The Robust Statistics method applied in the Kalman filter: theory and application.

The Robust Statistics method applied to the Kalman filter: theory and application

Alessandro Carosio, *IGP ETH – Swiss Federal Institute of Technology Zürich, Swiss*
Alberto Cina, *DITAG – Politecnico di Torino, Italy*
Marco Piras, *DITAG – Politecnico di Torino, Italy*

BIOGRAPHY

Alessandro Carosio is a full professor at the Institut for Geodesy and Photogrammetry of ETH Zurich. He is in charge of teaching and research on GIS and the Theory of Errors. Alberto Cina is a professor at the Politecnico di Torino. He works on GPS networks, Theory of Errors and Surveying. Recently, he has been developing GPS networks in Piedmont, Italy. Marco Piras is Ph.D student in Geodesy and Geomatic at the Politecnico di Torino. He works on real time deformation monitoring, the treatment of GPS measures and the Kalman Filter at the Politecnico di Torino.

ABSTRACT

This research deals with the possibility of using the robust methods, particularly the Huber-estimator, in a sequential approach. The Kalman filter is a frequently used sequential technique, which foresees the use of velocity filter in the algorithm. The disadvantage, as in all non robust methods, is that any possible errors , that are not removed in the original dataset, could lead to an incorrect solution. There are different methods in literature devoted to analyzing and detecting of outliers, that can be applied the Kalman filter. In robust statistics, the Huber estimator foresees three ways of verifying the basic conditions in the objective function: the first is to modify the normal matrix, the second is to modify the weight matrix and the last is to modify the observation vector. The last method has been applied to the Kalman filter and a particular Robust Kalman Filter has been generated. This method is based on sequential calculus using the Kalman filter applied to robust statistics. This operation makes the relative residual equal to the k parameter, for each modified observation. In this situation the Least Squares method can also be applied for these observations. The RKF has also been applied to a dataset with different outlier and the results obtained with other typical sequential techniques (Sequential Least Squares) have been compared. The method, using the robust statistics, suffers less from the presence of outlier because it has a breakpoint of about 0,2. This allows many errors to be present in the data, in the same epoch, but a correct solution to be obtained. The technique could play an important role in sequential applications, for example in the control of area correction parameters, calculated from a control center. which manages a GPS stations permanent network. Another possible case is that of real time deformation control. The implemented method could therefore be a good solution for outlier detection or for to calculate the estimated solution with a low dependence on the outliers.

INTRODUCTION

The Kalman filter is satisfactorily used in different disciplines (i.e. economy, engineering, etc) as it is a useful tool for the recursive treatment of dynamic linear systems. The value of an estimated parameter can usually change epoch by epoch, but not in number. These models are called "static systems". In other cases, called dynamic, the estimation of parameters [m] depends both on the observations and on the motion model (usually when there are deformations phenomena or estimation of the positioning of the vehicles). In these case, the unknown of motion model are added to the unknown parameters of system. The solution is then calculated by applying the Kalman Filter (KF). If there is a dynamic linear system, the state vector estimation problem is solved using the Kalman Filter through the linear observation model:

$$x_k = F_k x_{k-1} + \varepsilon_k \qquad (1)$$

$$y_k = A_k x_k + e_k \qquad (2)$$

where:
x_k: state vector at epoch k [m];
y_k: observation vector at epoch k [n];
F_{k-1}: transition matrix [m, m];
A_k : design matrix [n, m];
ε_k and e_k system and observation noise, where $E(\varepsilon_k\varepsilon_k^T)=C_{\varepsilon k}$; $E(e_k e_k^T)=C_{ek}$; $E(\varepsilon_k) = 0$; $E(e_k) = 0$. Assuming that the initial values of x_0 and its cofactor matrix Q_{x0} are known and ε_0 and e_0 are equal to white noise, which are uncorrelated to x_k and to other values, the estimated Kalman Filter is given one

prediction step and one correction step, for the model described in (1) and (2).
In the prediction step:

$$\hat{x}_{k/k-1} = F_{k/k-1}\hat{x}_{k-1/k-1} \tag{3}$$

$$Q_{k/k-1} = F_{k/k-1}Q_{k-1/k-1}F^T_{k/k-1} + C_{\varepsilon k} \tag{4}$$

in the correction step:

$$\begin{aligned}\hat{x}_{k/k} &= \hat{x}_{k/k-1} \\ &+ Q_{k/k-1}A_k^T(A_kQ_{k/k-1}A_k^T + C_{ek})^{-1}(y_k - A_k\hat{x}_{k/k-1})\end{aligned} \tag{5}$$

$$\begin{aligned}Q_{k/k} &= Q_{k/k-1} \\ &- Q_{k/k-1}A_k^T(A_kQ_{k/k-1}A_k^T + C_{ek})^{-1}A_kQ_{k/k-1}\end{aligned} \tag{6}$$

The classical KF foresees a normal distribution of the errors ε and e. The hypothesis of normalized residuals is not always verified, as in the case where an outlier or other contaminations of the data are present. The purpose of this research is to have a correct estimation of parameters in sequential way and a Robust Kalman Filter (RKF) is therefore required. In literature there are several different procedures to make the previous sequential technique robust. Some of these are very hard to apply in practical cases for example in geodetic problems. Some practical Robust Kalman Filter are here described and after a brief description of the Huber Estimator and the derived method, called BIBER Estimator, a new approach is proposed.

ROBUST KALMAN FILTER USING WEIGHTS

The first method that is here dealt was described by Kubik and Wang [32] and it is used in kinematic GPS positioning. This method is based on an algorithm which consider the variance-inflation outlier modelling. An outlier often means that the residual distribution ,v, has tails longer that those of the normal Gaussian distribution. One of these distributions is the variance-inflation model:

$$v_k \approx \alpha_1 N_v(0, C_{ek1}) + \alpha_2 N_v(0, C_{ek2}) \tag{7}$$

where $N_v(0, C_e)$ defines the normal probability density function (p.d.f) with zero-mean, variance matrix C_e, $0<\alpha_1<<1$, $\alpha_1 + \alpha_2 = 1$, $C_{ek2} > C_{ek1}$. This model was opportunely studied and modified by Kubik e Wang, so that it could be applied to geodetic problems.

