Uncertainty Quantification for Interpretable Constitutive Models using Genetic Programming based Symbolic Regression

Presenting Author: Donovan Birky*†

Co-Authors: Karl Garbrecht*†, John Emery†, Coleman Alleman†, Brian Lester†, Geoffrey Bomarito‡ Michael Buche†, Jacob Hochhalter*

*University of Utah, †Sandia National Laboratories, ‡NASA

July 26th, 2023
USNCCM

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

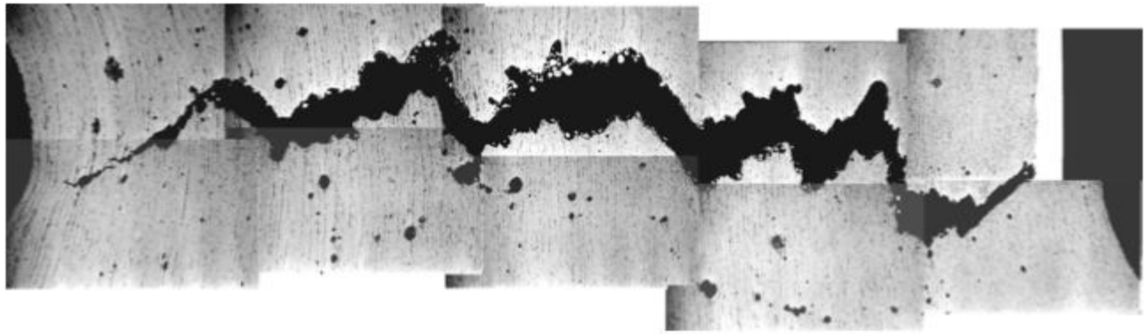
1. Background
2. Genetic Programming based Symbolic Regression (GPSR)
3. GPSR for Implicit Equations
4. Uncertainty Quantification (UQ) using Bayesian Statistics
5. GPSR for Implicit Equations with UQ
6. Verification Tests
7. Application to Constitutive Modeling

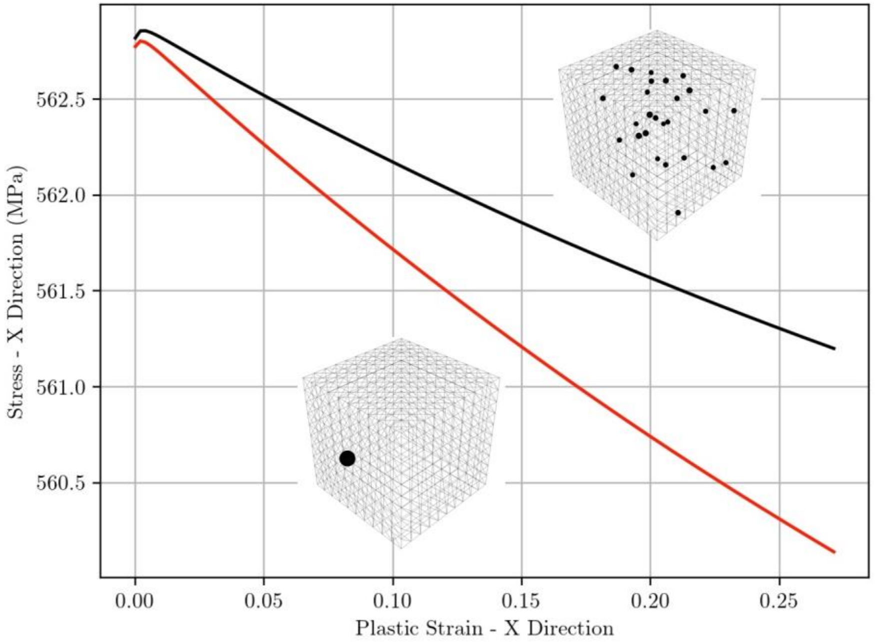
Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

