A National Ground - Water Data Base for South Africa

Development of the Data Base

V E Cogho ,J Kirchner and J W Morris

WRC Report No.150/1/89

A NATIONAL GROUND-WATER DATA BASE FOR SOUTH AFRICA

- DEVELOPMENT OF THE DATA BASE

by

V E COGHO
J KIRCHNER
J W MORRIS

Institute for Ground-water Studies

University of the Orange Free State
Bloemfontein, South Africa

WRC Report No. 150/1/89

Obtainable from:

Water Research Commission
P O Box 824
PRETORIA
0001

ISBN 0 947447 24 5
SET NO. 0947447 26 1

EXECUTIVE SUMMARY

INTRODUCTION

For the identification of ground-water resources, the development of new water-supply schemes and the siting of boreholes, it is necessary to utilize all relevant ground-water data, if manpower and funds are to be used in the most efficient way. The availability of this data also facilitates research and reduces its cost. The generation and collection of these data are very expensive, as it requires the drilling of boreholes and field observations. At present, a great deal of these data are not formally recorded. Therefore, the creation of a ground-water data base, where all these data can be stored, was stated as one of the priority tasks in the Master Plan for Research and Development Requirements in Geohydrology (Water Research Commission, 1981).

After preliminary evaluation of existing ground-water data bases, e.g. the United States Geological Survey's WATSTORE (USGS, 1980), the Australian data base and others (Unesco, 1983; Bundesanstalt, undated), had been made by the Institute for Ground-water Studies (IGS), which were followed by discussions with the Directorate of Geohydrology of the Department of Water Affairs (DWA), the Water Research Commission (WRC) awarded a contract in 1984 to the Institute for Ground-water Studies for the development of a ground-water data base on the mainframe computer of the Department of Agriculture in Pretoria.

OBJECTIVES

According to the agreement, concluded between the WRC and the IGS, the general aim of the new data base was to improve both:

i) the results of geohydrological field work, and

ii) the planning and management of South Africa's ground-water resources through better geohydrological information.

To this end, the following objectives were spelled out in the agreement between the WRC and the IGS:

- **The ground-water data base has to be implemented on the existing mainframe computer of the Department of Agriculture and it has to be linked to other data bases, e.g. the hydrochemical data base.**

This comprised the following tasks:

SYSTEM ANALYSIS: The structure of the data base shall be decided on in consultation with the Department of Water Affairs after a system analysis comprising of:

EVALUATION OF EXISTING DATA to determine requirements for the new data base;

REPORTING ON THE SUITABILITY of existing data and the kind and format of data to be entered in future;

PROPOSING A STRUCTURE for the data base and which software should be developed; and

OUTLINING THE HARDWARE REQUIREMENTS.

IMPLEMENTATION: The implementation phase comprised of three tasks. The time allocated for each task was as follows:

- Mainframe programming (one year).
- Processing software development (two years).
- Data entry (concurrently with the software development).

- **All existing geohydrological data in the Orange Free State, as far as possible, have to be entered into the data base.**

  The Orange Free State data, available at the IGS and the Directorate of Geohydrology, shall be entered by IGS; additional data shall be obtained from municipalities and other bodies; maps and LANDSAT photos shall be interpreted for tentative delineation of aquifers and additional field and reconnaissance investigations shall be carried out, if and where necessary.

- **Software for entering, processing and data out-put has to be developed.**

  Generated with the aid of the software processing programs, a geohydrological report shall be compiled to demonstrate the facilities of the processing package.

- **The Department of Water Affairs' personnel has to be trained in the operation of the data base.**

THE DATA BASE

1. SYSTEM ANALYSIS

1.1 Structure

After the evaluation of the existing data and after the requirements for the new data base had been determined, it was decided that this data base should be point-orientated and should contain unprocessed data. It appeared desirable to include as many parameters as possible to comply with the needs of a maximum number of users. It was necessary, on the other hand, that valuable historical data, which are generally limited in accuracy compared to what can be put in the data base, should also be accommodated and, finally, disk-space requirements and processing times had to be considered.

With this in mind, the parameters to be included in the data base were decided upon. In order to reduce the disk space necessary for the storage, the data were grouped in 32 data sets. These data sets comprise the items numbered between 10 and 119 in figure 1.

The basic information about a site is contained in the data set ENTRY. Of the 28 fields provided there (see Table 1), at least the allocated identification code, the code of the map on which the site is situated, latitude, longitude, the accuracy of the position, the drainage region, the site type and the reporting institution must be given.

The remaining data sets consist of headers like LOGS, CONSTRUCTION, etc. and repeating groups, e.g. 'Penetration rate', 'Aquifer' and 'Geology', which resort under their respective headers and which can be added, when required.

Only one ENTRY data set is allowed for each site. Headers and repeating groups, however, may be repeated. Headers contain general information which applies to all repeating groups which are to be stored under them. The validity of other new data in the various data sets, e.g. dates and water levels, is checked against information already stored in the data base. Certain constructional information in the data set ENTRY is updated, as more recent information is added in the repeating groups under the header CONSTRUCTION. Because most of the data are stored in repeating groups, it is possible to build up historical records for the different sites.

All parameters that can be accommodated in the data base are shown in table 1.

Apart from the basic information in ENTRY, *geohydrological data gained during drilling* are stored under the header LOGS, i.e. the penetration rate recorded, the thickness and type of the geological strata encountered and the aquifers struck.

Constructional information is grouped under the header CONSTRUCTION. Data regarding the contractor are placed under the header, while details about depths and diameters are stored under the repeating group 'Hole'. Similarly, the information regarding the casing can be entered under 'Casing', information regarding size and distribution of perforations under 'Openings' and information on filter pack or seals placed around the casing under 'Fill'.

Data regarding Installation are stored under the header INSTALLATION.

For various *operational data* of a site, provision is made under five headers and repeating groups: 'DISCHARGE RATE' and 'WATER LEVEL' for the recording of time of measurement and measured discharge rate or water level respectively and 'FIELD MEASUREMENT' and 'METER READING' for storage of times and a variety of different parameters or different meters respectively. To avoid duplication, the details of the chemical analyses are stored in the chemical data base of the Department of Water Affairs. Under 'WATER SAMPLE', only the date and time of sampling, as well as the sample number are recorded. Correlation between the two data bases is by means of the sample number, the identification code of the site and the coordinates.

'PUMPING TEST'. Deviating from the principle of accommodating only unprocessed data, provision is made to store calculated values for *aquifer parameters* and recommended abstraction rates. The original data, which were used to determine these parameters, are stored under the operational repeating groups DISCHARGE and WATER LEVEL.

Other data which do not fall under any of the previous headings are grouped under 'MISCELLANEOUS'. They comprise information on *site selection, ownership* and *numbers other than the site id code*. They refer to the kind and whereabouts of *other geohydrological data* which are not site specific, *site visits, special cases* like information recorded from tunnels, shafts, etc., *references* to more general information about the area of the site, the *search status*, i.e. information on the sources of the data entered into the data base and lengthy *comments* which cannot be recorded in the comment of the relevant data set.

THE DATA BASE STRUCTURE

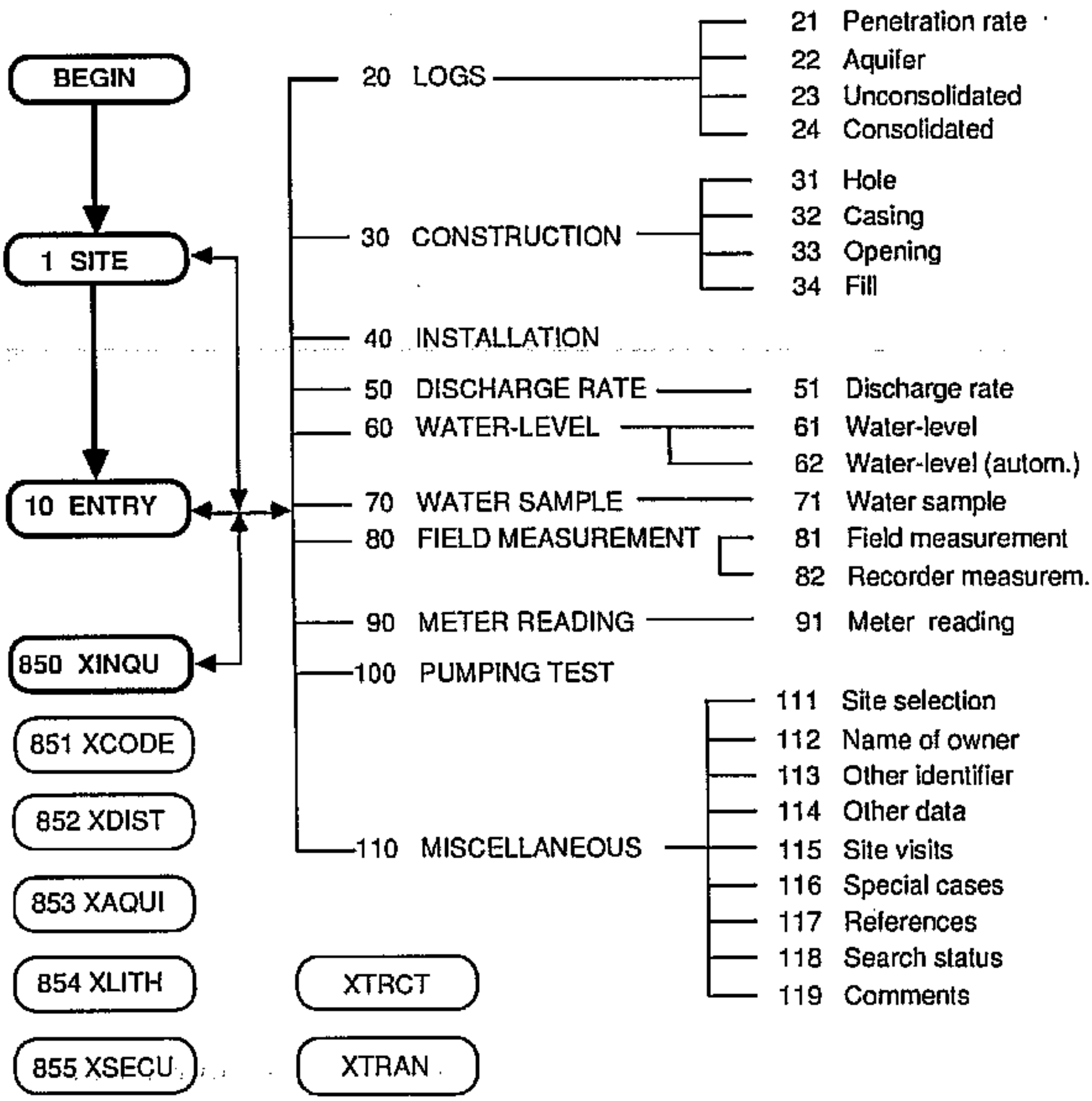


Figure 1: The structure of the Ground-Water Data Base

THE DATA BASE PARAMETERS

1.0 ENTRY
1.0. 1 Site identification code
1.0. 2 Map code
1.0. 3 Number on 1 : 50 000 map
1.0. 4 Province, district & farm no.
1.0. 5 Site name
1.0. 6 Topographic setting
1.0. 7 Latitude
1.0. 8 Longitude
1.0. 9 Coordinate accuracy
1.0.10 Drainage region
1.0.11 Altitude
1.0.12 Method of measurement
1.0.13 Site type
1.0.14 Site selector
1.0.15 Nominal diameter
1.0.16 Depth completed
1.0.17 Completion date
1.0.18 Depth information source
1.0.19 Status of site
1.0.20 Purpose of site
1.0.21 Water use: consumer
1.0.22 Water use: application
1.0.23 Reporting institution
1.0.24 Equipment
1.0.25 Potability
1.0.26 Date record entered
1.0.27 Date record updated
1.0.28 Comment

2.0 LOGS
2.0. 1 Site identification code
2.0. 2 Header sequence number
2.0. 3 Reporting institution
2.0. 4 Date record entered

2.1 Rep.group: Penetration rate
2.1. 1 Site identification code
2.1. 2 Header sequence number
2.1. 3 Repeating group number
2.1. 4 Depth to top of interval
2.1. 5 Depth to bottom of interval
2.1. 6 Diameter of interval
2.1. 7 Penetration rate
1.1. 8 Comment

2.2 Repeating group: Aquifer
2.2. 1 Site identification code
2.2. 2 Header sequence number
2.2. 3 Repeating group number
2.2. 4 Depth to top of aquifer
2.2. 5 Depth to bottom of aquifer
2.2. 6 Aquifer code
2.2. 7 Comment

2.3. Rep.group: Unconsolidated
2.3. 1 Site identification code
2.3. 2 Header sequence number
2.3. 3 Repeating group number
2.3. 4 Depth to top of interval
2.3. 5 Lithology code
2.3. 6 Primary colour
2.3. 7 Secondary colour
2.3. 8 Texture
2.3. 9 Primary feature
2.3.10 Secondary feature
2.3.11 Feature attribute
2.3.12 Sorting
2.3.13 Roundness
2.3.14 - 20 Sieve analysis
2.3.21 Comment

2.4 Repeating group: Consolidated
2.4. 1 Site identification code
2.4. 2 Header sequence number
2.4. 3 Repeating group number
2.4. 4 Depth to top of interval
2.4. 5 Lithology code
2.4. 6 Primary colour
2.4. 7 Secondary colour
2.4. 8 Texture
2.4. 9 Feature
2.3.10 Comment

3.0 CONSTRUCTION
3.0. 1 Site identification code
3.0. 2 Header sequence number
3.0. 3 Piezometer number
3.0. 4 Reporting institution
3.0. 5 Date record entered
3.0. 6 Date of construction
3.0. 7 Name of contractor
3.0. 8 Construction data source
3.0. 9 Method of construction
3.0.10 Type of finish
3.0.11 Method of development
3.0.12 Duration of development
3.0.13 Spec. development treatment
3.0.14 Cost of construction
3.0.15 Comment

3.1 Repeating group: Hole
3.1. 1 Site identification code
3.1. 2 Header sequence number
3.1. 3 Repeating group number
3.1. 4 Depth to top of interval
3.1. 5 Depth to bottom of interval
3.1. 6 Diameter of hole

2.3.2 Repeating group: Casing
3.2. 1 Site identification code
3.2. 2 Header sequence number
3.2. 3 Repeating group number
3.2. 4 Depth to top of casing interval
3.2. 5 Depth to bottom of interval
3.2. 6 Diameter of casing interval
3.2. 7 Casing material
3.2. 8 Wall thickness of casing
3.2. 9 Comment

3.3 Repeating group: Openings
3.3. 1 Site identification code
3.3. 2 Header sequence number
3.3. 3 Repeating group number
3.3. 4 Depth to top of open interval
3.3. 6 Diameter of open interval
3.3. 7 Material in this interval
3.3. 8 Type of openings in interval
3.3. 9 Length of openings
3.3.10 Width of openings
3.3.11 Hor. distance betw. openings
3.3.12 Vertical dist. betw. openings
3.3.13 Method of openings made

3.4 Repeating group: Fill
3.4. 1 Site identification code
3.4. 2 Header sequence number
3.4. 3 Repeating group number
3.4. 4 Depth to top of interval
3.4. 5 Depth to bottom of interval
3.4. 6 Type of fill
3.4. 7 Comment

4.0 INSTALLATION
4.0. 1 Site identification code
4.0. 2 Header sequence number
4.0. 3 Reporting institution
4.0. 4 Date record entered
4.0. 5 Date of installation
4.0. 6 Type of installation
4.0. 7 Depth to pump intake
4.0. 8 Type of power
4.0. 9 Power rating
4.0.10 Manufacturer
4.0.11 Serial number
4.0.12 Power meter number
4.0.13 Monitoring facility
4.0.14 Installation data source
4.0.15 Comment

5.0 DISCHARGE RATE
5.0. 1 Site identification code
5.0. 2 Header sequence number
5.0. 3 Reporting institution
5.0. 4 Date record entered
5.0. 5 Type of discharge
5.0. 6 Method discharge measured
5.0. 7 Discharge information source

5.1 Rep.group: Discharge rate
5.1. 1 Site identification code
5.1. 2 Header sequence number
5.1. 3 Repeating group number
5.1. 4 Date of measurement
5.1. 5 Time of measurement
5.1. 6 Discharge rate
5.1. 7 Comment

6.0 WATER-LEVEL
6.0. 1 Site identification code
6.0. 2 Header sequence number
6.0. 3 Piezometer number
6.0. 4 Reporting institution
6.0. 5 Date record entered
6.0. 6 Method of measurement
6.0. 7 Water-level status
6.0. 8 Collar elevation
6.0. 9 Water-level info. source

6.1 Repeating group: Water level
6.1. 1 Site identification code
6.1. 2 Header sequence number
6.1. 3 Repeating group number
6.1. 4 Date of measurement
6.1. 5 Time of measurement
6.1. 6 Water-level
6.1. 7 Comment

7.0 WATER SAMPLE
7.0. 1 Site identification code
7.0. 2 Header sequence number
7.0. 3 Reporting institution
7.0. 4 Date record entered

7.1 Repeating group: Water sample
7.1. 1 Site identification code
7.1. 2 Header sequence number
7.1. 3 Repeating group number
7.1. 4 Date of sampling
7.1. 5 Time of sampling
7.1. 6 Sample number
7.1. 7 Comment

8.0 FIELD MEASUREMENT
8.0. 1 Site identification code
8.0. 2 Header sequence number
8.0. 3 Repeating group number
8.0. 4 Date record entered

8.1 Rep.gr.: Field measurement.
8.1. 1 Site identification code
8.1. 2 Header sequence number
8.1. 3 Repeating group number
8.1. 4 Date of measurement
8.1. 5 Time of measurement
8.1. 6 Parameter measured
8.1. 7 Reading
8.1. 8 Comment

9.0 METER READING
9.0. 1 Site identification code
9.0. 2 Header sequence number
9.0. 3 Reporting institution
9.0. 4 Date record entered
9.0. 5 Type of meter read
9.0. 6 Comment

9.1 Repeating group: Meter reading
9.1. 1 Site identification code
9.1. 2 Header sequence number
9.1. 3 Repeating group number
9.1. 4 Date of reading
9.1. 5 Time of reading
9.1. 6 Reading
9.1. 7 Comment

10.0 PUMPING TEST
10.0. 1 Site identification code
10.0. 2 Header sequence number
10.0. 3 Reporting institution
10.0. 4 Date record entered
10.0. 5 Date pumping test started
10.0. 6 Method tested
10.0. 7 Depth to pump intake
10.0. 8 Recomm. abstraction rate
10.0. 9 Transmissivity
10.0.10 Storativity
10.0.11 Comment

11.0 MISCELLANEOUS
11.0. 1 Site identification code
11.0. 2 Header sequence number
11.0. 3 Reporting institution
11.0. 4 Date record entered

11.1 Rep. group.: Site selection
11.1. 1 Site identification code
11.1. 2 Header sequence number
11.1. 3 Repeating group number
11.1. 4 Description & recommendation

11.2 Repeating group: Owner
11.2. 1 Site identification code
11.2. 2 Header sequence number
11.2. 3 Repeating group number
11.2. 4 Ownership date
11.2. 5 Name of owner

11.3 Rep.group.: Other identifier
11.3. 1 Site identification code
11.3. 2 Header sequence number
11.3. 3 Repeating group number
11.3. 4 Other identifier
11.3. 5 Assignor

11.4 Repeating group: Other data
11.4. 1 Site identification code
11.4. 2 Header sequence number
11.4. 3 Repeating group number
11.4. 4 Data type
11.4. 5 Type of logger
11.4. 6 Location
11.4. 7 Format

11.5 Repeating group: Site visits
11.5. 1 Site identification code
11.5. 2 Header sequence number
11.5. 3 Repeating group number
11.5. 4 Date of visit
11.5. 5 Visitor

11.6 Rep. group: Special cases
11.6. 1 Site identification code
11.6. 2 Header sequence number
11.6. 3 Repeating group number
11.6. 4 Number of holes in group
11.6. 5 Depth of the deepest hole
11.6. 6 Depth of shallowest hole
11.6. 7 Method holes constructed
11.6. 8 Diameter of group
11.6. 9 Length of pond/tunnel/drain
11.6.10 Strike of pond/tunnel/drain
11.6.11 Dip of tunnel or drain
11.6.12 Depth of lateral collector
11.6.13 Length of lateral
11.6.14 Mesh of screen in lateral

11.7 Repeating group: References
11.7. 1 Site identification code
11.7. 2 Header sequence number
11.7. 3 Repeating group number
11.7. 4 Reference (map/secin./rep.)

11.8 Rep.group: Search status
11.8. 1 Site identification code
11.8. 2 Header sequence number
11.8. 3 Repeating group number
11.8. 4 Rec. type & period searched

11.9 Repeating group: Comments
11.9. 1 Site identification code
11.9. 2 Header sequence number
11.9. 3 Repeating group number
11.9. 4 Comment

Table 1: The parameters of the Data Base

1.2 Hardware

The data base is implemented on the mainframe computer at the Department of Agriculture. Terminals, printers and micro computers are linked, either directly or via the Transparent Government Network and concentrators. BURROUGHS' ET 1100 and other compatible products which can emulate the ET 1100 screen as, for instance, the Olivetti M24, the Qume QVT 109 and ADM 32A terminals, are all supported. For local processing, IBM compatible (Olivetti M24) and HP 17 micro computers with printer and plotter attachments have been used so far.

2. IMPLEMENTATION

2.1 Communication with the data base

Mainframe programming has been done in such a way that data are entered, updated and extracted interactively by terminals connected to the mainframe via the Government Network. For entry, change, deletion and retrieval of data, use is made of screens on the terminal which depict the coding forms that have to be filled in. All items, shown in figure 1, are screens and most of them correspond with the data sets in which the actual values are stored. The BEGIN screen is for logging on and SITE for determination of the site identification code for which information is to be updated. The screen numbered 850 is for inquiries. The higher numbered screens contain various code lists and are, together with the unnumbered screens, for the use of the data base administrator only. Details, describing the step-by-step procedures which a user must follow to enter, retrieve and process data, are contained in a users manual (Institute for Ground-water Studies, 1988a) and the technical specifications are given in a second volume (Institute for Ground-water Studies, 1988b).

2.2 Data collection, preparation and entry

Data to be entered into the data base must satisfy certain minimum requirements and must be presented in the correct format. Under normal circumstances, data are entered on coding forms for the different data sets. Details, describing the required format of the data and guide-lines regarding the completion of the coding forms, are contained in the users manual. A list of the fields and codes of the data base is also available.

Once an authorized user has logged onto the system, a choice of the SITE may be made. Other screens may be selected by entering transaction codes, data and next screen numbers, using the numbers given in figure 1. As an example of the layout of the screens, the one for the data set INSTALLATION is shown in figure 2.

```
<T]<139]<INSTL<000217]<30AUG85]
Accepted  codes: ADD, CHG, DEL or INQ - 'HOME' cursor  to  change  code [ADD]

                               I N S T A L L A T I O N

Site  id  code           <2424AD00001]        Record  sequence  number         ^    ]
Reporting  institution          [IGS ]        Date  entered                  <      ]
Date  of  installation      ^        0]       Type  of  installation             [ ]
Depth  to  intake           ^        0]       Type  of  power                    [ ]
Power  rating               ^        0]       Manufacturer    [                     ]
Serial  number          [            ]        Power  meter  number   [              ]
Monitoring  facility             [ ]          Data  source                       [ ]
Comment            [                 ]

                                                   Next  screen  number  ^   0][@]
```

Figure 2: Example of a screen: 40 INSTALLATION

2.3 Retrieval

A variety of options exist for the retrieval of data; when the user wishes to extract information using the data base programs, the inquiry screen XINQU is called up by entering 850 as the next screen number (see figure 3).

```
<T]<140]<XINQU]<00375]<30SEP88]^8620]
          N A T I O N A L  G R O U N D - W A T E R  D A T A  B A S E
                           I N Q U I R Y  S C R E E N
Title of request  [                                                        ]
Requested by      [                                         Date <19880930]
---------------------------------------------------------------------------
Single sites from [               ] to [              ]
      OR
Grid  = Latitude:      minimum ^ 0] ^      ]  maximum  ^ 0] ^      ]
Grid  = Longitude:     minimum ^ 0] ^      ]  maximum  ^ 0] ^      ]
      OR
Province [ ]  District [ ]  Farm [      ]  OR  Drainage region  [      ]
---------------------------------------------------------------------------
Site  type  [ ] Completion/Digitizing date from ^          ] to ^19880930]
---------------------------------------------------------------------------
Chose  option: mark with 'X' or number [ ] DATA           [ ] STATISTICS
[ ] ENTRY           [ ]  LOGS           [ ] CONSTRUCTION   [ ] INSTALLATION
[ ] DISCHARGE       [ ]  LEVEL          [ ] SAMPLE         [ ] FIELD MEASUREMENT
[ ] METER READING   [ ]  PUMPING TEST   [ ] MISCELLANEOUS  [ ] ALL DATA
---------------------------------------------------------------------------
Output to: [ ] Printer file <P> or Transmission file <T>       Go now [ ] <Y/N>
---------------------------------------------------------------------------
Return to BEGIN <leave blank for another inquiry, else Y [ ] [@]
```

Figure 3: The inquiry screen 850 XINQU

Here, data are selected according to identification codes or area. The area can be defined by either two coordinate pairs; by province, district or farm; or by drainage region. Optionally, the extracted data can further be sorted by specifying the site type, e.g. borehole, well or fountain and the completion date. Data can be printed as various printed lists or in statistically processed form, by choosing 'Printer file' from the menu. Printed lists contain tables, displaying all data of the data sets selected on the inquiry screen and are sorted according to either coordinates, identification codes, farm numbers or drainage regions.

If the option 'Statistics', instead of 'Data' is chosen, one of the following print-outs can be obtained:

1) the NUMBER OF ENTRIES within the specified area broken down, according to 5 accuracy classes of the position;
2) the NUMBER OF ENTRIES, the NUMBER OF THE HEADERS including 'Installation' and 'Pumping test' and the NUMBER OF EACH OF THE 22 REPEATING GROUPS within the selected area;
3) in addition to the information under point 2, the respective NUMBER OF REPEATING GROUP ENTRIES PER SITE within the selected area; and
4) SELECTED INFORMATION ABOUT EACH SITE intended amongst others for use of the technician or geologist in the field.

MACHINE-READABLE DATA FILES can be created by selecting the option 'Transmission files'.

If data are to be selected by other parameters than the options provided by the inquiry screen of the data base, special software like ERGO, INQUIRY and LIRC must be used which is provided on the mainframe.

3. DATA ENTRY

Data from a total of 9585 boreholes in the Orange Free State were traced and entered into the ground-water data base. Roughly, one-third of the data were extracted from internal reports of the Institute for Ground-water Studies and the Directorate of Geohydrology. The remaining two-thirds of the data were collected from existing borehole completion cards, site selection reports, etc., which, in numerous cases, rarely contained enough data per site, to warrant inclusion into the ground-water data base.

After consultation with the Directorate of Geohydrology, it was decided that data from Atlantis should be used to demonstrate the use of those programs for which no or insufficient data were available from the Orange Free State. Therefore, data from approximately 260 boreholes in Atlantis were also entered.

Furthermore, data from 186 boreholes west of De Aar (Karoo Aquifer Project) and from a number of boreholes in the Oudtshoorn area were entered by the IGS. The Directorate of Geohydrology has, apart from assisting the IGS with data entry from the Orange Free State and Atlantis, used the data base for data from the northern boundary areas of South Africa.

In total, data from 35 209 boreholes have been entered, since data entry began in January 1986. Of this number the IGS entered just over 10 000 boreholes; the balance has been entered by DWA.

4 THE SOFTWARE

4.1 The programs

The data base has been created with the aid of LINC II (Logic and Information Network Compiler), a fourth generation system development package. By means of special programs (or inquiry software programs like LIRC), data are processed on the mainframe for the creation of the transmission and printer files.

It has been found, however, that specialized geohydrological processing of the data is more effectively done on micro computers. Micro computers are more user-friendly and graphical output can easily be obtained. For the geohydrological processing and presentation of the data, micro computers linked to the mainframe are therefore used.

For the use of the data base, the IGS has developed software which facilitates the processing of geohydrological and geohydrochemical data. Programs for the presentation of chemical data (Piper-, Schoeller- Durov- and SAR-diagrams), graphical displays of geological and other borehole information in sections, on plans and contour maps, as well as a variety of single and multivariate statistical tests, were developed. The available programs are listed in figure 4.

4.2 Geohydrological processing of data

Data from the Orange Free State and from Atlantis have been used to demonstrate the output that can be obtained by using the processing package. The authors have also tried to show how interpretation of the results, obtained by means of the programs, can assist the professionals in the exploitation and management of the ground-water resources. Some examples of graphical presentation of processed data are shown in figure 5.

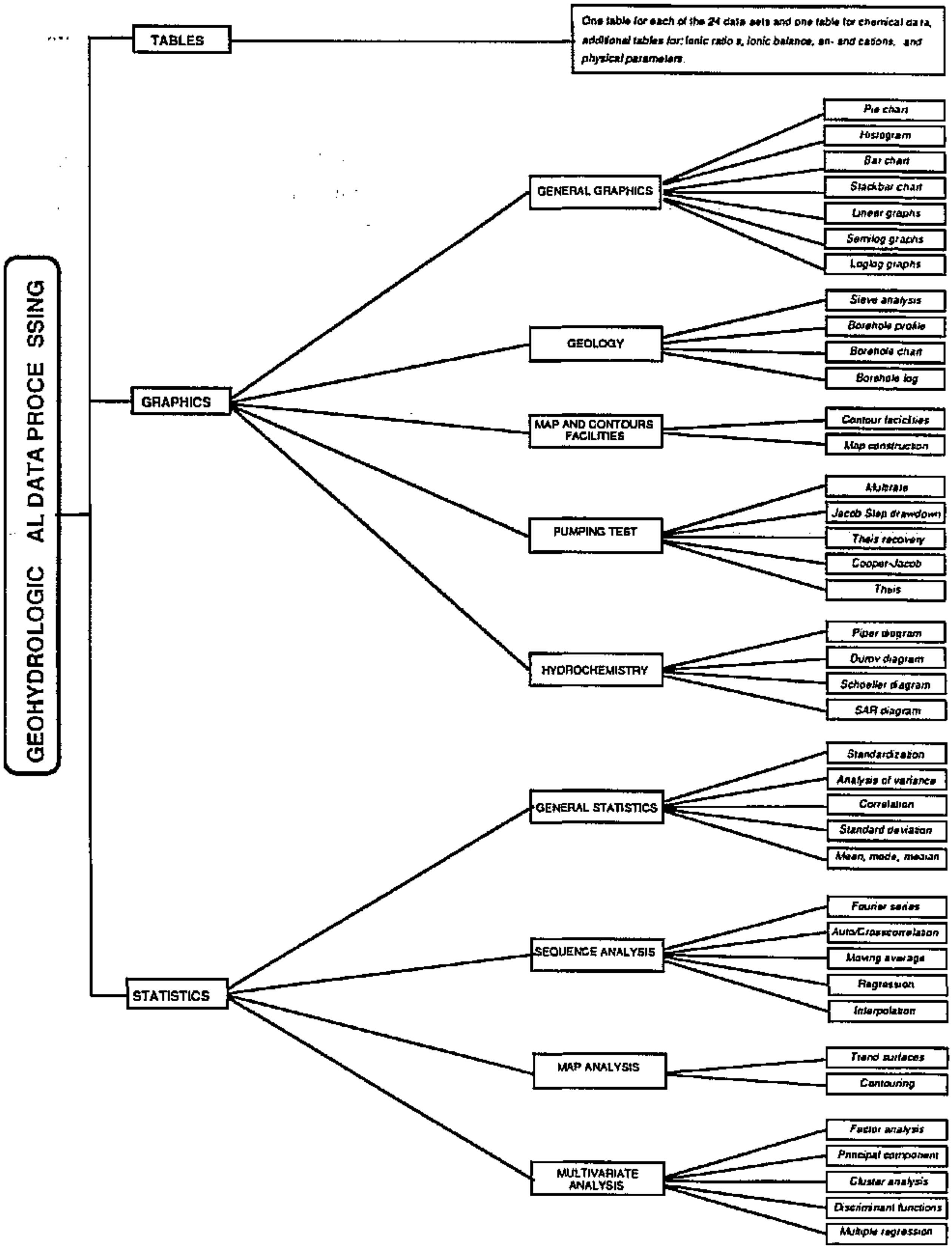
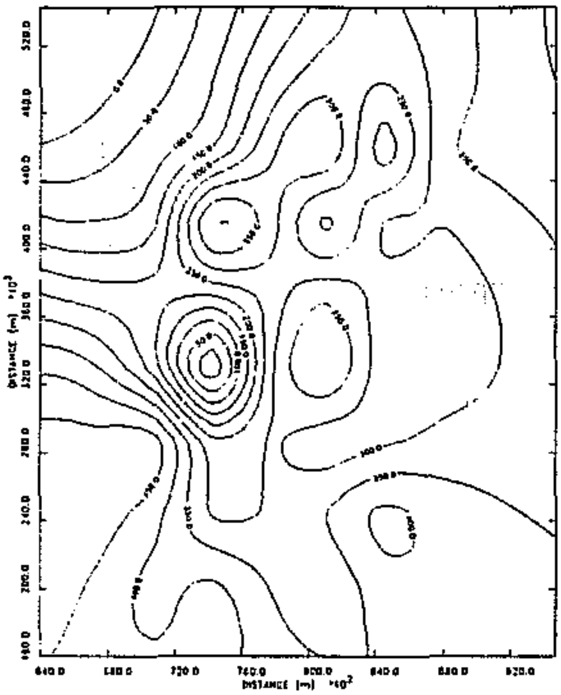
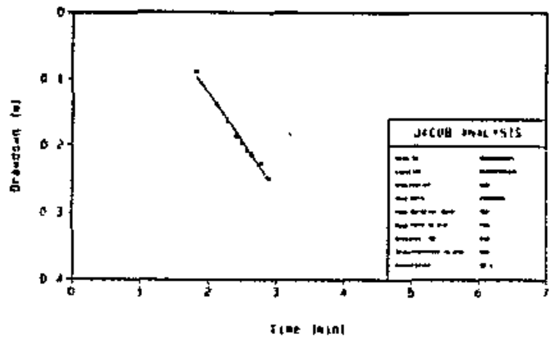


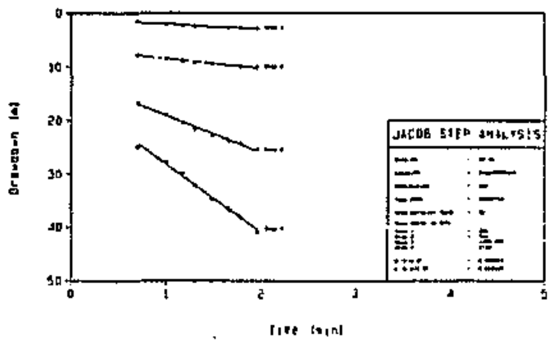
Figure 4: Geohydrological data processing programs.



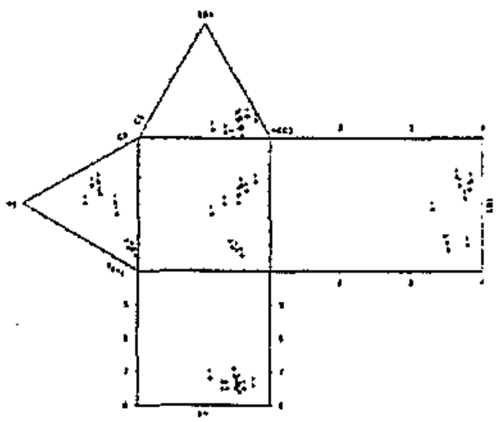
WATER-LEVEL CONTOURS



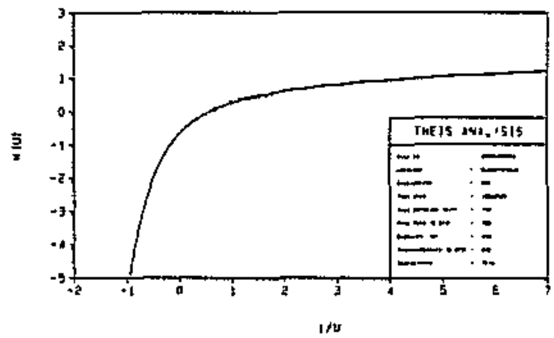
PUMPING TEST : JACOB



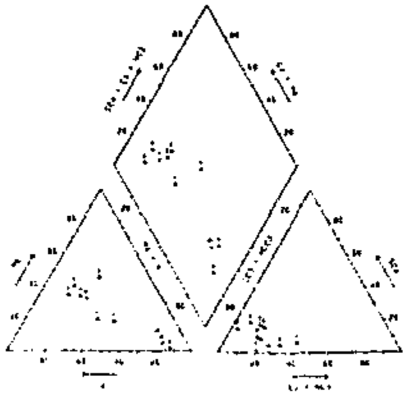
PUMPING TEST : STEP TEST



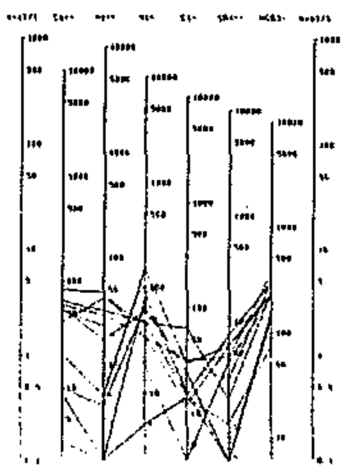
DUROV DIAGRAM



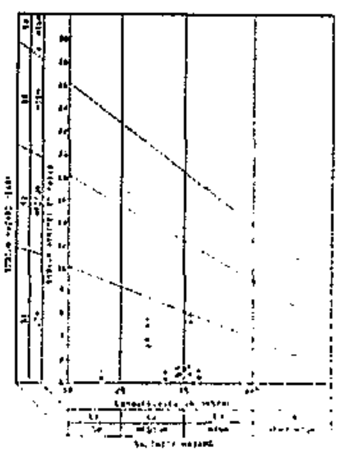
PUMPING TEST : THEIS



PIPER DIAGRAM



SCHOELLER DIAGRAM



SAR DIAGRAM

Figure 5: Examples of graphical presentation of processed data.

4. TRAINING OF WATER AFFAIRS' PERSONNEL

Since the beginning of 1986, when the data base became operative and the various facilities have become available, some of the personnel of the Directorate of Geohydrology have been trained in the use of the data base. Since the middle of 1986, the data base has been managed by the Department of Water Affairs. Training in the use of the processing programs has taken place during 1987, up to the beginning of 1988.

ON-GOING DEVELOPMENT

In the near future, it will also be possible to extract data from the data base via the newly founded Computer Centre for Water Research in Pietermaritzburg.

Future aims include batch loading facilities (under way for water-level recorders and geophysical borehole logs), the refinement and extension of the existing software and possibly the addition of further software modules, e.g. extended CAD (computer-aided design) facilities for preparation of maps.

With the support of the Water Research Commission, the Institute for Ground-water Studies is presently developing G-Base, a PC-operated data base compatible with the National Ground-water Data Base. G-Base facilitates further data processing and also allows data capturing on a much wider basis.

CONCLUSION

The Ground-water Data Base is a national facility. It has now been completed and is available for the use by geohydrologists, as well as various other professionals, working in the field of ground-water development, -supply and -research, for the government as well as the private sector.

All major objectives have been met (cf. page 3 of the agreement):

(i) The data base is implemented on the mainframe computer and linked to the chemical data base.

(ii) Existing ground-water data from the Orange Free State have been entered, as far as possible.

(iii) Software for the processing of data has been developed.

(iv) Personnel of the Directorate of Geohydrology has been trained in the use of the data base.

Additional objectives achieved were (cf. page 3 of the agreement):

(a) Orange Free State data *and additionally some Atlantis data* have been evaluated.

(b)*) Additional data have only been obtained from Dewetsdorp.

(c)*) LANDSAT photos and geological maps have only been used in the Dewetsdorp area. Aquifers have not been tentatively delineated.

(d)*) Field control of data occurred only at Dewetsdorp.

(e) Orange Free State *and Atlantis* data have been entered and processed.

(f)*) Additional field investigations, if and where necessary, have, with the exception of the Dewetsdorp area, not been done.

(g) A geohydrological report generated with the aid of the data base programs has been compiled.

The data base has been introduced to the community by means of a poster show at the exhibition of the National Water Borehole Association in Johannesburg in November, 1986. A brochure was also prepared for this occasion. In response to a publication, *Development of a national ground-water data base,* which was published in WATER SA, numerous requests for reprints have been received from South Africa, America, Asia and Europe.

REFERENCES

Bundesanstalt für Bodenforschung (undated).- Coding forms for Well Data Base, personal communication P.E. Groba, Hannover, West Germany.

Institute for Ground-water Studies (1988a).- National Ground-water Data Base, Volume 1: User Guide.

Institute for Ground-water Studies (1988b).- National Ground-water Data Base, Volume 2: Technical Specifications.

Kirchner J., V.E. Cogho & J.W. Morris (1987).- Development of a national ground-water data base. Water SA, *13*, p 165 -170

Unesco (1983).- Computer processing of ground-water data (Chapter 8.5) in *Ground-water studies.*

U.S. Geological Survey (1980).- National Water Data Storage and Retrieval System.

Water Research Commission (1981).- Research and Development Requirements in Geohydrology.

---

*) Not fully achieved, as it was anticipated that these tasks would be carried out at a number of places in the OFS.

ACKNOWLEDGMENTS

The research described in this report resulted from a project funded by the Water Research Commission entitled:

"The development of a National Data Bank for ground-water data."

The Steering Committee responsible for this project, consisted of the following persons:

Dr P J T Roberts	Water Research Commission (Chairman)
Mr H M du Plessis	Water Research Commission
Mr P Weideman	Water Research Commission (Secretary)
Prof H J Potgieter	University of the OFS
Mr J R Vegter	Directorate of Geohydrology, Dept of Water Affairs
Mr F G de Kock	Directorate of Computer Services, Dept of Water Affairs
Mr J Nel	Directorate of Computer Services, Dept of Water Affairs
Dr C P R Roberts	Directorate of Planning Services, Dept of Water Affairs
Mr M W Selibowitz	Directorate of Planning Services, Dept of Water Affairs
Mr R J Connelly	Steffen, Robertson & Kirsten
Prof J Kirchner	Institute for Ground-water Studies, University of the OFS
Prof F D I Hodgson	Institute for Ground-water Studies, University of the OFS

The financing of the project by the Water Research Commission and the contributions of the members of the Steering Committee are acknowledged gratefully.

CONTENTS

CHAPTER 1 1

INTRODUCTION 1

CHAPTER 2 5

DATA BASES - STATE OF THE ART 5

2.1 BASIC CONCEPTS 5

2.1.1 GENERAL DEFINITIONS.. 5

2.1.2 EXTENDED DATA BASE DEFINITIONS.. 7

2.2 DATA BASE MODELS.. 10

2.2.1 RELATIONSHIPS WITHIN DATA MODELS.. 12

2.2.2 FLAT FILE MODEL.. 15

2.2.3 RELATIONAL DATA MODEL.. 15

2.2.4 HIERARCHICAL MODEL.. 18

2.2.5 NETWORK MODEL.. 23

2.2.6 MODEL COMPARISON.. 24

2.3 EXISTING GROUND-WATER DATA BASES.. 27

2.3.1 DATA BASE BY THE BRITISH GEOLOGICAL SURVEY (U.K.).. 27

2.3.2 DATA BASE OF THE WATER RESOURCES SECTION OF THE PUBLIC WORKS DEPARTMENT OF WESTERN AUSTRALIA.. 33

2.3.3 BOREHOLE DATA BASE - S.W.A./NAMIBIA.. 38

2.3.4 DATA BASE BY THE U.S.G.S. (WATSTORE).. 39

CHAPTER 3 48

NATIONAL GROUND-WATER DATA BASE FOR SOUTHERN AFRICA 48

3.1 INTRODUCTION 48

3.2 SOFTWARE REQUIREMENTS FOR MAINFRAM DATA BASE DEVELOPMENT.. 49

3.3 STRUCTURE OF THE NATIONAL GROUND-WATER DATA BASE.. 52

3.4 DATA BASE PARAMETERS.. 58

3.5 DATA INPUT 65

3.6 DATA OUTPUT 68

3.7 HARDWARE REQUIREMENTS.. 73

3.8 COMMUNICATIONS.. 75

CHAPTER 4 78

DATA PROCESSING 78

4.1 INTRODUCTION 78

4.2 DATA TABULATION....79

4.3 GRAPHICAL REPRESENTATION....87

- 4.3.1 HYDROCHEMICAL DATA ANALYSIS....87
- 4.3.2 PUMPING TEST ANALYSIS....94
- 4.3.3 GEOLOGICAL ANALYSIS....104
- 4.3.4 MAP AND CONTOURING FACILITIES....111
- 4.3.5 GENERAL GRAPHICS....117

4.4 STATISTICS 117

- 4.4.1 GENERAL STATISTICS....117
- 4.4.2 SEQUENCE ANALYSIS....122
- 4.4.3 MAP ANALYSIS....127
- 4.4.4 MULTIVARIATE DATA ANALYSIS....128

CHAPTER 5 133

EVALUATION OF THE PROCESSING FACILITIES OF THE GROUND-WATER DATA BASE 133

5.1 INTRODUCTION 133

5.2 GEOLOGY 135

- 5.2.1 LOCAL ANALYSIS....135
- 5.2.2 REGIONAL ANALYSIS....147

5.3 AQUIFER ANALYSIS....147

- 5.3.1 LOCAL ANALYSIS....149
- 5.3.2 REGIONAL ANALYSIS....181

5.4 BOREHOLE CONSTRUCTION AND INSTALLATION....188

- 5.4.1 LOCAL ANALYSIS....188

5.5 GROUND-WATER QUALITY....199

- 5.5.1 LOCAL ANALYSIS....199
- 5.5.2 REGIONAL ANALYSIS....233

5.6 DEVELOPMENT AND MANAGEMENT OF AQUIFERS....246

5.7 DISCUSSION OF RESULTS....247

CHAPTER 6 249
DISCUSSION 249
6.1 EVALUATION OF THE NATIONAL GROUND-WATER DATA BASE 249
6.2 DATA BASE USERS 251
6.3 DATA BASE APPLICATION ASPECTS 256
6.3.1 AQUIFER MANAGEMENT 256
6.3.2 PLANNING ASPECTS 256
6.3.3 RESEARCH ASPECTS 256
6.3.4 MANPOWER ASPECTS 257
6.3.5 COST ASPECTS 257
6.3.6 SECURITY ASPECTS 258
6.4 EVALUATION OF THE DATA PROCESSING PACKAGE 259
6.4.1 TABLES 260
6.4.2 GRAPHICS 260
6.4.3 STATISTICS 265
6.5 CONCLUSIONS 267

REFERENCES 268
SUMMARY 275
APPENDIX I - USER'S GUIDE

LIST OF FIGURES

2.1 Simplified, schematic representation of a data base system (after Date, 1975) 6
2.2 Two stages of data independence (after Atre, 1980) 9
2.3 Data base design steps (after Atre, 1980) 11
2.4 Hierarchical tree structure 19
2.5 Access paths of a hierarchical data model 20
2.6 Representation of an aquifer data base 21
2.7 Representation of a site data base 22
2.8 Representation of a site data base with different depths to top and (all) the respective aquifers for a specific depth 23
2.9 Representation of four different network data modes (after Martin, 1975) 25
2.10 Simplification of complex network structures 25
2.11 Data base represented by a hierarchical and relational model (after Ullman, 1980) 26
2.12 Data base structure (after Harvey, 1973) 30
2.13 Fixed format input coding sheet (after Harvey, 1973) 31
2.14 Western Australia data base structure (after Harvey, 1978) 35
2.15(a) EDIT data base entry 36
2.15(b) EDIT data base retrieval 36
2.15(c) Rating curve data base entry 36
2.15(d) Rating curve data base retrieval (after Harvey, 1978) 36
2.16 Schematic representation of the WATSTORE information system (after Showen, 1976) 41
2.17 Data base structure (after WATSTORE, User's Guide, Volume 2) 43
3.1 Data set structure 54
3.2 (a) Basic data structure and (b) basic data structure converted into a simple tree structure 56
3.3(a) Basic tree structure of the ground-water data base 57
3.3(b) Ground-water data base structure, indicating the respective positions of each data set in the tree structure 57
3.4 Parameters of the ground-water data base 59
3.5 Administrative district subdivisions in the Transvaal 64
3.6 Data coding form example 67
3.7 Inquiry screen 69
3.8(a) Printer file output of site specific data 71
3.8(b) Printer file output of site data statistics 71
3.9 Transmission file output example 72

3.10	Elements of a Govnet Transparent Protocol Network	76
3.11	Networking via SAPONET-P	77
4.1	Diagrammatic representation of the data processing possibilities	80
4.2	Table output of the Entry data set	81
4.3	Table output of the Water-level data set	82
4.4	Table output of the Water-sample data set	83
4.5	Piper diagram	89
4.6	Durov diagram	90
4.7	Schoeller diagram	92
4.8	SAR diagram	93
4.9	Theis analysis	96
4.10	Cooper-Jacob analysis	98
4.11	Theis recovery analysis	100
4.12	Jacob Step Drawdown analysis	102
4.13	Multirate pumping test analysis	103
4.14	Graphical representation of a geological borehole log	106
4.15	Graphical representation of a borehole chart	107
4.16	Graphical representation of a borehole profile or cross-section	108
4.17	Grain-size distribution graph	110
4.18(a)	Two-dimensional areal map	112
4.18(b)	Two-dimensional contour map, produced by a plotter	114
4.18(c)	Two-dimensional contour map, produced by a printer	115
4.18(d)	Three-dimensional contour map of chlorine concentrations	116
4.19	Examples of general graphics displays	118
4.20	Diagrammatic representation of the linear interpolation technique	124
4.21(a)	Data sequence as a step function	124
4.21(b)	Equally spaced data after rectangular integration	124
5.1	Bar chart showing the distribution of borehole coordinate accuracies	134
5.2(a)	Simplified geological map of the Orange Free State (after Truswell)	136
5.2(b)	Simplified geological map of the Atlantis area (after Müller, 1984)	137
5.3	Borehole logs of the southern Orange Free State	139
5.4	Simplified borehole profile of boreholes displayed in figure 5.3	140
5.5	Borehole logs of the Hertzogville region	143
5.6	Borehole profile of the Hertzogville region	144
5.7	Borehole log of the Atlantis area	148
5.8(a)	Borehole position map of the southern Orange Free State	150
5.8(b)	Borehole position map of boreholes with recorded water levels in the southern Orange Free State	151

5.8(c)	Water-level contour map of the southern Orange Free State	152
5.8(d)	Topographical contour map of the southern Orange Free State	153
5.8(e)	Simplified geological map of the southern Orange Free State	154
5.9(a)	Histogram of yields in sedimentary aquifers in the southern Orange Free State	157
5.9(b)	Histogram of yields in dolerite associated aquifers in the southern Orange Free State	158
5.10	Frequency distribution of borehole yields in the southern Orange Free State	159
5.11(a)	Yield contour map of the southern Orange Free State	160
5.11(b)	Borehole depth contour map of the southern Orange Free State	161
5.11(c)	Depth water struck contour map of the southern Orange Free State	162
5.12	Borehole profile of the Silwerstroom area in the Atlantis region	167
5.13	Atlantis water-level contours (October, 1980)	168
5.14	Atlantis water-level contours (October, 1981)	169
5.15	Atlantis water-level contours (October, 1982)	170
5.16	Atlantis water-level contours (October, 1983)	171
5.17	Atlantis water-level contours (October, 1984)	172
5.18	Atlatnis water-level contours (July, 1986)	173
5.19	Well field delineation	175
5.20	Water-level hydrograph of borehole 3318CB00010 - starting July, 1978	176
5.21	Water-level hydrograph of borehole 3318CB00015 - starting July, 1978	177
5.22	Water-level hydrograph of borehole 3318CB00048 - starting September, 1979	178
5.23	Water-level hydrograph of borehole 3318CB00124 - starting August, 1977	179
5.24	Proposed borehole observation network for the Atlantis aquifer	180
5.25(a)	Frequency distribution of borehole yields in the Orange Free State	183
5.25(b)	Borehole yields of square degrees 2726, 2727, 2728 and 2729	184
5.25(c)	Borehole yields of square degrees 2825, 2826, 2827 and 2828	185
5.25(d)	Borehole yields of square degrees 2829, 2925, 2926 and 3025	186
5.25(e)	Borehole yields of square degrees 3026 and 3027	187
5.26	Frequency distribution of borehole depths in the region	190
5.27	Borehole chart of a typical southern Orange Free State borehole	191
5.28	Borehole chart of a typical Atlantis borehole	198
5.29	Distribution of chemical analysed boreholes in the Trompsburg area	200
5.30	Hydrochemical diagrams of water samples in the Trompsburg area	201
5.31	Piper analysis of randomly reduced water samples	203
5.32	TDS contours of the Trompsburg area in mg/l	204
5.33	Electrical conductivity contours of the Trompsburg area in mS/m	205
5.34	pH contours of the Trompsburg area	206
5.35	Alkalinity contours of the Trompsburg area in mg/l	207

5.36 Nitrate contours of the Trompsburg area in mg./l.....208

5.37 Chloride contours of the Trompsburg area in mg/l.....209

5.38 Cl/HCO3-ratio contours of the Trompsburg area.....210

5.39(a) R-mode cluster analysis of samples from the Trompsburg area.....217

5.39(b) Q-mode cluster analysis of samples from the Trompsburg area.....218

5.40 Graphical representation of the significant principal component scores.....220

5.41 Graphical representation of the significant rotated factor scores.....221

5.42 Piper diagram of water samples in the Atlantis region.....223

5.43 Durov diagram of water samples in the Atlantis region.....224

5.44 Distribution map of water sample locations in the Atlantis area.....225

5.45 Chloride contours in mg/l of water samples in the Atlantis area.....226

5.46 Sulphate contours in mg/l of water samples in the Atlantis area.....227

5.47 Total alkalinity contours in mg/l of water samples in the Atlantis area.....228

5.48 pH contours of water samples in the Atlantis area.....229

5.49 TDS contours of water samples in mg/l in the Atlantis area.....230

5.50 Conductivity contours of water samples in mS/m in the Atlantis area.....231

5.51 R-mode cluster analysis of the Atlantis water sample data.....234

5.52 Principal component analysis of the Atlantis water sample data.....235

5.53 Q-mode rotated factor matrix - Atlantis water sample data.....236

5.54 Factor score contours - Factor 1.....237

5.55 Factor score contours - Factor 2.....238

5.56 Distribution of the chemical characteristics of the southern Orange Free State and Dewetsdorp ground waters.....239

LIST OF TABLES

2.1 Two-dimensional data lay-out (flat file) 15
2.2 Example of a simple relation 16
2.3 Description of GWSI schema records (after Mercer and Morgan, 1982) 46
3.1 Order of letter elimination (after WATSTORE, User's Guide) 63
3.2 Output option differences 73
5.1 Tabulated geological detail of some of the boreholes in the southern Orange Free State 141
5.2 Water-level statistics 149
5.3(A) Tabulated statistics of raw borehole yield data between sedimentary and dolerite aquifers in the southern Orange Free State 155
5.3(B) Tabulated statistics of raw borehole yield data between sedimentary and dolerite aquifers in the southern Orange Free State 155
5.4 Borehole yield statistics of the southern Orange Free State 156
5.5 Summarized water applications 163
5.6 Aquifer types encountered in Atlantis boreholes 164
5.7 Tabulated pumping test results of the Atlantis area 166
5.8 Water-level statistics for the Orange Free State 182
5.9 Borehole yield statistics of the Orange Free State 182
5.10 Tabulated equipment distribution in the southern Orange Free State 189
5.11 Tabulated borehole depth statistics of the region 189
5.12 Tabulated borehole depth statistics of the Orange Free State 192
5.13 Tabulated Atlantis borehole depth statistics 192
5.14 Tabulated borehole diameter information of the Atlantis area 193
5.15 Penetration rate statistics 194
5.16 Tabulated Opening information of the Atlantis area 195
5.17 Tabulated Fill information of the Atlantis area 196
5.18 Tabulated Installation information of the Atlantis area 197
5.19 Statistic of the water analyses from the Trompsburg area 202
5.20 Analysed cations of the southern Orange Free State 212
5.21 Analysed anions of the southern Orange Free State 213
5.22 Analysed physical parameters of the southern Orange Free State 214
5.23 Ionic balance of the analysed samples in the southern Orange Free State 215
5.24 Tabulated ionic ratios of ground waters from the southern Orange Free State 216
5.25 Tabulated ionic ratios of the Atlantis region 232
5.26 Analysed cations of the Dewetsdorp region 240

5.27	Analysed anions of the Dewetsdorp region	241
5.28	Analysed physical parameters of the Dewetsdorp region	242
5.29	Ionic balance of the analysed samples in the Dewetsdorp region	243
5.30	Ionic ratios of the Dewetsdorp region	244

CHAPTER 1

INTRODUCTION

In Southern Africa it has been realized that refined quantitative answers are needed for optimal utilization of the available water resources. Surface and subsurface resources are developed at an ever-increasing pace to meet the needs of agricultural, industrial, domestic, commercial, recreational users, mining activities and power generation.

Competition for the available resources has brought about an awareness that the principal problem confronting the water suppliers is resource management. However, before water resources can be managed effectively, they must be quantitatively assessed. Thus, geohydrologists, hydrologists and engineers are increasingly called upon to determine how much water is available for further development and to evaluate what the consequences of such development will be.

The greater the development, the greater is the stress on available water and land resources. Likewise, the smaller the resource, the greater the stress and the greater the need for accurate and reliable knowledge to support its careful management and development (Harvey, 1978).

In Southern Africa, the existing water resources have to be managed very carefully to satisfy the future levels of demand. Accordingly, the demand for reliable hydrological related data has increased over the past decade and has led to an upgrading of water resources data collection.

To be of real value, the data must be reliably collected, stored, processed and readily available for data assessment. Due to the large number of hydrological data involved, they must be stored on computer to achieve fast access and comprehensive extraction.

The hydrological information system envisaged for South Africa will consist of eight related data bases comprising the following information, i.e. precipitation, flow, reservoirs, evaporation, sediment load, floods, chemical quality and ground water. In order to expedite the process of computerization, the Water Research Commission has funded the development of these data bases.

At the beginning of 1984, the Water Research Commission awarded a contract to the Institute for Ground-water Studies at the University of the Orange Free State, comprising the development of a National Ground-water Data Base for Southern Africa. The development of this data base was to be done in conjunction with the Directorate of Geohydrology of the Department of Water Affairs. This data base had to make provision for physical, geological and geohydrological information about any known ground-water source in Southern Africa. User-friendly software for the geohydrological processing of the data was to be created as part of the data base package.

According to the agreement concluded between the Water Research Commission and the Institute for Ground-water Studies, the aim of the new data base was to improve:

(i) the results of geohydrological field work and
(ii) the planning and management of South Africa's ground-water resources through better geohydrological information.

To this end, the following objectives were identified:

(i) To link the new ground-water data base with the other data bases in the hydrological information system and to implement it therefore on the central mainframe computer

(ii) To develop a structure for the data base in consultation with the Department of Water Affairs after an analysis comprising the following:

(a) Evaluation of existing data to determine requirements for the new data base.

(b) Reporting on the suitability of existing data and the kind and format of data to be entered in future.

(iii) To develop software for data entry, retrieval, processing and output.

(iv) To enter all Orange Free State data available in reports and records for testing of the data base and the processing programs.

(v) To outline the hardware requirements for the ground-water data base.

(vi) To train Water Affairs personnel in the use of the new data base and the application package.

In a system analysis, the potential users have to be borne in mind, in order to satisfy their requirements. Users which may benefit from the ground-water data base are:

(i) Geohydrologists and geohydrological technicians.

(ii) Engineers.
(iii) Drilling contractors and suppliers of pumping equipment.
(iv) Water supply organizations.
(v) Researchers.

The objective of this report is to discuss the scientific development of the ground-water data base, with special reference to its application in data processing. It is supplemented by a Users' Guide which contains detailed instructions for entering data into the data base, as well as extraction and data processing. This Users' Guide is available as a separate volume, WRC Report No. 150/2/89.

The existence of four basic data base structures, as well as other basic data base concepts, serves as a natural starting point for a discussion. Three of the structures are of particular importance and are discussed in chapter 2. They are:

(i) Relational data base model (section 2.2.3).
(ii) Hierarchical data base model (section 2.2.4).
(iii) Network data base model (section 2.2.5).

Furthermore, a number of existing ground-water data bases are discussed in this chapter.

Chapter 3 discusses the new ground-water data base. After an introduction, the software available on the mainframe for the creation of a data base is described in section 3.2 and the structure of the new data base in section 3.3. All the parameters accommodated in the data base are discussed in section 3.4. A data base would be of little value, if it cannot easily be accessed. Facilities created for input and output are described in sections 3.5 and 3.6. Hardware requirements and communication network are dealt with in sections 3.7 and 3.8.

An essential facet of the ground-water data base is its ability to perform data processing. Three basic facilities are included in the data base, namely extraction of data and output in the form of:

(i) Tables.
(ii) Graphical displays.
(iii) Statistical analyses.

These data processing techniques are discussed in chapter 4.

After discussion of the basic concepts, the development of the data base and its data processing facilities, the actual applications and the advantages of assessing site specific data are discussed in the following two chapters.

Chapter 5 describes the evaluation of geohydrological data from the Orange Free State and Atlantis region, already accommodated in the data base. Special reference is made to the scale of the evaluation, namely either:

(i) Regional or
(ii) Local.

Some of the results and difficulties experienced, while using the data base, are discussed in chapter 6. Furthermore, special attention is focussed on the utilization, with respect to aquifer management, planning, manpower, research and cost aspects of aquifer development.

CHAPTER 2

DATA BASES - STATE OF THE ART

2.1 BASIC CONCEPTS

In order to understand the terminology used for describing files and data bases, it is necessary to define certain basic terms. Data terminology, i.e. the description of data and of data relationships, can be divided into two groups, namely logical and physical. Physical data descriptions refer to the physical recording of data on the hardware, whereas logical data descriptions refer to the representation of data to the application programmer or user (Martin, 1975). Only the logical data description is contemplated hereunder.

2.1.1 General definitions

Enterprise:

Any large-scale commercial, scientific or technical organization. An enterprise must necessarily maintain numerous operational data.

Entity (logical record):

Information regarding any person, place or concept about which data are recorded; not necessarily the same as the stored data record (e.g. all information for one site).

Data base:

A data base is a collection of data, containing the relationships between data records, data groups and data items which are used by a particular enterprise. Figure 2.1 shows a simplified, diagrammatic view of a data base.

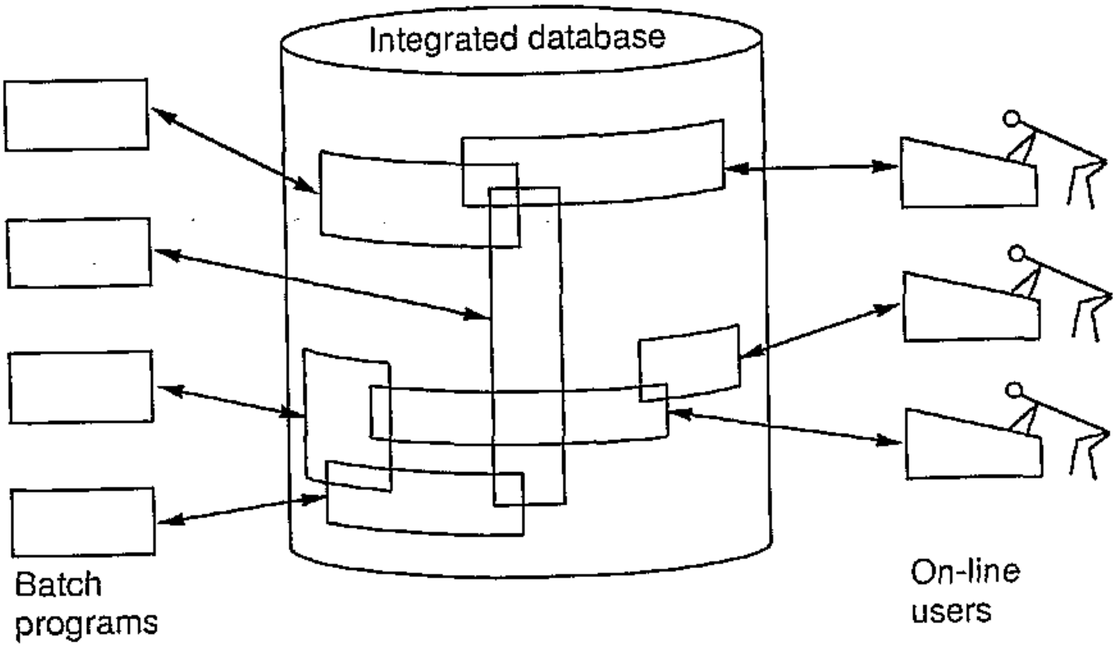


Figure 2.1.Simplified, schematic representation of a data base system (after Date, 1 1975).

Data record:

It represents a collection of related attributes, stored on some medium; in this case the storage device of a mainframe computer (e.g. all information for one site, same as Entity, if information for all data sets is available).

Data set (data group, data file):

It consists of a collection of attributes within a record and is referred to as a whole. Two types of data groups exist, i.e.:

(i) Vectors (arrays).
(ii) Repeating groups.

A vector is a one-dimensional, ordered selection of data items, whereas a repeating group is a collection of data which occurs multiple times within a specific record (e.g. the water-level record of one site). Another expression describing a data group is a data file.

Data item (attribute, data element):

As the smallest unit of named data, it forms part of various basic characteristics of a specific entity (e.g.all water levels of one site).

Domain:

The smallest unit in a relational model is an individual data item. A domain is a set of such values, all of the same type, e.g. the domain of aquifer codes.

Relation:

Is a subset of the Cartesian product of a list of domains.

Tuple:

The grouping of data generally represents a relationship between the respective data items. Such a related set of values is referred to as a tuple (see figure 2.11).

Schema:

The logical data base description is referred to as schema. It gives the names of the entities and attributes and specifies the relations between them (e.g. a complete tree diagram of a data base).

List:

A simple list structure is a sequence of data items or records linked by pointers.

Flat arrangement:

The simple two-dimensional lay-out of data items is referred to as a flat arrangement (e.g. the SWA data base).

Data base management system:

Is a software system capable of supporting and managing an integrated data base.

2.1.2 Extended data base definitions

A data base is not just a collection of related data records, however. The data must also be organized in such a way that efficient location of any data item or combination of items is possible. The persons responsible for the systematical arrangement of data in the data base are the system analyst, programmer and data base administrator.

In general, data bases provide any enterprise with a centralized control of its operational data. This has the following advantages (Date, 1975):

- (i) The amount of redundancy in stored data can be reduced.
- (ii) Inconsistency of stored data can be avoided.
- (iii) Stored data can be shared.
- (iv) Standards can be enforced.
- (v) Security restrictions can be applied.
- (vi) Data integrity can be maintained.
- (vii) Conflicting requirements can be balanced.

These advantages are fairly obvious. However, implied by several other points mentioned, data independence must be added to the list.

Data redundancy:

It is unavoidable that some data elements are used in a number of different applications. Since the data are required by multiple applications, they are often recorded in multiple data files and stored repeatedly. This leads to data redundancy.

Generally, a data base is defined as a non-redundant collection of data items. In reality, however, a certain measure of controlled data redundancy should always exist in a data base. This improves the access times and provides simpler addressing techniques. On the other hand, uncontrolled data redundancy has the following disadvantages (Martin, 1975):

- (a) Additional cost storing multiple copies.
- (b) Multiple updating of redundant copies.
- (c) Inconsistent information caused by different copies of data which may be in different stages of updating.

Data independence:

Data independence is one of the main requirements of a data base. It can be defined as the immunity of applications to changes in storage structure and access strategy. This means that the data and the application programs which use the data are independent from each other. Thus, if either one is changed, the other need not be altered. In practice, however, complete data independence rarely exists.

Due to two major reasons, a data base system is extremely undesirable which allows applications to be data dependent because:

(i) Different applications would need different views of the same data.

(ii) The data base administrator should have the freedom to change the storage structure or access strategy or both, in response to changing requirements, without the necessity of modifying existing applications.

Two levels of data independence exist, i.e.:

(i) Physical data independence.
(ii) Logical data independence.

Both levels will subsequently be described in more detail and are diagrammatically illustrated in figure 2.2.

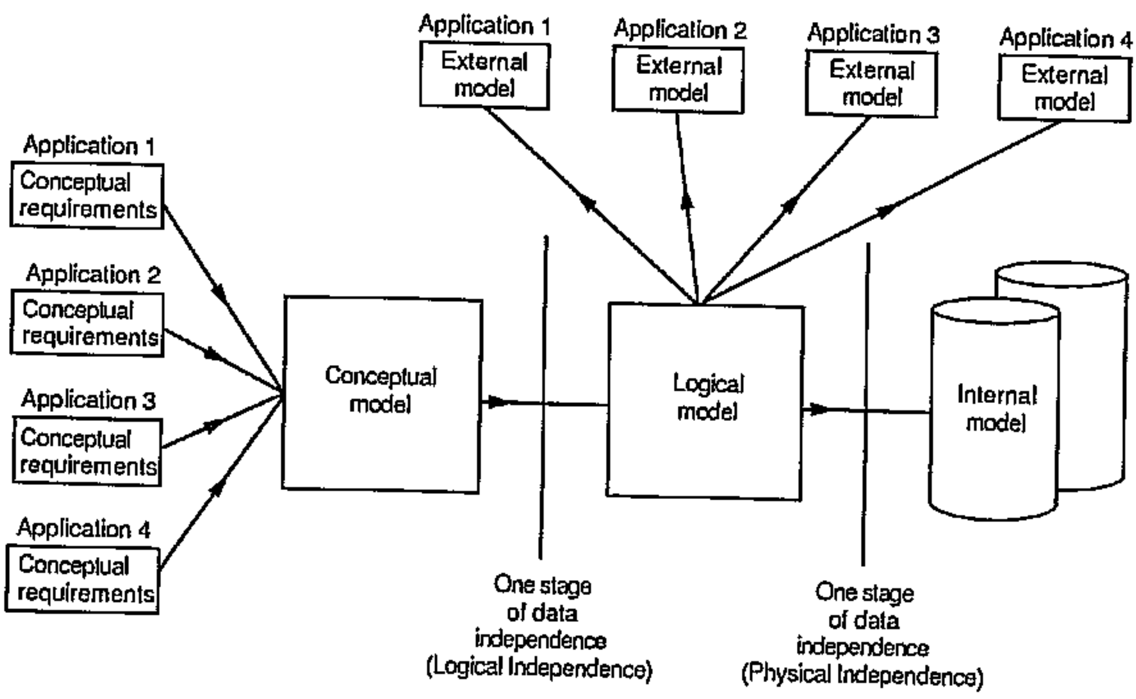


Figure 2.2. Two stages of data independence (after Atre, 1980).

(a) Physical data independence

The physical lay-out and organization of the data can be changed without changing either the logical structure of the data or the application programs. Any changes applied to the physical data base organization may, however, affect the efficiency of the application programs.

(b) Logical data independence

The overall logical structure of the data may also be altered without changing the application programs as long as no data are removed which are used by the applications.

The degree of data independence provided, depends on the thoroughness of the data base design. A thorough analysis of an enterprise's entities and their interdependence will minimize the impact of changing data requirements in one program on the requirements of others.

2.2 DATA BASE MODELS

Before discussing some of the existing data base models, the concept of data base design will be dealt with.

The task of data base design is to take the conceptual data model, which is the end user's views of the organization, and map these onto appropriate structures at the logical and physical level. The early design phase should produce a flexible data base structure, adaptable to a changing user environment, whereas the later design phase may emphasize optimization of performance.

Specific objectives and data base requirements should be obtained at the highest competent level in the enterprise. Among the requirements collected and documented, certain design objectives must be met:

(i) Efficiency.
(ii) Security.
(iii) Accuracy.
(iv) Compatibility.
(v) Implementability.
(vi) Maintainability.
(vii) Flexibility.
(viii) Portability.

Depending on the actual enterprise and its circumstances, a variety of these objectives will figure as being significant for any given logical data base model design.

Considering the above-mentioned objectives, figure 2.3 shows the steps to be taken when designing a data base.

The entities of an enterprise and the relationships between them can be represented by a data model. Data base management systems are based on either a specific data model or

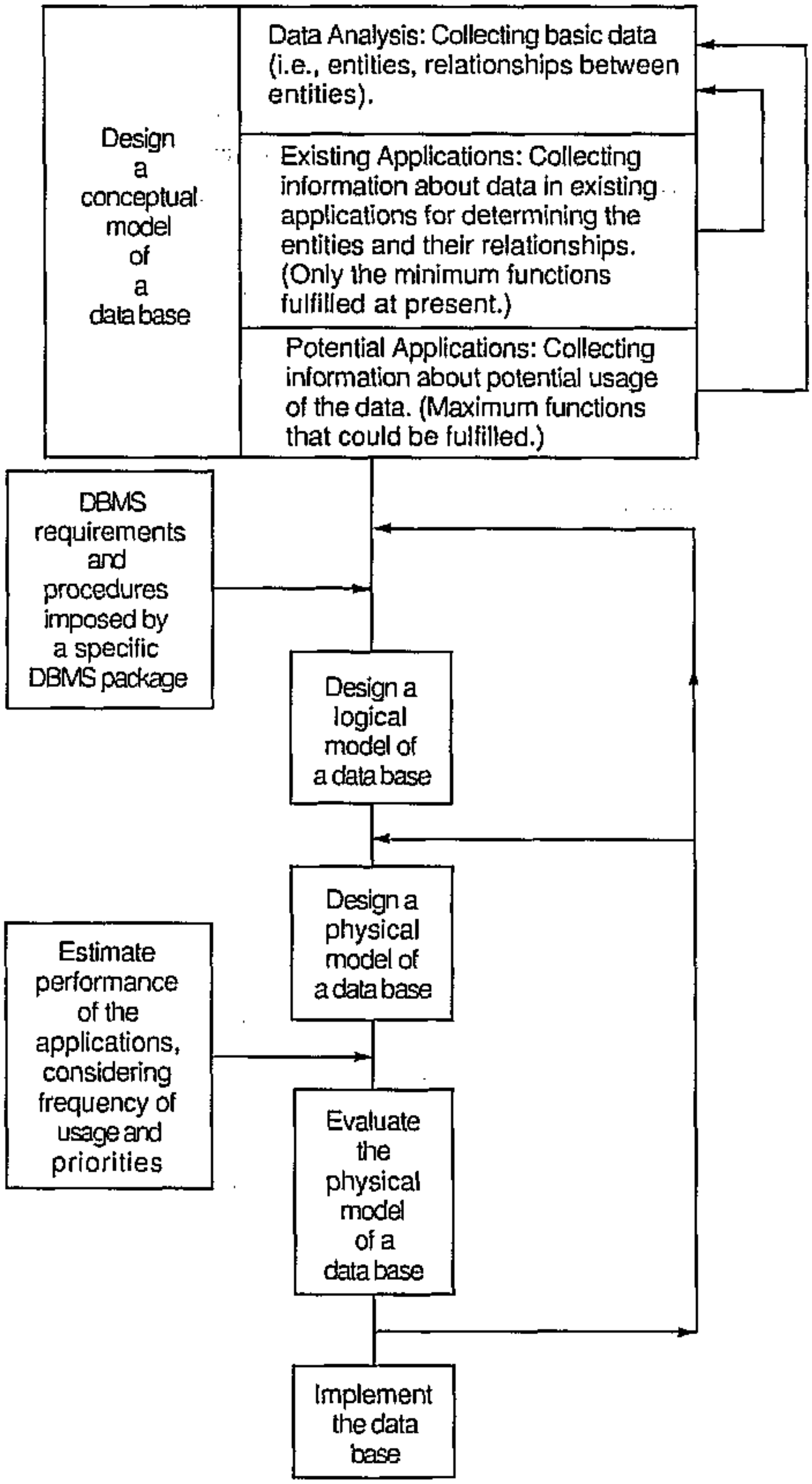


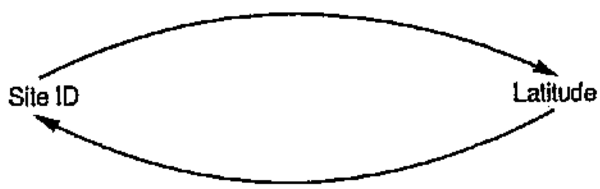
Figure 2.3. Data base design steps (after Atre, 1980)

a combination of data models. The main difference among the various models lies in the representation of the relationships between the entities.

2.2.1 Relationships within data models

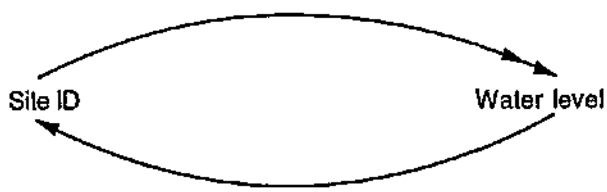
Associations between occurrences of data items or data sets are called relationships and can either be of a one-to-one, one-to-many or a many-to-many relationship.

One-to-One Relationship (between two data items):



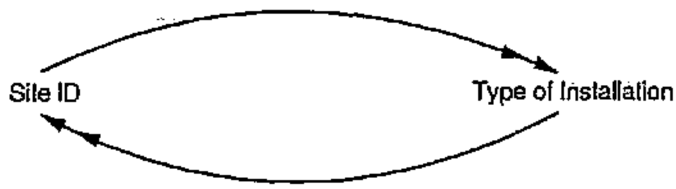
A borehole has been drilled. It has a specific latitude. This represents a one-to-one relationship between the site identification and the latitude.

One-to-Many Relationship (between two data items):



One or many different water levels may have been measured at a site as time increases. This constitutes a one-to-many relationship between the site identification and the water levels.

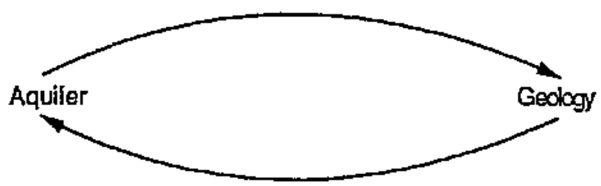
Many-to-Many Relationship (between data items):



During the course of time, one or more pumps may have been installed in a borehole or a pump might have been used in various boreholes. This is called a many-to-many relationship between the site identification and the type of installation.

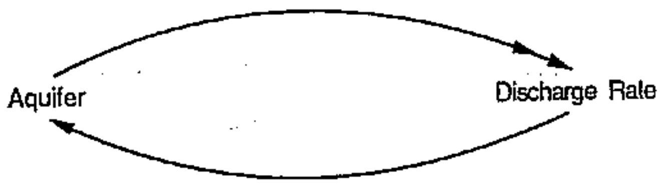
The relationships between data items are part of the conceptual model, and they have to be represented in a data base. Any number of data items can participate in a relationship. On the other hand, the same data sets can participate in a number of relationships.

One-to-one Relationship (between two data sets):



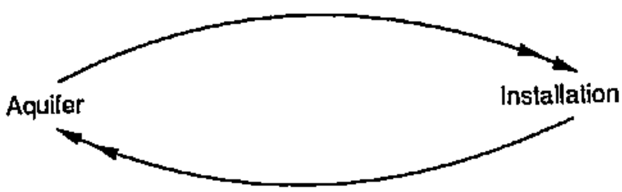
A specific aquifer is characterized by specific geology, while a combination of specific geological features leads to a specific aquifer type. This results in a one-to-one relationship between the Aquifer and Geology data sets.

One-to-Many Relationship (between two data sets):



Different discharge rates may have been measured in an aquifer at different depths or at different times. This constitutes a one-to-many relationship between the Aquifer and Discharge rate data sets.

Many-to-many Relationship (between two data sets):



Many installations can be made in an aquifer, while a specific installation may have been done in various aquifers. This is a many-to-many relationship between Aquifer and Installation data sets.

After the explanation of the various relationships, existing data models will be described. These are:

(i) Flat file model.
(ii) Relational model.
(iii) Hierarchical model.
(iv) Network model.

2.2.2 Flat file model

The most common way of associating a value with a data item and associating one data item with other data items of the respective entities, is to store the data items together in a fixed order, as illustrated in table 2.1.

The simple two-dimensional lay-out of data elements in this table is referred to as a flat arrangement or better known as a flat file. The table must be set up in such a way that no information about the relations between data items is lost. Tables or flat files have the following properties (Martin, 1975):

(i) Each entry in a table represents one data item; there are no repeating groups.

(ii) Flat files are column homogeneous; that is in any column all items are of the same kind.

(iii) Each column is assigned a distinct name.

(iv) All rows are distinct; duplicate rows are not allowed.

(v) Both rows and columns can be viewed in any sequence at any time without affecting either the information content or the semantics of any function using the table.

TABLE 2.1 TWO-DIMENSIONAL DATA LAY-OUT (FLAT FILE).

Borehole number	Farm name	Latitude (°)	Longitude (°)	Altitude (mamsl)	Water Level (m)	Yield (l/s)	Equip-ment
1	Rietvlei	25,1234	25,1000	1451,00	10,12	1,5	Mono
1	Spitskop	25,5900	25,1900	1456,00	15,00	2,6	Mono
1	Graslaagte	26,1000	25,5000	1396,00	9,50	10,5	Turbine
2	Graslaagte	25,9000	25,6000	1421,00	20,60	0,5	W.P.
1	Mooifontein	25,8000	25,7500	1501,00	34,10	1,0	W.P.

2.2.3 Relational data model

Relational data models are constructed from a flat file arrangement of data items.

The mathematical concept underlying the relational model is the set theoretic relation, which is a subset of the Cartesian product of a list of domains. A domain, however, is simply a set of values. For example, the set of integers is a domain. The Cartesian product of domains $D_1, D_2, ..., D_k$ written as $D_1 \times D_2 \times ... \times D_k$ is a set of all k-tuples $(v_1, v_2, ..., v_k)$ such that v_1 is in D_1, v_2 is in D_2 and so on. If $k = 2$, $D_1 = \{0,1\}$ and $D_2 = \{a,b,c\}$, then $D_1 \times D_2$ is $\{(0,a), (0,b), (0,c), (1,a), (1,b), (1,c)\}$. Thus, a relation is any subset of the Cartesian product of one or more domains. For instance, $\{(0,a), (0,b), (0,c)\}$ is a relation, a subset of $D_1 \times D_2$ defined above. The members of a relation are called tuples. Each relation that is a subset of $D_1 \times D_2 \times ... \times D_k$ is said to have arity (degree) k. A tuple $(v_1, v_2, ..., v_k)$ has k components, the i-th component being v_i. Often one can view a relation as a table where each row is a tuple and each column corresponds to one component, generally known as attributes (Ullman, 1980). For a graphical representation of a relational data model, see figure 2.11(b).

In table 2.2, a simple relation is defined, whose attributes are District, Drainage region and Number of boreholes.

TABLE 2.2. EXAMPLE OF A SIMPLE RELATION.

District	**Drainage region**	**Number of Boreholes**
Bloemfontein	352	1200
Pretoria	123	10000
Cape Town	720	5000
Durban	1960	12000

The arity (degree) of this relation is three.

If attribute names are attached to the columns of a relation, then the order of the columns becomes irrelevant. In mathematical terms, tuples are viewed as mappings from attribute names to values in the domains of the attributes, making certain relations equal which were not equal under the traditional definition of a relation (Ullman, 1980).

To summarize, the following important properties of relations have been observed (Date, 1975):

(i) All values within the relation are atomic data items; i.e. they cannot be split further without any loss of information.

(ii) Each column has a unique name.

(iii) No duplicate rows are allowed. The relation is a set with tuples as elements and by definition a set is not supposed to have any duplicate elements.

(iv) Row and column order become insignificant.

A relation, satisfying property (i), is normalized, i.e. any non-flat data structure is transformed by the data base designer into a set of flat relations that have no repeating groups. This is one of the main features of a relational data base. However, the tuples in the relation must have keys which uniquely identify them, i.e. certain data items may be repeated while serving as an identifier. This might increase redundancy to a certain extent, but it implies not necessarily increased storage requirements, because normalization is concerned with the logical structures and not with the physical implementation of the data storage.

Normalized data have certain advantages, namely:

(i) Simplicity - end users are confronted with a simplistic representation of data in the form of two-dimensional tables.

(ii) Flexibility - editing relations in such a way that the different logical files wanted, are given in the form requested.

(iii) Precision - relations are based on a well-developed and precise theory of relations.

(iv) Security - security controls are easily implemented.

(v) Relatability - attributes from different sets or flat structures can be related.

(vi) Ease of implementation - storing of flat structures can be less complex than storing tree or network structures.

(vii) Data independence - if in a normalized form, any changes to the data base relations can generally reduce major changes of application programs. This means that good data independence is more easily achieved.

(viii) Data manipulation language - the precise mathematical foundations of relations form a sound base for a data manipulation language.

As the data bases of an enterprise grow, the number of interrelationships also increase and make it difficult to efficiently represent these relationships with directed links (pointers). The constant requirement for growth and change makes it desirable to use normalized data structures, i.e. relational data bases.

Operating any relational data base include certain basic operations by which the data base is modified. They are the INSERT, DELETE and MODIFY instructions.

The implementation of these three operations will be described under the assumption that the relations are represented by files in which tuples and data records are in a one-to-one correspondence, which make these operations straightforward. To insert or delete a tuple, a relation is needed in which to insert, and the values of those attributes that form the key for the underlying file. The tuple is converted to a logical record, after which it is inserted or deleted in the file by the appropriate file organization. Many systems allow deletion without a key. This, however, necessitates a scan of the entire file, irrespective of the file organizations used. In the case of the modification of data, the relation, the values for the key attributes and new values are required. The attribute values are translated to field values, whereafter they are modified by the appropriate procedure. If no values for the key attributes are given, then the entire file is once again scanned.

2.2.4 Hierarchical model

This particular model has been, and is still quite frequently, used in many systems. It has its origin in storage structures, which were used when most data processing was performed with sequential media, e.g. magnetic tape or punched cards. The distinction between the data model and the storage structure is almost non-existent.

A hierarchy is a tree in which all links point in the direction from child to parent. Trees are generally used in both logical and physical data descriptions. A tree consists of a number of nodes (data elements) arranged in a hierarchy, where each node is linked to a node at a higher level. The latter is called the parent node and may have one or more nodes (elements) at the next lower level, which are the children. A child node cannot have more than one parent.

Nodes are interconnected by branches. The highest node in a particular tree structure is called the root node, while the nodes at the lowest level have no children and are called leaves. The height of the tree is dependent on the number of hierarchic levels involved. Figure 2.4 is an example of a hierarchical tree structure.

Each occurrence of a root node is the beginning of a new logical data record, forming a hierarchical data base.

The following conditions have to be satisfied when designing a hierarchical data base (Atre, 1980):

(i) The data base always has to start with a root node.

(ii) Nodes (data elements) consist of one or more attributes, describing the entity of the particular node.

(iii) Dependent nodes (children) can follow on succeeding levels. The node on the preceeding level becomes the parent node of the new child node. Dependent nodes can be added horizontally, as well as vertically, with no limitations. Level 1, however, can only have one node (root).

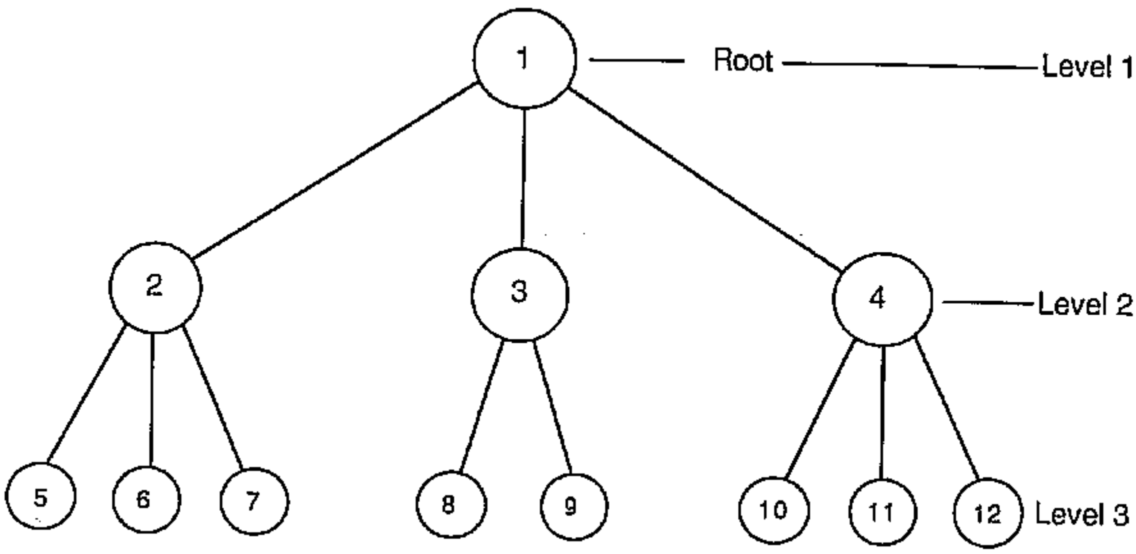


Figure 2.4. *Hierarchical tree structure.*

(iv) Nodes at level 2 are connected with one and only one node at level 1. Further, each node at level 3 is connected to one and only one node at level 2 and so forth. Due to the fact that there can only be one connection between any two nodes, the connections do not need labels.

(v) Parent nodes can have one or more dependent child nodes. If, however, it does not have any dependent nodes, it is not a parent node.

(vi) Any node can only be accessed through it's parent node. This results in a true hierarchy. Thus, a hierarchical data model consists only of linear paths (see figure 2.5), making the access path to each node within the structure, unique.

(vii) A node can occur a number of times at a certain level. However, each occurrence has to be connected with a parent node occurrence. Every node occurrence starts a logical record.

The choice of a hierarchical data model should be made with its performance in mind. One-to-one and one-to-many relationships are especially suitable for hierarchical data structures. To a certain extent, many-to-many relationships can also be implemented in this data model. This may, however, lead to unnecessary redundancy in the stored data. A certain measure of redundancy at a logical level does promote simplicity to the structure. Redundancy at the physical level, however, is totally undesirable.

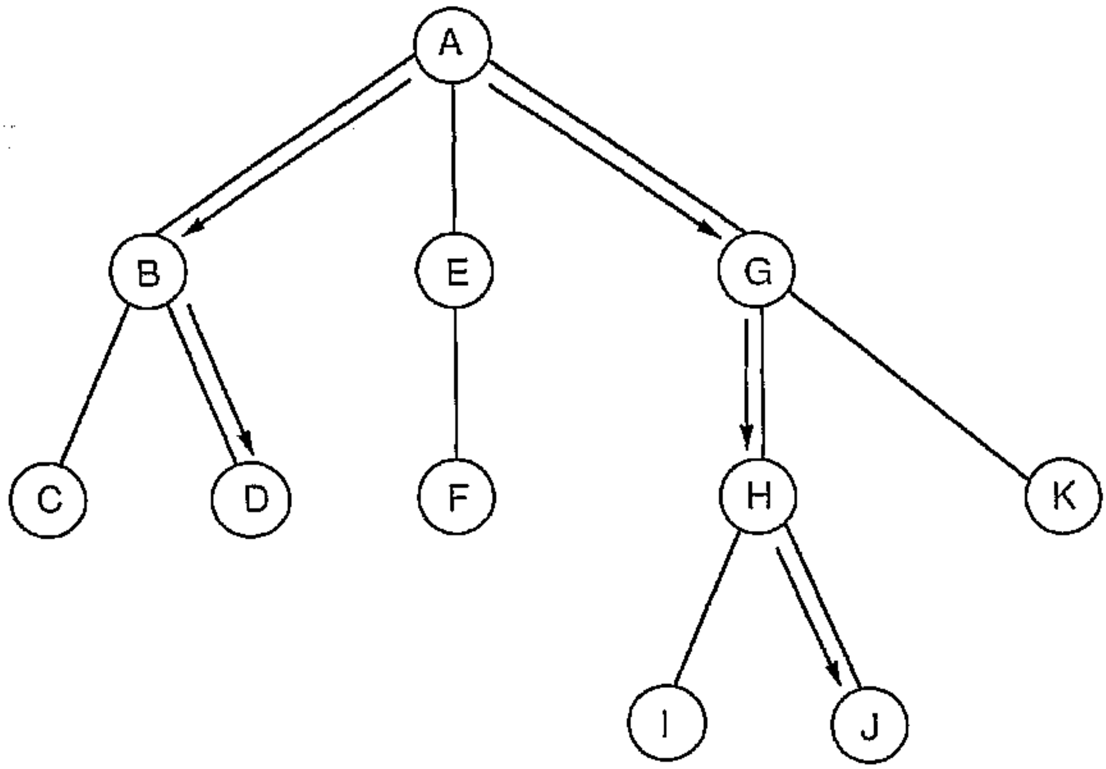


Figure 2.5. Access paths of a hierarchical data model.

Information in a hierarchical data model can be represented in various ways, one of which is shown in figure 2.6.

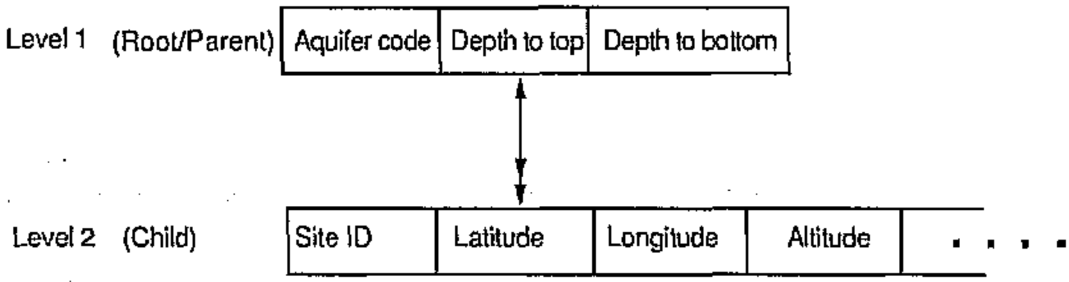
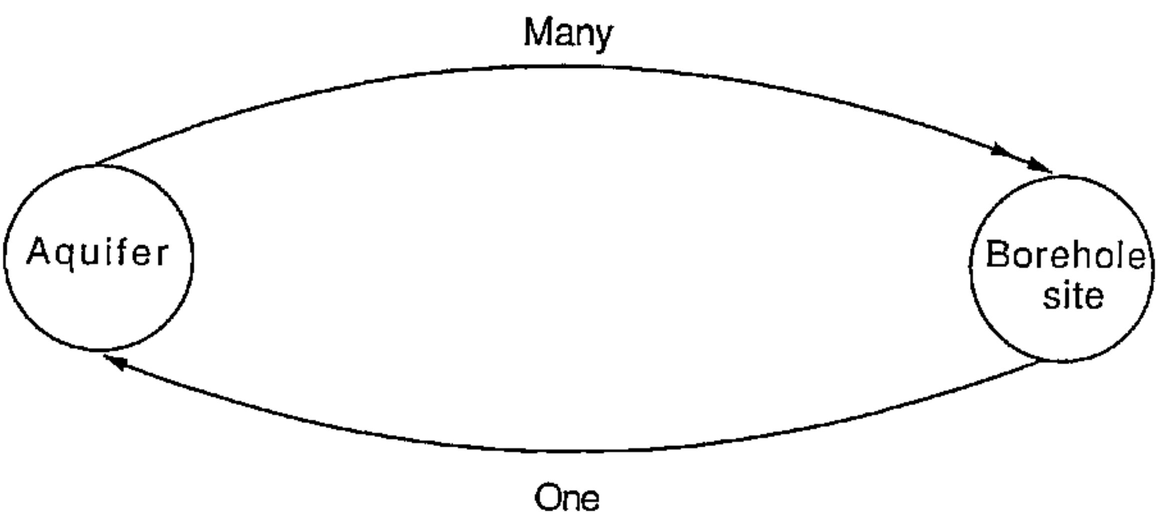


Figure 2.6. Representation of an aquifer data base.

Aquifer is the root node in the structure, shown in figure 2.6 This structure enables us to store information on many borehole sites for a specific aquifer, i.e.:



In another structure, the entity borehole site is the root or parent node and aquifer is the dependent node, i.e. a number of aquifers may occur at a specific site. Figure 2.7 is a diagrammatic representation of this particular structure.

