Chapter 6
A Multivariate Specification Development Approach Using Principal Component Analysis

Colin M. Haynes and Sean T. Kelly

Abstract The concept of a conservative, bounding environment for multiple-input, multiple-output (MIMO) vibration testing is inherently multivariate. However, some multivariate statistical methods proposed for MIMO specification development are very computationally expensive. This computational burden has motivated work to reduce the computational complexity of a specification with many control degrees of freedom (or variates). In this paper, a new method based on principal component analysis is presented that decomposes the multivariate specification problem into a univariate one while still accounting for the underlying correlation structure of the data. Numerical studies are used to illustrate the benefits of the new approach.

Keywords Multivariate · Specifications · MIMO · Environmental Testing · Conservatism · Principal Component Analysis

List of Notations and Abbreviations
Mahalanobis distance (MD)
multiple-input, multiple-output (MIMO)
principal component (PC)
principal component analysis (PCA)
probability density function (PDF)

Introduction

To ensure that structures can perform their functions as intended under uncertain loading conditions, it is necessary to characterize the loading they are expected to undergo. These estimates of the service environment can then be used to develop a statistically-conservative bounding environment to test one or more structures to and qualify them for service. This sort of environmental specification development and testing is common for shock and vibration environments.

Historically, such testing has been done on single-axis electrodynamic shakers by applying the bounding environment to one axis at a time. As test equipment and control software have improved, new multi-axis approaches have been considered for more environmental testing applications to produce responses more representative of the true service environment [1, 2].

Traditional univariate approaches to specification development cannot account for the correlations between variates, which in the case of multiple-input, multiple-output (MIMO) testing are assumed to be different channels (locations and directions) used for control of an environmental test. Previous work has demonstrated that these correlations exist and that naive approaches, such as applying univariate one-sided normal tolerance bounds to each variate separately, often result in either: 1) impossibly unlikely specifications with respect to the service environment, or 2) likely specifications which are not sufficiently conservative [3]. For the former, to ignore the correlations is to attempt a test that distorts the way the structure would respond in service.

Previous proposals for multivariate specifications have accounted for the correlation structure of the data, but often the computational cost at higher dimensions (e.g., up to 24 dimensions) is prohibitive [4]. This paper proposes an approach

Colin M. Haynes · Sean T. Kelly
Mail Stop C932, Los Alamos National Laboratory, Los Alamos, NM 87545
e-mail: cmhaynes@lanl.gov; stkelly@lanl.gov

*Colin M. Haynes and Sean T. Kelly contributed equally to this work.

Kristin B. Zimmerman, PhD (eds.), *Dynamic Evironments and Durability Testing, 2026, Vol. 7*, Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.1007/978-8-743-81446-7_6

based on transforming the data using principal component analysis (PCA), which enables much more efficient calculation of the maximum predicted environment by using univariate calculations on individual principal coordinates rather than on variates directly.

Properties of PCA as Applied to Specification Development

PCA Overview

PCA is a common data reduction and decomposition technique based on the singular value decomposition. It can be shown that the left singular vectors of the mean-centered data matrix identified for this linear operation are the orthogonal directions that successively maximize the explained variance [5]. A detailed description of the mathematics behind PCA is out of the scope of this paper. Instead, only the properties particularly useful for the proposed specification development approach will be highlighted.

Assumption of Lognormality

In line with common practice in the field, we assume that environments data (usually power spectral densities and shock response spectra) are lognormally distributed [6]. However, each variate at each frequency line has some correlation to (relationship with) the other variates based on the dynamics (modes) of the structure. Such data can often be easily transformed into data following a multivariate normal distribution by simply taking the logarithm.

One convenient property of PCA is that the resulting principal coordinates are uncorrelated [5]. However, it can also be shown that in the situation of uncorrelated multivariate normal data, the variates are also independent [7]. Thus, applying PCA to log-transformed spectra may reasonably be assumed to produce fully independent coordinates. Furthermore, data scaling is performed such that the resulting principal coordinates all have equal sample standard deviation $\overline{\sigma}_{\mathrm{pca}}$.

