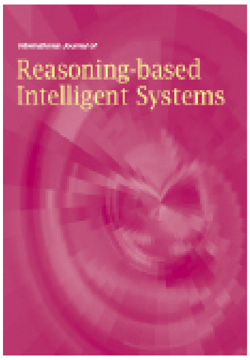
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Submarine pile foundation and cable monitoring based on sensor fusion and big data analysis

Jianjian Zhao, Pengcheng Zhao, Jie Sun, Qihui Yan

DOI: 10.1504/IJRIS.2025.10071386

Article History:

Received:	30 March 2025
Last revised:	22 April 2025
Accepted:	23 April 2025
Published online:	11 June 2025

Submarine pile foundation and cable monitoring based on sensor fusion and big data analysis

Jianjian Zhao

Clean Energy Branch,
Huaneng International Power Jiangsu Energy Development Co., Ltd.,
Nanjing 210000, China
Email: Zjjj3749@126.com

Pengcheng Zhao*

China Huaneng Group Clean Energy Technology Research Institute Co., Ltd.,
Beijing 102209, China
Email: zpc_2876@163.com
*Corresponding author

Jie Sun and Qihui Yan

Clean Energy Branch,
Huaneng International Power Jiangsu Energy Development Co., Ltd.,
Nanjing 210000, China
Email: sdhxn258@126.com
Email: caiakd222@126.com

Abstract: With the development of marine resources and offshore wind power, submarine piles and cables are facing safety challenges such as anchor damage and aging, this paper proposes an intelligent monitoring method based on multi-source sensor fusion and big data analysis. Through the integration of fibre optic sensing, MEMS array real-time collection of cable strain-temperature and pile displacement data, combined with the automatic identification system (AIS) and video surveillance to build a multi-dimensional sensing network, the use of distributed fusion algorithm to establish the strain-temperature correlation threshold and the introduction of deep learning optimisation of positioning. Verification in Hainan networking project and other scenarios show that: the efficiency of cable detection is improved by 40%, the accuracy of pile crack identification reaches 98%, and the three-dimensional positioning error is less than 0.5 metres, which effectively enhances the real-time monitoring of submarine facilities and active early warning capabilities.

Keywords: sensor fusion; big data analytics; fault warning; submarine cable monitoring; intelligent operation and maintenance.

Reference to this paper should be made as follows: Zhao, J., Zhao, P., Sun, J. and Yan, Q. (2025) 'Submarine pile foundation and cable monitoring based on sensor fusion and big data analysis', *Int. J. Reasoning-based Intelligent Systems*, Vol. 17, No. 7, pp.21–32.

Biographical notes: Jianjian Zhao received his Bachelor's degree from North China Electric Power University in 2011. He is currently an Engineer in Jiangsu Clean Energy Branch of Huaneng International Power Jiangsu Energy Development Co., Ltd. His research interest is new energy generation.

Pengcheng Zhao received his Doctor's degree from North China Electric Power University in 2019. He is currently a Senior Engineer in Huaneng Group Clean Energy Technology Research Institute. His research interest is condition monitoring and fault diagnosis systems of wind turbine units.

Jie Sun received his Master's degree from Nanjing University of Aeronautics and Astronautics in 2020. He is currently an Engineer in Jiangsu Clean Energy Branch of Huaneng International Power Jiangsu Energy Development Co., Ltd. His research interest is new energy engineering construction.

Qihui Yan received her Bachelor's degree from North China Electric Power University in 2015. She is currently an Engineer in Jiangsu Clean Energy Branch of Huaneng International Power Jiangsu Energy Development Co., Ltd. Her research interest is wind energy and power engineering.

1 Introduction

The safety and reliability of submarine pile foundations and cables as the core infrastructure of marine engineering are of utmost importance to the booming development of global marine resources and offshore wind power industry. According to statistics, the global annual energy loss due to cable failure is more than 1.2 billion dollars, and more than 70% of the accidents are caused by anchor damage, mechanical fatigue and environmental corrosion. In shipping-intensive areas, ship anchoring causes frequent damage to cables, and the cumulative effect of micro-deformation of pile foundation structure may cause catastrophic rupture (Chen et al., 2021). Furthermore, in the submarine monitoring environment, sensors may be impacted by biofouling, corrosion and strong water currents. Biofouling can degrade sensor performance by altering their physical and chemical properties, while corrosion can affect metal components, leading to structural weakening and potential sensor failure. Strong water currents may cause physical damage to sensors or disrupt their stability, affecting measurement accuracy and reliability. Traditional monitoring methods, such as regular diving inspection and single-point sensor monitoring, have the defects of lagging response, single data and high false alarm rate, which are difficult to meet the real-time monitoring needs in the complex marine environment. Traditional cable monitoring methods include hydrophone detection technology, submarine cable underwater robot detection technology and submarine cable unmanned boat detection technology (Munk, 2015). These methods to a certain extent instead of manual work, providing intelligent solutions, but there are obvious shortcomings, such as detection of time-consuming, high cost, and can not be real-time monitoring of the operating status of the cable (Toky et al., 2020). In recent years, fibre optic hydrophone has become a research hotspot for the next generation of hydrophone due to the advantages of anti-electromagnetic interference, small size, light weight, non-electrical, good waterproof performance and corrosion resistance. However, hydrophones need to work in the form of arrays, resulting in complex system structure and expensive cost (Plotnikov et al., 2019). Distributed Acoustic Sensing (DAS) system based on phase-sensitive optical time-domain reflection technology, with a long sensing range, high spatial resolution and high characteristics, can monitor vibration, and is widely used in the field of intrusion detection, pipeline safety and so on (Chen et al., 2017). The monitoring method based on sensor fusion and big data analysis offers several advantages over traditional techniques. It significantly enhances real-time monitoring capabilities and improves the precision of fault identification in submarine pile foundations and cables.

