EXPLORATORY STATISTICAL ANALYSIS OF COAL QUALITY AT THE HIGHVALE MINE, (ARDLEY COAL ZONE), CENTRAL ALBERTA

COAL GEOLOGY GROUP, ALBERTA GEOLOGICAL SURVEY

PREPARED BY: R. WONG
R. STROBL
R. KRZANOWSKI
N. CHIDAMBARAM

OPEN FILE REPORT OFR 1988-09

April 1988

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

FOREWORD

The Coal Geology Group of the Alberta Geological Survey is beginning studies on the variation in coal quality and the development of models to predict coal quality throughout the province. This work is part of a three-year contract funded by the Alberta Department of Energy and the Alberta Research Council, which continues through March 1989. A status report and workplan (McCabe et al., 1987) outlines four projects undertaken by the Coal Geology Group. The projects are interrelated and together will lead to a synthesis of coal quality in the plains, foothills and mountains regions.

Project one of the status report and workplan deals with the quality of the Ardley coal zone and the Drumheller coal zone (lower Horseshoe Canyon Formation) in the plains region. For the study of the Ardley coal zone, the Highvale area was selected because of the high level of geologic confidence in the available data and the abundance of coal quality information.

The variation in quality between seams as well as the lateral variation of individual seams are investigated. We hope that predictive models and statistical techniques developed in this study can be applied to further studies in other selected areas. We also hope that this work will serve as a base and means of comparison for the regional assessment of the coal quality of the Ardley coal zone planned in the second phase.

EXECUTIVE SUMMARY

The purpose of the present study is to examine the variability in coal quality of the Ardley coal zone. This first exploratory phase involves the study of distributional characteristics, correlations and multivariate comparisons. The data examined include both proximate and ultimate data, on a seam-by-seam basis.

The results indicate considerable variability from seam to seam. Generally, seams 1 and 2 are the most consistent, with relatively better quality. Seam 5 is the least consistent, and the quality is relatively low. The other seams are intermediate between these two extremes.

There is no unexpected relationships from the correlation analyses, although one has to use rank correlations for inference in most cases because the data distributions do not generally follow the normal form. The present study also provides quantitative information on data characteristics, which are useful to the second phase of analysis. Such information is particularly important in statistical model identification of subsequent analyses, as well as in the selection of appropriate statistical techniques.

Recommendations have been made regarding the approaches of the second phase of the present study. It is stressed that geological models and information are essential in guiding the statistical work. It is also suggested that a variability measure based on the absolute departures from the sample median is appropriate for the Highvale data.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to TransAlta Utilities for making the coal quality database from the Highvale Mine available for statistical analysis. Special thanks also go to TransAlta Utilities and Monenco Ltd. for making available the report, "Highvale and Whitewood, Extended Geologic Mapping" which is in preparation. The research project was jointly funded by the Alberta Office of Coal Research & Technology and the Alberta Research Council. Maureen FitzGerald typed the manuscript.

INTRODUCTION

The Highvale Mine is located 80 km west of Edmonton and lies immediately south of Wabamun Lake (see figure 1). It is the largest producing coal mine in Canada, supplying electricity to Alberta from two generating plants, the Sundance and Keephills. In 1982 approximately 8 million tonnes of coal were mined, increasing to about 13 million tonnes in the current year. The rank of the coal is subbituminous "B". The mine permit boundaries extend over 20 km, parallel to the outcrop edge of the Paskapoo Formation, where the Ardley coal zone is at shallow depths.

The geology of the Highvale area has been extensively studied, and with over 1000 wells drilled, represents one of the best explored coal deposits in Alberta. Coal quality data from 315 coreholes are available and the mining characteristics of the seams are documented in detail by previous studies. Because of the wealth of coal quality data available and the laterally extensive nature of the mineable seams, the Highvale Mine data was considered to be appropriate for the preliminary studies of the Ardley coal zone.

The approach to analysis is basically statistical. Statistical methods with and without considerations of spatial dependence will both be used. The former category includes such geostatistical techniques as variograms and kriging analyses. The latter category includes mostly classical statistical methods for basic data comparisons and predictive model building. Therefore, for the purpose of coal quality prediction, both autocorrelation and cross-correlation structures in the data will be involved. While the variogram represents the autocorrelation in the data filed, regression models based on the cross correlation between coal quality and a set of geological predictors, are useful representations of relationship between variables.

Two separate phases of analyses are planned. The first phase is exploratory in nature. The emphasis is on the basic descriptions and comparisons of the data. This phase also includes the compilation,

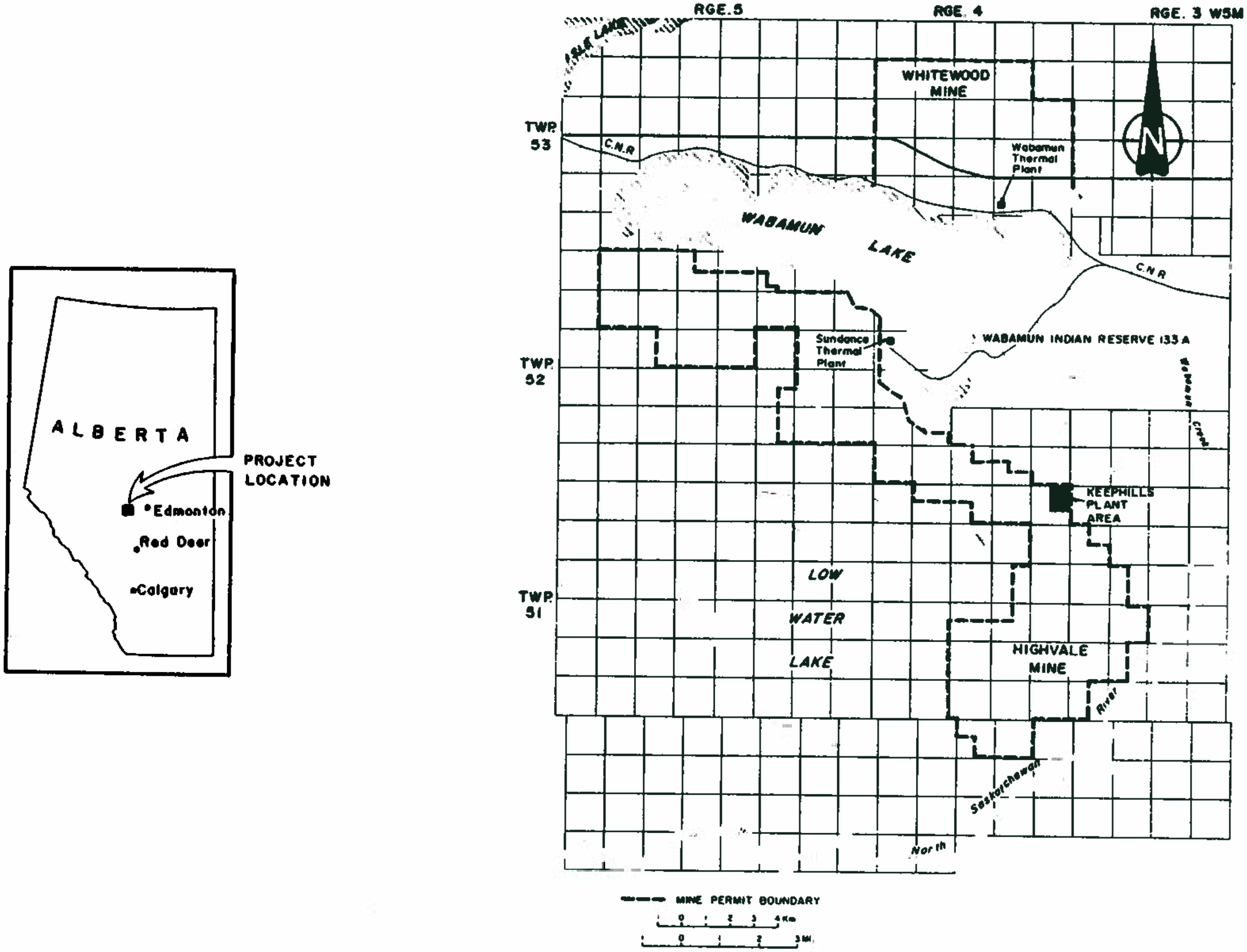


Figure 1. General location map of the Highvale Mine limits and locations of the Sundance and Keephills thermal electric plants (modified from Nikols, 1985).

pretreatment and screening of the Highvale data for analysis. The results provide the basis for the more advanced treatments and analyses of the second phase. The second phase is more of a confirmatory nature, with emphasis on more rigorous testing of hypothesis, model building and validation. The geological aspects will be further stressed in this phase. Also geological models of coal variability will provide the necessary guidelines for statistical analyses. For example, geologically, one would expect a difference in coal quality between coastal and alluvial depositional environments. Therefore, a statistical comparison between samples from these environments is planned for future studies. Basically, this will be a search for statistical evidence of the geological models, which may be a simple one such as the relationships between coal quality and the sedimentology of the mineable coal seams.

The first step was the separation of the data into geologically homogeneous groups. The data were subdivided according to the six seams that were correlated across the entire minesite. The preliminary results indicate that the coal quality varies considerably from seam to seam. In addition, data derived from coals along their subcrop edge must be treated separately or left out of the analyses because coals from these areas are commonly affected by weathering and/or oxidation processes.

Geological interpretation should always be made of the statistical results, since the geology of the deposit is closely related to the quality of coal. The identification of the geological factors affecting coal quality is of prime importance in this study. It is also useful to see if variations in coal quality are attributed to sampling errors, or the tectonic and depositional setting during the accumulation of coal-forming peats.

OBJECTIVES

The overall objectives of this study (phases 1 and 2) are to create a systematic method of characterizing Ardley coals and to quantify the variations in their quality. The work completed in phase 1 will provide a base for the comparison of the variation in coal quality typical of the Highvale minesite with the quality variation observed regionally. Phase 1 will also provide basic information to formulate statistical models for the characterization and prediction of coal quality over both sampled and unsampled areas. The following are specific purposes of phase 1.

(1) To determine the distributional properties of the variables. This is required in any statistical analysis. It allows for the selection of the most appropriate methods, and provides necessary information on the data.

(2) To determine the correlations between variables within the studied seams. This step will allow the determination of the type and strength of dependence among the coal quality parameters. Also, it allows the detection of any unrealistic relationships that exist. Such relationships would imply problems in the data.

(3) To determine the similarity of coal quality characteristics between the seams. This will allow us to detect and quantify differences between the seams based on the measured variables, either individually or collectively as a group.

Relationships between coal quality parameters and the controls on coal quality will also be investigated. A subproject currently underway is the comparison between petrographic, palynologic and chemical analyses of seam 2 at the Highvale Mine. Results of this study will lead to a better understanding of the relationship between coal quality and the environment of deposition, during the time of peat accumulation.

DATABASE

The database for the project was provided by TransAlta Utilities (TAU). Out of approximately 1700 drillholes in the TAU database, 315 contain coal quality information within the Highvale Minesite. The data was delivered on magnetic tape (1600 BPI) in the TAU "Verify" format (see appendix 1). For each of the 315 holes, the following information was supplied: the location, the drillhole identifier, the lithology, and, depending on the type of analytical data available, one or more of the following analyses: proximate, ultimate, elements in the ash, ash fusibility, equilibrium moisture, hardgrove grindability index (HGI), bulk density, and trace element concentrations. Data from the tape was extracted and stored on the Alberta Research Council VAX 8600 computer system. Separate files were created for each data subset (see table 1).

Screening of the data

Screening of the TAU data was required to produce meaningful data subsets and to isolate erroneous values. Other problems associated with the data included duplicate codes. In some records the same coding scheme was used to represent zero values, missing data and data field overflow. For certain records the stratigraphic codes were missing and for others the sample type codes (viz. Composite (C), Detailed (D) and Segmented (G)) were missing. Due to these problems, statistical analyses were performed on a data set from which the problematic entries had been removed.

Selection of the coordinate system

Two coordinate systems were provided in the TAU database: the Dominion Land Survey System (Township, Range, LSD) and a mine coordinate system. The mine coordinate system was more reliable and accurate, and will be especially useful for geostatistical studies. The origin (or zero point) of the mine grid is located at the northeast corner of Section 24, Township 52, Range 5 west of the 5th Meridian. Mine grid coordinates are expressed in metres, using positive and negative northings and eastings from the origin. An added advantage is that most of the information related to the geology of the area such as mining data, engineering data, mine planning, and pit mapping are based on the mine grid.

Proximate and ultimate analysis

Because coal is a heterogeneous mixture of organic compounds, moisture and mineral impurities, its composition determines its behaviour when used. Two sets of analytical data commonly used to determine coal quality are proximate and ultimate analyses.

The primary standard for grading the quality of coal is proximate analysis. Proximate analysis estimates the composition of the coal as percentages (by weight) of four main constituents: moisture (H_2O), fixed carbon (FXC), volatile matter (VOLM) and ash (ASH). Analyses are presented on an air-dried basis (AIR), which includes the residual moisture content and a moisture-free (DRY) basis. Moisture constitutes a significant portion of subbituminous coals and is retained primarily within the porosity. Moisture consumes energy in the powerplant operation, and it affects the handling and pulverization properties of the coal. Fixed carbon and volatile matter represent mainly the organic fraction of coal; ash is the residue remaining after combustion. These components affect essentially all powerplant systems. Heat of combustion (or calorific values) (MJKG) and total sulphur (S) are commonly included with the proximate analysis. These are used to measure combustion performance and potentially harmful emissions. In a general sense, the best thermal coals contain relatively high levels of fixed carbon and volatile matter, and relatively low levels of sulphur, moisture and ash.

Ultimate analysis is commonly performed on coal to determine the percent by weight of carbon, hydrogen, nitrogen, oxygen and sulphur. When ash content is added to the sum of these elements, the result will total to 100%. Ultimate analyses for different seams or zones are commonly compared on a dry, mineral-matter-free (DMMF) basis. Parameters commonly used in describing organic materials in the coal are the oxygen to carbon (O/C) and hydrogen to carbon (H/C) ratios (Tewalt, 1986).

Data records for other coal quality aspects such as trace element concentrations, equilibrium moisture, Hardgrove Grindability Index, specific gravity and bulk density are also important coal quality parameters. Many of these parameters affect problems associated with emissions, corrosion, slagging and fouling. Unfortunately, because of the limited amount of data for these variables in the Highvale database, not many statistical analyses can be performed at this time.

Selection of subsets for statistical analysis

Three sample types are contained in the database, namely Composite (C), Detailed (D) and Segmented (G). To satisfy the requirements of homogeneity, data sets were separated into these 3 groups. Composite samples, which represent analyses from the entire seam, were found to be the most common sample type and the most numerous in the database. Detailed samples were available only for a small area of the minesite. For D samples, however, the uppermost and lowermost 15 cm of the seam and every 30 cm of the middle portion were analyzed separately giving information of the vertical variation in coal quality. A study comparing the results of the C and D type samples would be useful in the second phase. Only a few G type samples were available so there were not enough data for statistical analysis.

Composite samples were further divided into air-dried (AIR) and oven-dried (DRY) sample types. Air drying is the partial drying of the sample to equilibrate with the room atmosphere, in which sample reduction and division take place. The sample is then oven-dried and reanalyzed as a DRY sample (the sample is dried until weight loss is not more than 0.1 percent per hour). Both types of analyses were done for the proximate data in the TAU database. In other databases such as the Energy Resources Conservation Board (ERCB) coal database, however, proximate analyses are mostly available on a DRY basis. In the regional study planned for phase 2, DRY sample analyses will likely be preferred for comparison.

AIR and DRY samples were further subdivided into six final subsets, one for each coal seam. The result is a series of analyses for Seams 1 through 6 for AIR and DRY samples.

Similarly, only the C samples were selected from the DRY ultimate data for statistical analysis.

GEOLOGY OF THE ARDLEY COAL ZONE

Regional overview

The Ardley coal zone is present throughout the Alberta plains, from Township 30 in the south to Township 66 in the north and from the outcrop edge to the deep subsurface (Richardson et al., 1988). It contains some of the thickest and most continuous coal seams in the province. Economically, it is one of Alberta's most important coal sources of thermal coal.

Based on the nomenclature proposed by Irish (1970), the Ardley coal zone lies near the base of the Paskapoo Formation and makes up the upper portion of the Scollard Member (see figure 2). The Battle Formation which lies at the base of the Scollard Member is a reliable datum which is often used to correlate Ardley coals, and to separate the similar lithology of the upper Horseshoe Canyon Formation from the lower Paskapoo Formation. Further details of the stratigraphy are available in Richardson et al. (1988), Holter et al. (1975) and Irish (1970).

Deposition of the Paskapoo Formation was in a foreland basin, (Beaumont, 1978) formed between the front of the Western Cordillera and the Precambrian Shield. Active thrusting in the Cordillera, during the late Mesozoic and early Tertiary, resulted in a downward flexure of the Alberta plains and foothills, and an asymmetric basin dipping towards the mountains developed. This tectonic setting and the variation in subsidence regimes throughout the basin are important controlling factors on the development of the Ardley coal zone as Richardson et al. (1988) point out. Paleomagnetic studies suggest that the paleolatitude of the central Alberta Basin during the Paleocene (time of deposition of the Ardley coal zone) was about 60 degrees north (Irving, 1979). Because of the absence of a polar ice cap at that time, the climate of the Alberta region was probably temperate.

Paleocene sediments of the Alberta basin are entirely non-marine in nature. The nearest marine sediments of this age belong to the

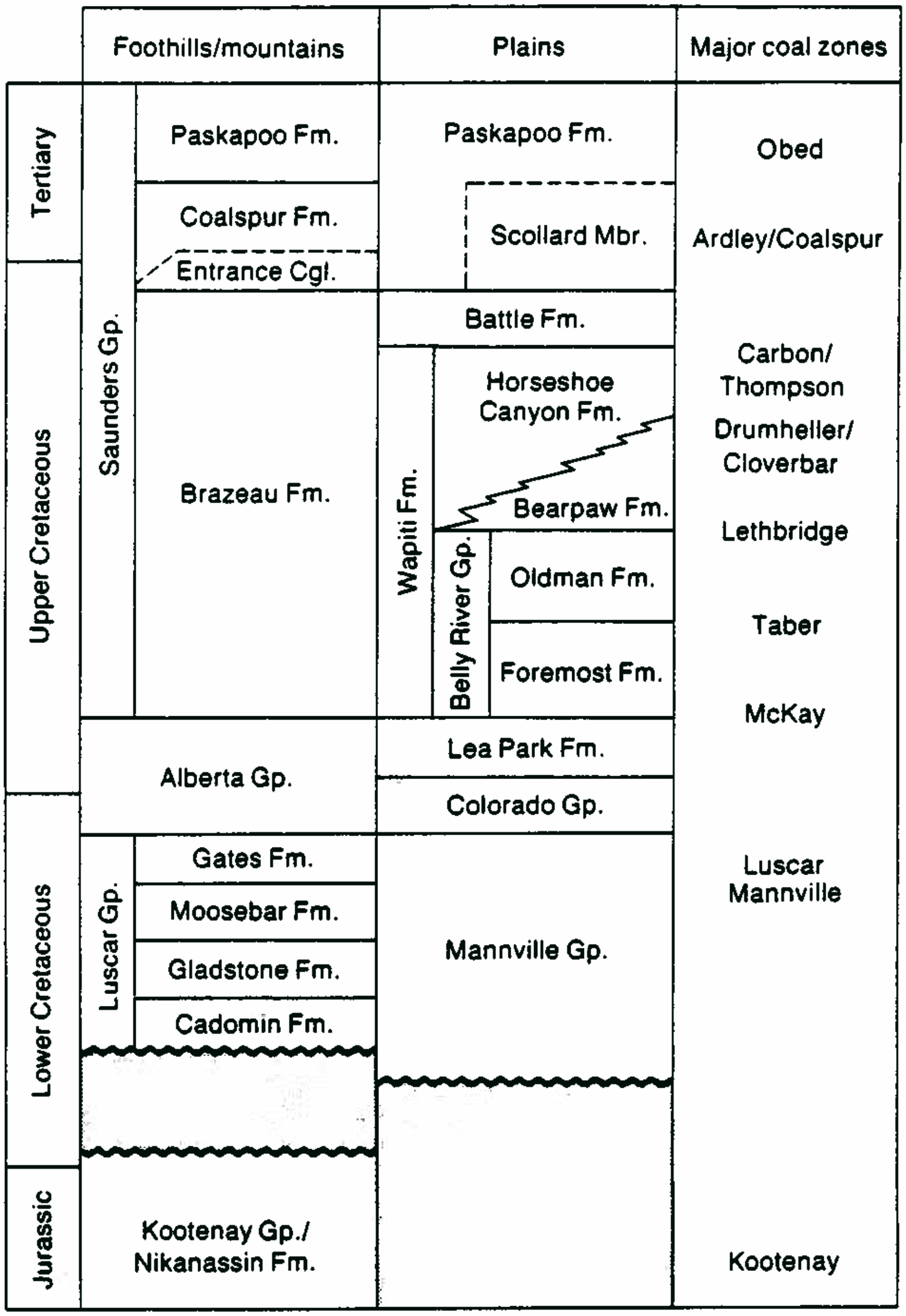


Figure 2. Table of coal-bearing formations and coal zones in Alberta. The Ardley coal zone is Tertiary in age and represents one of the youngest coal deposits in the province.

Cannonball Member of the Fort Union Formation in North Dakota. This means that Ardley coals formed at least 800 km inland from the nearest seaway. Most studies of the Ardley suggest that the environment of deposition was that of an alluvial plain dissected by a series of meandering rivers. When looking at the coal quality aspects of the Ardley zone, we will consider the regional tectonic setting, the effects of fluvial activity, and the geometry, lateral extent, and facies of the peat-forming mires that produced Ardley coals.

Geology of the Highvale area

The Highvale Mine is situated along the outcrop edge of the Paskapoo Formation (figure 3). At this location, the Ardley coal zone is mined at depths ranging from 20 to 80 m. Over 1000 development and exploration holes were drilled, 546 of which contain a full suite of geophysical logs. The TAU database contains over 35 000 picks, defining (a) the type and thickness of overburden, (b) marker beds such as bentonites above, below and within the coal zone, and (c) the top and base of each coal seam. Isopach maps of the mineable seams produced by Monenco Limited were used to determine lateral extent and thickness variation of each seam (Lyons et al., in prep.). The maps are based on a 50 m grid spacing for the northern half of the Highvale property, and a 100 m grid spacing for the southern half. It is estimated that 500 or more drillholes were used for contouring, offering a high degree of confidence in these maps. The drillhole density was higher in the northern half of the property which allowed for a closer grid spacing.

Description of seams in the Highvale area

Five seams are mined in the Highvale area (see figure 4). These seams are continuous and are easily identified. A generalized cross-section (figure 5) shows the consistency in thickness of the seams, and the unique geophysical log signatures. The seams are essentially flat lying, and dip gently to the southwest at 6.5 m/km. It should be noted that although the thicker seams, such as Seams 1 and 2 are continuous and contain only minor partings within the Highvale

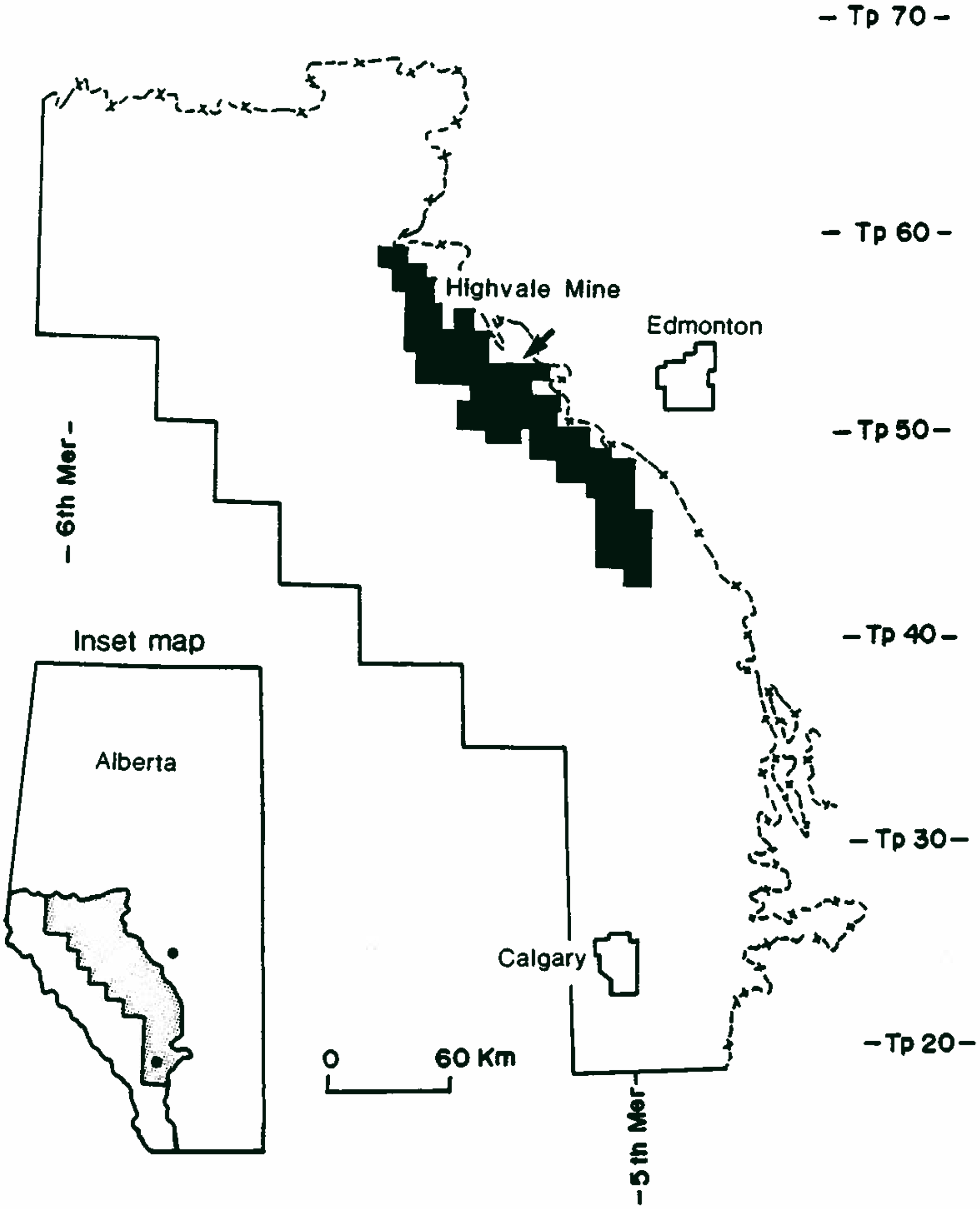


Figure 3. Geologic map of the plains region of southern Alberta. The Mayerthorpe, Wabamun and Wetaskiwin coal fields, in the northern, central and southern parts, respectively, are shown. Near-surface Ardley coals are present at these locations.

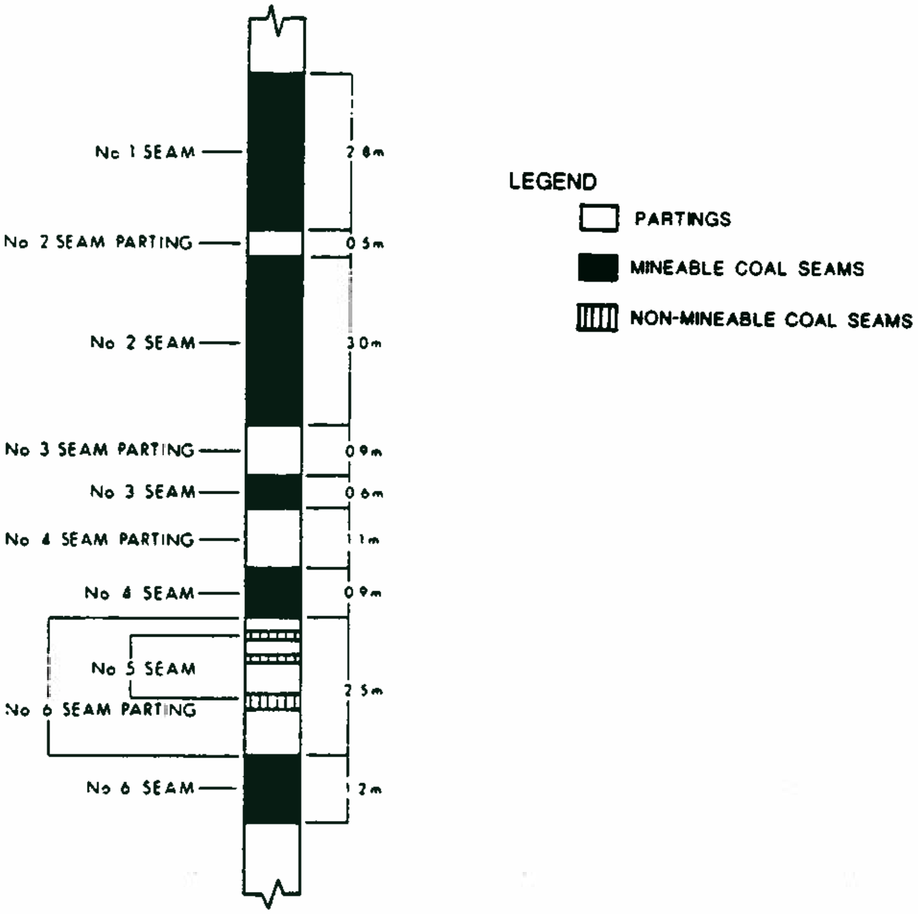


Figure 4. Numbering system and average thicknesses of the Ardley coal seams in the Highvale Mine. Seams are numbered 1 through 6, from the top to base, respectively (after Taylor, 1985).

