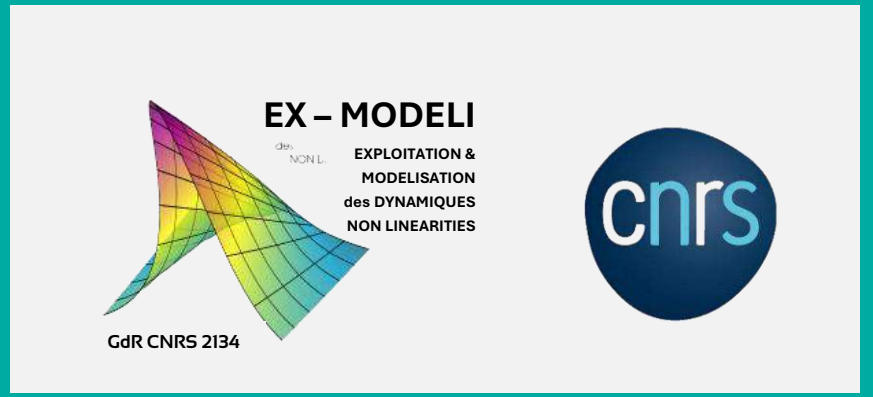
LEARNING MODEL DISCREPANCY USING SPARSE BAYESIAN POLYNOMIAL NARX

Eyoel ATNAFE

Rafael TELOLI, Emeline SADOULET-REBOUL, Scott COGAN



Octobre, 2026

Content

1 Context

2 State-of-the-art

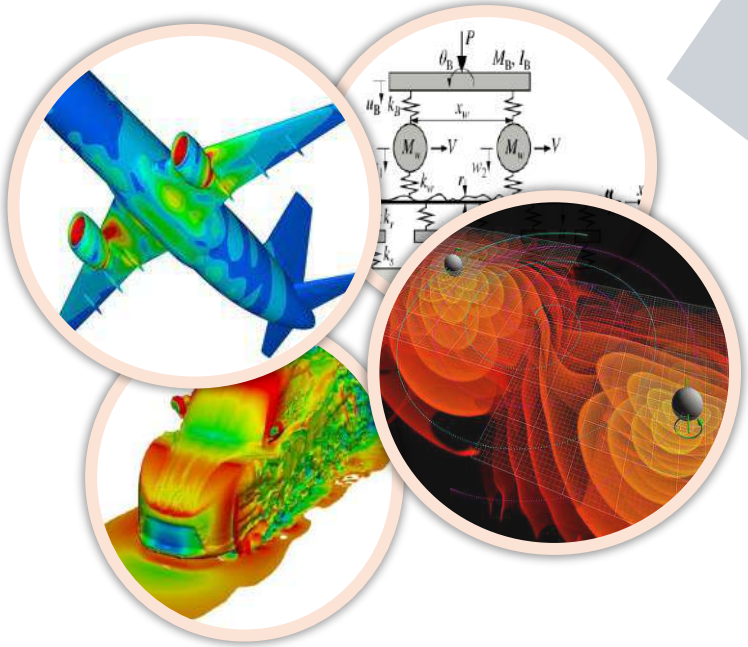
3 Proposed Methodology

4 Result and Discussion

5 Conclusion

- Linearized
- Limited Multiphysics interactions
- Required calibration
- Required understanding of detailed physics that are sometimes unknown

Physics-based model





Physical systems

- Nonlinear
- Hidden parameters
- Difficult to fully observe
- Multiphysics interaction
- Lack of knowledge

Mismatch

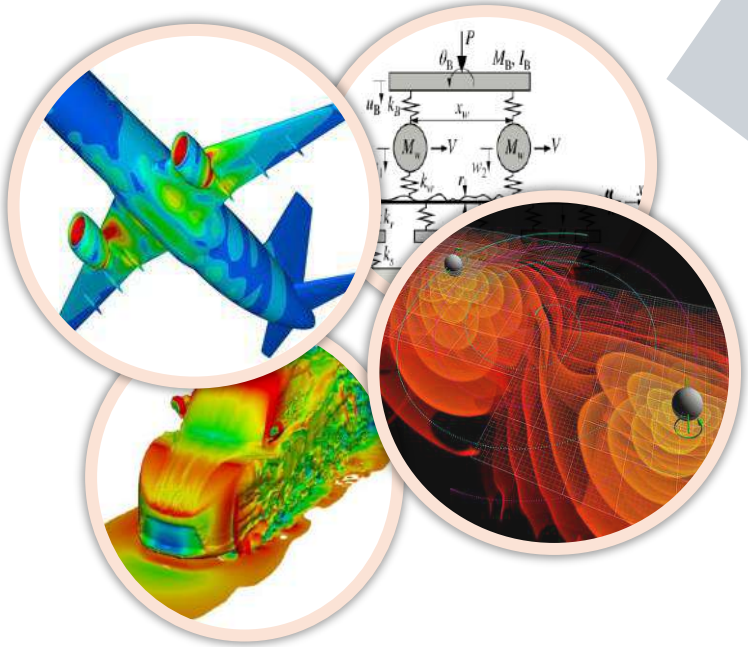
Model Discrepancy

A.k.a
Model Error
Model bias
Model Inadequecy



Physical systems

Physics-based model



Model Discrepancy

Where does this error come from?

Where does this error come from?

- Parametric Uncertainty
- **Model Inadequacy**
- Parametric variability
- Residual variability
- Code error
- Observation error

[1] Kennedy, Marc C., and Anthony O'Hagan, (2001).

θ – System parameters
x – State variables

Where does this error come from?

Objective

Build a discrepancy framework that
- ✓ "Captures" the missing physics
- ✓ Correct existing model
- ✓ Quantify Uncertainty

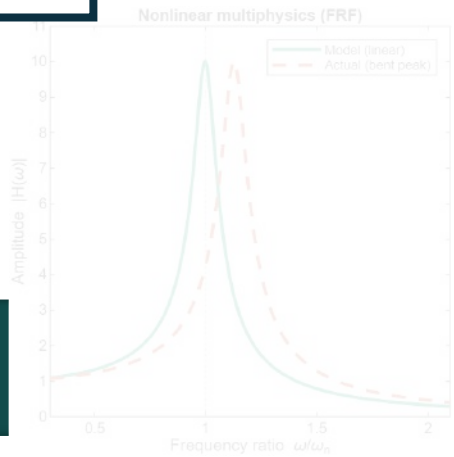
Model In

$+\delta(\phi, x) + \epsilon$

Partially Known Structure

Missing physics

- ⊗ Stiffening or hardening nonlinearities
- ⊗ Hysteresis Nonlinearities
- ⊗ Unaccounted nonlinear Multiphysics
- ⊗ Nonlinear contact, friction
- ⊗ Unmodeled viscoelasiticity

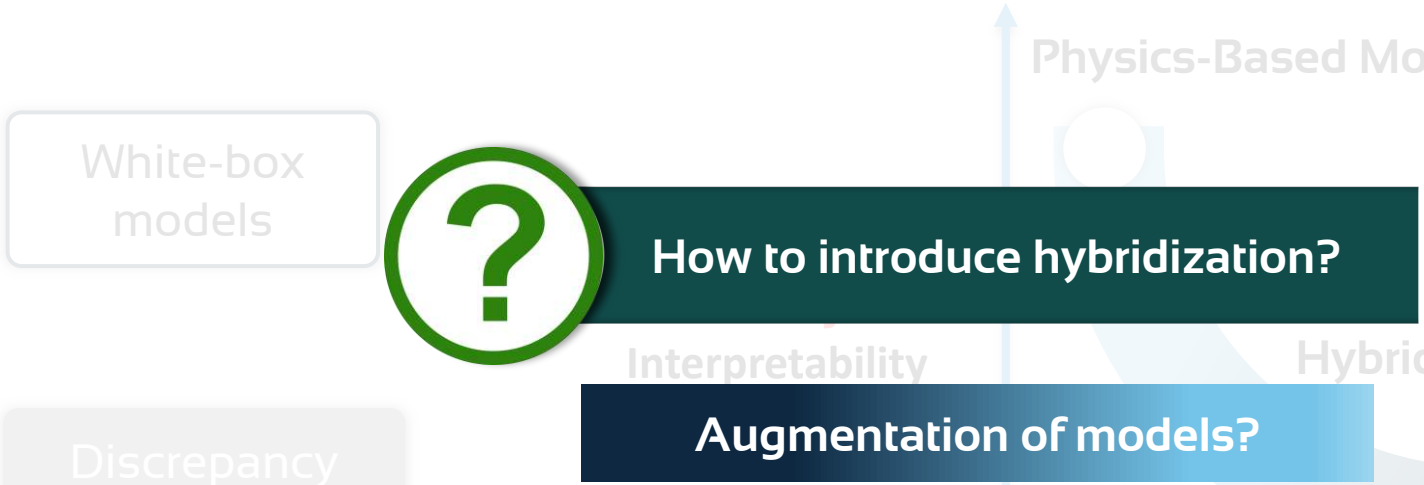


How to model or identify?

