Development and field application of a multivariate statistical process control framework for health-monitoring of transportation infrastructure

Yikai Chen\(^a\), Pablo L. Durango-Cohen\(^b,\ast\)

\(^a\)FedEx Services, Memphis, TN 38116, USA
\(^b\)Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, A332, Evanston, IL 60208, USA

**Abstract**

We present a two-part multivariate statistical process control framework to support health-monitoring of transportation infrastructure. The first part consists of estimation of regression and ARIMA–GARCH models to explain, predict, and control for common-cause variation in the data, i.e., changes that can be attributed to usual operating conditions, including traffic loads, environmental effects, and damage when present throughout the data. The second part of the framework consists of using multivariate control charts to simultaneously analyze the standardized innovations of the aforementioned models in order to detect possible special-cause or extraordinary events, such as unique/infrequent traffic, weather, or the onset of damage. The proposed approach revolves around construction of $T^2$ control charts as a framework to jointly monitor the evolution and contemporaneous correlation of a set of measurements. The approach provides significant practical/computational advantages over individual analysis of multiple structural properties, and addresses technical problems stemming from ignoring the relationships among them.

To illustrate the framework, we analyze strain and displacement data from the monitoring system on the Hurley Bridge (Wisconsin Structure B-26-7). Data were collected from 1 April 2010 to 28 June 2012. Analysis of 7 measurement sequences collected over the 27 month planning horizon revealed 6 possible special-cause events. In terms of outlier interpretation, we use Mason–Young–Tracy Decomposition to establish the contribution of (subsets of) the measurements. Also, we link the most significant special-cause events, in terms of magnitude and duration, to unusual changes in weather and traffic. To conclude, we compare the proposed approach and benchmark the empirical results with Principal Component Analysis, perhaps, the most common alternative appearing in the literature.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In Chen et al. (2014), we present a generally-applicable statistical process control (SPC) framework to support structural health monitoring (SHM) of transportation infrastructure. The framework links the literature on statistical performance modeling to support infrastructure/asset management with the literature on SHM, and consists of two parts: The first overlaps with the former, and involves estimation of statistical models to explain, predict, and control for common-cause variation in...
the data, i.e., changes that can be attributed to usual operating conditions, including traffic loads, environmental effects, and damage\(^1\) when present throughout the data. As explained in the literature, this analysis can be used to support a plethora of managerial decisions, e.g., planning maintenance (cf. Durango-Cohen (2007) and the references therein). The second part of the framework overlaps with the SHM literature, and consists of using control charts to analyze/monitor the prediction errors of the aforementioned statistical models in order to detect possible special-cause or extraordinary events. Special causes can include unique or infrequent traffic or weather events, as well as the onset of damage. In an online monitoring setting, detection of special causes could lead to scheduling (equipment or structural) inspections, road closures, etc. As applied herein, post hoc, this analysis provides forensic capabilities, and supports assessment and refinement of statistical performance models.

In Chen et al. (2014), our implementation of the second part of the framework used univariate control charts to analyze measurement sequences, one at a time. In contrast, in the current paper, we report on the implementation of multivariate statistical process control (MSPC) as a framework to simultaneously analyze/monitor multiple sequences. This improves the capabilities of the framework, as well as its (computational) appeal in cases where large numbers of measurement sequences are available. Specifically, MSPC provides a statistically-rigorous framework to:

- Aggregate response measurements in order to reduce data dimensionality, and thus, the complexity of simultaneously monitoring multiple sequences. The importance of this issue is growing as the availability of technologies that allow for reliable, long-term, high-frequency/continuous monitoring of various properties increases, and the associated costs to deploy SHM systems decreases. As an example to emphasize this point, we note that the SHM system installed on the St. Anthony Falls Bridge, completed in September of 2008 to replace the I-35W Bridge in Minneapolis, MN, consists of 328 sensors (Inaudi et al., 2009; Hedegaard et al., 2010).
- Describe, and subsequently, monitor the relationships among response measurements. A characterization of such relationships, i.e., the contemporaneous correlation among measurements, provides additional benchmarks to assess the consistency (in the progression) of a set of measurements, and thus, for the detection and interpretation of special-cause events.
  - From a managerial perspective, added capabilities to interpret special-cause events are important in SHM because inspection can be impractical/costly/dangerous. As an example, we note that, in field applications, sensing/data-collection exhibits variation from exposure to the environment, traffic, and other factors. In such situations, analysis of the relationships among measurements may shed light on whether special-cause variation can be attributed to changes in underlying structural responses, or to changes in the sensing/data-collection process, e.g., sensor failure or calibration changes, and explains the logic behind the deployment of redundant sensors. Importantly, this type of analysis can guide resource allocation decisions, such as the type of specialist—technician or bridge engineer—that should conduct further examination.
  - From a technical perspective, characterization of the relationships among the measurements is necessary to set the parameters, i.e., the Type I error rate, that determine the overall performance of the analysis/monitoring scheme. The Type I error rate determines the level of confidence with which outliers can be interpreted to be indications of special-cause events (as opposed to random variation), as well as the average run length (ARL), i.e., the expected number of measurements between outliers. Because measurements in a multivariate process can exhibit complex interactions, in general, it is not possible to control the performance of the overall monitoring scheme by adjusting the Type I error rates of univariate control charts. As a result, inferences drawn from monitoring measurements individually may be incorrect or unsubstantiated.

To illustrate the use of MSPC to support SHM of transportation infrastructure, we apply the framework to analyze strain and displacement data from the monitoring system on the Hurley Bridge (Wisconsin Structure B–26–7). Data were collected from 1 April 2010 to 28 June 2012 (820 days). As in Chen et al. (2014), our implementation of the first part of the framework consists of sequential estimation of regression and ARIMA–GARCH models to explain and control for common-cause variation. The regression models relate variability in the response measurements to changes in explanatory variables: ambient and steel temperature, humidity, and traffic. ARIMA–GARCH models are used to control for serial dependence, which is, perhaps, the most appealing approach to monitor multivariate processes (Mason and Young, 2002), i.e., the evolution and contemporaneous correlation of a set of quality characteristics. Analysis of 7 measurement sequences collected over the 27 month planning horizon revealed 6 possible special-cause events. In terms of outlier interpretation, we use Mason–Young–Tracy (MVT) Decomposition to establish the contribution of (subsets of) the measurements. In terms of attribution, we link the most significant special-cause events, in terms of magnitude and duration, to unusual changes in weather and traffic.

\(^1\) Wenzel (2009) defines damage as changes to material or geometric properties of a structure, including changes to the boundary conditions and connectivity. Damage affects current or future performance of such systems, i.e., load-bearing capacity. Herein, deterioration refers to the damage initiation and progression process.
To complement the analysis, we implement Multivariate Exponentially Weighted Moving Average (MEWMA) and Generalized Sample Variance (GSV) control charts. We use the former to assess the duration of special-cause events, as well as to monitor for possible small-magnitude, persisting events. The latter provides a framework to monitor the (evolution of within sample) covariance among measurements, i.e., their dispersion and contemporaneous relationships.

The remainder of this paper is organized as follows: We begin in Section 2 by positioning our work with respect to studies where SPC is used for damage detection in SHM, or where multiple structural properties are monitored/analyzed simultaneously. Section 3 provides a summary of the methodological underpinnings of the proposed framework. A description of the data, as well as the results of our analysis are presented and discussed in Section 4. Conclusions and additional opportunities are presented in Section 5. The paper also includes 3 appendices: In Appendix A, we present the results of the statistical tests on the standardized innovations to ensure compliance with the assumptions underlying the use of control charts. Appendix B describes the analysis that we use to link unusual traffic and weather to special-cause events. In Appendix C, we use Principal Component Analysis to benchmark the results presented in Section 4.

2. Related work

Technological advances in the last two decades have motivated the development and adaptation of instruments, equipment, as well as data processing and analysis methods for SHM. The literature documenting these efforts is vast and growing rapidly. In terms of technological aspects, as a starting point, we refer the reader to Balageas et al. (2006), who review and classify SHM systems based on their sensing technologies. Notably, what has improved in the last decade is sensor reliability, enabling continuous, long-term, simultaneous monitoring of various structural properties, and, in turn, motivates the need to develop rigorous and computationally-efficient methods with general applicability to process different data types emanating from multiple sources. In terms of data processing, following Kullaa (2011), at a high level, two approaches have emerged to support SHM: physics-based and empirical. We begin this section by providing background on physics-based approaches. We then position our work with respect to empirical methods to support SHM paying special attention to studies where (i) SPC, i.e., control charting, is used to detect (possible) damage, or (ii) where multiple response measurements are analyzed simultaneously.

The majority of physics-based methods to support SHM rely on analyzing vibrations induced by traffic or environmental loads, or by mechanical devices in the case of lab specimens. The premise is that a structure’s response, i.e., modal parameters such as resonant frequencies, mode shapes and modal damping are functions of its mass, damping and stiffness, and that changes, e.g., reductions in stiffness resulting from the onset of cracks or loosening of connections, cause detectable changes in the modal parameters characterizing the vibrations (Doebbling et al., 1998). Fritzen (2005) provides an overview of ancillary analytical and computational methods used to infer location, type and magnitude of damage. Analytical methods have performed well at characterizing damage in controlled environments, e.g., lab specimens or rotating equipment in manufacturing settings (Wowk, 1991). In terms of field applications, Friswell and Mottershead (1995) describe finite-element model updating as, perhaps, the most promising technique. It involves calibrating a finite-element model of a structure with response measurements collected for a reference, “undamaged” state. Subsequent deviations between measurements and predictions, lead to model updating to achieve agreement, and in turn, shed light on plausible damage scenarios. A fundamental limitation of physics-based methods is that, in general, they are not rich enough to explain (long-term) variations in response measurements exhibited by complex structures, even when they are undamaged and operating in normal conditions. This lack of explanatory capability hinders damage detection because it is difficult to separate damage from other sources of variation, and in turn, motivates the need to develop alternative methods to explain and predict the progression of measurements/properties.

In terms of empirical methods to support SHM, we follow Kullaa (2011) who describes the following 4 (overlapping) approaches to deal with common-cause variation in the progression of structural response measurements:

1. Consider response measurements or surrogates that are sensitive to damage, but insensitive to operational conditions;
2. Group measurements obtained under similar conditions;
3. Relate common-cause variation to observed/measured (changes in) operational factors; and
4. Describe common-cause variation in the response measurements without relating it to (measured) explanatory factors.

While seemingly trivial, the first approach is common in the literature, where, as in physics-based methods, the focus is on the analysis of structural vibrations. The second approach relies on the assumption that facilities operate in a finite set of (latent) reference states (i.e.: undamaged or combinations of damage scenarios and operational conditions), where they yield stationary measurements. A variety of statistical methods have been used, often in controlled environments such as The ASCE Benchmark Structure (Johnson et al., 2004), to cluster measurements and to make inferences about the underlying states, e.g., identifying damage levels. Kiremidjian (2009) provides a recent overview of this work.