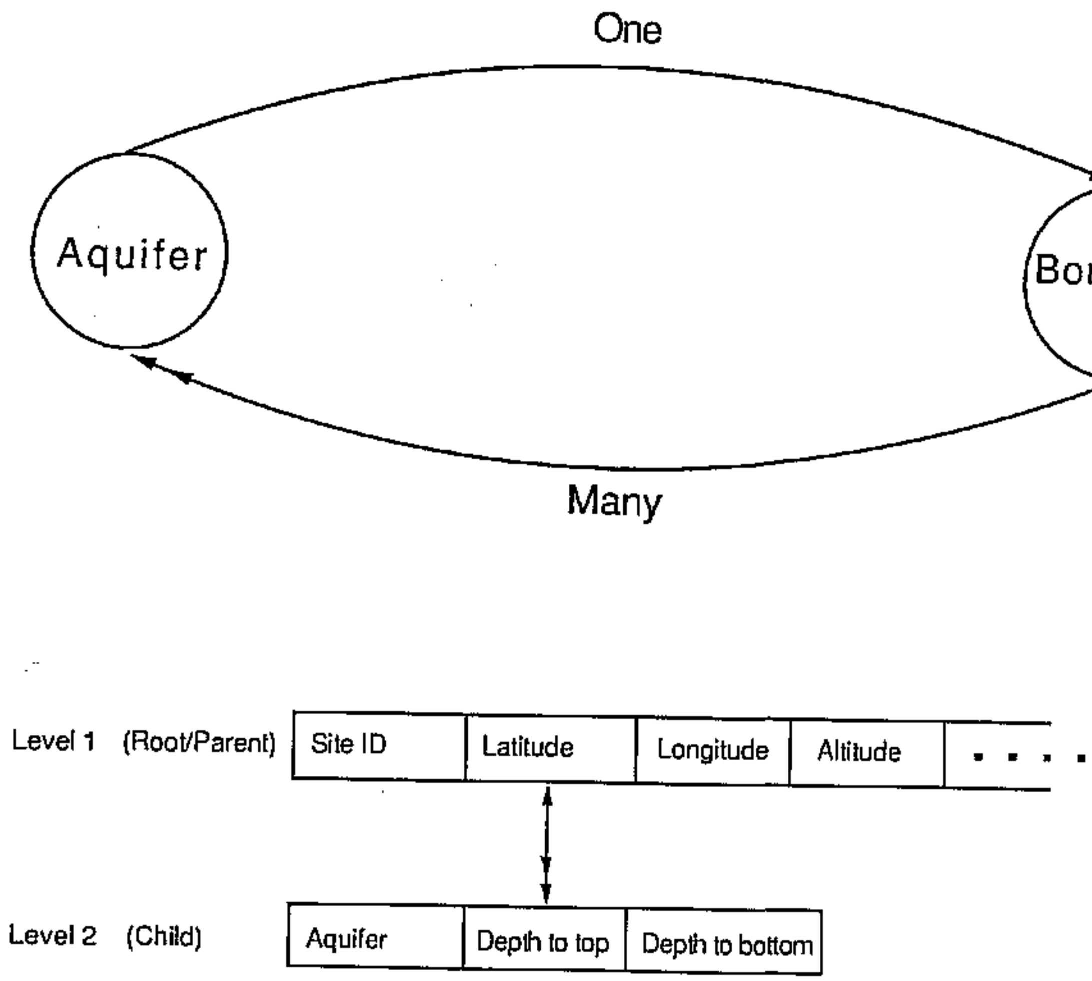
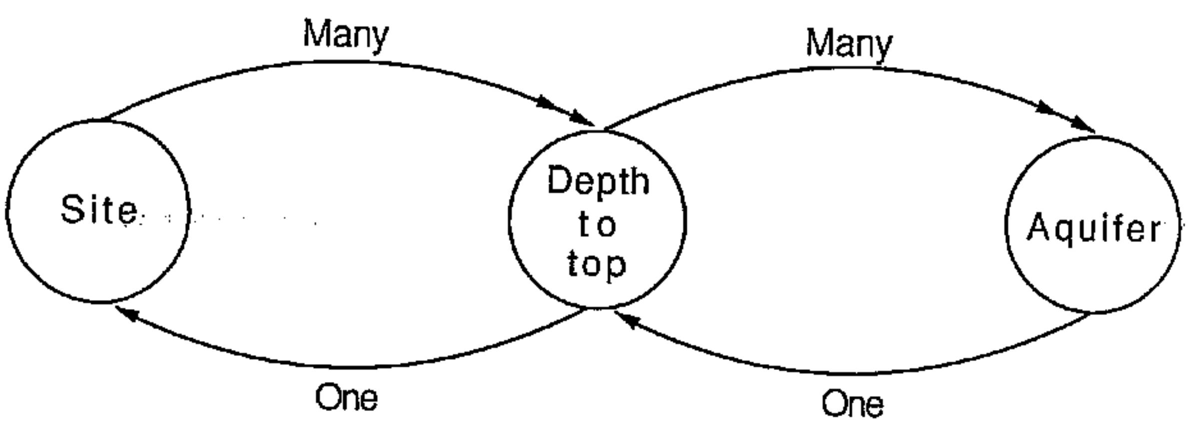


Figure 2.7. Representation of a site data base.



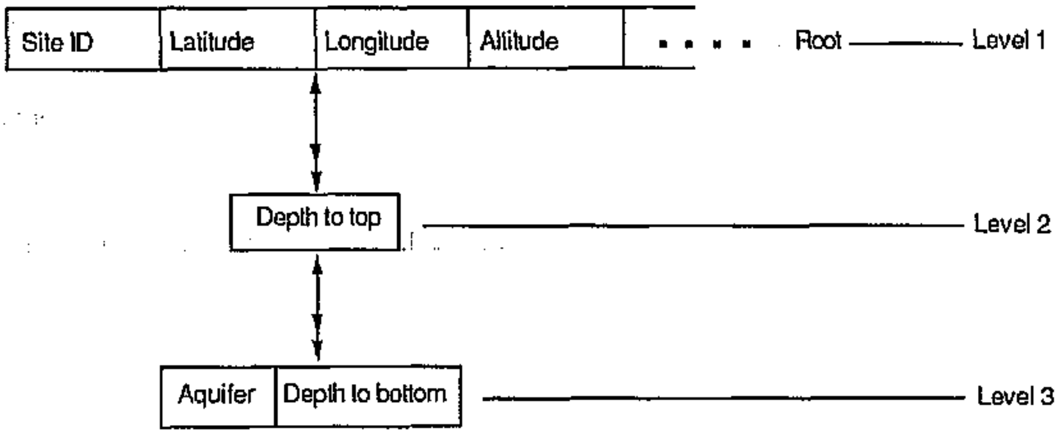


Figure 2.8. Representation of a site data base with different depths to top and (all) the respective aquifers for a specific depth.

A third way of organizing the same data is shown in figure 2.8. This structure allows many depths to top for a specific site and many (different) aquifers for the respective depths to top to be entered.

Insertion or addition of new data must occur through the relevant parent node, i.e. a child node cannot exist without a parent node occurrence. Deleting certain data may cause a loss of existing data, i.e. if a parent node is deleted, the depending nodes (children) are also deleted.

The reason for anomalies arising from some storage operations must be ascribed to the fact that a many-to-many relationship cannot be efficiently represented within a hierarchical structure. Such anomalies can be partially resolved by introducing two tree structures with an established relation. This may, however, lead to redundancy at a physical level.

2.2.5 Network model

The network data model is a more general structure than the hierarchical model, due to the fact that any child node may have more than one immediate superiors. This enables the structure to accommodate many-to-many relationships. In figure 2.9 four different network structures are shown.In all four diagrams it is illustrated that the child node has more than one parent node. This means that any item in a network structure can be linked to any other item. Many network structures show that the mapping between the parent and child node is similar to a hierarchical structure. The parent to child mapping is often complex, whereas the child to parent mapping is generally simple. An important distinction between a simple and complex network structure is that the complex structure needs more elaborate methods to link the data physically. This is not necessarily a disadvantage, because a complex structure can always be represented by simple structures, e.g. a tree structure or a set of tree structures, with redundant elements (see figure 2.10).

In all four diagrams it is illustrated that the child node has more than one parent node. This means that any item in a network structure can be linked to any other item. Many network structures show that the mapping between the parent and child node is similar to a hierarchical structure. The parent to child mapping is often complex, whereas the child to parent mapping is generally simple. An important distinction between a simple and complex network structure is that the complex structure needs more elaborate methods to link the data physically. This is not necessarily a disadvantage, because a complex structure can always be represented by simple structures, e.g. a tree structure or a set of tree structures, with redundant elements (see figure 2.10).

This redundancy does not mean physical redundancy. Physical data lay-outs that work well with tree structures do not necessarily perform likewise with network structures (Martin, 1975). They may require a rather complex structure. Such complexity demands a thorough knowledge of the logical structure.

As in the previous models, the basic commands to change data in network data base are INSERT, DELETE and MODIFY. According to Ullman (1980), any of the above operations performed on the data model must consider both logical record types and links. The principles for all three operations are the same, therefore only one command will be discussed.

In order to insert a logical record (entity) into the data base, a physical record has to be created from the logical record. The record is then inserted into a file representing the logical record type. An advantage when deleting an occurrence is that not all the dependent nodes are necessarily deleted, but that some are left intact.

2.2.6 Model comparison

For the evaluation of the three relevant data models, namely the relational, hierarchical and network models, the following criteria have to be compared by which their performance can be judged. They are:

(i) Ease of use.

(ii) Efficiency of implementation.

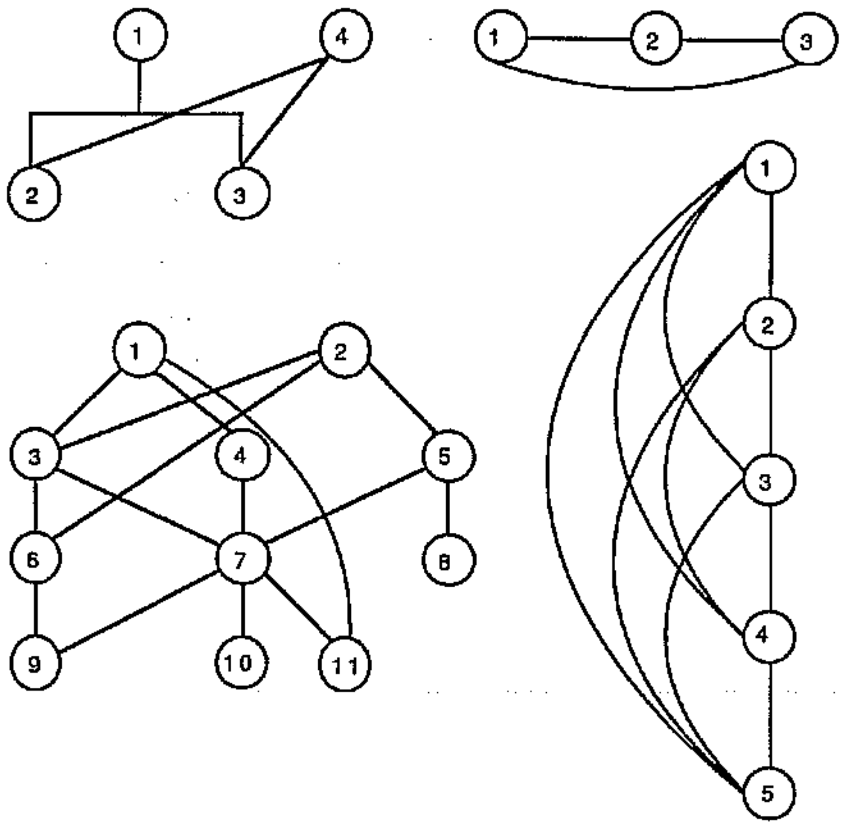


Figure 2.9. *Representation of four different network data modes (after Martin, 1975).*

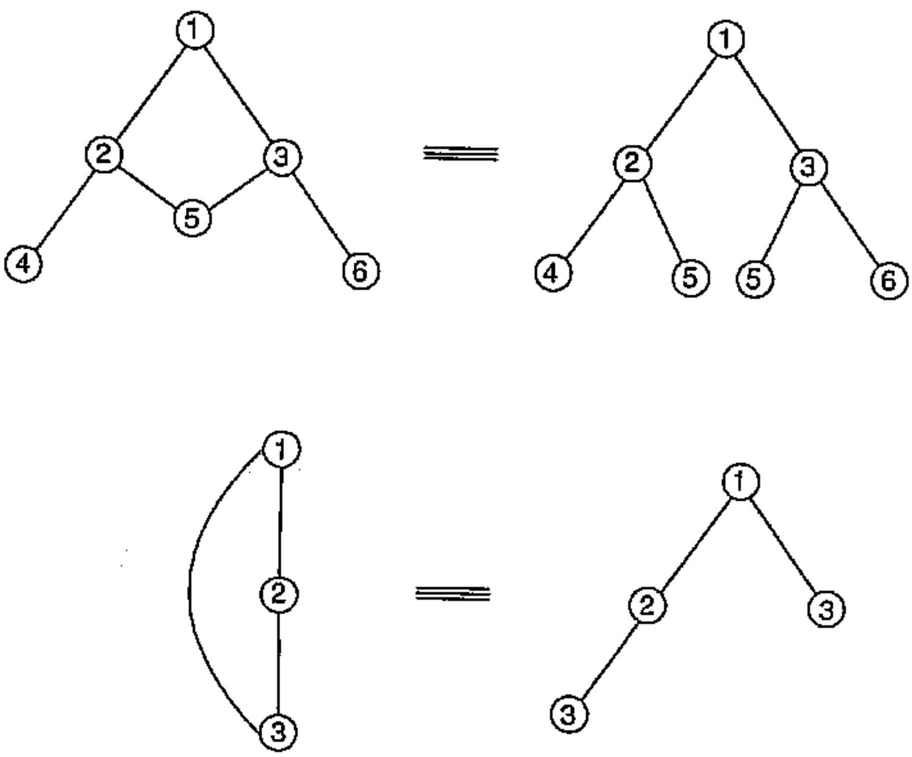


Figure 2.10. *Simplification of complex network structures.*

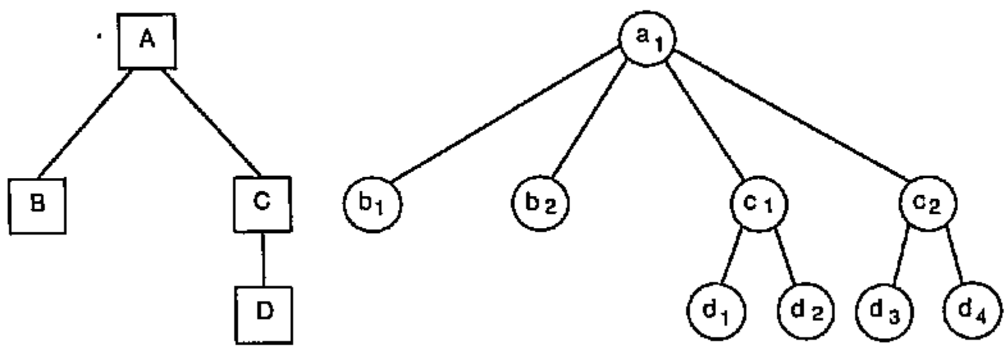
(iii) Efficient utilization.

Considering the first criterium, the *relational model* has the simpler structure of the three models. Furthermore, numerous high-level computer languages support the relational model, making systems based on it, available to persons with a low programming level.

The *network model*, on the other hand, requires an extensive knowledge of both record types and links and their relationships. Although networks are able to handle many-to-many relationships, the implementation thereof is not always straightforward. This also applies to *hierarchical models* which require extensive knowledge of the use of pointers.

Efficient implementation, on the other hand, suits the latter two models more, because data structures as the multi-list and the pointer-based implementation of variable length records, do not generally map to many-to-many relations. Since a relational model often represents many-to-many relations, efficient implementation of the data model can become difficult without specialized data structures.

(a)



(b)

A	B
a_1	b_1
a_1	b_2

A	C
a_1	c_1
a_1	c_2

C	D
c_1	d_1
c_1	d_2
c_2	d_3
c_2	d_4 → Tuple

Figure 2.11. Data base represented by a hierarchical and relational model (after Ullman, 1980).

Efficient storage utilization depends, to a certain extent, on the amount of data redundancy. Hierarchical models, if properly designed, tend to minimize data redundancy. For instance, figure 2.11 shows an example data base represented by a hierarchical model (2.11 (a)) and a relational model (2.11(b)). On the left side of figure 2.11(a), a basic tree structure is shown, which can be expanded to a more detailed structure, as depicted on the right side of this figure. If the tree structure in figure 2.11(a) is converted to a relational data structure, the basic data structure is represented by the capital letters, while the detail, resembling the second part of figure 2.11(a), is shown in small letters.

It is evident how many times a_1, c_1 and c_2 are repeated in the relational model, illustrating its data redundancy.

In the past, most data bases have been based on either the hierarchical or the network model. Relational models, however, have been receiving more attention lately, for two reasons. Firstly, design concepts originally applying to large data bases, are now used for medium and small data bases. Secondly, many of the existing apparent inefficiencies of the model can be eliminated by numerous optimization techniques (Ullman, 1980).

2.3 EXISTING GROUND-WATER DATA BASES

Before applying any of the previously mentioned principles for the design of the new ground-water data base, a closer study of some existing data bases has be made to avoid unnecessary design mistakes. Four data bases were studied in more detail and are subsequently discussed:

(i) Data base by the British Geological Survey (U.K.).

(ii) Data base by the Public Works Department (Western Australia).

(iii) Borehole data base - S.W.A. (Namibia).

(iv) WATSTORE - U.S.G.S.

2.3.1 Data base by the British Geological Survey (U.K.)

As early as 1835, well records have been collected on a voluntary basis by the Geological Survey of Great Britain. Since 1945, however, it has become mandatory to submit hydrological related data to the Geological Survey about all newly drilled wells. Up to 1973, a total of some 90 000 records of wells, springs and exploratory boreholes were in the possession of the Institute of Geological Sciences (former Geological Survey). A storage retrieval system was developed to ease problems arising with the processing of these data.

The main objective was to provide a method whereby a preliminary search of those wells or hydrological stations can be done by computer, which may need closer scrutiny of detailed observations and of data recorded in their original form (Harvey, 1973).

Recorded data may be extremely variable in both quantity and quality. They can be subdivided into five primary groups:

- (i) Record identity.
- (ii) Well construction.
- (iii) Geohydrological details.
- (iv) Stratigraphical information.
- (v) Hydrochemical information.

Certain restrictions were applied, allowing limited storage and retrieval of only the most basic data. Parameters, such as lithological information or continuous measurements of water levels and yields during a pumping test, were not included in the system. Instead, the stored data include the geological classification and thickness of the strata and yield and pumping water levels at equilibrium. Similarly, only selected chemical data have been included in the data base.

The data arrangement for each well or spring has been grouped into nine subject categories or data lists. Within each subject, the items are entered chronologically, where applicable. The input data are similar in content to those published in the Well Catalogue of the Water Supply Papers of the Geological Survey of Great Britain (Harvey, 1973).

Each well comprises the following data categories:

- (i) Well identity.
- (ii) Stratigraphical succession and thickness.
- (iii) Lining type and aquifer.
- (iv) Well dimensions.
- (v) Rest water level and date.
- (vi) Pumping water level and date.
- (vii) Yield and date.
- (viii) Hardness, chloride and date.
- (ix) Chemical data, mineral analysis and date.

Data storage structure:

Well identity is regarded as the basic common link among wells. For each well, on the other hand, the lining type and aquifer are used as common links to all lists, with the exception of the stratigraphical succession and thickness.

Development stages of a well are stored in separate units, linked by pointers at the succeeding stage. All data lists, with the exception of the well identity, can be extended to accommodate additional information. Figure 2.12 is a diagrammatic representation of the data structure.

Storage of data concerning groups of wells differs, in that no information on stratigraphy and thickness are included.

Data input:

Data for each site are entered in ten logical records (see figure 2.13). The first record indicates whether it is a single site or a group of sites. The remaining records are used to enter the data of each of the nine data lists. The input format for the data is defined by a series of input programs.

Logical records may consist of a number of 80-character cards, depending on the number of data sets recorded for each data list. However, the number of locations per well allocated in the well identity and well dimension list are fixed. Figure 2.13 shows the data input coding form.

Storage on magnetic tape:

Sorted geographically, according to the coordinates of the National Grid System, the data are stored on tapes. The grid system consists of a network of 10 km squares, which are taken as storage units respectively. Thus, all information about hydrological related sites within such a square is stored together in suites of blocks on tape (Harvey, 1973).

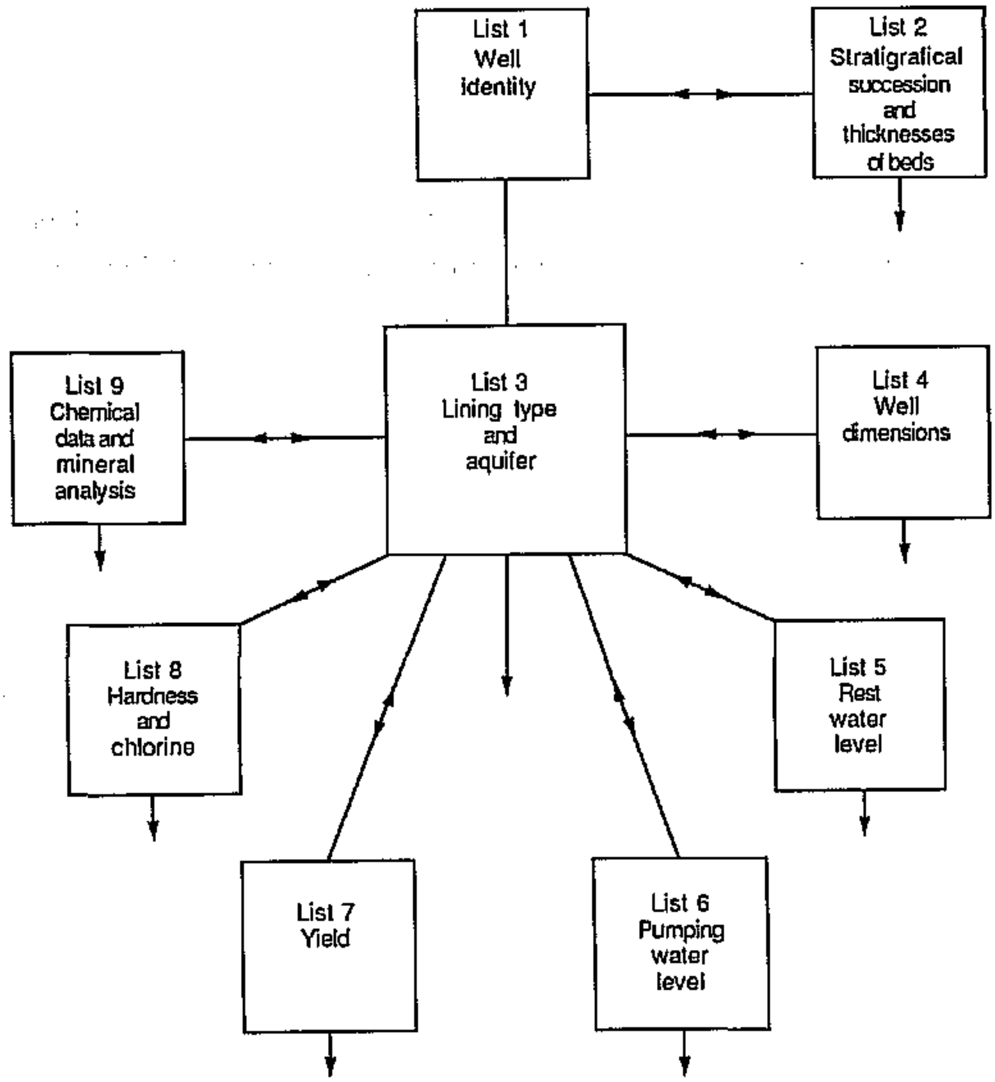


Figure 2.12. *Data base structure (after Harvey, 1973).*

Data retrieval:

Data retrieval is initiated by allowing a questioner to present questions in a coded form. It is very important that correct words and punctuation are used. The questions can be divided into two parts, i.e. one part defining what the user wants printed and, secondly, a part laying down various criteria by which the user selects the relevant data (Harvey, 1973).

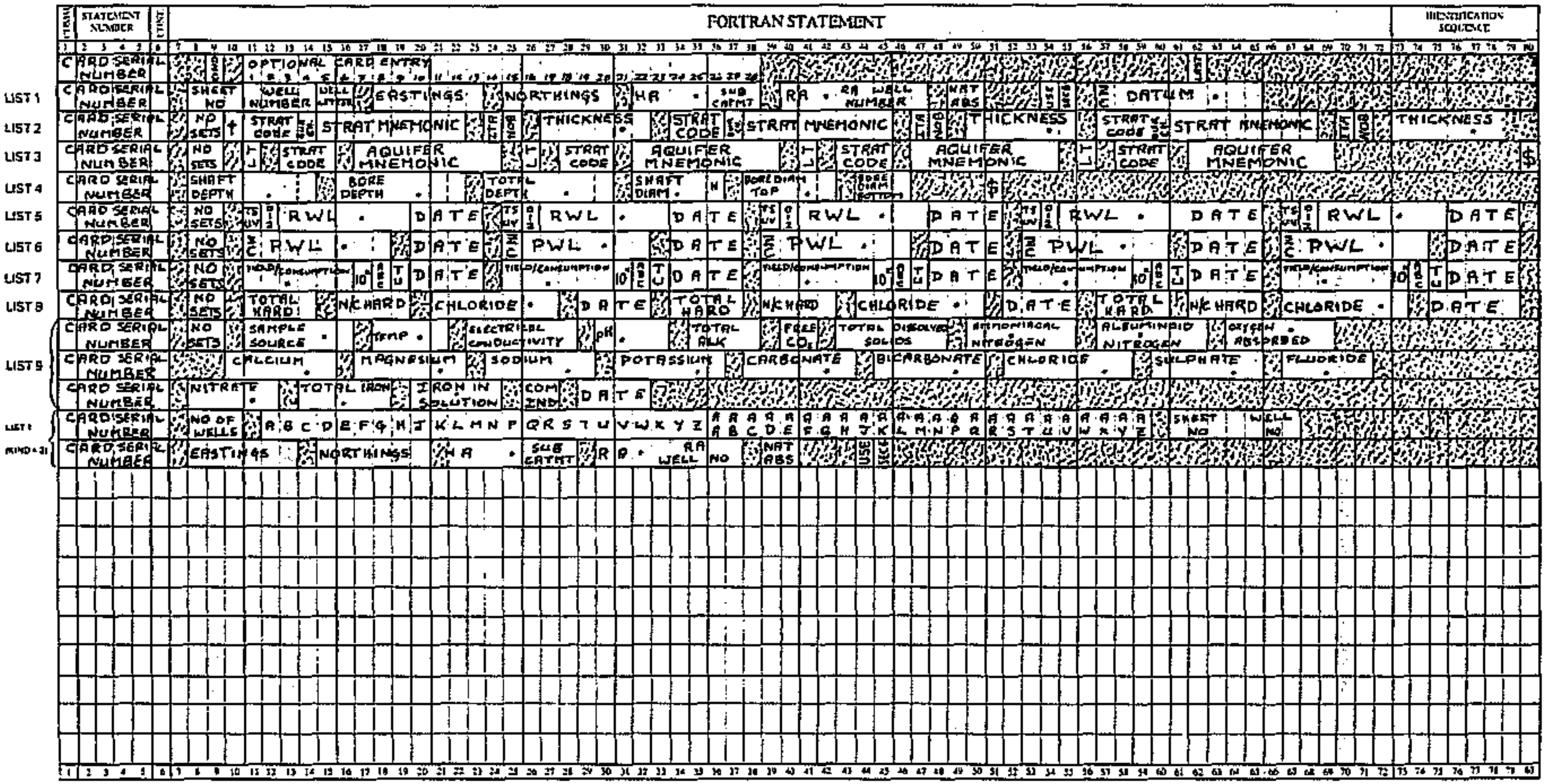


Figure 2.13. Fixed format input coding sheet (after Harvey, 1973).

Discussion:

The storage and retrieval programs of the data base were written in FORTRAN V, so that the basic machine code could be used to manipulate the magnetic tape and to evaluate the input. FORTRAN V is ideally suited for scientific calculations (numerical calculations). However, if used to manipulate various types of data, as well as big data files, it becomes less efficient.

The data base was implemented on an IBM 1130 mainframe computer. The data are stored on magnetic tape, which can only store and access data sequentially. Therefore, records have to be accessed from the beginning of the magnetic tape, when searching for any specific data. This makes data retrieval rather inefficient. The situation may have changed, however, as computer technology has vastly improved over the past years.

The data base structure resembles a hierarchical model and was created in such a way that additional information can be accommodated. Although the data accommodated in the data base include some of the more important borehole parameters, not enough basic data (e.g. use, potability or borehole installation) are accommodated in the data base. Codes have been used to some degree to allow for uniformity and to save storage space.

Extraction programs allow for the search and selection of data according to a wide range of options. These require a measure of programming skills and the knowledge of acceptable codes. Data output is provided in the form of line printer print-outs. Further data processing facilities are not provided.

Difficulties encountered during development and operation of the data base include the following:

Accommodation of both borehole and spring information in the same record.

Storage of data from single boreholes and group of boreholes.

Variable number of items and quantity for individual sites.

Missing data.

Recording of data from different stages of development of a particular well.

Facilities for later adding and updating of data.

2.3.2 Data base of the Water Resources Section of the Public Works Department of Western Australia

Water resources in Western Australia have to be managed carefully to satisfy the likely levels of demand. This has resulted in an increase of water resources data collection and has led to the development of a water resources management system.

The primary data accommodated in this information system are streamflow, pluviograph and water quality data.

Parameters:

The majority of the relevant data are recorded on a continuous basis, comprising the following parameters:

- (i) Stream-level data.
- (ii) Rainfall data.
- (iii) Reservoir level data.
- (iv) Borehole water levels.
- (v) Conductivity measurements.
- (vi) Temperature, evaporation, humidity and radiation data.

Data from a number of sites are only recorded intermittently by field personnel, due to their lack of fluctuation.

Data base structure:

Flexible data file structures have been designed which provide the links between the data of different data sets. They also cater for many different forms of data to be held. The data base system consists of four major storage files (i to iv), namely:

- (i) Station establishment and registration file.
- (ii) Edit data base.
- (iii) Water quality and non-continuous data base.
- (iv) Rating curve data base.

Two other file structures are also important:

- (i) Working file.
- (ii) Daily data file.

A flow diagram for the different components is displayed in figure 2.14.

Figures 2.15(a), (b), (c) and (d) show flow diagrams for the Edit- and Rating curve data base.

Data entry:

Data entry is restricted to measuring stations registered on the computer. Whenever a station is registered, it is assigned a number, based on the type of information collected at that specific site. Four station types exist, i.e.:

(i) Stream stations - Gauging stations where flow and water quality are continuously measured.

(ii) Sampling stations - Sampling of intermittent data.

(iii) Ground-water stations - Where water levels, water quality and borehole yields are measured continuously or non-continuously.

(iv) Metereological stations - Measurement of meteorological parameters.

Station establishment and registration file - All information comprising the station registration details is kept here. This file is constantly in use during any kind of data operation. It writes the station name in the header of the data listings, enables programs to produce map overlays and gives the station location (Harvey, 1978).

Edit data base - All continuously recorded data are accommodated in this specific file. However, not all the continuous data can be held resident on the file. Therefore, older and little accessed data are archived. Collected data, when initially entered, are first processed by a program EVCON, which produces a file in the PREEDIT format. These data are then processed by the program EDIT which enters data into the data base and retrieves it from the data base.

Data can be accessed in two ways, i.e. either by specifying the station and variable number and the range and time for which the data are required. Secondly, data can be retrieved by entering the number of the chart wanted (Harvey, 1978).

Water quality and non-continuous data base - Only the non-continuous data are stored here. Due to the many different data types stored in this file, it is very flexible,

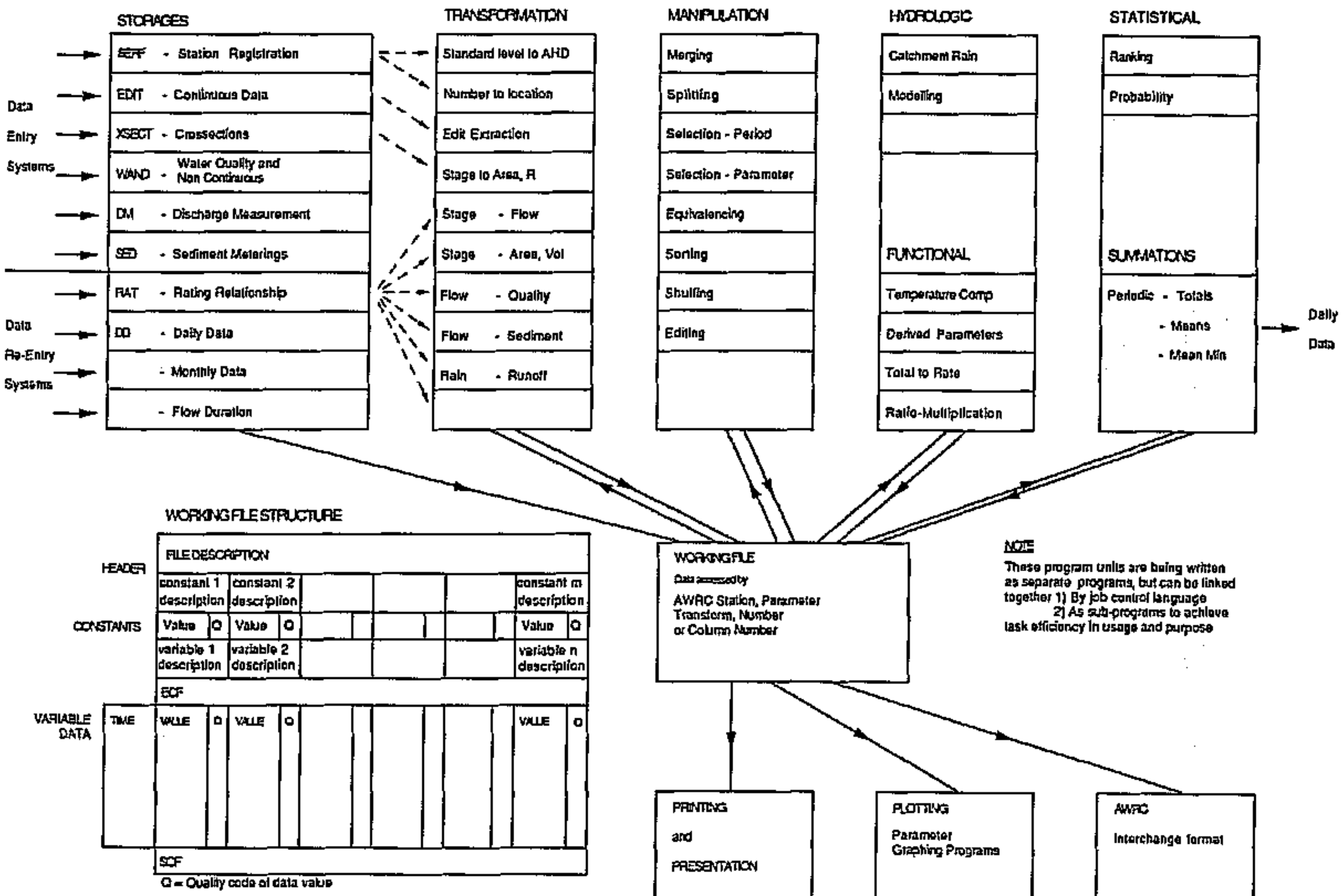


Figure 2.14. Western Australia data base structure (after Harvey, 1978).

(a).

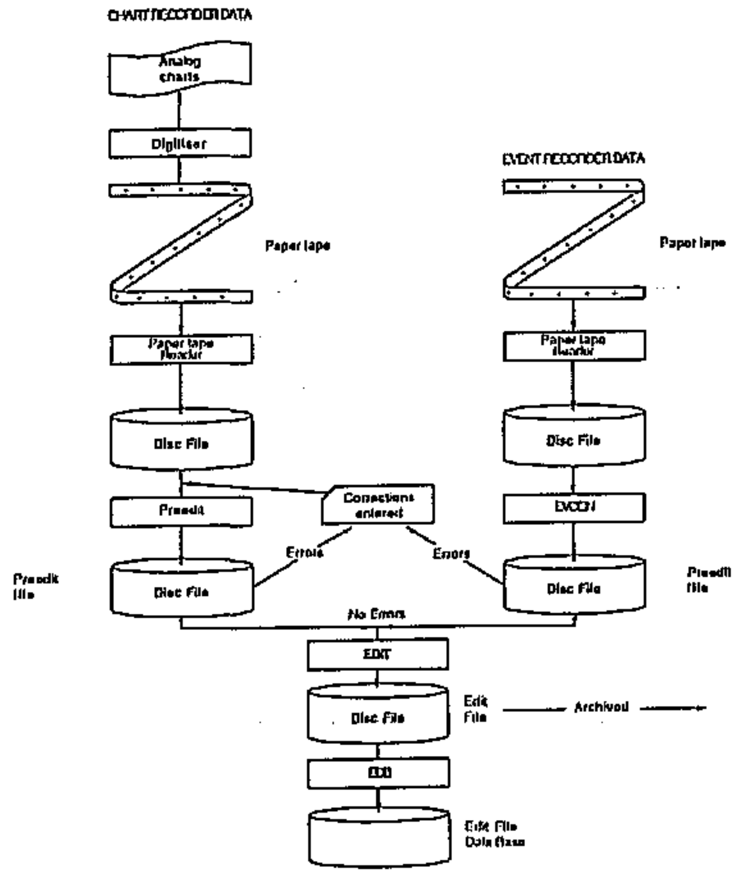


Figure 2.15(a). EDIT data base entry.

(b).

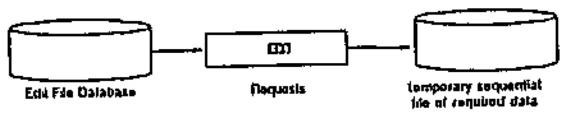


Figure 2.15(b). EDIT data base retrieval.

(c).

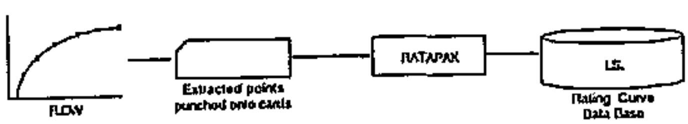


Figure 2.15(c). Rating curve data base entry.

(d).

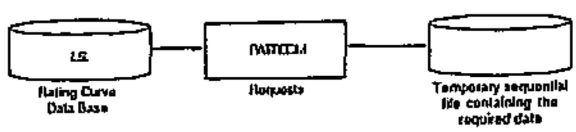


Figure 2.15(d). Rating curve data base retrieval (after Harvey, 1978).

storing the value and an identifying parameter. Sediment load data are also accommodated in this specific file structure (Harvey, 1978).

Rating curve data base - This file contains all the XY-coordinates of the stage, flow relationships for all gauging stations, in the information system. The file structure is in the form of an indexed sequential file. Data are entered by a program RATAPAK and data retrieval is done by a program RATREQM.

Each of the nearly 200 parameters that may be measured and processed by the data base system is defined in the data base dictionary. Codes used in this data base were based on the Australian standard (Harvey, 1978).

Two other file structures remain, namely:

Working format file - It forms an important link between the data from the different data sets. Similar to the non-continuous data file structure, it is very flexible, allowing the accommodation of many different types of data for further processing (Harvey, 1978).

Daily data format file - This file contains daily data in a very simple format, allowing easy data manipulation and archiving. It has a flexible structure, allowing the processing of any variable accommodated in the parameter dictionary. It also forms an important part of the data access for the users (Harvey, 1978).

All the above-mentioned data are stored on disc pack. Monthly back-up copies of the data on disc pack are made on magnetic tape.

Due to the ever-increasing demands on the water resources in this region, new parameters can be added to the water resources data bank, making this data base a constantly evolving data base.

Discussion

The Western Australian data base was implemented on a Cyber 172 mainframe computer. Adequate disk storage was provided in the form of two 216 Mb disk packs.

From the published information, it cannot be determined which data structure was used for the development of the data base, although a schematic representation of the file structures is given.

The primary data accommodated in this data base are hydrological data. The only parameters accommodated in the data base which refer to ground water, are water levels, yields and chemo-physical properties of the water. This data base is therefore not really suited for the storage of ground-water data.

Similar to the previous data base, this data base provides a facility to add new parameters, when needed. A wide range of data is stored in the form of codes.

Apart from the standard print-out facilities, this data base also provides a limited data processing facility in the form of tabular outputs, as well as several plotting routines. A further positive aspect of the data base is the built-in data quality control which ensures the use of reliable data.

Thus, although the data base accommodates very little ground-water data, it still serves as a good example, especially the water quality section thereof.

2.3.3 Borehole data base - S.W.A./Namibia

The Department of Water Affairs in South West Africa developed a simple borehole data base during the past decade. This data base was developed on a Hewlett Packard micro computer.

The data are stored on disk according to the 1:50 000 topographical maps, on which the sites occur. The data files in which these data are stored carry the name of the respective map. Data from ten boreholes are stored in four records on the data base disk. To avoid the loss of any data, back-up copies of the data base disk are made monthly. Data records are stored in three string arrays (A, B and C), a real array (S) and an integer array (I), i.e.:

Farm No.	A(X,1)	Main water	I(X,7)
Year	I(X,1)	Original water level	I(X,8)
Borehole No.	A(X,11)	Latest water level	I(X,9)
Equipment (code)	I(X,2)	Original yield	I(X,10)
Longitude	S(X,1)	Latest yield	I(X,11)
Latitude	S(X,2)	Geological information	B(X,1)
Collar height	S(X,3)	Analysis No.	C(X,1)
Depth of hole	S(X,4)	TDS	I(X,12)
Diameter	I(X,3)	Sulphate	I(X,13)
Total casing	I(X,4)	Nitrate	I(X,14)
Perforated casing	I(X,5)	Fluoride	I(X,15)
Water struck	I(X,6)		

'X' is a number between 0 and 9, denoting the borehole number in the respective record.

These arrays are thus also the basic structure of this data base. Such a structure was discussed in detail in section 2.2.1, and is generally known as a flat file structure. Special programs have been developed, allowing the user to enter, change or extract data accommodated in the data base.

Data processing programs have also been included in the data base, providing the user with various print-outs and data evaluation techniques such as map plots, pumping test analyses and hydrochemical analyses.

Discussion

This particular data base was developed in BASIC on a Hewlett Packard micro computer HP 30. The structure of the data base resembles the flat file model.

As in the case of the previous two examples, this data base accommodates only selected ground-water data. However, the parameters included, comprise a wider spectrum of important items than the previous two data bases.

Disadvantages encountered with this data base were:

Fixed space was allocated for each site. Changes in e.g. water level or yield could only be stored to a very limited degree (original and latest), otherwise older data were lost with each update.

It was developed on a micro computer with insufficient disk space, i.e. the data and programs were stored on exchangeable disks with rather limited capacity. This implied that data retrieval became less efficient as the amount of stored data increased, for instance, when data from different disks have to be extracted.

A worthwhile aspect of the data base is its data processing facility. It contains various data evaluation techniques and representation programs. These facilities served as a guide-line for the creation of the data processing programs for the National Ground-water Data Base.

2.3.4 Data base by the U.S.G.S. (WATSTORE)

The National Water Data Storage and Retrieval System forms part of four major information management programs of the United States Geological Survey (U.S.G.S.):

National Water Data and Retrieval System (WATSTORE).

National Water-use Information Program.

National Water Data Exchange (NAWDEX).

Water Resources Scientific Information Centre (WRSIC).

The data systems supporting these programs are maintained on a large mainframe computer (Amdahl) and are mainly batch-orientated (M.D. Edwards, undated). The system is presently replaced by a new National Water Information System (NWIS), which is designed as an interactive distributed data system and which will eventually run on 75 mini computers, linked by a telecommunication system, compatible with the "X.25" protocol.

WATSTORE was developed in the early 1970's to provide an efficient water resources information system. It comprises the processing, storage and extraction of surface and subsurface water resource data. The information system consists of a number of data bases, called files, in which related data are grouped and stored. Due to the systems flexible file structure, additional data files can easily be included, if necessary. The file structure is diagrammatically represented in figure 2.16.

Current data are stored on magnetic disk, while historical data are stored on magnetic tape. A brief description of the WATSTORE files is given below.

Station header file:

This file contains information about the identification, location and physical description of any site accommodated in the WATSTORE files. It also serves as an automated index, simplifying the retrieval of any of the stored data, without searching the large existing data files. Some of the information stored in this file is mandatory. This means that no data regarding this station can be entered into the data base, if any one of the mandatory items is missing.

Peak flow file:

All the streamflow and gauge height measurements are stored in this file.

Daily values file:

This file contains all the hydrological related data measured on a continuous basis. All the continuous data not in daily format are reduced to daily format before they are stored in this file. The flexible structure of the file allows many different types of data to be stored, such as streamflow, river stages, reservoir contents, water temperature, electrical conductivity, sediment concentrations, sediment discharges and ground-water levels.

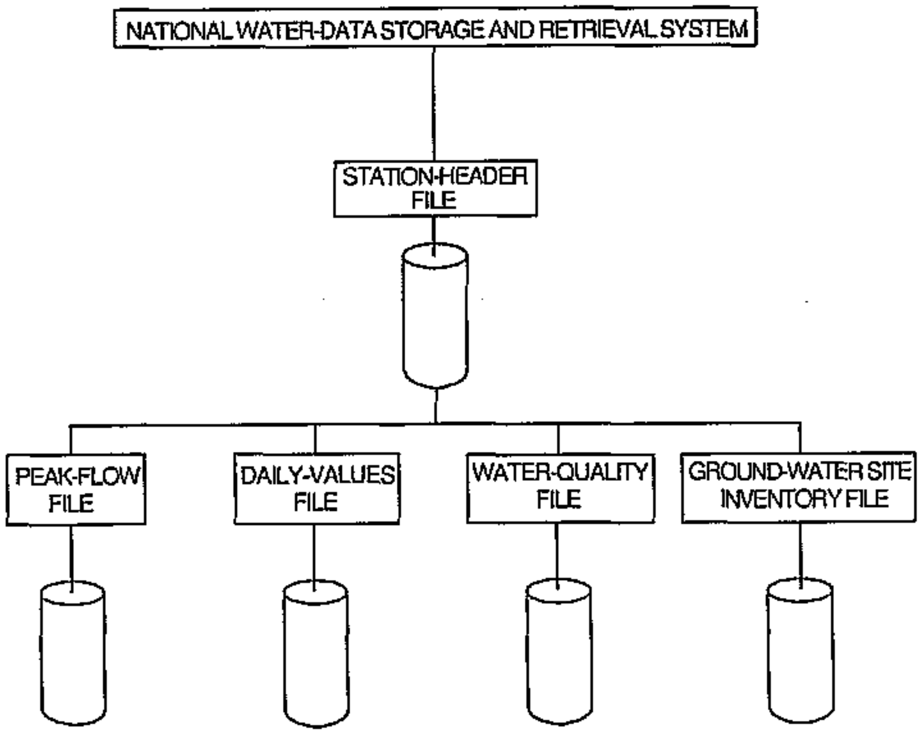


Figure 2.16. Schematic representation of the WATSTORE information system (after Showen, 1976).

Water quality file:

Chemical, physical, biological and radiochemical information about surface and subsurface water resources are stored in this file. Most of the data stored in the file are obtained through analytical techniques, performed in water quality laboratories.

Ground-water site inventory file:

The structure of this particular file will be discussed in more detail, because data contained in it, originate from wells, springs and other ground-water sources. Data items included are site location and identification, geohydrological and geological characteristics, well construction history and various field measurements, such as yield, water level and water temperature (Showen, 1976).

The Ground-water Site Inventory File (GWSI) is developed, using a data base management system known as SYSTEM 2000.

File structure

SYSTEM 2000 supports a hierarchical data base structure; the structure of the GWSI is therefore hierarchical. Figure 2.17 is a diagrammatic representation of this structure and shows all the data items accommodated by the data base. The root node contains a unique identification number and location information for each site. The SYSTEM 2000 uses various indexing techniques to keep track of the locations of data items stored randomly on a disk file. This feature simplifies the addition and retrieval of data from the GWSI (Mercer and Morgan, 1982).

In this structure, every schema record item belongs to a schema record on a higher level. Furthermore, a fundamental aspect of the schema record is that it can occur once, many times or not at all. Each schema record contains a number of mandatory items.

Uniquely identifying schema items, generally known as control fields, are normally mandatory (WATSTORE, User's Guide, Volume 2). Table 2.3, at the end of this chapter, lists a description of the GWSI schema records.

Data that have been coded must be checked for reliability, after which they can be entered into the data base. A specially developed program checks the entered data to see whether they satisfy required edit criteria. Three basic checks are done on the newly entered data, namely:

- (i) Syntax checking.
- (ii) Compatibility checking.
- (iii) Out-of-range checking.

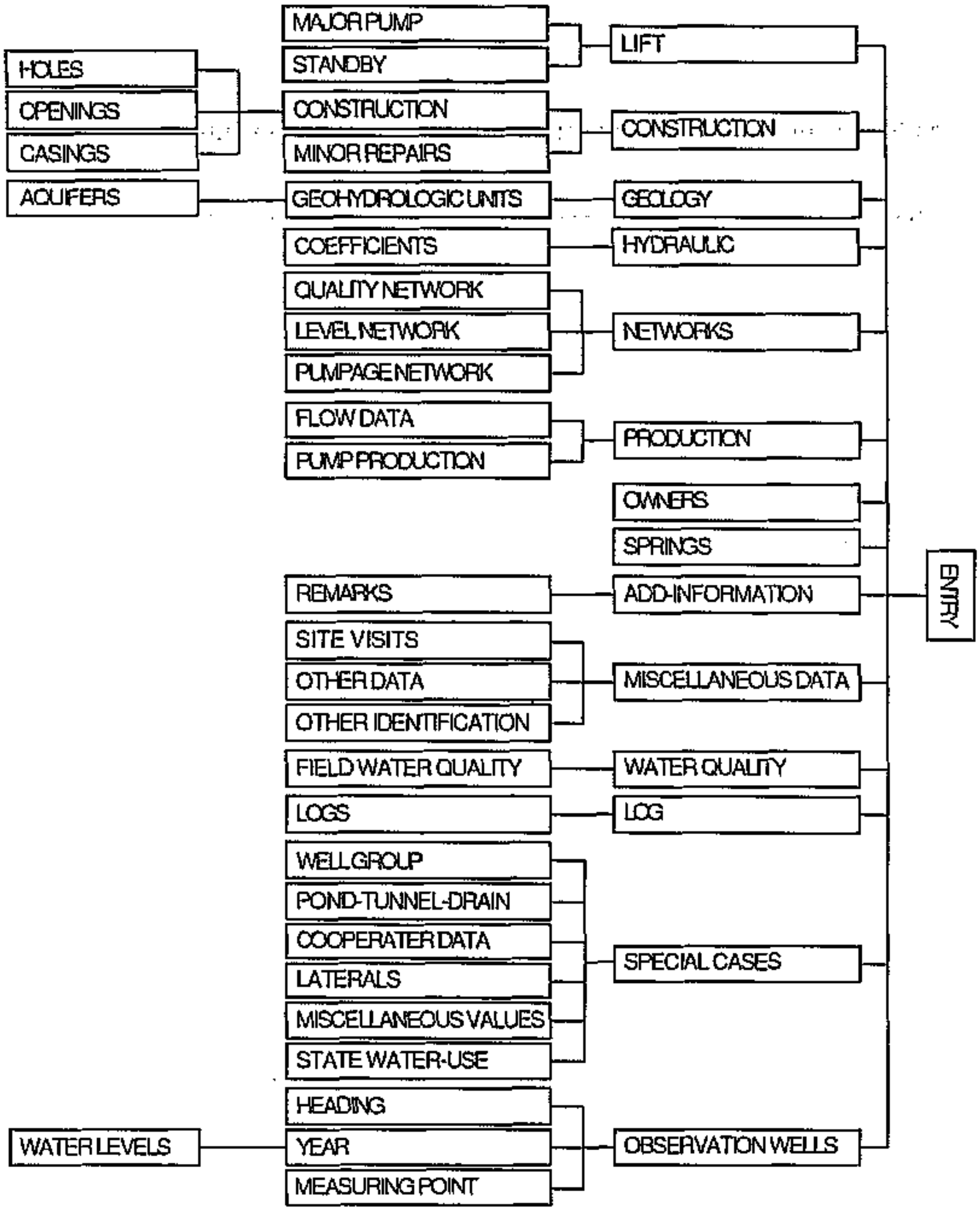


Figure 2.17. Data base structure (after Mercer & Morgan, 1982).

The update process is under complete control of the data base administrator (WATSTORE, User's Guide).

Data storage and retrieval

The data base management system offers a wide variety of capabilities for creating data bases, updating data (add, change, delete), data retrieval, data reporting and interfacing with other computer programs. Important attributes of computer systems are to accommodate large quantities of data, to rapidly search through these data and, when required, process it.

The data base management system is designed quite flexible, offering the user effective data retrieval and reporting capabilities (WATSTORE, User's Guide).

Data storage in the data base

Each component within the data base has a unique identification code which distinguishes it from the other components in the data base and identifies all the components present in the system. The amount of data stored in the GWSI is independent of the data base definition, meaning that the number of boreholes stored in the data structure are independent of the data base definition. Data items for a specific site are related and are kept from the data from another well, forming separate data sets (WATSTORE, User's Guide).

Data retrieval from the data base

The system supports a great variety of data retrieval commands and depends for efficiency almost entirely on the initial data base set up. Due to the existence of inverted files or indexes in GWSI structure, extraction of specific sites is very efficient. Not all the data items are indexed, thus subdividing data retrieval into two components, i.e. key and non-key components. When a data item is designated as a key, an index is created for that data item and it can be used to retrieve data fast. If a data item is designed as a non-key, no index is created and using it to retrieve data will probably be much slower. The designation of keys to data items is done by the data base manager (WATSTORE, User's Guide).

Data can be retrieved selectively by the retrieval language which forms a part of the data base management system. Programs, extracting selected data and prepare them for publication, plotting or statistical evaluation, have been written.

Discussion

The GWSI was designed and implemented in 1974, on a mainframe computer, using the SYSTEM 2000 data base management system. The main objectives were (i) to meet the need for storage of standardized ground-water data and (ii) to provide nation-wide computer access to the data.

The structure of a SYSTEM 2000 data base is usually hierarchical, while various indexing techniques are used to keep track of the locations of data, randomly stored on disk files. The indexing feature used in the data base management system simplifies the addition of new data and makes data retrievals more efficient. This advantage must be seen against the disadvantage of redundancy and increased time requirement for the ADD, CHANGE AND DELETE operations which must update the respective key lists.

A wide variety of geohydrological related data is accommodated in this data base (see figure 2.17). Most data in the GWSI are raw ground-water data and are collected and prepared for submission to the data base by the participating organizations. These data are then forwarded to the GWSI data base administrator, who controls the final update.

An important aspect of the data base is the inclusion of mandatory data items. This ensures that the data entered, comply with a certain minimum standard. In addition to that, a further quality control is implemented, which checks the reliability of the data. Special computer programs were developed for checking the syntax, compatibility and whether the data are out-of-range.

Due to the fact that the developers used an established data base management system, data retrieval is relatively simple. A great variety of data retrieval commands exist, which can be enhanced by an additional programming language extension. Due to the use of indexing, the data retrieval becomes very fast.

The GWSI includes a number of processing facilities in the form of tabular reports, X-Y plots and map plots for many of the data. However, the most important feature of the GWSI is the fact that data input is standardized and the techniques of manipulating the ground-water data are the same throughout the United States of America.

Problems inherent in batch-orientated and centralized mainframe computers on the one hand and improved computer technology on the other, led to the decision to switch to an interactive, decentralized minicomputer system (the file management system MIDAS will be used on the new Prime computers). A significant amount of commercially available software will be incorporated in the new system and attention will be given to "menus" and "HELP" functions for less experienced users.

Due to the GWSI's comprehensiveness, it was used as a basis to develop the National Ground-water Data Base for Southern Africa. The development of this data base will be discussed in the next chapter.

TABLE 2.3. DESCRIPTION OF GWSI SCHEMA RECORDS (AFTER MERCER AND MORGAN, 1982).

ENTRY	Site identifiers such as latitude and longitude,altitude, State, county and so forth.
LIFT	Type, such as pump or bucket; includes horsepower, intake setting and so forth.
MAJOR PUMP	Manufacturer, serial number, energy consumption, capacity and so forth.
STANDBY	Alternative power types.
CONSTRUCTION	
CONSTRUCTION	Date of completion, contractor, seal type, finish and seal bottom.
HOLES	Type of well and dimensions, including diameter of top and bottom of hole.
OPENINGS	Depth intervals of perforated zones, size and shape and screen material.
CASINGS	Type and material, top in reference to land surface, depth to bottom and diameter of each string.
MINOR REPAIRS	Repair information.
GEOLOGY	
GEOHYDROLOGIC UNITS	Name of formation, including unit identifier and its depth.
AQUIFERS	Includes static water level in aquifer.
HYDRAULIC	Includes the unit identifier.
COEFFICIENTS	Includes conductivity, diffusivity and leakance.
NETWORKS	
QUALITY NETWORK	Water-quality network, including name of agency that gathers samples at site.
LEVEL NETWORK	Water-level network, including name of agency that collects water-level measurements at site.
PUMPAGE NETWORK	Pumpage network, including name of agency that monitors water withdrawal at site.
PRODUCTION	
FLOW DATA	Information about springs, including flow period and discharge.
PUMP PRODUCTION	Production of the well, including production date and method.
OWNERS	The site's owner, name and ownership date.
SPRINGS	Spring data; for example, name and number of openings.
ADDITIONAL INFORMATION	
REMARKS	Additional remarks about the site.

MISCELLANEOUS DATA	Other references and sources of data.
SITE VISITS	Visits to the site, such as the inventory person and date of visit.
OTHER DATA	Location and formats of other data available about the site.
OTHER IDENTIFICATION	Other site identifiers.
WATER QUALITY	
FIELD WATER QUALITY	Field water-quality data, such as the sample date, the constituent, its measurement and source (aquifer name).
LOG	
LOGS	Type of geophysical or other logs available for the well, including type and source.
SPECIAL CASES	
WELL GROUP	Multiple wells that are manifolded to a single discharge pipe, including the number of wells and the deepest and shallowest wells in the group.
POND, TUNNEL, DRAIN	The length, width and depth of a pond, tunnel or drain.
COOPERATOR DATA	Data that cooperating local agencies need, such as cooperator's site identifier, registration number and so forth.
LATERALS	Information about Ranney wells, including the depth, length and diameter of the laterals that drain to the central well.
MISC.VALUES	Data for which no other schema record has been established.
STATE WATER USE	The State's use of water, including water type and the amount of water.
OBSERVATION WELLS	
HEADING	Textual information about the site.
YEAR	Specific year for data in the lower level schema record(s).
WATER LEVELS	Includes water-level measurements and respective dates.
MEASURING POINT	Includes the measuring point height and the date when the measurement was made.

CHAPTER 3

NATIONAL GROUND-WATER DATA BASE FOR SOUTHERN AFRICA

3.1 INTRODUCTION

Rainfall over Southern Africa varies from more than 1 000 mm/a in the eastern parts of the country to less than 200 mm in the west (*Management of the Water Resources of South Africa, Department of Water Affairs*), while the potential evaporation varies from a minimum of less than 1 400 mm along the Natal South Coast to over 2 400 mm in Gordonia and parts of Bushmanland. The South African subcontinent is therefore characterized by predominantly semi-arid climatic conditions with periodical droughts and has limited water resources. Accordingly, it is highly important that the existing surface and subsurface water resources are managed in the best possible manner. This task can considerably be facilitated if all hydrological related data are stored (in various data bases) on computer. In order to expedite the process of data computerization, the Water Research Commission has funded the development of such data bases which, when completed, will form a central hydrological information system (HIS).

HIS will consist of eight data bases which will contain data on the following topics, i.e. precipitation, flow, reservoirs, evaporation, sediment load, floods, chemical quality and ground water. The purpose of this chapter is to discuss the development of the ground-water data base for Southern Africa.

During the latter part of 1983, the Institute for Ground-water Studies (I.G.S.) of the University of the Orange Free State compiled a draft plan for the development of such a ground-water data base and submitted it to the Water Research Commission. After discussion of the draft version with the Directorate of Geohydrology of the Department of Water Affairs, the Water Research Commission awarded a four-year research contract to the Institute for Ground-water Studies. The contract comprised the development of a National Ground-water Data Base for Southern Africa and started in January 1984.

The objectives of the contract were:

(i) A detailed analysis of the data base requirements.

(ii) The design of an appropriate and effective data base structure.

(iii) The implementation of the structure on a mainframe computer, including data update and extraction facilities.

(iv) The development of an efficient data processing package and software enabling communication with the chemical data base.

(v) Assisting the Directorate of Geohydrology in the selection of appropriate computer hardware.

(vi) Training personnel of the Department of Water Affairs in the effective utilization of the data base.

The ultimate aim of the data base is to provide information on physical, geological and geohydrological aspects of any known ground-water source in Southern Africa.

3.2 SOFTWARE FOR MAINFRAME DATA BASE DEVELOPMENT

Development of a data base is a complex and comprehensive task. Various aspects have to be taken into account, like the data to be stored, the users' needs, the available hardware and programming software.

As the data base was to be developed on a mainframe computer, for which a special data base management system (DMSII) was available, it was important to ensure compatibility between the different data bases of the hydrological information system. Generally, all data base management systems consist of a set of integrated data structures or files which reside on some mass storage medium and are manipulated and managed by various specialized programs. There are therefore a number of advantages when using a data base management system instead of a set of discreet files. DMSII offers the following advantages:

(i) Data and program independence.
(ii) Ease of programming.
(iii) Wide choice of host languages.
(iv) Data consistency.
(v) Concurrent updating.
(vi) Data protection.

(vii) Data base development and evolution.
(viii) Economic system resources.
(ix) Ease of system design.

Another advantage of DMSII is that it supports LINC (Logic and Information Network Compiler). LINC is used to design and develop data bases. It consists of a definition language and a generator which produce programs for the data base. Such compilers are generally known as fourth generation system development compilers and are used to generate third generation computer codes, such as COBOL, automatically.

Three basic parts, i.e. components, events and profiles, form the LINC compiler system. Components are used to define fixed data elements. Events are the transactions that occur with the data and profiles are used to specify ways in which components and events of any system are to be combined and viewed.

The LINC compiler system is structured around components, events and profiles. The purpose of the LINC specification is to describe the various elements of the specific information system to the LINC compiler. The LINC specification is divided into various logical groups. The following list indicates the functions included in the LINC specifications and the steps to be taken when developing a data base with its aid.

1. Option specification

(i) Create a DMSII data base and a new system.
(ii) Reorganize/rebuild a DMSII data base originally created by LINC.
(iii) Create a test-version data base allowing an enhanced system to be developed separately from a production system.
(iv) Specify decimal entry on numeric fields.
(v) Specify log of transactions.
(vi) Generate listing of the specification.
(vii) Establish or rebuild a network.
(viii) Facilitate off-line input.
(ix) Specify Message Control System (MCS).
(x) Allow Generalize LINC Interface facility - allow LINC to be updated, when necessary.
(xi) Allow American date formats to be used.

2. User specification.

(i) Specify user identifier, data base name, system name.

(ii) Declare number of accounting periods.
(iii) Define data base distribution over the disk packs.

3. *Network specification.*

(i) Declare network configuration (line, channel, terminal, address, stations, etc.).
(ii) Define security privilege level.
(iii) Specify executable programs.

4. *Logical data base specification.*

(i) Declare logical data base for DMSII Inquiry.

5. *Profile specification.*

(i) Declare access paths to the data base.

6. *Global specification.*

(i) Define global logic.
(ii) Define global set-up data items.
(iii) Define keywords.

7. *Interface specification.*

(i) Specify screen format identification.
(ii) Specify automatic return of screen format after processing screen input.
(iii) Define help/teach functions.
(iv) Assign file security privileges.
(v) Specify repeat lines.
(vi) Identify source of data.
(vii) Define system volumes.

8. *Data specification.*

(i) Define screen formatting.
(ii) Describe data.
(iii) Define field definitions (type, size, decimals, data, justification, highlighting, required, etc.).
(iv) Define field usage (input only, input-output, output only).

(v) Define automatic data base access on ordinates.
(vi) Specify pre-screen logic.
(vii) Insert global logic.

9. Logic specification.

(i) Specify arithmetic functions.
(ii) Define user-controlled access and update of data base.
(iii) Specify logic commands.
(iv) Insert global logic.
(v) Define error and message handling.
(vi) Specify data editing and verification.

After selection of the data base management system to be used for the development of the ground-water data base, the information included in section 2.2 can be used to design an efficient data base structure.

3.3 STRUCTURE OF THE NATIONAL GROUND-WATER DATA BASE

With DMSII, it is possible to use any of the three basic data structures, namely the hierarchical model, the network model and the relational model. All these data structures have the advantage that related data are grouped together, which makes data manipulation simple and straightforward.

Certain data suit certain data structures better than others. In order to select the most appropriate data structure for the data base, a number of aspects listed below, had to be considered.

(i) Who will be using the data base and what type of enquiries must the data base be able to handle?

(ii) The approximate number of possible data sites, in order to allocate enough storage capacity for the data base.

(iii) The data base should be point-orientated and include data originating from boreholes, fountains and/or the other ground-water observation points.

(iv) The data base must primarily suit the geohydrologists, i.e. the Directorate of Geohydrology's requirements and, to a certain extent, the remaining sections of the Department of Water Affairs.

(v) The majority of sites contain only a few basic data items. Therefore, the structure of the data base must be such, that it can incorporate sparse, as well as detailed data, in such a way that data redundancy, storage requirements and search time are minimized.

Complying with the above-mentioned requirements, the relevant data were divided into two major groups:

(i) Basic data of a site.

(ii) Detailed additional data of a site.

All data were grouped into 11 data groups, comprising the following items:

(i) ENTRY - The basic information about a site.

(ii) LOGS - Data collected during making of a site.

(iii) CONSTRUCTION - The data related to the civil and mechanical construction of a site.

(iv) INSTALLATION - The data related to mechanical installations made on a site.

(v) DISCHARGE RATE - The data related to the yield of a site.

(vi) WATER LEVEL - The data related to the water level of a site.

(vii) WATER SAMPLE - Analysis number (H-number) referring to the analysis in the chemical data base.

(viii) FIELD MEASUREMENT - The data related to various parameters measured in the field at a site.

(ix) METER READING - The data originating from any kind of meter installed at a site.

(x) PUMPING TEST - The data derived from a pumping test on a site.

(xi) MISCELLANEOUS - The data related to miscellaneous aspects of the site.

Of these groups, ENTRY, INSTALLATION and PUMPING TEST are not subdivided into subgroups. The remaining eight groups, however, are subdivided into header and repeating groups, with the repeating groups resorting under their respective headers.

Figure 3.1 is the schematic representation of this structure. The remaining eight groups, divided into their headers and respective repeating groups, are listed below:

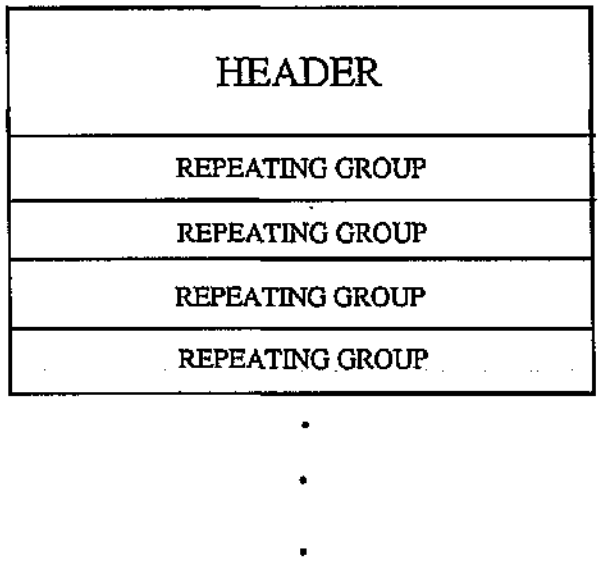


Figure 3.1. Data set structure.

(i) Logs
- (a) Penetration rate
- (b) Aquifer
- (c) Unconsolidated geology (especially for detailed soft rock description)
- (d) Consolidated geology

(ii) Construction
- (a) Hole
- (b) Casing
- (c) Openings
- (d) Fill

(iii) Discharge rate
- (a) Discharge rate

(iv) Water level
- (a) Water level

(v) Water sample
- (a) Water sample

(vi) Field measurement
- (a) Field measurement

(vii) Meter reading
- (a) Meter reading

(viii) Miscellaneous
- (a) Site selection
- (b) Name of owner
- (c) Other identifier
- (d) Other data
- (e) Site visits
- (f) Special cases
- (g) References
- (h) Search status
- (i) Comments

Thus, taking all the data groups into account, the data are now grouped into 32 data sets.

An important aspect of the data groups is that all the basic information of a specific site is contained in the ENTRY data set. Only one ENTRY data set is allowed for each site to avoid duplication. This is, of course, logical, as one site can e.g. only have one pair of coordinates with a specific altitude. Both headers and repeating groups may be repeated. Headers contain general information which applies to all of the up to 999 repeating groups, which can be stored under each of them.

Although the INSTALLATION and PUMPING TEST data sets are not subdivided into header and repeating groups, they may be repeated, e.g. more than one pumping test can be performed on a specific borehole site over a number of years and more than one pump be installed.

After evaluation of all the above-mentioned information, the data base structure was defined, as shown in figure 3.2(a). This data structure has three levels, namely basic information, header and respective repeating groups. It can be easily converted into a simple tree structure (see figure 3.2(b), because the relationships among the data sets are generally limited to a one-to-many relationship.

The data could also easily have been accommodated in a relational data model, which is the simpler model of the three existing models and thus the preferable choice. However, due to the relational model's positional independence, as well as its non-procedural access path, some of the data items, under consideration, would have to be stored repeatedly. This would have led to unnecessary data redundancy, even though the redundancy would have been controlled.

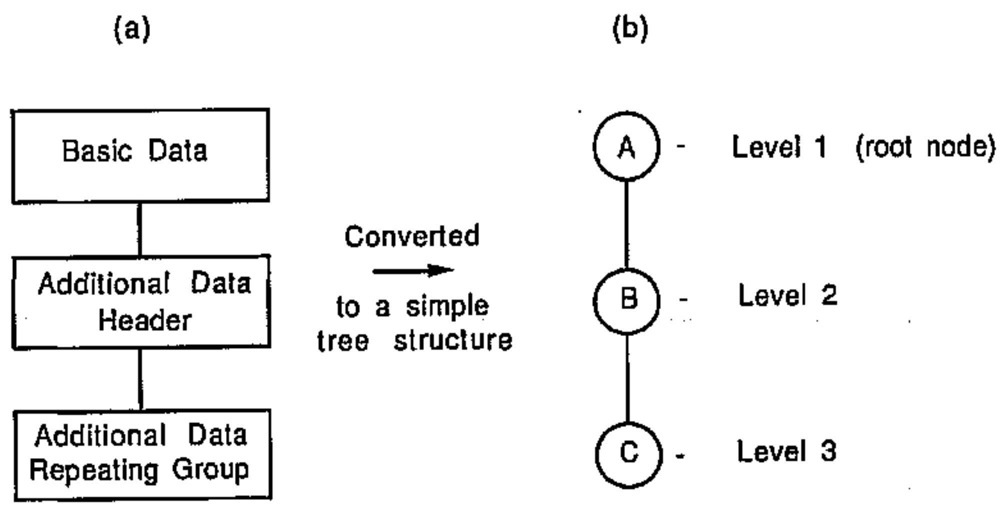


Figure 3.2(a). Basic data structure and (b) basic data structure converted into a simple tree structure.

The network data model is unsuitable for the type of data considered here, because of its complex structure. The complexities in the structure are caused by the network model's capability to handle many-to-many relationships, in addition to the simpler relationships (see section 2.2.4). This, again, leads to unnecessary data redundancy. However, due to the models' ability to handle many-to-many relationships, data access may, in some cases, be more efficient.

Thus, the logical choice of a data structure to be used for the data base, was the tree (hierarchical) data structure. Expanding the basic tree structure of figure 3.2(b) to the tree structure containing all the (32) data sets, figure 3.3 displays the structure with all the data sets contained in the new data base.

An apparent disadvantage of the tree structure is that every node, except the root node, has to be accessed through its parent node. This makes the access path to a specific node unique. However, DMSII allows to combine certain network features with features of the tree structure. Data access in the ground-water data base could therefore be made more efficient. A direct implication of this combination is that the user can access all the

(a)

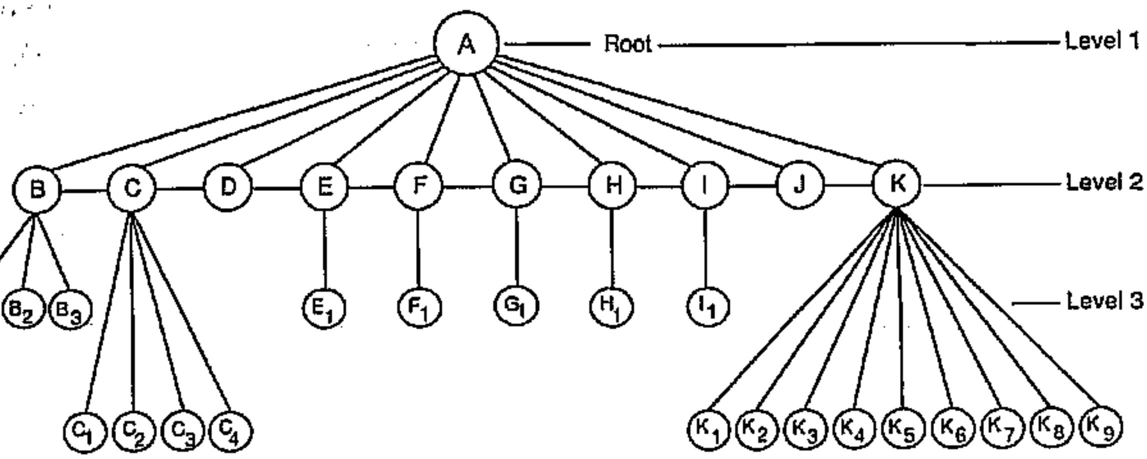


Figure 3.3 (a). Basic tree structure of the ground-water data base.

(b)

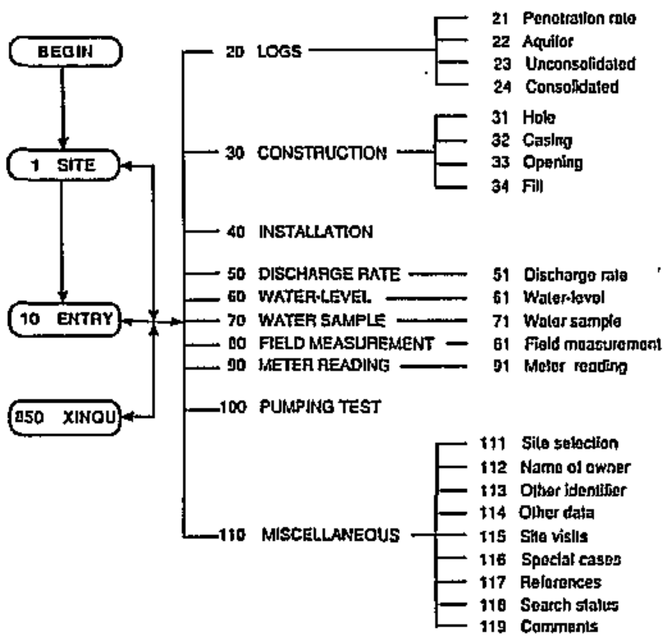


Figure 3.3 (b). Ground-water data base structure, indicating the respective positions of each data set in the tree structure.

data sets on the same level by entering the respective data set code. In the case of the ground-water data base, the network aspects cause a certain data redundancy. This becomes negligible if one considers the amount of physical storage capacity saved by the tree structure's low level of data redundancy.

3.4 DATA BASE PARAMETERS

Once the data base has been designed and created, complying with the predefined requirements, specific parameters (values/data items) may be entered into the data base. All the parameters accommodated in the data base are listed in figure 3.4.

A data base depends entirely on the type and quality of data added to it. In order to promote good quality of the data in the data base, certain parameters in each data set were declared mandatory. The mandatory items of the data set ENTRY are:

(i) Site ID Code
(ii) Map Code
(iii) Latitude
(iv) Longitude
(v) Coordinate Accuracy
(vi) Drainage Region
(vii) Site Type
(viii) Reporting Institution

The mandatory items about a site are the absolute minimum data that can be entered for any site. Mandatory items in the other data sets are printed bold in figure 3.4.

Much consideration was given to the parameters proposed for inclusion in the data base as well as their format. This was followed by a thorough discussion with the Department of Water Affairs to ensure that the selected parameters complied with the requirements as seen by the Directorate of Geohydrology For a detailed description of each parameter included in the data base, chapter 3 of the User's Guide can be consulted.

A considerable number of the parameters accommodated in the data base, are represented by standardized codes, of which the majority are one- or two letter codes as can be seen in figure 3.4. The remaining codes, namely:

(i) district codes,

10 ENTRY

* **1.0.1 Site identification code**
1.0.2 Map code
1.0.3. Number on 1 : 50 000 map
1.0.4 Province, district and farm number
1.0.5 Site name
1.0.6 Topographic setting
  A Alluvial fan
  D Dunes
  F Flat surface, plain
  H In or along sinkhole
  L Along dam, lake or swamp
  M On mountain or hill
  P In or along pan
  R In or along river
  S Hillside (slope)
  T Terrace
  V Valley
* **1.0.7 Latitude**
* **1.0.8 Longitude**
1.0.9 Coordinate accuracy
* 1.0.10 Drainage region
1.0.11 Altitude
1.0.12 Method of measurement
  A Altimeter
  L Levelled or surveyed
  M Interpolated from map
1.0.13 Site type
  B Borehole
  D Dug well
  F Fountain
  H Sinkhole
  I Interconnected or drainage well
  M Multiple boreholes
  S Seepage pond
  T Tunnel, shaft or drain
  W Well point
  Z Other
1.0.14 Site selector
  B-I Boring Inspector
  DIV Water diviner
  DRL Driller
  DWA Geohydrology, DWA
  GEO Geologist or Geophysicist (Consultant etc.)
  GSO Geological Survey
  OWN Owner
1.0.15 Nominal diameter
1.0.16 Depth completed
1.0.17 Completion date
1.0.18 Depth information source
  C Construction data
  D Driller's logs
  F Field checked by Reporting Institution
  G Geologist, technician, operator's record
  M Memory (Owner, Driller, Operator)
  O Owner's record
  R Report or file
  P Pump operators record
1.0.19 Status of site
  D Destroyed
  G In use
  U Unused
1.0.20 Purpose of site
  E Exploration or stratigraphic test hole
  M Mine drainage
  O Observation
  P Production (water supply)
  R Recharge
  S Standby
  W Waste disposal
  Z Other
1.0.21 Water use: consumer
  U Urban
  N Non-urban
1.0.22 Water use: application
  AD Agricultural and domestic use
  AI Agricultural - Irrigation only
  AS Agricultural - stock watering only
  DA Domestic - all purposes
  DG Domestic - garden only
  E Nature conservation
  P Public
  TC Industrial - commercial
  TI Industrial - industrial
  TM Industrial - mining
  TP Industrial - power generation
1.0.23 Reporting Institution
  D&M Dames & Moore
  GSP Geological Survey, Pretoria
  IGS Institute for Ground-water Studies
  NPRL NPRL of CSIR
  SRK Steffen, Robertson & Kirsten
  WA-B Boring, Water Affairs
  WA-G Geohydrology, Water Affairs
  WAGC Geohydrology, Water Affairs, Cape Town
  WA-P Permit Section, Water Affairs
1.0.24 Equipment
  A Airlift
  C Centrifugal pump
  H Hand pump
  J Jet
  M Mono-type pump
  N No equipment
  P Piston pump
  Q Powerhead
  R Recorder
  S Submersible pump
  T Turbine
  W Windpump
  X Windpump with powerhead
  Z Other
1.0.25 Potability
  V Very good
  G Good for human consumption
  M Marginal for human consumption
  A For animals only
  U Unsuitable for human or animal consumption
  F Fresh
  B Brack
  S Salty
1.0.26 Date record entered
1.0.27 Date record updated
1.0.28 Comment

20 LOGS

2.0.1 Site identification code
2.0.2 Header sequence number
2.0.3 Reporting Institution
2.0.4 Date record entered

2.1 REPEATING GROUP: PENETRATION RATE
2.1.1 Site identification code
2.1.2 Header sequence number
2.1.3 Repeating group number
2.1.4 Depth to top of interval
2.1.5 Depth to bottom of interval
2.1.6 Diameter of interval
2.1.7 Penetration rate
2.1.8 Comment

2.2 REPEATING GROUP: AQUIFER
2.2.1 Site identification code
2.2.2 Header sequence number
2.2.3 Repeating group number
2.2.4 Depth to top of aquifer
2.2.5 Depth to bottom of aquifer
2.2.6 Aquifer code
2.2.7 Comment

2.3 REPEATING GROUP: UNCONSOLIDATED
2.3.1 Site identification code
2.3.2 Header sequence number
2.3.3 Repeating group number
2.3.4 Depth to top of interval
2.3.5 Lithology code
2.3.6 Primary colour
  S Black
  B Blue
  C Brown
  G Green
  H Grey
  M Purple
  O Orange
  P Pink
  R Red
  W White
  Y Yellow
2.3.7 Secondary colour
  B Bluish
  C Brownish
  D Dark
  G Greyish
  L Light
  M Purple
  O Orange
  P Pinkish
  R Reddish
  V Olive
  Y Yellowish
2.3.8 Texture

00	Crypto	33	Medium
11	Very fine	34	Medium to coarse
12	Very fine to fine	35	Medium to very coarse
13	Very fine to medium	41	Coarse and very fine
14	Very fine to coarse	42	Coarse and fine
15	Very fine to very coarse	43	Coarse and medium

Figure 3.4. Parameters of the ground-water data base.

21 Fine and very fine
22 Fine
23 Fine to medium
24 Fine to coarse
25 Fine to very coarse
31 Medium and very fine
32 Medium and fine
44 Coarse
45 Coarse to very coarse
51 Very coarse and very fine
52 Very coarse and fine
53 Very coarse and medium
54 Very coarse and coarse
55 Very coarse

2.3.9 Primary feature
AG Argillaceous
AR Arenaceous
BK Baked
CA Calcareous
CB Carbonaceous
CE Cemented
CS Consolidated
FC Fractured
FE Ferrugenous
FS Feldspatic
GV Gravel-bearing
HD Hard
HM Heavy minerals bearing
JT Jointed
MC Micaceous
MS Massive
OO Oolitic
PO Phosphorite bearing
PT Peaty
SF Soft
SH Shelly
SI Silicified
SL Silty
UC Unconsolidated
WT Weathered

2.3.10 Secondary feature
2.3.11 Feature attribute
- Slightly
+ Very

2.3.12 Sorting
11 Unsorted
22 Poorly sorted
23 Poorly to moderately sorted
32 Moderately to poorly sorted
33 Moderately sorted
34 Moderately to well sorted
43 Well to moderately sorted
44 Well sorted

2.3.13 Roundness
11 Angular
12 Angular to subangular
13 Angular to subrounde
14 Angular to rounded
22 Subangular
23 Subangular to subrounded
24 Subangular to rounded
33 Subrounded
34 Subrounded to rounded
44 Rounded

2.3.14. - 20 Sieve analysis
2.3.21 Comment

2.4 REPEATING GROUP: CONSOLIDATED
2.4.1 **Site identification code**
2.4.2 Header sequence number
2.4.3 Repeating group number
2.4.4 **Depth to top of interval**
2.4.5 **Lithology code**
2.4.6 Primary colour
2.4.7 Secondary colour
2.4.8 Texture
2.4.9 Feature
2.4.10 Comment

30 CONSTRUCTION

3.0.1 **Site identification code**
3.0.2 Header sequence number
3.0.3 Piezometer number
3.0.4 **Reporting institution**
3.0.5 Date record entered
3.0.6 **Date of construction**
3.0.7 Name of contractor
3.0.8 Construction data source
3.0.9 Method of construction
A Air-rotary
C Cable-tool
D Dug
H Hydraulic rotary
J Jetted
P Air percussion
R Reverse rotary
Z Other

3.0.10 Type of finish
B Brickwork
C Porous concrete
F Filter (gravel pack with perforations)
G Gravel pack with screen
H Horizontal gallery
O Open bottom (partially cased)
P Perforated or slotted
S Screen
W Well point
X Open hole
Z Other

3.0.11 Method of development
A Pumped with air lift (no inductor)
B Bailed
C Compressed air
I Air-lifted with inductor
J Jetted or washed
N None
P Pumped
S Surged

3.0.12 Duration of development
3.0.13 Special development treatment
B Brushing
C Chemical (acid etc.)
D Dry ice
E Explosives
H Hydrofracturing
Z Other

3.0.14 Cost of construction
3.0.15 Comment

3.1 REPEATING GROUP: HOLE
3.1.1. **Site identification code**
3.1.2. Header sequence number
3.1.3 Repeating group number
3.1.4 **Depth to top of interval**
3.1.5 **Depth to bottom of interval**
3.1.6 **Diameter of hole**
3.1.7 Comment

3.2 REPEATING GROUP: CASING
3.2.1 **Site identification code**
3.2.2 Header sequence number
3.2.3 Repeating group number
3.2.4 **Depth to top of casing interval**
3.2.5 **Depth to bottom of casing interval**
3.2.6 **Diameter of casing interval**
3.2.7 Casing material
B Brass
C Concrete
D Copper
P PVC or other plastic
S Steel
X Stainless steel
Z Other material

3.2.8 Wall thickness of casing
3.2.9 Comment

3.3 REPEATING GROUP: OPENINGS
3.3.1 **Site identification code**
3.3.2 Header sequence number
3.3.3 Repeating group number
3.3.4 **Depth to top of open interval**
3.3.5 **Depth to bottom of interval**
3.3.6 **Diameter of open interval**
3.3.7 Material in this interval
3.3.8 **Type of openings in interval**
F Perforated or slotted with fibre mesh
L Perforated or slotted
M Mesh screen
P Perforated or slotted
S Screen
W Wellpoint
X Open hole
Z Other

3.3.9 Length of openings
3.3.10 Width of openings
3.3.11 Horizontal distance between openings
3.3.12 Vertical distance between openings
3.3.13 Method of openings made
A Sawn
D Drilled
M Machine cut
P Punched
S Screen
W Well point
X Wire wound
Z Other

3.3.14 Comment

3.4 REPEATING GROUP: FILL
3.4.1 **Site identification code**
3.4.2 Header sequence number
3.4.3 Repeating group number
3.4.4 **Depth to top of interval**
3.4.5 **Depth to bottom of interval**
3.4.6 **Type of fill**
B Bentonite or clay
C Cement
G Gravel (>2mm)
S Sand (≤2mm)
X Bottom closed

3.4.7 Comment

Figure 3.4. Continued.

40 INSTALLATION

4.0.1 **Site identification code**
4.0.2 Header sequence number
4.0.3 **Reporting institution**
4.0.4 Date record entered
4.0.5 **Date of installation**
4.0.6 **Type of installation**
4.0.7 Depth to pump intake
4.0.8 Type of power
- D Diesel engine
- E Electric motor
- H Hand
- W Windpump
- Z Other

4.0.9 Power rating
4.0.10 Manufactor
4.0.11 Serial number
4.0.12 Power meter number
4.0.13 Monitoring facility
- A Airline
- I Piezometer Inside casing
- O Piezometer Outside casing
- T Pressure Transducer

4.0.14 Installation data source
4.0.15 Comment

5.0 DISCHARGE RATE

5.0.1 **Site identification code**
5.0.2 Header sequence number
5.0.3 **Reporting institution**
5.0.4 Date record entered
5.0.5 **Type of discharge**
- F Flow
- P Pump

5.0.6 Method discharge measured
- C Current meter
- E Estimated
- F Flume
- M Totalling meter
- N Notch (V- or U-notch)
- S Submerged orifice
- V Volumetric measurement (container + stop-watch method
- U Venturi meter
- W Weir
- Z Other

5.0.7 Discharge information source

5.1 REPEATING GROUP: DISCHARGE RATE
5.1.1 **Site identification code**
5.1.2 Header sequence number
5.1.3 Repeating group number
5.1.4 **Date of measurement**
5.1.5 Time of measurement
5.1.6 **Discharge rate**
5.1.7 Comment

6.0 WATER-LEVEL

6.0.1 **Site identification code**
6.0.2 Header sequence number
6.0.3 Piezometer number
6.0.4 **Reporting institution**
6.0.5 Date record entered
6.0.6 Method of measurement
- A Airline
- E Electrical contact meter
- P Pressure gauge measurement
- R Recorder
- S Steel tape
- U Reported, unknown
- X Estimate

6.0.7 Water-level status
- A Water-level affected by nearby pumped hole or drilling
- D Dry
- F Flowing
- O Obstruction, no water-level measured
- P Pumped water-level
- R Recovering water-level
- S Static water-level

6.0.8 Collar elevation

6.1 REPEATING GROUP: WATER-LEVEL
6.1.1 **Site identification code**
6.1.2 Header sequence number
6.1.3 Repeating group number
6.1.4 **Date of measurement**
6.1.5 Time of measurement
6.1.6 **Water-level**
6.1.7 Comment

70 WATER SAMPLE

7.0.1 **Site identification code**
7.0.2 Header sequence number
7.0.3 **Reporting institution**
7.0.4 Date record entered

7.1 REPEATING GROUP: WATER SAMPLE
7.1.1 **Site identification code**
7.1.2 Header sequence number
7.1.3 Repeating group number
7.1.4 Date of sampling
7.1.5 Time of sampling
7.1.6 Sample number
7.1.7 Comment

80 FIELD MEASUREMENT

8.0.1 **Site identification code**
8.0.2 Header sequence number
8.0.3 **Reporting institution**
8.0.4 Date record entered

8.1 REPEATING GROUP: FIELD MEASUREMENT
8.1.1 **Site identification code**
8.1.2 Header sequence number
8.1.3 Repeating group number
8.1.4 **Date of measurement**
8.1.5 Time of measurement
8.1.6 **Parameter measured**
- C Conductivity (mS/m)
- P pH
- T Temperature (°C)
- X Tracer
- Z Other

8.1.7 **Reading**
8.1.8 Comment

9.0 METER READING

9.0.1 **Site identification code**
9.0.2 Header sequence number
9.0.3 **Reporting institution**
9.0.4 Date record entered
9.0.5 **Type of meter read**
- A Abstraction recorded
- H Hour meter
- P Power meter
- W Water meter
- Z Other

9.0.6 Comment

9.1 REPEATING GROUP: METER READING
9.1.1 Site identification code
9.1.2 Header sequence number
9.1.3 Repeating group number
9.1.4 **Date of reading**
9.1.5 Time of reading
9.1.6 **Reading**
9.1.7 Comment

100 PUMPING TEST

10.0.1 **Site identification code**
10.0.2 Header sequence number
10.0.3 **Reporting institution**
10.0.4 Date record entered
10.0.5 **Date pumping test started**
10.0.6 **Method tested**
- A Air test
- B Bailer test
- F Free flow test
- O Owner test
- P Controlled pumping test (step or main)
- R Controlled recovery test (step or main)
- S Slug test

10.0.7 Depth to pump intake
10.0.8 Recommended abstraction rate
10.0.9 Transmissivity
10.0.10 Storativity
10.0.11 Comment

110 MISCELLANEOUS

11.0.1 **Site identification code**
11.0.2 Header sequence number
11.0.3 **Reporting institution**
11.0.4 Date record entered

Figure 3.4. Continued.

11.1 REPEATING GROUP: SITE SELECTION
11.1.1 Site identification code
11.1.2 Header sequence number
11.1.3 Repeating group number
11.1.4 Description and recommendation

11.2 REPEATING GROUP: OWNER
11.2.1 Site identification code
11.2.2 Header sequence number
11.2.3 Repeating group number
11.2.4 Ownership date
11.2.5 Name of owner

11.3 REPEATING GROUP: OTHER IDENTIFIER
11.3.1 Site identification code
11.3.2 Header sequence number
11.3.3 Repeating group number
11.3.4 Other identifier
11.3.5 Assignor

11.4 REPEATING GROUP: OTHER DATA
11.4.1 Site identification code
11.4.2 Header sequence number
11.4.3 Repeating group number
11.4.4 Data type
C Conductivity
G Gamma, Gamma/gamma
K Caliper
N Neutron
P Self-potential
R Resistivity
S Sieve analysis
T Temperature
Z Other
11.4.5 Type of logger
11.4.6 Location
11.4.7 Format
F Files
M Machine readable
P Published
Z Other format

11.5 REPEATING GROUP: SITE VISITS
11.5.1 Site identification code
11.5.2 Header sequence number
11.5.3 Repeating group number
11.5.4 Date of visit
11.5.5 Visitor

11.6 REPEATING GROUP: SPECIAL CASES
11.6.1 Site identification code
11.6.2 Header sequence number
11.6.3 Repeating group number
11.6.4 Number of hole etc. in group
11.6.5 Depth of the deepest hole
11.6.6 Depth of the shallowest hole
11.6.7 Method of holes constructed
11.6.8 Diameter of group
11.6.9 Length of pond/tunnel/drain
11.6.10 Strike of pond/tunnel/drain
11.6.11 Dip of tunnel or drain
11.6.12 Depth of lateral collector
11.6.13 Length of lateral
11.6.14 Mesh of screen in lateral

11.7 REPEATING GROUP: REFERENCES
11.7.1 Site identification code
11.7.2 Header sequence number
11.7.3 Repeating group number
11.7.4 Reference (map/section/report)

11.8 REPEATING GROUP: SEARCH STATUS
11.8.1 Site identification code
11.8.2 Header sequence number
11.8.3 Repeating group number
11.8.4 Record type & period searched

11.9 REPEATING GROUP: COMMENTS
11.9.1 Site identification code
11.9.2 Header sequence number
11.9.3 Repeating group number
11.9.4 Comment

Figure 3.4. Continued.

(ii) aquifer codes and
(iii) lithology codes

are listed in tables 1, 2 and 3 in the appendix of chapter 3 of volume 1 of the User's Guide.

Codes were defined by using the U.S.G.S. method. Table 3.1 lists the order of letter elimination used to derive at the appropriate codes. Deviations from this method, however, occurred in order to accommodate similar names.

TABLE 3.1. ORDER OF LETTER ELIMINATION (AFTER WATSTORE, USER'S GUIDE).

1. A	10. T	19. G
2. E	11. N	20. P
3. I	12. S	21. K
4. O	13. R	22. B
5. U	14. L	23. V
6. W	15. D	24. X
7. H	16. C	25. J
8. Y	17. M	26. Q
9. Double letters (Delete one)	18. F	27. Z

District and lithology codes were solely derived from the respective names, whereas the aquifer codes consist of a numeric as well as a character portion.

Province and district codes

Relevant codes are derived from the names of the administrative districts. The administrative districts were chosen, because their boundaries do not change as opposed to the frequently changing demarcations of the magisterial districts.

In Transvaal, a coding system is used to identify the different districts. The whole province of Transvaal is subdivided into a network of square degrees. Each square degree is identified by a two-character code (e.g. KO), representing a specific district. This system is known as the Degree Square Index System. This system has been adopted in the data base. The province and district codes for the Transvaal are therefore based on codes used in the square index reference system. Figure 3.5 depicts the subdivision of administrative districts in the Transvaal.

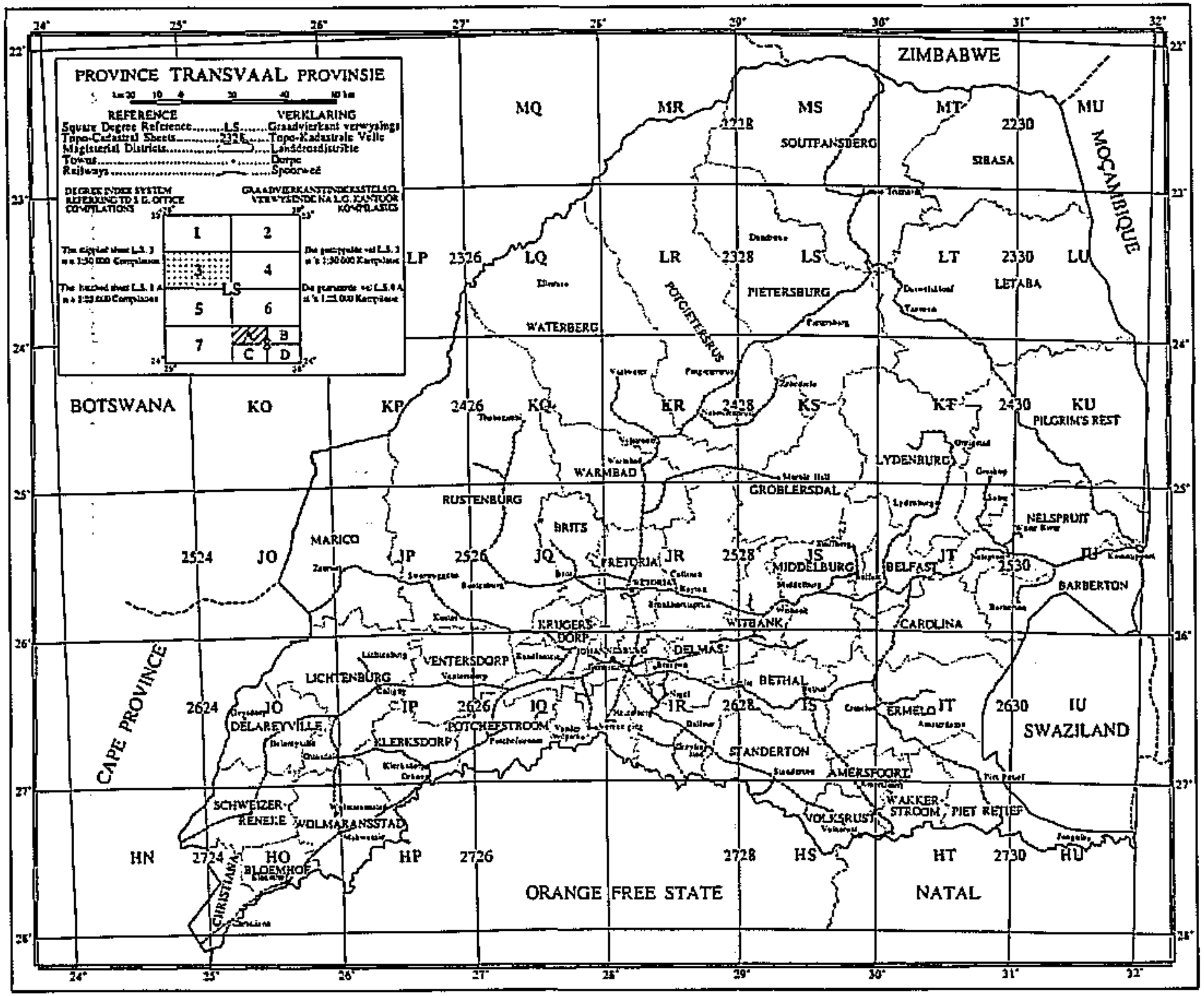


Figure 3.5. Administrative district subdivisions in the Transvaal.

