SWEDISH WATER AND AIR POLLUTION RESEARCH LABORATORY

INSTITUTET FÖR VATTEN- OCH LUFTVÅRDSFORSKNING

TELEPHONE 22 25 40
TELEGRAMS WATERRESEARCH

DROTTNING KRISTINAS VÄG 47
114 28 STOCKHOLM SWEDEN

ADP-System for Administration and Processing of Environmental Control Data.

I. Water Pollution Data.

Bo Åsell and Lars Landner

B 126
Stockholm
July
1972

Introduction

To form a correct opinion on the pollution situation in a recipient, it is usually necessary to have access to a very extensive primary material of water data. Data collected to illustrate the situation must be supplemented by earlier data, preferably collected over a period of many years, in order to provide information about long-term changes. Consequently the data required to draw correct conclusions complicates the work through their sheer multitude and difficulty of analysis. Today, therefore, the application of ADP technique appears to be a necessity for the creation of an effectively operating environmental control system.

The Swedish Water and Air Pollution Research Laboratory (IVL) has had under way since the beginning of 1971 a project for development of an ADP system for administration and processing of environmental control data, particulary within the water pollution field. The project has been supported by the FUSAM Committee of IBM and by the Foundation for Water and Air Pollution Research. An account will here be given of the work done hitherto on the administration of data from physicochemical water analyses.

Users of the ADP system

The primary object of the ADP-system is to serve our own organizatio which collects a large quantity of water data in conjunction with various research projects and also in conjunction with various programs for the control of pollution from industry and municipalities. Consequently the users of the system are secondarily also our various principals - industrial enterprises, municipalities, water conservancy associations etc. There is also a tertiary group of users, the regional and central control organs of government agencies.

Nordic cooperation

A joint group for data processing within the field of water conservancy, with members from Denmark, Finland, Norway and Sweden, has been formed under the auspices of the Nordic Council for Applied Research. The group will attempt to divide the work on development

of a data bank between the Nordic countries. It will also regularly issue information on what is being done in the respective countries.

The Water Research Institute in Denmark has plans for use of the system developed by IVL for water chemistry data.

Adaptation to terminal processing on a time-sharing basis

IVL's ADP system is adapted to terminal processing via a time-sharing system of type DDD/McGill. The main advantages of this system are that comparatively quick access to stored data is obtained at a fairly low cost and that the control and setting up of recorded data can be simply done at IVL and by persons acquainted with the respective data materials (this is especially important for the recording of older material). Furthermore the system, when extended to full capacity, will then be readily accessible to different users who have their own terminals.

Brief description of the ADP system

The hardware equipment for the time-sharing system consists of a central processor IBM 370/145 (256 KB).

Disc memory units IBM 2314. Magnetic tape stations IBM 2540. Line control equipment Memorex 1270.

At IVL the equipment consists of a teleprinter terminal Olivetti TE 300 and a magnetic tape encoder FRIDEN 4301. All data programs in IVL's system are written in IBM FORTRAN IV (G).

Briefly the data programs developed hitherto may be classified as follows (see Appendix 1):

a) Programs for input and error checking of data recorded on magnetic tape

b) Programs for certain types of transformation and restructuring of data

c) Programs for output of data in publishable form

d) Programs for storage of data in a register in which data are linked together on a geographical area basis

e) Programs for recovery of stored data according to a given search profile

f) Programs for statistical analyses of data

Recording of input data

Results of physicochemical water analyses of samples from the various recipients are entered on special forms at IVL's laboratories, together with identification data. The forms then constitute a basis for recording of data recordwise on magnetic tape. A record consists of an identification and the analyses from one sampling point on one occasion. The record structure is as follows: the first 24 positions certain 12 identification variables followed by the respective analyses, each preceded by an alphabetical code on 2 positions ID (12), CODE, MEASURED VALUE, CODE, MEASURED VALUE etc. The ID variables which identify the record are as follows:

1) General origin, 1 position (e.g. recipient data, marine data)
2) Category of data, 2 positions (e.g. water chemistry, hydrological)
3) Type of effect on the recipient (e.g. pulpmill, municipal)
4) User of the information (client), 2 positions
5) Internal designation of geographical area, 3 positions
6) Internal designation of station, 3 positions
7) Depth of water where sample collected (in dm), 3 positions
8-10) Date of sampling (year, month, day) 3x2 positions
11) Internal designation of season, 1 position
12) Code for deepest sampling points at the respective station, 1 position

Other data of the respective sampling points, such as geographical coordinates and watercourse, are recorded in separate registers.

Error check and setting up of recorded data

The data recorded on magnetic tapes are transferred to a disc memory file. An error check of the records can then be initiated via a terminal. Input data to the check program are permissible ID designations, types of measured values, maximum and minimum of measured values, and format for the analyses of the records. Input data for the respective geographical areas are given separately.

The resulting error lists are then checked against the form of analyses and the data file can thereafter be set up via a terminal.

Data transformations and output

The data file now constitutes the basis for restructuring of data. Certain new values, e.g. degree of oxygen saturation, are calculated and added to the records. The records are transferred to a non-formatted intermediate file and then printed out by a program which, via a conversion table, arranges the data in publishable lists.

Storage of data records

The data records are packed together and stored in a special disc storage register developed for this project. In this register the records are stored in a number of subregisters, each of which can be accessed and extended without needing to search through the entire register. A subregister consists of data of the same "origin" and "category" and from the same goegraphical area. Blocks within a subregister are all linked together through a reference in each block to the preceding and subsequent blocks in the subregister. This means that the blocks within the subregisters need not lie in a physical sequence, which is a great advantage when inserting new data in a subregister.

All contact with the storage register and administration of the list of contents is done by means of a special register handling program, which also arranges for a number of checks that the data in the register are the same as were once written into it.

