CANADA

DEPARTMENT OF TRANSPORT
METEOROLOGICAL BRANCH

# Modern Techniques in Canadian Climatic Data Processing

By

B.S.V. CUDBIRD

CANADA

**DEPARTMENT OF TRANSPORT**

**METEOROLOGICAL BRANCH**

# Modern Techniques in Canadian Climatic Data Processing

By

B.S.V. CUDBIRD

UDC 518.5 : 551. 506

CLI 2-68

January 30, 1968

# METEOROLOGICAL BRANCH – DEPARTMENT OF TRANSPORT – CANADA

## MODERN TECHNIQUES IN CANADIAN CLIMATIC DATA PROCESSING

By

*B.S.V. CUDBIRD*

### INTRODUCTION:

A basic objective in climatology is to make climatological information available in the form most appropriate for decision-making. This involves the collection, quality control, processing and distribution of climatological data, and frequently supplying consultation and advice on the application of these data.

Previous to 1950 all processing of climatological data was done by clerical means in the Climatology Division of the Canadian Meteorological Service. However in the late 1940's it was recognized that in view of the increasing volume of data being accumulated and of the more extensive and exacting requirements for these data, it was necessary to embark on a program of machine processing. Consequently the transfer of climatological data to punched cards for machine processing was begun in 1950.

There has been a steady expansion of both staff and equipment facilities in the Machine Processing Section since the modest beginning 17 years ago. During the 1950 to 1965 period specially modified unit record type equipment was utilized for all quality control and processing operations. However, in 1965, most of this equipment was replaced by the installation of an electronic computer.

The climatological archives of the Canadian Meteorological Service now contain more than 60,000,000 punched cards. This great resource for scientific research and economic planning has been generated and processed and is being exploited in a variety of ways. Present plans to convert the archives to magnetic tape represent a substantial procedural change and further changes will no doubt be introduced as new devices for mass data storage are developed.

While the basic objectives of climatological processing remain unchanged, technological developments promise new applications for exploiting the climatological archives. The development of new devices for consulting large files of data in a minimum of time will facilitate the utilization of the climatological archives for planning purposes and, at the same time, provide useful operating tools for the practicing meteorologists.

The discussion which follows presents a review of past experiences in machine processing of Canadian climatological data, an evaluation of the present methods and procedures, and an outline of plans for the future.

Data processing by machine methods in the Climatology Division has always followed the same basic system in spite of modifications and refinements over the years. The demand for the creation, preservation and exploitation of an accurate historical collection of basic weather data remains paramount.

There are four broad areas of activity in a data processing system:

1. Generation. 2. Quality control. 3. Processing. 4. Storage or archiving.

## GENERATION

The 80-column punched card is the basic input unit for machine processing climatological data in Canada. To generate the punched cards the Canadian Meteorological Service employs both manual and automatic methods.

Most cards are punched manually. Punching at Meteorological Branch Headquarters produces by far the greatest number of cards, presently about 300,000 cards each month. Begun in 1950, the Headquarters' punching unit has used types 016, 024, 026 and 029 IBM key punches to produce the 17 different card types displayed in Appendix III. Types 056 and 059 verifiers are also used in the punching unit for verification of certain fields in the various card types. The central card punching unit has a staff of 22.

A field punching program using IBM type 001 hand punches was begun in Canada in 1951. Under this program the personnel at approximately 60 principal meteorological stations transfer data from the hourly weather reports to the #1 cards and forward the punched cards along with the original records to Meteorological Branch Headquarters for processing at the end of each month. Fifty thousand cards per month are produced in this field punching program.

Automatic generation of punched cards for climatological purposes has been largely restricted to experimental projects where a tape-to-card punch, IBM type 047, is used to create cards from paper tape. Fischer-Porter automatic rain gauges have been installed at a number of locations in Canada. These gauges produce an incremental record of precipitation on 16-channel paper tape. Using a translating device and a type 047 tape-to- card punch, this information is converted to punched cards (#3 card) for processing.

Automatic data generating equipment has also been used in microclimatological projects. An example of this is the Forest Meteorology Project in which observational data from 96 sensors are recorded every five minutes when the project is operating. This information is recorded on 5-channel paper tape and is converted directly to punched cards using a type 047 tape-to-card punch.

## QUALITY CONTROL:

The checking of data for accuracy and completeness before publication and use in analytical studies is an essential task in every climatological service. With the advent, in 1950 of machine processing in the Canadian Service, the punched card became available for editing and quality control of the data. Initially the unit record equipment available in the Climatology Division including sorters, collators and tabulators, was used to check the accuracy and completeness of the transcription to punched cards of the observed data. It soon became apparent that the methods used to check the accuracy of transcription could be adapted to check the accuracy and consistency of the observation itself. Therefore a "feed back" program was initiated whereby the punched cards containing observing or reporting inconsistencies were sent back to the observing stations to provide information to the weather observers regarding the quality of their observing and the legibility of their records.

The Canadian system of quality control of the archives file of current climatological data developed gradually. Procedures and methods became more thorough and efficient as more sophisticated unit record equipment such as the type 101 electronic statistical machine and the calculating punch became available. Finally, 1,000 cards per minute editing became possible with the introduction of the type 108 card proving machine. The types 101 and 108 were the main quality control devices used in the Climatology Division for a decade. Each new card type made its own particular demands on the system for checking and correction.

With the introduction of the IBM system 360 computer with disc drives, a new philosophy of quality control had to be developed. With this equipment the data from the punched cards are transferred to a disc pack for quality control and processing. The disc pack is basically a data storage device used in conjunction with the computer. The pack is mounted on a disc drive which contains the electronic and mechanical elements that transfer the pack's information impulses to the computer. Data are read from the surface disc by a set of magnetic heads that move across the coated area as the disc pack spins. The disc, being a random access storage device, has significant advantages for editing and processing climatological data.

## PROCESSING – PAST METHODS – UNIT RECORD EQUIPMENT:

In this report "processing" of climatological data means the actual machine operations required for the preparation of summaries, frequency distributions or tabulations. Routine tabulations and reports are produced on a regular monthly basis after the quality control of the data has been completed and the card file has been corrected. Special projects are scheduled in available time.

During the period 1954-1965, tabulators, types 402 and 407, were used as well as electronic statistical machines, type 101, for printed output. The type 407 tabulator is an efficient printing device and performed as the chief printing output machine for the climatological publications from 1954 through 1965. The type 101 is one of the most efficient unit record devices available for processing climatological data. Its ability to output distributed counts to 240 print positions in 60 classes is still not equalled by any computer configuration. The ESM type 101 was used extensively for processing data for special projects in the Climatology Division during the past 10 years, and at one time four of these machines were in service. The last type 101 electronic statistical machine was released in the spring of 1967.

One of the main objectives of machine processing of climatological data is the acquisition of data in a machine processible medium for special analyses and research projects. By July 1, 1967, the punched card archives in the Climatology Division contained 60,000,000 cards. Unfortunately limitations on staff and equipment have restricted the exploitation of these archives for special projects. However, by 1964, the number and complexity of special machine projects had increased to the point where it was considered that the installation of a medium sized computer was justified.

## PROCESSING – PRESENT METHODS – CENTRAL COMPUTER:

Early in 1964 a survey was made of available medium sized computing systems available in Canada. A disc-oriented IBM system 360 model 30 computer was selected from among six proposals submitted by computer manufacturers. As delivery of this computer was not scheduled until September 1966, arrangements were made for the installation of a disc-oriented IBM 1440 computer in August 1965. Although the 1440 computer proved to be incapable of carrying out the full quality control and processing routines required, it did provide valuable training and experience for the staff. The transition from unit record operations to stored program operations was made gradually over a two-year period and by the time the system 360 computer was installed in September 1966, both programming and operating personnel were able to make the conversion without a break in the routine processing operations.

At the present time all quality control and processing for routine operations and analyses for special projects are carried out on the computer configuration detailed in Appendix I. The punched cards created in the field or centrally at Headquarters are read onto the disc

files via the type 2501 card reader or type 1442 card read punch. The quality control programs written for each card or file type are then brought from their residence on the first disc drive (2311–1), prepared for execution and placed in the core storage of the central processor (2030).

Quality control of the card file is so programmed that printouts produced by the type 1443 printer can be reviewed by Meteorological Technicians for error. Replacement cards are produced where either punching or observing errors are detected and corrected by the reviewers. The corrected cards are then read onto the disc file under control of a correction-write program. Since the original file is written on a concentric series of tracks on the disc, the whole of which is rotated at 1500 rpm, every record or observation is available for processing every 25 m/sec. Moreover, every record is addressable in random order. Thus, no matter which record in the file needs correcting it can be found (accessed) in a minimum of time and returned corrected to its original position on the disc file. Quality control and correcting of climatological files is very efficient using such direct access devices as discs.

The disc packs are also very efficient for sequential processing and, therefore, the the discs are used to prepare the tabulations used in the routine climatological publications of the Meteorological Branch. The type 1443 printer, though relatively slow, produces high quality print output and meets the standards necessary for photographic reduction of the tabulations for printing.

At the present time punched cards are the basic input medium through the 2501 card reader for special project work on the computer. Unfortunately the time involved in reading these cards onto the disc is substantial, involving 100 minutes for the transfer of 10 years of hourly data from card to disc. It is, therefore, desirable to retain as much archival data as possible on standby disc packs. The high cost of the disc packs is a limiting factor in this type of operation but with suitable organization and management, it has become possible to store about 180,000 records on a single 1316 disc pack. A single disc pack may thus contain 20 years of hourly records from a single station or two stations for a decade.

Processing speeds vary when running projects on a disc and depends upon programming complexity and depth of record. However some idea of the processing speeds may be obtained when it is noted that it takes about nine minutes to process a station decade (88,000 records) on a disc and print out a diurnal frequency distribution for any desired parameter.

## ARCHIVES:

The punched card archives in the Climatology Division contain more than 60,000,000 cards. Appendix II provides a listing of the volume and depth of these archives by card type.

In addition to the card archives the Canadian Meteorological Service now has an interim electronic archive of 40-station decades on disc packs. This "spinning archive" provides quick response whenever requests are received for information regarding hourly data available on these discs.

A library of 9-level magnetic tape records is also being built up through a conversion program from punched cards to magnetic tape. This tape library is currently quite small and contains tape records of hourly data for only British Columbia and the Yukon Territory for the period 1957–1966 inclusive.

## FUTURE PLANS:

### Data Generation:

There has been little real progress during the past ten years in improving methods of data generation. Automatic data generation of climatological information has so far proved disappointing. When the data are generated at the observational site the electronic or mechanical generator seldom has remained operational long enough to provide information of real climatological value. On the other hand, data transmitted over the telecommunications circuits have not met the rigid standards required for climatological analysis as transmission errors of appreciable proportions do occur. An improved telecommunications system will provide more reliable transmission of data for climatological purposes in future. Optical scanning which appeared to offer considerable promise a few years ago, now appears to be losing its appeal to systems designers.

