the background long wavelength variations of intensity, and then subtracts this layer from the initial image, leaving only the useful information. The user should input the vertical and horizontal standard deviations of the Gaussian distribution and then use the "*preview*" button to inspect the result. On the top of the figure that appears after pressing the "*remove background*" button is shown the initial image, and at the bottom the result of the background removal. In order to apply changes to the original image the user should press the "*apply*" button.

An older saved analysis (see also the Save results paragraph) can be loaded with the '*load analysis*' button, .

Common image navigation and viewing tools

DigitSeis incorporates some basic viewing tools that the user can activate at any time of the program execution. These tools include "*zoom in*" and "*zoom out*" functionality, , an image mouse-based "*panning*" tool , the "*data cursor*" tool , and the "*zoom reset*" tool . Additionally at the bottom left area of the program window, it is displayed the current position of the mouse in pixels coordinates, within the image.

Preparing for digitization

Cropping

Typically, the scanned images include around the seismogram a lot of additional empty space. This unnecessary space consumes memory and may introduce artifacts therefore it should be removed prior to any further process. This can be done by using the "*crop*" tool,  . The user should make sure not to crop the image to narrow but rather leave some space (approximately equal to the distance between two subsequent traces) around the traces.

Further correcting contrast by adjusting the histogram

DigitSeis provides to the user the option to further interactively correct, if it is needed, the contrast and the brightness of the image (Figure 4), by clicking the tool from the "***adjust contrast*" tool**,  , at the main toolbar. To apply the changes in pixel values in the target image, click the Adjust Data button, which remains unavailable until you make a change to the histogram of the image. In the ideal case seismogram should be represented with white and bright grays and the background should be as dark as possible. This tool is based upon MATLAB's emended Adjust contrast tool (type *help imcontrast* for more information).

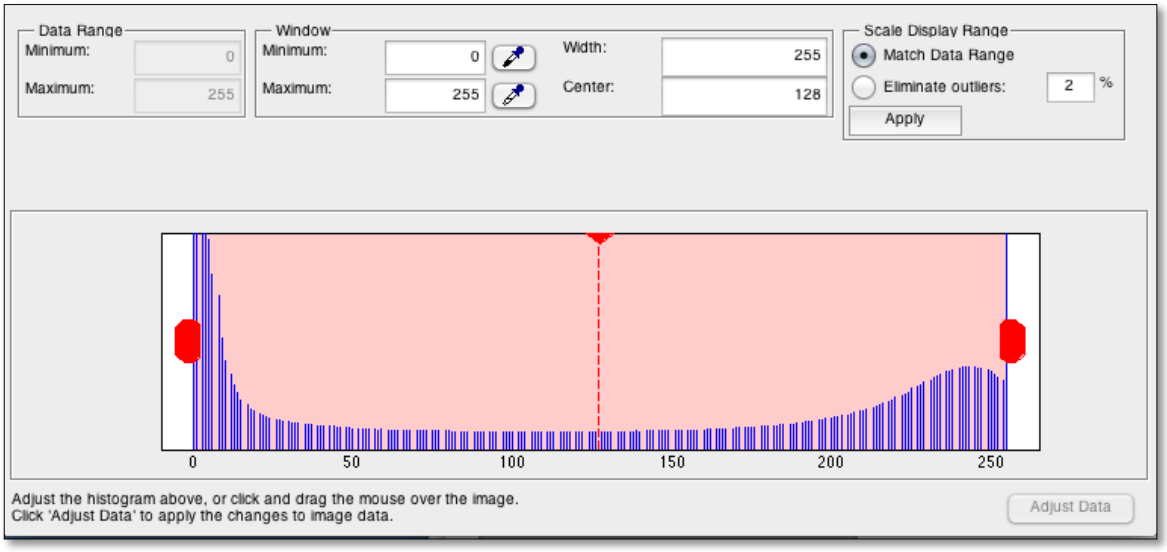


***Figure 4**: Interactive tool for further adjusting the contrast and the brightness of the scanned image.*

DigitSeis provides a shortcut to calling MATLAB's *imtool* function by clicking the "***view info and edit image*" tool**,  , for viewing information and interactively editing the image. By default the image that is loaded is the output of the *open tool*, however the user can use the provided interactive menus and tools to open and

explore the initial input image from the scanned seismogram file. This tool is independent from the rest of the program therefore, any processing done here is not reflected to the main program. However, the user can save the changed image and then reopen it from the main program.

Correcting for rotation

If the scanned image presents significant rotation from the horizontal, the user can use the "***correct rotation***" **tool**, , for correcting that. In most cases rotation corrections is not required. This tool uses radon transform of the image and attempts to identify automatically the optimum correction angle (Figure 5). Initially it uses searches every one degree, from -8 to +8 degrees with respect of the current rotation. Afterwards it refines the step to 0.1 degrees an repeats the procedure.
The user can accept the suggested rotation or input a different one. In general rotation correction is not needed and should be examined only if there is some systematic problem at the digitization outcome or severe rotation of the scanned image.

