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]

***Abstract*—We address the problem of adaptive information-optimal data collection in time series. Here a remote sensor or explorer agent throttles its sampling rate in order to track anomalous events while obeying constraints on time and power. This problem is challenging because the agent has limited visibility — all collected datapoints lie in the past, but its resource allocation decisions require predicting far into the future. Our solution is to continually fit a Gaussian process model to the latest data and optimize the sampling plan on line to maximize information gain. We compare the performance characteristics of stationary and nonstationary Gaussian process models. We also describe an application based on geologic analysis during planetary rover exploration. Here adaptive sampling can improve coverage of localized anomalies and potentially benefit mission science yield of long autonomous traverses.**

I. Introduction

This work addresses the problem of adaptive sensing of time series data. We present methods that enable self-throttling sensors to adjust their data collection rate, providing enhanced coverage during transient anomalies of special interest. This capability could benefit power-constrained environmental sensor networks. For example, a remote seismic sensor could conserve resources by limiting its measurements during normal conditions and increasing its sample cadence during rare earthquake events [1]. Similar methods could enhance robotic exploration where a sensor follows a fixed trajectory, such as an exploration rover transect or a deep space flyby. These agents commonly observe long homogeneous segments punctuated by short periods of rapid change. We focus here on adaptive sensing problems that have the following characteristics:

- An agent senses a single scalar value over time, subject to constant white measurement noise.
- The underlying signal usually comes from a background process whose behavior is slowly varying and whose parameters can be characterized in advance.
- There are rare anomalous periods during which the process may change more rapidly.
- Resource constraints limit the agent to a small total number of measurements, but the agent can analyze collected data to revise its sampling plan on the fly.
- We aim to optimize modeling fidelity, e.g. information gain with respect to the underlying noiseless values.

Our fidelity objective is independent of the absolute signal magnitude; an optimal policy will sample time steps whose values are uncertain, but it will not exhibit any preference for unusually high or low signal values. This problem equates to active learning [2] in which the agent allocates future measurements to reduce uncertainty as measured by the Shannon entropy of the process at all time steps. Common measures of information gain are the entropy of the measurements themselves [3] or their mutual information with respect to unobserved time steps [4]. Submodular optimization algorithms can efficiently solve a broad class of these cost functions [5].

Adaptive sampling of time series is related to classical spatial experimental design [6], but it also has unique challenges. One is *measurement imbalance* — all previous data lie in the past, but this region cannot be observed again. Planning future measurements requires predicting the future, extrapolating beyond the range of collected data. A second challenge is that information-driven sampling requires *accurate second order statistics* because the value of each new observation is related to its prediction certainty. A third challenge is *limited control*; the agent may only choose the length of time to wait before the next sample. Finally, time series sampling domains often require extreme *computational efficiency*; power-limited sensor motes must react in time to capture transient events.

There has been some limited previous work in this area. Researchers have investigated information-optimal adaptive time series sampling for Markov chains and graphs [7], [8]. These tests typically use batch optimizations rather than an on-line update, and are limited to the family of graphical models. An ARIMA-based method by Law et al. does adapt to local changes in real time [1], but this approach relies on heuristic rules and does not enforce a global budget on the number of samples. We seek a new approach to permit a true information-theoretic treatment, with on-line updates, over a broader class of probabilistic models.

Specifically, we perform adaptive online sampling of time series using Gaussian process (GP) time-series regression models [9]. These models are favorable since they define a full predictive posterior distribution and compute observation differential entropies in closed form. Most standard GP formulations are not appropriate for adaptive sampling because they are *stationary*, i.e. the covariance function is time-invariant. In such cases information gain predictions are completely independent of the measurement values [10], resulting in evenly-spaced samples. Consequently stationary models are not appropriate for localized anomalies. Additionally, GP model inference can be computationally expensive. Typically parameter fitting requires unbounded iterative procedures or Markov Chain Monte Carlo sampling, precluding fast reactions by power-limited sensor motes.

