

Intelligent Instrumentation

Alice M. Harper

University of Texas at El Paso, El Paso, TX 79968
and

Shirley A. Liebman

Aberdeen Proving Grounds, MD 21105-5066

Accepted: July 1, 1985

Feasibility studies on the application of multivariate statistical and mathematical algorithms to chemical problems have proliferated over the past 15 years. In contrast to this, most commercially available computerized analytical instruments have used in the data systems only those algorithms which acquire, display, or massage raw data. These techniques would fall into the "preprocessing stage" of sophisticated data analysis studies. An exception to this is, of course, are the efforts of instrumental manufacturers in the area of spectral library search. Recent firsthand experiences with several groups designing instruments and analytical procedures for which rudimentary statistical techniques were inadequate have focused efforts on the question of multivariate data systems for instrumentation. That a sophisticated and versatile mathematical data system must also be intelligent (not just a number cruncher) is an overriding consideration in our current development. For example, consider a system set up to perform pattern recognition. Either all users need to understand the interaction of data structures with algorithm type and assumptions or the data system must possess such an understanding. It would seem, in such cases, that the algorithm driver should include an expert systems specifically geared to mimic a chemometrician as well as one to aid interpretation in terms of the chemistry of a result. Three areas of modern analysis will be discussed: 1) developments in the area of preprocessing and pattern recognition systems for pyrolysis gas chromatography and pyrolysis mass spectrometry; 2) methods projected for the cross interpretation of several analysis techniques such as several spectroscopies on single samples; and 3) the advantages of having well defined chemical problems for expert systems/pattern recognition automation.

Key words: data systems, intelligent; instrumentation; multivariate algorithms, statistical and mathematical; pattern recognition; preprocessing; pyrolysis gas chromatography; pyrolysis mass spectrometry.

Modern computer hardware and software technologies have revolutionized the direction of analytical chemistry over the past 15 years. Standard multivariate statistical techniques applied to optimization and control of instrumentation as well as routine decision making are at the forefront of new instrumental methods such as biomedical 3-dimensional scanners and pyrolysis MS and GC/MS as well as more established measurement techniques. Despite these advances, little attention has been paid to the exploitation of intelligent computerized instrumentation in the design phase of chemical research.

Instrumental intelligence is the ability of a scientific instrument to perform a single or several intelligence func-

tions in such a way that operations normally performed by the scientist are completely under automated computer control and decision making. Under this definition, intelligent instruments are quite common. Indeed in recent years, manufacturers of small scientific equipment have used the term "intelligent" in conjunction with single purpose items such as recorders to describe the addition of software and/or programmability to the device. Concurrent with this, larger scientific instruments have been marketed with data systems hosting a wide variety of intelligence functions including control and optimization of instrumental variables, optional modes of experimental design, signal averaging, filtering and integration as well as post analysis data massaging and library search interpretation. Although instruments with the software to perform sophisticated intelligence operations exist, they are not so readily marketed as intelligent instruments. For example, modern pulsed Fourier transforms nuclear magnetic resonance spectrometers (NMR) have microcomputers built into the system, operate over a wide range

About the Authors: Alice M. Harper is with the Chemistry Department at the University of Texas at El Paso. Shirley A. Liebman is with the Ballistics Research Laboratories at Aberdeen Proving Grounds.

of NMR experimental designs, and control instrumental parameters; however, decision making is, for the most part, an operator based function. For this reason, such instruments have limited "intelligence" in comparison to the level of intelligence required to carry an experiment to completion without extensive interaction with the operator. In fact, it could be that more intelligent research instruments might also be less versatile. The limitation is the current state of technology of intelligence programming. The instrument, if used for routine analysis where problem statements can be well defined, could operate with no loss of utility as a totally automated and intelligent instrument.

Figure 1 gives insight into the problems arising in creating intelligence programs for instrumentation. Figure 1a) is an analytical chemist's perception of a totally automated exper-

imental design [1]¹. As can be seen, the experiment remains unspecified without an initial problem statement. The issue of intelligent data systems for single instruments is one of either defining the set of all possible problem statements in an evolutionary design or restricting the analysis to a single well defined problem. Figure 1b) is an example of a first stage multivariate data analysis system proposed for research applications in pyrolysis mass spectrometry [2]. The design is one that leaves the problem statement, data interpretation and decision making entirely in the hands of the scientist. What is really described in this figure is a statistical package residing on a microcomputer which receives data from the mass spectrometer. However, as Isenhour has

¹Figures in brackets indicate literature references.

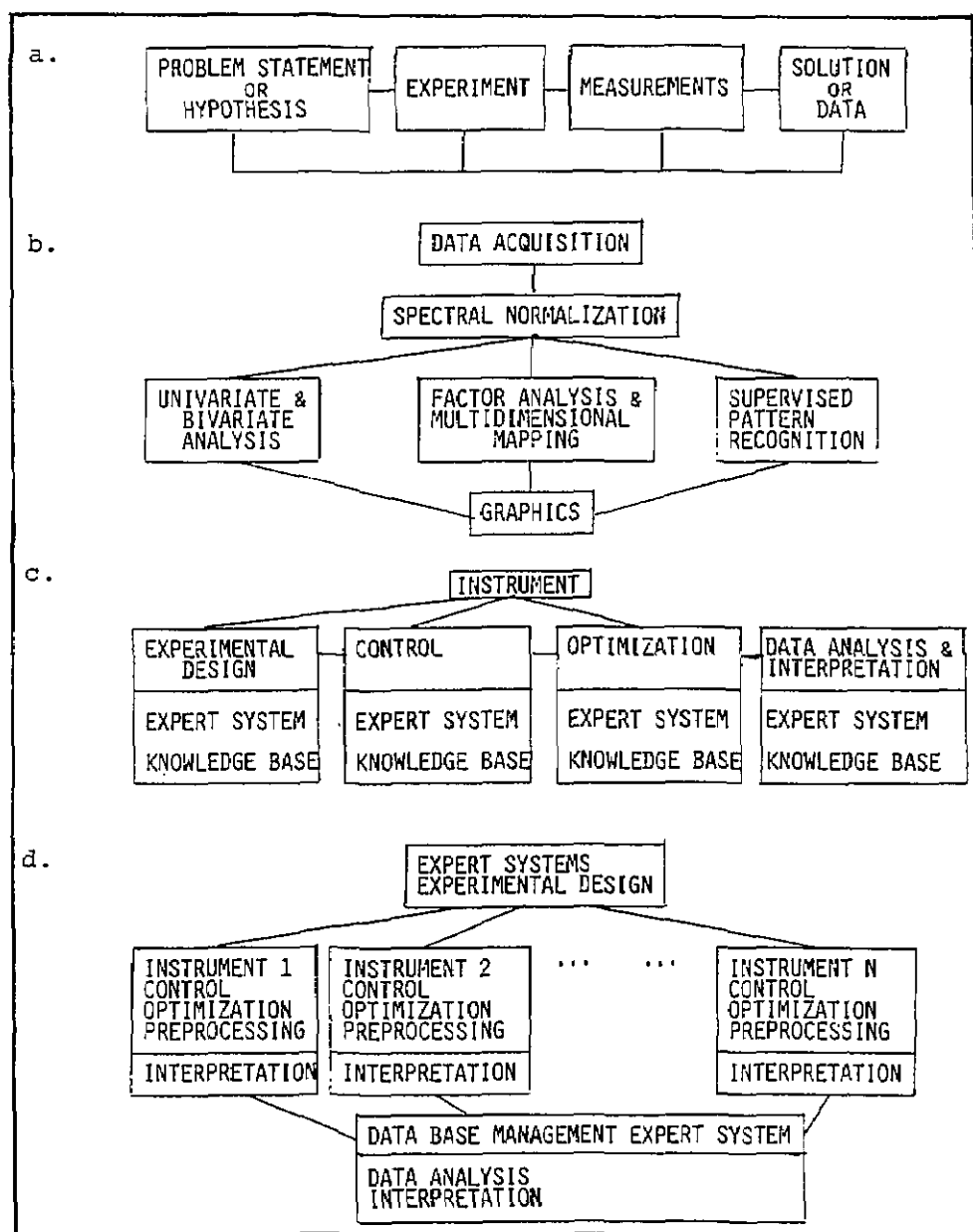


Figure 1—Four designs for instrument data systems.

