Aplicação da Análise de Componentes Principais na monitorização da integridade estrutural

Tiago Coelho¹ Min Xu² Luís O. Santos³

RESUMO

Com o avanço tecnológico, a monitorização estrutural passou a utilizar uma variedade significativa de sensores, resultando num grande volume de dados coletados em tempo real, o que exige um tratamento adequado para a deteção de danos estruturais. A aplicação de métodos robustos de tratamento estatístico é essencial para a identificação de padrões dentro desse conjunto de dados.

No âmbito das atividades de observação de estruturas no Laboratório Nacional de Engenharia Civil (LNEC), foi desenvolvida uma metodologia de análise estatística para o pré-processamento dos dados obtidos da monitorização estrutural. Essa metodologia permite a identificação de sensores com comportamentos anómalos, a eliminação de informações redundantes e o reconhecimento de padrões de comportamento ao longo do tempo.

Nesta comunicação apresenta-se uma aplicação da metodologia desenvolvida a múltiplas séries temporais de dados de diversas grandezas físicas, provenientes de sensores integrados no sistema de monitorização estrutural da Ponte 25 de Abril. O processo inclui a utilização de técnicas de análise estatística descritiva, abrangendo tanto análises estatísticas univariadas e multivariadas quanto a Análise de Componentes Principais (ACP).

Palavras-chave: Pontes; Monitorização estrutural; Análise de componentes principais

1. INTRODUÇÃO

A Análise de Componentes Principais (ACP) é uma técnica estatística usada para reduzir a dimensionalidade de um conjunto de dados, sem perda de informação, através da transformação de variáveis possivelmente correlacionadas num conjunto menor de variáveis linearmente não correlacionadas, denominadas de Componentes Principais [1]. Essa transformação é realizada de modo que a primeira componente principal tenha a maior variância possível, enquanto cada componente subsequente é projetada para maximizar a variância restante, sob a restrição de ser ortogonal às componentes anteriores. ACP é amplamente utilizada para simplificar a análise de dados, melhorar a visualização, e identificar padrões e estruturas subjacentes nos dados.

¹ Laboratório Nacional de Engenharia Civil, Lisboa, Portugal, tcoelho@lnec.pt

² Laboratório Nacional de Engenharia Civil, Lisboa, Portugal, xumin@lnec.pt

³ Laboratório Nacional de Engenharia Civil, Lisboa, Portugal, luisosantos@lnec.pt

Nesta comunicação, apresenta-se um estudo realizado com os dados obtidos através do sistema de monitorização estrutural da Ponte 25 de Abril⁴, aplicando as técnicas de análise estatística descritiva, com ênfase na Análise de Componentes Principais.

Foram analisados os dados de 52 sensores, obtidos durante um período de 12 meses. O principal objetivo é identificar padrões nos dados para melhorar o desempenho do processamento de dados e, em última instância, permitir a deteção atempada de alterações no comportamento estrutural que possam comprometer a segurança.

2. DESCRIÇÃO DA ESTRUTURA E DO SISTEMA DE MONITORIZAÇÃO ESTRUTURAL

A Ponte 25 de Abril é uma ponte suspensa que faz a ligação das margens do Rio Tejo entre Lisboa e Almada. Composta por seis tramos, a ponte tem um comprimento total de 2277,5 m, tendo os tramos suspensos laterais e central vãos de 483,4 m e de 1012,9 m, respetivamente (Figura 1) [2].

O tabuleiro da ponte, constituído por uma viga de rigidez metálica e suspenso por quatro cabos principais, suporta o tráfego rodoviário ao nível da corda superior e o tráfego ferroviário ao nível da corda inferior. A secção transversal da viga treliçada tem 12,0 m de altura e 21,5 m de largura. Os pilares P1, P2, P5 e P6 têm uma altura de aproximadamente 60,0 m. As torres P3 e P4 cerca de 180,0 m de altura e suportam o tabuleiro a uma altura de 60,0 m.

A Ponte 25 de Abril encontra-se munida de um sistema de monitorização estrutural constituído por conjunto de sensores dispostos ao longo do tabuleiro e na torre P3 [2]. Este sistema de monitorização permite registar um total de 210 grandezas físicas, medindo as condições ambientais e operacionais, nomeadamente, velocidade do vento, temperaturas e o tráfego ferroviário, e as respostas estruturais: acelerações, extensões, rotações e deslocamentos longitudinais. O registo das leituras é efetuado através 11 unidades de aquisição, sincronizados entre si, com uma taxa de aquisição de 500 Hz.



