INSTITUTE OF HYDROLOGY

COMPUTER TECHNIQUES IN GROUNDWATER RESOURCE STUDIES

1ST INTERIM REPORT

JULY 1978

SUMMARY

Computer techniques offer a basic facility to handle and analyse large quantities of data. One of the main aims of this project has been to develop general analytical methods which can be applied to groundwater resource studies.

The work described in this report has been carried out in the first year of a three year project funded by the Overseas Development Ministry (ODM).

CONTENTS

Key to Figures and Tables

Figures

Tables

INTRODUCTION

This is the first interim report on the work carried out in the ODM financed project concerned with computer techniques in groundwater resource studies which is aimed at providing groundwater hydrologists with an organised, efficient and modern approach to data. As such, it is not concerned with specific overseas groundwater resource studies but rather methods of 'universal application' in a common framework that will meet the requirements of most studies.

The reasons for such an approach to groundwater resource studies stem both from the type of data and their requirement for analysis. The data are often varied (water levels, water chemistry, pumping tests, etc) and even a modest study can lead to large quantities of diverse data. The analytical requirements often depend upon using elements of different data sets simultaneously whilst drawing as many scientifically justifiable conclusions from them as possible. Moreover, the presentation of the data in a unified and organised way plays an important role in such studies.

In fact, our experience has shown that the preparation and presentation of the data causes major problems at the reporting stage. It is evident from a review of groundwater resource reports that frequently presentation of this information is offered as a substitute for analysis or if analysis is made the extra effort of presenting the data proves too much for the authors.

In this report, we are concerned with introducing the concepts of the system, some of the language and the design philosophy. At this stage, we are not writing a manual for the operation of the programs that have been developed. The manual with the preliminary set of application programs and data system will be released later in 1978. At that stage, we envisage that the users of the system can co-operate in its future development by suggesting further analytical methods and testing the existing ones.

In the first section, we introduce a computer based data system that has been developed specifically for groundwater resource study data. In this way, we have been able to make the most use of our knowledge of the type of data and its requirements for analysis. Alternative general purpose data systems such as G-EXEC[1] and DMS1100[2] include general overheads which have been avoided in our tailor-made system. Obviously, the development of even a tailor-made system in a limited time scale from the beginning would have been unrealistic. Instead, we have developed the system from the combined experience gained in both the Oman data system and an applied hydrological data system. The development of these two earlier systems, over a period of some six years, has meant that progress could be conveniently compressed into the first year of the study in a similar time scale to the implementation of an alternative general purpose system. Moreover, the tailor-made system gains in efficiency because of the specific nature of its design.

[1] Scientific data base (originally a geological data base) supported by NERC Central Computing Group.

[2] Univac data base based on the recommendations of the codasyl (international data base standards) committee.

In the second section we deal with the application programs that have been developed to be used in conjunction with the data system. These programs are a preliminary set and are based on the methods used in previous studies although they have been completely rewritten to make maximum use of the structure of the new data system. Further methods will be added in the remaining two years corresponding to the requirements of field workers. In the final section we discuss a control langauge which is being developed to facilitate the use of the application programs with the data system.

