A NEURAL NETWORK ANALYSIS OF OPTIMAL CATPAC PARAMETERS

by

Marcya Newman Foster

A Masters Project Submitted to the
Interdisciplinary Degree Program in the Social Sciences in
Partial Fulfillment of the Requirements for the Degree
of
Master of Science

1996

Major Professor
Joseph Woelfel, Ph.D

Committee Member
Allan Canfield, Ph.D

Director of Graduate Studies
Norman Baker, Ph.D

A Neural Network Analysis of Optimal CATPAC Parameters

(abstract)

The data analysis community has always been torn between quantitative and qualitative analysis. There is an inordinate amount of data that remains unanalyzed. The tool to analyze text must be easy to grasp, simple to use, allow for varying degrees of involvement, require minimal preparation of raw data, and preferably have comparison capabilities through simultaneous testing and process measurements. Ideally, it must also have a broad application for varying organizational types and levels in different fields and possibly different languages. CATPAC, a self-organizing neural network, may offer a promising approach to this task. It provides multivariate analysis of several types. In experiments, two important quantitative results emerge: subjective judgements of the simulations and visualizations of those simulations. NEUROSHELL was employed to test and verify its prediction capabilities.

I would like to thank Joe Woelfel, Becky Omdahl, and Drew Campbell for their encouragement and inspiration, and Allan Canfield for his critical reading of this paper.

TABLE OF CONTENTS

I. INTRODUCTION

CATPAC Historical Aspect

The data analysis community has always been torn between quantitative and qualitative analysis. There are advantages and disadvantages to both. Quantitative procedures provide a precise determination, quick and easy, but require pre-existing coding schemes to show all possible outcomes. This requires detailed advance knowledge of the problem, and may rule out synergistic results. The richness of data is also compromised using this method.

Qualitative advantages include unstructured responses, which provide richness of data. Nuances in the data are easily communicated, and there are no bounds for data formats. The text may be produced by words, pictures, and sounds. But good qualitative analysis is extremely difficult and rarely achieved.

One aspect of qualitative analysis and probably the largest, is text analysis. Text may be the output of focus groups, questionnaires, or interviews, among other sources. Lexis-Nexis provides an inordinate amount of data. However, most qualitative data does not get analyzed. Large amounts of both verbal and textual data exist that are poorly analyzed or totally un-analyzed.

Over the past several years, devices were invented to analyze text. The General Inquirer is an early example of text analysis. It is cumbersome. Text analysis software is, for the most part, template matching with simple parsing techniques. Template matching is looking for similarity in words. Parsing is breaking up sentences into words and structures of two or more words, as clauses, phrases, etc. But these procedures are not satisfactory because they have to be precoded. The same problem exists as with other cases of quantitative analysis.

A technological development that offers a promising approach, however, is the Artificial Neural Network (ANN). "ANNs are mathematical models in two senses. First, they represent idealized models of biological neural systems, and are often used as tools to help better understand the functioning of biological systems. But more frequently, ANNs are used to model other processes, and it is these other kinds of uses that provide the basis for ANNs as an applied technology." (Woelfel,1993).

There are several types of ANNs modeled on neurological switch-like matrices (binary and analog neurons) interconnected by pathways that take on values.[1] The network learns and forgets patterns. It relies solely on changing weights. The neurons are multiplied by weights (transfer function). These connections or weights may be classified by the rules which govern the weights changes. Type 1: Hebb Rule - unsupervised or self-organizing network;[2] Type 2: Weights change in order to produce a predetermined output - supervised; and Type 3: Weights are directly measured by some process. Type 1 provides a promising analysis by enabling the user to uncover the underlying concepts of the data. It has the capability to recognize the re-occurrence of words and phrases in input data without the need of a pre-existing coding scheme. CATPAC is an example of the self-organizing neural network

[1]Matrices in computers are logic networks in the form of arrays of input and output leads with logic elements connected at some of the intersections.

[2]The Hebb Rule states that the connection between nodes, that are simultaneously activated, is strengthened. A network of this type can learn an internal representation of its external environment.

that has been optimized for reading text and doing content analysis.

CATPAC is a crude parser. CATPAC breaks text up into words. It looks for and counts words. More advanced parsers can recognize phrases, clauses, nouns, verbs, subjects, etc.