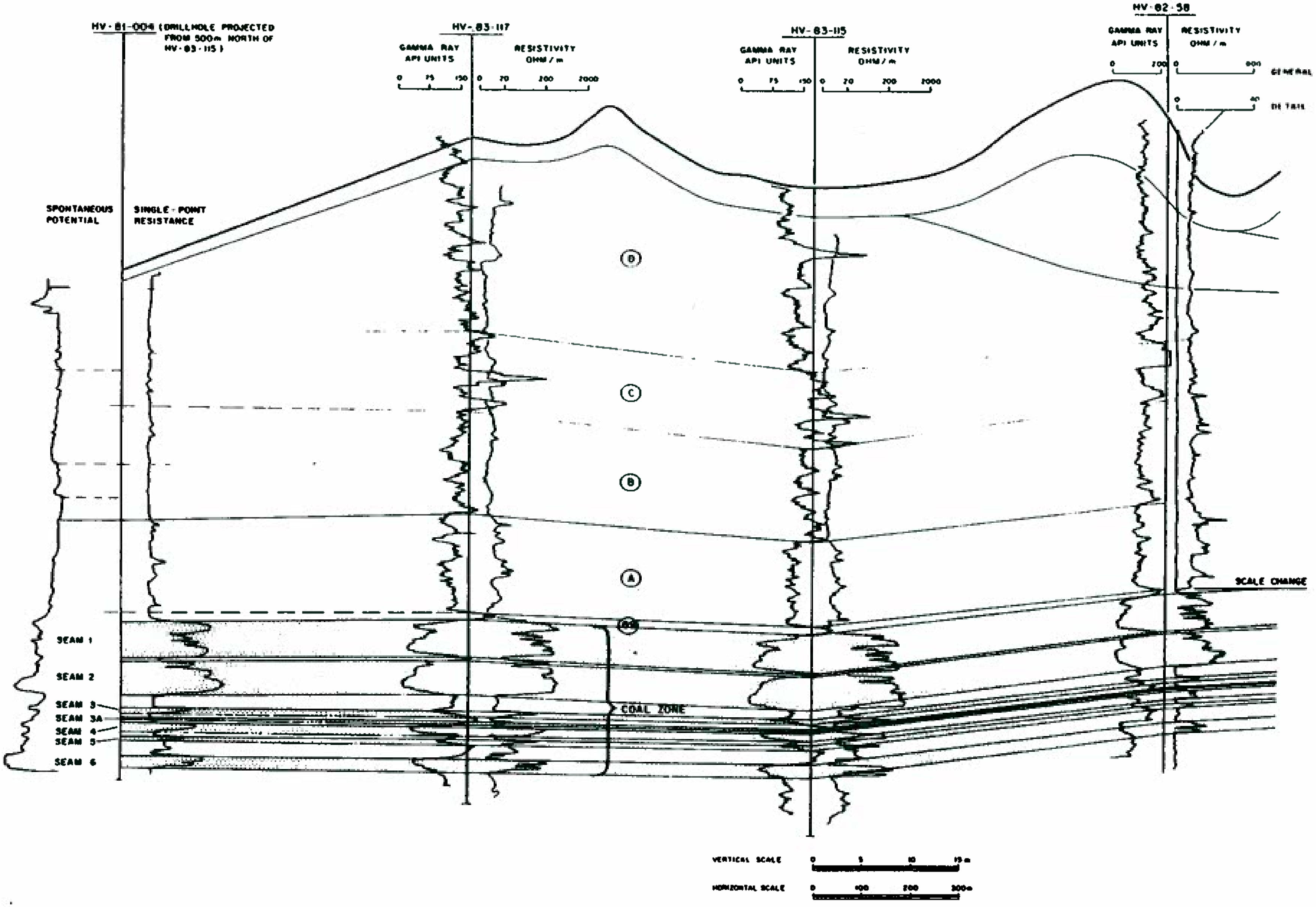


Figure 3 Generalized cross-section showing the major seams in the Highvale Mine. The distinct geophysical log responses and wide lateral continuity of the coal seams are highlighted (modified from Lyons et al., in prep.).

minesite, correlations made to the nearby Whitewood and Genesee Mines, to the north and south respectively, indicate that these seams thin and split regionally. Seam 5 is not mined at Highvale, and is considered waste material between the major economic seams. No isopach map is available for Seam 5, but abundant coal quality analyses were available so they were included in this study.

The following is a brief description of the six coal seams. For a summary of each seam and the nature of the parting between seams, see table 2.

Seam 6

Seam 6 is the lowest seam in the sequence. The average thickness in the northern half of the Highvale area is approximately 1.2 m. The seam progressively thins to 0.7 m in the southern half (figure 6). The seam appears to discontinue immediately south of the Highvale Mine limits, but exists again in the Genesee area.

Seam 6 consists of a lower and an upper subseam separated by a laterally continuous 5 to 6 cm bentonitic parting (figure 7). The bentonitic parting is included with the coal during mining, so relatively high ash contents are associated with composite samples of this seam. The apparent increase in ash content to the south may be attributed to the thickening of the upper subseam towards the north. The ash content in the south may appear higher than in the north because of higher parting to coal ratios.

The bentonitic parting and the unique geophysical log response associated with Seam 6 is particularly good for regional correlation. Based on these characteristics, Seam 6 equivalents are found in localities as far south as the Red Deer area and as far north as Fox Creek. In addition, palynologic and radiometric analyses suggest that the Cretaceous/Tertiary boundary commonly occurs along or near its base. Within the Highvale minesite, Seam 6 is often used as a datum and wells are commonly drilled 5 to 10 m below its base.

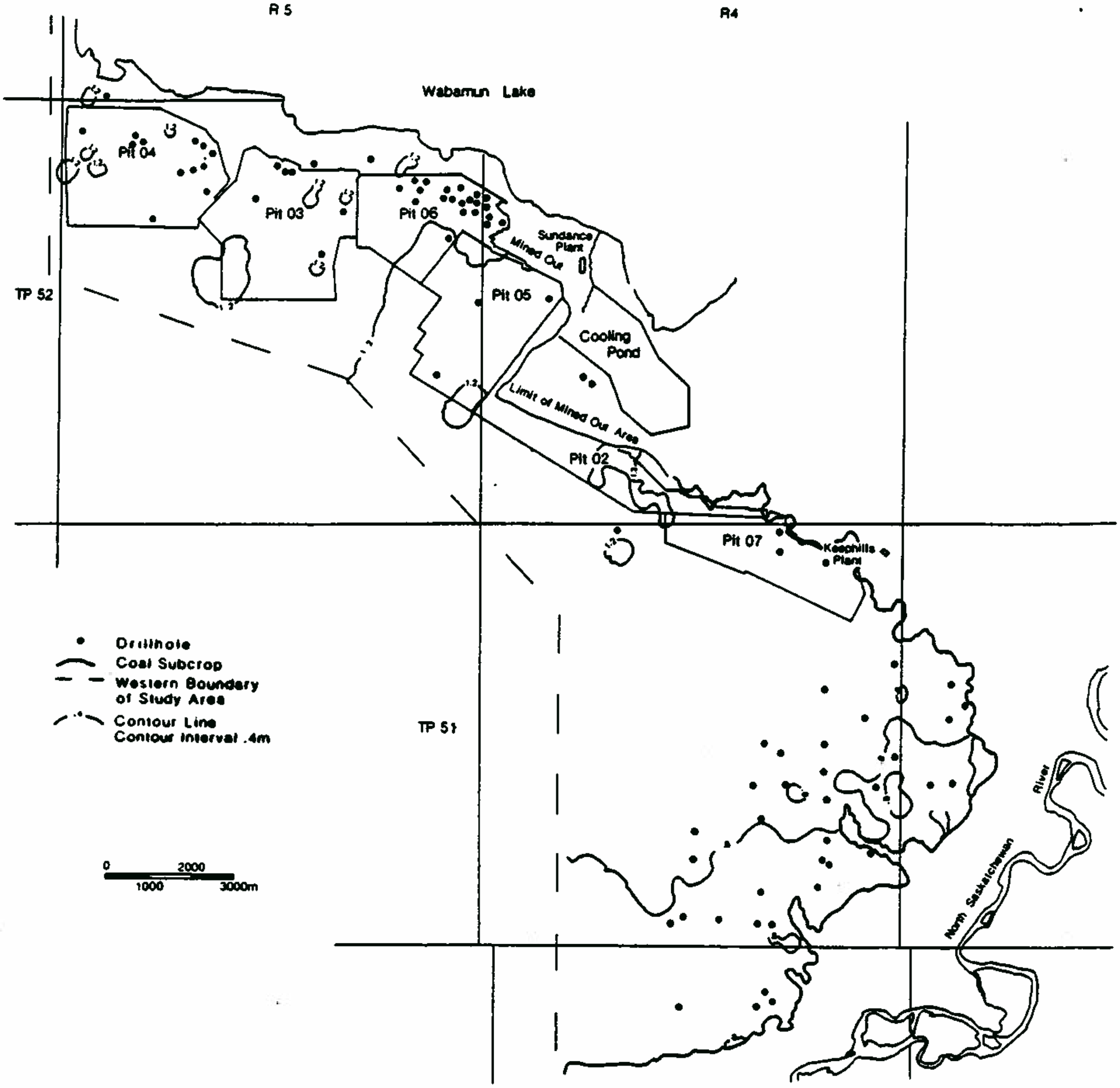


Figure 6. Isopach map showing the variation in thickness of Seam 6 in the Highvale Mine. Contour interval is 0.4 metres (from Lyons et al., in prep.).

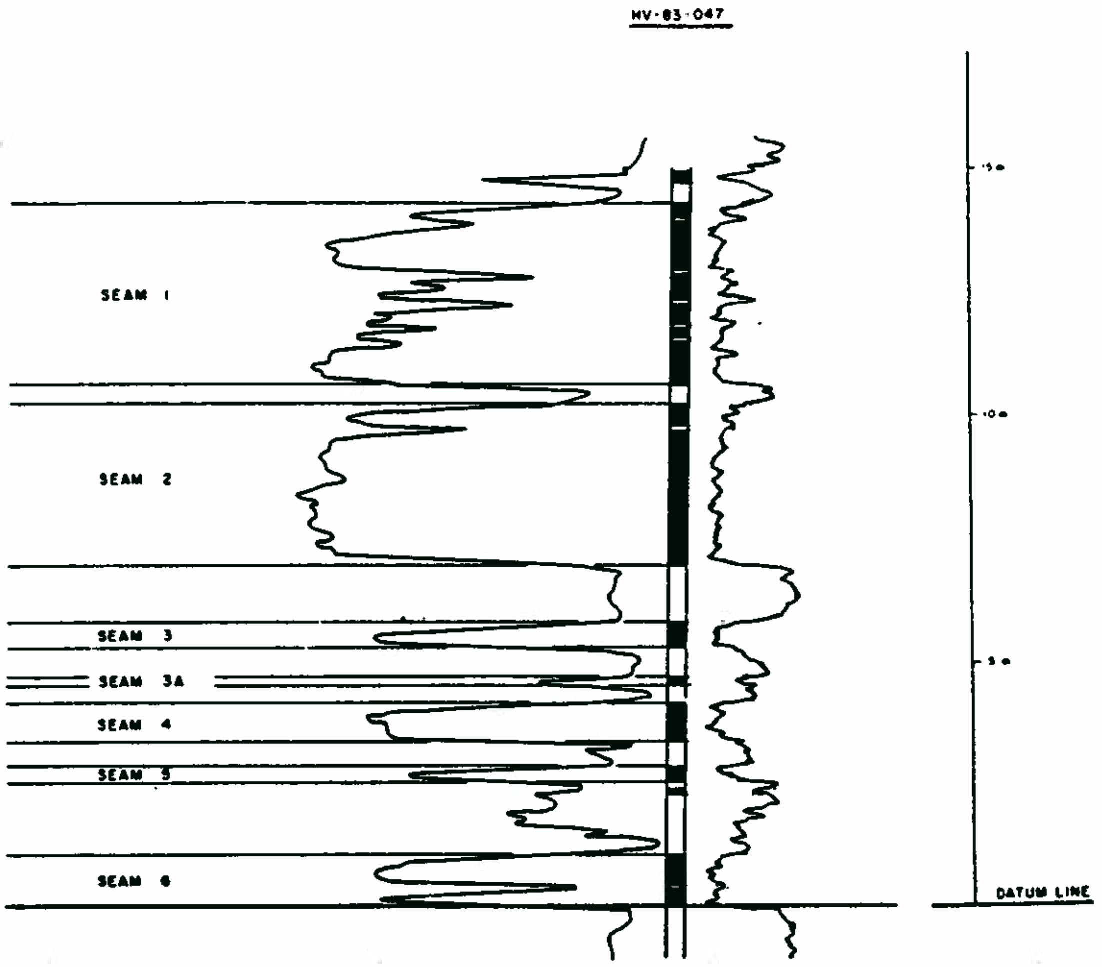


Figure 7. Detailed lithologic description of the Ardley coal zone in the Highvale Mine. The geophysical log pair shown are natural gamma ray (left) and bed resolution density (right). Black shaded areas represent coals. White areas represent partings (from Lyons et al., in prep.).

Seam 5

Seam 5 is less than 0.4 m thick over most of the minesite (figure 7). The seam has a high ash content due to the numerous partings and is not mined. Although it is not considered economic in the Highvale area, data of numerous analyses are available so it was included in the statistical study. It could be used for comparison to other seams in the mining sequence, and subsequently be compared to the equivalent seam in the Whitewood area, where it is considered to be economic and is mined.

Seam 4

Seam 4 is laterally persistent over the Highvale area and is considered economic (figures 7 and 8). In the northern portion of Highvale, it is typically 0.85 m thick and commonly contains up to 20 percent ash. In the southern portion of the Highvale, however, Seam 4 thins to 0.3 m and the ash increases to 50 percent. The seam is not considered economic in this area. Further south (i.e. outside of the mine property), Seam 4 grades into a carbonaceous mudstone. The ash estimates given are crude approximations, but the trend to higher ash in the south is significant.

Seam 3

Seam 3 is one of the most consistent and laterally continuous seams in the Highvale area. Typically it is 0.4 to 0.6 m thick throughout the Highvale Mine property (figures 7 and 9). A major split develops in the southern half of the Highvale property, where the seam is separated into an upper Seam 3 and a lower Seam 3a by a bentonic parting. The parting thickens further to the south.

Seam 2

Seam 2 is one of the two principal coal seams and in economic terms, is the most important seam in the sequence (figures 7 and 10).

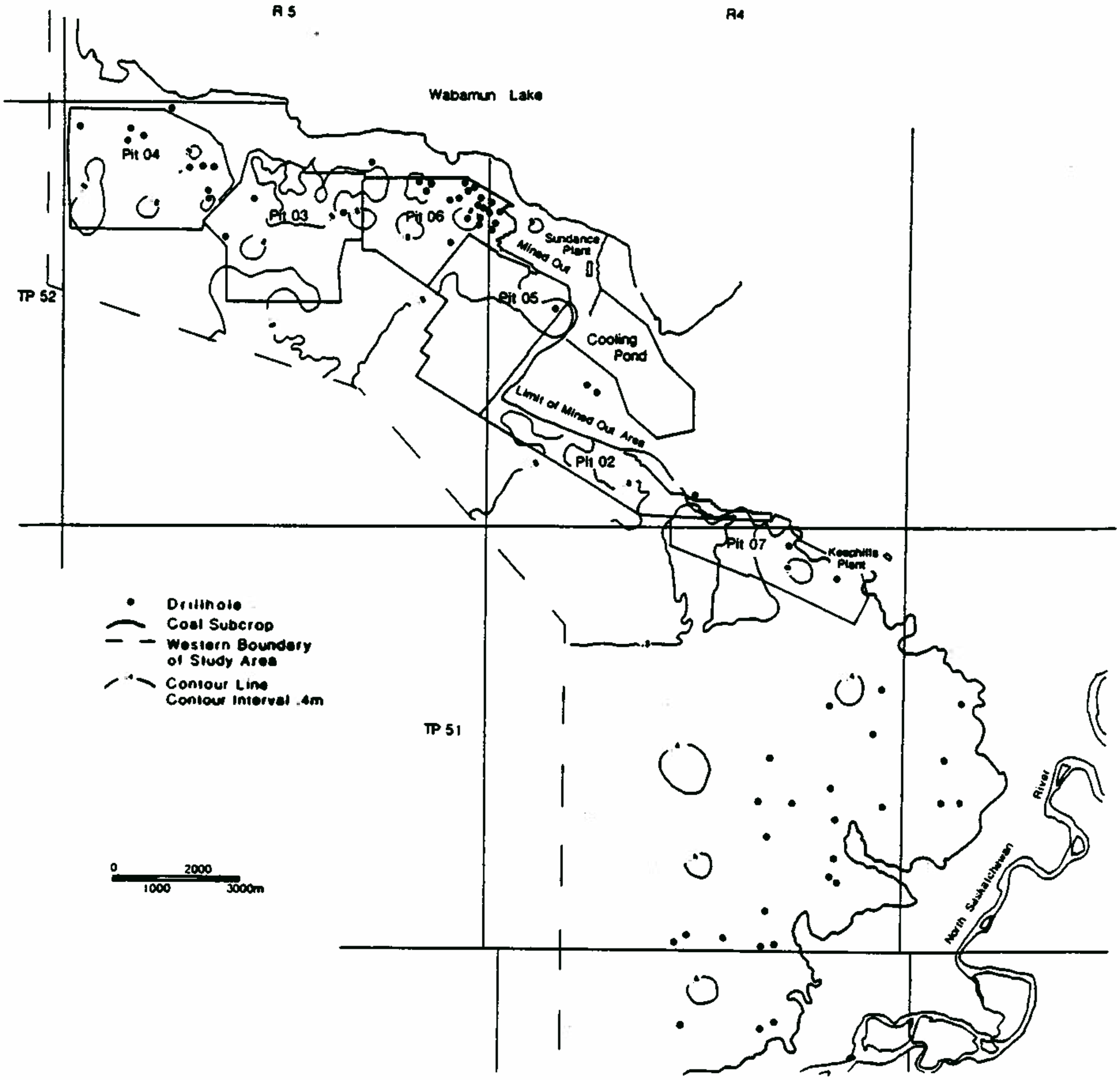


Figure 8. Isopach map showing the variation in thickness of Seam 4 in the Highvale mine. Contour interval is 0.4 metres (from Lyons et al., in prep.).

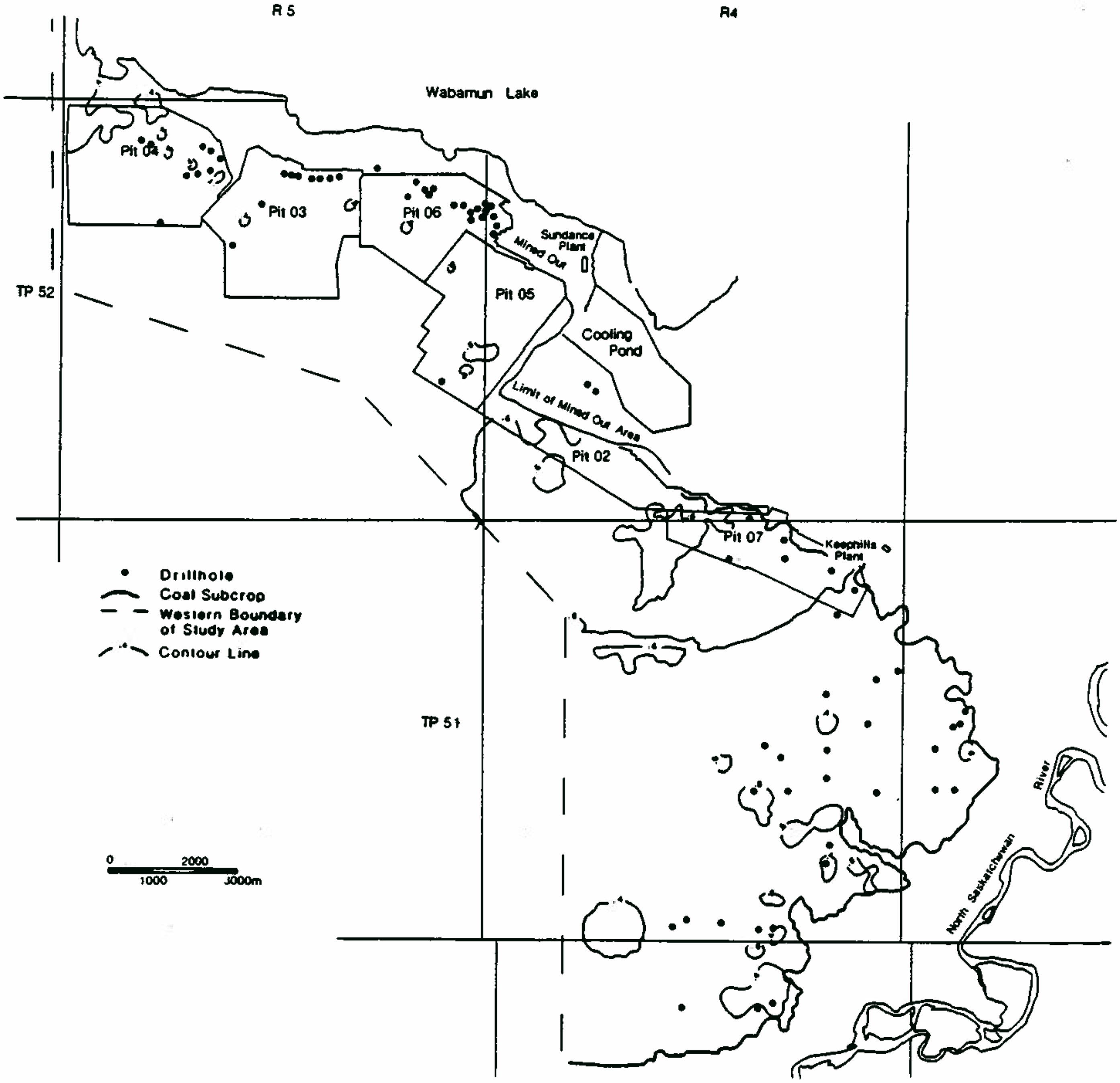


Figure 9. Isopach map showing the variation in thickness of Seam 3 in the Highvale mine. Contour interval is 0.2 metres (from Lyons et al., in prep.).

It is laterally extensive and the thickness ranges from 3.0 to 3.5 m throughout the Highvale area. No significant partings are present and it is one of the most consistent seams in terms of coal quality. A detailed study involving petrographic, palynological and chemical analyses of Seam 2 is currently underway to determine the vertical and lateral variations in the quality of this seam.

Seam 1

Seam 1 is the uppermost economic seam in the sequence. It is the second of the two principal coal seams in the Highvale Mine (figures 7 and 11). The thickness is highly variable, ranging from 1.0 to 4.0 m. This is due, in part, to local erosion of the upper portion of the seam by glacial effects and/or post-depositional channelling. One example of post-depositional channelling is a sharp linear thinning trend along the south shore of Wabamun Lake in Pit 03 (figure 11). Areas of thinning associated with glacial activity are commonly located near the subcrop boundaries of Seam 1. Excluding local erosion, an overall thinning trend of the coal seam is observed towards the southwest, due to deposit-ional controls during peat accumulation. The ash content of this seam appears to be more variable than that of Seam 2. The relatively large variation in ash (10 to 25 percent) throughout the minesite appears to be dependent on the thickness variation of the seam and/or to be caused by varying thicknesses of bentonitic and argillaceous partings.

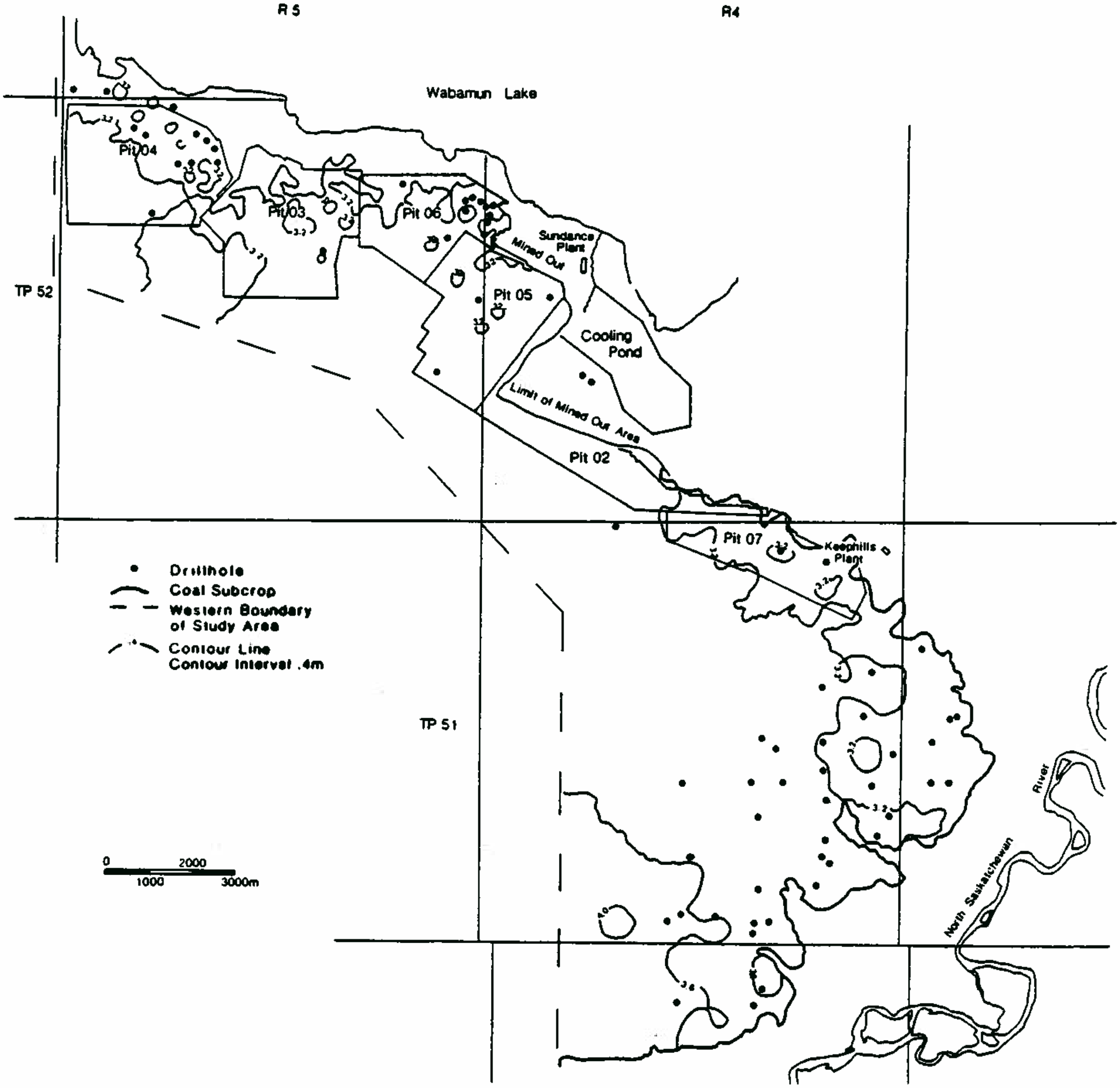


Figure 10. Isopach map showing the variation in thickness of Seam 2 in the Highvale mine. Contour interval is 0.4 metres (after Lyons et al., in prep.).

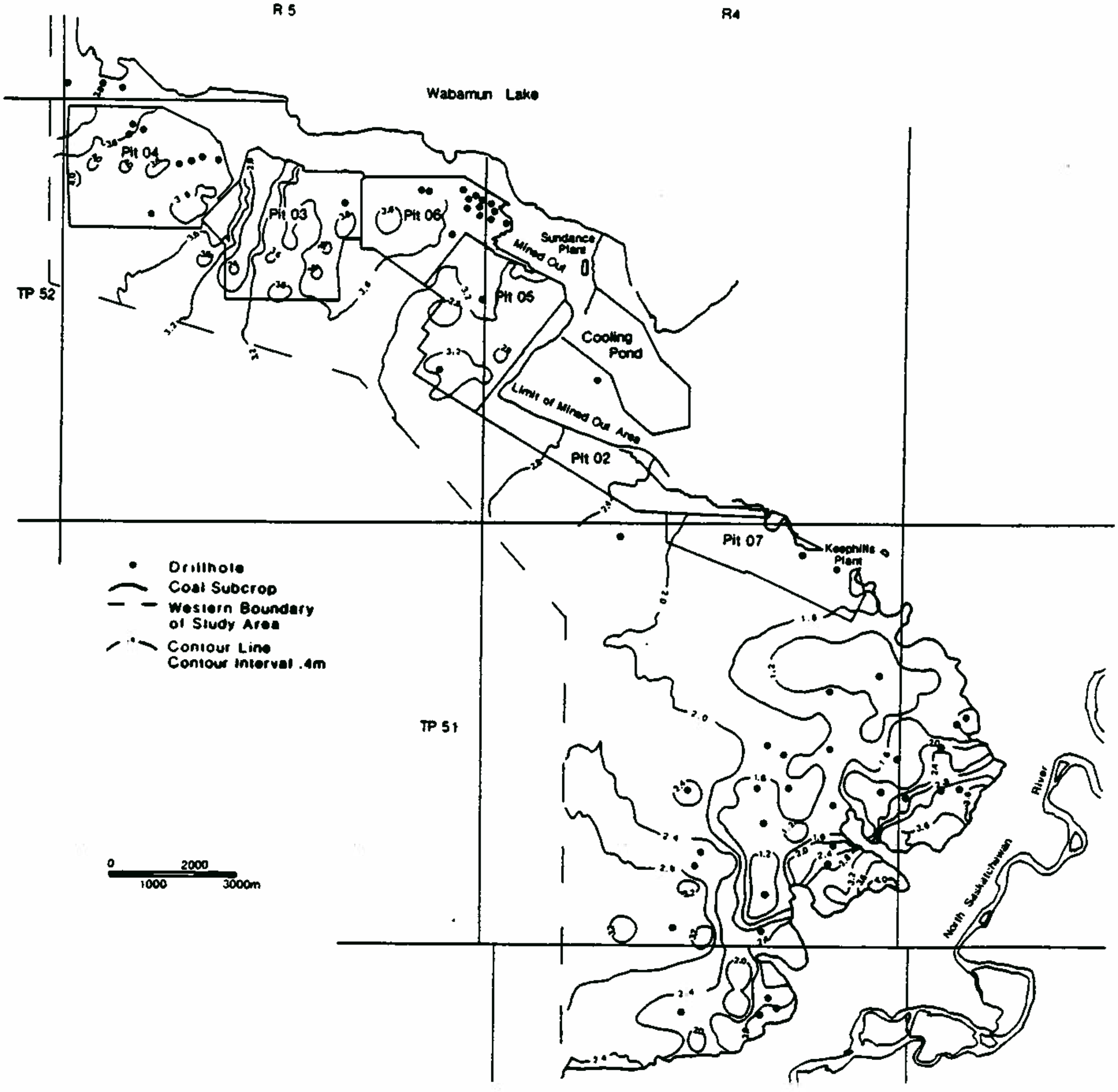


Figure 11. Isopach map showing the variation in thickness of Seam 1 in the Highvale mine. Contour interval is 0.4 metres (from Lyons et al., in prep.).