- Totally automated experimental design and decision making.
- Multivariate analysis for spectral instruments.
- Expert systems driven single purpose instrument.
- Laboratory automation using expert systems drivers.

demonstrated, multivariate factoring of spectral libraries offers advantages in the interpretation of complex single spectra [3]. It would seem that the incorporation of a multivariate statistics package in the data system is a key element in intelligence programming for many instruments. Figure 1c) incorporates the expert system approach to intelligence for instrumental systems. Under this design, the instrument can accommodate a series of problem statements and decision networks at various analysis stages and uses an intelligent driver for the multivariate analysis and interpretive stages of the analysis. Figure 1d) places such a system into the research laboratory controlling a variety of instruments and interpreting results based on one or more analyses.

Although the diagrams of figures 1b-1d are not comprehensive designs for total automation, they do provide a hierarchy for linking problem statements and decision making into multivariate research problems. After a brief discussion of components of intelligence designs, an example will be presented of the feasibility of developing expert system data reduction for pyrolysis analysis problems.

Problem Statements: Ideally a problem statement is analogous to the standard hypothesis in statistical analysis. Under the hypothesis a knowledge base can be collected and the hypothesis tested. For example, a patient does or does not carry a genetic trait [4]. Unfortunately, chemical research problems are often ill-specified and the problem statement may become a hierarchy of data investigations leading to one or more problem statements. For example, 1) are there differences in the chemical composition of a series of samples? if differences do occur; 2) what are the nature of the differences?; 3) do the chemical differences correlate with observed changes in physical properties?; and 4) can physical properties be predicted from the chemical differences? [5,6].

Knowledge Base: In order for an instrument to operate as an intelligent system, it must have the knowledge base necessary to arrive at a solution for each of its problem statements. This knowledge base may contain data, rules ("Mass peak 94 is phenol"), programs, and heuristic knowledge.

Consider the knowledge base required for setting up and operating routine analyses of polymer composition by pyrolysis gas chromatography. Possible problem statement areas include optimization of pyrolysis parameters, chromatographic conditions and interface characteristics, control of instrument and data acquisition parameters, data reduction, and data interpretation. The knowledge base must include all information necessary to each of the proposed problem statements. For optimization of parameters, rules governing the detection of an optimum, algorithms (e.g., "simplex" [7]) for efficiently moving toward an optimum, rules for hierarchical movements within the algorithm, rules for the detection of poor optimization surface structure, and representative previous optima data might be employed in the decision making. Such optima

will be determined, in part, by the polymer degradation characteristics and therefore will not, in this case, be independent of the samples used in the analysis. Instrumental control might employ a knowledge base of rules for the automation of events such as the initiation of sample pyrolysis and data collection. Data reduction for this method will require the rules and data necessary for baseline correction, chromatographic normalization, and peak matching. The knowledge base requires a "memory" of previously collected data to aid peak matching protocols, rules for baseline determinations, transformations for baseline correction, normalization rules and algorithms, and rules for acceptance or rejection of the chromatogram under consideration. Data interpretation on the other hand might require a library of previous chromatograms as well as rules and/or algorithms for interpretation of the current event based on a knowledge of past data with verified interpretations.

Development of the knowledge base is the expensive and time consuming operation in the development instrumental intelligence even when the application is highly specific. It must be remembered that each operation that a human might perform automatically from experience must be programmed into the data system. For this reason, among the attributes of the system, there needs to be evolutionary operation. In other words, in addition to long term knowledge, facts about the current data and new conjectures under consideration must also be easily accommodated.

Expert Systems Driven Multivariate Data Systems: The response of many modern chemical instruments (e.g., spectrometers and chromatographs) is inherently multivariate. For such instruments, data reduction and interpretation often consume a greater portion of analysis time than data collection, and requires scientific expertise. The time delay between data collection and decision making has become an acute problem for newer hyphenated techniques, such as gas chromatography-mass spectrometry and mass spectrometry-mass spectrometry, which are capable of collecting thousands of mass spectral peaks in a short period of time.

One possible solution to the problems imposed by large bodies of data is to incorporate into the instrument a data reduction system consisting of multivariate analysis methods. The problem encountered in the actual implementation of such a system is that few experts in instrumental analysis have the expertise for carrying multivariate statistical analyses. It has become increasingly apparent that instruments employing versatile multivariate based data systems should be capable of operating in a transparent data analysis mode. In order to accomplish this, the expertise of a chemometrician will need to be programmed into an expert systems driver for the instrument data system. Given a problem statement, and a knowledge base of rules from previous experience, the computer could decide from a variety of possibilities how to reduce the data and represent the results in a meaningful form. A relatively simplistic example of this is the problem

statement "Is there a correlation between two independent variables?" Invisible to the user would be the operations for determining the integrity of the variable distributions, a computation of the correlation and a determination of the significance of the correlation coefficient of the relationship. The result might take a simple form such as "there appears to be a significant correlation between the first variable and the logarithm of the second variable. Would you like to see the computed values?"

Demonstration of Expert System for Data Analysis in Curie-point Pyrolysis Mass Spectrometry

Pyrolysis mass spectrometry has come to the attention of both mass spectrometrists and chemometricians because of its utility in the analysis of polymeric materials and the complexity of the mass spectra produced by natural polymers and biopolymers. The technique involves degradation of the solid material by pyrolysis followed by mass spectrometry of the pyrolysis fragments. It has been demonstrated by the pioneering work of Meuzelaar (for example see [8]) that Curie-point pyrolyses of samples of biomaterials can produce profiles which, when properly normalized [2], are reproducible and diagnostic of the chemical similarities and dissimilarities among groups of samples and are quantitative under appropriate experimental designs [9]. Because the process of pyrolysis followed by mass spectrometry of the network polymer of natural heterogeneous biopolymers produces mass spectra with peaks which tend to be highly correlated, the interpretation problems created by the vast number of, and the overlap of, the masses are solved through multivariate analysis of the mass spectral profiles. A data system for a pyrolysis mass spectrometer would be of limited utility without multivariate statistical methods [2].

