## FAST MOVING FRONTS - 2009

**January 2009**

PDF

**Mia Hubert, Peter Rousseeuw, & Karlien Vanden Branden talk with *ScienceWatch.com* and answer a few questions about this month's Fast Moving Front in the field of Mathematics.**

**Article: ROBPCA: A new approach to robust principal component analysis**
Authors: Hubert, M;Rousseeuw, PJ;Vanden Branden, K
Journal: TECHNOMETRICS, 47 (1): 64-79, FEB 2005
Addresses: Katholieke Univ Leuven, Dept Math, B-3001 Louvain, Belgium.
Katholieke Univ Leuven, Dept Math, B-3001 Louvain, Belgium.
Univ Antwerp, Dept Math & Comp Sci, B-2020 Antwerp, Belgium.
Katholieke Univ Leuven, Dept Math, B-3001 Heverlee, Belgium.

---

**SW: Why do you think your paper is highly cited?**

Our paper offers a solution to an important problem in statistics and data analysis: how to perform data reduction when the observations may be contaminated with outlying values. This problem is especially important for the analysis of high-dimensional data sets, such as spectral data in chemometrics and genetic data in bio-informatics. We propose an algorithm which is highly robust and computationally feasible, and we also provide a graphical tool for outlier detection.

Moreover, our method serves as the cornerstone of new highly robust calibration methods (principal component regression and partial least squares regression), a robust classifier, and robust multi-way techniques. The availability of user-friendly software in our Matlab toolbox LIBRA has facilitated the practical use of our algorithm. So far applications of our method have been developed in chemometrics, bio-informatics, image analysis, face recognition methods, computer vision, sensory analysis, statistical quality control, and fault detection.

*Coautor*
*Peter Rousseeuw*

**SW: Does it describe a new discovery, methodology, or synthesis of knowledge?**

The new method combines elements of two existing approaches. The first is the minimum covariance determinant method, which dates back to 1984 and for which a fast algorithm was constructed in 1999 (Rousseeuw P and Van Driessen K, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics* 42: 212-23, 1991). This approach is quite accurate, but is not applicable when there are more dimensions than observations (as in the case of

spectra).

The second approach is principal component analysis by projection pursuit, as advocated by several authors. That approach can deal with high-dimensional data but is typically less accurate. Our proposed method, ROBPCA (Robust Principal Component Analysis), combines the advantages of both approaches by being robust, more accurate, and able to handle high-dimensional data.

*Coauthor*
*Karlien Vanden Branden*

**SW: Would you summarize the significance of your paper in layman's terms?**

Principal component analysis is the most popular technique for data reduction. It transforms data with many variables (columns) to a coordinate system with fewer variables, which often have a meaningful interpretation. Data reduction is extremely useful nowadays, as new data collection methods are widely applied and data storage has become much cheaper.

Consider, for example, microarray data containing gene expressions of thousands of genes, or online process measurements. Unfortunately, the more data are gathered, the more likely it becomes that outliers will be present. Hence, there is a need for statistical methods that perform data reduction while at the same time being robust against outliers, i.e., able to resist the ill effects of outlying cases and able to detect these cases. The proposed method is robust in this sense, and the outlying cases can be detected by the outlier map that is a part of the output.

**SW: Where do you see your research leading in the future?**

We continue to do work on constructing robust versions of other statistical techniques, with an eye toward computational feasibility.

**Do you foresee any social or political implications for your research?**

We don't expect a direct effect, but statistical methods are used in all fields of science (including sociology) as well as in business and for political decision-making. Therefore, developing better statistical tools holds the promise of improving the insights and conclusions of research work done in all these areas.

**Professor Mia Hubert**
**Department of Mathematics**
**Katholieke Universiteit Leuven**
**Leuven, Belgium**

**Professor Peter Rousseeuw**
**Department of Mathematics and Computer Science**
**University of Antwerp**
**Antwerp, Belgium**

**Karlien Vanden Branden, Ph.D.**
**Dexia Bank**
**Brussels, Belgium**

Keywords: high-dimensional data sets, spectral data in chemometrics, matlab toolbox libra, robust principal component analysis.

PDF

back to top

2009 : January 2009 - Fast Moving Fronts : Mia Hubert, Peter Rousseeuw, & Karlien Vanden Branden

Scientific Home | About Scientific | Site Search | Site Map