DETECTION OF DEFORMATIONS AND OUTLIERS IN REAL-TIME GPS MEASUREMENTS BY KALMAN FILTER MODEL WITH SHAPING FILTER

Lihua Li, Heiner Kuhlmann

Institute of Geodesy and Geoinformation, University of Bonn
Nußallee 17, 53115 Bonn, Germany

Abstract: Kalman filter is used to process the real-time deformation series, but it requires white noise. Because the GPS observations with high sampling rate are correlated, the kalman filter with shaping filter is applied. In order to monitor and control the quality of the GPS observation series, a new method of simultaneously detecting deformations and outliers is put forward. The deformation and outlier have some similarity but also some differences so that they can be detected and distinguished simultaneously. This new method is applied in a GPS experiment and the feasibility of this method is certified.

1. INTRODUCTION

Landslide hazard is one of the major natural hazards. In order to reduce any human losses and the economic damages it is necessary to develop an early warning system of landslides. It is a challenging and complex topic to develop an early warning system of landslides because the landslide hazard is caused by mutual interaction of various factors. To develop the deformation monitoring and analysis model is one of the most significant parts, providing an important basis for identifying a landslide. GPS is a useful tool to obtain real-time observations of landslides. The main task of the GPS real-time series analysis is to separate the measurement deviations from the observations and to detect the point deformation epochs with less time delay.

As we know, Kalman filter is a key tool to process the real-time deformation series, but it requires white noise. Kuhlmann (2003) has obtained some results that the GPS observations with high sampling rate are correlated, the kalman filter with shaping filter is applied in the GPS time series. Shaping filter can be used for eliminating the colored noise. The shaping filter describes the long term movement of correlating measurement deviations. More accurate results can be obtained after the data is processed by the kalman filter model with shaping filter. In order to monitor and control the quality of the GPS observation series, we propose a new method that can detect deformations and outliers simultaneously in this paper. The deformation and outlier have some similarity but also some differences so that they can be detected and distinguished simultaneously. It is discussed in detail how to determine the state vector when outlier and deformation occur. In order to verify the new method, we have carried out a GPS experiment.

2. DESCRIPTION OF THE GPS EXPERIMENT

The GPS experiment was carried out on the roof of the Institute of Geodesy and Geoinformation and the Max-Planck Institute in Bonn, Germany. The baseline was about 1.2 km. The GPS equipment consisted of a Trimble 5700 receiver and Zephyr antennas. A cut-off angle of 10° was chosen and the sampling rate Δt was 1 second. During the GPS kinematic measurements, the height of the rover station which was on the roof of the Institute of Geodesy and Geoinformation was changed with a crank every 30 minutes in steps of 12.5mm. The kinematic measurement lasted for 6 hours (see Figure 1). The same baseline was observed for 17 hours with both fixed antennas (see Figure 2). The same sampling rate and the same cut-off angle was chosen. The static observations are used for the estimation of the parameters of the stochastic model. Additionally the developed method can be checked because here no deformation should be detected.

