

Optimizing Experiments with Pyomo.DoE

`dowlinglab.github.io/pyomo-doe`

Alexander (Alex) Dowling, Ph.D.

Daniel Laky, Ph.D.

Chemical and Biomolecular Engineering
University of Notre Dame

PSE Stakeholder Summit
September 19, 2024

Contributors to Pyomo.DoE and ParmEst: Jialu Wang (ND), Dan Laky (ND), Hailey Lynch (ND), John Sirola (SNL), Bethany Nicholson (SNL), Miranda Mundt (SNL), Shawn Martin (SNL), Katherine Klise (SNL)



Power of Adaptive Sequential Optimal Experiments

Self-Driving Laboratories

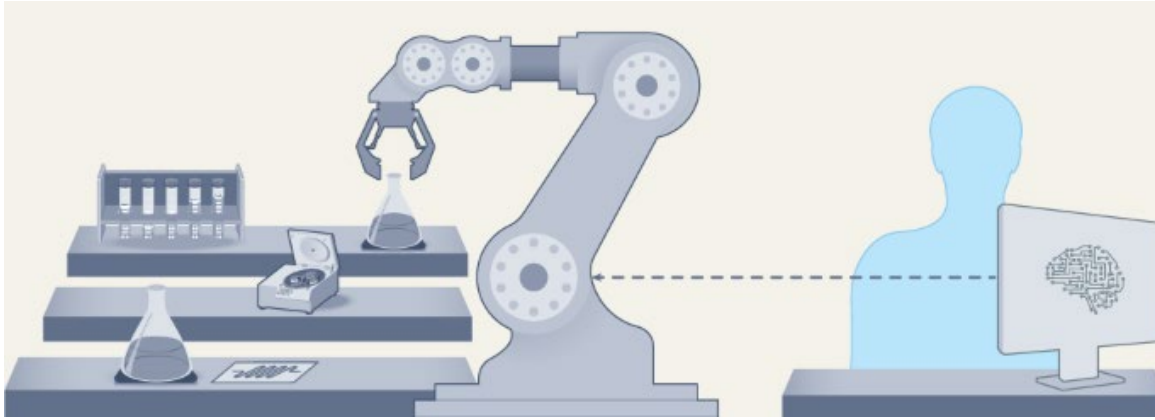


Figure: Abolhasani & Kumacheva (2023), *Nature Syn.*

Epps et al. (2022), *Advanced Materials*

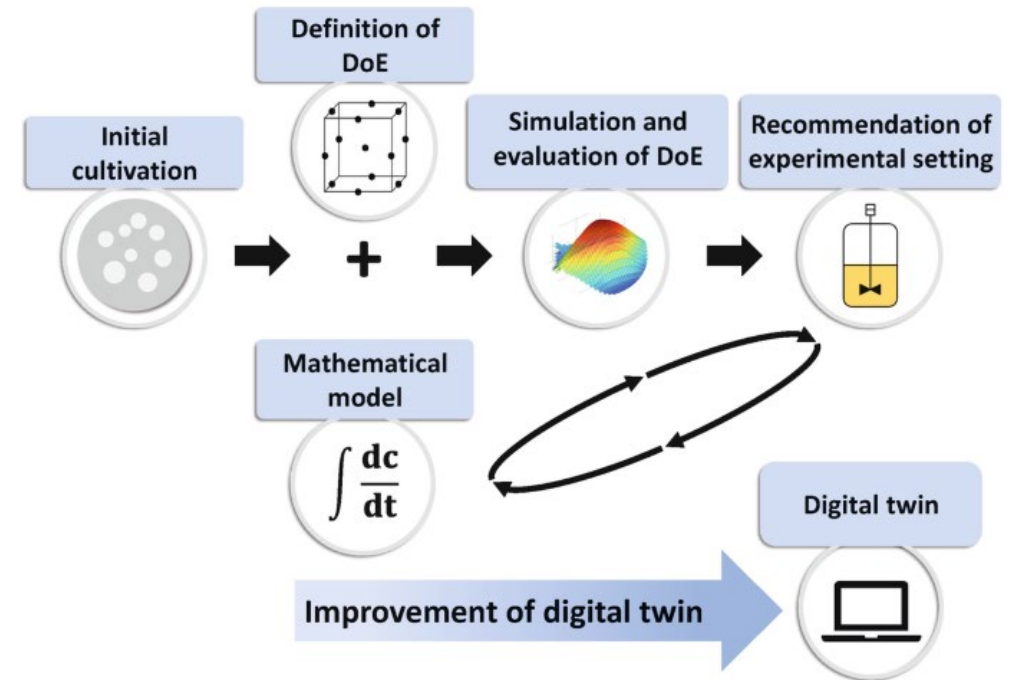
MacLeod et al. (2020), *Science Advances*

MacLeod et al. (2022), *Nature Communications*

Hase, Roch, Aspuru-Guzik (2019), *Trends in Chemistry*

Seifrid et al. (2022), *Acc. Chem. Res.*

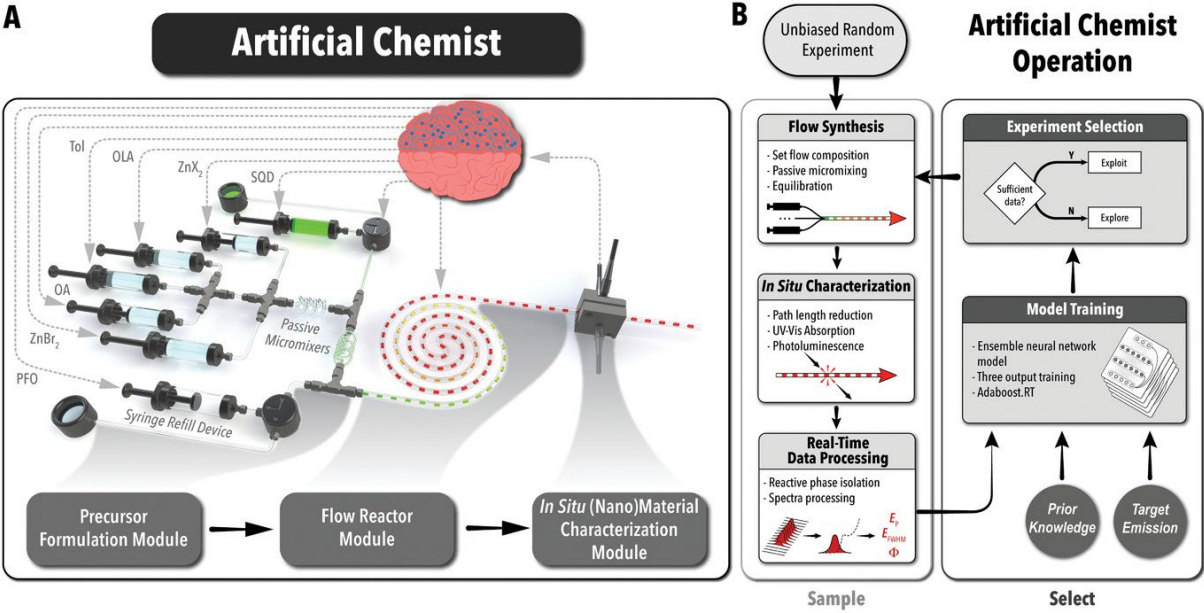
Automation + Model-Based Design of Experiments



Kuchemuller et al. (2020), *Digital Twins*

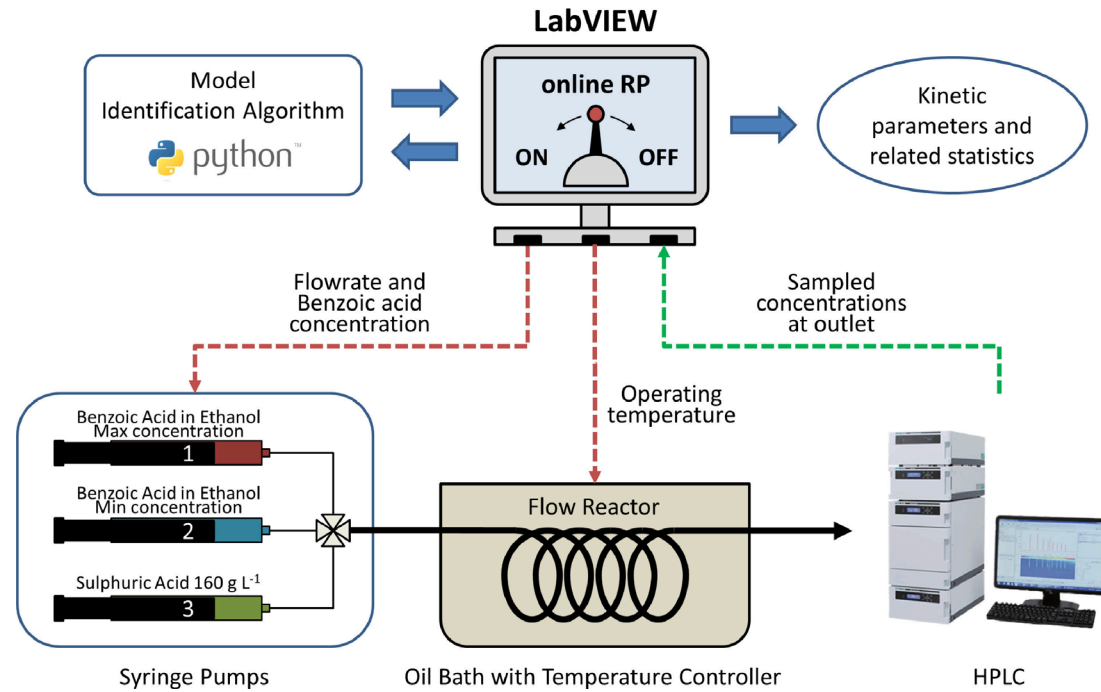
Many Recent Examples of Sequential Optimal Experiments

Quantum Dots (Machine Learning)



