PREVOST, X.M. Present status and applications of the National Coal Data Base. APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries. Volume 3: *Geostatistics.* Johannesburg, SAIMM, 1987. pp. 165 – 174.

Present Status and Applications of the National Coal Data Base

X.M. PREVOST

Fossil Fuels Division, Geological Survey of South Africa, Pretoria

A computer system to store, retrieve and process coal borehole and analytical data together with other geological, geographical or topographical information is described. As a result of the large volume and variety of data stored since its inception in 1979, as well as the basic versatility of the system, the National Coal Data Base has been performing a number of new tasks for which it was not originally designed.

As considerable sets of data are used on almost every project, the use of statistical application programs has become an important processing tool.

Introduction

In South Africa, the logs of all boreholes drilled for exploration purposes have to be submitted to the Department of Mineral and Energy Affairs, and the Geological Survey acts as custodian of this information. The logs of some holes drilled for mine-planning purposes are also submitted voluntarily to fill any possible gaps that the exploration holes might have left.

With the abundance of coal information accumulated on file, it became obvious that something other than manual retrieval and processing was imperative. In 1978 a trial attempt was made to computerise borehole data[1], but the resources available at the time were, however, insufficient. Only towards the end of 1979 was the creation of the actual National Coal Data Base initiated. The beginnings were difficult and much work was needed not only to design a comprehensive and flexible data base, but also to add a host of application packages and programs which would eventually make the processing of coal data possible[2-4]. One of the last but not the least of the requirements was the acquisition of a gridding and contouring package that could not only produce contours with a smooth minimum curvature between data points, but also provide gridding routines that would give an accurate surface reconstruction technique which would honour the original data. This was achieved after a lengthy test-and-trial process.

Finally, the information for the

last of the major coal fields began to be captured in 1986. A review of the work achieved and a forecast of future needs was then compiled[5]. The Data Base now contains approximately 25 000 boreholes comprising more than one million header, lithological and analytical records. It includes comprehensive information for the Witbank, Highveld, Eastern Transvaal, Northern Natal and Orange Free State coal fields and has become the major coal data processing facility in South Africa and one of the biggest in the world.

Data collection

The borehole data originate mainly from 15 large coal-prospecting and mining companies, and a multitude of small companies and individuals, most of which have only begun prospecting for coal recently, following the boom experienced in the coal industry. An average of 300 borehole logs arrive and are indexed monthly.

The source documents used to create the Data Base are not only diverse in their format but also use different nomenclature in their description of the lithologies and coal seams or zones intersected. To attain homogeneity, all logs therefore need to be scrutinised at the point of capturing. The most important part of this exercise is to arrive at a standard coal seam labelling, for if this is not achieved any attempt to calculate coal resources will be futile.

To complement the data contained in the boreholes logs, geological maps showing the presence or absence of coal-bearing strata and topographic maps displaying coal seam outcrop lines are digitised and then used to establish the limits of individual blocks of coal.

Coding

The information from the borehole logs is transcribed into three different type of documents.

The A form contains fields that identify and locate the boreholes, with a unique record number which contains a check digit at the end. The coordinates are recorded in the South African Lo system along with the collar elevation.

The form G is used to enter the detailed sequence of rock units intersected by the borehole, including lithological descriptions, depth and thickness. In the case of coal seams, seamlets or zones, the name of the seam is recorded and below it each individual sample is entered with a sample number that identifies it and links it to the analytical information obtained from it.

The form P contains all available analytical data for the seams intersected and sampled in the borehole. It comprises items such as proximate analysis, heat value, coking tests, sulphur and phosphorus. Samples not analysed but for which average values can be estimated from their sample descriptions are defaulted.

Data management

To create a comprehensive data base it is important first to decide what data elements must be stored to produce the output required. This is a difficult task at all times and it becomes more complex in view of continuous changes in output needed by a continuously changing technology. It is therefore essential to have a data base management system that permits structured changes to be performed easily and inexpensively. The National Coal Data Base has changed several times since its inception and recently, two data levels were removed and a new one created.