This work presents a simple but effective GP formulation to address the requirements of adaptive time series sampling. We introduce nonstationary behavior by fitting a local modulation to the background GP hyperparameters. It is tantamount to a constrained local warping of the input space [11] that is always centered on the current time step. We identify deviations from the background process by regularly recomputing the degree of local warping on the fly during data collection. The following section reviews classical Gaussian process prediction and sample selection. Section III describes our approach and explores its performance in simulation. Section IV describes potential applications for adaptive rover sampling during long traverses. We report results of an autonomous rover traverse incorporating autonomous data collection with a Visible Near Infrared (VNIR) spectrometer.

II. Background

A remote agent observes a time series at times $t \in T$. Each time is associated with a vector $\mathbf{x}_t \in \mathbb{R}^d$ of known independent variables. This could be the clock time, creating a 1D input space, or it could have more dimensions containing contextual information such as physical location. The agent samples the environment to yield scalar measurements $y_t \in \mathbb{R}$. We will write the set of independent variables as $X = \{\mathbf{x}_t\}_{t \in T}$ and the set of measurements as $Y = \{y_t\}_{t \in T}$.

The agent must select measurement times to balance the expected information gain of these samples and cost expended. At any given time the agent's only real decision is the time of the next observation. However this requires considering the entire budget for all samples and the likelihood of encountering anomalies later. Without loss of generality, we will formulate the adaptive sampling problem as extrapolating from the set of all previous measurements at times T to select future measurement times T', where the total number of measurements is bounded by a budget B, e.g. $|T \cup T'| \leq B$. The agent plans its future samples, waits until the next planned sample time, acquires a new measurement, and then revises its sampling plan before starting the process again. The following sections detail our time series modeling strategy, measurement selection, and nonstationary models capable of tracking anomalies.

A. Gaussian Process Prediction

We treat the environmental process under study as a function $f(\mathbf{x}) : \mathbb{R}^d \mapsto \mathbb{R}$. Each sample is a measurement of this process perturbed by Gaussian noise:

$$y = f(\mathbf{x}) + \mathcal{N}(0, \sigma^2) \tag{1}$$

We model the environment with a Gaussian process (GP), a prior over functions $f(\mathbf{x})$ such that the value of any set of samples is multivariate Gaussian-distributed [9]. The prior takes the form of a covariance function $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ parameterized by hyperparameters θ. This represents similarity between input locations, inducing a GP covariance matrix K with elements given by the covariance function evaluated between each pair of points. We will use the shorthand K_X to represent the covariance matrix for previous measurements X; it contains $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ in the ith row and the jth column. The prior distribution is $P(f(X)) = \mathcal{N}(0, K_X)$. Additive measurement noise contributes to the diagonal of the covariance matrix. With some abuse of notation, we will use Y to represent the column vector of measured values:

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

A popular form for the covariance function is the squared exponential [9], parameterized by the hyperparameters $\theta = \{\psi_1, \psi_2, w_1 \ldots w_d\}$ consisting of positive coefficients ψ_1 and ψ_2, as well as a length scale w_k for each input dimension:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \psi_1 + \psi_2 \exp\left\{ -\frac{1}{2} \sum_{k=1}^{d} \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{w_k^2} \right\} \tag{3}$$

Typically the agent learns hyperparameters from data using evidence maximization (maximum likelihood fitting) or Markov Chain Monte Carlo (MCMC) techniques. We refer the reader to [9] for further details.

We define the set of candidate observations at future times T', with independent variables $X' = \{\mathbf{x}'_t\}_{t \in T'}$. The matrix $K_{X'}$ contains pairwise covariance function values for these candidate observations. The matrices $K_{XX'}$ and $K_{X'X}$ represent covariance between old and new measurement locations. For example, to the element in the ith row and jth column of $K_{XX'}$ is the covariance function evaluated with the ith old location and the jth future location. This forms the combined matrix:

$$K_{X \cup X', X \cup X'} \;=\; \begin{bmatrix} K_X & K_{XX'} \\ K_{X'X} & K_{X'} \end{bmatrix} \tag{4}$$

The Gaussian process prediction equations [9] estimate future observations. We simply condition the full joint Gaussian on previous measured values in Y:

$$\mu_{X'}|Y \;=\; K_{X'X}(K_X + \sigma^2 I)^{-1} Y \tag{5}$$

$$K_{X'}|Y \;=\; K_{X'} - (K_{X'X} K_X + \sigma^2 I)^{-1} K_{XX'} \tag{6}$$

New collected data can augment this onboard model over time to improve certainty of future predictions.

