

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231258594>

Computer series. 20: Teaching chemometrics: A course on application of mathematical techniques to chemistry

ARTICLE *in* JOURNAL OF CHEMICAL EDUCATION · JULY 1981

Impact Factor: 1.11 · DOI: 10.1021/ed058p646

CITATIONS

10

READS

15

2 AUTHORS, INCLUDING:



Mike Delaney

Massachusetts Water Resources Authority

43 PUBLICATIONS 523 CITATIONS

[SEE PROFILE](#)

Teaching Chemometrics: A Course on Application of Mathematical Techniques to Chemistry

Michael F. Delaney¹ and F. Vincent Warren, Jr.²

Tufts University, Medford, MA 02155

Computers are becoming commonplace in modern chemical instrumentation due to their decreasing expense and increasing capabilities. Chemists are finding computers not only to be a valuable aid in the measurement and management of laboratory data, but to be helpful in interpreting and analyzing chemical information as well. Recent trends toward the "computerization" of chemical research have affected numerous areas. Modern instrumentation frequently incorporates computer technology, with the result that many measurements can be made more reliably, and more easily, than was previously possible. Advances in areas such as Fourier Transform Spectrometry would not even have been feasible without the availability of dedicated mini-computers at acceptable costs. In a great many chemical situations, computers allow rapid collection and processing of large quantities of data, thus enabling chemists to extract more information than ever before from significant experiments. It seems readily apparent, then, that if chemists are to take maximum advantage of the various benefits offered by the coupling of computers with chemical research, they must first acquire a working familiarity with computers. Our course in Chemometrics represents a significant effort to provide college chemistry majors with a better understanding of potential roles which computers might play in chemical research.

Chemometrics is a general term covering the application of mathematical techniques in chemistry (1). The field was "created" by people who realized that the impact of computers on chemical analysis would extend well beyond that of interfaced calculators. We perceived the need for a graduate level course in chemometrics. Our intention was not only to arouse a general interest in chemometrics but also to provide a foundation for students desired to pursue research topics in this area. Our goal was to survey this expanding and rapidly developing field in a vibrant, challenging fashion. This paper will describe our goals, format, and experiences. This "special topics" course was enthusiastically received by our students, and we feel that the material covered will increasingly be desirable in courses at other institutions.

Topic Selection

The course was formulated to be of a survey nature, with a balance sought between thorough coverage of each topic studied and a comprehensive overview of the field of chemometrics. In order to insure that each topic included in the course syllabus could be treated in reasonable detail, it was necessary to select a manageable list. Our approach in preparing the syllabus was to stress particularly those chemometrics techniques which are most widely used, and for which interesting and useful applications have been reported in the

¹ Author to whom correspondence should be directed. Present address: Department of Chemistry, Boston University, 685 Commonwealth Ave., Boston, MA 02215.

² Present address: Department of Chemistry, Boston University, 685 Commonwealth Ave., Boston, MA 02215.

Table 1. Chemometrics Topics

	Percent of Lectures
Simplex Optimization	20
Data Massaging (<i>S/M</i>)	10
Non-Parametric Statistics	5
Pattern Recognition	30
Library Searching	15
Graph Theory	15
Factor Analysis	5
Artificial Intelligence	
Simulation	
Computerized Organic Synthesis	
Information Theory	

literature. Every effort was made to choose a diverse selection of topics, which would fairly represent the breadth of chemometrics. A pragmatic consideration at this stage of planning was, of course, the instructor's familiarity with the various possible topics.

Since our goal was to excite students about chemometrics, as well as to acquaint them with the field, we deliberately de-emphasized coverage of statistics, even though an understanding of this discipline is critical to the full appreciation of some chemometric techniques. We felt that students would be more likely to pursue an understanding of statistics via other available courses, if we could succeed at interesting them in techniques which rely on statistics. In addition, the relegation of statistics to a minor role in the syllabus freed us to spend more time on other less traditional chemometric topics.

