

A Dynamic Latent Variable Model for Monitoring the Santa Maria del Fiore Dome Behavior

Bruno Bertaccini(✉), Silvia Bacci, and Federico Crescenzi

Department of Statistics, Computer Science, Applications "G.Parenti", University of Florence, Florence, Italy
{bruno.bertaccini,silvia.bacci,federico.crescenzi}@unifi.it

Abstract. A dynamic principal component analysis is proposed to monitor the stability, and detect any atypical behavior, of the Brunelleschi's Dome of Santa Maria del Fiore, in Florence. First cracks in the Dome appeared at the end of the 15th century and nowadays they are present in all the Dome's webs, although with an heterogenous distribution. A monitoring system has been installed in the Dome since 1955 to monitor the behavior of the cracks; today, it counts more than 160 instruments, such as mechanical and electronic deformometers, thermometers, piezometers. The analyses carried out to date show slight increases in the size of the main cracks and, at the same time, a clear relationship with some environmental variables. However, due to the extension of the monitoring system and the complexity of collected data, to our knowledge an analysis involving all the detected variables has not yet conducted. In this contribution, we aim at finding simplified structures (i.e., latent common factors or principal components) that summarize the measurements coming from the different instruments and explain the overall behavior of the Dome across the time. We found that the overall behavior of the Dome tracked by multiple sensors may be satisfactorily summarized with a single principal component, which shows a sinusoidal time trend characterized, in a one-year period, by an expansive phase followed by a contractive phase. We also found that some webs contribute more than others to the Dome's movements.

Keywords: Monument monitoring system · Missing-data imputation models · Dynamic factor analysis · Forecasting · Architectural heritage preservation

1 Introduction

Santa Maria del Fiore is the cathedral of Florence, begun in 1296 it was completed in 1436. The Dome, engineered by Filippo Brunelleschi, was built in 18 years. With more than 4 millions bricks, Brunelleschi's Dome was the greatest architectural feat in the Western world (because the Brunelleschi's idea was to

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 47–58, 2020.
https://doi.org/10.1007/978-3-030-58811-3_4

build an octagonal dome higher and wider than any that had ever been built). It is characterized by eight webs resting on an octagonal supporting tambour and converging to the lantern that crown its top. First cracks in some webs appeared at the end of the 15th century. The evolving of the issue gave rise of concern and already in 1695 the Grand Duke of Tuscany established a first commission with the task to investigate the stability of the Dome. Nowadays, cracks are present in all webs, although particularly numerous in the even ones, mainly in webs 4 and 6 that are located on the opposite sides of the nave (Fig. 1).

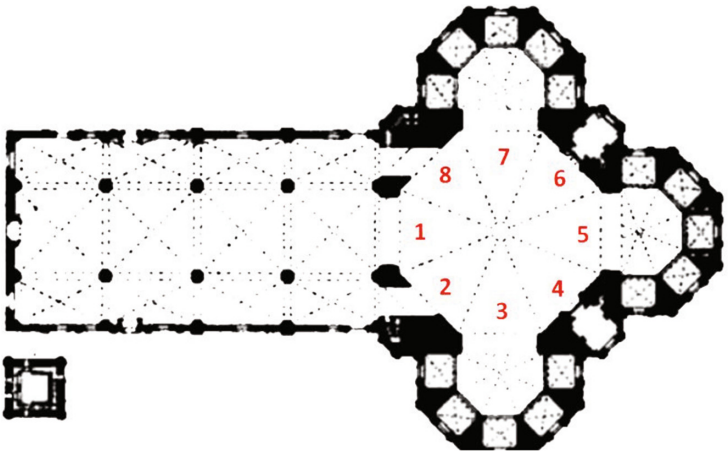

Fig. 1. Planimetry of the Santa Maria del Fiore church and dome, with numbered webs (source: Opera di Santa Maria del Fiore Foundation)

The monitoring system, installed in the Dome since 1955, today counts more than 160 sensors (mechanical and electronic deformometers, thermometers, piezometers, ...). Due to the variety and number of instruments involved, this monitoring system is one of the most accurate static control systems around the world installed on an historical building. The extension of the monitoring system and the amount of data collected by each sensor make it very difficult to obtain an overall evaluation of the static evolution of the Dome. However, the analysis carried out on data collected by single sensors applied upon the main cracks have allowed to highlight a constant pejorative increase in width over time, and a clear relationship with the main (exogenous) environmental (meteorological and seismic related) variables involved [5].

