Hydrocarbon show database:

Using algorithms for faster and objective analyses

Victor Larède

MSc Internship – Earth, Structure and Dynamics

Supervisors:
Guido Hoetz (EBN B.V.)
Fred Beekman (Utrecht University)

Student number:
3818705

Credits:
30 ECTS

First version:
25-02-2019
Second version:
08-03-2019

Contents

Abstract

For the last 2 years EBN has been compiling the hydrocarbon (HC) show database. HC shows are defined as significant occurrences of HC gases or fluids in combination with favourable lithology (i.e. rocks that can be produced from). The database in assembled from Log-, (SW-)core and Test data. One of the methods of filling this database is the 'Manual input' method (QC1) where a person analyses and processes all the data. Filling the database using this method will take at least another 3 years to with the current number of relevant boreholes in the Netherlands. In this internship project, an attempt has been made into automating the gas log analysis part from QC1 by using computer algorithms. A code has been written which can perform the gas analysis workflow used in QC1. Furthermore, an investigation has been done to find a decent method for determining/calculating the background gas signal. A dataset has been analysed both manually and automatically. Assuming that the manual analysis is perfect, a misfit through an RMS error has been calculated for the automated methods to compare how successful the different methods are. The 'drawn' background gas method performed the best with an RMS-error of 0.31. the automated analysis using this method takes around 10-15 minutes. The project has resulted in tool (GAT) with a user-interface that can be used to perform an automated gas log analysis with major time savings. To find a correct manner to calculate the background gas signal more testing and research is needed.

1 Introduction

EBN is owned by the Dutch government and, therefore, executes parts of the Dutch climate policy for the ministry of economic affairs and climate. Until 3 years ago EBN mainly focussed on the exploration, exploitation and storage of Dutch oil and natural gas. Nowadays, EBN has broadened its focus to other energy industry activities and established itself as one of the key players in the energy transition, with the aim of fulfilling the Dutch climate goals.

1.1 The Hydrocarbon show database

Over the last decades extensive hydrocarbon exploration has been conducted in the Netherlands, both on- and offshore. At least 470 gas fields and 50 oil fields have been discovered of which there are currently roughly 250, respectively 15 in production (NLOG, 2018).

The *MIJNBOUWWET* in the Netherlands states that all data from exploration and exploitation activities ought to be released after 5 years. TNO, Geologische Dienst Nederland, stores this data in a public database called NLOG. EBN is involved in nearly all oil- and natural gas-projects as a non-operating partner, providing itself with what may be the most extensive knowledge on the Dutch subsurface concerning and exploitation. Collaboration between EBN and the operating partners is of high importance (to assist in achieving the climate goals). As the exploration of the Dutch subsurface, for any purpose, becomes more challenging given its maturity, new and efficient analytical techniques can provide opportunities for future exploration and risk assessment.

The Hydrocarbon Show database has been developed to conveniently arrange, store and visualize EBNs and NLOGs data regarding observations of HC in the subsurface. It can serve as a tool for EBN and its partners to get a quick and uncluttered overview into what is known about the subsurface in an area of interest in terms of exploration and exploitation data.

1.1.1 Previous work

For the last 2 years EBN has been compiling the hydrocarbon (HC) show database. HC shows are defined as significant occurrences of HC gases or fluids in combination with favourable lithology (i.e. rocks that can be produced from).

Heerema (2016) started constructing the HC database and Kickken (2016) introduced the first 2D (Qgis) and 3D (Petrel) visualization method setups and workflow. Westerweel (2017) and Korevaar (2018) further refined the HC show database workflow and Qgis interface. Moreover, Korevaar (2018) also improved the Petrel interface. Meanwhile, Blom (2017) performed regional studies using the HC show database and recommended Spotfire as a tool for further/parallel analysis of the HC show database. In the end, the HC database project is there to make data on HC shows easily accessible in a user-friendly environment, for EBN and their partners. Klop (2018) further improved the workflow, see Figure 1, by introducing the ‘quick input’ (QC2) method which makes it possible to process a lot of HC show data in one go w.r.t. the ‘manual’ input (QC1). Furthermore, she followed up on the recommendation of Blom (2017) and created the Spotfire-interface.

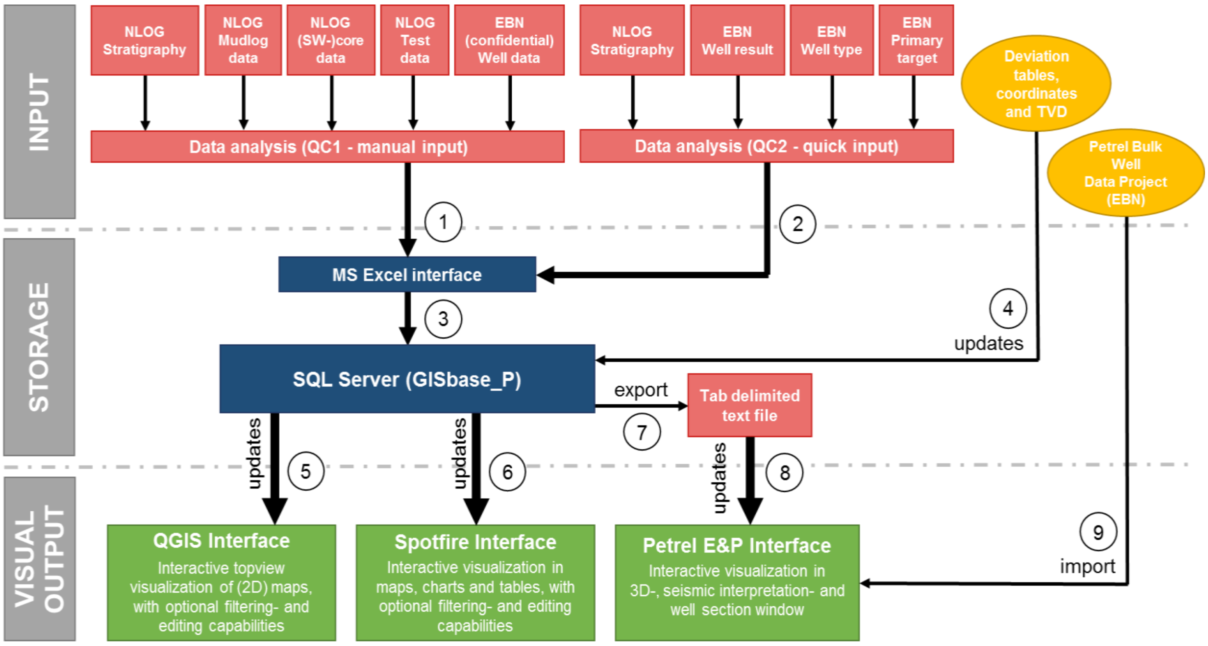


Figure 1. Box-model of the HC show database workflow (Klop, 2018)

1.2 Problem

Manual analysis of the data is more detailed than the 'quick input' method, but it is also very time consuming. An experienced individual can on average analyse 3 wells a day 'manually' (Klop, 2018).

According to 'EBNs basisregistratie boringen' there are 6573 boreholes in the Dutch on- and offshore of which approximately 2/3 contain information on HC. 755 have been analysed manually by 08-02-2019. Estimations say that it takes at least another 3 years to analyse all the remaining boreholes.

1.3 Objectives

It is therefore favourable to investigate and workout the opportunities of automating and fastening the manual workflow of this database. Furthermore, manual analysis easily leaves room for mistakes (Baud, 2018) and bias towards a result. Automation will hopefully result in a more objective analysis than is produced manually.

1.3.1 Approach

The previous stages of the Hydrocarbon show database project are listed below and visualized on a timeline in Figure 2**.** The approach for the current project will be according to the following points:

- Investigation of the possibilities of using machine learning as a tool to analyse well data
- Investigation of the available data-formats used in the HC show database; searching digital files and verify whether it is worth the effort to digitize non-digital files
- Analysing wells 'manually' according to the improved QC1 workflow (Klop, 2018) developed by Westerweel (2017) and Korevaar (2018) to get acquainted with the analysis and further expand the database. Improvements to the workflow will be implemented if necessary
- Once this is done, a code will be written and a user interface to automate the manual analyses will be designed and created

- Lastly, a validation of whether code can reproduce result from manual analyses will be performed

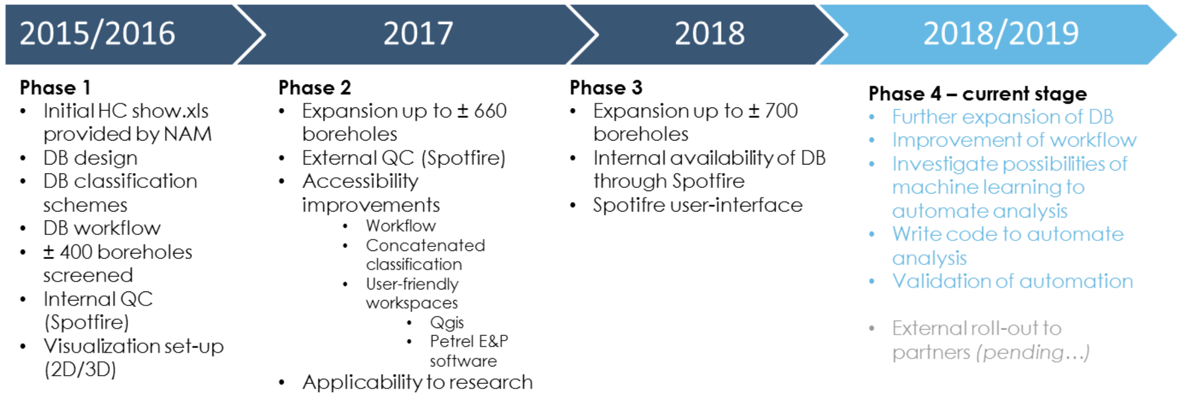


Figure 2. Timeline of the development of the HC show database. Steps that have been taken before the start of this project are denoted by the dark blue arrows and year, the light blue arrow denotes what will be done during this project and shortly therafter.

2 The Hydrocarbon database master file

The HC show master (excel-)file consists of 5 data sheets titled; 'info', 'General well data', 'Log data', '(SW-)core data' and 'Test data'.

2.1 Existing input methods

There are 2 methods used to add data to the master file. The first method is called the 'manual' input method (QC1) and the second one is called the 'quick' input method (QC2), see Figure 1.

QC1 is the most detailed method, because all the log, test and core data that is available on a borehole is used to classify the shows per stratigraphic interval, typically by reading and analysing logs reports etc.. However, as mentioned before, performing QC1 can be very time consuming. In QC2, the assumption is made that wells which have successfully produced oil and gas or where hydrocarbon shows were observed during drilling, according to 'EBN's basisregistratie boringen', can be added with a GOOD test result. Only the layers that are in the stratigraphic group of the 'primary target' of the well get this label. Using QC2 a large amount of boreholes could be added to the database in order to show where GOOD shows can be observed, but QC1 has not been performed yet.

3 Exploring possibilities of machine learning

It was stated that the analyses in QC1 can be very time consuming and that there is a lot of data left to be analysed (Klop, 2018). It is favourable to investigate the possibilities to fasten the analyses. Since we are dealing with a large database/set, the concept of *data mining* comes to mind, which uses the application of machine learning methods to large databases. "*Machine learning is programming computers to optimize a performance criterion using example data or past experience.*"
https://books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&ots=T5ejNL-elQ&sig=Rzyde6VOmKV_qzlsZPfg8w49DJk#v=onepage&q&f=false

3.1 Previous work

Baud (2018) investigated the possibilities of using Machine learning to automate the analyses for the Geo-Drilling Events (GDE) database. The data used to populate this database is mainly retrieved from End-of-well reports, daily-drilling reports and composite logs. The idea was that a tool, developed by an external party, would automatically analyse the reports and eventually populate the database with the classifications by using of an algorithm.

“The initial plan to provide Rolloos with 781 observations was modified as most references of the documents from which the observations were made were not readily available. Only 197 observations were handed over to Rolloos for the initial training set” (Baud, 2018)

Finally, Baud and EBN concluded that due to the lack of functionality of the tool, the lack of EBN’s capacity to provide the necessary assistance to the other party and the lack of perspective for the use of this tool in other projects and also the complexity of the analysis, that it was not feasible to go through with this project.

The Hydrocarbon show database is filled with data from ‘EBN basisregistratie boringen’, formation evaluation logs, geological logs, lithologs, masterlogs, mudlogs, composite logs, (sidewall-)core reports, end-of-well reports, final well reports and (production-) test data reports. Only a minor part of the data listed above is the same as the data used in the pilot for the Geo-Drilling Events database. After some discussion with Guido Hoetz, Oscar Berger, Jacco van de Put, Jessica Klop and myself we came to the conclusion that a similar approach as performed by Baud (2018) would likely end-up with the same result. However, at the same time it was noticed that gas logs could be digitally available for certain wells in LAS-format. Jan Lutgert, Kees van Ojik and Guido Hoetz were able to retrieve some LAS-files via their contacts in the industry. Furthermore, these files can also be found on NLOG and in Livelink, although very sparsely. Thus, it was decided that, in terms of automation of the analyses for the hydrocarbon show database, this project would be narrowed down to focus mainly on the automation of gas log analyses.

4 Data requirements

As mentioned in the section above, the decision was made to narrow this research down to the automation of the analyses of gas logs. <u>The manual analysis of gas log images is estimated to take around 2 hours per borehole.</u> The automation should not only deliver an accurate but also a quicker analysis. To do that, the data needs to be available in a digital format, of which a brief explanation is given below.

4.1 LAS- and other formats

"The LAS file format is a public file format for the interchange of 3-dimensional point cloud data between data users. Although developed primarily for exchange of lidar point cloud data this format supports the exchange of any 3-dimensional x,y,z tuplet. This binary file format is a alternative to proprietary systems or a generic ASCII file interchange system used by many companies." (Anon, 2019)

LAS-files frequently represent digital log data and are for example used in petrel. Nevertheless, these files can also be opened and adjusted in software such as Excel or Notepad, see Figure 3.

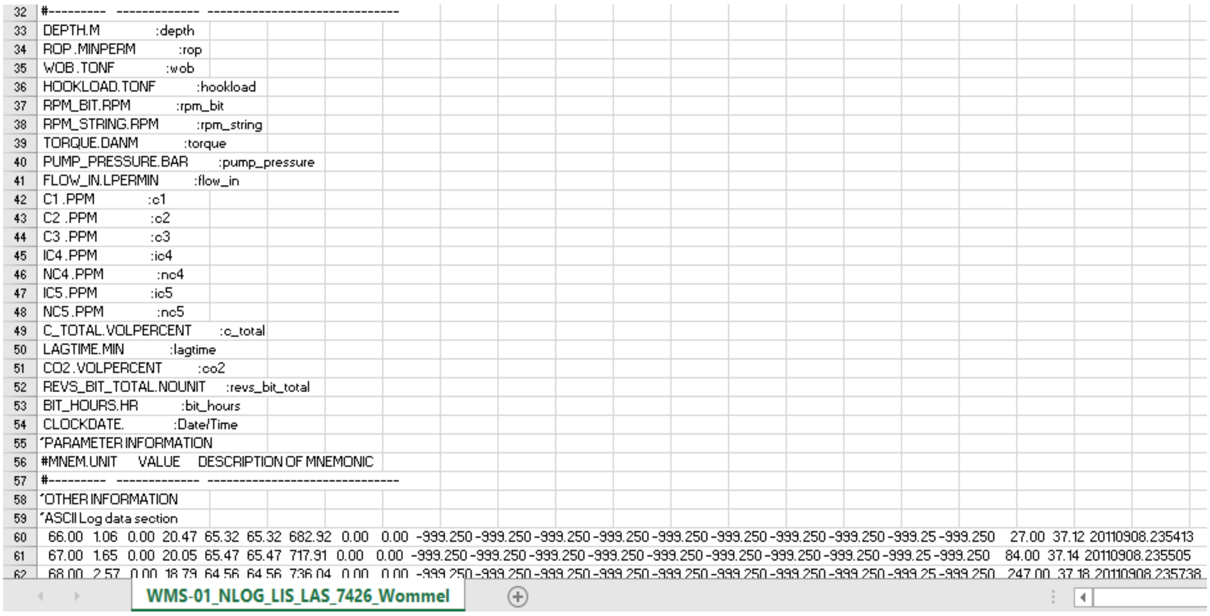


Figure 3. Snip of a mud log opened in Excel. The top represents the header containing the metadata and at the bottom the data is written.

Similarly, other formats such as ASCII- and .txt-files containing gas data were encountered.

Since recent log images are presented in .pdf-format rather than images or images of scanned paper logs, it makes sense that these LAS-, ASCII- and .txt-files are the input-data behind the .pdf-files of the logs. However, getting a hold on all the data-files constructing the PDF-format logs would take a lot of cooperation from the operator that produce them. Hence, it would be convenient to verify whether it is possible to quickly produce such data from log-files that can be found on NLOG or LiveLink.