CATPAC is able to identify the most important words in a text and determine their patterns of similarity based on their associations in the text. From this information, CATPAC is able to identify the main underlying concepts dealt with in the text. Its neural technology provides pattern matching capabilities a human might have. CATPAC creates its own matrix of neurons depending on the work required of it. "It is this process of communicating activation levels throughout a network of nodes that gives a neural network its information-processing capabilities--which include the ability to represent, store, retrieve, and associate patterns of arbitrary complexity, and to generalize information learned about a given pattern to other related patterns." (Woelfel, 1993).

CATPAC is designed to read and analyze text. It works much like the human mind. It provides enhanced qualitative analyses and performs hierarchical cluster analysis. CATPAC is multilingual (e.g., CATPAC handles English, French, Spanish, Italian, German and Dutch.)

CATPAC performs routine textual analysis. It lists words alphabetically and by frequency. It deals with the elimination of unnecessary articles, prepositions, and other superfluous words using an exclude file. It examines connections among remaining significant words throughout the text. CATPAC uncovers underlying patterns or concepts as clusters.

If programs like CATPAC and their more sophisticated successors should continue to prove effective, they solve the problem of providing objective quantitative data from qualitative data and, in doing so, provide a means to make a quantitative estimate. "Neural networks' capacity to read vast quantities of unprepared text and provide a brief and useful synopsis of their main concepts provides policy researchers with a useful tool not available with conventional technology." (Woelfel, 1993).

CATPAC Theory

CATPAC reads and analyzes textual material by reading any ASCII text file.[3] CATPAC assigns a neuron to each of the major words chosen in the text. A moving window slides through the text. When a word representing a certain neuron is in the window, that neuron is activated. When the neurons are simultaneously activated, they are strengthened by adding the Hebb Constant of .05, which is the default strengthening factor.[4] When a given input pattern of active neurons represents a pattern, repeated patterns will activate the same neurons, strengthening them. The network becomes a unit and it begins to recognize and learn patterns to remember.[5] Several variables may be manipulated by the user for this task, such as unique words requested,

---

[3]ASCII stands for the American Standard Code for Information Interchange. Characters are composed of 8-bit codes (ex.0000 0110.)

[4]Khanna (1990) stated that the Hebbian Rule is a learning strategy that suggests that when a cell A repeatedly and persistently participates in firing cell B, then A's efficiency in firing B is increased.

[5] "On the most fundamental level, neural networks perform pattern recognition and do so more effectively than any other technique known. Once the network has detected a pattern, the information can be used to classify, predict, and analyze the causes of the pattern. Neural networks have been successfully applied to signal processing, modeling and forecasting." (Coleman, 1992)

scanning window size, window slide size, clamping of nodes (1/0), and cycling. CATPAC counts, sorts, and lists words by frequency and alphabetically.

Unique words requested indicate how many words are considered necessary to evaluate. Three hundred (300) is the maximum possible with the system. CATPAC uses an exclude file to eliminate common words, articles, prepositions, and other repetitious words which would not add value to analysis. The exclude file may be modified by the user to add or delete entries. The unique words chosen by the user may not be equal to actual words found by the system. CATPAC begins with the words with the highest frequency and lists them. It then lists the words with the next highest frequency, and so on. If a frequency has more words than the total of requested unique words, it will not list that frequency, but will cut off at the end of the previous frequency. Therefore, CATPAC will never exceed the requested unique words but may fall short of them.

Scanning window size may be varied from one word upward. In its current configuration, scanning window size may grow to a maximum of twenty words. The scanning window indicates how many unique words are in each sequential scan.

Scanning window slide indicates how many words the scanning window will move each time. If you use a slide of one (1) and the scanning window size is 5, then the first five words will be in the window, then words 2-6, 3-7, and so on. If, however you choose a size of three (3) and a slide of five (5), certain words will never be in the window. For example, initially words 1-3 will be in the window, then words 6-8, and 11-13, and so on. Words four (4) and five (5), words nine (9) and ten (10), and every fourth and fifth word of each remaining slide will not be in the window and will remain unanalyzed.

Clamping of nodes (neurons) indicates whether to retain the strengthening factor as "on" (active) to produce tighter connections between nodes. Clamping of nodes prevents neurons that have been turned on, or activated, from being turned off, or deactivated. Lines between simultaneously stimulated nodes are tightened (Back Propagation Rule.)[6]

---

[6]With the Back Propagation Rule, errors can be expressed as a value of output nodes and output nodes as a function of weights.

Cycling

In a biological neural network, activation update processes go on in a continuous fashion. Signals are sent and received through a process called hysteresis (delay). They travel through the network at different rates of speed. Neurons become active and inactive depending upon this process.

