Adaptive Gaussian Process–Based Sampling for Energy-Efficient Aquatic Sensing with ASVs

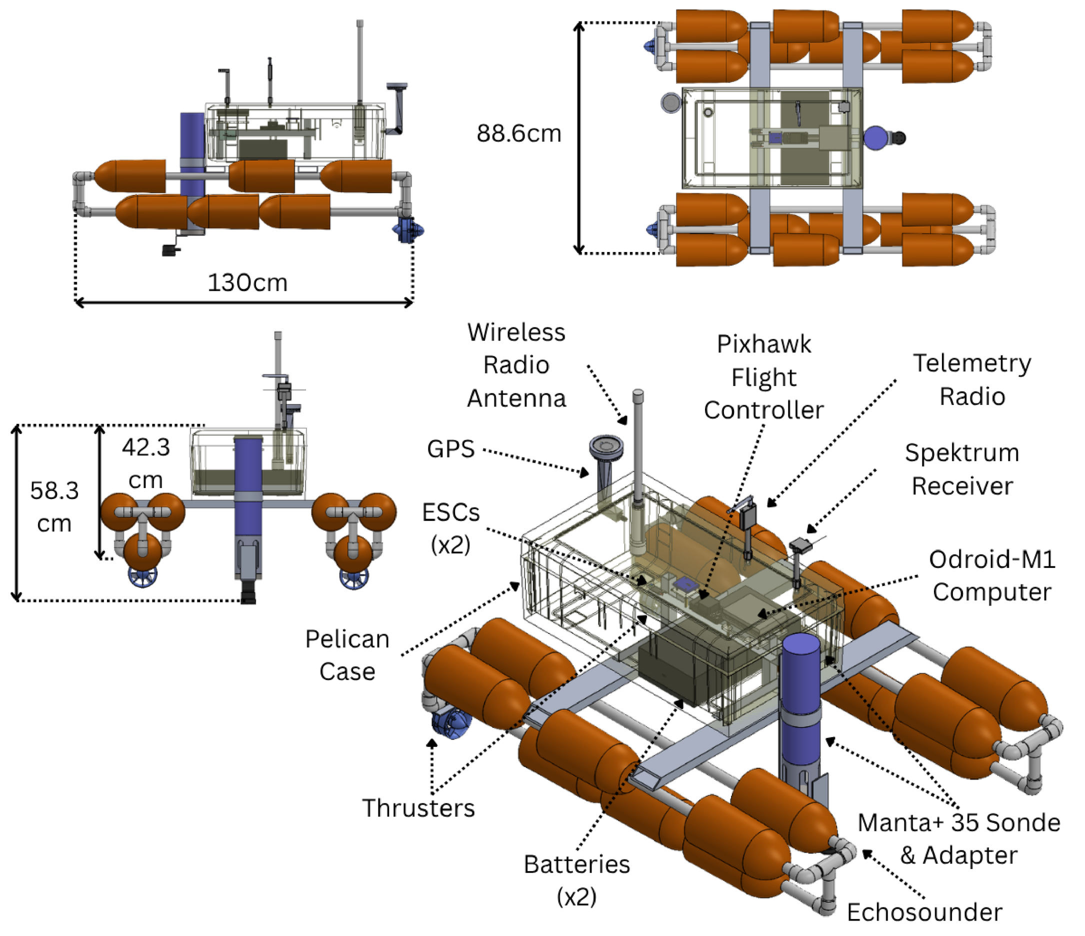
Rodney Staggers Jr. · Bharath Vedantha Desikan · Kanav Prashar · Carlos Torre · Jnaneshwar Das
DREAMS Lab • School for Engineering of Matter, Transport & Energy • Arizona State University

Introduction & Motivation

- Environmental aquatic monitoring is labor-intensive, spatially limited, and energy-constrained
- Autonomous Surface Vessels (ASVs) can help [1], but sparse trajectory data make it hard to:
  - Reconstruct *continuous* environmental fields
  - Separate real spatial features (algal blooms, thermal inflows) from artifacts
  - Maximize **information gain per unit energy**

