

Received July 3, 2020, accepted July 7, 2020, date of publication July 15, 2020, date of current version October 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009524

Monitoring of Fibre Optic Links With a Machine Learning-Assisted Low-Cost Polarimeter

MARTIN ŠLAPÁK¹, JOSEF VOJTĚCH¹, (Senior Member, IEEE),
ONDŘEJ HAVLÍŠ¹, (Member, IEEE), AND RADAN SLAVÍK^{1,2}, (Senior Member, IEEE)

¹Department of Optical Networks, CESNET, 16000 Prague, Czech Republic

²Optoelectronics Research Centre, University of Southampton, Southampton SO17 1BJ, U.K.

Corresponding author: Martin Šlapák (slapak@cesnet.cz)

This work was supported in part by the E-Infrastructure CESNET–Modernization Project under Grant CZ.02.1.01/0.0/0.0/16_013/0001797, and in part by the Ministry of Interior under Grant VI20152020045.

ABSTRACT The optical fibres widely used in telecommunication can be simultaneously used for (distributed) sensing or fibre network self-monitoring. In our work, we monitor changes in the fibre environment via monitoring changes in the state of light polarization without the utilization of methods based on back-scattered light. These changes can generate a vast amount of data, but it is generally not straightforward to extract useful information from them, e.g., future fibre break predictions or earthquake monitoring. We suggest using machine learning to solve this problem. However, since the measured data events are not labelled (i.e., we do not know in advance what fingerprint in the measured data corresponds to a future fibre break), unsupervised machine learning methods must be used. Here, we report a proof-of-concept approach in which we use a simple polarimetric technique and installed optical fibre, which we disturb with controlled vibrations, knocking on the fibre, and rack door closing near the fibre. Using a machine learning K-means algorithm, we distinguish between data generated with these controlled disturbances and data generated by noise due to common traffic. These results are the first step along the way to automated data labelling, which can be used for the classification of events.

INDEX TERMS Optical fibre sensors, event detection, K-means, machine learning, polarimeter.

I. INTRODUCTION

Intelligent fibre optic systems need to provide (among other benefits) self-monitoring, which, for example, includes early detection of future fibre breaks or detection of an intrusion. To perform the mentioned intentions in as large a portion of the fibre network as possible, these smart functionalities should be provided with hardware of minimal cost.

The above effects can be monitored via measuring changes in the properties of the light (phase, amplitude, or polarization) that propagates through the fibres, which are measured at the receiver side [1] or in the back-reflection (e.g., [2], [3]). The system operating in transmission has the advantage of simplicity (often not requiring any additional components [1]), but generally cannot provide spatially-resolved detection (with some exceptions, e.g., the relatively complex dual-wavelength method presented in [4]). Detection systems operating via reflection (e.g., optical time-domain

reflectometry, OTDR [5] or phase-resolved optical time-domain reflectometry ϕ -OTDR [6]) can achieve spatially-resolved detection, but generally require additional optical hardware and must cope with the detection of extremely low levels of the back-scattered signals. The additional hardware may significantly increase the cost of the system, e.g., ϕ -OTDR requires a dedicated, relatively high-power continuous-wave (CW) light source with significant demands on its coherence length, which must be at least twice the fibre optical length. Such a CW light source is not only expensive: its high power can also easily cause undesirable cross-talk when the fibre is simultaneously used for data signals.

For practical use, any measurement technique sensitive to fibre manipulation needs to be complemented with data analysis that identifies events of interest, e.g., ‘tapping on the fibre’ or ‘slow degradation of the fibre’. Such a technique must function effectively even in the presence of ‘noise’, e.g., events that are not of interest (a person walking down the corridor, playing music or talking, a truck passing on the road, etc.).

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues .

Methods of machine learning have recently emerged to address the classification of measured data; however, they have been almost-uniquely applied to back-scatter-based systems (based on OTDR or ϕ -OTDR) [7]–[9]. Tejedor *et al.* [7] provide an overview of the application of Distributed Acoustic Sensing (DAS) and pattern recognition to detect potentially dangerous events on gas pipelines. Makarenko [8] and Bublin [9] both utilize deep learning techniques for data acquired by DAS during perimeter monitoring on pipelines. Machine learning applied to simple systems operating in transmission has not been, to the best of our knowledge, reported yet. For the classification, the model must be learned from data which are already labelled. The most challenging question is how to obtain such labels for measured data.