4.1.1 Didger5

Didger5 is a software with which you can perform digitisations, image registrations & warping and coordinate conversions. It can be used to digitize maps, graphs, well logs and other types of data.

In theory this software seemed very promising to quickly convert gas data, from images of well logs, with only few input. At least, that is what the instruction movie implies. The general idea is you load the image containing your log and delimit the log area which you want to digitize with 4 points representing x_{min}, x_{max}, y_{min} and y_{max}, see Figure 4. Then, you select a part of the line representing the chromatograph in the log and the program should draw a polyline representing this chromatograph. This function can also distinguish between several plots within one graph, for example C1, C2, etc., on the basis of their colour. After this, only minor adjustments by hand should be made by hand to the polyline, before finally saving it to a format of choosing.

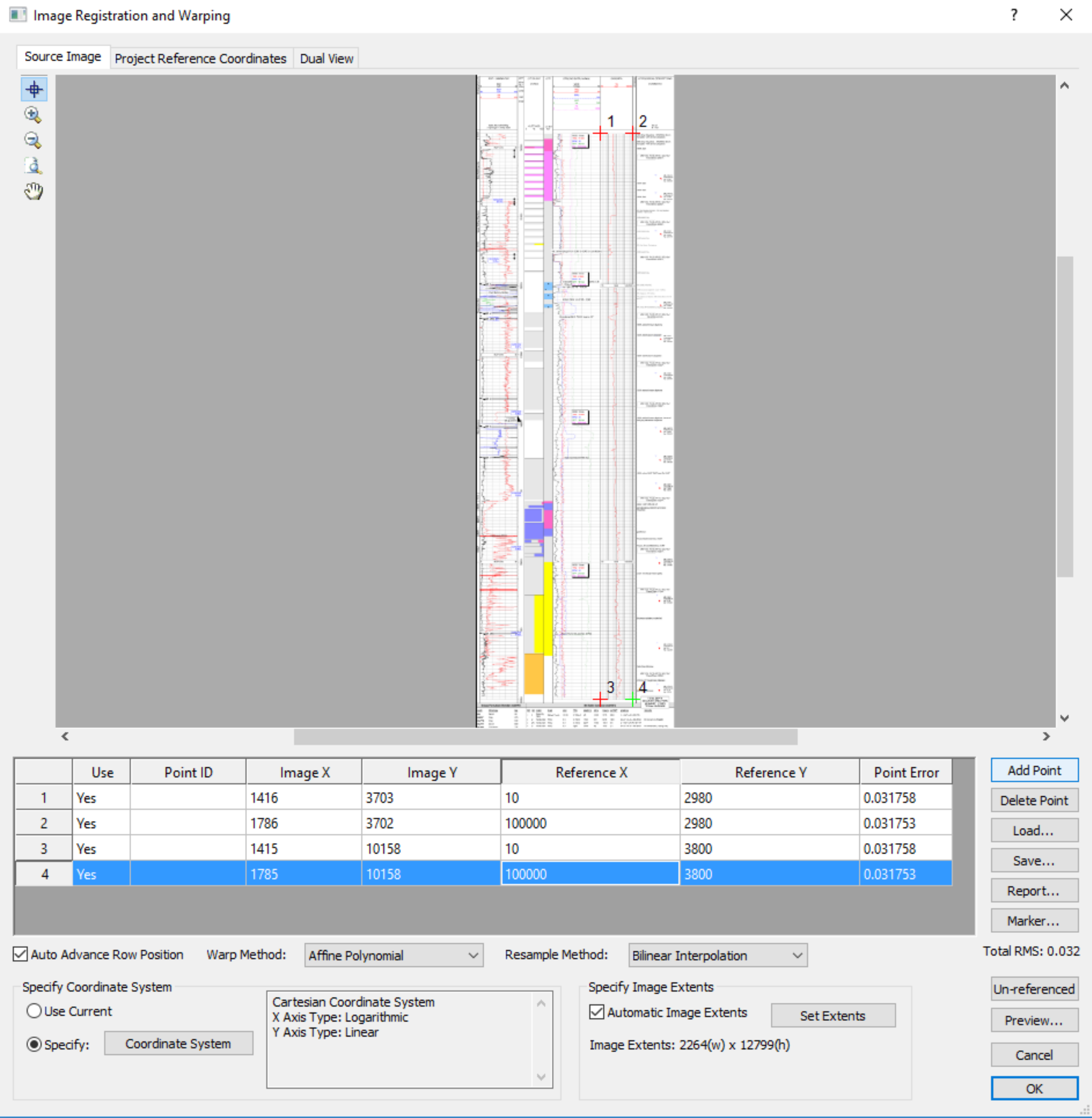


Figure 4. Didger5 interface where the chromatograph area of an imported log image can be indicated and calibrated with 4 points.

However, in practice using this software to digitise the logs is far more cumbersome and time-consuming. Images of the well logs are usually quite narrow and elongated, this makes them hard to visualize nicely in the didger5 user interface.

Furthermore, there are logs, particularly the older ones, that are visualized in black-and-white. The digitization of a black-and-white chromatograph with a ditto coloured grid on the background is problematic for the software, as it doesn't know what lines to connect and what lines to ignore.

And lastly, delimiting the graph area in the log doesn't keep the software from going through the whole document. So, if you want to digitize the, for example, a C1 plot that is represented by a red line, the software will connect all elements in the whole document which are red, rendering a crisscross of polylines which do not represent what needs to be digitized, see Figure 5.

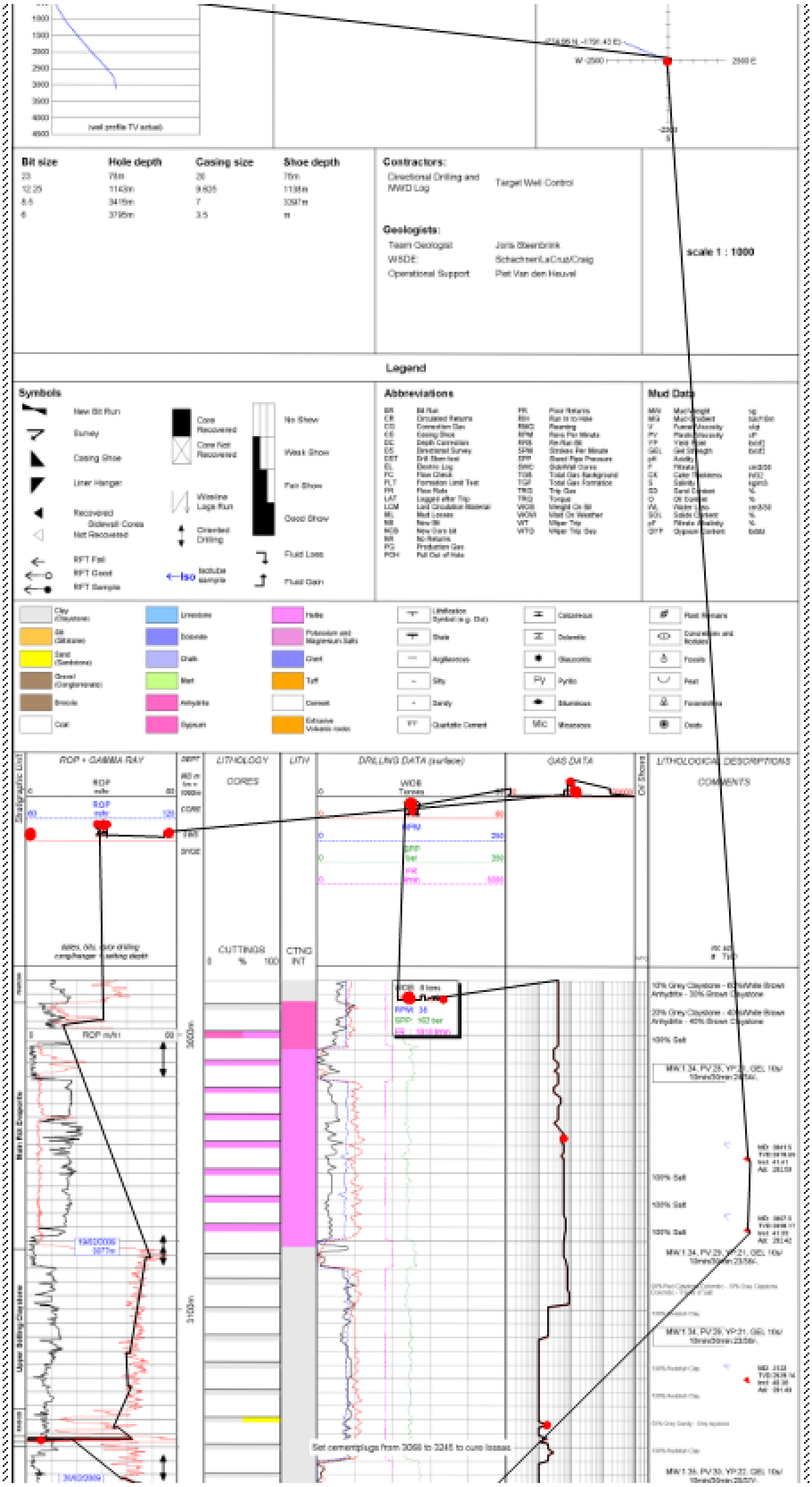


Figure 5. Snapshot of the result of an attempt to digitize the (red lined) C1 gas signal from a chromatograph of a log image in Didger5 by creating a polyline. Instead of being restricted to the chromatograph area, the software connects all the features on the log image that have red in them.

5 Automation and excel interface

5.1 Programming language

Throughout my bachelor and master studies, I have gained experience with several programming languages of which each one has its pros and cons. However, the HC show database is stored within an Excel-file and most people have at least some experience using excel. For this reason, the decision was made to code in Visual Basic for Applications (VBA), which is an object-based and event-driven programming language developed by Microsoft. It is directly coupled to Excel, this makes this project easy to pass on and to be used by other people.

5.2 Manual gas log analysis workflow, a brief explanation

To have a general understanding on how to perform the manual gas log analyses, a summary of the workflow from Klop (2018) is given below:

1. The stratigraphic intervals, the top depth and the base depth of the layers have to be exported from EBNs basisregistratie boringen and inserted in 3 columns of the 'Log data' sheet of the HC show database master file.
2. If the data is displayed in a chromatograph, the alkanes (C1' to 'nC5) that are measured are displayed in the legend at the top of the graph. There is a column for each alkane in the log data sheet. When an alkane is not mentioned, the cells of the according column are filled in with 'NOT MEASURED'.
3. Per stratigraphic interval the highest C1 gas reading (peak) is read off and noted in the cells of the C1 column corresponding the said stratigraphic interval. Also, the depth of the highest gas reading and the corresponding lithology is noted in the log data sheet. Peak that coincide with coal beds or bituminous layers are ignored. Furthermore, the values of the remaining alkanes (C2' to 'nC5) at the same depth are noted.
4. The sum of all alkanes of a stratigraphic interval is noted in the 'Total HC gas' column.
5. The background gas is the average gas value when ignoring the maxima and minima of a stratigraphic section. The background gas associated with the highest gas reading is estimated and noted down in the 'Background gas' column.
6. Once, the information listed above has been noted down correctly, the stratigraphic interval can be classified with a show label according to Figure 6. PtBR stands for 'Peak-to-Background-Ratio', which is the highest gas reading divided by the background.

Gas show classification					
		NO SHOW	POOR	FAIR	GOOD
Interpreted	**Peak gas**	< 500 ppm (0,05%)	> 500 ppm (0,05%)	> 500 ppm (0,05%)	> 1000 ppm (0,1%)
	PtBR	< 2	2 < PtBR < 3	> 3	> 5
	Lithology & composition / grading & thickness	Halite/anhydrite	Mudstone/shale/ claystone/marl	siltstone	Sandstone/dolomite/ limestone/chalk

Figure 6. existing rules of thumb for the gas show classification. The top two rows 'Peak gas' and 'PtBR' hand a 'Raw' show, the combination with a corresponding lithology hands an 'Interpreted' show.

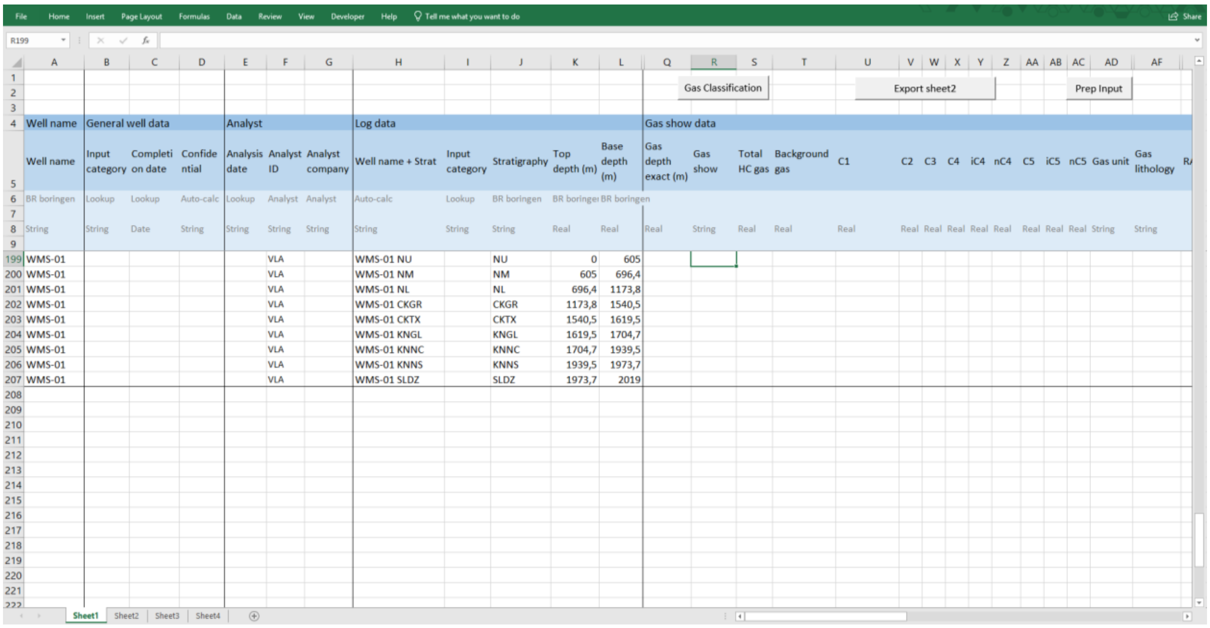
7. There are some exceptions to these rules of thumb, the most important ones:
    a. *"If the background reading in a sandstone is already in the 10000s of ppm, then a peak of 3 times the background reading can be considered a GOOD show, even though according to the above scheme it would be at the POOR/FAIR boundary."*
    b. *"Shows in the shallow North Sea Group (NU). As this stratigraphic group often comprises of unconsolidated material, shows might indicate shallow gas potential. The group has generally high porosity, even if the lithology is clay. Therefore, strong shows in clay in this formation should be classified as GOOD."*

5.3 Gas log Analyser Tool (GAT), workflow automated analyses

To automate the gas log analyses, I had to write a code that mimics the log data analyses workflow earlier created as part of QC1. I started at the core of this workflow which, to me, were the rules of thumb containing the gas show classification, see Figure 6.

Once an *Excel macro,* which could classify the data listed above, was established, I worked my way outside as much as possible. To main code can be found in the appendix in section 10.

The GAT works as a separate HC Show database 'Log data' sheet file, where the code can perform the analyses and write the classification eventually, see Figure 7. In the end, the automated classification data can be uploaded to the HC show database master file. The basic workflow of the created tool is visualized in a block scheme in Figure 14



Well name	Input category	Completion date	Confidential	Analysis date	Analyst ID	Analyst company	Well name + Strat	Input category	Stratigraphy	Top depth (m)	Base depth (m)
BR boringen	Lookup	Lookup	Auto-calc	Lookup	Analyst	Analyst	Auto-calc	Lookup	BR boringen	BR boringen	BR boringen
String	String	Date	String	String	String	String	String	String	String	Real	Real
WMS-01					VLA		WMS-01 NU		NU	0	605
WMS-01					VLA		WMS-01 NM		NM	605	696,4
WMS-01					VLA		WMS-01 NL		NL	696,4	1173,8
WMS-01					VLA		WMS-01 CKGR		CKGR	1173,8	1540,5
WMS-01					VLA		WMS-01 CKTX		CKTX	1540,5	1619,5
WMS-01					VLA		WMS-01 KNGL		KNGL	1619,5	1704,7
WMS-01					VLA		WMS-01 KNNC		KNNC	1704,7	1939,5
WMS-01					VLA		WMS-01 KNNS		KNNS	1939,5	1973,7
WMS-01					VLA		WMS-01 SLDZ		SLDZ	1973,7	2019

Figure 7. GAT 'Log data' sheet file for automated classification. In the top of the sheet 3 buttons are present. The 'Prep input' button lets the user select and pre-process input gas data of a borehole. The 'Gas Classification' button initiates the automated classification and the 'Export sheet2' offers the user the opportunity to store the used and pre-processed input data, which is temporarily written the Sheet2, to a separate file.

