

# LA-UR-22-31580

Approved for public release; distribution is unlimited.

**Title:** Selecting experiments to better inform predictive models for the Mars rover

**Author(s):** Klein, Natalie Elizabeth

**Intended for:** Meeting with CMU STAMPS group

**Issued:** 2022-11-01



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



# Selecting experiments to better inform predictive models for the Mars rover

Natalie Klein

Los Alamos National Laboratory

11/4/22

# Los Alamos mentors



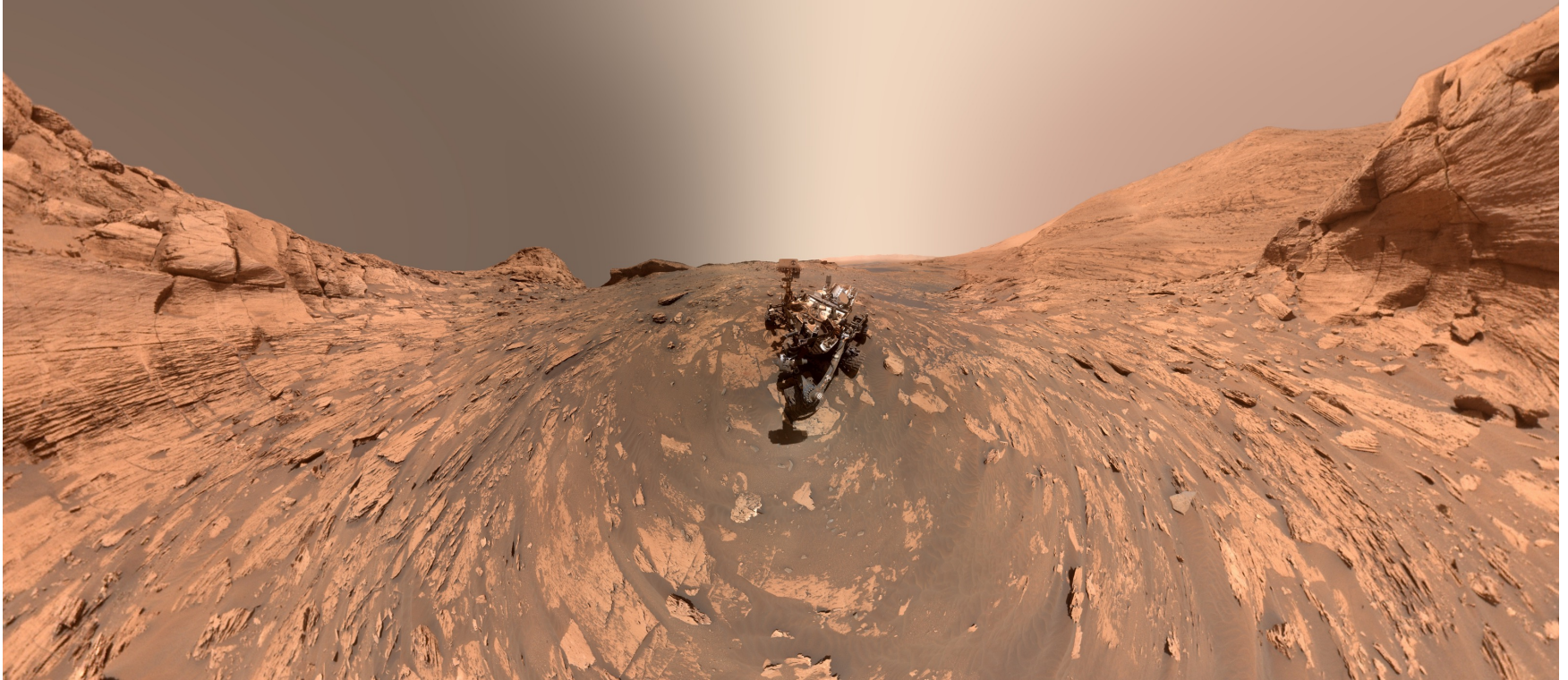
Natalie Klein  
Statistics/ML  
[neklein@lanl.gov](mailto:neklein@lanl.gov)

Patrick Gasda  
ChemCam  
[gasda@lanl.gov](mailto:gasda@lanl.gov)

