



## STATISTICAL ANALYSIS OF REINFORCED CONCRETE BRIDGES IN ESTONIA

Sander Sein<sup>1</sup>✉, Jose Campos Matos<sup>2</sup>, Juhan Idnurm<sup>3</sup>

<sup>1,3</sup>Tallinn University of Technology, Ehitajate tee, 5, Tallinn 19086, Estonia

<sup>2</sup>University of Minho, Campus de Azurém, Guimarães, 4800-058, Portugal

E-mails: <sup>1</sup>sander.sein@ttu.ee; <sup>2</sup>jmatos@civil.uminho.pt; <sup>3</sup>juhan.idnurm@ttu.ee

**Abstract.** This paper introduces a possible way to use a multivariate methodology, called principal component analysis, to reduce the dimensionality of condition state database of bridge elements, collected during visual inspections. Attention is paid to the condition assessment of bridges in Estonian national roads and collected data, which plays an important role in the selection of correct statistical technique and obtaining reliable results. Additionally, detailed overview of typical road bridges and examples of collected information is provided. Statistical analysis is carried out by most natural reinforced concrete bridges in Estonia and comparison is made among different typologies. The introduced multivariate technique algorithms are presented and collated in two different formulations, with contrast on unevenness in variables and taking into account the missing data. Principal components and weighing factors, which are calculated for bridges with different typology, also have differences in results and element groups where variation is retained.

**Keywords:** bridge management, condition assessment, multivariate analysis, principal component analysis, statistical analysis, visual inspections.

### 1. Introduction

Bridges play an essential role in an infrastructure network, a famous Serbian writer has introduced them as follows: “Of everything that man erects and builds in his urge for living nothing is in my eyes better and more valuable than bridges. They are more important than houses, more sacred than shrines. Belonging to everyone and being equal to everyone, useful, always built with a sense, on the spot where most human needs are crossing, they are more durable than other buildings, and they do not serve for anything secret or bad.” – Andrić (2015), Nobel Prize Winner in 1961.

Due to the economic and societal risk of failure of bridges, it is vital for asset managers and stakeholders to implement adequate management systems to ensure the risk of failure is under stated performance criteria (Mueller, Stewart 2011). Many authorities, including Estonian Road Administration, have implemented a management system to monitor the condition of their existing bridge network (Lauridsen *et al.* 1998). A Bridge Management System (BMS) is a systematic and rational approach to perform all management activities related to managing a bridge stock (Scherer, Glagola 1994). It usually includes planning intervention activities to fulfil the serviceability requirements and using computational tool to track, record and process the results of management actions (Lauridsen *et al.* 1998). Bridge Management System

comprises coordinated activities to realize their optimal value, which involves balancing of costs, risks, opportunities and performance requirements. This framework allows asset managers to plan further assessment and intervention (Matos *et al.* 2015) often rely on collecting condition ratings of structures using visual inspections (Das 1998; Estes, Frangopol 2005).

Unified visual inspections, as the most basic level in the assessment of existing road structures (Rücker *et al.* 2006), have been used on Estonian national road bridges for more than 10 years and more than 40 000 element condition states have been recorded during the period. Although there are only 995 bridges, it is difficult to discern, which bridge is the worst state or what kind of intervention it is possible to carry out. One solution, to make decision-making more efficient, it is possible to exploit modern computational methods to perform “big data” analysis, which allows the extraction of information from a large dataset for descriptive and predictive purposes, using statistical techniques (Manyika *et al.* 2011). The purpose of data-reduction in multivariate analysis is to represent the original data with suitable lower-dimensional space, which helps to enable visualisation and discover data structures and patterns (Martinez *et al.* 2010). This paper investigates the use and applicability of linear multivariate analysis method called Principal Component

Analysis (PCA), where different algorithms are applied on collected visual inspection database of most common reinforced concrete bridge typologies in Estonia to see the possibilities of statistical analysis and to point out elements with higher variance.

Hanley *et al.* (2015; 2016) have previously investigated Principal Component Analysis applicability to condition rating data of BMS by integrating this technique into a network of road bridges in Ireland and Portugal. In conclusion, the PCA is indicated to be a viable tool in the assessment of large data sets relating to engineering applications, and it is possible to convert results directly to weighing factors of condition ratings. It was also suggested to investigate other typologies with more detailed description (Hanley *et al.* 2016).

In this paper, visual inspection database and most ordinary reinforced concrete bridges of Estonian national road network are introduced. The difference with previous research is an additional comparison of two algorithms of PCA and element groups with greater variance are pointed out. Two different algorithms are compared to show the possibility of making wrong decisions based on results of PCA. This paper attempts to address this issue by choosing appropriate indicators and giving a rational basis for modelling technique and decision-making. From an engineering perspective, decisions are naturally the result of a well-structured reasoning which justifies the selection of the final solution. Decisions made because of a logical, scientifically structured process are considered as rational decisions (Sánchez-Silva, Klutke 2016).

