# Opto-Electronics Review

# Detection and elimination of unwanted spikes (artifacts) in rotational seismographic signals using artificial intelligence

Bartosz Sakowicz[1*] ⓘ, Marek Kamiński[1] ⓘ, Paweł Marciniak[1] ⓘ, Mariusz Jankowski[1] ⓘ,
Leszek. R. Jaroszewicz[2,3] ⓘ

[1] Department of Microelectronics and Computer Science, Lodz University of Technology, ul. Wólczańska 221, 93-005 Lódź, Poland
[2] Institute of Applied Physics, Military University of Technology, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
[3] Elproma Elektronika Sp. z o.o., ul. Duńska 2A, 05-152 Czosnów, Poland

**Abstract**

This paper presents methods for detecting and eliminating artifacts in signals recorded by the FOS6 rotational seismograph based on the Sagnac effect. A combination of classical threshold-based techniques and artificial intelligence (AI) algorithms was employed, aimed not only at detecting artifacts but also at improving the overall quality of the recorded data. Particular emphasis was placed on the deliberate use of AI – not as a direct filtering tool, but as a means of identifying regions of the signal that can be effectively smoothed or removed while preserving waveform integrity. The threshold-based algorithm mainly functioned as a source of training data for the AI models, enabling effective learning and testing of the approaches developed. Training data were obtained from the earlier FOS5 device, and verification was performed using recordings from both FOS5 and FOS6, enabling evaluation of the proposed methods under real-world conditions. To suppress artifacts, a simple linear interpolation method was proposed that preserves signal continuity and morphology while minimising distortion. The results show that this combined approach significantly increases the usability of the measurement system, enabling a more reliable analysis of seismic events and reducing the number of false alarms.

## 1. Introduction

The advanced rotational seismographs offer a comprehensive, six-component (three for translational and three for rotational) description of the seismic field at any given location. This enhanced data enables a deeper understanding of seismic wave propagation and the fundamental characteristics of earthquake sources. Furthermore, information on rotational ground motions is crucial for evaluating how building structures, particularly high-rise and large-scale constructions, respond to seismic events. By understanding these movements, engineers can design more resilient buildings and conduct more accurate seismic risk assessments. The device also presents significant potential in various other fields, including monitoring volcanic activity and explosions, as well as imaging subsurface geological structures, which is particularly valuable for oil and gas exploration.

The primary experimental data presented in this paper were obtained from a fibre-optic seismometer type 6 (FOS6), which meets all the requirements for rotational seismology [1]. It is a newly developed fibre-optic interferometer with high-resolution readout of the Sagnac phase shift [2], induced between two counter-propagating waves in a closed optical path when the plane of propagation undergoes angular velocity [3, 4]. This device is the result of 20 years of research during the development of four classes of fibre-optic rotational seismometers (FORSs): starting with an autonomous fibre-optic rotational seismometer (AFORS) [5, 6] through a fibre-optic system for rotational events and phenomena monitoring (FOSREM) type SS [7, 8] and type BB [9, 10], and up to fibre-optic seismographs (FOS5 and FOS6) [11–13]. The FOSREM-3D project has yielded the development of the world's most precise 3-axis rotational seismograph (Fig. 1)

**Fig. 1.** FOS6 device (left) and FOS6 device on test platform (right).
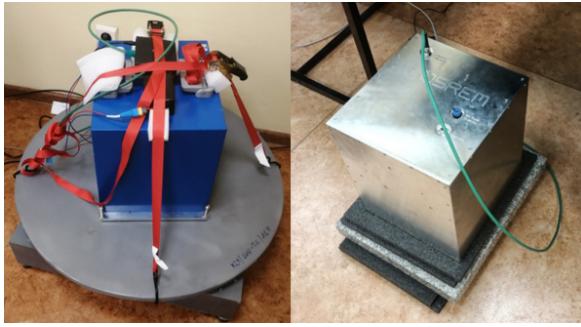
with a detection bandpass of 0.01–100 Hz, capable of measuring angular velocity with sensitivity in the range of 50 nrad/s in 1 Hz bandpass.

The general construction of the FOS6 consists of two interdependent parts: optical and electronic (Fig. 2). The optical heads are cylindrical, with a diameter of 250 mm and a height of less than 50 mm. They incorporate a fibre-optic Sagnac interferometer in the so-called minimal fibre-optic gyroscope (FOG) configuration.



**Fig. 2.** Simplified (showing one of the three optical heads) block diagram of the analysed FOS6 [15].
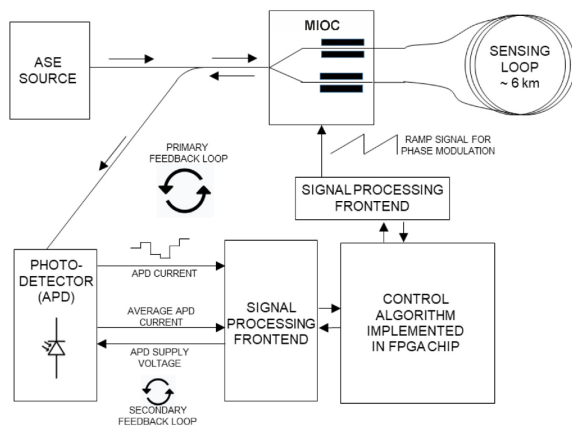
By employing a sensor loop in the form of a bifilar, quadrupolar winding of fibre with a nominal length of 6000 m and a total loss of approximately 1 dB, and by combining other elements ensuring an optical loss of the system at the level of 15 dB, supplying the system with a single ASE source with a power of 14 dBm ensures a theoretical sensitivity of around 5 nrad/s for a 1 Hz bandpass. All FOS6 use a single-mode optical fibre, SMF-28e+ (Corning Inc., New York, NY, USA), a multi-integrated optical circuit (MIOC) (IdealPhotonics Ltd., Kwun Tong, Hong Kong), and other telecommunication fibre-optic components. The triaxial optics (optical heads) installed in FOS6 work in conjunction with a dedicated electronic system that detects angular velocity by processing the Sagnac phase shift in a feedback loop, as initially proposed for the construction of FOG as a so-called closed-loop system [14].