[1] Kennedy, Marc C., and Anthony O'Hagan, (2001).

State-of-the-art

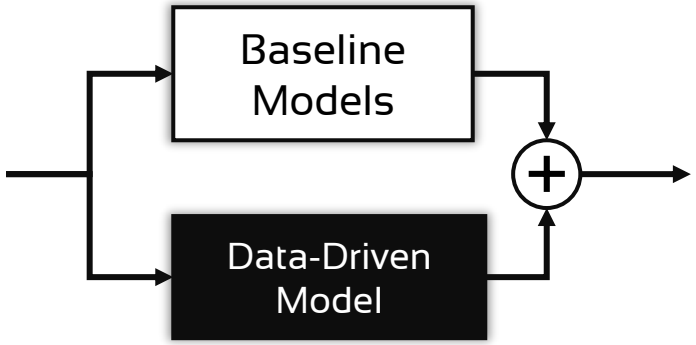
How to model or identify?



[2] Faraji, F., M. Reza, and A. Knoll. (2024)
[3] Maarten S., 2026

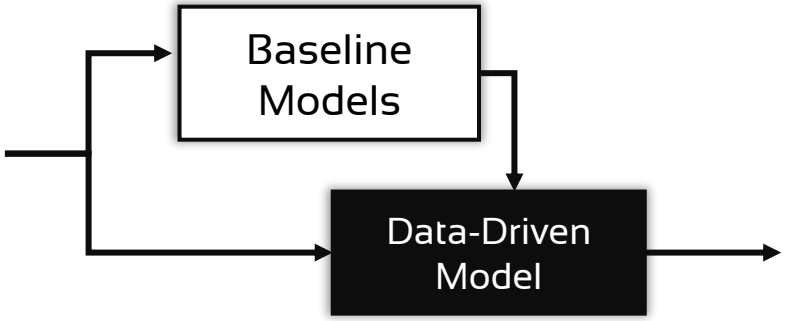
Model Augmentation

Parallel Connection



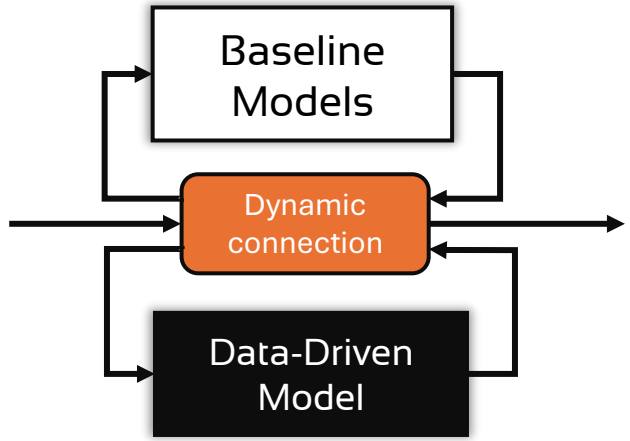
$$y = f(x) + \delta(x)$$

Series Connection



$$y = \delta(f(x))$$

Dynamic Connection



$$y = S(f(x), \delta(x))$$

What is $\delta(x)$

[2] Faraji, F., M. Reza, and A. Knoll. (2024)
[3] Maarten S., 2026
[4] Hoekstra, J. H. 2025
[5] Matthew E. Levine And Andrew M. Stuart, (2020)

System Identification

$\delta(x)$ Data-Driven Discrepancy Model

Representation

Input–Output Models

FIR/NFIR | NARX | NARMAX

e.g.: GP-NARX, NN-NARX

State-Space Models

Linear SS | NL SS | Physics-Structured SS

e.g.: pH, S4, RNN, MCK(2nd order)

Parameterization

Parametric Models

Polynomial | SINDy | MLP | BNNs

e.g.: Bayesian SINDy, pHNN, HNN/LNN

Data-Driven Nonlinear System Identification

Non-Parametric Models

Kernel methods: GP | RVM | SVR

Trees | Ensembles

Focus

- UQ
- Interpretability

[6] Nelles, Oliver. (2001)

System Identification

$\delta(x)$ Data-Driven Discrepancy Model

Input–Output Models

NARX

Parametric Models

SINDy

e.g.: Bayesian SINDy, pHNN, HNN/LNN

Data-Driven Nonlinear System Identification

Focus

UQ

- SINDy is linear in its parameters, so the Bayesian posterior is tractable.
- Sparse priors give uncertainty on both the coefficients and the selected terms.

Interpretability

Sparse Regression permits governing equation discovery and transparency to the regressors.

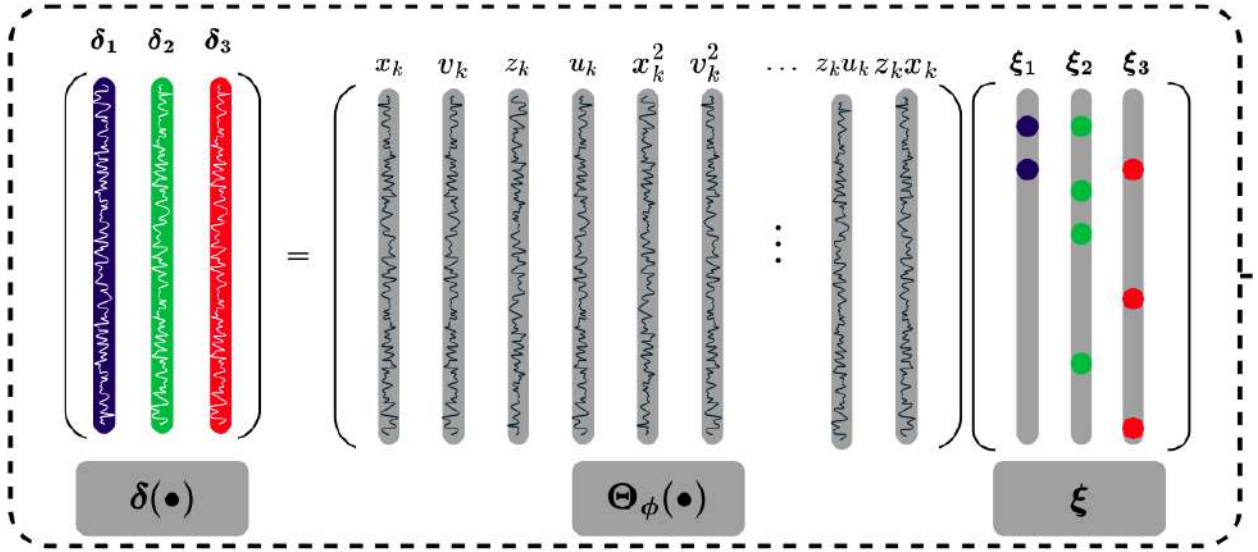
A few active terms keep each one **physically meaningful**.

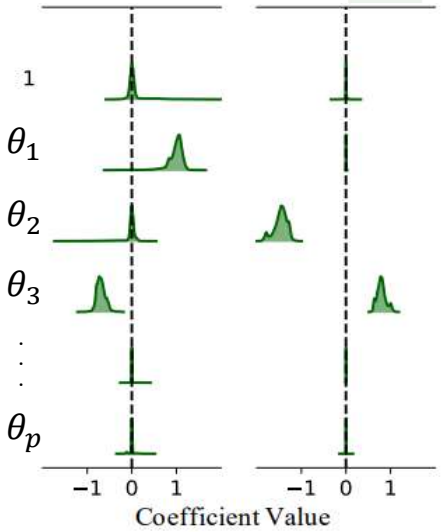
Parsimony means less overfitting and better extrapolation.