To test the hypothesis that an expert systems approach to data reduction is potentially helpful and feasible, an expert systems driver was implemented to mimic the data analysis portion of a Rocky Mountain Coal study done at Biomaterials Profiling Center in Utah. Detailed results of the original study can be found in [5,6] and of the numerical methods used in [2]. Briefly, 102 Rocky Mountain Coals were analyzed in quadruplicate by pyrolysis mass spectrometry. The pyrolysis profiles were added to a preexisting data set containing conventional measurements on the same coal samples (see table 1). After normalization of the profiles by the method of Eshwis et al. [10], average mass spectra were analyzed using multivariate analysis techniques.

Figure 2 is a minimal design for an expert systems driven data acquisition and analysis system for pyrolysis mass spectrometry. Figure 3 shows the design of the data bases for the present application. As diagrammed in figure 2, each

Table 1. Conventional measurement contained in the "old data" data base for the Rocky Mountain coals.

Conventional Measurements	
Vitrinite	% Potassium
Fusinite	% Phosphorus
Semifusinite	% Moisture
Macrinite	% Pyritic sulfur
Liptinite	% Mineral matter
Vitrinite Reflectance	% Volatiles
% Silicon	% Organic sulfur
% Aluminum	Calorific value
% Titanium	% Organic carbon
% Magnesium	% Organic hydrogen
% Calcium	% Organic sulfur
% Sodium	% Organic nitrogen
	% Organic oxygen

operation of the instrument requires an expert systems driver and decision module for its automation. A rudimentary expert systems was built to mimic the decision making used for statistical analysis of coal pyrolysis patterns. This system demonstrates the problems and pitfalls associated with intelligent instrumental development.

Normalization is rarely an option in pyrolysis techniques since the size of the sample undergoing electron impact is determined by the quantity of pyrolysates actually making their way to the ion source of the mass spectrometer. For this reason, spectra are normalized to place each sample on a relative quantitative basis. Furthermore, when replicates of a single sample are analyzed in detail, it is found that some peaks replicate better than others. For example, if an organic solvent is used during the sample preparation, the mass peaks due to this solvent replicate poorly. The same is true of contaminants absorbed to the sample matrix. On the other hand, because the sample size in terms of total ion counts is a variable in these experiments, often one or more replicate spectra will exhibit outlying tendencies when compared to other replicates of the same sample.

NORMA [2], developed by Meuzelaar's group at Utah, is designed to select peaks with stable variance characteristics for inclusion in the normalization process. With this routine an expert interacts with the computer in a loop of peak deletions and replicate spectra deletions until a set of peaks is defined which stabilizes the normalization process.

An expert systems approach to the software interaction represents peaks by their variances over the samples and sample replicates and spectra by Euclidean distances between replicates over the mass units. A library data base for normalization is established which contains spectra patterns for commonly used solvents and commonly encountered contaminants as well as the background spectrum from the mass spectrometer. The expert systems are initialized by computation of peak variances. The variances are ordered high to low and the shape of the plot of ordered variances is

Figure 2—Expert systems driven pyrolysis mass spectrometer. Each stage of data acquisition, analysis, and interpretation has decision protocols based on a knowledge base of an expert. Note that the interaction between mathematical methods and the expert system chemometrician.

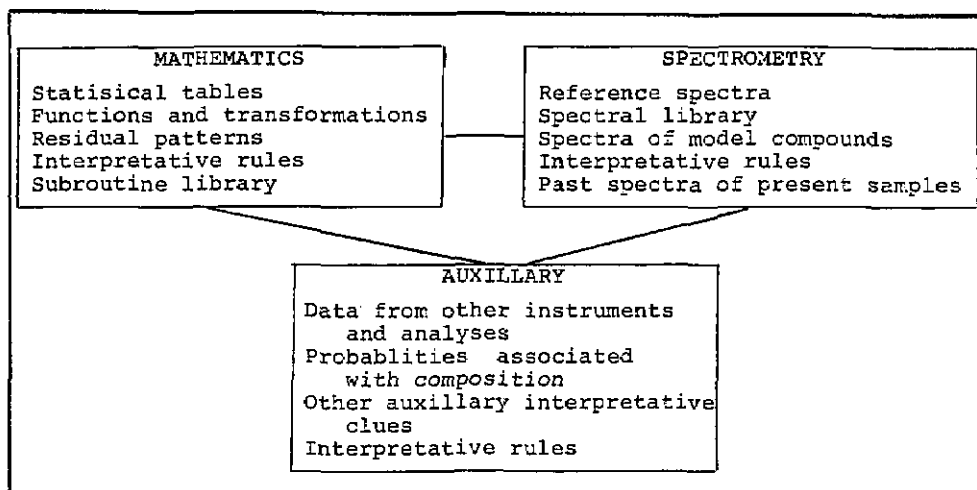
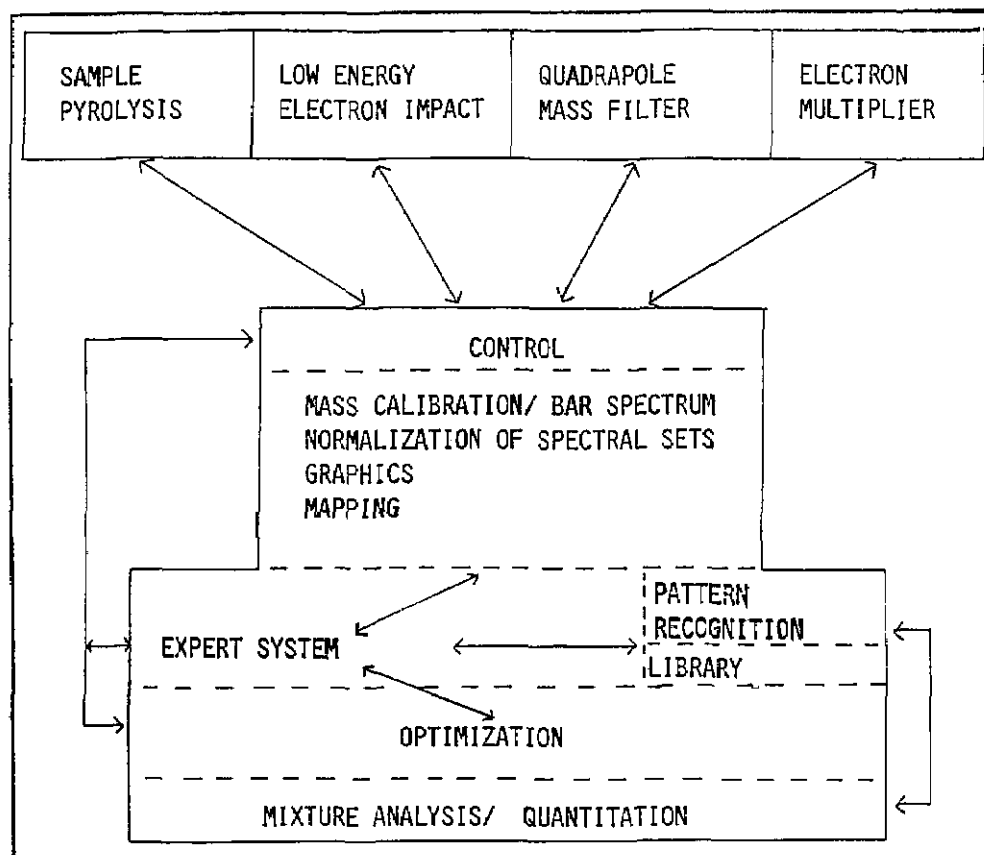


Figure 3—Types of data bases required for an application oriented expert system for instrumental analysis and interpretation. Knowledge representation takes two forms: 1. knowledge which when taken as a whole or in parts can be represented as a similarity or correlation; and 2. knowledge for which antecedent clauses must be satisfied in order for the interpretative/decision making process to occur.

analyzed. The library is searched under expert systems guidance to account for the peaks exhibiting high relative variance. For example, background spectrum is placed under consideration only when the total ion counts of the sample spectrum is less than a factor α of the background. After examination of the library, a decision is made as to which peaks will be deleted based on their expected contribution to the peak variation and with restrictions on the total number of peaks that can be deleted at this stage of analysis. This first deletion is a permanent deletion of peaks. None of the casual base peak deletions is reexamined at later stages.