DATA SYSTEM

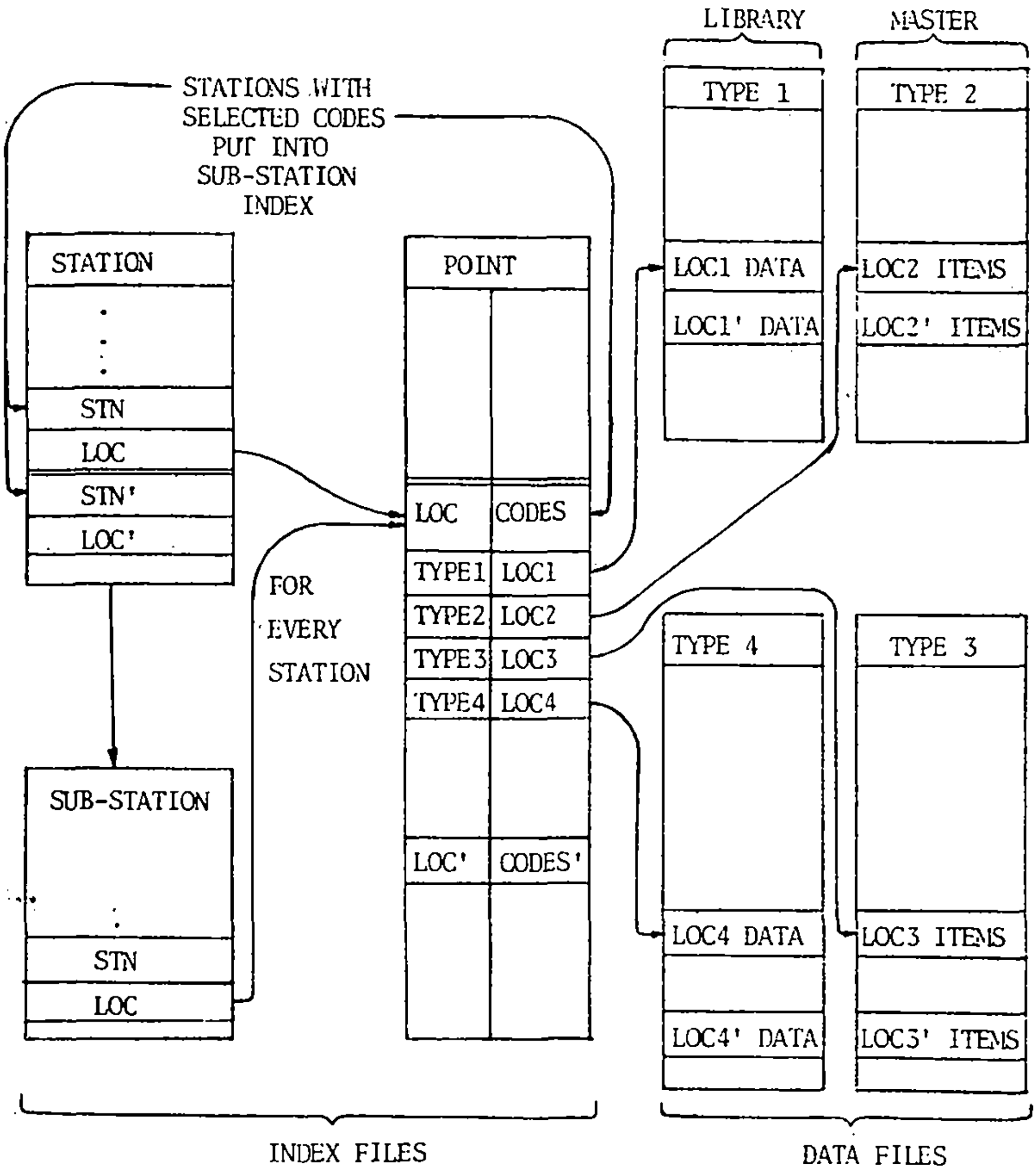
The data system is a collection of files of data which are particularly organised in such a way as to minimise the effort required to abstract or add any single piece of data. All this organisation has been achieved by using two index files, created automatically as data is put on the system. The first file, and the first point of entry in accessing the data, is a station-index. By station we mean any defined point in the study area (eg with a station number in the form of a grid reference) that has data associated with it. The station may be a well with water level data or a spring with chemistry sample data etc. The second file, point-index, contains information about the different types of data and includes codes such as source type, aquifer type etc.

The data itself is contained in additional files of two different types - (a) Library files or (b) Master files. This distinction is made because of a fundamental difference between two types of data collected in groundwater resource studies. For instance, water level, pumping test, spring discharge, lithological log data are all defined for distinct points and measured at different times (or in some cases depths). Whereas water chemistry sample analyses and well description are not only defined in space (and for chemistry analyses by sample as well) but several different variables are also measured. Data falling in the first category is generally stored in library files and those in the second on master files.

All the files are direct access; that is, knowing the location of the data, it can immediately be found without going through the file sequentially. The programs are all written in ascii standard Fortran, so that they are as little machine dependent as possible.

