Geophysical Research Abstracts, Vol. 7, 02212, 2005
SRef-ID: 1607-7962/gra/EGU05-A-02212


# How to retrieve physical information from highly contaminated data: a most robust and most efficent new method

**P.L. Xu**

Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

The findings of this paper are summarized as follows: (i) We propose a new robust estimation method, which can tolerate \$50\%\$ of data contamination and meanwhile achieve high, least-squares-comparable efficiency. Since the objective function is identical with least squares, the method may also be called robust least squares. An iterative version of the method has been implemented and shown to be capable of resisting against more than \$50\%\$ of contamination. As a by-product, a robust estimate of scale parameter can also be obtained. Unlike the least median of squares method and repeated medians, which use a least possible number of data to derive the solution, the new robust least squares method attempts to employ a maximum possible number of good data to derive the robust solution, and thus will not be affected by partial near multi-collinearity among part of the data or if some of the data are clustered together; (ii) Although M-estimates have been reported to have a breakdown point of \$1/(t+1)\$, we have shown that the weights of observations can readily deteriorate such results and bring the breakdown point of M-estimates to zero. The same zero breakdown point of the \$L_1\$-norm method is also derived, again due to the weights of observations; (iii) By assuming a prior distribution for the signs of outliers, we have developed the concept of subjective breakdown point, which may be thought of as an extension of stochastic breakdown by Donoho \& Huber but can be important in explaining real-life problems in Earth Sciences and image reconstruction; and finally, (iv) We have shown that the least median of squares method can still break down with a single outlier, even if no highly concentrated good data nor highly concentrated outliers exist.