Informative path planning in complex marine environments: Gap imputation and variance minimization in satellite data

Diajeng Wulandari Atmojo[1,2,3,*], Christoph Manss[2], Janina Schneider[1,2], and Oliver Zielinski[2,4]

[1]Carl von Ossietzky University of Oldenburg, Institute for Chemistry and Biology of the Marine Environment, Wilhelmshaven, Germany
[2]German Research Center for Artificial Intelligence, Research Department Marine Perception, Oldenburg, Germany
[3]Now at University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany and
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[4]Now at Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Warnemünde, Germany
[*]Correspondence: atmojo@uni-bremen.de

OCEANS 2024 - SINGAPORE | 979-8-3503-6207-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/OCEANS51537.2024.10752615

***Abstract*—Satellite remote sensing plays a critical role in operational oceanography by providing Near Real Time (NRT) measurements of essential climate variables (ECVs), such as Sea Surface Temperature (SST) and Salinity with extensive spatial coverage compared with in-situ measurements. However, these parameters are obtained through passive measurements with satellite data that often suffer from attenuation due to atmospheric effects, resulting in missing data in satellite recordings. To overcome this issue, we propose a novel approach towards intelligent autonomous measurement acquisition that outlines the potential use of Unmanned Surface Vehicles (USVs). The aim is to measure at locations with missing data by combining information-driven exploration and energy-efficient path planning. Information-driven exploration refers to identifying potential measurement points with high informativeness. We use differential entropy as information metric based on a Gaussian Process Regression (GPR) model that uncovers the variances in the satellite data. Subsequently, we plan a path of an USV by extending the Rapidly-exploring Random Tree Star (RRT*) algorithm by introducing an energy-dependent cost function considering Sea Surface Currents (SSC) to accurately capture the complex marine environment. By this, we introduce a modified information metric trading out informativeness and energy consumption in the path planning. The results demonstrate that informative and energy-efficient path planning can significantly reduce the USV's energy consumption, with potential savings of up to 88%.**

***Index Terms*—gap imputation, remote sensing, unmanned surface vehicle, autonomous navigation, path planning, RRT*, operational oceanography, North Sea**

I. Introduction

Operational oceanography encompasses long-term and systematic routine monitoring to achieve accurate ocean state forecasts, providing support for climate research, ecological studies, maritime stakeholders, and political decision-makers. Satellite remote sensing builds up one of the crucial components in the observation network of operational oceanography as it performs Near Real Time (NRT) measurements with extensive spatial coverage, surpassing the limitations of in-situ measurements [1], [2].

However, satellite passive observations are retrieved by measuring the naturally reflected sunlight from the earth's surface, which is challenging due to its attenuation in the atmosphere caused by cloud covers, unfavourable sun glints, rain, and high aerosol concentration [3], [4]. These atmospheric effects lead to missing values in the Level 2 (L2) satellite recordings. More than two-thirds of the ocean's surface is covered by clouds on average [5], highlighting the need for reliable and efficient gap imputation methods.

To address this issue, considerable efforts have been dedicated to developing post-hoc data-driven reconstruction methods that infer missing data values from available data to generate higher-level satellite products (Level 3 and 4) [3], [4]. Machine learning approaches such as support vector regression, random forest [6], and neural networks [7] have been applied for Sea Surface Temperature (SST) interpolation. Additionally, data fusion techniques have been proposed to fill the data gaps. These techniques involve merging satellite data with outputs from an ecohydrodynamic model [8] or combining data from multiple satellites and in-situ measurements [9].

In this paper, we introduce a novel approach that outlines the potential use of Unmanned Surface Vehicles (USVs) by integrating advanced autonomous driving techniques into the L2 satellite data reconstruction process to enhance the data quality of higher-level products. To the best of our knowledge, USVs have primarily been deployed for validation purposes of Level 3 or Level 4 satellite data [10], [11], so-called matchups, relying on expert knowledge to strategically select measurement points. Contrarily, we aim to develop a data-driven framework that lets the USV select measurement points autonomously based on the most recent data gaps in the satellite recordings in an intelligent and energy efficient manner. We achieve this through the following steps:

1) *Entropy-driven exploration*: Along with the work of [12], we identify potential measurement points with high informativeness based on a Gaussian Process Regression (GPR) model. As information metric, the normalized differential entropy [13] is derived from the variance estimation of the GPR model.
2) *Energy-efficient path planning*: We perform sampling-based path planning with a modified Rapidly-exploring

Random Tree Star (RRT*) algorithm towards potential measurement points with high informativeness. As cost function, we adopt the energy consumption model by [14] considering the Sea Surface Currents (SSC).
3) *Measurement point selection*: To incorporate the costs associated with the USV into the information metric, we select the next measurement point trading of informativeness and energy consumption.

This paper is structured as follows: Sec. II presents the datasets used in this work, Sec. III explains the methodology comprising methods for entropy-driven exploration, energy-efficient path planning, and measurement point selection. In Sec. IV, we show the simulation experiments and evaluation of the results. We discuss the applied methods and the simulation results in Sec. V and suggest ideas for future works. Finally, Sec. VI concludes our proposed approach.

II. Datasets

The data sources used in our simulation setup can be divided into three different subsets: model area with the static obstacles constraining where the USV operates, satellite data showing the data gaps, and environmental data providing information about the marine environmental conditions.

A. Model area

Fig. 1 shows the model area, which is the North Sea extending from 6.5°E to 9.15°E longitude and from 53.38°N to 55°N latitude. We consider the following static obstacles: land cover, offshore wind farms, and in-situ measurement platforms such as tide gauges and offshore installations for oil and gas exploitation. We neglect dynamic obstacles like ships as dealing with dynamic collision avoidance is beyond the scope of this work.

We construct the land cover using the Water Body Dataset from the Shuttle Radar Topography mission [15] by masking the area outside the water body. Moreover, rivers and bays are also treated as land cover. The offshore installations are available in the European Marine Observation and Data Network (EMODnet, https://emodnet.ec.europa.eu/en/human-activities).

B. L2 SST of SENTINEL-3A

The L2 SST data of SENTINEL-3A [16] lies the basis of our work. The SST is passively measured by the Sea and Land Surface Temperature Radiometer (SLSTR) instrument, which provides a spatial resolution of 1 km and a swath width of 1420 km. The model area is sensed twice a day, with daytime measurements taken in the descending direction and nighttime measurements in the ascending direction. Due to the diurnal warming effect and the skin effect introducing biases in the daytime SST data, we only consider the nighttime SST data for analysis [17], [18].