The model expressed in (7) is considered and if we assume the "a posteriori" density distribution, $p(x_1|Y_1)$ which is the conditioned probability of x_1 (the single parameter that has to be estimated) Y_1 is known (the observations vector) and this probability follows a Gaussian distribution, it is possible to show that for each k >> 1, $p(x_k|Y_{k-1})$ can be close to Gaussian density measured in terms of the Kullback - Leiber distance. Under this hypothesis, the foreseen y_k distribution, given Y_{k-1}, is:

$$\begin{aligned}p(y_k \mid Y_{k-1}) &= \int p(y_k \mid x_k)p(x_k \mid Y_{k-1})dx_k \\ &= \alpha_1 N_y(A_k x_{k/k-1}, R_{k1}) + \alpha_2 N_y(A_k x_{k/k-1}, R_{k2}) \\ &= c[\alpha_1(|R_{k1}|)^{-\frac{1}{2}}\exp(-0{,}5v_k^T R_{k1}^{-1}v_k) \\ &+ \alpha_2(|R_{k2}|)^{-\frac{1}{2}}\exp(-0{,}5v_k^T R_{k2}^{-1}v_k)]\end{aligned} \tag{8}$$

where

c is a constant,

$$R_{ki} = A_kQ_{k/k-1}A_k^T + C_{eki} \tag{9}$$

And the expected residual is:

$$v_k = y_k - A_kx_{k/k-1} \tag{10}$$

$|R_{ki}|$ is the determinant of R_{ki}

Equation (10) describes the term of innovation.
Thus

$$\begin{aligned}&\partial p(y_k \mid Y_{k-1})/\partial y_k = \\ &= c[\alpha_1(|R_{k1}|)^{-\frac{1}{2}}\exp(-0{,}5v_k^T R_{k1}^{-1}v_k)R_{k1}^{-1}v_k \\ &+ \alpha_2(|R_{k2}|)^{-\frac{1}{2}}\exp(-0{,}5v_k^T R_{k2}^{-1}v_k)R_{k2}^{-1}v_k]\end{aligned} \tag{11}$$

Placing

$$\alpha_{k1} = \left\{1 + (\alpha_2/\alpha_1)(|R_{k1}|/|R_{k2}|)^{\frac{1}{2}}\exp\left(0.5v_k^T\left(R_{k1}^{-1} - R_{k2}^{-1}\right)v_k\right)\right\}^{-1}$$

(12)

$$\alpha_{k2} = 1 - \alpha_{k1} \tag{13}$$

Considering what has been shown until now, equation (5) and (6) can be redefined as follows:

$$\begin{aligned}\hat{x}_{k/k} &= \hat{x}_{k/k-1} \\ &+ Q_{k/k-1}A_k^T(\alpha_{k1}R_{k1}^{-1} + \alpha_{k2}R_{k2}^{-1})(y_k - A_k\hat{x}_{k/k-1}) \\ &= \alpha_{k1}x_k^{(1)} + \alpha_{k2}x_k^{(2)}\end{aligned} \tag{14}$$

where

$$x_k^{(i)} = x_{k/k-1} + Q_{k/k-1}A_k^T(R_{ki})^{-1}(y_k - A_k\hat{x}_{k/k-1}) \tag{15}$$

The Covariance matrix is:

$$Q_{k/k} = Q_{k/k-1} - Q_{k/k-1}A_k^T G_k(y_k)A_kQ_{k/k-1} \tag{16}$$

where

G_k is the score function matrix.

This Robust Kalman Filter can be defined applying Masreliez's theorem, with a score function derived from variance-inflation model. In the previously mentioned algorithm, the state vector is estimated from a weighted summation of two standard KFs. The weights are described by α_{k1} and α_{k2} which are the "a posteriori" probability of events that the observations noise come from $N_1(0, C_{ek1})$ and $N_2(0, C_{ek2})$ respectively. The score function estimate is a non linear estimate of the observations as the weights α_1 and α_2 are non-linear functions of the observations like the estimates values (14) and (16).

ROBUST KALMAN FILTER AND HUBER ESTIMATOR

There are different robust estimation and outlier detection methods, where the purpose is both the definition and detection of the outliers and an estimation without being influenced by outlier effects. It is very important to have a special tool that is able to define them effectiveness of the method in relation to the different purposes. The Huber Estimator is classified as an optimal robust procedure to solve problems with outlier defined in the "y-direction". This means outliers located in direction y, considering the observation space (Fig. 1).

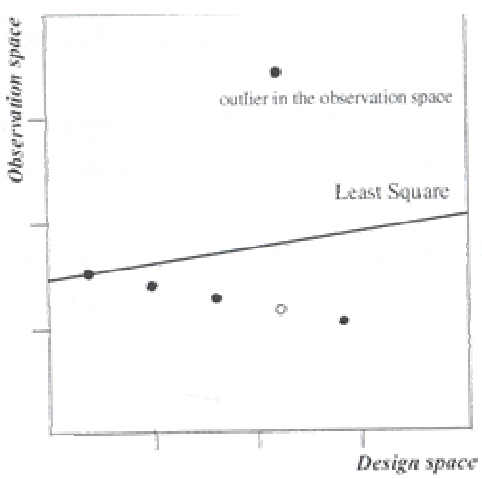

Fig. 1: Least Squares fit with outlier in the obs. space

The estimation of unknown x parameters is usually solved by the Least Squares, minimizing the sum of residual squares. Huber, instead, proposes to minimize:

$$\sum \rho \left(\frac{\sum_{i=1}^{n} \rho(\sum_{j=1}^{m} (a_{ij} x_j) - y_i)}{\sigma} \right) \qquad (17)$$

where ρ is the convex function of the residual, which increases less quickly than the squared function. The term σ represents the deviation standard of the observations. Making the partial derivative of previous function with respect to $x_1,..., x_m$ and defining ψ as the influence function, it is possible to obtain the following system of equations:

$$\sum_{i=1}^{n} \psi\left(\frac{v_i}{\sigma}\right) a_{i,j} = 0 \qquad for\ j = 1,...,m \qquad (18)$$

The solution of system (18) is called "M-Estimate" from x_1 to x_m of the Huber type, where the ρ function is an arbitrary function that must be chosen. Huber proposes the following function:

$$\rho(v) = \begin{cases} \frac{1}{2} v^2 & for\ |v| \le c \\ c|v| - \frac{1}{2} c^2 & for\ |v| > c \end{cases} \qquad (19)$$

The term c is called the tuning constant and its value depends on the data dispersion. This function is identical to the Least Squares objective function in the interval (-c; c) and outside this interval corresponds to a linear function. The estimator defined by functions ρ, ψ and equations (19) is called the Huber Estimator. If c → ∞ the Huber estimators coincides with the Classical Least Squares estimator. The ρ and ψ functions of the Least Squares and Huber Estimator are compared and drawn in (Fig. 2) and (Fig.3).

