Unmanned Systems, Vol. 14, No. 3 (2026) 549–560

DOI: 10.1142/S2301385026410025

UNMANNED SYSTEMS
https://www.worldscientific.com/worldscinet/us

Large-Scale Ocean Environment Field Perception for Underwater Robots with Limited Measurements

Lei Lei [*], Xi Chen [†], Ben M. Chen [‡]

Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong,
Hong Kong, P. R. China

Accurate perception of large-scale ocean environment fields is essential for marine monitoring, resource management, and climate research. However, direct observation is severely limited by the sparse, asynchronous, and resource-constrained measurements with mobile platforms such as underwater gliders and robots. To overcome this challenge, we propose a perception framework tailored for underwater robots operating with limited measurements. The method integrates three components: spatiotemporal decomposition to obtain compact and interpretable ocean features; multi-objective sparse sampling that jointly optimizes information gain, spatial coverage, and depth anchors; and spatiotemporal fusion to reconstruct full fields from sparse in-situ data. Experiments on a Pacific Ocean temperature dataset confirm that with only 1% sampling sparsity, the reconstructed fields achieve an average error below 0.5°C, while preserving thermocline variability and mesoscale structures. These results demonstrate that the proposed method enables efficient, accurate, and scalable ocean environment perception, offering significant potential for future autonomous missions with underwater robots.

Keywords: Environment perception; sparse measurement; environment uncertainty; mobile platforms; large-scale reconstruction.

1. Introduction

Oceans cover more than two-thirds of the Earth's surface and play a central role in regulating climate, sustaining biodiversity, and supporting socio-economic activities [1, 2]. While satellites provide wide-area coverage of surface variables [3] and in-situ networks such as Argo deliver basin-scale vertical profiles [4], the ocean interior remains sparsely observed in space and time, particularly below the mixed layer. In recent years, mobile platforms, including autonomous surface vessels and autonomous underwater vehicles or gliders [5–7], have expanded access to remote regions by enabling adaptive, in-situ sampling along long trajectories under tight resource budgets [8, 9]. However, their measurements are inherently sparse, asynchronous, and constrained by energy, navigation, and safety, making direct full-field observation impractical and motivating the development of methods for reconstructing large-scale environment fields from limited observations.

A central component of ocean environment perception is the construction of spatiotemporal models capable of capturing the evolution of high-dimensional fields from incomplete data. Existing approaches can be broadly categorized into physics-based and data-driven methods. Physics-based methods, grounded in numerical models and data assimilation schemes, combine sparse observations with governing dynamical equations to reconstruct full-field states [10, 11]. While they ensure physical consistency can generate multi-variable outputs, they are computationally demanding and depend heavily on uncertain parameterizations. In contrast, data-driven methods rely on statistical inference and machine learning to approximate field structures directly from observations. Classical approaches, such as

Received 2 September 2025; Accepted 17 November 2025; Published 8 December 2025. This paper was recommended for publication in its revised form by Special Issue Editors, Jie Chen, Ben M Chen, and Zongli Lin
Email Addresses: [*]*leilei001@cuhk.edu.hk*, [†]*xichen002@cuhk.edu.hk*, [‡]*bmchen@cuhk.edu.hk*
[†]Corresponding author.

empirical orthogonal functions and dynamic mode decomposition, extract dominant spatiotemporal modes from historical datasets [12, 13], while recent advances employ Gaussian processes, deep learning, and implicit neural representations to capture nonlinear correlations and improve reconstruction fidelity [14–16]. Although these methods offer flexibility and computational efficiency, they remain highly dependent on large-scale training data and often lack interpretability, raising concerns about transparency and generalizability in scientific applications.

In parallel, a growing body of research has explored sparse sampling strategies to maximize the value of limited measurements. Information-theoretic approaches, such as mutual information maximization and entropy reduction, prioritize observations that yield the greatest expected information gain [17–19]. Statistical and optimization-based methods adopt experimental design criteria, including A-optimality and D-optimality, to minimize reconstruction uncertainty [20, 21]. In oceanography, these concepts have been extended to adaptive path planning, where gliders or AUVs are directed to collect measurements that enhance coverage or reduce forecast errors [22–24]. While such methods improve sampling efficiency, they are often tailored to specific goals in isolation such as variance reduction, coverage, or trajectory feasibility. Moreover, depth-dependent structures, including thermoclines and oxygen minimum zones, are rarely incorporated [25, 26], even though they critically constrain three-dimensional reconstructions.

To balance accuracy with operational feasibility, recent studies highlight the importance of multi-objective optimization in sampling design. Among the most widely adopted objectives are information gain and uncertainty reduction, which directly influence the quality of reconstructed fields [27, 28]. Spatial coverage is another critical target, ensuring that samples are well distributed across large basins rather than clustered in localized regions [29, 30]. Beyond horizontal distribution, vertical structure constraints are increasingly recognized as essential, since key layers such as thermoclines, oxygen minimum zones, and deep-water masses shape ocean dynamics [31]. However, explicit depth-anchor constraints are seldom included, and vertical sampling often remains secondary.

Motivated by these challenges, we develop a large-scale ocean environment perception framework that directly addresses the constraints of underwater robots operating with sparse and limited measurements. Instead of relying on dense observation or computationally prohibitive models, the framework builds on three complementary ideas. First, a spatiotemporal decomposition is introduced to distill dominant modes of the ocean field, yielding a compact and interpretable representation of complex dynamics. Second, we design a multi-objective sampling strategy that balances information gain, spatial coverage, and depth-anchor constraints, ensuring that the limited sensor resources of mobile platforms are deployed in the most effective way. Finally, a spatiotemporal fusion scheme is employed to reconstruct full-field temperature structures by coherently integrating spatial patterns with temporal evolution. Evaluation on Pacific Ocean temperature data demonstrates the effectiveness of this approach. Even at a sampling sparsity of only 1%, the framework reconstructs three-dimensional fields with an average error below 0.5°C, while preserving key features such as thermocline variability and mesoscale dynamics. These results highlight not only the efficiency of the method but also its suitability for practical deployment, where underwater robots must operate under strict endurance and sensing constraints.

The remainder of this paper is structured as follows: Sec. 2 formalizes the problem. Section 3 introduces the proposed framework in detail. Section 4 presents experimental evaluations on Pacific Ocean data. Section 5 concludes the paper with discussions on implications for future autonomous systems.

2. Problem Description

The underlying dynamics of ocean environments can be described by $\mathcal{X} \in \mathbb{R}^{N_x \times N_y \times N_z \times N_t}$, where N_x, N_y, and N_z are the spatial resolutions along longitude, latitude, and depth, and N_t is the number of time steps. Each element $\boldsymbol{x}_{i,j,k,t}$ corresponds to the value of the environmental variable at location (i, j, k) and time t. Vectorizing space into $N_s = N_x N_y N_z$ locations yields

$$\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_t}] \in \mathbb{R}^{N_s \times N_t},$$

where $\boldsymbol{x}_t \in \mathbb{R}^{N_s}$ denotes the full snapshot at time t.