METHODS OF STATISTICAL ANALYSIS

This phase of the exploratory analysis of the Highvale data can be described in three separate categories, viz. (1) Univariate analysis, (2) Bivariate analysis and (3) Multivariate analysis. These represent, respectively, the analysis of individual variables for their distributional characteristics, the examination of pairs of variables for their relationships, and the comparisons and analyses of multivariate data. Figure 12 summarizes the methods used in the present study. The mathematical representations of the statistical methods are listed in Appendix 2. Only qualitative descriptions are given in this section.

Univariate analysis

This provides the details on the distributional characteristics of the variables observed from the samples.

The frequency distribution describes how often different values occur in the sample. The selection of proper statistical procedures for analysis is often based on the characteristics of the data frequency distribution. For example, statistical techniques that require normally distributed data are inappropriate if one finds that the frequency distribution of the data does not follow the normal (Gaussian) form. Another example is that the shape of the frequency distribution can be useful in identifying the type of data transformation suitable for a given analysis. Data anomalies, such as outliers and multiple populations, can also be observed.

For the purpose of visual comparisons, histograms and multiple box plots were used. A box plot is a simple diagram showing the maximum, minimum, the 75th percentile, the median, and the 25th percentile. The percentile is the value which is not exceeded by a certain portion of the sample. For example, the 25th percentile is the value of the variable not exceeded by 25 percent of the sample; the 30th percentile is that not exceeded by 30 percent of the sample, and so on.

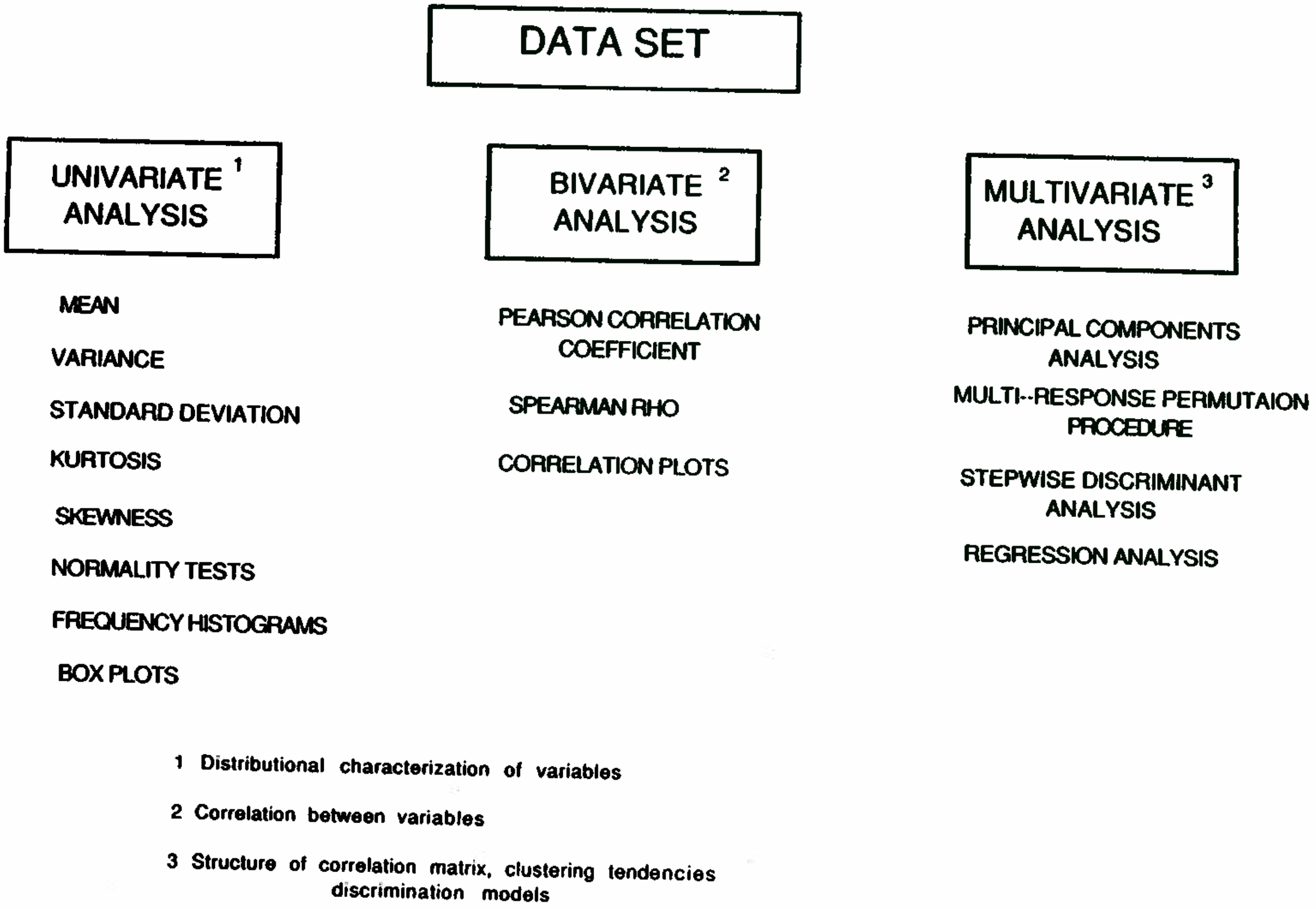
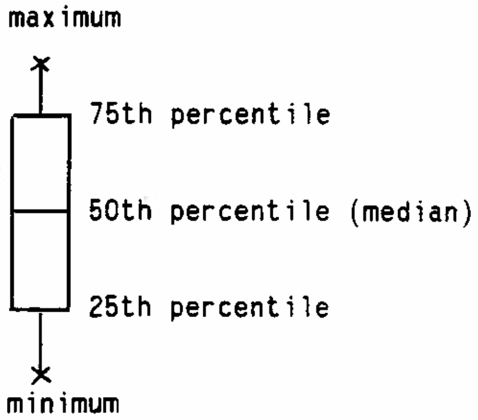


Figure 12. Methods of statistical analysis used in this study. For a description of each method see text.



Hence, when multiple box plots are used, the basic characteristics of the distributions of a set of variables can be summarized visually. In the present study, a multiple box plot was used for each variable from all six seams.

The statistical parameters computed include:

(1) The sample mean.
(2) The sample median.
(3) The sample mode.
(4) The sample variance.
(5) The coefficient of variation.
(6) The range.
(7) The interquantile range.
(8) The skewness coefficient.
(9) The kurtosis coefficient.

It is useful to compute some kind of centred value from a sample so that an estimate of the same for the data population, can be obtained. The mean, the median and the mode all are measures of central tendency. The mean or average of the sample values is usually the best estimate of the mean of the population from which the sample is drawn. The median is the halfway point in a set of values when they have been arranged in order. It is also known as the 50th percentile. Hence, the sample median is the 50th percentile of the data sample, estimating the 50th percentile of the data population. The population median and mean may not be equal. In the case of a symmetric probability distribution such as the normal (Gaussian) distribution, where the two are equal, the sample median may be used to estimate the population mean. The median is not as good an estimate for central tendency, as the sample mean, in the sense that it does not make use of the maximum amount of information from the data sample. However, the median is less affected by outliers, and therefore, a better measure of central tendency than the mean for skewed data distributions. A mode is a peak value of the frequency distribution. A single mode of relatively high frequency gives a quick estimate of central tendency, in the form of the most frequently-occurring value of the sample obtained. The mode is not a useful measure of central location when there are more modes in the data distribution, or when the mode is not clearly defined. For example, this occurs when there is a flat top, rather than a distinct peak in the frequency distribution. However, the mode can be a useful indicator where there are more data populations involved. For example, a bi-modal frequency distribution sometimes indicates that the sample is from two populations.

For measuring variability, the sample variance and coefficient of variation are commonly used parameters. The sample variance is the root mean square departure from the sample mean. It is a measure of the spread of the data and the standard deviation is its square root. The coefficient of variation is the ratio of the standard deviation to the mean. It therefore compares variability to the average value of the data. Other measures of variability include the difference between the maximum and minimum values (or the range), the difference between the 75th percentile and the 25th percentile (the inter-quantile range). In the case of a normal distribution, both these ranges can be used to approximate the population variance, particularly when the sample size is small. However, they are not as good as the sample variance for estimating the population parameter of variance.

Two other measures of distributional characteristics are skewness and kurtosis. The skewness coefficient S measures the degree of asymmetry in the data distribution, and the coefficient of kurtosis (K) is a measure of the degree of peakedness in the data distribution. A normal distribution has a skewness of zero and a kurtosis of 3.

Included in the univariate analysis are also two tests for normality. The null hypothesis for both tests is that the sample comes from a normal distribution. First is the Lilliefors test (Lilliefors, 1967), which is the classical Kolmogorov-Smirnov test applied to cases with unknown (estimated) population parameters. This test was used for large samples. For small sample sizes, ($n < 50$), the Shapiro-Wilk W statistic was used (Shapiro and Wilk, 1965). For the purpose of the present study, only the results of these tests in the form of P-values are presented. The P-value is the probability of obtaining a test statistic more extreme than the one computed from the data. A small P-value indicates an acceptable level of confidence to reject the null hypothesis involved.

Bivariate analysis

The degree of linear association between a pair of variables is examined by correlation analysis. For the purpose of the present exploratory study, only linear correlation is considered. Both the Pearson product moment correlation and the Spearman rank correlation (RHO) were used. The correlation coefficients range in values from -1 to +1. The test of significance for the Pearson product moment correlation requires the normality assumption. Hence, the Spearman rank correlations were also computed since for many cases, normality cannot be assumed. The tests of significance associated with the Spearman rank correlation are distribution-free (i.e. no underlying distributional assumption). It is useful to compare the results of the Spearman rank correlation with those of the Pearson product moment correlation.

A useful tool of bivariate analysis is the multiple scatterplot. The scatter diagrams for the various variable pairs are plotted together to present an overall visual description of the different relationships. It is arranged like a correlation matrix, except that the correlation coefficients are replaced by the respective scatter diagrams. Scatter plots are useful in many ways. They are the logical comparisons for regression and correlation analysis. The pattern of points plotted are visual representations of the strength of association, shape of the bivariate data distribution, the existence of outliers and the appropriateness of the linear model.

Bivariate analyses are useful in identifying predictors (the independent variable in a regression model). Although for the present study, most of the variables analyzed will be predictands (the dependent variables), it is possible to formulate predictive models using the measured variables. An example is the prediction of calorific value of coal from the different seams.

Multivariate analysis

A major technique used in this class of analysis is a non-parametric multiple comparison method termed the multi-response permutation procedure (MRPP). The interested reader is referred to Mielke et al. (1981), which contains further references for the theoretical details of this method. Only a summary of the procedure is given here. This method involves no distributional assumptions and is for testing the degree of clustering of groups of multivariate data. In this context, distinct grouping implies significant differences between the a priori groups. The test statistic for this procedure is based on the weighted average of the Euclidean distances between data points in the multidimensional space. The rejection of the null hypothesis means that the groups are significantly different, considering the variables (responses) observed.

There are certain advantages in using MRPP in geological applications. First the method is "distribution-free". It is appropriate for data which do not follow the multivariate normal distribution. Second, it is a permutation test which does not require independent random samples. This implies that the spatial dependence of data from geological fields do not create difficulties in the test results. Third, the test is based on the Euclidean distance measure. This means that the "congruence principle" is satisfied (Mielke, 1985, 1986). Most people visualize data in the Euclidean space. The congruence principle simply states that if one analyzes data in the same manner as one perceives it, more intuitive results can be obtained.

Another multivariate technique used in the present study is principal component analysis (PCA). The principal components are basically the eigen vectors of the sample variance-covariance matrix. It is a useful variable reduction technique. A large number of variables can be represented by a few principal components, which are independent of each other. Such independence is useful in the formulation of stepwise multiple regression models, where the principal components may enter as predictors (independent variables). However, for the present study, where the variables represent coal quality, rather than predictors of coal quality, the PCA is still useful as a variable reduction technique. The purpose is to see if some meaningful principal components can be identified. The spatial distribution of such principal components may be related to geological factors.

Multivariate analyses are used in the present study to provide overall assessments of similarity between samples, based on the observed variables. The results therefore supply a different scale of information about the data set, different from that of the univariate and bivariate analyses mentioned previously. Since these are simultaneous comparisons of a number of variables, the results are dependent on the set of vaiables selected. Also, overall assessment implies less emphasis on details, which are provided by analyses at the univariate and bivariate levels. Hence, analyses from all three levels should be used together for a more complete assessment of data.

RESULTS

In this section, only the completed analyses are reported. Some of the analyses mentioned in the previous section are in the process of being performed or the results are in the process of being interpreted. These will be reported at a later date.

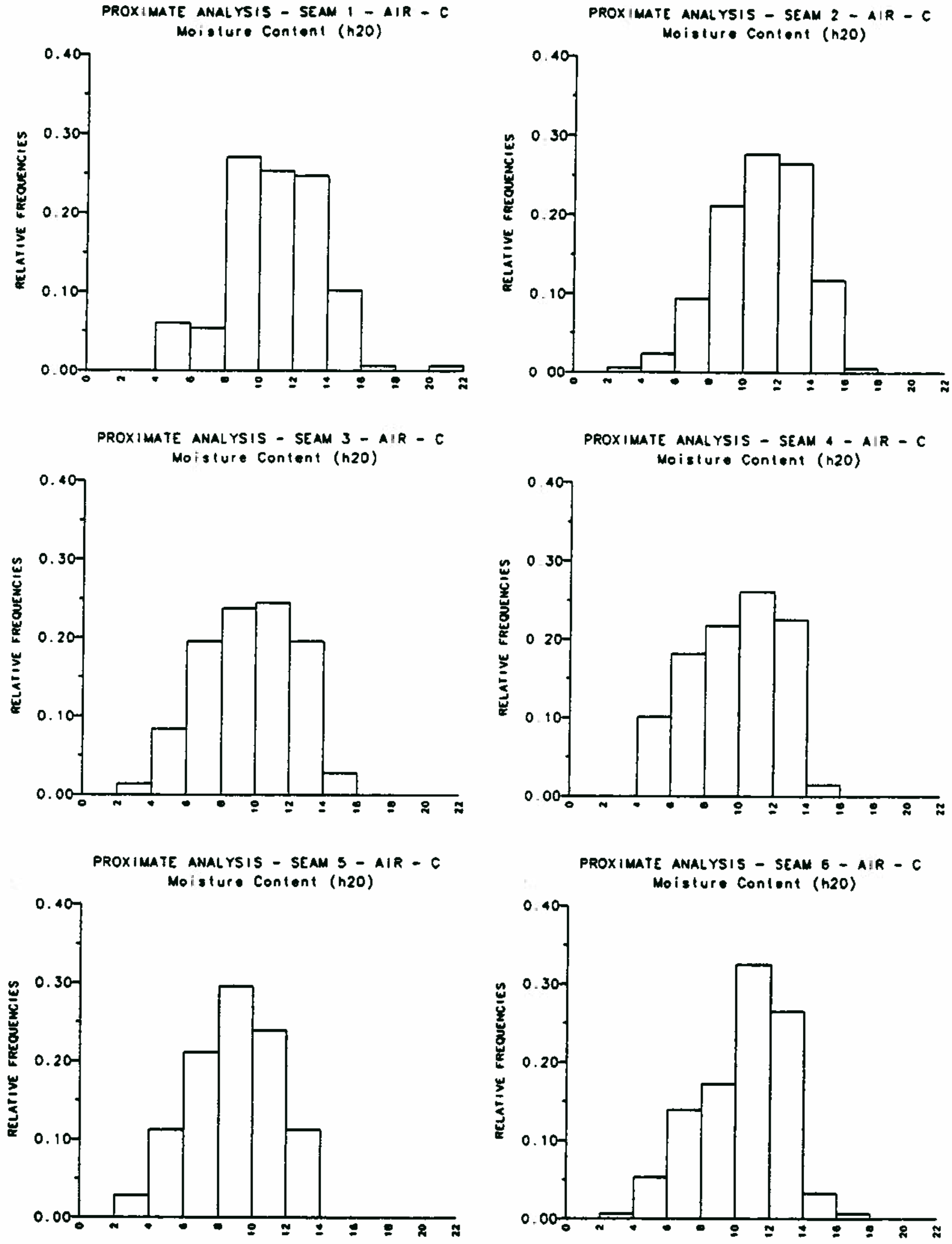
The analysis of the proximate data

In the present study, only the C samples were analyzed for the proximate group. Separate analyses were performed for the AIR and DRY categories with emphasis on the seam to seam variation of the Highvale data. The comparisons and the displays are mainly on a seam-by-seam basis.

Distributional characteristics:

Figures 13 to 23 contain histograms of the variables studied from both AIR and DRY cateogories. By plotting the six histograms from the six coal seams on the same page, interseam comparison is facilitated. These figures also provide a visual representation of the seam-to-seam variation of each of the variables involved. There are considerable differences in the central tendency and range of values from seam to seam. One general observation was that there were frequent departures from the normal distribution, which is assumed in many classical statistical techniques. This implies that care must be taken in any subsequent analysis relying on the normality assumption. Another general observation is that the skewness of the distribution for a given variable can also vary greatly (see figure 21). This means that one cannot apply any systematic technique to the data, to transform them so that normality can be assumed, for the transformation appropriate for one seam may not be so for another. It is implied from figures 13 to 23 that there is considerable non-homogeneity among the seams. This will be tested subsequently.

Meanwhile, box plots (figure 24 and 25) are used to summarize the



PROXIMATE ANALYSIS - SEAM 1 - AIR - C
Moisture Content (h20)

PROXIMATE ANALYSIS - SEAM 2 - AIR - C
Moisture Content (h20)

PROXIMATE ANALYSIS - SEAM 3 - AIR - C
Moisture Content (h20)

PROXIMATE ANALYSIS - SEAM 4 - AIR - C
Moisture Content (h20)

PROXIMATE ANALYSIS - SEAM 5 - AIR - C
Moisture Content (h20)

PROXIMATE ANALYSIS - SEAM 6 - AIR - C
Moisture Content (h20)

Figure 13. Frequency distributions of the moisture content of coal seams 1 to 6 for the Highvale area. Moisture is expressed in percent. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.

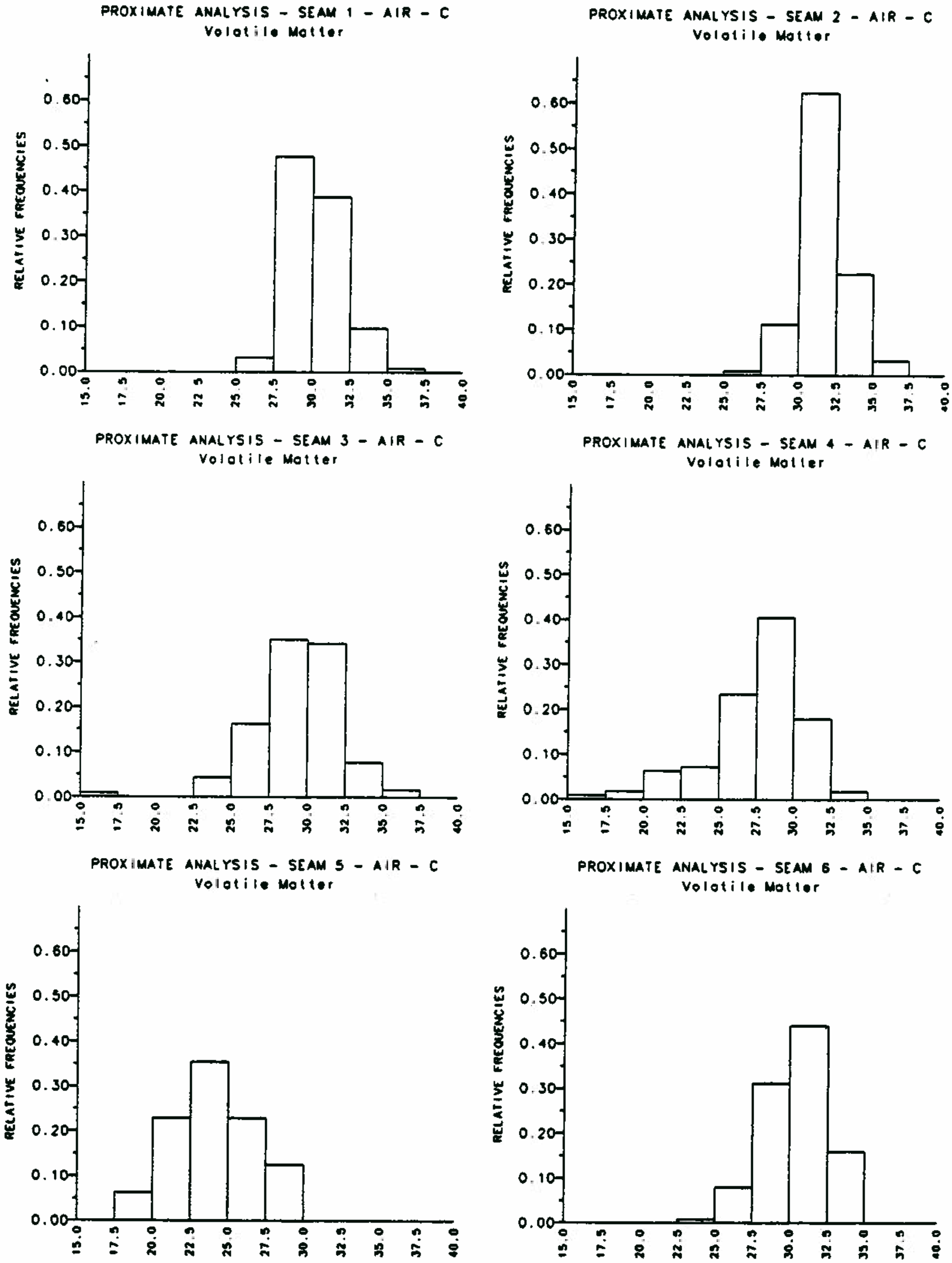


Figure 14. Frequency distributions of the volatile matter of coal seams 1 to 6 for the Highvale area. Volatile matter is expressed in percent. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.

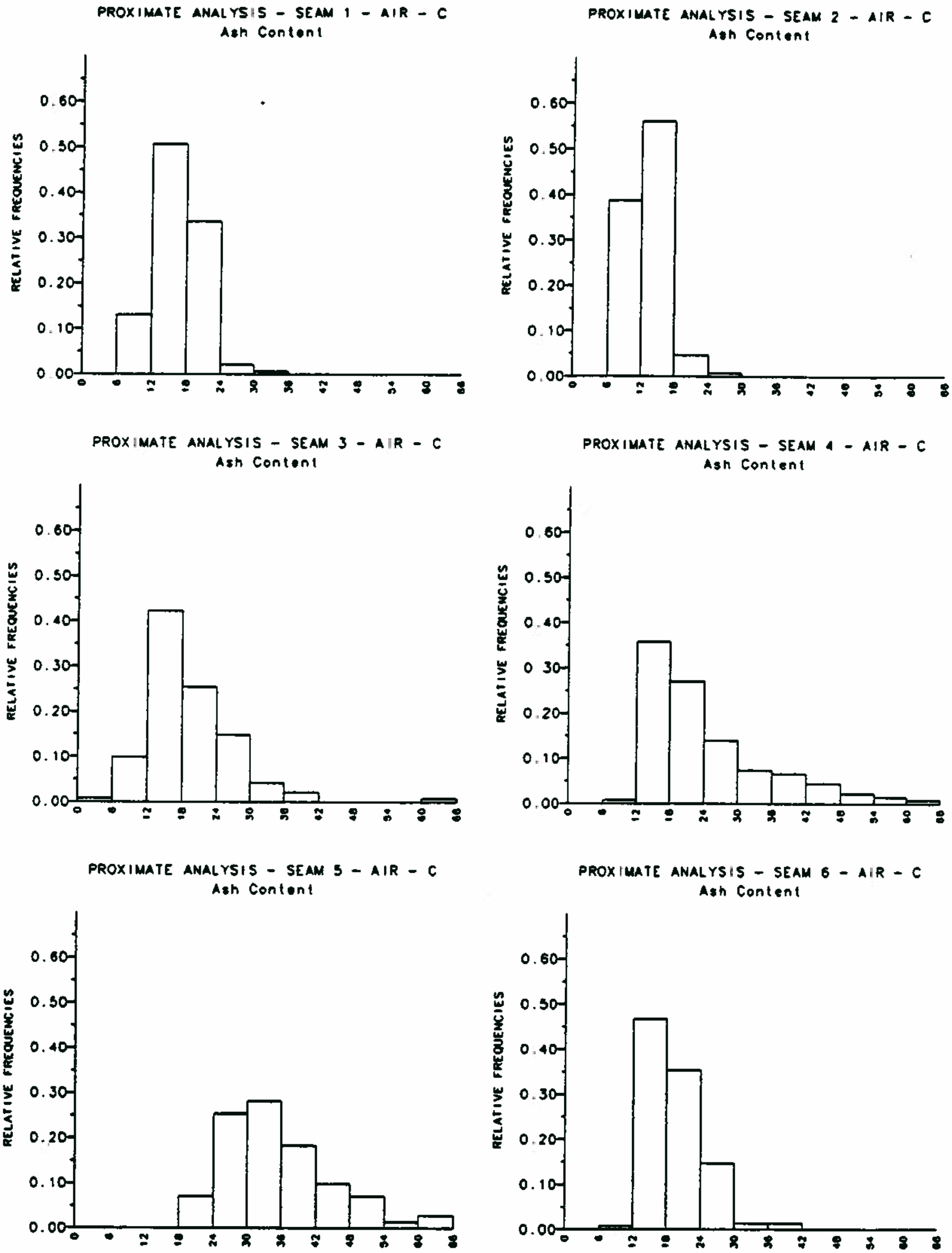
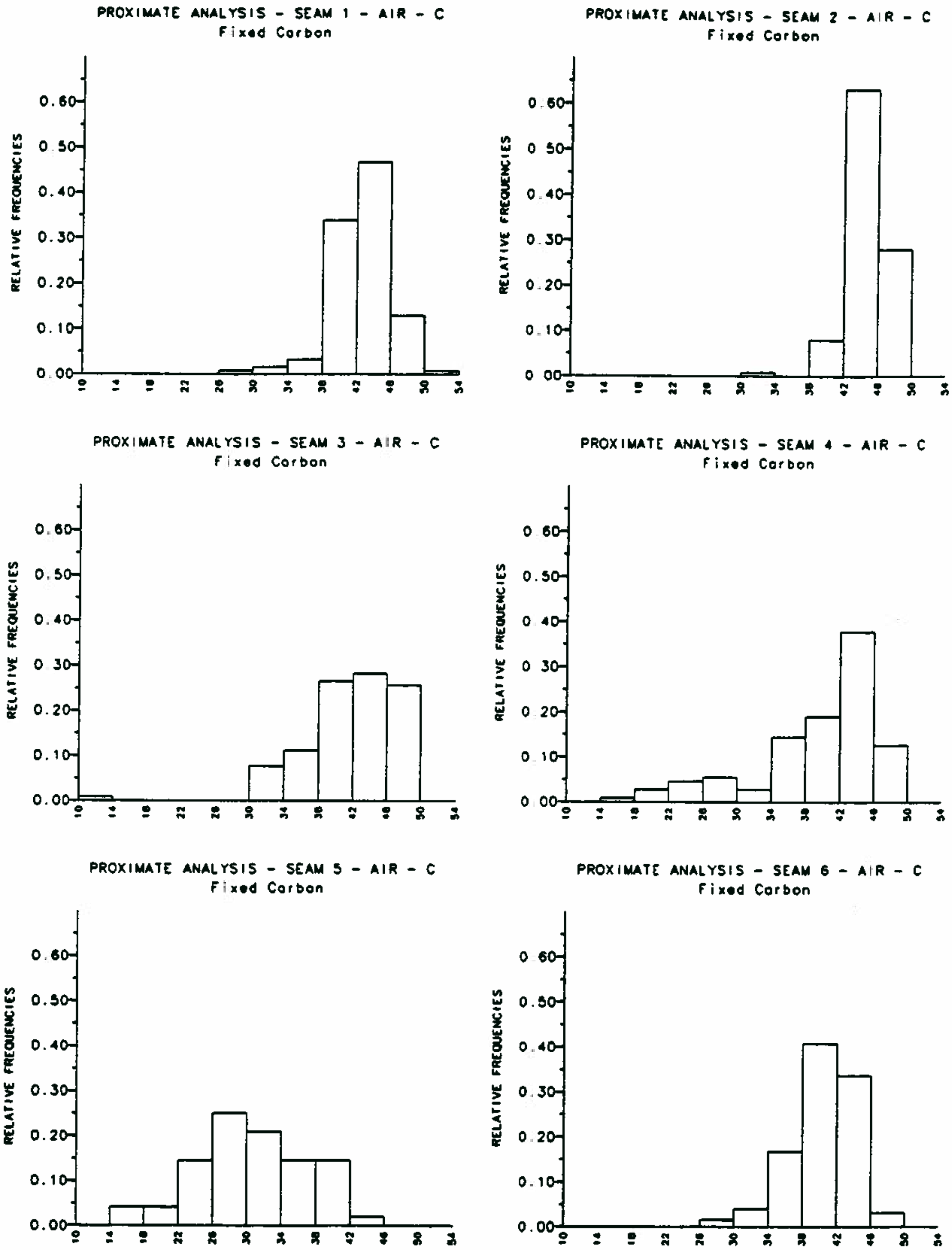


Figure 15. Frequency distributions of the ash content of coal seams 1 to 6 for the Highvale area. Ash content is expressed in percent. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.



PROXIMATE ANALYSIS - SEAM 1 - AIR - C
Fixed Carbon

PROXIMATE ANALYSIS - SEAM 2 - AIR - C
Fixed Carbon

PROXIMATE ANALYSIS - SEAM 3 - AIR - C
Fixed Carbon

PROXIMATE ANALYSIS - SEAM 4 - AIR - C
Fixed Carbon

PROXIMATE ANALYSIS - SEAM 5 - AIR - C
Fixed Carbon

PROXIMATE ANALYSIS - SEAM 6 - AIR - C
Fixed Carbon

Figure 16. Frequency distributions of fixed carbon of coal seams 1 to 6 for the Highvale area. Fixed carbon is expressed in percent. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.