Present indications are that for some time to come the 80-column punched card will continue to be the primary input into computing systems for quality control and processing of climatological data. However, there is no longer an absolute need for the unit record concept. Thus data may be created as "information strings" on a punched card or on a magnetic tape as a "spread item record". This procedure is feasible since the present generation of computers can utilize information in this form for processing on disc or in core. In addition it is expected that generalized input cards can be planned in such a form that the computer can automatically produce output records in the form required.

## QUALITY CONTROL:

Future planning for the quality control function should include gradual assumption by the computer of the editing and correction functions presently performed manually by technician and clerical staff. More sophisticated techniques involving inter-file comparisons, adoption of statistical filter techniques and other advanced mathematical tests must be developed.

## PROCESSING:

The new generation of computer systems has confirmed the replaceable disc pack file as an efficient low cost processing medium. In fact, the disc has emerged as the chief functional component of the latest computers and is likely to play an increasing role in routine processing of climatological data. Plans are now underway to make greater use of disc files which may be updated, randomly accessed and sequentially processed.

The Meteorological Service of Canada holds in its climatological archives the only extensive climatological data bank in the country. The exploitation of the data contained on these 60 million punched cards presents a tremendous challenge to those responsible for meeting the needs of government, industry, aviation, resources development, etc., to provide climatological statistics for planning and development. Arrangements must be made at an early date to satisfy user requirements by supplying copies of these data or by undertaking computer projects utilizing the data available in machine processible form.

## ARCHIVING:

Planning for permanently storing climatological data in the archives in a medium that is both secure and machine processible is quite flexible at the present time. Fosdic microfilming

of punched card files provides a permanent storage medium but the slow progress being made in the area of magnetic tape retrieval from microfilm is disappointing. On the other hand, the wholesale commitment of data on punched cards to magnetic tape, with its low security and high housekeeping costs, must be viewed with serious reservations.

The new generation of computers provides at least a partial solution to the problems associated with card storage. Since most processing is now done by computer it is no longer necessary to rigidly adhere to the unit record concept in card design. Certain types of cards may be designed to contain a number of unit records. For example, the #8 upper air card can be redesigned to contain not one level but five levels per card. Such composite cards are known as "archives spread item cards" or ASIC cards. A further compaction of the data may be accomplished by using packed decimal coding. By adopting these techniques it is possible to achieve an immediate reduction in card accession but the problem of card deterioration with time remains.

The success of disc processing in the Canadian Meteorological Service is leading to the investigation of other devices of the direct access family. Despite the fact that many files on disc may be read thousands of times without the loss of a bit or byte of information, the ultimate medium still appears to be some kind of microfilm – a film optical scanning device for input to computers. Hopefully there may be a major breakthrough in this area within the next few years. As of today a complete economic solution to the problem of archival security with machine processibility remains elusive. Meanwhile, the need to provide better service to the science of meteorology and to the national economy increases.

# MODERN TECHNIQUES OF DATA PROCESSING

## APPENDIX I

## A LIST OF IBM MACHINES CURRENTLY IN USE IN THE DATA PROCESSING SECTION OF THE CLIMATOLOGY DIVISION

**Type 024 – Numeric Key Punch**

This is the basic card punch of the Section. A skilled operator will produce 1200 No. 1 cards daily on this machine.

**Type 029 – Card Punch Model A**

Printing Card Punch which punches 64 characters in 360 punching code.

**Type 047 – Tape to Card Punch**

A machine by which 5 channel teletype tape is converted automatically to IBM cards. A development program to convert synoptic reports (1200Z) is now in progress.

**Type 056 – Card Verifier**

A machine by which the punching done on the Type 024 equipment is verified. A card which has been verified will contain a semi-circular nick on its upper right edge. A skilled operator will verify about 4,000 cards per day.

**Type 077 – Collator**

A device by which two files of cards may be compared in certain programming fields and interfiled (collated) on the basis of the comparison. Its operating speed is 240 cards per minute.

**Type 083 – Sorter**

The basic processing machine of the IBM card system. On this machine IBM cards are sorted into pockets according to digit punching in a column. Only one column at a time may be sorted. Operating speed is 1000 cards per minute.

**Type 108 – Electronic Card Proving Machine**

A transistorized device with core counters and increased editing capacity. This machine, has undertaken most of the quality control processes but this is now being taken over by the 360 system. Operating speed is 1,000 cards per minute.

**Type 514 – Reproducing Punch**

A device for reproducing and gang punching IBM cards. Speed is 100 cards per minute.

**Type 552 – Interpreter**

A machine which types along the upper edge of an IBM card the numeric or alphabetical equivalent of the punching in that card. Operating speed is 60 cards per minute.

**Type 867 – Output Typewriter**

A typewriter attached to the Type 108 to provide print out from its core counters. Operating speed is 1000 characters per minute.

**Type 360 – D30 Computer Configuration consists of:**

| | | |
|---|---|---|
| 1 | 2030 | Central processing unit with 16000 bytes of core storage. The cycle (access) time of this computer is 1.5 micro-seconds per byte. |
| 1 | 1052 | Console typewriter. This device provides for operator – system communication via a keyboard. |
| 1 | 2841 | Storage control unit. This device provides the interface between all direct access devices and the central processing unit. |
| 2 | 2311 | Disk drives. These direct access devices rotate at 1500 rpm and contain demountable disk packs of information. Each pack contains $7.25 \times 10^6$ bytes of data. |
| 1 | 6960 | Selector channel. This device provides a direct path between high speed input/output media (disks, drums, tapes etc) and the central processing unit. |
| 1 | 2415 | Tape unit. This compact device contains its own control unit and has two drives capable of reading or writing information (800 bytes per inch) at the rate of 15000 bytes per second. |
| 1 | 1443 | Printer. This device outputs printed lines at a rate dependent upon the character set installed. A 13 character set will print at 600 lines per minute; a 52 character set will print at 240 lines per minute. |
| 1 | 1442 | Card/read punch. This device reads information in IBM cards at 400 C.P.M. and punches IBM cards at a rate dependant upon the no. of columns punched. (about 90 cards per minute for 80 columns punched) |
| 1 | 2501 | This device is a high speed serial card reader using fiber optics as the reading medium. It reads information in IBM cards at the rate of 1000 cards per minute. |

## APPENDIX 2
## PUNCHED CARD ARCHIVES
## JANUARY 1, 1968.

**Hourly Surface Observations** – Card #1

Total volume in archives: 26,000,000.

Cards are punched for each regularly scheduled observation and are available for all synoptic and hourly observing stations from January 1950. Data from 9 selected stations have been punched back to 1947.

**Surface Synoptic Data** – Card #2

Total volume in archives: 90,000.

Contains coded surface synoptic data for 245 stations and these data are retained in the archives for a period of 2 years only.

**Recording Precipitation Gauge Data** – Card #3

Total volume in archives: 1,400,000.

These data are punched from recordings of the MSC tipping bucket rain gauge, and include hourly daily and monthly totals, intensities and dates. Data are now available from over 300 stations. Punched card data commenced in 1960 from 106 stations.

**Summary for the Climatological Day** – Card #4

Total volume in archives: 23,000,000.

Data are available from 2400 first order and climatological stations, 600 of which have a record of 30 years and 111 stations with observations covering a period of 60 years.

**Upper Air Standard Pressure Level Data** – Card #5

Total volume in archives 3,250,000.

These punched cards contain data for regularly scheduled upper air observations beginning in January 1954, at each of the standard pressure surfaces from 35 radiosonde stations.

**Upper Wind** – Card #6

Total volume in archives: 1,170,000.

Direction and speed data currently produced from 35 rawinsonde and 10 pibal stations and these data have been placed on punched cards from 1955.

**Radiosonde Special Data** – Card #7

Total volume in archives: 260,000.

This card contains special radiosonde data from 35 stations and are available on punched cards from 1954.

**Signigicant Level Data** – Card #8

Total volume in archives: 3,500,000.

Pressure, temperature, relative humidity and height of each significant level are punched in this card type from 35 reporting stations from 1957, for each regularly scheduled observation.

**Marine** – Card #9

Total volume in archives: 600,000.

The card format for the marine surface data punched cards conforms with the WMO requirements introduced in 1962. Coded marine observations have been transferred to punched cards since 1957 for selected ocean and Great Lakes Ships.

**Bright Sunshine** – Card #10

Total volume in archives: 800,000.

This card is a combined hourly and summary for the day of sunshine observations from the Cambell Stokes recorder and presently being punched for 210 stations. Data in the archives extend back to 1959.

**Hourly Radiation** – Card #11

Total volume in archives: 1,200,000.

Hourly radiation values placed on punched cards since 1959 for 28 stations. Net all wave radiation (RF #4) is reported from 9 stations.

**Soil Temperature** – Card #12

Total volume in archives: 50,000.

The volume of this card type has been considerably reduced because of the conversion to the new ASIC format. Soil temperatures for A.M.-P.M. observations at all levels are now on one card per day instead of, as previously one card per level. Data from 42 stations are available from 1958.

**Daily Evaporation** – Card # 13

Total volume in archives: 30,000.

Data punched includes net water loss from pan, daily wind mileage mean air temperature and computed lake evaporation, and readings are obtained from 73 stations beginning in 1962.

**Air Pollution** – Cards #14 & 14A

Total volume in archives: 950,000.

Air pollution data consists of the national Air Pollution card #14 which comprises fixed level data and the municipal Air Pollution card 14A which is data from movable surface stations. Hourly readings are available on punched cards from 67 stations beginning in 1960.

**Hourly Wind Data** – Card #15

Total volume in archives: 950,000.

