J Geod
DOI 10.1007/s00190-013-0690-8

ORIGINAL ARTICLE

Robust Kalman filtering based on Mahalanobis distance as outlier judging criterion

Guobin Chang

Received: 23 March 2013 / Accepted: 30 December 2013

Abstract A robust Kalman filter scheme is proposed to resist the influence of the outliers in the observations. Two kinds of observation error are studied, i.e., the outliers in the actual observations and the heavy-tailed distribution of the observation noise. Either of the two kinds of errors can seriously degrade the performance of the standard Kalman filter. In the proposed method, a judging index is defined as the square of the Mahalanobis distance from the observation to its prediction. By assuming that the observation is Gaussian distributed with the mean and covariance being the observation prediction and its associate covariance, the judging index should be Chi-square distributed with the dimension of the observation vector as the degree of freedom. Hypothesis test is performed to the actual observation by treating the above Gaussian distribution as the null hypothesis and the judging index as the test statistic. If the null hypothesis should be rejected, it is concluded that outliers exist in the observations. In the presence of outliers scaling factors can be introduced to rescale the covariance of the observation noise or of the innovation vector, both resulting in a decreased filter gain. And the scaling factors can be solved using the Newton's iterative method or in an analytical manner. The harmful influence of either of the two kinds of errors can be effectively resisted in the proposed method, so robustness can be achieved. Moreover, as the number of iterations needed in the iterative method may be rather large, the analytically calculated scaling factor should be preferred.

G. Chang (✉)
Tianjin Institute of Hydrographic Surveying and Charting, Tianjin, China
e-mail: guobinchang@hotmail.com

G. Chang
Naval University of Engineering, Wuhan, China

Keywords Kalman filter · Robust · Hypothesis test · Mahalanobis distance

1 Introduction

In the field of optimal estimation, the parameters and/or states are considered as random variables. In general, the statistical information of a random variable is completely captured by the probability distribution function or the density function. However, in most practical applications it is almost impossible and also unnecessary to know the exact distribution or density function. The first two moments of a random variable, i.e., the mean and covariance are often of more interest and relatively easier to get. The Kalman filter (KF) is the most celebrated real-time linear estimator of the mean and covariance, and has found extremely broad field of application (Kalman 1960). Although there were no assumptions of Gaussian-distributed process or measurement noises in the seminal paper (Kalman 1960), it can be proven that only when the above assumptions hold, KF is optimal, i.e., that it has minimal mean squared error, and is unbiased and consistent (Simon 2006, p 130). Gaussian distribution, due to its simplicity, i.e., its density function can be completely determined by the mean and covariance, and its several elegant properties, namely the Gauss–Markov theorem and central limit theorem (Huber 1972), is assumed in a great deal of estimating methods, including the classic least-square estimator and the KF. However, in practical application the Gaussian distribution may not necessarily hold. Actually the Gaussianity of a sample can be checked using its QQ-plot. For a sample, say $X_i, i = 1, 2, \ldots, N$, let its pth sample quantile be $\xi_{X,p}$ which can be calculated using the sample's order statistics, or its empirical distributions. Let the pth quantile of an exact standard Gaussian distribution be $\xi_{Z,p}$. If the

Published online: 11 January 2014

sample is Gaussian distributed even with unknown mean μ and variance σ^2, we have $\xi_{X,p} \approx \sigma\xi_{Z,p} + \mu$, note that we use "$\approx$" rather than "=" because $\xi_{X,p}$ is an approximated one even though $\xi_{Z,p}$ is exact. The plot of $\xi_{X,p}$ versus $\xi_{Z,p}$ is called the QQ-plot whose (approximate) linearity indicates the Gaussianity of the sample. In Hewer et al. (1987), the curvature of the normal QQ-plot of the glint noise in the angle tracking signal of mono-pulse radars indicates the heavy-tailed non-Gaussian distribution. Unfortunately, as a generalization of recursive weighted least-square estimator, KF fails to perform adequately when the noise follows a non-Gaussian distribution; furthermore the KF is also susceptible to outliers. To address the non-Gaussian noises, particle filter, which can be derived directly from the Bayesian estimator without adopting the linear framework as in KF, can be employed (Gordon et al. 1993). However, the shortcoming of particle filter used to address the problem discussed here is threefold. Firstly, the computational load is rather heavy, especially in the high-dimensional case, as the probability density function must be approximated by plenty of particles. Secondly, the distribution, though unnecessarily being Gaussian, should be exactly known. Thirdly, it cannot resist the effect of outliers.

Both the non-Gaussian distribution and outliers can be dealt with using a kind of method with the name as robust filter or estimator. The robustness has multiple meanings, and the *qualitative robustness* defined in (Hampel 1971) is rather intuitive, i.e., that small changes in the data should produce only small changes in the estimate, where "small" changes involve both large changes in a small fraction of the data and small changes in all the data. Besides the robustness, the statistical efficiency of an estimator is also a natural requirement implying that it should be as nearly optimal as possible when the nominal conditions hold. There are many forms of robust KF in the literature. The most straightforward method is the data-censoring strategy, i.e., the observations whose residuals are "too large" are rejected. However, this simple strategy may be very inefficient in some cases and will lead to a lack of continuity in the estimates of the covariance (Schick and Mitter 1994). Median-based filters can also be employed to make KF robust, e.g., in Chang and Wu (2000), this kind of filter may be highly robust, but not efficient, and it is implemented in a block-based batch-preprocessing manner which quite limits its real-time applications. By minimizing the estimation error in the worst case, H-infinity filter can be used to address the uncertainties in the observation noise (Duan et al. 2006). However, it breaks down in the presence of outliers (Gandhi and Mili 2010). In Boncelet and Dickinson (1983) and Kovacevic et al. (1992), the KF is constructed as a linear regression problem which is solved robustly by a probabilistic approach called M-estimation (Huber 1972). As a generalization of KF, the Bayesian estimator is robustified also using the M-estimation in the field of geodesy (Yang 1991). As a direct application of (Yang 1991), the M-estimation-based KF for rank deficient observation model was proposed in (Koch and Yang 1998). A robust and adaptive KF is proposed in (Yang et al. 2001) and applied to estimate the satellite clock offset in (Huang and Zhang 2012). In this method, at any epoch, an initial state estimate is firstly calculated using only the current observation through the robust M-estimation, so the robustness is ensured. Then this initial state estimate is used to modify the a priori estimate of the covariance, so the adaptiveness is also ensured. However, in this method the observation equation must be over-determined, or more exactly, excluding the observations which have outliers in them, the number of the remaining observations should be no less than the dimension of the state. So this method with both adaptiveness and robustness can only be applied in limited cases. The M-estimation-based KF is extended to the derivative-free nonlinear filters through the statistical linear regression perspective of the nonlinear transformation (Karlgaard et al. 2007; Wang et al. 2010). But these nonlinear robust KFs achieve robustness at the cost of reducing the accuracy of the nonlinear transformation itself, and are modified in (Chang et al. 2012a, 2013). The Huber-based unscented KF is generalized to cope with the case that outliers in both the prediction and observation co-exist simultaneously (Chang et al. 2012b). Essentially the M-estimation-based filters may be equivalent in some sense to the method of observation-noise-inflating or observation-trimming (Chang et al. 2012a, 2013), but with relatively computational complexities. Two heuristic methods can be employed to make the filter robust, one of which inflates the covariance of the observation noise, see, e.g., Hajiyev and Soken (2012), Soken and Hajiyev (2010), Yang et al. (2002b), while the other re-scale the filter gain, see, e.g., Kim et al. (2008, 2009). As can be seen in the following of this paper, two methods are essentially equivalent, both of which are simple but effective.

