NASA

Technical Memorandum 86152

A GENERAL SCIENTIFIC INFORMATION SYSTEM TO SUPPORT THE STUDY OF CLIMATE-RELATED DATA

Lloyd A. Treinish

SEPTEMBER 1984

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

A GENERAL SCIENTIFIC INFORMATION SYSTEM TO

SUPPORT THE STUDY OF CLIMATE-RELATED DATA

LLOYD A. TREINISH

NATIONAL SPACE SCIENCE DATA CENTER
DATA MANAGEMENT SYSTEMS FACILITY
CODE 634

SEPTEMBER 1984

NASA/GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

TABLE OF CONTENTS

I.   INTRODUCTION

During the late 1970's, a number of U.S. governmental, international, and academic reports emphasized the need for a national and world climate program as well as a climate monitoring system. These efforts culminated with Congress passing the National Climate Program Act in 1978. The National Aeronautics and Space Administration (NASA) has been a major participant in climate research for several decades through spacecraft and instrument development, and the handling, preparation, and analysis of data. It is logical that part of NASA's response to this congressional mandate was the establishment of a useable climate data base for environmental satellite measurements and other atmospheric parameters (GSFC - 1977).

Today, the climate research community has access to an unprecedented collection of environmental satellite and related ground and atmospheric measurements. The climate scientist frequently spends countless hours tracking down the availability, status, location, and accessibility of desired data sets before beginning actual research. NASA's Goddard Space Flight Center (GSFC) has developed the Pilot Climate Data System (PCDS) as a significant aid in supporting the climate scientist by providing a comprehensive data management and analysis capability.

NASA's Pilot Climate Data System (PCDS) offers a valuable new tool for researchers using climate-related data. The PCDS is being developed to serve as a focal point for managing and providing access to a large collection of actively used data for the earth, ocean and atmospheric sciences. The PCDS provides uniform data catalogs, inventories, and access methods for selected

1

NASA and non-NASA data sets. Appropriate data manipulation capabilities have been developed to enable scientific users to preview the data sets using graphical and statistical methods. The PCDS is designed to be an easy-to-use, generalized scientific information system for the support of researchers in the aforementioned disciplines. It has evolved from its original purpose as a climate data base management system in response to a national climate program, into an extensive package of capabilities to support many types of data sets from both space-borne and surface-based measurements with flexible data selection and analysis functions (OAO - 1979).

The PCDS supports NASA climate, weather and severe storm research, and application programs as well as various NASA scientists and NASA-funded researchers. These users can employ the PCDS to scan, analyze, manipulate, compare, display and study climate parameters from many different data sets. Data producers can use the system for validating, evaluating, and archiving data or maintaining account records and data inventory. Academic researchers, who may be working with limited budgets can obtain quick access to selected portions of larger data sets. In addition, information on data demands can be used by managers for planning data processing and analysis activities.

The PCDS is composed of a User Facility and an Update Facility. The PCDS User Facility, which is described in Section III, provides the means for an individual to access and work with data and information about that data (Reph et al. - 1984a). A technique for preparing a data-independent model of climate-related data via a file structure, which is described in Section IV, has been developed. It is used to provide data-independent display and analysis functions within the User Facility. The PCDS Update Facility, which

2

is described in Section V, provides the means for maintaining and enhancing
the data bases that compose the PCDS (Reph - 1983).

Figure 1 illustrates schematically the information flow in the PCDS. On
the left of the figure, are indicated several sources of climate-related data.
These sources provide data sets on magnetic tape as illustrated in the next
column, which reside in the PCDS Library. The PCDS manages these data via
on-line catalogs, inventories and data bases. A user can access this
information as shown on the right.

II. IMPLEMENTATION

The PCDS is implemented on a computer system based upon a Digital Equipment Corporation (DEC) VAX-11/780[1] located at the National Space Science Data Center (NSSDC) of NASA's Goddard Space Flight Center (GSFC) in Greenbelt, Maryland. It uses the Transportable Applications Executive (TAE) as a user interface (Helfer, et al. - 1981). The TAE was also developed at GSFC and provides a convenient vehicle for program development and operation. It provides a menu-driven, user-friendly interface for novice users, and a command language interface for experienced users. The TAE is used in the PCDS to provide a uniform user interface with extensive on-line help facilities for a variety of software. This enables a user to learn a single "language" to access many kinds of tools.

For cost effectiveness, the PCDS design integrates existing technology, including several commercial software packages, with GSFC-developed software, into a useful scientific information management facility. In addition to the TAE, the PCDS uses a commercial data base management system (ORACLE [2]), graphics package (TEMPLATE [3]), numerical mathematics and statistics package (PROTRAN [4]), symbolic mathematics (MACSYMA [5]), and network communications (DECNET [6]). These software systems were selected because of their

---

[1] VAX is a registered trademark of Digital Equipment Corporation.

[2] ORACLE is a registered trademark of Oracle, Inc. (Oracle - 1983).

[3] TEMPLATE is a registered trademark of Megatek, Inc. (Megatek - 1983a, Megatek - 1983b).

[4] PROTRAN is a registered trademark of IMSL, Inc. (IMSL - 1983).

[5] MACSYMA is a registered trademark of Symbolics, Inc. (MIT - 1983)

[6] DECNET is a registered trademark of Digital Equipment Corporation (DEC - 1983).

4

capabilities to meet the PCDS goals in the PCDS environment for the least total cost. These components are integrated with the GSFC-developed software under the TAE environment so the user need not master each software package. New technology can be incorporated continuously, as it is developed, because each software component can be replaced or enhanced with other commercial products or updates without affecting the user's view of the PCDS. By taking this approach, the PCDS was able to meet its goal of demonstrating a significant application of state-of-the-art data management techniques without spending unnecessary time or money developing special-purpose hardware or software.

III. USER FACILITY

The PCDS User Facility consists of five major components or subsystems that are described below. Each of these subsystems can operate independently (IMB - 1984). Briefly, the components are:

1. An extensive on-line catalog that uniformly describes many data sets (Catalog)

2. An on-line inventory of PCDS data holdings (Inventory)

3. A variety of accessible data sets and a range of data set selection capabilities to select desired data according to time or geographic areas (Data Access)

4. A set of data manipulation utilities (Data Manipulation)

5. A set of data display utilities (Graphics)

Figure 2 illustrates the data structure of the PCDS User Facility. It shows the five subsystems as well as the various collections of data or information about data, with which they interact. These data bases are also described in the following subsections. The data-independent Climate Data File (CDF) ties together the Data Access, Data Manipulation and Graphics Subsystems. CDF is described in Section IV.

III. A. Catalog

The PCDS Catalog Subsystem provides a central source of on-line information about a variety of data sets in a standard format. It can be used to determine the availability and location of data.

The Catalog describes climate-related data sets and their associated sensor measurements that are contained in the PCDS library as well as other data archives. The data descriptions include the characteristics, processing status, availability, quality and contacts for further information. Currently, this subsystem describes about 150 climate-related data sets with details on both existing and planned data sets and products. This information is at a fairly high level of aggregation (e.g., all Backscatter ultraviolet (BUV) radiance measurements from the Nimbus-4 satellite could comprise one data set), enabling a user to determine whether or not to retrieve data. The Catalog provides information for a user to learn about climate data prior to making inquiries as to its volume and availability, or actually studying the data.

The Catalog is completely accessible on-line through computer terminals and is occasionally printed in a hardcopy format (i.e., Reph – 1984). The on-line version consists of a Summary Section which can be queried using keywords, and a Detailed Section which can be browsed like a book. The Summary Section can provide a very compact output of one line per data set, or it can provide a more detailed description with a full terminal screen of information per data set. The data base supporting the Summary Section is managed by ORACLE. The Detailed Section can then be browsed to acquire more

information about any data set.

Table I illustrates an example of the full-screen output from the
Summary Section. It displays the spatial and temporal coverage, and available
products and their archival status for level II (i.e., orbital) data that
contains information about ozone as measured from the Nimbus-4 spacecraft.
The Catalog Subsystem shows that one experiment generated such data. The data
set descriptions in the detailed section are outlined in Table II.

A hardcopy utility is also available for use with the Catalog and the
Inventory Subsystems, which allows the user to print the contents of any
terminal screen.

The sources and parameters covered by the current contents of the
catalog are shown in Table III. Most of the sources in the list are from
various NASA and National Oceanic and Atmospheric Administration (NOAA)
spacecraft.

III. B. Inventory

The Inventory Subsystem provides detailed information about temporal coverage, data volume and specific data location for data sets available through the PCDS. This information is used by the Data Access Subsystem to automatically transfer data from the PCDS Library to a user data set when the user specifies a data type, time range, and geographic area. In this way, the mechanics of accessing a data subset are isolated from the users. The Inventory Subsystem also allows a user to directly query a data base, managed by ORACLE, which contains the detailed information about the data sets which are available via the PCDS. The Inventory describes the data holdings of the PCDS in sufficient detail to enable the PCDS to retrieve, locate and access the requested data or permit a user to decide to request particular data.