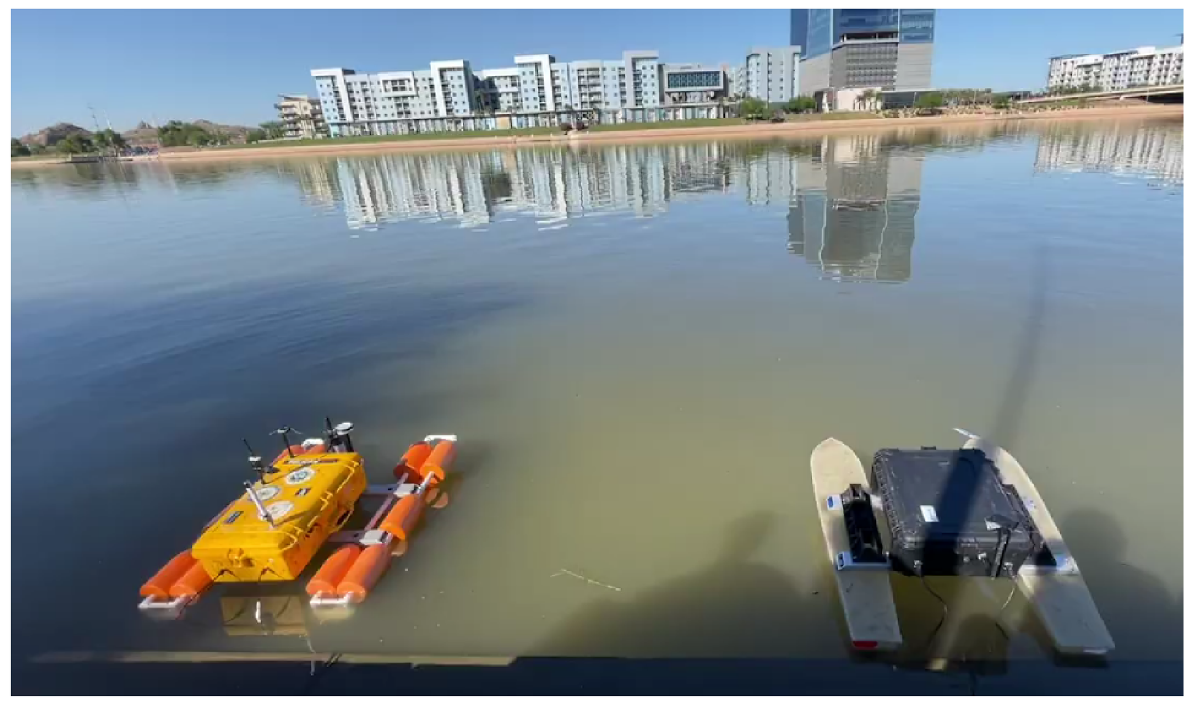
This work: GP regression (stationary + non-stationary) optimizes interpolation with Kac-Rice significance testing on a **sub-\$2,500 ASV**.

