Principal Component Analysis of Wavelet Transformed GPS Data for Deformation Monitoring

C.Ogaja, J. Wang, C.Rizos
School of Surveying and Spatial Information Systems,
The University of New South Wales, Sydney, NSW 2052, Australia

Abstract. In a continuous GPS deformation monitoring scheme, any action taken generally relies on a description of the state of the process or events as given by GPS measurements. Timely and correct interpretation of the GPS data is essential to improved quality control, safer system operations, and reduction in the number of false alarms. Unfortunately, GPS data are contaminated by both random and systematic errors of unknown sources. Instrument failure, poor or uncalibrated instrumentation, receiver noise and multipath, all can contribute to data problems. Without proper pre-treatment, the necessary interpretation is difficult, if not impossible. Data contaminated by outliers must be eliminated and noise levels reduced. In many cases, critical information occurs over a short duration, and hence is difficult to detect. Wavelets can be used to pre-process data in order to better locate and identify significant events. Combining this type of data pre-processing with multivariate statistics can generate useful insights into the problems of deformation monitoring, data analysis and data interpretation.

Multivariate techniques can be used to identify process variability and to develop models for on-line monitoring and control. By comparing new observations with a reference model that describes normal variability, simple control charts can be constructed to detect data inconsistencies and processing problems. This paper extends these concepts by demonstrating that correlated output of multiple sensors such as GPS receivers, accelerometers, anemometers or temperature sensors can be significantly improved with pre-filtering of the time series signals using a median filter, and a time-scale decomposition using a multi-resolution wavelet function. After the data are filtered and decomposed, the multivariate statistical method of Principal Component Analysis is used to develop a deformation monitoring model.

Keywords. FIR median hybrid filter, wavelet transformation, principal component analysis, GPS deformation studies

1 Introduction

Deformation monitoring includes the tasks to detect sensor malfunctions, measurement jumps and discontinuities, or other special events such as systematic biases, as early as possible, and then finding and removing (or mitigating) the factors causing those events. Some recent continuous GPS deformation studies (e.g. Ogaja *et al*, 2001a; Roberts *et al*, 2000) incorporates multiple GPS receivers and additional sensors. The measurement data are more frequently recorded due to the high sampling rate and Real-Time Kinematic (RTK) capability. In such cases, traditional techniques like Statistical Process Control (SPC) charts do not allow a sufficient detection speed, and Multi-variate Statistical Process Control (MSPC) methods (e.g. Jackson, 1980; Dunia *et al*, 1996) may be applied.

This paper aims at the application of two pre-processing techniques for the development of a monitoring model, namely:

- a Finite Impulse Response (FIR) Median Hybrid Filter (FMH), and
- the Haar Wavelet Transformation (HWT)

The MSPC method of Principal Component Analysis (PCA) is proposed and applied to data that have been pre-processed by both the FMH and HWT. For on-line monitoring, a control limit whose value is computed by means of either the chi-square-distribution or *F*-distribution (Ogaja *et al*, 2001b), is set for the statistic T^2 (whose realisation at the "start-up" stage is the *Q* statistic).

The proposed methodology is discussed in the following sections, followed by a test example and the concluding remarks.

2. Methodology

2.1 FIR Median Hybrid Filter

To extract useful information from on-line data, a robust method is needed to pre-process the data to eliminate noise prior to use in deformation monitoring model development. In contrast to the usual linear filter, the Finite impulse response Median Hybrid (or FMH) filter (Heinonen & Neuvo, 1987) is able to reduce the noise and to preserve sudden jumps in the data. Those jumps are the sudden events which have to be detected during the pre-processing.

For a time series $\{x_1, \ldots, x_N\}$, the $(K + 1)^{st}$ FMH filter output is given by :

$$x_{k+1}^{FMH} := median\{\hat{x}_{K+1}, \bar{x}_{K+1}, x_{K+1}\} \qquad (1)$$

where

$\hat{x}_{K+1}$ is the average of K past samples,
$\bar{x}_{K+1}$ is the average of K future samples.