Sparse Indetification of Nonlinear Dynamics

Θ – Library of Function
x – State variables

SINDy is a method designed to identify governing equations of dynamical systems directly from data using sparse regression of library of functions.





$$\frac{dx}{dt} = f(x)$$

Dynamical systems

$$\frac{dx}{dt} = f(x) = \Theta(x)\Xi$$

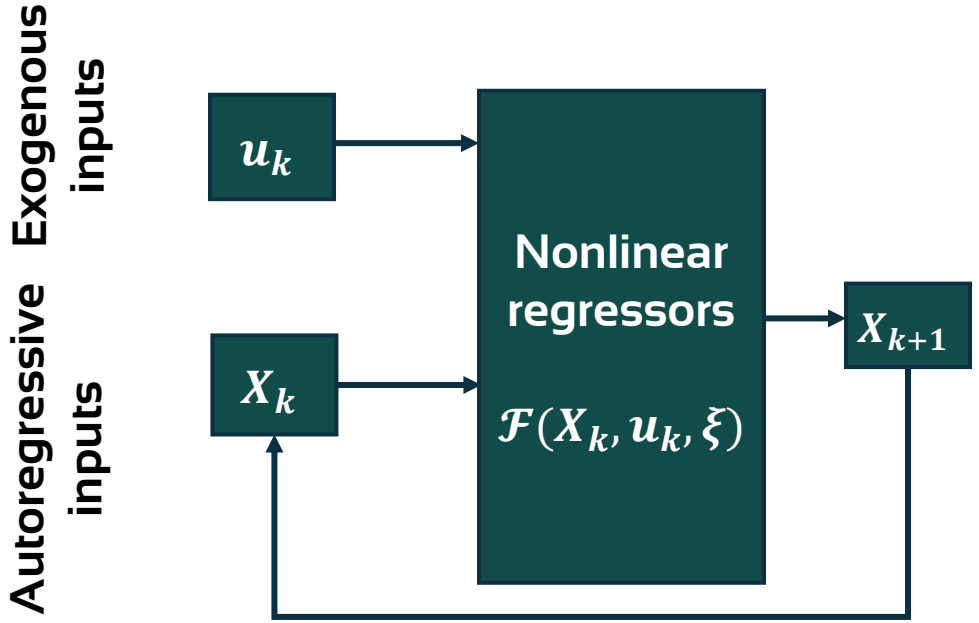
SINDy

$$p(\Xi \mid y_k, \sigma^2) = \frac{p(y|\Xi,\sigma^2)p(\Xi|\sigma^2)}{p(y|\sigma^2)}$$

Bayesian SINDy

[7] Brunton, Steven L.; Kutz, J. Nathan, (2016)

Nonlinear AutoRegressive with eXogenous input



Regressors: $\mathcal{F}(X_k, u_k, \xi)$

- Linear Functions (ARX)
- Neural networks (MLP-NARX)
- Polynomial functions (p-NARX)
- Radial Basis function (r-NARX)
- Gaussian Process (GP-NARX)

$X - [x, v, z]$ – **vector of abstract state**

All hidden states are modelled by the term z

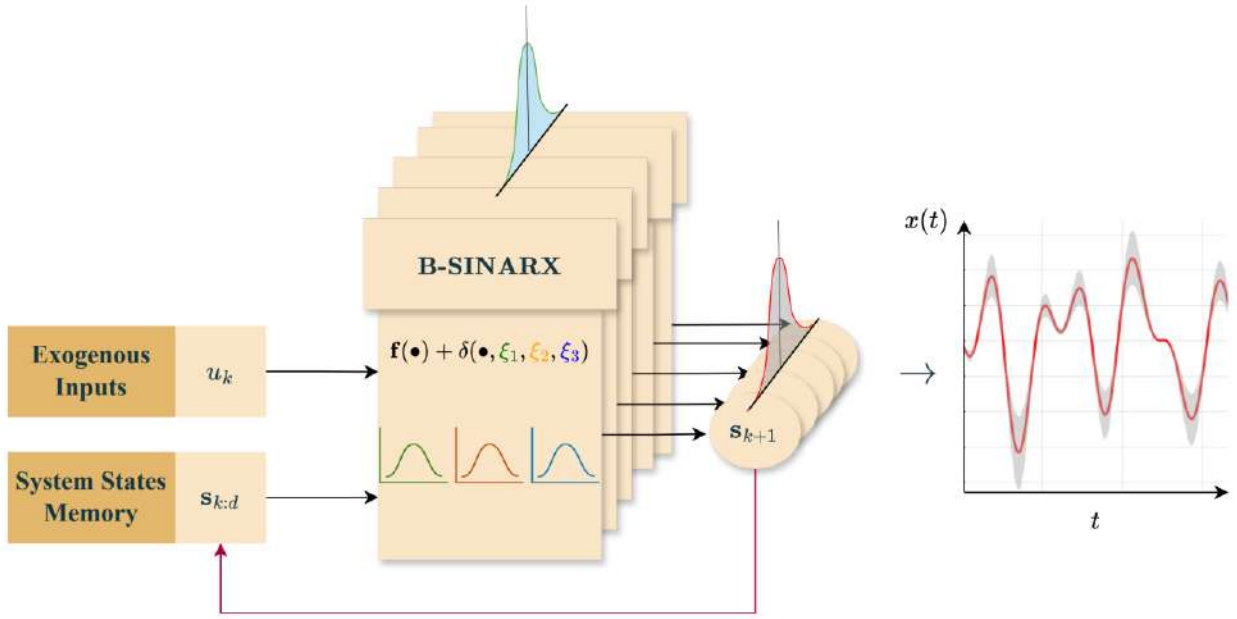
Proposed Methodology

Sparse Bayesian Polynomial NARX

Bayesian-SINARX Framework

Create a discrepancy model using Bayesian Nonlinear autoregressive augmentation to baseline low-fidelity model.

Learn an interpretable, parsimonious functions from scarce data.



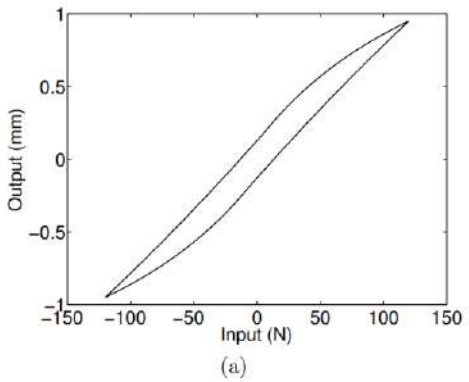
B-SINARX Framework

$$\mathbf{s}_{k+1} = \underbrace{\mathbf{f}\left(\phi_{k:d}\right)}_{\text{Physics based}} + \underbrace{\boldsymbol{\Theta}\left(\phi_{k:d}\right)\boldsymbol{\xi}}_{\text{Data-Driven}} + \boldsymbol{\varepsilon}_k,$$

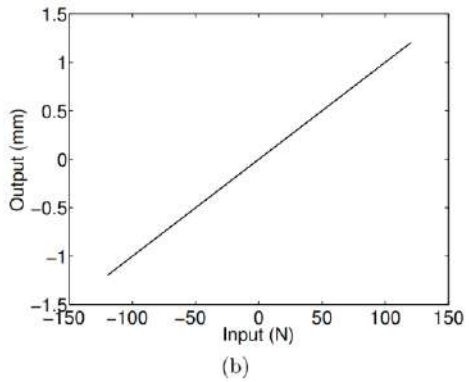
[7] Eyoel A. (2026, under view)

Problem Setup

The Bouc-Wen Hysteresis is used to model complex and nonlinear behaviors such as rate-dependent effect, memory and asymmetric responses.

