2

BN-SA--3680-HEDR

DE93 005667

Pre-Decisional Draft

Benchmarking Studies for the DESCARTES and CIDER Codes

Hanford Environmental Dose
Reconstruction Project

P. W. Eslinger
S. J. Ouderkirk
W. E. Nichols

January 1993

Letter report for
the Technical Steering Panel
and the Center for Disease Control
under Contract 200-92-0503(CDC)/18620(BNW)

Battelle
Pacific Northwest Laboratories
Richland, Washington 99352

PREFACE

The primary objective of the Hanford Environmental Dose Reconstruction (HEDR) Project is to estimate the radiation dose that individuals could have received as a result of emissions since 1944 from the U.S. Department of Energy's (DOE) Hanford Site near Richland, Washington. An independent Technical Steering Panel (TSP) directs the project which is conducted by Battelle Pacific Northwest Laboratories (BNW).

One of the major objectives of the HEDR Project is to develop several computer codes to model the airborne releases, transport and environmental accumulation of radionuclides resulting from Hanford operations from 1944 through 1972. In July 1992, the HEDR Project Manager determined that the computer codes being developed (DESCARTES, calculation of environmental accumulation from airborne releases, and CIDER, dose calculations from environmental accumulation) were not sufficient to create accurate models.

A team of HEDR staff members developed a plan to assure that computer codes would be developed to meet HEDR Project goals. The plan consists of five tasks: 1) code requirements definition, 2) scoping studies, 3) design specifications, 4) benchmarking, and 5) data modeling. This report defines the benchmarking studies done for the DESCARTES and CIDER computer codes.

The work documented in this report was performed concurrently with the other tasks. Information developed in the course of the work for this document influenced the course of other scoping studies and vice versa. In the interest of prompt interaction between the HEDR staff, the TSP, and the public, this report is being issued in its current form. The reader should note that recommendations on several design issues have been made to the TSP, in part as a result of the benchmarking described here. The recommendations pertinent to this report are reducing the number of nuclides from three to one (iodine-131), using weekly data for the iodine-131 emitted in the 1940s and monthly thereafter, and dividing the geographic area in which a person may have received a dose of radiation into 1064 sections instead of the 2091 suggested by the draft requirements.

The effects of these recommendations are thoroughly discussed in the following reports:

- "Determination of Radionuclides and Pathways Contributing to Cumulative Dose," BN-SA-3673 HEDR
- "Determination of Radionuclides and Pathways Contributing to Dose in 1945," BN-SA-3674 HEDR

- "Determination of the Spatial Resolution Required for the HEDR Dose Code," BN–SA-3677 HEDR

- Determination of the Temporal Resolution Required for the HEDR Dose Code," BN–SA-3682 HEDR

GLOSSARY

C	-	Programming language
CIDER	-	Calculation of Individual Doses from Environmental Radionuclides (computer code)
deminimus	-	Radionuclide concentration below which there are no measurable health effects
DESCARTES	-	Dynamic EStimates of Concentrations and Accumulated Radionuclides in Terrestrial_EnvironmentS (computer code)
dose code	-	Computer code used to calculate radionuclides doses
gigabyte(Gb)	-	One billion bytes
kilobyte (Kb)	-	One thousand bytes
modeling domain	-	Geographic area potentially affected by Hanford emissions
nodes	-	Units in a grid used to identify a local area of radionuclide concentration
Pilot	-	Computer code used to run case studies needed to develop the sensitivity/uncertainty analysis plan
RATCHET	-	Regional Atmospheric Transport Code for Hanford Emissions Tracking (computer code)
system clock time	-	Amount of time that the computer system is dedicated to solving an application
wall clock time	-	Actual elapsed time

CONTENTS

TABLES

INTRODUCTION

The Hanford Environmental Dose Reconstruction (HEDR) project is developing several computer codes to model the airborne release, transport, and environmental accumulation of radionuclides resulting from Hanford operations from 1944 through 1972. In order to calculate the dose of radiation a person may have received in any given location, the geographic area addressed by the HEDR Project will be divided into a grid. The grid size suggested by the draft requirements contains 2091 units called nodes.