In this paper, two cases are considered, i.e., that outliers are artificially introduced into the actual observations at some isolate epochs and that the nominal Gaussian distribution of the observation noise is contaminated by a small fraction of another zero-mean Gaussian distribution with larger covariance (as to make the resulted non-Gaussian distribution heavy-tailed). These two cases represent the above-mentioned two kinds of "small changes", respectively. If the linear system with Gaussian-distributed white noise is correctly modeled, the observation vector should be Gaussian distributed with mean and covariance being its prediction and the associate covariance. and the square of the Mahalanobis distance from the observation to its prediction, which is defined as the judging index in this contribution as to detect outliers, should be Chi-square distributed with the dimension of the observation vector as the degree of freedom. At any epoch, hypothesis test is performed with the nominal

assumption being the null hypothesis and the judging index as the testing statistic. Replace the observation vector with its realization, i.e., the actual observation in the judging index, we get the actual judging index. Calculate the p value which is the cumulative probability of the Chi-square distributed random value being larger than the actual judging index. If the p value is smaller than a predetermined significance level, a small value, say α, it is implied that the innovation vector is too large to emerge probably, then we can conclude with the probability $1-\alpha$ that some errors in the actual observations exist. After the errors are detected, two scalar factors can be used to adjust the covariance of the observation noise and the innovation vector, respectively, both resulting in a decreased filter gain, so the badly obtained observation can be less weighted and hence robustness against the outliers in the observation can be achieved. By forcing the actual judging index to be the α-quantile of the Chi-square distribution, one of the scaling factor can be calculated by solving a nonlinear equation using Newton's method (Dennis and Schnabel 1987, p 86) while the other can be calculated analytically.

The remaining of the paper is organized as follows. In Sect. 2, KF, the optimal estimator for a linear system with Gaussian noises, is introduced emphasizing the detection of outliers using the Mahalanobis distance. In Sect. 3, a scaling factor is introduced to inflate the covariance matrix of the observation noise or the filter gain so as to make the filter robust, and by forcing the square of the Mahalanobis distance to be the α-quantile, the scaling factor is calculated. In Sect. 4, the problem of kinematic positioning with assumed constant velocity (CV) model is studied, and two kinds of modeling error, i.e., outliers and contamination of the assumed Gaussian distribution of the observation noise, are introduced to the CV model, and the better performance of the proposed method is validated. The conclusion is given in Sect. 5.

Below, the symbols "^" and "∼" above a variables represent an estimate and a measurement; the superscripts "−" and "+" represent the a priori and a posteriori estimates, respectively; $N(x;\mu,P)$ denotes that the variable x obeys Gaussian distribution with mean μ and covariance P.

2 Kalman filter

Consider the discrete-time linear stochastic state space model with Gaussian-distributed process and measurement noise (Simon 2006, p 124),

$$x_k = F_{k-1}x_{k-1} + w_{k-1} \tag{1}$$
$$y_k = H_k x_k + v_k \tag{2}$$

where x_k is the n-dimensional state vector at epoch k and is to be estimated. The symbol y_k is the m-dimensional observation vector which can be measured at epoch k. The symbols F_{k-1}, H_k are the $n\times n$ dimensional propagation matrix and the $m\times n$ dimensional observational matrix, respectively. And w_{k-1}, v_k are process and observation noises, respectively. Without loss of generality, the two noises are assumed zero-mean uncorrelated Gaussian white noises satisfying $E[w_k w_j^{\mathrm{T}}] = Q_k\delta_{kj}$, $E[v_k v_j^{\mathrm{T}}] = R_k\delta_{kj}$, $E[w_k v_j^{\mathrm{T}}] = 0$, where Q_k and R_k are corresponding covariances, and δ_{kj} is the Kronecker delta function whose value is one when $k=j$, and zero otherwise. The initial state estimate is assumed to be Gaussian with mean $\hat{x}_0^+$ and covariance P_0^+, and uncorrelated with any process and measurement noises.

Systems (1) and (2) are solved by KF to give the a posteriori estimate of x_k, i.e., $\hat{x}_k^+$ and its covariance P_k^+ given the observations $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_k$. The kth Kalman filter equations are as follows,

$$\hat{x}_k^- = F_{k-1}\hat{x}_{k-1}^+ \tag{3}$$
$$P_k^- = F_{k-1}P_{k-1}^+F_{k-1}^{\mathrm{T}} + Q_{k-1} \tag{4}$$
$$\hat{y}_k^- = H_k\hat{x}_k^- \tag{5}$$
$$P_{\hat{y}_k^-} = H_k P_k^- H_k^{\mathrm{T}} + R_k \tag{6}$$
$$K_k = P_k^- H_k^T \left(P_{\hat{y}_k^-}\right)^{-1} \tag{7}$$
$$\hat{x}_k^+ = \hat{x}_k^- + K_k\left(\tilde{y}_k - \hat{y}_k^-\right) \tag{8}$$
$$P_k^+ = P_k^- - K_k H_k P_k^- \tag{9}$$

where K_k is the Kalman filter gain at epoch k, and $\hat{y}_k^-$ is the observation prediction vector stated above with $P_{\hat{y}_k^-}$ as its associate covariance matrix. Note that $n_k = \tilde{y}_k - \hat{y}_k^-$ in (8) is often called the innovation vector.

