Los Alamos
NATIONAL LABORATORY

From Earth to Mars: Statistical Challenges in Analyzing Rover Spectroscopy Data

Natalie Klein
Statistics Group, Los Alamos

November 7, 2025

LA-UR-25-30678

Advancing Machine Learning for Scientific Discovery

- Intense 10-12 week research internship at LANL
- Information at **m.lanl.gov/aml** (applications will open soon)
- Questions: email **aml-admin@lanl.gov**

Acknowledgements

Postdocs: Dr. Weizhi Li, Dr. Scott Koermer, Dr. Michael Geyer

Students: Mark Hinds, Tom Winckelman, Giosue Migliorini, Reid Morris, Francisco Holguin, Katiana Kontolati

Collaborators: Dr. Samuel Clegg, Dr. Diane Oyen, Dr. Elizabeth Sklute, Dr. Brendan J. Gifford, Dr. Patrick J. Gasda, Dr. Nishant Panda

Funding: Laboratory Directed Research and Development

Other areas for potential collaboration (not in this talk)

- Surrogate models for Density Functional Theory to infer nuclear magnetic resonance parameters from 3D molecular structures with the goal of inverse solutions (identify materials from spectral features)
- Multimodal data analysis for time series/sensor data with robustness to noisy and missing modalities, UQ, and interpretability
- Analysis of time series/sensor data under distribution shift, especially when data is sparse or events of interest are rare
- Scalable uncertainty quantification for machine learning or AI models
- Statistical analysis of AI models

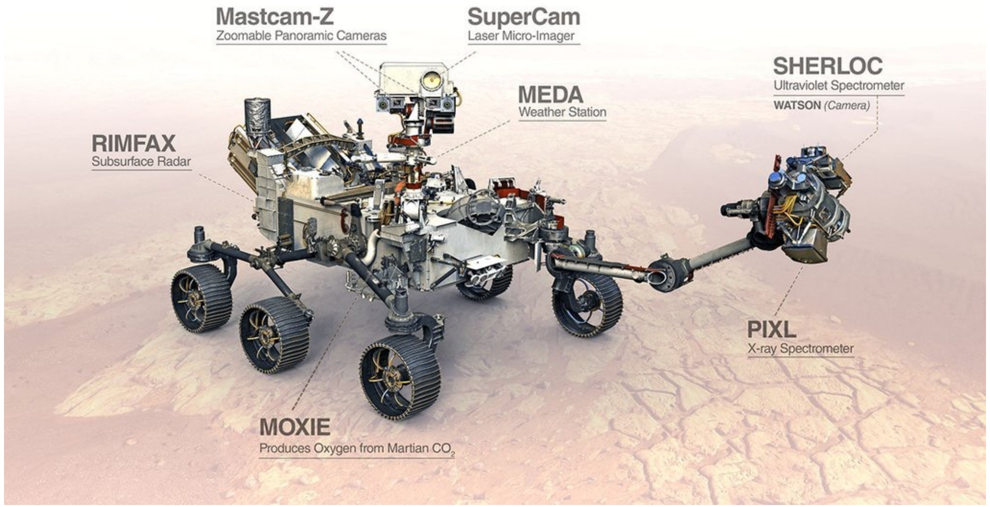
Outline

- Why Mars rover data?
- Landscape of data analysis
- Leveraging Earth and Mars data
- Multimodal data fusion
- Generative models for spectra
- Frontiers and ongoing work
- Conclusions and open questions

NASA *Curiosity* rover selfie

Why are we interested in Mars rover data?

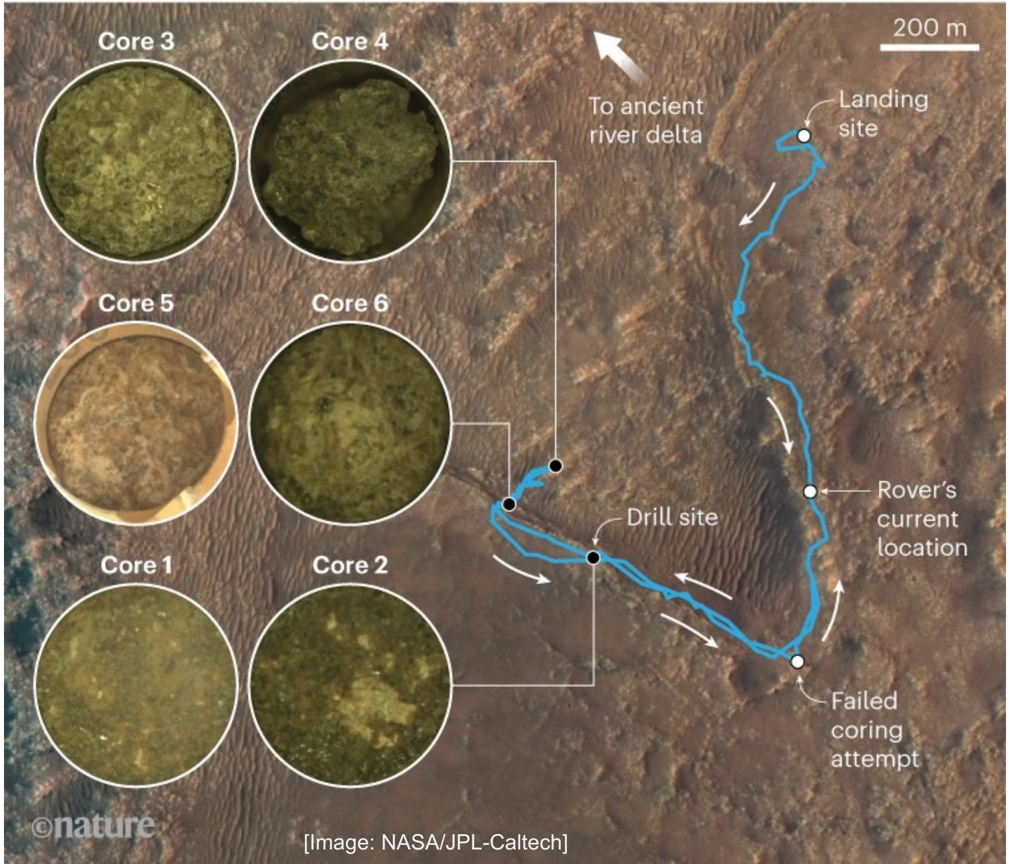
- Understand rock/soil composition
- Evaluate habitability (was there life on Mars?)
- Organics and biosignatures
- Analyze geochemistry to understand formation conditions (evidence of water?)
- Understand Earth's past and future climate



[NASA Perseverance rover]

SAMPLING MARS

NASA's Perseverance rover has drilled and collected three pairs of rock samples in Jezero Crater on Mars, the first Martian rocks ever intended to be brought back to Earth for scientific analysis. The rover will soon head to its ultimate destination, an ancient river delta, to look for signs of past life.



[Image: NASA/JPL-Caltech]

Potential biosignatures discovered! (September 2025)

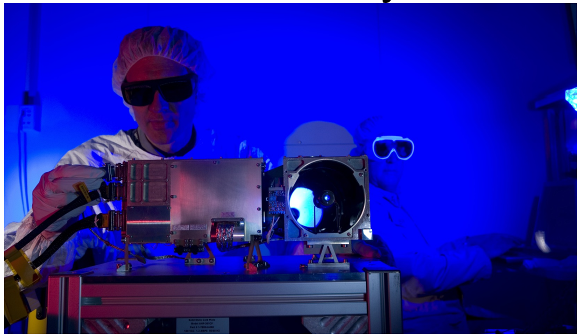
NASA's Perseverance rover discovered leopard spots on a reddish rock nicknamed "Cheyava Falls" in Mars' Jezero Crater in July 2024. Scientists think the spots may indicate that, billions of years ago, the chemical reactions in this rock could have supported microbial life; other explanations are being considered.

Credit: NASA/JPL-Caltech/MSSS

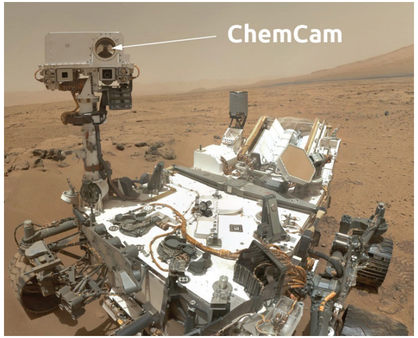
Statistical challenges in analyzing rover data

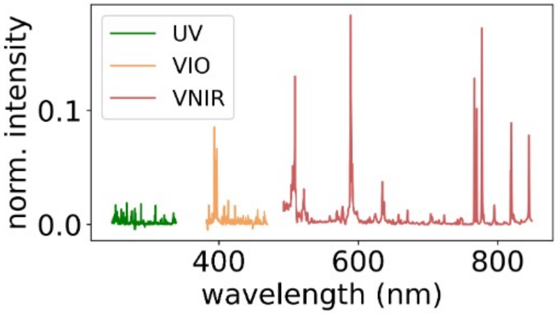
- Data is high-dimensional
- Data is multimodal (and many modalities are not yet quantitative)
- Data can be noisy
- Data is collected under shifting environmental conditions
- Data is collected by similar but distinct instruments

Laboratory



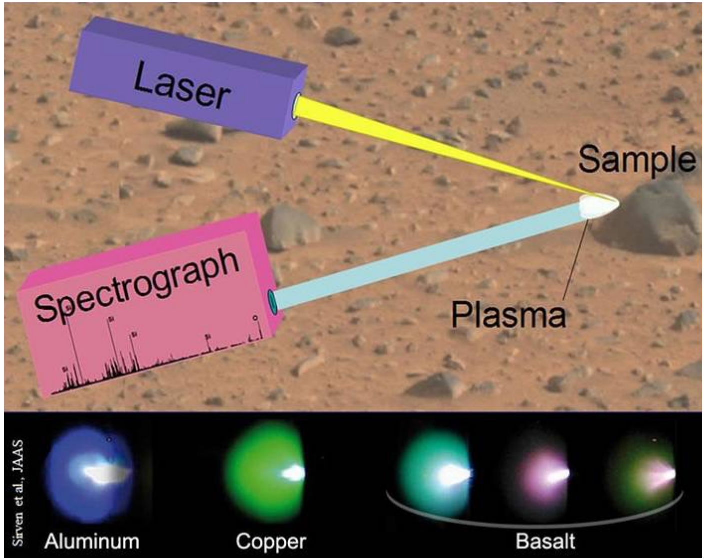
Mars

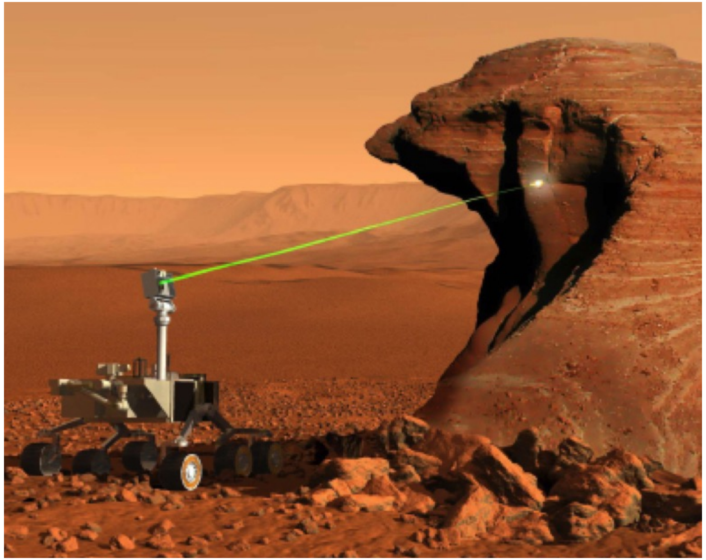




[Images: NASA/JPL-Caltech/LANL]

Much of our work focuses on ChemCam and SuperCam, which enable *elemental analysis*





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

Outline

- Why Mars rover data?
- Landscape of data analysis
- Leveraging Earth and Mars data
- Multimodal data fusion
- Generative models for spectra
- Frontiers and ongoing work
- Conclusions and open questions