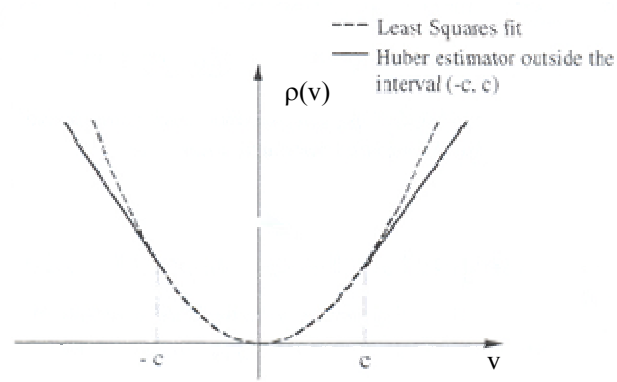

Fig. 2: objective function of Least Squares and Huber estimator

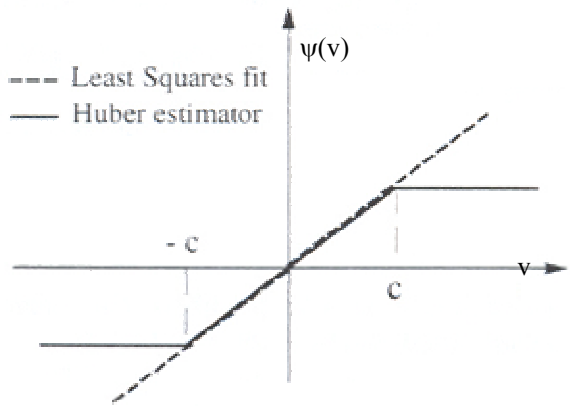

Fig. 3: influence function of Least Squares and Huber estimator

The normal equation can be defined by differentiating (19) from the unknown parameters and considering the partial derivative equal to zero. A system composed of three different cases: $v \leq -c$, $-c \leq v \leq c$ and $v \geq c$ is made through this operation. Initially it is not known in which of the previous intervals the residual v will be. The problem is numerically solved thanks to the help of some iterative solution of calculus. The first solution is calculated through an iterative process using the Least Squares estimator, but the result might not be satisfactory. The process continues, moving the value with an outlier towards the fixed values, modifying the values continually until convergence is obtained. Assuming σ equal to 1, the Huber estimator is defined as the solution of (18). Studying this approach, the authors understood that this method limits the influence of the outliers to just in y-direction, but it does not consider the outlier position, in observation space. Another approach was developed from the Huber Estimator and M estimator to give robustness to Kalman Filter. This method was described by Cipra and Romera [9]. The relation (5) can derive, known the values (3) and (4) and described the current values y_k, by the next minimizing procedure:

$$\hat{x}_{k/k} = \arg\min(x_k)\{(\hat{x}_{k/k-1} - x_k)^T (Q_{k/k-1})^{-1} (\hat{x}_{k/k-1} - x_k) + (y_k - A_k x_k)^T C_{ek}^{-1} (y_k - A_k x_k)\} \quad (20)$$

where argmin is defined in the whole $x_k \in \Re^n$. Argmin represents the value of the variable with which the function value concerned reaches its minimum. This procedure can be considered as the weighted Least Squares method, but in reality it is equivalent to a non-weighted Least Squares in the case of a linear regression model:

$$\begin{pmatrix} (Q_{k/k-1})^{-0.5} \hat{x}_{k/k-1} \\ C_{ek}^{-0.5} y_k \end{pmatrix} = \begin{pmatrix} (Q_{k/k-1})^{-0.5} \\ C_{ek}^{-0.5} A_k \end{pmatrix} x_{k/k} + \begin{pmatrix} w_k \\ \eta_k \end{pmatrix} \quad (21)$$

where the residuals w_k and η_k are defined as:

$$E(w_k) = 0; \quad E(\eta_k) = 0; \quad \mathrm{var}\begin{pmatrix} w_k \\ \eta_k \end{pmatrix} = I \quad (22)$$

Furthermore, comparing (21) with (1) and (2), the following equations are obtained:

$$\begin{aligned} w_k &= (Q_{k/k-1})^{-0.5} (\hat{x}_{k/k-1} - F_k x_{k-1/k-1} - \varepsilon_k) \\ \eta_k &= Q_{k/k}^{-0.5} e_k \end{aligned} \quad (23)$$

therefore a possible contamination of ε_k results in a contamination of w_k without affecting η_k. There is a similar behaviour in the second equation, (23), because a probable contamination of e_k damages η_k without corrupting w_k. The model defined in (21), can now be rewritten distinguishing for particular row as:

$$\begin{aligned} q_{ik} &= n_{ik} x_k + w_{ik}, \qquad i = 1,\ldots,n \\ c_{jk} &= r_{ik} x_k + \eta_{jk}, \qquad j = 1,\ldots,m \end{aligned} \quad (24)$$

where

$$(Q_{k/k-1})^{-0.5} \hat{x}_{k/k-1} = \begin{pmatrix} q_{1k} \\ \ldots \\ q_{nk} \end{pmatrix}, \qquad (Ce_k)^{-0.5} y_k = \begin{pmatrix} c_{1k} \\ \ldots \\ c_{mk} \end{pmatrix},$$

$$(Q_{k/k-1})^{-0.5} = \begin{pmatrix} n_{1k} \\ \ldots \\ n_{nk} \end{pmatrix}, \qquad (Ce_k)^{-0.5} A_k = \begin{pmatrix} r_{1k} \\ \ldots \\ r_{mk} \end{pmatrix},$$

$$w_k = \begin{pmatrix} w_{1k} \\ \ldots \\ w_{nk} \end{pmatrix}, \qquad \eta_k = \begin{pmatrix} \eta_{1k} \\ \ldots \\ \eta_{mk} \end{pmatrix}$$

