Science-based Design of Experiments

Alexander (Alex) Dowling

adowling@nd.edu

October 12, 2023

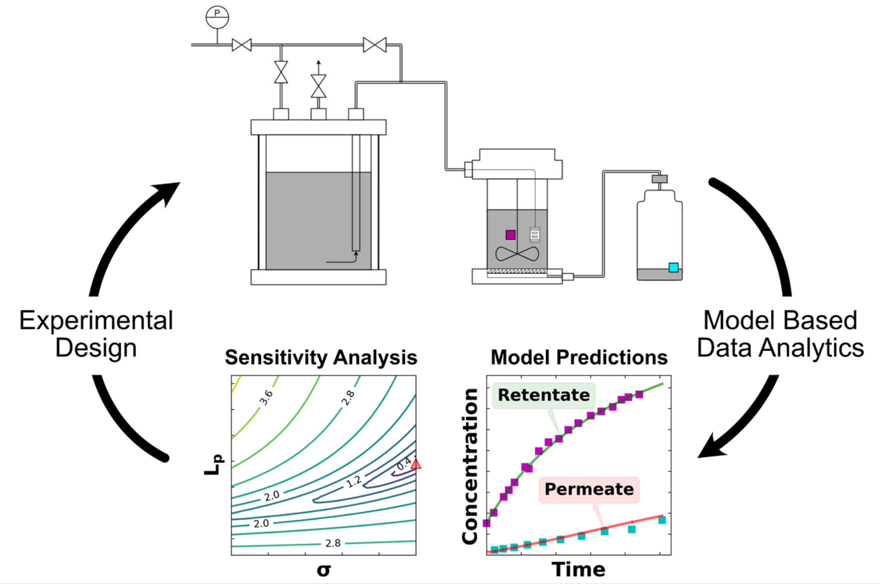
Advanced PSE Workshop, Washington D.C.

Presentation Goals

1. When to use science-based (model-based) design of experiment?

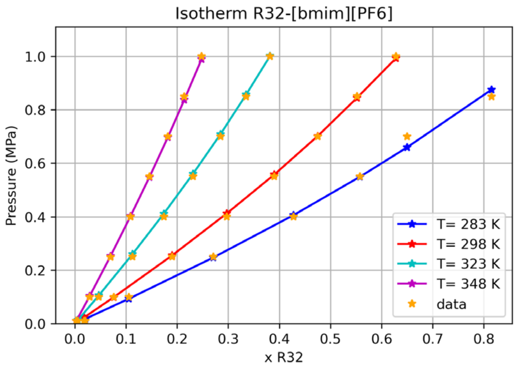
2. Illustrate key ideas of SBDoE using membrane example

3. Summarize `Pyomo.DoE` and development plan

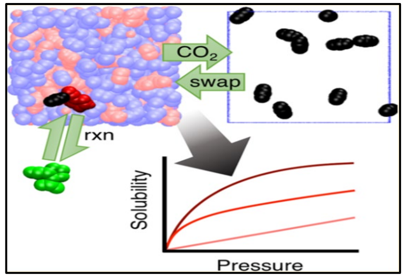


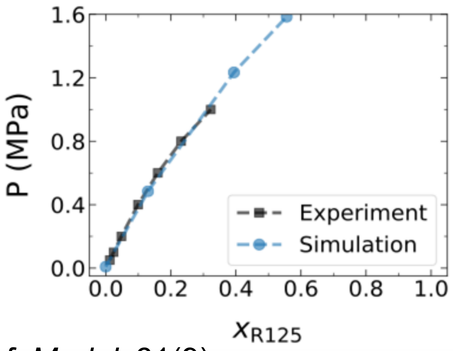
What data are most valuable to optimize systems?

Experimental thermophysical property measurements



Molecular simulation predictions





Befort et al (2021), *J. Chem. Inf. Model*, 61(9), 440-4414.

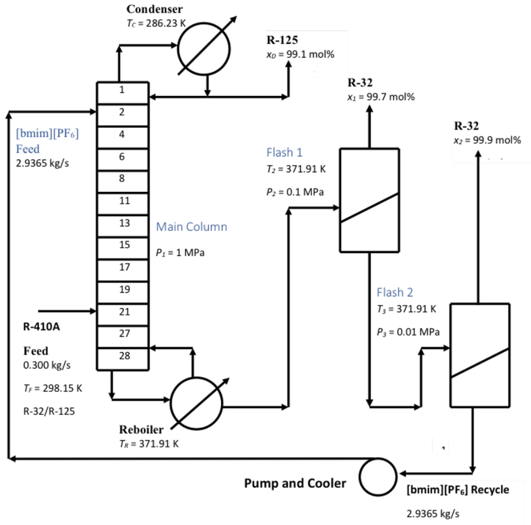
What uncertainties are acceptable?

Which model parameters to calibration from data?

How can optimization help straighten this tortuous path?

System optimization, e.g., design, operation, control

$$\begin{aligned} \min \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0 \\ & h(x,y) = 0 \\ & x \in \mathbb{R}^n, \; y \in \mathbb{Z}^m \end{aligned}$$



Shiflett, M. B. and Yokozeki, A. *Chimica Oggi - Chemistry Today*, 24(2). (2006).
Devaki, S. J. and Sasi, R. *IntechOpen*. (2007).

Design of Experiments ≠ One-Factor-at-a-Time

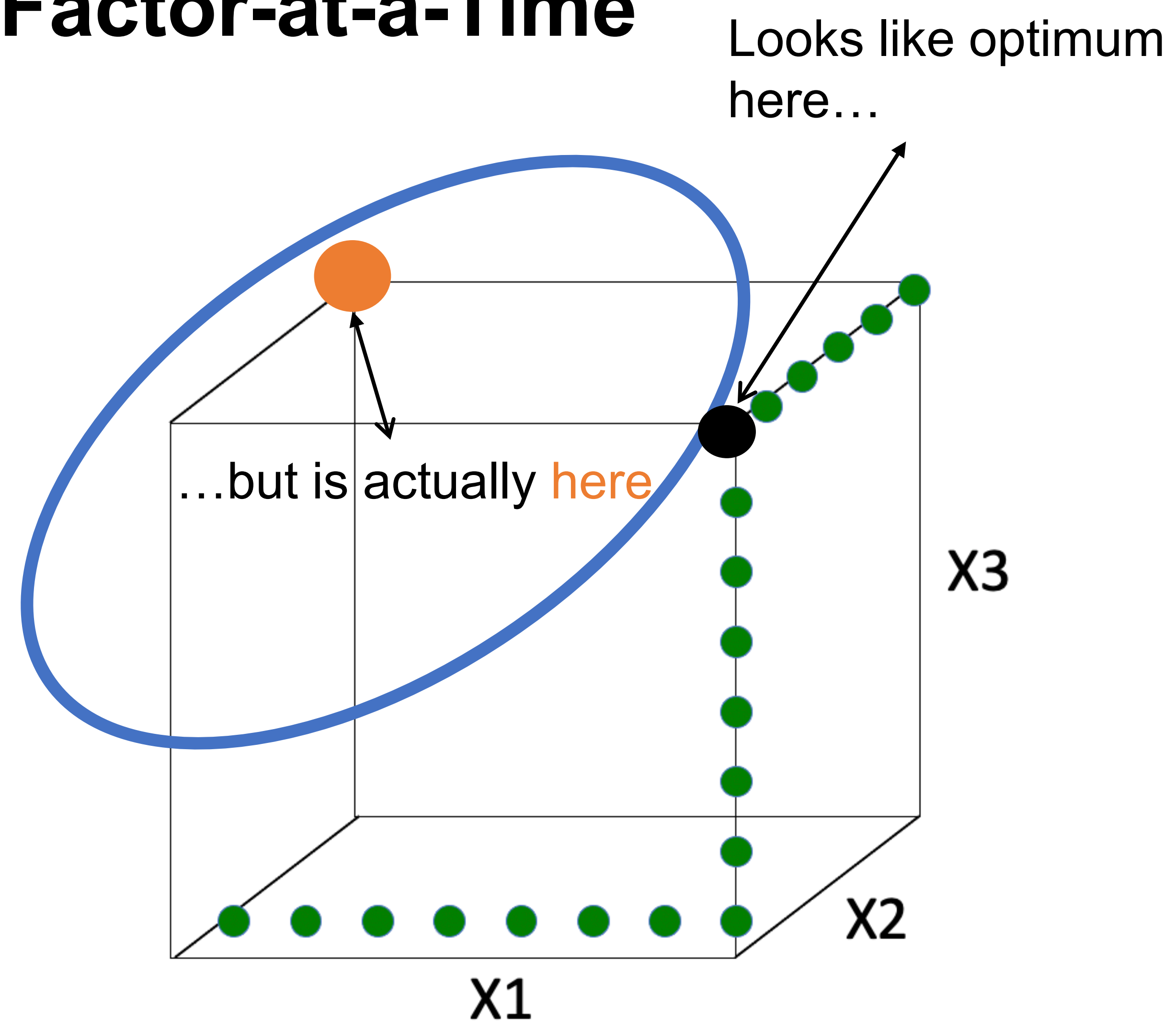
OFAAT strategy:

- Change only one input (factor) at a time
- Hold all others constant

Inefficient use of budget

Cannot identify interactions

- Effect of one factor changes when another factor changes
- Finding optimal operating conditions is unlikely



Design of Experiments ≠ One-Factor-at-a-Time

OFAAT strategy:

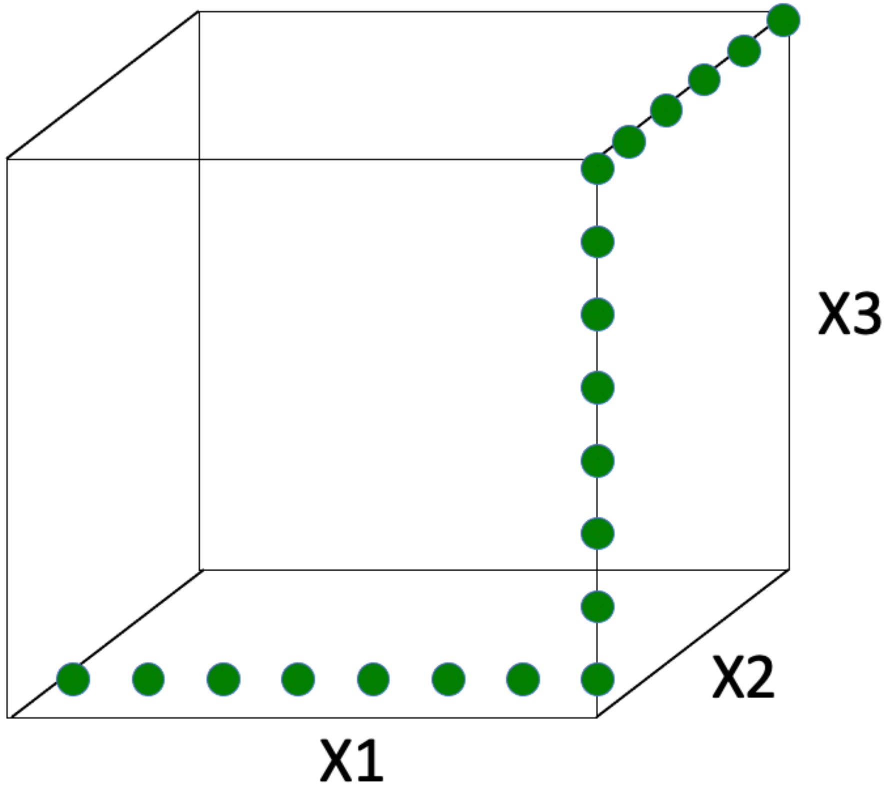
– Change only one input (factor) at a time
– Hold all others constant

Inefficient use of budget

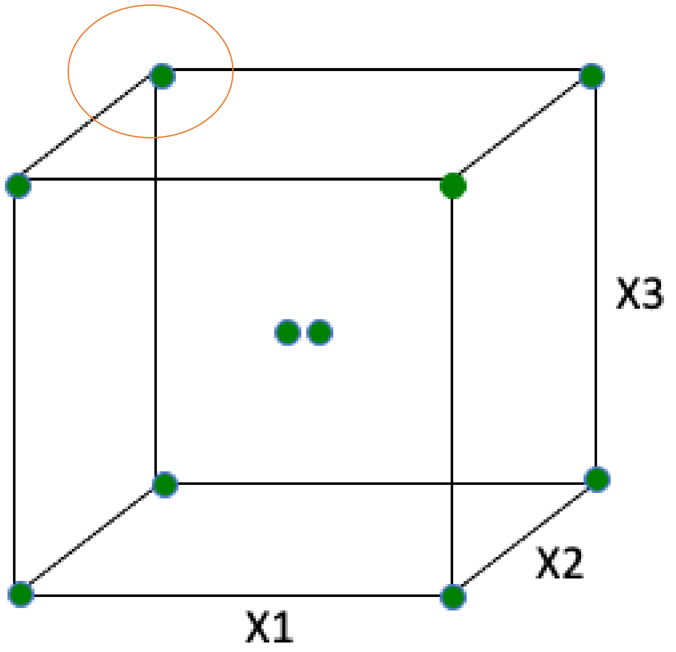
Cannot identify interactions

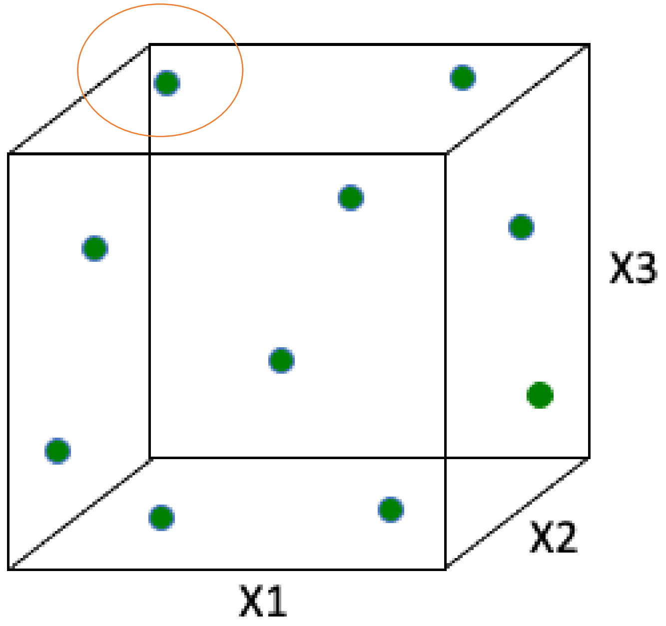
Not randomized

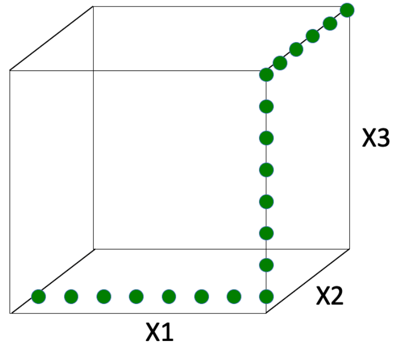
– Changing conditions can negatively affect the results