Because of the nature of the computer, an update must be planned and executed at intervals to carry out the latest activations. In CATPAC, this is called cycling. An update of all the weights takes place with each cycle. CATPAC may cycle two or three times and uncover second and third order relationships among words that otherwise may not have been considered. No cycling may only produce the more obvious associations between words and phrases. Excessive cycling, on the other hand, could homogenize the data and make it less meaningful.

The cycling variable allows a maximum of four iterations through the data. Each iteration is one cycle. Zero (0) cycles is an accepted choice.

Each connection is strengthened a maximum of 5% each cycle. Increasing this rate makes CATPAC learn faster. This number is applied to connections between nodes which are simultaneously activated. These connections are weights. The pattern of weights or connections among neurons forms a representation within CATPAC of the associations among the words in the text. This pattern of weights represents complete information about the similarities among all the words in the text.

The data are normalized to reduce the strengthening values to one (1) or less. Re-normalization occurs with each cycle so the strengthening factors grow smaller, always normalizing to one (1) or less. Some numbers may be negative. Zero (0) is the default threshold of activation of each node.

All of the communication activations (weights) of all connecting nodes are a result of the transfer function and the activation function.

Transfer Function

Neurons are turned on by input in two distinct ways in CATPAC: (1) by appearing in the moving window, and (2) by being connected to other neurons that have been turned on. These inputs are transformed by a transfer function. "CATPAC can use one of four transfer functions: a linear function varying between -1 and +1, a logistic function ranging between 0 and +1, a logistic function varying between -1 and +1, and a hyperbolic tangent function varying between -1 and +1." (Woelfel, 1991).

Activation Function

Inputs are summed after being transformed by the transfer function above. The neuron is activated if it goes over a certain threshold. If it does not exceed the threshold, the neuron remains inactive. "The weight represents the proportion of the activation value of the node that will be communicated to the connecting nodes. The default threshold is 0.0 which is appropriate for three of the four transfer functions. (.5 would be a more reasonable value for the logistic varying between 0 and +1.) By lowering the threshold, the more likely for neurons to become activated: by raising the threshold, the less likely for neurons to become activated." (Woelfel, 1991).

A decay rate which counteracts the activation of neurons dictates how quickly a neuron returns to its rest condition of 0.0. Each cycle is governed by a decay rate of the activation levels of neurons. If the default decay rate of .9 was used, each neuron would lose 90% of its activation with each cycle if it was not reactivated. This rate may be adjusted to increase or decrease the rate of speed of the activation levels.

A learning rate is also a variable of CATPAC. The default learning rate is .05. Each connection can only be strengthened a maximum of 5% for a given cycle. While increasing this rate may result in faster learning, the results may be difficult to interpret as new information is introduced.

CATPAC allows for the systematic manipulation of many independent variables such as unique words requested, scanning window size, window slide size, clamping of nodes (1/0), and cycling. It examines the connections between nodes simultaneously activated to uncover the underlying pattern or concepts in the data. Technically, this is a Hopfield Analog binary pattern of connections among neurons and is a complete paired comparison similarities matrix.[7] As such, it lends itself to the most powerful and sophisticated of statistical analyses. (Woelfel, 1991).

Cluster Analysis

A diameter method cluster analysis is automatically performed by CATPAC. The cluster analysis may be in the form of network analysis or a simpler and earlier version based on a concurrence matrix. After CATPAC processes, it displays through use of Johnson's hierarchical matrix dendogram the words which are closely associated with one another; more simply, the frequency they co-occur in the scanning window. The dendogram graphically depicts the frequency words occur in relation to each other. Clusters of co-occurrence appear as peaks on the dendogram.[8] At the lowest level of analysis. All words co-occur.

---

[7] The Hopfield networks store conceptual structures as patterns solely as a function of weights or relationships among nodes. All values are within +/- 1.

[8] Co-occurrence does not insist that the words are next to one another in the input data.

II. THE PROBLEM

CATPAC's performance depends on a group of interrelated, non-linear parameters whose default settings have been established by cut-and-try methods. CATPAC's parameter settings have not been exposed to a systematic evaluation (such as the controlled application of variables on specific texts.) Therefore, qualities of operating characteristics are virtually unknown.

III. METHODS

This research attempted to expose CATPAC's default parameters to several empirical tests to establishsome grounds in observation. To identify optimal CATPAC techniques, five (5) texts were chosen to systematically evaluate CATPAC. They were of various contents and lengths. Each text was subjected to a systematic battery of experiments, 102 in all, arbitrarily varying the following independent variables by the numbers indicated: Unique words of 30, 50, and 100, scanning window size from 1 - 9, window slide size from 1 - 3, clamping of nodes (on/off), and cycles from 1 - 4. Defaults were taken on threshold, restoring force, decay factor, and momentum factor for all the experiments.

