Principal Component Analysis for Condition Monitoring of a Network of Bridge Structures

Ciarán Hanley¹, Denis Kelliher², & Vikram Pakrashi¹

¹Dynamical Systems & Risk Laboratory (DSRL), School of Engineering, University College Cork, Ireland
²Research Unit for Structures & Optimisation (RUSO), School of Engineering, University College Cork, Ireland
E-mail: c.m.hanley@uemail.ucc.ie, d.kelliher@ucc.ie, v.pakrashi@ucc.ie

Abstract. The use of visual inspections as the primary data gathering tool for modern bridge management systems is widespread, and thus leads to the collection and storage of large amounts of data points. Consequently, there exists an opportunity to use multivariate techniques to analyse large scale data sets as a descriptive and predictive tool. One such technique for analysing large data sets is principal component analysis (PCA), which can reduce the dimensionality of a data set into its most important components, while retaining as much variation as possible. An example is applied to a network of bridges in order to demonstrate the utility of the technique as applied to bridge management systems.

1. Introduction

Bridge maintenance and management systems (BMS) are usually populated by bridge inspectors and the indicators of the condition of different components of the bridge may be used to comment on the condition of the entire network, to a certain extent. Often, the interpretation of the condition of the network, based on summary statistics using the condition ratings become the guiding factor behind possible intervention or investment of individual bridges, including prioritization of investment on one bridge as compared to another. While there exist studies on the choice of appropriate indicators and related modelling techniques, studies over actual networks are still scarce.

This paper presents a principal component analysis (PCA) of the ratings in a regional road network of bridge structures in Ireland. The network consists of ninety-four bridges and the locations overlap urban and rural zones, with varied exposure conditions. The system of ratings follows the same definitions as that of the national road network [1]. This statistical analysis method, first formalised by Pearson [2] and developed by Hotelling [3], identifies the component ratings that are more important than the others in terms of explained variability and reassesses these relative importance ratings of different factors from an engineering point of view, based on additional data. The method is widely used in a variety of different science disciplines [4], but its use in infrastructure management applications has not been explored. This paper attempts to address this issue by investigating the suitability of the approach on a BMS and its limitations in relation to extracting network-wide information.

The study creates a helpful benchmark in the regional road network of Ireland for a larger and more comprehensive study involving the entire national road network. The benchmarking
also allows for comparison and contrast with regional road networks of other European Union (EU) countries and countries with an established bridge maintenance management system [5]. The results can be used for the specific network for its evolution with time at a later stage to identify the change of relative importance of different parameters that define the performance of individual bridges based on inspection. The results can also be useful for comparing relative importance of structures, networks of structures or components when different indicators are proposed for such purpose.

2. Bridge Management Systems

Most modern bridge management systems (BMS) are based on the condition assessment of the network through visual inspection as a first step with which more detailed assessment can be planned for [6]. Each individual bridge element is visually assessed and assigned a condition rating based on a prescribed scale, of which is varied across different BMS. A popular method of processing the results of these visual inspections is the Markov chain method; used to predict the future condition states of elements if intervention actions are carried out or not. Such a method is used in the United States (US) systems PONTIS [7] and BRIDGIT [8]. Other times, visual inspections are merely used as a means of identifying areas for further assessment [9]; either through destructive/non-destructive testing or structural assessment, such as deterministic and probabilistic analysis [10].

In the BMS used for this analysis, visual inspection results are recorded for 13 elements in the bridge structure, with an additional result recorded for the condition of the overall structure in general (Table 1). This result is typically a summation of the bridge inspector’s impression of the bridge, and is not based on any linear combination of the other 13 results. Thus, this summary observation has been excluded from the input data, and the analysis will be conducted on the 13 elements only.

