Adaptive Gaussian Process–Based Sampling for Energy-Efficient Aquatic Sensing with Autonomous Surface Vessels

Rodney Staggers Jr., Bharath Vedantha Desikan, Kanav Prashar, Carlos Torre, Jnaneshwar Das
DREAMS Lab, Arizona State University

***Abstract*—This research presents an autonomous adaptive sampling framework for aquatic sensing using an Autonomous Surface Vessel (ASV). The system integrates a commercial flight controller, the ROS 2 middleware, and an auxiliary onboard computer to enable this real-time architecture. The ASV collects geospatial and biogeochemical data—such as Sea Surface Temperature—and reconstructs continuous environmental fields using Gaussian Process regression. Both stationary and non-stationary kernel formulations are employed, allowing the model to incorporate Gaussian noise and positional uncertainty in the measurements. To prioritize informative regions, statistical hotspot detection is applied to the reconstructed fields, identifying significant spatial extrema while accounting for model uncertainty. This approach aims to maximize the accuracy of environmental field reconstruction while minimizing unnecessary traversal and reducing informational entropy. By explicitly favoring high-information regions over uniform coverage, the framework improves information gain per unit distance traveled, enabling more efficient use of limited onboard energy resources and supporting larger-area deployments under fixed battery constraints. Initial validation is conducted in a PX4 Simulation-in-the-Loop (SITL) environment with expectations of implementing on two low-cost (appx. $2,500 USD) ASVs.**

I. Introduction

Autonomous environmental monitoring platforms can play a vital role in understanding and managing aquatic ecosystems. As opposed to repeated, manual topographic assessments by researchers, the development of commercialized remote-sensing tools has enabled geomorphologists to document data at such increased rates and accuracy. In addition to this, the development of robotic platforms equipped with integrated sensor suites has demonstrated the collection of high-resolution spatio-temporal data in large water bodies, allowing assessments of environmental health in areas ravaged by disasters or pollutants [1], [2]. Nonetheless, current aquatic monitoring methods are still labor-intensive, spatially limited, and offer low spatio-temporal resolution. Limitations in power allocation, time, and traversal constrain researchers to optimize the deployment sites of these platforms to estimate scalar field reconstructions.

The R/V *Karin Valentine*, as depicted in Fig. 1 and 2, is a low-size, weight, and power (SWaP) autonomous surface vessel (ASV) developed for this purpose, incorporating a robust set of hardware optimized for autonomous *in-situ* data collection and optimal sample return for *ex-situ* analysis. This system architecture supports real-time sensor integration, state estimation using an onboard Inertial Measurement Unit (IMU) and external Global Positioning System (GPS) module, and autonomous operation. Its payload—consisting of a 1D sonar for depth mapping and a multiprobe water-quality sonde—enables real-time bathymetric and biogeochemical analysis.

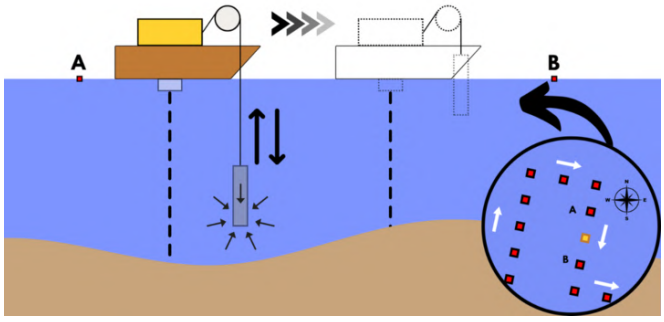


Fig. 1: GPS, bathymetry, and water quality data are collected as the R/V travels to subsequent waypoints

To transform raw spatial data into interpretable environmental insights, this work employs Gaussian Process (GP) Regression [3] to reconstruct continuous scalar fields [4], such as temperature distributions.

II. Related Work

The deployment of autonomous vehicles for water sampling is not a novel concept, as this area of research has experienced a strong resurgence since the turn of the millennium. Recent systems have integrated sonar, multiparameter sondes, satellites, LiDAR, and geospatial sensing equipment to generate dense spatio-temporal datasets over lakes, rivers, and coastal environments [2], [5]. In recent years, groups such as the University of Sevilla, have taken advantage of advancements in flight controllers to perform lithological and bathymetric measurements on aquatic environments using an autonomous surface vessel (ASV) [6], [7]. On the autonomy front, tasks such as station-keeping and obstacle avoidance have been demonstrated in aquatic environments [8], [9].

To improve sampling efficiency, researchers have investigated informative path planning and adaptive sampling strategies that select future measurements according to expected information gain. Such approaches have been demonstrated on a number of autonomous marine vehicles in estimates to guide exploration [10], [11]. Despite the progress, the application of Kac-Rice-based significance testing to non-stationary GP reconstructions remains largely unexplored in environmental monitoring contexts.

Fig. 2: Research vessel (sans winch) docked at Tempe Town Lake in Arizona, USA

III. Problem Statement and Contributions

A. Problem Statement

Mobile environmental sensing produces spatial measurements along sparse, irregular trajectories. The underlying environmental processes, however, form continuous spatial fields whose structure must be inferred from these limited observations. Standard interpolation methods provide point estimates but do not have associated uncertainty, and peak detection applied to interpolated fields typically ignores statistical significance.

In addition, stationary GP models impose a single global smoothness scale across the domain, which may obscure localized features or introduce interpolation artifacts. When peaks are identified in reconstructed fields, it is therefore unclear whether they correspond to genuine spatial structure or are expected fluctuations under the assumed prior.

