Environment Canada Imaging Cover Page

Report N.:

TEC-574

SKP Box Number: 672572427

CANADA

DEPARTMENT OF TRANSPORT
METEOROLOGICAL BRANCH

AUTOMATIC DATA PROCESSING

by

R. STRACHAN

TORONTO ONTARIO

CANADA - DEPARTMENT OF TRANSPORT - METEOROLOGICAL BRANCH
315 Bloor Street, West,
Toronto 5, Ontario.

AUTOMATIC DATA PROCESSING

by

R. Strachan

ABSTRACT

The role of automatic data processing in
numerical weather prediction is outlined. The general difficul-
ties of the problem are described as well as the organization of
the procedures being used at the Central Analysis Office.
Computational aspects are highlighted to illustrate how the
computer configuration dictates the techniques of data handling.

LE TRAITEMENT AUTOMATIQUE DES DONNÉES

par

R. Strachan

RÉSUMÉ

L'auteur esquisse à grands traits le rôle du
traitement automatique des données dans la prévision numérique
du temps. Il décrit les difficultés générales du problème ainsi
que l'organisation des méthodes utilisées au Service central
d'analyse. Il met en relief les divers aspects du calcul afin
d'illustrer comment la configuration des calculatrices dicte
les techniques de traitement des données.

AUTOMATIC DATA PROCESSING

by

R. Strachan

1. INTRODUCTION

General Purpose

The purpose of Automatic Data Processing, or ADP as it is commonly abbreviated, is to extract pertinent weather reports from the teletype traffic and to provide single station data in a format suitable for operational numerical weather prediction procedures. These station data are then objectively analysed into a regular grid network of values. Such transformation is necessary for the finite-differencing techniques of the prognostic time-step.

In order to permit adequate boundary conditions for the numerical forecast, the data area must be hemispheric and extend to meteorologically inactive low latitudes. The ADP program in use at the Central Analysis Office recognizes all upper air reports within a 1709-point octagonal grid network; provides unique height, temperature, dewpoint, wind speed and direction data at each of the ten mandatory pressure levels for these reporting stations; and ensures that the abstracted data have gross vertical consistency.

General Difficulty

In general, computer problems vary between two extreme classifications: (1) those which perform lengthy, rather complicated calculations on a small amount of initial data, and (11) those which perform brief, rather trivial arithmetical calculations on huge amounts of input data. The various prediction models in numerical weather prediction fall into the first classification. For example, the four-level baroclinic model executes highly re-iterative computer calculations for approximately one hour, all based on less than 7000 initial height values. Because of the iterative nature of the commands, the programs are rather small and simple to organize. Automatic Data Processing falls

primarily into the second category. Approximately one to one and a half million characters of data are processed in each synoptic period, in ten to fifteen minutes of computer time. The ADP program contains many complex-branching, decision-making commands, but not all of these commands are used on each station report.

To add to this complexity, the ADP program must be written in the basic machine language rather than in a more general language such as Fortran. These more general compilers do not as yet include logical, character-handling commands which are essential to an efficient, and effective, ADP program.

International meteorological codes change - sometimes with only short notice. As a consequence the ADP programs must be quite general and adaptable. The frequent revisions necessitated by these code changes must be introduced as modifications to the existing program rather than by rewriting the entire program.

General Solution

In order to meet the requirements for flexibility and adaptability, the ADP program cycle is organized into quite separate and distinct subprograms called subroutines. This separation is primarily on a functional basis. Functional roles are distinguished as to whether they are affected by a change in international code format, (primarily the logical recognition subroutines) or whether they deal exclusively with the data abstracted from the reports, (primarily the arithmetical code conversion subroutines and consistency checks). These latter subroutines may be further sub-divided into those which involve meteorological aspects, and those which deal exclusively with transformations which are purely computational, such as data transfers, magnetic tape movements, etc.

The separation of these functions is so complete in the operational program that individual card decks of program commands may be exchanged at run time. Thus, revisions may be confined to particular aspects of the operational program, resulting in simpler program modification and check-out technique.

Moreover, revisions may be planned simultaneously in more than one function by several persons without further extensive co-ordination.

