Adaptive Sensing of Time Series with Application to Remote Exploration

David R. Thompson[1]
Nathalie A. Cabrol[2]
Michael Furlong[3]
Craig Hardgrove[4]
Bryan Kian Hsiang Low[5]
Jeffrey Moersch[6]
David Wettergreen[3]

[1]Jet Propulsion Laboratory,
California Institute of Technology, USA
[2]SETI Institute, USA
[3]The Robotics Institute,
Carnegie Mellon University, USA
[4]Malin Space Sciences, Inc., USA
[5]National University of Singapore
[6]University of Tennessee, USA

A Portion of the research described in this presentation was carried out at the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration..

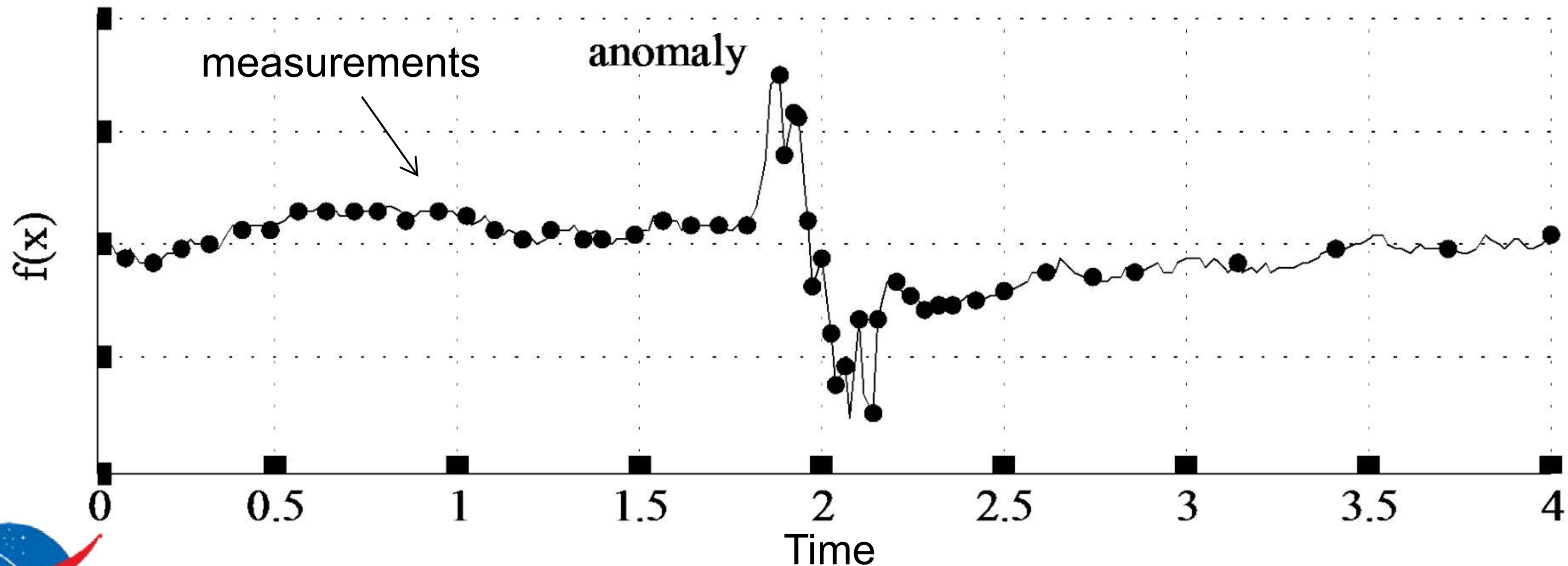
Why adaptive sampling?

Enable better robotic explorers like rovers and AUVs

Detect dynamic events with remote sensor networks

Handle uncertainty during spacecraft flybys, encounters

More adaptive "passenger" sensor payloads



Special challenges of time series data

Measurement imbalance Samples lie in the past, but we can only change the future. To plan we must extrapolate beyond the range of collected data.