B. Sample Selection

In our formulation the agent selects a fixed-size subset T' to maximize information gain, optimally reducing the entropy of the function everywhere in the environment. Shewry and Wynn show that under some weak assumptions this is equivalent to maximizing the entropy of the observations themselves [3], a procedure known as Maximum Entropy Sampling (MES). MES reflects the intuitive idea that we should collect data where the result is most uncertain. In Gaussian spatial processes the differential entropy of X' is related to the determinant of the covariance matrix $K_{X'}$, leading to the objective:

$$R(X'; X) = H(X'|X) = \frac{1}{2} ln \left\{ (2\pi e)^d \det(K_{X'}) \right\} \tag{7}$$

This objective has a closed form. While the entropy objective is formally NP-hard, it belongs to a class of submodular functions [5] for which greedy sample selection often performs

quite well. We note there exist other submodular measures of information gain such as mutual information with respect to a reference set [5]. A formal treatment of the mutual information criterion is beyond our scope, suffice to say one can often optimize it using similar methods [4].

Sample selection is comparatively simple for a 1D time series. Here we select a uniform *local* sampling rate that applies within the future time horizon α. The total sample budget is fixed, so this local rate fully determines the *distal* sampling rate used beyond. The agent performs a one-parameter optimization to find the best allocation of measurements between local and distal segments. This is faster than most standard MES algorithms such as pointwise greedy selection since the number of determinant calculations scales with the constant number of budget splits considered. It improves numerical stability since the domain is fully-populated with observations for each determinant calculation. Nevertheless, greedy or non-myopic MES are both reasonable alternatives when power resources permit.

C. Nonstationary Covariance

Note that equation 7 is entirely independent of collected measurements Y. As a consequence the optimal sample placement can be computed in advance [12]. The process is *stationary*, i.e. the covariance is the same everywhere in the environment. This leads to sampling at regular time intervals and is obviously undesirable for tracking anomalies.

Researchers have proposed several alternatives that provide nonstationary behavior. Paciorek et al. define a smoothly-varying covariance function based on the Mátern form, fitting its hyperpameters using MCMC techniques [13]. Alternatively one can define local length scales with a smooth process and fit them jointly with the GP hyperparameters using gradient descent [14]. Adams et al. propose a Gaussian process Product Model defined as the pointwise product of two latent Gaussian processes, and fit this using an Expectation Maximization approach [15]. In general, these latent scale approaches may involve more challenging optimization than is possible for limited computation power available to the remote agent. Additionally, it can be difficult to guarantee their stability or scale them to large datasets. To date these models have been applied to a very limited range of real world on line scenarios. Others have induced nonstationary behavior by partitioning the input space into many local Gaussian processes [16]. This is computationally effective, but one still needs a global covariance matrix to plan with the entire sample budget.

Alternatively, one can abandon the pure Gaussian process formulation; Low et al. demonstrate the use of a log-Gausssian process for mapping environments containing localized hotspots [10]. The entropy of log-GPs is similar to expression 7 but with an additional term based on the predictive mean, leading to the objective:

$$R_{\mathrm{LGP}}(X';X) = \frac{1}{2} ln \left\{ (2\pi e)^d \det(K_{X'}) \right\} + \mu_{X'} \quad (8)$$

This objective is dependent on observations through the $\mu_{X'}$ term. It favors adaptive sampling behavior with a high density of samples near hot-spot locations with high predicted mean. Following our objective of reconstruction fidelity (Section I), we use a nonstationarity form where the data may not be log-GP distributed, and any period of rapid change might be interesting regardless of the absolute signal magnitude.