Since there is more than one possible decision, for the better outcome the use of statistical techniques should also be considered. Many techniques have been developed for this purpose, but PCA is one of the oldest and most widely used (Jolliffe, Cadima 2016). It is first formalized by Pearson (Pearson 1901) and described in its algebraic form by Hotelling (Hotelling 1933). The technique identifies the component ratings which are more important than the others regarding explained variability and reassesses these relative importance ratings of different factors from an engineering point of view, based on additional data (Hanley *et al.* 2015). Principal Component Analysis is dimensionality-reduction method where a set of original variables are replaced by an optimal set of derived variables, called Principal Components (PCs) (Jolliffe, Cadima 2016).

**Table 1.** Description of condition state ratings in Estonian BMS

Condition state rating	Description, action
1	Element in good condition, no rehabilitation needed
2	Element in fair condition, minor rehabilitation, and local repair needed
3	Element in bad condition, repair needed
4	Element in critical condition, major repair or replacement needed

One solution is to use the results as an input for the bridge condition calculation or predictive models for different typologies of bridges by calculating the importance of relevant elements which define the performance of individual bridge based on visual inspections. The results are also useful for comparing relative components of bridges with different typology, age, traffic intensity, exposures or other environmental situations.

## 2. Description of dataset

The average age of Estonian national road bridges is 40 years, which indicates the necessity to have an overview of their condition and to make correct decisions to preserve them as long as possible. The implementation of unified visual inspections and management in Estonian national roads started in 2003; the system was based on United States system PONTIS, which was initially used to monitor the state of existing bridge stock and to predict the future condition of bridges (Thompson *et al.* 1998). First official inspections were carried out in 2005, and within this period every bridge is visually inspected twice and 400 of them already three times.

Assessment of a bridge consists of inspecting every element unit of the bridge and evaluating each with a condition rating based on a scale of damage present and necessary rehabilitation method (Table 1). An overall element condition index is calculated based on the overall quantity of units and state factors, as shown in Eq (1):

$$H_e = \frac{\sum_s k_s q_s}{\sum_s q_s} \cdot 100\% , \quad (1)$$

where  $H_e$  – condition state of element;  $s$  is condition state;  $k_s$  – coefficient of state and  $q_s$  is the amount of units in current state. Bridge condition index is calculated based on element condition index and weight factor. The overall condition rating of the bridge, which is often the primary decision criterion of investments, has misleading impact because different states of element deterioration possess equal condition ratings in overall.

A central element in life-cycle assessment (LCA) involves making predictions about the degradation of the system. It requires a clear understanding of the physical laws which define the system behaviour and possible uncertainties. The degradation of a bridge condition describes the process by which one or a set of elements lose value with time (Sánchez-Silva, Klutke 2016). In PONTIS, the future condition states of visual inspections are processed with Markov chain method, depending on the assumption whether the interventions are performed during the time frame between inspections (Thompson *et al.* 1998).

From an overall number of 995 Estonian national road bridges, 778 are constructed of reinforced concrete. Although the number of reinforced concrete bridges has decreased in last decade, by replacing them with soil-steel composite bridges, it is still most popular construction material in Estonia.

There are 4 main typologies for reinforced concrete bridges, which are divided into 7 different groups according to main girder and construction year, these are:

1. Mounted simply supported beams with diaphragms (constructed after 1956).
2. Mounted simply supported beams (constructed after 1963).
3. Cast in site simply supported slabs (constructed before 1960).
4. Mounted or cast in fragments simply supported slabs (constructed after 1960).
5. Cast in site simply supported beams (constructed before 1956).
6. Mounted frames (constructed after 1978).
7. Cast on site simply supported cantilevers (constructed before 1956).

These typologies make 98% of all reinforced concrete bridges. The primary concentration is on bridges that were constructed or repaired before 2005, when first unified visual inspections were carried out, because every kind of intervention is changing the natural degradation process. In last 10 years, the average amount of bridges being built or reconstructed is 41, and since the main criteria of multivariate analysis is to exclude bridges with intervention, there has been a filtering process before analysing most common typologies. After excluding repaired and reconstructed structures, only 5 main typologies with a dataset of 501 reinforced concrete bridges remain (Table 2). To be clear, Group No. 1 consists of both simple span bridge typologies constructed after 1956 and Group No. 4 consists of the simple span and cantilever bridges.

As mentioned before, the visual inspection results are recorded on an element basis. It is possible to select from 124 different elements, which are distinguished by type and overall dimensions. For example, there are 5 different elements for piers with different width.

Within this analysis, the elements are divided into 16 different groups (Table 3) to investigate overall patterns in specific typology. This division of element groups was made based on available element observations and structural integrity, which describes the bridge performance where structural components have bigger effect on the overall load bearing capacity than non-structural. The condition ratings for each group are calculated from every represented element condition and are on the scale of 1–4 in increasing order as condition rating for elements (Table 1). No weight factors are used in the calculation of group and condition ratings are average results of elements.