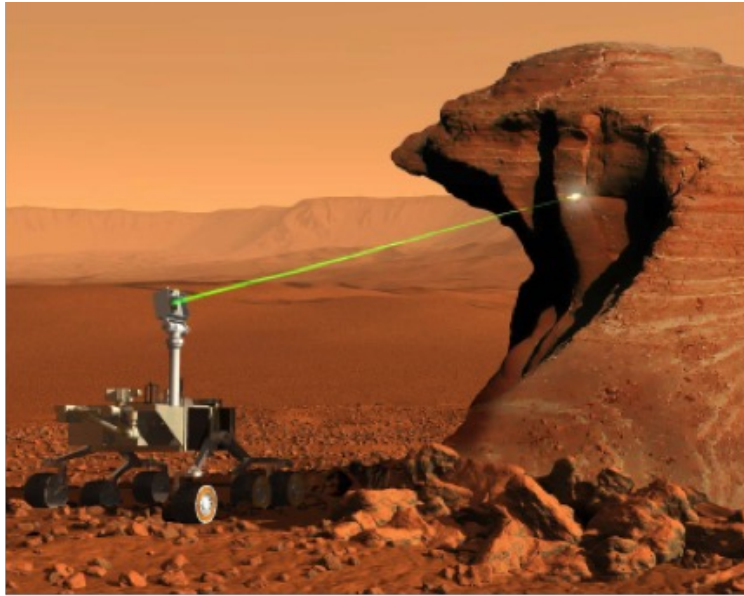
Giri Gopalan  
Statistics  
[ggopalan@lanl.gov](mailto:ggopalan@lanl.gov)



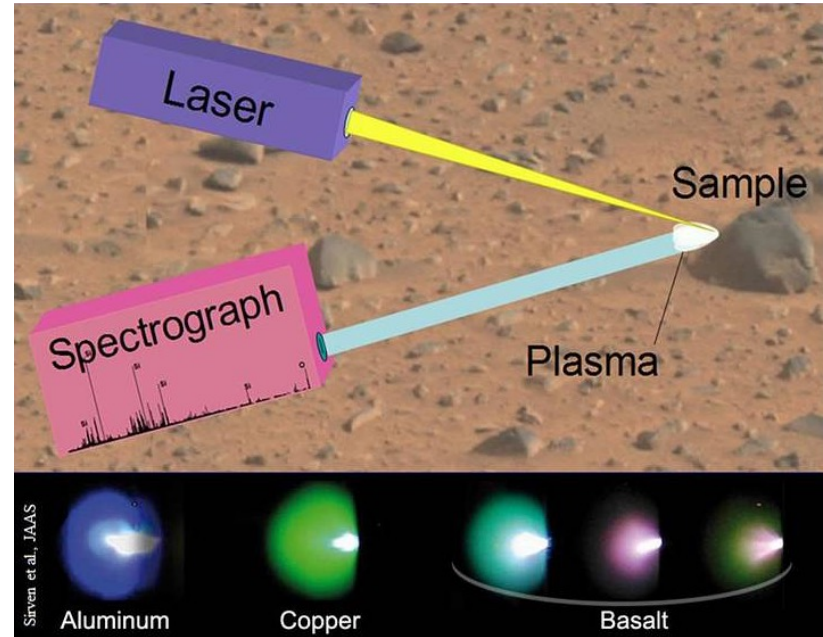
**We want to explore and characterize Mars geologically,  
but we can't go there or send back samples. (Yet.)**