An output correlation matrix of the independent variables was produced manually. A software product, Experimental Design Optimizer (EDO) was applied to the experiment criteria requesting squared effects on the range of independent variables. EDO produces a list of all admissible experiments in addition to a list of the smallest number of experiments which would tell the most about the data. EDO provides a prediction error ratio for each experiment. An average error of prediction closest to one (1) is best. Experiments may be added to or subtracted from the list for optimal results. An EDO Generated Design gives the average, maximum, and minimum error of prediction.

Output was initially evaluated in two ways: (1) CATPAC dendograms were analyzed at four (4) levels from the top of the dendogram. Levels 5, 10, 15, and 20 were arbitrarily chosen. Numbers of clusters and total words at these arbitrary levels were counted and recorded. (2) Each experiment was subjectively rated by two humans. These ratings were recorded for each of the experiments. The weights input network (.WIN) file, an output file of CATPAC, was used as input to ORESME, a software product in Galileo for comparison of resulting concepts with CATPAC dendograms. ORESME illustrates a more human-like understanding when applied to specific jobs (e.g., campaigns.) CATPAC is hierarchical and can only place a word in one cluster, whereas, ORESME may use a word in multiple concepts.

Multidimensional scaling and Perceptual Mapping (Galileo) are both options of analysis through the output of CATPAC. But for this experiment, five (5) training sets, or experiments, were randomly chosen from each of the five texts to be entered into NEUROSHELL, a neural network software package. The following independent variables were input into records: Unique words, words found, window size, window slide, clamping (1/0), cycles run, and text used. Dependent variables were input as: number of clusters at each of four (4) levels of analysis, total number of words counted at each of the four (4) levels of analysis, subjective rating by researcher "D" and subjective rating by researcher "M".

NEUROSHELL was then asked to learn the training sets. The program was allowed to run approximately 14 - 15 hours. All remaining experiments were entered into NEUROSHELL to test its prediction capabilities. These experiments are the test sets. NEUROSHELL was tested against the test sets in the areas of subjective rating, (other variables, etc.).

IV. RESULTS

The results of CATPAC provide multivariate analysis of several types. Cluster analysis revealed a tendency for clusters and total words to decrease as the scanning window size increased. This held true when the cycles were increased to 2 and 3. When the cycles reached 4, the pattern was random, the decreasing tendency of clusters and total words disappeared. Clamping of nodes (off) produced a stability of clusters and total words at a scanning window of 5 or less. This stability decreased precipitously when the window reached a size of 8 and 9. Clamping of nodes (on) tended to reduce the clusters in experiments of multiple cycles. These settings corresponded to the best values in a subjective rating. The ratings converted at r^2 values of $\backsimeq .8$.

Overall, experiments performed provided two important quantitative results: (1) subjective judgements of the simulations and (2) visualizations of those simulations.

NEUROSHELL was able to predict the test cases with a high degree of accuracy.

This research on CATPAC put forth only initial findings and is by no means conclusive. It provides a method to quantitatively measure results in text analysis. Further research is needed. There needs to be extensive testing of the different activation functions, as only default values were used in these experiments. Comparisons of the hierarchical dendogram of CATPAC with ORESME to test the percentage of difference of emerging concepts has only been touched upon. Most importantly, there is a need to provide systematic and quantitative comparisons of CATPAC.

V. SUMMARY AND CONCLUSION

In this study, we are trying to predict how good CATPAC predicts the outcome. On the subjective side, "D" and "M" rate how well CATPAC will predict the outcome, while NEUROSHELL predicts which of the parameters would be optimal, which writing is best and which parameter setting is best.

This thesis tested a new procedure. Artificial Neural Networks (ANNs) as models of other processes offer a promising approach for the analysis of unstructured raw data. CATPAC is a self-organizing neural network optimized for reading text and doing content analysis. CATPAC provides an objective quantitative analysis from vast quantities of unprepared qualitative data and allows for a quantitative estimate of its contents.

In this paper, CATPAC was analyzed by individuals and NEUROSHELL. This was done to evaluate its parameter settings in a systematic evaluation. These data were also run through EDO to test the error of prediction ratio of each experiment. This analysis used 5 different kinds of writing. The subjective rating by two individuals was inputted into NEUROSHELL. NEUROSHELL was then tested against the subjective ratings of the two individuals and against the test sets it learned. NEUROSHELL was able to predict the test cases with a high degree of accuracy.