The automatic processing of meteorological data is preceded by a preliminary operation. In this operation, the meteorological information coded in the conventional manner on paper tape is translated without manual editing into a different code on magnetic tape. This translation of codes is performed by an auxiliary computer, the buffer, without involving the principal computer, the central processor. The translation is necessary for computer calculations, and is performed as a separate operation so as not to tie up the high-speed central processor with relatively slow-speed paper tape reading. An illustration of the comparative speeds might be informative at this point. The teletype circuits have a maximum net accumulation rate of 50 characters/second as compared with a maximum optical paper tape reading rate of 500 characters/second. The transfer rate of the magnetic tapes into the main computer is 50,000 to 100,000 characters/second. Essentially then, the paper tape reader is ten times faster than the paper tape accumulation rate, and the magnetic tapes are one hundred times faster than the paper tape reader.

The second or main operation consists of the processing of all the transmitted traffic by the central processor. The input data are read from the magnetic tape provided by the first stage operation. Individual reports are recognized, and the abstracted data are stored in preassigned locations on a second magnetic tape. This second magnetic tape provides the input information for objective analysis procedures.

2. SUBROUTINES DEALING WITH METEOROLOGICAL ASPECTS

Subroutines which involve meteorological decisions have been classified as to whether they are purely arithmetical in nature, or whether they involve the logical recognition of the type of report. The purely arithmetical subroutines are iterative and process each new item of data in the same manner as the preceding item. The decision-making, logical subroutines, however, execute different sequences of operations depending upon the results of the preceding operations.

As an example of decision-making, a line of
teletype data is recognized as the start of a possible radiosonde
report if it contains a five character word beginning with either
an 85 or a 70. If this condition is not fulfilled, then the line is
dismissed from further radiosonde consideration. If this condi-
tion is fulfilled, then the report is examined for other five
character words beginning with 50, 40, 30, 25, 20, 15 and 10. If
at least three of these are recognized with no more than one of
the sequence missing between each one recognized, the report is
confirmed as a radiosonde; otherwise it is dismissed as before.
In this rather simplified illustration, not all lines are examined
for words beginning with 50, and some lines are subjected to
closer examination than others. The course of action followed in
each case is determined by the data content of the line.

Logical Subroutines

These subroutines recognize the format of the
report (radiosonde, pibal, RADAT, etc.), the data present in the
report, the station identifier, and the time of the report. The
details of the coding of these various elements is subject to
international agreement and revisions to these subroutines are
frequent. Moreover, any delay in program revision subsequent
to a code change can result in the loss of data.

Prior to 1964 the coding of meteorological reports
was implicit. With this type of code the individual elements of
wind, pressure, height, temperature etc., had to be recognized
before it could be determined that the report was a radiosonde,
or pibal. Also, the distinction between land and ship reports had
to be made frequently by trial examination of the appropriateness
of the potential quadrant, day-of-the-week identifier, and time
group. Nonpertinent messages underwent lengthy, involved
testing before it could be decided that they were not a radiosonde
or pibal.

On January 1, 1964 an explicit format type was
introduced. In this type of code a preliminary group states un-
ambiguously whether the report is a radiosonde, or pibal; and
whether it originates from a land station or from a ship. The

individual elements must still be recognized, but not necessarily
before the format type has been decided. This decision as to
format type is straight-forward, and conceivably, when all
reports are of this explicit type, messages in various formats
could be sorted as a preliminary step on a smaller auxiliary
computer. The most important consequence would be the
elimination of the trial procedures applied to non-pertinent
messages.

In many instances, ambiguities in the meteoro-
logical code still require an elaborate reductio-ad-absurdum
logic. For example in the six possible radiosonde reports that
follow, the x's represent any arbitrary number, and the under-
lined figures are indicator groups for the 850-, 700- and 500-mb
data groups.

<u>85</u>xxx	<u>70</u>xxx	xxxxx	xxxxx	xxxxx	xxxxx	<u>50</u>xxx
<u>85</u>xxx	<u>70</u>xxx	xxxxx	xxxxx	xxxxx	<u>50</u>xxx	≠50xxx
<u>85</u>xxx	<u>70</u>xxx	<u>70</u>xxx	≠0xxxx	<u>50</u>xxx	≠50xxx	≠50xxx
<u>85</u>xxx	<u>70</u>xxx	70xxx	0xxxx	<u>50</u>xxx	≠50xxx	≠50xxx
<u>85</u>xxx	<u>70</u>xxx	≠70xxx	xxxxx	<u>50</u>xxx	≠50xxx	≠50xxx
<u>85</u>xxx	<u>70</u>xxx	xxxxx	<u>50</u>xxx	≠50xxx	≠50xxx	≠50xxx