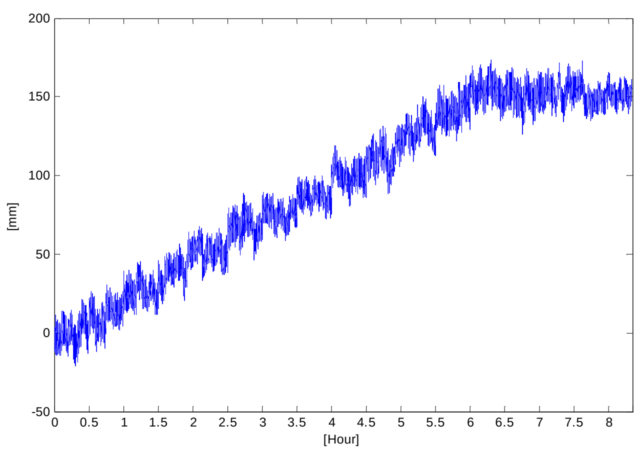

Figure 1 - Kinematic height observation time series

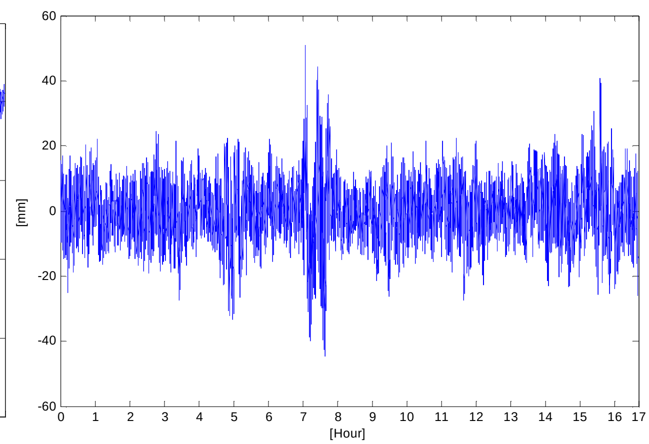

Figure 2 - Static height observation time series

3. DETERMINATION OF THE AUTOCORRELATION FUNCTION FOR CORRELATING ERRORS

As we know, because of the multipath effect and some other error sources, the GPS measurements with high sampling rate are influenced in a similiar way, resulting in autocorrelation. So GPS measurements contain colored noise. Firstly the autocorrelation function (Strang and Borre, 1997) is introduced.

3.1. Description of the Autocorrelation Function

The observation series is described as ($a_1, a_2, a_3,\ldots,a_i,\ldots a_n$) , which are made at equidistant time intervals Δt. n is the total number of the observations. Firstly, we compute the mean value m of all observations; secondly, we compute the autocorrelation coefficient $\hat{C}(k)$ of the observation series according to the definition below.

$$\hat{C}(0) = \sum_{1}^{n}(a_i - m)(a_i - m)/(n-1); \tag{1}$$

$$\hat{C}(1) = \sum_{2}^{n}(a_i - m)(a_{i-1} - m)/(n-2); \tag{2}$$

…

$$\hat{C}(k) = \sum_{k+1}^{n} (a_i - m)(a_{i-k} - m)/(n-k)\,; \qquad (3)$$

The normalized autocorrelation coefficient R_k is defined by the following formula,

$$R_k = \frac{\hat{C}(k)}{\hat{C}(0)} \qquad (4)$$

Where k is the normalized autocorrelation coefficient's index, the time lag between a_i and a_{i-k} is $k \cdot \Delta t$. The distribution of the autocorrelation coefficient of the GPS static observations is obtained by the formulas (1)-(4). The autocorrelation function of the GPS static height observations (see Figure 2) is described as follows.

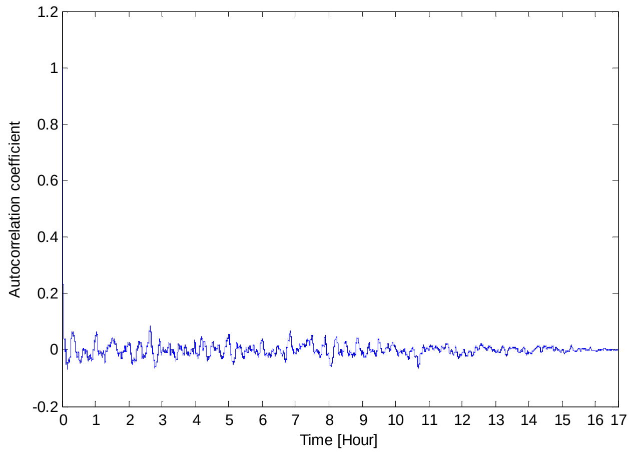

Figure 3 - Autocorrelation coefficient of all the GPS static height observations

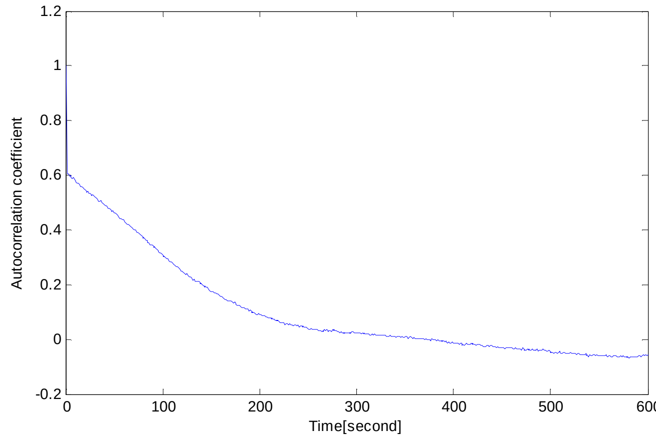

Figure 4 - Autocorrelation coefficient of first 600 GPS static height observations

In order to get a better understanding of the autocorrelation function, the autocorrelation coefficients of all the GPS static observations (see Figure 3) were zoomed to the autocorrelation coefficients of the first 600 GPS height observations (see Figure 4).