The electronic part consists of five modules: sensor and MIOC modules mounted in each optical head, a power module supplying power to all electronic modules, an ASE source, and a control module implementing closed-loop

detection. The sensor module uses an avalanche photodiode (APD), where, due to the non-linear operating characteristics and the very low level of the received optical signal, it was necessary to create a low-noise, thermally stable structure insensitive to external electromagnetic fields. The MIOC is a Y-type waveguide fabricated in planar technology on a lithium niobate crystal. Polarisation-maintaining optical fibres with a diameter of 80 micrometers are externally connected to the three branches of this structure. At the same time, electrodes located at the Y-junction enable a sawtooth phase modulation via an external voltage of approximately 4.5 V.

The control module is a key component for implementing closed-loop detection. This module implements the so-called four-step digital phase shift, optimised for each optical head's resonant frequency (with a fibre length accuracy of 1 m) and providing a stable operating point of $3\pi/4$. This module also provides the device ID and supports all other modules. The control module also includes input/output communication, enabling external setting of selected module parameters, as well as outputting a time-stamped, GPS-based, digitised rotational velocity value for each channel at a 1000 Hz sampling rate, in the form of a data frame.

The power control unit (PCU) is the second element of the FOSREM-3D system that has been created. The module can be powered from the mains, an external battery, or a photovoltaic panel. As a communication unit, the PCU performs remote configuration of selected FOS6 parameters and carries out direct processing of data acquired from FOS6, including: its archiving as raw data with a 1000 Hz sampling rate and storage in the form of 1-hour files in the miniSEED format, compatible with firmware commonly available in the seismology field.

In modern seismology, one of the significant challenges is distinguishing genuine seismic events from transient noise anomalies, hereafter referred to in the paper as artifacts – short-lived, high-amplitude spikes with non-seismic origins. These artifacts can be deterministic, resulting from electronic interference, such as the conversion of an analogue signal to a digital signal using an "imperfect" analogue-to-digital converter, transmission errors, or environmental noise. Additionally, it can be directly linked to a rotation rate, as artifacts exhibit a particular type of sharp, low-amplitude spikes that are regularly observed above a few Hz, with their amplitudes decreasing as the frequency increases [16]. Such artifacts are characteristic of all FOGs that use digital closed-loop processing and result from "ramp peaks". The electronic system employs a ramp voltage to counteract rotations, increasing and resetting to zero upon reaching $2\pi$, with a reset frequency depending on the detected rotation rate. This process creates a periodic, though not perfectly sinusoidal, signal that generates a spike at the primary frequency and harmonics at higher frequencies (see example in Fig. 3). For FOGs, where angular changes are detected via a suitable rotation rate integration, these peaks are negligible due to their averaged mathematical impact. However, in FORS, they can be confused and require removal.

The presence of the artifacts described above is a problem for two reasons. Firstly, signal distortion caused by them makes it difficult or impossible to analyse, as the
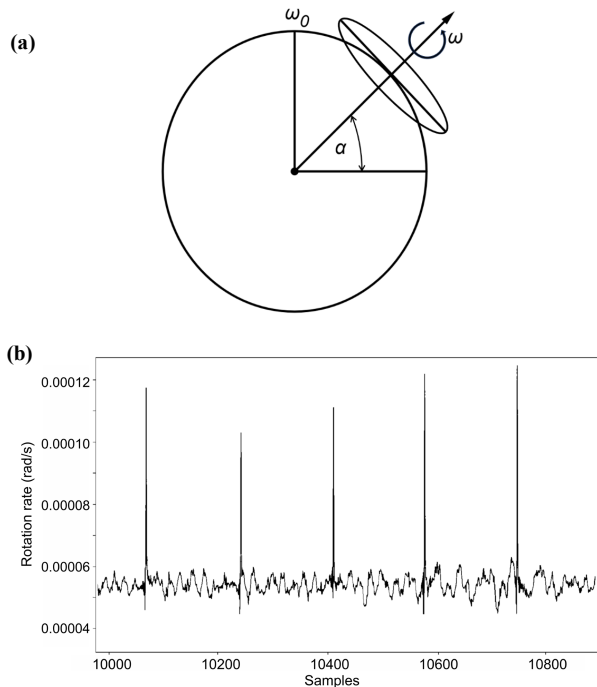
**Fig. 3.** The FOS6 positioned on the Earth's surface at a given latitude α with three axes aligned vertically, East-West, and North-South (a), and an example of artifacts as periodically repeating spikes on the upper-axis (Earth's angular velocity multiplied by the cosine of the latitude) (b).

distorted portion of the signal essentially becomes useless [17]. Of course, with short 3–4 millisecond spikes, this problem is less significant. The second reason is the negative impact on the signal recording and alarm reporting processes. A typical recorder consists of a digitizer connected to a computer, which can store all data continuously or store only seismic events [18]. In continuous monitoring systems, where the occurrence of expected phenomena must be recorded, the automatic start of data recording is crucial. This is the situation in seismology during the collection of seismic events, where the classical procedure of starting data collection when the signal exceeds a certain threshold is generally used. However, if the ADC part of the system generates pins (single or quasi-periodic data with a random value above the noise level), the above recording procedure starts an inefficient operation that can be classified as errors.

An algorithm that can distinguish real seismic events from artifacts and then remove them is a necessary and valuable tool in such situations. The primary purpose of detecting spikes is not only to identify artifacts but also to suppress or eliminate them, thereby improving the quality of recorded data and preventing false triggering of seismic event detection. Ideally, such an algorithm would already be in place at the signal generation stage, but this is not always possible (especially in the case of signals that have already been recorded with distortions). Therefore, a complementary version of the algorithm has been developed to process and clean up already distorted signals.

