DIMACS Center
Rutgers University

**Workshop on Data Depth: Robust Multivariate Analysis,
Computational Geometry and Applications**

Final Report

December 2003

Ia. Participants from the program

Participants:

PI: Fred Roberts

Workshop Organizers:
      Regina Liu, Rutgers University
      Robert Serfling, University of Texas at Dallas
      Diane Souvaine, Tufts University
      Yehuda Vardi, Rutgers University

Ib. Participating Organizations

Telcordia Technologies: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

AT&T Labs - Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshop.

NEC Laboratories America: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Lucent Technologies, Bell Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Princeton University: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Avaya Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

HP Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

IBM Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Microsoft Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

National Security Agency: Collaborative Research
Individuals from the organization participated in the program planning.

1c. Other Collaborators

The project involved scientists from numerous institutions in numerous counties. The resulting collaborations also involved individuals from many institutions in many countries.

II. Project Activities

Multivariate statistical methodology plays a role of ever increasing importance in real life applications, which typically entail a host of interrelated variables. Simple extensions of univariate statistics to the multivariate setting do not properly capture the higher-dimensional features of multivariate data, nor do they yield geometric solutions because of the absence of a natural order for multidimensional Euclidean space. A more promising approach is the one based on "data depth", which can provide a center-outward ordering of points in Euclidean space of any dimension. Extensive developments in recent years have generated many attractive depth-based tools for multivariate data analysis, with a wide range of applications. The diversity in approaches, emphases, and concepts, however, makes it necessary to seek unified views and perspectives that would guide the further development of the depth-based approach.

The concept of data depth provides new perspectives to probabilistic as well as computational geometries. In particular, the development of implementable computing algorithms for depth-based statistics has brought about many new challenges in computational geometry. This workshop created a unique environment for multidisciplinary collaboration among computer scientists, theoretical and applied statisticians, and data analysts. It brought together active researchers in these fields to discuss significant open issues, establish perspective on applications, and set directions for further research.

The DIMACS Workshop on Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications was held on May 14 - 16, 2003 at the DIMACS Center, Rutgers University, Piscataway, NJ. There were 92 participants, including 48 faculty, 13 researchers, 30 students and 1 post-doctoral fellow.

The workshop was presented under the auspices of the DIMACS Special Focus on Data Analysis and Mining and the DIMACS Special Focus on Computational Geometry and Applications.

Among the highlights of the talks was an overview on depth functions in nonparametric multivariate analysis by Robert Serfling, University of Texas at Dallas, and a talk on depth and arrangements by David Eppstein, University of California, Irvine. Karl Mosler, Universitat zu Koln, Germany, gave a very interesting presentation on data analysis by zonoid depth and Mario Romanazzi, "Ca' Foscari University of Venice, Italy gave one on data depth in multivariate analysis with emphasis on dependence, discrimination, and clustering. These talks were illustrative of the international flavor of the workshop. Government agencies were also represented. For instance, there was a talk on fast algorithms for frames and point depth in polyhedral hulls by Jose Dula of the US Census Bureau. Industry was also represented. Suresh Venkatasubramanian of AT&T Labs - Research addressed the hardware side of the visualization issue, explaining how depth contour plots can give a quick idea as to the shape of a data distribution, especially for time-varying data. He presented a series of examples demonstrating how the graphics pipeline can be used to compute depth contours for various depth measures in the fastest implementation known for points in two dimensions. Computer scientist Peter Meer, Rutgers University, in his talk on nonparametric clustering of high-dimensional data, presented

computationally efficient algorithms for the representation of visual information based on nonparametric clustering methods. This work resulted from the collaboration of Meer, Ilan Shimshoni, Technion-Israel Institute of Technology, Israel, and Bogdan Georgescu, Department of Computer Science, Rutgers University.

Additional talks were as follows:

A Note on Simplicial Depth
    Michael Burr, Tufts University

Maximal Depth Estimators of Regression Quantiles for Censored Data
    Steve Portnoy, University of Illinois

Data Depth and Mixture Models
    Ryan Elmore, Thomas Hettmansperger*, Fengjuan Xuan, Penn State University

Comparing Multivariate Scale Using Data Depth: Testing for Expansion or Contraction
    Kesar Singh, Rutgers University
A Test for Equal Variances Using Data Depth
    Karen McGaughey, Kansas State University

Computing the Center of Area of a Convex Polygon
    Pat Morin, Carleton University, Canada

Distance Problems on Points and Lines
    Ovidiu Daescu, Ningfang Mi, University of Texas at Dallas

Computation of Projection Depth and Related Estimators
    Yijun Zuo, Michigan State University

On some Probabilistic Algorithms for Computing Tukey's Half Space Depth
    Biman Chakraborty*, National University of Singapore and Probal Chaudhuri, Indian Statistical Institute

On Data Based Distances between Data Vectors and between Hyperplanes
    Hannu Oja, University of Jyvaskyla, Finland