The Inventory is the lowest level in the data description hierarchy. It describes data sets at a fairly low level of data aggregation (e.g., an image, orbital strip, or physical file could comprise one data set). The information about the data maintained in the Inventory is at a resolution consistent with each individual data set. For example, if a data set contains one day of data per physical file on magnetic tape, then the Inventory maintains its information at a resolution of one day. All data sets described in the Inventory are available via the PCDS. (No planned or future products are described in the Inventory.) Since the Inventory is maintained on-line in a codified format, a user can easily query it, specifying keywords to limit the information listed.

The Inventory Subsystem consists of several programs which allow a user

to query the on-line inventory. Output is provided in tabular listings or a graphical format. The queries have been designed to allow a user to progress from simple queries to more detailed ones by using the information returned from the simpler queries as inputs to the more detailed queries. For example, the Inventory can list what kind of climate information is available. If a user is interested in ozone, ozone would be in that list. From this point, the Inventory can identify those data sets in the PCDS that contain ozone. For each data set, the Inventory can provide characteristics of the data coverage, a list of tapes and files that contain ozone, etc. Thus, a user can start with an interest in a parameter like ozone, which was perhaps initiated by browsing the Catalog, and determine via the Inventory queries the volume and availability of actual ozone data. Another query is also available to identify additions to the Inventory since a specified date. Table IV shows an example of tabular output from the Inventory Subsystem. It shows a list of the available data in the PCDS Library for the detailed ozone profile (DPFL) dataset from the Nimbus-4 BUV instrument. It illustrates the number of tapes, files, dates and orbits covered, and some history of the tapes for 1972. Some of this inventoried information is displayed graphically in Figures 3 and 4. Figure 3 shows a data coverage map (i.e., orbit tracks for the Nimbus-4 spacecraft when data from the BUV instrument are available in the DPFL data set) for January 1, 1972. Figure 4 shows the data rate for the DPFL data set as a line chart for all of 1972.

The current capabilities of the Inventory Subsystem include the following functions:

  o List all available climate parameters

o List available data types and corresponding climate parameters

o List information about each selectable logical unit of data (e.g., maps, grids, profiles, etc.)

o List a summary of information about available data types, such as number of tapes, number of files, orbit range covered, and the time range covered

o List information at the tape level about the available data, providing tape identification, number of files, orbit range, generation date, inventory date and time range

o List information at the file level about the available data, providing tape identification, file number, time range, orbit range and size

o List a history of the tapes inventoried in the PCDS

o Provide a hardcopy utility identical to the one used in the Catalog Subsystem

o Graphically summarize the Inventory contents by using plots of data rate as a function of time or by showing data coverage overlaid on a world map

The availability of graphical products from the Inventory Subsystem represents an interface between the ORACLE database management system and the TEMPLATE graphics package. In response to user queries submitted via the TAE, ORACLE is requested to retrieve specific information concerning the amount or availability of data in the PCDS. The data are then reformatted and passed to TEMPLATE for graphical display. Several representation schemes are available for data rate plots (e.g., line plot, bar chart and pie chart). For geographic displays the user can set a window for the world coastlines.

11

Table V lists the sources of data sets currently supported by the Inventory Subsystem. Selected data sets from each experiment are available on the PCDS.

III. C. <u>Data Access</u>

The Data Access Subsystem allows selection of data subsets primarily on the basis of time and/or geographic location. Data sets, which are supported by this subsystem, are kept in a library of magnetic tapes or disks within the PCDS. The tapes are typically in the original format provided by the data set producer. The Data Access Subsystem allows a user to select a subset of any data set and output it in one of several formats. A user will typically learn about climate data and its availability via the Catalog and Inventory Subsystems prior to employing the Data Access Subsystem to extract actual data of interest.

The input data to the Data Access Subsystem can be a standard tape from the PCDS library, a tape that was created earlier using this utility, or it can be a subset that was earlier placed on a disk. The user can copy a data subset to another magnetic tape or to magnetic disk, or can simply get a hardcopy listing. The most important output option of the Data Access Subsystem is to convert the data subset selected by the user into a self-describing, data-independent format which can then be employed by the Data Manipulation and Graphics Subsystems of the PCDS. This format is called Climate Data File (CDF), which is described in Section IV.

The Data Access Subsystem employs the PCDS Inventory as a data dictionary to automatically locate data without the user being aware of the details. (Refer to figure 2 for a schematic of the relationship between the Data Access and Inventory Subsystems.) The user never has to know anything about tapes, files, data formats, etc. The Inventory and special software

13

take care of all the details that are required to locate, read, and translate each of the different data types to the CDF format. To provide this ease of use, the Data Access Subsystem supports only those data sets that are held in the PCDS Library, and for which custom access software has been written. The data sets currently supported with Data Access are from the experiments listed in Table V.

Since the Data Access Subsystem isolates the user from the details of the original data source, it could be used to access data available on media other than magnetic tape or disk and at locations other than the PCDS. This capability has been demonstrated through the access of on-line packetized spacecraft data via another computer on a remote network (i.e. DECNET) node (Green - 1983). A request for that "remote" data is of the same form as a request for "local" data. The PCDS Inventory keeps track of where the data are located and Data Access "knows" how to access them.

III. D. <u>Data Manipulation</u>

The Data Manipulation Subsystem provides a set of utilities that can be used with any data set that is structured in the CDF format. These utilities allow the user to customize a data set before working with it in the Graphics Subsystem, or before transferring it to another computer for further work. These utilities can work on a data set that was extracted via the Data Access Subsystem, created via a previous application of the Data Manipulation Subsystem or generated independently of the PCDS. Since the tools in the Data Manipulation Subsystem are data-independent, they can be used to support parallel analyses on heterogeneous data, an important capability for interdisciplinary studies like climate. These utilities (i.e., the first seven on the following list) are designed to input a CDF and output another CDF so that they can be concatenated to perform any desired combination of functions. The utilities currently available are:

   o CDF Subset – to further subset data. This utility allows the
      customization of an output data set by selecting any fields from a
      CDF and editing (filtering) the data set based on any combination of
      fields within the CDF.

   o CDF Merge – to merge two CDF's. This utility allows two data subsets
      to be merged together. One important application of this utility is
      to produce overlaid data sets for comparison.

   o CDF Combine – to combine data elements algebraically. This utility
      produces a new CDF with additional fields based on a user-specified

15

combination of the fields within the original CDF. Any algebraic
combination can be specified.

o CDF Average - to apply statistics to data elements. This utility
   supports the calculation of means and variances from any time period
   within a CDF. Running time averages, ensemble or canonical averages
   and data compression are supported.

o CDF Ungrid - to ungrid a gridded data set. This utility can split a
   map grid into its data, latitude and longitude components and place
   them in an ungridded format in a new CDF to permit comparisons with
   other ungridded data.

o CDF Grid - to create or reformat a gridded dataset. This utility can
   take ungridded data as triples of information (latitude, longitude
   and data), and grid them into a new CDF via a user-selected meshing
   algorithm to permit comparison with other gridded data. In addition,
   the capability to take an extant gridded data set and regrid it or
   change its coordinate system is provided.

o CDF Anomaly - to compute deviations (or anomalies) between parallel
   data sets. A utility is provided to remove a signature from a data
   set that masks its information. For example, a simple mathematical
   model (e.g., a CDF containing some representation of the average or
   ensemble distribution of the data) could be subtracted from a data
   set. The results would be placed within a new CDF.

o   CDF Listing — to create listings of a CDF. This utility can generate

    a listing whose format can be tailored by selecting only the fields

    of interest or editing (filtering) to display only records of

    interest from the CDF.

o   Access to the PROTRAN numerical mathematics and statistics software

    (IMSL — 1983).

o   Access to the MACSYMA symbolic mathematics software (MIT — 1983).

III. E. Graphics

The Graphics Subsystem provides graphical display utilities for use with any data that are structured in the CDF format. This subsystem can be used to preview a data set, or it can be used to get hardcopy output for further analysis or for publication. Extensive representations of data stored in a CDF are provided. In addition, the generation of text charts is provided. The utilities of the Graphics Subsystem employ the TEMPLATE package to generate all graphical displays. TEMPLATE provides the framework for graphical output independent of the graphics device in use, highly flexible tools for graphics generation and quick response time for the creation of requested displays. The Graphics Subsystem provides displays on a variety of hardware with a wide selection of capabilities. A user can employ low-cost graphics terminals for quick-look displays, or access more sophisticated terminals for rapid display and analysis of complex data. In addition, high quality, hardcopy products suitable for publication can be produced in both black-and-white and color.

All of the graphic displays emphasize annotation for the proper labelling and identification of plotted data using information derived from the CDF containing the displayed data. The displays are placed in a common format to permit visual comparisons of different data represented in similar fashions. For all data, multiple representation schemes, each with several options, are provided to permit the user to gather as complete an understanding of the data as is possible via computer graphics.