Sequential DoE Avoids These Drawbacks and Is Always More Efficient







Uses **20 runs**

Two Different Sequential DoE Approaches
Each uses **10 runs**

Power of Adaptive Sequential Optimal Experiments

Self-Driving Laboratories

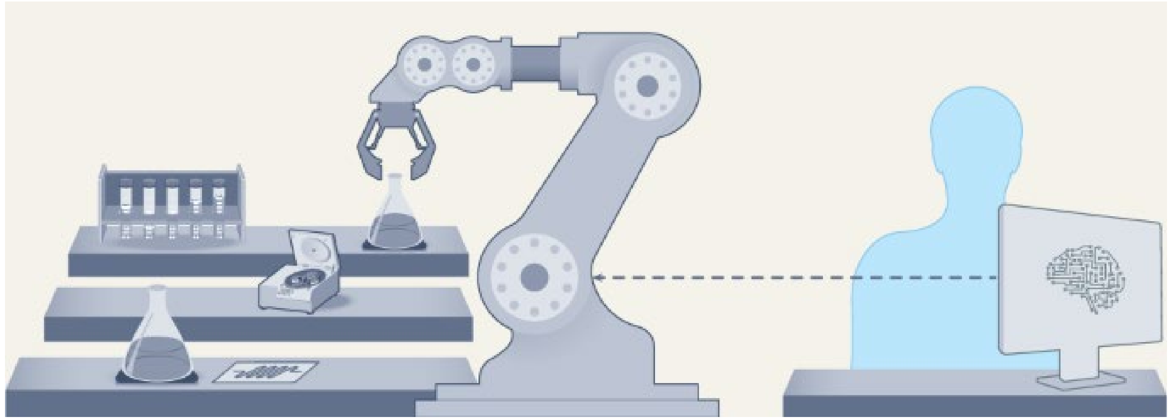
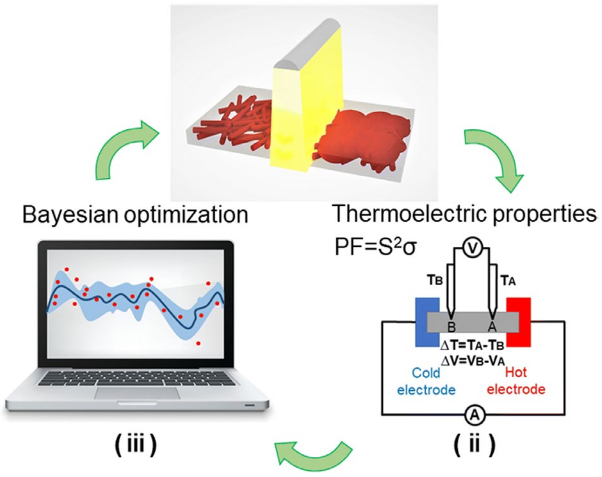


Figure: Abolhasani & Kumacheva (2023), *Nature Syn.*

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.*

Bayesian Optimization



Saeidi-Javash et al. (2022), *Eng. Env. Sci*

CO_2 Capture Technology Scale-Up

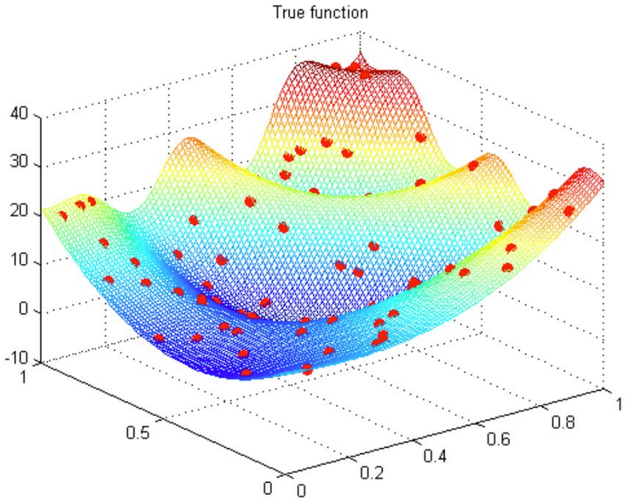
Morgan et al (2020), *Applied Energy*

Photo: National Carbon Capture Center (AL)

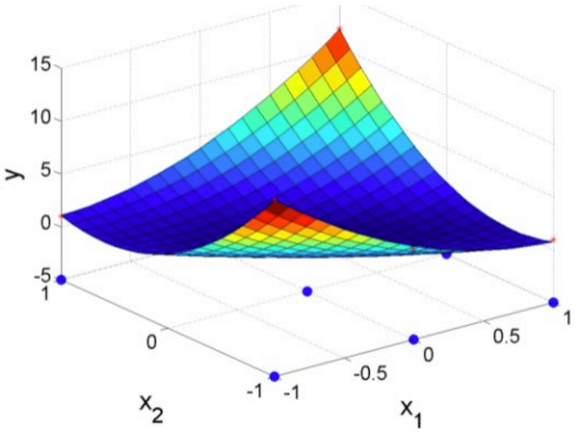
Optimizing Computational & Physical Experiments

Classical DoE

Space Filling



Response Surface

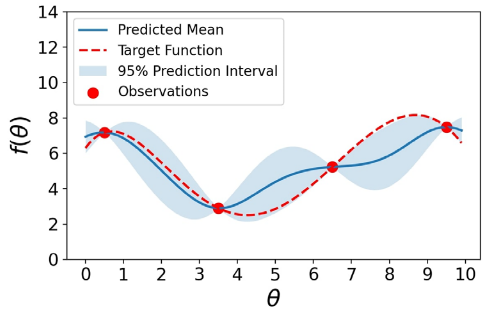


+ Decades of success,
  staple of applied statistics

- Static/fixed design that is not updated
  as data are collected

Modern (Adaptative) DoE

Bayesian Optimization



with a data-driven model

Model-based

$$\rho\, C_p \frac{\partial T(t,\, X)}{\partial t} + \rho\, C_p\, V \cdot \nabla T(t, X) + \nabla \cdot (-k\, \nabla T(t, X)) = Q$$

with a scientific model

Left figures from https://www.datadvance.net/product/pseven-core/generic-tool-for-design-of-experiments/
Right figure: Adapted from Wang and Dowling (2022) by M. Carlozo.

Science-based (Model-Based) DoE History

Jialu Wang

Steady-state system
1959

Dynamic systems,
time-invariant decisions
1974

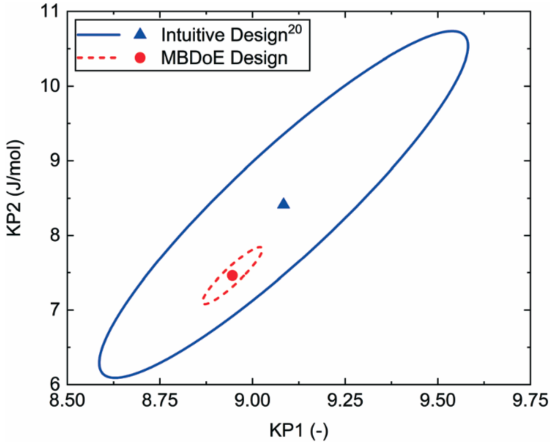
Dynamic systems,
Dynamic decisions
1994

1935
Factorial methods

Kinetics applications

1989, Biological applications

MBDoE parameter confidence regions are…
- 84% smaller than intuitive design
- 40% smaller than factorial design



Waldron, C., et al.., 2020, *Reaction Chemistry & Engineering*, *5*(1), pp.112-123.

Robust Formulations:

Bruwer, M.J. and MacGregor, J.F., 2006. Robust multi-variable identification: Optimal experimental design with constraints. *Journal of Process Control*, *16*(6), pp.581-600.

Dette, H., Melas, V.B., Pepelyshev, A. and Strigul, N., 2005. Robust and efficient design of experiments for the Monod model. *Journal of theoretical biology*, *234*(4), pp.537-550.

Online (Automated) MBDoE:

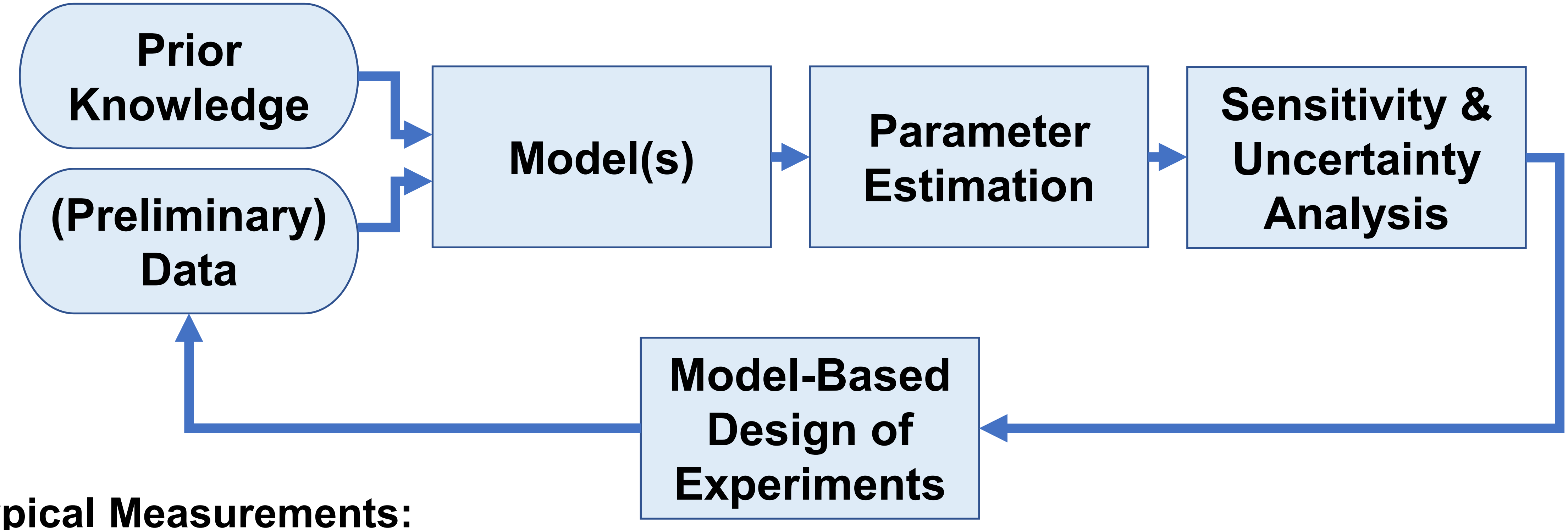
Quaglio, M., Waldron, C., Pankajakshan, A., Cao, E., Gavriilidis, A., Fraga, E.S. and Galvanin, F., 2019. An online reparametrisation approach for robust parameter estimation in automated model identification platforms. *Computers & Chemical Engineering*, *124*, pp.270-284.

Waldron, C., Pankajakshan, A., Quaglio, M., Cao, E., Galvanin, F. and Gavriilidis, A., 2019. Closed-loop model-based design of experiments for kinetic model discrimination and parameter estimation: Benzoic acid esterification on a heterogeneous catalyst. *Industrial & Engineering Chemistry Research*, *58*(49), pp.22165-22177.

Recommended Review:

Franceschini, G. and Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), pp.4846-4872.

Science-based Modeling Workflow

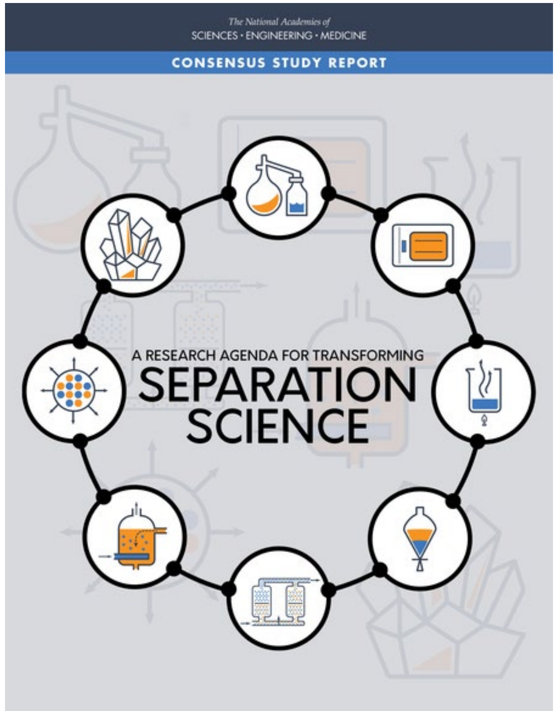


Typical Measurements:
- Pressures
- Temperature
- Composition
- Flowrate

Typical Models: Partial Differential Algebraic Equations

Transport Thermodynamics Reactions

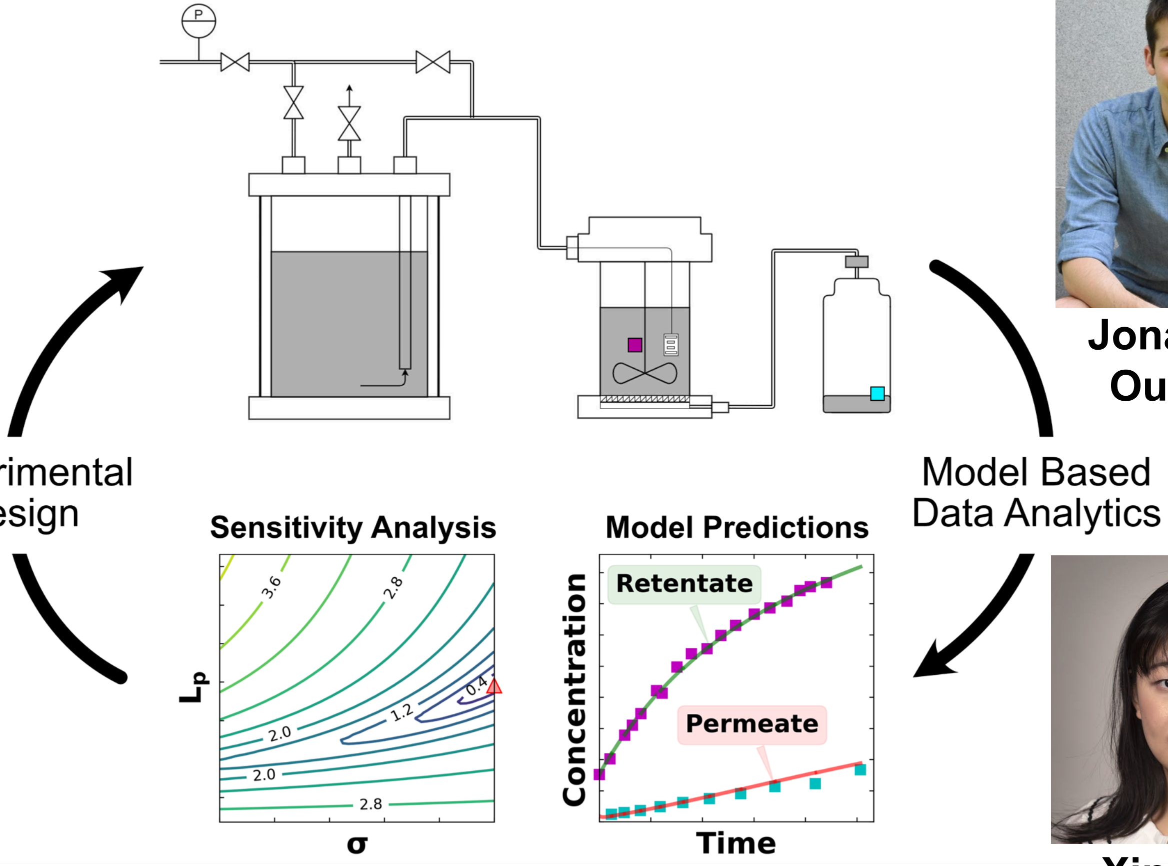
DATA: Diafiltration Apparatus for high-Throughput Analysis