Background

- Traditional material models often require idealizations of complex microstructures to develop closed-form solutions, e.g., Gurson [1], Cocks-Ashby [2]
- Relaxation of these assumptions can introduce stochastic variables
  - Random inclusions or voids
- Models can represent these stochastic responses through UQ

[3]



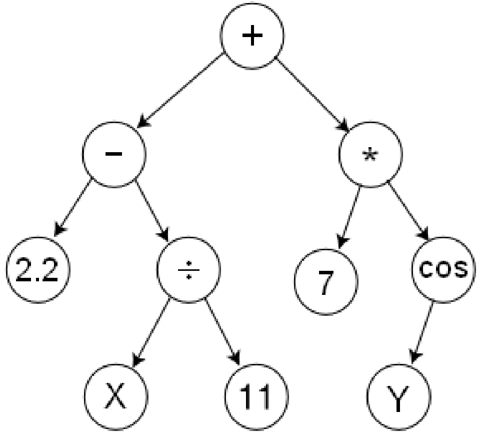


[1] A. L. Gurson. Continuum theory of ductile rupture by void nucleation and growth: Part i—yield criteria and flow rules for porous ductile media.
[2] F. Cocks and M. F. Ashby, "Creep fracture by coupled power-law creep and diffusion under multiaxial stress," no. September, pp. 395–402, 1980.
[3] A. A. Benzerga, J.-B. Leblond, A. Needleman, and V. Tvergaard. Ductile failure modeling. International Journal of Fracture, 201(1):29–80, 2016. doi: 10.1007/s10704-016-0142-6.

GPSR

- Symbolic regression searches space of known equations via combinations of variables and weights
- Genetic programming evolves equations based on fitness with data

[3]

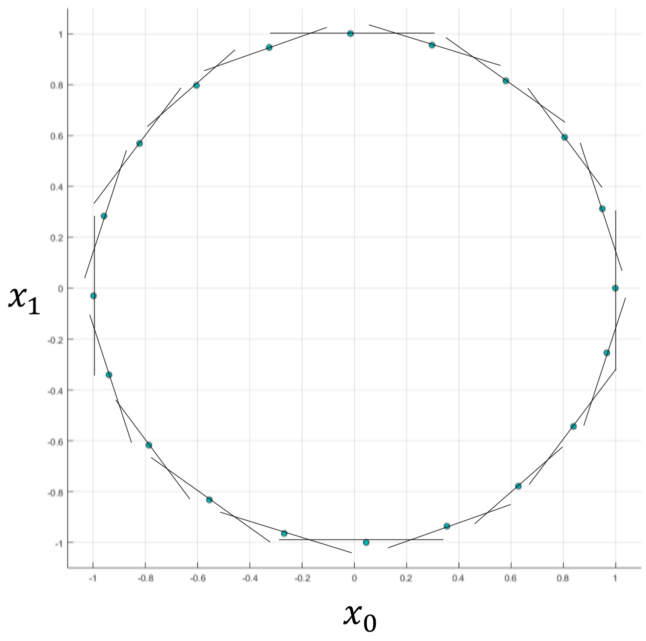


$$\left(2.2 - \left(\frac{X}{11}\right)\right) + \left(7 * \cos(Y)\right)$$

[3] G. Bomarito, T. Townshend, J. Hochhalter, E. Adams, K. Esham, and D. Vera, “Bingo,” 2018. [Online]. Available: https://github.com/nasa/bingo.

GPSR for Implicit Equations

- Implicit models of the form $f(X) = C$ prevent the use of traditional error metrics due to arbitrary solutions, e.g., $x_1 - x_1 + C = C$
- Instead, fitness is defined as the difference in partial derivatives between the model and data