When the first word after the 85 group begins with a 70 it may
represent either a $-20^{\circ}C$ temperature at the 850-mb pressure
level (first three lines), or a 700-mb indicator group with the
850-mb temperature intentionally missing (last three lines).
Which of these two possibilities is the case can only be determined
by an examination of the format of the rest of the message. Most
of these code ambiguities result from the occasional deletion of
optional reporting groups in a message. If optional groups were
made mandatory (i. e. missing winds reported as 0////), or if
optional groups were coded in an unambiguous manner (i. e. pre-
fixed with 999 indicators), then the present elaborate logic clould
be very much simplified.

The first of these gross error checks relates the reported thickness to the reported temperatures for each mandatory pressure stratum. A computed thickness is calculated from the mean temperature for the stratum. The difference between this calculated thickness and the reported thickness is compared to a tolerance calculated for the stratum. The lowest reported height is assumed correct and the thicknesses for successive mandatory stratam are checked sequentially upward. If a thickness check exceeds the tolerance then the inconsistency is attributed to the upper level data. In this case the check is repeated on a double layer using data above and below the suspected level. A revised height may be substituted for a garbled height by substituting the calculated thickness in place of the reported lower thickness. This substitution is possible only if the double layer check is less than a calculated tolerance. Otherwise the temperature at the suspected level is also questionable, and the calculated thickness for the lower layer uncertain. Thus, a garbled temperature causes both height and temperature to be rejected at that level.

Since the calculated thickness is based upon a linear lapse rate joining the two boundary temperatures of the layer, some allowance must be made for the occurrence of significant level temperatures in the layer. Significant temperatures are statistically less likely to occur when the mean lapse rate is near either the isothermal, or the dry-adiabatic; and more likely when the lapse rate is between these values. A discrepancy between the calculated and the reported thicknesses due to significant temperatures is accepted only if it is less than a prescribed tolerance. This thickness tolerance varies with the mean lapse rate reported in a stratum. The tolerance varies linearly, and symmetrically, from minimum values for near isothermal, and near dry-adiabatic lapse rates, to a maximum value for a lapse rate of half the dry adiabatic.

A second consistency check relates wind speed with direction, and requires that strong winds have a small vertical variation of direction. Wind data for each mandatory level in a report are arbitrarily classified into one of three speed classes; light (less than 16 knots), moderate (greater than 15 but less than 31 knots), and strong (greater than 30 knots). The

direction of a strong wind must be within 30° of the direction interpolated from the two adjacent levels. A moderate wind must be within 60° of the direction interpolated except if one of the adjacent levels reports a light wind. A light wind, or a moderate wind with an adjacent light wind, does not undergo a directional consistency check. Wind data, both direction and speed, which fail these checks are discarded.

More detailed consistency checks, such as the smoothing of the wind profile, extrapolation and interpolation of the data for missing level data, have not been done for two reasons. First, the original priority was allocated to a basic operational program; and more fundamentally, it was intended that the program provide data in a raw form. Any data fitting or biasing should properly be left to the consumer (an analysis program, or research experiment).

3. SUBROUTINES DEALING WITH PURELY COMPUTATIONAL ASPECTS

In the previous section, procedures arising from meteorological assumptions or code agreements were outlined. These assumptions and agreements dictate the particular procedure to be employed, and any change necessitates a modification of the procedure. The general meteorological procedure used is applicable to any computer and computer configuration. Indeed, the preliminary ADP procedures were formulated on the IBM 650 at McGill, and some of the programs tested on the Bendix G-20 at Chalk River prior to the installation of the computer at the Central Analysis Office.

In the subroutines dealing with the computational aspects, however, consideration of the particular computer configuration is essential. These subroutines formulate the techniques for data handling used throughout the entire ADP program. In general, they are not affected by changes in code or meteorological assumptions. The translation of the meteorological reports from paper tape into computer code can be used as an illustration of these purely computational procedures. The pertinent paper tape characters are transformed into an alphanumeric character representation on magnetic tape as outlined in the accompanying table. Each number or alphabetic character in the first column is represented in the computer by the configuration of 6 'bits' listed in the second column.