This work addresses the problem of uncertainty-aware spatial field reconstruction and statistically principled peak detection from sparse trajectory-based measurements, including the extension from stationary to non-stationary Gaussian Process models.

B. Contributions

The contributions of this work are:

1) A scalable, cost-effective (sub-$3,000) autonomous limnological research platform integrating water-quality sensing, bathymetric mapping, and autonomous navigation.
2) Extension of Kac–Rice peak detection from the stationary RBF case to the non-stationary Paciorek kernel using Gaussian conditioning and numerical kernel derivatives for active processing of incoming environmental data.

IV. Design Parameters

A. Hardware

The surface vessel, as depicted in Fig. 3 and 4, is designed in the form of a catamaran boat measuring $130\,\text{cm} \times 88.6\,\text{cm} \times 42.3\,\text{cm}$, with a total mass of $21.3\,\text{kg}$. It is powered by two $12\,\text{V}$, $20\,\text{Ah}$ $LiFePO_4$ batteries connected in parallel, and typically exerts 75-151 Wh of energy for a typical mission (the parameters of which will be described later). With the use of two T200 Blue Robotics thrusters, which can provide a maximum thrust of 9.02/7.04 kgf (FWD/REV), the vehicle can achieve a cruise speed ranging from 0.7 to $2\,\text{m/s}$.

The boat's avionics package includes a Pixhawk flight controller running v1.16.1 PX4 autopilot stack for low-level control, and an ARM processor companion computer that runs ROS2 for high-level decision making and data handling. This structure leverages the PX4 simulation-in-the-loop (SITL) and "ros2 bag" storage system to ascertain the aforementioned data. Additionally, it is able to take advantage of the embedded packages and scripts needed to cache the

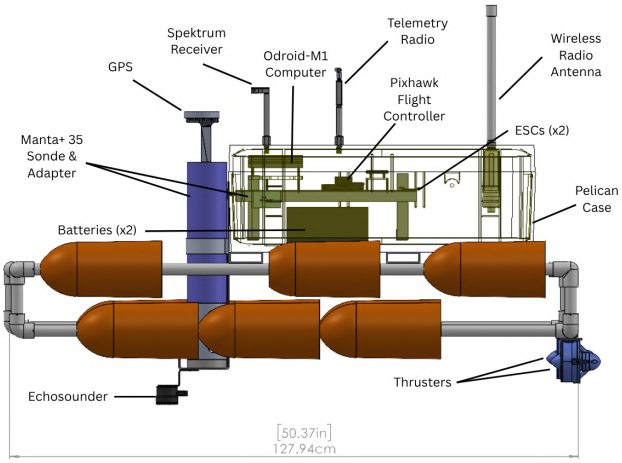


Fig. 3: CAD model of the autonomous surface vessel

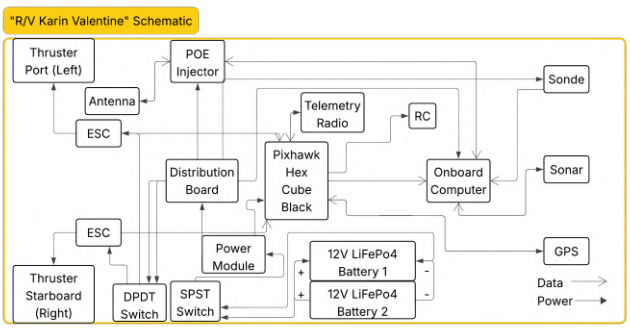


Fig. 4: Schematic demonstrating the boat's device connections

geospatial, biogeochemical, and bathymetric data (which will be discussed later).

As the vessel traverses the region, it collects depth and hydrological measurements using an echosounder and multiparameter sonde. Bathymetry is measured with a Blue Robotics 1D Ping echosounder, which has an operating range of 0.3–100 m and a transducer frequency of 115 kHz. The device has a 25° beam width and connects to the onboard computer via a UART–USB adapter.

A Manta+35 multiparameter sonde measures the following biogeochemical features: Temperature [°C], pH, Colored Dissolved Organic Matter (CDOM) [ppb], Chlorophyll [μg/L], Turbidity [NTU], Conductivity [μS/cm], Dissolved Oxygen [mg/L], and (most recently) Salinity [PSU]. Table I outlines the resolution and functional range for each device. This system utilizes a RS232-USB adapter, with a +12V VDC, to integrate to the ODROID-M1. The noise parameters of the instrument can be included into the covariance matrix of the Gaussian process regression to accurately depict the uncertainties.

Geospatial positioning is provided by a CubePilot Here+ V2 RTK GNSS receiver, enabling centimeter-level accuracy. Given the use of RTK (Real-Time Kinematics), the precision of its pose estimation is optimized to a minimum error of 0.025 m. This is plugged into the ”GPS1” serial port of the Pixhawk, and relayed to the ODROID-M1 via the ”TELEM2” port.

On average, GPS data (running through an Extended Kalman Filter), sonar data, and sonde data are collected at 50, 10 and 1 Hz, respectively.

B. Communications

The vehicle uses three communication links with the ground station.

The ground station connects to the PX4's ground control application, QGroundControl, via a pair of 915 MHz Telemetry radios. These MAVLink-supported single half-duplex devices provide easy access to planning boustrophedon surveys.

The Spektrum DSMX Quad Race Serial Receiver, which operates off of an agile 2.4 GHz bandwidth, enables vehicle teleoperation with a RC transmitter at the ground station. Via this communication, the execution of various control modes (i.e. Manual or Mission mode) can be initiated by the researcher.