Before running the automated classification digital gas data of a borehole from LAS-, ASCII- or .txt-files needs to be opened in excel, see Figure 8. This data is going to be the input for the code.

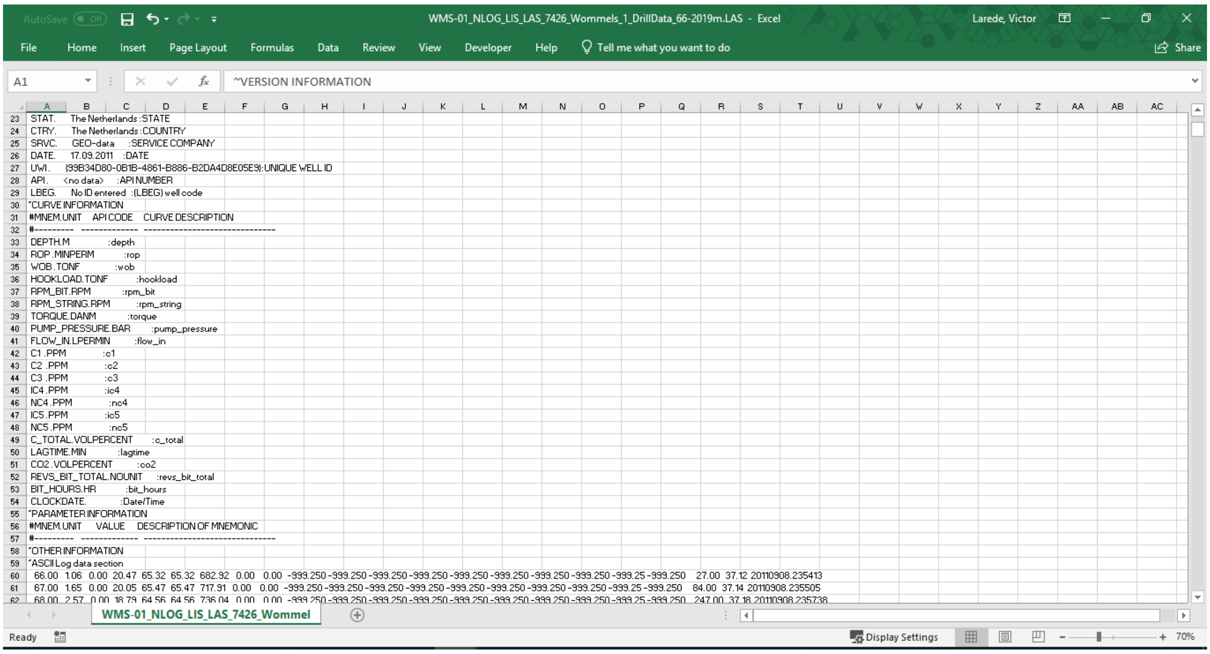


Figure 8. LAS-file opened in Excel. Note that all the metadata and measurement data is stored in the first column.

Often the data needs to be sorted into separate columns, see Figure 8. This can be done via the excel 'Data-tab' with 'text to columns', see Figure 9.

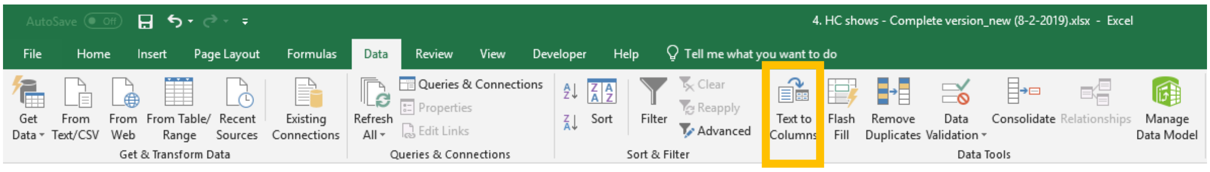


Figure 9. Excel data-tab, text data can be sorted in columns with the 'Text to Columns' button (inside yellow box).

If a log image is available, it is advisable to screen it for indications of events that may cause deflections in the measurements, such as equipment breaking down, connection- or trip gas etc. Usually, these events are pointed out in the log images. They can be seen in the chromatograph as large sudden maxima or minima, see Figure 10.

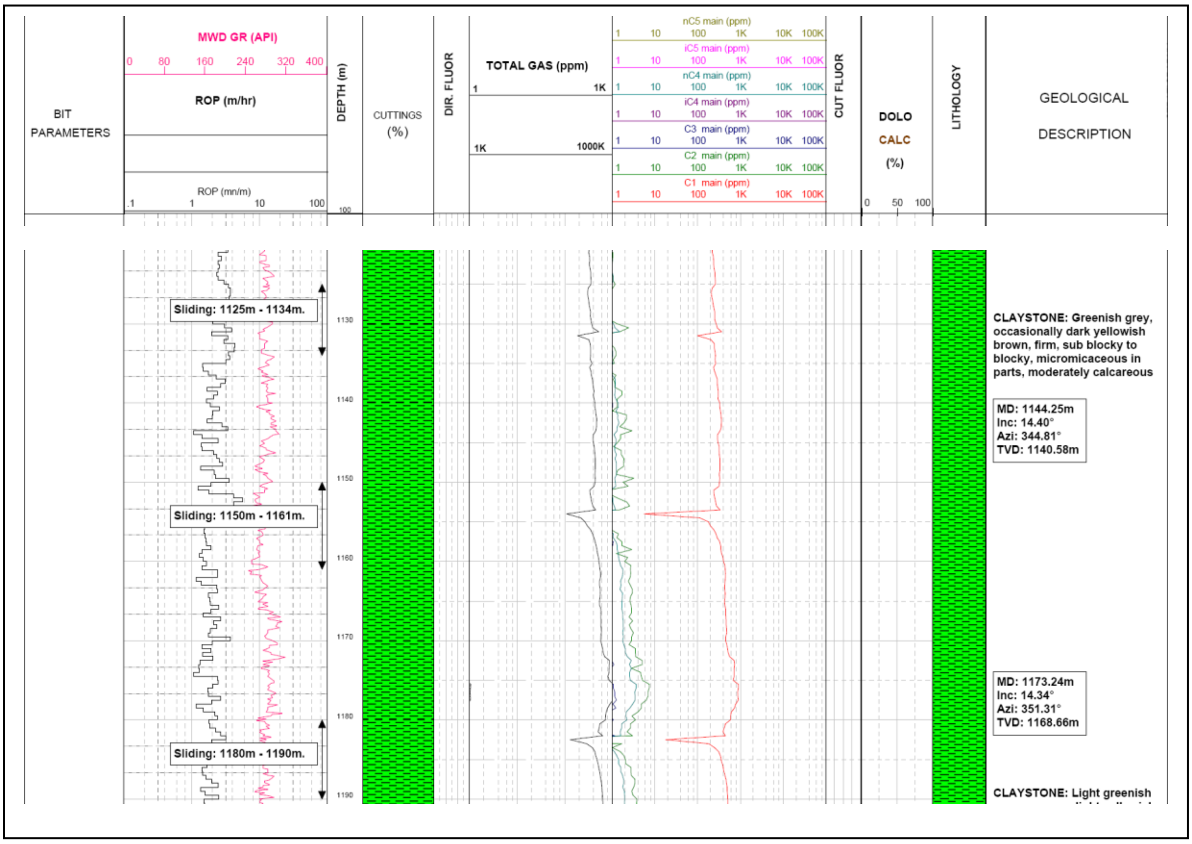


Figure 10. Example of deflections in the gas-measurements. On the left side of the log image is indicated at which depths sliding of the bit took place; at the locations of these depths' abrupt wiggles in the C1 gas chromatograph are observed.

If needed, these deflections can then be adjusted so that they will not influence the automated analysis, for example by replacing the deviating data with an average in line with the preceding and succeeding data.

By clicking on the 'Prep input' button in the automated classification file an input-window pops up in which you can select the data to be analysed, see Figure 11. The code needs at least input data on depth, C1 and lithology of the borehole. Furthermore, a window interval over which the background can be calculated needs to be inserted.

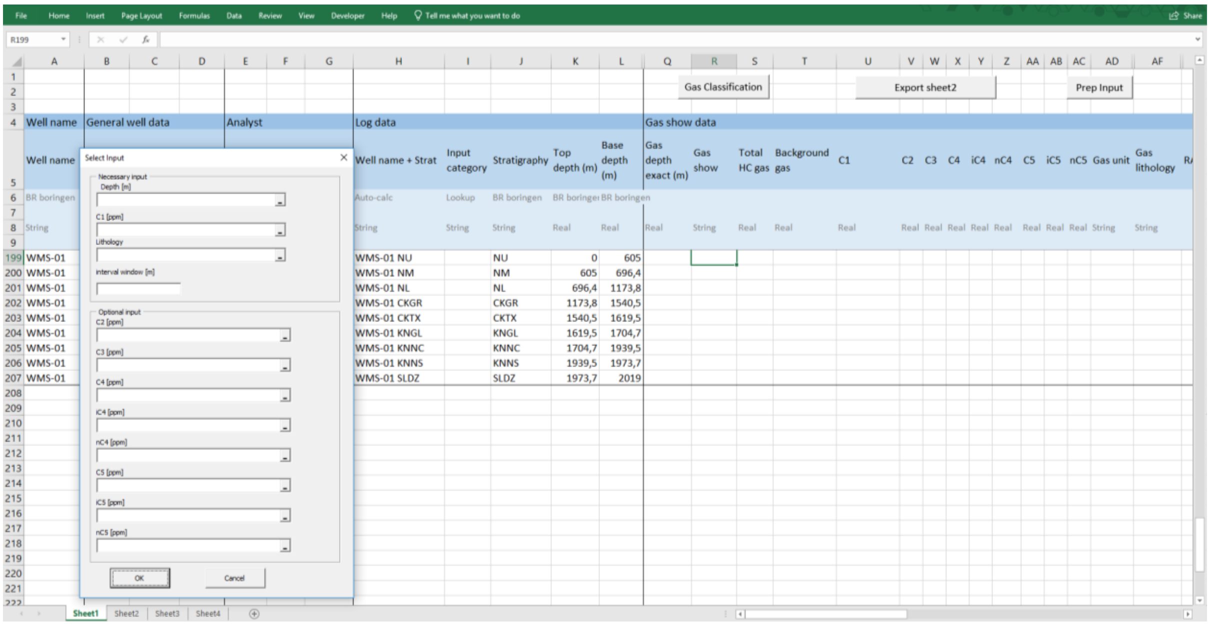


Figure 11. Pop-up window in which the user can select the data to be analysed. The popup-up window is called by clicking on the 'Prep input' button.

After inserting the columns containing the data, the program copies it to the second sheet of the 'Automated-classification' file, see Figure 12.

- A background gas signal is calculated, this will be discussed in detail in section Error! Reference source not found..
- The C1 signal is filtered by the calculated background, the filtered signal is set to 1 where the background exceeds the C1 signal.
- Then, the filtered signal is filtered by lithology. Setting the signal to 1 where associated lithologies would lead to a NO SHOW classification, so that peaks at thes3 depths will be ignored by the classification algorithm.
- The C1 gas data vs. depth are plotted on a chart, as the C1 gas log is leading for the classification. Likewise, the calculated background signal and C1 signal filtered by lithology and background is plotted.

The reason for copying the input data to Sheet2 of the 'Automated-classification' file is to give the user the opportunity to view the (pre-processed) data in a well-organized manner (the digital files often contain a lot of data that is not used in the classification). Furthermore, it offers the chance to adjust the data that is to be analysed, if necessary.

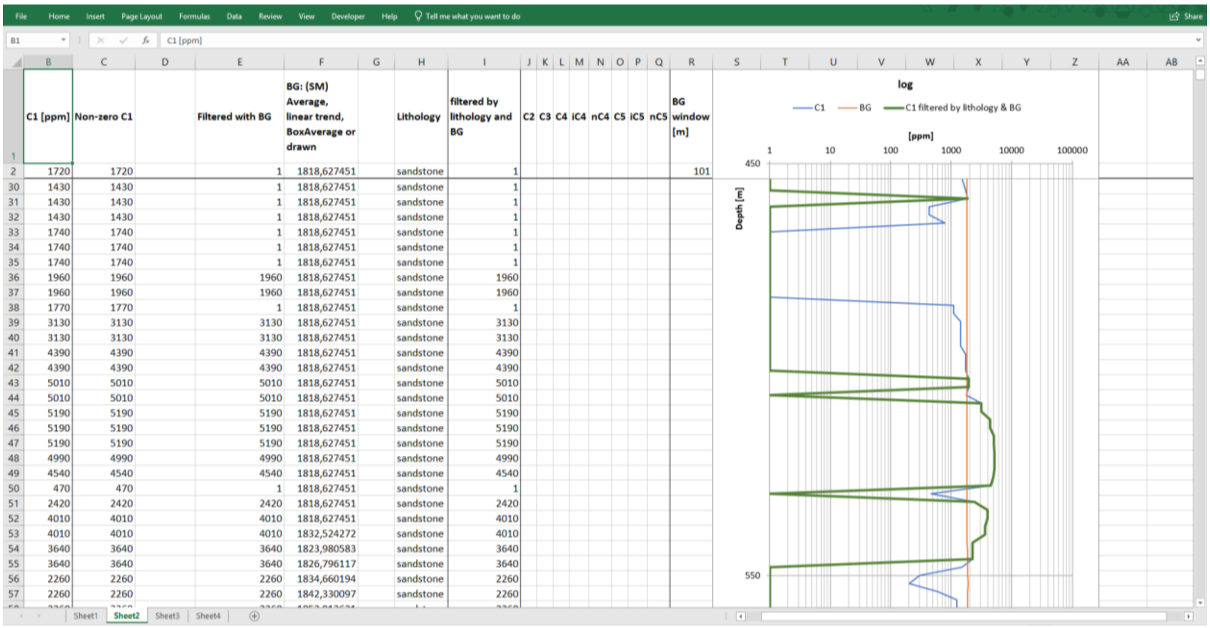


Figure 12. Sheet2 of the 'automated classification' file. Here, the input data is stored during the classification. Furthermore, pre-processed data such as a calculated background and the filtered gas signal is stored here as well. A chart shows the unfiltered C1 signal, the background signal and the filtered signal. Creating this sheet offers the user to screen the data a last time before finalising the classification. After finishing the classification, this sheet can be exported and stored in a database to be able to look up the details of the classification later.

When all of this is done, the real classification can begin, see Figure 13. By clicking on the 'Gas classification' button a pop-up window appears. The user needs to select the borehole range in the 'Gas show' column of the 'Log data' sheet (Sheet1 in Figure 13). The code requires this selection because it uses the selected range as a reference to fetch the data on the stratigraphic intervals needed for the classification per layer from the neighbouring columns in Sheet1. It also uses this range to write the output from the classification to the cells in the 'Gas show data' columns.

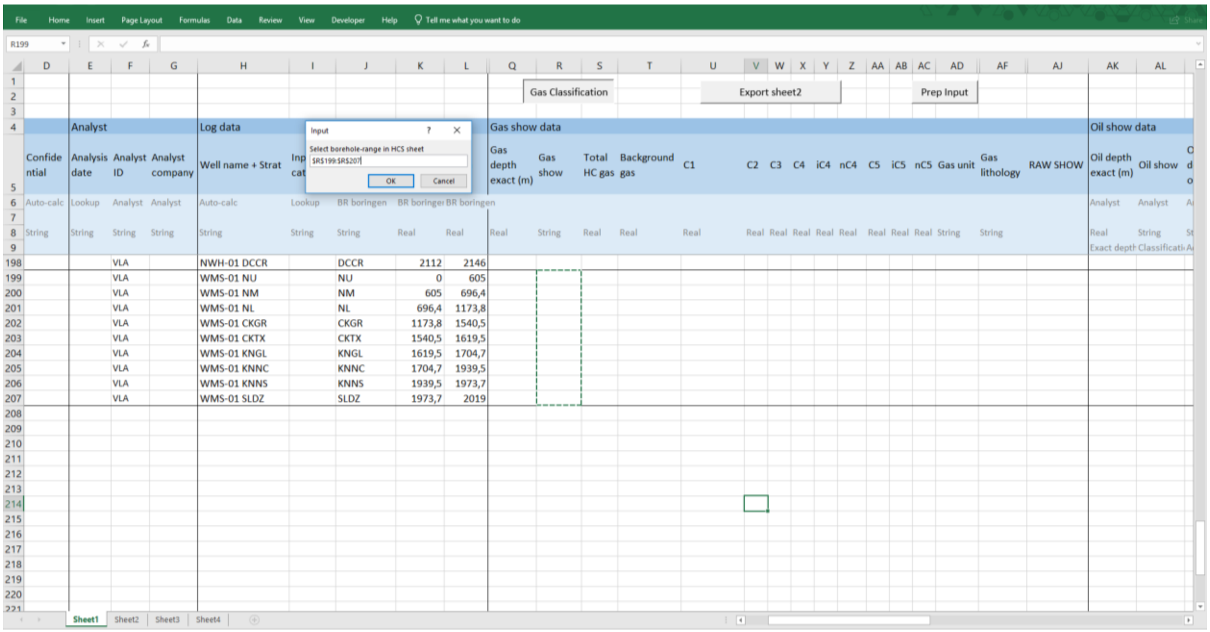


Figure 13. After pre-processing the data, the stratigraphic intervals can be classified in the Gas show data columns. Clicking the 'Gas classification button' and then selecting the boreholes range in the sheet initiates the final stage of the automated classification.