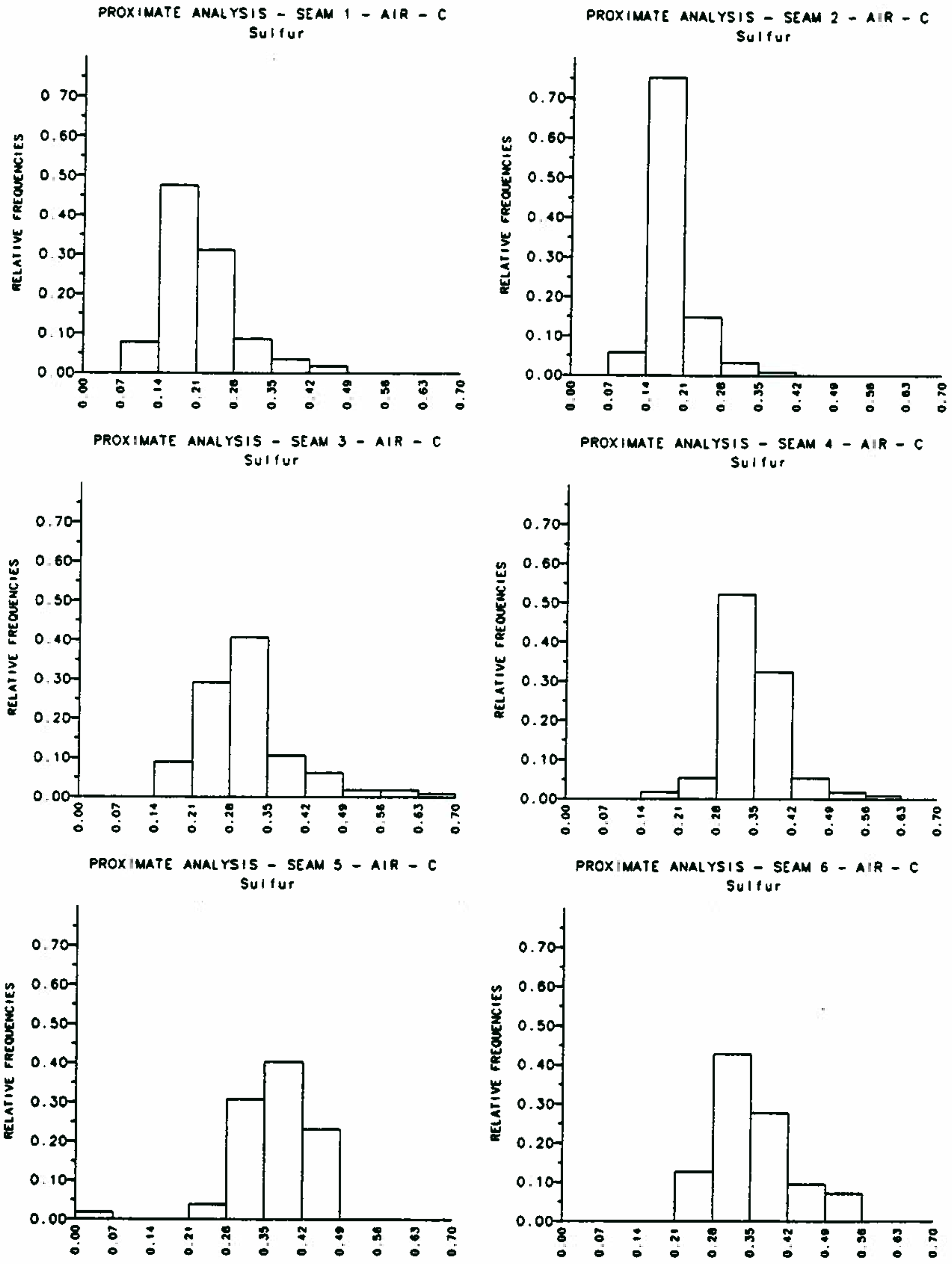
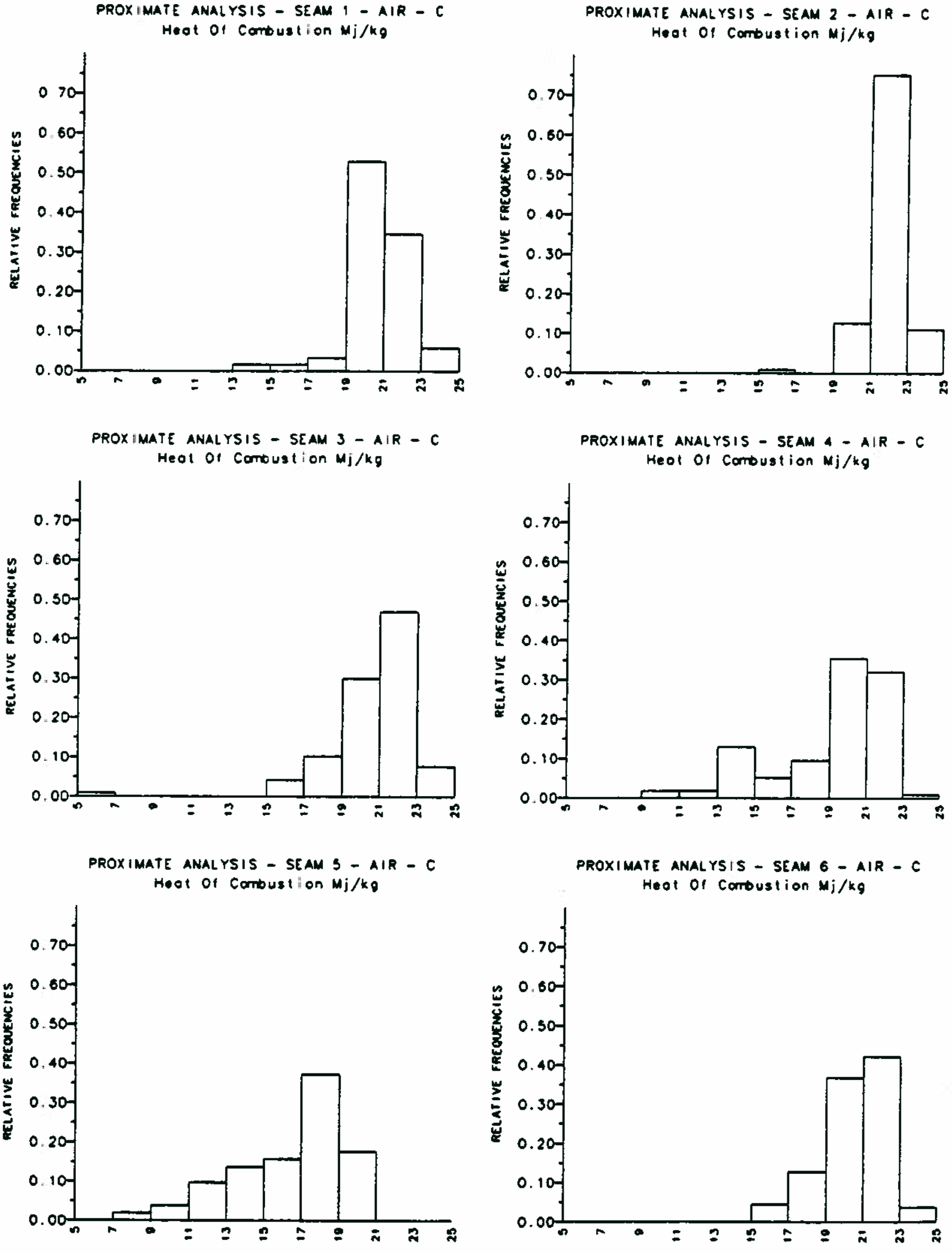


Figure 17. Frequency distributions of total sulphur of coal seams 1 to 6 for the Highvale area. Sulphur is expressed in percent. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.



PROXIMATE ANALYSIS - SEAM 1 - AIR - C
Heat Of Combustion Mj/kg

PROXIMATE ANALYSIS - SEAM 2 - AIR - C
Heat Of Combustion Mj/kg

PROXIMATE ANALYSIS - SEAM 3 - AIR - C
Heat Of Combustion Mj/kg

PROXIMATE ANALYSIS - SEAM 4 - AIR - C
Heat Of Combustion Mj/kg

PROXIMATE ANALYSIS - SEAM 5 - AIR - C
Heat Of Combustion Mj/kg

PROXIMATE ANALYSIS - SEAM 6 - AIR - C
Heat Of Combustion Mj/kg

Figure 18. Frequency distributions of heat of combustion of coal seams 1 to 6 for the Highvale area. Heat of combustion is expressed in Mj/kg. The data are from sampling category C (composite samples) on an air-dried (AIR) basis.

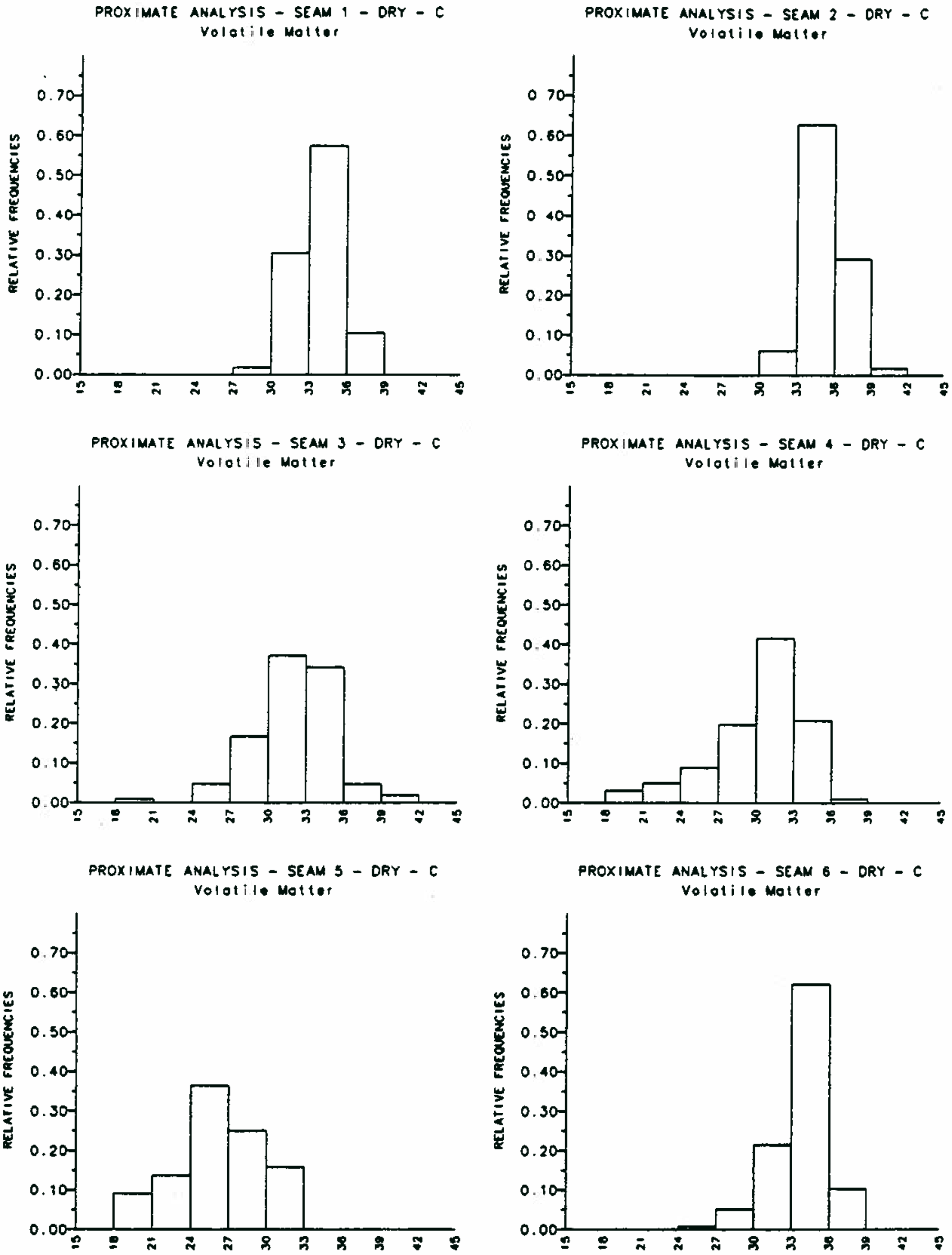


Figure 19. Frequency distributions of volatile matter of coal seams 1 to 6 for the Highvale area. Volatile matter is expressed in percent. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis.

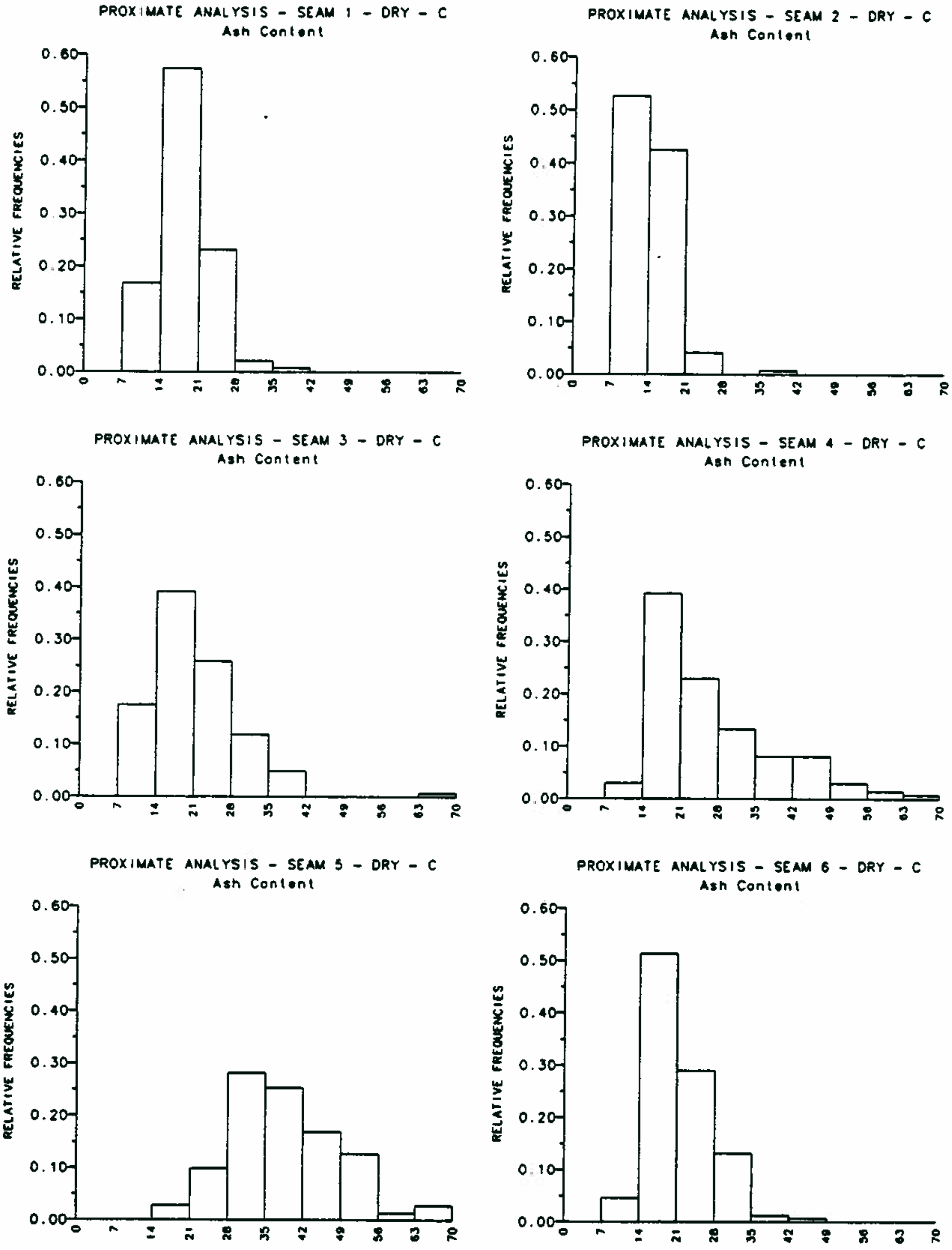


Figure 20. Frequency distributions of ash content of coal seams 1 to 6 for the Highvale area. Ash is expressed in percent. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis.

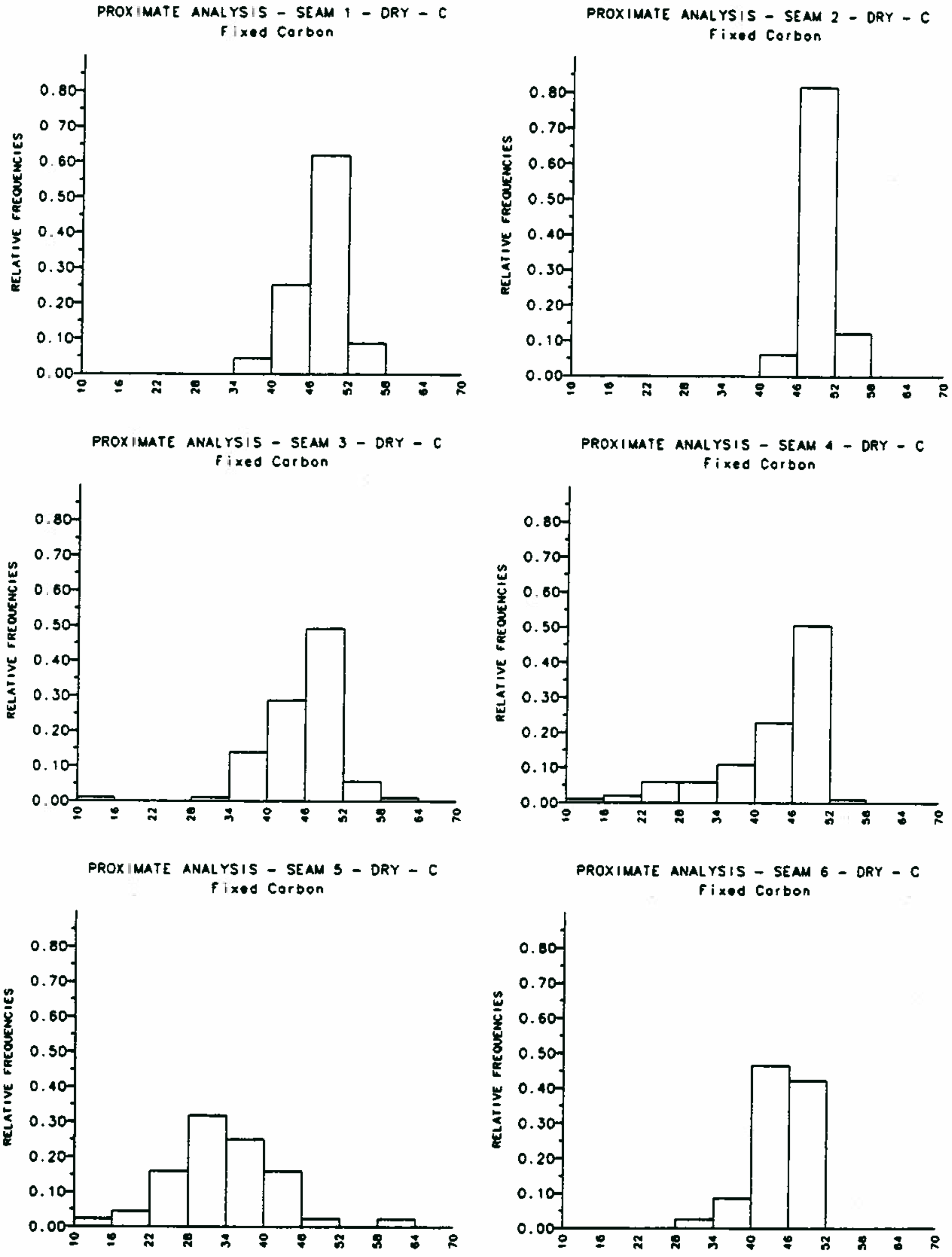


Figure 21. Frequency distributions of fixed carbon of coal seams 1 to 6 for the Highvale area. Fixed carbon is expressed in percent. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis.

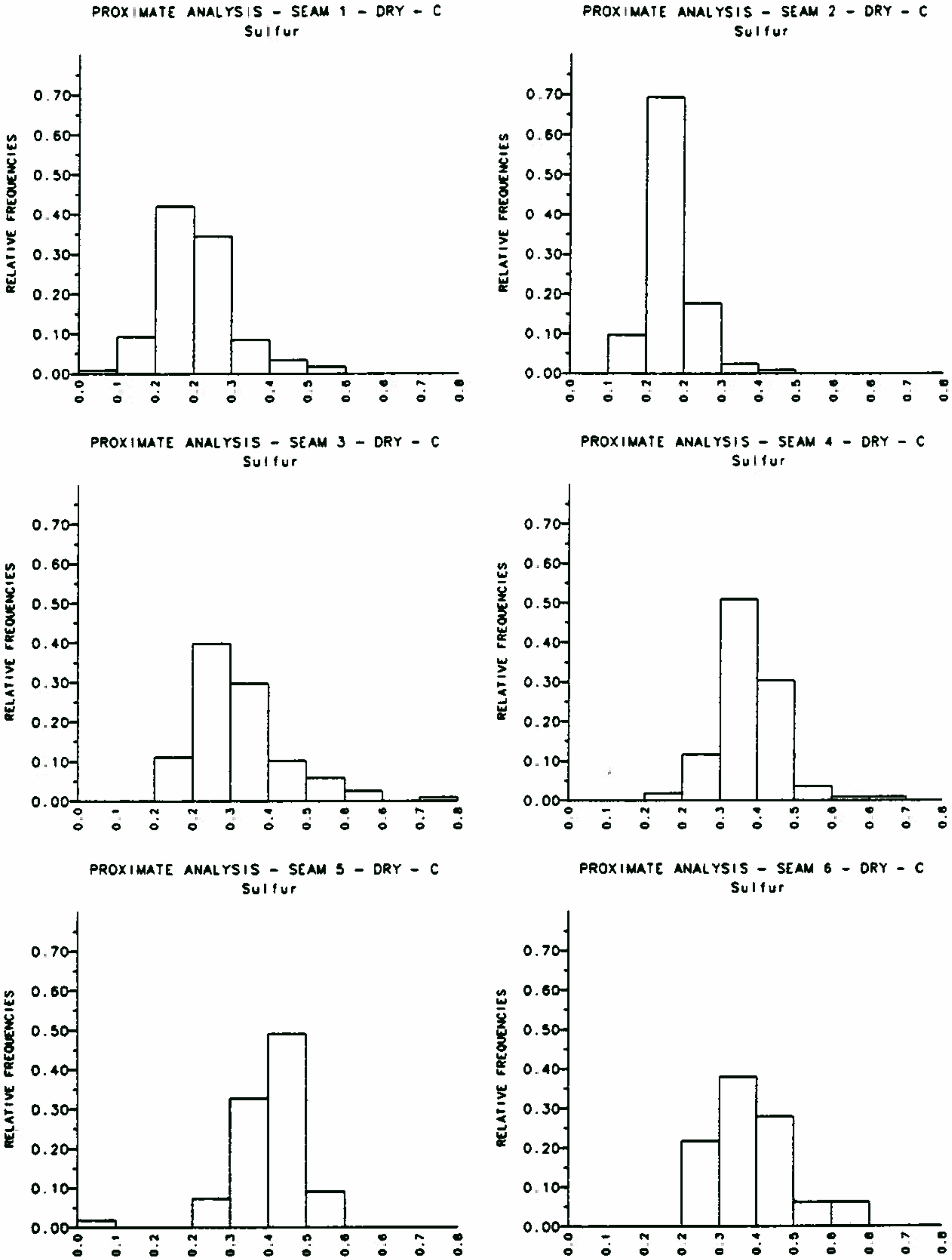


Figure 22. Frequency distributions of total sulphur of coal seams 1 to 6 for the Highvale area. Sulphur is expressed in percent. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis.

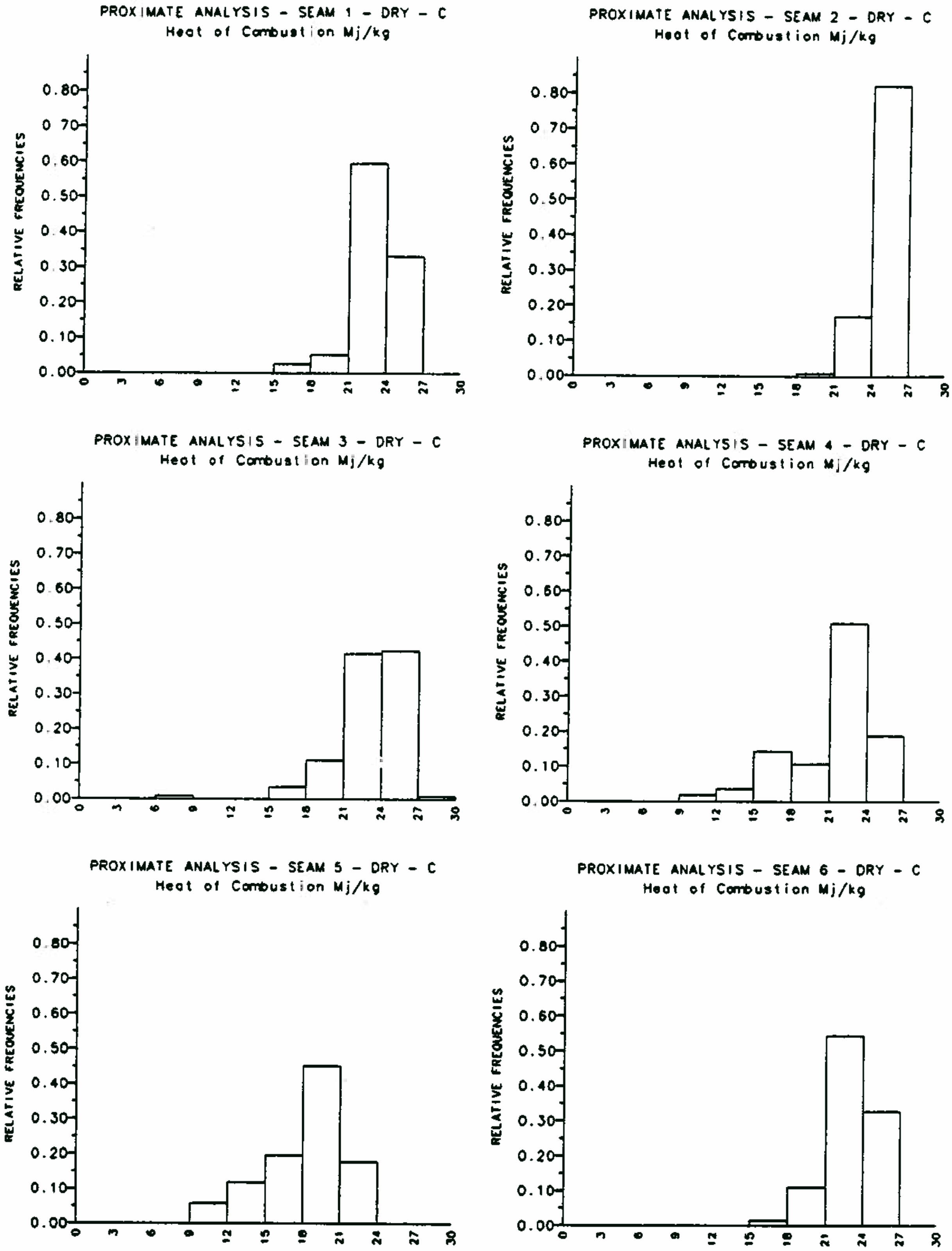


Figure 23. Frequency distributions of heat of combustion of coal seams 1 to 6 for the Highvale area. Heat of combustion is expressed as Mj/kg. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis.

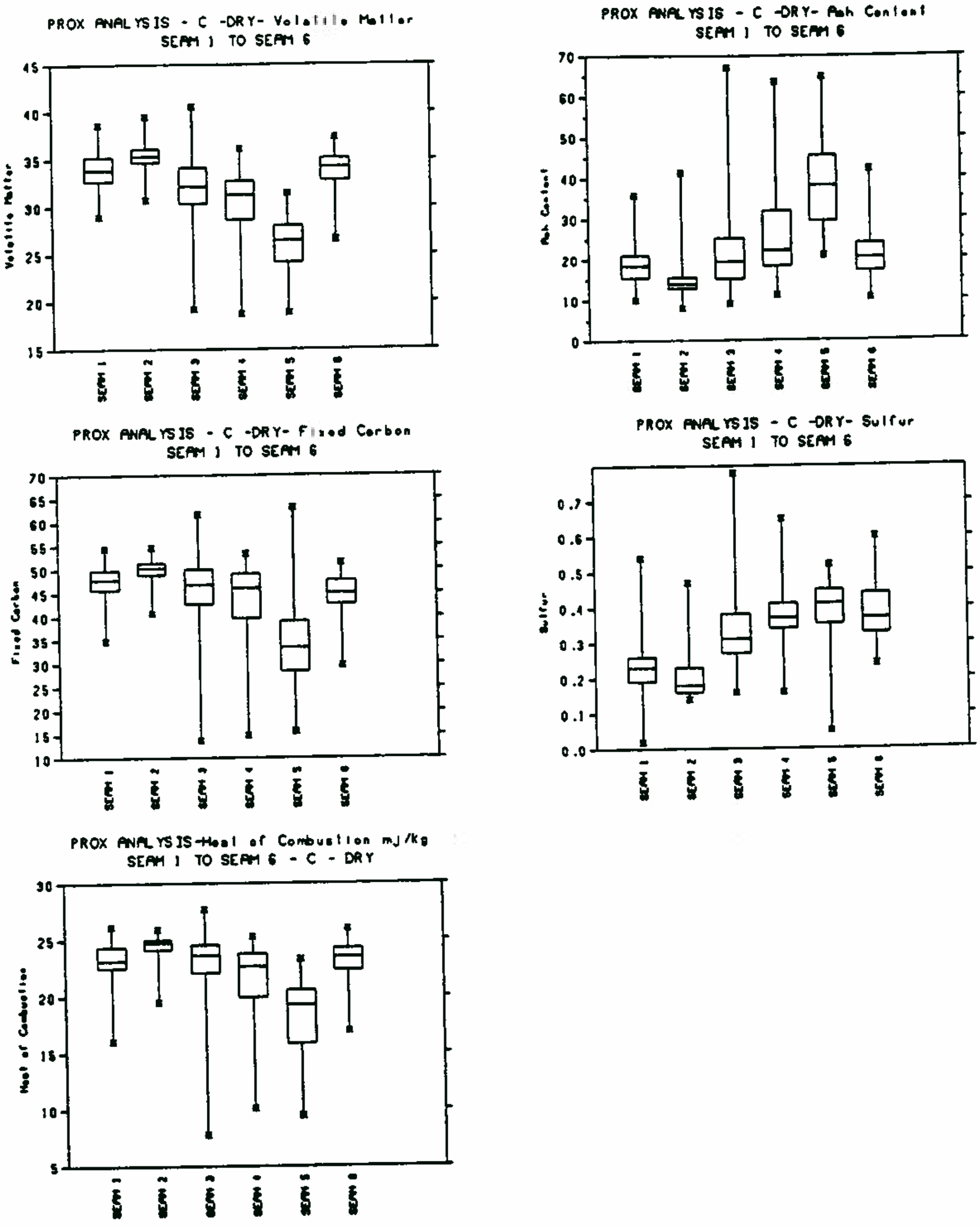


Figure 24. Multiple box plots of proximate analyses for seams 1 to 6 in the Highvale area. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis. See text for details.