For example, two-dimensional graphics utilities provide standard x-y

plots with one or more dependent variables on the same plot. Scatter

diagrams and vector plots are supported. Plots can be made using rectangular

or polar coordinates with linear or logarithmic axes. Linear, polynomial, and

spline curve fits can be performed on the data set before plotting. Several

types of statistics can be provided, including data smoothing, means and

standard deviations of the data. In addition, histograms of a single variable

with statistical options are available. Several different type fonts are

available for output, including several publication quality fonts.

Three-dimensional graphics utilities provide contour plots, surface

diagrams, and pseudo-color images. Map grids with outlines of the world

coastlines in various map projections can be overlaid on these plots of

geographic data. These utilities also provide a large number of options to

tailor the output to specific user requirements for format, quality, and

content.

A graphics post-processing utility permits the redirection of any output

from the other graphics utilities to various graphics devices. A user can

store the plots and other graphics information on magnetic disk and recall

them at any time via the post-processor. Depending on the user's

requirements, options are provided for combining and reformatting plots, and

generation of publication or presentation quality output (e.g., camera-ready

copy, 35mm slide).

Figures 5 and 6 illustrate sample two- and three-dimensional data

representation graphical products, respectively, as generated from the PCDS.

The graphics post-processing utility was used to create the actual displays in

those figures from plots previously archived on the PCDS computer system. All of the plots in the two figures were generated from CDFs created by the Data Access Subsystem which were further processed by the Data Manipulation Subsystem. Figure 7 shows the statistical distribution of surface temperatures for Taipei, China from seventy-two years of data averaged to monthly resolution (1900 through 1972). The inner cross-hatched histogram bars show the portion of the distribution within one standard deviation of the mean. The other bars are outside of that regime. A curve for a gaussian (normal) distribution corresponding the mean and standard deviation of the data is drawn for reference. Although, the distribution of data displayed is obviously non-gaussian, the curve and hatched bars are drawn to illustrate the versatility of this graphics option. Figure 8 shows the distribution of geopotential heights over the north pole as calculated for January 1, 1975 from 0:00 to 12:00 (GMT). The data are represented as a three-dimensional surface in a north polar stereographic map projection. A plane indicating the geographic coverage of the data with coastlines and fiducial markings is drawn below the surface.

IV.   CLIMATE DATA FILE

Climate Data File (CDF) is a self-describing, self-consistent data structure which provides the applications software in the PCDS Data Manipulation (Section III. D.) and Graphics (Section III. E.) Subsystems with true data independence. In other words, the programs in the Data Manipulation and Graphics Subsystems do not have specific knowledge of the data with which they are working. This permits a user to apply the same function to different data without concern - the user relies on his own intelligence to interpret the results, a critical feature for multi-disciplinary studies like climate. These programs use the information in CDF to inform the user as to the contents, history and structure of CDF, and provide sufficient annotation of a program's results (e.g., a graph).

As CDF isolates the details of the structure of a dataset from a user of the data in various applications software, the programmer of such applications does not need to know the details of the storage of CDF, which is implemented as a data abstraction (cf. Liskov, et al. - 1977). The programmer deals only with a collection of operations on CDF (e.g., access, manipulate) that are maintained in an Interface Routine Library. This isolation permits enhancements to the CDF implementation as new software and hardware technology permits, with no changes to applications software. The user simply perceives improved performance. In addition, the CDF concept is extensible in the programmer's perspective by the addition of new operations to the interface library.

As shown in Figure 7, the CDF structure is supported by several files

(cf. Chatterjee - 1984). Each CDF consists of one header file and at least

one data file. The header functions as a data dictionary for one or more data

files by describing the information in such data files (i.e., organization,

structure, history, parameters, units of the parameters, formatting

information). Figure 8 illustrates the relationship between the data

descriptions in the header and the actual data. The contents of the header

are employed to inform a user of a CDF's contents for selection, manipulation,

etc. in an applications program, or for labelling the results produced by an

application. The header information maps to the data file(s) for proper

selection of actual data values. This mapping can be altered for use in an

application via a view file.

The view file permits a data file to be looked at in a logical

organization different from its physical organization. For example, a data

file may contain records of several parameters each, which are sorted by the

first parameter. If, in some application the data need to be sorted by the

second parameter, then a view file containing pointers to the order of the

second parameter can be employed so that the data file need not be

regenerated. In addition, the view file can be employed to give a restricted

or filtered look at the data file. The interaction between these files is

also illustrated in Figure 8.

The concept of using a data dictionary to describe the contents of a

data file is not new for the purpose of achieving a data-independent,

transportable standard, especially in the geophysics community (cf. Thomas and

Guertin - 1981). However, the CDF is at variance with those standards by

being oriented towards the end user—the researcher, rather than the

programmer that wants to read someone's data set on a tape. This is

illustrated by the provision of different logical views of a single data set

for different applications. Another more important illustration of the

difference between CDF and conventional data standards is in the data

description in the CDF header file. Each of these descriptors not only

defines the name of each data parameter as well as its units (e.g.,

TEMPERATURE [DEGREES KELVIN]), but the organization of the data into data

constructs. The CDF supports various data constructs that can be concatenated

into more complex structures that may be needed to describe data. Such

constructs are a mechanism for storing data that constitute some entity of

interest to a user of the data (e.g., an atmospheric temperature profile--a

collection of temperatures at various levels in the atmosphere). The simplest

such construct would represent one dimension of data, a collection or vector

of numbers. The next level would imply two dimensions of data as two parallel

vectors like an atmospheric profile with a vector of values and a vector of

levels, each of which corresponds to a value. A three-dimensional construct

implies a matrix of values and a matrix of auxiliary data like a map of values

at specific latitude-longitude locations. Another example would be the

combination of two, two-dimensional constructs - a time history of values and

a profile of values to yield a profile history, i.e., a collection of

information as a function of time and atmospheric height. Table VI shows this

progression of data constructs, gives some examples from climatology, and

illustrates ways of viewing such entities graphically. It should be

emphasized that although the examples are from the atmospheric sciences, the

techniques apply to ALL data. In Figure 8, a parameter in the data file

implies some data construct (i.e., scalar, vector, etc.), while the

descriptors in the header and view files define the nature of the construct

(i.e., dimensionality, type, units, etc.).

Climate Data Files are in general initiated by the Data Access Subsystem. Additional CDFs are generated by the Data Manipulation Subsystem which acts on extant CDF, or by custom software that is independent of the PCDS. Any CDF is viewable via the Graphics Subsystem. (Refer to Figure 2 for a schematic of the relationship between CDF and these subsystems.) Once a user has generated a CDF containing data of interest, that user may choose to migrate the CDF to another computer facility via network protocols or off-line media. Once the data are available on the other computer, the user can analyze the data with custom, data-dependent algorithms to meet a specific research objective. The results could be placed into another CDF and returned to the PCDS for graphics, further manipulation, or archiving. In this way, potentially cpu-intensive or data-dependent analyses can be distributed to appropriate facilities other than the PCDS, which is devoted to information management. In addition, a version of the CDF Interface Routine Library would be provided on those other computers to permit the applications programmer to work with data migrated from the PCDS via CDF.

V. UPDATE FACILITY

The PCDS Update Facility serves as a support tool for the PCDS User Facility by providing the means for the creation and maintenance of the data bases employed by the Catalog, Inventory and Data Access Subsystems as well as miscellaneous utilities. As illustrated in figure 2, there are four basic data bases (Detailed Catalog, Summary Catalog, Inventory and Data Library), which are logically related to each other. They contain information about data or the data themselves. These data bases, which are described in previous sections, can be accessed with appropriate user commands via the aforementioned subsystems. The ability of a PCDS user to access data information or data must be independent of the data themselves. This concept of data-independence, which is implicit throughout the PCDS, is predicated upon the ability to create and maintain identical kinds of information in underlying data bases.

The support of the Catalog Subsystem is divided into two parts; one for the Detailed Section, and the other for the Summary Section. In both cases, the ability to examine the current data bases as well as to update them is provided. For the former, the information may be updated via a text editor. To update the latter, the tools of the ORACLE data base management system (cf. ORACLE - 1983) are utilized.

The support of the Inventory Subsystem is also divided into two parts. The first is concerned with the creation of a summary of the contents of each data tape (or data source) which is to be entered into the library of PCDS data holdings. This summary is created according to a data-independent model.

25

This model contains, for example, a description of the contents of each tape and file in a data set (e.g., tape/file identification, start and stop times, type of data, etc.). As with the Data Access Subsystem, this tool contains custom software isolated from the user for reading a particular supported data tape. The second portion of the Inventory Subsystem support is the entry or loading of the model of a data tape's contents into the Inventory database. Again, the tools provided by ORACLE are involved to accomplish this task (i.e., software which employs ORACLE's host language interface). To add new tapes of a supported data set to the PCDS Inventory, the Update Facility is simply used to enter a summary of its contents to the Inventory database. However, to add a new data set to the PCDS, the following must be accomplished--although not necessarily in this order, nor sequentially:

o   Gather information about data sets so as to understand the data to a
    sufficient level to prepare an entry into the Detailed Section of the
    Catalog.

o   Prepare a summary of the detailed Catalog entry for the Summary
    Section.

o   Have principal investigator/producer of data set review new Catalog
    entries.

o   Update the Detail and Summary Sections of the Catalog by adding the
    new entries.

o   Prepare software that can read and interpret the data tapes.

o   Create the data-independent summary model of the contents of the
    data tapes.

o   Update the Inventory data base with all pertinent information about
    the data.

o   Establish the information to define a CDF of the data including the
    definition of all constructs and their descriptions.

o   Prepare software to translate selected portions of data to CDF.

o   Expand the data reading, interpretation and CDF generation software
    to connect to the Data Access Subsystem. This includes software for
    copying and subsetting from tape to tape, tape to disk, tape to CDF
    and tape to listings.