Due to limited sensing resources and platform endurance, only a small subset of locations can be observed. Let $\Omega \subseteq \{1, \dots, N_s\}$ denote the selected sensor set, with $|\Omega| = M \ll N_s$. The sampling operator $\mathcal{P}_\Omega \in \{0,1\}^{M \times N_s}$ extracts the measurements at Ω, yielding

$$\boldsymbol{y}_t = \mathcal{P}_\Omega \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, N_t, \tag{1}$$

where $\boldsymbol{y}_t \in \mathbb{R}^M$ is the measurement vector and $\boldsymbol{\varepsilon}_t$ represents sensor noise. Unlike full-field data, $\{\boldsymbol{y}_t\}$ are sparse, non-uniform, and insufficient for direct analysis.

The objective is to reconstruct the full spatiotemporal field $\widehat{\boldsymbol{X}}$ from sparse measurements $\{\boldsymbol{y}_t\}_{t=1}^{N_t}$ by designing: (i) a compact spatiotemporal representation of $\boldsymbol{X}$, (ii) a one-shot sensor layout Ω chosen to maximize information gain, coverage, and depth constraints under budget M, and (iii) a reconstruction mapping $\mathcal{F}$ that fuses sparse observations with the representation:

$$\widehat{\boldsymbol{X}} = \mathcal{F}(\{\boldsymbol{y}_t, \Omega\}_{t=1}^{N_t}). \tag{2}$$

Formally, the reconstruction problem is posed as

$$\min_{\mathcal{F},\Omega} \sum_{t=1}^{N_t} \ell(\widehat{\boldsymbol{x}}_t, \boldsymbol{x}_t)$$
$$\text{s.t.} \quad \boldsymbol{y}_t = \mathcal{P}_\Omega \boldsymbol{x}_t, \ |\Omega| = M, \tag{3}$$

where ℓ is a reconstruction loss. The goal is to recover $\boldsymbol{X}$ with minimal error given sparse and fixed measurements.

3. Methodology

3.1. *Framework*

The proposed framework integrates spatiotemporal decomposition, sparse location optimization, and spatiotemporal fusion into a unified pipeline for reconstructing ocean environment fields from limited measurements, as illustrated in Fig. 1.

- The high-dimensional field $\boldsymbol{X}$ is first decomposed into a set of dominant spatial modes $\boldsymbol{\Phi}$ and their temporal coefficients $\boldsymbol{A}$, yielding a compact low-rank representation that captures the essential background structures and dynamic variations of the environment.
- Based on the extracted spatial modes, a one-time sensor layout Ω^* is optimized under a fixed budget M. The selection jointly maximizes information gain, spatial coverage, and depth-anchor constraints, ensuring that mobile platforms collect representative and stable observations despite limited endurance and payload capacity.
- During operation, sparse measurements $\boldsymbol{y}_t$ collected at Ω^* are mapped into the reduced coefficient space, estimated via regularized least squares, and subsequently fused with $\boldsymbol{\Phi}$ to reconstruct the full spatiotemporal field $\widehat{\boldsymbol{X}}$.

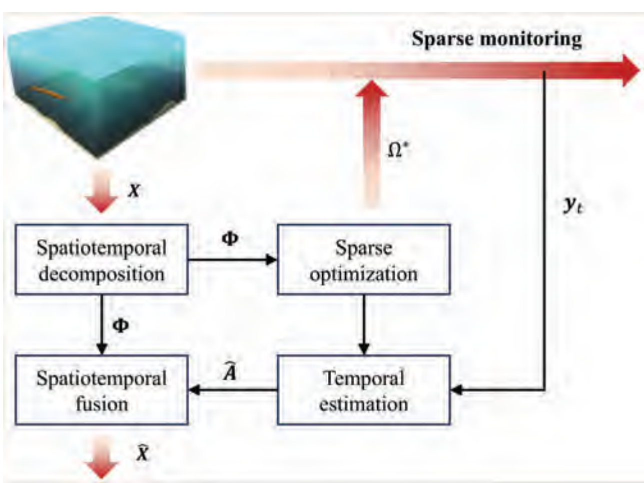


Fig. 1. Proposed framework for large-scale ocean field perception. The method consists of three stages: (i) spatiotemporal decomposition, (ii) sparse location optimization, and (iii) spatiotemporal fusion.

3.2. *Spatiotemporal decomposition*

The spatiotemporal field $\boldsymbol{X} \in \mathbb{R}^{N_s \times N_t}$ is high-dimensional, where N_s is the number of spatial grid points and N_t is the number of time steps. Direct manipulation of $\boldsymbol{X}$ is computationally prohibitive and sensitive to noise. To obtain a compact and interpretable representation, we perform spatiotemporal decoupling by separating spatial structures from their temporal evolutions.

We seek a rank-R factorization of the form

$$\boldsymbol{X} \approx \boldsymbol{\Phi}\boldsymbol{A} = \sum_{r=1}^{R} \boldsymbol{\phi}_r \boldsymbol{a}_r^\top, \tag{4}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_R] \in \mathbb{R}^{N_s \times R}$ contains R dominant spatial modes, $\boldsymbol{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_R]^\top \in \mathbb{R}^{R \times N_t}$ stores the temporal coefficients, $R \ll \min(N_s, N_t)$ ensures dimensionality reduction and denoising.

Thus, each snapshot $\boldsymbol{x}_t$ can be approximated as a linear combination

$$\boldsymbol{x}_t \approx \sum_{r=1}^{R} a_r(t)\, \boldsymbol{\phi}_r, \tag{5}$$

where $\boldsymbol{\phi}_r$ captures a recurrent spatial pattern (e.g. basin-scale circulation, vertical gradients), and $a_r(t)$ tracks its temporal activation.

To obtain the decomposition, singular value decomposition is applied:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top, \tag{6}$$

where $\boldsymbol{U} \in \mathbb{R}^{N_s \times r_{\max}}$ contains orthonormal spatial bases, $\boldsymbol{V} \in \mathbb{R}^{N_t \times r_{\max}}$ contains temporal patterns, and $\boldsymbol{\Sigma}$ is diagonal with singular values $\{\sigma_i\}$.

The decomposition can be interpreted as

$$\boldsymbol{X} = \sum_{r=1}^{R} \sigma_r\, \boldsymbol{u}_r \boldsymbol{v}_r^\top, \tag{7}$$

where $\boldsymbol{u}_r$ is the rth column of $\boldsymbol{U}$ and $\boldsymbol{v}_r$ is the rth column of $\boldsymbol{V}$.

Each $\boldsymbol{u}_r \in \mathbb{R}^{N_s}$ represents a dominant spatial structure of the field. Collecting the first R modes gives

$$\boldsymbol{\Phi} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_R] \in \mathbb{R}^{N_s \times R}. \tag{8}$$

The temporal evolution of each mode is obtained from $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$. Specifically, the rth temporal coefficient series is

$$\boldsymbol{a}_r = \sigma_r \boldsymbol{v}_r \in \mathbb{R}^{N_t}, \tag{9}$$

where $\boldsymbol{v}_r$ is the rth column of $\boldsymbol{V}$. Stacking across all modes yields

$$\boldsymbol{A} = \boldsymbol{\Sigma}_R \boldsymbol{V}_R^\top \in \mathbb{R}^{R \times N_t}, \tag{10}$$

where $\boldsymbol{V}_R$ denotes the first R columns of $\boldsymbol{V}$, and $\boldsymbol{\Sigma}_R$ is the truncated diagonal matrix with the top R singular values.