3. DADOS EM ANÁLISE

O conjunto de dados em análise é composto por registos horários obtidos durante um período de 12 meses, entre 2017-10-01 e 2018-09-30, correspondendo a 52 variáveis (Figura 2) e 8581 registos horários, sendo que cada registo corresponde ao valor mediano num período de uma hora.

Como ações, foram incluídas a velocidade do vento (W) nas três direções ortogonais, registadas na secção de meio-vão ("0"), temperaturas da estrutura (T) medidas nas secções do tabuleiro (0 e 66N) e da torre P3, a uma cota de 90,0 m (P3C). As respostas da estrutura são representadas por extensões (e) medidas nas secções 0, 66N e P3C, rotações transversais (ct) e longitudinais (cl) nos topos da torre P3 e do pilar P5, e deslocamentos longitudinais (d) na junta do P7. A localização dos sensores nas secções está apresentada na Figura 3.

⁴ Obra gerida pela Infraestruturas de Portugal, S.A.



Figura 3. Disposição dos sensores no tabuleiro e pilares

4. ANÁLISE DE COMPONENTE PRINCIPAIS

4.1 Metodologia aplicada

A aplicação do método de Análise de Componentes Principais (ACP) aos dados obtidos do SHM tem como objetivo de identificar padrões dos dados que possam detetar alterações provocadas por danos da estrutura.

A ACP é uma técnica estatística que reduz a dimensionalidade dos dados, convertendo variáveis originais num conjunto menor de variáveis não correlacionadas, chamadas componentes principais, que explicam a maior parte da variância dos dados originais [1]. Isso permite simplificar a análise sem perda significativa de informação.

Inicialmente, é realizada uma análise bivariada para uma eventual remoção de variáveis redundantes. De seguida é feita a remoção dos efeitos das ações ambientais nas respostas estruturais através regressão linear múltipla (RLM), de acordo com a seguinte expressão:

$$Z_{n,p} = \beta_{n,p}^0 + Q_{n,d}\beta_{d,p} + \epsilon_{n,p},\tag{1}$$

onde *n* corresponde ao número de registos, $Z_{n,p}$ representa a matriz das *p* variáveis dependentes ("saída"), $Q_{n,d}$ representa a matriz das *d* variáveis independentes ("entrada"), $\beta_{n,p}^0$ o vetor das ordenadas na origem, $\beta_{d,p}$ a matriz dos coeficientes de regressão e $\epsilon_{n,p}$ os erros residuais que explica a diferença entre os valores observados e os valores previstos pelo modelo.

A ACP será efetuada para o conjunto de dados das ações ambientais e para o conjunto de dados das respostas estruturais. A transformação pode ser expressa como:

$$\mathbf{Y}_{n,q} = \mathbf{X}_{n,p} \mathbf{U}_{p,q},\tag{2}$$

onde *n* corresponde ao número de registos, $X_{n,p}$ representa a matriz de dados original com *p* variáveis, $Y_{n,q}$ representa a matriz de componentes principais com *q* componentes principais e $U_{p,q}$ a matriz dos vetores próprios (pesos) associados às componentes principais.

Como resultado, serão identificadas as componentes principais das ações e das respostas estruturais, permitindo efetuar uma RLM com um número reduzido de variáveis independentes, as componentes principais.

Por último, será efetuada uma segunda análise que passa por avaliar a eficácia da utilização da Análise de Componentes Principais na caracterização e remoção de ações e efeitos ambientais independentes utilizando apenas as variáveis correspondentes às respostas estruturais [3–6]. Esta abordagem tem por base o facto de que, por vezes, a caracterização dos efeitos ambientais não é suficientemente abrangente e eficaz e existe a necessidade de aplicar métodos robustos de análise estatística para os caracterizar.

4.2 Análise bivariada

A análise estatística bivariada permite determinar relações entre diversas grandezas e identificar padrões, calculando coeficientes de correlação entre pares.