Fig. 1 also depicts the SST data from the NRT product, which are available three hours after sensing. The presented measurements were taken on November 2, 2021, from 21:39:59.19 until 21:40:30.10 UTC. The spatial gaps observed

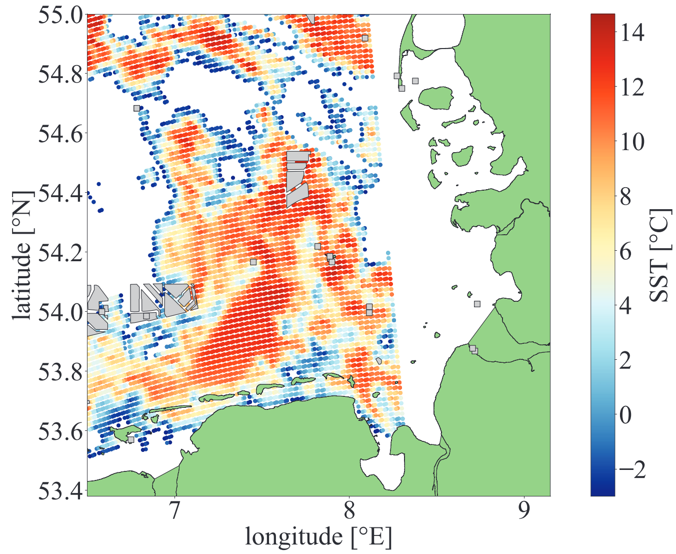

Fig. 1: Model area with land cover (green) and offshore installations (grey) as static obstacles and L2 SST data of SENTINEL-3A in ascending direction on November 2, 2021. The North Sea was sensed between 21:39:59.19 and 21:40:30.10 UTC.

in the measurements are attributed to atmospheric effects, whereas the sharp cutoff on the right border is due to the swath width of the SLSTR instrument.

C. Ocean model BSH-HBMnoku

The marine environment is characterised by various dynamic conditions such as ocean currents and winds impacting the energy demand of an USV [19]. For simplicity, we neglect the winds and focus on understanding the influence of the SSC on the energy consumption of the USV as the aim of our work is to give a general comprehensive framework on how to utilize USVs for gap imputation.

We use the SSC field of the operational numerical model system BSH-HBMnoku [20] of the Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie) which performs four times daily ocean forecasts on two different structured grid resolutions. For this work, we use the model output of the finer horizontal grid resolution of 0.9 km and a temporal resolution of 15 minutes.

III. Methodology

Fig. 2 outlines the workflow of this work. The methods are divided into three parts: entropy-driven exploration based on a GPR model, energy-efficient path planning in complex marine environments with the RRT* algorithm, and selection of the next measurement point trading out informativeness and energy consumption.

A. Entropy-based exploration based on a GPR model

The objective of exploration is to minimize uncertainty about unknown environments and maximize the acquisition of information. In our case, the locations of gaps in the L2 SST

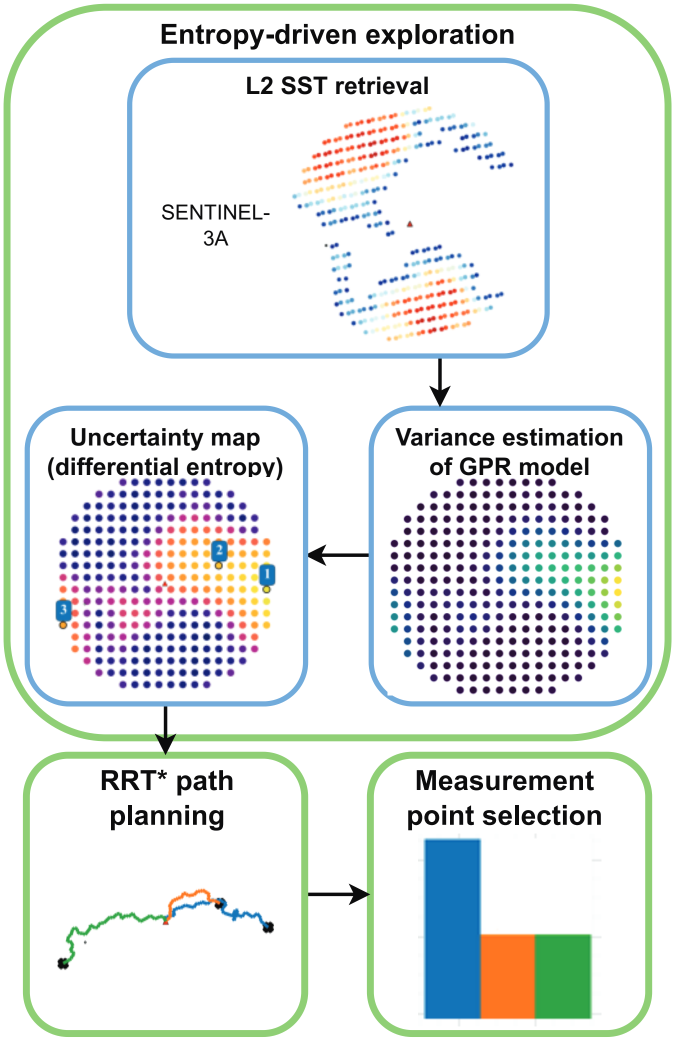

Fig. 2: Outline of the workflow, highlighting the contributory elements of this work. The green boxes showcase the essential components of our approach. The blue boxes represent a detailed schematic of the methods used for entropy-driven exploration.

data represent the unknown environment and the USV should be guided to locations where the SST is uncertain. We do that in three steps: First, to quantify the uncertainty, we estimate the SST field of a specific area with a GPR model by using L2 SST data of SENTINEL-3A as training data for the GPR model. Second, based on the variance estimates provided by the GPR model, we compute the differential entropy [13] of each potential measurement point. Third, we identify potential measurement points with high informativeness.

1) Fundamentals of Gaussian Processes (GPs): A GP is a stochastic process in which every finite collection of random variables has a multivariate Gaussian distribution [21]. It is a probability distribution over functions with a continuous domain, such that the distribution is the joint distribution of those random variables. A GP is specified by its mean function $m(\cdot)$ and covariance function $k_\theta(\cdot,\cdot)$ of a physical process $y(\mathbf{x}) \in \mathbb{R}$, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$ denoting the position vector in a two-dimensional space:

$$y(\mathbf{x}) \sim GP(m(\mathbf{x}), k_\theta(\mathbf{x}, \mathbf{x}')), \tag{1}$$
$$m(\mathbf{x}) = \mathbb{E}[y(\mathbf{x})], \tag{2}$$
$$k_\theta(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(y(\mathbf{x}) - m(\mathbf{x}))(y(\mathbf{x}') - m(\mathbf{x}'))], \tag{3}$$

where $m(\mathbf{x})$ is the mean function indicating the expected process value at input $\mathbf{x}$, i.e. the average of all functions in the distribution at point $\mathbf{x}$. The covariance function $k_\theta(\cdot,\cdot)$ defines the correlation between process values at distinct input points $\mathbf{x}$ and $\mathbf{x}'$ with the hyperparameters $\boldsymbol{\theta}$.