If the assumptions about (1) and (2) hold, the observation should be Gaussian distributed with mean and covariance being $\hat{y}_k^-$ and $P_{\hat{y}_k^-}$, respectively, i.e., with the probability density function

$$\rho(y_k) = N(y_k;\hat{y}_k^-, P_{\hat{y}_k^-}) = \frac{\exp\left(-\frac{1}{2}\left(y_k-\hat{y}_k^-\right)^{\mathrm{T}}\left(P_{\hat{y}_k^-}\right)^{-1}\left(y_k-\hat{y}_k^-\right)\right)}{\sqrt{(2\pi)^m\left|P_{\hat{y}_k^-}\right|}} \tag{10}$$

where $|P_{\hat{y}_k^-}|$ is the determinant of $P_{\hat{y}_k^-}$. However, if there is some violation to the assumption about (1) and/or (2), e.g., that there are some outliers in the observation or that the Gaussian distribution of the observation noise is contaminated with some other distributions, (10) will no longer hold. On the other hand, if (10) does not hold, some violation to the assumption or some kind of modeling errors can be thought to exist. Specifically hypotheses test can be carried out to detect the modeling errors as follows.

Actually the purpose of performing the hypothesis test is to check if the actual observation is compatible with the assumed model, or in other word, the null hypothesis. Firstly, we take (10) as the relevant null distribution which does hold under the assumed model. Secondly, twice the minus exponent in (10) is treated as the relevant test statistic. This test statistic is used as the judging index to detect the modeling errors. Actually the test statistic or the judging index is just the square of the Mahalanobis distance from observation y_k to its mean $\hat{y}_k^-$ (Izenman 2008, p 60), i.e.,

$$\gamma_k = M_k^2 = \left(\sqrt{\left(y_k - \hat{y}_k^-\right)^{\mathrm{T}} \left(P_{\hat{y}_k^-}\right)^{-1} \left(y_k - \hat{y}_k^-\right)}\right)^2 = \left(y_k - \hat{y}_k^-\right)^{\mathrm{T}} \left(P_{\hat{y}_k^-}\right)^{-1} \left(y_k - \hat{y}_k^-\right) \tag{11}$$

where $M_k = \sqrt{(y_k - \hat{y}_k^-)^{\mathrm{T}}(P_{\hat{y}_k^-})^{-1}(y_k - \hat{y}_k^-)}$ is the Mahalanobis distance. Thirdly, the distribution of the test statistic under the null hypotheses is decided, or specifically, assuming the null hypotheses is true, γ_k should be Chi-square distributed with degree of freedom m. Fourthly, a significance level α, a probability threshold below which the null hypothesis will be rejected, is selected. α is a small value, in this contribution 1 % is adopted. Fifthly, substitute the actual observation $\tilde{y}_k$ to replace y_k in γ_k and get the actual judging index $\tilde{\gamma}_k$, in other word a realization of the judging index, determine the probability of γ_k being larger than $\tilde{\gamma}_k$ under the Chi-distribution, i.e., the p value. If the p value is smaller than α, the null hypotheses should be rejected and it is concluded that there exist some kind of violations to the basic assumptions, that is to say outliers are detected. An alternative process to the fifth step can also be used: instead of the p value, the α-quantile χ_α of the Chi-square distribution is predetermined, e.g., that for the 2-degree-of-freedom Chi-square distribution with the significance level being 1 %, it is 9.2. Then we have

$$\Pr\left(\gamma_k > \chi_\alpha\right) = \alpha \tag{12}$$

where $\Pr(\cdot)$ represents the probability of a random event, i.e., that the probability of γ_k being larger than χ_α should be very small, say α. So if the actual $\tilde{\gamma}_k$ is larger than this α-quantile, the null hypotheses can be rejected and again it can be concluded that there exist some kind of violations to the basic assumptions. This alternative process bears the merit that a less computational effort is needed as the α-quantile (so long as the dimension of the observation at any epoch stays the same which is always the case in practice) can be predetermined and need not to be calculated online, moreover, the α-quantile in this alternative process can be directly used in the calculation of the scaling factor as shown in the following section.

3 A novel robust Kalman filter

In this paper a robust KF with a scaling factor is introduced to address the modeling errors. If the actual judging index $\tilde{\gamma}_k$ is not greater than χ_α, the standard KF is carried out according to (3)–(9), otherwise a scaling factor λ_k is introduced to inflate the observation noise covariance R_k

$$\bar{R}_k = \lambda_k R_k \tag{13}$$

So the following equation can be satisfied

$$\bar{\gamma}_k = n_k^{\mathrm{T}} \left(\bar{P}_{\hat{y}_k^-}\right)^{-1} n_k = n_k^{\mathrm{T}} \left(H_k P_k^- H_k^{\mathrm{T}} + \bar{R}_k\right)^{-1} n_k = \chi_\alpha \tag{14}$$

i.e.,

$$f(\lambda_k) = n_k^{\mathrm{T}} \left(H_k P_k^- H_k^{\mathrm{T}} + \bar{R}_k\right)^{-1} n_k - \chi_\alpha = 0 \tag{15}$$

Equation (15) is nonlinear in λ_k, even when n_k is a scalar. The λ_k can be solved iteratively using Newton's method (Dennis and Schnabel 1987, p 86).

$$\lambda_k(i+1) = \lambda_k(i) - \frac{f\left(\lambda_k(i)\right)}{f'(\lambda_k(i))} \tag{16}$$

where i represents the ith iteration. In (16), the derivative of a quadratic form or the derivative of an inverse matrix should be carried out. As in Simon (2006, p 14), it gives

$$\frac{\mathrm{d}}{\mathrm{d}t}(A^{-1}) = -A^{-1}\frac{\mathrm{d}A}{\mathrm{d}t}A^{-1} \tag{17}$$

where A is an invertible matrix as a function of t. So substituting (17) into (16) gives

$$\lambda_k(i+1) = \lambda_k(i) + \frac{\bar{\gamma}_k(i) - \chi_\alpha}{n_k^T(\bar{P}_{\hat{y}_k^-}(i))^{-1} R_k (\bar{P}_{\hat{y}_k^-}(i))^{-1} n_k} \tag{18}$$

The initial value is set to be 1, i.e., $\lambda_k(0) = 1$, and the iterations terminate when the judging index $\bar{\gamma}_k(i) \le \chi_\alpha$.

The implementation of the proposed adaptive method is given step by step in Table 1.