Retrieval of stored data

Via a terminal optional parts of the stored data can be extracted and presented on an output data file and/or a terminal. The search is administered by a main program which utilizes the register handling program. The search profile may be of a very varying character. Search can be done on all ID variables and the search profile offers all combinations of them. Furthermore the types of analysis one wishes to retrieve can be indicated, and limitations can be set on these as well so that, for example, oxygen contents only between 2.0 and 5.0 mg/l are extracted. If any ID variable is not included in the search profile, all values of that variable are accepted.

Statistical calculations

Statistical analyses can be made on input data stored in the intermediate file and on data extracted from the register to an output data file.

Most analyses are desired for different time intervals, stations, depths and types of measured values separately. For most statistical analyses programs, therefore, there is a special "classification profile" according to which the calculations are made. The input data required are then the times, stations and depths which are to constitute the classification categories.

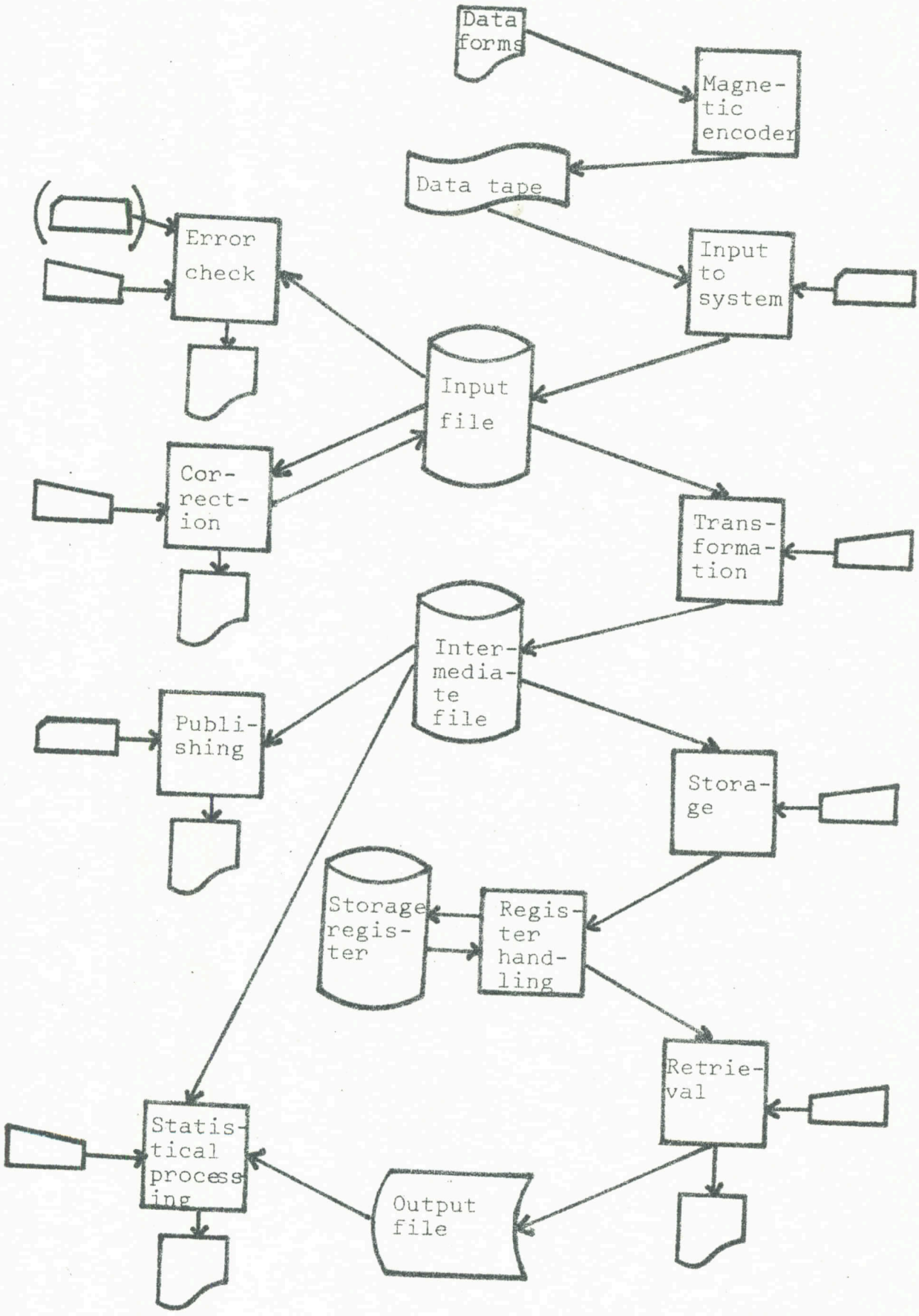
The types of analyses made within classification categories are elementary statistical calculations such as means, standard error,

maxima and minima, and trend calculations, particulary linear regression analyses. Comparisons are also made usually between classification categories, e.g. means and trends are compared by means of t-tests. The results are presented either on a terminal or, in the case of a voluminous output, on a printer.

Future development of the project

In order to obtain an effective control and understanding of the environment, many different types of data must be taken into account. Therefore work is continuing on extending the system to comprise also other types of data than physicochemical. The types of data considered are additional information on the investigated areas, such as hydrological, meteorological, sedimentological and biological data, and information on the emittents, such as position and discharge conditions.

It is our opinion that a system which integrates many different types of data will be of great assistance in, for example, work on simulation models.


Data forms
Magne-tic encoder
Data tape
Error check
Input to system
Input file
Cor-rect-ion
Trans-forma-tion
Inter-media-te file
Publi-shing
Stora-ge
Storage regis-ter
Regis-ter hand-ling
Retrie-val
Statis-tical process-ing
Output file