PA3 A GENERAL-PURPOSE, TIME-SHARED PROBLEM ANALYSIS LANGUAGE

D. G. Ebeling
Corporate Engineering
General Electric Co.
Schenectady, N.Y.

and

E. G. Hurst, Jr.
Wharton School
Univ. of Pennsylvania
Philadelphia, Pa.

Abstract

A general-purpose, time-shared language for modeling and solving problems which exhibit uncertainty in the input variables is presented. The language is described in terms of some comparative measures of simulation systems. A simple example is worked to demonstrate the major features of the language. Examples of actual problems which have been solved are described to illustrate the breadth of possible applications.

1. INTRODUCTION

A Problem Analysis language, the third version of which (PA3) the current paper describes, has been developed in response to the need for a general-purpose, time-sharing package which allows the user to obtain quick answers to a problem in which one or more of the factors may be uncertain The language has now grown so that large and complex problems can be handled, and the model can grow easily as the implications of the original model are explored using the features available. The original version of the language was developed when the senior author discovered he was doing virtually the same initializing, sampling, bookkeeping, and outputting on several of the problems he was solving by Monte Carlo methods on the time-shared computer. This first version was a single program which incorporated the common needs of various problems while allowing the underlying structure to be tailored to an individual problem. It was distributed widely in the General Electric Company, and still provides the basis in several internal courses for instruction in probabilistic modeling and simulation. The second version added greater identification ability, and the ability to perform sensitivity analyses. In several adaptions, it is being used by many of the product departments of the General Electric Company for annual budgeting. The third and current version has added greater flexibility in its sampling options, printing options, change options, and so forth, and has grown to approximately 35 time-shared files linked together by a master executive routine.

Through all versions, one important feature has remained: the requirement that the user state the structure of his particular problem. This is accomplished in a "kernel" of one or more statements in the BASIC language which relate a set of input variables called X(·) to another set of desired output variables called Y(·). Deterministic and, if appropriate, probabilistic data on the X(·) must be provided, along with identifying names of both the X(·) and Y(·). Using these basic inputs, the language interactively guides the user through a series of options which may be chosen or declined in response to computer-asked questions. The options currently available are:

(1) Data verification

(2) Deterministic analysis

(3) Sensitivity analysis on each input variable

(4) Monte Carlo simulation from a wide range of distributions yielding summary statistics along with selected automatically-scaled histograms

(5) Temporary or permanent changes in data

(6) Structural changes in the model logic

In the remainder of the paper, the language is described further by relating it to several criteria by which simulation programs are compared. A simple example is used just to demonstrate the conversational use of some of the available options. The paper concludes with a brief description of several representative examples illustrating the breadth of problems which have been solved using this language.

2. DESCRIPTION OF PA3 FEATURES

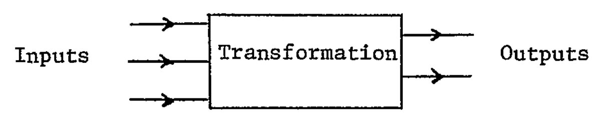
In this section of the paper, some of the detailed features of the language are presented in the context of some of the usual measures of comparison of simulation programs. These measures have been distilled and hybridized from those given in the references found at the end of the paper.

2.1 SOPHISTICATION OF INTENDED USER

The PA3 language is intended for the analyst who knows little or nothing about simulation. It was expected, and has been borne out by the applications to date, that the analyst could learn to use the language in very little additional time over and above the time required to learn the rudiments of the BASIC language and the time-shared computer. For users already familiar with these tools, almost no additional set-up time has been required to gain a working knowledge of PA3. In short, the language is intended for the direct hands-on use of the person with the problem, not a specialist to whom the problem must be described.

2.2 WORLD VIEW

The language assumes all problems to be of the form



where:

(1) The nature of the transformation is known.

(2) Some of the input variables may not be known with certainty.

(3) The uncertainty can be expressed by probability distributions.

(4) The input variables may be related to each other.