We are confident that the static "behavior" of the Dome and its ability to answer to the environmental action can be correctly interpreted jointly analyzing all the information gathered by the monitoring system in all these years. To our knowledge, such an analysis has never been presented in any of the studies conducted in the past. A first tentative in this direction was made by Bertaccini [2] who estimated a structural equation model to explain the behavior of the

fourth web (in which 13 deformometers and 3 thermometers are installed) in response to the external events (of exceptional and non-exceptional nature).

In this paper, we exploit the use of Generalized Dynamic Principal Components (GDPC) [7] to summarize the "breathing mechanism" of the entire Dome. Similarly to other methods of factor/principal component analysis, the GDPC carries out a dimensionality reduction of the vector of observations coming from different sources (in our case, the deformometers installed upon cracks), taking explicitly into account the correlation between consecutive observations. Thus, the amount of information characterizing the time series of the cracks width is synthesized in a small number of factors (principal components), which should simplify the understanding of the static evolution of the Dome as a whole.

The rest of the paper is organized as follows. After a brief presentation of the functioning of electronic sensors that constitute the monitoring system installed on the Dome, particular attention will be paid to the statistical techniques employed to impute missing data in case of sensors fault (Sect. 2). In general, a complete data matrix is needed when statistical models have to be estimated on the basis of a multivariate time series structure. The imputation process will be propedeutic to the estimation of a dynamic principal components model whose results will be presented and discussed in Sect. 3. Some final remarks and further perspective of analysis will be discussed in the conclusive section (Sect. 4).

2 Data from the Monitoring System

The electronic instruments taken into account in the analyses presented in this work are the 66 deformometers and 56 air and masonry thermometers installed by ISMES (Istituto Sperimentale Modelli e Strutture[1]) in 1987. These sensors record data every six hours (4 measures per day) since January 8, 1988. This means just less than 47.000 measures per sensor, for a total of about 6 million measurements. Data used in this work were limited to about the last 20 years of monitoring activity, from January 1, 1997 to February 28, 2017. Information (metadata) on the location of instruments are available and will be of primary importance in building the statistical model (for the sake of brevity, Fig. 2 shows only the location of the instruments installed on web 4).

The whole dataset (measurements and metadata) was provided by the Opera di Santa Maria del Fiore, in the context of a scientific collaboration with the Department of Statistics, Computer Science, Applications "G. Parenti" of the University of Florence under the responsibility of Prof. Bruno Bertaccini.

Given the extent of the monitoring system in terms of installed instruments, it was decided to reduce the dataset dimension using the daily average of the valid measures acquired by each sensor. The variations of temperature between night and day within the 24 h are considered irrelevant because the masonry is not able to react instantly to changes in air temperature, especially those within

[1] http://www.asim.it/ismes/IsmesUk/default.htm.

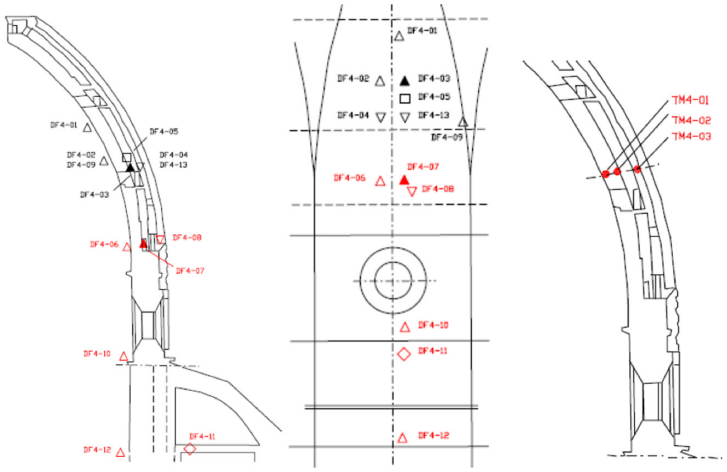

Fig. 2. Location of deformometers and thermometers on web 4 (source: Opera di Santa Maria del Fiore Foundation)

a day; thus, a very slight variation range is expected for the daily detections produced by each deformometer. Under these assumptions, arithmetic mean does not present drawbacks and has the advantage to solve any cases in which daily measurements, for temporary faults, are less than 4.