The filter (1) is applied iteratively to enhance noise reduction. It involves the determination of the median of three elements, where K is the window half-width. $K > 0$ but smaller than half the length of the shortest expected region of constant value in the data. This is referred to as a steady-state region. Large values of K will result in the distortion of a signal whose constant region is of a duration shorter than K, however, a large K does contribute to improving the noise suppression capabilities. Ideally, K should be chosen as large as possible without losing steady-state information.

Figure 1 illustrates the performance of the FMH filter on a signal that is corrupted with Gaussian noise. Using a window half-width of 9 and 50 consecutive passes of the filter, the noise is essentially filtered from the signal while preserving the steady-state regions.

2.2 Wavelet Transformation

Wavelet transformation involves representing general functions in terms of simple, fixed blocks at different scales and positions. These blocks are a family of wavelet functions generated from a prototype function, called "mother" wavelet, by translation and scaling operations. The prototype function must satisfy certain admissibility conditions (Chui, 1992).

One way to distinguish between the various types of wavelet transforms is whether the time and scale parameters are continuous or discrete. Continuous signals leads to Continuous Wavelet Transforms (CWT) if the parameters are continuous, and to a Wavelet Series Expansion if the parameters are discrete. Discrete signals leads to Discrete Wavelet Transforms (DWT) if the parameters are chosen to be discrete. The type of "mother" wavelet applied can also be used to describe the transformation process.

In this work, we consider the use of a Haar Wavelet Transform (HWT) whose continuous form is given as (Olivier & Vetterli, 1991; Wang, 1995):

$$Tf(s,t) = \int \psi_s(t-u) f(u) du \qquad (2)$$

The notation ψ in (2) is the prototype function, and the transform $Tf(s,t)$ is a function of both the scale, or frequency, s and the spatial position, or time, t. The plane defined by the pair of variables (s, t) is called the scale-space, or time-frequency plane, and the value of $Tf(s,t)$ depends upon the value of function f in the neighbourhood of t of size proportional to the scale s. At small scales, $Tf(s,t)$ provides localised information.

In practice, the measured data are normally observed only at discrete values. The discrete Haar Wavelet Transform (DHWT), a discretised version of the continuous transform in (2), is therefore applied. In that case, the simple prototype Haar function, developed around 1910, is used (Chui, 1992):

$$\psi_H(t) := \begin{cases} 1 & 0 \le t \le 1/2 \\ -1 & 1/2 < t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and the time-scale parameters are discretised. Thus, if $s = 2^j$ and $u = k2^j$, the corresponding wavelets become a function of two integer parameters, j and

k. For this case, the wavelets form a dyadic series whose coefficients are easily calculated for combinations of the parameters j and k.

The dyadic form of the DHWT is applied to the mean-centred data after the data are pre-filtered by the FMH filter to remove the high frequency (noise) and gross outliers. Sudden changes which have to be detected are indicated by large values of the wavelet coefficients on the finest scale.