Here, we suggest performing automatic recognition of the fibre changes of interest (events) from data acquired with our cost-effective transmission-based polarization monitoring method [3] using unsupervised machine learning. To achieve this, we demonstrate the first necessary step experimentally, which is grouping (clustering) data that are deemed to correspond to the same/similar event. We suggest and use the unsupervised machine learning algorithm K-means [10] for this task.

Although our measurement method [3] requires additional hardware on the receiver side (a low-cost polarization analyzer, which we describe later), it can be used in conjunction with the extremely inexpensive transmission of amplitude-modulated data. Thus, the overall system cost can be lower than for techniques requiring a coherent receiver [1], where the cost of optical hardware to transmit and detect the data on its own is significantly higher than for simple amplitude-modulated transmission.

Our results represent the first step towards the future automated identification of various events from the measured data. Achieving this final goal will require training a classification model for categorization of events (i.e., to evaluate which footprint data correspond to ‘tapping on the fibre’ and which group of data corresponds to ‘slow fibre degradation’, etc.). This will allow for the final data classification of ‘tapping on the fibre’ or ‘slow fibre degradation’ directly from the newly measured yet unseen data.

The structure of this paper is as follows: section II describes the experimental setup and provides definitions of the experiments. The following section III aims to provide an accurate explanation of the data processing pipeline and the utilized algorithms. Section IV summarizes the results and explains the estimations of the parameters for all testing scenarios. The performance and evaluation of results are discussed in section V, where possible future work is also mentioned. All is concluded in the final section VI.

II. EXPERIMENTAL SETUP

We used an optical fibre polarimeter, which consists of four polarization-sensitive fibre gratings inscribed along a short piece of optical fibre. The light passing through the gratings is scattered out with the scattering intensity

polarization-dependent. The four projections of the input State Of Polarization (SOP) onto four linearly-polarized signals with the polarization axis projection defined by the polarization axis of the four gratings [11] is then detected by four InP photodiodes. Subsequently, the signals from photodiodes are amplified with transimpedance amplifiers and digitized using analogue-to-digital converters; for technical details, see [11], [12].

Our experimental testbed, Fig. 1, consists of two laboratories (‘A’ and ‘B’) in different buildings which are approximately 1 km apart from each other. The installed G.652 single-mode fibre that connects these two laboratories is installed underground. Within the two buildings, it passes through several rooms with switches in which the fibre is connected. This setup thus represents a real-world scenario, especially when compared to other reports that address pure in-laboratory conditions, e.g., [13], [14].

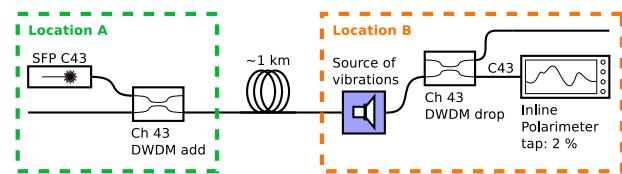


FIGURE 1. Testbed schematics depict SFP C43 (Small Form-factor Pluggable transceiver at $\lambda = 1542.94$ nm) as a light source and inline polarimeter with 2% tap as measurement endpoint.

In laboratory A, a data signal is sent down the Fibre Under Test (FUT) using a Small Form-factor Pluggable (SFP) transceiver operating at 1542.94 nm (ITU channel 43). It is detected in laboratory B with the polarimeter that samples the signal at a fixed frequency of 20 kHz, and data are recorded continuously with a computer.

During the measurement, we introduce controlled disturbances as follows:

1) A loudspeaker creating pulses of discrete frequencies (PDF): we glued a small section of the FUT in a 2 mm diameter protective buffer directly onto a diaphragm of a loudspeaker. We introduced pulses of discrete frequencies of 10, 20, 30, 40, 50, 75, 100, 200, and 500 Hz interleaved by 3-s long silences. After that, we introduced 5-s long ‘heartbeat’ signals that consisted of repeated pairs of two signals with 100% and 20% of loudspeaker amplitude, respectively. The scenario PDF contains ten types of different events in total. The overall length of the PDF sequence was 95 s.