# ChemCam is a laser-induced breakdown spectroscopy instrument deployed on Mars for over 3,600 sols.

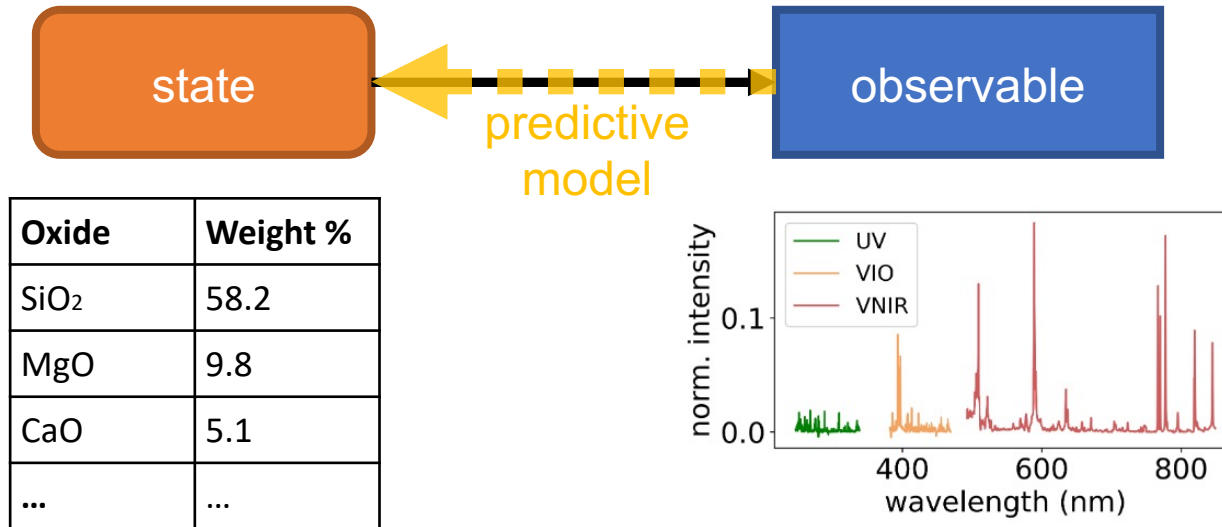


Images: NASA/JPL-Caltech/LANL; Sirven et al 2007



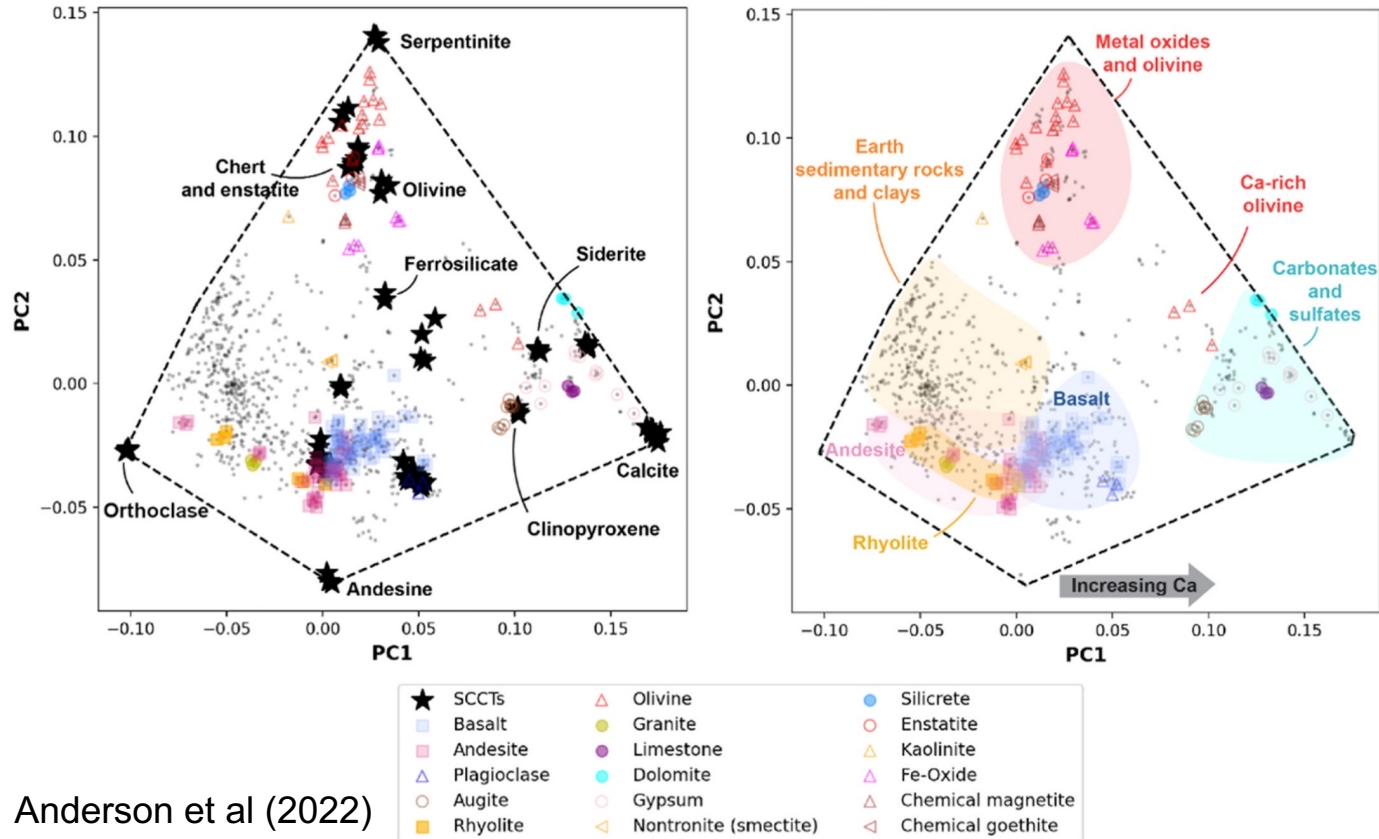
# Statistical learning models are trained to predict composition from spectral observations.