NASA *Curiosity* rover selfie

Los Alamos NATIONAL LABORATORY

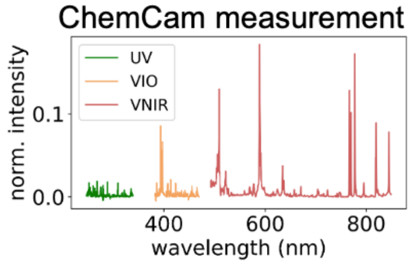
Classic approaches for elemental analysis

- Elemental analysis is commonly treated as a prediction problem

Prediction Problem

Chemical Composition

Oxide	Weight %
SiO_2	58.2
MgO	9.8
CaO	5.1
…	…



Common technique: partial least squares [e.g., Clegg et al 2017]

$$\mathbf{X} = \mathbf{Z}\mathbf{V}^{\mathsf{T}} + \mathbf{E}$$

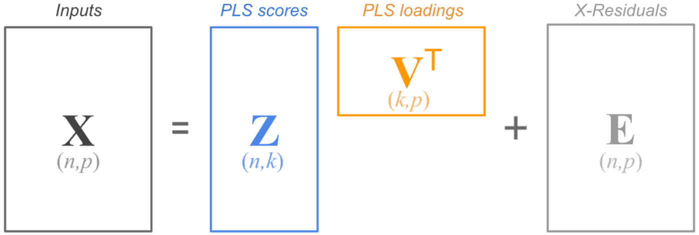


Figure 16.1: Matrix diagram for inputs

$$\mathbf{y} = \mathbf{Z}\mathbf{b} + \mathbf{e}$$

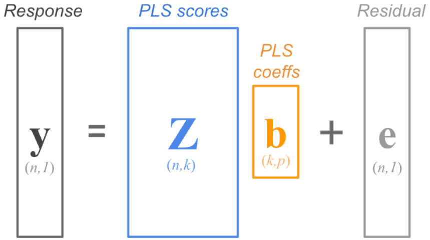
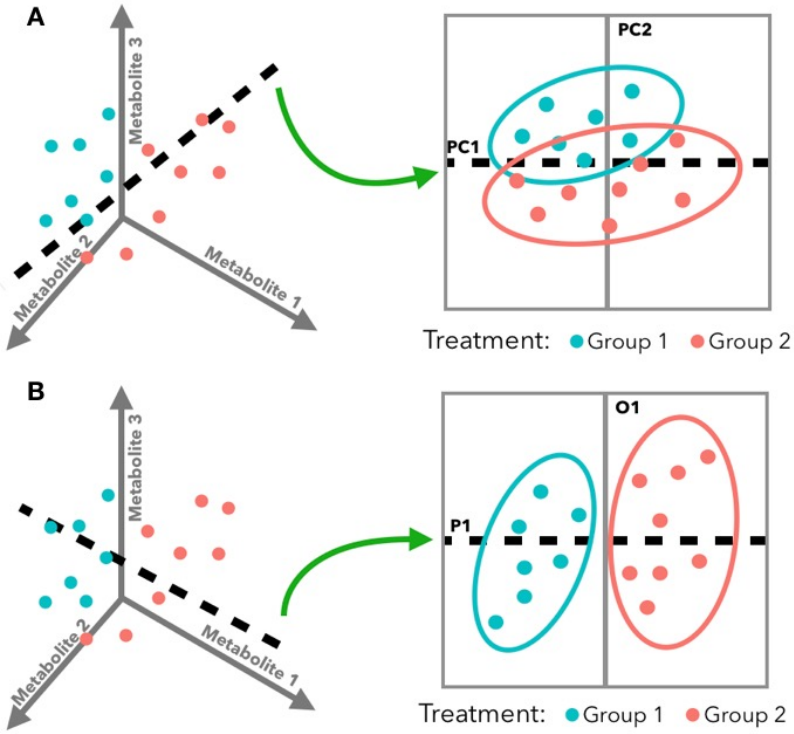


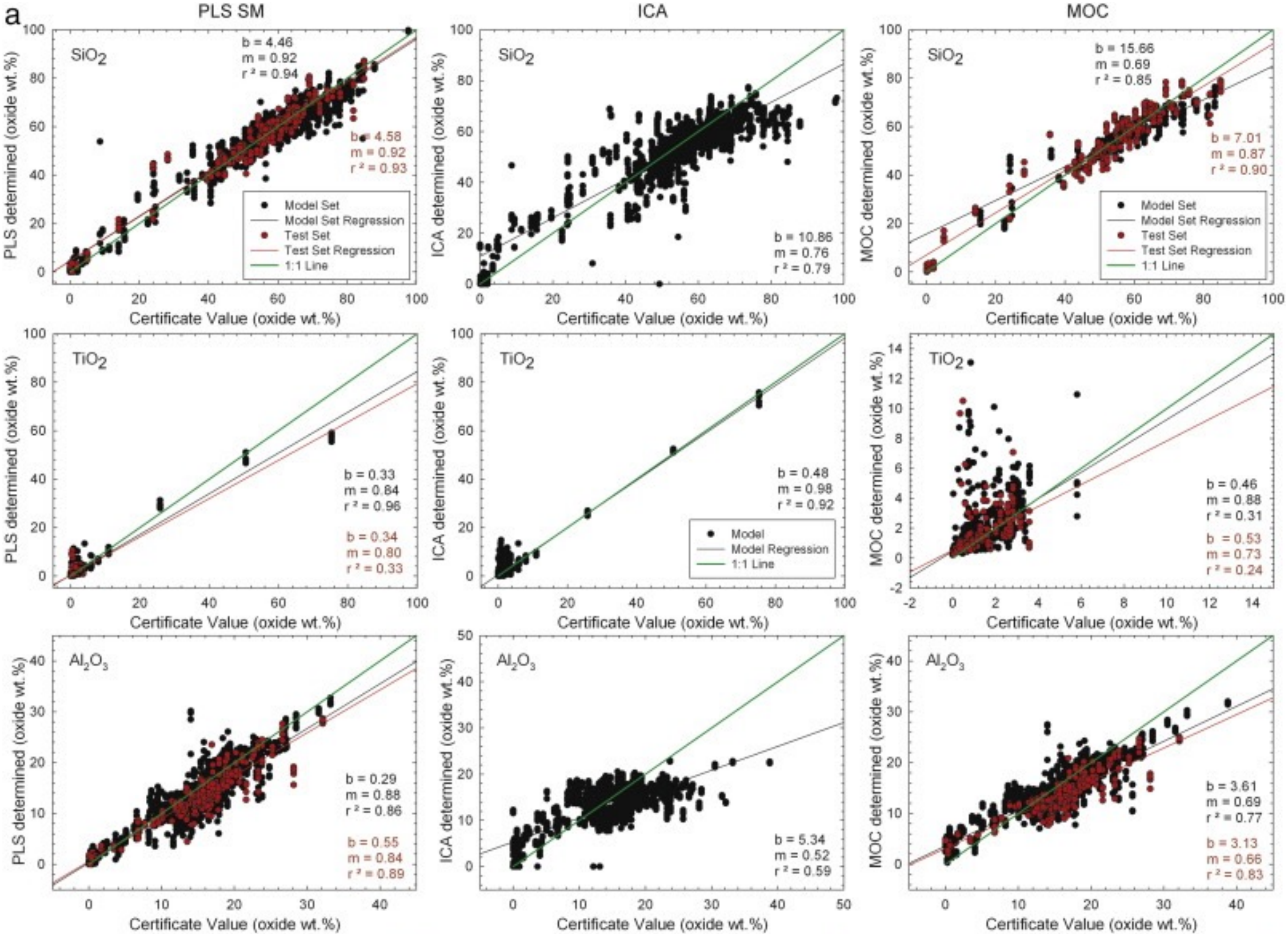
Figure 16.2: Matrix diagram for response

https://allmodelsarewrong.github.io/pls.html



https://ericrscott.com/project/pls-ecology/

Los Alamos NATIONAL LABORATORY

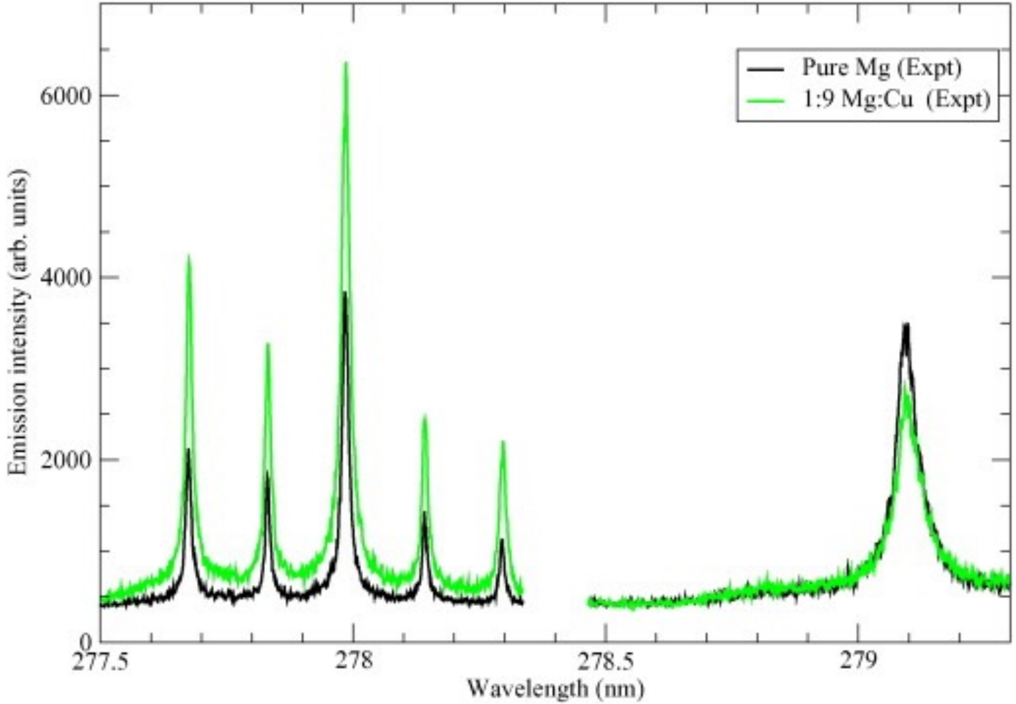


Los Alamos NATIONAL LABORATORY

[Clegg et al 2017]

Why go beyond partial least squares?

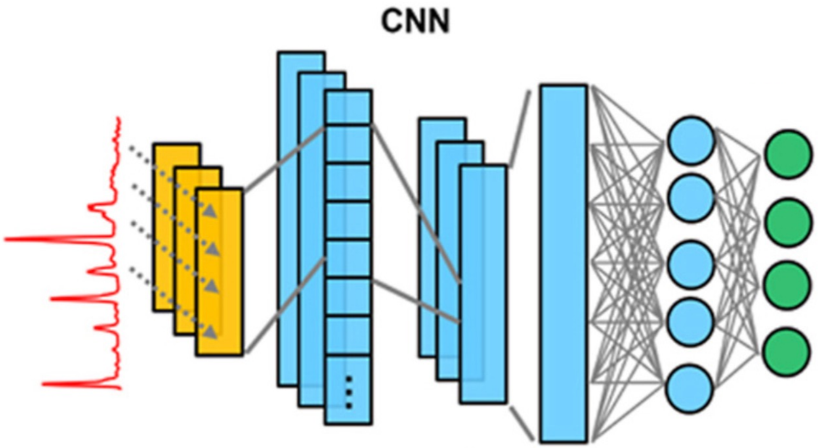
- Matrix effects (nonlinear interactions between elements)



Judge et al 2016 https://doi.org/10.1016/j.sab.2016.06.004

Modern approaches for elemental analysis

Algorithm	Abbreviation
Ordinary Least Squares	OLS
Partial Least Squares	PLS
Least Absolute Shrinkage and Selection Operator	LASSO
Ridge regression	Ridge
Elastic Net	ENet
Orthogonal Matching Pursuit	OMP
Support Vector Regression	SVR
Random Forest	RF
Gradient Boosting Regression	GBR
Local Elastic Net	Local ENet
Blended submodels	Blend



Official SuperCam post-landing calibration
[Anderson et al 2022 https://doi.org/10.1016/j.sab.2021.106347]

Challenges with modern approaches

- Need to balance accuracy and interpretability
- Numerous choices (model selection, data preprocessing)
- Want to understand uncertainty (expect some extrapolation)
- Data sets are small (~500-800 unique targets)