Notice that feedback and the automatic advancement of time are not implicit features of the language. Indeed, accepting the definition of simulation in Meier, Newell and Pazer[8], the program does not perform simulation but, rather, model sampling. Of course, simulation characteristics can be included by appropriate model definition, as was done in the process control example described below, but these features are not an integral part of the current system. The main intention of the language is not to simulate the temporal behavior of dynamic systems; it is intended to manipulate a system whose model is a perhaps complex set of algebraic relationships between input and output variables.

2.3 LANGUAGE

PA3 is written in the BASIC language. In spirit, PA3 resembles the GASP language in that it uses a well-known, general-purpose language and supplements it with a number of special functions. The user is required to know only enough BASIC to adequately represent his problem as a series of statements in that language. BASIC was chosen as the user's language because it was felt that most potential users would be acquainted with either BASIC or FORTRAN, and that the FORTRAN users could easily learn enough BASIC to state their problems. The use of BASIC has required some restrictions on the local variable names available to the user because of the global requirements for names, particularly array names, but this does not seem to have caused undue hardship on any user to date. The system is currently implemented only on the GE Mark II time-sharing system, but could likely be implemented on any time-sharing system having the BASIC language with string variable, chaining, and permanent file capability. Because of the nature of the language and its intended use, its implementation in a batch computation mode is not contemplated.

2.4 LEVEL OF MODEL DEFINITION

The model is entirely user supplied, with no problem structure provided by the language. Some functions are available to the user, and more are contemplated, but the user must essentially build his model from the statement types available in BASIC. The model is compiled at execution time, and unanticipated changes in logic require the changing of the kernel and subsequent recompilation. If one has saved different versions of the model logic, any of these can be exercised using the same data base by requesting a logic changing option. In the interactive environment characteristic of time-sharing, however, the practical difference between this and an interpretive mode of operation becomes much less noticeable, particularly since any change in kernel is automatically accompanied by a debug check and printout. Only when one reaches the extremely interpretive level of the OPS-3 system is the difference between interpretation and compilation felt.

2.5 LEVEL OF DATA DEFINITION

At the outset, the data are saved in a file in much the same manner as the kernel. During execution, a change to another previously saved compatible data file can be made by exercising a load option. Alternatively, one or more pieces of data in the current data file can be temporarily or permanently changed by the use of a change option.

2.6 PROGRAMMING VERSUS EXECUTION TIME

As should be clear from the above discussion, the emphasis in PA3 is entirely on reducing programming time at the expense of execution time. Given the extra inefficiencies of time-shared computation, particularly in the case where a series of chained programs is passed across a relatively constant model and data, the unit cost of useful computation would be expected to be higher. In most problems, especially those where repeated execution is not necessary, the total computation cost is so small compared to the total manpower used that the slight increment is more than compensated for by the increase in flexibility and decrease in turn-around time. This is, of course, an argument for time-shared computation in general, but it also applies to a special purpose language like PA3 where much of the bookkeeping common to many problems is performed by the language.

2.7 SAMPLING PROCEDURES

Where Monte Carlo methods are called for, PA3 permits a variety of distribution shapes. The central distribution is the normal, with special transformations used to capture skewness in either direction. Also currently available are variants of the Weibull distribution bounded on either the left or right, and a uniform distribution with general range. By using special functions in the kernel, truncated, folded, and other variations on the basic distributions can be constructed. Although no discrete distributions are built in, they can easily be constructed in the kernel using the uniform distribution as the basic sampling distribution.

The characteristics of each probability distribution are inputted using four numbers. The first of these numbers specifies the type of distribution, and the last three specify three points on the cumulative distribution, usually the tenth, fiftieth, and ninetieth percentiles.

2.8 OUTPUT OPTIONS

All outputs from each analysis are titled by date, time, and other relevant information for ease of identification later. In addition, the user is given a choice of items to be examined, depending on the type of output.

(1) Data verification: the user can examine any subset of the input data.

(2) Deterministic analysis: the user can look at a subset of the output variables computed at the deterministic values of the input variables.

(3) Sensitivity analysis: for any single input variable, the user can examine the effect of its changes on a subset of the output variables.