Figure 3 shows data from the deformometers installed in the even webs. The figure clarifies what is commonly referred as the "breathing" mechanism of the Dome in time: the cracks tend to expand and shrink cyclically according to seasons. This behavior is due to the structure of the Dome that may be assimilated to a physic "closed system", where the structural constraints define the relationship of forces between the various cracks, which in turn are subjected to the action of meteorological and other environmental variables [3].

The preliminary analysis of the measures acquired by the monitoring system highlighted the presence of missing data and outliers in almost all the series of data acquired by the deformometers as well as by the air and masonry thermometers. This phenomenon is generally due to storms and blackouts; it is probable that certain sensors (see, for instance, those represented in Fig. 4) have gone out of calibration due to the action of lightning, producing anomalous oscillations for a shorter or longer period. Due to their lack of (correct) information [1], outliers have been assimilated to missing data and preliminary deleted.

Unfortunately, statistical models for time series generally require a complete data matrix; otherwise, it would not be possible to take full advantage from time related information. Hence, 9 deformometers having more than 10% of missing observations have not been taken into account. With the detections acquired by the remaining 57 deformometers, we designed a statistical algorithm for the imputation of missing values.

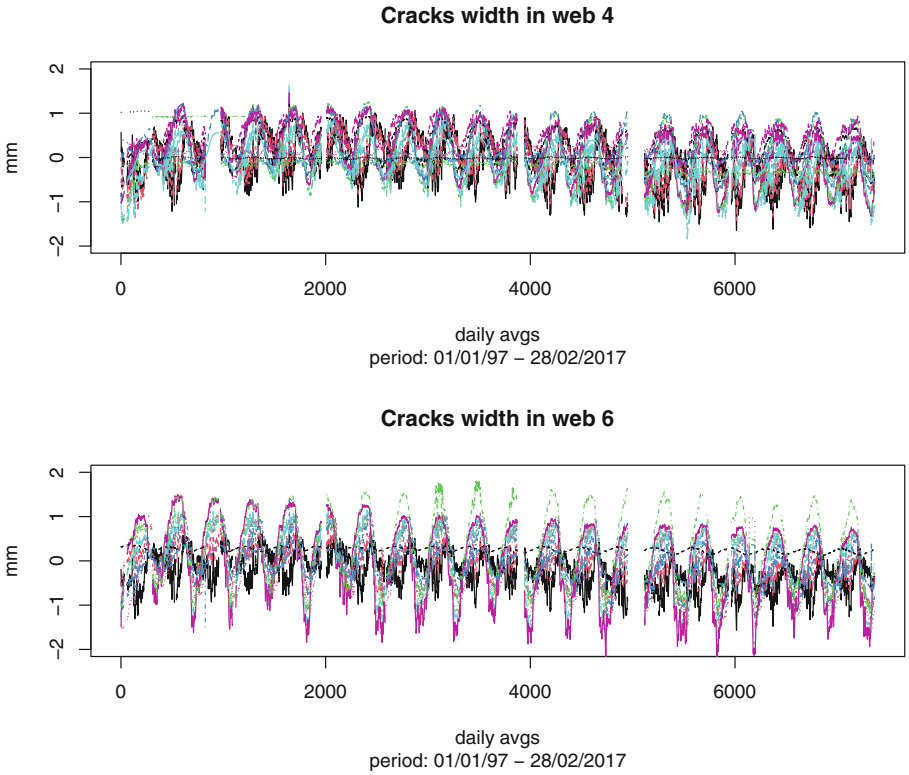

Fig. 3. Deformometers installed in the 4th and 6th webs: cracking pattern over time (width in *mm*)

On clean data, a five-parameters quadratic-sinusoidal regression model was fitted for each type of sensor (air thermometers, masonry thermometers, and deformometers) imposing a sine of period equal to the length of one year, to take into account both the seasonal fluctuation of the measurements or the possible variations produced by systematic changes in the structural framework. The regression model estimated for each sensor is thus formalized as:

$$y = \alpha + \beta t + \gamma t^2 + \delta \sin(\eta + t\frac{2\pi}{365}) + \epsilon, \tag{1}$$

where y is the measure detected by each sensor (i.e., air or masonry temperature and crack width), t is the number of days elapsed from the activation of the monitoring system, and ϵ is an error component that accounts for the stochastic nature of the relationship between y and t.