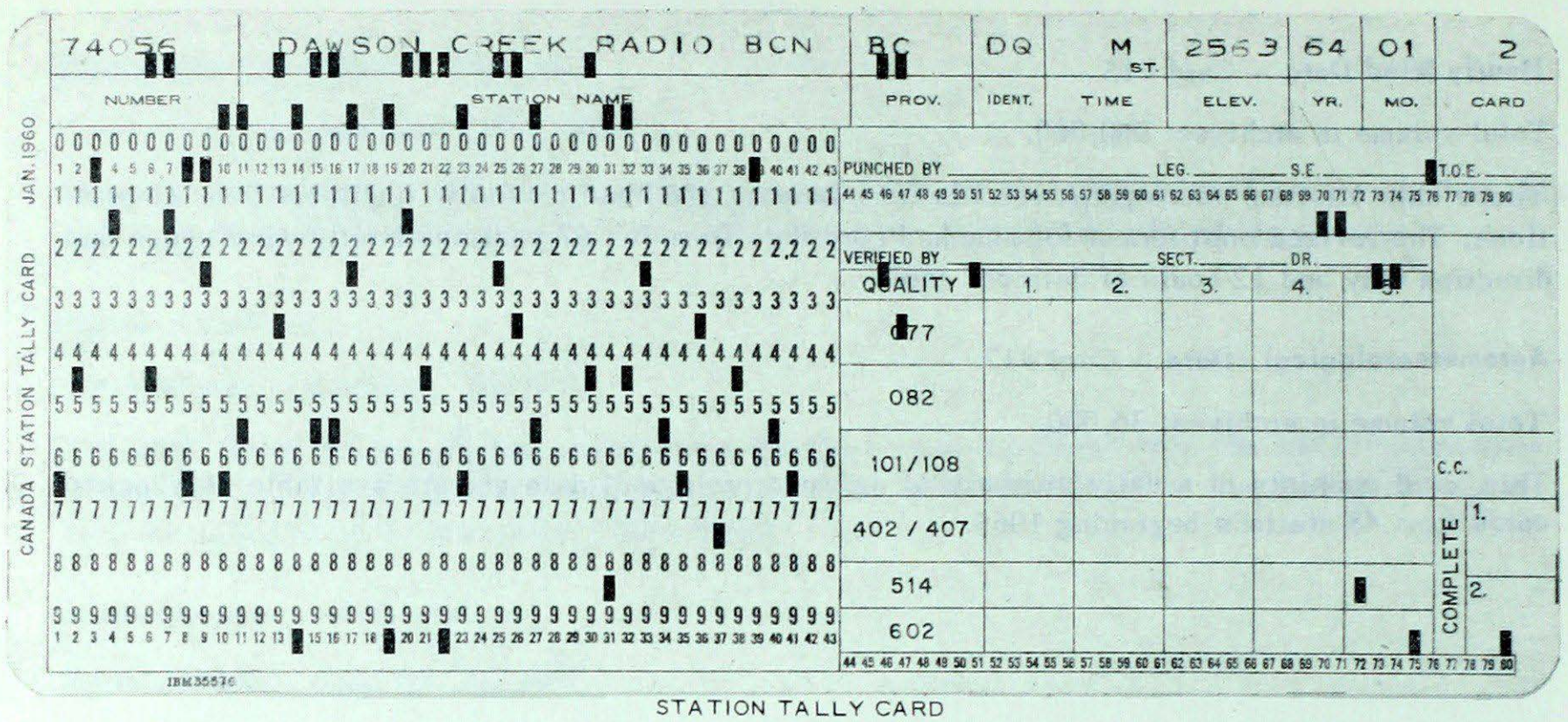
These data are from anemograph records from January 1963 and presently available from 214 stations. The revised card format for punched card data from Jan 67 contains hourly wind speed and direction only and 12 hours of data per card.

**Agrometeorological Data** – Card #17

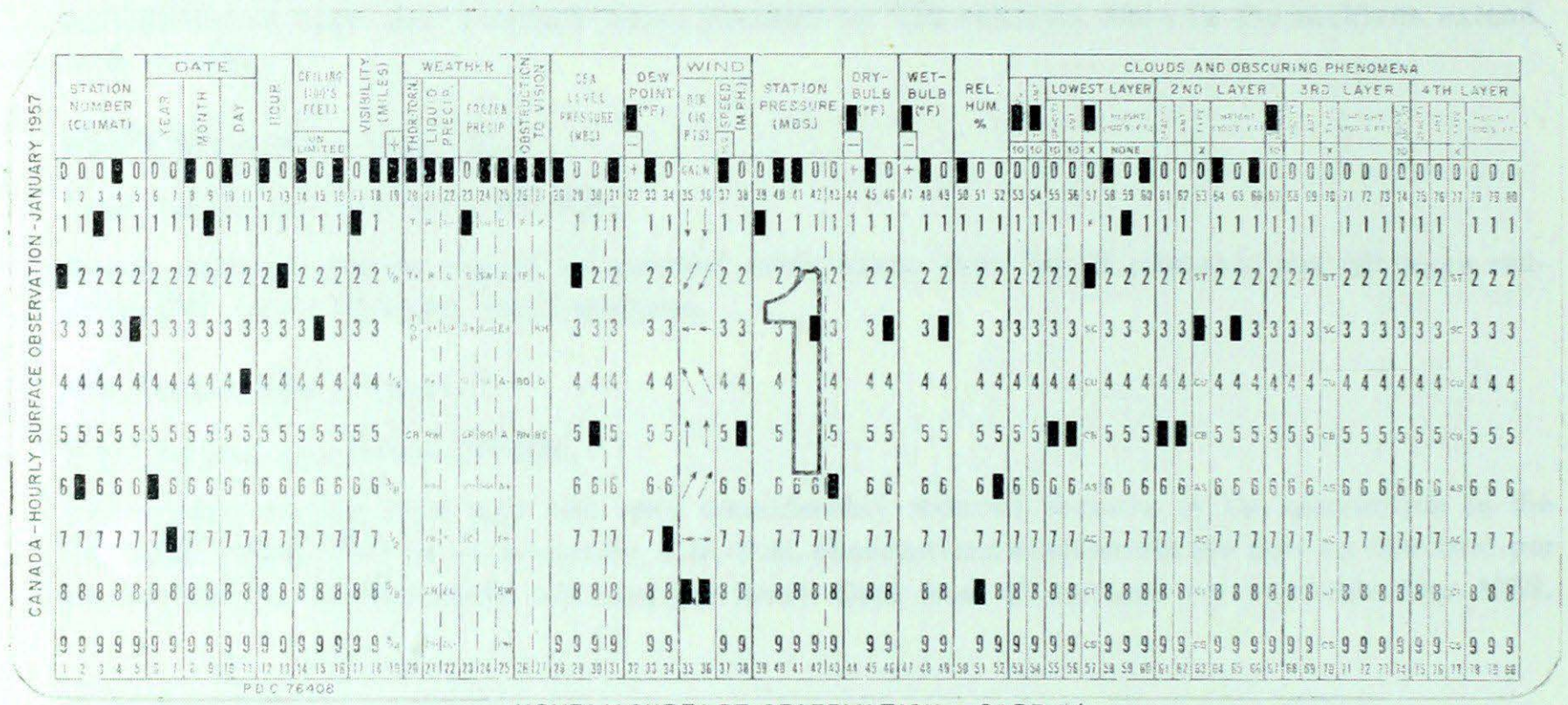
Total volume in archives: 30,000.

This card consists of a daily summary of agrometeorological data and are available on punched cards from 48 stations beginning 1965.