---

2 This limitation does not undermine the analysis of structural vibrations because modal parameters are (relatively) stationary and insensitive to operational variation, including minor damage (Farrar et al., 2005). Additional advantages of analyzing vibrations are the simple requirements in terms of instrument type, (i.e., accelerometers,) and location (because vibrations are global properties that can be monitored remotely). Location flexibility facilitates deployment of tethered instruments, which enable data and power transmission, and contributes to the (historical) appeal of these methods for field applications. As a pioneering example, we cite Vincent (1958, 1962) who analyzed the vibrations of the Golden Gate Bridge.
The last two approaches consist of characterizing common-cause variation in the progression of measurements. The ensuing models are frequently used in conjunction with control charts to detect (possible) unexplained measurements. Kullaa (2011) describes approach 3, where (variation in) response measurements are explained by (changes in) measured explanatory variables, as the most obvious; however, he recognizes the potential for alternative specifications due to complex relationships among the variables. He also points out that unobserved explanatory variables can lead to difficulties in describing or predicting common-cause variation, which has motivated the development of models relying exclusively on the response measurements, i.e., approach 4. Examples of such models include polynomial regression models (Peeters and De Roeck, 2001; Worden et al., 2002), pure time series models (Sohn et al., 2000b,a; Fugate et al., 2001), Principal Component Analysis (Yan et al., 2005), factor analysis (Kullaa, 2003), and neural networks (Sohn et al., 2003). Peeters and De Roeck (2001) and Worden et al. (2002) are parts of series of seminal studies of environmental effects on the resonant frequencies of the ZZ4 Bridge in Switzerland and the Alamosa Canyon Bridge in USA, respectively. As is done herein, both studies mention the use of polynomial or exponential regression models to capture trends in the responses. Also overlapping with the work herein, availability of temperature data allowed for estimation of auto-regressive with exogenous inputs (ARX) models capturing serial dependence, including the dynamic effects of temperature (gradients). Control charts are used in subsequent studies to detect the effects of deliberate and significant damage scenarios on the models’ prediction errors. In Sohn et al. (2000b,a), Fugate et al. (2001), pure time series models are used to control for autocorrelation in acceleration measurements collected from lab specimens in controlled testing situations. These studies are important because they were first to develop control charts for damage detection in SHM. Sohn et al. (2000a) is related to Yan et al. (2005) and is highly relevant because multiple measurements are analyzed simultaneously. As discussed below, projection techniques, such as PCA, are used to construct statistics that aggregate measurements. Finally, in factor analysis and neural networks, response measurements are represented as manifestations of (a small number of) latent variables representing environmental factors. These relationships are first calibrated with training data, and subsequently monitored. Interested readers are referred to Sohn (2007), Farrar and Worden (2013) where these studies are reviewed in great detail.

The two-part framework of Chen et al. (2014) and used herein builds on Alwan and Roberts (1988), and has elements of approaches 3 and 4. The first part of the framework consists of formulation and estimation of statistical models to explain, predict and control for common-cause variation in the response measurements. For simplicity, our implementation consists of sequential estimation of regression and ARIMA–GARCH models. Regression models are used to relate variability in the measurements to changes in measured explanatory variables. We then use ARIMA–GARCH models to control for serial dependence, which is a significant source of common-cause variation, not explained by the regression models. While the sequential approach is adequate to control for common-cause variation, it is relevant to mention synergistic and highly-sophisticated (panel data) models appearing (recently) in the statistical performance modeling literature, and where, aside from predicting deterioration, the objective is to estimate the effect of structural design, maintenance, traffic, or environmental conditions, including their interactions, on various performance criteria, including damage initiation and progression, functional, economic or environmental impact. In particular, limitations associated with sequential estimation are addressed by the closely-related works of Chu and Durango-Cohen (2008), Kobayashi et al. (2015). The ARIMAX-GARCH model of Kobayashi et al. (2015) is most relevant. Significantly, the authors develop a sophisticated Bayesian approach to reduce the computational effort associated with simultaneous estimation. Chu and Durango-Cohen (2008) deals with pooling data across a panel of facilities, and estimating the (carryover) effects of exogenous variables, and is closely-related to the recent work of Anastasopoulos and Manningker (2015). Other recent papers addressing relevant technical issues, such as dealing with measurement errors, include Kobayashi et al. (2014), Lethanh et al. (2015).

Our earlier implementation of the second part of the framework used univariate control charts to analyze the ensuing prediction errors, i.e., the standardized innovation sequences, for detection of (possible) special-cause events. A fundamental assumption underlying the use of control charts is that the quality characteristics being analyzed are generated by a stationary and random process, i.e., that the data follow a distribution that is stationary, serially-independent and identically-distributed. While the idea of using SPC to support SHM is not original to Chen et al. (2014), verification of compliance with this assumption, and overall assessment/diagnosis of the results, e.g., average run lengths, are rarely reported in the literature. Among other contributions, this led to the estimation of the first model of conditional heteroscedasticity, i.e., time-varying volatility, in the context of transportation infrastructure. Far from an obscure technical point, we present empirical examples showing that this level of detail can be necessary to ensure reliable performance of control charts. It also contributes (along with high-quality data collection) to meaningful results, where despite analyzing distinct structural properties of bridge elements at separate locations, we report consistent detection (across multiple response measurements) of special-cause events (with transient effects), and are able establish correlations with the incidence of extraordinary traffic or weather. Benchmark studies are often concerned with the analysis of structural vibrations, and use control charts to detect large, permanent shifts in measurement levels caused by deliberate damage to structures/lab-specimens.
We also observe that the quality characteristics analyzed in Chen et al. (2014) differ from those used in other studies. We analyze the progression of prediction errors, i.e., standardized innovations. Control charts can, therefore, be used as tools to assess the adequacy of, and when necessary correct/refine model specifications used to characterize common-cause variation. Moreover, given an adequate model and a predetermined confidence level, it is correct to interpret outliers as indications of special-cause events. Significantly, this approach is applicable to arbitrary response-measurements/structural-properties. In contrast, several studies have focused on the analysis of established, though highly-specific “damage sensitive features”, e.g., the coefficients of auto-regressive models used to describe acceleration profiles of a structure’s vibrations (Sohn et al., 2000a).

In this paper, we report on the development and field application of a MSPC framework to support SHM of transportation infrastructure. The proposed framework addresses practical and technical problems in Chen et al. (2014), where univariate control charts are used to analyze measurement sequences, one at a time. Here, we use Hotelling $T^2$ control charts to aggregate (into a scalar) and simultaneously monitor the progression of a set of prediction errors obtained from statistical models of common-cause variation. In terms of innovative application of state-of-the-art methodologies, we highlight the use of Mason–Young–Tracy (MYT) Decomposition to identify (subsets of) measurements contributing to outliers (Mason et al., 1995). This is not only appealing, but can have significant practical implications in the interpretation of special-cause events, and ensuing resource allocation decisions. To complement the analysis, we implement Multivariate Exponentially Weighted Moving Average (MEWMA) and Generalized Sample Variance (GSV) control charts. We use the former to assess the duration of special-cause events, as well as to monitor for possible small-magnitude, persisting events. The latter chart provides a framework to monitor the (evolution of sample) covariance among measurements, i.e., their dispersion and contemporaneous relationships. The study by Kullaa (2009) is the most relevant benchmark. He uses a variety of MSPC tools, including $T^2$ and MEWMA Charts to analyze the vibrations of a vehicle crane tested under controlled damage scenarios and operational loads.

There have been a handful of recent studies focused on simultaneous analysis of multiple structural properties. The main approach involves using PCA to project a set of contemporaneous measurements onto a subspace of orthogonal principal components of the covariance matrix of the set of measurements (Sohn et al., 2000a; Yan et al., 2005). Subsequently and typically, univariate control charts are used to monitor the progression of the projected measurements/errors, one dimension at a time; though to improve the practical/computational appeal and to avoid technical problems (stemming from the relationships among the measurements), the analysis is usually restricted to a single dimension, i.e. the most significant component. One of the key disadvantages of analyzing projections is that it is not (always) possible to make inferences that apply in the original space, which means that it is not always possible to identify (subsets of) measurements contributing to outliers. It is also relevant to note that (raw) projections may not comply with the assumptions for control charting. Nevertheless, we implement PCA as a benchmark, and discuss its weaknesses in Appendix C.

3. Methodology

We begin this section by presenting a conceptual overview of our implementation of the framework of Chen et al. (2014) consisting of the following 2 steps:

1. Estimation of statistical models to control for common-cause variation; and
2. Use of SPC tools to detect special-cause events.

In terms of our implementation, the approach to characterize common-cause variation consists of sequential estimation of regression and ARIMA–GARCH models. Because the analysis overlaps with Chen et al. (2014), and because it depends on the data and relies on statistical tools that are familiar to transportation researchers, details, including the model specifications and assessment, are relegated to the empirical study in Sections 4.2 and A.9 We highlight that, in spite of limitations partly stemming from the implementation’s simplicity, the results of the tests presented in Appendix A confirm the adequacy of the approach to yield statistics complying with the assumptions underlying the use of control charts, and thus, necessary for the reliable detection of special-cause events.

Our implementation of the second part of the framework, presented in Section 4.3, revolves around the use of the Hotelling $T^2$ control chart to simultaneously analyze the standardized innovations of the aforementioned ARIMA–GARCH models. As discussed in MacGregor and Kourtì (1995), Lowry and Montgomery (1995), Mason and Young (2002), the $T^2$ control chart is, perhaps, the most appealing approach to monitor continuous, steady-state, multivariate processes. In addition to providing an integrated approach to analyze/monitor the central tendency, dispersion, and contemporaneous correlation of a set of quality characteristics, the $T^2$ control chart is appealing because it supports complementary analyses.

---

7 Readers interested in comprehensive reviews of the MSPC literature are referred to Lowry and Montgomery (1995), Bersimis et al. (2007) and Chen (2014). MacGregor and Kourti (1995), Mason and Young (2002) are examples of references reporting industrial applications of MSPC. Significantly, there appears to be a dearth of applications where data display common-cause variation.

8 Similar studies are reviewed in Mujica et al. (2008).

9 The implementation in Kosnik et al. (2014) relies on seasonal ARIMA models to control for common-cause variation, and thus, illustrates the dependence of the analysis on the data.
In particular, there are a number of procedures to identify (subsets of) measurements contributing to outliers.\footnote{The most common alternatives for MSPC involve the use projection methods, e.g., PCA, to aggregate the measurements. As discussed earlier and in Appendix C, one of the key disadvantages of analyzing projections is that it is not (always) possible to make inferences that apply to the original measurement space. It is also relevant to note that (raw) projections may not comply with the assumptions for control charting.} In this paper, we follow Mason et al. (1995) and apply Mason–Young–Tracy (MYT) Decomposition to separate the quality characteristic into independent components associated with the individual measurements, and with the relationships among them. Also, following Lowry and Woodall (1992), we consider MEWMA control charts (along with independent components associated with the individual measurements, and with the relationships among them) to (detect possible small-magnitude changes, as well as to) assess the nature of special-cause events, i.e., if they are transient or persisting.\footnote{As discussed in MacGregor and Kourti (1995), Multivariate CUSUM charts are an alternative for this analysis. We selected MEWMA charts because the analysis involves construction of \( T^2 \) charts.} Also, we use GSV charts as an approach to focus on detection of special-cause events that producing changes in the (evolution of within sample) covariance among measurements, i.e., their dispersion and contemporaneous relationships.