## 2. Hardware methods of artifact detection and removal

There are two solutions for effectively removing artifacts generated by imperfect MIOC work. The first of these,

referred to as a "passive" approach, concerns the system's electronic components, which are based on FOG technology, and involves applying the deramping procedure to miniSEED files containing raw data. This method has proven effective for both single-axis and three-axis FORS-blueSeis-3A devices [19]. It works by correlating the system output signal with data from the MIOC control signal, thereby eliminating the ramp peaks from the raw measurements (see Fig. 4).
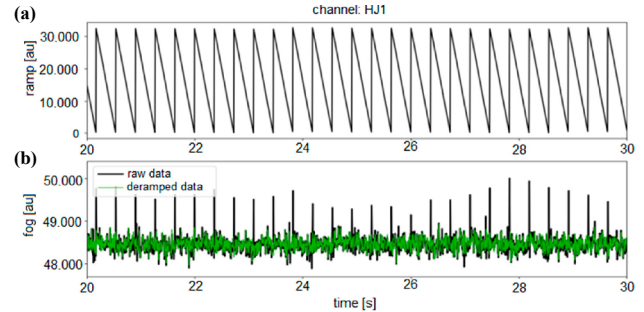


**Fig. 4.** Example of a deramping procedure protected against the effective removal of ramp peaks from the raw data: an example of the ramp signal driving the MIOC (a), and a simulation of the raw and deramping data using the Bernauer script (b) [19].

The second approach, known as the "active" method, involves modifying the configuration of a field-programmable gate array (FPGA). This approach leverages precise knowledge of when the MIOC signal reaches a $2\pi$ threshold, enabling the problematic sample to be removed and replaced with an interpolated value derived from adjacent samples. A higher signal digitisation rate is recommended to enhance the effectiveness of this method. In summary, the correct choice of algorithm is much more complicated due to the more complex structure of artefacts or other types of imperfections. These are easy to recognise visually, but more challenging to identify automatically. In such situations, an artificial intelligence (AI)-based solution may be a suitable choice.

## 3. State of the art of AI methods for artifact detection

With the exponential growth in seismic data volume, especially from dense networks and distributed acoustic sensing (DAS) systems, traditional thresholding or rule-based filtering approaches are no longer sufficient. In this context, AI, particularly machine learning (ML) and deep learning (DL) techniques have emerged as powerful tools for automating and refining artifact spike detection in real-time.

Convolutional neural networks (CNNs) are among the most widely used methods in seismic signal analysis due to their strong pattern-recognition capabilities. For instance, the DASEventNet model achieved 100% accuracy in microseismic detection from DAS waveforms, highlighting the potential of CNNs to learn complex waveform morphologies [20]. Nevertheless, CNNs are prone to overfitting, particularly when artifact spikes closely resemble genuine seismic events. A lack of representative noise samples in training datasets further increases the risk of frequent misclassifications [17].

Another important family of models are recurrent neural networks (RNNs), including their variants with long short-term memory (LSTM). These architectures can model temporal dependencies, which makes them suitable for analysing signal sequences where contextual information is essential. A bidirectional LSTM-based design has proven effective in distinguishing complex transient signals in medical data, and this capability can be translated to the domain of seismic artifact detection [21]. However, LSTM networks may suffer from vanishing gradients and instability when trained on long sequences, which reduces their ability to detect delayed or overlapping artifacts.

Transformer architectures have also been introduced into seismology. Models such as QuanFormer [22], which leverage attention mechanisms to achieve global context awareness are increasingly applied in seismic phase picking [23]. Their ability to differentiate contextual patterns enables them to highlight subtle anomalies indicative of artifacts. The main drawback, however, lies in their demand for extensive labelled data and significant computational resources, which makes them less feasible for real-time deployment in resource-constrained environments.

Beyond supervised learning, autoencoders and other unsupervised learning approaches have been investigated. Autoencoders are trained to reconstruct clean signals and are therefore adept at identifying outliers that deviate from learned norms, such as unwanted spike artifacts. They have already been successfully applied in bio-signal denoising and demonstrate promising potential for seismic anomaly detection. Yet, their lack of specificity can lead to the misclassification of weak but real seismic events as noise, particularly at low signal-to-noise ratios (SNR).

Another promising direction is transfer learning, where large regional datasets, such as the Texas Earthquake Dataset (TXED), enable adapting global models to local seismic conditions. This approach improves model performance in regions with distinctive seismic and noise characteristics [17]. However, without domain-specific fine-tuning, transferred models risk underperformance due to environmental and geophysical discrepancies.

Despite their demonstrated potential, all these AI-based techniques face several methodological challenges. Class imbalance is a critical issue because artifact spikes are rare compared to regular seismic events, resulting in highly skewed datasets. This often results in models exhibiting high precision but low recall for the artifact class. The second limitation relates to the inability to explain. In critical systems, such as earthquake early warning, model decisions need to be interpretable; however, most deep learning models act as "black boxes," which reduces user trust and complicates debugging [24]. Ultimately, the quality of training data remains a crucial factor in determining model performance. Variations in annotation standards across different seismic networks can introduce inconsistency and reduce the generalisability of AI systems [25].

Future research directions therefore include the development of hybrid models that combine CNN, Transformer, and RNN modules to leverage both local and global features; the application of explainable AI techniques, such as attention map or saliency visualisation, to improve interpretability of artifact detection; the creation of region-specific benchmarks, such as TXED, to support fine-tuning in unique geological settings; and the use of data augmentation and class balancing methods, including synthetic artifact generation and weighted loss functions, to address dataset imbalance.

## 4. Threshold method

However, the effectiveness of all these approaches ultimately depends on the availability of reliable training data. Preparing accurately classified sets of spike artifacts and undistorted signals is both time-consuming and inefficient when performed manually, yet it remains indispensable for training robust models. To address this bottleneck, we developed a deterministic threshold-based algorithm for signal classification. This algorithm is not intended as an alternative to AI methods, but rather as a practical and efficient source of high-quality training data, ensuring that advanced models receive consistent, trustworthy input.

