Physically-Based Data-Driven Approach for Anomaly Detection in Long-Term Continuous Monitoring Systems: Application to a Case Study

EMMA GHINI, MATTEO MARRA, MICHELE PALERMO, GIADA GASPARINI, STEFANO SILVESTRI and TOMASO TROMBETTI

ABSTRACT

Smart management of cultural heritage structures benefits from long-term continuous monitoring. Continuous monitoring collects large dataset of measurements that significantly improve the possibility to provide a correct interpretation of the structure's behaviour and to detect evolutionary trends and anomalies. By referring to structural features which are systematically extracted from the monitored dataset and identified as Reference Quantities (RQ), a physically-based data-driven approach is developed in this research work and applied to the case study of the Two Towers in Bologna (Italy). The proposed approach streamlines data processing, correlates structural behaviour, accounts for environmental effects, and supports the establishment of alert thresholds for an effective structural monitoring for the considered case study.

INTRODUCTION

The preservation of cultural heritage buildings is particularly relevant due to their historical, artistic, and societal significance. These structures often represent national and local identity and attract global attention. However, due to ageing materials, environmental conditions, and external forces, they are susceptible to structural degradation and need for preservation.

In this regard, continuous monitoring is often implied to detect early signs of damage and prevent irreversible failures. Cultural heritage buildings benefit from the installation of continuous Structural Health Monitoring (SHM) systems since the collected data can be used to identify material degradation and ensure long-term conservation. Continuous SHM is also effective in detect structural anomalies such as cracks, deformation, and tilting before they become critical. In addition, the continuous acquisition and elaboration of data on the structure is resourceful in optimizing management decisions based on real-time data analysis.

The present research work proposes a novel approach which combines real data (data-driven) and physical quantities (physically-based) with the aim of enhancing the accuracy of trend identification and data-driven anomaly detection, providing a more

Emma Ghini, PhD Candidate, Email: emma.ghini2@unibo.it. Department of Civil, Chemical, Environ- mental and Materials Engineering (DICAM), University of Bologna, Bologna, Italy

reliable framework for decision-making in structural management.

MONITORING DATA IN LONG-TERM CONTINUOUS SHM SYSTEMS

Long-term continuous SHM involves the observation and the analysis of a construction over time using periodically sampled response measurements to monitor changes to the material and geometric properties. The monitoring activity is developed through the following phases: (i) *Data Collection*: sensors installed on the structure continuously acquire real-time data; (ii) *Data Storage*: the collected data are indexed and classified; (iii) *Data Processing*: collected information is processed to remove noise and extract relevant features. Smart SHM systems in addition commonly implies the automation of the aforementioned processes and also include algorithms to identify trends, detect anomalies, and support decision-making.

SHM system can be grouped into different categories depending on the duration and frequency of the measurements, the measured quantities, and the techniques or installed instruments. Variations of crack opening, inclinations, settlements, and deformations due to external actions such as external temperature, humidity, wind, or connected to material degradation can be considered as quasi-static phenomena since they vary slowly over time. Systems that record such quantities, generally two to six times per hour for automated setups, are commonly referred to as Static SHM (SSHM) systems. In recent years, several structures, and in particular historical monuments, have been equipped with permanent static monitoring systems able to automatically record data over long periods of time and, theoretically, for the entire life of the structure.

Although in the last years, several studies have been focused on the interpretation of data recorded by dynamic monitoring systems installed in historical buildings [1–7], the development of reliable procedures for the interpretation of the data acquired by static monitoring systems is still a challenging issue.

To this purpose, the present research work has the two-fold objective of (i) defining a set of physically-based parameters that can be used for the interpretation of the recorded data and the detection of evolutionary trend (*Reference Quantities*) and (ii) formulating thresholds to define the intensity level of the anomaly in the monitored data (*normalized dataset for the data-driven anomaly detection*).