The distance matrix of replicates is generated using the peaks remaining in the analysis and samples are deleted based on a comparison of their distances from the expected distance generated as a mean distance over the samples with variance σ_d . (Note that such a formulation may ignore systematic error among the sample replicates, and won't perform at the expert level under such a condition).

The next stage on analysis involves computation of both peak variances and sample distances. The rank order of the peak variances is compared. If variance reduction seems to be exhibited by a small set of peaks upon deletion of the

samples, the peaks involved are temporarily deleted, the samples previously deleted are brought back into analysis and the distances reexamined. The expert systems decide which will be deleted at this stage, peaks or samples based on the recomputed distances. The iterative decision making—statistical computation continues until a key set of peaks remain activated for normalization over all spectra. A diagram of the proposed normalization experts system is shown in figure 4.

The system, as described, does not completely emulate an expert for all applications of pyrolysis mass spectrometry. It has already been noted that the data evaluation does not address errors which arise from time dependent or systematic error. In addition to this, the decision making process, while designed to emulate the human decision process based on statistical results, does not necessarily operate on a one-to-one correspondence with a human expert. The problem is that two experts working on the same set of data may arrive at approximately the same results via slightly different procedural routes. The same is true of the expert system when compared to a human expert. The major deviations from a

human procedure found when working with this system is the lack of "intuition" or "fuzzy logic" that is used by the expert. The expert systems converges in more steps than is necessary by human interaction with the statistical algorithms and for some data bases, has trouble defining convergence at the solution. For example, the optimal cut off parameters for variance and distance change between data sets. These and other problems are best solved by training the expert systems to recognize the structure of a good spectral solution in addition to the structure of good statistics.

Correlation Based Hypothesis Testing

Perhaps the most often asked question of large data sets involves finding relationships between the variables or between the variables and an external parameter. Table 2 lists the form of the problem statements included in our data system for this demonstration. The term "relate" invokes one of several multivariate algorithms for the study of correlations in the data. The possible responses are the Pearson correlation coefficient, linear regression, factor analysis or canonical correlation analysis.

Consider the problem statement: What is the relationship of peak 34 (H_2S) in the mass spectrum of coal to the total organic sulfur from the conventional data matrix. "Relate current data, mass 34 to old data, organic sulfur over samples, all" results in a computation of the correlation coefficients, a estimate of significance and the confidence interval about the correlation. "Relate current data, mass 34 to old data, organic, sulfur over samples, all; Interpret using old data, organic sulfur" results in the additional computation of $(\text{organic sulfur}) = a(\text{mass 34}) + b$ with residuals E_i . The residual pattern is tested for randomness. A failure results in a search through the library for a reference residual pattern with similarities to the computed residual pattern.

The next relate function asks for a study of the relationships among variables in the current data set. "Relate current data, all with current data, all over samples, all" results in the factor analysis of the data correlation matrix. The loadings of the factor analysis are interpreted for the variable relationships seen along the orthogonal axes of the original rotation. This interpretation is an experts systems based analysis of the major peak series and will be discussed later.

Interpretation of the factor score is a difficult problem. Consider the two dimensional factor score projection of these data. Figure 5a is a projection without labels. Figure 5b is the same projection after a human expert has assigned sample labels corresponding to the geological source of the samples and an interpretation. Figure 5a is representative of the information about the factor scores stored by the computer. The addition of labels is readily accomplished but, without training, patterns formed by the sample labels

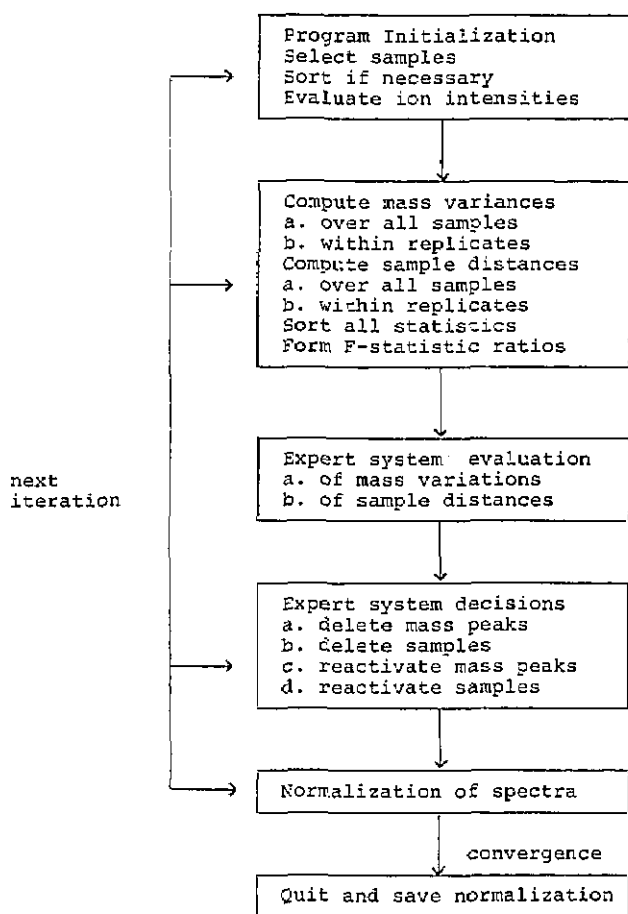


Figure 4—Interaction of expert systems and statistical computation for a rational normalization process for pyrolysis mass spectrometry.

Table 2. Variations of the expert system commands RELATE and INTERPRETE USING. The FORTRAN subroutine calling protocols are based on the data types used in each variable location of the command and on the command sequence. Note that the same command strings and rules can accommodate computation on questions about the data base transpose matrix when "...OVER samples ..." is replaced by "...OVER variables..."