*The research community should use **data science**, modeling, and simulation with **experimental measurements** to develop a fundamental understanding of separation materials in **complex environments** and at multiple scales.*

William Phillip

Alexander Dowling



Jonathan Ouimet

Xinhong Liu

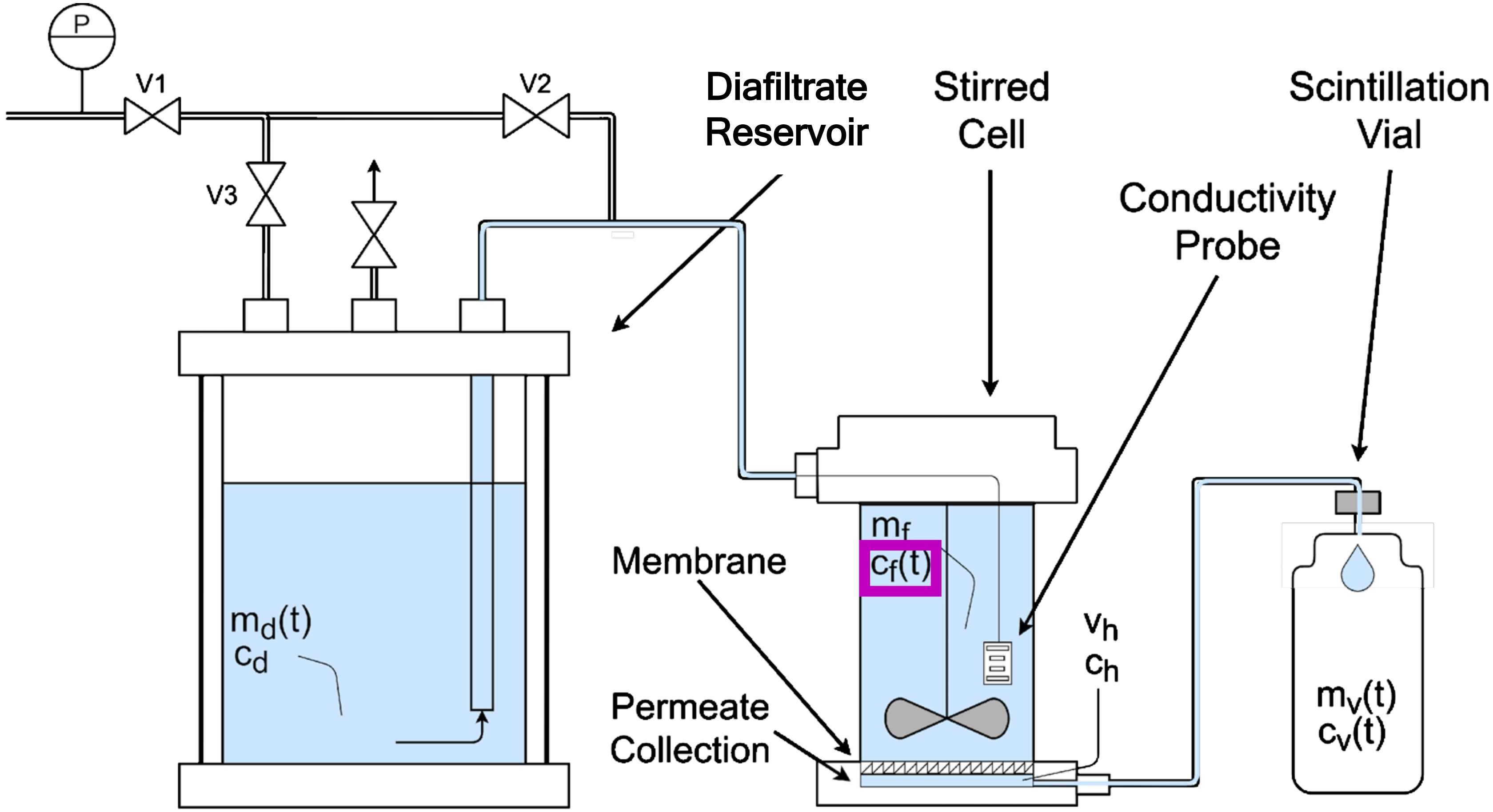
How to develop, accelerate, and improve the membrane characterization technique?

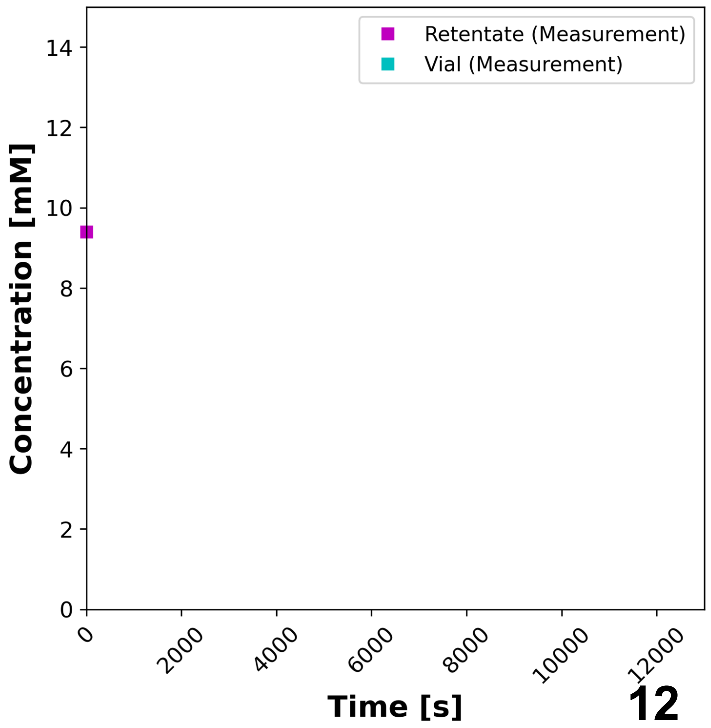
How to rapidly and precisely elucidate governing phenomena and mechanisms?

[Left Fig.] National Academies of Sciences, Engineering, and Medicine. (2019). The National Academies Press.
[Right Fig.] Ouimet, J. A. et. al., (2022). Journal of Membrane Science, 641, 119743.

Dynamic Diafiltration Experiment Sweeps Larger Concentration Space

Experimental degrees of freedom: temperature, pressure, feed concentration, diafiltration concentration