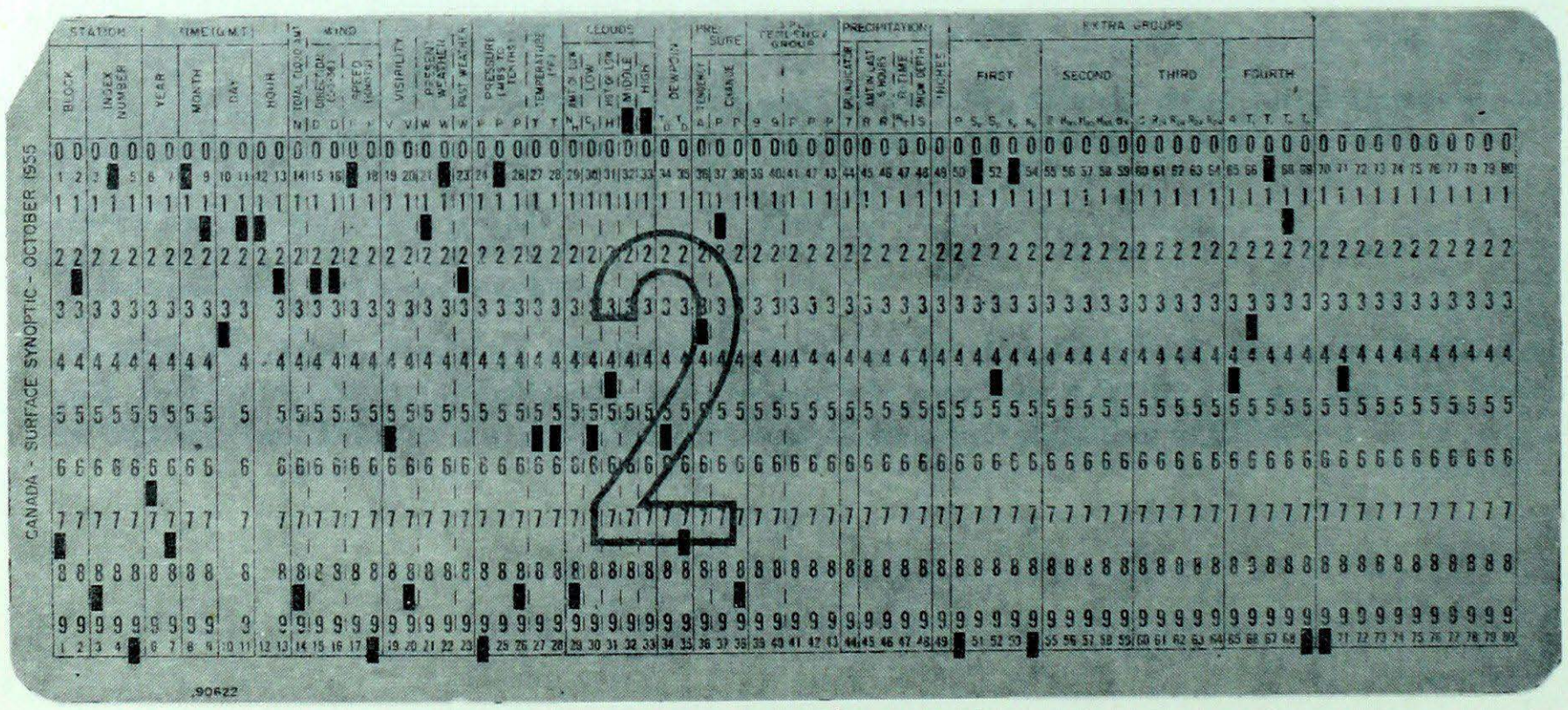
# APPENDIX 3
## PUNCHED CARD TYPES IN THE CLIMATOLOGY DIVISION ARCHIVES



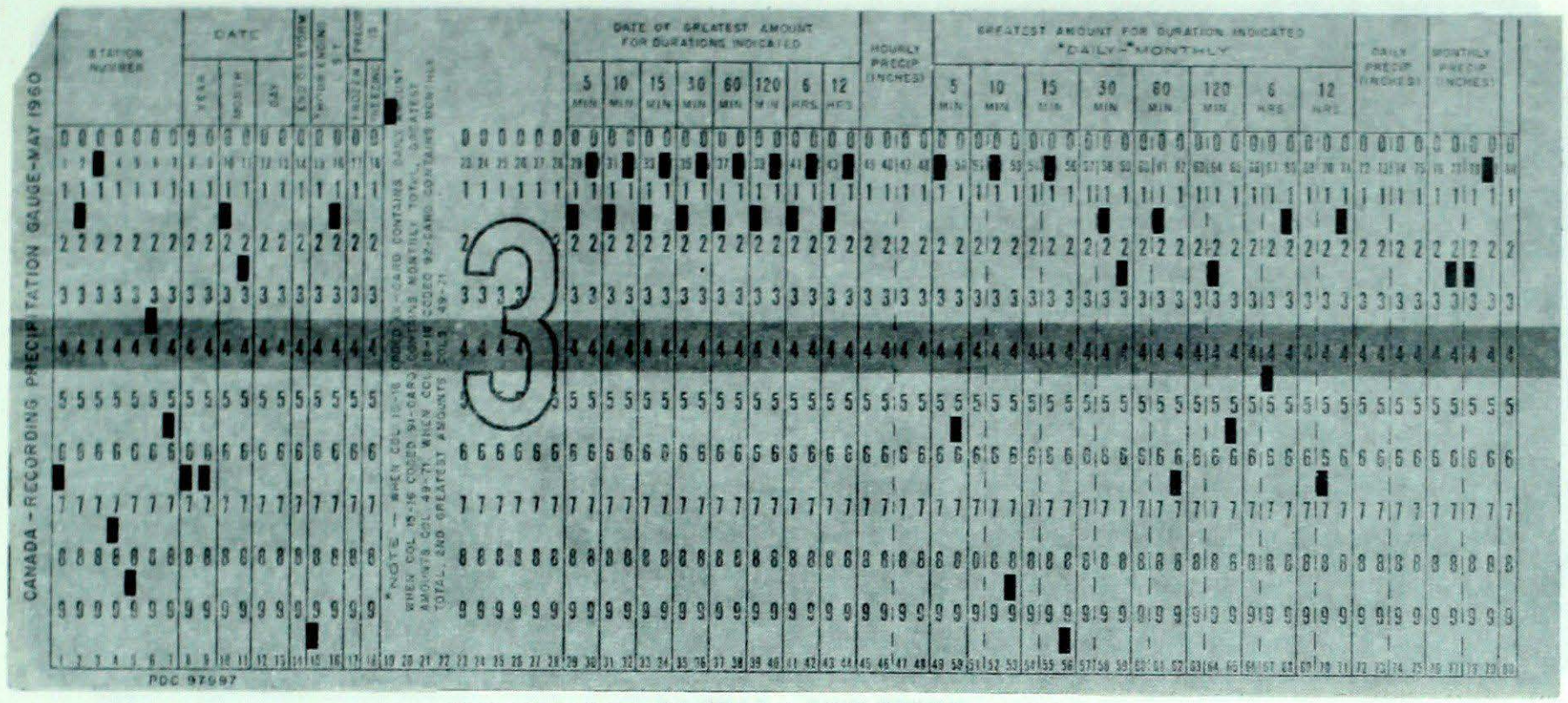

STATION TALLY CARD



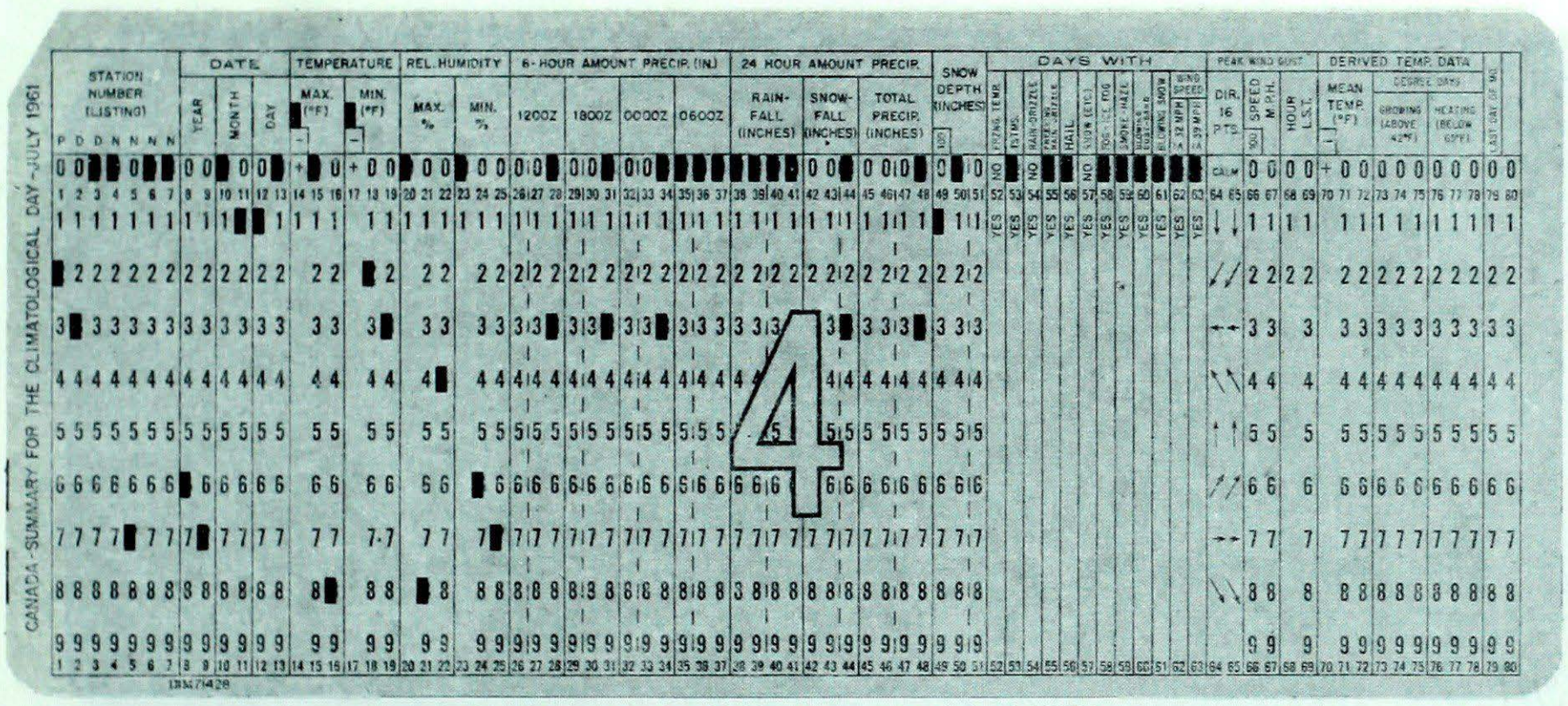

HOURLY SURFACE OBSERVATION – CARD # 1