0	100 000	A	000 001	K	001 011	U	010 101
1	100 001	B	000 010	L	001 100	V	010 110
2	100 010	C	000 011	M	001 101	W	010 111
3	100 011	D	000 100	N	001 110	X	011 000
4	100 100	E	000 101	O	001 111	Y	011 001
5	100 101	F	010 000	P	010 000	Z	011 010
6	100 110	G	010 001	Q	010 001	All other symbols	
7	100 111	H	001 000	R	001 000		
8	101 000	I	001 001	S	010 011		
9	101 000	J	001 010	T	010 100		

Each time a teletype space character is en-
countered on paper tape, the subsequent pertinent characters are
stored along with a flag bit in a new word of computer memory (or
magnetic tape record). Each time a line feed teletype character
is detected a second flag bit is set. Five or more consecutive
line feets characters sets both flag bits. The following message:

NNNN

.
.
.
.
.
.

THIS IS AN EXAMPLE

.

72722
12/// 85XXX

is transformed into alphanumeric format in core, or on magnetic
tape, as indicated below.

Line feed flag	Space flag	Five "six bit" characters					
1	0	000 000	N 001 110	N 001 110	N 001 110	N 001 110	First computer word in core
1	1	000 000	T 010 100	H 001 000	I 001 001	S 010 011	Second " " " "
0	1	000 000	000 000	000 000	I 001 001	S 010 011	Third " " " "
0	1	000 000	000 000	000 000	A 000 001	N 001 110	Fourth " " " "
0	1	E 000 101	X 011 000	A 000 001	M 001 101	P 010 000	Fifth " " " "
0	0	000 000	000 000	000 000	L 001 100	E 000 101	Sixth " " " "
1	0	7 100 111	2 100 010	7 100 111	2 100 010	2 100 010	Seventh " " " "
1	0	1 100 001	2 100 010	011 111	011 111	011 111	Eighth " " " "
0	1	8 101 000	5 100 101	X 011 000	X 011 000	X 011 000	Ninth " " " "

It will be obvious that the particular technique of data storage must vary from computer to computer, since not all computers have 32-bit words, nor flag bits, etc. However, some procedure of both numeric and alphabetic character representation must be employed in any computer.

It will also be appreciated that the alphanumeric character representation of all data must be transformed into a purely numeric form before any arithmetic operations can be

performed. That is, 127 in alphanumeric characters is 100 001, 100 010, 100 111 and must be converted to 1111111 in binary in order to simplify the 'carry' operation of arithmetic.

Comparison of the magnetic tape sorting procedures employed in ADP at Suitland, Offut and Dorval underlines the primary role of tape hardward or other auxiliary storage in determining the nature of the program. It is a meteorological requirement that all data be represented unambiguously and only once for objective analysis. This may be accomplished by merging the data from all messages from a single station after ADP, or during ADP. Because of the addressable magnetic tape feature of the Central Analysis Office computer installation, the necessary tape movements are made while the data are being processed and an appreciable saving of overall processing time results. Other magnetic tapes do not possess this addressable tape feature and consequently the merging of station reports into a single report for each station must be performed after ADP.

