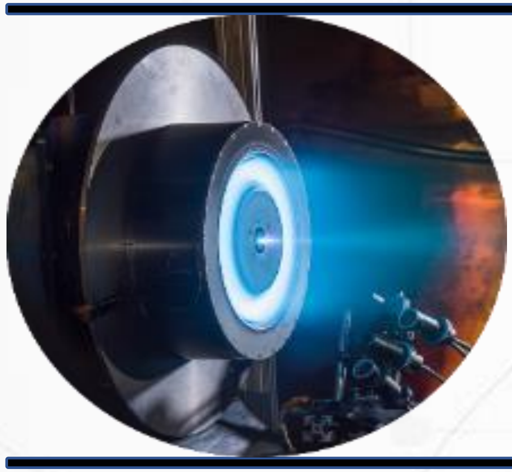
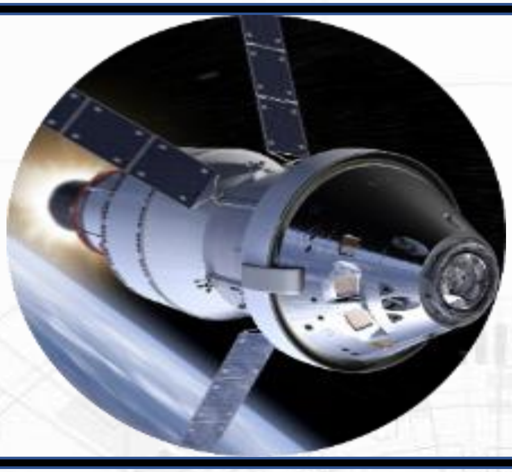
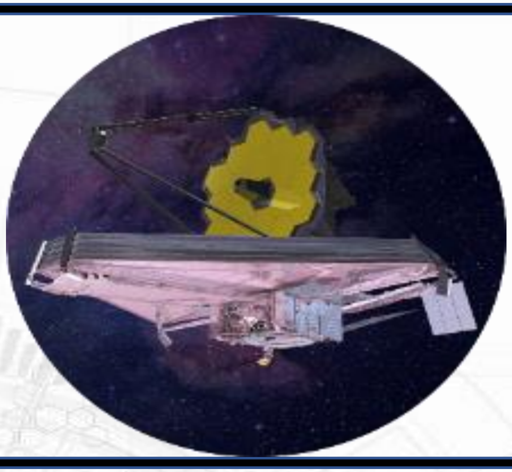
FY21 CIF/IRAD Final Review
November 2-4, 9, & 10, 2021

National Aeronautics and
Space Administration


Title: Building Trust in Artificial Intelligence: Fusing Interpretable Machine Learning and Uncertainty Quantification
PI Name: Geoffrey Bomarito (D309, RD)
CO-I Name: Patrick Leser (D309, RD)
Partners: Jacob Hochhalter & Nolan Strauss (University of Utah)
Intern: Bohan Li (University of Arizona)
Date: 11/03/2021

Why Does NASA Need to Trust Machine Learning (ML) and Artificial Intelligence (AI)?

Problem:
"...entrusting important decisions to a system that cannot explain itself presents obvious dangers." – A. Adadi and M. Berrada

"[Black-box ML] is problematic because it can adversely affect the understanding, trust, and management of ML algorithms" – A. Seeliger, M. Pfaff, and H. Krcmar

Mission Need:
- Autonomy (STIP 5 - Trusted Autonomy, ARMD Thrust 6 - assured autonomy)
- Certifiability (STIP 1 - certification, ARMD Thrust Ultra-Efficient Commercial Vehicles)
- Low risk tolerance -> need to trust models!

- **Goal:**
  - Investigate recent developments in explainable ML/AI and learn how to start the process of trust building.
  - Infuse this knowledge into NASA.