Another limitation of data usage is due to the peculiarity of bridge inspections, which are carried out in different time frame. Bridges have been inspected seasonally or in so-called cycles, where assessment of all structures is done in 3 or 4 years. The first cycle was in 2005–2007, second in 2010–2013 and the third one started in 2015. An example of average condition indexes of all inspection cycles is shown in Fig. 1, where calculated average conditions are shown.

Due to the nature of deterioration process, without any intervention, the element state has to be higher, and within the period between first two assessments, the condition index is higher or slightly better. Improvement of the condition is normally caused by the subjectivity of inspector or unrecorded maintenance works. Nevertheless, differences between condition state of the second and third cycle are much more significant, and they describe the situation where most of the elements are getting into better condition state without any intervention. The differences are mainly emerging because of insufficient information, caused by the situation where all the bridges are inspected twice, and only half of them three times. In further analysis, the primary attention is paid to 501 bridges with two different results.

### 3. Formulation of principal component analysis and algorithms in case of missing data

The primary purpose of the PCA is to reduce the dimensionality of a set of data and redefine the input variables as principal components (PCs). It is a linear combination of the original variables, having fewer variables than the original dataset while preserving most of the information (Hotelling 1933; Jolliffe 2002). The first principal component  $Y_1$  is defined as Eq (2):

**Table 2.** Main typologies of Estonian national road bridges without interventions

Group No.	Type	Construction year	Amount
1	Simple span beam	1956–...	195
2	Slabs	...–1959	86
3	Slab in fragments	1960–...	118
4	Simple span and cantilever beam	...–1955	34
5	Frames	1978–...	68
Total			501

**Table 3.** Element groups divided by position

Non-structural elements	Number	Structural elements	Number
Overlay	1	Deck plate	9
Barriers	2	Edge beam	10
Handrails	3	Piles and columns	11
Drainage	4	Abutment cap, crossbeam	12
Slopes	5	Wing wall, front wall, foundation, abutments	13
Deformation joints	6	Diaphragms	14
Other (river bed, snow nets, signs)	7	Main girder	15
Waterproofing	8	Bearings	16

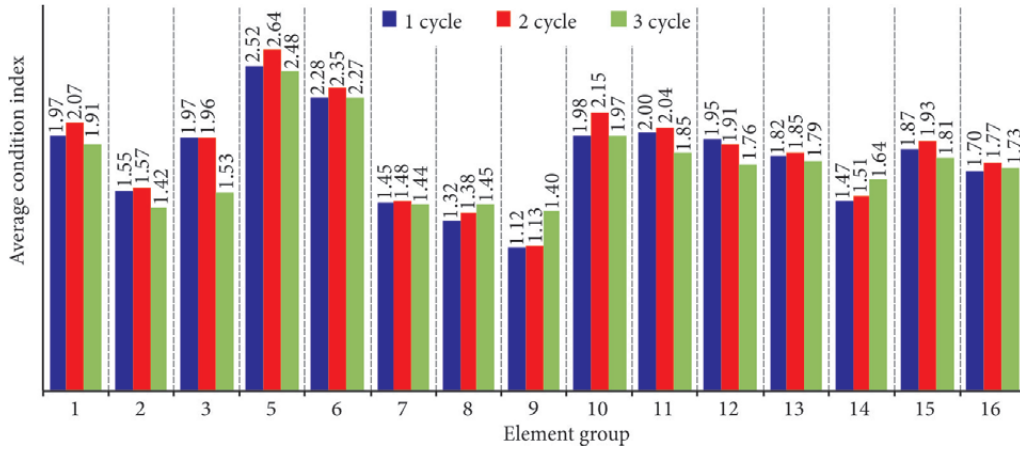


Fig. 1. Average condition indexes of element groups for simply supported beams (Group No. 1)

$$Y_1 = \alpha'_1 x = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1n}x_n = \sum_{j=1}^p \alpha_{1j}x_j, \quad (2)$$

where  $\alpha'_1 x$  – a linear function of the elements;  $x$  having maximum variance;  $\alpha$  – a vector of  $p$  coefficients  $\alpha$ .

The first principal component is the direction along which the data set shows the largest variation (Ringnér 2008), and the second component is determined under the constraint of being orthogonal to the first component and to have the largest variance (Abdi, Williams 2010). The second principal component  $Y_2 = \alpha'_2 x$  is found in a similar manner to the first principal component, and so on for the subsequent principal components up to  $p$  PCs. Recommendation is given for the number for PCs accounting the variance in the data set, which must be significantly lower than all the calculated components (Jolliffe 2002).

As described by Hanley *et al.* (2015), it is possible to use the sum of the square of PC coefficients  $\alpha_j$  for each variable, because the sum of coefficients are equal to unity and it is a better indicator in comparison of results. Weighing factors are derived from coefficients as shown in Eq (3):

$$\zeta = \sum_{j=1}^p \lambda_j x_j, \quad (3)$$

where  $\zeta$  – a combination of weighing factors based on  $\lambda_j = \alpha_{1j}^2 \cdot 100\%$  and original condition ratings  $x_j$ .