| Variable | Description                        | Variable | Description                             |
|----------|------------------------------------|----------|-----------------------------------------|
| 1        | Bridge surface                     | 8        | Piers                                   |
| 2        | Expansion joints                   | 9        | Bearings                                |
| 3        | Footways/median                    | 10       | Deck/slab                               |
| 4        | Parapets/guardrails                | 11       | Beams/girders/transverse beams          |
| 5        | Embankments/slopes                 | 12       | Riverbed                                |
| 6        | Wing walls/retaining walls         | 13       | Other elements                          |
| 7        | Abutments                          | 14       | Structure in general                    |

The condition ratings are evaluated and assigned for each variable present or accessible in the bridge structure, and is assigned on a scale of 0-5 in increasing order of level of deterioration/damage. The condition ratings are defined to be a certain state and the inspector assigns these ratings on an observational basis (Table 2).

3. Principal Component Analysis

Principal component analysis (PCA) is a multivariate analysis technique [12], the primary purpose of which is to reduce the dimensionality of a set of data [13]. A background to the theory is presented here for the sake of completeness, and further information on the method can be found in the referenced texts.

The desired outcome of the analysis is to redefine the input variables as principal components (PC), being a linear combination of the original variables, but having a magnitude less than the original data set, but while preserving most of the information. This is accomplished by
Table 2. Condition rating descriptions [11]

| Rating | Description |
|--------|-------------|
| 0      | No or insignificant damage. |
| 1      | Minor damage but no need of repair. |
| 2      | Some damage, repair needed when convenient. Component is still functioning as originally designed. Observe the condition development. |
| 3      | Significant damage, repair needed very soon. i.e. within next financial year |
| 4      | Damage is critical and it is necessary to execute repair works at once, or to carry out a detailed inspection to determine whether any rehabilitation works are required. |
| 5      | Ultimate damage. The component has failed or is in danger of total failure, possibly affecting the safety of traffic. It is necessary to implement emergency temporary repair work immediately or rehabilitation work without delay after the introduction of load limitation measures. |

highlighting the variables that demonstrate the most variance in the data set. The first principal component \( Y_1 \) is defined as:

\[
Y_1 = \alpha_1'x = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j \tag{1}
\]

Where \( \alpha_1'x \) is a linear function of the elements \( x \) having maximum variance, and \( \alpha \) is a vector of \( p \) coefficients \( \alpha \). The sum of the square of the coefficients \( \alpha_i \) is equal to unity, and is a better indicator of the influence the coefficient has than the raw value:

\[
\sum_{i=1}^{p} \alpha_i^2 = \alpha'\alpha = 1 \tag{2}
\]

The first principal component is the direction along which the data set shows the largest variation [14], and the second component is determined under the constraint of being orthogonal to the first component and to have the largest variance [15]. 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. It is, however, desired that most of the variance in the data set is accounted for in the PCs \( \ll p \). In order to locate the principal components, it is necessary to determine the covariance matrix \( \Sigma \) of the vector of random variables \( x \). It can then be shown that \( \alpha_k \) is an eigenvector of \( \Sigma \) corresponding to its \( k \)th largest eigenvalue \( \lambda_k \) [13].

The above can be discussed in matrix terms where a PCA can be conducted through an eigenvalue decomposition (EVD) or a more robust and generalized singular value decomposition (SVD) [16]. For a data matrix \( X \) of \( n \) observations on \( p \) variables measured about their means:

\[
X = U兰A'
\tag{3}
\]

Where \( L \) is an \( (r \times r) \) diagonal matrix, and \( U \) and \( A \) are \( (n \times r) \) and \( (p \times r) \) matrices, respectively, with orthonormal columns, and \( r \) is the rank of \( X \). It has been observed that SVD approach to PCA is a computationally efficient and generalised method to determining the PCs.

It has been suggested that PCA should only be conducted on continuous variables that conform to a Gaussian distribution [17], and that its application to discrete data, such as condition ratings of a BMS, is inappropriate. However, so long as inferential techniques that require the assumption of multivariate normality are not invoked, there is no necessity for the variables in the data set to have any associated probability distribution [4].