Lastly, sheet2 can be exported to a separate file after the classification with the click on and 'Export' button. This functionality is included, so that only the data used for the classification can be stored in one of EBN's databases, for example Livelink, to be able review the data later for any kind of reason.

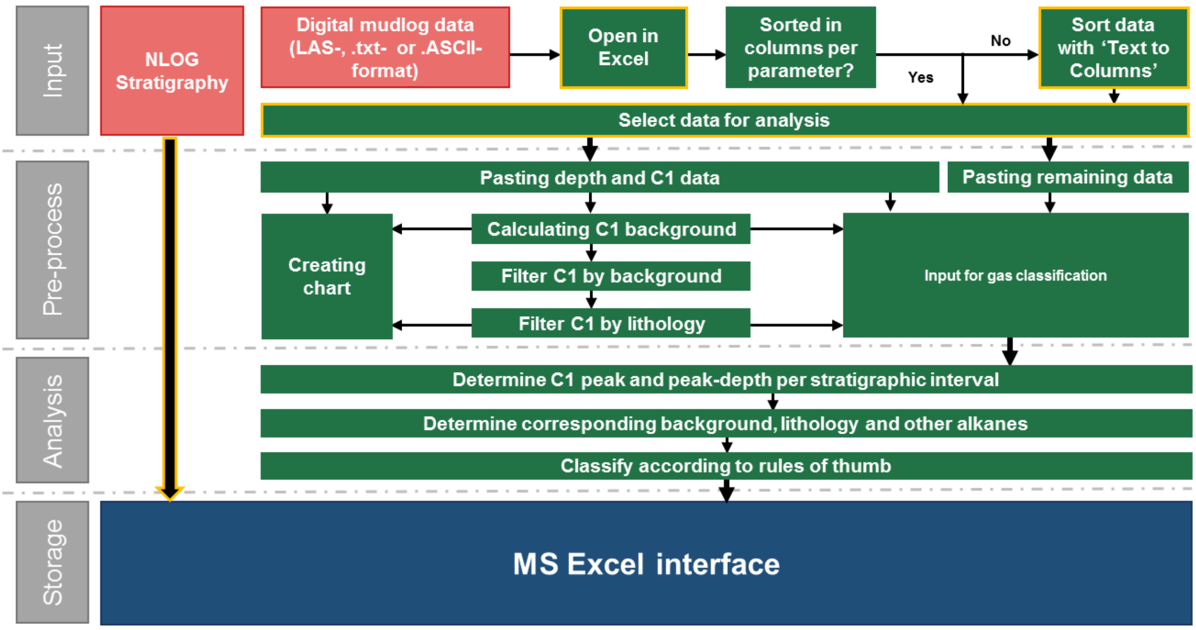


Figure 14. Block scheme of the workflow used by GAT and code performing the automated classification.

6 Background gas

"Background gas is the relative baseline against which all other gas shows are compared." (Heerema, 2016.)

"The background gas is the average value of a stratigraphic section when ignoring the peaks and minima." (Klop, 2018)

Just two statements from former internship reports defining what background gas is. The amount to pretty much the same but are also not defined very strictly.

The background gas signal can be influenced by several factors, like; mud type, mud weight vs. pore pressure, the amount of sediment and lithology.

Determining the background gas automatically, is the most difficult task in the automation, as its description is not very straightforward and it might require a lot of (subjective) geological insight to properly determine the signal. In this research, a few possible techniques to determine the background gas, have been explored. They are listed in

Table 1 and explained in the sections below.

Table 1. list of Background gas methods that were tested.

	Background method	subgroup
1	Linear trendline	
2	Constant average	
3	Average per stratigraphy	
	Simple moving average	
4		101 [m]
5		251 [m]
6		501 [m]
	Quartile filtered average	
7		Average between q1 and q2
8		Average between q1 and q3
	Quartile filtered average over fixed intervals	
9		101 [m]
10		251 [m]
11		501 [m]
	Moving quartile filtered average	
12		101 [m]
13		251 [m]
14		501 [m]
15	Drawn	

To e validate the GAT, the gas logs of 9 different well locations (Figure 16) were analysed manually as well as automatically testing 8 methods to calculate/determine the background gas. The wells contained data on a total of 107 stratigraphic intervals. For all methods used a root mean square (RMS) error was calculated according to equation 1 (see Figure 15 for a clarification of the variables):

$$RMS_{error} = \sqrt{\frac{\sum(observed - predicted)^2}{number\ of\ observations\ -\ 1}} \qquad \text{eqn(1)}$$

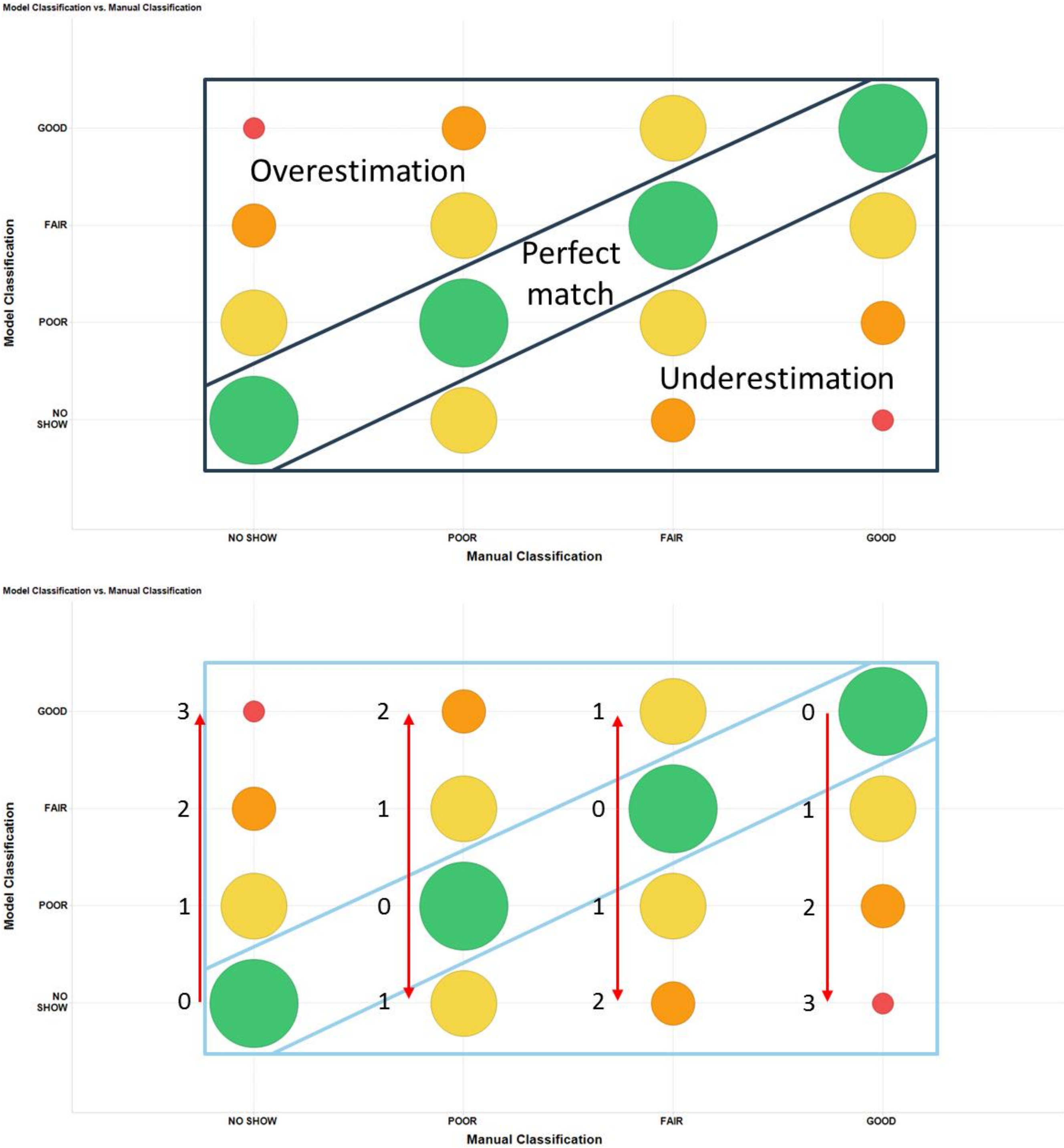


Figure 15. Images clarifying how the rms error is calculated. Manual classification on the x-axis versus Model classification on the y-axis. In the top of the figure one can see that perfect matches (green circles) plot on the bottom-left to top-right diagonal. Model overestimation plots in the top-left corner indicated in the square and model underestimation plots in the bottom-right corner. Error are outcomes that deviate from the 'perfect match'-diagonal. Under the assumption that the manual classification is was conducted perfectly, they are considered as observations. The outcomes from the model are then the predictions. The errors are then calculating the absolute value the subtraction of the predicted values from the observed values, denoted by the numbers along the red arrows.

The comparison of manual classification versus automated classification will be visualized in Figure 17 to Figure 22 and Figure 24 to Figure 32 and quantified in Table 2 to Table 16. Obviously, the better the results of an automated method match the manual results, the more confidence there will be in the associated algorithm.

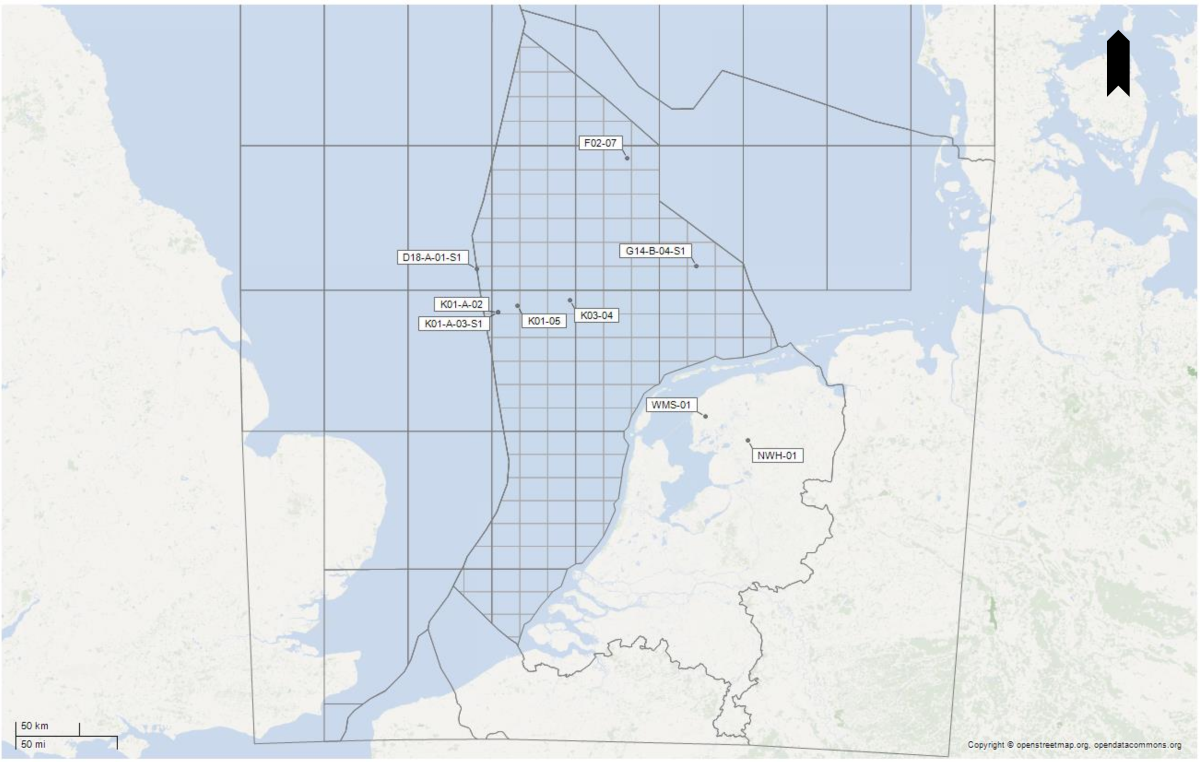


Figure 16. Map of the Dutch on- and offshore including the location of the 9 boreholes that are used to test the background scenario's, labelled with their names.

6.1 Fourier analysis

Gas logs can show a lot of oscillations or variations, which can have multiple causes, e.g. varying lithologies etc. Fourier series are built up out of sines and cosines and may be able to represent such a signal. Using a Discrete Fourier transform (DFT), it would then be possible to filter out the dominant frequency of the gas signal. This could represent the background gas signal of the gas measurement.

In practice this method cannot easily be applied. Excel does offer the possibility to perform Fourier analysis with the 'Data Analysis Tools' package. However, to make this work, the number of input data values must be a power of 2 and must not exceed 4096. As most of the datasets do not have a number of data points which is a power of 2 or do have a number of datapoints which exceeds 4096, this tool was found not suitable for the automation. Reducing the dataset to a correct number of datapoints does not fit the purpose of the automation as you could lose valuable information for the analysis. For example, the highest gas reading in a layer could then, by accident, be left out and missed. Therefore, the decision was made to test simpler methods to determine the background and observe how well they approach 'reality'.

6.2 One linear trendline or a constant average as background

The first two methods use all the gas-data of a borehole and calculate a background as a constant average value or linear trendline. These methods do not meet the criterion as given by Klop (2018), but it is interesting to look how well they perform.

6.2.1 Simple linear regression

A simple linear regression is performed to calculate a linear trendline with the gas values in a borehole to represent the background according to equation 2:

$$y = \beta_0 + \beta_1 x \qquad \text{eqn.2}$$

Here *y* represents the background gas [ppm] and *x* represents depth [m], β_0 is the y-intercept and β_1 is the slope. The gas samples are the dependent variables y_n and the depth values are the independent variables x_n. First, the means $\bar{x}$ and $\bar{y}$ of both the dependent and independent variables are determined. All linear regression lines go through the point where these two means cross. Second, the difference between the observations and the means is calculated:

$$x - \bar{x}$$

$$y - \bar{y}$$

Then, $x - \bar{x}$ is squared $(x - \bar{x})^2$ and $(x - \bar{x})(y - \bar{y})$ is calculated. Now β_1 can be determined.

$$\beta_1 = \frac{(x - \bar{x})(y - \bar{y})}{(x - \bar{x})^2} \qquad \text{eqn.3}$$

Since all linear regression lines go through the point where the two means cross, β_0 can now be calculated as well by rewriting equation 2 to equation 4:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \qquad \text{eqn.4}$$

$$y_n: \textit{dependent variables}$$

$$x_n: \textit{independent variables}$$

$$\bar{x}: \textit{mean of } x_n$$

$$\bar{y}: \textit{mean of } y_n$$

$$\beta_0: y - \textit{intercept}$$

$$\beta_1: \textit{slope}$$

6.2.1.1 Result linear trendline

Figure 17. 'Manual classification' on the x-axis versus the 'Automated classification' with a a linear trendline as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 2. 'Automated classification' with a linear trendline as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		Linear Trend				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	58	3	0	0	0.95
	POOR	11	7	0	3	0.33
	FAIR	6	0	4	3	0.31
	GOOD	5	0	0	7	0.58

RMS error: ~0.96

6.2.1.2 Result constant average

Figure 18. 'Manual classification' on the x-axis versus the 'Automated classification' with a constant average as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 3. Automated classification' with a constan average as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		Constant average				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	57	4	0	0	0.93
	POOR	9	9	0	3	0.43
	FAIR	6	0	3	4	0.23
	GOOD	1	1	1	9	0.75

RMS error: ~0.80

6.3 Average per stratigraphic interval (APS)

An average gas value is calculated per stratigraphic interval. This comes close to the definition in (Klop, 2018).