Epps et al. (2022), *Advanced Materials*

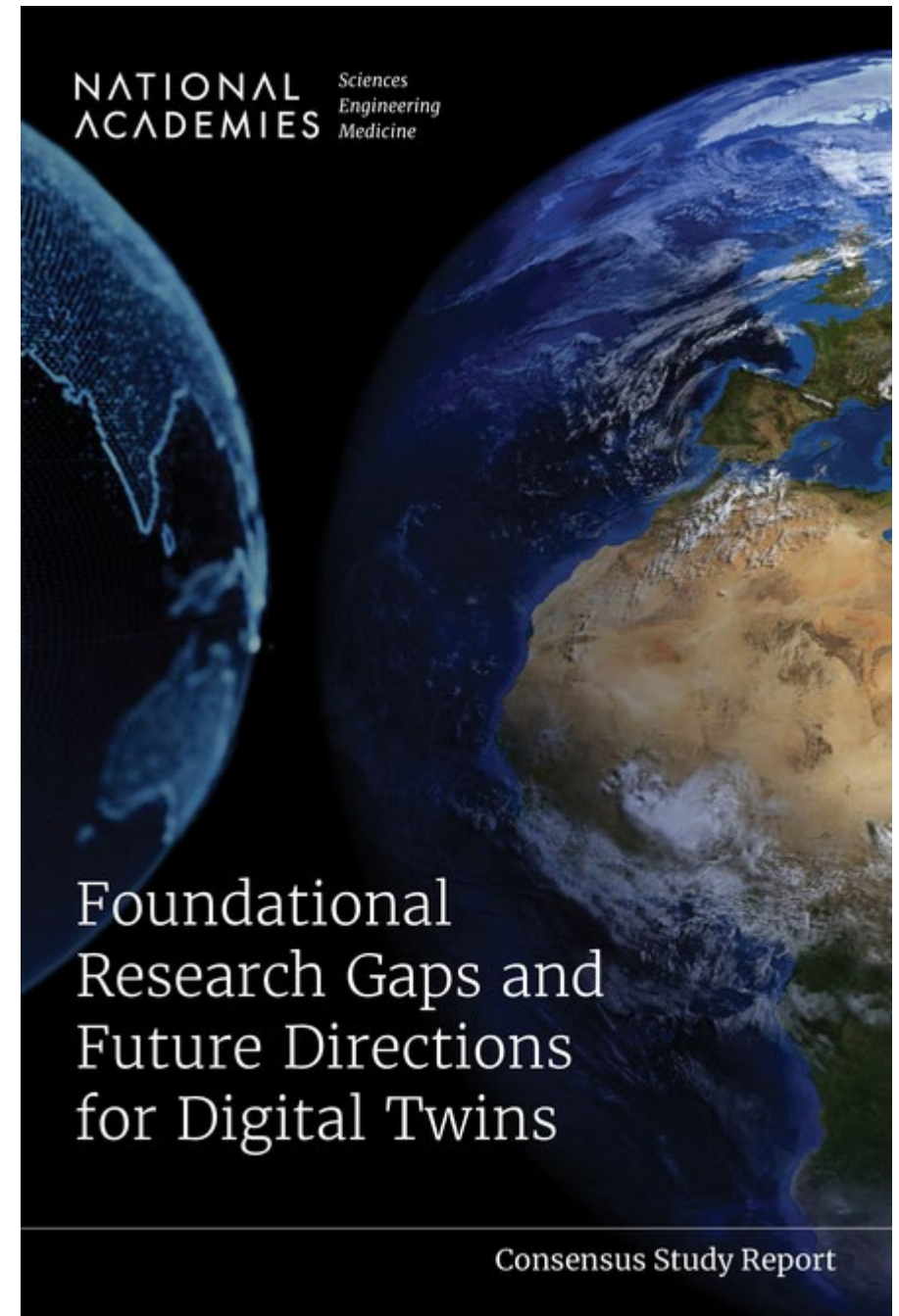
Reaction Engineering (Science-based Models)



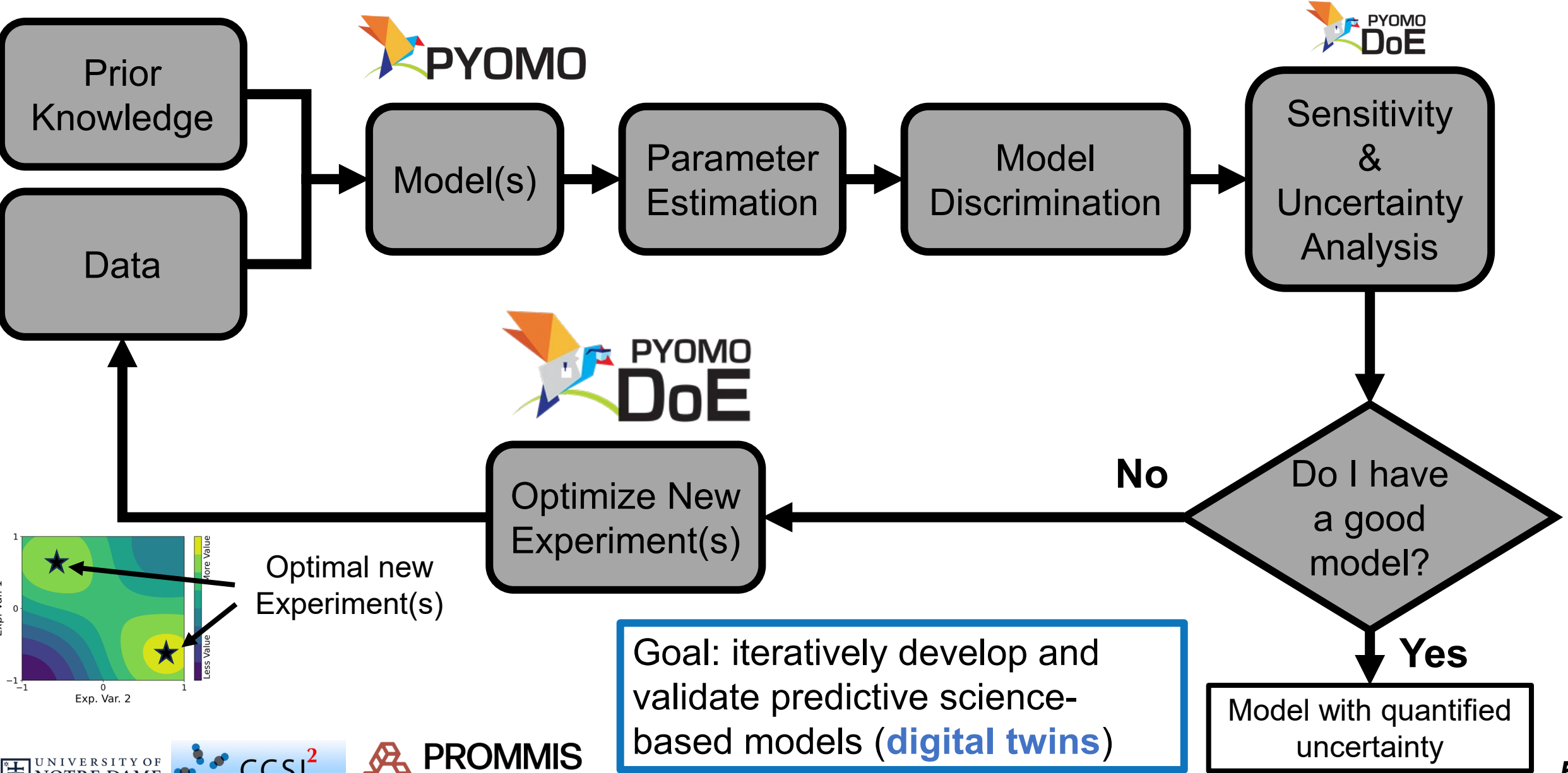
Quaglio et al. (2019), *Comp. & Chem. Eng.*

What is a **Digital Twin**?

*A **digital twin** is a set of **virtual information constructs** that mimics the structure, context, and behavior of a **natural, engineered, or social system** (or **system-of-systems**), is **dynamically updated** with data from its physical twin, has a **predictive capability**, and **informs decisions that realize value**. The **bidirectional interaction** between the virtual and the physical is central to the digital twin.*

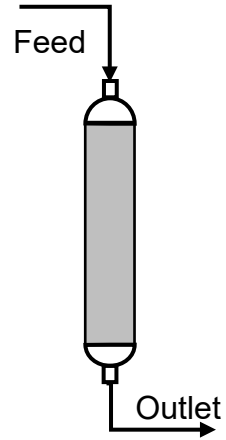


Science-based Data Analytics Workflow

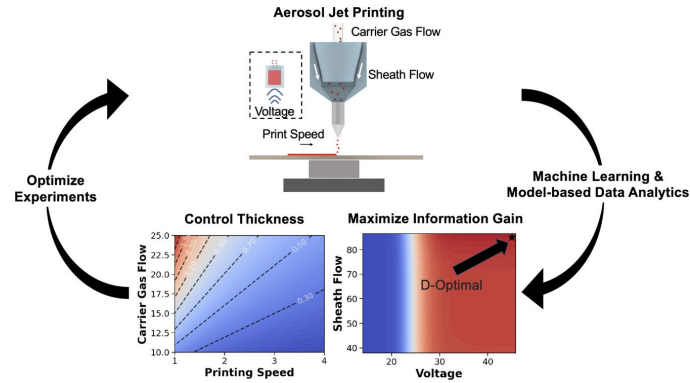


MBDoe Facilitates Collaborations

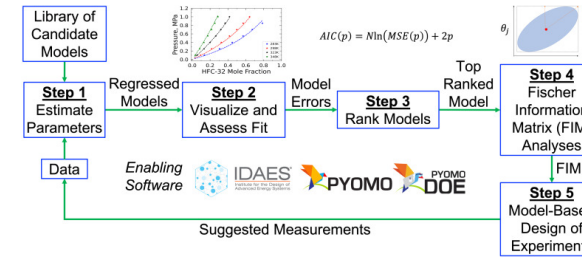
CO₂ Capture



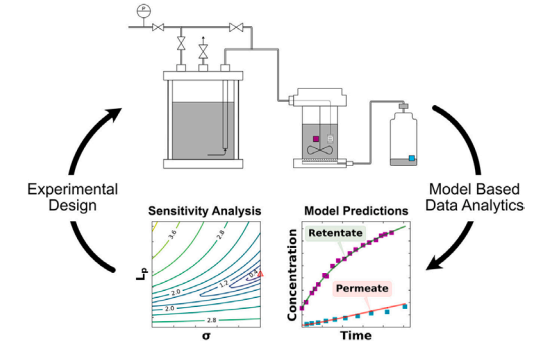
Additive Manufacturing of Thermoelectric Devices



Thermodynamic Modeling (Refrigerants)



Rapid/Automated Membrane Characterization



Jialu Wang



Ke Wang



Dr. Bridgette Befort



Xinhong Liu



Wang, J. and Dowling, A.W. (2022), *AIChE J.* e17813.

Wang K., Zhang M., Wang, J., Shang, W., Zhang, Y., Luo, T., Dowling, A.W. (2023), *Digital Chemical Engineering*

Befort, B.J., Garciadiego, A., Wang, J., Wang, K., Maginn, E.J., Dowling, A.W. (2023), *Fluid Phase Equilibria*.

Ouimet, J.A, Xinhong, L., Brown, D.J., Eugene, E.A., Pops, T., Muetzel, Z.W., Dowling, A.W., Phillip, W.A., (2022). *J. Membrane Science*.