The subjects selected for study, listed in Table 1, span a range bordering directly on the data acquisition phase of an experiment (Simplex Optimization) to areas which are more extrapolatory (Graph Theory). A review of chemometrics publications has recently appeared (2). The breakdown of

The Computer Series attempts to delineate the current state of the art of computer usage in chemical education while accommodating the needs and background of readers who have little or no computer expertise. In addition to full-length articles, short descriptions of specific applications of computers in classrooms or laboratories are published as "Bits and Pieces," so that readers who have appropriate computer systems can obtain and use the programs described. The editor's intention is that the Computer Series be understandable for beginners but at the same time interesting for experts.

John W. Moore received his AB from Franklin and Marshall College and his PhD from Northwestern University, concentrating in physical inorganic chemistry. Following NSF-sponsored postdoctoral work at the University of Copenhagen he has taught at Indiana University and Eastern Michigan University.



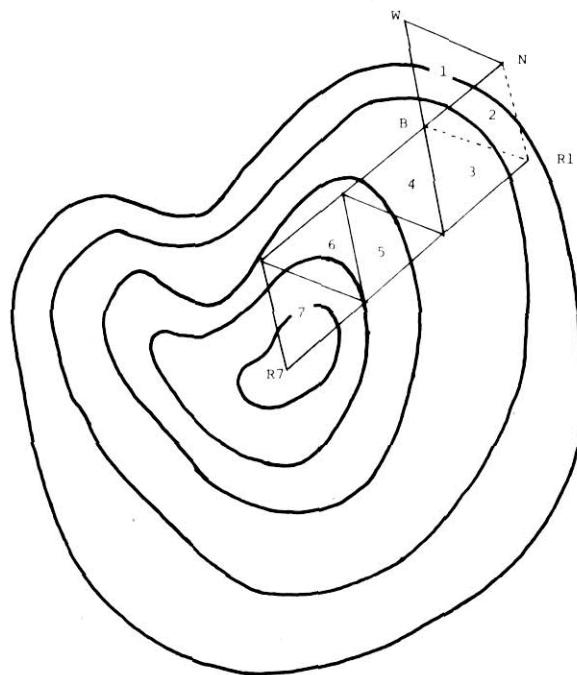


Figure 1. A simplex operating on a contoured response surface.

topic areas presented by Kowalski is seen to differ from those covered in this course. This demonstrates the youth of the field. We were surprised that Kowalski neglected Library Searching and Graph Theory, considering their importance in handling large chemical data bases, while we did not at first fully appreciate the importance of Information Theory in defining the efficiency of analytical methodologies.

Brief Introduction to Selected Topics

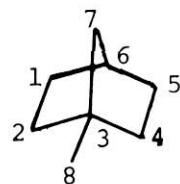
For the benefit of those readers who are unfamiliar with the topics listed in Table 1, a very brief introduction to six of the topics considered in our course is presented below. For each of the six topics, the literature which we found to be most helpful and instructive is cited. The reader is strongly advised to consult those references for the clarification of any of the concepts mentioned.

Simplex Optimization

Simplex optimization (3–5) is a particularly convenient method for determining the optimal numerical values for any number of variables which may influence a dependent response. The method is successful especially when these variables interact, and can be used in both experimental and numerical applications. To optimize N variables, an N -dimensional simplex, having $N + 1$ vertices, is needed. The coordinates of each vertex are a particular set of values which the variables will be made to assume. To illustrate the principles of simplex optimization, a 2-dimensional simplex, as shown in Figure 1, is most helpful, since a simplex of four or more dimensions cannot be readily visualized. The equations used in the simplex process can, however, be extended to any number of dimensions, despite the visualization problem.

The basic simplex process is as follows: (1) Determine the response for each vertex. (2) Label the vertices as B (best), N (next to worst), and W (worst). (3) Reflect the simplex away from W, as indicated by the dotted lines in Figure 1. (4) Check the response at R1, the reflected vertex. (5) Relabel the vertices of the next simplex and continue.

The simplex is finished when no further improvement in the response can be made. A modified simplex, as well as a super modified simplex, has been developed, and each of these offers certain advantages over the basic version (4, 6).



ATOM
NUMBERS 1 2 3 4 5 6 7 8

	1	2	3	4	5	6	7	8
1	0	1	0	0	1	0	0	0
2	1	0	1	0	0	0	0	0
3	0	1	0	1	0	0	1	1
4	0	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	0
6	1	0	0	0	1	0	1	0
7	0	0	1	0	0	1	0	0
8	0	0	1	0	0	0	0	0

Figure 2. A hydrogen suppressed chemical graph and connectivity table for 3-methyl norbornane.

