Automated Tools for Layer Recognition (BMPix) and Counting (PEAK)

General remarks and further information

The tools presented here are detailed in the publication:
Weber, M.E., Reichelt, L., Kuhn, G., Pfeiffer, M., Korff, B., Thurow, J., and Ricken, W. (2010): The BMPix and PEAK tools: New methods for automated laminae recognition and counting – Application to glacial varves from Antarctic marine sediment. Geochemistry, Geophysics, Geosystems, 11(1), doi:10.1029/2009GC002611.

The original idea for the tools came from Lucia Reichelt:
Entwicklung neuer Methoden zur hochauflösenden Untersuchung von Umweltvariationen an spätquartären Sedimenten aus dem Weddellmeer, Antarktis. Unpublished diploma thesis, University of Cologne, 123 pp., 2007).

The source code was written by Björn Korff.

Please check this website for updates!

System requirements

The tools consist of a set of macros written in Visual Basic. They can only be used in conjunction with Microsoft Excel. Thorough and reliable tests have been performed for Excel 2000 (version 9) through 2007 (version 12) in a Windows XP environment. Functionality should also be given for various Macintosh versions of Excel (except for Excel 2008), although this platform was not extensively tested. The vertical screen resolution has to be at least 768 pixel.

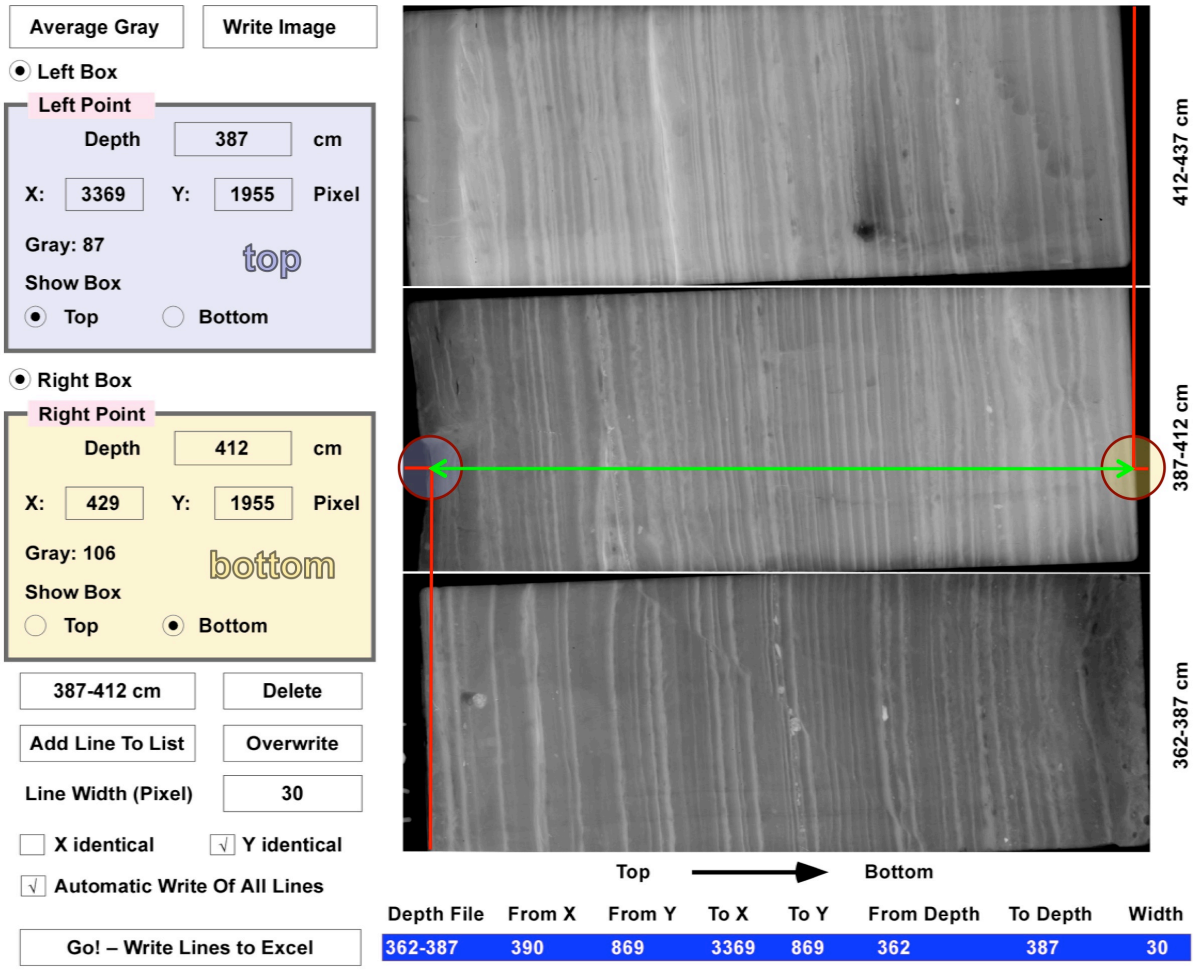
Overview

1. The BMPix Tool

General purpose

The BMPix macro linearly extracts red, green, blue (R/G/B), and gray values at pixel resolution along a profile line between freely adjustable start and end points (circles in Figure 1) from any given image that is stored in the BMP file format. This means that color and gray values can be generated along a profile line from surface images (e.g., photos, line-scan images) and from transmitted light images (e.g., x-radiographs, thin sections). Note that when the term “core” is used in the description, it applies to an ocean sediment site since most of the measurements have been carried out on this type of climate archive. However, the same is true if you apply the method to different archives, e. g., limnic cores, ice cores, tree rings, and corals.

Figure 1: *Input mask of the BMPix tool. RGB and gray-scale data are generated at pixel resolution along a green profile line between pre-defined start (top, blue) and end (bottom, yellow) points for any given BMP image. Here, 2,940 data points (pixel 429 – 3369) were measured over a 25-cm long x-radiograph image (387-412 cm), i.e., the resolution is 12 measurements/mm or 85 µm. White numbers in the list to lower right contain measurement data from previous profile line in section 362-387 cm. Upon completion, gray-scale data will be extracted from all profile lines added to the list.*