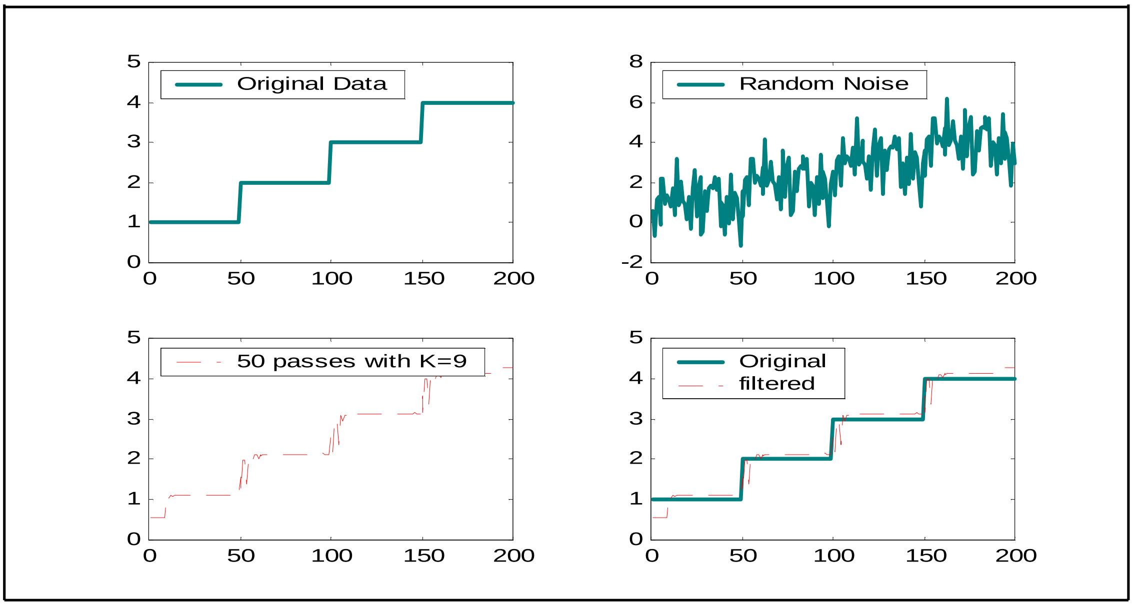

Fig. 1: FMH/Median hybrid filter example. **Top left panel:** original data signal; **top right panel:** signal with Gaussian random noise, **bottom left panel:** filtered signal after 50 consecutive passes and a window half-width of 9; **bottom right panel:** a superposition of the filtered and the uncorrupted signal.

2.3 Principal Component Analysis

The principal component analysis (PCA) is a MSPC method for monitoring simultaneously the correlated output of sensor or multi-sensor measurements. In this method, a T^2 statistic is applied to test for significant deviation of the current output from the normal process characteristics. The realisation of T^2 at the "start-up" stage gives rise to another statistic, the Q statistic. A control limit is established prior to the start of the measurement process, during which the reference mean value and covariance matrix of the correlated output are estimated from some "clean" past samples.

To derive the control limit for the PCA, assume that the correlated output comprising of p variables follow a p-dimensional multivariate normal distribution with the mean vector, $(\mu_1, \ldots, \mu_p)^T$ and covariance matrix, Σ. The notation μ_i is the mean for the i^{th} variable and Σ is a $p \times p$ matrix consisting of the variances and covariances of the p variables. This assumption is valid at the "start up" stage solely for the purpose of deriving the control limit. After this, the data are assumed to be reasonably normally distributed.

In the "start up" stage, m past samples are available to estimate the parameters $(\mu_1, \ldots, \mu_p)^T$ and Σ. If the t epoch observation of the p variables from the m reference samples is represented by the vector X_t, and their estimated mean vector containing the means of each variable is $\bar{X}_t$, the estimated covariance matrix is given by:

$$S_m = \frac{1}{m}\sum_{t=1}^{m}(X_t - \bar{X}_m)(X_t - \bar{X}_m)^T$$

To construct a multivariate control chart based on the Hotelling's T^2 statistic for observation X_t one uses the charting statistic:

$$Q_t = (X_t - \bar{X}_m)^T S_m^{-1}(X_t - \bar{X}_m). \qquad (3)$$

where the distribution of Q_t is approximated with a chi-square or an F distribution to obtain the control chart limits (Ogaja *et al*, 2001b).

Once the control limits are established at the "start up" stage, monitoring of future samples is realised by the charting statistic:

$$T_f^2 = (X_f - \overline{X}_m)^{\mathrm{T}} S_m^{-1} (X_f - \overline{X}_m) \qquad (4)$$

where X_f denotes the p-dimensional vector of future observations on the p variables, $\overline{X}_m$ is the p-dimensional mean vector of the m observations in the "clean" reference samples, S_m is the $p \times p$ covariance matrix associated with these observations.