Data Massaging

The improvement of experimental data using computerized techniques (data massaging) can facilitate increases in the signal to noise ratio (S/N) which improves the precision of the data interpretation process. Published studies (7–12) have primarily focused on the use of low-pass digital filters to smooth noisy data. Other methods include the use of ensemble averaging (signal averaging) (13) in which repetitive scans of a spectrum are summed together to yield an increased S/N, and the use of Fourier transforms to estimate parameters for fused peaks (14).

Pattern Recognition

The many forms of pattern recognition (15–19) which have been developed are well-represented by the simple linear learning machine (LLM). Consider the problem of separating a collection of digitized mass spectra into two groups: those which contain oxygen, and those which do not. Each spectrum of N m/e values is treated as a set of coordinates for the vertex of a vector in N -dimensional space. Working on the assumption that the vertices corresponding to a particular class of spectra will tend to cluster together, the LLM attempts to position an N -dimensional linear surface, called a hyperplane, between the vertices representing members of a "training set" of spectra, for which the LLM knows the group membership of each spectrum. The validity of the resulting decision surface is generally tested by presenting the LLM with a "test set" of unknown spectra. Although any given LLM can make only one binary decision, a variety of LLM's may be trained and implemented, so that a series of increasingly specific binary questions may be asked about the members of a data set.

Library Searching

Library searching systems (20–24) are designed for the purpose of retrieving from a large file of digitized spectra those members which are most similar to an unknown digitized spectrum. Generally, the best five or ten matches from the library are reported, as an aid to identification of the unknown

compound. The least efficient method of library searching, both from the standpoint of speed and of storage requirements, is a point-by-point comparison of the complete spectra, and a great many alternatives to this style of search have been proposed. Condensation of the spectra is one approach to simplifying both the storage and comparison requirements. The intensity associated with each spectral element may be reduced from full resolution all the way down to a binary (peak/no peak) representation. The number of elements per spectrum may likewise be reduced by a variety of methods. Of further concern in designing library searching systems is the feasibility of "pre-filters" (simple tests which narrow the portion of the library to be searched.) Finally, a variety of techniques for the evaluation of library searching performance have been suggested.

Graph Theory

Techniques for representing chemical structures (graphs) in computer-compatible form are sorely needed, if computers are to play a truly central role in interpreting chemical information. Chemists have drawn upon various aspects of graph theory in their efforts to tackle this key problem. Randic has been particularly active in this area (25-30), and we chose to discuss a number of his ideas in class.

Molecular skeletons can be represented in topological matrix (connectivity matrix) form, as shown in Figure 2. For an N -membered molecule, an N -by- N matrix is generally constructed, in which each "1" signifies a pair of connected vertices. For example, a "1" in row 1, column 2, of a topological matrix indicates that atoms 1 and 2 are bonded. Unfortunately, the same molecular skeleton can be numbered in a variety of ways, and therefore an assortment of matrices arises. Several possible solutions to this problem of uniqueness were discussed in class. Applications of topological matrices to artificial intelligence and library searching were also considered.

The notion of a connectivity index for a given molecular skeleton has been proposed by Randic. Here, each bond of a framework is given a rating, according to the "degree of branchedness" of each of the terminal atoms. For the example shown in Figure 2, atom 3 would be rated "4", since this atom is connected to four non-hydrogenic atoms. Atom 4, by similar reasoning, would be rated "2". Then, the contributions from each bond in the molecule are summed up, to give the overall connectivity index. In the case of Figure 2, the sum of the square roots of the reciprocal ratings yields a connectivity index of 3.7857. By this procedure, an entire chemical structure can be represented by a single number, which is related to the branchedness of the molecule.

Atomic and molecular path codes are a third technique for reducing chemical structures to a numerical representation. To generate an atomic path code, the following values are listed in sequence: the number of paths (bonds) of length one, the number of paths of length two, and so on. For the case of the terminal atom of *n*-butane, the atomic path code would be 1,1,1,1. To generate a molecular path code, all atomic path codes for the molecule are summed, and the result divided by two (since each path will have been counted twice.) Molecular path codes have been considered for sub-structure searching and the unique atom-numbering of chemical compounds (28-30).