Figure 1 shows a schematic view of the data system. The arrows indicate typical passages requiring either access of data or selection of the station index. The diagram indicates how the station-index, containing only the station number (STN) and location (LOC) of further information, is linked to the point-index containing code information and locations of the pertinent data on the master and library files for the station. The station index is nominally restricted to 3000 stations to allow all the stations to be held in core. The point index can contain the locations and descriptions of up to 28 different types of data for each station. The diagram shows a typical configuration of two library and two master named data files. Thus, to access, say, type 2 data for a particular station number; the passage from the station to the point-index would be via LOC and from point-index to type 2 master data file would be via LOC2.

Figure 1   Schematic View of Data System



Our experience of operating data systems has shown us that the selection on a catchment, sub-catchment or aquifer type (etc) basis is a vital facility needed during the development of understanding of the groundwater system and its problems. This is achieved by selecting a subset of stations based on a selection parameter (code), or within a predescribed area, to produce a sub-station index as shown on the diagram.

1.1 Index Files

(a) Station Index

This file contains both the station number (grid reference) and the location where further information about the station can be found in the point index file. The station index is kept as small as possible so that it may be held entirely in core; that is, within the program. This index is both constructed and searched using a binary chop process. That is, the station numbers are arranged in ascending order and the appropriate position is found to be below or above the mid-point of the index. The process is repeated in the correct half until either the station number or its position for insertion is found. The file structure is shown in Table 1.

(b) Point Index

This file contains the locations, lengths and processing date-times for each of the types of data for each station. The date-times not only give a diary of the data development, but they can also be interrogated before overwriting information for a particular station. This latter use is well exemplified in the SPLINE program in Section 2. The file structure is shown in Table 2.

Table 1 Station Index

Word	Description	Example data
	Previous end[1] of STATION INDEX	2059
	Current end[2] of STATION INDEX	2323
	10 - identifier for STATION INDEX	10
4	Free	
(5	Station number 1	64716324[3]
(6	Location on POINT INDEX for secondary index for Station 1	1051[4]
(7	Station number	64817324
(8	Location on POINT INDEX for secondary index for Station 2	1251[4]
(2N+3	Station number N	64818742
(2N+4	Location on POINT INDEX for secondary index for Station N	151[4]

Notes

[1] Previous end of any file refers to the next free location in file the last time it was used; and

[2] Current end of any file refers to the next free location at present.

[3] In this example the station numbers we use are eight figure grid references; the first four figures are northings and the last four are westings. In fact the data system can adapt to any grid reference precision by associating a compound number for which the station number can be ordered and the grid reference extracted.

[4] Although the grid references are numerically ordered the locations referred to in the point index are not.

Table 2 Point Index

Word	Description		Example Data
	Previous end of POINT INDEX		16151
	Current end of POINT INDEX		18251
	9 - identifier for POINT INDEX		9
4	Location of next free type number		5
5-7	Name of type 1 file (12 characters)		WATER-LEVELS
...			
86-88	Name of type 28 file (12 characters)		
89	Block length of type 1 file		
...	...		
116	Block length of type 28 file		
117	Sample size of type 1 file		
...	...		
144	Sample size of type 28 file		
145-150	Free		
151	Station number 1		64818742
152-166	Codewords[1] for station number 1		000000001100011
(167	Location of type 1 data)	) for station	5
(168	Length of type 1 data)	) number 1	124
(169	Date-time of type 1 data)		7812101242[2]
(...			
(248	Location of type 28 data)	) for station	
(249	Length of type 28 data)	) number 1	
(250	Date-time of type 28 data)		
100N+51	Station number N etc		64817342

Notes

1. These codewords (up to 15) have assigned meanings, for instance, a well source in one of a group of aquifers in a particular sub-catchment etc.

78 (year), 12 (month), 10 (day) and 12:42 (time).

1.2 <u>Data Files</u>

(a) Library files

These files are designed to hold long strings of data for each station, for instance water level records for a particular well. The data are arranged in blocks of two words with the date(key) in the first word and the water level measurement in the other. This allows insertion of new data in the correct chronological position. Library files are identified by having the file flag equal to zero. Table 3 shows the library file structure. However, data with several different values corresponding to measures of different attributes cannot easily be put into library file form. For this reason master files have also been developed.