Step-by-step instruction

Double-clicking the “BMPix_macro_2009.xls” icon will launch MS Excel. If not, it can be manually started through Tools/Makro/Makros/Run. Note that macros have to be enabled in the Excel preferences. The macro asks for the BMP image file that should be used for the analysis. Note that the image file name has to have exactly nine digits, whereby the last digit needs to be an underscore, followed by the extension (e. g., “AB-cd123_.bmp”). Select the image in the opening dialogue and click ok. The input mask will load (see Figure 1), containing the image to the right and a number of possible settings to the left (note that Figure 1 shows an 8-bit gray-scale image). In the title window the dimensions of the image (in pixel) is displayed.

Start by choosing the “Left Point” in the “Left Box” in pixel, i. e., the starting (“Top”) point for the analysis. This can be achieved by either using the vertical and horizontal scroll bars or by typing in the desired coordinates for “X” (3369) and “Y” (1955). In both cases, the conjunction of the red bars will display the position of the starting point with the “Gray” value (87) at that point (note that a 24-bit BMP image file will display RGB values here). The next step is to associate this point with a “Depth” (387). The same should then be done for the “Right Point” in the “Right Box”, i. e., the ending (“Bottom”) point for “X” (429), “Y” (1955), and “Depth” (412). Note that “Show Box” allows you change “Top” and “Bottom” to account for a reverse orientation of the image.

The chosen depth interval (387-412 cm) is automatically displayed underneath the two box fields. If desired, the name for the interval can be changed. Note than replicate names are not allowed. You can “Delete” these entries and redo the depth assignments, or you can add the data from 387-412 cm (“Add Line To List”) to the list below (highlighted here in blue), which already contains data from a previous measurement from 362 – 387 cm. Using this procedure, it is possible to perform multiple measurements for a single image with different settings, to segment lines along a single image into multiple parts, or to add lines from consecutive images in order to achieve a complete profile.