The original probability density function (PDF) isocontours of representative bivariate synthetic data with a correlation ρ of -0.6719 at a given frequency line are shown on the left side of Figure 1. Note that the correlation structure between variates causes these isocontours to appear as ellipses in the two-dimensional case and generalize to hyperellipsoids in higher dimensionalities. Once transformed through PCA, because of the properties just mentioned, the PDF isocontours become circles in two dimensions and hyperspheres in higher dimensions. The right side of Figure 1 illustrates the result of this data standardization in the PCA space when the data are projected onto both principal components (PCs).

Note that Figure 1 includes more information than just the PDF isocontours in illustrating the effect of the workflow that is outlined in the following section. Also plotted are the 95/50 (i.e., 95% coverage and 50% confidence) one-sided univariate tolerance limits for each variate independently, expressed as purple lines. As shown in Figure 1 and discussed in the Introduction, the intersection of these lines (purple dot) is significantly unlikely to occur based on its proximity to the data distribution. This intersection is also shown when projected in the PCA space, further demonstrating its distance from the data population when viewed in independent principal coordinates with equal sample standard deviation. PC 1 and PC 2 are also depicted with scaling added for ease of visualization, as well as the point selected by calculating the one-sided univariate tolerance limit for each principal coordinate in the PCA space (this intersection point is denoted $\mathbf{x}_{\mathrm{pca,s}}$), as described in the following section.

Proposed PCA Specification Development Workflow

PCA Factorization

The first step of the proposed derivation approach is to form the data matrix $\mathbf{A}$ (dimension m variates by n samples) and perform PCA factorization. This factorization is given by $\mathbf{A}{=}\mathbf{W}_{\mathrm{pca}}\mathbf{X}_{\mathrm{pca}}$, where $\mathbf{W}_{\mathrm{pca}}$ contains the left singular vectors scaled by their respective singular values as columns (both calculated from the mean-centered data matrix), and the rows of $\mathbf{X}_{\mathrm{pca}}$ are the principal coordinates (accounting for if $\mathbf{A}$ is not mean-centered by default). Note that for this work we consider a PC to refer to a column of $\mathbf{W}_{\mathrm{pca}}$, rather than a left singular vector alone.

Here we assume that n is greater than m—that is, there are more samples of the intended service environment than intended control channels to estimate—as this is the situation that tends to produce reasonable variance estimates when calculating estimates of bounding environments. However, this approach is not limited to this situation and does not inherently impose a constraint on whether $n > m$ or $n \leq m$.