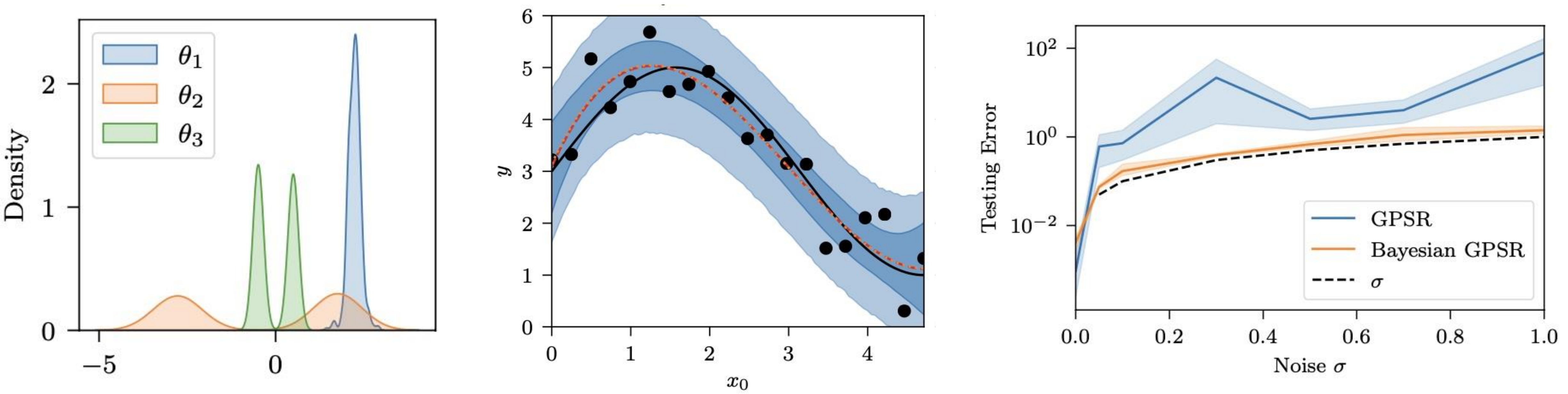
Implicit Fitness Function

$$\Phi(f, X) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\sum_{j=1}^{P}\frac{\partial f}{\partial x_j}\frac{\Delta x_j}{\Delta t}}{\sum_{k=1}^{P}\left|\frac{\partial f}{\partial x_k}\frac{\Delta x_k}{\Delta t}\right|}\right|$$ [4]

[4] G. Bomarito, T. Townsend, K. Stewart, K. Esham, J. Emery, and J. Hochhalter. Development of interpretable, data-driven plasticity models with symbolic regression.

UQ for Explicit GPSR

- Bayes' theorem gives the posterior probability of parameters, θ, given observations, D, and model, f
- Compute marginal likelihood using sequential Monte Carlo (SMC)
- Has been shown to decrease model complexity and overfitting [5]



[5] G. Bomarito, P. Leser, N. Strauss, K. Garbrecht, and J. Hochhalter. Bayesian model selection for reducing bloat and overfitting in genetic programming for symbolic regression.
[6] Anthony O'Hagan. Fractional bayes factors for model comparison.

UQ for Implicit GPSR

- Multi-objective approach using implicit fitness and computed marginal likelihood