Two of the codes being developed are DESCARTES and CIDER. The DESCARTES code will be used to estimate the concentration of radionuclides in environmental pathways from the output of the air transport code RATCHET. The CIDER code will use information provided by DESCARTES to estimate the dose received by an individual.

The requirements that Battelle (BNW) set for these two codes were released to the HEDR Technical Steering Panel (TSP) in a draft document on November 10, 1992. It was not apparent whether the DESCARTES and CIDER codes could be designed to meet all of the draft requirements. This document reports on the preliminary work performed by the code development team to determine if the requirements could be met. The document contains three major sections: 1) a discussion about the merits of using a commercial database manager to handle some of the data storage and retrieval functions, 2) studies concerning disk size and computation time for the DESCARTES computer code, and 3) studies concerning computation time for the CIDER computer code.

USE OF A COMMERCIAL DATABASE MANAGER

Early in the preliminary design considerations for DESCARTES and CIDER the possibility of using a commercial database manager was considered. The database manager would have been used to store, retrieve, and perform reporting for all input and output data for the DESCARTES and CIDER codes. After some consideration, the code development team decided that a commercial database manager will not be used. Instead, the codes will be designed to access data directly from structured files. This decision was supported by all the software professionals on the external peer review panel of the code development task that met on November 5th and 6th, 1992, in Richland.

The decision to use structured files instead of a commercial database manager took into consideration the client's needs, the cost, the disk storage capacity, and the amount of time required to execute the computer code. The principal reasons for this decision are:

- Delivery of the dose code computer codes used to determine the radiation dose (referred to as the "dose code" for short) in the "turnover package" to the client with an embedded software product would require the client to purchase a license for the embedded product. This approach does not appear to be currently acceptable to the client.

- The use of structured files permits all needed data to be read in CIDER in one- (or a few-) bit streams which allows direct input from disk to memory at near the maximum rate possible for the HEDR Sun 690 computer. This approach would likely not be possible with a database manager thus it could have and would, therefore, result in a much slower data transfer period.

- A database manager would use approximately an additional 20 percent of disk space. Since the amount of available disk space is already a major factor in the development effort, any additional use of disk space is a major concern.

- The data set for the dose code is static once it is built. The database manager would be used only for storing and retrieving data. No data manipulations would be performed. Thus, some of the major strengths of a database manager would not even be needed.

- Use of a database manager would add significant costs to the project. The major costs would be incurred in purchasing the product and training staff to use the product. Also, it could take longer to develop the computer code necessary to access the database functions from DESCARTES and CIDER than if a structured file approach were used.

BENCHMARKING FOR THE DESIGN OF DESCARTES

The computer code development team conducted two benchmark tests of the DESCARTES code to provide assurance that the code could be designed to meet the draft requirements. The first benchmark dealt with determining the disk space storage space required to make a production run for a single radionuclide. There is some concern that the 24 Gb of on-line storage available on the HEDR Sun 690 computer will be inadequate to support production runs of the DESCARTES code. The second benchmark dealt with estimating the computation time required to perform the environmental accumulation calculations. The timing requirement being addressed by this benchmark test is draft requirement 14.1:[a]

> "The environmental software shall be capable of producing a full environmental accumulation database for use by the dose code (3 radionuclides, all media, 2091 nodes, years 1944 through 1972) in 30 days wall clock time."

DISK STORAGE SPACE REQUIREMENTS

The DESCARTES data size benchmark was designed to estimate the disk storage space required to operate the code in a production mode. A single DESCARTES run will compute the environmental accumulation of a single radionuclide over the time period from 1944[b] through 1972 for 100 realizations. The storage requirements for this benchmark case were determined through a paper study by the development team.

All data size estimates are directly dependent upon the number of nodes (geographic units) in the modeling domain and the time increment to be used for data output by DESCARTES that will be used in CIDER. For this benchmark, the modeling domain is considered to contain 2091 nodes and data will be output by DESCARTES in daily increments for the time period 1945 through 1949 and weekly thereafter until 1972. An alternative considered is for the output interval to contain weekly increments from 1945 through 1949 and monthly increments thereafter until 1972 in a domain containing 1064 nodes.