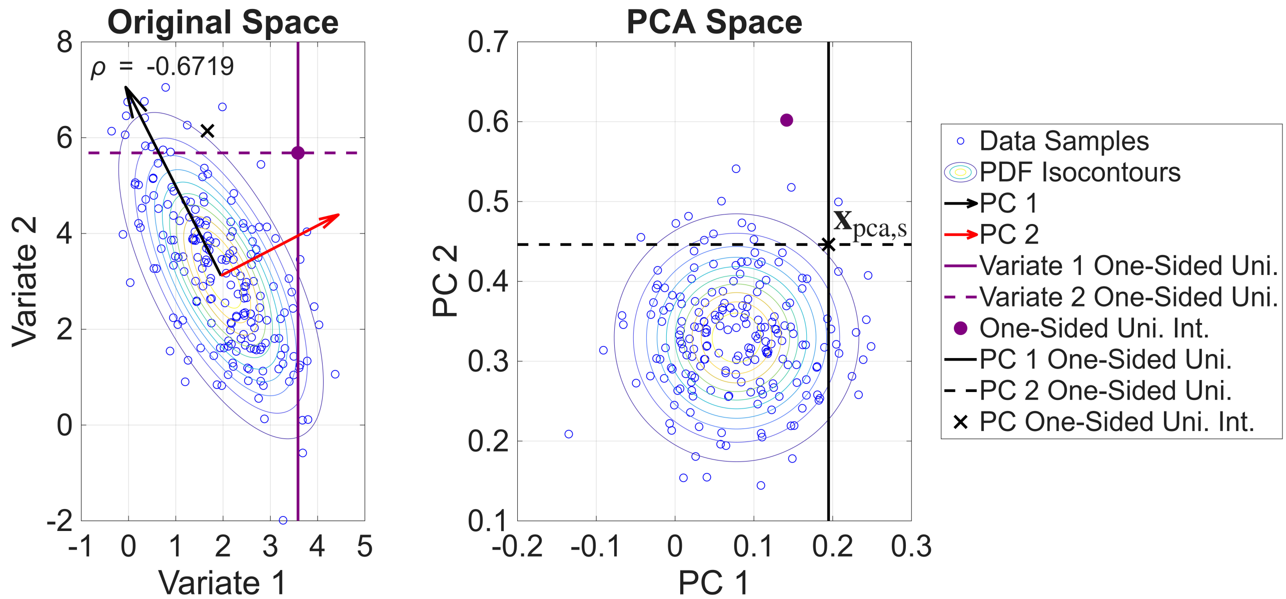

Fig. 1 Comparison of data samples and ellipses (PDF isocontours) in (left) original space (i.e., physical coordinates) vs. (right) PCA space to show the standardization. Variate 1 and 2 one-sided univariate upper tolerance limits and their intersection are shown as purple lines and a dot, respectively. Principle coordinate 1 and 2 one-sided univariate upper tolerance limits and their intersection are shown as black lines and an $\times$, respectively, with intersection denoted $\mathbf{x}_{\text{pca,s}}$.

Because PCA is often used for data reduction and reconstruction, it is common in many applications to use fewer principal components p than the number of dimensions (i.e., $p < m$). However, for the sake of brevity, in this paper we only consider the case where all principal components are retained (i.e., $p = m$)

Compute Univariate Intersection of Principal Coordinates

As mentioned previously, once the data are transformed into PCA space, each principal coordinate becomes independent. Statistical independence allows the principal coordinates to be treated separately in computing a one-sided upper tolerance bound. In fact, here, because the data are standardized in the PCA space to have the same sample standard deviation $\overline{\sigma}_{\text{pca}}$, only one calculation is necessary for a given coverage-confidence level. The mathematical form of this operation is shown in Eq. 1.

$$\mathbf{x}_{\text{pca,s}} = \begin{bmatrix} x_{\text{pca,s},1} \\ x_{\text{pca,s},2} \\ \vdots \\ x_{\text{pca,s},m} \end{bmatrix} = \begin{bmatrix} \overline{\mu}_{\text{pca},1} \\ \overline{\mu}_{\text{pca},2} \\ \vdots \\ \overline{\mu}_{\text{pca},m} \end{bmatrix} + k\left(n,\ \beta,\gamma\right)\overline{\sigma}_{\text{pca}}\mathbf{1} \tag{1}$$

In Eq. 1, $\mathbf{x}_{\text{pca,s}} \in \mathbb{R}^m$ is a column vector containing the univariate one-sided upper tolerance bounds for each principal coordinate (i.e., the intersection of the univariate bounds shown in the right plot in Figure 1), $x_{\text{pca,s},i}$ is the upper bound for principal coordinate i, $\overline{\mu}_{\text{pca},i}$ is the sample mean of principal coordinate i (stored in the principal coordinate sample mean vector $\overline{\boldsymbol{\mu}}_{\text{pca}}$), $k\left(n,\ \beta,\gamma\right)$ is the one-sided k-factor (or tolerance factor [8]) for number of samples n and prescribed coverage β and confidence γ, $\overline{\sigma}_{\text{pca}}$ is the sample standard deviation of each principal coordinate (equal due to data scaling) as mentioned previously, and $\mathbf{1}$ is an m-dimensional column vector of ones. Note that if one selects $p < m$, additional modifications to Equation 3.1 are needed to account for selected vs. unselected principal components. While the case of $p < m$ is not considered here, this case is permissible with this approach and is to be included as part of future work.

Mahalanobis Distance as Severity Surrogate

In the univariate specification development case, "severity" is usually taken simply as the higher value of the spectrum being considered. The fact that all values reside on a number line implies the ordering of different values very clearly. However, ordering is not so simple in the multivariate case, requiring the definition of some metric that represents increasing severity across multiple variates. Note that only considering univariate specification development does not alleviate this problem. Instead, such an approach means deciding that only one variate needs to be considered while allowing responses at all other locations and directions to vary arbitrarily.