Algorithm 1 Multi-objective Implicit SMC-GPSR Algorithm

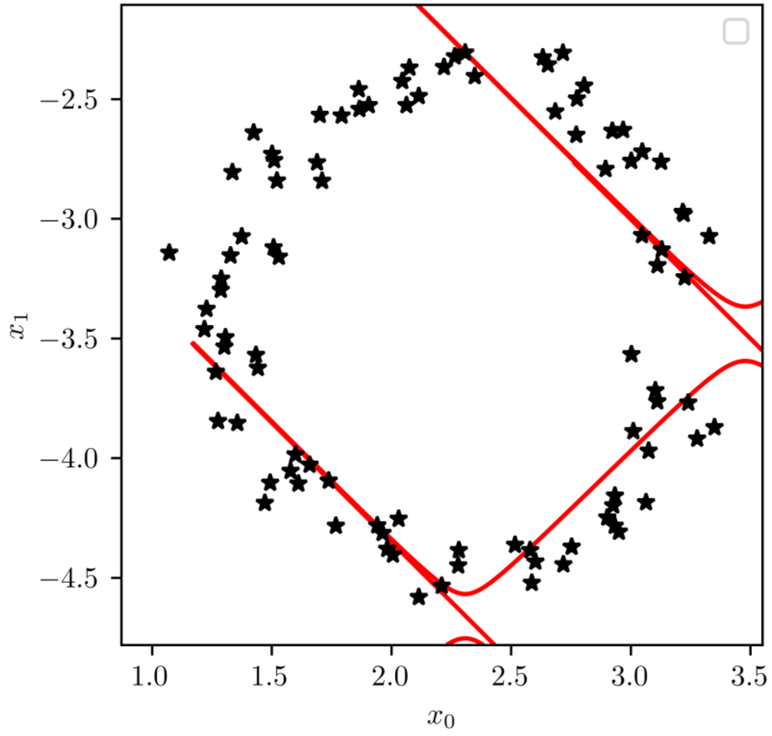
1. Propose a model, $f(X)$.
2. Calculate model fitness, $\Phi(f(X))$, based on implicit fitness.
3. Calculate marginal likelihood, $L = \pi(f|X)$ using SMC with $y = C$, where C is the constant for the proposed implicit model, $f(X) = C$.
4. Normalize the marginal likelihood, $\bar{L} \in \{0, 1\}$.
5. Calculate final model fitness, $F(f(X)) = \Phi + \bar{L}$.

Test Results

- Circle with added noise, $\mathcal{N}(0, 0.1)$ - $f = (x_0 - 2.3)^2 + (x_1 + 3.4)^2 = 1$

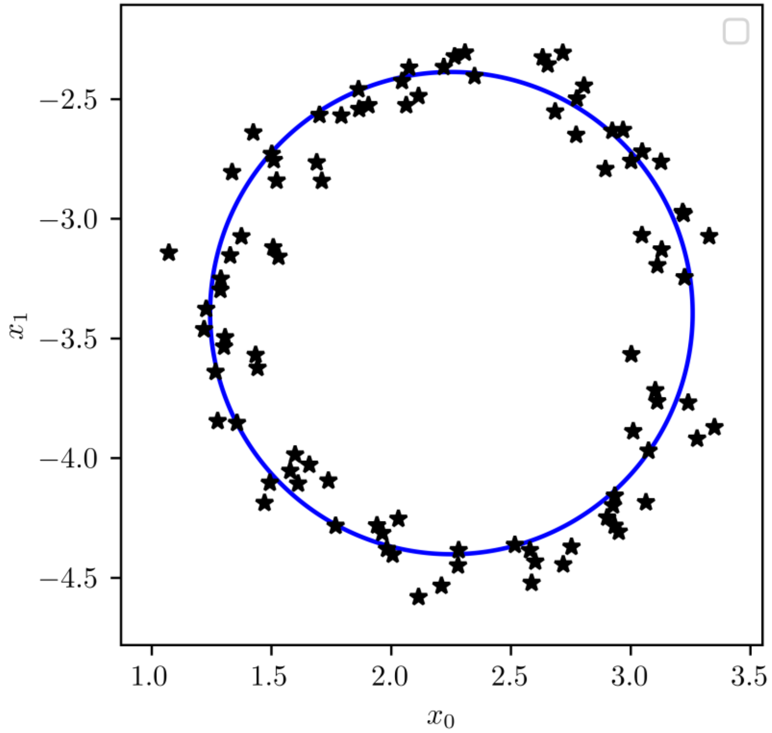
Implicit Algorithm

$$f = 268.94x_0 + \frac{x_0}{x_0 + x_1 + 2.35} + \frac{x_0}{x_0 + x_1} - 268.94x_1 = 0$$



