

Evaluation of Reinforcement Learning-Based Adaptive Modulation in Shallow Sea Acoustic Communication

Yifan Qiu^{1,2}, Xiaoyu Yang^{1,2}, Feng Tong^{1,2} and Dongsheng Chen^{1,2}

Received: 10 October 2024 / Accepted: 02 December 2024
© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

While reinforcement learning-based underwater acoustic adaptive modulation shows promise for enabling environment-adaptive communication as supported by extensive simulation-based research, its practical performance remains underexplored in field investigations. To evaluate the practical applicability of this emerging technique in adverse shallow sea channels, a field experiment was conducted using three communication modes: orthogonal frequency division multiplexing (OFDM), M-ary frequency-shift keying (MFSK), and direct sequence spread spectrum (DSSS) for reinforcement learning-driven adaptive modulation. Specifically, a Q-learning method is used to select the optimal modulation mode according to the channel quality quantified by signal-to-noise ratio, multipath spread length, and Doppler frequency offset. Experimental results demonstrate that the reinforcement learning-based adaptive modulation scheme outperformed fixed threshold detection in terms of total throughput and average bit error rate, surpassing conventional adaptive modulation strategies.

Keywords Adaptive modulation; Shallow sea underwater acoustic modulation; Reinforcement learning

1 Introduction

The technology of underwater acoustic communication (UAC) plays an increasingly vital role in observing marine life, monitoring water pollution, exploring resources, surveying natural disasters, and ensuring national security (Gussen et al., 2016; Tang et al., 2020; Huang et al., 2018). Underwater communications face more severe challenges, such as frequency-selective fading, large Doppler shifts, and randomly varying multi-path propagation, than terres-

trial communications (Stojanovic and Preisig, 2009). Moreover, shallow sea underwater acoustic communication is further complicated by strong reflections from the sea surface and seafloor, as well as the influence of various types of anthropogenic noise (Bulut and Ergin, 2021).

Conventional UAC systems, which usually use a single modulation mode, struggle to adapt to the hostile time-spatial-frequency variations of the underwater acoustic (UWA) channel (Fan and Wang, 2021). To address this challenge, adaptive modulation technology has drawn significant attention from the community by adjusting modulation parameters or switching modulation modes to accommodate the adverse variations of the UWA channel. Recently, growing investigations have been conducted on the adaptive modulation of UAC.

Benson et al. (2000) proposed an adaptive modulation system to select between coherent modulation and incoherent modulation based on time delay spread and Doppler spread. Mani et al. (2008) proposed a variable-rate adaptive modulation coding technique that uses an achievable information rate and equalized signal-to-noise ratio (SNR) as criteria for selection and verified its feasibility through experiments. Radosevic et al. (2014) introduced an adaptive modulation technique for orthogonal frequency division multiplexing (OFDM) that uses adaptive bit and power allocation. Wan et al. (2015) proposed a new index for assessing communication performance.

Barua et al. (2022) designed a real-time OFDM-based adaptive modulation UWA communication system. The

Article Highlights

- A Q-learning based adaptive modulation scheme, using the signal-to-noise ratio, multipath spread length, and Doppler frequency offset to evaluate channel quality, is proposed for shallow sea acoustic communication.
- An at-sea RL adaptive modulation UWA communication experiment is conducted to evaluate the practical effectiveness of RL in optimizing adaptive modulation UWA communication.
- Experimental results demonstrate that the Q-learning based adaptive modulation scheme outperformed fixed threshold detection in terms of total throughput and average bit error rate, surpassing conventional adaptive modulation strategies.

✉ Feng Tong
ftong@xmu.edu.cn

¹ College of Ocean and Earth Sciences, Xiamen University, Xiamen 361100, China

² National and Local Joint Engineering Research Center for Navigation and Location Service Technology, Xiamen 361100, China

proposed cluster-based adaptive modulation schemes use estimated SNR and predefined thresholds to switch modulation sizes for clusters. Experiments conducted in tanks and rivers demonstrated that the schemes can achieve superior performance for a nonstationary, time-varying UWA channel.

However, conventional methods for classifying thresholds based on channel state information (CSI) generally require a large amount of prior filed data to determine the corresponding thresholds, and different environments may require distinct thresholds. Moreover, adaptive modulation requires a feedback link to transmit the current CSI and demodulation BER back to the transmitter, providing a reference for switching to the next modulation mode. However, the time delay associated with round-trip handshaking (Su et al., 2022) can render the feedback information outdated, leading to lagging modulation adaptation and performance degradation.

