

## Tutorial – Clustering and Lithology Module

# Contents

Part 1 – Cluster analysis .....	4
1.1 Cluster analysis of log data .....	4
1.2 Cluster methods available in CycloLog .....	4
1.3 The clustering menu and toolbar .....	5
1.4 K-Means Clustering .....	6
1.5 Native Clustering .....	9
1.6 Hierarchical Clustering (Constrained).....	13
1.7 Hierarchical Clustering (Unconstrained) .....	16
Part 2 - The lithology library .....	19
2.1 Working with the lithology library.....	19
2.2 Property grid of Histogram.....	22
2.3 Displaying lithology patterns in a log .....	24
2.4 Saving the Histogram.....	24

Copyright © 2018 PanTerra Geoconsultants B.V. All rights reserved

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without written consent of PanTerra Geoconsultants. Under the law, reproducing includes translating into another language or format. As between the Parties, PanTerra Geoconsultants retains title to, and ownership of, all proprietary rights with respect to the software contained within its products. The software is protected by copyright laws. Therefore, you must treat the software like any other copyrighted material (e.g., a book or sound recording). Every effort has been made to ensure that the information in this user guide is accurate. PanTerra Geoconsultants is not responsible for printing or clerical errors. Information in this document is subject to change without notice.

CycloLog and INPEFA are registered trademarks of ENRES Energy Resources International BV and PanTerra Geoconsultants BV.

# Part 1 – Cluster analysis

## 1.1 Cluster analysis of log data

Cluster analysis in general is a method for exploring ways of arranging multivariate data into groups, or clusters. For the well geologist, it is a way of analyzing log data into electrofacies, each electrofacies being a way of grouping together data points with a similar range of properties. The variables in the case of log data are the various logs, any number of which can be selected for inclusion in the analysis.

The simplest form of clustering is the classification of a single log into a number of 'bins' of equal size, usually used for a colour display of the GR log in a way that suggests the range of lithologies represented. However, this method takes no account of the way the data points are distributed among the arbitrarily defined bins, and this section of the tutorial describes how to use the more sophisticated clustering tools available in CycloLog.

## 1.2 Cluster methods available in CycloLog

Cluster analysis has a very wide range of applications in many different disciplines, and many different clustering methods exist. Of these, CycloLog offers two *non-hierarchical* methods, and one *hierarchical* method.

The two non-hierarchical methods in CycloLog are:

- Native clustering
- K-means clustering

**Native clustering** is quite intuitive, resembling the way that the eye tries to find clusters among points on a 2-D plane. It looks for groups of points that are each closely clustered around some centre point, at the same time trying to find group centres that are as far apart from each other as possible.

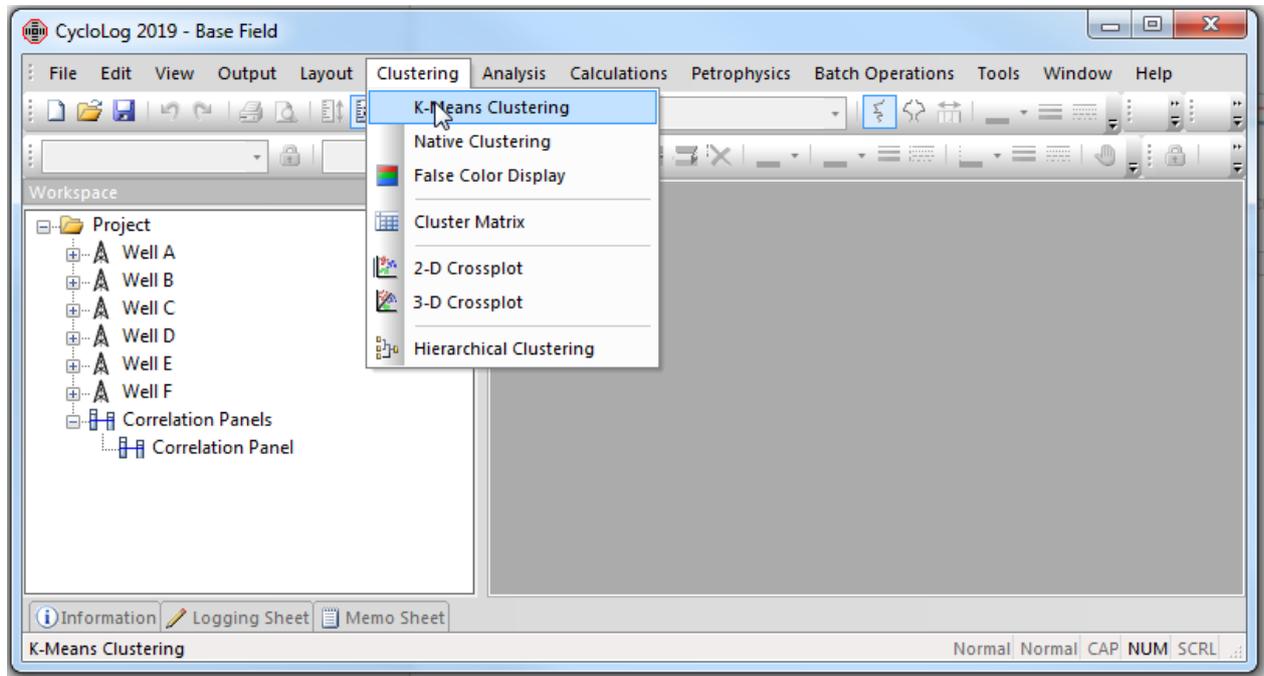
**K-means clustering** seeks the best way of dividing the data-points among some number (K) of clusters such that each data-point belongs to the cluster with the nearest mean value.

**Hierarchical clustering** in CycloLog works from the bottom-up, starting by assigning each data point to a separate cluster, then progressively adding clusters together based on their similarity. (Other hierarchical methods work from the top-down, successively subdividing the data into a larger and larger number of clusters.)

It is not easy to suggest criteria for choosing one method or another, and the user is invited to experiment with all of them. The following gives introductory instructions for each method.

### 1.3 The clustering menu and toolbar

Cluster analysis functions are grouped together in the **Clustering** menu (in the main menu bar, at the top of the CycloLog screen):



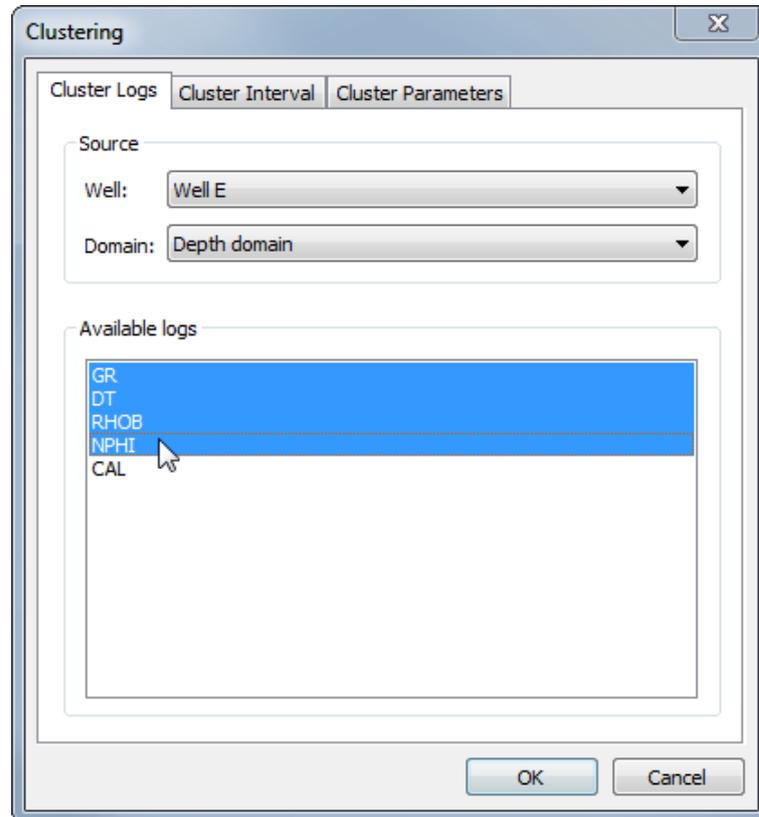
A small number of clustering functions can also be accessed from the **Clustering** toolbar. Position the cursor over the icons on the toolbar to reveal their functions.