Having discussed the framework’s implementation, the remainder of the section is organized as follows: First, we introduce basic notation in the context of providing an intuitive explanation of the use of control charts to detect special-cause events. We then describe the construction of \( \bar{x} \) and \( s \) control charts, which we use as benchmarks to draw insights from the multivariate analysis. We then overview the use of \( T^2 \) control charts as a rigorous statistical approach to monitor the measurements and to detect special-cause events. We also discuss the use of the MYT Decomposition to identify sets of measurements contributing to outliers. Finally, we discuss MEWMA and GSV control charts to complement our analysis.

### 3.1. Use of control charts

Control charts are “graphical displays” of the evolution of quality characteristics over time. They provide a statistical basis to detect special-cause events, which result in departures from a stationary and random data-generating process. Various control charts have been designed to test for particular types of departures. To illustrate, following Montgomery (2009), we present a generic control chart in Fig. 1.

In the example, \( x_w^t \) is a random variable, with mean \( \mu_w \) and standard deviation \( \sigma_w \), that represents a measurement of quality characteristic \( w \) collected at time \( t \). In addition to displaying the time series, \( \{x_w^t\}_{t=1}^T \), control charts include reference levels: a Center Line (CL) set at \( \mu_w \), Upper and Lower Control Limits (UCL and LCL) set at \( \mu_w \pm L\sigma_w \), \( L \), a distance from the CL measured in multiples of the process standard deviation, is chosen so that measurements outside the control limits, i.e., outliers, indicate possible deviations to the assumption of a stationary process, i.e., changes in the mean or increases in the variance of the process. It is common practice to set \( L = 3 \), which means that for \( iid \) measurements following a normal distribution, the probability that a measurement falls outside of the control limits is 0.027% (0.135% in each direction). The choice of \( L \) trades off the probability of Type I and Type II errors. Type I errors, false positives, arise when one incorrectly concludes that the process has deviated from stationary, \( iid \) behavior; whereas Type II errors, false negatives, arise when deviations are not detected. Using a larger \( L \) decreases Type I, but increases Type II errors.

### 3.2. Single variable control charts

\( \bar{x} \) Charts. are used to monitor the mean value of a quality characteristic. \( \{\bar{x}^T\}^T_{t=1} \) is the time series being monitored, where \( N \) corresponds to the number of measurements, each identified with argument \( n \), collected in sampling period \( t \).\footnote{For simplicity, we consider a constant sample size, \( N \).} \( \bar{x}^T \), therefore, corresponds to the average measurement of characteristic \( w \) taken over the number of units/measurements collected in sampling period \( t \); that is,

\[
\bar{x}^w_t = \frac{\sum_{n=1}^N x^n_w(n)}{N} \tag{1}
\]
Again, following the assumption that quality characteristic $w$ follows a normal distribution with mean $\mu_w$ and standard deviation $\sigma_w$, $\bar{x}_w$ is a Normal random variable with mean $\mu_w = \mu_w$ and standard deviation $\sigma_w = \frac{\sigma_w}{\sqrt{n}}$. The reference levels in $\bar{x}$ charts are analogous to those used in the construction of Fig. 1. That is, $CL = \mu_w$ and $UCL = \mu_w + L\sigma_w$. There is often a need to estimate $\mu_w$ and $\sigma_w$. In this paper, we rely on the following estimates presented in Montgomery (2009):

$$\hat{\mu}_w = \bar{x}_w = \frac{\sum_{t=1}^T x_w^t}{T}$$  \hspace{1cm} (2)

$$\hat{\sigma}_w = \frac{\hat{\sigma}_w}{\sqrt{N}} = \frac{\tilde{y}_w}{c_4\sqrt{N}}$$, where:

$$s_w^2 = \frac{\sum_{n=1}^N (x_w^t(n) - \bar{x}_w^t)^2}{N - 1} \hspace{1cm} (4)$$

$$\tilde{s}_w = \frac{\sum_{t=1}^T s_w^t}{T}$$  \hspace{1cm} (5)

where $s_w^t$ is the standard deviation of type $w$ measurements collected in $t$, and $\tilde{s}_w$ is the average over all sampling periods. $c_4$ in Eq. (3) is a bias correction factor calculated as:

$$c_4 = \sqrt{\frac{2}{N - 1} \frac{\Gamma(N/2)}{\Gamma((N - 1)/2)}}$$  \hspace{1cm} (6)

where $\Gamma(\cdot)$ is the Gamma Function.

$s$ Charts are complementary charts used to monitor the (progression of the within sample) standard deviation of quality characteristic $w$, i.e., $(s_w^t)_{t=1}^T$. Relying on the assumption of a normally distributed quality characteristic, we consider the random variable:

$$\lambda_t = \frac{\sqrt{N - 1} \cdot s_w^t}{\sigma_w}$$  \hspace{1cm} (7)

which follows a Chi distribution with $(N - 1)$ degrees of freedom. It can be shown that $\mu_w = c_4\sigma_w$ and $\sigma_w = \sigma_w\sqrt{1 - c_4^2}$, which, combined with the estimate in Eq. (3), yields the following reference levels for the $s$ control chart:

$$UCL = \tilde{s}_w + \frac{\tilde{s}_w}{c_4} \sqrt{1 - c_4^2}$$  \hspace{1cm} (8)

$$CL = \tilde{s}_w$$  \hspace{1cm} (9)

$$LCL = \tilde{s}_w - \frac{\tilde{s}_w}{c_4} \sqrt{1 - c_4^2}$$  \hspace{1cm} (10)

### 3.3. Hotelling $T^2$ control charts

Rather than monitoring $w$ measurement sequences, one at a time, the development of multivariate control procedures/charts relies on the construction of aggregate statistics. Various statistics have been devised to monitor the central tendency, dispersion and correlation structure in multivariate processes. The $T^2$ control chart, a multidimensional analog of the $\bar{x}$ Chart, is, perhaps, the most appealing one (Mason and Young, 2002). In terms of notation, the problem involves monitoring $j$ variables $x_1, x_2, \ldots, x_j$ and constructing the following statistic, which corresponds to the square of the Mahalanobis Distance (Mahalanobis, 1930):

$$D^2_{M}(t) = N \cdot ([\bar{x}_t - \bar{M}]^\prime \Sigma^{-1} [\bar{x}_t - \bar{M}])$$  \hspace{1cm} (11)

Again, for simplicity, we consider a constant sample size of $N$ for all measurements in all periods. Under the assumption that the $w$-dimensional measurements are drawn from a multivariate normal distribution with mean $\bar{M}$ and covariance $\Sigma$, $D^2_{M}(t)$ follows a Chi-square distribution. As was the case earlier, there is often a need to estimate process parameters, $\bar{M}$ and $\Sigma$, from the data. In these cases, the analogous distance measure is referred to as a $T^2$ statistic. In period $t$, the statistic $T^2(t)$ is calculated as follows:

$$T^2(t) = N \cdot ([\bar{x}_t - \bar{M}]^\prime \Sigma^{-1} [\bar{x}_t - \bar{M}])$$  \hspace{1cm} (12)

where $\bar{M} = [\bar{x}_1, \ldots, \bar{x}_w]$. Hereafter, we drop the argument $t$ from $T^2(t)$ to simplify the notation. Estimates of the covariance between variables $w_1$ and $w_2$ within the $t$-th subgroup are calculated as follows:

$$S_{t}^{w_1,w_2} = \frac{1}{N - 1} \sum_{n=1}^N (x_{w_1}^t(n) - \bar{x}_{w_1}^t) \cdot (x_{w_2}^t(n) - \bar{x}_{w_2}^t)$$  \hspace{1cm} (13)
We note that when \( w_1 = w_2, s_{tw}^{w_1} = s_{tw}^{w_2} \) is measurement \( w_1 \)'s sample variance in period \( t, (s_{tw}^{w_1})^2 \). Overall estimates of the covariances are constructed by averaging over the \( T \) periods. The estimates can be organized in the form of a \( W \times W \) symmetric matrix, \( S \), as follows:

\[
S = \begin{bmatrix}
    \bar{s}_{11} & \bar{s}_{12} & \ldots & \bar{s}_{1W} \\
    \bar{s}_{21} & \bar{s}_{22} & \ldots & \bar{s}_{2W} \\
    \vdots & \vdots & \ddots & \vdots \\
    \bar{s}_{W1} & \bar{s}_{W2} & \ldots & \bar{s}_{WW}
\end{bmatrix}
\]  

(14)

Under the assumption that the measurements are drawn from a multivariate normal distribution, it can be shown that:

\[
T^2 \sim \frac{(T - 1) \cdot (N - 1) \cdot W}{T \cdot (N - 1) - W + 1} \mathcal{F}_{W,T(N-1)-W+1}
\]  

(15)

where \( \mathcal{F}_{W,T(N-1)-W+1} \) follows an \( \mathcal{F} \) distribution with degrees of freedom \( W \) and \( T \cdot (N - 1) - W + 1 \). This in turn leads to the definition of a control limit for the \( T^2 \) chart:

\[
UCL = \frac{(T - 1) \cdot (N - 1) \cdot W}{T \cdot (N - 1) - W + 1} \mathcal{F}_{x,W,T(N-1)-W+1}
\]  

(16)

where \( x \) is a pre-specified Type I error rate, and \( \mathcal{F}_{x,W,T(N-1)-W+1} \) denotes the \( (1 - x) \) quantile of the underlying \( \mathcal{F} \) distribution. Herein, \( x \) is set to 0.27% for consistency with the \( L = 3 \) control limits for univariate control charts described earlier.

### 3.3.1. Outlier Interpretation: Mason–Young–Tracy Decomposition

Measurements contributing to outliers are not directly identifiable in multivariate control charts. Approaches to address this problem are reviewed in Lowry and Montgomery (1995), Bersimis et al. (2005, 2007). Here, we implement MYT Decomposition to identify measurements’ contributions to (large) \( T^2 \) values. In MYT Decomposition, a given \( T^2 \) value is divided into (orthogonal) terms representing contributions from different measurements (Mason et al., 1995). Terms exceeding a threshold indicate that the associated measurements contribute significantly to the outlier. In the simple bivariate case, where \( x_1^2 \) and \( x_2^2 \) are measurements of two interdependent quality characteristics, the MYT method yields:

\[
T^2 = T^2_1 + T^2_{2|1}
\]  

(17)

where \( T^2_1 \) represents the marginal contribution of \( x_1^2 \) to the total \( T^2 \) value, and \( T^2_{2|1} \) represents the conditional contribution of \( x_2^2 \) given \( x_1^2 \). The calculation of the two components is given by:

\[
T^2_1 = N \cdot \frac{(\bar{x}_1^2 - \hat{\mu}_1)^2}{(\bar{s}_1)^2}
\]  

(18)

\[
T^2_{2|1} = N \cdot \frac{(\bar{x}_2^2 - \hat{\mu}_{2|1})^2}{(\bar{s}_{2|1})^2}
\]  

(19)

where \( \hat{\mu}_{2|1} \) and \( (\bar{s}_{2|1})^2 \) are estimates of the conditional mean and variance of \( x_2^2 \) given \( x_1^2 \). Large marginal components are associated with outliers in the corresponding measurements, whereas a large conditional components indicate either outliers or breaks in the relationship between measurements. In order to determine a threshold for consistent judgment, Mason et al. (1995, 1997) showed that both marginal and conditional components follow \( \mathcal{F} \) Distributions scaled by a constant:

\[
T^2_* \sim \frac{T^2_1}{T} \mathcal{F}_{1,T(N-1)}
\]  

(20)

where * represents any one of the terms in the bivariate decomposition.

