UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 1/MS – 2007

Statistical analysis: a tool for understanding monitoring data

by

Domenico Bruzzi, Alessandro Fassò, Orietta Nicolis, and Giorgio Pezzetti

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

Statistical analysis: a tool for understanding monitoring data.

Domenico Bruzzi[1], Alessandro Fassò[2], Orietta Nicolis[3], Giorgio Pezzetti[4]

[1]President, SISGEO S.r.l., 12345 Masate - MI – Italy , domenico@sisgeo.com
[2]Professor of Statistics, University of Bergamo, Italy, 24046 Dalmine BG, alessandro.fasso@unibg.it
[3]Assistant professor of Statistics, University of Bergamo, Italy, 24046 Dalmine BG, orietta.nicolis@unibg.it
[4]Technical Director, FIELD S.r.l., 24040 Lallio BG, Italy, giorgio.pezzetti@fieldsrl.it

ABSTRACT: Monitoring data often need deep analysis in order to understand the real behaviour of a structure or a soil mass which may be overshadowed by unknown phenomena or boundary conditions. Temperature is one of the most common parameters which affect field measurements and create problems in data interpretation and use.
One of the most powerful tool which can be used to perform these analyses is represented by statistical approach. The analysis starts from the knowledge of the mechanics of the considered phenomena and related physical assumptions in order to separate the contribution of the different components.
Statistical techniques allow to identify the different contributions and to reduce the uncertainty of a system. Moreover, by means of a hierarchical multivariate detector (HMD) based on statistical control charts it is possible to analyze the behaviour of complex systems considering all the measuring instruments and their data in order to fix threshold values and to check the tendency of the system to change its status from a safety situation to a warning or alert one.
This paper describes the statistical background of the approach with reference to a specific application on a stayed bridge.

1. INTRODUCTION

Monitoring data very often contain a number of information that is difficult to separate and to understand. The monitored phenomena are influenced by other parameters as well as by environmental conditions which can be considered as part of the investigation analysis or as a disturb, modifying or overshadowing the monitored parameters.
In order to solve the problem of separating the effects of undesired variables from the parameters under observation, a statistical approach is proposed. The choice of the statistical approach is based on the assumption that in most cases it is the only possibility to deal such complex problems and to find the relationships among different quantities.
The monitoring activity starts from the most simple analyses such as the research of the correlation between the variables (parameters which are under monitoring) and covariates (quantities that influence the parameters) and leads to the development of an hierarchical multivariate detector based on the use of the so called “control charts” which enable to analyse the effect of an “anomaly” on the whole set of data which means on the monitored system (Fuchs C. and Kenett 1998; Montgomery, 2005). These control charts can be used to “predict” the behaviour of a system when one or more parameters change according to a specific law.

The approach is presented with reference to a specific application on a road bridge.
The bridge is a three span (45 – 90 – 45 m) cable stayed bridge which support a heavy traffic and is a crucial point for the whole mobility of the city of Milan.
The Authority in charge of the bridge asked for a monitoring system which measures a number of parameters to be used as reference for planning the maintenance of the bridge as well as for the global safety of the structure.
The monitoring system was designed and installed at the beginning of 2005 and data have been collected starting from April 2005.
The activity of data analysis has been carried out by means of a specific statistical program developed by SISGEO, FIELD and the University of Bergamo and started from the analyses of the behaviour of the instruments in order to find out the correlation with other parameters or environmental quantities such as temperature and ended up with the set-up a multivariate control chart for the whole system.
In this work, first, we propose a model for describing the correlation structure and the dynamics of the data in normal conditions, that is when there aren't anomalies. Then, we use the residuals of the model to get a hierarchical structure of control charts which allows to identify the instrument affected by the anomaly, starting from a general statistic of the overall system. In this application, we define a hierarchical multivariate detector (HMD) based on the multivariate Exponentially Weighted Moving Average control chart (EWMA)(Lowry et al. 1992). As EWMA charts are particularly sensitive to small shifts, the HMD allows to monitoring the bridge in the long-term period in accordance with the maintenance program. In order to show the capability of the proposed monitoring system to detect slow structural variations, we consider some examples where we simulate certain structural anomalies.

