P-04-26

Oskarshamn site investigation

Calculation of Fracture Zone Index (FZI) for HSH01

Lennart Lindqvist, Bergsten & Co i Värnamo AB

Hans Thunehed, GeoVista AB

January 2004

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co
Box 5864
SE-102 40 Stockholm Sweden
Tel 08-459 84 00
+46 8 459 84 00
Fax 08-661 57 19
+46 8 661 57 19

ISSN 1651-4416
SKB P-04-26

Oskarshamn site investigation

Calculation of Fracture Zone Index (FZI) for HSH01

Lennart Lindqvist, Bergsten & Co i Värnamo AB

Hans Thunehed, GeoVista AB

January 2004

Keywords: Fracture zone index, FZI, Fracture frequency, Borehole geophysics, Logging.

This report concerns a study which was conducted for SKB. The conclusions and viewpoints presented in the report are those of the authors and do not necessarily coincide with those of the client.

A pdf version of this document can be downloaded from www.skb.se

Abstract

The aim of this work was to carry out a multivariate calculation of a Fracture Zone Index, FZI, along the percussion borehole HSH01 in the Simpevarp area. This will generalize and integrate information from geophysical logs, geological mapping and manual classification to a numerical description of the fracture properties of the rock.

The available data have been joined into a matrix with common and uniform section lengths through averaging, interpolation, resampling and manual coding in order to create comparable sections along the borehole.

A manual classification (GFZI) of the borehole in three types of classes was performed in order to define the properties that FZI is supposed to describe. These types are core of fracture zone (GFZI=2), transition zone (GFZI=1) and normal unaffected rock (GFZI=0).

Relations between objects and variables were analyzed with Principal Component Analysis (PCA) and outliers were identified and excluded from the modelling of the data set.

A regression model that describes the relation between the significant input variables and the manual classification, GFZI, was established with Projection to Latent Structures (PLS). FZI was then calculated for all sections along the borehole based on the PLS-model.

Contents

1 Introduction

Multivariate statistics in the form of Principal Component Analysis, PCA, and Projection to Latent Structures, PLS, is well documented analysis techniques. The methods for multivariate analysis that were developed during the 1970's are described in detail in /1/.

Multivariate analysis has become a popular tool within several sciences, including geoscience where it has been used for a long time for e.g. analysis and evaluation of chemical and petrophysical variables in exploration /2, 3, 4/.

Large amounts of data have been created during the last years in investigation sites managed by SKB. These data are well suited for multivariate analysis and applications are described by /5, 6, 7/. The methods are also used in some countries where assessments of rock volumes for localisation of a repository for spent nuclear waste is ongoing, e.g. /8/.

Recently a calculation of Fracture Zone Index was performed for the cored drillhole KSH01A at the Oskarshamn investigation site, /9/.

The calculation of FZI was initially described in /7/.

Analysis was performed by Bergsten & Co in Värnamo AB and GeoVista AB in accordance with the instructions and guidelines from SKB (activity plan AP PS 400-03-094 and method description MD 810.003, SKB internal controlling documents) and under supervision of Leif Stenberg, SKB.

2 Objective and scope

The aim of this work was to carry out a multivariate calculation of a Fracture Zone Index, FZI, along the borehole HSH01 in the Simpevarp area. This will generalize and integrate information from geophysical logs, geological mapping and manual classification to a numerical description of the fracture properties of the rock in a robust and objective way.

The calculation of FZI with multivariate techniques is based on measured and observed quantities along the borehole. The PLS-model used to calculate FZI can be used on data from other boreholes provided that they are from a similar geological environment and have the same set of variables.

The most important prerequisite for this analysis is the definition of what FZI is supposed to describe. This will be the quantity that the measured and observed data will try to model and predict. The borehole is therefore divided into sections of three discrete intensities of fracturing (GFZI) based on manual classification.

3 Equipment

Multivariate statistical calculations have been performed with Simca-P version 8.0 (Umetrics AB). Grapher v 4.0 (Golden Software) has been used for presentation of the final results.

4 Execution and results

4.1 Pre processing of data

The pre processing of input data is summarized in Table 4-1.

The HSH01 were percussion drilled for 198 m. The drillhole has been logged to 194 m.

A common section length of one metre was chosen for all variables in this work. This choice was partly based on previous experiences but also on the fact that e.g. the core-mapped fracture frequency is available in one metre sections. This choice is not critical and appears sound since significant fracture zones often have a width of several metres.

Longer sections of e.g. 5 to10 metres would probably create mixing of different zone classes and the borders between zones would be blurred.

Shorter section lengths would not create any technical problems and would even be advantageous for some variables like e.g. the sonic log which show short wave-length anomalies for fractures.

The geophysical logs (magnetic susceptibility, natural γ-radiation, caliper, 16" normal resistivity, fluid resistivity, single point resistance) were all initially measured at 0.1 metres intervals. Average values for one metre sections were calculated. The electrical logs and the caliper log were deconvolved to give weighted discrete source indications. These were summed in one metre sections.

Missing values are indicated by the number –999 in the data file.

Table 4-1. Pre processing of data for calculation of FZI.

Processed primary data	Pre processing	Resulting data file
Mapped fracture frequency from SICADA	Geophysical logs: Averaging in one metre sections, resampling	HSH01_alla_var_01.xls (MS Excel)
Geophysical logs from SICADA	Electrical and caliper logs: Deconvolution and summation	
Penetration rate from SICADA	Subtraction of trend from caliper data	
	Creation of common data matrix	

4.2 Variables for analysis

A total of 11 input variables were available in the data matrix resulting from the pre processing. Additionally there was a column describing the manual classification of fracture intensity (GFZI), a column for section identity and a column for length along the borehole.

The GFZI variable was slightly modified so that 5 sections on either side of a fracture zone were given the value 0.1 to indicate the proximity to the zone. These sections are called the near zone.

The list below shows the acronyms that have been used for the various variables in the text and in figures in the rest of this report.

Acronym

Id = Identity consisting of borehole number plus depth in metres
e.g. 1013 means hole #1 (HSH01) and depth 13–14 metres

Le = Length along hole (m), e.g. 13.5 for the above section

GFZI = Geological Fracture Zone Index
= 2 core of fracture zone
= 1 transition zone
= 0.1 near zone
= 0 normal rock

Cpd = Calip decon

N1d = Normres16 decon

Srd = Single point resistivity decon

Cp = Calip_resid

Fr = Fluid resistivity

L_N1 = Logarithm of Norm 16

Ng = Natural Gamma

L_Sr = Logarithm of Spr

Ms = Magnetic susceptibility

Ff = Fracture frequency

Pr = Penetration rate

Output variables

NGFZI = Numerical Geological Fracture Zone Index calculated with PLS-technique

FZI = Continuous Fracture Zone Index based on NGFZI

FZI can also be assigned discrete values according to:

= 2	core of fracture zone,	NGFZI > 1.5
= 1	transition zone,	0.5 < NGFZI < 1.5
= 0	unaffected rock,	NGFZI < 0.5

4.3 Incomplete sections and outliers

For some sections there are missing variables. These sections are indicated to the analysis by setting to –999.

For detection of outliers the multivariate PCA analysis was used. A total of 12 sections were removed with Id 1018, 1020, 1041, 1045, 1046, 1047, 1162, 1163, 1164, 1165, 1166, and 1167. These sections seem to be a part of a trend but with a significant difference from its neighbours – hence they are excluded to leave a more homogenous data set.

The results from the PCA-analysis from the final model M4, gives a reliable and robust view of the data matrix where the influence from the outliers have been removed.

For the PLS analysis no extra outliers were identified.

The modelling with PLS is also performed with cross validation, where parts of the data matrix are alternately removed to test the significance and repeatability of the model and its components.

4.4 Definition of FZI

Initially a manual classification of information from the borehole called GFZI was performed. This classification describes the intensity of fracturing in discrete levels.

The core of the fracture zone has been given the value GFZI=2 whereas the transition zone to normal rock has been given the value GFZI=1. The vicinity of the transition zone have been complemented with a near zone where GFZI=0.1. Rock out side these zones, normal unaffected rock, have been given the value GFZI=0.

The approach is that of looking for a model that describes the rock in terms of fracturing only and where other properties like e.g. lithology do not interfere. The model might consist of several components with the common properties that they correlate with GFZI and that they contribute to a more robust description of FZI. Those properties of the rock mass that are reflected in the variables but not correlate to GFZI are automatically eliminated.

4.5 General relations between variables

The general relations between data have been analyzed with PCA. No attempt has been made at this stage to make any kind of prediction of FZI. A seek of an understanding of the relations within the data set was performed.

The correlations between variables are visualised in a number of variable loading plots below. Variables that plot close to each other show correlation in these plots whereas variables on the opposite side of the center of the plot show reverse correlation. Variables in the distal parts of the plot show strong correlation whereas variables close to the origin shows weak correlation with the other variables. The horizontal and vertical direction vectors are orthogonal and by definition uncorrelated to each other.

4.5.1 Principal Component Analysis

Principal Component transformation was performed using all variables and objects in the data matrix. The normalized distance to the centre of the first model is shown in Figure 4-1 and the PCA-analysis variable loadings in Figure 4-2 and object scores in Figure 4-3. Some outliers are evident in the plot.

Sections with DmodX(PS),N >3.0, i.e. those that have a normalized distance to the centre of the model greater than 3 standard deviations, were omitted from further analysis by two step PCA-analysis. The outliers were identified as Id 1018, 1020, 1041, 1045, 1046, 1047, 1162, 1163, 1164, 1165, 1166, 1167 and 169 sections remained for the analysis. The outliers contain anomalous values for the geophysical logs. The removal thus means that only geophysical data with background and moderately anomalous values are included in the analysis.

PCA for model M3

After excluding the 12 outliers, a PCA-model was calculated for the remaining data set called model M3. The normalized distance from the model centre for each section is plotted in Figure 4-4.

The PC-analysis of the data matrix revealed two significant components that describe 55.5% of the total variation in the data. The first component (first component – horizontal) shows the properties related to the depth and the penetration rate versus Fr, L_Sr, L_N1. The second component (second component – vertical) shows the properties related to fracturing and lithology (Figure 4-5 and 4-6).

