Detection and elimination of strain reading anomalies in distributed strain sensing readings

Lisa Ulbrich[1]

[1]Hentschke Bau GmbH, Zeppelinstr. 15, 02625 Bautzen, Germany

E-Mails: ulbich.lisa@hentschke-bau.de

Abstract: Distributed strain sensing (DSS) is increasingly used in structural health monitoring (SHM). Unlike conventional measurement technology, DSS allows for quasi-continuous measurement of strains with high spatial resolution, enabling the recording of global strain changes in the structure. However, DSS readings are subject to process-related interference, such as dropouts and strain reading anomalies (SRA). SRA can hinder data analysis during postprocessing or even falsify the results. Because of high amount of data produced by DSS, it is crucial to detect and remove SRA from DSS readings during automated preprocessing. Outlier removing methods from literature were analysed for their potential application to DSS readings. Based on this analysis, a sliding median z-score had been proposed, implemented in the *fosanalysis* framework and applied to various DSS readings for validation and to determine suitable parameters.

Keywords: Strain Reading Anomalies, Distributed Fibre Optic Sensors, Distributed Strain Sensing

Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: 10.15480/882.13497

1 Introduction

1.1 Distributed strain sensing in structural health monitoring

Distributed fibre optical sensors (DFOS) become increasingly relevant in structural health monitoring (SHM). Unlike traditional measurement techniques (e.g. strain gauges), DFOS can capture global changes in strain within the structure, making them suitable for both event-related and event-independent SHM purposes. Of particular interest is distributed strain sensing (DSS), based on coherent optical frequency domain reflectometry (cOFDR) of Rayleigh backscatter [1]. This measurement technique provides strain readings with an accuracy of up to 1 µm/m, measurement rates of up to 250 Hz, and spatial resolution of up to 0.65 mm [2] and therefore allows the monitoring of crack patterns and crack development in concrete structures [3] and facilitates the detection of deformations [4] within a structure. Hence, DSS has enormous potential for predictive maintenance [5]. For more detailed understanding of the principal of DSS, please refer to [3].

1.2 Strain Reading Anomalies (SRA)

Due to the application of cOFDR, process-related disturbances in the DSS readings like dropouts and SRA occur [6, 7]. Dropouts are measurement points for which no strain reading is available, while SRA manifest as significant local peaks in strain readings [7]. A strain peak is considered a SRA if it exhibits a considerable deviation from the previous and following measurement, in space and/or time (i.e. space and/or time domain) [8].

In [7], SRA in DFOS strain readings on reinforcing bars were analysed using a ODiSI-A OBR. It was shown, that SRA often occur at neighbouring sensor positions and thus form so-called anomalistic areas [7]. These anomalistic areas occur preferably in highly strained sensor sections or sections with high strain gradients [7]. Furthermore, SRA have been classified into harmful and harmless SRA [7]. Harmful SRA (HF-SRA) occur at the same sensor position at all measurement times, while harmless SRA (HL-SRA) only occur at certain times. In contrast to HL-SRA, HF-SRA only exhibit an abrupt increase along the time axis and then remain constant, which means that no more actual strain values are registered at this sensor positions [7, 8].

The presence of SRA can impede accurate data analysis during postprocessing or even lead to false results. Thus, elimination of strain reading anomalies is vital to ensure accurate and reliable results in SHM using DSS [1, 7]. However, the ideal approach is to prevent or minimize SRA through technical measures e.g. selecting an appropriate sensor [8]. If technical measures fail to prevent SRA, detecting and removing SRA is necessary as part of the data preprocessing [8].

1.3 Outlier elemination methods

In the following, elimination methods from other fields of application are also considered, which is why the term "outliers" is used instead of SRA. Usually, there are two strategies to deal with outliers:

1) **Smoothing methods** can align outliers with the measurement but do not remove them. Common smoothing methods are sliding average [1, 9], sliding median [1, 10], locally weighted scatterplot smoothing (LOWESS), spline interpolation, and Savitzky-Golay filter [9]. However, smoothing methods are disadvantageous for typical postprocessing applications (e.g., crack width calculation), as they replace actual measurement values with calculated (other) values and thus could falsify postprocessing results [9].

2) **Detection methods** use thresholds to detect and subsequently remove outliers. Detection methods are, e.g., geometrical threshold method (GTM) [8], polynomial interpolation comparison method (PICM) [8], and z-value-based methods [11]. Defining a threshold can be challenging as it should neither erroneously exclude genuine values (false positive) nor disregard outliers (false negative).