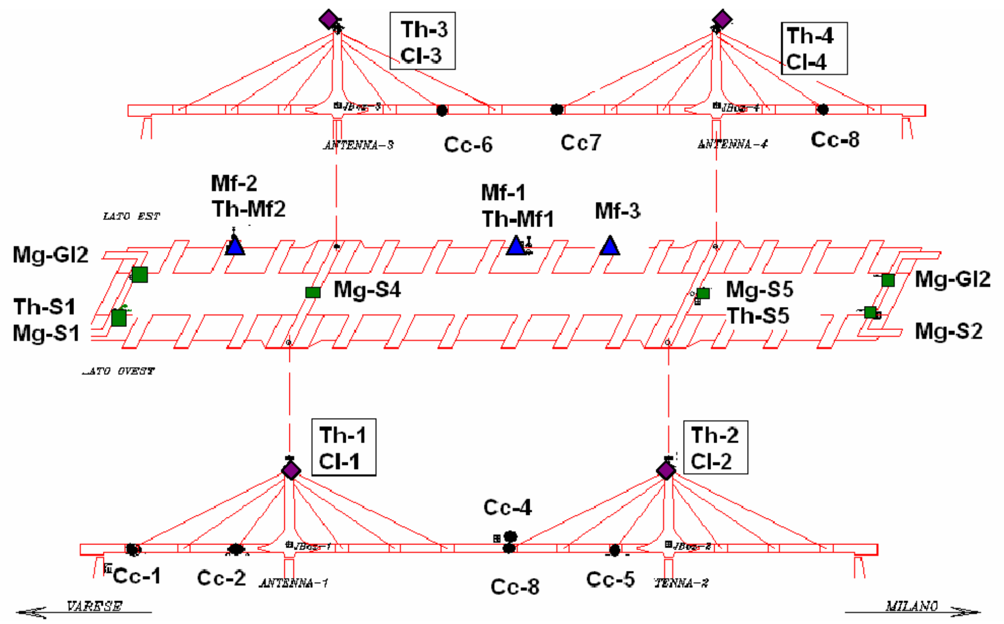

FIG. 1. Planimetry of the stayed road bridge and instruments: load cells (circles) ; deflection meters (triangles); joint meters (squares) and clinometers (diamond).

The work is organized as follows. Section 2 describes the instruments installed for monitoring the bridge and a statistical preliminary analysis of the measurements is done in Section 3. The statistical methodology for monitoring the structure is dealt in Section 4 and 5: in Section 4 an integrated statistical model for estimating the dynamics of the measurements is proposed and Section 5 describes the HMD based on the EWMA control charts. In Section 6, two examples of simulated anomalies are considered. Finally, conclusions are provided in Section 6.

2. THE MONITORING DATA SET

The monitoring system of the bridge considered in our application is shown in the planimetry of Fig. 1. It is composed by four different measurement instruments: the joint meters (Mg), the deflection meters (Mf), the clinometers (Cl) and the load cells (Cc). Nearby some of these instruments the thermometers have been installed. These instruments are set in different points of the bridge and measure some important characteristics for the monitoring process: the clinometers have been installed on the top of the cable towers and measure the slopes in two different directions (ChA and ChB); the joint meters have been fixed at the base of the bridge and measure the horizontal movements of the structure; the deflection meters have been installed at the center of two spans of the bridge for measuring the vertical movements of the structure; the load cells have been installed, at the construction stage, on some cable anchors and measure the load into the cables.
It is important to consider all these measurements jointly in the monitoring process because some observed values from one instrument could be strongly correlated to the others. For example, an anomalous increase of the deflection measurements could be correlated with a decrease of the cables tension, caused by a settlement of an anchorage. All these instruments are connected to data acquisition unit which stores all data at regular frequency. The recorded values can be read through a GSM modem. Even if the monitoring system has started to store the data from April 2005 with a regular frequency of four hours, not all the instruments have been installed at the same time because the installation procedure has followed a particular maintenance program. In this work we consider the daily average measurements from April, 5 - 2005 to May 7 - 2006 for the following instruments: 3 clinometers (Cl-1 chA, Cl-1 chB, and Cl-2 chA); 6 joint meters (Mg-S1, Mg-S2, Mg-S4, Mg-S5, Mg-GL1 and Mg-GL2) ; 5 load cells (Cc-1, Cc-2, Cc-3, Cc-4 and Cc-5). For the temperatures, the five thermometers (T-Cl1, TCl2, T-S1, T-S5 and T-Mf2) have been considered for the same period.

3. PRELIMINARY ANALYSIS OF THE DATA

In order to identify a model for a long term statistical monitoring, a preliminary statistical analysis has been done considering daily average measurements observed in the period from April 5, 2005 to the 7, 2006. To do this, the statistical software "Fieldstat" has been used (Fassò *et al.*, 2004).
As the data have a different variability we rescaled the original measurements respect to their standard deviations. Figg. 2, 3 and 4 show the standardized daily measurements for the joint meters, clinometers and load cells, respectively.
As expected, the readings of the same type of instrument have a similar behavior

which can be represented by an high correlation value, from the statistical point of view. Observing the joint meter measurements in Fig. 2 it is possible to note a strong cyclical trend, mainly due to the effect of meteorological variables. This is more evident if we plot these measures against temperatures: in Fig. 5 we show an example of the correlation between the measures of Mg-S1 and the temperatures given by its thermometer, T-S1. In this case, the linear regression captures more than the 97% of the variability of the data. This analysis has been done for all the instruments and, except for instrument Mg-S4, the R^2 indices resulted higher than 0.96, showing a strong correlation with temperatures. The instrument Mg-S4 has been installed on a fixed point of the bridge and, for this reason, its measures are less influenced by meteorological variables. The linear regression and other advanced statistical methods, such as local polynomial regressions, can be used to reduce the variability of the data mainly due to some exogenous variables (temperatures, traffic, wind, etc.) and extract more precise measures (see, for example, Fassò *et al.*, 2004 and 2005). This correlation with the temperature is lower for the clinometers and the load cells, even if the correlation values for the last ones are slightly higher than those obtained using the clinometers.