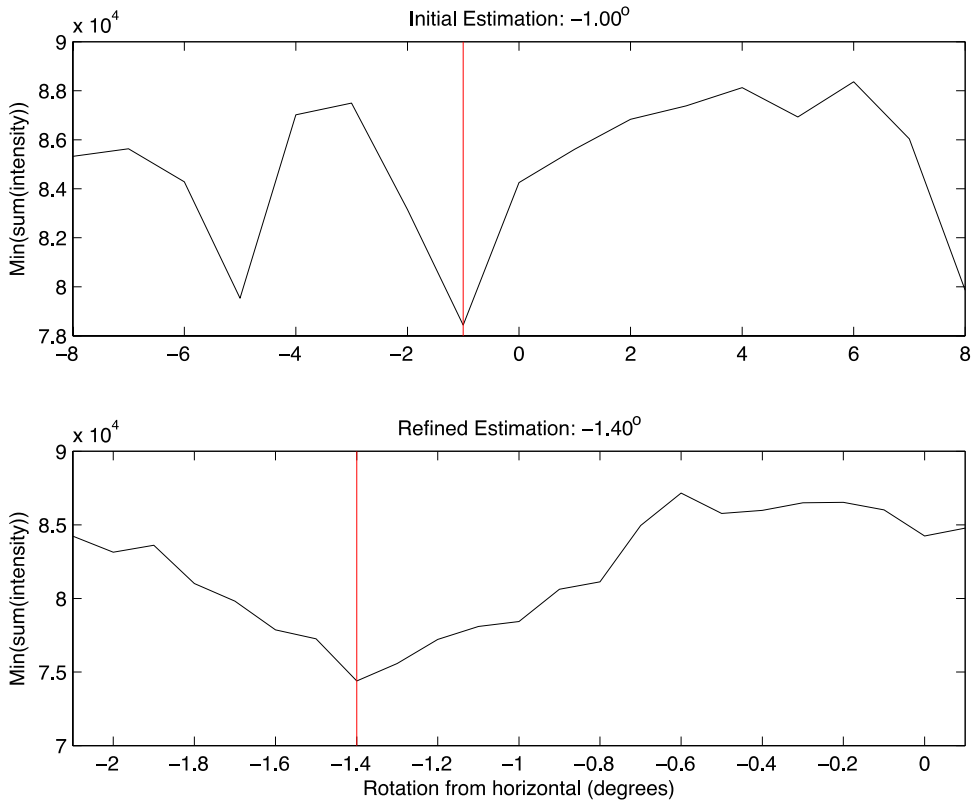


Figure 5*: Example figure from the rotation correction evaluation for different rotation angles. The first row corresponds to the initial search every 1 degree while the second, to the refined search around the solution yielded the previous stage. The Optimum rotation angle for each stage is marked with the vertical line.*

Removing regions

It is common that the scanned images include dark pixels (show bright at the working version), which are not part of the seismogram but they rather related with other factors such as timing notes that have been written between the traces, stains on the paper and noise from the scanning process. Some of these issues can be resolved with the contrast adjustment procedure while others are automatically discarded during the digitization process, with the condition that they are faint enough compared with the seismogram. However, the rest should be removed using the "***remove region"* tool,** , or to be marked as objects to be rejected from within the edit classification window (see Classification paragraph) to avoid infecting the digitization (Figure 6).