RELATE data base 1, variable list 1 TO data base 2, variable list 2 OVER samples, sample list

Examples:

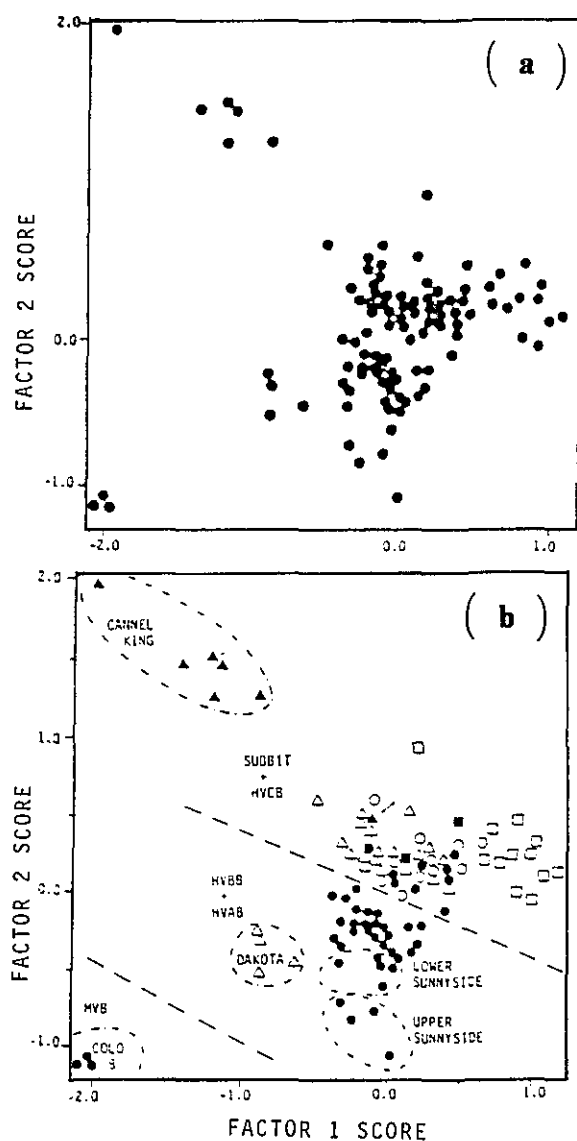
1. RELATE current data, mass 34 TO old data, organic sulfur OVER samples, match
(Results: correlation coefficient and significance test)
 2. RELATE current data, mass 34 TO old data, organic sulfur OVER samples, match; INTERPRETE USING old data, organic sulfur
(Results: correlation computed at least squares fit, significance test, and residual pattern evaluation)
 3. RELATE current data, all TO current data, all OVER samples, all
(Results: Factor analysis of current data and loading interpretation)
 4. RELATE current data, all TO old data, calorific value OVER samples, match, active; INTERPRETE USING old data, calorific value
(Results: Factor analysis of current data followed by target rotation to calorific value and interpretation of loadings)
 5. RELATE current data, all TO old data, all OVER samples, match
(Results: Canonical correlation analysis of two data bases and interpretation of mass spectral loadings)
-

within the space have little meaning. A generalized solution to this problem does not seem likely. Each study would require an elaborate knowledge base specific to the samples in order to interpret trends seen in this picture.

For the coal data, the old data matrix of conventional measurements provides a more easily implemented route for hypothesis testing on the pyrolysis mass spectral factors. Consider once more the "relate" command. After matching the two data sets by the logic used in [5], factor analysis of the PY/MS set followed by "Relate *current data 2, all* with *old data, calorific value* using *samples, active*; interpret using *old data, calorific value*" results in the regression analysis of $\text{calorific value} = \sum W (\text{factor scores})_{\text{py-ms}} + b$ producing both the variable and the sample relationships of the rotation of the Py/MS data to calorific value. The results are given in figure 6 and discussed in [6].

The last example of a correlation based statement is "Relate *current data, all* with *old data, all* over *samples, all*." This results in a canonical correlation analysis factoring both data matrices in such a way that the overlap in information between the sets is maximized. Four factors are extracted. The first two of these are shown in figures 7 through 9. This analysis showed these data sets to be approximately 80% correlated over the coal samples. The major chemistry of the conventional measurement trends are given above the spectral representation from the Py/MS factors.

Figure 5—Comparison of computer representation of factor scores (a.) to the same plot after interpretation by an expert (b.) Symbols represent sample specific relationships deduced by the expert. Dotted lines are a separation of the samples into classes not available to the computer in the pyrolysis data base. The knowledge base required to mimic the expert would form a reference book of information about the samples. The generation of such a knowledge base would be a formidable task.



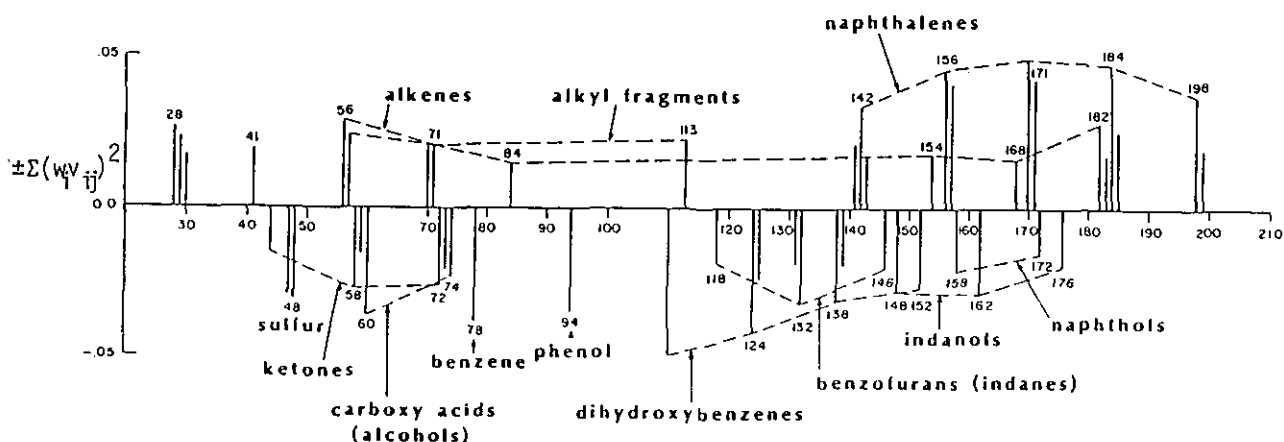


Figure 6—Chemical interpretation of the mass spectral peaks associated strongly with a targeted least squares rotation of pyrolysis factor scores to the external parameter calorific value. Peak interpretations were accomplished by the system described in figure 10. Results on chemical interpretation are identical to those of the original study, demonstrating that chemical information is more readily implemented as an expert system than sample specific information for this instrumental method.