Pyomo.DoE Example: Temperature Control Lab (TC Lab)

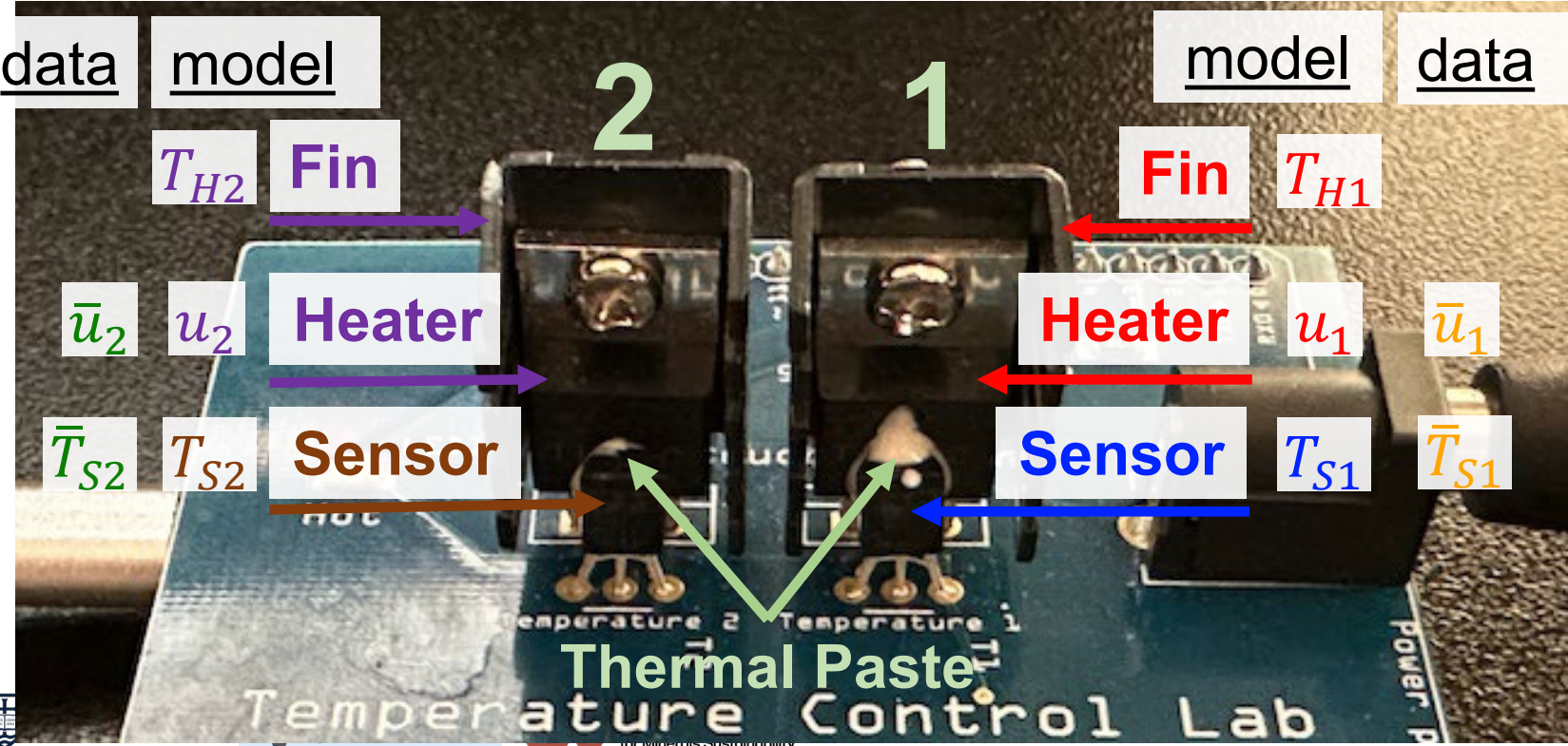
dowlinglab.github.io/pyomo-doe/notebooks/tclab_model.html

$$C_p^H \frac{dT_{H,1}}{dt} = U_a(T_{amb} - T_{H,1}) + U_b(T_{S,1} - T_{H,1}) + \alpha P_1 u_1$$

$$C_p^S \frac{dT_{S,1}}{dt} = U_b(T_{H,1} - T_{S,1}), \quad \theta = (U_a, U_b, C_p^H, C_p^S)^\top$$

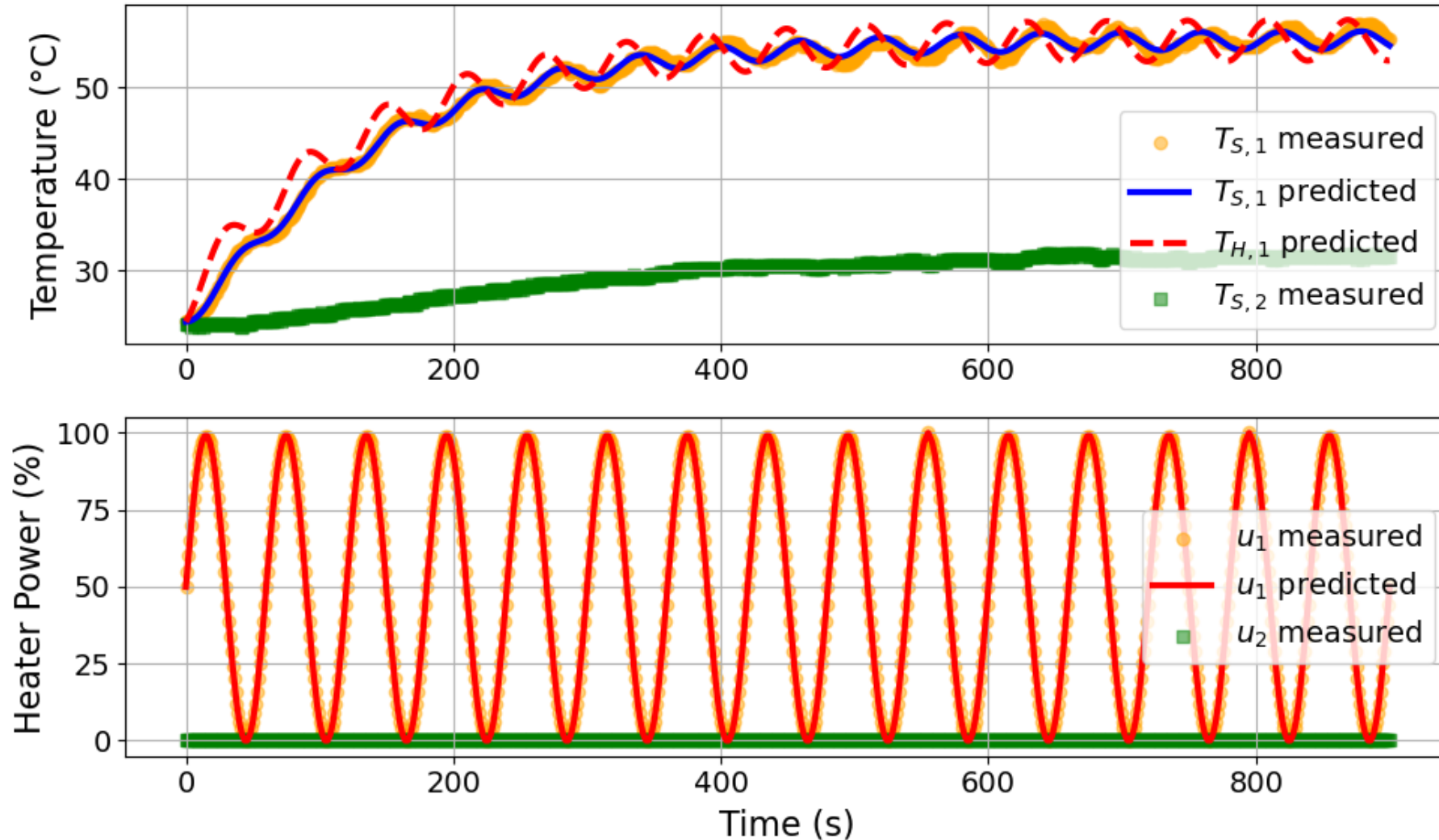


Thank you to Prof. Jeff Kantor (1954-2023) for the TCLab example and so much more.



TC Lab: Data and Parameter Estimation

Hands-On Tutorial: dowlinglab.github.io/pyomo-doe

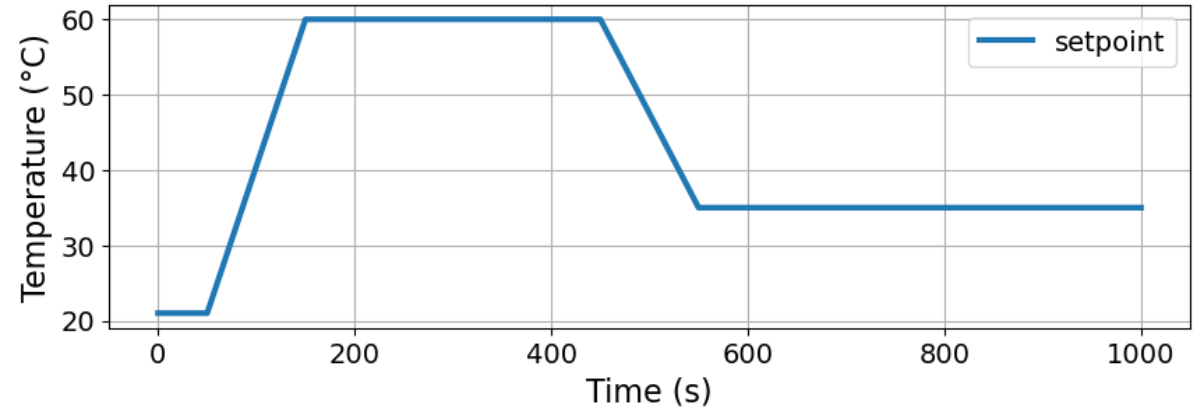
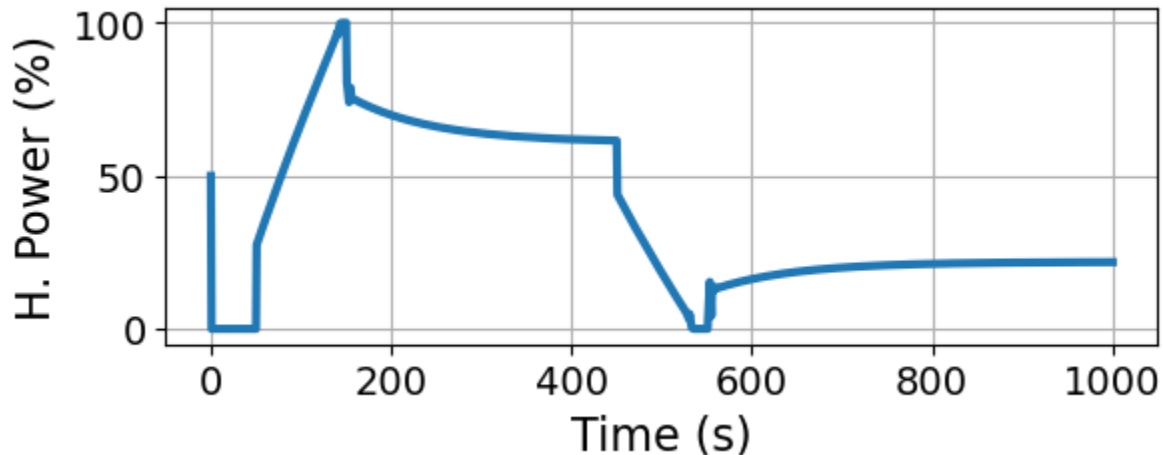