The “Line Width (Pixel)” defines the number of pixel over which the measurement is averaged perpendicular to the scanning direction. The minimum value is 1. Higher values can be chosen depending on layer geometry and stratification. “X identical” and “Y identical” allow for locking the coordinates from one measurement to the next.

The “Write Image” field at the top will save a copy of the BMP image file containing the centered position of the profile line. This step is required if PEAK (see below) is to be used. This line is only 1 pixel wide, but the actual width is indicated at the top and bottom as a crossline. The file name will add the depth interval (387-412 cm) to the original file name of the image and store it in the same folder. The “Automatic Write Of All Lines” field in the lower left will automatically save the copy of the image once the “Go! – Write Lines to Excel” button is pressed. You can also store the average gray value of a predefined area, for instance for calibration purposes, by clicking the “Average Gray” field at the top, adjusting both “Left Point” and “Right Point”, clicking “Average Gray” again, and waiting for the value to display.

Once all desired lines are added to the list, the command “Go! Write Lines to Excel” will execute the macro and extract RGB and gray values for all profile lines that have been defined. The data will be written to a separate Excel spreadsheets after a file name has been specified for the workbook (note that the name has to have the same nine digits as above, e. g., “AB-cd123_.xls”). Also, the list containing the start and end points will automatically be stored in the same folder as “AB-cd123_.bmp.list”.

The name of each spreadsheet reflects the depth interval (362-387 cm, 387-412 cm, etc.) and has to be unique. Each spreadsheet contains X and Y coordinates, the depth that has been linearly interpolated along the profile line between the top and the bottom, extracted gray and R/G/B values, as well as normalized gray values. The format for R/G/B was chosen to allow for direct color or gray-scale visualization in the PANGAEA software PanPlot. The normalized gray value is important if the PEAK tool is to be used.

We recommend leaving all resulting files in the folder where the original image is stored. In case the analysis has be modified or redone, all required information (i. e., the copy of the image containing the profile line(s), and the list file) is ploaded again. Changing a profile line, for instance, can then be achieved by starting the macro again, loading the original image, double-clicking the profile line in the list, modifying its settings, and clicking “Overwrite” and then “Go! – Write Lines to Excel”. In case a profile line should be deleted, click the line in the list, and then click “Delete”.

Trouble shooting

Although problems may vary depending on the Microsoft Excel version and operation system used, there are a few tips we can provide:

- If the BMPix macro does not run, make sure that the depth intervals have no replicate names.
- If the input mask or the BMP image is not entirely displayed, try increasing the screen resolution.
- If only gray values are extracted, the image is only 8 bit instead of 24 bit. Try converting into 24 bit if desired.
- If Excel does not open macros you have to change the security setting under Tools/Makro/Security to at medium or low, and allow for macros when asked in an opening dialogue. Alternatively, you if you use Excel 2007, it will not support macros at all. Then you will have to switch to an older (2000-2002) version.

2. The PEAK Tool

General purpose

The PEAK macro is designed to provide automated counts of the overall amount of layers present in an archive, based on the variability documented in gray-scale curves. Through defining the minimum height and length of a gray-scale peak, the macro counts all layers present. “PEAK_max_count” counts every bright maximum of a bright and dark layer succession (couplet), whereas “PEAK_zero_count” counts every bright and dark layer.