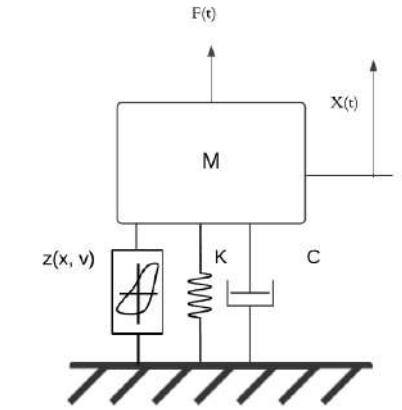


Nonlinear Behavior



Linear Behavior

Nonlinear Benchmarks



m [kg]	c [Ns/m]	k [N/m]	α [N/m]	β	γ [m^{-1}]	δ [m^{-1}]	ν
2	10	5×10^4	5×10^4	10^3	0.8	−1.1	1

$$m\,\ddot{x} + c\dot{x} + kx + z = u$$

$$\dot{z}(x,\dot{x}) = \alpha\,\dot{x} - \beta(\gamma|\dot{x}||z|^{n-1}z + \delta\dot{x}|z|^{n}$$

Objective

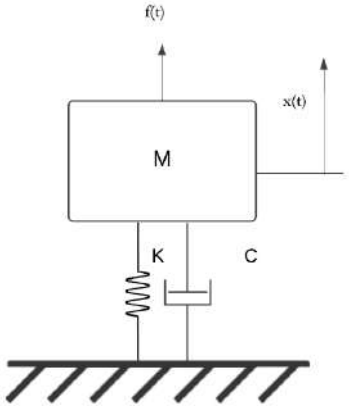
Build a discrepancy model that predicts the residual caused by hysteresis force from a baseline linear model

[8, 9] M. Schoukens and J.P. Noël, 2016,2017,

Problem Setup

Baseline Model

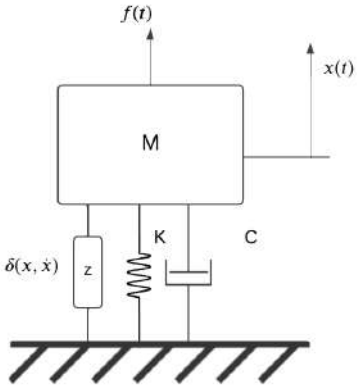
$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = f(t)$$



Discrepancy Model

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x + \boldsymbol{\delta(x,\dot{x})} = f(t)$$

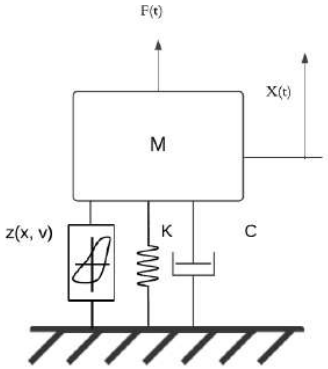
$$\boldsymbol{\delta(x,\dot{x}) = z = f(t) - f_{lin}}$$



True Model

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x + \boldsymbol{z(x,\dot{x})} = f(t)$$

$$\boldsymbol{\dot{z} = \alpha\dot{x} - \gamma|\dot{x}||z|^{n-1}z - \delta\dot{x}|z|^n}$$



Data Setup

Excitation
- Training data
  - **Multi-sine**: bandwidth 5–150 Hz
- Test Data
  - **Chirp** (sine sweep): [5 150] Hz
  - **Multi-sine**: bandwidth 5–150 Hz

Input:
- System States: $\boldsymbol{s(t)} = [\boldsymbol{x(t)}, \dot{\boldsymbol{x}}(\boldsymbol{t}), \boldsymbol{z(t)}]$
- Excitation:f($\boldsymbol{t}$)

5% noise was add to the data

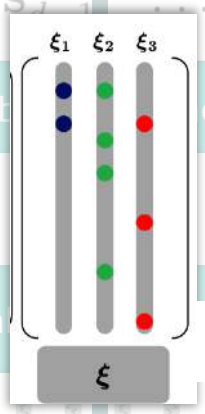
Performance measurement

EEARTH (Enhanced error assessment of response time histories)

Training Data

Solve Nonlinear ODE

Henkel Matrix



Evaluate Residual

$$\mathcal{R} := \mathcal{S} - \mathcal{F} \in \mathbb{R}^{(N-d+1)\times 3}$$

Identification

Priors

$$\pi(\boldsymbol{\xi}|\boldsymbol{\alpha}) = (2\pi)^{-M/2} \prod_{m=1}^{M} \alpha_m^{1/2} \exp\left(-\frac{\alpha_m \xi_m^2}{2}\right),$$

Sparse Bayesain Regression

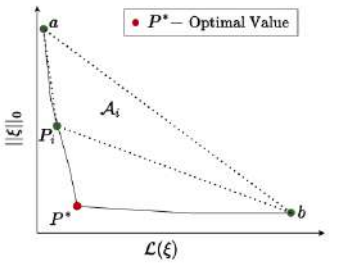
$$\pi(\boldsymbol{\mathcal{R}}|\boldsymbol{\xi}, \boldsymbol{\alpha}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{\|\boldsymbol{\mathcal{R}} - \boldsymbol{\Theta}_{\boldsymbol{\Phi}}\boldsymbol{\xi}\|^2}{2\sigma^2}\right).$$

$$\pi(\boldsymbol{\xi}|\boldsymbol{\alpha}, \sigma^2, \mathcal{D}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\gamma_i = 1 - \alpha_i \Sigma_{ii} \qquad \sigma^{-2}\boldsymbol{\Theta}^\top\boldsymbol{\Theta}),$$

«well-determinedness»

Evaluate



- Simulation Performance
- Hystersis Identification

Result and Discussion

Identified Model

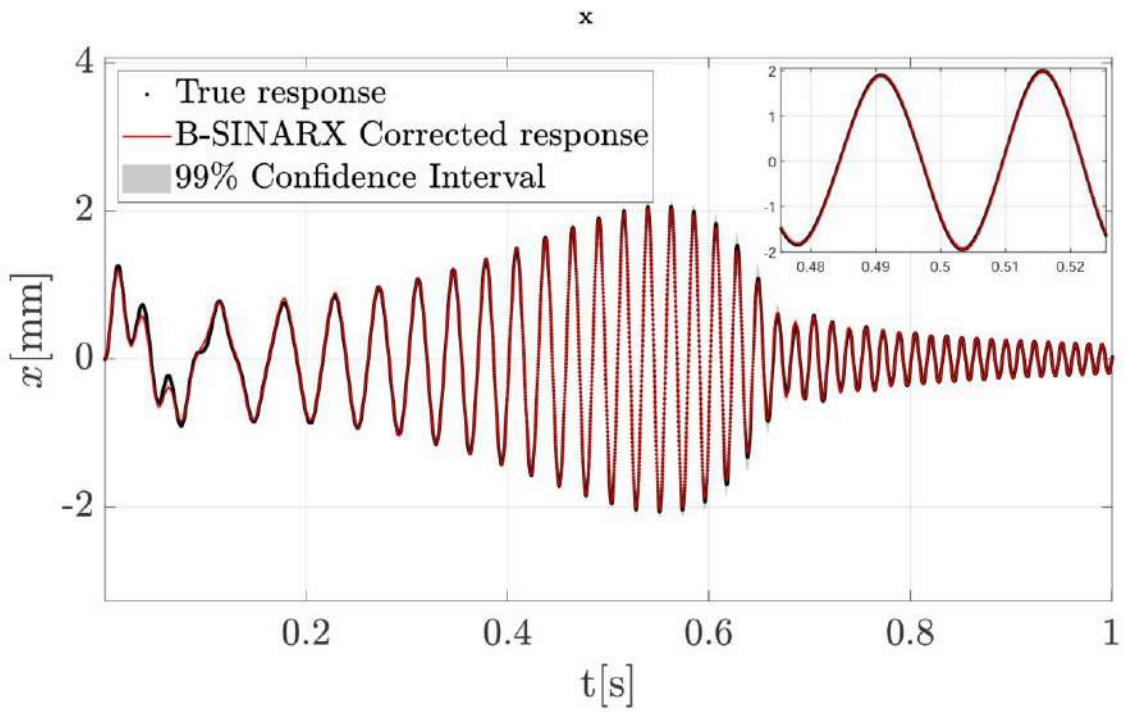
State	Correct Model	B-SINARX	Error
x_{k+1}	$x_k + 1.3 \times 10^{-4} v_k$	$0.997x_k + 1.3 \times 10^{-4} v_k + 2.16 \times 10^{-8} z_k$	0.3%
v_{k+1}	$0.999v_k - 3.3x_k + 6.67 \times 10^{-5} u_k - 6.67 \times 10^{-5} z_k$	$0.997v_k - 3.68x_k + 6.71 \times 10^{-5} u_k - 6.16 \times 10^{-5} z_k$	6.6%
z_{k+1}	$z_k + 6.67v_k - 0.107\|v_k\|z_k + 0.147v_k\|z_k\|$	$0.998z_k + 6.63v_k - 0.105\|v_k\|z_k + 0.148v_k\|z_k\|$	1.05%