From Figure 4, we can see that the GPS measurement deviations are composed of white noise and colored noise. The colored noise follows the exponential distribution. When the time lag $k \cdot \Delta t$ is larger, for example 400 seconds, the autocorrelation of the observations is not so obvious. But when the time lag is smaller, for example 1 second, the autocorrelation coefficient between these two observations is becoming larger.

3.2. Stochastic Model's Determination

The GPS measurements deviations can be divided into correlating errors Δ and non-correlating errors δ (Schwieger, 1999). The correlating errors and non-correlating errors is seperately being described by the standard deviations σ_Δ and σ_δ . As mentioned above, correlating deviations follow a Gauß-Markov-process with correlation function $R(\tau) = e^{-\alpha|\tau|}$. According to the estimation of the stochastic model (Kuhlmann, 2003), we can get the results, $\sigma_\delta = 4.5545\,mm$, $\sigma_\Delta = 5.7241\,mm$ and $\alpha = 0.0063$.

4. KALMAN FILTER WITH SHAPING FILTER

The Kalman Filter, first derived by Kalman (1960) for use in electrical control systems, is a widely used method for deformation analysis nowadays. The kalman filter model (Welch and Bishop, 2006) can be described as follows,

The system equation is described by

$$x_k = \Phi_{k,k-1} x_{k-1} + w_k \tag{5}$$

where x_k, x_{k-1} are the state vectors at different epochs; $\Phi_{k,k-1}$ is the system transition Matrix; ω_k is the system noise.

The measurement equation is given by:

$$l_k = H_k x_k + \varepsilon_k \tag{6}$$

where l_k is the measurement vector at epoch *k*; H_k is the observation transition matrix; ε_k is the measurement noise.

There are two groups in the implementation of the kalman filter: time update equations and measurement update equations (Welch and Bishop, 2006).

The time update equations are

$$\overline{x}_k = \Phi_{k,k-1} \hat{x}_{k-1} \tag{7}$$

$$P_{\overline{x}_k} = \Phi_{k,k-1} P_{\hat{x}_{k-1}} \Phi^T_{k,k-1} + Q_k \tag{8}$$

where $\overline{x}_k$ is the predicted value of the state vector and $\hat{x}_{k-1}$ is the optimal estimator of the state parameters at the previous epoch $k-1$; $P_{\overline{x}_k}$ is the covariance of $\overline{x}_k$ and $P_{\hat{x}_{k-1}}$ is the covariance of $\hat{x}_{k-1}$; Q_k is the variance of the system noise w_k.

The measurement update equations are

$$G_k = P_{\overline{x}_k} H_k^T (H_k P_{\overline{x}_k} H_k^T + R)^{-1} \tag{9}$$

$$d_k = l_k - H_k \overline{x}_k \tag{10}$$

$$\hat{x}_k = \overline{x}_{k-1} + G_k d_k \tag{11}$$

$$Q_{d_k} = H_k P_{\overline{x}_k} H_k^T + R \tag{12}$$

$$P_{\hat{x}_k} = (I - G_k H_k) P_{\overline{x}_k} \tag{13}$$

where G_k is the gain matrix; R is the covariance of the observation noise; d_k is the innovation and Q_{d_k} is the covariance of d_k.

In this GPS experiment there are no forces at the input and the deformation changed very slowly, so the kalman filter model can be described as the identity model. Because of the colored noise in GPS measurements, a shaping filter is used. The state vector is augmented by another variable x_2 that is used to describe the long term movement of correlating measurement deviations.

The new state vector is

$$x_k = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$$

with the state vector of the height change at epoch k $x_1(k)$ and the movement of the correlation measurement deviations $x_2(k)$.

The new system equation and the new measurement equation are obtained as follows.

$$\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\alpha\Delta t} \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & e^{-\alpha\Delta t} \end{bmatrix} \begin{bmatrix} w_1(k) \\ w_2(k) \end{bmatrix} \tag{14}$$

$$l(k) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \varepsilon(k) \tag{15}$$

5. ALGORITHM OF DETECTING DEFORMATIONS AND OUTLIERS

5.1. Idea of the Algorithm

This research aims at developing algorithms which can detect the deformation and the outlier with short time delay.