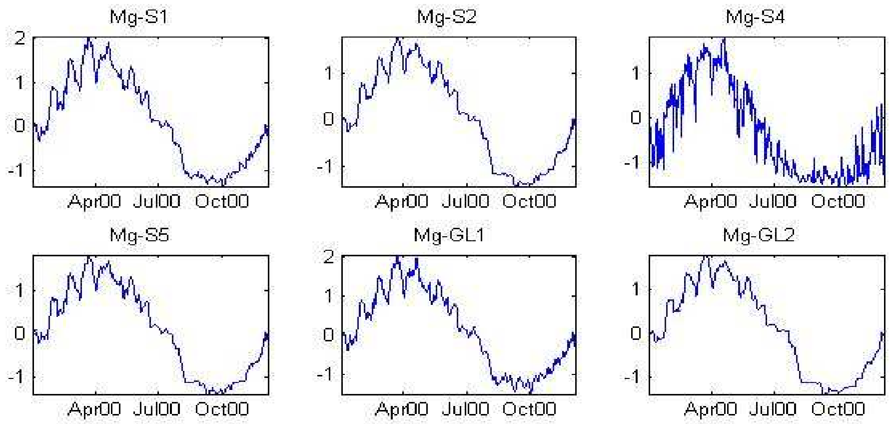

FIG. 2. Daily average measurements of the joint meters from April 5, 2005 to the May 7, 2006

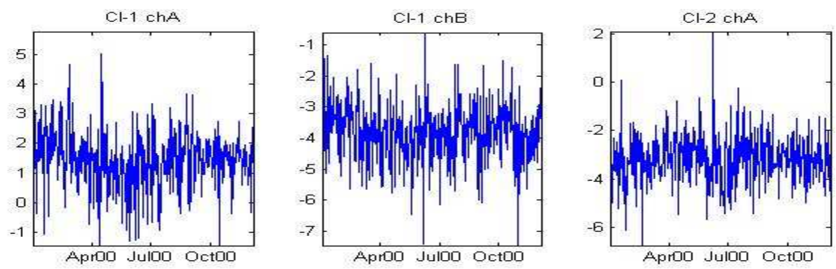

FIG. 3. Daily average measurements of the clinometers from April 5, 2005 to the May 7, 2006

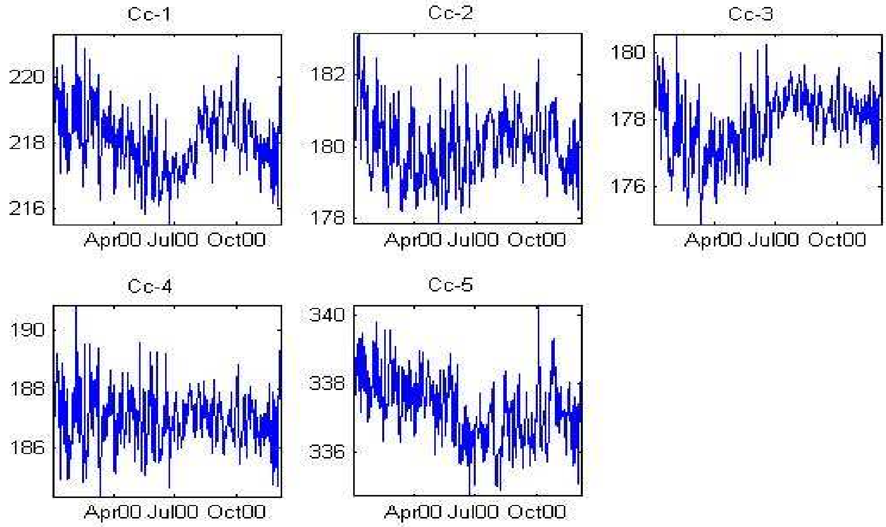

FIG. 4. Daily average measurements of the load cells from April 5, 2005 to the May 7, 2006

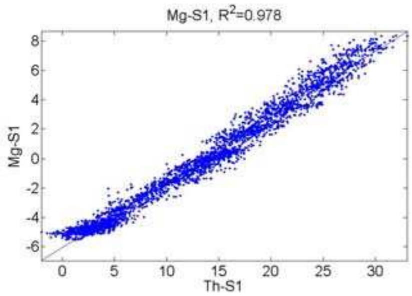

FIG. 5. Measurements of Mg-S1 versus temperature (T-S1): R^2=0.978.