The rank R is chosen by an energy threshold η

$$R^* = \min \left\{ R\colon \frac{\sum_{i=1}^{R} \sigma_i^2}{\sum_{i=1}^{r_{\max}} \sigma_i^2} \geq \eta \right\}. \tag{11}$$

This guarantees that the retained modes capture a prescribed portion of the field variance.

3.3. *Sparse location optimization*

The accuracy of spatiotemporal reconstruction depends critically on the selection of observation sites. Since only $M \ll N_s$ sensors can be deployed, the chosen sensor set Ω must maximize the information content of the measurements while ensuring physical representativeness and numerical stability.

Given the spatial basis $\boldsymbol{\Phi} \in \mathbb{R}^{N_s \times R}$ obtained from decomposition, the measurement at time t is

$$\boldsymbol{y}_t = \mathcal{P}_\Omega \boldsymbol{x}_t \approx \boldsymbol{\Phi}_\Omega \boldsymbol{a}(t), \tag{12}$$

where $\mathcal{P}_\Omega \in \{0,1\}^{M \times N_s}$ extracts rows of $\boldsymbol{\Phi}$ corresponding to Ω, and $\boldsymbol{\Phi}_\Omega \in \mathbb{R}^{M \times R}$ is the reduced sensing matrix. The invertibility and conditioning of $\boldsymbol{\Phi}_\Omega$ directly affect the stability of coefficient recovery and thus the quality of reconstruction.

We cast the sensor placement problem as a multi-objective optimization:

$$\begin{aligned} \Omega^* = \arg \max_{\substack{\Omega \subset \mathcal{S} \\ |\Omega| = M}} & \underbrace{\log\det(\boldsymbol{\Phi}_\Omega^\top \boldsymbol{\Phi}_\Omega + \gamma \boldsymbol{I})}_{\text{information}} \\ & + \alpha_1 \underbrace{\mathcal{G}(\Omega)}_{\text{coverage}} + \alpha_2 \underbrace{\sum_{i \in \Omega} s(i)}_{\text{anchors}}. \end{aligned} \tag{13}$$

The three terms play complementary roles:

- Information gain: The log-determinant term is analogous to a D-optimal design criterion, ensuring that the reduced matrix $\boldsymbol{\Phi}_\Omega$ is well-conditioned and supports accurate coefficient recovery.
- Spatial coverage: $\mathcal{G}(\Omega)$ promotes dispersion of sensors in the horizontal plane, preventing redundant measurements in clustered regions and improving basin-scale representativeness. A typical choice is the minimum pairwise distance under anisotropic scaling:

$$\mathcal{G}(\Omega) = \min_{\substack{i \neq j \\ i,j \in \Omega}} \|\boldsymbol{r}(i) - \boldsymbol{r}(j)\|_{\boldsymbol{D}}, \tag{14}$$

$$\|\boldsymbol{u}\|_{\boldsymbol{D}}^2 = \boldsymbol{u}^\top \operatorname{diag}(d_x^2, d_y^2, d_z^2) \boldsymbol{u}. \tag{15}$$

- Anchor point: $s(i) \geq 0$ encodes the importance of placing sensors in slowly changing areas to capture vertical heterogeneity and provide an anchor point for robustness.

The weights α_1, α_2 balance the emphasis among information, coverage, and vertical representation.

The optimization in (13) is NP-hard, as it involves combinatorial subset selection. Exhaustive search is infeasible when N_s is large. In practice, we adopt a greedy strategy: starting from an empty set, iteratively add the location that yields the maximum increase in the objective. This approach offers near-optimal guarantees under submodularity assumptions and scales efficiently to large domains.

3.4. *Spatiotemporal fusion*

Once the spatial basis $\boldsymbol{\Phi}$ and sensor layout Ω^* are determined, the final step is to reconstruct the full spatiotemporal field from the sparse observations. This process fuses spatial modes with temporal dynamics in order to obtain coherent, full-field reconstructions.

At each time step t, the measurements are given by

$$\boldsymbol{y}_t = \mathcal{P}_{\Omega^*} \boldsymbol{x}_t \approx \boldsymbol{\Phi}_{\Omega^*} \boldsymbol{a}(t), \tag{16}$$

where $\boldsymbol{\Phi}_{\Omega^*} \in \mathbb{R}^{M \times R}$ is the restriction of the spatial modes to the chosen sensor sites. To recover the temporal coefficients $\boldsymbol{a}(t) \in \mathbb{R}^R$, we solve a regularized least-squares problem:

$$\boldsymbol{a}^*(t) = \arg\min_{\boldsymbol{a}} \ \|\boldsymbol{y}_t - \boldsymbol{\Phi}_{\Omega^*} \boldsymbol{a}\|_2^2 + \gamma \|\boldsymbol{a}\|_2^2 \tag{17}$$

with closed-form solution

$$\boldsymbol{a}^*(t) = \left(\boldsymbol{\Phi}_{\Omega^*}^\top \boldsymbol{\Phi}_{\Omega^*} + \gamma \boldsymbol{I}\right)^{-1} \boldsymbol{\Phi}_{\Omega^*}^\top \boldsymbol{y}_t, \tag{18}$$

where $\gamma > 0$ is a regularization parameter that balances noise suppression and fidelity.

Given the estimated coefficients $\boldsymbol{a}^*(t)$, the full-field snapshot is reconstructed as

$$\widehat{\boldsymbol{x}}_t = \boldsymbol{\Phi} \boldsymbol{a}^*(t), \quad t = 1, \dots, N_t. \tag{19}$$

Stacking across all time steps yields the reconstructed field

$$\widehat{\boldsymbol{X}} = \boldsymbol{\Phi} \widehat{\boldsymbol{A}}, \quad \widehat{\boldsymbol{A}} = [\boldsymbol{a}^*(1), \boldsymbol{a}^*(2), \dots, \boldsymbol{a}^*(N_t)]. \tag{20}$$

3.5. *Theoretical analysis*

3.5.1. *Observability*

The reconstruction relies on estimating temporal coefficients $\boldsymbol{a}(t)$ from sparse measurements:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_\Omega \boldsymbol{a}(t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\Phi}_\Omega \in \mathbb{R}^{M \times R}. \tag{21}$$

For unique recovery, $\boldsymbol{\Phi}_\Omega$ must have full column rank, i.e.

$$\operatorname{rank}(\boldsymbol{\Phi}_\Omega) = R. \tag{22}$$

Remark 3.1. The number of sensors must satisfy $M \geq R$ to ensure observability of the reduced-order dynamics. Moreover, the robustness of the inversion depends on the conditioning of $\boldsymbol{\Phi}_\Omega$, as the error amplification factor scales with its condition number $\kappa(\boldsymbol{\Phi}_\Omega)$.