Both strategies can be applied to space or time domains or simultaneously to both domains. For DSS readings the strain state can vary significantly over time, whereas a change in strain over distance is usually limited due to the close proximity of the measuring points in the physical structure [8].

Z-score-based methods, including z-value, modified z-score and Whitaker-Hayes z-score [11], differ from other detection methods by increasing the exaggeration of an outlier, making it easier to determine a threshold. Additionally, these methods can be fully automated, unlike PICM, making them highly efficient for analyzing large datasets. Due to these favorable properties, z-score-based methods show great potential for detecting SRA and will be elaborated on further.

While z-score is based on the mean, modified z-score uses the median [11]. Since the median itself is more stable against outliers, the difference between modified z-score of actual measurement values z_{AMV} and z-score of SRA-values z_{SRA} is more significant. Thus, modified z-score has a greater potential for SRA detection in DSS than z-score [11]. The modified z-score z_i is calculated as follows:

$$z_i = \left|0.6745 \cdot \frac{y_i - \tilde{y}}{\mathrm{MAD}}\right| \tag{1}$$

Where y_i is the measurement value i, $\tilde{y}$ is the median of all measured values y. The term $y_i - \tilde{y}$ is called absolute deviation (AD). MAD is the median of all AD's. The multiplier 0.6745 compensates for the asymptotic distortion when calculating MAD from standard distribution [12].

The Whitaker-Hayes z-score according to [13] uses the median of $\Delta_i = y_i - y_{i-1}$. Since in DSS readings Δ_i is significant between SRA and their predecessor as well as their successor, there might be significant z_{AMV}, that can be classified false positive. Another issue appears with anomalistic areas. Up to the values of neighbouring SRA, Δ_i might be small, resulting in small z_{SRA}. In this case, the result is false negative. Last but not least, dropouts next to SRA lead to missing AD values and thus to missing z_{SRA}, resulting in false negative classification.

All in all, modified z-score shows the greatest potential for SRA detection in DSS readings. However, DSS readings with wide measurement value range, i.e. due to cracks, lead to z_{AMV} close to z_{SRA} due the shift of the median resulting in higher AD. Therefore, the modified z-score is not reliable for SRA detection in DSS readings with a wide measurement value range.

2 Sliding modified z-score

In order to overcome the issues of modified z-score described before, a combination of sliding median and modified z-score called sliding modified z-score sz_i, is proposed as follows:

$$sz_i = \left|0.6745 \cdot \frac{\varepsilon_i - \widetilde{\varepsilon_{(\iota \pm r)}}}{\mathrm{MAD}}\right| \tag{2}$$

where ε_i is the strain measurement value i and $\widetilde{\varepsilon_{(\iota \pm r)}}$ is the median of all strain measurement values in the radius r (sliding median). Thereby, term $\varepsilon_i - \widetilde{\varepsilon_{(\iota \pm r)}}$ is called the AD from the sliding median. MAD is the median of all AD's from the sliding median.

In contrast to median $\tilde{y}$, the sliding median $\widetilde{\varepsilon_{(\iota \pm r)}}$ leads to a median close to ε_i and thus to smaller AD. In order to reduce the AD for ε_{AMV} but not for ε_{SRA}, r has to be choosen small enough to get $\widetilde{\varepsilon_{(\iota \pm r)}}$ close to ε_{AMV}, but wide enough minimize influence by SRA.

Due to the sliding median, AD can possibly tend towards 0. Thus, MAD might be 0 and sz_i cannot be calculated. In that case, MAD is replaced by the mean of all AD's from the sliding median (MeanAD). In order to compensate for asymptotic distortion when calculating MeanAD from a standard distribution, the multiplier has to be replaced by 0.7979 [12].

In order to classify and sort out SRA, a threshold t has to be defined. If sz_i exceeds t, ε_i is identified as SRA and replaced by a dropout:

$$sz_i > t \rightarrow SRA \tag{3}$$

In order to validate the algorithm and find a suitable r, various DSS readings where processed in the space domain with different r. All readings were from an ODiSi 6100 from Luna Innovation Inc. using monolithic Epsilon Sensors from Nerve-Sensors. The performance of the sliding modified z-score was validated regarding the resulting distance Δsz between maximum sliding modified z-score the actual measurement values $\max z_{AMV}$ and minimum sliding modified z-score of the SRA's $\min z_{SRA}$. A high Δsz allows determination of a reliable t, which neither erroneously excludes genuine values (no false positive) nor disregards outliers (no false negative). In case of $\Delta sz \leq 0$ definition of t is not possible.