Na Figura 4 está presente a matriz de correlações entre todos as variáveis, na qual se pode verificar que as temperaturas têm valores de correlação elevados entre si. Por sua vez, os extensómetros situados na corda inferior da viga do tabuleiro, os deslocamentos longitudinais e algumas rotações apresentam igualmente valores de correlação elevados com as temperaturas. É possível também observar alguns grupos de correlação elevada tais como, os extensómetros da corda inferior do tabuleiro com os deslocamentos longitudinais e rotações longitudinais e também os extensómetros das laterais da torre P3 com a sua rotação transversal.



Figura 4. Matriz de correlações de todas as variáveis

No que respeita às variáveis da velocidade do vento, verifica-se que não existe qualquer correlação com as restantes. De facto, a ação do vento sobre a estrutura é considerada uma ação dinâmica, enquanto, as restantes variáveis representam o comportamento quase-estático da estrutura. Assim, as velocidades do vento serão excluídas das análises seguintes.

As temperaturas medidas nas diferentes secções da estrutura fornecem uma representação do campo térmico ao longo da mesma. Os coeficientes de correlação acima de 0,70 indicam uma distribuição de temperatura praticamente homogénea, sendo que as diferenças entre as variáveis se devem à localização

dos sensores (altura da secção e exposição solar). Verifica-se que a maioria dos elevados valores de correlação, superior a 0,99, correspondem à mesma localização do sensor nas secções 0 e 66N e também na torre P3, do lado montante e jusante. Deste modo, foram removidas 10 variáveis redundantes com coeficiente de correlação superior a 0.99, mais especificamente todas a variáveis da secção 66N e duas da torre P3, secção C.

4.3 Regressão Linear Múltipla

Como se verificou, as variáveis das respostas estruturais apresentam uma correlação significativa com as variáveis relativas à temperatura, significando que o seu comportamento está fortemente influenciado pelo efeito térmico. Deste modo, a utilização de uma regressão linear múltipla permite remover em parte o efeito da temperatura nas respostas estruturais e identificar, com maior clareza, eventuais alterações estruturais devido a outras ações ou a eventuais danos estruturais. Para este efeito, foi utilizado o modelo de acordo com a Eq. (1), considerando as respostas estruturais como variáveis dependentes e as temperaturas como variáveis independentes, designado por modelo RML_T.

A título de exemplo, apresenta-se na Figura 5 a série temporal dos erros residuais, ϵ , resultantes da regressão da variável e1.P3mC (extensómetro na face sul da torre P3). Na mesma figura está igualmente presente a série de dados original a cinzento. É possível observar que grande parte das variações diárias é devida à temperatura, entretanto removida.



Figura 5. Remoção do efeito da temperatura na variável e1.P3mC

No Quadro 1 apresentam-se os valores dos coeficientes de determinação ajustado do RLM [7] das variáveis em análise. Analisando os valores obtidos, verifica-se que, na sua maioria, os modelos de regressão RML T resultaram em bons ajustes, com valores acima de 0,7, permitindo remover efetivamente as variações diárias relacionadas com os efeitos da temperatura.

| Quadro 1. Regressão linear multipla: coeficiente de determinação ajustado | | | | | | | | |
|---|---------|---------|---------|--------|---------|--|--|--|
| Sensor | R_a^2 | Sensor | R_a^2 | Sensor | R_a^2 | | | |
| e1.0 | 0,915 | e1.P3mC | 0,969 | dm.P7 | 0,988 | | | |
| e4.0 | 0,901 | e2.P3mC | 0,985 | dj.P7 | 0,988 | | | |
| e6.0 | 0,764 | e3.P3mC | 0,841 | ct.P3m | 0,513 | | | |
| e7.0 | 0,788 | e4.P3mC | 0,895 | cl.P3m | 0,725 | | | |
| e1.66N | 0,913 | e1.P3jC | 0,947 | ct.P3j | 0,957 | | | |
| e4.66N | 0,912 | e2.P3jC | 0,958 | cl.P3j | 0,891 | | | |
| e6.66N | 0,612 | e3.P3jC | 0,788 | ct.P5m | 0,886 | | | |
| e7.66N | 0,589 | e4.P3jC | 0,962 | cl.P5m | 0,902 | | | |
| | | | | ct.P5j | 0,864 | | | |

4.4 Análise de Componentes Principais das ações impostas à estrutura

Nesta primeira análise de componentes principais, o quadro de dados é composto por 14 variáveis de temperatura, representando as ações impostas à estrutura. A determinação das componentes principais deste conjunto de dados é feita de acordo com a Eq. (2).