The first disk storage requirement is for the data from the RATCHET code. Two parameters, air concentration and air deposition, will be required to perform runs of the

(a) Letter, Software Requirements Specification for the Hanford Environmental Dose Reconstruction (HEDR) Project Air Pathway Environmental Accumulation and Dose Codes," from D. B. Shipler (BNW), to J. E. Till (TSP) and M. R. Donnelly (CDC), November 10, 1992.

(a) For the purposes of estimating data sizes we ignore the limited data in 1944.

DESCARTES code. The storage requirements for the air input data are delineated in Table 1. For the baseline, the total storage required for air data is over 5 Gb. If the number of nodes is reduced to 1064, the total storage required is about 2.53 Gb.

DESCARTES is required to store 31 spatially and temporally distributed parameters, although five of these are not required as input for CIDER. Each parameter may have a different number of nodes and/or a different time increment associated with it. Table 2 outlines the total data storage requirements for these 31 parameters with baseline time increments. Table 3 outlines the total data storage required for these parameters under the alternative time-increment scheme. From Table 2, the total disk storage required to store one radionuclide environmental accumulation data output set is roughly 45 Gb as compared to approximately 4 Gb for the alternative domain size and time increment scheme (Table 3).

The data size estimates given in Tables 2 and 3 assume that every node in the domain will have nonzero data associated with it for every time increment simulated between 1945 and 1972. Given fluctuating wind patterns and declining source-term releases after 1951, many nodes may have an effective zero concentration associated with them for some time periods. The code is being designed to take advantage of the zero data and reduce disk storage requirements accordingly.

The total disk space requirement during a DESCARTES production run will be the sum of the input and output file sizes. We assume for this benchmark that the other input files (run control, commercial distribution, animal diet, and plant-specific parameters) will require less than 1 Gb of disk space regardless of the time-increment chosen.

TABLE 1. Baseline RATCHET Data Size

			1945-1949			1950-1972		
Parameters	Number of Nodes	Time Incre-ments per Year	Number of Years	Subtotal Gbytes	Time Incre-ments per Year	Number of Years	Subtotal Gbytes	Total Gbytes
Air Concentration	2091	365	5	1.53	52	23	1.00	2.53
Aerial Deposition	2091	365	5	1.53	52	23	1.00	2.53
				3.06			2.00	5.06

TABLE 2. Baseline DESCARTES Data-Output Size

			1945-1949			1950-1972		
Parameters	Number of Nodes	Time Increments per Year	Number of Years	Subtotal Gbytes	Time Increments per Year	Number of Years	Subtotal Gbytes	Total Gbytes
Air Concentration	2091	365	5	1.526	52	23	1.0	2.526
Upper Soil	2091	365	5	1.526	52	23	1.0	2.526
Root Zone Soil	2091	365	5	1.526	52	23	1.0	2.526
Other Vegetables	2091	90	5	0.376	13	23	0.250	0.626
Grain	2091	1	5	0.004	1	23	0.019	0.023
Milk - Regime 1 Ind.	200	365	5	0.146	52	23	0.096	0.242
Milk - Regime 1 Herd	200	365	5	0.146	52	23	0.096	0.242
Milk - Regime 2 Ind.	200	365	5	0.146	52	23	0.096	0.242
Milk - Regime 2 Herd	200	365	5	0.146	52	23	0.096	0.242
Milk - Regime 3 Ind.	2091	365	5	1.526	52	23	1.0	2.526
Milk - Regime 3 Herd	2091	365	5	1.526	52	23	1.0	2.526
Milk - Regime 4 Ind.	2091	365	5	1.526	52	23	1.0	2.526
Milk - Regime 4 Herd	2091	365	5	1.526	52	23	1.0	2.526
Grocery Milk - Urban	2091	365	5	1.526	52	23	1.0	2.526
Grocery Milk - Rural	2091	365	5	1.526	52	23	1.0	2.526
Goat Milk	2091	365	5	1.526	52	23	1.0	2.526
Creamery Milk	26	365	5	0.019	52	23	0.012	0.031

TABLE 2. (contd)