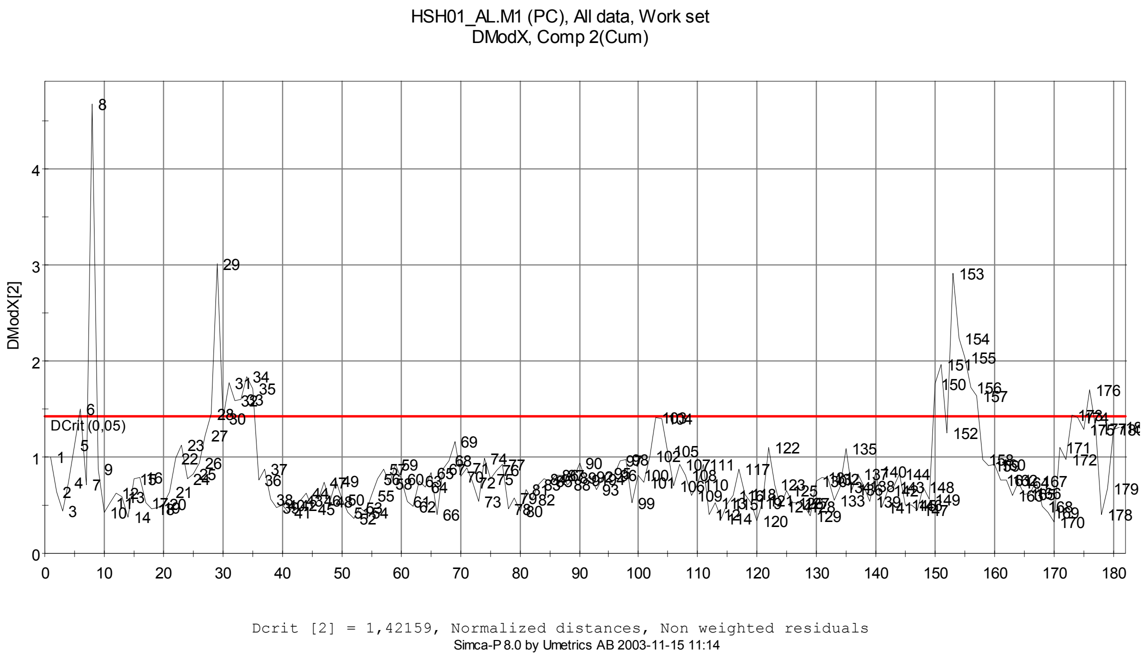

Figure 4-1. *Normalized distance (in standard deviations) of 1-m sections in the data matrix, to the centre of the initial PCA-model.*

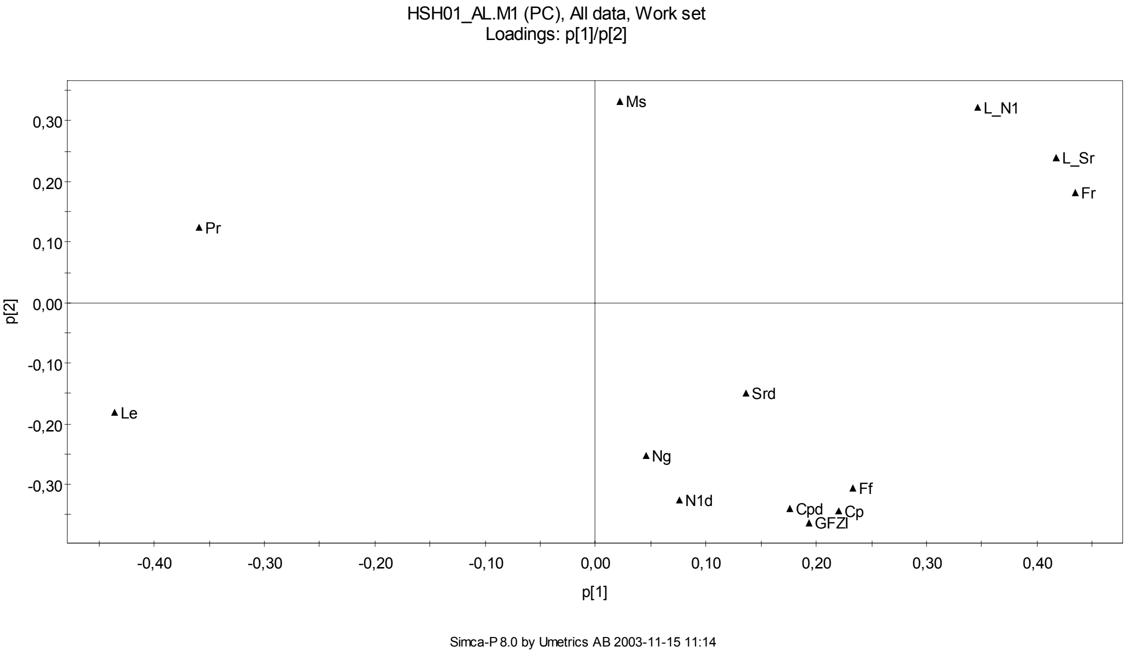

Figure 4-2. *PCA variable loading for all data for the first model.*

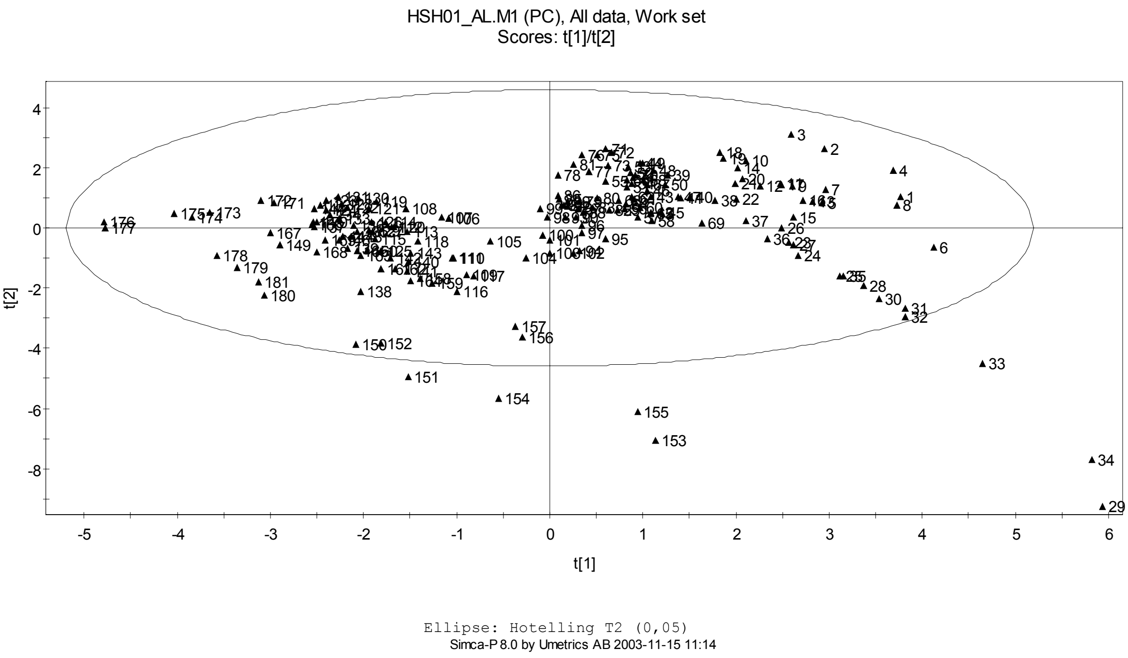

Figure 4-3. *PCA object score for all data for the first model.*

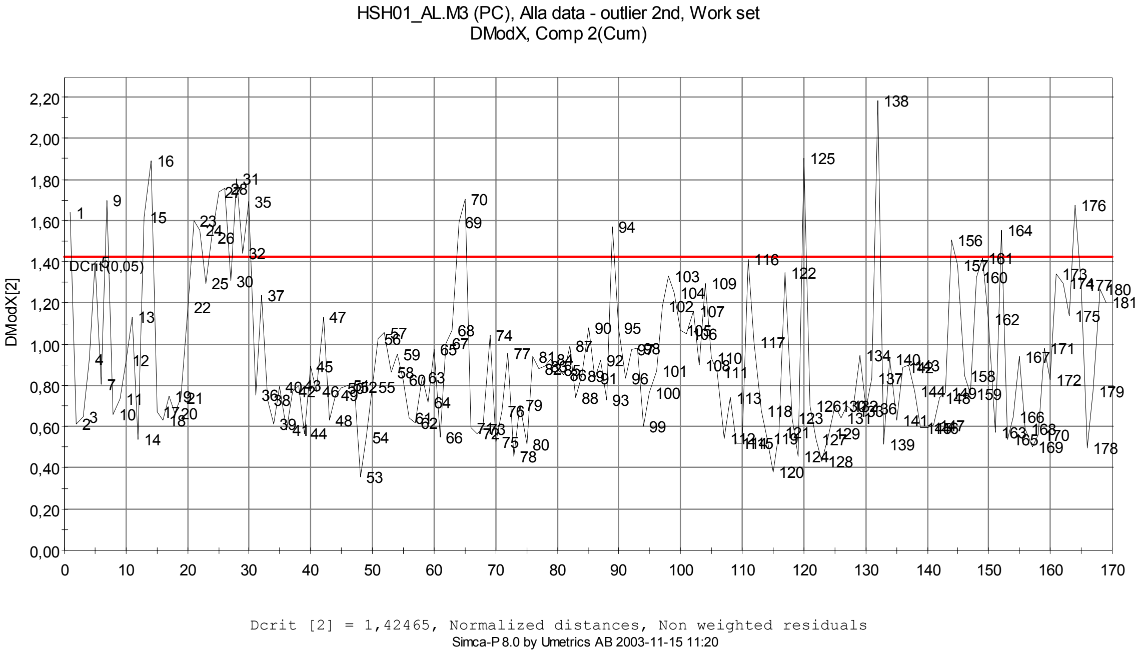

Figure 4-4. *Normalized distance (in standard deviations) of sections in the data matrix from the centre of the PCA-model M3.*