TABLE I: Initial Sonde Sensor Measurement Specifications

Category	Parameter	Range	Resolution	Accuracy
Temperature	temperature	-5 to 50 °C	0.01	±0.1 (within 10 °C of calibration)
pH/ORP	pH	0 to 14 units	0.01	±0.1
Turbidity	turbidity	0 to 1000 FNU	0.01	±0.3 FNU or ±2%
	turbidity	1000 to 4000 FNU	0.1	±4%
Dissolved Oxygen	concentration	0 to 20 mg/L	0.01	±0.1
	concentration	20 to 30 mg/L	0.1	±0.15
	concentration	30 to 50 mg/L	0.1	±5%
	% saturation	0 to 500%	0.01	same as conc.
Conductivity	spec. cond.	0 to 5000 µS/cm	0.1	±0.5% or ±1 µS/cm
Pressure	depth	0 to 25 m	0.01	±0.05
	depth	0 to 200 m	0.01	±0.4
Fluorescence	Chl (blue)	0 to 100 µg/L	0.01	-
	CDOM	0 to 300 ppb	0.01	-

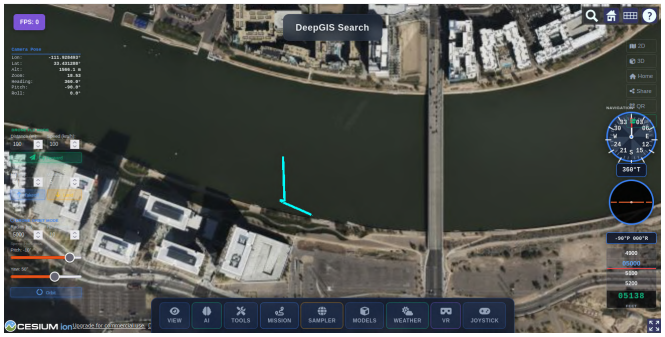


Fig. 5: The segment of the vehicle's pose from a December 6th, 2024 mission is visualized with DeepGIS

Finally, a pair of 2.4 GHz Ubiquiti Bullet wireless radios (one on the boat and one at the ground station) provide an IP link that enables:

- SSH access to the onboard computer for initializing packages and executing shell scripts,
- Real-time monitoring of sensor data streams,
- Upload of collected data to the DeepGIS-XR visualization platform (Fig. 5).

DeepGIS-XR supports semantic mapping over satellite imagery and simulation-based environmental visualization for digital twin integration.

C. Software Implementation

The execution of autonomous missions relies on several software packages. The MAVROS package enables communication between the onboard ROS2 system and the PX4 flight controller. It bridges messages between the flight controller, onboard computer, and ground station via the MAVLink protocol. MAVLink is a lightweight publish-subscribe protocol suited to systems with limited computational resources, such as the flight controller (256 KB RAM). QGroundControl uses MAVLink to configure and execute waypoint-based boustrophedon surveys. The system initially operated on a Holybro Pixhawk 4 (FMUv5) and now runs on a Hex Cube Black (FMUv3 firmware). The vessel operates under the PX4 “Differential Rover” airframe configuration.

A motorized winch is integrated to enable vertical profiling of the water column. Upon reaching a sampling location, the vessel drifts while the sonde is lowered in 1 m increments with 3–5 s dwell times. The winch continues to deploy until the probe is within 1 m of the bathymetric depth measured by the sonar. The sonde is then retrieved to the surface before the vessel proceeds to the next waypoint (Fig. 1). A feedback loop to estimate the length of dispersed rope is monitored by the use of dual-channel hall effect sensor (efficient for determining the speed and direction of the winch) and 20 magnets equally spaced on the side of the spooled system.

Because MAVROS no longer provides direct access to low-level PWM or normalized thrust commands for the PX4 infrastructure, integration of the winch required an alternative control strategy. Although PX4 exposes actuator control

through its internal uORB messaging system, this interface is not accessible via MAVLink. Therefore, the winch is controlled using MAVROS service calls within a callback loop to achieve reliable deployment and retrieval.

In addition to the MAVROS library, the ping_sonar_ros library was used to handle the UART Serial communication between the onboard computer and the echosounder. A python script was written to decode the serial port and convert the data into a readable string that is stored (along with the bathymetry, GPS, pose, velocity, and vehicle status data) in a ros2 bag.

V. Methodology

A survey mission resembles a boustrophedon pattern with five-meter turns in a 100 m $\times$ 100 m section of the lake; the waypoint system uses a 0.5 m threshold radius.

This research is conducted at Tempe Town Lake, a man-made lake. The lake has a mean depth, width, and length of 5.5 m, 240 m, and 3.22 km, respectively. Ideally, our operations would seek to adaptively sample a predetermined region (2,500-12,000 m^2) to identify optimal regions of interest. The platform enables 4D modeling of ecosystem dynamics by integrating bathymetric and biogeochemical data, fostering sustained ecological monitoring. Field deployments in lake environments validate its consistency and adaptability across diverse conditions. Using Gaussian process regression, we generate 2D and 3D water-quality maps rendered interactively through Folium. Fig. 6 shows that in post-processing, asynchronously collected data are fused to align biogeochemical characteristics with their geospatial coordinates. Subsequently, interpolations are performed with Gaussian process regression, to estimate the parameters' behavior in the observed region [12].