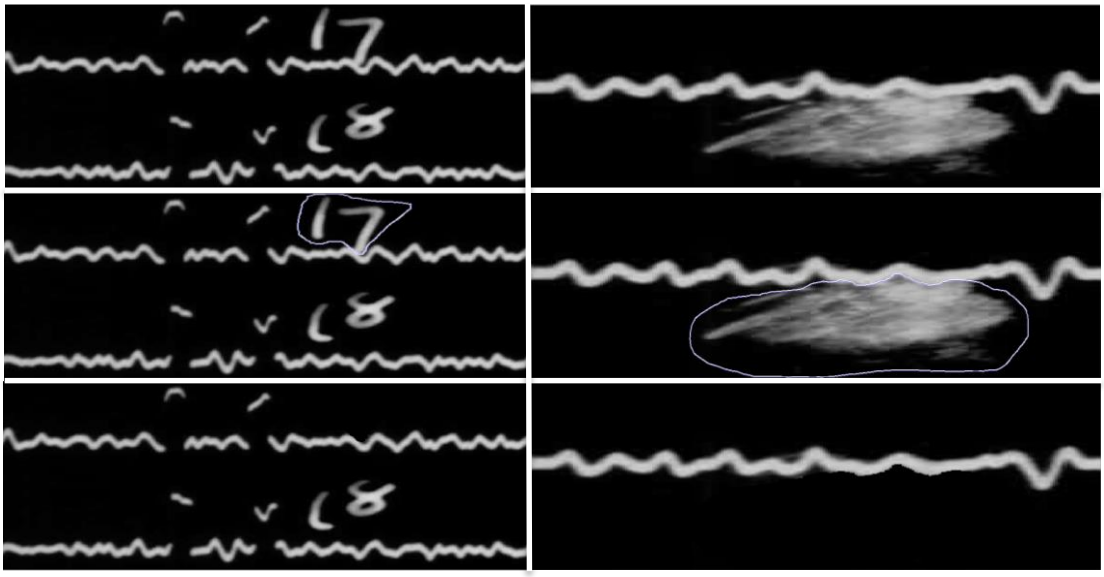


Figure 6*: figure showing two examples with regions that the user may remove prior to digitization. The left column shows a case with timing notes and the right column a case with a stain on the paper. First row shows the initial image, second column shows the region selected by the user to be removed, and the last row shows the image after removing the selected region.*

Finally, make sure that you have properly set the *toggle button that indicates the direction of the time mark offsets.*

Identify traces and time marks

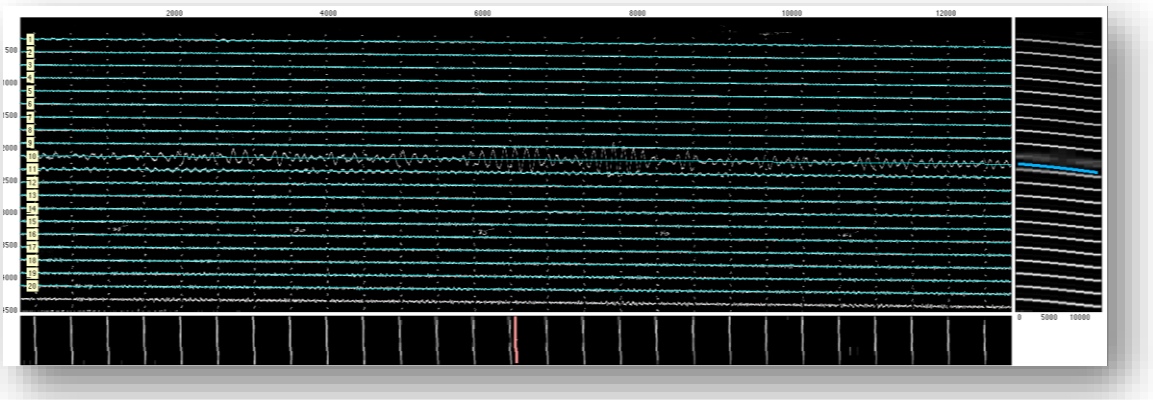
By pressing the *"**Identify traces & time marks**"* button, DigitSeis prepares the digitization of the image by performing the following tasks:

1) Automatically detects the position and the number of seismic traces (trace 0-lines) within the scanned image. Furthermore, it estimates the horizontal distortion of the seismograms that is to be corrected during the digitization process.

2) Classifies the objects shown in the scanned image within the following three categories: Main trace, time mark, object that should not be digitized (e.g., stains, notes etc.).
3) Identifies the vertical distortion of the seismograms that also should be corrected during the digitization and the timing process.
4) Assigns the objects that correspond to either normal trace or time mark to traces.

The output is plotted as curves that mark the traces locations (Figure 7). Additionally, two diagnostic plots are created to show the trace and the time mark distortion at the horizontal and the vertical directions. The user should also enter the approximate length of the time marks in pixels, the distance between successive time marks both in pixels and in seconds, and finally, the time separation between traces in hours (Figure 7b).

(a)



(b)

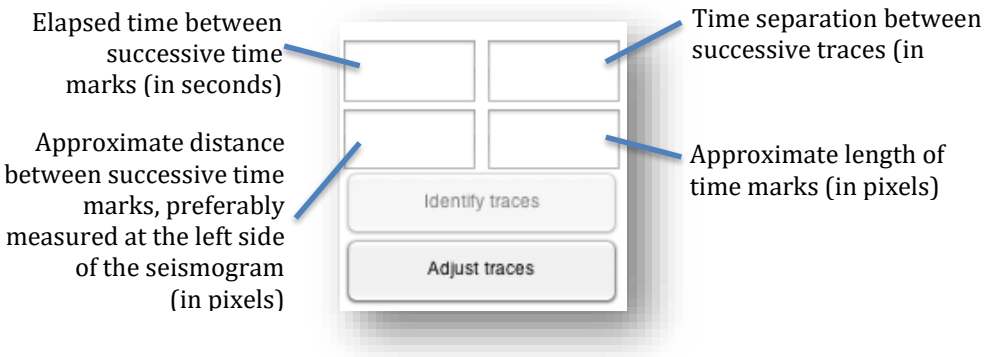


Figure 7: *a) This figure shows the seismogram and on top the identified traces (cyan lines) with their labels. The bottom panel shows the distortion of the time marks modeled with the red curve. The right panel shows the horizontal distortion modeled with the blue curve. b) The fields where the user should enter information that will be used for the classification and the timing of the traces.*