Need for accurate second order statistics The value of a new observation is related to prediction certainty

Limited control options The agent may only choose the length of time to wait before the next sample

Computational efficiency Power-limited sensors must react quickly to capture transient events.

Relevant background work

Active learning in general [Cohn et al., 1996]

D-optimal spatial experimental design [Cressie 1991, Shewry & Wynn, 1987]

Information-optimal sampling for Markov chains [Krause & Guestrin 2005, Thompson et al., 2008]

ARIMA-based adaptive time series models [Law et al., 2009]

Our assumptions

1. The agent measures a **background process** that is:
   - slowly-varying
   - scalar
   - with additive white noise
2. The process varies more rapidly during rare **anomalies**
3. The number of measurements is limited by **resource constraints**
4. The agent can analyze collected data and revise its sampling plan on the fly.
5. We want to optimize **information gain** with respect to the underlying noiseless values.

Formulation

An agent observes a time series $t \in T$

Each timestep has independent variables $\mathbf{x}_t \in \mathbb{R}^d$

Data collection yields scalar measurements $y_t \in \mathbb{R}$

Assume the process is generated by an underlying function

$$f(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$$

$$y = f(\mathbf{x}) + \mathcal{N}(0, \sigma^2)$$

Underlying function　　Measurement noise

Gaussian process model

Assume measured values are jointly Gaussian-distributed

Place a prior P(Y) over all observations using a *covariance function* $\kappa(\mathbf{x}_i, \mathbf{x}_j)$

Evaluated between all observation pairs, it forms a covariance matrix K such that:

$$P(Y) \sim \mathcal{N}(0, K_X + \sigma^2 I) \;\propto\; e^{-\frac{1}{2} Y^T (K_X + \sigma^2 I)^{-1} Y}$$

A common choice is the squared exponential:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \psi_1 + \psi_2 \exp\left\{ -\frac{1}{2} \sum_{k=1}^{d} \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{w_k^2} \right\}$$

hyperparameters

Length scale hyperparameter

Gaussian process inference

For a set of observed locations X we can compute the joint predictive distribution of new candidate locations X'

The combined covariance matrix:

Measured locations

$$K_{X \cup X', X \cup X'} = \begin{bmatrix} K_X & K_{XX'} \\ K_{X'X} & K_{X'} \end{bmatrix}$$

Candidate locations

The predictive distribution, conditioned on observations

$$\begin{aligned} \mu_{X'}|Y &= K_{X'X}(K_X + \sigma^2 I)^{-1} Y \\ K_{X'}|Y &= K_{X'} - (K_{X'X} K_X + \sigma^2 I)^{-1} K_{XX'} \end{aligned}$$

Information-optimal design

- To maximize information gain, just maximize the entropy of future observations:

$$R(X';X) = H(X'|X) = \frac{1}{2} ln \left\{ (2\pi e)^d \det(K_{X'}) \right\}$$

- Known as Maximum Entropy Sampling (MES) [Shewry & Wynn 1987]
- Greedy observation selection often works well for this class of problems
- Other information-theoretic objectives are possible

Our key innovation: Nonstationarity

The squared exponential is the same throughout the input space, so it *cannot represent* local deviations!

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \psi_1 + \psi_2 \exp\left\{ -\frac{1}{2} \sum_{k=1}^{d} \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{w_k^2} \right\}$$

We introduce a *new term z(x)* that warps the covariance function symmetrically around the current timestep to model local anomalies in progress