Remark 1 As shown in the numerical simulations in the next section, at some epochs many iterations may be needed for the above method to converge, sometimes more than 10. This rather low convergence rate shows the bad computational efficiency, and this will be problematic especially in the case of real-time applications. Fortunately, as can be seen in the forthcoming remark, we can construct an even more efficient alternative where an analytical solution exists and no iteration is needed.

Table 1 Robust Kalman filter

Initiation: initiate the filter with $\hat{x}_0^+$, P_0^+;

Filtering: For k=1, 2, ⋯

1 State prediction through (3) and (4);

2 Innovation vector calculating through (5) and (6);

3 $\lambda_k(0)=1$ and calculating the criterion index $\bar{\gamma}_k$ through (14);

4 State update

4.1 If $\bar{\gamma}_k \le \chi_\alpha$:

Standard Kalman filter through (7), (8), and (9);
If the filter does not terminate:
k → k+1, and go to 1;
Else:
Go to 5

4.2 Else:

Update λ_k through (18);

Update $\bar{R}_k$ through (13);

Update $\bar{\gamma}_k$ through (14);

Go to 4;

5 End of the filter.

Remark 2 From (6), an inflated covariance of the observation noise results in an inflated covariance of the observation prediction. Other than (13), a scalar factor, used to adjust the covariance of the observation prediction, can also be introduced to ensure the robustness, i.e.,

$$\bar{P}_{\hat{y}_k^-} = \kappa_k P_{\hat{y}_k^-} \tag{19}$$

and κ_k can be calculated analytically rather than iteratively as in (18), i.e.,

$$\kappa_k = \begin{cases} 1 & \text{if} \quad \tilde{\gamma}_k \le \chi_\alpha \\ \frac{\tilde{\gamma}_k}{\chi_\alpha} & \text{else} \end{cases} \tag{20}$$

Both λ_k and κ_k can be adopted in the correction stage of KF as in (5)–(9). Besides (9), the a posteriori estimate of the covariance can also be calculated as

$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^{\mathrm{T}} + K_k R_k K_k^{\mathrm{T}} \tag{21}$$

As pointed out in Simon (2006, p 129), Eq. (21), named the Joseph stabilized version of the covariance correction equation, is more numerically stable than (9), because it guarantees the symmetry and positive definiteness of P_k^+ as long as P_k^- is symmetric and positive definite. Being a covariance of a random vector, P_k^+ itself should be symmetric and positive definite ideally. And in some filtering algorithms, symmetry and positive definiteness are required, as the square root of the covariance matrix is required, e.g., in unscented KF (Julier et al. 2000). So (21) should be preferred to. However, in (21) R_k is explicitly used which is not available explicitly in (19), this can be obtained through

$$\bar{R}_k = \bar{P}_{\hat{y}_k^-} - H_k P_k^- H_k^{\mathrm{T}} = (\kappa_k - 1) H_k P_k^- H_k^{\mathrm{T}} + \kappa_k R_k \tag{22}$$

Remark 3 From (7) an inflated $P_{\hat{y}_k^-}$ results in a decreased filter gain, so robustness can also be achieved by rescaling the filter gain with a scalar factor smaller than one. Actually the inverse of the factor κ_k in Remark 1 can directly used to rescale the filter gain. In Kim et al. (2008, 2009), an adaptive fading Kalman filter with rescaling filter gain is proposed to address the error in the observation equation. Although with the name "adaptive", the filter-gain-rescaling scheme is essentially a robust method.

Remark 4 In the proposed method, if the actual judging index $\tilde{\gamma}_k$, i.e., the square of the Mahalanobis distance from actual observation to predicted one, is greater than χ_α, modeling errors exist probably. When the scaling factor is 1, it boils down to the standard KF. As the scaling factor increases (greater than 1), the covariance of the observation prediction is enlarged, then the Kalman filter gain will decrease, which means a decreased contribution of the observation in update of filter. By employing the factor λ_k or κ_k, the harmful effect of the modeling error in the observation is restrained. The proposed method has the merit of being insensitive to the modeling errors as outliers in the actual observation and the contamination in the nominal Gaussian distribution of the observation noise.

Remark 5 In this study, any abnormally large innovation vector is assumed to be due to the errors in the observation. However, from (5), it can be seen that besides errors in $\tilde{y}_k$, errors in $\hat{y}_k^-$ can also lead to an abnormally large n_k. Errors in $\hat{y}_k^-$ implies that the observation prediction is incorrectly obtained, i.e., there are some modeling errors in the process model. These modeling errors in process model can be addressed by another methodology with "adaptive filter" as its name (Kim et al. 2008, 2009). In adaptive filters a factor is introduced to rescale the covariance of the a priori state estimate P_k^- other than to rescale R_k in the robust filters. Adaptive filters have the merits of being insensitive to the errors in process models, or in other word, having a strong tracking ability to the irregularly varied states (Zhou and Frank 1996). Similar to robust filters, the merit of adaptive filters hold only when some prerequisites are met, i.e., the actual observation $\tilde{y}_k$ is correctly obtained. Actually the adaptive method can also be regarded as a special case of the robust method in a broad sense, as it has the robustness against errors in the state process models.

Remark 6 The robust Bayesian estimator developed in Yang (1991) actually gives a unified structure to get either adaptiveness or robustness, because this filter can modify not only R_k but also P_k^-. If $\hat{x}_k^-$ is assumed correct and taken as the initial value of the iteration, R_k will be modified, and the robustness can be assured, similar to the approach in this paper. On the other hand, if observation is assumed to be correct, and the state estimate calculated from only the current observation is taken as initial value, P_k^- will be modified, and adaptiveness will be assured. However, just as stated in Yang (1991), it cannot cope with the errors simultaneously co-existing in both the observation and state prediction. And this estimator cannot decide by itself whether state prediction or observation is correct, so it can be either adaptive or robust, but cannot be both adaptive and robust. Similarly the unified structure in Kim et al. (2008, 2009) can also be either adaptive (the P_k^--rescaling form) or robust (the K_k-rescaling form), and again this structure itself cannot make a choice between the adaptive and the robust. The method proposed by Yang et al. (2001) is essentially a kind of filter with both adaptiveness and robustness, but under special conditions, and this adaptive and robust method is currently employed in Huang and Zhang (2012) to estimate the satellite clock offset in satellite positioning. In their method, at any instance, an initial state estimate is firstly calculated using only the current observation through robust least square method, so the robustness is assured. Then this initial state estimate is used to modify P_k^-, so the adaptiveness is obtained. However, in this method the observation equation must be over-determined, or more exactly, with excluding the observations which have outliers in them, the number of the remaining observations should be not less than the dimension of the state. So this method with both adaptiveness and robustness can only be applied in limited cases. The filter with both adaptiveness and robustness for the general case is still an open problem and beyond the scope of this study.