If there are no changes in the time series, for example no deformation or no outlier in the GPS measurements, the filtered results should, from a statistical point of view, follow the normal distribution with mean μ and variance σ^2. The variance σ^2 is obtained from the static height observations time series by the formula

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{1f} - u)^2 \tag{16}$$

where x_{1f} denotes the filtered result of the static observations time series; u is the mean value of the filtered results in the time series and n is the number of the chosen static observations.

Usually the time series contains deformations or outliers. Therefore, a test factor

$$T = \frac{\left| x_{1f} - x_0 \right|}{\sigma} \tag{17}$$

is used to detect these changes. Here x_{1f} is the filtered result of the time series and x_0 is the initial value of the state vector x_1 which can be computed from the former measurements. It is known that T doesn't follow the normal distribution when the deformation or outlier occurs in the time series. With a given significance level α, we can get the boundary P_α of T from the normal table. By the hypothesis testing, it is possible to detect values which do not follow the normal distribution. That means the abrupt changes (deformation or outlier epoch) can be detected, but it is impossible to distinguish outliers from deformations by T.

The difference between the deformation and the outlier is that the outlier occurs isolated, that means the test should be accepted at the following epochs. The test factor T will change suddenly and then still follow the normal distribution for the next following epochs.

The situation is different when a deformation occurs. If there is a deformation, the observations at this epoch and the next following epochs are all changed and as a consequence the test factor T will not follow the normal distribution at the following epochs until x_0 is changed to the new initial value.