In order to understand the dynamics of the system it is very important to estimate the autocorrelations of the instrumental data, that is the correlation between a measure *y* and the same one delayed of *j* days. As the autocorrelation behavior resulted very similar for the readings of the same type of instrument, we show in Fig. 6 only one for each instrument. It is apparent that the clinometers have a behavior similar to a white noise, the joint meters show a strong inertia, probably due to the effect of temperature, and the load cells have some significant correlations although lower than the joint meters.

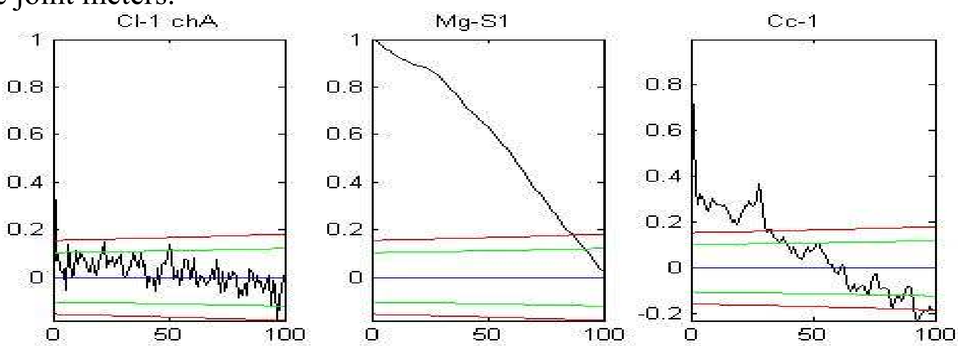

FIG. 6. Autocorrelations and their $\pm 2\sigma$ and $\pm 3\sigma$ confidence intervals for the clinometer Cl1-ChA (left), the joint meters Mg-S1 (center), and the load cell Cc-1 (right). On X-axis there is the j-delay and on the Y-axis the correlation values.

4. STATISTICAL MODEL FOR NORMAL DYNAMICS

We now introduce a statistical model which analyses the dynamics of the bridge in normal conditions, that is when the measurements are not affected by anomalies. The model considers the influence of the temperature on the data and estimates jointly the overall dynamics, showing the structural correlations, the residual dynamics and the measurement errors, which will be used in the next section to define the hierarchical multivariate detector (HMD).

Denoting by y_t the set of measurements at day *t*, we assume that the dynamics of the monitoring system can be modelled by the following equations:

$$\begin{aligned} y_t &= \mu_t + e_t \\ \mu_t &= d + Dx_t + \eta_t \\ \eta_t &= A\eta_{t-1} + \varepsilon_t \end{aligned} \qquad (1)$$

The component μ_t is the "signal" or the "state" of the physical system, correlated with the temperatures, or other exogenous variables which we denote by x_t.

The sets *d* and *D* are the mean levels of the measurements and the coefficient matrix, respectively. The component η_t is the "state of the system" adjusted for the effect of the temperature. It could represent the inertia of the system or the effect of other unobserved variables (such as traffic, wind, etc.). As we assume that *A* is a diagonal matrix, the latent variable η_t has a first order autoregressive structure. The components e_t and ε_t are Gaussian white noises with a cross-correlation structure given by Σ_e and Σ_η, respectively.

With regard to the estimation procedure of the model of Eq. (1) several methods can be used. In order to have a better identification of the model, in this work, we have chosen to estimate the model in two steps: first, we pre-estimated the *D* matrix via the ordinary least square method (OLS) and then we use the error prediction method (Ljung, 1999) for estimating the remaining parameters of the model.

In our application, the exogenous variables are represented by five sets of temperatures (T-Cl1, TCl2, T-S1, T-S5 and T-Mf2) which are characterized by a seasonal component and by a strong cross-correlation. As it is well known that this correlation can cause some computational problems in the estimation procedure, we decomposed the data in orthogonal components through the principal component analysis (Joliffe, 1999) and we used these "principal components" (PC) as the regressors of the model in the matrix x_t.

Table 1 shows the resulting estimates for the mean vector *d* and for the matrices *D* and *A* (the non significant coefficients are omitted). Concerning the estimation of the set *D*, we can note that the joint meters are positively correlated with the principal components of the temperature matrix, as expected. The estimates of the diagonal matrix *A*, being less then one, show a stable behaviour for the latent component η_t. Applying the Kalman filter algorithm to the model of Eq. (1), we get the estimates of the residuals e_t and their standard deviations (last column of Tab.1). The cross-correlation matrix of the residuals is given in Table 2. We note an approximately block diagonal structure, where the clinometers seem uncorrelated with the rest of the system. This correlation structure will be then used in the next Section for getting an hierarchical multivariate detector (HMD) which allow to provide a signal if a instrument is affected by anomaly.