3 Demonstration strain readings from single crack

A dataset of DSS readings from a single crack in a prestressed concrete girder in 4-point bending is used for demonstration. The dataset contains a single SRA at x = 0,039 m and a variety of dropouts in the neighborhood. The SRA was varied regarding the difference S between the latest available ε_{AMV} and ε_{SRA} in a range between +5000 µm/m and -5000 µm/m, see *figure 1*.

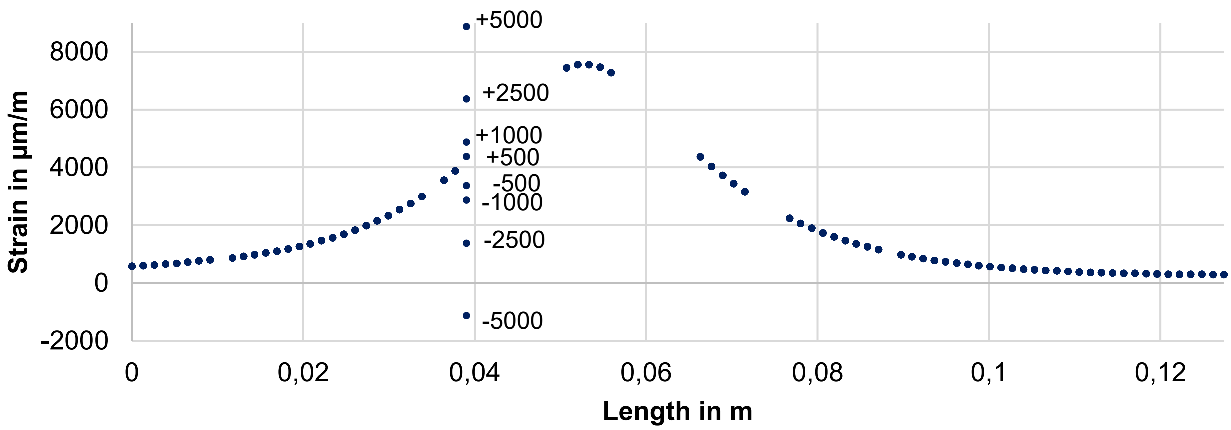

Figure 1: Predefined Dropouts and SRA's in the crack area.

Sliding modified z-score was applied to the DSS reading for each SRA variation with radius $r =$ 1…15. Examples for the resulting distances Δsz are shown in *Figure 2*.

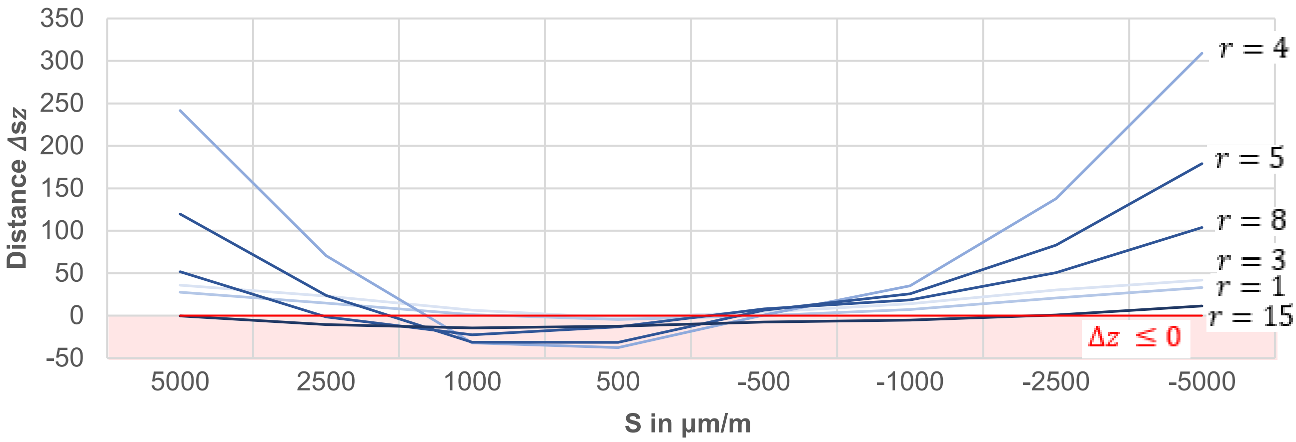

Figure 2: Distance Δz between $\max z_{AMV}$ and $\min z_{SRA}$.