Factor Analysis

Factor analysis is an established statistical method for extracting a limited number of constructs, which explain the intercorrelations among the variables of a data set. A mathematical and theoretical treatment of the technique is available in several texts (31-34). Chemists have found several forms of factor analysis to be useful for solving chemical problems. Principal component analysis has been used for the deconvolution of overlapped chromatographic peaks (35), for the analysis of the mass spectra of mixtures (36), and for studying

retention in gas chromatography (37). Target transformation factor analysis has been applied to the qualitative and quantitative determination of the components of mixtures, via mass spectral analysis (38).

Texts

Selection of a suitable text for the course was seriously hampered by the complete lack of a single volume covering any significant fraction of the desired material. While such a text will probably appear in the future, we attempted to meet the need for students to have core reading material by using two texts, liberally supplemented by original journal articles. The first book utilized was "Chemometrics: Theory and Practice" (39), edited by Bruce Kowalski, which is a collection of papers from an A.C.S. Symposium. Unfortunately, the symposium format was not particularly beneficial to students due to the tendency of the papers to briefly cover a specialized topic using undefined technical terms. However, with suitable introductory lecture material, the students could appreciate these chemometrics successes from a wide array of chemical research endeavors. The second text selected was "Chemical Applications of Pattern Recognition" (15) by Jurs and Isenhour. This source does provide a solid presentation of the key elements of pattern recognition by describing the initial efforts of chemometrists working with this technique, but little is added to the authors' original publications from which the book was prepared. Several other texts were found useful for the preparation of lecture material (31, 40, 41). Extensive use of the original literature was made, both for lecture preparation and for reading assignments. In addition to the general analytical chemistry journals, chemometrics articles were found in the *Journal of Chemical Information and Computer Science, Computers and Chemistry*, and the applied analytical journals (chromatography, spectroscopy, etc.).

Requirements

In addition to attending lecture, students were required to participate in four types of activities: homework assignments, examinations, an independent project, and a term paper.

Examinations

Two exams were given during the course to probe the student's knowledge of chemometrics. The questions mostly involved essay or simple calculation answers which primarily covered the basic factual material from the lectures. Other than letting the students demonstrate their comprehension of the material, it was difficult to develop questions which induced the students to extrapolate beyond the basic subject matter. One aspect of the exams which could be better developed would be to pose questions based on published applications of chemometrics. These could include many instrumental techniques, (e.g., chromatography, electrochemistry, or spectroscopy) or subject areas such as environmental, clinical, or pharmaceutical analysis.

Assignments

Our students showed considerable variation in their familiarity with computers, and with chemistry, since they ranged from undergraduates majoring in chemistry or computer science to graduate students in several divisions of chemistry. To meet the challenge of teaching students of such varied ability, a set of interactive, computer-based assignments was developed to supplement the lectures and reading. Students were able to involve themselves with a variety of chemometric techniques by modifying and using a library of computer programs compiled by the instructor. Students with greater ability and interest tended to go beyond the minimum requirements of the assignment, but all students gained valuable hands-on experience in using computers for chemical purposes.

Although use of the computer played an integral role in the learning process, the requirement for student programming

Table 2. Chemometrics Assignments

Familiarization with Computer, Fortran Program
Bomber Optimization
Simplex Optimization
Graphics Terminal Usage
Signal-to-Noise Enhancement
Chemical Compound Data Set Preparation
Hyperplane Separation
Weight Sign Feature Selection with Hyperplane Separation
Using Arthur for Pattern Recognition
Examination of Molecular Path Codes
Independent Project
Term Paper

was deliberately kept to a minimum. This focused attention on the subject matter rather than on the headaches of writing programs. While we did require a familiarity with FORTRAN (or a similar language) as a prerequisite to the course, this was done primarily to insure that students would not associate an aura of mystery with the computer. Our main interest was to train chemists to use computers rather than to attempt any detailed consideration of how computers work, or of how to write programs. A computer, like an analytical balance, is a tool which many chemists will find useful. Just as few chemists worry about whether their analytical balance works by adding weights to, or removing weights from, the balance beam, we feel that for practical purposes, a chemist need not concern himself with the question of where the bits go when an array is zeroed.