Aquifer codes

The codes used are based on formations occurring in South Africa and neighbouring countries (Stratigraphy of South Africa, Handbook 8, 1980). The first two characters of each code are numeric, indicating the relative age of the different geological groups, complexes, formations or units. The oldest stratigraphical group is represented by the digits 01 and the youngest group by digits 46. The last four characters of the code are derived from the formation name.

Lithology codes

These codes are based on the different rock types normally found in South Africa.

The aquifer and lithology code lists contain codes for all major formations and the commonly occurring rock types. New codes can be added to the lists by the Data Base Administrator, if the need arises.

After review of the parameters accommodated in the data base and the codes used, data entry of site specific data is dealt with in the following section.

3.5 DATA INPUT

Having created the structured environment of a data base for storage of geohydrological data and defined all the relevant data items accommodated in the data base, data can be entered into this structure.

It is the responsibility of the user to decide which data he enters. Data entered should be feasible, reliable and site specific. Field data, contained in reports or in a raw format, must be thoroughly checked and edited before they are stored.

Geohydrological data can be subdivided into five groups. Coding forms and screen outlays for data entry (on the terminals) have been designed and created accordingly. The subdivision of groups is as follows:

- (i) General information pertaining to the site.
- (ii) Information gathered during the making of the site.
- (iii) Data regarding the construction and installation of the site.

(iv) Observations made during testing and operation of the site.

(v) Information which does not fall under one of the above categories.

Data may consist of either text, code, integer, number, date or time. Chapter 4 of volume 1 of the User's Guide contains the rules and guide-lines for the completion of the coding forms.

Six different data coding forms are provided for the entry of the 32 data sets (see chapter 3 of the User's Guide, Volume 1). Mandatory items are marked on the coding forms, i.e. the name of the specific data item is boxed. Figure 3.6 shows as an example the ENTRY data coding form currently in use.

After completion of the coding form, data are ready to be entered into the computer by the user. Numerous tests have been implemented in the programs for syntax-, compatibility- and out-of-range checks of the incoming data. This does, however, not relieve the data collector from providing accurate data. Deductions or estimates should not be included, if they may be regarded as measured data (they may be placed in the Comments).

The basic information (ENTRY) regarding any specific site, must always be entered before any other information for the site can be accommodated.

Establishing contact with the mainframe computer is comprehensively described in chapter 5 of the User's Guide. Once contact has been made, a number of pages or screens can be paged through (depending on the terminal up to 10 screens). The LINC system, however, uses only the first two pages of the terminal. Page one is the most important and accommodates all data screens. Page 2 is used for error messages, warnings, help menus and direct communication with LINC.

After logging onto the data base, the first operational screen is used to enter the site identification code, the default reporting institution, current date and instructions. A next screen number field which appears on most of the screens provides the facility to select the screen to be added, changed, deleted or wanted for inquiry. The screen numbers coincide with the data set numbers (see figure 3.3).

Four possible transactions exist:

(i) ADD - to add new data (default)

(ii) CHG - to change existing data

(iii) DEL - to delete data

(iv) INQ - to inquire about specific data

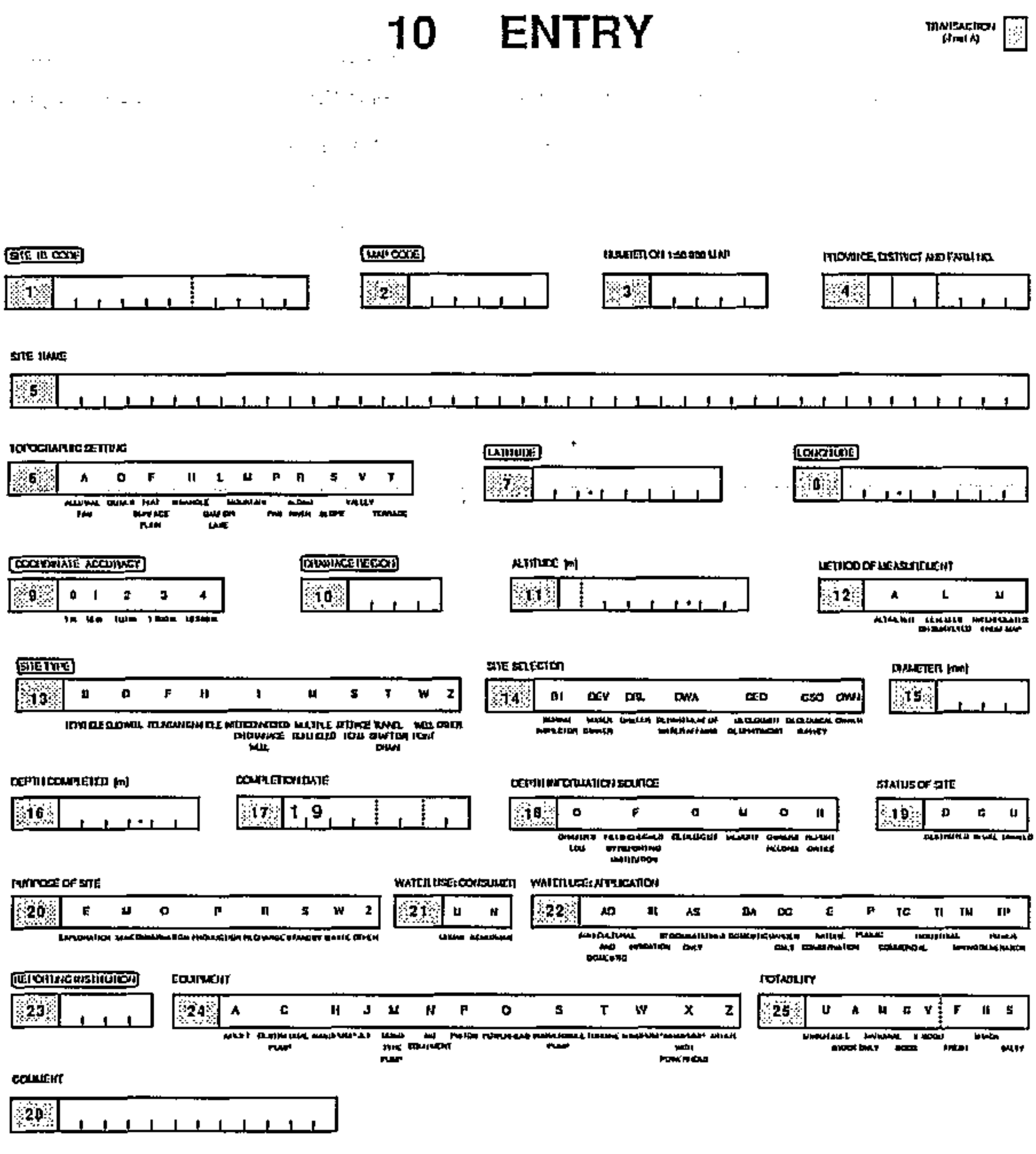


Figure 3.6. Data coding form example.

(v) ESC - return to first operational screen

The ESC option can be used to abort a screen and to return to the first operational data base screen. Having chosen the relevant screen, one of the above-listed commands has to be selected before any data can be handled. If a record is added to a data set that contains a header, then the header will always be displayed first and the header has to be completed before the relevant repeating group screen is displayed. The system automatically proceeds from the header to the requested repeating groups. Headers with the repeating groups under them have therefore no next screen field (see chapter 5 of the User's Guide for further detail).

The data fields on the screens are arranged in the same order as on the corresponding data coding forms. Until the transmit key is pressed, it is possible for the user to revise the contents of any of the unprotected fields and correct them where required.

The deletion of records from the data base consists of two phases. The first being the marking of the faulty records which is done under LINC, whereas the second phase comprises the actual deletion of the records, being done under the Command and Edit Language (CANDE). This task can only be carried out by the data base administrator.

To delete a specific record, it must be retrieved from the data base by the INQ transaction. Once the record appears on the screen, the transaction code is changed to DEL and the record is transmitted. The record remains in the data base, but is no longer visible to the user.

If an ENTRY data set is marked for deletion, the ENTRY data set as well as all the other data sets resorting under that specific ENTRY, are deleted. This ensures that no repeating group exists without a header and no header exists without an ENTRY.

Once data entry is completed, the session can be terminated by following the proper prescribed log-off procedures (see chapter 5 of the User's Guide).

3.6 DATA OUTPUT

Storage of geohydrological related data is the primary task of a data base. Just as important, however, is the retrieval of the data for further evaluation purposes.

Data can be extracted from the data base in four ways:

(i) Screen output.
(ii) Data extraction via printer file.

(iii) Statistics of data.

(iv) Data extraction via transmission file.

The extraction occurs by means of the software provided in the data base. Similar to the deletion procedure, inquiries followed by extraction of data can be subdivided into two distinct stages. At first, the inquiry is specified under LINC, after which a print-out or a transmission file is generated by means of the Logic and Information Report Compiler (LIRC) run under CANDE.

Both the above-mentioned stages will be discussed subsequently in more detail.

LINC stage

A great advantage for the user of a data base is to select data sets for extraction to satisfy his specific requirements. A special inquiry screen was developed to assist the user in selecting the data wanted. Figure 3.7 shows the inquiry screen.

```
<T}<140]<XINQU]<D0375]<30SEP88]^8620]
          N A T I O N A L   G R O U N D - W A T E R   D A T A   B A S E
                              I N Q U I R Y   S C R E E N

Title of request  [                                                          ]
Requested by       [                                            Date <19880930]
---------------------------------------------------------------------------------
Single sites from [                ] to [             ]
       OR
Grid = Latitude:     minimum ^ 0] ^       ]  maximum  ^ 0] ^       ]
Grid = Longitude:    minimum ^ 0] ^       ]  maximum  ^ 0] ^       ]
       OR
Province [ ]  District [ ]  Farm [      ]   OR   Drainage region  [      ]
---------------------------------------------------------------------------------
Site  type  [  ]  Completion/Digitizing date from ^              ] to ^19880930]
---------------------------------------------------------------------------------
Chose option: mark with 'X' or number [ ] DATA              [ ] STATISTICS
[ ] ENTRY           [ ]  LOGS             [ ] CONSTRUCTION    [ ] INSTALLATION
[ ] DISCHARGE       [ ]  LEVEL            [ ] SAMPLE          [ ] FIELD MEASUREMENT
[ ] METER READING   [ ]  PUMPING TEST     [ ] MISCELLANEOUS   [ ] ALL DATA
---------------------------------------------------------------------------------
Output to: [ ] Printer file <P> or Transmission file <T>         Go now [ ] <Y/N>
---------------------------------------------------------------------------------
Return to BEGIN <leave blank for another inquiry, else Y [ ] [@]
```

Figure 3.7. Inquiry screen.

This screen can be called up by entering 850 as the next screen number. The inquiry screen can be divided into five parts:

(i) Heading
(ii) Site selection
(iii) Site type
(iv) Data set selection
(v) Output options

Data can be extracted according tot the options provided under (ii) to (iv).

Heading - consists of title under which the inquiry can be recognized at a later stage, as well as the name of the requestor.

Site Selection - comprises the selection of sites by either one of the following parameters:

(i) Single site(s)
(ii) Grid
(iii) Province, district and/or farm number
(iv) Drainage region

Site Type - selects a specific site by entering the relevant code in the appropriate field. If, however, the site type field is left blank, data for all types will be provided.

Data Selection - A further two options are provided, i.e. either the raw data or the statistical information (4 different output options) about the data sets in the data base are extracted. If raw data are chosen, the opportunity exists to extract any one of the major data sets, shown in figure 3.7, or all the data. Detail regarding any of the chosen data sets can be selected, by using one of the following codes:

X - All the data of the selected data sets are extracted.
H - Only the headers of the selected data are extracted.
R - Only the repeating groups of the selected data sets are selected (use numbers 1 to 6 in cases where only one repeating group, resorting under one header, is required).

Output Options - consist of two data output selections, either as printer file or as transmission file. Printer files are chosen for print-out of data in lists, while transmission files are used to extract data for transfer and further processing on microcomputer. The main difference between the two file types is summarized in table 3.2.

Figures 3.8(a) and (b) are examples of the printer file output, whereas an example of the transmission file can be seen in figure 3.9.

(a)

```
LOGS HEADER          HDR REP I   DATE E

      3025BB00109     1 IGS   19851021 A
      3025BB00332     1 IGS   19860116 A
      3025BB00335     1 IGS   19860116 A
      3025BB00336     1 IGS   19860116 A


AQUIFER              HDR REP     TOP BOTTOM    CODE COMMENT

      3025BB00109     1     1  13.70    0.00 40KRDL BFRT/DOL       A
      3025BB00332     1     1  26.60    0.00 40KRDL BFRT/DOLSILL A
      3025BB00335     1     1   7.60    0.00 40KRDL BFRT/DOLDYKE A
      3025BB00336     1     1   7.60    0.00 40KRDL BFRT/DOLDYKE A


DISCHARGE HEADER  HDR REP I DATE E METH
                                 TYPE SRCE

      3025BB00109     1 IGS   19851021 P      A
      3025BB00324     1 IGS   19860116 P      A
      3025BB00335     1 IGS   19860116 P      A


LEVEL HEADER       HDR  REP I    DATE E ME    COLLR


                    PZ                    STAT       SRCE

      3025BB00090    1 0 IGS   19851016        99.99    A
      3025BB00098    1 0 IGS   19851016        99.99    A
      3025BB00109    1 0 IGS   19851021        99.99    C
      3025BB00130    1 0 IGS   19851212        99.99    A
      3025BB00131    1 0 IGS   19851212        99.99    A
      3025BB00132    1 0 IGS   19860103        99.99    A
      3025BB00133    1 0 IGS   19860103        99.99    A
      3025BB00335    1 0 IGS   19860116        99.99    A


WATER-LEVEL        HDR REP    DATE M TIME    LEVEL COMMENT

      3025BB00090    1    1 19770321 2400     0.50                   A
      3025BB00098    1    1 19770323 2400    11.00                   A
      3025BB00109    1    1 19770321 2400     7.50                   A
      3025BB00130    1    1 19770321 2400    11.90                   C
      3025BB00131    1    1 19770321 2400    12.00 CHECK DATE        C
      3025BB00131    1    2 19770321 2400     9.00                   C
      3025BB00132    1    1 19770321 2400     3.50                   A
      3025BB00133    1    1 19770321 2400     1.00                   A
      3025BB00335    1    1 19570000 2400     0.50                   A
```

Figure 3.8 (a). Printer file output of site specific data.

(b)

```
NATIONAL GROUND-WATER DATA BASE              19860722

REQUEST:  STATS 1                                              00000000 TO 19860722
BY:       KI                           DATE REQUESTED: 19860722  SITE TYPE:
TYPE: PROV. & DIST. & FARM NO.  O F S              TR       167
LATITUDE 00.00000 - 00.00000  LONGITUDE 00.00000 - 00.00000


THERE ARE      17 ENTRIES WITH THE FOLLOWING COORDINATE ACCURACIES:

ACCURACY         0        1        2        3        4     TOTAL
                 0        0       17        0        0        17
```

Figure 3.8 (b). Printer file output of site data statistics.

```
0103025AA000013025AA000010PH  14GROOTFONTEIN
F30.1894425.17083204311380.00 MB    0200000.0000000000RGPNASIGS W
1986092419860924
0503025AA00001  1IGS 19860924P
0513025AA00001  1  1197802282400  0.38
0603025AA00001  10IGS 19860924  99.99
0613025AA00001  1  1197802282400  7.10
0103025AA000023025AA000020PH  14GROOTFONTEIN
F30.1969425.17056204311385.00 MB    0200000.0000000000RGPNASIGS W
1986092419860924
0503025AA00002  1IGS 19860924P
0513025AA00002  1  1197803282400  0.38
0603025AA00002  10IGS 19860924  99.99
0613025AA00002  1  1197802282400 10.80
0103025AA000033025AA000030PH 498GELUK
F30.2111125.18417204311399.00 MB    0200000.0000000000RGPNASIGS W
1986092419860924
0503025AA00003  1IGS 19860924P
0513025AA00003  1  1197802272400  0.44
0603025AA00003  10IGS 19860924  99.99
0613025AA00003  1  1197802272400 10.90
0103025AA000043025AA000040PH 498GELUK
F30.2177825.18944204341414.00 MB    0200000.0000000000RUPNASIGS W
1986092419860924
0603025AA00004  10IGS 19860924  99.99
0613025AA00004  1  1197802282400 18.15
0103025AA000053025AA000050PH 498GELUK
F30.2286125.18417204311400.00 MB    0000000.0000000000RGPNASIGS W
1986092419860924
0503025AA00005  1IGS 19860924P
0513025AA00005  1  1197802272400  2.52
0603025AA00005  10IGS 19860924  99.99
0613025AA00005  1  1197802272400  9.50
0103025AA000063025AA000060PH  14GROOTFONTEIN
F30.1922225.14750204311376.00 MB    0200000.0000000000RGPNASIGS W
1986092419860924
0503025AA00006  1IGS 19860924P
0513025AA00006  1  1197802282400  1.26
0603025AA00006  10IGS 19860924  99.99
0613025AA00006  1  1197802282400 13.10
0103025AA000073025AA000070PH  14GROOTFONTEIN
F30.1994425.13722204341365.00 MB    0000000.0000000000RGPNASIGS W
1986092419860924
0503025AA00007  1IGS 19860924P
0513025AA00007  1  1197802282400  0.76
0603025AA00007  10IGS 19860924  99.99
0613025AA00007  1  1197802282400 10.95
```

Figure 3.9. Transmission file output example.

TABLE 3.2. OUTPUT OPTION DIFFERENCES.

PRINTER FILE	TRANSMISSION FILE
132-character records	78-character records
5-character names	3-digit record id. codes
one space between each data field	no spaces between data fields
expanded inquiry header	no inquiry header

CANDE stage

Printer and transmission files are initially output to printer back-up files under LINC. The printer back-up files are temporary files. Print-outs of the back-up files are possible. To secure the data in a more permanent fashion, the respective printer back-up files may be stored in a permanent file on a back-up disk pack. The transmission files must always be converted to ordinary data files before they can be transmitted to the micro computer. This file conversion is also done under CANDE. Once the data have been transmitted to a micro computer for further processing or are no longer needed, they should be removed from the back-up disk pack.

A detailed description of the data extraction procedure is given in the User's Guide, chapter 6. More detail information about completing the inquiry screen, including the selection of the relevant data items, is provided in chapter 6 of the User's Guide.

3.7 HARDWARE REQUIREMENTS

The data base is implemented on a mainframe computer with the usual peripherals. Terminals, printers and microcomputers are linked either directly or via the Govnet Transparent Protocol Network (GTPN) to the mainframe.

Due to the dual purpose of the data base package, i.e. data storage as well as data processing at a later stage, the required hardware can be subdivided into two parts. The first being for the access of the National Ground-water Data Base (NGDB) on the mainframe computer, whereas the second part includes all the peripherals necessary for the processing on (IBM compatible) micro computer.

Access to the mainframe computer is normally gained by means of Govnet certified terminals through GTPN, using the Post Office's SAPONET-P network as the main carrier. A wide variety of micro computers, especially IBM compatibles, may also be connected to the network, provided they have been certified by the Govnet Organization. Govnet only provides the necessary emulation software for genuine IBM compatible personal computers of the XT and AT varieties. These have the distinct advantage of either emulating a mainframe terminal or operating as a micro computer. For file transfer, the PC/XT must also be linked to the Govnet Standard Protocol Network (GSPN). The GSPN is a host independent network, providing file transfer facilities between a large variety of mini- and mainframe computers used by the Central Government.

In order to obtain any printer file output, a wide variety of printers is supported by both networks. The printers must also be certified by Govnet if connected in an addressable mode from the mainframe computer.

The minimum hardware requirements for a data base work station are an IBM compatible microcomputer, a printer, a terminal emulator and GTPN network equipment.

Data processing is mainly done by micro computer. The type of micro computer currently in use is an IBM compatible (MS.DOS) personal computer. In order to ensure efficient use of such a machine during data processing routines, certain minimum hardware requirements are mandatory:

(i) 640 Kb RAM (minimum).
(ii) 8087 Math co-processor.
(iii) 1 x 360 Kb Floppy disk drive.
(iv) 1 x 20 Mb Hard disk.
(v) 1 x Serial interface.
(vi) 1 x Parallel interface.
(vii) Intercom 1000 communications card (mainframe terminal emulator).
(viii) 80-Character per line dot matrix printer.
(ix) External plotting device (minimum paper size A3).
(x) Appropriate digitizer (optional, minimum size A1).

The GTPN network is specially designed to provide interactive processing with full screen emulation capability for a variety of terminals of various dissimilar types of hosts. The network equipment and software controlling the network are locally designed and manufactured and are based on a host independent network.

Due to model and vendor changes, it is not possible to supply brand names or suppliers, but a diagrammatic representation of a basic network is provided in Figure 3.10.

The Govnet organization is a bureau of the Commission for Administration that provides a data communication service for Central Government Departments.

3.8 COMMUNICATIONS

Due to the distribution of work stations in South Africa, a transparent network was developed. This network is connected to the mainframe computer and other host computers under SAPONET-P protocols and supports remote devices by means of data concentrators and distributors.

There are two ways of interconnecting the various remote elements. The first way is to interconnect elements using a high performance bus compatible area networking system, while the second way makes use of leased or public data networks such as SAPONET-P. These two interconnection schemes can be used simultaneously in different configurations. Attention will be focussed on the latter option. Such a network configuration is generally also known as a wide area network.

In such a network, the host controller and site controllers form the basis of the system. When messages are received from either a host computer or any remote device, the master node assembles the individual characters into one or more packets of data, generally 256 characters in length. Before they are transmitted, the packets are placed into frames, which respectively contain addressing, control and sequence numbering fields to check on the destination of the packets and to verify their correct sequencing. To ensure the transfer of data between communication devices, using a shared data network, the packet switching technique is used. The master node collects data packets from different types of hosts and deliver these packets through the available links.

Network links connecting the nodes in the network system may be configured to use public data networks or dialled connections. The network under discussion uses SAPONET circuits via X.25 protocols. When using the X.25 network, SAPONET takes care of the data routing. Thus, should any port fail, the transmission of data cannot take place. Figure 3.11 shows a diagrammatic representation of a network.

The GTPN communications network ensures data integrity at all times by the high level end-to-end protocol within the system, such that all data transported through the

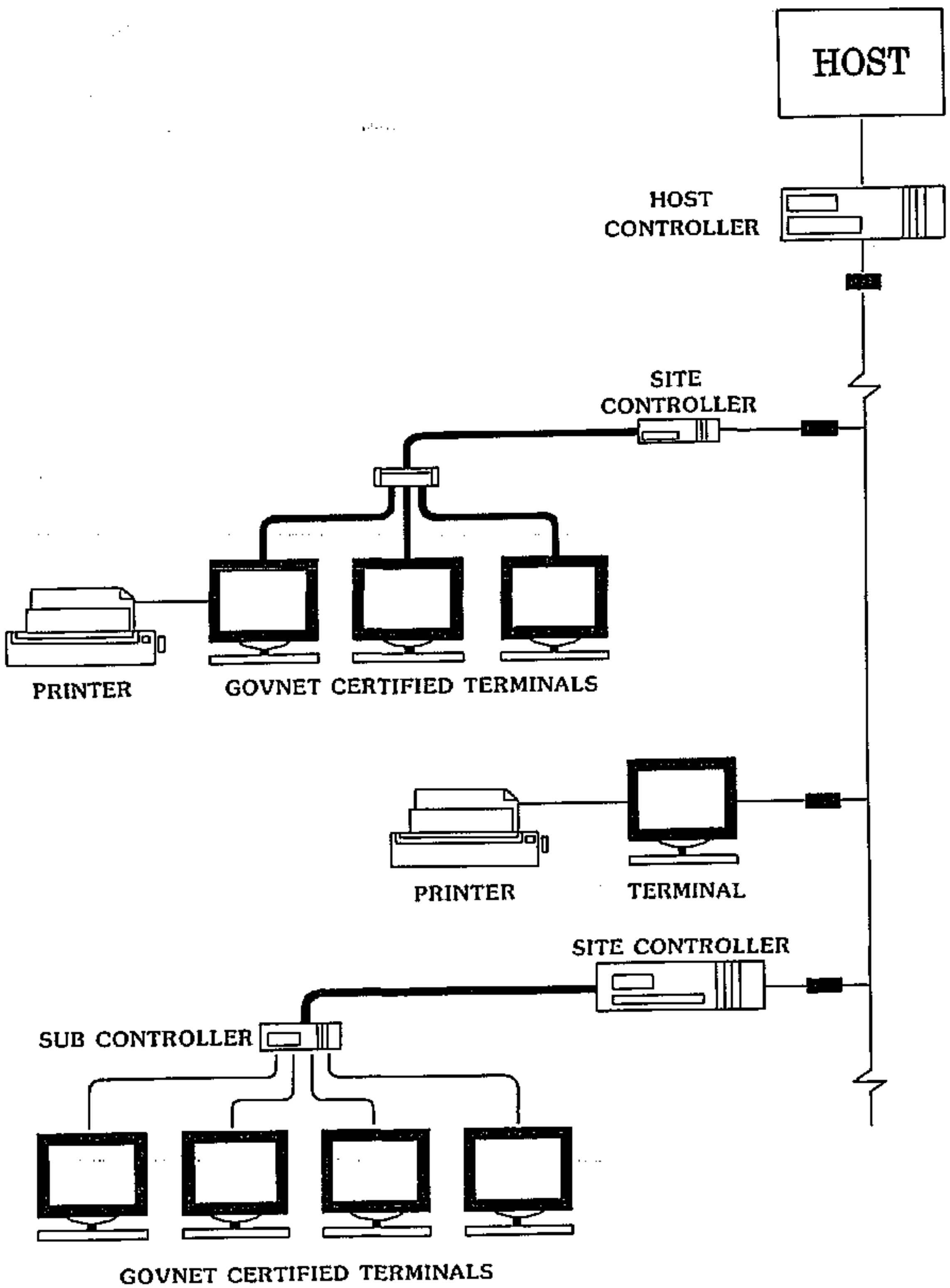


Figure 3.10. Elements of a Govnet Transparent Protocol Network.

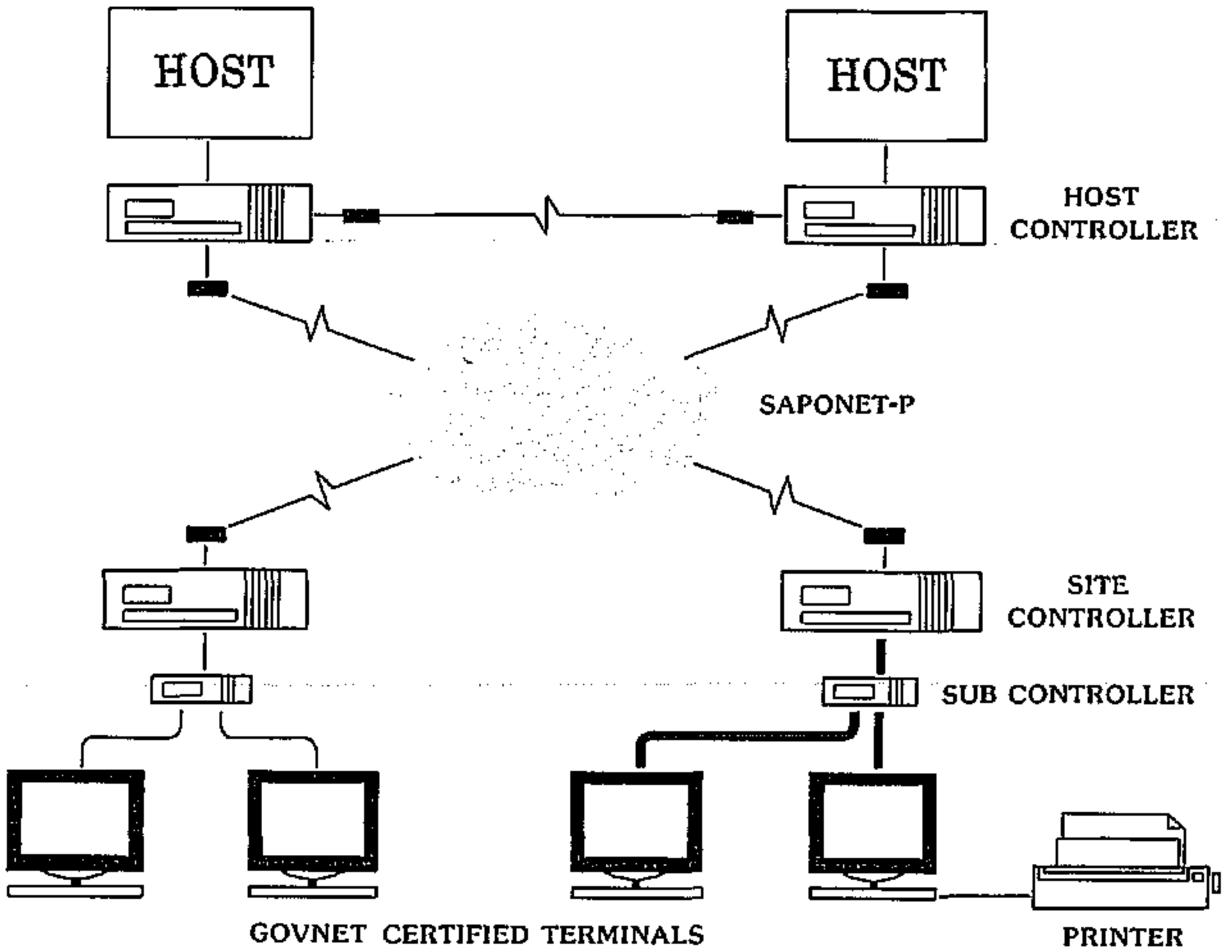


Figure 3.11. *Networking via SAPONET-P.*

network are handled in the same way, whether it is asynchronous or synchronous, individual or packets of information.

This concludes the communications between the host computer and remote terminals. However, at this stage a micro computer is still unable to form part of the network system. This is only possible if the computer can effectively emulate a specific computer terminal. The INTERCOM 1000 communications card and software provide such a facility. After installing the communications card and software on the micro computer, it can operate when activated as a TD830 computer terminal. Simultaneously, software is provided to perform complete CANDE file transfer, creating the facility to export data files from the mainframe computer to the micro computer.

CHAPTER 4

DATA PROCESSING

4.1 INTRODUCTION

Ground water and the geological formations through which it flows constitute a complex dynamic system in which any change of one part of the system is reflected in the others. Therefore, when such a dynamic system is evaluated, attention must be focussed on the interaction which the various components have on each other. One must therefore understand the interactions well to develop, exploit and manage the available ground-water resources to the best advantage of the country.

To develop the ground-water potential fully, the resources must be evaluated quantitatively. A major part of the data base project was to provide various computer programs, which assist the user in a wide variety of data processing techniques.

Apart from certain statistical inquiries, all further data processing is done on IBM compatible computers. Before any data in the National Ground-water Data Base can be evaluated by any of the techniques provided in the data processing package, they have to be transferred from the mainframe computer to the microcomputer (see chapter 6.3 of the User's Guide).

All data required for a certain investigation are transferred once and in one file. This is not only done because it is difficult to provide transmission files for each of the various data files required by the processing programs, but mainly to save on the time and cost of data transmission. The transmission file is unravelled by a special program and the various selected data sets are then stored in binary files on the hard disk of the mini computer.

A second program allows the user to select data items from these binary files and to combine them in ASCII data files required by the various processing programs. Thus, the whole process which creates the data files is automated to a large extent. This enables the user to use the programs without detailed knowledge of data manipulation on a computer. Information about the data selection is contained in chapter 6 of the User's Guide.

Figure 4.1 shows the available data processing possibilities in the form of a diagram. They comprise three basic options :

(i) Data tabulation.
(ii) Graphical presentation.
(iii) Statistical analyses.

4.2 DATA TABULATION

Certain data or combinations of data cannot adequately be represented by graphical display nor can they be statistically analysed, but they can be presented in tables. Tabular data output is subdivided into two groups, namely: (a) the data output facility of the mainframe computer and (b) data output of processed data by micro computer.

The first option has been discussed in some detail in section 3.5 and is also dealt with in chapter 6 of the User's Guide. Here attention will be focussed on tables created by micro computer routines.

General tables

Depending on the requirements of the user, the number of tables are virtually unlimited. A selection had therefore to be made. At present, selected data from all 24 data sets can be printed in tables. These data sets are:

(i) Entry.
(ii) Penetration rate.
(iii) Aquifer.
(iv) Geology.
(v) Hole.
(vi) Casing.
(vii) Openings.
(viii) Fill.
(ix) Installation.
(x) Discharge Rate.
(xi) Water Level.
(xii) Water Sample.
(xiii) Field Measurement.
(xiv) Meter Reading.
(xv) Pumping Tests.
(xvi) Site Selection.
(xvii) Name of Owner.
(xviii) Other Identifier.
(xix) Other Data.
(xx) Site Visits.
(xxi) Special Cases.
(xxii) References.
(xxiii) Search Status.
(xxiv) Comments.

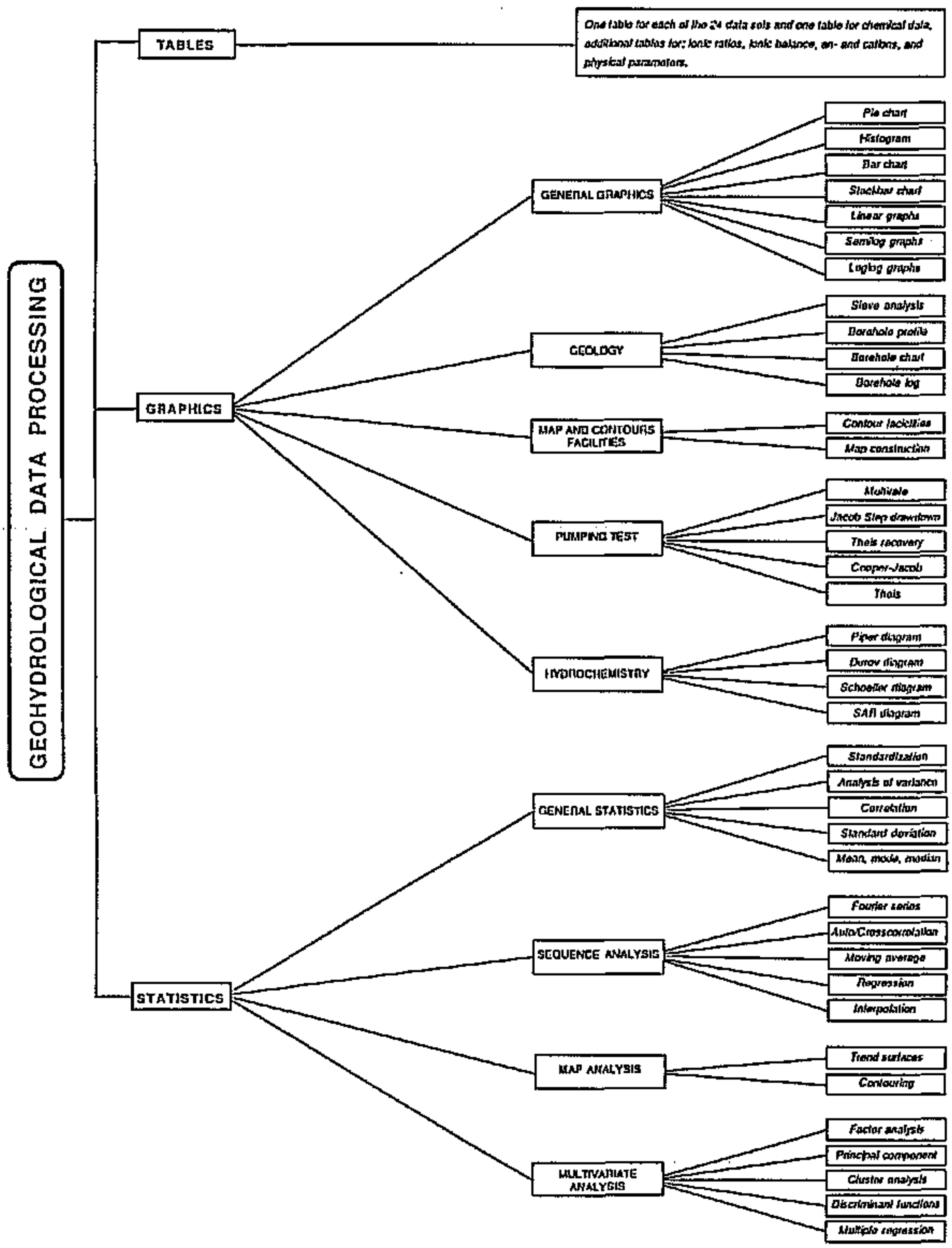


Figure 4.1. Diagrammatic representation of the data processing possibilities.

Site Id.	Topo. Setting	Latitude	Longitude	Altitude (m)	Site Type	Site Selector	Nominal Diameter (mm)	Depth Comp. (m)	Site Status	Site Purpose	Water Use	Water Appl.	Equip-ment	Pota-bility	Reporting Inst.
3318AD000008	D	33.37834	18.25694	0.00	B	DWA	165	58.00	G	E	U				WAGC
3318AD000009	D	33.40222	18.27000	0.00	B	DWA	165	02.00	G	E	U				WAGC
3318AD000010	D	33.45639	18.30472	0.00	B	DWA	165	68.00	G	E	U				WAGC
3318AD000011	D	33.42834	18.20056	0.00	B	DWA	165	52.00	G	E	U				WAGC
3318AD000012	D	33.44583	18.29361	0.00	B	DWA	165	48.00	G	E	U				WAGC
3318AD000066	D	33.45861	18.37473	109.66	B	DWA	0	29.00	G	E	U				WAGC
3318AD000067	D	33.40167	18.28111	63.97	B	DWA	0	54.50	G	E	U				WAGC
3318AD000068	D	33.46945	18.31556	36.00	B	DWA	0	21.00	G	E	U				WAGC
3318AD000069	A	33.44416	18.32000	71.89	B	DWA	0	51.50	G	O	U				WAGC
3318AD000073	A	33.40167	18.28111	64.18	B	DWA	0	50.30	G	E	U				WAGC
3318AD000074	D	33.45139	18.32723	67.67	B	DWA	0	49.00	G	E	U				WAGC
3318AD000077	A	33.46250	18.35003	75.50	B	DWA	165	42.00	G	E	U				WAGC
3318AD000078	A	33.46139	18.33167	40.81	B	DWA	0	40.00	G	E	U				WAGC
3318AD000079	A	33.45417	18.29416	43.98	B	DWA	0	66.58	G	E	U				WAGC
3318AD000081		0.00000	0.00000	0.00			0	0.00							
3318AD000087	D	33.46556	18.30583	0.00	B	DWA	0	50.00	G	E	U				WAGC
3318AD000089		0.00000	0.00000	0.00			0	0.00							
3318AD000090		0.00000	0.00000	0.00			0	0.00							
3318AD000094	D	33.39444	18.28278	0.00	B	DWA	165	59.00	G	E	U				WAGC
3318AD000095		0.00000	0.00000	0.00			0	0.00							
3318AD000113	D	33.44416	18.32000	71.93	B	DWA	0	51.50	G	P	U				WAGC
3318AD000120	D	33.36889	18.36809	62.75	B	DWA	380	58.00	G	Z	U				WAGC
3318AD000121	D	33.44416	18.32000	71.93	B	DWA	0	51.50	G	P	U				WAGC
3318CB000001	D	33.62699	18.44198	54.67	B	DWA	203	35.00	G	P	U	P	S		WAGC
3318CB000002	D	33.63458	18.44074	52.17	B	DWA	203	42.00	G	P	U	P	S		WAGC
3318CB000003	D	33.62296	18.43367	53.11	B	DWA	203	45.00	G	P	U	P	S		WAGC
3318CB000004	D	33.58014	18.37804	38.73	B	DWA	152	26.00	G	O	U	P	S		WAGC
3318CB000005	D	33.57911	18.37225	33.39	B	DWA	152	28.00	G	O	U	P	S		WAGC
3318CB000006	D	33.62959	18.39980	53.44	B	DWA	0	50.00	D	E	U	P	S		WAGC
3318CB000007	D	33.57352	18.37654	38.42	B	DWA	152	24.00	G	O	U	P	S		WAGC
3318CB000008	D	33.58167	18.37667	37.74	B	DWA	175	45.00	G	P	U	P	S		WAGC
3318CB000009	D	33.59743	18.37379	26.65	B	DWA	152	15.00	G	O	U	P	S		WAGC
3318CB000010	D	33.57305	18.46603	163.73	B	DWA	152	0.00	G	O	U	P	S		WAGC
3318CB000011	D	33.63063	18.44013	53.12	B	DWA	200	30.00	G	P	U	P	S		WAGC
3318CB000012	D	33.63210	18.42840	39.40	B	DWA	200	38.00	G	P	U	P	S		WAGC
3318CB000013	D	33.57455	18.39111	47.73	B	DWA	152	30.00	G	O	U	P	S		WAGC
3318CB000014	D	33.58995	18.49999	41.89	B	DWA	152	36.00	G	O	U	P	S		WAGC
3318CB000015	D	33.60170	18.43163	60.42	B	DWA	152	30.00	G	O	U	P	S		WAGC
3318CB000016	D	33.58167	18.37695	37.94	B	DWA	200	30.00	U	O	U	P	S		WAGC
3318CB000017	D	33.58156	18.45713	140.76	B	DWA	152	33.00	G	O	U	P	S	V	WAGC
3318CB000018	D	33.59647	18.45505	112.10	B	DWA	152	30.00	G	O	U	P	S		WAGC
3318CB000019	D	33.61945	18.42445	51.30	B	DWA	200	36.00	G	O	U	P	S	G	WAGC
3318CB000020	D	33.61402	18.43989	56.76	B	DWA	152	30.00	G	O	U	P	S		WAGC
3318CB000021	D	33.63472	18.44055	52.29	B	DWA	200	30.40	G	O	U	P	S	G	WAGC
3318CB000022	D	33.61945	18.42445	51.14	B	DWA	203	40.00	G	P	U	P	S	V	WAGC

Figure 4.2. Tabular output of the Entry data set.

Site Id.	Method Measured	Water Level Status	Collar Height (m)	Date Measured	Time Measured	Water Level (m)	Reporting Inst.
3318AD000067	E	S	0.00	19850610	2400	10.37	WAGC
3318AD000067	E	S	0.00	19850710	2400	10.32	WAGC
3318AD000067	E	S	0.00	19850807	2400	10.39	WAGC
3318AD000067	E	S	0.00	19850904	2400	10.22	WAGC
3318AD000067	E	S	0.00	19851001	2400	10.19	WAGC
3318AD000067	E	S	0.00	19851126	2400	10.20	WAGC
3318AD000067	E	S	0.00	19851212	2400	10.33	WAGC
3318AD000067	E	S	0.00	19860122	2400	10.44	WAGC
3318AD000067	E	S	0.00	19860213	2400	10.49	WAGC
3318AD000067	E	S	0.00	19860318	2400	10.56	WAGC
3318AD000067	E	S	0.00	19860414	2400	10.56	WAGC
3318AD000067	E	S	0.00	19860505	2400	10.57	WAGC
3318AD000067	E	S	0.00	19860614	2400	10.54	WAGC
3318AD000067	E	S	0.00	19860711	2400	10.44	WAGC
3318AD000067	E	S	0.00	19861003	2400	10.25	WAGC
3318AD000067	E	S	0.00	19861120	2400	10.30	WAGC
3318AD000068	E	S	0.00	19850508	2400	4.29	WAGC
3318AD000068	E	S	0.00	19850610	2400	3.75	WAGC
3318AD000068	E	S	0.00	19850710	2400	3.86	WAGC
3318AD000068	E	S	0.00	19850807	2400	3.68	WAGC
3318AD000068	E	S	0.00	19850904	2400	3.62	WAGC
3318AD000068	E	S	0.00	19851001	2400	3.66	WAGC
3318AD000068	E	S	0.00	19851126	2400	3.70	WAGC
3318AD000068	E	S	0.00	19851212	2400	3.81	WAGC
3318AD000068	E	S	0.00	19860122	2400	3.86	WAGC
3318AD000068	E	S	0.00	19860213	2400	3.91	WAGC
3318AD000068	E	S	0.00	19860318	2400	3.95	WAGC
3318AD000068	E	S	0.00	19860414	2400	3.97	WAGC
3318AD000068	E	S	0.00	19860505	2400	3.98	WAGC
3318AD000068	E	S	0.00	19860614	2400	3.90	WAGC
3318AD000068	E	S	0.00	19860711	2400	3.64	WAGC
3318AD000068	E	S	0.00	19861003	2400	3.60	WAGC
3318AD000068	E	S	0.00	19861119	2400	3.69	WAGC
3318AD000078	E	S	0.00	19850419	2400	2.59	WAGC
3318AD000078	E	S	0.00	19850508	2400	2.62	WAGC
3318AD000078	E	S	0.00	19850610	2400	2.64	WAGC
3318AD000078	E	S	0.00	19850904	2400	2.42	WAGC
3318AD000078	E	S	0.00	19851001	2400	2.43	WAGC
3318AD000078	E	S	0.00	19851126	2400	2.50	WAGC
3318AD000078	E	S	0.00	19851212	2400	2.54	WAGC
3318AD000078	E	S	0.00	19860122	2400	2.60	WAGC
3318AD000078	E	S	0.00	19860213	2400	2.63	WAGC
3318AD000078	E	S	0.00	19860318	2400	2.67	WAGC
3318AD000078	E	S	0.00	19860414	2400	2.69	WAGC
3318AD000078	E	S	0.00	19860506	2400	2.71	WAGC

Figure 4.3. *Tabular output of the Water-level data set.*

Site Id.	Depth to Top of Aquifer (m)	Depth to Bottom of Aquifer (m)	Aquifer Code
3310CB00001	0.01	30.00	45BRDP
	30.00	35.00	45VRSR
	35.00	42.00	32TGBG
3310CB00002	0.01	30.50	45BRDP
	30.50	42.00	32TGBG
3310CB00003	0.01	30.00	45BRDP
	30.00	45.00	32TGBG
3310CB00004	0.01	26.00	45BRDP
3310CB00005	0.01	26.00	45BRDP
3310CB00007	0.01	24.00	45BRDP
3310CB00008	0.01	30.00	45BRDP
	30.00	45.00	32TGBG
3310CB00009	0.01	9.00	45BRDP
	9.00	13.50	45VRSR
	13.50	15.00	32TGBG
3310CB00011	0.01	31.00	45BRDP
3310CB00012	0.01	27.00	45BRDP
	27.00	36.00	45VRSR
	36.00	38.00	32TGBG
3310CB00013	0.01	29.00	45BRDP
	29.00	30.00	32TGBG
3310CB00014	0.01	14.00	45BRDP
	14.00	31.00	45VRSR
	31.00	36.00	32TGBG
3310CB00015	0.01	24.00	45BRDP
	24.00	30.00	32TGBG
3310CB00016	0.01	29.00	45BRDP
	29.00	30.00	32TGBG
3310CB00018	0.01	22.00	45BRDP
	22.00	30.00	32TGBG
3310CB00019	0.01	36.00	45BRDP
3310CB00020	0.01	29.00	45BRDP
	29.00	30.00	32TGBG
3310CB00021	0.01	30.00	45BRDP
	30.00	30.40	32TGBG
3310CB00022	0.01	36.00	45BRDP
	36.00	40.00	32TGBG
3310CB00023	0.01	28.00	45BRDP
	28.00	37.00	45VRSR
	37.00	38.00	32TGBG
3310CB00024	0.01	31.00	45BRDP
	31.00	37.00	45VRSR
	37.00	38.00	32TGBG
3310CB00025	0.01	41.00	45BRDP
	41.00	43.00	32TGBG

Figure 4.4. Tabular output of the Aquifer data set.

In these tables, specific data of either a single site or of a number of sites can be printed. Figures 4.2, 4.3 and 4.4 are examples of some of these tables. More details about the remaining tables are shown in chapter 6.3 of the User's Guide.

Hydrochemical tables

All ground waters contain a number of solubles. Depending on their electric charge, these solubles can be divided into anions and cations. Various options are offered for the tabulation of chemical data. They can either be tabulated in mg/l or in mval/l. An important feature is the print-out of the ionic balance. Thus, before a hydrochemical evaluation is done, the ionic balance of the respective chemical analyses may be printed, if there is doubt about their reliability. Unreliable chemical analyses are depicted in the print-out by the percentage error in the ionic balance. A cut-off percentage error of 5% is used to distinguish between reliable and unreliable chemical analyses of water samples. Based on the ionic balance table, all the unreliable chemical analyses should be omitted, when any of the hydrochemical evaluation techniques are used.

In addition to the chemical tables, a special routine was developed to calculate specific chemical ionic ratios and also print them in tabulated form.

Ionic ratios are useful in determining certain characteristics of ground water. When evaluating ionic ratios, one must not attach too much significance to the order of magnitude of the values of the respective ratios, but rather compare the values of the respective ratios of different analyses with each other, in order to identify possible anomalies. Furthermore, dependent on the chemical composition of the ground water in a specific region, only certain ratios of a group of water samples may be significant.

Calculation of the following ratios is included in the data processing package:

(i) Cl/HCO3
(ii) SO4/Cl
(iii) (Cl - Na)/Cl
(iv) (Na + Cl + SO4)/(Ca + Mg + HCO3)
(v) Ca/Mg
(vi) (Na - Cl)/SO4
(vii) (Cl - Na)/Mg
(viii) (Cl - Na)/(SO4 + HCO3 + NO3)

Their application to geohydrological evaluations, as proposed by Dundon (1977), will be discussed in some detail hereunder.

Cl/HCO_3-ratio

This ratio generally indicates the relative distance from the source of recharge, i.e. the relative age of the water. The higher values are found in regions with a higher TDS further away from the recharge area, whereas low ratios are usually associated with recharge areas.

SO_4/Cl-ratio

Low ratios are often indicative of a reducing environment, i.e. infiltrating water passes through a bed, containing organic matter. Further away from the recharge area, the SO_4/Cl-ratio generally becomes smaller. However, high values may also occur as a result of gypsum dissolution. A SO_4/Cl-ratio > 1 is seen as a left-sloping line between Cl and SO_4 on the Schoeller diagram.

(Cl - Na)/CL-ratio

This ratio is generally known as the base exchange index. It means that the bivalent Ca^{++} and Mg^{++} ions are adsorbed onto clay minerals, in exchange for Na^{+} ions. Negative ratios usually indicate base exchange. The (Cl - Na)/Cl ratio can also be deduced from the Schoeller diagram. A negative ratio represents a right-sloping line between Na and Cl. A negative base exchange index value also indicates a dominance of Na over Ca, and gives an indication of the proximity to the recharge source.

$(Na + Cl + SO_4)/(Ca + Mg + HCO_3)$-ratio

This ratio distinguishes between hydrochemical samples of various compositions and can be used to indicate different groups of water.

Ca/Mg-ratio

This ratio is almost always less than one, except in dolomitic or basic igneous rock environments. A ratio > 1 indicates therefore probable recharge areas where these rock types occur.

$(Na - Cl)/SO_4$-ratio

For positive ratios, there is a sodium excess and for negative ratios, chloride prevails over sodium. Likewise, for absolute ratio values < 1 sulphate predominates or else there is a sodium or chloride surplus.

	$(Na - Cl) > 0$	$(Na - Cl) < 0$
$\frac{(Na - Cl)}{SO_4} < \lvert 1 \rvert$	$Na(HCO_3)SO_4$ (or NaCl) water	e.g.$Ca(Mg)CO_3$ water
$\frac{(Na - Cl)}{SO_4} > \lvert 1 \rvert$	$Na\text{-}HCO_3$ water	e.g.Na(Ca,Mg)Cl water

When the ratio is less than one, the water is a $NaSO_4$ (or NaCl)-water. At large negative ratios, the water will predominantly characterize a NaCl-water. Such a ratio indicates relatively old water. If the ratio is greater than one, the water is classified as a $NaHCO_3$-water. This is characteristic for dynamic water. The waters can be further subdivided into groups, such as:

(i) water with a dominant primary alkalinity,
(ii) water with a dominant secondary alkalinity, and
(iii) water with a dominant secondary salinity.

(Cl - Na)/Mg-ratio

In order to avoid misleading interpretations, this ratio should be used in conjunction with the Cl/HCO_3 and the $\frac{(Na + Cl + SO_4)}{(Ca + Mg + HCO_3)}$ -ratios. If the ratio (Cl - Na) / Mg is less than one, the analysed water is classified as a Mg - Cl-water. Such waters are mostly unsuitable for any type of consumption and are often stagnant. If the ratio is greater than one, the water is classified as a Ca - Cl-water, which also resorts under the stagnant waters. Both types of water can be subdivided into further groups, portraying the following characteristics:

(i) water with a dominant primary alkalinity, and
(ii) water with a dominant secondary alkalinity.

It should be noted that these waters are only stagnant if the $\frac{(Na + Cl + SO4)}{(Ca + Mg + HCO3)}$ - and the Cl/HCO3-ratios are considerably greater than one.

(Cl - Na)/(SO_4 + HCO_3 + NO_3)-ratio

Like the (Na + Cl + SO_4) / (Ca + Mg + HCO_3)-ratio, this ratio varies for different groups of waters and should only be used in conjunction with other ionic ratios.

4.3 GRAPHICAL REPRESENTATION

A frequently used and informative manner of presenting data is through graphical displays. The graphics provided by the data base processing software can be subdivided into five different categories:

(i) Hydrochemical analysis.
(ii) Pumping test analysis.
(iii) Geological analysis.
(iv) Map and contouring facilities.
(v) General graphics.

Each of these groups can be further subdivided into various graphical evaluation techniques. All graphical techniques included in the data processing package are discussed subsequently. Numerous examples of the application of these techniques, using data extracted from the data base, can be found in chapter 6.3 of the User's Guide.

Two graphical packages were acquired to support the limited micro computer graphics facilities:

(i) PROCAD.
(ii) PLOT88.

PROCAD is a computer-aided design (CAD) package, which is used to generate the different graphical displays. PLOT88, on the other hand, provides the contouring facilities for the construction of two- and three-dimensional plots.

4.3.1 Hydrochemical data analysis

A number of valuable techniques uses graphical methods to simplify the evaluation of hydrochemical data and their trends. To represent data for interpretation purposes, two types of displays exist, namely maps showing the geographical distribution of suitable parameters and secondly, parameter relationship diagrams. Both types of graphical representations should be used concurrently. In this section, attention will be focussed on the relationship diagrams. Maps will be discussed in section 4.3.4.

The results of analyses can either be expressed in absolute values or as percentages, depending on the type of diagram used. Of the many techniques available, some of the more popular techniques were computerized. They are:

(i) Piper diagram.
(ii) Durov diagram.
(iii) Schoeller diagram.
(iv) SAR diagram.

Except for the SAR diagram, all these diagrams are multivariate diagrams. Graphical display of various constituents from different ground-water samples permits direct comparison of various ground-water types. By using different symbols, it is possible to distinguish between waters from stratigraphically different aquifers, from different depths or of different evolutionary status (Matthess, 1982).

Piper diagrams

In the early 1940's, Hill (1940) and Piper (1944) represented hydrochemical ground-water analyses on trilinear diagrams. In these diagrams, the major cations and anions of the analyses are plotted in the two base triangles as cation and anion percentages of milli-equivalents/litre. The respective cation and anion positions of each analysis are projected onto a diamond field which represents the total ion relationship. Figure 4.5 shows the Piper trilinear diagram. An important parameter, the actual ion concentration, is (normally) not depicted in the Piper diagram. This is a major shortcoming.

Durov diagram

This is also a trilinear diagram and has been developed by Durov in the late 1940's. Similarly to the Piper diagram, it is based on percentage milli-equivalent values. However, additional features like TDS, conductivity and pH are also incorporated into this diagram. Again the respective ions are plotted in the appropriate triangles. These positions are now projected into the adjoining square field. They can be further projected into TDS-, conductivity- and pH fields. Figure 4.6 represents such a diagram.

Generally, trilinear relationship diagrams, such as Piper and Durov diagrams, not only permit comparison between numerous types of ground waters, but also reflect the effect of certain geochemical processes, e.g. mixing of different waters. If no precipitation takes place, mixtures of two different waters usually lie on a straight line (Matthess,

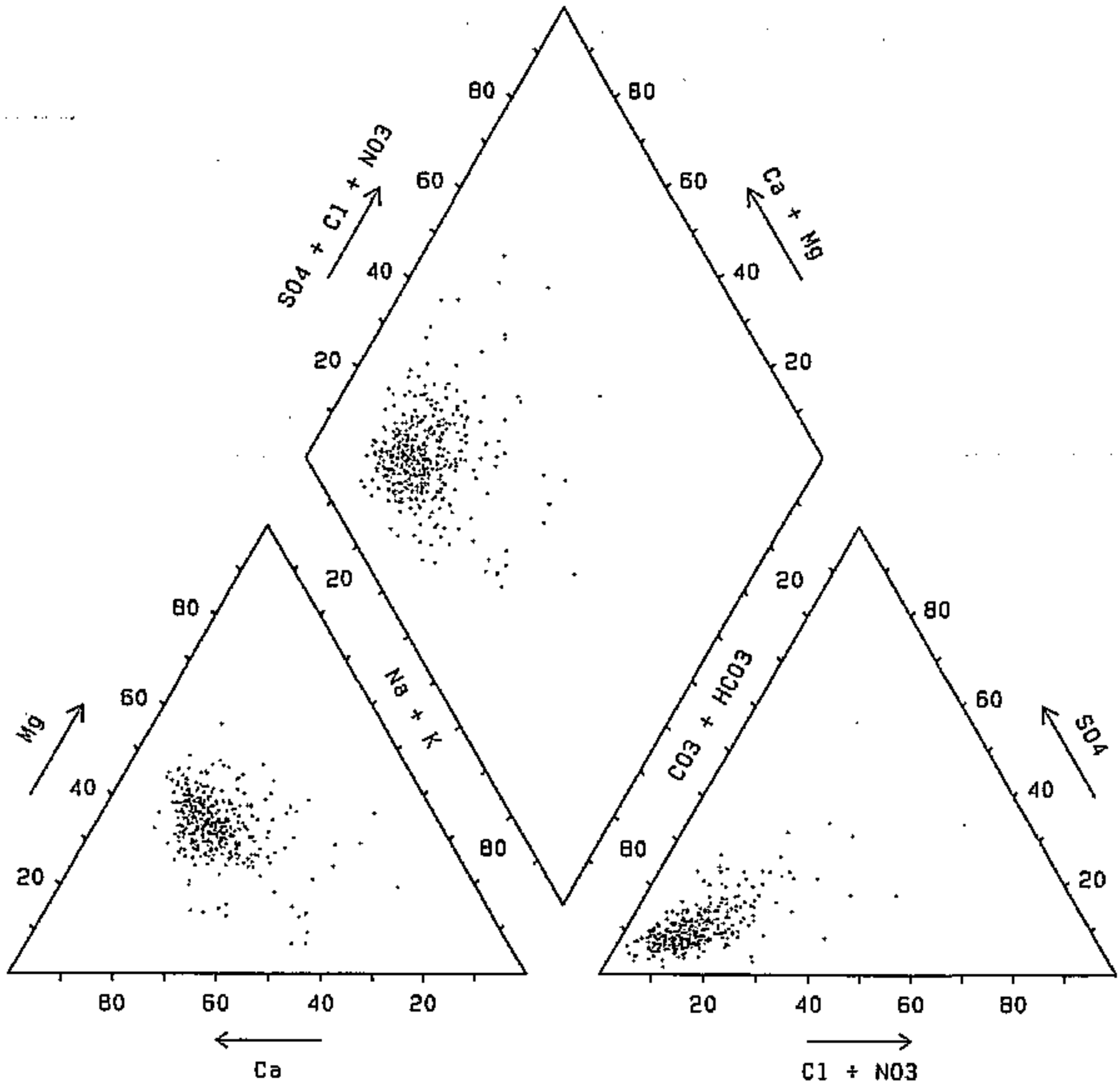


Figure 4.5. *Piper diagram.*

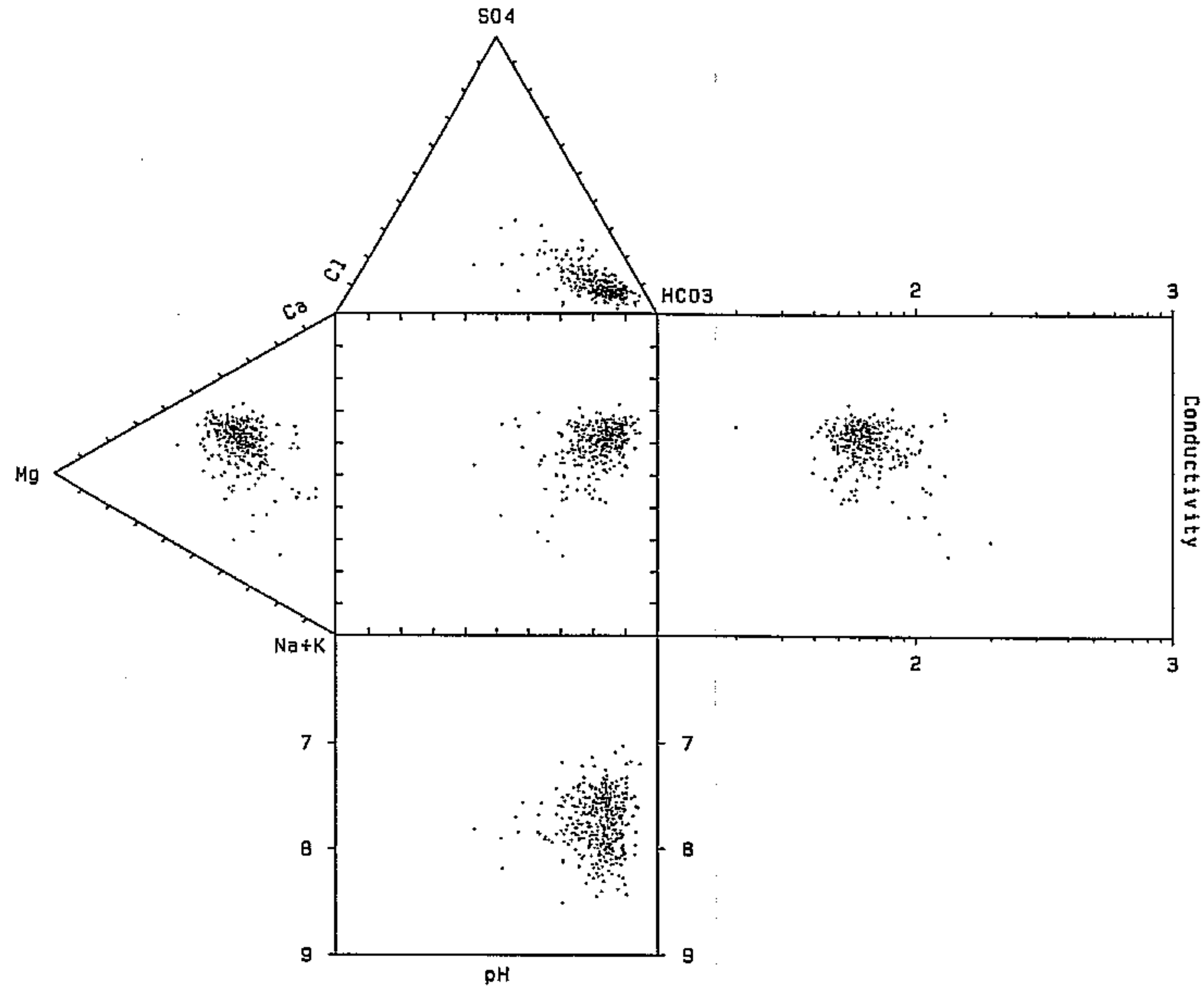


Figure 4.6. *Durov diagram.*

1982). Three-component mixtures take place within a triangle of which the corners represent the plot positions of the respective components.

Schoeller diagram

The Schoeller diagram is a widely used vertical scale diagram which assists in identifying different types of ground water. Ion concentrations of a number of different analyses can be displayed on one diagram. The concentrations are plotted either in mval/l (or in mg/l with off-set logarithmic scales) in the sequence Ca^{++}, Mg^{++}, Na^{+}, Cl^{-}, SO_4^{-} and HCO_3^{-}. Points on the respective vertical scales which belong to the same analysis, are connected by straight lines. Parallel lines between two axes indicate the same ionic ratios at different concentrations (Matthess, 1982). Figure 4.7 is an example of such a diagram.

Only a limited number of analyses may be plotted on a diagram, to be able to distinguish between the respective analyses. It has, on the other hand, the advantage over trilinear diagrams that actual concentrations are shown. Furthermore, Schoeller diagrams can easily be adopted to accommodate other combinations of parameters, for example pH or some of the heavy metals.

SAR diagram

The sodium adsorption ratio diagram is used to assess the suitability of water for irrigation purposes. Irrigation water with a high sodium content can replace the exchangeable cations, Ca^{++} and Mg^{++}, of clay minerals in the soil. Soils which become saturated with sodium tend to lose their fertility and their permeability. The ability of water to exchange Ca^{++} and Mg^{++} with Na^{+} is expressed by the sodium adsorption ration (Richards, 1954):

$$SAR = \frac{Na^{+}}{\sqrt{\frac{Ca^{++} + Mg^{++}}{2}}}$$

The ion concentrations in the equation are expressed in mval/l. A SAR diagram is shown in Figure 4.8. In the diagram, which is subdivided into 16 different fields, the electrical conductivity is plotted against the sodium adsorption ratio. Each field indicates different salinity and sodium exchange hazards.

Depending on soil types and irrigation practices, the diagram gives an indication which crops can be grown under the circumstances. In irrigation areas where ground-water levels are shallow, evaporation of the irrigated water can lead to an increase in the salt

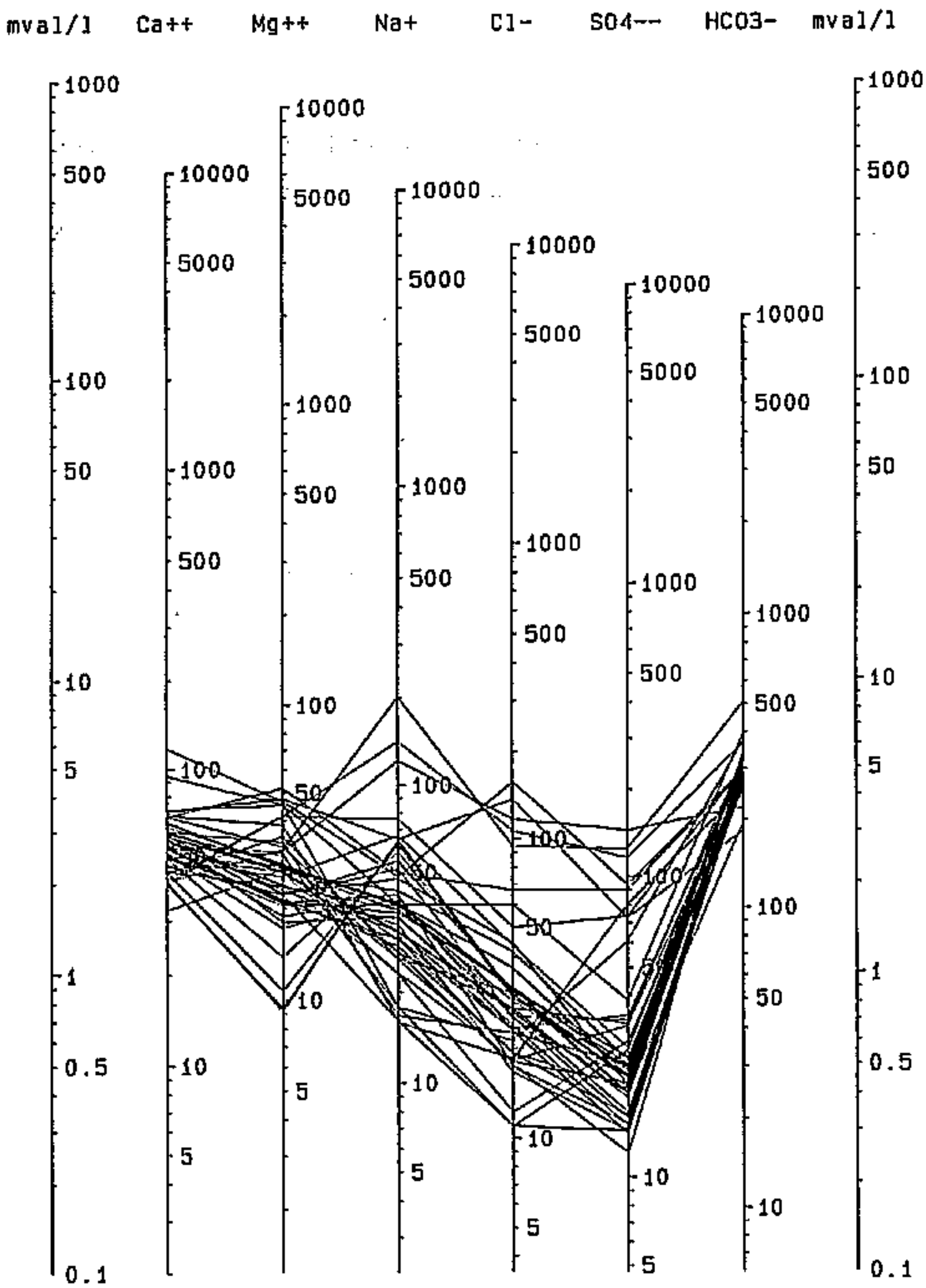


Figure 4.7. *Schoeller diagram.*

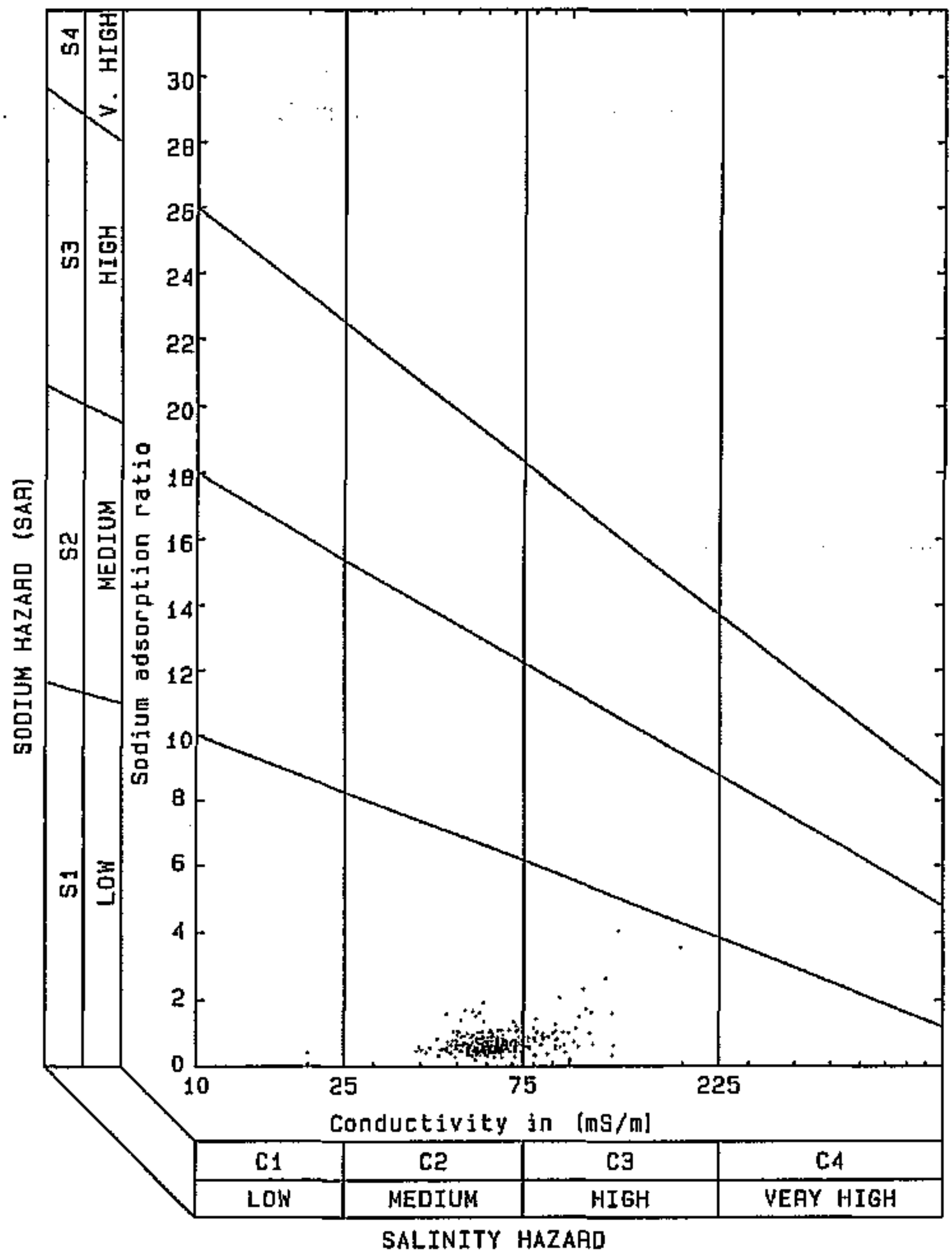


Figure 4.8. *SAR diagram.*

concentration of the ground water. If irrigation water is obtained from the aquifer itself, recirculation of the water takes place and the salt concentration in the water may become too high for further irrigational use.

Apart from the hydrochemical relationship diagrams discussed, hydrochemistry can also be represented by other graphical techniques, such as X-Y plots, hydrochemical maps and cross-sections. These techniques will be discussed later in this chapter.

4.3.2 Pumping test analysis

For the correct assessment of the available ground-water resources, the knowledge of the hydraulic properties of the aquifer is essential. Pumping tests are the most commonly used means to determine the hydraulic characteristics of aquifers. A number of existing pumping test evaluation techniques have been computerized. They are listed below:

- (i) Theis analysis.
- (ii) Cooper-Jacob analysis.
- (iii) Theis Recovery analysis.
- (iv) Jacob Step Drawdown test analysis.
- (v) Multirate analysis.

These techniques can be divided into two major groups, namely constant yield test evaluation techniques (i - iii) and variable yield test evaluation techniques (iv - v).

Theis analysis

The Theis analysis is used to evaluate constant yield pumping tests subject to specific conditions of the aquifer, which are listed in the relevant literature (Kruseman and De Ridder, 1979).

The radius of influence of a borehole, pumped at a constant rate, extends laterally with time. Due to the fact that the water is released from storage within the aquifer, the water level declines. Therefore, the decrease in water level, multiplied by the storage coefficient and summed over the area of influence, equals the discharge. The rate of decline decreases as the area of influence increases, eventually becoming negligible (Kruseman and De Ridder, 1979). This whole relation is contained in the following equation by Theis:

$$s = \frac{Q}{4\pi kD} \int_{u}^{\infty} e^{-y} \frac{dy}{y} = \frac{Q}{4\pi kD} W(u)$$

where

$u = \frac{r^2S}{4kDt}$

s = the drawdown (m) in an observation well a distance r (m) from the pumped well

Q = the discharge rate of the pumped well (m^3/day)

S = storage coefficient

kD = transmissivity of the aquifer (m^2/day)

t = the time since the pumping test has started (days)

$W(u) = \int_u^\infty e^{-y}\frac{dy}{y}$ (Theis' well function)

approximated by $0,5772 - \ln(u) + u - u^2/2.2! + u^3/3.3! - u^4/4.4! + \ldots$

Using the above equation, Theis devised a type curve graphical method to evaluate a constant yield pumping test. A master curve is constructed by plotting values of W(u) against values of 1/u on logarithmic paper. Values of drawdown (s) are plotted against time-distance (t/r^2) values on the same scale logarithmic paper. The time-drawdown curve is superposed on the master curve and the W(u) and 1/u values of a chosen match point are noted (see figure 4.9). These values are inserted in the Theis equation for calculation of the hydraulic parameters S and kD.

Cooper-Jacob analysis

This method is based on the Theis equation and thus automatically subjected to the same conditions, plus additional restrictions as listed in Kruseman and De Ridder (1979). It differs from the Theis formula in that the exponential integral is expanded as a convergent series, resulting in the following relation:

$$s = Q / 4\pi kD\,(-0,5772 - \ln u + u - u^2/2.2! + u^3/3.3! \ldots)$$

For small values of u (u < 0,01), the drawdown can be expressed by the asymptote (Kruseman and De Ridder, 1979), reducing the above equation to a linear relation, expressed as:

$$s = \frac{2.30Q}{4\pi kD}\log\frac{2.25kDt}{r^2S}$$

Symbols used in both expressions are the same ones as those used in the Theis equation. Thus, the drawdown (s) can be plotted against the logarithm of time. A straight line is

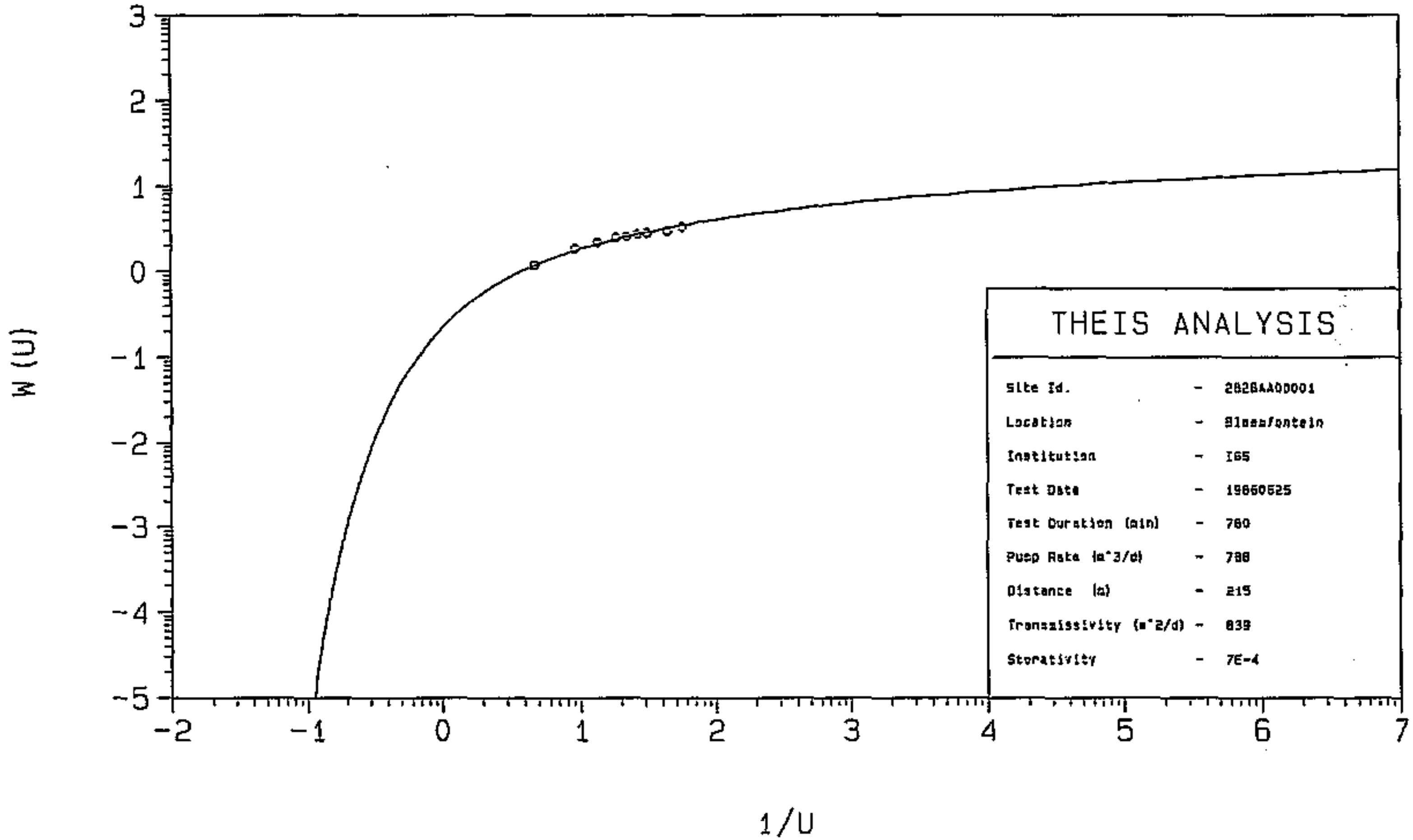


Figure 4.9. *Theis analysis.*

fitted through the data points and the transmissivity kD is calculated by using the equation

$$kD = \frac{2.3Q}{4\pi\Delta s}$$

where Δs = the drawdown per log cycle.

Extending this line to intercept the time axis (shown in figure 4.10), where $s = 0$, the (t_0) value is substituted in the deduced relation

$$S = \frac{2.25\,kDt_0}{r^2}$$

to obtain the storage coefficient (S).

Theis recovery analysis

After completion of a constant yield test, it is equally important to observe the water-level recovery of the pumped well itself and of observation wells in the vicinity thereof. The analysis, according to this method, is subjected to the same conditions as the Cooper-Jacob method. A residual drawdown s" is measured during the recovery period and can be expressed as (Kruseman and De Ridder, 1979):

$$s'' = \frac{Q}{4\pi kD}\left(\ln\frac{4kDt}{r^2S} - \ln\frac{4kDt''}{r^2S''}\right)$$

where

s"	=	residual drawdown (m)
r	=	distance (m) from the pumped well
S"	=	storage coefficient during the recovery period
S	=	storage coefficient during the pumping test
t	=	time since pumping started (d)
t"	=	time since pumping stopped (d)
Q	=	discharge rate (m^3/d)

Assuming S and S" are constant and equal and u is very small, the above expression reduces to

$$s'' = \frac{2.30Q}{4\pi kD}\log\frac{t}{t''}$$

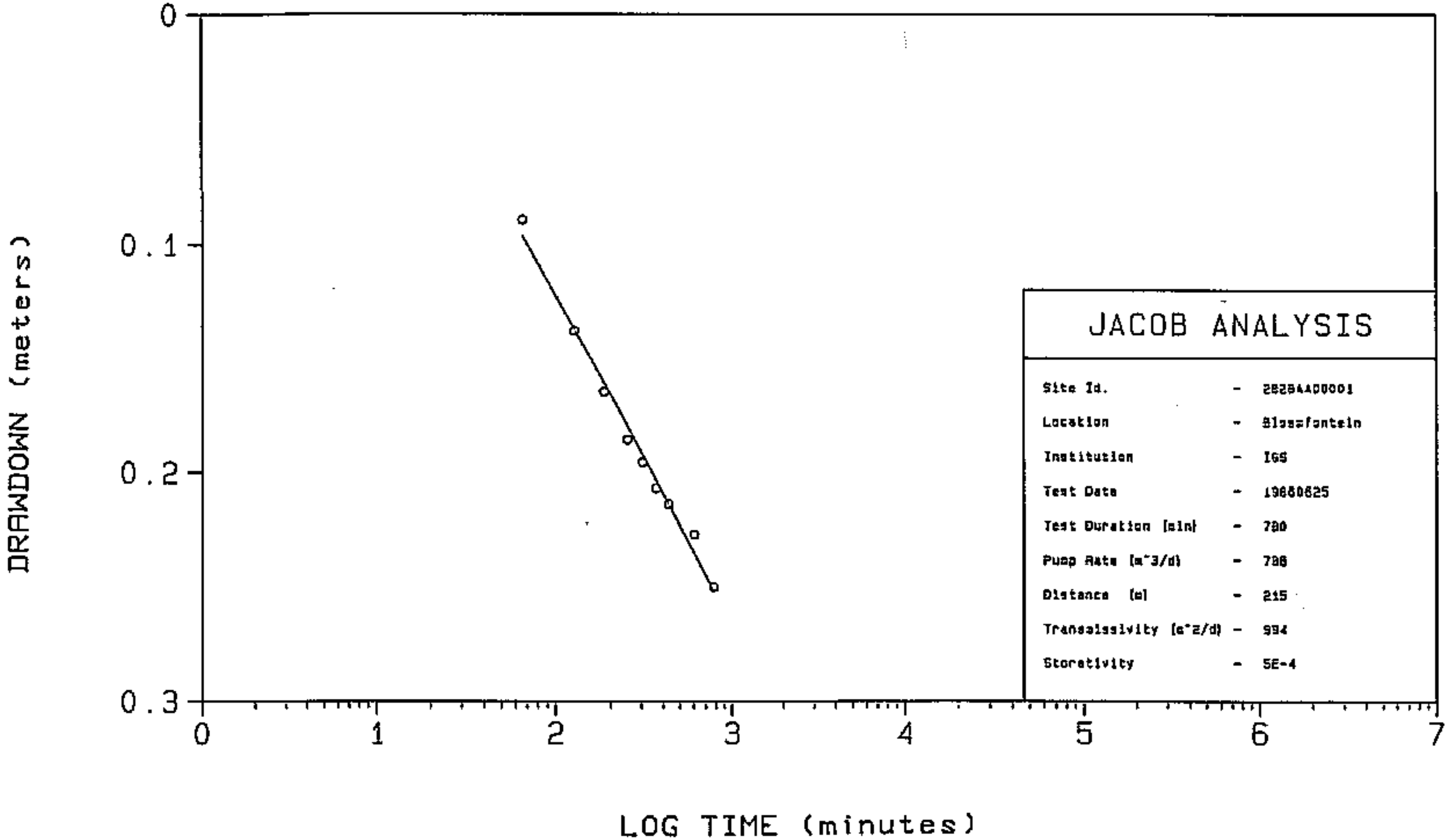


Figure 4.10. *Cooper-Jacob analysis.*

Similar to the Cooper-Jacob analysis, the data are plotted on semi-log paper, i.e. the residual drawdown (s") is plotted against the logarithm of the time ratio (t/t") (see figure 4.11). The slope per log cycle is determined and inserted in the following mathematical expression to calculate the transmissivity:

$$T = \frac{2.30Q}{4\pi\Delta s''}.$$

All the methods discussed are part of the constant yield evaluation techniques and are widely applied to observation boreholes.

Jacob Step Drawdown analysis

The step drawdown analysis resorts under the second group, which comprises variable yield pumping tests. These tests offer the possibility to establish the efficiency of a specific production well for various discharge rates and to determine the optimal discharge rate for that well.

Drawdown in a production well can be divided into two components, namely:

(i) Aquifer loss.
(ii) Well loss.

The resistance to laminar flow within the aquifer, caused by matrix properties of the geological formations, is called aquifer loss. Well loss, on the other hand, results from the resistance due to turbulent flow around the well and is mainly caused by the construction of the hole.

In 1947, Jacob postulated that the well loss in a production well is proportional to the square of the discharge rate and the relationship between the two components of drawdown is:

$$s_w = BQ + CQ^2$$

where

s_w = drawdown in the pumped well (m^3/d)
BQ = aquifer loss
CQ^2 = well loss

He devised a pumping test method determining the well loss, as well as the effective radius of the pumped well. In such a test, the drawdown in the production well is observed, while the discharge rate is increased in steps. The basic assumption in this

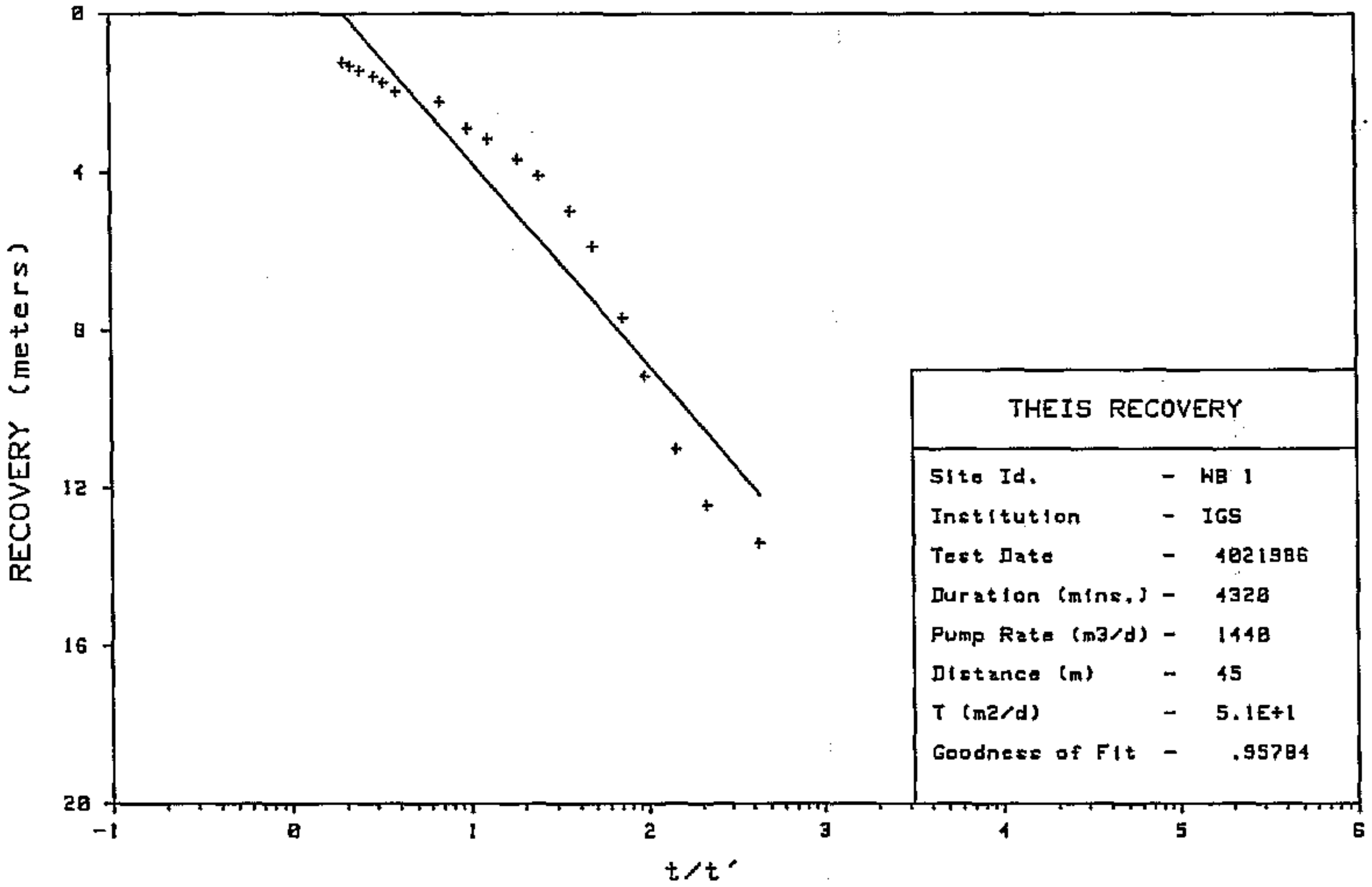


Figure 4.11. *Theis recovery analysis.*

method is that the increase in discharge rate for each step is equivalent to an additional pump, starting to pump in the specific well (Clark, 1977). The drawdown (s_w) is plotted against the logarithm of the time, as shown in figure 4.12.

The analysis of step tests provides the discharge rate vs. drawdown relationship of the pumped well over the range of yields tested. Once the aquifer- and well losses have been calculated, the values can be used to determine the well efficiency for a specific discharge rate. The well efficiency is a measure of the effectiveness of the well screen or the wall of the borehole as a transmitting medium between the aquifer and the hole. The well efficiency (E) and/or the development of the hole are quantitatively expressed as:

$$E = \frac{100\,BQ}{BQ + CQ^2} \quad \text{or} \quad E = \frac{100}{1 + \frac{CQ}{B}}$$

As can be seen in the above equation, the well efficiency is not a constant, but varies inversely with discharge, i.e. efficiency is maximum for low discharges and minimum for high discharges (Roscoe Moss Company, 1982).

Well efficiency calculations are useful for comparisons between wells of similar depths in the same well field (Clark, 1977). They must not be used to compare boreholes under different geological conditions.

Multirate pumping test analysis

Multirate pumping tests differ from the step drawdown tests in that the water level is left to recover after each step. The basic evaluation technique is similar to the Jacob Step Drawdown analysis. Figure 4.13 is a graphical representation of such a test and the parameters calculated.

The multirate pumping tests conclude the explanation of the most often used pumping test evaluation techniques that have been computerized. When the determined aquifer parameters are used to calculate ground-water reserves or the ground-water balance for a certain area, it must be realized that the geohydrological conditions in the area will, to a larger or lesser degree, deviate from the conditions on which the evaluation techniques are based on. The behaviour of fractured aquifers, common in South Africa, differs from that of porous aquifers on which the evaluation techniques are based. Double porosity and delayed yield are other factors which may cause even further deviations of the calculated parameters from the actual or representative values. The programs developed, free the geohydrologist from the burden of standard plots and calculations. His skills in the interpretation of the results are still essential.

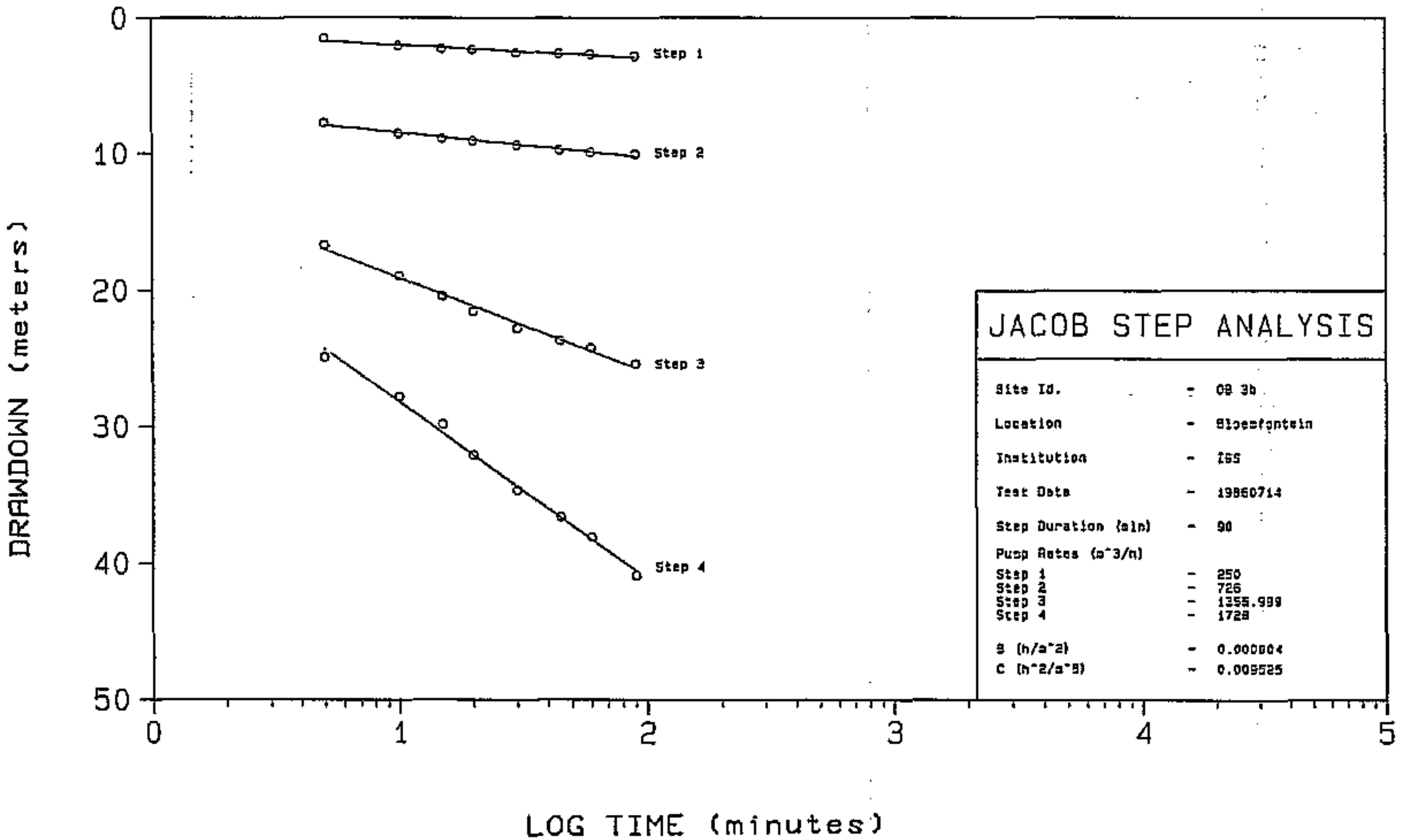


Figure 4.12. *Jacob Step Drawdown analysis.*

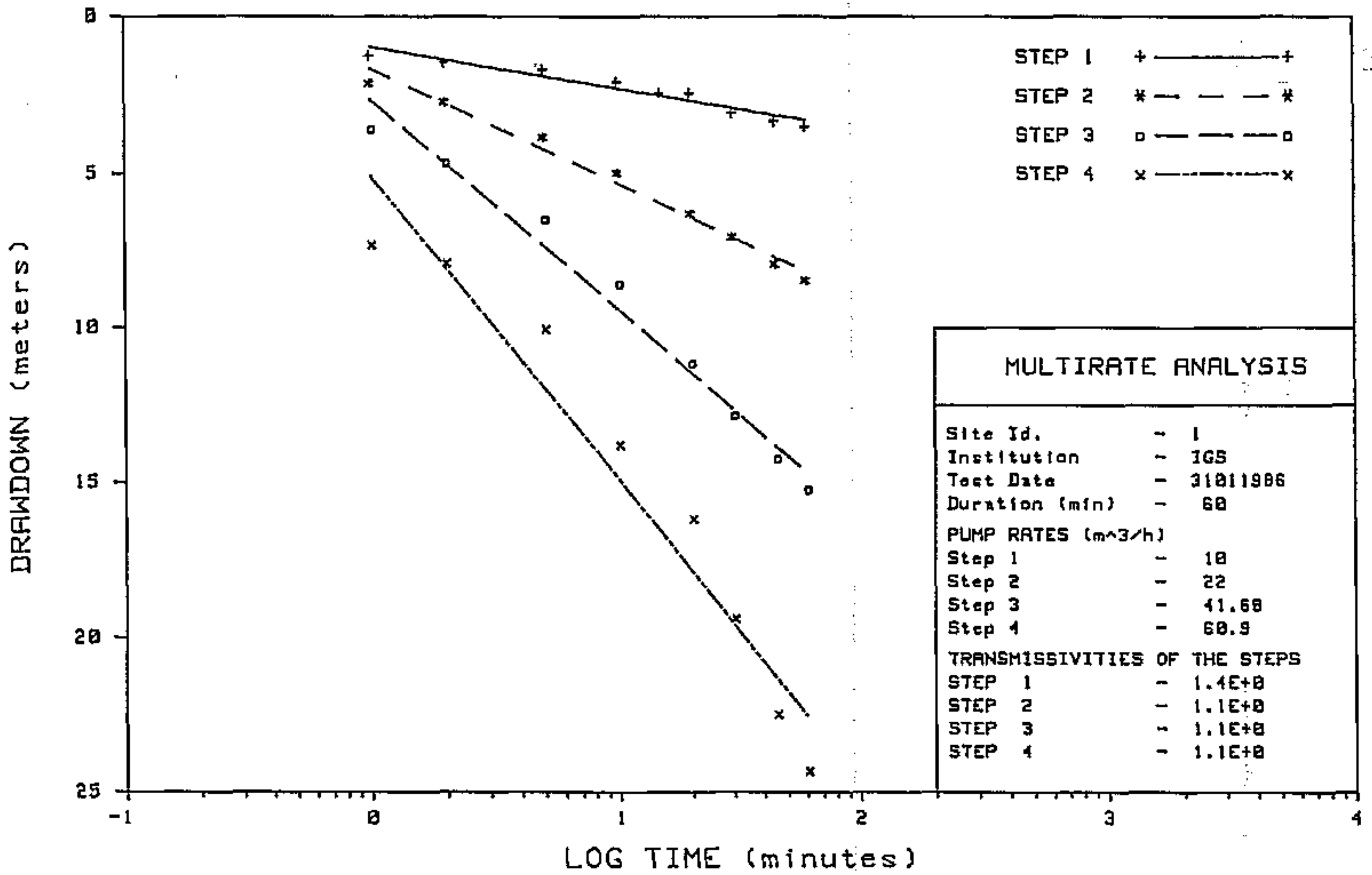


Figure 4.13. Multirate pumping test analysis.