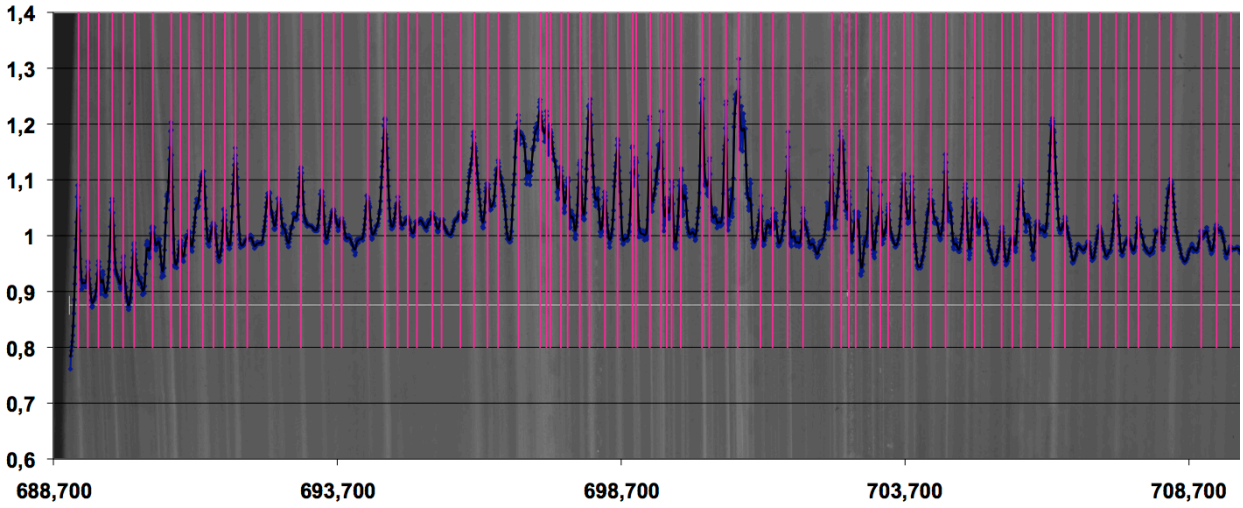
Step-by-step instruction

Some preparations have to be made before PEAK can be used. First, gray-scale curves have to be available. Since we use PEAK in conjunction with BMPix, the data must be in the format provided by BMPIX (i.e., at least depth has to be in column C, and normalized gray value in column F). Second, since we use visual control, all copies of the BMP images containing the position of the profile line (e.g., AB-cd123_387-412 cm.bmp; see description BMPix tool) should be converted to JPG (note that the file name has to be preserved; e. g., AB-cd123_387-412 cm.jpg) and stored in a newly created folder called “JPG”. This folder has to be placed in the same folder as the Excel workbook containing the spreadsheets (e. g., 387-412 cm) with the gray scale curves.
When the preparations are made you should load the PEAK_max_count macro. Then you should open the Excel workbook (e. g., PS1599_1) containing the gray-scale data that is separated into individual spreadsheets. Note that depth and normalized gray values have to be in columns C and F, respectively, in every single spreadsheet. To execute the macro, choose Tools/Makro/Makros/Run. A window dialogue opens and asks for the three settings that define the rules which have to be fulfilled in order to count as a maximum: “Area for Averaging” is the full width half maximum (FWHM) of the Gaussian smoothing. Try half the amount of pixel of the average thickness of a peak for starters (5 is a good value for thin marine varves when scanned at 300 dpi). The “Minimum Cycle Length” defines the minimum thickness of a layer that will be regarded as a peak. The minimum value is 1 (only good for very fine layers). The “Minimum Amplitude” of the normalized gray-scale curve defines the minimum height of a peak that will be regarded as a layer (0.01-0.05 are good for starters).

Once all settings are adjusted, you should click “OK” and the “PEAK_max_count” macro will be executed. After a while a window will open and display a floating chart containing the image with the gray-scale curve (in blue), and the amount of layers (in purple) count along the profile line (in white). The spreadsheet underneath shows the input data (column A to I), the layer count (column H, as a jump from 0.8 to 1.4), and in column K the “Number of Maxima”, as well as peaks that were either “Too small” or “Too flat” to count as a layer according to your settings. Column M and N provide additional statistical information. If you are not happy with the results, run the macro again with changed settings. You can either run all spreadsheets of a workbook consecutively with individual settings, which will be more precise, or you use the Makro “PEAK_max_count_all” to run all spreadsheets at once, which will be

faster. Note that the description above is given for the “PEAK_max_count” macro, which counts every bright peak of a layered sequence consisting of bright and dark layers.