## 1.4 K-Means Clustering

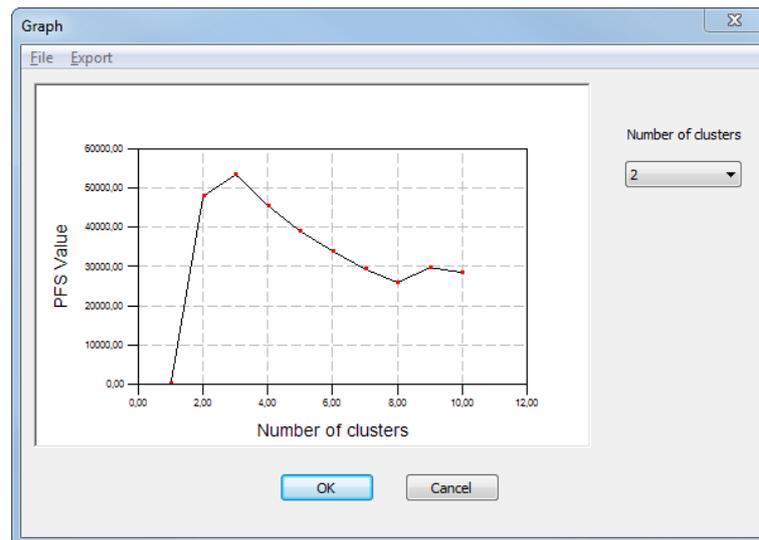
To run a K-means clustering:

- Go to the main menu bar and select: **Clustering → K-Means Clustering**. The following dialog box opens:



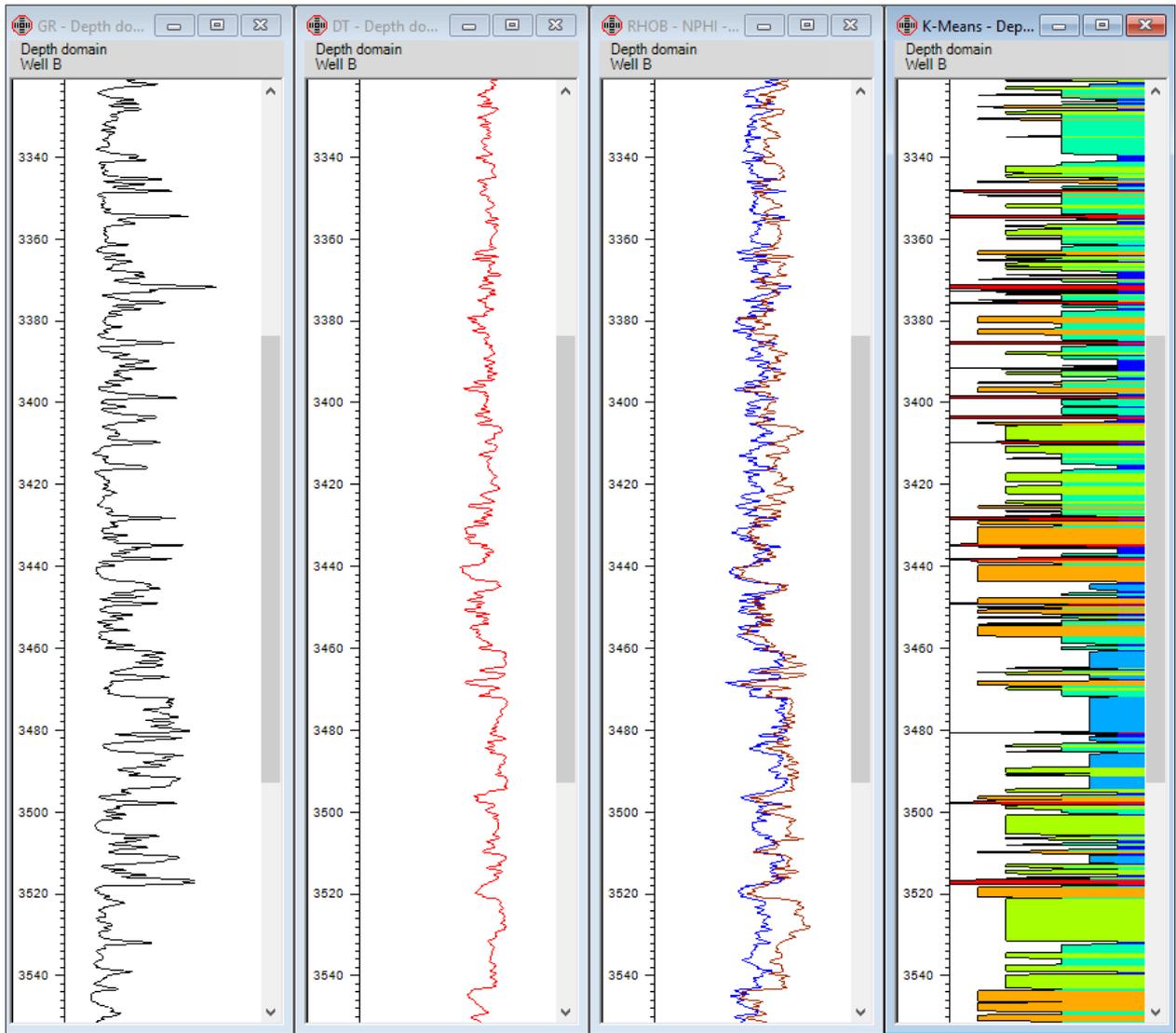
- In the **Cluster Logs** tab, select the **Well** and **logs** you wish to run a K-means clustering.
- In the **Cluster Interval** tab, define the **Top** and **Bottom** depth of the interval to be analysed.
- In the **Cluster Parameters** tab, select the number of clusters (default = 10): *this should be larger than the expected final number of clusters.*
- Click **OK** to run the analysis.

The analysis runs, and the following graph is displayed:



The graph shows an evaluation of the outcome for numbers of clusters up to the number you defined (Default of 10 is used in the example above).

- In the **Number of clusters** drop-down list, select the number corresponding to the point at which the graph begins to flatten (from left to right). In the illustrated example, you might choose 8.
- Click **OK**.
- A cluster log called **K-Means** is added to the workspace and can now be displayed.
- To display the colour fill of the K-means cluster log, right-click over the K-means data pane log and select **Display → Fill → Right → Log → K-Means**. OR open the Properties toolbar (see next page, below). In the Properties pane, use the Log Fill options. Choose *Log* in the **Fill with** entry and select the cluster log.

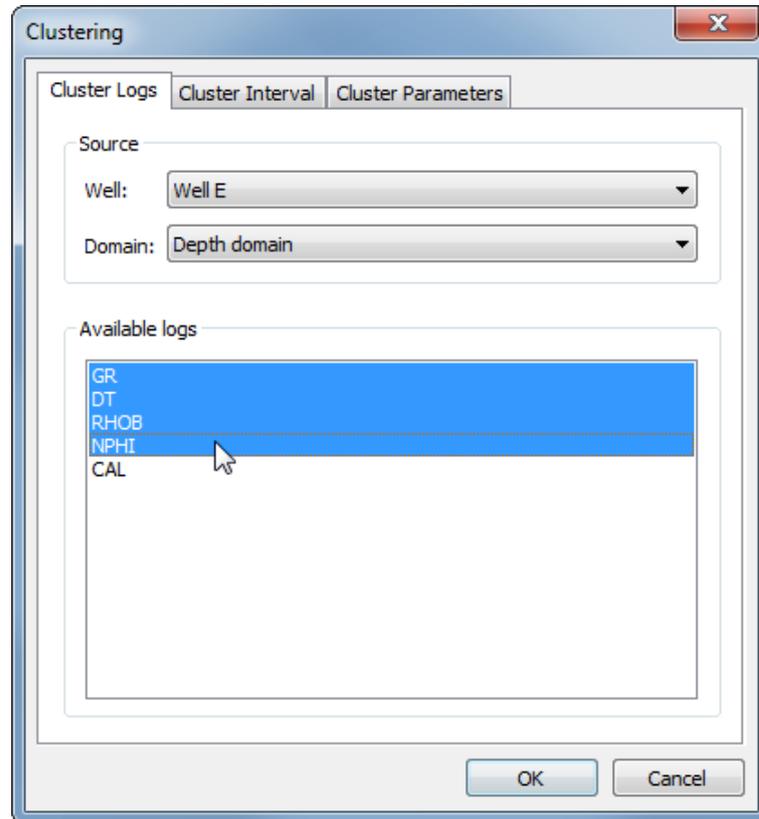


The **Cluster Matrix** can be displayed and modified, as described for the Native method below.