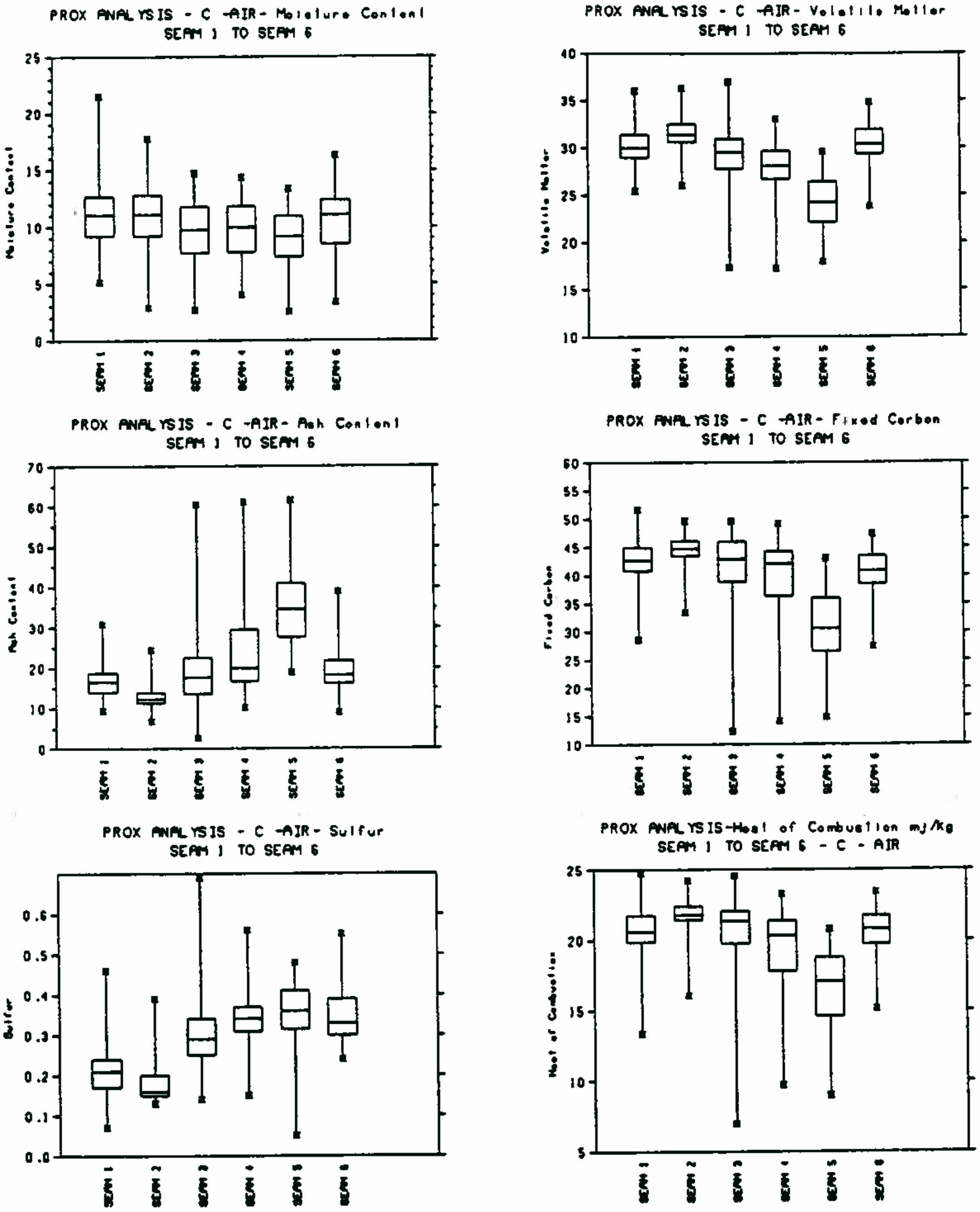


Figure 25. Multiple box plots of proximate analyses for seams 1 to 6 in the Highvale area. The data are from sampling category C (composite samples) on an air-dried (AIR) basis. See text for details.

seam-to-seam variations of the different variables from the proximate data. These provide further indication of the non-homogeneity amongst the six coal seams. For example, seam 5 appears to stand out in all analyses as having a different distribution. Basically, the box plots reveal that the central tendency and shape of the distributions from seam 5 are often different from the rest of the seams. From a geologic perspective, we would expect seam 5 to be different. It is not mined at Highvale because it has an extremely high ash content, is highly variable in thickness, and generally of poor quality. In many parts of the mine this seam is not even a coal (by definition, a coal contains at least 50% organic matter, by weight). These characteristics are typical of coal seams formed proximal to an active fluvial source, perhaps as a low-lying mire affected by frequent flooding events. In the Whitewood mine immediately north of Highvale, seam 5 is considered economic and has a lower ash content. Towards the south, seam 5 grades into a carbonaceous mudstone, suggesting that the fluvial source was to the south during peat accumulation.

The relatively low variability of seams 1 and 2 is also obvious from the smaller sizes of the boxes. The positions of the boxes indicate the relative levels of the variables. For example, seam 2 has the highest level of volatile matter, fixed carbon, heat of combustion (calorific value) and the lowest level of ash and sulphur. Seams 1 and 2 likely resulted from the accumulation of peats in laterally extensive mires isolated from the effects of fluvial systems at least in the Highvale area.

The skewness of the distribution in box plots is represented by the spacings between the horizontal bars of the box and the lengths of the "whiskers" that link the box to the maximum and minimum points. For example, in the plots for sulphur (on a DRY basis) the distribution from seam 5 is observed to be negatively skewed, whereas those from seams 2 and 3 are positively skewed (See figures 24 and 25). This means that there are more larger values (negatively skewed) or more smaller values (positively skewed) in the data distribution.

A summary of the univariate statistics for each seam is listed in tables 3A and 3B. The listing consists of (1) sample size (N), (2) sample mean ($\bar{X}$), (3) sample median or the 50 percentile (Q_{50}), (4) skewness S, (5) kurtosis K, (6) mode and (7) the P-value for normality tests. In addition, tables 3C and 3D list the values of the different variability measures for the same data. These consist of the maximum, the minimum, the range, the interquantile range, the variance and the coefficient of variation, of each variable for each seam.

From the above analyses, it can be seen that normality of the data distribution is not usually satisfied in the Highvale data. This therefore casts doubts on the usefulness of the sample mean as a measure of central tendency, and the appropriateness of many statistical methods relying on the assumption that the data distribution follows a Gaussian form. For this reason, the emphasis on the use of the median as the measure of central tendency is appropriate. This probably implies that the variance is not an appropriate measure of variability in the present case, because the variance is by definition, the root mean square departure from the mean. It is still used in this exploratory study because most readers are familiar with it. A different measure for variability based on the departures from the sample median will be examined in Phase 2 of the present study. There should also be considerations for the use of "distribution-free" statistical techniques. The following is a seam by seam summary of observations of univariate analyses for the DRY category (tables 3B and 3D).

Seam 1: This is a seam with fairly low variability in most of the variables analyzed. It has a medium to high level of volatile matter (median 33.9) relative to other seams, and is relatively low in ash and sulphur (median 18.4 and 0.25 respectively). Fixed carbon and calorific value are both high (median 47.7 and 23.2 respectively) and are close to those of seam 2.

Seam 2: This is by far the most consistent seam in terms of coal quality. The variance is the lowest of all variables analyzed. Of all six seams, it has the highest level of volatile matter (median 35.4), fixed carbon (median 50.2) and consequently the highest calorific value (median 24.8). It also has the lowest level of sulphur (median 0.18) and ash (median 13.9).

Seam 3: This seam has a medium level of volatile matter (median 32.2), ash (median 19.3) and fixed carbon (median 46.8) among the six coal seams. The variances of these variables are also intermediate relative to the other seams. This seam has the highest variance in sulphur (0.009) although the sulphur level (median 0.31) is not particularly high in comparison with the other seams.

Seam 4: This is the seam with generally high variance in most variables studied. It has the highest variability in volatile matter (12.8), ash (132.1) and calorific value (12.1) among the six seams, although the variance of sulphur is low (0.004). The level of all variables in this seam is intermediate relative to the other seams.

Seam 5: This is also a seam with generally high variance in all variables: Volatile matter (10.2, second highest), ash (108.2, second highest), fixed carbon (71.9, highest), sulphur (0.006, second highest), calorific value (11.9, second highest). It also has the lowest level of calorific value (median 19.2) and volatile matter (median 26.4) and the highest level of sulphur (median 0.4) and ash (median 38.0) of all six seams. For all variables, either the skewness S or kurtosis K of the data from this seam is different from those of the other seams. For example, in volatile matter, K is negative for seam 5, while all other seams have positive K. The same holds for ash. The skewness of the distribution for fixed carbon of seam 5 is positive while all other seams are negative, whereas the sulphur of seam 5 is the only set that is negatively skewed. These indicate that seam 5 is a different population in the statistical sense.

Seam 6: This is an intermediate seam like seam 1. The level of volatile matter is second to seam 2, with a median value of 34.1, although it ranks fourth in terms of variance in this variable. It has the second highest level of sulphur (median 0.37), whereas the variability of sulphur in this seam is the same as those of seams 1 and 5. The other variables are intermediate in terms of both level and variability, among the seams.

Also included in tables 3A and 3B are results of tests for normality for each of the variables. The variables which are tested to be normally distributed are listed in table 4. It can be seen that less than 30 percent of the variables were found to be normally distributed. Hence, care must be taken in the general application of classical statistical techniques, where normality is assumed. Moisture content (H_2O) and volatile matter (VOLM) are tested to be normally distributed in most cases. However, there is enough heterogeneity among the seams that exceptions are not difficult to find. This means that for the general analysis of the data set, nonparametric methods with no distributional assumptions are likely to be more appropriate. The major disadvantages of nonparametric methods are the compromise in power and perhaps a slight loss of information in some cases.

Correlations

The correlation coefficient was computed between each pair of variables for each seam. Tables 5 to 10 contain the correlation coefficients for each of the six seams for both DRY and AIR categories. There are two values for each pair of variables in tables 5 to 10. The second entry represents that from the DRY category. Based on the analyses described previously, the correlation coefficients which were computed from a pair of normally distributed variables are indicated using an asterisk. This means that the normality assumption for the significance test was satisfied. The correlation coefficients which are significant at the usual 5 percent level are denoted with the same asterisks in parentheses. These are the linear correlations showing statistically significant relationships for the given variable pairs.

The tables show that there are only a few cases where such significance can be obtained. This is because of both the existence of non-normality in the data and the lack of relationships between some of the variables. There is also considerable variation in data characteristics from seam to seam. It also supports the use of a nonparametric (distribution-free) approach in subsequent data analysis. Some of the significant relationships are represented by low correlation coefficients. This implies that although relationships exist, they are relatively weak and may not be useful for prediction.

Without rigorous statements of statistical inference, one may still make observations regarding the pattern and magnitude of the correlations listed in tables 5 to 10. One observes, first of all, a close correspondence in values between the DRY and AIR samples for all seams. One also sees a consistently high positive correlation between FXC and MJKG and a consistently high negative correlation between FXC and ASH, or MJKG and ASH. Such relationships are what one may expect physically. However, the present preliminary analysis cannot establish statistical significance in these cases, because of violating the normality assumption.

Figures 26 to 37 (in pocket) are visual descriptions of the relationships among the variables. These are useful supplements to the statistical results already described. For example, one can see that the relationship between FXC and MJKG, although not supported by the previously described significance test, is fairly strong for all seams. The plots also show the existence of outliers in the relationship between MJKG and ASH. The skewed data distribution is manifested as uneven clustering of points in these plots.

The Spearman rank correlation was computed for the variables and the results are presented for comparison. Normality is not required and a similar pattern of association was observed. The results are listed in tables 11 to 22 for both DRY and AIR analyses.

Here, statistically significant correlations can be identified with those having a P-value of less than 0.05. One can see a consistent negative relationship between ASH and VOLM, and ASH and FXC throughout all seams for both AIR and DRY samples. Similarly, there is a consistent positive correlation between FXC and MJKG and VOL and MJKG. The relationships associated with S seem to vary with seams, in magnitude, sign and significance level. As mentioned previously, these relationships are not unexpected. Nevertheless, they show that data quality is of a reasonable level and careful analyses should provide reasonable results.

Multivariate Comparisons

Multivariate comparisons were performed for the proximate data. The purpose of such comparisons was to see if the samples tested were the same based on the attributes involved. Namely, one is interested in knowing if the seams are similar, considering the variables measured for each seam.

The previous analyses have indicated that there are considerable variations in the frequency distributions of the variables contained in the data set. Consequently, statistical techniques based on multivariate normality are not appropriate. For both the DRY and AIR categories of the proximate data the MRPP was applied to test for homogeneity among the six seams based on the measured values of H_2O, VOLM, ASH, FXC, S and MJKG (H_2O was not available in DRY category, therefore only five variables were used in those cases). The same comparison was made between all pairs of seams. The results give P-values from 10^{-4} to 10^{-34}, indicating statistical significance. This means that it is safe to conclude that the six seams of coal in the Highvale have different characteristics from each other. The same conclusion can be reached for samples from the DRY and AIR categories.

A preliminary run of principal components analysis using the AIR samples was attempted. A principal component analysis was performed for each seam. As mentioned in the methodology section, the purpose of principal component analysis is basically for variable reduction and orthogonalization (the components are orthogonal, ie. independent of each other). There is variation in the results from seam to seam. Generally, the first two to three principal components explain approximately 90 percent of the variance. The first component in all six seams explains between 54 and 68 percent. Seams 1 and 2 rank the lowest in terms of the percentage of variance explained by the first component (54 percent approximately in both cases). This first component appears to be associated with the energy content of the various seams. Table 23 lists the weights on each variable for the first component from each seam. These weights can be considered as correlation coefficients between the first component and the specific variable. It appears that approximately equal positive coefficients are obtained for MJKG, FXC and VOLM, while negative coefficients are obtained for ASH. The second component explains between 17 to 24 percents for the six coal seams. It appears to be associated with water and sulphur but there is inconsistency from seam to seam. The third component explains approximately 12 to 16 percent of the total variance. There is some relationship with sulphur and water as in the second component but again there is inconsistency between seams.

The significance of the above results is uncertain at this point. The use of principal components analysis as a variable-reduction technique may be more relevant in the case of a large number of independent variables in a regression model. The set of variables examined in this case are more likely to be dependent variables which are individually significant in the geologic sense. No orthogonalization is needed. The use of principal components analysis here is simply for variable reduction to obtain a representation of energy contents in the seams. It is possible to map the pattern of this first principal component and relate it to the geological structure. The pattern for each seam can also be compared.

The analysis of the ultimate data

The emphasis is again on the seam to seam variations. The comparisons and exhibits are therefore seam-by-seam. The data set was analyzed on two bases. One was as determined (AD), i.e., this is the original data from TransAlta. The other was dry ash free (DAF) which was derived from the original data following ASTM (American Society for Testing and Materials) procedures.

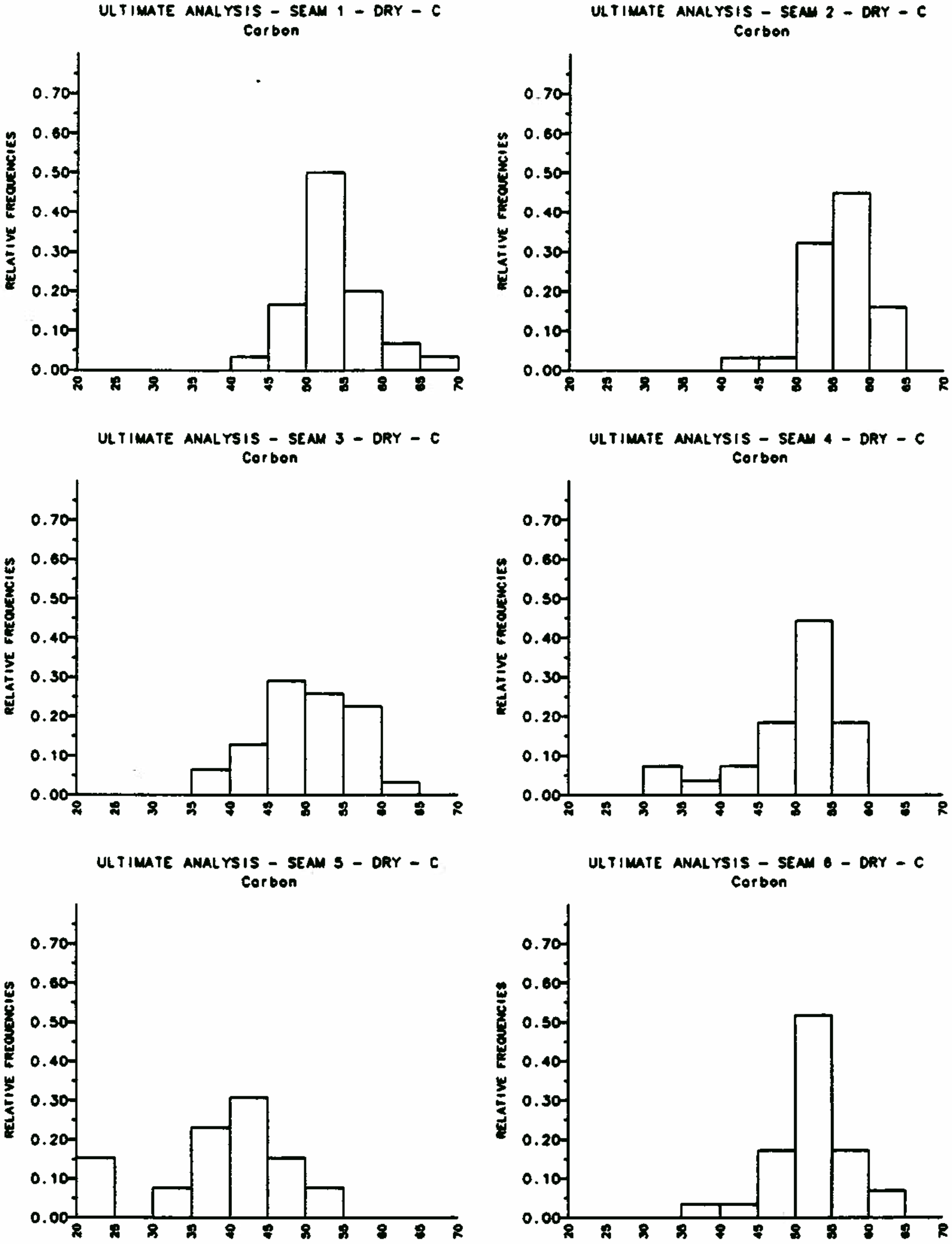
Distributional characteristics (AD data)

Figures 38 to 43 are histograms of the different variables from ultimate analysis. The multiple box plots are shown in figure 44. There is again considerable variation from seam to seam in all the variables involved. For example, the change in skewness for carbon content distribution from seam to seam suggests that the data probably come from different populations. Nitrogen, on the other hand, has fairly symmetric distribution shape, except in the case of seam 1. Water content shows a wide range of skewness and kurtosis in the shape of its frequency distribution with seam 5 having the highest peakedness. The high variability in frequency distribution can also be observed in ash, hydrogen and sulphur. However, negative skewness exists in all six seams for hydrogen content distribution and positive skewness exists in all six seams for sulphur content distribution.

The multiple box plots provide another visual representation of variability of the ultimate data. An interesting observation from the box plots of figure 44 is that seam 5 again appears to stand out in all variables, either in terms of the level of the particular variable (represented by the median) or the variability (represented by the range, or the length of the box and whisker diagrams). The same can be said of seam 2, perhaps with some exceptions. In any case, seams 1 and 2, have the highest level of carbon and lowest level of ash.

Figure 44 shows that the maximum levels of moisture in seams 1, 2, 3 and 6 are all having the value of 20 percent. This is an imposed upper limit assigned to moisture content in the data acquisition process.

Table 24 contains the summary statistics for the AD ultimate data. Seams 1 and 2 have the highest levels (medians) of moisture and carbon. Seam 6 has the lowest level of moisture, whereas seam 5 has the lowest level of carbon. These are associated with high frequencies of



ULTIMATE ANALYSIS - SEAM 1 - DRY - C
Carbon

ULTIMATE ANALYSIS - SEAM 2 - DRY - C
Carbon

ULTIMATE ANALYSIS - SEAM 3 - DRY - C
Carbon

ULTIMATE ANALYSIS - SEAM 4 - DRY - C
Carbon

ULTIMATE ANALYSIS - SEAM 5 - DRY - C
Carbon

ULTIMATE ANALYSIS - SEAM 6 - DRY - C
Carbon

Figure 38. Frequency distribution of the carbon content of coal seams 1 to 6 for the Highvale area. Carbon content is expressed in weight percent. The data are from ultimate analysis and DRY-C samples.

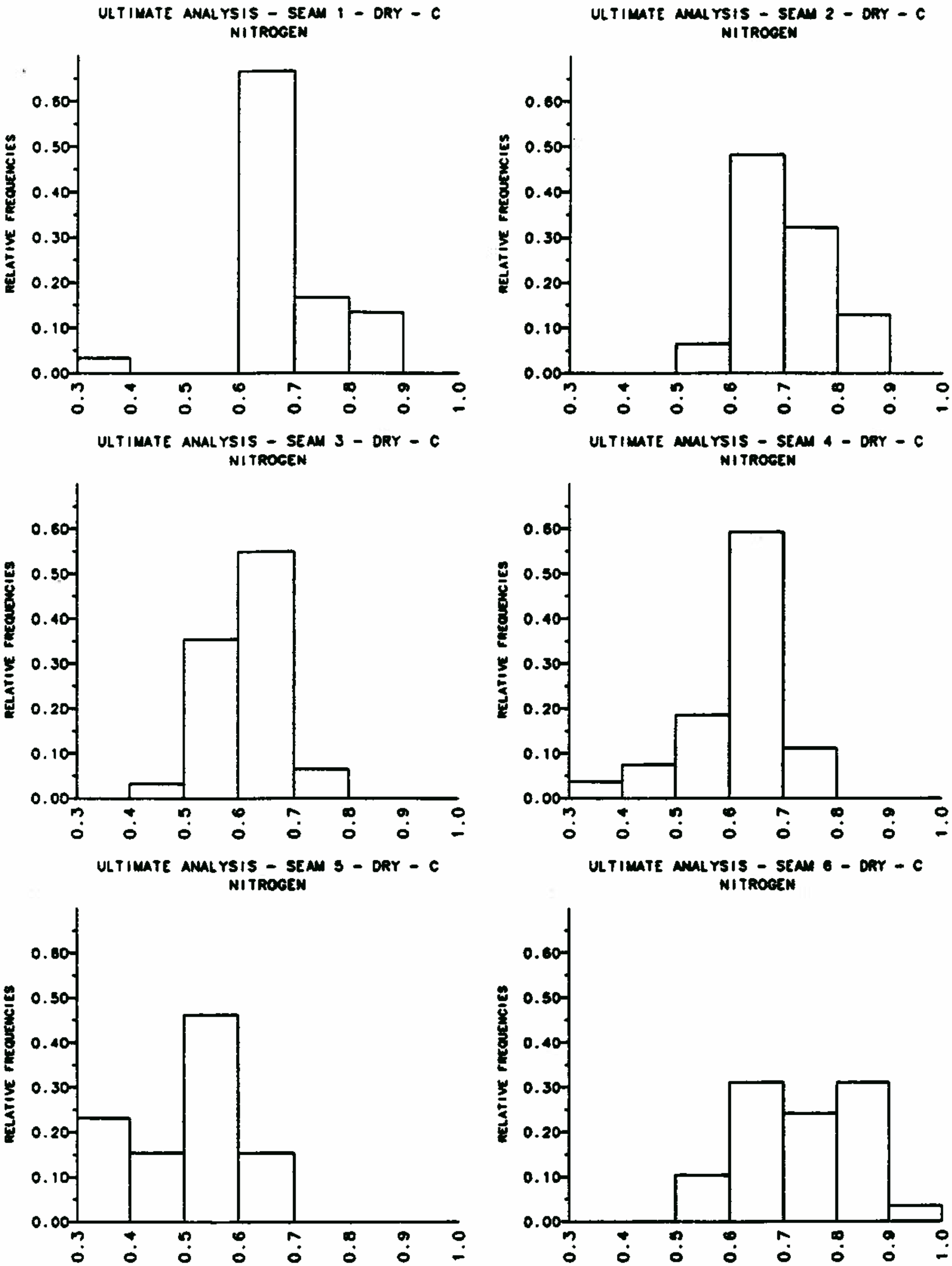


Figure 39. Frequency distribution of the nitrogen content of coal seams 1 to 6 for the Highvale area. Nitrogen content is expressed in weight percent. The data are from ultimate analysis and DRY-C samples.

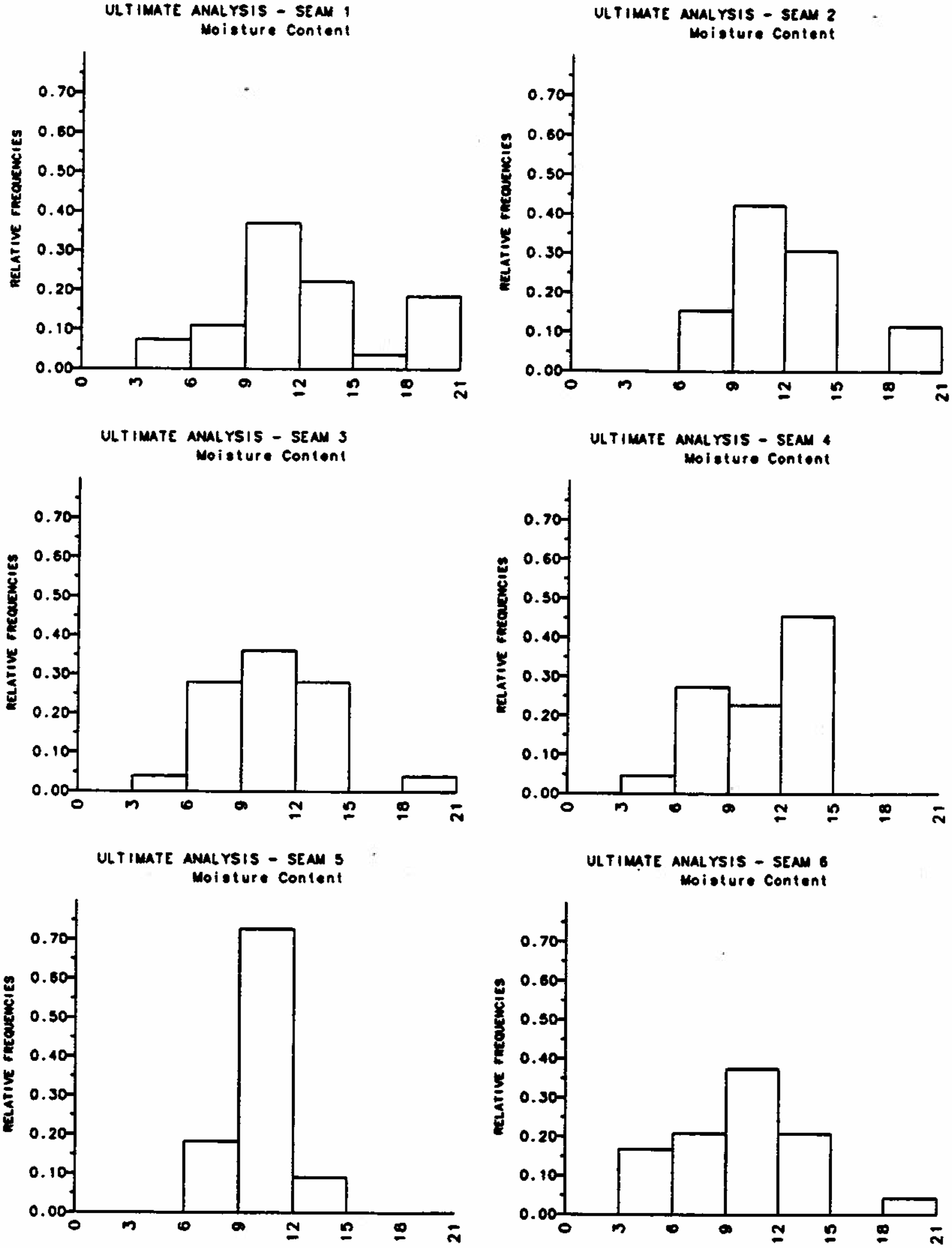


Figure 40. Frequency distribution of the moisture content of coal seams 1 to 6 for the Highvale area. Moisture content is expressed in weight percent. The data are from ultimate analysis and as-received composite samples.