The threshold method was designed specifically for generating such datasets. The fundamental challenge lies in preparing learning sets, and manual selection of waveforms is inefficient given the demands of multilayer neural networks (some comments on how to assess this issue). There is no fixed formula to determine the number of samples required to achieve reliable results, as this depends on multiple factors [26–29]. Model complexity matters as deep neural networks require more data than shallow models such as linear regression, and larger models are more prone to overfitting on small datasets. Data quality is equally essential: clean, informative, and diverse datasets can outperform much larger but noisy ones, while lower noise levels generally reduce the required sample size. Overfitting is typically revealed when high training accuracy is accompanied by low test accuracy, indicating that the model has insufficient data or inadequate regularisation. Finally, class imbalance further increases the need for larger datasets when certain classes are underrepresented.

In summary, there is no "magic number" for the ideal dataset size in neural network training. It is always a balance between problem complexity, model complexity, and data characteristics, and in practice it is often an iterative process. One begins with a reasonable dataset size and then monitors model performance, expanding the training data as needed.

To support this process, a threshold-based algorithm was applied to classify signals based on several conditions. Samples were marked as disturbances (spikes) if a local extreme signal value was present, if the value exceeded the mean value of the modulus of the original signal (after removal of the DC component) several times, if no signal of comparable amplitude and opposite sign was found in the immediate neighbourhood, and if several consecutive potential spikes appeared at repeated distances, indicating quasi-periodicity. Threshold and signal characteristic values (e.g., DC components) were calculated within selected time windows (Fig. 5). As a result, the algorithm performed worst in transition areas, reflecting changes in the observed phenomena (for example, pure noise *vs.* vehicle movement near the sensor).

The adopted methodology inevitably produces errors, both by omitting some noise and eliminating valid signal segments (Figs. 6 and 7). Nonetheless, it provides a sys-

tematic and reproducible framework that can replace manual selection with a deterministic process. Most importantly, it serves as a robust source of labelled data for training advanced AI models, thereby linking classical thresholding with modern deep learning approaches in a complementary manner.

## 5. Methodology

A suitable development environment for time series classification was established by installing Python (version 3.12) and creating an isolated virtual environment using *venv* procedure to manage project-specific dependencies. Essential libraries, including NumPy, Pandas, Matplotlib, and Scikit-learn, were installed for data manipulation, visualisation, and classical ML. PyTorch was used for DL implementations, and additional libraries were installed to support statistical time-series analysis and feature extraction. Code development was carried out using the VS Code IDE, with Git employed for version control and collaborative development.

To support data preparation, a dedicated application was developed to visualise time series data and enable manual selection of specific samples for subsequent learning iterations. This provided a mechanism for refining the network model through targeted data curation.

During algorithm development, the FOS6 device had not yet been deployed for regular operation so that no learning data could be obtained from it. It was not even entirely clear what the spike artifacts would look like, as the control algorithms were still under development. For this reason, signals recorded with a single-axis FOS5 device were used to develop AI models. The FOS5 is similar in design to the FOS6 but is an earlier, less advanced version.

The data were recorded on October 10th, 2021, using an instrument installed in an underground laboratory located in a bunker in Książ (in Lower Silesian Voivodeship, Poland). Figure 8 represents a 2-hour dataset comprising approximately 720 000 data points. Due to this density, the detected spikes, indicated by red markers, visually merge and become difficult to distinguish in detail. The figure is intended for illustrative purposes; therefore, close-up views are provided in other figures (Figs. 10–12) for clarity. It should be noted that the description refers to Fig. 8. In practice, we used two 2-hour recordings (including the one shown in Fig. 8) as training data for the algorithm. The obtained signals exhibited high variability in the observed phenomena. The night recordings were practically free of distorting signals, whereas the daytime data were strongly affected by visitors to the bunker. Both signal types were included in the learning and test datasets.

From approximately two hours of recorded signals, 1-second windows containing either spike artifacts or clean signals (without spikes) were extracted automatically using a threshold-based algorithm. All extracted samples were subsequently reviewed to eliminate potential misclassifications. The final curated dataset consisted of:

1. 509 signals without spikes (duration 1.001 s),
2. 235 signals with spikes (duration 1.001 s).

For samples containing spikes (Fig. 9), the extraction window was adjusted to centre the artifact within the time frame.
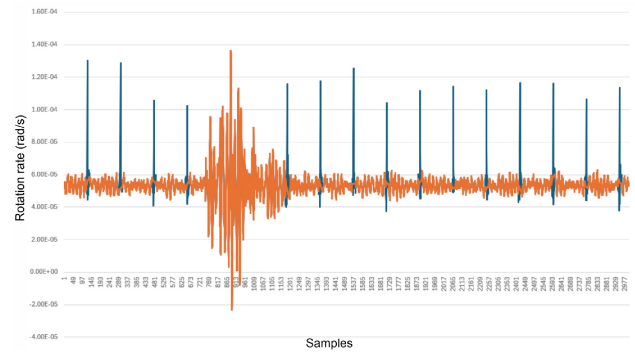


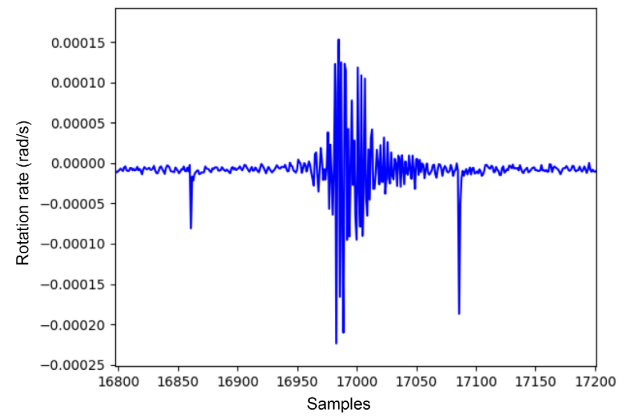**Fig. 5.** Example of correct spike detection with threshold algorithm.