The model described in (24) is comparable to the Gauss-Markow model, where the system solution is:

$$\hat{x}_{k/k} = \arg\min\{\sum_{i=1}^{n} (q_{ik} - n_{ik} x_k)^2 + \sum_{j=1}^{m} (c_{jk} - r_{jk} x_k)^2\} \quad (25)$$

that can be robustifyied using (25) with:

$$\hat{x}_{k/k} = \arg\min\{\sum_{i=1}^{n} \rho_{1i} (q_{ik} - n_{ik} x_k)^2 + \sum_{j=1}^{m} \rho_{2j} (c_{jk} - r_{jk} x_k)^2\} \quad (26)$$

where ρ_{1i} and ρ_{2j} are opportune robustness functions with derivative equal to $\psi_{1i}(i=1,...,n)$ and $\psi_{2j}(j=1,...,m)$ as it is used in the M-estimator method.
The application of the robustness function in (26) leads to the contamination effect of ε_k and e_k being eliminated. The normal function of $\hat{x}_{k/k}$, given by (26), is:

$$\sum_{i=1}^{n} n_{ik}^T \psi_{1k}\left(q_{ik} - n_{ik}\hat{x}_{k/k}\right)^2 + \sum_{j=1}^{m} r_{ik}^T \psi_{2k}\left(c_{jk} - r_{jk}\hat{x}_{k/k}\right)^2 = 0 \tag{27}$$

which can only be solved explicitly in some special cases. In general, it is possible to apply the following approximate normal equation:

$$\sum_{i=1}^{n} w_{1ik} n_{ik}^T \left(q_{ik} - n_{ik}\hat{x}_{k/k}\right)^2 + \sum_{j=1}^{m} w_{2jk} r_{jk}^T \left(c_{jk} - r_{jk}\hat{x}_{k/k}\right)^2 = 0 \tag{28}$$

where the weights w_{1ik} (i = 1,..., n) and w_{2jk} (i = 1,..., m) are defined as:

$$w_{1ik} = \frac{\psi_{1i}(q_{ik} - n_{ik}\hat{x}_{k/k-1})}{q_{ik} - n_{ik}\hat{x}_{k/k-1}} \qquad w_{2jk} = \frac{\psi_{2j}(c_{jk} - r_{jk}\hat{x}_{k/k-1})}{c_{jk} - r_{jk}\hat{x}_{k/k-1}} \tag{29}$$

Equations (29) follow from (27) only if $\hat{x}_{k/k}$ is approximated with $\hat{x}_{k/k-1}$. This method can be considered as a recursive variant of the normal equation of the IWLS (Iterated Weighted Least Square).
Using (26) and applying some algebraic operations, the next robust relation is extracted from equations (5) and (6):

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5} A_k^T (A_k Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5} A_k^T + C_{ek}^{0.5} W_{2k}^{-1} C_{ek}^{0.5})^{-1} (y_k - A_k \hat{x}_{k/k-1}) \tag{30}$$

$$Q_{k/k} = Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5} - Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5} \cdot A_k^T (A_k Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5} A_k^T + C_{ek}^{0.5} W_{2k}^{-1} C_{ek}^{0.5})^{-1} \cdot (A_k Q_{k/k-1}^{0.5} W_{1k}^{-1} Q_{k/k-1}^{0.5}) \tag{31}$$

where $\hat{x}_{k/k-1}$ and $Q_{k/k-1}$ are given in (3) and (4) and W_{1k} and W_{2k} are respectively:

$$W_{1k} = diag\{w_{1k},....., w_{nk}\} \qquad W_{2k} = diag\{w_{2k},....., w_{mk}\} \tag{32}$$

The covariance matrix estimation by means of an non approximate approach is much more complex, thus it is usually better to use equation (6). The most frequent case of contaminated data is called ε-contaminated normal data, where a normal distribution with an acceptable variance is contaminated by a small value of ε (usually 0.05) with a symmetrical distribution and heavy tails. For this kind of data, with an ε-contaminated distribution N(0, 1), Huber's function ψ_H is:

$$\psi_H(v) = \begin{cases} v & for\,|v| \le c \\ c \cdot \mathrm{sgn}(v) & for\,|v| > c \end{cases} \tag{33}$$

The constant value c depends on ε (it is advisable to use c=1.645, for ε=0.05). In the case without any errors, the classical Least Squares approach is used, where:

$$\psi_H(v) = v \tag{34}$$

If the Huber Estimator is applied the following considerations can be made. Relation (1) and (2) are generally considered in the Kalman Filter. In the normal equation, (24), $\psi_{Hi} = \psi_{HL}$(i=1,...,n) and $\psi_{21} = \psi_H$ are considered. A new recursive robust relation, applying to KF, can be described,:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + Q_{k/k-1} A_k^T v_k^{0.5} \psi_H\left(\frac{v_k^{0.5}(y_k - A_k\hat{x}_{k/k-1})}{A_k Q_{k/k-1} A_k^T + v_k}\right) \tag{35}$$

$$Q_{k/k} = Q_{k/k-1} - \frac{Q_{k/k-1} A_k^T A_k Q_{k/k-1}}{A_k Q_{k/k-1} A_k^T + v_k} \tag{36}$$

With (30), relation (32) can be rewritten as:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + \begin{cases} \dfrac{Q_{k/k-1} A_k^T}{A_k Q_{k/k-1} A_k^T + v_k}(v_p) & (IA) \\ Q_{k/k-1} A_k^T v_k^{-0.5} c \cdot \mathrm{sgn}(v_p) & (IIA) \end{cases} \tag{37}$$

where:

v_p = expected residual = $(y_k - A_k \hat{x}_{k/k-1})$

(IA.37) is used when the following relation is verified:

$$\left|v_p\right| \le c v_k^{-0.5}\,(A_k Q_{k/k-1} A_k^T + v_k) \qquad (38)$$

In the other cases (IIA.37) is used. If we consider a general function ψ instead to ψ_H, in according with (27) and (28) the solution can now be definied as:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + \frac{Q_{k/k-1} A_k^T}{A_k Q_{k/k-1} A_k^T + \frac{v_k}{g_k}} (v_p) \qquad (39)$$

$$Q_{k/k} = Q_{k/k-1} - \frac{Q_{k/k-1} A_k^T A_k Q_{k/k-1}}{A_k Q_{k/k-1} A_k^T + \frac{v_k}{g_k}} \qquad (40)$$

where:

$$g_k = \frac{\psi(v_k^{-0.5} v_p)}{v_k^{-0.5} v_p} \qquad (41)$$

There is the recursive estimation of parameter x, in a linear regression model (1) and (2), with observations y affected by errors. The easiest case could have a transition matrix F, equal to the unitary matrix.

ROBUST KALMAN FILTER WITH BIBER ESTIMATOR

The proposal here is to give robustness to the Kalman filter starting from BIBER estimator. This estimator belongs to M-estimator family, in particular the Schweppe type.