But what is the training data???





# Spectral training sets are created by measuring known targets in the laboratory on Earth.



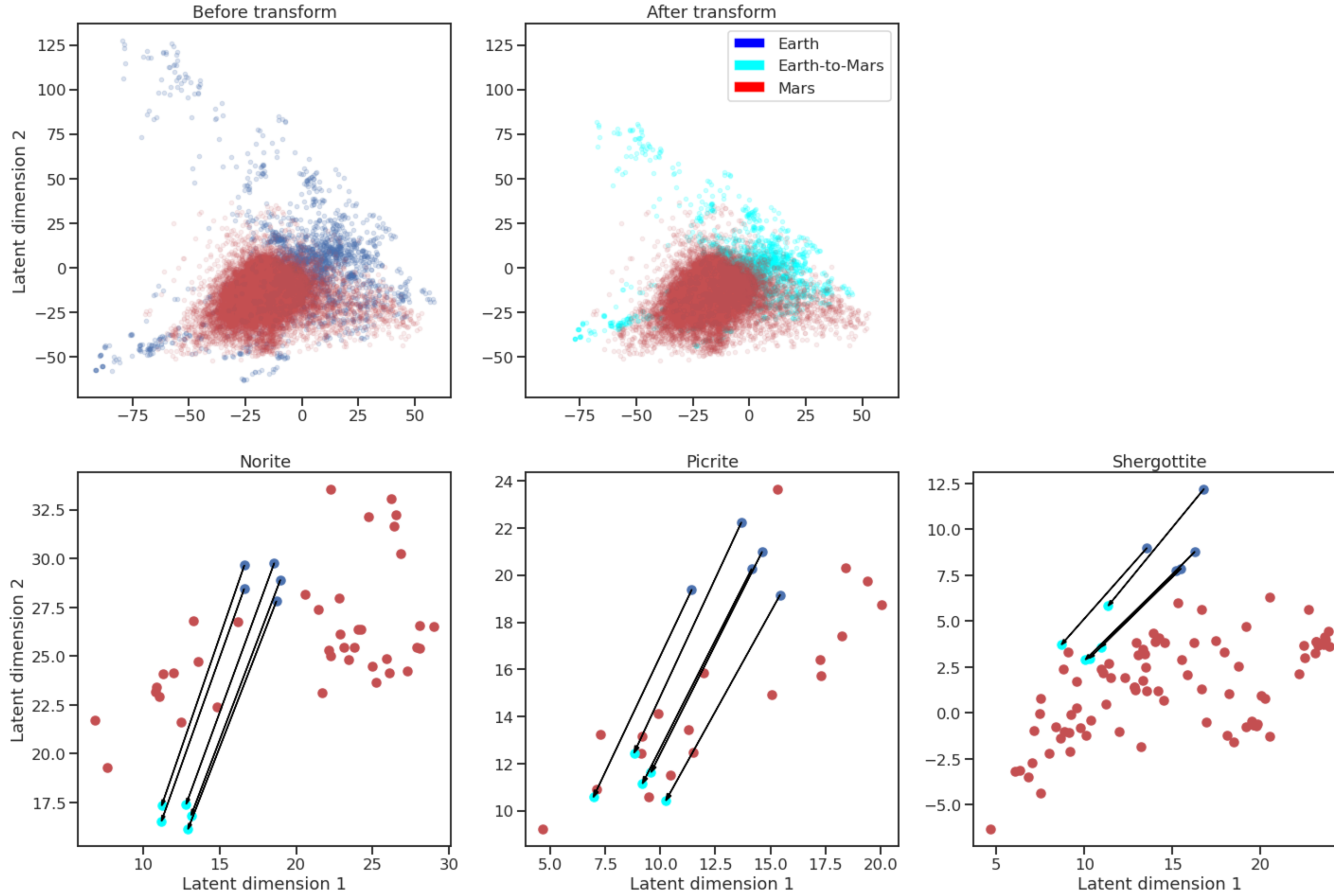
Anderson et al (2022)