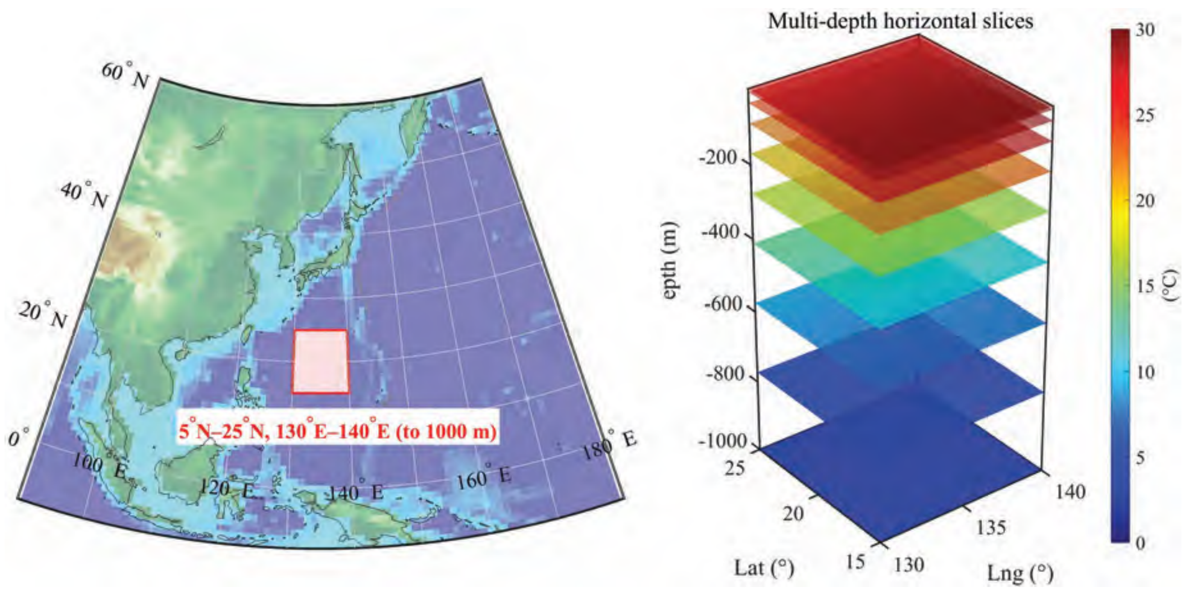


Fig. 2. Experimental area and temperature spatiotemporal distribution.

3.5.2. *Convergence of coefficient estimation*

The coefficients are estimated via regularized least squares:

$$\boldsymbol{a}^*(t) = (\boldsymbol{\Phi}_\Omega^\top \boldsymbol{\Phi}_\Omega + \gamma \boldsymbol{I})^{-1} \boldsymbol{\Phi}_\Omega^\top \boldsymbol{y}_t. \tag{23}$$

Let $\boldsymbol{a}^\dagger(t)$ be the true coefficients, and assume additive noise $\boldsymbol{\varepsilon}_t$ with variance σ^2. The estimation error satisfies:

$$\|\boldsymbol{a}^*(t) - \boldsymbol{a}^\dagger(t)\|_2 \le \kappa(\boldsymbol{\Phi}_\Omega)\, \|\boldsymbol{\varepsilon}_t\|_2 + \mathcal{O}(\gamma), \tag{24}$$

where $\kappa(\boldsymbol{\Phi}_\Omega)$ denotes the condition number of the sensing matrix.

This relation indicates that the convergence rate of the estimator depends linearly on the noise level and the regularization strength. A smaller γ reduces bias but increases sensitivity to noise, while a larger γ enhances stability at the cost of bias. In practice, γ is selected to balance these effects so that the total mean-square error is minimized. When $\boldsymbol{\Phi}_\Omega$ is well conditioned and both γ and σ^2 approach zero, the bias and noise terms vanish, leading to asymptotic convergence $\boldsymbol{a}^*(t) \to \boldsymbol{a}^\dagger(t)$. Therefore, the convergence and stability of coefficient estimation are jointly governed by the conditioning of $\boldsymbol{\Phi}_\Omega$ and the proper choice of γ.

3.5.3. *Error bound for reconstruction*

The reconstructed field $\widehat{\boldsymbol{x}}_t = \boldsymbol{\Phi} \boldsymbol{a}^*(t)$ yields error

$$\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2 \le \|\boldsymbol{\Phi}\|_2 \cdot \|\boldsymbol{a}^*(t) - \boldsymbol{a}^\dagger(t)\|_2. \tag{25}$$

Remark 3.2. The reconstruction accuracy depends on two factors: the representational capacity of $\boldsymbol{\Phi}$ (mode truncation) and the stability of inversion through $\boldsymbol{\Phi}_\Omega$.

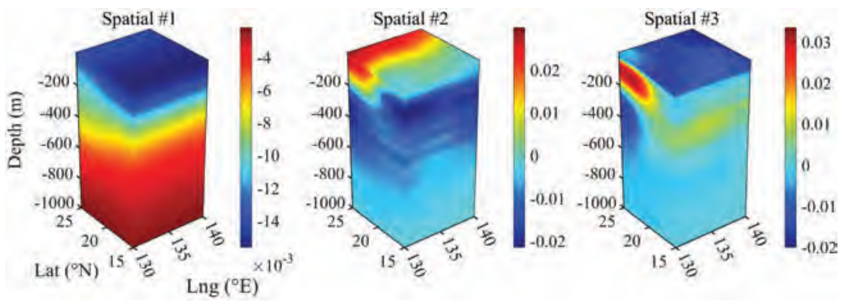
4. Experiment

4.1. *Dataset description*

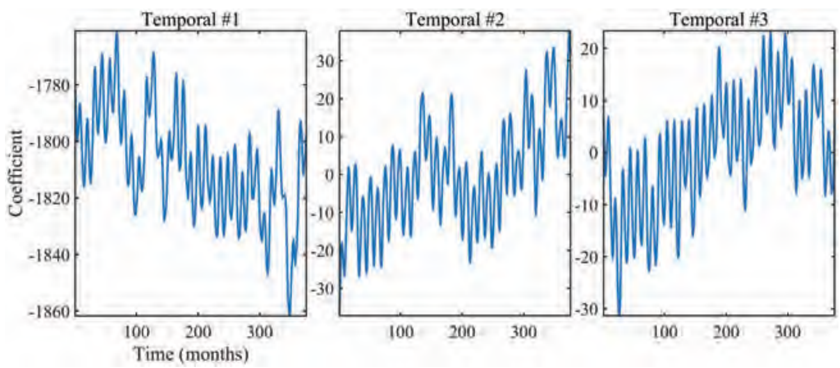
The Pacific temperature dataset [32] in Fig. 2 was employed to evaluate the proposed framework. From the global archive, we extract a subregion covering 15°N–25°N and 130°E–140°E, extending from the surface to 1000 m depth. The spatial domain is discretized into 21×21 horizontal grid points with 41 vertical levels, resulting in $41 \times 21 \times 21$ grid points per snapshot. The temporal dimension consists of $N_t = 731$ monthly snapshots. This region lies within the western Pacific subtropical gyre, characterized by pronounced thermocline variability, mesoscale eddies, and intermittent upwelling events. Such dynamics make it a challenging benchmark for sparse sensing and spatiotemporal field reconstruction. Each entry $x_{i,j,k,t}$ of the resulting ST tensor represents temperature at longitude i, latitude j, depth k, and time t, yielding a data tensor of size $N_x = 21$, $N_y = 21$, $N_z = 41$, $N_t = 731$. The dataset provides a realistic test case for in-situ mobile platform observations where high-resolution structures coexist with strong temporal variability, requiring both informative sensor placement and robust reconstruction for effective ocean field perception.