Approach: Background

- Explanations vs. Experience
  - Asking “Why?”
  - Faster trust building

- Interpretable Machine Learning (IML)

“Why Are We Using Black Box Models in AI when We Don’t Need To?” - Rudin and Radin. 2019

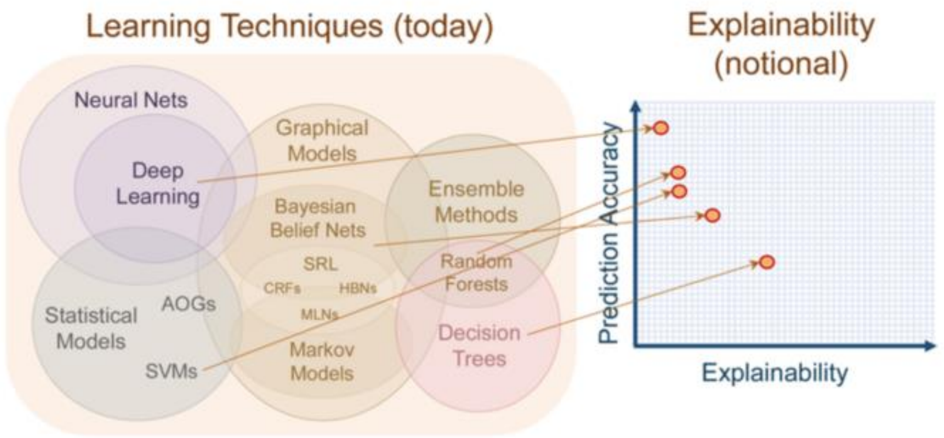
- Uncertainty Quantification (UQ)

“Treat[] confidence as complementary to explanation…” - Bhatt et al. 2020

Motivation for Explanations

Justification	Control
Improvement	Discovery

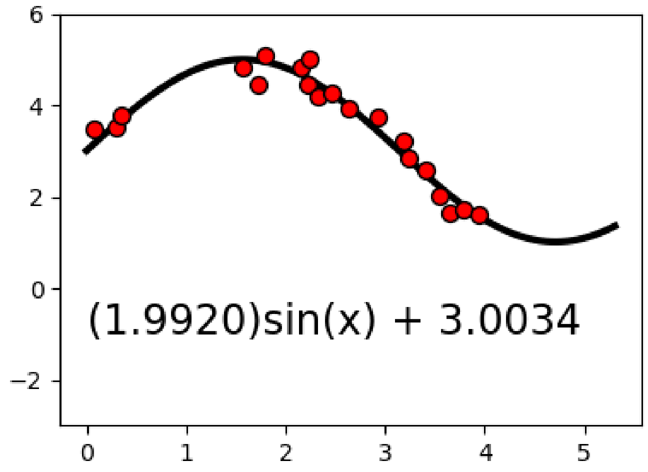
Common Misconception of Accuracy-interpretability Tradeoff