It is also imperative to correct most of the data errors before the information enters the data base. For this reason the data verification and validation routines are probably its most time-consuming aspect, even though most of the checking is done by the computer logic unit. The verification is done in two stages. Firstly, the information captured is compared with table files which check the validity of each alpha field or against an upper and lower limit in the case of numerical data. The same data are then recaptured and compared with the first set to eliminate typing errors. Validation is performed once the data files are sent to the mainframe to be loaded on the Data Base. The errors corrected there are more difficult to detect and usually pass unnoticed in the verification check.

The data base

The information stored in the Data Base is identified by a unique number, a set of X and Y coordinates and the collar elevation (Figure 1).

The level where these data are located is called BORREC (borehole records). These records contain what is known as header data. The detailed lithological descriptions of the rocks intersected are entered on the GEOREC level, or geological records.

The analytical results obtained for all coal samples tested for their chemical or physical characteristics are loaded on the level called SAMREC (sampling records) or analytical data.

These three types of records,

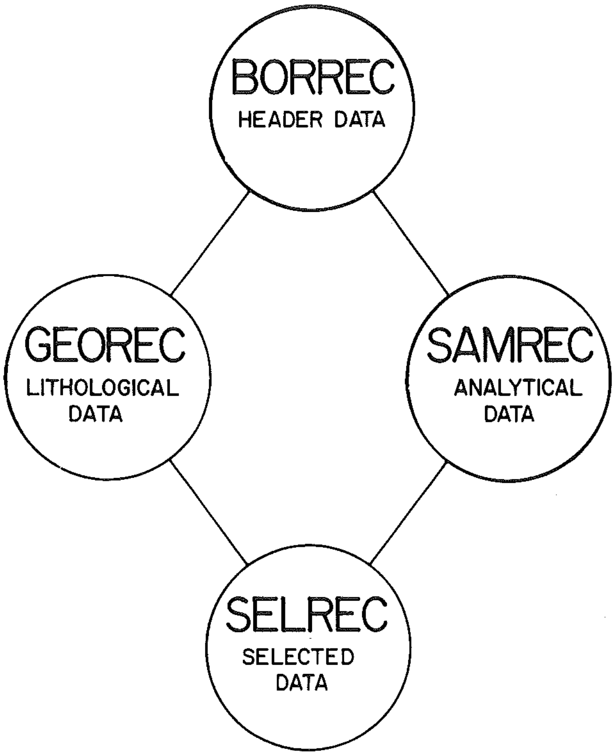

FIGURE 1. Hierarchical structure and record definition of the Data Base

BORREC, GEOREC and SAMREC, comprise the area in the Data Base where raw or unprocessed information is stored; it is also called the Primary Data Base.

There is a fourth type of record contained in the SELREC level; these records are the result of processing involving GEOREC and SAMREC records and the Vertical Selection program and contain, for each coal seam present in the borehole, a 'mineable thickness' and a set of average quality parameters calculated from the various samples composing the selected seam horizon.

Data processing

Apart from the Data Base which is the main source of information, the system also has a number of application programs. These are used to process the borehole, geological and topographical data and to produce the output required by the users.

The most important application of the Data Base is the calculation of coal resources and reserves on a regional or local scale. This is done to assist in the formulation of a coal policy and planning infrastructure development.

The following stages are necessary to produce a reserve report.

Data extraction

The Data Base resides on a Univac 1100-80 mainframe computer linked to the capturing station by direct and dial-up lines and modems. Apart from this, a stand-alone microcomputer workstation, TI Business Pro BSP-2, is available.

To extract information, a set of instructions is entered using the Data Base query language. Once the instructions are executed, a borehole data file is created. This file can be stored on the mainframe, and processing then takes place there using the mainframe version of the system. Alternatively the file can be transmitted to the microcomputer where the processing is done using the micro version of the system. In practice some