2) Periodic closing and opening of a rack door (RACK): this test represents events that occur in actual operation. It consisted of tapping on the FUT two times, followed by closing and opening the door of the rack in which the fibre was installed. The open-close event was performed five times during the acquisition interval, which was approximately 44 s long.

The experimental data were then visualized by time-frequency 2-D maps (known as spectrograms). For each time point (x-axis), the frequency response (y-axis) was calculated

from the time domain data acquired over a short time interval using Fast Fourier Transform (FFT) [15].

The FUT, which is partially buried underground, primarily detects frequencies below 1 kHz. This is because higher frequencies are strongly attenuated by most materials, including soil [16]. We have found that most FUT detection occurs for frequencies below 200 Hz.

Due to the fixed polarimeter sampling frequency, the resolution of the time-frequency map is limited. Better frequency resolution (larger FFT window) reduces time resolution, and vice-versa. Time resolution that is required for data acquisition depends on the duration of the disturbance events we are monitoring. In our experiment, we chose the FFT window size of 12,288 samples, which led to a relatively low time resolution of ~ 0.625 s (1 column of the spectrogram equals 0.625 s), which, however, was sufficient to distinguish between two consecutive events. This choice provides excellent frequency resolution, which should allow for effective distinction between different FUT disturbances, which is the key interest in our study.

III. DATA PROCESSING

In this section, we start with a brief description of the critical components of the K-means algorithm. Next, the data processing pipeline will be explained.

A. K-MEANS ALGORITHM

The K-means method allows division of measured data (events) into K clusters, with each cluster containing experimental observations that are deemed to correspond to the same event (in our case, fibre disturbance). Finding parameter K is not straightforward and is generally performed experimentally or with the help of some metric measuring the K-means performance. We discuss our approach for K estimation later.

The K-means algorithm works iteratively: see Fig. 2. In the initialization step, we place one measured data sample into each cluster. In the first iterative step, K-means sorts the rest of the measured data events into the K clusters based on their similarity, which is calculated via the least-squared Euclidean distance (mean) between the measured data event and the cluster's centroid. In the second iteration step, K-means calculates new centroids of each cluster. Consequently, the centroid represents all measured data events that are in the cluster. The iterative process continues until there are no more re-assignments of measured data events to the clusters between the iterative steps. This process assigns events of a similar nature to the same cluster. In practice, we utilized K-means implemented with Python's *scikit.learn* package.

B. DATA PROCESSING PIPELINE

The data processing pipeline schematic is depicted in Fig. 3. It starts with an acquired signal, which is subsequently transformed into a spectrogram using FFT. Firstly, we normalize the values of the spectral content to be within $[0, 1]$

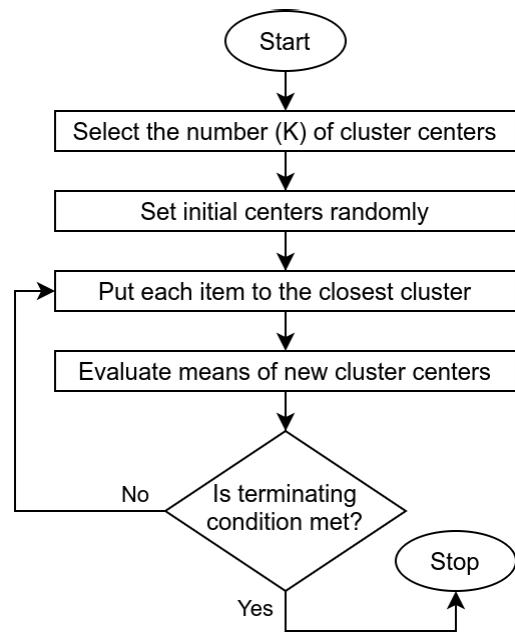


FIGURE 2. The schematic diagram of the K-means algorithm.