For \( W \)-variate situations, the MYT Decomposition is generalized as follows:

\[
T^2 = T^2_1 + T^2_{2|1} + T^2_{3|2} + \ldots + T^2_{W|1,2,\ldots,W-1}
\]  

(21)

where the conditional terms \( T^2_{w|1,2,\ldots,w} \) represent the contribution of variable \( x_{w+1}^2 \) for given \( x_1^2, x_2^2, \ldots, x_w^2 \). We note that Eq. (21) is one of \( W! \) possible decompositions, and that because the order of conditional terms does not matter, e.g., \( T^2_{w|1,2,\ldots,w} = T^2_{w|1,2,\ldots,w} \), there are \( W \cdot 2^{W-1} \) unique terms appearing in the possible decompositions. The \( T^2 \) components can be shown to follow \( \mathcal{F} \) Distributions:

\[
T^2_{w|1,2,\ldots,w} \sim \frac{(T^2_1 \cdot (T - 1)}{T \cdot (T - W - 1)} \mathcal{F}_{1,T(N-w)}
\]  

(22)

where \( w \) is the number of variables conditioned on.
3.4. Supplementary charts

Even though changes in a process’s variability/correlation can lead to outliers in $T^2$ charts, it can be of interest to monitor variability directly. Also, in SHM it is of significant interest to detect persisting changes to a process that may not result in outliers in $T^2$ charts due to their small magnitude. In this paper, we implement Generalized Sample Variance (GSV) and Multivariate Exponentially-weighted Moving Average (MEWMA) charts to achieve the aforementioned objectives.

**GSV Control Charts** are used to monitor the variability in the set of measurements (Alt, 1984). The chart is analogous to the $s$ chart, and relies on the determinants of the sample covariance matrix in period $t$. $|S_t|$, $S_t$ is an unbiased estimate of $\Sigma$ if the underlying process is in control. We plot $|S_t|$ along the horizontal axis and set the control limits at $\pm 3$ standard deviations $\sqrt{\text{Var}(|S_t|)}$ from its mean level $E(|S_t|)$. It can be shown (cf. Montgomery (2009)) that:

$$E(|S_t|) = b_1 \cdot |\Sigma|$$

$$\text{Var}(|S_t|) = b_2 \cdot |\Sigma|^2$$

where:

$$b_1 = \frac{1}{(N - 1)^W} \prod_{w=1}^{W} (N - w)$$

$$b_2 = \frac{1}{(N - 1)^{2W}} \prod_{w=1}^{W} (N - w) \cdot \left[ \prod_{w=1}^{W} (N - w + 2) - \prod_{w=1}^{W} (N - w) \right]$$

and an unbiased estimate of $|\Sigma|$ can be obtained by $|S^*|/b_1$ where $|S^*| = \frac{1}{T} \sum_{t=1}^{T} |S_t|$. Therefore the upper and lower control limits are defined as:

$$\text{UCL} = \frac{|S^*|}{b_1} \left( b_1 + 3\sqrt{b_2} \right)$$

$$\text{LCL} = \frac{|S^*|}{b_1} \left( b_1 - 3\sqrt{b_2} \right)$$

To conclude, we note that construction of the GSV chart requires a sample size greater than the number of variables ($N > W$). Otherwise $b_1 = b_2 = 0$.

**MEWMA Charts** are designed to monitor the following multivariate statistics (Lowry and Woodall, 1992):

$$Z_t = \Lambda \bar{X}_t + [1 - \Lambda]Z_{t-1}$$

where $Z_1 = \bar{X}_1$, the weight matrix $\Lambda$ is diagonal with elements $[\lambda_1, \lambda_2, \ldots, \lambda_W]$, and where $0 < \lambda_w \leq 1, w = 1, 2, \ldots, W$ are a set of tuning parameters. Component $w$ in the vector $Z_t$ corresponds to an exponentially-weighted moving average of the measurement in $t$ and the history of the process up to $t - 1$. That is, $Z_{w}^w = \lambda_w X_{w}^w + (1 - \lambda_w)Z_{w-1}^w$. Without prior knowledge to make the weights of past observations different across the $W$ variables, we adopt a uniform weight $\lambda_1 = \lambda_2 = \ldots = \lambda_W = \lambda$ for simplicity. Eq. (29), therefore, reduces to $Z_t = \lambda \cdot \bar{X}_t + (1 - \lambda) \cdot Z_{t-1}$.

Following the logic motivating the use of the $T^2$ control chart, the moving averages, $Z_t$, are aggregated into statistics as follows:

$$T^2 = Z_t' \cdot \Sigma Z_t^{-1}$$

where $\Sigma_{Z_t}$ is the covariance matrix of $Z_t$, and is given by:

$$\Sigma_{Z_t} = \frac{\lambda [1 - (1 - \lambda)^2]}{2 - \lambda} \cdot \Sigma$$

which converges to $|\lambda/(2 - \lambda)| \Sigma$ as $t$ increases. $\Sigma$ is as in Section 3.3. $T^2$ values exceeding a pre-specified threshold, $h\_{\text{MEWMA}}$, indicate possible special-cause events. Various calibration approaches, including simulation, are used to obtain a threshold yielding a desired Average Run Length (ARL).

4. Empirical examples

We describe our implementation of the framework presented in the previous section to process data from the SHM system on the Hurley Bridge (Wisconsin Structure B-26-7). We begin by describing the data used for analysis. We then report on the statistical modeling to control for common-cause variation. Finally, to highlight key capabilities of the proposed framework, we present 4 numerical examples of the application of MSPC for detection of special-cause events.
4.1. Data

The Hurley Bridge carries westbound traffic on US Route 2 over the Montreal River from Ironwood, Michigan to Hurley, Wisconsin. The bridge consists of 3 spans supported by a steel structure consisting of 5 girders. The installation of the SHM system was motivated by WisDOT’s concern that loads associated with logging trucks traveling from Michigan into Wisconsin are causing premature deterioration of the structure, with excessive movement being one of the specific concerns. In terms of instrumentation, the SHM system consists of 2 displacement transducers and 13 strain gages. Ambient temperature, relative humidity, and temperature on the steel girders are also collected. A weigh-in-motion system, located upstream (East) of the bridge, provides information about traffic loads. A schematic of the bridge, including sensor identification and location, is presented in Fig. 2.

In terms of the design of the SHM system, the displacement transducers are placed at the location subject to the greatest expected movement, Girder 1. The girder supports on the Wisconsin side (West) are fixed, whereas the supports on the Michigan side (East) are flexible (rollers). There are two vehicle lanes on the bridge deck, and Girder 4 serves as the major support for the right lane where most of the truck traffic is expected, which explains the large number of strain gages on this girder. For reference, strain gages were installed on each of the 5 girders at the mid-point of Span 2. Additional details regarding the instrumentation, including the labels that we use to identify each of the measurements, appear in Table 1.

To illustrate the proposed framework, we analyze the response measurements from the following 7 sensors: Displ-L, Displ-T, G2-S2-M, G3-S2-M, G4-S2-M, G4-WCP-SL, and G4-ECP-SL. Data were collected from 1 April 2010 to 28 June 2012, for an analysis period of 820 days. Measurements, both responses and explanatory factors, were collected at a frequency of 100 Hz. Hourly averages, transmitted by a computer on-site, were used to construct and process averages for each day in the analysis period. Thus, the data used for analysis correspond to the 7 sequences of 820 measurements presented in Fig. 3.14 Seasonal variation and (small) increasing trends in Displ-L and Displ-T are the most visible features of the measurements as appearing in the figure. While transforming the data or rescaling the plots has potential to reveal additional features/insights, the lack of visual evidence of unusual (longitudinal or cross-sectional groups of) measurements serves as motivation to apply MSPC.

Data from the weigh-in-motion system includes information about every vehicle traversing the bridge during the analysis period, an average of 5200 vehicles/day. These data are described further in Appendix B and in Chen (2014).

4.2. Statistical performance modeling

In Section 4.2.1, we describe the formulation and estimation of a set of regression models to relate variation in exogenous variables to the response measurements. In Section 4.2.2, we describe the formulation and estimation of ARIMA–GARCH models to control for serial dependence in the residuals of the regression models.

4.2.1. Regression analysis

We formulate 7 commonly-specified regression models to capture the (contemporaneous) effect of the explanatory variables on the response measurements. The models also include linear trend terms as supplementary predictors. Prior to specifying the common equation, we investigated the multicollinearity among the predictor variables using the Variance Inflation Factor Method. The analysis, reported in Chen et al. (2014), led to the removal of ambient temperature from the regression equations. In addition, preliminary analysis did not yield a model where (cumulative) traffic effects were statistically-significant, and thus, traffic was also excluded as an explanatory variable.15 The regression equations were, therefore, specified as follows:

\[ y_t^w = \beta_0^w + \beta_1^w t + \beta_2^w \text{Humid}_t + \beta_3^w \text{ST}_t + \epsilon_t^w \]  

(32)

where \( t = 1, \ldots, T \) and \( w = 1, \ldots, W \) are, respectively, used to index the \( T = 820 \) days in the analysis period, and each of the \( W = 7 \) measurements. The variables Humid and ST correspond to the daily average relative humidity and steel temperature on day \( t \). The steel temperature corresponds to that of the girder closest to the relevant instrument (temperature on Girder 1 was used for measurements on Girder 2, and temperature on 3 for measurements on Girder 4). The coefficients \( \beta_0^w \) and \( \beta_3^w \) are the intercept and the slope of the linear trend in measurement sequence \( w \). \( \beta_2^w \) and \( \beta_3^w \) capture the effects of relative humidity and temperature on \( w \), and \( \epsilon_t^w \) denote the residuals.

The estimation results are presented in Table 2. For each coefficient \( \beta_j^w \), a t-stat with absolute value larger than 1.96 (and p-value close to 0) indicates that the associated variable is statistically-significant at the 95% confidence level (d.f.=818). Coefficients not deemed statistically-significant are enclosed in parentheses. In terms of goodness-of-fit, the models for the displacement measurements, Displ-L and Displ-T, resulted in high \( R^2 \). Over 85% of the variation in these measurements is explained by corresponding changes in the explanatory variables together with the linear trend. With the exception of

13 Other measurements were excluded due to sensor failures and other problems. Additional discussion and analysis appear in Chen (2014).
14 The optimal treatment of temporal resolution among measurements can be complex. We adopt a homogeneous specification here for simplicity. Despite possible limitations, we find it good enough for our objective to control common-cause variation and detect special-cause events.
15 One plausible explanation is that the transmission of average hourly response measurements (and the construction of daily averages) filters the effect of traffic loads (on the strain measurements). For the purpose of this analysis, the exclusion provides a set of events that could be related to special-cause variation.
G4-WCP-SL, the models show lower, though reasonable levels of fit with $R^2$ values ranging from 40% to 70%. Basic data transformations did not lead to significant improvements in the goodness-of-fit, which means that either the relationships between measurements and explanatory variables are more sophisticated, that additional variables are needed to explain changes in the measurements, or that the remaining variation is random. These directions are not explored further because they are not the focus of the paper.