In the missing data imputation process, the quadratic-sinusoidal regression model allows us to impute an average value consistent with the information that each sensor has detected in the period (one or more days of the year) to which missing data refers to. The fitting indicators produced by each quadratic-

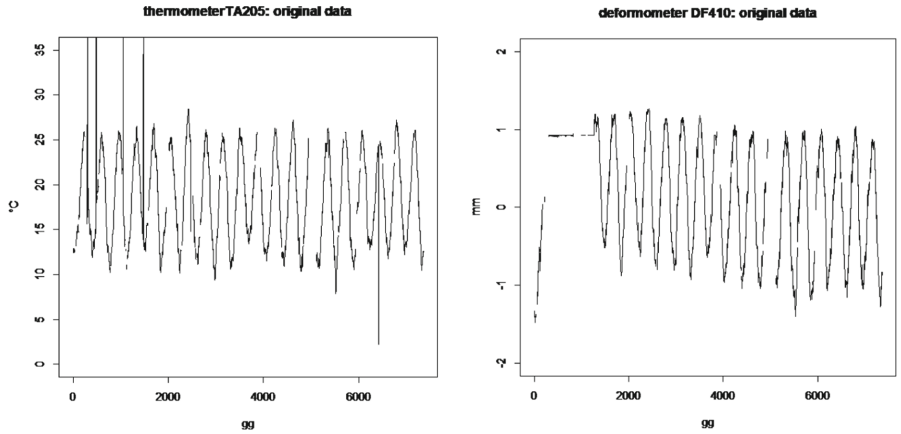

Fig. 4. Outliers and missing data in the cracking pattern over time measured by sensors TA2-05 and DEF4-10 (respectively, temperature in °C and width in *mm*)

sinusoidal regression model are very high. However, limiting the imputation process to the application of this model to each sensor independently of the values detected by the others was considered not completely satisfactory, because of the strong relationships among air temperature, web walls temperature, and cracks width.

For this reason, we adopted a cascade procedure that provided, through the use of linear regression models, the preliminary imputation of missing values present in the air thermometers time series. In turn, the complete series of air thermometers was used to impute missing data in the time series of the masonry thermometers. Finally, the complete series of masonry thermometers of each web were used to predict missing values of the deformometers installed in that web. In particular, for each air thermometer, the week of the year to which data refers to, the minimum, average and maximum daily temperature recorded in the city and the values estimated by the respective quadratic-sinusoidal regression model were provided as covariates (explicative variables) of the model used to impute missing values present in the corresponding series. For each masonry thermometer series, missing values have been imputed using the following covariates: the week of the year to which data refers to, the complete series of the air thermometers installed in the relative web and the estimations produced by the quadratic-sinusoidal regression model applied on the masonry thermometer series that has to be imputed. Similarly, missing values in the deformometer series have been imputed using a regression model with covariates: the week of the year to which data refers to, the complete series of the masonry thermometers installed in the relative web and the estimations produced by the quadratic-sinusoidal regression model applied on the deformometer series that has to be imputed. For the sake of brevity, the results of the imputation process are presented only for some sensors.

Figure 5 shows the result of the imputation process for the air thermometers TA7-03 installed on web 7: the imputation model shows an outstanding R-squared goodness of fit index ($R^2 = 98.15\%$).

Figure 6 shows the result of the imputation process for the masonry thermometers TM4-03 installed on web 4: the R-squared fitting index is excellent also in this case ($R^2 = 98.48\%$).

Figure 7 shows the result of the imputation process for the first three deformometers (DF2-01, DF2-02, DF2-03) installed on web 2. The fitting indices for those models were, respectively, 92,12%, 87,30% and 96,60%.