The practical applications of PCA are many, and include [18]:
• the examination of correlations between variables;
• the reduction of the basic dimensions of the variability in the measured set to the smallest number of meaningful dimensions;
• the elimination of variables which contribute relatively little extra information;
• the examination of the grouping of individuals in n-dimensional space;
• determination of the objective weighting of measured variables in the construction of meaningful indices;
• the allocation of individuals to previously demarcated groups;
• the recognition of misidentified individuals;
• orthogonalization of regression calculations.

It is often considered wise to use the correlation matrix for a PCA, as the standardized variates are dimensionless and can be more readily compared [13]. However, when the variables are measured in the same units and have a low variance, using the covariance matrix is sometimes appropriate, and can be beneficial when statistical inference is important. In this case, when the condition ratings are already dimensionless, it is not entirely necessary to standardise the variables. A biplot [19] is most often used to visualise the output of a PCA, as it can handle a matrix of a higher rank than two by approximating it as a matrix of rank two. The biplot displays the orthogonal component coefficients for each variable and the principal component scores for each observation. An aspect of the biplot is the plotting of variable vectors, the direction and length of each indicate how each variable contributes to the two principal components in the plot. Further discussion of the method can be seen in depth with the additional references [20, 21, 22, 23, 24, 25].

4. Results
A network of 94 bridges was assessed overlapping rural and urban areas. These bridges were predominantly observed to be masonry arch bridges (rural) and reinforced concrete (RC) slab bridges (urban), and so the results are presented for these two bridge types, as the sample size for the remaining bridges in the network was insufficient. As the condition ratings for each of the variables are dimensionless and assessed under the same ratings scale, the PCA was conducted using the correlation matrix for the data.

The distribution of the condition ratings for the individual elements for both bridge types is shown in a box plot (Figure 1). In this plot, the edges of the box represent the 25th and 75th percentiles of the data, with the central marker showing the median. The whiskers extend from the boxes to the extreme values of the data, and outliers are plotted individually. It can be seen that the condition ratings for the masonry arch bridges tend to deviate between 0 and 2, with extreme values reaching a rating of 3 (Figure 1a), whereas the RC bridges are much more bunched around 0-1 with an extreme rating of 2, with the exception of a few outliers (Figure 1b). It can be seen that the widest distribution of condition ratings for a particular element are for the parapets/guardrails (4). This is not unexpected as these are typically the elements that are most likely to suffer vehicular impact, or made from lower grade materials than used for the structural elements.

The number of important components can be visualised with a scree plot of the eigenvalues [26]. This can be a useful tool for determining what components are important and what components can be discarded from the data set. As the slope of the scree plot begins to flatten, the components become less important as they have retain less variance then the previous components. Another method of determining which components to retain is by eliminating any components with an eigenvalue less than the average. For this analysis, it can be seen that there is a progressive decrease in the influence of the higher-order eigenvalues λ, and it has been
determined that the first three components retain most of the variance for both the masonry arch and RC bridges; accounting for 75.19% and 70.8% of the data, respectively (Figure 2). The inclusion of a fourth PC would have retained 86.1% and 83.4% of the variance, however its inclusion, in both cases, would’ve violated the practice of only retaining eigenvalues greater than the average.

The six bridge variables/elements for the masonry and RC bridges are represented by vectors on a bi-plot of the principal component space (Figure 3). The bi-plot displays vectors of the correlation coefficients $\alpha$ (Table 3) between the bridge elements and the PCs, as well as a scatter plot representing the level of correlation between two PCs for each bridge structure. It can be seen that the six relevant bridge variables are all positively correlated for the first PC, with the exception being the road surface (1) for masonry bridges being negatively correlated for this PC (Figure 3). The correlation coefficients with the highest absolute values indicate how much the PC represents certain variables, and the sign describes the relationship between the PC to the actual state. For the masonry bridges in the network, it can be seen that first PC is most positively correlated with the state of the abutments (7), the deck/slab (10), and the parapets/guardrails (4), and negatively correlated with the bridge surface (1) (Figure 3a). It should be noted that because a higher condition rating corresponds to a deteriorated/failed...
state, the positive PC coefficients represent a negative state, and thus a negative coefficient corresponds to a more favourable state. Thus, it can be said that an individual masonry bridge structure in the network with a high score for the first PC is one that has a bridge surface in a favourable state and whose abutments, deck/slab, and parapets/guardrails are in a deteriorated state. Similarly, it can be seen that the second PC is most highly correlated with positive values for coefficients relating to wing walls/retaining walls (6), the bridge surface, and the riverbed (12). Thus, bridges with a high score for the second PC can be described as having these three variables in a poor state. It should be noted that the bridge surface is described in contradictory terms for the first two PCs, and so it expectantly observed that there is little correlation between the first two PC scores for masonry bridges.