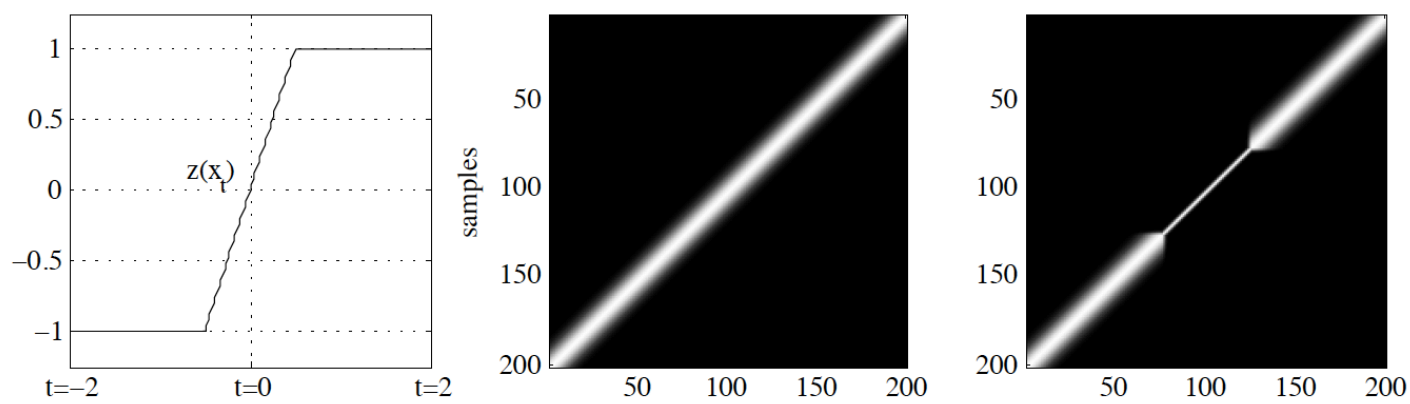
$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \psi_1 + \psi_2 \exp\left\{ -\frac{(z(\mathbf{x}_i) - z(\mathbf{x}_j))^2}{2\beta^2} - \sum_{k=1}^{d} \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{2w_k^2} \right\}$$

Forming the z term

z(x) has a constant slope within a neighborhood of the current time step:

$$z(\mathbf{x}_t) = \begin{cases} -1, & t < (t_{now} - \alpha) \\ \frac{t - t_{now}}{\alpha}, & (t_{now} - \alpha) \leq t \leq (t_{now} + \alpha) \\ 1, & (t_{now} + \alpha) < t \end{cases}$$