TC Lab: Dynamic Optimization in Pyomo

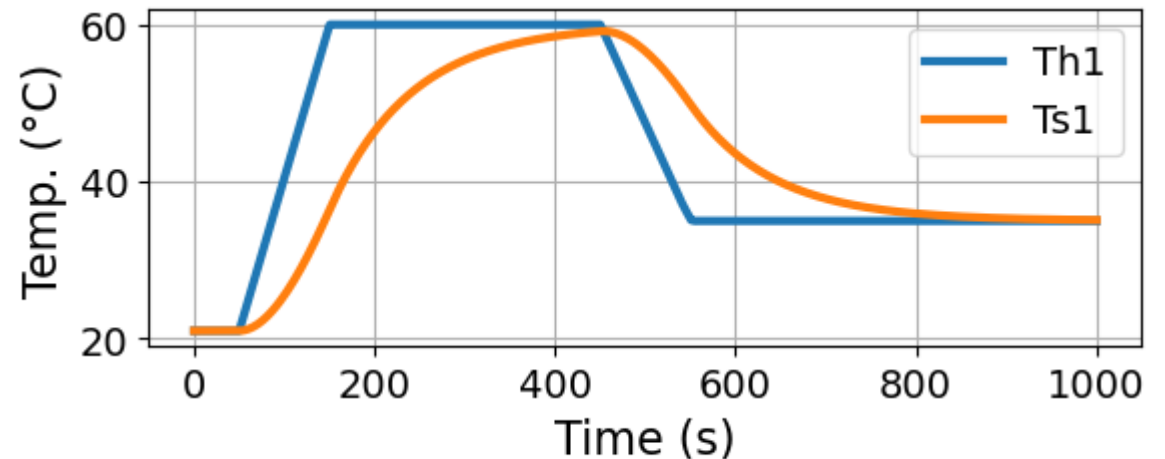
dowlinglab.github.io/pyomo-doe/notebooks/pyomo_simulation.html

$$\begin{aligned} \min_{u(t)} & \int_{t_0}^{t_f} \| SP(t) - T_H(t) \|^2 dt \\ \text{s. t.} & C_p^H \frac{dT_H}{dt} = U_a(T_{amb} - T_H) + U_b(T_S - T_H) + \alpha P u(t) \\ & C_p^S \frac{dT_S}{dt} = U_b(T_H - T_S) \\ & T_H(t_0) = T_{amb} \\ & T_S(t_0) = T_{amb} \end{aligned}$$

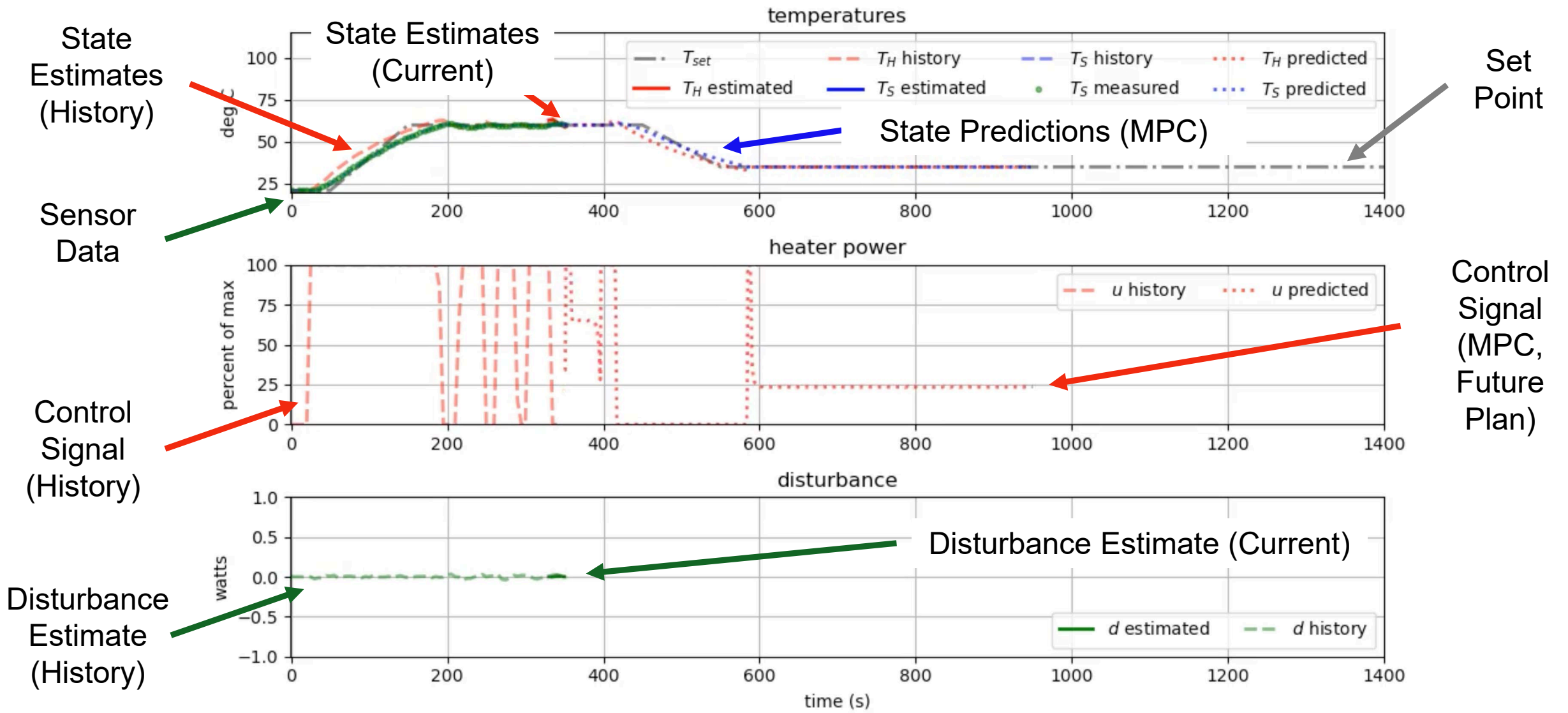
Optimal control profile



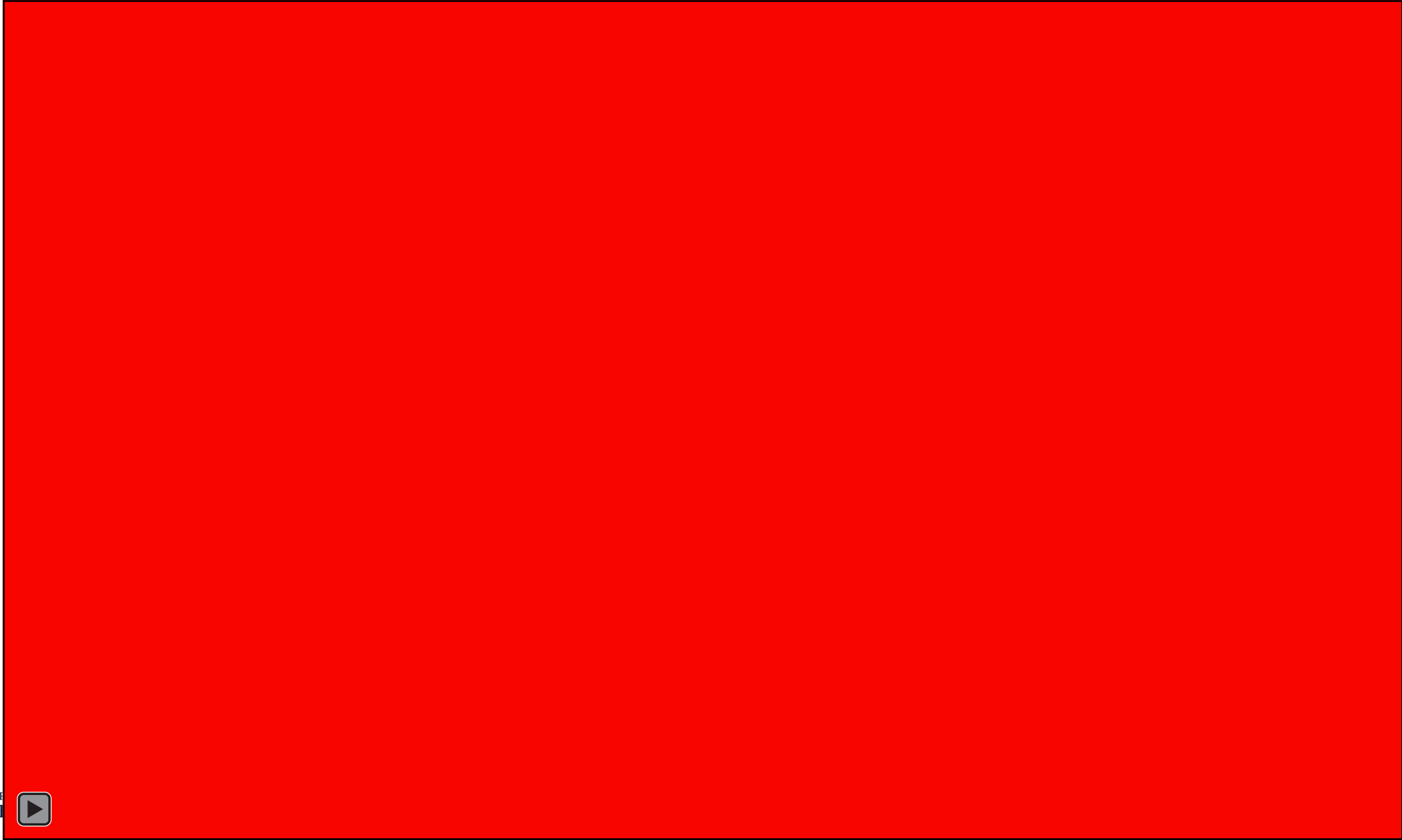
T_{h1} matches the setpoint using optimal control



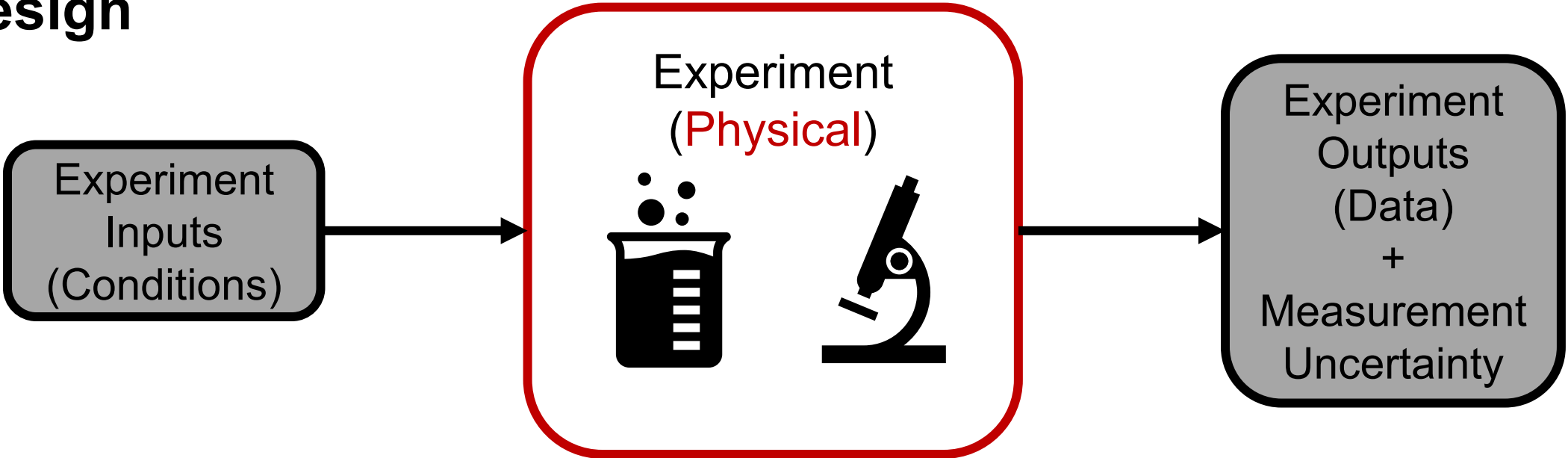
Teaching Digital Twins (MPC, State Estimation)