Implicit + SMC Algorithm

$$f = (x_0 - 2.245)^2 + (x_1 + 3.377)^2 = 1$$



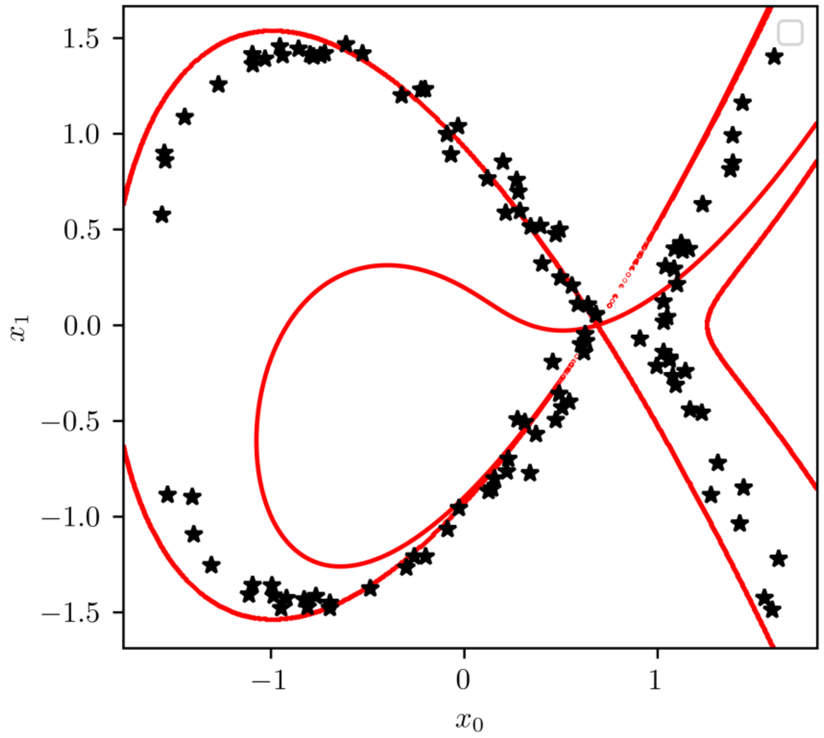
[7] M. Schmidt and H. Lipson. Symbolic regression of implicit equations. In Genetic Programming Theory and Practice VII, pages 73–85. Springer, 2010. ISBN 9781441916266. doi: 10.1007/978- 1-4419-1626-6.

Test Results

- Elliptic curve with added noise, $\mathcal{N}(0, 0.05)$ $f = x_0^3 - 2x_0 - x_1^2 + 1 = 0$
  -