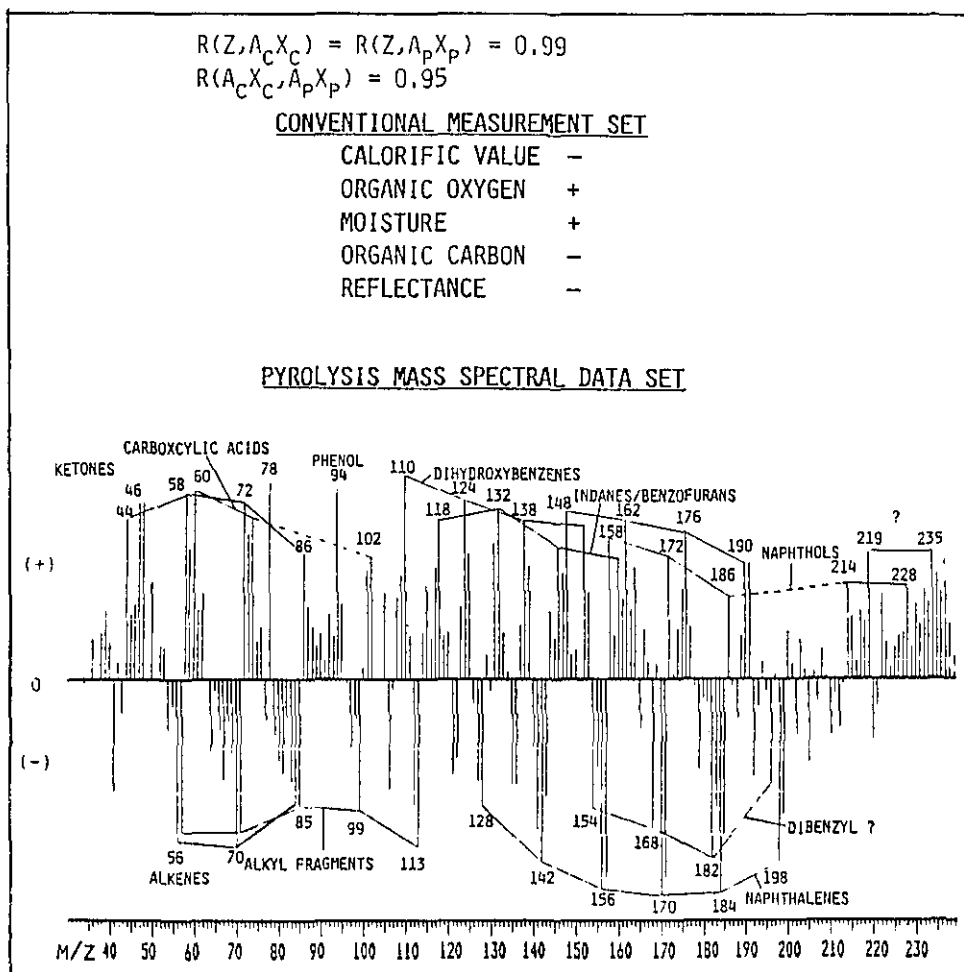


Figure 7—First canonical variate loadings for the rotation of pyrolysis mass spectra of coal samples to the conventional data matrix described in table 1. The correlation of the data bases to the derived variate (Z) is 0.99 and to each other is 0.95. A_cX_c and A_pX_p are the linear composites of the conventional and pyrolysis data bases respectively. Only the signs of the strongly correlated variables of the conventional data are given. Loadings for the pyrolysis data are given as positive and negative. Interpretations of mass peak were accomplished using the system described in figure 10.

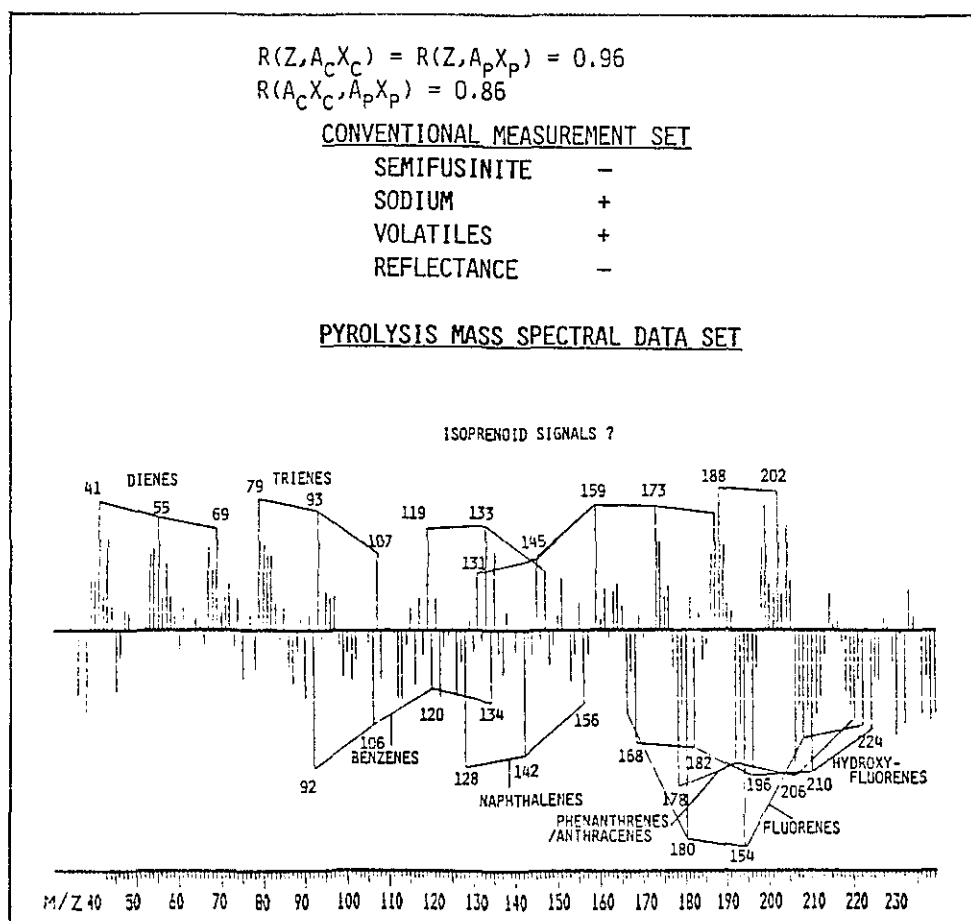


Figure 8—Second canonical variate loadings for the rotation of pyrolysis mass spectra to the conventional data matrix described in table 1. See figure 7 for an explanation of the symbols used in this figure. Interpretation by the system described in figure 10 failed for the positive loadings labeled as isoprenoid signals?. These were assigned by the authors.