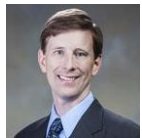
Teaching Digital Twins (MPC, State Estimation)



“Experiment” Abstraction Streamlines Closed-Loop Experiment Design



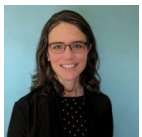
Dr. Bethany Nicholson



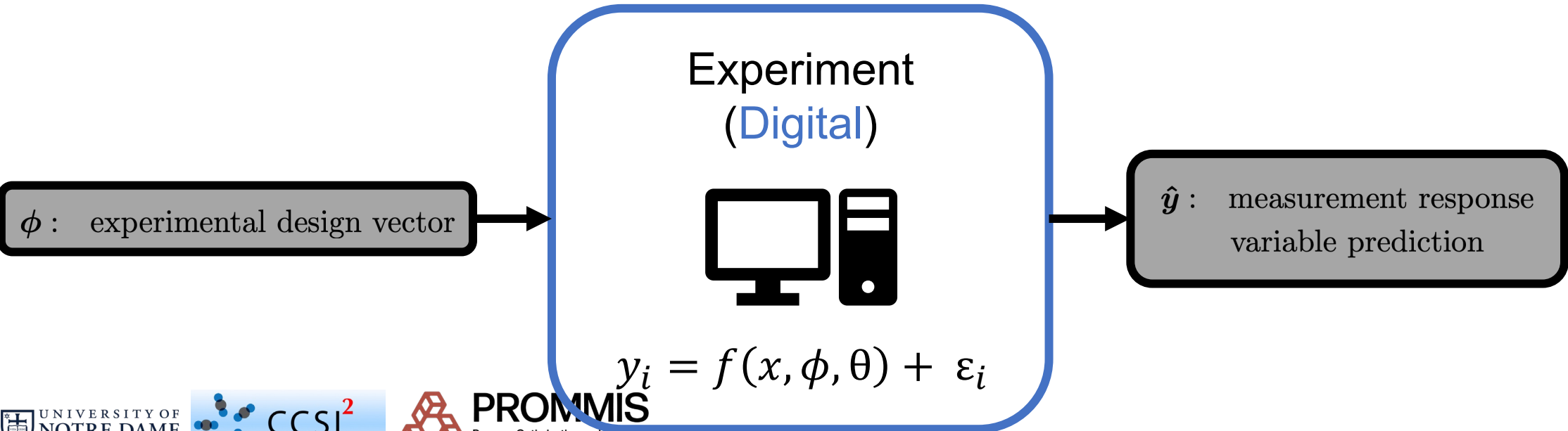
Dr. John Siirola



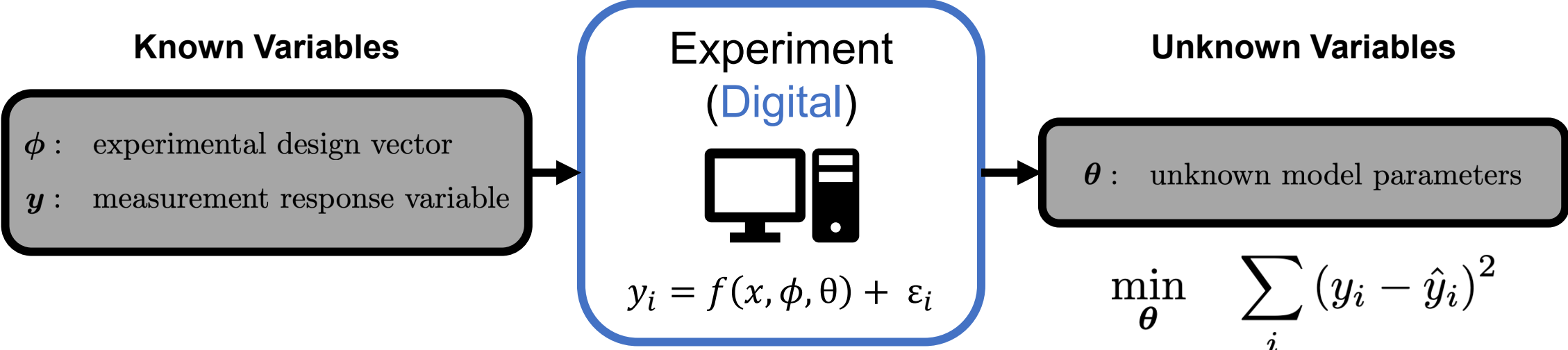
Dr. Shawn Martin



Katherine Klise



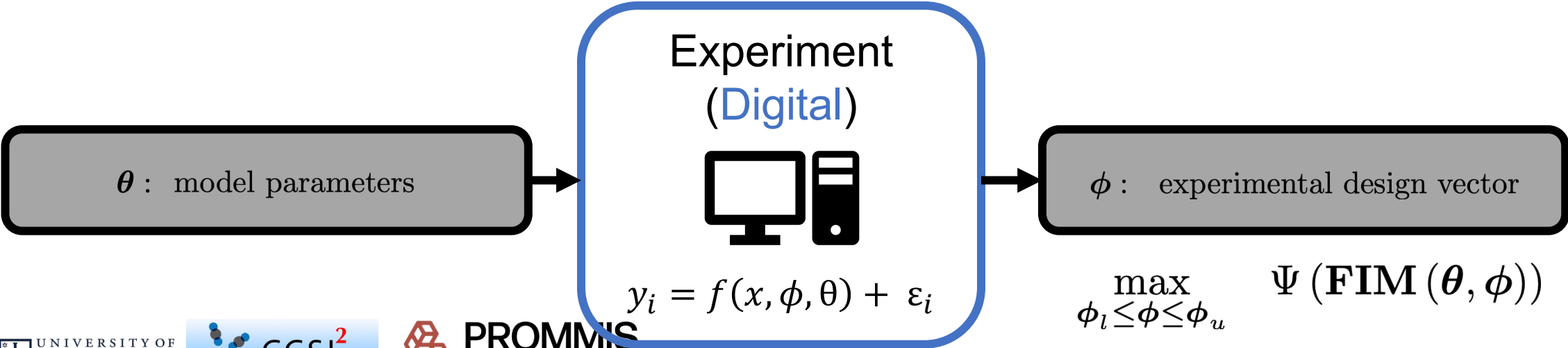
“Experiment” Abstraction Streamlines Closed-Loop Experiment Design



Dr. Bethany Nicholson



Dr. John Siirola



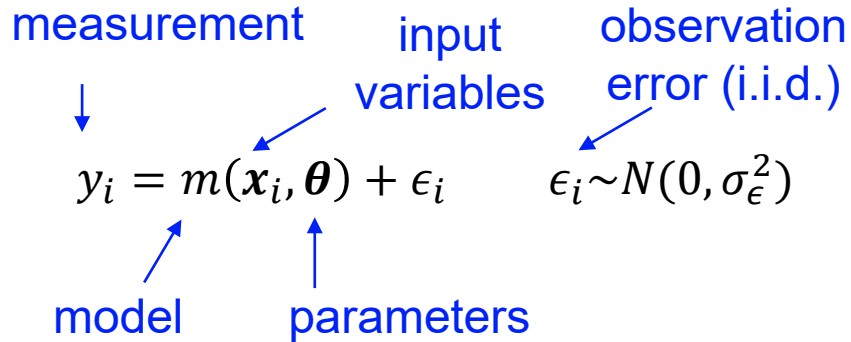
Dr. Shawn Martin



Katherine Klise

Parameter Estimation and Uncertainty Basics

Assume a model and error structure:



What values of model parameters θ best fit the data X and y ?

$$\hat{\theta} = \arg \min_{\theta} \Psi := \frac{1}{2} \sum_i [y_i - m(x_i, \theta)]^2$$

best fit estimates

How sensitive are the least-squares objective Ψ to perturbations in θ ?

$$H = \begin{bmatrix} \frac{\partial^2 \Psi}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \Psi}{\partial \theta_n \partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \Psi}{\partial \theta_1 \partial \theta_m} & \cdots & \frac{\partial^2 \Psi}{\partial \theta_m^2} \end{bmatrix} \quad Q(\theta) = \begin{bmatrix} \frac{\partial m(x_1, \theta)}{\partial \theta_1} & \cdots & \frac{\partial m(x_1, \theta)}{\partial \theta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial m(x_n, \theta)}{\partial \theta_1} & \cdots & \frac{\partial m(x_n, \theta)}{\partial \theta_m} \end{bmatrix}$$

Hessian matrix

$$H \approx Q^T Q$$

sensitivity matrix

How does measurement uncertainty ϵ propagate into uncertainty about the regressed parameters $\hat{\theta}$?

covariance matrix for $\hat{\theta}$

$$V_{\hat{\theta}} \approx \sigma_\epsilon^2 H^{-1} \approx \sigma_\epsilon^2 (Q^T Q)^{-1}$$

Fisher information matrix for $\hat{\theta}$

$$M_{\hat{\theta}} \approx V_{\hat{\theta}}^{-1} \approx \frac{1}{\sigma_\epsilon^2} (Q^T Q)$$

Extensions not shown: sophisticated error structures, Bayesian or MLE inference, ...

Bard (1974)
Bates and Watts (1988)
Pirnay, Lopez-Negrete, Biegler (2012)

TCLab: Eigendecomposition of the Fisher Information Matrix

ParmEst: `dowlinglab.github.io/pyomo-doe/notebooks/parmest.html`

FIM: `dowlinglab.github.io/pyomo-doe/notebooks/doe_exploratory_analysis.html`

FIM:

```
[[517225.40941304    1360.01262476 -66404.72541298   -1002.47319402]
 [   1360.01262476    5004.3737258    12379.2662576    5238.40389773]
 [-66404.72541298    12379.2662576    65481.16908635    14190.01468139]
 [-1002.47319402     5238.40389773    14190.01468139    5526.94375493]]
```

eigenvalues:

```
[5.26802218e+05  6.26035823e+04  3.83207978e+03  1.61037063e-02]
```

eigenvectors:

U_a	[[-9.89752804e-01	-1.35949591e-01	4.36702406e-02	-7.52086327e-05]	U_a
U_b	[8.63262440e-04	-2.26164575e-01	-6.85698047e-01	-6.91857665e-01]	U_b
$1/C_p^H$	[1.42671125e-01	-9.31600001e-01	3.33329462e-01	-2.56487437e-02]	$1/C_p^H$
$1/C_p^S$	[5.79584008e-03	-2.49977462e-01	-6.45602485e-01	7.21578207e-01]]	$1/C_p^S$

Difficult to uniquely estimate $U_b C_p^S$ with this experiment!