4.3.3 Geological analysis

The geological information about aquifers, as well as the non-water-bearing formations, forms an important part of the borehole information, accommodated in the National Ground-water Data Base, because the aquifer characteristics are mainly dependent on the geological properties of the respective formations. Geological data are stored in either of two data sets:

(i) Unconsolidated geology.
(ii) Consolidated geology.

Unconsolidated geology makes provision for a more detailed description of the strata. Both sets, and the graphical techniques resorting under them, will be discussed hereunder.

Consolidated geology accommodates the basic geological data and allows a short description of especially hard-rock formations.

Additional parameters provided for unconsolidated geology relate especially to unconsolidated formations. The list below contains the parameters provided for the entry of geological data. Items (v) - (viii) refer to the unconsolidated geology only:

(i) Depth to top of formation.
(ii) Lithology.
(iii) Primary and secondary colour.
(iv) Texture.
(v) Primary and secondary features, as well as a feature attribute.
(vi) Sorting.
(vii) Roundness.
(viii) Sieve analysis.

With the exception of sieve analyses, the same basic graphical evaluation techniques are used to display the geological characteristics for both consolidated and unconsolidated geology.

A number of graphical evaluation methods are provided in the processing package:

(i) Borehole logs.
(ii) Borehole chart.
(iii) Borehole profile.
(iv) Sieve analysis.

The following symbols, ∇ and $\triangleleft$, are encountered in the borehole logs and charts. Respectively, they indicate the water level and depth at which water was struck in the borehole.

Borehole logs

This graph displays a stratigraphical column of the strata encountered, together with the uncoded description thereof. The diagram is shown in figure 4.14.

The water level and the depths at which water has been struck are also shown, so that the hydrogeologist obtains a visual impression of the prevailing geohydrological conditions and allows him to identify the water-bearing formations (aquifers).

Borehole chart

This display is based on the previous diagram. However, it contains additional features, such as information on the penetration rate, diameter drilled and diameter of casing in the hole. Figure 4.15 is a graphical representation of the diagram.

The second diagram is especially useful for the hydrogeologist, engineer, operator or driller, because they summarize and display the more important geological, geohydrological and technical features of the borehole.

Borehole profile

It is often important to obtain a geological cross-section of the investigated area. Such a cross-section is provided in the borehole profile diagram. A maximum of 15 boreholes may be used to construct the diagram. If the respective boreholes have been logged in sufficient detail, it is possible to correlate various lithological units with one another by interconnecting the boundaries between them. Figure 4.16 depicts such a diagram.

The boreholes used in the profile have only a vertical scale. They are equally spaced along the horizontal axis. It is necessary that the boreholes selected for the diagram lie more or less along a straight line.

Sieve analysis

When a well is designed in unconsolidated formations, a sieve analysis is important, because grain size and its distribution are needed for the design of an efficient production well. In these formations, screens or perforated casings are used to retain the surrounding sediments and to ensure optimal flow into the borehole. The size of the screen openings or the perforations in the casing are selected according to the grain-size

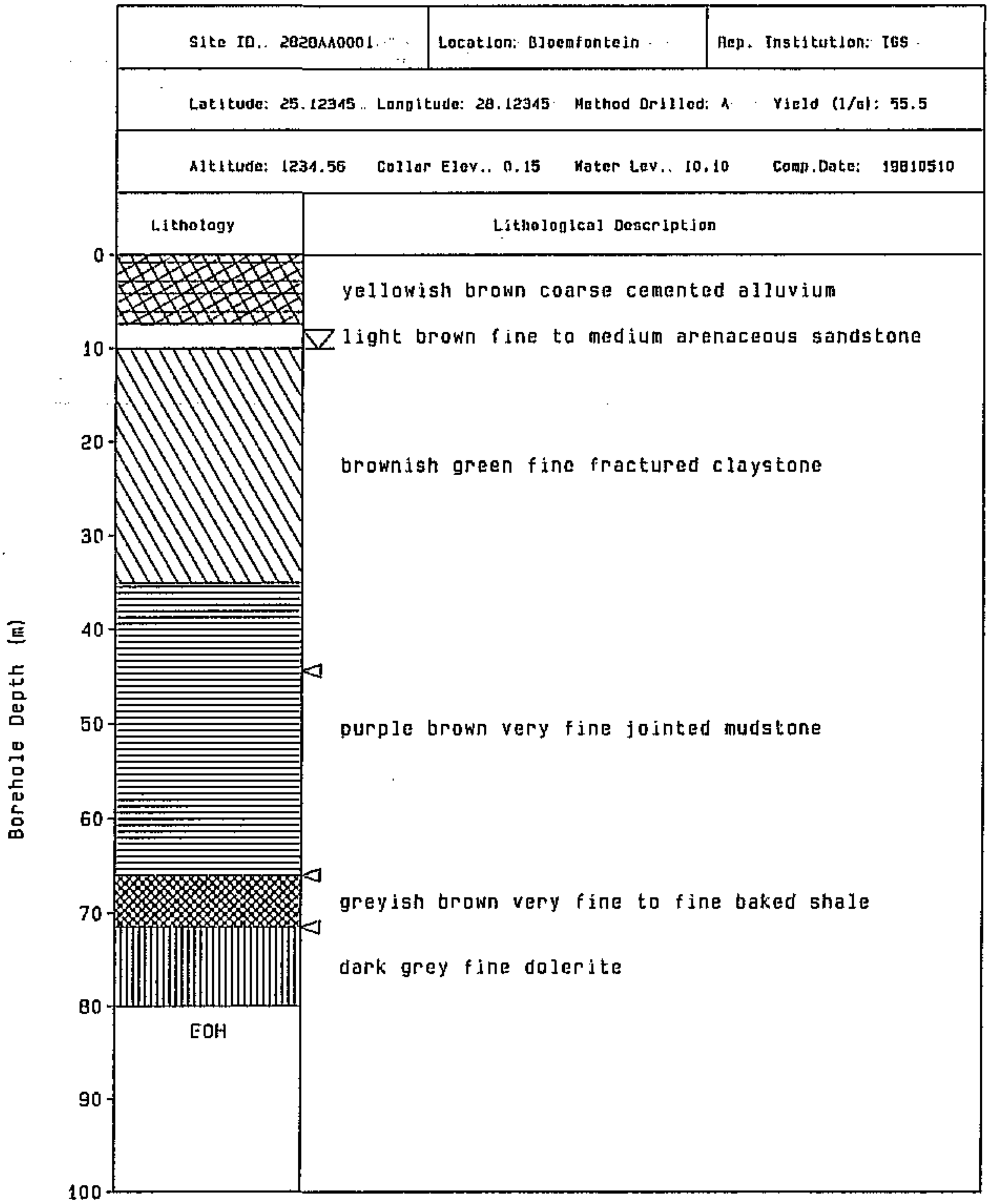


Figure 4.14. *Graphical representation of a geological borehole log.*

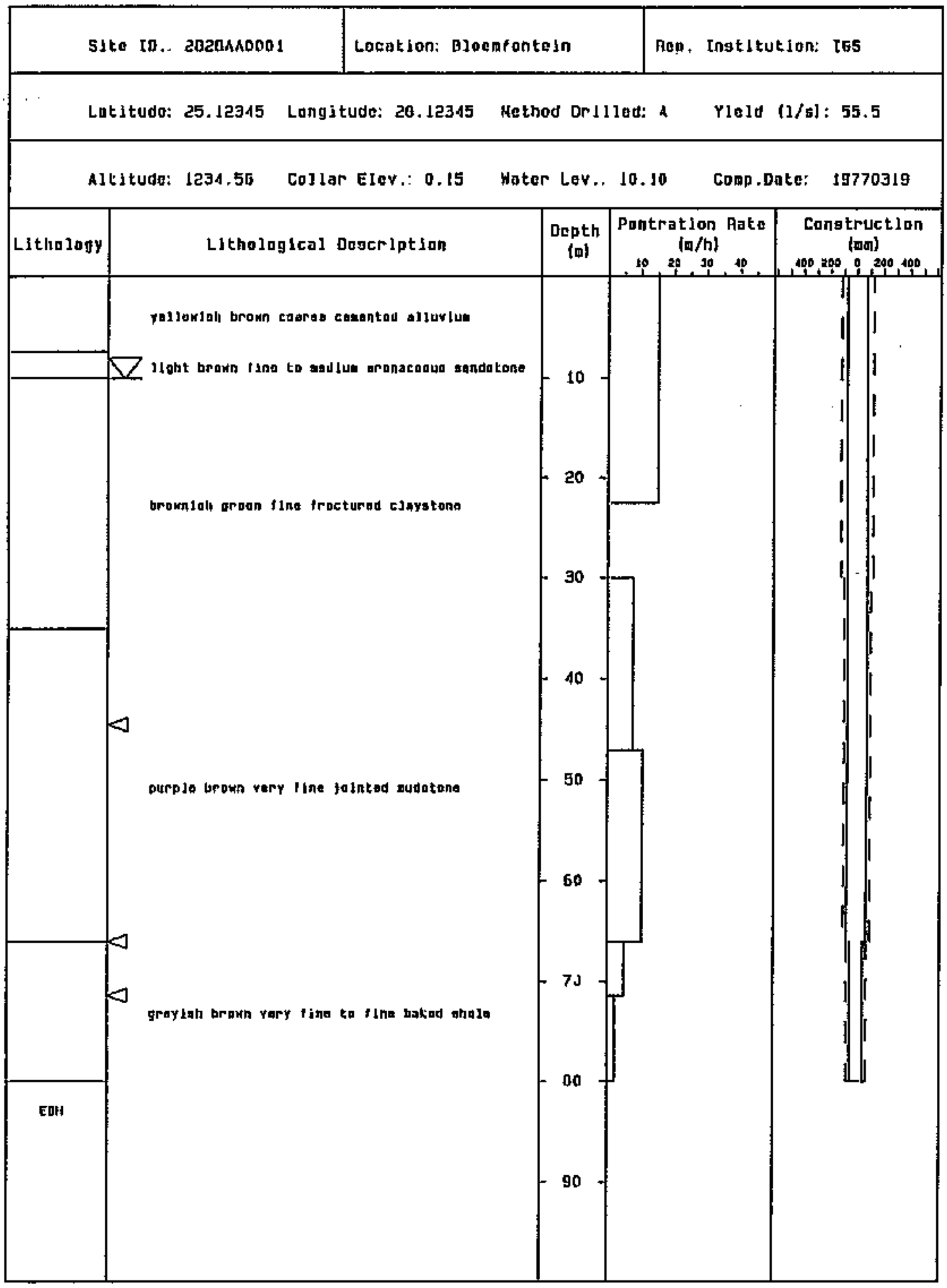


Figure 4.15. Graphical representation of a borehole chart.

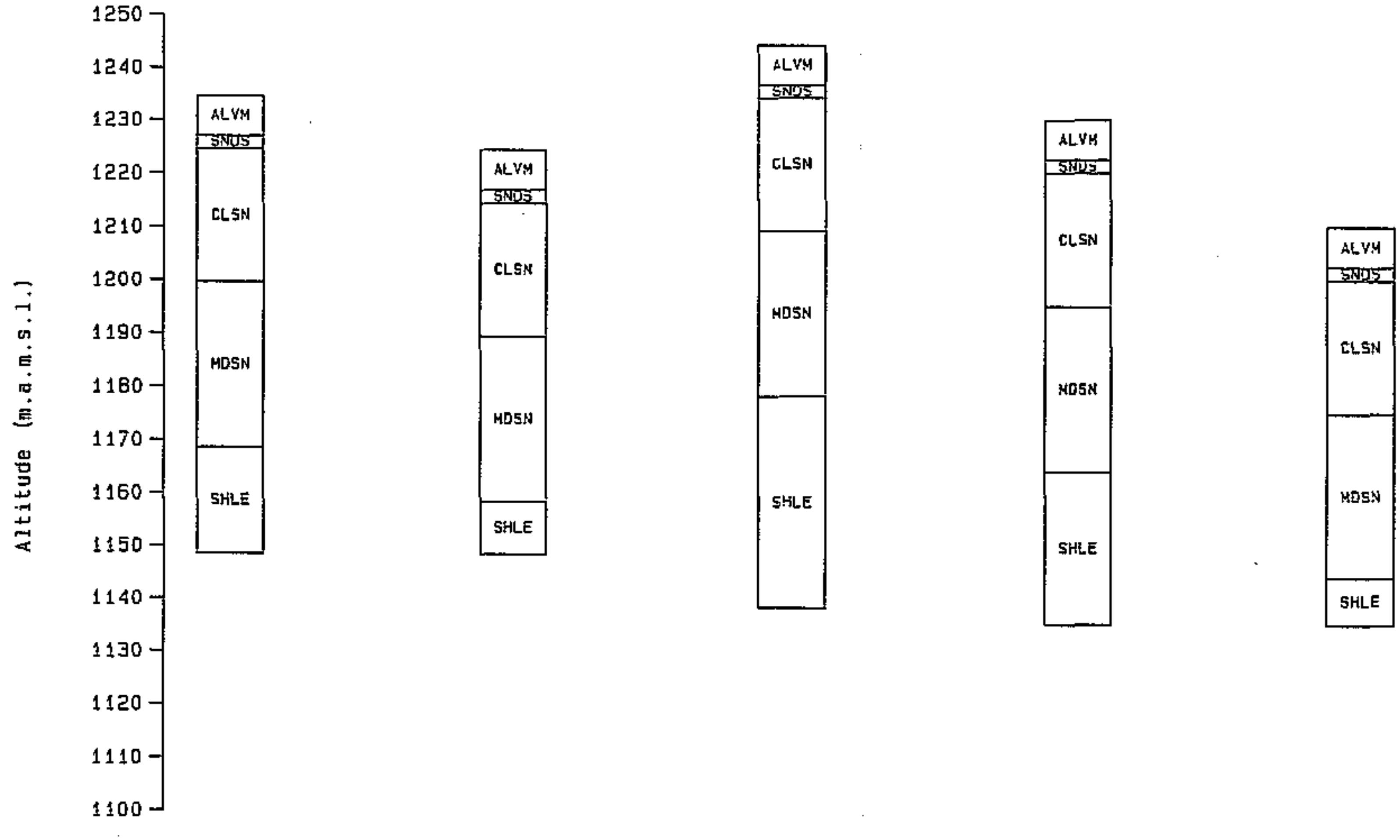


Figure 4.16. *Graphical representation of a borehole profile or cross-section.*

distribution of the surrounding unconsolidated medium. The depth(s) at which the screens are placed also depends on the grain-size analysis. Good, economic well design is needed for low-friction, undisturbed entrance of ground water into the borehole without excessive head loss.

According to Johnson (1975), for a homogeneous formation that consists of fine sand, the screen opening size is selected to retain from 40 to 50% of the sand matrix. Generally, the 40% size is chosen when the ground water is not particularly corrosive and when there is little doubt as to the reliability of the sieve analysis of the sample.

During the drilling process of a borehole in loose formations, drilling samples must be taken frequently. The samples are dried and sifted through a series of sieves. The retained fraction for a each sieve is weighed and expressed as a percentage of the total sample weight. A grain-size distribution graph is constructed by plotting the cumulative percentages of each retained fraction against the logarithm of the sieve openings (see figure 4.17).

The sieve openings used in the National Ground-water Data Base are given in millimeters. To express these sizes logarithmically, the following relation is used:

$$\phi = -\log_2 d$$

where

ϕ	=	the logarithmic value of the sieve opening or particle size
d	=	the sieve opening or particle diameter in millimeters.

Various coefficients can be deduced from the graph (Friedman, 1978) such as:

(i) Graphic mean = $(\phi_{16} + \phi_{50} + \phi_{84}) / 3$

(ii) Inclusive graphic standard deviation = $\dfrac{(\phi_{84} - \phi_{16})}{4} + \dfrac{(\phi_{95} - \phi_5)}{6.6}$

(iii) Simple sorting measure = $0{,}5\ (\phi_{95} - \phi_5)$

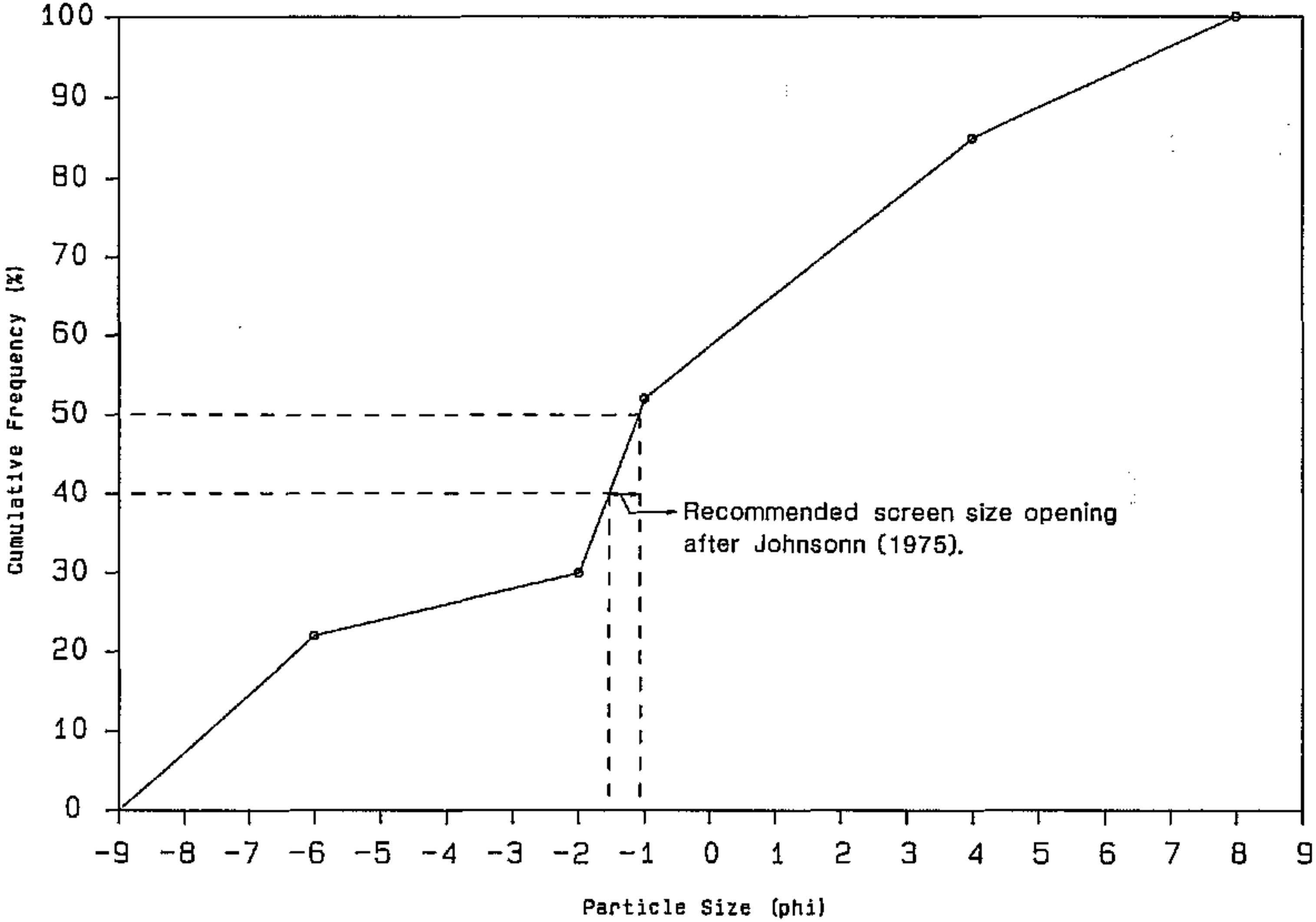
(iv) Inclusive graphic skewness

$$= \frac{(\phi_{84} + \phi_{16} + 2\ \phi_{50})}{2\ (\phi_{84} - \phi_{16})} + \frac{(\phi_{95} + \phi_5 - 2\ \phi_{50})}{2\ (\phi_{95} - \phi_5)}$$

(v) Simple skewness measure = $(\phi_{95} + \phi_5) - 2\phi_{50}$

(vi) Graphic kurtosis = $\dfrac{(\phi_{95} - \phi_5)}{2.44\ (\phi_{75} - \phi_{25})}$

where ϕ_n denotes the n-th percentile of the fraction.



Graphic Mean = 3.545

Simple Sorting Measure = 13.348

Simple Skewness Measure = 22.152

Inclusive Graphic Standard Deviation =0.228

Inclusive Graphic Skewness = 0.475

Graphic Kurtosis = -9.628

Figure 4.17. *Grain-size distribution graph.*

4.3.4 Map and contouring facilities

One of the main advantages of geohydrological and contour maps is the fact that they portray geohydrological conditions, which are not necessarily apparent from any other graphical technique nor from a data print-out. They are, therefore, especially useful to the hydrogeologist. This group contains two sections:

(i) Map facilities.
(ii) Contour facilities.

Map facilities

Maps are generally used to depict features, such as topography, boundaries of geological formations, borehole positions, farms and drainage regions of a specific area.

Before any map of a specific area can be constructed, coordinates of all the features to be shown on the map must be known. If not, these features must first be digitized. This can be done with a specific computer program, supported by PROCAD. Once they have been obtained, a map can be constructed at the desired scale (see figure 4.18(a)).

Contour facilities

Contours can be subdivided into two groups:

(i) Two-dimensional contours.
(ii) Three-dimensional display.

The latter serves a more illustrative purpose, which cannot be used for quantitative interpretation purposes.

Two-dimensional contouring

These contours show the areal distribution of any numerical parameter in a two-dimensional plane. They can then be used for both quantitative and qualitative interpretations. A special contouring package (PLOT 88) is included in the processing package.

The contouring program is based on a two-dimensional Cartesian coordinate system, meaning that the x- and y-coordinates which are real numbers in a world coordinate system, are transformed to device coordinates, based on a selected scale factor.

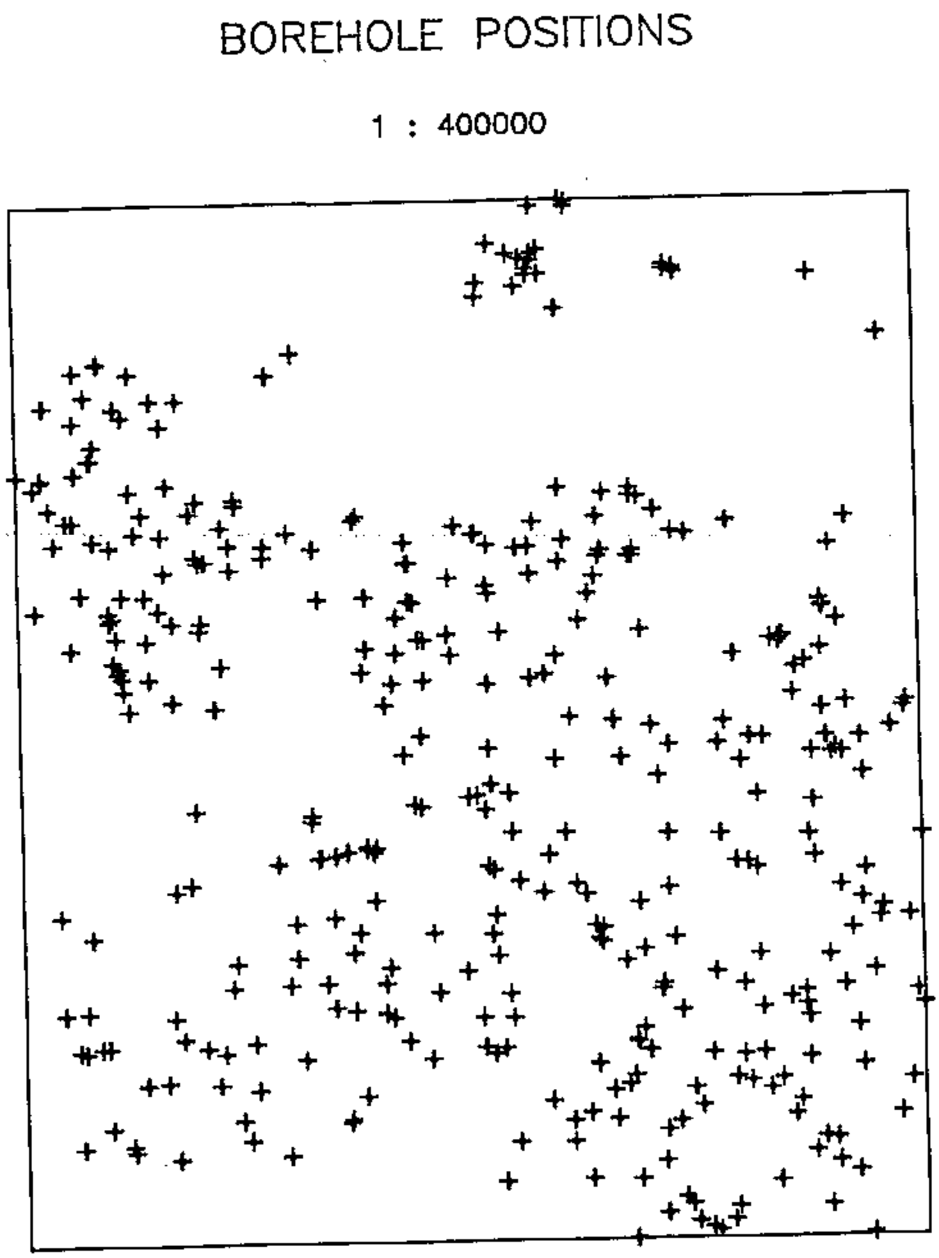


Figure 4.18(a) Two-dimensional areal map.

When constructing a two-dimensional contour map of a specific area, certain constraints have to be met to ensure proper results. Firstly, a variable grid matrix is constructed over the surface. The size of the grid depends on the distribution and number of observation points, spread over the area. At each node of the grid, a z-coordinate is calculated by using either Laplace- or cubical spline interpolation or a combination of the two, which is represented by the following equation:

$$(\nabla^2 x(z) + \nabla^2 y(z)) - (cay\,(\nabla' x(z) + \nabla' y(z)) = 0$$

The first part of the equation comprises the Laplace interpolation and the second part the cubical spline interpolation. Depending on the value of cay (see PLOT 88 manual), which may vary between 0,0 and 10^6, either one of the interpolation schemes, or a combination of them, is used. Laplace interpolation tends to produce a rather spiky surface, with sharp peaks and dips. Cubical spline interpolation, on the other hand, produces a smoother surface.

Having completed the interpolation of values for the grid control points, the gridded control points can be further smoothed by applying Laplacian smoothing. Too much smoothing, however, can filter out important relevant features, and leave the user with a trend surface only.

Figure 4.18(b) displays a two-dimensional contour map of chloride concentration, found in ground-water samples from the southern Orange Free State.

The processing package contains a second contouring program, which is used mainly in conjunction with the statistical data processing package. This program produces contour maps on a line printer. The contour intervals are represented by different characters (see figure 4.18(c)). Although the finished maps are crude in appearance, compared to those drawn by plotters, they are inexpensive and adequate for many purposes.

Three-dimensional display

This facility is mainly used for qualitative illustration. The procedure to be followed for the construction of a three-dimensional display is similar to the one used in two-dimensional contouring. The main difference lies in the last part of the procedure, where the grid control points are transformed into a three-dimensional surface, shown in figure 4.18(d).

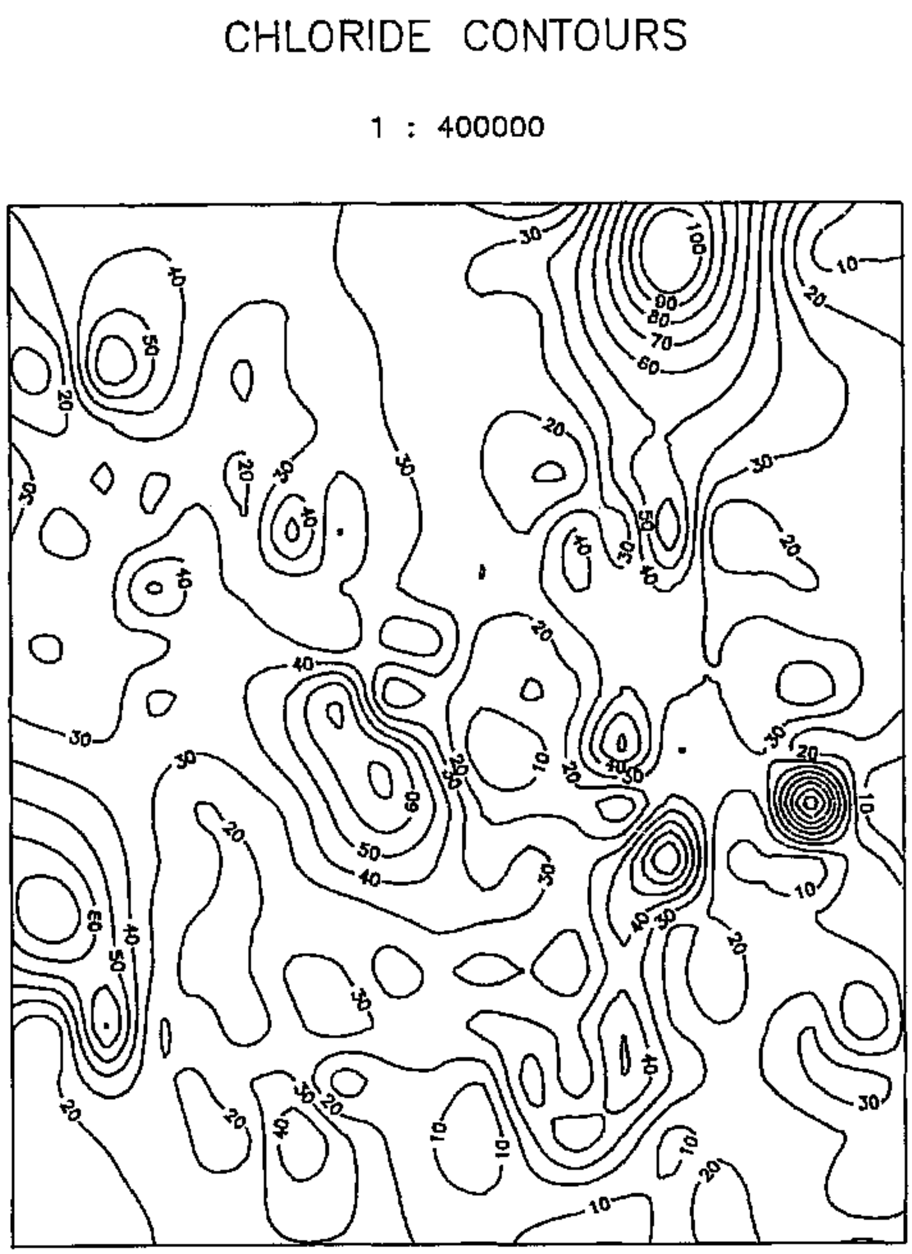


Figure 4.18(b) Two-dimensional contour map, produced by a plotter.

```
CONTOUR PROGRAM
===============

   MATRIX OF GRID VALUES


             1111     1     3  33333333333333333      33333
                          33333333333333 33  3333    33333
                 3    333333333333          333333333333
               333333333         5         33333333333333333
               3333333     55555555 555    33333333333333333
              333333     55555555555555    333333   33333333
             333333  5555555     55 55555          333333333
 33333333333333    555555 77    777      55555     33333333
  333     333     55     777777777777     55555   33333333
33333     333    55   777777 7  77777777   55555 333 333333
3333333333333  555    77777       777 7777    555 3333333
3333333333    555  777777        9      777    5     333333
33333333     55   777     9999999999999      77 5     3333
333333333   5  7 7777    99999999999999     9 7  5    333  1
33        555  77777     999999999999999      7  55   3333
3333  5555557777777    999999999999999999     77 5555  33
3     5555   77777    999999999999999999     777  555  33333
33   55      7777    9999999999999999999     77   555 333333
33   555     77777    99 999999999        77777   555  33333
33    55555 7777777                      77777   5555  33333
333   555555 77777777    777    777 77777       55555   3333
 333    55555       77777777777777777777       555      3333
  33      5555       777 777  7 777 77       55555      3333
33333      5555555    5                   555555      333333
33333333   3       55555555555555555555555        333333333
333333333333               5555555555555        333333333333
3333333  33333333     3            5   5    33  333   333333
```

```
LEGEND OF SYMBOLS
-----------------

SYMBOL - 1   =       .4     TO        2.7
SYMBOL -     =      2.7     TO        4.9
SYMBOL - 3   =      4.9     TO        7.2
SYMBOL -     =      7.2     TO        9.4
SYMBOL - 5   =      9.4     TO       11.7
SYMBOL -     =     11.7     TO       14.0
SYMBOL - 7   =     14.0     TO       16.2
SYMBOL -     =     16.2     TO       18.5
SYMBOL - 9   =     18.5     TO       20.7
```

Figure 4.18(c). Two-dimensional contour map, produced by a printer.

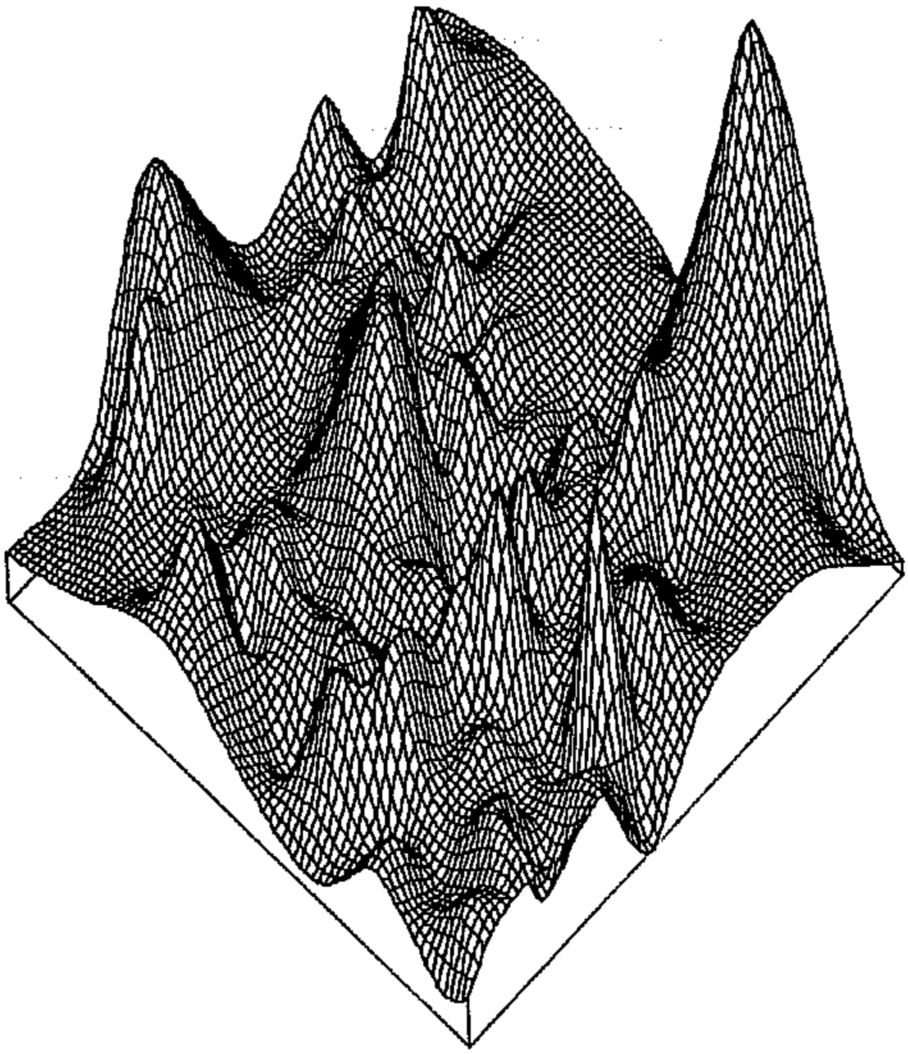


Figure 4.18(d). *Three-dimensional contour map of chloride concentrations.*

4.3.5 General graphics

Of all the graphical techniques, general graphics are used frequently to represent relationships between various parameters, e.g. water-level graphs, and need no further explanation. The different techniques supported are listed below:

(i) XY linear graphs.
(ii) XY log-linear graphs.
(iii) XY log-log graphs.
(iv) Bar charts.
(v) Stacked bar charts.
(vi) Pie charts.
(vii) Histograms.

Figure 4.19 shows examples of some of the above-mentioned diagrams.

4.4 STATISTICS

Due to the fact that geohydrologists depend generally on relatively few observations, which are made at different times and/or places and are subject to a varying measure of uncertainty, numerous statistical analysis techniques have been developed to process these data.

Statistical data processing comprises a number of disciplines, such as:

(i) General statistics.
(ii) Sequence analysis.
(iii) Map analysis.
(iv) Multivariate analysis.

All four groups will be discussed below:

4.4.1 General statistics

Each population has a specific distribution. The characteristics of each sample in the population indicate a certain spread and symmetry of the distribution. Such characteristics are known as parameters, if they describe populations, or as statistics, if they describe samples. Thus, statistics can be used to estimate parameters of populations

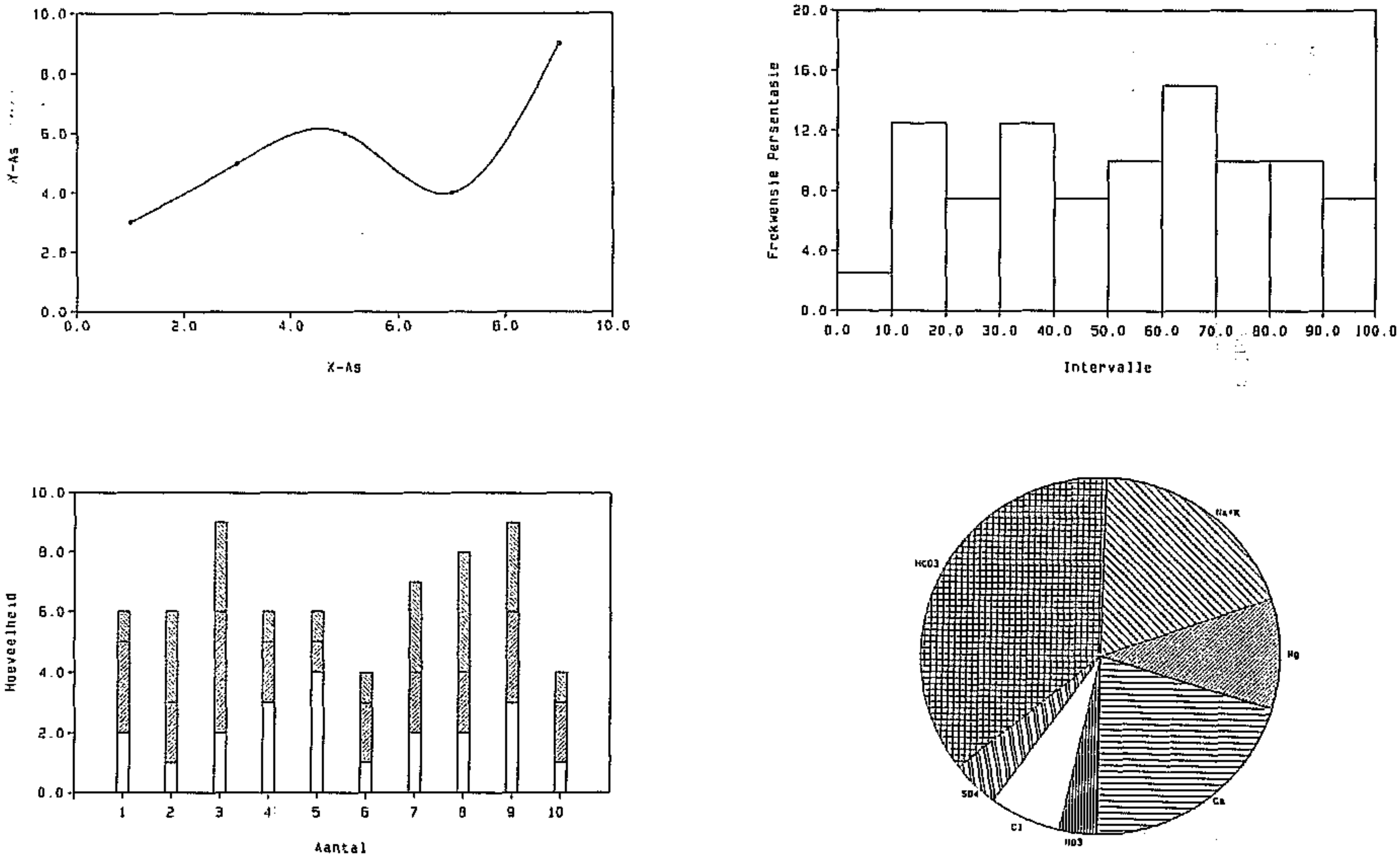


Figure 4.19. *Examples of general graphics displays.*

and to test hypotheses about populations. The statistics under discussion in this section are listed:

(i) Mean calculation.

(ii) Variance and standard deviation calculation.

(iii) Covariance and correlation calculation.

(iv) Analysis of variance.

Mean calculation

This are the most basic and frequently used measure describing the tendency of the sample population. The arithmetic mean of a distribution is expressed as:

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

where

$\overline{x}$ = mean value

x_i = sample value

n = number of samples

The median represents the mid-point of the distribution, while the mode represents the value in the distribution with the greatest frequency.

Variance and standard deviation calculation

Although it is easy to determine the mean of any number of observations, the obtained value may have little significance on its own. The spread or dispersion of observation values about the mean is expressed by two measures, namely the variance and the standard deviation. The variance (S^2) is the average squared deviation of all sample values from the population mean and is expressed by:

$$S^2 = \quad = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}$$

However, to provide a statistic, describing the spread of data around the mean and in the same units as the sampled data, the standard deviation is calculated which is the square root of the variance. A small standard deviation is indicative of a tightly clustered population around the population mean (Van Tonder, 1981). Again, it is important not to attach too much significance to the numerical value of the variance or the standard deviation. The variance or standard deviations of different samples should be compared.

Covariance and correlation calculations

Different variables are often measured simultaneously. This results in a measure of mutual interaction. Such an interaction is described by the covariance, which is the joint variation of two variables about their common mean and is expressed as:

$$COV_{jk} = \frac{1}{n\,(n-1)} \left(n \sum_{i=1}^{n} x_{ij}\, x_{ik} - \sum_{i=1}^{n} x_{ij} \sum_{i=1}^{n} x_{ik} \right)$$

where COV_{jk} = covariance between variable j and k.

Only if the units of the populations or samples considered are identical, then the covariance of different data sets may be compared. The correlation coefficient, on the other hand, is used to determine the degree of interrelation between variables of populations or samples measured in different units. The correlation is defined as the ratio of the covariance of the two variables and the product of their standard deviations. It can be expressed as:

$$r_{ij} = \frac{COV_{ij}}{S_i S_j}$$

Analysis of variance

The previous techniques can only be used to compare two samples with each other. However, many tasks involve numerous groups of observations, for example, ten water samples with different chemical compositions. To determine whether their bicarbonate content is similar, a comparison can be done by the analysis of variance, i.e. separating the total variance in a collection of measurements into various components or sources. The tests of equality operate by simultaneously considering both differences in means and in variances (Davis, 1973).

The equivalence of a number of samples can be determined by one-way analysis of variance through testing the hypothesis and its alternative

$H_0 : \mu_1 = \mu_2 = ... = \mu_n$
H_1 : at least one mean is different

The method is subject to a number of assumptions:

(a) Each group of observations represents a random sample from the respective population(s), e.g. conductivity measurements of a borehole.

(b) Each parent population is normally distributed.

(c) Each population has the same variance.

A formulation was developed for one-way analysis of variance which is listed below:

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F-test
Among groups	SS_A	m-1	MS_A	MS_A/MS_W
Within groups	SS_W	N-m	MS_W	
Total variation	SS_T	N-1		

where

N = number of groups

m = number of samples to be compared

If the value calculated for the F-test lies within a certain distance from the value assumed by the hypothesis, it is accepted. The width of the interval within which the hypothesis is accepted, depends on the significance level and the degrees of freedom.

Thus one-way analysis of variance is suitable to test the hypothesis that various populations, represented by random samples, are identical. However, a one-way model cannot assess the magnitude of the introduced variance, because it is mixed with other sources of variance in the sum of the squares (Davis, 1973). Two-way analysis of variance, on the other hand, enables assessment of the variation and its magnitude. As in the case of one-way analysis of variance, a formulation was developed for the specific technique. It is listed below:

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F-test
Among groups	SS_A	m-1	MS_A	MS_A/MS_e *1)*
Within groups	SS_B	n-1	MS_B	MS_B/MS_e *2)*
Error	SS_e	(m-1)(n-1)	MS_e	
Total variation	SS_T	N-1		

1) test of significance of difference between groups
2) test of significance of difference within groups

In the two-way analysis of variance, two hypotheses and their alternatives are considered, while four assumptions about the respective parent populations are made (Till, 1974):

(a) Each group is a random sample, drawn from different populations.
(b) Each parent population is normal distributed.
(c) Each parent population has the same variance.
(d) There is no interaction between different treatments and different samples.

4.4.2 Sequence analysis

The purpose of this section is to demonstrate different ways of evaluating data, which are characterized by their position along a line and thus form a sequence. The following techniques will be described:

(i) Equal spacing and interpolation.
(ii) Linear and curvilinear polynomial regression analysis.
(iii) Moving average smoothing.
(iv) Auto- and cross-correlation.
(v) Fourier series analysis.

Equal spacing and interpolation

Often field measurements are arbitrarily distributed in space and time. Various sequence analysis techniques require equally spaced data. Two interpolation techniques, which perform equal spacing interpolation on arbitrary spaced measurements, have been included in the package, namely:

(i) Linear interpolation of data.
(ii) Rectangular integration.

Linear interpolation

The method is demonstrated in the following illustration (see figure 4.20).

Assuming that there exists a linear relationship between the values y_1 and y_2, an intermediate value y' can be calculated by the following expression:

$$y' = \frac{(y_2 - y_1)(x' - x_1)}{(x_2 - x_1)} + y_1$$

The closer any interpolated point is to an observed point, the closer is its value to that of the observation.

Rectangular integration

While linear interpolation considers only the two points on either side of the interpolated point, rectangular integration considers all observations between successive points of interpolation. The original data are regarded as a step function in which the observation interval is constant (see figure 4.21(a)).

To obtain an equally spaced data sequence, a new step function of rectangles is generated which is equal in total area and equally spaced (see figure 4.21(b)).

Linear and curvilinear regression analysis

The most frequently used statistical method in geohydrological investigations is the simple linear regression analysis. This method is based on the assumption of a linear relationship between two variables and provides a means of predicting or estimating one dependent variable from the knowledge of a second independent variable (Haan, 1977). Such a relationship is described by the following expression:

$$\hat{y}_i = b_0 + b_1 x_i$$

where

$\hat{y}_i$	=	dependent variable
x_i	=	independent variable (regressor)
b_0	=	y-axis intercept
b_1	=	slope of the linear relationship (regression line)

If data cannot be fitted by a first degree relation according to the correlation coefficient, it should be fitted by a regression curve of a higher order. This is generally done by polynomial expansions of the following form:

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + b_3 x_i^3 + \ldots + b_n x_i^n$$

Moving average smoothing

This technique forms part of the generally known filter- or time trend analysis techniques. Of all the techniques resorting under this group, it is the most basic and generally used smoothing method. Irregularly spaced data can be smoothed by successive linear or curvilinear fits. This can be rather time-consuming. By interpolating

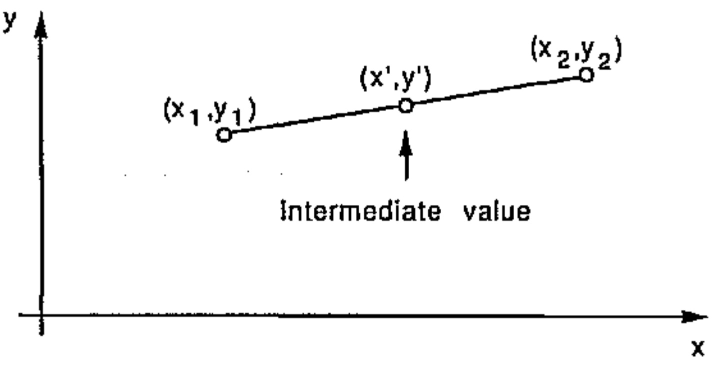


Figure 4.20. *Diagrammatic representation of the linear interpolation technique.*

(a)

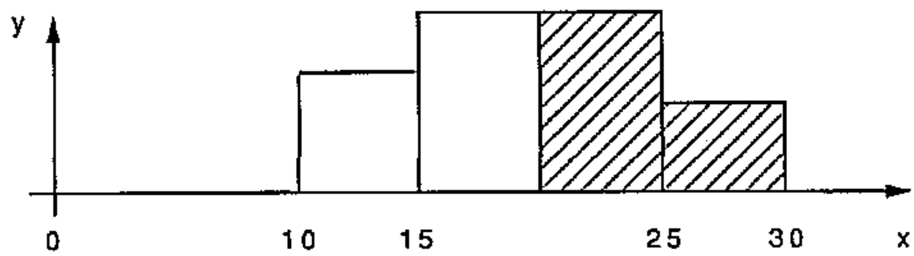


Figure 4.21(a). *Data sequence as a step function.*

(b)

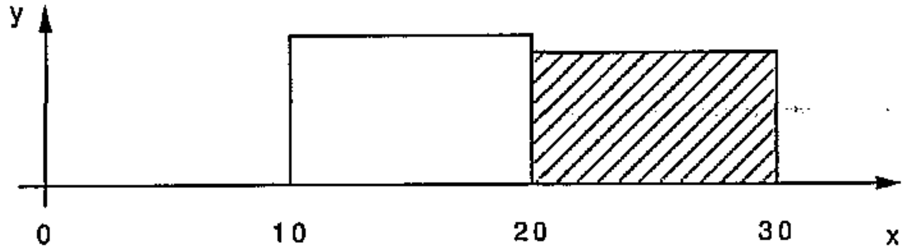


Figure 4.21(b). *Equally spaced data after rectangular integration.*

arbitrarily spaced data to an equally spaced grid, data can be smoothed by applying a simple moving average. Moving averages tend to reduce the variance of the original sequence. This may result in a shift of peaks and troughs between the smoothed curve and the original curve. However, the method proves to be valuable in many cases, e.g. eliminating the short-term variance in the data, so that the data become easier to evaluate. In its simplest form, the moving average method is mathematically expressed as:

$$y_i = \sum_{j=i-k}^{i+k} \frac{y_j}{m}$$

where

$$k = \frac{m-1}{2}$$

m = length of the smoothing interval

For proper functioning, m should be an odd number.

Auto- and cross-correlation

Data sequences, e.g. borehole hydrographs, sometimes tend to repeat a specific pattern. The repetition may not be exact, but it might give a cyclic impression. These so-called cycles can be compared with each other by calculating a measure of similarity between the segments. The process of auto-correlation compares an entire data sequence with itself at all possible positions. By doing this, not only is the degree of similarity determined, but also the degree of dissimilarity in the data sequence. The comparison of the sequence is done by computing the auto-correlation coefficient, r_L, defined as the linear correlation between a time series of the same series at a later interval of time (Davis, 1973) and can be expressed as:

$$r_L = \frac{COV(y_i, y_{i+L})}{(S_y)^2}$$

where

L = the time lag between the two series being compared.

When the auto-correlation has been calculated at every lag position, the auto-correlation is complete and a correlogram can be plotted, which represents the auto-correlation versus the lag. At lags with a high coincidence of sample values, the correlogram will show high auto-correlation.

While the auto-correlation function compares a sequence with itself, the cross-correlation compares two different data sequences with one another. This technique differs slightly from the previous method, in that the sequences compared may not be of equal length. Furthermore, the one sequence is moved past the other, matching the data with one another. The cross-correlation coefficient, r_m, can be expressed in mathematical terms as:

$$r_m = \frac{COV_{1,2}}{S_1 S_2}$$

where

$COV_{1,2}$ = covariance of the overlapped positions of data sequences.

Similar to the auto-correlation method, match positions are numbered and a cross-correlogram is constructed by plotting the match positions against the cross-correlation. In some instances, data smoothing may emphasize underlying characteristics, aiding the cross-correlation technique. However, data smoothing generally reduces data independence (Davis, 1973).

Fourier series

Fourier analysis is the decomposition of a signal into its harmonic constituents. Fourier analysis is applied to any type of equally spaced data sequence, e.g. water-level charts. The technique divides any curve into three basic components, namely:

(i) Linear trend.
(ii) Periodic components.
(iii) Random component.

If a trend exists in the data, it should be detected and possibly be removed. The primary purpose of this evaluation technique is to extract the dominant periodic components of a specific curve, thus filtering the background, high frequency noise from the curve. Periodic functions are generally represented by sinusoidal wave-, or better known as Fourier series, and can be expressed as (Spiegel, 1974):

$$y_i = \frac{a_0}{2} + \sum_{i=1}^{n} \frac{a_n \cos(n\pi x_i)}{\lambda} + \frac{b_n \sin(n\pi x_i)}{\lambda}$$

where

a_0, a_n, b_n	=	the cosine and sine phase angles (Fourier coefficients)
λ	=	the wave length

In basic terms, the amplitude y_i at a point x_i is determined by the sum of the amplitudes of the component sine and cosine waves at a distance x_i from the origin.

This method is also frequently used to separate regional and residual anomalies from each other, often encountered in geophysics.

4.4.3 Map analysis

Data sequences represent data on profile lines. However, the two-dimensional representation of data in the form of a map is also a good means to display data. Two facilities have been included in the software package:

(i) Contouring (on a line printer).
(ii) Trend surfaces.

For more detail about contouring, see section 4.3.4.

Trend surfaces

Trend surfaces are defined as linear functions of the geographic coordinates of arbitrarily distributed data, such that the squared deviations from the trend are minimized. Polynomial expansions are used for geohydrological or geological trend surface analyses. Due to their extreme flexibility, they can be expanded to sufficiently high orders that may fit complex surfaces. For example, the mathematical equation for a second-degree trend surface (Krumbein *et al.*, 1965) is:

$$z = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy$$

This technique is used to separate various map components from one another, thus dividing the data into a regional and local component. Generally, trend surfaces are used to determine parameter distribution. A very useful aspect of trend analyses is that differences between surfaces of different degrees can indicate areas of geohydrological anomalies. A trend surface analysis can provide an initial guide for interpretation of arbitrarily distributed data. In cases where the geohydrological data are not sufficiently well-distributed, trend surface analyses tend to produce unrealistic patterns (Lloyd and Heathcote, 1985).

In order to get meaningful results, the number of data points must exceed the number of coefficients in the polynomial expansion. In practice, only polynomials up to a fourth degree are used. Furthermore, if statistical tests are done on the data, the number of points determines the degrees of freedom, which must be large to ensure a meaningful F-test. The shape of the area also affects the form of the polynomial fit. That is, if the data are not evenly distributed over the area, the trend surface tends to be elongated parallel to the pattern of the data points. Thus, it is good practice to have data over a larger area than the area to be mapped.

4.4.4 Multivariate data analysis

All the previously discussed methods considered single variables. However, especially in hydrochemistry, it is often required to look at the changes of several properties simultaneously. This is accomplished by multivariate data analysis, which is divided into some of the following evaluation methods, namely:

- (i) Multiple linear regression analysis.
- (ii) Discriminant function analysis.
- (iii) Cluster analysis.
- (iv) Principal component analysis.
- (v) Factor analysis.

Multiple linear regression

When one dependent variable and various independent variables are available, multiple regression is commonly used to find a linear model for predicting unobserved values for the dependent variable. The model eventually used does not necessarily contain all the independent variables. A general linear model was adopted in the processing package and can be mathematically expressed as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon$$

where

y = the dependent variable
$x_1,..,x_n$ = the independent variables
ε = the random error

The first step in a multiple regression analysis is to compute the correlation matrix between independent variables.

Retaining values in a regression equation that are highly correlated, complicates the evaluation of the regression coefficients. A common practice in selecting a multiple linear regression model is to perform several regressions on a given data set, using different combinations of the independent variables. The regression analysis with the best data fit is then selected. The criterion generally used for this purpose is the goodness of fit, yielding the largest value.

All the variables retained in a regression should make a significant contribution to the regression, unless there is an overriding reason for retaining a non-significant variable. All the variables retained should have a physical significance.

The ultimate test for the chosen model depends on its ability to predict the dependent variable for observations of the independent variables that were not used in estimating the regression coefficients (Haan, 1977).

When the multiple regression analysis of the software package is applied, the user is provided with two sets of regression coefficients, namely:

(i) Partial regression coefficients.
(ii) Standardized partial regression coefficients.

The partial regression coefficients represent the rate of change (slope) in the dependent variable for a unit change in the independent variable. The standardized partial regression coefficients are the partial regression coefficients expressed in units of the standard deviation. These coefficients provide a means of assessing the importance of individual variables in the regression, which may not have the same units.

Discriminant function analysis

The main objective of the discriminant function analysis is to provide a numerical criterion for classifying variables into statistically distinguishable groups. It is generally applied for two different purposes. Firstly, to assess the statistical significance of the stated hypothesis and, secondly, to classify variables into specific groups (Seyhan, 1981).

The discriminant function transforms sample measurements into a single discriminant score, where the score represents the sample's position along a line defined by the discriminant function. This transformation represents the minimum ratio of the difference between the means of two data groups and the variance within the groups.

A correlation matrix of the relevant variables is formed and is then tested pair by pair to establish which pairs show a correlation not significantly greater than zero. By pivotal condensation, the matrix is reduced, resulting in a linear discriminant function. The discriminant index is then used to compare the two groups of variables with each other, but only after those variables are eliminated which do not contribute to the discrimination.

When applying linear discriminant analysis to groups of data, it should be noted that the populations must be different. To establish whether the populations differ, a T^2-test can be applied to the data sets. Thereafter, the discriminant function analysis can be carried out to establish a useful index for assignment of new data to either one of the relevant data populations.

Cluster analysis

Cluster analysis is ideally suited to subdivide a given population into a number of groups, generally known as clusters. Cluster analysis selects sets of mutually similar variables from a matrix of many such variables, as based on the similarity coefficient matrix defining the interrelationships among the variables (Seyhan, 1981).

Raw data generally consist of a number of observations with one or more parameters or characteristics measured during each observation. Thus, the data can be represented by a n x m matrix. The purpose of the technique is to establish whether a measure of resemblance exists between the observed characteristics. In the processing package, the similarity matrix is constructed by either one of two methods, namely:

- (i) Correlation coefficient method.
- (ii) Distance coefficient method.

Similarities are determined between the variables (or observations) by arranging the variables in such a way that the elements with the highest mutual similarity are placed together. These groups or clusters of elements are now associated with other groups, which they closely resemble, until all the groups have been classified. Such a classification can be graphically represented by a dendrogram.

Cluster analysis is sometimes used as a prelude to factor analysis, although it has certain advantages in that the identity of the original variables is maintained.

Principal component analysis

Often the user is confronted with the problem of having to simplify and describe the relationships between complex variables of a large population. Principal component analysis and factor analysis are ideally suitable to represent such relationships, by simpler relationships amongst fewer variables, making large complex relationships more comprehensible.

The objective of the principal component analysis is to find a linear transformation of a given set of variables, such that specific statistical properties are satisfied. The transformed data will have the property of being uncorrelated and chosen in such a way that each succeeding variable will have a smaller variance (Fenner *et al.*, 1969).

The transformed variables are linear combinations of the original variables and are known as the principal components. The number of principal components will be equal to the number of variables used. Generally, the first few principal components form a large part of the total variance. Furthermore, principal components represent the coordinate axes of an ellipsoid, formed by the scatter of the field data. Once the ellipsoid axes have been obtained, their magnitudes are measured.

Factor analysis

Factor analysis is used to analyse relationships within a set of variables or observations, by means of the analysis of all relevant data matrices. The analysis is done in order to determine whether the observed data set can be simplified.

Two modes of factor analysis exist, namely:

(i) R-mode.
(ii) Q-mode.

The R-mode factor analysis investigates the relationships between variables, whereas the Q-mode determines the relationships between the various observations. Both methods can be performed on the same data set, however, with different results. No matter what mode is used, four steps should be followed in order to ensure significant analysis of the data, namely:

(i) Preparation of input-data matrix.
(ii) Extraction of initial factors.
(iii) Factor matrix rotation.
(iv) Factor score calculation.

R-mode factor analysis

Let the observed data set be contained in matrix, $X_{N,n}$. The objective is to determine m linear combinations of the original n variables that describe the geohydrological features without loss of information. The above-mentioned linear combinations are generally known as factors. The technique does not use the original data matrix, $X_{N,n}$, for computational purposes, but rather a correlation coefficient or distance coefficient matrix, derived from the original data matrix. The data are always standardized, in order to ensure that all variables are in the same units. Once a correlation between the variables has been established, the factor scores are calculated by the following expression (McCammon, 1975):

$$z_{kj} = a_{1j} f_{k1} + a_{2j} f_{k2} + ... + a_{mj} f_{km} + a_j e_{kj}$$

where

z_{kj}	=	the j-th variable of variable k (standardized variables)
a_{mj}	=	the factor loadings of the j-th variable and m = 1,2,3,...
f_{km}	=	the factor scores of the variables and m = 1,2,3,...
e_{kj}	=	unique factor to variable z_{kj} and is analogous to the error term in a regression equation

Unlike the factor scores, the factor loadings do not change from variable to variable. The factor loadings can be rotated by Kaiser's Varimax scheme, in order to obtain even better results.

Q-mode factor analysis

Due to the fact that the relationship between samples is determined with this method, the method differs slightly from the R-mode, in that the input data matrix is transposed. The remaining computations are executed in the same way as previously discussed under the R-mode analyses.

To discuss the multivariate statistical analyses techniques, as well as all the other statistical techniques in detail in this report, would be a rather tedious exercise. The following books were consulted during the compilation of this chapter and are recommended for those readers who want more detail about the respective techniques: Davis (1973), Haan (1977), Krumbein and Graybill (1965), Till (1974), Miller and Kahn (1962), Seyhan (1981), McCammon (1975) and Jöreskog, Klovan and Reyment (1976).

The data processing techniques discussed, are demonstrated by data processing examples of site specific data in Appendix I.

CHAPTER 5

EVALUATION OF THE PROCESSING FACILITIES OF THE GROUND-WATER DATA BASE.

5.1 INTRODUCTION

As part of the project, borehole data from the Orange Free State and Atlantis were entered into the National Ground-water Data Base, in order to test its functioning and that of the processing package. Some of the data were entered by the Institute for Ground-water Studies and others by the Directorate of Geohydrology of the Department of Water Affairs.

Roughly, one-third of the Orange Free State data were extracted from internal reports of the Institute for Ground-water Studies and the Directorate of Geohydrology. The data available at this stage cover rather restricted areas of the Orange Free State and, with a few exceptions, only basic data have been captured. To what extent the data variety is limited, will be seen in the subsequent discussion of the various data sets from the Orange Free State.

The remaining two-thirds of the data were extracted from borehole completion cards which, in most cases, contained rarely enough site specific data to allow inclusion in the ground-water data base. The main set-back was the uncertainty about the exact borehole positions, when only the names of the farms were known on which the holes were drilled. This uncertainty is illustrated in figure 5.1, where two-thirds of all borehole sites, which were entered for the Orange Free State, have a coordinate accuracy of "4". This means that the position of the holes is accurate within 10 km.

Borehole data from the Atlantis area were included to provide data for the test of those data sets that were not sufficiently covered amongst the Orange Free State data.

Apart from testing the performance of the processing package, the purpose of this chapter is to demonstrate the applicability of the processing programs in the evaluation of ground-water resources. For this purpose, the data were divided into four topics:

(i) Geology.

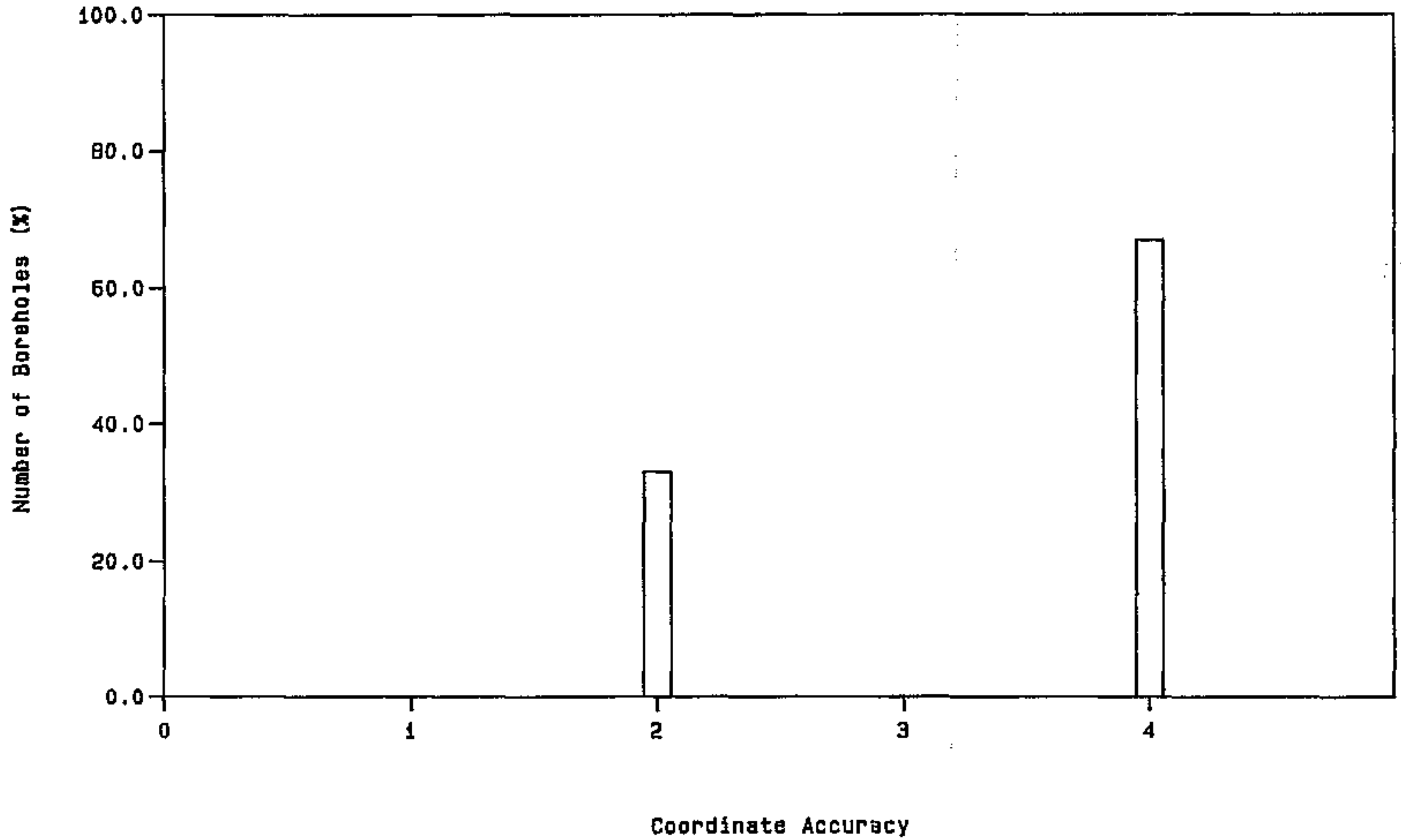


Figure 5.1. *Bar chart showing the distribution of borehole coordinate accuracies.*

(ii) Aquifer analysis.

(iii) Borehole construction and installation.

(iv) Ground-water quality.

5.2 GEOLOGY

Figure 5.2(a) shows a generalized map of the geology of the Orange Free State. Almost the whole of the Orange Free State is underlain by sedimentary rocks of the Karoo Sequence, which are intruded by dolerite sills and dykes. The only region where other than Karoo sediments occur, lies in the northern part of the province. Here, the Vredefort dome is found which consists of Ventersdorp lava, Witwatersrand sediments and granite.

But for a few exceptions, the whole Karoo Sequence is present in the Orange Free State. This can be seen when a section, from Kimberley in the west through Bloemfontein to Maseru in the east, is considered.

Figure 5.2(b) is a simplified, geological map of the Atlantis area. On this map, it can be seen that the greater part of the area is covered by unconsolidated Cenozoic deposits, which in turn are underlain by sediments of the Malmesbury Group.

The geological data from these areas will be evaluated, both on a local scale and a regional scale. The geology data set is divided into two groups:

(i) Unconsolidated geology.

(ii) Consolidated geology.

As discussed in the previous chapter, the unconsolidated data set refers especially to the unconsolidated formations. Examples of such formations are the alluvial, colluvial and aeolian deposits in the Orange Free State, as well as the Cenozoic deposits of the Atlantis area.

Consolidated geology refers to hard rock, namely the igneous, metamorphic and consolidated sedimentary rocks.

5.2.1 Local analysis

Borehole data from the Orange Free State are unevenly distributed. Because of their present deficiencies regarding the geological logs, the geological data should rather be

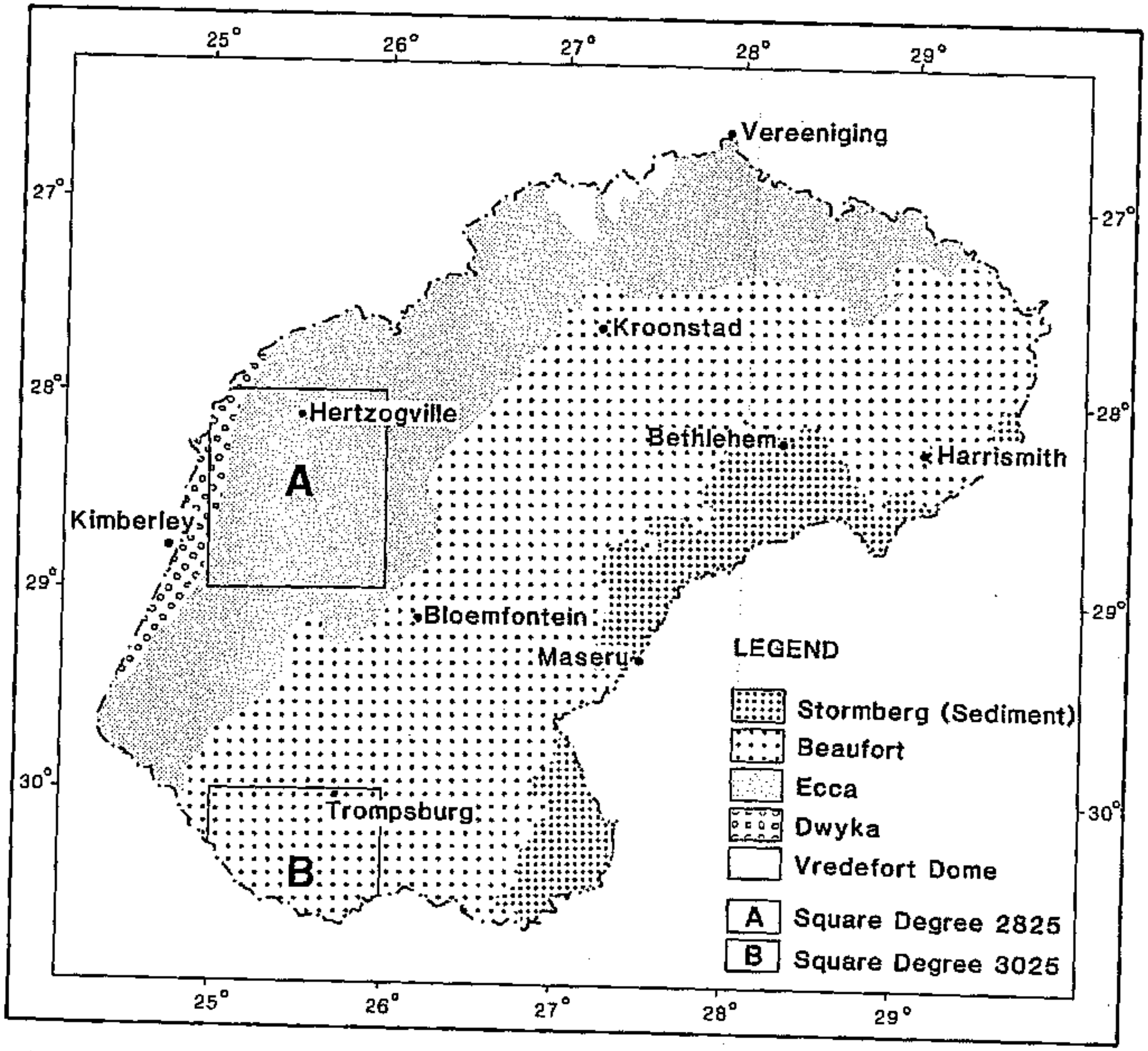


Figure 5.2(a). Simplified geological map of the Orange Free State (after Truswell).

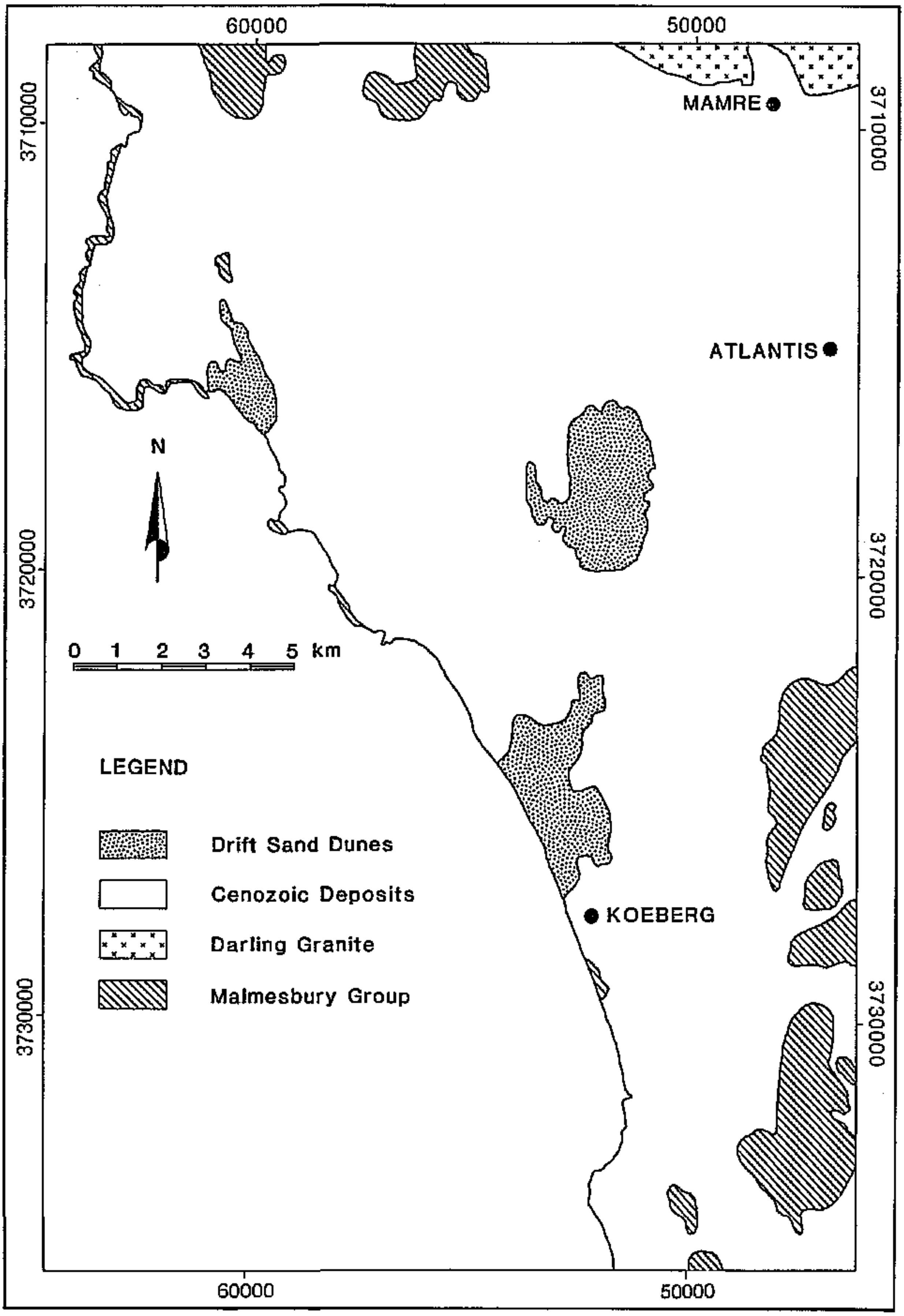


Figure 5.2(b). Simplified geological map of the Atlantis area (after Müller, 1984).

evaluated locally than regionally. It is, however, not the aim of this chapter to evaluate the whole Orange Free State on a local scale. Two areas, containing a fair amount of data in the consolidated geology data set, were chosen as an example. Both the areas comprise a square degree. They are:

(i) Square degree 3025.

(ii) Square degree 2825.

A third area, the Atlantis area, will be used to evaluate the unconsolidated geology data set.

Square degree 3025

This square degree lies in the southern Orange Free State in the vicinity of Trompsburg, Smithfield and Bethulie. It is predominantly underlain by the Beaufort Group of the Karoo Sequence. Of the 1840 boreholes entered into the data base, 85 boreholes were geologically logged. On its own, the geology data set can divide the region into the various lithological units. If, however, the geology data set is combined with the aquifer data, a rather complete lithological subdivision of the geology can be made. Figure 5.3 shows four typical borehole logs of this area. The symbols $\underline{\nabla}$ and $\lhd$ indicate the water level and depths at which water was struck in the boreholes. A good visual impression of the geological strata encountered in the holes is obtained. Figure 5.4 shows the borehole profile with the four boreholes displayed in figure 5.3. It should be noted that the boreholes used in the profiles have only a vertical scale, while horizontally the boreholes, which should lie more or less on a straight line, are equally spaced.

It is hardly possible to write general programs which depict successions of thin layers and successive thick formations equally well. In figure 5.4, it can be seen that the description of the lithology of the successive thin layers plotted, becomes indistinguishable. As it is undesirable to disregard all thin beds in the section, the problem can be overcome if either the particular geological log is simplified by combining a number of thin layers in one representative layer or by including them in the adjoining thicker strata. If this is not acceptable, only those logs in which the minimum thickness of the individual layers exceeds $\pm$ 2,5 m, may be presented on borehole profiles.

While the geological description in the borehole logs is given in uncoded form and a legend for the different formations is added to the borehole profiles, the tabulated geological information in table 5.1 requires the use of the code lists (due to practical reasons which are discussed in chapter 6.3.1.1 of the User's Guide). Because of the

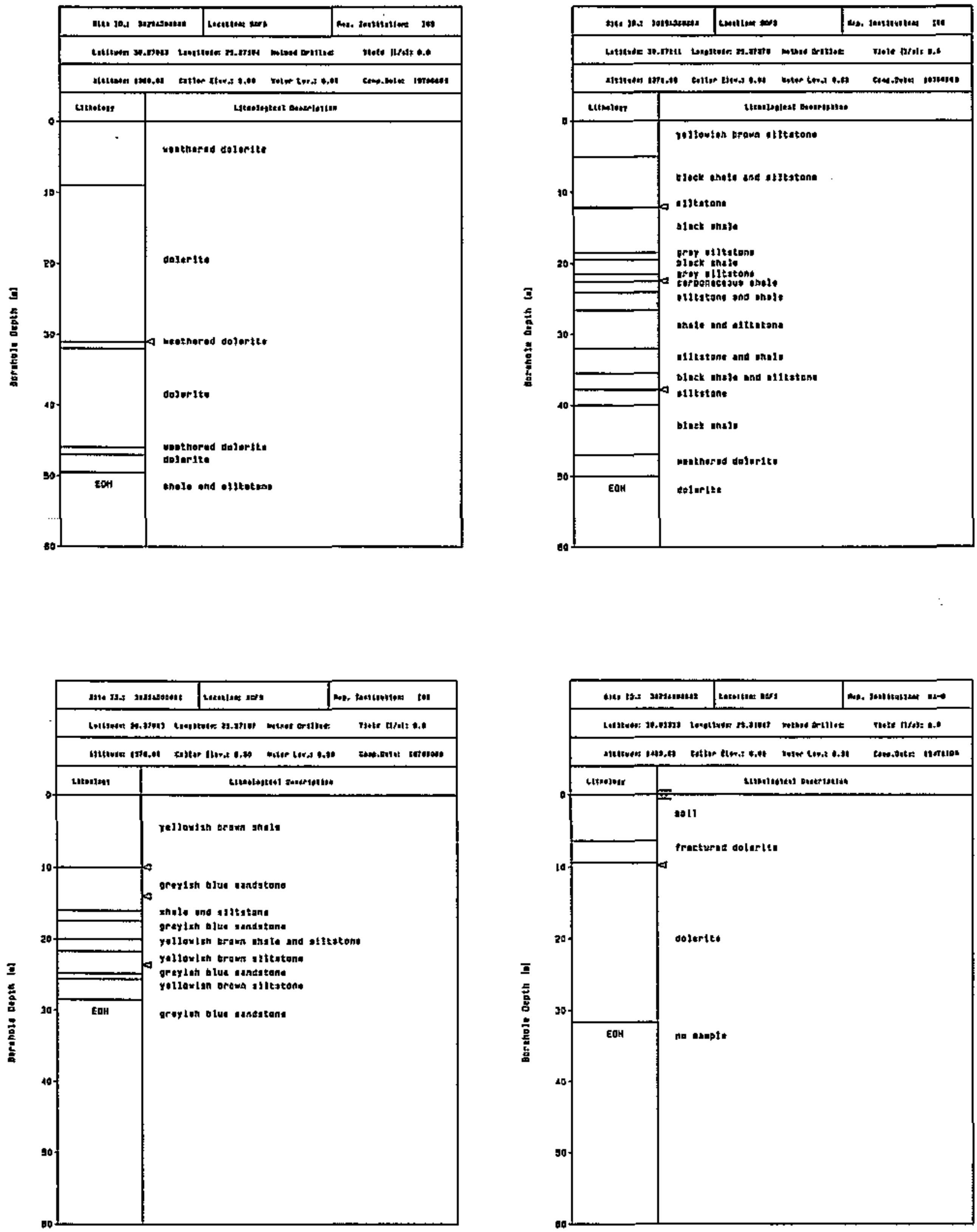


Figure 5.3. *Borehole logs of the southern Orange Free State.*

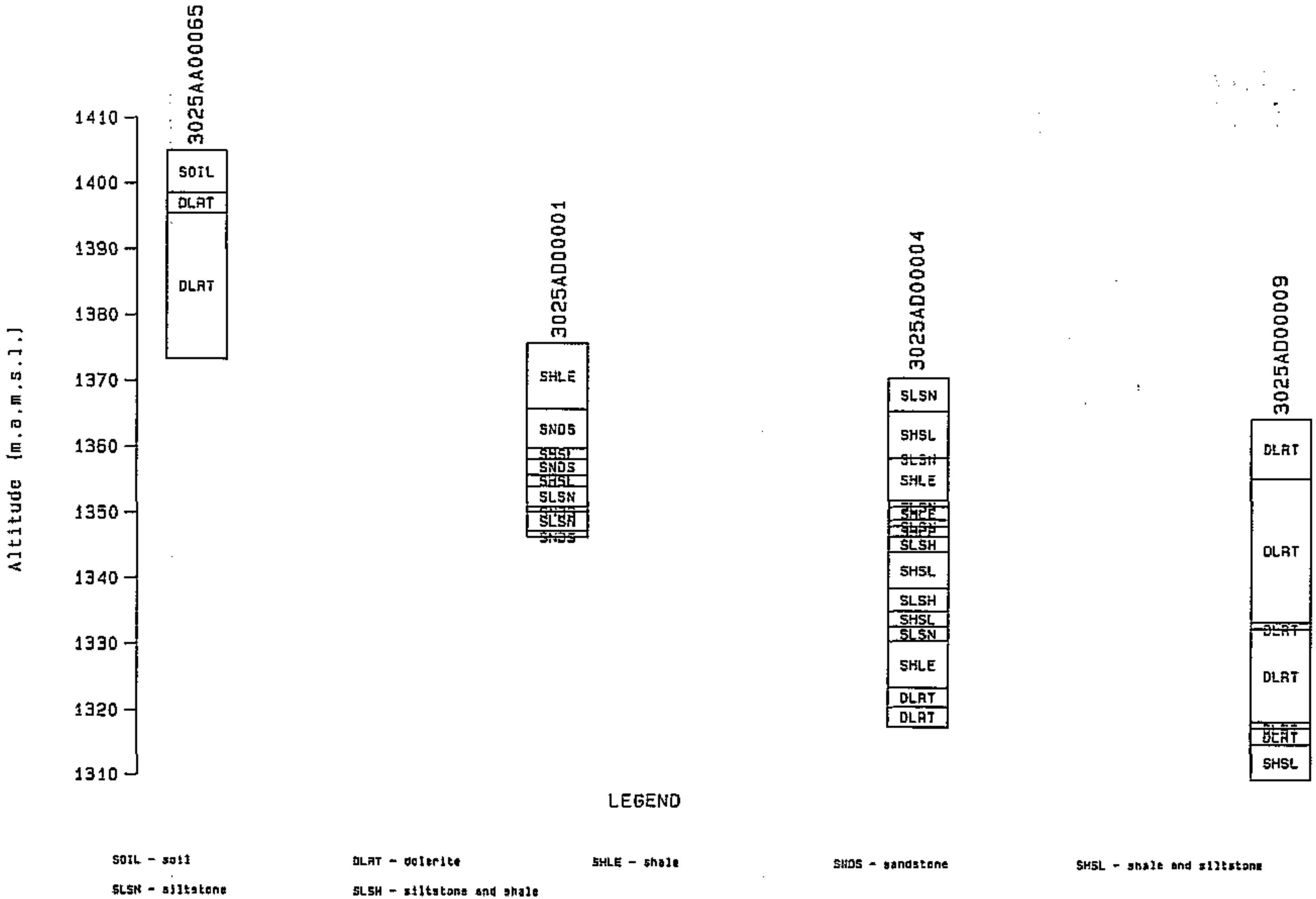


Figure 5.4. *Simplified borehole profile of boreholes displayed in figure 5.3.*

TABLE 5.1. TABULATED GEOLOGICAL DETAIL OF SOME OF THE BOREHOLES IN THE SOUTHERN ORANGE FREE STATE.

Site Id.	Depth to Top (m)	Lithology Code	Primary Color	Secondary Color	Texture	Feature
3025AD00001	10.00	SNDS	B	H		
	16.00	SHSL				
	17.50	SNDS	B	H		
	20.00	SHSL	C	Y		
	21.70	SLSH	C	Y		
	24.00	SNDS	B	H		
	25.60	SLSH	C	Y		
	28.50	SNDS	B	H		
3025AD00002	0.00	SLSH	C	Y		
	7.00	SHLE	B			
	10.00	SNDS				WT
	15.00	DLRT				WT
	23.00	DLRT				WT
	26.00	DLRT				
3025AD00003	0.00	SLSH	C	Y		
	6.50	SHSL	S			
	15.50	SNDS	B	H		
	15.60	SHSL				
	18.50	SNDS				
	18.60	SHSL				
	20.00	SNDS	H			
	20.20	SHSL				
	24.90	SNDS				
	24.20	SHSL				
	25.30	DLRT				WT
	31.50	DLRT				
3025AD00004	0.00	SLSH	C	Y		
	5.00	SHSL	S			
	12.00	SLSH				
	12.20	SHLE	S			
	18.50	SLSH	H			
	19.50	SHLE	S			
	21.50	SLSH	H			
	22.50	SHLE				CO
	24.00	SLSH				
	26.50	SHSL				
	32.00	SLSH				
	35.50	SHSL	S			
	37.80	SLSH				
	40.00	SHLE	S			
	47.00	DLRT				WT
	50.00	DLRT				
3025AD00005	0.00	DLRT				WT
	13.50	DLRT				WT
	15.00	DLRT				

restricted recorded geological detail in table 5.1, a subdivision of the Beaufort Group into distinct lithological horizons is not possible.

Analysing the geological data from 85 boreholes, using either the borehole logs or the data print-outs, a monotonous sequence of shale, siltstone and mudstone, with interbedded sandstone, is found. The wide-spread intrusion of dolerite sills and/or dykes can also clearly be seen. In 65 boreholes, dolerite was encountered at some stage during the drilling process. The sedimentary succession can be classified as argillaceous, because of the predominant shaly deposits in the succession. The colours of the rocks vary from black, greyish blue, yellowish brown to grey and white. Black carbonaceous shale was also frequently found, confirming the rich organic (fossil) content of the Beaufort succession in this area. The sandstone normally encountered, weathers to a yellowish coloured rock and is often feldspathic. Colluvium, variable in thickness, is frequently encountered in the region at and near the surface. By comparing the lithology code and depth to top data sets with the topo setting, one finds that the colluvium is seldom thicker than 11 m and that it occurs in the flat areas and the lower hill slopes of the region.

Square degree 2825

This area lies in the western Orange Free State, in the vicinity of Hertzogville. It is underlain by sediments of the Ecca Group. Of the 367 boreholes in the area, 242 boreholes were logged geologically. As more than half of the logged boreholes have a coordinate accuracy of "4", the data lend themselves better to an analysis on a smaller scale, where the exact positions of boreholes are less important.

Figure 5.5 shows some logs of the formations encountered around Hertzogville. According to the geological description of 120 boreholes in the district, the formations consist of a sequence of bluish-black shale with interbedded sandstone. Similar to the southern Orange Free State, the sediments are intruded by dolerite. Figure 5.6 shows the same boreholes in a profile.

Combining the information on the geological logs with the aquifer data set, it is confirmed that the sequence forms part of the Ecca Group, as described by Nel (1977) and Behounek (1980). Further subdivision of the Ecca into distinct horizons, based on different physical properties, is not possible.

It is also observed that the top part of the succession frequently consists of various unconsolidated sediments, such as sand, soil, clay, sand and gravel. According to Behounek (1980) and Grobler *et al* (1987), most of the sand deposits in the region stem from wind erosion of the surrounding region and are classified as aeolian deposits.

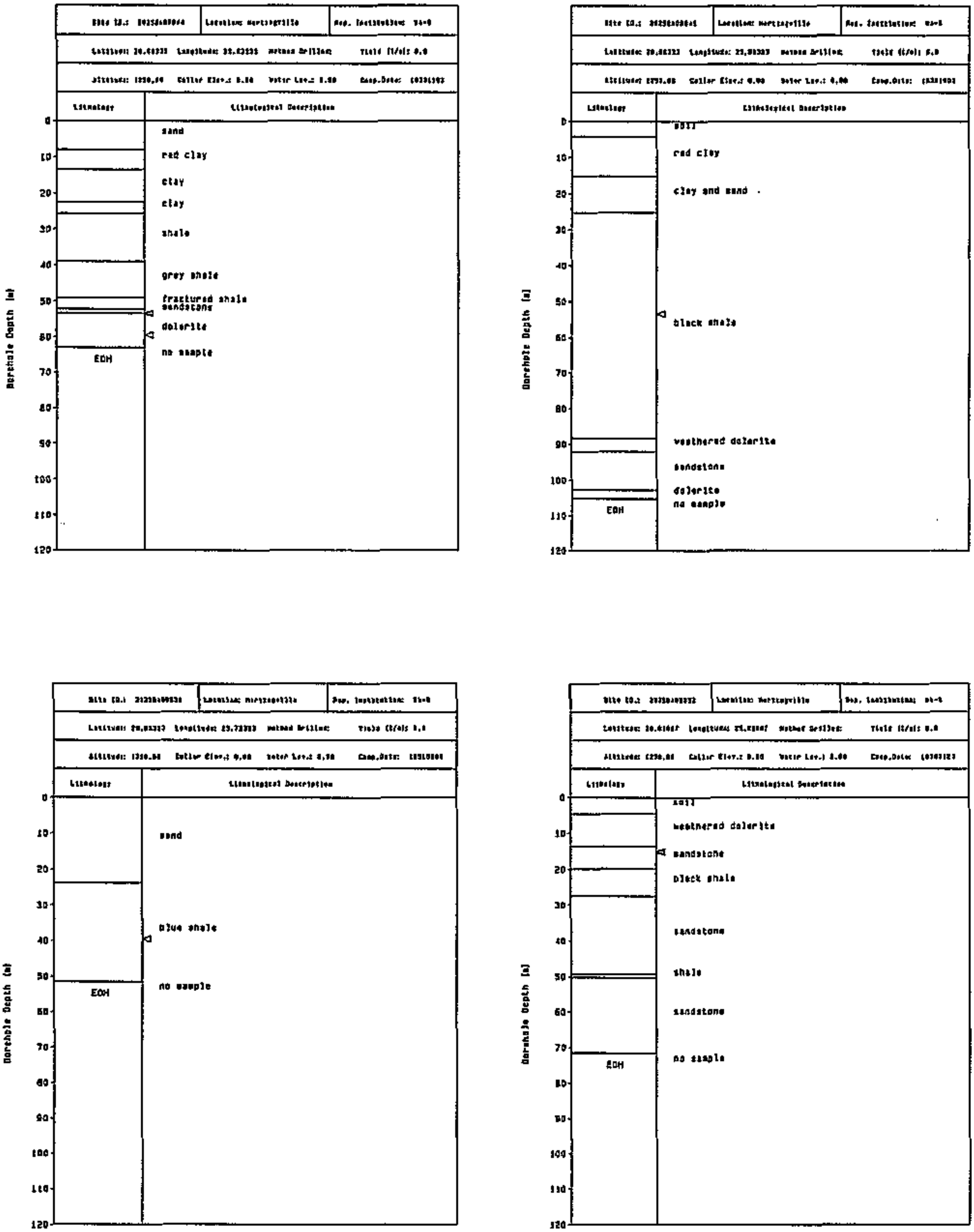


Figure 5.5. *Borehole logs of the Hertzogville region.*

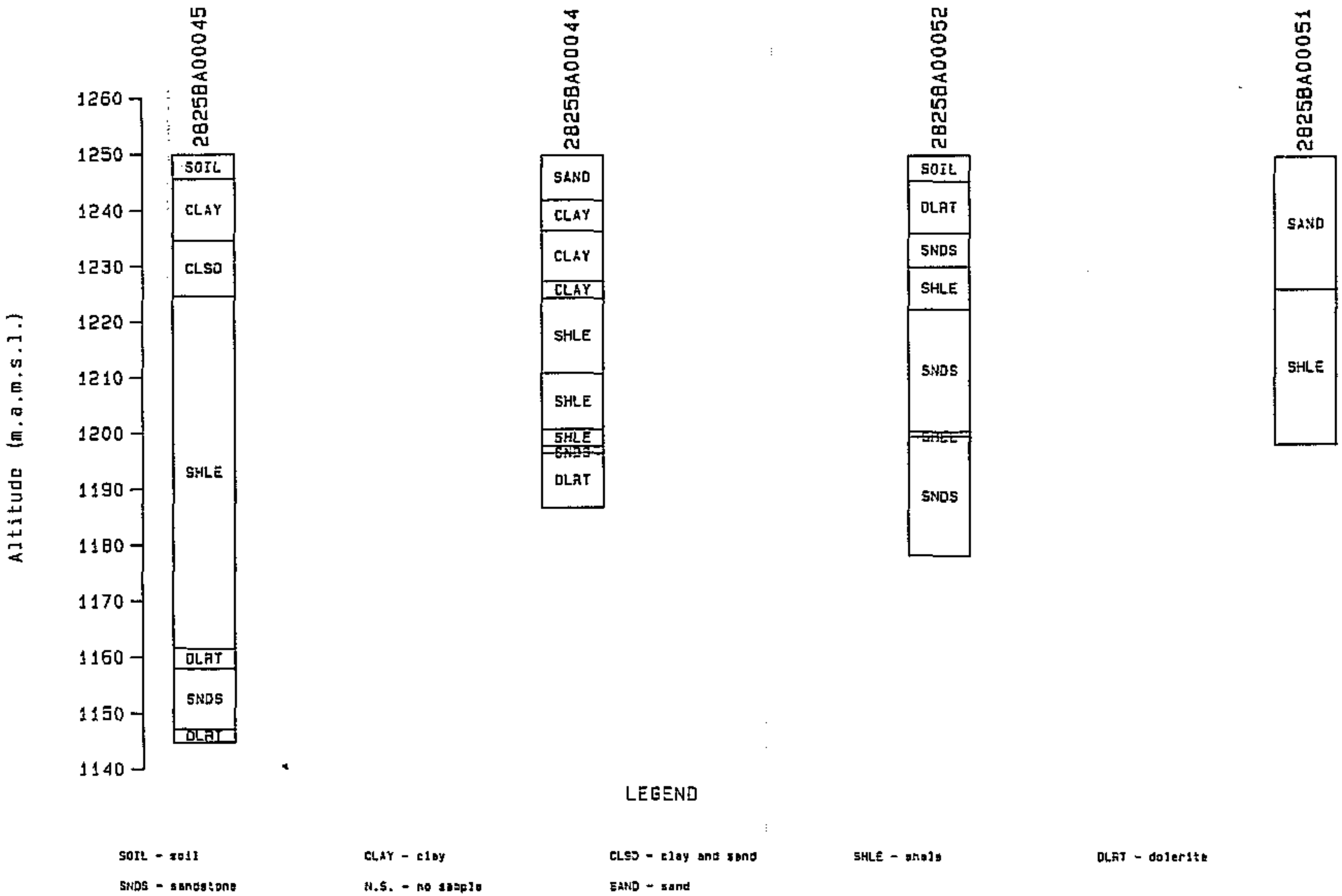


Figure 5.6. Borehole profile of the Hertzogville region.