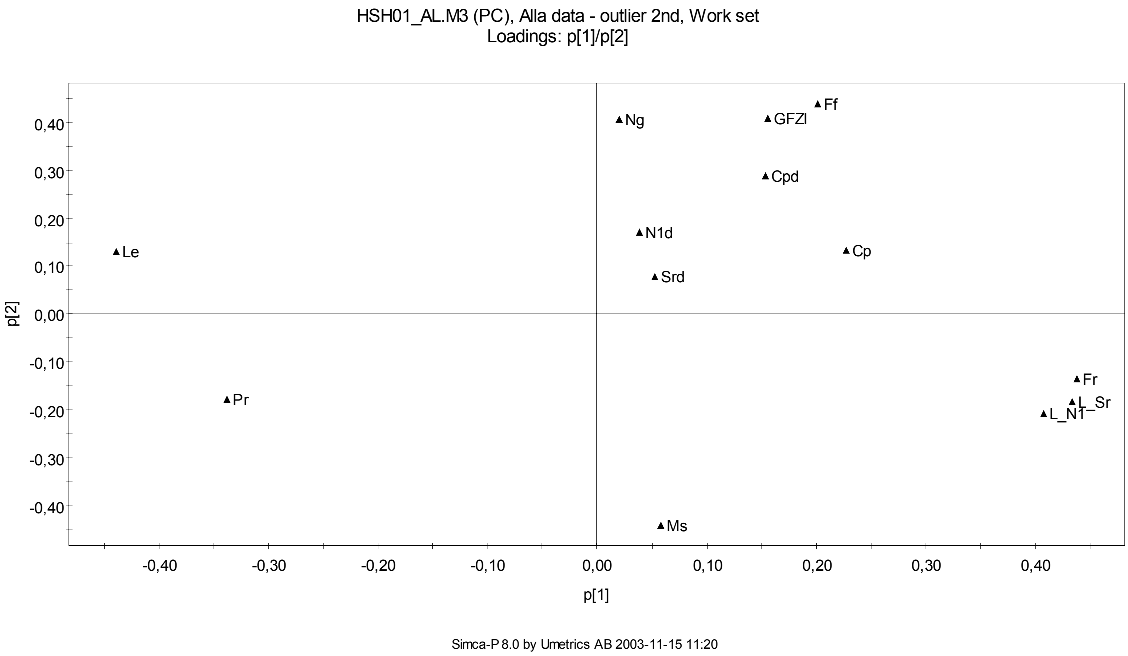

Figure 4-5. *Variable loadings for the first two PC's for model M3.*

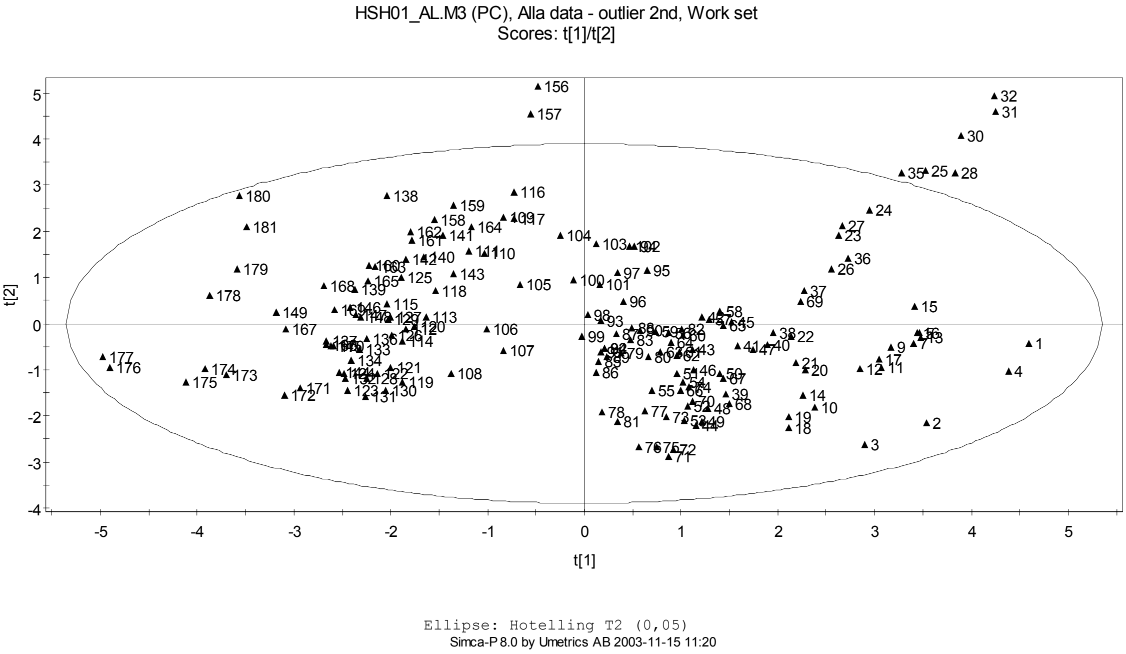

__Figure 4-6.__ Object scores for the first two PC's for model M3.

Model PCA_M3: 1st PC – 36,3% of total variation

The first component (horizontal direction in Figure 4-5) shows properties related to the depth but also correlation to some extent with fracturing. This mixture appears since the average fracture frequency is larger in the upper part of the hole. The resistivity of the borehole liquid decreases with depth and this is the reason why the electrical logs show correlation with depth.

Positive direction (right)
Fr, L_Sr, L_N1 and to some extent Cp, Ff, GFZI, Cpd

Uncorrelated
Ms, Sr, N1d, Ng

Negative direction (left)
Le, Pr

Model PCA_M3: 2nd PC – 19,2% of total variation

The vertical direction in the variable loading plot (Figure 4-5) indicates different types of lithology and its relation to fracture frequency.

Rocks with high values of natural gamma radiation and low magnetic susceptibility are correlated to sections with high values of fracture frequencies and the the manual classification of GFZI to the top, versus rocks with low fracture frequency correlated to high magnetic susceptibilty to the bottom. The correlation between fracture frequency and the geophysical logs is rather weak. This might be explained by the fact that the sections with the largest geophysical anomalies have been identified as statistical outliers that are not included in the analysis. Compare variable loadings in Figure 4-5 with loadings in Figure 4-2 (ignore change of direction of second component).

Positive direction (top)
Ng, Ff, GFZI, Cpd

Uncorrelated
All other variables

Negative direction
Ms

Model PCA_M3: PC1 & PC2 – 55.5% of total variation

The object score plot in Figure 4-6 (= borehole sections) shows a fairly homogeneous data set with trends towards the extreme values in the core of the fracture zone.

A second PC-analysis were applied on the data set where the variables Pr and Le were excluded since they dont seems to be correlated with the fracture frequency. The name of this model was M4.

Model PCA_M4: PC1 – 30.5% of total variation

The first component is still indicated by the Fr, L_Sr, L_N1 to the left and the second vertical componet shows the same pattern for the varaibles related to rock lithology and the fracture frequency. The variable loadings can be seen in Figure 4-7 and the object scores in Figure 4-8. The change of the direction of the first component, with Fr to the left compared to Figure 4-5, is just a random behaivor of the analysis and has no significance.

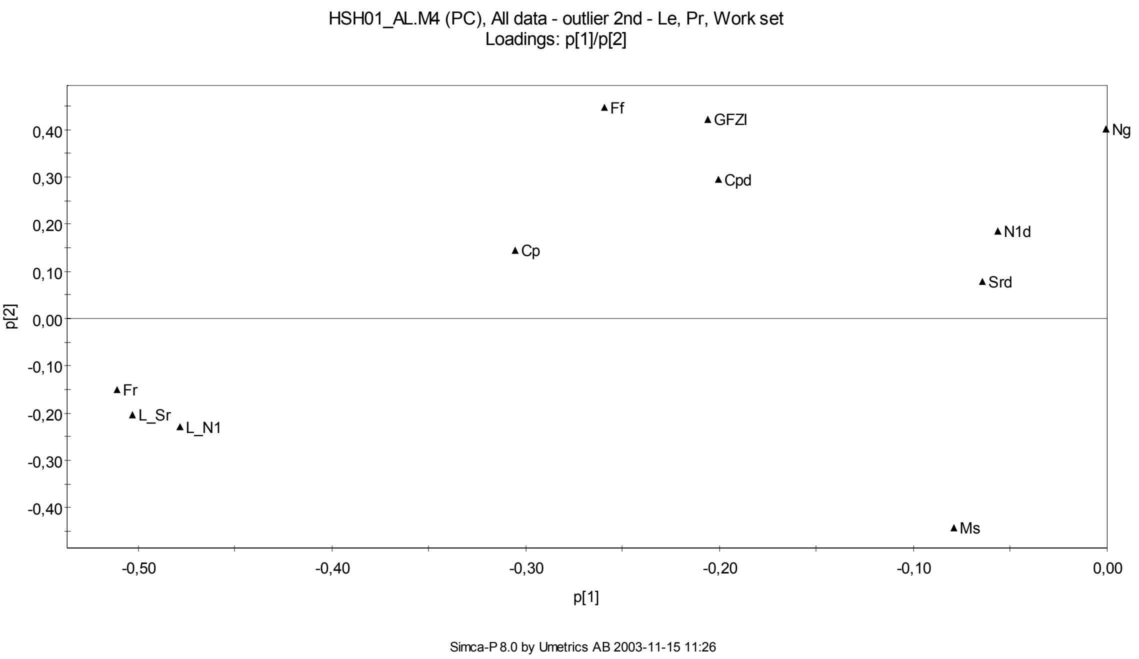

Figure 4-7. *Variable loadings for PC1 and PC2 for model M4.*

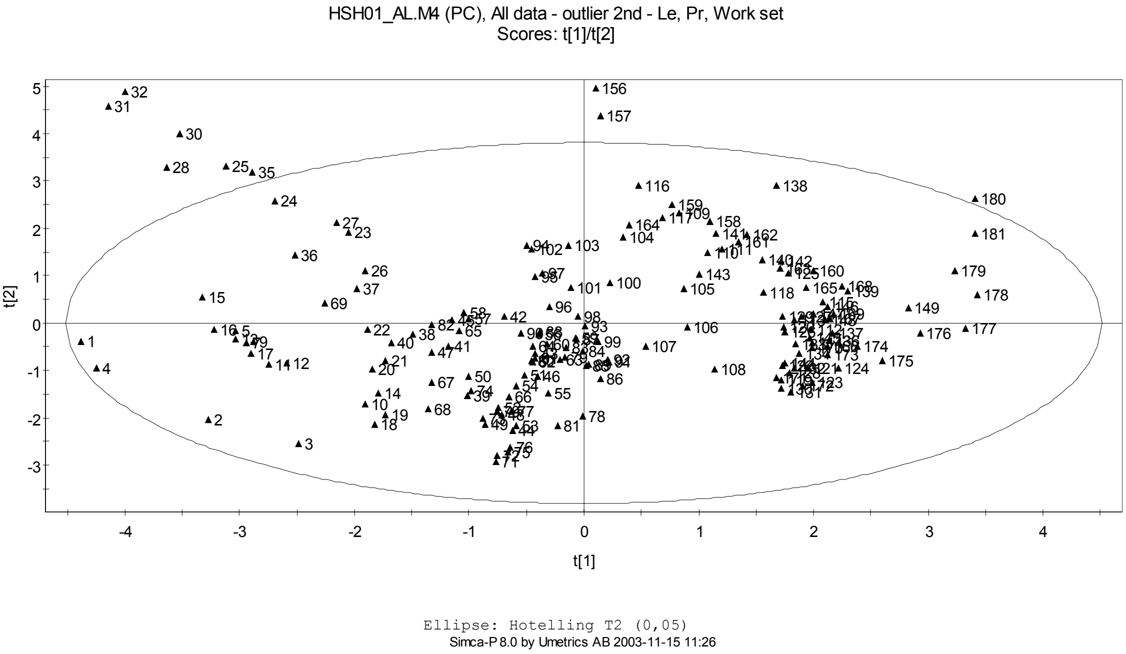

***Figure 4-8.** Object scores for PC1 and PC2 for model M4.*

Modell PCA for M3 and M4

The explained variation of these models is just above 50.0% with a clear relation to fracturing and lithology in the second component.

The variables Cp, N1d, Srd, Fr, L_Sr, L_N1 do not show any significant correlation with fracturing or lithology and are located along the first component or close to the center. These variables are independent of the fracturing of the rock on this general level where the outliers have been excluded. When the extreme sections close to the core of the fracture zone where included, Figure 4-2, also Cp, N1d and L_N1 were related to the Ff and GFZI.

The significant variables are Ff, GFZI, Ng and to some extent Cpd versus Ms for model M3 and M4.

4.5.2 Summary of PC-analysis

The PC-analysis is summarized in Table 4-2.

The results from the PC-analysis generate two significant components. The position along these components for every section is denoted object score, t1 and t2. The object scores are also plotted against length along the borehole in Figure 4-9.

Additionally, the normalized distance to the model centre has been calculated for every borehole section in the form of a standard deviations, DModX(PS),N.

A probability value, PModX(PS), is also calculated for every section that indicates the probability for the section to be within the confidence limits of the model. If this probability is greater than 5% we can assume that the section is within the confidence limits.

Table 4-2. PC-analysis of borehole data.