Recently, machine learning has received considerable attention as a key enabler for advancing terrestrial wireless communications, offering a novel approach to underwater adaptive modulation communication (Huang et al., 2022). Zhu et al. (2023) proposed a learning model called AttLstmPreNet to predict UWA channels. This model combines an attention mechanism with LSTM-based architectures to capture sparsity in time-delay scales and coherence in the gap-time scale. The proposed method demonstrates superior prediction accuracy to adaptive channel predictors and the plain LSTM model. Reinforcement learning (RL) focuses on enabling learning agents to take actions in a certain environment based on feedback to maximize cumulative rewards over time. It is considered an effective approach to making wise decisions in a dynamic environment. RL has shown great potential in AUV path planning (Bhopale et al., 2019), UWA network routing (Zhang et al., 2021), and UWA communication. Wang et al. (2018) developed a model-based RL adaptive transmission algorithm for time-varying UWA channels. The approach formulates the adaptive problem as a partially observable Markov decision process.

Fu and Song (2018) investigated a Dyna-Q-based adaptive modulation algorithm designed to predict the next state and communication throughput according to the effective SNR. The results were validated in a mobile AUV scenario. Su et al. (2019) proposed an adaptive modulation and coding scheme using RL to improve communication efficiency. This approach selected a transmission strategy based on the quality of service requirements without requiring a priori knowledge of the channel model.

However, most previous research has relied on numerical simulations, revealing a considerable lack of field investigations under practical UWA channel conditions to evaluate the performance enhancements of reinforcement learning-based modulation optimization. Huang and Diamant

(2020) used realistic experimental data to validate a throughput classification approach. However, due to the drifting of the two receivers and the time-varying channel, the channel state (SNR) varied across different modulation schemes. Sweta et al. (2024) proposed a reinforcement learning-based automated modulation switching algorithm designed to enhance data transmission efficiency and communication quality. The study presents a cost-effective acoustic modem design using commercially available components, implemented on a Raspberry Pi, enabling full-duplex underwater communication. Although a number of RL models have been successfully applied to adaptive communication, validation experiments based on real marine environment data are limited. A field experiment in a representative shallow water channel is conducted to assess the practical effectiveness of RL in shallow sea underwater acoustic communication. The evaluation involved three modulation modes, namely, OFDM, MFSK, and DSSS; it provided a performance analysis and comparison of Q-learning-driven adaptive modulation schemes using field verification data.

2 System model

Without loss of generality, a node-to-node UAC link is considered for evaluation of the Q-learning adaptive modulation scheme, and the underwater acoustic channel is assumed to be stationary within the duration of each packet. The flowchart of the Q-learning adaptive modulation UWA acoustic communication system is shown in Figure 1.

The adaptive modulation acoustic communication system consists of two components: a single-input/single-output acoustic transceiver and an optimization process. For each period, N bits of information are generated and modulated into a packet signal. The signal is transmitted to the RX, which initially performs channel estimation to acquire the CSI knowledge \mathbf{h}_j , where $\mathbf{h}_j = [R, T_m, D]^T$ indicates the j th measured CSI feature that contains three CSI features, including the received SNR, the multipath time delay expansion length, and the Doppler shift value. The multipath delay expansion is defined as follows:

$$T_m = t_2 - t_1 \quad (1)$$

$$t_2 = \underset{\tau}{\operatorname{argmax}} A(\tau) \geq 0.3 \times A_{\max}(t) \quad (2)$$

where t_1 represents the time of the arrival of the direct signal, $A(t)$ indicates the matched filter output, and A_{\max} refers to the maximum peak amplitude of the matched filter output.

Then, the \mathbf{h}_j and the bit-error-rate (BER) is sent back to the TX. According to \mathbf{h}_j , an appropriate modulation mode $m_i \in M, i \in \{1, 2, \dots, I\}$ is selected under a specific policy π . The target of π is expressed as follows:

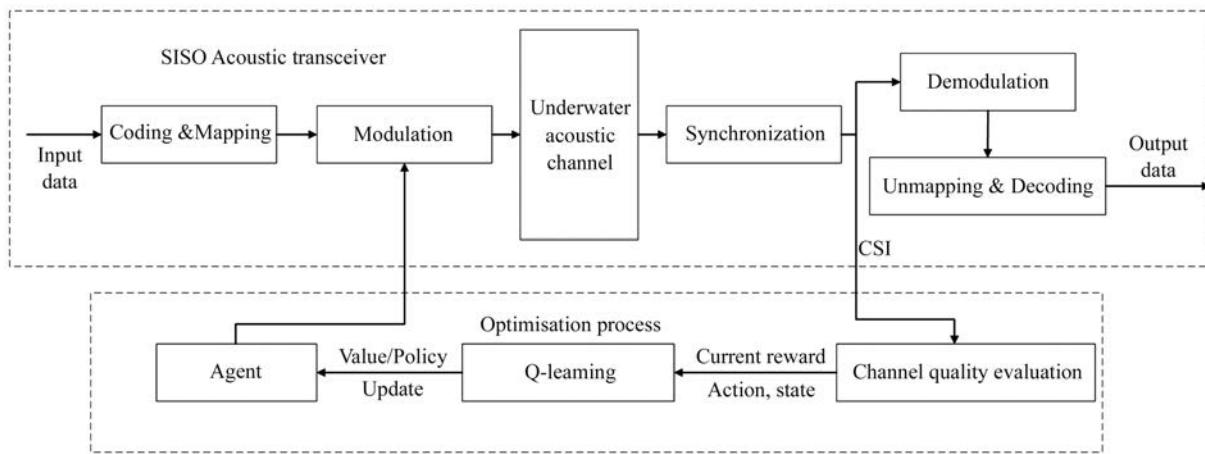


Figure 1 Illustration of Q-learning adaptive acoustic communication system

$$\begin{aligned} & \max \sum_{j=1}^{\infty} \max F_j^i, 1 \leq j \leq \infty, 1 \leq i \leq I \\ & \text{s.t. } e_j < 0.03 \end{aligned} \tag{3}$$

where $F_j^i = (1 - e_j) \times V^i$ indicates the throughput of the j th packet signal, e_j denotes the bit error rate of the j th signal, and V^i refers to the bit data rate of the m_i modulation mode.

We formulate the communication process as a Markov decision process and use Q-learning to solve the problem.

Then, the total throughput is defined by the following equation:

$$F_{\text{all}} = \sum_{k=1}^N F_k^i \tag{4}$$

where N indicates the total number of the transmitted packets, and k denotes the number of the transmitted packet.

The action space and state space are defined as $A = \{m_1, m_2, \dots, m_I\}$ and $S = \{h_1, h_2, \dots, h_j, \dots\}$, respectively. The immediate reward function is the instantaneous throughput of the receiver for any given state S_j and action A_j expressed as follows:

$$r(S_j, A_j) = dF_j^i \tag{5}$$

Parameter d equals 0 if transmission fails or 1 if transmission succeeds. The TX updates S_j when the feedback message is sent back and then calculates the instant reward using Equation (5).

Subsequently, we use the Q-learning algorithm to calculate the Q-values according to the current state and reward, as follows:

$$\begin{aligned} Q(S_j, A_j) & \leftarrow Q(S_j, A_j) + \\ & \alpha [r(S_j, A_j) + \gamma \max Q(S_{j+1}, A_{j+1}) - Q(S_j, A_j)] \end{aligned} \tag{6}$$

where $\alpha \in (0, 1)$ is a positive learning rate to control the weighting of current experience, and $\gamma \in (0, 1)$ indicates the discount factor.

The equation iteratively refines the Q-values by incorporating new experiences. The current estimate $Q(S_j, A_j)$ is updated toward a more accurate estimate that reflects the immediate reward $r(S_j, A_j)$ and the expected future rewards, discounted by γ .

This adjustment process incrementally aligns the Q-values with the true expected cumulative rewards. The term in square brackets represents the TD error, quantifying the discrepancy between the current Q-value and the new estimate based on observed rewards and estimated future rewards. This error guides the updates, ensuring Q-values adapt to deviations from expected outcomes.

Repeated applications of the update equation across numerous iterations drive the Q-values toward an optimal estimation of the action-value function. This iterative refinement is critical for the algorithm to accurately represent expected cumulative rewards, enabling the development of an effective policy.

Convergence occurs as updates decrease in magnitude, signaling that the Q-values are stabilizing near their optimal values. During exploration, the agent experiments with various actions, enriching its experiences and influencing Q-value updates. By contrast, exploitation relies on learned Q-values to select actions expected to maximize rewards.