Figure 2: *Snap-shot containg the gray-scale curve (blue) and the profile line (white) along which the layers (purple) were counted. A graph like this will be displayed as a floating chart on top of a data file once the “PEAK_max_count” macro has been executed.*



The “PEAK_zero_count” macro basically follows the same logic. Here, however, every bright and every dark layer are counted by determining the transition between the two, i. e., the point, where the gray-scale curve crosses the average gray value. You will have to adjust three settings before the macro should be executed: the minimum dynamics of a layer (in percent of the standard deviation), the length of the moving average, and the minimum width of a layer.

That’s basically it! Additional macros help you to extract, condense and format the resulting data for further use (see “Collection of Additional Macros”).

Trouble shooting

Although problems may vary depending on the Microsoft Excel version and operation system that are used, there are a few tips we can provide:

- If no image is displayed after the “PEAK_max_count” macro has been executed, the file name of the image may not be according to the specification given above.
- If the layer counts (purple bars) do not display at the appropriate place, try to re-adjust the scale for the graph to the exact dimensions of the underlying image.
- If Excel does not open macros you have to change the security setting under Tools/Makro/Security to at medium or low, and allow for macros when asked in an opening dialogue. Alternatively, you if you use Excel 2007, it will not support macros at all. Then you will have to switch to an older (200-2002) version.

3. Collection of Additional Macros

General purpose

The macros of this collection are not essential for the use of either BMPix or PEAK, but the implementation allows for faster and much more convenient data processing and handling. Most of them come into play after the PEAK tool has been used. In each case, the macros as well as the files referred to, have to be launched into Microsoft Excel first.

3.1. Collect_workbooks

“Collect_workbooks” will collect all open Excel workbooks (without itself) and write them into a new Excel workbook.

3.2. Total_layer_count

“Total_layer_count” will summarize entries in column K, row 2 from all Excel spreadsheet of an open Excel workbook and display the result in a opening window. Accordingly, the total amount of layers counted in a core will be displayed.

3.3. Extract_Column_Multi_Sheet

“Extract_Column_Multi_Sheet” will extract specified columns of all Excel spreadsheets of an Excel workbook and write the data consecutively into a spreadsheet of a new Excel workbook. If more than 65,000 data points are extracted the macro will add another spreadsheet and write the remaining data in there. Upon execution a dialogue will open and ask for the columns to be extracted. Note that the marked data has to be arranged in the same columns for all spreadsheets. The data can then be saved as “.xls” for use within Excel, or as “.tab” for plotting with PanPlot (see PANGAEA data library).

“Extract_Columns_Single_Sheet” will also extract specified columns of all Excel spreadsheets of an Excel workbook, but instead of writing the data consecutively into a single spreadsheet, it will write the data separately into new spreadsheets.

3.4. Depth_adjustment

“Depth_adjustment” allows for the extraction and combination of various data sets of a core, even though when collected at different depth intervals. All data sets should be stored in different Excel spreadsheets with depth in ascending order as the first column. Upon execution, the macro will extract the information into a new Excel spreadsheet, write depth (in ascending order) and associate measured parameters in the following columns. The benefit of this format is, for instance, the ability to export and plot the data directly in PanPlot for visual comparison.

3.5. Condense_data

"Condense_data" will reduce the overall amount of data by averaging. This macro is specifically useful for data sets that are distributed over several Excel spreadsheets (i. e., data sets with more than 65,000 data points). In the opening window, you will have to specify the number of columns (beginning with A) for each of which the averaging will apply, and the number of data points, i. e., the reduction factor, that should be averaged.

3.6. RGB_for_PanPlot

"RGB_for_PanPlot" searches the active Excel spreadsheets of an Excel workbook for R, G, B, or Red, Green, Blue, extracts the data and combines them in a new column in the format R/G/B. The new column will have to be defined first by typing R/G/B into the first row of the Excel spreadsheet. Upon execution of the macro, the data will be converted into text format. This macro is useful to provide the required format for the colored visualization of RGB-values in PanPlot.

3.7. Separate_RGB

"Separate_RGB" basically reverses the "RGB_for_PanPlot" macro. It searches for "R/G/B" in the selected column, extracts the data, and inserts three additional columns containing R, G, and B data. All you have to do is to mark the first row of the column containing the word R/G/B, before executing the macro.