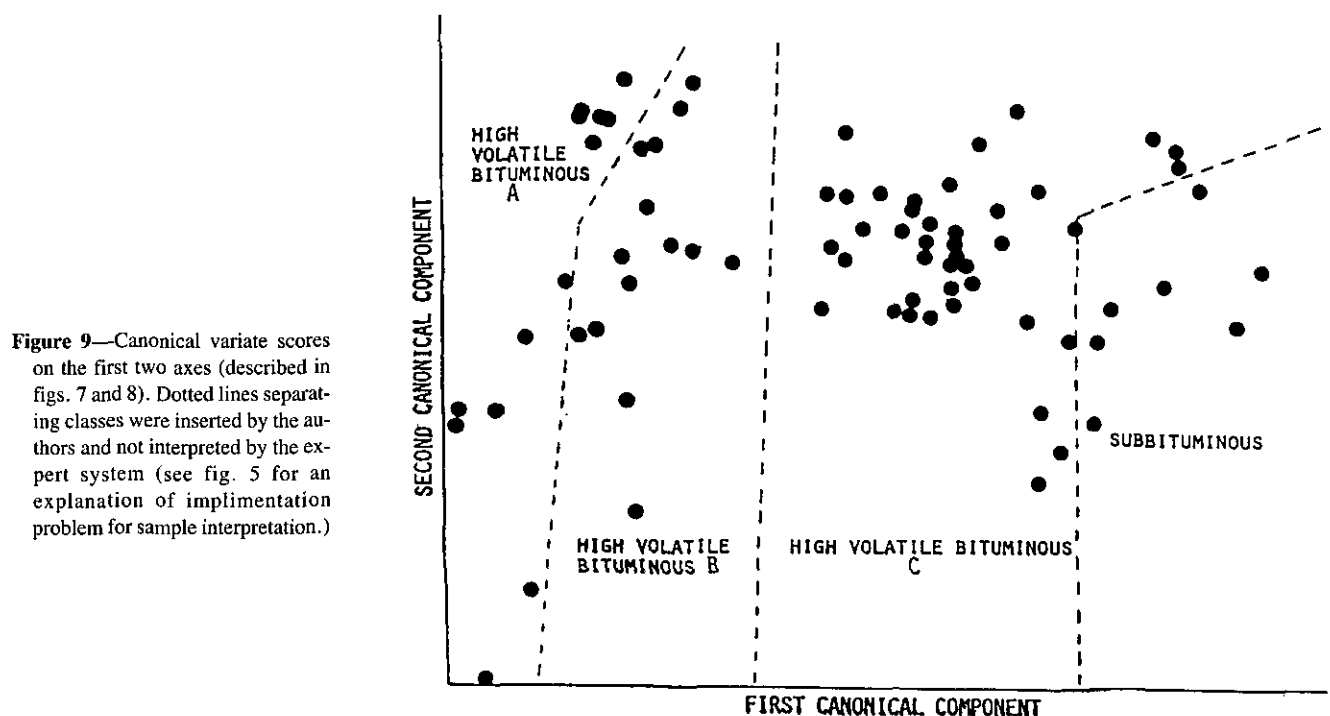


Figure 9—Canonical variate scores on the first two axes (described in figs. 7 and 8). Dotted lines separating classes were inserted by the authors and not interpreted by the expert system (see fig. 5 for an explanation of implimentation problem for sample interpretation.)

The assignment of chemical interpretations for mass spectra and for factor loadings is accomplished by an expert systems interpreter that can be used for any study in which the samples are coal. The data base for the interpretation includes commonly encountered chemical species in coal, the major peaks expected from their presence and, given two chemical species with similar patterns, the probability of their contribution as a major component to a coal pattern. Also included is a routine for generating molecular species from C,N,O, and S given the base peak molecular weight and the ion series. The operation of the expert system along with the results of each iteration are given in figure 10. Because Py-MS spectra contain many low intensity masses of questionable interpretation, and because factor loadings are rarely pure, only the most intense ion series patterns are interpreted. The program is initialized by setting an initial threshold limit (TL3). The ions (m/z) above the threshold are collected, sorted and given temporary chemical assignments. The threshold limit is lowered each time by a lesser factor until it encounters the "grassy" region of the spectrum. At this point, permanent ion series interpretations are assigned. The "?" appearing in the figure for Mass 104 means that no library interpretation of this peak was found and that the number of peaks in the ion series was below the limit set for generation of hypothetical molecular species. The entire system is diagramed in figure 11.

Discussion

The correlation work described in this paper along with similar considerations of other algorithms seem to support the possibility that, for a given instrument, an expert systems driver for data analysis can be developed which is independent of the nature of the chemical problem. The way to accomplish this is to build expert systems drivers to interpret the problem statement and to interpret the data analysis results. Viewed in this manner, an instrument-dependent expert system can generate the experimental design and optimization from the problem statement and pass to the data analysis driver the statistical elements necessary for its decision on the proper data reduction protocol. Learning is initiated when the data analysis system receives an unrecognizable set of elements. Otherwise, the expert system selects from the knowledge base the algorithm sequence for data reduction.

We are currently extending these concepts to the construction of an expert system, EXMAT, for experimental design, optimization, data reduction and interpretation of measurements made on a sample by a variety of thermal analysis instrumental techniques. TIMM® (General Research Corp., Mclean, VA) a FORTRAN-based expert system generator, has been enabled development of a heuristically-linked set of expert systems for material analy-

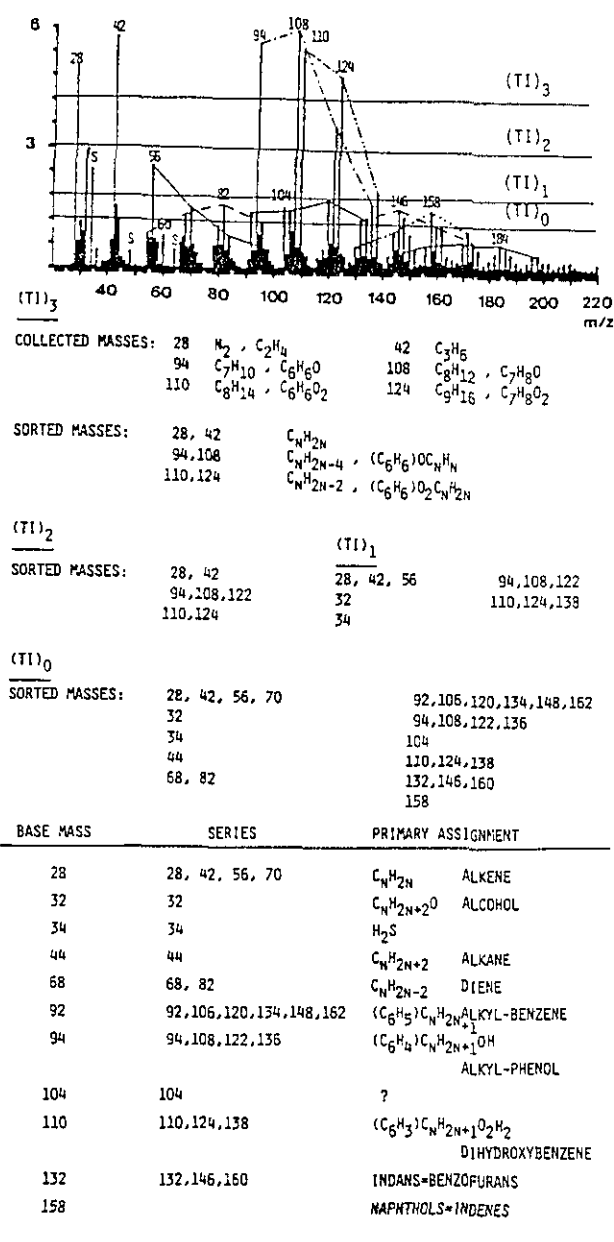


Figure 10—Operation of the expert system used to interpret the pyrolysis mass spectra and spectral loadings in the demonstration. The rules used to sort and interpret the peaks are based on ion series produced by mass differences of 14 (CH_2). Base peaks help terminate series for resolution of multiple interpretations. ? is a peak not present in the data base.

sis. The attributes of the TIMM® system are listed in table 3. In order to accomplish the analytical goals of this project, we have combined the concepts of specific instrumental intelligence with the goals set forth in figure 1a to provide a data system capable of experimental designs utilizing one or several analysis techniques. Each instrument retains its own control, design, preprocessing and interpretative expert systems unit but the data analysis unit has, of necessity, been generalized for analysis of data from a single or from

Figure 11—Steps in the expert system interpretation on pyrolysis mass spectra of coals.