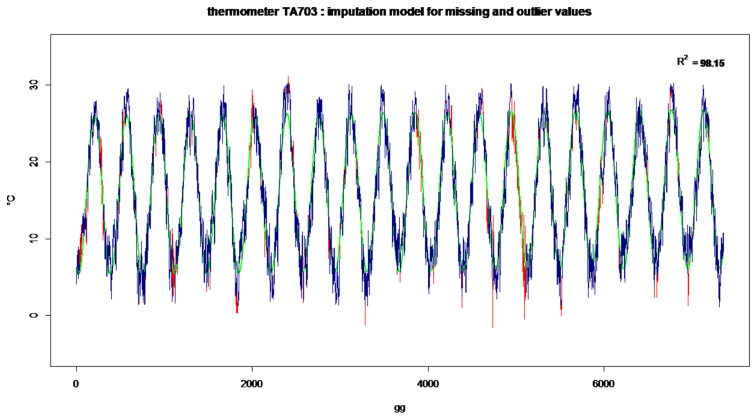

Fig. 5. Air-thermometer TA7-03: result of the imputation process (respectively, in green the seasonal fluctuation from the quadratic-sinusoidal regression model, in red the imputed values)

The R-squared goodness of fit indices computed for all the deformometers that compose the ISMES monitoring system are reported in Table 1. Almost all the indices show a very good performance of the relative model used to impute missing data. In very few cases, the fitting is unsatisfactory: these cases happen when the pattern of detections is quite inconsistent with the typical fluctuation of temperatures (i.e. when the sinusoidal regression model does not fit properly).

3 Modeling the Dome's Breathing Behavior

In this section, we present a suitable statistical approach to reduce the dimensionality of sensors measurements in order to facilitate the comprehension of the overall static evolution of the Dome. The adopted approach is based on Generalized Dynamic Principal Components (GDPC) which should be able to discover some synthesised (latent) trends of the Dome from the time series of the 57 deformometers installed on the eight webs. The suitability of the approach was verified by limiting the analysis to the measurements acquired in the last year.

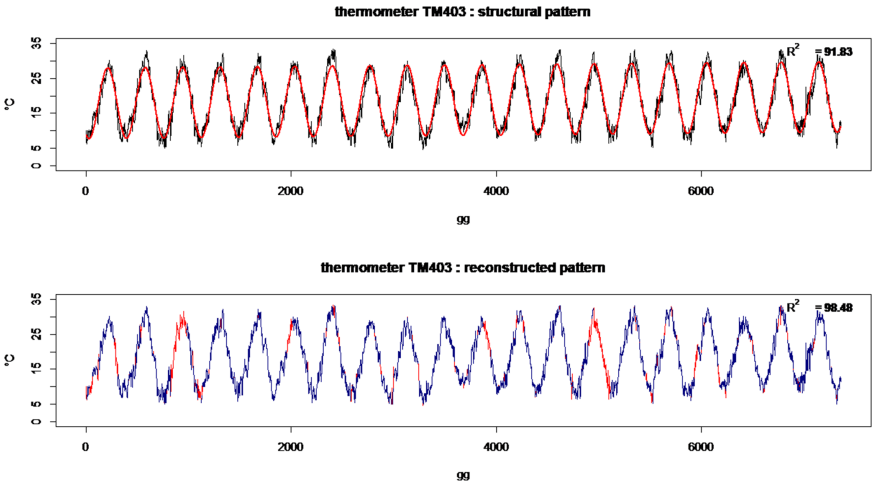

Fig. 6. Masonry thermometer TM4-03: the structural pattern of temperatures over time from the quadratic-sinusoidal regression model (top panel) and the result of the imputation process (bottom panel, in red the imputed values) (Color figure online)

Table 1. R-squared goodness of fit indices computed for all the ISMES deformometers

web 1		web 2		web3		web 4		web 5		web 6		web 7		web 8	
DF	R^2	DF	R^2	DF	R^2	DF	R^2	DF	R^2	DF	R^2	DF	R^2	DF	R^2
101	*.945*	201	*.921*	301	*.938*	401	*.818*	502	*.861*	601	*.643*	701	*.961*	801	*.807*
102	*.920*	202	*.873*	302	*.934*	402	*.722*	503	*.877*	604	*.718*	702	*.929*	802	*.941*
103	*.891*	203	*.966*	303	*.606*	404	*.741*	504	*.933*	605	*.872*	703	*.320*	803	*.923*
104	*.773*	204	*.849*			405	*.861*			606	*.790*			804	*.878*
105	*.957*	205	*.847*			406	*.787*			607	*.878*			805	*.941*
106	*.789*	206	*.948*			407	*.881*			608	*.911*			806	*.947*
		207	*.943*			408	*.861*			609	*.928*			807	*.902*
		208	*.943*			409	*.923*			610	*.906*			808	*.941*
		209	*.961*			410	*.798*			611	*.956*			809	*.688*
		210	*.937*			411	*.935*			612	*.920*			810	*.923*
						412	*.931*								
						413	*.672*								