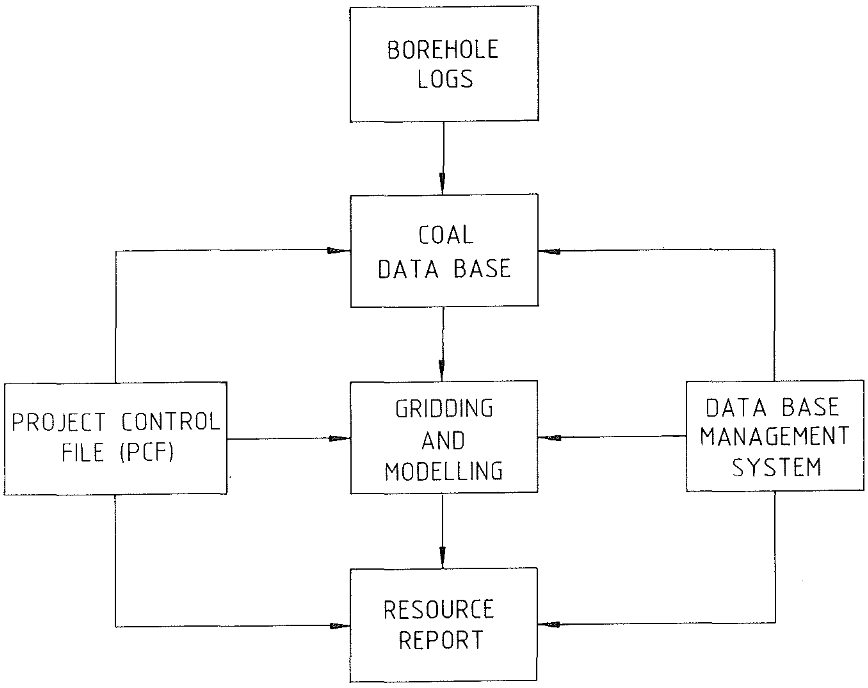

FIGURE 2. Data base processing

programs are executed on the mainframe, especially those that involve large sets of data and heavy number-crunching, and the rest are done on the micro.

Another file used for borehole extraction on the mainframe is the so-called Project Control File, or PCF, which contains tables used for data validation and also serves to constrain the extracted information and eliminate all values not acceptable in a resource/reserve report.

In the micro the PCF only contains the names of all the projects and their maximum and minimum X and Y coordinates. A second file called LOOKUP contains remainder of the information of the mainframe PCF.

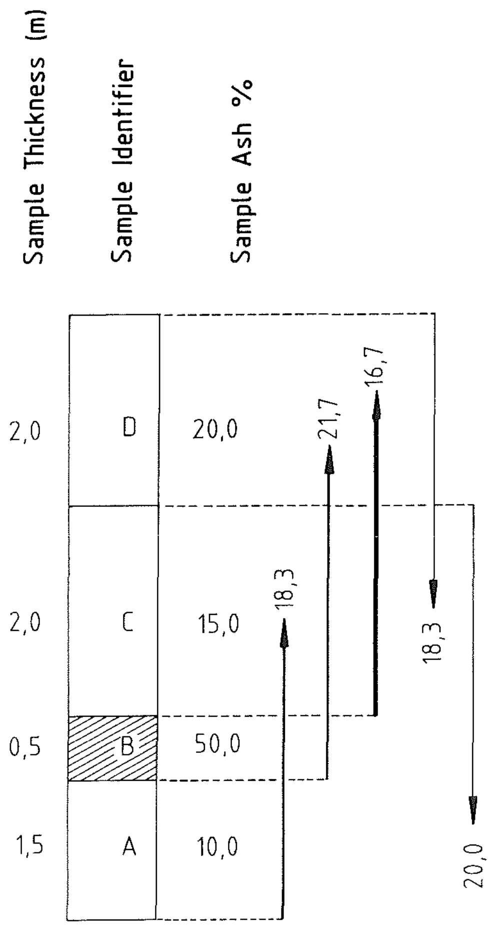

FIGURE 3. Vertical selection

Vertical selection (Figure 3)

Coal seams are usually composed of

units or layers that may vary in quality extensively. These units are sampled and analysed separately, and in order to find the optimal mineable horizon according to user-defined contraints and to determine the set of quality parameters associated to such a horizon, the Vertical Selection program is used.

This program calculates and optimizes the values for all the parameters needed to decide which portion of the seam is considered to be mineable. Its output consists of a coal thickness and a weighted average of all the quality parameters available.