Atlantis area

As in the case of the Orange Free State, the geology and aquifer data sets from the Atlantis area have to be used in combination to subdivide the unconsolidated geological data into the proper lithostratigraphical units. Four different units can then be recognized, i.e.:

(i) Varswater Formation.

(ii) Bredasdorp Formation.

(iii) Darling granite.

(iv) Tygerberg Formation.

The Bredasdorp and Varswater Formations are of Cenozoic (Tertiary/Quaternary) age, while the Tygerberg Formation forms part of the Late-Precambrian Malmesbury Group.

Tygerberg Formation

The Tygerberg Formation underlies the Cenozoic deposits of the whole area. Outcrops of the formation are frequently encountered, especially along the coast and in the south-eastern parts of the Atlantis area. The formation consists of a predominantly pelitic and semi-pelitic succession, interbedded with massive medium- to fine-grained, sometimes phyllitic, greywacke and impure quartzite.

It is characteristic for this formation, that it is generally deeply weathered beneath the younger sand deposits and can therefore easily be misinterpreted as Cenozoic sediments.

Darling Granite

Outcrops of the Darling Granite, which intrudes into the Malmesbury Group, are found in the northern parts of the Atlantis area, as shown in figure 5.2(b). Very little is known regarding the water-bearing characteristics of the granite.

Varswater Formation

The Varswater Formation is a marine deposit and is divided into two members:

(i) Duynefontein Member.

(ii) Silwerstroom Member.

Duynefontein Member

The Duynefontein sand deposits usually overlie the Silwerstroom deposits. Where the latter deposits are absent, the Duynefontein deposits rest on the Tygerberg Formation. This member consists of fine well-sorted sand, with minor pelletal phosphorites.

Silwerstroom Member

According to Bredenkamp and Vandoolaeghe (1982) and Van der Merwe (1983), the Silwerstroom Member is a poorly sorted deposit, consisting of a phosphate-bearing mud, gravel and sand. It also contains large quantities of shell fragments. A further distinctive property is the fact that the Silwerstroom Member always occurs beneath the sea level (Rogers, 1980).

Bredasdorp Formation

In the Atlantis area, the Bredasdorp Formation is of aeolian origin. It can be subdivided into three members:

- (i) Witzand Member.
- (ii) Mamre Member.
- (iii) Springfontein Member.

Witzand Member

This member is a shelly, calcareous, aeolian sand deposit. According to Bredenkamp and Vandoolaeghe (1982), it forms the top part of the Bredasdorp Formation and usually occurs above the water table of the Atlantis area.

Mamre Member

The Mamre Member is a salty, peaty variant of the Springfontein Member. A further characteristic of the member is its high silt content. The silt in the formation does adversely affect the hydraulic conductivity of the aquifer.

Springfontein Member

This member consists of well-sorted fine to coarse sand and is by far the most permeable member (Van der Merwe, 1980). Shell fragments are found at the base of the member in the form of a basal gastropod bed, whereas in the remaining part of the deposit no shell fragments are found.

Figure 5.7 shows a borehole log of formations typically encountered in the area. Where data items such as Water level, Collar elevation, Yield or Method constructed are missing, these particular data were not entered into the data base.

5.2.2 Regional analysis

All geological and aquifer logs of the Orange Free State, accommodated in the data base, were extracted to see how they compare with the regional geology of the Orange Free State, which was illustrated in figure 5.2(a). The data are relatively well-suited for qualitative regional analysis and allow the user to form a good idea of the formations expected in regions where geological data are recorded. Insufficient detail in both the geology and aquifer data sets, however, prevents differentiation between regions with similar geology. The information contained in the data base provides an overview of the basic geology of an area, if the data are available.

A similar analysis of the available geological data can, of course, also be done without the help of a computer. The major benefit derived from the computerized analysis is the fast and comprehensive access to the relevant data, supported by the processing facilities, providing data print-outs, logs, profiles and maps, as required. For any area of interest, the user can, with the aid of the data base, extract large volumes of geohydrological data in a short time. These data can then be processed to produce, e.g. print-outs or borehole logs, ready for inclusion in reports. This increases productivity as well as insight.

5.3 AQUIFER ANALYSIS

As it is the purpose of the data base and the data processing package to assist in the detection, development and operation of the ground-water resources in Southern Africa and to facilitate planning and management thereof in this section, the data accommodated in various data sets are combined and evaluated to show how they can help to assess the ground-water potential of the Orange Free State and that of the Atlantis area. As in the case of geology, the section is divided into a local and regional analysis.

In general, the ground water in the Orange Free State is associated with dolerite structures, although ground water is also frequently found in the Karoo sediments itself. The potable ground water in the Atlantis area stems from the unconsolidated Cenozoic sediments.

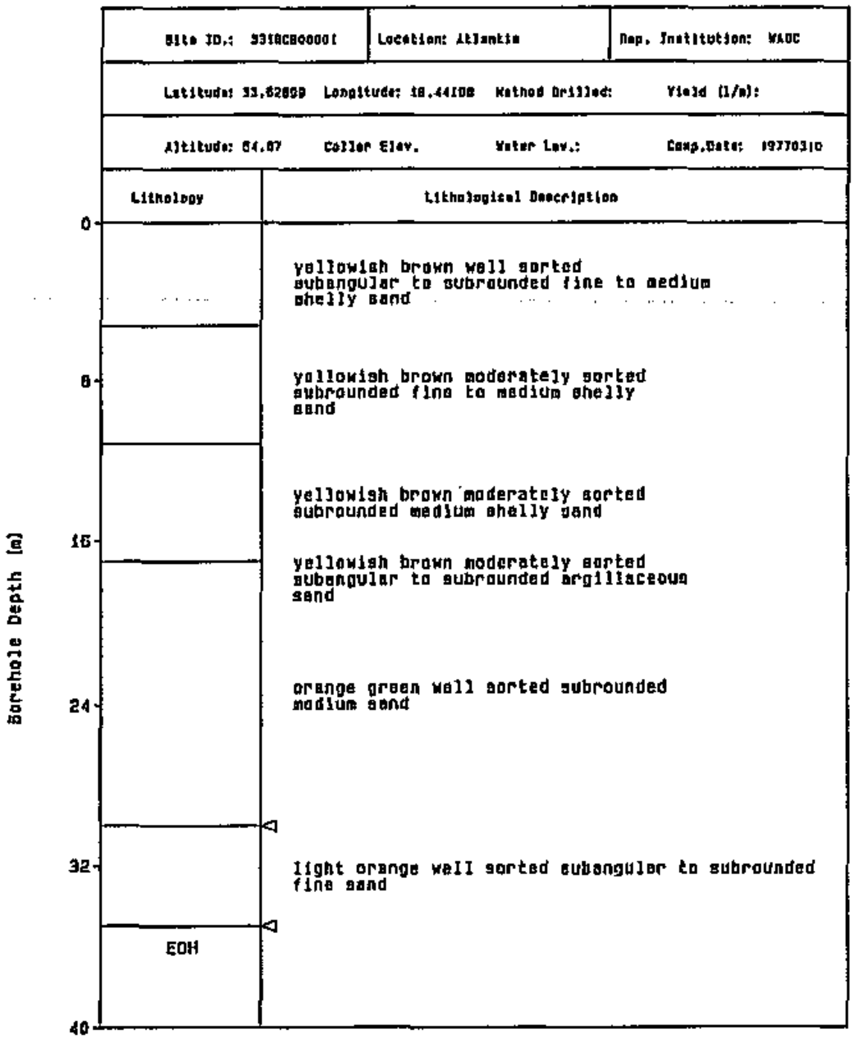


Figure 5.7. Borehole log of the Atlantis area.

5.3.1 Local analysis

Two areas were selected for the local analysis, namely square degree 3025 and the Atlantis area.

Square degree 3025

There are 1840 boreholes in the area of which 1138 have registered borehole depths, 765 have registered yields, 845 have registered water levels and 217 have registered depths at which water was struck plus an aquifer code. Figures 5.8(a), (b), (c), (d) and (e) represent the borehole positions, positions of boreholes with recorded water levels, water-level contours, altitude contours and geology of the area under discussion. None of the boreholes in the region has a long-term record, therefore water-level fluctuations in the region could not be studied. The water-level contour map in figure 5.8(c) is based on all available water levels. It illustrates that water levels closely follow the surface topography. Little difference can be seen when the water-level contours are compared with the altitude contours. However, if the areas, marked A, of the water-level contour map and altitude contour map (see figures 5.8(c) and (d)) are compared to each other, it is observed that the water levels rise above the topography, which is incorrect in this case. If the same areas (marked A) in figures 5.8(a) and (b) are compared to each other, the reason for this situation becomes obvious, as the number of observation points in the enclosed areas varies considerably. The lack of water-level observation points in this area is directly responsible for the inaccurate contours (see figure 5.8(b)). Thus, when constructing contour maps of a region, it is important to identify areas where little or no data are available, in order to minimize above-mentioned inaccuracies. Possible solutions to such problems are to refine the grid matrix or to use other interpolation techniques.

Long-term water-level measurements or contoured water-level changes would, however, give an indication of seasonal variations and long-term water-level trends. In table 5.2, the water level statistics are shown: water levels are comparatively shallow and show little deviations. Differences between water level- and altitude contours can therefore be expected in the in the decimetre- and metre range, but not in the 10 m range.

TABLE 5.2. WATER-LEVEL STATISTICS.

Number of observations	=	811
Mean water level	=	5,37 m
Standard deviation of water levels	=	2,20 m

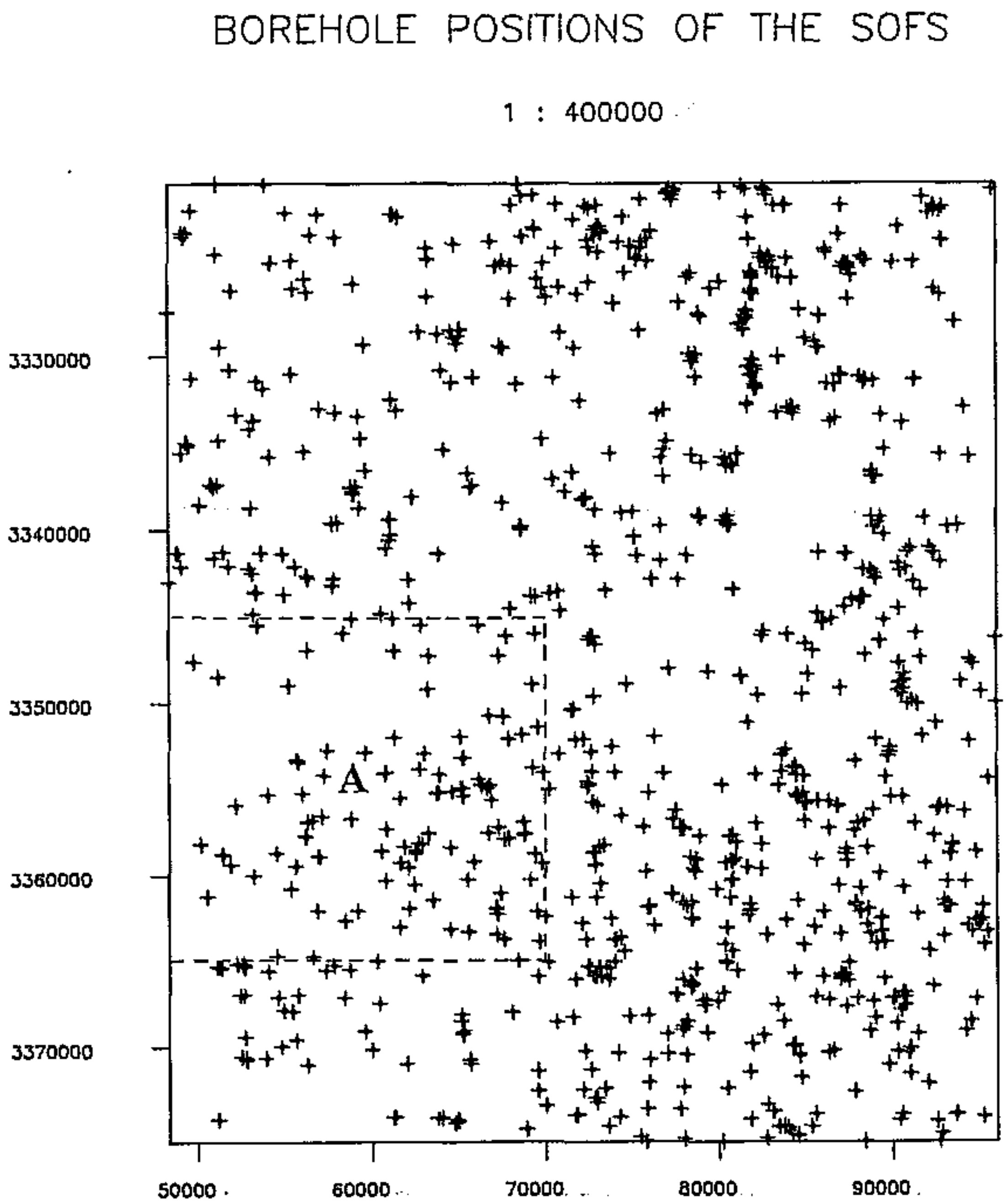


Figure 5.8(a). Borehole position map of the southern Orange Free State.

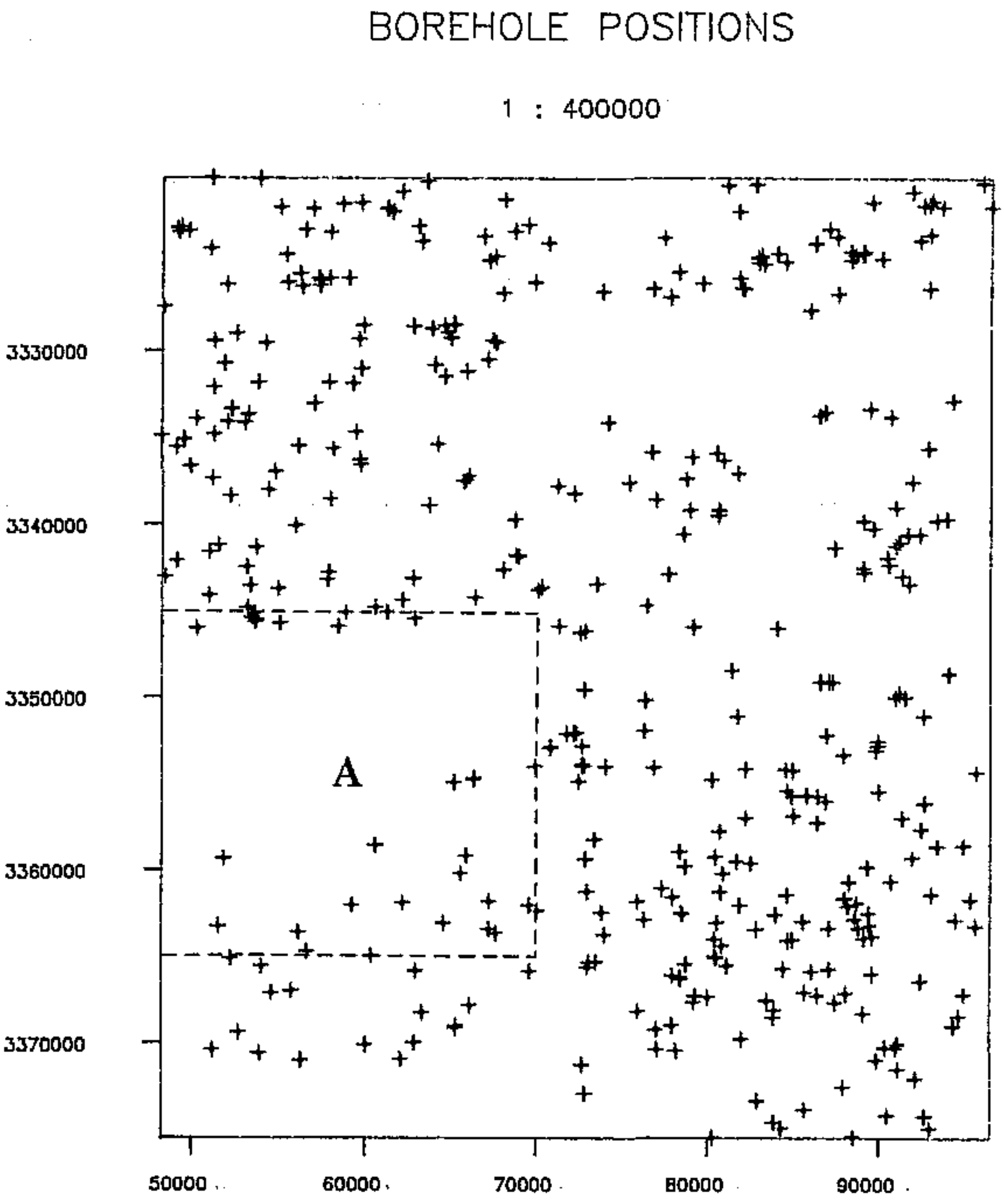


Figure 5.8(b). *Borehole position map of boreholes with recorded water levels in the southern Orange Free State.*

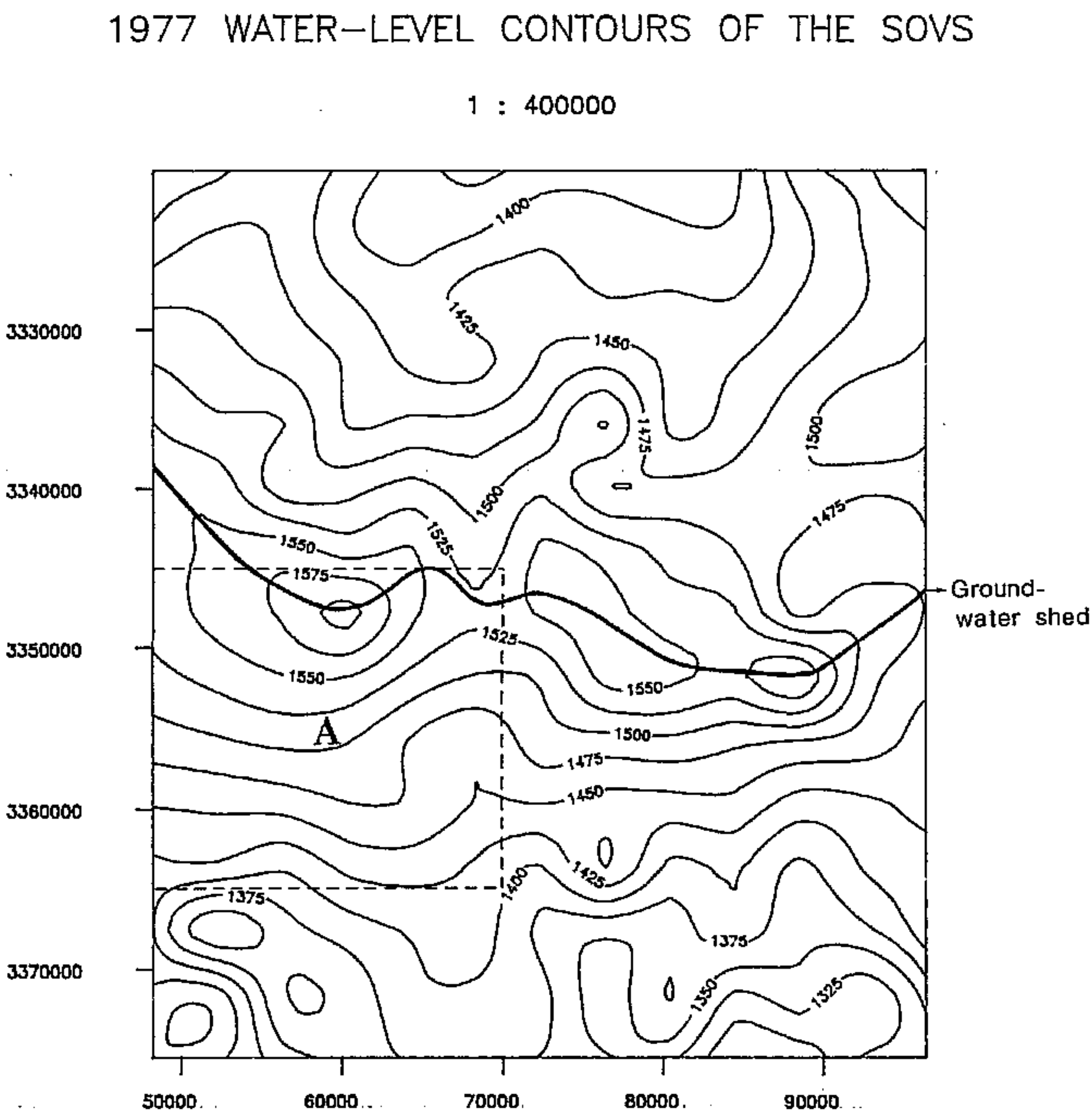


Figure 5.8(c). Water-level contour map of the southern Orange Free State

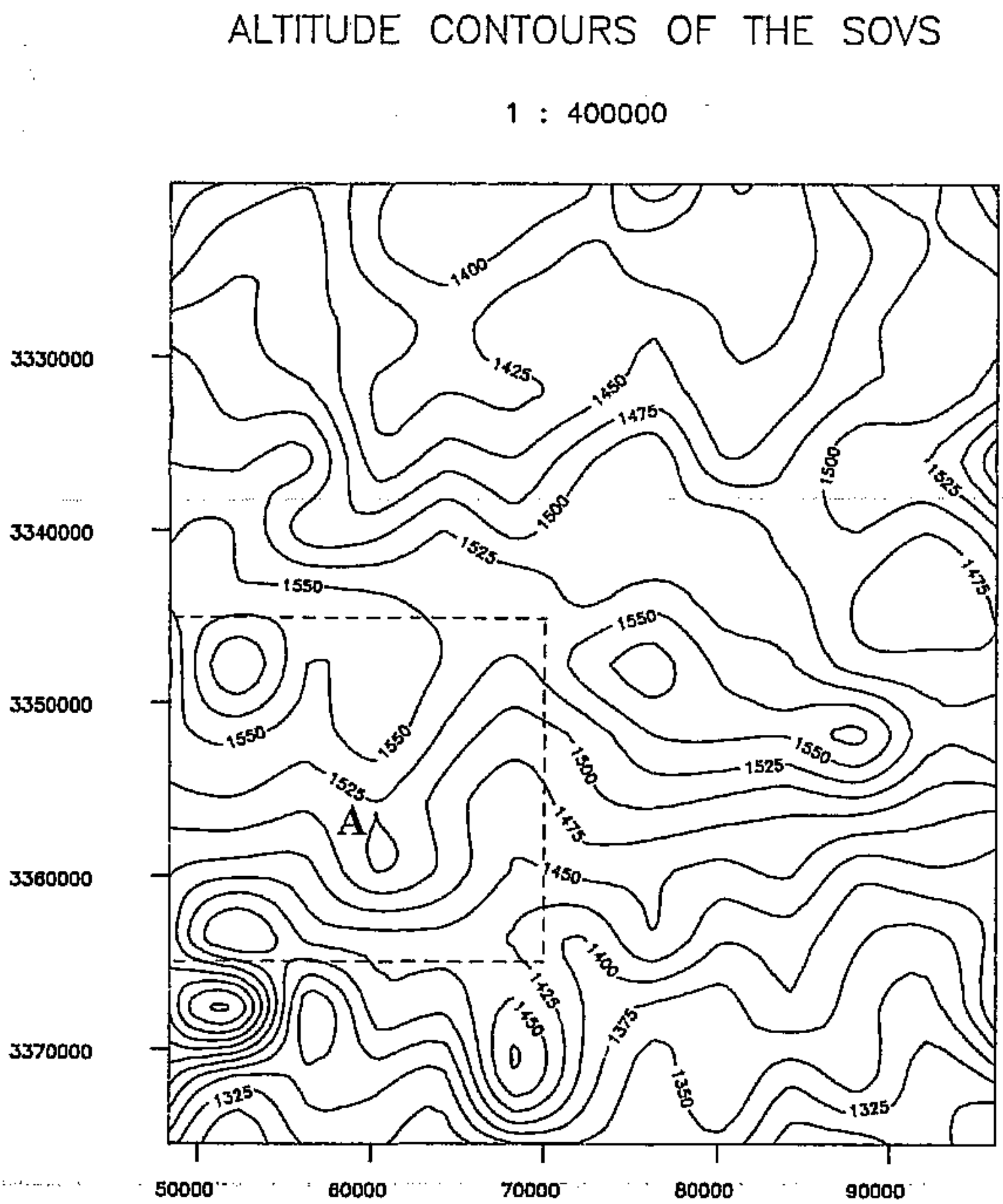


Figure 5.8(d). Topography contour map of the southern Orange Free State.

SIMPLIFIED GEOLOGICAL MAP OF THE SOFS

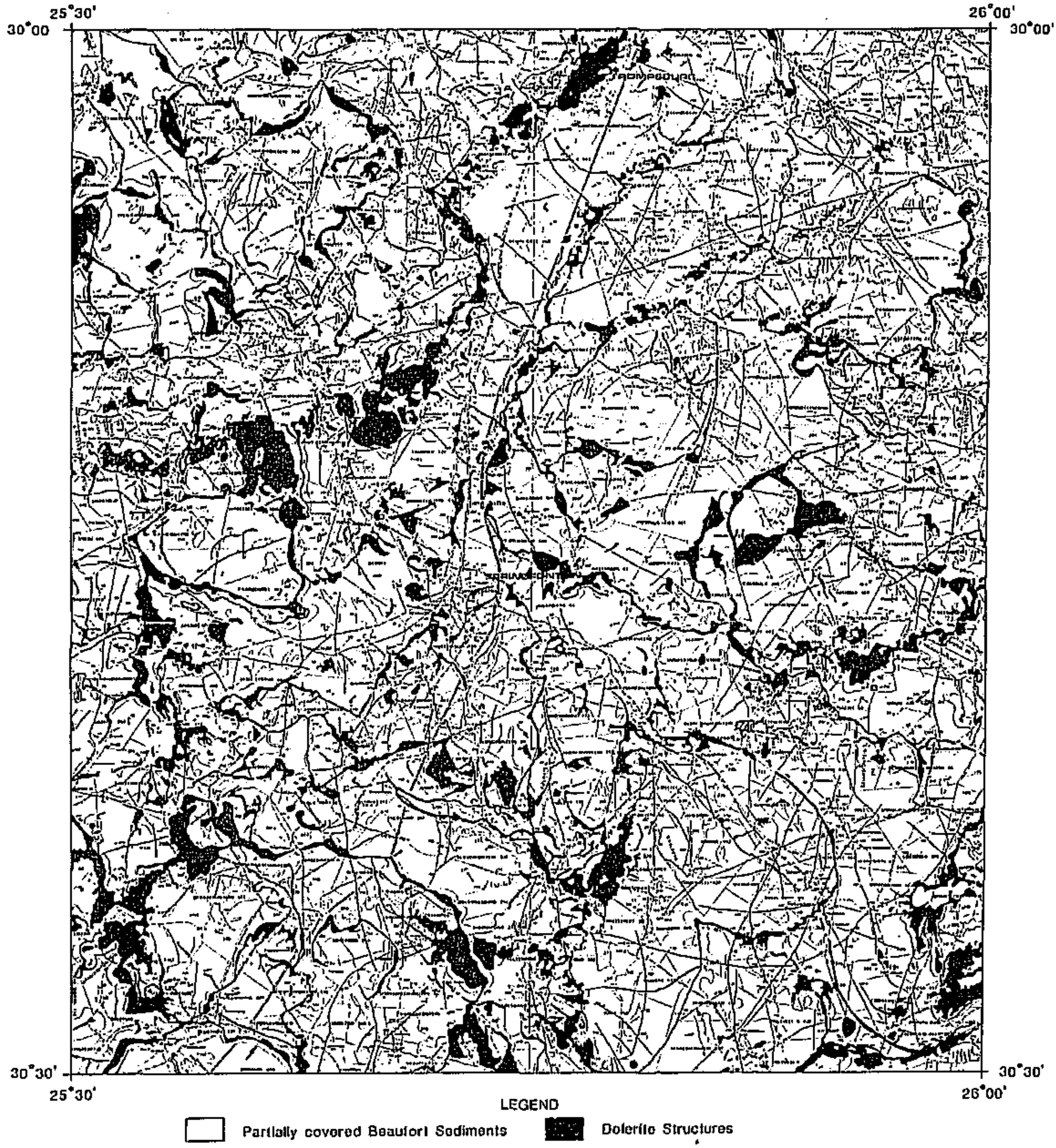


Figure 5.8(e). Simplified geological map of the southern Orange Free State.

Of the 217 boreholes in which the aquifer type was identified, 119 were drilled in Beaufort sediments, 94 in dolerite and 4 in sediments and dolerite. Data files, containing depths to the top of the aquifer and yield, were compiled for both the boreholes in sediments and in dolerite. Table 5.3(a) shows the basic statistics of both groups.

TABLE 5.3(a). TABULATED STATISTICS OF BOREHOLE YIELD DATA FOR SEDIMENTARY AND DOLERITE ASSOCIATED AQUIFERS IN THE SOUTHERN ORANGE FREE STATE.

YIELDS OF SEDIMENTARY AQUIFERS

Number of observations	=	72
Mean yield	=	4,67 l/s
Median	=	2,52 l/s
Standard deviation	=	6,17 l/s

YIELDS OF DOLERITE ASSOCIATED AQUIFERS

Number of observations	=	56
Mean yield	=	5,92 l/s
Median	=	3,25 l/s
Standard deviation	=	6,13 l/s

If the standard deviation in table 5.3(a) is subtracted from the mean value, a negative borehole yield is obtained. This indicates that the borehole yield data are not normally distributed. If figures 5.9(a) and (b) are consulted, one sees that in both cases the distribution of data is skew and follow probably a log-normal distribution. Thus, the standard deviation calculated for the ordinary data is insignificant. The log yield data are evaluated in table 5.3 (b):

TABLE 5.3(b). TABULATED STATISTICS OF THE LOGARITHM OF BOREHOLE YIELD DATA BETWEEN SEDIMENTARY AND DOLERITE ASSOCIATED AQUIFERS IN THE SOUTHERN ORANGE FREE STATE.

YIELDS OF SEDIMENTARY AQUIFERS

Number of observations	=	72
Geometric mean yield	=	2,70 l/s
Standard deviation	=	2,80 l/s

YIELDS OF DOLERITE ASSOCIATED AQUIFERS

Number of observations	=	56
Geometric mean yield	=	3,30 l/s
Standard deviation	=	3,20 l/s

It is noted from the above tables that the median is closer to the geometric mean of the yield distribution and therefore more suitable to indicate the yield, which can be expected from a borehole. It is also noteworthy that the mean yields in the dolerite associated aquifers are slightly higher than those in sedimentary aquifers. Figures 5.9(a) and (b) illustrate the yield distribution of both aquifers in the form of histograms.

It is generally accepted that the ground-water occurrence in the Beaufort sediments is associated with dolerite structures. This may not be necessarily true for all regions, as indicated by the statistics of the Southern Orange Free State. Furthermore, if the lithology and the depth of water struck are correlated with each other, it is found that the ground water present in the sediments cannot be associated with any particular lithological horizon.

According to Vegter (1984), the Beaufort sediments have a primary porosity ranging from 0,012 to 0,014 and a permeability of less than 1×10^{-9} to 1×10^{-8} m/s, i.e. virtually no primary porosity. Water encountered in the sedimentary aquifers of the Beaufort Subgroup is therefore mainly derived from secondary aquifers, i.e. from fractures and joints in the sedimentary rocks.

In table 5.4 and figure 5.10, similar statistics are displayed for all yield data in the region. A yield contour map is also constructed with the same data (see figure 5.11(a)). Figures 5.11(b) and (c) show contour maps of the borehole depth and the depth at which water was struck respectively.

TABLE 5.4. BOREHOLE YIELD STATISTICS OF THE SOUTHERN ORANGE FREE STATE.

Number of observations	=	814
Geometric mean yield	=	2,20 l/s
Standard deviation	=	3,50 l/s

If the mentioned contour maps are studied and compared with the geological and the water-level map, shown in figures 5.8(b) and (d), the following can be deduced:

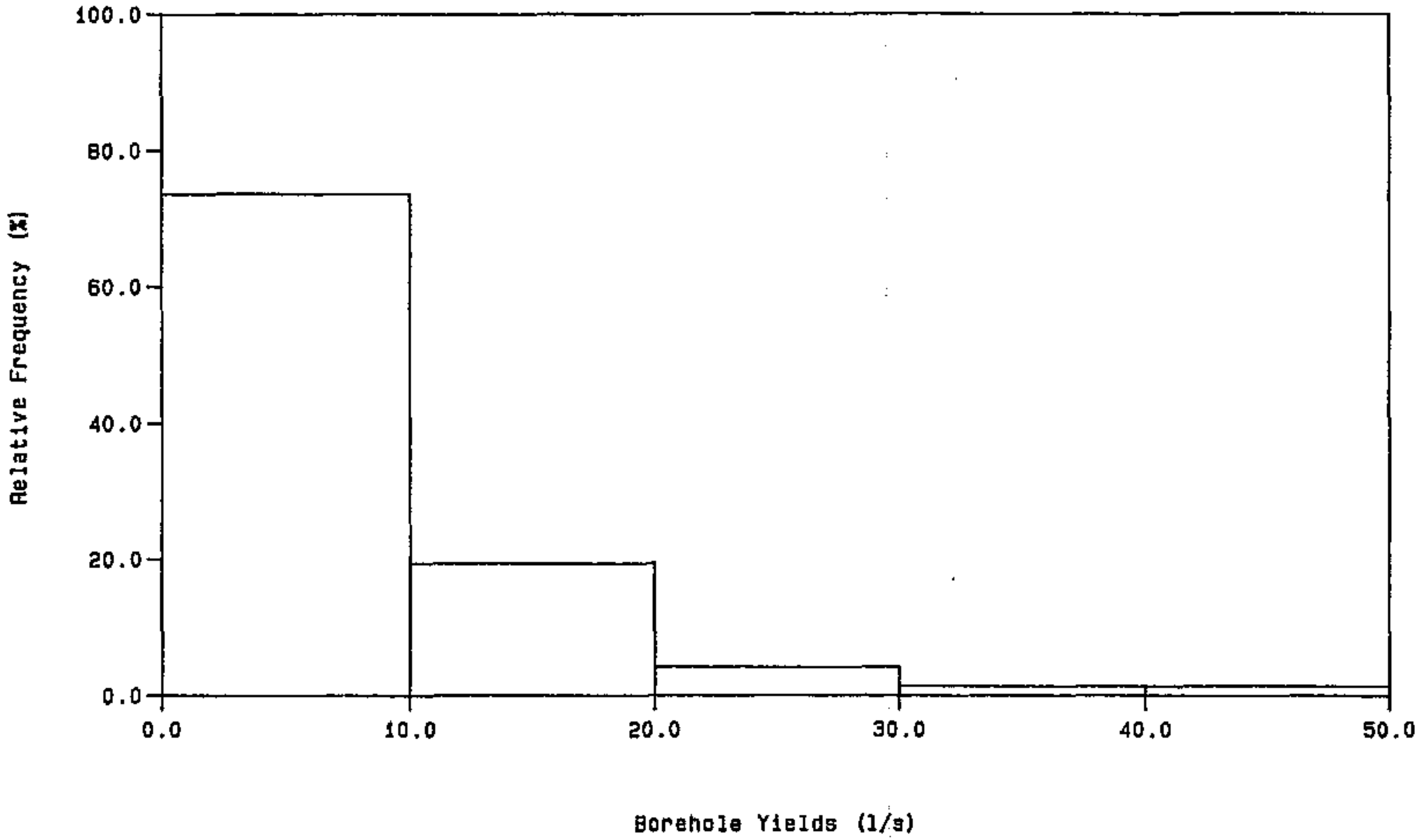


Figure 5.9(a). Histogram of yields in sedimentary aquifers in the southern Orange Free State

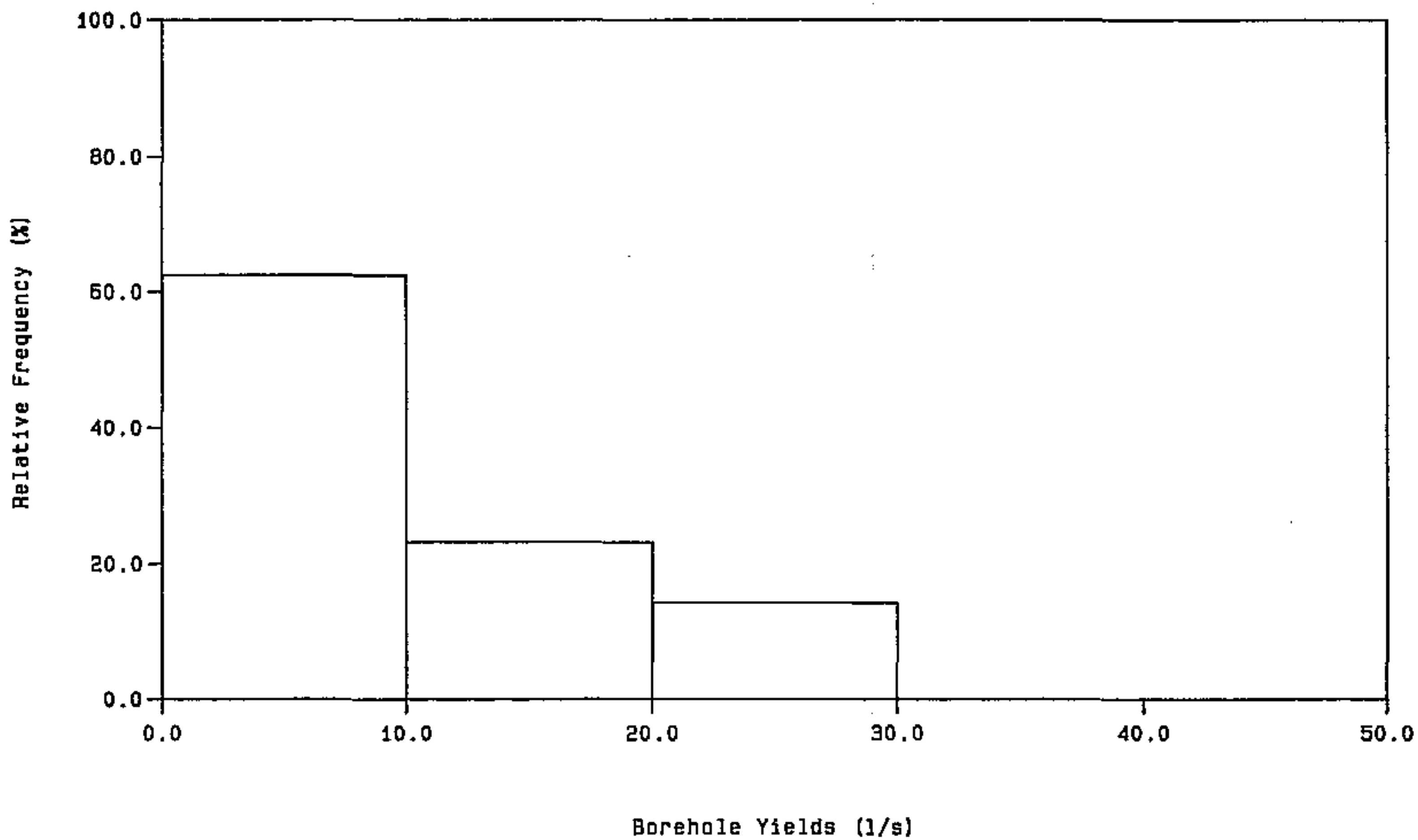


Figure 5.9(b). *Histogram of yields in dolerite associated aquifers in the southern Orange Free State.*

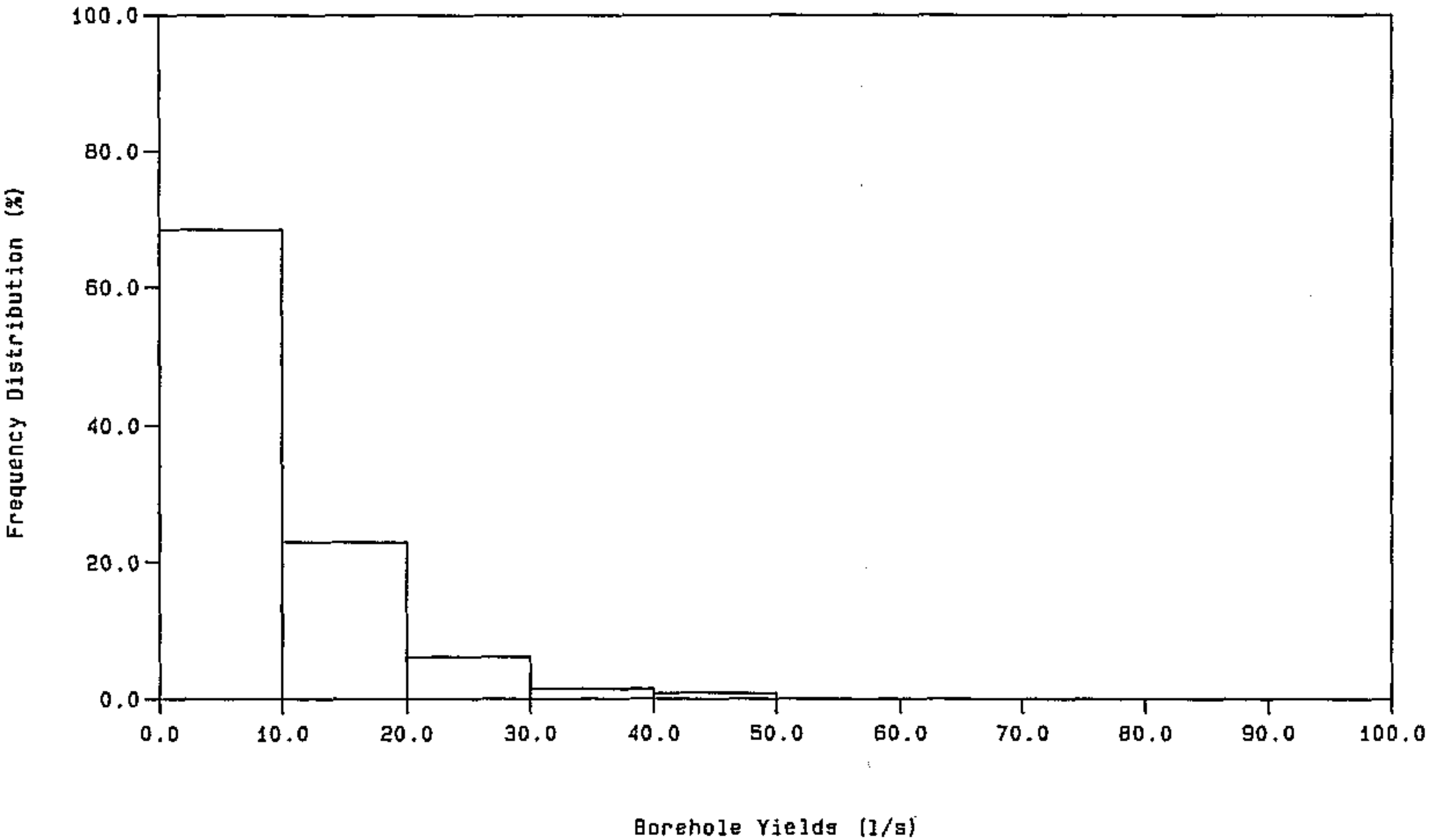


Figure 5.10. *Frequency distribution of borehole yields in the southern Orange Free State.*

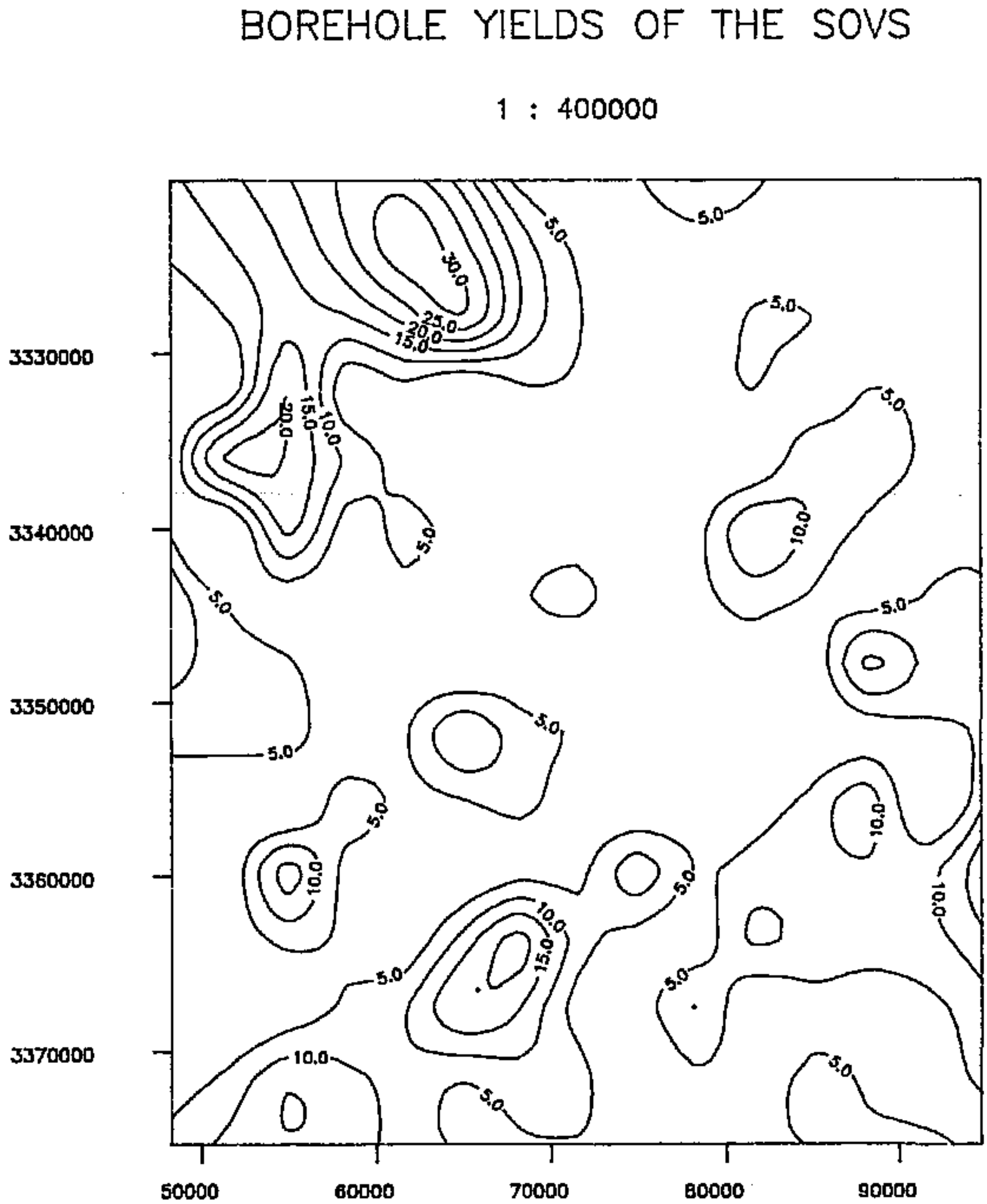


Figure 5.11(a). *Yield (l/s) contour map of the southern Orange Free State.*

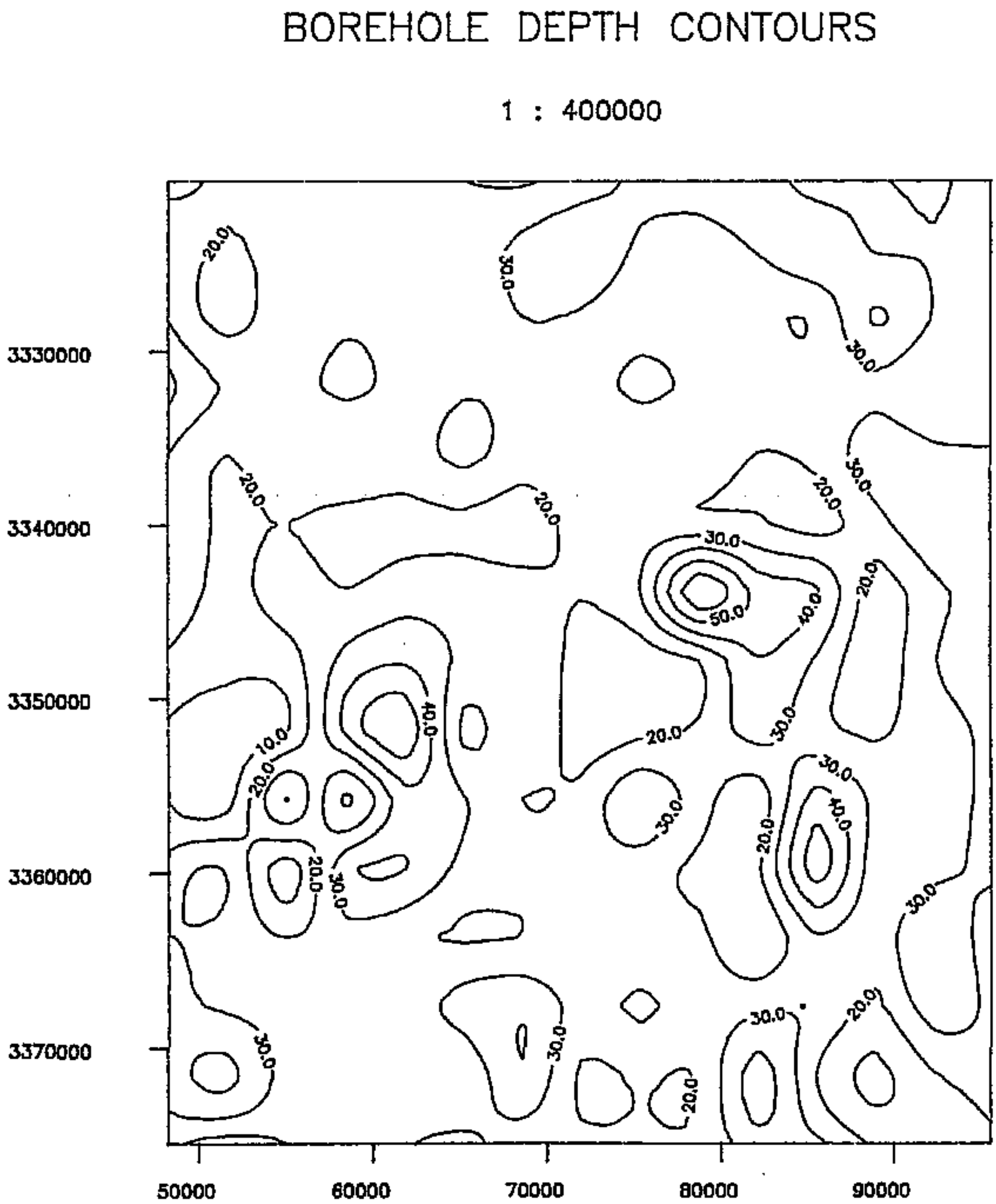


Figure 5.11(b). Borehole depth (m) contour map of the southern Orange Free State.

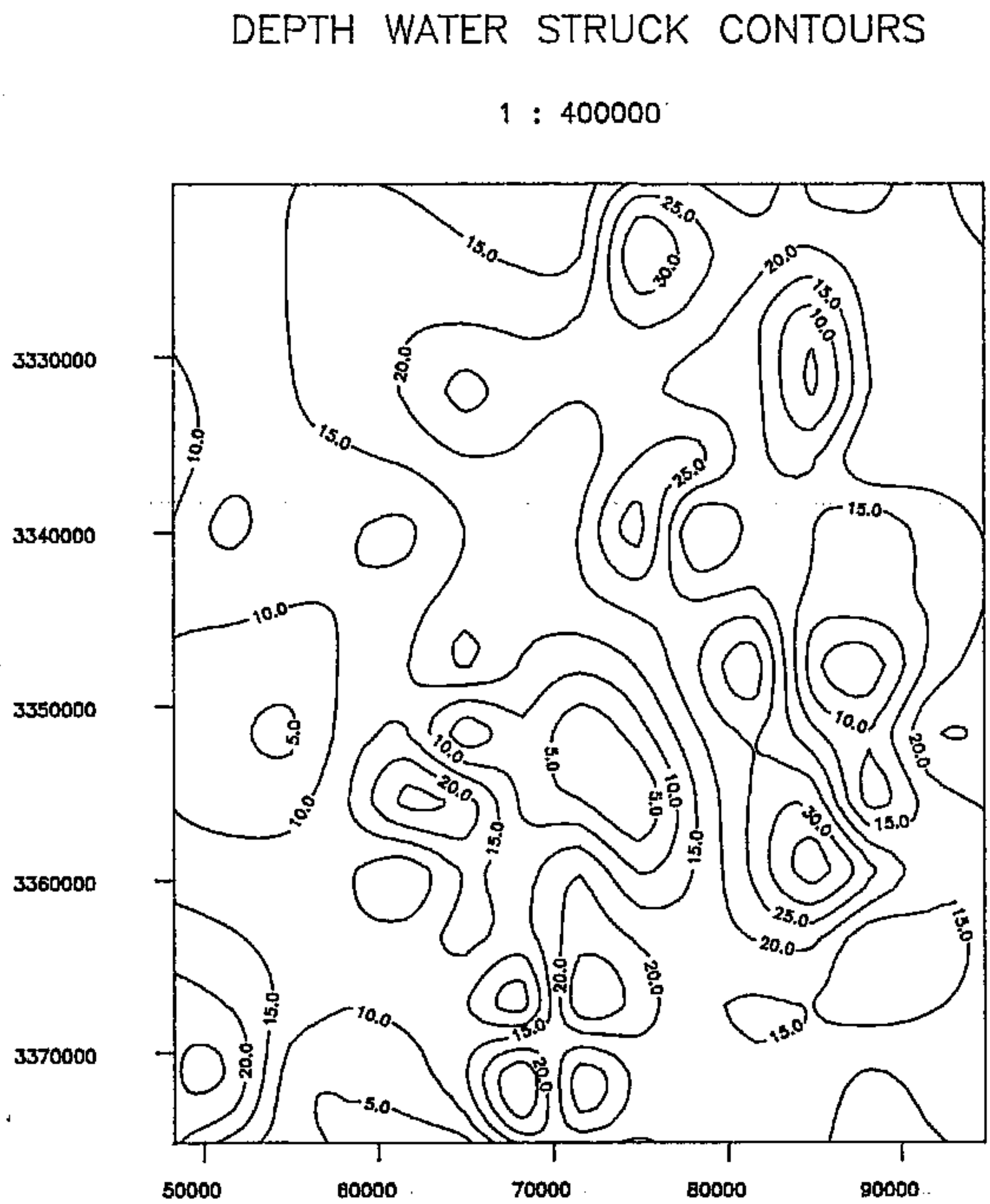


Figure 5.11(c). *Depth water struck (m) contour map of the southern Orange Free State.*

(a) At the scale portrayed on the maps there is no correlation between borehole yield and geology.

(b) There is no correlation (correlation coefficient -0,0597) between borehole yield and borehole depth in the area.

(c) There is no correlation (correlation coefficient -0,3291) between borehole yield and the depth at which water was struck in the area.

(d) The lower borehole yields tend to correlate with the watershed in the area, as shown in figure 5.8(b).

These deductions support the assumption that aquifers in this area are of secondary nature, i.e. water is found in fractures, joints and dolerite contact zones.

Boreholes in the region are mainly used for agricultural and household purposes. Table 5.5 summarizes the water applications.

TABLE 5.5. SUMMARIZED WATER APPLICATIONS.

Water Application	Number of boreholes
Agricultural and domestic (AD)	78
Agricultural - irrigation only (AI)	264
Agricultural - stock watering only (AS)	676
Domestic - all purposes (DA)	82

When data from the southern Orange Free State were evaluated, data from one pumping test were available. A representative value or the range of values for transmissivity and storativity of the area can therefore not be given. The recorded pumping test was performed in the surrounding area of Phillippolis. Five observation boreholes and the production borehole were used to calculate the hydraulic parameters. Transmissivity values for this test ranged from 48 m^2/d to 100 m^2/d, whereas the storativity varied considerably between $1,2 \times 10^{-2}$ and $1,0 \times 10^{-5}$.

In conclusion, it may be said that the information contained in the data base provides a good overview on a local scale, regarding the distribution of boreholes, borehole yields, depth at which water was struck, geology and borehole use in the area. This information is quite useful, when new borehole sites must be selected or when a new ground-water project is initiated in the area. A sufficient number of (long-term) water-level measurements and pumping-test data for this region have yet to be entered into the data base. These are necessary for a proper evaluation of the ground-water resources of square degree 3025.

Atlantis area

If the aquifer data set of the Atlantis area is evaluated, one finds that 105 boreholes have recorded aquifers. Three different aquifers were identified:

(i) Bredasdorp Formation.

(ii) Varswater Formation.

iii) Tygerberg Formation.

Thus, with the exception of the Darling Granite, all geological units in the area are aquifers. Depending on their respective hydraulic conductivities, the aquifers can be dealt with as three separate aquifers or be combined into one aquifer. If, however, the hydraulic conductivity of a specific formation is much lower than that of others, this formation must be regarded as an aquitard.

In 85 of the 105 boreholes, more than one aquifer was recorded, whereas in the remaining 20 boreholes, 19 boreholes were drilled in Cenozoic deposits and one in the Tygerberg Formation.

The Bredasdorp Formation is water bearing in 103 boreholes (see table 5.6). Due to its occurrence in the majority of the boreholes, this formation is the most wide-spread aquifer in the region. According to Bredenkamp and Vandoolaeghe (1982), the Springfontein Member of the Bredasdorp Formation was identified as the main water-bearing stratum in the area, because it consists of coarse-grained sediments (see section 5.2.1).

TABLE 5.6. AQUIFER TYPES ENCOUNTERED IN ATLANTIS BOREHOLES

Identified aquifer types per borehole	Number of boreholes
Bredasdorp Formation Varswater Formation Tygerberg Formation	39
Bredasdorp Formation Varswater Formation	14

Bredasdorp Formation Tygerberg Formation	31
Varswater Formation Tygerberg Formation	1
Bredasdorp Formation	19
Tygerberg Formation	1
TOTAL	105

When the data were evaluated, only a limited number of pumping-, bail- and slug tests were entered into the ground-water data base. As can be seen from table 5.7, the transmissivity values in the Atlantis area vary between 140 m^2/d and 730 m^2/d, whereas the storativity ranges between $1{,}6 \times 10^{-2}$ and $2{,}9 \times 10^{-2}$.

For a better estimation of the potential of the different aquifers, data from the Discharge rate- and Pumping test data sets must also be taken into account. Borehole yields from Atlantis have, however, not yet been recorded in the National Ground-water Data Base. Due to the fact that the Atlantis aquifers are phreatic aquifers, layer thickness and hydraulic conductivity can be used to rate the different aquifers. Of the three identified aquifer units, the Bredasdorp Formation is the thickest with a mean thickness of 27,3 m, in contrast to the Varswater Formation's mean thickness of 8,5 m. If the transmissivity values of the pumping-test data set are divided by the aquifer thickness, an average value for the hydraulic conductivities of the respective formations can be derived at. The calculated hydraulic conductivity of the Bredasdorp Formation then becomes 14 m/d, whereas the corresponding value for Varswater Formation is 5,5 m/d. Based on the information contained in the data base, no hydraulic conductivity value can be given for the Tygerberg aquifer. Consequently, the Bredasdorp Formation is to be regarded the main water-bearing formation in the Atlantis area; a conclusion which coincides with the findings of Bredenkamp and Vandoolaeghe (1982).

Of the approximately 260 recorded boreholes in the Atlantis area, water levels have been entered for 155 boreholes. Recorded water levels are, however, not evenly distributed over the area. Many of the observation boreholes are clustered around the major abstraction areas and the sites at which water-levels were repeatedly measured, have also changed considerably over the years. Contour maps which show the water-level elevations for a number of successive years are therefore difficult to compile and compare. This is illustrated by the series of water-level contour maps of the Atlantis area, which depict water levels between 1980 and 1986 (see figures 5.13, 5.14, 5.15, 5.16, 5.17 and 5.18; water levels for the year 1985 are not shown, because there were not sufficient measurements for this year). If the borehole positions, indicated by crosses on

TABLE 5.7. TABULATED PUMPING TEST RESULTS OF THE ATLANTIS AREA.

Site Id.	Date Test Started	Method Tested	Pump Intake Depth (m)	Rate Rec. (l/s)	Transmissivity (m^2/d)	Storativity	Reporting Inst.
3318AD00094	19861212	S	0.00	0.0	141	0	WAGC
3318CB00001	19830000	P	0.00	5.0	0	0	WAGC
3318CB00001	19790709	P	23.36	9.7	0	0	WAGC
3318CB00002	19800111	P	22.30	17.1	0	0	WAGC
3318CB00003	19790502	P	28.29	20.0	0	0	WAGC
3318CB00012	19840106	P	0.00	12.0	0	0	WAGC
3318CB00019	19790619	P	20.70	25.0	0	0	WAGC
3318CB00021	19810217	P	23.40	15.0	0	0	WAGC
3318CB00022	19790520	P	25.20	15.5	0	0	WAGC
3318CB00024	19840120	P	37.00	0.0	0	0	WAGC
3318CB00025	19840130	P	26.50	34.0	0	0	WAGC
3318CB00029	19831103	P	39.00	10.0	0	0	WAGC
3318CB00030	19831026	P	33.00	27.0	0	0	WAGC
3318CB00031	19840119	P	41.00	81.0	0	0	WAGC
3318CB00032	19831110	P	38.80	18.0	0	0	WAGC
3318CB00035	19830105	P	29.00	18.0	611	292	WAGC
3318CB00036	19840126	P	37.40	0.0	0	0	WAGC
3318CB00037	19790910	P	21.45	14.0	421	162	WAGC
3318CB00038	19821221	P	0.00	19.0	0	0	WAGC
3318CB00045	19790322	P	23.05	21.3	0	0	WAGC
3318CB00046	19790402	P	18.45	20.0	0	0	WAGC
3318CB00047	19790628	P	10.75	0.0	0	0	WAGC
3318CB00048	19800125	P	22.40	17.7	0	0	WAGC
3318CB00049	19830302	P	0.00	19.0	731	292	WAGC
3318CB00050	19830420	P	36.00	19.0	0	0	WAGC
3318CB00051	19830309	P	26.50	17.0	0	0	WAGC
3318CB00052	19830314	P	23.00	12.0	301	232	WAGC
3318CB00053	19830831	P	0.00	0.0	0	0	WAGC
3318CB00054	19840113	P	30.00	12.0	0	0	WAGC
3318CB00055	19840124	P	37.60	35.4	0	0	WAGC
3318CB00056	19840201	P	40.00	18.3	0	0	WAGC
3318CB00057	19831208	P	41.00	70.1	0	0	WAGC
3318CB00080	19860903	S	0.00	0.0	0	0	WAGC
3318CB00088	19860529	S	0.00	0.0	0	0	WAGC
3318CB00090	19860827	S	30.00	0.0	0	0	WAGC
3318CB00091	19860903	S	0.00	0.0	0	0	WAGC
3318CB00104	19860602	S	39.00	0.0	0	0	WAGC
3318CB00110	19830317	P	0.00	15.0	611	272	WAGC
3318CB00112	19781031	P	21.50	13.0	0	212	WAGC
3318CB00116	19840705	P	30.00	24.3	0	0	WAGC
3318CB00126	19780713	P	19.27	0.0	0	0	WAGC
3318CB00189	19830823	P	40.00	65.0	0	0	WAGC
3318CB00193	19860801	S	0.00	0.0	714	0	WAGC
3318CB00196	19860730	S	0.00	0.0	271	0	WAGC
3318CB00197	19860801	S	0.00	0.0	254	0	WAGC
3318CB00201	19860801	S	0.00	0.0	0	0	WAGC
3318CB00204	19860701	S	0.00	0.0	0	0	WAGC
3318CB00205	19860321	S	0.00	0.0	159	0	WAGC
3318CB00209	19860801	S	0.00	0.0	211	0	WAGC
3318CB00213	19860729	S	0.00	0.0	0	0	WAGC
3318CB00215	19860730	S	0.00	0.0	209	0	WAGC
3318CB00216	19860801	S	0.00	0.0	0	0	WAGC
3318CB00218	19860801		0.00	0.0	0	0	WAGC
3318CB00219	19860701	S	0.00	0.0	0	0	WAGC

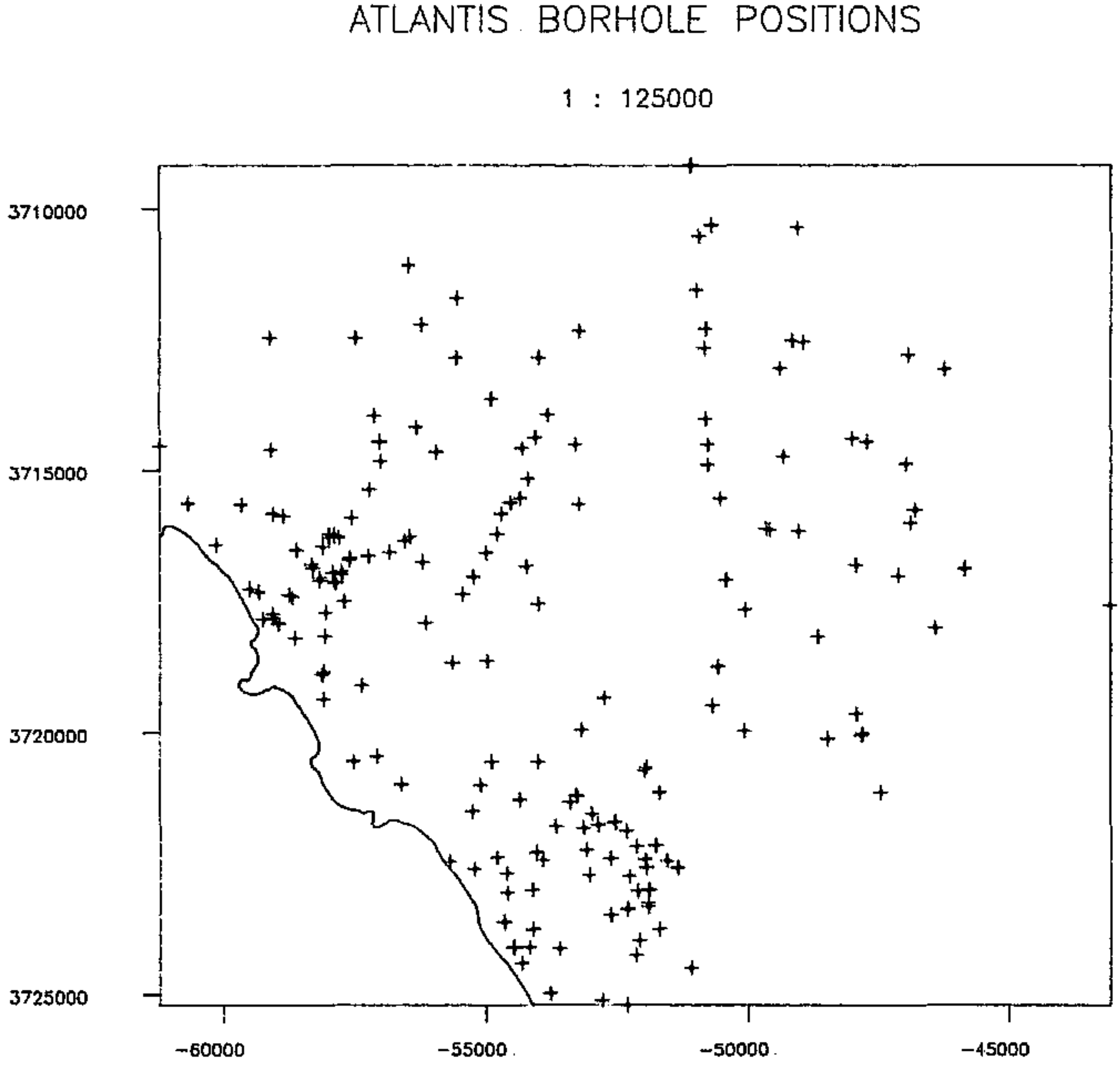


Figure 5.12. Atlantis borehole positions.

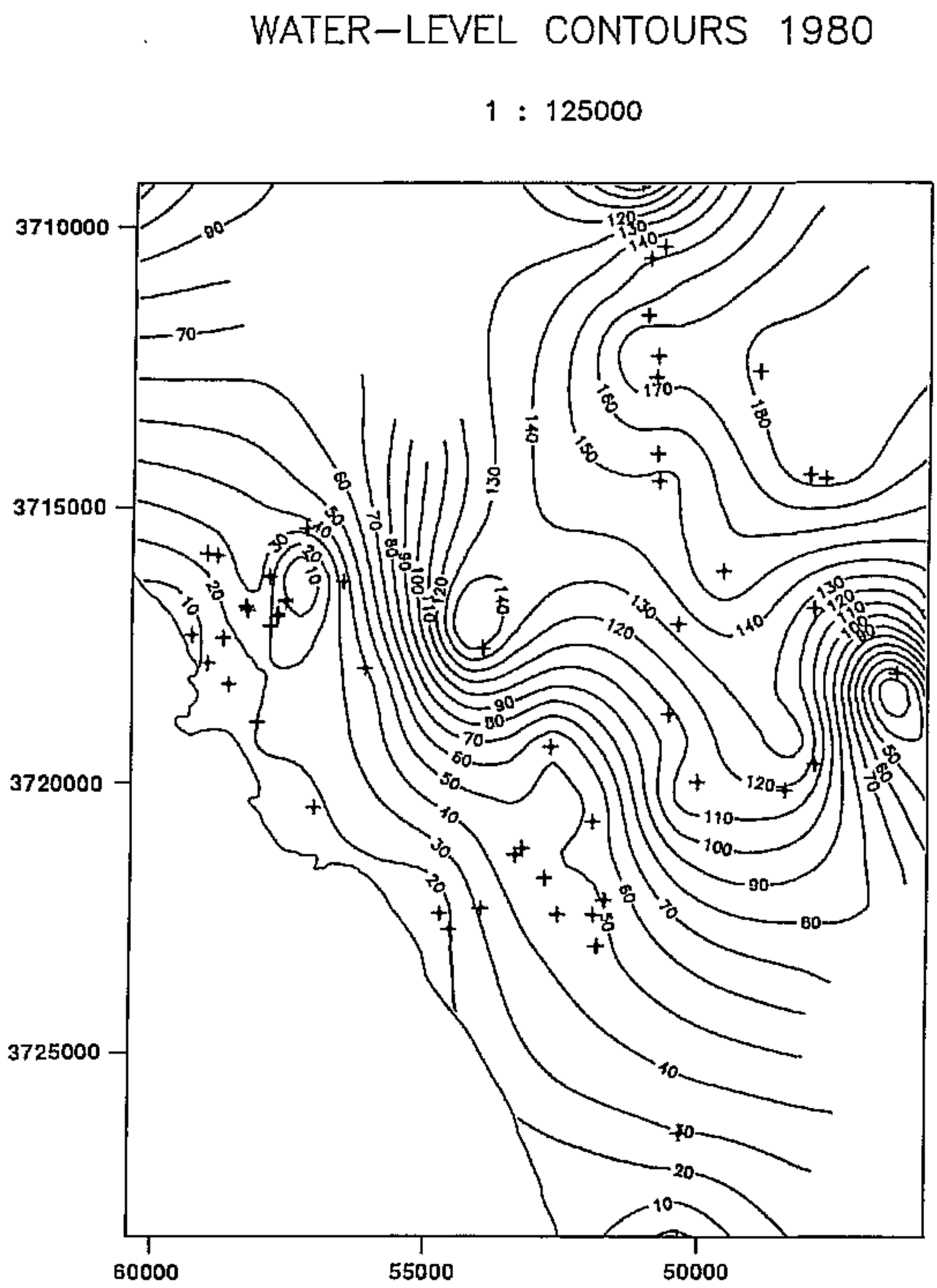


Figure 5.13. *Atlantis water-level contours (October, 1980).*

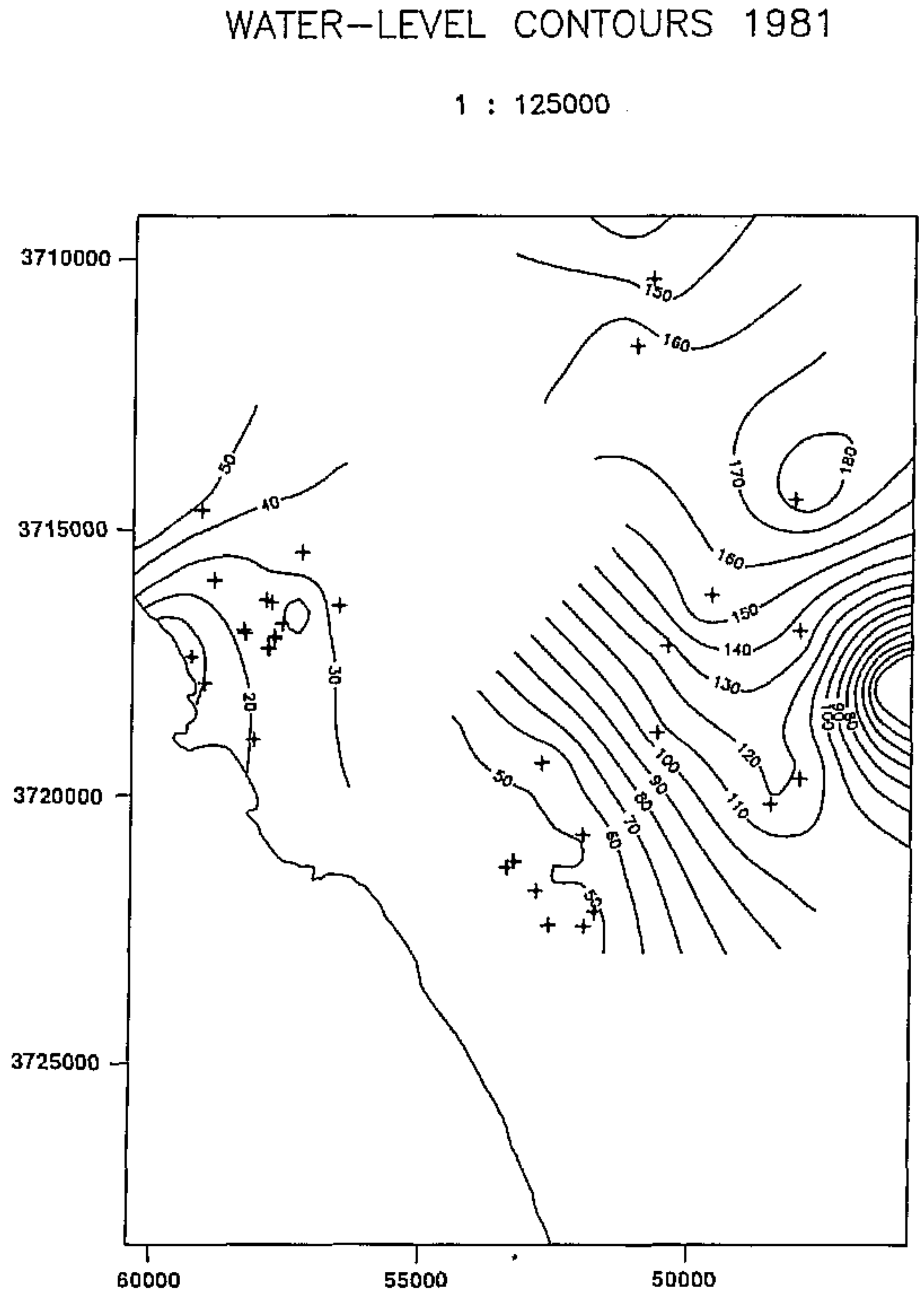


Figure 5.14. Atlantis water-level contours (October, 1981).

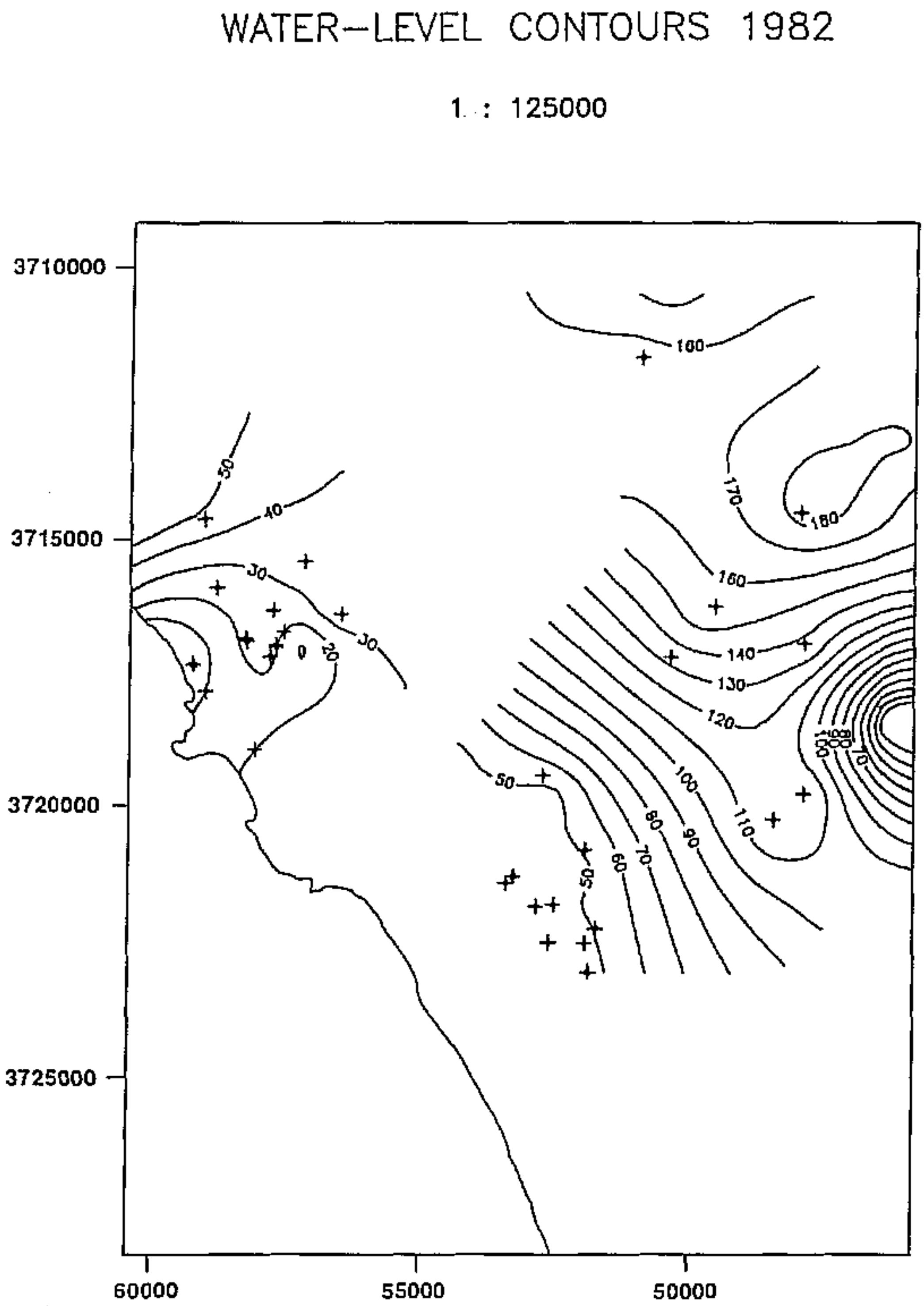


Figure 5.15. Atlantis water-level contours (October, 1982).

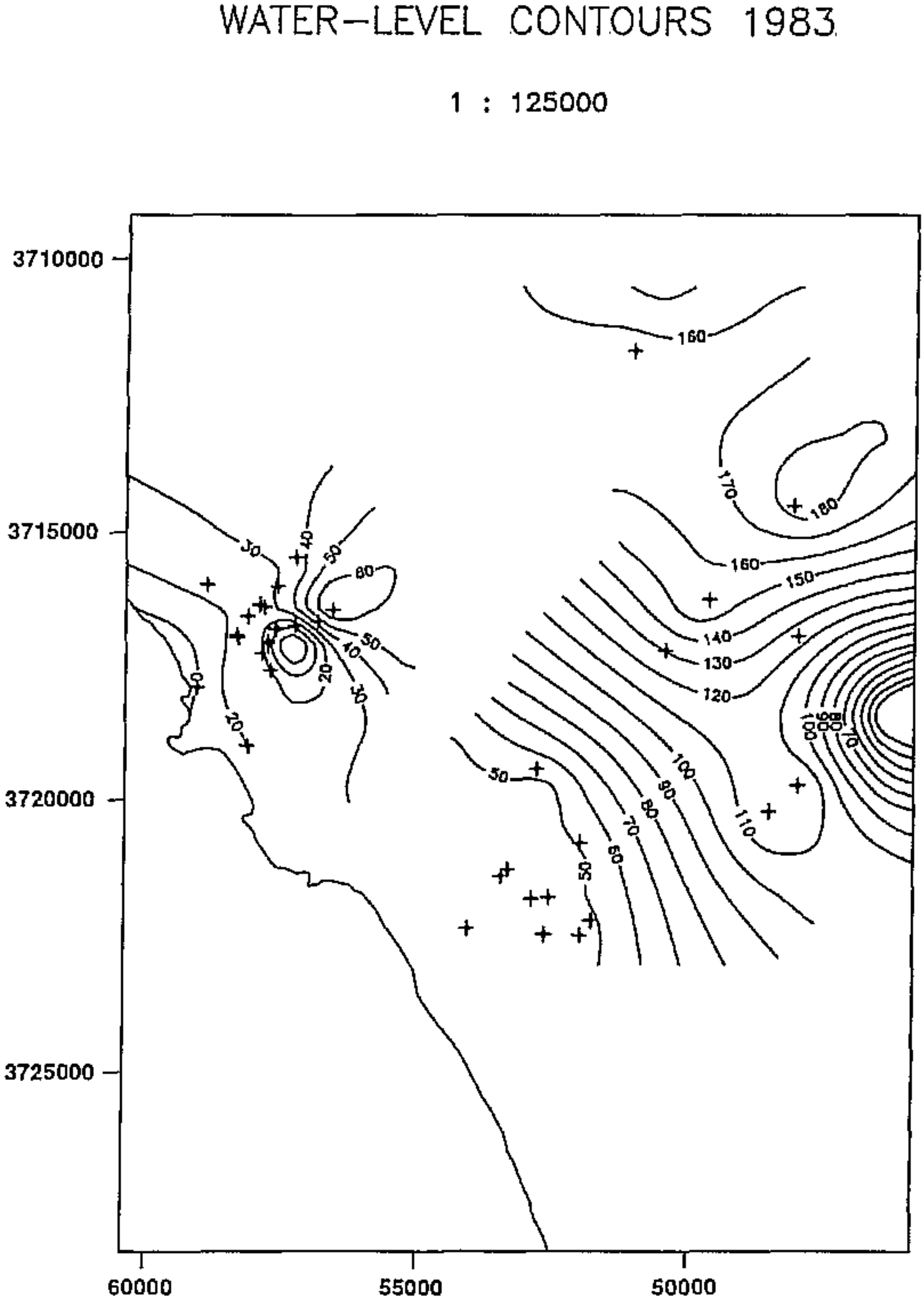


Figure 5.16. Atlantis water-level contours (October, 1983).

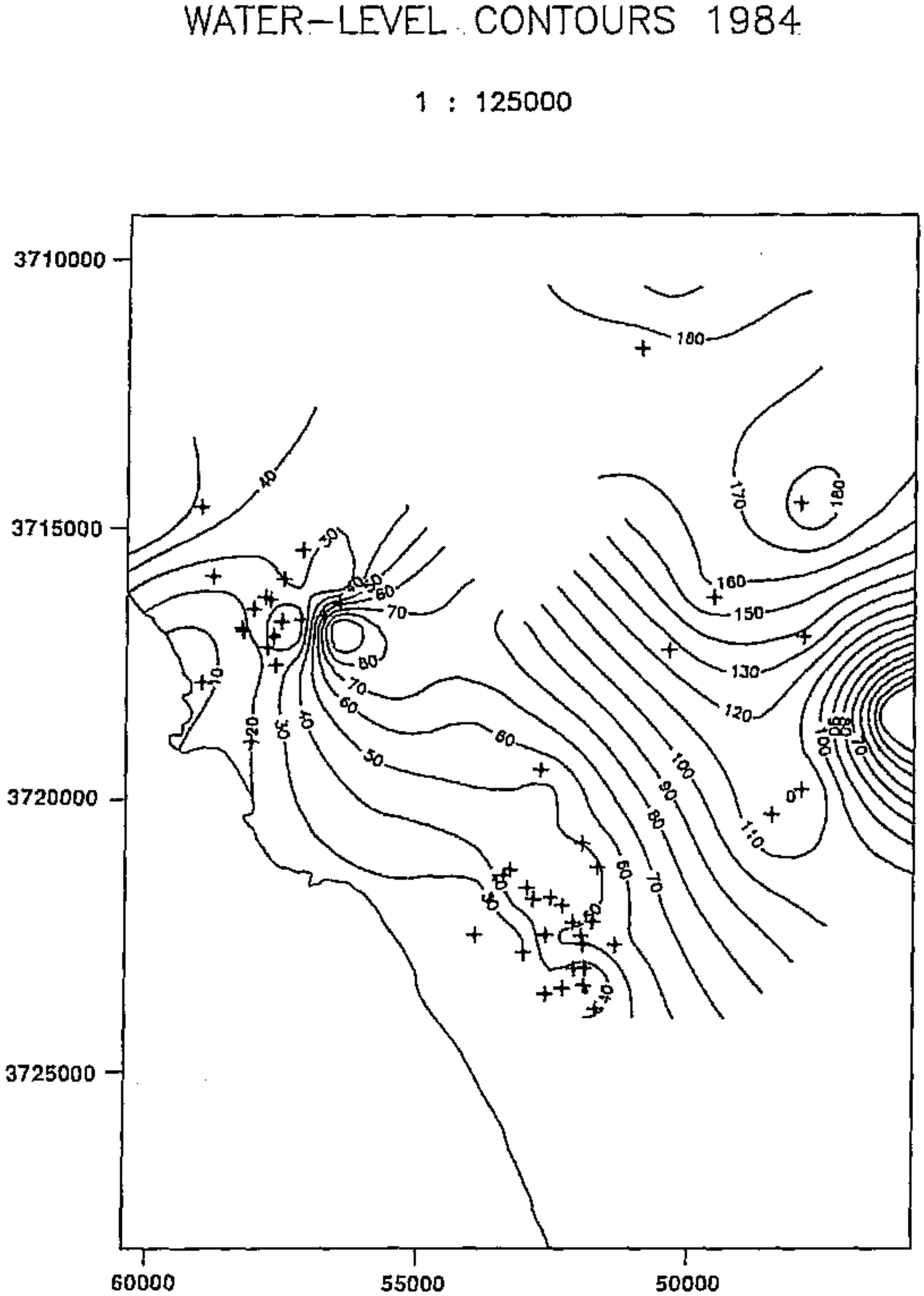


Figure 5.17. Atlantis water-level contours (October, 1984).

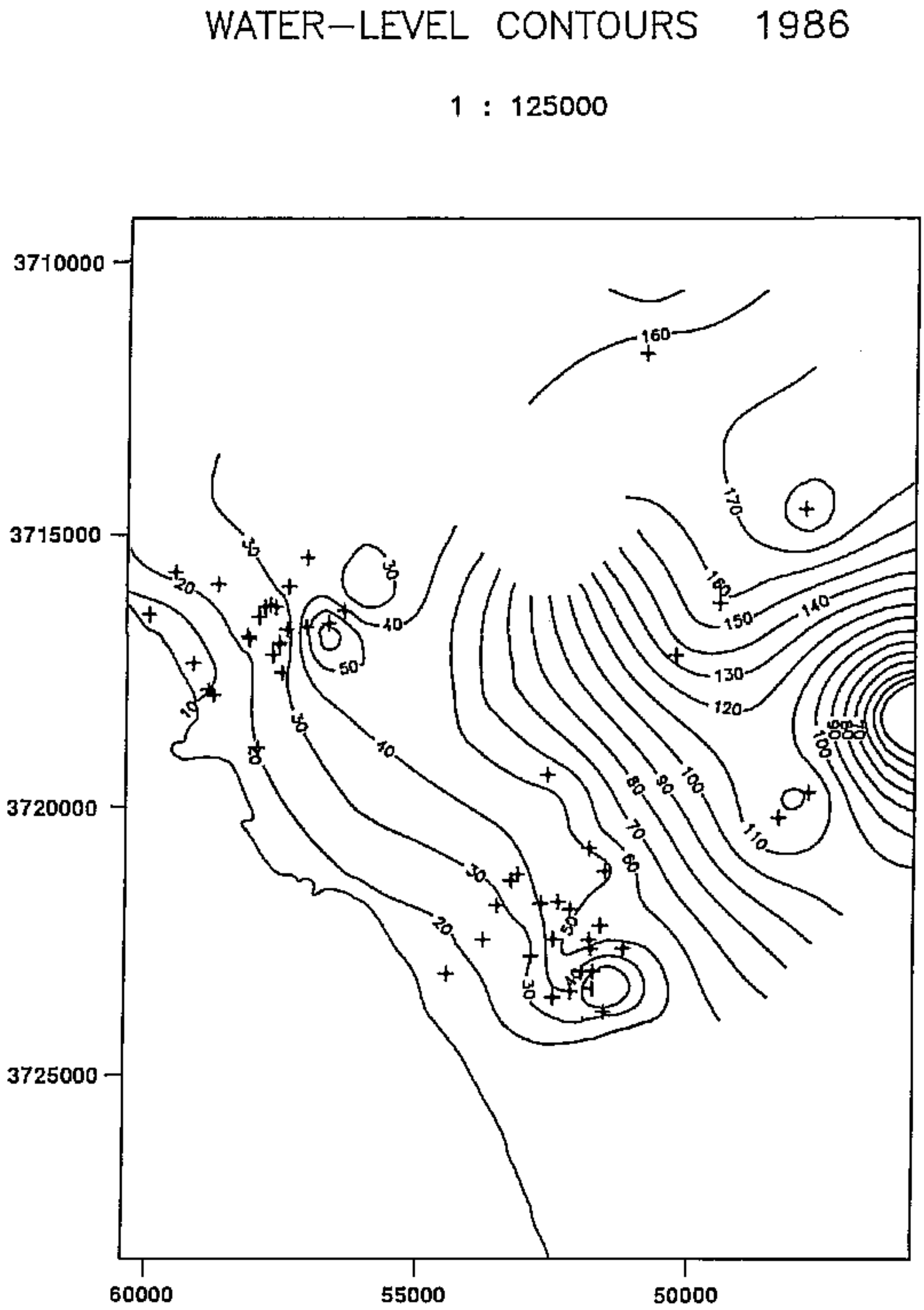


Figure 5.18. Atlantis water-level contours (July, 1986).

the respective water-level contour maps, are compared with those on figure 5.12, the uneven distribution and clustering of the observation boreholes become evident. Because of this uneven and varying distribution, any deductions, based on the water-level maps, must be regarded with caution.

A relatively steep ground-water gradient (see figures 5.13 and 5.18) is observed almost parallel to the coast-line, which separates the Atlantis-Wesfleur production area from the Silwerstroom- and Witzand production areas. The map in figure 5.19 delineates these production fields and also shows the position of the four boreholes for which water-level hydrographs were reproduced (see figures 5.20, 5.21, 5.22 and 5.23). Water-level fluctuations (figure 5.22) reflect the different extraction and recovery periods, monitored in borehole 3318CB00048. Except for the effects of pumping, the hydrographs show little water-level variations over the years and support the impression obtained from the contour maps.

The water-level records from the area depict rather small, long-term variations inland, but a noticeable response to pumping in the coastal production areas with slow recovery in the Silwerstroom area was observed. In spite of the abstractions taking place, a relatively steep gradient is observed between the coastal production areas and the sea. The steep gradient minimizes the danger of possible sea-water intrusion.

The data can also be used to identify deficiencies in the operation of the well field and the proper exploitation of the aquifer. While it is not the aim of the study to re-evaluate management techniques currently used in the exploitation and operation of the Atlantis well field, it can be demonstrated how the data base and its processing facility can be utilized to design and optimize, e.g. the monitoring network, using only the data at hand, i.e. data from the ground-water data base. A good observation network has the advantage of protecting the aquifer through:

- early warnings in the case of overpumping,
- early warnings in the case of intrusion of sea-water or salty ground water from the underlying Malmesbury Group, and
- indication of the best extraction rates for long-term usage of boreholes and aquifers.

For example, if the distribution of the observation boreholes in figures 5.13 and 5.18 is compared with one another, it is noted that the distribution of the observation boreholes has changed remarkably. Both maps encompass relatively large areas where no data have been collected. This leads to an uncertainty regarding the water level and the ground-water

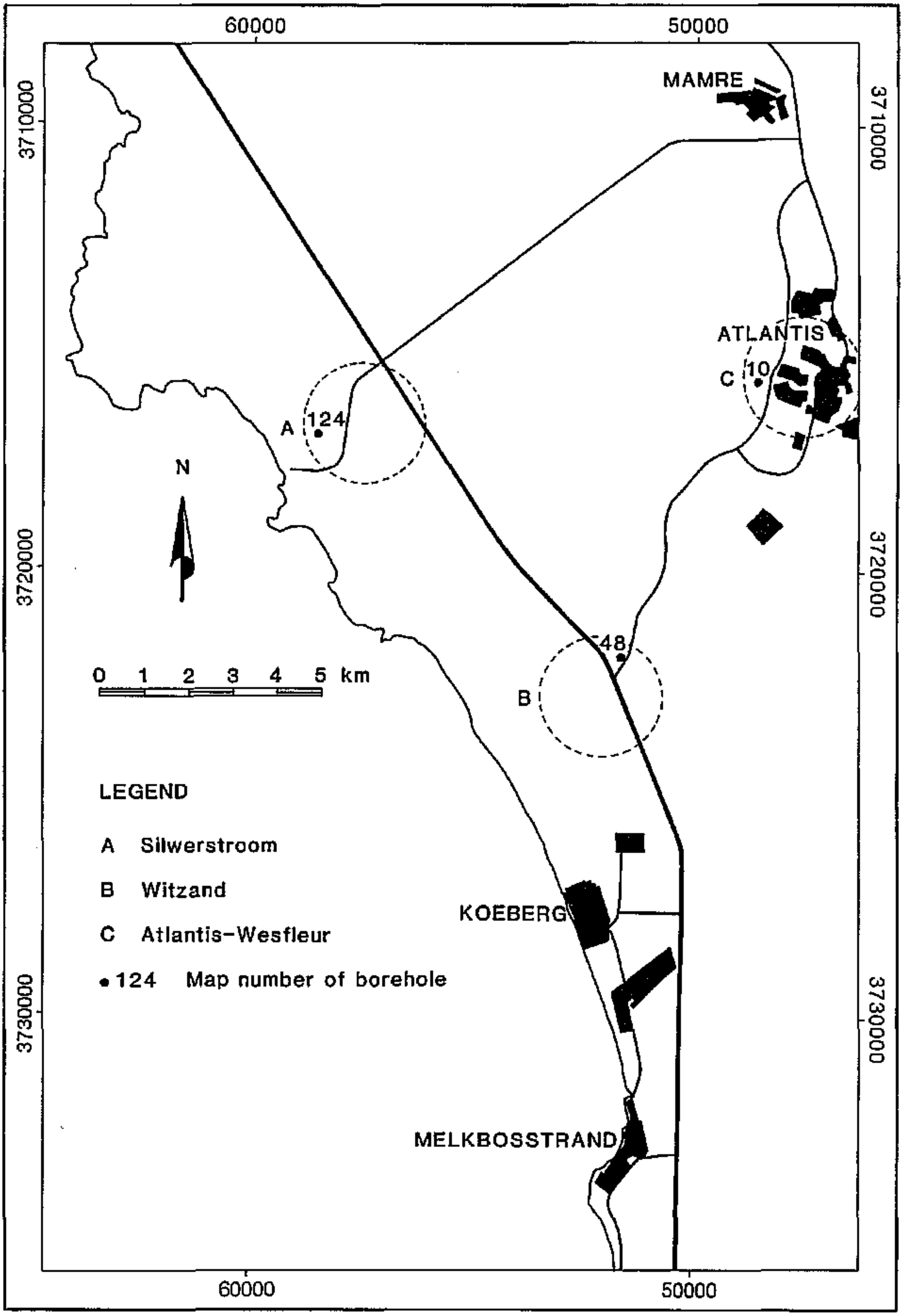


Figure 5.19. Well field delineation.