Several methods have been proposed in the literature to deal with dimensionality reduction in time series analysis building on Dynamic Factor Analysis (DFA; examples are provided by [4,10] and references therein). DFA models are closely related to GDPC models because they assume that one important part of the original series can be explained in a dynamic way by a relatively small number of common factors. However, while DFA relies on a series of assumptions

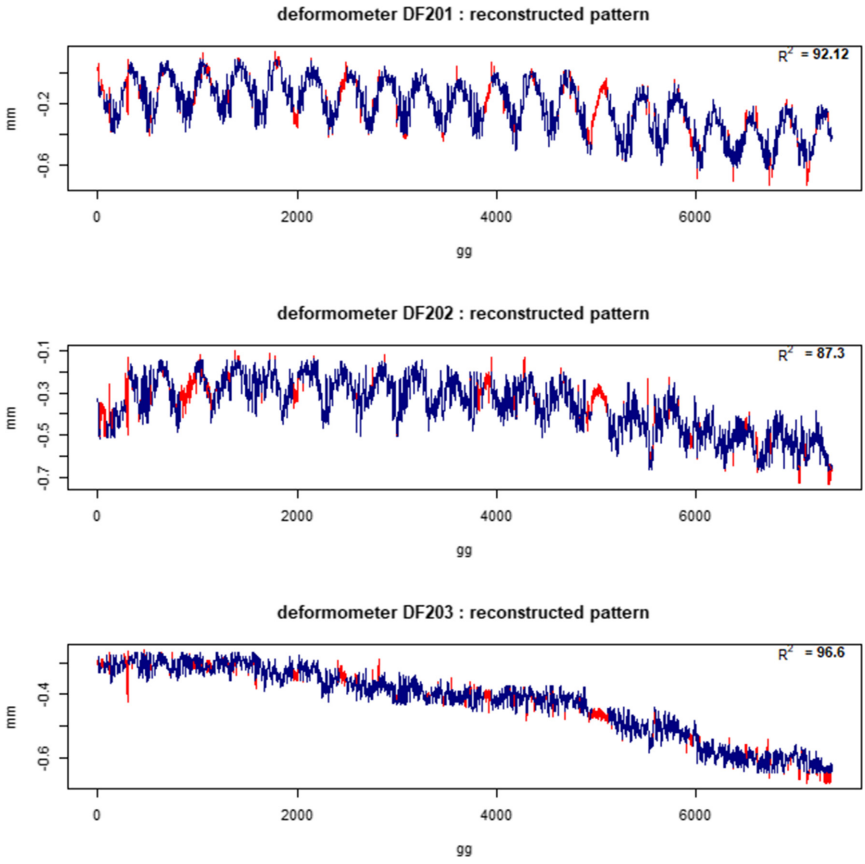

Fig. 7. Deformometers DF2-01 to DF2-03: result of the imputation process (in red the imputed values) (Color figure online)

regarding the distribution of error terms and on the variance-covariance matrices of the state-space equations, GDPC is more flexible because it only assumes that the time series share a common latent factor.

Let $z_t = (z_{1t}, ..., z_{mt})'$ denote a time series vector, where m is the number of series and $1 \leq t \leq T$. The first dynamic principal component with k lags is defined as the vector $\boldsymbol{f}_t = (f_t)_{-k+1 \leq t \leq T} = (f_{t-k}, ..., f_{t-1}, f_t)'$ such that the reconstruction of the series z_{jt} $(1 \leq j \leq m)$, defined as a linear combination of $f_{t-k}, \ldots, f_{t-1}, f_t$, is optimal with respect to the mean squared error (MSE) criterion. More precisely, given a factor $\boldsymbol{f}_t$ of length $(T+k)$, an $m \times (k+1)$ matrix of coefficients or loadings $\boldsymbol{\beta} = (\beta_{jh})_{1 \leq j \leq m, 1 \leq h \leq k+1}$ and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)'$, the reconstruction $\tilde{z}_{jt}$ of original series z_{jt} is defined as $(1 \leq j \leq m;\ 1 \leq t \leq T)$

$$\tilde{z}_{jt}(\boldsymbol{f}_t, \boldsymbol{\beta}, \alpha_j) = \alpha_j + \sum_{h=0}^{k} \beta_{j,h+1} f_{t-h}. \tag{2}$$