Table 1. Estimates of the vector *d*, the matrix *D*, the diagonal of *A* and the standard deviations of the estimated residuals. The non significant values are omitted.

		Estimate of D						
	d	PC1	PC2	PC3	PC4	PC5	Diagonal(A)	StdDev(e_i)
Cl-1 chA	1.427			0.173				0.976
Cl-1 chB	-3.828					-0.182		0.971
Cl-2 chA	-3.158							0.992
Mg-S1	-1.558	0.986	0.133	0.027	-0.027		0.748	0.067
Mg-S2	-1.643	0.982	0.170				0.815	0.064
Mg-S4	-1.735	0.927	-0.078	0.063	-0.069	0.058	0.804	0.242
Mg-S5	-1.609	0.984	0.159			-0.015	0.844	0.051
Mg-GL1	-1.545	0.991	0.110	0.024	-0.025		0.704	0.057
Mg-GL2	-1.604	0.977	0.191				0.802	0.062
Cc-1	218.225	-0.233	0.373	0.312	-0.172		0.899	0.631
Cc-2	180.342	-0.409	0.551		0.137		0.808	0.494
Cc-3	178.543	-0.634	0.581		0.106	-0.107	0.829	0.468
Cc-4	186.729	-0.112	0.727	0.209			0.847	0.620
Cc-5	336.687	0.267	0.205	0.312			0.832	0.657

Table 2. Correlation matrix of the residuals. The non significant values are omitted.

	Cl-1B	Cl-2A	Mg-S1	Mg-S2	Mg-S4	Mg-S5	Mg-GL1	Mg-GL2	Cc-1	Cc-2	Cc-3	Cc-4	Cc-5
Cl-1A	0.18	-0.10											
Cl-1B	1		0.11										
Cl-2A		1											
Mg-S1			1	0.81	0.39	0.71	0.90	0.78	0.38			0.22	0.10
Mg-S2				1	0.19	0.90	0.72	0.95	0.33		0.12	0.25	0.10
Mg-S4					1	-0.04	0.47	0.21	-0.28	-0.15	-0.50	-0.38	-0.18
Mg-S5						1	0.61	0.84	0.36		0.21	0.27	0.11
Mg-GL1							1	0.65	0.33			0.20	
Mg-GL2								1	0.26			0.19	
Cc-1									1	0.56	0.79	0.91	0.69
Cc-2										1	0.68	0.73	0.78
Cc-3											1	0.92	0.71
Cc-4												1	0.77
Cc-5													1

5. HIERACHICAL MULTIVARIATE DETECTOR

Once the model of Eq. (1) has been estimated in normal conditions, a statistical monitoring system of the bridge is provided by an hierarchical multivariate detector (HMD) based on control charts.

In particular, we evaluate some daily statistics, based on the residuals e_t, given from the pre-estimated model of Eq. (1), for finding anomaly in a specific instrument, starting from the detection of the overall system.

First, we smooth the residuals e_t using the multivariate exponentially weighted moving average, EWMA (Lowery *et al.,*1992):

$$z_t = z(e_t) = \lambda e_t - (1-\lambda) z_{t-1}, \qquad (2)$$

where $z_0 = 0$ is the initial value of the system and $0 < \lambda \le 1$ is a smoothing parameter which reduce the largest stochastic oscillations. Then, we summarize the stability of the whole structure, based on z_t, by the following statistic

$$X_t^2 = z_t' \Sigma^{-1} z_t \qquad (3)$$

where Σ^{-1} is the inverse of the estimated variance-covariance matrix of z_t and the prime indicates transposition. The statistic X_t^2 represents the distance of any point from the mean target zero, that is the mean in stable conditions. Increasing values of X_t^2 suggest that the monitored system is going away from its target and it is probably going out of safety. The EWMA control charts allow to detect small shifts.

As the covariance matrix Σ is approximately a block diagonal matrix, the statistic in Eq. (3) can be approximately written as a sum of the statistics of each type of monitoring gauge,

$$X_t^2 = X_{(Cl)t}^2 + X_{(Mg)t}^2 + X_{(Cc)t}^2 \qquad (4)$$

In the following, we will call z_t the first level detector; $X_{(Cl)t}^2, X_{(Mg)t}^2$ and $X_{(Cc)t}^2$ the second level detectors, and X_t^2 the third level detector.

In this work, we use a hierarchical procedure for detecting and localizing structural anomalies which is able to both strongly control the false alarms rate (Lowry et al., 1992) and to diagnose which structure component is the source of the anomaly (Fassò, 1991; Fassò, 1992; Fassò, 1997; Hochberg and Tamhane, 1987) .

First, we consider the overall or third level detector. When this detector signals an anomaly, in order to understand its source, we check the second level detectors which are designed to detect the instrument responsible for the signal. After this, we check the first level detectors in order to identify which particular instrument is drifting.