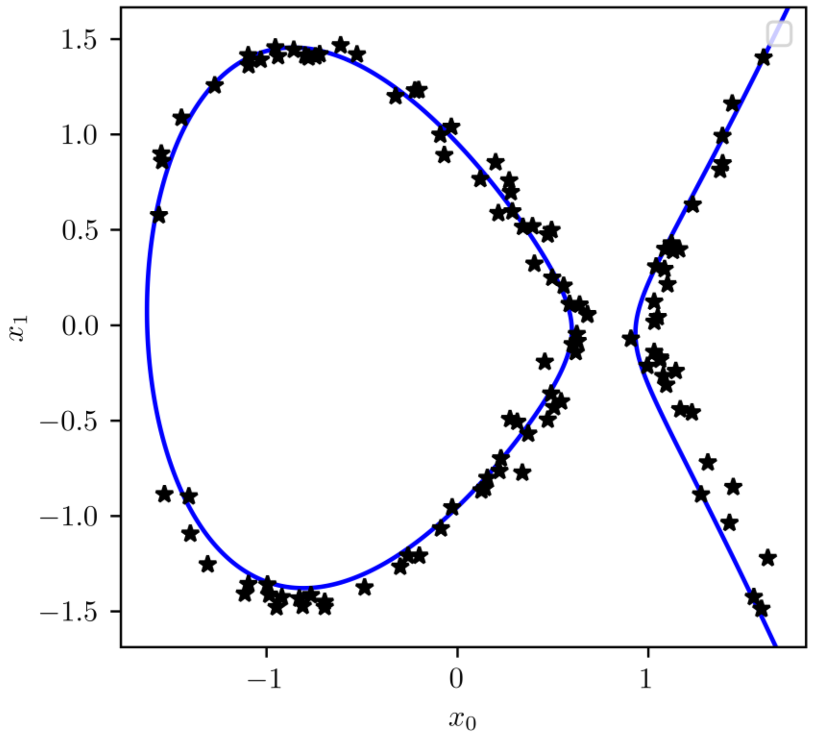
Implicit Algorithm

$$f = \frac{-x_0 + \frac{0.6x_1^2}{(x_0-0.7)^2} + 0.7}{\frac{x_0}{-x_0+\frac{0.6x_1^2}{(x_0-0.7)^2}+0.63} - 1.89} = 0$$



Implicit + SMC Algorithm

$$f = 0.92x_0^3 - 1.9x_0 - x_1^2 + 1 = 0$$



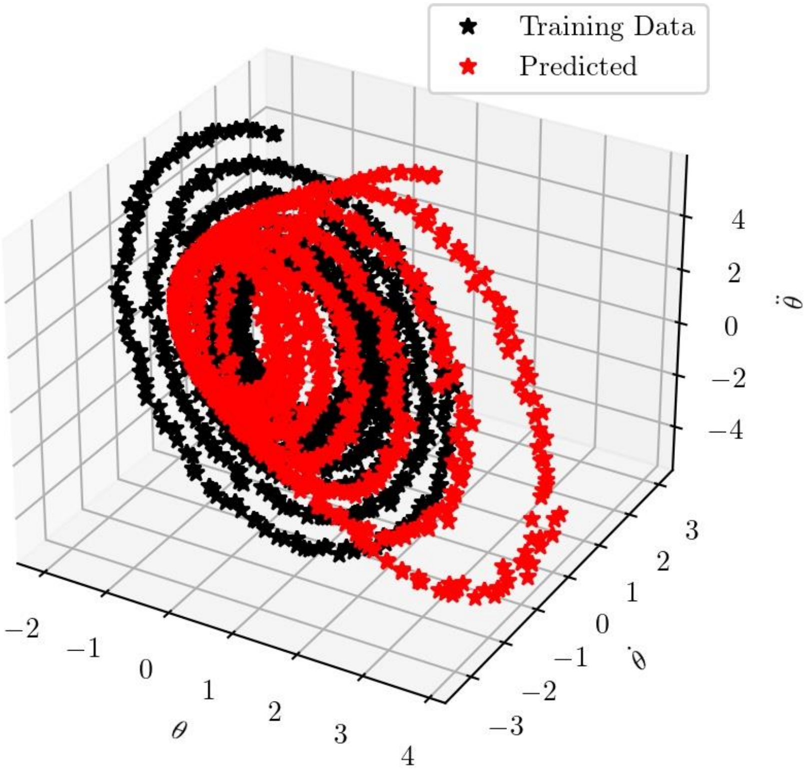
[7] M. Schmidt and H. Lipson. Symbolic regression of implicit equations. In Genetic Programming Theory and Practice VII, pages 73–85. Springer, 2010. ISBN 9781441916266. doi: 10.1007/978- 1-4419-1626-6.

Test Results

- Harmonic oscillator with added noise, $\mathcal{N}(0, 0.05)$ - $f = \ddot{\theta} - 0.1\dot{\theta} + 3\theta = 0$