For r = 1-3 all SRA can be detected, except for S = 500 µm/m. However, Δsz is quiet small for all SRA. The highest Δsz can be reached with $r = 4$ or 5. For $r = 4$ or 5 only SRA with S = 500 µm/m and S = 1000 µm/m can not be detected. Thus, $r = 4$ and 5 are most promising for removing SRA. While with $r = 4$, SRA still have high impact on sliding median $\widetilde{\varepsilon_{(\iota \pm r)}}$, with $r = 5$ SRA have less impact. For higher r, Δsz becomes smaller and fewer SRA can be detected.

4 Validation on strain readings with multiple cracks

In order to validate algorithm and parameters, the proposed modified z-score has been applied to DSS readings of prestressed concrete beams with multiple cracking, see *figure 3*. Due to pre-stressing and multiple cracking strain readings show a wide range between appr. -220 µm/m and +12 000 µm/m. While reading 1 has only one SRA (a1) at x = 2.7963 m, reading 2 has an accumulation of four SRA (a2 to d2) next to each other at x = 0.7475 m … 0.7527 m forming an anomalistic area, with one dropout between. Parameter r was set to 4 and 5.

For reading 1, it delivers a high Δsz with both r, see *Figure 4*. For reading 2, the resulting Δsz are much smaller, see *Figure 5*. Furthermore, the resulting Δsz differ up to $|S|$ of the SRA. Thereby, $|S|$ is the absolute difference between the latest available ε_{AMV} and ε_{SRA}. SRA with $|S| > 6900$ µm/m (a2 and d2) result in positive Δsz and are thus possible to recognise and eliminate by a suitable threshold *t*. SRA with $|S| < 1500$ µm/m (b2 and c2) result in negative Δsz and can not be detected anymore. $r = 4$ results in higher Δsz also in reading 2.

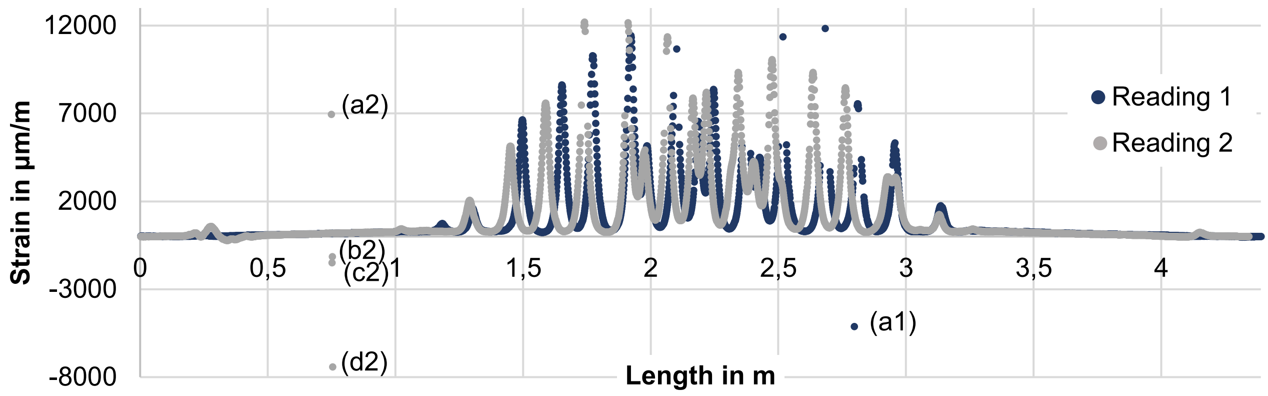

Figure 3: DSS reading 1 and 2.