Processed primary data	Processing	Resulting data file
HSH01alla_var_01.xls	Creation of initial PCA-model	HSH01_pca_M3_01.xls (MS Excel)
	Removal of outliers	
	Creation of final PCA-model, M3	
	Calculation of principal component scores (t1, t2), distance to model [DModX(PS),N] and probability of belonging to model [PModX(PS)].	

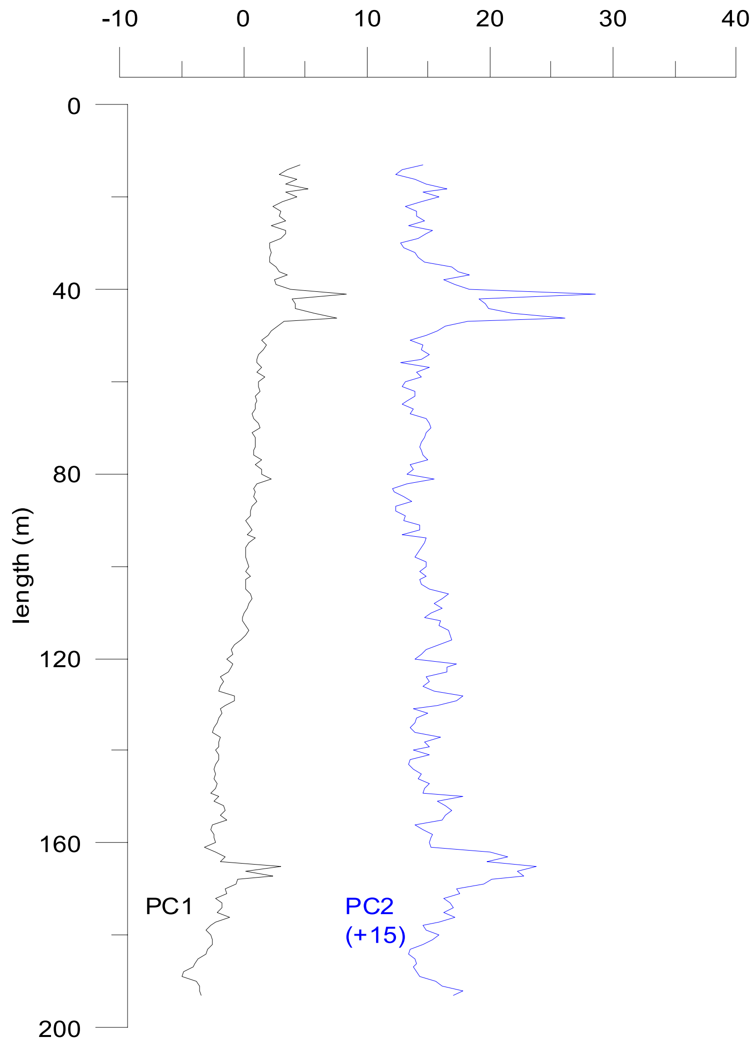

__Figure 4-9.__ Principal component object scores plotted versus length along the hole. Positive scores in PC1 are due to high electrical resistivities and also inversely related to the depth in the hole. Some correlation with fracturing can be seen since the average fracture frequency is higher for the upper part of the hole compared to the lower part. Positive scores in PC2 indicate fracturing that also correlates with presence of rocks with high gamma radiation (probably felsic) and low magnetic susceptibility (possibly due to alteration). A constant value has been added to PC2 for presentation purpose in this graph.

The two parameters that indicate whether a section complies with the model are defined as:

DModX(PS),N

Distance to the model in X space after n components for the observations used to fit the model. The distance is the standard deviation of the obersvations with scaling and centering. N stands for normalized distance.

PmodX(PS)

Probability of belonging to the model in the X-space for observations used to fit the model. Observations with probability of belonging of less than 5% are considered to be non members i.e. they are diffrent from the normal observations used to build the model.

4.6 PLS-modelling of GFZI

4.6.1 Manual classification of GFZI for PLS-analysis

GFZI is a manual classification of the rock into fracture zones by taking geological and geophysical information into account. The numerical value that is assigned to each class is our predefinition achieved with FZI.

In this case the data set is quite small, hence no restriction is used for the fracture frequence in each class before the PLS modelling.

4.6.2 PLS-model for GFZI and prediction of NGFZI

The normalized distance, DmodX, for each section to the central part of the model is shown in Figure 4-10. No new outliers were detected by the PLS-analysis and the PLS-model is based on the same data set as for the PCA analysis with excluded outliers. After the outliers have been deleted, the distance to the model for each section shows a fairly homogenous data set, Figure 4-11. Compare distance to model for the PCA-analysis in Figure 4-1 and Figure 4-4.

The PLS analysis is now directed to the prediction of GFZI with the remaining variables and objects. This is done by assigning GFZI as Y-values and the other variables as X-values in a model that can be seen as a stepwise regression analysis using the object scores in the X-space to model the Y variable. Two data blocks are defined in this way, X and Y. For each added PLS-component the amount of explained variation in the two blocks are given, i.e. how much of the variation in the X-block is used to explain a certain amount of variation in the Y-block. Step-by-step a new component is added from the X-space to model the remaining variation of Y as long as they are significant due to cross-validation criteria.

The analysis indicated two significant components after cross-validation. The model is called M2 in the discussion, tables and figures below:

PLS_M2	**X**	**Y**	
Comp.1	27.6%	38.1%	explained variation
Comp.2	25.1%	12.7%	explained variation
Total	**52.7%**	**50.8%**	**explained variation**

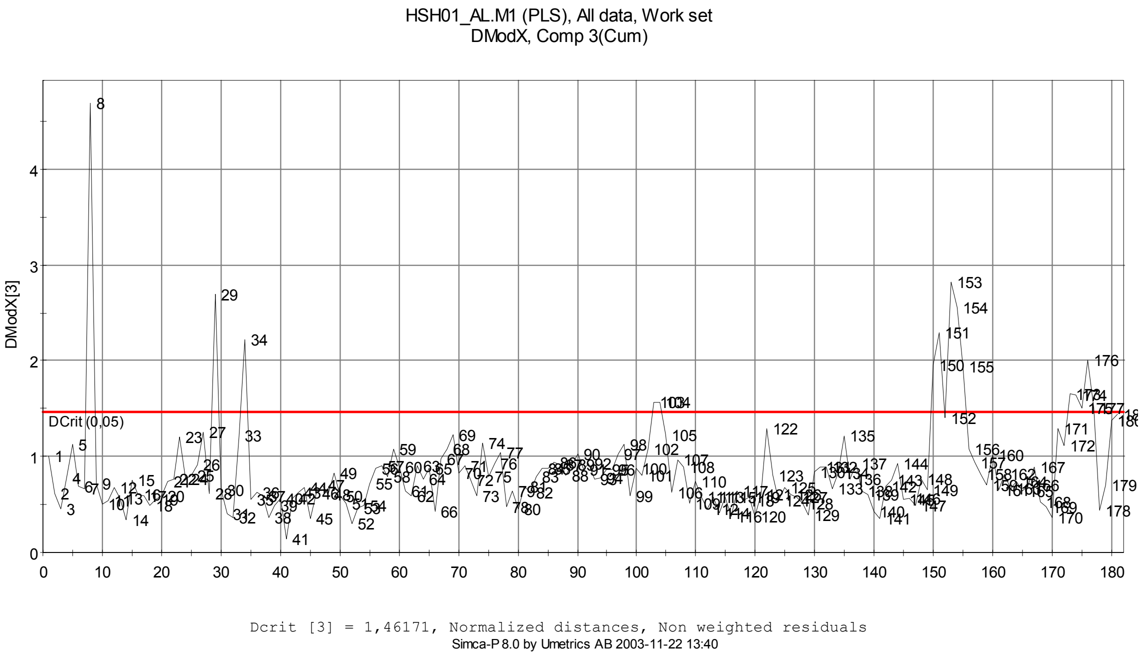

Figure 4-10. *Normalized distance to model centre for each section included in the initial PLS model estimation before the outliers has been excluded.*

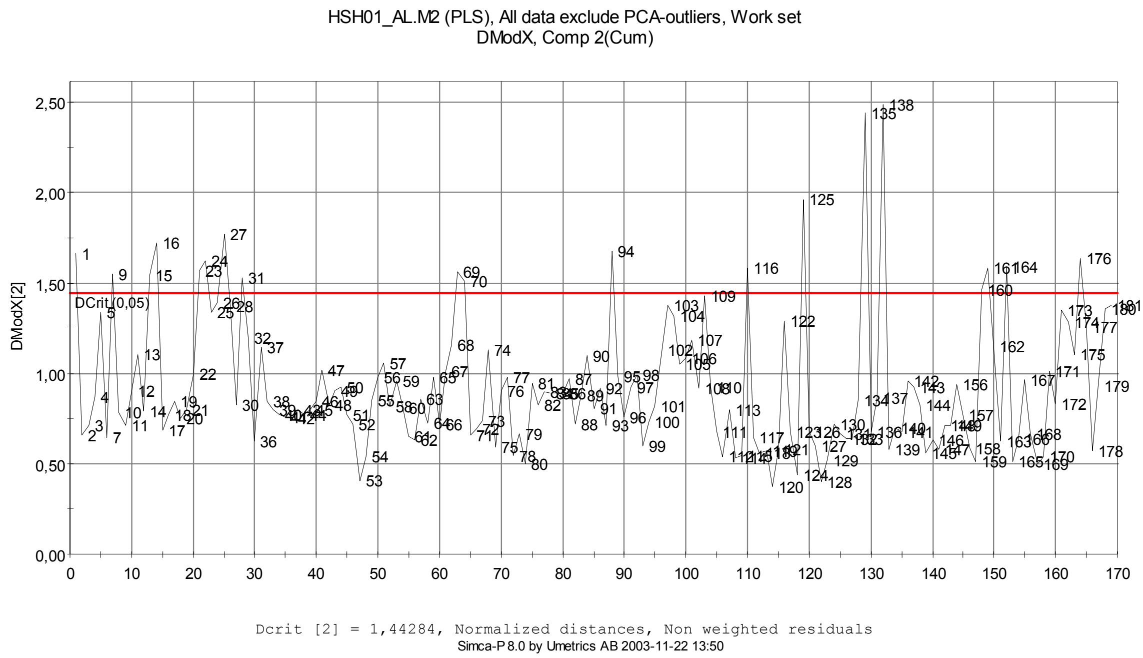

Figure 4-11. *Normalized distance to model centre for each section after outliers has been deleted.*

With the first component, 27.6% of the variation in X is used to explain 38.1% of the variation in Y (= GFZI). For component 2, 25.1% of the remaining variation in X is used to explain another 12.7% of the remaining variation in Y.

A total of 52.7% of the variation in X was used to explain 50.8% of the variation in Y with two significant components.

The PLS components are interpreted as follows:

PLS component 1 – X: 27.6% – Y: 38.1%

The variables Ff is the most important variable fror the prediction of GFZI and on the opposite side the variables Ms and Pr are most important as shown in Figure 4-12.

PLS component 2 – X: 25.1% – Y: 12.7%

The variable Ff, Figure 4-13, is also here strongly correlated with GFZI. Additionally the variable Le has changed side indicating that the depth in the hole is related to GFZI. This might be an artefact since a fracture zone is located close to the bottom of the hole and another one close to the top and might not indicate a true universal correlation between the depth and the GFZI. Caution should be taken to use this component as a basis for prediction of the GFZI on a general basis.