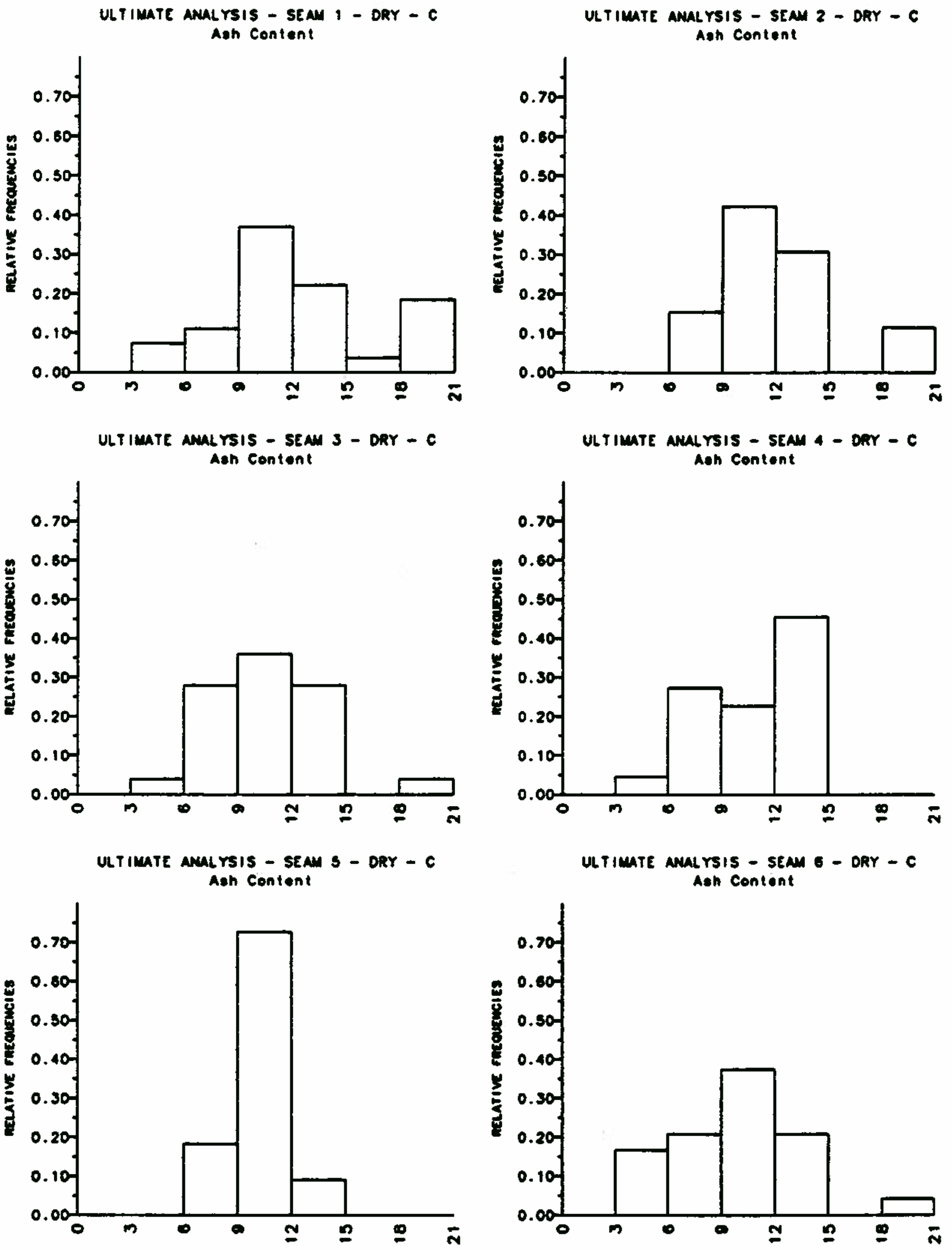


Figure 41. Frequency distribution of the ash content of coal seams 1 to 6 for the Highvale area. Ash content is expressed in weight percent. The data are from ultimate analysis and DRY-C samples.

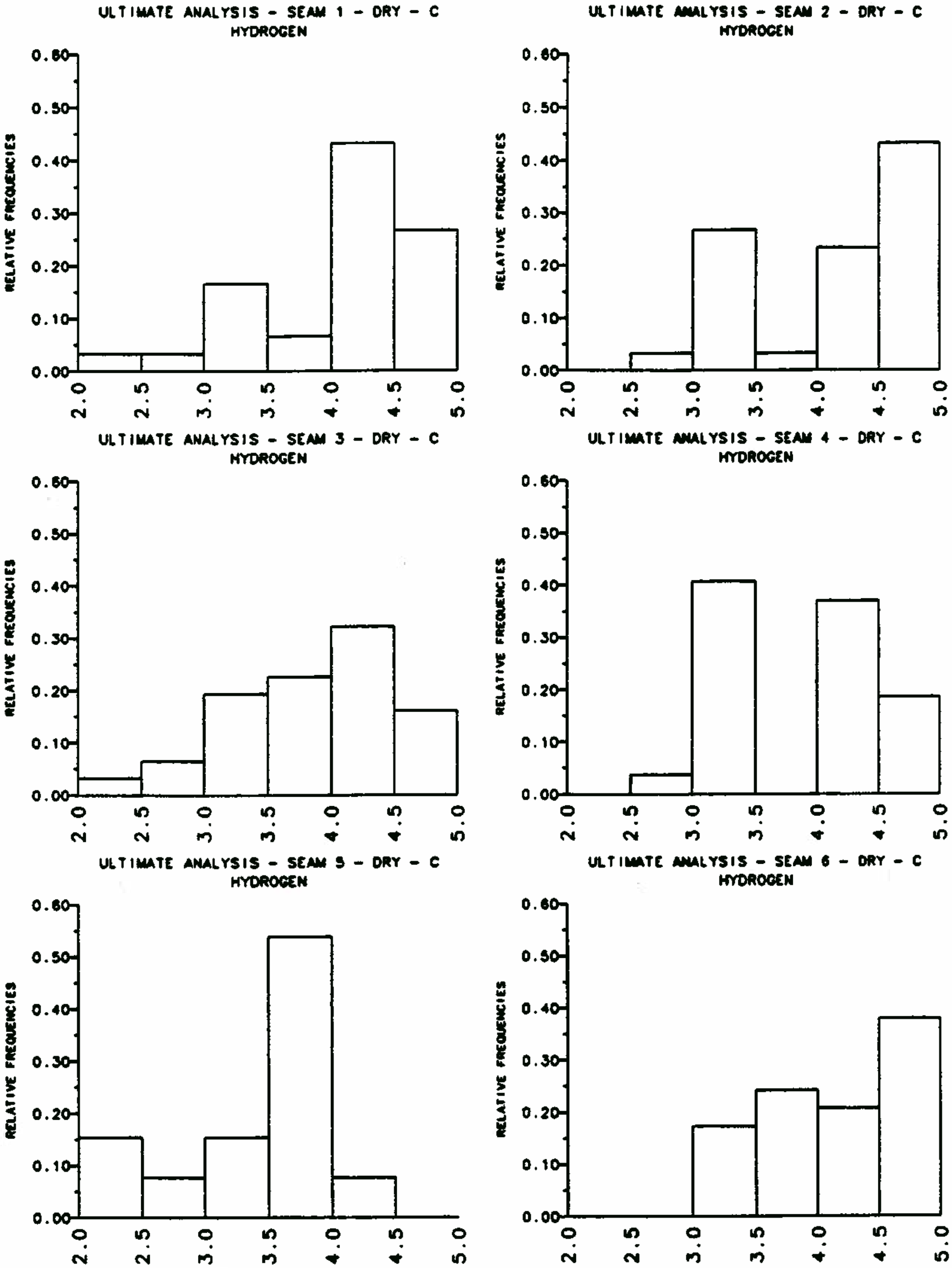


Figure 42. Frequency distribution of the hydrogen content of coal seams 1 to 6 for the Highvale area. Hydrogen content is expressed in weight percent. The data are from ultimate analysis and DRY-C samples.

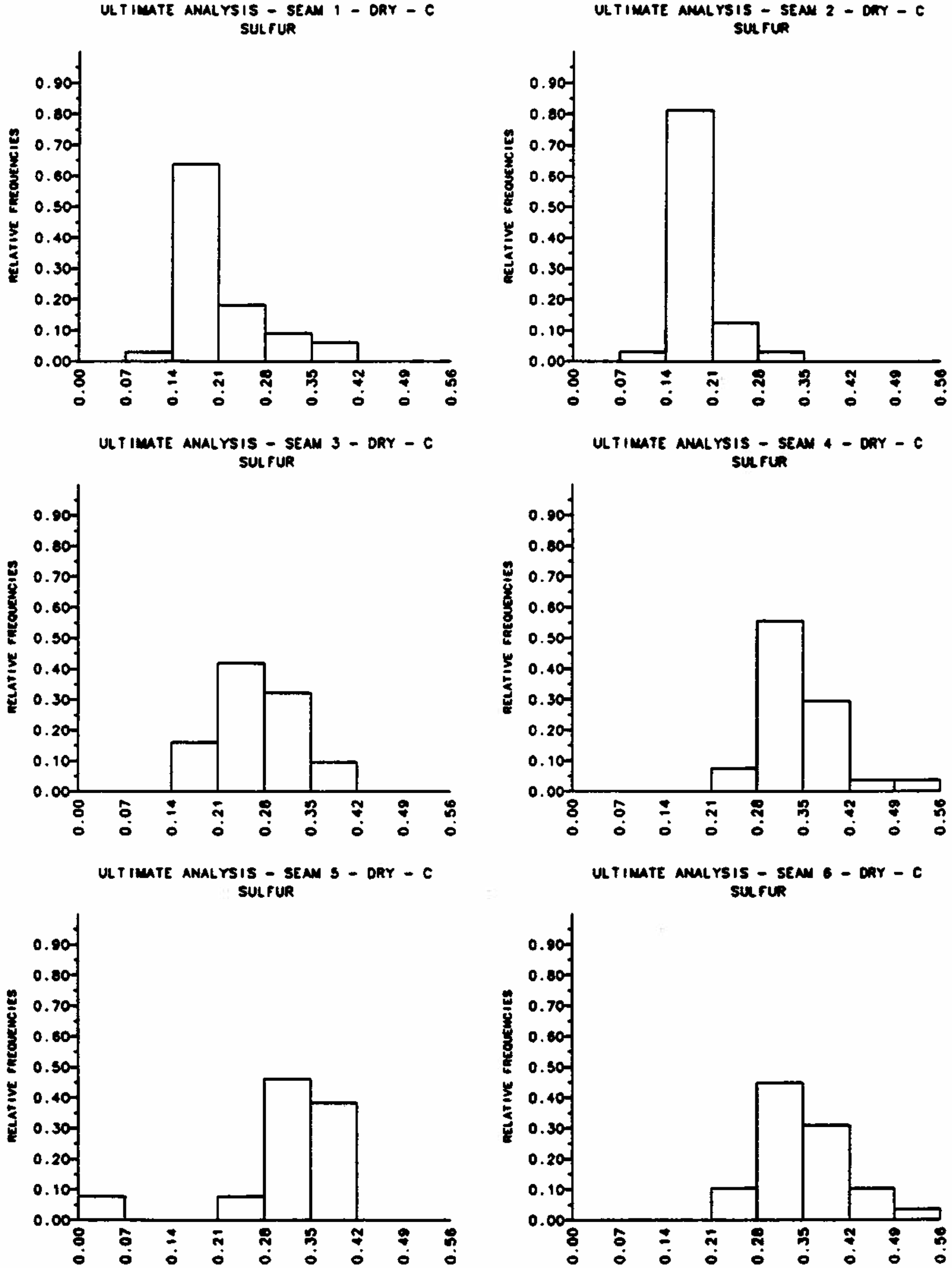


Figure 43. Frequency distribution of the sulphur content of coal seams 1 to 6 for the Highvale area. Sulphur content is expressed in weight percent. The data are from ultimate analysis and DRY-C samples.

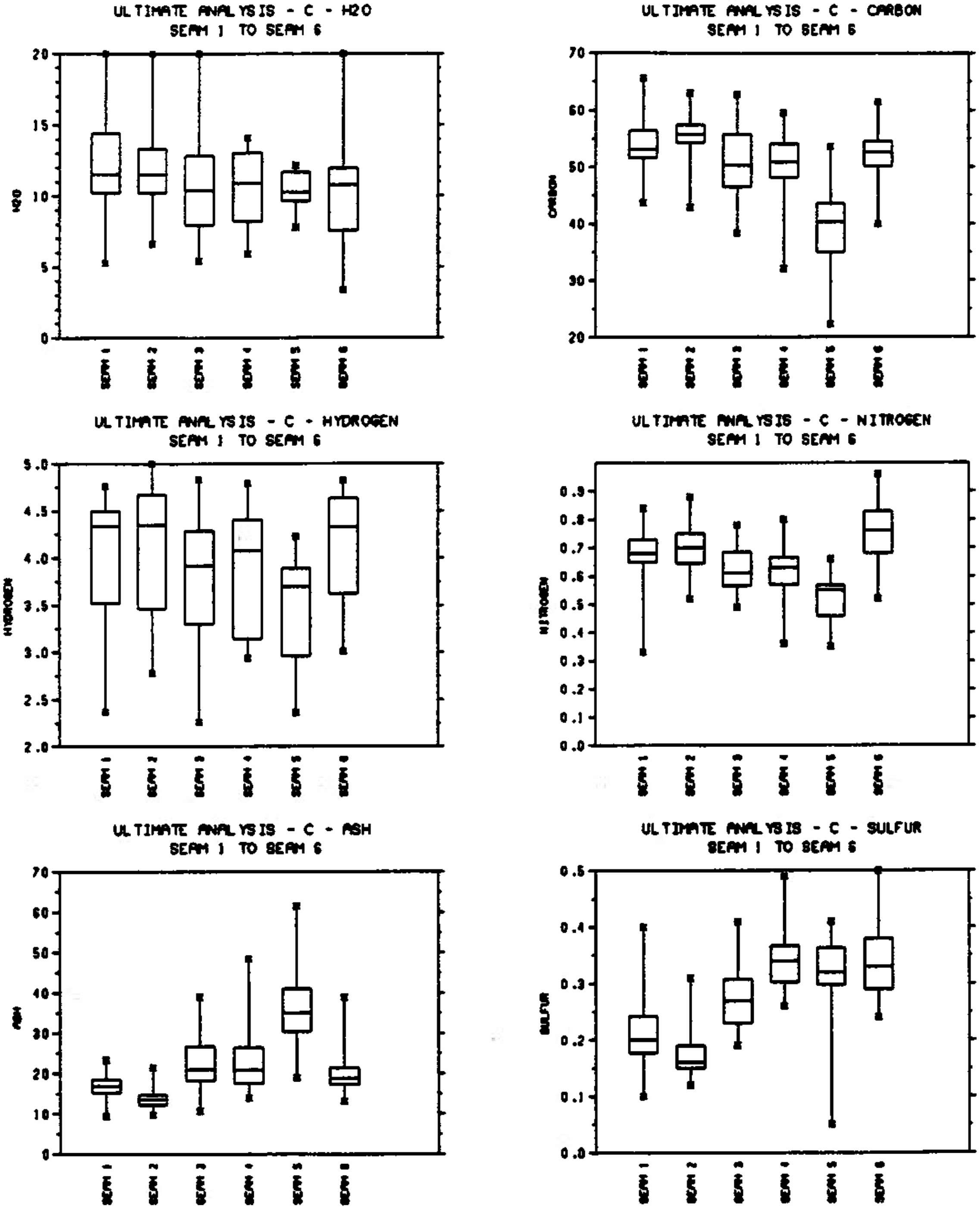


Figure 44. Multiple box plots of ultimate analyses for seams 1 to 6 in the Highvale area. The carbon, hydrogen, nitrogen, ash and sulphur are reported on a dry basis. The data are from composite samples.

occurrence of low value observations as shown in the modes of the respective variables. Ash content is highest in seam 5 and lowest in seams 1 and 2. In fact, the ash content of seam 5 is more than double that of either seam 1 or 2. The ash contents in seams 3, 4 and 6 are intermediate. The sulphur contents of seams 1 and 2 are again the lowest of all seams. Particularly, the sulphur in seam 2 is approximately half of that in either seam 4, 5 or 6. In fact, seam 4 and seam 6 have the highest sulphur contents. Hydrogen is highest in seams 1 and 2, followed closely by seam 6. Seam 5 has the lowest hydrogen level. As for nitrogen, a departure from the patterns of the other variables exists in the form of seam 5 having the lowest level and seam 6 the highest. The other seams are in between.

0.10 Regarding the shapes of the data distributions, represented numerically by the skewness S, the kurtosis K and the P-value for normality tests, considerable variations exist among variables and among seams. The distribution for carbon is negatively skewed except for seam 1, while the distribution for ash is positively skewed, again with the exception of seam 1. The skewness for the distribution of hydrogen is positive for all seams, whereas negative skewness generally exists in sulphur distributions (except seam 5). The other variables have both positive and negative skewness indicating considerable variabilities. Similar situations occur with kurtosis. Carbon and ash have mostly positive kurtosis, both with the exception of seam 3, whereas hydrogen has mostly negative kurtosis except for seam 1. The kurtosis for sulphur is all positive but less consistent patterns exist for nitrogen and moisture. The results of the normality tests varies considerably among the seams for each variable. The only consistent result is that normality is satisfied for seam 3 in all variables observed. The list of normal variables is shown in table 25.

Table 26 lists the variability measures for each ultimate AD variable. Discussion is again based on the variance as a variability measure, recognizing that for practically all cases, the median is close to the mean, and that a better measure of variability using the median as the centre of moments will be examined in detail later. Moisture is most variable in seam 1 and least variable in seam 5. As mentioned earlier, there is a maximum possible value of 20 percent on all moisture observations, which appears to be some kind of imposed limit due to either measurement or data processing. Both ash and carbon contents are most variable in seam 5 and least variable in seams 1 and 2 as expected, whereas the variation in sulphur content is highest in seam 5 and lowest in seam 2. Variation in both hydrogen and nitrogen is about the same in all six seams.

Correlations (AD data)

The correlation coefficients (both Pearson's and Spearman's) between the ultimate AD variables are listed in tables 27 to 32. Since the data distributions contain significant departures from the Gaussian form, the use of the Spearman correlation coefficient is more relevant. Generally, there is a significant negative correlation between carbon and ash, as expected. In fact, this is the consistent relationship through the six seams in tables 33 to 38. The range of variations from seam to seam includes both variations in the nature of relationships and the level of significance obtained. The ultimate AD data contains considerable variability.

The data sets were then converted to dry ash free (DAF) basis. The conversion follows standard ASTM procedures. Oxygen was obtained by subtracting the total proportion represented by other elements from 100 percent. Both oxygen and hydrogen were given in values which include the hydrogen and oxygen in associated water. Therefore, these were first converted to dry basis, from which the values for dry ash free basis were derived. Other elements were converted by applying the appropriate conversion factors as outlined in the ASTM guidelines.

Distributional characteristics (DAF data)

The variability in distributional characteristics for the DAF data is very similar to that of AD. The data distribution for any variable from a given seam shows varying form with little consistency. For brevity, only the main features are described for the DAF data. The major elements obtained are oxygen, hydrogen, carbon, nitrogen and sulphur. The summary statistics for each of these elements are shown in table 39. An important feature is that after eliminating samples with one or more missing data, a smaller sample size was obtained for the elements involved. In particular, seam 5 only contains 10 data points, which is quite small for rigorous statistical examination.

It appears that the highest level of oxygen is in seam 3, whereas seam 5 has the lowest level and also the largest variability. In fact, seam 5 contains the highest variability of all seams in all elements on DAF basis except for nitrogen. It also has the lowest level of carbon and the highest sulphur and hydrogen contents of the six seams. Despite the small sample size, seam 5 still stand out as being different from the other seams, based on the DAF data. Less obvious is the data from seams 1 and 2. Our previous analyses showed that these are the 'better" coal seams of the Highvale area. The DAF data, however, do not show the same level of consistency as the data examined before. Seam 1 still has the lowest variability in DAF carbon, but the carbon level is not particularly high compared to other seams. However, table 42 does show that seams 1 and 2 are the lowest in sulphur level as well as the least variable in sulphur content.

Correlations (DAF data)

The Spearman correlation coefficients for the DAF data are shown in tables 40 to 45. The correlation pattern of the ultimate DAF variables also contains considerable variation from seam to seam. The only consistent and significant correlations are the negative relationships between carbon and oxygen and carbon and hydrogen, if we consider significance as having P-values of less than, say 0.10. Consistency means that the same negative relationship exists throughout all six seams. The relationship between sulphur and oxygen is significant only when it is positive. This occurs in seams 1, 2, 4 and 6 and suggests that data noise is of a considerable level in the other two seams. Significant positive relationships also exist between nitrogen and hydrogen for seams 1 to 4. Other than these features, there is little consistency in the pattern of correlation from seam-to-seam.

The analysis of the hydrogen/carbon ratio

The hydrogen to carbon ratio (H/C) was computed for each sample of the ultimate data. The AD values of hydrogen and carbon were divided by 1.007 and 12.02, the respective molecular weights, before the ratios were analyzed on a seam-by-seam basis.

The frequency distribution of the ratios is generally not normal, as expected, with only those from seam 1 approximating normality at the 10 percent significance level. The summary statistics of the ratios for each seam are listed in table 46. Using the median (Q_{50}) again as the measure of central tendency, one may conclude that seams 1 and 2 are the lowest in H/C level and seam 5 is the highest in H/C level, with the other seams in between. With a median of 1.1146, seam 5 is approximately 18 percent higher than seam 2 in the ratio of H/C. Seam 5 is also the most variable seam regarding H/C, with the highest variance and the widest range. Seams 1 and 2 are the least variable in terms of the variance of the data, although the range is the smallest in seam 3.

Note that the above statistics were computed using a sample with no missing data. The estimates for seam 5 were obtained with a sample size of only 13. This is less than half the sample size of any other seam. Although bias with such a small sample is likely, the statistics seem to confirm our knowledge of seam 5 as a thin, low quality seam. On the other hand, seams 1 and 2 emerge as ones with the lowest H/C and most consistency.

CONCLUSIONS AND RECOMMENDATIONS

The Highvale data were analyzed for statistical properties and basic relationships using univariate, bivariate and multivariate techniques. The following conclusions can be made regarding the analyses of both proximate and ultimate data.

(1) There is high variability from seam to seam for all variables measured. Variations are in terms of levels (central tendencies), variance, frequency distributions and relationships. Statistically, the implication is that the same variable from different seams may come from different populations. Geologically, this means that one has to differentiate the coal samples clearly as to which seam they come from. In other words, it is probably unwise to mix coal samples from different seams and treat the combination as a single population.

It is recognized that there are situations where seam-by-seam differentiation is not practicable. For example, in the Highvale-Ardley comparisons, one probably has to combine the Highvale seams in some manner because the regional Ardley data were not differentiated by seams. In such situations, it would be necessary to try various ways of combining the Highvale data, making use of the geological information available from the Highvale minesite.

Geological information is also useful in assisting the search for explanation in the observed seam-to-seam variation. This in turn facilitates the formulation of geological models regarding the formation of coal. Geological models are important to the design of statistical analyses. In particular, the observed differences among the Highvale coal seams suggest that some geological hypothesis should be formulated before further rigorous statistical analysis is applied. The statistical analysis can then be used to modify and ultimately verify the geological hypothesis.

The high seam-to-seam variability also suggests that some form of statistical smoothing may be relevant in studying the Highvale data.

The principal components analysis, for example, can be used in identifying the spatial patterns of the data from each seam and possibly relating them to the geological information of the Highvale mine.

It is clear from the results of this exploratory study that there is little room for classical statistical methods based on normal theory in the analysis of the Highvale data. In fact, one should not expect that a single class of statistical procedure can be used for all variables in all seams. Statistical methods should be carefully chosen, based on the associated geological models, and the results of this exploratory work.

(2) It is suggested from both the proximate and the ultimate (AD) data that seams 1 and 2 are the most consistent in term of coal quality. The same pattern was not as obvious from the ultimate (DAF) data. One possible reason is that there are certain inaccuracies associated with the derivation of the ultimate (DAF) data files. In any event, seams 1 and 2 emerge as the seams with relatively high quality coal in the Highvale minesite, based on proximate and ultimate (AD) analyses. These two seams are shown to have relatively high carbon levels and are relatively low in ash. Variability within these two seams is also generally low.

While such results are not unexpected, the present study provides detailed quantitative measures on all observed variables. This information is useful not only to those who will further analyse the data, but also to those who are interested in the nature of the different seams.

(3) Seam 5 emerges on the other extreme as a thin, low quality seam. Variability is generally high for most observed variables. It is low in carbon level and high in ash. The statistics for this seam show why it is not economic. An interesting observation is that the data samples from seam 5 are usually small for the different analyses. Yet, there is no significant bias in the results obtained. In other words, those small samples appear to be showing what one would expect seam 5 to be.

This is related to an interesting problem in geostatistics: How many samples does one need to show coal quality in the Highvale mine, or in the Ardley region for that matter? Perhaps, the number is fairly small.

(4) The other seams are in between these two extremes. Although not specific to the Highvale data, there are two important considerations for the statisical analyses of data of this type. One is the problem of multiplicity. In any data rich environment, of which coal geology is certainly one, it is very tempting to look for statistical significance by performing many analyses on the same data set. Multiplicity implies that if one looks long enough, one will see statistical significance even when it does not exist. For example, in 100 analyses of correlation, one observes five significant cases. One should be careful not to conclude that the relationships for these five cases are real. The concept of type I error and statistical significance at, say the 5% level, implies that out of 100 uncorrelated cases, five will erroneously show correlation. Hence, it is recommended that analyses should be carefully planned, and the procedures towards model building should be executed in a logical way.

The second important concern is the spatial dependence. There is a strong spatial dependence in geological data. This means that the assumption of independent random samples of classical statistics is not appropriate in many geological studies. The effect of the non-randomness on the analysis using classical statistics may vary from case to case. Depending on the nature of the geologic controls behind this spatial correlation and the availability of data, it is possible to perform some kind of assessment of this effect using Monte Carlo or other simulation techniques. For example, one may compare the characteristics of simulated samples with and without spatial dependence. It will be interesting to see the result of such an investigation for the Highvale data.

The following is a proposed approach for the second phase of analysis of the Highvale data. The first step will be the characterization of variability. Consider the simple case where coal quality is represented by a variable X. The simplest way to characterize the variability of X is to use the statistical parameter variance. The variance is the root mean square departure from the mean, which is a measure of central tendency. The use of the variance of X to characterize the variability of X is basically a method of parameterization, i.e. to represent variability by one or more parameters, or numbers, which can be related to the number of effects.

There are other more detailed ways to parameterize variability in a spatial data field, such as that of the coal quality variable X. As in the case of the variance, the basic concept is still to measure the departures of the data from a central value. Since most variables have been found not to follow the Gaussian distribution, the sample median should be preferred to the sample mean as the appropriate measure of central tendency in this case. We will define a measure based on the absolute departure (i.e. neglecting the sign of the departure) from the sample median of X. For the entire sample, one may compute the absolute departures for each value of X obtained. A frequency distribution of these absolute departures can be constructed to represent the variability of X.

Several parameters may be used to describe such a frequency distribution. For simplicity, we will consider the case where a probability model can be fitted to the frequency distribution of the departures observed. For example, the gamma distribution can have two parameters for its shape and scale. The shape parameter measures the symmetry of the probability distribution and the scale parameter represents the magnitude of the variate in question, i.e. the magnitudes of the departures in this case. Hence, when the gamma distribution is fitted to the distribution of the absolute departures, one obtains two parameters for the variability of X, viz. the scale and shape parameters obtained.

Other parameters are possible, depending on the outcome of the analysis results. Alternatives are the area under the curve for the frequency distribution of the absolute departures, or the mean, median or modal values of the departures. However, probability models are useful representations, and other density functions than the gamma distribution are also possibilities.

After the process of parameterization, one may proceed to study variability on the local (Highvale minesite) and the regional scale (entire Ardley zone). Based on the geological modal of the factors affecting coal quality, several types of analyses may be performed:

(A) Comparisons of variability between the local data and the regional data.
(B) If differences are detected, identify the geological factors associated with such differences.
(C) Attempt to relate variability to geologic factors and identify any possible relationships.
(D) If relationships exist, attempt comparisons of relationships between the local and regional scenes.

The details of methodologies will depend on the outcome of the analyses, but we can envisage the usage of a wide range of statistical techniques. Exploration of relationships is not restricted to continuous data. Categorical or ordinal data very often provide useful information when analyzed appropriately. An example of categorical data is for the differentiation of the depositional environments A versus B. An example of analysis is to examine the association of coal quality X with the kind of depositional environment using the Chi-square test of contingency tables. An example of ordinal data will be the different levels of thickness in an Ardley seam, which can be incorporated in a regressional model for the level of ash.

So far we have been considering one hypothetical coal quality variable X. But coal quality is not likely to be dependent on a single variable. Other additional variables may be investigated or better still, incorporated statistically in a coal quality index. The simplest example is to develop an index based on the linear combination of several variables, with appropriate weights given to the variables used.

The scope of the analyses associated with this project is large and more ideas will develop as analyses progress. The above are only basic outlines which will serve as a guide.

LIST OF REFERENCES

Beaumont, C., (1978): The evolution of sedimentary basins on a visco-elastic lithosphere: theory and examples, Geophysical Journal of the Royal Astronomical Society, v. 55, p. 471-497.

Dawson, M., McCandlish, K., Godfrey, T., Nikols, D. and Fenton, M. (1985): Coal Subdivision of the CSPG Field Trip #1, The Ardley Coal Zone, Canadian Society of Petroleum Geologists.

Heap, M.P, Folsom, B.A. and Pohl, J.H., (1986): Coal Properties: Their Impact Upon Utilization, Paper presented at the Pacific Rim Coal Conference, Hong Kong.

Holter, M.E., Yurko, J.R. and Chu, M., (1975): Geology and Coal Reserves of the Ardley Coal Zone of Central Alberta, Albertal Research Council, Earth Science Report 75-7, 41 pp.

Irish, E.J.W., (1970): The Edmonton Group of South-Central Alberta, Bulletin of Canadian Petroleum Geology, v. 18, pp. 125-155.

Irving, E., (1979): Paleopoles and paleolatitudes of North America and speculations about displaced terrains, Canadian Journal of Earth Science, v. 16, p. 669-694.

Langenberg, C.W., Macdonald, D.E. and Richardson, R.J.H., (1986): Geological Studies in Alberta: A Status Report, Alberta Research Council Open File Report 1986-14, 106pp.

Lilliefors, H.W., (1967): On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, Journal of the American Statistical Association, p. 399-402.

Lyons, B., Arguile, S., Nikols, D., Godfrey, T. and Ellison, A., in prep., Highvale and Whitewood, Extended Geologic Mapping, prepared for TransAlta Utilities Corporation by Monenco Ltd., 128 pp.