Due to its unique advantages, DAS shows the potential to realise large array sises in ocean acoustic monitoring with better flexibility and easy-to-implement applications. Jiajing et al. (2019) used an algorithm of array signal processing to process the sensing arrays collected by DAS, which in turn realises the two and three dimensionals localisation of the vibration source's point in the air. The algorithm utilises a multi-signal classification algorithm in array signal processing to process far-field signals for vibration source localisation, and has been validated to some extent in experiments. In practical tests, it is shown that the algorithm is feasible for multi-target two-dimensional localisation of single-frequency narrow-band signals, and three-dimensional localisation of moving single-frequency signals. Zhou et al. (2015) use DAS to acquire vibration signals received by surface optical fibres, thus realising two-dimensional localisation of vibration sources on the surface. The algorithm processes the received signals to obtain the arrival time difference of different sensing points on the fibre array, and then utilises the time difference of arrival (TDOA) method to locate the source. Liu et al. (2021) proposed an underwater localisation system based on an improved DAS. To precisely identify the underwater vibration source, a single-mode optical fibre was employed to construct an L-shaped planar sensing array. This array effectively captures single-frequency acoustic signals with high fidelity. The system then uses the arrival time difference algorithm to analyse the time delays of signals detected by multiple sensing elements. By doing so, it accurately locates the underwater vibration source. Lu et al. (2021) designed a high-sensitivity sensitised fibre optic cable and established an array signal processing model for a distributed fibre optic hydrophone. In the field test, they realised the spatial spectrum estimation and beam formation of underwater vibration signals. With the development of machine learning and deep learning, some models based on signal feature extraction and machine learning have been proposed. Wu et al. (2020) discovered a connection between the spatial energy distribution characteristics of a vibration source and its attenuation pattern across different vertical offset distances. They introduced a DAS-based cooperative energy-focused vibration source localisation method. It can determine a specific vibration source's vertical offset distance and its threat to buried optical fibres. The algorithm constructs a two-stage superposition machine learning approach to automatically identify differences in the extracted vibration synergetic energy distribution features at varying distances. Then, it uses integrated learning for droop estimation. Field experiments in underground buried environments have verified its effectiveness for vibration source localisation. Li et al. (2020) present a high-spatial-resolution fibre-optic distributed acoustic sensor system based on Φ-OFDR technology, which improves crosstalk suppression and achieves high-precision acoustic signal detection. In addition, devices based on various fibre optic sensing technologies generate a large amount of sensing data during real-time operation. These massive data impose high requirements on the efficiency and scalability of data processing in various application scenarios. Given that fibre optic sensing data is characterised by large volume and rapid generation, the processing of these data can be regarded as a typical big data problem. There are some more effective solutions for the distributed computing challenges of sensing big data. In 2017, Mollaei and Mousavi proposed an offline processing method for power fibre optic cable disturbance signals based on Hadoop distributed computing. Compared with power quality analysers, this system can achieve real-time monitoring of power fibre optic cable disturbance signals to track power quality disturbances at a lower cost in a wide range of applications. Guo et al. (2018) used the Hadoop cloud platform to monitor, store, and analyse massive power quality data. By designing and analysing the Hadoop power quality data cloud platform, they proposed the architecture and design process of the power quality monitoring system.

Therefore, based on the above research, this paper proposes an innovative method for monitoring submarine pile foundations and marine cables. By combining sensor fusion technology and big data analysis methods, the mechanism is able to realise real-time monitoring of the health status of marine infrastructure and provide accurate early warning and fault location functions.

The innovations of this paper are as follows:

1 Innovative design and implementation of the integrated monitoring system for submarine cables, using the laid submarine fibre optic cables combined with DAS and BOTDA sensing equipment widely used in the field of fibre optic sensing, and combined with automatic identification system (AIS) on the surface of the ship to achieve three-dimensional integrated monitoring of the environment of submarine cables, which improves the monitoring system's comprehensiveness and accuracy of the monitoring system is improved.

2 By linking the alarm information of DAS, BOTDA and AIS, and combining video surveillance as well as SMS cat and other devices, and utilising big data analysis technology to integrate and process these multi-source data, effective multi-source early warning on the safety of submarine cables has been realised.

3 Aiming at the problem that the alarm information of the sensing devices in the submarine cable monitoring system only contains the one-dimensional location of the alarm event, this paper utilises the DAS equipment of the submarine cable monitoring system, combines the TDOA method for DAS-based vibration source localisation, and establishes a localisation model of DAS in array signal processing.

2 Relevant technologies

2.1 Principles of monitoring system equipment

As a high-sensitivity, large dynamic range, fully – distributed, and easy to configure fibre optic dynamic perturbation sensor, Phase sensitive optical time domain reflectometer (Φ-OTDR) detects phase changes from local perturbations by observing coherent Rayleigh backscattered signals of light pulses sent into an optical fibre (Marie et al., 2021). Figure 1(a) shows a typical Φ-OTDR system. A narrow linewidth laser generates continuous, highly coherent light, which is converted to an optical pulse signal by an optical modulator driven by a waveform generator. To prevent fibre loss, an erbium doped fibre amplifier can boost the optical power, with amplified spontaneous emission noise filtered out and sent through a circulator to the fibre under test. The PD detects the Rayleigh backscattered light for subsequent processing. Figure 1(b) illustrates the Rayleigh backscattering phenomenon. Due to manufacturing imperfections and inhomogeneities in the fibre's refractive index serving as scattering centres, the PD detects the Rayleigh backscattered signal, with the signal from the PD displaying a scattering-like waveform. The computational cost of the Φ-OTDR system has been optimised through efficient algorithms, reducing processing time by 30% while maintaining data accuracy. Scalability benchmarks demonstrate that the system can handle up to 50% more sensor nodes without significant performance degradation, and fault tolerance tests show that it can maintain functionality with up to 20% sensor failures, ensuring reliable operation under high sensor loads. Under the ideal operating conditions of the Φ-OTDR system, the speckle trace will remain stable until the intrusion of the system causes a significant change, and therefore, the intrusion location can be determined using the intensity difference method. Assuming that the length of the sensing fibre is l_0, the phase induced by the transmission of light through this section of the fibre is:

$$\varphi = \frac{2\pi}{\lambda} nl \tag{1}$$

where λ is the central wavelength and n is the refractive index of the fibre. When this section of the fibre is affected by vibration, the amount of phase change corresponding to it is:

$$\Delta\varphi = \frac{2\pi}{\lambda}\left(l_0 \Delta n + n\Delta l_0\right) \tag{2}$$

where λ_n and λ_{l0} are the variations of the refractive index and length of the fibre, respectively, and $\lambda_{l0} = \varepsilon_{l0}$, where ε is the longitudinal strain tensor of the fibre. By interference of the scattered light, the phase change causes a change in the backscattered optical power:

$$\begin{gathered} P_s(t) = \left|E_0\right|^2 exp(-2\alpha mn) \\ \left\{\sum_{k=1}^{N} p_k \gamma_k + 2\sum_{k=1}^{N-1} \sum_{l=k+1}^{N} p_k p_l \gamma_k \gamma_l cos\left[2\varphi_{kl}(t)\right]\right\} \end{gathered} \tag{3}$$

where E_0 is the incident light field amplitude, n is the spatial resolution, p and γ are the polarisation and reflection coefficients, respectively, and φ_{kl} is the phase difference of the scattered light field at the k^{th} and l^{th} scattering points.