III. Method

Here we propose a simple method for introducing nonstationarity to permit adaptive sampling in GP time series. It is related to previous work that warps the input space with a monotonic function, as in [11], or augments Gaussian processes with additional input dimensions, as in [17]–[19]. These modifications create covariance relationships that are nonstationary with respect to original time index. Here, we warp the time series symmetrically about the current time step to model a local anomaly. We then fit the degree of warping at run time using maximum likelihood estimation. This simplification is effective for time series applications because nonstationarity in the distant past and future has little practical effect on the adaptive sampling. In a 1D time series, the only real decision is the amount of time to wait until the next measurement. Our proposed model always returns to the baseline stationary process in the far future.

The local warping takes the form of a piecewise linear function $z(\mathbf{x}_t)$ based on the difference between the time index t and the current time step t_{now}. It has a constant slope within a local neighborhood α of the current time step:

$$z(\mathbf{x}_t) = \begin{cases} -1, & t < (t_{now} - \alpha) \\ \frac{t - t_{now}}{\alpha}, & (t_{now} - \alpha) \leq t \leq (t_{now} + \alpha) \\ 1, & (t_{now} + \alpha) < t \end{cases} \quad (9)$$

We define a new covariance function based on Equation 3, with a new length scale hyperparameter w_z determining the local warping's influence.

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \psi_1 + \psi_2 \exp\left\{ -\frac{(z(\mathbf{x}_i) - z(\mathbf{x}_j))^2}{2\beta^2} - \sum_{k=1}^{d} \frac{(\mathbf{x}_{ik} - \mathbf{x}_{jk})^2}{2w_k^2} \right\} \quad (10)$$

Large β values reduce the covariance function to the standard stationary squared exponential form. Small values reduce covariance in the neighborhood of t_{now}, increasing its information value which favors dense local sampling. Figure 1 shows a visual example. Small values of β pinch the covariance structure and reduce the local temporal correlation. The constant α is fixed in advance, and controls the size of the temporal neighborhood. It is tantamount to a smoothing factor determining the rate at which the process alternates between anomalous and baseline segments. Large temporal neighborhoods cause the agent to change its behavior more slowly; it will be less susceptible to noise, but also slower to recognize new anomalies and to return to its baseline sampling rate after the anomaly has ended.

We fit hyperparameters by maximizing the marginal data log likelihood:

$$\log p(Y|X) = -\tfrac{1}{2} \{ \, Y^T (K_X + \sigma^2 I)^{-1} Y + |K_X + \sigma^2 I| + n \log(2\pi) \, \} \quad (11)$$

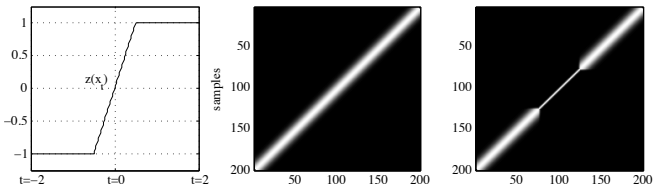

Fig. 1. Local temporal warping Left: $z(\mathbf{x})$ is a piecewise linear function that is symmetric about the current time step. Center: We fit the length scale associated with $z(\mathbf{x})$ to define a new covariance matrix. A large β corresponds to a near-stationary covariance structure. Bright pixels represent large-valued elements. This covariance matrix shows a sequence with 200 hypothetical sampling locations, evaluated when the agent reaches the midway point (sample 100). Right: a small value of β reduces temporal correlation, as would be the case for a local anomaly.

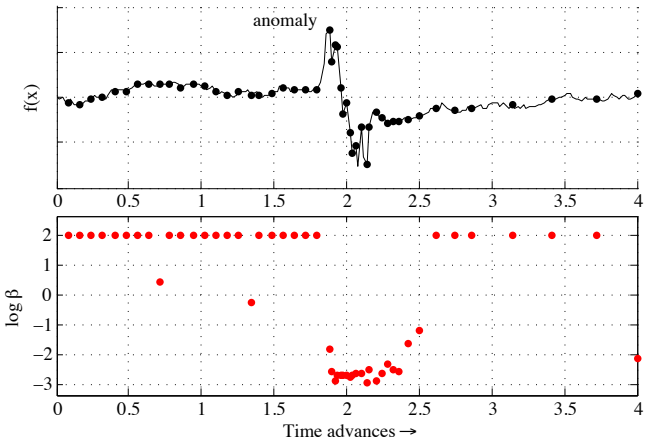

Fig. 2. Adaptive time series sampling with local temporal warping. Top: an agent sequentially samples a random walk, the rate of which increases dramatically in the segment labeled "anomaly." Bottom: The agent performs maximum likelihood estimation of the local warping parameter β at each time step. It increases the local warping near $t = 2$, which in turn increases sample density to cover the anomaly feature.