4.2. *Parameter settings*

For fair evaluation, several parameters are fixed across all experiments. The spatiotemporal decomposition employs singular value decomposition, and the number of retained modes R is determined by an energy threshold of $\eta = 99.99\%$, ensuring compactness while preserving dominant



(a)



(b)

Fig. 3. The first three spatial and temporal features, (a) spatial modes and (b) temporal coefficients.

dynamics. In sparse location optimization, the sensor budget is varied from 0.5% to 20% of the total spatial grid points (M/N_s). During reconstruction, temporal coefficients are estimated via regularized least squares. This configuration provides a consistent basis for comparing different sampling strategies (uniform, random, and proposed) under various sparsity ratios and validates the framework's robustness to practical constraints of mobile platform observations. For underwater robot simulation, a glider-like sawtooth trajectory with adaptive sensing is adopted, where 5% of the grid points along the path are selected according to the proposed optimization scheme.

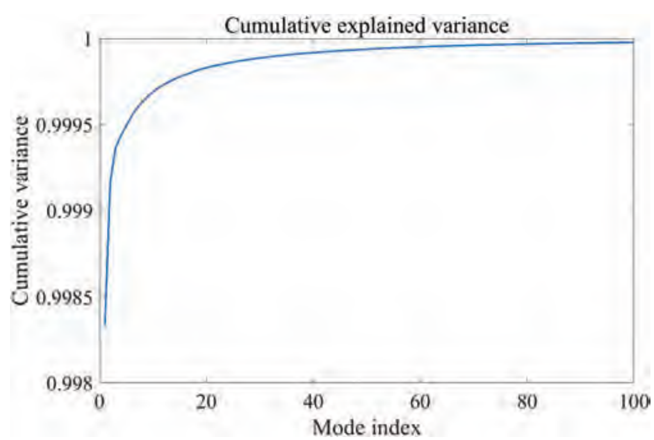


Fig. 4. Cumulative explained variance of spatiotemporal features.

4.3. *Performance metrics*

To assess the effectiveness of the proposed framework, we employ several quantitative metrics that evaluate reconstruction accuracy, statistical consistency, and physical reliability of the recovered spatiotemporal fields.

Root Mean Square Error (RMSE) measures the average deviation between reconstructed and true fields:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_s N_t}\sum_{t=1}^{N_t}\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2^2}, \tag{26}$$

where $\widehat{\boldsymbol{x}}_t$ and $\boldsymbol{x}_t$ are the reconstructed and true snapshots. Lower RMSE indicates higher fidelity.

Mean Absolute Error (MAE) measures the average magnitude of reconstruction errors, independent of direction:

$$\mathrm{MAE} = \frac{1}{N_s N_t}\sum_{t=1}^{N_t}\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_1. \tag{27}$$

Coefficient of determination (R^2) evaluates the proportion of variance in the reference field that is explained by the

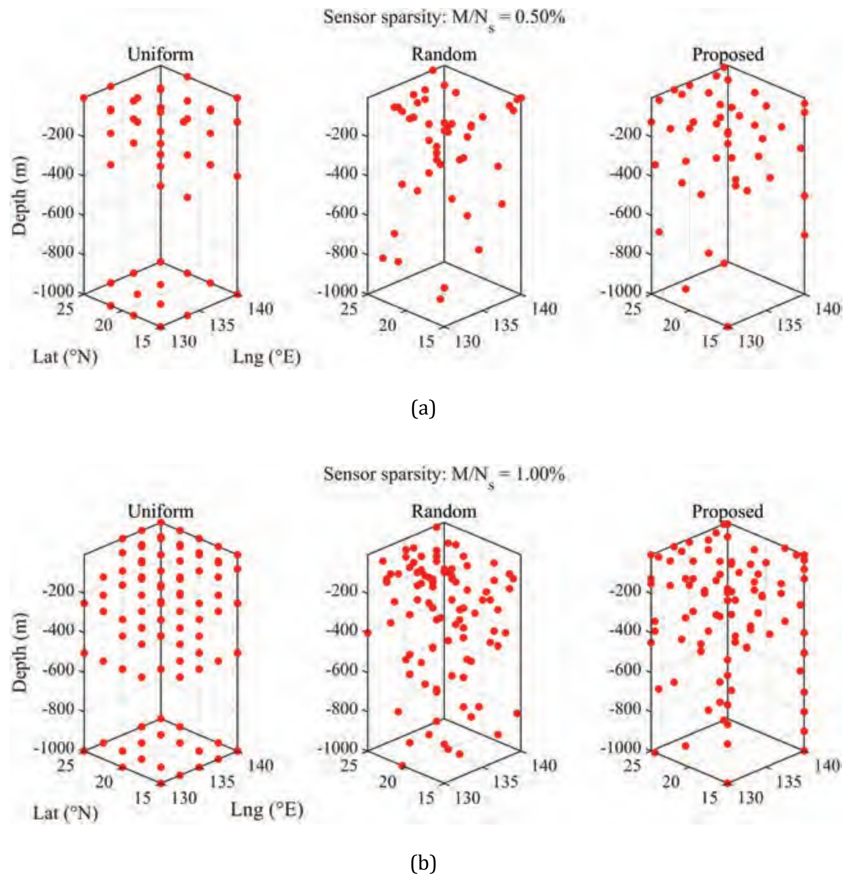


Fig. 5. Sensor distributions under different sparsity ratios, (a) $M/N_s = 0.5\%$ and (b) $M/N_s = 1.0\%$.

reconstruction:

$$R^2 = 1 - \frac{\sum_{t=1}^{N_t} \|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2^2}{\sum_{t=1}^{N_t} \|\boldsymbol{x}_t - \bar{\boldsymbol{x}}\|_2^2}, \tag{28}$$

where $\bar{\boldsymbol{x}}$ denotes the temporal mean of the ground truth field. Higher R^2 indicates better consistency with reference variability.

Relative Error (RE) normalizes the reconstruction error with respect to the magnitude of the true field:

$$\text{RE} = \frac{\sum_{t=1}^{N_t} \|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2}{\sum_{t=1}^{N_t} \|\boldsymbol{x}_t\|_2}. \tag{29}$$

4.4. *Analysis of spatiotemporal decoupling*

To examine the representational efficiency of the proposed decomposition, we extract the dominant spatiotemporal features on the Pacific temperature dataset. Figure 3 illustrates the first three spatial modes and their corresponding temporal coefficients. The spatial modes (Fig. 3(a)) capture physically meaningful structures of the ocean field. The first mode reflects the large-scale background stratification, dominated by vertical gradients between the surface mixed layer and deeper waters. The second and third modes highlight regional variations such as thermocline fluctuations and mesoscale eddy-like anomalies, with clear localized signatures in both horizontal and vertical directions. These modes indicate that the decomposition can efficiently capture both large-scale background states and higher-order perturbations with only a few basis vectors. The temporal coefficients (Fig. 3(b)) describe the evolution of each mode over time. The first coefficient follows long-term variability with seasonal and interannual oscillations, consistent with known Pacific climate fluctuations. The second and third coefficients exhibit more pronounced variability, reflecting the time-varying strength of thermocline undulations and eddy activities. Importantly, these temporal series are smooth and interpretable, confirming that the modal decomposition disentangles spatial structures from temporal dynamics in a compact and meaningful way.