The properties of a GP are fully specified by the choice of the covariance function and its hyperparameters $\boldsymbol{\theta}$, where the latter are inferred from the training data and the former is a design parameter. The Log-Marginal-Likelihood is then defined as

$$\begin{aligned}\log\, p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}) = &-\frac{1}{2}\mathbf{z}^T(\mathbf{K} + \sigma_\epsilon^2\mathbf{I})^{-1}\mathbf{z}\\ &-\frac{1}{2}\log|(\mathbf{K} + \sigma_\epsilon^2\mathbf{I})|\\ &-\frac{N}{2}\log 2\pi.\end{aligned} \tag{4}$$

To find the optimal hyperparameters $\boldsymbol{\theta}_*$, (4) is maximized with respect to $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}_* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}(\log\, p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z})), \tag{5}$$

where $\mathbf{z} \in \mathbb{R}^N$ denotes the N noisy observations, $\sigma_\epsilon^2 \in \mathbb{R}$ the noise variance, and $\mathbf{K} \in \mathbb{R}^{N\times N}$ is the covariance matrix of N observations constructed by applying the covariance function. $\mathbf{I}$ is the identity matrix. For the maximization of the Log-Marginal-Likelihood with respect to $\boldsymbol{\theta}$, the first and second terms in 4 are only relevant, as the other term is constant for a given $\mathbf{z}$ and N. $\mathbf{X} = [\mathbf{x}^{[1]}\ \mathbf{x}^{[2]}\ \cdots\ \mathbf{x}^{[N]}]^T \in \mathbb{R}^{N\times 2}$ summarises all observation points N in a two-dimensional space as a matrix. See [21] for further details on the derivation of the Log-Marginal-Likelihood.

2) Generation of an uncertainty map: We use the module `GaussianProcessRegressor` of `scikit-learn` [22] to implement a GPR model. For the covariance function, we choose the Matérn kernel [21], which is stationary and anisotropic and has two control parameters: the characteristic length-scale $\mathbf{l} \in \mathbb{R}^D$, which defines the distance in which two samples $\mathbf{x}$ and $\mathbf{x}'$ correlate in D-dimensional space, and $\nu \in \mathbb{R}^+$, which controls the smoothness of the function.

The `GaussianProcessRegressor` only optimizes $\mathbf{l}$ and requires a predefined ν. Therefore, we perform a 10-fold cross-validation (CV) using SST data of the ocean model BSH-HBMnoku as training data since BSH-HBMnoku provides a spatially fully covered estimation of the SST. We set ν values in the range of 0.1 and 5 and select the ν value for which the CV results in the lowest root mean squared error (RMSE). The Matérn kernel exhibits the lowest μ_{RMSE} at $\nu = 0.2485$ such that we use this value in the next steps for optimizing the hyperparameters $\boldsymbol{\theta}$.

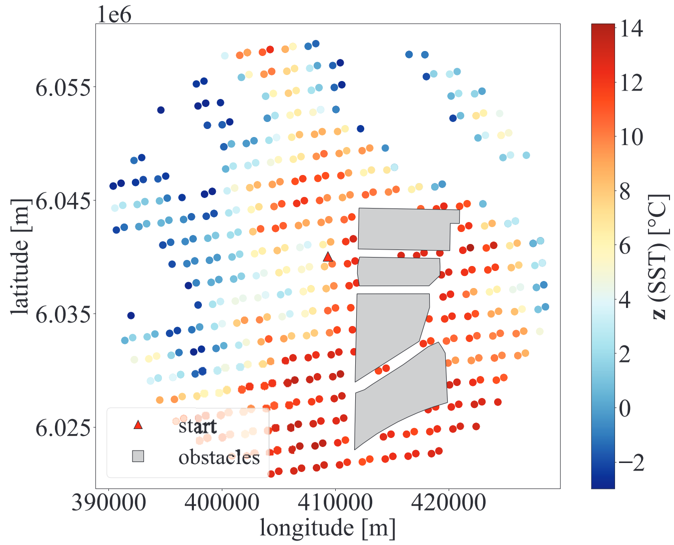

Fig. 3: Local L2 SST as training data in the GPR model within the search area of an USV located at coordinates 7.6 °E and 54.5 °N with a search radius of 20 km. Spatial gaps in the satellite recordings are attributed to atmospheric effects. Coordinate Rerefernce System: ETRS89 / UTM Zone 32.

Let an USV be located within the model area. We assume the USV to acquire the reduced NRT product of the L2 for its region of interest SST three hours after sensing, precisely on November 3, 2021, at around 00:40 UTC. Specifically, the USV retrieves the L2 SST data within a radius of 20 km from its current position, which we refer to the search area in the following. Fig. 3 provides an example of the L2 SST retrieval for an USV located in 7.6 °E, 54.5 °N. Again, gaps in the L2 SST data result from atmospheric effects attenuating the reflected sunlight from the earth's surface. Subsequently, we fit the GPR model on the locally retrieved L2 SST data with the most suited ν-value selected from the CV to infer the optimal hyperparameters $\boldsymbol{\theta}_*$.

Once the GPR model is trained, the next step is to identify P potential measurement points of high uncertainty. The aim is to generate an uncertainty map $\mathbf{h} \in \mathbb{R}^P$ which serves as a guide for the USV to navigate towards points with high informativeness. Alg. 1 outlines the workflow of the generation of an uncertainty map.

Algorithm 1 Uncertainty map

Input: trained GPR-MODEL, $\mathbf{X}_*$
Output: $\mathbf{h}_{norm}$

1: $P \leftarrow \texttt{len}(\mathbf{X}_*)$
2: **for** i = 1, ..., P **do**
3: $\quad \mu_*, \sigma_* \leftarrow \texttt{predict}(\text{GPR-MODEL}, \mathbf{X}_*^{[i]})$
4: $\quad \mathbf{h}^{[i]} \leftarrow \texttt{differential\_entropy}(\sigma_*)$
5: **end for**
6: $\mathbf{h}_{norm} \leftarrow \texttt{normalize}(\mathbf{h})$

Let $\mathbf{X}_* = [\mathbf{x}_*^{[1]} \; \mathbf{x}_*^{[2]} \; \cdots \; \mathbf{x}_*^{[P]}]^T \in \mathbb{R}^{P\times 2}$ be a two-dimensional grid of P potential measurement points with a spatial resolution of 1.2 km over the search area. With the trained GPR model, we predict the process value $y_*^{[i]}$ with $i = 1, 2, ..., P$ at each potential measurement point $\mathbf{x}_*^{[i]}$ with the respective model variance σ_* (Alg. 1, line 3). Subsequently, we compute the differential entropy H [13] for each potential measurement point (Alg. 1, line 4):

$$\mathbf{h}^{[i]} = H(y_*^{[i]}|\mathbf{X}) = \frac{1}{2}\log(2\pi e \sigma_*^2). \tag{6}$$

normalizing the uncertainty map facilitates the comparison of differential entropy values with other search areas (Alg. 1, line 6):

$$\mathbf{h}_{norm} = \frac{\mathbf{h} - \texttt{min}(\mathbf{h})}{\texttt{max}(\mathbf{h}) - \texttt{min}(\mathbf{h})}, \tag{7}$$

where $\mathbf{h}_{norm}$ denotes the normalized uncertainty map, $\texttt{max}(\mathbf{h})$ the maximum and $\texttt{min}(\mathbf{h})$ the minimum differential entropy within the search area, respectively. For better readability, we will refer to the normalized differential entropy as entropy in the following.