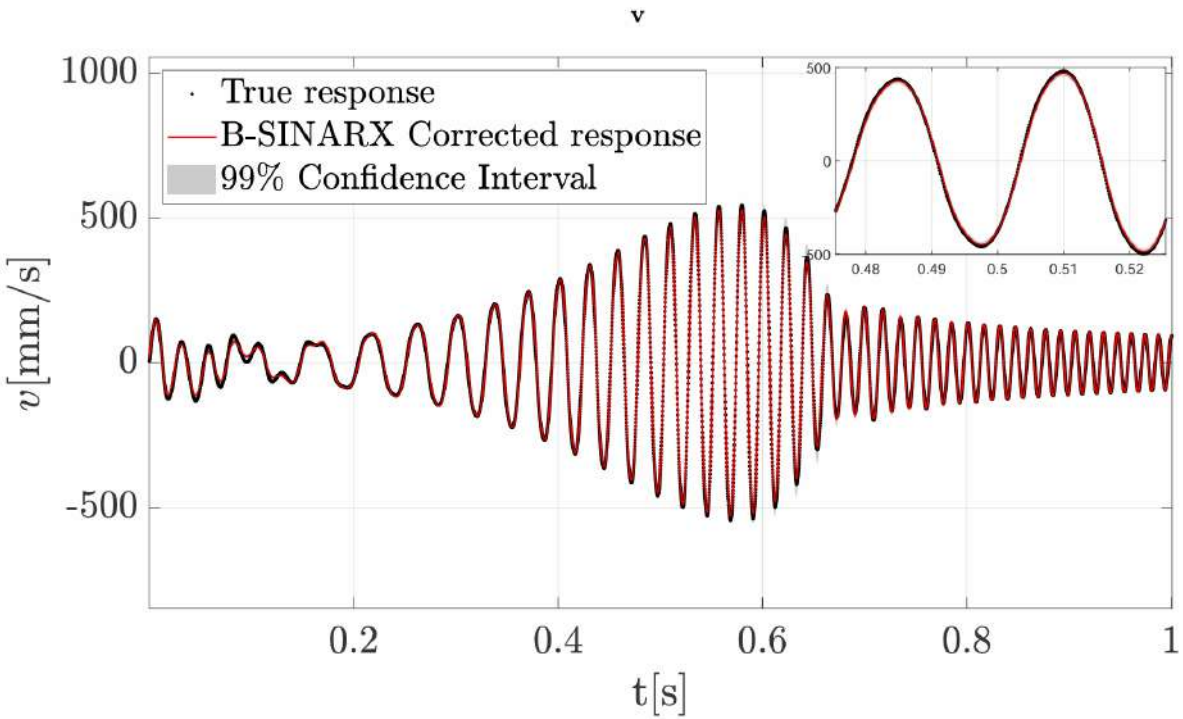
– Incorrectly identified – Correctly identified (< 2% error) – Identified (> 2% error)

- Missing terms are identified with high accuracy
- Hysteresis force Identified accurately
- Adjustment of already existing "parameters"