In Figure 4-14 the relation between the variables and GFZI is graphically indicated for the two component model. Variables close to GFZI has strong influence on GFZI as variables on the opposite side of the center indicate inverse correlation.

The importance of the X-variables can be shown as a line plot, Figure 4-15, indicating the strength of each variable to predict the GFZI.

After scaling the model can be transfered to show the scaled weigths for of each variable for the prediction of GFZI as in Figure 4-16.

This graph can also be expressed as the coefficients for the regression model including the constant term and sorted in the order of the size of the coefficients for each variable, Table 4-3.

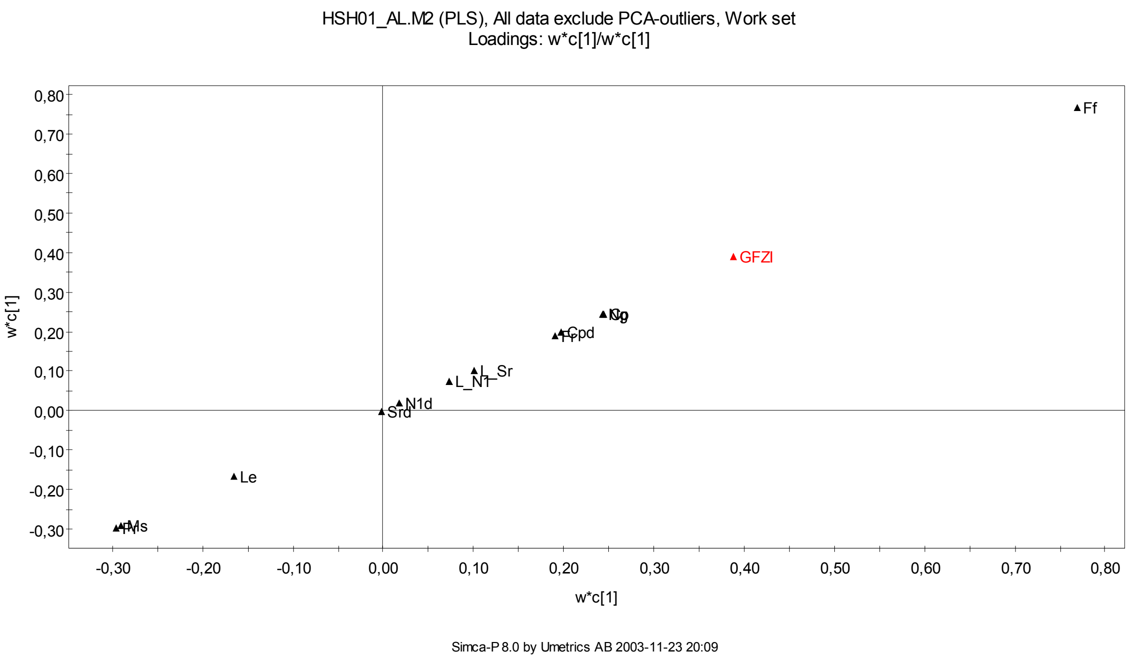

__Figure 4-12.__ Model PLS-M2. First component influence on GFZI.

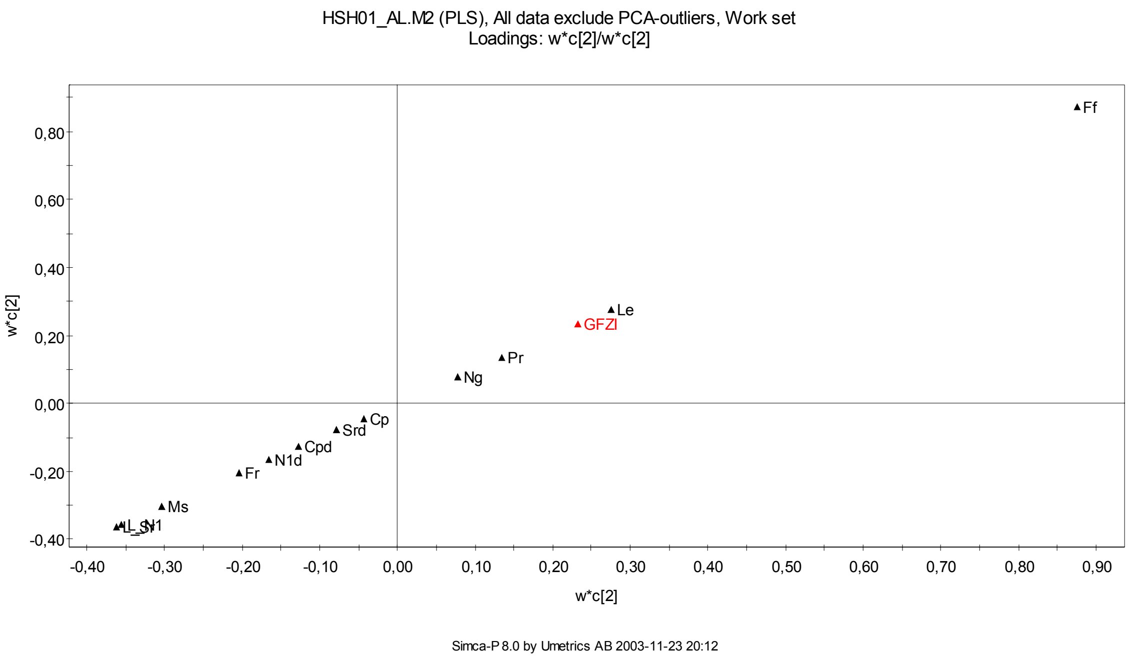

Figure 4-13. *Model PLS-M2. Second component influence on GFZI.*

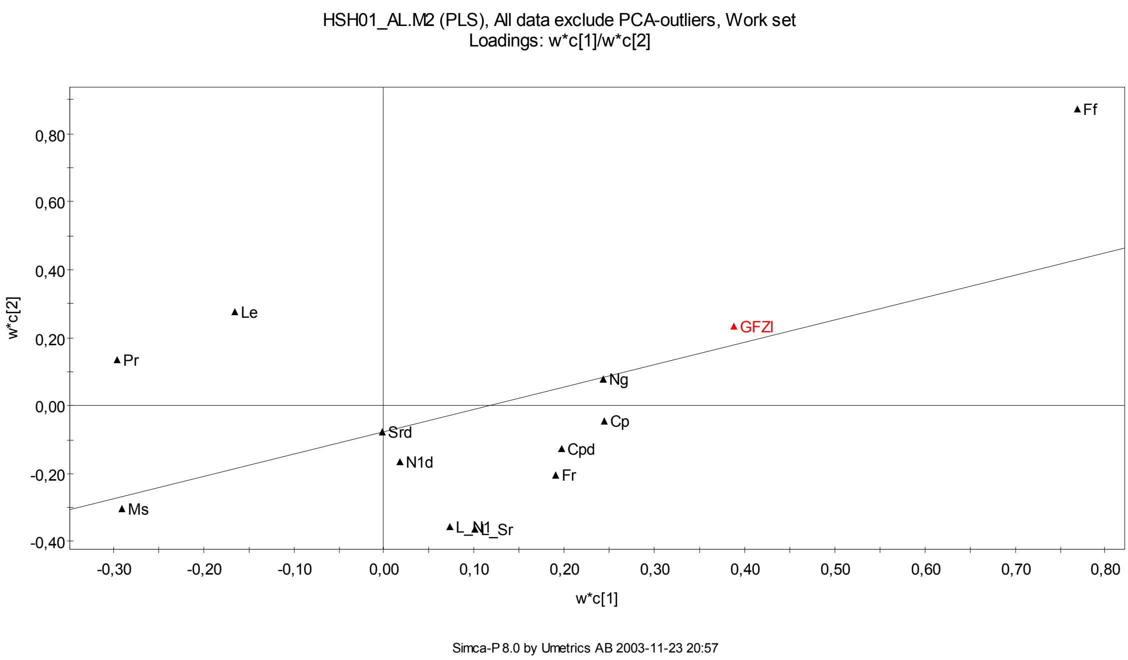

Figure 4-14. *Model PLS-M2. The influence of variables on GFZI for a two component model.*

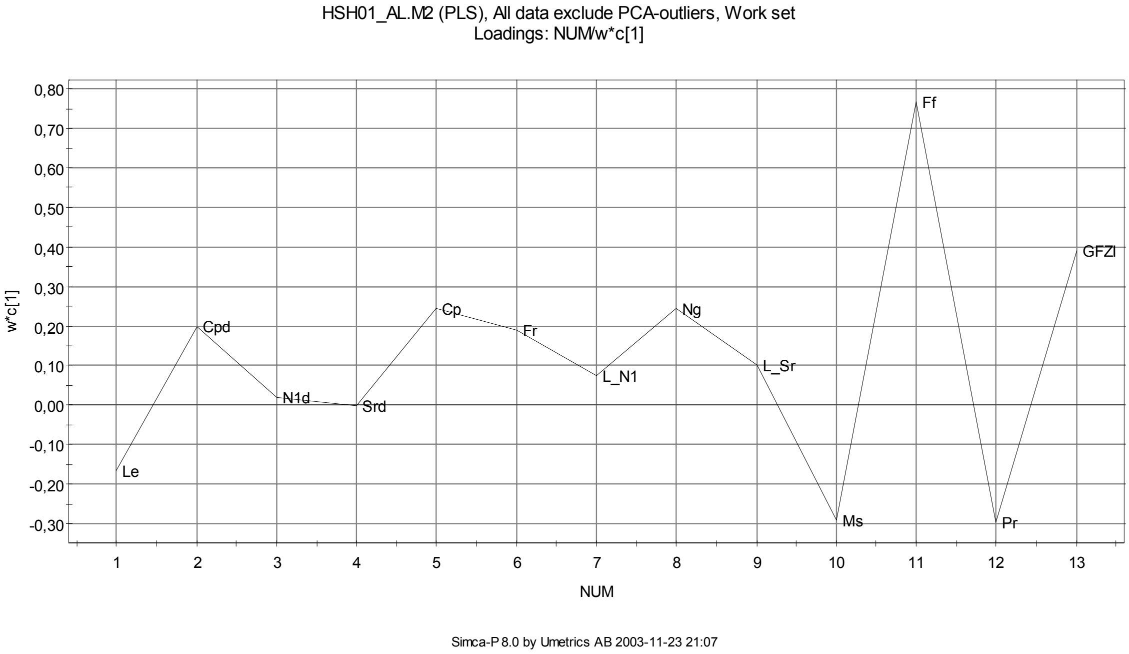

Figure 4-15. *Model PLS-M2. The indication of the influence of each variable on GFZI for a two component model.*