(4) Probabilistic analysis: the sample mean, sample standard deviation, and approximate tenth and ninetieth fractile are accumulated during the Monte Carlo trials for all output variables; a subset of these can be printed. In addition, every sample point on up to five of the output variables can be saved for the more detailed output options. These options include the printing of the sample high, sample low, sample tenth, fiftieth, and ninetieth percentiles, as well as printing of a detailed histogram to any interval scale and midpoint desired. On the first pass, PA3 automatically chooses the scale and midpoint of the histogram based on the sample values observed.

The next section of the paper demonstrates these options in the context of a simple example.

3. EXAMPLE OF SYSTEM OPERATION

The following highly simplified model for a "New Product" introduction demonstrates the use of the major options currently available. While the system can accept up to 400 input variables and calculate up to 200 result variables, for this illustration we have only 4 inputs

X(1) =	Selling Price	$	=	900
X(2) =	Unit Cost	$	=	500
X(3) =	Sales Volume	Units	=	5000
X(4) =	Overhead	$1000	=	1100

and 3 results

Y(1) =	Margin	$/unit
Y(2) =	Profit	$1000
Y(3) =	Net ATax	$1000

which are related by the following logic:

Margin = Price - Unit Cost (1)

Profit = Margin * Sales - Overhead (2)

Net = Profit * .52 (3)

The following reproductions show the simplicity and speed of entering and analyzing this problem statement. To enter any problem, the line number assignment rule is:

1000 -	Logic Statement
5000	# Inputs, # Results
5100 -	Deterministic Values
6100 -	Probabilistic Values
7000	Problem Name, Problem #
7100 -	Input Names
8100 -	Result Names

```
                20:19       09/22/69

READY

NEW
NEW FILE NAME-- L001
READY
1000 LET Y(1)=X(1)-X(2)
1010 LET Y(2)=Y(1)*X(3)/1000-X(4)
1020 LET Y(3)=Y(2)*.52
SAVE
READY
```

```
NEW
NEW FILE NAME-- D001
READY

5000 4, 3
5100 900, 500, 5000, 1100
6100 1, 820, 900, 940
6110 2, 450, 500, 540
6120 3, 4000, 5000, 5800
6130 4, 1000, 1100, 1160
7000 NEW PRODUCT PLAN, P001
7100 SELLING PRICE $, UNIT COST $
7110 SALES VOL UNITS, OVERHEAD $
8100 MARGIN $/UNIT, PROFIT $1000
8110 NET A TAX $1000
SAVE

READY
```

At this point, the system is ready to operate, and Figures 1 through 5 show how the following options were exercised:

DEBUG checks format and internal consistency of data base and loads system. Note provision for listing all data for a detailed check of values and names.

LOGIC checks format and internal consistency of logic statement and loads system. Note automatic sequencing for calculation of deterministic results.

SENS performs sensitivity analysis for any (or all) results as any single input variable is stepped through any range using any step size.

PROB performs probabilistic analysis using 1 to 500 Monte Carlo trials. Provides up to 5 result histograms, if specified.

CHANGE provides opportunity to change any deterministic or probabilistic entries.

In these figures, the conversational replies of the user are underlined.

In spite of the simple nature of this example and its unrealistic assumptions, enough of a flavor for the options available in the language has been given to indicate its capabilities in attacking more complex problems. In the next section, some of the problems which have been attacked are described.

4. DESCRIPTION OF ACTUAL PROBLEMS

The four examples here are chosen from the four primary functional areas of a business: engineering, finance, manufacturing, and marketing. Each of them was suggested by the person with the problem, in all cases a person with little or no modeling and simulation experience. While they may seem simple to the experienced analyst, they represent a big step forward to people not used to thinking in these terms. The typical involvement of the authors was a few hours to acquaint the user with PA3 and help him structure his initial model, followed by occasional phone conversations to help him iron out minor problems. In a week or two of elapsed time, the user typically had obtained and documented his initial results, and was on his way to embellishing the model for later analyses. In some cases, significant savings were shown; in others, no savings could be substantiated because the problem had never been viewed in these terms before.