PHYSICALLY-BASED DATA-DRIVEN APPROACH FOR DATA INTERPRETATION

A SSHM system returns several time series $x(t)$ recorded by the installed instruments that can be decomposed into two main components [**?**, 8–10]:

$$x(t) = x_1(t) + x_2(t) \tag{1}$$

Typically, for an historical building, the principal monitored quantities $(x(t))$ include deformation parameters (relative displacements between two points), cracks openings, inclinations, and strains.

In Equation 1, $x_1(t)$ is the "pseudo-periodic" component of $x(t)$, indicating the oscillatory component due to the external actions including self-weight, gravitational loads, environmental actions like thermal effects, wind, humidity, etc.. $x_2(t)$ is the "evolutionary" component of $x(t)$ mainly associated to material degradation, time-varying irreversible soil settlements, etc. It is responsible for the evolution of the state of conservation typically due to damage accumulation. In most cases, the relative contribution of components $x_1(t)$ to the total recorded signal $x(t)$ is predominant with respect to components $x_2(t)$. Therefore, a simple visual inspection of the recorded time histories is not sufficient to detect the eventual underlying evolution of the state. This is particularly true for the case of historic buildings constructed with massive masonry walls whose usual time response (in absence of extreme events) is mainly governed by thermal effects.

In light of all these issues, it clearly appears that a meaningful data analysis is required to characterize the main properties of the two components $x_1(t)$ and $x_2(t)$. For this aim, specific descriptors, hereafter referred to as Reference Quantities (RQs) are here introduced. Then, their variation in time is investigated to characterize the main features of the two components.

The numerical values of the introduced RQs can be systematically processed through automatized procedures considering different time spans and then stored in a database. Then, such automatized procedure could be easily applied to several buildings belonging to a given typology (such as towers, churches, and palaces) in order to establish meaningful ranges for the "reference values" of different parameters (such as deformations, cracks openings, and inclinations). In detail, with reference to the j-th generic day, the following Reference Quantities are here introduced:

- ***daily amplitude*** representing the maximum daily variation on the recorded measurements:
$$\delta_{\text{day},j} = \max x(t_i) - \min x(t_i) \quad \forall t_i \in j\text{-th day} \tag{2}$$

- ***mean daily value*** representing the average value of measurements on a daily basis:
$$\mu_{\text{day},j} = \frac{1}{n_j}\sum_{i=1}^{n_j} x(t_i) \quad \forall t_i \in j\text{-th day} \tag{3}$$
where n_j is the total number of measurements of the j-th day;

- ***absolute daily residual of the mean value*** representing the difference between the mean value, recorded in the j-th day of the k-th year, and that recorded in the same j-th day of the k_0-th year (i.e. the reference year, generally corresponding to the first year of monitoring):
$$r_{\mu_{\text{day},j}}(k-k_0) = \mu_{\mu_{\text{day},j}}(k) - \mu_{\mu_{\text{day},j}}(k_0) \tag{4}$$

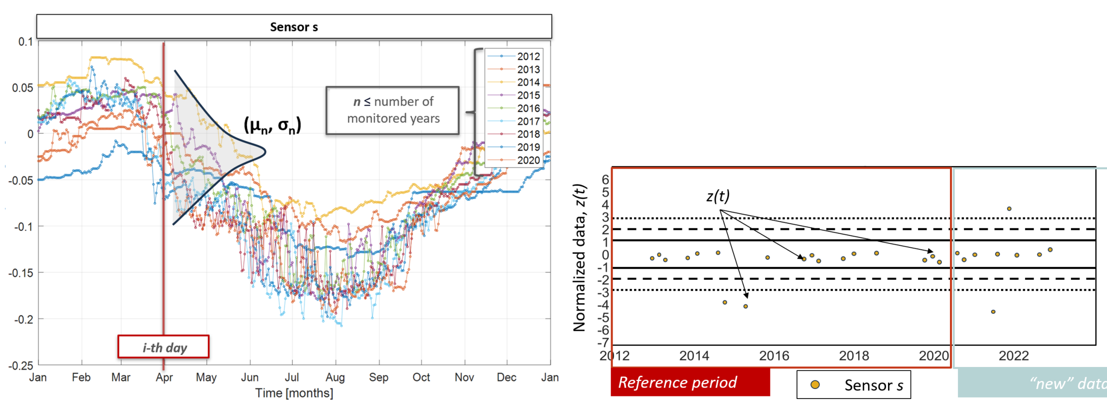
- ***progressive daily residual of the mean value*** representing the difference between the mean value, recorded in the j-th day of the $k_{+1}-th$ year, and that recorded in the same j-th day and the $k-th$ year:
$$rp_{\mu_{\text{day},j}}(k+1-k) = \mu_{\text{day},j}(k+1) - \mu_{\text{day},j}(k) \tag{5}$$