R/V *Karin Valentine*

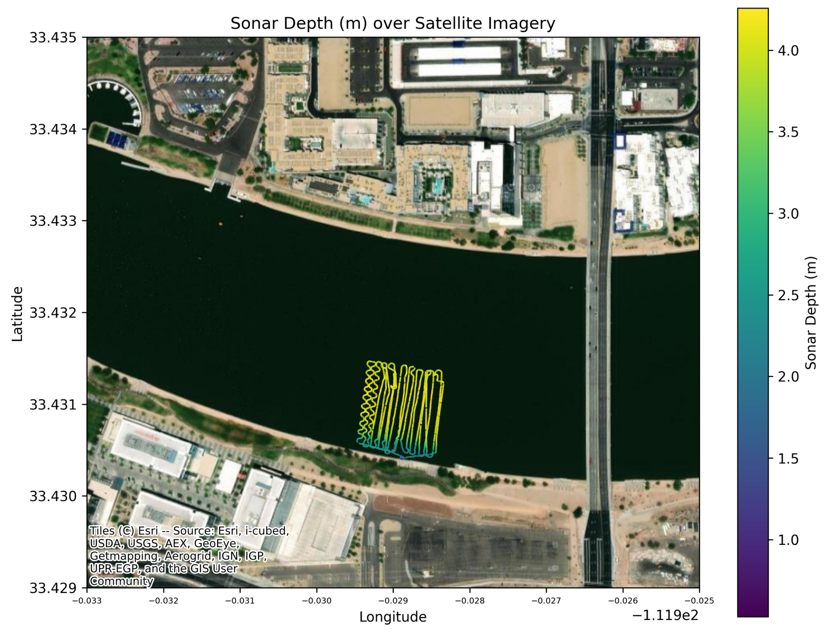


CAD model of the autonomous catamaran ASV.

Autopilot	Pixhawk / PX4 v1.16.1
Computer	ODROID-M1 + ROS 2
Power	2×12 V / 20 Ah $LiFePO_4$
Speed	0.7–2.0 m/s
RTK GNSS	±2.5 cm accuracy
Sonde	℃, pH, DO, Chl, CDOM, NTU, Conductivity, Depth
Sonar	0.3–100 m echosounder
Winch	Vertical column profiling



Deployment of a fleet of ASVs in Tempe Town Lake, Arizona, USA



Bathymetric survey overlaid on satellite imagery, Tempe Town Lake, Dec 2024.

Gaussian Process Regression

Given observations $\mathbf{y}$ at locations $\mathbf{X}$, the GP posterior at $\mathbf{x}_*$ is:

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

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

Sonde noise σ_n^2 enters the covariance matrix directly.

Stationary RBF Kernel

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

Single global lengthscale ℓ (80 Adam steps, MLL). Good for homogeneous fields; may miss sharp gradients.

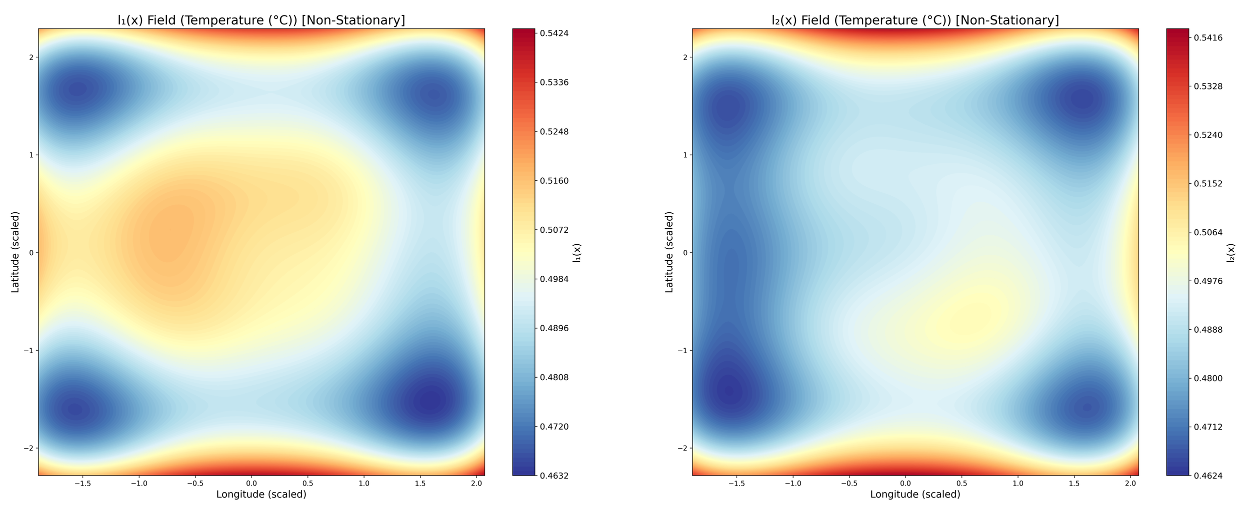
Non-Stationary Paciorek Kernel [2]

Water quality has *spatially varying* smoothness. A local 2×2 matrix encodes direction-dependent lengthscales:

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

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

ℓ_1, ℓ_2, θ vary spatially via **25 RBF basis functions** on a 5×5 grid (76 params, 150 Adam steps).