**Fig. 6.** Example of incorrect detection of spike by threshold method.
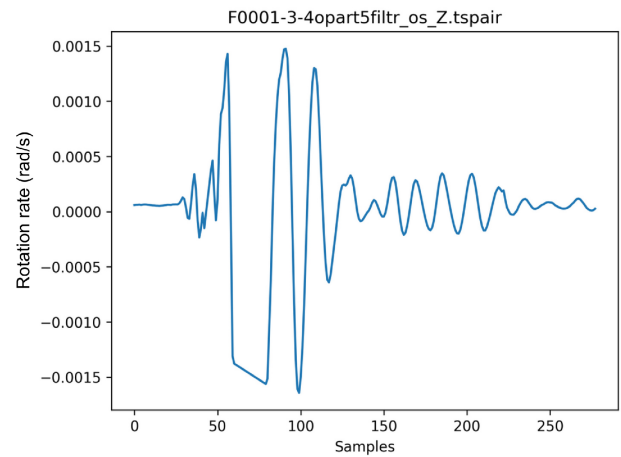


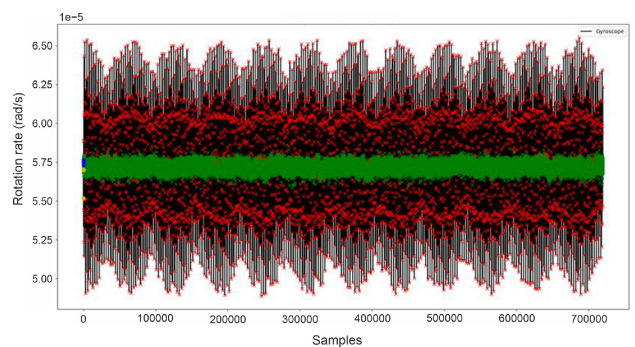**Fig. 7.** Incorrectly removed oscillations mistaken for a spike.



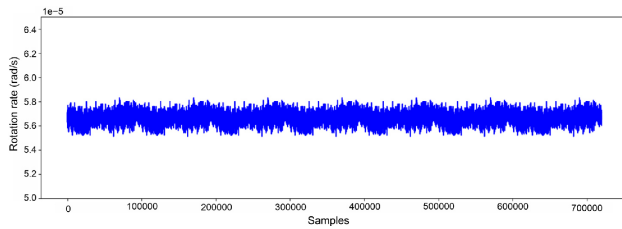**Fig. 8.** FOS5 signal recorded on October 10th, 2021 (only at night).

**Fig. 9.** Example of correct spike detection with threshold algorithm.

It should be emphasised that the algorithm was designed to detect a spike located at the centre of the time window (centre sample). The development of the spike detection algorithm based on AI methods followed the steps outlined below:

1. **Verification of the prepared dataset** – the dataset consisted of 744 1-second samples. For signals containing spikes, each window was extracted to include 0.5 s before the spike, the spike itself, and 0.5 s after the spike.

2. **Preparation of datasets with varying window lengths** – to examine the influence of temporal context on detection performance, four datasets were prepared with different window lengths: 1.001 s (originally extracted windows without modification), 0.501 s (0.25 s before the spike, the spike, and 0.25 s after), 0.251 s. (0.125 s before, spike, 0.125 s after), and 0.101 s (0.05 s before, spike, 0.05 s after). The subsequent steps were carried out independently for each dataset to determine which window length yielded the most accurate spike detection.

3. **Data augmentation** – three augmentation techniques were applied: symmetrical reflection of the waveform around its median, shifting the window one sample to the left and right to generate spike less signals, and adding noise with amplitudes ranging from 5 to 50% of the standard deviation. To balance the dataset, noise was applied only to samples containing spikes.

4. **Spike detection algorithm development** – the research focused on neural network-based approaches combined with statistical parameters. Several models were tested, including multilayer perceptrons (MLPs) using raw waveforms with the spike positioned at the centre of the window, MLPs using extracted statistical descriptors of the signal such as mean, median, standard deviation, extrema, skewness, and kurtosis, recurrent LSTM networks suited for sequential dependencies, and 1D CNNs adapted to capture local patterns. For all

models, hyperparameters such as the number of layers and units, dropout rate, normalisation, kernel size, filters, and batch size were tuned. Among the approaches considered, the second method – MLP with statistical features – yielded the best visual results and was therefore selected for further analysis.

5. **Verification on FOS5 data** – the performance of the algorithms described in point 4 was verified using data from the FOS5 device. Tests were conducted for different window lengths defined in step 2, enabling a direct comparison of detection efficiency under controlled conditions.

6. **Verification on FOS6 data** – the detection algorithm was subsequently evaluated under real-world conditions using the first incoming data from the FOS6 device. Four instruments were buried in the ground, and approximately one hour of operation was recorded. The signals proved to be largely undistorted, resembling the night-time recordings obtained from the FOS5 device, except for several artificially induced short disturbances such as the explosion of a firecracker or the drop of a weight from a height of five meters. For these recordings, both algorithms – the threshold-based and the AI-based – were applied, and the number and position of detected spikes were compared. When the algorithms produced different outcomes, the corresponding waveform segments were manually reviewed and evaluated. The results of this comparison are summarised in Table 1, which is presented in the next section.

7. **Elimination of unwanted spikes** – for each maximum previously identified as an artifact, a linear interpolation procedure was applied to eliminate the unwanted jump from the signal. A fixed area was defined around each detected peak – extending both before and after the maximum value – and replaced with values interpolated between stable parts of the waveform on both sides of the spike. The boundaries of this area were chosen so that the total duration of the removed segment did not exceed 3 ms, which corresponds to the maximum length of a single spike observed in the data. This ensured effective suppression of transient pulses while minimising distortion of adjacent signal components. This method ensured smooth continuity of the reconstructed waveform after spike removal, preserving its overall morphology and amplitude balance. Due to its simplicity and low computational costs, it can be easily implemented in various seismic data acquisition and processing systems.