3. Online Procedure

The techniques discussed in the previous section are combined to an online procedure (see below). For that procedure, the time series data for each variable are normalised, that is, adjusted to the zero mean and unit variance. Those data are then pre-processed in order to develop an online monitoring model. The proposed procedure is as follows:

- Apply a FMH filter to the individual sensor output variables (or components) to remove the "unwanted" high frequency components (noise) and gross outliers.
- Select a uniform wavelet transformation scale L, for all the measured variables (or components). To each variable, apply the DHWT to generate the wavelet coefficients vector, $G_j x\{j = 1, \ldots, L\}$, and a scaling function coefficients vector, $H_L x$. G_j and H_L represents filters and x is the original data.
- Apply the PCA to the wavelet spectra (coefficients) of all the variables of interest. This gives rise to the current value of the T^2 (or Q) statistic of the FMH-filtered wavelet transformed data.
- To the T^2 (or Q) statistic obtained at the selected scale, apply a control limit obtained with a chi-square or F distribution (see previous section).

The data representing the current condition is updated by moving a time-window step-by-step, and scaled by using the mean and the variance obtained at the initial step.

The advantage of the proposed procedure is that the variability defined by the wavelet coefficients is better visible in the wavelet spectrum than in the time history of the data itself. For instance, small shifts in several variables may be significant for establishing the occurrence of an event, a system fault or upset. However, if the time duration is small, this contribution is lost when compared to more significant variations. The wavelet transformation concentrates this variability into a few coefficients, thereby enhancing their effect in the wavelet domain as compared to the time domain.

4. Test Example

The proposed procedure is applied to the correlated output of sensors. Figure 2 represents GPS observations (GPSN & GPSE), wind velocity (UE & UW), accelerometer response (δE & δN) and temperature (T) recorded on the rooftop of Singapore's tallest building, the Republic Plaza Building, during a pilot experiment executed between 17th and 25th February 2000. The original GPS data were recorded using two Trimble 4700 receivers sampling at 1Hz (resampled in Fig. 2), while the wind velocity data and accelerometer response were recorded using two UVW anemometers and B1 accelerometers respectively. All sensors were installed on the rooftop. The reference station for the GPS sub-system was located on the rooftop of a nearby low-rise building, approximately 1km away. The experiment covered two periods of maximum wind conditions reaching up to 10m/s, during the 19th and 22nd of February 2000.

With respect to the GPS vectors, the significant events that occurred during the 19th and 22nd are not revealed in the time series of the GPS data. Figure 3 shows the results of applying the DHWT to FMH filtered GPS vectors, each containing 1112 samples. In each of the panels, the top plot indicates the original GPS vectors, the middle plot are the results after applying the FMH filter using 20 passes and with a half window size of 9, and the bottom plot indicates the empirical wavelet coefficients from a DHWT of the FMH filtered data. Note that the steady state conditions are preserved and additional filtering was not necessary as it did not improve the signals.

According to Figure 4, applying PCA to the wavelet transformed data revealed three significant events including the two wind conditions of 19th and 22nd February 2000. The plots of the Q-statistic both of the original unfiltered GPS signals and the wavelet transform of the FMH filtered GPS signals are shown, the horizontal lines representing the control limits.