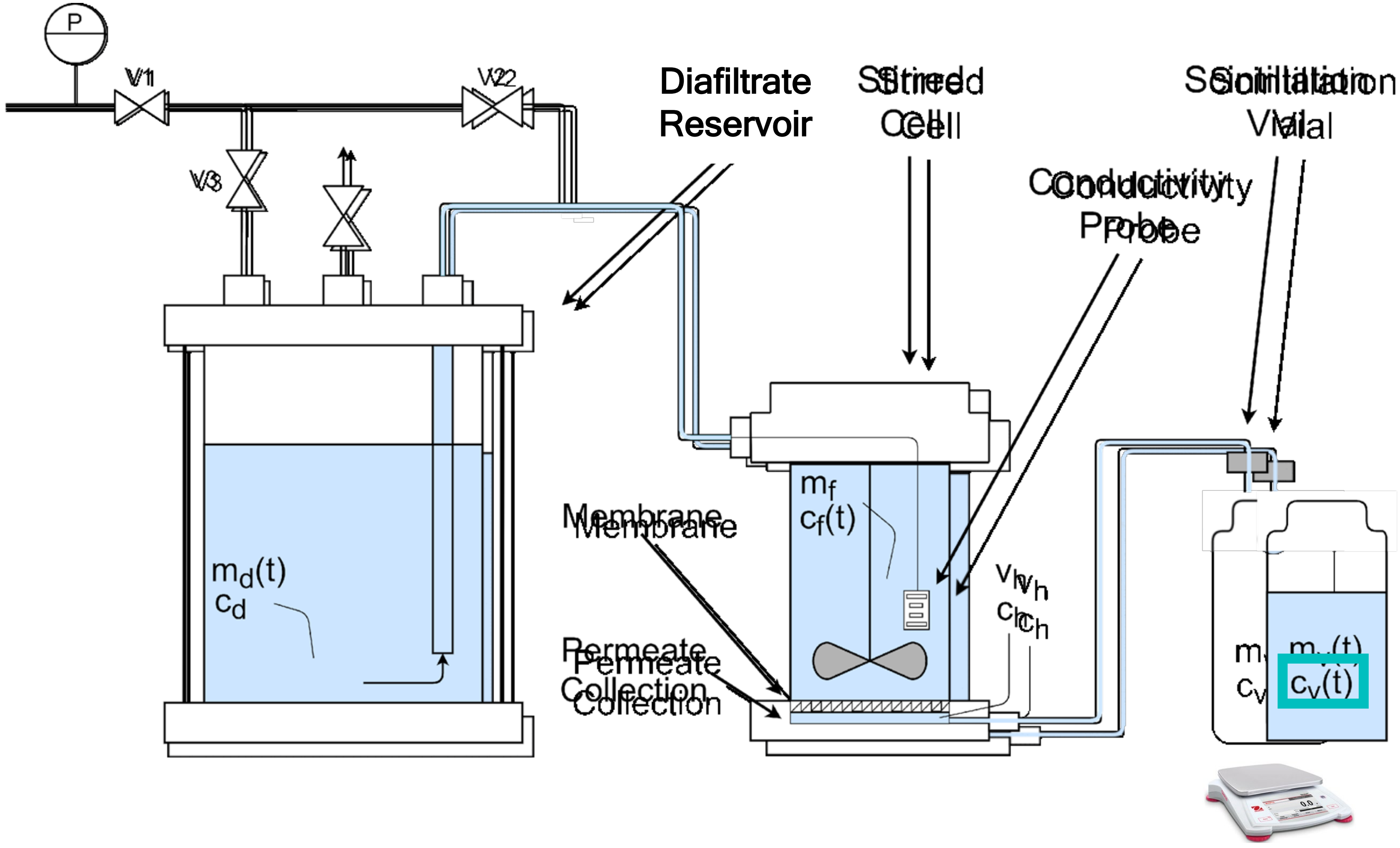


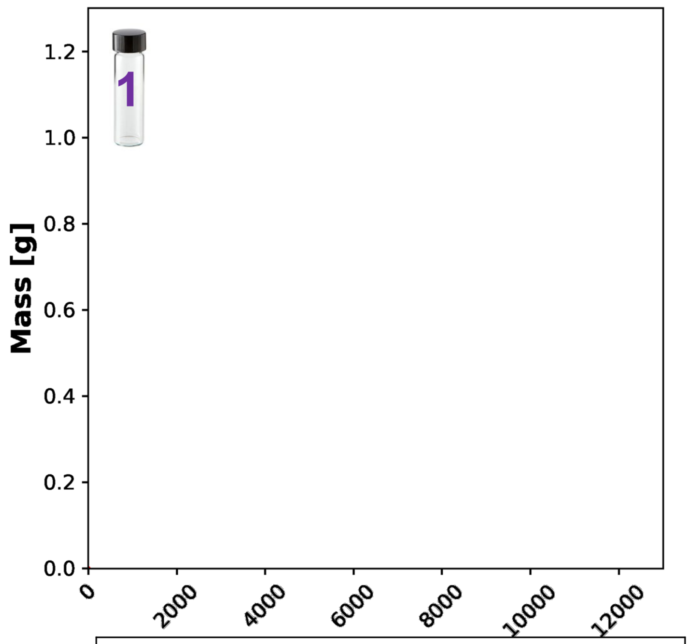


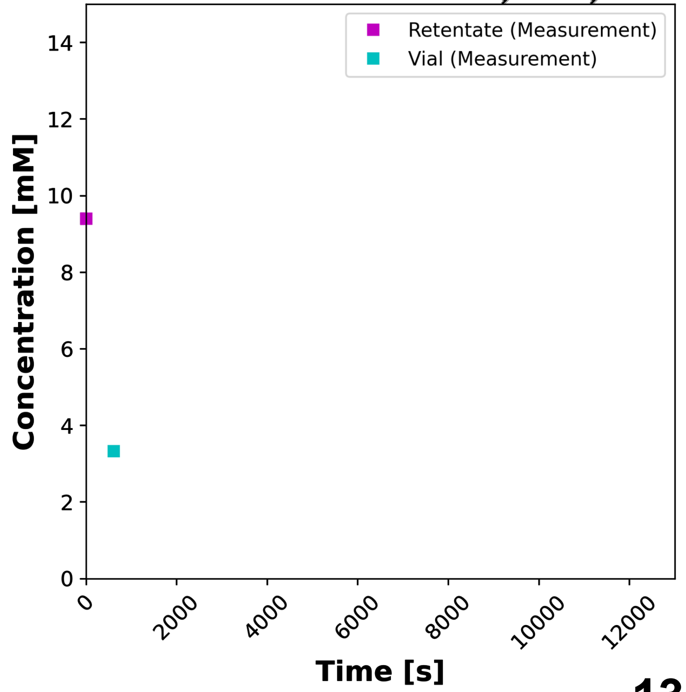
Animation: experiment start-up

UNIVERSITY OF NOTRE DAME

Dynamic Diafiltration Experiment Sweeps Larger Concentration Space

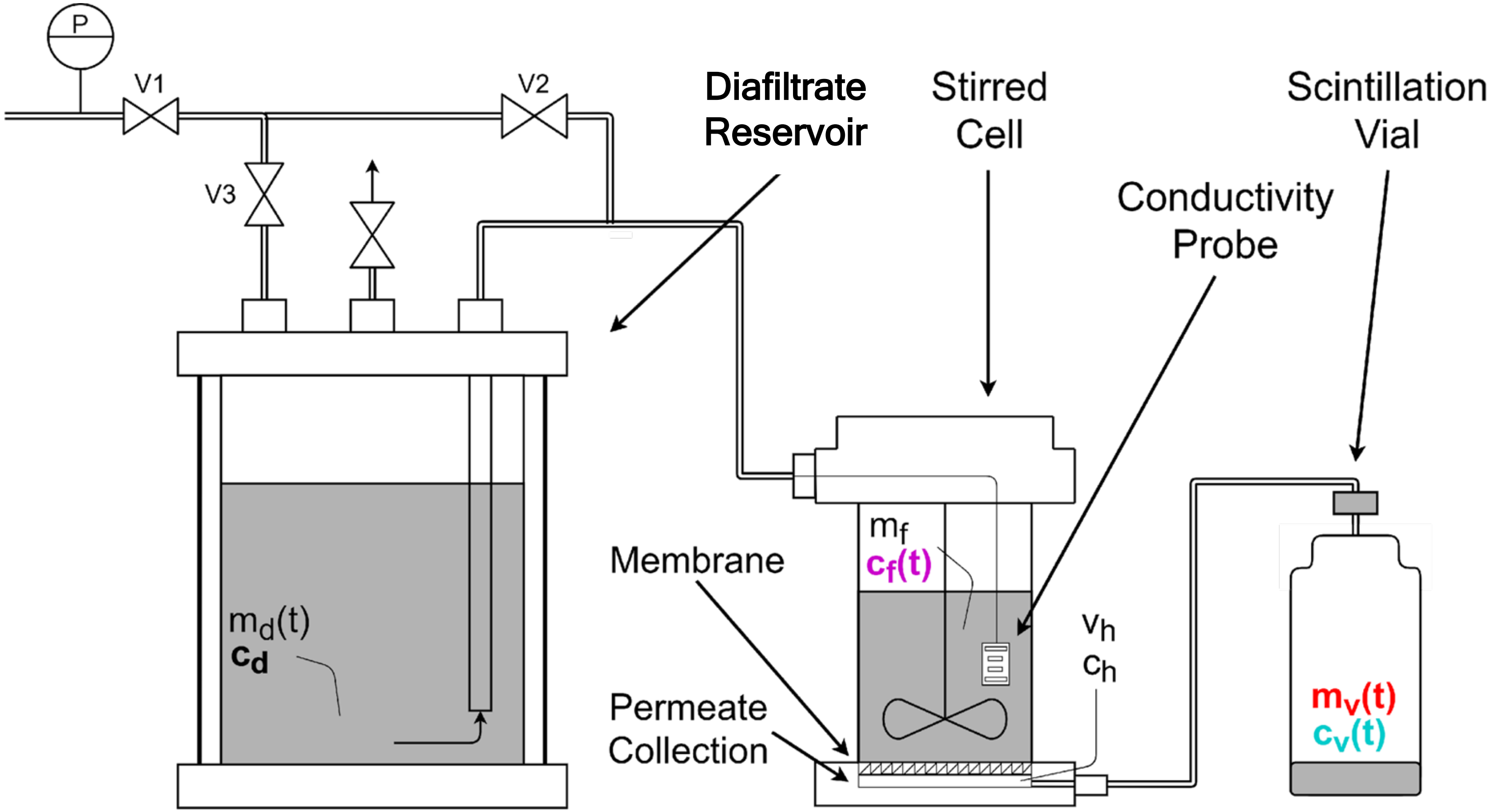


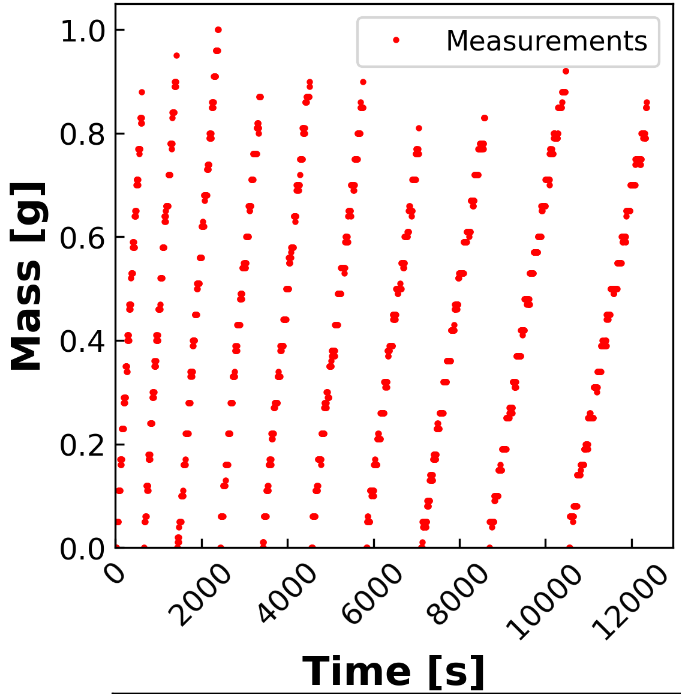


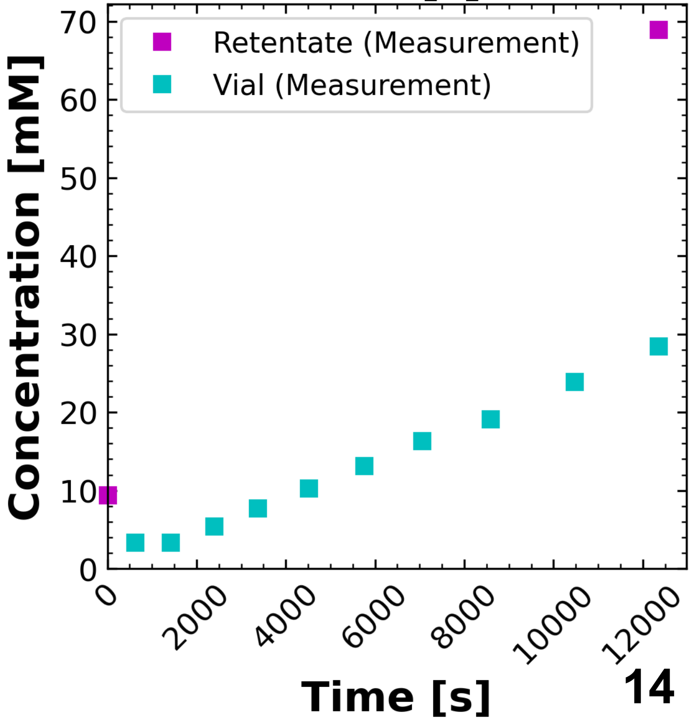


UNIVERSITY OF NOTRE DAME

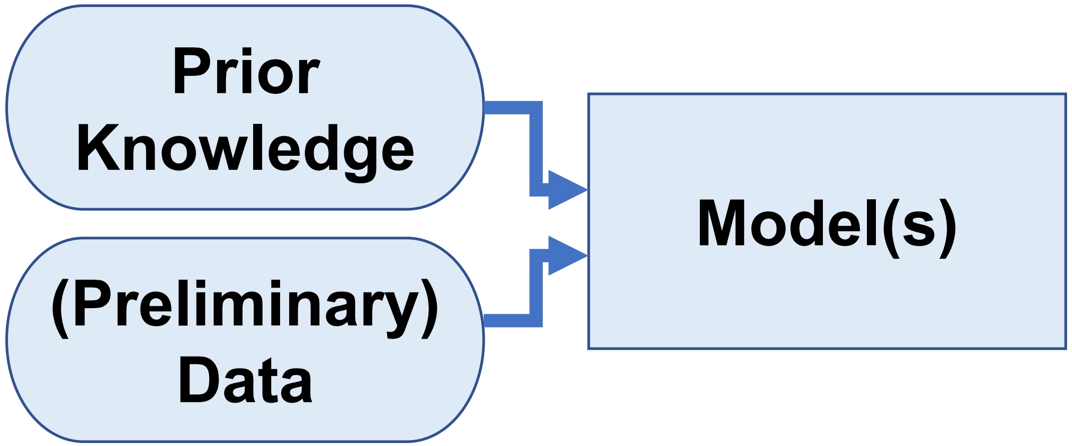
Dynamic Diafiltration Experiment Sweeps Larger Concentration Space 5x Faster Than Filtration Experiments







Science-based Modeling Workflow



Lumped Parameter Model Infers Membrane Properties

$\boxed{}$ **Fitted Parameters**

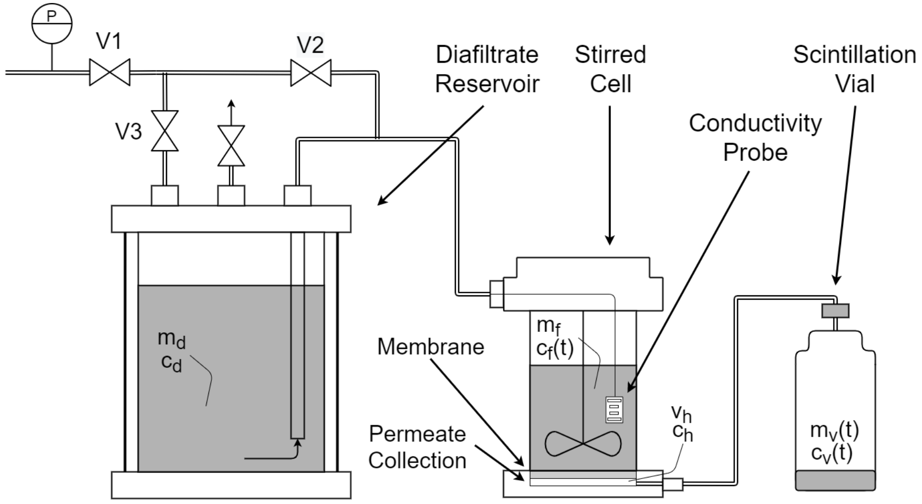
L_p: hydraulic permeability

B : solute permeability coefficient

σ : reflection coefficient

ΔP: applied pressure drop
n : number of dissolved species
R : gas constant
T : temperature
A_m: area of the membrane
ρ : density of solution

k : Mass transfer coefficient of interest
b : Diameter of the stirred cell
D : Diffusion coefficient of the solute in the solvent
ν : Kinematic viscosity of the solvent
v^0: Average velocity within the system



Water flux across the membrane

$$J_w = \boxed{L_p}(\Delta P - \boxed{\sigma}\Delta\pi)$$

$$\Delta\pi = nRT(c_{in} - c_h)$$

Solute flux across the membrane

$$J_s = \boxed{B}(c_{in} - c_h)$$

Concentration polarization

$$\frac{c_{in} - c_h}{c_f - c_h} = exp\left(\frac{J_w}{k}\right)$$

$$\frac{kb}{D} = 0.23\left(\frac{bv^0}{\nu}\right)^{0.57}\left(\frac{\nu}{D}\right)^{0.33}$$

Mass balance in dialysate reservoir

$$\frac{dm_d}{dt} = -A_m\rho J_w$$

Constant concentration in dialysate reservoir

$$\frac{dc_d}{dt} = 0$$

Constant mass in feed cell

$$\frac{dm_f}{dt} = 0$$

Solute balance in feed cell

$$\frac{d(c_f m_f)}{dt} = A_m\rho(J_w c_d - J_s)$$

Constant mass in holdup

$$\frac{dm_h}{dt} = 0$$

Solute balance in holdup

$$\frac{d(c_h m_h)}{dt} = A_m\rho J_s + \frac{dm_d}{dt}c_h$$

Mass balance in collecting vial

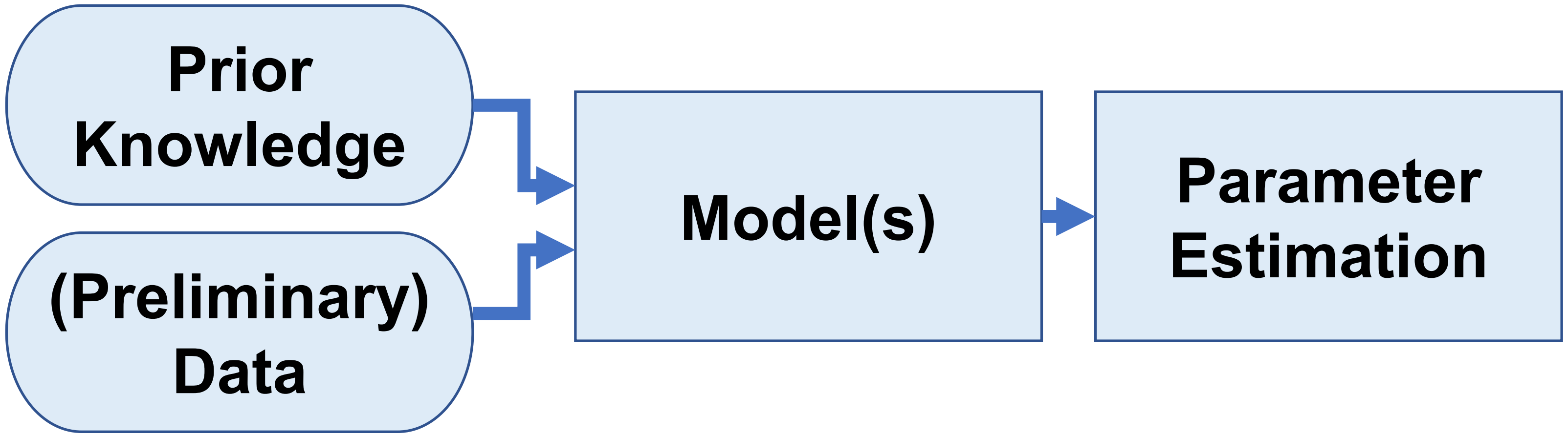
$$\frac{dm_v}{dt} = -\frac{dm_d}{dt} = A_m\rho J_w$$

Solute balance in collecting vial

$$\frac{d(c_v m_v)}{dt} = -\frac{dm_d}{dt}c_h = A_m\rho J_w c_h$$

[Fig.] Ouimet, J. A. et. al., (2022). Journal of Membrane Science, 641, 119743.