The vertical selection is executed for each seam intersection individually, and no attempt is made to correlate selected horizons from one seam in one borehole to another. The optimisation is done only vertically, not horizontally.

To select a coal horizon, many factors are used in conjunction with the quality values. One of these is the presence of rock partings within the seam. These layers are recognised by a high ash content of 50% or more, or, if the sample is not analysed, by a default ash value.

These defaults are provided because when a coal seam is sampled and one or more bands of inorganic material are found in it, these bands are usually not analysed. During the process of vertical selection for each band or unit within the coal seam a value is needed for its use to average the quality of the seam or horizon. These default values are entered in the Parameter File and linked to a lithological code in the LOOKUP File. For example, if a unit labelled with the code SST=sandstone is found not analysed, a default value of ash=100% will be read and used for the calculation.

To perform the Vertical Selection, the parameters controlling the process are defined for the specific coal quality stored in the Parameter Files. If necessary, a Parameter File can be created for each coal resource/reserve project, or for each coal quality. These files contain parameters for up to 20 seams. All the major seams, that is seams for which resource estimates are needed, must have a complete set of parameters, while minor seams or seamlets for which calculations are not required are included for correlation purposes only. The seams are entered in a chronological order or in order to deposition, with the first entered being the oldest. Constraints for each of the parameters are entered and these will finally determine the level of refinement the coal-resource output. For instance, if the Parameter File is named TS = total seam, there will be no constraint for thickness, or ash or volatiles; if the file is HG = high grade steam, the constraint for ash is $\leqq$ 35%, if the file is LC = lean coal the constraint for volatiles is $\leqq$ 12%.

Grid generation

A square grid is established over the project area defined by the data extracted. The size of the cells is a function of the data density and can be calculated roughly by the formula:

$$d = \sqrt{(Xmax-Xmin)(Ymax-Ymin)^{N}}$$

where N = number data points. The grid interval, G is such that:

$$\frac{d}{4} \leqq G \leqq \frac{d}{2}$$

The gridding package only requires that in the area of interest, X and Y dimensions should be exactly divisible by the grid-cell size. After gridding the area, a value is interpolated from all the data points around each cell and placed at the node or centre-point of the cell.

The gridding algorithms available are: piecewise linear and quadratic least squares with and without corrector, linear and quadratic projection with and without surface smoothing, moving average, discrete or step, distance and density.

Grids are only as reliable as the values that generated them and often anomalously high or low values are found. These values are not easily detected from the grid itself, and to find and correct them, the grid editor is used which creates a quick contour plot where highs and lows produce a suspicious bunching of contour lines. Centring the cursor on those areas the grid node values can be displayed and if a data error is detected, a correction of the grid can be performed (without having to revise the data file). A new contour can be generated, tested again for anomalies and if proven correct, stored.

Another way of fixing grids with the grid editor is to use the cursor and make the contour lines follow a certain path; this operation automatically will change the values of the grid nodes of the cells which lie in the path of those contours and when new contours are generated using the repaired grids all angularities and lack of parallelism between contours will be smoothed out.

Grid manipulation

The calculated grids are not only

used to produce isolines which show the behaviour of the different coal parameters, but also calculate a matrix of values for each of the coal qualities needed for the reserve reporting. Using the grid manipulation routine, the grid cells are limited to only acceptable values within the given constraints.

The resulting grids are then employed for seam modelling. There is a routine that merges all the coal-quality grids and builds the seam model. During the seam modelling the panel demarcation is executed. This program divides the area of interest into individual coal blocks separated from each other by geological features such as dykes or channelling, or by areas where the coal seam is thinner or of poor quality. While executing this program, a minimum block size is given, and in the reserve report only blocks of this size or bigger will be indicated. The output of this routine is a graphic display that shows the position and shapes of all reserve blocks within the area. The blocks are also colour-coded and numbers are assigned to them.

Reserve reporting

This program needs two files, the model file or file with the merged grids and the demarcation file or file that gives the limit of each of the coal blocks of which reserves will be calculated.