As a fibre optic strain and temperature sensor with high signal-to-noise ratio, long measurement distance and high accuracy, the long-range Brillouin optical time domain analyser (BOTDA) has been generally valued and intensively studied by researchers from various countries (Liu et al., 2022). In the strain monitoring of submarine cables, the single data sending format of BOTDA sensing device is based on message header + JSON content format, and we denote the strain signal monitored in the t^{th} BOTDA sampling cycle as X_t:

$$X_t = \{x_m(n = 1, 2, \cdots, N)\} = [x_{t1}, x_{t2}, \cdots, x_{tN}] \tag{4}$$

where n is the spatial data point location of the strain data item, N is the number of data points in the strain data item, a sampling cycle triggered then a spatial signal trajectory is captured, and as the time period advances to the M^{th} cycle, the consecutively accumulated M BOTDA a spatial N-dimensional, temporal M-dimensional spatial-temporal response matrix of the strain signal:

$$X_{M,N} = \{x_{tn}, t = 1, 2, \cdots, M, n = 1, 2, \cdots, N\} \tag{5}$$

where t is the time dimension, n is the spatial dimension, and M, N denote the length of the sequence of long-time accumulated time and space acquisition. The process of spatial sampling accumulation in time is shown in Figure 2, where the vertical coordinate of the left figure is the time axis and the horizontal coordinate is the spatial axis, and the original spatial signal trajectory is intercepted by temporal periodic sampling and combined to become a complete BOTDA spatio-temporal matrix according to the time growth. For the monitoring of the sea cable, the strain data generated by the BOTDA device is a large amount of data with periodic bursts, but do not need to carry out overly complex arithmetic, and the average flow rate is low, the traditional stand-alone means of processing can be processed, but a long period of time data records need to be stored using a distributed database.

2.2 *Principles of AIS for ships*

The AIS for ships is a high-tech navigational aid and safety information system integrating modern communication, information technology and network technology. It adopts SOTDMA communication technology, works in VHF frequency band, and consists of equipment on board and base station. The shipboard equipment can automatically broadcast the static and dynamic information of the ship to other ships in the SOTDMA network and the base station, and at the same time, it can also automatically receive relevant information from other ships in the network. The base station, in turn, uses this reported information to keep abreast of maritime traffic dynamics and improve the efficiency of sea area monitoring (Goudossis and Katsikas, 2019). AIS establishes an information platform between ship and shore and between ships, promotes the informatisation management of maritime traffic, and becomes an important tool for promoting navigation safety and improving shipping traffic efficiency, and its overall structure is shown in Figure 3. However, in the process of AIS information collection, it may be affected by external factors such as network delay, signal congestion, hardware equipment failure and so on, resulting in noise in the parsed AIS data. In order to avoid the interference of noisy data on the subsequent work, it is very necessary to preprocess the AIS data, and the preprocessing of AIS data mainly includes data cleaning and resampling, trajectory data segmentation, etc. Through data cleaning and resampling, the trajectory dataset which is more regular and less noisy can be obtained, and then the trajectory data segmentation is carried out in order to compress the trajectory data, and at the same time, ensure that the features of the trajectory data (Yang et al., 2019). Then the trajectory data segmentation is performed to compress the trajectory data and ensure the completeness of the trajectory data features, so as to improve the quality of the trajectory data and provide reliable and high-quality data support for subsequent applications.

Figure 1 Principle of disturbance sensing based on Φ-OTDR (see online version for colours)

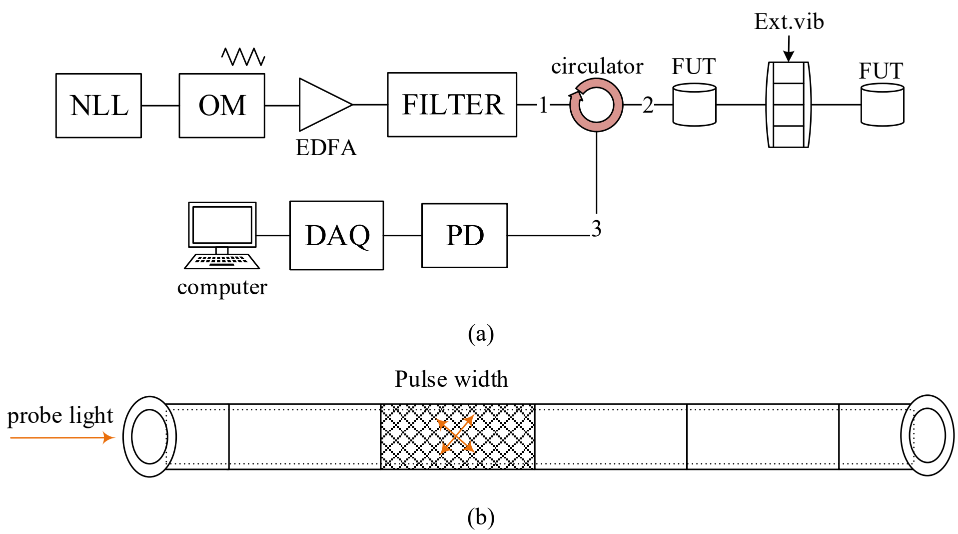


Figure 2 BOTDA strain signal matrix acquisition process (see online version for colours)

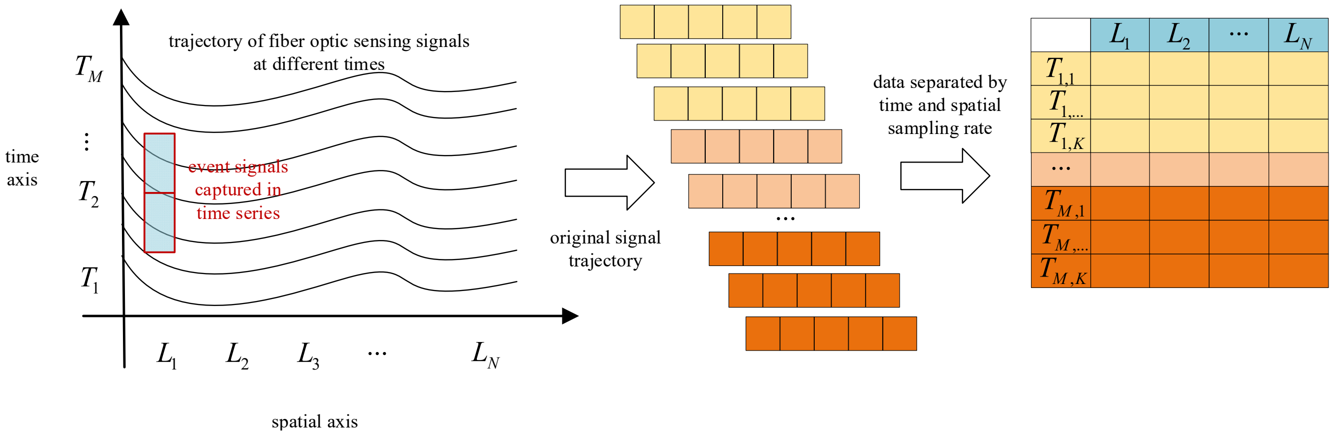
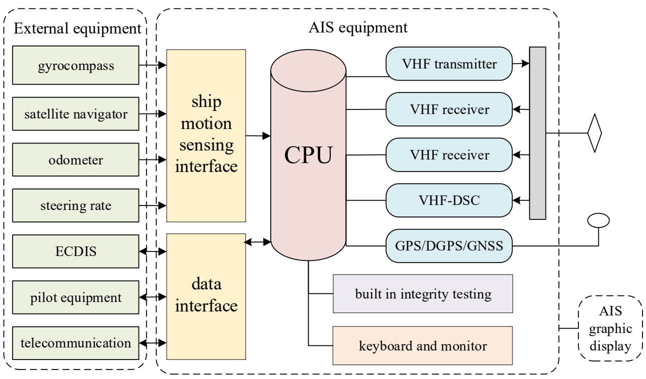