**Table 3.** Principal component coefficients $\alpha$ and squared coefficients $\alpha^2$ for network

|       | Masonry Arch |            |            | RC        |            |            |
|-------|--------------|------------|------------|-----------|------------|------------|
|       | PC 1 $\alpha$ | PC 2 $\alpha$ | PC 3 $\alpha$ | PC 1 $\alpha$ | PC 2 $\alpha$ | PC 3 $\alpha$ |
| x     | $\alpha$     | $\alpha^2$ | $\alpha$     | $\alpha$     | $\alpha^2$ | $\alpha$     |
| 1     | -0.350       | 0.122      | 0.556       | -0.239     | 0.057      | -0.354       |
| 4     | 0.455        | 0.207      | 0.173       | -0.477     | 0.228      | -0.584       |
| 6     | 0.250        | 0.063      | 0.561       | -0.311     | 0.097      | 0.619        |
| 7     | 0.521        | 0.272      | -0.065      | 0.417      | 0.173      | -0.294       |
| 10    | 0.473        | 0.224      | -0.370      | 0.137      | -0.393     | 0.238        |
| 12    | 0.334        | 0.112      | 0.457       | 0.209      | 0.539      | 0.083        |
| $\Sigma$ | 1.000      | 1.000      | 1.000      | 1.000      | 1.000      | 1.000        |

**Figure 3.** Correlation coefficients for PCs 1 & 2 for masonry (a) and RC (b) bridges in network.

For the RC bridges in the network (Figure 3b), the first PC is most highly correlated with the abutments (7), the deck/slab (10), and the wing walls/retaining walls. As these all have positive correlation coefficients for the first PC, this component can be said to describe how deteriorated these elements are in a RC bridge structure, again with a high PC score describing a more deteriorated state than a lower PC score. The second PC is correlated mostly with the bridge surface (1) and to an extent the parapets/guardrails (4). However, the bridge surface is the most dominant variable in this PC and so the second PC can be described as a measure of the state of deterioration of the bridge surface and, to an extent, the parapets/guardrails. Alternatively, this PC can be described as a measure of the “topside” of the RC bridge structures in the network.
However, as has been noted previously, the third PC also retains a notable amount of variance (Figure 2), and so it cannot be expected that the first two PCs satisfactorily define the data alone. It is possible to display the relationship between the coefficients for three PCs in a three-dimensional bi-plot, however it cannot be optimally interpreted. Thus, additional bi-plots for the second and third PCs have been presented (Figure 4). When compared to the plots for the first two PCs (Figure 3), it can be seen that the new plot confirms the description of the second PC, as the most correlated vectors are demonstrating a consistent relationship with the PC coefficients $\alpha$. For the masonry bridges in the network, it can be seen that the third PC is highly correlated (positively) to the deteriorating condition of the riverbed (12) and the abutments (7), and to the favourable condition of the parapets/guardrails (7) and the deck/slab (10). Thus, this PC can be said to describe a discrepancy between the conditions of the substructure and the superstructure of the masonry bridges; where a bridge with a high positive PC score describing a bridge with a good superstructure and a poor substructure, and a “high” negative score describing a bridge with a poor superstructure and good substructure. For the RC bridges, it is evident that the third PC correlates mostly to the unfavourable condition of the riverbed (12), and to the favourable condition of the deck/slab (10), and bridges that exhibit corresponding condition ratings will be possess a high score for this PC.