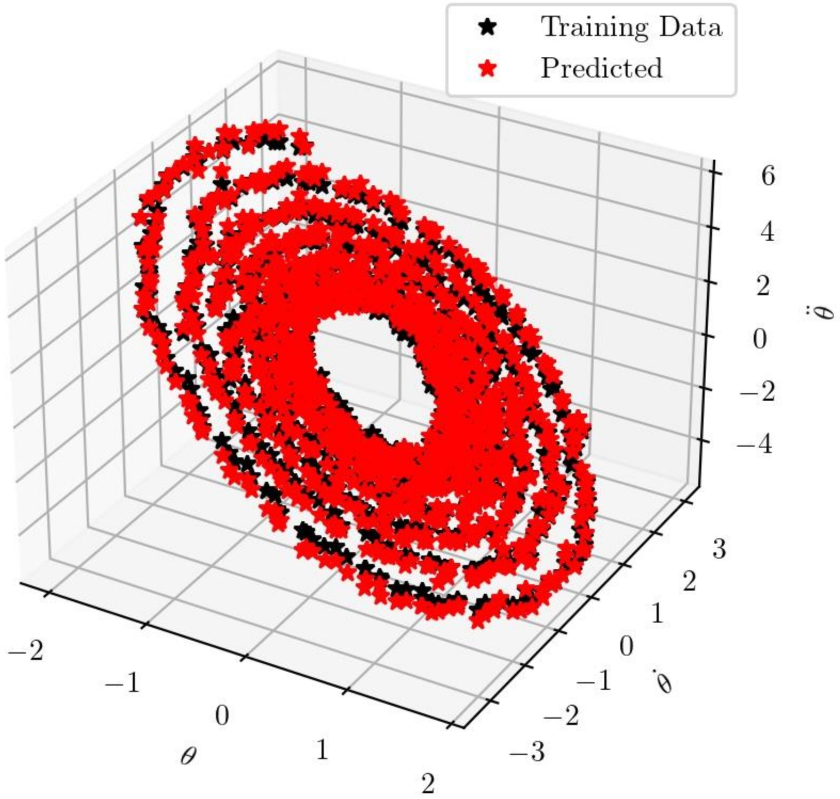
Implicit Algorithm

$$f = 678.47\theta + 226.16\dot{\theta}^2 + 77.86\ddot{\theta}^2 - 227.16\ddot{\theta} - 1026.27 = 0$$



Implicit + SMC Algorithm

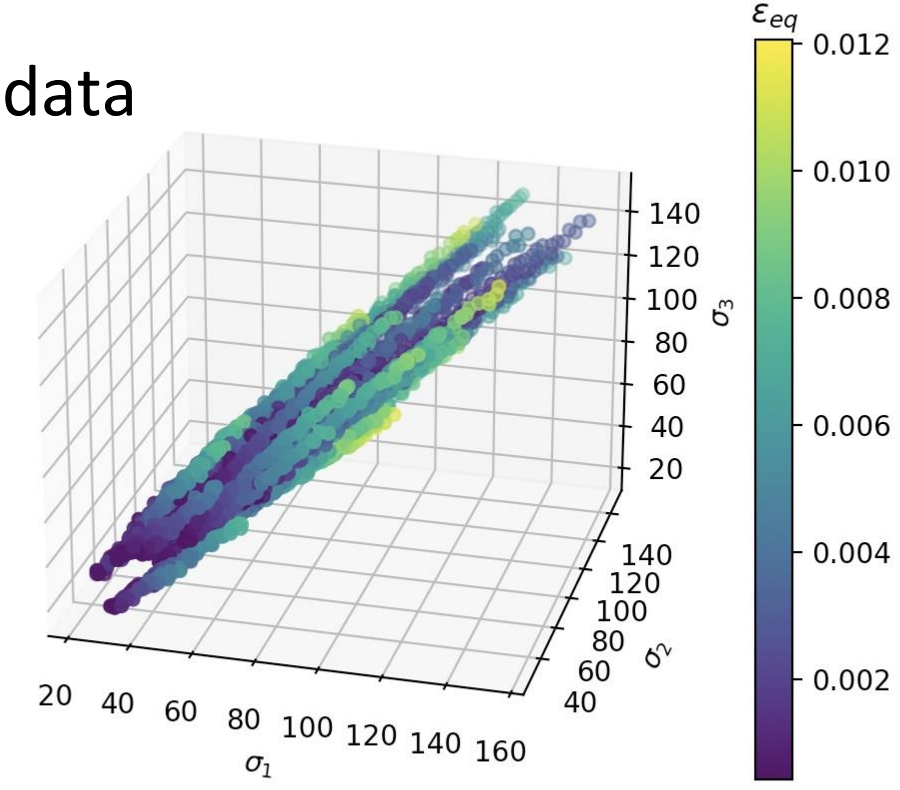
$$f = \ddot{\theta} - 0.11\dot{\theta} + 2.9\theta = 0$$



[7] M. Schmidt and H. Lipson. Symbolic regression of implicit equations. In Genetic Programming Theory and Practice VII, pages 73–85. Springer, 2010. ISBN 9781441916266. doi: 10.1007/978- 1-4419-1626-6.