To further quantify the compactness of the decomposition, Fig. 4 reports the cumulative explained variance of the spatiotemporal modes. The curve shows a rapid increase in the first few modes, indicating that most of the variance

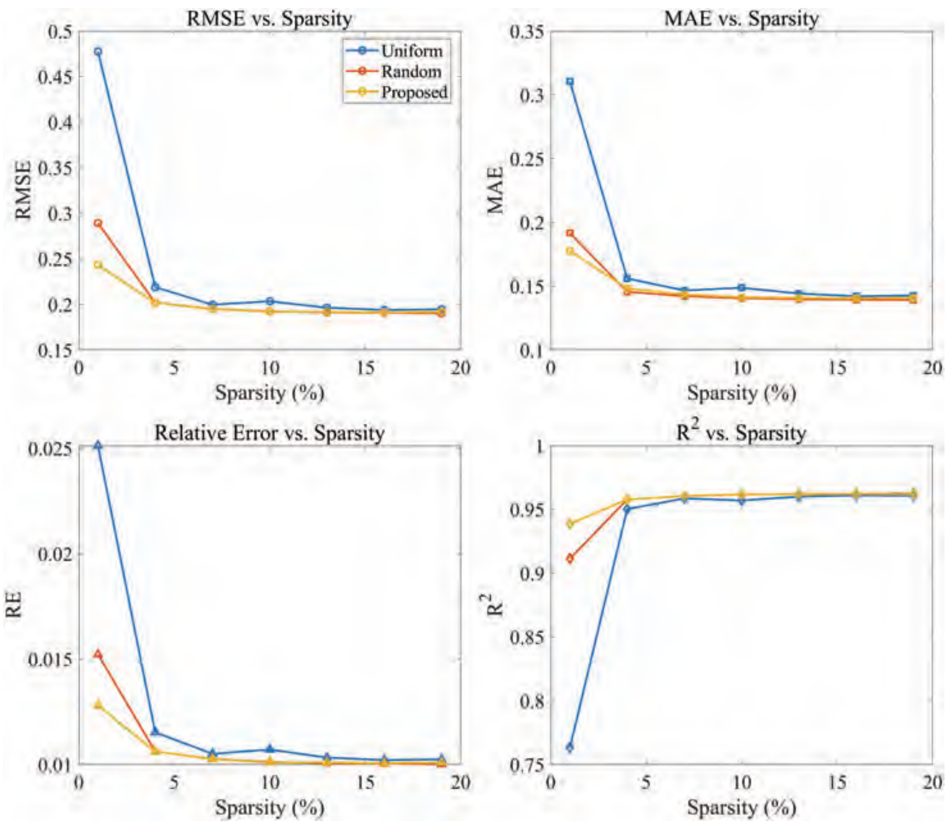


Fig. 6. Reconstruction performance for different sensor sparsity levels, comparing uniform, random, and proposed strategy.

is concentrated in a low-dimensional subspace. Specifically, fewer than 10 modes capture more than 99% of the total variance, and around 25 modes explain over 99.9%. Beyond this point, the marginal contribution of additional modes becomes negligible. This observation demonstrates that the Pacific temperature field exhibits a strong low-rank structure, where a small set of dominant modes is sufficient to describe the major dynamics. Such efficiency is particularly advantageous for sparse sensing, since accurate field reconstruction can be achieved with limited sensors if the selected measurements are informative for these leading modes. Consequently, the modal decomposition not only provides interpretability but also offers a compact and computationally tractable representation, forming a solid foundation for the subsequent sparse location optimization and spatiotemporal fusion.

4.5. *Analysis of sparse measurements*

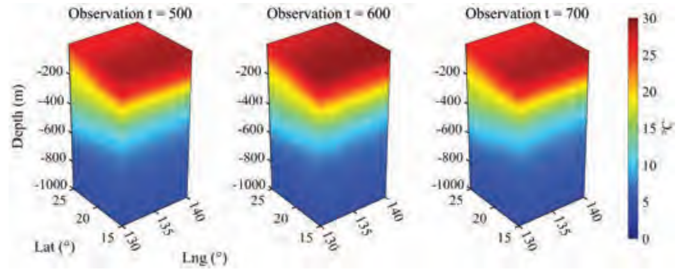
Figure 5 compares sensor layouts generated by three strategies: uniform, random, and the proposed optimization, under sparsity ratios of 0.5–1.0%. Uniform sampling enforces a regular grid, which guarantees coverage but often overlooks dynamically important regions such as the thermocline. Random sampling provides variability but leads to clustering and leaves large unsensed gaps, particularly in depth. In contrast, the proposed method distributes sensors more adaptively: surface and thermocline regions are more densely covered while deeper layers are anchored by sentinel sensors. This balanced design captures both energetic upper-layer dynamics and deep constraints, providing a more informative sampling pattern. Furthermore, the adaptive layout reflects a physically meaningful strategy: energy-rich regions where mesoscale eddies and thermocline shifts dominate receive higher sensor density, while relatively stable deep-water masses are sparsely but strategically monitored. This aligns with operational priorities in real ocean missions, where underwater robots must maximize scientific return under strict energy and time budgets. The proposed optimization also reduces redundancy compared with random placement and avoids oversampling in uniform grids, leading to better use of sensor resources.

The quantitative performance is shown in Fig. 6. Across all sparsity levels, the proposed method consistently outperforms both uniform and random baselines in terms of RMSE, MAE, and RE, while achieving higher R^2. At extremely low sparsity (0.5–1.0%), the advantage is most pronounced: reconstruction errors of the proposed method are significantly lower, and R^2 exceeds 0.9, whereas uniform sampling still suffers from poor fidelity. As sparsity increases beyond 5%, all methods gradually converge toward similar accuracy, but the proposed approach remains slightly superior. This trend demonstrates that intelligent sensor placement is especially critical in the sparse regime, while at higher sampling densities even naive strategies benefit from redundancy. Another important observation is the stability of the proposed method: its performance curves are smooth and monotonic, while random placement shows fluctuations and larger variance, reflecting its sensitivity to chance clustering. Uniform sampling, though stable, fails to exploit flow-dependent variability and therefore underperforms across all tested sparsity levels. Taken together, these results demonstrate that the proposed optimization not only improves mean accuracy but also enhances robustness and reliability, which are essential for deployment in unpredictable ocean environments.

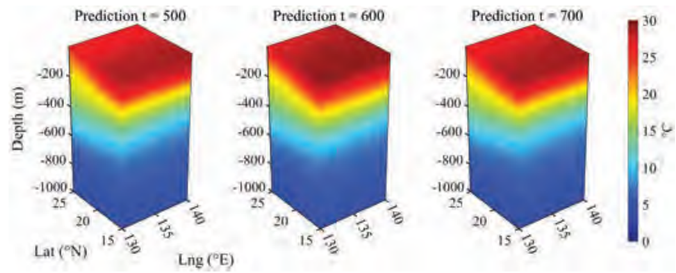
4.6. *Analysis of environment perception*

Figure 7 illustrates the reconstruction performance of the proposed spatiotemporal fusion framework at three representative time steps ($t = 500$, $t = 600$, and $t = 700$). Panel (a) shows the ground-truth observations, panel (b) presents the corresponding reconstructed fields, and panel (c) gives the point-wise reconstruction error. First, the reconstructed temperature fields visually match the ground truth across different depths and time steps. The large-scale stratification, characterized by strong vertical gradients between the warm surface layer and cold deeper waters, is well captured. Regional features such as the thermocline position and mesoscale fluctuations are also preserved, confirming that the spatiotemporal fusion effectively combines spatial structures with temporal dynamics to recover realistic three-dimensional patterns. Second, the point-wise error maps indicate that reconstruction errors are generally small, with magnitudes below 0.5°C across the domain. Most discrepancies appear in dynamically active regions, particularly near the surface and thermocline, where temperature gradients are sharp and mesoscale processes induce localized variability. In contrast, deeper layers exhibit negligible error, reflecting the relative stability of deep-water structures and the robustness of the sentinel-based anchoring in sparse sensor placement. Third, the temporal consistency of results across $t = 500$, 600, and 700 demonstrates the generalization capability of the framework. Despite being evaluated at different phases of seasonal and interannual variability, the reconstructed fields remain accurate, and error patterns are spatially coherent rather than random noise. This highlights the ability of the fusion scheme to disentangle mode dynamics and adapt them to sparse observations.