Approach: Fusing Interpretable ML and UQ

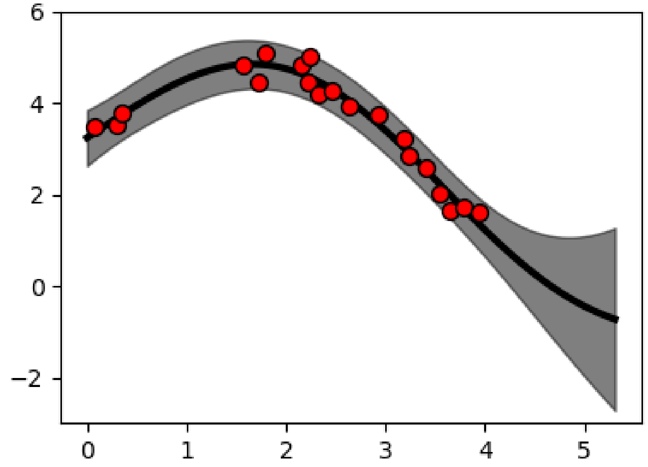
State of the Art

Interpretable ML



e.g.
Symbolic Regression
Decision Trees

black-box ML with UQ

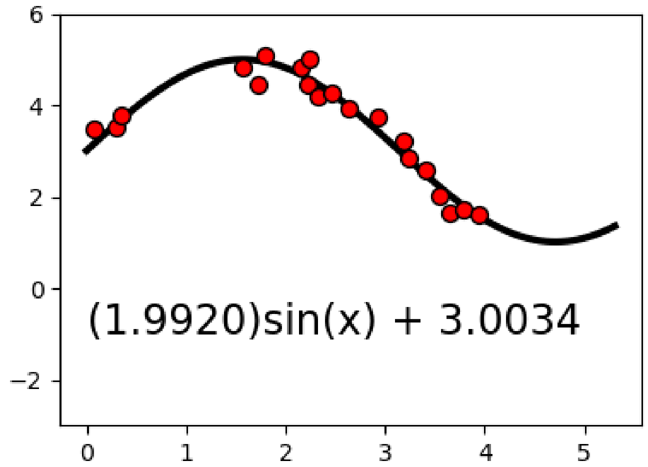


e.g.
Gaussian Processes
Bayesian Neural Nets

Approach: Fusing Interpretable ML and UQ

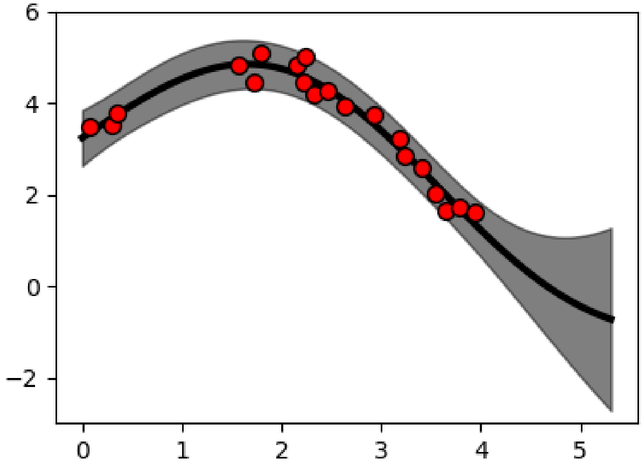
State of the Art

Interpretable ML



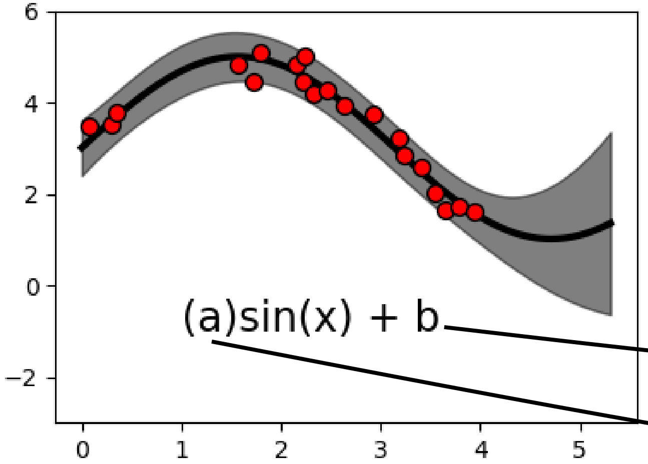
e.g.
Symbolic Regression
Decision Trees

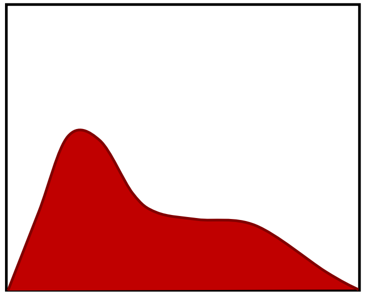
black-box ML with UQ



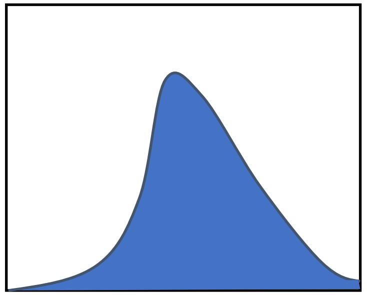
e.g.
Gaussian Processes
Bayesian Neural Nets

Interpretable ML + UQ





b



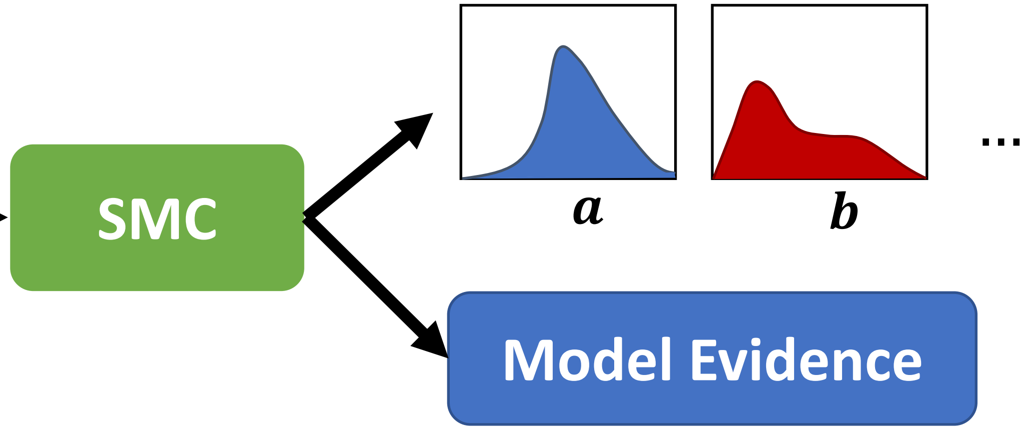
a

Approach: Fusing Interpretable ML and UQ

- Interpretable ML: Symbolic Regression (SR)
  - Finds an equation to best fit data
  - Tests many arbitrarily complex equations
- Auto UQ: Sequential Monte Carlo (SMC)
  - Estimates uncertainty in model parameters and noise level
  - Estimates model evidence
  - Does this through numerous model evaluations