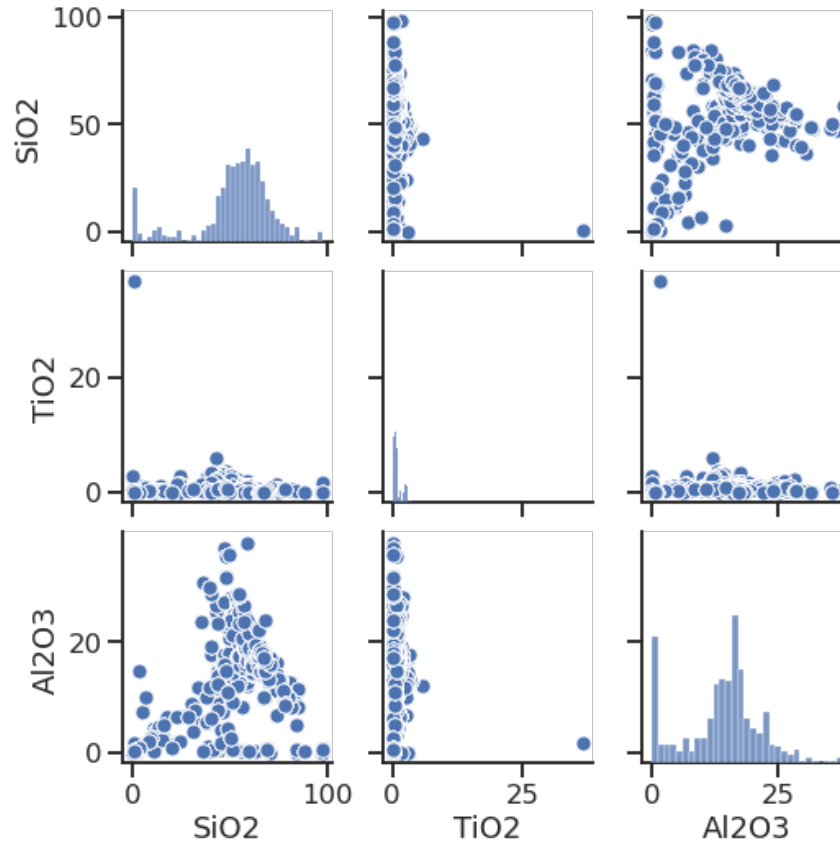
**We have very little labeled data from Mars, limited to a set of calibration targets mounted on the rover.**



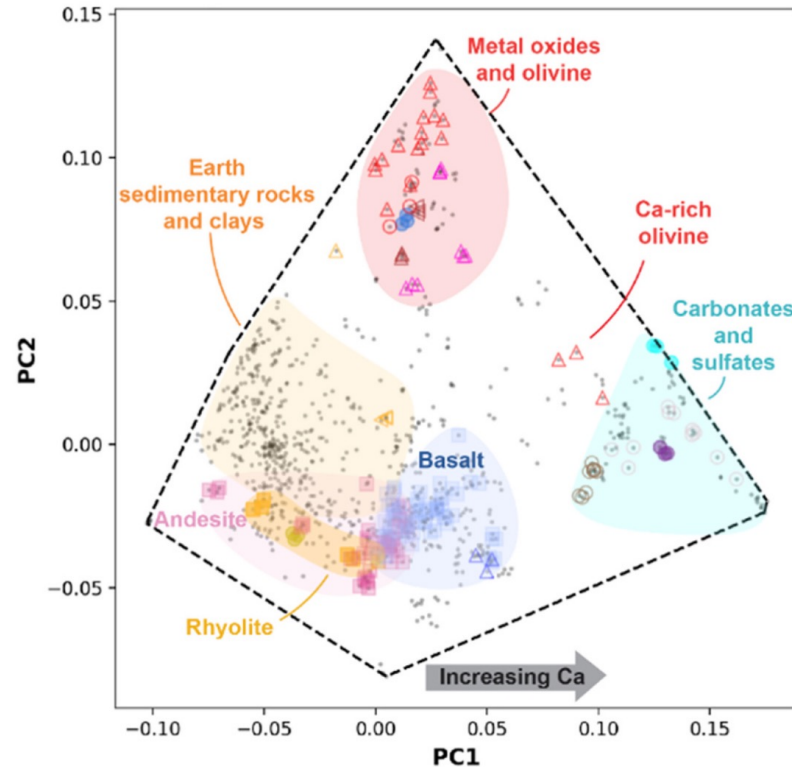
# Question 1: how well do the current Earth and Mars samples align?