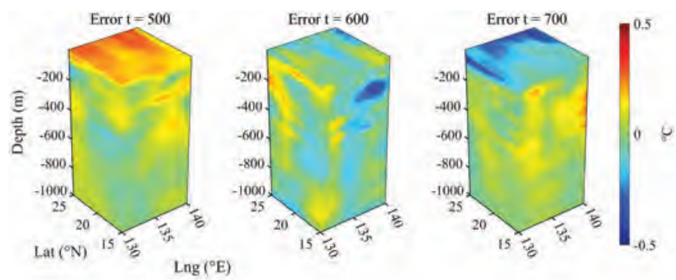
To further demonstrate the practical relevance of the proposed framework, we conduct a simulation mimicking the operation of an underwater robot performing sparse adaptive sensing. As shown in Fig. 8, the vehicle collects in-situ observations, with only 5% of the full spatial measurements sampled, with red markers indicating the selected locations optimized by the proposed sparse sensing strategy.



(a)



(b)



(c)

Fig. 7. The temperature field reconstruction, (a) ground truth, (b) prediction results, and (c) point-wise error.

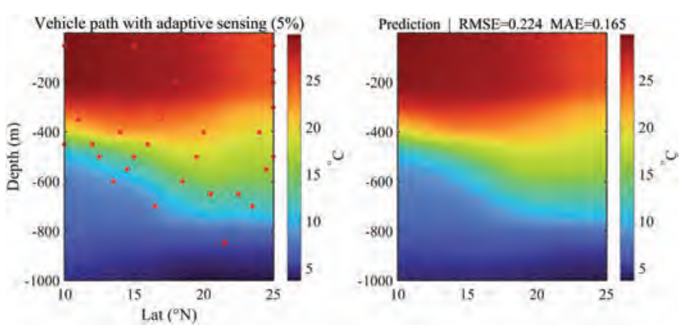


Fig. 8. Underwater robot sparse sensing at 5% sparsity.

The reconstructed temperature field is presented in the right panel. Despite the strong sparsity, the reconstruction achieves high fidelity with an RMSE of 0.224 and MAE of 0.165, accurately capturing both large-scale stratification and depth-dependent gradients. These results highlight that the proposed approach can robustly recover environmental fields even when constrained by realistic sampling patterns of mobile platforms. Importantly, the integration of sparse sensing ensures that dynamically important layers, such as the thermocline, are well represented, while deep waters are anchored with sentinel observations.

5. Conclusion

This paper presented a large-scale ocean environment perception framework that reconstructs spatiotemporal fields from sparse and heterogeneous measurements. The proposed approach integrates three key components: (i) a spatiotemporal modal decomposition that provides a compact and interpretable representation of dominant ocean structures; (ii) a multi-objective sparse location optimization strategy that balances information gain, spatial coverage, and anchor constraints; and (iii) a spatiotemporal fusion scheme that reconstructs the full environmental field by coupling spatial modes with temporal dynamics. Experimental validation on the Pacific temperature demonstrates that the framework can achieve high reconstruction accuracy under extreme sparsity, with average errors below 0.5°C at sampling ratios as low as 1%. The analysis confirms that the method effectively captures large-scale stratification, thermocline variability, and mesoscale anomalies, while maintaining robustness across different temporal phases. In summary, the proposed framework provides a principled and practical solution to sparse ocean environment perception. It offers a pathway toward more intelligent and resource-efficient autonomous observation and paves the way for future integration with adaptive mission planning, multi-platform cooperation, and real-time decision-making in complex marine environments.

Future work will extend the proposed framework to jointly optimize both the number and placement of sensors to balance problem scale and information loss, and to generalize the spatiotemporal modeling and fusion scheme beyond temperature to multi-variable environments such as salinity and oxygen, enabling comprehensive, efficient, and interpretable perception of complex ocean processes.

Acknowledgments

This study was supported by the Research Grants Council of Hong Kong SAR (Grants: 14217922, 14209623, 14209424, and 14200524), the Environment and Conservation Fund of Hong Kong SAR (Project No. 142/2023), and the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government via the Hong Kong Centre for Logistics Robotics.

ORCID

Lei Lei
https://orcid.org/0000-0002-3630-3359
Xi Chen
https://orcid.org/0000-0003-2168-9057
Ben M. Chen
https://orcid.org/0000-0002-3839-5787

References

[1] P. Choudhary *et al.*, Empowering blue economy: From underrated ecosystem to sustainable industry, *J. Environ. Manag.* **291** (2021) 112697.
[2] A. M. Eger *et al.*, The value of ecosystem services in global marine kelp forests, *Nat. Commun.* **14**(1) (2023) 1894.
[3] I. Olthof and T. Rainville, Dynamic surface water maps of canada from 1984 to 2019 landsat satellite imagery, *Remote Sens. Environ.* **279** (2022) 113121.
[4] D. Roemmich *et al.*, On the future of argo: A global, full-depth, multi-disciplinary array, *Front. Mar. Sci.* **6** (2019) 439.
[5] M. Lu, S. Liu, Y. Luo, Q. Qi, Y. Sun and J. Li, Modeling and output feedback tracking control of self-developed autonomous underwater vehicle, *Unmanned Syst.* **13**(6) (2025) 1639–1660.
[6] Z. Yan and J. Yan, Self-triggered formation control of an asynchronous multiple underwater unmanned vehicle under communication delay, *Unmanned Syst.* (2025) 1–19.
[7] L. Lei, Y. Zhou and J. Zhang, From extended environment perception towards real-time dynamic modeling for long-range underwater robot, *IEEE Trans. Robot.* **41** (2025) 3423–3441.
[8] Y. H. Tan and B. M. Chen, Survey on the development of aerial–aquatic hybrid vehicles, *Unmanned Syst.* **9**(3) (2021) 263–282.
[9] D. R. Yoerger *et al.*, A hybrid underwater robot for multidisciplinary investigation of the ocean twilight zone, *Sci. Robot.* **6**(55) (2021) eabe1901.
[10] R. M. Holmes, J. D. Zika, S. M. Griffies, A. M. Hogg, A. E. Kiss and M. H. England, The geography of numerical mixing in a suite of global ocean models, *J. Adv. Model. Earth Syst.* **13**(7) (2021) e2020MS002333.
[11] J. Jin, M. Espino, D. Fernàndez-Garcia and A. Folch, Coupling of numerical groundwater–ocean models to improve understanding of the coastal zone, *Ocean Sci.* **21**(4) (2025) 1407–1424.
[12] J.-M. Beckers and M. Rixen, Eof calculations and data filling from incomplete oceanographic datasets, *J. Atmos. Ocean. Technol.* **20**(12) (2003) 1839–1856.
[13] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* **656** (2010) 5–28.
[14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2006).
[15] X. Luo, W. Xu, Y. Ren, S. Yoo and B. Nadiga, Continuous field reconstruction from sparse observations with implicit neural representations, preprint (2024), arXiv:2401.11611.
[16] J. E. Santos *et al.*, Senseiver: Efficient reconstruction of spatial fields from sparse measurements with attention-based neural networks, *Nat. Mach. Intell.* **5**(11) (2023) 1317–1325.