**Table 1.**
Comparison of the number of spikes detected by the threshold method and the AI method (X channel).

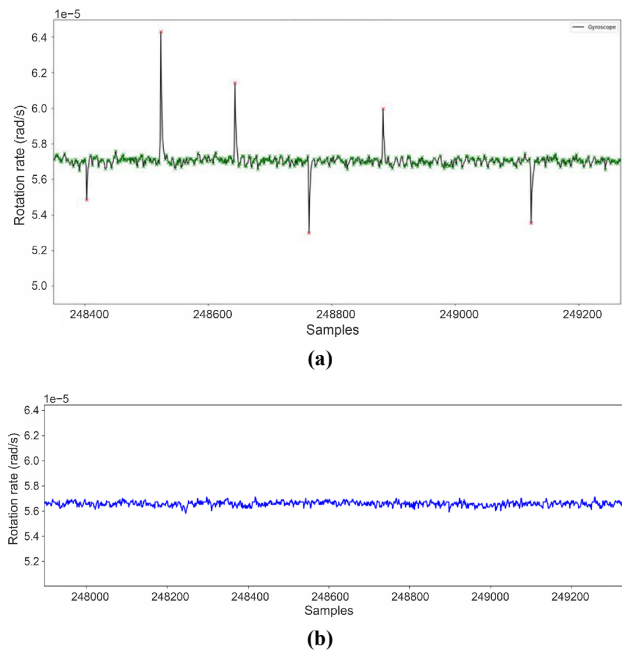| Signal | Real value | Correctly detected by AI method | Incorrectly detected by AI method | Detected by threshold method |
|--------|-----------|-------------------------------|---------------------------------|------------------------------|
| F0004-1-2 | 139 | 134 | 0 | 51 |
| F0004-3-4 | 154 | 145 | 0 | 49 |
| F0004-5-6 | 140 | 133 | 5 | 29 |
| F0004-7-7 | 85 | 81 | 0 | 35 |
| F0004-8-8 | 95 | 92 | 1 | 63 |
| F0004-9-11 | 150 | 143 | 1 | 149 |

**(a)**



**(b)**

**Fig. 10.** Analised signal (10.10.2021) from FOS5: signal (zoom) before processing (a), signal (zoom) filtered using AI detection algorithm and linear interpolation procedure (b).

# 6. Results

The results of the experiments are presented in Table 1 and Figs. 10–12. Table 1 summarises the detection performance of the threshold-based and AI-based algorithms applied to the first data recorded by the FOS6 device. Figures 10–12 illustrate representative examples of the correct spike detection for both FOS5 and FOS6 signals under different window lengths and noise conditions, providing a visual assessment of the effectiveness of the applied methods. A 2-hour experiment was recorded with the FOS5 device.

Figure 10 presents a signal recorded by FOS5. This device is continuously exposed to interference during operation. Figure 10(a) shows a segment of the signal before filtering, where disturbances (spikes) are clearly visible. Figure 10(b) illustrates the same segment after applying AI algorithms for spike detection and linear interpolation. It is worth noting that this dataset was used during the training of the neural network, which explains a high-detection accuracy observed in this case.

The developed algorithm was subsequently validated on test data (not used during training) obtained from a new device, FOS6. The corresponding signals are shown in Figs. 11 and 12. These data originate from tests conducted at a military training ground, where the devices were tasked with recording events, specifically firecracker explosions. Each event was captured by four FOS6 devices positioned at varying distances from the explosion site.

Figures 11(a) and 12(a) present the signals recorded by the FOS6 device located furthest from the explosion (device ID: F0004). Figure 11(a) corresponds to explosions 1 and 2, while Figure 12(a) corresponds to explosion 7. Figures 11(b) and 12(b) show the same signals after applying the detection and interpolation algorithms. As indicated in Table 1 (columns 2 and 3) and demonstrated in Figs. 11(b), 12(b), and 12(c), they provide a zoomed-in view of the explosion itself. Importantly, as shown in
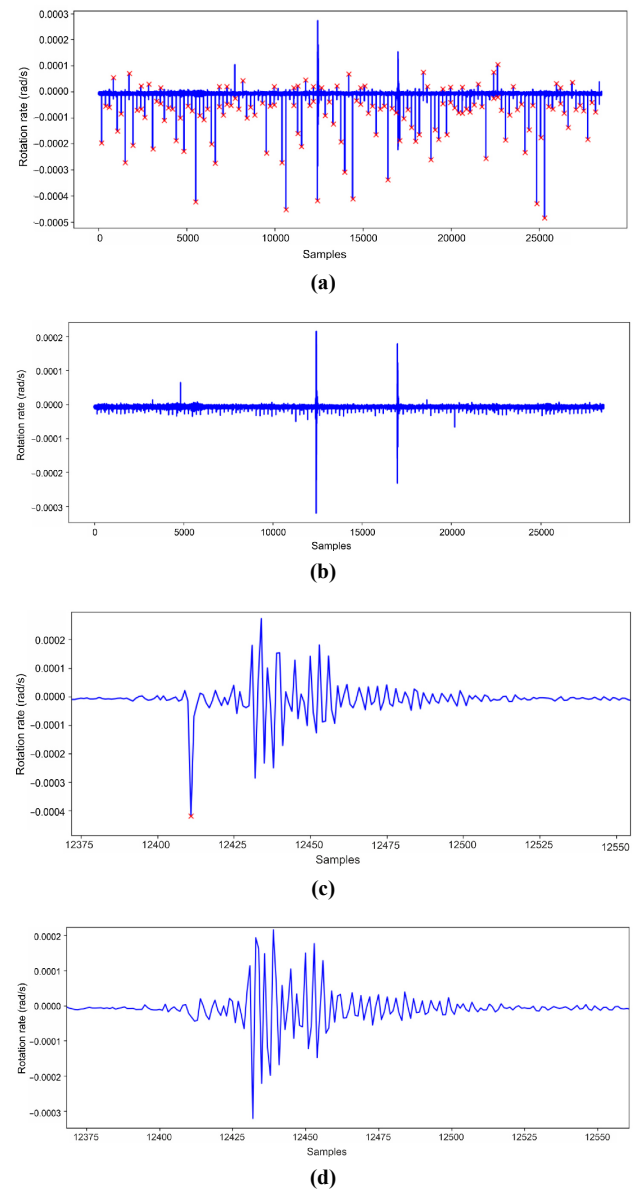


**(a)**



**(b)**



**(c)**



**(d)**

**Fig. 11.** Analised signal F0004-1-2: before processing (a), signal filtered using AI detection algorithm and linear interpolation procedure (b), signal (zoom) (c), signal (zoom) filtered using AI detection algorithm and linear interpolation procedure (d).