## 1.5 Native Clustering

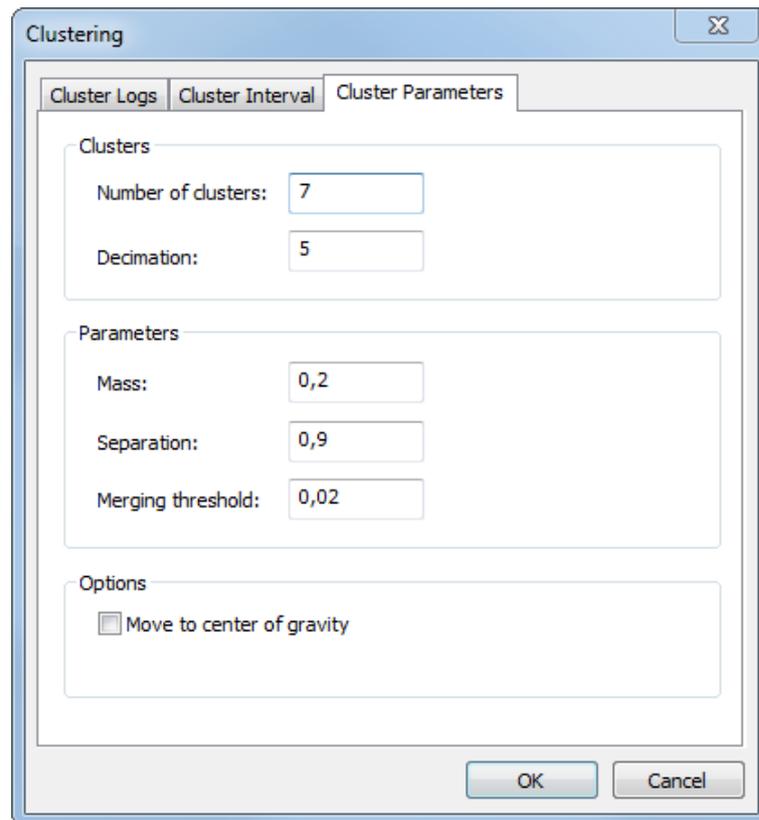
To run a Native clustering:

- Go to the main menu bar and select: **Clustering → Native Clustering**. The following dialog box opens:



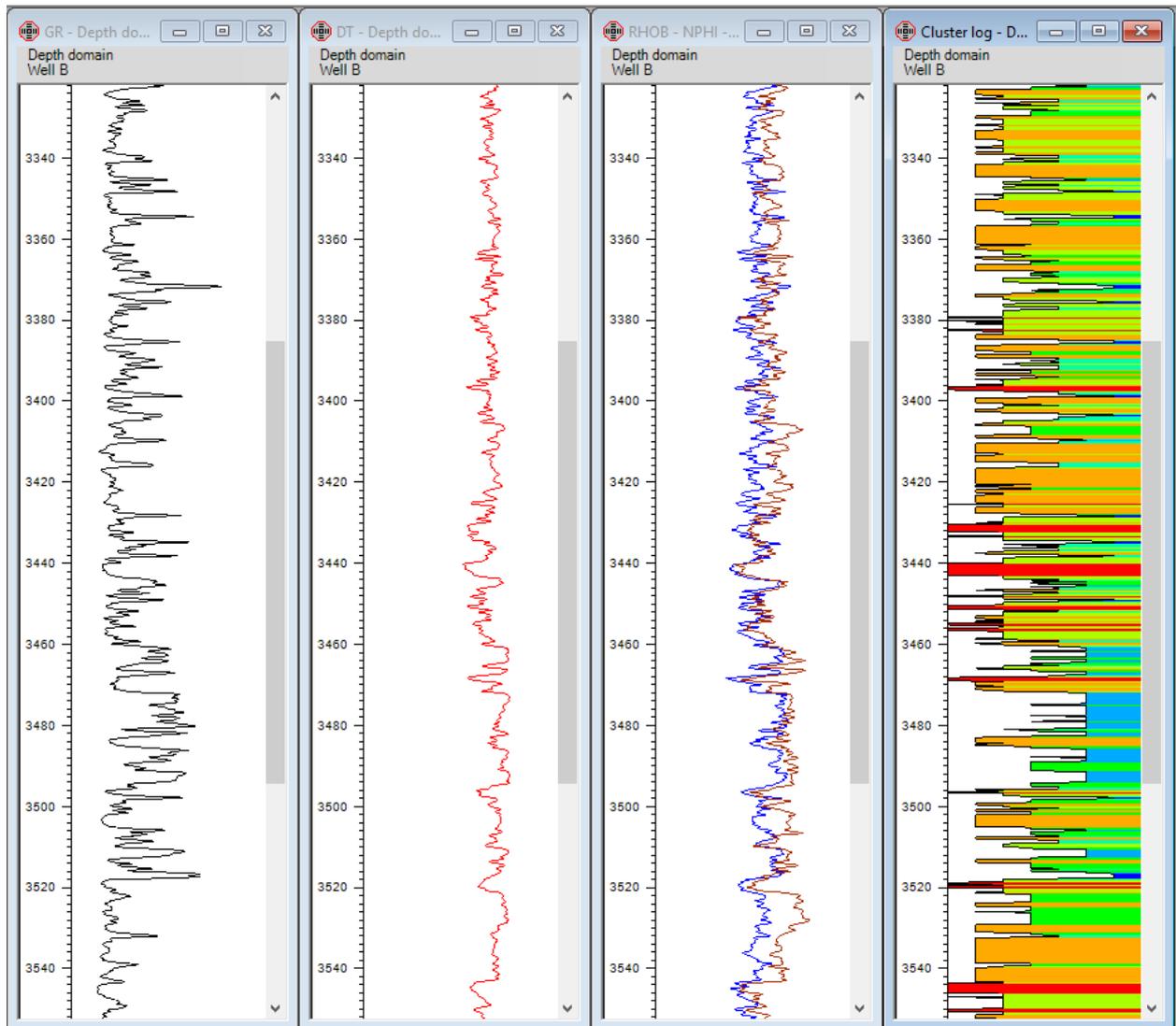
- In the **Cluster Logs** tab, select the well and the logs to be included in the cluster analysis (see example in the figure above).
- In the **Cluster Interval** tab, define the **Top** and **Bottom** depth of the interval to be analysed.

- Open the **Cluster Parameters** tab:



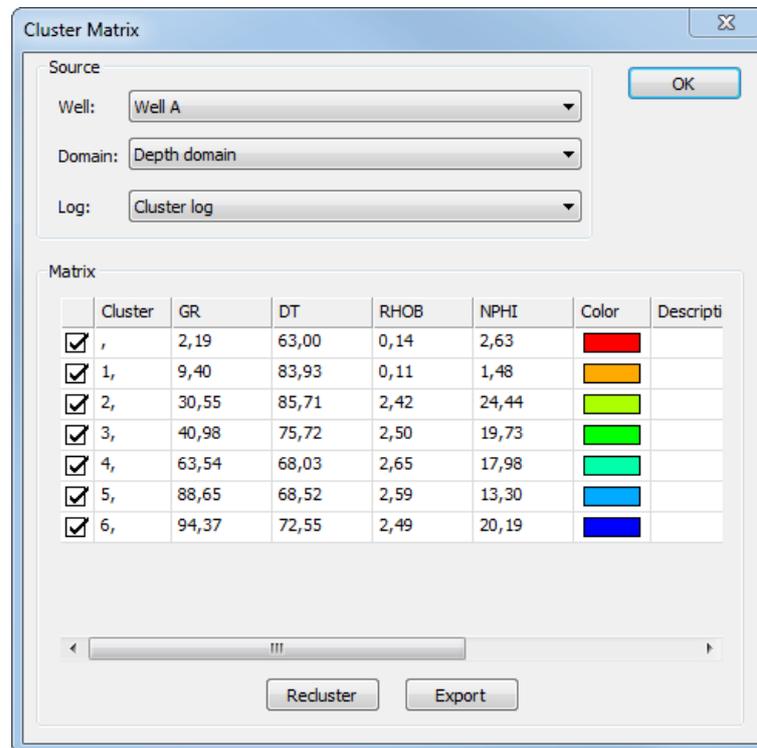
- For an initial run, accept the default parameters and click **OK**.
- The analysis runs, resulting in a new **Cluster Log**, which is added to the workspace.
- *It is a good idea to rename the cluster log: right-click over its name in the workspace, select **Rename**, and type the new name, then press **Enter**.*
- To view the cluster log, double-click on its name in the workspace.