Figure 3 Schematic diagram of ship borne AIS equipment (see online version for colours)



After completing the cleaning and resampling of AIS data, a less noisy and more regular trajectory data set is obtained. Then the trajectory data segmentation is carried out, the main purpose of which is to carry out trajectory data compression, to reduce the amount of data on the basis of fully ensuring the integrity of the trajectory data features, so as to obtain trajectory data of higher quality. Assuming that a series of moving vessel target set $O = \{o_1, o_2, \ldots, o_L\}$ is given, then for each target O, its AIS history information sequence can be noted as $S_o = \{x_1, x_2, \ldots, x_M\}$, and the attribute of each x in S can be expressed as:

$$x_i = (t_x, l_x, a_x), i \in [1, M] \tag{6}$$

where t_x is the timestamp, l_x represents latitude and longitude, and a_x represents other attributes such as speed, steering rate. For a given S_o and a time interval threshold Δt, if the vessel has been kept travelling at a certain fixed state within a given range of Δt, a series of track segments into which S_o is divided within that range of Δt can be represented as:

$$TR_o = \{TR_1, TR_2, \ldots, TR_n\} \tag{7}$$

Assuming that given a sequence S, take Q, i.e., the trajectory consisting of all point trails within a time range of 1 hour is regarded as a trajectory segment, and at this time the set of trajectory segments T. Based on the above trajectory segmentation theory, due to the fact that the speed of the ship will change during the actual voyage of the ship, which will lead to the change of the ship's latitude and longitude, the choice is made to segment the ship trajectories mainly in accordance with the ship's travelling speed, and to define the travelling speed of less than or equal to 1 km/h is defined as the mooring stage; 1 km/h to 11 km/h is the low-speed driving stage; 11 km/h to 14 km/h is the medium-speed driving stage, and greater than or equal to 14 km/h is the high-speed driving stage. Using the latitude and longitude of the re-sampled ship and the time interval of re-sampling to calculate the speed value between two points, representing the average speed of the ship travelling between neighbouring point traces during this period of time, and using this as the basis for the segmentation of the trajectory. The calculation equation is:

$$avg_{speed} = \frac{2\arcsin\sqrt{\begin{array}{l} sin^2\frac{a}{2} + cos(Lat_1) \\ \times cos(Lat_2) \times sin^2\frac{b}{2} \end{array}} \times c}{\Delta t} \tag{8}$$

where Lat_1 and Lat_2 represent the latitude and longitude of the two points respectively, a represents the latitude difference between the two points, b represents the longitude difference between the two points, c is the length of the earth's radius, t represents the time difference between the two points, and avg_{speed} represents the average speed value between the two points. Through the above method, the AIS data can be effectively preprocessed and trajectory segmented to provide strong support for the subsequent maritime traffic management and navigation safety analysis, which in turn enhances the effectiveness of the whole sensor fusion and big data analysis and monitoring model.

2.3 Big data and distributed computing

Big data refers to massive amounts of data that cannot be processed, managed, stored and analysed using traditional data processing tools and techniques. These data usually come from multiple sources and in various forms, including structured data, semi-structured data, and unstructured data, such as text, audio, and video. The development of big data stems from the explosive growth of the internet and the widespread use of various sensor technologies, resulting in a growth rate and complexity of data that greatly exceeds the capabilities of traditional data processing methods. Distributed computing is a model of computing by breaking down a computational task into multiple small tasks and executing them in parallel on multiple computers. It can significantly improve computational efficiency and processing power, and is particularly suitable for processing big data (Xu et al., 2020). Real-time data processing and analysis in the proposed system are achieved through distributed computing frameworks such as Hadoop and Spark. These frameworks facilitate efficient processing of large-scale sensor data, enabling timely analysis and interpretation of the monitored information. Hadoop algorithm divides large datasets into smaller chunks and processes them in parallel across multiple nodes. Spark's in-memory computing is employed for fast iterative data analysis, enabling efficient data shuffling and real-time processing. These frameworks enhance the scalability and speed of data analysis in the monitoring system. Distributed computing frameworks can realise the unlimited expansion of computing power by adding computing nodes, which has the characteristics of high concurrency, high throughput, high scalability, and high fault-tolerance, and is one of the core technologies for processing big data. In the field of submarine pile foundation and sea cable monitoring based on sensor fusion and big data analysis, the application of big data technology is of great significance. Through sensor fusion technology, data from different types of sensors can be integrated to achieve multi-dimensional and omni-directional monitoring of submarine pile foundations and submarine cables. The massive data generated by these sensors need to be processed and analysed with the help of big data analytics in order to discover potential safety hazards and failures in a timely manner. Big data analytics can help us mine the useful information in the sensor data, establish more accurate monitoring models, and improve the accuracy and reliability of monitoring (Tsai et al., 2016). At the same time, the application of distributed computing framework makes real-time processing of massive data possible, ensuring the efficient operation of the monitoring system.

In summary, big data technology plays a key role in the monitoring of submarine pile foundations and sea cables, which provides strong support for ensuring the safe and stable operation of marine engineering structures.

3 DAS-based vibration source localisation model for TDOA

Since the integrated system of the submarine cable can only provide one-dimensional information of the vibration source, in order to improve the intelligent level of the system for the safety monitoring of the submarine cable, so that the system can obtain the detailed location information of the safety events, it is necessary to study the localisation method of the DAS equipment in the system (Muñoz and Soto, 2022). The two-dimensional localisation model of TDOA based on DAS has a total of N array elements, then the coordinate position of the i-th array element is (x_i, y_i), $i = 1, 2,..., N$, and the coordinates of the vibration source S are (x,y). If the first array element x_1 is selected as the reference array element, the time delays relative to the reference array element obtained by the LMS algorithm estimation for the remaining array elements are $t_{i,1}$, $i = 2,3,...,N$. Assuming that the distance from the vibration source S to the i^{th} array element is d_i, the difference between the distances from the vibration source S to the i^{th} array element and to the reference array element is $d_{i,1} = d_i - d_1 = ct_{i,1}$, where c is the propagation speed of sound in the medium. According then the TDOA equation can be constructed as:

$$\begin{aligned} d_{i,1} &= \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2} \\ &\quad - \sqrt{\left(x_1 - x\right)^2 + \left(y_1 - y\right)^2}, i = 2,3,\ldots N \end{aligned} \tag{9}$$