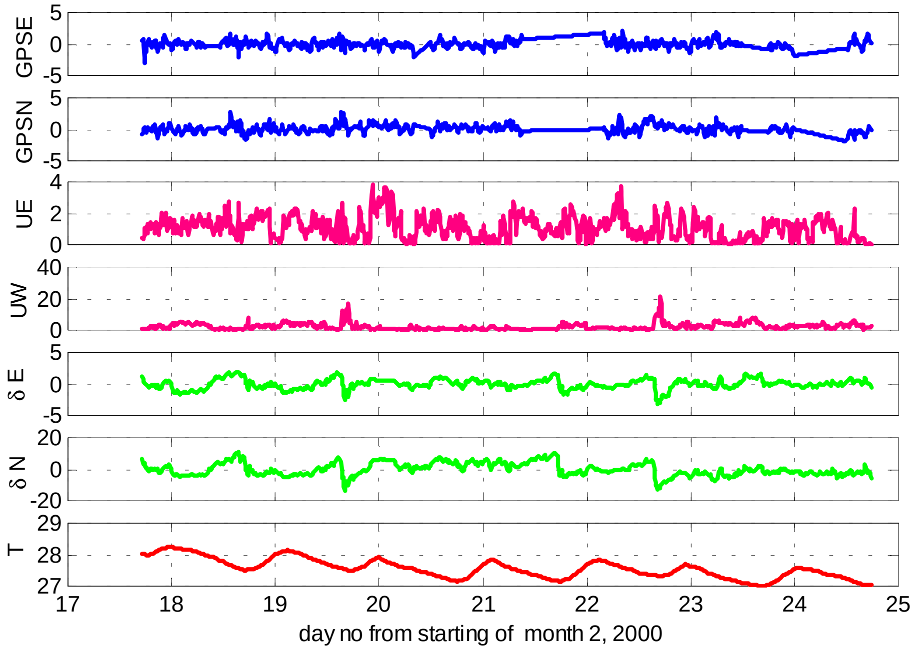

Fig. 2: Simultaneous GPS and Auxilliary Sensor data. From Top to bottom: **Panel 1 & 2:** GPS signals; **Panel 3 & 4:** Wind velocity vectors; **Panel 5 & 6:** Acceleration response vectors; **Panel 7:** Differential temperature.

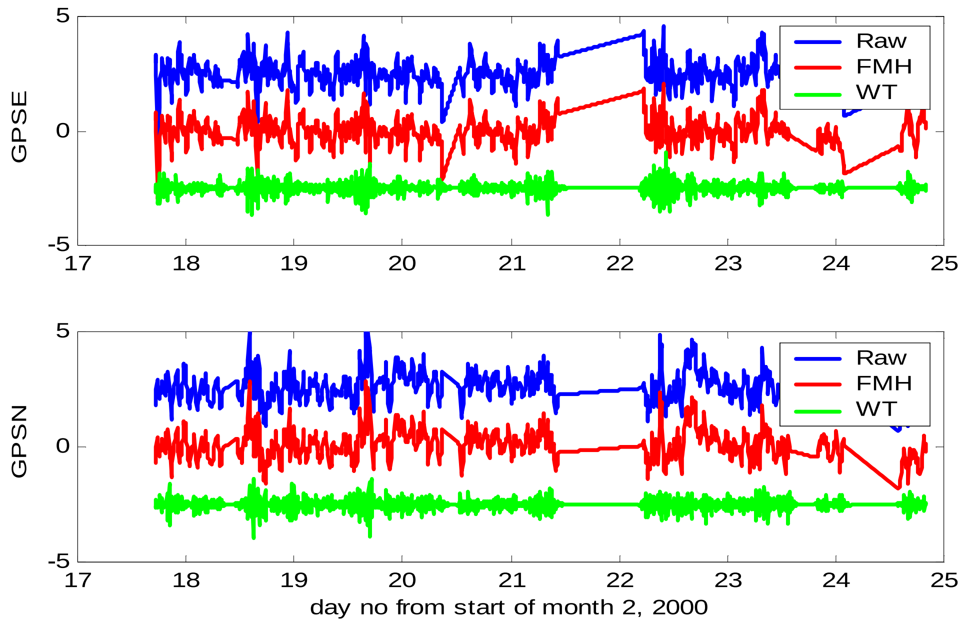

Fig. 3: Wavelet Transform of FMH filtered GPS data signals. **Top panel:** DHWT of FMH filtered GPSE data signals; **bottom panel:** DHWT of FMH filtered GPSN data signals.