6.3.1 Result average per stratigraphy

Figure 19. 'Manual classification' on the x-axis versus the 'Automated classification' with a average gasvalue per stratigraphic interval as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified Circles are sized by the amount of data.

Table 4. Automated classification' with an average gas value per stratigraphy as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		Average per Stratigraphy				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	55	4	2	0	0.90
	POOR	3	14	2	2	0.67
	FAIR	1	3	8	1	0.62
	GOOD	1	0	0	11	0.92

RMS error: ~0.63

6.4 Simple moving average (SMA)

A simple moving average is the average value $\bar{y}_{SM}$ calculated over a window-range of *n* data points in the centre of that range. So, for example, if you would calculate the single moving average with a window of 3 data points, namely point 1, 2 and 3, then the average over these points will be calculated in point 2. The window-range keeps the same length, although the window which represents the range will move over a whole dataset.

$$\bar{y}_{SM} = \frac{y_M + y_{M-1} + \cdots + y_{M-(n-1)}}{n}$$

In this research, 3 window ranges of 101, 251 and 501 meters are tested on how well they replicate the background gas.

The background signal cannot be calculated with a simple moving average at the beginning and the end of a borehole, because the window-range has not yet been reached or the end of the window has already reached the end of the data set. At these depths, the background is calculated as the average over the preceding or remaining data.

6.4.1 Result SMA

6.4.1.1 101 [m]

Figure 20. 'Manual classification' on the x-axis versus the 'Automated classification' with a simple moving average of 101 [m] (SMA101) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 5. Automated classification' with a simple moving average interval of 101 meter (SMA101) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		SMA101				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
manual	NO SHOW	60	1	0	0	0.98
	POOR	3	18	0	0	0.86
	FAIR	1	6	6	0	0.46
	GOOD	2	1	2	7	0.58

Rms error: ~0.66

6.4.1.2 251 [m]

Figure 21. 'Manual classification' on the x-axis versus the 'Automated classification' with a simple moving average of 251 [m] (SMA251) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 6. Automated classification' with a simple moving average interval of 251 meter (SMA251) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches Succes ratio per manual classification label is in the outer-right column.

		SMA251				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	57	4	0	0	0.93
	POOR	4	15	2	0	0.71
	FAIR	2	1	10	0	0.77
	GOOD	1	1	1	9	0.75

RMS error: ~0.56

6.4.1.3 501 [m]

Figure 22. 'Manual classification' on the x-axis versus the 'Automated classification' with a simple moving average of 501 [m] (SMA501) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 7. Automated classification' with a simple moving average interval of 501 meter (SMA501) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		SMA 501				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	57	3	1	0	0.93
	POOR	4	15	2	0	0.71
	FAIR	3	1	9	0	0.69
	GOOD	1	0	2	9	0.75

RMS error: ~0.59

6.5 Quartile filtered average

Sort the gas data of a single stratigraphic interval. Taking the 1st, 2nd and 3rd quartiles of the data provides information on where the minima and maxima of a layer must lie, see Figure 23. The 2nd quartile is the median of the data. The 1st quartile is the median between the lowest value of the data and the 2nd quartile and the 3rd quartile is the median between the highest value between the highest value and the 2nd quartile. So, basically the 1st and 3rd quartile denote the boundaries of the 25% lowest and highest values of the data. Setting the background to the average of the values between the 1st and 2nd quartile or 1st and 3rd quartile, and thus neglecting the outliers, comes closest to the definition of the background as described in (Klop, 2018).

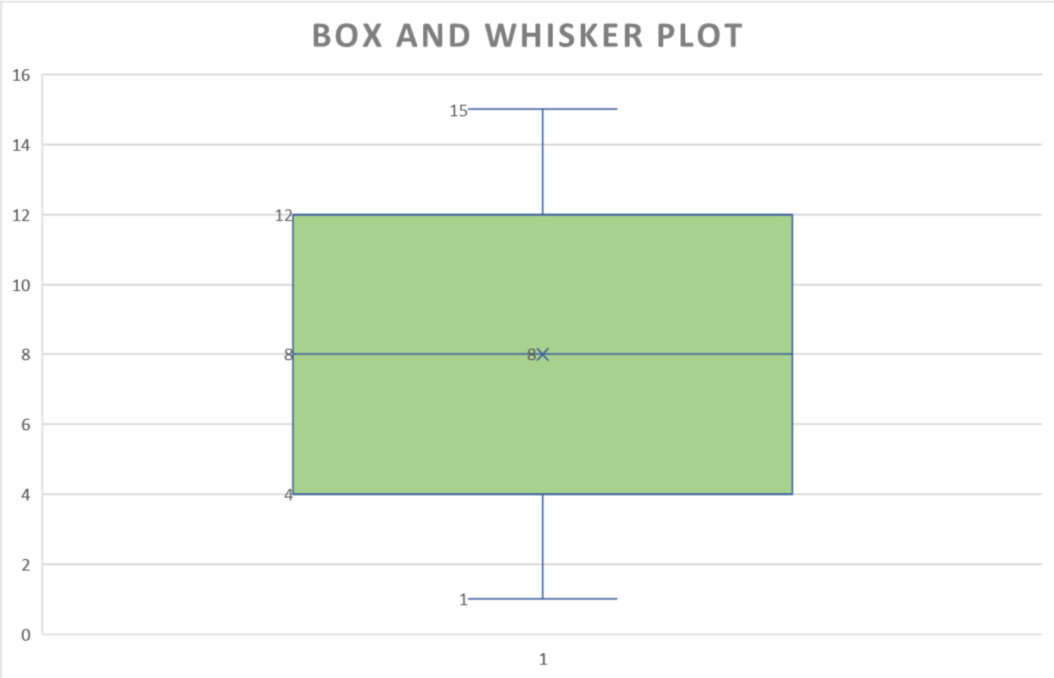


Figure 23. A box and whisker plot of a fictional dataset of values 1 to 15. The box is vertically bounded by quartile 1 and quartile 3, and centred by quartile 2. The vertical outliers denote the absolute minimum and maximum of the dataset.

6.5.1 Results Quartile filtered average

6.5.1.1 Average between quartile 1 and quartile 2 (BOX)

Figure 24. 'Manual classification' on the x-axis versus the 'Automated classification' with an average gas value per stratigraphic interval between the 1st and 2nd quartile (BOX) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 8. Automated classification' with an average gas value per stratigraphic interval between the 1st and 2nd quartile (BOX) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		BOX				
		NO SHOW	POOR	FAIR	GOOD	SUCCES
Manual	NO SHOW	53	6	2	0	0.87
	POOR	1	15	2	3	0.71
	FAIR	0	2	9	2	0.69
	GOOD	2	0	0	10	0.83

RMS error: ~0.67

6.5.1.2 Average between quartile 1 and quartile 3 (BOX2)

Figure 25. 'Manual classification' on the x-axis versus the 'Automated classification' with a an average gas value per stratigraphic interval between the 1st and 3rd quartile (BOX2) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 9. Automated classification' with an average gas value per stratigraphic interval between the 1st and 3rd quartile (BOX2) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		BOX2				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	54	5	2	0	0.89
	POOR	2	15	1	3	0.71
	FAIR	1	2	8	2	0.62
	GOOD	2	0	0	10	0.83

RMS error: ~0.66

6.6 Quartile filtered average over fixed intervals

To further investigate the functionality of using quartile filtering. Taking a quartile filtered average over a fixed depth interval is the next step. In this case we only look at an average taken over the 1st quartile-3rd quartile range with 3 intervals of 101, 251 and 501 m.

6.6.1 Result quartile filtered average over fixed interval

6.6.1.1 101 [m]

Figure 26 'Manual classification' on the x-axis versus the 'Automated classification' with a quartile filtered average over a fixed interval of 101 meter (WindowBOX101) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 10. Automated classification' with a quartile filtered average over a fixed interval of 101 meter (WindowBOX101) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		WindowBOX101				
		NO SHOW	POOR	FAIR	GOOD	Success Ratio
Manual	NO SHOW	56	5	0	0	0.91
	POOR	2	18	0	1	0.86
	FAIR	1	0	8	4	0.61
	GOOD	1	0	0	11	0.92

RMS error: ~0.51

6.6.1.2 251 [m]

Figure 27. 'Manual classification' on the x-axis versus the 'Automated classification' with a quartile filtered average over a fixed interval of 251 meter as background (WindowBOX251) on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 11. Automated classification' with a quartile filtered average over a fixed interval of 251 meter as background (WindowBOX251) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		WindowBOX251				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	54	6	1	0	0.89
	POOR	1	16	4	0	0.76
	FAIR	0	1	7	5	0.54
	GOOD	1	0	2	9	0.75

RMS error: ~0.55

6.6.1.3 501 [m]

Figure 28. 'Manual classification' on the x-axis versus the 'Automated classification' with a quartile filtered average over a fixed interval of 501 meter (WindowBOX501) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 12. Automated classification' with a quartile filtered average over a fixed interval of 501 meter (WindowBOX501) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		WindowBox501				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	54	4	3	0	0.89
	POOR	0	17	1	3	0.81
	FAIR	0	0	7	6	0.54
	GOOD	0	0	0	12	1

RMS error: ~0.57

6.7 Moving quartile filtered average

Another possible way to calculate the background using quartile filtering is to calculate the background with a moving window, in the same way as with the simple moving averages. In this case we only look at an average taken over the 1st quartile-3rd quartile range with 3 intervals of 101, 251 and 501 m.

6.7.1 Result moving quartile filtered average

6.7.1.1 101 [m]

Figure 29. 'Manual classification' on the x-axis versus the 'Automated classification' with a moving quartile filtered average over a fixed interval of 101 meter (MovingWindowBOX101) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 13. Automated classification' with a moving quartile filtered average over a fixed interval of 101 meter (MovingWindowBOX101) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		MovingWindowBOX101				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	60	1	0	0	0.98
	POOR	3	18	0	0	0.86
	FAIR	1	6	6	0	0.46
	GOOD	2	1	2	7	0.58

RMS error: ~0.59

6.7.1.2 251 [m]

Figure 30. 'Manual classification' on the x-axis versus the 'Automated classification' with a moving quartile filtered average over a fixed interval of 251 meter (MovingWindowBOX251) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 14. Automated classification' with a moving quartile filtered average over a fixed interval of 251 meter (MovingWindowBOX251) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		MovingWindowBOX251				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	57	4	0	0	0.93
	POOR	4	15	2	0	0.71
	FAIR	2	1	10	0	0.76
	GOOD	1	1	1	9	0.75

RMS error: ~0.56

6.7.1.3 501 [m]

Figure 31. 'Manual classification' on the x-axis versus the 'Automated classification' with a moving quartile filtered average over a fixed interval of 501 meter (MovingWindowBOX501) as background on the y-axis. Green circles denote a perfect match, yellow dots denote a misfit of 1, orange circles denote a misfit of 2 and red circles denote a misfit of 3. The white labels denote how many times a match or mismatch has been classified. Circles are sized by the amount of data.

Table 15. Automated classification' with a moving quartile filtered average over a fixed interval of 501 meter (MovingWindowBOX501) as background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches. Succes ratio per manual classification label is in the outer-right column.

		MovingWindowBOX501				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	27	3	1	0	0.93
	POOR	4	16	2	0	0.73
	FAIR	3	1	9	0	0.69
	GOOD	1	0	2	9	0.75

RMS error: ~0.59

6.8 Draw/trace background with GetData Graph Digitizer 2.2.6

Another way to determine the background is to 'draw' it on a log. Using *GetData Graph Digitizer 2.2.6* you can import a log image. After defining the log axes, you can select data point representing the background gas which can then be exported to, for example, and excel-file. Although this may be a step back in terms of automation, it does not necessarily mean that it will be less efficient.

6.8.1 Results drawn background

Figure 32. 'Manual classification' on the x-axis versus the 'Automated classification' with a digitally drawn signal as background on the y-axis. Green circles denote a perfect match and yellow dots denote a misfit of 1. The white labels denote how many times a match or mismatch has been classified.

Table 16. Automated classification' with a 'drawn' background versus 'Manual classification'. The top-left to bottom-right diagonal denotes prefect matches, the rest denotes misfits.

		Drawn background				
		NO SHOW	POOR	FAIR	GOOD	Success ratio
Manual	NO SHOW	57	4	0	0	0.93
	POOR	2	18	1	0	0.86
	FAIR	0	0	9	4	0.69
	GOOD	0	0	0	12	1

RMS error: ~0.31

6.9 RMS errors

Table 17 lists all the RMS errors of the automated classification with different backgrounds a visualization is show in the bar chart of Figure 33.

The method where a background gas signal is drawn/traced performs the best. The analysis takes around 10-15 minutes which is faster than the estimate 2 hours estimated for the manual analyses mentioned in section 4.

Table 17. Overview of all the background methods that have been tested and their RMS error.

	Background method	**subgroup**	**RMS error**
1	Linear trendline		0.96
2	Constant average		0.80
3	Average per stratigraphy		0.63
	Simple moving average		
4		101 [m]	0.66
5		251 [m]	0.56
6		501 [m]	0.59
	Quartile filtered average		
7		Average between q1 and q2	0.67
8		Average between q1 and q3	0.66
	Quartile filtered average over fixed intervals		
9		101 [m]	0.51
10		251 [m]	0.55
11		501 [m]	0.57
	Moving quartile filtered average		
12		101 [m]	0.59
13		251 [m]	0.56
14		501 [m]	0.59
15	Drawn		0.31

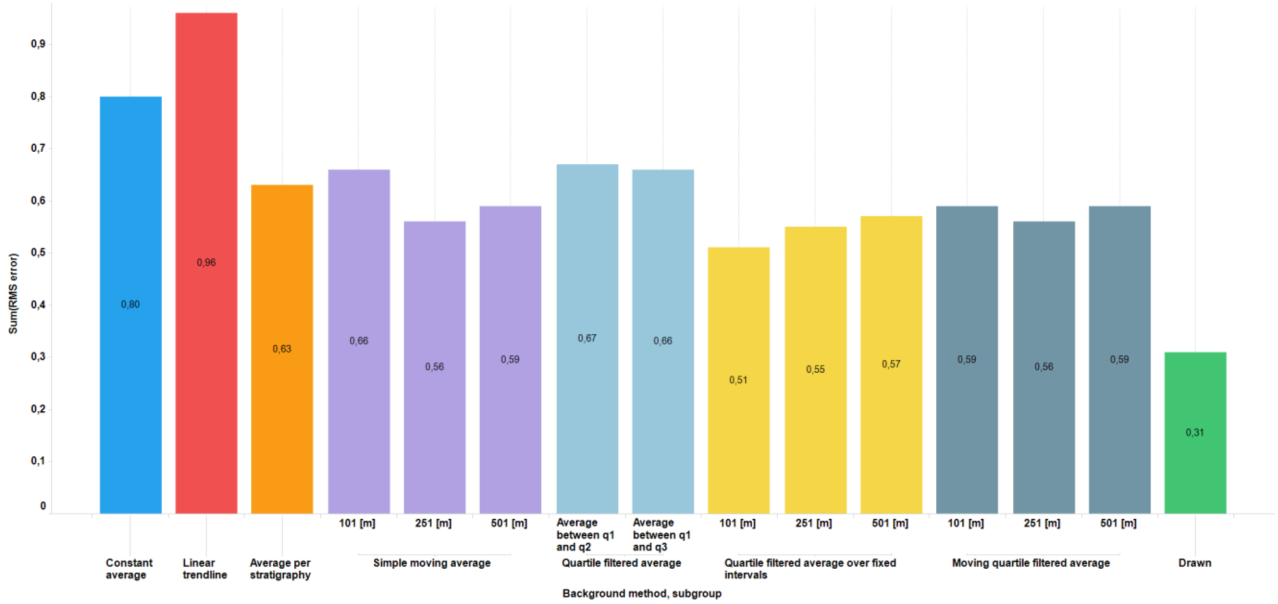


Figure 33. Bar chart of the RMS-errors listed in Table 17, methods on the x-axis and RMS-error value on the y-axis.

7 Discussion and recommendations

Referring to the general goals stated in the introduction, this internship has been successful on numerous points and has led to the following outcomes and deliverables:

- At the beginning, it was concluded that using machine learning techniques to automate the analyses for the HC show database was beyond the scope of this project and that the automation should be narrowed down to gas log analysis.
- A macro that can determine a gas show label according to EBNs gas show classification rules has been designed and tested.
- A code that can filter gas data with a calculated background signal and lithology, and which can perform the gas show classification per stratigraphic interval has been created and tested
- The drawn background gas method performed the best out of all the tested methods with an RMS-error of 0.31. It takes 10-15 minutes to perform the gas log analyses using this methods, which is much quicker than the estimated 2 hours for the manual analysis.
- The Gas log Analyser Tool (GAT) containing a user interface to run the automated gas classification has been developed.
- With *GetData Graph Digitizer 2.2.6*, a tool for digitizing an observed background has been found.
- Metrics to compare several methods have been developed.