Science-based Modeling Workflow



Nonlinear Regression as a 2-Stage Stochastic Program

Pyomo: Bynum, Hackebeil, Hart, Laird, Nicholson, Siirola, Watson, Woodruff. Springer, 2021.
Parmest: Klise, Nicholson, Staid, Woodruff. *Computer Aided Chemical Engineering*, 47 (2019): 41-46.
Pyomo.DAE: Nicholson, Siirola, Watson, Zavala, and Biegler. *Mathematical Programming Computation* 10(2) (2018): 187-223.

$$\min \quad \sum_{i=1}^{N_E} \sum_{t \in \boldsymbol{t}_i} \left(\tilde{\boldsymbol{y}}_i(t) - \boldsymbol{y}_i(t)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_y \left(\tilde{\boldsymbol{y}}_i(t) - \boldsymbol{y}_i(t)\right)$$

$$\text{s.t.} \quad \left.\begin{aligned} \dot{\boldsymbol{x}}_i(t) &= \boldsymbol{f}(\boldsymbol{x}_i(t), \boldsymbol{z}_i(t), \boldsymbol{u}_i(t), \overline{\boldsymbol{w}}_i, \boldsymbol{\theta}) \\ \boldsymbol{g}(\boldsymbol{x}_i(t), \boldsymbol{z}_i(t), \boldsymbol{u}_i(t), \overline{\boldsymbol{w}}_i, \boldsymbol{\theta}) &= \boldsymbol{0} \\ \boldsymbol{y}(t) &= \boldsymbol{h}(\boldsymbol{x}_i(t), \boldsymbol{z}_i(t), \boldsymbol{\theta}) \end{aligned}\right\} \text{DAE System}$$

$$\left.\begin{aligned} \boldsymbol{f}^{\boldsymbol{0}}(\dot{\boldsymbol{x}}_i(t_0), \boldsymbol{x}_i(t_0), \boldsymbol{z}_i(t_0), \boldsymbol{u}_i(t_0), \overline{\boldsymbol{w}}_i, \boldsymbol{\theta}) &= \boldsymbol{0} \\ \boldsymbol{g}^{\boldsymbol{0}}(\boldsymbol{x}_i(t_0), \boldsymbol{z}_i(t_0), \boldsymbol{u}_i(t_0), \overline{\boldsymbol{w}}_i, \boldsymbol{\theta}) &= \boldsymbol{0} \\ \boldsymbol{y}^{\boldsymbol{0}}(t_0) &= \boldsymbol{h}(\boldsymbol{x}_i(t_0), \boldsymbol{z}_i(t_0), \boldsymbol{\theta}) \end{aligned}\right\} \text{Initial Conditions}$$

Stage 2
$\forall i \in \{1, \ldots, N_E\}$

`Parmest`

N_E	Number of experiments
$\boldsymbol{t}_i$	Measurement times for experiment *i*
$\boldsymbol{\Sigma}_y$	Covariance matrix for measurement errors
$\boldsymbol{\theta}$	Estimated parameters (stage 1)

<u>Stage 2 (for experiment *i*)</u>

$\tilde{\boldsymbol{y}}_i$ & $\boldsymbol{y}_i$	Measurements & model responses (parameter, variable)
$\boldsymbol{x}_i$	Time-dependent differential state variables
$\boldsymbol{z}_i$	Time-dependent algebraic state variables
$\boldsymbol{u}_i$	Time-varying control (parameter)
$\overline{\boldsymbol{w}}_i$	Time-invariant control (parameter)

Weighted Least-Squared Parameter Estimation

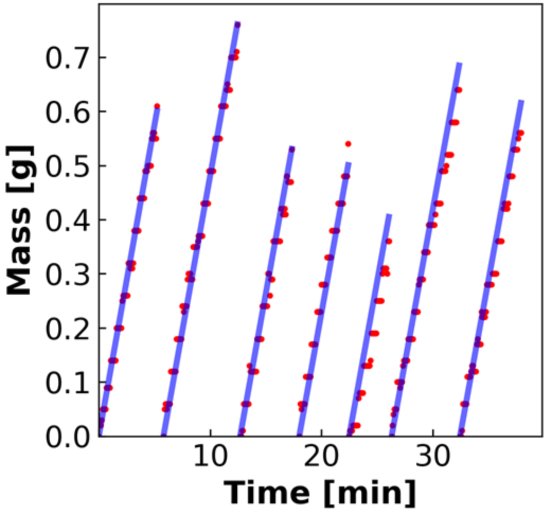
$$\hat{\theta} = \arg\min_{\theta} \underbrace{\sum_i w_{m_v,i}\left(m_{v,i} - \hat{m}_{v,i}\right)^2}_{\text{Mass collected}}$$

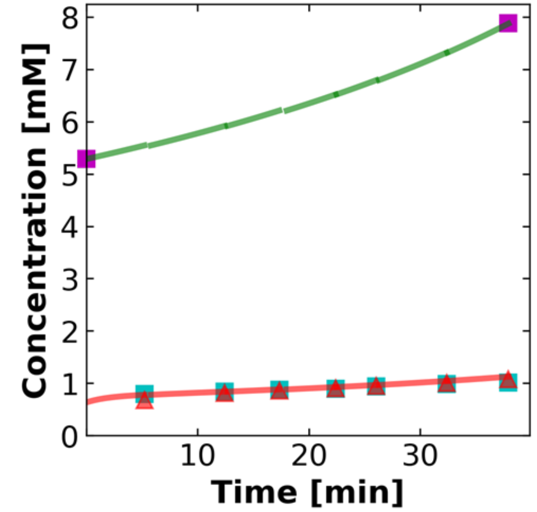
$$+ \underbrace{\sum_j w_{c_v,j}\left(c_{v,j} - \hat{c}_{v,j}\right)^2}_{\text{Vial concentration}}$$

$$+ \underbrace{\sum_k w_{c_f,k}\left(c_{f,k} - \hat{c}_{f,k}\right)^2}_{\text{Retentate concentration}}$$

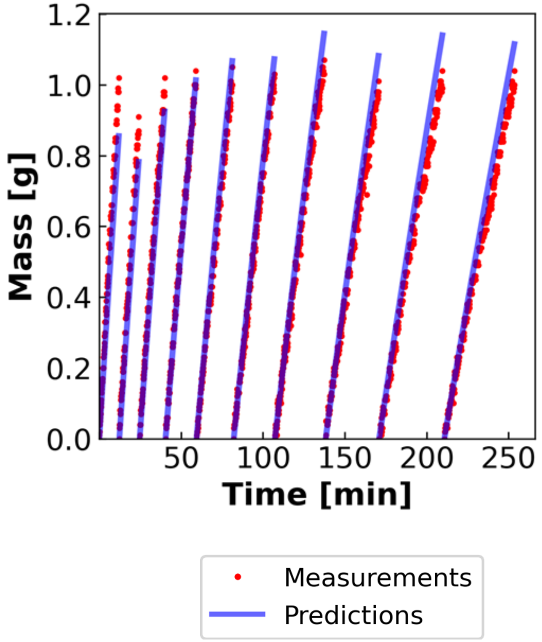
	L_p	B	**σ**
Filtration	4.30 L/(m^2-hr-bar)	0.67 μm/s	0.83 dimensionless
Diafiltration	3.25 L/(m^2-hr-bar)	0.31 μm/s	1 dimensionless

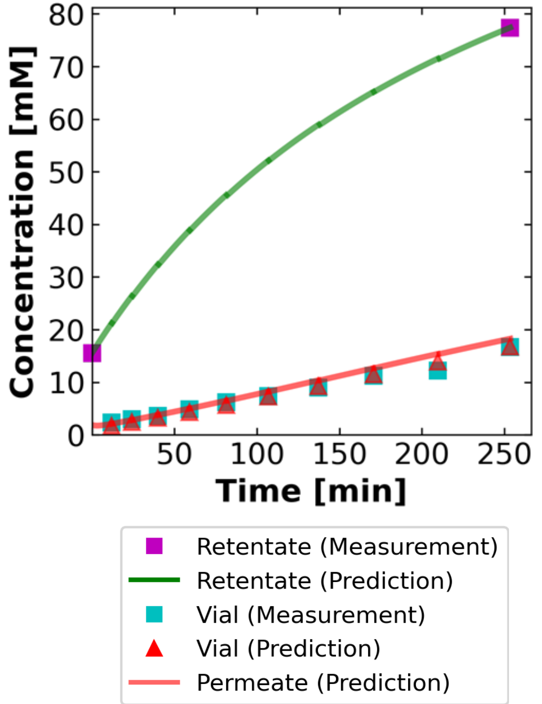
Filtration mode
Feed: 5mM KCl



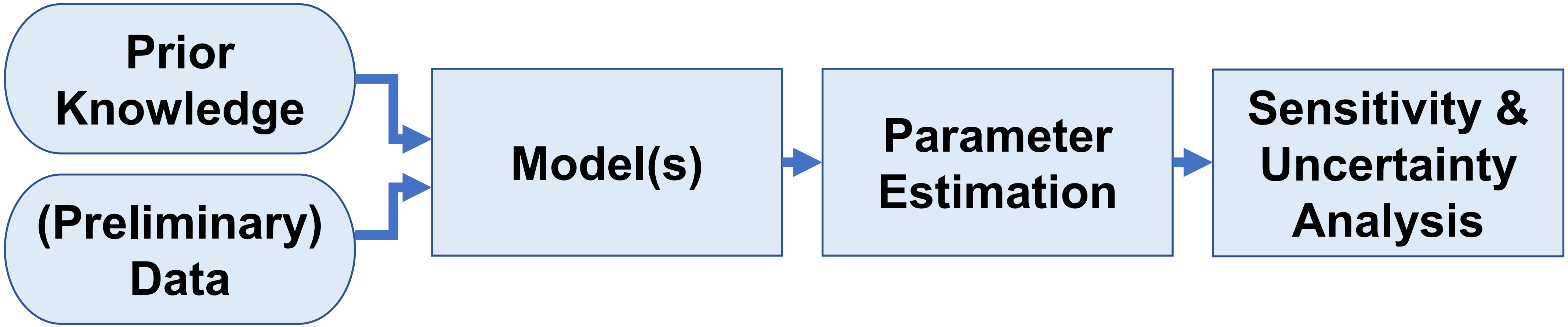


Diafiltration mode
Feed: 5mM KCl
Dialysate: 80mM KCl





Science-based Modeling Workflow



Parameter Estimation and Uncertainty

Assume a model and error structure:

$$y_i = f(\boldsymbol{x_i}, \boldsymbol{\theta}) + \epsilon_i \qquad \epsilon \sim N(0, \sigma_\epsilon^2)$$

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

$$\widehat{\boldsymbol{\theta}} = \text{argmin}_{\boldsymbol{\theta}} \sum_i [y_i - f(x_i, \boldsymbol{\theta})]^2$$

best fit estimates

How sensitivity is the least-squares objective Ψ to perturbations in θ?

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

Hessian matrix $\quad \boldsymbol{H} \approx \boldsymbol{Q}^T \boldsymbol{Q} \quad$ sensitivity matrix

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

covariance matrix for $\hat{\theta}$ $\qquad \boldsymbol{V}_{\widehat{\boldsymbol{\theta}}} \approx \sigma_\epsilon^2 \boldsymbol{H}^{-1} \approx \sigma_\epsilon^2 (\boldsymbol{Q}^T \boldsymbol{Q})^{-1}$

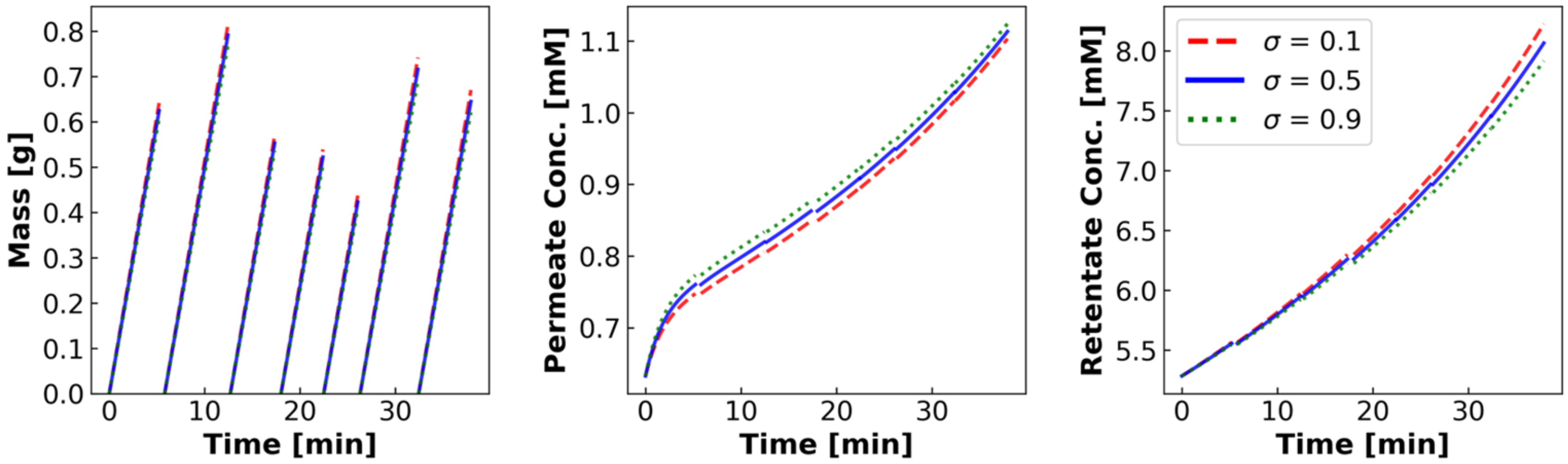
Key Insight. If model predictions $f(\boldsymbol{x_i}, \boldsymbol{\theta})$ are insensitive to θ_j, then:

- sensitivity $\boldsymbol{Q}$, Hessian $\boldsymbol{H}$, and covariance $\boldsymbol{V}_{\widehat{\boldsymbol{\theta}}}$ matrices are (numerically) rank deficient
- data $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot identity θ_j in model f
- large uncertainty in θ_j (and corresponding elements of $\boldsymbol{V}_{\widehat{\boldsymbol{\theta}}}$)

Bard (1974)
Bates and Watts (1988)

Finding 1: Diafiltration Experiments Identify the Reflection Coefficient σ

Filtration experiment (low concentrations)