Examination of the estimates presented in Table 2 indicates that linear trends are small in magnitude but significant components of all of the measurement sequences. For example, the results show that Girder 1 drifted East at an average speed of 0.12 mils/day (98.4 mils over the analysis period). This movement is in excess of that associated with seasonal variation in temperature and humidity, and is a likely indication of deterioration of the bridge supports. Persistent movement in the transverse direction is also significant (57.4 mils over the analysis period). In the case of the strains, linear trends could indicate changes in the material properties, e.g., fatigue, the sensor calibration, or the loading conditions.

The inclusion of intercept terms in the regression Eq. (32) yields sequences of residuals, each with zero mean, capturing unexplained variation. Serial-dependence, i.e., carryover effects from earlier measurements, possibly influenced by earlier changes in the explanatory variables, can be a significant source of unexplained common-cause variation. Modeling serial dependence, therefore, leads to refinements in forecasting the progression of the measurement sequences. Moreover, the

---

Table 1

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measurement</th>
<th>Unit</th>
<th>Component</th>
<th>Location</th>
<th>Axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Displ-L</td>
<td>Displacement</td>
<td>mils</td>
<td>Girder 1</td>
<td>East abutment</td>
<td>Long.</td>
</tr>
<tr>
<td>Displ-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Trans.</td>
</tr>
<tr>
<td>G1-S2-M</td>
<td>Girder 1</td>
<td></td>
<td>Span 2, mid-point</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G2-S2-M</td>
<td>Girder 1</td>
<td></td>
<td>Span 2, mid-point</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G3-S2-M</td>
<td>Girder 3</td>
<td></td>
<td>Span 2, mid-point</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-S2-M</td>
<td>Girder 4</td>
<td></td>
<td>Span 2, mid-point</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G5-S2-M</td>
<td>Girder 5</td>
<td></td>
<td>Span 2, mid-point</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-P2-TF</td>
<td>Girder 4 top flange</td>
<td></td>
<td>Pier 2</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-WCP-SL</td>
<td>Strain</td>
<td>$\mu$strain</td>
<td>Pier 2, west end south edge</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-WCP-ST</td>
<td>flange cover plate</td>
<td></td>
<td>Pier 2, east end south edge</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-WCP-CL</td>
<td>Girder 4 bottom</td>
<td></td>
<td>Pier 2, east end center</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>G4-ECP-CL</td>
<td>flange cover plate</td>
<td></td>
<td>Pier 2, east end center</td>
<td>Long.</td>
<td></td>
</tr>
<tr>
<td>Expl. factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST-G1</td>
<td>Steel</td>
<td>$^\circ$F</td>
<td>Girder 1</td>
<td>Span 2, mid-point</td>
<td></td>
</tr>
<tr>
<td>ST-G3</td>
<td>Steel</td>
<td>$^\circ$F</td>
<td>Girder 3</td>
<td>Span 2, mid-point</td>
<td></td>
</tr>
<tr>
<td>ST-G5</td>
<td>temperature</td>
<td>$^\circ$F</td>
<td>Girder 5</td>
<td>Span 2, mid-point</td>
<td></td>
</tr>
<tr>
<td>Temp</td>
<td>Air temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humid</td>
<td>Relative humidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 2. Schematic of the Hurley Bridge.
analysis of special-cause variation (Section 4.3) is predicated on the assumption of randomness, and, in turn, motivates the need to control for sources of systematic variation, such as serial dependence. In the following section, we describe the formulation and estimation of ARIMA–GARCH models, a class of time series models, to control for serial dependence that is not captured in the regression models.

Table 2

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Level</th>
<th>Linear trend</th>
<th>Humidity</th>
<th>Steel temp.</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_0$</td>
<td>$t$-stat</td>
<td>$b_1$</td>
<td>$t$-stat</td>
<td>$b_2$</td>
</tr>
<tr>
<td>Displ-L</td>
<td>4782.8</td>
<td>490.0</td>
<td>0.12</td>
<td>31.7</td>
<td>(-0.06)</td>
</tr>
<tr>
<td>Displ-T</td>
<td>4183.0</td>
<td>725.2</td>
<td>0.07</td>
<td>31.3</td>
<td>0.27</td>
</tr>
<tr>
<td>G2-S2-M</td>
<td>423.8</td>
<td>68.6</td>
<td>0.06</td>
<td>26.1</td>
<td>-0.69</td>
</tr>
<tr>
<td>G3-S2-M</td>
<td>218.8</td>
<td>30.7</td>
<td>0.07</td>
<td>26.2</td>
<td>-0.89</td>
</tr>
<tr>
<td>G4-S2-M</td>
<td>-115.2</td>
<td>-18.4</td>
<td>0.06</td>
<td>26.9</td>
<td>-0.66</td>
</tr>
<tr>
<td>G4-WCP-SL</td>
<td>420.7</td>
<td>37.5</td>
<td>0.05</td>
<td>10.7</td>
<td>-1.70</td>
</tr>
<tr>
<td>G4-ECP-SL</td>
<td>-237.8</td>
<td>-27.9</td>
<td>0.08</td>
<td>25.7</td>
<td>-1.11</td>
</tr>
</tbody>
</table>

Fig. 3. Daily (average) measurements.
4.2.2. Time series analysis

We use commonly-specified ARIMA–GARCH models, Eqs. (33)–(36), to simultaneously capture/control for serial dependence in the mean and variance of the residual series. The first step, i.e., the “integrated” part of the model, is used to construct stationary series by differencing. Application of the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test suggests that first-order differences \((d=1)\), shown in Eq. (33), are sufficient for each of the residual series. In the second step, we use maximum likelihood estimation to find the ARMA-GARCH coefficients in Eqs. (34) and (36) to capture serial dependence in the ensuing series. The estimation results appear in Table 3.

\[
\begin{align*}
\hat{z}_t^w &= \hat{e}_t^w - \hat{e}_{t-1}^w \\
\hat{z}_t &= \sum_{i=1}^p \psi_i^w \cdot \hat{z}_{t-i} + \theta_0^w + \epsilon_t^w + \sum_{j=1}^q \theta_j^w \cdot \hat{e}_{t-j}^w \\
\hat{e}_t^w &= \chi_t^w \cdot \sigma^w_r(t), \\
\left(\sigma^w_r(t)\right)^2 &= \sum_{k=1}^r \alpha_k^w \cdot \left(\sigma^w_r(t-k)\right)^2 + \beta_0^w + \sum_{i=1}^s \beta_i^w \cdot \left(e^w_{t-i}\right)^2
\end{align*}
\]

\(\hat{e}_t^w\) is the time \(t\) regression residual of measurement sequence \(w\). \(\hat{z}_t^w\) is the first-order difference, and \(\hat{e}_t^w\) is the associated error term. Eq. (34) is an Autoregressive Moving Average (ARMA) model where \(\hat{z}_t^w\) is a combination of its historic values, \(\hat{z}_{t-1}^w, \ldots, \hat{z}_{t-p}^w\), and the error terms, \(\epsilon_t^w, \ldots, \epsilon_{t-q}^w\). Each error term, \(\epsilon_t^w\), consists of a time-dependent standard deviation, \(\sigma^w_r(t)\), and white noise, \(\chi_t^w\). Eq. (36) is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) part capturing the dependence in the series variances. It relates \(\left(\sigma^w_r(t)\right)^2\) to its historic values, \(\left(\sigma^w_r(t-1)\right)^2, \ldots, \left(\sigma^w_r(t-k)\right)^2\), and previous squared error terms, \((\epsilon_{t-1}^w)^2, \ldots, (\epsilon_{t-p}^w)^2\). Satisfactory estimation yields a standardized innovation series, \(\{\chi_t^w\}_{t=1}^T\), where \(\chi_t^w = \epsilon_t^w / \sigma^w_r(t)\), that follows a standard normal distribution. In particular, the orders of the ARMA and GARCH components, \(p, q, r, s\), are selected so that the standardized innovations series, \(\{\chi_t^w\}\), satisfy the normality assumption, and so that they do not display significant autocorrelation or conditional heteroscedasticity. With the exception of Displ-T, serial dependence in the residual sequences can be effectively captured by ARIMA\((p=1, d=1, q=1)\)-GARCH\((r=1, s=1)\) models. Displ-T requires a second order autoregressive model \((p=2)\). For all series, we used the Ljung-Box, Kolmogorov-Smirnov, and Engle Tests (at the 95% confidence level) to verify compliance with the respective assumptions. The results of the statistical testing are reported in Appendix A.

Fig. 4 displays the standardized innovations and autocorrelation functions (ACFs) for each of the response measurements. The standardized innovations are presented because the statistics that we analyze to detect possible special-cause events are derived from them. The statistical tests discussed in the previous paragraph are used to verify compliance with the assumptions underlying the application of control charts. The ACFs in Fig. 4 constitute a visual aid. Because the autocorrelation values are within 2 standard deviations from the mean, we conclude (at the 95% confidence level) that the innovations do not display (significant) serial dependence. Additional details are presented in Chen et al. (2014).

4.3. Analysis of special-cause variation

In this section, we consider different subsets of the response measurements to illustrate the capabilities of MSPC for SHM. The first example corresponds to the analysis of the longitudinal and transverse displacements: Displ-L and Displ-T. Example 2 consists of monitoring two strain measurements on the cover plates at the joint between Girder 4 and Pier 2: G4-WCP-SL and G4-ECP-SL. The third example involves monitoring the mid-span strains on Girders 2, 3 and 4: G2-S2-M, G3-S2-M and G4-S2-M. These girders provide direct support to the two traffic lanes on the bridge deck. Finally, the last example involves simultaneous analysis of the 7 measurements.

To conduct the analysis, we organized the standardized innovations associated with each measurement into weekly subgroups/samples, i.e., the sample size \(N = 7\) and \(T = 820/7 = 117\). As discussed in the construction of the \(X\) Chart, grouping reduces the variability in the statistics being monitored, which increases the sensitivity to detect small changes in a process. Weekly subgroups are in line with sample sizes reported in synergistic applications, they yield intuitive and homogeneous samples, and provide a reasonable balance between detection of persisting small changes and yielding timely signals. In general, of course, the sampling scheme can be designed to address economic or technical concerns, e.g., error rates associated with specific/expected process changes.