THRESHOLD DEFINITION FOR DATA-DRIVEN ANOMALY DETECTION

In this section, we explore the potentialities of the proposed data-driven interpretation approach with the aim of defining thresholds for the anomaly detection.

Given the definition of the RQs in the previous section, the monitoring data can be clustered into a new smaller dataset of physical quantities. Over the monitoring period, the reference quantities referred to as the time series $y(t)$ are systematically extracted from the dataset for each sensor j-th. The time series $y(t)$ inherits the effect of the seasonal variations which affect the monitoring data with an oscillatory component on a yearly basis. In order to account for this aspect, the time series $y(t)$ can be described as the k-th realization of a stochastic process, representative of the k-th year. Given a monitoring period defined over N years, for a fixed i-th day of the year (e.g., April 1st), as depicted in Fig.1a, there are n reconstructed values related to a given reference period P. Over the number of years identified by the reference period ($n \leq N$), the statistics (e.g. the mean and standard deviation (μ_n, σ_n) of the n values) are computed. Assuming a Gaussian probability density function for the distribution of $y(t)$, the dataset is normalized using the statistic descriptors (mean value and standard deviation).

The normalized dataset $z(t)$ expresses a statistical distance from the mean computed over the distribution of the referenced period. Hence, for each considered s-th sensor, given a reference period over n years, chronological control plots of the normalized dataset $z(t)$ - as represented in Fig.1b - are built and serve as an effective data-driven interpretation tool. In fact, with reference to the normalized values of RQ, the interval $-1 < z(t) \leq 1$ is associated to the 68% of the reference data distribution over the reference period. Analogously, the intervals $-2 < z(t) \leq 2$ and $-3 < z(t) \leq 3$ are associated to the 95% and 99.7% of the reference data distribution, respectively. According to the definition of the Cumulative Density Function (CDF) in a Standard Normal Distribution, the normalised values may be interpreted as threshold values with the following meaning. If the new actual RQ normalised value is equal to 1, it means that that specific value has been already either achieved or exceeded in the past monitoring period in 16 cases out of 100 (probability of exceedance P_E = 16%). If it is equal to 2, it means that that specific value has been exceeded only in 2.3 cases out of 100 (P_E =



(a) (b)

Figure 1. (a) Example of RQ dataset ($y(t)$) over several monitoring years, (b) Control plot of the normalized dataset.

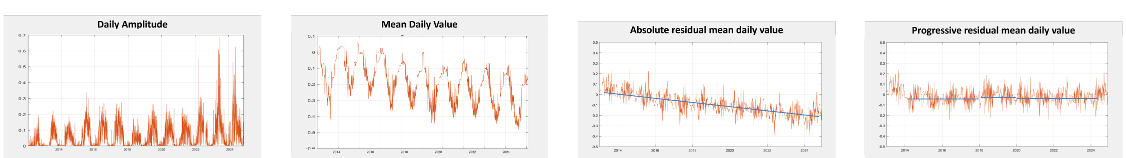
2.3%, i.e. rare occurence). If it is equal to 3, that specific value corresponds to a P_E = 0.13%, thus reflecting an extremely rare occurrence of the value. In the framework of practical applications related to structural monitoring, a low probability of exceedance is correlated to an high level of anomaly inherited in the considered feature (inverse correlation), meaning that values larger than these thresholds indicate potential structural anomalies requiring intervention.