In particular, first, we plot the third level detector X_t^2 on a control chart and we compare it with a threshold value $h^0\left(X^2\right)$. If the monitoring system is affected by an anomaly, the third level detector shows a persistent exceedance of this threshold value, starting from a temporal point $t = t^*$. Then, we plot each second level detector $X_{(i),t}^2$ and we compare it with its threshold value, $h^0\left(X_{(i)}\right)$. If $\left|X_{(i),t}^2\right| > h^0\left(X_{(i)}^2\right)$, we identify the source of anomaly in the type of instrument *i*. Lastly, by checking the first level detectors, we localize the anomaly for the instruments exceeding the corresponding threshold $h^0\left(z_j\right)$.

At each level of this hierarchical procedure, the threshold value is chosen in a such way that the probability of false detection is very low (Lowry *et al.*, 1992).

6. SIMULATION OF ANOMALIES

In order to illustrate the method and to show its ability to detect anomalies, two examples are considered. First, we affect some instrumental measurements by additive artificial shifts, starting from a certain day, then, we apply the monitoring procedure of Section 5 on the residuals of the pre-estimated model in Eq. 1.

In all cases, we assume that each anomaly starts from the 167^{th} day (18-09-2005) with a linear transient of 30 days. This means that the shifts increase day by day and they spend 30 days to reach their maximum value. In order to draw each control chart, we consider a smoothing parameter $\lambda = 0.01$ and a probability of false detection $p = 0.001$.

Such simulations are not intended to represent real situations of structural settlements, but to show the response of this monitoring system to a particular type of long term variations in the data.

6.1 Anomalies of the joint meters

The first example of simulation is based on a variation of +1.5mm and -1.5mm for the joint meters Mg-S2 and Mg-GL2, respectively, as in the Fig. 7. Such anomaly regards two sensors on two different sides of the bridge and therefore it could represent an effect of compensation in a small settlement.

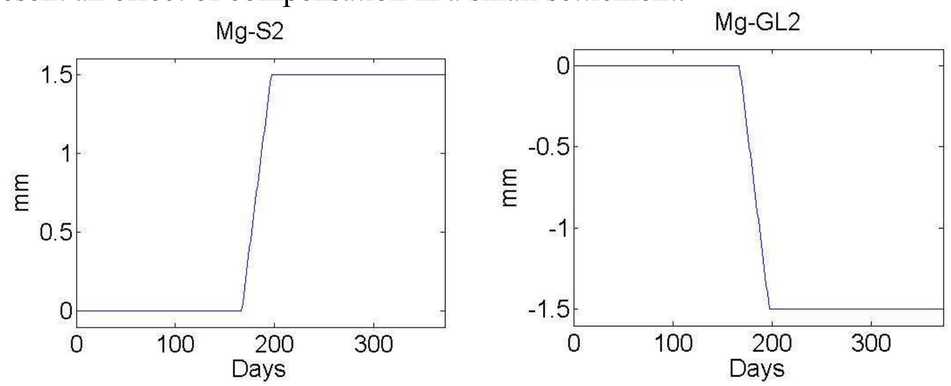

FIG.7. Anomalies of the joint meter Mg-S2 (left) and Mg-GL2 (right).

The results of the hierarchical control chart procedure are the followings.

- Fig. 8 (a) shows the behaviour of the third level *EWMA* detector. This detector exceeds the threshold value (straight line), starting from day 193. This means that it detects the anomaly when the transient have reached the 90% of the maximum settlement.
- The diagnostics based on the second level detectors (in Fig. 8 b, c, d) shows that the anomaly clearly comes from the joint meters.
- From the plot of the first level detectors (Fig. 9) it is easy to see that the origin of the anomaly is in the joint meters Mg-S2 and Mg-GL2, the only ones above the threshold.