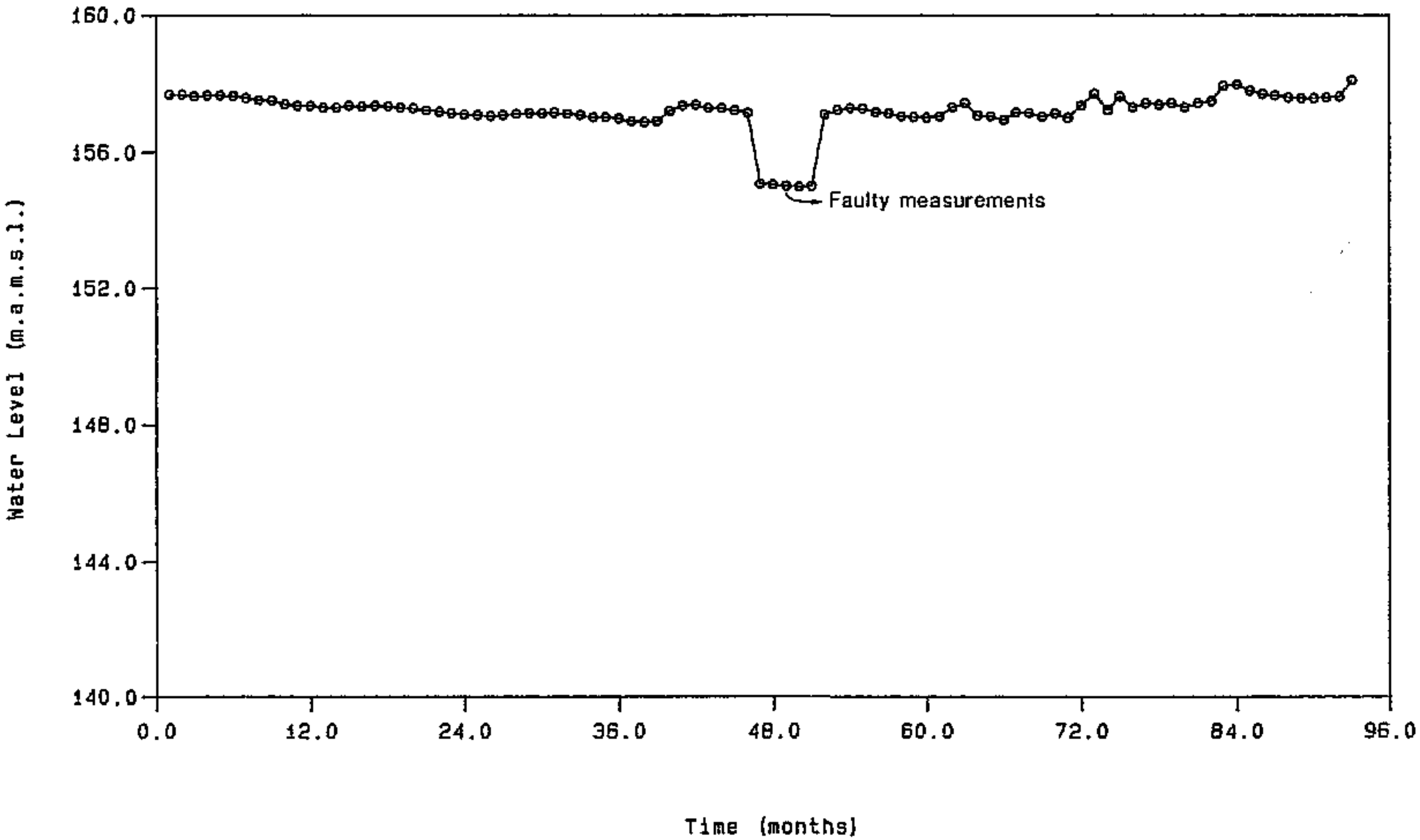


Figure 5.20. Water-level hydrograph of borehole 3318CB00010 - starting July, 1978.

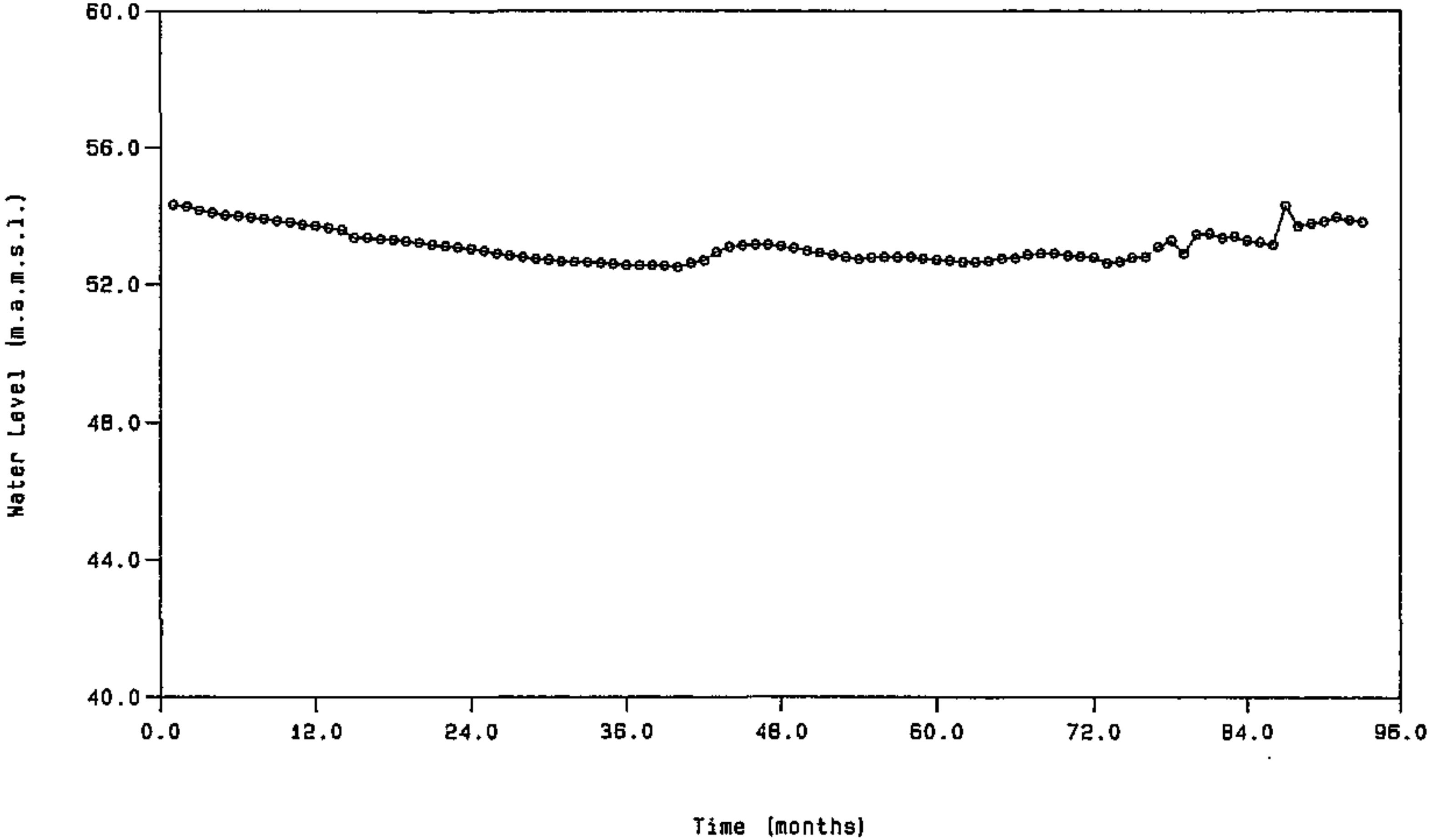


Figure 5.21. Water-level hydrograph of borehole 3318CB00015 - starting July, 1978.

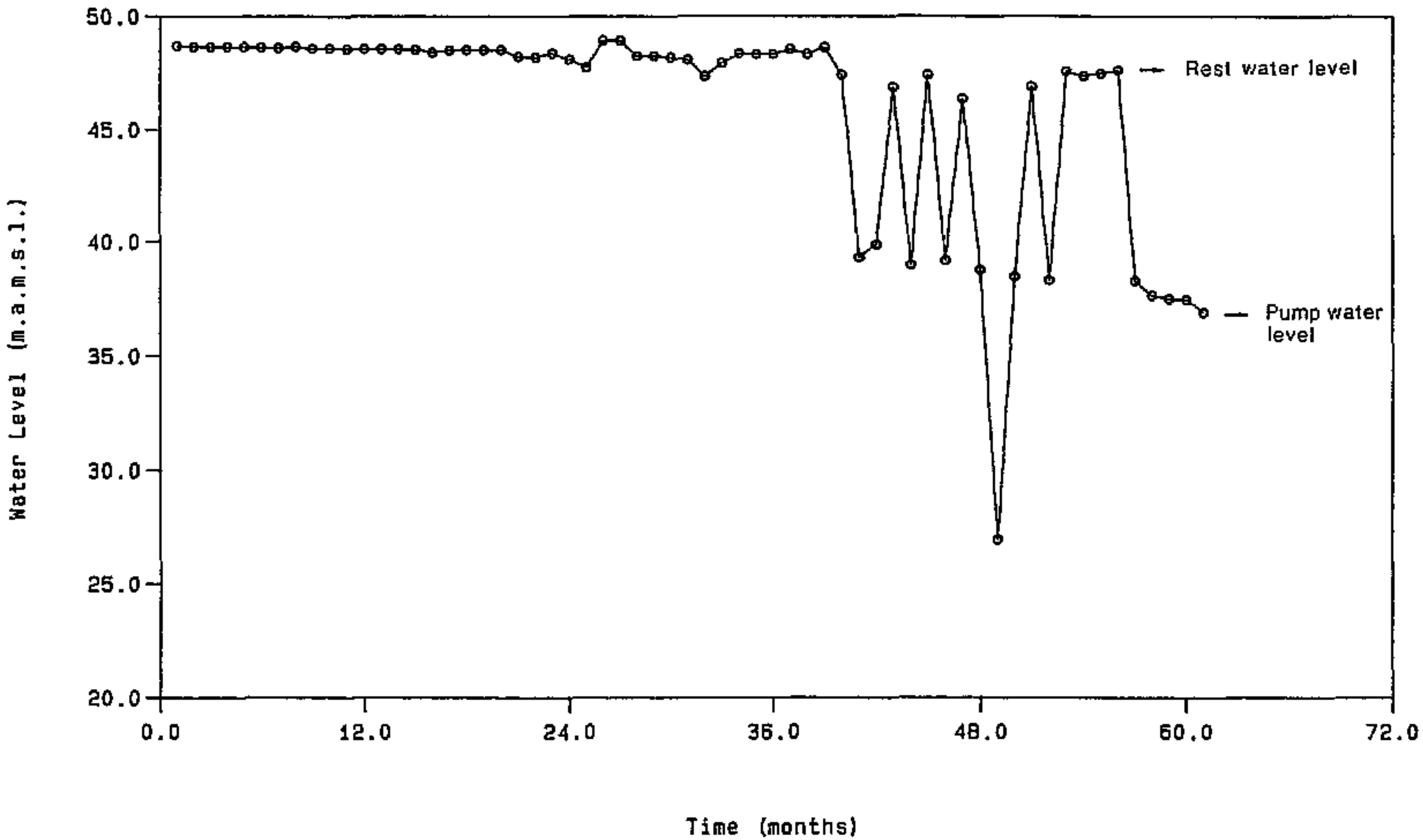


Figure 5.22. Water-level hydrograph of borehole 3318CB00048 - starting September, 1979

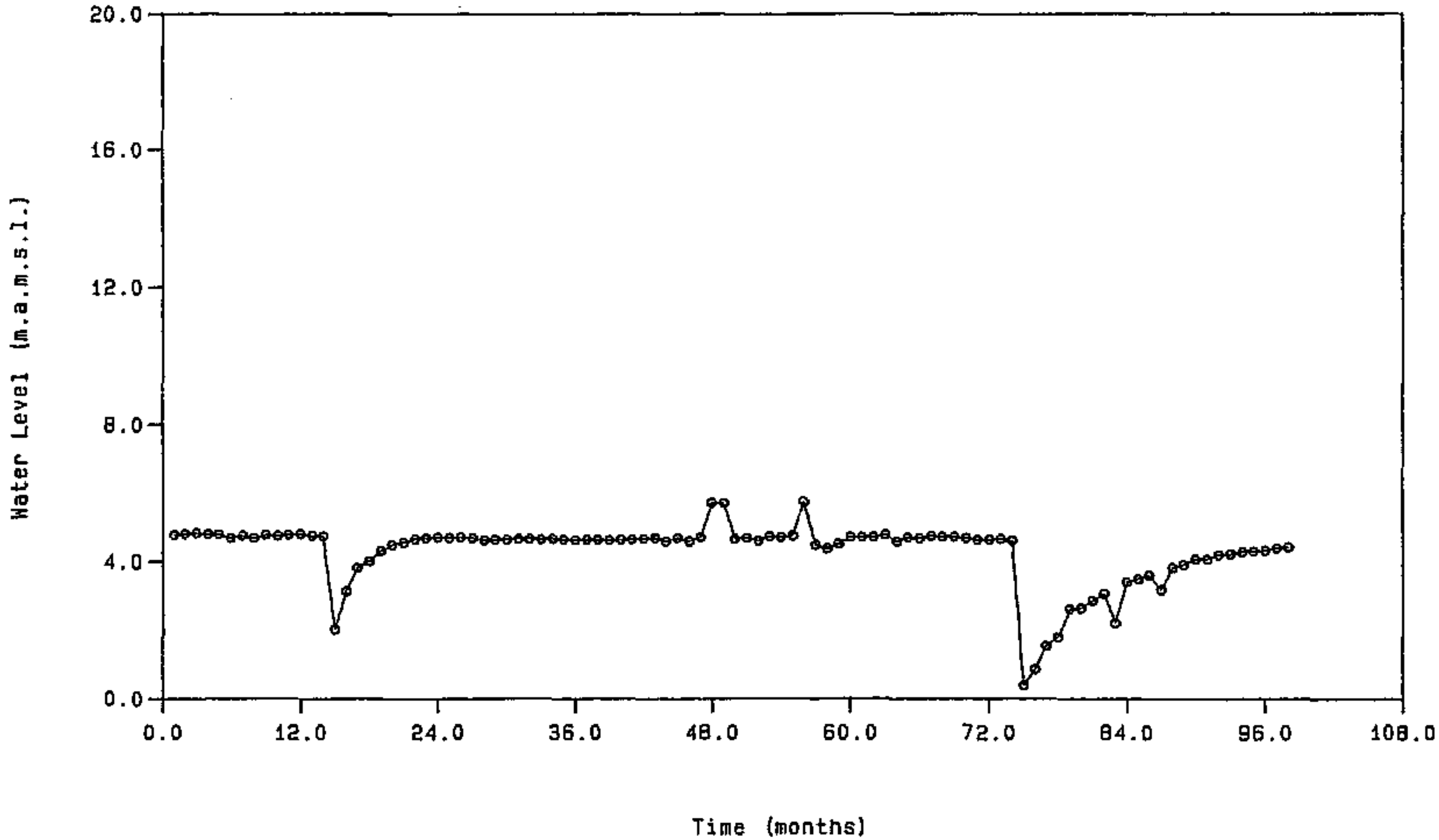


Figure 5.23. *Water-level hydrograph of borehole 3318CB00124 - starting August, 1977.*

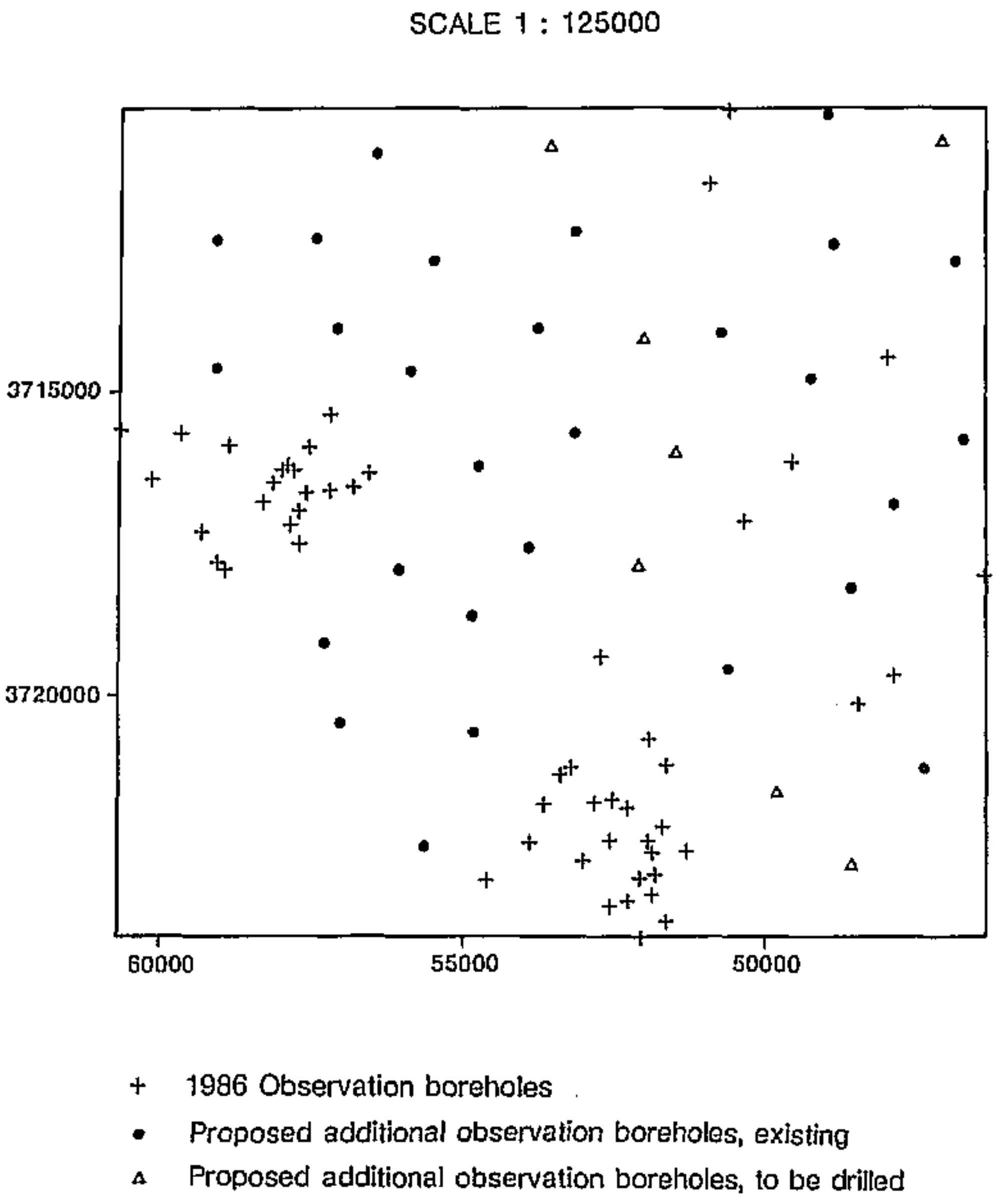


Figure 5.24. Proposed borehole observation network for the Atlantis aquifer.

movement in these areas, as illustrated by the discontinued contours in certain parts of the maps. The maps illustrated in figures 5.14, 5.15 and 5.16 are most effected by the reduced number of observation boreholes.

If the observation boreholes in 1980 (fig. 5.13) and 1986 (fig. 5.18) are compared with the map in figure 5.12 which shows the positions of the existing boreholes, it is obvious that sufficient boreholes have been drilled which could be used to monitor water levels and electrical conductivities in and around the well field. The density of the observation boreholes depends on the aquifer parameters. If there is more than one production area in an aquifer, it is desirable to design the observation network in such a way that not only the response of the various production areas is monitored, but also the reaction and behaviour of the whole aquifer.

In figure 5.18, the production areas are well covered by observation holes, but not the aquifer as a whole. Thus, the total aquifer response and the well field interactions are not monitored. Based on the 1986 observation borehole distribution and considering the steep water-level gradients (see figure 5.13) in the well field, figure 5.24 shows an observation network that would cover the whole Atlantis aquifer. Although the proposed observation network seems dense, the majority of the boreholes need to be measured only once or twice per annum, considering the long-term water-level fluctuations which are small over most of the well field.

A network which is too dense is not desirable, because of the unnecessary expenses to capture the data. Considering only the observed water-level variations and the detail generated on the contour maps, it might well be possible to decrease the number of observation holes in the densely populated production fields. On the other hand, operational aspects or considerations regarding the chemistry of the water may require data from all the holes monitored.

5.3.2 Regional analysis

Similarly to the geological data, the data necessary for the aquifer evaluation which are presently stored in the data base and can be used for a regional analysis, are insufficient. As discussed earlier, water-level contour maps on a regional scale, are of little value, because in the majority of cases they reflect the surface topography only. Only in areas with little relief where geological structures, such as dykes, are responsible for steep water-level gradients or discontinuities, water-level contours will show a different pattern. Contoured water-level changes depict recharge regions, if a proper time interval is chosen and they also assist in water balance calculations. Such maps can only be compiled, however, if long-term water-level measurements of the respective region exist. The available water-level data are poorly distributed and long-term observations do not exist for any of the sites.

Interesting is the comparison of the water-level statistic for the whole of the Orange Free State in table 5.8 and the southern Orange Free State (table 5.2), where a 7,38 m shallower mean water level of 5,37 m was observed. This difference can partly be ascribed to the fact that the levels in the Trompsburg area were measured during an above normal rainfall cycle, whereas the remaining water-level measurements of the Orange Free State were taken at various times and probably represent average conditions. Mean water levels in the Dewetsdorp area rose from 14,30 m during the drought in 1985 to 7,7 m after the high rainfalls of the 1987/88 season.

TABLE 5.8. WATER-LEVEL STATISTICS FOR THE ORANGE FREE STATE

Number of observations	=	5152
Mean water level	=	12,75 m
Standard deviation of water levels	=	11,65 m

A parameter which can be evaluated on a regional scale in the Orange Free State, is the borehole yield. Borehole depths will be discussed in the next section. Table 5.9 summarizes the borehole yield statistics of all the boreholes in the Orange Free State, which have recorded borehole yields.

TABLE 5.9. BOREHOLE YIELD STATISTICS OF THE ORANGE FREE STATE.

Number of observations	=	4988
Geomentric mean yield	=	1,72 l/s

Yields are log-normal distributed. The standard deviation of the geometrical mean has no practical importance. Figure 5.25(a) represents these statistics in a histogram.

To see whether any significant differences in borehole yield existed in various parts of the Orange Free State, the province was subdivided in square degrees. Figures 5.25(b) to (e) show histograms of the respective square degrees. If the graphs are analysed, it is found that more or less 90% of all the boreholes in the Orange Free State have yields smaller than 10 l/s. Boreholes in the central and southern parts of the province tend to have slightly higher yields than those in the northern parts, probably because of the occurrence of larger and more frequent dolerite outcrops in the south. The dolerite may also have an influence on the recharge in these areas which, in spite of the lesser rainfall, may be noticeably higher than in the northern parts of the Orange Free State.

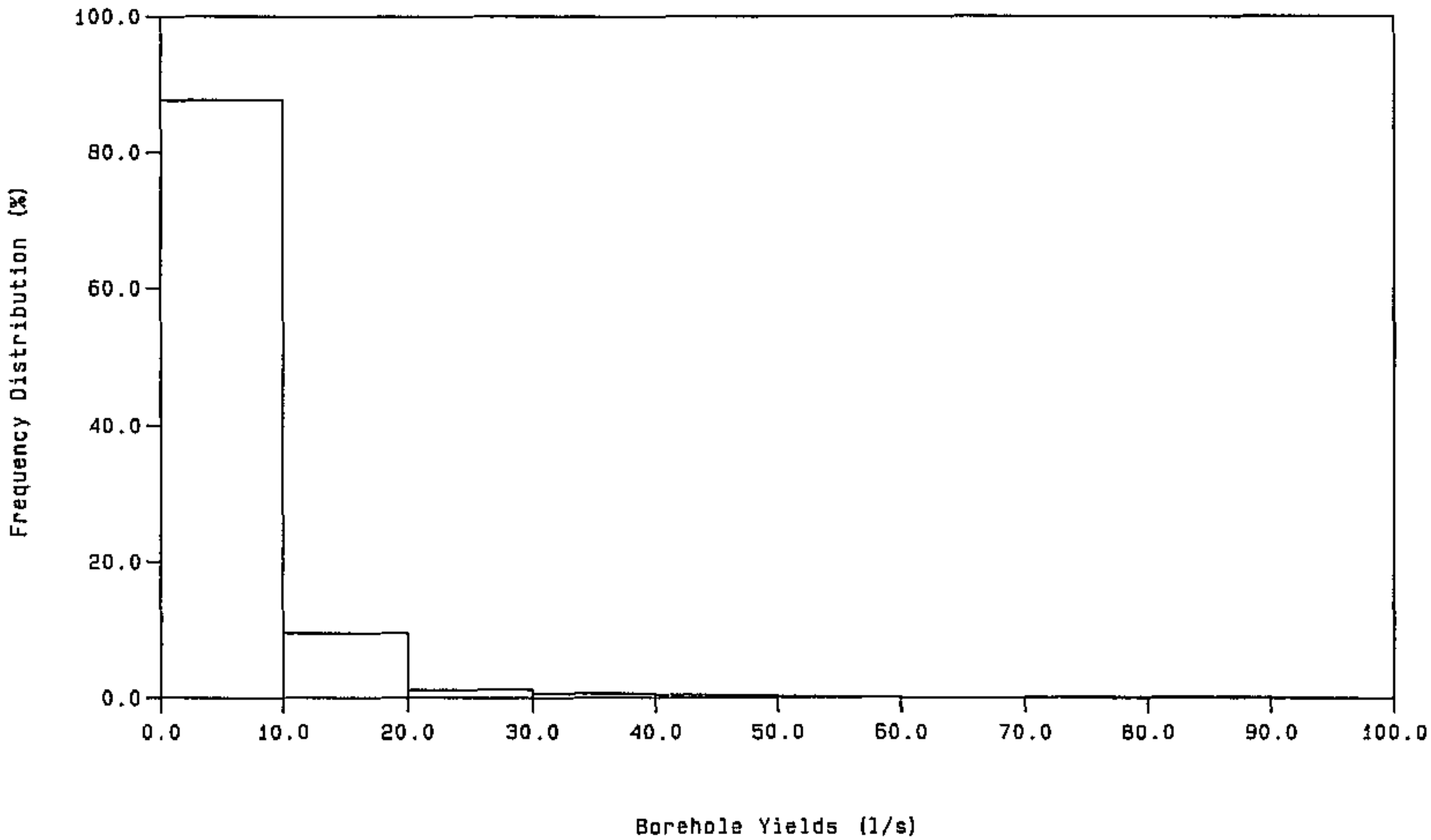


Figure 5.25(a). Frequency distribution of borehole yields in the Orange Free State.

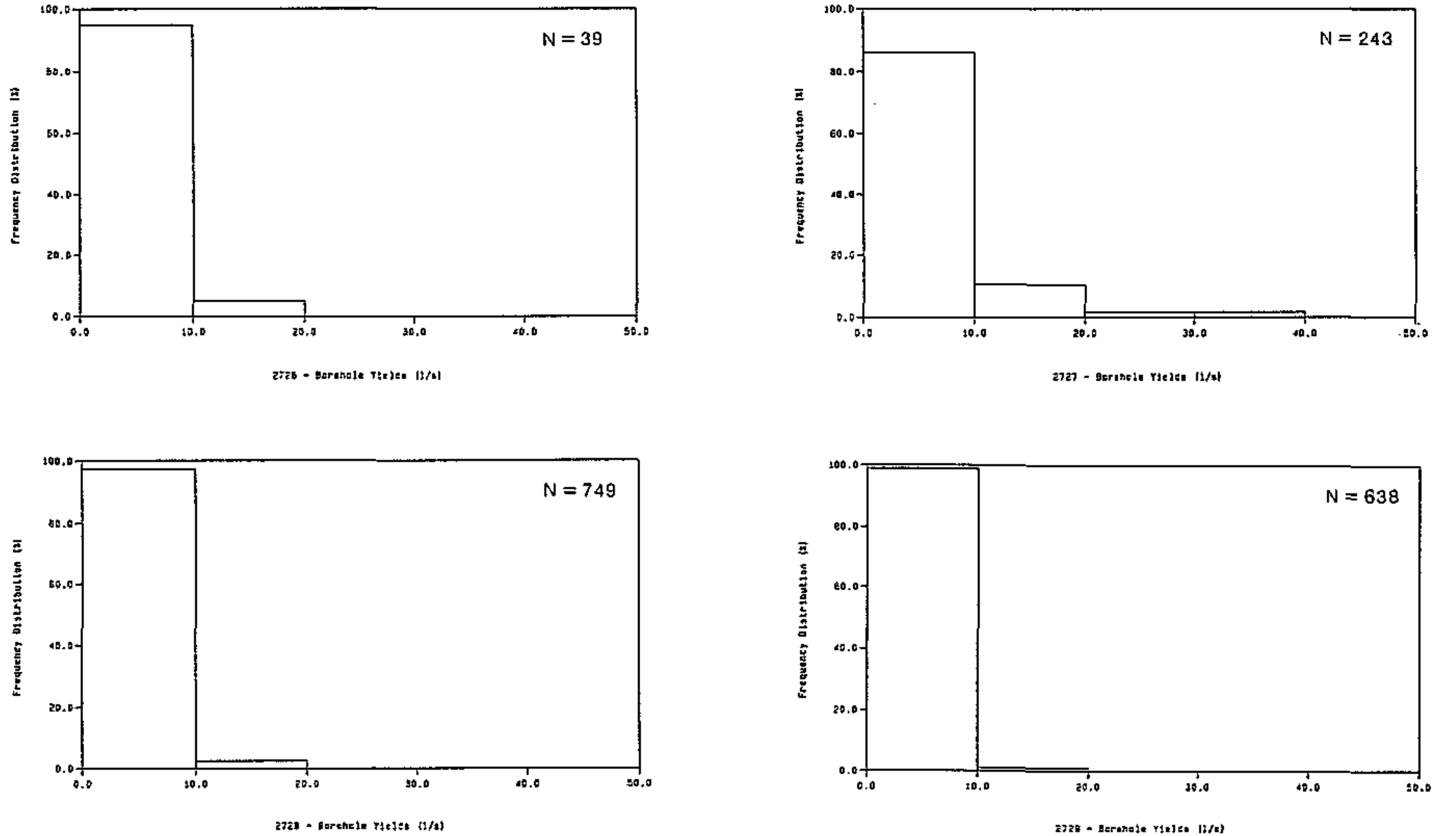


Figure 5.25(b). *Borehole yields of square degrees 2726, 2727, 2728 and 2729.*

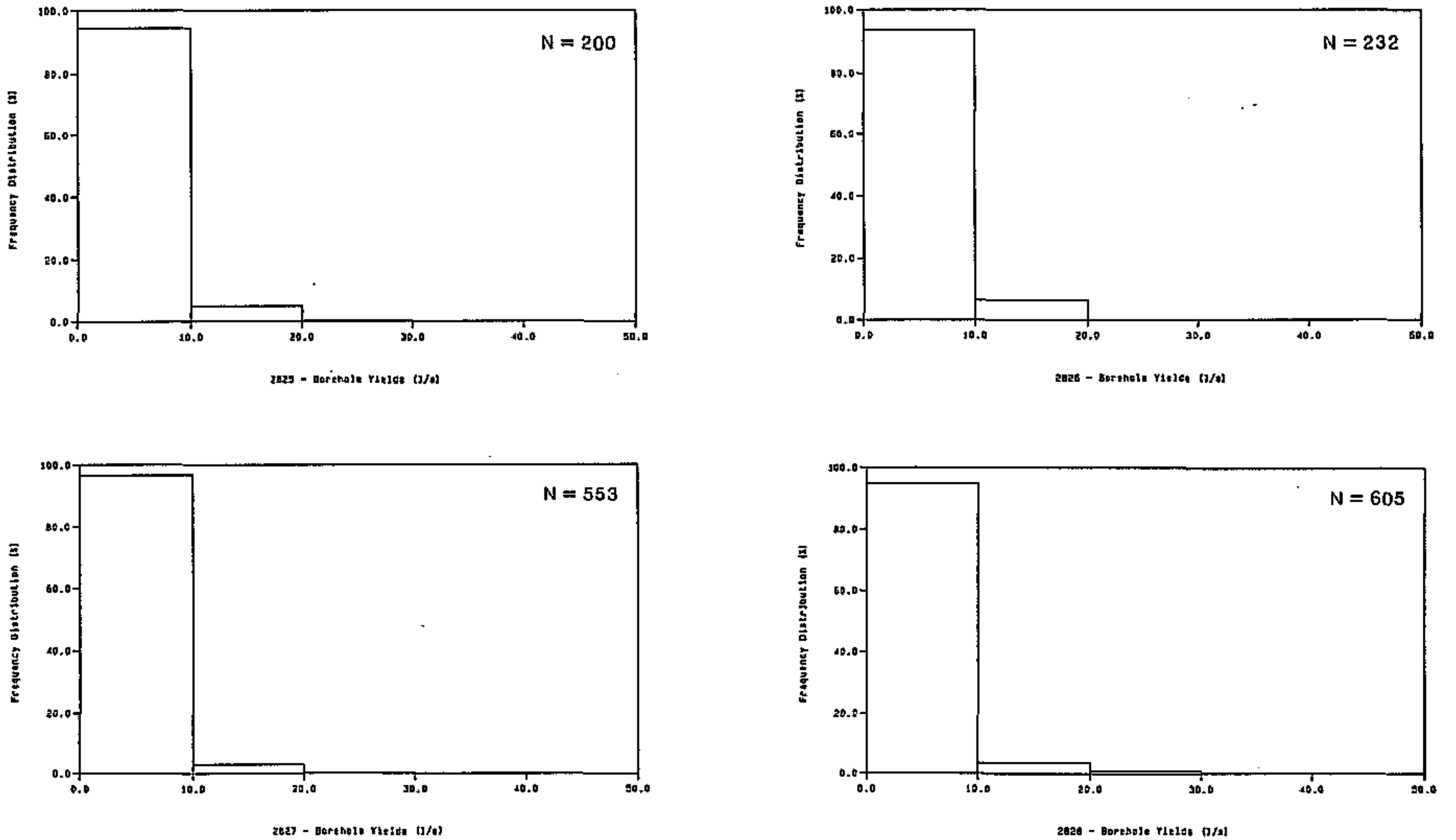


Figure 5.25(c). *Borehole yields of square degrees 2825, 2826, 2827 and 2828.*

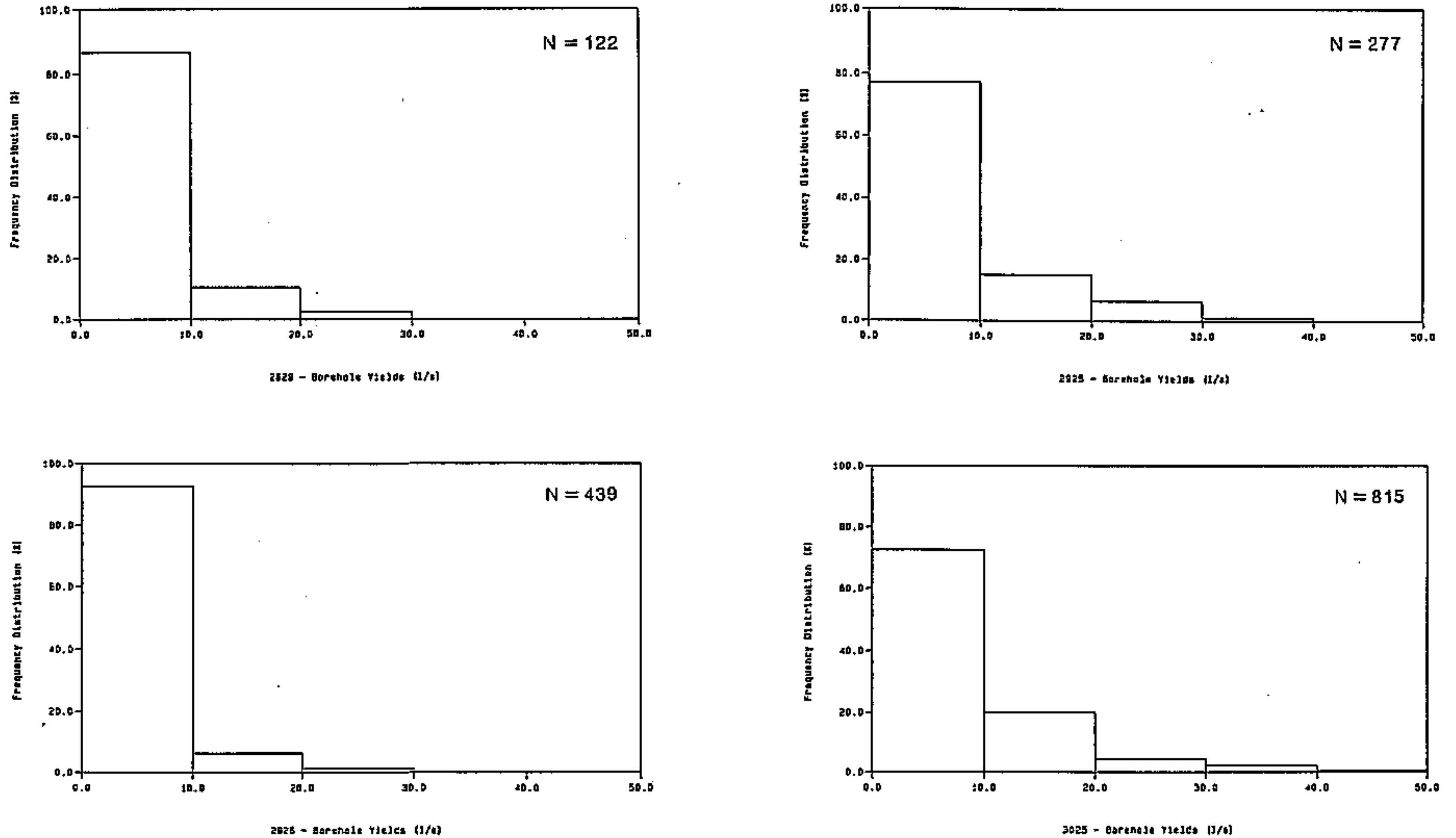
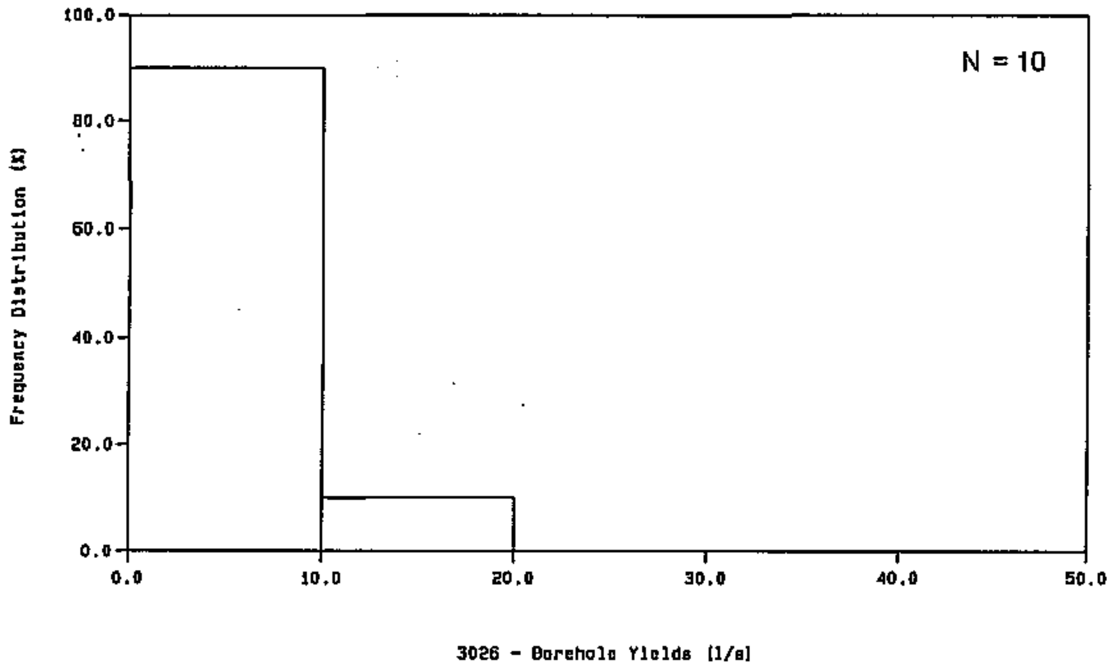


Figure 5.25(d). *Borehole yields of square degrees 2829, 2925, 2926 and 3025.*



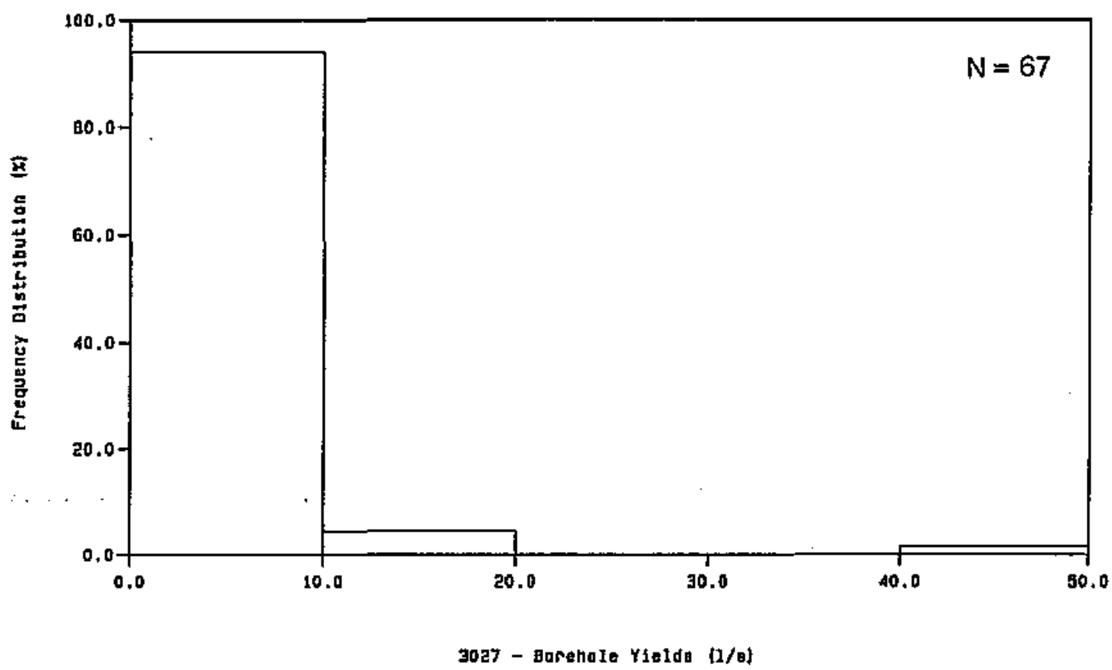


Figure 5.25(e). *Borehole yields of square degrees 3026 and 3027.*

A total of 20 pumping tests, spread over six 1:50 000 topographical maps, were available for the Orange Free State. Valid comparisons of aquifer parameters in the various regions or a correlation between the hydraulic parameters and the lithology can therefore not be made at this stage. Reported transmissivities varied from 260 m^2/d to as low as 9,4 m^2/d, while the storativities ranged between 6,0 x 10^{-3} and 6,2 x 10^{-6}.

Thus, to summarize, the lack of detail, inaccurate borehole positions and, in the Orange Free State, the poorly distributed data, make deductions from the existing data difficult. Yet, it could be shown that processing of the available data revealed interesting information, e.g.:

- Skew distribution of borehole yield data.
- Yield statistics: lower yields in the northern parts of the Orange Free State, higher yields (and recharge) associated with dolerite.
- Atlantis aquifer evaluation based on statistical data processing.
- Monitoring network design.
- Statistics on water application.

Possible advantages in processing long-term water-level records for aquifer balance calculations and recharge estimation could not be shown.

5.4 BOREHOLE CONSTRUCTION AND INSTALLATION

Orange Free State construction data, present in the data base, refer to the Hole- and Casing data sets only. Depth completed and the nominal diameter of the respective boreholes are obtained from the Entry data set. Where Construction and Installation data of the Orange Free State were insufficient, similar data from the Atlantis area were evaluated.

5.4.1 Local analysis

Square degree 3025

Again, square degree 3025 in the southern Orange Free State was selected, because of the relatively good data coverage. Of the 1840 boreholes in the region, 1329 boreholes were equipped. Table 5.10 shows the distribution of the equipment of these boreholes.

TABLE 5.10. TABULATED EQUIPMENT DISTRIBUTION IN THE SOUTHERN ORANGE FREE STATE.

Type of equipment	Number of installations
Wind pump (W)	1185
Turbine pump (T)	60
Centrifugal pump (C)	12
Submersible pump (S)	28
Mono pump (M)	29
Wind pump with power head (X)	15

If the borehole equipment is related to borehole yields in the area, one finds that almost 70% of the boreholes have yields < 10 l/s, whereas 89% of the boreholes in the area are equipped with wind pumps. Thus, there exists quite a good relationship between borehole equipment and borehole yields in the southern Orange Free State. Boreholes are generally shallow in the area. The mean depth is 26,75 m and wind pumps can be used nearly everywhere. Therefore, no definite relationship exists, if borehole equipment is related to borehole depth. This pattern is also observed when the other Orange Free State data are inspected.

The borehole depth distribution in the region is tabulated in table 5.11 and is illustrated by the histogram in figure 5.26.

TABLE 5.11. TABULATED BOREHOLE DEPTH STATISTICS OF THE REGION.

Number of holes	=	1138
Mean depth	=	26,75 m
Standard deviation	=	13,34 m

Only 1138 boreholes in the region had a recorded borehole depth.

As seen from the histogram, the majority of borehole depths vary between 20 m to 40 m, correlating well with the mean borehole depth of 26,75 m. Due to the fact that the mean static water level in the region is 5,37 m and that most of these water levels were measured during an above normal rainfall cycle (1977), farmers in the area may be confronted with drying up of shallow boreholes during droughts.

A typical borehole construction of the region is illustrated in figure 5.27 . The dotted line represents the diameter of the borehole, whereas the solid line indicates the casing in the borehole.

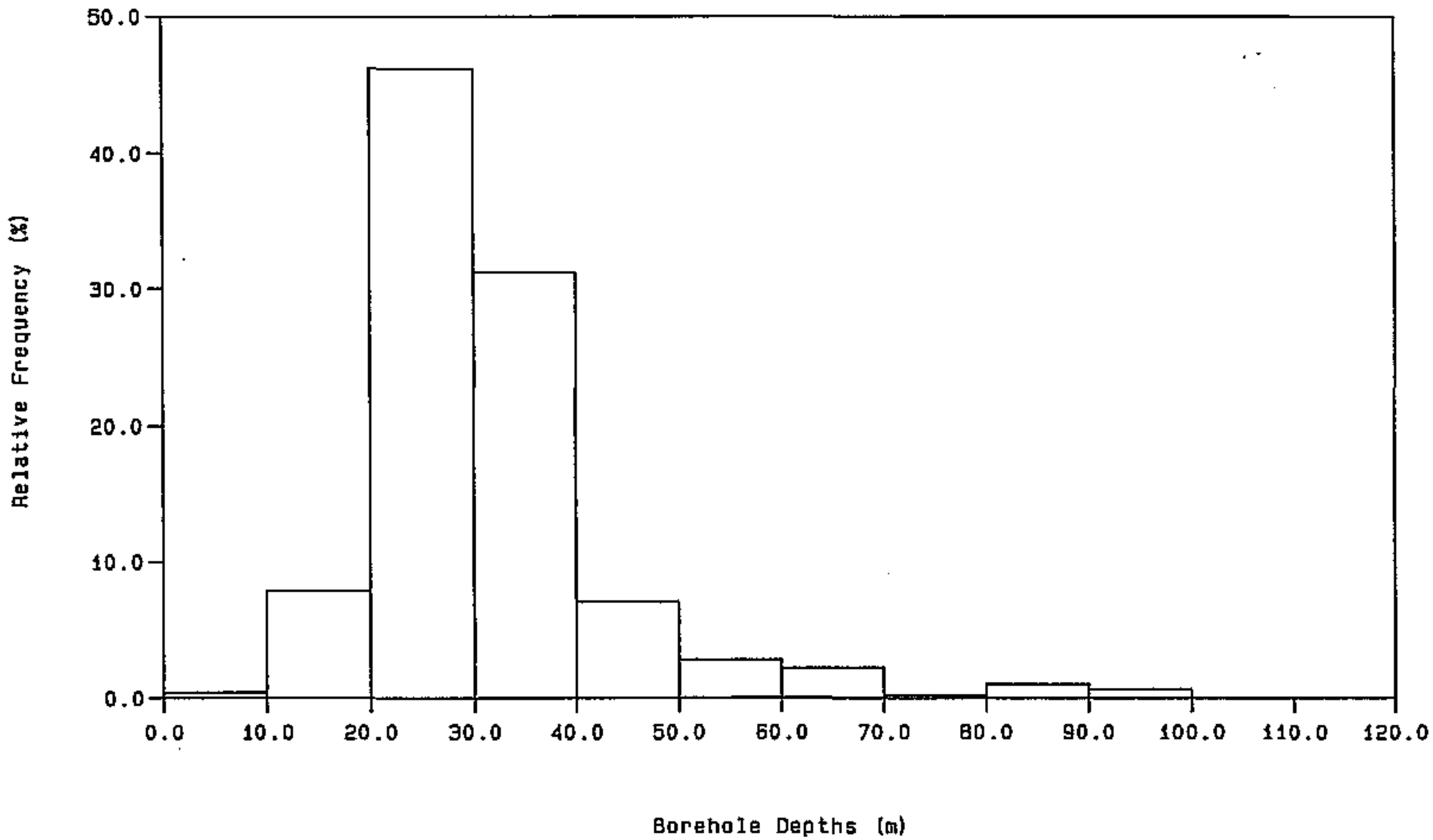


Figure 5.26. *Frequency distribution of borehole depths in the region.*

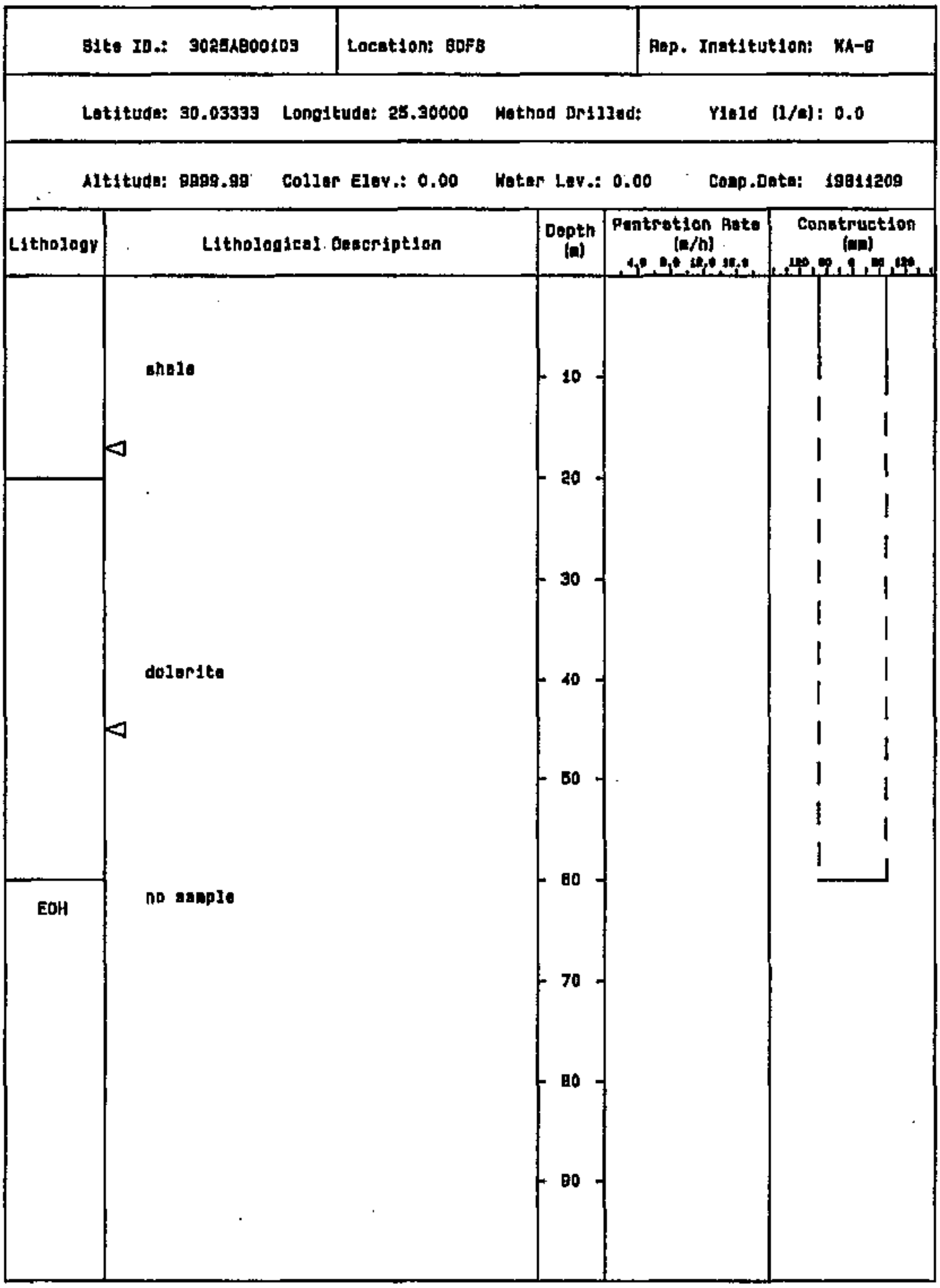


Figure 5.27. *Borehole chart of a typical southern Orange Free State borehole.*

No data exist in the data base applicable to casing perforations and borehole fill. Boreholes are drilled in consolidated rock and although some occasionally collapse, perforated casing (and gravel packs) are seldomly required. Sufficient construction data are recorded about the boreholes in the region to assist the hydrogeologist and engineer in future planning and construction of new boreholes. A sound estimate of the construction costs involved can be based on the existing data. In most cases, steel casings were encountered, while casing depths rarely exceeded 20 m.

In June 1987, for example, the local percussion drilling costs for a 165 mm diameter borehole varied from R14 to R17 in soft rock and R19 to R25 in hard rock. The maximum casing depth in the region is about 20 m, while the cost of casing installed per meter length was R18. If the maximum drilling rates for both soft rock and hard rock are considered, then the construction costs of a 40 meter borehole with 20 m casing, can vary between R1040 to R1360, depending on the amount of soft rock encountered during the drilling process.

Table 5.12 shows the statistics for borehole depths in the whole of the Orange Free State.

TABLE 5.12. TABULATED BOREHOLE DEPTH STATISTICS OF THE ORANGE FREE STATE.

Number of holes	=	8219
Mean depth	=	44,62 m
Standard deviation	=	24,57 m

Atlantis area

Numerous borehole construction data of the Atlantis area were entered into the groundwater data base. Of the 260 boreholes in the region, 216 boreholes have recorded borehole depths. The statistics of the borehole depths are listed in table 5.13.

TABLE 5.13. TABULATED ATLANTIS BOREHOLE DEPTH STATISTICS.

Number of boreholes	=	216
Total (recorded) depth drilled	=	7195,80 m
Mean depth	=	33,31 m
Standard deviation	=	19,45 m

Of the 260 boreholes in the region, 148 boreholes have recorded nominal diameters which range from 75 mm to 380 mm. Almost all boreholes were drilled with the same diameter from top to bottom. Table 5.14 displays some of the borehole construction information of

TABLE 5.14. TABULATED BOREHOLE DIAMETER INFORMATION OF THE ATLANTIS AREA.

Site Id.	Depth to Top of Interval (m)	Depth to Bottom of Interval (m)	Diameter of Hole (mm)	Method of Construction	Type of Finish	Method of Development	Duration of Developement	S. Development Treatment	Cost of Construction
3310AD00011	0.01	52.00	165				0		0.00
3310AD00066	0.01	29.00	250				0		0.00
3310AD00067	0.01	35.50	250				0		0.00
3310AD00068	0.01	21.00	250				0		0.00
3310AD00120	0.01	58.00	380				0		0.00
3310AD00121	0.01	50.00	300				0		16675.86
3310CB00001	0.00	35.00	250	Z			0		0.00
	0.00	42.00	304	C	P	C	0		0.00
3310CB00002	0.00	42.00	304	C	P	C	3		0.00
3310CB00003	0.01	31.10	152				0		0.00
	31.10	36.80	152				0		0.00
3310CB00004	0.00	26.00	152	C	P		0		0.00
3310CB00005	0.01	28.00	203	C		C	0		0.00
3310CB00006	0.01	50.00	152	C	P	C	0		0.00
3310CB00007	0.00	24.00	203			C	40		0.00
3310CB00008	0.01	45.00	304	C	P	C	0		0.00
3310CB00009	0.01	15.00	203	C	P	C	0		0.00
3310CB00010	0.01	43.00	152	C	P	C	0		0.00
3310CB00011	0.01	30.00	300	C	P	C	0		0.00
3310CB00012	0.01	38.00	200	C	P	C	0		0.00
3310CB00013	0.00	30.00	203	C	P	C	2		0.00
3310CB00014	0.01	36.00	203	C	P	C	0		0.00
3310CB00015	0.01	30.00	152	C	P		0		0.00
3310CB00016	0.01	30.00	152	C	P	C	0		0.00
3310CB00017	0.01	33.00	203	C	P	C	0		0.00
3310CB00018	0.01	30.00	152	C	P	C	0		0.00
3310CB00019	0.01	36.00	304	C		C	0		0.00
3310CB00020	0.01	30.00	203			C	18		0.00
3310CB00021	0.01	30.40	304				0		0.00
3310CB00022	0.01	48.00	304	C	P		0		0.00
3310CB00023	0.01	30.00	304				0		0.00
3310CB00024	0.01	38.00	304	C		C	0		0.00
3310CB00025	0.01	43.00	304				0		0.00
3310CB00026	0.01	35.00	203	C	P	C	0		0.00
3310CB00027	0.01	43.00	300				0		0.00
3310CB00028	0.01	45.00	152	C	P	C	0		0.00
3310CB00029	0.01	37.00	254	C			0		0.00
3310CB00030	0.01	40.00	200				0		0.00
3310CB00031	0.01	41.00	304				0		0.00
3310CB00032	0.01	40.00	254	C	P	C	0		0.00
3310CB00033	0.01	25.00	203				0		0.00
3310CB00034	0.01	30.00	250	C	P	C	0		0.00
3310CB00036	0.01	41.00	304				0		0.00
3310CB00037	0.01	48.00	200	C	P	C	0		0.00
3310CB00039	0.01	42.50	300	C	P	C	0		0.00

the region (for explanation of the codes in tables 5.14, 5.16, 5.17 and 5.18, consult section 3 in the User's Guide).

Of the 114 boreholes entered in the Hole data set, 73 boreholes have a recorded Method of construction. Of these boreholes, 69 were drilled by the cable-tool method, while only four boreholes were drilled by a hydraulic rotary drill.

During the drilling process, penetration rates were recorded in 26 boreholes. As shown in table 5.15, the predominantly soft formations gave rise to high penetration rates.

TABLE 5.15. PENETRATION RATE STATISTICS.

Number of observations	=	861
Mean penetration rate	=	102,80 m/h
Standard deviation	=	67,71 m/h

After completion, almost all the boreholes were developed by compressed air. In most cases, the duration of the development was not recorded. At sites where the duration of the development was recorded, it varied between 3 and 40 hours.

Because the formations encountered were predominantly unconsolidated, almost all the boreholes were cased, generally down to bed-rock. Where water-bearing zones were struck either perforated casing or screens were used. Table 5.16 lists some of the Opening information of a number of boreholes in the region.

In combination with the installation of perforated casing or screens, a gravel pack was placed around the casings. Up to date, no sieve analyses are recorded in the data base. Sieve analyses of the encountered unconsolidated formation are important for the selection of the proper size of perforations or screens and the grain size of the gravel pack. Table 5.17 lists the types of fill used in the boreholes in the region.

Of the 260 boreholes in the region, 12 boreholes are equipped with submersible pumps, whereas 13 boreholes have water-level recorders installed. For the remaining holes, no equipment was recorded in the ground-water data base. Table 5.18 lists the installations done on boreholes in the Atlantis area.

In figure 5.28 a borehole chart of a typical Atlantis borehole is shown.

TABLE 5.16. TABULATED OPENING INFORMATION OF THE ATLANTIS AREA.

Site Id.	Depth to Top of Interval (m)	Depth to Bottom of Interval (m)	Diameter of Hole (mm)	Material in Interval	Type of Openings	Length of Openings (mm)	Width of Openings (mm)	Hor. Dist. Openings (mm)	Vert. Dist. Openings (mm)	Method Open. made
3318AD000011	33.00	52.00	90		X	0	0.0	0	0	
3318AD000066	10.00	22.00	128		F	0	0.0	0	0	
3318AD000068	12.00	18.00	128		F	0	0.0	0	0	
3318AD000120	32.00	52.00	200	P	F	0	0.0	0	0	
3318CD000001	15.00	33.00	195	P	F	100	3.0	13	260	A
	24.00	30.00	203	S	S	0	0.0	0	0	
3318CB000002	24.20	30.20	152	S	F	0	0.0	0	0	
3318CB000003	31.10	36.80	152	S	S	0	0.0	0	0	
3318CB000004	24.50	25.50	152	S	F	0	0.0	0	0	
3318CB000008	10.69	20.39	175	P	F	57	4.0	0	0	
3318CB000009	9.00	11.00	152	S	W	0	0.0	0	0	
3318CB000010	27.00	30.00	127	S	F	0	0.0	0	0	
3318CB000011	36.00	38.00	200	P	F	0	0.0	0	0	
3318CB000012	36.00	38.00	200	P	F	0	0.0	0	0	
3318CB000014	21.95	23.95	152	S	W	0	0.0	0	0	
3318CB000016	11.80	27.00	175	P	F	0	0.0	0	0	
3318CB000017	26.50	29.50	127	S	F	0	0.0	0	0	
3318CB000018	10.50	20.50	152	S	F	0	0.0	0	0	
3318CB000019	21.12	35.67	180	P	F	0	0.0	0	0	
3318CB000020	20.00	22.00	152	S	F	0	0.0	0	0	
3318CB000021	20.60	30.34	200	P	F	0	0.0	0	0	
3318CB000022	26.20	29.20	150	S	F	0	0.0	0	0	
	30.70	33.70	150	S	F	0	0.0	0	0	
3318CB000023	26.00	30.00	200	P	F	0	0.0	0	0	
3318CB000024	36.00	30.00	200	P	F	0	0.0	0	0	
3318CB000025	41.00	43.00	200	P	F	0	0.0	0	0	
3318CB000026	21.00	23.00	152	S	F	0	0.0	0	0	
3318CB000027	30.00	43.00	200	P	F	0	0.0	0	0	
3318CB000028	39.00	45.00	200	P	F	0	0.0	0	0	
3318CB000029	11.00	35.00	200	P	F	0	0.0	0	0	
3318CB000031	39.00	41.00	200	P	F	0	0.0	0	0	
3318CB000033	14.00	17.00	127	S	F	0	0.0	0	0	
3318CB000034	24.00	30.00	200		P	0	0.0	0	0	
3318CB000036	39.00	41.00	200	P	F	0	0.0	0	0	
3318CB000037	16.75	26.50	175	S	F	0	0.0	0	0	
3318CB000039	36.00	42.00	200	P	F	0	0.0	0	0	
3318CB000040	29.30	31.38	152	S	L	0	0.0	0	0	
3318CB000041	6.00	8.00	152	S	F	0	0.0	0	0	
3318CB000042	26.00	28.00	152	S	F	0	0.0	0	0	
3318CB000043	13.00	15.00	152	S	F	0	0.0	0	0	
3318CB000044	18.50	20.50	200	P	F	0	0.0	0	0	
3318CB000045	24.00	30.00	203	S	S	0	0.0	0	0	
3318CB000046	24.00	30.00	150	P	S	0	0.0	0	0	
3318CB000047	19.50	31.50	203	S	F	0	0.0	0	0	
3318CB000048	26.30	32.30	152	S	S	0	0.0	0	0	

TABLE 5.17. TABULATED FILL INFORMATION OF THE ATLANTIS AREA.

Site Id.	Depth to Top of Interval (m)	Depth to Bottom of Interval (m)	Type of Fill
3318A0000011	33.00	52.00	X
3318CB000001	35.00	35.00	X
3318CB000011	38.00	38.00	X
3318CB000012	38.00	38.00	X
3318CB000022	26.20	35.00	G
3318CB000023	30.00	38.00	X
3318CB000024	38.00	38.00	X
3318CB000025	43.00	43.00	X
3318CB000027	43.00	43.00	X
3318CB000028	45.00	45.00	X
3318CB000031	41.00	41.00	X
3318CB000034	30.00	30.00	X
3318CB000036	41.00	41.00	X
3318CB000039	42.00	42.00	X
3318CB000045	24.00	30.00	G
3318CB000046	24.00	30.00	G
3318CB000048	26.30	33.70	G
3318CB000054	32.00	32.00	X
3318CB000055	41.00	41.00	X
3318CB000056	46.00	46.00	X
3318CB000057	42.00	42.00	X
3318CB000058	18.00	18.00	X
3318CB000077	16.50	17.00	X
3318CB000078	33.00	33.00	X
3318CB000080	40.00	40.00	X
3318CB000083	33.00	33.00	X
3318CB000084	55.00	55.00	X
3318CB000085	18.00	18.00	X
3318CB000086	25.00	25.00	X
3318CB000087	40.00	40.00	X
3318CB000088	40.00	40.00	X
3318CB000089	28.00	28.00	X
3318CB000090	31.00	31.00	
3318CB000091	43.00	43.00	X
3318CB000094	49.00	49.00	X
3318CB000095	50.00	50.00	X
3318CB000097	35.00	35.00	X
3318CB000099	28.00	28.00	X
3318CB000100	32.00	36.00	X
3318CB000103	34.00	34.00	X
3318CB000114	43.00	43.00	X
3318CB000115	26.00	26.00	X
3318CB000117	20.00	20.00	X
3318CB000118	27.00	27.00	X
3318CB000125	51.00	51.00	X

TABLE 5.18. TABULATED INSTALLATION INFORMATION OF THE ATLANTIS AREA.

Site Id.	Install. Date	Install. Type	Pump Intake Depth (m)	Power Type	Power Rating (kw)	Manufacturer	Serial Number	Power Meter Number	Mon. Facility	Reporting Ins.
3318CB00001	19820703	S	23.50	E	15				I	WAGC
3318CB00002	19811204	S	20.00	E	15				I	WAGC
3318CB00003	19811220	S	0.00	E	15				I	WAGC
3318CB00004	19780820	R	0.00		0					WAGC
3318CB00005	19780820	R	0.00		0					WAGC
3318CB00007	19780820	R	0.00		0					WAGC
3318CB00035	19830105	S	0.00	E	15				I	WAGC
3318CB00036	19840304	R	0.00		0					WAGC
3318CB00037	19801010	S	0.00	E	15				I	WAGC
3318CB00038	19830104	S	0.00	E	15				I	WAGC
3318CB00040	19800402	R	0.00		0					WAGC
3318CB00041	19800402	R	0.00		0					WAGC
3318CB00042	19800402	R	0.00		0					WAGC
3318CB00044	19800402	R	0.00		0					WAGC
3318CB00045	19811119	S	22.00	E	15				I	WAGC
3318CB00046	19800402	R	0.00		0					WAGC
3318CB00047	19800402	R	0.00		0					WAGC
3318CB00048	19840108	S	20.00	E	15				I	WAGC
3318CB00101	19790912	S	0.00	E	15				I	WAGC
3318CB00106	19790118	S	19.50	E	15				I	WAGC
3318CB00107	19800402	R	0.00		0					WAGC
3318CB00112	19800117	S	22.50	E	15				I	WAGC
3318CB00116	19831010	R	0.00		0					WAGC
3318CB00166	19800402	R	0.00		0					WAGC
3318DA00105	19781117	S	0.00	E	15				I	WAGC

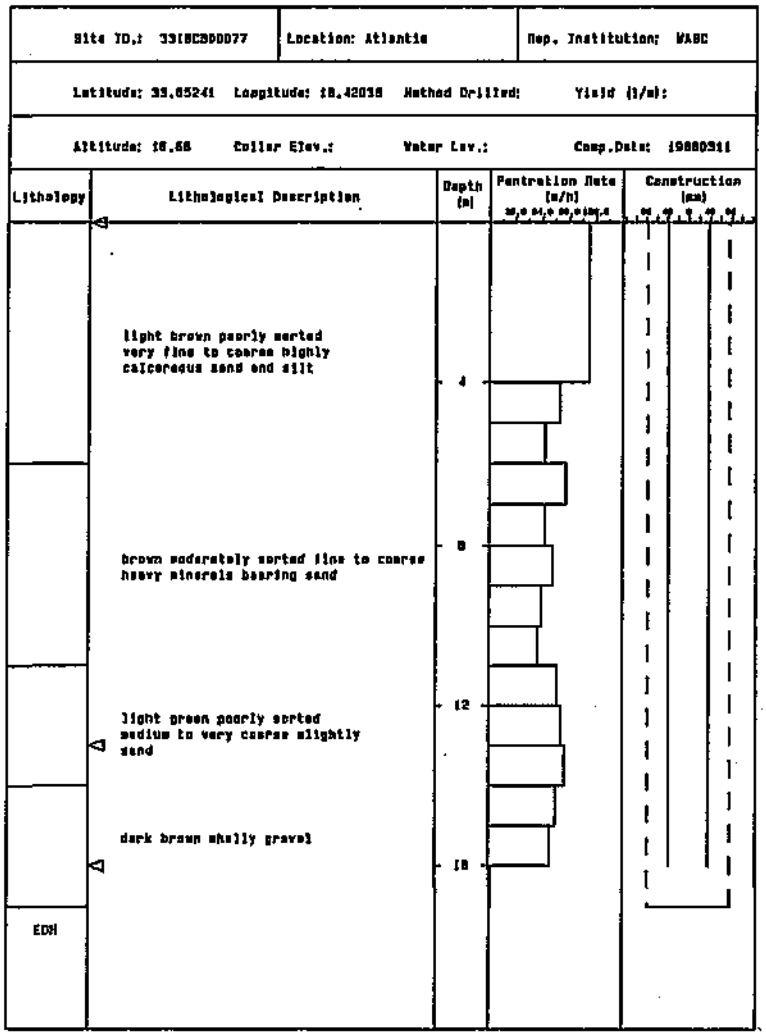


Figure 5.28. Borehole chart of a typical Atlantis borehole.

5.5 GROUND-WATER QUALITY

One of the most important facets of geohydrological evaluations is the hydrochemical assessment of the ground-water resources in a region. In this section, the hydrochemical data will be evaluated on a local as well as a regional basis.

5.5.1 Local analysis

Two areas are used to illustrate the hydrochemical data processing, namely square degree 3025 and the Atlantis area.

Square degree 3025

Water samples from a total of 463 boreholes were chemically analysed in this square degree. Only a part of these analyses (385 samples), i.e. those in the Trompsburg area, were used to assess the ground-water quality of the area. Figure 5.29 shows the distribution of the boreholes from which the samples were taken. The analyses were plotted on different hydrochemical diagrams, illustrated in figure 5.30.

In the diagrams, one finds that, except for a few analyses, the ground water belongs to one group. The chemical properties of ground water in this region are generally dominated by the alkaline earths and weak acids, i.e. the waters found are predominantly calcium-carbonate waters which were recently recharged. The SAR diagram shows that the majority of samples have a medium electrical conductivity and a very low sodium adsorption ratio. This means that the water can be used for irrigation purposes on almost any soil, with little danger of the development of harmful levels of exchangeable sodium.

Further statistical processing was done on a reduced sample of 100 analyses. This reduction was necessary to suit the arrays used in the statistical package.

These 100 analyses were randomly selected from the region and were again plotted on the different hydrochemical diagrams, to test the representativeness of the random sample. According to Krumbein and Graybill (1965), Till (1974) and Miller and Kahn (1962), a random sample of the size n from a population of N is valid, if the population mean and variance are approximately equal to the sample mean and variance. Table 5.19 shows the population mean and variance in comparison to the sample mean and variance. The respective values are close to each other and indicate that the sample is representative of the population.

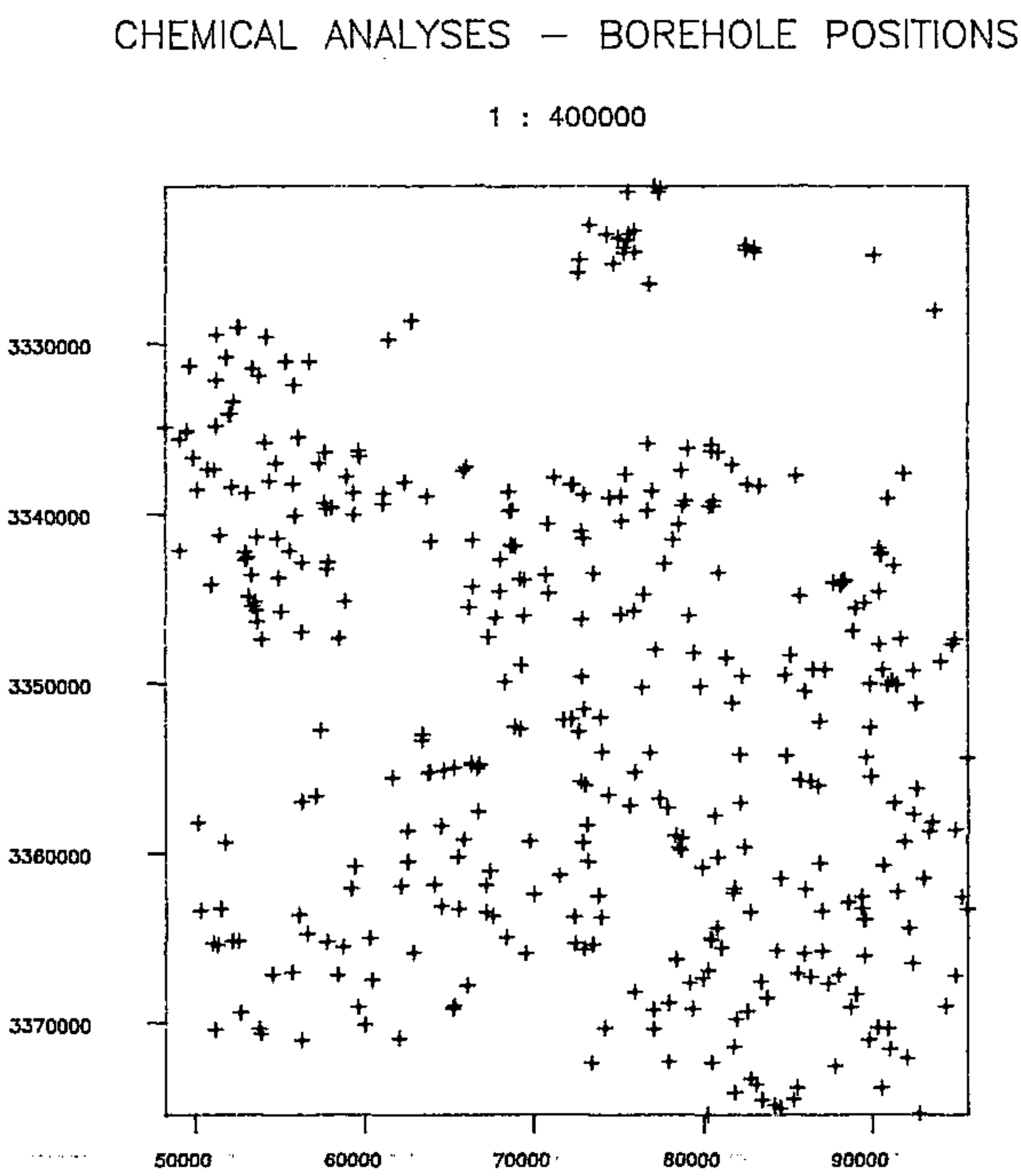


Figure 5.29. Distribution of chemical analysed boreholes in the Trompsburg area.

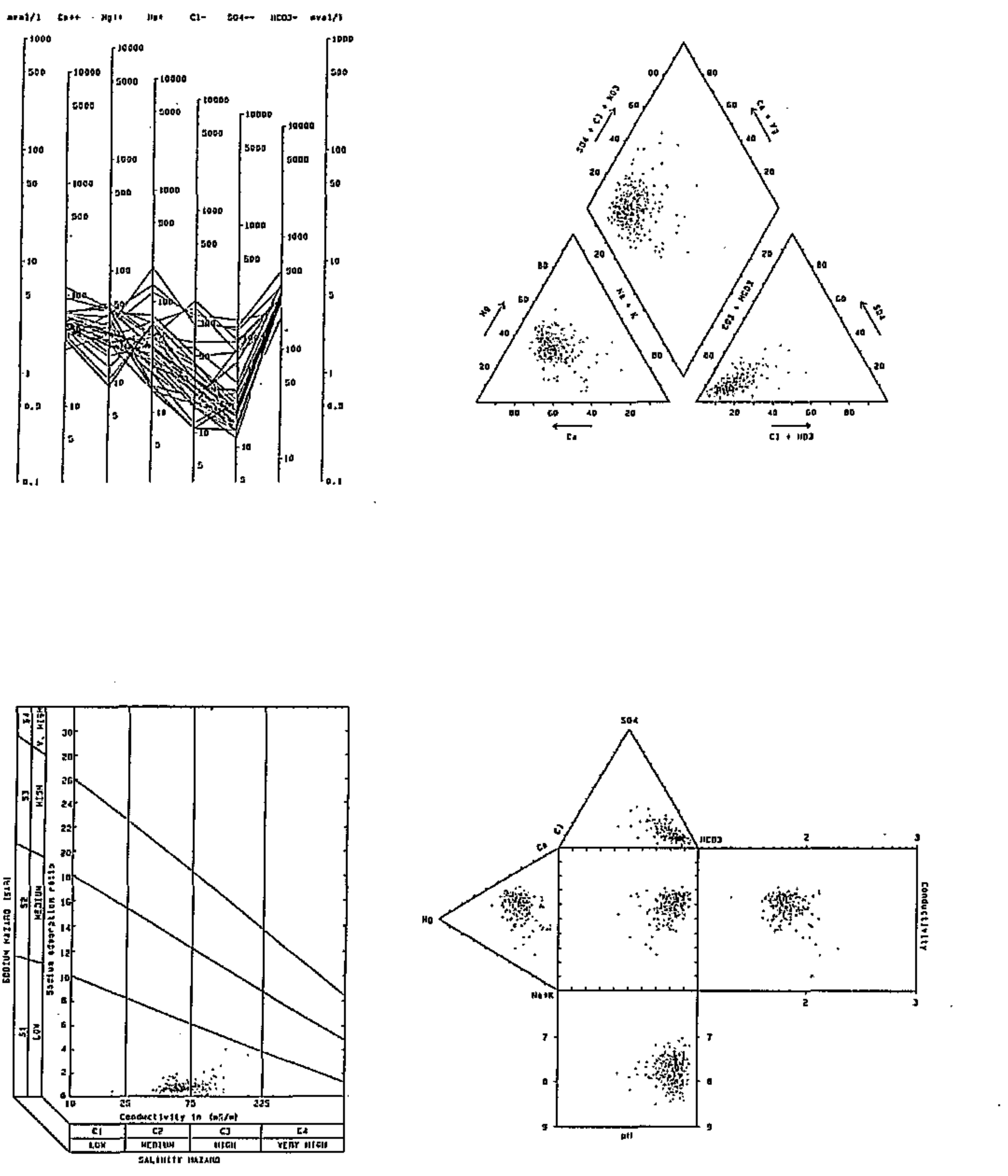


Figure 5.30. Hydrochemical diagrams of water samples in the Trompsburg area.

TABLE 5.19. STATISTIC OF THE WATER ANALYSES FROM THE TROMPSBURG AREA.

	POPULATION (N)	SAMPLE (n)
Number of Ca and Mg analyses	456	100
Mean of Ca analyses (mg/l)	27,59	26,64
Standard deviation of Ca analyses (mg/l)	9,22	8,55
Mean of Mg analyses (mg/l)	58,85	57,84
Standard deviation of Mg analyses (mg/l)	13,19	13,44

Figure 5.31 shows the Piper analysis of the reduced sample. If compared with the original data, it can be seen that there is little difference between the two plots, confirming visually the findings of the statistic in table 5.19.

Contour maps (based on 385 analyses) of a number of parameters were drawn to see whether any one of the selected parameters displayed a definite pattern or trend. Figures 5.32 - 5.38 show the different contour maps.

In figure 5.32, it is observed that the TDS-values tend to be slightly higher along the watershed (figure 5.8 (c)). This observation is difficult to explain. Possibly sheet flow occurs more often on the slopes and in the valleys, dissolving and carrying away more of the salts produced by weathering of the rocks. Electrical conductivity contours (see figure 5.33) follow the pattern of the TDS, while for the alkalinity contours (see figure 5.35) this is less evident. The chloride contours in figure 5.37 correlate also reasonably well with the TDS- and electrical conductivity contours.

In the nitrate contours (see figure 5.36), four anomalies are depicted. They can most probably be ascribed to point source pollution.

Generally, small values of the Cl/HCO_3-ratio are indicative of possible recharge areas. In figure 5.38, these ratios do not portray a specific pattern. Recharge occurs probably rather widespread.

To summarize, none of the above contour maps shows definite trends. This was to be expected, because water from Beaufort sediments and from dolerites are both classified as Ca/Mg-HCO_3 waters and, with a few exceptions, the hydrochemical diagrams show only one group.

Ionic ratios may help in the classification of waters, detection of anomalies and lead to better understanding of different geohydrochemical processes. Not all ionic ratios may be

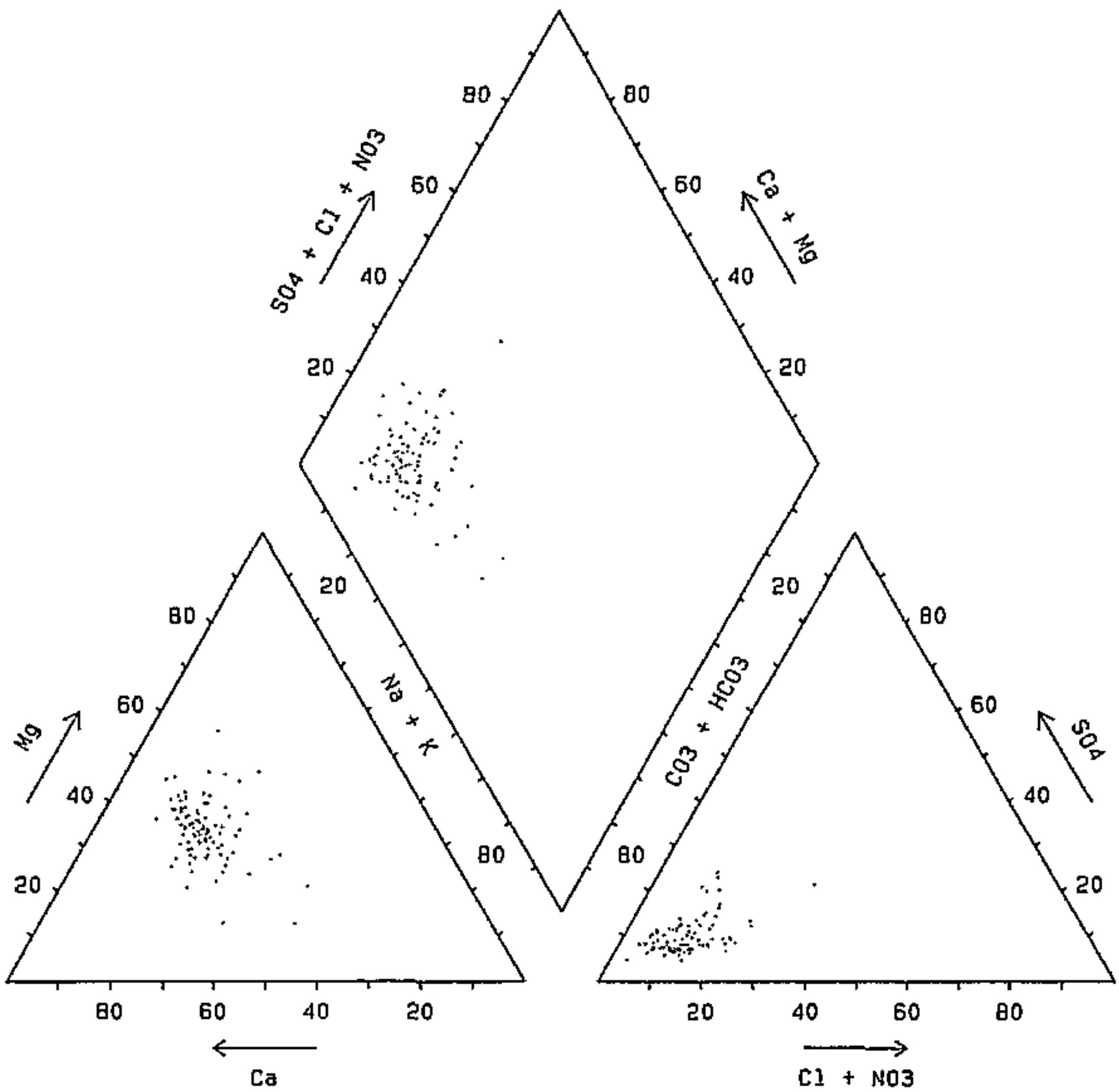


Figure 5.31. Piper analysis of the randomly reduced water samples.

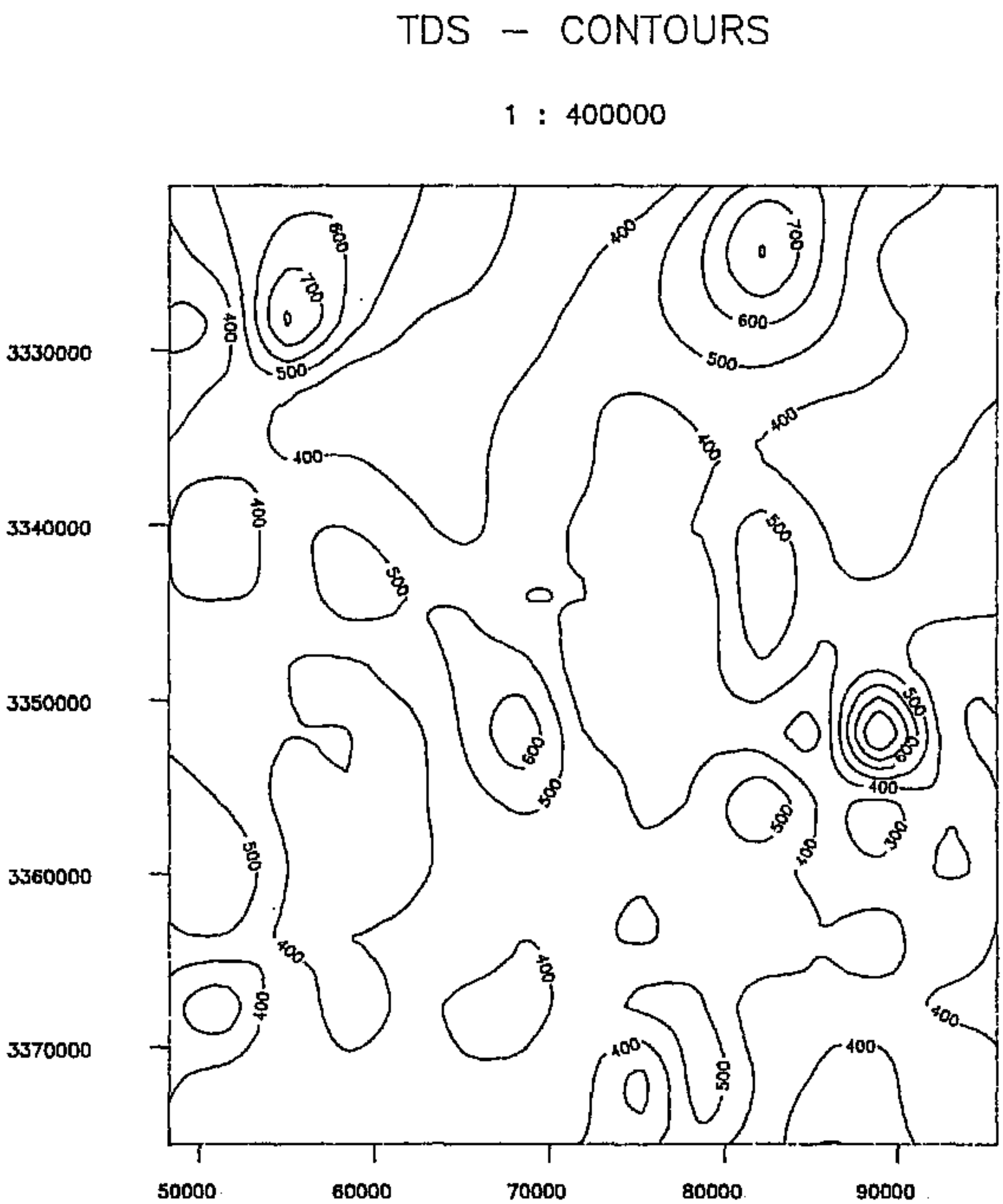


Figure 5.32. TDS contours of the Trompsburg area in mg/l.

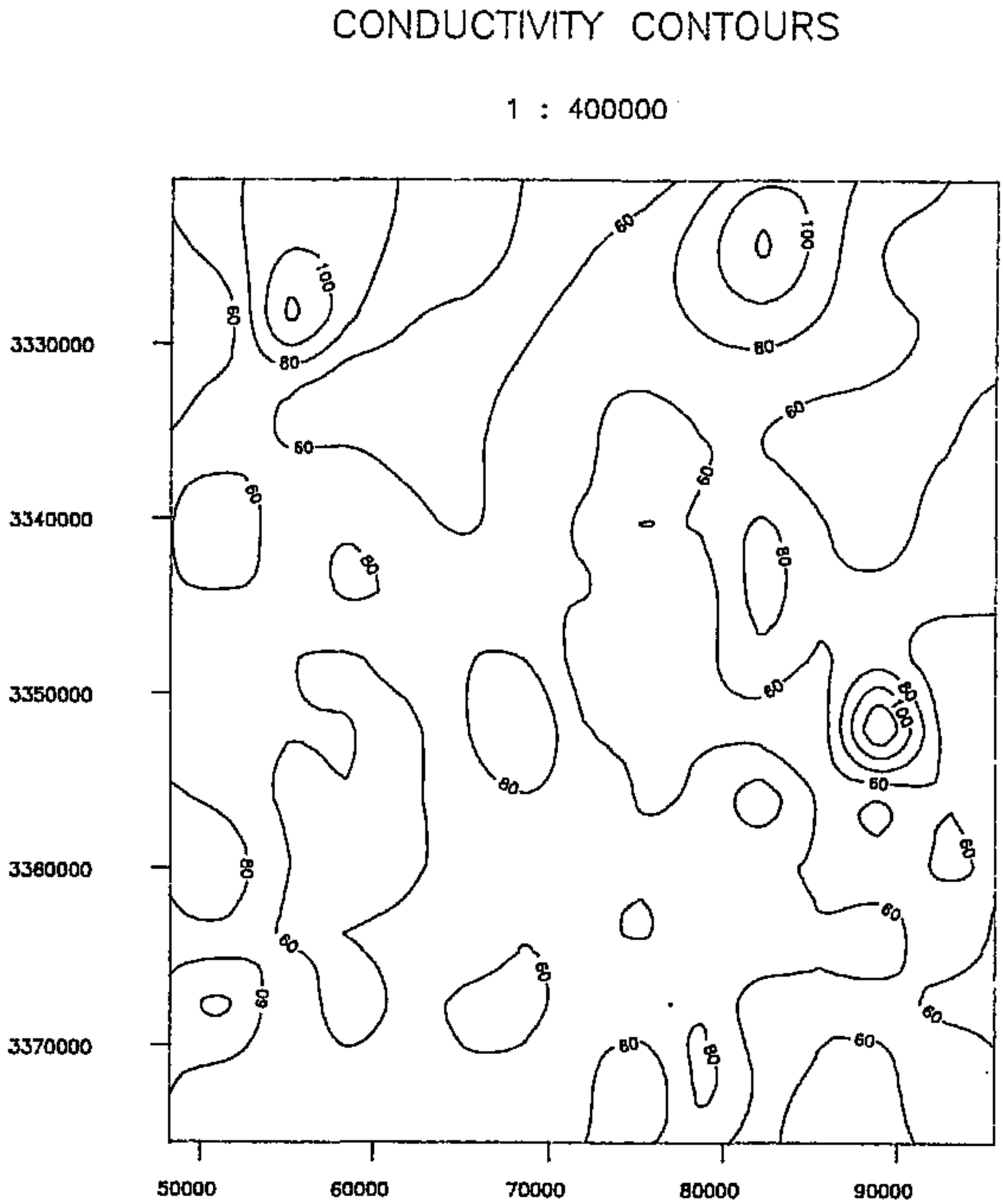


Figure 5.33. Electrical conductivity contours of the Trompsburg area in mS/m.

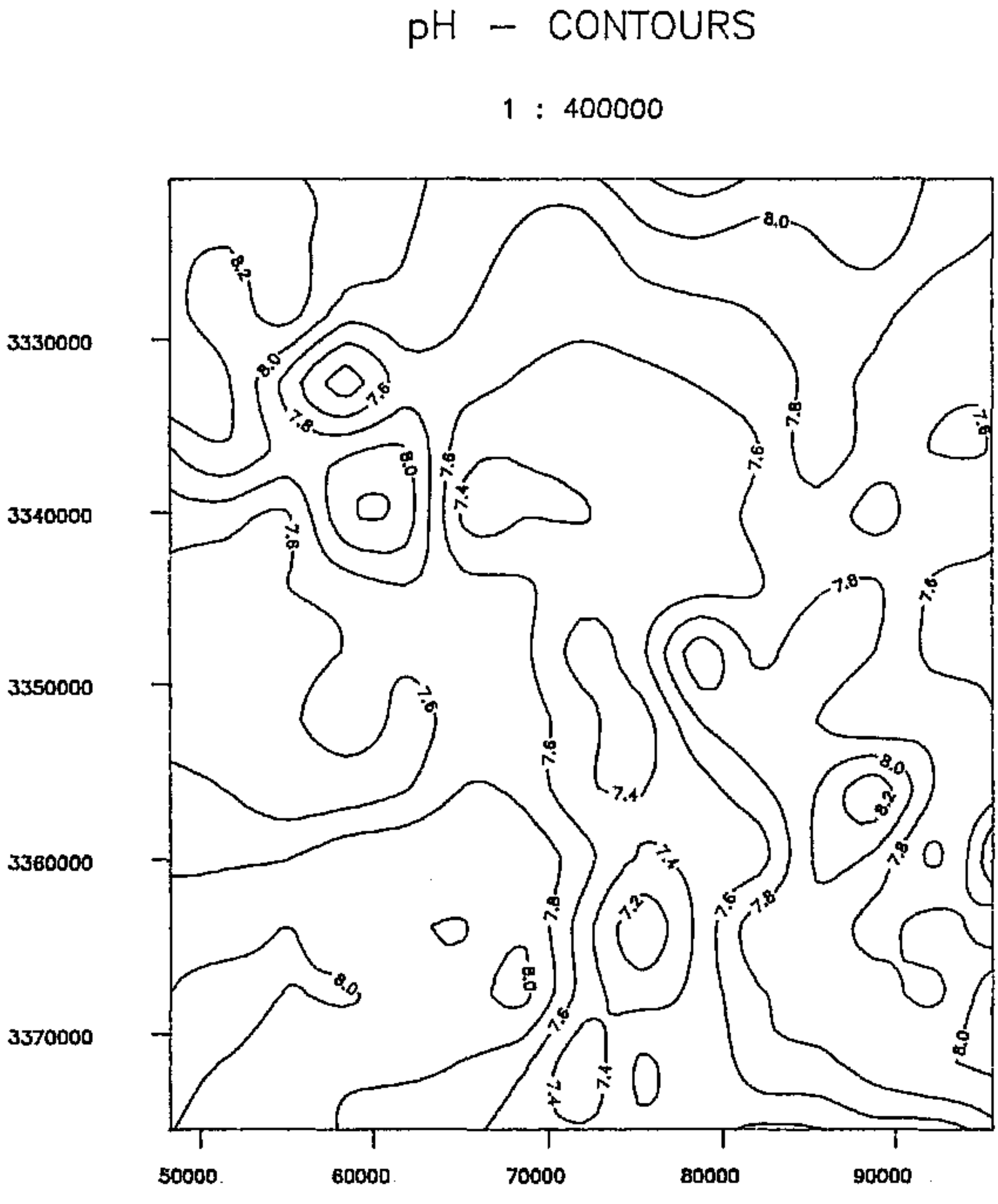


Figure 5.34. pH contours of the Trompsburg area.