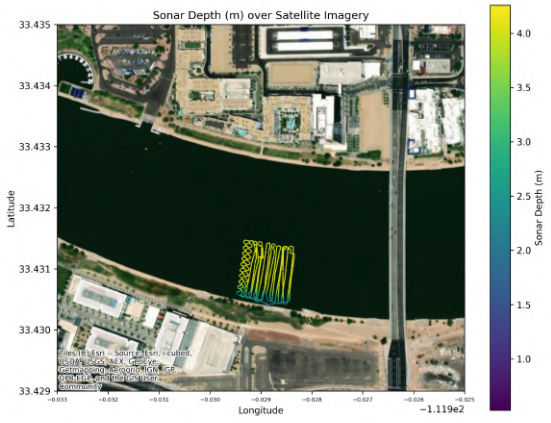


Fig. 6: Satellite imagery with bathymetric survey data plotted against GPS tracks of explored region

VI. Gaussian Process Spatial Reconstruction

This section describes the spatial reconstruction of water quality variables from sparse boat-mounted sensor data using Gaussian processes. We use two kernel formulations: a stationary RBF kernel and a non-stationary Paciorek kernel with spatially varying lengthscales. Detected peaks in the reconstructed fields are tested for statistical significance using the Kac-Rice formula.

A. Gaussian Process Regression

A Gaussian process (GP) is a distribution over functions where any finite collection of function values is jointly Gaussian [13]. We use GPs to build continuous spatial maps from the sparse sensor readings collected by R/V Karin Valentine.

Given N observations $\mathbf{y}$ at spatial locations $\mathbf{X}$, the GP gives us two things at any unobserved location $\mathbf{x}_*$: a best estimate (the posterior mean) and a measure of how uncertain that estimate is (the posterior variance). The mean is a weighted combination of the observed values, with weights determined by how close $\mathbf{x}_*$ is to the training points. The variance is small near data and large in regions where we have no observations.

Both are computed from the kernel matrix K (which encodes how correlated any two locations are) and the noise variance σ_n^2:

$$\mu(\mathbf{x}_*) = \mathbf{k}_*^\top (K + \sigma_n^2 I)^{-1} \mathbf{y} \tag{1}$$

$$\sigma^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (K + \sigma_n^2 I)^{-1} \mathbf{k}_* \tag{2}$$

where $\mathbf{k}_*$ holds the covariances between $\mathbf{x}_*$ and the training points.

1) Stationary RBF Kernel: The kernel function determines how the GP generalizes from observed to unobserved locations. We use the radial basis function (RBF) kernel:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2} \right) \tag{3}$$

where σ_f^2 controls the overall amplitude and ℓ is the lengthscale. A larger ℓ means the GP treats distant points as correlated, producing smoother maps. A smaller ℓ allows more spatial variation. This kernel is called *stationary* because it depends only on the distance between points, not on where they are.

The hyperparameters (ℓ, σ_f^2, σ_n^2, and a mean constant) are learned by maximizing the log marginal likelihood (MLL) with 80 steps of Adam optimization. A log-normal prior on the lengthscale keeps it from shrinking too small when data is dense.

B. Non-Stationary GP with Paciorek Kernel

A stationary kernel applies the same smoothness everywhere, but water quality fields can have sharp gradients near inflows and smooth behaviour in open water. The Paciorek kernel [14] handles this by letting the lengthscale vary across space.

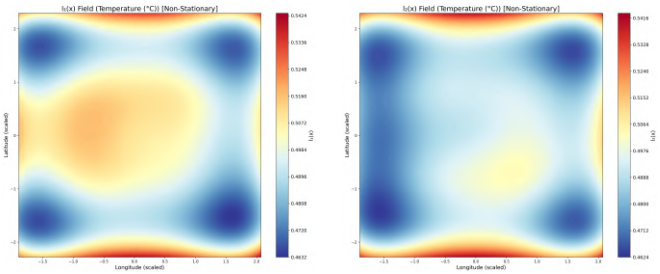


Fig. 7: Learned lengthscale fields $\ell_1(\mathbf{x})$ (left) and $\ell_2(\mathbf{x})$ (right) for the non-stationary GP, December 6 temperature. Shorter lengthscales (blue) mean finer spatial detail.

At each location $\mathbf{x}$, a local 2×2 covariance matrix $\boldsymbol{\Sigma}(\mathbf{x})$ describes how the kernel stretches in different directions. It is built from two directional lengthscales $\ell_1(\mathbf{x})$, $\ell_2(\mathbf{x})$ and a rotation angle $\theta(\mathbf{x})$:

$$\boldsymbol{\Sigma}(\mathbf{x}) = R(\theta)\,\mathrm{diag}(\ell_1^2, \ell_2^2)\,R(\theta)^\top \tag{4}$$

The kernel then combines the local covariance from both points:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2\,\frac{|\boldsymbol{\Sigma}_\mathbf{x}|^{1/4}\,|\boldsymbol{\Sigma}_{\mathbf{x}'}|^{1/4}}{|\bar{\boldsymbol{\Sigma}}|^{1/2}}\,\exp\!\Big(-\tfrac{1}{2}\mathbf{d}^\top \bar{\boldsymbol{\Sigma}}^{-1}\mathbf{d}\Big) \tag{5}$$

where $\bar{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}_\mathbf{x} + \boldsymbol{\Sigma}_{\mathbf{x}'})/2$ is the average and $\mathbf{d} = \mathbf{x} - \mathbf{x}'$.

