Automated Experiment in SPM: Bayesian Optimization for efficient searching of parameter space to maximize functional response

Rama Vasudevan[1], Kyle Kelley[1], Jacob Hinkle[2], Hiroshi Funakubo[3], Sergei Kalinin[4], Stephen Jesse[5] and Maxim Ziatdinov[6]

[1]Oak Ridge National Laboratory, United States, [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States, [3]Tokyo Institute of Technology, United States, [4]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States, [5]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, United States, [6]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

One of the primary goals of microscopy is to correlate the functional property of materials with microstructural features of the sample. Measuring the functionality on highly heterogeneous samples in microscopy typically requires the use of spectroscopy, where parameters are varied whilst the probe is fixed in space, to measure the functional response at that location. When built up over a grid of points, the resulting hyperspectral datasets contain relevant information on the spatial variability of functional response of the sample [1].

There are primarily two downsides to this approach. The first is that spectroscopy is typically time-expensive, as grid-based spectroscopy scales as O(N2). This makes dense grid-based spectra time consuming and can limit the applicability for unstable, beam-sensitive or bias-sensitive samples, or for situations where drift is very large (e.g., at higher temperatures). The second downside is that much of the data captured is highly redundant: adjacent spatial locations are likely to produce reasonably similar functional responses. As such, methods to improve the efficiency of the sampling mechanism are attractive.

Here, we explore the use of Gaussian process (GP) regression combined with Bayesian optimization to improve the efficiency of spectroscopy and implement it on a commercial atomic force microscope (AFM) to produce an automated experimental system, for the task of finding regions of high electromechanical response in a model ferroelectric thin film. The sample is inserted into the microscope, and a random subset of pixels is chosen for voltage spectroscopy, where each pixel is associated with a hysteresis loop acquired at that location. The loop area is calculated for each of the measured pixels, and gaussian process regression is used to interpolate the unobserved pixels, providing a method of not only prediction but also uncertainty. This is used within a Bayesian optimization framework that utilizes an acquisition function (in this case, expected improvement [2]) to choose the best locations to sample during the next iteration. This process is repeated until the user terminates the process. It should be noted that the GP regression and Bayesian optimization were run on a dedicated GPU server in direct connection with the AFM, given the large processing requirements of GP based algorithms. Results for this process are shown in Fig. 1 for (a,b) simulated example and Fig. 1 (c,d) real implementation on the microscope. These results indicate that this approach can rapidly find areas of maximal loop area, although can suffer from limited exploration. We then extended this method to incorporate prior knowledge in the form of high-resolution images before the spectroscopy to better guide the optimization process. Although implemented here for scanning probe microscopy, this method is general and can be employed within electron microscopy and virtually any other scanning modality.
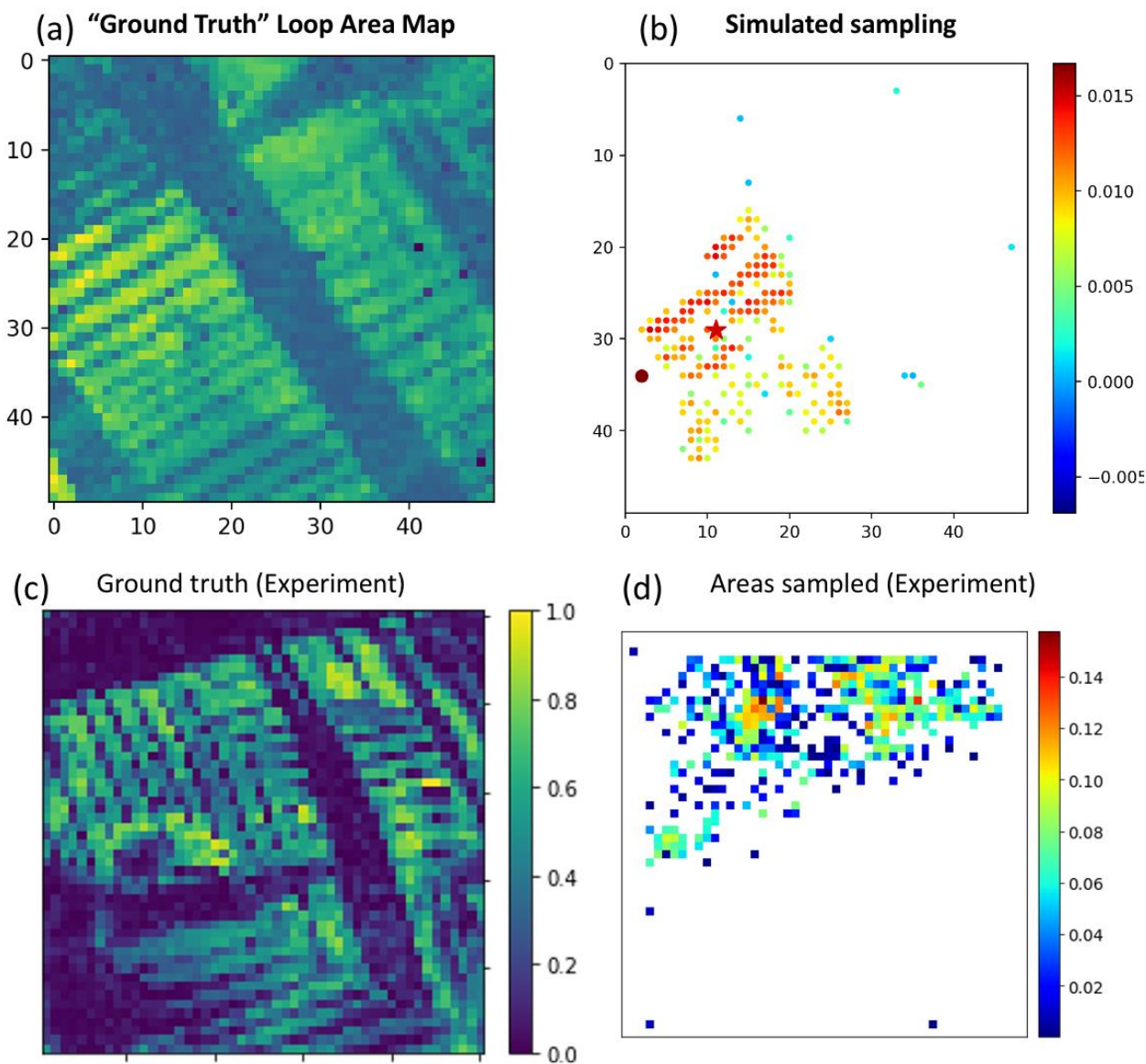
Figure 1. Figure 1: (a) Ground truth dataset. At each pixel, a hysteresis loop is captured, and the loop area is calculated. This results in a 2D map of loop area. High response is seen in the center-left part of the map. (b) Gaussian Processing - Bayesian optimization framework used to determine the next sampling locations. As can be seen, the sampled pixels are closely aligned with those from the high response region in (a), suggesting the process is quick to find the regions of interest. (c) ground truth from experiment. (d) Pixels sampled on the microscope, after 40 iterations of the GP-Bayesian optimization cycle, where on each iteration 10 new pixels are selected to measure spectroscopy at. The high response regions near the top of the image are clearly found by this method.

References
[1] S. Jesse et al. Ann. Rev. Phys. Chem. **65**, 519 (2014).
[2] B. Shahriari et al. Proc. IEEE **104**, 148 (2015).
[3] This research was conducted at and supported by the Center for Nanophase Materials Sciences, which is a US DOE Office of Science User Facility.