Model-Based DoE Optimization Formulation

$$\begin{array}{ll}
 \max_{\varphi} & \Psi[M(\hat{\theta}, \varphi)] \\
 \text{s. t.} & \dot{x}(t) = f(x(t), z(t), u(t), \bar{w}, \hat{\theta}) \\
 & g(x(t), z(t), u(t), \bar{w}, \hat{\theta}) = \mathbf{0} \\
 & y(t) = h(x(t), z(t), \hat{\theta}) \\
 & f^0(\dot{x}(t_0), x(t_0), z(t_0), u(t_0), \bar{w}, \hat{\theta}) = \mathbf{0} \\
 & g^0(x(t_0), z(t_0), u(t_0), \bar{w}, \hat{\theta}) = \mathbf{0} \\
 & y^0(t_0) = h(x(t_0), z(t_0), \hat{\theta})
 \end{array}
 \left. \begin{array}{l} \\ \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} \text{DAE System} \\ \\ \\ \text{Initial} \\ \text{Conditions} \end{array}
 \left. \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \right\} m(x(t), y(t), z(t), u(t), \bar{w}, \hat{\theta}) = \mathbf{0}$$

- y Measurements (model responses)
- $\hat{\theta}$ Estimated parameters
- x Time-dependent differential state variables
- z Time-dependent algebraic state variables
- u Time-varying control variables
- \bar{w} Time-invariant control variable

Fisher information matrix (FIM):

$$M \approx V_{\hat{\theta}}^{-1} \approx \sigma_{\epsilon}^{-2} H \approx \sigma_{\epsilon}^{-2} Q^T Q$$

MBDoe Decisions:

$$\varphi = (u(t), x(t_0), z(t_0), \bar{w}, \mathbf{t})$$

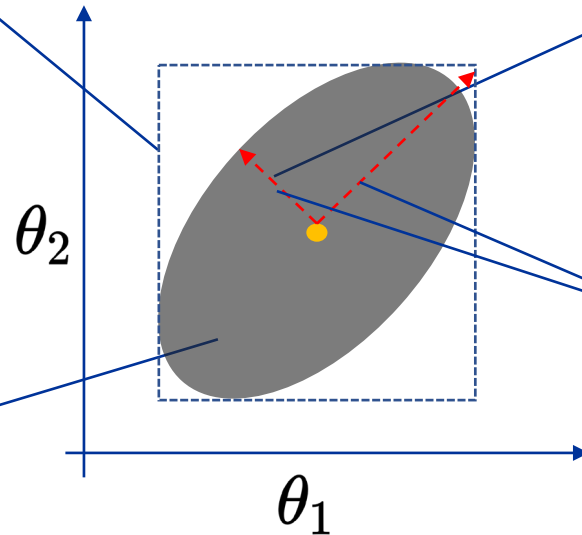
Alphabetic Design Criteria Measure Information Content

Figure adapted from: Franceschini, G., & Macchietto, S. (2008). *Chem. Eng. Sci.*, 63(19), 4846-4872.

A-optimality
max trace(\mathbf{M})
enclosing box volume
poor choice for highly correlated θ

D-optimality
max det(\mathbf{M})
ellipsoid volume
robust to linear transformations

confidence ellipsoid for
covariance matrix $\mathbf{V} = \mathbf{M}^{-1}$



E-optimality
max min(eig(\mathbf{M}))
major axis
recommended if \mathbf{M} is ill-conditioned

ME-optimality
min $\kappa(\mathbf{M}) = \max(\text{eig}(\mathbf{M})) / \min(\text{eig}(\mathbf{M}))$
ratio of major to minor axes
recommended if \mathbf{M} is ill-conditioned

Joint Parameter Precision and Model Discrimination

Alberton, A.L., Schwaab, M., Lobão, M.W.N. and Pinto, J.C., 2011. Experimental design for the joint model discrimination and precise parameter estimation through information measures. *Chemical Engineering Science*, 66(9), pp.1940-1952.

Galvanin, F., Cao, E., Al-Rifai, N., Gavriilidis, A. and Dua, V., 2016. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. *Computers & Chemical Engineering*, 95, pp.202-215.

Galvanin, F., Cao, E., Al-Rifai, N., Dua, V. and Gavriilidis, A., 2015. Optimal design of experiments for the identification of kinetic models of methanol oxidation over silver catalyst. *Chimica Oggi-Chemistry Today*, 33(3), pp.51-56.

Pankajakshan, A., Waldron, C., Quaglio, M., Gavriilidis, A. and Galvanin, F., 2019. A Multi-Objective Optimal Experimental Design Framework for Enhancing the Efficiency of Online Model Identification Platforms. *Engineering*, 5(6), pp.1049-1059.

Model Discrimination

Hunter, W.G. and Reiner, A.M., 1965. Designs for discriminating between two rival models. *Technometrics*, 7(3), pp.307-323.

Buzzi-Ferraris, G. and Forzatti, P., 1983. A new sequential experimental design procedure for discriminating among rival models. *Chemical engineering science*, 38(2), pp.225-232.

Ferraris, G.B., Forzatti, P., Emig, G. and Hofmann, H., 1984. Sequential experimental design for model discrimination in the case of multiple responses. *Chemical engineering science*, 39(1), pp.81-85.

Pyomo.DoE Formulation: MBDoE as 2-Stage Program

max $\log \det(\mathbf{M}(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi})) = 2 \sum_{i=1}^{N_p} \log L_{ii}$ **D-optimality**

s.t. $\mathbf{M} = \sum_r \sum_{r'} \tilde{\sigma}_{r,r'} \mathbf{Q}_r^T \mathbf{Q}_{r'}$ **Stage 1**

$\mathbf{M} = \mathbf{L}\mathbf{L}^T, L_{ii} \geq \epsilon$ **Cholesky factorization**

$q_{r,p}(t) = \frac{y_{r,p}^+(t) - y_{r,p}^-(t)}{2\epsilon_p}$ **Central finite difference**

$\mathbf{m}(x_p^+(t), y_p^+(t), z_p^+(t), \mathbf{u}(t), \bar{\mathbf{w}}, \boldsymbol{\theta}_p^+) = \mathbf{0}$ **Two model evaluations**

$\mathbf{m}(x_p^-(t), y_p^-(t), z_p^-(t), \mathbf{u}(t), \bar{\mathbf{w}}, \boldsymbol{\theta}_p^-) = \mathbf{0}$ **Two model evaluations**

$\boldsymbol{\theta}_p^+ = \hat{\boldsymbol{\theta}} + \mathbf{e}_p \epsilon_p$ **Up and down perturbations**

$\boldsymbol{\theta}_p^- = \hat{\boldsymbol{\theta}} - \mathbf{e}_p \epsilon_p$ **Up and down perturbations**

Stage 2

$\forall p \in \{1, \dots, N_p\}$

Model Sensitivity

$$\mathbf{Q}_r = \begin{bmatrix} \frac{\partial y_r(t_1)}{\partial \theta_1} & \dots & \frac{\partial y_r(t_1)}{\partial \theta_{N_p}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_r(t_n)}{\partial \theta_1} & \dots & \frac{\partial y_r(t_n)}{\partial \theta_{N_p}} \end{bmatrix} = [\mathbf{q}_{r,1} \quad \dots \quad \mathbf{q}_{r,N_p}]$$

$$\mathbf{q}_{r,p} = \left[\frac{\partial y_r(t_1)}{\partial \theta_p} \quad \dots \quad \frac{\partial y_r(t_n)}{\partial \theta_p} \right]^T$$

- \mathbf{y} Measurements (model responses)
- \mathbf{Q}_r Dynamic sensitivity for response r
- $\mathbf{m}(\cdot)$ DAE model
- $\hat{\boldsymbol{\theta}} \in \mathbb{R}^P$ Estimate for parameters
- $\mathbf{M} \in \mathbb{R}^{P \times P}$ Fisher information matrix
- $\mathbf{L} \in \mathbb{R}^{P \times P}$ Lower triangular Cholesky factorization
- ϵ_p Small perturbation for parameter p
- $\mathbf{e}_p \in \mathbb{R}^P$ Unit vector with "1" in position p

Temperature Control Lab (TC-Lab) – Closed-Loop Experimental Design with New Experiment Abstraction

Parameter Estimation

$$C_p^H \frac{dT_{H,1}}{dt} = U_a(T_{amb} - T_{H,1}) + U_b(T_{S,1} - T_{H,1}) + \alpha P_1 u_1$$

$$C_p^S \frac{dT_{S,1}}{dt} = U_b(T_{H,1} - T_{S,1})$$

$\theta \equiv \{C_P^H, C_P^S, U_a, U_b\}$
 $y \equiv \{T_{S,1}\}$
 $\phi \equiv \{u_1\}$



Optimal Experiment Design/Optimal Control

Digital Components ■

Physical Components ■

Physical System

TC Lab: DoE, Exploratory Analysis

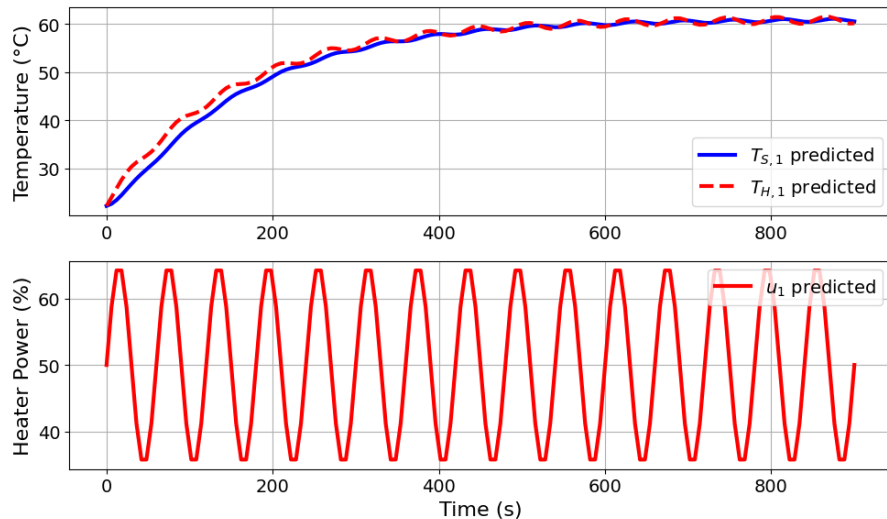
dowlinglab.github.io/pyomo-doe/notebooks/doe_exploratory_analysis.html

Sensitivity of the FIM to experimental design.