To balance exploitation and exploration, the Boltzmann exploration strategy is used (Abdallah and Kaisers, 2016). The policy $\pi(S, A)$ is defined as a function of Q-values and τ , formulated as follows:

$$\pi(S_j, A_j) = \frac{e^{\frac{Q(S_j, A_j)}{\tau}}}{\sum_{A_i} e^{\frac{Q(S_j, A_i)}{\tau}}} \tag{7}$$

where $\pi(S_j, A_j)$ indicates the probability of selecting action A_j at state S_j , and τ denotes the temperature. While a higher τ suggests a more random strategy, a lower τ corresponds to a more greedy strategy. From Equation (7), we can observe that the action with higher Q-values will have a greater probability of being selected. Based on Equations (6) and (7), the transmission strategies for the next state can be optimized accordingly. The Q-table continuously iterates until it stabilizes.

During training, the channel state is influenced by multiple factors, preventing a strict mapping of modulation methods to specific states. Instead, an iterative training and trial-and-error process allows the agent to develop a learning strategy, enabling the selection of the most appropriate approach.

3 Experimental configuration

The field experiment was conducted in Wuyuan Bay, Xiamen, China. It is a semienclosed bay with an average depth of approximately 10 m. The signal was transmitted from a transceiver fixed at a depth of 5 m, and the transmitted signal was received and collected by a receiver suspended at the same depth, with a sampling rate of 75 kHz. The frequency bandwidth of the transducers was 13–18 kHz. The distance between the two terminals was 820 m, as shown in Figure 2. Remarkable impulse noise can be observed in the collected signal waveform due to frequent boat traffic passing by the experimental location, as illustrated in Figure 3.



Figure 2 Map of the field experiment

Three typical modulation modes were adopted as candidate modulation modes, namely, OFDM, MFSK, and DSSS for evaluating the performance of adaptive modulation. Among them, OFDM is suitable for high data rate transmission in favorable channel conditions, whereas DSSS

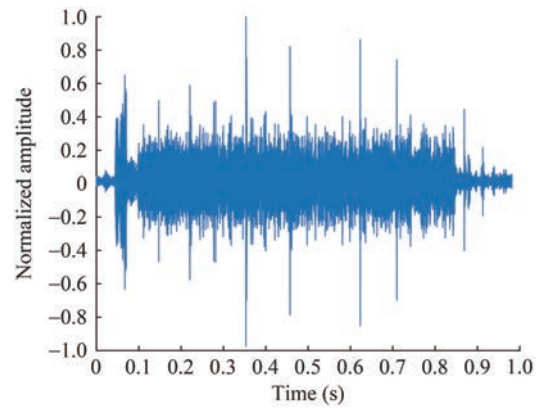


Figure 3 Waveform of experimental signal

offers high reliability at low data rates under adverse conditions. MFSK provides a balance between data rate and communication reliability.

Figure 4 illustrates the structure of the packet for UWA communication associated with different communication modes. This framework consists of a sync preamble, guard time, and information symbols. The sync preamble is composed of a linear frequency modulation signal, which can be used to obtain the SNR and multipath time delay expansion length. The SNR is calculated using Equation (8).

$$\gamma = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \tag{8}$$

where γ indicates the received SNR, P_{signal} denotes the signal power, and P_{noise} represents noise power.

The Doppler shift value and the multipath delay expansion length are obtained using the methods described in Li et al. (2019a, 2019b).

With the same configuration of sync preamble and guard time, the information symbols use the modulation schemes of OFDM, DSSS, and MFSK. The specific parameters for each communication mode are provided in Table 1.

In the sea experiment, three types of UWA communication signals were transmitted and collected every 10 min from 4:30 pm to 8:50 pm throughout the experiment duration. The channel impulse response measured during the experiment duration is shown in Figure 5. The figure demonstrates that the multipath pattern experienced evident variations.