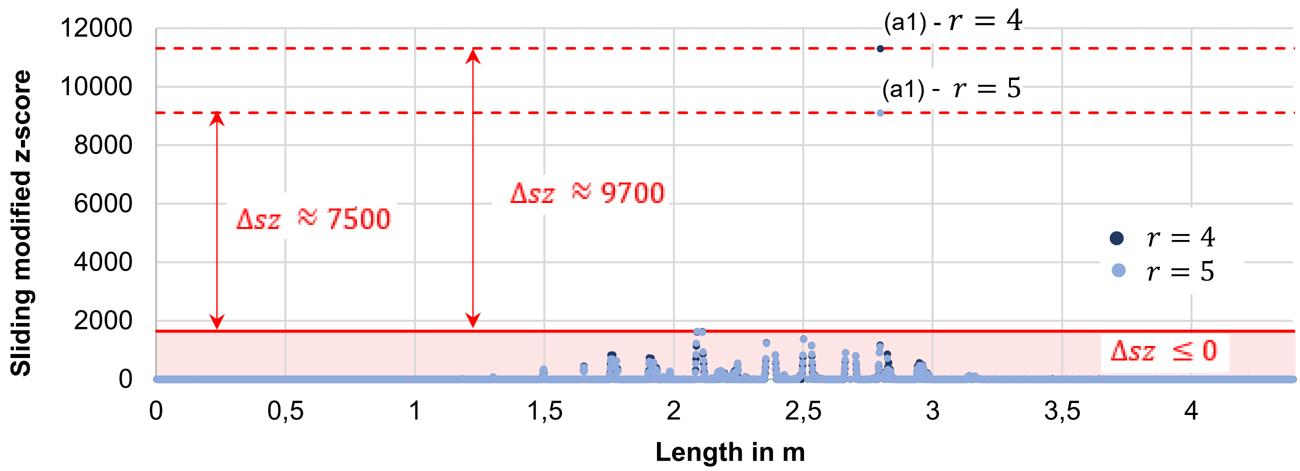

Figure 4: Sliding modified z-score for DSS reading 1 with $r = 4$ and 5.

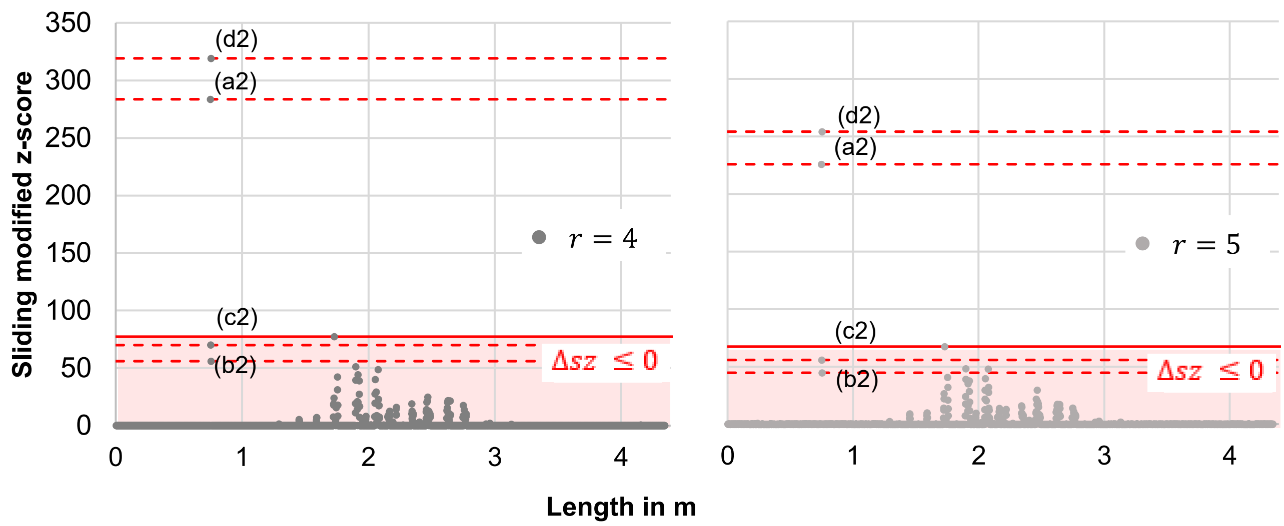

Figure 5: Sliding modified z-score for DSS reading 2 with $r = 4$ and 5.

5 Discussion

The application of the sliding modified z-score on a single crack shows, that $r = 4$ and 5 are most promising for SRA detection. $r > 5$ increases AD, resulting in a smaller Δz. For $r < 4$ sliding median is influenced by SRA, resulting in small AD and a smaller Δsz. Furthermore, the closer an SRA is to the actual measurment values, the smaller is Δsz. The smaller Δsz, the less likely is a threshold t for DSS readings with a wide range of measured values, which neither ignores outliers nor sorts out real measured values.

The validation shows, that sliding modified z-score works well with $r = 4$ and 5 for DSS readings with a wide range of measured values and with high SRA. It allows defining a threshold, which neither erroneously exclude genuine values nor disregard outliers. In reading 2, the anomalistic area influences the median, resulting in a low AD for SRA. Additionally, MAD becomes 0 and has been replaced by MeanAD, which is, due to the impact of SRA, higher than MAD. Thus, for smaller SRA Δsz got negative. In order to avoid this, r has to be increased, when there are two or more SRA next to each other.

