

---

# Charles E. Davidson, Ph.D.

1672 N. Longfellow St.  
Arlington, VA 22205

cell: 315-569-2508

home: 703-533-7546

charles.davidson@att.net

<http://home.att.net/~charles.davidson/>

---

**OBJECTIVE:** To obtain a position utilizing my skills in multivariate data analysis and data mining, not necessarily limited to the chemometric, chemoinformatic, or bioinformatic fields. Particularly interested in algorithm development and software prototyping.

**SKILLS:** Analytical chemist knowledgeable with chemometric/statistical software such as Pirouette, SCAN, and Minitab, but I prefer writing my own programs in MATLAB. Knowledgeable with molecular simulation packages Spartan, QUANTA, and MOE. Experienced in Windows and Linux operating systems, HTML, and  $\text{\LaTeX}$ ; some experience with SVL, C<sup>++</sup>, Perl, and Javascript.

---

## Experience

RESEARCH ASSISTANT, 2000–2003

CLARKSON UNIVERSITY, POTSDAM, NY

Developed a software package for multivariate data analysis based on projection methods (e.g., principal component analysis, partial least squares, Kohonen neural networks) and genetic algorithms in MATLAB. The idea underlying this system is that pattern recognition methods will work well when the problem is simple. By identifying the appropriate features, a difficult problem can be reduced to a simple one. This research involved the following topics:

*Supervised Learning Approach to Pattern Recognition:* Principal components analysis (PCA) reduces the dimensionality of multivariate data by finding new axes that track independent sources of variation. Integrating PCA with a genetic algorithm (GA) allowed for the identification of features whose information is primarily about class differences. Other projection methods that take advantage of specific data structure were substituted for PCA, such as partial least squares (PLS), and Kohonen self-organizing maps (SOMs). The SOM enables successful handling of outliers and nonlinear relationships within the data.

Spectroscopic and chromatographic chemical data, as well as biological and physical data sets were successfully analyzed. Examples from spectroscopy include the classification of hard, soft, and tropical woods, classification of recyclable plastics, and quality control of pharmaceutical tablets by near IR spectroscopy. Problems in chemical communication and fuel spill identification showed the applicability of this methodology to chromatographic data. Examples of biological data include DNA microarray data sets.

*Unsupervised Learning Approach:* A GA to select features that maximize clustering of the data was designed. This tool serves as a data microscope, elucidating interesting structure and relationships within the data, and helping to identify confounding variables. This methodology aids the analysis of SAR data, and enables the analysis of exceedingly large and complex data sets from molecular dynamics (MD) simulations without relying upon previous knowledge. This method can identify atoms involved in substrate binding, and has been validated using sperm whale myoglobin simulation data.

*Transverse Learning Approach:* A combination of supervised and unsupervised learning approaches improved classification and prediction of multivariate data sets, especially those that contain missing or uncertain class labels. This method searches for features that contain information about class differences and simultaneously maximize clustering in the projection space, allowing the information present in unlabeled data to be used to guide classification, and prevent overfitting. Transverse learning also allows for the identification of sub-classes in the data.

---

Charles E. Davidson, Ph.D.  
1672 N. Longfellow St.  
Arlington, VA 22205

charles.davidson@att.net  
cell: 315-569-2508  
home: 703-533-7546

---

*(Transverse Learning Continued)* The transverse learning GA aids the analysis of DNA Microarray data. It also serves an important role in a larger methodology for structure-activity relationship (SAR) analysis. Past SAR analyses suffer from fragment-based descriptors that are satisfactory for homologous molecules, but which are less effective when applied to data sets with a great deal of structural variation. Difficulties in modeling the underlying chemical process, which can be quite complex, also contribute to the mixed success of past SAR efforts. This new methodology, involving an enhanced version of Breneman's Transferable Atom Equivalent (TAE) descriptors that contain pertinent shape and electronic information, uses the transverse learning GA to help determine the underlying chemical processes involved. This methodology has been applied successfully in musk odor SAR studies.

*Ordinate Pattern Recognition:* Some pattern recognition problems involve the differentiation of classes that are related to each other, e.g., good, better, best. A fitness function for the pattern recognition GA was developed that captures information about the existence of these relationships in the data. This ordinate fitness function serves as a bridge between pattern recognition and calibration.

*Multivariate Curve Resolution:* Research independent of the GA originally designed for deconvolution of co-eluting species in chromatography was adapted to image analysis. A Varimax extended rotation (VER) followed by alternating least squares (ALS) estimates the concentration and spectral profiles of each component. The efficacy of this approach was demonstrated through the analysis of Raman image data of water-in-oil emulsions.