The previous formula, discussed in matrix terms, where a PCA is conducted through an Eigenvalue Decomposition (EVD) or a more robust and generalized Singular Value Decomposition (SVD) (Chambers 1977).

For a data matrix  $X$  of  $n$  observations on  $p$  variables measured by their means Eq (4):

$$X = ULA', \quad (4)$$

where  $L$  – an  $(r \times r)$  diagonal matrix;  $U$  and  $A$  –  $(n \times r)$  and  $(p \times r)$  matrices, respectively, with orthonormal columns, and  $r$  – the dimensionality of  $X$ .

SVD approach to PCA is shown to be computationally efficient and generalized method to determining the PCs.

As in previous investigations by Hanley *et al.* (2016) in using PCA for analysing BMS data, there is a problem with “missing data” in data-subsets, which describes a situation where a statistical difference of observations in among element groups appears. In multivariate analysis, it is often possible to use the existing structure of the data to estimate the missing data and complete a dataset, for example using Alternating Least Square (ALS) algorithm.

The algorithm alternates between imputing the missing values in and applying standard PCA to the in-filled (complete) data matrix (Jolliffe 2002). Initially, the missing values are replaced by the row-wise means of the previous matrix. The covariance matrix of the complete data are then estimated without the problems of principal components being more abundant than estimated variances and situation where the covariance matrix is negative, as in Eq (5)–(6) (Ilin, Raiko 2010).

The Alternating Least Square algorithm alternates among the updates:

$$X = (A'A)^{-1} A'U', \quad (5)$$

$$A = U'X'(XX')^{-1}, \quad (6)$$

where  $X$  –  $(p \times n)$  matrix of principal components;  $U$  and  $A$  –  $(n \times r)$  and  $(p \times r)$  matrices, as stated in Eq (3).

This iteration is efficient when only a few principal components are needed, so the number of principal components must be significantly lower than dimensionality of data vectors. (Roweis 1998). Alternating Least Square alternates among imputing the missing values in updated matrix, by replacing the values with row-wise mean values of original matrix. The covariance matrix of updated data is estimated using bias term  $M$  and matrix  $A$  will be computed using EVD. Principal components are calculated using Eq (7) (Ilin, Raiko 2010):

$$X = A^T (Y_c - M), \quad (7)$$

where  $Y_c$  states centred principal component and  $M$  states row-wise mean values of original matrix.



For better estimation of missing values, the computation is reconstructed as in Eq (8) (Ilin, Raiko 2010):

$$Y_c = \begin{cases} Y \\ AX + M' \end{cases}, \quad (8)$$

where  $Y$  will be used for observed values and  $AX + M$  for missing values.

In case of condition rating for a bridge element it is considered inappropriate to complete the dataset with algorithms, because unfilled rating usually indicates missing element (Hanley *et al.* 2016).

The suggestion has also been made for variables, which means PCA should only be conducted on continuous variables conforming to a Gaussian distribution (Qian *et al.* 1994), and its application to discrete data, such as element condition state ratings, are inaccurate. However, so long as inferential techniques requiring the assumption of multivariate normality are without reference, there is no necessity for the variables in the data set to have any associated probability distribution (Jackson 2003).

It is often considered wise to use the correlation matrix for a PCA, as the standardized varieties are dimensionless and more readily compared (Jolliffe 2002). However, when the variables are measured in the same units and have a low variance, using the covariance matrix is sometimes appropriate, and it is beneficial when statistical inference is essential. In case, when the condition ratings are already dimensionless, it is unnecessary to entirely standardise the variables (Hanley *et al.* 2015). Since condition ratings in Estonian BMS are already dimensionless, the raw data is used as input for a PCA.

#### 4. Results

The Principal Component Analysis was conducted on five most common reinforced concrete bridge types shown in Table 2. In every specified typology, there was a different number of element groups present, and in some typologies, there were only a few records available on the

condition of drainage, deformation joints or bearings, so these elements were excluded from analysis.

The comparison was made in two different steps. At first, two different algorithms of PCA were compared, concentrating on some useful PC and coefficients. Secondly, differences in PCA results between different typologies were compared and possible input for weighing factors suggested.

#### 4.1. Comparison of algorithms

To compare different algorithms with identical data, suggested script by Matlab® was used. An example of results using ALS and SVD algorithms is presented in Fig. 2 and Table 4 where the mean values of condition states and included some observations of element groups are shown. As stated in the previous chapter, ALS algorithm fills missing data with row-mean values and input for calculating PCs are based on artificial data, which is different from SVD algorithm.

Differences in Fig. 2 are small when considering the condition state is within the limits of 1–4, but within the period of two different inspections, the condition state has changed in similar calibre, which means the decision-making process is influenced by artificial results.

