Geophysical Research Abstracts
Vol. 14, EGU2012-5117-1, 2012
EGU General Assembly 2012

An optimized Leave One Out approach to efficiently identify outliers

L. Biagi (1), S. Caldera (2), and D. Perego (1)

(1) Politecnico di Milano, DIIAR, Geomatics Laboratory at Como Campus, Como, Italy (ludovico.biagi@polimi.it), (2) GRED, Como, Italy (stefano.caldera@gmail.com)

Least squares are a well established and very popular statistical toolbox in geomatics. Particularly, LS are applied to routinely adjust geodetic networks in the cases both of classical surveys and of modern GNSS permanent networks, both at the local and at the global spatial scale.

The linearized functional model between the observables and a vector of unknowns parameters is given. A vector of N observations and its apriori covariance is available. Typically, the observations vector can be decomposed into n subvectors, internally correlated but reciprocally uncorrelated. This happens, for example, when double differences are built from undifferenced observations and are processed to estimate the network coordinates of a GNSS session. Note that when all the observations are independent, n=N: this is for example the case of the adjustment of a levelling network.

LS provide the estimates of the parameters, the observables, the residuals and of the a posteriori variance. The testing of the initial hypotheses, the rejection of outliers and the estimation of accuracies and reliabilities can be performed at different levels of significance and power. However, LS are not robust. The a posteriori estimation of the variance can be biased by one unmodelled outlier in the observations. In some case, the unmodelled bias is spread into all the residuals and its identification is difficult.

A possible solution to this problem is given by the so called Leave One Out (LOO) approach. A particular subvector can be excluded from the adjustment, whose results are used to check the residuals of the excluded subvector. Clearly, the check is more robust, because a bias in the subvector does not affect the adjustment results. The process can be iterated on all the subvectors. LOO is robust but can be very slow, when n adjustments are performed.

An optimized LLO algorithm has been studied. The usual LS adjustment on all the observations is performed to obtain a 'batch' result. The contribution of each subvector is subtracted from the batch result by algebraic decompositions, with a minimal computational effort: this holds for the parameters, the a posteriori residuals and the variance. Therefore all the n subvectors of residuals can be checked. The algorithm provides exactly the same results of the usual LOO but it is significantly faster, because it does not require any iteration of the adjustment. In some way, this is an inverse application of the well known sequential LS where the parameters are estimated sequentially by adding the contribution of new observations as they are available. In the presentation, the optimized LOO is discussed. Its application to a very simple example of a levelling network is discussed and compared to the usual approaches for outliers identification, in view of a further study for the application to the real time quality check of positioning services.