Many of the problems attacked to date are solvable exactly by special analytical techniques (e.g., the process problem is a simple Markov process), but these are neither transparent nor within the ken of the person with the problem. If he were required either to learn the special technique or to take his problem to someone acquainted with the technique, the problem would never get solved. An approximate technique which allows the user to get a feeling for his problem in a short time is regarded as being preferable to that alternative.

4.1 ENGINEERING TOLERANCES ANALYSIS

The production of electron gun assemblies for use in color television tubes requires very close mechanical tolerances on the critical grid-cathode diameters, thicknesses, and spacings. If these geometric requirements are not held closely, the operating cut-off voltages will not match the electrical characteristics of the driving circuit. On the other hand, production losses increase sharply with these high precision requirements, and so engineering conservatism could result in excessive manufacturing costs. Using this PA3 system, it was a relatively simple matter to make a probabilistic evaluation of a reasonably complex algebraic formula, and thus the expected percentage of failures could be found as a function of the design tolerances.

4.2 FINANCIAL MODEL OF A BUSINESS

One product department is representing its total business by a time-shared model in an attempt to choose product direction and forecast various financial measures. As more has been learned about the business, the detailed model has become extremely enriched; at this writing, it relates 392 input variables to 178 output variables. In spite of its size and complexity, the analysts involved continue to use the time-shared model because they feel the flexibility for minor changes and the short turn-around time justify the slight premium they must pay for not going to a batch operation.

4.3 PROCESS CONTROL

The production of silicone products involves many processes which are performed in sequence. Typically, both the capacity and the quality of the product produced by any stage is dependent on the qualities of the input streams. Further, the output from any stage typically must be directed (sorted) to several different processes, many of which may be salvage and reworking operations.

Using a model of this process, the manufacturing engineer was able to simulate its transient and steady state operation under various control options. From this, he obtained the optimal values of the control variables, values which produced a significant financial improvement in the process operation. The model is now being used for further exploration in equipment selection and other related areas.

4.4 SALES AND PROFIT FORECASTING

One of our businesses involves the installation of large systems (power plants, industrial equipment, etc.). To forecast the over-all business situation, it is necessary to combine the industrial sales estimates of each of the many different product lines and, also, to estimate the profit contribution of each of these lines using both historical knowledge and any special information about each situation. Recognizing the inherent uncertainties in these situations, the department management requested that a program be written to process the probabilistic distributions assigned to each of the product line's sales and profitability estimates. This program was conceived and perfected in a few hours. A later modification was prepared in another half-day session that permitted a complete specification of the conditional probability relations between the "% gross income" and the "sales volume" in each of the product lines.

5. CONCLUSIONS

Transient documentation has been given on a continuing effort to write a "....special purpose, problem-oriented simulation language that reduce(s) the distance between a person with a problem and an executable computer program," an objective stated by Kiviat.[5] This language has by no means reached steady state, with options being added by its author in response to its users' needs, much in the spirit of future simulation developments suggested by Krasnow and Merikallio.[6] More special facilities will continue to be added, but the ability for the beginner with little modeling experience to get answers for his simple problem quickly will be retained.

BIBLIOGRAPHY

1. D.G.Ebeling, "A General Purpose, Time-Sharing Computer Program for Probabilistic Analysis" (October 1967).
2. D.G.Ebeling, "Probability Concepts, Monte Carlo Technique and Explanation of BASIC Program PA1" (February 1968).
3. D.G.Ebeling, "A General Purpose, Conversational Time-Sharing Program for Probabilistic Analysis (PA3 - 1969 version of PA1)" (August 1969).
4. M.Greenberger, M.M.Jones, J.H.Morris, and D.N.Ness, On-Line Computation and Simulation: The OPS-3 System, The MIT Press: Cambridge, Mass. (1965).
5. P.J.Kiviat, "Development of New Digital Simulation Languages" The Journal of Industrial Engineering, 17, 11 (November 1966).
6. H.S.Krasnow and R.A.Merikallio, "The Past, Present, and Future of General Simulation Languages" Management Science, 11, 2 (November 1964).
7. M.R.Lachner, "A General Simulation Capability" Data Processing (November 1962).
8. R.C.Meier, W.T.Newell, and H.L.Pazer, Simulation in Business and Economics, Prentice-Hall: Englewood Cliffs (1969).
9. T.H.Naylor, J.L.Balintfy, D.S.Burdick, and K.Chu, Computer Simulation Techniques, John Wiley & Sons: New York (1966).
10. D.Teichrow and J.F.Lubin, "Computer Simulation - Discussion of the Technique and Comparison of Languages" Communications of the ACM, 9, 10 (October 1966).
11. K.D.Tocher, "Review of Simulation Languages" Operational Research Quarterly, 16, 2 (June 1965).