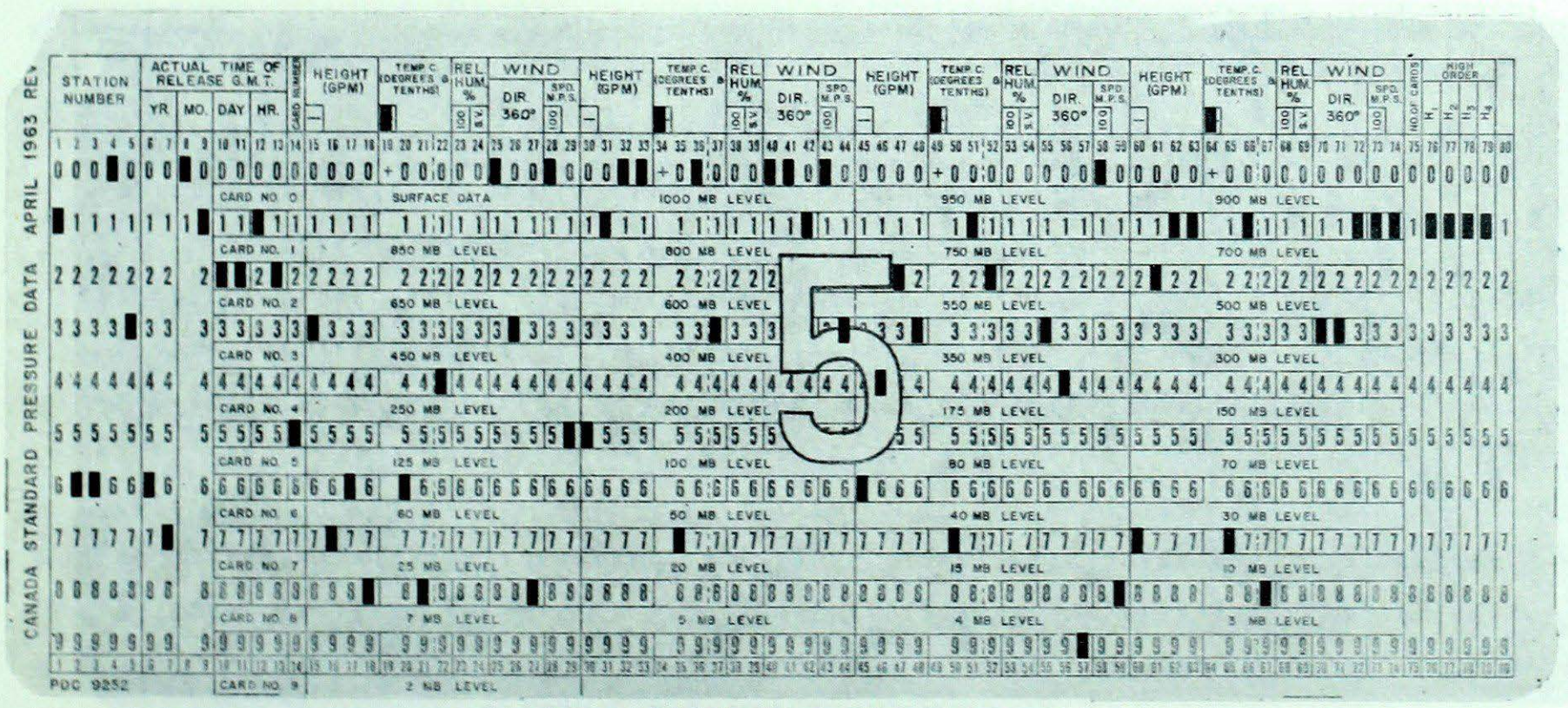

SURFACE SYNOPTIC – CARD # 2



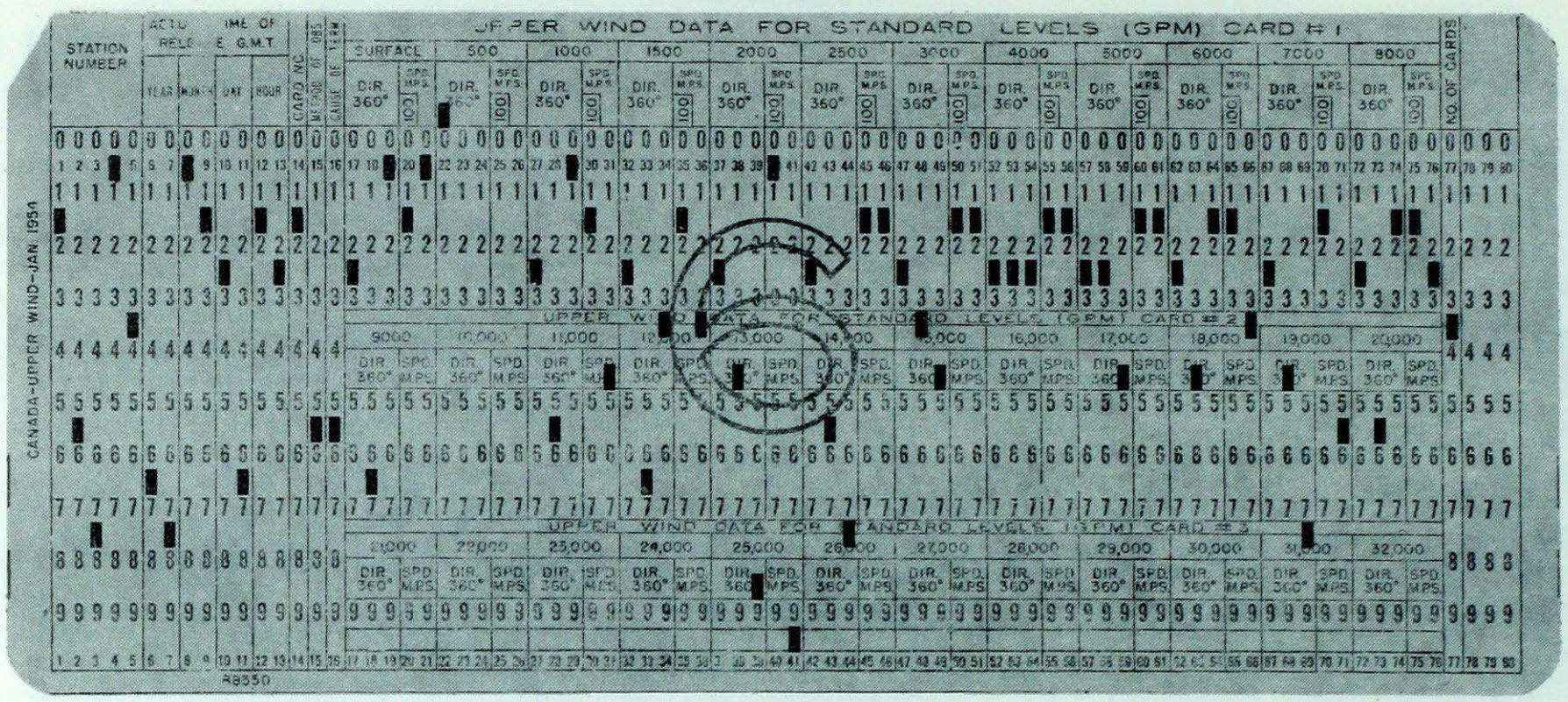

RECORDING PRECIPITATION GAUGE – CARD #3



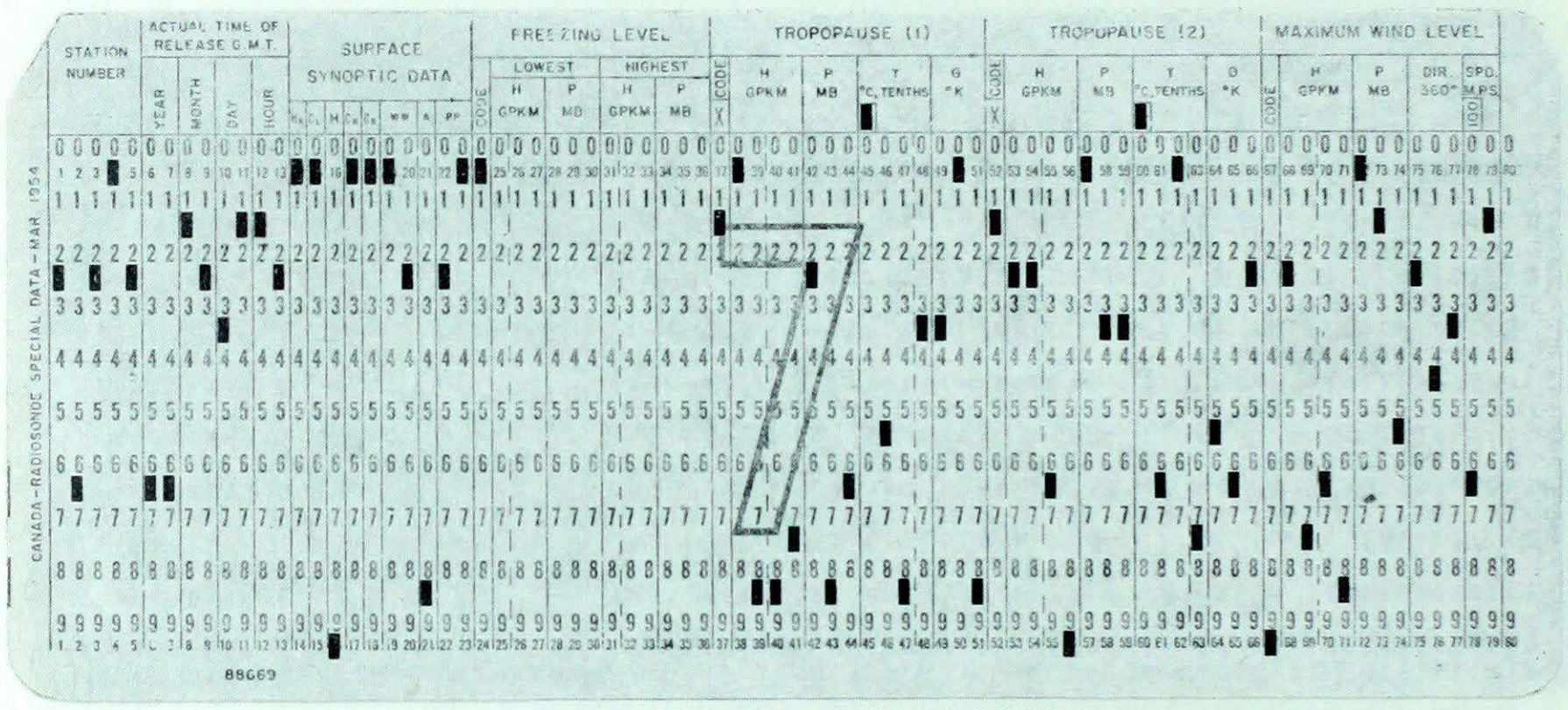

SUMMARY FOR THE CLIMATOLOGICAL DAY – CARD #4



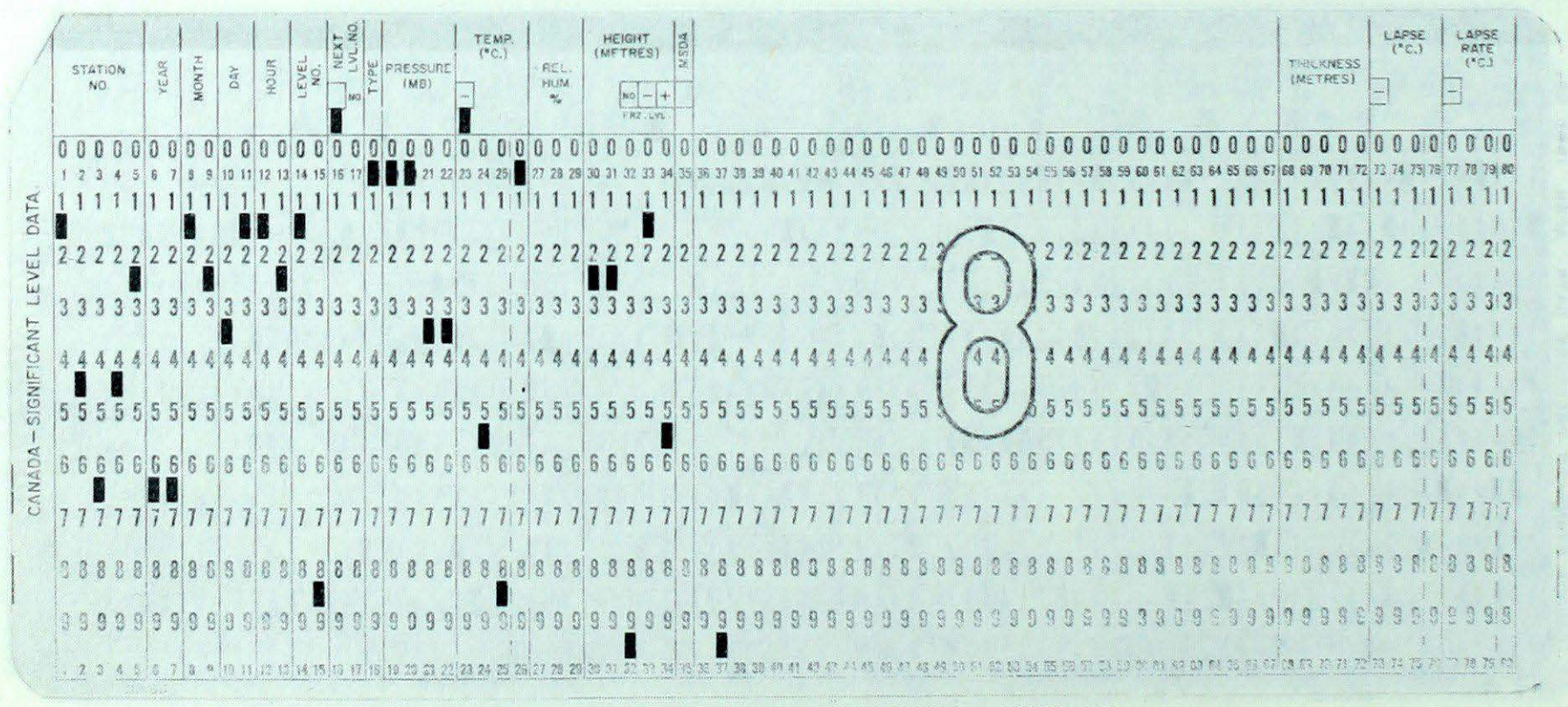