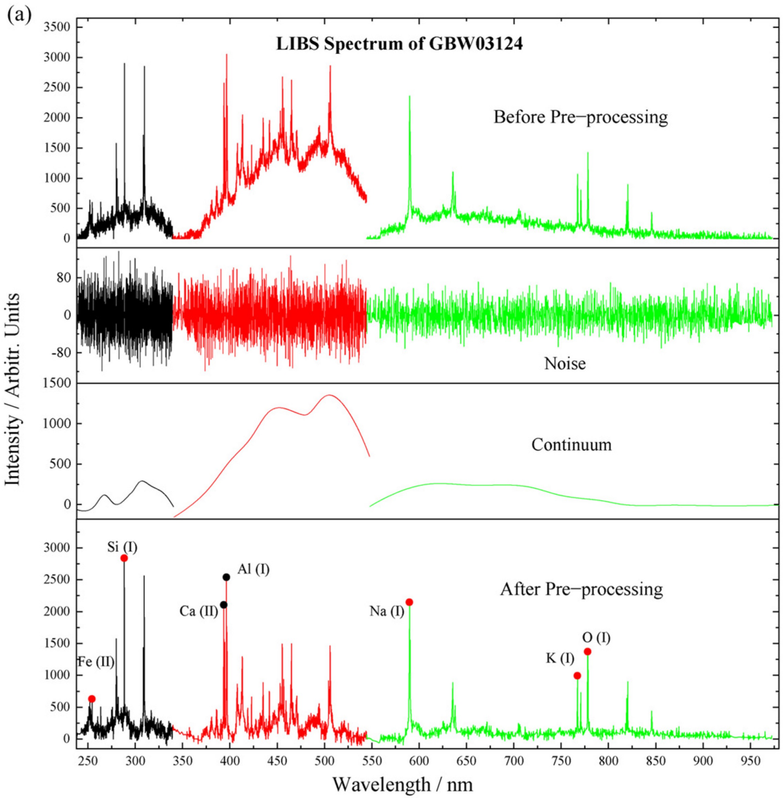
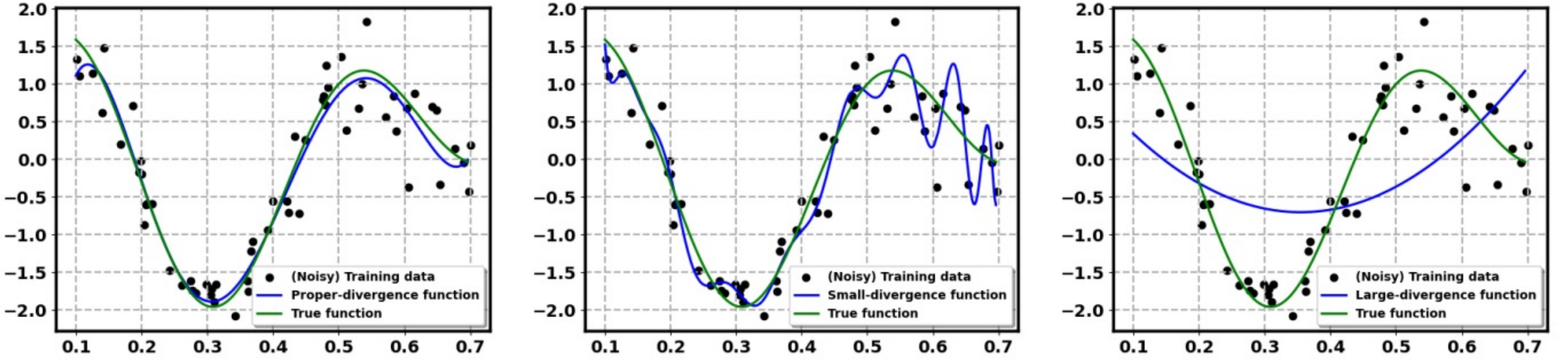


Image: Liu et al 2021, https://doi.org/10.3390/rs13234773

We've explored the role of regularization

- CNNs with regularization perform better on out-of-sample prediction

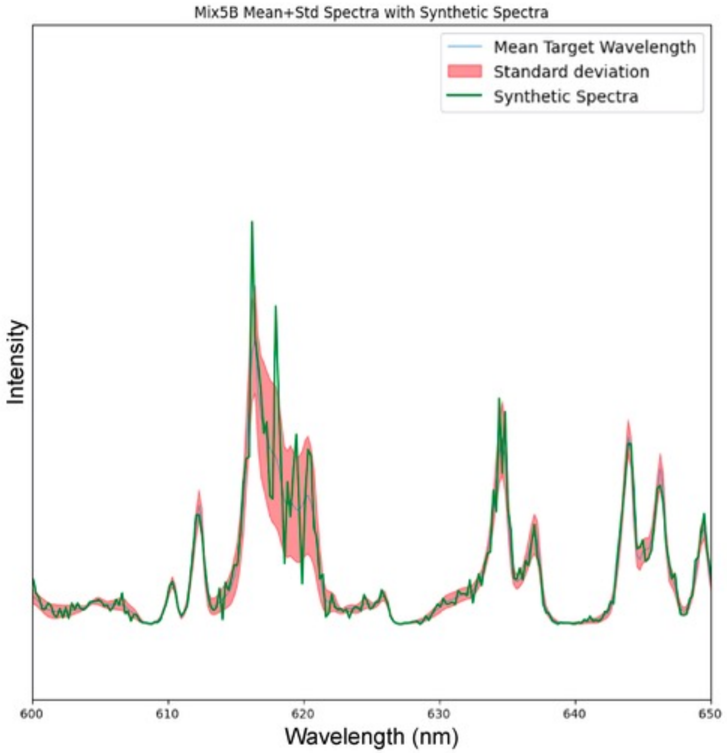


(a) Training MSE: 0.13 (**Proper divergence**) Test MSE: **0.17**
(b) Training MSE: 0.09 (Small divergence) Test MSE: 0.24
(c) Training MSE: 0.88 (Large divergence) Test MSE: 1.02

Fig. 1. True function and various approximation functions are compared, along with their training and test mean squared errors (MSE). The training MSE *implicitly* quantifies the divergence between the training targets and predictions. Consequently, the function approximation in (a), which maintains an appropriate level of divergence, achieves a smaller test error compared to the approximations in (b) and (c), where the divergences are too small and large, respectively.

[Li et al 2025 https://www.arxiv.org/abs/2502.03755]

We’ve explored data augmentation approaches



Data augmentation approaches

- Resampling based on observed spectral variation
- Wavelength shifts

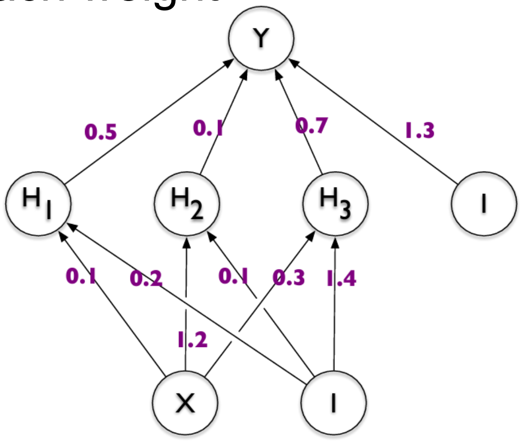
Results

- Increased robustness to noise and wavelength miscalibration
- Ability to train more complex models

[Morris et al, poster at LPSC 2024]

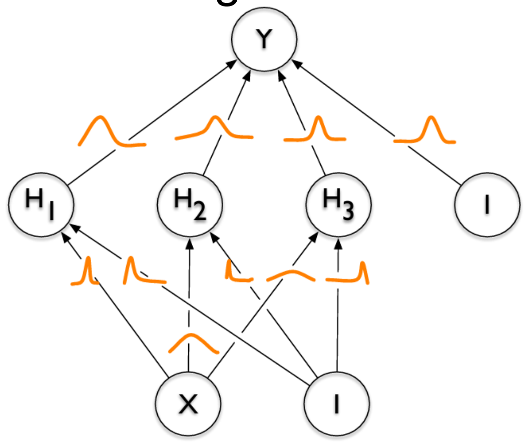
We’ve explored UQ with Bayesian neural networks

Standard neural network: find a good *single value* for each weight



Result: one prediction for each input

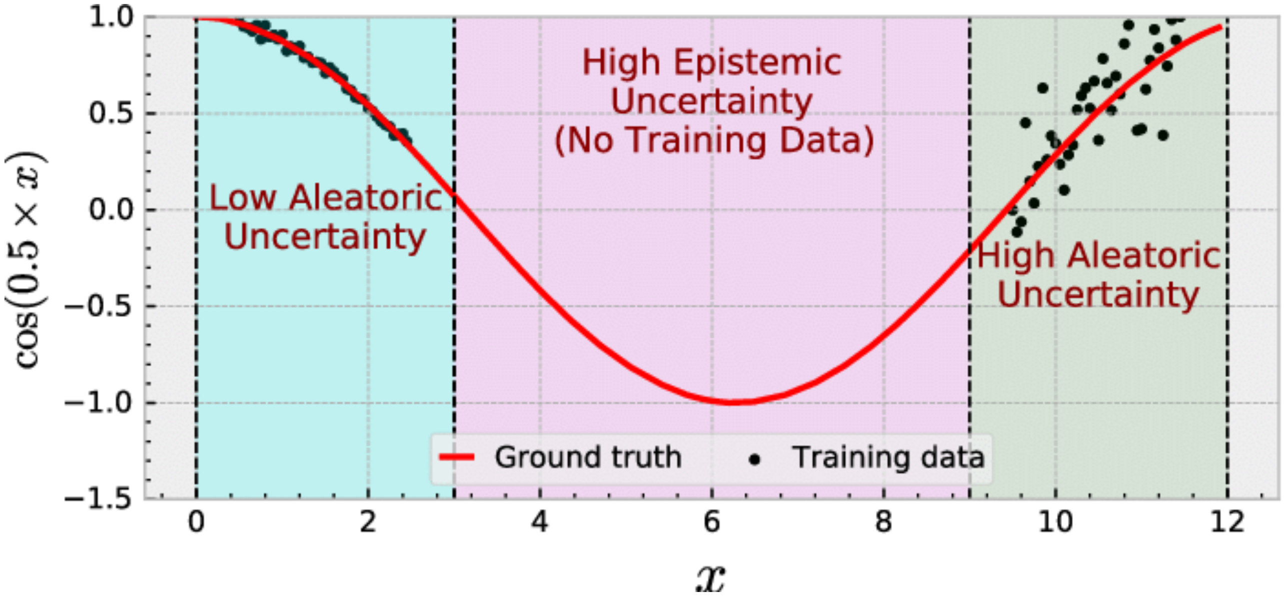
UQ-aware neural network: find a plausible *distribution* for each weight



Result: predictive uncertainty

We find it critical to distinguish epistemic and aleatoric uncertainty.

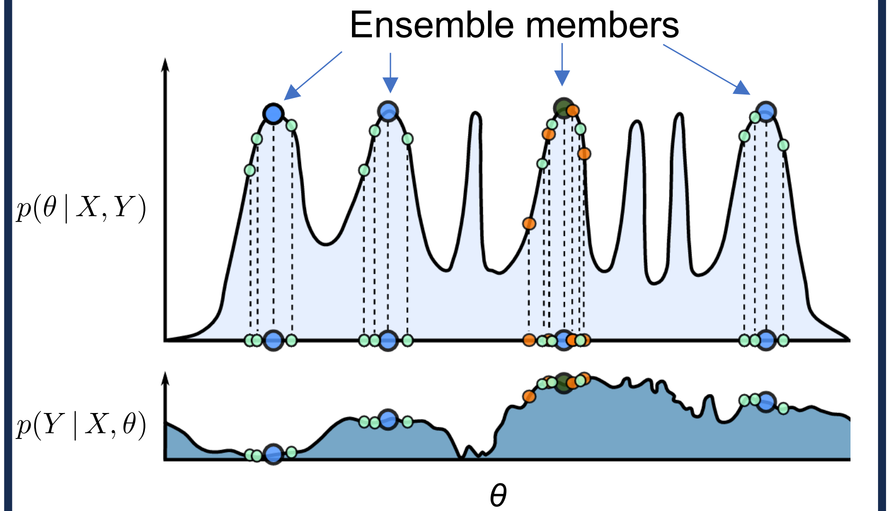
- *"Why is the model uncertain and how does that affect decision-making?"*



Tuna, O. F., Catak, F. O., & Eskil, M. T. (2023). Uncertainty as a Swiss army knife: new adversarial attack and defense ideas based on epistemic uncertainty. *Complex & Intelligent Systems*.