1) Spatial Parameterization: The functions $\ell_1(\mathbf{x})$, $\ell_2(\mathbf{x})$, and $\theta(\mathbf{x})$ are parameterized using 25 RBF basis functions on a 5×5 grid. The lengthscales are passed through a sigmoid to stay within bounds $[\ell_{\min}, \ell_{\max}]$. This gives 76 learnable parameters total (25 each for ℓ_1, ℓ_2, θ, plus the log signal variance), optimized via MAP with 150 Adam steps and per-parameter learning rates.

Fig. 7 shows the learned lengthscale fields for a representative survey.

C. Kac-Rice Peak Detection

The GP posterior mean is a smooth surface with peaks and valleys. Some of these correspond to real spatial features (a warm inflow, an algal bloom), but others are just interpolation artifacts. We need a principled way to separate the two.

The Kac-Rice formula [15] (Theorem 11.2.1) gives us exactly that. It counts how many local maxima above a threshold u we would *expect* from a Gaussian random field based on its kernel alone:

$$\mathbb{E}[M_u] = \int_T p_{\nabla X}(\mathbf{0}) \times \mathbb{E}\Big[\big|\det \nabla^2 X\big|\;\mathbb{1}_{\nabla^2 X \prec 0}\;\mathbb{1}_{X \geq u}\;\Big|\;\nabla X = \mathbf{0}\Big]\,d\mathbf{t} \tag{6}$$

In words: at each point, the formula asks how likely the gradient is to be zero there (i.e. a critical point exists), and if so, how likely the Hessian is negative definite (a true peak, not a saddle) and the value exceeds u. Each sample is weighted by $|\det \nabla^2 X|$, so sharper peaks count more.

The expected count under the fitted kernel provides a natural reference distribution against which observed peaks can be compared. Since the kernel hyperparameters are themselves learned from data via marginal likelihood maximization, this comparison is empirical Bayes in nature: it measures whether an observed peak is taller than what the data-fitted kernel's intrinsic peak-height distribution would typically produce. We therefore treat the quantity below as a peak-prominence score rather than a strict posterior test [16]:

$$p\text{-value} = \frac{\mathbb{E}[\text{peaks above } x_{\text{obs}}]}{\mathbb{E}[\text{all peaks}]} \tag{7}$$

A small score means the peak is unlikely under the fitted GP's own peak-height distribution, and is therefore unlikely to be an interpolation artifact of that kernel structure.

1) How it works for the non-stationary Paciorek kernel: For the Paciorek kernel, two key challenges arise [16].

First, the gradient is no longer independent of the value and Hessian [17]. Knowing that the gradient is zero at a peak changes what we know about the curvature there. We handle this with standard Gaussian conditioning (the Schur complement):

$$\boldsymbol{\Sigma}_{\text{cond}} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\,\boldsymbol{\Sigma}_{BB}^{-1}\,\boldsymbol{\Sigma}_{BA} \tag{8}$$

where A = (value, Hessian) and B = (gradient). For the stationary RBF, $\boldsymbol{\Sigma}_{AB} = \mathbf{0}$ and the correction vanishes.

Second, the Paciorek kernel has no closed-form derivatives. The covariance entries must be computed numerically via finite differences on $k(\mathbf{t}, \mathbf{t}')$ at each peak location, since the kernel is different everywhere. Peaks where the resulting matrix fails a positive semi-definite check are conservatively assigned $p = 1.0$.

Algorithm 1 summarizes this.

Because the non-stationary GP adapts its lengthscale to the data, it tends to produce smoother posterior fields, and fewer peaks reach significance compared to the stationary RBF.

This formulation uses the fitted GP as the reference distribution, while the peak being scored comes from the posterior mean conditioned on the same data. A posterior-predictive variant, computing expected peaks under $X \mid \text{data}$, would

Algorithm 1 Non-Stationary Paciorek Peak Detection

Require: Paciorek kernel $k(\mathbf{t}, \mathbf{t}')$; posterior mean grid
Find candidate peaks via local maximum filter
for each peak at $\mathbf{t}_0$ with height x_{obs} **do**
  Compute 6×6 covariance at $\mathbf{t}_0$ via finite differences
  Condition on gradient $= 0$ via (8)
  if $\boldsymbol{\Sigma}_{\text{cond}}$ not positive semi-definite **then**
    Assign $p = 1.0$ (skip this peak)
  end if
  Standardize: $z \leftarrow (x_{\text{obs}} - \mu_0)/\sqrt{k(\mathbf{t}_0, \mathbf{t}_0)}$
  if $p < 0.05$ **then**
    Mark as significant (hot if $z > 0$, cold if $z < 0$)
  end if
end for

offer a stronger Bayesian interpretation and is left to future work.

D. Application: Tempe Town Lake

We apply the GP reconstruction and Kac-Rice peak detection to water quality data from Tempe Town Lake, Arizona. A sensor-equipped boat traverses the lake recording GPS-tagged measurements of the parameters mentioned in Section IV. Six surveys from September 2024 through February 2025 are processed independently.