Table 3 <u>Library File</u>

Word	Description	Example Data	
1	Previous end of LIBRARY FILE	5143	
2	Current end of LIBRARY FILE	5473	
3	Type number +10 - identifier for LIBRARY FILE	11	
	BLOCK LENGTH	2	
	1st word of data for station A, as given in point index (arranged in blocks of data)	781011	(Date)
		96.34	(water level)
		781012	(Date)
...		...	
4 + length	Last word of data for station A, as given in point index	124	
5 + length	1st word of data for station B	070978	
	(repeated in strings of data referenced by the location and length given in the point index)		

(b) Master files

Master files are designed to hold either a single or small multiple

(< 100) of several different measured variables (items) for each station. Water chemistry data, for instance, where one or more samples may be taken at each station for a variety of determinations (eg anions, cations, pH, electrical conductivity etc) considered as the items.

There are two modes of operation for master files which are relevant to different types of water resource data. The first mode, chosen by declaring a file flag equal to one, is the single data case; in this case it is assumed that new data from the same station replaces existing data. This mode is useful in cases where data is changed regularly and only one value for each item exists (e.g. aquifer parameter estimates, results of statistical analysis, etc). The second mode is one in which new data is appended to the existing data (the small multiple data case). A good example is that of water chemistry data described above. This mode is chosen by setting the sample size parameter to be greater than one. Table 4 shows the master file structure.

1.3 Operation and Additional Features

Information is added to the data system by execution of a program, whereas it is accessed in an application program by a subroutine. This gives a highly efficient approach to data storage and access, being on the one hand safe in updating existing files and on the other flexible in allowing use of several different data types in one application program. Moreover, when the control langauge is used in conjunction with these programs it removes the chore of executing several of them one after another. The program for basic data handling uses a subroutine with the formats for all the different

Table 4 Master File

Word	Description	Example Data
1	Previous end of MASTER FILE	6105
	Current end of MASTER FILE	7205
	Type number + 10 identifier of Master file	12
	BLOCK length (maximum foreseeable words/station)	100
	Next free ITEM location	74
)) 5+2*BLOCK)))	Space reserved for ITEM NAMES (two words/item to allow for 8 character names)	SAMPLE DATE
6+2*BLOCK)) 5+3*BLOCK)	Space reserved for relative location of each of them within the data for each station	1 3 ...
6+3*BLOCK))) 5+4*BLOCK)	Space reserved for length of each item	1 1 ..
6+4*BLOCK	1st word of data referenced by the location given in the point index for a station. A total of BLOCK such words are reserved for each sample set for each station. Only locations for which an Item value has been declared are filled. The remaining location are set to -999 to distinguish from that data with a zero value.	651 (sample number) 781218 (date)

Repeated from 6+4*BLOCK and referenced in the POINT INDEX types of data. This allows the system to store data in any appropriate format; relieving the necessity (typical of many other systems) for standard forms for the collection of water resource study data. Coupled with this program are a set of commands which allow for insertion, deletion and replacing data, synchronization of the date-time and adding new code information to the point-index.

To retrieve data from a library file, two subroutines are used. The first finds the type number corresponding to the named data file and assigns it to the system. It also returns the block length and the file flag which is set to zero for library files. The second subroutine is used to access the data with the station number used as an input argument; it fills a common block with the data. On return, an error switch is set to zero if the transfer of data has been successful. If the station number is set to a default value, the subroutine will access the data from all the stations on the station-index or sub-index.

To retrieve data from a master file three subroutines are used. The first is the same as for the library files, but the file flag is returned with 1 for single data and greater than one for multiple data. The second subroutine returns the locations and lengths of requested items and the third one accesses the data with, as before, an error switch and the ability to set a default value to get all the stations on an index.

Additional subroutines have been written to access the pertinent information from the point-index for data held on both library and master files. These routines are useful when the applications programs need to select stations on certain code data, for instance just well sites, springs or aquifer characteristics. They can also be used to find the date-time of the file creation, thus aiding efficient file updating.