[17] N. Wang, D. Xu, H. Zhang and X. Kong, Information theory of compressive sensing in sensor array, *IEEE Trans. Aerosp. Electron. Syst.* **61**(2) (2024) 4480–4491.
[18] C. Robbiano, M. R. Azimi-Sadjadi and E. K. Chong, Information-theoretic interactive sensing and inference for autonomous systems, *IEEE Trans. Signal Process.* **69** (2021) 5627–5637.
[19] G. Zhang, P. Wang, W. Wang, Y. Mu, Y. Li, J. Tang, H. Song, H. Wen and S. Mumtaz, An information theoretic approach to distributed detection for mobile wireless sensor networks under byzantine attack in entirely unknown or complicated environment: Design, analysis, and evaluation of the attack strategy, *IEEE Internet Things J.* **12**(4) (2024) 3689–3706.
[20] X. Fang, J. Li, S. Zhang, W. Chen and Z. He, Optimal aoa sensor-source geometry with deployment region constraints, *IEEE Commun. Lett.* **26**(4) (2022) 793–797.
[21] K. Nakai, Y. Sasaki, T. Nagata, K. Yamada, Y. Saito and T. Nonomura, Nondominated-solution-based multi-objective greedy sensor selection for optimal design of experiments, *IEEE Trans. Signal Process.* **70** (2022) 5694–5707.
[22] L. Lei, X. Chen and B. M. Chen, Environment uncertainty-aware dynamic modeling for autonomous deep-sea robots, *Unmanned Syst.* **13**(5) (2025) 1295–1306.
[23] J. Li, W. Zhang, B. Guo, F. Yao, M. Zhang and X. Shao, Dynamic positioning for autonomous underwater vehicles: A tube model predictive control approach, *Unmanned Syst.* **13**(6) (2024) 1573–1583.
[24] D.-Z. Dong, Y. Xu, Z.-Y. Guo, K.-H. Miao, Z.-H. Gong and D.-L. Luo, Adaptive cooperative formation control of autonomous underwater vehicles based on event-triggered mechanism, *Unmanned Syst.* **13**(1) (2025) 91–104.
[25] J. Ma, X. Li, J. Song, L. Wen, Q. Wang, K. Xu, J. Dai and G. Zhong, The effects of seawater thermodynamic parameters on the oxygen minimum zone (omz) in the tropical western pacific ocean, *Mar. Pollut. Bull.* **187** (2023) 114579.
[26] E. Romero, L. Tenorio-Fernandez, E. Portela, J. Montes-Aréchiga and L. Sánchez-Velasco, Improving the thermocline calculation over the global ocean, *Ocean Sci.* **19**(3) (2023) 887–901.
[27] M. O. Berild, Y. Ge, J. Eidsvik, G.-A. Fuglstad and I. Ellingsen, Efficient 3d real-time adaptive auv sampling of a river plume front, *Front. Mar. Sci.* **10** (2024) 1319719.
[28] S. Bai, T. Shan, F. Chen, L. Liu and B. Englot, Information-driven path planning, *Curr. Robot. Rep.* **2**(2) (2021) 177–188.
[29] E. Galceran and M. Carreras, A survey on coverage path planning for robotics, *Robot. Auton. Syst.* **61**(12) (2013) 1258–1276.
[30] L. Nam, L. Huang, X. J. Li and J. Xu, An approach for coverage path planning for uavs, in *2016 IEEE 14th Int. Workshop Advanced Motion Control (AMC)* (IEEE, 2016), pp. 411–416.
[31] T. B. von See, J. Greinert and T. Meurer, Detection and tracking of ocean layers using an auv with ukf based extremum seeking control in the baltic sea, *Sci. Rep.* **14**(1) (2024) 20629.
[32] South China Sea and Adjacent Seas Data Center, National Earth System Science Data Center, National Science and Technology Infrastructure of China, http://ocean.geodata.cn.

Lei Lei received his B.E. degree in Naval Architecture and Ocean Engineering from the Department of Ship Engineering, Harbin Engineering University, Harbin, P. R. China in 2016, M.E. degree in Mechanical Engineering from the Department of Mechanical Science and Engineering, Huazhong University of Science and Engineering, Wuhan, P. R. China in 2019, and Ph.D. degree from the Department of Systems Engineering, City University of Hong Kong, Hong Kong in 2024.

Dr. Lei is currently a Postdoctoral Fellow at the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong. His research interests include underwater unmanned systems and ocean big data.

Xi Chen received his B.S. Degree in Energy and Environment Systems Engineering from Zhejiang University, in 2009, MPhil and Ph.D. degrees in Building Environment and Energy Engineering from the Hong Kong Polytechnic University, in 2011 and 2017, respectively.

Dr. Chen is now Research Assistant Professor in the Chinese University of Hong Kong, and a Chartered Energy Engineer in Engineering Council (UK). He has over 10 years of experience in sustainable building technologies related to green building, built environment labeling and smart inspection. He has led or managed multiple research projects including Australian Research Council (ARC), Environment, and Conservation Fund (ECF), Innovation and Technology Fund (ITF), Research Grants Council (RGC) Hong Kong and industrial consultancy projects with an awarded sum over 10M HKD. He published over 70 papers in peer-reviewed international journals/conferences and coauthored two books in green building, energy efficiency and renewable application areas with an h-index of 32.

Dr. Chen has been awarded the DECRA (Discovery Early Career Researcher Award) Fellow in the Australian Research Council and Fulbright Scholar in the Lawrence Berkeley National Laboratory. He also serves as the editorial board member of Buildings, Energies, Unmanned Systems and Advances in Applied Energy.

Ben M. Chen is Professor and Chairman of the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong (CUHK). Prior to joining CUHK in 2018, he was a Provost's Chair Professor in the Department of Electrical and Computer Engineering at the National University of Singapore, where he worked for 28 years. He also served as Assistant Professor at the State University of New York at Stony Brook from 1992 to 1993 and worked as a Software Engineer at South China Computer Corporation from 1983 to 1986. His current research focuses on unmanned systems and their applications.

Dr. Chen is an IEEE Fellow and a Fellow of the Academy of Engineering, Singapore. He has authored or co-authored hundreds of research articles and ten research monographs in control theory, unmanned systems, and financial market modeling. He has served on the editorial boards of several international journals, including Automatica and IEEE Transactions on Automatic Control. Currently, he is the Editor-in-Chief of Unmanned Systems and an Editor of the International Journal of Robust and Nonlinear Control.