It should be noted that the tools of the PCDS can be used to verify the
implementation of a new data set from browsing the catalog, checking the
contents of tapes, to listing and plotting data.

VI. CONCLUSION AND FUTURE PLANS

NASA's Pilot Climate Data System provides extensive data management and analysis facilities for use by climate researchers. These facilities are packaged into a system that is extremely easy to use and very flexible. The use of the PCDS requires no knowledge of data formats or programming languages, nor does it require any special programming to support a given application.

The PCDS currently provides direct support to climate researchers at Goddard Space Flight Center. It is expected that this support will be expanded to researchers outside of Goddard, including universities. The system is constantly being improved in response to investigator requirements. New data sets and general-purpose capabilities based on user requirements are incorporated continuously. For example, the International Satellite Cloud Climatology Project (ISCCP) is generating several new data sets that will contain parameters of great interest to the climate research community. Some of these data sets will be archived in the PCDS for availability to NASA's climate community (WCP - 1982). Table VII summarizes the current research activities supported by the PCDS. Several users employ the PCDS from remote locations via networks (e.g., DECNET). It should be noted that research and development activities are conducted in data management techniques and the results are applied to improve the PCDS. The capability of the PCDS will be expanded through the augmentation of the computer facility at Goddard's National Space Science Data Center (NSSDC). During the next year, the installation of greater memory and disk storage, more computing power, transfer of the tape library to optical disk, network communications to

additional computer facilities (e.g., large mainframe and laboratory computers) and additional graphics devices is planned. Concurrent with this effort and responses to user requirements, the software will be augmented to support more statistical analyses and data manipulations, faster response time, and greater flexibility and variety of additional graphical representations including modelling (e.g., MOVIE.BYU, BYU-1984), multidimensional data representations, animation and imaging. In addition, improvements in software components like TAE, ORACLE and TEMPLATE are applied as they become available. Finally, studies of alternate approaches to the techniques employed in the PCDS are on-going. For example, benchmarking of back-end database machines as possible replacements for software data base management systems are underway (cf. BTS - 1983). It is expected that such a database machine will be added to the NSSDC computer facility next year.

The PCDS provides state-of-the-art data management techniques to support comprehensive scientific research. In addition to supporting large data sets from many different sources, it will serve as a model for future systems. It is anticipated that these future data management systems will be integrated into larger network facilities serving many different scientific disciplines.

VII. REFERENCES

Brigham Young University (BYU). MOVIE.BYU – A General Purpose Computer
   Graphics System, Version 5.1.   January 1984.

Business and Technological Systems, Inc. (BTS). A Survey of Data Base Machine
   Applicability to the Pilot Climate Data System (BTS8-83-02/rb).
   November 1983.

Chatterjee, Lisa Conners. Redesign for the Content and Structure of a PCDS
   Climate Data File (CDF). NASA/Goddard Space Flight Center, April 1984.

Digital Equipment Corporation (DEC). DECnet-VAX User's Guide. May 1982.

Green, James L. "Data Systems Technology Program (DSTP) Data Base Management
   System/Mass Memory Assembly (DBMS/MMA)." NETWORDS, 9, no. 3, pp. 1-14,
   November 1983.

Helfer, Dorothy P., Howell, David R., Owings, Jan, Szczur, Martha R. and
   Dalton, John T. "A Transportable Executive for Interactive
   Applications." Proceedings of the Harvard Computer Graphics Conference,
   July 1981.

IMSL Corporation. PROTRAN User's Manual (Version 1.1). January 1983.

Liskov, Barbara, Snyder, Alan, Atkinson, Russel, and Schaffert, Craig.
   "Abstraction Mechanisms in CLU." CACM, 20, p. 564, 1977.

Massachusetts Institute of Technology (MIT). A Brief Overview of MACSYMA.
   Mathlab Group, Laboratory for Computer Science, MIT, March 18, 1983.

Megatek Corporation. TEMPLATE Version 3.0 Training Manual. April 1983.

Megatek Corporation. TEMPLATE Version 3.0 Reference Manual. December 1983.

NASA/Goddard Space Flight Center (GSFC). Proposed NASA Contribution to the
   Climate Program. July 1977.

Novak, Lawrence V., Treinish, Lloyd A., and Smith, Paul H. "A New Tool for
   the Management and Study of Climate-Related Data." Presented at the
   joint session of the Eighth Conference on Probability and Statistics in
   the Atmospheric Sciences and the Third Conference on Applied
   Climatology, November 17, 1983, American Meteorology Society.

OAO Corporation (OAO). User Requirements for a NASA Climate Data Base
   Management System. A report to NASA/GSFC, October 1979.

Oracle Corporation. Oracle User's Guide (Version 3). March 1983.

Reph, Mary G. (Ed.). NASA Climate Catalog. NASA/Goddard Space Flight Center,
   February 1984.

Reph, Mary G. (Ed.). <u>Pilot Climate Data System Update Facility User's Guide</u>.
    NASA/Goddard Space Flight Center, June 1983.

Reph, Mary G., Treinish, Lloyd A., and Bloch, Lewis. <u>Pilot Climate Data
    System User's Guide</u>. NASA/Goddard Space Flight Center, May 1984.

Reph, Mary G., Treinish, Lloyd A. and Smith, Paul H. "The Pilot Climate Data
    System," To be presented at the Ninth William T. Pecora Memorial Remote
    Sensing Symposium, October 1984.

Thomas, Valerie and Guertin, Florian E. "Standardization of Computer
    Compatible Tape Formats for Remote Sensing Data." 1981 IEEE
    International Geoscience and Remote Sensing Symposium (IGARSS '81),
    June 8-10, 1981, <u>IEEE Catalog No. 81CH1656-8</u>, pp. 834-839.

Treinish, Lloyd A. "A General Display System for the Study of Climate-Related
    Data." Presented at the First Conference of the TEMPLATE User Network,
    January 25, 1984.

Treinish, Lloyd A., Novak, Lawrence V. and Smith, Paul H. "A New Data
    Management Tool to Support the Study of Climate-Related Data."
    Presented at the Spring 1983 Conference of the American Geophysical
    Union, June 2, 1983.

World Climate Project (WCP). <u>The International Satellite Cloud Climatology
    Project</u>. ICSU-WMO, Washington, D.C., 1982.

VIII. ACKNOWLEDGEMENTS

CR Categories and Subject Descriptors: E.1[Data Structures], H.2.4[Database Management]: Systems, H.2.7[Database Management]: Database Administration, H.3.0[Information Storage and Retrieval]: General, H.4.2[Information System Applications]: Types of Systems, I.3.4[Computer Graphics]: Graphics Utilities - Application packages, J.2[Physical Sciences and Engineering]: Earth and atmospheric sciences.

General Terms: Information System, Data Management, Data Analysis.

Additional Key Words and Phrases: Scientific Information Management, Scientific Data Analysis, Climatology.

FIGURES

33

PILOT CLIMATE DATA SYSTEM (PCDS)