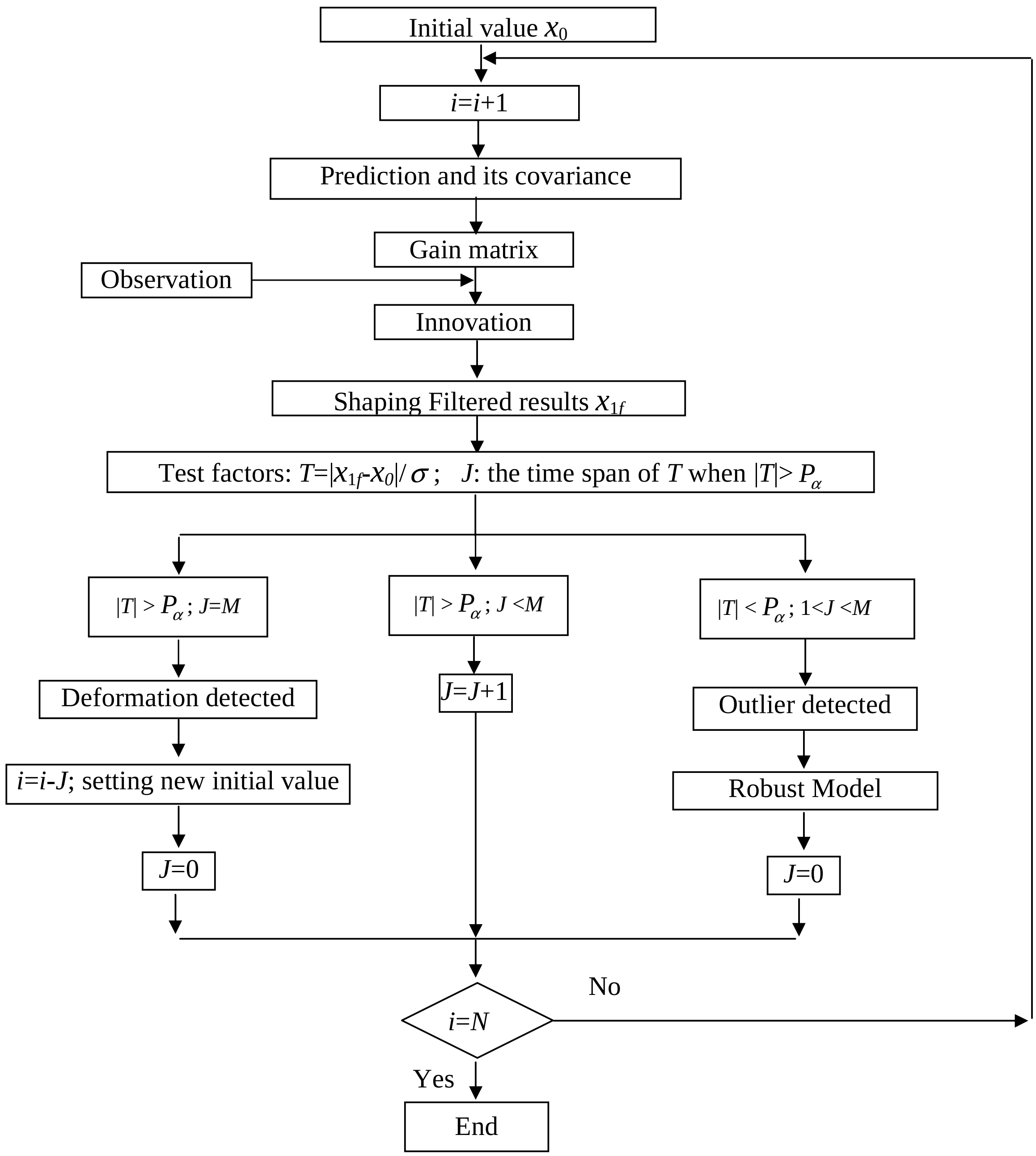

Figure 5 - Flowchart of the test procedure for the modified kalman filter

To make use of the described behavior another factor J is used. J is the number of continuous rejected tests. If J is smaller than a chosen boundary M it means that an outlier occurs. If J is larger than the boundary a deformation is found.

So the test factors T and J are two factors to distinguish outliers and deformations. The test factors T and J should detect the deformation as soon as possible, with short time delay and less false alarms (when a deformation is detected, the system will give an alarm). So a suitable decision for the factors is important.

The flowchart of this algorithm is described in Figure 5. The proposed procedure is applicable to any dynamic system.

In this example the significance level α is selected as 5%, the boundary of J is selected as M = 3 epochs, σ is 2.5 mm obtained from the static height series. Of course according to different significance levels, M is selected as different numbers.

5.2. Modification of the filtered Results when the Outlier is detected

The filtered results will be deteriorated by the outlier. Therefore, after the detection of the outlier we should reduce the outlier's influence on the filtered results. The gain matrix must be modified, because the outlier affects the filtered results by the gain matrix. In this paper, the method accepted is based on the idea of the equivalent weights function (Y.Yang, 2002). If an outlier occurs the modified kalman gain matrix G_k is construted as follows:

$$\gamma_{i,j} = \begin{cases} 1 & |S_k| < k_0 \\ \dfrac{k_0}{|S_k|}\left(\dfrac{k_1 - |S_k|}{k_1 - k_0}\right) & k_0 < |S_k| < k_1 \\ 0 & |S_k| > k_1 \end{cases} \tag{18}$$

$$G_k = G_k \cdot \gamma_{i,j} \tag{19}$$

where k_0 and k_1 are two constants, usually chosen as 2.0-3.0 and 4.5-8.5 respectively; $S_k = |d_k| Q_{d_k}^{-\frac{1}{2}}$.

5.3. Determination of the initial Value at the Epoch when the Deformation is detected

As we know, if the deformation is detected, the initial state value x_1 is changed to a new value which should be equal to the new deformation result $x_{1new}(k)$. There are 4 different methods to determine the new deformation value as the new initial mean value.

5.3.1. 1st method

According to the idea of the equivalent weight function, we can change the weight of the observations. When deformation occurs, the observations play a main role. Hence, the weight of the observations should be increased. Because the gain matrix can be considered as the weight of the observations, the gain matrix of $x_1(k)$ is modified as follows:

$$G_{1k} = \left(\frac{\hat{x}_1(k) - x_{1initial}}{\sigma}\right) \cdot G_{1k} \tag{20}$$

where $x_{1initial}$ is the initial value of $x_1(k)$.