3) Exploration of most informative points: In the final step, we identify local maxima based on clustering the uncertainty map. Clusters represent areas where uncertainty values are grouped together based solely on their spatial locations. Filtering local maxima is summarised in Alg. 2.

Algorithm 2 Filter local maxima

Input: $\mathbf{h}_{norm}, k$
Output: $\mathbf{h}_{max}$

1: $\mathbf{h}_{0.75} \leftarrow \texttt{filter\_quantil}(\mathbf{h}_{norm}, 0.75)$
2: $[\mathbf{h}_{cluster}^{[1]}, ..., \mathbf{h}_{cluster}^{[k]}] \leftarrow \texttt{cluster}(\mathbf{h}_{0.75}^{[k]})$
3: **for** i = 1,..,k **do**
4: $\quad \mathbf{h}_{max}^{[i]} \leftarrow \texttt{max}(\mathbf{h}_{cluster}^{[i]})$
5: **end for**
6: $\mathbf{h}_{max} \leftarrow \texttt{sort}(\mathbf{h}_{max})$

To narrow down the focus on areas with relatively higher uncertainty, we filter potential measurement points falling within the upper 75% percentile, denoted as $\mathbf{h}_{0.75}$ (Alg. 2 line 1). Among $\mathbf{h}_{0.75}$, we apply k-means clustering [23] with $k = 3$ (Alg. 2 line 2) to identify the potential measurement points with the highest uncertainty of each cluster $\mathbf{h}_{max}^{[i]}$ for $i = 1, ..., k$ (Alg. 2 line 4). Eventually, we rank the local maxima according to their entropy value in a descending order (Alg. 2 line 5).

In the subsequent analysis, we will incorporate the costs associated with the USV into the information metric. This comprehensive consideration of costs and information will facilitate the selection of the most efficient and informative measurement point for the USV's exploration strategy.

B. Energy-efficient path planning in complex marine environments

To plan the path of an USV in an energy-efficient manner, the complex environmental conditions need to be considered, encompassing the ocean currents and the winds. In our approach, we reduce the complexity of the environment by introducing an energy consumption model proposed in [14] that only takes into account the ocean currents. As presented in the following, we use this energy consumption model as cost function for the RRT* algorithm [24] to find feasible and optimal paths from the start to the goal position.

1) Energy consumption model: We define the cost as the energy consumption E of the USV depending on the prevailing SSC. Let $\mathbf{v} \in \mathbb{R}^2$ be a velocity vector in two-dimensional space and $|\mathbf{v}| \in \mathbb{R}_0^+[\frac{\mathrm{m}}{\mathrm{s}}]$ the respective speed. Assume an USV navigating from point $\mathbf{x}_i$ to point $\mathbf{x}_{i+1}$ with a total velocity $\mathbf{v}_g$ that is the resultant vector of the relative velocity $\mathbf{v}_u$, contributed by the engine of the USV, and the velocity of the SSC $\mathbf{v}_c$:

$$\mathbf{v}_g = \mathbf{v}_u + \mathbf{v}_c. \tag{8}$$

Then, $\mathbf{v}_g$ can be inserted into the energy consumption model E such that

$$E = \alpha |\mathbf{v}_u|^3 \frac{\|\mathbf{x}_i - \mathbf{x}_{i+1}\|_2}{|\mathbf{v}_g|}, \tag{9}$$

where $\alpha \in \mathbb{R}^+[\frac{\mathrm{kg}}{\mathrm{m}}]$ represents the water density, drag coefficient, and the reference area of the USV [14]. For our analysis, it is sufficient to set $\alpha = 1$. Furthermore, we assume the USV to drive constantly with $|\mathbf{v}_g| = 10\frac{\mathrm{km}}{\mathrm{h}}$. Consequently, only the relative speed $|\mathbf{v}_u|$ varies with the velocity of the SSC $|\mathbf{v}_c|$, and, thus, E linearly depends on $|\mathbf{v}_u|^3$. The simplified energy consumption model is defined as follows:

$$E = |\mathbf{v}_u|^3 \frac{\|\mathbf{x}_i - \mathbf{x}_{i+1}\|_2}{10\frac{\mathrm{km}}{\mathrm{h}}}, \tag{10}$$

2) RRT algorithm:* The Rapidly-exploring Random Tree (RRT) algorithm [25] and its' computationally more efficient version RRT* [24] are stochastic sampling-based path planning algorithms commonly used in complex high-dimensional spaces with obstacles. It builds a searched tree $\mathcal{T} = \{\mathbf{t}_0, ..., \mathbf{t}_M\}, \mathbf{t} \in \mathbb{R}^2$ of M positions to find a feasible trajectory between the start point $\mathbf{x}_0$ and the goal point $\mathbf{x}_{end}$. Here, we utilize the energy consumption E (Eq. 9) as cost function to link the SSC to the path planning.

We use the RRT* algorithm provided by the open-source software `PythonRobotics` [26]. To take into account the complex geometries of the obstacles in the model area, we extend the `RRTStar` module by adding geographic information system features to the obstacles. Collision avoidance is guaranteed by creating safety boundaries with a distance of 500 m around the obstacles. Furthermore, we neglect the physical characteristics of the USV, assuming the USV to be a frictionless particle.

Since the SSC is variable in space and time, we update the SSC field every 15 minutes, which is the model time step of the BSH-HBMnoku ocean model [20]. As the RRT* algorithm stochastically generates a path, we generate a set of 20 path segments $\mathcal{S} = \{\mathbf{S}_1, ..., \mathbf{S}_{20}\}, \mathbf{S}_j \in \mathbb{R}^{L \times 2}, \forall j = 1, \ldots, 20$ with the RRT* algorithm in each time step, where L denotes the number of positions of each path segment $\mathbf{S}$. The altered RRT* selects then the path segment with the lowest cost c based on E. In the next time step, the end position of the previous selected path segment is then used as start position and 20 candidate paths are generated again. This procedure is repeated until the USV reaches the goal position. Alg. 3 summarises the path planning steps under the spatiotemporal variablity of the SSC.

Algorithm 3 Path planning in complex marine environments

Input: $\mathbf{x}_0, \mathbf{x}_{end}$, SSC
Output: $\mathcal{T}$
$\mathcal{T} \leftarrow \mathcal{T} \cup \mathbf{x}_0$ {initialize final path with start position}
$i = 0$ {initialize model time step}
while $\mathcal{T}^{[end]} \neq \mathbf{x}_{\mathbf{end}}$ **do**
  $\mathcal{S} \leftarrow NULL$
  for j = 1, ..., 20 **do**
    $\mathbf{S}_j \leftarrow \texttt{RRT}\star(\mathcal{T}^{[end]}, \mathbf{x}_{end}, \mathrm{SSC}_i)$
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathbf{S}_j\}$
  end for
  $\mathbf{S}_{min} \leftarrow \texttt{costmin}(\mathcal{S})$ {segment with lowest cost}
  $\mathcal{T} \leftarrow \mathcal{T} \cup \{\mathbf{S}_{min}\}$ {append segment to final path}
  $i = i + 1$
end while

C. Measurement point selection

After generating the uncertainty map, the USV aims to drive to the test point with the highest entropy. However, considering the challenging marine environment, it is crucial to incorporate the costs associated with the USV when choosing the next measurement points. While navigating towards the point with the highest entropy would yield the highest information gain, it may not always be feasible if the path requires a high cost in terms of the heuristic distance or SSC. In such cases, an alternative approach is to select measurement points that offer a relatively low path cost while still maintaining a relatively high entropy compared to the entire search area.