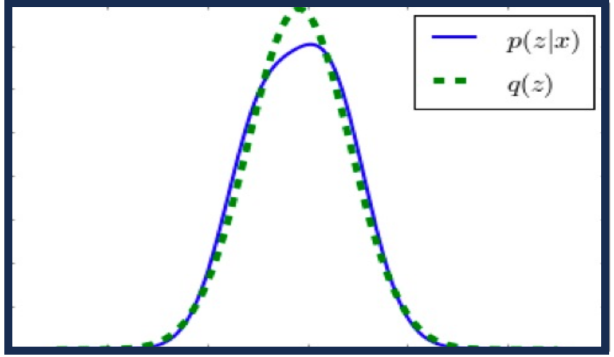
Some scalable methods for epistemic uncertainty: ensembles and variational Bayesian inference.

Deep Ensemble

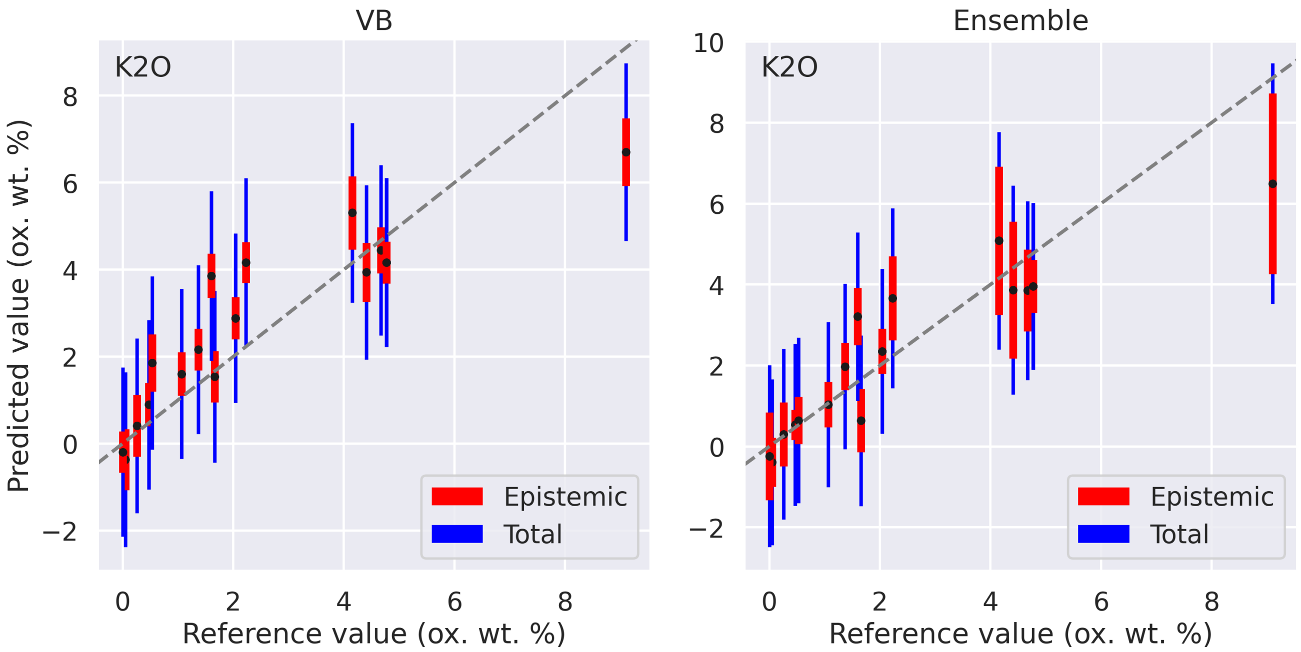


Adapted from [Wilson and Izmailov 2020]

Variational Bayes

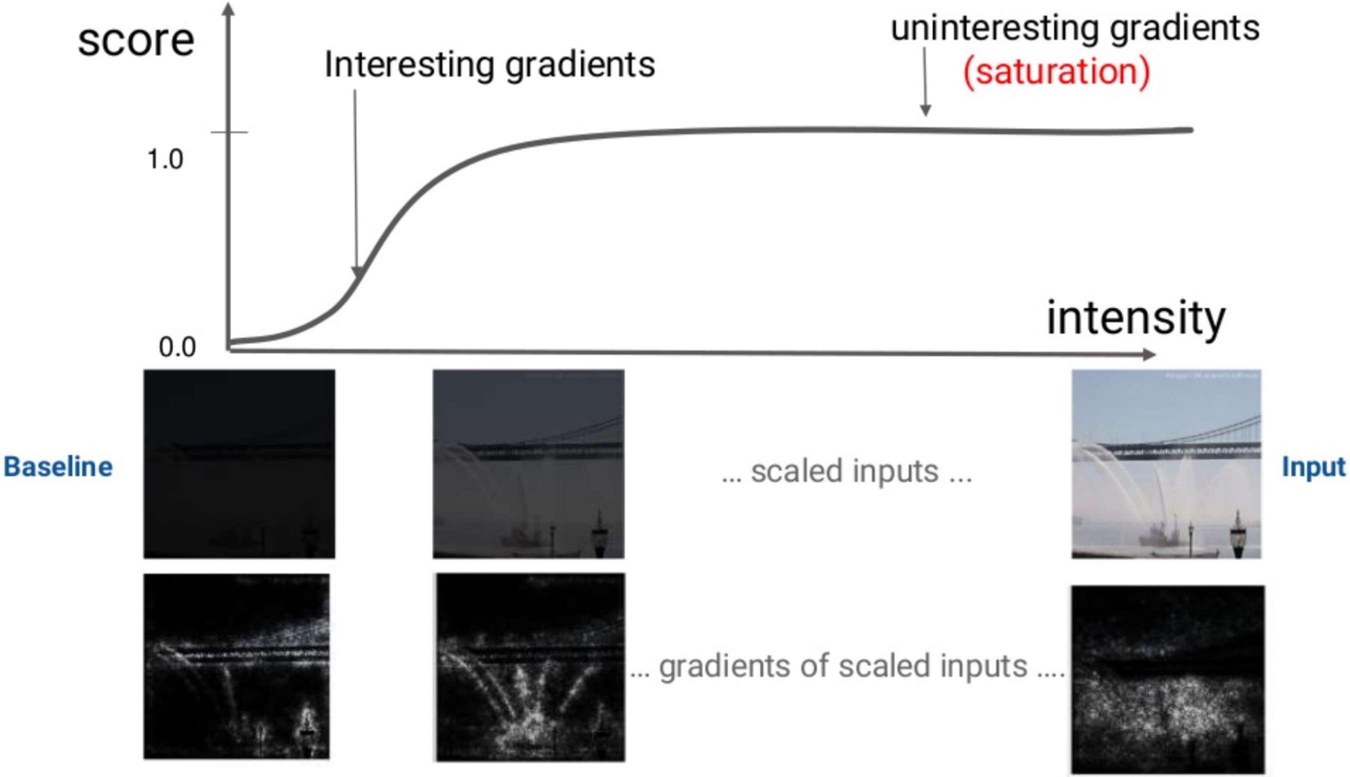


Bayesian neural networks can give well-calibrated UQ



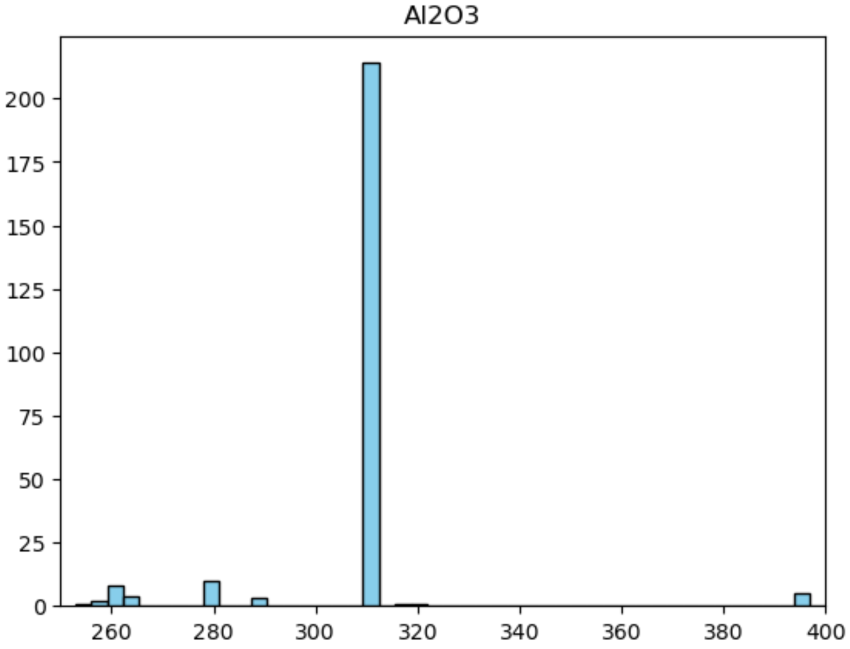
[Klein et al 2025 https://doi.org/10.61278/itea.46.3.1006]

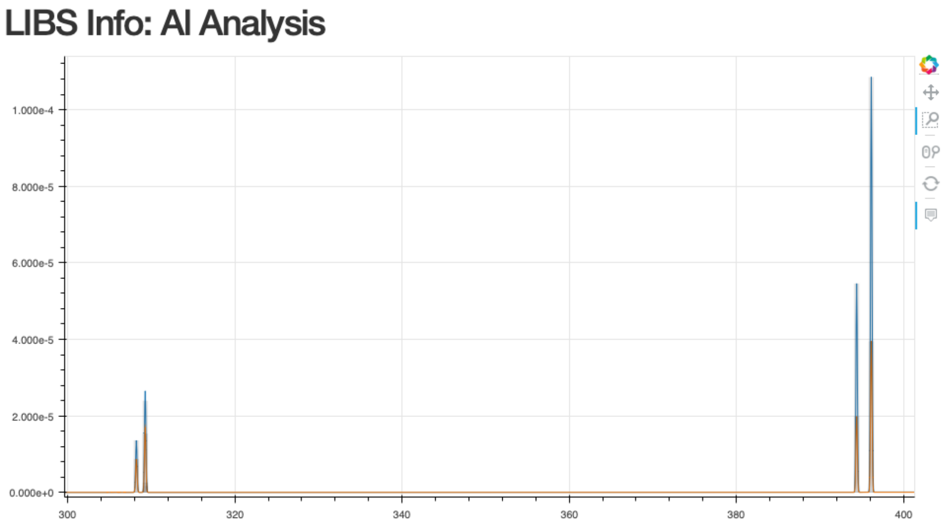
We are exploring interpretability approaches for CNNs.



[Image by Krishnaram Kenthapadi: Explainalbe AI in Industry (KDD 2019 Tutorial)]

Initial results show good alignment of strong integrated gradient scores with physically-relevant features.