Beef	2091	365	5	1.526	52	23	1.0	2.526
Poultry	2091	365	5	1.526	52	23	1.0	2.526
Eggs	2091	365	5	1.526	52	23	1.0	2.526
Outer Leafy Veg.	2091	180	5	0.753	26	23	0.500	1.253
Outer Leafy Veg. - Comm.	200	180	5	0.072	26	23	0.048	0.120
Inner Leafy Veg.	2091	180	5	0.753	26	23	0.500	1.253
Inner Leafy Veg. - Comm.	200	180	5	0.072	26	23	0.048	0.120
Inner Fruit	2091	60	5	0.251	8	23	0.154	0.405
Outer Fruit	2091	60	5	0.251	8	23	0.154	0.405
Alfalfa	2091	3	5	0.013	3	23	0.058	0.071
Pasture	2091	365	5	1.526	52	23	1.0	2.526
Grass Hay	2091	1	5	0.004	1	23	0.019	0.023
Silage	2091	1	5	0.004	1	23	0.019	0.023
Sage-brush	2091	365	5	1.526	52	23	1.0	2.526
				26.046			17.146	45.338

TABLE 3. Alternative DESCARTES Data-Output Size

			1945-1949			1950-1972		
Parameters	Number of Nodes	Time Incre-ments per Year	Number of Years	Subtotal Gbytes	Time Incre-ments per Year	Number of Years	Subtotal Gbytes	Total Gbytes
Air Concentration	1064	52	5	0.111	12	23	0.117	0.228
Upper Soil	1064	52	5	0.111	12	23	0.117	0.228
Root Zone Soil	1064	52	5	0.111	12	23	0.117	0.228
Other Vegetables	1064	13	5	0.028	2	23	0.020	0.048
Grain	1064	1	5	0.002	1	23	0.010	0.012
Milk - Regime 1 Ind.	200	52	5	0.021	12	23	0.022	0.043
Milk - Regime 1 Herd	200	52	5	0.021	12	23	0.022	0.043
Milk - Regime 2 Ind.	200	52	5	0.021	12	23	0.022	0.043
Milk - Regime 2 Herd	200	52	5	0.021	12	23	0.022	0.043
Milk - Regime 3 Ind.	1064	52	5	0.111	12	23	0.117	0.228
Milk - Regime 3 Herd	1064	52	5	0.111	12	23	0.117	0.228
Milk - Regime 4 Ind.	1064	52	5	0.111	12	23	0.117	0.228
Milk - Regime 4 Herd	1064	52	5	0.111	12	23	0.117	0.228
Grocery Milk - Urban	1064	52	5	0.111	12	23	0.117	0.228
Grocery Milk - Rural	1064	52	5	0.111	12	23	0.117	0.228
Goat Milk	1064	52	5	0.111	12	23	0.117	0.228
Creamery Milk	26	52	5	0.003	12	23	0.003	0.006
Beef	1064	52	5	0.111	12	23	0.117	0.228

TABLE 3. (contd)

Poultry	1064	52	5	0.111	12	23	0.117	0.228
Eggs	1064	52	5	0.111	12	23	0.117	0.228
Outer Leafy Veg.	1064	26	5	0.055	6	23	0.059	0.114
Outer Leafy Veg. - Comm.	200	26	5	0.010	6	23	0.011	0.021
Inner Leafy Veg.	1064	26	5	0.055	6	23	0.0[illegible]	0.114
Inner Leafy Veg. - Comm.	200	26	5	0.010	6	23	0.011	0.021
Inner Fruit	1064	9	5	0.019	2	23	0.020	0.039
Outer Fruit	1064	9	5	0.019	2	23	0.020	0.039
Alfalfa	1064	3	5	0.006	3	23	0.029	0.035
Pasture	1064	52	5	0.111	12	23	0.117	0.228
Grass Hay	1064	1	5	0.002	1	23	0.010	0.012
Silage	1064	1	5	0.002	1	23	0.010	0.012
Sage-brush	1064	52	5	0.111	12	23	0.117	0.228
				1.960			2.105	4.065

The output of the DESCARTES code will be used to build a fast-access data set for the CIDER code. Double storage will be required for many variables when building this data set. The disk size required for the entire set of operations in DESCARTES is summarized in Table 4 for both the baseline data size and the alternative data size. The size estimates in this table do not take advantage of data size reductions that would result by eliminating nodes containing no contamination.