Before proceeding, it is necessary to discuss compliance with the assumptions underlying the use of the MSPC procedures. In particular, the construction of the \(T^2\) control chart builds on the assumption of a set of quality characteristics jointly drawn from a multivariate normal distribution, yielding \(T^2\) statistics following \(F\) distributions (for \(N > 1\)). Unfortunately, conclusive tests to verify the assumption of multivariate normality are not available.\(^{16}\) Thus, we follow the quantile-quantile (Q–Q) plot approach suggested by Mason and Young (2002) to conduct the compliance test for our case studies in this section and report the results in Appendix A.

\(^{16}\) The exact distribution of the test statistics is unknown in the procedures used to assess multivariate normality, and as a result, the critical values can only be approximated. The procedures, therefore, only serve as indicators of possible multivariate normality (Seber, 1984; Looney, 1995).
4.3.1. Example 1: Biaxial Girder displacements

In this example, we analyze the sequences corresponding to the longitudinal and transverse displacements, Displ-L and Displ-T, on the East end of Girder 1. The $T^2$ and GSV control charts ($z = 0.27\%$) appear in Fig. 5. The $T^2$ control chart displays outliers at weeks #46 and #95, the former with a much larger magnitude. The GSV Chart yields signals at weeks #8, #54, #59, #62 and #107.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>$\hat{\psi}_2$</th>
<th>Stat</th>
<th>$\phi_0$</th>
<th>Stat</th>
<th>$\phi_1$</th>
<th>Stat</th>
<th>$a_1$</th>
<th>Stat</th>
<th>$b_0$</th>
<th>Stat</th>
<th>$b_1$</th>
<th>Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displ-L</td>
<td>0.47</td>
<td>12.5</td>
<td>0.02</td>
<td>0.6</td>
<td>-0.94</td>
<td>-64.2</td>
<td>0.95</td>
<td>44.6</td>
<td>4.92</td>
<td>1.2</td>
<td>0.03</td>
<td>2.5</td>
</tr>
<tr>
<td>Displ-T</td>
<td>0.74</td>
<td>18.5</td>
<td>-0.25</td>
<td>-6.8</td>
<td>0.00</td>
<td>-0.1</td>
<td>-0.86</td>
<td>-32.3</td>
<td>0.96</td>
<td>51.6</td>
<td>0.66</td>
<td>1.5</td>
</tr>
<tr>
<td>G2-S2-M</td>
<td>0.32</td>
<td>7.0</td>
<td>-0.02</td>
<td>-0.5</td>
<td>-0.87</td>
<td>-35.0</td>
<td>0.96</td>
<td>51.1</td>
<td>1.35</td>
<td>1.4</td>
<td>0.03</td>
<td>2.6</td>
</tr>
<tr>
<td>G3-S2-M</td>
<td>0.37</td>
<td>7.5</td>
<td>-0.01</td>
<td>-0.3</td>
<td>-0.86</td>
<td>-31.7</td>
<td>0.95</td>
<td>52.6</td>
<td>1.96</td>
<td>1.7</td>
<td>0.03</td>
<td>3.1</td>
</tr>
<tr>
<td>G4-S2-M</td>
<td>0.35</td>
<td>6.9</td>
<td>0.00</td>
<td>-0.1</td>
<td>-0.85</td>
<td>-29.3</td>
<td>0.95</td>
<td>55.2</td>
<td>1.64</td>
<td>1.7</td>
<td>0.03</td>
<td>3.3</td>
</tr>
<tr>
<td>G4-WCP-SL</td>
<td>0.31</td>
<td>7.0</td>
<td>-0.05</td>
<td>-0.8</td>
<td>-0.89</td>
<td>-40.6</td>
<td>0.94</td>
<td>58.0</td>
<td>4.63</td>
<td>1.8</td>
<td>0.04</td>
<td>3.7</td>
</tr>
<tr>
<td>G4-ECP-SL</td>
<td>0.30</td>
<td>6.7</td>
<td>-0.01</td>
<td>-0.3</td>
<td>-0.89</td>
<td>-41.2</td>
<td>0.96</td>
<td>44.5</td>
<td>3.28</td>
<td>1.1</td>
<td>0.02</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Fig. 4. (a) Standardized innovation series and (b) autocorrelation functions.
the outliers are likely related to special-cause variation in the transverse displacement.

general, can be set according to a desired Type I error chart, we observe that the event at #46 has a persisting effect.

observation in Example 1, even though the physical properties, sensors and location are all different. From the MEWMA #59, #62 and #80. Interestingly, we notice that the outlier at week #46 is (and those at #59 and #62 are) consistent with the persisting effect from week #72 to #75.

As shown in Fig. 9, the marginal and conditional components of Displ-L and Displ-T. We observe that both the marginal and conditional contributions of Displ-T (i.e., $T^2_L$ and $T^2_T$, respectively, denote the marginal and conditional contributions of Displ-L and Displ-T. We observe that both the marginal and conditional components of Displ-T (i.e., $T^2_L$ and $T^2_T$) are significant. In contrast, the Displ-L components are insignificant. We conclude, therefore, that the outliers are likely related to special-cause variation in the transverse displacement.

We benchmark the results by constructing $x$ and $s$ control charts ($x = 0.27\%$) for Displ-L and Displ-T. They are presented in Fig. 6. We observe that the two signals in the $T^2$ chart corresponds to outliers in the $x$ Chart of Displ-T, confirming the inference obtained from the MYT Decomposition. We also observe an outlier in the $s$ chart of Displ-L coincides with the most significant outlier appearing in the GSV Chart.

Bivariate examples allow for graphical representation and interpretation of the results. Fig. 7 is a scatter plot of the samples in the data set. Displ-L innovations are measured on the horizontal axis, and Displ-T on the vertical axis. The rectangle corresponds to the intersection of the univariate control regions. The innovations in weeks #46 and #95 fall below the rectangle’s lower border, indicating that they fall below the LCL of Displ-T’s $x$ chart. The control region for the $T^2$ chart is represented by the ellipse. Points outside the ellipse correspond to outliers in the $T^2$ chart.

Finally, in Fig. 8, we present MEWMA control charts for different combinations of the tuning parameters $x$ and $h$, which, in general, can be set according to a desired Type I error or, alternatively, to detect changes of a given magnitude. We adopted the tabulated $h$ values from Lowry and Woodall (1992) based on $ARL = 200$ ($x = 0.5\%$), which is why $h$ varies. By reducing $x$, we observe that the effect of the outlier at week #46 is persisting, whereas the effect of the outlier at #95 is transient, i.e., its effect dissipates as $x$ is reduced. We also observe evidence of a significant special-cause event with a small magnitude, but persisting effect from week #72 to #75.

4.3.2. Example 2: Strain pair on cover plates of Girder 4

We analyze 2 longitudinal strains on cover plates located at the joint between Girder 4 and Pier 2: G4-WCP-SL, G4-ECP-SL. As shown in Fig. 9, the $T^2$ chart has outliers at weeks #1, #46 and #81, while the GSV chart has outliers at weeks #22, #58, #59, #62 and #80. Interestingly, we notice that the outlier at week #46 is (and those at #59 and #62 are) consistent with the observation in Example 1, even though the physical properties, sensors and location are all different. From the MEWMA chart, we observe that the event at #46 has a persisting effect.

The MYT Decompositions of the $T^2$ values are presented in Table 5. They show that the marginal contributions of the individual measurements are small. The conditional contributions, on the other hand, exceed their critical thresholds. This suggests that the outliers are triggered by inconsistencies in the relationship between the measurements.

The scatter plot in Fig. 10 confirms the above analysis. In particular, we observe that the $3 T^2$ outliers are within the univariate control region, indicating that an inconsistent relationship between the measurements is the likely cause of the signals. This is also apparent from the $x$ charts, shown in Fig. 11, where it is easy to see temporary inconsistencies between the

\[ T^2 = T^2_L + T^2_T = T^2_L + T^2_T \]  are the two possible decompositions.

The inclination of the ellipse is related to the correlation between the variables. The length of the ellipse’s axes depends on the variance in the data.

<table>
<thead>
<tr>
<th>Signal week</th>
<th>$T^2$</th>
<th>Marginal</th>
<th>$T^2_L$</th>
<th>$T^2_T$</th>
<th>Conditional</th>
<th>$T^2_{LT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#46</td>
<td>30.72*</td>
<td>0.52</td>
<td>23.26*</td>
<td>30.20*</td>
<td>7.46</td>
<td></td>
</tr>
<tr>
<td>#95</td>
<td>14.24*</td>
<td>1.52</td>
<td>12.89*</td>
<td>12.72*</td>
<td>1.34</td>
<td></td>
</tr>
</tbody>
</table>

* Indicates components exceeding 99.7% confidence level.

Fig. 5. Biaxial displacements: (a) $T^2$ control chart and (b) GSV control chart.
Fig. 6. $\bar{x}$ and $s$ control charts of Girder 1's displacement: (a) Displ-L and (b) Displ-T.

Fig. 7. Elliptical control region of Biaxial displacements.

Fig. 8. MEWMA control chart of Biaxial displacements on Girder 1 ($\alpha = 0.5\%$).
Table 5
MYT Decomposition of $\tau^2$ statistic for G4-WCP-SL & G4-ECP-SL.

<table>
<thead>
<tr>
<th>Signal week</th>
<th>$\tau^2$</th>
<th>$\tau^2_{WCP}$</th>
<th>$\tau^2_{ECP}$</th>
<th>Conditional $\tau^2_{ECP/WCP}$</th>
<th>Conditional $\tau^2_{WCP/ECP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>22.60$^a$</td>
<td>1.54</td>
<td>0.62</td>
<td>21.06$^a$</td>
<td>21.98$^a$</td>
</tr>
<tr>
<td>#46</td>
<td>13.21$^a$</td>
<td>1.33</td>
<td>0.15</td>
<td>11.88$^a$</td>
<td>13.06$^a$</td>
</tr>
<tr>
<td>#81</td>
<td>22.89$^a$</td>
<td>0.02</td>
<td>3.49</td>
<td>22.87$^a$</td>
<td>19.4$^a$</td>
</tr>
</tbody>
</table>

$^a$ Indicates components exceeding 99.7% confidence level.

Fig. 9. Multivariate control charts of cover plate strains.

Fig. 10. Elliptical control region of G4-WCP-SL and G4-ECP-SL.

Fig. 11. $\bar{x}$ and $s$ control chart of cover plate strains: (a) WCP-SL and (b) ECP-SL.
measured earlier, but not in this example. Outliers at #12 and #33 appear in this example for the first time, which may be related to lack of consistency among measurements, which also suggests special-cause variation. The GSV Chart does not apply in this example because it is for events as opposed to random variation. Except for #81, the MEWMA chart shows that these events have persisting effects most significant in terms of magnitude. Consistent detection throughout the analysis suggests outliers reflect special-cause

Finally, we observe that the $s$ chart for the west cover plate displays excessive variance at weeks #97 and #104, although they do not coincide with other outliers in this example.

4.3.3. Example 3: Mid-span strains

In this example, we consider the mid-span longitudinal strains on Girders 2, 3 and 4: G2-S2-M, G3-S2-M and G4-S2-M. Fig. 12 shows $T^2$, MEWMA ($\lambda = 0.1$ and $\alpha = 0.5\%$) and GSV control charts.

