**Principal Component Analysis for Photoacoustic Experiment Optimization: Evaluating Correlation vs. Covariance Approaches**

S.M. Kovacevic1, M.N. Popovic1, M. Jordovic Pavlovic2, D.D. Markushev1, D.K. Markushev1

1*Institute of Physics, Belgrade, Serbia*

2*Faculty of Mechanical Engineering and Civil Engineering in Kraljevo, University of Kragujevac, Kragujevac, Serbia*

e-mail: slavica.kovacevic@ipb.ac.rs

**Background** – Photoacoustic material characterization is a non-destructive technique for assessing samples' optical, thermal, and mechanical properties. In this paper we utilize two approaches to PCA, covariance-based and correlation-based, to identify key modulation frequencies for neural network (NN) input, enabling accurate predictions. Using simulated data from a composite piston model [1], we analyzed thermoelastic photoacoustic signals from aluminum samples at varying thicknesses (l) and cut-off frequencies (f₀), across 300 modulation frequencies (10 Hz–10 MHz) in an open-cell setup (Fig 1a).

**Results** – The data pre-processing of input variables involved transforming amplitudes into decibels. For the covariance approach, the input data to PCA was scaled using maximum absolute (MA) scaling while for the correlation-based approach we used Z-scaled data as input, since $Cor(MA-scaled)=Cov(Z-scaled)$[].When using covariance-based PCA, five phases were selected with modulation frequencies ranging from 1 kHz to 38 kHz. In contrast, the correlation-based approach captured one amplitude and four phases, with frequencies between 22 kHz and 363 kHz.

The NN used had an input layer of 5 nodes, 5 and 10 nodes in the hidden layers and an output layer with 2 nodes, corresponding to *l* and *f0*. For better performance, logarithmic values of output features were used, and both input and output features of the NN were scaled using MA scaling. NN trained on the correlation-based set has smaller mean relative errors (0.03% for *l* and 0.05% for *f0*) compared to the NN trained on the covariance-based set (0.1% for *l* and 0.18% for *f0*) on the independent test data (data outside of the original set used for training).

Figure 1. Open-cell experimental setup (a) where the sample placed directly on the microphone

a)

b)

c)

 is periodically heated by a current modulated LED. Independent test results for *l* (b) and *f0* (c).

**Conclusion** – Current analysis shows that an NN with 5 thermoelastic signal points, selected using either of the two approaches, as input features predicts values of *l* and *f0* with adequate precision and accuracy (<5% relative error). While the correlation-based selected variables cover a wider range of modulation frequencies, covariance selected variables capture only phases at significantly lower frequencies, closer to the working range of photoacoustic, from 20 Hz to 20kHz, and thus is a better fit.

REFERENCES

[1] McDonald, F. A. & Wetsel, G. C. Generalized theory of the photoacoustic effect. J. Appl. Phys. 49, 2313–2322 (1978).

[2] Djordjević, К. Lj. et al. Influence of data scaling and normalization on overall neural network performances in photoacoustics. Opt. Quantum Electron. 54, 501 (2022).

[3] Choi, J. & Yang, X. Asymptotic properties of correlation-based principal component analysis, J. Econometrics, 229 (2022)