To address the trade-off between maximizing the informativeness and minimizing the energy consumption, we introduce a modified information metric taking into account the costs associated with the USV's operations. Dividing the entropy by the cost, a high value indicates a point with high informativeness and low cost. We compute the heuristic distance and the total energy consumption of each path generated by the RRT* planner and compare the following metrics: entropy, entropy/distance, and entropy/energy. For each search area, we choose the point that yields the highest value within each metric category. Finally, we compare the total energy consumption towards the selected measurement points.

IV. Results

We analyze three search areas within the model area (Fig. 4). For exemplary purposes, we only show the exploration and path planning results of search area 3 because it showcases both obstacles and data gaps. Subsequently, we compare different metrics based on the resulting energy consumption in each search area. Eventually, we investigate the total energy consumption of the USV varying travel direction and start time of the USV.

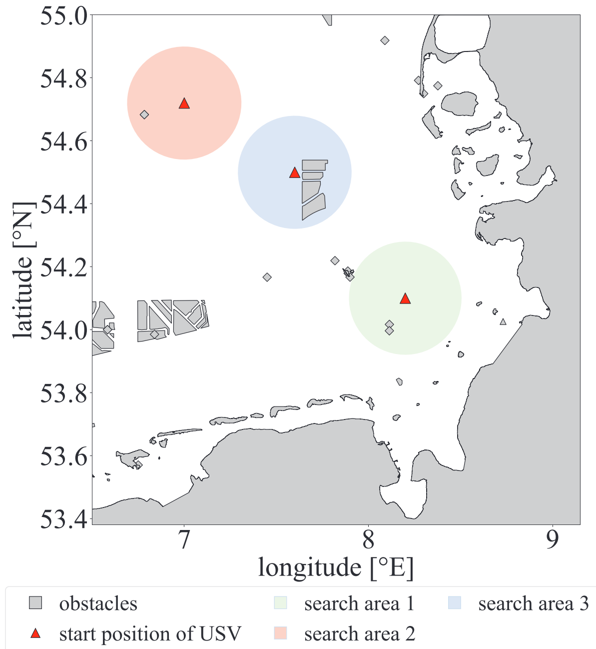

Fig. 4: Search areas investigated within the model area. Each search area has a radius of 20 km. In the following, the coordinates of the start positions are indicated. Search area 1: 7.0 °E, 54.72 °N. Search area 2: 7.6 °E, 54.5 °N. Search area 3: 8.2 °E, 54.1 °N.

A. *Entropy-based exploration and path planning with energy constraints: search area 3 as an example*

Clustering the resulting uncertainty map of Alg. 1 leads to local maxima depicted in Fig. 5. The local maxima are ranked according to their entropy values.

After identifying the most informative measurement points, we perform the SSC dependent RRT* planner for each local maximum to further analyze the energy consumption of the paths. Fig. 6 presents the final paths of an USV travelling with a constant total speed of $|\mathbf{v}_g| = 10\frac{\text{km}}{\text{h}}$ starting on November 3, 2021, at 00:45 UTC, which is three hours after the acquisition time when the NRT data are available. The cumulative sum of the energy consumption of each path is highlighted in Fig. 7.

Path 3 reveals to be the most energy efficient path with a straight trajectory and a total energy consumption of 87.8 kJ. On the other hand, paths 1 and 2 make a curve towards the Northwest. Fig. 8 displays the mean direction of the SSC in search area 3 between 00:45 UTC and 23:45 UTC on November 3, 2021. The Northwest direction of path 3 corresponds to the direction of the prevailing SSC at the start time with 321°. This indicates that the modified RRT* planner indeed searches for a path that follows the SSC to minimize the cost.

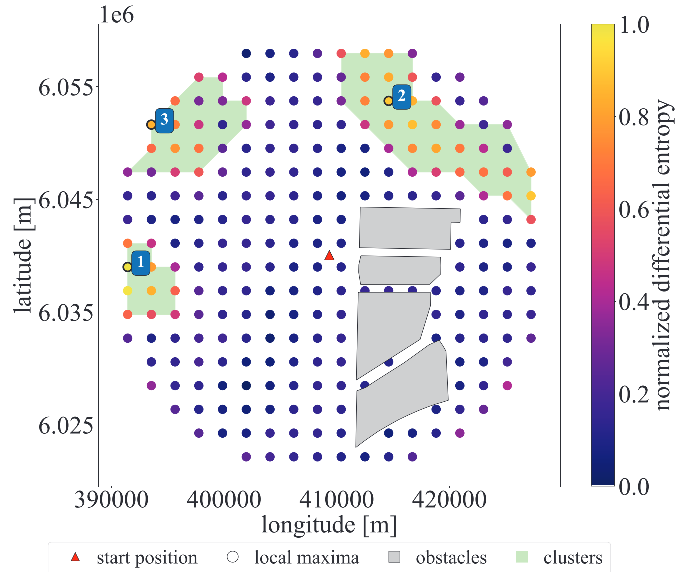

Fig. 5: Search area 3: Clustering the uncertainty map. The clusters are indicated in green, resulting in three local maxima. The local maxima are ranked according to their entropy value in a descending order and the labels denote the ranking.

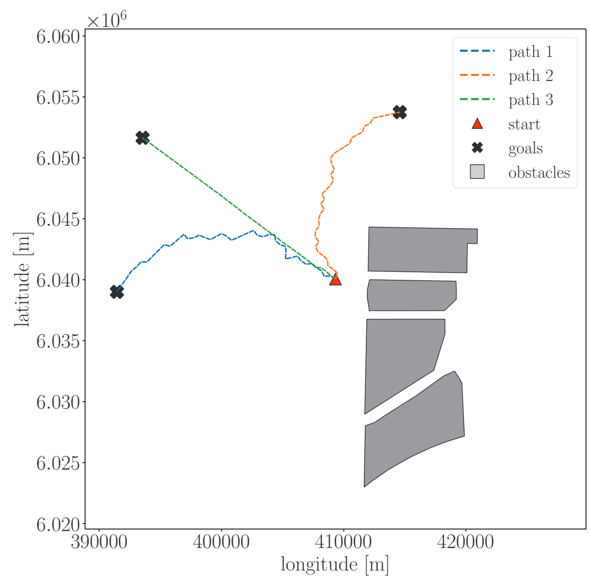

Fig. 6: Search area 3: Final paths towards the first (blue line), second (yellow line), and third (green line) local maximum. The USV travels with a constant total speed of 10 $\frac{\text{km}}{\text{h}}$ starting on November 3, 2021, at 00:45 UTC.