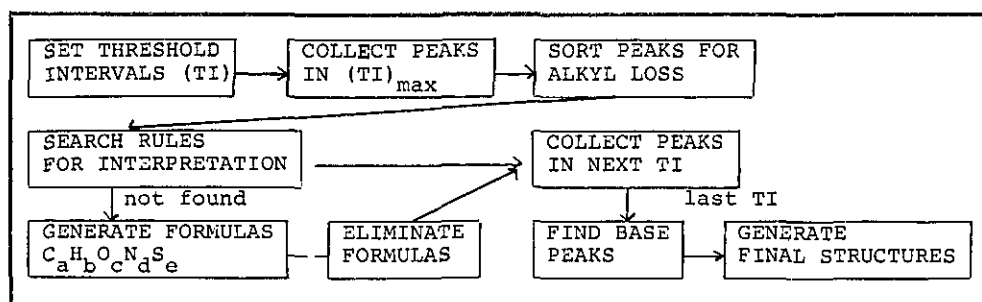


Table 3. Attributes of the TIMM^R expert system. TIMM^R was adapted to our application because addition of FORTRAN subroutine library calls is more readily accomplished than in other systems and because the decision logic can already use computationally based rules as well as chaining logic rules.

TIMM^R: A FORTRAN - BASED EXPERT SYSTEMS APPLICATIONS GENERATOR, General Research Corporation

Forward/backward chaining using analog rather than propositional representation

Knowledge base divided into two sections: declarative knowledge and knowledge body

Pattern-matching using a nearest neighbors search algorithm to compare current situation with antecedent clauses

Unique similarity metric computed from order information in declarative knowledge giving distance metric over all classes

Decision structure and knowledge body readily developed and modified by expert in any domain

Heuristically-linked expert system using implicit and explicit method permitting processing of "microdecisions" that are part of "macrodecisions"

Each system independently built, trained, exercised, checked for consistency and completeness and then generalized

multiple measurement techniques. Table 4 gives an outline of the decision making process for the experimental design expert system, and for data interpretation. Rule generation under the system is demonstrated in figure 12. The expert system strategy for the chemometrics portion of the system is similar to that described previously in the Py-MS example with the added feature that the system generate the protocol of analysis based on the initial problem statement and the available data. The system is in its earliest stages of development so any or all aspects of the proposed design are subject to modifications as experience is gained in training the system. Nevertheless, we feel that an expert systems such as ours offers a strategy for automation of laboratory instrumentation and interpretation under an expert systems approach.

```

Rule_1
If:
  SCOPE           IS  R&D
  SAMPLE AMT      IS  TRACE
  SAMPLE FORM     IS  POWDER
  SAMPLE PROCESS  IS  *
  SAMPLE HISTORY  IS  *
  INSTR. AVAIL    IS  ALL
Then:
  ANAL STRATEGY   IS  SPECTRMS(100)
  
```

Figure 12—Example of rule generation in EXMAT using the TIMM[®] expert system. Rule is from the expert system for analytical strategy.

Table 4. Example taken from overall organization structure of EXMAT, an expert systems for materials analysis.

ANALYTICAL STRATEGY	DATA INTERPRETATION
CHOICES: CHROMGC CHROMLC SPECTRFTIR SPECTRMS THERMTA ELEMEL	CHOICES: PATTERN RECOGNITION SAMPLE ID GROUP ID PEAKS ID NO ID COMBINE DB EXTEND DB CORRELATION
FACTORS: 1. SCOPE QUAL QUANT PURITY QUAL/QUANT TIME/FUND LIMIT TRACE R&D CORRELATION SCREEN	FACTORS: 1. DATA GENERATE FTIR DB MS DB LC DB TA DB GC-FTIR DB GC-MS DB GC DB 2. GC DB TREATMENT DIRECT COMPARE CHEMOMETRICS DATA SET ID 3. FTIR DB TREATMENT DIRECT COMPARE PAIRS SEARCH 4. ETC. FOR EACH DATA BASE (DB)

References

- [1] Harper, A.M., Polymer Characterization Using Gas Chromatography and Pyrolysis, S. Liebman and E.J. Levy, Marcel Dekker, Inc. (1985).
- [2] Harper, A.M.; H.C.L. Meuzelaar, G.S. Metcalf, and D.L. Pope, Analytical Pyrolysis Techniques and Applications, K. Voorhees, ed. 157-195 (1984).
- [3] Owens, P.M.; R.B. Lamb and T.L. Isenhour, Anal. Chem., **54** 2344 (1982).
- [4] McMurry, J.E.; J.A. Pino, P.C. Jurs, B. Lavine, and A.M. Harper, Anal. Chem., vol. 57, 295-302 (1985).
- [5] Meuzelaar, H.L.C., and A.M. Harper, Characterization and Classification of Rocky Mountain Coals by Curie-Point Mass Spectrometry, Fuels, vol. 63, 639-652 (1984).
- [6] Harper, A.M.; H.L.C. Meuzelaar, and P.H. Given, Fuels, vol. 63, 793-799 (1984).
- [7] Deming, S.N.; S.L. Morgan, and M.R. Willcott, American Laboratory, October, 13 (1976).
- [8] Meuzelaar, H.L.C.; J. Haverkamp, and F.D. Hileman, Curie Point Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam (1982).
- [9] Windig, W.; P.G. Kistemaker, and J. Haverkamp, J. Anal. Appl. Pyrol., **3**, 199 (1982).
- [10] Eshwis, W.; P.G. Kistemaker and H.L.C. Meuzelaar, Analytical Pyrolysis, CER, Jones, ed., Elsevier, Amsterdam, 151-166 (1977).

DISCUSSION

of the Harper-Liebman paper, Intelligent Instrumentation

Richard J. Beckman

Los Alamos National Laboratory

There has been an instrumentation revolution in the chemical world which has changed the way both chemists and statisticians think. Instrumentation has lead chemists to multivariate data—much multivariate data. Gone are the days when the chemist takes three univariate measurements and discards the most outlying.

Faced with these large arrays of data the chemist can become somewhat lost in the large assemblage of multivariate methods available for the analysis of the data. It is extremely difficult for the chemist—and the statistician for that matter—to form hypotheses and develop answers about the chemical systems under investigation when faced with large amounts of multivariate chemical data.

Professor Harper proposes an intelligent instrument to solve the problem of the analysis and interpretation of the data. This machine will perform the experiments, formulate the hypotheses, and “understand” the chemical systems under investigation.

What impact will such an instrument have on both chemists and statisticians? For the chemist, such an instrument will allow more time for experimentation, more time to think about the chemical systems under investigation, a better understanding of the system, and better statistical and numerical analyses. There would be a chemometrician in

every instrument! For the statistician, the instrument will mean the removal of outliers, trimmed data, automated regressions, and automated multivariate analyses. Most important, the entire model building process will be automated.

There are some things to worry about with intelligent instruments. Will the chemist know how the data have been reduced and the meaning of the analysis? Instruments made today do some data reduction, such as calibration and trimming, and the methods used in this reduction are seldom known by the chemist. With a totally automated system the chemist is likely to know less about the analysis than he does with the systems in use today.

The statistician when reading the paper of Professor Harper probably asks what is the role of the statistician in this process? Will the statistician be replaced with a microchip? *Can* the statistician be replaced with a microchip? In my view the statistician will be replaced by a microchip in instruments such as those discussed by Professor Harper. This will happen with or without the help of the statistician, but it is with the statistician's help that good statistical practices will be part of the intelligent instrument.

Professor Harper should be thanked for her view of the future chemical laboratory. This is an exciting time for both the chemist and the statistician to work and learn together.