TEACHING ASSISTANT, 2000–2003 CLARKSON UNIVERSITY, POTSDAM, NY  
Taught general chemistry, spectroscopy, and instrumental laboratories. Received the Outstanding Teaching Award for Graduate Students for the 2002-2003 school year. Managed the revision and updating of the instrumental laboratory manual.

## Education

2000–2003 CLARKSON UNIVERSITY, POTSDAM, NY  
Ph.D., Analytical Chemistry, with minor in Informatics.  
Dissertation: "Genetic Algorithms for Data Mining and Multivariate Data Analysis."  
Advisor: Barry K. Lavine. GPA: 3.9/4.0

1996–1999 CLARKSON UNIVERSITY, POTSDAM, NY  
B.S., Chemistry, with Distinction, ACS Accredited. GPA: 3.6/4.0, 3.7/4.0 in major. Received the George L. Jones, Jr., Award for Excellence in Chemistry, and the CRC Science Achievement Award. Three-time Presidential Scholar.

## Activities

Exchange Student at Monash University, Melbourne, VIC, Australia  
Member of the Clarkson University Hearing Committee on Discipline & Disorders  
Graduate Advisor and Webmaster for the Racquetball Club  
Attended the Intercollegiate Racquetball National Tournaments, 2000–2003

---

Charles E. Davidson, Ph.D.  
1672 N. Longfellow St.  
Arlington, VA 22205

charles.davidson@att.net  
cell: 315-569-2508  
home: 703-533-7546

---

## Publications

1. B. K. Lavine, D. Brzozowski, A. J. Moores, C. E. Davidson, and H.T. Mayfield, "Genetic Algorithm for Fuel Spill Identification," **Anal. Chim. Acta**, 2001, 437(2), 233–246.
2. B. K. Lavine, C. E. Davidson, A. J. Moores, and P. R. Griffiths, "Raman Spectroscopy and Genetic Algorithms for the Classification of Wood Types," **Appl. Spectrosc.**, 2001, 55(8), 960–966.
3. B. K. Lavine, C. E. Davidson, and A. J. Moores, "Innovative Genetic Algorithms for Chemoinformatics," **Chemom. Intell. Lab. Syst.**, 2002, 60(1), 161–171.
4. B. K. Lavine, C. E. Davidson, and A. J. Moores, "Genetic Algorithms for Spectral Pattern Recognition," **Vib. Spectrosc.**, 2002, 28(1), 83–95.
5. B. K. Lavine, C. E. Davidson, Robert K. Vander Meer, S. Lahav, V. Soroker, and A. Hefetz, "Genetic Algorithms for Deciphering the Complex Chemosensory Code of Social Insects," **Chemom. Intell. Lab. Syst.**, 2003, 66(1), 51–62.
6. B. K. Lavine, C. E. Davidson, C. Breneman, and W. Katt, "Electronic Van der Waals Surface Property Descriptors and Genetic Algorithms for Developing Structure-Activity Correlations in Olfactory Databases," **J. Chem. Inf. Comput. Sci.**, 2003, 43, 1890–1905.

— *Accepted for Publication* —

7. B. K. Lavine, C. E. Davidson, C. Breneman, and W. Katt, "Genetic Algorithms for Clustering and Classification of Olfactory Stimulants," in **Chemoinformatics: Methods and Protocols**, J. Bajorath (Ed.), Humana Press, IN PRESS.
8. B. K. Lavine and C. E. Davidson, "Classification and Pattern Recognition," in **Practical Handbook of Chemometrics, 2nd Edition**, Paul Gemperline (Ed.), Marcel Dekker Press, IN PRESS.
9. B. K. Lavine, C. E. Davidson, J. P. Ritter, D. Westover, and T. Hancewicz, "Varimax Extended Rotation Applied to Multivariate Spectroscopic Image Analysis," **Microchem. J.**, IN PRESS.

— *Submitted for Publication* —

10. B. K. Lavine, C. E. Davidson, S. Hawkins, and T. M. Hancewicz, "Denoising of Ballistometer Data Using Fourier Filtering and Pattern Recognition Techniques," **International Journal of Cosmetic Science**, submitted.
11. B. K. Lavine, C. E. Davidson, C. Breneman, and W. Katt, "Development of Structure–Activity Olfactory Correlations using Electronic Van der Waals Surface Property Descriptors and Genetic Algorithms," in **Chemometrics and Chemoinformatics**, B. K. Lavine (Ed.), **ACS Symposium Series**, submitted.
12. B. K. Lavine, J. Workman, and C. E. Davidson, "Chemometrics: Past, Present, and Future," in **Chemometrics and Chemoinformatics**, B. K. Lavine (Ed.), **ACS Symposium Series**, submitted.
13. B. K. Lavine, C. E. Davidson, and W. T. Rayens, "Machine Learning Based Pattern Recognition Applied to Microarray Data," **Combinatorial Chemistry & High Throughput Screening**, submitted.