Figs. 11(d) and 12(d), and confirmed by Table 1 (column 4), no false detections distorted the segment of interest. This contrasts with the threshold-based algorithm, which either leaves spikes immediately preceding the explosion or introduces distortions due to false detections. In this application, AI proved to be more effective than fixed-threshold approaches.

# 7. Discussion and conclusions

This paper presents a method for detecting spikes in FOS6 seismograph waveforms using the Sagnac effect. The AI based approach is challenging to implement and requires substantial training data. However, it was shown that waveforms recorded with the earlier FOS5 device can be successfully used as training data for evaluating signals from the FOS6 device, yielding entirely satisfactory results. As with any classification method, both type I (non-detection
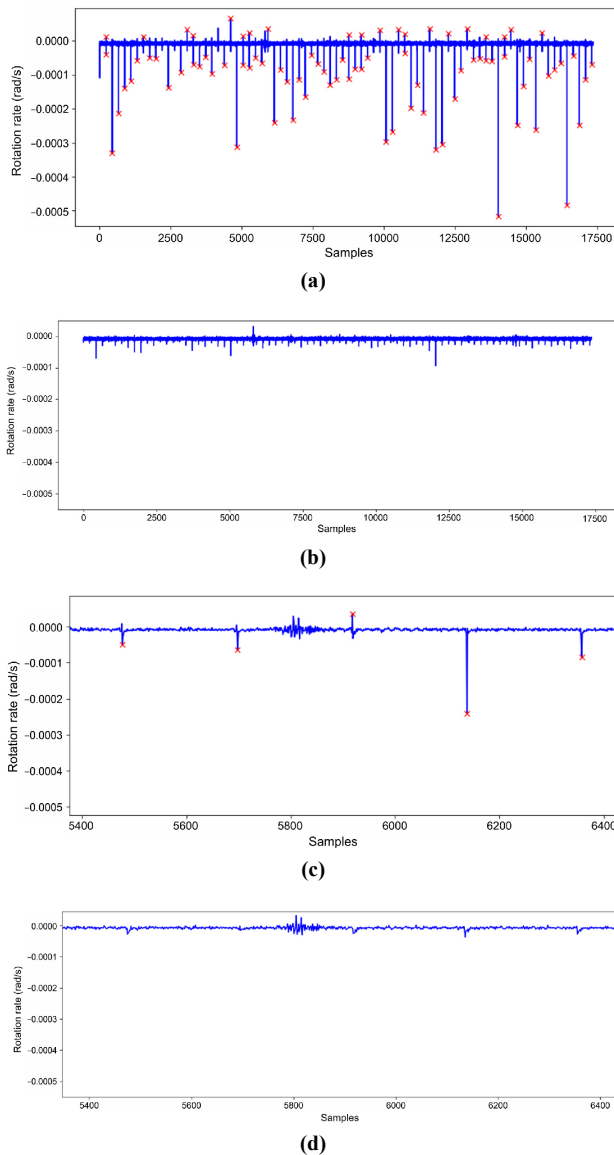
**(a)**



**(b)**



**(c)**



**(d)**

**Fig. 12.** Analised signal F0004-7-7: before processing (a), signal filtered using AI detection algorithm and linear interpolation procedure (b), signal (zoom) (c), signal (zoom) filtered using AI detection algorithm and linear interpolation procedure (d).

of an existing spike) and type II (false spike detection) errors occur; however, their frequency is low, and these errors can be leveraged in subsequent learning processes to improve performance gradually. In this way, neural network-based solutions can refine their classification parameters over time.

By contrast, the threshold-based approach performed significantly worse. Its main limitation lies in the difficulty of adjusting thresholds as signal characteristics change. As a result, errors produced by this method tend to appear at the most undesirable moments, for example, during or immediately before firecracker explosions. While thresholds can be tuned, the process is like pulling on a blanket that is too short: covering one edge case always leaves another exposed. Awareness of errors alone does not enable their elimination without simultaneously introducing new ones. Consequently, threshold-based detection will remain a less flexible and less adaptive tool compared to AI.

The proposed method is a post-processing approach that effectively reduces the impact of artifacts on recorded seismic waveforms. The resulting signal is easier to analyse with threshold-based event detectors, thereby reducing the number of false alarms and improving the overall reliability of the results. An additional advantage of this solution is that it can be applied to archived data that cannot be corrected using hardware-level signal generation methods.

## Authors' statement

## Acknowledgements

## References

[1] Jaroszewicz, A. *et al*. Review of the usefulness of various rotational seismometers with laboratory results of fibre-optic ones tested for engineering applications. *Sensors* **16**, 2161 (2016). https://doi.org/10.3390/s16122161

[2] Arditty, H. J. & Lefèvre, H. C. Sagnac effect in fiber gyroscopes. *Opt. Lett*. **6**, 401–403 (1981). https://doi.org/10.1364/OL.6.000401

[3] Lefèvre, H. C. Principle of the Interferometric Fiber-Optic Gyroscope. in *The Fiber-Optic Gyroscope*, *2nd ed*. (Artech House, 2014).