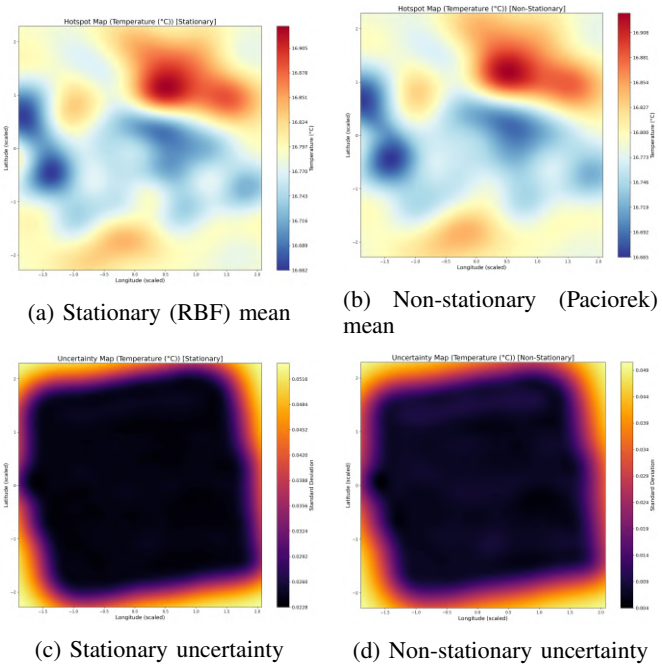
1) Data Processing: GPS coordinates (WGS84) are projected to UTM Zone 12N via `pyproj`, giving metric coordinates. These are then standardized (zero mean, unit variance) with `scikit-learn`'s `StandardScaler`. Target values are mean-centered ($y \leftarrow y - \bar{y}$) but not scaled, so the GP posterior stays in original measurement units.

Both GP models subsample to a maximum of 800 training points per survey. For the non-stationary Paciorek kernel this is necessary because exact GP inference scales as $O(N^3)$, and beyond 800 points the conjugate gradient solver in GPyTorch tends to diverge. The stationary GP also subsamples to 800 points to prevent the lengthscale from collapsing to very small values with dense data. Predictions are made on a 150×150 grid covering the lake domain.

2) Reconstruction Results: Fig. 8 shows the GP posterior mean and uncertainty for a representative temperature survey. The stationary RBF produces a smooth interpolation with uncertainty growing away from the sampling trajectory. The non-stationary Paciorek kernel adapts its spatial resolution across the domain.

3) Hotspot Detection Results: Fig. 9 shows the Kac-Rice results. Red stars mark significant hot peaks ($p < 0.05$), blue stars mark significant cold spots ($p < 0.05$), and grey dots are non-significant local extrema. The p-value annotations give a quantitative confidence measure for each detected feature.

4) Implementation: Both GP models use GPyTorch [18] with GPU acceleration. The stationary RBF model runs 80 Adam steps for MLL optimization; the non-stationary Paciorek model runs 150 steps with per-parameter learning rates. The Monte Carlo Kac-Rice computation draws 100,000 samples per peak.



(a) Stationary (RBF) mean

(b) Non-stationary (Paciorek) mean

(c) Stationary uncertainty

(d) Non-stationary uncertainty

Fig. 8: GP reconstruction of temperature, December 6th, 2024 survey. Top: posterior mean in original units. Bottom: posterior standard deviation.

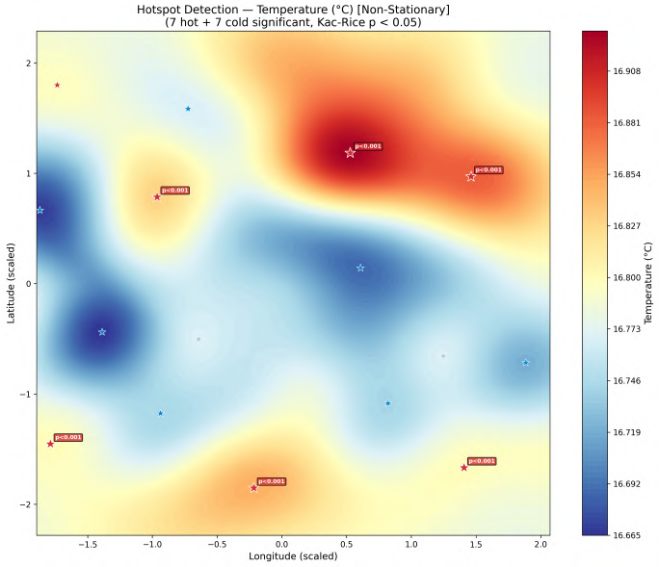


Fig. 9: Kac-Rice hotspot detection on Non-stationary GP reconstructions for Temperature, December 6th, 2024 survey. Red stars: significant hot peaks. Blue stars: significant cold spots. Grey dots: non-significant extrema.

VII. Integration of Adaptive Sampling in PX4 SITL

In the Gazebo simulation environment (depicted in Fig. 10), a vehicle running the same firmware and configuration as the deployed ASVs is used, ensuring consistency between simulated and real-world flight stacks. This process investigates autonomous adaptive sampling for aquatic environmental monitoring using Gaussian Process–based informative path planning implemented on a PX4 rover. Similar processes have been performed in the past for online aquatic reconstruction, specifically for Autonomous Underwater Vehicles (AUVs) [19], [20]. Within the environment, a PX4 R1 rover traverses a 25 m × 25 m domain containing a synthetic Gaussian temperature field. The system initializes by collecting measurements at 10 predefined waypoints, after which a Gaussian Process regression model is fit to estimate a continuous spatial field along with its associated predictive uncertainty. At each iteration, the planner selects the next sampling location by maximizing an information-theoretic acquisition function based on mutual information, penalized