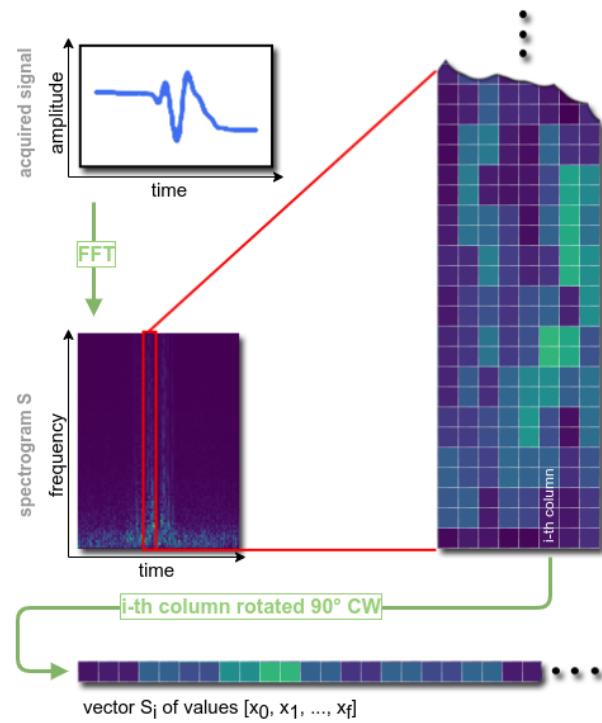


FIGURE 3. The overview of the data processing pipeline. The acquired signal is first transformed using FFT from time to frequency domain. Next, the whole spectrogram is normalized, and the columns are used as input vectors for the K-means algorithm.

throughout the entire spectrogram. We then describe the normalized spectrogram by a matrix S , in which the spectral content of the i -th measurement sample is in the i -th column of S , denoted as S_i .

Subsequently, we take all columns S_i as input vectors for the K-means algorithm. The initial K-means clusterization result showed that most of the measured events (in the *RACK* scenario) occurred within a single time slot of 0.625 s (corresponding to a single column in S). However, few measured events were longer (in particular, from the *PDF* test), spreading over several columns in the spectrogram. To account for them as a single measurement event when processing them with K-means, we pre-processed them with the *unwrap* method.

This method takes the u consequent data lines and then unwraps them to the side. We demonstrate this with the following example:

```

A = [[0 0]
      [1 1]
      [2 2]
      [3 3]
      [4 4]]
B = unwrap(A, 3)
B:
[[ 0.   0.   1.   1.   2.   2.]
 [ 1.   1.   2.   2.   3.   3.]
 [ 2.   2.   3.   3.   4.   4.]]

```

After the application of *unwrap* to the transposed spectrogram, we obtain new matrix S^T , which we transposed back. Subsequently, the input vector is defined as $V_i = \text{unwrap}(S^T)^T{}_i$. For example, $V_0 = [S_0, \dots, S_4]$ for $u = 5$.

We used *unwrap* for data events during the *PDF* test. From the experimental data, we estimated that each event lasted for approximately five samples (5×0.625 s), which is the value we used for *unwrap*.

As we have mentioned earlier, establishing how many clusters K should be used for the K-means method is non-trivial. Sometimes, this is known, e.g., when we know how many types of events we expect. However, when not known, we can, for example, use learning algorithms applied to the measured data or data extracted from them ('meta-data'). This approach is commonly referred to as 'meta-learning'. Meta-learning requires a way to evaluate the K-means clustering performance, i.e., to evaluate how similar events within the same cluster are, and how different they are in different clusters. This is done either by a human or by using metrics (or an index) that quantifies the clustering performance. There are many indices to perform this task, e.g., Silhouette index [17], Davies-Bouldin [18], Calinski-Harabasz [19], Dunn index [20], and R-squared index [21]. Establishing which one is the most applicable to our data is beyond the scope of this paper. Later, we will discuss how to assess their suitability using an example of the Silhouette index.