$$f(x) = a\sin(bx) + ab$$
$$f(x) = a$$
$$f(x) = ax + bx^2 + \frac{c}{cos(dx)}$$
$$\vdots$$

$$f(x) = a\sin(bx) + ab \ldots \rightarrow$$



Approach: FY21 Focus

- Interpretable ML: Symbolic Regression (SR)
  - Finds an equation to best fit data
  - Tests many arbitrarily complex equations
- Auto UQ: Sequential Monte Carlo (SMC)
  - Estimates uncertainty in model parameters and noise level
  - Estimates model evidence
  - Does this through numerous model evaluations

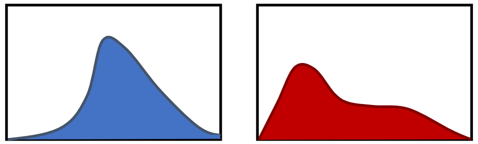
Project Goals (FY21)

1. *Increase the robustness of automated SMC*
2. *Increase the computational efficiency of SR with integrated SMC*

Results: Robustness of Automated SMC

$$\boldsymbol{f(x) = a\sin(bx) + ab} \ldots \rightarrow \text{SMC}$$

SMC → **Model Evidence**



$\boldsymbol{a}$ $\boldsymbol{b}$ …

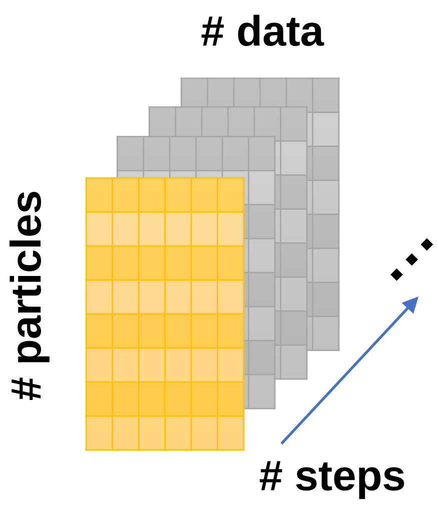
Hurdles	Solutions
• No knowledge of parameter values →	Fractional Bayes Factor
• Multi-modal fits →	Multi-start parameter proposals
• Hand tuned hyperparameters →	Automation, heuristic optimization
• Noisy estimates →	Model re-evaluation

Robustness on test set 46% → 99%

Results: Computational Efficiency

$$f(x) = a\sin(bx)$$

# data	x	# particles	x	# steps	=	evaluations per equation
~100		~1,000		~100		**~10,000,000**

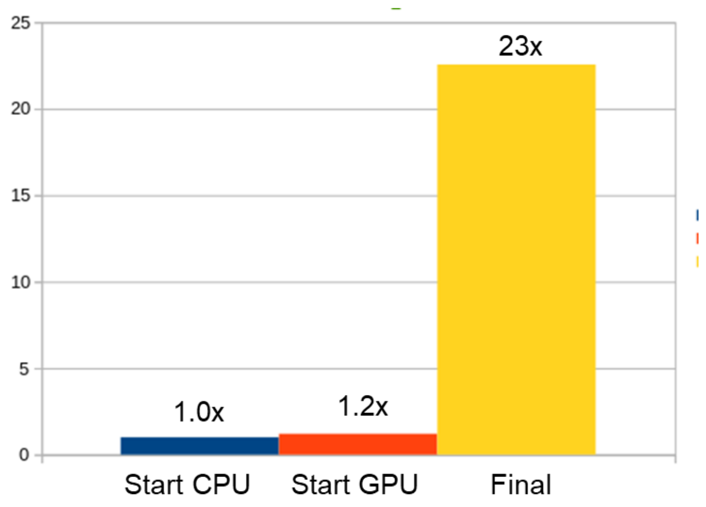


Hurdles

- Large number of evaluations for each equation
- Many equations

Solutions

- Leverage independent nature of evaluations (vectorization, parallelization)
- GPU implementation (Intern, NASA-NVidia GPU Hackathon)