In Table 4, only element group “Other” results have remained unchanged. The reason is the same amount of

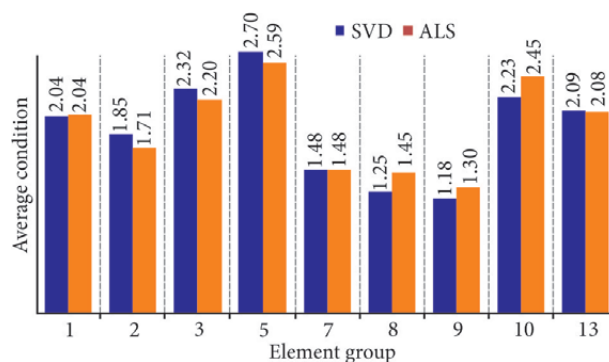
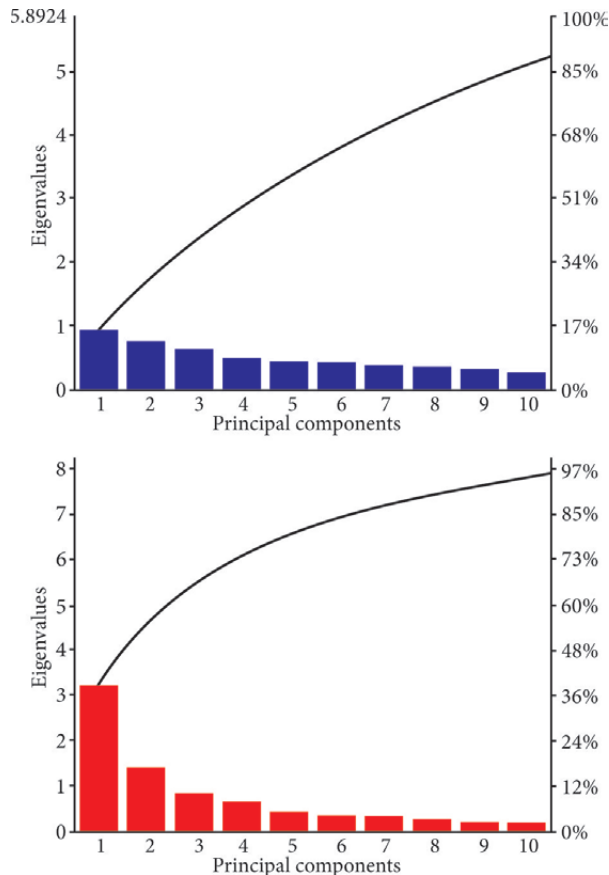


Fig. 2. Mean values of Group No. 2 (slab bridges) data of first inspection cycle

Table 4. Comparison of different PCA algorithm results

Element group		Overlay	Barriers	Handrails	Slopes	Other	Waterproofing	Deck plate	Edge beam	Substructure
Singular Value Decomposition (SVD)	Mean	2.04	1.85	2.32	2.70	1.48	1.25	1.18	2.23	2.09
	Observations	177.00	63.00	79.00	86.00	265.00	90.00	90.00	89.00	177.00
Alternating Least Square (ALS)	Mean	2.04	1.71	2.20	2.59	1.48	1.45	1.30	2.45	2.08
	Observations	265.00	265.00	265.00	265.00	265.00	265.00	265.00	265.00	265.00
Alternating Least Square (ALS) difference in % from Singular Value Decomposition (SVD)	Mean	0.30	-7.60	-5.20	-4.10	0.00	15.60	9.80	9.70	-0.30
	Observations	49.70	320.60	235.40	208.10	0.00	194.40	194.40	197.80	49.70
Variance		0.46	0.66	0.76	0.46	0.47	0.33	0.03	0.16	0.14

observations in both algorithms, which in other element groups are filled with artificial data calculated with ALS algorithm. Differences in mean values of condition states differ from 7.60% to 15.60%, which is considered to be within the limits of human subjectivity, but still, have the



**Fig. 3.** Scree plot of Group No. 1 using SVD (left) and ALS (right) algorithm

**Table 5.** Conclusion of principal components and retained variation from different algorithms

Group	Inspection cycle	Singular Value Decomposition (SVD)		Alternating Least Square (ALS)	
		Principal Components PCs	VAR retained	Principal Components PCs	VAR retained
1	1	6	63.8	4	73.7
	2	6	66.6	4	77.7
2	1	4	74.3	2	77.6
	2	5	79.2	2	74.5
3	1	5	82.7	3	74.6
	2	4	71.0	3	73.8
4	1	5	79.9	2	96.6
	2	4	77.3	2	97.2
5	1	4	83.0	2	71.9
	2	3	75.7	3	81.7

additional influence on results. Differences in mean values and number of observations are directly uncorrelated, but elements with more missing fields have a more significant difference in mean condition states. On the other hand, when looking the ranking of element groups, the list remained unchanged.

In conclusion, according to Hanley *et al.* (2016) arguments and overall results of the comparison, in asset management, it is correct to use data with original results to prevent using false information of elements, which are missing. Due to incorrect information, interventions must be carried out also on missing elements.