Assuming, that the manual classification was done perfectly. Several methods for an automated calculation of the background gas level have been tested:

- The linear trend line method showed the highest RMS error. This is probably due to the fact, that if the background is calculated this way, it can only go up, go down or remain constant, which is not necessarily the case with background gas. Another shortcoming of this method is that, depending on the data that is used to calculate it, a linear trendline can return negative values. In reality, this cannot happen with background gas.
- Calculating the background as a constant average was slightly better than calculating it as a linear trendline, but still an RMS error of 0.80 remains high. A constant average as a background for an entire borehole evidently doesn't account for nuances in gas signal per stratigraphic interval or any trends that might be in the data.
- A gas average per stratigraphic interval already matches the description of the background better compared to the two methods mentioned above. The RMS error of 0.63 is much lower. But still this method doesn't account for any variations that might be associated with a peak.
- An attempt has been done to account for such variations by calculating the background as a simple moving average. Three intervals have been tried of which the results were varying. An interval of 101 meter had a slightly higher RMS error of 0.66, whereas using intervals of 251 and 501 meter handed lower RMS errors of 0.56 and 0.59, respectively. The disadvantages of using a simple moving average is that, if there are more than 1 peak within the window range of the interval, the calculated background might be too elevated to get proper show classifications. In that case the GAT will underestimate the shows. The same problem will appear when the interval is too small. Wider peaks will then be classified with a background that is too elevated and the shows will thus be underestimated as well. If the interval is very large, it might happen that the calculated background is too low, consequently the show could be overestimated.
- A quartile filtered average per stratigraphic interval might be the closest this project has come incorporating for the definition of the background. However, it is debatable how

much of the data can be considered a maxima and minima. In this case the background was calculated, first, as an average per stratigraphic interval neglecting 25 % of the lowest data values as minima and 50 % of the highest as maxima, and second, as an average neglecting 25 % of the lowest data values as minima and 25 % of the highest as maxima. Both cases didn't show promising results in terms of RMS errors, which were 0.67 and 0.66. This is higher than the result when simply taking an average value per stratigraphic interval.

- By setting a fixed interval over which the quartile filtered average could be calculated, an attempt was made to lower the RMS error. This time only an average over Q1-Q3 range (25 % of the lowest data values as minima and 25 % of the highest as maxima were neglected) was tested. In this case, the RMS error got higher as the fixed interval became larger (0.51-57).
- In another attempt to account for variations in the background associated with a peak, a moving quartile filtered average was tried. Unfortunately, this didn't result in better RMS errors.
- Lastly, a method to digitize an observed background was used, where a user can draw points on a calibrated log. Even when using this method an RMS error of 0.31 was observed.

None of the methods used to calculate background is convincingly better than all the others. The lowest error (0.51) was observed when using a quartile filtered average over a fixed interval of 101 meter. However, it too early to state that this is the best method, being that the number of 107 stratigraphic intervals that were used out of 9 boreholes were used as a dataset is too small for such statements. A larger dataset, containing more boreholes and with a larger spread on and offshore, is needed to properly test the GAT. But within the timespan of this internship, this was not feasible.

I recommend letting a group of several experienced analysers perform the manual gas log analyses on the same dataset. Compare their result and see what the RMS-error would be. This should be leading when stating whether a certain method for calculating the background gas by the GAT is best.

Regarding the gas show classification rules, I advise reconsidering the lithologies. In the current classification peaks that coincide with coal or bituminous layers must be ignored. However, this is not the case for peaks that coincide with halite or anhydrite layers, although they can only give NO SHOW classifications. In that case it might be better to ignore those peaks as well.

The focus and approach of this project was more on the automation of a pre-existing workflow rather than checking its validity w.r.t. to the geological basis. It, therefore, is advisable to investigate what geological knowledge could be added to improve the gas classification and automation, such as, for example, using the current HC show database to improve the workflow for determining the background gas in gas logs. Or investigate, using the HC show database Spotfire tool, what a common background value is per stratigraphic group, starting with the most important ones (Zechstein, Rotliegend, Trias etc.) and see whether there is a common trend. This could be added to the current classification workflow.

Furthermore, taking into account that the Dutch on- and offshore have been explored and exploited for decades already, and it thus can be reasonable to argue that currently the highest potential for gas exploration and exploitation lies within stratigraphic intervals that have been classified with POOR/FAIR show labels. Assuming that the majority of the NO SHOW and GOOD show stratigraphic intervals has been found. Therefore, it might be considered that the approach to work further on this project should be to first find methods that have a high success ratio in classifying POOR and FAIR shows and then, afterwards, finetuning the NO SHOW and GOOD show classifications.

Assuming that the majority of the gas log with digital files also have good quality log images, it is advisable to add a code that can screen the image files, as a pre-processing step, for indications of measurement deflections or unwanted data for the analyses, such as trip- and connection gas, which are commonly indicated on logs.

8 Acknowledgements

Finalising my internship, I want to thank Bastiaan Jaarsma for introducing me to Guido Hoetz who offered me the opportunity to work as an intern at EBN, for which I'm very grateful.

I want to thank Guido Hoetz for his supervision which I've experienced to be very pleasant. He led me carry out my work independently and guided me with his ideas and vision when I would get stuck.

After a successful collaboration during my master thesis, Fred Beekman once again was one of my supervisors throughout this research project. I want to thank him for being very much up-to-date and involved the Hydrocarbon show database project, which resulted in great help and recommendations.

I would also like to thank Walter Eikelenburg, Kees van Ojik, Jan Lutgert, Audrey Roustiau, Jacco van de Put, Sabine Korevaar, Thomas Kopetsky, Sander de Waal and Abdul Hamid for their help and advice in this project.

Lastly, a special thanks to all my fellow (former) interns Jessica Klop, Peter Schilder, Alex Marks, Patrick Reinhard and Vibhuti Wani for their support during my internship.

9 References

Anon, (2019). [online] Available at: https://www.asprs.org/divisions-committees/lidar-division/laser-las-file-format-exchange-activities [Accessed 26 Feb. 2019].

Baud, A. (2018). Analysis and Seismic Integration of the Geo-Drilling Events Database (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

Blom, C.A.H. (2017). Hydrocarbon Shows database: the power of systematic analysis (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands

Heerema, C. (2016). Inventory of Hydrocarbon Shows in the Northern offshore (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

Kickken, Y. (2016). Inventory of Hydrocarbon Shows in the Northern Dutch offshore (Part 2) (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

Klop, J. (2018). Hydrocarbon show database: EXPLORING VISUALIZATION OPTIONS IN TIBCO SPOTFIRE (Internship report as a part of Msc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

Korevaar, S. (2018). HC Show Database: correlating HC shows with structural attributes (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

Westerweel, J. (2017). Hydrocarbon Show Database: the benefits of a well-defined methodology (Internship report as part of MSc Earth Structure and Dynamics, Utrecht University). EBN B.V., Utrecht, Netherlands.