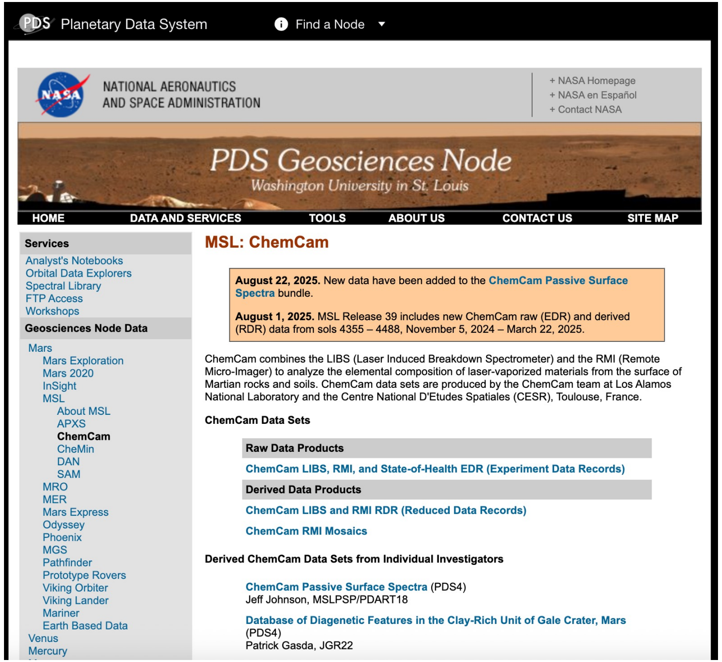




[Led by Brendan J. Gifford]

Mars rover data is an interesting testbed for methods

- Balance accuracy, interpretability, uncertainty
- LIBS data is available on the PDS (along with a lot of other data!)
  - ChemCam: https://pds-geosciences.wustl.edu/missions/msl/chemcam.htm
- Beyond elemental predictive models
  - Leveraging data from different environments or instruments
  - Leveraging multimodal data
  - Generative spectral models



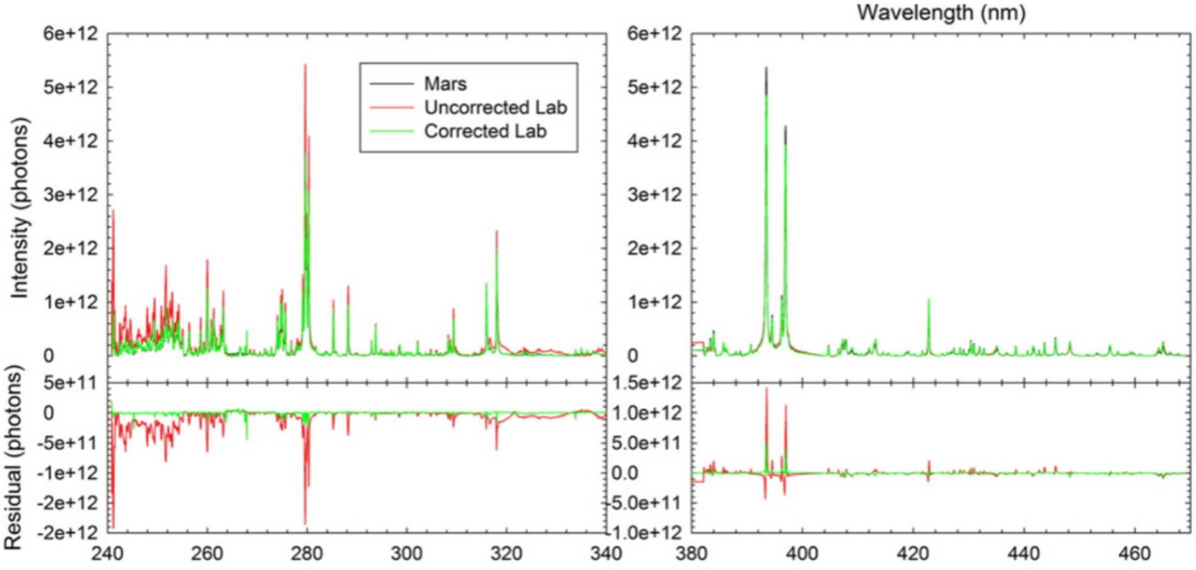
Outline

NASA *Curiosity* rover selfie

Data shifts across Earth and Mars

How to account for differences and represent uncertainty in the mapping?

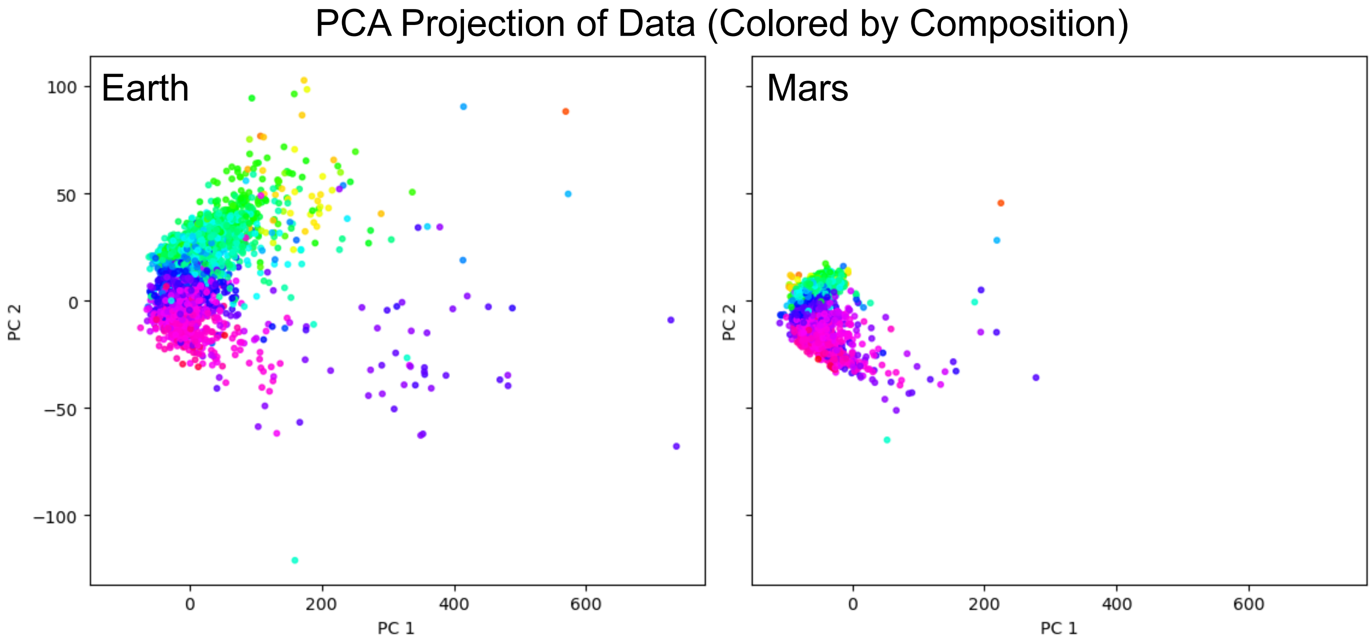
Johnson et al (2015)



Clegg et al (2017)

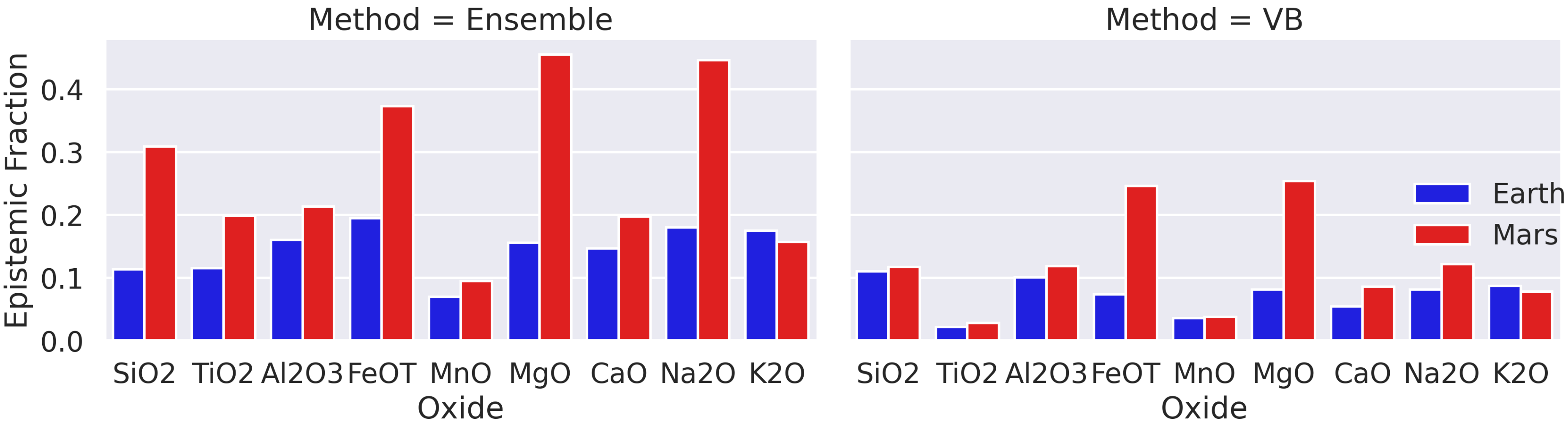
There are evident differences between data collected under different environmental conditions.

- SuperLIBS data in Earth and Mars atmospheric chambers



Los Alamos
NATIONAL LABORATORY

Epistemic uncertainty could indicate out of distribution data (on Mars)

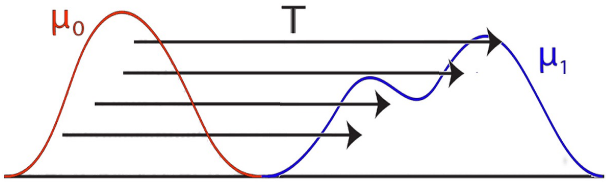


[Klein et al 2025 https://doi.org/10.61278/itea.46.3.1006]

Very little labeled Mars data!

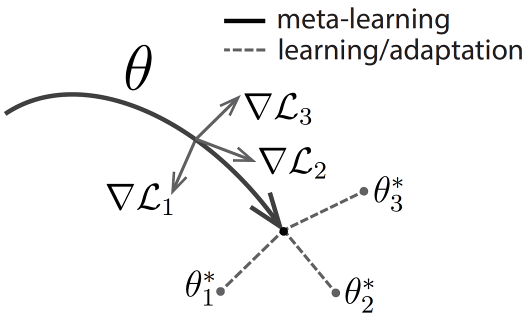
Approaches to handling domain shift

- Data-level strategies
  - Transform samples from one environment to better match samples from another environment, potentially using label or pairing information
  - Pseudo-labeling or self-training

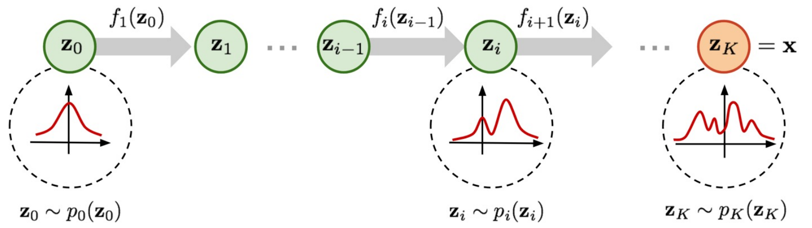


N. Papadakis, Optimal Transport for Image Processing, habilitation à diriger des recherches, Université de Bordeaux, Dec. 2015

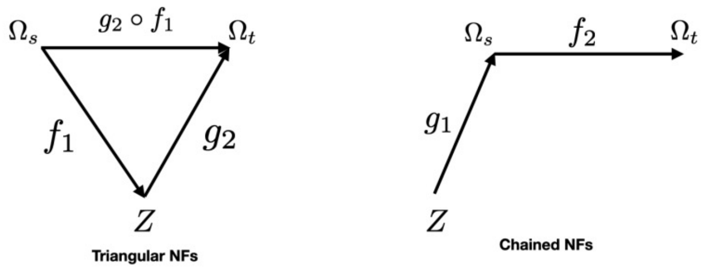
- Model-level strategies
  - Domain-invariant models (e.g., adversarial domain adaptation)
  - Importance weighting
  - Meta-learning or domain-specific adapters



We developed semi-supervised data-centric transformations based on normalizing flows (parOT).