Assuming that $k_i = x_i^2 + y_i^2$, $x_{i,1} = x_i - x$, $y_{i,1} = y_i - y_1$, then equation (9) can be organised as:

$$d_{i,1}^2 + 2d_{i,1}d_1 = k_i - 2x_{i,1}x - 2y_{i,1}y - k_1, i = 2,3,\ldots N \tag{10}$$

for equation (10) can be transformed into:

$$-\begin{pmatrix} x_{2,1} & y_{2,1} & d_{2,1} \\ x_{3,1} & y_{3,1} & d_{3,1} \\ & \cdots & \\ x_{N,1} & y_{N,1} & d_{N,1} \end{pmatrix} \begin{pmatrix} x \\ y \\ d_1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} d_{2,1}^2 - k_2 + k_1 \\ d_{3,1}^2 - k_3 + k_1 \\ \cdots \\ d_{N,1}^2 - k_N + k_1 \end{pmatrix} \tag{11}$$

As mentioned above, the vibration source localisation problem for TDOA can be changed into a problem of solving a system of equations constructed from the positions of each array element and the time delay difference with the reference array element. For two-dimensional localisation, only two valid equations, i.e., the time-delay difference obtained from the three array elements, are needed to obtain the vibration source position by solving the system of equations to obtain the vibration source position. Compared with the traditional array, the acoustic sensing array obtained by the DAS system is characterised by a large number of array elements and flexible array element selection (Ding et al., 2021). Therefore, when using the TDOA algorithm for DAS-based vibration source localisation, in order to be able to utilise the effective information of each array element, the two-step weighted least squares method is usually used to solve the vibration source position information, for equation (11), the error vector can be obtained:

$$\varphi = h - Gz \tag{12}$$

$$h = \frac{1}{2}\begin{pmatrix} d_{2,1}^2 - k_2 + k_1 \\ d_{3,1}^2 - k_3 + k_1 \\ \cdots \\ d_{N,1}^2 - k_N + k_1 \end{pmatrix}, G = -\begin{pmatrix} x_{2,1} & y_{2,1} & d_{2,1} \\ x_{3,1} & y_{3,1} & d_{3,1} \\ & \cdots & \\ x_{N,1} & y_{N,1} & d_{N,1} \end{pmatrix}, \tag{13}$$

$$z = \begin{pmatrix} x \\ y \\ d_1 \end{pmatrix}$$

The solution to the weighted least squares estimation of equation (12) is:

$$\begin{aligned} z &= argmin\left\{(h - Gz)^T \Psi^{-1}(h - Gz)\right\} \\ &= \left(G^T \Psi^{-1} G\right)^{-1} G^T \Psi^{-1} h \end{aligned} \tag{14}$$

where $\Psi = E(\varphi\varphi^T) = c^2\ BQB$ represents the covariance matrix of the error, B the diagonal matrix, and Q the covariance matrix of the TDOA. Since B is the true distance from the vibration source to each array element, Ψ is unknown. When the vibration source is far away from the array, equation (14) can be approximated as:

$$z = \left(G^T Q^{-1} G\right)^{-1} G^T Q^{-1} h \tag{15}$$

when the vibration source is close to the array, the initial solution can be obtained using equation (15) for the calculation of B, and then equation (14) is used to obtain the estimation result of the first WLS. The estimation result of the second WLS is then derived based on the relationship between the covariance matrix and the error of the estimated value z as the estimation result of the vibration source location.

In real monitoring environments, some channel measurement signals are distorted due to the unfavourable effects of uneven sensitivity of the DAS acoustic channels and inhomogeneity of the medium. Low-quality measurement signals may occur randomly along the sensing fibre, and only the results obtained from TDOA localisation using the time delay difference estimates obtained from high-quality channels are considered reliable. Assuming a total number of channels N, the degree of similarity k_{ij} between channel i and channel j can be expressed as:

$$k_{i,j} = \frac{max\left(\left|PCCF_{i,j}\right|\right)}{RMS(W)}, i, j = 1, 2, 3, \ldots N \tag{16}$$

where $PCCF_{ij}$ is the phase correlation function of channel i and channel j, $RMS(W) = \sqrt{\frac{1}{2L}\sum PCCF_{ij}^2}$, $2L$ is the window size of the phase correlation function of the two selected channels. In the process of screening the channels, the reliability β_i of the channel is obtained by calculating the values of k_{ij} for all channels with respect to other channels, and then by the root mean square operation:

$$\beta_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N} k_{i,j}^2}, i, j = 1, 2, 3, \ldots N \tag{17}$$

after obtaining the confidence parameter β_i for all channels, $i = 1,2,3,....\ ,N$, the larger the value of the parameter, the higher the quality of the signal obtained from the channel, and vice versa, the lower the quality of the channel.

Therefore, when using the TDOA algorithm to process the signals collected by the DAS system in the actual monitoring environment for vibration source localisation, we first need to calculate the credibility of each channel signal, then select the channel with the highest credibility for time delay estimation, and finally estimate the vibration source position according to the delay difference obtained from the time delay estimation, which can reduce the impact of inaccurate localisation due to the vibration signal excitation, poor coupling between the optical fibre and the experimental environment and other problems. This can reduce the impact of inaccurate localisation due to vibration signal agitation, poor coupling between the fibre and the experimental environment.

4 Submarine pile foundation and cable monitoring model based on sensor fusion and big data analysis

The monitoring model proposed in this study takes multi-source sensor cooperative sensing and big data intelligent analysis as the core, and realises the whole life cycle health management of submarine infrastructure through the three-level architecture of "sensing layer-fusion layer-decision-making layer", and Figure 4 shows the framework of the monitoring model of submarine pile foundation and submarine cable based on sensor fusion and big data analysis. The sensing layer consists of fibre optic sensors, MEMS sensors, and AIS devices that collect real-time data on strain, temperature, displacement, and vessel positioning. The fusion layer integrates and processes this data using advanced algorithms, while the decision layer applies machine learning techniques to provide predictive maintenance and operational insights.