Three additional programs have been written to give a resumé of the information on the data-file, to pack the data file (removing deleted data) and to select stations from the station-index on certain codes or within a quadrilateral boundary. After selecting stations on some criteria, it is then possible to create a sub-index and use this new index with the data system subroutines, thus accessing the relevant data with the default station number.

APPLICATION PROGRAMS

The set of application programs that we are presenting after the first year of this study are obviously based on our knowledge and experience of resource studies. Thus, necessarily, they will appear unbalanced in the sense that areas in which we have had little experience may only be covered in a superficial way; whereas areas in which we have a great deal of involvement will appear relatively complex. This position will obviously alter as we not only gain experience in these areas, but also find the additional requirements of other studies.

Methods of analysis can be divided into various categories, depending on the different types of data:

(i) Water level data - drilled and hand dug wells dipped on a routine well round or from water level recorders. Data in general consists of a station number (grid reference), a date as well as possibly a time, and a reading of depth to standing water level normally measured in metres.

(ii) Water chemistry data - the results of chemical analyses of water samples taken at wells, springs, surface water channels etc. Data consists of a station number, a sample number, date, possibly depth and values for the different chemical determinations (eg pH, cations, trace metals etc).

(iii) Well data - descriptive data concerning the nature and geographical location of the well. The nature of the well can be described by many values such as depth, screen details, discharge, etc.

(This type of descriptive data applies to all water sources)

(iv) Boundary data - describing the areal limitation of the study area as well as catchments within this area. The data consists of strings of grid references following closely the outline of geological and topographic boundaries.

(v) Pumping test data - results of either step drawdown or constant discharge tests carried out at a drilled borehole. Data consists of time from test starting and drawdown with discharge details. (pre-test levels would be stored in the water level file).

(vi) Lithological log data - collated from observations and measurements describing the nature and extent of the material the drill passes through and the progress of the drill through it.

To these data categories, several others could be added, depending on the requirements of the study and the nature of the area (eg isotope data, geophysical logs, falaj (qanat) measurements etc). What is immediately apparent is the diversity of information and the implication of the frequent need to analyse two or more data categories simultaneously. This really is the essence of the project; to provide analytical flexibility coupled with the ability to handle large quantities of diverse data.

The first set of application programs deal with water levels and offers nearly complete analysis of water level data going up to modelling but not, at the moment, dealing with it.

2.1 Water Levels

The different aspects and needs to analyse water level data can be conveniently represented in the form of a block diagram, Figure 2. In this diagram we have represented the transition from raw water level data to final results as moving from left to right. We have reserved the circles to indicate programs; the data system files (which are all of the library type) as boxes. An asterisk signifies files which are temporary in a sense which will be expanded later. Thus, there are eight programs and four data files as well as the station and point index which deal in an integrated way with a variety of different analyses. Also there are links that can be used for further analysis, for instance digital modelling and statistical analysis. In fact, throughout the application program development we have tried to avoid duplicating statistical analysis that is adequately covered in the ASCOP[1] statistical package, and can be conveniently used with the data system.