Application to Constitutive Modeling

- Testing on von Mises surface first demonstrated by Bomarito et al. [4]
  - Data generated from a FE modeled RVE
  - von Mises material with hardening
- Adding Gaussian noise, $\mathcal{N}(0, 0.1)$, scaled to data



Target Equations

$$f = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2 - 39600\epsilon_{eq} - 1960200\epsilon_{eq}^2 = 0$$

$$\dot{\epsilon}_{ep} = \sqrt{\frac{2}{9}\left[(\dot{\epsilon}_{p1} - \dot{\epsilon}_{p2})^2 + (\dot{\epsilon}_{p2} - \dot{\epsilon}_{p3})^2 + (\dot{\epsilon}_{p1} - \dot{\epsilon}_{p3})^2\right]}$$

[4] G. Bomarito, T. Townsend, K. Stewart, K. Esham, J. Emery, and J. Hochhalter. Development of interpretable, data-driven plasticity models with symbolic regression.

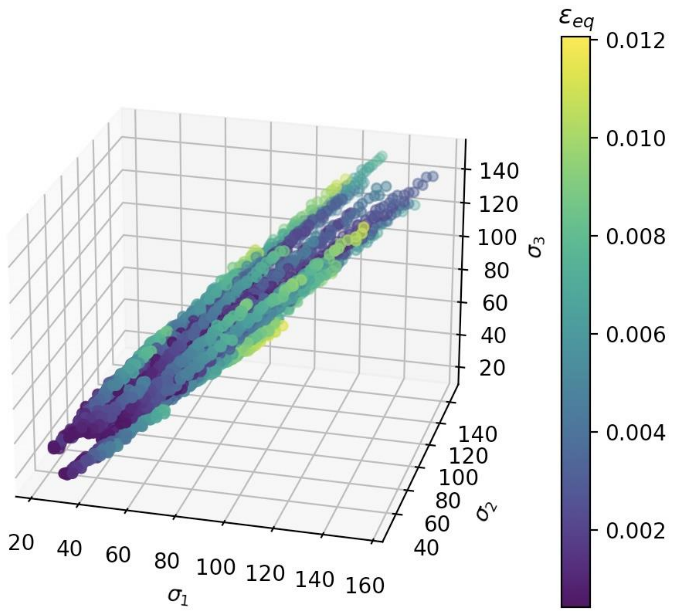
Preliminary Results

- Increased dataset size and larger models lead to much slower training times than the test cases
- Testing with parallel evaluations, fitness predictors, pytorch c++ implementations

<u>Target Equations</u>

$$f = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2 - 39600\epsilon_{eq} - 1960200\epsilon_{eq}^2 = 0$$

$$\dot{\epsilon}_{ep} = \sqrt{\frac{2}{9}\left[\left(\dot{\epsilon}_{p1} - \dot{\epsilon}_{p2}\right)^2 + \left(\dot{\epsilon}_{p2} - \dot{\epsilon}_{p3}\right)^2 + \left(\dot{\epsilon}_{p1} - \dot{\epsilon}_{p3}\right)^2\right]}$$



Conclusions

- Multi-objective approach enables UQ for implicit models, enabling learning of yield surface models
- SMC approach enables algorithm to fit noisy data better than conventional implicit approach in test cases
- Future work
    - Speeding up the algorithm for larger datasets/models
    - Training on datasets generated from random void representative volume elements (RVEs)