Water flux across the membrane

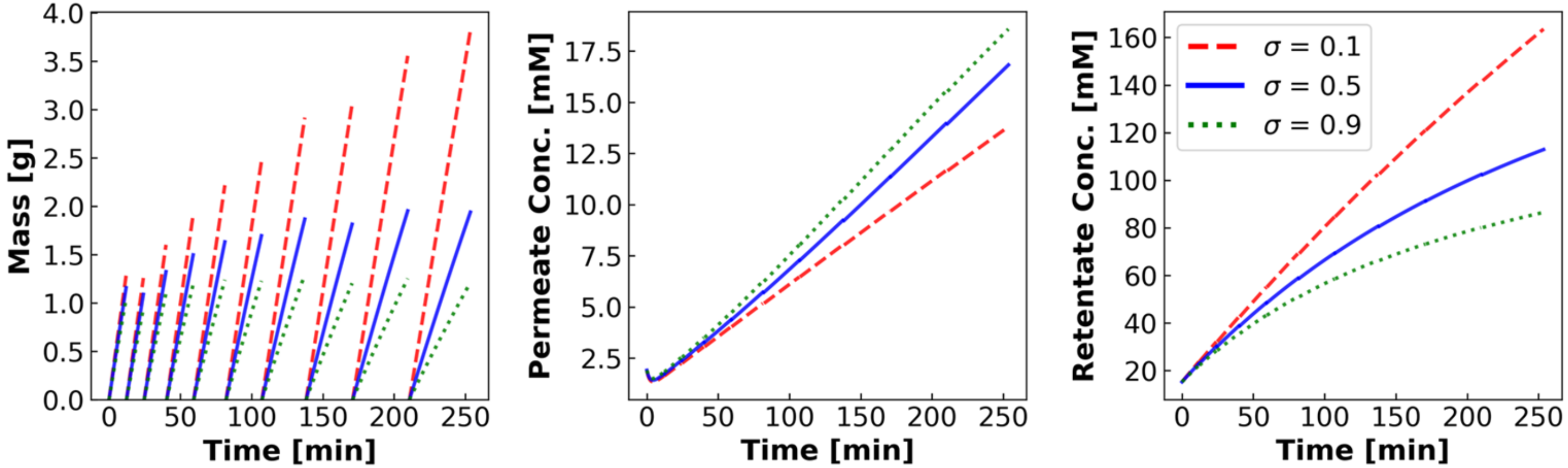
$$J_w = L_p(\Delta P - \sigma\,\Delta\pi)$$

$$\Delta\pi = nRT(c_{in} - c_h)$$

Solute flux across the membrane

$$J_s = B\,(c_{in} - c_h)$$

Diafiltration experiment (high concentrations)



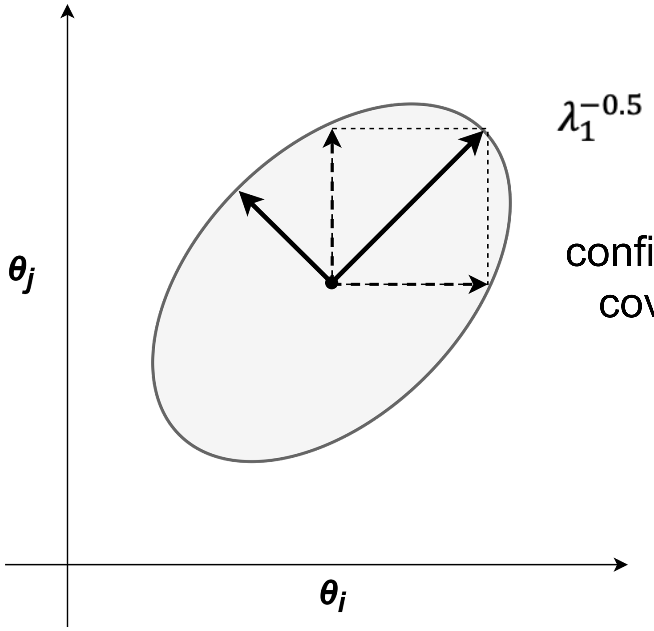
[Fig.] Ouimet, J. A. et. al., (2022). Journal of Membrane Science, 641, 119743.

Finding 1 continued: Fisher Information Matrix (FIM) Quantifies the Information Content

$$\boldsymbol{M} \approx \boldsymbol{V}_{\widehat{\boldsymbol{\theta}}}^{-1} \approx \sigma_\epsilon^{-2}\boldsymbol{H} \approx \sigma_\epsilon^{-2}\boldsymbol{Q}^T\boldsymbol{Q}$$

$$\boldsymbol{Q}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial f(x_1,\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial f(x_1,\boldsymbol{\theta})}{\partial \theta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_n,\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial f(x_n,\boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix}$$

$\boldsymbol{M}$	Fisher information matrix
$\widehat{\boldsymbol{\theta}}$	Estimated parameters
$\boldsymbol{V}_{\widehat{\boldsymbol{\theta}}}$	Covariance matrix for $\widehat{\boldsymbol{\theta}}$
$\boldsymbol{H}$	Hessian matrix for regression opt.
$\boldsymbol{Q}$	Model sensitivity matrix
$\boldsymbol{\sigma}_\epsilon^2$	Measurement error variance



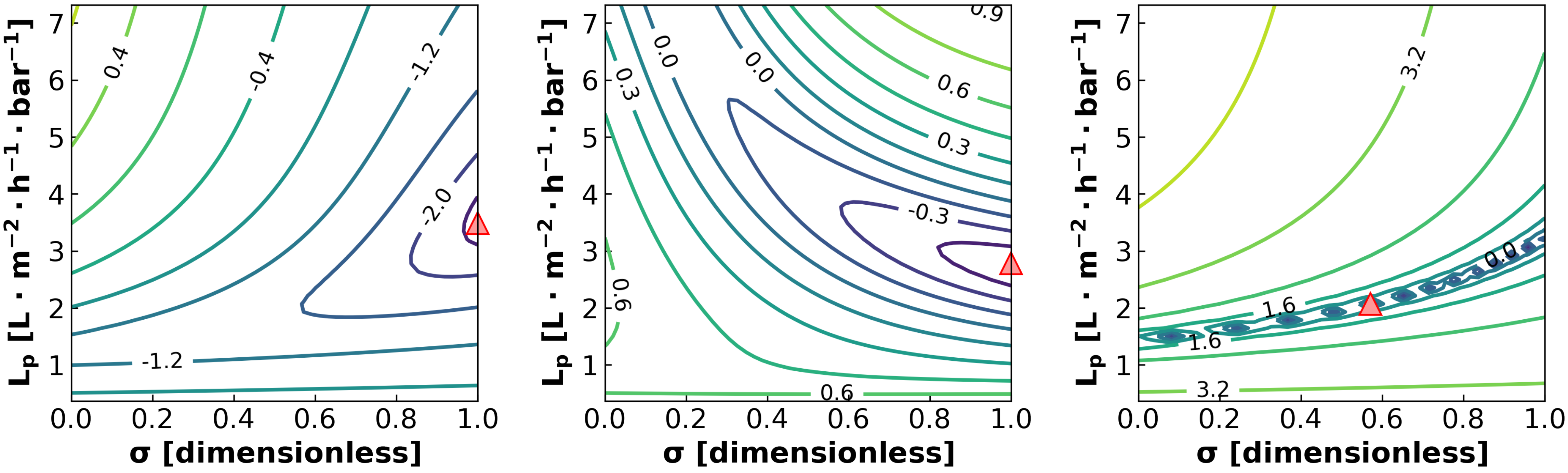
confidence ellipse for covariance matrix
V = M^{-1}

	FIM	Eigen values	Eigenvectors		
			L_p	B	**σ**
Filtration	$\times 10^{9}$ $\begin{bmatrix} 4.3352 & 0.0074 & -0.3742 \\ 0.0074 & 0.0163 & -0.0007 \\ -0.3742 & -0.0007 & 0.0328 \end{bmatrix}$	4.85×10^5	0.086	0.0033	0.9963
		1.63×10^7	0.002	-1	0.0032
		4.37×10^9	-0.9963	-0.0017	0.086
Diafiltration	$\times 10^{10}$ $\begin{bmatrix} 1.7779 & 0.3004 & -1.8364 \\ 0.3004 & 0.5567 & -1.9033 \\ -1.8364 & -1.9033 & 7.1384 \end{bmatrix}$	1.96×10^8	-0.1405	-0.9468	-0.2894
		1.27×10^{10}	0.9503	-0.211	0.2289
		8.18×10^{10}	-0.2778	-0.2429	0.9294

UNIVERSITY OF NOTRE DAME

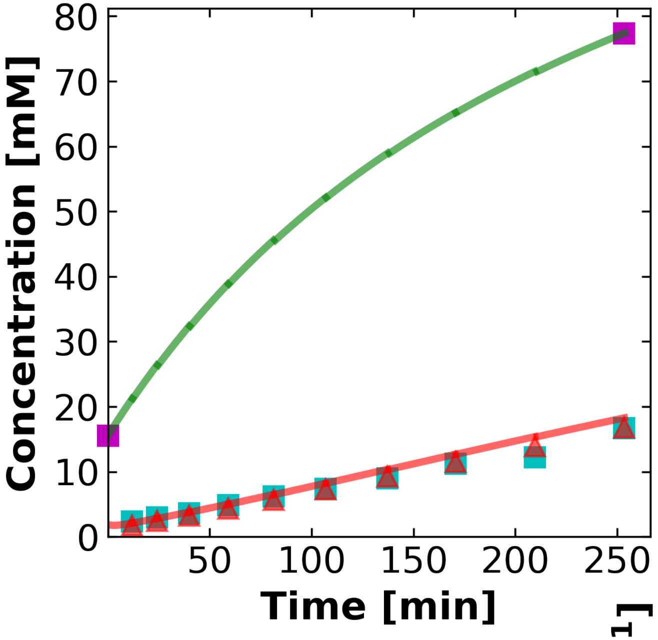
Multi-objective Parameter Estimation Trade-offs

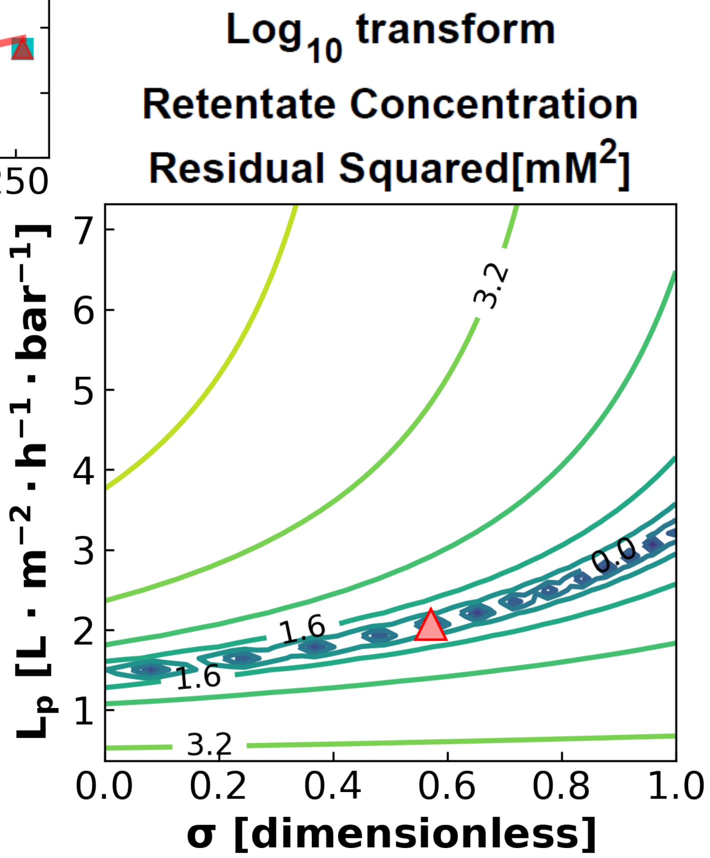
Mass Residual Objective [log g²]	Vial Conc. Residual Objective [log mM²]	Retentate Conc. Residual Objective [log mM²]
$\log_{10} \sum_i \left(m_{v,i} - \widehat{m}_{v,i}\right)^2$	$\log_{10} \sum_j \left(c_{v,j} - \hat{c}_{v,j}\right)^2$	$\log_{10} \sum_k \left(c_{f,k} - \hat{c}_{f,k}\right)^2$



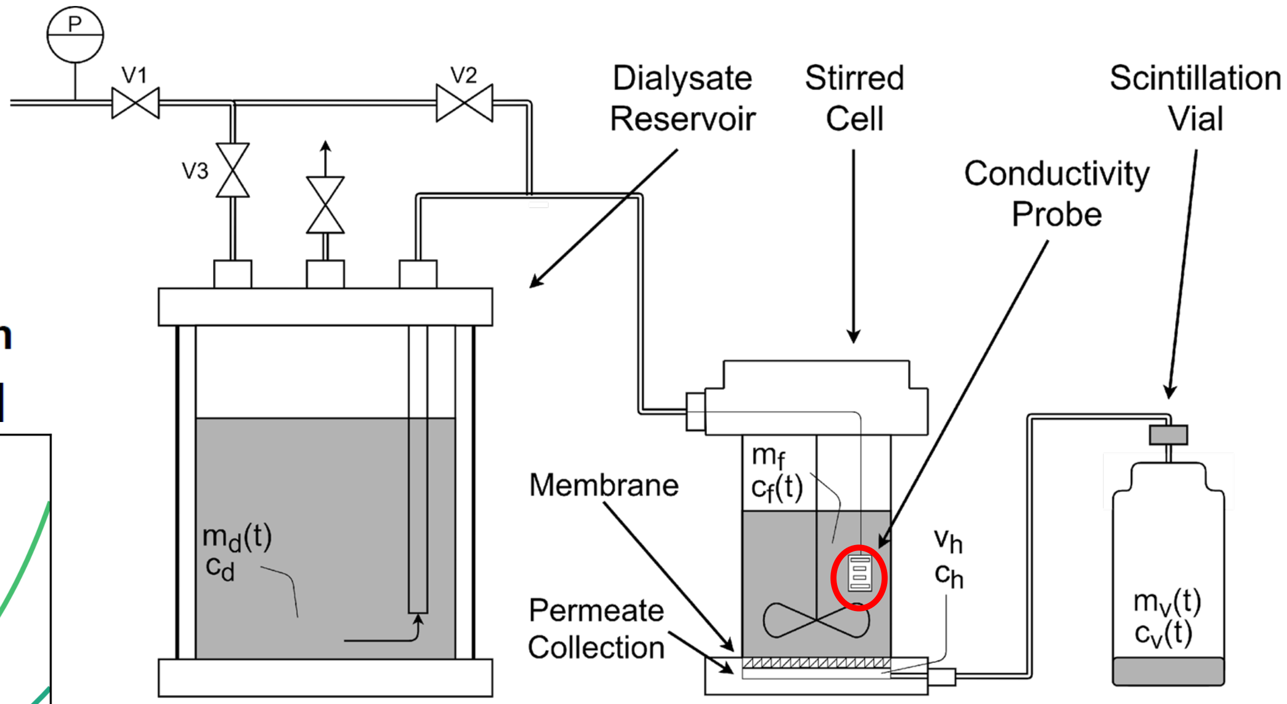
[Fig.] Ouimet, J. A. et. al., (2022). Journal of Membrane Science, 641, 119743.