In this paper, we propose using Mahalanobis distance (MD) as a severity surrogate for defining an environment specification. To the authors' knowledge, while the MD and Mahalanobis-squared distance have been commonly used as part of damage detection methodologies for structural health monitoring [9, 10], MD has not been explicitly used for assessing severity when defining, comparing, and evaluating environment specifications.

For the multivariate lognormal data considered here, the MD, denoted d_M, quantifies the distance between a point $\mathbf{a}$ and a distribution specified by a mean $\boldsymbol{\mu}_\mathrm{a}$ and covariance matrix $\boldsymbol{\Sigma}_\mathrm{a}$ as:

$$d_M(\mathbf{a};\ \boldsymbol{\mu}_\mathrm{a},\ \boldsymbol{\Sigma}_\mathrm{a})\ = \sqrt{(\mathbf{a}-\boldsymbol{\mu}_\mathrm{a})^T\boldsymbol{\Sigma}_\mathrm{a}^{-1}\,(\mathbf{a}-\boldsymbol{\mu}_\mathrm{a})} \tag{2}$$

This formulation allows for a simple way to represent the ordering of different potential multivariate points, with larger d_M values being more "severe" relative to smaller values. However, any ordering will not necessarily be unique—that is, multiple points may have the same MD. Also note that this metric is agnostic to any damage mechanisms, just as the corresponding univariate approach does not imply any particular failure model or damage metric.

Notably, MD is invariant under the PCA transformation (and affine transformations in general), meaning a selected point or contour in the PCA space represents the same distance (or severity) as the point transformed back to physical coordinates.

Project Result Back to Physical Coordinates

Once the univariate calculation has been performed in the PCA space at the chosen level of coverage and confidence, the hypersurface of points with the same MD can be determined. The selected specification point (here, $\mathbf{x}_{\mathrm{pca,s}}$) may then simply be transformed back to physical coordinates to produce the multivariate environmental specification.

Note that any point on the hypersphere in PCA space is equally valid according to this approach, though clearly some will be more damaging than others in practice. Metrics are currently in development which aim to select the point on the hypersphere that represents the most severe environment. These metrics are omitted for brevity and will be a subject of future work.

This approach has been described to this point assuming a given data matrix $\mathbf{A}$ at a single frequency line ω. However, to define an environment specification over a frequency range, one simply needs to repeat this process for all frequency lines of interest.

Figure 2 shows an overview of the process presented so far.

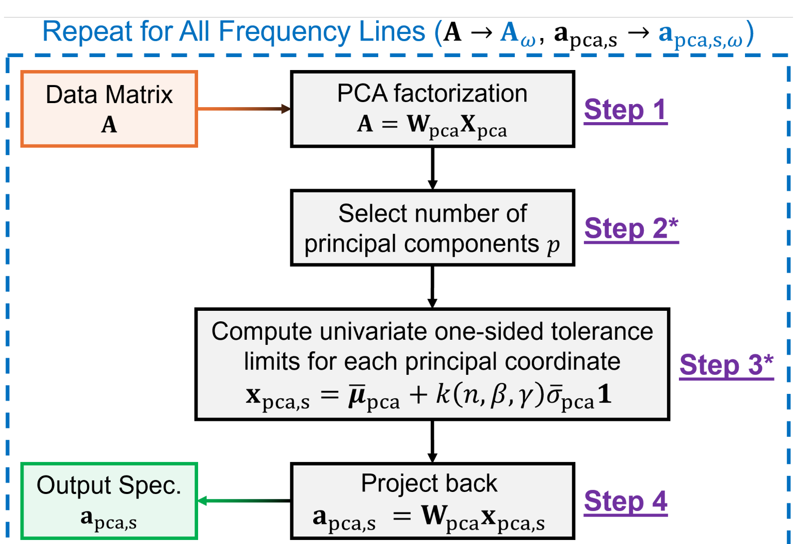

Fig. 2 Flowchart illustrating the proposed specification development process. *It is important to note that, as shown, the equation for $\mathbf{x}_{\mathrm{pca,s}}$ in Step 3 assumes $p = m$ (i.e., all principal components are used). However, for completeness, we include Step 2 where one may select $p < m$. If one selects $p < m$, Step 3 requires a minor modification to the equation for $\mathbf{x}_{\mathrm{pca,s}}$. This modification is a subject of future work, and we confine discussion here to the case of $p = m$.