10 Appendix: Main code

```
Sub UFuse2()

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Save

UserForm1.Show

Application.ScreenUpdating = True

MsgBox "and we're back!"

End Sub

Sub shape2(ByVal depth As Range, C As Range, lithology As Range, interval, Optional ByVal C2 As Range = Nothing, Optional ByVal C3 As Range = Nothing, Optional ByVal C4 As Range = Nothing, Optional ByVal iC4 As Range = Nothing, Optional ByVal nC4 As Range = Nothing, Optional ByVal C5 As Range = Nothing, Optional ByVal iC5 As Range = Nothing, Optional ByVal nC5 As Range = Nothing)


MsgBox "shape2"


'------------------------------

' FOR USERFORM

'------------------------------


'sub to use function which subtracts cell-values to check shape of gas-chromatogram

'--------------------------------------

'necessary input

'--------------------------------------

Dim NonZeroC As Range


Dim store1 As Double

Dim store2 As Double


Dim cnt As Integer

Dim i As Integer

Dim j As Integer

'---------------------------------------

'interval = Application.InputBox("interval for moving average", Type:=1)


Application.ScreenUpdating = False

cnt = 0

For Each cell In C

cnt = cnt + 1

Next


'----------------------

'clearing sheet2

'----------------------
```

```
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

On Error Resume Next 'if there is no data ór chart in sheet to delete go to next line of code

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2:X1048576")
.ClearContents

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").ChartObjects.Delete

On Error GoTo 0 'disables on error resume next


'------------------------------

'paste data

'------------------------------

Cells(2, 18) = interval


For i = 2 To cnt + 1

Cells(i, 1) = depth(i - 1)

Cells(i, 2) = C(i - 1)

    If Cells(i, 2) = 0 Then            'non-zero values for plot with logarithmic scale

    Cells(i, 3) = 1

    ElseIf Cells(i, 2) = "NO USE" Then

    Cells(i, 3) = ""

    Else

    Cells(i, 3) = Cells(i, 2)

    End If

Cells(i, 8) = lithology(i - 1)

Next i


'-------------------------

'checking whether the signal is increasing or decreasing

'-------------------------

'For i = 2 To cnt + 1

'    If i = 2 Then

'    Cells(i, 4) = 0 'first cell

'    Else

'    Cells(i, 4) = PlusMinEqual(Cells(i - 1, 2), Cells(i, 2))

'    End If

'Next i


'For i = 2 To cnt + 1

'    If i = 2 Then

'    Cells(i, 5) = 0

'    Else

'    Cells(i, 5) = UpDownConc(Cells(i - 1, 4), Cells(i, 4))
```

```
'    End If
'Next i
'------------------------

'-----------------------
'pasting optional data
'-----------------------
If Not C2 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 10) = C2(i - 1)
    Next i
End If

If Not C3 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 11) = C3(i - 1)
    Next i
End If


If Not C4 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 12) = C4(i - 1)
    Next i
End If

If Not iC4 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 13) = iC4(i - 1)
    Next i
End If

If Not nC4 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 14) = nC4(i - 1)
    Next i
End If

If Not C5 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 15) = C5(i - 1)
```

```
    Next i
End If

If Not iC5 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 16) = iC5(i - 1)
    Next i
End If

If Not nC5 Is Nothing Then
    For i = 2 To cnt + 1
    Cells(i, 17) = nC5(i - 1)
    Next i
End If

End Sub
Sub DataSheet2()
'MsgBox "datasheet2 sub"
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

Dim depth As Range
Dim C As Range
Dim NonZeroC As Range
Dim lithology As Range

Dim interval As Double

Dim C2 As Range
Dim C3 As Range
Dim C4 As Range
Dim iC4 As Range
Dim nC4 As Range
Dim C5 As Range
Dim iC5 As Range
Dim nC5 As Range

Dim PrepC As Range

Dim cnt As Integer
'------------------------------------
'resetting ranges with same input data, but now in other file
```

```
'-----------------------------------

'hier misschien een do with loop?



Set depth =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2").End(xlDo
wn))

Set C = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("B2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("B2").End(xlDo
wn))

Set NonZeroC =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("C2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("C2").End(xlDo
wn))

Set lithology =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("H2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("H2").End(xlDo
wn))

interval = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(2,
18).Value



Set C2 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("J2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("J2").End(xlDo
wn))

Set C3 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("K2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("K2").End(xlDo
wn))

Set C4 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("L2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("L2").End(xlDo
wn))

Set iC4 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("M2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("M2").End(xlDo
wn))

Set nC4 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("N2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("N2").End(xlDo
wn))

Set C5 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("O2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("O2").End(xlDo
wn))

Set iC5 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("P2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("P2").End(xlDo
wn))

Set nC5 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("Q2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("Q2").End(xlDo
wn))

'Set PrepC =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("I2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("I2").End(xlDo
wn))




Call CntSub(depth, C, lithology, interval, NonZeroC, C2, C3, C4, iC4, nC4, C5, iC5, nC5)
```

```
End Sub

Sub CntSub(ByVal depth As Range, C As Range, lithology As Range, interval, NonZeroC As Range,
Optional ByVal C2 = Nothing, Optional ByVal C3 = Nothing, Optional ByVal C4 = Nothing,
Optional ByVal iC4 = Nothing, Optional ByVal nC4 = Nothing, Optional ByVal C5 = Nothing,
Optional ByVal iC5 = Nothing, Optional ByVal nC5 = Nothing)


Dim cnt As Double

Dim MA As Range


cnt = 0

For Each cell In depth

cnt = cnt + 1

Next


'--------------------------------

'       BACKGROUND SCENARIOS

'--------------------------------


'Call SMA(depth, C, interval, cnt)

'Call LinReg(depth, C, cnt)

'Call ConstantAverage(C, cnt)

'Call AvPerStrat(C, depth)

'Call DrawnBG

'Call BOXbg(C, depth)

'Call BOXbg2(C, depth)

'Call WindowBOXbg(depth, C, interval)


Call MovingWindowBOX(depth, C, interval)


'Call dif(C, cnt)

'---------------------------------


Set MA = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("F2")

Call filter(C, MA, cnt)                                                       'filtering
chromatograph with BG

Call preprocess2(cnt)                                                         'filtering by
lithology

Call CreateLogChart2(cnt)                                                     'chart


                '   Call showrange2(depth, C, lithology, MA, cnt, PrepC, C2, C3, C4, iC4, nC4,
C5, iC5, nC5)


End Sub
```

```
Function PlusMinEqual(cell1, cell2) As String

If cell2 - cell1 > 0 Then
PlusMinEqual = "+"
ElseIf cell2 - cell1 < 0 Then
PlusMinEqual = "-"
Else
PlusMinEqual = "NO"
End If

End Function
Function UpDownConc(one, two) As String
UpDownConc = one & "" & two
End Function
Sub MA(ByVal depth As Range, C As Range, interval, cnt)
'----------------------
moving Average
'----------------------
Dim step As Double
Dim TopGap As Double

step = depth(3) - depth(2) 'depthstep in meters
TopGap = interval / step   'determining the number of steps for the gap at the top of the well
due to the moving average interval

For i = 2 To cnt + 1
store1 = 0
'depth(i).Select
store2 = 0
If i >= TopGap Then
    For j = i - TopGap + 1 To i
        If C(j) <> "NO USE" Then
        store2 = store2 + C(j)
        End If
    Next
End If

If depth(i).Value < depth(1) + interval Then
    For j = 1 To TopGap
        If C(j) <> "NO USE" Then
        store1 = store1 + C(j)
```

```
        End If
    Next
    Cells(i, 5) = store1 / interval
ElseIf depth(i).Value >= depth(1) + interval Then
Cells(i, 5) = store2 / interval
Else
End If

Next

End Sub
Sub SMA(ByVal depth As Range, C As Range, interval, cnt)
'----------------------
'Simple moving average
'----------------------
Dim step As Double
Dim TopGap As Double

'Set depth = Application.InputBox("select steps", Type:=8)
'Set C = Application.InputBox("select gas", Type:=8)
'interval = Application.InputBox("interval", Type:=1)

step = depth(3) - depth(2) 'depthstep in meters
TopGap = interval / step   'determining the number of steps for the gap at the top of the well
due to the moving average interval

For i = 1 To cnt
store1 = 0
'depth(i).Select
store2 = 0
store3 = 0
If i >= TopGap / 2 Then
    For j = i - TopGap / 2 + 1 To i + TopGap / 2
        If C(j) <> "NO USE" Then
        store2 = store2 + C(j)
        End If
    Next
End If

If depth(i).Value < depth(1) + interval / 2 Then
    For j = 1 To TopGap / 2
```

```
        If C(j) <> "NO USE" Then
        store1 = store1 + C(j)
        End If
    Next j
    Cells(i + 1, 6) = store1 / (TopGap / 2)

ElseIf depth(i).Value >= depth(1) + interval / 2 And depth(i).Value <= depth(cnt) - interval /
2 Then
Cells(i + 1, 6) = store2 / TopGap

ElseIf i >= cnt - TopGap / 2 Then
    For k = cnt - TopGap / 2 To cnt
        If C(k) <> "NO USE" Then
        store3 = store3 + C(k)
        End If
    Next k
    Cells(i + 1, 6) = store3 / (TopGap / 2)
End If

Next i


End Sub
Sub LinReg(ByVal depth As Range, C As Range, cnt)

'Dim signal As Range
'Dim steps As Range

'Set steps = Application.InputBox("Select steps", Type:=8)
'Set signal = Application.InputBox("Select signal", Type:=8)


Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Range("A2:X1048576")
.ClearContents
Dim avsteps As Double
Dim avsignal As Double

'cnt = 0
'For Each Cell In steps
'cnt = cnt + 1
'Next
```

```
Sumsteps = 0

For i = 1 To cnt

Sumsteps = Sumsteps + depth(i)

Next i


avsteps = Sumsteps / cnt


sumsig = 0

For i = 1 To cnt

sumsig = sumsig + C(i)

Next


avsignal = sumsig / cnt


'Dim XminAvX As Object

'Dim YminAvY As Object

'Dim sqXminAvX As Object

'Dim XYmultip As Object


For i = 1 To cnt

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 1) =
depth(i)

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 2) =
C(i)


Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 3) =
depth(i) - avsteps    'x-X

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 4) =
C(i) - avsignal  'y-Y

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 5) =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 3) ^ 2
'sq(x-X)

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 6) =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 3) *
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 4) '(x-
X)(y-Y)

Next i


Application.ScreenUpdating = False

Dim SumsqXminAvX As Double

Dim SumXYmultip As Double


SumsqXminAvX = 0

SumXYmultip = 0
```

```
For i = 1 To cnt

SumsqXminAvX = SumsqXminAvX +
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 5)

SumXYmultip = SumXYmultip +
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 6)

Next


Dim b_1 As Double

Dim b_0 As Double


b_1 = SumXYmultip / SumsqXminAvX


b_0 = avsignal - b_1 * avsteps


For i = 1 To cnt

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 7) =
b_0 + b_1 * depth(i)

Next


For i = 1 To cnt

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(i + 1, 6) =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet3").Cells(i + 1, 7)

Next i

Application.ScreenUpdating = True


End Sub

Sub ConstantAverage(ByVal C As Range, cnt)


Dim sum As Double

Dim av As Double

Dim i As Integer

sum = 0

For i = 1 To cnt

sum = sum + C(i)

Next i


av = sum / cnt


For i = 2 To cnt + 2

Cells(i, 6) = av

Next i


End Sub
```

```
Sub AvPerStrat(ByVal C As Range, depth As Range)


Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet1").Activate 'in order
to work further in different sheet without getting error messages

Application.ScreenUpdating = True

Set ShowColRange = Application.InputBox("Select borehole-range in HCS sheet", Type:=8)

Application.ScreenUpdating = False


cnt = 0

For Each cell In depth

cnt = cnt + 1

Next


'counting number of show cells that are selected

cnt2 = 0

For Each cell In ShowColRange

cnt2 = cnt2 + 1

Next


'selecting each cell in pre-selected gas show column, as a reference for code (Active cell)

For i = 1 To cnt2

ShowColRange(i).Select

topstrat = ActiveCell.Offset(0, -7).Value

basestrat = ActiveCell.Offset(0, -6).Value


av = Gas4_2(topstrat, basestrat, cnt, depth, C)


    For j = 1 To cnt

    If depth(j).Value >= topstrat And depth(j).Value < basestrat Then

    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j + 1, 6)
= av

    End If

    Next j




Next i

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

End Sub

Sub DrawnBG()


Set X = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2",
Range("A2").End(xlDown))
```

```
For Each cell In X

cnt = cnt + 1

Next


Application.ScreenUpdating = True

Set KnownX = Application.InputBox("depth data", Type:=8)

Set KnownY = Application.InputBox("Drawn background data", Type:=8)

Application.ScreenUpdating = False


For Each cell In KnownX

cnt1 = cnt1 + 1

Next




For i = 1 To cnt1

    For j = 1 To cnt

    If X(j) >= KnownX(i) And X(j) <= KnownX(i + 1) Then

    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j + 1, 6)
= WorksheetFunction.Forecast_Linear(X(j), Range(KnownY(i), KnownY(i + 1)), Range(KnownX(i),
KnownX(i + 1)))

    End If

    Next j

Next i




End Sub

Sub BOXbg(ByVal C As Range, depth As Range)

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet1").Activate 'in order
to work further in different sheet without getting error messages

Application.ScreenUpdating = True

Set ShowColRange = Application.InputBox("Select borehole-range in HCS sheet", Type:=8)

Application.ScreenUpdating = False


cnt = 0

For Each cell In depth

cnt = cnt + 1

Next


'counting number of show cells that are selected

cnt2 = 0

For Each cell In ShowColRange
```

```
cnt2 = cnt2 + 1
Next

For i = 1 To cnt2
ShowColRange(i).Select
topstrat = ActiveCell.Offset(0, -7).Value
basestrat = ActiveCell.Offset(0, -6).Value

begin = 1
Do Until depth(begin).Value >= topstrat
begin = begin + 1
Loop

ending = begin
Do Until depth(ending).Value >= basestrat
If basestrat > depth(cnt) Then
Exit Do
Else
ending = ending + 1
End If
Loop



If ending - begin > 3 Then
q1 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 1)
q2 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 2)
q3 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 3)
End If

    counter = 0
    gascount = 0
    For k = 1 To cnt
    If ending - begin > 3 Then
        If depth(k).Value >= topstrat And depth(k).Value < basestrat Then
            If C(k) > q1 And C(k) < q2 Then
            counter = counter + 1
            gascount = gascount + C(k)
            End If
        End If
    End If
    Next k
```

```
    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If

    For j = 1 To cnt
    If ending - begin > 3 Then
        If depth(j).Value >= topstrat And depth(j).Value < basestrat Then
            If C(j) > q1 And C(j) < q2 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            ElseIf C(j) <= q1 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            ElseIf C(j) >= q2 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            End If
        End If
    Else
        If depth(j).Value >= topstrat And depth(j).Value < basestrat Then
        Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j + 1,
6) = C(j)
        End If
    End If

    Next j


Next i
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

End Sub
Sub BOXbg2(ByVal C As Range, depth As Range)
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet1").Activate 'in order
to work further in different sheet without getting error messages
Application.ScreenUpdating = True
Set ShowColRange = Application.InputBox("Select borehole-range in HCS sheet", Type:=8)
Application.ScreenUpdating = False

cnt = 0
```

```
For Each cell In depth
cnt = cnt + 1
Next

'counting number of show cells that are selected
cnt2 = 0
For Each cell In ShowColRange
cnt2 = cnt2 + 1
Next

For i = 1 To cnt2
ShowColRange(i).Select
topstrat = ActiveCell.Offset(0, -7).Value
basestrat = ActiveCell.Offset(0, -6).Value

begin = 1
Do Until depth(begin).Value >= topstrat
begin = begin + 1
Loop

ending = begin
Do Until depth(ending).Value >= basestrat
If basestrat > depth(cnt) Then
Exit Do
Else
ending = ending + 1
End If
Loop



If ending - begin > 3 Then
q1 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 1)
q2 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 2)
q3 = WorksheetFunction.Quartile_Exc(Range(C(begin), C(ending)), 3)
End If

    counter = 0
    gascount = 0
    For k = 1 To cnt
    If ending - begin > 3 Then
        If depth(k).Value >= topstrat And depth(k).Value < basestrat Then
```

```
            If C(k) > q1 And C(k) < q3 Then
            counter = counter + 1
            gascount = gascount + C(k)
            End If
        End If
    End If
    Next k

    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If

    For j = 1 To cnt
    If ending - begin > 3 Then
        If depth(j).Value >= topstrat And depth(j).Value < basestrat Then
            If C(j) > q1 And C(j) < q3 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            ElseIf C(j) <= q1 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            ElseIf C(j) >= q3 Then
            Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j
+ 1, 6) = av
            End If
        End If
    Else
        If depth(j).Value >= topstrat And depth(j).Value < basestrat Then
        Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(j + 1,
6) = C(j)
        End If
    End If

    Next j


Next i
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

End Sub
Sub WindowBOXbg(ByVal depth As Range, C As Range, interval)
```

```
For Each cell In depth

cnt = cnt + 1

Next


step = depth(3) - depth(2) 'depthstep in meters

TopGap = interval / step   'determining the number of steps for the gap at the top of the well
due to the moving average interval



Dim NumberOfWindows As Integer


h = 0

Do Until cnt2 >= cnt

h = h + 1

cnt2 = h * TopGap

Loop




NumberOfWindows = (h - 1) 'number of windows that fits in borehole



old_j = 1


For i = 1 To NumberOfWindows


j = i * TopGap


q1 = WorksheetFunction.Quartile_Exc(Range(C(old_j), C(j)), 1)

q2 = WorksheetFunction.Quartile_Exc(Range(C(old_j), C(j)), 2)

q3 = WorksheetFunction.Quartile_Exc(Range(C(old_j), C(j)), 3)


counter = 0

gascount = 0


    For k = old_j To j

        If C(k) > q1 And C(k) < q3 Then

        counter = counter + 1

        gascount = gascount + C(k)
```

```
        End If
    Next k

    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If

    For l = old_j To j
    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(l + 1, 6)
= av
    Next l



old_j = j + 1

Next i

cnt3 = NumberOfWindows * TopGap

remainder = cnt - cnt3

If remainder > 3 Then
q1 = WorksheetFunction.Quartile_Exc(Range(C(cnt3), C(cnt)), 1)
q2 = WorksheetFunction.Quartile_Exc(Range(C(cnt3), C(cnt)), 2)
q3 = WorksheetFunction.Quartile_Exc(Range(C(cnt3), C(cnt)), 3)

    For m = cnt3 To cnt
        If C(m) > q1 And C(m) < q3 Then
            counter = counter + 1
            gascount = gascount + C(m)
        End If
    Next m

    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If
```

```
    For n = cnt3 To cnt

    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(n + 1, 6)
= av

    Next n


End If





End Sub

Sub MovingWindowBOX(ByVal depth As Range, C As Range, interval)


For Each cell In depth

cnt = cnt + 1

Next


step = depth(3) - depth(2) 'depthstep in meters

TopGap = interval / step   'determining the number of steps for the gap at the top of the well
due to the moving average interval


check = WorksheetFunction.IsEven(TopGap)


If check = False Then

TopGap = TopGap + 1

End If



For i = 1 To cnt



If i > TopGap / 2 Then

    q1 = WorksheetFunction.Quartile_Exc(Range(C(i - TopGap / 2), C(i + TopGap / 2)), 1)

    q2 = WorksheetFunction.Quartile_Exc(Range(C(i - TopGap / 2), C(i + TopGap / 2)), 2)

    q3 = WorksheetFunction.Quartile_Exc(Range(C(i - TopGap / 2), C(i + TopGap / 2)), 3)


    counter = 0

    gascount = 0

    For j = i - TopGap / 2 To i + TopGap / 2

        If C(j) > q1 Or C(j) < q3 Then

        counter = counter + 1

        gascount = gascount + C(j)

        End If
```

```
    Next j

    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If

    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(i + 1, 6)
= av

End If

If i < TopGap / 2 Then
    q1 = WorksheetFunction.Quartile_Exc(Range(C(1), C(TopGap / 2)), 1)