Put differently, the USV exhibits a preferred travel direction associated with the prevailing SSC. Furthermore, comparing path 2 and 3 demonstrates that the heuristic distance between the start and the goal position is inadequate for defining the cost since it does not account the dynamic environment of the ocean.

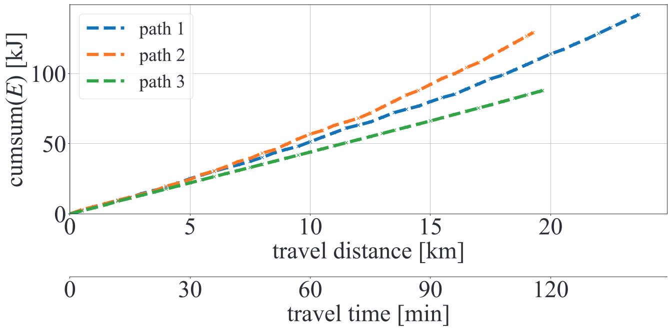

Fig. 7: Search area 3: Cumulative sum of the energy consumption [kJ] as a function of travel distance and time.

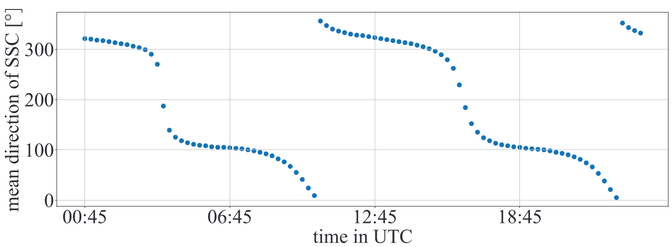

Fig. 8: Search area 3: Mean direction of the SSC between 00:45 UTC and 23:45 UTC on November 3, 2021.

B. Assessment of the new metrics on the energy consumption

Balancing out both informativeness and costs in our decision-making process, we examine the newly introduced information metrics explained in Sec. III-C. Fig. 9 shows the information metrics in each search area and the total energy consumption of the respective paths.

In search area 1, the most informative point coincides with the highest energy consumption of 277.5 kJ. Both the entropy/distance and entropy/energy information metrics lead to the same test point selection, indicating that this specific path represents the shortest and most energy-efficient trajectory. In search area 2, the entropy/distance metric reveals the selection of a path that consumes the highest amount of energy, followed by the entropy metric. The entropy/energy metric showcases the lowest energy consumption. In search area 3, the entropy metric results in the highest energy consumption, followed by the entropy/distance metric. The entropy/energy metric exhibits the lowest energy consumption of 87.8 kJ.

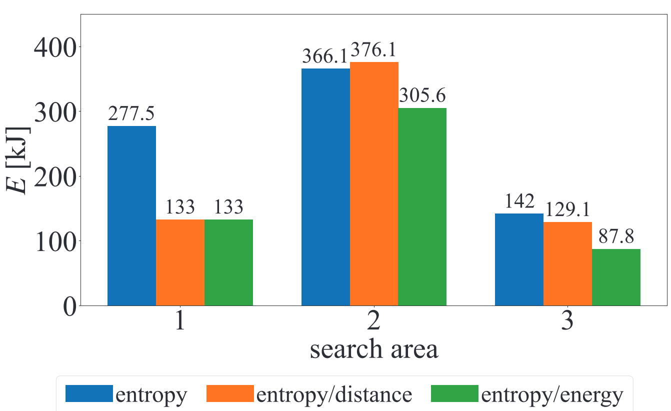

Fig. 9: Comparison between the metrics in each search area. Number above the bar indicates the total energy consumption. In all search areas, the entropy/energy metric stands out to be the most energy-efficient strategy, trading out informativeness and energy consumption.

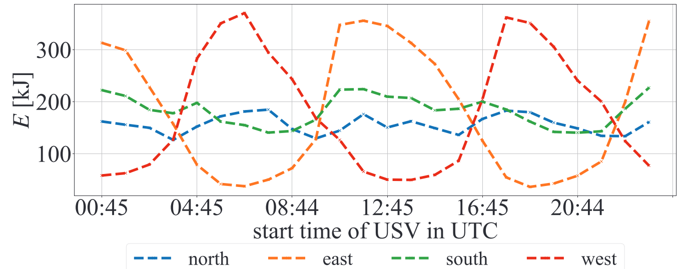

Fig. 10: Search area 1: The dependency of the total energy consumption [kJ] on the start time of the USV with a constant total speed of 10 $\frac{\text{km}}{\text{h}}$. Four goals are simulated directed to the north (blue), east (yellow), south (green), and west (red).

C. Time-dependent energy consumption

In Sec. IV-B, the USV showcases a preferred travel direction associated with the prevailing SSC. Since the SSC is not only variable in spatial dimension, but also in temporal dimension, the operating time of the USV plays an important role when assessing the energy consumption. Depending on the direction of the USV towards its goal position and the prevailing SSC, distinct paths resulting from the modified RRT* planner and energy consumption arise. It is therefore worthwhile to analyze the total energy consumption as a function of start time of the USV and the distinct travel directions. To neglect obstacles in this analysis, we show the time dependency of the total energy consumption in search area 1 as an example (Fig. 10, simulating an USV travelling towards the north, east, south, and west.

The energy consumption of the east and west paths follow a periodic behaviour with a periodicity of about 12 to 12.5 h and a phase difference of half a period to each other, corresponding to the semi-diurnal tidal waves dominating the North Sea [27]. The north and south paths display an almost steady energy consumption.

To further analyze the energy consumption of different start times, we compute the energy savings ΔE of two consecutive local extremes as follows:

$$\Delta E[\%] = \frac{\mathtt{max}_{\mathtt{local}}(E) - \mathtt{min}_{\mathtt{local}}(E)}{\mathtt{max}_{\mathtt{local}}(E)} \tag{11}$$

where $\mathtt{min}_{\mathtt{local}}(E)$ denotes a local minimum of the total energy consumption E and $\mathtt{max}_{\mathtt{local}}(E)$ the local maximum. Computing the energy savings ΔE for the east path, we get a value of 88.1%.

V. Discussion and Future Work

The energy assessment in Sec. IV-B highlights the limitation of relying solely on heuristic distance as cost metric, as it does not necessarily guarantee the most energy-efficient path due to the dynamics of the SSC. Our introduced entropy/energy information metric stands out to be the most suitable metric, trading out informativeness and energy consumption. Therefore, we recommend utilizing the entropy/energy metric in the decision-making process for optimal path selection in marine environments.

analyzing the time-dependent energy consumption in Sec. IV-C demonstrates the significant potential for energy savings and underscore the importance of scheduling USV start times based on tidal currents. By leveraging the natural flow of the tides, we improve the energy efficiency of the USV. A related work to these findings is the concept of active current selection for langrangian profilers introduced by [28]. This concept is inspired by the Tidal-Stream Transport observed in marine animals [29] to achieve energy-efficient navigation.