All hyperparameters except β are set ahead of time by training on the baseline process. This corresponds to a strong point mass prior on hyperparameters. At runtime Equation 11 then reduces to a one-parameter optimization of β. If the observations do not depart significantly from the baseline model, β will be large and the Gaussian process is stationary. We perform this optimization once per time step by bracket search, though gradient descent could also be used.

Figure 2 shows an example. Here an agent measures a univariate random walk process. We use a held out data set to fit all hyperparameters except β. The sequence begins with a small set of 10 evenly-spaced measurements to seed the GP prediction (Figure 2). The agent then enters an adaptive loop in which it first evaluates new values $\{z(\mathbf{x}_t)\}_{t\in T}$, then fits the degree of temporal warping β using Equation 11, then computes a new covariance matrix, and finally plans the observation spacing for the remaining budget using the Maximum Entropy Sampling procedure described above. The process of data collection, model fitting, and planning repeats for each new simulation time step (Algorithm 1).

Input: Initial samples $X = \{x_t\}_{t\in T}$,
Measurements $Y = \{y_t\}_{t\in T}$,
Temporal neighborhood α,
Hyperparameters $\theta = \{\phi_1, \phi_2, w_1 \ldots w_n\}$,
Total sample budget B

while $t_{now} < t_{end}$ and $|X| < B$ **do**
  compute $Z = \{z(\mathbf{x}_t)\}_{t\in T}$ via Eqn. 9;
  find β to maximize $\log p(Y|X)$ via Eqns. 11-10;
  use β to compute covariance $K_X|Y$ via Eqn. 6;
  plan new samples X' with greedy selection,
    to maximize $\det(\mathrm{K}_{X'}|\mathrm{Y})$,
    subject to $|X \cup X'| \leq B$;
  wait until next sample time step t_{next};
  collect datum, $X = X \cup \mathbf{x}_{t_{\mathrm{next}}}$, $Y = Y \cup y_{t_{\mathrm{next}}}$;
  $t_{now} \leftarrow t_{next}$
end

Algorithm 1: Adaptive time series sampling.

The random walk rate increases at $t = 2$, and Maximum likelihood estimation effectively detects the local anomaly. After completely passing the anomaly the agent resumes its baseline sampling rate near $t = 3$. The agent begins to increase its sample spacing around $t = 2.5$ because it has nearly exhausted its budget and must plan to cover the entire time sequence. The local warping model assumes a permanent return to the baseline state in the future, so it spends samples agressively. This is appropriate if anomalies are rare or short; otherwise, an effective *ad hoc* remedy is to plan based on a fixed fraction of the true budget remaining.

We simulated 100 random walks similar to Figure 2, hiding in each an anomalous segment that was 10% of the total sequence length. We then reconstructed each time series using the agent's sparse measurements, linearly interpolating the collected data. The mean squared difference between the reconstruction and the actual time series serves as a performance score. Figure 3 shows a typical anomaly segment reconstructed from samples collected by stationary and nonstationary models. There is a visible benefit due to the adaptive system's ability to detect, represent, and react to the anomaly. This implementation uses unoptimized code on a modern laptop processor. The entire process of refitting the GP model and replanning the sample rate requires just 0.023 seconds on average. This suggests the technique is feasible for sub-second sample rates by power-constrained platforms.

Figure 4 (Right) shows the resulting reconstruction fidelity scores for various values of α. Box plots show the squared reconstruction error for the traverse as a fraction of the error incurred by the stationary alternative. The adaptive method significantly outperforms stationary sampling for a wide range of α values. Performance is robust within a factor of 2-3 around the optimal value, and degrades gracefully outside this range. Very large values of α reduce to the stationary case. In practice, one can set α in advance using prior knowledge of feature timescales or tune it using cross-validation error on an analogue dataset.