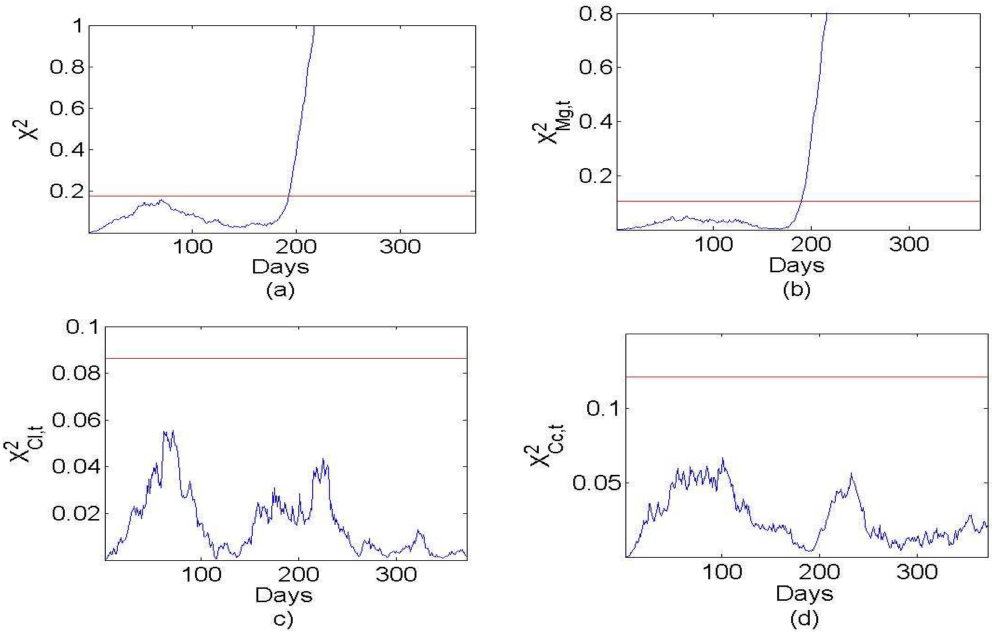

FIG. 8. MEWMA detectors (with $\lambda = 0.01$): third level detector with detection at *t=193* (a); second level detector for the joint meters with detection at *t=190* (b); second level detector for the clinometers (c) and second level detector for the load cells (d).

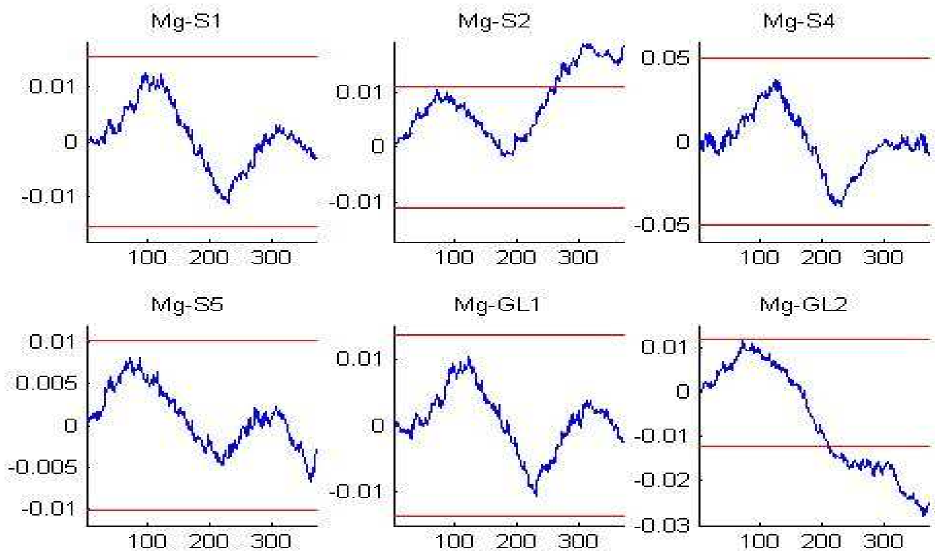

Figure 9: First level detector for each joint meter

6.2 Anomalies of the clinometers

The second simulation example is based on a variation of σ, that is equal to the standard deviation of the data. In particular, we added +σ to the mean of the clinometer Cl1-ChA and -σ to the clinometer Cl1-ChB.
In this case, the third level *MEWMA* detects the anomaly at the 208^{th} day such as in Fig. 10 (a). Moreover, the second level detector localizes the problem in the clinometer set (in Fig. 10 b, c, d). Finally, the first level detector (in Fig. 11) clearly shows the anomaly in the clinometers CL1-ChA and ChB.