$\ell_1(\mathbf{x})$ $\ell_2(\mathbf{x})$
Learned lengthscale fields — blue (short) = finer spatial detail.

Kac-Rice Peak Significance

Rather than flagging every local maximum, we test each against the kernel's own peak-height distribution [3]:

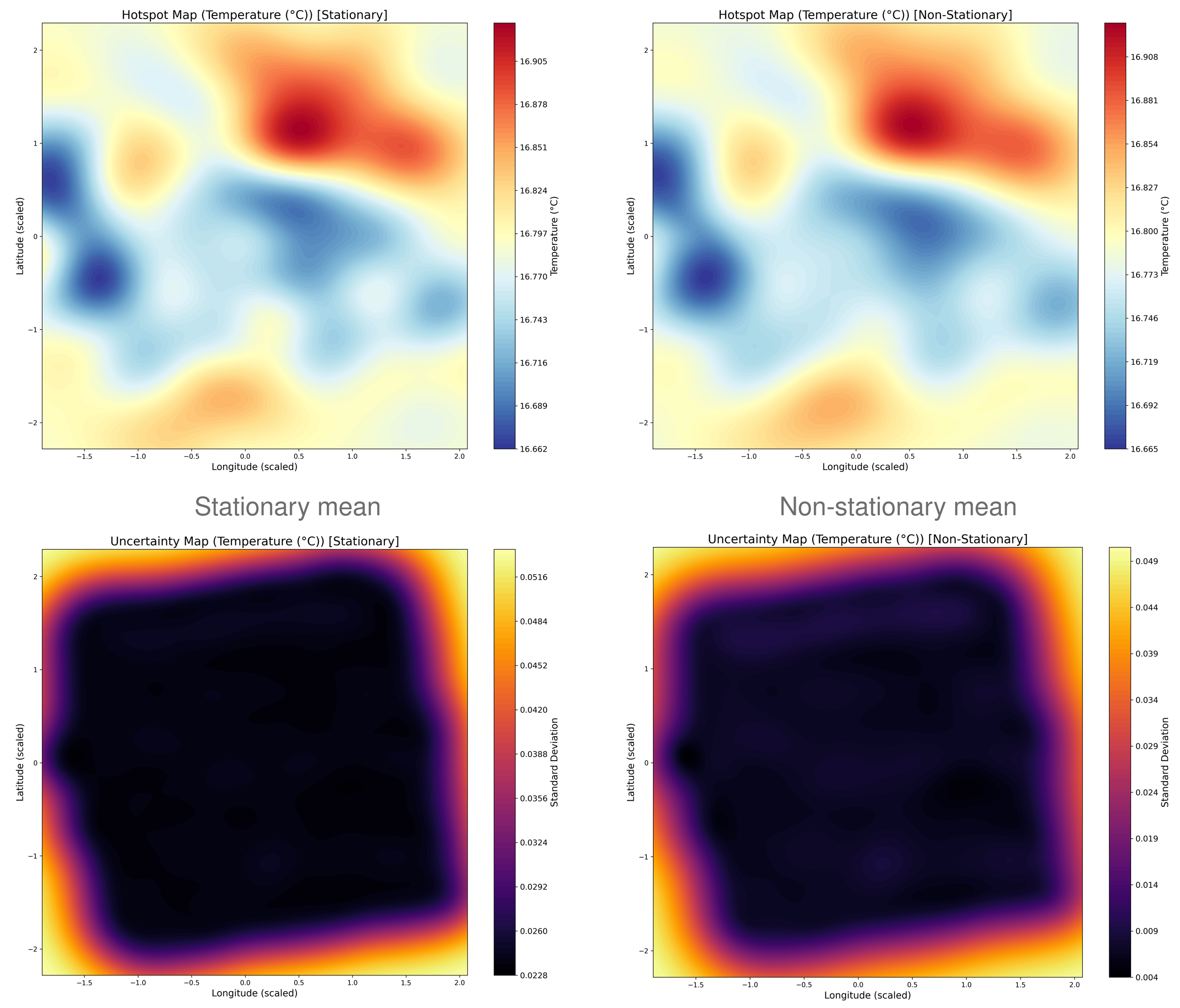
$$\mathbb{E}[M_u] = \int_T p_{\nabla X}(\mathbf{0})\, \mathbb{E}\left[\left|\det \nabla^2 X\right| \mathbb{1}_{\nabla^2 X \prec 0}\, \mathbb{1}_{X \geq u} \,\middle|\, \nabla X = \mathbf{0}\right] dt$$