STANDARD PRESSURE DATA – CARD #5



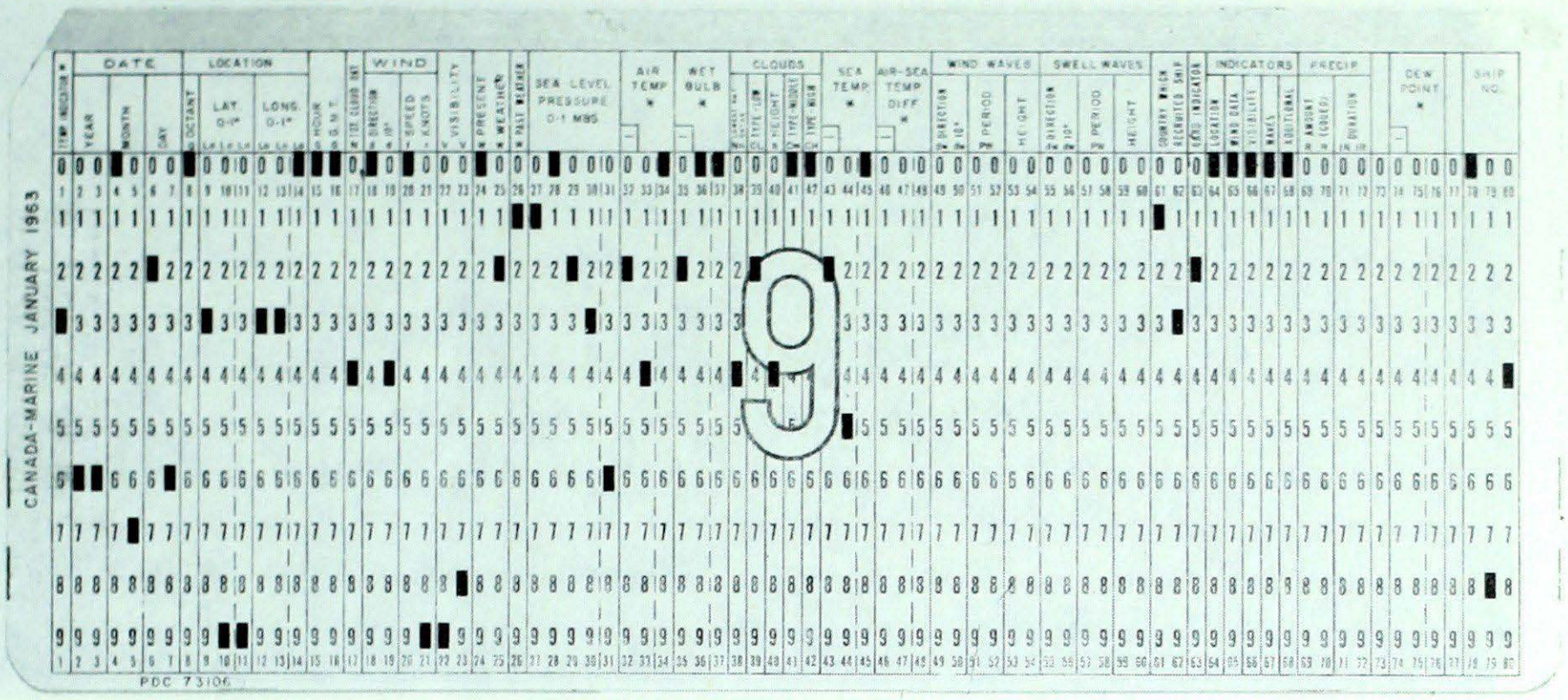

UPPER WIND – CARD # 6



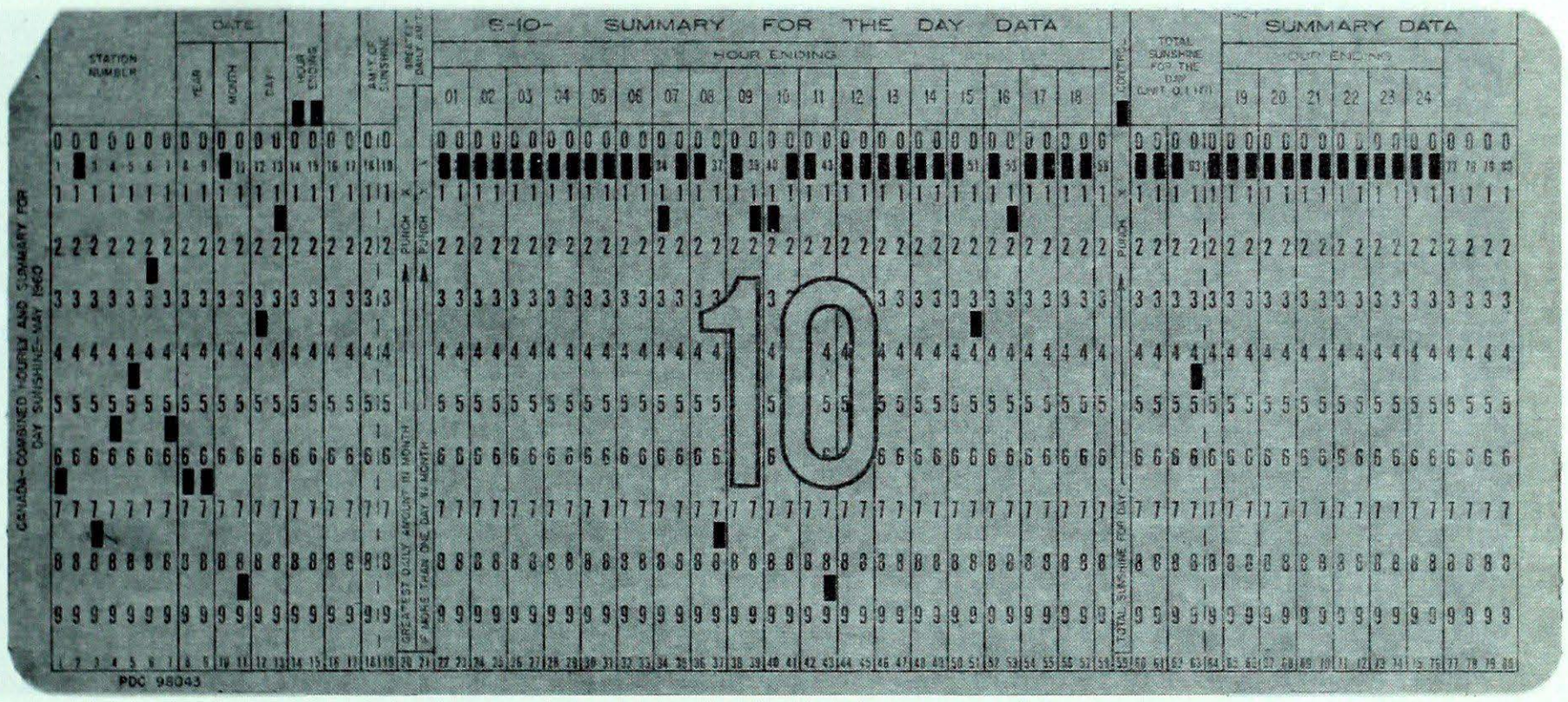

RADIOSONDE SPECIAL DATA – CARD # 7



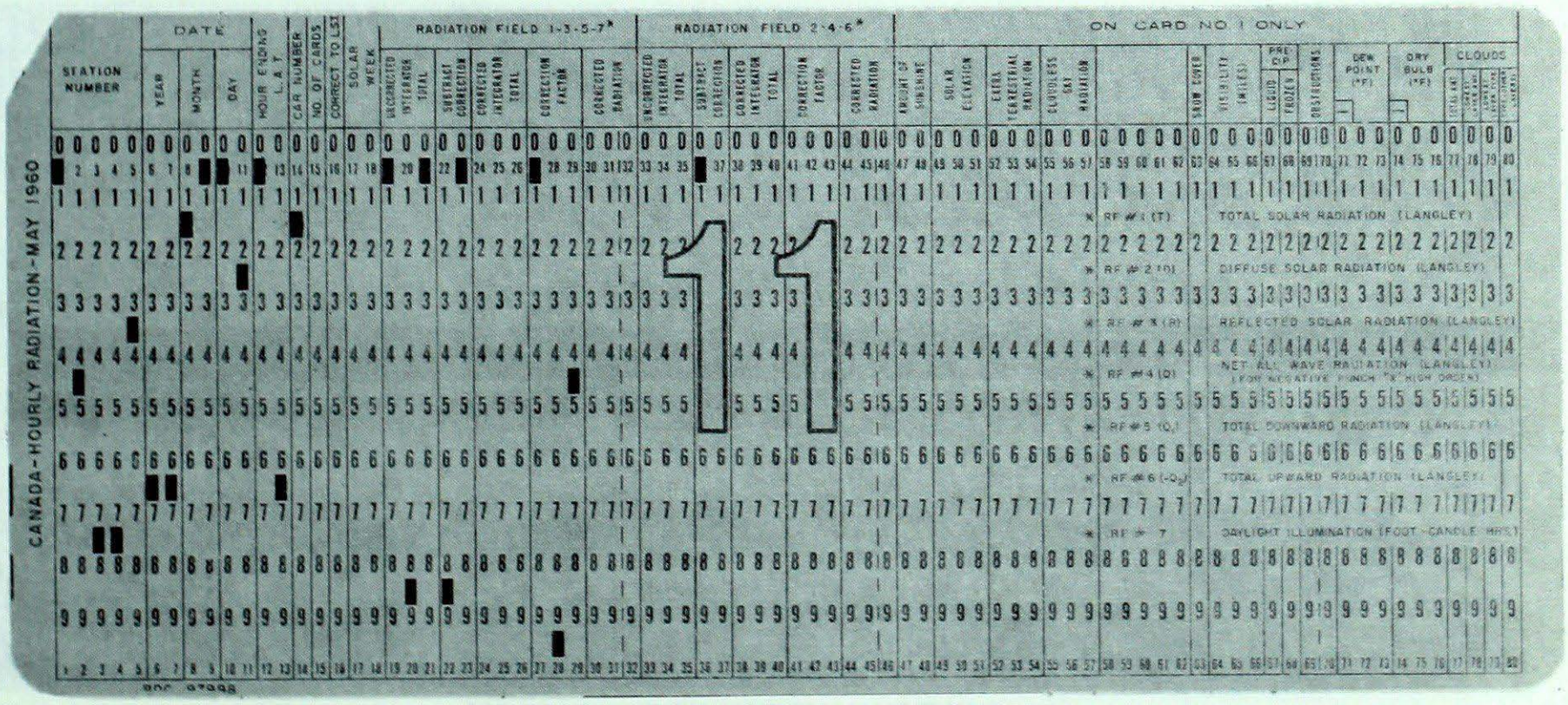

SIGNIFICANT LEVEL DATA – CARD # 8