The only outlier appears on the $T^2$ control chart at week #46 which coincides with signals in the previous examples, even though the property, sensors and location are different. The GSV chart displays 5 outliers at weeks #40, #46, #58, #85 and #104 indicating unusual volatility in the within sample variance and showing partial correspondence to earlier analysis. The MYT decomposition of the $T^2$ statistic for week #46 is presented in Table 6. It shows that the marginal components, $T^2_{G2}$ and $T^2_{G4}$, are less than the critical value. In contrast, both conditional components, $T^2_{G2|G4}$ and $T^2_{G4|G2}$, significantly exceed the critical value, suggesting inconsistencies in their contemporaneous correlation at week #46.

The corresponding univariate control charts are presented in Fig. 13 for reference. We notice that the outliers do not coincide with discernible events on the corresponding charts.

4.3.4. Example 4: Analyzing 7 measurements

Here, we consider simultaneous analysis of the 7 measurements. The $T^2$ and MEWMA control charts are shown in Fig. 14. The $T^2$ chart displays 6 outliers at #1, #12, #33, #46, #58 and #81. In descending order, outliers at #46, #1, #58 and #81 are most significant in terms of magnitude. Consistent detection throughout the analysis suggests outliers reflect special-cause events as opposed to random variation. Except for #81, the MEWMA chart shows that these events have persisting effects on the measurements, which also suggests special-cause variation. The GSV Chart does not apply in this example because $N = W = 7$.

To assess the contribution of the different measurements to the outliers, given the large number of possible non-trivial MYT Decompositions ($7 \cdot 2^{7-1} = 448$), we proceed by grouping the measurements. In the decomposition presented in Table 7, we partition the 7 measurements into two subsets: the first with G2-S2-M (hereafter G2), G3-S2-M (G3), G4-S2-M (G4), Displ-L (L), and Displ-T (T); the second with G4-WCP-SL (W) and G4-ECP-SL (E). $T^2_{(G2,G3,G4,L,T)}$ corresponds to the combined contribution from the 5 measurements in the first subset. Its UCL at the 99.7% confidence level is 18.35, meaning that 5 of the 6 measurement sets resulted in $T^2$ outliers if we had excluded the strain pair (W, E) from the analysis.

Similar arguments hold true for the other subset. The overall $T^2$ accumulates the contribution from both subsets as well as their relationships, which are listed in the "Conditional" columns in Table 7. Components exceeding their respective (and reported) UCLs are labeled (A), (B), (C) and (D), and further decomposed in Tables 8 and 9.

---

19 The overall correlation coefficient between the measurements is $\rho = 0.91$.

20 Outliers at #1, #46, and #81 appear in earlier $T^2$ charts; and #58 appears in the GSV charts in examples 2 and 3. The outlier at #95 is the only $T^2$ outlier detected earlier, but not in this example. Outliers at #12 and #33 appear in this example for the first time, which may be related to lack of consistency among measurements that had been analyzed separately.
Fig. 13. x and s control chart of mid-span strains: (a) G2, (b) G3, and (c) G4.

Fig. 14. Multivariate control charts of Displ-L, Displ-T, G2-S2-M, G3-S2-M, G4-S2-M, G4-WCP-SL, and G4-ECP-SL.

Table 7
MYT Decomposition of $T^2$ statistic for 7 measurements.

<table>
<thead>
<tr>
<th>Week</th>
<th>T^2 Subset</th>
<th>T^2_{W,G2,G3,G4}</th>
<th>T^2_{W,L,E}</th>
<th>Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>40.82^a</td>
<td>13.51</td>
<td>22.6^a</td>
<td>23.47^a</td>
</tr>
<tr>
<td>#12</td>
<td>22.85^a</td>
<td>3.90</td>
<td>0.08</td>
<td>5.26</td>
</tr>
<tr>
<td>#33</td>
<td>23.18^a</td>
<td>12.33</td>
<td>7.33</td>
<td>9.37^a</td>
</tr>
<tr>
<td>#46</td>
<td>85.79^a</td>
<td>50.93^D</td>
<td>13.21^a</td>
<td>17.65^a</td>
</tr>
<tr>
<td>#58</td>
<td>37.48^a</td>
<td>16.75</td>
<td>4.55</td>
<td>0.36</td>
</tr>
<tr>
<td>#81</td>
<td>32.71^a</td>
<td>9.15</td>
<td>22.89^a</td>
<td>7.53</td>
</tr>
<tr>
<td>UCL</td>
<td>22.11</td>
<td>18.35</td>
<td>11.84</td>
<td>9.34</td>
</tr>
</tbody>
</table>

^a Indicates components exceeding UCL at 99.7% confidence level.

Table 8
Selected $T^2$ components in MYT Decomposition.

<table>
<thead>
<tr>
<th>Week</th>
<th>Components of (A)</th>
<th>Components of (B)</th>
<th>Components of (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T^2_{W,A}</td>
<td>T^2_{W,B}</td>
<td>T^2_{W,C}</td>
</tr>
<tr>
<td>#1</td>
<td>21.98^a</td>
<td>21.06^a</td>
<td>1.54</td>
</tr>
<tr>
<td>#12</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#33</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#46</td>
<td>13.06^a</td>
<td>11.88^a</td>
<td>1.33</td>
</tr>
<tr>
<td>#58</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#81</td>
<td>19.40^a</td>
<td>22.87^a</td>
<td>0.02</td>
</tr>
</tbody>
</table>

^a Indicates components exceeding UCL at 99.7% confidence level.
To summarize, we organize all the results in Table 10, where L, M and H indicates low, medium and high levels of contribution to the $T^2$ outliers. While it is not always possible to identify assignable causes, Appendix B shows analysis of traffic and weather data suggesting that the most significant outlier at #46 may be explained by unusually large temperature changes. We also conclude that outliers at #1 and #58 may be, respectively, explained by significant changes in weather and traffic.

As stated, a common approach used in the literature to analyze multiple measurements simultaneously involves using PCA to monitor their projections onto subspaces. In Appendix C, we present the results of PCA and compare to it the analysis of $T^2$ control charts. The approach detects the most significant outlier at week #46 and confirms a significant contribution of unusual Displ-T measurements.

5. Conclusions and research directions

In Chen et al. (2014) we present a two-part SPC framework linking the literature on statistical performance/deterioration modeling of transportation infrastructure/assets with the literature on SHM. The first part of the framework, overlapping with the former, consists of estimation of statistical models to explain, predict, and control for common-cause variation in the data, i.e., changes that can be attributed to usual operating conditions, including traffic loads, environmental effects, and damage when present. The second part of the framework, overlapping with the latter, consists of using control charts to analyze/monitor the standardized innovations, i.e., prediction errors, of the aforementioned models in order to detect possible special-cause or extraordinary events.

Our earlier implementation of the second part of the framework used univariate control charts to analyze measurement sequences, one at a time. Motivated by technological advances that allow for reliable, high-frequency/continuous, long-term data collection from multiple sources, in this paper, we present MSPC as a framework to simultaneously analyze sets of response-measurements/structural-properties. The proposed approach revolves around construction of $T^2$ control charts as a framework to aggregate (into a scalar) and monitor the evolution and contemporaneous correlation of a set of measurements. The proposed approach can be especially appealing when large sets of measurements/properties are collected. Significantly, characterization and analysis of the relationships among measurements provides additional capabilities to interpret special-cause events, and thus to support resource-allocation decisions.

To illustrate the capabilities of the proposed framework, we analyze 4 different subsets of 7 response measurements from the SHM system on the Hurley Bridge. The first 3 examples consider non-overlapping subsets of measurements. As in Chen et al. (2014), despite analyzing distinct structural properties of bridge elements at separate locations, we report consistent detection (across multiple response measurements) of possible special-cause events, which is an indication of the reliability of the framework and the thoroughness of the implementation. In example 4, we analyzed the 7 measurement sequences collected over the 27 month planning horizon, and detect 6 possible special-cause events, and link the most significant ones, in terms of magnitude and duration, to extraordinary changes in weather and traffic. In terms of interpretation, we use Mason–Young–Tracy (MYT) Decomposition to establish the contribution of (subsets of) the measurements each of the outliers. This constitutes an innovative application of a sophisticated methodology, and to the best of our knowledge, is the first

---

21 Somewhat arbitrarily, we label contributions H if the $3 \times UCL \leq T^2$; M if $2 \times UCL \leq T^2 < 3 \times UCL$; and L if UCL $\leq T^2 < 2 \times UCL$. We note that using these criteria, the outlier at #46 is H, the one at #1 is M, and all others are L.
MSPC study in the SHM context where the problem of identifying contributions to outliers is considered. The vast majority of MSPC studies in SHM rely on projection methods that preclude inferences that apply in the original measurement spaces. We use Principal Component Analysis (PCA), the most common projection technique appearing in the literature, both to benchmark the results, as well as to illustrate the aforementioned shortcomings.

While the idea of using SPC to support SHM is not original to Chen et al. (2014) or this paper, our implementation differs in terms of the rigor of implementation, as well as the type and source of data. In particular, benchmark studies are often concerned with the analysis of structural vibrations, which are relatively insensitive to operational factors. Moreover, data are often collected from structures/lab-specimens in controlled environments where damage is deliberate. The use of high-quality field data to validate the framework is, therefore, an important contribution from an engineering perspective. Verification of compliance with underlying assumptions, and overall assessment/diagnosis of the results are rarely reported in the literature. Among other contributions, this led to the estimation of the first model of conditional heteroscedasticity, i.e., time-varying volatility, in the context of transportation infrastructure. It also contributes (along with high-quality data collection) to meaningful results, where we are able to establish correlation between outlier detection and the incidence of extraordinary traffic or weather.

5.1. Research directions

In terms of highlighting limitations of the present study constituting possible research opportunities, we note:

- The homogeneous temporal resolution of the measurements (daily averages) may lead to (short-term) loss of information. In part, the problem originates from the data collection protocols, which result in the transmission and recording of hourly averages. Ideally, the temporal resolution ought to be tailored for different distresses to match their physical nature and reflect managerial concerns;
- In the same spirit, while the use of commonly-specified statistical models is adequate (to control for common-cause variation and detect special-cause events), ideally, specifications should be tailored to the measurements;
- The sequential estimation of linear regression and time series models may introduce bias to the estimation results in a strict sense. Although this is common in practical applications, one needs to be aware and careful of its negative implications. The adoption of state-space specifications of time series models as in Chu and Durango-Cohen (2007, 2008), and notably Kobayashi et al. (2015), which is closest to our work, seems like a promising approach to address this problem;
- Among other advantages, additional data availability (beyond the 27 months used in the present study) would:
  - provide an opportunity to test the framework’s online monitoring capabilities. As is done in online applications of SPC (cf. Montgomery (2009)), the analysis period would be divided into two phases: a calibration phase that would allow for the estimation of performance models to control for common-cause variation; and a control/monitoring phase dedicated to special-cause event detection.
  - allow us to explore larger sample sizes, e.g., 30, which would improve the capabilities to analyze large subsets of measurements, e.g., construct the Generalized Sample Variance Chart for example 4.

Acknowledgments

The research was conducted while the first author was a graduate student at Northwestern University. The authors acknowledge the technical support of David Corr and David Kosnik from the Infrastructure Technology Institute at Northwestern University. The research was supported by grants from the Infrastructure Technology Institute through projects 60027736 and 60025238.