Prominence: $p = \dfrac{\mathbb{E}[\text{peaks} > x_{\text{obs}}]}{\mathbb{E}[\text{all peaks}]}$

For the **Paciorek kernel**, gradient $\not\perp$ (value, Hessian) — corrected via Schur complement: $\mathbf{\Sigma}_{\text{cond}} = \mathbf{\Sigma}_{AA} - \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}\mathbf{\Sigma}_{BA}$

Kernel derivatives via finite differences; 100k MC samples per peak.

GP Reconstruction — Sample Temperature Survey

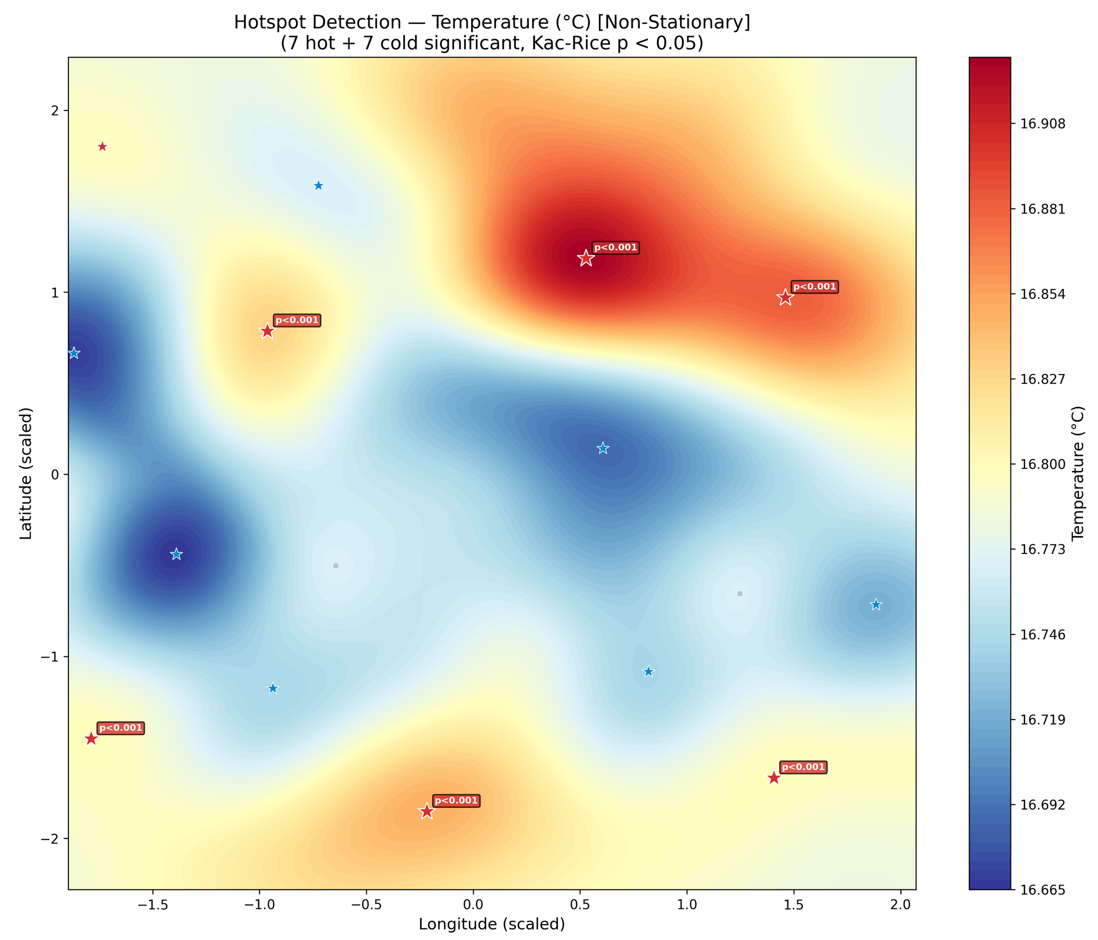


Stationary mean — Non-stationary mean

Stationary σ — Non-stationary σ

Non-stationary GP adapts smoothness to local data density. Sample from December 6th, 2024 dataset of Tempe Town Lake

Hotspot Detection — Kac-Rice Results



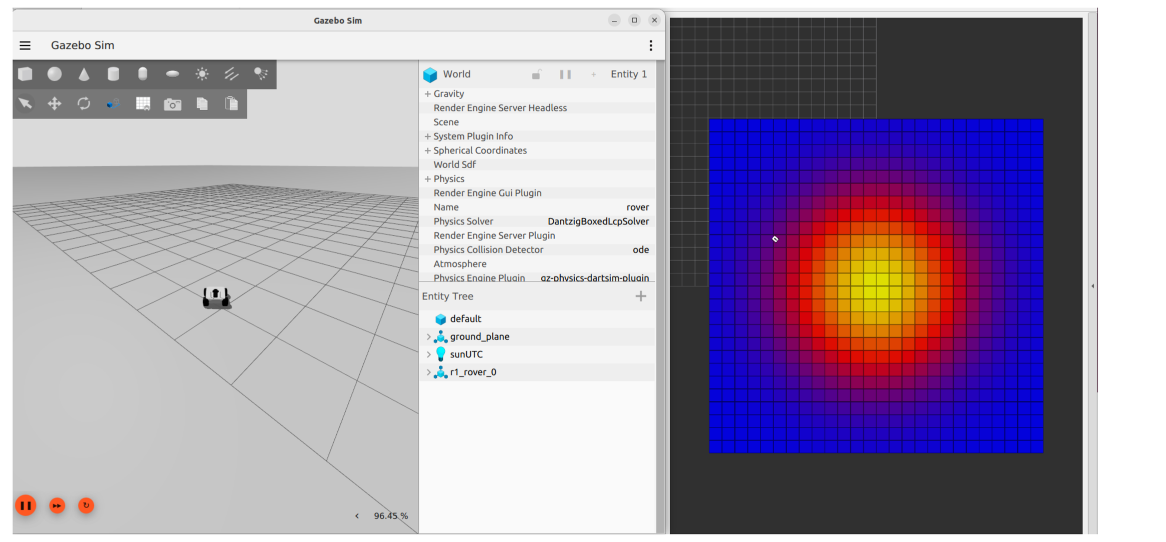
Non-stationary GP hotspot detection (Dec 6 temperature). Red = significant hot peak ($p<0.05$). Blue = cold spot. Grey = non-significant.