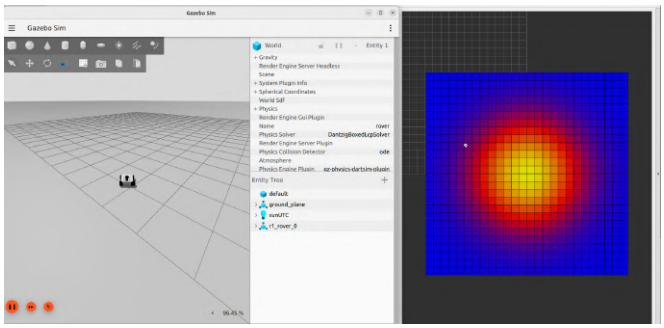


Fig. 10: PX4 R1 Rover explores domain with Gaussian temperature field, while computing next waypoint from mutual information function

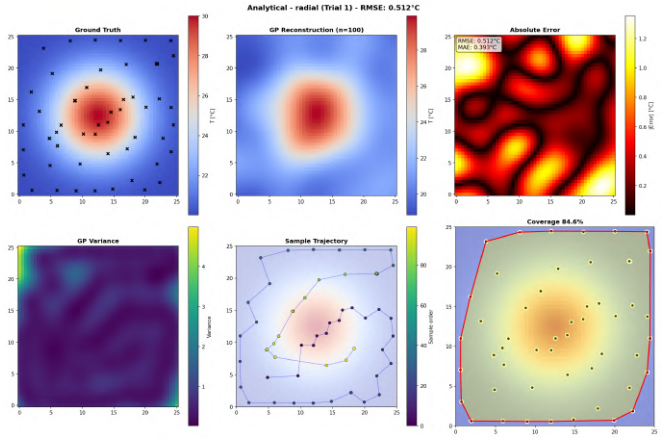


Fig. 11: Starting from upper left and continuing clockwise: (a) the Ground Truth, (b) Reconstructed Field, (c) Absolute Error, (d) Path coverage, (e) Trajectory, and (f) Variance of simulated rover adaptively sampling in geofence domain

by a travel cost term. This objective balances information gain with traversal efficiency, discouraging unnecessary motion. The mission terminates after 100 samples, at which point the reconstructed field is evaluated against ground truth using standard error metrics. In regions where candidate locations yield similar information gain, the planner favors shorter paths, thereby reducing total traversal distance and associated power consumption. Fig. 11 depicts how the selection of 110 (10 predefined and 100 computed from sampling) were able to cover approximately 84.6 % of the field and Root Mean Square Error of $0.512°C$. Solutions are being explored to minimize uncertainty along the field edges.

VIII. Conclusion

This work presented a Gaussian Process–based approach for reconstructing continuous spatial fields from sparse, trajectory-based water-quality measurements collected by a low-cost autonomous surface vessel. Stationary RBF and non-stationary Paciorek kernels were evaluated on multiple Tempe Town Lake surveys to study how kernel structure affects reconstructed spatial patterns.

Local extrema in the reconstructed fields were evaluated using the Kac–Rice formulation, producing model-consistent p-values for peak heights under the implied Gaussian random field. Across surveys, the non-stationary model modified local smoothness relative to the stationary baseline and, in turn, changed the set of extrema that reached significance.

Simulation results demonstrate the feasibility of integrating probabilistic environmental modeling directly into the autonomous sensing workflow. These capabilities establish a foundation for information-driven environmental monitoring in which future measurements are selected according to their expected contribution to model accuracy rather than predetermined coverage patterns. Such adaptive approaches have the potential to increase information gain per unit energy expenditure, extend mission effectiveness, and improve characterization of spatially heterogeneous aquatic environments.

The resulting pipeline supports both retrospective survey analysis and identification of candidate regions for targeted follow-up sampling during subsequent deployments.

IX. Future Work

We aim to evaluate sampling efficiency by measuring information gain (e.g., reduction in posterior variance) per unit distance or energy in both heterogeneous and mottled environments, thereby identifying conditions under which adaptive non-stationary GP sampling provides the greatest modeling benefit. The ASV first collects synchronized temperature (T_i) and GPS measurements along an initial coverage path (x_i, y_i), which are used to fit a non-stationary Gaussian Process model with a Paciorek kernel. The model then evaluates an acquisition function within a predefined geofence to select subsequent sampling locations that maximize expected information gain while minimizing travel cost. Newly acquired measurements are incorporated into the dataset, the GP is updated, and the process repeats until a mission termination criterion is reached. We will additionally investigate the effects of grid resolution and local maximum detection on the Kac–Rice Monte Carlo estimator in regions containing closely spaced extrema.

As shown in Fig. 12, secondary missions with an informative path can be estimated after the first lawnmower survey, by maximizing a utility function. Additionally, multi-robot mapping can be carried out with a second autonomous vessel (under construction) to both increase the coverage area of the experiment and reduce mapping time.

As the winched system conducts repeated limnological analyses over an extended period of time, the framework to estimate geospatial, bathymetric, and biogeochemical variables is expanded to enable a 4D spatio-temporal model. We integrate existing tools for coral semantic mapping and water column visualization. In situ measurements are compared with archived satellite observations (e.g., Landsat, MODIS). [21]. For example, temperature profiles collected by the ASV can be aligned with satellite-derived sea surface temperature to estimate vertical temperature gradients.