We modify β to adaptively "pinch" the covariance function



Large β - stationary Small β - locally uncorrelated

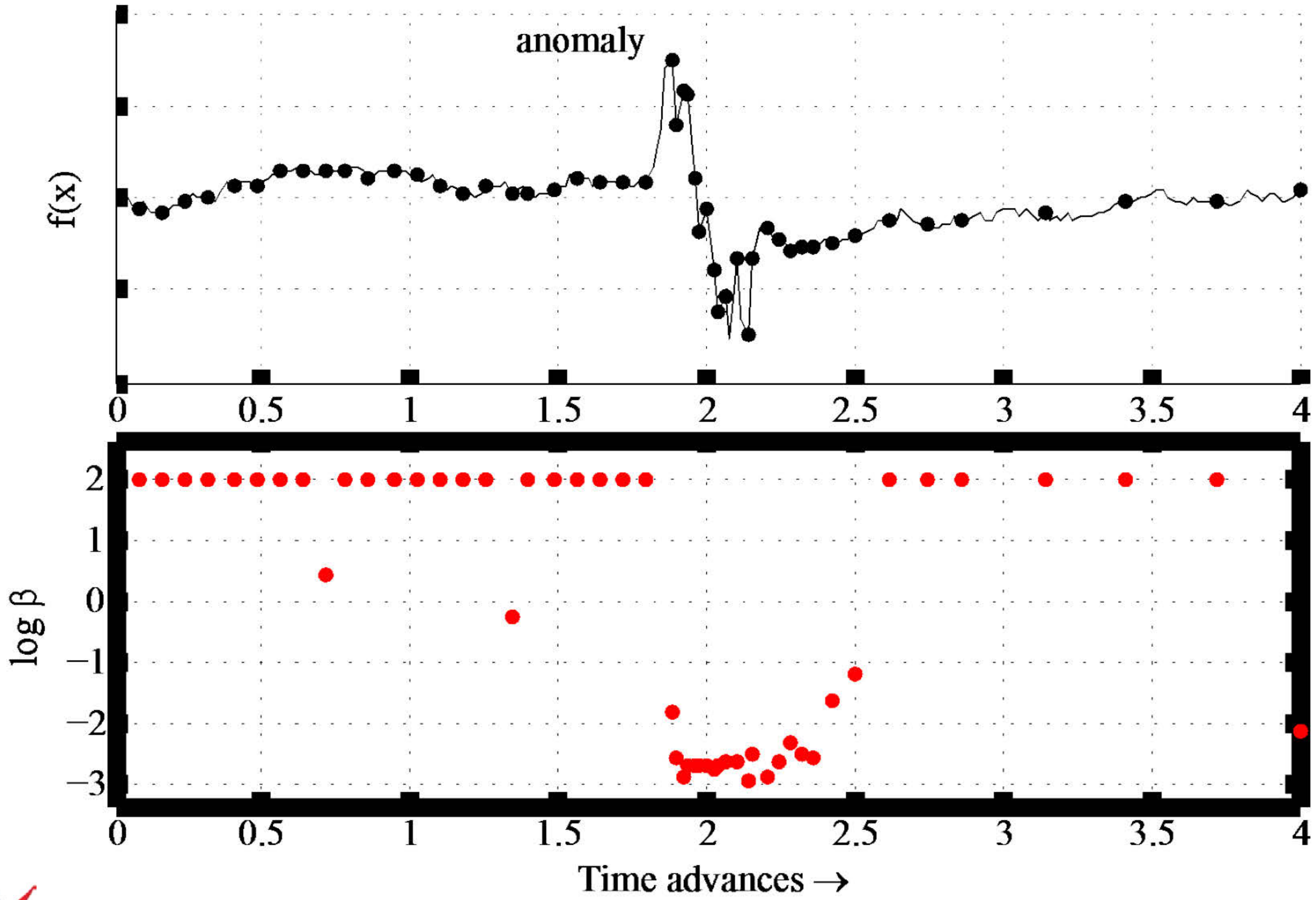
Hyperparameter learning

- Set z's neighborhood size manually, to control spatial smoothing
- All other free parameters set through maximum likelihood estimation. The log likelihood is:

$$\log p(Y|X) = -\tfrac{1}{2}\left\{ Y^T(K_X + \sigma^2 I)^{-1}Y \right. \\ \left. +|K_X + \sigma^2 I| + n\log(2\pi) \right\}$$

- **Offline** fit lengthscales by training the model on the background process
- **Online** estimate β for each new timestep
  - **A cheap one-parameter optimization** that depends only on local values!