#### 4.2. Number of principal components

The number of useful components was visualized with a scree plot of the eigenvalues (Cattell 1966). When it is necessary to determine, which PCs are essential and which must be discarded from the data set, then it is effective to use scree plot as a tool. As the components become less influential the slope of the scree plot begins to flatten because they have retained less variance than the previous components (Hanley *et al.* 2016). From the example of beam bridges shown in Fig. 3, the PC, at which the plot begins to flatten out, occurs for SVD at the fourth PC and ALS at the third PC. There is also a difference in the retained percentage of the variation in the data, being respectively 48% and 66% for different methods with substantial PCs. For Singular Value Decomposition, the retaining variation percentage is still under half of all data variation. Without artificial data, it is necessary to use results of six PCs to have a similar amount of variation described. In both cases, the inclusion of these PCs would not violate the practice of retaining eigenvalues higher than average value.

The comparison of different typologies, shown in Table 5, is made with the assumption where the variation of deterioration model is described by PCs, which have higher eigenvalues than average. Although in all cases it is possible to get minimal results with fewer PCs using ALS, some patterns are stressed out. Group No. 1 has 15 element groups represented, so the lack of principal components is understandable. In Group No. 5, there are only nine element groups represented and a subset of missing data is lower, which gives similar results from both algorithms. Differences and similarities in variables are described with a specialty of ALS algorithm, where missing values are filled with row-mean values of the original matrix, and as a result, the variation of less presented components are lower.

If ALS algorithm makes the whole database artificial, then it is necessary to make some changes in data collection during visual inspection of different bridges by levelling the amount of data in different groups. One possibility is to spread these groups based on the amount of data – one with more elements and other with less. Since ALS algorithm retains most of the variation in false data, further results are based only on SVD algorithm.

The coefficients of Principal Component indicate the relationship between the bridge elements and principal component (Hanley *et al.* 2016). As results vary from

positive to negative, the positive values indicate advanced damage for the element groups in the bridge, and a negative values indicates these element groups, which are in favourable conditions.

For first principal component the largest coefficients of PC 1 in every typology were for the non-structural elements indicating most advanced damage for types in handrails, for old slab typology in barriers and for frame typology in other elements. Situation is explained by the circumstance where most of bridges have old deteriorated handrails or barriers. As most of element groups with the largest coefficients in PC 1 are indicating elements with shorter life cycle, there is also more variation in visual inspection results. Unfortunately, these elements are non-structural and it are considered as irrelevant elements of a bridge, because they incorporate minimal risk to structural load carrying capability. Since the retained variation, percentage is relatively high in following components then it is necessary to include results from them.

For second principal component in most cases the largest coefficients were also for non-structural elements indicating most correlation to previous elements in slopes, deformation joints and waterproofing.

Since different typologies retain variation differently and do not have a correlation in respective principal components, then only example of one bridge typology is presented (Figs 4–5) and further discussion of possible weighing factors for every typology is presented in subchapter number 4.3.

Example of principal components is based on slab bridges Group No. 2, which represents the most common typology constructed before 1960's since the lengths needed to span were relatively low, and it was easier to cast in slabs than beams. The results are influenced by non-structural elements in every principal component. The background of the results is explained by adequate building quality of structural components, and naturally, non-structural elements deteriorate faster. The most variance is retained in safety barriers. The first structural component is waterproofing. Results of the same element group have different signs, for example, overlay, results in the first cycle show deterioration, but in second cycle it shows better condition. The reason is described with the situation, where on the first inspection, the attention was on other structural elements, and this element group was inspected superficially. The same pattern is applicable for other element groups.

### 4.3. Weight factors of element groups

According to Eq (3) and results of PCA, it is possible to obtain weighing factors for every specified bridge typology. Results were calculated with results combined with two inspection cycles, where squares of PCs were taken before averaging the results. In this way, different signs of coefficients were eliminated, and element groups with higher variation remained as principal elements.

Overall results are presented in Table 6. Results are similar for most of the typologies, where weight factors are higher in non-structural element groups, and only some basic element groups influence overall results.

Weighting factors are also representing the overall results of PC 1, which shows where the most variance of visual inspection data is retained in safety barriers,

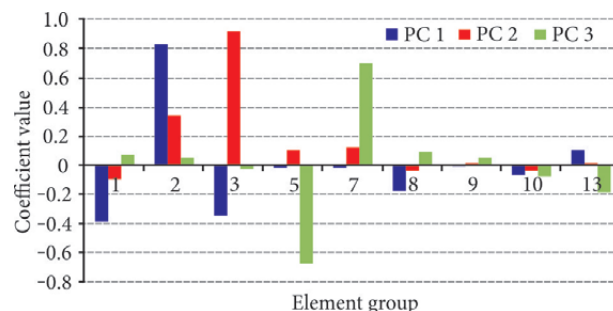


Fig. 4. Principal Component coefficients of first inspections of slab bridge typology