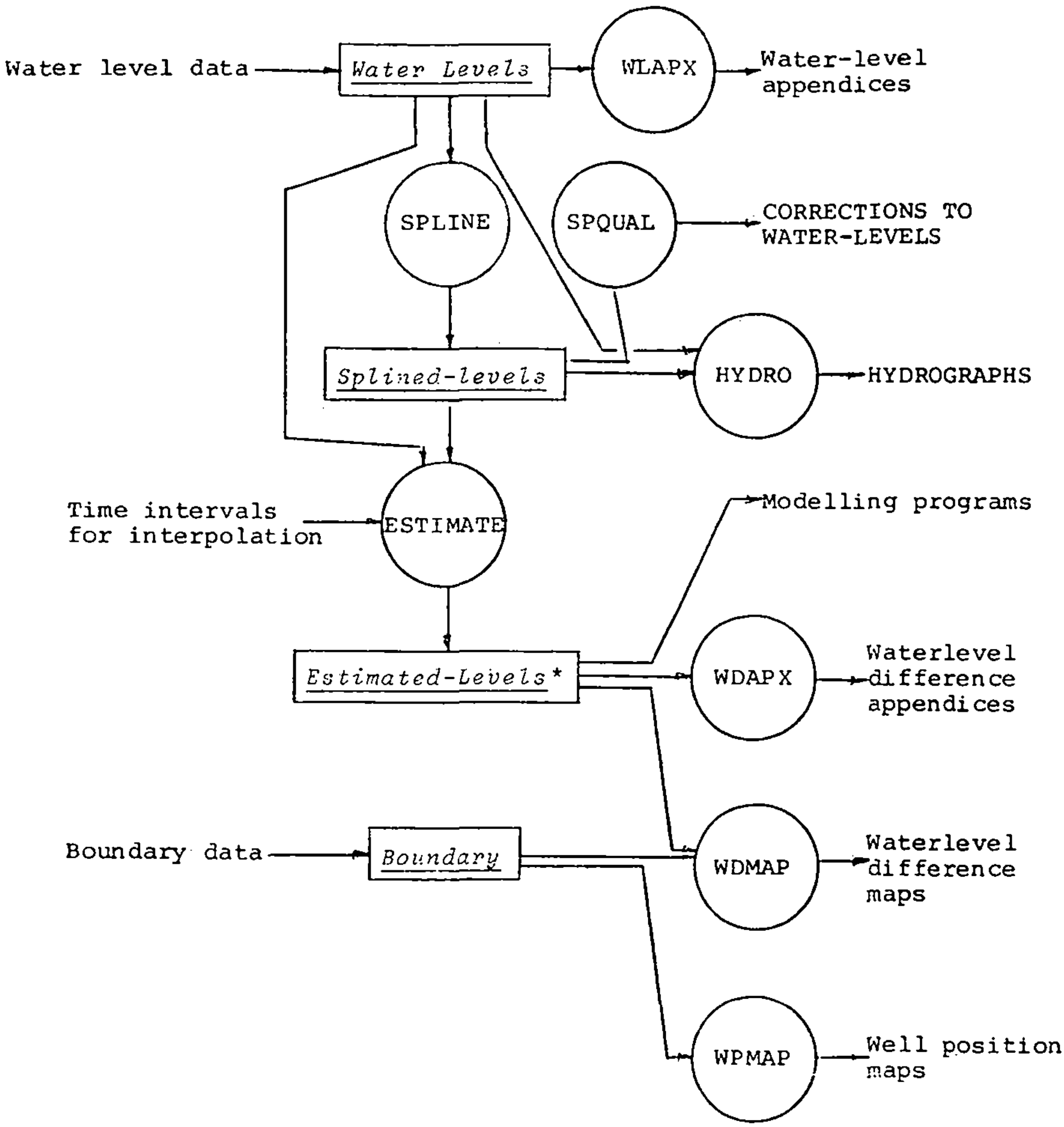
Supplementary raw data in the form of boundary grid references (iv) may also be required for mapping water level differences and well positions.

<u>The purpose of the programs can be described as follows:</u>

SPLINE Fits a smooth curve through irregular temporally spaced water level values using the cubic spline method (cubic segments fitted to two points in turn assuring smoothness at the joins). The program creates a data file of values (*splined-levels*) so that a value for any time can be interpolated.

[1]A set of computer programs developed by the National Computing centre to facilitate easy statistical analysis (eg multiple regression, analysis of variance etc.)

Figure 2 Water-Level Analysis



SPQUAL	Quality control check for water levels. The method uses the information calculated by the SPLINE program to check that a water level observation lies within a prescribed tolerance of the smooth interpolated estimate using points either side of the value. The program prints a diagnostic and does not correct the file automatically since under some conditions quite large and abrupt changes in water levels occur naturally.
WLAPX	Produces pages of output suitably formatted for inclusion as appendices in reports of water level data. The data is organised into station number order and using the selection facility can be sorted for any of the codes such as aquifer source, catchments or with any chosen quadrilateral boundary.
HYDRO	Plots well hydrographs; graphs with water level measurements as ordinates and the times as abscissae. To fit smooth curves through measurements, the program uses the *splined-levels*. Variable scale and time interval facilities are included.
ESTIMATE	Creates a temporary library file of estimated water levels given details about the time intervals. The *estimated-levels* file is temporary in the sense that it is overwritten as soon as a new set of estimated water levels for different dates are requested. So only the most recently used

estimates are retained on file.