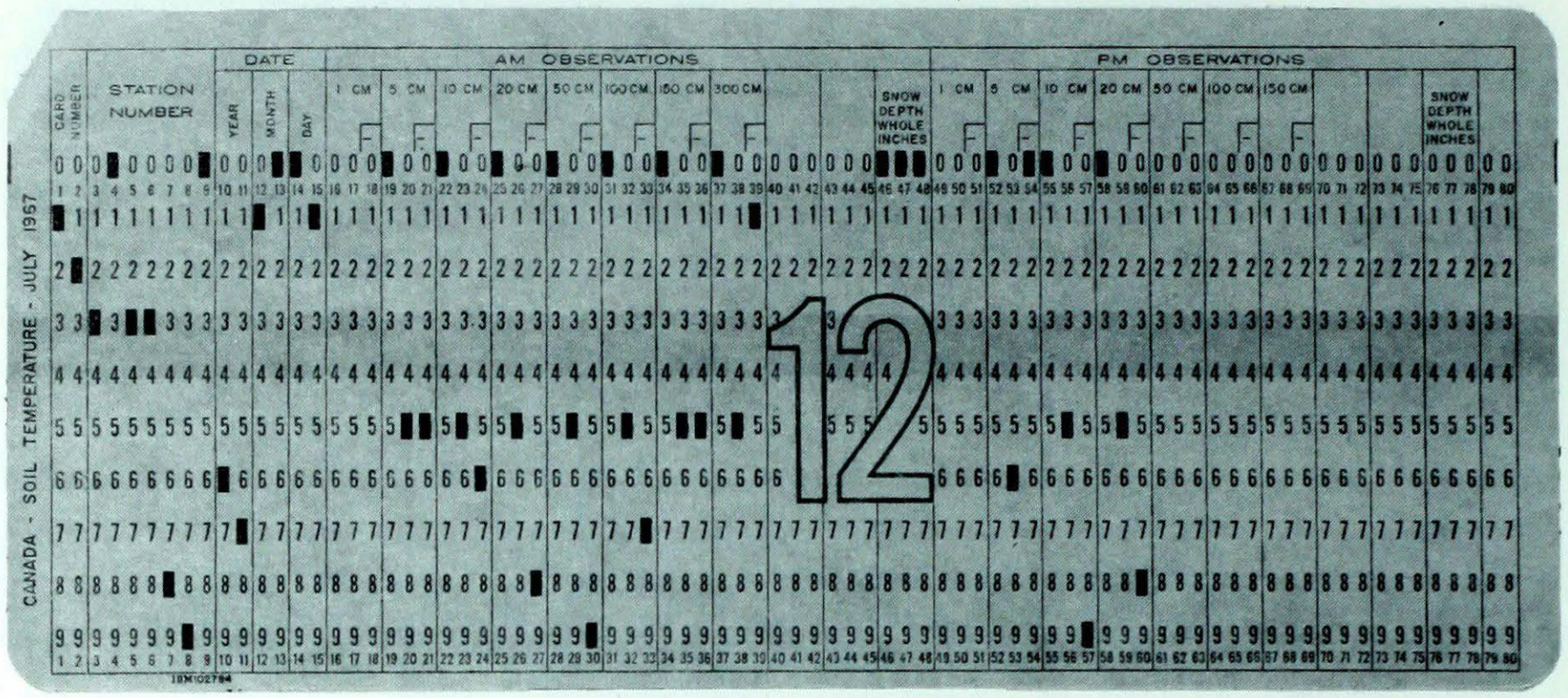

MARINE – CARD # 9



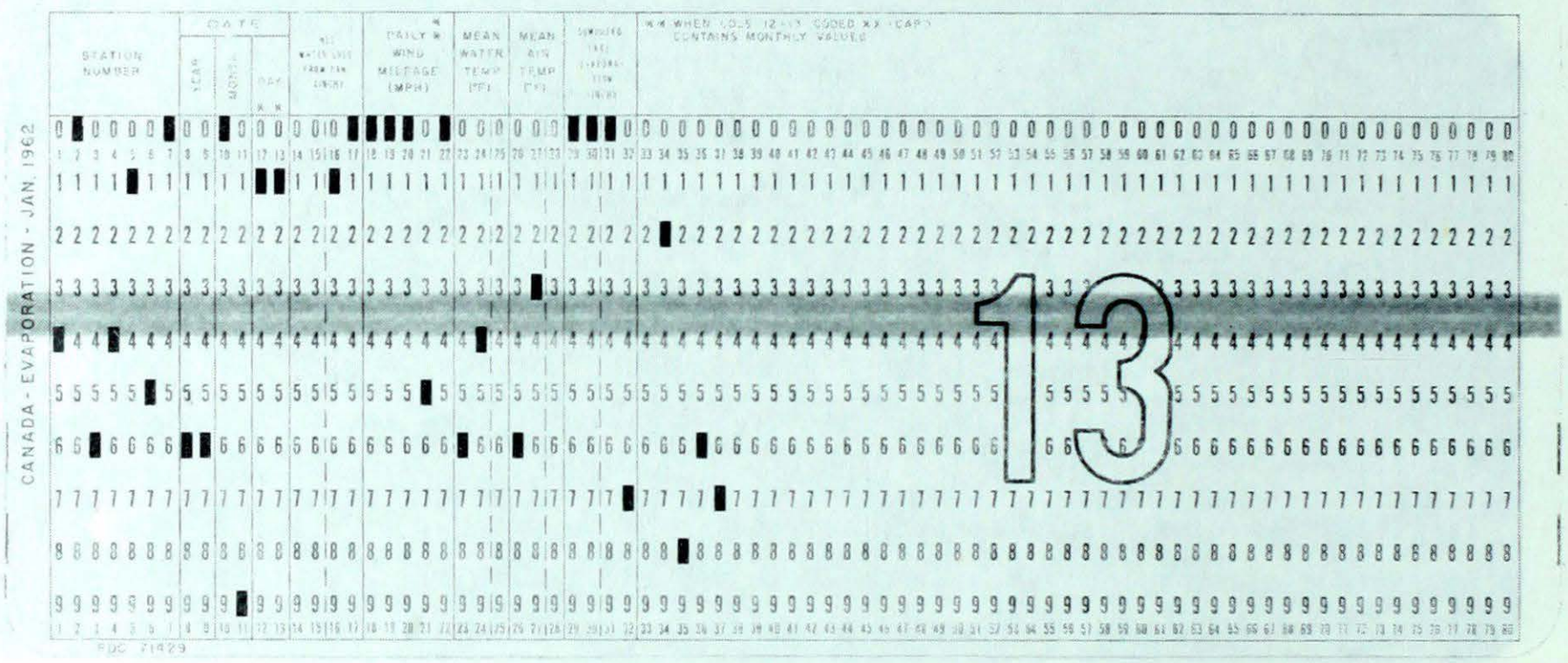

COMBINED HOURLY AND SUMMARY FOR DAY SUNSHINE – CARD # 10



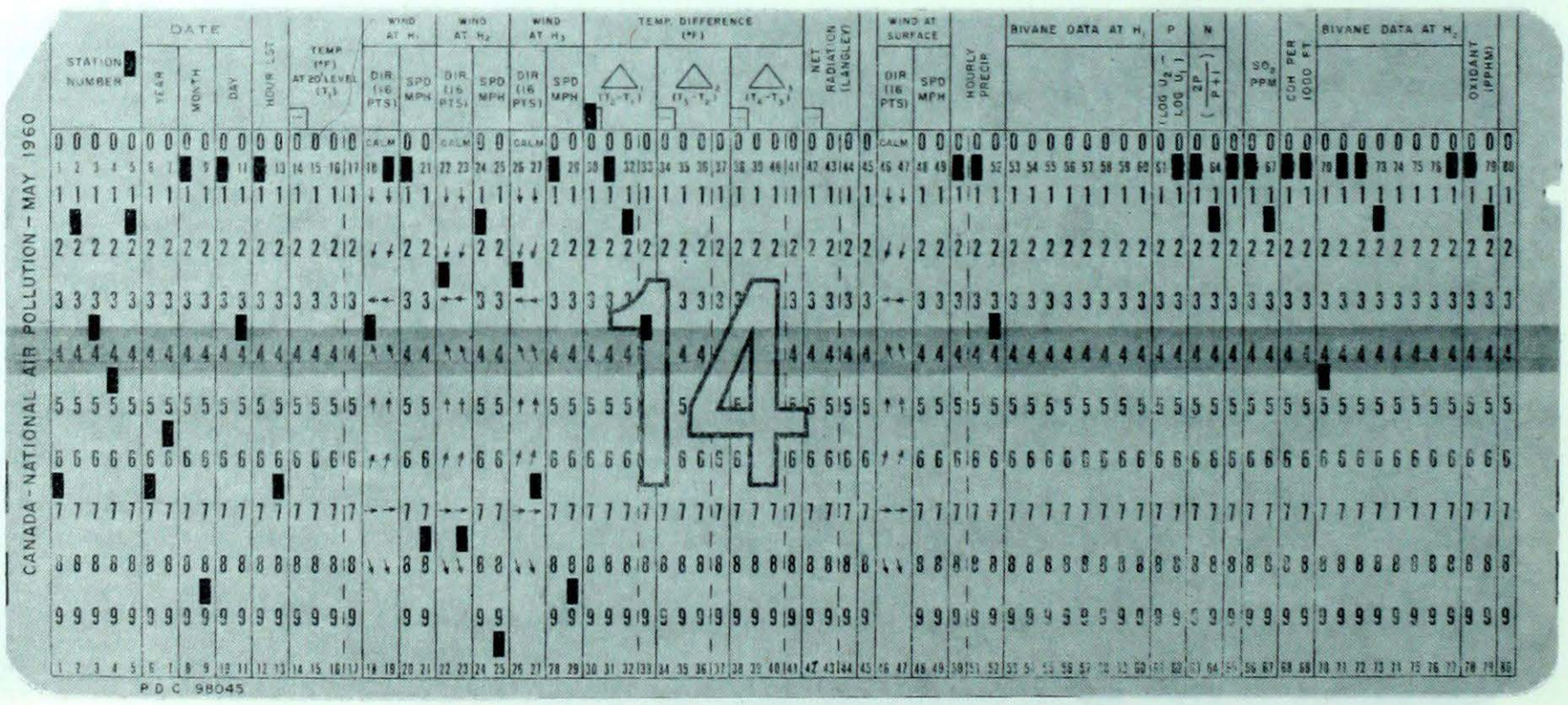

HOURLY RADIATION – CARD # 11



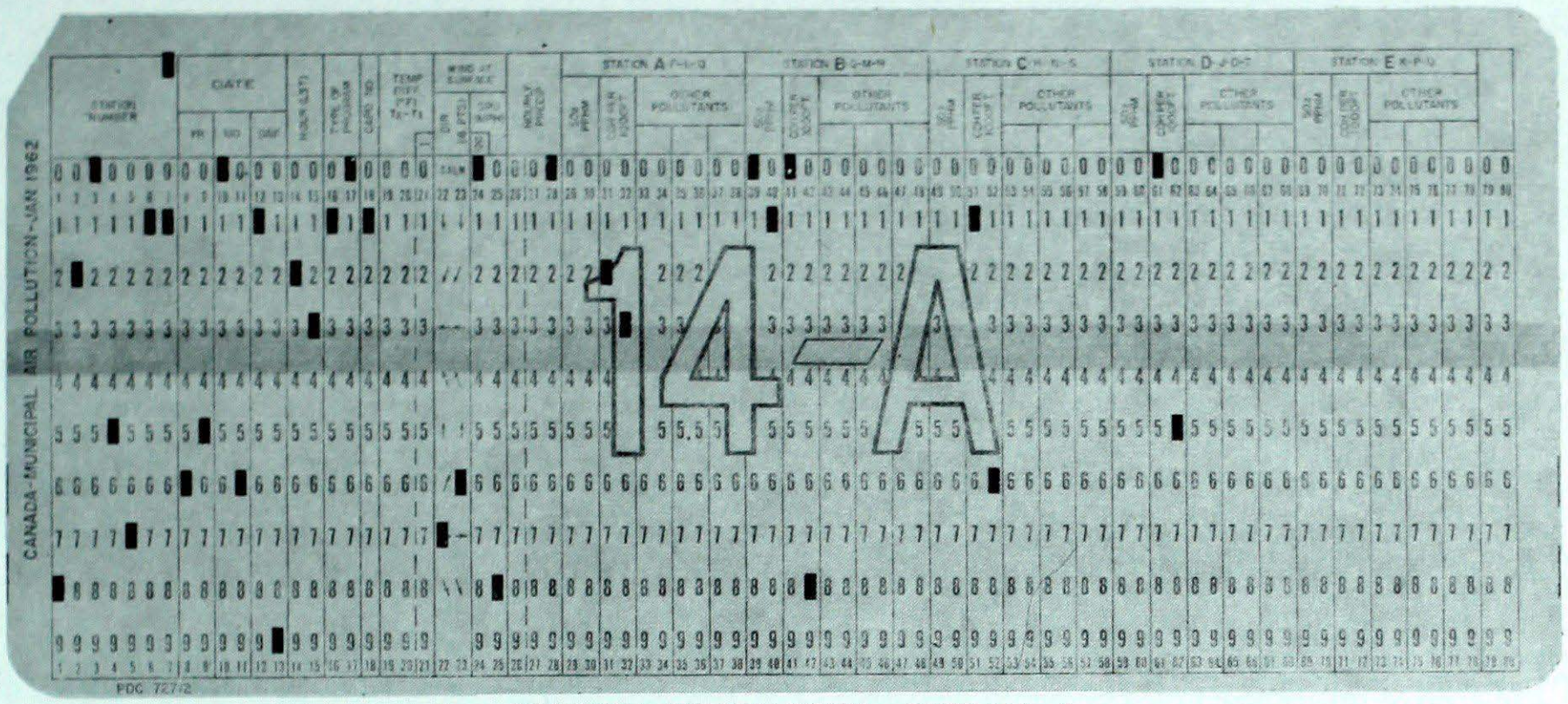