— *In Preparation* —

14. B. K. Lavine and C. E. Davidson, "Learning from Expression Data," **Bioinformatics**, in preparation.
15. B. K. Lavine, C. E. Davidson, and W. T. Rayens, "Genetic Algorithms for Data Mining—Profiting from the Past," **J. of Chemom.**, in preparation.
16. B. K. Lavine and C. E. Davidson, "Genetic Algorithms That Emulate Human Pattern Recognition Through Machine Learning for Database Mining and Knowledge Discovery," **J. Chem. Inf. Comput. Sci.**, in preparation.
17. B. K. Lavine, C. E. Davidson, R. K. Vander Meer, D. Carlson, S. Lahav, V. Soroker, and A. Hefetz, "Gas Chromatography/Pattern Recognition Techniques Applied to Taxonomy and Chemical Communication," **Microchem. J.**, in preparation.

---

Charles E. Davidson, Ph.D.  
1672 N. Longfellow St.  
Arlington, VA 22205

charles.davidson@att.net  
cell: 315-569-2508  
home: 703-533-7546

---

## Presentations

— *Author and/or Presenter* —

1. “Innovative Genetic Algorithms for Pattern Recognition of Chemical Data,” at the 220th National Meeting of the American Chemical Society, Washington, DC, 2000, and at the Clarkson University Department of Chemistry Seminar, 2000.
2. “Genetic Algorithms,” at Clarkson University, part of the Freshman Seminar Series, 2001.
3. “Gene Expression and DNA Microarray Technology,” Clarkson University Department of Chemistry Seminar, 2003.
4. “Mining Microarray Data,” Clarkson University Department of Chemistry Seminar, 2003.
5. “Genetic Algorithms for Data Mining and Multivariate Data Analysis,” Dissertation Defense, Clarkson University, Dec. 2003.
6. “Genetic Algorithms for Pattern Recognition and Multivariate Calibration,” Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) Conference, Detroit, MI, October 11, 2001.
7. “New Methods in Multivariate Spectrometric Image Analysis I,” FACSS, Detroit, MI, October 12, 2001.
8. “Genetic Algorithms for Database Mining,” Center for Process Analytical Chemistry (CPAC) Meeting, Seattle, WA, November 6, 2001.
9. “Genetic Algorithms for Developing Structure–Activity Correlations in Large Olfactory Databases,” National ACS Meeting, Boston, MA, August 22, 2002.
10. “Varimax Extended Rotation (VER) Applied to Multivariate Spectroscopic Image Analysis,” Chemometrics in Analytical Chemistry (CAC) Conference, Seattle, WA, September 23, 2002.
11. “Genetic Algorithms for Pattern Recognition and Multivariate Calibration,” CAC Conference, Seattle, WA, September 24, 2002.
12. “Supervised Learning from Gene Expression Data,” Advanced Topics in Microarray Analysis, NIH Workshop, January 22, 2003 (Poster Presentation).
13. “Genetic Algorithms that Emulate Human Pattern Recognition Through Machine Learning,” National ACS Meeting, New Orleans, LA, March 26, 2003.

---

Charles E. Davidson, Ph.D.  
1672 N. Longfellow St.  
Arlington, VA 22205

---

charles.davidson@att.net  
cell: 315-569-2508  
home: 703-533-7546

---

## References

. . . . .

*References are available on file, upon request. Please contact the Clarkson University Career Center:*

---

Career Center	career@clarkson.edu
8 Clarkson Avenue	day: 315-268-6477
Potsdam, NY 13699-5620	fax: 315-268-7616

---

*For more details, please contact the following people, directly. They are familiar with my work and character.*

Associate Professor Dr. Barry K. Lavine (Thesis advisor)  
Analytical Chemistry, Clarkson University

---

Prof. Barry K. Lavine	bklab@clarkson.edu
8 Clarkson Avenue	day: 315-268-2394
Potsdam, NY 13699-5610	fax: 315-268-6610

---

Research Associate Professor Dr. Dan V. Goia  
Center for Advanced Materials Processing, Clarkson University

---

Prof. Dan V. Goia	goiadanv@clarkson.edu
8 Clarkson Avenue	day: 315-268-4411
Potsdam, NY 13699-5814	fax: 315-268-6567

---