IV. Application to Adaptive VNIR Spectroscopy

Here we illustrate adaptive acquisition of Visible Near Infrared (VNIR) spectra by a rover, during a long traverse

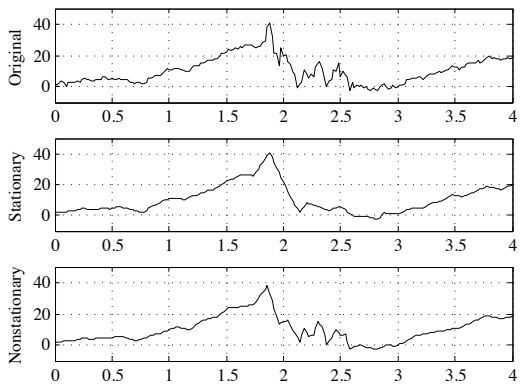

Fig. 3. Reconstruction error for a simulated test segment. The nonstationary model uses the same number of samples, but improves the reconstruction fidelity.

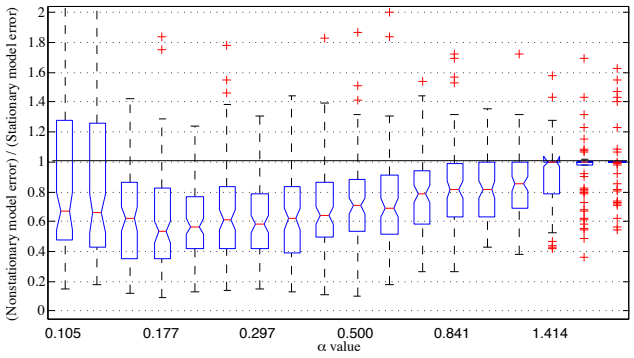

Fig. 4. Performance as a function of the smoothing parameter α on the simulated test segments. We performed 100 random trials and compared the reconstruction errors to a stationary alternative. The box-and-whisker diagram shows the extrema, middle half, and median of the errors. Single outlier datapoints appear as red "+" symbols. The central notch identifies a 90% confidence interval on the median. Boxes show median and quartiles of the performance distribution, with whiskers indicating extrema and red "+" symbols the extreme outliers.

that took place during field experiments at Amboy Crater [17]. The rover carried a mast-mounted VNIR spectrometer capable of measuring the 400-2000 nm range. The Amboy terrain contains regions of basalt and sediment having different spectral signatures (Figure 5). The rover travelled under autonomous navigation with obstacle avoidance, and meanwhile acquired densely spaced (2.5m interval) spectral measurements from the terrain directly in front of the rover.

The traverse area was mostly homogeneous clay sediment, but two distinctive basalt platforms showed much greater spatial and spectral variation. These spectra can be challenging to interpret. Typical analyses use diagnostic absorption features, matching against spectral libraries, or with *linear unmixing* based on endmember spectra. Here a much simpler strategy is sufficient. We use the ratio of 600nm and 410nm bands as a single index describing the abundance of the two material types. This yields a univariate process sampled at 278 time steps. We then test adaptive sampling by simulating traverses that subsample these measurements.

The terrain can vary on short spatial scales of less than 5

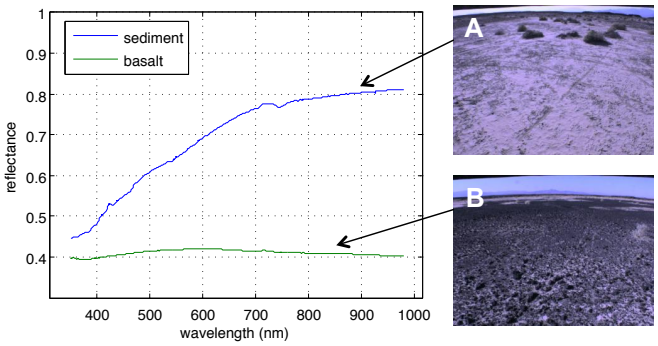

Fig. 5. VNIR spectra acquired autonomously by a mast-mounted rover instrument. Two key endmembers are shown, illustrating a significant difference in slope across the 400-900nm range. Insets show navigation camera context images associated with these spectra. (A) Context image taken during acquisition of the sediment endmember spectrum. The spectrometer field of view lies in the lower center of the image. (B) Context image for the basalt endmember spectrum.