The assignments (Table 2) spanned most of the topics encountered in the lectures. They were structured around computer programs written or modified by the instructor. The students executed the programs on the university's DEC-system-10 time-sharing system, entering parameters or data when needed, and interpreting and summarizing the results. On some occasions students needed to make minor modifications of the programs, usually of a data-formatting nature, to tailor the execution to their individual circumstances.

The first assignment helped the students develop a familiarity with the computer system by trying out various system commands and operations and by writing, running, and debugging a simple FORTRAN program. The next assignment considered various optimization procedures. The students executed the game BOMBER, in which the player needs to find a bomb in a million room building. The user has a "bomb distance meter" and must find the bomb before it explodes. The student is asked to play with the program and to develop a strategy to find the bomb in an organized way. The lecture material considered the development of algorithms and heuristics. The next assignment exposed the student to Simplex Optimization. The student was asked to execute the program SIMPLE, to locate the maximum response to a stored function provided with the program, and then to a function chosen by the student.

The next assignment showed the students how to use and program a Tektronix graphics terminal for plotting data. Students were asked to execute two programs which plotted two arrays (a random distribution of two dimensional points, and a Gaussian peak) and were then asked to modify the program to display a graph of their own design. This plotting capability is useful in demonstrating the operation of signal-to-noise enhancement filters. The user could vary parameters and visually observe the effect on the experimental data.

To facilitate the next several assignments the students were asked to go to the literature and collect five physical constants: molecular weight, melting point, boiling point, density, and refractive index, for two sets of twenty-five chemical compounds. Each set was to contain members of a single functional group class. With this data for a total of 50 compounds the next assignments investigated the use of pattern recognition with chemical data. The hyperplane separation tech-

Table 3. Chemometrics Independent Projects

Improved S/N Filter
Examination of Clustering Algorithms
Pattern Recognition of Molecular Path Codes
Cross Terms in Pattern Recognition of VPIR Spectra
Propagation Method for Library Searching
Combined Chemical Compound Data
Normalized Thresholds for Hyperplane Separation
S/N Filter Graphic Demonstration
Pattern Recognition Examination of McReynold's Constants

nique was used to try to separate the compounds into the two functional group classes. Forty compounds were randomly assigned to the training set, and the ten remaining compounds constituted the test set. The students then used the "weight sign feature selection" approach, along with hyperplane separation, to eliminate features which were not useful in separating the compounds into classes. The students were asked to discuss the results in terms of chemical principles. In addition, the students executed these programs with a synthetic set of data which exhibits subtle behavior which might not be seen with the chemical compound data, and to create a synthetic data set of their own, for use with the pattern recognition programs.

The next assignment used the ARTHUR pattern recognition system (39, 42) developed by Kowolski and co-workers. Many of the routines contained in ARTHUR were discussed in lecture, and files to format data and execute the various routines were provided for the students either to use directly or to modify to meet their needs. The ARTHUR routines included normalization, weighting, hyperplane separation, *K*th nearest neighbor classification, Bayes' classification, feature projection, and clustering algorithms.

The last class assignment used the ALPATH routine, developed by Randic and co-workers (29) which computes a molecular path code from the connection table of a chemical compound graph. The students were asked to use the routine to compute path codes for a variety of molecules.

Independent Projects

One of the most stimulating parts of the course was the independent projects. Unlike laboratory-based student projects, which can cause the destruction of facilities, the risking of human lives, and the exasperation of a teaching assistant beyond the point of no-return, a computer-based independent project can be relatively risk free, while still maintaining the possibility for creative chemistry. In this course, students were given a free hand to select a project of interest to them, from any area covered in the lectures. This allowed them to explore areas more carefully, especially in fields of their particular expertise. Discussions were held with the instructor to map out a strategy for the project and to assess feasibility. A short paper describing the project's background and goals, and summarizing the results, was expected upon completion of the project.