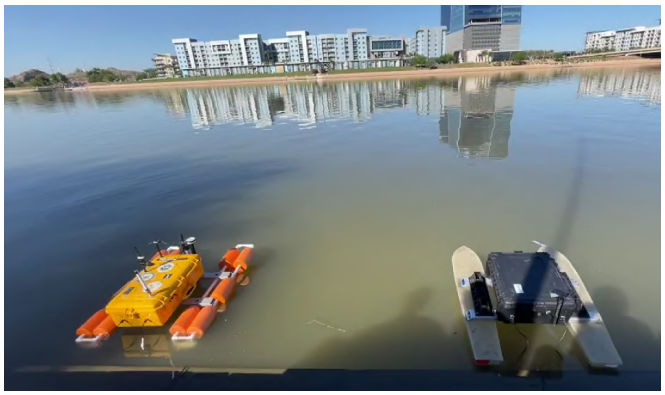


Fig. 12: Future missions seek to expand the efficacy of ecosystem monitoring by implementing a synergistic fleet of vehicles

REFERENCES

[1] K. C. Cavanaugh, T. W. Bell, K. E. Aerni, J. E. Byrnes, S. McCammon, and M. M. Smith, "New technologies for monitoring coastal ecosystem dynamics," *Annual Review of Marine Science*, vol. 17, no. Volume 17, 2025, pp. 409–433, 2025. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-marine-040523-020221

[2] J. C. Brock and S. J. Purkis, "The emerging role of lidar remote sensing in coastal research and resource management," *Journal of Coastal Research*, vol. 53, no. SI, pp. 1–5, 2009.

[3] Williams, Christopher and Rasmussen, Carl, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds., vol. 8. MIT Press, 1995. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf90577442771720a370c3c723-Paper.pdf

[4] P. B. Jónsson, J. Wang, and J. Kim, "Scalar field reconstruction based on the Gaussian process and adaptive sampling," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 2017, pp. 442–445.

[5] L. Somoza, V. Dıaz-del Rıo, R. León, M. Ivanov, M. Fernández-Puga, J. Gardner, F. Hernández-Molina, L. Pinheiro, J. Rodero, A. Lobato *et al.*, "Seabed morphology and hydrocarbon seepage in the gulf of cadiz mud volcano area: Acoustic imagery, multibeam and ultra-high resolution seismic data," *Marine geology*, vol. 195, no. 1-4, pp. 153–176, 2003.

[6] L. M. Díaz, S. Y. Luis, A. M. Barrionuevo, D. S. Diop, M. Perales, A. Casado, S. Toral, and D. Gutiérrez, "Towards an Autonomous Surface Vehicle Prototype for Artificial Intelligence Applications of Water Quality Monitoring," 2024. [Online]. Available: https://arxiv.org/abs/2410.05892

[7] S. Manjanna, A. Q. Li, R. N. Smith, I. Rekleitis, and G. Dudek, "Heterogeneous multi-robot system for exploration and strategic water sampling," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4873–4880.

[8] A. Pereira, J. Das, and G. S. Sukhatme, "An experimental study of station keeping on an underactuated asv," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 3164–3171.

[9] F. Sotelo-Torres, L. V. Alvarez, and R. C. Roberts, "An unmanned surface vehicle (USV): Development of an autonomous boat with a sensor integration system for bathymetric surveys," *Sensors*, vol. 23, no. 9, p. 4420, 2023.

[10] T. Wilson and S. B. Williams, "Adaptive path planning for depth-constrained bathymetric mapping with an autonomous surface vessel," *Journal of Field Robotics*, vol. 35, no. 3, pp. 345–358, 2018.

[11] 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. 8, pp. 235–284, 2008. [Online]. Available: http://jmlr.org/papers/v9/krause08a.html

[12] T. He, M. M. Tsui, A. B. Mayfield, P.-J. Liu, T.-H. Chen, L.-H. Wang, T.-Y. Fan, P. K. Lam, and M. B. Murphy, "Organic ultraviolet filter mixture promotes bleaching of reef corals upon the threat of elevated seawater temperature," *Science of The Total Environment*, vol. 876, p. 162744, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969723013608

[13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005. [Online]. Available: https://doi.org/10.7551/mitpress/3206.001.0001

[14] C. Paciorek and M. Schervish, "Nonstationary covariance functions for gaussian process regression," in *Advances in Neural Information Processing Systems*, S. Thrun, L. Saul, and B. Schölkopf, Eds., vol. 16. MIT Press, 2003. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2003/file/326a8c055c0d04f5b06544665d8bb3ea-Paper.pdf

[15] R. J. Adler and J. E. Taylor, *Random Fields and Geometry*, ser. Springer Monographs in Mathematics. Springer New York, NY, 2007.

[16] D. Cheng, "On local maxima of smooth gaussian nonstationary processes and stationary planar fields with trends," *Stochastic Processes and their Applications*, vol. 181, p. 104560, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304414924002680

[17] Y. Zhao, D. Cheng, S. Davenport, and A. Schwartzman, "On the peak height distribution of non-stationary gaussian random fields: 1d general covariance and scale space," 2025. [Online]. Available: https://arxiv.org/abs/2502.12452

[18] J. R. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson, "GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration," in *Advances in Neural Information Processing Systems*, 2018.

[19] S. Yan, Y. Li, X. Feng, S. Li, Y. Tang, Z. Li, and M. Yuan, "An auv adaptive sampling path planning method based on online model prediction," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 323–328, 2019, 12th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896319322128

[20] G. E. Berget, T. O. Fossum, T. A. Johansen, J. Eidsvik, and K. Rajan, "Adaptive sampling of ocean processes using an auv with a gaussian proxy model," *IFAC-PapersOnLine*, vol. 51, no. 29, pp. 238–243, 2018, 11th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896318321980

[21] J.-C. Jang and K.-A. Park, "High-resolution sea surface temperature retrieval from landsat 8 oli/tirs data at coastal regions," *Remote Sensing*, vol. 11, no. 22, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/22/2687