FIGURE 1:
Pilot Climate Data System (PCDS) -
Overview
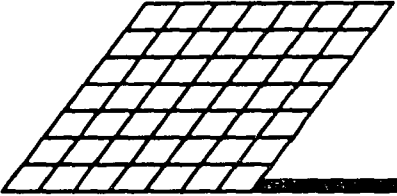
PILOT CLIMATE DATA SYSTEM (PCDS)
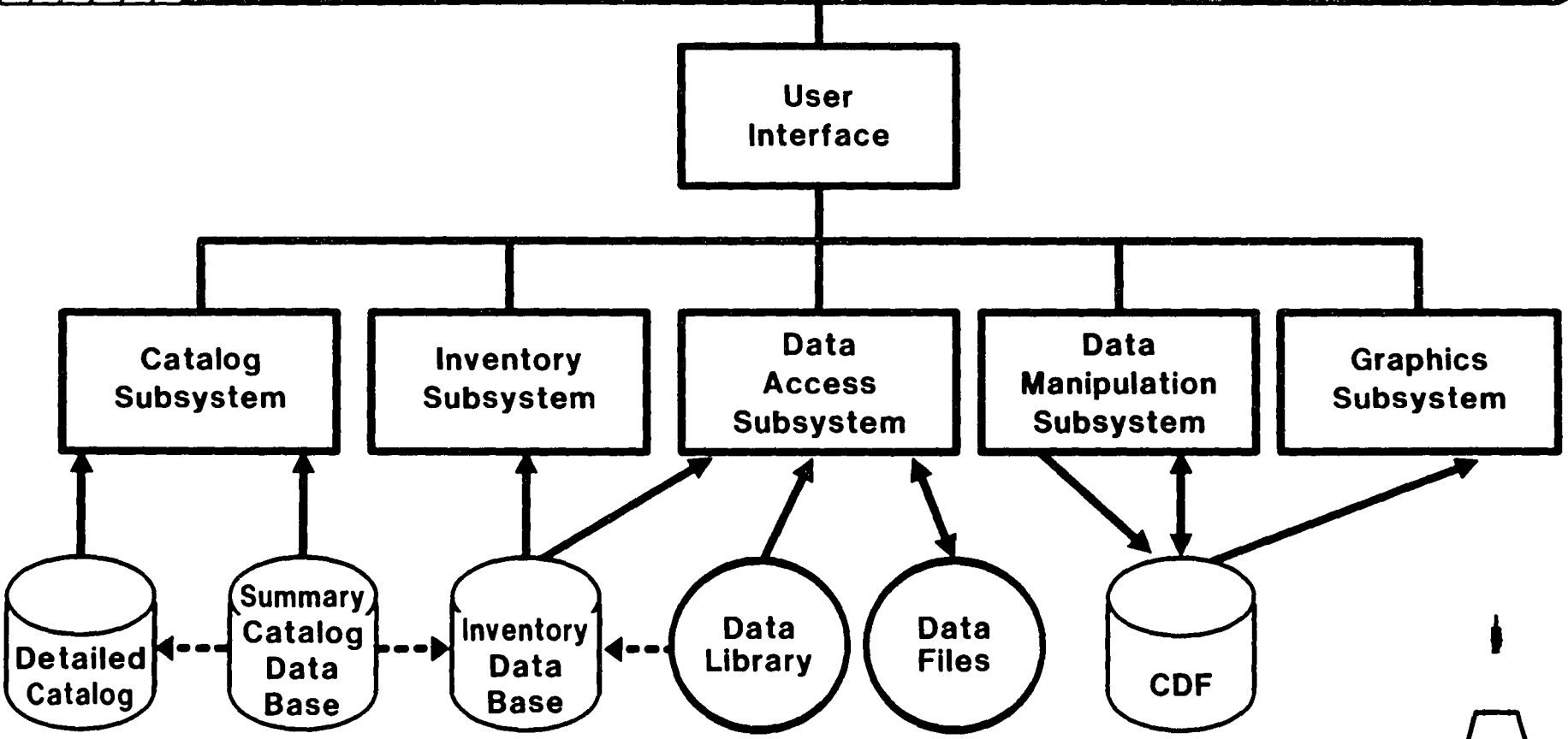
Data Structure of the PCDS

FIGURE 2:

Structure of the PCDS
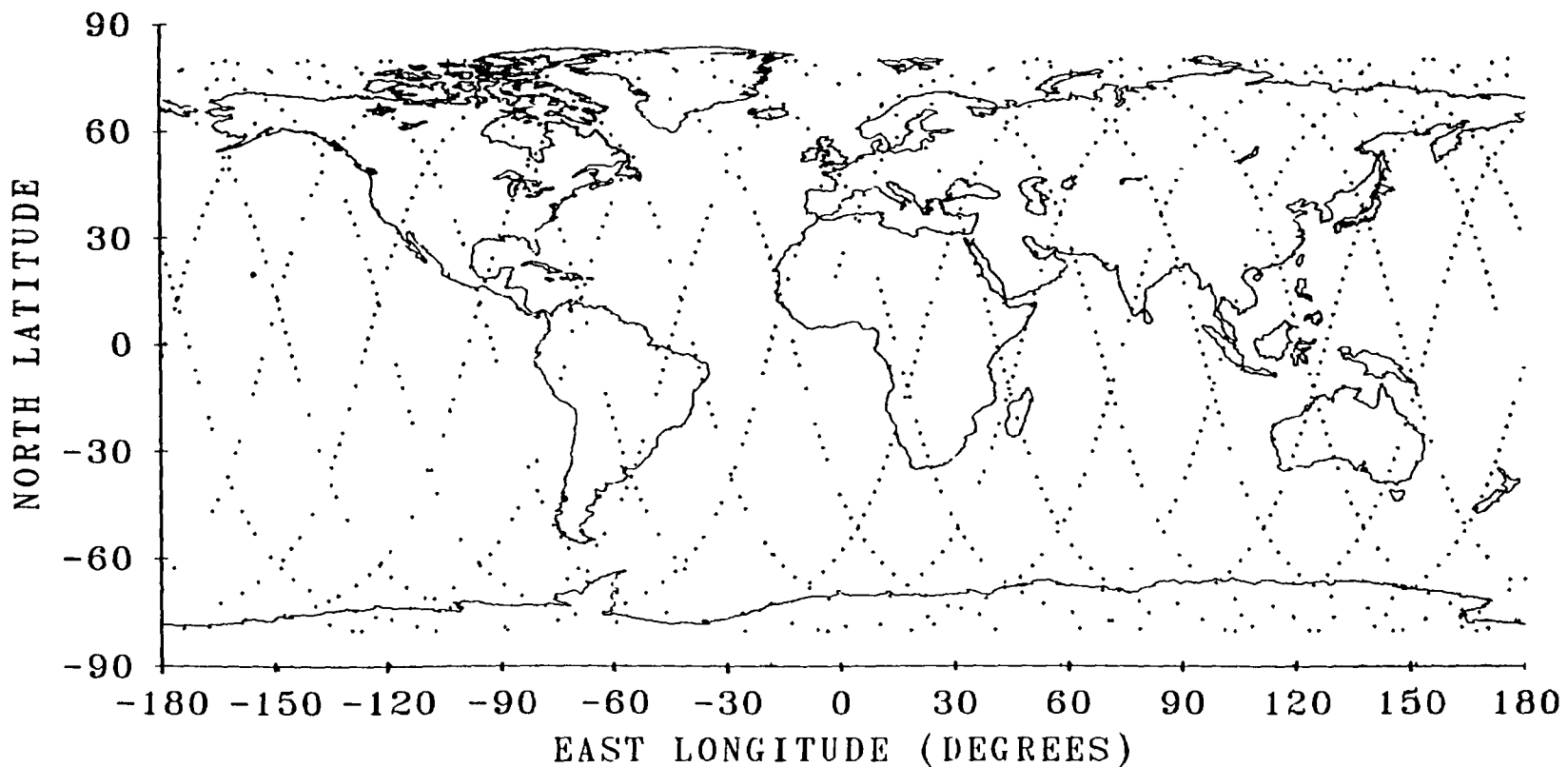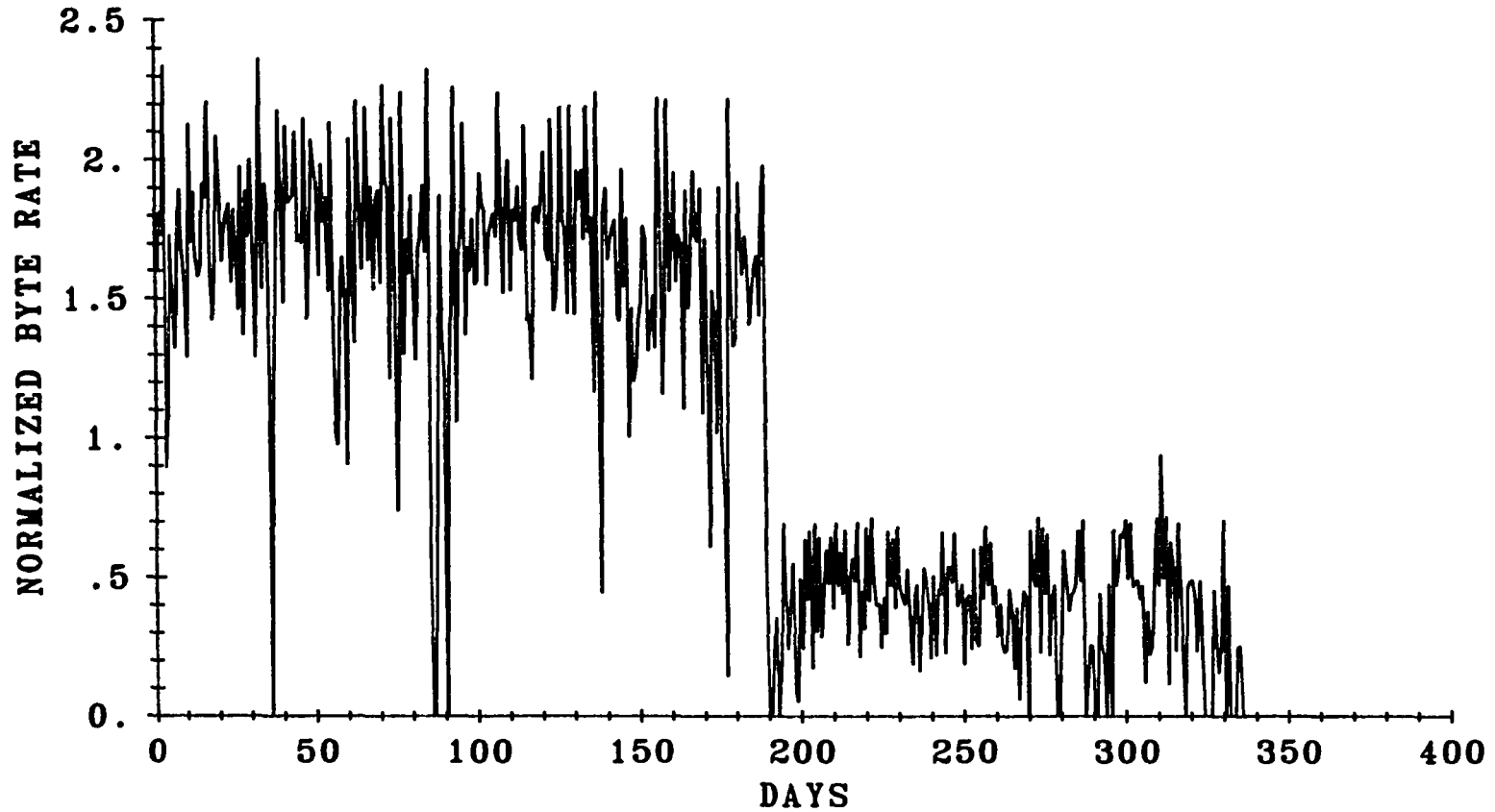
FIGURE 4:
Summary of the 1972 Inventory of the
Nimbus-4 BUV Detailed Profile Data
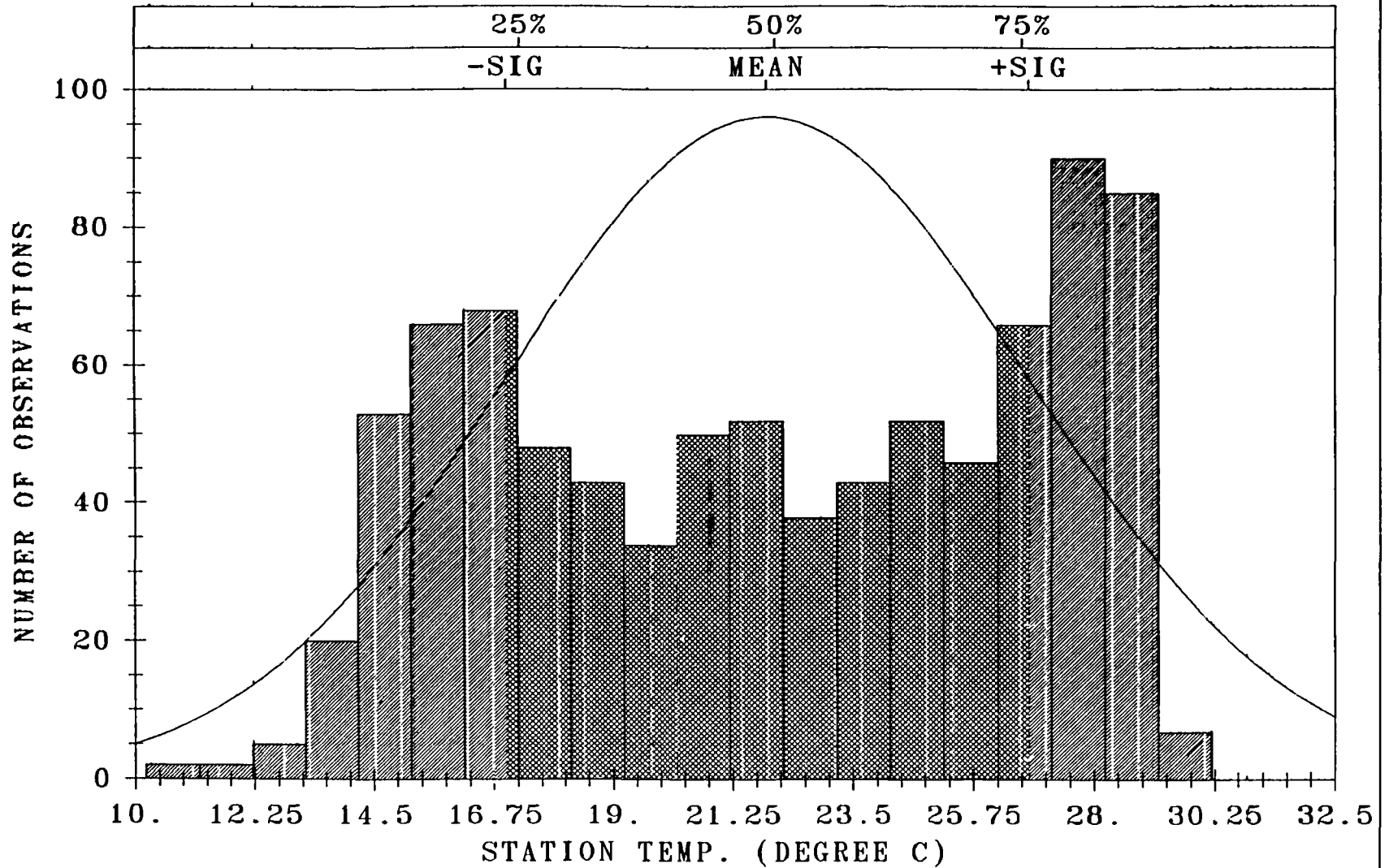
FIGURE 5:
Surface Temperature for Taipei China:
January 1900 until December 1972
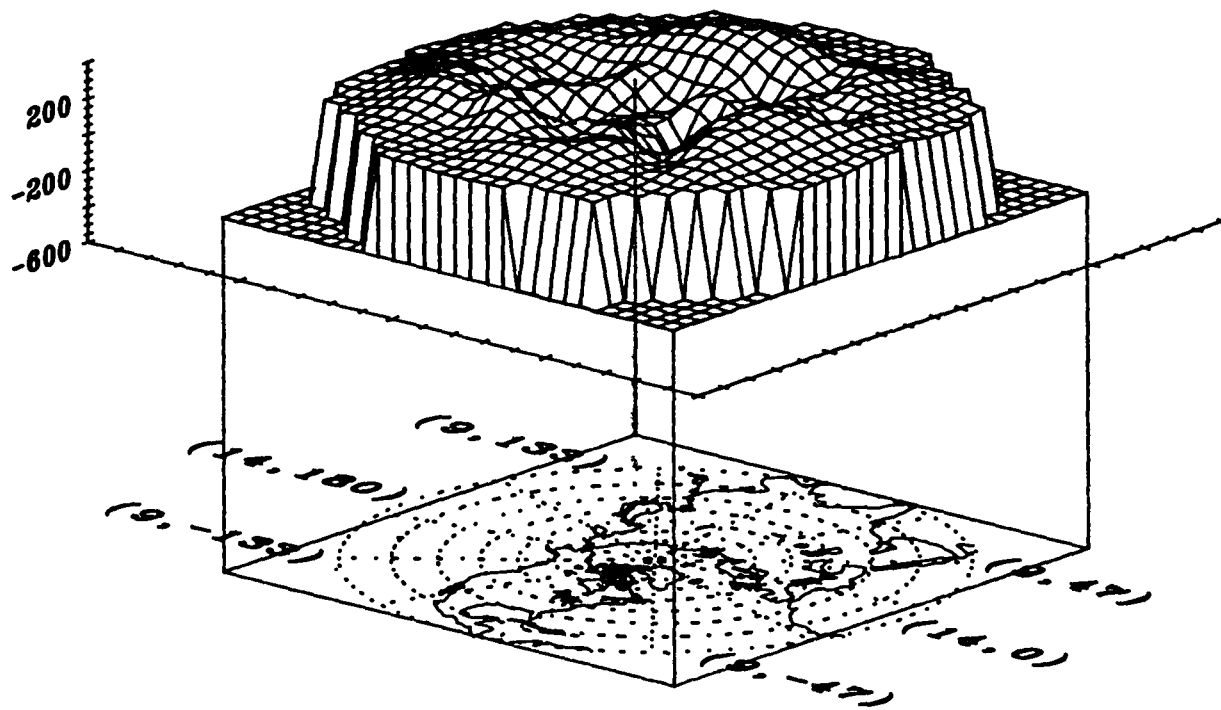
FIGURE 6:
Geopotential Height (Meters): Distribution
over the North Pole for January 1, 1975
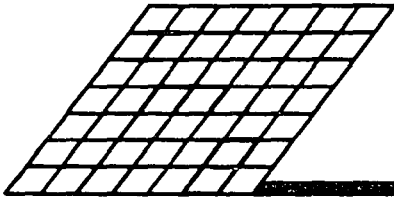
PCDS CLIMATE DATA FILE (CDF)