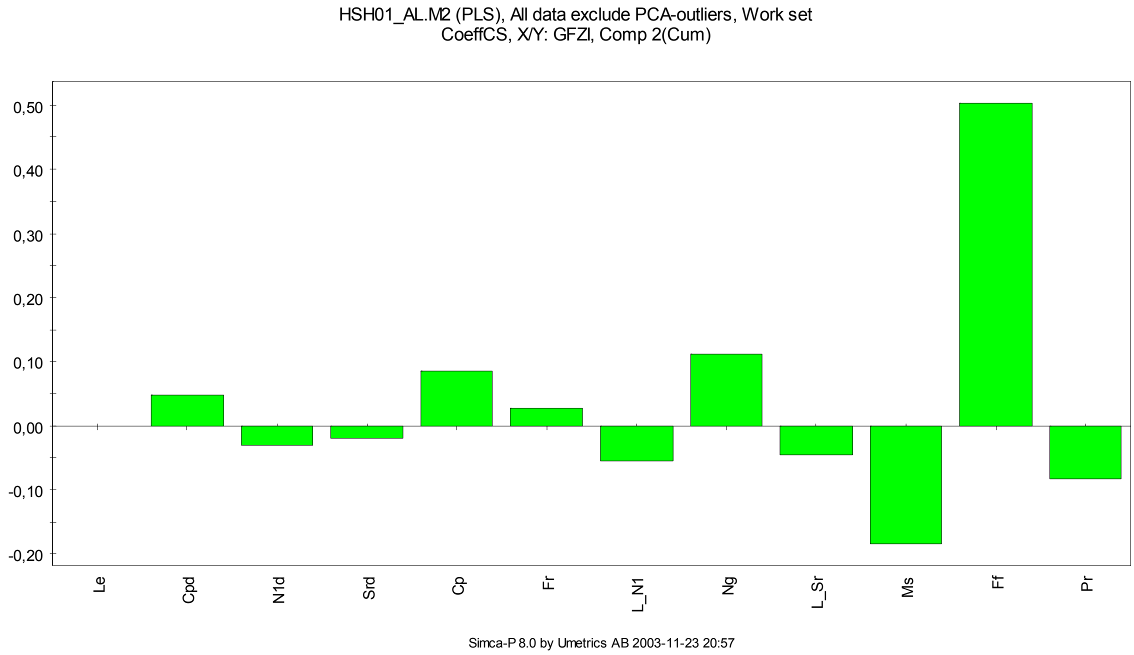

Figure 4-16. *Model PLS-M2. The importace of the variables for prediction of GFZI.*

Table 4-3. Model PLS-M2. ThePLS regression coefficients corresponding to centered and scaled X, and scaled GFZI, for the two significant components.

CoeffCS[2]	Const	Ff	Ng	Cp	Cpd	Fr	Le	Srd	N1d	L_Sr	L_N1	Pr	Ms
GFZI	0,332	0,503	0,113	0,085	0,047	0,027	0,000	–0,019	–0,032	–0,045	–0,055	–0,083	–0,184

4.6.3 PLS-model for GFZI without hole depth and penetration rate

An attempt was made to calculate NGFZI without using the two variables Le (=hole depth) and Pr (=penetration rate) as X variables. The new model is called M3 and was based on the same sections as for model M2. The influence of the variables for the first and the second component are shown in Figure 4-17.

To show more in detail the relation between the variables and the GFZI, the two components are shown separately in Figure 4-18 and Figure 4-19.

The corresponding variable influence for the M3 model are shown in Figure 4-20, and the regression coefficents after scaling in Figure 4-21 and Table 4-4. Object scores for the two significant PLS-components of model M3 are plotted against length along the hole in Figure 4-22.

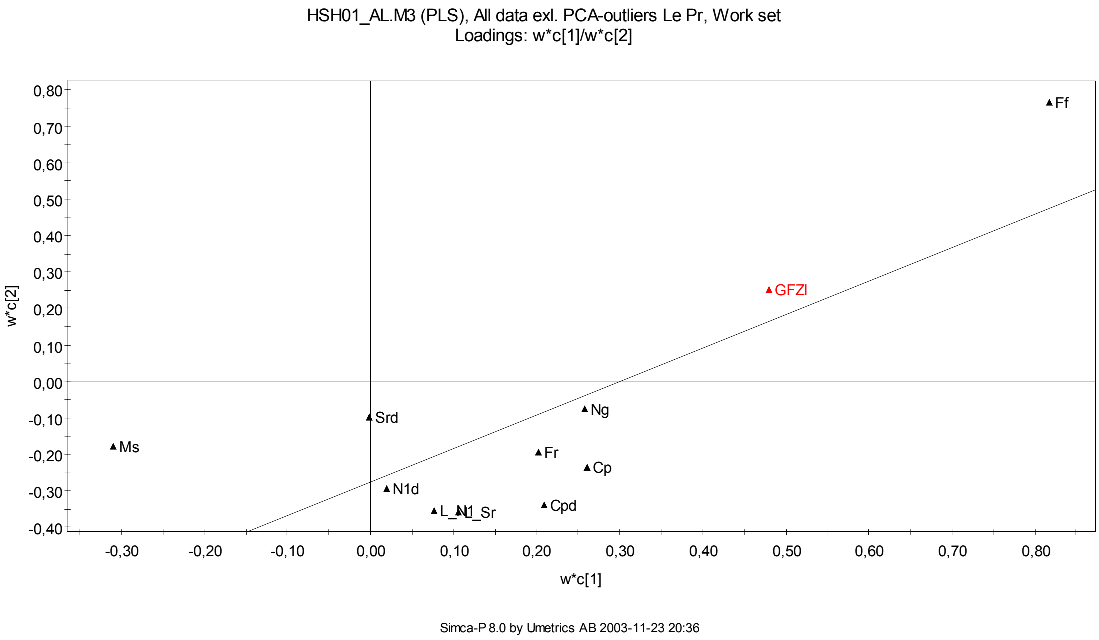

Figure 4-17. *Model PLS-M3. The influence of variables on GFZI for a two component model with Le and Pr excluded.*

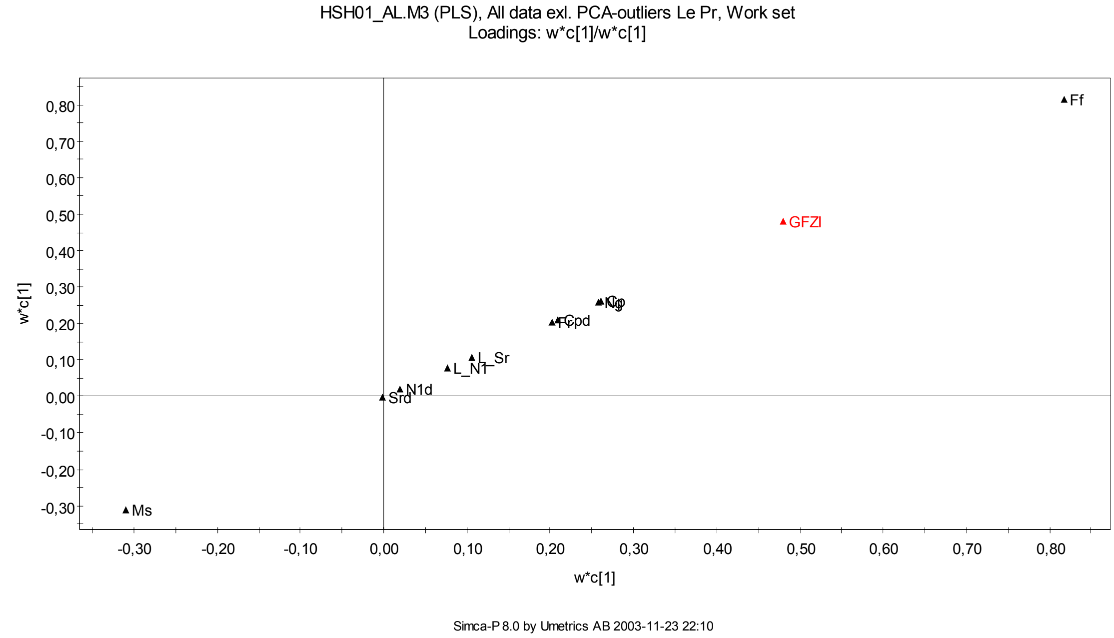

Figure 4-18. *Model PLS-M3. The influence of the variables on GFZI for the first component.*

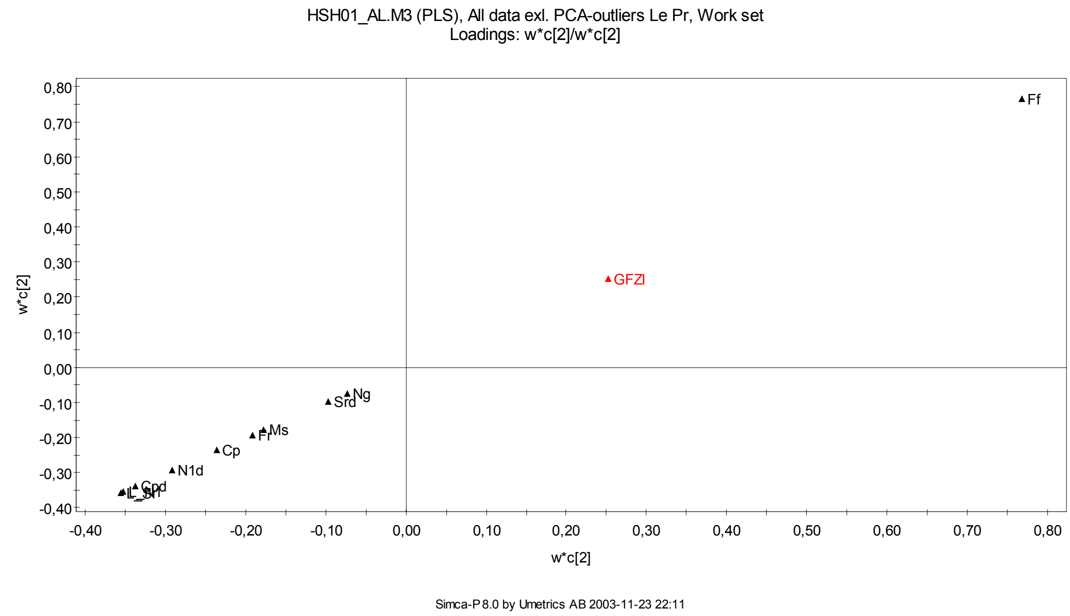

Figure 4-19. *Model PLS-M3. The influence of the variables on GFZI for the second component.*