[4] Michelson, A. A. & Gale, G. G. The effect of the Earth's rotation on the velocity of light. *Nature*, **115**, 566 (1925). https://doi.org/10.1038/115566a0

[5] Jaroszewicz, L. R., Krajewski, Z. & Teisseyre, K. P. Usefulness of AFORS – autonomous fibre-optic rotational seismograph for investigation of rotational phenomena. *J. Seismol*. **16**, 573–586 (2012). https://doi.org/10.1007/s10950-011-9258-3

[6] Kurzych, A., Jaroszewicz, L. R., Krajewski, Z., Teisseyre, K. P. & Kowalski, J. K. Fibre-optic system for monitoring rotational seismic phenomena. *Sensors* **14**, 5459–5469 (2014). https://doi.org/10.3390/s140305459

[7] Kurzych, A. T. & Jaroszewicz, L. R. A review of rotational seismology area of interest from a recording and rotational sensors point of view. *Sensors* **24**, 7003 (2024). https://doi.org/10.3390/s24217003

[8] Kurzych, A. *et al*. Fibre-optic Sagnac interferometer in a FOG minimum configuration as instrumental challenge for rotational seismology. *J. Light. Technol*. **36**, 879–884 (2018). https://doi.org/10.1109/JLT.2017.2769136

[9] Kurzych, A., Kowalski, J. K., Sakowicz, B., Krajewski, Z. & Jaroszewicz, L. R. The laboratory investigation of the innovative sensor for torsional effects in engineering structures' monitoring," *Opto-Electron. Rev*. **24**, 134–143 (2016). https://doi.org/10.1515/oere-2016-0017

[10] Kurzych, A. T. *et al*. Two correlated interferometric optical fiber systems applied to the mining activity recordings. *J. Light. Technol*. **37**, 4851–4857 (2019). https://doi.org/10.1109/JLT.2019.2923853

[11] Bernauer, F. *et al*. Rotation, strain and translation sensors performance tests with active seismic sources. *Sensors* **21**, 264 (2021). https://doi.org/10.3390/s21010264

[12] Kurzych, A. T., Jaroszewicz, L. R., Kowalski, J. K. & Sakowicz, B. Investigation of rotational motion in a reinforced concrete frame construction by a fiber optic gyroscope. *Opto-Electron. Rev*. **28**, 69–73 (2020). https://doi.org/10.24425/opelre.2020.132503

[13] Jaroszewicz, L. R., Dudek, M., Kurzych, A. T. & Teisseyre, K. P. A test performance of fibre-optic sensors for real-time investigations of rotational seismic events: A case study in laboratory and field conditions. *Opto-Electron. Rev*. **29**, 213–219 (2021). https://doi.org/10.24425/opelre.2021.140102

[14] LeFèvre, H. C. *The Fiber-Optic Gyroscope, 3rd ed*. (Artech House, 2022).

[15] Zając, P. *et al*. Self-noise reduction in a FOG-based rotational seismometer confirmed by Allan variance analysis. *Opto-Electron. Rev*. **32**, e152766 (2024). https://doi.org/10.24425/opelre.2024.152766

[16] Murray-Bergquist, L., Bernauer, F. & Igel, H. Characterization of six-degree-of-freedom sensors for building health monitoring. *Sensors* **21**, 3732 (2021). https://doi.org/10.3390/s21113732

[17] Chen, Y. *et al*. TXED: The Texas earthquake dataset for AI. *Seismol. Res. Lett*. **95**, 2013–2022 (2024). https://doi.org/10.1785/0220230327

[18] Havskov, J. & Alguacil, G. *Instrumentation in Earthquake Seismology, 2nd ed*. (Springer, 2016). https://doi.org/10.1007/978-3-319-21314-9

[19] Bernauer F. Blueseis–Sandbox (as of June 6th, 2019). (Accessed: Oct. 4th, 2022). [Online]. https://github.com/fbernauer/blueseis_sandbox

[20] Yu, P., Zhu, T., Marone, C., Elsworth, D. & Yu, M. DASEventNet: AI-based microseismic detection on distributed acoustic sensing data from the Utah FORGE Well 16A (78)-32 hydraulic stimulation. *J. Geophys. Res. Solid Earth* **129**, e2024JB029102 (2024). https://doi.org/10.1029/2024jb029102

[21] Fan, Q. *et al*. A multi-domain feature fusion epilepsy seizure detection method based on spike matching and PLV functional

networks. *J. Neural Eng*. **22**, 016025 (2025). https://doi.org/10.1088/1741-2552/adaef3

[22] Zhang, Z., Yang, H., Wang, Y., Zhang, L. & Lin, S.-H. QuanFormer: A transformer-based precise peak detection and quantification tool in LC-MS-based metabolomics. *Anal. Chem*. **97**, 2698–2706 (2025). https://doi.org/10.1021/acs.analchem.4c04531

[23] Corsaro, M. *et al*. Seismic Phases Picking with Artificial Intelligence: A Novel Approach for Distributed Acoustic Sensing Data Analysis. in *Galileo Conference: Fibre Optic Sensing in Geosciences* (2024). https://doi.org/10.5194/egusphere-gc12-fibreoptic-82

[24] Diefenthaler, M. *et al*. AI-assisted detector design for the EIC (AID(2)E). *J. Instrum*. **19**, C07001 (2024). https://10.1088/1748-0221/19/07/C07001

[25] Lomax, A. *et al*. Effects on a deep-learning, seismic arrival-time picker of domain-knowledge based preprocessing of input seismograms. *Seismica* **3**, 1–23 (2024). https://seismica.library.mcgill.ca/article/view/1164/1492

[26] Balcan, M.-F. *et al*. How Much Data Is Sufficient to Learn High-Performing Algorithms? Generalization Guarantees for Data-Driven Algorithm Design. in *STOC'21 Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* 919–932 (STOC'21, 2021). https://doi.org/10.1145/3406325.3451036

[27] Neyshabur, B., Bhojanapalli, S., McAllester, D. & Srebro, N. Exploring Generalization in Deep Learning. in *31st Conference on Neural Information Processing Systems* 1–10 (TTIC, 2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/10ce03a 1ed01077e3e289f3e53c72813-Paper.pdf

[28] Frénay, B. & Verleysen, M. Classification in the presence of label noise: A survey. *IEEE Trans. Neural Netw. Learn. Syst*. **25**, 845–869 (2014). https://doi.org/10.1109/TNNLS.2013.2292894

[29] Harari, O., Bingham, D., Dean, A. & Higdon, D. Computer experiments: Prediction accuracy, sample size and model complexity revisited. *Stat. Sin*. **28**, 899–919 (2018). https://www3.stat.sinica.edu.tw/statistica/oldpdf/A28n216.pdf