- In the figure below, the cluster log has been coloured. Open the Properties pane and use the Log Fill options. Choose *Log* in the **Fill with** entry and select the cluster log.



To view the numerical properties of the clusters:

- Open the **Cluster Matrix** either from the Clustering menu, or from the Clustering toolbar.
- The cluster matrix shows the mean log values that defines the centre of each cluster; this may be helpful for interpreting the geological or petrophysical significance of each cluster.



The **Cluster Matrix** can be used to analyse and modify the cluster results:

- To change the colours assigned to a cluster, click on a colour in the **Color** column and select the desired colour.
- To add a text description to a cluster, click in the **Description** column of the cluster matrix, and type in your text.
- The **Export** button can be used to export the cluster matrix to an ASCII file (\*.cls).
- Note that the cluster log results can be used to colour fill other logs listed in the workspace (except special logs). Open a log from the workspace and make sure that it is the active log. Open the Properties pane and select **Fill with** → **Log** and subsequently select the cluster log from the drop-down menu.

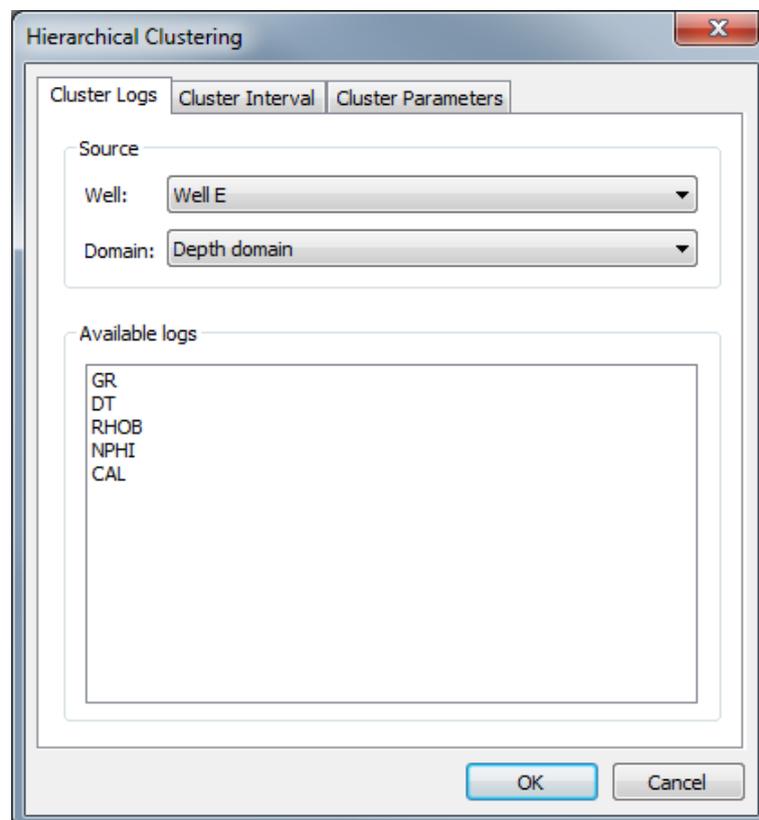
## 1.6 Hierarchical Clustering (Constrained)

There are two different approaches to **Hierarchical Clustering** in CycloLog: **Unconstrained**, and **Constrained**. (Examples are illustrated below.)

- In the **Constrained** method, the analysis is forced to maintain the original order of the samples, so that the resulting dendrogram reflects the similarity between adjacent samples. This method may be also be called stratigraphic clustering.
- In the **Unconstrained** method, the analysis is not restricted to maintain the original order of the samples. Samples from the top of the well could from part of the same cluster as samples from the bottom of the well.

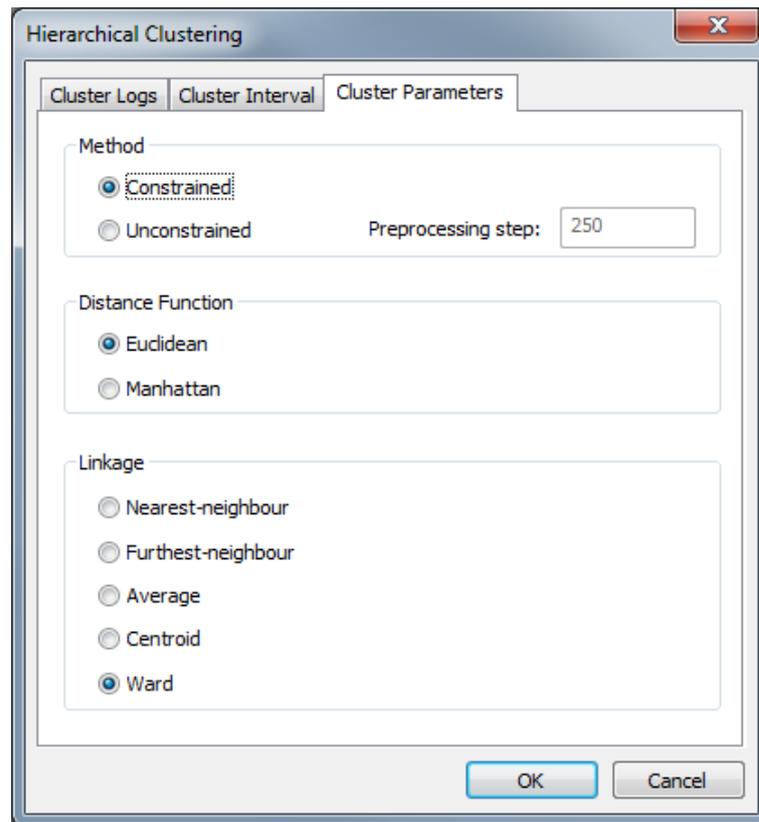
To run Hierarchical Clustering:

Go to the main menu bar and select: **Clustering → Hierarchical Clustering**. The following dialog box opens:



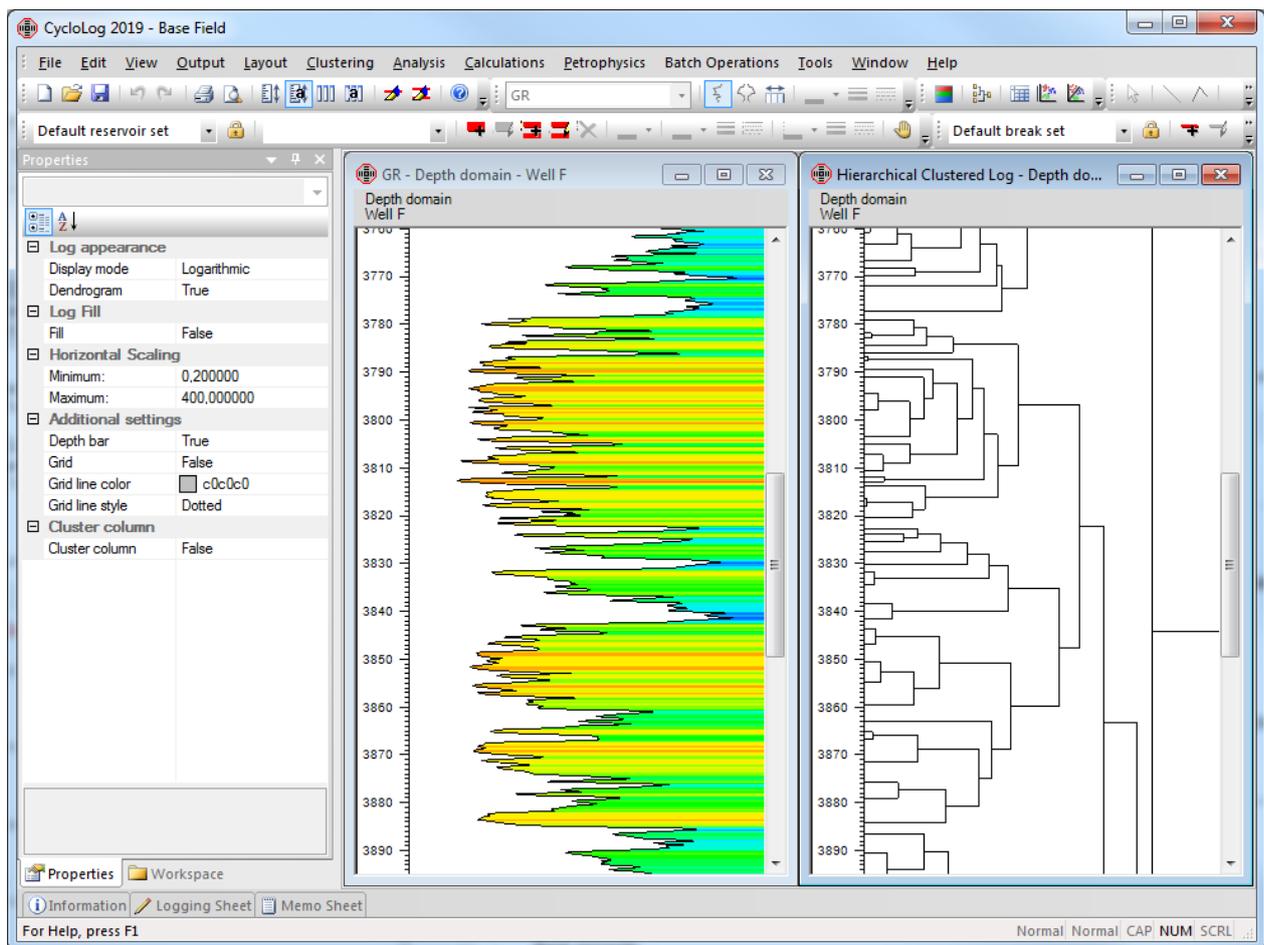
- In the **Cluster Logs** tab, select the **Well** and **Logs**.
- In the **Cluster Interval** tab, define the **Top** and **Bottom** depth of the interval to be analysed.