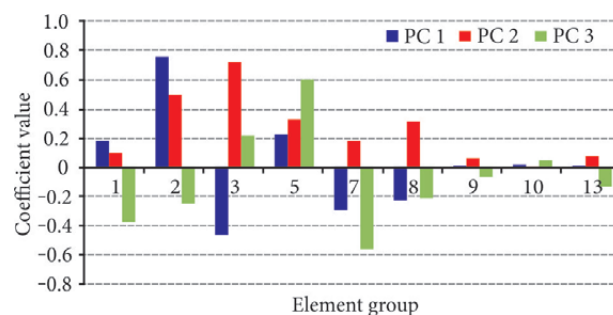


Fig. 5. Principal Component coefficients of second inspections of slab bridge typology

Table 6. Weighting factors λ of different bridge typologies

Element group	Typology Group No.				
	1	2	3	4	5
1	0.64%	8.84%	1.92%	2.92%	1.79%
2	2.08%	62.27%	7.70%	9.09%	5.09%
3	77.93%	16.95%	72.85%	56.68%	NA*
4	NA*	NA*	NA*	NA*	NA*
5	4.16%	2.51%	1.36%	8.10%	2.27%
6	4.70%	NA*	NA*	1.25%	NA*
7	0.34%	4.51%	6.26%	3.99%	86.08%
8	0.10%	4.15%	9.25%	2.07%	0.33%
9	0.03%	0.00%	0.27%	0.13%	0.01%
10	0.17%	0.20%	0.09%	0.36%	0.65%
11	0.10%	NA*	0.20%	0.29%	0.38%
12	0.24%	NA*	0.03%	NA*	NA*
13	0.31%	0.57%	0.06%	3.20%	3.39%
14	0.05%	NA*	NA*	0.04%	NA*
15	1.93%	NA*	NA*	5.52%	NA*
16	7.21%	NA*	NA*	6.34%	NA*

Note: \*data not available

handrails or other elements. Three different reasons describe the background of these results. Firstly, there have been two different inspectors, each carrying out inspections in one year. Both inspectors were inexperienced at the beginning and gained knowledge during inspections. Secondly, these elements are widely damaged by traffic and snow ploughing. Also, the usage of de-icing salts makes steel elements deteriorate faster. Thirdly, the barriers and handrails are widely missing on older bridges, and due to safety restrictions, there have been recorded conditional state four of barriers instead of listing it as missing element.

Another critical finding in comparison of element groups among typologies there are differences in weight factors. It is incorrect to use same weighing factors for all the bridges in one network and even when overall typology is same, like Groups No. 2 and No. 3, the results are different.

The findings of this work should not be used directly in predictive models because the results are not taking into account the risk of failure or impact to load carrying capacity. One opportunity is to combine the results with initial weights or expert judgement using weighted PCA, but this topic needs additional investigation. On the other hand, the results endorse using PCA as a useful data reduction tool, because putting same weight factors to all structural components influences decision-making process.

## 5. Conclusions

There are three meaningful conclusions to be drawn from the statistical analysis of visual inspection data of reinforced concrete bridges in Estonian national roads.

1. In a comparison of different Principal Component Analysis algorithms, using Singular Value Decomposition algorithm is suggested because it uses the original data. Alternating Least Square algorithm results are useable with less principal components and with more specified coefficients. The main variation of Alternating Least Square results is hidden under missing elements, and condition states of element groups differ up to 15.6% from the original state and Singular Value Decomposition results. The difference in mean condition states of compared algorithms is increasing when the amount of missing information is higher.

2. Regarding different coefficients in first principal components, the main variation is retained in non-structural elements. According to this investigation, it is inappropriate to use Principal Component Analysis results directly in predictive models and for decision making it is important to consider additional circumstances as the risk of failure and influence to overall load capacity.

3. Reinforced concrete bridges in Estonia have similarities in principal components, but the importance of same elements are unequal. To be clear then although the main variance in Principal Components is retained in the same element, then according to results, different typologies retain variance in different elements, and due to fundamental distinction, it is incorrect to make decisions based on only construction material or typology.

A general conclusion is that Principal Component Analysis is suitable to be used as a statistical tool for data reduction and additional analysis of visual inspection data to filter out most significant components, but it is necessary to use additional weighing factors to emphasize the real influence of structural elements.

For future research, it is essential to cluster bridges based on a similar number of variables and add circumstances as additional weighting factors. It helps to provide relevant information without prioritizing common elements.

## Acknowledgments

The authors acknowledge the supporting of TU1406 – Quality Specifications for Roadway Bridges, standardization at a European level (BridgeSpec), a COST Action supported by EU Framework Programme Horizon 2020.