To create a reserve report, the coordinates of the area of interest are again requested as well as the names of the coal parameters to be read, the upper and lower constraints for the parameters constrained and the minimum number of cells needed to define a coal block. The Reserve Report program outputs firstly a reserve print-out map showing the area of interest with all its grid cells. A symbol qualifies the cells 0 = accepted, X = rejected and . = outside block. Secondly, a print-out is made of the reserve tabulations with a histogram for each parameter and finally a résumé indicating the total reserves of the area by block.

In summmary, the Data Base processing entails working in five important fields (Figure 2). The Data Base, where all raw information is stored; the Data Base Management System, that provides the facility to access the Data Base and prepare the information for further processing; the Project Control File that serves to constrain the input/output to and from the Data Base; the Data Modelling and Gridding routines that calculate and manipulate the final grid, and the Reserve Reporting routine that produces the reserve report.

Data base applications

The information stored in the Data Base does not only describe the coal seams and qualities but also contains a large amount of qualitative and quantitative sedimentological data. This information has for example been used to draw contour maps of different rock horizons or groups of horizons, to study the relationship between the sulphur content and the deposition environment of coal seams within the basin.

One of the most striking uses of the Data Base has been study of the across-basin correlation of coal seams[6] by using coal chemical properties such as ash, volatiles and moisture content and physical properties like elevation and thickness. It was found that no single property or combinations of them provided sufficient resolution to characterise individual seams.

Discriminant analysis was then used to form linear combinations or discriminant functions of the properties which statistically provide the best possible resolution between groups. The study has proved to a high degree of certainty that by using this method coal seams can be correlated horizontally between boreholes and even adjacent coal fields. Geostatistical methods were used to study the lateral variation of coal properties when it was found that the average range of influence is approximately 15 km. The anomalous values resulting from data entry errors and the devolatilisation or oxidation of coal were identified by means of a normal probability graph and the Seyler diagram for South African coals.

Conclusions

After seven years of compiling the National Coal Data Base, the information contained in it is sufficient to produce reliable coal resource estimates for the major coal fields of the Republic, but apart from this usage, the possible applications for the Data Base seem to be endless and are only beginning to be formulated.

The degree of sophistication achieved within the Data Base and all its application programs is such that it is hoped that many other users of similar systems will benifit from them. The Data Base was mainly designed for coal and other tabular-shaped deposits, but with very few alterations it could easily be applied to other ore bodies of different shape.

The system's main feature is probably its capability to process in one operation an unlimited number of data points and to produce results within a short period of time using a mainframe computer. Its other advantage is that it can be easily scaled-down and used on a microcomputer for smaller jobs with comparable speed and the same degree of accuracy.

References

1. PRéVOST, X.M. Numerical Surface Techniques applied to the regional coal studies. Rep. Geol. Surv. S. Afr. 1978-0088 (Unpubl.), 1978.

2. DU PLESSIS, M.D., FOX, T.J. and PRéVOST, X.M. Project 'RCRRS Trial Run 2629AC': An evaluation of CORES, a computerised system for regional coal reserve/resources calculations and the creation and maintenance of a coal geology data base. Rep. Geol. Surv. S. Afr. 1979-0208 (Unpubl.), 1979.

3. PRéVOST, X.M. and DU PLESSIS, L. Evaluation of the coal potential for an area in the Kinross-Secunda sector of the Highveld/Witbank coalfield. Rep. Geol. Surv. S. Afr. 1980-0009 (Unpubl.), 1979.

4. PRéVOST, X.M. An approach to a computerised coal reserves and resources evaluation as applied at the Geological Survey. Rep. Geol. Surv. S. Afr. 1980-0029 (Unpubl.), 1980.

5. PRéVOST, X.M. The South African National Coal Data Base: A computer-based system for resource estimation and geological research. GEOCONGRESS '86. Biennial Congress of the Geological Society of S.A. Extended Abstracts, 1986. pp. 687-690.

6. LE ROUX, J.P. Die eienskappe en korrelasie van steenkoollae in die Witbank-Bethal-Breyten-gebied. M.Sc. thesis, Univ. Pta (Unpubl.), 1985.