A Data-Independent, Self-Describing and Self-Consistent File Structure for the Support of Generalized Data Handling
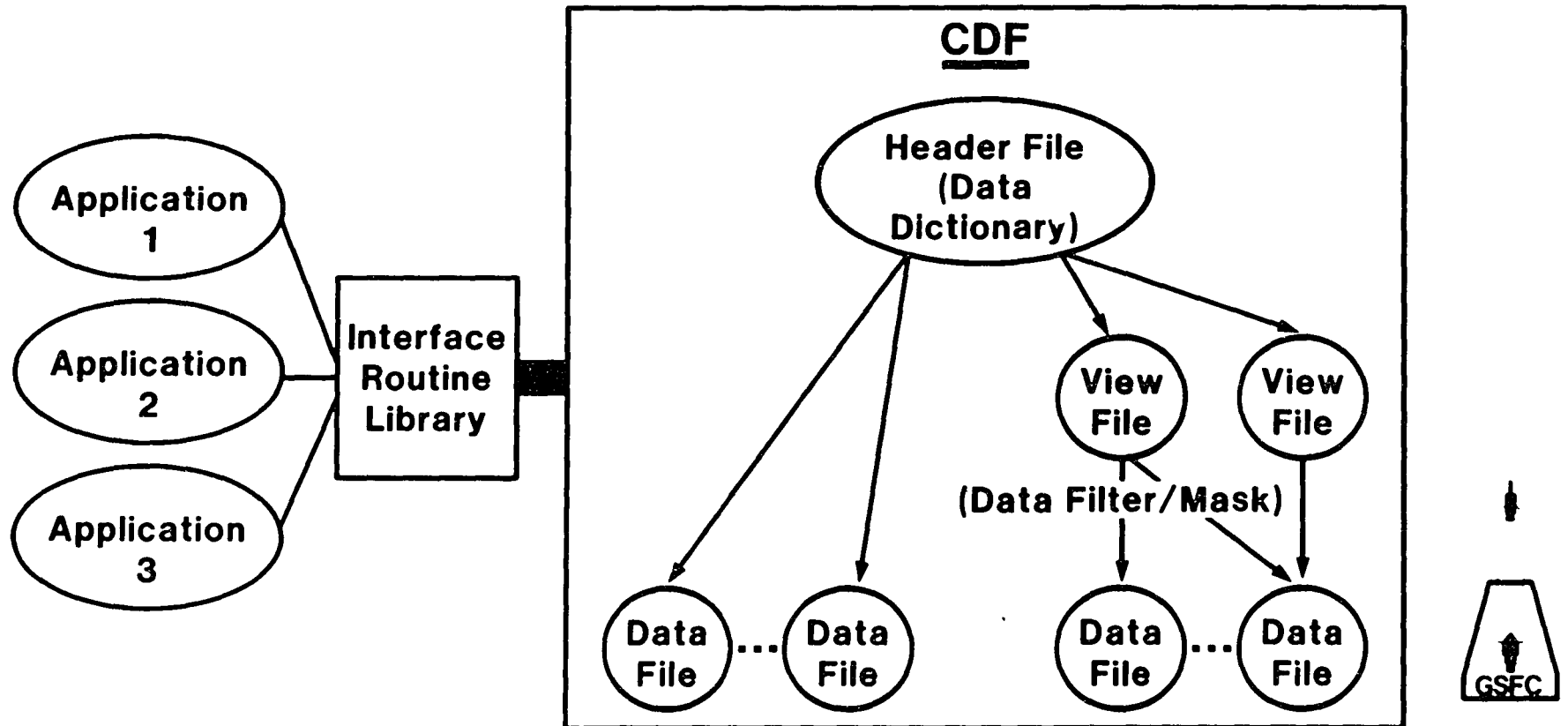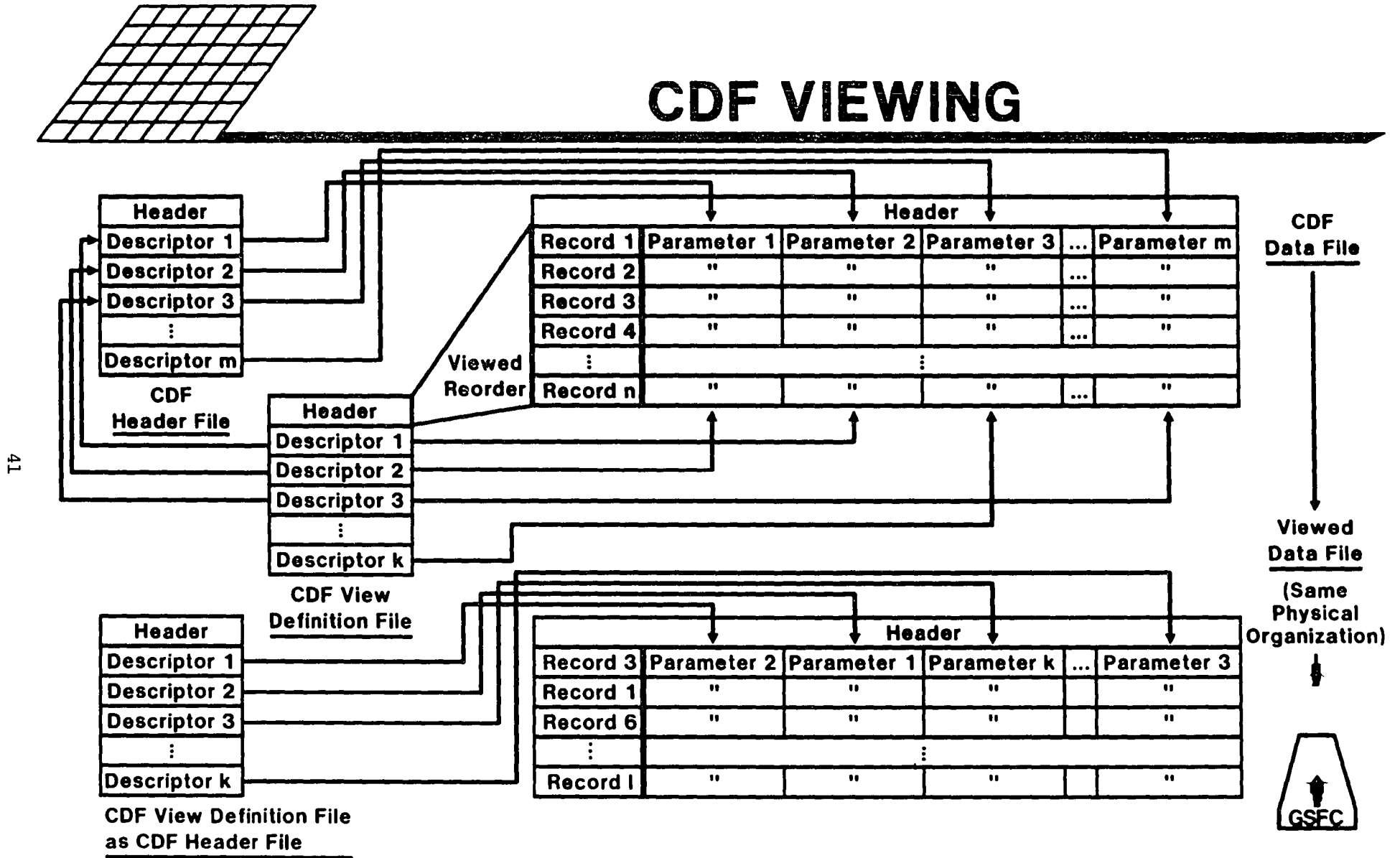
FIGURE 7:

PCDS Climate Data File (CDF)

CDF VIEWING

Header					
Record 1	**Parameter 1**	**Parameter 2**	**Parameter 3**	**...**	**Parameter m**
Record 2	"	"	"	...	"
Record 3	"	"	"	...	"
Record 4	"	"	"	...	"
⋮			⋮		
Record n	"	"	"	...	"

Header					
Record 3	**Parameter 2**	**Parameter 1**	**Parameter k**	**...**	**Parameter 3**
Record 1	"	"	"		"
Record 6	"	"	"		"
⋮			⋮		
Record l	"	"	"		"

Header
Descriptor 1
Descriptor 2
Descriptor 3
⋮
Descriptor m

CDF Header File

Header
Descriptor 1
Descriptor 2
Descriptor 3
⋮
Descriptor k

CDF View Definition File

Header
Descriptor 1
Descriptor 2
Descriptor 3
⋮
Descriptor k

CDF View Definition File as CDF Header File

Viewed Reorder

CDF Data File

Viewed Data File

(Same Physical Organization)

GSFC

FIGURE 8:

CDF Viewing

41

TABLES

TABLE I

Sample Summary Listing from the Catalog Subsystem

PARAMETER: OZONE
          (Total & Vertical Distributions)

LEVEL: II   SENSOR: BUV   MISSION: NIMBUS-4

SPATIAL COVERAGE   Global, 40mb – 0.4mb; horizontal: 200km x 200km,
 AND RESOLUTION: vertical: 2.5km

TEMPORAL COVERAGE   START TIME: 04/1970   END TIME: 05/1977
 AND RESOLUTION: 6 days for global coverage, daylight only;
                     32 sec/ observation