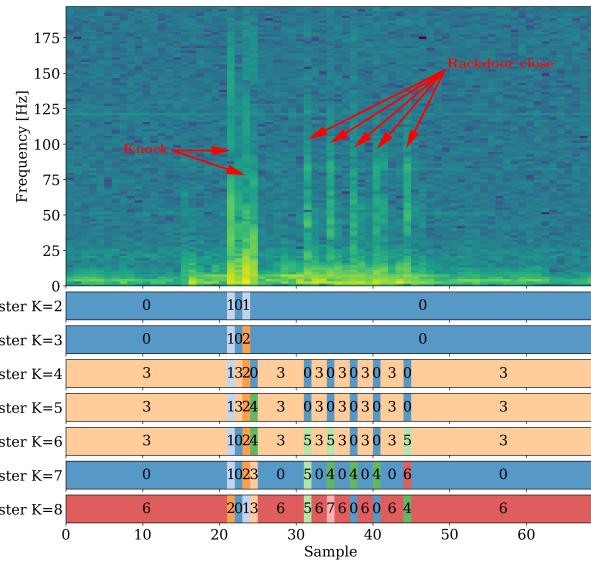
IV. RESULTS

Table 1 summarizes the results for both use-cases, *RACK* and *PDF*. We can see that the success rate in the *RACK* scenario for $K = 5$ is 100 % for event identification. Still, only 6 of 7 events are assigned into the correct cluster, and at the same

TABLE 1. Result summary of both use-cases, *RACK* and *PDF*. The lowest acceptable value of K' row includes such K values that at least 50 % of different events are distinguished from background. The last two rows are evaluated with respect to the lowest acceptable K .

	use-case	
	PDF	RACK
# of events	10	7
# of different event types	10	2
the lowest acceptable K	8	5
max. correctly identified events	5 (50 %)	7 (100 %)
cluster homogeneity	40%	85.7%

time, each type of event has only one assigned cluster (this is referred to as homogeneity). Specifically, the 2nd knock is misplaced into cluster number 2 instead of the same cluster as the 1st knock (cluster number 1). In the *PDF* use-case (for $K = 8$), the best event identification achieved 50% success, and cluster homogeneity was achieved for 4 of 10 events.



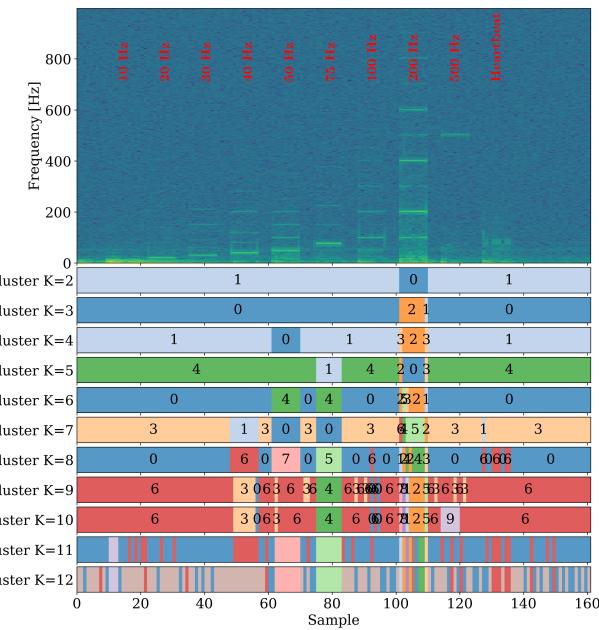


FIGURE 5. Measured and normalized spectrogram of PDF test with corresponding K-means clusterization for different $K = [2, 12]$. One sample is equivalent to 0.625 s, and the total time duration is 100 s. For a clearer view, the last two rows of clustering results lack the number of clusters.

allowed the identification of all expected events (background, knocking, and rack closing) is $K = 5$: five door closing events (cluster 0), knocking (clusters 1 and 2), and background (cluster 3). However, cluster 4, which follows the 2nd knock, is erroneously not evaluated as background (cluster 3).

In the *PDF* test, Fig. 5, we can see that the polarimeter is capable of capturing the frequency of the disturbance (as we increased it from 10 through 20, 30, 40, 50, 75, 100, 200, and 500 Hz), including its harmonics. Thus, we expect that K-means should recognize all of these events. Since we expect nine different frequencies, heartbeat knocking, and background, K should be 11.

For all K values from 2 to 10, K-means did not distinguish between background and footprints of 10, 20, and 30 Hz frequencies. Later, for $K = 11$ and $K = 12$, there is evidence of excessive granularity in ‘silence’ segments, and only 40, 50, and 75 Hz frequencies were identified in their full length. The two results ($K = 9$ and $K = 10$) are very similar. The results for $K = 10$ identified one more cluster (number 9 for 500 Hz frequency) than the results for $K = 9$, where there are also traces of excessively extensive granularity (short intervals with cluster number 3).