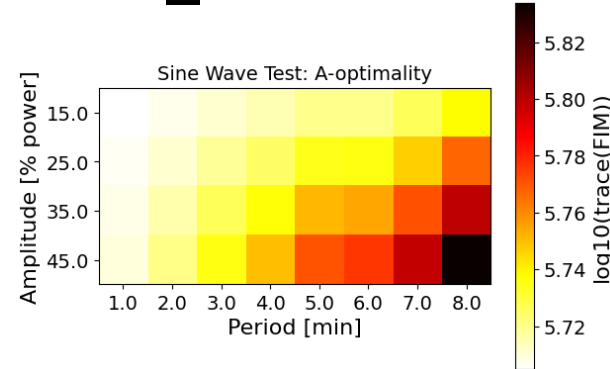
Example: Sine Wave

Vary:

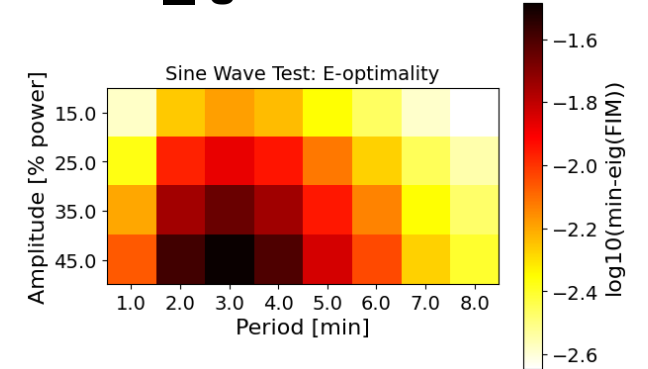
- period (from 1 to 8 minutes)
- amplitude (from 15% to 50%)



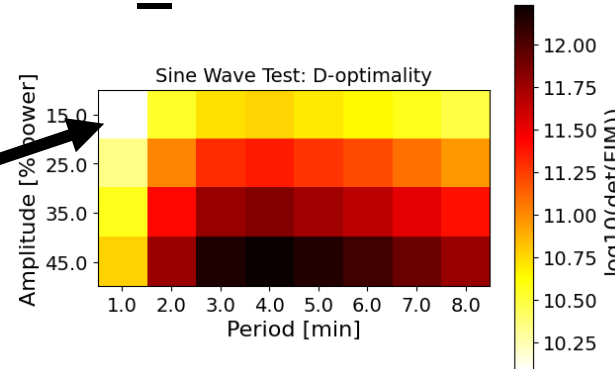
trAce



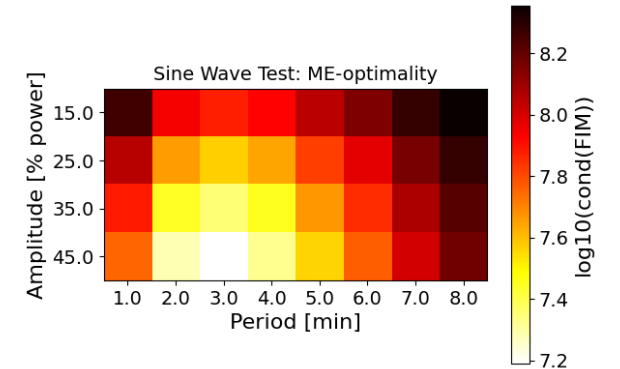
Min. Eigenvalue



Determinant

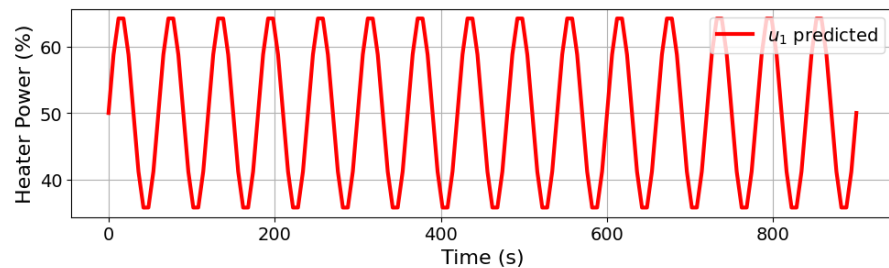
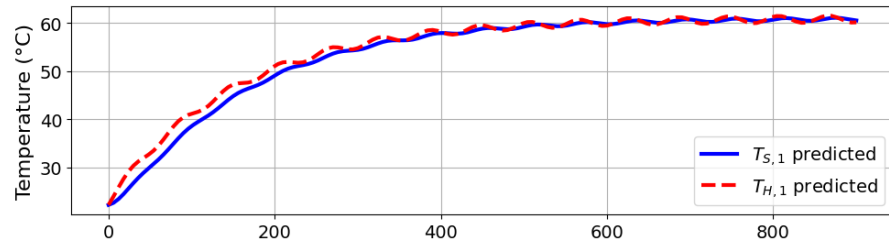
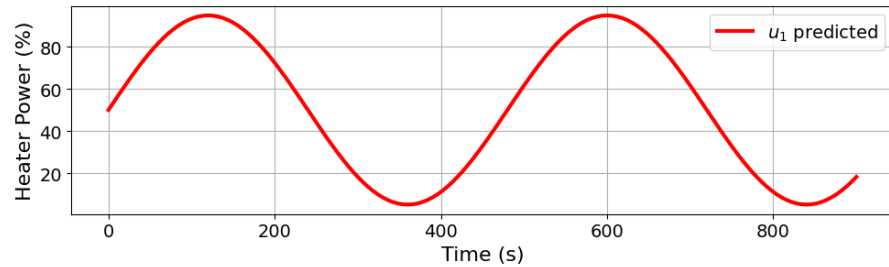
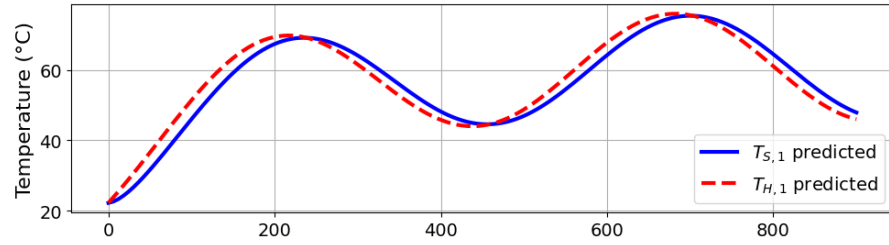


Condition Number

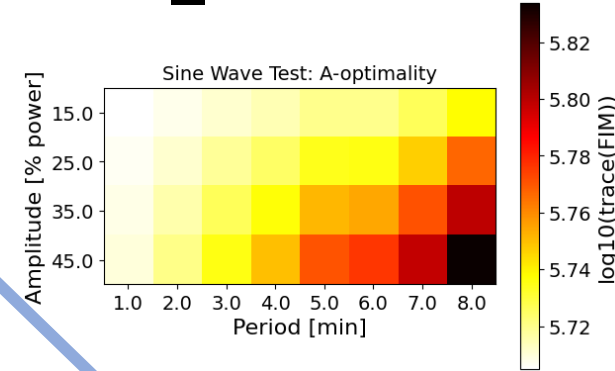


TC Lab: DoE, Exploratory Analysis

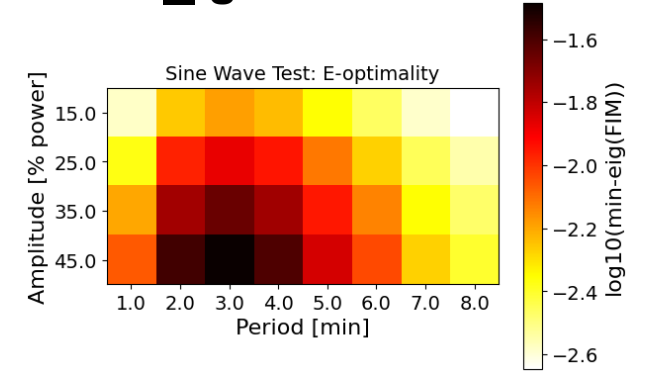
dowlinglab.github.io/pyomo-doe/notebooks/doe_exploratory_analysis.html



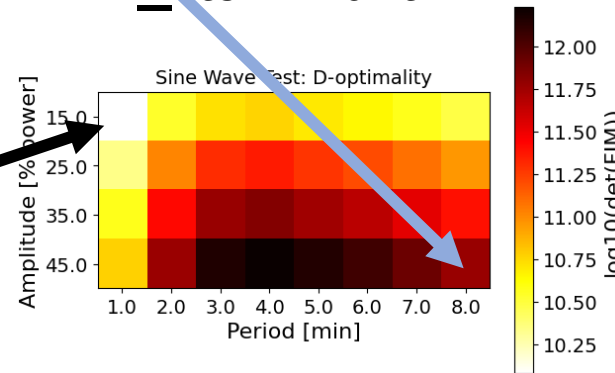
trAce



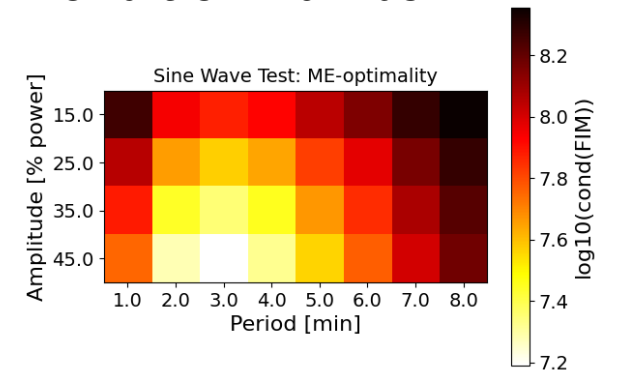
Min. Eigenvalue



Determinant



Condition Number



TC Lab: A-Optimal Next Experiment

dowlinglab.github.io/pyomo-doe/notebooks/doe_optimize.html

$$\max_u \log \text{trace}(\mathbf{M}(u) + \mathbf{M}_0)$$

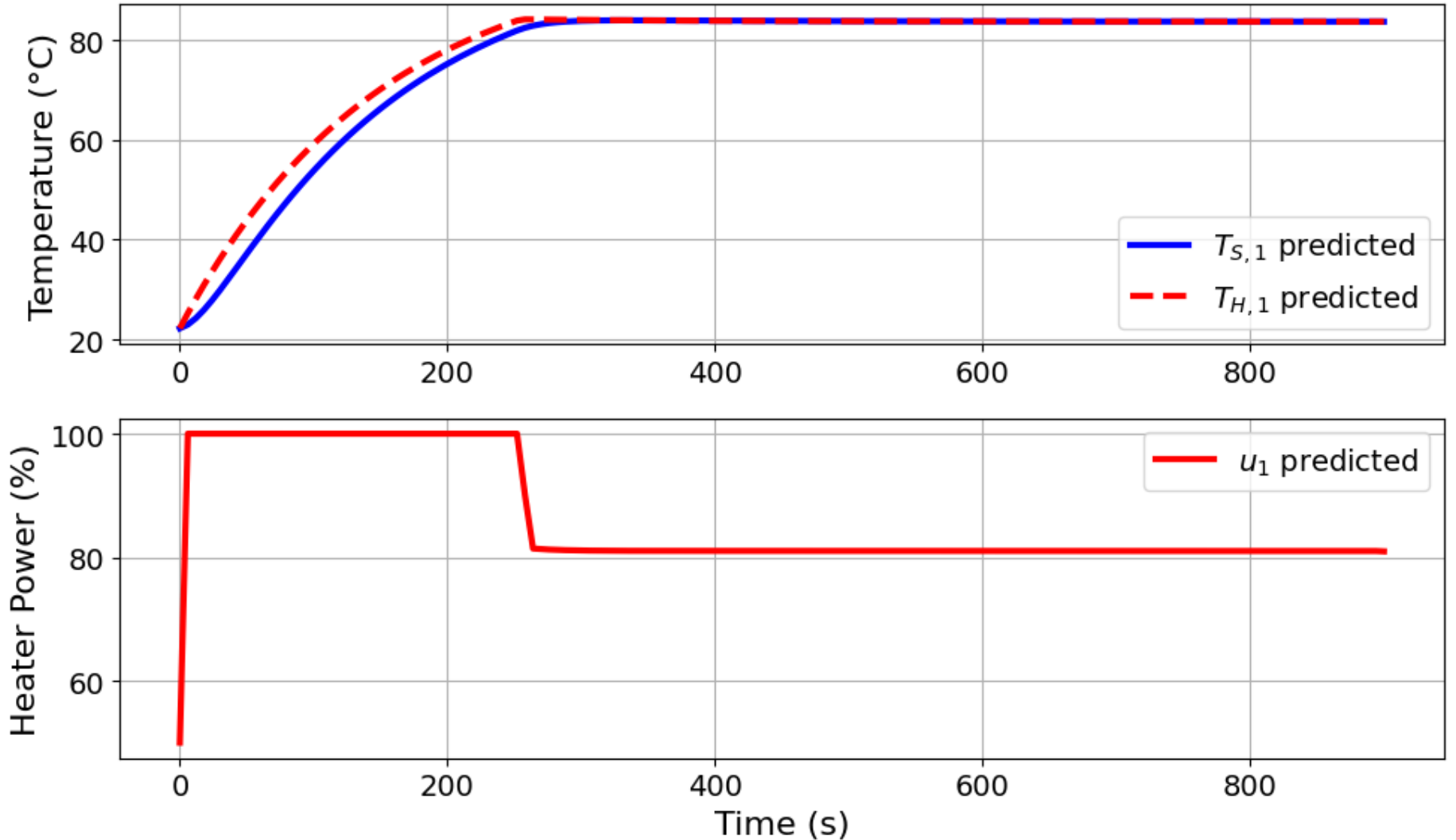
$$\text{s.t. } C_p^H \frac{dT_H}{dt} = \dots$$