Remark 7 The proposed method, or all the robust filters, achieves the robustness at the cost of decreasing the efficiency, i.e., it may mistake a good $\tilde{y}_k$ as erroneous one. Just as mentioned above, even when the innovation vector is relatively large, i.e., the actual judging index is greater than χ_α, it may still be a Gaussian-distributed random multivariable but with a much low probability, i.e., not greater than α. So the efficiency of the proposed method is only slightly affected if α is assigned a quite small value, say 1 %, it might as well be said that our method has the efficiency of $1-\alpha$, say 99 %. Moreover, this efficiency can be adjusted through selecting the significance level according to the specific conditions.

Remark 8 In Hajiyev and Soken (2012), to make the filter robust against errors in the observation, the observational errors are detected according to the inequality $\mathrm{trace}(n_k n_k^{\mathrm{T}}) \geq \mathrm{trace}(P_{\hat{y}_k^-})$, i.e., Eq. (20) in their paper. As $\mathrm{trace}(n_k n_k^{\mathrm{T}}) = (\|n_k\|)^2$ and $\mathrm{trace}(P_{\hat{y}_k^-}) = \sum_{i=1}^{m} P_{\hat{y}_k^-}(i,i)$, $P_{n_k}(i,i)$ is the ith diagonal elements of $P_{\hat{y}_k^-}$, it gives $(\|n_k\|)^2 \geq \sum_{i=1}^{m} P_{\hat{y}_k^-}(i,i)$. Compared to the proposed method in this contribution, this error-detecting strategy has two shortcomings: firstly, the off-diagonal elements of $P_{\hat{y}_k^-}$ are neglected, secondly, the theoretical foundation is not clear and the significance level cannot be adjusted reasonably.

Remark 9 The judging index γ_k and the factor λ_k in the proposed method are calculated based on the information of the current epoch only. In many other papers, the factors are calculated based on both the current information and the historical information, and the windowing method is employed to select the historical information used. Similar to the case of adaptive filters in Yang and Gao (2007), our method has

the merit of lower computational load and being more sensitive in reflecting the observational errors in the present epoch.

Remark 10 Frankly speaking, the proposed method, as well as many other robust filters with a single scaling factor, may bear a shortcoming that if certain individual observation is contaminated by outlier, the whole contribution of the observation vector is unreasonably decreased including the contribution of good observations. However, this shortcoming may be compensated through sequential KF. In sequential KF, instead of processing the measurement as a vector, the KF observation-update equation is implemented to account for only one element of the observation vector at a time, after all the m elements are processed the a posteriori estimate is obtained and fed to the prediction at the next epoch. As the observations become scalars in the sequential update, the Chi-square hypothesis is exactly equivalent to the Gaussian hypothesis, moreover Eq. (15), though nonlinear, can be solved analytically. It is noted that the sequential KF can be implemented only when the elements of the observation are un-correlated with each other, i.e., the covariance matrix of the observation noise should be diagonal. If it is not the case, we can firstly de-correlate the observation vector through Choleski decomposition then apply the sequential update. Another way to address the correlated observation is to apply the bifactor method proposed in Yang et al. (2002a), specifically in the sequential update, if one element of the observation is detected to be outlier, not only the variance of this observation element should be inflated, but also the covariances between it and the other remaining observation elements should be inflated.

4 Simulation

The kinematic positioning problem with constant velocity model is employed (Yang et al. 2001). The kinematic model along the north axis in continuous time domain is as follows,

$$\dot{x}(t) = \begin{bmatrix} \dot{p}_N(t) \\ \dot{v}_N(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_N(t) \\ v_N(t) \end{bmatrix} + \begin{bmatrix} 0 \\ a_N(t) \end{bmatrix} = Ax(t) + Gw(t) \tag{23}$$

where $p_N(t)$ and $v_N(t)$, which are the north position and north velocity at epoch t, respectively, are to be estimated in real time. The north acceleration $a_N(t)$ is modeled as random noise.

Equation (23) can be converted into discrete-time form as (Bekir 2007, p 186)

$$x_k = Fx_{k-1} + w_k \tag{24}$$

where $x_k = x(k\Delta t)$ with Δt as the sampling interval.

$$F = \exp(A\Delta t) = I + A\Delta t \tag{25}$$

$$Q_k = E(w_k w_k^{\mathrm{T}}) \approx \exp\left(\frac{1}{2}A\Delta t\right) GQG^{\mathrm{T}} \exp\left(\frac{1}{2}A^{\mathrm{T}}\Delta t\right) \tag{26}$$

with $Q = E(w(t)(w(t))^{\mathrm{T}})$.

The positions along three axes are observed, so the observation equation is

$$y_k = Hx_k + \eta_k \tag{27}$$

where H is the observational matrix, and η_k is the observational noise.

The accelerations in the state process equation are considered as Gaussian-distributed white noise with nominal standard deviation of 0.01 m/s^2, and the observational noise is assumed to be Gaussian-distributed white noise with standard deviation of 0.1 m. and these parameters along with the state process model (24) and observation model (27) are employed to carry out KF and robust KF (RKF). Note that the RKF can be carried out using either the iterative scheme shown in Table 1 or the analytical method presented in Remark 2. The filtering estimate of the both schemes are almost the same, so in the following the estimation errors of the RKF are illustrated without telling which scheme gives these errors.

Two different cases are studied:

Case 1: outlier case: positioning errors with the value of 2 m are artificially introduced into (27) every 30 min. The position estimation errors of KF and RKF are shown in Fig. 1, and the velocity estimation errors are shown in Fig. 2. The numbers of iterations needed in this case are illustrated in Fig. 3.

Case 2: contaminated Gaussian-distribution case: the nominal Gaussian distribution of the observation noise is perturbed by another distribution (Maronna et al. 2006, p 19), i.e., the actual probability density function is

$$\rho_{\text{actual}} = (1-\varepsilon)\rho_{\text{nominal}} + \varepsilon\rho_{\text{perturbing}} \tag{28}$$

where the ratio of the perturbing distribution $0 \le \varepsilon \le 0.1$.