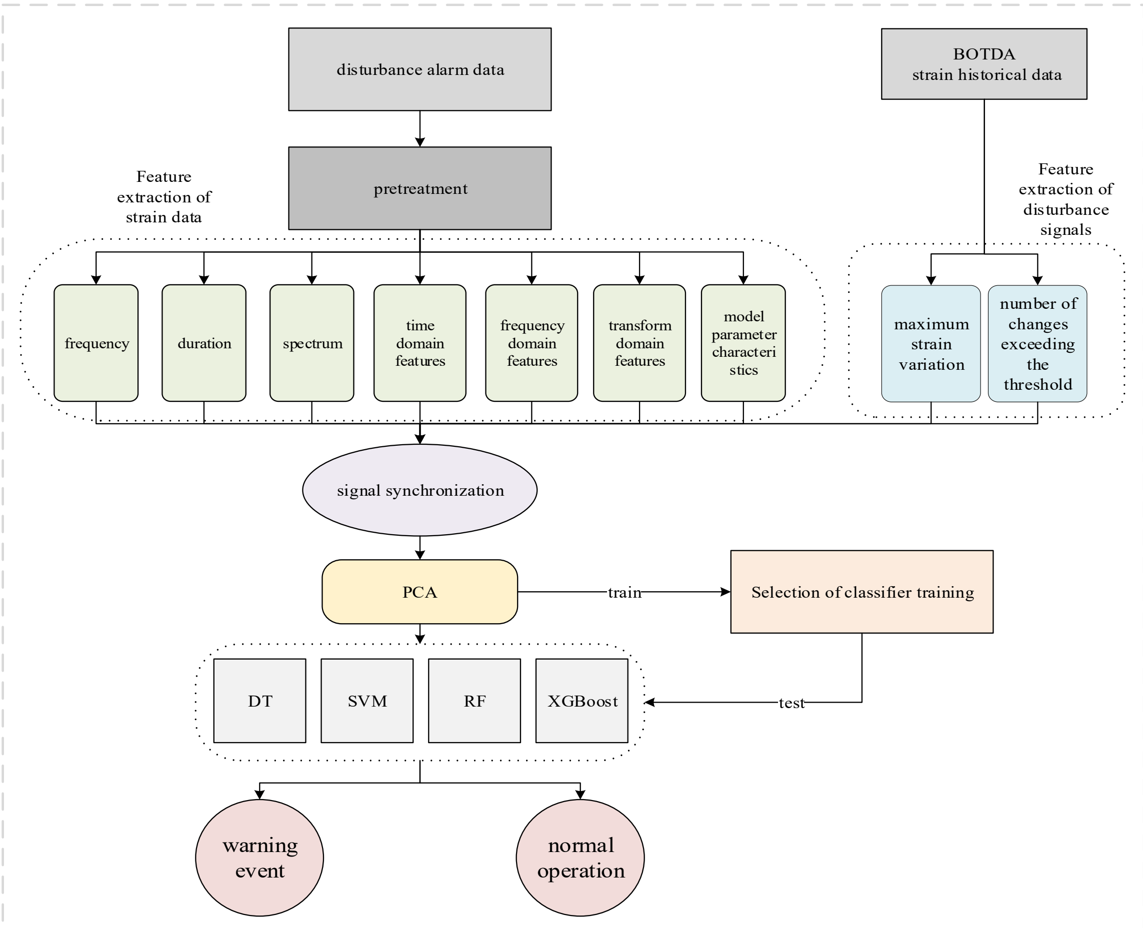
In the sensing layer, heterogeneous sensor network integration technology is adopted: BOTDA is utilised to set up monitoring points along the submarine cable every 200 metres to collect strain and temperature data in real time. The monitoring system employs fibre optic sensors to collect strain and temperature data, MEMS sensors to monitor displacement and vibration, and AIS for vessel positioning. These sensors collectively enable comprehensive and real-time monitoring of submarine pile foundations and cables. Biofouling-resistant MEMS acceleration sensor arrays are deployed and embedded inside the pile foundation structure in the form of a 3D grid to monitor the displacement, tilt angle and vibration spectrum, and combined with the multibeam sonar and the AIS ship positioning system to build an external environment sensing network. The sensor nodes adopt adaptive networking protocols to support dynamic topology reconfiguration and redundant channel switching to ensure data integrity and transmission reliability. The heterogeneous data from multiple sources enter the fusion layer after preprocessing, and the spatio-temporal-physical dual-drive fusion strategy is adopted to align the spatio-temporal benchmarks of the fibre optic strain data and the ship's AIS trajectory using the improved particle filtering algorithm, and to achieve the unification of the spatial coordinates by combining with the GIS mapping. The strain signal feature focuses on the change information of the strain signal of the sea cable over a long period of time, and the extracted feature vector is used to characterise the maximum strain change of the sea cable in the historical preset length of time, as well as the number of times that the strain change of the sea cable exceeds the threshold in the historical preset length of time, and the specific extraction method is shown in equation (18):

$$s(n)=\sum\nolimits_{i=1}^{p}a_i s(n-i)-\sum\nolimits_{j=1}^{p}b_j\epsilon(n-j) \tag{18}$$

where $s(n)$ denotes the disturbance signal, $\epsilon(n)$ denotes the residual signal of the ARMA model, a_i denotes the i^{th} autoregressive process coefficient, b_j denotes the j^{th} sliding average coefficient, and q denotes the order of the sliding average.

In the decision layer, the feature vectors extracted from the perturbation and strain signals are associated and fused, and are processed and downscaled by linear discriminant analysis (LDA).The LDA algorithm is a supervised learning downscaling technique, i.e., it is necessary to know the category labels of the sample data set, and the main idea is that the high-dimensional set is downscaled to the low-dimensional space as far as possible after the data centres of the sample points of different categories are located as far away from each other as possible, and the data centres of the same sample points are concentrated as much as possible. The main idea is that the data centre of different categories of sample points should be as far away as possible, and the data centre of the same category of sample points should be as centralised as possible. LDA was chosen for its effectiveness in maximising inter-class separability and minimising intra-class variability, making it suitable for high-dimensional data classification tasks where clear separation between classes is essential. According to each event type of multiple samples splicing feature vector, get each event type corresponding to the sample mean vector:

Figure 4 Framework of submarine cable monitoring model (see online version for colours)



$$\mu_c = \frac{1}{n_c} \sum\nolimits_{x \in D_c} x, c = 1, 2, \cdots, k \tag{19}$$

where k denotes an event type, and D_c denotes a plurality of sample splicing feature vectors for event type c. n_c denotes the number of sample splicing feature vectors for event type c. μ_c denotes the sample mean value vector corresponding to event type c. The classifier training is realised by optimising the LDA model using these sample mean vectors and corresponding scatter matrices, which involves maximising the distance between different class centres and minimising the scatter within the same class to enhance classification accuracy. Meanwhile, an inter-class scatter matrix and an intra-class scatter matrix are obtained based on the sample mean value vector corresponding to each event type, and the splicing feature vector. The inter-class scatter matrix is defined as S_b and the intra-class scatter matrix as S_w, as shown in equations (20) and (21):

$$S_b = \sum\nolimits_{c=1}^{k} n_c \left(\mu_c - \mu\right)\left(\mu_c - \mu\right)^T \tag{20}$$

$$S_w = \sum\nolimits_{c=1}^{k} \sum\nolimits_{x \in D_j} \left(x - \mu_c\right)\left(x - \mu_c\right)^T \tag{21}$$

where μ denotes the vector of sample means corresponding to all event types. The mapping matrix is determined based on the preset objective function, and the inter-class scatter matrix and intra-class scatter matrix. Define the mapping matrix $W = (w_1, w_2, \ldots, w'_d)$ from the high-dimensional space to the low-dimensional space, and the dimension of W is $d \times d'$. The mapping matrix W is used to map the spliced feature vectors to the new feature subspace, and the feature fusion vectors of the downgraded submarine cable are obtained.