Here, we decided to select $K = 8$ as a trade-off between the number (5) of correctly identified clusters for frequencies and clear and homogeneous identification of background. The footprints of 40, 50, 75, and 200 Hz (cluster numbers 6, 7, 5, and 4) were correctly identified. The result for $K = 7$ is similar to the results for $K = 8$ (perhaps slightly better in the case of the 200 Hz footprint), but it failed in the identification of ‘heartbeat’. In our opinion, it is better to assign ‘heartbeat’

into the same cluster as 40 Hz rather than into the cluster of background (which is what happened for $K = 7$). For $K = 8$, in Fig. 5, we see separated clusters of blocks corresponding to frequencies (40, 50, 75, and 200 Hz), as well as ‘heartbeat’ knocking. Unfortunately, lower frequencies of 10, 20, and 30 Hz were not identified as separate clusters. However, we obtained good consistency in background noise identification for one type of footprint (cluster 0).

V. DISCUSSION

In Section IV, we have found that $K = 5$ captured all of the events in the *RACK* test, although we expected to need only $K = 3$. As we have mentioned earlier, the task of finding the optimum K , which we performed manually, should be performed automatically in the future with the help of metrics or an index. We mentioned that there is a large number of such metrics or indices. However, as we show below with the example of the Silhouette index, a careful approach must be adopted when using these metrics or indices.

A. SILHOUETTE INDEX

The Silhouette index ranges from -1 to $+1$, where a high value indicates that events are well matched to their cluster and poorly matched to neighbouring clusters.

TABLE 2. Silhouette index for *RACK* and *PDF* tests for various K .

Silhouette index		
K	<i>RACK</i>	<i>PDF</i>
2	0.835	0.470
3	0.807	0.447
4	0.642	0.112
5	0.623	0.109
6	0.526	0.106
7	0.585	0.086
8	0.526	0.084
9		0.046
10		0.071
11		0.034
12		0.012

Table 2 shows Silhouette index values for both experiments: *RACK* and *PDF*. For *RACK*, the highest Silhouette index is achieved for $K = 2$. As shown in Fig. 4, with $K = 2$, K-means recognized only one of the two events: it recognized all knocks, but no closings of the rack. Thus, a metrics or index other than Silhouette must be used to recognize both events of interest.

For the *PDF* test, we expected 11 clusters (9 frequencies, background, and ‘heartbeat’), but similarly to the *RACK* test, the Silhouette index is the highest for $K = 2$. As shown in Table 2, there is a local maximum of the Silhouette index at $K = 10$, which is slightly lower than the expected $K = 11$. We speculate that the footprint of the 10 Hz signal is too similar to the ‘background’. Analyzing local maxima of the Silhouette index gives a K value close to what we would expect, showing the usefulness of this index for the *PDF* test.

B. FUTURE WORK

Once the properties of clusters are established, we would assign labels to them (e.g., knowing that cluster number 3 corresponds to events describing opening a rack with fibres, we could label this cluster as ‘Rack opening’). The output pairs (input vector-label) will next be used as input data for training of machine learning classification algorithms like Support Vector Machines, Artificial Neural Networks or K-Nearest Neighbours (which would later place unseen data events into different classes).

Following that, we could provide automated identification of measured events with a classification algorithm. For practical application of classifier training, we will need at least hundreds to thousands of labeled data. It can be expected that the footprints of a given event will not be entirely identical because of differences in the background noise, environment and conditions at the specific time of occurrence. Thus, K-means requires numerous input vectors (repeated measurements of the same event) to converge a cluster mean to a stable and reliable value. In practice, the selected optical link will be measured for a reasonably long period, and the K-means will process the acquired data.

Our initial experiments shown here provided us with data corresponding to limited types of events. The main advantage of the machine learning approach is that a trained classifier is capable of rapidly evaluating new data and assigning them into categories learned from training data, enabling real-world events to be accommodated even with the in-lab developed system.