https://lilianweng.github.io/posts/2018-10-13-flow-models/



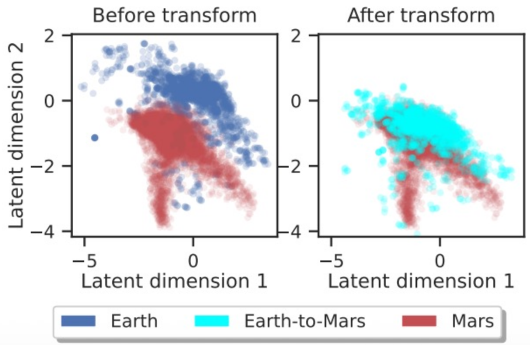
TRIANGLE NF	
TASK	PROCEDURE
SAMPLE Ω_s	SAMPLE Z AND APPLY g_1
SAMPLE Ω_t	SAMPLE Z AND APPLY g_2
LIKELIHOOD $x \in \Omega_s$	$\rho_Z(f_1(x))\lvert\det Df_1(x)\rvert$
LIKELIHOOD $x \in \Omega_t$	$\rho_Z(f_2(x))\lvert\det Df_2(x)\rvert$
PUSHFORWARD MAP	$g_2 \circ f_1$
CHAINED NF	
TASK	PROCEDURE
SAMPLE Ω_s	SAMPLE Z AND APPLY g_1
SAMPLE Ω_t	SAMPLE Z AND APPLY $f_2 \circ g_1$
LIKELIHOOD $x \in \Omega_s$	$\rho_Z(f_1(x))\lvert\det Df_1(x)\rvert$
LIKELIHOOD $x \in \Omega_t$	$\rho_Z(f_1(g_2(x)))\lvert\det D(f_1 \circ g_2)(x)\rvert$
PUSHFORWARD MAP	f_2

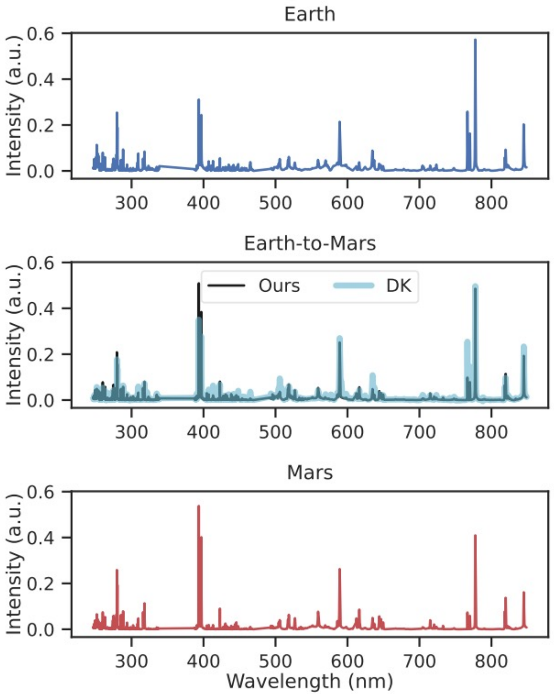
[Panda et al 2023 https://arxiv.org/abs/2304.08673]

For ChemCam data, parOT was successful in learning a transformation from Earth to Mars data.

We showed:

- Improved performance on Mars calibration targets
- Ability to incorporate auxiliary information (e.g., paired data)
- Ability to detect outliers via likelihood scores





[Panda et al 2023 https://arxiv.org/abs/2304.08673]

We have explored importance weighting for combining datasets collected under different conditions.

- Consider Earth and Mars data with the goal of training a model with limited Mars data to minimize the Mars risk:

$$R_{\text{Mars}}(\theta) = \mathbb{E}_{p_{\text{Mars}}(x,y)}\left[\ell(f_\theta(x), y)\right]$$

- This can be rewritten

$$R_{\text{Mars}}(\theta) = \int\int \ell(f_\theta(x), y) p(y|x) p_{\text{Earth}}(x) \frac{p_{\text{Mars}}(x)}{p_{\text{Earth}}(x)}\, dx\, dy$$

$$= \mathbb{E}_{p_{\text{Earth}}(x,y)}\left[\frac{p_{\text{Mars}}(x)}{p_{\text{Earth}}(x)} \ell(f_\theta(x), y)\right]$$

Earth data
- Abundant
- Easy to collect
- Labeled

Mars data
- Labeled is scarce
- Unlabeled is more abundant
- Expensive to collect labels

Density ratios can be estimated via classification.

- Train a classifier to separate Earth and Mars data:

$$s(x) = P(z = 1|x) = \frac{1}{1 + \exp(h(x))}$$

- By Bayes' Rule,

$$\frac{p_{\text{Mars}}(x)}{p_{\text{Earth}}(x)} = \frac{p(z=0)}{p(z=1)} \frac{s(x)}{1 - s(x)}$$

$z = 0$

Earth data
- Abundant
- Easy to collect
- Labeled

$z = 1$

Mars data
- Labeled is scarce
- Unlabeled is more abundant
- Expensive to collect labels

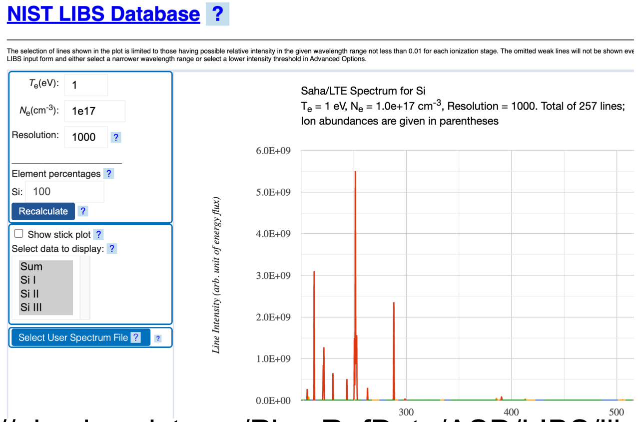
Importance weighting improves neural network models for Mars data with limited labels.

- Estimating weights: 2,000 Earth/Mars (no labels needed)
- Predictive model: 1,000 Earth (labeled), 100 Mars (labeled)

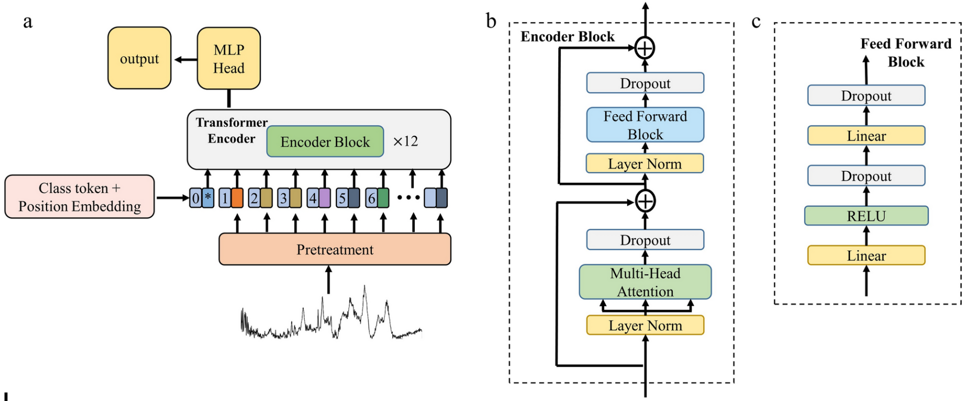
Training	R-squared (Mars test set)
Train on Earth only	0.23
Train on limited labeled Mars only	0.001
Train on Earth+Mars (no weighting)	0.39
Train on Earth+Mars (weighting)	0.51

Big outstanding challenge: avoid lengthy data collections and recalibrations.

- Leverage related but distinct data to inform new data collections
- Preliminary idea for LIBS: transformer model pre-trained on synthetic data
  - Ability to handle different wavelengths
  - Incorporate different physical parameters and line widths during training
  - "Foundation model" for LIBS analysis



https://physics.nist.gov/PhysRefData/ASD/LIBS/libs-form.html

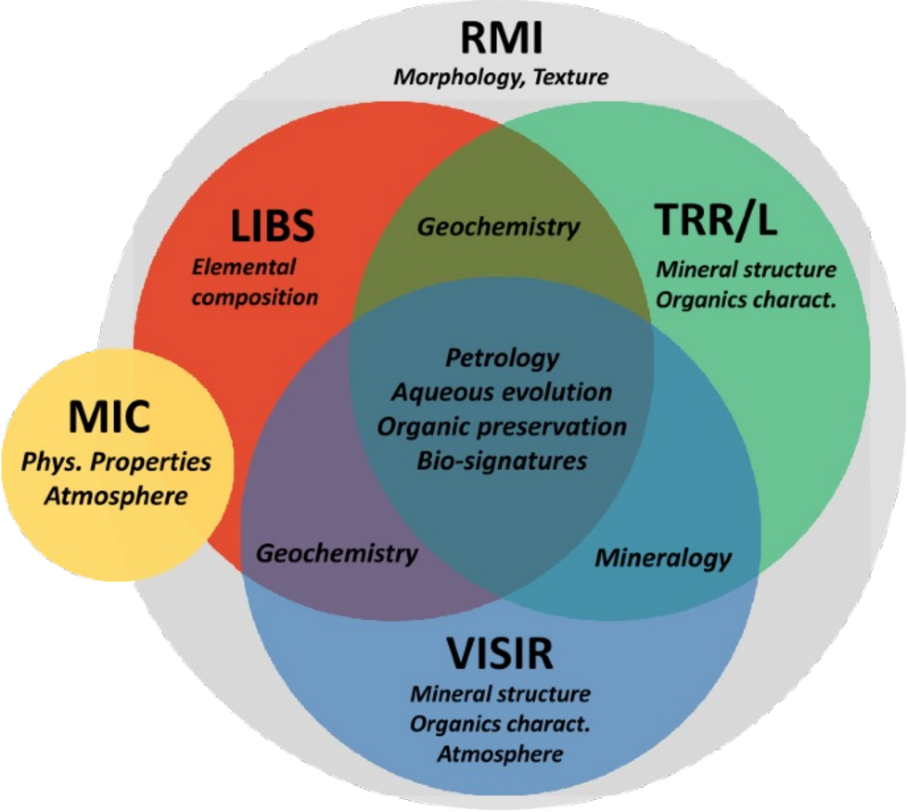


https://doi.org/10.1038/s41598-023-28730-w

Outline

NASA *Curiosity* rover selfie

Multimodal fusion across instruments is of great interest

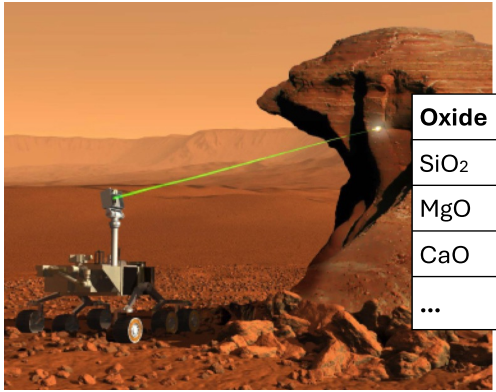


RMI: Remote micro-imaging
LIBS: Laser-induced breakdown spectroscopy
TRR/L: Time-resolved Raman and Luminescence
VISIR: Visible and infrared spectroscopy
MIC: Microphone

[Maurice et al 2021 https://doi.org/10.1007/s11214-021-00807-w]

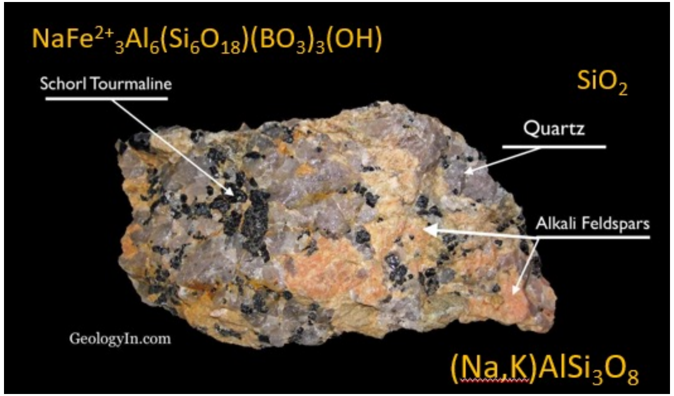
Motivation behind data fusion for planetary applications

Elemental analysis
(e.g., LIBS)



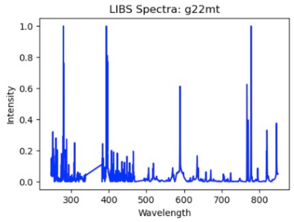
Oxide	Weight %
SiO_2	58.2
MgO	9.8
CaO	5.1
…	…

Mineralogy
(e.g., Raman or VISIR)