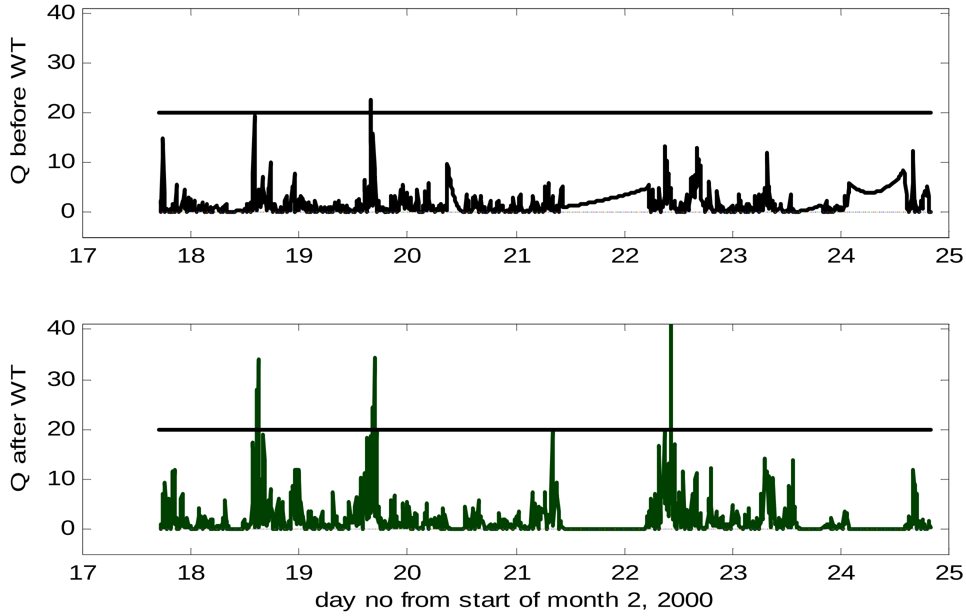

Fig. 4: Monitoring by *Q*-statistic. **Top panel:** *Q*-statistic of original GPSE and GPSN signals; **bottom panel:** *Q*-statistic of wavelet transformed FMH filtered GPSE and GPSN signals.

5. Concluding Remarks

A discussion is presented on the suitability of applying a multivariate technique to the wavelet transformed GPS data in order to monitor events in a GPS deformation monitoring scheme. The study demonstrates that pre-filtering of the data with a Finite impulse response Median Hybrid filter (FMH), and developing the calibration model from a set of coefficients obtained from a discrete Haar wavelet transform (DHWT) of the FMH-filtered data, produces a calibration model with better (sharper) classification features.

Acknowledgements

The principal author completed this work while undertaking PhD studies at the University of New South Wales (Australia) as an International Postgraduate Research Scholar (IPRS). The second author is the holder of an Australian Research Council Postdoctoral Fellowship.

References

Chui, C. K. (1992). An Introduction to Wavelets. Academic Press, New York.

Dunia, R., S.J. Qin, T.F. Edgar & T.J. McAvoy (1996). Identification of faulty sensors using principal component analysis. *AIChE J.*, 42, 2797-2812.

Heinonen, P., & Y. Neuvo (1987). Fir-median hybrid filters. *IEEE Trans. Acoustics, Speech & Signal Processes*, 35, 832-838.

Jackson, J.E. (1980). Principal components and factor analysis: Part I – principal components. *Journal of Quality Technology*, 12, 201-213.

Ogaja, C., C. Rizos, J. Wang, & J. Brownjohn (2001a). High precision dynamic GPS system for on-line structural monitoring. *5th Int. Symp. on Satellite Navigation Technology & Applications*, Canberra, Australia, 24-27 July, paper 35, CD-ROM Proc.

Ogaja, C., J. Wang, C. Rizos & J. Brownjohn (2001b). Multivariate monitoring with GPS observations and auxiliary multi-sensor data. *To be published in GPS Solutions*.

Olivier, R. & M. Vetterli (1991). Wavelets and signal processing, *IEEE Sig. Processing*, October, 14-38.

Roberts, G.W., X. Meng & A.H. Dodson (2000). Structural dynamic and deflection monitoring using integrated GPS and triaxial accelerometers. *13th Int. Tech. Meeting of the Satellite Division* *of the U.S. Inst. of Navigation GPS ION 2000*, Salt Lake City, Utah, 19-22 September, 59-68.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika*, 82, 2, 385-397.