VI. CONCLUSION

Our approach demonstrates the suitability of unsupervised machine learning methods like K-means to sort measured data with event footprints into a defined number of categories (clusters). Subsequently, these categories can be used as target labels of data instances for training a classification model.

The key task is to find the optimum number of categories, especially when adopting methods of machine learning. We demonstrate this using the Silhouette index, which requires careful analysis to allow for estimation of this number.

To make this practical, it will be necessary to carry out further steps, including the evaluation of events with respect to the knowledge of the original event occurrences or with learned classifiers.

REFERENCES

- [1] J. E. Simsarian and P. J. Winzer, “Shake before break: Per-span fiber sensing with in-line polarization monitoring,” in *Proc. Opt. Fiber Commun. Conf.*, Washington, DC, USA: Optical Society of America, Mar. 2017, Paper M2E.6. [Online]. Available: <https://www.osapublishing.org/abstract.cfm?uri=OFC-2017-M2E.6>
- [2] C. Kirkendall, “Distributed acoustic and seismic sensing,” in *Proc. OFC/NFOEC Conf. Opt. Fiber Commun. Nat. Fiber Optic Engineers Conf.*, Mar. 2007.
- [3] M. Slapák, J. Vojtěch, and P. Munster, “Polarization changes as early warning system in optical fiber networks,” in *Proc. 42nd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2019, pp. 597–600.

- [4] O. B. Wright, “Stabilized dual-wavelength fiber-optic interferometer for vibration measurement,” *Opt. Lett.*, vol. 16, no. 1, pp. 56–58, Jan. 1991. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-16-1-56>
- [5] M. K. Barnoski, M. D. Rourke, S. M. Jensen, and R. T. Melville, “Optical time domain reflectometer,” *Appl. Opt.*, vol. 16, no. 9, pp. 2375–2379, Mar. 1977.
- [6] R. Sifta, P. Munster, P. Sysel, T. Horvath, V. Novotny, O. Krajsa, and M. Filka, “Distributed fiber-optic sensor for detection and localization of acoustic vibrations,” *Metrology Meas. Syst.*, vol. 22, no. 1, pp. 111–118, Mar. 2015.
- [7] J. Tejedor, J. Macias-Guarasa, H. Martins, J. Pastor-Graells, P. Corredera, and S. Martin-Lopez, “Machine learning methods for pipeline surveillance systems based on distributed acoustic sensing: A review,” *Appl. Sci.*, vol. 7, no. 8, p. 841, Aug. 2017.
- [8] A. V. Makarenko, “Deep learning algorithms for signal recognition in long perimeter monitoring distributed fiber optic sensors,” in *Proc. IEEE 26th Int. Workshop Mach. Learn. for Signal Process. (MLSP)*, Sep. 2016, pp. 1–6.
- [9] M. Bublin, “Machine learning for distributed acoustic sensors, classic versus image and deep neural networks approach,” 2019, *arXiv:1904.11546*. [Online]. Available: <http://arxiv.org/abs/1904.11546>
- [10] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [11] V. Mikhailov, S. Dunn, and P. S. Westbrook, “Robust remote calibration of fiber polarimeters,” in *Proc. Opt. Fiber Commun. Conf./Nat. Fiber Optic Eng. Conf.*, 2011, pp. 1–3. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/5875651>
- [12] V. Mikhailov, S. Dunn, and P. S. Westbrook, “In-line, high speed fibre polarimeter with large calibration bandwidth and accurate reference-free calibration procedure,” in *Proc. 36th Eur. Conf. Exhib. Opt. Commun.*, Sep. 2010, pp. 1–3.
- [13] Z. Wang, J. Zeng, J. Li, M. Fan, H. Wu, F. Peng, L. Zhang, Y. Zhou, and Y. Rao, “Ultra-long phase-sensitive OTDR with hybrid distributed amplification,” *Opt. Lett.*, vol. 39, no. 20, pp. 5866–5869, Oct. 2014.
- [14] L. D. van Putten, A. Masoudi, and G. Brambilla, “100-km-sensing-range single-ended distributed vibration sensor based on remotely pumped erbium-doped fiber amplifier,” *Opt. Lett.*, vol. 44, no. 24, pp. 5925–5928, 2019.
- [15] E. O. Brigham and R. E. Morrow, “The fast Fourier transform,” *IEEE Spectr.*, vol. 4, no. 12, pp. 63–70, Dec. 1967.
- [16] M. L. Oelze, W. D. O’Brien, and R. G. Darmody, “Measurement of attenuation and speed of sound in soils,” *Soil Sci. Soc. Amer. J.*, vol. 66, no. 3, pp. 788–796, May 2002.
- [17] K. Wang, B. Wang, and L. Peng, “CVAP: Validation for cluster analyses,” *Data Sci. J.*, vol. 8, pp. 88–93, 2009, Art. no. 0904220071.
- [18] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
- [19] T. Calinski and J. Harabasz, “A dendrite method for cluster analysis,” *Commun. Statist. - Simul. Comput.*, vol. 3, no. 1, pp. 1–27, Jan. 1974.
- [20] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973.
- [21] S. Sharma and S. Sharma, *Applied Multivariate Techniques*. New York, NY, USA: Wiley, 1996.