- Open the **Cluster Parameters** tab:



- Select Constrained.
- For an initial run at least, accept the default Distance Function and the default Linkage method.
- Click OK to run the analysis.
- Open the resulting Hierarchical Clustered Log that appears in the workspace.

The following figure shows a constrained hierarchical clustering of an example well (i.e., Well F and logs GR, DT, RHOB and NPHI). The display is in the form of a dendrogram, in which the horizontal axis shows the degree of similarity of adjacent data-points, from most similar (left) to least similar (right).



The display of this Hierarchical Clustered Log has been made easier to read by:

- Make sure the Hierarchical Clustered Log is open and the active pane.
- Open the **Properties** pane.
- In **Log appearance**, change the **Display mode** to *Logarithmic*.
- In **Horizontal Scaling**, change the minimum to 0.2 (you may test what is the preferred view, by using a lower or higher value). Don't change the **Maximum** value.

It is now easier to see the relationships between samples. Note, for example, that the pair of adjacent samples that are most different from each other are above 3780m (because they belong to the two largest clusters). In this way you can determine a hierarchy of relationships between adjacent samples.

For additional options, see section 18.6 Hierarchical Clustering in the CycloLog Help manual.

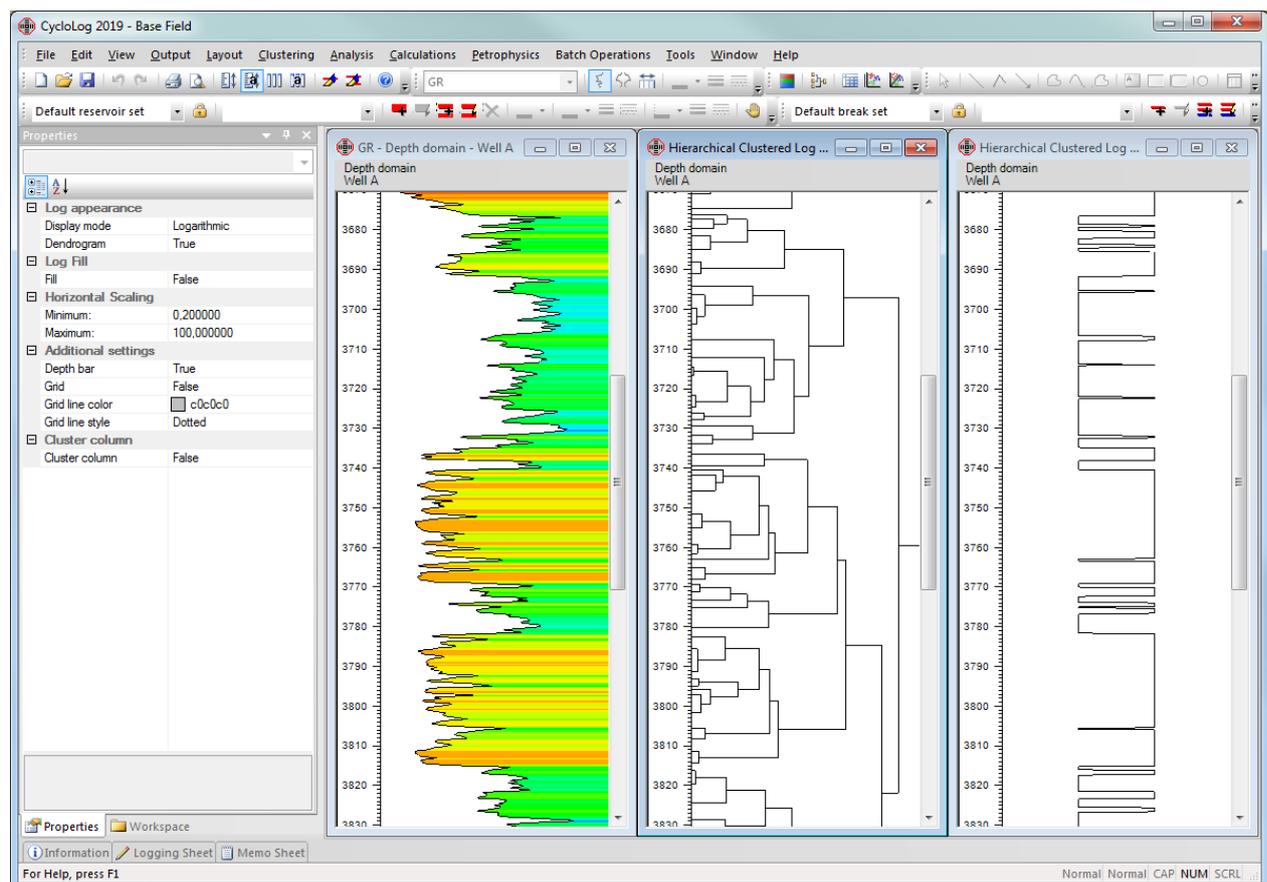
## 1.7 Hierarchical Clustering (Unconstrained)

In Unconstrained Hierarchical Clustering, the clustering is unconstrained by the relative stratigraphic (depth) positions of the data points.

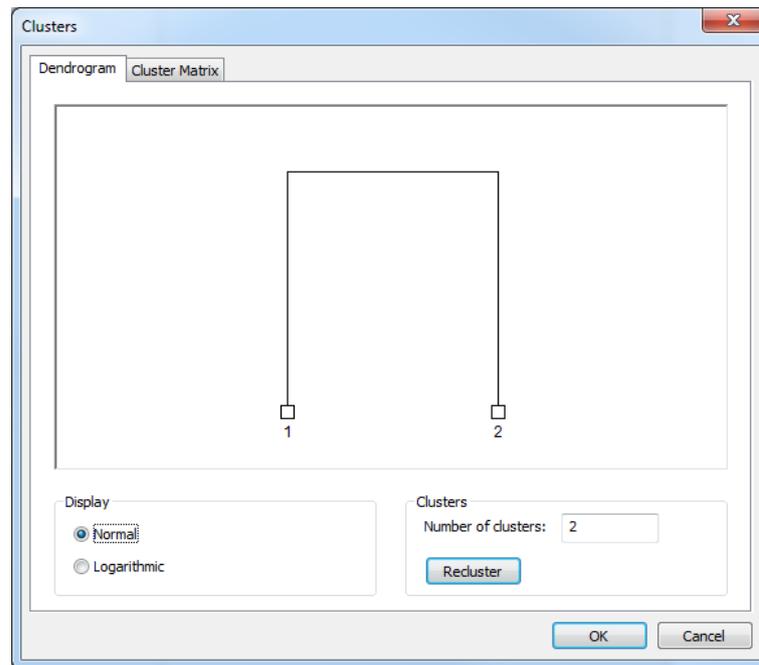
CycloLog does the clustering in two stages (to avoid unnecessarily long computing time). The first 'pre-processing' stage is a K-means clustering, which provides the part-processed data on which the hierarchical clustering subsequently operates. The number of pre-processing steps can be defined by the user.