Finding 2: Add Retentate Concentration Measurements





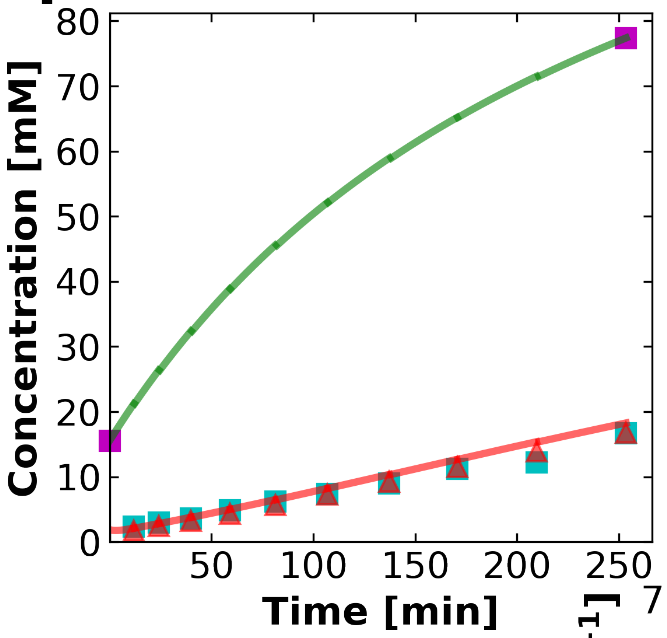
Several sets of parameters could give good predictions



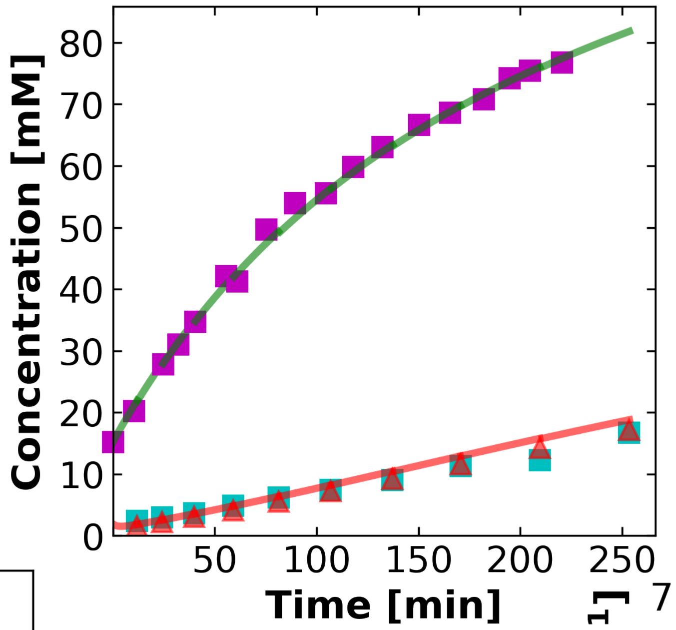
The conductivity probe enables more data collected

[Fig., adapted] Ouimet, J. A. et. al., (2022). Journal of Membrane Science, 641, 119743.

Finding 2: Add Retentate Concentration Measurements improves precision of σ estimation

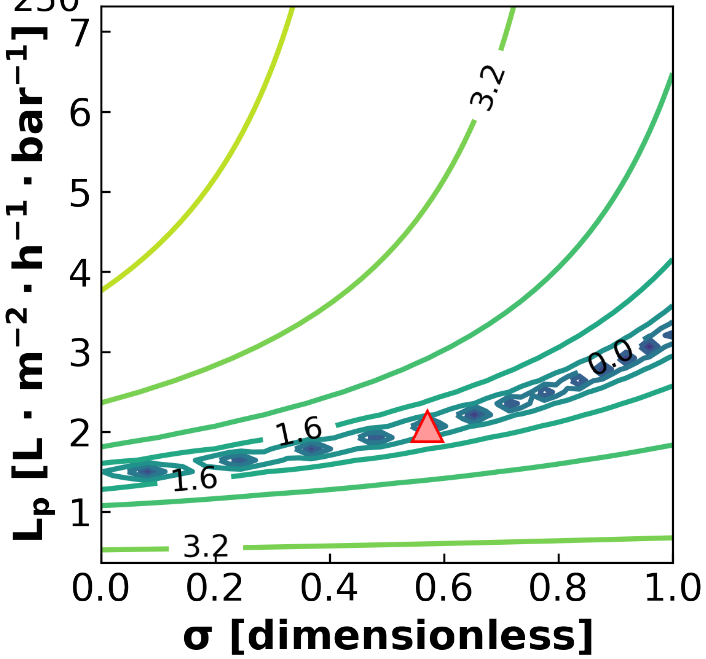


Retentate concentration data added



Retentate Conc. Residual Objective [log mM²]

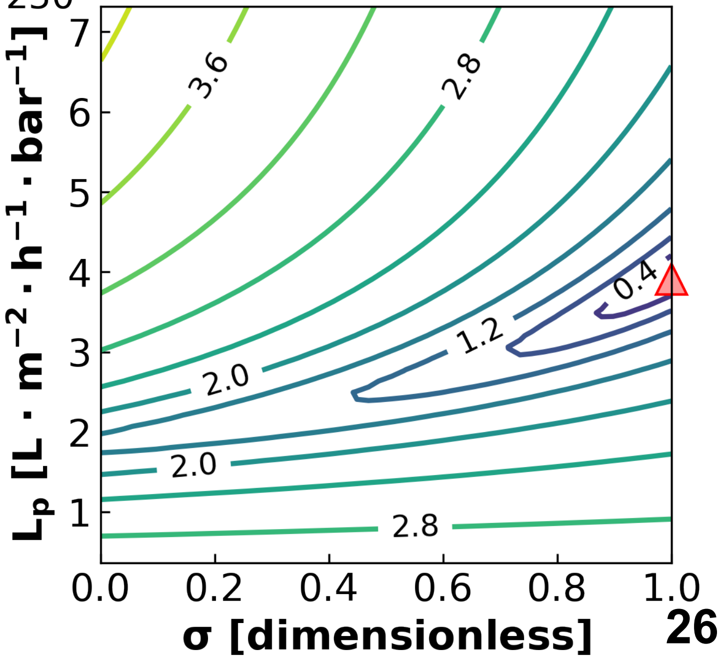
$$\log_{10} \sum_k \left(c_{f,k} - \hat{c}_{f,k}\right)^2$$



Local minima eliminated

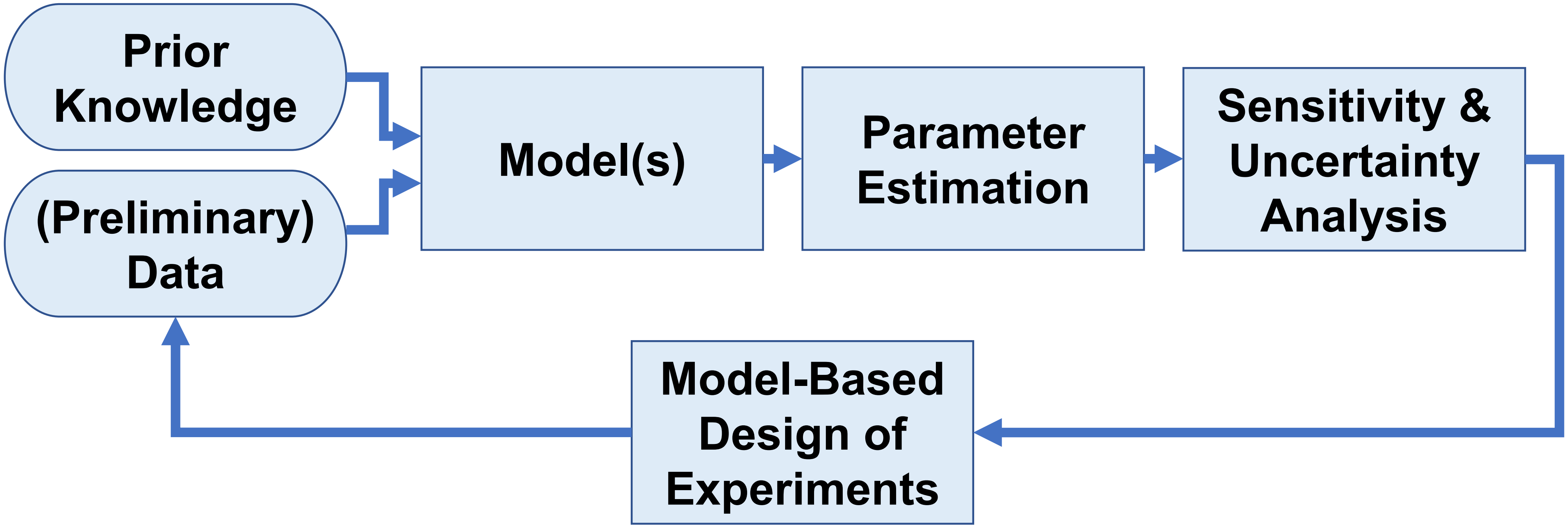
Retentate Conc. Residual Objective [log mM²]

$$\log_{10} \sum_k \left(c_{f,k} - \hat{c}_{f,k}\right)^2$$



UNIVERSITY OF NOTRE DAME

Science-based Modeling Workflow



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(M)

enclosing box volume

poor choice for highly correlated **θ**

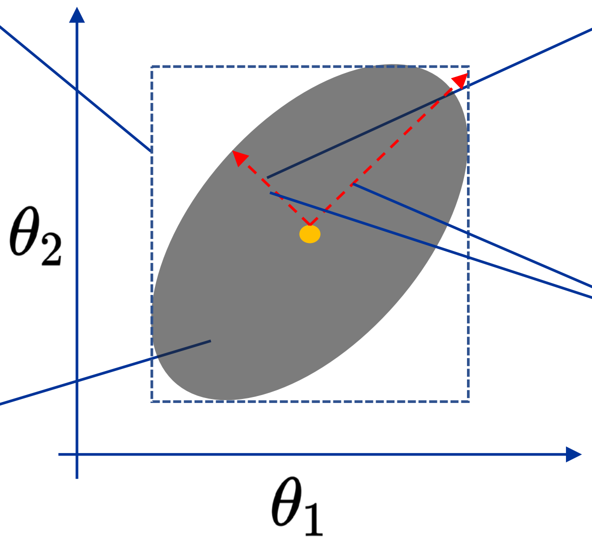
D-optimality

max det(M)

ellipsoid volume

robust to linear transformations

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



E-optimality

max min(eig(M))

major axis

recommended if M is ill-conditioned

ME-optimality

min κ(M) = max(eig(M)) / min(eig(M))

ratio of major to minor axes

recommended if 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.

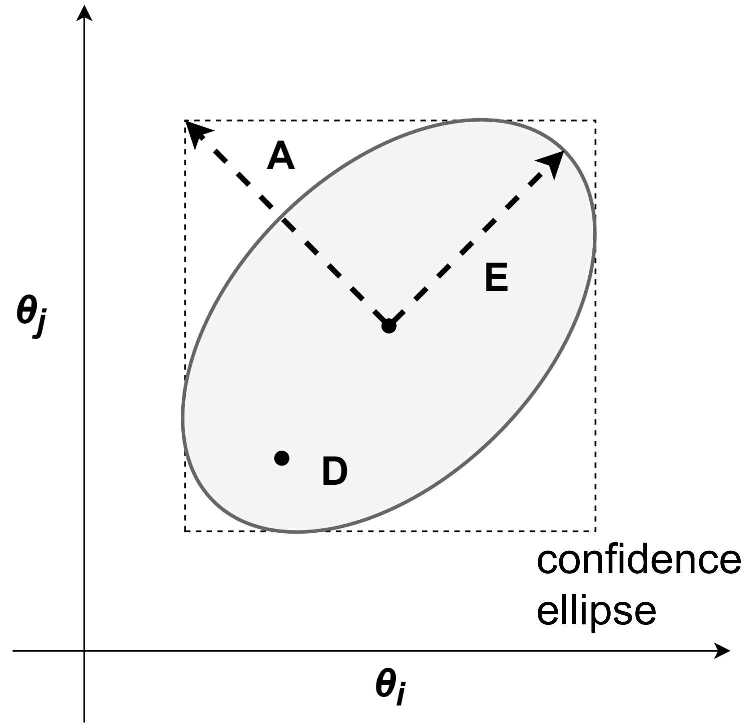
Model Discrimination

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.

Finding 2 Continued: Fisher Information Matrix Quantifies Benefits of New Retentate Concentration Measurements (Conductivity Probe)



A		**dimensions of the enclosing box**
D		**area or volume**
E		**size of major axis**
ME		**ratio of minor axis over major axis**

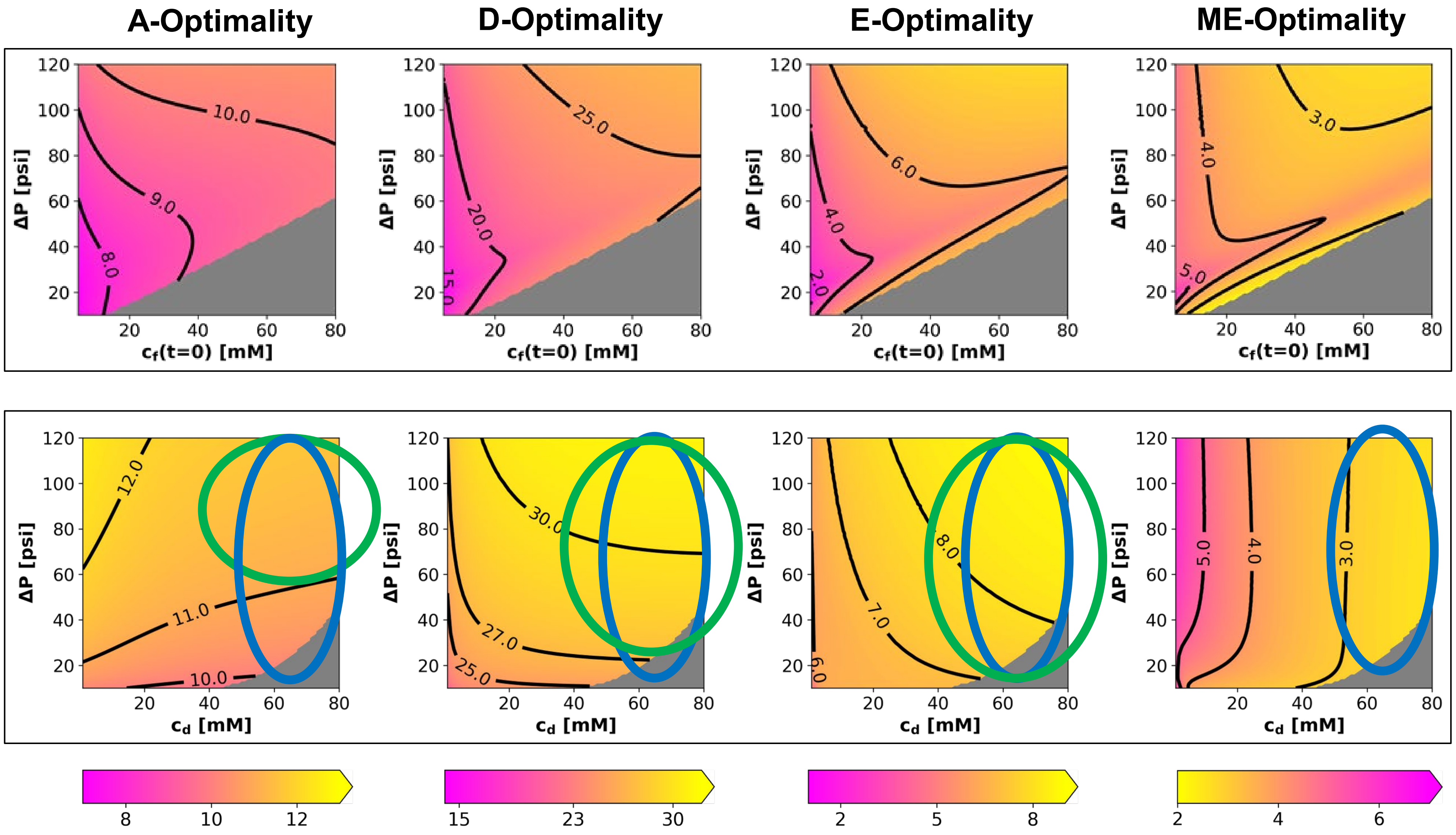
	A-optimal Maximize	D-optimal Maximize	E-optimal Maximize	ME-optimal Minimize
	Trace	**Determinant**	**Smallest eigenvalue**	**Condition number**
w/o conductivity probe	9.47×10^{10}	2.03E+29	1.96E+08	417.745
w/ conductivity probe	1E+11 **6%↑**	2.69E+29 **32%↑**	2.17E+08 **11%↑**	393 **6%↓**