A list of typical projects is given in Table 3. One student was intrigued with a publication mentioned in class which described a method for an improved low pass filter. The article (43) claimed that one could filter an experimental signal using a time constant which significantly distorts the data, and then reconstruct the original waveform with an improved signal-to-noise ratio using a mathematical reconstruction procedure. Another student desired a better understanding of the parameters used in the clustering algorithms provided in ARTHUR. The project consisted of trying the clustering routines on the obsidian artifact elemental analysis data (42) provided with the ARTHUR system. A quite interesting project consisted of applying pattern recognition to classify compounds based on their molecular path code. The student typed in a binary connection table for a set of 50 compounds. Then the

ALPATH program was used to calculate the molecular path codes, which were formatted for input to ARTHUR. The classification problem was to decide if the compounds were cyclic or acyclic. Although the pattern recognition classification was quite successful it was later realized that this problem has an analytical solution based on the first two elements of the molecular path code (the number of atoms, and the number of bonds, respectively).

Another project consisted of using the propagation method, proposed by Isenhour and co-workers (24), to study library searching performance for a collection of vapor phase infrared spectra (22,23). In the propagation technique, a seed compound is searched through the library to yield a list of closest matches. Each new compound on the best match list is subsequently searched against the library to find its best matches. The effectiveness of the library searching procedure can be assessed by how rapidly this list of closest matches expands and by the chemical similarity of the best matches.

One student studied the use of cross terms for the improvement of pattern recognition interpretation of vapor phase infrared spectra (44, 45). Cross terms are composed of a combination of the intensity values at a pair of wavelengths. For example, for identifying an ester functionality it would be useful to have a cross term made up of the carbonyl stretching wavelength and the carbon-oxygen single bond stretching wavelength, since an ester will have both types of absorptions. The project consisted of comparing several approaches for automatically selecting useful cross terms and evaluating their effect on the performance of the pattern recognition interpretation.

One student pooled together the class's chemical compound data sets, described above, and attempted a classification using ARTHUR. Another student used ARTHUR to study the McReynold's Constants for gas chromatographic stationary phases (46). One project was concerned with the effect of using normalized weight vectors for hyperplane separations using non-zero thresholds. It is observed (47) that when the weight vectors are not normalized, the use of increasingly larger thresholds results in larger values of the weight vector components without actually increasing the width of the dead zone.

An exceedingly useful project for subsequent semesters comprised the development of a graphical demonstration program for signal-to-noise enhancement filters (11). The student took the filter subroutines used in the course and combined them into a program package which allows a user to vary the filter and signal waveform parameters, and then to visually observe their effect on the data processing. The program generates a gaussian peak with noise based on the selected values. The user chooses a type of filter and a suitable set of parameters. The program then displays simultaneously the data array before and after the filtering process, and finally calculates the signal-to-noise enhancement, peak maximum shift and peak half-width increase.

Term Papers

The students were asked to write a paper on a topic of their choice to induce them to pursue an area of chemometrics in more detail. Being able to lucidly write about a subject demonstrates a competence and familiarity with the material. Most of the students chose to examine applications of the techniques discussed in the lectures to a chemical problem. For example, papers were written on protein secondary structure prediction, drug design, and oil spill identification. Other topics for papers included a discussion of a survey of student's feelings about chemometrics, and one student wrote the user's manual for the Signal-to-Noise Filter Graphical Demonstration program, which he prepared for his independent project.

Discussion

On the whole the course proceeded remarkably well. Although bugs in the programs were constantly being noticed,

these were generally of a trivial nature, since for the most part the programs had been used by others, for research or teaching purposes. The project-programmer structure of the DEC-system-10 operating system facilitated the sharing of program and data resources by the participants, while maintaining the security and integrity of privileged information.

Student response to the course was quite encouraging, as revealed by the responses to a questionnaire administered by one of the students. In addition to finding the study of chemometrics enjoyable, our students uniformly felt that they would benefit from their knowledge of the field, and that they would be likely to use chemometric techniques in their future research, should the opportunity arise. They strongly agreed with the suggestion that more chemists, and especially analytical chemists, should learn more about chemometrics.

Many of the comments and responses we received were helpful in pointing out the strengths and weaknesses of the present course, and in indicating some ideas for improvements. Of the topics covered, students especially appreciated pattern recognition, library searching and graph theory, but had little affection for non-parametric statistics or factor analysis. There was general agreement that ours should not be a course in statistics, but rather a computer applications course for chemists. Five of the topics covered in the course received a very high approval rating for inclusion in a future chemometrics course: simplex optimization, data massaging, pattern recognition, ARTHUR, and library searching. Graph theory received mixed reviews, while factor analysis and non-parametric statistics found very little support.