The estimator is defined as the solution of the system:

$$\sum_{i=1}^{n} \psi\left(\frac{v_i}{w_i}\right) \mu_i a_{i,j} = 0 \qquad for\ j = 1,\dots,m \qquad (42)$$

where the deviation standard of residuals is employed as additional weight μ_i. This weight is:

$$\mu_i = \sigma_{vi} = \sigma_0 \sqrt{(q_{vv})_{ii}} \qquad (43)$$

The functions ρ and ψ are represented by:

$$\rho\left(\frac{v_i}{\sigma_{vi}}\right) = \begin{cases} \frac{1}{2\sigma_{vi}} v_i^2 & for \left|\frac{v_i}{\sigma_{vi}}\right| < c \\ \frac{c}{\sigma_{vi}} \left|v_i\right| - \frac{1}{2} c^2 & for \left|\frac{v_i}{\sigma_{vi}}\right| \ge c \end{cases} \qquad (44)$$

$$\psi\left(\frac{v_i}{\sigma_{vi}}\right) = \begin{cases} \frac{v_i}{\sigma_{vi}} & for \left|\frac{v_i}{\sigma_{vi}}\right| < c \\ \mathrm{sgn}(v_i) \cdot c & for \left|\frac{v_i}{\sigma_{vi}}\right| \ge c \end{cases} \qquad (45)$$

there are three different cases, that depend on the standardized residual:

$$\begin{aligned} &\text{(I)} && \left(\frac{v_i}{\sigma_{vi}}\right) \le \mathrm{c} \\ &\text{(II)} && -\mathrm{c} < \left(\frac{v_i}{\sigma_{vi}}\right) < \mathrm{c} \\ &\text{(III)} && \left(\frac{v_i}{\sigma_{vi}}\right) \ge \mathrm{c} \end{aligned} \qquad (46)$$

Case (II) coincides with the Least Squares method. The user chooses the value of constant c, but the recommended value in literature between 2,5 and 4 generally falls.
Equation (45) can only depend on the residual v_i:

$$\psi_{ki}(v_i) = \begin{cases} v_i & for\ |v_i| < k_i \\ \mathrm{sgn}(v_i) \cdot k_i & for\ |v_i| \ge k_i \end{cases} \qquad (47)$$

where $k_i = c \cdot \sigma_{vi}$. This limit results to be different for each residual, but it is unchanged during the process of computational process, once σ_{vi} "a priori" is estimated. The solution is sought by minimizing function (47), in three different intervals. The calculus procedure is composed of three steps:

- Least squares fit
- Choice of the observation with the largest $\left(\frac{v_i}{\sigma_{vi}}\right)$
- Calculation of the unknown parameters and new residuals

Assuming the number of outliers not too high, a first compensation can be effected by Least Squares method, which offers the chance of classifying all the observations inside interval (II). In the second step, the observation with the greatest absolute standardized residual value, is removed from interval (II). When the residuals are correlated, it is difficult to individualize the observation affected by outlier, thus the one with the most elevated absolute residual value is chosen. When this observation has been identified, it is possible to estimate the unknown parameters and the new residual, thus giving a reduced influence to the marked observation. The different influence can be considered through one of the following methods:

- Recalculation of some parts of the normal matrix;
- Correction observation y_i;

- Decrease of the weight of observation y_i.

These methods lead to the same result. In the case of corrected observations, an attempt is made to modify the observations which are outside the interval (II.46). In this way the residual of those observations correspond exactly to - k_i or $+k_i$. The modified observation, $y_{i(mod)}$ has therefore less influence on the calculation of the unknown parameters. From an algebraic point of view, the problem is described with the following equations. The residuals of y_i can be expressed as:

$$v_i = -(q_{vv})_{i1} y_1 - \ldots - z_i y_i - \ldots - (q_{vv})_{in} y_n \quad (48)$$

and for the modified observations, becomes:

$$\mathrm{sgn}(v_i) k_i = -(q_{vv})_{i1} y_1 - \ldots - z_i y_{i(\mathrm{mod})} - \ldots - (q_{vv})_{in} y_n$$

(49)

Placing d_i:

$$d_i = v_i - \mathrm{sgn}(v_i) k_i \qquad \text{for } |v_i| > k_i \quad (50)$$

and comparing (48) with (49), the modified observation results to be equal to:

$$y_{i(\mathrm{mod})} = y_i - \frac{d_i}{-z_i} \quad (51)$$

After the observation has been modified, the new parameter x_{rob} and the respective residual can be estimated. In this way it is not necessary to invert the normal matrix, because the draw matrix and the weights matrix do not change.
The unknown parameters and the correspondent residuals are:

$$x_{k(rob)} = x - (A^T A)^{-1} A^T \begin{pmatrix} 0 \\ \ldots \\ \frac{d_i}{-z_i} \\ \ldots \\ 0 \end{pmatrix} \quad (52)$$

$$v_{k(rob)} = A x_{(rob)} - (y - \begin{pmatrix} 0 \\ \ldots \\ \frac{d_i}{-z_i} \\ \ldots \\ 0 \end{pmatrix}) \quad (53)$$

After this first calculation, it is necessary to establish whether the new residual falls into the interval (II.46). This procedure is applied iteratively to all observations with residuals outside this interval. After this iterative process, the parameters are estimated through a robust technique. The residuals calculated for each observation y_i, step by step, are:

$$\Delta d_i = \frac{1}{-z_i} \sum_{i=1}^{s-1} d_i \quad (54)$$

where s is the number of iterations.
The total residual is:

$$r_{i(tot)} = k_i - \Delta d_i \quad (55)$$

The σ_0^2 estimation obtained from the Least Squares is:

$$\sigma_0^2 = \frac{1}{n-m} \sum_{i=1}^{n} v^2 \quad (56)$$

In the same way the proposal with the BIBER estimator is:

$$\sigma_0^2 = \frac{1}{(n-m)\beta} \sum_{i=1}^{n} \psi_{ki}^2 (v_i) \quad (57)$$

where β is a constant that is defined by:

$$\beta = (c^2 + (1+c^2)(2\Phi(c) - 1) - 2c\varphi(c)) \quad (58)$$

In (58), Φ is the normal standard distribution, while φ is its density. An attempt was made to apply the BIBER estimator principle to the Kalman filter. Starting from the correct solution of Kalman filter (5), it can be rewritten as:

$$\hat{x}_{k/k} = N_k^{-1} (A_{k-1}^T P_{k-1} A_{k-1} + A_k^T P_k Y_k) \quad (59)$$

equation (59) can also be reduced to:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + K_k (Y_k - A_k \hat{x}_{k/k-1}) \quad (60)$$

where K_k is called gain matrix. The technique that needs to be applied in the BIBER estimator is that of defining solution (60) with a classical approach of estimating the residuals (48) and classifying them according to (46). An iterative process is followed which leads the observations to be modified, solving the problem with Least Squares, at the last step. Recalling (5) and using K term equal to:

$$K_k = Q_{k/k-1} A_k (A_k Q_{k/k-1} A_k^T + C_{ek})^{-1} \quad (61)$$

a solution is obtained that is described by:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + K_k (Y_{k(\mathrm{mod})} - A_k \hat{x}_{k/k-1}) \quad (62)$$