Time Response

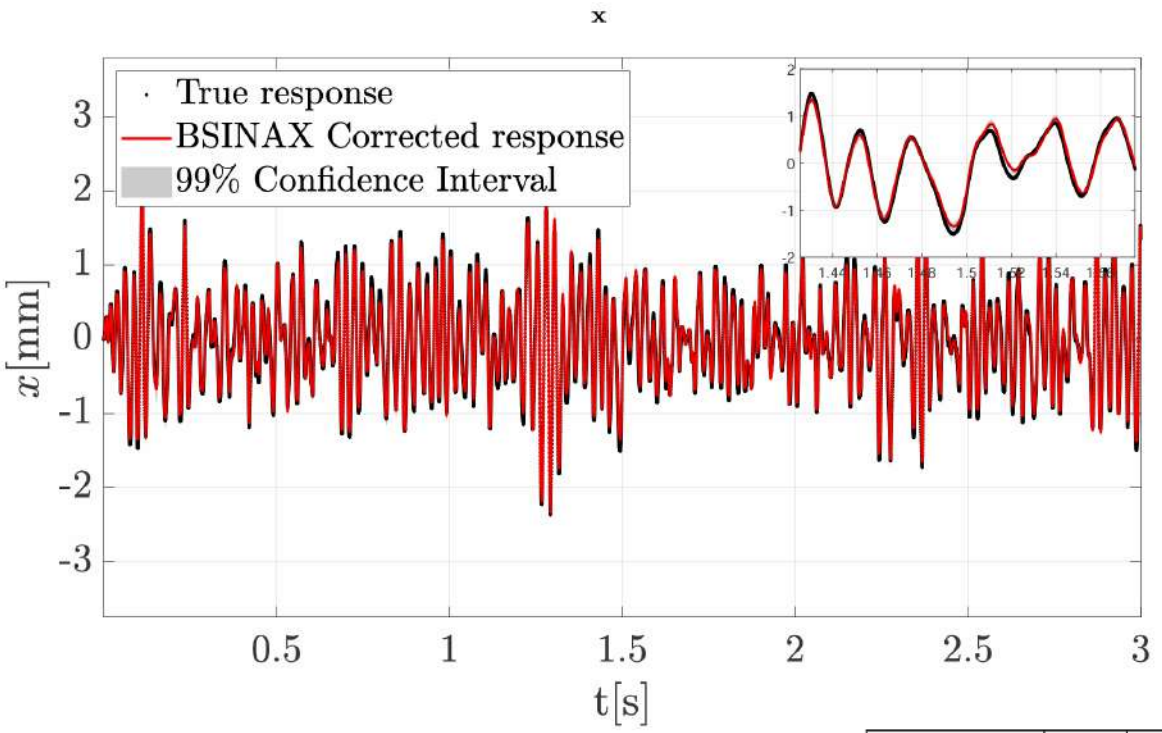
Chirp Input

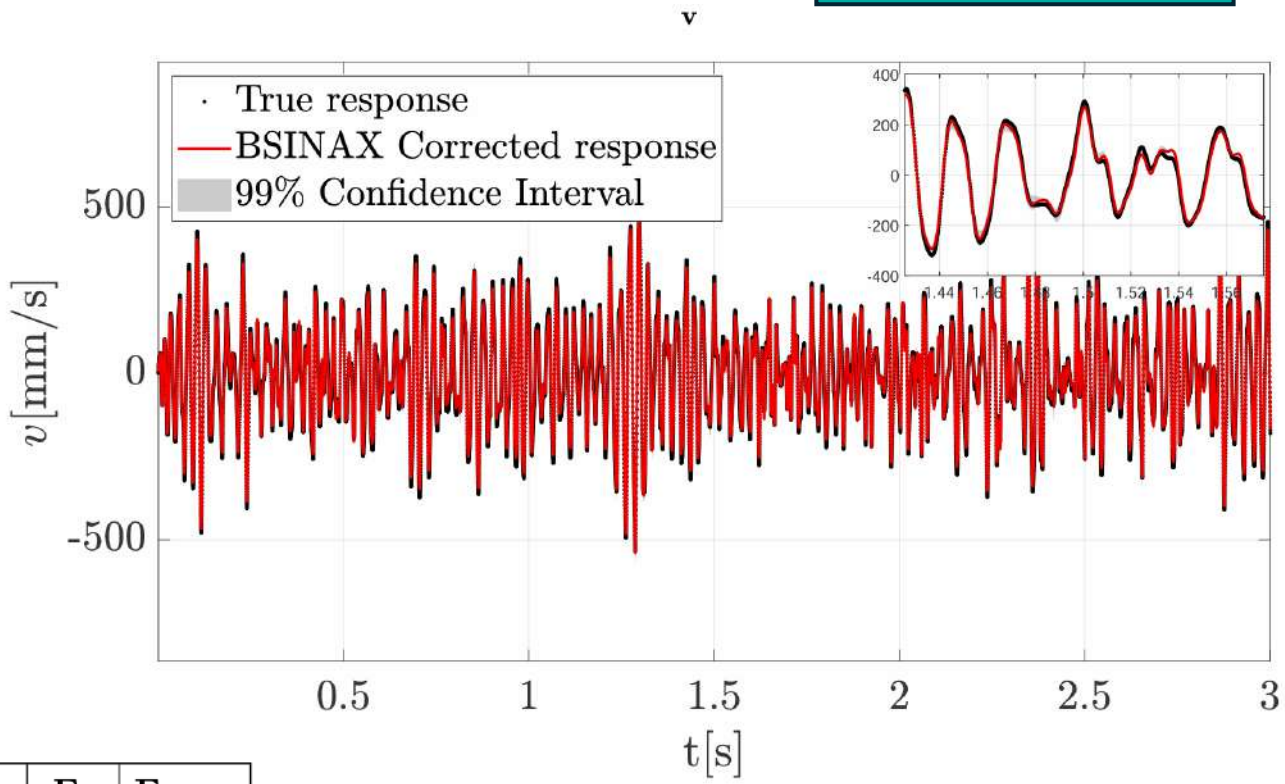




Time Response

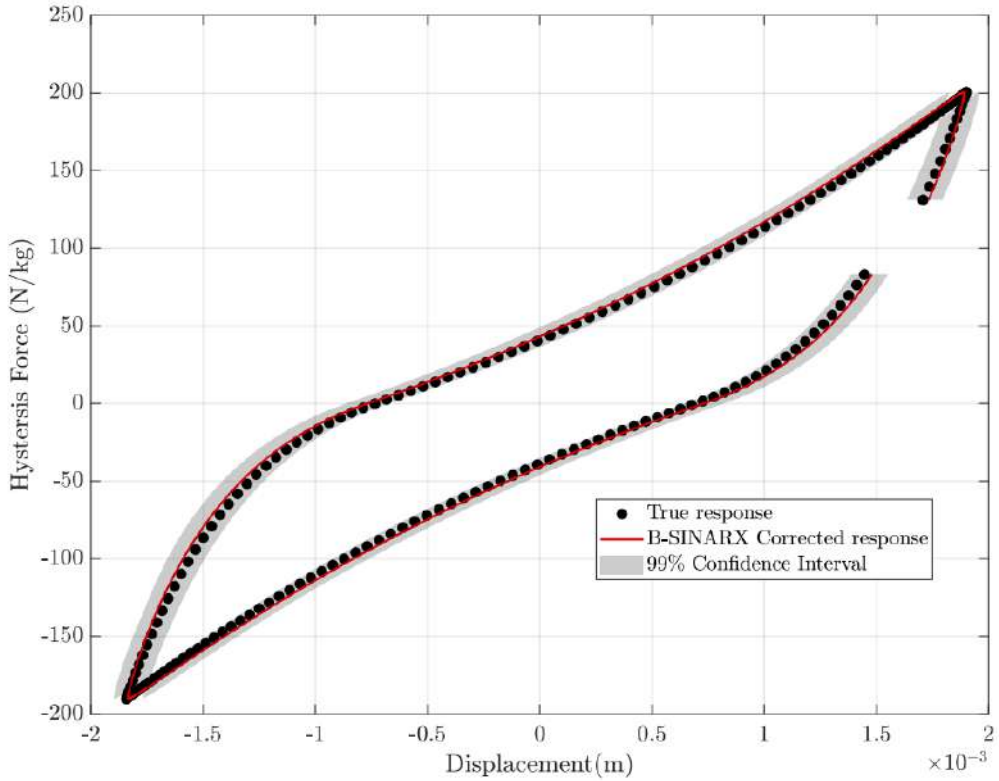
Multi-sine Input





Dataset	E_p	E_m	E_s	E_{Total}
Chirp	100	96.3	96.4	**97.1**
Multi-sine	100	97.9	98.3	**98.5**

Hysteresis plot



Correctly identified hysteresis behaviour with a good confidence interval



Discussion

It is shown that, it is possible to learn an interpretable parameterized NARX model with in the discrepancy framework.

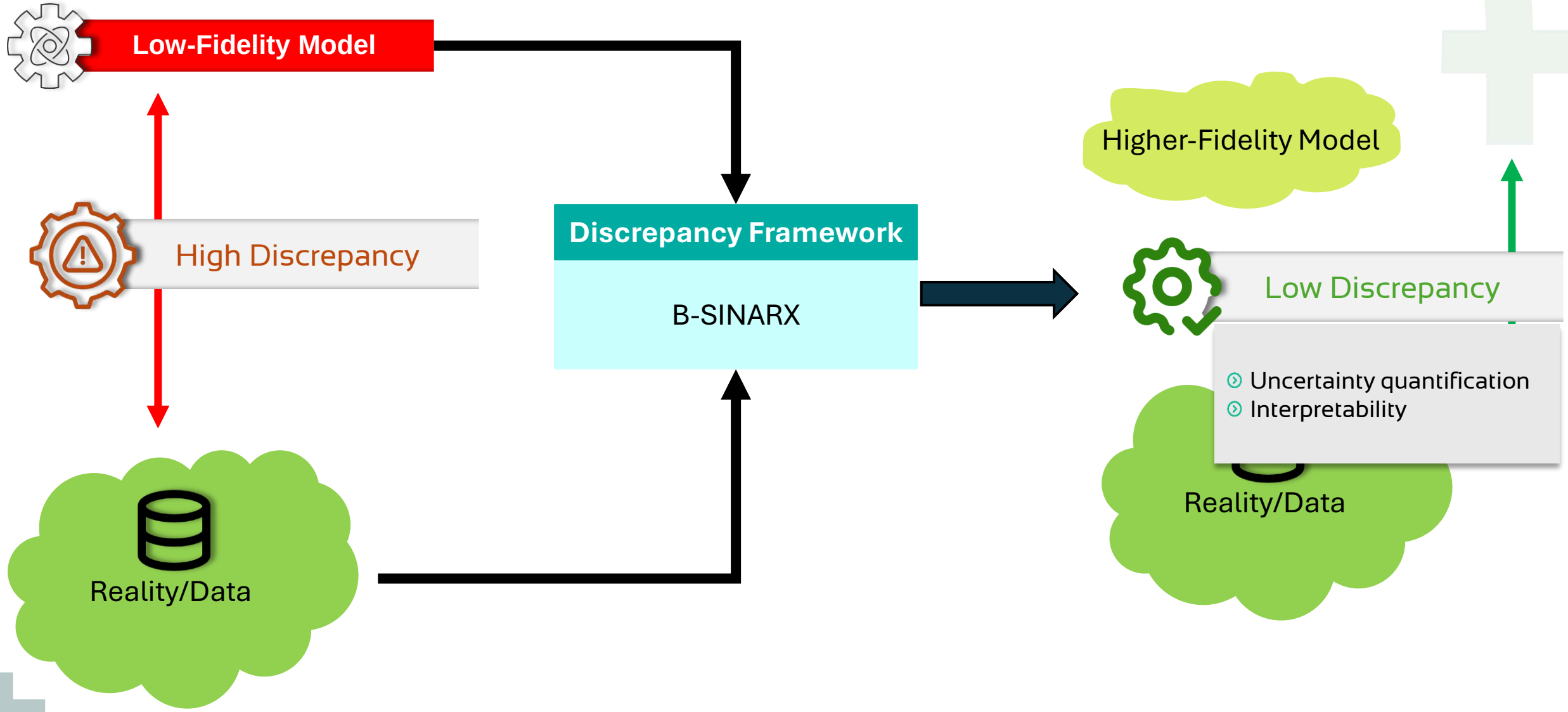
Parsimony and Generalizability to inputs it hasn't seen make this approach even powerful for cases where data is scarce

This approach enables target model correction rather than full model replacement

Since uncertainty quantification is included in the approach we can see when our mode is able to predict well or not.

Conclusion and Insight

Why is this useful?