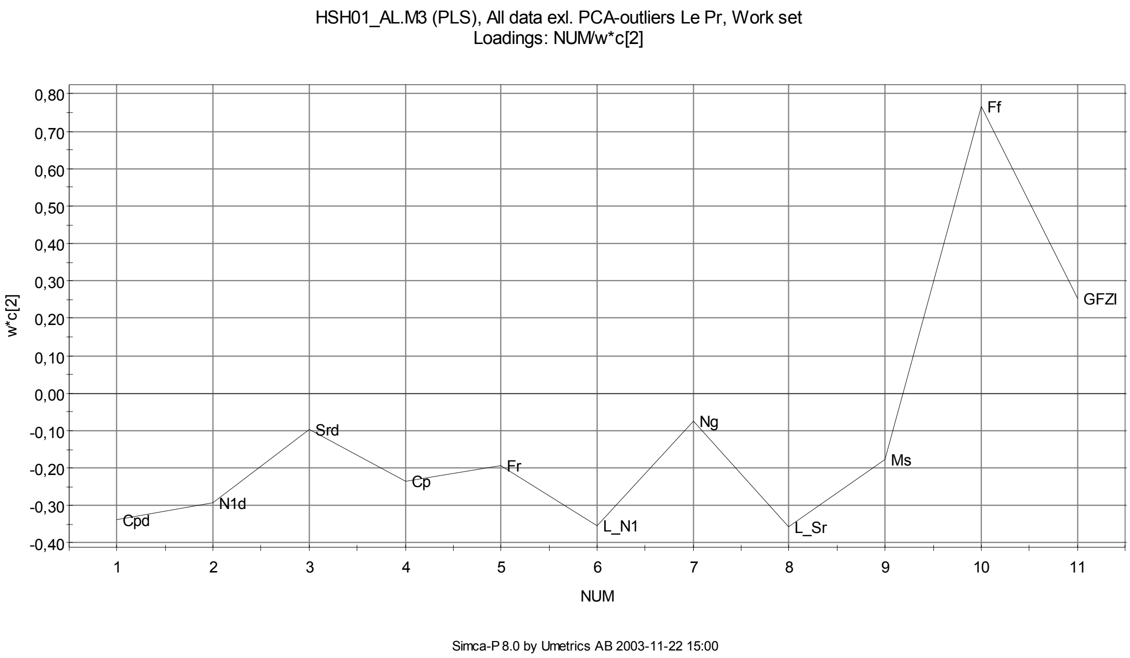

Figure 4-20. *Model PLS-M3. The indication of the influence of each variable on GFZI for a two component model without variables Le and Pr.*

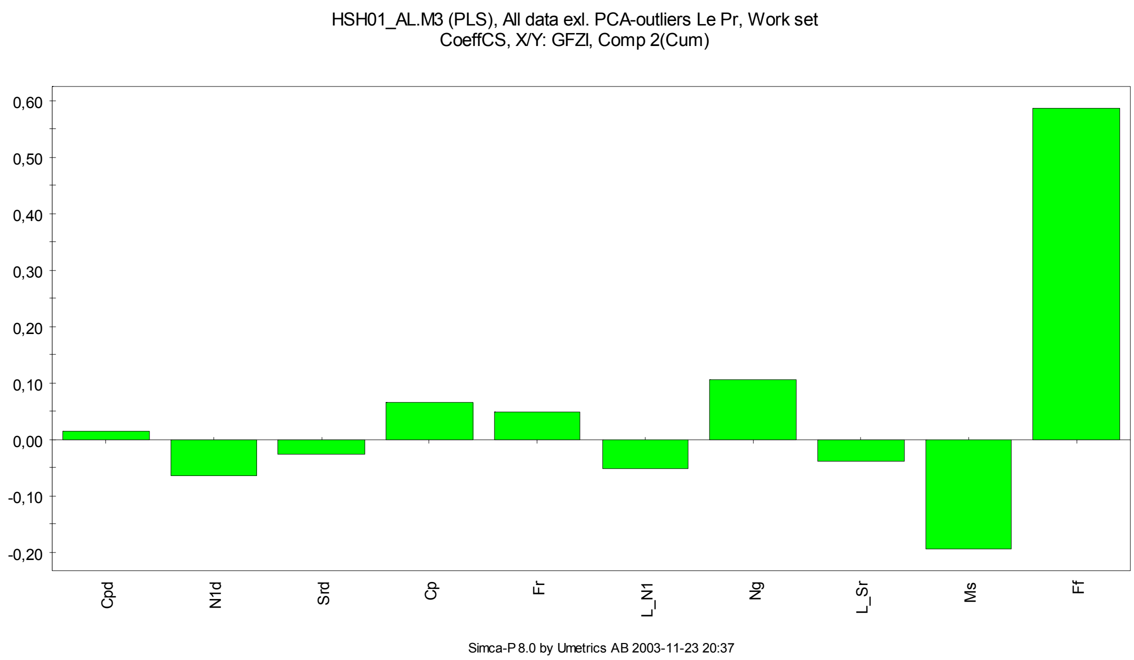

Figure 4-21. *Model PLS-M3. The importance of the variables for prediction of GFZI.*

Table 4-4. Model PLS-M3. ThePLS regression coefficients corresponding to centred and scaled X and scaled GFZI for the two significant components. Variables Le and Pr were excluded for the model calculation.

CoeffCS[2]	Const	Ff	Ng	Cp	Fr	Cpd	Srd	L_Sr	L_N1	N1d	Ms
GFZI	0,332	0,587	0,105	0,065	0,048	0,015	–0,025	–0,039	–0,052	–0,065	–0,193

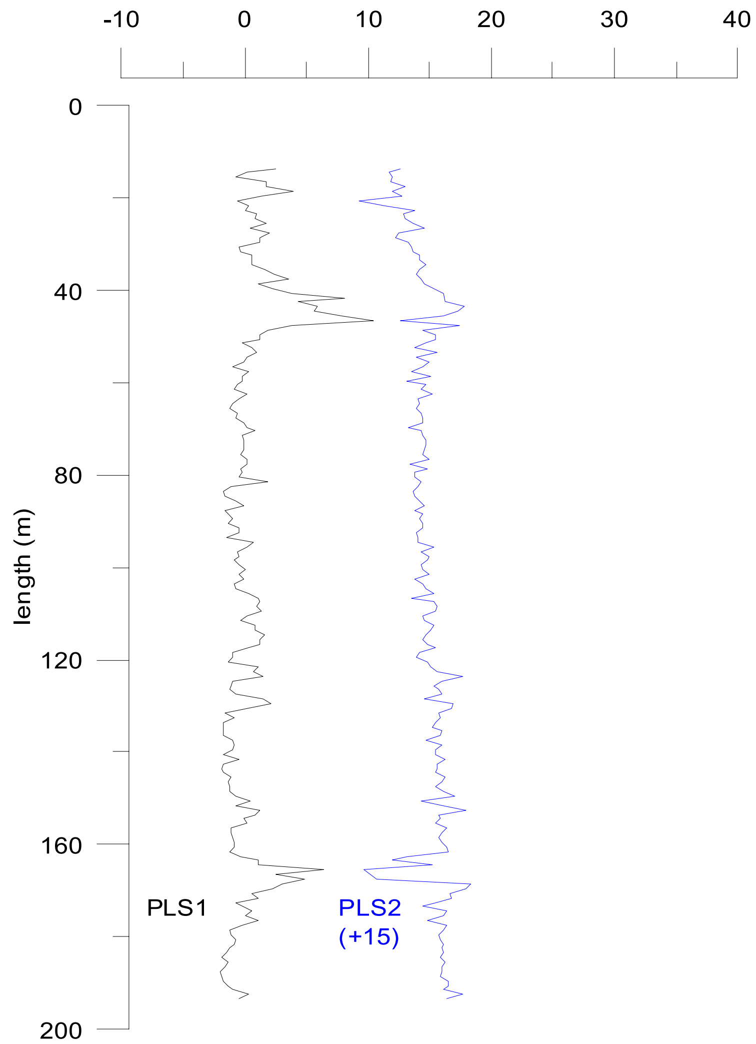

Figure 4-22. *Object scores for the two significant PLS-components of model M3 plotted against length along the hole. A constant value has been added to PLS2 for the graphical presentation. Note that PLS1 is very similar to PC1 in Figure 4-9, indicating that the two methods can provide similar results.*

4.6.4 Residual between observed GFZI and predicted NGFZI

The result from the prediction of NGFZI with model M2 and M3 has been analyzed in the form of a normal probability plot for the residual between the observed value of GFZI and the predicted value of NGFZI, Figure 4-23 and Figure 4-24. A fairly straight line for the probability plot indicates that the residuals are close to a normal distribution. However there is a brake in the straight line that indicates that data consist of two distributions that is interpreted as, the unaffected rock data and data from the core of the fracture zone.

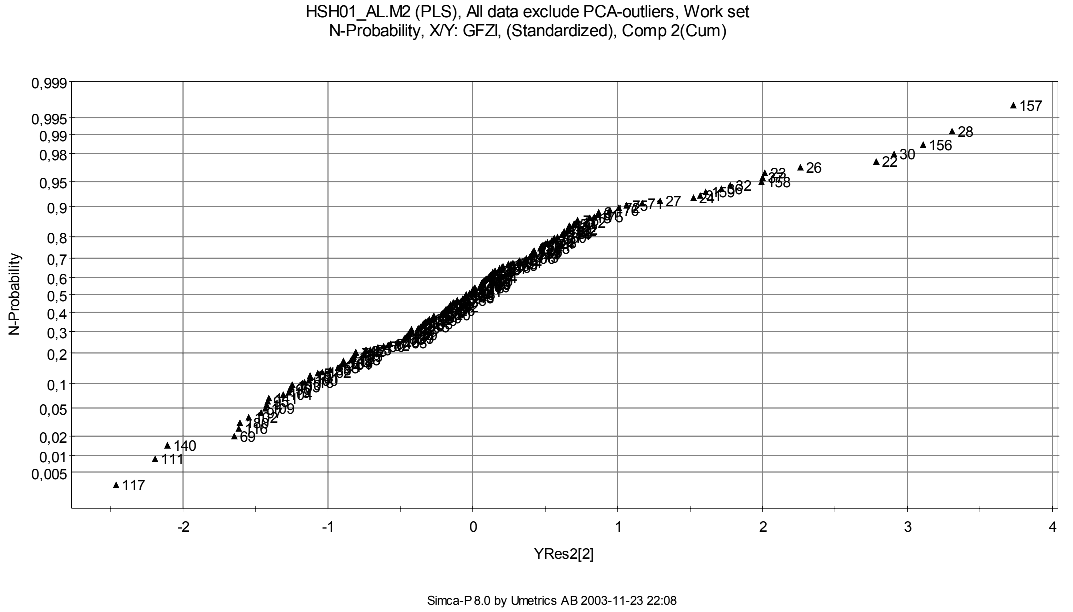

Figure 4-23. *Normal probability plot for the residual from model M2, $Y_{obs} - Y_{pred}$, ($GFZI_{obs} - NGFZI_{pred}$)*

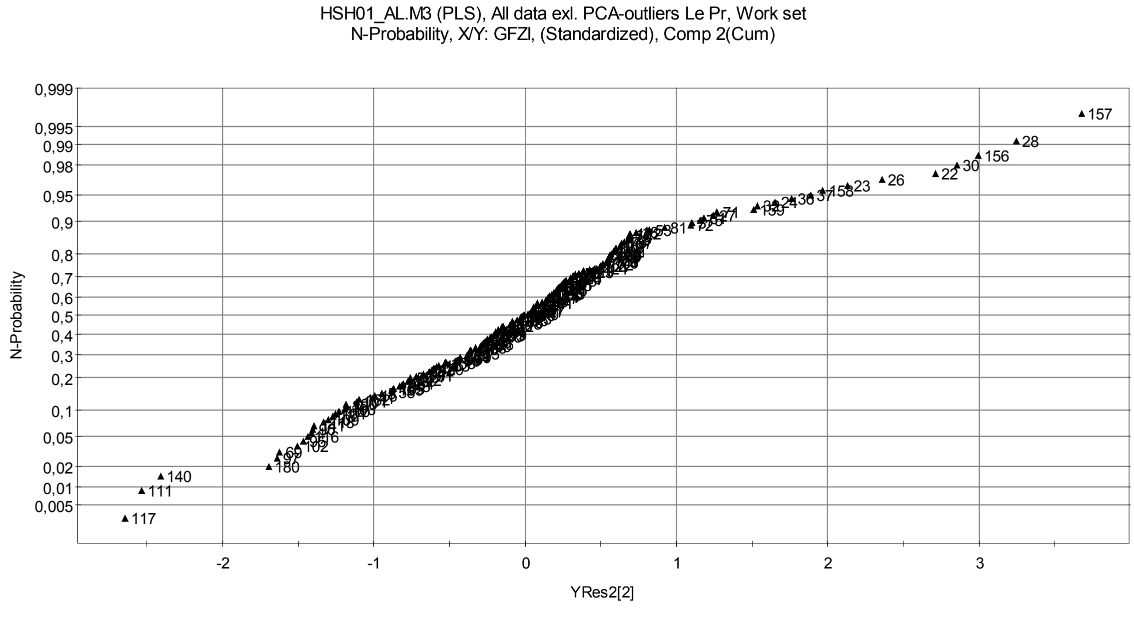

Figure 4-24. *Normal probability plot for the residual from model M3, $Y_{obs} - Y_{pred}$, ($GFZI_{obs} - NGFZI_{pred}$)*

4.6.5 Conclusions for PLS-modelling of NGFZI

The general conclusion is that only Ff has a major influence on the variation of GFZI. The modelling power of the X-variables is not fully satisfactory, describing approximately 53% of the variation of GFZI.