Na Figura 6 a) apresentam-se as percentagens de variância de cada componente principal (CP) e variância acumulada. Verifica-se que a primeira componente principal (CP1) consegue explicar quase a totalidade da variância dos dados (89,7%). Por outro lado, a CP2 e a CP3 apresentam coeficientes de correlação superiores a 0,40 em algumas variáveis (Figura 6 b)), indicando que estas componentes poderão ter alguma relevância na representação da ação das temperaturas na estrutura. Como resultado, as três primeiras CP permitem explicar 98% da variância dos dados originais.



Figura 6. a) Variância explicada (ações impostas) e b) matriz correlação entre as CP e as ações impostas

Para avaliar a importância das componentes principais na caracterização da resposta estrutural, aplicou-se o método de regressão linear múltipla às variáveis de resposta, considerando como variáveis independentes as combinações da 1^a à 6^a componentes principais de acordo com a Eq. (1), designado por RLM_CP. Este resultado foi comparado com o do modelo RLM_T, referido no ponto 4.3.

A correlação entre os erros residuais do modelo RLM_CP e os erros residuais do modelo RML_T é apresentada na Figura 7 a). Na Figura 7 b) o gráfico apresentado representa o valor médio do coeficiente de correlação de todas as variáveis de resposta. Verifica-se que é possível constituir um modelo RLM eficiente com apenas as quatro primeiras componentes principais, obtendo um coeficiente de correlação médio de 82%.



Figura 7. a) Matriz de correlação entre os resíduos da RLM com temperaturas e RLM com componentes principais b) Valor médio do coeficiente de correlação

Como exemplo, a Figura 8 compara as séries temporais dos erros residuais das regressões para a variável dm.P7 (deslocamento longitudinal no pilar P7), usando 2 tipo de RLM, RML_T (gráfico superior) e RLM_CP com as quatro componentes principais (CP1 a CP4) (gráfico inferior). O coeficiente de correlação entre os erros residuais destas duas séries temporais é de 0,908, indicando que os modelos de RLM são idênticos.

No Quadro 2 apresentam-se os valores dos coeficientes de correlação entre erros residuais dos dois modelos RLM referidos, aplicados às variáveis da resposta estrutural. Neste quadro observa-se que, a

utilização do modelo de RLM com as 4 primeiras componentes principais revelou-se eficaz na remoção do efeito da temperatura, para coeficientes de correlação acima de 0,8.

Com este resultado, verifica-se que é possível reduzir o quadro de dados de catorze variáveis de temperatura para quatro variáveis de componentes principais, para representar a ação térmica global, tornando a análise do comportamento estrutural mais eficiente.



Figura 8. Comparação entre resíduos da RML_T com RLM_CP, exemplo da variável dm.P7

| Sensor | ρ | Sensor | ρ | Sensor | ρ |
|--------|-------|---------|-------|--------|-------|
| e1.0 | 0,824 | e1.P3mC | 0,729 | dm.P7 | 0,908 |
| e4.0 | 0,780 | e2.P3mC | 0,676 | dj.P7 | 0,903 |
| e6.0 | 0,901 | e3.P3mC | 0,642 | ct.P3m | 0,967 |
| e7.0 | 0,900 | e4.P3mC | 0,718 | cl.P3m | 0,961 |
| e1.66N | 0,870 | e1.P3jC | 0,791 | ct.P3j | 0,748 |
| e4.66N | 0,781 | e2.P3jC | 0,734 | cl.P3j | 0,940 |
| e6.66N | 0,915 | e3.P3jC | 0,632 | ct.P5m | 0,857 |
| e7.66N | 0,961 | e4.P3jC | 0,576 | cl.P5m | 0.880 |
| | | | | ct.P5j | 0,905 |

Quadro 2. Coeficientes de correlação entre erros residuais de RLM_T com RLM_CP

4.5 Análise de Componentes Principais da resposta estrutural

A análise de componentes principais das variáveis associadas às respostas estruturais tem como finalidade identificar e remover os efeitos associados às ações impostas à estrutura, sem recorrer às variáveis relacionadas com as mesmas. Neste sentido esta análise utilizará apenas as variáveis respeitantes às respostas estruturais.