## Question 2: how do empirical distributions of the training library in composition space affect the models?



# Question 3: if we could collect more training data on Earth, which samples would most impact the models?



Anderson et al (2022)





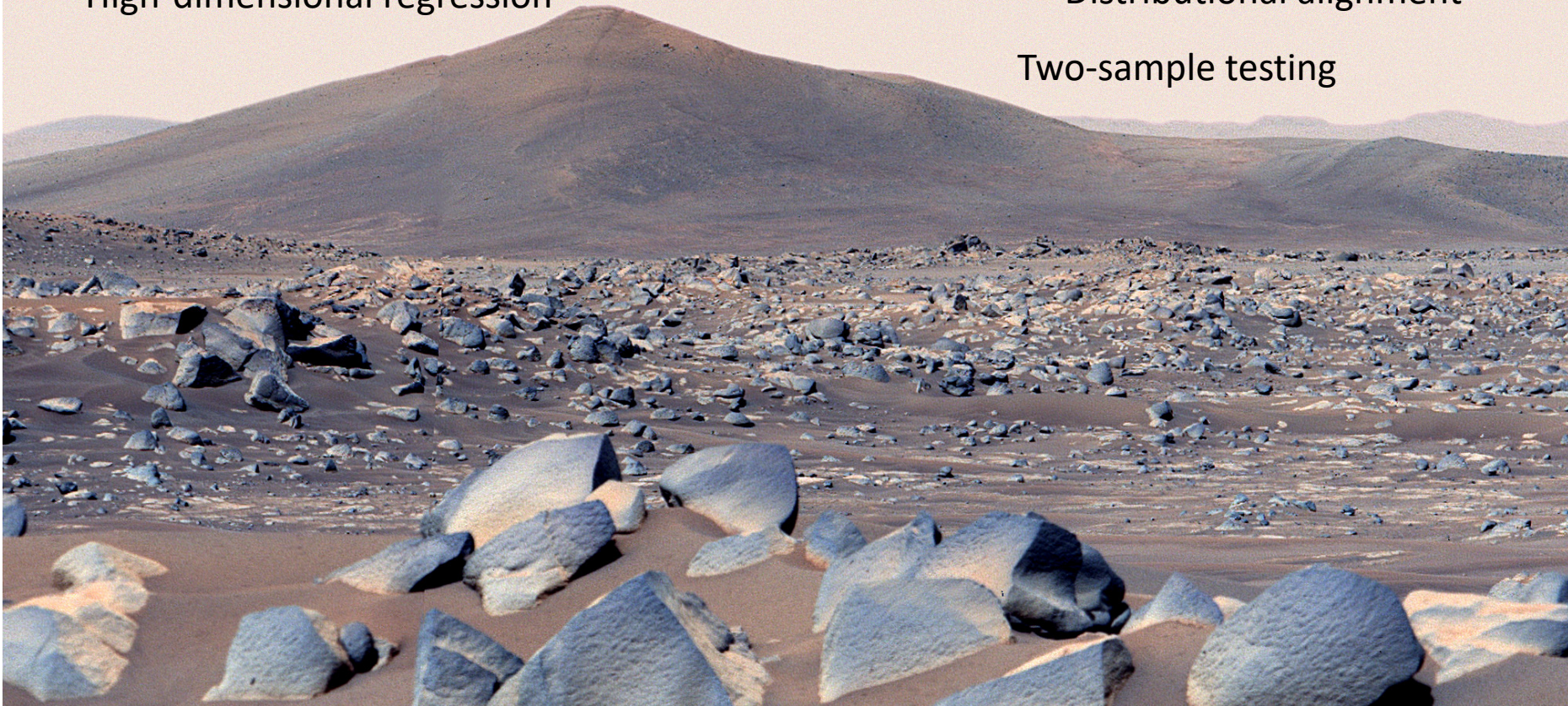
Statistical learning

Active learning and experimental design

High-dimensional regression

Distributional alignment

Two-sample testing



NASA/JPL-Caltech/MSSS