When constructing the model area, we only consider static obstacles such as land cover and offshore installations. To ensure safe autonomous navigation, our work should be extended by considering the dynamic nature of the North Sea and human influence with increasing volume of ship traffic. As the North Sea is dominated by semi-diurnal tidal waves, the water depth is a crucial parameter for safe navigation and feasibility of the USV's path. For a more realistic setup, the RRT* algorithm should include the spatiotemporal variability of low tide areas by incorporating the bathymetry of the model area. Apart from the environmental conditions in the North Sea, future work should incorporate dynamic obstacle avoidance strategies based on Automatic Identifications System (AIS) data.

To generate the uncertainty map introduced in Sec. III-A2, we compute the differential entropy [13] from the model variance σ_* of the GPR prediction. The differential entropy serves as an information metric for quantifying the absolute amount of information that each potential measurement point holds. However, it does not take into account the reduction of uncertainty after the measurement is taken [30]. To address this limitation, mutual information (MI) [31] serves as a preobservation indicator of the potential usefulness of acquiring information through a specific measurement [30] by taking into account the cross-correlations of the potential measurement points [32]. For future work, we recommend to use MI as it is the state-of-the-art metric to exploit information for GP-based models [12], [33]. The focus of our work is primarily on introducing a general concept of integrating USVs into operational oceanography using informative path planning methods. Therefore, we use differential entropy as the most basic information metric.

When identifying local maxima in the uncertainty map, we set the number of clusters k fixed to three which we choose based on the data gaps in search area 3. As each search area encompasses different data gaps, k should not be set fixed. Instead, finding the optimal value of k for each search area automatically can be accomplished by performing the elbow curve method [34].

Another crucial aspect currently not considered in our path planning is the constraint posed by energy consumption. Whether the USV is fuel-, solar- or wind-powered, there are limitations on the available resources. To address this, energy constraints should be employed in the cost function within path optimization. One approach is to control the relative velocity $\mathbf{v}_u$ of the USV. Since the energy consumption according to [14] is assumed to be linearly dependent on $\mathbf{v}_u^3$, $\mathbf{v}_u$ can emerge to infinity, which is technically not realistic. In a more realistic setup, the relationship between the energy consumption and $\mathbf{v}_u$ is non-linear.

Linked to the energy consumption model proposed by [14], we compute the cost of the USV based solely on the SSC. To develop a more realistic path planning approach and achieve more accurate trajectories, future works should incorporate winds into the energy consumption model, e.g. by [19]. By considering the influence of winds in the cost function of the RRT* planner, the model can capture a broader range of environmental factors that affect the energy consumption.

Furthermore, we simplify the USV to be a frictionless mass point. When operating in complex marine environments, the motion of the USV can become highly uncertain due to factors such as hydrodynamic forces and moments, and marine environmental interference forces [35]. Several works have addressed the stochastic motion of an USV such as [36] who proposed a stochastic optimal control.

Our approach focuses exclusively on employing a single USV for exploration tasks. To cover the entire North Sea and reduce the variance in satellite data to the minimum, employing multi-agent systems by using multiple USVs should be further investigated. Examples of such research are the works by [37]–[39]. While [37] developed a search algorithm based on particle swarm optimization and inertia Levy-flight for submarine exploration, [38] and [39] proposed an entropy-driven swarm exploration under sparsity constraints, evaluating various coordination strategies for a network of interconnected mobile agents.

The goal of our work is to find the most energy-efficient path towards an informative single measurement point. Going beyond, the next question is how many points could be measured on average until the next satellite acquisition under energy constraints. Rather than solely looking at a single measurement point, a broader sight is to consider the entire set of informative points, i.e. finding the most-energy efficient path that covers high-information locations along a path. A key aspect that needs to be addressed is the time dependency of the costs, which arises from the spatiotemporal dynamics of the SSC. This problem can be treated as a variation of the travelling salesman problem. Various algorithms have been developed to tackle the travelling salesman problem, such as ant colony optimization algorithms, particle swarm optimization algorithms, and genetic algorithms [40].

VI. Conclusion

This paper demonstrates the potential of USVs for gap imputation in satellite oceanographic data by combining energy consumption assessment and entropy-driven exploration. L2 SST data of SENTINEL-3A in the North Sea serve as training dataset in the GPR model to accurately estimate the SST revealing locations of high uncertainty. The differential entropy as information metric enables the identification of measurement points with high informativeness. Subsequently, we perform the modified RRT* planner, which is extended by the energy consumption model as cost function, to search for feasible and optimal paths considering the prevailing SSC, ensuring save and energy-efficient navigation of the USV in complex marine environments.

Simulation results reveal that the USV exhibits a preferred travel direction, and choosing this direction leads to the most energy-efficient path. The preferred travel direction is influenced by both space and time since it is based on the spatiotemporal variation of the SSC driven by the tidal currents that dominate the North Sea. Scheduling the start time of the USV according to the tidal currents can lead to energy savings up to 88%. Furthermore, applying information-driven exploration highlights the complexity of the decision-making process for the USV in selecting measurement points. The trade-off between informativeness and energy consumption poses a significant challenge. To address this, we recommend utilizing the entropy/energy metric as a means of achieving optimal path selection in marine environments.

Acknowledgment

The DFKI Lower Saxony (DFKI NI) is funded by the Lower Saxony Ministry of Science and Culture and the Volkswagen Foundation.

Information contained in this publication has been derived from data that is made available under the European Marine Observation Data Network (EMODnet) Human Activities project (https://emodnet.ec.europa.eu/en/human-activities), funded by the European Commission's Directorate-General for Maritime Affairs and Fisheries (DG MARE).

References

[1] O. Zielinski, J. A. Busch, A. D. Cembella, K. L. Daly, J. Engelbrektsson, A. K. Hannides, and H. Schmidt, "Detecting marine hazardous substances and organisms: sensors for pollutants, toxins, and pathogens," *Ocean Science*, vol. 5, no. 3, p. 329–349, Sep. 2009. [Online]. Available: http://dx.doi.org/10.5194/os-5-329-2009

[2] A. Schiller, B. Mourre, Y. Drillet, and G. Brassington, "An overview of operational oceanography," in *New Frontiers in Operational Oceanography*, E. P. Chassignet, A. Pascual, J. Tintoré, and J. Verron, Eds. GODAE OceanView, 2018.

[3] L. Ćatipović, F. Matić, and H. Kalinić, "Reconstruction methods in oceanographic satellite data observation—a survey," *Journal of Marine Science and Engineering*, vol. 11, no. 2, p. 340, 2023.

[4] A. Stock, A. Subramaniam, G. L. van Dijken, L. M. Wedding, K. R. Arrigo, M. M. Mills, M. A. Cameron, and F. Micheli, "Comparison of cloud-filling algorithms for marine satellite data," *Remote Sensing*, vol. 12, no. 20, p. 3313, 2020.

[5] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 3826–3852, 2013.

[6] S. Sunder, R. Ramsankaran, and B. Ramakrishnan, "Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from modis data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 228–240, 2020.