    q2 = WorksheetFunction.Quartile_Exc(Range(C(1), C(TopGap / 2)), 2)
    q3 = WorksheetFunction.Quartile_Exc(Range(C(1), C(TopGap / 2)), 3)

    counter = 0
    gascount = 0
    For k = 1 To TopGap / 2
        If C(k) > q1 Or C(k) < q3 Then
        counter = counter + 1
        gascount = gascount + C(k)
        End If
    Next k

    If gascount = 0 Or counter = 0 Then
    av = q2
    Else
    av = gascount / counter
    End If

    Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(i + 1, 6)
= av


ElseIf i > cnt - TopGap / 2 Then
    q1 = WorksheetFunction.Quartile_Exc(Range(C(cnt - TopGap / 2), C(cnt)), 1)
    q2 = WorksheetFunction.Quartile_Exc(Range(C(cnt - TopGap / 2), C(cnt)), 2)
    q3 = WorksheetFunction.Quartile_Exc(Range(C(cnt - TopGap / 2), C(cnt)), 3)
```

```
    counter = 0

    gascount = 0

    For l = cnt - TopGap / 2 To cnt

        If C(l) > q1 Or C(l) < q3 Then

        counter = counter + 1

        gascount = gascount + C(l)

        End If

    Next l


        If gascount = 0 Or counter = 0 Then

        av = q2

        Else

        av = gascount / counter

        End If

        Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(i + 1,
6) = av

    End If


Next i



End Sub

Sub filter(ByVal C As Range, MA As Range, cnt)

'--------------------------------------------

'filtering chromatograph with moving average

'--------------------------------------------

For i = 2 To cnt + 1

    If Cells(i, 6) > Cells(i, 2) Then

    Cells(i, 5) = 1

    Else

    Cells(i, 5) = Cells(i, 2)

    End If

Next

End Sub

Sub CreateLogChart2(ByVal cnt)


Dim xrng As Range

Dim yrng As Range

Dim MArange As Range

Dim FiltSing As Range

Dim chartrange As Range
```

```
'Dim cht As Object

Dim cht As Chart


'----------------------------------------------

'Set Newbook = Workbooks.Add

'With Newbook

'.SaveAs Filename:="temporary_analysis_file.xslx"

'End With

'----------------------------------------------


'data

Set xrng =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range(Cells(2, 3),
Cells(cnt + 1, 3)) 'Non-zero C1-range

Set yrng =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range(Cells(2, 1),
Cells(cnt + 1, 1)) 'depth

Set MArange =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range(Cells(2, 6),
Cells(cnt + 1, 6)) 'moving average

Set FiltSing =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range(Cells(2, 9),
Cells(cnt + 1, 9)) 'filtered signal


Set chartrange =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range(Cells(1, 19),
Cells(1000, 26))


'create chart

Set cht = ActiveSheet.Shapes.AddChart.Chart


With cht

.ChartType = xlXYScatterLinesNoMarkers


'do this to NOT get an extra empty series in chart

    Do While .SeriesCollection.count > 0

    .SeriesCollection(1).Delete

    Loop


'new series for chart

.SeriesCollection.NewSeries

    With .SeriesCollection(1)

    .XValues = xrng

    .Values = yrng

    End With
```

```
.SeriesCollection(1).Name = "C1"

.SeriesCollection.NewSeries
    With .SeriesCollection(2)
    .XValues = MArange
    .Values = yrng
    End With
.SeriesCollection(2).Name = "Simple Moving Average"

.SeriesCollection.NewSeries
    With .SeriesCollection(3)
    .XValues = FiltSing
    .Values = yrng
    End With
.SeriesCollection(3).Name = "C1 filtered Lith & peaks"



'Format chart
.HasLegend = True
.SetElement (msoElementLegendTop)
    '.Legend.Top = 20
.HasTitle = True
.ChartTitle.Text = "log"
.ChartTitle.Format.TextFrame2.TextRange.Font.Size = 12
.ChartTitle.Format.TextFrame2.TextRange.Font.Bold = True



'.Parent.Height = xrng.Height + 100
'.Parent.Top = xrng.Top - 100
    '.PlotArea.Top = xrng.Top
    '.Parent.Top = Range("S2").Top
'.Parent.Left = Range("S2").Left

'---------------------------------
.Parent.Top = chartrange.Top
.Parent.Left = chartrange.Left
.Parent.Height = chartrange.Height
.Parent.Width = chartrange.Width
'---------------------------------

'x-axes
```

```
.Axes(xlCategory).ScaleType = xlScaleLogarithmic
.Axes(xlCategory).HasMajorGridlines = True
.Axes(xlCategory).HasMinorGridlines = True
.Axes(xlCategory).MaximumScale = 100000
.Axes(xlCategory).MinimumScale = 1
.Axes(xlCategory).HasTitle = True
.Axes(xlCategory, xlPrimary).AxisTitle.Text = "[ppm]"
    '.Axis(xlCategory).AxisTitle.Top = 2

'y-axes

.Axes(xlValue).ReversePlotOrder = True
.Axes(xlValue).HasMajorGridlines = True
.Axes(xlValue).MajorUnit = 100
'.Axes(xlValue).HasMinorGridlines = True
'.Axes(xlValue).MinorUnit = 1
.Axes(xlValue).HasTitle = True
.Axes(xlValue, xlPrimary).AxisTitle.Text = "Depth [m]"
.Axes(xlValue).MinimumScale = 0
End With

End Sub
Sub Classification_Button()
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Activate

Dim depth As Range
Dim C As Range
Dim NonZeroC As Range
Dim lithology As Range

Dim interval As Double

Dim C2 As Range
Dim C3 As Range
Dim C4 As Range
Dim iC4 As Range
Dim nC4 As Range
Dim C5 As Range
Dim iC5 As Range
Dim nC5 As Range
```

```
Dim PrepC As Range

Dim BG As Range


Set BG = Application.InputBox("Select Background", Type:=8)


Dim cnt As Integer

'-----------------------------------

'resetting ranges with same input data, but now in other file

'-----------------------------------

'hier misschien een do with loop?


Set depth =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("A2").End(xlDo
wn))

Set C = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("B2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("B2").End(xlDo
wn))

Set NonZeroC =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("C2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("C2").End(xlDo
wn))

Set lithology =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("H2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("H2").End(xlDo
wn))

interval = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Cells(2,
18).Value


Set C2 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("J2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("J2").End(xlDo
wn))

Set C3 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("K2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("K2").End(xlDo
wn))

Set C4 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("L2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("L2").End(xlDo
wn))

Set iC4 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("M2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("M2").End(xlDo
wn))

Set nC4 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("N2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("N2").End(xlDo
wn))

Set C5 = Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("O2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("O2").End(xlDo
wn))

Set iC5 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("P2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("P2").End(xlDo
wn))

Set nC5 =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("Q2",
```

```
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("Q2").End(xlDown))

Set PrepC =
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("I2",
Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Range("I2").End(xlDown))


cnt = 0

For Each cell In depth

cnt = cnt + 1

Next


Call showrange2(depth, C, lithology, BG, cnt, PrepC, C2, C3, C4, iC4, nC4, C5, iC5, nC5)




End Sub

Sub showrange2(ByVal depth As Range, C1 As Range, lithology As Range, MA As Range, cnt, PrepC
As Range, Optional ByVal C2 = Nothing, Optional ByVal C3 = Nothing, Optional ByVal C4 =
Nothing, Optional ByVal iC4 = Nothing, Optional ByVal nC4 = Nothing, Optional ByVal C5 =
Nothing, Optional ByVal iC5 = Nothing, Optional ByVal nC5 = Nothing)


Dim ShowColRange As Range


'counters

Dim i As Integer

Dim cnt2 As Integer


Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet1").Activate 'in order
to work further in different sheet without getting error messages

Application.ScreenUpdating = True

Set ShowColRange = Application.InputBox("Select borehole-range in HCS sheet", Type:=8)

Application.ScreenUpdating = False

'counting number of show cells that are selected

cnt2 = 0

For Each cell In ShowColRange

    cnt2 = cnt2 + 1

Next


'selecting each cell in pre-selected gas show column, as a reference for code (Active cell)

For i = 1 To cnt2

ShowColRange(i).Select


'voorwaarde hier, of in volgende sub?

Call GasAuto(depth, PrepC, lithology, MA, cnt, C1, C2, C3, C4, iC4, nC4, C5, iC5, nC5) 'PrepC
was C
```

```
Next


MsgBox ("Classification has been completed")

Application.ScreenUpdating = True

End Sub

Sub GasAuto(ByVal depth As Range, PrepC As Range, lithology As Range, MA As Range, cnt,
Optional ByVal C1 As Range = Nothing, Optional ByVal C2 As Range = Nothing, Optional ByVal C3
As Range = Nothing, Optional ByVal C4 As Range = Nothing, Optional ByVal iC4 As Range =
Nothing, Optional ByVal nC4 As Range = Nothing, Optional ByVal C5 As Range = Nothing, Optional
ByVal iC5 As Range = Nothing, Optional ByVal nC5 As Range = Nothing)


'counter, top of stratigraphy, base of stratigraphy, name of stratigraphy


Dim topstrat As Double

Dim basestrat As Double

Dim strat As String

Dim peaklocation As Integer


Dim check As Variant


'(l)ithology, (p)eak & (b)ackground at a certain depth

Dim l As String

Dim p As Double

Dim b As Double


'gas-values at depth

Dim d_c1 As Double

Dim d_c2 As Double

Dim d_c3 As Double

Dim d_c4 As Double

Dim d_ic4 As Double

Dim d_nc4 As Double

Dim d_c5 As Double

Dim d_ic5 As Double

Dim d_nc5 As Double


'--------------------------------------------------------------------


'cnt = Gas1(depth) dit werkt gek genoeg niet



'MsgBox cnt, , "cnt terug van Gas1="
```

```
strat = ActiveCell.Offset(0, -8).Value

topstrat = ActiveCell.Offset(0, -7).Value

basestrat = ActiveCell.Offset(0, -6).Value


check = Gas1_2(topstrat, basestrat, cnt, depth, C1) 'check difference between min and max in strat

check2 = Gas1_3(topstrat, basestrat, cnt, depth, MA, PrepC)

peaklocation = Gas2_1(topstrat, basestrat, cnt, depth, PrepC)

'MsgBox (check)


If check = "no show" Or check2 = 0 Then

ActiveCell.Value = "NO SHOW"

ActiveCell.Offset(0, 1) = 0

ActiveCell.Offset(0, 2) = Gas4_2(topstrat, basestrat, cnt, depth, C1) 'background

ActiveCell.Offset(0, 3) = 0

ActiveCell.Offset(0, 4) = 0

ActiveCell.Offset(0, 5) = 0

ActiveCell.Offset(0, 6) = 0

ActiveCell.Offset(0, 7) = 0

ActiveCell.Offset(0, 8) = 0

ActiveCell.Offset(0, 9) = 0

ActiveCell.Offset(0, 10) = 0

ActiveCell.Offset(0, 11) = 0


Else


ActiveCell.Offset(0, 3) = Gas2(topstrat, basestrat, cnt, depth, PrepC)

ActiveCell.Offset(0, -1) = Gas3_1(depth, peaklocation) 'Gas3(topstrat, basestrat, cnt, depth, C1) first adjustment in code

ActiveCell.Offset(0, 2) = Gas4_3(MA, peaklocation)  'Gas4_2(topstrat, basestrat, cnt, depth, C1)

ActiveCell.Offset(0, 14) = Gas5_1(lithology, peaklocation) 'Gas5(topstrat, basestrat, cnt, C1, depth, lithology)


l = ActiveCell.Offset(0, 14).Value

p = ActiveCell.Offset(0, 3).Value

b = ActiveCell.Offset(0, 2).Value


ActiveCell.Offset(0, 4) = Gas9(C2, peaklocation)

ActiveCell.Offset(0, 5) = Gas9(C3, peaklocation)

ActiveCell.Offset(0, 6) = Gas9(C4, peaklocation)

ActiveCell.Offset(0, 7) = Gas9(iC4, peaklocation)
```

```
ActiveCell.Offset(0, 8) = Gas9(nC4, peaklocation)
ActiveCell.Offset(0, 9) = Gas9(C5, peaklocation)
ActiveCell.Offset(0, 10) = Gas9(iC5, peaklocation)
ActiveCell.Offset(0, 11) = Gas9(nC5, peaklocation)


d_c1 = ActiveCell.Offset(0, 3).Value
d_c2 = ActiveCell.Offset(0, 4).Value
d_c3 = ActiveCell.Offset(0, 5).Value
d_c4 = ActiveCell.Offset(0, 6).Value
d_ic4 = ActiveCell.Offset(0, 7).Value
d_nc4 = ActiveCell.Offset(0, 8).Value
d_c5 = ActiveCell.Offset(0, 9).Value
d_ic5 = ActiveCell.Offset(0, 10).Value
d_nc5 = ActiveCell.Offset(0, 11).Value

'SHOW
ActiveCell.Value = Gas6(strat, l, p, b)
ActiveCell.Offset(0, 18) = Gas7(p, b)

'total gas
ActiveCell.Offset(0, 1) = Gas8(d_c1, d_c2, d_c3, d_c4, d_ic4, d_nc4, d_c5, d_ic5, d_nc5)
End If

End Sub
Function Gas1(ByVal depth As Range) 'to be fixed, doesn't count the range

'Application.ScreenUpdating = False
Dim cnt As Integer
cnt = 0
For Each cell In depth
cnt = cnt + 1
Next
Gas1 = cnt
'MsgBox Gas1, , "Gas1 (cnt) ="
'Application.ScreenUpdating = True
End Function
Function Gas1_2(topstrat, basestrat, countAcheck, depth As Range, gas As Range)

Dim i As Integer
```

```
Dim maxwaarde As Double
Dim minwaarde As Double
Dim max As Double
Dim min As Double

'MsgBox countAcheck, , "countAcheck ="
'MsgBox topstrat, , "top of stratigraphy"
'MsgBox basestrat, , "base of stratigraphy"
'MsgBox depth(1), , "depth column starts with"
'MsgBox gas(1), , "gascolumn starts with"

'looping over cells in stratigraphy to find maximum value
maxwaarde = 0
For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) > maxwaarde Then
            maxwaarde = gas(i)
            End If
        End If
    End If
    max = maxwaarde
Next i
'MsgBox max, , "maxwaarde="

'searching min-value
minwaarde = max
For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) < minwaarde Then
            minwaarde = gas(i).Value
            End If
        End If
    End If
    min = minwaarde
Next i

'MsgBox minwaarde

'checking whether gas-value within certain range are suitable to give shows
```

```
'MsgBox max - min, , "max-min = "


If max - min <= 500 Then

Gas1_2 = "no show"

End If


'------------------------------------

'a statement that checks the shape of the chromatograph within a statigraphy should be
incorporated here as well

'------------------------------------


End Function

Function Gas1_3(topstrat, basestrat, countAcheck, depth As Range, BG As Range, gas As Range)


'does gas-measurement rise above set bg level?

Dim i As Integer


Dim maxwaarde As Double



maxwaarde = 0

For i = 2 To countAcheck

    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then

        If gas(i).Value <> "NO USE" Then

            If gas(i).Value > BG(i) Then

            maxwaarde = 1

            End If

        End If

    End If

Next i

Gas1_3 = maxwaarde


End Function

Function Gas2(topstrat, basestrat, countAcheck, depth As Range, gas As Range)

Application.ScreenUpdating = False


'MsgBox ("gas 2")


Dim maxwaarde As Double
```

```
'looping over cells in stratigraphy to find maximum value
maxwaarde = 0
For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) > maxwaarde Then
            maxwaarde = gas(i).Value
            End If
        End If
    End If
    Gas2 = maxwaarde
Next i

Application.ScreenUpdating = True
End Function
Function Gas2_1(topstrat, basestrat, countAcheck, depth As Range, gas As Range)
Application.ScreenUpdating = False

Dim maxwaarde As Double
Dim peaklocation As Integer

'looping over cells in stratigraphy to find LOCATION maximum value
maxwaarde = 0
For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) > maxwaarde Then
            maxwaarde = gas(i).Value
            peaklocation = i
            End If
        End If
    End If
    Gas2_1 = peaklocation
Next i

Application.ScreenUpdating = True
End Function
Function Gas3(topstrat, basestrat, countAcheck, depth As Range, gas As Range)
Application.ScreenUpdating = False

'MsgBox ("gas 3")
```

```
Dim maxwaarde As Double
Dim gasdepth As Double

For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) > maxwaarde Then
            maxwaarde = gas(i).Value
            gasdepth = depth(i).Value
            End If
        End If
    End If
    Gas3 = gasdepth
Next i

Application.ScreenUpdating = True
End Function
Function Gas3_1(depth As Range, PeakLoc)
Application.ScreenUpdating = False

Gas3_1 = depth(PeakLoc)

Application.ScreenUpdating = True
End Function
Function Gas4_2(topstrat, basestrat, countAcheck, depth As Range, gas As Range) As Double
Application.ScreenUpdating = False

'MsgBox ("gas 4_2, temporary background")

Dim i As Integer
Dim count As Double
Dim gascount As Double
Dim counter As Double
Dim gastot As Double

count = 1
gascount = 0

For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
```

```
    count = count - 1 + 2 'to not divide by 0
        If gas(i).Value <> "NO USE" Then
        gascount = gascount + gas(i).Value
        End If
    End If
    counter = count
    gastot = gascount
Next i

Gas4_2 = gastot / counter

Application.ScreenUpdating = True
End Function
Function Gas4_3(MA As Range, PeakLoc)
Gas4_3 = MA(PeakLoc)
End Function
Function Gas4_4(topstrat, basestrat, countAcheck, depth As Range, gas As Range)

i = 1
Do Until depth(i).Value >= topstrat
i = i + 1
Loop
begin = i

j = i
Do Until depth(i).Value >= basestrat
j = j + 1
Loop
ending = j





Next i
End Function
Function Gas5(topstrat, basestrat, countAcheck, gas As Range, depth As Range, lith As Range)
Application.ScreenUpdating = False

'MsgBox ("gas 5")
```

```
Dim maxwaarde As Double
Dim lithloop As String
Dim lithology As String

'looping over cells in stratigraphy to find lithology
maxwaarde = 0

For i = 2 To countAcheck
    If depth(i).Value >= topstrat And depth(i).Value < basestrat Then
        If gas(i).Value <> "NO USE" Then
            If (gas(i).Value) > maxwaarde Then
            maxwaarde = gas(i).Value
            lithloop = lith(i)
            End If
        End If
    End If
    'hands lithology at peak
    Gas5 = lithloop
Next i

Application.ScreenUpdating = True
End Function
Function Gas5_1(lithology As Range, PeakLoc)
Gas5_1 = lithology(PeakLoc)
End Function
Function Gas6(strat, lithology, peak, background)
Application.ScreenUpdating = False

Dim PtBR As Double

PtBR = (peak / background)

'gas classification
If lithology = "coal" Or lithology = "anhydrite" Or lithology = "halite" Or lithology = "rock salt" Or lithology = "mudstone" Or lithology = "claystone" Or lithology = "shale" Or lithology = "marl" Or lithology = "siltstone" Or lithology = "sandstone" Or lithology = "limestone" Or lithology = "chalk" Or lithology = "dolomite" Then
    If peak < 500 Or PtBR < 2 Or lithology = "coal" Or lithology = "anhydrite" Then
    Gas6 = "NO SHOW"
    ElseIf PtBR >= 2 And PtBR < 3 Then
    Gas6 = "POOR"
    ElseIf PtBR >= 3 Then
```

```
        If lithology = "sandstone" And background >= 10000 Then 'Very BG exception

        Gas6 = "GOOD"

        ElseIf strat = "NU" Then 'North Sea Group (NU) exception

            If lithology = "mudstone" Or lithology = "shale" Or lithology = "marl" Then

            Gas6 = "POOR"

            ElseIf lithology = "siltstone" Then

            Gas6 = "FAIR"

            ElseIf lithology = "claystone" Or lithology = "sandstone" Or lithology =
"limestone" Or lithology = "dolomite" Or lithology = "chalk" Then

                If peak < 1000 Or PtBR < 5 Then

                Gas6 = "FAIR"

                Else

                Gas6 = "GOOD"

                End If

            End If

        Else

            If lithology = "claystone" Or lithology = "mudstone" Or lithology = "shale" Or
lithology = "marl" Then

            Gas6 = "POOR"

            ElseIf lithology = "siltstone" Then

            Gas6 = "FAIR"

            ElseIf lithology = "sandstone" Or lithology = "limestone" Or lithology = "chalk"
Or lithology = "dolomite" Then

                If peak < 1000 Or PtBR < 5 Then

                Gas6 = "FAIR"

                Else

                Gas6 = "GOOD"

                End If

            End If

        End If

    End If

Else

Gas6 = "combination not yet in classification"

End If


Application.ScreenUpdating = True

End Function

Function Gas7(peak, background) As String


Dim PtBR As Double


PtBR = peak / background
```

```
If peak = Empty Or PtBR = Empty Then
Gas7 = "1 or 2 parameters missing"
ElseIf PtBR < 2 Or peak < 500 Then
Gas7 = "NO SHOW"
Else
    If PtBR >= 2 And PtBR < 3 Then
    Gas7 = "POOR"
    ElseIf PtBR >= 3 And PtBR < 5 Then
    Gas7 = "FAIR"
    ElseIf PtBR >= 5 Then
        If peak < 1000 Then
        Gas7 = "FAIR"
        Else
        Gas7 = "GOOD"
        End If
    End If
End If

End Function
Function Gas8(C1, C2, C3, C4, iC4, nC4, C5, iC5, nC5)
'calculating total gas

Gas8 = C1 + C2 + C3 + C4 + iC4 + nC4 + C5 + iC5 + nC5

End Function
Function Gas9(gas As Range, PeakLoc)
Gas9 = gas(PeakLoc)
End Function
Sub SaveSheet()

Dim question As String
Dim answer As String

question = "Save Sheet2 as file?"

answer = MsgBox(question, vbQuestion + vbYesNo, "Save")

If answer = vbYes Then
'opens new workbook
Set Newbook = Workbooks.Add
```

```
'copies input data sheet and pastes it to the new book

Workbooks("Automated_HCS_sheet_MovingWindowBOX501.xlsm").Sheets("Sheet2").Copy
before:=Newbook.Sheets("Sheet1")

Else

End If


End Sub

Sub preprocess1(ByVal lithology As Range, cnt)


For i = 2 To cnt + 1

If lithology(i) = "coal" Or lithology(i) = "lignite" Or lithology(i) = "bitumen" Then

Cells(i + 1, 8) = 1

Else

Cells(i + 1, 8) = 0

End If

Next


End Sub

Sub preprocess2(ByVal cnt)  'filtering by lithology


For i = 2 To cnt + 1


If Cells(i, 8) = "coal" Or Cells(i, 8) = "lignite" Or Cells(i, 8) = "bitumen" Or Cells(i, 8) =
"salt" Or Cells(i, 8) = "rock salt" Or Cells(i, 8) = "anhydrite" Or Cells(i, 8) = "halite"
Then

Cells(i, 9) = 1

Else

Cells(i, 9) = Cells(i, 5) 'which is already filtered with BG

End If

Next


End Sub
```