The disk storage requirement depends directly on the time increment used in the production DESCARTES runs. If daily environmental concentrations for the 1945 to 1949 period and weekly environmental concentrations thereafter are required, approximately 90 Gb of disk space will be required to run DESCARTES for a single radionuclide without using tape drives. If weekly output for the 1945-1949 period followed by monthly output for 1950-1972 will suffice, the production runs will require about 21 Gb. If the domain contains only 1064 nodes, then the storage requirement is reduced to about 10.3 Gb.

The HEDR Sun 690 computer currently has 24 gigabytes of direct on-line storage without resorting to tape drives. Production runs of DESCARTES and also concurrent operation of CIDER seem readily achievable if the time increments are performed on a weekly/monthly basis and the number of nodes in the domain is reduced to 1064.

TABLE 4. Disk Storage (Gb) Required for DESCARTES Production Runs

	Baseline Daily/Weekly Data	Alternative Weekly/Monthly Data
From RATCHET	5.06	2.57
In DESCARTES	45.338	4.07
To CIDER	40.26	3.55
Other Data	1.00	0.50
Total	91.658	10.69

COMPUTATION TIME REQUIREMENTS

An additional benchmark was conducted for DESCARTES to determine the time required to perform representative environmental pathway calculations. The approach based the time estimate on J. C. Simpson's Pilot computer code.

The benchmark case for the Pilot code estimates the time it takes to calculate the environmental concentrations of a single radionuclide in two soil layers, eight plant products, and eleven animal types for 1000 nodes. It also provides details on the computation time required for each module (biomass, soil, plant products, and animal products). The total time required to run this case using the Pilot code on a dedicated system (including all I/O) was 245.0 hours (wall clock time) and 39.7 hours (system clock time).

Table 5 summarizes the scaling used to extrapolate the Pilot code timing requirements to DESCARTES production run estimates assuming 1064 nodes in the modeling domain. The time estimates double if 2091 nodes are considered. Each area of computational effort is scaled separately based on the increased number of radionuclides, nodes, time increments, and categories of the particular module. The result is that for the baseline time-increment scheme, 145 days of wall clock time and 23.4 days of system clock time are required to prepare the full environmental concentration data set.

The numbers in Table 5 are only indicators based on the Pilot code, which, according to its author J. C. Simpson, is "extremely input/output bound." Significant gains will be realized in performance for the final DESCARTES code. Still, an improvement factor of 15 is necessary in wall clock time to meet software requirement 14.1. Such an improvement cannot be guaranteed. Also, the Pilot code does not perform some tasks required of DESCARTES which will consume more computer time, such as the commercial distribution of food products between nodes.

This computation time summary indicates that execution of the environmental pathway accumulation equations will probably exceed the 1-month wall clock requirement if the production run is performed for 3 radionuclides with daily output for 1945 through 1949 and weekly output from 1950 through 1972. DESCARTES would have to run at least 15 times faster than Pilot to meet the limit, and this improvement factor neglects some tasks DESCARTES will have to perform that Pilot does not, such as commercial distribution of food products.

As a mitigating factor, this timing estimate also neglects the benefit of implementing a deposition cutoff level below which computation of environmental accumulations are unnecessary. Preliminary scoping studies have indicated that most nodes can be eliminated from further consideration for time periods after 1951.

TABLE 5. DESCARTES Time Estimates

Category	Pilot Number	DESCARTES Baseline Number	DESCARTES Alternative Number
Radionuclides	1	3	1
Nodes	1000	2091	1064
Time Increments	365	2978	2978
Plant Types	8	15	15
Soil	1	1	1
Animal Types	11	15	15
Wall Clock Time Calculation Module	Pilot Code Time (hr)	DESCARTES Code Time (hr)	DESCARTES Code Time (hr)
Biomass	37.8	1,209.1	615.2
Soil	13.3	226.9	115.4
Plants	102.2	3,269.2	1,663.4
Animals	91.7	2,133.3	1,085.4
Total	245.0	6,838.5	3,479.4
Total (days)	10.2	284.9	145.0
CPU Clock Time Calculation Module	Pilot Code Time (hr)	DESCARTES Code Time (hr)	DESCARTES Code Time (hr)
Biomass	8.9	284.7	144.9
Soil	2.2	37.5	19.1
Plants	13.3	425.4	216.4
Animals	15.3	355.9	181.1
Total	39.7	1,103.5	561.5
Total (days)	1.7	46.0	23.4