Appendix A. Testing innovations for compliance

Here we present the results of the hypothesis tests applied to the innovation series to ensure compliance with the assumptions underlying control charting. First, we discuss the tests to verify that the individual innovation series are stationary, iid Normal. Then, we present Q–Q plots to assess the multivariate normality of the sets of innovation series used in the examples presented in Section 4.3.

- The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test assesses the null hypothesis that a univariate series is stationary. It decomposes the series into a deterministic trend, a random walk plus a stationary error, and constructs a score statistic to test whether the random walk has zero variance. Since the statistic follows nonstandard distribution, we interpolate the critical values and p-values following Monte Carlo simulations conducted by Kwiatkowski et al. (1992).
- The Ljung-Box test evaluates the null hypothesis that a univariate series exhibits no autocorrelation within a fixed number of lags. It computes a Q statistic which, under the null hypothesis, asymptotically follows a chi-square distribution with \((d – c)\) degrees of freedom where \(d\) is the fixed lag and \(c\) is the total number of coefficients used in the time series model to generate the residuals. In this paper we use \(d = 20\) for all series.
- The (one-sample) Kolmogorov–Smirnov test is a non-parametric test to assess the null hypothesis that a given sample series is drawn from a standard Normal distribution. It compares the population c.d.f. of the sample series against the
The most distinguishable impact from traffic on the structural integrity. We observe a spike in the cumulative weight over the

is expected to follow a Beta distribution with degrees of freedom

when implementing the multivariate

in Example 2, one outlier in Example 3 and four outliers in Example 4). The result indicates that the measurement data set

trend except for a few extreme outliers in the upper tail of each plot (respectively, two outliers in Example 1, three outliers

distribution to the reference distribution.

approximately straight lines in the plot with a slop of 1 and intercept of 0 indicates an excellent fit of the empirical data

Hypothesis testing for innovations series.

which is a tractor or straight truck power unit.

temperature ranges likely contributed to the most recognizable and consistent outlier among the

porary features within the series may provide supplementary insights. In particular, a pronounced spike at week #46 of all

extraordinary factors and the special-cause events detected in the analysis. Fig. B.16 presents the weekly peak-to-valley

temporal consistency that connects weights from heavy trucks with some special-cause events.

Although the entire series of cumulative and maximum traffic turns out to be insignificant in the regression model, we found

the unusual southbound movement identified in Example 1.

We also investigated traffic loading on both lanes of the Hurley Bridge as another contributing factor to the deterioration.

Fig. B.17 illustrates the weekly cumulative and maximum weight from all Class-13 vehicles,22 which potentially represent

the most distinguishable impact from traffic on the structural integrity. We observe a spike in the cumulative weight over the

hypothesized c.d.f. (of a standard Normal distribution), and the test statistic equals to the supremum of their distance set.

Since the critical value is approximated by interpolation, the test decision is determined by comparing the p-value with the

significance level \( \alpha \).

The Engle's ARCH test assesses the null hypothesis that a residuals series has no conditional heteroscedasticity. It

regresses the variance of the sample at current time using the variance of \( d \) past samples (i.e., an ARCH(\( d \)) model). The

test statistic is a Lagrange multiplier statistic that, under null hypothesis, follows a chi-square distribution with \( d \) degrees of

freedom.

The results of the tests at the 95% (\( \alpha = 0.05 \)) confidence level are reported in Table A.11, which indicates that the 7 inno-

vation series comply with necessary statistical assumptions.

In addition, we use quantile–quantile (Q–Q) plots as a visual aid to determine whether the measurement data conform to

the multivariate normality assumption. We construct the statistic:

\[
B = \frac{T^2}{(T/1)(N/1)} \left( 1 + \frac{T^2}{(T/1)(N/1)} \right)
\]

where \( T^2 \) is defined in Eq. (12). As stated in Section 4.3, the Multivariate Normality assumption guarantees that

\( \frac{T(N/1-W)}{W(T/1)(N/1)} \) \( T^2 \) follows an \( F \) distribution with degrees of freedom \( W \) and \( T \cdot (N - 1) - W + 1 \). Thus it can be shown that \( B \)

is expected to follow a Beta distribution with degrees of freedom \( \frac{W}{2} \) and \( \frac{T(N/1-W)}{W(T/1)(N/1)} \). The Q–Q plot is a graphical tool that com-

pares the quantiles of the calculated \( B \) statistic against the corresponding quantiles of the reference Beta distribution. An

approximately straight lines in the plot with a slop of 1 and intercept of 0 indicates an excellent fit of the empirical data

distribution to the reference distribution.

The Q–Q plots for each of the 4 case studies in this paper are presented in Fig. A.15. We observe approximately linear

trend except for a few extreme outliers in the upper tail of each plot (respectively, two outliers in Example 1, three outliers

in Example 2, one outlier in Example 3 and four outliers in Example 4). The result indicates that the measurement data set

conform to the multivariate normality assumption while, on the other hand, we would expect the implied number of outliers

when implementing the multivariate \( T^2 \) control charts.

Appendix B. Assignable causes

We analyzed both environmental conditions and traffic loading to identify possible assignable causes, i.e., links between

extraordinary factors and the special-cause events detected in the analysis. Fig. B.16 presents the weekly peak-to-valley

range of related environmental conditions. Although the entire measurement series were used in the regression model, tem-

porary features within the series may provide supplementary insights. In particular, a pronounced spike at week #46 of all

temperature ranges likely contributed to the most recognizable and consistent outlier among the \( T^2 \) control charts in our

examples. Practically speaking, a sharp climb of almost 40°F in ambient/steel temperature within the second week of Febru-

ary 2011 could be a plausible cause of all or some of the unusual structural responses in the same week as described in Section

4.3. Similar link can be argued between the second largest spike in both steel temperature range plots at week #95 and the

unusual southbound movement identified in Example 1.

We also investigated traffic loading on both lanes of the Hurley Bridge as another contributing factor to the deterioration.

Although the entire series of cumulative and maximum traffic turns out to be insignificant in the regression model, we found

temporal consistency that connects weights from heavy trucks with some special-cause events.

Fig. B.17 illustrates the weekly cumulative and maximum weight from all Class-13 vehicles,22 which potentially represent

the most distinguishable impact from traffic on the structural integrity. We observe a spike in the cumulative weight over the

\footnote{In FHWA definition, Class-13, the heaviest vehicle category, corresponds to vehicles with seven or more axles consisting of three or more units, one of which is a tractor or straight truck power unit.}
month-long period starting from the second week of May to the second week of June in 2011 (week #58 to #63). It is then followed by another spike spanning the entire August of 2011 (week #71 to #74). The first spike overlaps with the unusual volatility across multiple distresses, while the second spike overlaps with a persistent shift in the displacements as detected by the MEWMA chart.

Appendix C. Principal component analysis

The intuition behind PCA in MSPC is to reduce the dimensionality of the analysis of special-cause variation by projecting a set of contemporaneous quality characteristics, i.e., standardized innovations, onto a lower-dimensional subspace. In Example 4, we consider a subspace spanned by the 3 principal components, i.e., eigenvectors with 3 largest eigenvalues, of the estimated covariance matrix, $\mathbf{S}$. As is shown in Fig. C.18(a), collectively these 3 components account for over 90% of the variation in the 7 innovation series. The 3 eigenvectors were normalized, and are presented in Table C.12. Projections onto each one of the 3 components are obtained by combining the innovations in a set using the associated coefficients appearing in Fig. A.15. Q–Q Plot: (a) Example 1, (b) Example 2, (c) Example 3, and (d) Example 4.

Fig. A.15. Q–Q Plot: (a) Example 1, (b) Example 2, (c) Example 3, and (d) Example 4.

Fig. B.16. Weekly peak-valley range: (a) G1 steel temperature, (b) G3 steel temperature, (c) air temperature, and (4) relative humidity.

Fig. B.16. Weekly peak-valley range: (a) G1 steel temperature, (b) G3 steel temperature, (c) air temperature, and (d) relative humidity.
In terms of the coefficients, we observe that the second and the third principal components have significantly larger contributions from Displ-L and Displ-T, while the first principal component combines the effects of the strains. Significantly, we observe that Displ-T contributes to the latter two components with an opposite sign.

The control charts for the 3 principal components are presented in Fig. C.18(b)–(d). The charts for principal components 2 and 3 display outliers in week #46, which coincide with the most significant outlier in our earlier analysis. The outlier on the chart for component 2 is above the UCL, whereas corresponding outlier is below the LCL on the chart for component 3. Given the magnitudes and signs of the coefficients in Table C.12, in this instance, we conclude that Displ-T contributed to the outlier. Unlike the MYT Decomposition used in Example 4, we are unable to identify other (statistically-significant) contributions, i.e., the lack of consistency between the strain measurements at the mid-span of Girders 2 and 3, and the strain at the East Cover Plate – G2-E and G3-E. In addition to the difficulties inherent in identifying subsets of measurements contributing to outliers, we also observe that in this example, PCA is less sensitive than the $T^2$ control chart analysis. This may be explained partially by the autocorrelation in the projections.

### Table C.12

<table>
<thead>
<tr>
<th>Component</th>
<th>G2-S2-M</th>
<th>G3-S2-M</th>
<th>G4-S2-M</th>
<th>G4-WCP-SL</th>
<th>G4-ECP-SL</th>
<th>Displ-L</th>
<th>Displ-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.45</td>
<td>0.44</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
<td>0.15</td>
<td>-0.01</td>
</tr>
<tr>
<td>#2</td>
<td>0.01</td>
<td>-0.17</td>
<td>-0.08</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.65</td>
<td>0.73</td>
</tr>
<tr>
<td>#3</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.11</td>
<td>-0.09</td>
<td>0.74</td>
<td>-0.66</td>
</tr>
</tbody>
</table>


In terms of the coefficients, we observe that the second and the third principal components have significantly larger contributions from Displ-L and Displ-T, while the first principal component combines the effects of the strains. Significantly, we observe that Displ-T contributes to the latter two components with an opposite sign.

The control charts for the 3 principal components are presented in Fig. C.18(b)–(d). The charts for principal components 2 and 3 display outliers in week #46, which coincide with the most significant outlier in our earlier analysis. The outlier on the chart for component 2 is above the UCL, whereas corresponding outlier is below the LCL on the chart for component 3. Given the magnitudes and signs of the coefficients in Table C.12, in this instance, we conclude that Displ-T contributed to the outlier. Unlike the MYT Decomposition used in Example 4, we are unable to identify other (statistically-significant) contributions, i.e., the lack of consistency between the strain measurements at the mid-span of Girders 2 and 3, and the strain at the East Cover Plate – G2-E and G3-E. In addition to the difficulties inherent in identifying subsets of measurements contributing to outliers, we also observe that in this example, PCA is less sensitive than the $T^2$ control chart analysis. This may be explained partially by the autocorrelation in the projections.

### References


Vincent, G., 1962. Correlation of predicted and observed suspension bridge behavior. Transactions American Society of Civil Engineers 127 (P8), 646–666.