- Run as for the Constrained method, except:
- Select **Unconstrained** in the **Cluster Parameters** tab of the Hierarchical Clustering dialog box.
- Either accept the default (=250) **Pre-processing steps**, or change to your preferred number.
- Click **OK**.
- CycloLog performs the clustering and adds a new **Hierarchical Clustered Log** to the workspace. (You may change the name by right-click with the cursor on the log's name in the workspace and then select Rename).
- Open the new log by double-clicking on its name in the workspace.

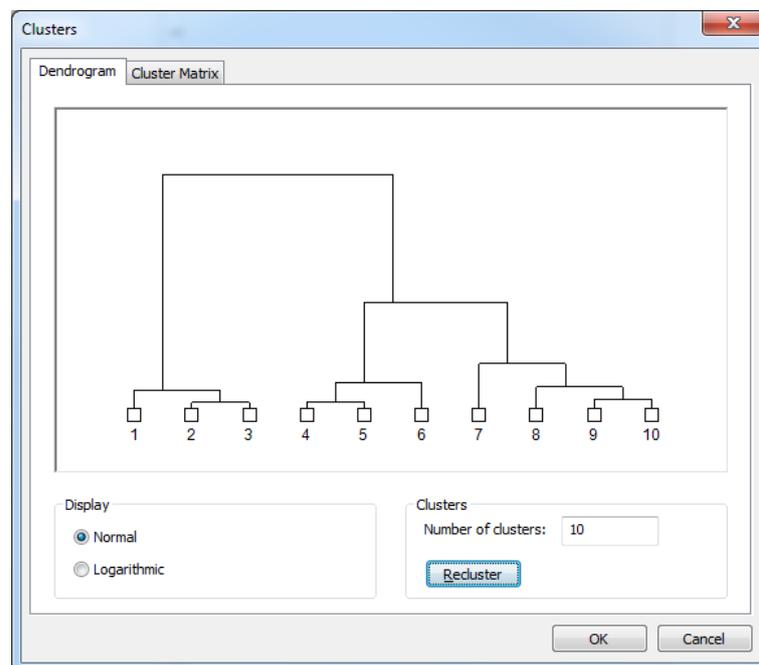
Note that the target number of clusters is *not* defined before the analysis. Initially, the display will show the data grouped into only two clusters, like the right log data pane in the illustrated example:



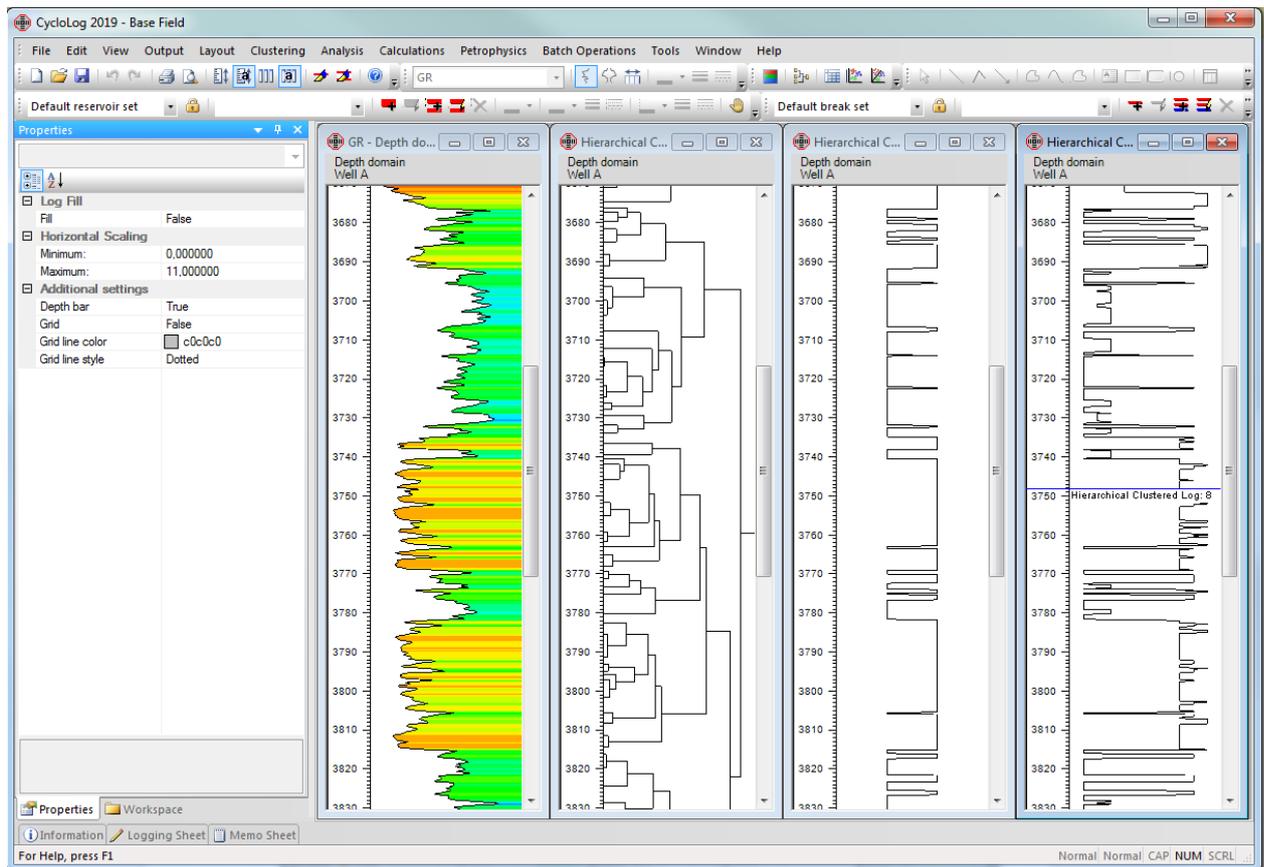
To change the number of clusters, right-click with the cursor over the relevant display pane, and click on **Set Clusters** to open the following, initial, Clusters dialog window:



- This window shows the **dendrogram** for the clustered data, with only two clusters (initially).
- Reset the **Number of clusters** to 10.
- Click on **Recluster**. The same data organized into ten clusters is shown below.



- The dendrogram shows the relative similarity/dissimilarity between the clusters. In the figure below, the curves of the 2 and 10 number of clusters are shown in the data panes on the right.



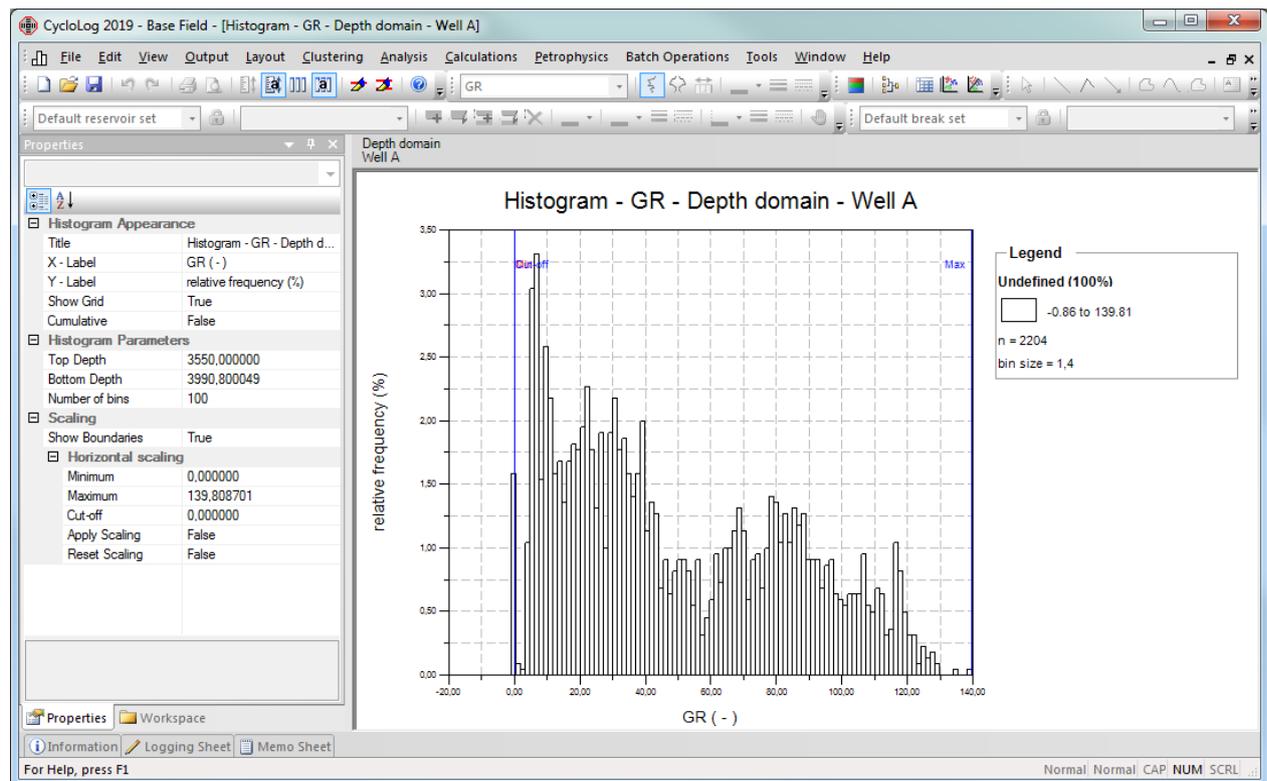
To find out which cluster each data-point has been assigned to, place the cursor over the **Hierarchical Clustered Log** pane, hold down the left mouse button, and move the cursor up or down to the required depth. Press the key-board letter L and the cluster number will be shown (see blue line in image above). It also will be shown in the **Status Bar** at the bottom of the CycloLog screen.