The geophysical logs have a fairly small impact on FZI. This might be explained by the fact that the sections with the largest geophysical anomalies were classified as statistical outliers and for that reason they were removed from the PLS-model generation. Keeping the outliers for the analysis might in this case have created a more local model for this borehole with a stronger correlation between GFZI and e.g. Cp, N1d and reversible L_N1 as could have been indicated by the initial PCA analysis without the removal of the outliers.

To include or remove the variables Le and Pr does not have a major influence on the results and the residual between observed GFZI and predicted NGFZI.

A Fracture Zone Index, FZI, has been calculated with model M3 that consists of two significant components. This value is here called Numerical Geological Fracture Zone Index, NGFZI. This index was calculated for all sections along the borehole and describes the subdivision of the rock into units of different fracturing with a continuous value from the core of the fracture zone to unaffected normal rock. This value is FZI after adjustments for intervals and outliers.

4.7 Calculation of FZI

NGFZI is a continuous variable varying between –0.5 and +3.5 along the borehole. Some minor adjustments have been done since the model is not aware of the limitations on the index that was predicted.

Sections that are outside the confidence limits of the model according to the variable, DmodX, have been assigned a “missing value” during the analysis. In order to avoid missing predictions in the final FZI vector, these values have been replaced by the original GFZI value for the corresponding section.

The final NGFZI value has been truncated so that the minimum value is 0.0 and the maximum is 2.5. This constitutes the final continuous value of FZI. The index can be split into discrete classes according to e.g.:

Core of fracture zone	$1.5 < FZI < 2.5$
Transition zone	$0.5 < FZI < 1.5$
Unaffected rock	$0.0 < FZI < 0.5$

The PLS-modelling and FZI-calculation is summarized in Table 4-5. The result of FZI calculation is shown in Figure 4-25.

Table 4-5. PLS-analysis and FZI-calculation.

Processed primary data	Processing	Resulting data file
HSH01_alla_var_01.xls	Creation of initial PLS-model	HSH01_pls_01.xls (MS Excel)
	Removal of outliers	
	Creation of PLS-models M2 and M3	
	Calculation of NGFZI, DmodX and regression coefficients	
	Back substitution of GFZI-values for outliers (M3)	
	Truncation to 0<FZI<2.5 (M3)	

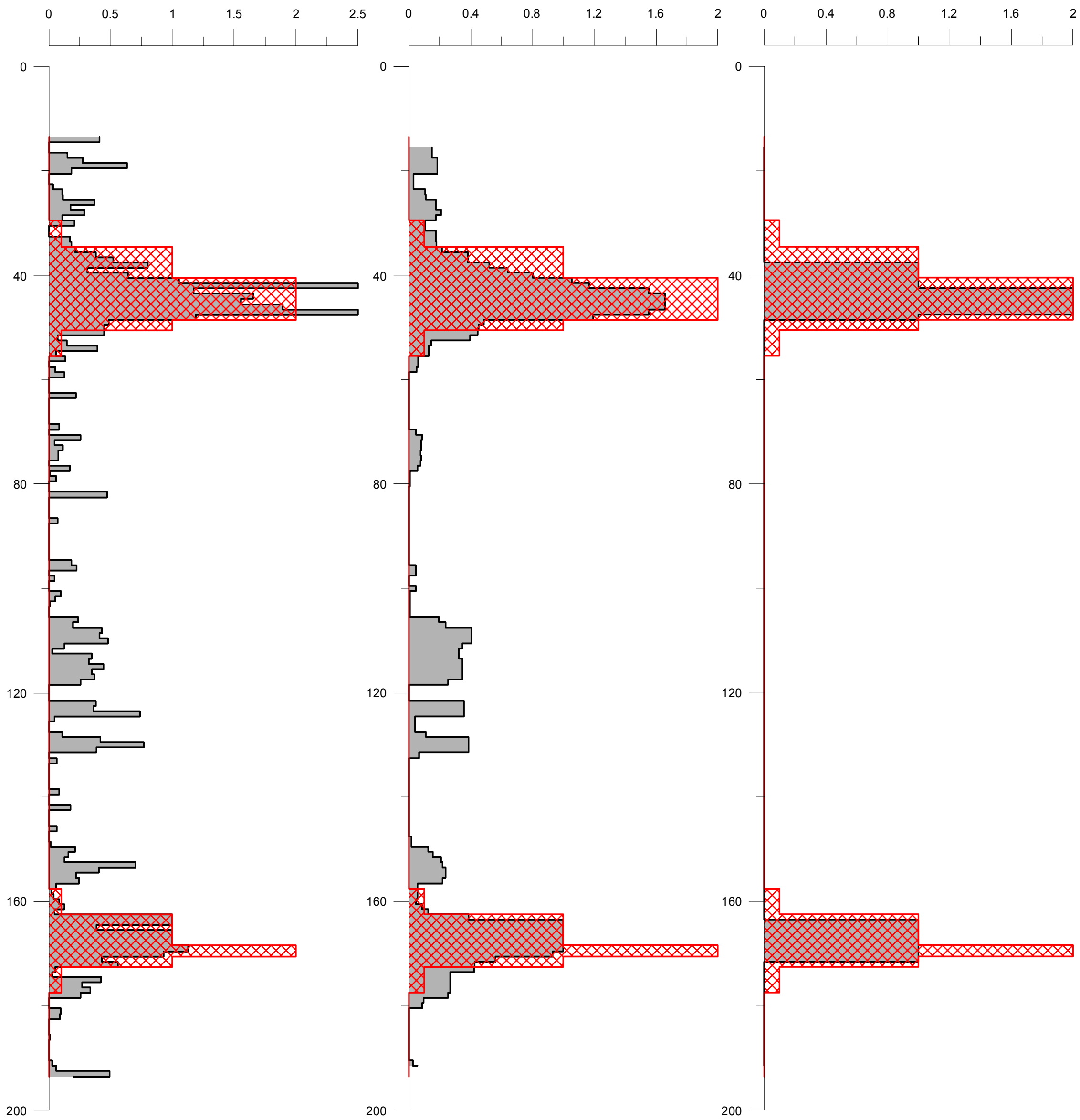

Figure 4-25. *Result of the FZI calculation in HSH01. The left graph shows the predicted FZI in grey and the original classified GFZI with red cross-hatching versus depth along the borehole. The central graph shows predicted FZI in grey after five point median filtering. The right graph shows FZI in grey as discrete classes according to the intervals FZI<0.5 → 0; 0.5<FZI<1. 5 →1; FZI>1. 5 →2.*

5 Summary and discussion

Measured and mapped variables along the borehole HSH01 have been evaluated with multivariate techniques with the purpose of calculating a Fracture Zone Index, FZI. This index should subdivide the rock into classes with information that supports interpretation of deformation zones.

By using multivariate techniques several variables can be considered simultaneously and only relevant correlated information from the variables are used for calculation of FZI. This gives a robust estimate and random and manual operator introduced noise not correlated with FZI will automatically be filtered away.

The models are based on a definition where the rock has been manually classified into three classes – core of fracture zone, transition zone and unaffected rock. This manual subdivision of the rock into classes is called Geological Fracture Zone Index, GFZI.

It is important that the input variables describe the entire range from the core of fracture zone to unaffected rock since this range is described by the model. The models will therefore contain variables that are not just indicative of fracture zones but also of unaffected rock. With the help of the model a Numerical Geological Fracture Zone Index, NGFZI, is calculated for all sections along the borehole, including the core of fracture zones, transition zones and the unaffected rock.

Only 53% of the variation in GFZI was explained by the model. The remaining 47% of the variation of GFZI could not be described by the present variables. Some part of this variation may be a noise component introduced in GFZI due to generalization during the manual classification that resulted in GFZI.

The result of FZI calculation for HSH01 is shown in Figure 4-25. The delivered data have been inserted in the database (SICADA) of SKB. The SICADA reference to the present activity is Field note No. 221.

References

/1/ **Wold S, Esbensen K, Geladi P, 1987.** Principal Component Analysis, Chemometrics and Intelligent Laboratory Systems, 2, p.37–52. Reprinted in Chemometrics Tutorials, 1990, 209–224, Elsevier Science Publishers.

/2/ **Lindqvist L, Lundholm I, 1985.** Effektivare prospektering med hjälp av dataanalys, Jernkontorets Annaler nr 4.

/3/ **Lindqvist L, Lundholm I, Nisca D, Esbensen K, Wold S, 1987.** Multivariate Geochemical Modelling and Integration with Petrophysical Data, Journal of Geochemical Exploration, 29, p.279–294, Elsevier Science Publishers.

/4/ **Esbensen K, Lindqvist L, Lundholm I, Nisca D, Wold S, 1987.** Multivariate Modelling of Geochemical and Geophysical Exploration Data, Chemometrics and Intelligent Laboratory Systems, 2, p.161–175, Elsevier Science Publishers.

/5/ **Andersson J E, Lindqvist L, 1988.** Prediction of Hydraulic Conductivity and Conductive Fracture Frequency by Multivariate Analysis of Data from the Klipperås Study Site, SKB Technical Report 89-11.

/6/ **Carlsten S, Lindqvist L, Olsson O, 1989.** Comparison between Radar Data and Geophysical, Geological and Hydrological Borehole Parameters by Multivariate Analysis of Data, SKB Technical Report 89-15.

/7/ **Black J, Olsson O, Gale J, Holmes D, 1990.** Site Characterisation and Validation – Stage 4 – Preliminary Assessment and Detail Predictions, Stripa Project 91-08, SKB Technical Report.

/8/ **Korkealaakso J, Vaittinen T, Pitkänen P, Front K, 1994.** Fracture Zone Analysis of Borehole Data in Three Crystalline Rock Sites in Finland – The principal Component Analysis Approach, Nuclear Waste Commission of Finish Power Companies, Report YJT-91-11.

/9/ **Lindqvist L, Thunehed H, 2003.** Oskarshamn site investigation. Calculation of Fracture Zone Index (FZI) for KSH01A. SKB P-03-93.