MARTIN ŠLAPÁK received the M.Sc. degree in software engineering and the Ph.D. degree in informatics from Czech Technical University, Prague, in 2010 and 2019, respectively. He joined the Research and Development Department, CESNET, Prague, in 2011, where he is currently with the Research and Software Development in czech research and educational network. He has participated with the EU projects GN3, GN3plus, FI-PPP Xifi, CLONETS, and TIFOON. His research interests include application of optical fibres as distributed sensor systems, signal processing, and machine learning.



JOSEF VOJTEČH (Senior Member, IEEE) received the B.Sc. degree in pedagogy, the M.Sc. degree (Hons.) in computer science, and the Ph.D. degree in all-optical networking from Czech Technical University, Prague, in 2001, 2003, and 2009, respectively. He has been active in the field of photonics networks, since 2003. He was responsible for the Development of a Family of Open Photonic Devices. He currently leads the Research Department of Optical Networks, CESNET. He also leads a Contractual Research Project for delivery of ultrastable coherent optical frequency for sensing of nuclear power plant containment stability. He has been co-organized the Customer Empowered Fibre Networks Workshop, since 2004. He has been with the Special Section on Photonic Networks and their Services, Conference on Telecommunications and Signal Processing, since 2016. He holds over 13 patents (including five U.S. patents) and multiple utility models. His record includes more than 120 scientific publications. He has participated with international projects, such as CLONETS, COMPLETE, FI-PPP XIFI, GN4, GN3+, GN3, GN2, Porta Optica Study, and SEEFIRE. He is a Senior Member of OSA and SPIE. He is also a member of ION. He received the Research Prize from the Czech Minister of Education, Miroslav Karásek and Jan Radil, in 2007.



ONDŘEJ HAVLÍŠ (Member, IEEE) received the master's degree in communications and informatics science from the Brno University of Technology, in 2012, and the Ph.D. degree from the Faculty of Electrical Engineering and Communication Brno University of Technology. He joined the Optical Networks Department, CESNET, in 2012. He was actively involved with the projects GN3 and GN3plus. He participates with the project DOBI. He holds two patents and several

utility models. His current research interests include research, design, and verification of photonic services in optical networks, transmission of ultra-stable quantities (T/F), development of photonics testbeds, and open line system for czech light (SDN DWDM).



RADAN SLAVÍK (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in optics and optoelectronics from the Faculty of Mathematics and Physics, Charles University, Prague, Czechia, in 1996 and 2000, respectively, and the D.Sc. degree from the Czech Academy of Sciences, Prague, Czechia, in 2009. From 1995 to 2000, and from 2004 to 2009, he was with the Institute of Photonics and Electronics, Czech Academy of Sciences. From 2000 to 2003, he was with the Centre d'Optique, Photonique et Laser, Université Laval, Quebec, QC, Canada, as a Postdoctoral Research Fellow. Since 2009, he has been with the Optoelectronics Research Centre, University of Southampton, U.K. His research interests include optical and optics-assisted signal processing, study of environmental sensitivity of optical fibres, and hollow-core optical fibres and their applications in telecommunications, metrology, and sensing. He is a Fellow Member with the Optical Society of America.

• • •