TAPE PRODUCTS: Detailed total ozone (DTOZ): 16 tapes/ 7 yrs; Compressed
                     total ozone (CTOZ): 4 tapes/ 7 yrs; Detailed profiles with
                     intermediate products (DPFL): 37 tapes/ 7 yrs: Compressed
                     profiles (CPFL): 4 tapes/ 7 yrs

ARCHIVE: NSSDC & PCDS   CATALOG REFERENCE: OZ/BN
ARCHIVE STATUS: Available

TABLE II

Outline of PCDS Catalog Descriptions

1. TYPE OF DATA
   1.1 Parameter/Measurement
   1.2 Unit of Measurement
   1.3 Data Source
   1.4 Data Set Identification

2. SPATIAL CHARACTERISTICS
   2.1 Spatial Coverage
   2.2 Spatial Resolution

3. TEMPORAL CHARACTERISTICS
   3.1 Temporal Coverage
   3.2 Temporal Resolution

4. INSTRUMENT DESCRIPTION
   4.1 Mission Objectives
   4.2 Key Satellite Flight Parameters
   4.3 Principles of Operation
   4.4 Instrument Measurement Geometry

5. DATA PROCESSING SEQUENCE
   5.1 Processing Steps and Data Sets
   5.2 Derivation Techniques/Algorithms
   5.3 Special Corrections/Adjustments
   5.4 Processing Changes

6. QUALITY ASSESSMENT
   6.1 Data Validation by Producer
   6.2 Confidence Level/Accuracy Judgment
   6.3 Usage Guidance

7. CONTACTS FOR DATA PRODUCTION INFORMATION

8. OUTPUT PRODUCTS AND AVAILABILITY
   8.1 Tape Products
   8.2 Film Products
   8.3 Other Products

9. DATA ACCESS
   9.1 Archive Identification
   9.2 Procedures for Obtaining Data
   9.3 PCDS Status/Plans

10. CONTACTS FOR ARCHIVE/DATA ACCESS INFORMATION

11. REFERENCES
    11.1 Satellite/Instrument/
         Data Processing
         Documentation
    11.2 Journal Articles and
         Study Reports
    11.3 Archive/DBMS Usage
         Documentation

12. RELATED DATA SETS

13. SUMMARY/SAMPLE

14. NOTES

TABLE III

Data Sources and Parameters Currently Described in the PCDS Catalog

<u>SOURCES</u>

AEM-2
First Global Atmospheric Research
   Program Global Experiment (FGGE)
GOES (1-6)
ITOS-1
LANDSAT
Nimbus-4
Nimbus-5
Nimbus-6

Nimbus-7
National Meteorology Center
   (NMC) analyses
National Oceanic and
   Atmospheric Administration
   (NOAA) analyses
NOAA missions (1-7)
OSTA-1
SEASAT
TIROS-N
World Meteorological
   Organization (WMO) surface
   stations

<u>PARAMETERS</u>

Albedo
Carbon dioxide
Chlorophyll concentration
Cloud cover
Forest cover
Geopotential height
Humidity
Ice sheet
Nitric acid
Nitrogen dioxide
Ozone
Precipitation

Radiation budget
Sea ice concentration
Sea surface elevation
Sea surface temperature
Snow coverage
Solar flux
Stratospheric aerosols
Surface pressure
Temperature
Wave height
Weather variables
Wind speed

45

TABLE IV

Sample Tape Listing from the Inventory Subsystem

***************************** DTYPE = DPFL   **************************

ITEM	PARM	PARAMETER		MISSION & SENSOR
SCAN	OZONE	OZONE		NIMBUS-4 BUV

TAPEID	ORBIT RANGE	--TAPE TIME RANGE-- START/END	# OF FILES	-GENERATION TIME-	-INVENTORY TIME-
P0515	8493	1972/01/01 01:26:08	365	80/05/26 00:00:00	83/08/23 14:08:10
	8907	1972/01/31 21:54:09			
P0516	8909	1972/02/01 00:41:37	336	80/05/26 00:00:00	83/08/23 14:39:29
	9294	1972/02/29 17:14:58			
P0517	9299	1972/03/01 01:56:02	329	80/05/26 00:00:00	83/08/23 15:15:40
	9696	1972/03/30 17:29:23			
P0518	9715	1972/04/01 01:24:03	377	80/05/26 00:00:00	83/08/23 15:51:31
	10115	1972/04/30 22:38:44			
P0519	10118	1972/05/01 01:38:28	380	80/05/26 00:00:00	83/08/23 16:22:44
	10532	1972/05/31 22:23:49			
P0520	10533	1972/06/01 00:07:49	325	80/05/26 00:00:00	83/08/23 16:51:22
	10934	1972/06/30 22:21:58			
P0521	10936	1972/07/01 01:20:38	512	80/05/26 00:00:00	83/08/23 17:14:42
	12985	1972/11/30 15:59:08			

NUMBER OF TAPES =   7   NUMBER OF FILES = 2624

TABLE V

Experiments Supported by the Catalog, Inventory and Data Access Subsystems

o Applications Explorer Mission-2 Stratospheric Aerosol and Gas Experiment (SAGE)

o Nimbus-4 Backscatter Ultraviolet (BUV)

o Nimbus-4/5 Selective Chopping Radiometer (SCR)

o Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR)

o Nimbus-7 Limb Infrared Monitor of the Stratosphere (LIMS)

o Nimbus-7 Solar Backscatter Ultraviolet (SBUV)

o Nimbus-7 Total Ozone Mapping Spectrometer (TOMS)

o Nimbus-7 Temperature Humidity Infrared Radiometer (THIR)

o Nimbus-7 Earth Radiation Budget (ERB)

o Nimbus-7 Stratospheric Aerosol Measurement (SAM II)

o Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR)

o National Meteorological Center (NMC) Daily Analyses of Atmospheric Parameters

o World Monthly Surface Station Climatology

o First Global Atmospheric Research Program Global Experiment (FGGE)

o National Oceanic and Atmospheric Administration Heat Budget Data

o Middle Atmosphere Electrodynamics (MAE - miscellaneous sounding rocket data sets)

47

TABLE VI

CDF Data Constructs
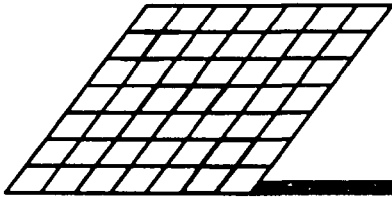
CDF DATA CONSTRUCTS

Dimensions of Data Supported	Data Type	Graphic Examples
1	"Flat" Data	Histogram
2	Time Histories Atmospheric Profiles Zonal Means	X-Y Plot
3	Grids/Images Zonal Profiles Zonal Histories Profile Histories	Contour Plot 3D Surface Diagram Pseudo-Color Image
4	Grid/Image Histories Gridded Profiles Zonal Profile Histories	Animation 3D Surface with Pseudo-Color
5	Gridded Profile Histories	Animated 3D Surface with Pseudo-Color
.	.	.
N	Multiple Grid/Image Planes/ Profiles/Histories	

GSFC

TABLE VII

NASA/GSFC Research Supported by the PCDS

o Statistical Climatology

o International Satellite Cloud Climatology Project

o Solar Flux

o Global Ozone Distribution

o Earth Radiation Budget Studies

o Middle Atmosphere Electrodynamics

o Stratospheric Photochemistry

o Distribution of Gravity/Magnetic Anomalies

o Vegetation Biomass in North American

o Land Surface/Soil Moisture Distribution

BIBLIOGRAPHIC DATA SHEET

1 Report No TM-86152	2 Government Accession No	3 Recipient's Catalog No.
4 Title and Subtitle A General Scientific Information System to Support the Study of Climate-Related Data		**5. Report Date** September 1984
		6. Performing Organization Code 634
7 Author(s) Lloyd A. Treinish		**8 Performing Organization Report No.**
9 Performing Organization Name and Address National Space Science Data Center Data Management Systems Facility NASA/Goddard Space Flight Center Greenbelt, MD 20771		**10 Work Unit No**
		11. Contract or Grant No
		13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address		
		14. Sponsoring Agency Code

15 Supplementary Notes

16 Abstract

The Pilot Climate Data System (PCDS) is being developed by NASA to manage a large collection of climate-related data of interest to the research community. The PCDS provides uniform data catalogs, inventories, access methods, manipulation tools and graphical displays for selected NASA and non-NASA data sets. The current capabilities of the PCDS include many tools for the survey of climate data. Utilities are available to browse, scan, study, select, filter, list, compress, perform statistics and display any data of interest. The PCDS now supports approximately twenty different data sets and is being used by several different in-house research groups. These capabilities are integrated into a flexible, easy-to-use information system.

17. Key Words (Selected by Author(s)) Scientific Information Management, Scientific Data Analysis, Data Base Management, Data Structures, Computer Graphics and Climatology	**18 Distribution Statement**	

19 Security Classif. (of this report) UNCLASSIFIED	20. Security Classif (of this page) UNCLASSIFIED	21 No of Pages 51	22. Price*

End of Document