4 Experimental results and analysis

The SNR, multipath spread, and Doppler shift (Wan

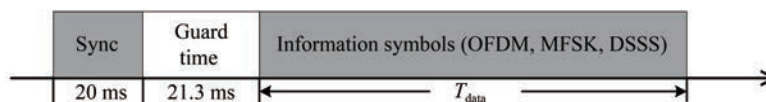


Figure 4 Structure of the packet

Table 1 Key modulation parameters of each communication mode

OFDM			DSSS			
Number of subcarriers	Data rate (bits/s)	T_{data} (ms)	Chip rate (chip/s)	Carrier frequency (kHz)	Data rate (bits/s)	T_{data} (s)
200	1 831	587.6	2 343	$f_c = 15.5$	55	14.028
MFSK						
Symbol duration (ms)		Number of frequency bin		Data rate (bits/s)		T_{data} (s)
$T = 13.65$		$M = 16$		188		6.458

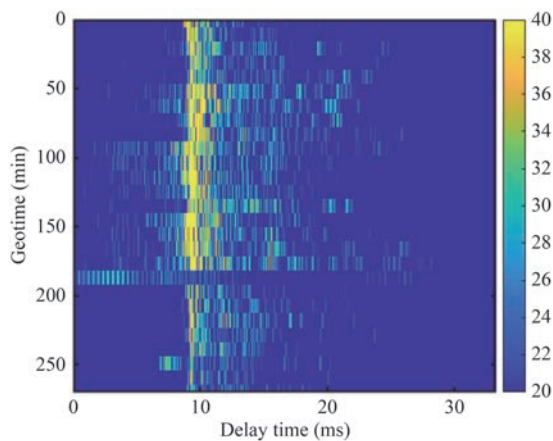
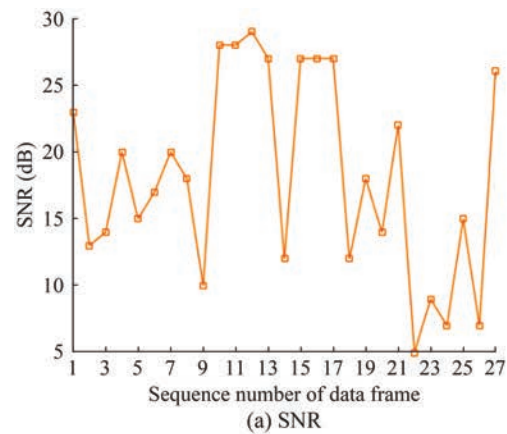


Figure 5 Estimated CIR during the experiment duration

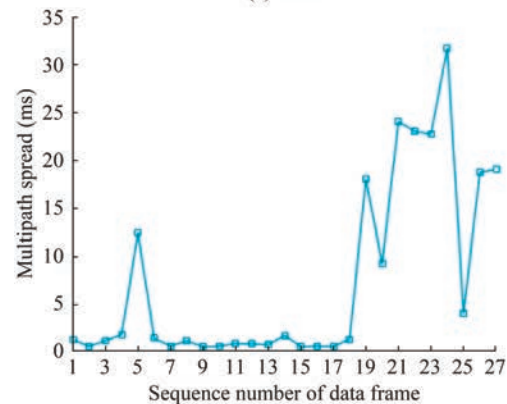
et al., 2012) measured during the experiment are provided in Figures 6(a), (b), and (c) respectively. On this basis, the CSI of the collected information data was calculated. As shown in Figure 6(a), the SNR of all packets exhibits a descending trend with considerable fluctuations, ranging from 5 to 29 dB. Meanwhile, the multipath spread is generally higher after the start of the 19th sequence, as revealed in Figure 6(b). The behavior of SNR and the multipath spread indicate that the channel quality deteriorated in the second half of the experiment duration. The reason for this finding is that the depth of Wuyuan Bay decreased due to the tide. Notably, compared with SNR and multipath spread, the Doppler value appeared more stochastic in Figure 6(c). Given that the transducers at both sides were fixed to wooden stacks with ropes, a smaller relative motion was observed, resulting in overall low Doppler values for all data packets.

According to each packet’s CSI, 27 states can be obtained during the experiment, namely, h_1, h_2, \dots, h_{27} . To optimize the transmission mode and enable adaptive modulation, the Q-learning algorithm, as described in Section 2, is adopted to train the collected data. The discount factor γ , learning rate α , and temperature parameter τ are set to 0.1, 0.1, and 100, respectively.