Note from Eq. 2 that f is not depending on the specific series j. The MSE to minimize is

$$MSE(\boldsymbol{f}_t, \boldsymbol{\beta}, \alpha_j) = \frac{1}{mT} \sum_{j=1}^{m} \sum_{t=1}^{T} [z_{jt} - \tilde{z}_{jt}(\boldsymbol{f}_t, \boldsymbol{\beta}, \alpha_j)]^2. \tag{3}$$

The minimization routine is efficiently implemented in the `gdpc` R package [6]. The original and reconstructed time series for deformometers DF101-104 are displayed in Fig. 8.

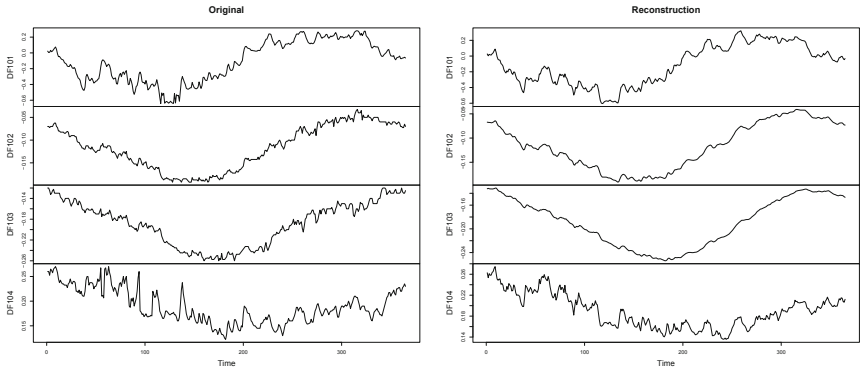

Fig. 8. Original and reconstructed time series for deformometers DF1-01 to DF1-04.

For the last year data the procedure took approximately 3 days on a 16 vCPU virtual machine with a RAM of 64 GB. The goodness of the reconstruction is evident, being the MSE of the reconstruction equal to 0.066. Indeed, the fitted model with one dynamic principal component with four lags, which represents the common factor underlying the time series of the 57 deformometers, explains the 93.4% of the whole variance in the dataset. The estimated time series of the common factor as well as the loadings for the 57 deformometers are displayed in Fig. 9.

We interpret the smooth behavior of the factor as the breathing mechanism of the Dome. Figure 9, top left panel, shows the values of the first principal component along the time, suggesting that, during a one year period, the Dome has a sort of sinusoidal trend with a period of expansion, followed by a period of contraction. This is corroborated by the estimates of loading parameters β_{jh} ($j = 1, \ldots, m;\ h = 1, \ldots, k+1$) that move together, alternating positive and negative values (Fig. 9). More in detail, measures of deformometers detected at 0, 2, and 4 lags (Fig. 9, panels "0 loading", "2 loading", and "4 loading") contribute, with a few exceptions, to the expansive phase, whereas measures detected at 1 and 3 lags (Fig. 9, panels "1 loading" and "3 loading") contribute to the contractive phase. Also, we notice that the highest loadings are due to cracks recorded by the deformometers in webs 4 (blue lines in Fig. 9) and 6 (pink