$$C_p^S \frac{dT_S}{dt} = \dots$$

$$0\% \leq u(t) \leq 100\%$$

$$T_H(t_0) = T_{amb}$$

$$T_S(t_0) = T_{amb}$$



TC Lab: D-Optimal Next Experiment

dowlinglab.github.io/pyomo-doe/notebooks/doe_optimize.html

$$\max_u \quad \log \det(\mathbf{M}(u) + \mathbf{M}_0)$$

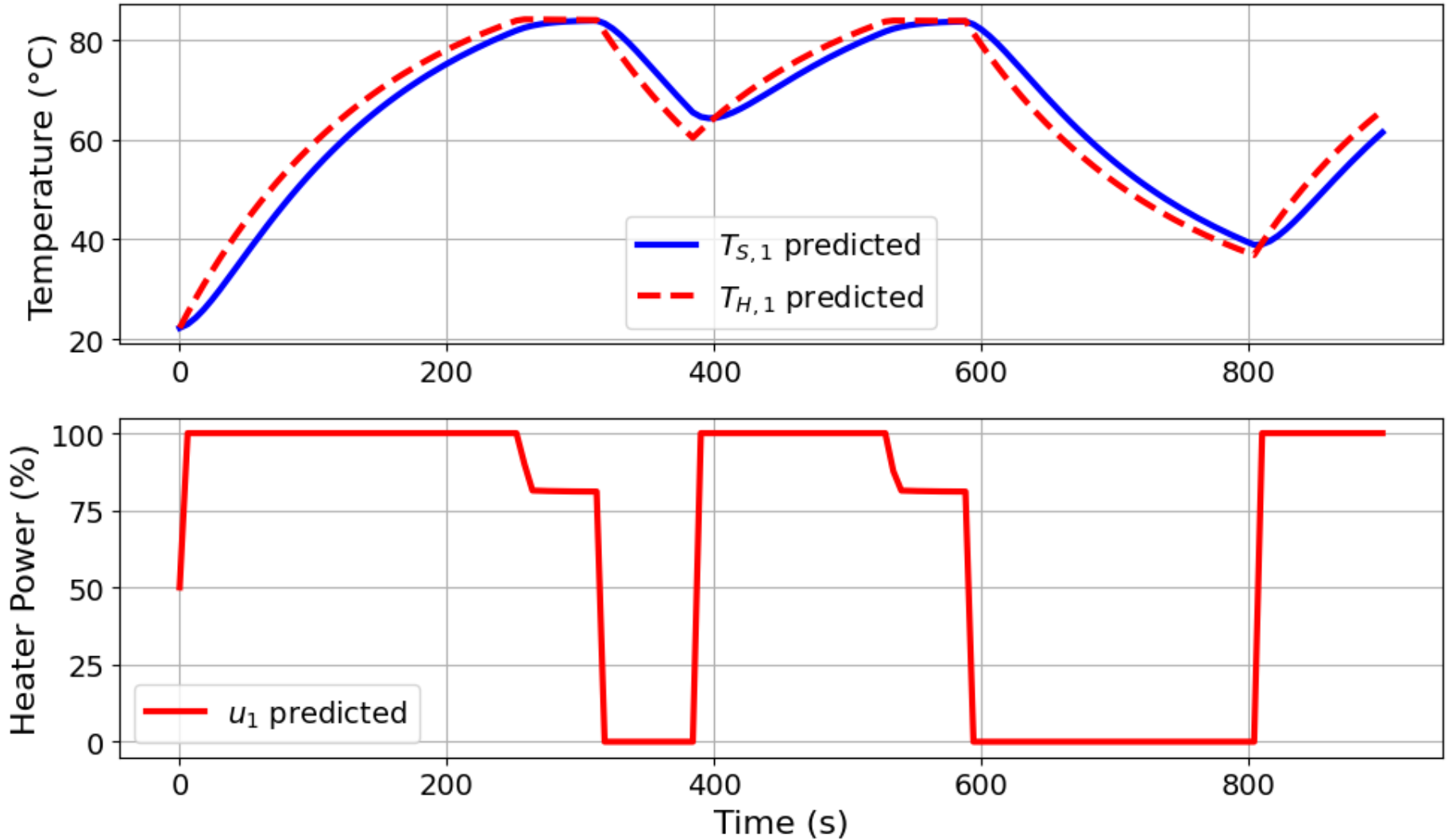
$$\text{s.t.} \quad C_p^H \frac{dT_H}{dt} = \dots$$

$$C_p^S \frac{dT_S}{dt} = \dots$$

$$0\% \leq u(t) \leq 100\%$$

$$T_H(t_0) = T_{amb}$$

$$T_S(t_0) = T_{amb}$$



$U_b C_p^S$ is more estimable with two experiments (sine wave test, D-optimal)! 23

Getting Started with Pyomo.DoE

Documentation: https://pyomo.readthedocs.io/en/stable/contributed_packages/doe/doe.html

Community Detection for Pyomo models

- Pyomo.DoE
 - Methodology Overview
 - Pyomo.DoE Required Inputs
 - Pyomo.DoE Solver Interface
 - Pyomo.DoE Usage Example
- GDPopt logic-based solver
- Infeasible Irreducible System (IIS) Tool
- Incidence Analysis
- MindtPy Solver
- MPC
- Multistart Solver
- Nonlinear Preprocessing Transformations
- Parameter Estimation with parmest
- PyNumero
- PyROS Solver
- Sensitivity Toolbox
- Trust Region Framework Method Solver
- MC++ Interface
- z3 SMT Sat Solver Interface

Read the Docs v: stable

[Home](#) / [Third-Party Contributions](#) / Pyomo.DoE [Edit on GitHub](#)

Pyomo.DoE

Pyomo.DoE (Pyomo Design of Experiments) is a Python library for model-based design of experiments using science-based models.

Pyomo.DoE was developed by **Jialu Wang** and **Alexander W. Dowling** at the University of Notre Dame as part of the **Carbon Capture Simulation for Industry Impact (CCSI2)** project, funded through the U.S. Department Of Energy Office of Fossil Energy.

If you use Pyomo.DoE, please cite:

[Wang and Dowling, 2022] Wang, Jialu, and Alexander W. Dowling. "Pyomo.DOE: An open-source package for model-based design of experiments in Python." *AICHE Journal* 68.12 (2022): e17813. <https://doi.org/10.1002/aic.17813>

Methodology Overview

Model-based Design of Experiments (MBDoe) is a technique to maximize the information gain of experiments by directly using science-based models with physically meaningful parameters. It is one key component in the model calibration and uncertainty quantification workflow shown below:

```
graph LR; A([Prior knowledge, preliminary data]) --> B[Model]; B --> C[Exploratory analysis]; C --> D[Parameter estimation]; D --> E[Uncertainty analysis]; E --> F([Model with quantified uncertainty]); C --> B; E --> B;
```


ParmEst and Pyomo.DoE Development Plans

Coming soon:

- Improved initialization
- Improved optimization performance
 - NLP decomposition
 - Grey-box objective calculations
- Improved modeling abstraction
 - Multiple experiments (e.g., planning batches)
 - Parameter uncertainty
- More applications, examples, and collaborations
- End-to-end uncertainty workflow (via interface with PyROS)



Acknowledgements

Pyomo.DoE development: *We graciously acknowledge funding from the U.S. Department of Energy, Office of Fossil Energy and Carbon Management, through the Carbon Capture Program (CCSI²) and Office of Resource Sustainability (PrOMMiS).*

This project was funded by the Department of Energy, National Energy Technology Laboratory an agency of the United States Government, through a support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

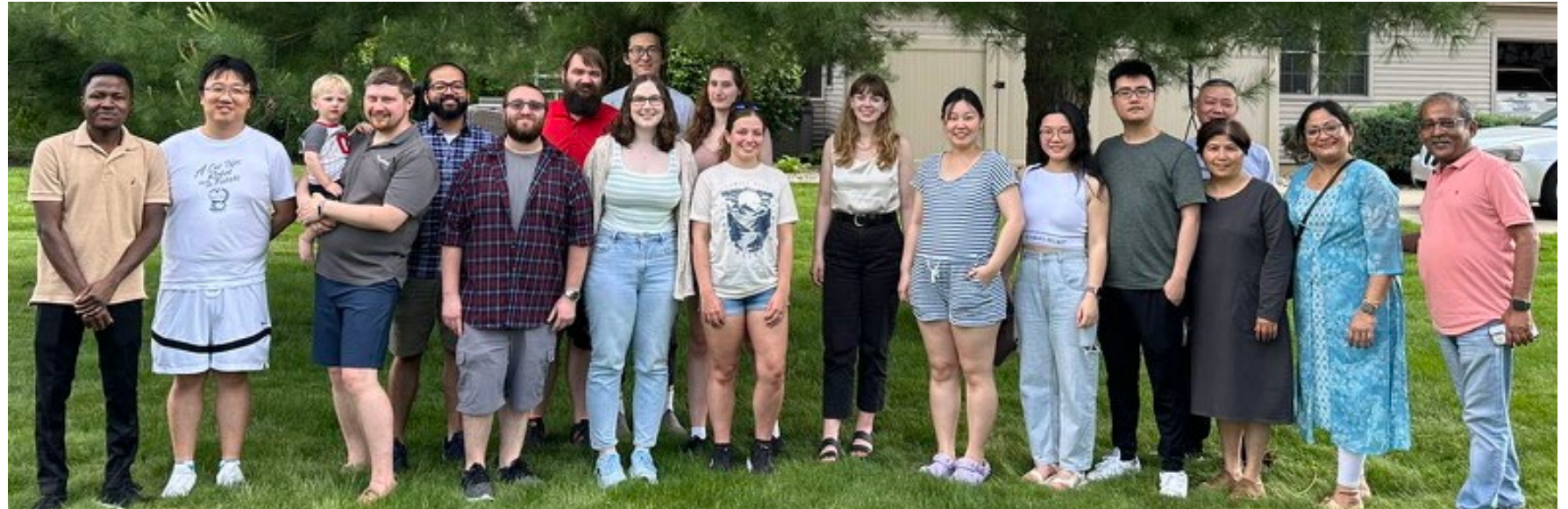


Pyomo Team (SNL):

- Bethany Nicholson
- John Sirola
- Miranda Mundt

Contributors (ND):

- Dr. Jialu Wang
- Dr. Dan Laky
- Hailey Lynch



Thank you to Prof. Jeff Kantor (1954-2023) for the TCLab examples and so much more.