This distribution is called contaminated Gaussian distribution. The perturbing distribution $\rho_{\text{perturbing}}$ can be any distribution and always a symmetric one. If $\rho_{\text{perturbing}}$ is also a Gaussian one but with a larger standard deviation, the actual distribution is also called a Gaussian mixture. In this study, the standard deviation of the perturbing distribution is assumed four times that of the nominal distribution, and the ratio of the perturbing distribution is firstly assumed 0.1, and

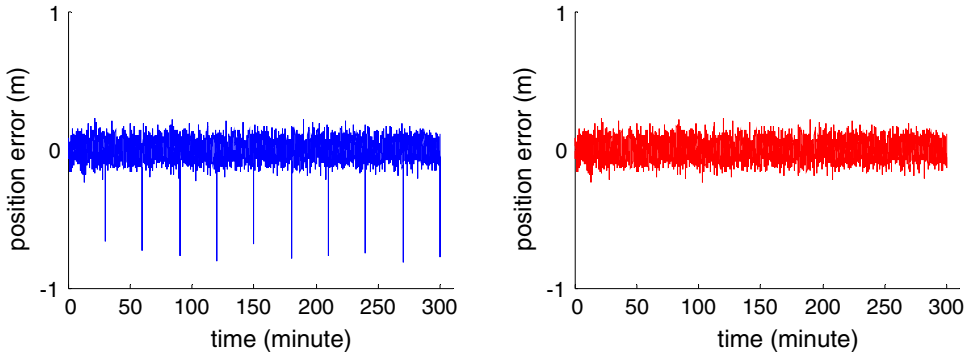

Fig. 1 Position estimation errors when outliers exist in actual observations, *left* KF, *right* RKF

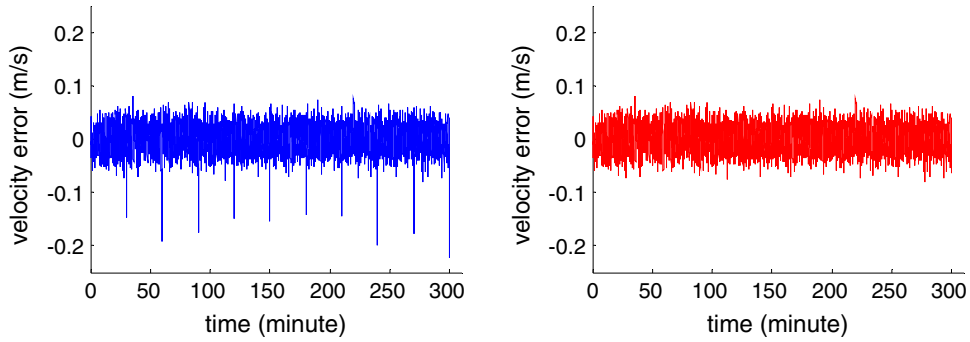

Fig. 2 Velocity estimation errors when outliers exist in actual observations, *left* KF, *right* RKF

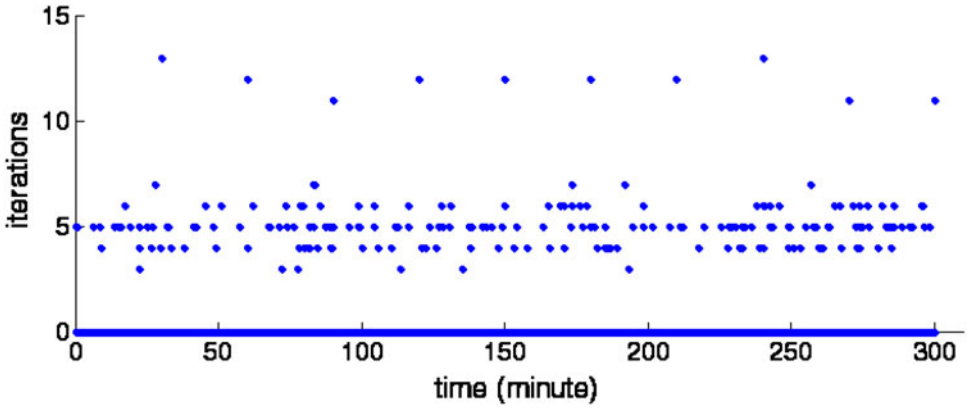

Fig. 3 Number of iterations needed for the iterative method when outliers exist in actual observations

the position estimation errors of KF and RKF are shown in Fig. 5 and the velocity estimation error in Fig. 5. The numbers of iterations needed for the iterative methods in this case are illustrated in Fig. 6.

In case 2, besides the estimation error, the root mean square (RMS) of the estimation errors is also calculated to evaluate the performance of different filters,

$$\mathrm{RMS}_{\mathrm{filter}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\hat{x}_k^+ - x_{\mathrm{true}}\right)^2} \tag{29}$$

where the subscript "filter" represent KF and RKF, respectively. The ratios of the contamination are assumed with the value of 0, 0.02, 0.04, 0.06, 0.08, and 0.1. And the RMSs of the position and velocity error versus different contaminating ratios are shown in Figs. 7 and 8, respectively. It is noted that the case with 0 ratio is just the model with the nominal parameters mentioned above.

From Figs. 1 and 2, it is apparent that the proposed RKF can give superior performance comparing to the standard KF in both position estimation and velocity estimation when there is, though a small fraction of, outliers in the actual observations. The proposed RKF can effectively resist the influence of the outliers. From Fig. 3, it is founded that though in most epochs, no iteration is needed, there are also many epochs at which 4–6 iterations are needed, and at a few

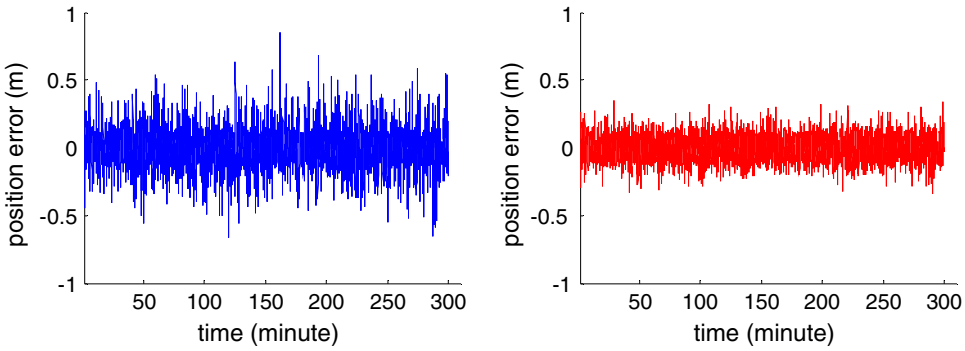

Fig. 4 Position estimation error when the Gaussian distribution of observation noise is contaminated, *left* KF, *right* RKF