The new state vector x_{knew} can be obtained as

$$x_{knew} = \begin{bmatrix} x_{1knew}(k) \\ x_{2knew}(k) \end{bmatrix} = \begin{bmatrix} \bar{x}_1(k) + \left(\dfrac{\hat{x}_1(k) - x_{1initial}}{\sigma}\right) \cdot G_{1k} \cdot d_k \\ 0 \end{bmatrix}.$$

5.3.2. 2nd method

If a deformation occurs at epoch k, the system equation does not describe the transformation between two neighbouring state vectors correctly. But the state vector's value at epoch k can be obtained from the measurement equation. As we know, the shaping filter is correlated and follows the exponential function; we can get the predicted shaping filter's value at epoch k $x_2(k)$. That is $x_2(k) = \mathrm{e}^{-\alpha t}\ x_2(k-1)$.

From the measurement equation,

$$l(k) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \varepsilon(k) = x_1(k) + x_2(k) + \varepsilon(k)$$

we can get the state vector's approximate value at epoch k,

$$x_{2new}(k) = \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1) \tag{21}$$

$$x_{1new}(k) = l(k) - x_{2new}(k) \tag{22}$$

The new state vector x_{knew} can be obtained as

$$x_{knew} = \begin{bmatrix} x_{1knew}(k) \\ x_{2knew}(k) \end{bmatrix} = \begin{bmatrix} l(k) - \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1) \\ \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1) \end{bmatrix}.$$

5.3.3. 3rd method

Because the shaping filter $x_2(k)$ follows the exponential distribute, $x_2(k)$ can be obtained $x_2(k) = \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1)$. Furthermore, the new state vector $x_{1new}(k)$ can be obtained by

$$x_{1new}(k) = \hat{x}_1(k-1) + velocity$$

The velocity was determined by the observation equations,

$$\begin{aligned} velocity &= x_1(k) - x_1(k-1) \\ \mathrm{M} \quad &= l(k) - x_2(k) - \varepsilon(k) - l(k-1) - x_2(k-1) - \varepsilon(k-1) \\ \mathrm{M} \quad &= l(k) - \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1) - \varepsilon(k) - l(k-1) - \hat{x}_2(k-1) - \varepsilon(k-1) \end{aligned} \tag{23}$$

The new state vector x_{knew} at this epoch k can be described as follows,

$$x_{knew} = \begin{bmatrix} x_{1knew}(k) \\ x_{2knew}(k) \end{bmatrix} = \begin{bmatrix} \hat{x}_1(k-1) + velocity \\ \mathrm{e}^{-\alpha t}\ \hat{x}_2(k-1) \end{bmatrix}.$$

5.3.4. 4th method

$x_{1new}(k)$ can still be determined by the formula $x_{1new}(k) = \hat{x}_1(k-1) + velocity$, but the method to determine the velocity is different from the 3rd method. The velocity is determined by the movement velocity observed by tacheometer. In this study, the observed velocity is equal to 1.5 times the difference between the filtered result $\hat{x}_1(k)$ and $\hat{x}_1(k-1)$. That is

$$velocity = 1.5 \cdot (\hat{x}_1(k) - \hat{x}_1(k-1)) \tag{24}$$

Thus, the state vector at the epoch k can be determined,

$$x_{knew} = \begin{bmatrix} x_{1knew}(k) \\ x_{2knew}(k) \end{bmatrix} = \begin{bmatrix} x_1(k-1) + 1.5 \cdot (x_1(k) - x_1(k-1)) \\ 0 \end{bmatrix}.$$

6. RESULTS ANALYSIS

6.1. Standard Deviation

The standard deviation of a random variable *x* is defined as:

$$\sigma = \sqrt{\frac{\sum \Delta\Delta}{n-1}} \tag{25}$$

where Δ is the error, i.e. the difference between the filtered results and the mean value; $n-1$ is the number of degrees of freedom.

After the static data series was processed by the kalman filter model with a shaping filter, the standard deviation σ was obtained as 2.5mm according to the equation (25).

6.2. Deformation Epochs detected

The method to detect the deformation epoch was applied to the static data series and the kinematic data series in order to check whether it works.

6.2.1. Static data processing results

	Detected Epochs (Height changed)		Standard Deviation (mm)
1st method	25339	25413	2.67
2nd method	25339		2.68
3rd method	25339		2.72
4th method	25339	25408	2.53
True epoch	No deformation epoch		

Table1 - static data processing results from 4 different methods

As mentioned above, there are 4 different methods to process the time series, so 4 different method results are obtained. As we know, the data series processed was static data series; there should be no deformation epochs detected.