SOIL TEMPERATURE – CARD # 12



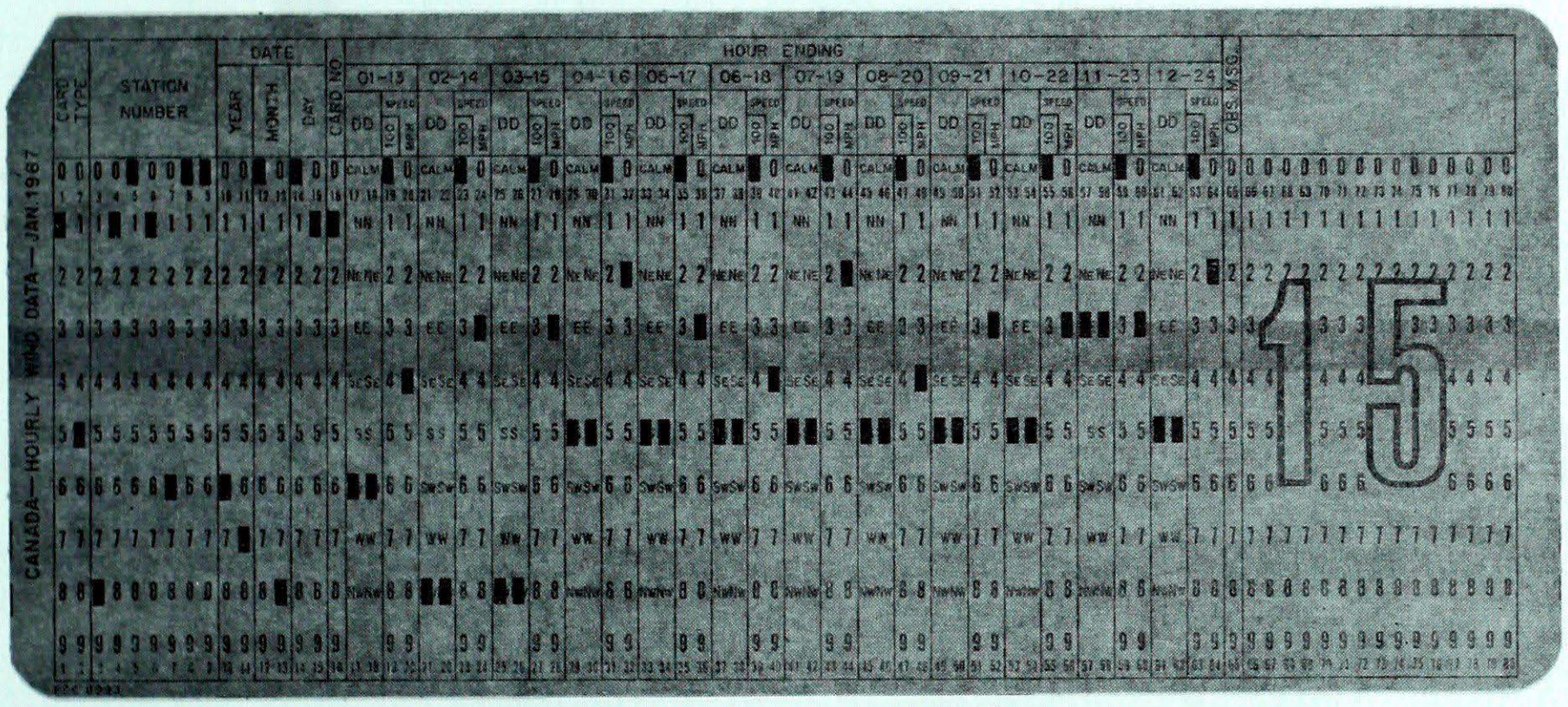

EVAPORATION – CARD # 13



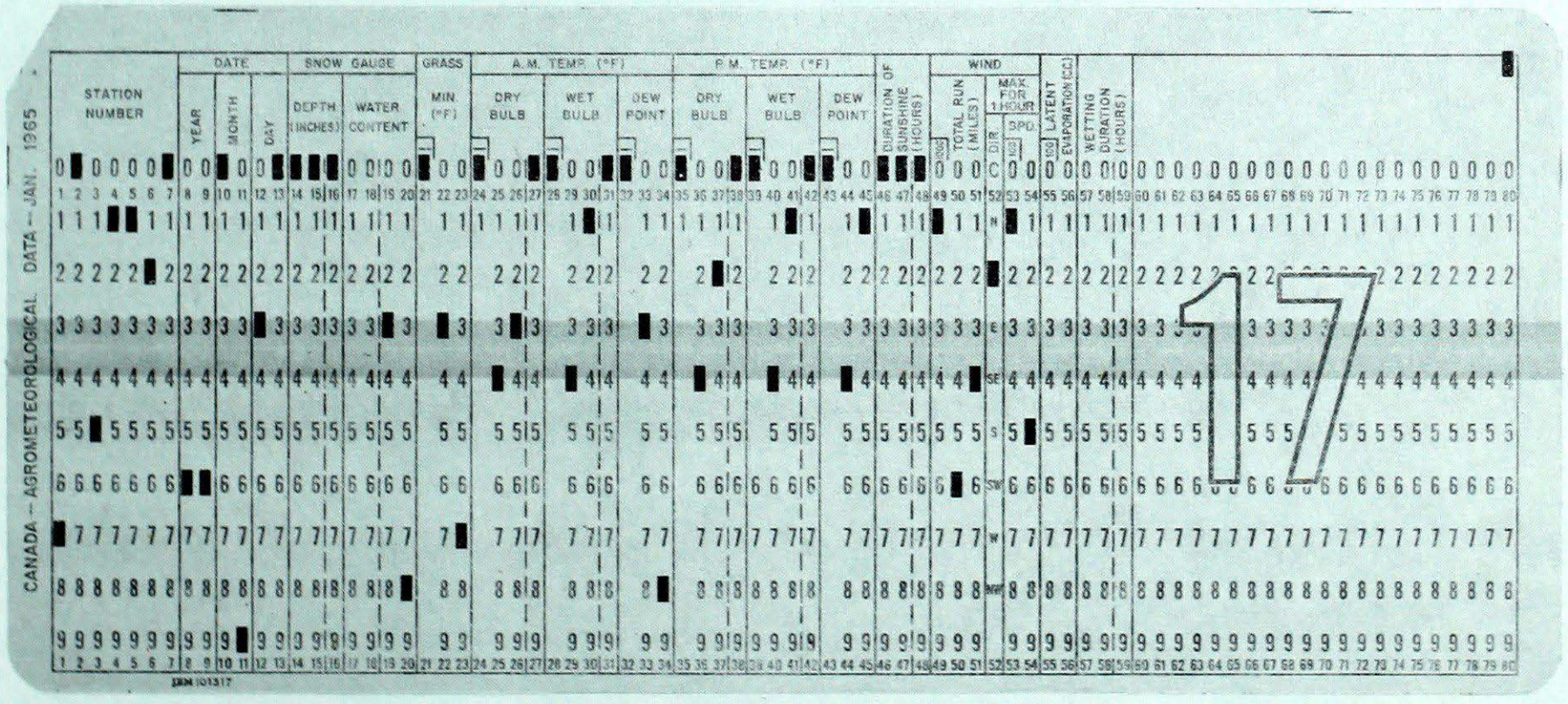

NATIONAL AIR POLLUTION – CARD # 14




MUNICIPAL AIR POLLUTION – CARD # 14 – A




HOURLY WIND DATA – CARD # 15




AGROMETEOROLOGICAL DATA – CARD # 17