BIOGRAPHIES

DR. D. G. EBELING joined the metallurgical department of the Homestead plant of U.S. Steel upon graduation from Rensselaer Polytechnic Institute in 1940. As ordnance officer at the U.S. Naval Proving Grounds, he was engaged in ballistic testing and development. Joining the General Electric Company after the war, he has had assignments in the Chemical Department, Research Laboratory, Turbine and Generator Department, Knolls Atomic Power Laboratory. In his present position as consultant on the Corporate Engineering staff, he has been active in the development of their Modern Engineering Course for engineering managers and, more recently, in the development of modern decision techniques for technical and business analyses.

DR. E. G. HURST, JR., is Assistant Professor of Industry, Wharton School, University of Pennsylvania. Formerly associated with the Power Transmission Division of the General Electric Company in Philadelphia applying various management science techniques to Division planning and operating problems, he has been active in a number of industrial educational programs in the field of management science and has published in the areas of inventory theory, production scheduling, and adaptive control. He holds an S.B. degree in civil engineering, S.M. in operations research, and Ph.D. in management from Massachusetts Institute of Technology. Dr. Hurst is a member of Tau Beta Pi, Sigma Xi, Institute of Management Sciences, Operations Research Society of America, and American Statistical Association.

FIGURE 1: CHECKING DATA BASE, PROBLEM LOGIC, AND DETERMINISTIC RESULTS

```
OLD
OLD FILE NAME--PA3--$
READY
RUN

PA3--$        20:31       09/22/69


NEXT OPTION? DEBUG
WHICH DATA FILE? D001


DATA IS READABLE --  WE HAVE STORED:
     4 INPUT NAMES AND DETERMINISTIC VALUES
     3 RESULT NAMES
     4 PROBABILISTIC DISTRIBUTIONS

WHICH ASSUMPTIONS? 1-4


09/22/69                                                     L089
 20:32                                                       D001

               NEW PRODUCT PLAN  ASSUMPTIONS

INPUT VARIABLE            DET VALUE TYPE LOW       MID       HIGH

 1    SELLING PRICE $      900       0    820       900       940
 2    UNIT COST $          500       0    450       500       540
 3    SALES VOL UNITS      5000      0    4000      5000      5800
 4    OVERHEAD  $          1100      0    1000      1100      1160

NEXT OPTION? LOGIC
WHICH LOGIC FILE? L001
LOGIC IS COMPUTABLE

WHICH RESULTS? 1-3


09/22/69                                    L001
 20:34                                      D001

              NEW PRODUCT PLAN
          DETERMINISTIC PREDICTIONS

      I    VARIABLE               VALUE

      1    MARGIN  $/UNIT          400
      2    PROFIT  $1000           900
      3    NET A TAX $1000         468


DETERMINISTIC ANALYSIS COMPLETED

YOU CAN RESTART BY CALLING OLD PROGRAM: PA3--$


USED    4.09 UNITS
```

FIGURE 2: RUNNING SENSITIVITY ANALYSIS OPTION

```
OLD
OLD FILE NAME--PA3--$
READY
RUN

PA3--$        20:35       09/22/69


NEXT OPTION? SENS
WHICH OUTPUT VARIABLES? 2,3
WHICH INPUT VARIABLE?, LOW VALUE?, HIGH VALUE?, STEP SIZE?
? 3,4000,6500,500


09/22/69                                                        L001
 20:36                                                          D001

      SENSITIVITY ANALYSIS  NEW PRODUCT PLAN


ASSUMPTION            RESULT                RESULT
SALES VOL UNITS       PROFIT   $1000        NET A TAX $1000

 4000                  500                   260
 4500                  700                   364
 5000                  900                   468
 5500                 1100                   572
 6000                 1300                   676
 6500                 1500                   780


SENSITIVITY ANALYSIS COMPLETED


USED    2.36 UNITS
```