meters, so we provide an allowance of 140 samples. This is half of the measurements that were actually performed, so the agent can only sample 50% of the potential sites. As before we consider two alternatives: an *even sampling* strategy corresponding to the stationary GP, and the adaptive system. We fit the noise and temporal lengthscale parameters of the GP using a short initial time segment and hold them constant thereafter. Then the agent waits until the next planned sample, adds it to the set of collected data, and then recomputes the complete plan for the remaining budget.

We performed a series of 20 trials to simulate adaptive sampling during this traverse, starting each virtual traverse at a different time step. These are not independent experimental trials, but they do mitigate the intrinsic variability of the score due to chance measurements and noise spikes. Figure 6 (Left panel) shows a typical result. Here the adaptive approach recognizes the anomaly and increases the measurement rate in the basalt platform. This improves performance with dense sampling of the anomaly that resolves the signal's positive and negative "spikes." These correspond to transitions across small concentrations of basalt and sediment. It comes at the cost of undersampling the tails of the traverse as the agent starts to run out of samples. Figure 6 (right panel) shows reconstruction error. The even sampling method underperforms adaptive sampling, with a median error over 50% larger. Even spacing undersamples the abrupt discontinuities and steep slopes in the anomalous region.

V. Discussion and Conclusions

Adaptive time series sampling is a challenging problem due the agent's limited visibility, limited control, and real-time computational requirements. This work presents a novel adaptive sampling method based on maximizing the information gain with respect to a Gaussian process model. Here a new covariance function represents local nonstationarity by reducing temporal correlations within a symmetrical neighborhood of the current time step. The approach shows improved performance on varied simulated

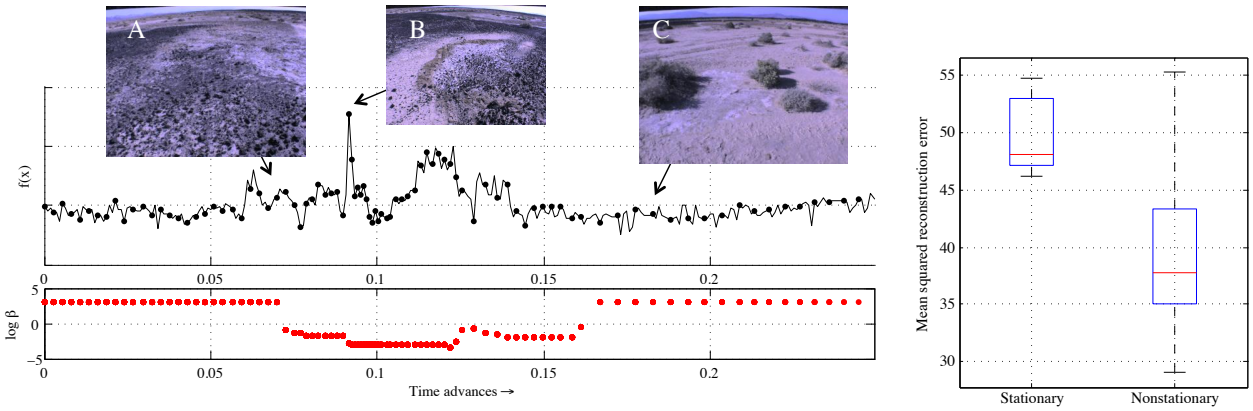

Fig. 6. Amboy traverse sequence. The rover estimates $f(x)$, a diagnostic band ratio used to discriminate basalt from sediment. Left top: basalt index and sample locations. Dots show measurement locations selected by adaptive data collection. Insets show navigation imagery from the traverse. (A) The rover enters the basalt platform area. (B) A bright, high-albedo patch of unknown material appears in the middle of the basalt platform region. (C) Homogeneous sediment region dominates the second half of the traverse. Left bottom: Optimal β values selected online (log scale). The adaptive approach detects an anomaly in the center of the traverse. Right: Reconstruction error for 10 initializations, showing the extrema, middle half, and median of the data.