Student opinion about the assignments was uniformly supportive. The opportunity to implement a wide variety of prepared computer programs was seen as an effective and stimulating vehicle for learning chemometrics. Students found the computer assignments to be interesting and challenging, and most said they had felt sufficiently intrigued to go beyond the minimum requirements for at least one assignment. Our decision to minimize required student programming was applauded, since this did help to focus attention on the subject matter. The independent projects seemed to allow more than ample opportunity for elective programming. Class sentiment was strongly positive toward the projects and papers, which indicates the degree to which the course stimulated an independent interest in the various chemometric techniques.

Most students recommended the original journal articles as a good source of information about the topics covered in the course. Students claimed, on the average, to have read over half of the 25 journal articles referenced in lecture. They found the pattern recognition text, by Jurs and Isenhour (15), to be clear and readable, and recommended (with a few reservations) that this book be used for future chemometrics courses. The text assembled by Kowalski (39) was not widely appreciated as an introductory text, however. Nonetheless, students did agree that there was sufficient reading material available as a complement to the lectures and assignments.

Future improvements for the assignments will include the incorporation of larger chemical data bases as they become available, and clarification of the input/output sections of the instructional programs to be more easily understandable by new users. A continuing critical evaluation of the topics covered, and of the depth to which they are explored seems important. When the course is given again, we will almost certainly include a discussion of simulation as it applies to areas such as electrochemistry and chromatography. Computerized organic synthesis, and information theory will also find a larger role in the syllabus.

Analytical chemistry is increasingly an information science, and recent publications (48-51) have shown how analytical procedures can be described, optimized, and understood in terms of information theoretical concepts. We feel that the teaching of chemometric principles will be increasingly desirable in contemporary chemical curriculum.

Acknowledgements

The authors are grateful for the participation and cooperation of the students involved in developing this course. This work was partially supported by a grant from the A.C.S. Petroleum Research Fund, and was presented, in part, in the Computers in Chemical Education Symposium at the 10th Northeast Regional A.C.S. Meeting, 31 June 1980, Potsdam, N.Y.