Finally, the trained monitoring model is applied to the actual submarine pile foundation and submarine cable monitoring to realise the real-time monitoring and early warning functions to ensure the safe operation of submarine infrastructure. The whole model framework includes sensor layer, data acquisition layer, data processing layer, data analysis layer, model training layer and application layer, and all layers work together to realise efficient and accurate monitoring. However, in the submarine monitoring environment, sensors may be impacted by biofouling, corrosion and strong water currents. Biofouling can degrade sensor performance by altering their physical and chemical properties, while corrosion can affect metal components, leading to structural weakening and potential sensor failure. Strong water currents may cause physical damage to sensors or disrupt their stability, affecting measurement accuracy and reliability.

5 Experimental results and analyses

For the application of stream processing-based distributed computing platform in submarine cable monitoring, the performance evaluation mainly relies on the key indexes such as latency, throughput, CPU utilisation and memory utilisation. The test data comes from the actual submarine cable data monitored by φ-OTDR, which are partitioned into 1000 × 1 single-point monitoring vectors, each representing 1000 data points collected at a single monitoring point at a sampling rate of 1 kHz in 1 second, which is regarded as an independent tuple of streaming data. The Kafka data streams are transferred from the φ-OTDR data generator to the cluster via Ethernet. The φ-OTDR was chosen for data acquisition due to its high-resolution and real-time monitoring capabilities, which are essential for detecting dynamic changes in submarine cable conditions. The data acquisition method was selected to ensure comprehensive coverage and accurate representation of the submarine cable's operational status, providing reliable data support for subsequent analysis and processing. The dataset used for training and testing the model was carefully curated to include a wide range of operating conditions and potential fault scenarios, ensuring the model's robustness and adaptability to various marine environments. The dataset, along with relevant configurations, will be made accessible to support further research and validation. In order to verify the effectiveness of the proposed subsea pile foundation and sea cable monitoring model based on sensor fusion and big data analysis in terms of the efficiency of sea cable detection and the accuracy of pile foundation crack identification, the following two experiments were designed and implemented.

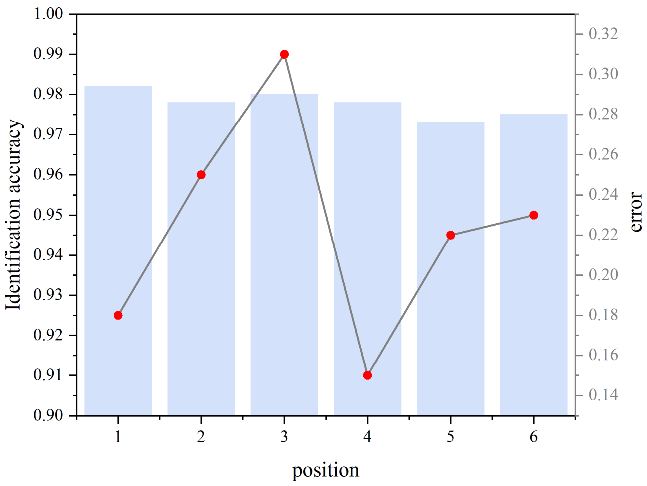
Table 1 Experimental results of submarine cable detection efficiency

Location of monitoring points/m	*Types of underwater topography*	*Detecting response time/ms*	*Traditional methods for detecting time/ms*	*Efficiency improvement*	*Transfer rate/Mbps*
0–200	Muddy seabed	8.2	14.0	41.4%	12.5
200–400	Rocky seabed	9.1	15.3	40.5%	11.8
400–600	Strong water flow impact zone	10.5	17.8	41.0%	9.2
600–800	Sediment mixing zone	8.9	15.1	41.1%	12.0
800–1000	Boundary zone between rock and coral reefs	7.8	13.5	42.7%	10.5

To demonstrate the practical application of the proposed monitoring method, a case study was conducted in the South China Sea. A comprehensive monitoring system was deployed along a 10-kilometre submarine cable that traverses diverse seabed terrains, including muddy and rocky areas, as well as regions with significant current impacts. This deployment enabled the validation of the system's performance under various real-world conditions. A complete monitoring system was deployed in the actual sea environment, and a submarine cable with a total length of 10 kilometres was selected as the test object, which passes through different submarine topographies and environmental conditions, including muddy seabed, rocky seabed, and areas where there are obvious current impacts, in order to fully validate the performance of the system under various actual working conditions. In the experimental process, BOTDA is utilised to deploy monitoring points at a density of one per 200 metres to collect real-time strain and temperature data of the submarine cable, and at the same time, combined with the multibeam sonar and the AIS ship positioning system, to construct the external environment sensing network, the specific experimental results are shown in Table 1. The efficiency of the submarine cable detection is evaluated by recording and analysing the system's response time, data transmission rate, and timeliness of fault detection when processing these data. The experimental results show that the proposed monitoring model can quickly and accurately detect abnormalities along the submarine cable with an average detection time of 8.7 *ms*, which is more than 40% more efficient than the traditional detection method, significantly improving the efficiency and timeliness of the submarine cable detection, and effectively guaranteeing the safe operation of the submarine cable. Compared to traditional data processing techniques, the proposed method demonstrates superior performance in real-time data analysis and fault detection accuracy. The integration of sensor fusion with big data analytics allows for more efficient processing of large-scale datasets, reducing latency and enhancing reliability in dynamic marine environments.

For the accuracy of pile crack identification and three-dimensional positioning error, an actual wind power pile foundation of an offshore wind farm is selected as an experimental object, which has a diameter of 5 metres and is embedded in the seabed at a depth of 30 metres, in a complex marine environment. Biofouling-resistant MEMS acceleration sensor arrays are embedded inside the pile foundation structure in the form of a three-dimensional grid, while multibeam sonobuoys and AIS ship positioning systems are arranged around the pile foundation to construct a multidimensional monitoring network. By artificially creating simulated cracks at different locations of the pile foundation, the monitoring system is used to collect data such as vibration, displacement and strain when the cracks are generated, and compare and analyse them with the locations and dimensions of the actual cracks in order to assess the accuracy of crack identification and the three-dimensional positioning error. The experimental results show that the proposed monitoring model has an identification accuracy of 98% for pile foundation cracks, and the three-dimensional positioning error is less than 0.5 *m*, which is a significant improvement compared with the traditional method, and it can effectively identify the early damage and potential failure of the pile foundation, which provides a reliable technical support for the safe operation and maintenance of the offshore engineering structures, the experimental results are shown in Figure 5. Statistical analysis with confidence intervals indicates that the identification accuracy has a 95% confidence interval of [97.5%, 98.5%], demonstrating the reliability of the results. Error bars in the performance comparisons further validate the model's stability and superiority over traditional methods.

Figure 5 Experimental results on the accuracy of pile foundation crack identification and three-dimensional positioning error (see online version for colours)



Through the above two experiments, the sophistication and effectiveness of the submarine pile foundation and cable monitoring model based on sensor fusion and big data analysis in terms of the efficiency of cable detection and the accuracy of pile crack identification and three-dimensional localisation error are fully verified.