Results of this study were that using a scanning window size of five (5) or less was optimal when the clamping of nodes parameter was set at "off." Clamping of nodes "on" tended to reduce the clusters in experiments of multiple cycles. Two or three cycles could be run with stable results, but when four (4) cycles were run, the patterns tended toward randomness.

While this study tested several different kinds of writings and was able to identify optimal settings for these specific writings, the subject is open to further analysis. Additional types of documentation must be analyzed through CATPAC and analyzed using NEUROSHELL to test whether the results hold true across the spectrum of other unique compilations of data. Further research is needed to compare these findings against these unique data.

BIBLIOGRAPHY

Barnett, G.A. & Woelfel, J. (1988). Readings in the Galileo System. Dubuque, IA: Kendall/Hunt.

Byrnes, Daniel J. "More on the approaching neural net." PI Quality. January/February 1994. pg. 14

Coleman, Kevin G. & Susan Watenpool, "Neural Networks in Knowledge Acquisition." AI Expert. January 1992.

Cox, Earl. "The Great Myths of Fuzzy Logic." AI Expert. January 1992. pp.

Fink, Edward L. (1990) "Mathematical Models for Communication: An Introduction." Journal of Communication. University of Maryland. Vol 43(1) Winter.

Hillman, David. "Knowledge-Based Systems on Cascading Neural Nets." AI Expert. December 1991. pp.

Khanna, Tarun. (1990). Foundations of Neural Networks. New York: Addison-Wesley.

Klein, H.A. (1974). The Science of Measurement. Toronto: General Publishing Co.

Rodriquez, Suzanne M. "Neural Networks Demystified." Systems. February 1994. pp.62-68.

Saltiel, J. & Woelfel, J. (1975). "Accumulated Information as a Basis for Attitude Stability." Human Communication Research.

Sekaran, U. 1984. Research Methods for Management. New York: Wiley.

Woelfel, Joseph, "Artificial Neural Networks in Policy Research: A Current Assessment." Journal of Communication. 43(1).

Woelfel, Joseph. (1990). Communication and Science. Buffalo: Department of Communication, SUNYAB.

Woelfel, J. & Danes, J. (1949). The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois.

Woelfel J., & Fink, E.L. (1990) Communication and Science. Buffalo: Department of Communication, SUNYAB.

Woelfel, J. & Fink, E.L. (1980). The Measurement of Communication Processes. New York: Acedemic Press.

Woelfel, J. & Saltiel, J. (1975). "Cognitive processes as motions in a multidimensional space." In F. Casmir (Ed.) International and Intercultural Communication. New York: University Press. (pp. 105-130)

Woelfel, J., Stoyanoff, N., & Danielsen, S. (1992). Catpac User Manual. Troy, NY: Terra Research and Computing Co.

APPENDICES

APPENDIX I - Examples of CATPAC Cluster Analysis

Original ASPECTS.DOC analysis for the Communication Department at SUNY

GALILEO PROJECT - CATPAC - CLUSTER ANALYSIS

	UNIQUE WORDS	WINDOW SIZE	CYCLES	Top 10 levels # clusters/ total # wds	Top 15 levels # clusters/ total # wds	clamping y/n	values y/n
Aspects.Doc	50	2	1	9/19	12/27	y	n
	50	3	1	9/19	12/27	y	n
	50	4	1	7/17	10/25	y	n
	50	5	1	5/15	7/22	y	n
	50	5	2	4/16	5/22	y	n
	50	6	2	4/15	5/21	y	n
	50	7	2	2/12	3/18	y	n
	50	8	1	2/12	2/17	n	n
	50	12	1	1/11	2/17	n	n

Major Clusters that emerged are listed below. The Aspects document contains references to these clusters using the number of each concept. Multiples of the concept may occur in the document.

1. Small Groups
2. Communication Theory Classes
3. Public Speaking
4. Real World
5. Student Interaction
6. Small Communication Classes Rewarding
7. Non-verbal communication class
8. Practical Theory
9. Writing/Advertising

To summarize, students liked small groups with student interaction, like the non-verbal communication class. They found small communications classes rewarding. They preferred practical theory to Communication Theory. Writing, advertising classes, and public relations that reflected the real world were most helpful.