Low-Fidelity Model
High Discrepancy
Reality/Data
Discrepancy Framework
B-SINARX
Higher-Fidelity Model
Low Discrepancy
Uncertainty quantification
Interpretability
Reality/Data

Experimental Validation

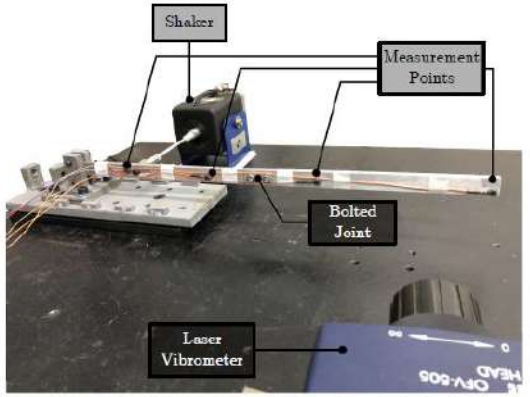
Interpretable Discrepancy Learning in Dynamical Systems Using Bayesian SINDy-NARX: The B-SINARX Framework

Eyoel Atnafe[a], Rafael Teloli[b], Emeline Sadoulet-Reboul[a], Scott Cogan[a]

[aa] Université Marie et Louis Pasteur, CNRS, Institut FEMTO-ST, , Besançon, F-25000, , France
[b b] Université Marie et Louis Pasteur, SUPMICROTECH, CNRS, Institut FEMTO-ST, , Besançon, F-25000, , France



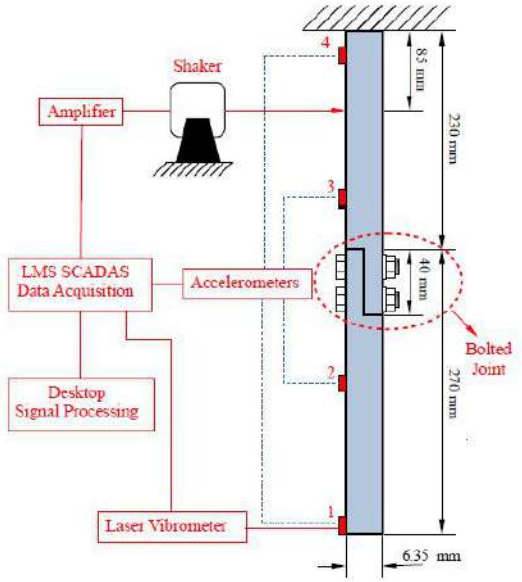
Date Final Disposition Set	Final Disposition
10 Oct 2026	Accept



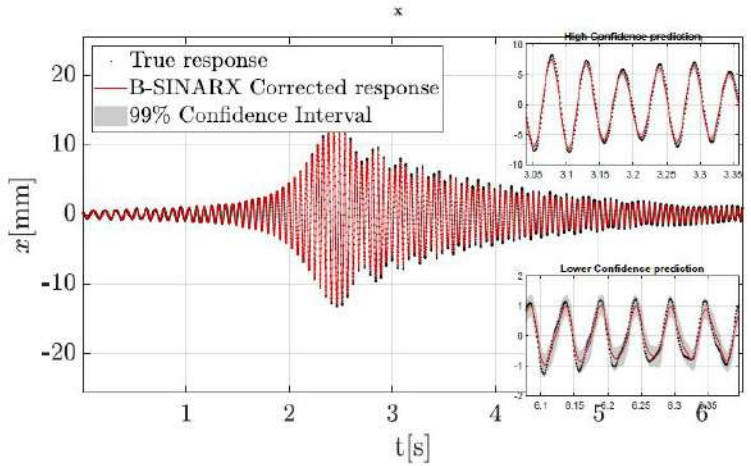
Top view of the bolted joint:



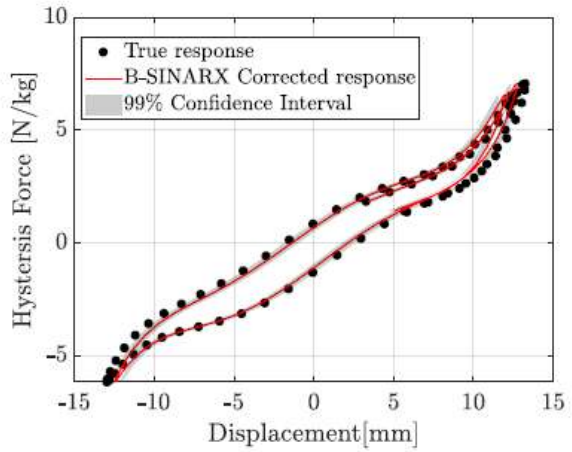
(a) Experimental setup.



(b) Schematic top view of the experimental setup.



(a) Comparison between the actual and predicted displacement time response from E approach for 15 mV chirp excitation



(b) Hysteresis Loop

References

J. Brynjarsdottir, A. O'Hagan, Learning about physical parameters: The importance of model discrepancy, Inverse problems 30 (11) (2014) 837 114007

A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, Q. Malartic, A com846 parison of combined data assimilation and machine learning methods for offline and online model error correction, Journal of computational science 55 (2021) 101468

Maarten Schoukens, Improving Baseline Models through Model Augmentation: Model Structures and Learning Approaches, Nonlinear Benchmark Workshop, Sheffield, UK, 2026

M. Levine, A. Stuart, A framework for machine learning of model error in dynamical systems, Communications of the American Mathematical Society 2 (07) (2022) 283–344

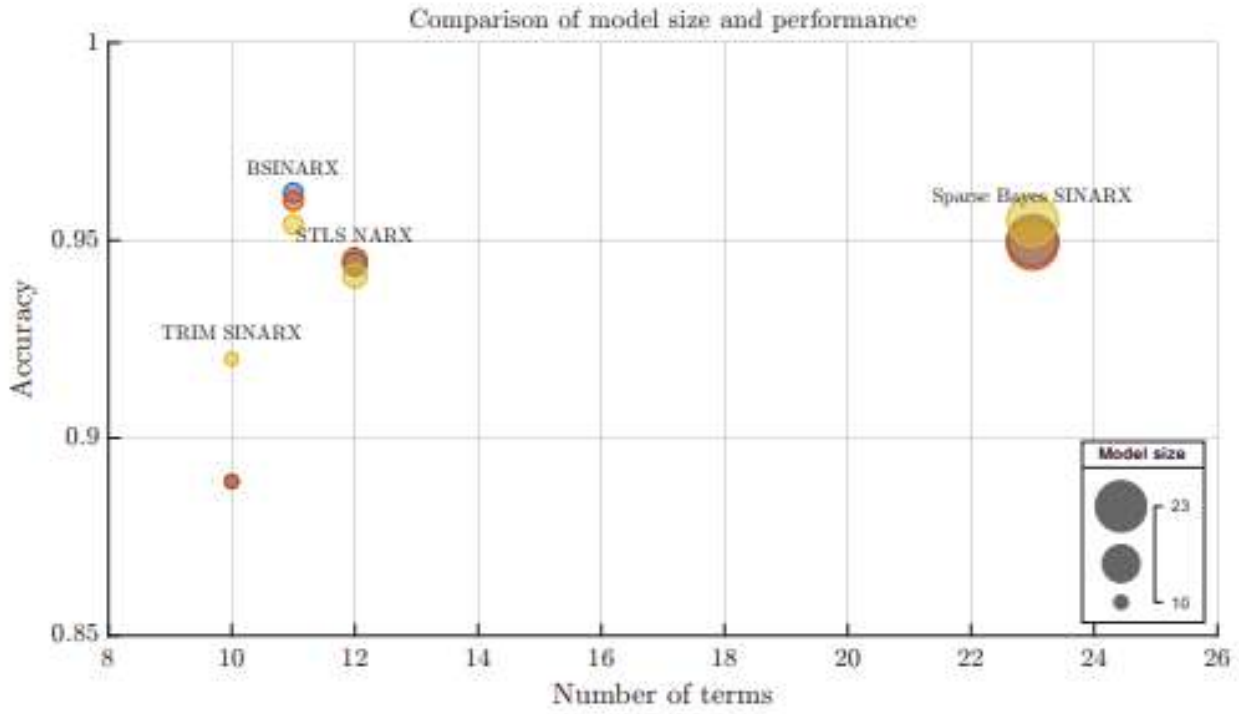
S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems Proceedings of the national academy of sciences 113 (15) (2016) 3932– 858 3937

Atnafe, E., Teloli, R., Sadoulet-Reboul, E., Cogan, S. Interpretable Discrepancy Learning in Dynamical Systems Using Bayesian SINDy-NARX: The B-SINARX Framework, Under Review, Mechanical Systems and Signal Processing.

Thank you!