The calculus of the covariance matrix can be developed using (6), because no term is modified. Equations (56) and (57) need to be recalculated for the σ_0^2 estimations.
In the sequential approach, relation (56) can be rewritten as:

$$\sigma_0^2 = \frac{1}{n-m} \sum_{i=1}^{n} v^2 = \frac{1}{n-m} \left[\left(\hat{v}^T P \hat{v} \right)_{old} + \Delta \left(\hat{v}^T P \hat{v} \right)_{new} \right] \quad (63)$$

where

$$\Delta(\hat{v}^T P\hat{v})_{new} = (Y_{k(\mathrm{mod})} - A_k \hat{x}_{k-1})^T T (Y_{k(\mathrm{mod})} - A_k \hat{x}_{k-1}) \quad (64)$$

with T obtained from:

$$T = (P_k^{-1} + A_k Q_{xx(k-1)} A_k^T)^{-1} \quad (65)$$

Rewriting (57), using (63) the following is obtained:

$$\sigma_0^2 = \frac{1}{(n-m)\beta}\{(\sum_I v_{k-1}^2 + c^2 \sum_{II} z_{k-1}^2) + (Y_{k(\mathrm{mod})} - A_k \hat{x}_{k-1})^T T (Y_{k(\mathrm{mod})} - A_k \hat{x}_{k-1})\} \quad (66)$$

This method is very adapt in sequential calculations, in particular with KF. Using the KF, the normal matrix is only sequentially adjourned and not recalculated. Some applications of this method, concerning problems of GPS positioning, are now shown.

APPLICATION AND CASE STUDIES

The previously described algorithm has been developed in MATLAB and applied to some sequential problems in Geodesy, particularly when operating with measurements GPS. The first problem concerns the parameter estimation of differential corrections in a network of GPS permanent stations. In recent, the number of GPS stations has increased, leading to necessity to build a control network that can manage and check the correctness of operations. It performs a real time service, which offers different kinds of real time positioning (VRS, FKP, MAX, etc), using the network information. The main concept consists in using a correction model to replace single corrections generated from each station. The reliability of the data transmitted is very significant because users are not able to decodify the received message and there is not always the possibility of obtaining raw data. The FKP method is an example, which through a particular polynomial, all differential correction generated by a single stations of the network are interpolated. It is possible for various reasons, to have an outlier to remove that has to be removed, so as not to have a wrong real time positioning. The main cause of this incorrect parameters estimation is the absence of one or several measurements that arrive at the control center. This absence can be caused both by the absence of differential correction or high latency in the communication network. With this problem the data cannot be used, because the control centre receives the single differential correction when it is unusable. Without spending too time to explain the principles of this method, the interpolation model, using FKP method, is represented in the following and the kind of phenomenon that is obtained when the data are affected by outlier is also given.

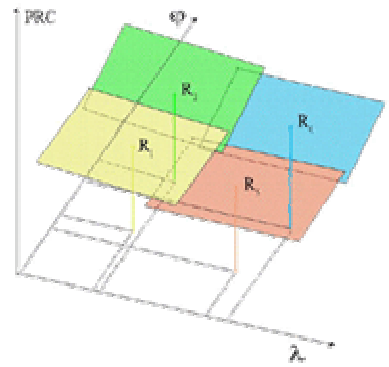

Fig. 4: Interpolation model FKP (Flächen-Korrektur-Parameter)

The interpolation model is composed of different parameters which considers the signal GPS behaviour along the satellite-receiver track.
The absence of one or several differential corrections leads to a different interpolation model, thus the parameters, which describe the model changes brusquely. Most users are not able to understand what has happened. Let us simplify this model and consider a bidimensional case, where a series of GPS stations are lined up. Each station transmits the differential corrections. In a normal situation there is a regular model (blue) defined by all the corrections, while in the case where a correction (ref2) is missing, the model is again recalculated (red).

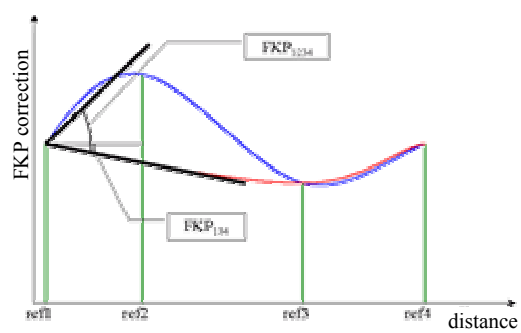

Fig. 5: FKP definition: with and without ref2

If one parameter is analysed which defines the model, it can be observed that in the normal case, while there are in some points some spikes, which are caused by phenomena previously described.
This phenomenon can be seen in fig. 6.

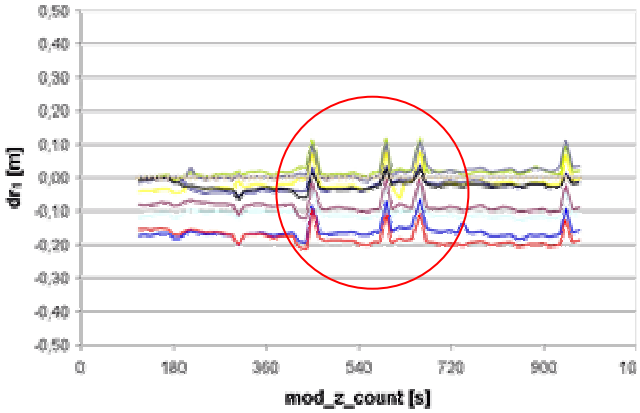

Fig. 6: dr_1 vs time

The parameters estimation can be defined by a sequential method. The problem has been solved both with sequential Least Squares and Robust Kalman Filter (RKF). Raw data have been generated in order to have an "exact" solution available and to be able to control the position and number of the outliers. The raw data represent the trend of one of the parameters of interpolation model, which is calculated using the sequential approach. In the observations (around 100), 10% are outliers. The implemented RKF has a breakpoint equal to 0,20. The same breakpoint is contained as the case of the Huber Estimator. The entity and position of the outliers change but these, considering the results, do not influence the final solution. Let ε be the value correlated to the problem. A comparison has been made between Least Squares solution, that it coincide with classical Kalman Filter, and RKF method. The traditional Kalman Filter might not be influenced by the presence of outliers if the outlier is occurred many epochs before and, above all, if it is a rare case. The problem becomes worse when the number of outlier increases and the distribution is random, therefore each solution could be contaminated by gross error.
Comparing the solutions, the following is obtained:

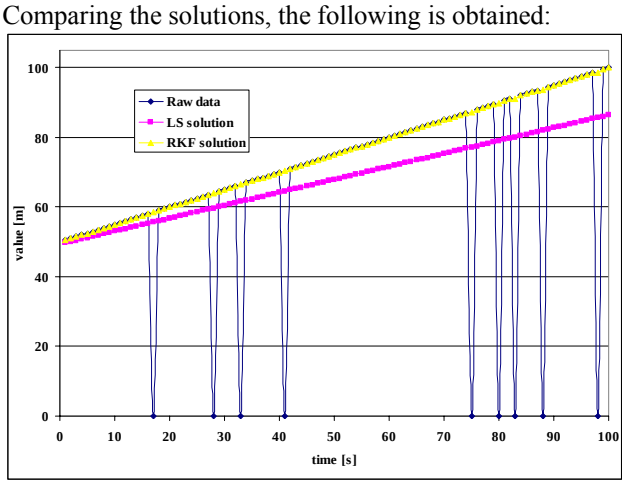

Fig. 7: LS solution vs RKF solution

The estimated solution, epoch by epoch, using the Robust Kalman Filter is almost perfectly overlapped by the raw data trend, while the Least Squares solution results, during the time, to be worse and leads to an incorrect solution.

Another possible application exists, which is also related to real time GPS positioning, but which is in this case, applied to deformation monitoring. This case considers a particular approach developed at the Politecnico di Torino, which foresees the use of a special phase combination, called triple differences, to obtain a positioning and the identification of movements in the first step. After, it is necessary to apply a more precise method to control the deformation: the double differences are used. Triple differences are more convenient because they constitute an easy combination to use during the data treatment. Triple differences, in fact, do not suffer from phase ambiguity and they also have a more regular trend, which permits eventually cycle slips phenomena to be identified in the raw data. This phenomenon is described by an high peak. The RKF method is also fundamental in this case, because it allows the triple differences trend to be redefined, without the influence of small noises or gross error. In pre-analysis, where triple differences are used, the real time correct position estimation of the monitored point is very important, to avoid false alarm process or some panic situations. The Kalman Filter is a good sequential tool, but it is not sufficient. It is also necessary to have a good outlier identification tool. In the raw data, in particular when the triple differences are used, there is both cycle slips and some errors caused by external phenomena. The triple differences trend is the following:

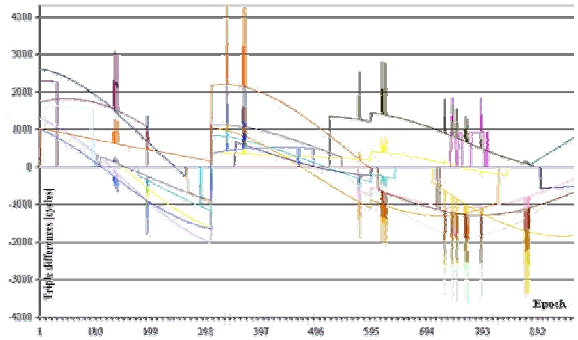

Fig. 8: Triple differences

The regularity is interrupted by the presence of outlier, which if not preventively individualized, could cause an incorrect position estimation.

CONCLUSION

Many different kinds of methods exist in literature to make a Robust Kalman Filter. Not all approaches are easy to apply

in geodesy, in particular concerning real time GPS positioning. It is clear that it is very important to have both a sequential calculus and a robust method there are not influenced by outliers or other effects in the measurements. Frequently, many process are developed in real time, therefore the main purpose is a good data analysis, to remove all the gross error in the raw data or to have a good tool which also works in a non perfect Gaussian distribution.

ACKNOWLEDGMENTS

The authors would like to thank the Italian Ministry of University and Research concerning the project "Scuola Interpolitecnica" which has made funds available to set up a collaboration between the Politecnico di Torino and the Polytechnic of Zurich, on robust statistics. The authors would in particular like to thank those people who helped during the correction of this paper.

REFERENCES

Conditioned probability

A and B are events from random test with P(B)>0. The conditioned probability of A, given B, is defined as:

$$P(A \mid B) = P(A \cap B) / P(B) \qquad (67)$$

This concept is based on the axiomatic definition of probability. The conditioned probability concept is analyzed, starting from the less formal notion and the most intuitive relative frequency. In a particular event C, with $Nn(C)$ number of times, C has been verified in the first n test. If $Nn(B)$ is high, the conditioned probability that A will be verified B having been verified, $Nn(B)$ must be close to the conditioned relative frequency of A given B. In other words, the relative frequency of A, when B has been verified, is:

$$Nn(A \cap B) / Nn(B) \qquad (68)$$

However, another application of relative frequency concept is considered:

$$Nn(A \cap B) / Nn(B)$$
$$= [Nn(A \cap B) / n] / [Nn(B) / n] \rightarrow \qquad (69)$$

$P(A \cap B) / P(B)$ as $n \rightarrow \infty$.

This leads to the same definition.
Sometimes, the conditioned probability can be directly calculated.

Kullback-Leibler distance

The Kullback-Leibler distance (KL-distance) is a natural distance function from a "true" probability distribution, p, to a "target" probability distribution, q. It can be interpreted as the expected extra message-length per datum due to using a code based on the wrong (target) distribution compared to using a code based on the true distribution. For discrete (not necessarily finite) probability distributions, $p=\{p_1, ..., p_n\}$ and $q=\{q_1, ..., q_n\}$, the KL-distance is defined to be

$$KL(p, q) = \Sigma_i\, p_i \,.\, \log_2(\, {}^{p_i}/_{q_i}\,) \qquad (70)$$

For continuous probability densities, the sum is replaced by an integral.

$$KL(p, p) = 0 \qquad (71)$$
$$KL(p, q) \geq 0$$

Note that the KL-distance is not, in general, symmetric.