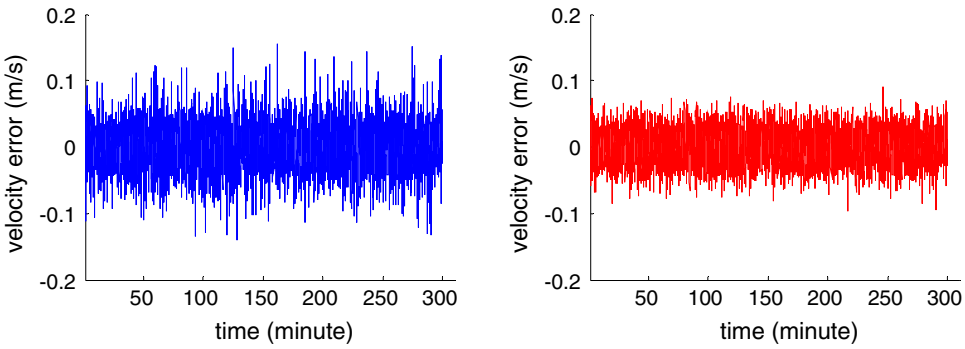

Fig. 5 Velocity estimation error when the Gaussian distribution of observation noise is contaminated, *left* KF, *right* RKF

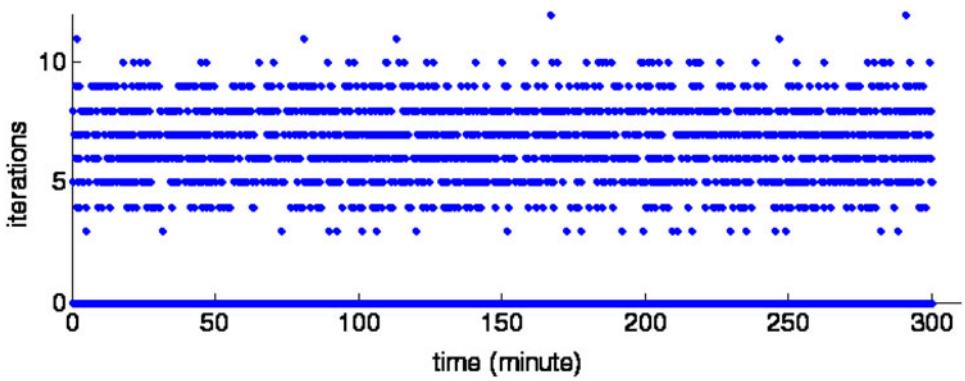

Fig. 6 Number of iterations needed for the iterative method when the distribution of the observation noise is contaminated

epochs, more than 10 iterations are needed (by inspection, it can found that these epochs correspond to the time when the outliers exist). This clearly shows the low computational efficiency of the iterative method. So it is rather apparent here that the analytical method should be preferred to the iterative method. Another phenomenon which can be found in Fig. 3 is noted that even at epochs at which no outliers exist, iterations are also needed. This is due to the statistical efficiency issue of the robust method. Actually, the robustness is achieved at the cost of some statistical efficiency loss, i.e., that, it is possible to mistake good observations as outliers.

From Figs. 4 and 5, it is apparent that the proposed RKF can give much superior performance comparing to the standard KF in both the position and velocity estimation when the nominal Gaussian distribution is perturbed by other distributions, though this perturbing distribution only accounts for 10 %. From Fig. 6, even though at most epochs, no iteration is needed for the iterative method, there are also many epochs (more than that in the first case) at which 4–10 iterations are needed, this again shows the low computational efficiency of the iterative method.

From Figs. 7 and 8, when the perturbing ratio is 0, i.e., that the system is correctly modeled, the RMS of the estimating error of KF is rather low, meaning that KF performs well in this situation. However, when the ratio deviates from 0, i.e., that some modeling errors exist, both the position and velocity estimating accuracies of the standard KF degrade seriously. Actually from simulation, the degradation happens even when the contamination only accounts for 0.01 % which illustrates the lack of robustness of the KF. On the other hand

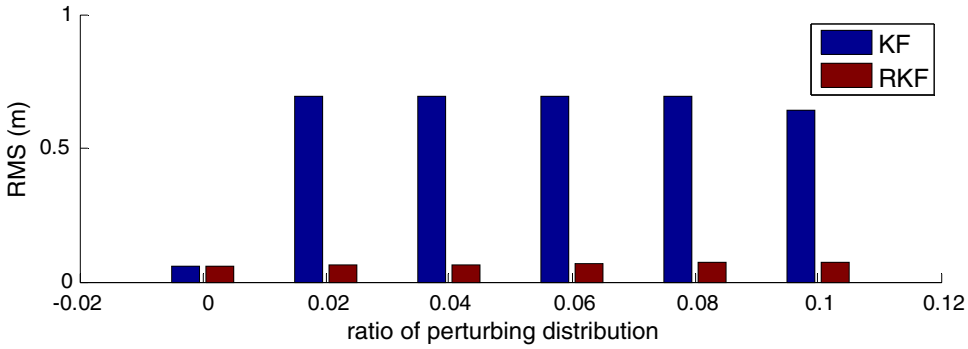

Fig. 7 RMS of position estimation errors versus different contaminating ratios

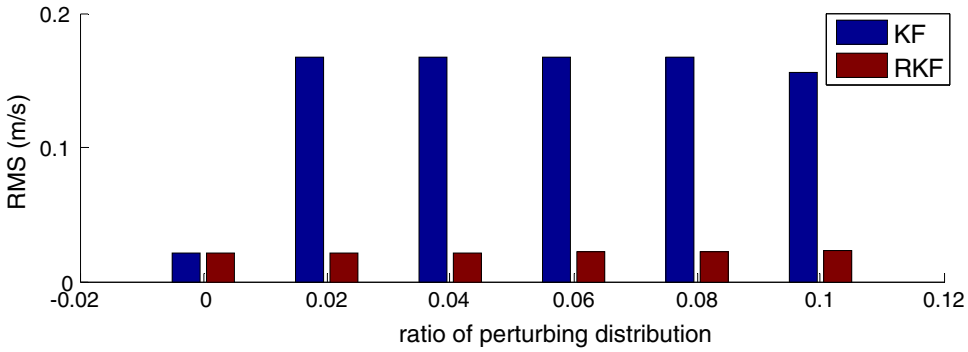

Fig. 8 RMS of velocity estimation errors versus different contaminating ratios

the accuracy of the proposed robust filter is rather stable no matter the nominal distribution is perturbed or not and no matter how much it is perturbed. So it can be concluded that the proposed filter is insensitive to such modeling errors as incorrect noise distribution, i.e., that the robustness of the proposed method is validated.