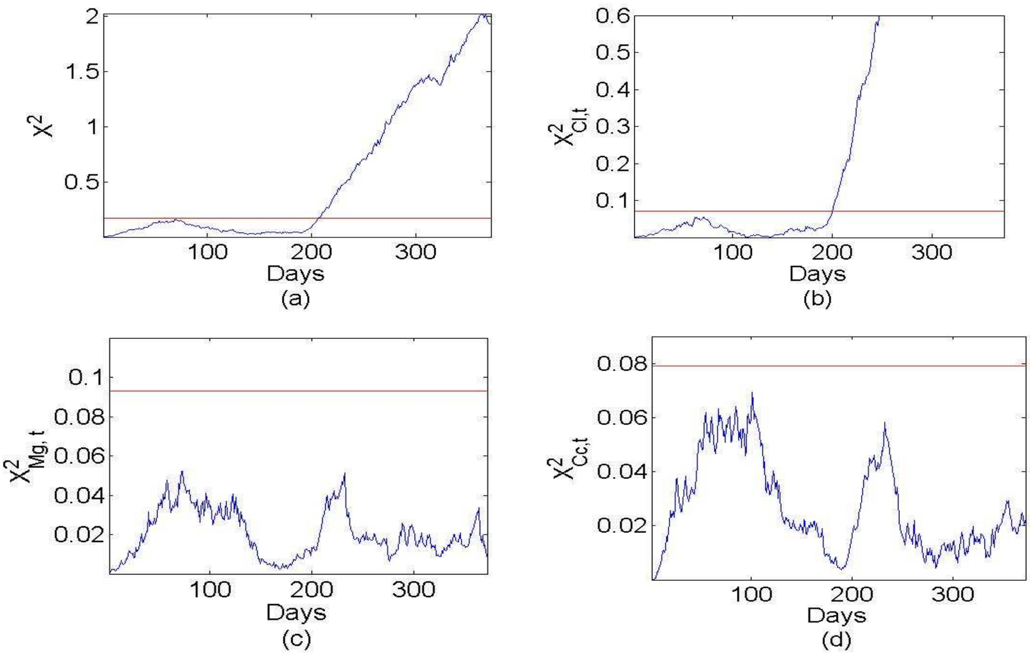

FIG. 10. MEWMA detectors (with $\lambda = 0.01$): third level detector with detection at *t=208* (a) and second level detector for clinometers with detection at *t=201* (b); second level detector for the joint meters (c) and second level detector for the load cells (d).

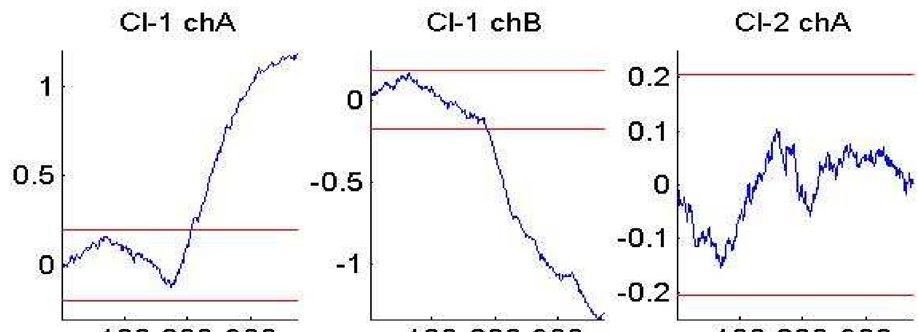

FIG. 11. First level detector for each clinometer

7. CONCLUSIONS

The study has shown that by a suitable statistical monitoring system it is possible to jointly monitor instruments that provide data with different dynamics.
From the preliminary analysis we have noted that the joint meters have a remarkable quote of variability which is due to temperature effects; the clinometers have not significant correlation with the temperature and the behaviour of the cargo cells are in a intermediate position between the other cited instruments.
Then, we have proposed a statistical monitoring procedure called hierarchical multivariate detector (HMD) which utilizes the residuals of an integrated model, estimated in normal conditions. The hierarchical procedure is based on control charts at different levels. The example developed on three levels strongly controls the probability of false signals and allows to detect small structural drifts and to localize the anomaly, after adjusting for correlation and environmental effects.
From the results of the control chart examples analyzed, we can conclude that our approach allows to successfully detect small shifts in the data.

REFERENCES

Fassò A. (1991). System Identification and Rupture Diagnostic: a Case Study. Riv. Statist. Appl., Vol. 3, n.4, 7-21.

Fassò A. (1992). Localizing Ruptures in Block Stochastic Systems. *J. Ital. Statist. Soc.*, Vol. 1, n.2.

Fassò A. (1997). On some control charts for nonlinear ruptures. *Italian J. Appl. Statist.*, Vol. 9, n.1, 123-141.

Fassò A., Locatelli S. (2007) Asymmetric Monitoring of Multivariate Data with Nonlinear Dynamics. Journal of German Statistical Society (AStA) . Vol.91:1, 23-27.

Fassò A., Nicolis O., Bruzzi D., Pezzetti G. (2004) Statistical modelling and uncertainty reduction of monitoring data in geomechanics, Working paper n.3/MS, Dept. IGI, University of Bergamo (www.unibg.it)

Fassò A., Nicolis O., Bruzzi D., Pezzetti G. (2005) Modelling and reducing uncertainty of field monitoring data in geomechanics by computerized statistical methods. In Proceeding of 11th IACMAG Conference, Torino 19-24 June 2005. Vol.1, 595-602. Edited by Barla G. and Barla M. (2005) *Prediction, analysis and design in geomechanical applications*. Pàtron, Bologna.

Fuchs C., Kenett R.S. (1998) Multivariate quality control. Theory and Applications. Marcel Dekker, New York.

Hochberg Y., Tamhane A.C. (1987). *Multiple comparison procedures*. Wiley, New York.

Joliffe I. T. (1986). *Principal Component Analysis. Series in Statistics*. Springer.

Ljung, L (1999). System Identification, Theory for the User, 2nd ed. Prentice Hall.

Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Chart, *Technometrics*, 34, 46-53.

Montgomery, D.C. (2005). *Introduction to statistical quality control*, Fifth Edition. John Wiley and Sons, New York.