6 Conclusion

Sliding modified z-score has shown to allow reliably detecting and removing SRA in DSS readings. Currently, the definition of a proper threshold *t* seems to be specific to the dataset. An automatic definition of threshold *t* would enable an automated SRA removal in DSS readings. Since small SRA have only minor or neglible impact on postprocessing results, automatic SRA removal can be limited to large and medium SRA's in most practical applications. The impact of SRA on postprocessing applications, such as crack detection, crack width calculation and others, has to be investigated.

Sliding modified z-score will be published in fosanalysis v0.4 .The software fosanalysis is available as free/libre open source software at https://github.com/TUD-IMB/fosanalysis.

Acknowledgements

This article presents results of the research project IDA-KI (Automated assessment of monitoring data for infrastructure constructions using AI and IoT), funded by the Federal Ministry for Digital and Transport, Germany, within the innovation program mFUND (funding reference: 19FS2013C).

References

[1] B. Richter, M. Herbers, and S. Marx, "Crack monitoring on concrete structures with distributed fiber optic sensors – Toward automated data evaluation and assessment," *Structural Concrete*, vol. 25, no. 2, pp.1465-1480, Nov. 2023. doi: 10.1002/suco.202300100.

[2] Luna Innovations Inc., *Optical Distributed Sensor Interrogator Model ODiSI 6: User's Guide,* 2020. [Online] Available: https://lunainc.com/sites/default/files/assets/files/resource-library/ODiSI%206100%20User%20Guide.pdf. [Accessed: Jun. 8, 2024.]

[3] O. Fischer, S. Thoma and S. Crepaz, „Distributed fibre optic sensing for crack detection in concrete structures“, *Civil Engineering Design,* vol. 1, Jul. 2019. doi: 10.1002/cend.201900008.

[4] C. Monsberger and W. Lienhart, "Distributed Fiber Optic Shape Sensing of Concrete Structures", *Sensors,* vol. 21, no. 18, Sept. 2021. doi: 10.3390/s21186098.

[5] M. Herbers, B. Richter, B. Schwarz, and S. Marx, „Messtechnische Überwachung von Spannbetonbauteilen mit faseroptischen Sensoren,“ In Proc. 12. Symposium Experimentelle Untersuchungen von Baukonstruktionen, 2023. doi: 10.25368/2023.143.

[6] S. Poe and M. Davis, Luna Innovations Inc., Personal emails (Jan. 2024).

[7] M. F. Bado, J. R. Casas, and A. Barrias, "Performance of Rayleigh-Based Distributed Optical Fiber Sensors Bonded to Reinforcing Bars in Bending," *Sensors,* vol. 18, no. 9, Sept. 2018. doi: 10.3390/s18093125.

[8] M. F. Bado, J. R. Casas, and J. G. i Esteve, "Post-processing algorithms for distributed optical fiber sensing in structural health monitoring applications," *Structural Health Monitoring,* vol. 20, no. 2, Jun. 2020. doi: 10.1177/1475921720921559.

[9] T. Janiak, H. Becks, B. Camps, M. Classen, and J. Hegger, "Evaluation of distributed fibre optic sensors in structural concrete," *Materials and Structures,* vol. 56, no. 159, Oct. 2023. doi: 10.1617/s11527-023-02222-9.

[10] T. O'Haver, *A pragmatic Introduction to Signal Processing with application in scientific measurements*, 2024. [Online]. Available: https://terpconnect.umd.edu/~toh/spectrum/. [Accessed: May 3, 2024.]

[11] N. Coca, "Removing Spikes from Raman Spectra with Anomaly Detection: Whitaker-Hayes algorithm in Python", *Towards Data Science,* Nov. 2019. [Online]. Available: https://towardsdatascience.com/removing-spikes-from-raman-spectra-8a9fdda0ac22. [Accessed: Feb. 12, 2024.]

[12] IBM, "Modified z score", 2024. [Online]. Available: https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-modified-z-score. [Accessed: Mar. 16, 2024.]

[13] D. A. Whitaker and K. Hayes, "A simple algorithm for despiking Raman spectra," *Chemometrics and Intelligent Laboratory Systems,* vol. 179, Jun. 2018. doi: 10.1016/j.chemolab.2018.06.009.