It can also be concluded from Figs. 7 and 8 that if there are no modeling errors about the noise distribution, all the observations should be good ones. The proposed filter gives an un-degraded performance comparing to the standard filter in this situation implying that the contribution of these good observations is not dropped. This illustrates the high efficiency of the proposed filter.

5 Conclusion

The standard Kalman filter is an optimal estimator only when some prerequisites hold, e.g., with a Gaussian-distributed observation noise with exactly known mean and covariance. However, in many practical applications, the observation noise may be a heavy-tailed one (heavier than the nominal Gaussian distribution), or the actual observation may be susceptible to outliers. The standard Kalman filter, as a generalization of classic least squared estimator, is not robust, or in other word is sensitive to the deviation from the nominal parameters such as heavy-tailed noise distributions or outliers, even when these deviations only account for a much small fraction. So in this paper, a robust version of the Kalman filter is proposed. At any epoch, if the square of the Mahalanobis distance from the innovation to the zero is greater than a predetermined quantile of a Chi-square distribution, implying the innovation is abnormally large, it is concluded that there are some kinds of errors in the observations. Then a scaling factor is introduced to rescale the covariance of the observation noise or of the innovation, resulting in decreased filter gain. So the actual observation is less weighted, and the information from the process model is more weighted. The effect of the heavy-tailed distribution of the observation noise and the outliers in the actual observation is effectively resisted in the proposed robust Kalman filter.

Acknowledgments I am grateful to the associate editor and three anonymous reviewers for their valuable comments. This work was supported by the National Basic Research Program of China (973 Program, No. 2012CB719902) and the National Natural Science Foundation of China (No.41274013).

References

Bekir E (2007) Introduction to modern navigation systems. World Scientific, Beijing

Boncelet CG, Dickinson BW (1983) An approach to robust Kalman filtering. In: The 22nd IEEE conference on decision and control, IEEE, pp 304–305

Chang D, Wu W (2000) Feedback median filter for robust preprocessing of glint noise. IEEE Trans Aerospace Electron Syst 36:1026–1035

Chang L, Hu B, Chang G, Li A (2012a) Huber-based novel robust unscented Kalman filter. IET Sci Meas Technol 6:502–509

Chang L, Hu B, Chang G, Li A (2012b) Multiple outliers suppression derivative-free filter based on unscented transformation. J Guid Control Dyn 35:1902–1907

Chang L, Hu B, Chang G, Li A (2013) Robust derivative-free Kalman filter based on Huber's M-estimation methodology. J Process Control 23:1555–1561

Dennis J, Schnabel R (1987) Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall, New Jersey

Duan Z, Zhang J, Zhang C, Mosca E (2006) Robust H2 and H∞ filtering for uncertain linear systems. Automatica 42:1919–1926

Gandhi MA, Mili L (2010) Robust Kalman filter based on a generalized maximum-likelihood-type estimator. IEEE Trans Signal Process 58:2509–2520

Gordon N, Salmond D, Smith A (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc F Radar Signal Process 140:107–113

Hajiyev C, Soken HE (2012) Robust estimation of UAV dynamics in the presence of measurement faults. J Aerospace Eng 25:80–89

Hampel F (1971) A general qualitative definition of robustness. Ann Math Stat 42:1887–1896

Hewer G, Martin R, Zeh J (1987) Robust preprocessing for Kalman filtering of glint noise. IEEE TransAerospace Electron Syst 23:120–128

Huang G, Zhang Q (2012) Real-time estimation of satellite clock offset using adaptively robust Kalman filter with classified adaptive factors. GPS Solutions 16:531–539

Huber P (1972) Robust statistics: a review. Ann Math Stat 43:1041–1067

Izenman AJ (2008) Modern multivariate statistical techniques: regression, classification, and manifold learning. Springer, Berlin

Julier S, Uhlmann J, Durrant-Whyte H (2000) A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Trans Autom Control 45:477–482

Kalman R (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82:35–45

Karlgaard CD, Schaub H, Crassidis J, Psiaki M, Chiang R, Ohlmeyer EJ, Norgaard M (2007) Huber-based divided difference filtering. J Guid Control Dyn 30:885–891

Kim K, Jee G, Song J (2008) Carrier tracking loop using the adaptive two-stage Kalman filter for high dynamic situations. Int J Control Autom Syst 6:948–953

Kim K, Lee J, Park C (2009) Adaptive two-stage extended Kalman filter for a fault-tolerant INS-GPS loosely coupled system. IEEE Trans Aerospace Electron Syst 45:125–137

Koch K, Yang Y (1998) Robust Kalman filter for rank deficient observation models. J Geod 72:436–441

Kovacevic B, Durovic Z, Glavaski S (1992) On robust Kalman filtering. Int J Control 56:547–562

Maronna R, Martin R, Yohai V (2006) Robust statistics. Wiley, Chichester

Schick I, Mitter S (1994) Robust recursive estimation in the presence of heavy-tailed observation noise. Ann Stat 22:1045–1080

Simon D (2006) Optimal state estimation: Kalman, H infinity, and nonlinear approaches. Wiley-Interscience, New Jersey

Soken H, Hajiyev C (2010) Pico satellite attitude estimation via robust unscented Kalman filter in the presence of measurement faults. ISA Trans 49:249–256

Wang X, Cui N, Guo J (2010) Huber-based unscented filtering and its application to vision-based relative navigation. IET Radar Sonar Navig 4:134–141

Yang Y (1991) Robust bayesian estimation. J Geod 65:145–150

Yang Y, Gao W (2007) Comparison of two fading filters and adaptively robust filter. Geo Spatial Inf Sci 10:200–203

Yang Y, He H, Xu G (2001) Adaptively robust filtering for kinematic geodetic positioning. J Geod 75:109–116

Yang Y, Song L, Xu T (2002a) Robust estimator for correlated observations based on bifactor equivalent weights. J Geod 76:353–358

Yang Y, Song L, Xu T (2002b) Robust parameter estimation for correlated geodetic observations. Acta Geod Cartogr Sin 31:95–99

Zhou D, Frank P (1996) Strong tracking filtering of nonlinear time-varying stochastic systems with coloured noise: application to parameter estimation and empirical robustness analysis. Int J Control 65:295–307