WDAPX — Produces a summary for each time interval of the estimated differences in water level for each station.

WDMAP
WPMAP — Produces on the graph plotter maps of water level differences or positions of stations with either all or selected boundary information and a grid overlay for easy identification by position.

Notes

1. The cubic-spline algorithm is part of the Harwell subroutine library developed by the theoretical physics Division of the United Kingdom Atomic Energy Authority.

Time interval options in ESTIMATE include water levels at either regularly spaced intervals (days, months or years) starting from a specified date or at a specified series of irregularly spaced dates.

Water level and water level difference contour maps can be drawn by using the estimated levels file and the SACM (surface approximation and contour mapping) automatic contouring package written by Application Consultants Incorporated, Houston, Texas.

2.2 Water Chemistry

The number and type of determinations of this data varies from study to study and consequently it is very difficult to produce programs with the necessary degree of generalisation to deal with all the results of any particular water resource study. However, there are a number of aids to the interpretation of water chemistry which are likely to be generally applicable.

These include Piper and Stiff diagrams and chemical ratios. For these and other similar comparisons between relative concentrations of various ions it is necessary to have some information about the reliability of the data. Ions for which determinations are commonly made are Ca, Mg, Na, K, Fe, HCO_3, CO_3, SO_4, Cl and NO_3; analyses which include determinations of the major ions (Ca, Mg, Na, HCO_3, SO_4 and Cl) are called complete.

A check of the quality of the data of these complete analyses is to compare the balance between cations and anions when all the determinations are converted to me/l. Also a crude classification scheme can be devised to group water chemistry samples on their domestic and agricultural suitability based on the World Health Organization criteria, the sodium adsorption ratio (a combination of Na, Ca and Mg measurements) and the electrical conductivity. It is quite probable that an industrial supply suitability classification can be added as well. Without dwelling on the implications of these classifications on the nature of the water, it is certain that they may, coupled with ion balance information, offer a valuable insight to the reliability of the data, since they highlight water chemistries with unusual characteristics. To this end we have developed three programs to utilize this information to begin with in the context of the reliability (or quality control) of the water chemistry data.

The first program automatically appends the balance and class information to the water chemistry master file. A positive number for the balance indicating the percentage difference for complete analyses, or a negative number indicating incomplete analysis. This means that the balance and class information need only be calculated once and since it is with the chemistry data, can be used when it is required. Obviously, the data description at the beginning of the master file is appended so that the position of this data is known to other programs. The other two programs in the group of reliability checking programs take this balance and class information and display it in such a way that it is simple to locate possible erroneous values. For instance, in the balance checking program, a range of percentages is supplied so that all the chemical analyses with a balance falling in that range can be arranged into descending order and displayed. Similarly with the agricultural and domestic suitability classification checking program; all the combinations of chemistries with unusual classifications can be displayed.

The interpretation of water chemistry data is normally carried out using methods such as Stiff and or Piper diagrams. The Stiff diagram method has already been implemented with the data system and the Piper diagram method will follow shortly. These schemes are used essentially as visual aids for interpreting chemical data so that a unique classification of water type can be made. However, most schemes so far devised are subject to misleading interpretation because of assumptions made in the hydrochemistry in representing the data in diagrammatical form. Indeed the diagram itself can be misleading. In essence these schemes treat the chemical concentrations as numbers without regard to their underlying hydrochemical properties. This approach is adequate for the synthesis of data for a large number of chemical elements where water quality is concerned. However, they can breakdown when applied to hydrochemical problems (eg mixing of different water types, conservancy of chemical elements). Thus, it is necessary to include schemes such as Piper and Stiff diagrams to classify water types as options within the framework of analytical techniques. In addition, it is envisaged that a method such as clustering or multi-dimensional scaling will be used to classify water types based on their statistical properties with reference to both the concentration and concentration relative to a conservative element (eg chloride) of the elements. This will impose a more rigorous scheme for classifying water type and the necessary research will be carried out in the second year of the study although some preliminary work has already been carried out.