## Part 2 - The lithology library

### 2.1 Working with the lithology library

The lithology library can be accessed through the Histogram pane of a log. To open the Histogram:

- Right-click with the mouse button on the log pane and select Analysis → Histogram.
- Alternatively, you can access the histogram from the right-click menu on a log in the workspace (select View Histogram).
- The Histogram for the selected log is shown in a pane.



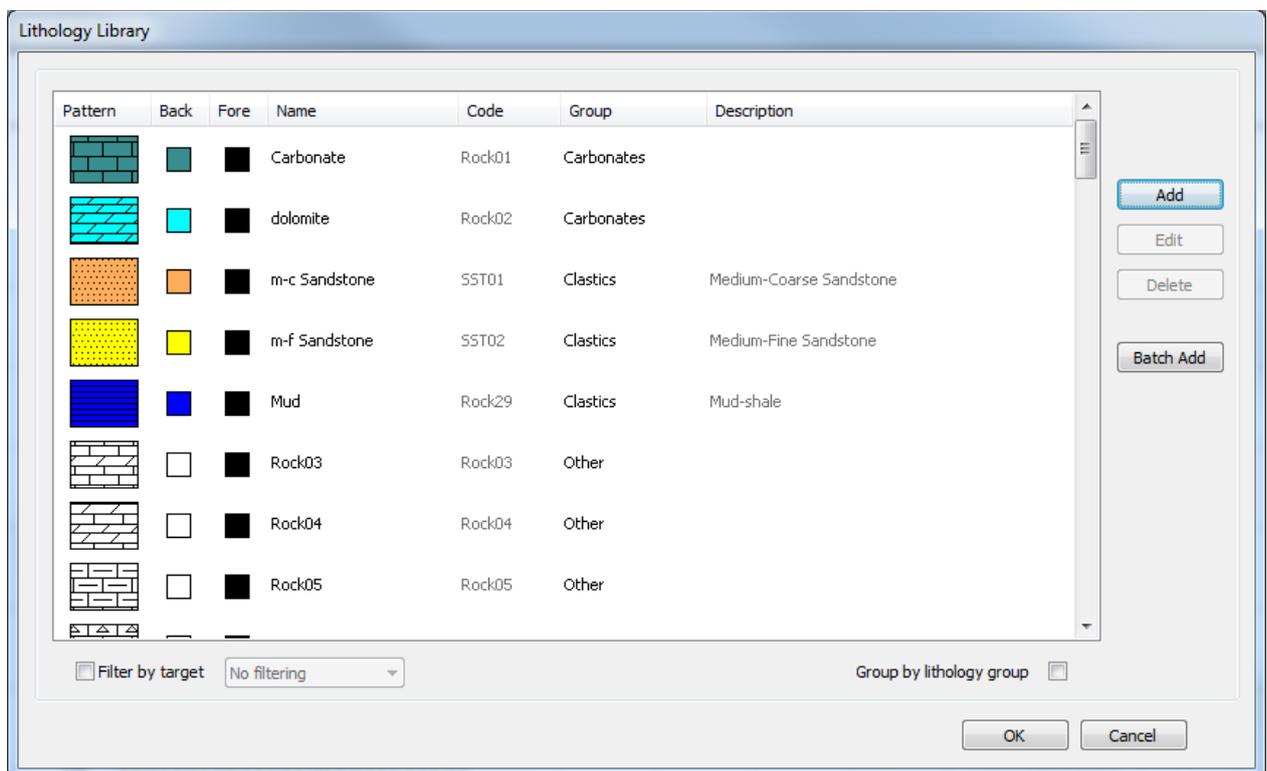
By default, vertical lines representing the minimum and maximum value of the data and the cut-off for toning are displayed. To change their values, they can be dragged along the log range. The interval defined by the minimum and maximum is used to calculate the relative frequency. Relative frequency is recalculated based on a newly defined minimum and maximum when **Apply Scaling** is set at **True** in the **Properties** pane. The Cut-off value is used for toning of the log which is displayed in the histogram. Toning can be applied via the **Properties** pane.

The log range in the histogram can be divided into subranges. Each subrange can then be assigned to a lithology. To make subranges, add pegs on the horizontal axis of the histogram:

- To set a peg, move the mouse pointer below the horizontal axis of the histogram and double click the left mouse button. A green coloured peg will be added.
- Press the left mouse button on a peg if its position needs readjusting. This allows dragging it along the log axis.
- Right click on a peg to remove the peg and delete the left or right range.
- When you define or delete a range by setting or removing a peg, the legend of the histogram is updated accordingly.

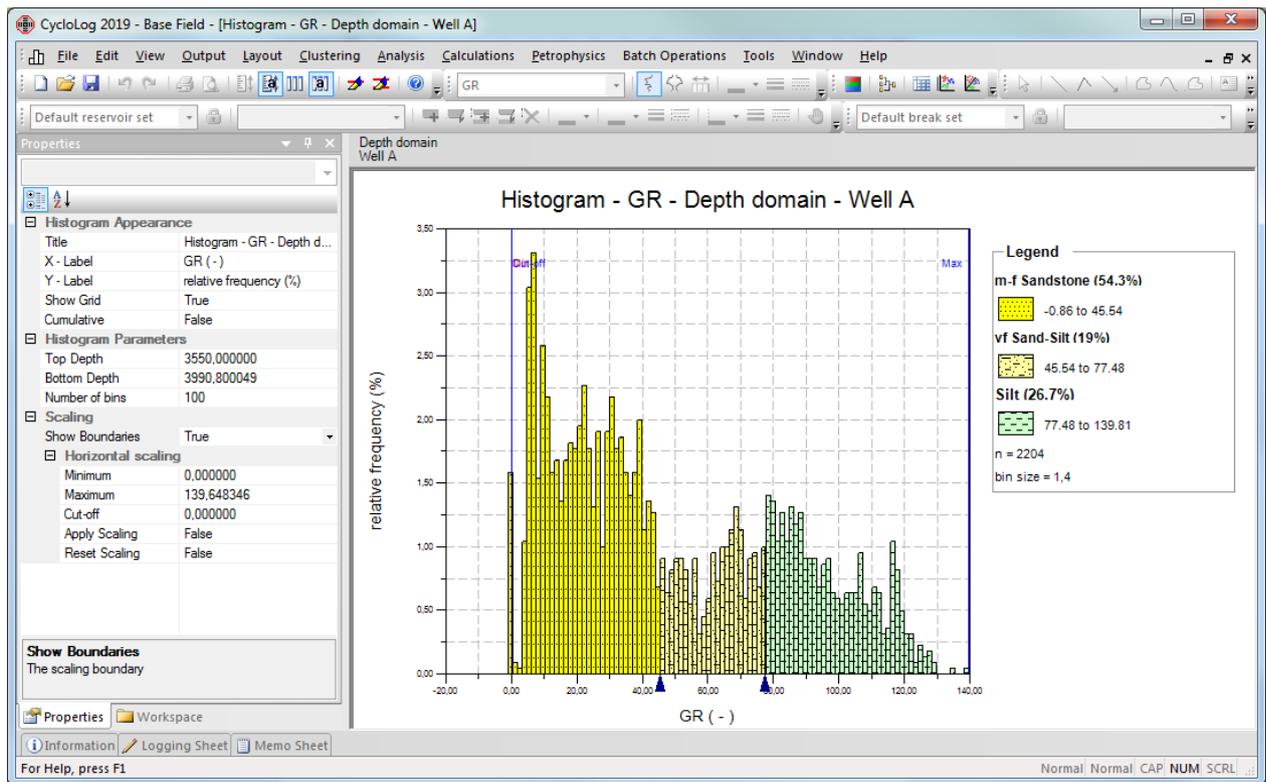
To each subrange, a lithology pattern can be assigned:

- Click on the lithology pattern in the legend of the histogram. This will open the Lithology Library:



- Select a lithology pattern in the lithology library.
- Click **OK** to apply the pattern to the selected subrange in the histogram.