References

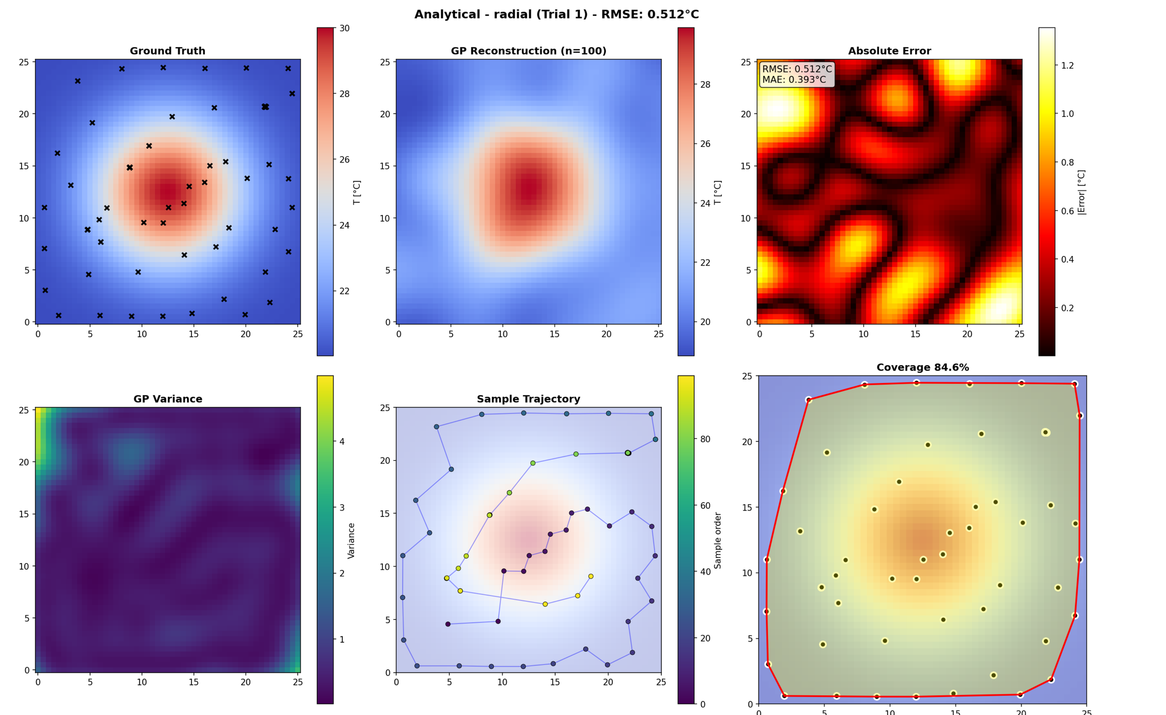
- L. M. Díaz, S. Y. Luis, A. M. Barrionuevo, D. S. Diop, M. Perales, A. Casado, S. Toral, and D. Gutiérrez, "Towards an Autonomous Surface Vehicle Prototype for Artificial Intelligence Applications of Water Quality Monitoring," 2024. [Online]. Available: https://arxiv.org/abs/2410.05892
- C. Paciorek 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
- R. J. Adler and J. E. Taylor, *Random Fields and Geometry*, ser. Springer Monographs in Mathematics. Springer New York, NY, 2007.

Adaptive Sampling — PX4 SITL

A PX4 R1 rover explores a 25×25 m Gazebo domain with a synthetic Gaussian temperature field. At each step the planner maximizes **mutual information** penalized by travel cost.



PX4 rover in Gazebo SITL computing the next waypoint via mutual information.



After 110 samples: **84.6% coverage**, RMSE $= 0.51\,^\circ$C.
(a) truth (b) reconstructed (c) error (d) coverage (e) trajectory (f) variance.

Key Contributions

1. **Sub-\$2,500 limnological ASV** — 8-parameter sonde, RTK GNSS, ROS 2 autonomy, motorized winch for vertical profiling
2. **Non-stationary Kac-Rice** — first extension of peak-detection significance testing from the RBF kernel to the Paciorek kernel via Gaussian conditioning and numerical derivatives

Conclusions

- Paciorek kernel modifies local smoothness, changing which extrema reach $p<0.05$
- Kac-Rice gives model-consistent empirical-Bayes p-values; non-stationary GP produces fewer false peaks
- SITL results confirm feasibility of information-driven autonomous surveys

Future Work

- **Multi-robot deployment** with second ASV (under construction)
- **4D spatio-temporal** modeling via repeated winch profiles over time
- Satellite cross-validation against Landsat / MODIS SST
- Posterior-predictive Kac-Rice under full $X \mid$ data distribution