View – edit classification

The user can see graphically and edit interactively the classification result by pressing the button with label "***Edit classification***". A new figure opens showing a pseudo-color version of the scanned image (Figure 8) where the colors indicate the classification of each object (white for main trace, green for time mark, and red for objects that should not be considered to the digitization).

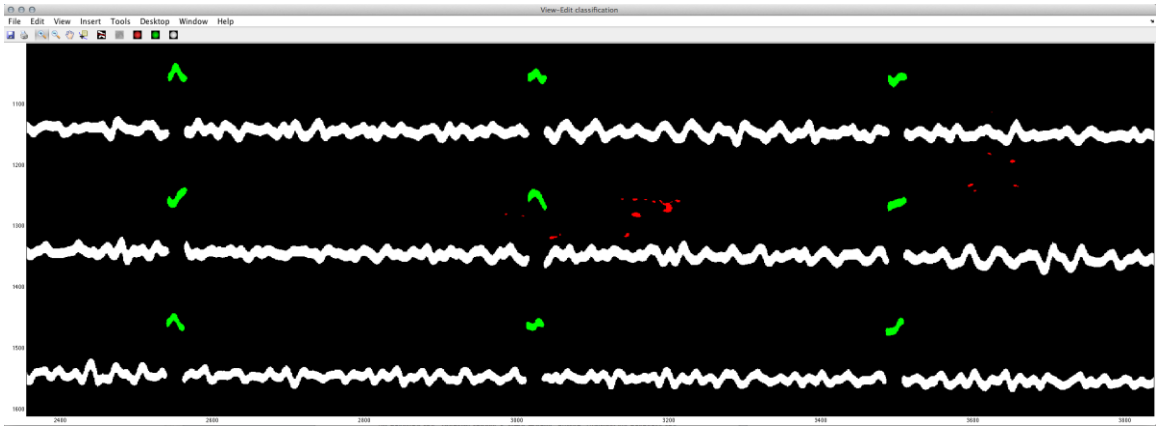


Figure 8*: Example of the view – edit classification figure.* Colors show the classification of each object i.e., white for main trace, green for time mark, and red for objects that should not be considered to the digitization.

The toolbar of this figure includes some common tools that can be used to view the current classification, but also some additional tools that can be used for editing the classification. More specifically, the **"*remove pixels along path* tool",**  , can be used to separate two objects that currently considered as a single one (e.g. a time mark that connects with the main trace). When this option is engaged, the user can draw an arbitrary path by moving the mouse and pressing the left button at the same time. When the left button is released, the pixels that are along the drawn path are set to zero and the automatic classification process is repeated (it may take some seconds). **After completing the separation of objects** that shouldn't be connected, the next stage allows manually correcting for classification errors that may exist from the automatic algorithm. Each object that is mistakenly tagged, can be manually re-classified to one of the three categories mentioned above, by first clicking at the appropriate among the three classification buttons in the toolbar i.e.,  for main trace,  for time mark,  for object to be rejected, and then by clicking within the object to be re-classified. During this stage the classification is not repeated for the whole image, therefore the update is faster. Note that if the user activates the "*remove pixels along path*" tool the selections that done manually are lost. Note also that none of the operations performed in this stage reflect the original image but only the classification masks are updated. In order to reset the classification edits the user can press again the *"Identify traces & time marks"* button located at the main window of the program. The button  , can be used if the user

wishes to use a different from the original image for the classification procedure. Note that the two images should have identical dimensions and they should be co-registered i.e., a pixel in one should have the same coordinates as the same pixel in the other image. There is also an alternative, color blind friendly color scheme that can be selected using the  button. In this case time marks appear with a light blue color.

Edit object assignments.

To digitize the traces DigitSeis assigns objects to traces based upon the distance of the center of their bounding box from the traces 0-lines (see also the Digitization paragraph). The user has the option to view the assignments in the classification window, by clicking on the checkbox at the right side of the toolbar. The assignments appear as numbers in the center of the bounding box of each object. The number corresponds to the trace that the object is been assigned to. The user can change the initial assignment of an object by clicking on its label and then type in the new number. The  button can be used if the user wishes to recalculate the assignments from the beginning.