McCabe, P.J., Langenberg, C.W., Macdonald, D.E., Richardson, R.J.H. and Strobl, R.S., (1987): Coal Geology Research Program, Status Report and Workplan For 1987 - 1989, Internal Report, 107 pp.

Mielke, P.W. Jr., Berry, K.J. and Brier, G.W., (1981): Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns, Monthly Weather Reviews, v. 109, p. 120-126.

Mielke, P.E., 1985. Geometric concerns pertaining to applications of statistical tests in the atmospheric sciences. Journal of Atmospheric Science, v. 42, p. 1209-1212.

Mielke, P.W., 1986: Non-metric statistical analysis: Some metric alternatives. Journal of Statistical Planning and Inference, v. 13, p. 377-387.

Nikols, D. and Lyons, B., (1985): Geology of the Highvale Mine, unpublished report presented at the CIM District 5 Meeting, 1985, 19 pp.

Nurkowski, J.R., (1985): Coal Quality and Rank Variation within the Upper Cretaceous and Tertiary Sediments Alberta Plains Region, Alberta Research Council Earth Science Report 85-1, 39 pp.

Rendu, J., (1982): Application of Geostatistics to Coal Evaluation, Proceedings Utah Geological and Mineral Survey Bulletin 118, 5th ROMOCO Symposium, pp. 251-261.

Richardson, R.J.H., Strobl, R.S., Macdonald, D.E., Nurkowski, J.R., McCabe, P.J. and Bosman, A., (1988): An evaluation of the coal resources of the Ardley coal zone to a depth of 400 m, in the Alberta Plains Area, Open File Report 1988-02, 96pp.

Shapiro, S.S. and Wilk, M.B., (1965): An Analysis of Variance Test for Normality (Complete Sample), Biometrika, v. 52, p. 591-611.

Taylor, J. (1985): Highvale Mine _in_ Coal in Canada, edited by T.H. Patching, Special Volume 31, Canadian Institute of Mining and Metallurgy, p. 164-167.

Tewalt, S.J., (1986): Chemical Characterization of Texas Lignite, Geological Circular 86-1, Bureau of Economic Geology, The University of Texas at Austin, 54 pp.

Table 1. Separation of TAU data into thematic files for statistical analysis. Note that each file is keyed using the unique drillhole number supplied for each record.

FILE NAME	CONTENTS
HV-ANALPROX.DAT	Results of proximate analysis. These data usually included moisture, volatile matter, ash, fixed carbon, sulphur, and calorific value on an "air dried" basis and "oven dried" basis.
HV-ANALEQUI.DAT	Equilibrium moisture, HGI, specific gravity, and bulk density values were given if available.
HV-ANALULT.DAT	Results of ultimate analysis. These data included determination in % of moisture, carbon, hydrogen, ash, sulphur, and oxygen by difference on an "air dried" basis.
HV-ASHFUSE.DAT	Results of analytical ash fusibility in an oxygenated atmosphere.
HV-DLS.DAT	Contains the drillhole identifier, and location in UTM and mine grid coordinates.
HV-GEOLITH.DAT	Lithology data including the stratigraphic codes for entire interval from surface to base of each drilling record.
HV-TRACE.DAT	Trace element analysis of S, Cl, F, Fe, Ca, TiO2, ... Zr (see appendix 1 for full listing).
HV-LATLONG.DAT	Contains the latitude and longitude coordinates for each drilling record.

Table 2. Description of Highvale coal seams and interburden (modified from Lyons et al., in prep).

INTERVAL	GENERALIZED DESCRIPTION	THICKNESS
SEAM 1	Coal, contains several bentonitic partings (0.1 to 0.4m). Highly variable ash content.	1.0 to 4.0 m
INTERBURDEN BETWEEN SEAM 1 and 2	Silty mudstone, grades laterally into a sandy siltstone. Thickness increases to the north and northeast.	0.2 to 0.6 m
SEAM 2	Coal, clean, laterally persistent	3.0 to 3.5 m
INTERBURDEN BETWEEN SEAMS 2 AND 3	Silty mudstone. Sharp increase in thickness to the north.	0.2 to 5.5 m
SEAM 3	Coal, splits into seams 3 and 3a to the south.	0.4 to 0.6 m
INTERBURDEN BETWEEN SEAMS 3 AND 4	Mudstone, silty, bentonitic with thin coal stringers up to 0.35 m, bentonitic at base.	1.0 to 1.2 m
SEAM 4	Coal, argillaceous interbeds, thins towards the south. Parting develops in the south.	0.3 to 0.85 m
INTERBURDEN BETWEEN SEAMS 4 AND 5	Mudstone, carbonaceous	0.5 to 0.8 m
SEAM 5	Coal, argillaceous, not economic to mine.	0.2 to 0.6 m
INTERBURDEN BETWEEN SEAMS 5 AND 6	Mudstone, carbonaceous, bentonitic.	1.2 to 1.6 m
SEAM 6	Coal, comprised of 2 subseams with 0.05 m bentonitic parting.	0.8 to 1.2 m

TABLE 3A

SUMMARY STATISTICS FOR PROXIMATE DATA (AIR-C)

H_2O	N	$\bar{X}$	Q_{50}	S	K	Mod	P-value
Seam 1	166	10.31	11.03	0.14	0.76	11.2	0.15
2	170	11.02	11.10	-0.27	-0.08	8.9	0.14
3	143	9.66	9.73	-0.19	-0.62	9.7	0.15
4	138	9.72	9.96	-0.28	-0.97	6.9	0.01
5	71	8.86	9.16	-0.39	-0.49	5.0	0.15
6	151	10.40	11.04	-0.49	-0.46	11.5	0.01

VOLM	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	124	30.25	29.97	0.50	0.64	28.8	0.10
2	125	31.57	31.40	0.17	1.31	30.7	0.05
3	117	29.29	29.53	-0.57	2.25	30.6	0.15
4	111	27.60	28.08	-1.09	1.31	28.8	0.01
5	48	24.14	24.18	-0.21	-0.37	20.4	0.46
6	125	30.39	30.43	-0.53	0.22	29.56	0.11

Table 3A. Summary statistics for moisture (H_2O) and volatile mater (VOLM) for seams 1 to 6. The data are from sampling category C (composite samples) on an air-dried basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3A (continued)

ASH	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	146	16.40	16.53	0.43	1.44	18.6	0.06
2	150	12.94	12.37	1.59	4.35	11.2	0.01
3	142	19.09	17.60	1.63	5.69	13.1	0.01
4	137	24.14	19.90	1.34	1.22	14.9	0.01
5	71	35.49	34.64	0.74	0.16	40.0	0.04
6	150	19.37	18.19	1.11	1.49	17.3	0.01

FXC	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	124	42.52	42.63	-0.96	2.62	42.3	0.05
2	125	44.65	44.68	-0.97	3.53	43.3	0.02
3	117	41.81	42.89	-1.63	5.89	46.0	0.01
4	111	39.53	42.05	-1.43	1.63	35.3	0.01
5	48	30.50	30.54	-0.32	-0.24	14.8	0.51
6	125	40.37	40.80	-0.73	0.65	35.3	0.13

Table 3A. Summary statistics for ash (ASH) and fixed carbon (FXC) for seams 1 to 6. The data are from sampling category C (composite samples) on an air-dried basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3A (continued)

S	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	116	0.22	0.21	1.25	2.65	0.20	0.01
2	121	0.18	0.16	1.68	4.09	0.14	0.01
3	113	0.30	0.29	1.47	3.36	0.28	0.01
4	111	0.34	0.34	0.37	3.05	0.33	0.01
5	52	0.36	0.36	-1.35	5.18	0.31	0.15
6	126	0.35	0.33	0.85	0.13	0.30	0.01

MJKG	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	121	20.71	20.62	-1.3	4.52	19.7	0.01
2	126	21.83	21.77	-1.43	7.94	21.6	0.01
3	117	20.76	21.35	-2.36	11.38	20.5	0.01
4	115	19.21	20.33	-1.21	0.63	14.1	0.01
5	51	16.34	17.07	-0.84	-0.04	17.0	0.01
6	133	20.58	20.8	-0.85	0.70	19.7	0.01

Table 3A. Summary statistics for total sulphur (S) and heat of combustion (MJKG) for seams 1 to 6. The data are from sampling category C (composite samples) on an air-dried basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3B

SUMMARY STATISTICS FOR PROXIMATE DATA
(DRY-C)

VOLM	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	115	33.91	33.9	0.003	0.27	34.0	0.15
2	116	35.36	35.4	-0.30	2.17	35.7	0.01
3	108	32.13	32.2	-0.65	2.06	31.8	0.15
4	101	30.30	31.3	-1.22	1.16	29.5	0.01
5	44	26.08	26.4	-0.41	-0.17	24.5	0.19
6	116	33.76	34.1	-0.93	0.90	30.2	0.01

ASH	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	143	18.29	18.43	0.59	1.55	13.3	0.08
2	148	14.64	13.87	3.60	21.98	13.0	0.01
3	143	21.11	19.34	1.76	6.29	21.9	0.01
4	135	26.73	22.15	1.20	0.68	14.7	0.01
5	71	38.46	37.99	0.49	-0.26	28.5	0.01
6	152	21.33	20.34	0.96	0.89	16.3	0.01

Table 3B. Summary statistics for volatile matter (VOLM) and ash content (ASH) for seams 1 to 6. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3B (continued)

FXC	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	115	47.41	47.73	-0.87	1.60	46.4	0.07
2	116	49.83	50.21	-1.30	3.02	49.6	0.01
3	108	45.78	46.82	-1.59	6.48	46.6	0.01
4	101	42.96	46.02	-1.42	1.49	46.0	0.01
5	44	33.12	33.39	0.69	2.66	30.0	0.13
6	116	44.63	45.03	-1.13	1.69	47.9	0.01

S	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	119	0.24	0.25	1.15	2.90	0.2	0.01
2	124	0.19	0.18	2.00	0.48	0.2	0.01
3	118	0.33	0.31	1.54	3.59	0.3	0.01
4	112	0.37	0.37	0.53	3.32	0.4	0.01
5	55	0.39	0.41	-1.66	6.29	0.4	0.15
6	129	0.38	0.37	0.81	0.09	0.3	0.01

Table 3B. Summary statistics for fixed carbon (FXC) and total sulphur (S) for seams 1 to 6. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3B (continued)

MJKG	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	118	23.21	23.18	-1.32	3.49	23.0	0.01
2	124	24.52	24.75	-2.36	9.44	24.1	0.01
3	118	23.04	23.65	-2.49	11.63	22.5	0.01
4	112	21.31	22.68	-1.29	0.84	22.0	0.01
5	51	18.11	19.24	-0.82	-0.11	12.4	0.02
6	134	23.09	23.49	-1.15	1.46	23.5	0.01

Table 3B. Summary statistics for heat of combustion (MJKG) for seams 1 to 6. The data are from sampling category C (composite samples) on a moisture-free (DRY) basis. The statistics are sample size (N), mean ($\bar{X}$), median (Q_{50}), skewness (S), kurtosis (K), mode and P-value (normality tests).

TABLE 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA
(AIR C)

H_2O	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	21.50	5.12	16.38	3.52	6.97	24.50
Seam 2	17.77	12.86	14.91	3.65	6.44	23.02
Seam 3	14.75	2.70	12.05	4.08	6.99	27.35
Seam 4	14.39	4.00	10.39	4.04	6.70	26.62
Seam 5	13.31	2.53	10.78	3.78	6.55	28.87
Seam 6	16.3	3.35	12.95	3.94	6.93	25.31

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA
(AIR C)

VOLM	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	36.05	25.45	10.60	2.38	3.17	5.89
Seam 2	36.31	26.00	10.31	1.89	2.68	5.19
Seam 3	37.00	17.31	19.69	3.21	8.03	9.67
Seam 4	33.04	17.18	15.86	3.04	9.16	10.69
Seam 5	29.5	17.89	11.61	4.38	8.19	11.85
Seam 6	34.85	23.77	11.08	2.57	4.09	6.65

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA
(AIR C)

ASH	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	30.89	9.35	21.54	4.9	13.64	22.51
Seam 2	24.46	6.83	17.63	2.43	6.19	19.24
Seam 3	60.52	2.6	57.92	9.04	56.70	39.43
Seam 4	61.04	10.2	50.84	12.98	111.27	43.69
Seam 5	61.5	18.81	42.69	13.56	92.02	27.02
Seam 6	38.94	8.97	29.97	5.87	24.99	25.80

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA
(AIR C)

FXC	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	51.59	28.6	22.99	4.14	62.42	8.29
Seam 2	49.65	33.44	16.21	2.69	5.74	5.36
Seam 3	49.61	12.32	37.29	7.20	30.10	13.12
Seam 4	49.1	14.08	35.02	8.00	51.47	18.49
Seam 5	43.01	14.83	28.18	9.74	43.70	21.67
Seam 6	47.32	27.36	19.96	5.14	14.26	9.35

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

Table 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA
(AIR C)

S	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	0.46	0.07	0.39	0.07	0.004	31.22
Seam 2	0.39	0.13	0.26	0.05	0.002	24.65
Seam 3	0.69	0.14	0.55	0.09	0.007	28.78
Seam 4	0.56	0.15	0.41	0.06	0.003	17.25
Seam 5	0.48	0.05	0.43	0.10	0.005	20.22
Seam 6	0.55	0.24	0.31	0.09	0.005	20.87

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3C

VARIABILITY STATISTICS FOR PROXIMATE DATA (AIR C)

MJKG	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	24.75	13.39	11.36	1.89	2.95	8.29
Seam 2	24.23	16.05	8.18	1.02	1.02	4.63
Seam 3	24.56	6.98	17.58	2.32	5.04	10.82
Seam 4	23.26	9.72	13.54	3.63	9.53	16.07
Seam 5	20.76	8.98	11.78	4.26	8.47	17.81
Seam 6	23.42	15.13	8.29	2.05	2.80	8.13

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3D

VARIABILITY STATISTICS FOR PROXIMATE DATA
(DRY C)

VOLM	Max.	Min.	Range.	I.Q.R.	VAR.	C.V.
Seam 1	38.60	28.99	9.61	2.49	3.01	5.11
Seam 2	39.50	30.73	8.77	1.63	2.18	4.18
Seam 3	40.55	19.20	21.35	3.88	9.89	9.78
Seam 4	36.14	18.74	17.40	4.20	12.78	11.79
Seam 5	31.45	18.86	12.59	4.01	10.21	12.25
Seam 6	37.27	26.56	10.71	2.39	4.05	5.96

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: coefficient of variation

TABLE 3D

VARIABILITY STATISTICS FOR PROXIMATE DATA
(DRY C)

ASH	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	35.69	9.92	25.77	5.67	17.67	22.97
Seam 2	41.23	7.89	33.34	2.79	12.55	24.19
Seam 3	67.13	9.03	58.10	9.98	65.94	38.45
Seam 4	63.59	11.09	52.50	13.62	132.14	43.14
Seam 5	64.74	20.82	43.92	16.18	108.24	27.04
Seam 6	42.06	10.42	31.64	6.77	33.28	27.04

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3D

VARIABILITY SATISTICS FOR PROXIMATE DATA (DRY C)

FXC	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	54.47	34.90	19.57	4.18	12.49	7.58
Seam 2	54.58	40.67	13.91	2.47	5.28	4.61
Seam 3	61.65	13.67	47.98	7.41	36.49	13.19
Seam 4	53.37	14.66	38.71	9.54	66.34	18.95
Seam 5	63.11	15.61	47.5	10.86	71.88	25.59
Seam 6	51.46	29.56	21.9	5.07	15.98	8.95

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3D

VARIABILITY STATISTICS FOR PROXIMATE DATA (DRY C)

S	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	0.54	0.02	0.52	0.07	0.006	33.04
Seam 2	0.47	0.14	0.33	0.07	0.003	25.23
Seam 3	0.78	0.16	0.62	0.11	0.009	29.58
Seam 4	0.65	0.16	0.49	0.07	0.004	18.36
Seam 5	0.52	0.05	0.47	0.10	0.006	19.47
Seam 6	0.60	0.24	0.36	0.12	0.006	20.80

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 3D

VARIABILITY STATISTICS FOR PROXIMATE DATA
(DRY C)

MJKG	Max.	Min.	Range.	I.Q.R.	Var.	C.V.
Seam 1	26.17	16.05	10.12	1.87	3.00	7.47
Seam 2	25.96	19.52	6.44	0.88	0.75	3.53
Seam 3	27.68	7.75	19.93	2.46	6.19	10.79
Seam 4	25.28	10.12	15.16	3.89	12.06	16.29
Seam 5	23.28	9.44	13.84	4.78	11.92	19.24
Seam 6	25.98	16.91	9.07	1.94	3.08	7.60

Max. Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V: Coefficient of variation

TABLE 4

LIST OF NORMALLY DISTRIBUTED VARIABLES FROM PROXIMATE DATA

PROXIMATE AIR-C	PROXIMATE DRY-C
Seam 1 H_2O	Seam 1 VOLM
Seam 2 H_2O	Seam 1 ASH
Seam 2 VOLM	Seam 1 FXC
Seam 3 H_2O	Seam 3 VOLM
Seam 3 VOLM	Seam 5 VOLM
Seam 5 H_2O	Seam 5 ASH
Seam 5 VOLM	Seam 5 FXC
Seam 5 FXC	Seam 5 S
Seam 5 S	
Seam 6 VOLM	
Seam 6 FXC	

TABLE 5

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	-0.49	-0.10	-0.52	-0.08	-0.50
VOLM		1.00 1.00	-0.53 -0.56(*)	0.42 0.20(*)	0.04 -0.04	0.42 0.10
ASH			1.00 1.00	-0.79 -0.92(*)	-0.16 -0.16	-0.72 -0.88
FXC				1.00 1.00	0.29 0.22	0.94 0.93
S					1.00 1.00	0.35 0.34
MJKG						1.00 1.00

Table 5. The Pearson product moment correlation matrix for seam 1, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 6

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	-0.60(*)	-0.15	-0.56	-0.10	-0.66
VOLM		1.00	-0.50	-0.48	0.12	0.64
		1.00	-0.52	0.25	0.09	0.33
ASH			1.00	-0.70	0.24	-0.57
			1.00	-0.67	0.13	-0.64
FXC				1.00	-0.27	0.88
				1.00	-0.33	0.85
S					1.00	-0.12
					1.00	-0.27
MJKG						1.00
						1.00

Table 6. The Pearson product moment correlation matrix for seam 2, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 7

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	-0.22(*)	-0.21	-0.17	0.06	-0.15
VOLM		1.00	-0.82	0.75	0.23	0.86
		1.00	-0.89	0.65	0.25	0.86
ASH			1.00	-0.92	-0.27	-0.91
			1.00	-0.87	-0.26	-0.98
FXC				1.00	0.32	0.95
				1.00	0.29	0.81
S					1.00	0.26
					1.00	0.24
MJKG						1.00
						1.00

Table 7. The Pearson product moment correlation matrix for seam 3, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 8

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	0.08	-0.25	0.16	0.13	-0.06
VOLM		1.00 1.00	-0.89 -0.94	0.87 0.88	0.07 0.13	0.93 0.94
ASH			1.00 1.00	-0.97 -0.99	-0.00 0.04	-0.96 -0.99
FXC				1.00 1.00	0.03 0.10	0.98 0.98
S					1.00 1.00	-0.04 -0.05
MJKG						1.00 1.00

Table 8. The Pearson product moment correlation matrix for seam 4, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 9

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	-0.02*	-0.43	0.13*	-0.13*	0.23
VOLM		1.00	-0.89	0.89(*)	0.29*	0.94
		1.00	-0.95*	0.79*	0.31*	0.95
ASH			1.00	-0.97	-0.17	-0.95
			1.00	-0.87(*)	-0.26*	-0.99
FXC				1.00	0.24*	0.98
				1.00	0.37*	0.84
S					1.00	0.23
					1.00	0.26
MJKG						1.00
						1.00

Table 9. The Pearson product moment correlation matrix for seam 5, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 10

	H_2O	VOLM	ASH	FXC	S	MJKG
H_2O	1.00	-0.50	-0.06	-0.32	-0.08	-0.42
VOLM		1.00	-0.65	0.66(*)	-0.07	0.84
		1.00	-0.82	0.63	-0.12	0.81
ASH			1.00	-0.88	0.35	-0.87
			1.00	-0.96	0.37	-0.96
FXC				1.00	-0.29	0.92
				1.00	-0.30	0.93
S					1.00	-0.21
					1.00	-0.30
MJKG						1.00
						1.00

Table 10. The Pearson product moment correlation matrix for seam 6, (proximate analysis). The first number in each cell is for AIR samples. The second number is for DRY samples. The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses.

TABLE 11

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.55136 0.0001 115	0.26925 0.00036 115	-0.1539 0.8848 91	0.32008 0.0021 90
ASH		-0.92429 0.0001 115	-0.19565 0.0330 119	-0.93838 0.0001 118
FXC			0.27494 0.0083 91	0.91205 0.0001 90
S				0.26816 0.0041 113

Table 11. Spearman correlation coefficient for the proximate DRY-C data of seam 1. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 12

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.59909 0.0001 116	0.15867 0.0889 116	0.02123 0.8408 92	0.25172 0.0155 92
ASH		-0.81374 0.0001 116	0.05717 0.5282 124	-0.75376 0.0001 124
FXC			-0.11702 0.2666 92	0.80123 0.0001 92
S				-0.04337 0.6396 119

Table 12. Spearman correlation coefficient for the proximate DRY-C data of seam 2. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 13

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.88820 0.0001 108	0.69097 0.0001 108	0.21942 0.0463 83	0.81308 0.0001 83
ASH		-0.90649 0.0001 108	-0.18057 0.0504 118	-0.97339 0.0001 118
FXC			0.28082 0.0101 83	0.86821 0.0001 83
S				0.17051 0.0710 113

Table 13. Spearman correlation coefficient for the proximate DRY-C data of seam 3. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 14

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.92051	0.83117	0.04648	0.90001
	0.0001	0.0001	0.6861	0.0001
	101	101	78	78
ASH		-0.97707	0.07908	-0.98841
		0.0001	0.4094	0.0001
		101	111	111
FXC			0.01933	0.95939
			0.8666	0.0001
			78	78
S				-0.05986
				0.5402

Table 14. Spearman correlation coefficient for the proximate DRY-C data of seam 4. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 15

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.94002 0.0001 44	0.86817 0.0001 44	0.41484 0.0282 28	0.91478 0.0001 24
ASH		-0.97082 0.0001 44	-0.21893 0.1083 55	-0.99063 0.0001 51
FXC			0.46704 0.0122 28	0.91565 0.0001 24
S				0.20814 0.1427 51

Table 15. Spearman correlation coefficient for the proximate DRY-C data of seam 5. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 16

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	ASH	FXC	S	MJKG
VOLM	-0.79499 0.0001 116	0.62753 0.0001 116	-0.13994 0.1834 92	0.76850 0.0001 98
ASH		-0.96056 0.0001 116	0.40998 0.0001 128	-0.97438 0.0001 133
FXC			-0.27997 0.0069 92	0.91838 0.0001 98
S				-0.33102 0.0001 128

Table 16. Spearman correlation coefficient for the proximate DRY-C data of seam 6. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 17

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	S	MJKG
H_2O	-0.46035 0.0001 124	-0.14485 0.0811 146	-0.40627 0.0001 124	-0.07922 0.3979 116	-0.38431 0.0001 121
VOLM		-0.51810 0.0001 124	0.45672 0.0001 124	-0.06417 0.5389 94	0.60193 0.0001 99
ASH			-0.81302 0.0001 124	-0.22657 0.0145 116	-0.79464 0.0001 121
FXC				0.32047 0.0016 94	0.88871 0.0001 94
S					0.29645 0.0013 115

Table 17. Spearman correlation coefficient for the proximate AIR-C data of seam 1. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 18

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	SULPHUR	MJKG
H_2O	-0.60891 0.0001 125	-0.18571 0.0229 150	-0.54870 0.0001 125	-0.25747 0.0044 125	-0.67472 0.0001 126
VOLM		-0.47380 0.0001 125	0.53382 0.0001 125	0.13580 0.1871 96	0.74946 0.0001 181
ASH			-0.62821 0.0001 125	0.14347 0.1164 121	-0.43184 0.0001 126
FXC				0.00393 0.9697 96	0.83150 0.0001 101
S					0.15072 0.0989 121

Table 18. Spearman correlation coefficient for the proximate AIR-C data of seam 2. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 19

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	S	MJKG
H_2O	-0.23795 0.0098 117	-0.25377 0.0023 142	-0.22227 0.0160 117	0.08816 0.3531 113	-0.25096 0.0063 117
VOLM		-0.79632 0.0001 117	0.73617 0.0001 117	0.17466 0.1036 88	0.81669 0.0001 92
ASH			-0.88035 0.0001 117	-0.18347 0.0518 113	-0.83375 0.0001 117
FXC				0.30819 0.0035 88	0.92654 0.0001 92
S					0.21252 0.0245 112

Table 19. Spearman correlation coefficient for the proximate AIR-C data of seam 3. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 20

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	S	MJKG
H_2O	-0.03157	-0.16657	-0.03937	0.01213	-0.22811
	0.7422	0.0517	-6816	0.8994	0.0142
	111	137	111	111	115
VOLM		-0.83803	0.83284	0.09635	0.90687
		0.0001	0.0001	0.3833	0.0001
		111	111	84	88
ASH			-0.90268	0.05226	-0.89551
			0.0001	0.5877	0.0001
			111	110	114
FXC				0.01816	0.93133
				0.8698	0.0001
				84	88
S					0.03415
					0.7232

Table 20. Spearman correlation coefficient for the proximate AIR-C data of seam 4. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 21

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	S	MJKG
H_2O	-0.04745 0.7488 48	-0.36355 0.0018 71	0.08034 0.5873 48	-0.17318 0.2195 52	0.13031 0.3621 51
VOLM		-0.87277 0.0001 48	0.87271 0.0001 48	0.41321 0.0259 29	0.91076 0.0001 28
ASH			-0.95832 0.0001 48	-0.12021 0.3959 52	-0.92794 0.0001 51
FXC				0.37699 0.0438 29	0.97769 0.0001 28
S					0.24553 0.0824 51

Table 21. Spearman correlation coefficient for the proximate AIR-C data of seam 5. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 22

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS

	VOLM	ASH	FXC	SULPHUR	MJKG
H_2O	-0.53075 0.0001 125	-0.02312 0.7788 150	-0.38731 0.0001 125	-0.08218 0.3603 125	-0.47200 0.0001 133
VOLM		-0.63575 0.0001 125	0.68542 0.0001 125	-0.04484 0.6578 100	0.82021 0.0001 107
ASH			-0.83471 0.0001 125	0.37998 0.0001 125	-0.81128 0.0001 125
FXC				-0.22771 0.0227 100	0.91929 0.0001 107
S					-0.22543 0.0112 126

Table 22. Spearman correlation coefficient for the proximate AIR-C data of seam 6. The second number in each entry represents the P-value from the significance test. The third number represents the sample size.

TABLE 23

	H_2O	VOLM	ASH	FXC	S	MJKS
Seam 1	-0.3342	0.3118	-0.4323	0.5184	0.2287	0.5322
Seam 2	-0.3658	0.4018	-0.3996	0.5011	-0.1076	0.5314
Seam 3	-0.0365	0.4510	-0.4979	0.4896	0.2156	0.5120
Seam 4	0.1756	0.4744	-0.5034	0.4915	0.0440	0.4972
Seam 5	0.2242	0.4654	-0.4916	0.4835	0.1466	0.4860
Seam 6	-0.1963	0.4440	-0.4687	0.4958	-0.1578	0.5233

Table 23. Coefficients for the different variables on the first principal component of the proximate AIR-C data.