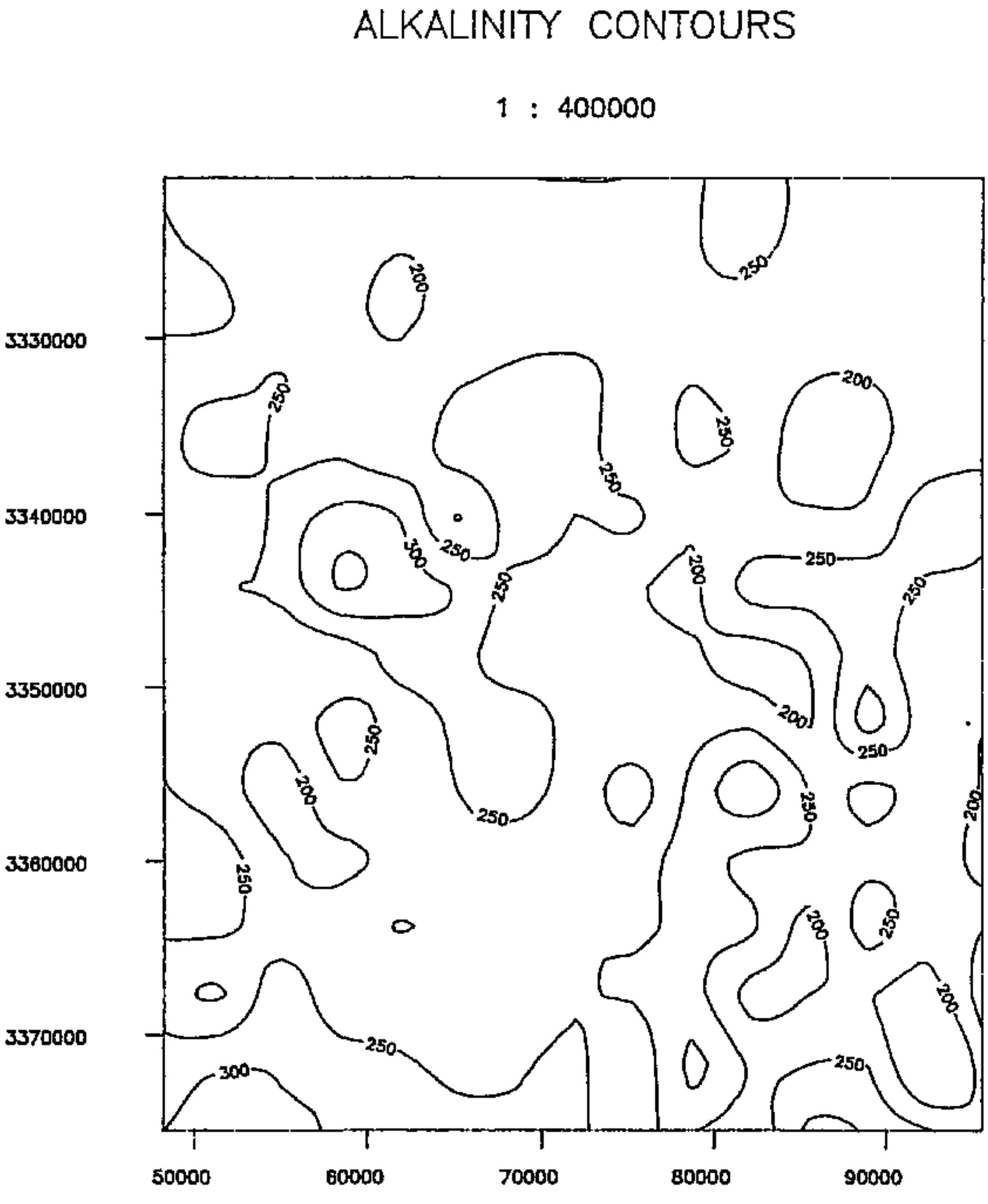


Figure 5.35. Alkalinity contours of the Trompsburg area in mg/l.

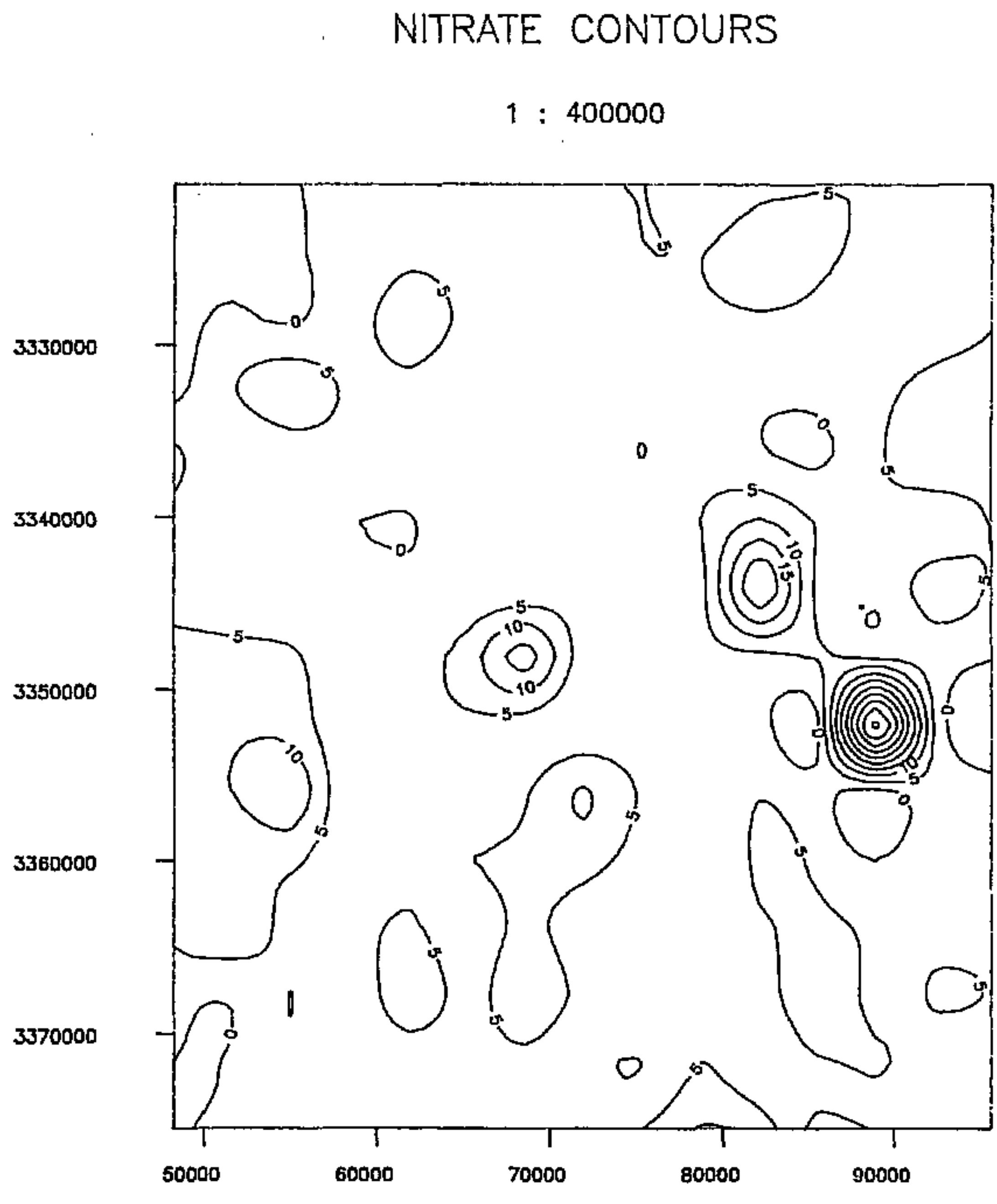


Figure 5.36. Nitrate contours of the Trompsburg area in mg/l.

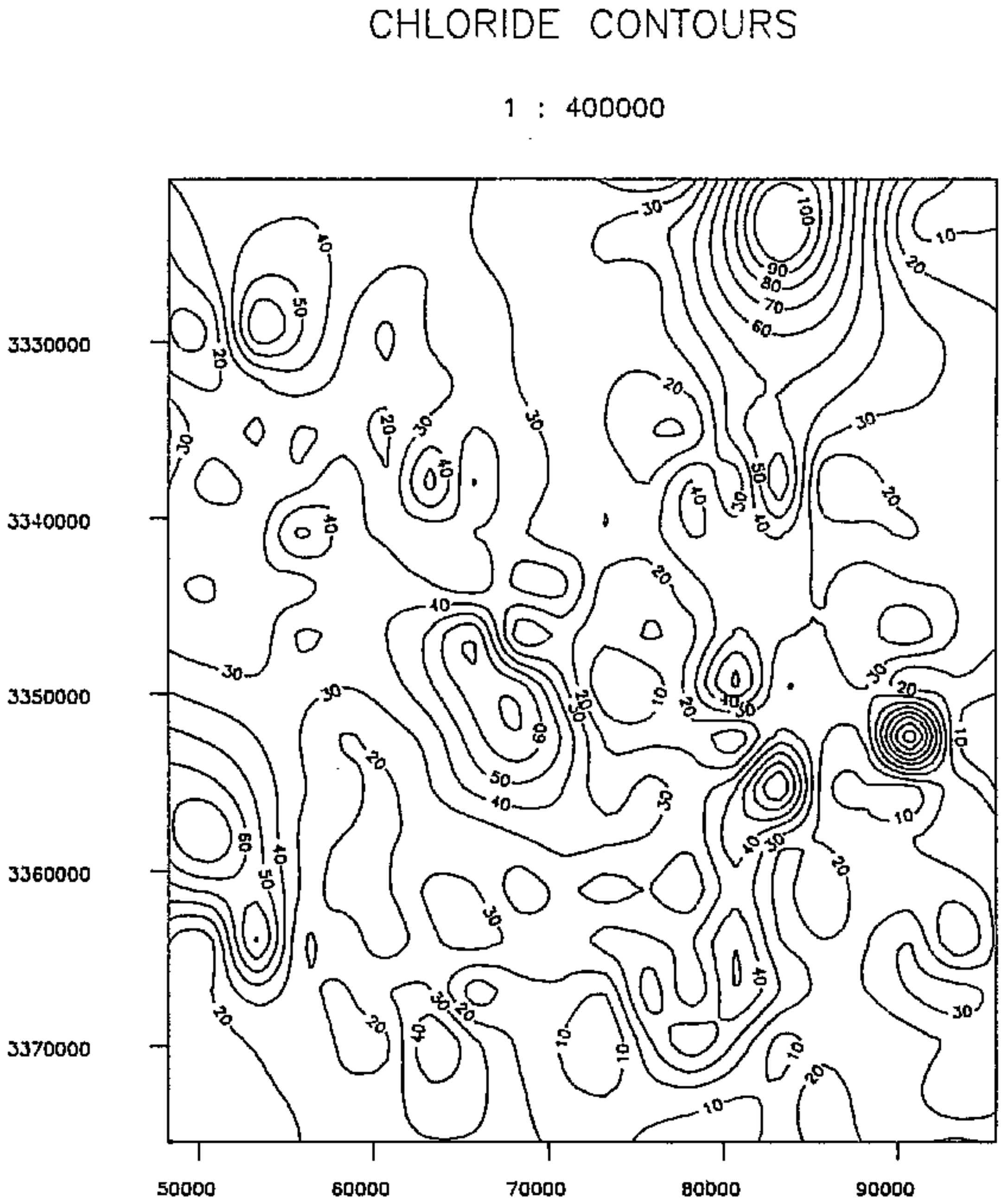


Figure 5.37. Chloride contours of the Trompsburg area in mg/l.

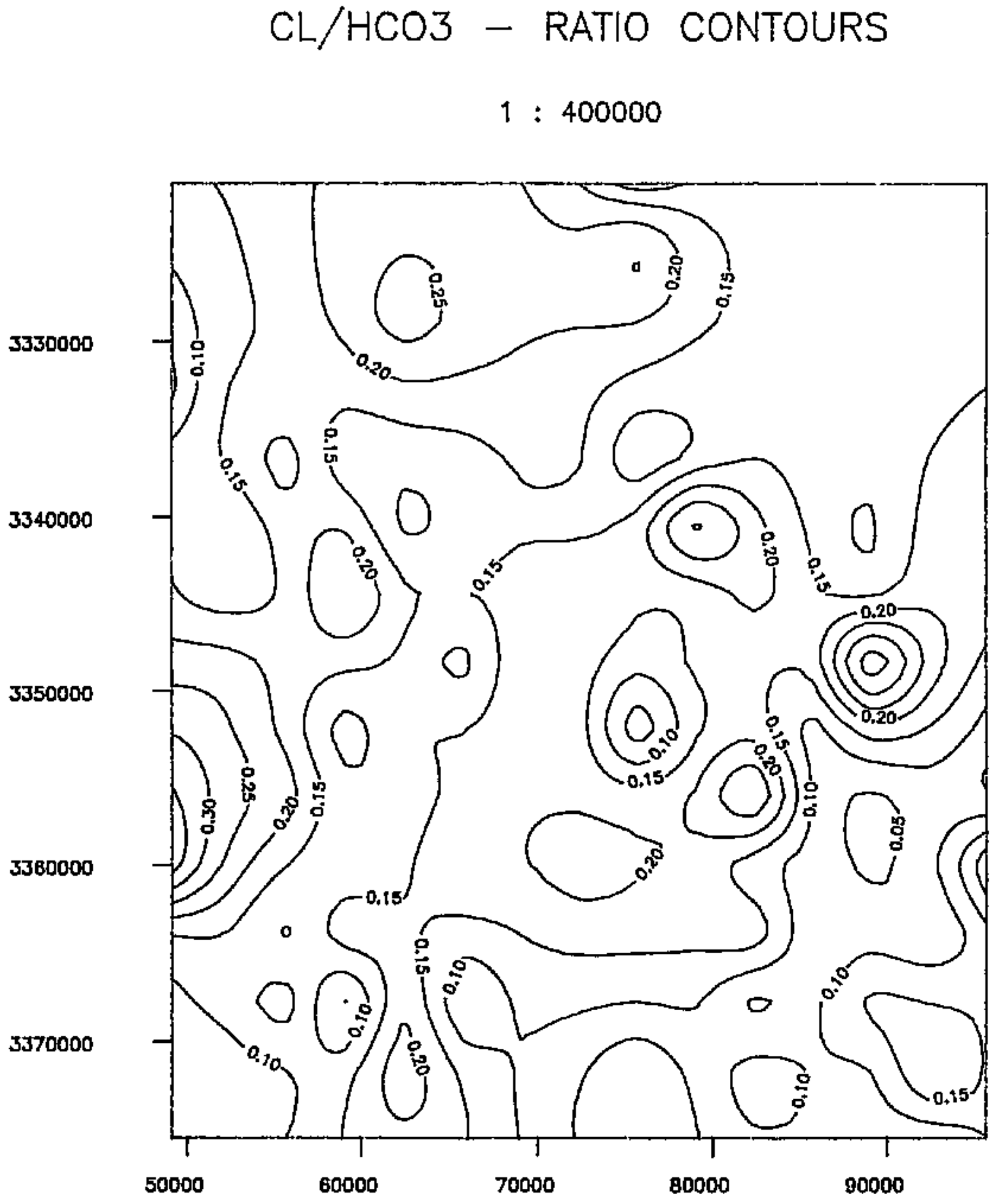


Figure 5.38. Cl/HCO3-ratio contours of the Trompsburg area.

of interest in a certain area. Tables 5.20 to 5.23 list the chemical components and ionic balances of the analyses used to calculate the ionic ratios in table 5.24. The following ionic ratios were evaluated:

(i) Cl/HCO_3
(ii) $(Cl - Na)/Cl$
(iii) $(Na + Cl + SO_4)/(Ca + Mg + HCO_3)$
(iv) $(Cl - Na)/Mg$

The reason for evaluating these ratios was to see whether it was possible to confirm, or even to improve on, the existing hydrochemical classification. When evaluating ionic ratios, one should not attach too much significance to the absolute values of the specific ratio, but rather compare the different values of the specific ratio with one another, in order to identify possible anomalous waters. However, no anomalous patterns were observed. Ground water in this area generally has a Ca/Mg-CO_3-character, which is portrayed in the low Cl/HCO_3-, $(Na + Cl + SO_4)/(Ca + Mg + HCO_3)$- and $(Cl - Na)/Mg$-ratios. The $(Cl-Na)/Cl$-ratio, generally known as the base exchange index, of almost all analyses is negative. This indicates that some base exchange is taking place. Anomalous ratio values tend to coincide with the few analyses, falling outside the Ca-HCO_3 water group.

Statistical analysis was also applied to assess the hydrochemical classification of water in the area. Three techniques were applied to the data, namely:

(i) Cluster analysis.
(ii) Principal component analysis.
(iii) Factor analysis.

Cluster and factor analysis are divided into:

(i) R-mode analysis - correlation coefficient method.
(ii) Q-mode analysis - distance coefficient method.

The R-mode analysis of the cluster and factor analyses correlate the parameters with each other, while the Q-mode compares the different water samples with each other. Both modes were calculated. Before starting with a further analysis, it is important to decide in which units to work, either mg/l or meq/l. In this example, concentrations are given in meq/l. Figures 5.39, 5.40 and 5.41 display the results of the three techniques. Because of their inherent differences, the R-mode and Q-mode analyses have to be discussed separately.

TABLE 5.20. ANALYSED CATIONS OF THE SOUTHERN ORANGE FREE STATE.

ANALYSIS OF MACRO ELEMENTS (mg/l) AND ASSOCIATED PARAMETERS

CATIONS

Site Id.	Ca	Mg	Na	K	Si
3025BA00055	32.60	27.10	25.00	1.10	16.00
3025BA00072	53.90	18.90	40.10	1.70	12.40
3025BA00076	52.00	30.40	17.00	0.90	19.50
3025BA00077	49.10	20.50	30.70	2.60	13.60
3025BA00111	44.30	23.00	37.40	1.10	12.50
3025BA00115	60.20	32.30	26.90	1.00	17.20
3025BA00120	58.10	26.60	31.00	1.00	19.10
3025BA00123	61.80	29.70	30.80	0.80	17.40
3025BA00134	63.20	31.40	22.40	1.30	19.20
3025BA00135	50.00	22.20	28.40	0.20	15.40
3025BA00137	62.00	34.60	22.40	1.10	21.50
3025BA00142	71.00	41.20	52.60	0.60	15.30
3025BA00151	55.00	29.50	10.40	0.90	20.30
3025BA00160	53.90	31.60	10.10	1.70	27.30
3025BA00161	52.70	23.90	30.10	0.60	20.00
3025BA00171	95.60	64.00	41.00	0.00	23.20
3025BA00180	53.70	22.50	25.50	0.70	17.70
3025BA00184	62.90	35.20	22.50	1.20	18.70
3025BA00185	47.20	20.60	23.50	0.90	16.60
3025BA00188	57.20	33.40	29.90	1.30	16.50
3025BA00210	60.00	20.90	10.10	1.90	18.50
3025BA00212	48.10	24.10	38.20	2.40	15.50
3025BB00008	50.60	26.60	22.90	1.30	0.00
3025BB00035	48.90	25.60	66.60	0.70	13.70
3025BB00062	56.90	24.10	23.10	1.00	16.30
3025BB00143	46.30	22.30	16.00	1.10	14.00
3025BB00151	59.50	26.60	25.30	1.00	14.90
3025BB00155	57.80	24.50	31.80	2.20	16.20
3025BB00161	74.30	35.20	49.20	2.10	13.50
3025BB00162	28.20	11.70	49.60	0.40	11.20
3025BB00170	57.20	24.40	30.50	3.20	14.60
3025BB00174	73.10	26.50	29.30	1.40	13.90
3025BB00177	65.90	25.20	29.80	0.80	13.90
3025BB00194	66.60	10.20	51.40	1.00	9.60
3025BB00200	60.60	25.10	28.70	2.40	15.50
3025BB00209	57.50	31.30	17.20	1.60	19.10
3025BB00210	91.90	38.10	19.60	3.60	16.20
3025BB00219	22.80	23.50	23.10	2.20	23.10
3025BB00221	55.50	19.50	32.10	2.30	14.10
3025BB00231	92.20	44.80	44.00	2.50	16.50
3025BB00252	61.60	38.10	15.10	0.80	19.40
3025BB00253	59.70	30.40	40.20	1.60	13.30
3025BC00039	70.70	33.20	37.50	0.90	17.80
3025BC00044	53.00	10.70	51.60	0.80	11.60
3025BC00048	50.00	23.70	39.00	0.60	13.00
3025BC00068	54.50	26.40	21.80	1.50	19.60
3025BC00076	34.40	16.10	42.30	0.40	13.10
3025BC00113	76.00	34.00	26.30	0.90	16.30
3025BC00117	40.20	20.80	17.80	1.10	16.00
3025BC00123	72.50	35.90	59.10	2.30	9.60
3025BC00128	61.00	24.40	31.40	1.40	14.80
3025BC00129	55.30	23.30	34.20	1.70	11.90
3025BC00131	41.80	26.00	20.30	0.30	16.90
3025BC00134	61.60	28.30	26.60	2.60	18.10
3025BC00144	63.00	27.00	32.60	0.70	16.20
3025BC00146	62.10	17.50	42.80	1.00	14.00
3025BC00148	63.50	19.50	31.00	1.30	12.80
3025BC00150	62.30	28.20	32.40	0.80	14.30
3025BC00155	46.10	15.60	22.20	0.70	16.30
3025BC00158	45.60	10.50	23.00	0.60	14.50
3025BC00165	65.10	31.00	19.60	0.70	16.80
3025BC00168	73.40	36.50	25.20	1.10	18.50
3025BC00170	64.60	28.20	30.40	1.10	14.50
3025BC00228	63.70	28.80	27.10	1.90	15.90
3025BC00294	79.20	64.50	51.20	0.50	28.90
3025BC00305	60.30	21.20	28.30	0.60	17.70
3025BC00356	61.60	24.40	32.50	1.20	15.60
3025BD00002	46.60	27.10	23.70	1.10	16.60
3025BD00005	50.10	28.50	25.70	2.60	16.70
3025BD00008	51.20	29.30	25.50	1.20	15.90
3025BD00013	52.80	27.00	20.00	0.40	19.20
3025BD00021	56.10	20.40	33.40	1.50	12.20
3025BD00028	39.80	24.60	26.80	1.00	7.40
3025BD00030	40.30	23.20	22.50	0.80	15.30
3025BD00033	54.40	24.10	24.80	1.40	10.00
3025BD00034	104.10	33.40	75.50	3.20	12.80
3025BD00036	35.00	30.40	16.30	1.50	24.30
3025BD00037	45.80	22.00	13.90	1.00	15.10
3025BD00046	52.00	27.20	25.80	0.60	16.80
3025BD00049	61.60	17.50	32.00	0.30	14.30
3025BD00064	30.50	25.20	35.70	0.40	12.10
3025BD00071	63.30	22.90	30.30	1.20	17.00
3025BD00087	94.40	26.40	36.70	1.60	11.20
3025BD00103	51.40	26.30	27.00	1.80	16.30
3025BD00108	46.00	23.00	26.40	4.50	12.10
3025BD00110	70.20	23.50	34.20	4.30	11.10
3025BD00121	61.80	28.80	34.10	2.00	15.00
3025BD00125	67.10	20.70	26.60	1.00	13.80
3025BD00133	44.00	10.30	25.10	0.00	14.90
3025BD00140	44.20	16.30	33.70	0.40	10.00
3025BD00142	51.10	26.20	13.70	0.80	23.70
3025BD00148	61.70	23.80	43.40	1.10	12.80
3025BD00155	51.20	32.30	27.20	0.90	15.90
3025BD00156	41.50	8.50	62.30	0.30	9.50
3025BD00157	69.30	27.70	25.50	5.20	15.90
3025BD00169	54.10	23.40	21.70	1.00	16.30
3025BD00171	65.30	25.30	39.10	0.00	12.00

TABLE 5.21. ANALYSED ANIONS OF THE SOUTHERN ORANGE FREE STATE.

ANIONS
======

Site Id.	CO3	HCO3	SO4	Cl	NO3	F
3025BA00055	0.00	0.00	20.20	30.40	6.40	0.30
3025BA00072	0.00	0.00	19.90	26.30	5.50	0.60
3025BA00076	0.00	0.00	27.80	11.00	0.10	0.20
3025BA00077	0.00	0.00	16.50	25.40	6.20	0.50
3025BA00111	0.00	0.00	21.90	31.30	1.50	0.50
3025BA00115	0.00	0.00	34.00	27.10	5.80	0.40
3025BA00120	0.00	0.00	20.30	22.90	2.50	0.30
3025BA00123	0.00	0.00	24.40	25.60	0.10	0.50
3025BA00134	0.00	0.00	21.60	27.10	3.00	0.30
3025BA00135	0.00	0.00	20.40	21.80	2.80	0.50
3025BA00137	0.00	0.00	25.10	16.00	0.00	0.30
3025BA00142	0.00	0.00	32.00	46.60	0.00	0.60
3025BA00151	0.00	0.00	16.00	20.90	1.00	0.30
3025BA00160	0.00	0.00	34.90	22.70	2.60	0.20
3025BA00161	0.00	0.00	19.20	23.70	3.30	0.50
3025BA00171	0.00	0.00	40.80	65.90	5.20	0.40
3025BA00180	0.00	0.00	19.90	25.00	1.50	0.40
3025BA00184	0.00	0.00	21.60	49.40	2.20	0.30
3025BA00185	0.00	0.00	16.40	20.00	2.20	0.30
3025BA00188	0.00	0.00	26.40	28.90	2.10	0.40
3025BA00210	0.00	0.00	33.30	19.60	1.20	0.40
3025BA00212	0.00	0.00	22.00	30.20	6.80	0.80
3025BB00008	0.00	0.00	29.00	12.30	5.00	0.10
3025BB00035	0.00	0.00	28.40	43.50	0.10	0.20
3025BB00062	0.00	0.00	33.50	18.30	4.10	0.90
3025BB00143	0.00	0.00	14.90	11.00	0.50	0.10
3025BB00151	0.00	0.00	19.30	24.10	2.50	0.30
3025BB00155	0.00	0.00	23.70	24.10	7.30	0.50
3025BB00161	0.00	0.00	90.00	95.10	3.00	0.80
3025BB00162	0.00	0.00	28.20	15.90	0.10	0.90
3025BB00170	0.00	0.00	14.40	30.40	3.80	0.80
3025BB00174	0.00	0.00	29.90	40.60	10.00	0.60
3025BB00177	0.00	0.00	22.10	25.70	2.90	0.40
3025BB00194	0.00	0.00	25.20	45.30	6.20	0.90
3025BB00200	0.00	0.00	22.00	20.30	2.70	0.60
3025BB00209	0.00	0.00	20.00	15.00	0.90	0.10
3025BB00210	0.00	0.00	32.70	37.50	21.60	0.10
3025BB00219	0.00	0.00	14.10	17.50	0.60	0.40
3025BB00221	0.00	0.00	15.50	23.30	3.60	0.60
3025BB00231	0.00	0.00	57.00	77.20	7.40	0.50
3025BB00252	0.00	0.00	26.20	29.80	2.80	0.40
3025BB00253	0.00	0.00	28.50	32.60	2.40	0.70
3025BC00039	0.00	0.00	62.40	42.20	1.90	0.50
3025BC00044	0.00	0.00	25.00	21.00	1.80	0.40
3025BC00048	0.00	0.00	21.10	26.50	1.50	0.60
3025BC00068	0.00	0.00	25.90	20.60	2.50	0.40
3025BC00076	0.00	0.00	16.30	25.30	0.00	0.50
3025BC00113	0.00	0.00	72.80	26.80	4.00	0.60
3025BC00117	0.00	0.00	15.70	16.00	1.70	0.30
3025BC00123	0.00	0.00	88.30	43.00	0.00	0.30
3025BC00128	0.00	0.00	43.30	31.90	5.00	0.40
3025BC00129	0.00	0.00	19.90	28.80	3.00	0.50
3025BC00131	0.00	0.00	27.40	14.20	2.50	0.50
3025BC00134	0.00	0.00	28.80	25.90	1.90	0.80
3025BC00144	0.00	0.00	42.00	24.70	6.70	0.60
3025BC00146	0.00	0.00	27.60	28.70	7.20	0.70
3025BC00148	0.00	0.00	30.30	20.50	7.70	0.60
3025BC00150	0.00	0.00	45.60	26.60	6.90	0.60
3025BC00155	0.00	0.00	23.90	8.00	2.00	0.30
3025BC00158	0.00	0.00	23.40	8.00	1.50	0.40
3025BC00165	0.00	0.00	23.10	16.10	0.10	0.30
3025BC00168	0.00	0.00	40.50	23.40	5.10	0.40
3025BC00170	0.00	0.00	42.90	39.30	2.00	0.60
3025BC00228	0.00	0.00	39.30	29.40	4.00	0.60
3025BC00294	0.00	0.00	74.30	90.30	5.70	0.60
3025BC00305	0.00	0.00	23.40	26.30	10.60	0.40
3025BC00356	0.00	0.00	33.40	29.40	7.90	0.40
3025BD00002	0.00	0.00	24.60	23.30	4.20	0.50
3025BD00005	0.00	0.00	25.90	24.40	2.50	0.60
3025BD00008	0.00	0.00	17.20	24.70	4.30	0.50
3025BD00013	0.00	0.00	21.90	8.10	0.40	0.40
3025BD00021	0.00	0.00	21.40	25.00	3.70	0.60
3025BD00028	0.00	0.00	19.40	16.20	0.00	0.50
3025BD00030	0.00	0.00	35.40	17.20	0.00	0.50
3025BD00033	0.00	0.00	27.80	23.30	2.70	0.40
3025BD00034	0.00	0.00	41.70	70.10	10.00	0.60
3025BD00036	0.00	0.00	29.60	11.40	1.90	0.30
3025BD00037	0.00	0.00	10.20	5.40	0.00	0.30
3025BD00046	0.00	0.00	19.80	15.60	2.00	0.40
3025BD00049	0.00	0.00	67.00	22.70	0.10	0.50
3025BD00064	0.00	0.00	46.10	20.30	2.40	0.50
3025BD00071	0.00	0.00	27.40	22.40	5.50	0.70
3025BD00087	0.00	0.00	93.40	34.30	1.30	0.70
3025BD00103	0.00	0.00	22.00	21.60	4.60	0.50
3025BD00108	0.00	0.00	22.30	10.00	5.10	0.60
3025BD00118	0.00	0.00	41.30	18.50	9.70	1.10
3025BD00121	0.00	0.00	33.70	20.00	5.70	0.70
3025BD00125	0.00	0.00	24.40	17.00	3.00	0.70
3025BD00133	0.00	0.00	10.20	6.50	2.50	0.80
3025BD00140	0.00	0.00	47.60	15.20	1.80	0.50
3025BD00142	0.00	0.00	20.20	7.10	0.10	0.40
3025BD00148	0.00	0.00	24.10	21.30	0.70	0.70
3025BD00155	0.00	0.00	26.20	15.10	0.40	0.50
3025BD00156	0.00	0.00	40.30	25.60	1.30	0.90
3025BD00157	0.00	0.00	29.80	11.10	10.10	1.10
3025BD00169	0.00	0.00	21.20	9.00	1.70	0.40
3025BD00171	0.00	0.00	40.30	23.50	0.40	0.80

TABLE 5.22. ANALYSED PHYSICAL PARAMETERS OF THE SOUTHERN ORANGE FREE STATE.

PHYSICAL PARAMETERS

Site Id.	pH	Cond	Alk	TDS	TEMP
3025BA00055	7.05	59.10	154.00	159.90	16.50
3025BA00072	7.47	58.10	220.90	179.40	18.00
3025BA00076	8.34	53.80	244.10	160.60	19.50
3025BA00077	8.11	58.80	213.90	173.10	18.50
3025BA00111	8.18	57.60	213.50	174.30	17.00
3025BA00115	7.95	70.00	260.20	212.90	18.00
3025BA00120	8.10	61.40	250.10	181.80	18.00
3025BA00123	7.38	63.50	279.30	191.10	0.00
3025BA00134	7.35	61.80	256.00	190.50	0.00
3025BA00135	7.43	55.70	232.10	161.70	16.30
3025BA00137	8.00	63.40	284.30	184.60	16.00
3025BA00142	7.96	85.50	357.00	260.00	15.00
3025BA00151	7.41	55.00	246.90	163.90	17.00
3025BA00160	7.47	55.90	219.90	185.10	16.20
3025BA00161	7.60	58.30	240.50	173.50	18.00
3025BA00171	7.79	105.50	400.30	327.70	14.80
3025BA00180	7.58	55.90	220.40	166.90	17.00
3025BA00184	7.42	64.70	233.90	214.00	16.00
3025BA00185	7.51	49.70	207.30	147.70	17.40
3025BA00188	7.59	64.40	272.40	196.10	19.50
3025BA00210	7.73	60.00	227.60	182.60	0.00
3025BA00212	7.63	62.80	214.90	188.10	16.00
3025BB00000	8.15	52.90	207.40	148.70	0.00
3025BB00035	7.99	70.70	260.00	227.80	18.80
3025BB00062	8.02	56.20	217.50	178.20	16.60
3025BB00143	7.48	44.60	211.10	126.20	0.00
3025BB00151	7.52	50.30	251.80	173.50	17.50
3025BB00155	7.62	59.90	227.90	180.10	0.00
3025BB00161	7.56	92.00	203.40	364.00	17.00
3025BB00162	7.87	51.20	174.50	146.20	17.30
3025BB00170	7.44	61.40	259.80	187.30	0.00
3025BB00174	7.65	68.10	215.50	225.30	0.00
3025BB00177	7.67	61.00	254.10	186.70	17.30
3025BB00194	7.52	64.40	208.10	217.20	18.00
3025BB00200	7.64	58.10	242.60	177.90	0.00
3025BB00209	7.56	56.30	252.50	162.70	18.00
3025BB00210	7.54	81.20	249.70	261.30	0.00
3025BB00219	7.31	39.60	163.70	127.40	0.00
3025BB00221	7.44	54.90	220.10	166.50	16.00
3025BB00231	7.71	96.90	305.40	342.10	17.50
3025BB00252	7.79	63.00	264.30	194.20	17.00
3025BB00253	7.85	68.20	290.20	209.40	17.00
3025BC00039	7.90	75.90	265.20	267.10	0.00
3025BC00044	7.72	61.60	263.00	183.90	16.00
3025BC00048	7.55	59.30	244.70	184.00	16.00
3025BC00068	7.56	57.00	232.40	173.30	0.00
3025BC00076	8.06	40.80	182.00	140.40	0.00
3025BC00113	7.88	73.90	252.50	258.50	0.00
3025BC00117	7.99	48.40	193.30	138.40	0.00
3025BC00123	7.76	85.30	291.90	311.00	0.00
3025BC00128	8.07	61.80	213.20	213.60	0.00
3025BC00129	8.11	60.40	229.40	179.40	0.00
3025BC00131	7.97	50.70	191.50	149.90	0.00
3025BC00134	7.82	60.60	246.70	194.60	0.00
3025BC00144	7.90	66.50	233.70	213.50	0.00
3025BC00146	7.87	61.90	231.50	201.60	0.00
3025BC00148	7.98	59.30	214.10	187.20	0.00
3025BC00150	8.07	66.40	227.60	217.70	0.00
3025BC00155	8.03	45.90	195.90	135.10	0.00
3025BC00158	7.90	42.30	173.10	127.50	0.00
3025BC00165	8.02	60.60	278.30	172.80	0.00
3025BC00168	7.91	71.40	290.50	224.10	0.00
3025BC00170	7.84	60.30	250.90	232.40	0.00
3025BC00228	7.87	63.30	236.70	210.70	0.00
3025BC00294	7.60	102.60	359.20	395.20	16.00
3025BC00305	7.51	55.50	105.40	188.80	0.00
3025BC00356	7.67	62.50	212.30	206.40	0.00
3025BD00002	8.07	54.00	200.90	167.70	16.00
3025BD00005	7.57	56.30	205.50	177.10	17.50
3025BD00008	8.01	58.00	220.10	169.80	16.00
3025BD00013	7.18	54.10	239.30	150.20	15.50
3025BD00021	7.70	58.20	217.10	174.50	16.00
3025BD00028	7.87	49.80	215.70	136.60	0.00
3025BD00030	7.59	51.40	200.40	163.20	17.00
3025BD00033	7.39	56.40	218.30	176.90	16.20
3025BD00034	7.45	104.40	363.50	351.40	16.00
3025BD00036	8.15	54.20	229.60	158.70	15.80
3025BD00037	7.61	44.80	200.30	113.70	12.30
3025BD00046	7.69	56.30	230.50	160.20	17.00
3025BD00049	7.66	58.90	182.20	216.00	17.30
3025BD00064	8.09	55.40	183.60	189.20	0.00
3025BD00071	7.61	62.00	234.50	191.60	17.20
3025BD00087	7.84	77.40	266.60	300.10	17.50
3025BD00103	7.63	58.00	217.30	171.60	15.00
3025BD00108	8.09	52.70	208.10	150.40	20.40
3025BD00110	7.79	68.20	255.20	214.00	17.80
3025BD00121	7.87	65.90	242.00	209.90	18.60
3025BD00125	7.46	58.70	233.10	176.00	18.00
3025BD00133	7.76	46.90	192.30	131.20	10.20
3025BD00140	7.91	49.70	158.90	170.30	19.50
3025BD00142	7.52	49.40	218.90	143.60	19.60
3025BD00148	7.67	63.90	271.10	189.60	17.00
3025BD00155	7.75	60.40	260.10	169.70	18.60
3025BD00156	7.76	57.50	191.80	190.30	21.00
3025BD00157	7.43	64.60	244.00	195.70	17.00
3025BD00169	7.20	53.40	223.60	148.80	16.60
3025BD00171	7.34	64.50	261.00	200.50	16.00

TABLE 5.23. IONIC BALANCE OF THE ANALYSED SAMPLES IN THE SOUTHERN ORANGE FREE STATE.

PERCENTAGE CATION - ANION BALANCE ERROR

Site Id.	(%) Ion Balance Error	
3025BA00055	4.90	
3025BA00072	0.80	
3025BA00076	0.34	
3025BA00077	2.99	
3025BA00111	0.72	
3025BA00115	2.52	
3025BA00120	1.79	
3025BA00123	-0.50	
3025BA00134	1.55	
3025BA00135	-2.45	
3025BA00137	1.16	
3025BA00142	-0.34	
3025BA00151	0.16	
3025BA00160	-1.23	
3025BA00161	-1.30	
3025BA00171	1.88	
3025BA00180	-1.46	
3025BA00184	2.69	
3025BA00185	-0.92	
3025BA00188	-0.33	
3025BA00210	2.29	
3025BA00212	2.36	
3025BB00008	4.03	
3025BB00035	2.13	
3025BB00062	0.99	
3025BB00143	-0.87	
3025BB00151	0.01	
3025BB00155	3.07	
3025BB00161	0.00	
3025BB00162	-0.83	
3025BB00170	0.57	
3025BB00174	5.86	Unreliable Analysis
3025BB00177	1.82	
3025BB00194	2.18	
3025BB00200	2.83	
3025BB00209	1.67	
3025BB00210	9.19	Unreliable Analysis
3025BB00219	-0.24	
3025BB00221	2.46	
3025BB00231	2.58	
3025BB00252	0.27	
3025BB00253	-1.56	
3025BC00039	-0.27	
3025BC00044	-0.77	
3025BC00048	2.58	
3025BC00068	-0.39	
3025BC00076	0.94	
3025BC00113	1.52	
3025BC00117	1.33	
3025BC00123	0.91	
3025BC00128	1.58	
3025BC00129	1.78	
3025BC00131	1.79	
3025BC00134	1.60	
3025BC00144	2.50	
3025BC00146	1.42	
3025BC00148	3.65	
3025BC00150	2.87	
3025BC00155	-2.17	
3025BC00158	-1.53	
3025BC00165	0.22	
3025BC00168	1.64	
3025BC00170	0.36	
3025BC00228	1.59	
3025BC00294	-0.24	
3025BC00305	7.13	Unreliable Analysis
3025BC00356	4.16	
3025BD00002	2.36	
3025BD00005	2.58	
3025BD00008	4.00	
3025BD00013	1.25	
3025BD00021	2.65	
3025BD00028	-0.57	
3025BD00030	-0.11	
3025BD00033	0.50	
3025BD00034	3.90	
3025BD00036	-0.18	
3025BD00037	2.79	
3025BD00046	3.17	
3025BD00049	1.21	
3025BD00064	-0.03	
3025BD00071	2.36	
3025BD00087	0.72	
3025BD00103	3.06	
3025BD00108	3.40	
3025BD00118	1.93	
3025BD00121	3.16	
3025BD00125	3.55	
3025BD00133	2.89	
3025BD00140	3.67	
3025BD00142	2.05	
3025BD00148	2.08	
3025BD00155	-0.41	
3025BD00156	-0.24	
3025BD00157	6.75	Unreliable Analysis
3025BD00169	2.66	
3025BD00171	1.53	

TABLE 5.24. TABULATED IONIC RATIOS OF GROUND WATERS FROM THE SOUTHERN ORANGE FREE STATE.

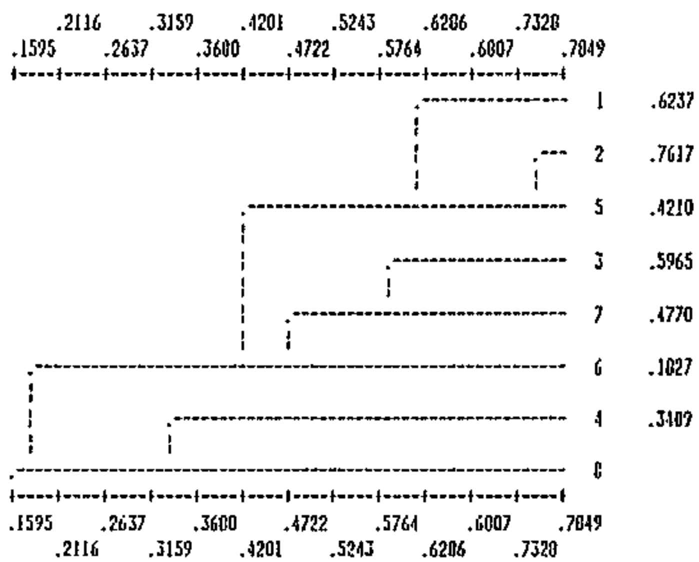
Site Id.	Cl / HCO3	SO4 / Cl	Cl-Na / Cl	Na+Cl+SO4 / Ca+Mg+HCO3	Ca / Mg	Na-Cl / SO4	Cl-Na / Mg	Cl-Na / SO4+HCO3+NO3
3025BA00055	0.27	0.49	-0.07	0.35	0.73	0.13	-0.03	-0.015
3025BA00072	0.16	0.56	-1.16	0.32	1.73	2.00	-0.60	-0.166
3025BA00076	0.06	1.06	-1.22	0.16	1.05	0.65	-0.17	-0.068
3025BA00077	0.16	0.40	-1.05	0.32	1.45	2.19	-0.51	-0.156
3025BA00111	0.20	0.51	-0.69	0.35	1.13	1.34	-0.34	-0.126
3025BA00115	0.14	0.92	-0.34	0.23	1.20	0.37	-0.11	-0.043
3025BA00120	0.13	0.65	-0.02	0.24	1.32	1.26	-0.28	-0.095
3025BA00123	0.13	0.70	-0.72	0.23	1.26	1.03	-0.23	-0.084
3025BA00134	0.15	0.59	-0.12	0.20	1.22	0.21	-0.04	-0.016
3025BA00135	0.13	0.69	-0.92	0.25	1.37	1.34	-0.32	-0.109
3025BA00137	0.08	1.10	-0.01	0.17	1.10	0.74	-0.15	-0.061
3025BA00142	0.18	0.51	-0.66	0.30	1.05	1.30	-0.27	-0.108
3025BA00151	0.12	0.59	-0.21	0.17	1.15	0.35	-0.06	-0.023
3025BA00160	0.14	1.13	0.40	0.19	1.03	-0.35	0.11	0.049
3025BA00161	0.14	0.60	-0.00	0.25	1.32	1.33	-0.30	-0.099
3025BA00171	0.23	0.46	0.12	0.26	0.80	-0.27	0.05	0.025
3025BA00180	0.15	0.59	-0.46	0.24	1.45	0.70	-0.19	-0.063
3025BA00184	0.29	0.32	0.39	0.27	1.08	-1.22	0.22	0.104
3025BA00185	0.13	0.60	-0.65	0.23	1.39	1.00	-0.24	-0.080
3025BA00188	0.15	0.67	-0.48	0.24	1.04	0.71	-0.15	-0.064
3025BA00210	0.12	1.25	-0.23	0.21	1.26	0.18	-0.06	-0.024
3025BA00212	0.19	0.54	-0.71	0.34	1.21	1.33	-0.35	-0.123
3025BB00008	0.08	1.78	-1.64	0.22	1.15	0.92	-0.20	-0.116
3025BB00035	0.23	0.48	-1.12	0.47	1.16	2.33	-0.73	-0.233
3025BB00062	0.12	1.35	-0.74	0.24	1.43	0.55	-0.22	-0.074
3025BB00143	0.07	1.00	-1.07	0.16	1.26	1.07	-0.20	-0.071
3025BB00151	0.13	0.59	-0.49	0.21	1.36	0.83	-0.17	-0.060
3025BB00155	0.15	0.72	-0.76	0.27	1.43	1.05	-0.30	-0.078
3025BB00161	0.65	0.70	0.24	0.63	1.20	-0.34	0.24	0.107
3025BB00162	0.13	1.31	-3.53	0.53	1.46	2.70	-1.75	-0.301
3025BB00170	0.16	0.35	-0.80	0.20	1.42	2.20	-0.37	-0.121
3025BB00174	0.26	0.54	0.07	0.31	1.67	-0.13	0.05	0.016
3025BB00177	0.14	0.63	-0.61	0.23	1.59	0.96	-0.24	-0.077
3025BB00194	0.30	0.41	-0.60	0.40	3.96	1.46	-1.00	-0.157
3025BB00200	0.12	0.80	-0.90	0.23	1.46	1.13	-0.29	-0.095
3025BB00209	0.08	0.98	-0.54	0.15	1.11	0.55	-0.10	-0.041
3025BB00210	0.21	0.64	0.37	0.22	1.46	-0.50	0.16	0.064
3025BB00219	0.15	0.59	-0.71	0.20	0.59	1.20	-0.22	-0.096
3025BB00221	0.15	0.49	-0.87	0.26	1.73	1.77	-0.40	-0.117
3025BB00231	0.35	0.54	0.21	0.37	1.25	-0.39	0.14	0.061
3025BB00252	0.16	0.65	0.29	0.18	0.98	-0.45	0.09	0.041
3025BB00253	0.16	0.64	-0.85	0.29	1.19	1.31	-0.32	-0.119
3025BC00039	0.22	1.09	-0.28	0.36	1.29	0.25	-0.13	-0.049
3025BC00044	0.11	0.88	-2.62	0.35	1.72	2.99	-1.06	-0.261
3025BC00048	0.15	0.59	-1.02	0.29	1.40	1.74	-0.44	-0.139
3025BC00060	0.12	0.93	-0.47	0.22	1.25	0.51	-0.14	-0.051
3025BC00076	0.19	0.47	-1.32	0.42	1.30	2.77	-0.79	-0.231
3025BC00113	0.15	2.00	-0.37	0.30	1.36	0.18	-0.11	-0.041
3025BC00117	0.12	0.69	-0.43	0.20	1.41	0.62	-0.14	-0.047
3025BC00123	0.20	1.51	-1.00	0.45	1.23	0.66	-0.44	-0.156
3025BC00128	0.21	1.00	-0.36	0.34	1.52	0.36	-0.18	-0.061
3025BC00129	0.17	0.51	-0.66	0.29	1.44	1.29	-0.31	-0.103
3025BC00131	0.10	1.42	-0.91	0.23	0.98	0.64	-0.20	-0.080
3025BC00134	0.14	0.82	-0.40	0.24	1.32	0.49	-0.14	-0.052
3025BC00144	0.15	1.25	-0.79	0.30	1.42	0.63	-0.20	-0.075
3025BC00146	0.17	0.71	-1.00	0.35	2.15	1.53	-0.67	-0.161
3025BC00148	0.13	1.09	-1.02	0.28	1.98	0.94	-0.43	-0.115
3025BC00150	0.16	1.26	-0.65	0.31	1.34	0.52	-0.24	-0.086
3025BC00155	0.06	2.20	-2.97	0.22	1.79	1.35	-0.56	-0.148
3025BC00158	0.06	2.15	-3.07	0.25	2.63	1.43	-0.87	-0.171
3025BC00165	0.08	1.06	-0.71	0.16	1.27	0.68	-0.14	-0.053
3025BC00168	0.11	1.27	-0.40	0.21	1.22	0.38	-0.12	-0.046
3025BC00170	0.22	0.80	-0.40	0.35	1.39	0.50	-0.21	-0.073
3025BC00220	0.17	0.78	-0.27	0.28	1.34	0.28	-0.11	-0.040
3025BC00294	0.35	0.61	0.19	0.39	0.74	-0.32	0.10	0.055
3025BC00305	0.20	0.65	-0.30	0.29	1.73	0.46	-0.16	-0.051
3025BC00356	0.19	0.84	-0.45	0.31	1.53	0.53	-0.22	-0.072
3025BD00002	0.16	0.78	-0.36	0.26	1.04	0.46	-0.12	-0.050
3025BD00005	0.16	0.70	-0.40	0.27	1.72	0.51	-0.19	-0.050
3025BD00008	0.15	0.51	-0.35	0.23	1.06	0.60	-0.12	-0.049
3025BD00013	0.05	1.99	-2.32	0.16	1.19	1.17	-0.27	-0.099
3025BD00021	0.16	0.63	-0.83	0.29	1.67	1.31	-0.39	-0.118
3025BD00028	0.10	0.88	-1.46	0.24	0.98	1.66	-0.34	-0.139
3025BD00030	0.12	1.51	-0.84	0.26	1.26	0.55	-0.23	-0.084
3025BD00033	0.15	0.88	-0.46	0.25	1.37	0.53	-0.17	-0.060
3025BD00034	0.27	0.44	-0.48	0.40	1.89	1.09	-0.39	-0.112
3025BD00036	0.07	1.91	-0.92	0.17	0.55	0.48	-0.11	-0.055
3025BD00037	0.04	1.39	-2.37	0.12	1.26	1.71	-0.23	-0.084
3025BD00046	0.09	0.93	-1.11	0.20	1.18	1.19	-0.26	-0.095
3025BD00049	0.17	2.17	-0.96	0.42	2.14	0.44	-0.47	-0.120
3025BD00064	0.21	1.20	-0.81	0.43	0.93	0.60	-0.34	-0.136
3025BD00071	0.13	0.90	-0.82	0.26	1.60	0.91	-0.31	-0.095
3025BD00087	0.18	2.00	-0.56	0.37	2.17	0.28	-0.26	-0.073
3025BD00103	0.14	0.75	-0.65	0.24	1.19	0.87	-0.21	-0.080
3025BD00108	0.07	1.64	-2.54	0.22	1.21	1.55	-0.44	-0.149
3025BD00110	0.10	1.64	-1.61	0.27	1.81	0.98	-0.47	-0.135
3025BD00121	0.16	0.89	-0.64	0.29	1.30	0.72	-0.24	-0.088
3025BD00125	0.10	1.06	-1.09	0.22	1.97	1.03	-0.35	-0.098
3025BD00133	0.05	2.06	-4.09	0.21	1.46	1.99	-0.50	-0.173
3025BD00140	0.13	2.30	-1.98	0.42	1.60	0.86	-0.70	-0.199
3025BD00142	0.04	2.09	-1.47	0.13	1.10	0.70	-0.16	-0.060
3025BD00148	0.11	0.83	-1.84	0.20	1.57	2.20	-0.62	-0.182
3025BD00155	0.08	1.20	-1.50	0.20	0.96	1.23	-0.27	-0.111
3025BD00156	0.10	1.16	-2.57	0.63	2.96	2.22	-2.79	-0.309
3025BD00157	0.06	1.90	-1.88	0.19	1.52	0.95	-0.32	-0.102
3025BD00169	0.06	1.73	-2.20	0.10	1.40	1.27	-0.34	-0.111
3025BD00171	0.12	1.26	-1.34	0.30	1.57	1.07	-0.47	-0.144

SIMILARITY MATRIX

	Ca 1	Mg 2	Na 3	K 4	HCO3 5	SO4 6	Cl 7	NO3 8
1	1.00	.53	.29	.23	.72	.54	.61	.44
2	.53	1.00	.02	.07	.76	.37	.59	.13
3	.29	.02	1.00	.21	.35	.40	.60	.13
4	.23	.07	.21	1.00	.08	.13	.24	.34
5	.72	.76	.35	.08	1.00	.27	.53	.11
6	.54	.37	.40	.13	.27	1.00	.56	.06
7	.61	.59	.60	.24	.53	.56	1.00	.24
8	.44	.13	.13	.34	.11	.06	.24	1.00

OBSERVATIONS COMBINED INTO CLUSTERS		SIMILARITY LEVEL OF CLUSTERING
2	5	.7617
1	2	.6237
3	7	.5965
4	8	.3409
3	6	.4770
1	3	.4210
1	4	.1027

DENDROGRAM



VALUES ALONG X-AXIS ARE SIMILARITIES

Figure 5.39(a). R-mode cluster analysis of samples from the Trompsburg area.

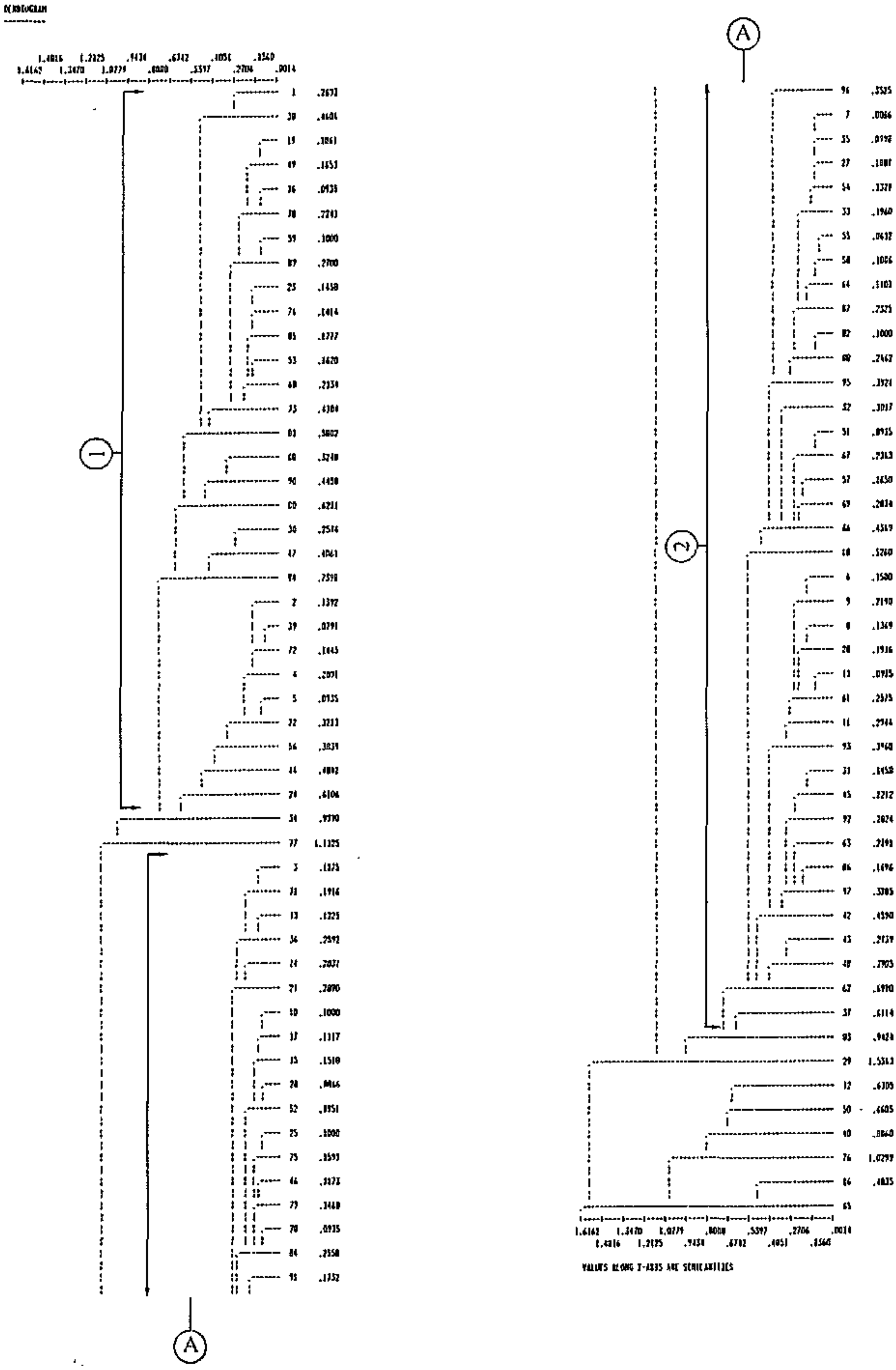


Figure 5.39(b). Q-mode cluster analysis of samples from the Trompsburg area.

R-mode analysis

This technique compares the variables with each another. In this analysis, the variables are the major ions of the respective water samples. They comprise the ions: Ca^{++}, Mg^{++}, Na^{+}, K^{+}, HCO_3^{-}, SO_4^{--}, Cl^{-} and NO_3^{-}, which are (in this order) numbered 1 to 8 in figure 5.39. In figure 5.39(a), it can be noticed that certain ions tend to correlate with each other and coincide well with specific ground-water types. For example, Mg^{++} and HCO_3^{-} are combined into one cluster (Group I), as well as Ca^{++} and HCO_3^{-} (Group II) (see figure 5.39(b)). If the significant principal component scores, as well as the significant rotated factor scores, are plotted against each other, a similar pattern, as observed in the cluster analysis, will be found, namely that variables 1 and 5 (Ca-HCO_3) will group together, as well as variables 2 and 5 (Mg-HCO_3). These results corroborate the results obtained in the hydrochemical diagrams, where the ground water displayed a Ca(Mg)-HCO_3 character.

Q-mode analysis

During a Q-mode analysis, the different water analyses are compared with each other, in order to depict various groups of ground water. Again, the 100 water analyses from the Trompsburg area were used. Studying the results of the Q-mode analysis, the same two groups, a Mg-HCO_3-group and a Ca-HCO_3-group, can be discerned, which were seen on the dendrogram of the cluster analysis. In the dendrogram, the numbers starting from 1 onwards, refer to the consecutive order of the water analyses in table 5.20. Figures 5.40 and 5.41 are plots of the principal component and rotated factor scores respectively.

The two groups cannot be discerned on the central diagram of the Piper- and Durov plots, where the sum of Ca^{++} and Mg^{++} is plotted against HCO_3^{-} . They, however, show on the Schoeller diagram. The samples which do not fall in any one of the two groups (Ca-HCO_3-group and Mg-HCO_3-group) coincide with the scattered samples in the Piper- and Durov diagrams.

No long-term hydrochemical data of the region are available in the ground-water data base for possible determination of trends or cycles.

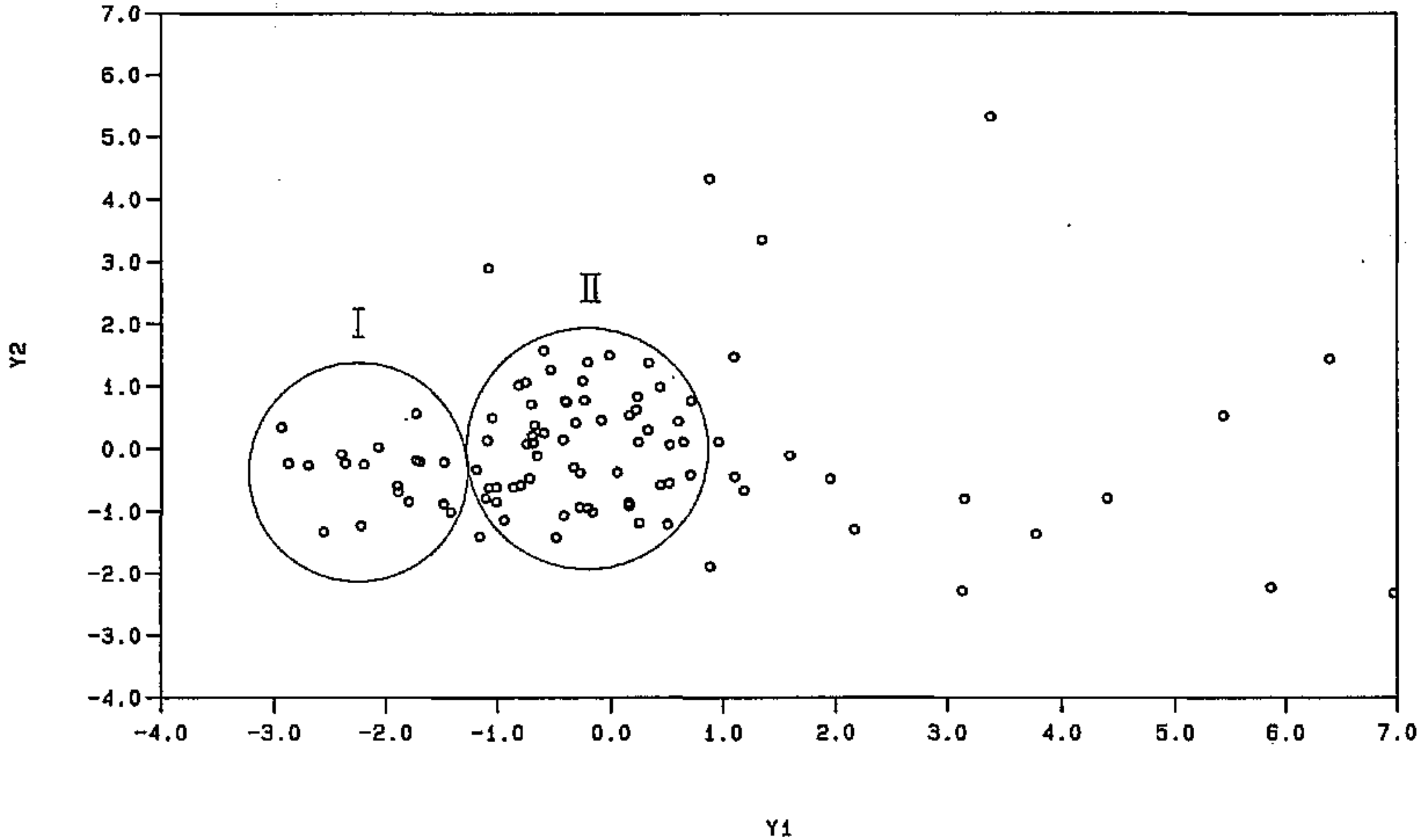


Figure 5.40. Graphical representation of the significant principal component scores.

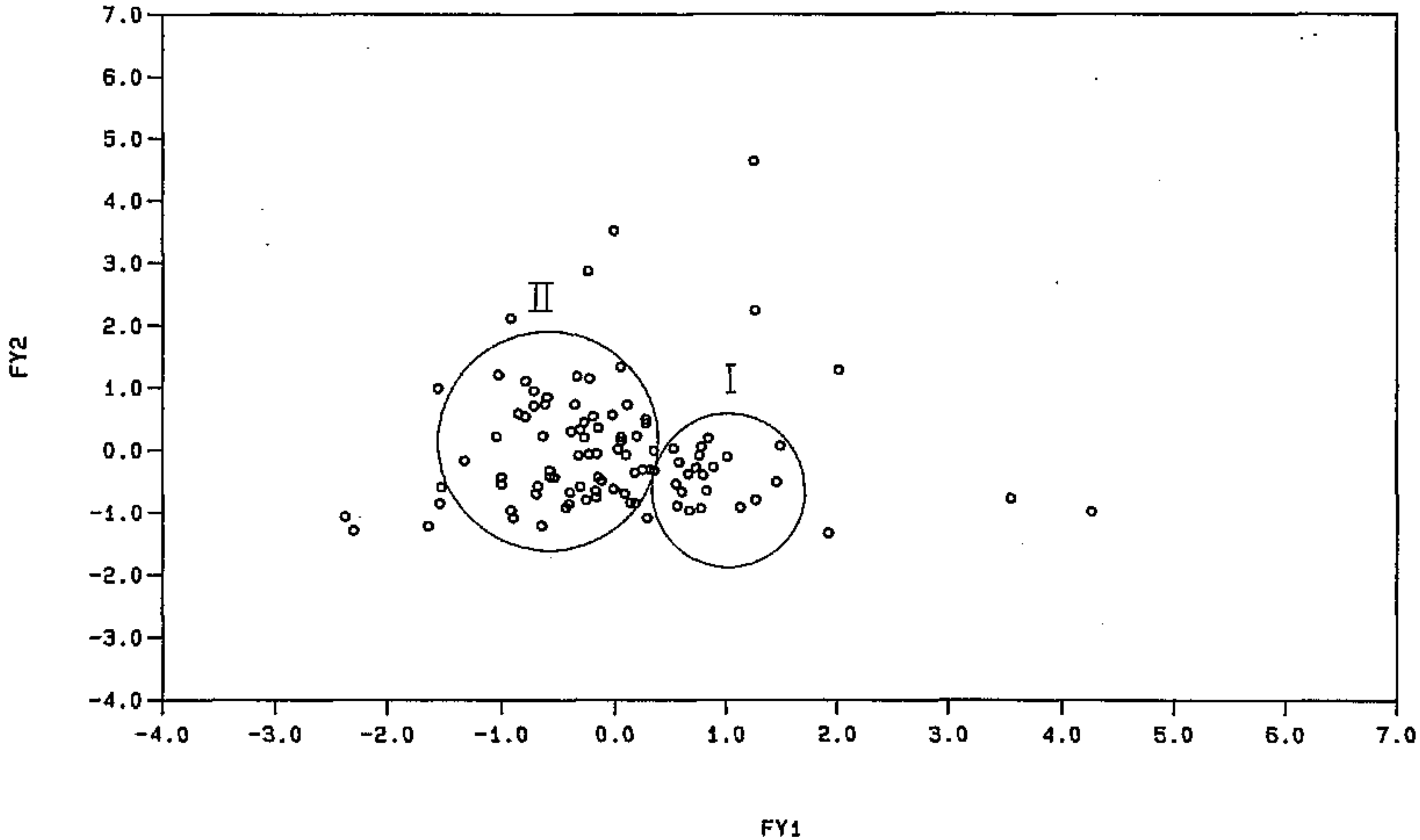


Figure 5.41. *Graphical representation of the significant rotated factor scores.*

Atlantis area

As discussed previously, one of the more important aspects of a geohydrological evaluation is the assessment of the chemical composition of the water. In the case of the Atlantis aquifer, this facet is especially important, because a whole community depends on the ground-water resources of the Atlantis area.

Figures 5.42 and 5.43 show the chemical analyses on Piper and Durov diagrams. According to these diagrams, the ground water can be divided into two groups. Group I waters display a $Ca\text{-}HCO_3$-character and seem to be subject to dissolution and mixing with ground water from Group II. Group II, on the other hand, has a NaCl-character.

According to Van der Merwe (1983), the general quality of the ground water in the Atlantis area is satisfactory. This is, to a certain extent, depicted in the Durov diagram (see figure 5.43), where the TDS-values (according to electrical conductivities) range between ± 200 mg/l and 1 400 mg/l.

Similar to the water-level observation boreholes, hydrochemical sample locations were unevenly distributed over the area (see figure 5.44). The hydrochemical contour maps do, therefore, not cover the entire area equally well (see figures 5.45 - 5.50), which may influence the significance of the evaluation. Two potential risk areas fall within the area under investigation:

(i) Silwerstroom area.
(ii) Witzand area.

From both areas, ground-water is abstracted at a large scale for the water supply of Atlantis. Ground water in these areas should therefore be periodically sampled to identify any possible contamination of the aquifers by sea-water intrusion in the case of the Silwerstroom area, or from the infiltration pans in the Witzand area.

Table 5.25 represents various ionic ratios of the water samples taken in the Atlantis area. If the ratios are studied, the two ratios $(Na+Cl+SO_4)/(Ca+Mg+HCO_3)$ and Cl/HCO_3 support the previous classification of ground water in the Atlantis area. The higher numerical values indicate the NaCl-waters, whereas the $Ca\text{-}HCO_3$-waters have values <1. The intermediate values (indicated in area A on figure 5.42) depict possible dissolution of NaCl or mixing of the two ground-water types.

By using statistical evaluation techniques, e.g.:

(i) cluster analysis,

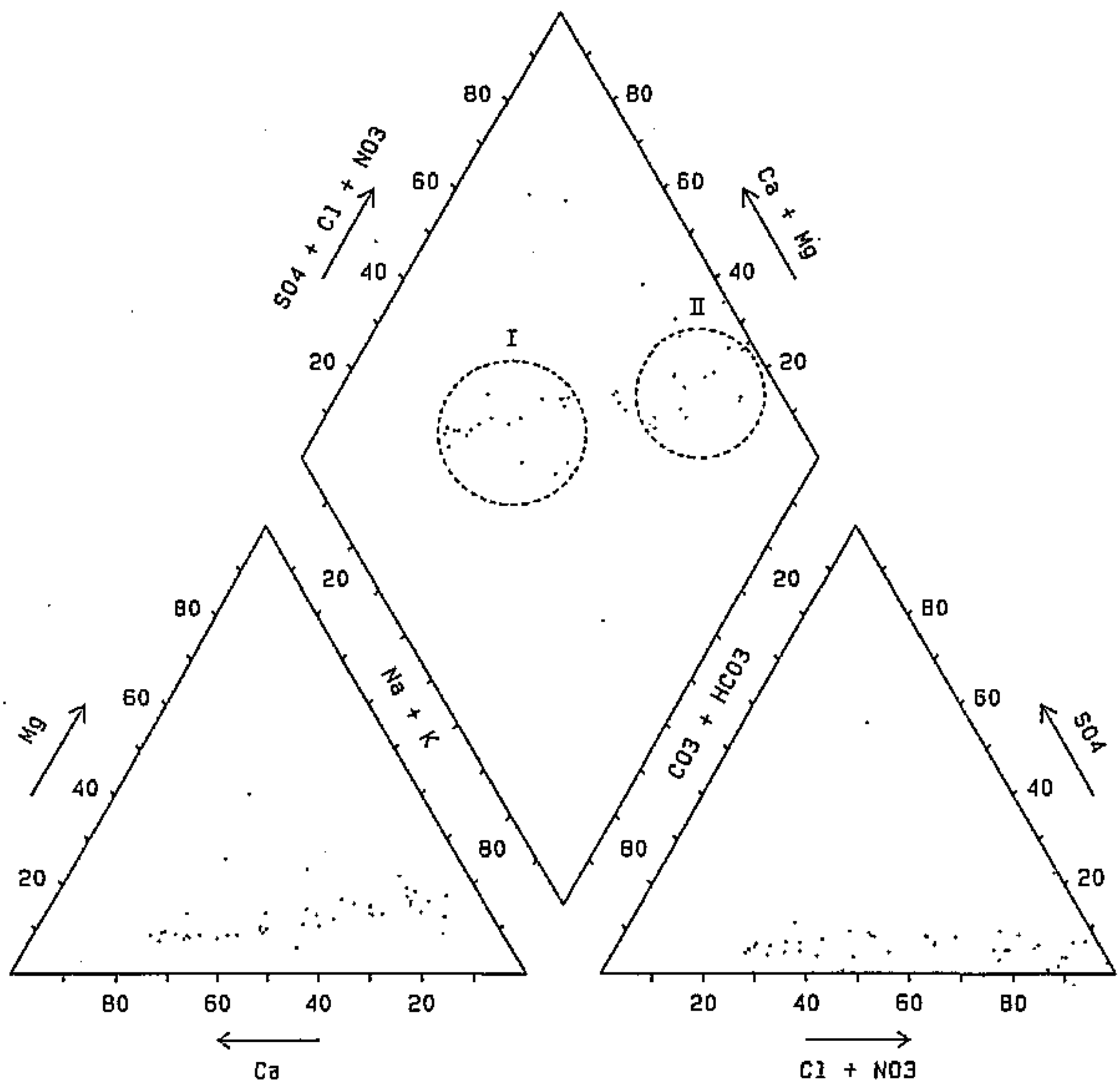


Figure 5.42. *Piper diagram of water samples in the Atlantis area.*

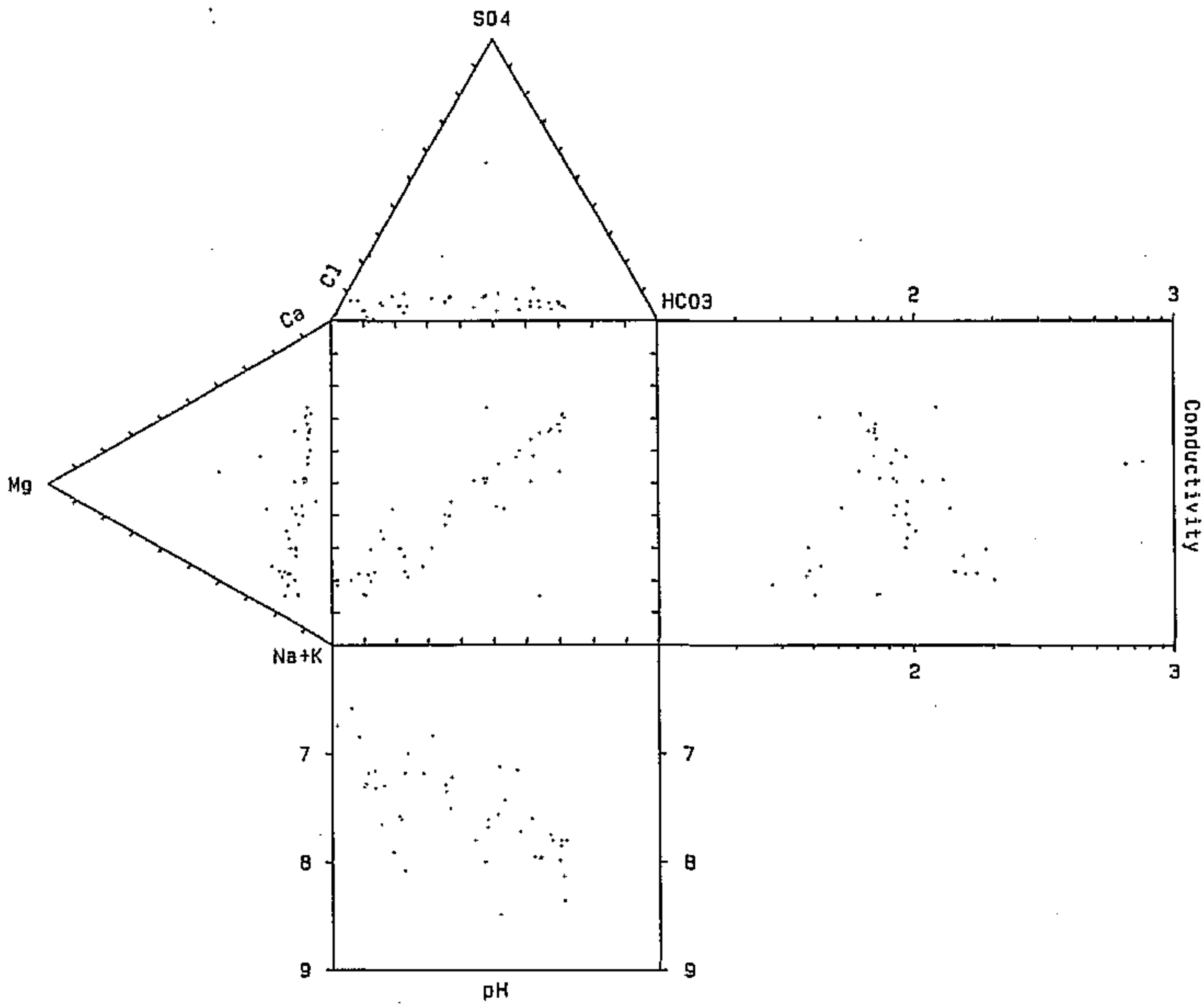


Figure 5.43. *Durov diagram of water samples in the Atlantis area.*

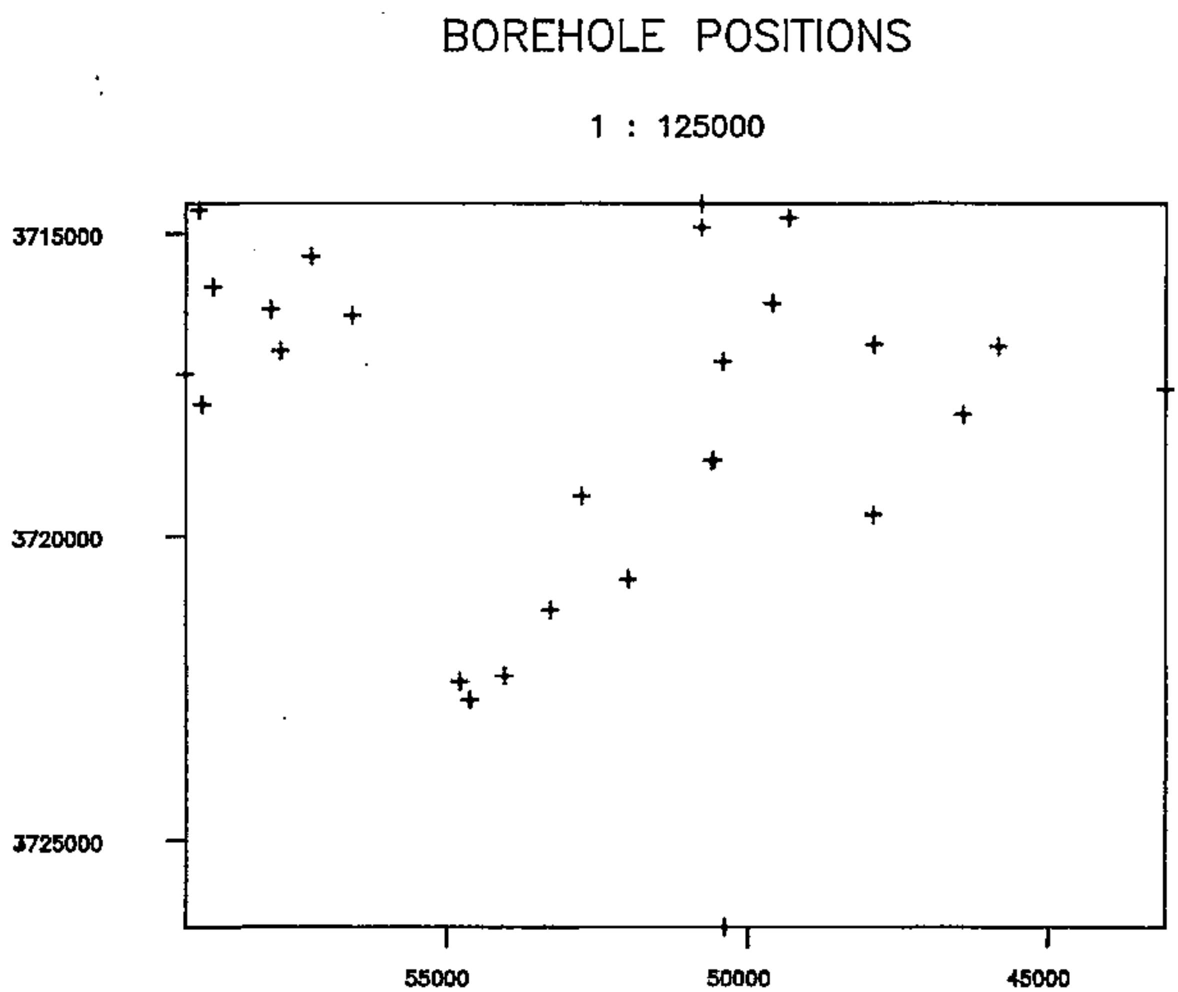


Figure 5.44. *Distribution map of water sample locations in the Atlantis area.*

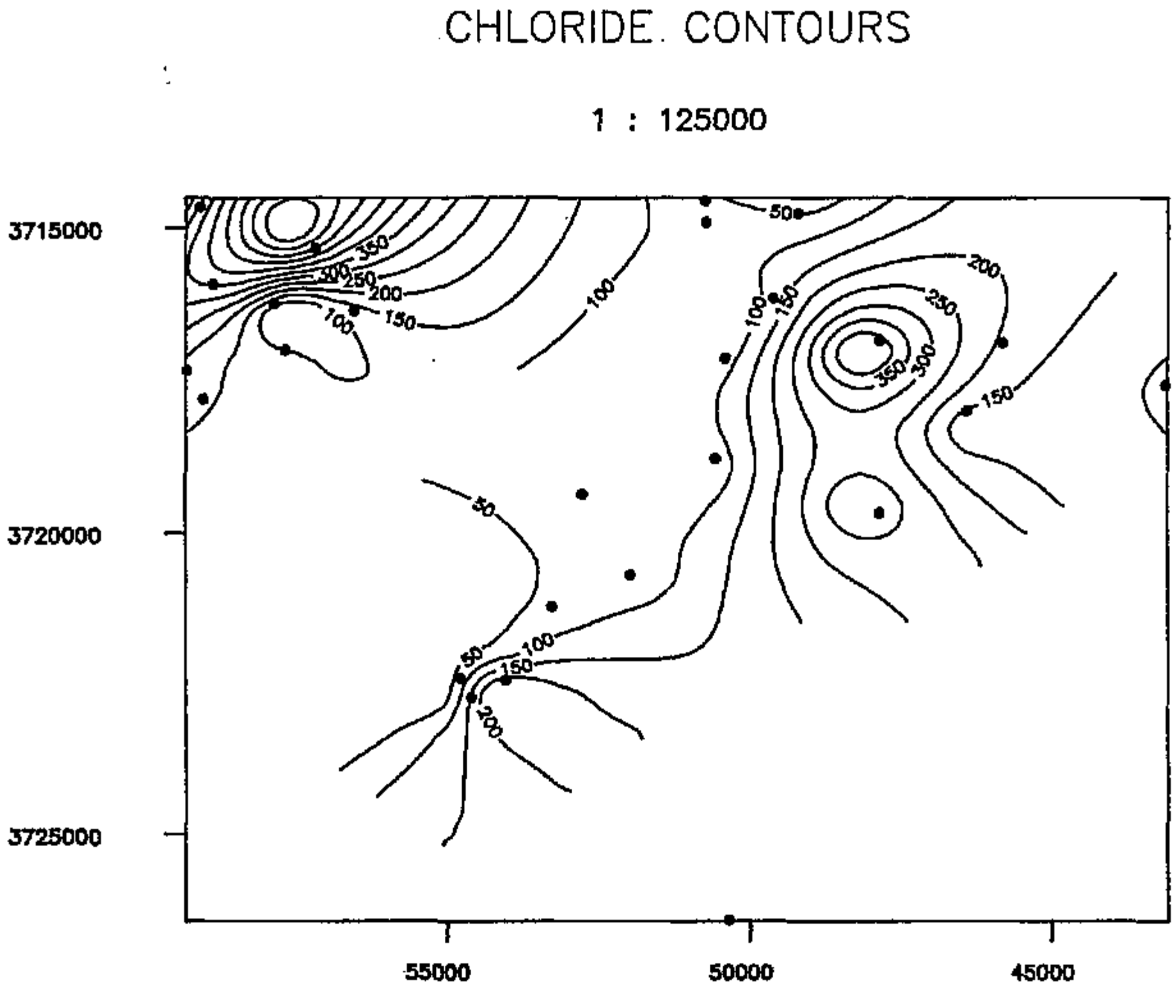


Figure 5.45. Chloride contours in mg/l of water samples in the Atlantis area.

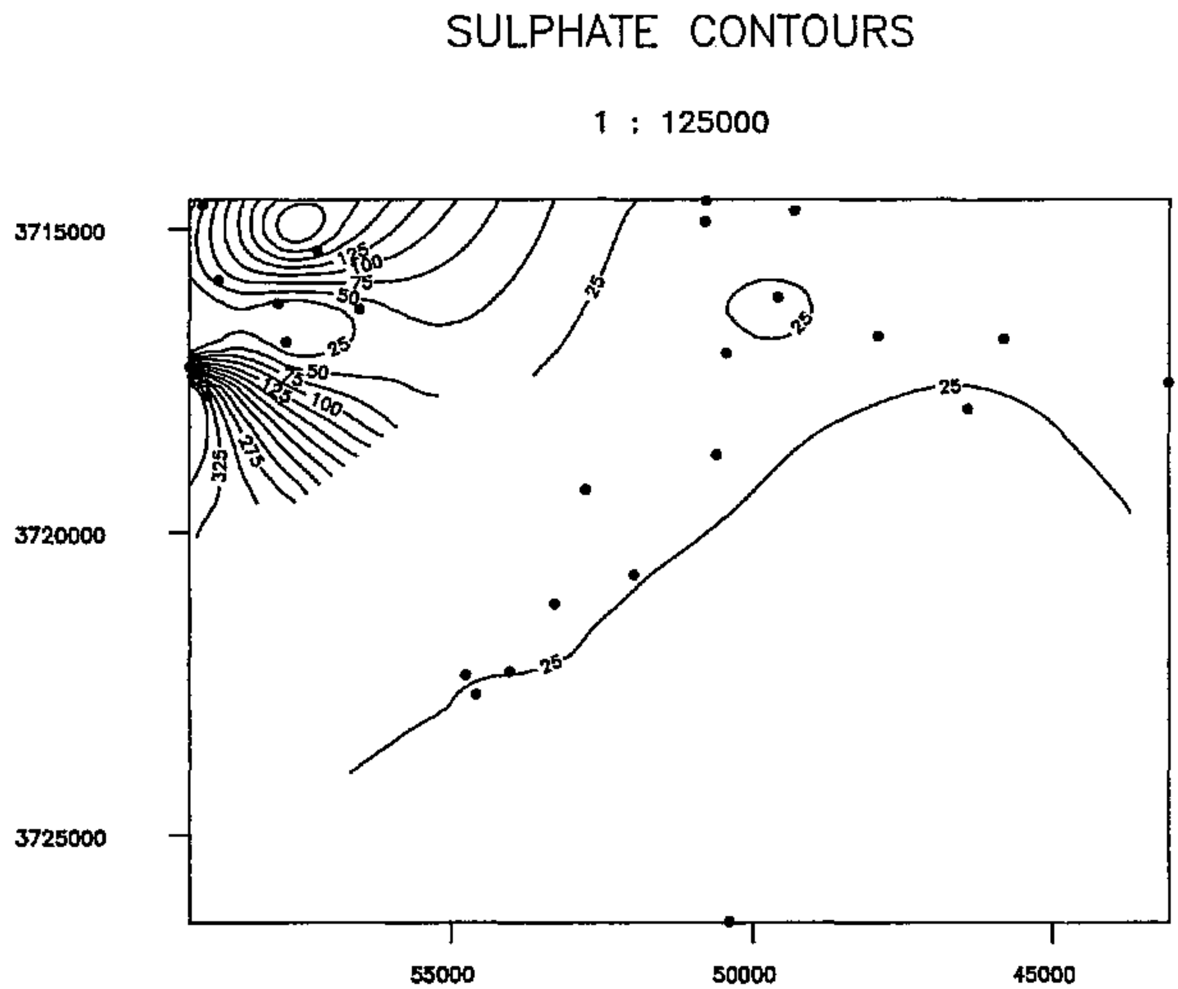


Figure 5.46. Sulphate contours in mg/l of water samples in the Atlantis area.

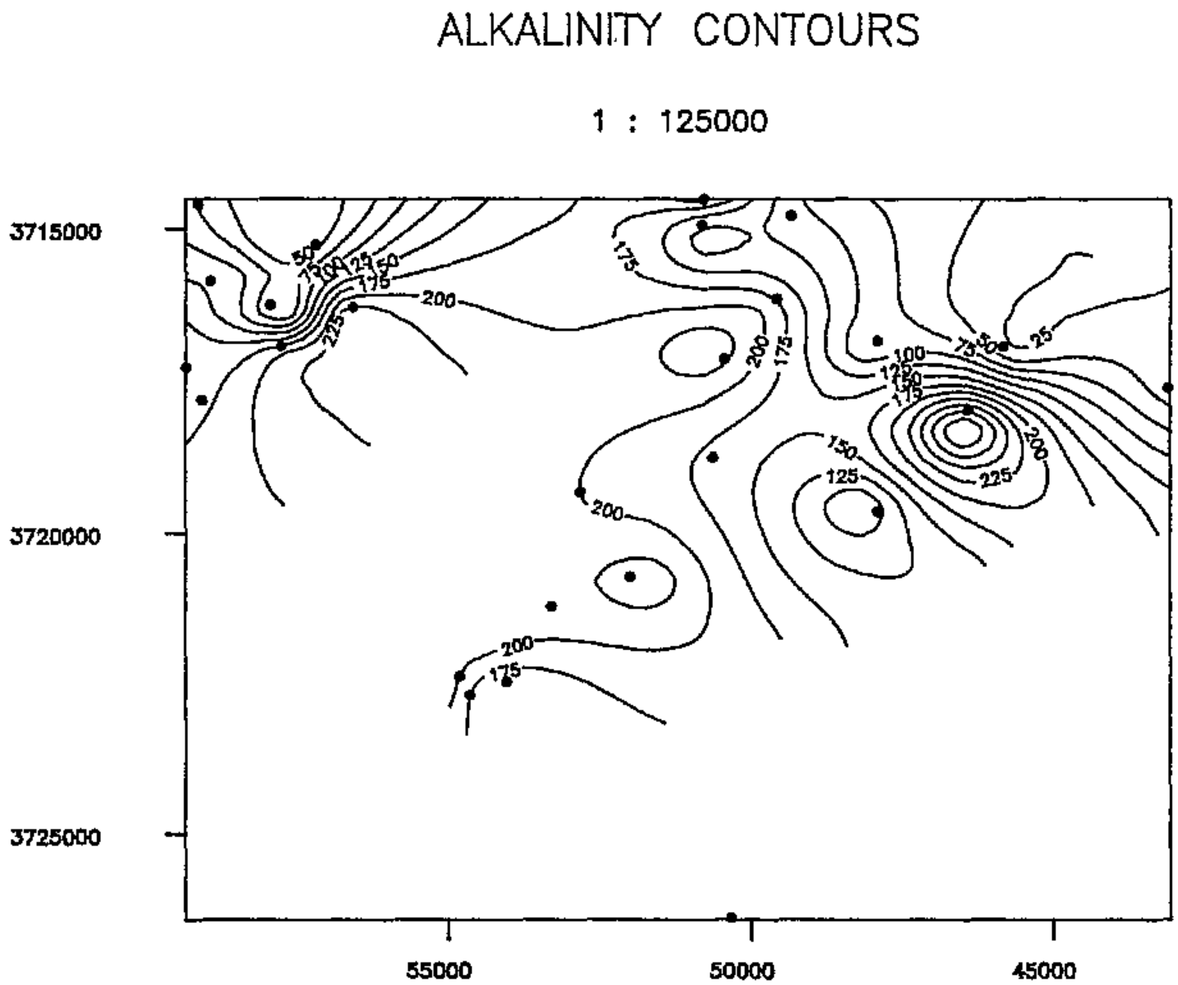


Figure 5.47. Total alkalinity contours in mg/l of water samples in the Atlantis area.

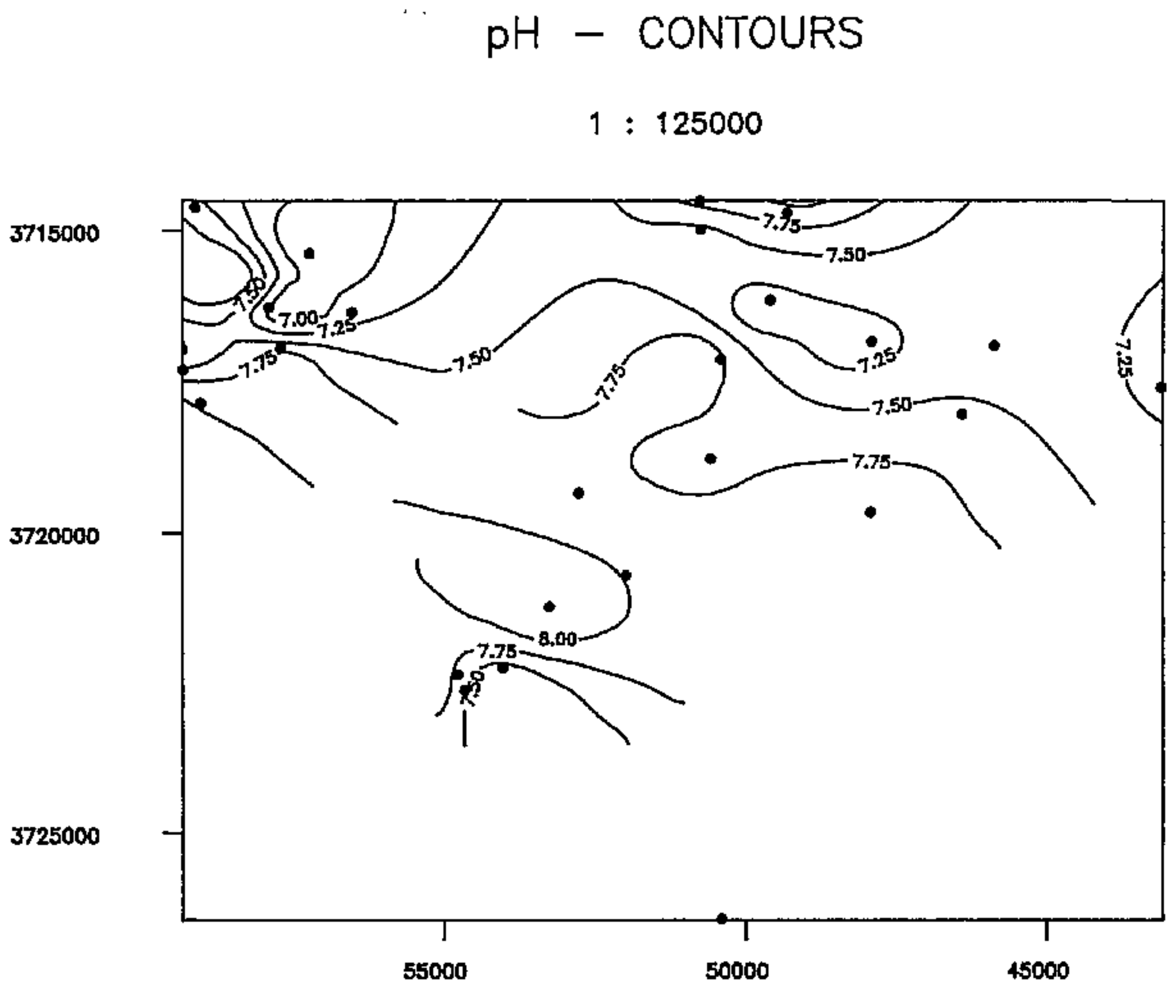


Figure 5.48. pH contours of water samples in the Atlantis area.