FIGURE 3: RUNNING PROBABILISTIC ANALYSIS OPTION

```
OLD
OLD FILE NAME--PA3--$
READY
RUN

PA3--$          20:37       09/22/69


NEXT OPTION? PROB
FOR WHICH RESULTS DO YOU WANT HISTOGRAMS? 2,3
N MC TRIALS? 100


09/22/69                                                      L001
 20:38                                                        D001

                          NEW PRODUCT PLAN
               PROBABILISTIC PREDICTIONS ASSUMING NORMALITY

RESULT                     Y10%         MEAN        Y90%        STD DEV
WHICH RESULTS? 1-3

 100 TRIALS:

 1   MARGIN  $/UNIT        318          393.5       469          59
 2   PROFIT  $1000         362          885.3       1410         409
 3   NET A TAX $1000       188          460.4       733          213

09/22/69                                                      L001
 20:40                                                        D001

                       NEW PRODUCT PLAN
        Y 2     PROFIT  $1000                   SCALE  1000  100

                                *
                                * *
                                * *
                            * * * *
                          * * * * * *
                        * * * * * * *
                        * * * * * * *
                      * * * * * * * * *
                      * * * * * * * * *
                * * * * * * * * * * * * * *
              * * * * * * * * * * * * * * *
  *       * * * * * * * * * * * * * * * * * * * * *
  ↑         ↑         ↑         ↑         ↑         ↑         ↑
-500        0        500      1000      1500      2000      2500
LOWEST=-470.68                                HIGHEST= 1895
```

FIGURE 4: PROBABILISTIC ANALYSIS OPTION continued

```
09/22/69                                                           L001
 20:42                                                             D001

                      NEW PRODUCT PLAN
      Y 3     NET A TAX $1000                   SCALE   500   100

                                  *
                                  *
                                *  *
                                *  *  *
                                *  *  *
                                *  *  *
                                *  *  *
                                *  *  *
                              *  *  *  *
                              *  *  *  *  *
                              *  *  *  *  *
                            *  *  *  *  *  *
                            *  *  *  *  *  *
                            *  *  *  *  *  *
                          *  *  *  *  *  *  *  *
                          *  *  *  *  *  *  *  *
                        *  *  *  *  *  *  *  *  *  *
                        *  *  *  *  *  *  *  *  *  *
                    *  *  *  *  *  *  *  *  *  *  *  *  *
    ↑         ↑         ↑         ↑         ↑         ↑         ↑
  -1000     -500        0        500      1000      1500      2000
LOWEST=-244.76                                        HIGHEST= 985.4


HISTOGRAM ANALYSIS COMPLETED

YOU CAN RESTART BY CALLING OLD PROGRAM: PA3--$

USED     9.32 UNITS
```

FIGURE 5: CHANGING DATA BASE AND RUNNING
NEW DETERMINISTIC ANALYSIS

```
OLD


OLD FILE NAME--PA3--$
READY
RUN

PA3--$          20:47        09/22/69


NEXT OPTION? CHANGE
ENTER 1,OR MORE, NEW DETERMINISTIC VALUES: VAR? VALUE? VAR? VALUE? ...

? 1,800,3,6000
 1    SELLING PRICE $ 800
 3    SALES VOL UNITS 6000

ENTER NEW PROB ASSUMPTIONS:  VAR? TYPE? LOW? MID? HIGH?
? 0

NEXT OPTION? DET

WHICH RESULTS? 1-3


09/22/69                                      L001
 20:48                                        D001

             NEW PRODUCT PLAN
          DETERMINISTIC PREDICTIONS

      I    VARIABLE                VALUE

      1    MARGIN   $/UNIT         300
      2    PROFIT   $1000          700
      3    NET A TAX $1000         364


DETERMINISTIC ANALYSIS COMPLETED

YOU CAN RESTART BY CALLING OLD PROGRAM: PA3--$

USED    2.08 UNITS
```