Beyond Multivariate Location: Halfspace Depth in Various Models of Data Analysis
    Ivan Mizera, University of Alberta

A Definition of Depth for Functional Observations
    Sara López-Pintado, Juan Romo*, Universidad Carlos III de Madrid, Spain

A Depth_Based Kurtosis Functional
    Jin Wang, University of Texas at Dallas

A Computational Tool for Depth-Based Statistical Analysis
   Eynat Rafalin, Tufts University

Tukey Depth-based Trimmed Means
   Jean-Claude Masse, Universite Laval, Quebec, Canada

Clustering and Cluster validation via the Relative Data Depth
   Rebecka Jornsten, Rutgers University

On Aspects of Regression Depth and Methods based on Convex Risk Minimization for Data Mining
   Andreas Christmann, University of Dortmund, Germany

Functional Samples and Bootstrap for Predicting SO_2 Levels
   Belen Fernandez-de-Castro*, S.Guillas, W. Gonzalez Manteiga,Universidade de Santiago de Compostela, Spain

Exact, Adaptive, Parallel Algorithms for Data Depth Problems
   Vera Rosta, McGill University, Canada

Analyzing The Number of Samples Required for an Approximate Monte-Carlo LMS Line Estimator
   David M. Mount, University of Maryland

Optimizing Depth Functions
   Stefan Langerman, Universite Libre de Bruxelles

On the Computation and Robustness of Some Data Depth Notions
   Greg Aloupis, McGill University

III. Project Findings

Workshop Volume

A volume resulting from this workshop is in preparation, and is scheduled to be published by the American Mathematical Society in June 2004 as a volume in the DIMACS book series.
This volume will contain all the papers presented in the workshop as well as the results of some new collaborative research generated through the contacts established during the workshop.

New collaborations and research directions

The main goal of the project was to stimulate new collaborations and research discussions. A number of communications have begun among various workshop participants and several collaborations, of an interdisciplinary nature have begun. The following are two examples.   Since the workshop, Eynat Rafalin and Diane Souvaine (Computer Science, Tufts University) and Regina Liu (Statistics, Rutgers

University) have begun their collaboration on bridging the gap of the sample version of depth function values in order to broaden the applicability of the depth function to data analysis. Robert Serfling (Statistics) and Ovidiu Daescu (Computer Science) have begun work on extending depth function methods to develop some new foundational contributions to computational geometry. Both Serfling and Daescu are at University of Texas at Dallas, but it took participation in the workshop to launch their collaboration. They are currently preparing a joint research proposal for submission to CISE.

IV. Project Training/Development

The workshop offered some student scholarships to reimburse expenses for travel, registration, and lodging at the workshop. The following students were supported:

Greg Aloupis, McGill University
Michael Burr, Tufts University,
Elvan Ceyhan, Johns Hopkins University
Ryan Elmore, Penn State University
Karen McGaughey, Kansas State University
Eyant Rafalin, Tufts University
Fracisco Vazquez-Grande, Universidad Autonoma de Madrid
Jin Wang, University of Texas at Dallas
Iris Y. Wei, University of Illinois at Urbana-Champaign
Fengjuan Suan, Penn State University
Jianhui Zhou, Michigan State University

Five of these students presented papers in the workshop and their papers will be published in the proceedings. [Brenda: you can't refer to section numbers. These are our own numbers.]

V. Outreach Activities

VI. Papers/Books/Internet

Main web site for Workshop on Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications
http://dimacs.rutgers.edu/Workshops/Depth/

The proceedings for this workshop is in preparation, and is scheduled to be published by the American Mathematical Society in June 2004 as a volume in the DIMACS book series.

VII. Other Products

Main web site for DIMACS 2001-2004 Special Focus on Data Analysis and Mining
http://dimacs.rutgers.edu/SpecialYears/2001_Data/

Main web site for DIMACS 2002-2005 Special Focus on Computational Geometry and Applications
http://dimacs.rutgers.edu/SpecialYears/2002_CompGeom/

VIII. Contributions within Discipline

We ultimately expect exciting research results because of the many collaborations that were established and we expect these collaborations to flourish and be productive for years.

Many of the workshop participants hoped that this workshop should be the first meeting of many as they developed new collaborations.

IX. Contributions -- other Disciplines

This was an inherently interdisciplinary project. Connections between computer science, statistics, and other disciplines were brought to light. New collaborations between computer scientists, theoretical and applied statisticians, and data analysts resulted.

X. Contributions -- Human Resource Development

One of the key components of the workshop was the strong participation of students. The workshop included five papers that were co-authored and/or presented by female graduate students. Several of the students were subsequently invited by other participants to present their work elsewhere or to carry out future collaboration on depth related subjects. The workshop also provided substantial travel support for eleven students.

XI. Contributions to Resources for Research and Education

XII. Contributions Beyond Science and Engineering