23x performance speedup

Accomplishments and Issues

- Overguide project awarded mid-FY → Rescoped
  - 0.3 FTE → 0 FTE
  - 12 Mo → 6 Mo

- Milestones
  - Software development: SMC embedded in SR
  - Reach an acceptable level of robustness in automated SMC (99%)
  - Initial GPU implementation (23x performance increase)

Highlights

- Participation in NASA/NVidia GPU hackathon
- Partnership with University of Utah
- Conference presentation
  - *Garbrecht et al. (2021) "Bayesian Genetic Programming Based Symbolic Regression with Preferential Search". 16th U.S. National Congress on Computational Mechanics (USNCCM)*
- Invited presentation at AFRL
  - Bomarito (2021) "Symbolic Regression with Genetic Programming"

Future Work

- FY22 IRAD
  - Complete proof-of-concept (journal paper planned)
  - Application to linking material properties to microstructural information (conference presentation/paper planned)
- LTC/OCT Support and engagement
  - Agency-wide AI/ML Collaboration

Funding

0 FTE

Procurement funds available	68.4k
External Collaborator (University of Utah)	45.3k
Summer Intern	13.0k
Fall Intern (partial)	8.4k
Total	66.7k

Building trust in artificial intelligence: fusing interpretable machine learning and uncertainty quantification

National Aeronautics and Space Administration

FY21 CIF IRAD Final Review

PI: Geoffrey Bomartio   Co-I: Patrick Leser

Annual Funding: $68.4K

Problem and Impact to NASA

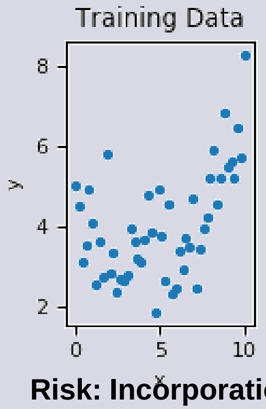
- Autonomy and certification are both areas of interest to NASA that can benefit significantly from artificial intelligence (AI) and machine learning (ML).
- However, they are both high-risk applications that *require trust* in the models being developed, and "trust through experience" is often not acceptable
- *Interpretable machine learning* (IML) could provide a solution to this problem by providing "trust through understanding"
- IML is an active area of research, but the most promising areas still lack robust methods for *quantifying uncertainty* in the presence of noisy data, as will typically be the case in real-world scenarios of interest to NASA

Goal/Objective

- Pursue the building of trust though better *understanding of our AI/ML* models
- Develop an interpretable AI/ML solution that can convey
  - How it makes predictions
  - How uncertain those predictions may be

Innovation

- Augment a *symbolic regression* (SR) code that learns interpretable models in the form of analytical equations with the ability to quantify uncertainty in model parameters (i.e., constants in the learned equations)
- Equation fitness (a measure of how well the model fits the data) will be based on *marginal log likelihood* (MLL) a Bayesian measure that considers uncertainty in the model and data
- MLL will be estimated using *sequential Monte Carlo* (SMC)

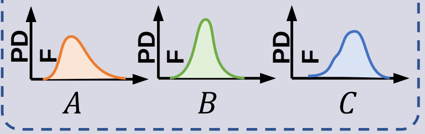


SR:

$$y = 0.12(x - 4.1)^2 + 3.01$$

SR+SMC:

$$y = A(x - B)^2 + C$$



Risk: Incorporation of computationally intensive uncertainty quantification (UQ) methods may cause intractable learning*

**Preliminary results suggest this is not the case*

Mitigation: Existing NASA codes benefit from development even if proposal deemed infeasible.

Technical Approach

- Combine existing SR and SMC NASA codes to rapidly develop proof-of-concept
- Study SR+SMC to determine pros/cons and evaluate potential for NASA benefit

New Project

Start TRL/End TRL: 2/2

LSTIP G.C.: 3) Revolutionary Airspace Solution for Future Dense, Heterogeneous Operations

Partners: University of Utah

Primary TX: 10) Autonomous Systems

Key Milestones/Deliverables

Fuse IML with UQ to produce models that are interpretable and make predictions that are robust to uncertainty

1. Software development: SMC embedded in SR
2. Reach an acceptable level of robustness in automated SMC (99%)
3. Initial GPU implementation (23x performance increase)

Questions?