Como resultado desta análise de CP, verifica-se que as três primeiras CP explicaram 80,6% da variância dos dados analisados (Figura 9 a)). A matriz de correlações mostra que a CP1 se correlaciona fortemente com as variáveis relativas às extensões medidas na corda superior da viga (e6 e e7), ao deslocamento longitudinal (dm.P7 e dj.P7), e à maioria das variáveis rotação. Relativamente à CP2, verifica-se que esta tem uma maior correlação com as extensões medidas na corda inferior da viga (e1 e e4) e a rotação transversal ct.P3j. Salienta-se que as extensões medidas na torre estão correlacionadas com CP2 e CP3, mais especificamente, as extensões medidas nas faces jusante e montante tem uma maior correlação com a CP2 e as extensões medidas nas faces sul e norte, com a CP3.

A influência das CP1 e CP2 nas variáveis medidas também pode ser observada na Figura 10, através do círculo de correlações das variáveis no plano CP1-CP2. Do ponto de vista do comportamento da estrutura, pode-se inferir que a CP1 está relacionada com movimentos longitudinais da estrutura e a CP2 com movimentos transversais.

Chama-se atenção que a *ct.P3j* está alinhada com a segunda componente principal, tal como as *e4.P3jC*, *e2.P3mC*, *e4.P3mC* e *e2.P3jC*, localizados nas faces nascente e poente. As correlações altas entre estas variáveis também foram comprovadas na análise bivariada (Figura 4).

A distribuição dos registos horários em função da exposição solar, divididos em três períodos, apresentada no gráfico *Biplot* da Figura 10, permite verificar que a CP1 varia sobretudo com a temperatura média ao longo do período (gráfico esquerdo) e a CP2 varia com a exposição solar na estrutura (gráfico direito). Os valores positivos da CP2 estão correlacionados com os registos do período da tarde e os negativos com os de manhã. Valores com CP2 próximo de zero representam os registos relativos ao período da noite. Uma análise adicional revela que a CP1 apresenta a variação sazonal de temperatura, enquanto a CP2 e CP3 se relacionam a variação diária.



Figura 9. Variância explicada: ACP das respostas estruturais



Figura 10. *Biplot* no plano 1-2 relativo às respostas estruturais, identificação dos registos pela temperatura média (esquerda) e pelo período do dia (direita)

Definido o modelo de componente principais, procedeu-se à reconstituição do quadro de dados com as variáveis originais da resposta estrutural, através do processo inverso da ACP, removendo componentes principais associadas às ações impostas à estrutura. A eficácia do modelo ACP é comparada com o modelo RLM utilizando as temperaturas, através do coeficiente de correlação de *Pearson* entre o quadro de dados reconstituído e os erros residuais obtidos no modelo RLM_T.

Esta análise foi efetuada através de um processo iterativo no qual se foram alterando as componentes principais a remover. Foram testados modelos com combinações de componentes principais da CP1 até CP14, uma vez que a partir da CP15 a correlação com as componentes principais das variáveis de temperatura é baixo (Figura 11). Na Figura 11, é também possível observar valores de correlação elevados entre as primeiras 4 CP das respostas e as CP de temperatura.

O resultado deste processo iterativo encontra-se exposto na Figura 12, na qual se apresentam, por ordem crescente, as dez combinações com um coeficiente de correlação médio mais elevado.



Figura 11. Matriz de correlação entre componentes principais das temperaturas e das respostas estruturais

O modelo no qual se removeram as componentes CP1, CP2 e CP3 é o que apresenta o maior coeficiente de correlação com o modelo RLM_T. No entanto, verifica-se que em algumas variáveis a remoção das três primeiras componentes principais ainda não é suficiente, apresentando um coeficiente de correlação baixo, indicando efeitos de outros fatores nas variações observadas, como se pode observar na Figura 13 (*dm.P7*).



Figura 12. Matriz de correlação entre resíduos da RLM com temperaturas e quadro de dados reconstituído com várias combinações de CPs



Figura 13. Comparação entre resíduos da RLM e o quadro de dados reconstituído para dm.P7

5. CONCLUSÕES

Neste estudo foi realizada uma Análise de Componentes Principais (ACP) para um conjunto de dados da monitorização estrutural da Ponte 25 de Abril em duas vertentes, uma relativa às ações impostas à estrutura e outra relativa às respostas estruturais.