But from the results above, epoch 25339 was detected in every method results. Therefore, it should be paid more attention to this epoch. It was discovered that at epoch 25339 the satellite geometry became poor. Geometric Dilution of Precision (GDOP) was about 20. That is why the result at this epoch is not accurate and this epoch is taken as the deformation epoch.

Compared to the standard deviations, the results obtained from the fourth method are better than the results from the other methods.

6.2.2. Kinematic data processing results

For the kinematic data series, the height was changed by the crank every 30 minutes. The deformation epochs were known. *J*=3 is the best chosen number compared to the other numbers and σ is computed as 2.5mm.

Epochs when height was changed					
True Epochs	Detected epoch results from 4 methods				Earliest detection
	1st method	2nd method	3rd method	4th method	
1800	1847	1847	1847	1847	1847
3600	3633	3601	3604	3604	3601
5400	6015	5833	5833	6015	5833
7200	7243	7476	7525	7211	7211
9000	9314	9379	9367	9271	9271
10800	Not detected	11026	10932	10876	10876
12600	12776	13465	12811	12710	12710
14400	14540	14512	14612	14540	14512
16200	16477	16377	16893	16457	16377
18000	18128	18327	18297	18273	18128
19800	20011	19899	20675	20222	19899
21600	21650	22098	22530	21808	21650
Max time delay	615s	865s	930 s	615 s	433s

Table 2 - Kinematic data processing results from 4 different methods

For the kinematic data series, it is more important to determine the new deformation value. Because the new deformation value will be set as the new initial value in order to detect the next deformation epoch. We can compare the results from 4 different methods in order to obtain more accurate epochs with short time delay. Table 2 shows the epochs detected by 4 different methods.

In this experiment, the height was changed every half an hour. The epoch should be detected at epochs1800, 3600, 5400,…, 21600, which are described in the first column of the table 2. In fact, the epochs are detected with different time delay by different methods. The principle to get more accurate epochs is to choose the earliest epoch among the 4 different detected results. For example, at the epoch 3600, the height was changed, but from the results, the detected results with different time delay were 3633, 3601, 3604 and 3604. We chose the earliest epoch 3601 with the least time delay as the best result. The last column of the table 2 shows the results with the least time delay when the height was changed everytime. From the last column of the table 2, we can see that the fastest detection is 1 second time delay, the slowest detection is 433 seconds time delay.

The processed filtered results can describe the deformation tendency more precisely than the observations because the noise in the time series was deleted from the observation time series.

7. CONLUDING REMARKS

Because of the colored noise in the GPS measurements, kalman filter with a shaping filtered is used in the real-time series. And in this paper the proposed method can be used to detect and distinguish deformations and outliers simultaneously. How to determine the state vector value when outliers and deformations occur is discussed in detail. An application to the GPS static and kinematic time series demonstrates that the method proposed can get the results with short time delay. This proposed method is useful to analyse the time series and make the right decision when deformations occur.

References

Kalman R. E. (1960) A New Approach to Linear Filtering and Prediction Problems, ASME Journal of Basic Engineering, Vol. 82D, PP.35-45.

Kuhlmann, H. (2003). Kalman-filtering with coloured measurement noise for deformation analysis, 11th FIG International Symposium on Deformation Measurements, Santorini, Greece, 2003 (http://www.fig.net/figtree/commission6/santorini).

Strang, G.and Borre, K. (1997). Linear Algebra, Geodesy, and GPS. Wellesley-Cambridge Press.

Schwieger, V.(1999) Ein Elementarfehlermodell für GPS-Überwachungsmessungen-Konstruktion und Bedeutung interepochaler Korrelationen. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr.231

Welch,G. and Bishop,G(2006) An introduction to the Kalman Filter, University of North Carolina at Chapel Hill, TR95-041, 2006 (http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf)

Yang, Y. (2002) Robust estimator for correlated observations based on bifactor equivalent weights, Journal of Geodesy(2002) 76:353-358.

Corresponding author contacts

M.S.c Lihua Li
lihua.li@uni-bonn.de
Institute of Geodesy and Geoinformation, University of Bonn
Nußallee 17, 53115 Bonn
Germany

Prof. Dr.-Ing. Heiner Kuhlmann
Heiner.Kuhlmann@uni-bonn.de
Institute of Geodesy and Geoinformation, University of Bonn
Nußallee 17, 53115 Bonn
Germany