[7] J. Wang and Z. Deng, "Development of modis data-based algorithm for retrieving sea surface temperature in coastal waters," *Environmental monitoring and assessment*, vol. 189, no. 6, p. 286, 2017.

[8] M. Konik, M. Kowalewski, K. Bradtke, and M. Darecki, "The operational method of filling information gaps in satellite imagery using numerical models," *International Journal of Applied Earth Observation and Geoinformation*, vol. 75, pp. 68–82, 2019.

[9] T. M. Chin, J. Vazquez-Cuervo, and E. M. Armstrong, "A multi-scale high-resolution analysis of global sea surface temperature," *Remote Sensing of Environment*, vol. 200, pp. 154–169, 2017.

[10] K. Koutantou, P. Brunner, and J. Vazquez-Cuervo, "Validation of nasa sea surface temperature satellite products using saildrone data," *Remote Sensing*, vol. 15, no. 9, p. 2277, Apr. 2023. [Online]. Available: http://dx.doi.org/10.3390/rs15092277

[11] J. Vazquez-Cuervo, J. Gomez-Valdes, M. Bouali, L. Miranda, T. van der Stocken, W. Tang, and C. Gentemann, "Using saildrones to validate satellite-derived sea surface salinity and sea surface temperature along the california/baja coast," *Remote Sensing*, vol. 11, no. 17, p. 1964, 2019.

[12] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies." *Journal of Machine Learning Research*, vol. 9, no. 2, 2008. [Online]. Available: https://www.jmlr.org/papers/volume9/krause08a/krause08a.pdf

[13] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[14] H. Niu, Z. Ji, A. Savvaris, and A. Tsourdos, "Energy efficient path planning for unmanned surface vehicle in spatially-temporally variant environment," *Ocean Engineering*, vol. 196, p. 106766, 2020.

[15] Earth Resources Observation And Science (EROS) Center, "Srtm water body dataset," 2017. [Online]. Available: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-water-body

[16] C. Donlon, B. Berruti, A. Buongiorno, M.-H. Ferreira, P. Féménias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra, "The global monitoring for environment and security (gmes) sentinel-3 mission," *Remote Sensing of Environment*, vol. 120, pp. 37–57, 2012.

[17] P. Cornillon and L. Stramma, "The distribution of diurnal sea surface warming events in the western sargasso sea," *Journal of Geophysical Research*, vol. 90, no. C6, p. 11811, 1985.

[18] G. A. Wick, W. J. Emery, and P. Schluessel, "A comprehensive comparison between satellite-measured skin and multichannel sea surface temperature," *Journal of Geophysical Research*, vol. 97, no. C4, p. 5569, 1992.

[19] Y. Zhang, G. Shi, and J. Liu, "Dynamic energy-efficient path planning of unmanned surface vehicle under time-varying current and wind," *Journal of Marine Science and Engineering*, vol. 10, no. 6, p. 759, May 2022. [Online]. Available: http://dx.doi.org/10.3390/jmse10060759

[20] T. Brüning, X. Li, F. Schwichtenberg, and I. Lorkowski, "An operational, assimilative model system for hydrodynamic and biogeochemical applications for german coastal waters," *Hydrographische Nachrichten*, no. 118, pp. 6–15, 2021.

[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian process for machine learning*. London, England: The MIT Press, 2006.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] J. McQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967*, 1967, pp. 281–297.

[24] S. Karaman and E. Frazzoli, "Sampling-based algorithms for optimal motion planning."

[25] S. M. LaValle and J. J. Kuffner, "Randomized kinodynamic planning," *The International Journal of Robotics Research*, vol. 20, no. 5, pp. 378–400, 2001.

[26] A. Sakai, D. Ingram, J. Dinius, K. Chawla, A. Raffin, and A. Paques, "Pythonrobotics: a python code collection of robotics algorithms."

[27] H. Klein and A. Frohse, "Oceanographic processes in the german bight," *Die Küste*, no. 74, p. 60, 2008. [Online]. Available: https://hdl.handle.net/20.500.11970/101594

[28] J. Jouffroy, Q.-Y. Zhou, and O. Zielinski, "On active current selection for lagrangian profilers," *Modeling, Identification and Control: A Norwegian Research Bulletin*, vol. 34, no. 1, pp. 1–10, 2013.

[29] R. B. Forward and R. A. Tankersley, "Selective tidal-stream transport of marine animals," in *Oceanography and Marine Biology*. CRC Press, 2001, pp. 313–362.

[30] K. Yang, S. Keat Gan, and S. Sukkarieh, "A gaussian process-based rrt planner for the exploration of an unknown and cluttered environment with a uav," *Advanced Robotics*, vol. 27, no. 6, pp. 431–443, 2013.

[31] W. F. Caselton and J. V. Zidek, "Optimal monitoring network designs," *Statistics & Probability Letters*, vol. 2, no. 4, pp. 223–227, 1984.

[32] A. Viseras, "Distributed multi-robot exploration under complex constraints," Ph.D. dissertation, April 2018. [Online]. Available: https://elib.dlr.de/120427/

[33] S. Bai, T. Shan, F. Chen, L. Liu, and B. Englot, "Information-driven path planning," *Current Robotics Reports*, vol. 2, no. 2, pp. 177–188, 2021.

[34] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, p. 267–276, Dec. 1953. [Online]. Available: http://dx.doi.org/10.1007/BF02289263

[35] M. T. Vu, M. Van, D. H. P. Bui, Q. T. Do, T.-T. Huynh, S.-D. Lee, and H.-S. Choi, "Study on dynamic behavior of unmanned surface vehicle-linked unmanned underwater vehicle system for underwater exploration," *Sensors (Basel, Switzerland)*, vol. 20, no. 5, 2020.

[36] P. Patrinos, P. Sopasakis, H. Sarimveis, and A. Bemporad, "Stochastic model predictive control for constrained discrete-time markovian switching systems," *Automatica*, vol. 50, no. 10, pp. 2504–2514, 2014.

[37] C. Tholen, T. A. El-Mihoub, and L. Nolle, "On a novel search strategy based on a combination of particle swarm optimisation and levy-flight," in *ECMS 2018 Proceedings edited by Lars Nolle, Alexandra Burger, Christoph Tholen, Jens Werner, Jens Wellhausen*. ECMS, May 2018. [Online]. Available: http://dx.doi.org/10.7148/2018-0190

[38] C. Manss, D. Shutin, A. Viseras, and G. Leus, "Coordination methods for entropy-based multi-agent exploration under sparsity constraints," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 490–494.

[39] C. Manss, I. Kuehner, and D. Shutin, "Experimental validation of entropy-driven swarm exploration under sparsity constraints with sparse bayesian learning," *Entropy*, vol. 24, no. 5, 2022. [Online]. Available: https://www.mdpi.com/1099-4300/24/5/580

[40] C. Dahiya and S. Sangwan, "Literature review on travelling salesman problem," *International Journal of Research*, vol. 5, no. 16, pp. 1152–1155, 2018.