See example below after having defined a lithology for each subrange.



Within the lithology library, lithology patterns can be added, edited and deleted:

- Open the **Lithology Library** by double clicking on a pattern in the legend of the histogram.
- To create a new lithology pattern, click **Add**.
- The **Create Lithology Unit** dialog box opens:

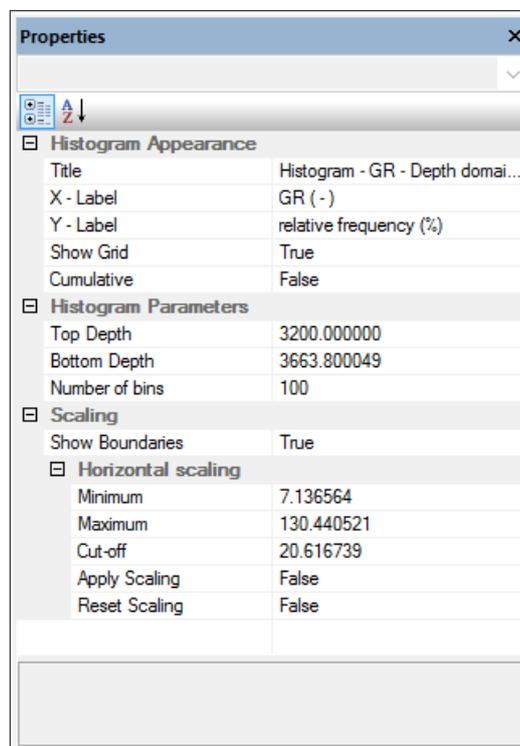
The "Create Lithology Unit" dialog box is shown. It has the following fields and options:

- Name: [Text input field]
- Code: [Text input field]
- Description: [Text input field]
- Group: [Dropdown menu, currently set to "Other"]
- Targets: Log analysis , Core analysis , Project specific
- Pattern: [Pattern selection area]
- Background: [Color selection dropdown]
- Foreground: [Color selection dropdown]
- Buttons: Load, OK, Cancel

- Insert a **Name** and **Code** for the new pattern.
- Adding a **Description** is optional.
- You can assign the new pattern to a lithology **Group** (The groups available in CycloLog are: Carbonates, Chemical, Clastics, and Other).
- Selecting a **Target** for the new pattern is optional.
- **Load** a pattern. The pattern file can have either the "bitmap" file format (\*.bmp) or the "tagged image file" format (\*.tif, \*.tiff).
- Adjust **Back-** and **Foreground** colour.

## 2.2 Property grid of Histogram

- The property grid of the Histogram can be accessed by opening the **Properties** bar (right click over the toolbar area and select Properties, if it is not already open).



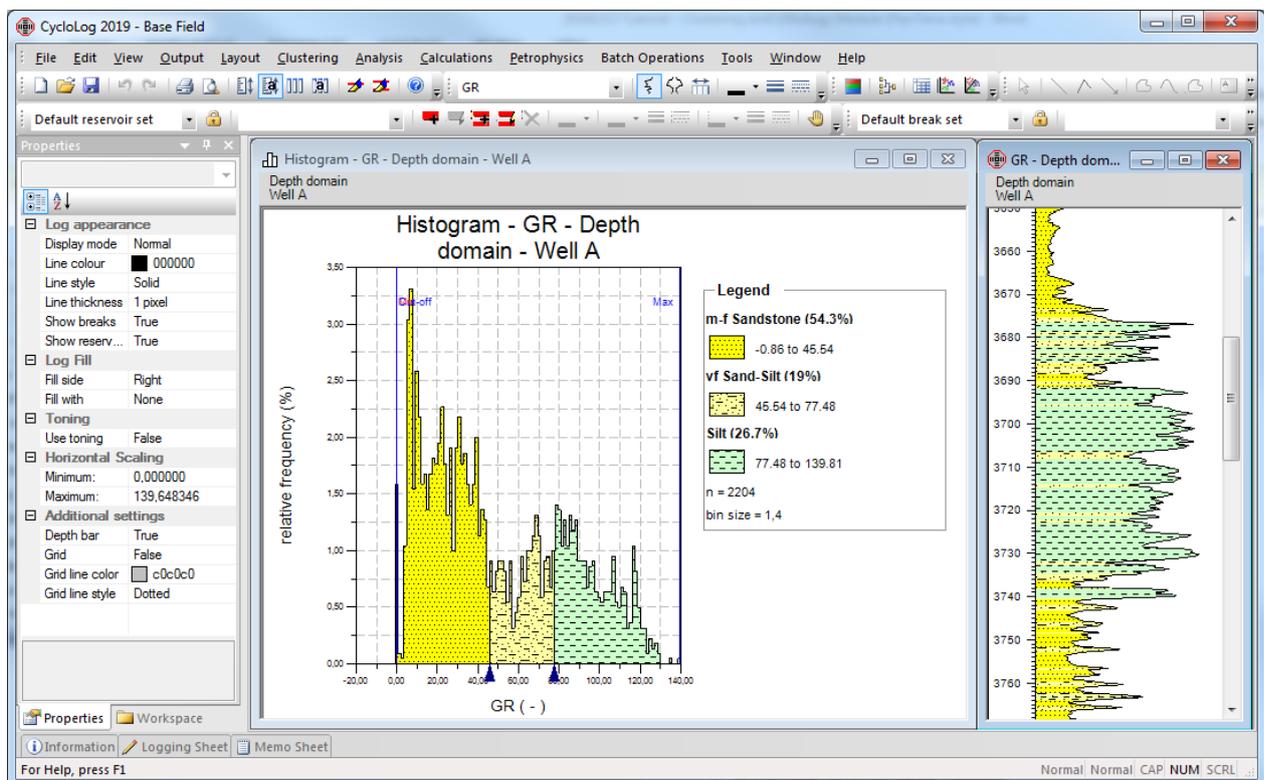
- In the **Histogram Appearance** node, the **Title** of the histogram, and the **X-** and **Y- Label** of the histogram axes can be edited. **Show Grid** and the **Cumulative** display of the histogram can be turned on or off by selecting *True* or *False*, respectively.

- In the **Histogram Parameters** node, the analysis interval of the Histogram can be adjusted (thereby not changing the displayed interval of the log itself). Therefore, edit the **Top Depth** and **Bottom Depth** values. The **Number of bins** (default = 100) can be specified.
- In the **Scaling** node, you can choose to **Show Boundaries** (i.e., the Minimum, Cut-off and Maximum vertical lines in the Histogram pane). By default, **Show Boundaries** is set at *True*.
- In the **Horizontal scaling** node, interval defined by the **Minimum** and **Maximum** is used to calculate relative frequency for the data points within the defined values. Set **Apply Scaling** to *True* to recalculate relative frequency based on the newly defined data interval. You can reset this scaling by setting **Reset Scaling** to *False*. The **Cut-off** value is used for the log toning.

## 2.3 Displaying lithology patterns in a log

To display the lithologies defined for a log (based on its histogram):

- Open the log from which the histogram and lithologies were generated by double clicking on it in the workspace tree.
- Go to, or open, the **Properties** pane of the log (in the example below, the GR log has been selected).
- Go to the **Log Fill** node and select **Fill with** → **Histogram**.
- The Log is filled with the lithology patterns defined above, in part 2.1.



The display of the lithology in the log pane is interactive: if the pegs are moved in the histogram, the log fill display is updated directly.

## 2.4 Saving the Histogram

- The Histogram will be saved when saving and closing the CycloLog project.
- A Histogram will not be added as a separate pane to the workspace. It can only be accessed via the active pane of a selected log.
- A Histogram cannot be exported.