Literature Cited

- (1) Kowalski, B. R., *Anal. Lett.*, **A11**, xi (1978).
- (2) Kowalski, B. R., *Anal. Chem.*, **52**, 112R (1980).
- (3) Deming, S. N. and Morgan, S. L., *Anal. Chem.*, **45**, 278A (1973).
- (4) Deming, S. N., and Morgan, S. L., *Anal. Chem.*, **46**, 1170 (1970).
- (5) Shavers, C. L., Parsons, M. L., and Deming, S. N., *J. CHEM. EDUC.*, **56**, 307 (1979).
- (6) Routh, M. W., Swartz, P. A., Denton, M. B., *Anal. Chem.*, **49**, 1422 (1977).
- (7) Hietjte, G. M., *Anal. Chem.*, **44**(6), 81A (1972).
- (8) Hietjte, G. M., *Anal. Chem.*, **44**(7), 69A (1972).
- (9) Savitsky, A., and Golay, M. J. E., *Anal. Chem.*, **36**, 1627 (1964).
- (10) Steiner, J., Termonia, Y., and Deltour, J., *Anal. Chem.*, **44**, 1906 (1972).
- (11) Delaney, M. F., and Uden, P. C., *Anal. Lett.*, **A12**, 673 (1979).
- (12) Ackroyd, M. H., 'Digital Filters', Butterworths, London, 1973.
- (13) Ernst, R. R., *Rev. Sci. Instrum.*, **36**, 1689 (1965).
- (14) Binkley, D. P., and Dessy, R. E., *Anal. Chem.*, **52**, 1335 (1980).
- (15) Jurs, P. C., and Isenhour, T. L., "Chemical Applications of Pattern Recognition," Wiley, New York, 1975.
- (16) Jurs, P. C., Kowalski, B. R., Isenhour, T. L., and Reilley, C. N., *Anal. Chem.*, **41**, 690 (1969).
- (17) Jurs, P. C., Kowalski, B. R., Isenhour, T. L., and Reilley, C. N., *Anal. Chem.*, **41**, 1949 (1969).
- (18) Kowalski, B. R., Jurs, P. C., Isenhour, T. L., and Reilley, C. N., *Anal. Chem.*, **41**, 1945 (1969).
- (19) Sjostrom, M., and Kowalski, B. R., *Anal. Chim. Acta*, **112**, 11 (1979).
- (20) Grotch, S., *Anal. Chem.*, **46**, 526 (1974).
- (21) Grotch, S., *Anal. Chem.*, **47**, 1285 (1975).
- (22) Delaney, M. F., and Uden, P. C., *Anal. Chem.*, **51**, 1242 (1979).
- (23) Delaney, M. F., and Uden, P. C., *J. Chrom. Sci.*, **17**, 428 (1979).
- (24) Rasmussen, G. T., Isenhour, T. L., *J. Chem. Info. Comput. Sci.*, **19**, 179 (1979).
- (25) Randic, M., *J. Chem. Phys.*, **60**, 3920 (1974).
- (26) Randic, M., *Intl. J. Quant. Chem.*, **5**, 245 (1978).
- (27) Randic, M., *J. Chrom.*, **161**, 1 (1978).
- (28) Randic, M., *J. Chem. Info. Comput. Sci.*, **19**, 23 (1979).
- (29) Randic, M., Brissey, G. M., Spencer, R. B., and Wilkins, C. L., *Comp. Chem.*, **3**, 5 (1979).
- (30) Randic, M. and Wilkins, C. L., *J. Chem. Info. Comput. Sci.*, **19**, 23 (1979).
- (31) Comrey, A. L., "A First Course in Factor Analysis," Academic Press, New York, 1973.
- (32) Cattell, R. B., "Factor Analysis," Harper, New York, 1952.
- (33) Lawley, D. N., Maxwell, A. E., "Factor Analysis as a Statistical Method," Butterworth, London, 1963.
- (34) Harman, H. H., "Modern Factor Analysis," Univ. of Chicago Press, Chicago, 1976.
- (35) Macnaughton, D., Jr., Rogers, L. B., and Wernimont, G., *Anal. Chem.*, **44**, 1421 (1972).
- (36) Ritter, G. L., Lowry, S. R., Isenhour, T. L., and Wilkins, C. L., *Anal. Chem.*, **48**, 591 (1976).
- (37) Weiner, P. H., and Howry, D. G., *Anal. Chem.*, **44**, 1189 (1972).
- (38) Malinowski, E. R., and McCue, M., *Anal. Chem.*, **49**, 284 (1977).
- (39) Kowalski, B. R., (Editor) "Chemometrics: Theory and Practice," A.C.S. Symp. Series, No. 52, 1977.
- (40) Lynch, M. F., Harris, J. M., Town, W. E., and Ash, J. E., "Computer Handling of Structure Information," Amer. Elsevier, New York, 1971.
- (41) Hartigan, J. A., "Clustering Algorithms," Wiley, New York, 1975.
- (42) ARTHUR, Version 1.9-77, avail. from A. M. Harper, Dept. of Chem., Univ. of Georgia, Athens, GA 30602.
- (43) Mohos, B., Zobrist, M., Von Zelewsky, A., and Galambos, G., *Anal. Chem.*, **48**, 231 (1976).
- (44) Kowalski, B. R., Jurs, P. C., Isenhour, T. L., and Reilley, C. N., *Anal. Chem.*, **41**, 695 (1969).
- (45) Jurs, P. C., *Appl. Spec.*, **25**, 483 (1971).
- (46) McReynolds, W. O., *J. Chrom. Sci.*, **8**, 685 (1970).
- (47) Wangen, L. E., Frew, N. M., and Isenhour, T. L., *Anal. Chem.*, **43**, 845 (1971).
- (48) Eckschlager, K., and Stepanek, V., "Information Theory as Applied to Chemical Analysis," Wiley, New York, 1979.
- (49) Edwards, E., "Information Transmission," Chapman and Hall, London, 1969.
- (50) Ritter, G. L., Lowry, S. R., Woodruff, H. B., and Isenhour, T. L., *Anal. Chem.*, **48**, 1027 (1976).
- (51) Eckschlager, K., *Anal. Chem.*, **49**, 1265 (1977).