APPENDIX I - 2

Original GOALS.DOC analysis of the students of SUNY

GALILEO PROJECT - CATPAC - CLUSTER ANALYSIS

	UNIQUE WORDS	WINDOW SIZE	CYCLES	Top 10 levels # clusters/ total # wds	Top 15 levels # clusters/ total # wds	clamping y/n	values y/n
Goals.Doc	50	2	1	9/19	13/28	y	n
	50	3	1	6/16	9/24	y	n
	50	4	1	5/15	9/24	y	n
	50	5	2	1/11	3/19	y	n
	50	6	2	3/15	2/19	y	n
	50	7	2	2/15	3/21	y	n
	50	8	1	3/13	2/17	n	n
	50	12	1	2/13	2/18	n	n

Major Clusters that emerged are listed below. The Goals document contains references to these clusters using the number of each concept. Multiples of the concept may occur in the document.

1. Business Major
2. Media/Broadcasting Classeses
3. Journalism
4. Writing Skills
5. Advertising
6. Learn Public Relations
7. Specific Job/Specific Career
8. Learn communication/become communicator

To summarize, many students were considering a business major to be most helpful in reaching goals. They were interested in the fields of journalism, advertising, and public relations. They wanted to write and speak better, and needed specific job/career oriented classes to help them achieve these goals.

APPENDIX I - 3

Original CHANGES.DOC analysis the students of SUNY would make to create a better Communications Department

GALILEO PROJECT - CATPAC - CLUSTER ANALYSIS

	UNIQUE WORDS	WINDOW SIZE	CYCLES	Top 10 levels # clusters/ total # wds	Top 15 levels # clusters/ total # wds	clamping y/n	nodes y/n
Changes.Doc	50	2	1	7/17	11/26	y	n
	50	3	1	6/16	10/25	y	n
	50	4	1	4/14	7/24	y	n
	50	6	2	2/12	2/17	y	n
	50	7	2	2/16	3/21	y	n
	50	8	1	2/12	2/17	n	n
	50	12	1	1/11	1/16	n	n

Major Clusters that emerged are listed below. The Changes document contains references to these clusters using the number of each concept. Multiples of the concept may occur in the document.

1. Communication Theory Classes
2. Public Relations
3. Learn Advertising
4. Real World
5. Offer Hands-on Research
6. Different Curriculum

To summarize, changes that students thought would help the Department more were to replace theory with more 'hands-on' research. Offer classes which would prepare them for careers in public relations and advertising. Students felt a different curriculum which stressed the "real world" would be extremely helpful.

APPENDIX II - Example of CATPAC Cluster Analysis with sujective ratings included.

CLUSTER ANALYSIS of LEADER7 Document

EXPs: <u>(USING SLIDE)</u>

Initial Runs with additional level of analysis

RUN	UNIQUE WORDS	WORDS FOUND	WINDOW SIZE	CYCLES	Levels of analysis are given as # CLUSTERS/TOTAL # WORDS Level 5	Level 10	Level 15	Level 20	Window slide	Clamping y/n	Ref. Run	Rate Subjective Drew/Marcy
1.	30	21	2	1	5/10	8/18	6/21	1/21	1	y	Exp1	60/80
2.	100	68	9	4	2/7	3/34	4/40	5/58	3	y	Exp2	25/40
3.	100	68	2	1	5/10	10/20	15/30	20/40	3	n	Exp3	55/60
4.	50	42	2	4	5/10	10/20	13/28	14/33	1	n	Exp4	55/75
5.	30	21	5	4	4/9	5/15	3/18	1/21	2	n	Exp5	60/50
6.	30	21	2	3	5/10	8/18	6/21	1/21	2	n	Exp6	70/80
7.	50	40	9	1	2/7	7/17	11/26	13/33	2	n	Exp7	75/85
8.	50	40	5	1	3/8	8/18	13/28	17/37	3	n	Exp8	70/85
9.	50	40	9	3	3/8	3/14	3/19	3/24	3	y	Exp9	85/45
10.	100	68	5	1	5/10	8/18	11/26	13/33	2	y	Exp10	80/85
11.	100	68	5	2	3/8	7/17	12/27	16/36	1	n	Exp11	80/90
12.	30	21	9	1	2/7	3/13	2/17	1/21	1	n	Exp12	75/55

APPENDIX III - Example of EDO - Experimental Design Optimizer results on partial runs.