Masreliez's theorem

Assuming that $p(x_k|Y_{k-1})$ is a Gaussian density, with the mean equal to $x_{k/k-1}$ and covariance matrix $Q_{k/k-1}$ and with $E(e_k e_k^T)= C_{ek}$, the conditioned value of x_k and its conditioned covariance $Q_{k/k}$ satisfy:

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + Q_{k/k-1} A_k^T g_k(y_k) \qquad (72)$$

$$Q_{k/k} = Q_{k/k-1} - Q_{k/k-1} A_k^T G_k(y_k) A_k Q_{k/k-1} \qquad (73)$$

where $g_k(y_k)$ is also called the "score function", in vector form, with components:

$$(g_k(y_k))_i = -\left[\frac{\partial p(y_k \mid Y_{k-1})}{\partial (y_k)_i}\right]\left[p(y_k \mid Y_{k-1})\right]^{-1} \qquad (74)$$

and G_k is a matrix with elements:

$$(G_k(y_k))_{ij} = \left[\frac{\partial g_k(y_k)_i}{\partial (y_k) j}\right] \qquad (75)$$

if equation (69) and (72) are compared with (3) and (9), it can be seen that this theorem preserves the recursive Kalman filter structure, but it replaces the linear score function with a non-linear score function $g_k(y_k)$. The density not Gaussian observations can be manipulated by (71) and (72).

Bibliography

[1] Andrews D. F., Bickel P., Hampel F. R., Huber P. J., Rigers W., Turkey J. W. (1972): *Robust Estimation of Location,* Princepton University Press, New Jersey;

[2] Borre, K., Strang, (1997) – "*GPS Geodesy and GPS*". Wellesley-Cambridge Press.

[3] Brown R. G, Hwang P. Y. C. (1997): *Introduction to Random Signals and Applied Kalman Filtering* - Third edition. ISBN 0 - 471- 12839-2;

[4] Carosio, A. (1996): *Les mèthodes de la statistique robuste au service de la compensation de réseaux*

géodésiques. IGP-Bericht Nr. 257f;

[5] Carosio, A. (2001): *The methods of robust statistics for applications in geodesy and GIS*. Tutorial, First Internat. Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS, ETH Zürich. IGP-Bericht Nr. 296

[6] Carosio, A. (2003): *Méthode de la statistique robuste en géodésie*. EPFL Lausanne

[7] Cina, A. Manzino, A. Piras, M. (2002) – *Controllo delle deformazioni con GPS in tempo reale*. Proceedings 6° Conferenza ASITA.

[8] Cina, A. Manzino, A. Roggero, M. (2002) – *Aspetti del posizionamento GPS differenziale in tempo reale. Un prototipo di stazione master e rover*. Proceedings 6° Conferenza ASITA.

[9] Cipra T., Romera R. (1991) *Robust Kalman Filter and Its Applications in Time Series Analysis*, Kybernetica 27; 481-494;

[10] Ershov A. A., Lipster R. S. (1978): *Robust Kalman filter in discrete time*. Automat, Remote control 39, 359-367;

[11] Ge L., Chen H. Y., Han S., Rizos C. (2000): *Adaptive filtering of continuous GPS results*. Journal of Geodesy, 572-580;

[12] Hofmann-Wellenhof, B. Lictenegger, H. Collins J. (1997) – *GPS Theory and Practice*. New York, Springer-Verlag.

[13] Huber J. P. (1981) *Robust statistics*, Wiley, New York;

[14] Huber J. P., Ronchetti M. E., Rousseeuw P. J., Stahel W. A. (1986) *Robust statistics: the approach based on influence functions*, Wiley, New York;

[15] Kalman R.E. (1960): *A new approach to linear filtering and prediction problems.* Trans. ASME J. Basic Engr. 35-45 ;

[16] Kanani, E. (2000): Robust Estimators for Geodetic Transformations and GIS. Dissertation Nr. 13521, IGP- Mitteilungen Nr. 70;

[17] Kremer E. (1994) *Robust Credibility via Robust Kalman Filtering,* ASTIN Bulletin, vol 24, No.2

[18] Kubik K., Wang Y. J. (1991): *Comparison of Different Principles for outlier detection.* Australian Journal of Geodesy, Photogrammetry and Surveying 54, 67-80;

[19] Leick, Alfred (1995) - *GPS satellite surveying*. New York, Wiley & Son.

[20] Manzino, A (2002) – "Introduzione all'uso del Filtro di Kalman in Geodesia". *Seminario internazionale: "La Fotogrammetria nell'era inerziale". Pavia, 13-14 giugno 2002.*

[21] Martin R. D. (1979): *Robust estimation for time series autoregressions.* Robustness in Statistics, Academic press, NY, 107-110;

[22] Masreliez C. J., Martin R. D. (1977): *Robust Bayesian estimation for the linear model and robustifying the Kalman filter.* IEEE Trans. Automat. Control AC-22, 361-371;

[23] Meinhold R. J., Singpurwalla N. D (1989) *Robustification of Kalman Ffilter models,* Journal of the American Statistical Association 84, 479-486;

[24] Mohinder S. G., Angus P. A. (2003): *Kalman Filtering. Theory and practice using Matlab*. Second Edition John Wiley & Sons. ISBN 0-471-39254-5;

[25] Morrison N. (1969): *Introduction to sequential smoothing and prediction*. Mc Graw Hill New York;

[26] Pena D., Guttman J. (1989): *Optimal collapsing of mixture distributions in robust recursive estimation.* Comm. Statist. Theory Methods 18, 817-833;

[27] Remondi W. B., Brown G. (2000):*Triple differencing with Kalman Filtering: Making it work.* GPS Solution Volume 3, number 3 pagg. 58-64;

[28] Schaffrin B. (1990): *A Least –Squares Approach to Robust Kalman filtering.* IAG proc. 107: kinematc System in Geodesy, Surveying and Remote Sensing, Springer Verlag, NY.

[29] Strang G., Borre K. (1997): *Linear Algebra, Geodesy and GPS.* Wellesly-Cambridge Press.

[30] Teunissen P. J. G., Salzman M. A. (1988): *Performance analysis of Kalman Filter.* Report 88.2. Geodetic Computing centre, Delft University of Technology;

[31] Tsai C., Kurz L. (1983): *An Adaptive Robustizing Approach to Kalman filtering.* Automatica 19, 279-288;

[32] Wang Y. J., Kubik K. K. (1993) *Robust Kalman filter and its geodetic applications*, Manuscripta Geodetica, No. 18, 349-354;

[33] Wübbena G., Willgalis S. (2001): *State space approach for precise real time positioning in GPS reference networks*, Proceedings of International Symposium on Kinematic Systems in Geodesy , Geomatics and Navigation, Banf f , Canada, 5-8.