Q&A

Ablation



Methods compared:

- STLS
- Trimmed Sparse Regression
- Vanila Sparse Bayesian Regression

Method	30 mV	20 mV	15 mV	# Terms
STLS NARX	0.944	0.945	0.941	12
Trimmed SINARX	0.889	0.889	0.920	**10**
Sparse Bayes NARX	0.950	0.949	**0.955**	23
B-SINARX	**0.962**	**0.960**	0.954	11

Change of Library

The choice of candidate function library has a decisive effect on how specific dynamic phenomena are characterized.

A deliberate, **physics-informed bias** in the **library** is advantageous, since it steers the identification toward models consistent with the presumed true physics.

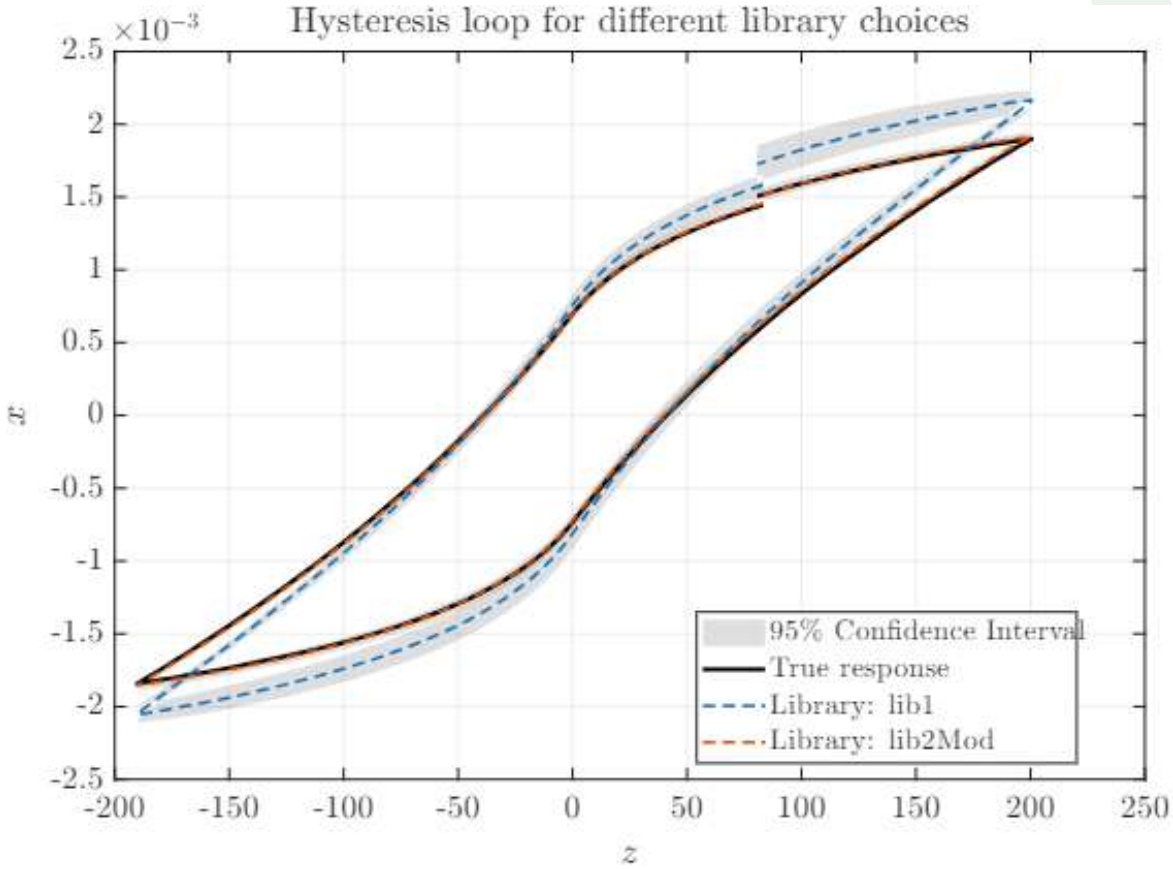


Table 5: Identified model using polynomial basis functions (`lib1.m`)

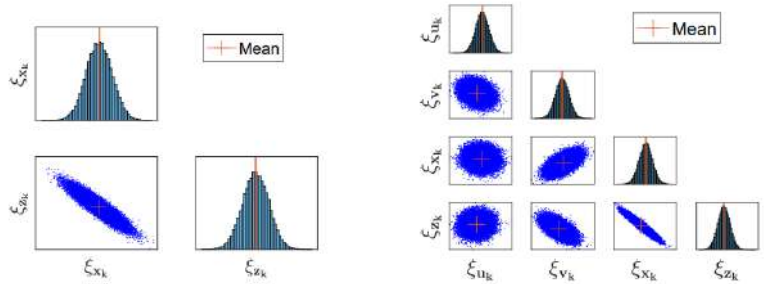
Basis Library	System state	Identified terms
Polynomial (`lib1.m`)	z_{k+1}	$0.992\, z_k + 9.60\, v_k + 0.086\, \mathbf{v_k z_k^2} - 9.15 \times 10^{-4}\, \mathbf{v_k^2 z_k}$
Poly + Abs (`lib2.m`)	z_{k+1}	$0.999\, z_k + 6.63\, v_k + 0.146\, \mathbf{v_k} \lvert \mathbf{z_k} \rvert - 1.01 \times 10^{-1}\, \lvert \mathbf{v_k} \rvert \mathbf{z_k}$

Identified Model

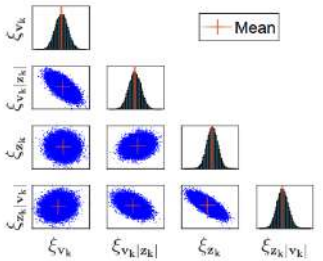
State	Correct Model	B-SINARX	Error
x_{k+1}	$x_k + 1.3 \times 10^{-4} v_k$	$0.997x_k + 1.3 \times 10^{-4} v_k + 2.16 \times 10^{-8} z_k$	0.3%
v_{k+1}	$0.999v_k - 3.3x_k + 6.67 \times 10^{-5} u_k - 6.67 \times 10^{-5} z_k$	$0.997v_k - 3.68x_k + 6.71 \times 10^{-5} u_k - 6.16 \times 10^{-5} z_k$	6.6%
z_{k+1}	$z_k + 6.67v_k - 0.107\lvert v_k \rvert z_k + 0.147 v_k \lvert z_k \rvert$	$0.998z_k + 6.63v_k - 0.105\lvert v_k \rvert z_k + 0.148 v_k \lvert z_k \rvert$	1.05%

– Incorrectly identified – Correctly identified (< 2% error) – Identified (> 2% error)

State	ξ	True	B-SINARX(mean)	Error(%)
x_{k+1}	ξ_{x_k}	1.000	0.997	0.3
	ξ_{v_k}	1.3×10^{-4}	1.3×10^{-4}	0.0
	ξ_{z_k}	0	2×10^{-8}	–
v_{k+1}	ξ_{u_k}	6.67×10^{-5}	6.71×10^{-5}	0.6
	ξ_{z_k}	-6.67×10^{-5}	-6.16×10^{-5}	10.4
	ξ_{v_k}	1.000	0.997	0.3
	ξ_{x_k}	−3.3	−3.68	7.6
z_{k+1}	ξ_{z_k}	1.000	0.998	0.2
	ξ_{v_k}	6.67	6.63	0.6
	$\xi_{v_k \lvert z_k \rvert}$	−0.107	−0.105	1.9
	$\xi_{z_k \lvert v_k \rvert}$	0.147	0.148	0.7



(a) (b)



(c)

Library of Function

Power

$$\Theta_p(x, v, z, u) = \{x^a, v^a, z^a, z^a\}$$
$$s.t \quad a \in 1, 2, \ldots, p$$

Polynomials

$$\Theta_p(x, v, z, u) = x^{a_1} v^{a_2} z^{a_3} z^{a_4}$$
$$s.t \sum_{i=1}^{4}(a_i) \leq p, \quad a_i \ \epsilon \mathbb{N}_0$$

Custom/Biased

$$\Theta_p(x, v, z, u) = x^{a_1} v^{a_2} z^{a_3} z^{a_4} |x|^{b_1} |v|^{b_2} |z|^{b_3} |u|^{b_4}$$
$$s.t \quad \sum_{i=1}^{4}(a_i + b_i) \leq p, \quad a_i, b_i \ \epsilon \mathbb{N}_0$$