EDO -- EXPERIMENTAL DESIGN OPTIMIZER
VERSION 5.0

TITLE:- test
Mon Mar 09 13:39:01 1992

VAR_NO	VAR_NAME	N_LEVELS	LEVELS			
1	window	4	2	4	6	8
2	cycles	4	1	2	3	4
3	clamp	2	1	2		
4	words	2	30	50		

Squared Variables:

VAR_NO	SQUARED_EFFECTS
1	YES
2	YES

The total time taken is 0.15 minutes

The Number of Experiments in the Design is 9

Prediction Error Ratio for all admissible expts:

EOP	window	cycles	clamp	words
0.958	2.00	1.00	1.00	30.00
1.039	2.00	1.00	1.00	50.00
0.846	2.00	1.00	2.00	30.00
0.832	2.00	1.00	2.00	50.00
1.077	2.00	2.00	1.00	30.00
0.942	2.00	2.00	1.00	50.00
1.041	2.00	2.00	2.00	30.00
0.792	2.00	2.00	2.00	50.00
1.082	2.00	3.00	1.00	30.00
0.929	2.00	3.00	1.00	50.00
1.073	2.00	3.00	2.00	30.00
0.812	2.00	3.00	2.00	50.00
0.919	2.00	4.00	1.00	30.00
0.948	2.00	4.00	1.00	50.00
0.902	2.00	4.00	2.00	30.00
0.827	2.00	4.00	2.00	50.00
0.977	4.00	1.00	1.00	30.00
1.181	4.00	1.00	1.00	50.00
0.844	4.00	1.00	2.00	30.00
0.985	4.00	1.00	2.00	50.00
1.013	4.00	2.00	1.00	30.00
1.017	4.00	2.00	1.00	50.00
0.955	4.00	2.00	2.00	30.00
0.857	4.00	2.00	2.00	50.00
0.996	4.00	3.00	1.00	30.00
0.981	4.00	3.00	1.00	50.00
0.967	4.00	3.00	2.00	30.00
0.849	4.00	3.00	2.00	50.00
0.862	4.00	4.00	1.00	30.00
1.038	4.00	4.00	1.00	50.00
0.821	4.00	4.00	2.00	30.00
0.907	4.00	4.00	2.00	50.00
0.942	6.00	1.00	1.00	30.00
1.193	6.00	1.00	1.00	50.00
0.812	6.00	1.00	2.00	30.00
1.006	6.00	1.00	2.00	50.00
0.958	6.00	2.00	1.00	30.00
1.009	6.00	2.00	1.00	50.00
0.903	6.00	2.00	2.00	30.00
0.856	6.00	2.00	2.00	50.00
0.945	6.00	3.00	1.00	30.00
0.979	6.00	3.00	1.00	50.00
0.921	6.00	3.00	2.00	30.00
0.854	6.00	3.00	2.00	50.00
0.843	6.00	4.00	1.00	30.00
1.067	6.00	4.00	1.00	50.00
0.810	6.00	4.00	2.00	30.00

1.097	8.00	1.00	1.00	50.00
0.772	8.00	1.00	2.00	30.00
0.926	8.00	1.00	2.00	50.00
0.933	8.00	2.00	1.00	30.00
0.939	8.00	2.00	1.00	50.00
0.913	8.00	2.00	2.00	30.00
0.813	8.00	2.00	2.00	50.00
0.957	8.00	3.00	1.00	30.00
0.944	8.00	3.00	1.00	50.00
0.968	8.00	3.00	2.00	30.00
0.853	8.00	3.00	2.00	50.00
0.888	8.00	4.00	1.00	30.00
1.061	8.00	4.00	1.00	50.00
0.893	8.00	4.00	2.00	30.00
0.975	8.00	4.00	2.00	50.00

The Average error of prediction is
0.944

The Maximum error of prediction is 1.193
at 6 1 1 50
The Minimum error of prediction is 0.772
at 8 1 2 30
The total number of admissible expts. are : 64

The EDO Generated Design is ...

	window	cycles	clamp	words
1	2	1	1	30
2	8	4	1	30
3	2	4	2	50
4	8	1	2	50
5	6	3	1	50
6	8	2	2	30
7	6	1	2	30
8	2	2	2	50
9	4	4	2	30

OPTION CHOSEN --- SUBTRACT

The Number Of Expts.Deleted From The Design Are - 1

OPTION CHOSEN --- EVALUATE

The Average error of prediction is
1.037

The Maximum error of prediction is 1.403
at 2 3 2 30
The Minimum error of prediction is 0.778
at 8 1 2 30
The total number of admissible expts. are : 64