2.3 Pumping Tests

The data defined in (v) fits into the library file category. Basically there are two types of pumping test, step-drawdown and constant-discharge. The step-drawdown test is to record the well level drawdown as the discharge is increased in steps as opposed to a constant discharge test. Two programs have been written to plot the results on the graph plotter on both log-log and log-linear axes for the step-drawdown and constant discharge test. The water level hydrograph from an observation well can also be plotted. Research has been undertaken to investigate the use of optimization methods in conjunction with pumping test results to find the aquifer parameters of transmissivity and storage coefficient.

2.4 Lithological Logs

This data defined in (vi) again fits into the library file category. The program that has been developed is a lithological log graph plotting program that conveniently encapsulates the data in one graph to be used for both analysis and appendix.

2.5 Appendices

Although the appendix programs are not strictly a separate category, since they form a part of both of the previous classes, they are sufficiently important to be considered separately.

An important part in the final stages of any project is the organisation and presentation of results in tabular form for inclusion in reports. There are a number of distinct types of appendix format which have been repeatedly required (see below) and it is proposed to generalise each of these types as far as possible and to write one program which will deal with each type of format; thus any data from the data system which is amenable to display in one of the standard layouts could be summarised without difficulty.

The great advantage of this approach to appendix writing is that almost all the output which is likely to be required could be generated by one of a small number of standard programs; only in a few cases where a very specialised format is required will it be necessary to write a new appendix program. As well as simplifying the preparation of reports, it will be much easier at any intermediate stage in a project to obtain summaries of particular types of data which have been stored on the data system.

The appendix programs which have already been written, and discussed in Section 2.1, to deal with specific output from the data system will be absorbed into the general appendix-writing scheme in due course.

Types of appendix format:

<u>Type 1</u> (single quantity arranged in station order)
For example, water level differences summary. Summaries for successive time periods, each summary has data for many stations. Data arranged in columns:

STATION NUMBER   DIFFERENCE

Generalises to : n quantities per station, m stations per line (where n, m are sufficiently small to fit available line

<u>Type 2</u> (single quantity arranged in keyword order)
For example, water level summary. Summary for each station: each summary has data for many different dates. Data arranged in columns:

DATE   WATER LEVEL

Generalises to: n quantities per date, m dates per line.

Type 3 (multiple quantity arranged in both sample and item order)

For example, chemistry appendix. Many items for each sample (too many to use type 2). Items in rows, samples in columns.

	SAMPLE 1	SAMPLE 2	SAMPLE n
DETERMINATION A			
DETERMINATION B			
...			
DETERMINATION X			

Type 4 (multiple quantities with miscellaneous ordering)
For example, well description appendix. One station per page; whole series of parts of information about the station, not necessarily arranged regularly in rows or columns.

These four basic types of appendix format can be conveniently considered as those using library files, types 1 and 2 and those using master files, types 3 and 4.

CONTROL LANGAUGE

As has been illustrated in the previous section the design philosophy of the application programs has been to write self-contained programs that do a specific task. Thus it is evident that for anything but the most trivial analytical requirement several program executions will be required. As an example consider plotting water level differences;

a typical program sequence would be as follows:

Program (1) - put new water level data on to the file;

Program (2) - calculate new spline values;

Program (3) - put new splined levels on the appropriate file;

Program (4) - check quality of the new water levels;

Program (5) - estimate interpolated water levels for the given time interval;

Program (6) - put estimated-levels on the appropriate file;

Program (7) - plot water level difference map given boundary information.

This sequence of seven program executions would not only be tedious but also prone to error if each one had to be executed in turn with the correct input data, flags etc. Thus to facilitate the running of standard sequences of programs, a control program has been written enabling the user to set a chain of programs in motion using simple control commands. The list of commands is added to as each new application program is implemented onto the system and will eventually be used to create a comprehensive langauge catering specifically for water resource analysis.

Obviously it will be possible to execute the programs individually outside the control language for problems not falling into the standard application program category.