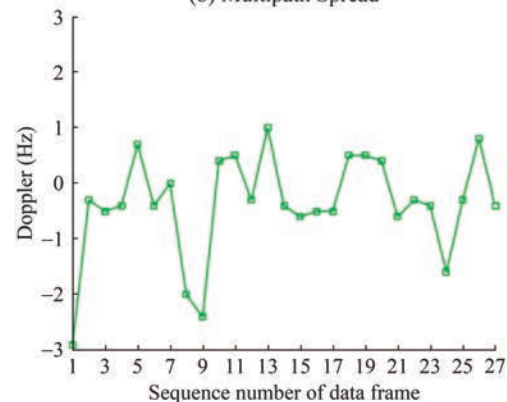
For performance comparison with respect to the Q-learning scheme (Scheme 1), two other conventional UWA transmission selection schemes, i.e., the fixed threshold adaptive modulation scheme (Barua et al., 2022) (Scheme 2)



(a) SNR



(b) Multipath Spread



(c) Doppler

Figure 6 CSI of all data packets

and randomly choosing adaptive transmission scheme (Wang et al., 2018) (Scheme 3), were adopted. For Scheme 2, the

assumption is that the modulation mode can be efficiently selected based on the feedback CSI and the pre-defined threshold, indicating that the selected mode matches the state from the previous moment. In addition to the three adaptive modulation schemes, each individual modulation mode was adopted for performance comparison.

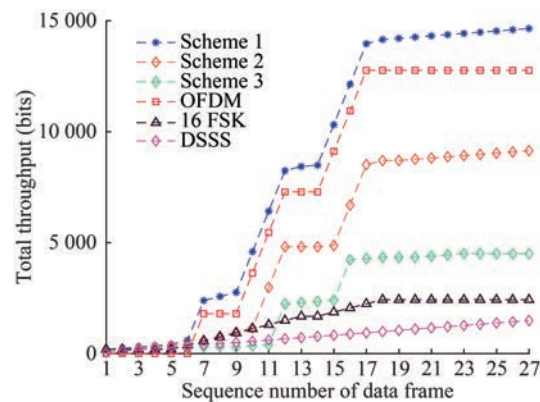
The performance of different single modulation modes, as well as adaptive modulation strategies, is shown in Figures 7(a) and 7(b), respectively. The figures illustrate that Scheme 1 achieves the highest throughput F_{all} , corresponding to a total throughput of 14 645.3 bits, which is 14.8% higher than that of the second highest throughput achieved using the single OFDM mode at 12 762.1 bits. A sudden increase in throughput was observed at certain sequences because the mode has been transformed into the OFDM, whose bit rate is much higher than the other. The total throughput of Schemes 2 and 3 is 9 139.4 and 8 332.3 bits, respectively. The 16 FSK and DSSS modes produce the lowest throughput of 2 425.9 and 1 485 bits, respectively.

However, in terms of communication reliability, as indicated by the BER, although the single OFDM mode achieves the second highest total throughput, it generated the worst BER results for most of the experimental duration, as indicated in Figure 7(b). By contrast, Scheme 1 and the single DSSS mode yielded the best BER performance. The single MFSK corresponded to an increasing BER as the channel quality degraded after packet sequence 19.

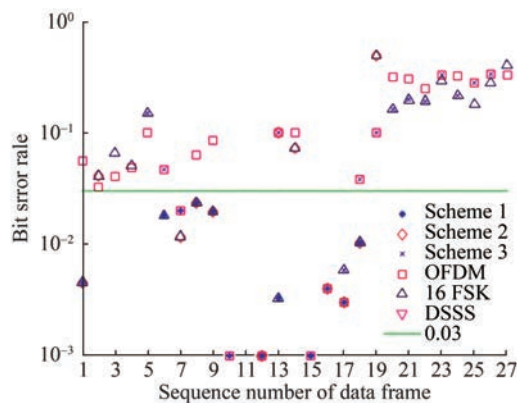
To specify the performance comparison among the three adaptive modulation schemes, the modulation mode selected for each transmission packet is shown in Figure 8. The figure demonstrates that the switching of transmission modes is dependent on the variation of the channel state. Specifically, under large Doppler shifts or low SNR, the transmission strategy tends to switch from OFDM to MFSK. However, when the multipath spread exceeds a certain threshold, or the SNR becomes extremely low, resulting in poor performance of OFDM and MFSK, the transmission strategy switches to the most robust option, namely, DSSS.

The channel conditions from the previous period no longer satisfy the modulation strategy transmission conditions of the current moment due to the channel's variation during the experiment. For Scheme 1, the RL algorithm was used to select the appropriate action for the next moment based on the current reward and Q value of the previous moment. This approach is applicable when dealing with a specific sea area.

In the case of Scheme 2, the mode was selected using the CSI from the previous moment. Thus, the transmission strategy may not be suitable for the next moment, thereby deteriorating the performance. In Scheme 3, as the transmission strategy was randomly selected, the corresponding performance experiences significant fluctuations, as revealed in Figure 7.