Lastly, the PCA specification development approach outlined here is compatible with other specification development techniques. Notably, it can be combined with both frequency smoothing and spatial averaging schemes [11].

Conclusion

This paper proposes a novel approach to multivariate environmental specifications that retains the correlation structure of the underlying data while also retaining the computational simplicity of existing methods based on the univariate one-sided tolerance bound. This approach is enabled by the observation that projecting jointly normal source data onto principal components results in independent coordinates. In addition, data scaling results in the principal coordinates being standardized—that is, $\overline{\sigma}_{\mathrm{pca}}$ and $k(n, \beta, \gamma)$ each only needs to be computed once for given $\mathbf{A}$ at a given frequency line.

MD is proposed as a surrogate for assessing environment severity in the absence of information on damage modes when defining environment specifications. Notably, MD is invariant under the PCA transformation and therefore can be computed in either physical space or PCA space.

Future work is planned to extend this method into the selection of different points on the PCA hypersphere that may be of practical interest in testing or may better target damage modes of interest.

References

1. G. D'Elia, U. Musella, E. Mucchi et al., "Analyses of drives power reduction techniques for multi-axis random vibration control tests," Mechanical Systems and Signal Processing, vol. 135, pp. 106395, 2020. https://doi.org/10.1016/j.ymssp.2019.106395
2. P. M. Daborn, P. Ind, and D. Ewins, "Enhanced ground-based vibration testing for aerodynamic environments," Mechanical Systems and Signal Processing, vol. 49, no. 1-2, pp. 165-180, 2014. https://doi.org/10.1016/j.ymssp.2014.04.010
3. C. Haynes, T. Thompson, and A. Watts (2025), "Motivating Multivariate Specifications for Multiple-Input, Multiple-Output Vibration Testing." Proc. In International Modal Analysis Conference, Orlando, FL, Feb. 2025.
4. A. Watts, T. Thompson, and D. Harvey, "Confidence Intervals on Multivariate Normal Quantiles for Environmental Specification Development in Multi-axis Shock and Vibration Testing," arXiv preprint arXiv:2404.06565, 2024. https://doi.org/10.48550/arXiv.2404.06565
5. M. Greenacre, P. J. Groenen, T. Hastie et al., "Principal component analysis," Nature Reviews Methods Primers, vol. 2, no. 1, pp. 100, 2022. https://doi.org/10.1038/s43586-022-00184-w
6. MIL-STD-810H "Environmental Engineering Considerations and Laboratory Tests", Method 514.8, ANNEX F, APPENDIX B, pg. 4, January 2019.
7. D. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 1 ed.: Athena Scientific, 2002.
8. K. Krishnamoorthy and T. Mathew, *Statistical Tolerance Regions: Theory, Applications, and Computation*, 1 ed. John Wiley & Sons, Ltd, 2009. https://doi.org/10.1002/9780470473900
9. H. Sarmadi and A. Karamodin, "A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects," Mechanical Systems and Signal Processing, vol. 140, pp. 106495, 2020. https://doi.org/10.1016/j.ymssp.2019.106495
10. Y. L. Zhou, E. Figueiredo, N. Maia et al., "Damage detection in structures using a transmissibility-based Mahalanobis distance," Structural Control and Health Monitoring, vol. 22, no. 10, pp. 1209-1222, 2015. https://doi.org/10.1002/stc.1743
11. R. Schultz and G. Nelson, "Techniques for Modifying MIMO Random Vibration Specifications," Sensors and Instrumentation, Aircraft/Aerospace and Dynamic Environments Testing, Volume 7: Proceedings of the 40th IMAC, A Conference and Exposition on Structural Dynamics, pp. 71-83, 2022. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-05415-0_7