## References

- Abdi, H.; Williams, L. J. 2010. *Principal Component Analysis, Wiley Interdisciplinary Reviews: Computational Statistics* 2(4): 433–459. <https://doi.org/10.1002/wics.101>
- Andrić, I. 2015. *Signs by the Roadside*. Sezam book. 472 p.
- Cattell, R. B. 1966. The Scree Test for the Number of Factors, *Multivariate Behavioural Research* 1(2): 245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)
- Chambers, J. M. 1977. *Computational Methods for Data Analysis*, A Wiley Publication in Applied Statistics. New York: Wiley. 268 p.
- Das, P. C. 1998. New Developments in Bridge Management Methodology, *Structural Engineering International* 8(4): 299–302. <https://doi.org/10.2749/101686698780488956>
- Estes, A. C.; Frangopol, D. M. 2005. *Life-Cycle Evaluation and Condition Assessment of Structures*, Structural Engineering Handbook, 2<sup>nd</sup> edition, chapter 36, 1–36.
- Hanley, C.; Kelliher, D.; Pakrashi, V. 2015. Principal Component Analysis for Condition Monitoring of a Network of Bridge Structures, *Journal of Physics: Conference Series* 628:1–8. <https://doi.org/10.1088/1742-6596/628/1/012060>
- Hanley, C.; Matos, J. C.; Kelliher, D.; Pakrashi, V. 2016. Integrating Multivariate Techniques in Bridge Management Systems for Life-Cycle Prediction, in *Civil Engineering Research in Ireland 2016*. Ed. by Nanukuttan, S.; Goggins, J., 29–30 August 2016, Galway, Ireland, 237–242.
- Hotelling, H. 1933. Analysis of a Complex of Statistical Variables into Principal Components, *Journal of Educational Psychology* 24(6): 417–441. <https://doi.org/10.1037/h0071325>
- Ilin, A.; Raiko, T. 2010. Practical Approaches to Principal Component Analysis in the Presence of Missing Values, *Journal of Machine Learning Research* 11: 1957–2000.
- Jackson, J. E. 2003. *A User's Guide to Principal Components*. Wiley-Interscience. 592 p.
- Jolliffe, I. T. 2002. *Principal Component Analysis*. 2<sup>nd</sup> edition. Springer. 518 p.
- Jolliffe, I. T.; Cadima, J. 2016. Principal Component Analysis: a Review and Recent Developments, *Philosophical Transactions* 374(2065): 1–16. <https://doi.org/10.1098/rsta.2015.0202>



- Lauridsen, J.; Bjerrum, J.; Andersen, N. H.; Lassen, B. 1998. Creating a Bridge Management System, *Structural Engineering International* 8(3): 216–220. <https://doi.org/10.2749/101686698780489117>
- Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A. H. 2011. *Big Data: the Next Frontier for Innovation, Competition, and Productivity*. McKinsey & Company. 156 p. Available from the Internet: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
- Martinez, W. L.; Martinez, A. R.; Solka, J. 2010. *Exploratory Data Analysis with MATLAB*. 2<sup>nd</sup> edition. Chapman & Hall/CRC. 508 p. <https://doi.org/10.1201/b10434>
- Matos, J. C.; Casas Rius, J. R.; Chatzi, E. N.; Høj, N. P.; Strauss, A.; Stipanovic, I.; Hajdin, R. 2015. *COST Action TU 1406*. eBook of the 1st Workshop Meeting, Geneva, 21–22 September 2015. ETH-Zürich. <https://doi.org/10.3929/ethz-a-010578903>
- Mueller, J.; Stewart, M. G. 2011. *Terror, Security, and Money: Balancing the Risks, Benefits, and Costs of Homeland Security*. Oxford University Press. 280 p. <https://doi.org/10.1093/acprof:osobl/9780199795758.001.0001>
- Pearson, K. 1901. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2(11): 559–572. <https://doi.org/10.1080/14786440109462720>
- Qian, G. Q.; Gabor, G.; Gupta, R. P. 1994. Principal Components Selection by the Criterion of the Minimum Mean Difference of Complexity, *Journal of Multivariate Analysis* 49(1): 55–75. <https://doi.org/10.1006/jmva.1994.1013>
- Ringnér, M. 2008. What is Principal Component Analysis?, *Nature Biotechnology* 26(3): 303–304. <https://doi.org/10.1038/nbt0308-303>
- Roweis, S. T. 1998. EM Algorithms for PCA and SPCA, *Advances in Neural Information Processing Systems*: 626–632.
- Rücker, W.; Hille, F.; Rohrmann, R. 2006. *Guideline for the Assessment of Existing Structures*. SAMCO Final Report. Federal Institute of Materials Research and Testing (BAM), Division VII.2 Buildings and Structures. 48 p.
- Scherer, W. T.; Glagola, D. M. 1994. Markovian Models for Bridge Maintenance Management, *Journal of Transportation Engineering* 120(1): 37–51. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1994\)120:1\(37\)](https://doi.org/10.1061/(ASCE)0733-947X(1994)120:1(37))
- Sánchez-Silva, M.; Klutke, G. A. 2016. *Reliability and Life-Cycle Analysis of Deteriorating Systems*. Springer International Publishing. 355 p. <https://doi.org/10.1007/978-3-319-20946-3>
- Thompson, P. D.; Small, E. P.; Johnson, M.; Marshall, A. R. 1998. The Pontis Bridge Management System, *Structural Engineering International* 8(4): 303–308. <https://doi.org/10.2749/101686698780488758>

Received 27 January 2017; accepted 25 May 2017