and real datasets with varying noise levels and different forms for the anomalous signals. GPs offer a principled solution for adaptive sampling; they can incorporate additional inputs or alternative covariance functions, fitting these relationships directly from training data. Moreover, they can model the joint probability distribution of all future measurements to support information-optimal sampling. An information theoretic perspective decouples the challenges of modeling, (e.g. creating the covariance matrix) from the challenge of planning an experimental design with limited resources. These attributes make the nonstationary GP model a compelling solution for adaptive sensing of time series.

Acknowledgement

A portion of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology. 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, with program executive Mary Voytek.

References

[1] Y. Law, S. Chatterjea, J. Jin, T. Hanselmann, and M. Palaniswami, "Energy-efficient data acquisition by adaptive sampling for wireless sensor networks," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*. ACM, 2009, pp. 1146–1151.

[2] Cohn, D.A. and Ghahramani, Z. and and Jordan, M.I., "Active Learning with Statistical Models," *JAIR*, vol. 4, pp. 129–145, 1996.

[3] M. C. Shewry and H. P. Wynn, "Maximum entropy sampling," *Journal of Applied Statistics*, vol. 14, no. 2, p. 165170, 1987.

[4] C. Guestrin, A. Krause, and A. Singh, "Near-optimal sensor placements in gaussian processes," in *ICML*, 2005, pp. 265–272.

[5] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *JMLR*, vol. 9, pp. 235–284, 2008.

[6] N. A. Cressie, *Statistics for Spatial Data*. Wiley NY, 1991.

[7] A. Krause and C. Guestrin, "Optimal nonmyopic value of information in graphical models-efficient algorithms and theoretical limits," in *IJCAI*, vol. 19, 2005, p. 1339.

[8] D. Thompson, T. Smith, and D. Wettergreen, "Information-optimal selective data return for autonomous rover traverse science and survey," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 968–973.

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[10] P. K. Kian Hsiang Low, John M. Dolan, "Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing," *Proceedings of the International Conference on Automated Planning And Scheduling*, 2009.

[11] E. Snelson, C. Rasmussen, and Z. Ghahramani, "Warped gaussian processes," *NIPS*, vol. 16, pp. 337–344, 2004.

[12] B. K. H. Low and S. Chien and J. Dolan and J. Chen and D. R. Thompson, "Decentralized Active Robotic Exploration and Mapping for Probabilistic Field Classification in Environmental Sensing," *AAMAS*, 2012.

[13] C. Paciorek and M. Schervish, "Nonstationary covariance functions for gaussian process regression," *Advances in Neural Information Processing Systems*, vol. 16, pp. 273–280, 2004.

[14] C. Plagemann, K. Kersting, and W. Burgard, "Nonstationary gaussian process regression using point estimates of local smoothness," *Machine Learning and Knowledge Discovery in Databases*, pp. 204–219, 2008.

[15] R. Adams and O. Stegle, "Gaussian process product models for nonparametric nonstationarity," in *Proc. 25th International Conference on Machine Learning*. ACM, 2008, pp. 1–8.

[16] R. Urtasun and T. Darrell, "Sparse probabilistic regression for activity-independent human pose inference," in *CVPR*, 2008, pp. 1–8.

[17] D. R. Thompson and F. Calderón P. and D. Wettergreen, "Autonomous science for large-scale robotic survey," *Journal of Field Robotics*, vol. 28, no. 4, 2011.

[18] G. Hollinger, B. Englot, F. Hover, U. Mitra, and G. Sukhatme, "Uncertainty-driven view planning for underwater surface inspection," *Intl. Conference on Robotics and Automation*, 2012.

[19] T. Pfingsten, M. Kuss, and C. Rasmussen, "Nonstationary gaussian process regression using a latent extension of the input space," *URL http://www. kyb. mpg. de/~ tpfingst*, 2006.