OPTION CHOSEN --- LIST PRINTER

	window	cycles	clamp	words
1	2	1	1	30
2	8	4	1	30
3	2	4	2	50
4	8	1	2	50
5	6	3	1	50
6	8	2	2	30
7	6	1	2	30
8	4	4	2	30

OPTION CHOSEN --- SUBTRACT

The Number Of Expts.Deleted From The Design Are - 1

OPTION CHOSEN --- EVALUATE

The Average error of prediction is
1.148

The Maximum error of prediction is 1.453
at 2 3 2 30
The Minimum error of prediction is 0.916
at 8 4 2 30
The total number of admissible expts. are : 64

OPTION CHOSEN --- LIST PRINTER

window cycles clamp words

APPENDIX III - 3

2	8	4	1	30
3	8	4	2	50
4	8	1	2	50
5	6	3	1	50
6	8	2	2	30
7	4	4	2	30

OPTION CHOSEN --- ADD

The Number Of Expts. Added To The Design Are - 1

OPTION CHOSEN --- EVALUATE

Prediction Error Ratio for all admissible expts:

EOP	window	cycles	clamp	words
0.982	2.00	1.00	1.00	30.00
1.083	2.00	1.00	1.00	50.00
0.950	2.00	1.00	2.00	30.00
0.982	2.00	1.00	2.00	50.00
0.993	2.00	2.00	1.00	30.00
1.195	2.00	2.00	1.00	50.00
0.845	2.00	2.00	2.00	30.00
0.993	2.00	2.00	2.00	50.00
1.007	2.00	3.00	1.00	30.00
1.223	2.00	3.00	1.00	50.00
0.824	2.00	3.00	2.00	30.00
1.007	2.00	3.00	2.00	50.00
0.926	2.00	4.00	1.00	30.00
1.121	2.00	4.00	1.00	50.00
0.778	2.00	4.00	2.00	30.00
0.926	2.00	4.00	2.00	50.00
1.296	4.00	1.00	1.00	30.00
1.242	4.00	1.00	1.00	50.00
1.402	4.00	1.00	2.00	30.00
1.296	4.00	1.00	2.00	50.00
1.065	4.00	2.00	1.00	30.00
1.098	4.00	2.00	1.00	50.00
1.099	4.00	2.00	2.00	30.00
1.065	4.00	2.00	2.00	50.00
0.969	4.00	3.00	1.00	30.00
1.037	4.00	3.00	1.00	50.00
0.975	4.00	3.00	2.00	30.00
0.969	4.00	3.00	2.00	50.00
0.952	4.00	4.00	1.00	30.00
0.980	4.00	4.00	1.00	50.00
1.000	4.00	4.00	2.00	30.00
0.952	4.00	4.00	2.00	50.00
1.281	6.00	1.00	1.00	30.00
1.209	6.00	1.00	1.00	50.00
1.403	6.00	1.00	2.00	30.00
1.281	6.00	1.00	2.00	50.00
1.038	6.00	2.00	1.00	30.00
1.052	6.00	2.00	1.00	50.00
1.093	6.00	2.00	2.00	30.00
1.038	6.00	2.00	2.00	50.00
0.952	6.00	3.00	1.00	30.00
1.000	6.00	3.00	1.00	50.00
0.980	6.00	3.00	2.00	30.00
0.952	6.00	3.00	2.00	50.00
0.969	6.00	4.00	1.00	30.00
0.975	6.00	4.00	1.00	50.00
1.037	6.00	4.00	2.00	30.00
0.969	6.00	4.00	2.00	50.00
0.926	8.00	1.00	1.00	30.00
0.969	8.00	1.00	1.00	50.00
0.960	8.00	1.00	2.00	30.00
0.926	8.00	1.00	2.00	50.00
0.907	8.00	2.00	1.00	30.00
1.056	8.00	2.00	1.00	50.00
0.824	8.00	2.00	2.00	30.00
0.907	8.00	2.00	2.00	50.00
0.960	8.00	3.00	1.00	30.00
1.130	8.00	3.00	1.00	50.00
0.645	8.00	3.00	2.00	30.00
0.960	8.00	3.00	2.00	50.00
0.982	8.00	4.00	1.00	30.00
1.112	8.00	4.00	1.00	50.00
0.916	8.00	4.00	2.00	30.00
0.982	8.00	4.00	2.00	50.00

The Average error of prediction is
1.037

The Maximum error of prediction is 1.403
at 6 1 2 30
The Minimum error of prediction is 0.778
at 2 4 2 30
The total number of admissible expts. are : 64