APPLICATION TO A CASE STUDY

The methodology described in the previous sections is here applied to the case study of the Asinelly Tower (Bologna, Italy). Constructed between 1109 and 1119 by a wealthy family, as a symbol of their power, the Asinelli Tower stands as the tallest masonry tower in Bologna, with an height of 97 meters. During its lifespan it served different functions. During the XIII century, its ownership was transferred to the Municipality of Bologna as a civic tower, while during World War II, the tower was converted into a watchtower.

Nowadays the tower shows an overall inclination of approximately 1.7° with respect to the vertical and it tilts 2.23 meters to the southwest. Enclosed by a small ”stronghold” built in 1488, which currently hosts some artisanal shops, the plinth is built with a selenitic stone, as typically occurred for monumental buildings at that time in the area. The tower's outer walls consist of solid brick outer layers and a rubble stone masonry core, with the masonry wall thickness gradually diminishing from 3.15 meters at the base to 0.45 meters at the summit.

Since 2012, a continuous static SHM system has been installed to monitor the structural health of the tower. The system layout comprises **n°8 Long Base Deformeters (A-F)** deployed to monitor the deformation of the masonry close to the main wall width-reduction cross-sections; **n°5 Short Base Deformeters (A-D)** positioned at critical locations to track the width variations of the main cracks; **n°8 Biaxial Inclinometers (A-I)** measuring the in-plane and out-of-plane inclination of the masonry walls; **n°2 Termometers and 6 Hygrometers (A-T)** recording the temperature and the internal and external environmental condition affecting the tower. The installed system automatically collect measurements every 15 minutes. The monitoring activity has continued consistently for 13 years, hence the monitoring dataset approximately comprises more than 450'000 samples for each sensor.



(a) (b) (c) (d)

Figure 2. Daily RQs over the monitoring period for a given sensor: (a) daily amplitude, (b) mean daily value, (c) absolute residual mean daily value, and (d) progressive residual mean daily value. All data are reported in millimeters (mm).