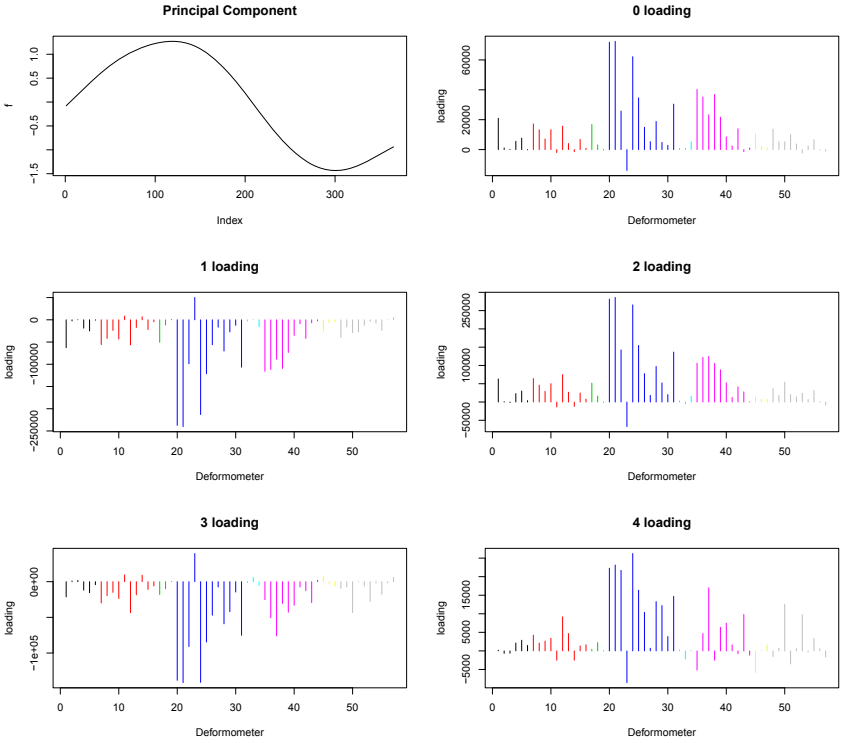

Fig. 9. GDPC results

lines in Fig. 9). In other words, webs 4 and 6 provide the maximum contribution to the movements of the Dome.

4 Conclusions

In this contribution, we performed an analysis of the web cracks reported by the sensors installed on the Brunelleschi's Dome of Santa Maria del Fiore. The complex data structure, resulting by the 57 deformometers over a period of 20 years with some random blackouts, suggested to adopt an approach based on two steps. First, we imputed the missing observations adopting a quadratic-sinusoidal regression model in order to build the complete time series for each deformometer. Second, a dynamic principal component model has been estimated to identify the overall breathing mechanism of the Dome and the contribution of each web crack to the overall movement. This model showed an excellent goodness of fit, clearly revealing the underlying breathing mechanism of the Dome. Compared to the dynamic factor analysis performed in [2], where the algorithm did not converged but stopped after reaching the maximum number of iterations, the GDPC provided much better convergence properties (results not shown here). The proposed analysis represents a useful instrument for the professionals (architects

and structural engineers) to globally evaluate the behavior of the Brunelleschi's Dome. In fact, this approach gave satisfactory results in financial applications (see [9]) and its consistency has been investigated in [8]. For the future developments of the work, we intend to relate the movements of the webs each other as well as to assess the impact of external observable (exogenous) variables on the evolution of the cracks.

Acknowledgements. Authors thank the Opera di Santa Maria del Fiore Foundation for providing the data acquired by the monitoring system installed on the Dome. Authors also acknowledge the financial support provided by the "Dipartimenti Eccellenti 2018–2022" Italian ministerial funds.

References

1. Barnett, V.: Outliers in statistical data. Technical report (1978)
2. Bertaccini, B.: Santa Maria del Fiore dome behavior: statistical models for monitoring stability. Int. J. Architectural Heritage **9**(1), 25–37 (2015)
3. Chiarugi, A., Bartoli, G., Bavetta, F.: La meccanica della cupola. In: Cupola di Santa Maria del Fiore, il Cantiere di Restauro 1980–1995, pp. 47–62. Istituto Poligraco e Zecca Dello Stato (1995)
4. Forni, M., Hallin, M., Lippi, M., Reichlin, L.: The generalized dynamic-factor model: identification and estimation. Rev. Econ. Stat. **82**(4), 540–554 (2000)
5. Ottoni, F., Blasi, C.: Results of a 60-year monitoring system for Santa Maria del Fiore dome in Florence. Int. J. Architectural Heritage **9**(1), 7–24 (2015)
6. Peña, D., Smucler, E., Yohai, V.J.: gdpc: an R package for generalized dynamic principal components. J. Stat. Softw. **92**(1), 1–23 (2020)
7. Peña, D., Yohai, V.J.: Generalized dynamic principal components. J. Am. Stat. Assoc. **111**(515), 1121–1131 (2016)
8. Smucler, E.: Consistency of generalized dynamic principal components in dynamic factor models. Stat. Probab. Lett. **154**, 108536 (2019)
9. Trucíos, C., Hotta, L.K., Pereira, P.L.V.: On the robustness of the principal volatility components. J. Empir. Finance **52**, 201–219 (2019)
10. Zurr, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R., Beukema, J.J.: Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics **14**, 665–685 (2003)