Aplicação da análise de Componentes Principais na monitorização da integridade estrutural

A ACP aplicada às ações teve como principal objetivo a redução da dimensão dos dados através da criação de novas variáveis representativas do comportamento das originais (componentes principais). Este novo conjunto de dados de componentes principais permitiu efetuar uma regressão linear múltipla com apenas quatro variáveis independentes de modo a ajustar o modelo de RLM às respostas estruturais (variáveis dependentes). A ACP das respostas estruturais pretendeu avaliar a eficácia da utilização de um número pequeno das Componentes Principais das respostas estruturais na caracterização e remoção dos efeitos de ações ambientais independentes. Os resultados de ambas as análises foram comparados com os resultados obtidos a partir de uma regressão linear múltipla, diretamente com as variáveis das ações (RLM_T).

Concluiu-se que a ACP das ações permitiu reduzir o quadro de dados de catorze para quatro variáveis de componentes principais e em simultâneo remover de forma eficaz do efeito da temperatura nas respostas estruturais, utilizando o modelo RLM_CP.

A ACP das respostas estruturais mostrou ser possível, na maioria das situações, remover os efeitos associados às ações relacionadas com o efeito da temperatura, sem recorrer a uma RLM, utilizando as temperaturas como variáveis de entrada. Este processo foi efetuado reconstituindo o quadro de dados, removendo as três primeiras componentes principais, sendo este o modelo que apresentou maior coeficiente de correlação com o modelo RLM_T. No entanto, em algumas variáveis a reconstituição dos dados não devolveu bons resultados na remoção dos efeitos das ações, indicando efeitos dos outros fatores nas variações observadas.

A utilização desta técnica permitiu ainda a identificação de sensores com um comportamento anómalo através da análise de padrões e comparação direta com as restantes variáveis.

AGRADECIMENTOS

Os autores agradecem à Infraestruturas de Portugal o apoio prestado através do contrato celebrado com o LNEC para monitorização da Ponte 25 de Abril, que possibilitou o desenvolvimento do presente estudo e a sua publicação.

REFERÊNCIAS

- [1] Tibaduiza, D. A.; Mujica, L. E.; Rodellar, J. (2011). Structural Health Monitoring based on principal component analysis: damage detection, localization and classification, *Advances in Dynamics, Control, Monitoring and Applications, Universitat Politècnica de Catalunya, Departament de Matemàtica Aplicada*, Vol. 3, No. 1, 8–17
- [2] Santos, J.; Oliveira, C.; Silveira, P. (2020). Monitorização estrutural da Ponte 25 de abril, *Construção Magazine*, No. 95, 24–29
- [3] Santos, J. P.; Crémona, C.; Orcesi, A. D.; Silveira, P. (2013). Multivariate statistical analysis for early damage detection, *Engineering Structures*, Vol. 56, 273–285. doi:10.1016/j.engstruct.2013.05.022
- [4] Laory, I.; Trinh, T. N.; Posenato, D.; Smith, I. F. C. (2013). Combined Model-Free Data-Interpretation Methodologies for Damage Detection during Continuous Monitoring of Structures, *Journal of Computing in Civil Engineering*, Vol. 27, No. 6, 657–666. doi:10.1061/(ASCE)CP.1943-5487.0000289
- [5] Zhang, H.; Guo, J.; Xie, X.; Bie, R.; Sun, Y. (2013). Environmental effect removal based structural health monitoring in the internet of things, *Proceedings 7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2013*, 512–517. doi:10.1109/IMIS.2013.91
- [6] Posenato, D.; Kripakaran, P.; Inaudi, D.; Smith, I. F. C. (2010). Methodologies for model-free data interpretation of civil engineering structures, *Computers and Structures*, Vol. 88, Nos. 7–8, 467–482. doi:10.1016/j.compstruc.2010.01.001
- [7] Rousselet, G. A.; Pernet, C. R. (2011). *Robust Statistics Show No Evidence for a Relationship between Fiber Density and Memory Performance, Proceedings of the National Academy of Sciences of the United States of America* (Vol. 108). doi:10.1073/pnas.1109188108