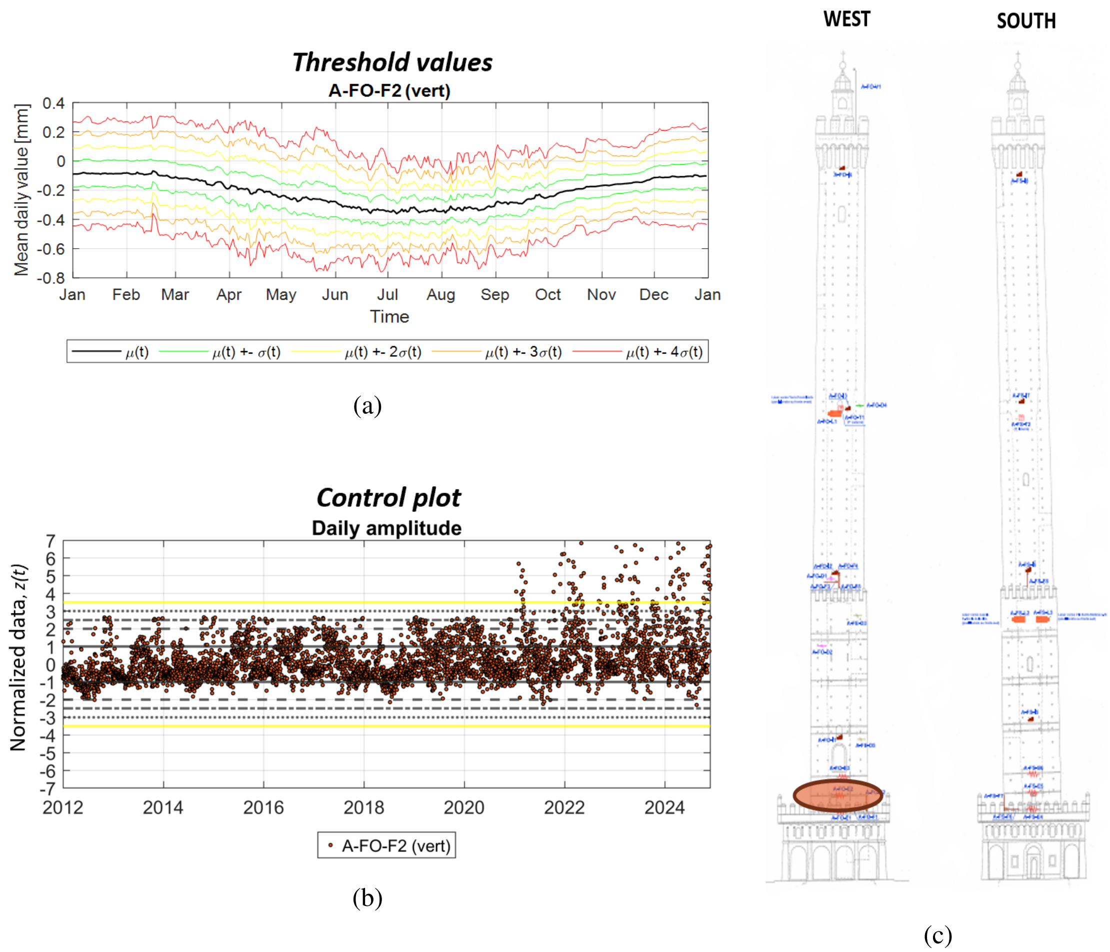


Figure 3. Example of threshold values and control plots for a given sensor for the considered case study: (a) threshold values, (b) control plot of daily amplitude, and (c) monitoring setup on the tower.

Following the proposed procedure, the Reference Quantities are evaluated for the complete measurement dataset through Eqs. (2)-(5), for each sensor. For illustrative purposes, the daily RQs for the long base deformeter installed at the base of the tower, namely sensor A-FO-F2, are reported in Fig.2.

The analysis of the daily amplitude (Fig.2a) highlights the seasonal influence of temperature on the data. Indeed, the amplitude is one order of magnitude greater in summer than in winter season. The analysis of the mean daily value (Fig.2b) over the monitoring period shows that the portion of masonry at the sensor location exhibits cyclic shortenings (summer-to-winter) and elongations (winter-to-summer) with a mean excursion of 0.24 mm over an instrumental base length of 150.6 cm. Moreover, a significative variation is observed in the values of the daily amplitude in the last years of monitoring and a decreasing trend can be qualitatively recognised in the plot of the mean daily value. However, since these quantities are specifically representative of the measured response of each single year, an immediate quantification of the identify phenomena over the entire monitoring period is not always easily derived. This demonstrates that the daily RQs expressing the daily amplitude and the mean daily value are particularly useful to trace the seasonal behaviour over the year, but they may not be particularly straightforward in quantifying the entity of an evolutionary trend. On the contrary the analysis of the absolute residual (Fig.2c) clearly highlights the presence of an evolutionary state that shows a mean cumulative increment of -0.36mm from the beginning of the monitoring period. Consistently, the analysis of the progressive residual (Fig.2d) shows that the detected trend has almost evolved constantly over the years with an average increment of -0.03 mm. This demonstrates that by referring to RQs expressed in term of residuals, (e.g. residual on the mean daily value) and assuming that the environmental effect is cyclic on an annual basis, this methodology allow to isolate and clearly define a quantitative measure of an evolutionary trend in the data.

To move forward in the direction of a smart management of data from an owner's perspective (e.g. real-time anomaly detection), the RQ dataset of the case study was normalized, as described in the previous section with respect to different reference periods. As previously anticipated, once a reference period is defined, the statistical parameters of the resulting distribution are used to define the threshold values for each sensor in each day of the year, as shown in Fig.3a for the illustrative example of the long base deformeter A-FO-F2. Contextually, the control plots showing the behaviour of normalized data $z(t)$ were also predisposed. Fig.3b shows an example of a control plot for the daily amplitude. In this case, to test the efficiency of the proposed methodology, the reference period was set from 2012 to 2019 ($n = 9$). Consistently with the analyses discussed in the previous paragraph and in comparison with Fig.2a, an anomaly in the original dataset is clearly identified in the control plot ($z(t) \geq 3$) for the considered sensor. This demonstrates that this approach can also be applied to process new monitoring data (e.g. on a platform collecting "real-time" measurements with the integration of automated data processing). Moreover, since the position and the ID of the sensor are known (Fig.3c), the anomaly detection can also lead to a direct prompt for on-site inspections.