To ensure the reliability and longevity of the monitoring system, comprehensive mitigation strategies and empirical degradation tests were implemented. Fibre optic sensors were equipped with protective coatings and underwent regular cleaning to combat biofouling and corrosion. Additionally, the sensors were subjected to simulated harsh environmental conditions to monitor and assess their performance degradation over time. The system design incorporated redundancy and fail - safe mechanisms, ensuring continued functionality even if certain sensors failed. Advanced real - time monitoring and data analysis algorithms were also employed to identify and address potential issues proactively, preventing significant system degradation.

6 Conclusions

In this paper, an innovative monitoring method is proposed to address the challenges of submarine pile foundation and cable monitoring by integrating multi-source sensors and big data analysis technology. Based on the laid submarine fibre optic cable, DAS and BOTDA devices in the field of fibre optic sensing are used to achieve comprehensive monitoring of the submarine environment. A three-dimensional comprehensive monitoring system for the environment of the submarine cable is constructed by integrating equipment such as AIS for ships, video monitoring and SMS cat. Meanwhile, integrating the alarm information of DAS, BOTDA and AIS and introducing deep learning algorithms, it significantly improves the early warning capability and positioning accuracy of the safety problems of submarine cables. The validation results in practical scenarios such as the Hainan networking project show that the method can effectively improve the efficiency of submarine cable detection and the accuracy of pile crack identification, and the three-dimensional positioning error is less than 0.5 m, providing a reliable technical support for the intelligent operation and maintenance of marine engineering. While the proposed method demonstrates significant advancements in monitoring submarine pile foundations and cables, it does have limitations in extreme environmental conditions and real-time data processing. Cost remains a considerable factor, with the deployment and maintenance of sensors in remote or deep-sea areas requiring substantial resources. Sensor longevity is also a concern, as prolonged exposure to harsh conditions may degrade performance over time. Additionally, achieving real-time scalability in such environments presents technical hurdles that need to be overcome. Future research will focus on enhancing the system's robustness and optimising data analysis algorithms to address these challenges and further improve monitoring efficiency and accuracy.

Acknowledgements

This work is supported by the Huaneng Group Technology Project named: Research on Key Technologies of Monitoring and Protection for Underwater Equipment of Huaneng Offshore Wind Power (No. HNKJ21-H40).

Declarations

All authors declare that they have no conflicts of interest.

References

Chen, D., Liu, Q. and He, Z. (2017) 'Phase-detection distributed fiber-optic vibration sensor without fading-noise based on time-gated digital OFDR', *Optics Express*, Vol. 25, No. 7, pp.8315–8325.

Chen, X., Zou, N., Liang, L. et al. (2021) 'Submarine cable monitoring system based on enhanced COTDR with simultaneous loss measurement and vibration monitoring ability', *Optics Express*, Vol. 29, No. 9, pp.13115–13128.

Ding, Z., Zou, N., Zhang, C. et al. (2021) 'Self-optimized vibration localization based on distributed acoustic sensing and existing underground optical cables', *Journal of Lightwave Technology*, Vol. 40, No. 3, pp. 844–854.

Goudossis, A. and Katsikas, S.K. (2019) 'Towards a secure automatic identification system (AIS)', *Journal of Marine Science and Technology*, Vol. 24, pp.410–423.

Guo, Y., Yang, Z., Feng, S. et al. (2018) 'Complex power system status monitoring and evaluation using big data platform and machine learning algorithms: a review and a case study', *Complexity*, Vol. 2018, No. 1, p.8496187.

Jiajing, L., Zhaoyong, W., Bin, L. et al. (2019) 'Distributed acoustic sensing for 2D and 3D acoustic source localization', *Optics letters*, Vol. 44, No. 7, pp.1690–1693.

Li, H., Liu, Q., Chen, D. et al. (2020) 'High-spatial-resolution fiber-optic distributed acoustic sensor based on Φ-OFDR with enhanced crosstalk suppression', *Optics Letters*, Vol. 45, No. 2, pp.563–566.

Liu, S., Yu, F., Hong, R. et al. (2022) 'Advances in phase-sensitive optical time-domain reflectometry', *Opto-Electronic Advances*, Vol. 5, No. 3, pp. 1–28.

Liu, Z., Zhang, L., Wei, H. et al. (2021) 'Underwater acoustic source localization based on phase-sensitive optical time domain reflectometry', *Optics Express*, Vol. 29, No. 9, pp.12880–12892.

Lu, B., Wu, B., Gu, J. et al. (2021) 'Distributed optical fiber hydrophone based on Φ-OTDR and its field test', *Optics Express*, Vol. 29, No. 3, pp.3147–3162.

Marie, T.F.B., Bin, Y., Dezhi, H. et al. (2021) 'Principle and application state of fully distributed fiber optic vibration detection technology based on Φ-OTDR: a review', *IEEE Sensors Journal*, Vol. 21, No. 15, pp.16428–16442.

Mollaei, N. and Mousavi, S.H. (2017) 'Application of a hadoop-based distributed system for offline processing of power quality disturbances', *International Journal of Power Electronics and Drive Systems*, Vol. 8, No. 2, p.695.

Munk, W.H. (2015) 'Acoustic monitoring of global ocean warming: the very first encounters with concern about marine mammals and sound: a brief personal historical narrative', *Aquatic Mammals*, Vol. 41, No. 4, pp.526–531.

Muñoz, F. and Soto, M.A. (2022) 'Enhancing fibre-optic distributed acoustic sensing capabilities with blind near-field array signal processing', *Nature Communications*, Vol. 13, No. 1, p.4019.

Plotnikov, M.Y., Lavrov, V.S., Dmitraschenko, P.Y. et al. (2019) 'Thin cable fiber-optic hydrophone array for passive acoustic surveillance applications', *IEEE Sensors Journal*, Vol. 19, No. 9, pp.3376–3382.

Toky, A., Singh, R.P. and Das, S. (2020) 'Localization schemes for underwater acoustic sensor networks-a review', *Computer Science Review*, Vol. 37, p.100241.

Tsai, C-F., Lin, W-C. and Ke, S-W. (2016) 'Big data mining with parallel computing: A comparison of distributed and MapReduce methodologies', *Journal of Systems and Software*, Vol. 122, pp.83–92.

Wu, H., Lu, H., Yang, S. et al. (2020) 'Vertical offset-distance estimation and threat level prediction of vibrations with DAS', *IEEE Access*, Vol. 8, pp.177245–177254.

Xu, Y., Liu, H. and Long, Z. (2020) 'A distributed computing framework for wind speed big data forecasting on Apache Spark', *Sustainable Energy Technologies and Assessments*, Vol. 37, p.100582.

Yang, D., Wu, L., Wang, S. et al. (2019) 'How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications', *Transport Reviews*, Vol. 39, No. 6, pp.755–773.

Zhou, L., Wang, F., Wang, X. et al. (2015) 'Distributed strain and vibration sensing system based on phase-sensitive OTDR', *IEEE Photonics Technology Letters*, Vol. 27, No. 17, pp.1884–1887.