(a) Total throughput



(b) BER

Figure 7 Performance of each scheme under actual sea trial data experiments

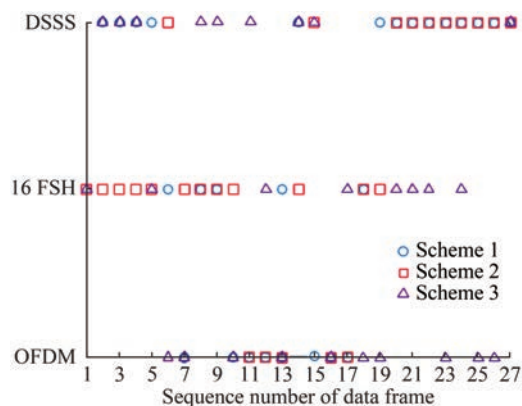


Figure 8 Selected transmission strategies in different schemes

5 Conclusions

While reinforcement learning-based adaptive modulation has been recognized as a promising approach to address the challenges posed by harsh UWA channels, experimental investigations from the perspective of field data-driven evaluation and comparison are limited. This gap causes difficulty in quantifying its performance improvement compared with traditional single communication modes or

conventional adaptive modulation strategies.

In this study, an at-sea RL adaptive modulation UWA communication experiment is conducted to evaluate the practical effectiveness of RL in optimizing adaptive modulation UWA communication.

Specifically, the experimental adaptive modulation UWA communication system includes three types of communication modes, with optimal switching driven by the Q-learning strategy according to the quantitative channel state information. Meanwhile, traditional single-mode modulations and adaptive modulation schemes are used as reference points for performance comparison.

The practical performance analysis and comparison verified that for the specific shallow water channel that exhibits evident time delay/Doppler variations, reinforcement learning-based underwater acoustic adaptive communication outperformed other single-mode or adaptive modulation approaches. This method shows strong potential for application in adverse shallow water channels. In the future, more comprehensive field experiments will be conducted to further evaluate its practical performance.

Acknowledgement The authors are grateful for funding from the National Key Research and Development Program of China (No. 2018YFE0110000), the National Natural Science Foundation of China (No. 11274259, No. 11574258), and the Science and Technology Commission Foundation of Shanghai (21DZ1205500) in support of the present research.

Competing interest Feng Tong is an editorial board member for the Journal of Marine Science and Application and was not involved in the editorial review, or the decision to publish this article. All authors declare that there are no other competing interests.

