

2018 Annual Meeting of the APS Four Corners Section

Volume 63, Number 16

**Friday–Saturday, October 12–13, 2018; University of Utah,
Salt Lake City, Utah**

[Session J04: Computational Physics 3](#)

8:00 AM–9:24 AM, Saturday, October 13, 2018
CSC Room: 10/12

Chair: Remi Dingreville, Sandia National Lab

Abstract ID: BAPS.2018.4CS.J04.4

Abstract: J04.00004 : Optimal Data Collection for Machine Learning*

8:48 AM–9:00 AM

 [Abstract](#) 

Presenter:

Casie Gaza
(Brigham Young University)

Authors:

Mark Transtrum
(Brigham Young University)

Kent L Gee
(Brigham Young Univ - Provo)

Katrina Pedersen
(Brigham Young University)

Brooks Butler
(Brigham Young University)

Casie Gaza
(Brigham Young University)

Collaborations:

Mark Transtrum, Kent Gee, Katrina Pedersen, Brooks Butler

Machine learning refers to a collection of computational techniques for identifying or learning patterns in data. Although existing techniques are most effective on large data sets, there is growing interest in applying methods on smaller ones. We consider the application of machine learning to predicting ambient sound levels in the contiguous United States from GIS data. The challenge is limited availability of training data from which to construct a model--data collection in this case is both cost and time expensive. This leads us to consider two questions: First, how to best validate a machine learning model with limited training data and two, given additional data can we measurably improve the accuracy of the model. We create an ensemble of models that perform equally well as measured by leave-one-out cross validation on our initial training set. However, these models give wildly different predictions for areas in the central region of the country. By collecting additional data in cropland areas in Utah, we were able to improve the predictions of our machine learning model to other, geographically similar regions of the country.

*National Science Foundation Grant 1557998 Brigham Young University Physics and Astronomy

To cite this abstract, use the following reference: <http://meetings.aps.org/link/BAPS.2018.4CS.J04.4>