CONCLUSIONS

The research results in the definition of a physically-based data-driven approach for data interpretation in long-term continuous static monitoring systems. The approach relies on a set of physically-based parameters (Reference Quantities) that are used for the analysis of the measured dataset and to detect evolutionary trends. Also, the description of the data in the normalized space allowed for a preliminary definition of threshold values associated with the statistical significance of probability of exceedance (and thus straightforwardly conveying a preliminary indication for identifying anomalies in monitoring data). Consequently, the proposed approach includes the description of the procedure for the generation of control plots to be used for data interpretation for a smart management of the monitored structure. In conclusion, to prove the effectiveness of the proposed approach, the methodology was applied to the case study of the Asinelli Tower in Bologna, showing significant improvement in the interpretation of the current condition of the tower and in the possibility of detecting the level and the location of anomalies in the tower's structural behaviour.

ACKNOWLEDGMENTS

The financial support provided through the S-MoSES (Smart Monitoring for Safety of Existing Structures and Infrastructures) Project (Research Project of National Interest, PRIN2020) by the MUR (Italian Ministry of University and Research) is gratefully acknowledged.

REFERENCES

1. Azzara, R. M., M. Girardi, V. Iafolla, D. M. Lucchesi, C. Padovani, and D. Pellegrini. 2021. "Ambient vibrations of age-old masonry towers: Results of long-term dynamic monitoring in the historic centre of Lucca," *International Journal of Architectural Heritage*, 15(1):5–21.
2. Beskhyroun, S., L. D. Wegner, and B. F. Sparling. 2012. "New methodology for the application of vibration-based damage detection techniques," *Structural Control and Health Monitoring*, 19(8):632–649.
3. Ivorra, S., F. J. Pallarés, and J. M. Adam. 2011. "Masonry bell towers: dynamic considerations," *Proceedings of the Institution of Civil Engineers-Structures and Buildings*, 164(1):3–12.
4. Magalhães, F., J. Pacheco, and A. Cunha. 2016. "Illustrating the relevance of SHM in a case study: the Foz Tua centenary railway bridge," in *8th European workshop on structural health monitoring, Bilbao, Spain*, pp. 5–8.
5. Robert-Nicoud, Y., B. Raphael, O. Burdet, and I. Smith. 2005. "Model identification of bridges using measurement data," *Computer-Aided Civil and Infrastructure Engineering*, 20(2):118–131.
6. Sohn, H., C. R. Farrar, F. M. Hemez, D. D. Shunk, D. W. Stinemates, B. R. Nadler, and J. J. Czarnecki. 2003. "A review of structural health monitoring literature: 1996–2001," *Los Alamos National Laboratory, USA*, 1(16):10–12989.
7. Tyagunov, S. and Y. Petryna. 2016. "Structural health monitoring and vulnerability assessment of buildings in earthquake prone areas," in *8th European Workshop on Structural Health Monitoring (EWSHM 2016)*, pp. 5–8.
8. Baraccani, S., G. Gasparini, M. Palermo, S. Silvestri, T. Trombetti, et al. 2014. "A possible interpretation of data acquired from monitoring systems," *CIVIL-COMP PROCEEDINGS*, 106:1–12.
9. Baraccani, S., M. Palermo, G. Gasparini, and T. Trombetti. 2021. "A time domain approach for data interpretation from long-term static monitoring of historical structures," *Structural Control and Health Monitoring*, 28(5):e2708.
10. Trombetti, T., S. Silvestri, G. Gasparini, M. Palermo, and S. Baraccani. 2015. "Reference quantities and values for a possible interpretation of the data acquired from monitoring system of historical buildings," *Computational Methods, Seismic Protection, Hybrid Testing and Resilience in Earthquake Engineering: A Tribute to the Research Contributions of Prof. Andrei Reinhorn*:111–126.