References

- Abdallah S, Kaisers M (2016) Addressing environment non-stationarity by repeating Q-learning updates. *Journal of Machine Learning Research* 17(46): 1-31
- Barua S, Rong Y, Nordholm S, Chen P (2022) Real-time adaptive modulation schemes for underwater acoustic OFDM communication. *Sensors* 22(9): 3436. <https://doi.org/10.3390/s22093436>
- Benson A, Proakis J, Stojanovic M (2000) Towards robust adaptive acoustic communications. In: OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No. 00CH37158). Piscataway: IEEE, 1243-1249
- Bhopale P, Kazi F, Singh N (2019) Reinforcement learning based obstacle avoidance for autonomous underwater vehicle. *Journal of Marine Science and Application* 18: 228-238. <https://doi.org/10.1007/s11804-019-00089-3>
- Bulut S, Ergin S (2021) Effects of temperature, salinity, and fluid type on acoustic characteristics of turbulent flow around circular cylinder. *Journal of Marine Science and Application* 20: 213-228. <https://doi.org/10.1007/s11804-021-00197-z>
- Fan C, Wang Z (2021) Adaptive switching for multimodal underwater acoustic communications based on reinforcement learning. In: The 15th International Conference on Underwater Networks & Systems. ACM, Shenzhen 1-2
- Fu Q, Song A (2018) Adaptive modulation for underwater acoustic communications based on reinforcement learning. In: OCEANS 2018 MTS/IEEE Charleston. Piscataway: IEEE, 1-8
- Gussen CMG, Diniz PSR, Campos MLR, Martins WA, Costa FM, Gois JN (2016) A survey of underwater wireless communication technologies. *Journal of Communication and Information Systems* 31(1): 242-255. <https://doi.org/10.14209/jcis.2016.22>
- Huang L, Wang Y, Zhang Q, Han J, Tan W, Tian Z (2022) Machine learning for underwater acoustic communications. *IEEE Wireless Communications* 29(3): 102-108. <https://doi.org/10.1109/MWC.2020.2000284>
- Huang J, Diamant R (2020) Adaptive modulation for long-range underwater acoustic communication. *IEEE Transactions on Wireless Communications* 19(10): 6844-6857. <https://doi.org/10.1109/TWC.2020.3006230>
- Huang JG, Wang H, He CB, Zhang QF, Jing LY (2018) Underwater acoustic communication and the general performance evaluation criteria. *Frontiers of Information Technology & Electronic Engineering* 19: 951-971. <https://doi.org/10.1631/FITEE.1700775>
- Li B, Zheng S, Tong F (2019a) Bit-error rate based Doppler estimation for shallow water acoustic OFDM communication. *Ocean Engineering* 182: 203-210. <https://doi.org/10.1016/j.oceaneng.2019.04.045>
- Li G, Wu J, Tang T, Chen Z, Chen J, Liu H (2019b) Underwater acoustic time delay estimation based on envelope differences of correlation functions. *Sensors* 19(5): 1259. <https://doi.org/10.3390/s19051259>
- Mani S, Duman TM, Hursky P (2008) Adaptive coding/modulation for shallow-water UWA communications. *Journal of the Acoustical Society of America* 123: 3749-3749. <https://doi.org/10.1121/1.2935305>
- Radosevic A, Ahmed R, Duman TM, Proaki JG, Stojanovic M (2014) Adaptive OFDM modulation for underwater acoustic communications: design considerations and experimental results. *IEEE Journal of Oceanic Engineering* 39(2): 357-370. <https://doi.org/10.1109/JOE.2013.2253212>
- Stojanovic M, Preisig J (2009) Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Communications Magazine* 47(1): 84-89. <https://doi.org/10.1109/MCOM.2009.4752682>
- Su W, Lin J, Chen K, Xiao L, Em C (2019) Reinforcement learning-based adaptive modulation and coding for efficient underwater communications. *IEEE Access* 7: 67539-67550. <https://doi.org/10.1109/ACCESS.2019.2918506>
- Su Y, Liu Y, Fan R, et al (2022) A cooperative jamming scheme based on node authentication for underwater acoustic sensor networks. *Journal of Marine Science and Application* 21: 197-209. <https://doi.org/10.1007/s11804-022-00277-8>
- Tang N, Zeng Q, Luo D, Xu Q, Hu H (2020) Research on development and application of underwater acoustic communication system. *Journal of Physics: Conference Series* 1617: 012036. <https://doi.org/10.1088/1742-6596/1617/1/012036>
- Sweta T, Ruthrapriya S, Sneka J, John SRA, Rohith G, Mangal D (2024) Reinforcement learning-based automated modulation switching algorithm for an enhanced underwater acoustic communication. *Results in Engineering* 23: 102791. <https://doi.org/10.1016/j.rineng.2024.102791>
- Wan L, Wang Z, Zhou S, Yang TC, Shi Z (2012) Performance comparison of doppler scale estimation methods for underwater acoustic OFDM. *Journal of Electrical and Computer Engineering*

- 2012: 1-11. <https://doi.org/10.1155/2012/703243>
- Wan L, Zhou H, Xu X, Huang Y, Zhou S, Shi Z (2015) Adaptive modulation and coding for underwater acoustic OFDM. *IEEE Journal of Oceanic Engineering* 40(2): 327-336. <https://doi.org/10.1109/JOE.2014.2323365>
- Wang C, Wang Z, Sun W, Fuhrmann DR (2018) Reinforcement learning-based adaptive transmission in time-varying underwater acoustic channels. *IEEE Access* 6: 2541-2558. <https://doi.org/10.1109/ACCESS.2017.2784239>
- Zhang Y, Zhang Z, Chen L, Wang X (2021) Reinforcement Learning-Based Opportunistic Routing Protocol for Underwater Acoustic Sensor Networks. *IEEE Transactions on Vehicular Technology* 70(3): 2756-2770. <https://doi.org/10.1109/TVT.2021.3058282>
- Zhu Z, Tong F, Zhou Y, Zhang Z, Zhang F (2023) Deep learning prediction of time-varying underwater acoustic channel based on LSTM with attention mechanism. *Journal of Marine Science and Application* 22: 650-658. <https://doi.org/10.1007/s11804-023-00347-5>