View and edit the identified traces

After engaging the *"Identify traces & time marks"* button the program estimates automatically the number of the traces and their approximate location. Nevertheless, there is the possibility some traces not to be recognized (e.g., short traces that start very close to the right edge of the paper, faint traces etc.) or to falsely recognize traces that do not exist (e.g., if the paper includes horizontal elongated stains). All these can be corrected by pressing the **"Adjust traces"** button. This shows a new window (Figure 9) from which the user can add traces (using the + button), remove traces (by unchecking the corresponding box of last column), and type new $x-$ and $y-$coordinates for the first point of the trace and/or new x coordinate for the last point of the trace. The $y-$coordinate of the last point (and all others except first) is calculated automatically based upon the horizontal distortion of the traces.

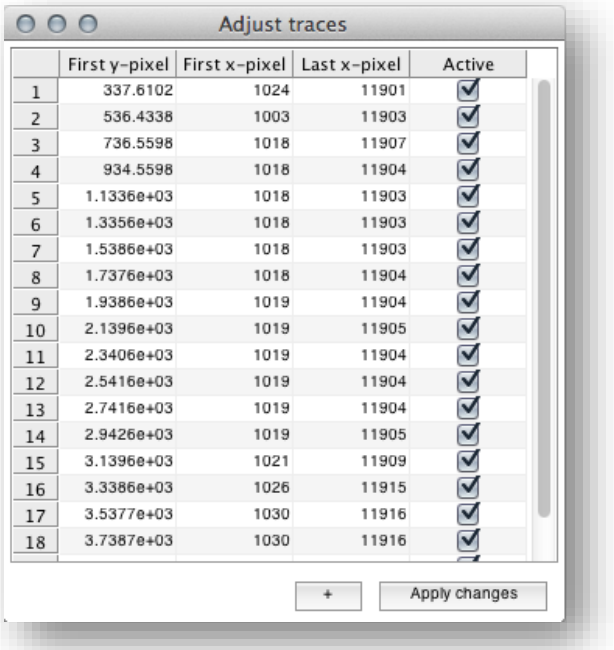


Figure 9: *Example of the "Adjust traces" window where the user can manually change the number, and the location of the traces. Each row corresponds to a different trace. First column shows the a/a number of the trace, second column the starting $y-$coordinate, third column the starting $x-$coordinate, the fourth column shows the ending $x-$coordinate and the last column is used to mark the traces that should be kept and those that should be removed.*

Timing of the traces

Initially the user should determine accurately the exact starting and ending pixel of each trace. This is done semi-automatically by clicking on the "***Mark start of traces***" and "***Mark end of traces***" buttons for the left and the right boundary of the traces respectively. In each case, DigitSeis estimates the position of the rest of the starting/ending points automatically based on the trace positions and the change intensity that occurs in the beginning and the ending of the traces. Finally, it plots an interactive polygon line that includes one vertex for each trace. The user has the ability to correct the position of the vertices by dragging each one of them to the right starting/ending position of the corresponding trace (Figure 10). This procedure finalizes by double-clicking on any of the points. The boundary is drawn with green color for the starting pixels and with red color for the ending pixels. Changes can be made by activating the "*edit*" buttons, which are located next to the "*Mark start of traces*" and "*Mark end of traces"* buttons respectively.