![Figure 4](image)

**Figure 4.** Correlation coefficients for PCs 2 & 3 for masonry (a) and RC (b) bridges in network.

The descriptions of the PCs were compared against the original condition rating data, and the masonry arch and RC bridges with upper and lower bound PC scores are presented below (Table 4). It is evident that there is a clear difference between the elements important to each PC for the upper and lower limits of the PC score. This indicates that the use of the PCA method is a viable tool for the assessment of a BMS based on condition ratings.

### 5. Conclusion
A principal component analysis was conducted on a BMS for a network of bridges in an urban and rural environment. The principal components obtained allowed for the bunching of bridge types based on their state of disrepair, and this indicates that the use of PCA can be a viable tool in the assessment of large data sets relating to engineering applications. An examination of a larger data-set over a bigger network would afford the opportunity to additionally assess meta-data for the bridges in a PCA; such as age, exposure, time since last inspection/repair, etc. This, coupled with the relative ease with which it can be applied to inspection-based systems, show its potential utility for effective application to an existing BMS.
Table 4. Condition ratings for upper and lower limit PC scores

| PC Limit | Masonry | RC |
|----------|---------|----|
|          | 1 4 6 7 | 10 12 |
| Upper    | 0 5 1 2 | 2 1    |
| Lower    | 2 0 0 0 | 0 0    |
| Upper    | 2 2 1 1 | 1 5 2 |
| Lower    | 0 0 1 2 | 0 1    |
| Upper    | 0 1 2 1 | 3 1    |
| Lower    | 1 5 1 2 | 0 1    |

Acknowledgements

The authors would like to gratefully acknowledge the financial support from the John Sisk Postgraduate Research Scholarship in Civil Engineering.

References

[1] Duffy L 2004 Proceedings of the ICE - Bridge Engineering 157 139–146 ISSN 1478-4637
[2] Pearson K 1901 Philosophical Magazine Series 6 2 559–572 ISSN 1941-5982
[3] Hotelling H 1933 Journal of Educational Psychology 24 417–441
[4] Jackson J E 2003 A User’s Guide to Principal Components (Wiley)
[5] Znidarić A, Pakrashi V, O’Brien E J and O’Connor A 2011 A review of road structure data in six European countries
[6] Das P C 1998 Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE) 8 299–302 ISSN 10168664
[7] Thompson P D, Small E P, Johnson M and Marshall A R 1998 Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE) 8 303–308 ISSN 10168664
[8] Hawk H and Small E P 1998 Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE) 8 309–314 ISSN 10168664
[9] Schellenberg K, Vogel T, Chèvre M and Alvarez M 2013 Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE) 23 402–410 ISSN 10168664
[10] Pakrashi V, O’Connor A and Salta M 2012 Assessment and Maintenance Planning of Infrastructure Networks Transportation Research Board (TRB) 91st Annual Meeting (Washington, D.C.)
[11] NRA 2008 EIRSPAN System Manual No. 3, Principal Inspection National Roads Authority, Ireland
[12] Mardia K, Kent J and Bibby J 1979 Multivariate Analysis (Academic Press) ISBN 9780124712522
[13] Jolliffe I 1971 Principal Component Analysis (Springer)
[14] Gabriel K R 1966 The Annals of Mathematical Statistics 34 122–148
[15] Anderson T W 1963 The Annals of Mathematical Statistics 36 55–75
[16] Rao C R 1964 Biometrika 58 325–338 ISSN 00064444
[17] Anderson T W 1963 The Annals of Mathematical Statistics 34 122–148
[18] Wold S, Esbensen K and Geladi P 1987 Chemometrics and Intelligent Laboratory Systems 2 37–52 ISSN 01697439
[19] Cattell R B 1966 Multivariate Behavioral Research 1 245–276 ISSN 0027-3171