Example: 2-state random walk

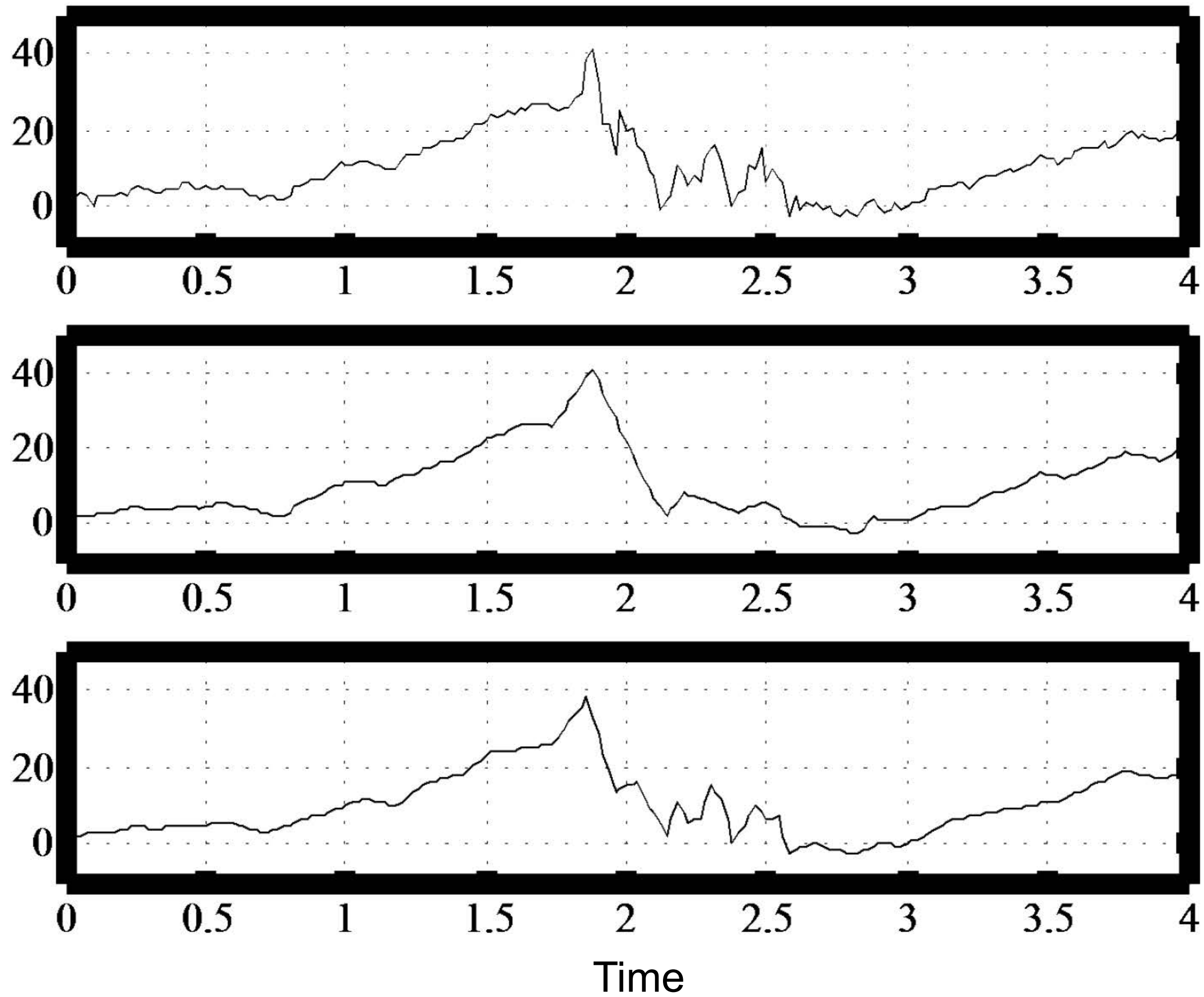


Example: 2-state random walk

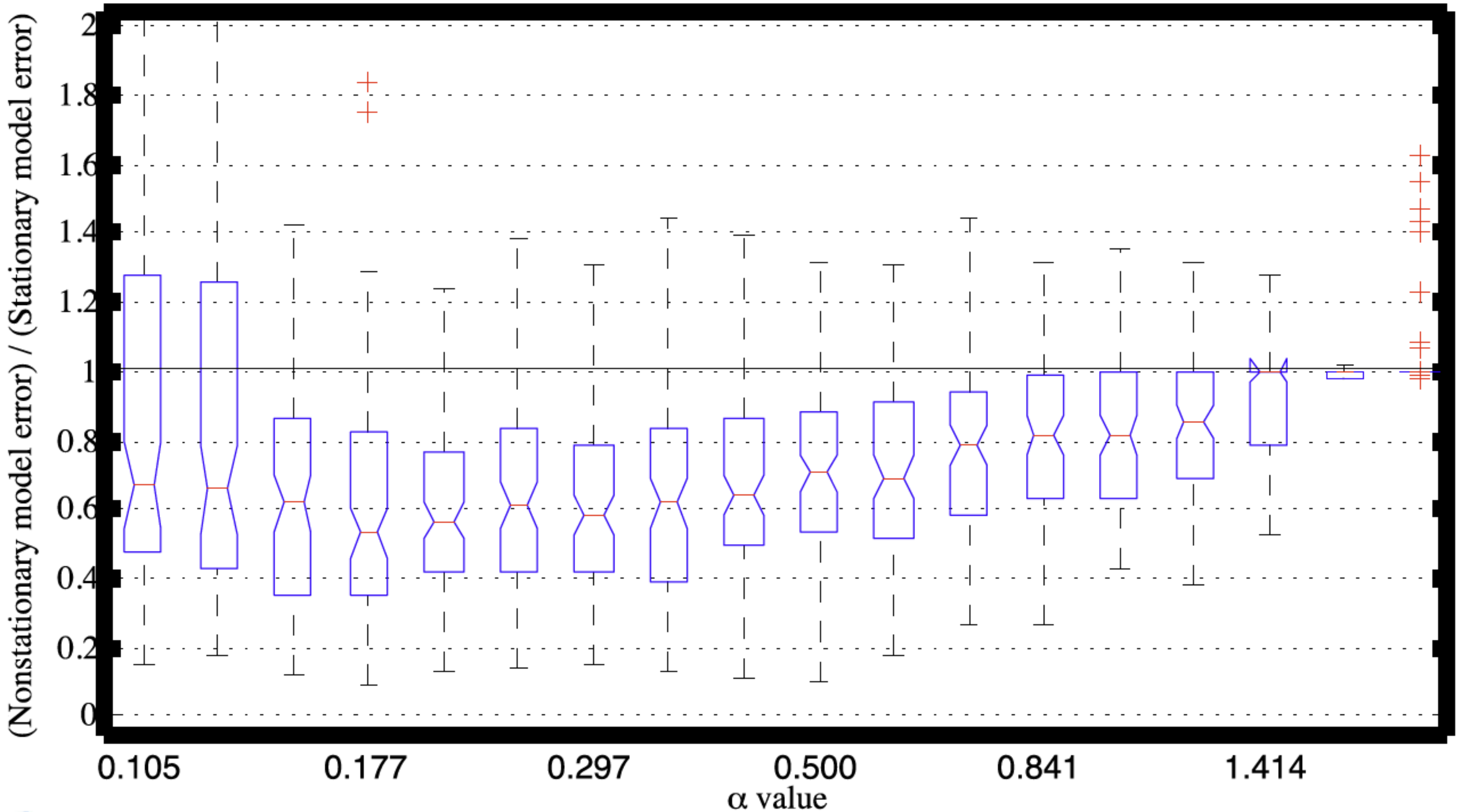
Underlying hidden process

Reconstruction from samples selected with stationary model

Reconstruction from nonstationary adaptive sampling

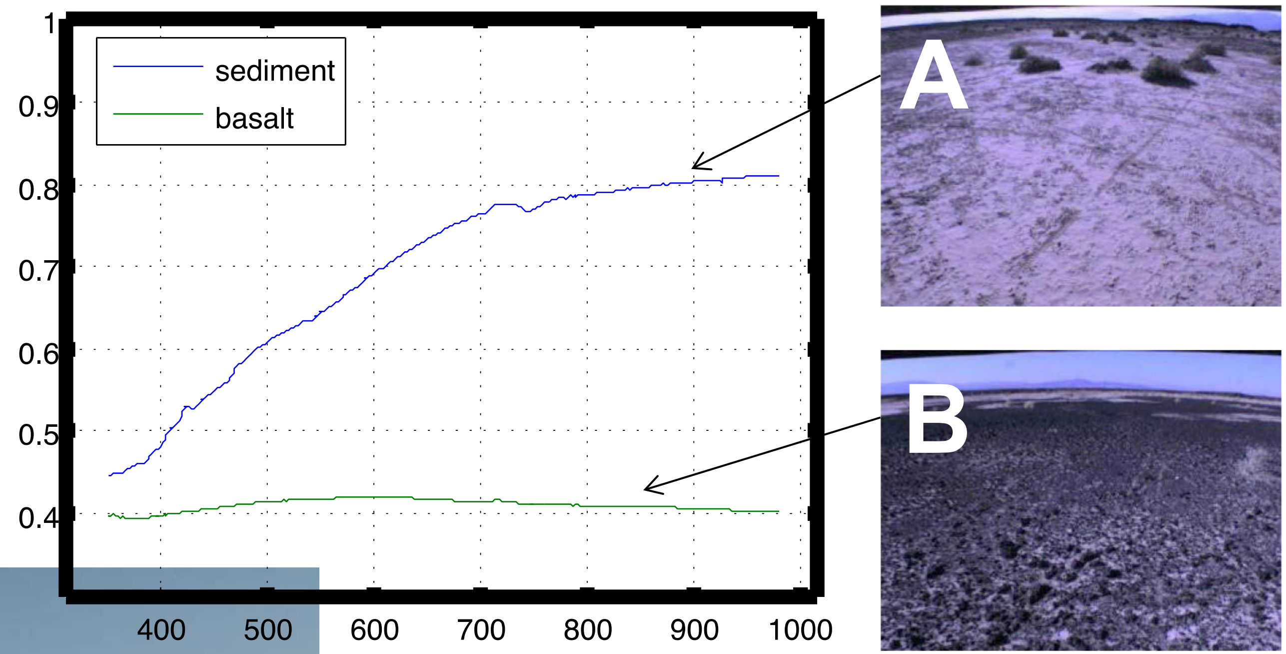


Sensitivity to neighborhood



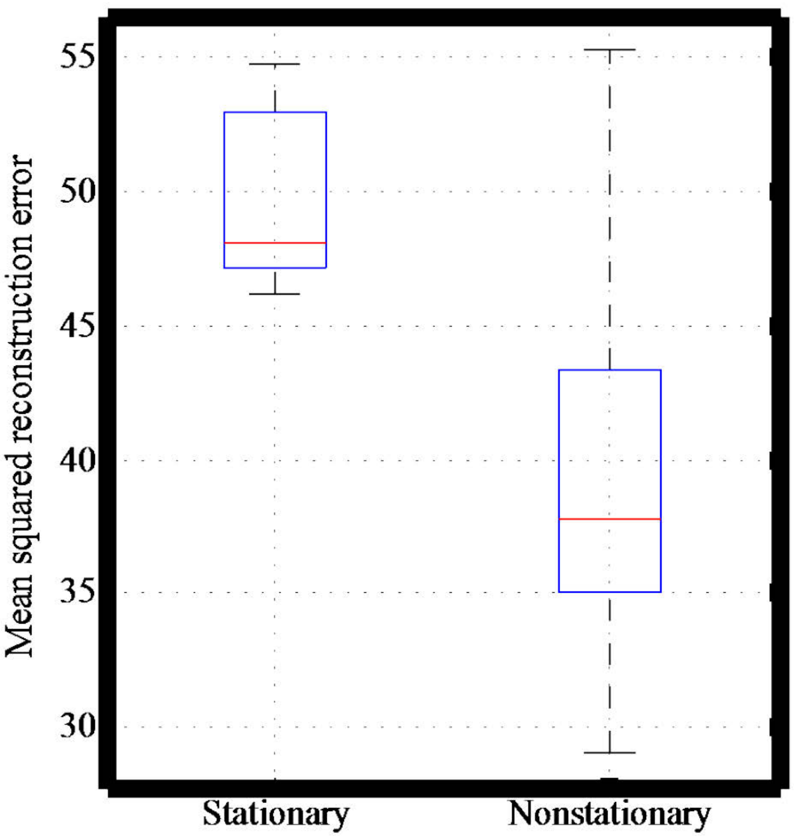
Amboy crater experiments

Amboy crater experiments