Based on the following assumptions, it appears that DESCARTES could possibly meet the 10-day time requirement (30 days divided by the three nuclides) if:

- The modeling domain contains only 1064 nodes.
- The DESCARTES code runs 5 times faster on a wall clock basis than the Pilot code. Execution time will be reduced from that of the Pilot code if the data from DESCARTES is output on a weekly basis for 1945-1949 and a monthly basis for 1950-1972.
- The deminimus dose level allows elimination of two-thirds of the nodes in the domain, averaged over the entire 28 years.

If these assumptions are not met, DESCARTES will almost certainly not meet the 10-day elapsed time requirement for a single nuclide.

BENCHMARKING FOR THE DESIGN OF CIDER

The code development team decided that two aspects of the CIDER code needed to be benchmarked. The first benchmark dealt with the time required to move data from disk to memory in the computer, and the second benchmark dealt with the time required to perform the dose calculations once the data was in memory. The timing requirement being addressed by these benchmarks tests is requirement 14.2:

> The dose software shall be able to calculate and output data for a "map" run of 2091 nodes on a dedicated machine for a single year and a single radionuclide with an elapsed wall-clock time of no more than 1045.5 minutes (30 seconds average per node).

TIME REQUIRED TO MOVE DATA FROM DISK TO MEMORY

The CIDER disk access benchmark addressed the time required to move data from disk to memory. The objective of the benchmark was to determine the amount of time it would take to move 2.2 Mb of information from disk to memory. DESCARTES generates an exposure data file. The 2.2 Mb size was an estimate of the amount of data that would be loaded from the exposure data file into memory by CIDER to calculate a year's worth of doses for an individual in a given node. A "fake" data set similar to the real data set was constructed in an attempt to obtain a realistic time for the data transfer.

Three separate codes were created and executed on the HEDR Sun 690 computer during the CIDER disk access benchmark. These codes were all written in the computer language C to take advantage of the input/output features not available in FORTRAN 77. The first code generated gigabyte sized data files in the format proposed for the DESCARTES environmental output data. A second code generated an index for the environmental data file. Finally, a third code determined the amount of time required to seek a location in the data file and transfer blocks of data from the file to memory.

The first program (MK_EXP) was designed to create two sample environmental files of 2.0 Gb and 1.0 Gb, respectively. However, due to a lack of file space on the disk, the program terminated early with only 95 percent of the data generated. Since a complete file was not required for the other benchmark tests we decided to use these partial files. The alternative was to force other users to delete files to free disk space and then re-run the code.

The second program (BLD_NDX) generated an index file associated with the environmental file. The purpose of the index file is to keep the address of the location of the data blocks in the large environmental data file. This code has the potential to work unmodified for the actual environmental data files created by DESCARTES.

The third program (BENCH_1) estimated seek and load times. A total of six runs of this code were performed over two days to obtain an estimate of disk performance with the proposed environmental file structures. The results are presented in Table 6.

The initialization time is the time required to create the data structures needed to hold the file specification information, load the index file into memory, and initialize arrays. Less than one second is required to perform this function ten times.

The seek time is the time required to access, read, and verify that the correct location was read for 100 locations. This operation takes approximately 0.03 seconds per access. Since this is a small fraction of the time required to load the data, the load times are a reasonable estimate of a combined seek and load time.

The load times are the times required to load sequential blocks of size 1, 2, 4, 8, and 16 Mb from the disk to memory. The values in the table are the total time for ten iterations. The results indicate that the highest possible data throughput rate is obtained with smaller block transfer sizes (approx 3 Mb per second for 1-Mb blocks) with the throughput leveling off at about 1 Mb per second for the larger block sizes.