TABLE 24

SUMMARY STATISTICS FOR ULTIMATE DATA (DRY-C) (AD)

WATER	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	27	12.47	11.52	0.50	-0.33	20.0	0.02
Seam 2	26	11.98	11.49	1.02	1.15	20.0	0.01
Seam 3	25	10.48	10.37	0.82	1.39	5.4	0.26
Seam 4	22	10.46	10.86	-0.38	-1.39	5.9	0.01
Seam 4	11	10.35	10.25	-0.54	-0.58	7.8	0.50
Seam 6	24	10.00	10.75	0.35	1.42	3.4	0.10

CARBON	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	30	53.85	53.06	0.47	0.99	43.75	0.63
Seam 2	31	55.73	55.61	-0.73	2.91	42.86	0.04
Seam 3	31	50.40	50.16	-0.08	-0.58	38.27	0.69
Seam 4	27	49.77	50.83	-1.11	0.89	50.03	0.01
Seam 5	13	38.77	40.26	-0.63	0.53	22.21	0.29
Seam 6	29	52.11	52.61	-0.49	0.71	39.84	0.66

TABLE 24 (continued)

HYDROGEN	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	30	4.06	4.34	-1.17	0.57	4.34	0.01
Seam 2	31	4.15	4.35	-0.66	-1.08	4.66	0.01
Seam 3	31	3.83	3.92	-0.53	-0.15	4.02	0.43
Seam 4	27	3.83	4.08	-0.10	-1.71	3.06	0.01
Seam 5	13	3.44	3.70	-0.80	-0.64	2.36	0.09
Seam 6	29	4.11	4.33	-0.40	-1.25	4.36	0.01

NITROGEN	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	30	0.69	0.68	-1.45	6.06	0.66	0.01
Seam 2	31	0.70	0.70	0.13	0.14	0.71	0.67
Seam 3	31	0.62	0.61	0.39	-0.18	0.69	0.44
Seam 4	27	0.62	0.63	-0.67	1.40	0.65	0.13
Seam 5	13	0.52	0.55	-0.48	-0.61	0.55	0.38
Seam 6	29	0.75	0.76	-0.19	-0.56	0.68	0.80

TABLE 24 (continued)

ASH	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	30	16.74	16.85	-0.31	1.01	9.40	0.76
Seam 2	31	14.05	13.59	1.10	1.12	9.80	0.02
Seam 3	31	22.29	20.93	0.57	-0.06	10.67	0.29
Seam 4	27	23.79	20.87	1.52	1.48	13.88	0.01
Seam 5	13	37.31	35.07	0.95	0.75	18.89	0.13
Seam 6	29	20.31	18.73	1.87	4.92	13.10	0.01

SULPHUR	N	$\bar{X}$	Q_{50}	S	K	Mode	P-value
Seam 1	33	0.22	0.20	1.16	1.55	0.20	0.01
Seam 2	32	0.17	0.16	1.69	3.04	0.15	0.01
Seam 3	31	0.27	0.27	0.69	0.09	0.23	0.21
Seam 4	27	0.34	0.34	1.12	1.73	0.34	0.03
Seam 5	13	0.31	0.32	-2.19	6.08	0.32	0.01
Seam 6	29	0.34	0.33	0.64	0.18	0.28	0.38

TABLE 25

LIST OF NORMALLY DISTRIBUTED VARIABLES FROM ULTIMATE DRY-C

Seam 3 water

Seam 1 carbon

Seam 3 carbon

Seam 5 carbon

Seam 6 carbon

Seam 3 hydrogen

Seam 2 nitrogen

Seam 3 nitrogen

Seam 4 nitrogen

Seam 5 nitrogen

Seam 6 nitrogen

Seam 1 ash

Seam 3 ash

Seam 5 ash

Seam 3 sulphur

Seam 6 sulphur

TABLE 26

VARIABILITY STATISTICS FOR ULTIMATE DATA (DRY-C) (AD)

WATER	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	20.00	5.28	14.71	4.49	19.25	35.17
Seam 2	20.00	6.61	13.39	3.16	12.81	29.88
Seam 3	20.00	5.39	14.61	5.15	10.85	31.43
Seam 4	14.08	5.91	8.17	4.97	8.17	27.31
Seam 5	12.12	7.77	4.35	2.16	1.97	13.55
Seam 6	20.00	3.35	16.65	4.88	13.03	36.11

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R.: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V.: Coefficient

TABLE 26 (continued)

VARIABILITY STATISTICS FOR ULTIMATE DATA
(DRY-C) (AD)

CARBON	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	65.46	43.75	21.71	5.05	19.45	8.19
Seam 2	62.91	42.86	20.05	3.19	15.70	7.11
Seam 3	62.63	38.27	24.36	10.02	37.00	12.07
Seam 4	59.44	32.04	27.40	6.26	49.87	14.19
Seam 5	53.49	22.21	31.28	9.96	79.11	22.94
Seam 6	61.41	39.84	21.57	5.05	22.70	9.14

Max: Maximum
Min: Minimum
Range: Maximum minus minimum
I.Q.R.: Interquantile range, 75th percentile minus 25th percentile
Var: Variance
C.V.: Coefficient of variation

TABLE 26 (continued)

VARIABILITY STATISTICS FOR ULTIMATE DATA (DRY-C) (AD)

HYDROGEN	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	4.76	2.37	2.39	1.00	0.37	14.90
Seam 2	5.00	2.78	2.22	1.22	0.44	16.02
Seam 3	4.83	2.26	2.57	0.99	0.40	16.50
Seam 4	4.79	2.94	1.85	1.29	0.41	16.73
Seam 5	4.23	2.36	1.87	1.02	0.37	17.72
Seam 6	4.82	3.01	1.81	1.02	0.32	13.80

Max:	Maximum
Min:	Minimum
Range:	Maximum minus minimum
I.Q.R.:	Interequantile range, 75th percentile minus 25th percentile
Var:	Variance
C.V.:	Coefficient of variation

TABLE 26 (continued)

VARIABILITY STATISTICS FOR ULTIMATE DATA
(DRY-C) (AD)

NITROGEN	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	0.84	0.33	0.51	0.09	0.01	13.84
Seam 2	0.88	0.52	0.36	0.12	0.01	12.06
Seam 3	0.78	0.49	0.29	0.13	0.01	11.60
Seam 4	0.80	0.36	0.44	0.10	0.01	15.68
Seam 5	0.66	0.35	0.31	0.14	0.01	18.65
Seam 6	0.96	0.52	0.44	0.15	0.01	14.54

Max:	Maximum
Min:	Minimum
Range:	Maximum minus minimum
I.Q.R.:	Interquantile range, 75th percentile minus 25th percentile
Var:	Variance
C.V.:	Coefficient of variation

TABLE 26 (continued)

VARIABILITY STATISTICS FOR ULTIMATE DATA
(DRY-C) (AD)

ASH	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	23.33	9.40	13.93	3.53	7.58	16.44
Seam 2	21.49	9.84	11.65	2.54	7.34	19.29
Seam 3	39.03	10.67	28.36	8.67	52.40	32.47
Seam 4	48.41	13.88	34.53	9.43	90.14	39.91
Seam 5	61.50	18.89	42.61	12.70	155.68	33.44
Seam 6	38.94	13.10	25.84	4.50	26.94	25.55

Max:	Maximum
Min:	Minimum
Range:	Maximum minus minimum
I.Q.R.:	Interquantile range, 75th percentile minus 25th percentile
Var:	Variance
C.V.:	Coefficient of variation

TABLE 26 (continued)

VARIABILITY STATISTICS FOR ULTIMATE DATA
(DRY-C) (AD)

SULPHUR	MAX	MIN	RANGE	I.Q.R.	VAR	C.V.
Seam 1	0.40	0.10	0.30	0.07	0.004	30.39
Seam 2	0.31	0.12	0.19	0.04	0.002	23.70
Seam 3	0.41	0.19	0.22	0.08	0.003	20.55
Seam 4	0.49	0.26	0.23	0.07	0.003	16.35
Seam 5	0.41	0.05	0.36	0.07	0.008	29.03
Seam 6	0.50	0.24	0.26	0.09	0.004	12.85

Max:	Maximum
Min:	Minimum
Range:	Maximum minus minimum
I.Q.R.:	Interquantile range, 75th percentile minus 25th percentile
Var:	Variance
C.V.:	Coefficient of variation

TABLE 27

	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	-0.73	-0.10	-0.58	0.18	-0.068
CARBON		1.00	-0.06	0.42	-0.63(*)	0.43
HYDROGEN			-1.00	0.19	0.16	-0.09
NITROGEN				1.00	-0.34	0.39
ASH					1.00	-0.20
SULPHUR						1.00

Table 27. The Pearson correlation matrix for seam 1 (ultimate analysis). The correlation coefficients from a pair of normally distributed variables are indicated with an asterick. The correlation coefficients which are significient at the usual 5% level are denoted with the same asterisks in parentheses. The data are from DRY-C samples.

TABLE 28

	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	-0.51	-0.01	-0.44	-0.19	-0.39
CARBON		1.00	0.03	0.43	-0.56	0.31
HYDROGEN			1.00	0.29	-0.10	-0.41
NITROGEN				1.00	-0.33	0.09
ASH					1.00	-0.22
SULPHUR						1.00

Table 28. The Pearson correlation matrix for seam 2 (ultimate analysis). The correlation coefficients froma pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses. The data are for DRY-C samples.

TABLE 29

	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	0.06	0.40	-0.12	-0.44	0.08
CARBON		1.00	0.20	0.21	-0.89	0.36
HYDROGEN			1.00	0.49	-0.41	-0.11
NITROGEN				1.00	-0.24	-0.18
ASH					1.00	-0.26
SULPHUR						1.00

Table 29. The Pearson matrix for seam 3 (ultimate analysis). The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses. The data are from DRY-C samples.

TABLE 30

	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	0.49	0.71	0.22	-0.71	-0.37
CARBON		1.00	0.15	0.49	-0.94	-0.01
HYDROGEN			1.00	0.23	-0.43	-0.27
NITROGEN				1.00	-0.53	0.14
ASH					1.00	0.10
SULPHUR						1.00

Table 30. The Pearson correlation matrix for seam 4 (ultimate analysis). The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses. The data are from DRY-C samples.

TABLE 31

WATER	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	0.75	0.87	0.66	-0.81	0.36
CARBON		1.00	0.46	0.88(*)	-0.99(*)	0.04
HYDROGEN			1.00	0.45	-0.54	0.43
NITROGEN				1.00	-0.87(*)	0.24
ASH					1.00	-0.08
SULPHUR						1.00

Table 31. The Pearson correlation matrix for seam 5 (ultimate analysis). The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses. The data are from DRY-C samples.

TABLE 32

	WATER	CARBON	HYDROGEN	NITROGEN	ASH	SULPHUR
WATER	1.00	-0.21	0.41	-0.01	-0.35	-0.48
CARBON		1.00	-0.25	0.18*	-0.71	0.21*
HYDROGEN			1.00	0.25	-0.14	-0.17
NITROGEN				1.00	-0.44	0.23*
ASH					1.00	-0.10
SULPHUR						1.00

Table 32. The Pearson correlation matrix for seam 6 (ultimate analysis). The correlation coefficients from a pair of normally distributed variables are indicated with an asterisk. The correlation coefficients which are significant at the usual 5% level are denoted with the same asterisks in parentheses. The data are from DRY-C samples.

TABLE 33

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO=0/NUMBER OF OBSERVATIONS=30

	C	H	N	ASH	S
H_2O	-0.73137	-0.09278	-0.57699	0.17741	-0.68262
	0.0001	0.6258	0.0009	0.3483	0.0001
C		-0.5744	0.41947	-0.62759	0.43304
		0.7630	0.0210	0.0002	0.0168
H			0.19227	0.15829	-0.08656
			0.3087	0.4035	0.6492
N				-0.34993	0.38803
				0.00580	0.0341
ASH					-0.19920
					0.2913

Table 33. Spearman correlation coefficients for ultimate AD (DRY-C) data of seam 1. The second number in each entry represents the P-value from the significance test.

TABLE 34

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS=31

	C	H	N	ASH	S
H_2O	-0.53724	0.34221	-0.31963	-0.17746	-0.43182
	0.0018	0.0595	0.0796	0.3395	0.0153
C		0.00081	0.47053	-0.56653	0.27854
		0.9966	0.0076	0.0009	0.1292
H			0.25177	0.01835	-0.37890
			0.1718	0.9220	0.0355
N				-0.26529	0.03237
				0.1492	0.8628
ASH					-0.11121
					0.5514

Table 34. Spearman correlation coefficients for ultimate AD (Dry-C) data of seam 2. The second number in each entry represents the P-value from the significance test.

TABLE 35

SPEARMAN'S CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO=0/NUMBER OF OBSERVATIONS=31

	C	H	N	ASH	S
H_2O	-0.05665	0.61838	0.00528	-0.30552	0.04261
	0.7621	0.0002	0.9775	0.0946	0.8200
C		0.12038	0.25890	-0.85887	0.37566
		0.5189	0.1596	0.0001	0.0373
H			0.46850	-0.30487	-0.05874
			0.0079	0.0954	0.7536
N				-0.26618	-0.25467
				0.1478	0.1668
ASH					-0.21472
					0.2461

Table 35. Spearman correlation coefficients for ultimate AD (Dry-C) data of seam 3. The second number in each entry represents the P-value from the significance test.

TABLE 36

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO=0/NUMBER OF OBSERVATIONS=27

	C	H	N	ASH	S
H_2O	0.14454 0.4719	0.81268 0.0001	0.23591 0.2362	-0.47091 0.0132	-0.34824 0.0751
C		-0.01924 0.9241	0.46325 0.0150	-0.87193 0.0001	0.15691 0.4344
H			0.35340 0.0706	-0.28432 0.1506	-0.19522 0.3290
N				-0.56378 0.0022	0.24347 0.2210
ASH					-0.08672 0.6671

Table 36. Spearman correlation coefficients for ultimate AD (Dry-C) data of seam 4. The second number in each entry represents the P-value from the significance test.

TABLE 37

SPEARMAN'S CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS=13

	C	H	N	ASH	S
H_2O	0.37689 0.2043	0.88308 0.0001	0.39724 0.1789	-0.45392 0.1192	0.33058 0.2699
C		0.35714 0.2309	0.73279 0.0044	-0.97253 0.0001	0.18432 0.5466
H			0.35262 0.2373	-0.44505 0.1275	0.41541 0.1581
N				-0.74105 0.0038	0.49103 0.0884
ASH					-0.27785 0.3580

Table 37. Spearman correlation coefficients for ultimate AD (Dry-C) data of seam 5. The second number in each entry represents the P-value from the significance test.

TABLE 38

SPEARMAN'S CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0/NUMBER OF OBSERVATIONS=29

	C	H	N	ASH	S
H_2O	-0.32099	0.56547	-0.00594	-0.23901	-0.56634
0.0895	0.0895	0.0014	0.9756	0.2118	0.0014
C		-0.40143	0.26185	-0.60985	0.33457
		0.0309	0.1700	0.0004	0.0761
H			0.22630	-0.12962	-0.29558
			0.2378	0.5028	0.1195
N				-0.50691	0.23639
				0.0050	0.2170
ASH					-0.16988
					0.3783

Table 38. Spearman correlation coefficients for ultimate AD (Dry-C) data of seam 6. The second number in each entry represents the P-value from the significance test.

TABLE 39

SUMMARY STATISTICS OF ULTIMATE (DAF) DATA

OXYGEN	N	$\bar{X}$	S	Q_{50}	MIN	MAX
Seam 1	26	10.69	8.98	6.87	0.74	27.13
Seam 2	28	10.49	9.04	7.77	0.89	27.88
Seam 3	27	11.26	9.76	9.89	0.21	31.00
Seam 4	21	11.36	10.56	9.32	0.01	30.12
Seam 5	10	9.61	13.47	3.13	0.13	34.87
Seam 6	25	16.81	9.27	8.81	0.32	27.60

HYDROGEN	N	$\bar{X}$	S	Q_{50}	MIN	MAX
Seam 1	26	4.16	0.95	4.38	1.05	5.46
Seam 2	28	4.14	0.93	4.43	2.37	5.30
Seam 3	27	4.18	0.92	4.45	2.47	5.49
Seam 4	21	4.25	0.98	4.34	2.66	5.91
Seam 5	10	4.69	1.18	4.77	2.27	6.24
Seam 6	25	4.46	0.85	4.74	2.76	5.60

TABLE 39 (continued)

SUMMARY STATISTICS OF ULTIMATE (DAF) DATA

CARBON	N	$\bar{X}$	S	Q_{50}	MIN	MAX
Seam 1	26	72.73	3.14	74.03	66.35	77.47
Seam 2	28	72.93	3.47	74.22	66.01	77.96
Seam 3	27	72.56	4.03	74.33	62.77	77.21
Seam 4	21	72.70	5.20	74.93	62.11	79.73
Seam 5	10	70.39	6.65	72.80	57.29	75.16
Seam 6	25	72.95	3.84	74.29	65.09	78.81

NITROGEN	N	$\bar{X}$	S	Q_{50}	MIN	MAX
Seam 1	26	0.92	0.11	0.95	0.48	1.00
Seam 2	28	0.92	0.09	0.93	0.64	1.03
Seam 3	27	0.89	0.10	0.89	0.67	1.07
Seam 4	21	0.90	0.13	0.90	0.58	1.15
Seam 5	10	0.98	0.07	1.00	0.84	1.06
Seam 6	25	1.04	0.14	1.07	0.66	1.18

TABLE 39 (continued)

SUMMARY STATISTICS OF ULTIMATE (DAF) DATA

SULPHUR	N	$\bar{X}$	S	Q_{50}	MIN	MAX
Seam 1	26	0.29	0.07	0.29	0.20	0.50
Seam 2	28	0.23	0.05	0.22	0.16	0.39
Seam 3	27	0.39	0.08	0.36	0.28	0.55
Seam 4	21	0.52	0.12	0.48	0.41	0.93
Seam 5	10	0.60	0.20	0.65	0.08	0.76
Seam 6	25	0.48	0.08	0.48	0.37	0.71

TABLE 40

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO=0 NUMBER OF OBSERVATIONS=26

	H	C	N	S
O	0.31469 0.1174	-0.50265 0.0089	-0.04728 0.8186	0.50034 0.0092
H		-0.51334 0.0073	0.47112 0.0151	0.12836 0.5320
C			-0.39499 0.0458	-0.37759 0.0572
N				-0.14333 0.4849

Table 40. Spearman correlation coefficients for ultimate DAF data of seam 1. The second number in each entry represents the P-value from the significance test.

TABLE 41

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO:0 NUMBER OF OBSERVATIONS=28

	H	C	N	S
O	0.23926 0.2201	-0.69787 0.0001	-0.16201 0.4101	0.46977 0.0117
H		-0.36600 0.0554	0.40348 0.0332	-0.08695 0.6600
C			-0.08936 0.6511	-0.33118 0.0852
N				-0.27308 0.1597

Table 41. Spearman correlation coefficients for ultimate DAF data of seam 2. The second number in each entry represents the P-value from the significance test.

TABLE 42

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER H_0:RHO=0 NUMBER OF OBSERVATIONS=27

	H	C	N	S
O	0.25561 0.1981	-0.63675 0.0004	-0.021357 0.9071	-0.30219 0.1255
H		-0.49229 0.0091	0.54892 0.0030	-0.42993 0.0252
C			-0.29966 0.1289	0.17617 0.3794
N				-0.21438 0.2829

Table 42. Spearman correlation coefficients for ultimate DAF data of seam 3. The second number in each entry represents the P-value from the significance test.

TABLE 43

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0:RHO=0 NUMBER OF OBSERVATIONS=21

	N	C	N	S
O	0.33517 0.1375	-0.66234 0.0011	0.21059 0.3595	0.45381 0.0388
H		-0.68659 0.0006	0.39532 0.0761	0.42652 0.0538
C			-0.29509 0.1941	-0.40157 0.0712
N				0.12713 0.5829

Table 43. Spearman correlation coefficients for ultimate DAF data of seam 4. The second number in each entry represents the P-value from the significance test.

TABLE 44

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0: RHO=0 NUMBER OF OBSERVATIONS=10

	H	C	N	S
O	0.24848 0.4888	-0.60000 0.0667	-0.65855 0.0384	-0.27356 0.4444
H		-0.76970 0.0092	0.25610 0.4751	0.71125 0.0211
C			0.25610 0.4751	-0.34651 0.3267
N				0.39450 0.2593

Table 44. Spearman correlation coefficients for ultimate DAF data of seam 5. The second number in each entry represents the P-value from the significance test.

TABLE 45

SPEARMAN CORRELATION COEFFICIENTS/PROB>|R| UNDER
H_0: RHO=0 NUMBER OF OBSERVATIONS=25

	H	C	N	S
O	0.21047	-0.69359	-0.20259	0.46072
	0.3126	0.0001	0.3315	0.0205
H		-0.49009	0.18406	0.16059
		0.0129	0.3784	0.4432
C			-0.15586	-0.33385
			0.4569	0.1029
N				-0.00097
				0.9963

Table 45. Spearman correlation coefficients for ultimate DAF data of seam 6. The second number in each entry represents the P-value from the significance test.

TABLE 46

SUMMARY STATISTICS OF THE HYDROGEN/CARBON RATIO

	N	$\bar{X}$	Q_{50}	S^2	MAX	MIN	RANGE
Seam 1	30	0.9065	0.9534	0.0252	1.1841	0.5868	0.5973
Seam 2	31	0.8939	0.9446	0.0237	1.1530	0.5999	0.5531
Seam 3	31	0.9146	0.9755	0.0264	1.1369	0.6532	0.4837
Seam 4	27	0.9333	1.0224	0.0307	1.1363	0.6001	0.5362
Seam 5	13	1.0926	1.1146	0.0386	1.3006	0.6092	0.6914
Seam 6	29	0.9524	1.0241	0.0292	1.3002	0.6350	0.6652

APPENDIX 1. HIGHVALE COAL QUALITY DATA IN THE TRANSALTA UTILITIES "VERIFY FORMAT"

```
Drill Hole Number                  :   HV 86 19

Location : Area                    :   HV
         : NTS                     :   83G

         : Lsd-Sec-T-R-M           :   8-25-52-5-5      from the NE corner
                                       of section 24 ;  0000689.63 meters N , and
                                                       -0000128.69 meters E

         : UTM : Reference Meridian :  000 ;  0000000.00 meter North
                                               0000000.00 meter East

Elevation (m AMSL)                 : 0755.07

Use                                :  DEV

Drill Data : Contractor            :  MCAUL
           : Drilling Method       :  ROTCORW

           : Hole Particulars      : Casing : Diameter (cm)  :  000
                                            : Depth (m)      :  00.0

                                   : Hole Diameter (cm)      :  014

                                   : Directional  : Dip      :  00
                                                  : Deviation :  00
                                                  : Azimuth  :  000

                                   : Total Depth (m)         :  049.00

                                   : Completion Date         :  1986-5-08

                                   : Depth of Till (m)       :  001.20

                                   : Core Recovery % : Coal  :  100.0
                                                     : Rock  :  089.0

Geophysical Data
           : Geophysical Contractor  :  BPB

           : Log Run

                      Caliper
                      Resistivity
                      Density(Long/Linear spaced)
                      Density(Short spaced)
                      Dual Neutron
                      Sonic (Short Spacing)
                      Sonic (Long Spacing)
                      Gamma
Analytical Data:
           : Analytical Laboratory(ies):  BIRTLEY

           : Analyses Run

                      Proximate
                      Coal Ash Analysis
                      Heating Value
                      Sulfur
                      Trace Elements
```

Drill Hole Number : HV 86 18

7-Mar-1987

ANALYTICAL PROXIMATE

DEPTH													
FROM	TO	INTcm	STRAT CODE	SAMPLED SEGMENT	METHOD	H2O	VOLM	ASH	FxC	S	Btu/lb	MJ/kg	Cal/gm
6.35	9.56	321	2	C	RCD	00.00	00.00	00.00	00.00	0.00	00000	00.00	0.00
					AIR	14.10	30.70	11.50	43.70	0.23	09238	21.49	5173.28
					DRY	00.00	35.74	13.39	50.87	0.27	10754	25.01	6022.24
					ADJ	00.00	00.00	00.00	00.00	0.00	00000	00.00	0.00
11.89	12.68	79	4	C	RCD	00.00	00.00	00.00	00.00	0.00	00000	00.00	0.00
					AIR	13.00	27.60	18.20	41.20	0.38	08550	19.89	4788.00
					DRY	00.00	31.72	20.92	47.36	0.44	09828	22.86	5503.68
					ADJ	00.00	00.00	00.00	00.00	0.00	00000	00.00	0.00

Core Recovery

Geological Data:

Geologists' Interpreted Lithologic Log Completed
by (Interpreter) : BL
on (Date) : 1986-5-11

Hydro Data:

Water Table Depth has been determined.
There were 02 piezometers set in the drill hole.

Soil Data:

Drill Hole Remarks:

Drill Hole Number : HV 86 18

7-Mar-1987

Analytical Equilibrium H2O ; HGI

DEPTH FROM	TO	INTcm	STRAT CODE	SAMPLED SEGMENT	EQUILIBRIUM H2O	HGI	SPECIFIC GRAVITY	BULK DENSITY
6.35	6.55	20	2	G	000.00	000	001.64	000.00
6.35	9.56	321	2	C	000.00	000	001.44	000.00
6.55	7.68	113	2	G	000.00	000	001.46	000.00
7.68	8.95	127	2	G	000.00	000	001.39	000.00
8.95	9.56	61	2	G	000.00	000	001.37	000.00
11.89	12.68	79	4	C	000.00	000	001.48	000.00

```
Drill Hole Number                     :  HV 86 18
 7-Mar-1987
                                                  Analytical Trace Elements
                                                                                     NOTE: S,CI,F,As,B,Br,Hg,I and Se are
                                                                                          determined from 'COAL'; the remainder
      DEPTH                     STRAT  SAMPLED                                            are determined from 'ashed coal'
  FROM       TO        INTcm   CODE   SEGMENT  <------%------>      PPM (parts per million)

   6.35      9.56       321      2       C      S     000.000  As    000.000    Mo    000.000
                                                CI    000.007  Ag    000.000    Mn    000.000
                                                F     000.000  Be    000.000    Ni    000.000
                                                Fe    000.000  B     000.000    Pb    000.000
                                                Ca    000.000  Br    000.000    Se    000.000
                                                Ti02  000.000  Cd    000.000    Sr    000.000
                                                               Cr    000.000    Th    000.000
                                                               Cu    000.000    U308  000.000
                                                               Hg    000.000    V     000.000
                                                               I     000.000    Zn    000.000
                                                                                Zr    000.000

  11.89     12.68        79      4       C      S     000.000  As    000.000    Mo    000.000
                                                CI    000.004  Ag    000.000    Mn    000.000
                                                F     000.000  Be    000.000    Ni    000.000
                                                Fe    000.000  B     000.000    Pb    000.000
                                                Ca    000.000  Br    000.000    Se    000.000
                                                Ti02  000.000  Cd    000.000    Sr    000.000
                                                               Cr    000.000    Th    000.000
                                                               Cu    000.000    U308  000.000
                                                               Hg    000.000    V     000.000
                                                               I     000.000    Zn    000.000
                                                                                Zr    000.000
```

Drill Hole Number : HV 86 18

7-Mar-1987

Analytical Ash Analysis

DEPTH															
FROM	TO	INTcm	STRAT CODE	SAMPLED SEGMENT	SiO2	AlO3	TiO2	Fe2O3	CaO	MgO	Na2O	K2O	P2O5	SO3	Undet
6.35	9.56	321	2	C	000.00	000.00	00.00	00.00	000.00	00.00	00.21	00.41	00.00	00.00	00.00
11.89	12.68	79	4	C	000.00	000.00	00.00	00.00	000.00	00.00	01.29	00.34	00.00	00.00	00.00

Drill Hole Number : HV 65 203 A

26-Feb-1987

Analytical Ash Fusibility

DEPTH			STRAT	SAMPLED									
FROM	TO	INTcm	CODE	SEGMENT		IDT-F	IDT-C	ST-F	ST-C	HT-F	HT-C	FT-F	FT-C
18.07	19.08	101		C	RED ATM	0000	0000	0000	0000	0000	0000	0000	0000
					OXY ATM	0000	0000	2375	1302	0000	0000	2450	1343
19.08	20.80	172		C	RED ATM	0000	0000	0000	0000	0000	0000	0000	0000
					OXY ATM	0000	0000	2150	1177	0000	0000	2200	1204
21.34	24.51	317		C	RED ATM	0000	0000	0000	0000	0000	0000	0000	0000
					OXY ATM	0000	0000	2300	1260	0000	0000	2450	1343
24.90	25.40	50		C	RED ATM	0000	0000	0000	0000	0000	0000	0000	0000
					OXY ATM	0000	0000	2250	1232	0000	0000	2350	1288

Drill Hole Number : HV 63 192

26-Feb-1987

Analytical Ultimate

DEPTH			STRAT	SAMPLED	<——IN % , AIR DRIED——>						OXY BY	CORE
FROM	TO	INTcm	CODE	SEGMENT	H2O	C	H	N	ASH	SUL	DIFF	REC %
10.80	13.90	310		C	12.70	49.40	03.30	0.60	18.60	00.20	15.00	00.00
14.40	17.50	310		C	13.20	54.80	03.60	0.60	09.40	00.15	18.20	00.00

Drill Hole Number : HV 86 18

7-Mar-1987

Geological Lithology

DEPTH			STRAT								
FROM	TO	INTcm	CODE	FRAME	MATRX	GNSZ	IND	CEM	STRUCT	COLOR	MATRIX %
0.00	0.18	18	20	101						746	
0.18	1.40	122	20	102	202	306				735	
1.40	5.89	449	1	115					619	741	
5.89	6.35	46		108	207	306				755	
6.35	9.56	321	2	115						741	
9.56	10.29	73		107		306			604	746	
10.29	10.90	61	3	115						741	
10.90	11.51	61		107		306			601	717	
11.51	11.58	7	31	115						741	
11.58	11.89	31		107	206	306			601	746	
11.89	12.68	79	4	115						741	
12.68	13.00	32		107		306			620	745	
13.00	13.23	23		109	208	305				746	
13.23	13.53	30	5	115						741	
13.53	13.75	22		107		306				746	
13.75	14.65	90		107		306				755	
14.65	14.90	25		114	207	306				744	
14.90	15.15	25		109	208	305			604	746	
15.15	15.86	71	6	115						741	
15.86	15.90	4	6	107	207	306				755	
15.90	16.27	37	6	115						741	
16.27	16.33	6		107					620	755	
16.33	20.00	367		109		305	411			754	

APPENDIX 2 MATHEMATICAL REPRESENTATION OF THE STATISTICAL METHODS USED.

Appendix 2: Statistical Equations

In this appendix, the statistical equations are listed for reference. The actual computations are performed on the Alberta Research Council Computer VAX computers, using existing software packages including SAS. Most equations are discussed either in the main text of this report or in standard statistical textbooks, therefore, no detailed discussions are included in this section.

The sample mean :

$$\overline{x} = \sum_{i=1}^{n} x_i/n \tag{1}$$

The sample median : This is represented by Q_{50}, which can be defined as

$$\int_{-\infty}^{Q_{50}} f(x)dx = \int_{Q_{50}}^{\infty} f(x)dx = 0.5 \tag{2}$$

The sample variance :

$$s^2 = \sum_{i=1}^{n}(x-\overline{x})^2/(n-1) \tag{3}$$

The coefficient of skewness S:

$$S = \hat{\mu}_3/(s^2)^{3/2} \tag{4}$$

The coefficient of kurtosis K:

$$K = \hat{\mu}_4/(s^2)^2 \tag{5}$$

where $\hat{\mu}_i$ is the sample estimate of the ith central moment.

The Kolmogorov-Smirnov test for normality is also called the Lilliefor's test when the mean and variance are estimated from the sample rather than known. The test statistic is:

$$D = \sup_{x} |S_n(x) - F_X(x)| \tag{6}$$

where $S_n(x)$ is the sample distribution function and $F_X(x)$ is the normal distribution function having the estimated mean and variance.

The Shapiro-Wilk test for normality is based on the statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{7}$$

where a_i are weights derived from the n × n covariance matrix of the n observations.

The sample correlation ceofficient (Pearson's product moment) :

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2)(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2)}} \tag{8}$$

The sample Spearman's rank correlation coefficient (Spearman's RHO) is calculated using the equation for the Pearson's product moment correlation coefficient, except that the values of the observations are replaced with the corresponding ranks from 1 to n for a sample of n pairs.

The principle components are eigen vectors $\vec{a}_i$ of the sample covariance matrix C, satisfying the equation

$$|C - \lambda_i I| \vec{a}_i = \vec{0} \tag{9}$$

where I is the identity matrix, and $\vec{0}$ the null vector.

The multi-response permutation procedure (MRPP) is a multivariate comparison technique based on the Euclidean distance between data points in a multi-dimensional space. The test statistic is

$$\delta = \sum_{h=1}^{g} C_h \xi_h \tag{10}$$

where C_h are the appropriate weights and ξ_h are average distances between data points from a total of g groups.