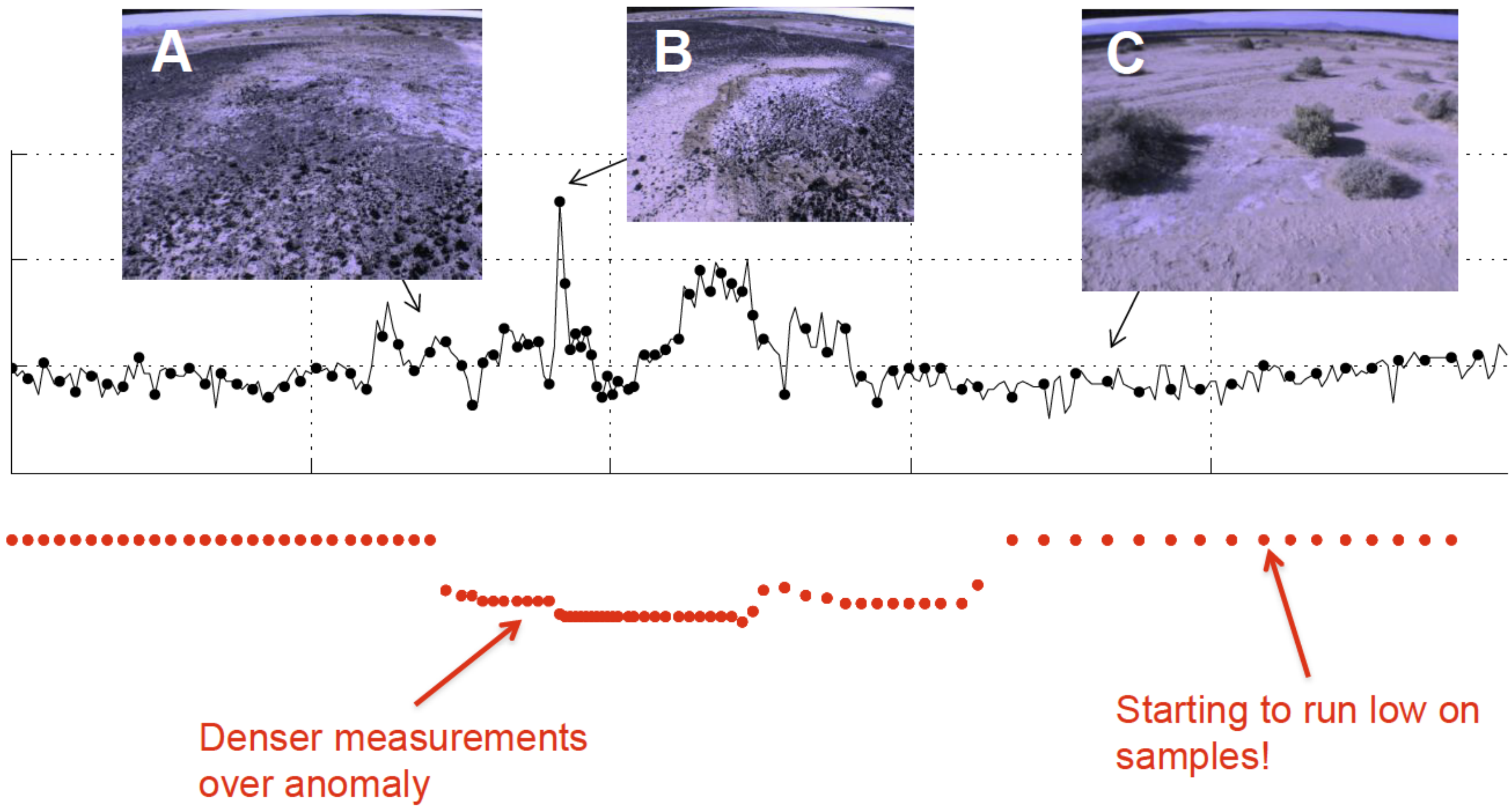


Amboy crater experiments

- Rover-mounted ASD field spectrometer measuring 0.5-2.5 microns
- Use endmember materials to create a scalar “basalt index”
- Goal: reconstruct surface composition over a rover traverse spanning 100s of meters



Amboy crater experiments



Closing thoughts

- Adaptive time series sampling is a challenging problem
  - Limited visibility
  - Limited control
  - Nonstationary structures
- Our new method: a principled, information-theoretic approach
  - Can incorporate new covariance relationships, additional independent variables
  - Models joint distribution over all future measurements, permitting globally-optimal planning
- A new covariance function represents local nonstationarity by reducing temporal correlations within near the current time step.
  - Simple and fast for ***real time*** inference and planning

Thanks!

The Amboy data was collected under NASA ASTEP Grant NNG04GB66G “Science on the Fly” and performed at Carnegie Mellon University.

More generally, this work is supported under the Life in the Atacama project by NASA Astrobiology Science and Technology for Exploring Planets (ASTEP) Grant NNX11AJ87G.

A portion of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology. U.S. government support acknowledged.