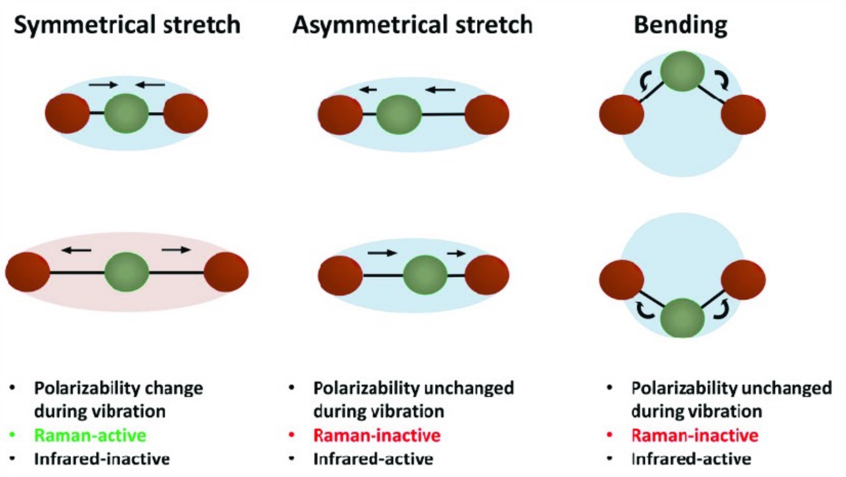


Elemental analysis + Mineralogy = Understanding formation environment

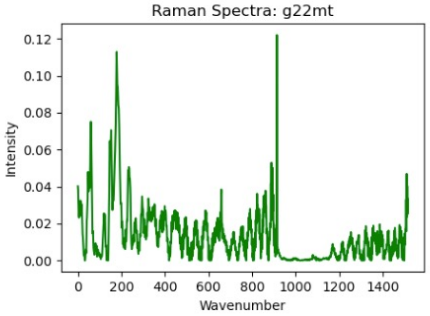
We seek to combine elemental and vibrational spectra

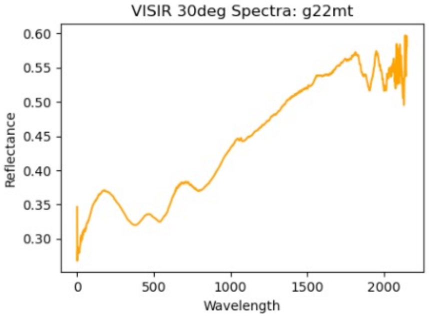


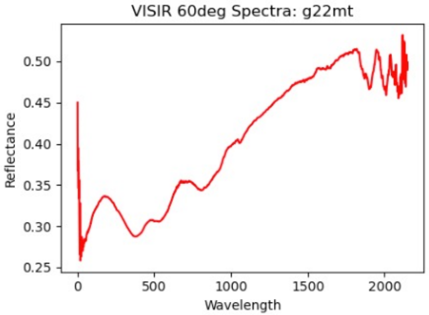
LIBS: Elemental

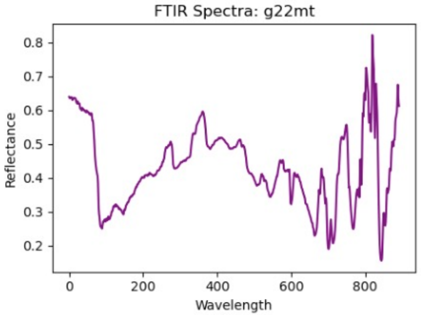


10.13140/RG.2.2.36752.48641



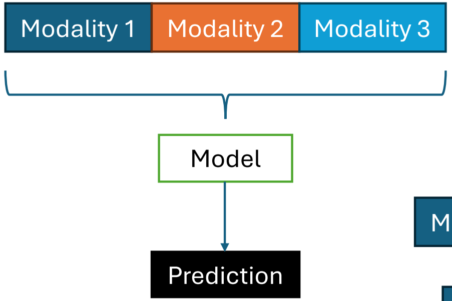




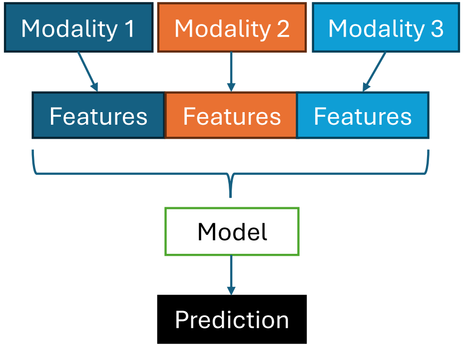


Data fusion approaches

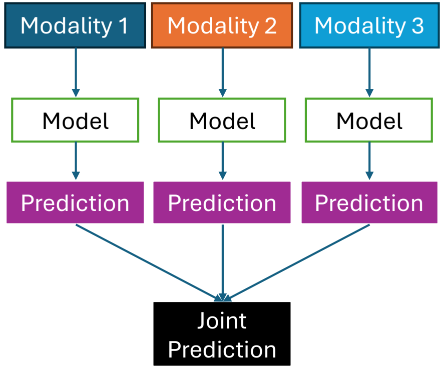
Low-level



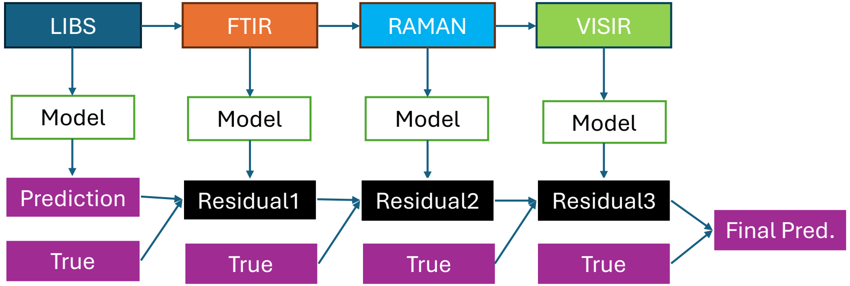
Mid-level



High-level

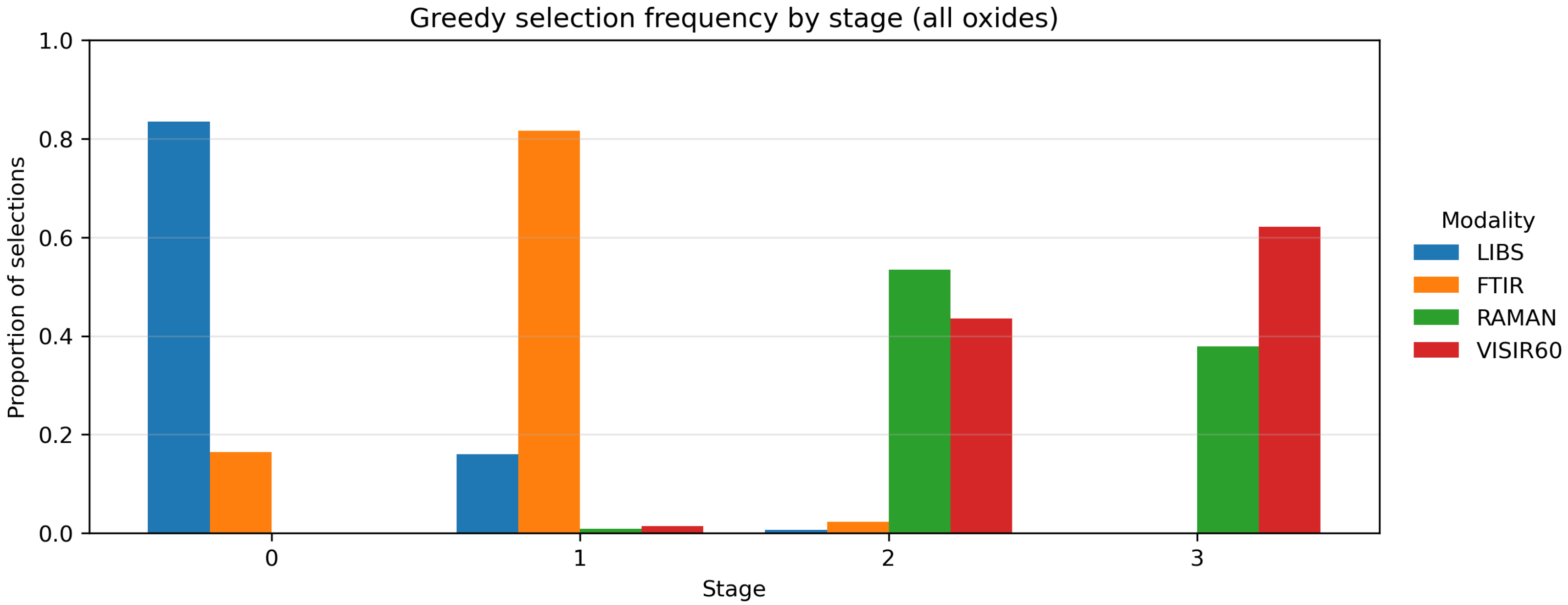


Boosted



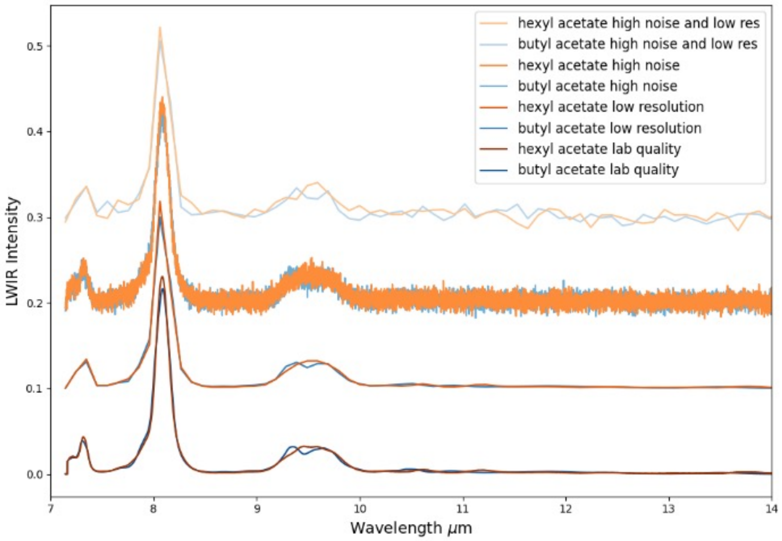
Interpretability

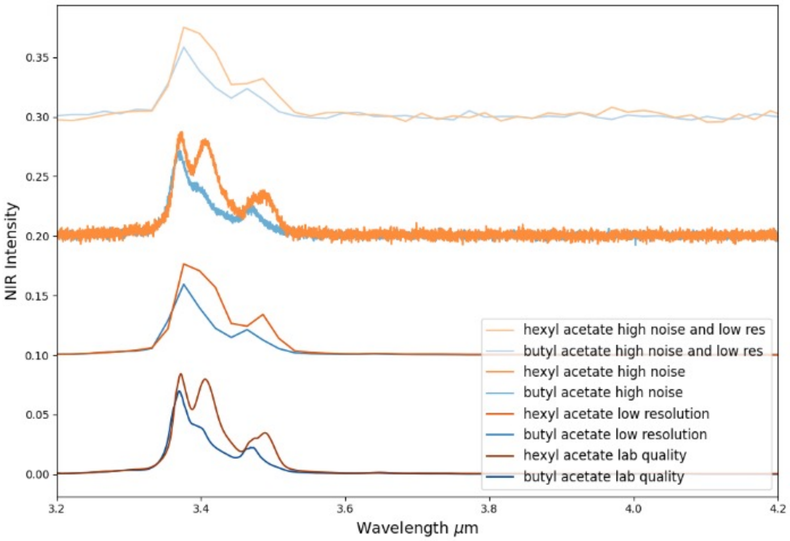
- Which modalities contribute most to the prediction?



Ongoing work: how to incorporate uncertainty on missing or noisy modalities and how to predict which modalities are most informative?