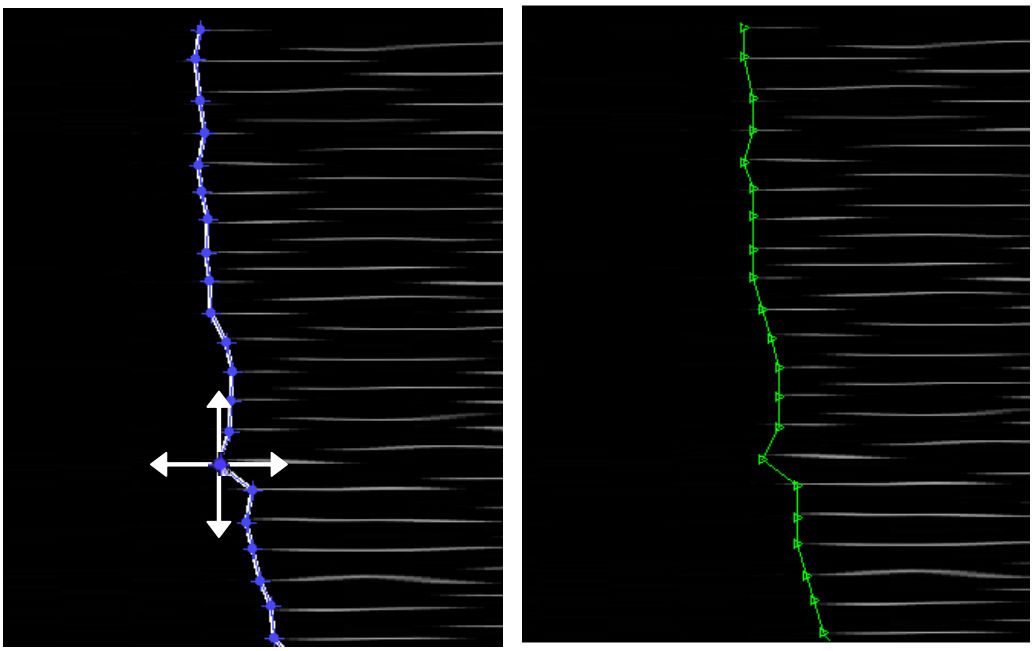


Figure 10: *Example from the marking of the start of the traces task. After clicking the first pixel of the first trace an inderactive polygon line is drawn. The user can apply corrections be draging each point to the first pixel of each trace. The marking is finalized when the user double-clicks at any point. Similar is the marking of the end point of each*

Next, the user should select two columns of time marks. Ideally, these should be the earliest and the latest, but the user should ensure that the selected reference columns include as many usable time marks as possible. If the selected column includes many problematic time marks such as those that are washed out and are truncated and appear on both sides of the seismogram, the user should select the next (or previous) column. The selection is performed by clicking the "**1st time mark**" and "**Last time mark**" buttons respectively. In each case, the user should click with the mouse at the last pixel of the earliest time mark of the selected column (i.e., the top one). Then the program approximates the position of the rest of the time marks by using the estimated distortion, and draws an interactive polygon line that includes one vertex for each time mark. The user has the ability to correct the position of the vertices by dragging each one of them to the right ending pixel of the corresponding time-mark, in a similar way to the marking of the start and the end of the traces (e.g., Figure 10). By double-clicking on the polygon, the positions are finalized. Changes can be made by activating the "*edit*" buttons, which are located next to the "1st time mark" and "Last time mark" buttons respectively. The user should also provide the number of time marks columns between the reference ones (including them) in the appropriate field (e.g., Figure 11).

The main information about the timing of the seismogram comes from the time marks in combination with the start and sometimes the end time and date that are usually written somewhere on the original seismograms (usually in the back side of

the paper). These reference times and dates are approximate and they serve as a guide for the timing of the first and the last time mark time and consequently, the rest of the time marks. DigitSeis requires as input from the user the date and time of the first time mark from the left reference column (i.e., the "*1st time mark*" column). The date and time can be entered using the corresponding "*Reference time*" button within the *Timing tools* frame (Figure 11).

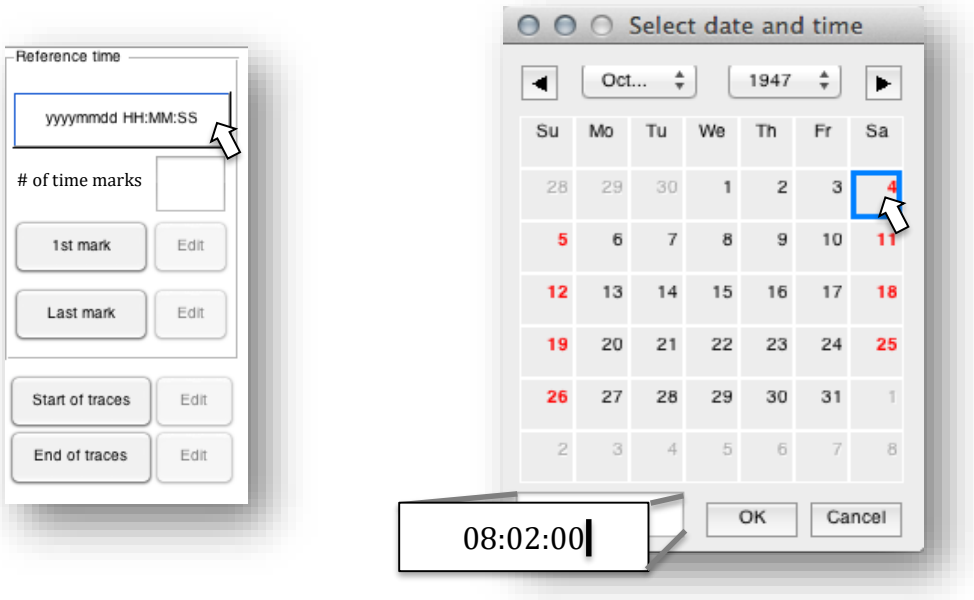


Figure 11*: The reference time button (left), and window that allows the user to ender the reference date and time (right). After the calendar window is closed, the selected date and time (if any) is shown on the button in the format: yyyymmdd HH:MM:SS. The buttons for selecting and editing the reference time mark columns, and also the start and the end of the traces are also shown on the left. The field where the user enters the number of time mark columns between the reference columns is shown also shown (# of time ticks).*

When all required information are entered, the timing is evaluated by pressing the "*Calculate timing*" button. The result will show on the seismogram as shown in Figure 12.