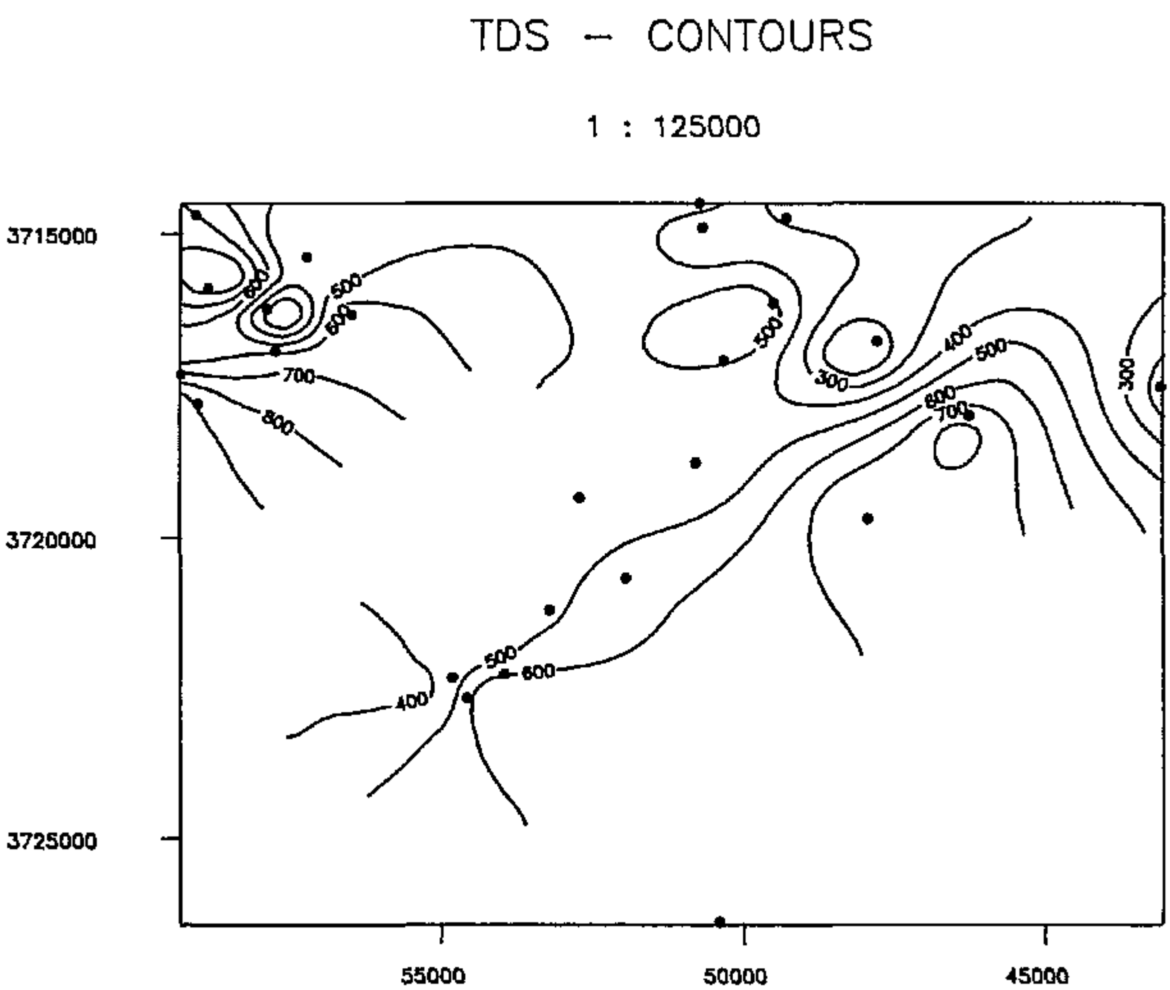


Figure 5.49. TDS contours of water samples in mg/l in the Atlantis area.

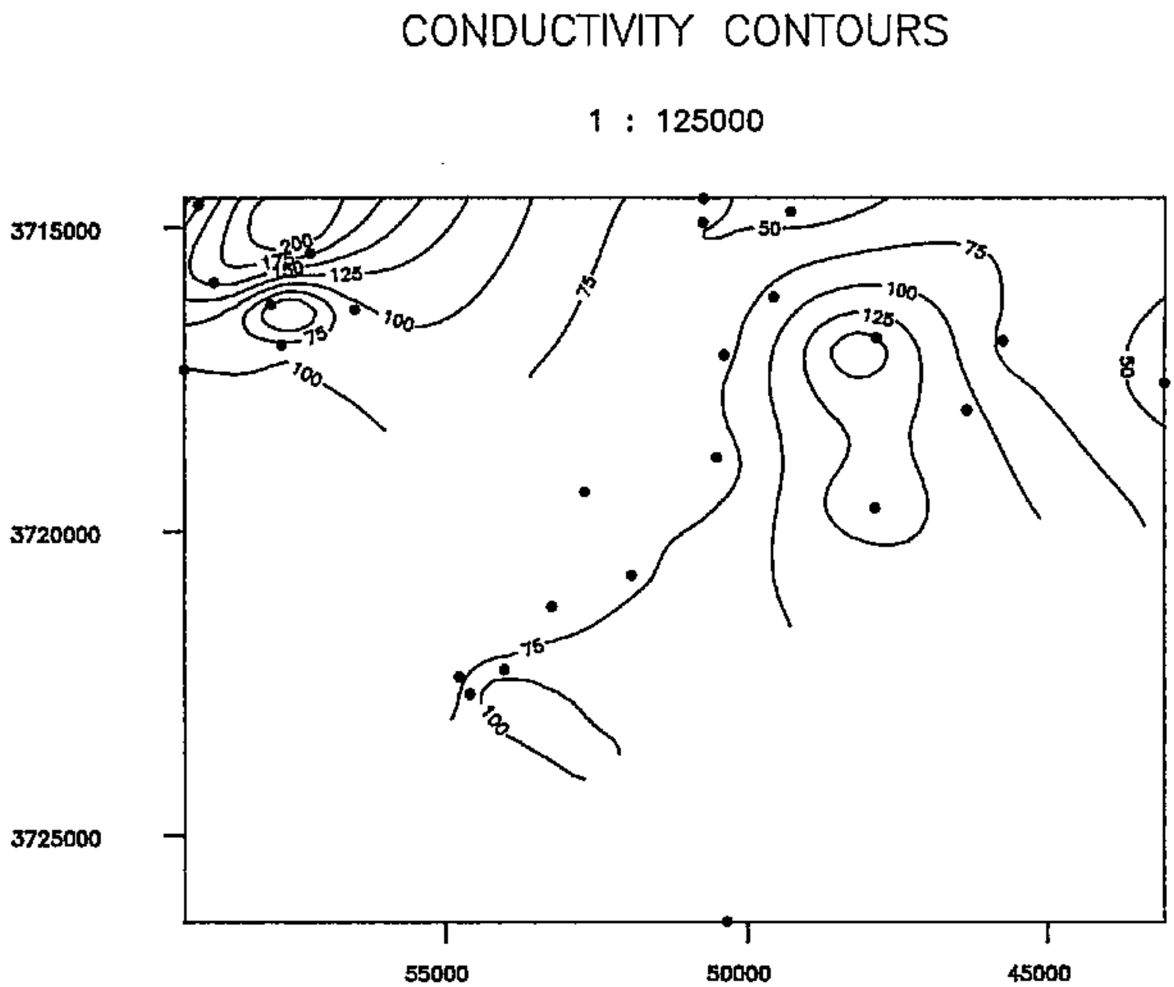


Figure 5.50. Conductivity contours of water samples in mS/m in the Atlantis area.

TABLE 5.25. TABULATED IONIC RATIOS OF THE ATLANTIS AREA.

IONIC RATIOS
============

Site Id.	Cl / HCO3	SO4 / Cl	Cl-Na / Cl	Na+Cl+SO4 / Ca+Mg+HCO3	Ca / Mg	Na-Cl / SO4	Cl-Na / Mg	Cl-Na / SO4+HCO3+NO3
3318CB00007	2.40	0.12	0.14	2.15	1.75	-1.14	0.76	0.259
3318CB00010	0.94	0.21	0.12	0.86	6.00	-0.57	0.63	0.096
3318CB00013	0.74	0.20	0.05	0.73	6.23	-0.27	0.24	0.035
3318CB00014	0.61	0.19	-0.15	0.73	3.22	0.77	-0.40	-0.082
3318CB00014	0.55	0.14	0.14	0.52	7.73	-1.00	0.60	0.071
3318CB00015	0.47	0.13	0.08	0.47	8.05	-0.63	0.33	0.037
3318CB00018	0.62	0.14	0.16	0.57	6.02	-1.19	0.69	0.093
3318CB00020	0.41	0.24	0.03	0.44	7.29	-0.12	0.09	0.011
3318CB00026	10.21	0.08	0.23	4.81	0.94	-2.92	1.35	1.720
3318CB00026	39.06	0.08	0.21	5.80	0.96	-2.63	1.37	1.909
3318CB00026	101.00	0.29	0.20	4.50	1.76	-0.60	1.22	0.662
3318CB00041	4.07	0.13	0.11	2.84	1.57	-0.81	0.50	0.278
3318CB00042	1.29	0.09	0.19	1.06	5.04	-2.21	1.15	0.225
3318CB00059	5.25	0.11	0.42	1.93	1.28	-3.81	1.41	1.403
3318CB00060	1.84	0.14	0.11	1.57	3.22	-0.80	0.61	0.160
3318CB00060	1.12	0.17	0.11	1.02	4.53	-0.62	0.52	0.100
3318CB00060	1.15	0.17	0.14	1.01	4.79	-0.86	0.73	0.139
3318CB00060	1.22	2.30	0.13	1.04	7.89	-0.06	0.37	0.043
3318CB00071	0.39	0.20	0.11	0.39	7.10	-0.57	0.33	0.041
3318CB00071	0.30	0.20	0.09	0.39	9.26	-0.45	0.34	0.033
3318CB00101	8.45	0.02	0.21	3.94	0.86	-13.22	1.18	1.591
3318CB00101	6.73	0.01	0.21	3.54	0.81	-24.95	1.08	1.356
3318CB00101	6.74	0.01	0.21	3.54	0.81	-26.02	1.08	1.370
3318CB00112	0.72	0.10	0.13	0.67	7.10	-1.38	0.69	0.089
3318CB00123	0.47	0.21	0.07	0.48	7.24	-0.31	0.23	0.020
3318CB00124	1.95	0.11	0.18	1.53	2.57	-1.58	0.83	0.200
3318CB00124	1.93	0.10	0.12	1.65	2.54	-1.13	0.59	0.186
3318CB00124	1.11	0.19	0.13	0.99	4.45	-0.69	0.60	0.119
3318CB00126	6.21	0.07	0.31	2.73	1.59	-4.57	1.67	1.350
3318CB00126	6.81	0.08	0.33	2.61	1.62	-4.32	1.68	1.490
3318CB00126	4.03	0.07	0.20	2.53	1.52	-2.90	1.02	0.627
3318CB00126	4.10	0.07	0.21	2.56	1.45	-3.06	1.05	0.679
3318CB00127	0.43	0.01	0.13	0.39	4.39	-26.19	0.30	0.057
3318CB00128	0.97	0.08	0.00	1.03	3.40	-0.01	0.00	0.001
3318CB00131	1.81	0.14	0.14	1.50	7.10	-0.96	1.43	0.199
3318CB00131	1.81	0.14	0.14	1.50	7.10	-0.96	1.43	0.199
3318CB00133	5.25	0.11	0.42	1.93	1.28	-3.81	1.41	1.403
3318CB00138	0.89	0.01	-0.04	0.96	2.40	3.70	-0.14	-0.036
3318CB00138	0.57	0.35	0.14	0.57	1.77	-0.41	0.18	0.069
3318CB00141	0.55	0.21	-1.06	1.30	0.38	4.92	-2.15	-0.515
3318CB00163	0.30	0.10	0.12	0.30	6.77	-0.69	0.34	0.044
3318CB00167	0.40	0.22	0.08	0.41	0.83	-0.38	0.06	0.030
3318DA00102	11.10	0.04	0.15	5.69	0.72	-3.73	1.03	1.126
3318DA00102	10.12	0.04	0.13	5.61	0.74	-3.34	0.97	0.960
3318DA00132	197.37	0.02	0.39	6.88	0.65	-17.25	2.76	14.080
3318DA00132	3.60	0.04	0.25	2.73	0.69	-6.46	1.14	0.802
3318DA00132	3.52	0.07	0.22	2.80	0.66	-3.09	1.01	0.616
3318DA00132	2.54	0.00	0.17	2.15	0.61	-227.97	0.61	0.443

(ii) principal component analysis and
(iii) factor analysis,

additional information is gained. The following cations and anions were used in the statistical analysis: Ca, Mg, Na, K, HCO_3, SO_4, Cl and NO_3. The results of the respective methods are shown in figures 5.51 - 5.53.

If the results of the R-mode cluster analyses, the principal component analyses and factor analyses are studied, two major groups are depicted:

(i) NaCl-water (Factor 1).
(ii) Ca-HCO_3-water (Factor 2).

Contouring the factor scores of each factor of the R-mode factor analysis, the areal distribution of the two groups is depicted (see figures 5.54 - 5.55). They resemble the chloride- and alkalinity contours respectively.

It is interesting to note that both factors are very prominent in the Silwerstroom and Witzand areas. Factor 2, which is a Ca-HCO_3-factor, usually depicts recharge areas, thus fresh water is recharged in both the Silwerstroom and Witzand areas. However, factor 1 (NaCl-water) is also prominent in these areas. This indicates that dissolution of NaCl or mixing of the two ground-water types occurs on a local scale.

5.5.2 Regional analysis

Very few chemical analyses are entered in the ground-water data base for the remainder of the Orange Free State. However, in the Dewetsdorp area, 44 analyses have been recorded. By comparing the Dewetsdorp samples with those of the southern Orange Free State, two regions in the Orange Free State can be hydrochemically compared with each other. Figure 5.56 shows the distribution of the major cat- and anions on Piper- and Durov diagrams in the southern Orange Free State and the Dewetsdorp area respectively. In tables 5.26 to 5.30, the chemical analyses, ion balance and ionic ratios of the Dewetsdorp analyses are listed.

As can be seen from the diagrams, the chemical composition of the ground water of the two areas differs considerably, although both are underlain by Beaufort sediments and are intruded by dolerite. The sediments may, however, be part of different units in the Beaufort Group.

CLUSTER ANALYSIS

SIMILARITY MATRIX

	Ca 1	Mg 2	Na 3	K 4	HCO3 5	SO4 6	Cl 7	NO3 8
1	1.00	.00	-.08	.13	.66	.61	-.07	.03
2	.00	1.00	.71	.28	-.27	.22	.78	.17
3	-.08	.71	1.00	.39	-.44	.21	.98	.29
4	.13	.28	.39	1.00	-.19	.41	.35	-.01
5	.66	-.27	-.44	-.19	1.00	-.07	-.48	.01
6	.61	.22	.21	.41	-.07	1.00	.20	.07
7	-.07	.78	.98	.35	-.48	.20	1.00	.23
8	.03	.17	.29	-.01	.01	.07	.23	1.00

OBSERVATIONS COMBINED INTO CLUSTERS		SIMILARITY LEVEL OF CLUSTERING
1	5	.6629
3	7	.9790
2	3	.7459
4	6	.4060
2	4	.2674
2	8	.1226
1	2	-.0091

DENDROGRAM

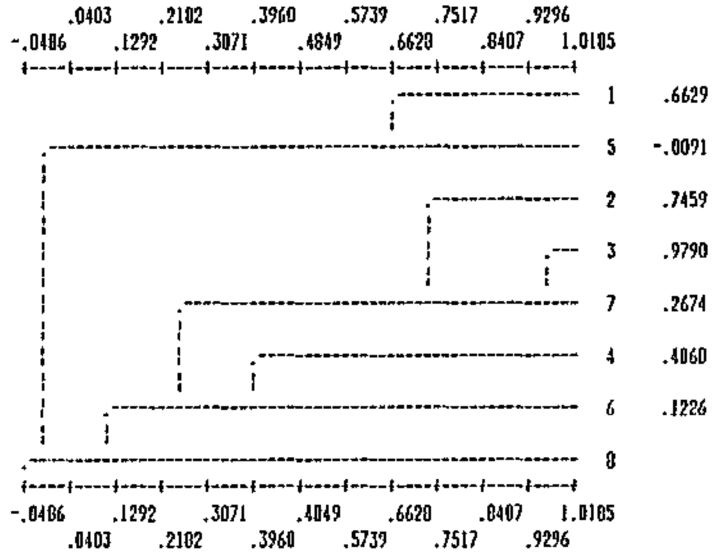


Figure 5.51. R-mode cluster analysis of the Atlantis water sample data.

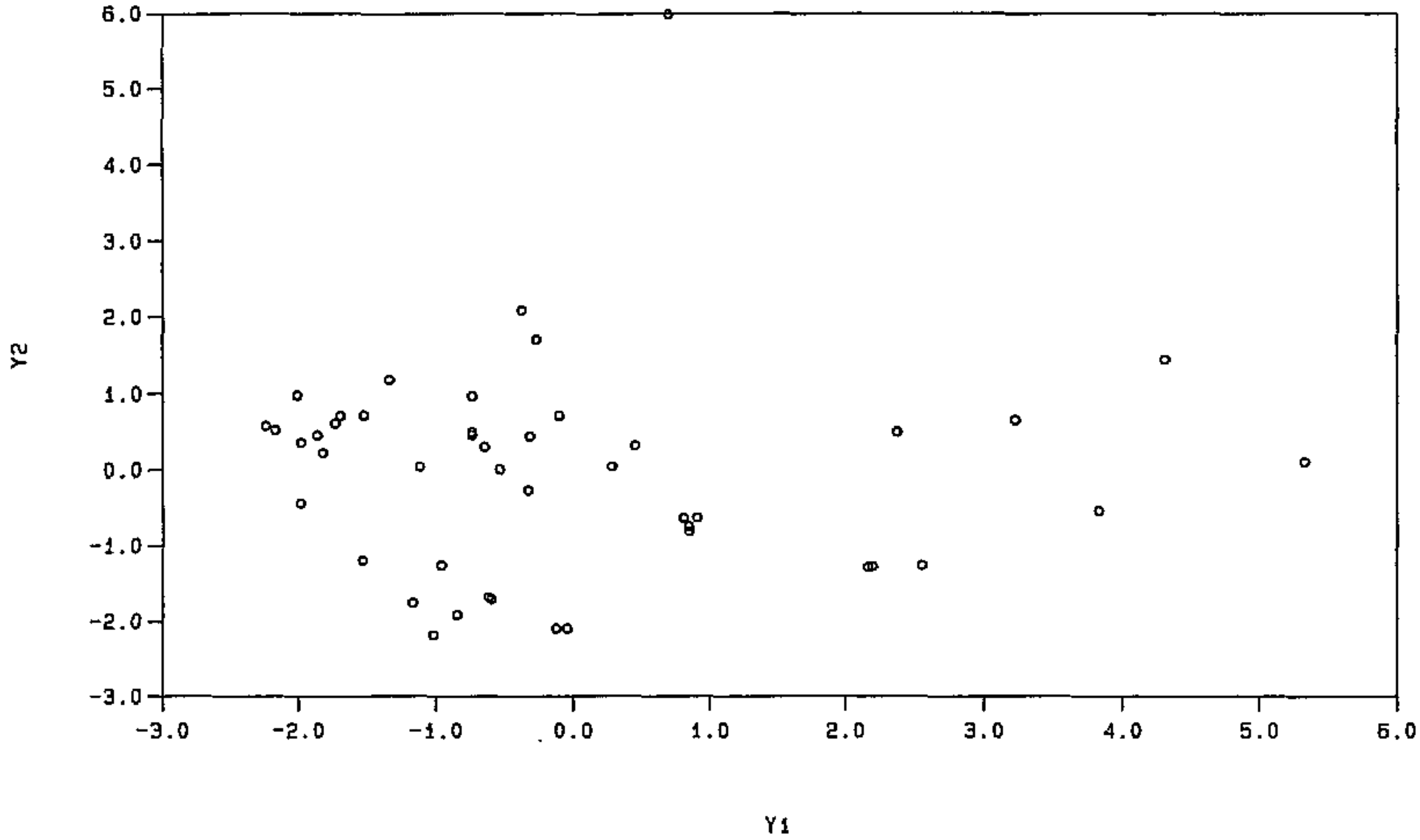


Figure 5.52. Principal component analysis of the Atlantis water sample data.

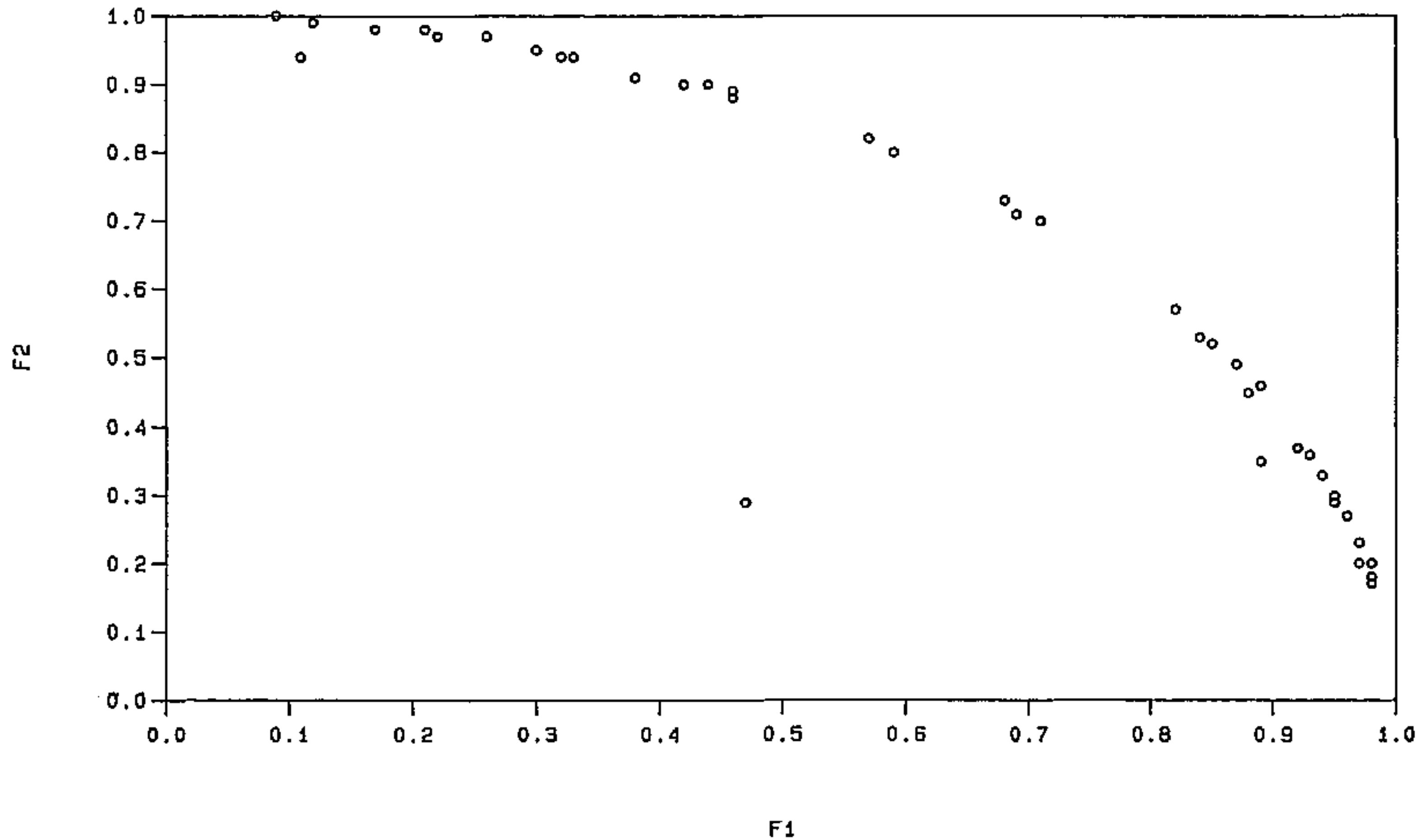


Figure 5.53. Q-mode rotated factor matrix - Atlantis water sample data.

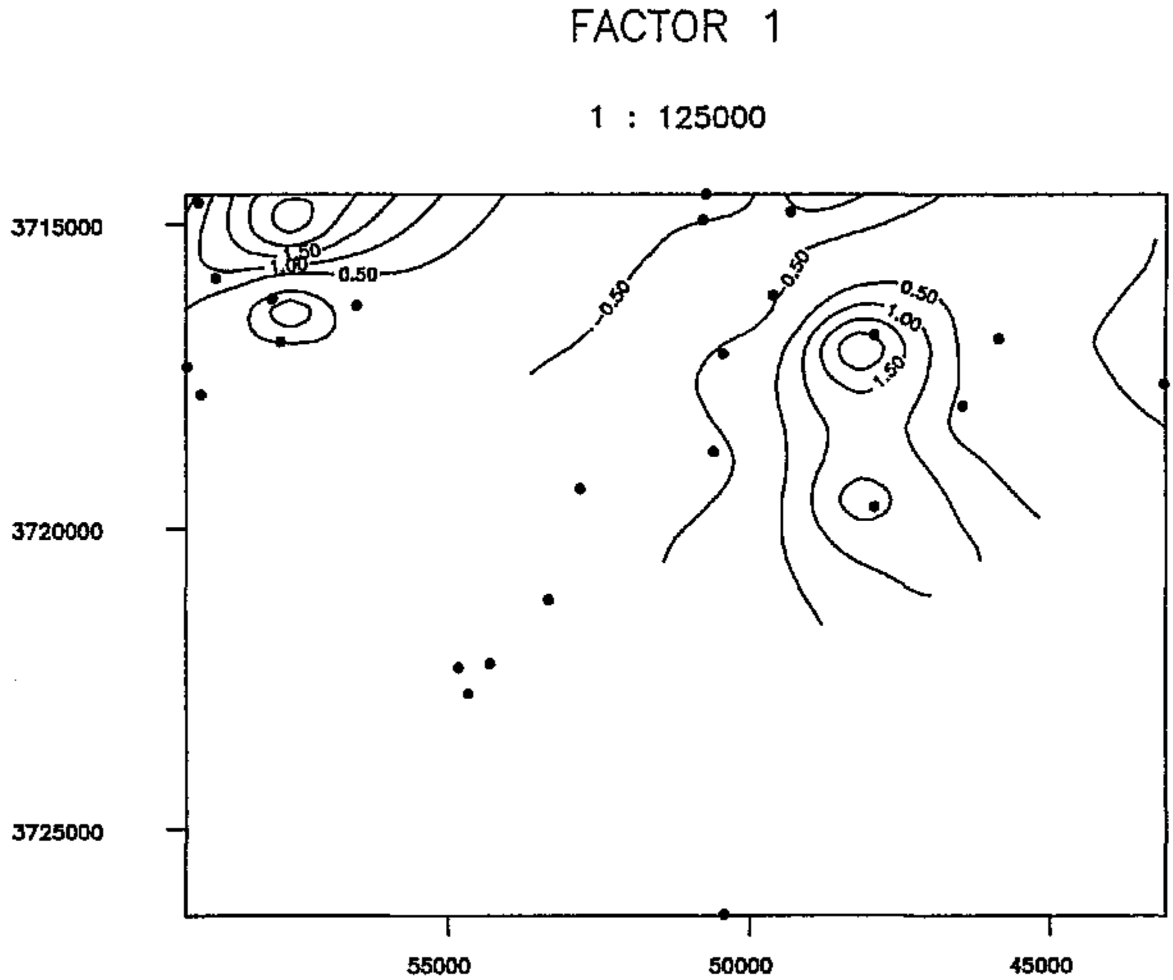


Figure 5.54. *Factor score contours - Factor 1.*

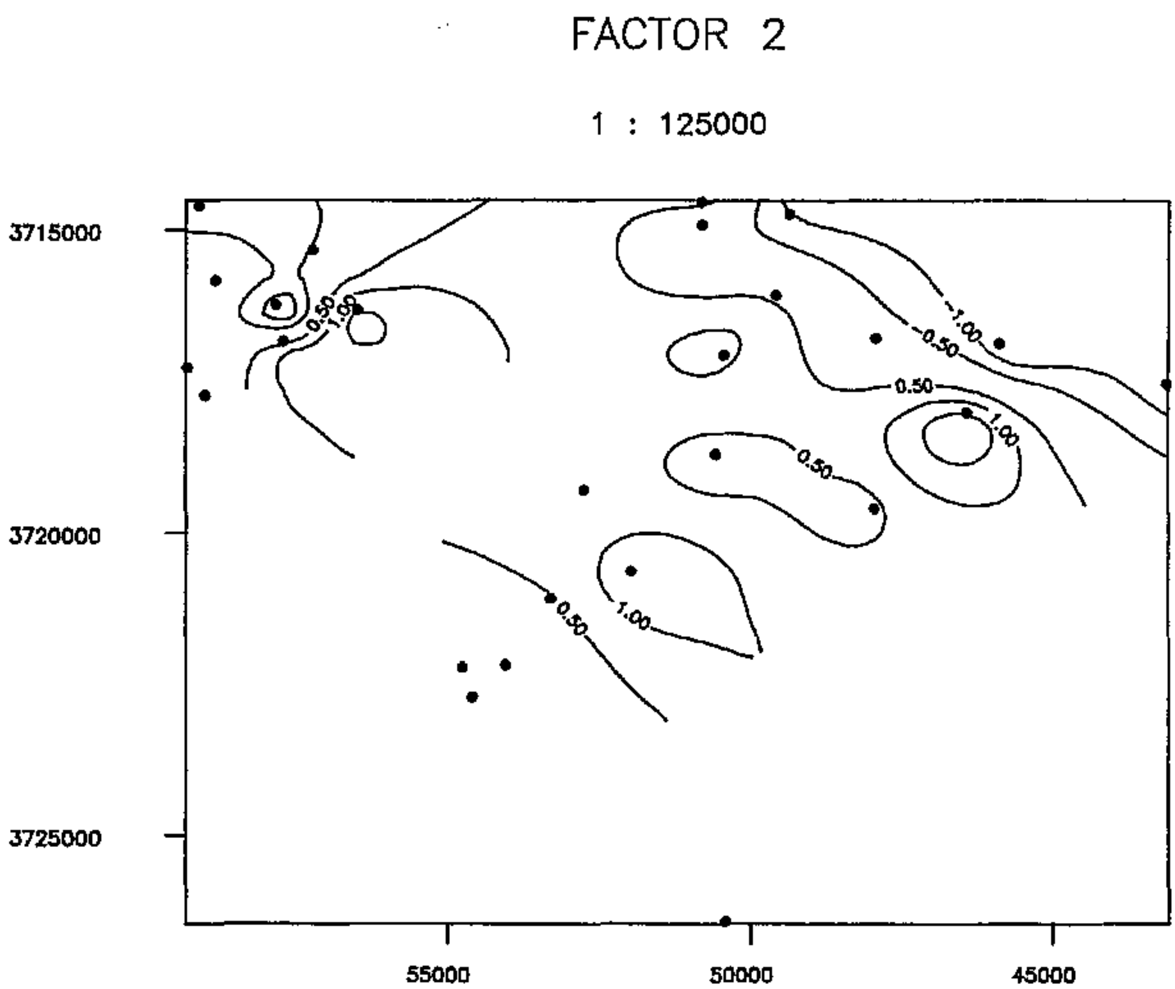


Figure 5.55. Factor score contours - Factor 2.

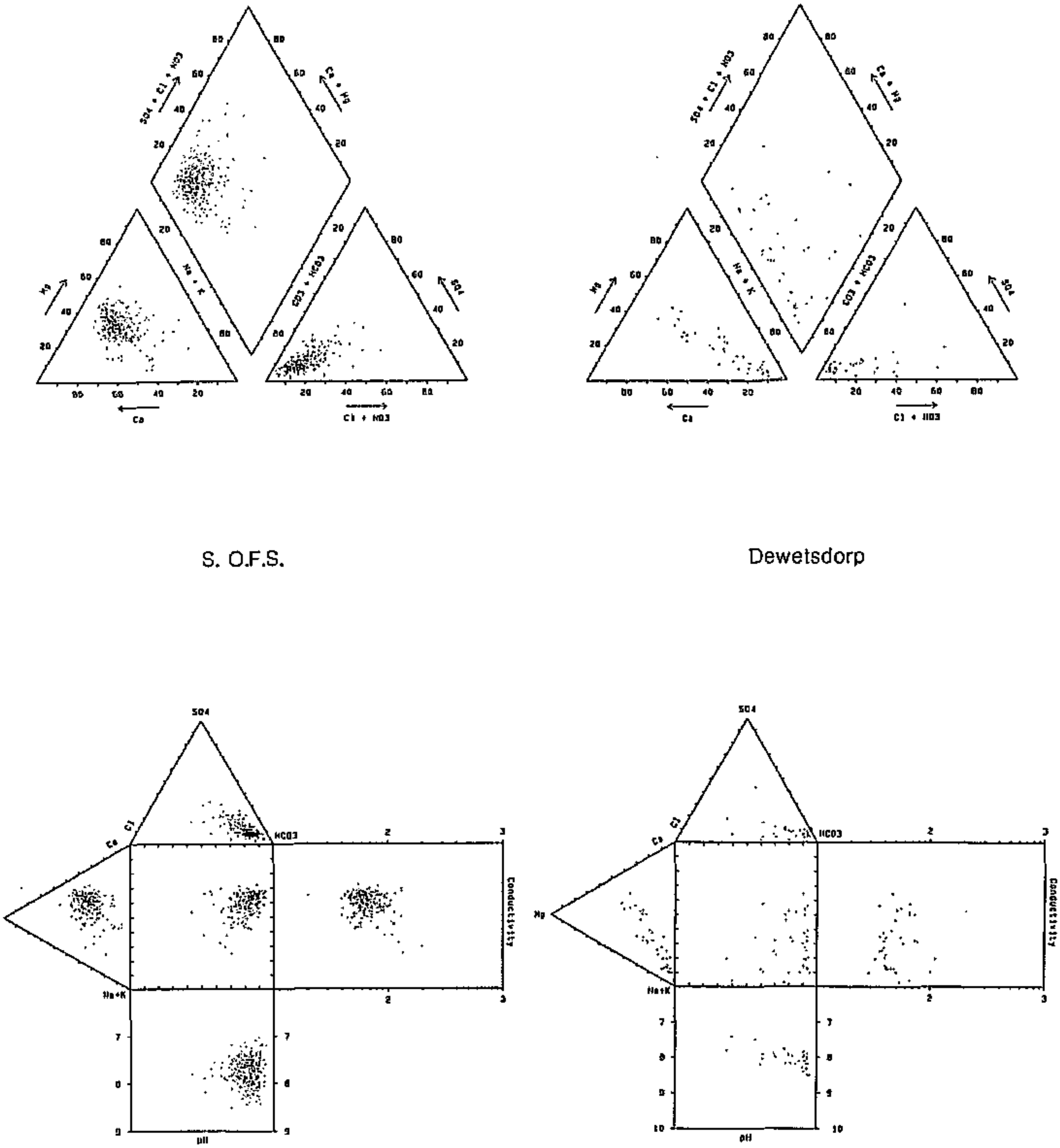


Figure 5.56. *Distribution of the chemical characteristics of the southern Orange Free State and Dewetsdorp ground waters.*

TABLE 5.26. ANALYSED CATIONS OF THE DEWETSDORP REGION.

ANALYSIS OF MACRO ELEMENTS (mg/l) AND ASSOCIATED PARAMETERS

CATIONS

Site Id.	Ca	Mg	Na	K	Si
2926CA00335	21.00	15.00	50.00	2.60	0.00
2926CA00352	57.70	24.60	60.70	6.60	17.30
2926CA00358	56.90	24.60	60.60	6.60	17.60
2926DA00001	17.00	3.40	66.70	4.80	4.00
2926DA00001	35.30	6.60	75.80	12.70	5.00
2926DA00002	25.30	10.10	53.00	3.70	5.10
2926DA00002	12.30	3.80	93.80	5.50	7.40
2926DA00002	12.00	3.20	105.80	15.50	7.90
2926DA00003	19.60	5.90	68.00	2.10	5.00
2926DA00005	18.50	15.00	60.40	1.90	8.80
2926DA00006	11.60	2.30	102.70	8.20	1.60
2926DA00007	14.10	7.60	72.80	3.30	4.60
2926DA00007	5.00	1.70	62.40	7.50	4.80
2926DA00008	16.30	7.50	67.20	2.40	9.00
2926DA00009	34.10	16.50	24.40	4.30	12.60
2926DA00011	12.70	3.20	65.80	6.80	6.30
2926DA00013	17.00	2.30	80.50	5.60	5.90
2926DA00015	6.00	1.00	72.00	1.50	5.20
2926DA00026	9.90	6.40	110.00	2.90	7.70
2926DA00026	7.80	7.30	106.20	2.10	7.90
2926DA00091	9.90	6.40	110.00	2.90	7.70
2926DA00091	7.80	7.30	106.20	2.10	7.90
2926DA00097	29.80	12.40	175.70	1.60	5.30
2926DA00106	4.60	0.90	133.20	1.30	0.10
2926DA00121	2.50	1.00	127.00	0.60	5.40
2926DA00125	39.50	17.70	117.00	1.60	9.10
2926DA00136	48.40	34.50	31.80	0.90	13.00
2926DA00156	16.30	6.20	201.80	1.30	4.30
2926DA00237	7.80	3.10	103.40	2.90	6.00
2926DA00239	34.20	16.70	99.50	1.00	9.50
2926DA00248	10.20	4.50	92.00	4.90	5.80
2926DA00250	10.60	4.10	100.80	5.00	6.30
2926DA00253	24.20	16.60	83.60	1.20	11.00
2926DA00254	17.00	16.40	49.30	1.70	9.50
2926DA00259	29.10	12.50	55.60	5.50	8.00
2926DA00263	12.00	5.10	73.90	3.50	6.70
2926DA00265	19.60	13.30	62.70	1.10	0.00
2926DA00267	41.50	16.10	40.00	1.90	10.60
2926DA00268	33.10	17.10	45.70	1.80	11.40
2926DA00269	37.00	27.50	32.80	1.30	12.80
2926DA00271	54.00	19.40	60.00	1.70	13.30
2926DA00272	43.10	25.70	36.80	0.60	14.70
2926DA00275	51.90	23.20	50.50	5.20	10.70
2926DA00276	44.40	21.00	17.80	2.20	14.50

TABLE 5.27. ANALYSED ANIONS OF THE DEWETSDORP REGION.

ANIONS
======

Site Id.	CO3	HCO3	SO4	Cl	NO3	F
2926CA00335	0.00	215.20	18.00	22.00	3.40	1.20
2926CA00352	0.00	327.60	42.10	44.50	0.90	0.80
2926CA00358	0.00	328.00	42.70	43.90	0.90	0.80
2926DA00001	0.00	150.90	12.10	36.40	0.00	2.80
2926DA00001	0.00	124.20	128.40	47.10	0.00	2.20
2926DA00002	0.00	209.40	19.80	6.20	0.00	0.70
2926DA00002	0.00	264.50	1.40	16.10	0.10	2.70
2926DA00002	0.00	297.70	2.70	12.50	0.40	1.90
2926DA00003	0.00	212.10	20.60	4.90	0.20	1.20
2926DA00005	0.00	233.70	10.80	7.20	0.10	0.90
2926DA00006	0.00	161.90	4.60	63.20	0.10	8.20
2926DA00007	0.00	158.90	7.70	36.40	0.10	9.90
2926DA00007	0.00	132.40	3.20	18.20	0.20	7.10
2926DA00008	0.00	215.60	10.20	6.70	0.10	1.10
2926DA00009	0.00	199.10	12.20	4.40	0.20	0.30
2926DA00011	0.00	193.80	12.10	7.70	0.20	1.30
2926DA00013	0.00	239.90	11.30	5.30	0.10	1.40
2926DA00015	0.00	176.70	7.10	3.10	0.10	1.10
2926DA00026	0.00	294.10	19.20	5.80	0.20	0.80
2926DA00026	0.00	278.20	15.20	6.20	0.10	1.10
2926DA00091	0.00	294.10	19.20	5.80	0.20	0.80
2926DA00091	0.00	278.20	15.20	6.20	0.10	1.10
2926DA00097	0.00	148.90	83.00	164.30	22.80	0.60
2926DA00106	0.00	125.40	18.30	113.20	2.70	0.70
2926DA00121	0.00	228.90	27.00	32.70	1.80	2.40
2926DA00125	0.00	391.00	35.60	36.50	1.50	0.60
2926DA00136	0.00	224.50	13.10	79.30	2.80	0.20
2926DA00156	0.00	312.00	73.20	98.30	9.20	0.30
2926DA00237	0.00	260.60	23.40	13.80	0.20	3.10
2926DA00239	0.00	284.50	43.60	61.10	0.50	0.90
2926DA00240	0.00	240.60	8.00	12.00	0.10	1.60
2926DA00250	0.00	262.10	15.50	6.70	0.00	1.00
2926DA00253	0.00	337.50	17.80	16.30	0.60	1.00
2926DA00254	0.00	211.20	11.10	5.80	0.10	0.60
2926DA00259	0.00	274.60	10.20	8.10	0.20	0.70
2926DA00263	0.00	204.20	16.70	4.40	0.10	1.10
2926DA00265	0.00	238.10	18.40	4.40	0.10	1.10
2926DA00267	0.00	235.70	23.40	26.90	2.00	0.60
2926DA00268	0.00	213.90	22.20	27.20	4.10	0.50
2926DA00269	0.00	294.50	11.20	11.50	1.30	0.30
2926DA00271	0.00	372.20	13.00	16.40	0.40	0.60
2926DA00272	0.00	323.20	12.60	14.00	0.70	0.40
2926DA00275	0.00	336.10	16.40	28.30	3.10	0.60
2926DA00276	0.00	251.50	15.20	9.40	3.00	0.20

TABLE 5.28. ANALYSED PHYSICAL PARAMETERS OF THE DEWETSDORP REGION.

PHYSICAL PARAMETERS

Site Id.	pH	Cond	Alk	TDS	TEMP
2926CA00335	8.17	41.00	172.20	356.40	0.00
2926CA00352	7.77	73.50	262.10	582.90	0.00
2926CA00358	7.76	73.00	262.40	582.70	0.00
2926DA00001	7.90	36.30	120.70	299.70	0.00
2926DA00001	7.50	55.20	99.40	437.30	0.00
2926DA00002	8.00	39.00	167.50	333.40	0.00
2926DA00002	8.00	39.70	211.60	407.70	0.00
2926DA00002	8.05	40.90	238.20	460.00	0.00
2926DA00003	8.09	39.60	169.70	339.60	0.00
2926DA00005	8.30	40.20	187.00	357.30	0.00
2926DA00006	8.20	40.50	179.50	364.40	0.00
2926DA00007	8.00	40.90	127.10	315.50	0.00
2926DA00007	7.90	28.50	105.90	242.70	0.00
2926DA00008	8.20	38.10	172.50	336.10	0.00
2926DA00009	8.00	35.60	159.30	308.10	0.00
2926DA00011	7.85	33.10	155.00	309.90	0.00
2926DA00013	8.20	40.00	191.90	369.30	0.00
2926DA00015	8.50	29.00	141.40	273.90	0.00
2926DA00026	8.34	46.00	235.30	457.00	0.00
2926DA00026	8.50	42.80	222.60	432.10	0.00
2926DA00091	8.34	46.00	235.30	457.00	0.00
2926DA00091	8.50	42.80	222.60	432.10	0.00
2926DA00097	7.80	112.40	119.10	645.90	0.00
2926DA00106	7.41	66.40	100.30	400.40	0.00
2926DA00121	7.75	53.90	183.10	426.30	0.00
2926DA00125	7.97	78.00	312.80	650.10	0.00
2926DA00136	8.03	66.80	179.60	452.60	0.00
2926DA00156	7.91	101.10	249.60	723.70	0.00
2926DA00237	8.11	50.30	208.50	424.30	0.00
2926DA00239	7.96	73.90	227.60	551.50	0.00
2926DA00248	8.40	41.30	192.50	379.70	0.00
2926DA00250	8.30	45.00	209.70	412.10	0.00
2926DA00253	8.29	58.10	270.00	509.80	0.00
2926DA00254	8.30	36.30	169.00	322.70	0.00
2926DA00259	7.90	46.30	219.70	404.50	0.00
2926DA00263	8.20	32.50	163.40	327.70	0.00
2926DA00265	8.30	39.00	190.50	366.80	0.00
2926DA00267	7.94	51.50	188.60	398.90	0.00
2926DA00268	8.17	49.80	171.10	377.00	0.00
2926DA00269	8.11	50.90	235.60	430.20	0.00
2926DA00271	7.94	65.60	297.80	551.80	0.00
2926DA00272	7.83	55.50	258.60	471.80	0.00
2926DA00275	8.18	66.20	268.90	526.00	0.00
2926DA00276	7.85	46.80	201.20	379.20	0.00

TABLE 5.29. IONIC BALANCE OF THE ANALYSED SAMPLES IN THE DEWETSDORP REGION.

PERCENTAGE CATION - ANION BALANCE ERROR

Site Id.	(%) Ion Balance Error	
2926CA00335	3.09	
2926CA00352	1.27	
2926CA00358	0.96	
2926DA00001	5.50	Unreliable Analysis
2926DA00001	-0.96	
2926DA00002	5.57	Unreliable Analysis
2926DA00002	3.27	
2926DA00002	5.07	Unreliable Analysis
2926DA00003	5.02	Unreliable Analysis
2926DA00005	6.30	Unreliable Analysis
2926DA00006	9.08	Unreliable Analysis
2926DA00007	9.36	Unreliable Analysis
2926DA00007	8.94	Unreliable Analysis
2926DA00008	5.72	Unreliable Analysis
2926DA00009	7.43	Unreliable Analysis
2926DA00011	3.74	
2926DA00013	4.04	
2926DA00015	6.26	Unreliable Analysis
2926DA00026	4.37	
2926DA00026	5.69	Unreliable Analysis
2926DA00091	4.37	
2926DA00091	5.69	Unreliable Analysis
2926DA00097	5.19	Unreliable Analysis
2926DA00106	3.80	
2926DA00121	2.42	
2926DA00125	2.09	
2926DA00136	3.25	
2926DA00156	2.88	
2926DA00237	0.62	
2926DA00239	0.86	
2926DA00240	5.87	Unreliable Analysis
2926DA00250	5.60	Unreliable Analysis
2926DA00253	-1.05	
2926DA00254	6.39	Unreliable Analysis
2926DA00259	0.94	
2926DA00263	6.16	Unreliable Analysis
2926DA00265	4.50	
2926DA00267	0.39	
2926DA00268	2.92	
2926DA00269	1.48	
2926DA00271	0.62	
2926DA00272	-0.72	
2926DA00275	0.94	
2926DA00276	0.22	

TABLE 5.30. IONIC RATIOS OF THE DEWETSDORP REGION.

Site Id.	Cl / HCO3	SO4 / Cl	Cl-Na / Cl	Na+Cl+SO4 / Ca+Mg+HCO3	Ca / Mg	Na-Cl / SO4	Cl-Na / Mg	Cl-Na / SO4+HCO3+NO3
2926CA00335	0.18	0.60	-2.06	0.60	0.85	4.75	-1.52	-0.450
2926CA00352	0.23	0.70	-0.90	0.46	1.42	1.30	-0.62	-0.102
2926CA00350	0.23	0.72	-0.94	0.46	1.40	1.31	-0.63	-0.185
2926DA00001	0.42	0.24	-1.52	1.10	3.18	6.21	-6.26	-0.574
2926DA00001	0.65	2.01	-1.50	1.68	3.24	0.75	-3.64	-0.423
2926DA00002	0.05	2.35	-10.40	0.49	1.52	4.43	-2.53	-0.475
2926DA00002	0.11	0.06	-7.22	0.82	1.96	112.77	-11.46	-0.753
2926DA00002	0.07	0.16	-10.43	0.79	2.27	65.62	-15.95	-0.746
2926DA00003	0.04	3.09	-17.85	0.67	2.02	5.77	-5.77	-0.633
2926DA00005	0.05	1.10	-9.79	0.47	0.75	8.87	-1.93	-0.492
2926DA00006	0.67	0.05	-1.26	1.77	3.06	23.47	-13.15	-0.817
2926DA00007	0.40	0.16	-1.80	1.07	1.13	11.55	-3.26	-0.669
2926DA00007	0.24	0.13	-3.76	1.20	1.78	29.04	-15.33	-0.864
2926DA00008	0.05	1.12	-12.01	0.60	1.32	10.72	-4.37	-0.607
2926DA00009	0.04	2.04	-5.67	0.21	1.25	2.70	-0.66	-0.200
2926DA00011	0.07	1.16	-10.75	0.76	2.41	9.29	-9.95	-0.682
2926DA00013	0.04	1.57	-19.96	0.73	4.40	12.72	-17.62	-0.718
2926DA00015	0.03	1.69	-29.49	0.90	3.64	17.50	-36.82	-0.849
2926DA00026	0.03	2.44	-24.72	0.84	0.94	10.15	-0.74	-0.777
2926DA00026	0.04	1.00	-21.64	0.82	0.65	12.00	-7.35	-0.778
2926DA00091	0.03	2.44	-24.72	0.84	0.94	10.15	-0.74	-0.777
2926DA00091	0.04	1.00	-21.64	0.82	0.65	12.00	-7.35	-0.778
2926DA00097	1.90	0.37	-0.46	2.82	1.46	1.24	-2.36	-0.471
2926DA00106	1.56	0.12	-0.69	3.84	3.10	5.77	-31.86	-0.887
2926DA00121	0.25	0.61	-4.41	1.64	1.52	7.25	-52.50	-0.939
2926DA00125	0.16	0.72	-3.57	0.68	1.35	4.98	-2.73	-0.514
2926DA00136	0.61	0.12	0.46	0.45	0.85	-3.77	0.41	0.250
2926DA00156	0.54	0.55	-1.94	1.97	1.59	3.54	-11.36	-0.795
2926DA00237	0.09	1.25	-10.28	1.08	1.53	8.23	-16.07	-0.843
2926DA00239	0.37	0.53	-1.37	0.89	1.24	2.61	-1.02	-0.424
2926DA00240	0.09	0.49	-9.25	0.85	1.37	18.06	-9.76	-0.764
2926DA00250	0.04	1.70	-19.06	0.86	1.57	11.20	-12.35	-0.782
2926DA00253	0.08	0.80	-6.64	0.54	0.80	8.26	-2.31	-0.510
2926DA00254	0.05	1.41	-9.76	0.41	0.63	6.93	-1.44	-0.434
2926DA00259	0.05	0.93	-0.87	0.40	1.41	9.50	-2.11	-0.431
2926DA00263	0.04	2.79	-20.96	0.76	1.43	7.50	-7.31	-0.706
2926DA00265	0.03	3.08	-10.15	0.51	0.89	5.90	-2.36	-0.527
2926DA00267	0.20	0.64	-1.13	0.41	1.56	1.76	-0.70	-0.196
2926DA00268	0.22	0.60	-1.27	0.48	1.17	2.11	-0.79	-0.242
2926DA00269	0.07	0.72	-2.95	0.22	0.82	4.11	-0.47	-0.189
2926DA00271	0.08	0.62	-4.22	0.31	1.69	6.02	-1.32	-0.307
2926DA00272	0.07	0.66	-2.78	0.23	1.02	4.19	-0.56	-0.198
2926DA00275	0.15	0.43	-1.56	0.33	1.36	3.66	-0.70	-0.212
2926DA00276	0.06	1.19	-1.62	0.17	1.20	1.36	-0.28	-0.096

Whereas the Trompsburg water samples are predominantly Ca/Mg-HCO_3-waters which were recently recharged, the samples from Dewetsdorp are mainly of the $NaHCO_3$-type, classified by Johnson (1975) as dynamic waters.

The two areas can also be compared by studying the ionic ratios of the respective analyses (tables 5.23 and 5.25). Three of the ratios indicate the difference in the ground-water composition. These ratios are:

(i) (Cl - Na)/Cl (base exchange index).
(ii) (Na - Cl)/SO4.
(iii) (Cl - Na)/Mg.

(Cl - Na)/Cl-ratio

This ratio is generally known as the base exchange index. The relatively high negative values of this ratio in the Dewetsdorp area indicate that base exchange is taking place; with increasing sodium percentage, the ground water has moved away from the recharge area. As Dewetsdorp lies on a watershed and the samples were taken within comparatively short distance from the watershed, this implies that the aquifers in the vicinity are confined or semi-confined and that noticeable ion exchange takes place over very short distances.

The much lower negative ratios of the Trompsburg region indicate that little or no base exchange occurs (see figure 5.56).

(Na - Cl)/SO4-ratio

Water is classified as $NaSO_4$-water (relative old water), if the (Na - Cl)/SO4-ratio is less than one (Matthess 1982). In the Dewetsdorp area, this ratio is, in the majority of the cases, considerably greater than one, i.e. the water is a $NaHCO_3$-water, displaying the characteristics of dynamic waters. In the Trompsburg region, this ratio tends to be greater than 1, although a few slightly smaller values are observed.

(Cl - Na)/Mg-ratio

This ratio should be used in conjunction with others, e.g. with the Cl/HCO_3- and the (Na+Cl+SO_4)/(Ca+Mg+HCO_3)-ratios. It is then found that the ground water in both regions reveals a relatively young age. However, this ratio shows the difference in the water composition of the two areas, because the sodium-nature of the Dewetsdorp waters is highlighted by the higher ratios.

To summarize, the processing facilities provided for evaluation of the hydrochemical data illustrate the chemical characteristics of waters from different regions, classify them and make geohydrological deductions regarding aquifers or the occurrence of certain waters. With the aid of the statistical package, it was occasionally possible to further differentiate between waters which could not easily be discerned on the Piper- or Durov diagrams or to provide other additional geohydrochemical information.

5.6 DEVELOPMENT AND MANAGEMENT OF AQUIFERS

Orange Free State

The purpose of the evaluation of geohydrological data is to efficiently and optimally develop and manage the ground-water resources of an area. To be able to do this, sufficient reliable long-term data are needed. Although almost 10 000 boreholes of the Orange Free State are accommodated in the Ground Water Data Base, about two-thirds of the boreholes contain insufficient or partly unreliable data. This means that, at this stage, the stated purpose cannot be fully achieved.

The data are, however, well-suited for the planning and development of new ground-water schemes in specific areas. If previous work was done in the area and the data are entered, the information is now easily accessible and can be processed in the required form. This gives the persons involved in the planning and development of a new project or extensions to existing projects, the necessary background in the geohydrology of the area under investigation.

Atlantis area

The data processing facilities, provided by the Ground-water Data Base, can assist with the development and management of the Atlantis ground-water resources. Apart from the availability of sufficient and reliable data, the fairly intimate knowledge of geohydrological principles, in general, and of the aquifer in particular, are a prerequisite for such a task.

For the Atlantis aquifer, most of the necessary data are available. Aquifer evaluations can be done. For aquifer management purposes, the monitoring network should be extended, as already indicated. Once existing data are supplemented by data around the present observation network and operational data are included, recommendations can be made regarding optimal pumping rates. With data updated on a regular basis, pumping rates may be adjusted according to variations in water-level gradients and possible changes in the chemical composition of the water. In combination with the rainfall data base, all the data necessary for recharge evaluations by means of modelling, or otherwise, can be supplied by the data base in machine readable format.

The existing data are well-suited for both scientific investigations and the planning of possible further extensions of the present well field.

5.7 DISCUSSION OF RESULTS

Data from the Orange Free State and the Atlantis area were processed to demonstrate the functioning of the ground-water data base and the use of its data processing facilities. The following deductions can be made:

(i) GEOLOGICAL DATA: The data are presented in a form, suitable for reports. Insufficient detail in both the geology and aquifer data sets, however, prevents differentiation between regions with similar geology. Yet, the information contained in the data base, provides an easy overview of the basic geology of any area and the availability of data.

(ii) AQUIFER DATA: Depending on the objective of the investigation carried out with the aid of the data base, there are distinct data requirements. Satisfactory results were obtained in demonstrating borehole statistics on water levels, yields (which are log-normal distributed), water use and pumping tests from Atlantis. At the scale the analysis was done, there was no correlation between borehole yield and geology, none between borehole yield on the one side and the depth water was struck or the depth of the hole on the other. However, an indication of lower yields along the watershed was observed.

(iii) BOREHOLE CONSTRUCTION AND INSTALLATION DATA: Statistical processing provides, amongst others, all information on dimension and construction of boreholes which is required for planning and tender purposes.

(iv) CHEMICAL DATA: By means of diagrams, contours and statistics, the programs provided, assist in the classification of aquifers (e.g. confined or unconfined), help to identify recharge areas or indicate mixed waters and delineate possible risk or pollution areas. They also assist in the general understanding of the geohydrochemical processes. With the aid of the statistical package, it was occasionally possible to further differentiate between waters, which could not easily be discerned on the Piper- or Durov diagrams or to provide other additional geohydrological information.

Although there is a shortage of high quality data in the data base (exact position of boreholes and shortage of construction data, sufficient long-term water-level measurements, extraction data, pumping test data), the available information also helps in the estimation of the aquifer potential and the design and optimization of monitoring networks.

CHAPTER 6

DISCUSSION

Now, that the ground-water data base is implemented, one must look whether, or to what degree, it meets all the objectives set at the beginning of the project. The discussion in this chapter will deal with the following aspects :

(i) National Data Base.
(ii) Data base users.
(iii) Data base applications.
(iv) Data processing package.

6.1 EVALUATION OF THE NEW DATA BASE

The discussion of the National Ground-water Data Base will be divided into five topics, namely:

(i) Structure.
(ii) Parameters.
(iii) Data input.
(iv) Data output.
(v) Hardware requirements and communications.

Structure

The four data structures, flat file, relational structure, hierarchical structure (tree structure) and network structure, were described in chapter 2.2. The data in the Ground-water Data Base are accommodated in three levels, consisting of basic information, header and respective repeating groups. Such a structure is easily converted into a tree structure, which has the advantage of minimizing data redundancy. As ground-water data are dealt with as a one-to-many relation, the tree structure is a suitable model. The data also suit a relational structure. However, the use of such a model would lead to unnecessary data redundancy.

The tree structure has the distinctive disadvantage that the data have to be accessed through the hierarchy, which makes the access of certain data relatively inefficient. This situation was, however, rectified by combining a few network features with the features of the tree structure, which lead to more efficient data retrieval.

To summarize, the data base was structured in such a way that the user has, within the constraints of the mainframe computer, easy and efficient access to the data.

Parameters

The parameters present in the data base were included, after an evaluation of the parameters included in other existing ground-water data bases and thorough discussion with the Directorate Geohydrology of the Department of Water Affairs, to adapt the data base to suit especially Southern African conditions.

A data base depends on the type and the quality of data added. To promote entry of good quality data, certain data items were declared mandatory. In addition, as many logical data checks as practically possible, have been included in the data base programs (chapter 3 of the technical specifications), e.g. latitude and longitude are compared with the map number.

One shortcoming in the parameter list is that hydrochemical parameters are not included. They have been omitted, because of the existence of a separate hydrochemical data base. Provision has, however, been made that the hydrochemical data can be accessed from the ground-water data base. As ground-water recharge occurs preferably through the unsaturated zone, data from this zone can well be regarded as ground-water related data. Yet, such data were not accommodated in the data base, although, to a certain extent, these data can be entered under the Field measurement data set.

Data input

At present, data are entered interactively into the data base, via numerous terminals. To date, batch load facilities exist only for digitized water levels and field measurement data. The aim, however, is to generate a complete batch load facility, which will enhance the data input, especially for remote work stations which are not permanently connected to the mainframe computer.

Data output

Data can be extracted from the data base in three ways:

(i) Statistics of data.
(ii) Data extraction via printer file.
(iii) Data extraction via transmission file.

A certain degree of data selection was incorporated in the data reporting and extraction facilities:

(i) Site.
(ii) Type.
(iii) Completion or Digitizing Date.
(iv) Data Set.

Due to the processing speed advantage a mainframe computer has over a micro computer, the data set selection (iv) could have been designed in such a way that conditional extraction of data was possible (e.g. all holes with diameter ≥ 200 mm and yield > 10 m^3/h, etc.). Such a facility would limit data selection and sorting on the micro computer. However, inclusion of such a user-friendly facility in the data base would have required very extended programming, because the type of requests a geohydrologist or other user might have, is nearly unlimited. For the experienced mainframe users, the ERGO facility provides the possibility to set conditions for data extraction.

One should, at this stage, also consider that the data base with its data processing package was designed around project orientated investigations, where various data are required for different aspects of the task. In cases where the same data are repeatedly used, there is a saving on access time, data transmission time and transmission cost. For instance, when the same request is made with varied conditions or when pumping test evaluation, water-level graphs and/or water-level contours are required.

Hardware requirements and communications

The data base is implemented on a mainframe computer. In order to access the computer, either a terminal or a micro computer which emulates the mainframe terminal, can be used. The fact that a micro computer can be used as a terminal is a major advantage, because of its data processing and storage abilities.

Remote work stations are connected to the mainframe (host computer) by a transparent network via X.25 protocol. If the network facility is used in conjunction with a micro computer, file transfer between the mainframe and micro computer is fast and easy. In the present set-up, this is a major advantage, because most of the geohydrological data processing are done on the micro computer.

6.2 DATA BASE USERS

When designing a data base, the potential users have to be borne in mind in order to satisfy their requirements. Users which may benefit from the ground-water data base are:

(i) Geohydrologists.

(ii) Geohydrological technicians.
(iii) Engineers.
(iv) Drilling contractors and suppliers of pumping equipment.
(v) Water supply organizations.
(vi) Researchers.

Geohydrologists

The tasks of the geohydrologist comprise geohydrological surveys, carried out to determine the ground-water inventory like water levels, yields of boreholes, aquifer parameters and quality, to explore and develop the ground-water resources and possibly manage the operational exploitation of well fields or aquifers, to investigate pollution aspects or other topics like effluent injection and heat exchange.

Geohydrological and geophysical field work, geohydrological control of drilling operations, data processing, literature study, report writing, recharge- or ground-water balance calculations and modelling may be carried out, subsequent to the survey or be done independently.

In the public service, the geohydrologist may also have ground-water related administrative tasks, which include topics as advice in drilling matters, issue of permits for drilling or for groundwater abstraction and control of such operations.

For any new project, the geohydrologist can now access the data base, extract and evaluate all the existing data at the touch of a button. This data availability and the available processing facilities assist him in the compilation of efficient working programs for the proposed investigations. Additionally, it gives him the tool to identify possible target areas for new ground-water development schemes and to evaluate whether the available data are sufficient. This knowledge will influence his investigation and allow him to tailor his working program (best aquifers, best quality, best exploration methods, compilation of drilling tender and drilling supervision requirements).

Once all the data of an investigation are entered into the data base, reporting can be done, using the facilities of the processing package. The reports enable the geohydrologist to delineate areas where additional work may or should be done. It also contains data for the construction of aquifer management models which may assist in the optimal utilization of the ground-water resources.

Geohydrological technicians

Technicians can either assist the hydrogeologist in geohydrological investigations or do borehole selections for e.g. the farming community and other small- or large scale users. The advantages of the data base for borehole site selection are discussed here.

Borehole selection requires detailed regional and local geohydrological knowledge of the area to be investigated. This means that the technician sent to select borehole sites in a particular region, has first to collect as much basic and detail information as possible, regarding existing boreholes in the area.

Information such as yields, depths, geology, aquifer, quality and possibly geophysics must be retrieved and evaluated. This may help him to identify specific ground-water related structures or to depict potential target areas, on which he can concentrate his activities. The information in the data base also helps to reduce duplication of geophysical and geohydrological observations.

Engineers

Engineers usually fill the medium and top-level management positions in an organization such as the Department of Water Affairs. It is in their hands to develop and manage aquifers or well fields, so that a long-term reliable supply of the necessary water to the communities can be guaranteed. In various parts of the country, e.g. at Atlantis, De Aar or Grootfontein (Western Transvaal), whole communities depend exclusively on the ground-water resources of the aquifers in the region.

For the development and management of ground-water schemes, the manager must have access to a variety of data such as:

(i) Water demand.

(ii) Ground-water potential (aquifer parameters, recharge, losses, present use, yields, quality, water levels and water-level changes).

(iii) Development cost (geohydrological investigation, drilling and development, installation, civil works).

(iv) Operational cost.

The National Ground-water Data Base accommodates some of the parameters under (ii). The geohydrologist can supply most of the remaining information required there by processing and evaluating the water-level-, discharge rate-, meter reading-, pumping-test- and possibly geology- and aquifer data sets. He can also indicate the cost of the geohydrological investigation based on various information contained in the data base.and supply information regarding drilling and development cost (iii).

It has already been mentioned that data required for aquifer management- and ground-water flow models, which can support the manager in long-term water resource management of the aquifer, are accommodated in the data base.

Mechanical engineers, designing borehole installations, can also use the ground-water data base in the design process. The following information:

- (i) borehole diameter and depth,
- (ii) borehole construction,
- (iii) borehole yield,
- (iv) water levels,
- (v) pumping test results and
- (vi) installations of existing boreholes

is stored in the data base and can be used for the most suitable borehole installation.

Engineers in charge of water supply schemes may store operational data like pumping- and rest-water levels, production figures, power consumption or quality data (e.g. electrical conductivity) in the data base.

Drilling contractors and Suppliers of pumping equipment

Information such as penetration rates, geology, borehole construction and installation are accommodated in the data base. If a drilling contractor tenders for the drilling of boreholes in an area, certain information like:

- (i) formations encountered,
- (ii) borehole depths and diameter,
- (iii) depths at which water was struck,
- (iv) type of casing used,
- (v) length of casing,
- (vi) drilling costs,
- (vii) drilling method used and
- (viii) penetration rates

is of importance to him, because it supplies him with the data he can base his tender prices on.

Data, contained in the data base, provide the supplier of pumping equipment with information regarding the type and size of equipment used and date installed. These data are useful for stock-keeping and for market research purposes.

Water supply organizations

Bodies such as municipalities, water boards and divisional councils, exploiting ground-water resources, resort under this group. These institutions are seldom involved in the actual exploration and development of an aquifer. However, the operation of ground-water supply schemes is their responsibility. Thus yields, water levels and water quality, as well as monthly abstractions and e.g. power meter readings, are required by them. Installation data, like depth of pump intake, type of installation or installation date, are also important for a smooth operation of the scheme and may help to prevent pumps from running dry or may be used for estimation of required spares or stand-by equipment.

Good aquifer management requires periodic control of ground-water levels, pumping rates and the ground-water quality in production and observation boreholes. Analysis of these data may prevent over-exploitation and help to detect possible faults in the system.

Scientists (researchers)

Ground-water research is mainly done at universities, research institutes or other research organizations, such as the CSIR. Other institutions, like the Department of Water Affairs or the AEC, are also involved in research. To conduct thorough geohydrological research, a great variety of geohydrological data is needed and most of these data are accommodated in the National Ground-water Data Base or, as in the case of hydrochemical data, can be accessed via it.

The access to a great number of geohydrological data is facilitated by the ground-water data base. The data base provides the scientist access to these data which he might need to conduct his studies. Research topics or data required might be:

(i) Aquifer potential (determination of aquifer parameters, recharge, etc.).

(ii) Geohydrochemical studies (regarding water quality, pollution, etc.).

(iii) Borehole geophysics.

(iv) Borehole development and construction.

(v) Aquifer modelling.

Data about the unsaturated zone are not separately stored in the ground-water data base.They can, however, be entered under field measurement. Practically every geohydrological parameter that might be needed by the researcher can be accommodated in the data base.

From the examples given, it can be deduced that identified data base user groups can find most, if not all, geohydrological and related data they need, in the data base.

6.3 DATA BASE APPLICATION ASPECTS

6.3.1 Aquifer management

If we do not want to deplete the aquifers in Southern Africa, the abstraction from the aquifer and the losses due to natural discharge and evapotranspiration may not exceed its recharge. Even, if an aquifer is to be mined, it is important to know how long the resource will last. In order to maintain the balance between recharge and discharge or to control the lowering of the water table of the aquifer, constant surveillance of a number of geohydrological parameters is necessary. A facility to store these data and retrieve different combinations of these variables for evaluation purposes is therefore important for good aquifer management. The National Ground-water Data Base contains all the necessary information, provided the relevant data were captured and entered into the data base.

6.3.2 Planning aspects

Occasionally, existing ground-water resources of a region are over-utilized, however, more often the ground-water potential of a new region is to be explored.

The ground-water data base can play an important role by providing geohydrological data about the area. These data can be evaluated and potential ground-water target areas be delineated. A geohydrological investigation can also be properly planned for the specific area. Based on previous experience, the most promising exploration methods can be chosen and duplication of surveys and other work already done, can be eliminated.

Supplementing the geohydrological investigation of the area, a comprehensive drilling program, including selection of the best drilling methods, optimal borehole construction and development, can be planned. A good estimate of the total costs of such a drilling program can be made. If these costs are combined with the estimation of the costs of the geohydrological investigation, the hydrogeologist can compile a realistic budget for the project.

6.3.3 Research aspects

With the aid of the data base, the scientist can either determine which area to use because of data availibility or assess where, what type of data and how many additional data are required for his research. Using information in the data base, the cost of geohydrological data generation can also be better estimated. The available data may also indicate where and what type of research should be done.

6.3.4 Manpower aspects

Although additional personnel is required to capture and enter historical data into the ground-water data base, more efficient deployment of manpower in the field and in office is possible. The cost of entering historical data should be set off against savings at a later stage when these data are used.

Field

Geohydrological investigations can be planned better with the aid of the ground-water data base, because shortcomings in the existing data can be identified and the tasks of the project adapted accordingly. Optimal deployment of field personnel for the duration of the investigation can therefore be achieved and unnecessary field work be minimized. This will result in increased productivity and cost-efficiency.

Office

In order to constantly update the ground-water data base with new geohydrological data, additional job opportunities were created for data typists. Geohydrologists and technicians are, on the other hand, to a large degree, relieved from compiling tables and drawing graphs, profiles and maps which may now be done by micro computer. Typing of tables and draughting will therefore also be reduced.

People associated with tasks, like data evaluation and data presentation, can become more efficient, because of the data processing and draughting facilities of the National Ground-water Data Base. Professions, such as draughtsmen and typists, can support a greater number of geohydrologists and technicians.

6.3.5 Cost aspects

Initially, the development and implementation of the ground-water data base was an expensive venture. However, as the amount of data stored and the number of data base users increase, the data base becomes more cost-efficient. An important aspect is the achieved increased efficiency and productivity of the personnel associated with a project. This has a cost-saving effect on any project during which noticeable use is made of the ground-water data base. The success rate of borehole site selection is also likely to rise, if more complete data regarding previous work and improved documentation are readily available.

Another advantage of the electronic storage and network facilities provided by the data base, which has a bearing on the cost of an investigation, is the accessibility of the data over long distances. Currently, it is still a rather time-consuming task to obtain data which are kept in files, stored at the head office or some other distant location. Users, who want data of an area, have to contact the respective organization or division and request the data. It then takes some time before the data are located, possibly duplicated and send to the user.

6.3.6 Security aspects

Security refers to various aspects:

(i) Security of the data against loss.
(ii) Access to the data-base.
(iii) Restricted access to qualified data.

Data security against loss is the responsibility of the manufacturer of the mainframe and the organization operating it. This aspect is not dealt with here.

Access to the data base: Unauthorized access to the mainframe is also outside the scope of this report. Provision is made within the data base programs by means of passwords, user- and reporting institution codes to control the access to the data base. Depending on the codes:

(i) inquiries may be made,
(ii) data may be added and updated or
(ii) code lists may be accesssed and changed.

The decision who may access the data base rests with the Department of Water Affairs as administrator of the data base and, at present, not only custodian, but owner of practically all the data .

To create the best possible data base facility, all persons and organizations dealing with any ground-water aspect should be accommodated. However, one problem has still to be solved or decided upon, namely who owns the data. In any data base, environment shared by multiple users, this is essential, because (some of) the data have an economical value and must therefore be protected against unauthorized access. Otherwise, the data will not be entered in the data base and be completely lost to the community. The Department of Water Affairs, as administrator of most of the data bases falling under HIS, has developed the general guidelines, regarding data access which applies to all data bases.

Generally, each data set or data item in a data base can be separately controlled, so that only authorized users can read, enter, or update and delete data. A good data base management system has the capability to enforce strict security control in this regard.

In many cases, one institution may enter certain data. It creates the records updates them and is responsible for their validity and accuracy. Although the data are shared, no other organization or individual is permitted to alter them.

According to Martin (1983), the following essentials should be taken into account when implementing a security system for the data base:

(i) Data base users should be identifiable, i.e. users must be positively identifiable before they may use the data base.

(ii) All their actions should be authorized and monitored, so that the network management may be able to check whether actions are authorized and monitored.

(iii) Data, hardware and software should be protected from fire, theft or other forms of damage.

(iv) Data should be reconstructable, auditable and tamper-proof, i.e. it should be easy to regenerate data, if something goes wrong and it should be regularly audited to prevent crime.

(v) Transmission should be fail-proof and private., i.e. it should be fail-safe when errors occur and the data should be protected to prevent eaves dropping.

(vi) Vital computer centres should be catastrophe-proof and replicated, i.e. they should be duplicated and data saved on another computer.

Thus, to accommodate all possible users in the data base, effective security control must be implemented. This will also encourage hesitating organizations to participate in the ground-water data base and add to its value for the whole community.

6.4 EVALUATION OF THE DATA PROCESSING PACKAGE

One of the important features of the National Ground-water Data Base is the comprehensive data processing package. This package provides data out-put in the form of tables, graphics and in processed statistical form.

The purpose of this discussion is to show potential users the applicability of the programs and also point out some of the traps the data processing techniques pose, when uncritically applied to ground-water data.

6.4.1 Tables

A series of tables has been provided for out-put of all the data sets in the ground-water data base. In addition, tables for chemical analyses, a number of chemical ratios and statistical information were developed.

The tables, showing the ground-water data base data sets, have one shortcoming because the codes are not translated. This was not done, because space had to be saved as the tables were developed for out-put on 80-character dot matrix printers. For the same reason, the data items, *farm name* and *comments*, are not included in the tables. However, the tables offer the user a neat data print-out of (most) data accommodated in the respective data sets. Each page of tabulated data contains a heading and 45 lines of data which can be included in a report. Complete data lists can be obtained by requesting printer out-put on the inquiry screen of the terminal.

Information regarding chemical analyses is tabulated in various forms and provides for out-put in mg/l or meq/l. The print-out of the ionic balance allows to identify unreliable or incomplete chemical analyses. A cut-off percentage error of 5% is used to distinguish between the two.

Additionally, a routine was developed which calculates specific chemical ionic ratios and prints them in tabulated form. Ionic ratios can be useful in identifying certain ground-water characteristics. When evaluating ionic ratios to identify possible anomalies, the differences between the individual analyses should be looked at rather than the order of magnitude of the values. It should be borne in mind that not all ratios are necessarily significant.

Ionic ratios are well-suited to compare ground-water qualities on a regional basis. When used in conjunction with the other hydrochemical and statistical analyses, classification of ground-water types for different regions may be possible.

6.4.2 Graphics

Graphic display is ideally suited to both identification of certain properties of the data and transfer of information The graphics included in the data processing package comprise the following categories:

(i) Unconsolidated/consolidated geology borehole profiles and cross-sections.
(ii) Hydrochemical diagrams.
(iii) Pumping test evaluations.
(iv) General graphics.
(v) Map and contour facilities.

Each of the five categories comprises of various graphical displays, which will be discussed with respect to their use in the evaluation of geohydrological data.

Geology - Consolidated/Unconsolidated

The geology forms an important part in any geohydrological investigation, because the aquifer characteristics are mainly dependent on the geological properties of the respective geological formations. In the data base, geology is divided into *consolidated* and *unconsolidated geology*. Both are further subdivided into a number of graphical displays:

(i) Borehole logs.
(ii) Borehole chart.
(iii) Borehole profile.

The unconsolidated groups has a further display *sieve analysis*.

Lithostratigraphic subdivision, grain-size distribution, penetration rate and borehole construction features are contained in these displays. A variety of descriptive features are provided to describe the geological formation in detail and provide the user with ample facilities to represent the geology in a graphical way. Where the geological description contains thin layers, the borehole log- and borehole chart displays have the disadvantage that the description of one layer is plotted on top of the previous thin layer. This can, to a certain extent, be rectified by editing the geological log before plotting.

Areal maps of the distribution of surface geology of a specific region can also be drawn. Two-dimensional contour maps or trend surface analyses can be used to show, for example, bed-rock elevations.

Hydrochemical analyses

Water quality evaluation requires a knowledge of the processes controlling natural ground-water composition and development.

The evaluation of chemical analyses is usually based on the relation of the different ions which determine the different types of water. The way in which the data are represented depends on the data to be processed and the questions asked. They can be represented on maps, showing the distribution of the respective parameters and/or on parameter relationship diagrams.

Four frequently used parameter relationship diagrams are available:

(i) Piper diagram.
(ii) Durov diagram.

(iii) Schoeller diagram.
(iv) SAR diagram.

The plotting positions of different ground-water analyses permit identification of various ground-water types. If different symbols are used, it is possible to distinguish between ground water from different stratigraphical units or of different evolutionary status.

Piper diagram

The Piper diagram represents hydrochemical analyses on a trilinear diagram. Depending on the plotting position in the central diagram, water may be identified as recent, dynamic and stagnant (Johnson 1975). Most people only use the central part of the diagram (the diamond-shaped field) to reach their conclusion on the water classification. This is incorrect. The total diagram must be considered to eliminate possible wrong conclusions. The Piper diagram has the shortcoming that only the major an- and cations are portrayed. Moreover, the cations, sodium and potassium, and the anions, chloride, nitrate and fluoride, are respectively combined, but neither the actual ion concentration nor any other parameters, e.g. TDS or pH, can be shown. The diagram does therefore neither bring out the total concentration nor critical ratios between certain constituents, for instance, between Na and K, Ca and Mg, CO_3 and HCO_3 and SO_4 and Cl.

Durov diagram

This is also a trilinear diagram. An advantage it has over the Piper diagram lies in the representation of additional physical parameters. It has, on the other hand, the serious disadvantage that two totally different chemical analyses may plot in the same position in the adjacent square (see section 6.3, Appendix I). This deficiency is partly overcome by the expanded Durov diagram, developed by Burdon and Mazloum (1958) and Lloyd (1965). This diagram has the advantage over the Piper diagram that it provides more detailed information on the an- and cation percentages and a better display of different ground-water types and hydrochemical processes.

The ordinary Durov diagram is not as informative as the Piper diagram in the classification of different ground-water types. Its major advantage lies in the fact that additional parameters may be included in the diagram, which can be significant when working in a mining environment.

The trilinear relationship diagrams permit the comparison between numerous ground waters and depict also the effect of mixing and dissolution of different ground-water types.

Schoeller diagram

The Schoeller diagram is a vertical scale diagram. Because the ions are plotted in meq/l, the actual parameter concentration is portrayed. Different water types are also distinctively depicted on the diagram. If the interconnecting lines of the different analyses between respective axes are parallel, the same ionic ratios at different concentrations are indicated. A further advantage of this diagram is that it can be easily adopted to accommodate any combination of parameters.

It is a disadvantage of the Schoeller diagram that only a few water samples can be illustrated on one diagram. The comparison of numerous water samples with one another is virtually impossible. If the ground-water types have successfully been identified by means of other hydrochemical evaluation techniques, a representative sample of each ground-water type can be plotted on the diagram, displaying the differences in the composition of the various ground-water types.

SAR diagram

This diagram is not particularly relevant for ground-water investigations, but is useful for irrigation purposes. It indicates the probability of salinity hazards and undesirable ion exchange for different soil types.

Apart from the hydrochemical relationship diagrams, hydrochemistry can also be represented and evaluated by using other graphical or statistical evaluation techniques. For example, the ionic ratios of an area can be contoured to indicate possible recharge areas. X-Y plots show relationships between various chemical parameters such as chloride and electrical conductivity or other parameters, e.g. TDS and ^{18}O. Thus, the user is not limited by the few hydrochemical diagrams provided, but may also use the other available techniques to present the different hydrochemical relationships.

Pumping test analyses

For the assessment of the ground-water resources of a region, a knowledge of the hydraulic properties of the water-bearing formations is needed. These data are usually obtained by means of pumping test analyses. The five most frequently used techniques are included in the package:

(i) Theis analysis.
(ii) Cooper-Jacob analysis.
(iii) Theis Recovery analysis.
(iv) Jacob Step Drawdown analysis.
(v) Multirate analysis.

General graphics

General graphics are used to illustrate two-dimensional relationships between various parameters, e.g. water-level graphs. General graphics may be used to supplement the other graphical evaluation techniques.

Map and contouring facilities

This group is divided into two sections:

(i) Map facilities.
(ii) Contouring facilities.

Map facilities

Maps are generally used to depict features, such as boundaries of geological formations, borehole positions, farm boundaries or drainage regions. The map facility included in the package is fully supported by PROCAD. Thus, the construction of very detailed geohydrological maps is possible, due to the 286 drawing levels and other inherent PROCAD functions. This means different geohydrological or other features can be drawn on different levels, giving the user the option to draw different combinations of levels.

Contour facilities

Geohydrological contour maps may portray conditions, which are not necessarily apparent, if other graphical techniques are used. When constructing a contour map, certain constraints have to be met to ensure proper function.

Depending on the distribution of data points, a variable reference grid must be constructed. The z-ordinate is calculated, using two different interpolation schemes. After the interpolation of the grid control points, they can be smoothed. Too much smoothing can filter out important features and leave the user with a trend surface. Depending on the number of samples points and their distribution, a suitable reference grid matrix must be chosen to ensure good representative contours.

It is recommendable to plot a map, showing the borehole positions and the respective z-values on the same scale as the contour map.. After the contours are constructed, both maps are superposed to control the correctness of the contours. If incorrect contours are identified, the reference grid matrix must be made finer and a new contour map be drawn. Correctness of contours is solely dependent on the user's judgement of the situation.

A facility to construct three-dimensional displays is also available. This facility is mainly for illustrative purposes.

6.4.3 Statistics

Because geohydrologists depend mainly on relatively sparse observations made at different times and/or places and in which a certain portion of uncertainty may exist, numerous statistical techniques have been developed to process these data. Statistical data processing provided, comprise four topics:

- (i) General statistics.
- (ii) Sequence analyses.
- (iii) Map analyses.
- (iv) Multivariate analyses.

It is essential that the use of the package is restricted to the appropriately skilled person. If uncritically employed, statistics may produce totally incorrect analyses. Statistical evaluation techniques are, to a great extent, sensitive to the distribution of the sample population. In the case of the data accommodated in the National Ground-water Data Base, many geohydrological related data (yields, transmissivities, grain-size distribution, etc.) display a log-normal distribution. Linear scales are seldom acceptable to nature. A millimetre difference between the diameter of two boulders is insignificant, but a millimetre difference between one sand grain and another is large and an important inequality. The natural scale for size classification, therefore, is logarithmic (Krige, 1951). It is not surprising that the distribution of, e.g. transmissivities in an aquifer, tends to be normal when the values are measured on a logarithmic scale (Delhomme, 1979).

This means that, before any statistical analyses can be done on such data, the respective data values must be transformed to their logarithmic value. An example of what may happen if data are evaluated without transformation (where necessary), is shown in tables 5.3(a) and (b), section 5.3.1. Here, the borehole yields were taken without transforming them into log values. A mean value was calculated which is not significant. If the standard deviation is looked at, a relatively large negative yield would be obtained which does not exist.

If the frequency distribution of particular borehole yields (figures 5.9(a) and 5.9(b)) are considered, it can be noted that both distributions are skew, most probably because of a log-normal distribution. Thus, the statistics in table 5.3(a) are of no significance.

Another fact to be remembered, is that geohydrological data have a multivariate nature. It is, therefore, good practice to start the statistical evaluation with the analysis of multivariate data. Here again, it is important to get rid of insignificant data by first constructing a correlation matrix of the respective data combinations.

The multivariate evaluation techniques are well-suited for the use in local and regional evaluation of hydrochemical data (see section 5.5). Multivariate analyses are advantageous for hydrochemical interpretations, because they are independent of the number or type of variables used, e.g. hydrochemistry can be correlated with borehole yields, depths or aquifer types. Furthermore, they avoid problems of closed-number systems, inherent in traditional techniques, such as the trilinear diagrams.

Graphical interpretations of ground-water quality are limited because

(a) there are a finite number of variables that can be considered,

(b) the variables are generally limited by convention to major ions and

(c) false relationships may be introduced by use of certain of the procedures (Dalton and Upchurch, 1978).

The spurious results are often an artefact of the use of closed-number systems. For example, in calculating the anion proportions for the Piper diagram, chloride, sulphate and bicarbonate, plus carbonate, should sum to 100%. Any increase in concentration in one or two variables, therefore, forces an apparent, but perhaps false, decrease in the third variable. By using raw data rather than percentage data, the problems inherent in the closed-number systems are avoided.

Furthermore, specific geohydrologic processes, such as the recharge of an aquifer, can be related, using multivariate statistical techniques. Areas impacted by slow natural recharge can be delineated by the use of hydrochemical factor analysis (Lawrence and Upchurch, 1982).

Another statistical technique which is useful in geohydrological evaluation is trend surface analysis. Trend surface analyses are frequently used for determining parameter distribution. Such analyses can provide an initial guide for interpretation, but tend to produce unrealistic geohydrological patterns, owing to data distribution bias (Lloyd and Heathcote, 1985). A useful aspect of trend surface analysis is that differences between surfaces of different degrees can depict anomalous areas.

To conclude, the use of the ground-water data base, in conjunction with the data processing facilities, can only be beneficial for development, planning and management of subsurface water resources in Southern Africa. The use of statistical methods does not replace good data, it may, however, improve the results. It is, therefore, once more stressed that the data base is only as good as the data entered. Equally important is that the user has a thorough knowledge of the techniques he uses when he analyses data.

Lastly, the data base and its processing facility were neither developed to replace logical thinking nor the ground-water expert's knowledge and abilities, but to serve as a tool increasing productivity and quality.

6.5 CONCLUSIONS

Today, the National Ground-water Data Base is one of the more comprehensive ground-water data bases in the world. At present, the data base is only accessible via the GOVNET network which is perhaps a disadvantage as it limits the number of potential users. This situation could be improved by establishing access via the CCWR mainframe computer in Pietermaritzburg.

However, the ideal situation for South African conditions would be to have a centralized ground-water data base, connected to a network of micro computer operated data bases, such as GBASE. This would ensure extensive use of the National Ground-water Data Base and at the same time country-wide entry of geohydrological data.

REFERENCES

ATRE, S. (1980). Data Base. Structured techniques for design, performance and management. John Wiley and Sons, New York, p. 433.

BAKER, C.H. and FOULK, D.G. (1980). Watstore User's Guide. Volume 2, Ground-water File, Open-file Report 75-799, U.S.G.S.

BAKER, C.H. and FOULK, D.G. (1980). Watstore User's Guide. Appendix, Open-file Report 75-426, U.S.G.S.

BEHOUNEK, N.J. (1980). The geology of an area west of Welkom, O.F.S. Unpubl. M.Sc. thesis, U.O.F.S., p. 162.

BREDENKAMP, D.B. and VANDOOLAEGHE, M.A.C. (1982). Die ontginbare grondwaterpotensiaal van die Atlantisgebied. Tegniese Verslag Gh 3227, p. 43.

BURDON, D.J. and MAZLOUM, S. (1958). Some chemical types of groundwater from Syria. UNESCO Symposium, Teheran, pp. 73-90.

BURGER, C.A.J., HODGSON, F.D.I. VAN DER LINDE, P.J. (1981). Hidrouliese eienskappe van akwifere in die Suid-Vrystaat. Institute for Ground-water Studies, U.O.F.S., Bloemfontein, South Africa, p. 154.

BURROUGHS CORPORATION (1977). Burroughs DMS II Reference Manual. Burroughs Corp., Detroit, Michigan.

BURROUGHS CORPORATION (1984). Burroughs Reference Manual, Logic and Information Network Compiler (LINC), Burroughs Corp., Detroit, Michigan.

CLARK, L. (1977). The analysis and planning of step drawdown tests. Q. Jl. Engng. Geol. Vol. 1010, pp. 125-143.

DALTON, G. and UPCHURCH, S.B. (1978). Interpretation of hydrochemical facies by factor analysis. Ground Water, Vol. 16, No. 4, pp. 228-233.

DATE, C.J. (1975). An introduction to database systems. Addison-Wesley Publishing Company, United States of America, p. 366.

DAVIS, J.C. (1973). Statistics and data analysis in geology. John Wiley and Sons, New York, p. 550.

DELHOMME, J.P. (1979). Spatial variability and uncertainty in ground-water flow parameters; a geostatistical approach. Water Resources Research, Vol. 15, No. 2, pp. 269-280.

DEPARTMENT OF WATER AFFAIRS (1986). Management of the water resources of the Republic of South Africa. CPT Book Printers, Cape Town.

DUNDON, P.J. (1977). Groundwater resource study of the Francistown area. M.Sc. thesis (Unpublished), U.O.F.S., Bloemfontein, South Africa.

DUROV, S.A. (1948). Klassifikacija prirodnych vod i graficeskoje izobrazenie ich sostava. Nauk SSSR, Vol. 59, No. 1.

FENNER, P. (1969). An introduction to mathematical geology. Models of geologic processes. AGI/CEGS Short Course Lecture Notes, Philadephia.

FRIEDMAN, G.M. and SANDERS., J.E. (1978). Principles of Sedimentology. John Wiley and Sons, New York, p. 792.

GROBLER, N.J., BEHOUNEK, N.J. and LOOCK, J.C. (1987). Development of pans in paleodrainage in the north-western Orange Free State. Paleoecology of Africa, Vol. 19 (in press).

HAAN, C.T. (1977). Statistical methods in hydrology. The Iowa State University Press, Ames, p. 378.

HARVEY, B.I. (1973). A computer system for storage and retrieval of hydrogeological data from well records. Institute of Geological Sciences, Report No. 73/18, p. 34.

HARVEY, R.A. (1978). Overview of the computer system for processing water resource information. Water Resources Section, Planning, Design and Investigation Branch, Public Works Department, Western Australia, p. 83.

HEM, J.D. (1970). Study and interpretation of the chemical characteristics of natural water. U.S. Geol. Surv. Water-Supply Paper 1473, Washington, p. 363.

HILL, R.A. (1940). Geochemical patterns in Coachella Valley. Trans. Am. Geophys. Un., 21, pp. 46-53.

JACOB, C.E. (1947). Drawdown test to determine effective radius of artesian well. Trans. Am. Soc. Civil Eng., 112, pp. 1047-1070.

JOHNSON DIVISION (1975). Ground water and Wells. A reference book for the water-well industry. Johnson Division, UOP Inc., St. Paul, Minnesota, p. 440.

JOHNSON, J.H. (1974). Hydrochemistry in groundwater exploration. Ground-water Symposium, Bulawayo.

JÖRESKOG, K.G., KLOVAN, J.E. & REYMENT, R.A. (1976). Geological factor analysis. Elsevier Scientific Publishing Company, Amsterdam, p. 178.

KRIGE, D.G. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand. Unpubl. M.Sc. thesis, Univ. of the Witwatersrand, Johannesburg.

KRUMBEIN, W.C. & GRAYBILL, F.A. (1965). An introduction to statistical models in geology. McGraw-Hill Book Company, New York, p. 475.

KRUSEMAN, G.P. & DE RIDDER, N.A. (1979). Analysis and evaluation of pumping test data. ILRI, Bulletin 11, Wageningen, p. 200.

LLOYD, J.W. (1965). The hydrochemistry of the aquifers of north-eastern Jordan. J. Hydrol., 3, pp. 319-330.

LLOYD, J.W. & HEATHCOTE, J.A. (1985). Natural inorganic hydrochemistry in relation to groundwater. An introduction. Clarendon Press, Oxford, p. 296.

MARTIN, J. (1975). Computer data-base organization. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, p. 558.

MARTIN, J. (1983). Managing the data base environment. Prentice Hall, Inc., New Jersey, p. 766.

MATTHESS, G. (1982). The properties of groundwater. John Wiley & Sons, New York, p. 406.

McCAMMON, R.B. (1975). Concepts in geostatistics. Springer-verlag, Berlin, Heidelberg, New York, p. 168.

MERCER, M.W. & MORGAN, C.O. (1982). Storage and retrieval of groundwater data at the U.S. Geological Survey. U.S.G.S. Circular 856, p. 9.

MILLER, R.L. & KAHN, J.S. (1962). Statistical analysis in the geological sciences. John Wiley and Sons, Inc., New York, p. 483.

MÜLLER, J.L. and BOTHA, J.F. (1986). A preliminary investigation of modelling the Atlantis aquifer. Bull. Inst. Ground-water Studies, Bloemfontein, 14, p. 174.

NEL, L. (1977). Die geologie van 'n gebied suid van Hopetown. Unpubl. M.S. thesis, U.O.F.S., p. 171.

PIPER, A.M. (1944). A graphic procedure in the geochemical interpretation of water analyses. Trans. Amer. Geophys. Un., 25, pp. 914-923.

PIPER, A.M. (1953). A graphic procedure in the geochemical interpretation of water analyses. U.S. Geol. Surv., Ground Water Notes, 25, pp. 914-923.

PROCAD SYSTEMS (1987). Procad Reference Manual.

RICHARDS, L.A. (1954). Diagnosis and improvement of saline and alkaline soils. United States Salinity Laboratory Staff, Washington, United Department of Agriculture, Agricultural Handbook, No. 60, p. 160.

ROGERS, J. (1980). First report on the Cenozoic sediments between Cape Town and Elands Bay. Tech. Report, 1980-0249, Geol. Surv. S. Afr., pp. 16-25.

ROSCOE MOSS COMPANY (1982). A guide to water well casing and screen selection. Roscoe Moss Company, Los Angeles, California, p. 76.

SACS (1980). Stratigraphy of South Africa. Handbook 8, Part 1 : Lithostratigraphy of the Republic of South Africa, South West Africa/Namibia and the Republics of Bophuthatswana, Transkei and Venda. Government Printer, Pretoria, p. 690.

SEYHAN, E. (1981). Introduction to multivariante statistical analysis in hydrology. Free University, Institute of Earth Science, Amsterdam, Netherlands.

SHOWEN, C.R. (1976). Storage and retrieval of water-resources data. U.S.G.S. Circular 756, pp. 20-25.

SPIEGEL, M.R. (1974). Fourier analysis with applications to boundary value problems. McGraw-Hill Book Company, New York, p. 191.

TILL, R. (1974). Statistical methods for the earth scientist. Department of Geology, University of Reading, Macmillan Press Ltd., London, p. 154.

TREDOUX, G. (1987). The application of hydrochemical techniques in groundwater investigations. Proceedings of the 1987 Hydrological Sciences Symposium, Vol. 1, Rhodes University, pp. 207-222.

TRUSWELL, J.F. (1970). An introduction to the historical geology of South Africa. Purnell, Cape Town, p. 167.

ULLMAN, J.D. (1980). Principles of database systems. Pitman Publishing Limited, London, p. 379.

VAN DER MERWE, A.J. (1983). Eksplorasie, ontginning en evaluering van grondwater van die sandafsetting in die Atlantisgebied vir watervoorsiening aan die Atlantisgroeipunt. Unpubl. M.Sc. thesis, U.O.F.S., Bloemfontein, p. 334.

VAN TONDER, G.J. (1981). Statistics and data analysis in geohydrology. Unpubl. M.Sc. thesis, U.O.F.S., Bloemfontein, p. 165.

VAN TONDER, G.J. & HODGSON, F.D.I. (1981). Statistics and data analysis in geohydrology. Institute for Ground-water Studies, U.O.F.S., Bloemfontein, South Africa, p. 161.

VEGTER, J.R. (1984). The role of ground water in South Africa's water supply. The proceedings of the International Conference on Ground Water Technology, Vol. 1, National Water Well Association, Worthington, Ohio, pp. 3-14.

WALTON, W.C. (1970). Groundwater resource evaluation. McGraw-Hill, Kogakusha, p. 664.

YOUNG, T.L. and VAN WOERT, M.L. (1986). PLOT88 Software Library Reference Manual. Plotworks, Inc. La Jolla, CA., p. 270.

SUMMARY

In 1984, the Water Research Commission awarded a contract to the Institute for Ground-water Studies at the University of the Orange Free State, comprising the development of a National Ground-water Data Base for Southern Africa. The data base had to make provision for storage of physical, geological and geohydrological information about any known ground-water source in Southern Africa. Also, it was to provide user-friendly software for the geohydrological processing of the data in the data base.

Three basic data structures were evaluated:

(i) Relational structure.

(ii) Hierarchical structure (tree structure).

(iii) Network structure.

The data in the ground-water data base are accommodated in three levels, consisting of basic information, header and respective repeating groups. Such a structure is easily converted into a tree structure which has the advantage of minimizing data redundancy.

The tree structure has the distinctive disadvantage that the data have to be accessed through the hierarchy which makes the access of certain data relatively inefficient. This situation was, however, rectified by combining a few network features with the features of the tree structure, which lead to more efficient data retrieval.

A data base depends, next to the type of data, on the quality of data added. To promote entry of good quality data, certain data items were declared mandatory. In addition, as many logical data checks as practically possible, have been included in the data base programs.

The parameters present in the data base were included, after evaluation of the parameters included in other existing ground-water data bases and after thorough discussion with the Directorate Geohydrology of the Department of Water Affairs to adapt the data base to especially suit Southern African conditions.

A number of data items and descriptions tend to be repeatable. Such items were coded, ensuring fast data entry and reducing space requirements on the storage device.

Four ground-water data bases were evaluated, before designing the National Ground-water Data Base:

(i) Data base by the Institute of Geological Sciences (U.K.).

(ii) Data base by the Public Work Department (Western Australia).

(iii) Borehole data base - S.W.A. (Namibia).

(iv) WATSTORE - U.S.G.S. (U.S.A.).

If the National Ground-water Data Base is compared with these data bases, it corresponds well with the Ground-water Inventory File of the National Water Data Storage and Retrieval System (WATSTORE) of the U.S.G.S.

The National Ground-water Data Base for Southern Africa has, however, the distinct advantage over the WATSTORE system that it is fully supported by data processing facilities. Data can easily be accessed and extracted from the ground-water data base and transferred to a micro computer. Here, the data can be evaluated by a comprehensive data processing package.

Data processing is divided into three categories, comprising of tabulated output, graphical output and statistical analyses. The graphical output includes geological, geohydrological and hydrochemical processing facilities, while the statistical analyses include general statistics, sequence analysis, map analysis and multivariate analysis.

To test the application and efficiency of both the data base and the processing facility, data of the Orange Free State and Atlantis were entered into the data base and evaluated with the data processing facility. Generally, good results were obtained. However, where problems were encountered, adjustments were made to eliminate these problems.

Thus, the National Ground-water Data Base is one of the most comprehensive existing data bases in the world, due to its ability to store an extensive range of geohydrological related data and its ability to process the stored data.