4. CONCLUSION

The procedures described in this paper and summarized in a block diagram in the appendix cannot be considered the ultimate in automatic data processing. Program deficiences must be corrected when detected. Changes in code format will introduce program modifications. Long-term changes in international data exchange agreements will surely require greater centralization in national and regional ADP. It is believed that the present ADP program has the flexibility to handle the short-term modifications, and that sufficient experience will be gained with this program to plan and implement a more highly centralized data exchange.
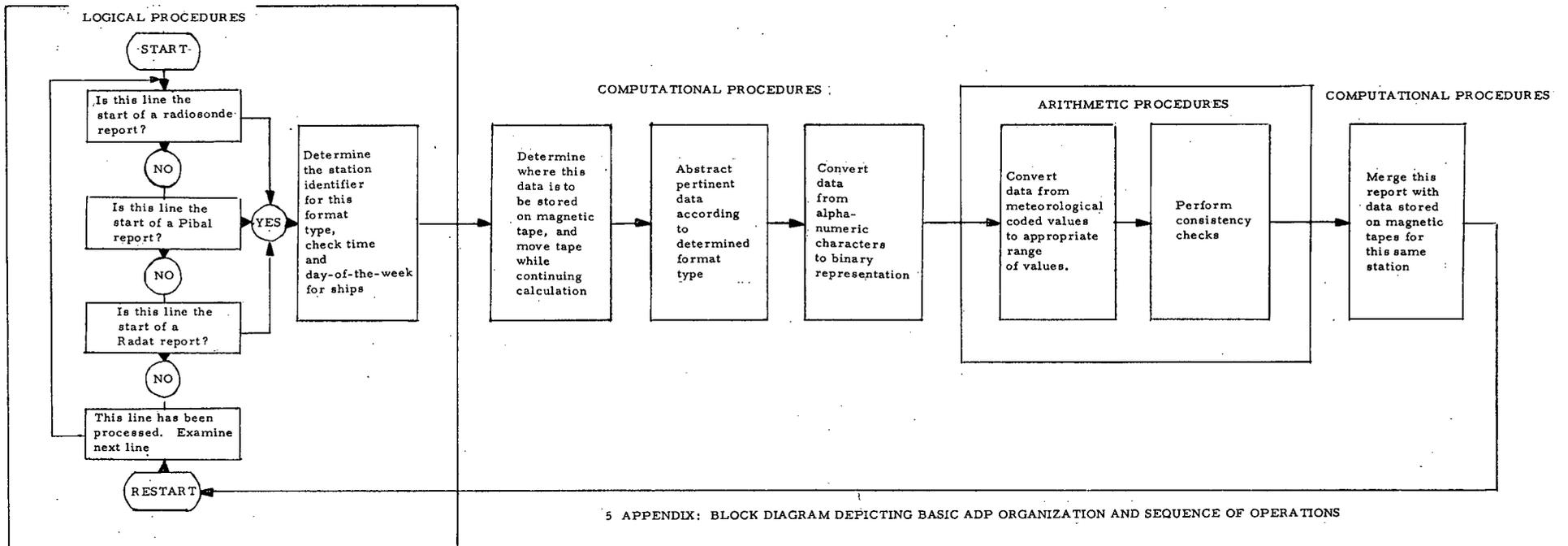
APPROVED,

J. R. H. Noble,
Director,
Meteorological Branch.

7. REFERENCES

(1) Cressman, G. P. and H. A. Bedient, 1957: "An Experiment in Automatic Data Processing." Mon. Wea. Rev., Vol. 85, No. 10, pp. 333-340.

(2) Kwizak, M., et al, 1960: "Application of Electronic Computers to Provision of Operational Meteorological Services in the Meteorological Branch," Queens Printer, p. 107.

(3) Strachan, R. A., 1962: "Automatic Data Processing for Numerical Weather Prediction". Proceedings of the Third Conference of the Computing and Data Processing Society of Canada, University of Toronto Press, pp. 78-88.

LOGICAL PROCEDURES

START

Is this line the
start of a radiosonde
report?

NO

Is this line the
start of a Pibal
report?

NO

YES

Is this line the
start of a
Radat report?

NO

This line has been
processed.   Examine
next line

RESTART

Determine
the station
identifier
for this
format
type,
check time
and
day-of-the-week
for ships

COMPUTATIONAL PROCEDURES

Determine
where this
data is to
be stored
on magnetic
tape, and
move tape
while
continuing
calculation

Abstract
pertinent
data
according
to
determined
format
type

Convert
data
from
alpha-
numeric
characters
to binary
representation

ARITHMETIC PROCEDURES

Convert
data from
meteorological
coded values
to appropriate
range
of values.

Perform
consistency
checks

COMPUTATIONAL PROCEDURES

Merge this
report with
data stored
on magnetic
tapes for
this same
station

5 APPENDIX: BLOCK DIAGRAM DEPICTING BASIC ADP ORGANIZATION AND SEQUENCE OF OPERATIONS

CIR-4262   UDC: 551.509.313
TEC-574
29 June 65.

CANADA
Department of Transport - Meteorological Branch
315 Bloor St., W., - Toronto 5, Ontario

Automatic Data Processing
by R. Strachan

11 pps.   1 fig.   3 refs.

Subject reference: 1. Numerical Weather Prediction

ABSTRACT: The role of automatic data processing in numerical weather prediction is outlined. The general difficulties of the problem are described as well as the organization of the procedures being used at the Central Analysis Office. Computational aspects are highlighted to illustrate how the computer configuration dictates the techniques of data handling.