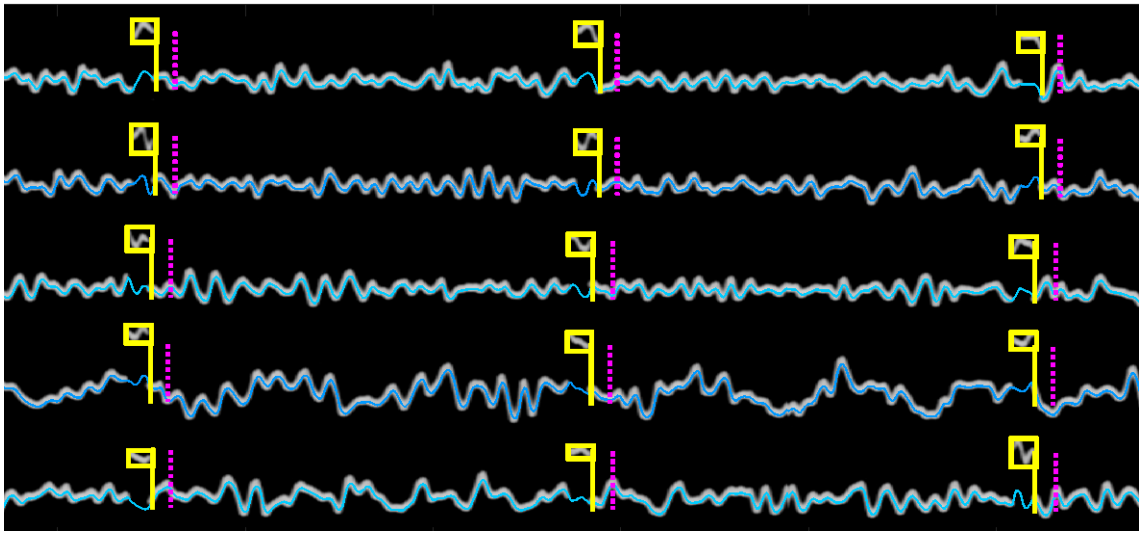


Figure 12: *Example from the timing of the traces. Yellow rectangles, indicate the identified time marks, magenta dotted lines show the theoretical timing that estimated using the distortion model, and yellow solid lines show the corrected timing based upon the location of the time marks.*

Digitization

Before digitizing the user should define a luminance threshold between 0 and 1 that controls how bright should be a pixel to be included in the digitization. Values close to 0, may yield noisy images. On the contrary, very large values (~1) can cause some faint parts of the seismogram to be ignored. The user can try different values and see which is appropriate for the specific case.
Additionally, the user should define the digitization-offset limits (the same for all traces) as a pair of numbers from 0 to 1 within brackets (default [0.3 0.3]). The first number controls the offset lower and the upper extend of the strip that is calculated from the trace 0-line. The offsets are calculated as a portion of the average distance between succeeding traces.

By clicking the "Digitize" button, DigitSeis initiates the automatic digitization procedure. The user can watch the progress of the digitization from the small clock that appears at the bottom right corner of the program. Similarly, with other tasks the user is informed from messages presented within the output message box. For each trace that is finished the corresponding seismogram is plotted on top of the image so the user is able to evaluate the digitization quality (Figure 13).

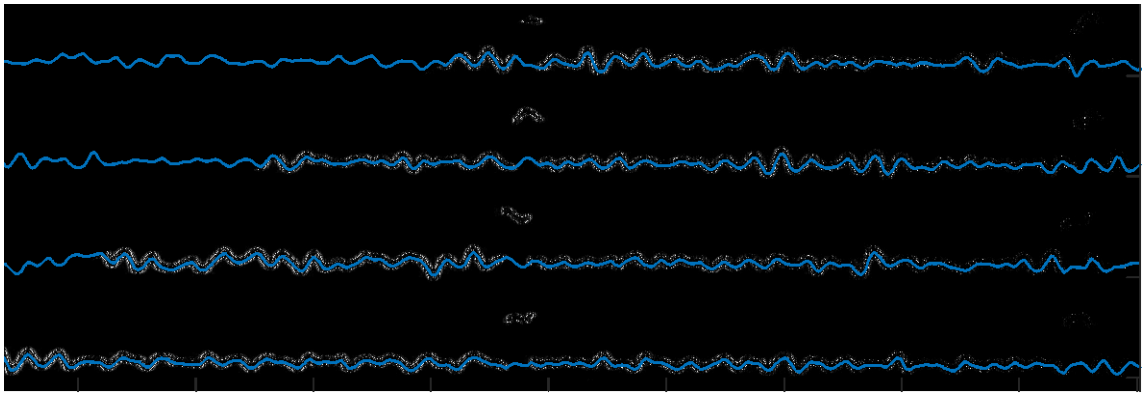


Figure 13*: Example of the digitization result, plotted on top of the seismogram image. Note that time marks are also digitized, and the offsets have been corrected.*