TABLE 6. Total Elapsed Time for CIDER Data Transfers

(Data in Seconds)

Description	No. of Iterations	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
Initialization time	10	0	0	0	0	0	1
Seek time	100	3	4	3	2	3	3
Load time for 1 Mb	10	3	5	4	7	9	11
Load time for 2 Mb	10	19	13	6	11	17	22
Load time for 4 Mb	10	42	43	38	29	39	44
Load time for 8 Mb	10	87	74	76	78	83	89
Load time for 16 Mb	10	167	167	169	161	168	171
Seek and load 2.2 Mb	10	20	17	16	13	12	25

The <u>seek and load</u> test is considered the most realistic test of performance for a real data file. The seek and load time is the total amount of time required to seek and load 10 non-contiguous blocks of data from the environmental data file to memory. A 2.2 Mb data block (365 days of data for a single node) was selected as the size most likely to be loaded from the real data file. The results indicate that the data throughput rate will be about 1.8 Mb per second.

The system was not dedicated to performing these tests. The lowest values obtained for all runs will most closely reflect the worst likely performance of a dedicated system. The average time for the six runs provides an indication of performance if dose code production runs are done concurrently with other activities on the system.

The highly variable test results are due to the fact that the system was not dedicated to this task and the operating system is attempting to perform additional system tasks. These results indicate that the time required to load the environmental data into memory on the average will be less than 2 seconds per year/node. The time required to transfer data from disk to memory does not appear to be large enough to cause difficulty in meeting the 30-second/node/year computation limit for a dedicated machine.

COMPUTATION TIME ESTIMATES

The second CIDER benchmark was conducted to estimate the time required to perform representative dose calculations after data were loaded into memory. The first approach was to make a time estimate using the Pilot code.

This benchmark case for the Pilot code estimates the time it takes to load data into memory, calculate the dose for 24 individuals with reference to diets and lifestyles for one year (365 days at a single node with seven pathways and one organ (thyroid)), and write the results to disk. Only a single load of exposure data for a node/year was required for this benchmark.

The total time required to run this case in the Pilot code on a dedicated system (including all I/O) was 242.29 seconds (system clock time) and 244 seconds (wall clock time). This time yields an average of about 10 seconds per node/year/individual.

If the time required to load and expand the data is removed from the estimate, the time required is 166.59 seconds (system time). This time yields an average of seven seconds per node/year/individual.

This benchmark indicates that the annual dose for <u>one</u> organ and <u>seven</u> pathways for an individual at a given node can be calculated in 7 seconds. However, there are at least ten possible pathways (all are likely in an adult diet and lifestyle) and three organs that will require dose calculations. If this time is linearly extrapolated, it will take 30 seconds to calculate and report an annual dose for a node/year/individual.

This average time for the calculation is based on a code that is not optimized for speed. We decided to determine what performance could be expected on the HEDR Sun 690 computer for the entire set of equations if the code were optimized for speed. An additional test was performed that consisted of coding one of the dose functions (CID-3) in C and calculating it 1.1 millon times (10 pathways x 3 organs x 100 realizations x 365 increments/year). Based on the time it took the HEDR Sun 690 computer to perform these calculations, the calculation will take about 7 seconds per node/year/individual for 10 pathways and three organs.

The computation time test indicates that execution of the dose equations, once all environmental and individual choice data have been placed in memory, should not greatly exceed 7 seconds. This time estimate does not include a wide variety of other functions that must be performed (e.g., error checking and obtaining age-dependent dose factors).

The total computation time required to estimate the annual dose for an individual includes time for the following actions: loading environmental data, loading creamery data (not in the data transfer timing test), performing error checking, generating stochastic values for several variables, reading diet information and dose factors from libraries, calculating the dose equations, and writing the output data to disk. The two benchmark cases performed address the computational components considered by the development team to consume most of the time in the CIDER code. Based on these tests, we feel comfortable that we can build the CIDER code to meet the 30-second time requirement.

CONCLUSIONS

The following conclusions were reached by these benchmark activities:

- The current code requirements for DESCARTES will need up to 90 Gb of disk space to run. Because the system only has 24 Gb of disk available, additional disks will have to be purchased to run in this configuration.
- By reducing of the number of grid nodes from 2091 to 1064 and the time increments from daily/weekly to weekly/monthly no more than 11 Gb will be required.
- The requirement that DESCARTES calculations be completed in 10 days cannot be achieved unless these suggested changes are accepted.
- The CIDER code will run in the current disk space assuming that less than 40 percent of the nodes are active.
- The CIDER code will run under the 30-second/node/year dose calculation.

END

DATE
FILMED
3/19/93