- Some modalities will not show distinct signatures and may vary in quality





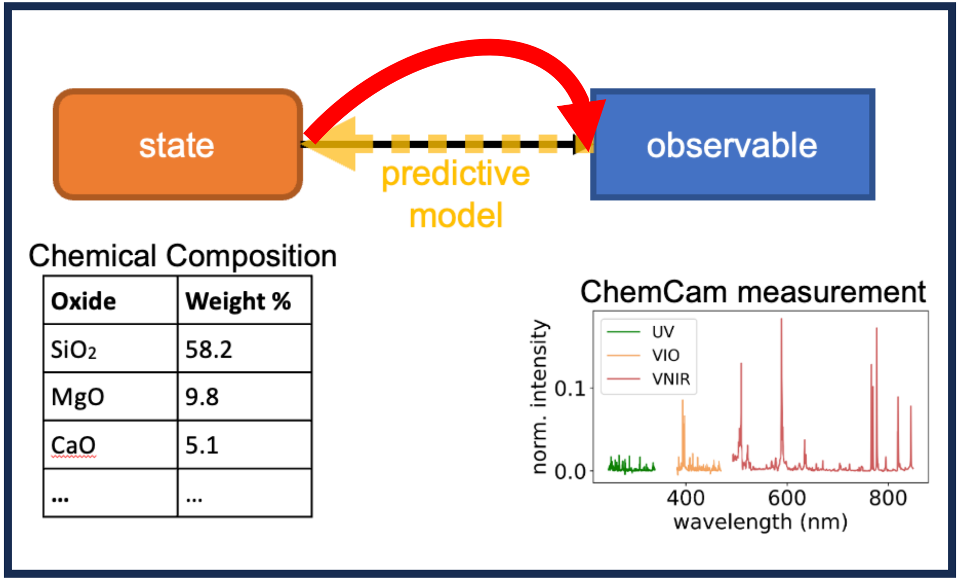
Outline

- Why Mars rover data?
- Landscape of data analysis
- Leveraging Earth and Mars data
- Multimodal data fusion
- Generative models for spectra
- Frontiers and ongoing work
- Conclusions and open questions

NASA *Curiosity* rover selfie

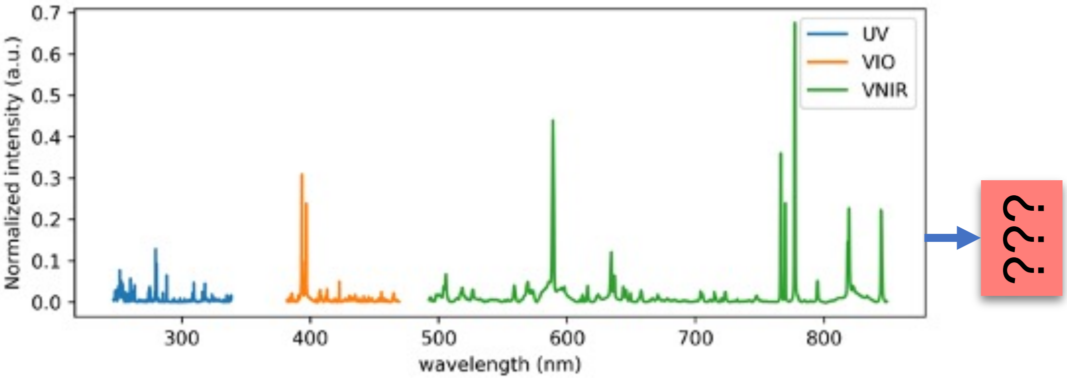
Idea: model spectra as function of elemental composition

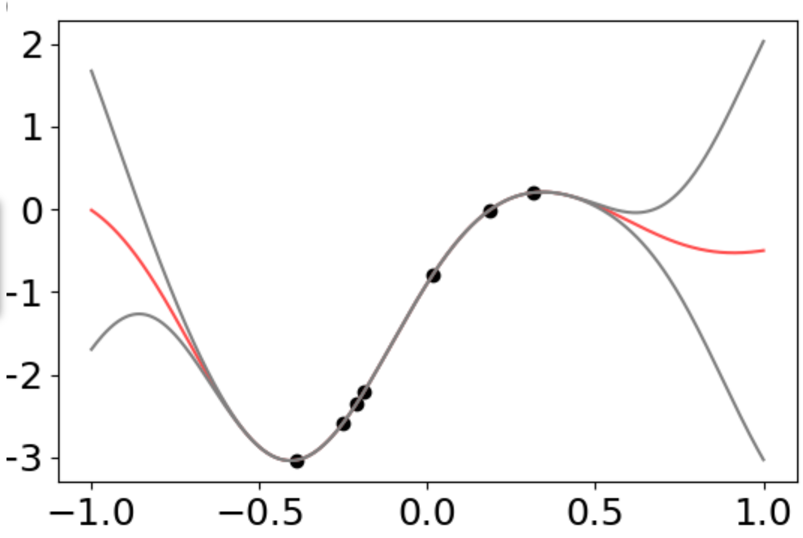
- May be more natural for statistical modeling, UQ
- Challenge: high-dimensional structured output



Generative models for LIBS spectra

- Goal: model LIBS spectra as function of ~10 inputs

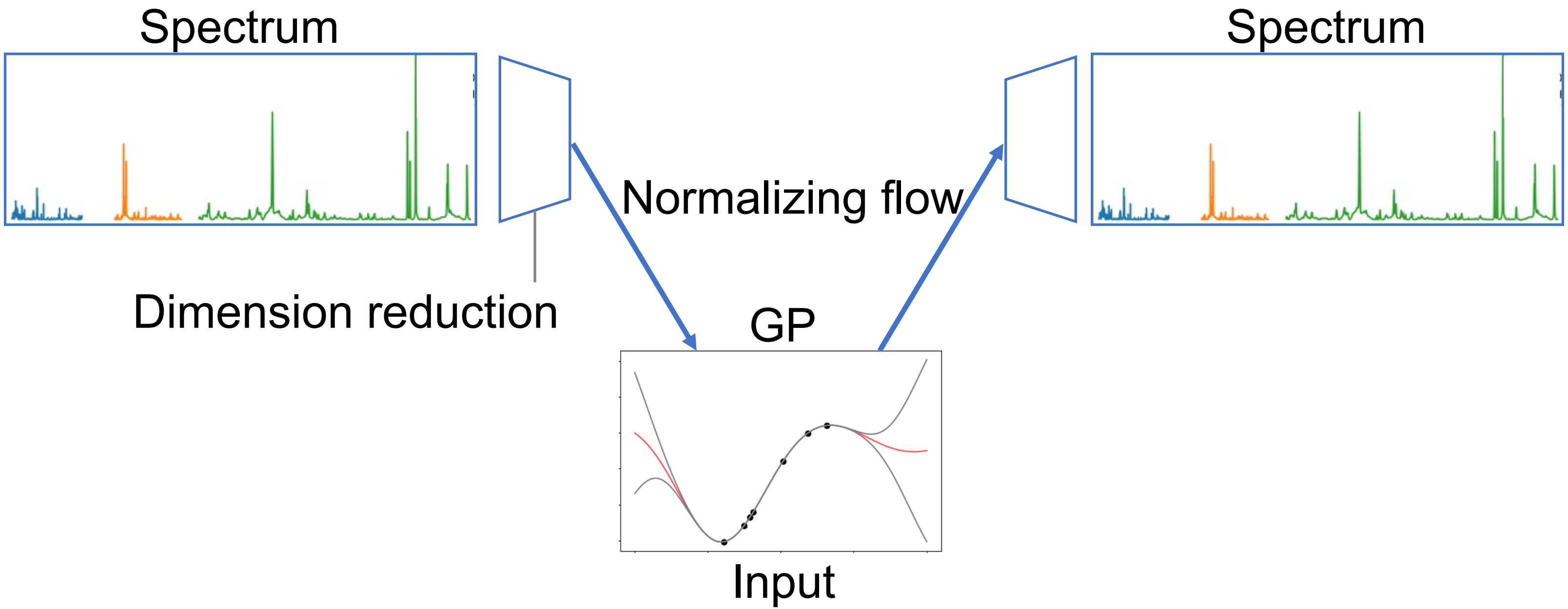




Input value
(e.g., plasma temperature)

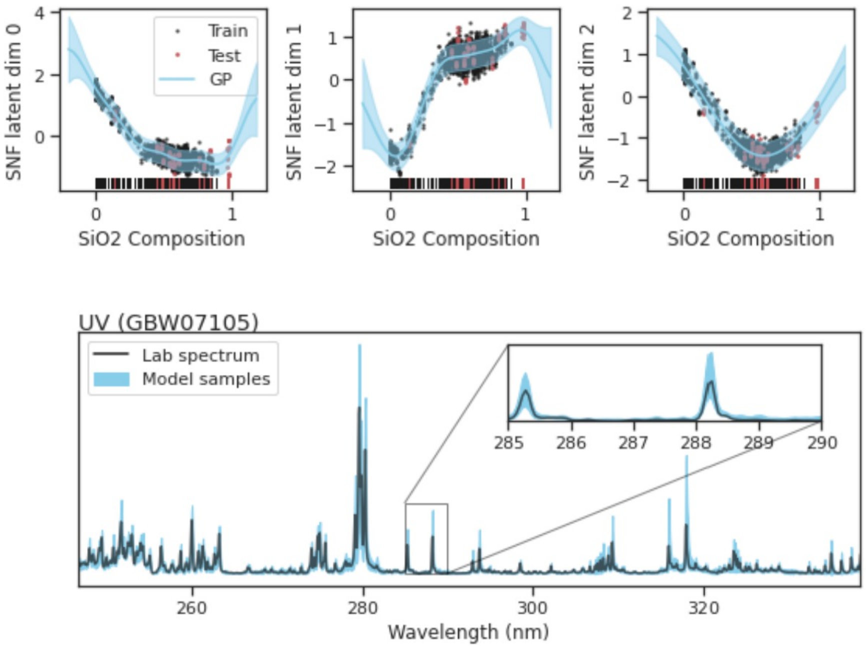
How to represent complex and high-dimensional spectral features?

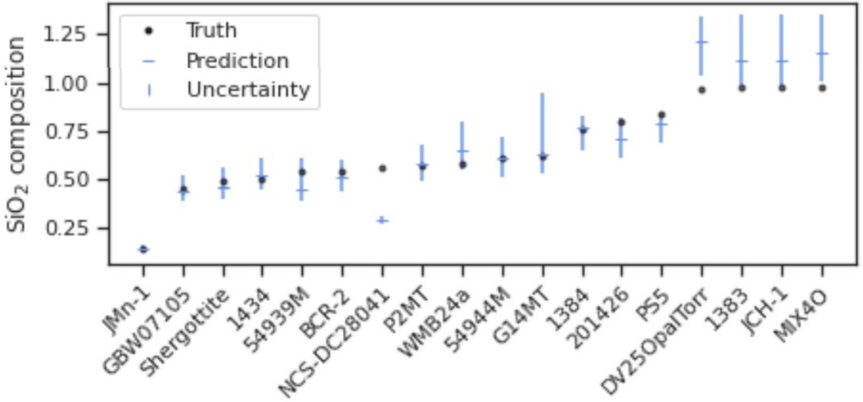
We combine dimension reduction, normalizing flows, and Gaussian process models.



Short workshop paper: https://arxiv.org/abs/2212.07554

Normalizing flow Gaussian process models reproduce spectral data and provide useful inverse calibration.





[Klein et al 2022 https://arxiv.org/abs/2212.07554]

Outline

- Why Mars rover data?
- Landscape of data analysis
- Leveraging Earth and Mars data
- Multimodal data fusion
- Generative models for spectra
- Frontiers and ongoing work
- Conclusions and open questions

NASA *Curiosity* rover selfie

Frontiers and ongoing work

- Simulation-based inference: fast approximate simulators for LIBS
  - Leverage partially accurate simulators and real data
  - Enable adaptation to different instruments/environments
  - Need to handle nuisance parameters
- Improved multimodal modeling
- Modeling tradeoffs: accuracy/interpretability
- Robust UQ
- Moving beyond composition prediction to incorporate mineralogy
- Physics-based modeling

Outline

- Why Mars rover data?
- Landscape of data analysis
- Leveraging Earth and Mars data
- Multimodal data fusion
- Generative models for spectra
- Frontiers and ongoing work
- Conclusions and open questions

NASA *Curiosity* rover selfie

Los Alamos
NATIONAL LABORATORY

Conclusions and open questions

- Progress: modern modeling, interpretability, UQ
- Challenge: incorporating available data
  - Different instruments
  - Simulation/experimental
  - Different environmental conditions
  - Different modalities
- Using all available information for coordinated prediction
  - Is composition consistent with mineralogy?
- Discovering new signatures