In cases where some trace exceeds the digitization limits (e.g., large amplitudes that cross with adjacent traces) the user may need to assist the digitization of this specific trace. By clicking the "**Correct trace**" button a message suggests the user to click at the center (mean value line) of the trace that is to be digitized separately. By doing that, a vertical line is plotted that marks the vertical limits of the trace strip. The user may have to adjust these limits by dragging the ending points of the line accordingly. The final position is set be double clicking to any of the points, and then a new window appears that shows the selected trace (Figure 14).

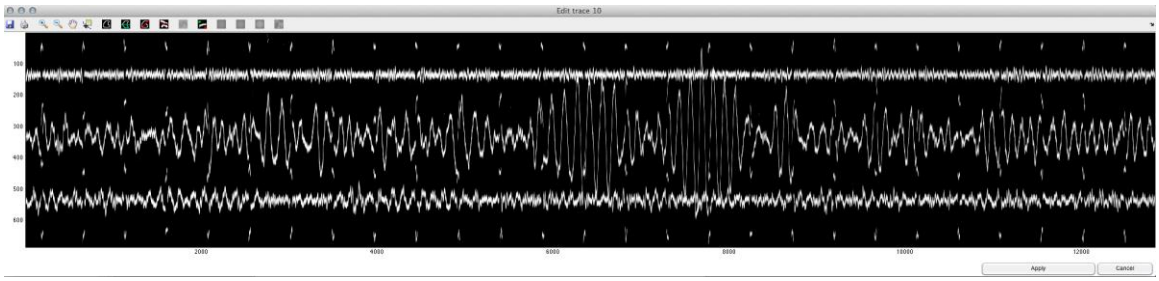


Figure 14*: Example from the single-trace digitization window. The tool bar includes interactive tools that allow the user to remove the pixels that do not belong to the selected trace.*

The toolbar of the new window includes tools similar to those described in the preparation section, that allow the user initially to remove all pixels that do not belong to the specific trace, and afterwards to correct the automatic classification of the objects if it is needed. Finally, by pressing the "*digitize*" tool, the digitization of the trace is performed. If it is acceptable it is applied to the rest of the analysis by pressing the "*Apply*" button. Pressing the "*Cancel*" button will have as a result to return to the main analysis window, ignoring the work done on the specific trace.

Save results

The results are saved in both SAC and MATLAB's "mat" format. The analysis can be saved at any stage by pressing the *save analysis* button,  . In the window that appears, the user should ender the path and the name of the mat-file that will store the analysis. After saving the analysis in "mat" format, a new window appears that asks the user if SAC files should be generated also. This option is valid only after the digitization and the timing of the trace has been concluded. If it is selected, an interactive window shows the headers of sac files and allows the user to complete various fields that are related with the station, the instrument, and other information. To produce SAC files DigitSeis uses the function writesac.m that is included in SAC distribution (under sac/utils).

he 'mat' file can be later used to load a saved analysis through the "*load analysis"* button ,  .

Work Flow

1) *Load and prepare the image (crop, correct contrast, remove noise, stains, notes, etc.).*
   These tasks can be done in an image processing software as well and then use the rectified image as an input for DigitSeis.
2) *Fill the fields shown in Figure 7b, and press Identify traces button. Make sure that you have properly set the*  *toggle button that indicates the direction of the time mark offsets.*
3) *Use Adjust traces button to correct add-remove zero-lines if needed.*
4) *Press the Edit classification button to review and correct the classification.*
5) *Mark the start of the traces, double-click to finalize, and afterwards the end of the traces as well.*
6) *Mark the very left column of time-marks, double-click to finalize it and afterwards the very right column as well.*
7) *Complete the date and time of the top time-mark of the left column you have marked. Complete also the number of time marks in each row including the columns you have selected. Double time marks that indicate the hour should be counted as one.*
8) *Press Calculate timing and inspect for problems*
9) *Press Digitize and inspect for problems.*
10) *Correct Digitization of specific traces if needed.*
11) *Save results and SAC files.*

Acknowledgment

This project was supported by the U.S. Geological Survey Earthquake Hazard Programs G14AP00016 and G15AS00037.