$\underbrace{\text{A-optimal, D-optimal, E-optimal}}_{\textbf{improved parameter precision}}$ $\underbrace{\text{ME-optimal}}_{\textbf{less correlation}}$

Finding 3: Recommend Diafiltration Experiments with $c_d \geq 50$ mM and $\Delta P \geq 45$ psi

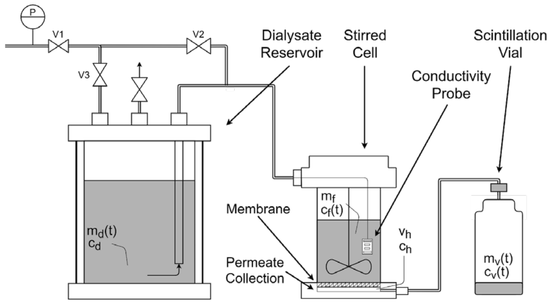
Filtration mode

Diafiltration mode



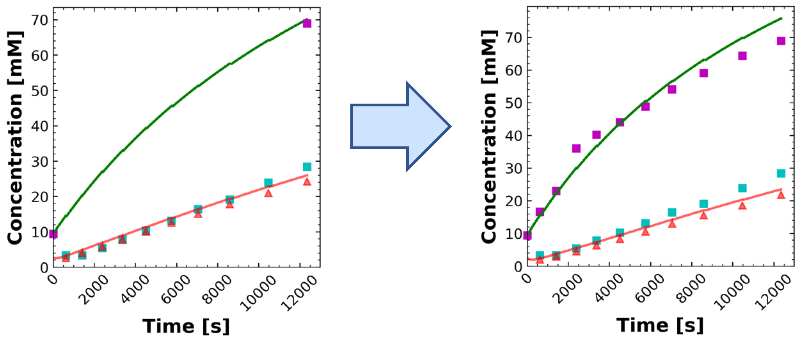
UNIVERSITY OF NOTRE DAME

Membrane Tutorial Conclusions



FIM analysis quantifies benefits of new sensors *a priori*

FIM eigendecomposition gives insights into model identifiability



Ouimet, J. A., Liu, X., Brown, D. J., Eugene, E. A., Popps, T., Muetzel, Z. W., Dowling, A. W., & Phillip, W. A. (2022). **DATA: Diafiltration Apparatus for high-Throughput Analysis**. *Journal of Membrane Science*, *641*, 119743.

Liu, X., Wang, J., Ouimet, J. A., Phillip, W. A., & Dowling, A. W. (2022). **Accelerating Membrane Characterization with Model-Based Design of Experiments.** *Computer Aided Chemical Engineering*, *PSE2021+*.

The Patrick and Jana Eiler's Graduate Student Fellowship (E. Eugene)

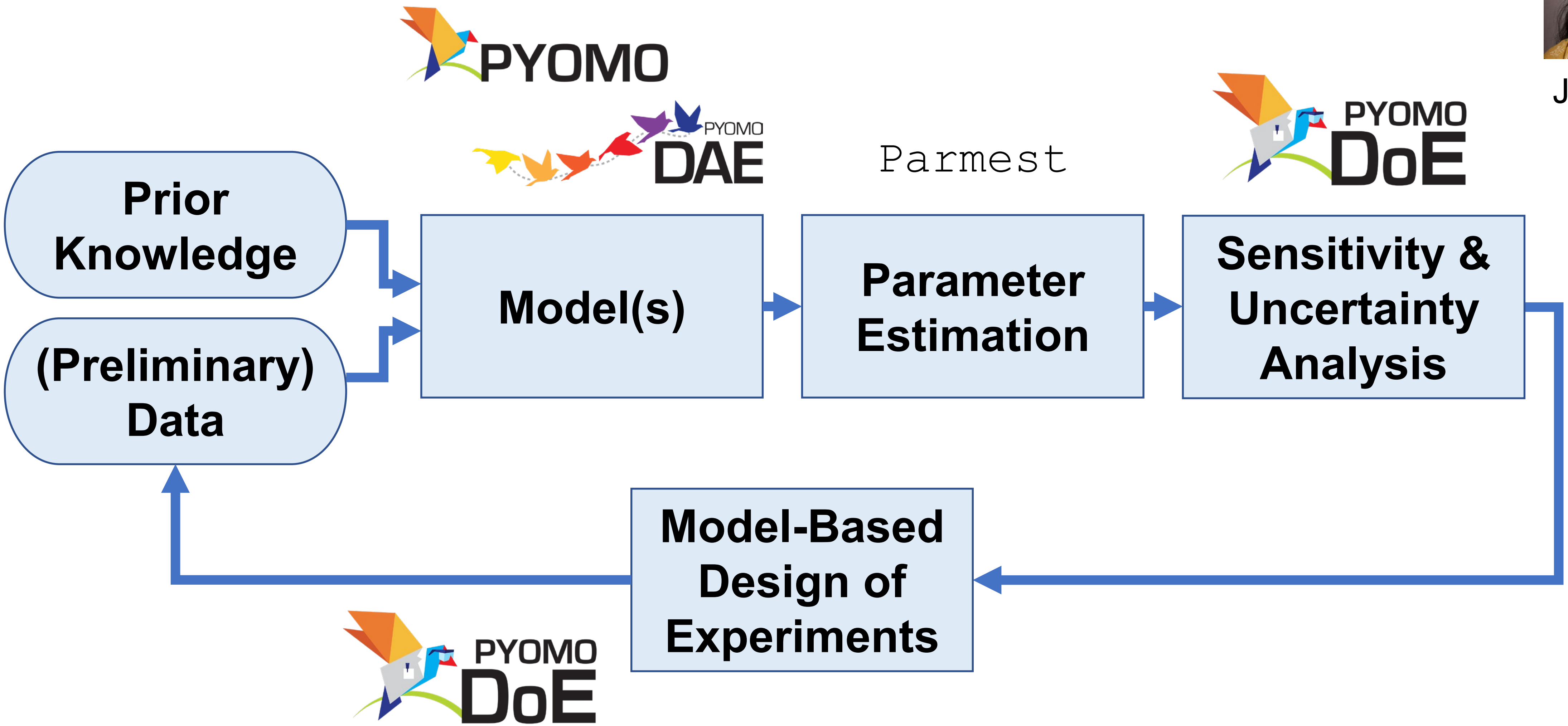
The Vincent P. Slatt Fellowship for Undergraduate Research in Energy Systems and Processes (N. Wamle)

Award CBET-1941596

Award CMMI-1932206-S001

Science-based Modeling Workflow

Jialu Wang



Wang, J. & Dowling A. W. (2022). Pyomo.DOE: An Open-Source Package for Model-based Design of Experiments in Python. *AIChE Journal,* e17813.

SBDoE Optimization Formulation

Jialu Wang

$$\max_{\boldsymbol{\varphi}} \quad \Psi\left[\, \boldsymbol{M}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}\right) \right]$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}}(t) = \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{z}(t), \boldsymbol{u}(t), \bar{\boldsymbol{w}}, \hat{\boldsymbol{\theta}} \right)$$

$$\boldsymbol{g}\left(\boldsymbol{x}(t), \boldsymbol{z}(t), \boldsymbol{u}(t), \bar{\boldsymbol{w}}, \hat{\boldsymbol{\theta}} \right) = \boldsymbol{0}$$

$$\boldsymbol{y}(t) = \boldsymbol{h}\left(\boldsymbol{x}(t), \boldsymbol{z}(t), \hat{\boldsymbol{\theta}} \right)$$

DAE System

$$\boldsymbol{f}^{\boldsymbol{0}}\left(\dot{\boldsymbol{x}}(t_0), \boldsymbol{x}(t_0), \boldsymbol{z}(t_0), \boldsymbol{u}(t_0), \bar{\boldsymbol{w}}, \hat{\boldsymbol{\theta}} \right) = \boldsymbol{0}$$

$$\boldsymbol{g}^{\boldsymbol{0}}\left(\boldsymbol{x}(t_0), \boldsymbol{z}(t_0), \boldsymbol{u}(t_0), \bar{\boldsymbol{w}}, \hat{\boldsymbol{\theta}} \right) = \boldsymbol{0}$$

$$\boldsymbol{y}^{\boldsymbol{0}}(t_0) = \boldsymbol{h}\left(\boldsymbol{x}(t_0), \boldsymbol{z}(t_0), \hat{\boldsymbol{\theta}} \right)$$

Initial Conditions

$$\boldsymbol{m}\left(\boldsymbol{x}(t), \boldsymbol{y}(t), \boldsymbol{z}(t), \boldsymbol{u}(t), \bar{\boldsymbol{w}}, \hat{\boldsymbol{\theta}} \right) = \boldsymbol{0}$$

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

Fisher information matrix (FIM):

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

MBDoE Decisions:

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

CCSI2 Carbon Capture Simulation for Industry Impact

UNIVERSITY OF NOTRE DAME

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846-4872.

Pyomo.DoE Formulation: MBDoE as 2-Stage Stochastic Program

max

$$\text{logdet}\left(\mathbf{M}(\hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}) \right) = 2 \sum_{i=1}^{N_p} \log L_{ii}$$ D-optimality

s.t.

Stage 1

$$\mathbf{M} = \sum_{r} \sum_{r'} \tilde{\sigma}_{r,r'} \mathbf{Q}_r^{\mathrm{T}} \mathbf{Q}_{r'}$$

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

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

$$\boldsymbol{m}\left(\boldsymbol{x}_p^{+}(t), \boldsymbol{y}_p^{+}(t), \boldsymbol{z}_p^{+}(t), \boldsymbol{u}(t), \overline{\boldsymbol{w}}, \boldsymbol{\theta}_p^{+} \right) = \mathbf{0}$$

$$\boldsymbol{m}\left(\boldsymbol{x}_p^{-}(t), \boldsymbol{y}_p^{-}(t), \boldsymbol{z}_p^{-}(t), \boldsymbol{u}(t), \overline{\boldsymbol{w}}, \boldsymbol{\theta}_p^{-} \right) = \mathbf{0}$$

Two model evaluations

$$\boldsymbol{\theta}_p^{+} = \hat{\boldsymbol{\theta}} + \boldsymbol{e}_p \epsilon_p$$

$$\boldsymbol{\theta}_p^{-} = \hat{\boldsymbol{\theta}} - \boldsymbol{e}_p \epsilon_p$$

Up and down perturbations

Stage 2

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

Model Sensitivity

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

$$\mathbf{q}_{r,p} = \begin{bmatrix} \frac{\partial y_r(t_1)}{\partial \theta_p} & \cdots & \frac{\partial y_r(t_n)}{\partial \theta_p} \end{bmatrix}^{\mathrm{T}}$$

$\boldsymbol{y}$	Measurements (model responses)
$\mathbf{Q}_\mathbf{r}$	Dynamic sensitivity for response *r*
$\boldsymbol{m}(\,)$	DAE model
$\hat{\boldsymbol{\theta}}$	Estimate for parameters
$\mathbf{M}$	Fisher information matrix
$\mathbf{L}$	Lower triangular Cholesky factorization
ϵ_p	Small perturbation for parameter *p*
$\boldsymbol{e}_p$	Unit vector with “1” in position *p*

New: Pyomo.DoE Extends `ParmEst` Interface

`create_model`

Create Pyomo model for DAE
Compatible with parmest

`DesignVariables`

Specify the MBDoE degrees of freedom and their bounds

`MeasurementVariables`

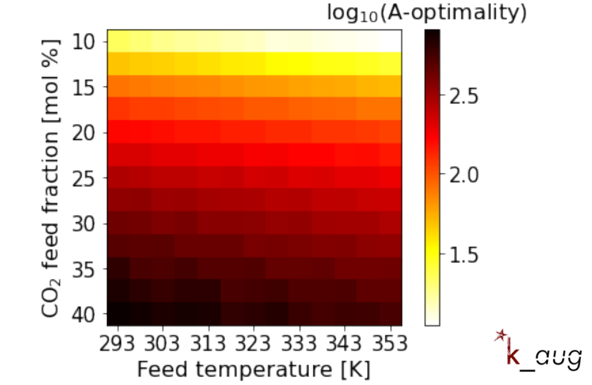
Specify the MBDoE measurement variables and observation error covariance matrix

`DesignOfExperiment`

PYOMO DoE

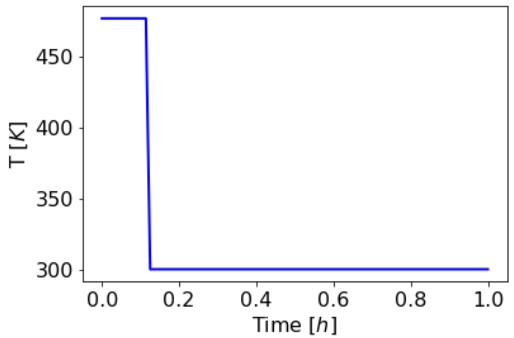
`compute_FIM`

Fast exploratory analysis



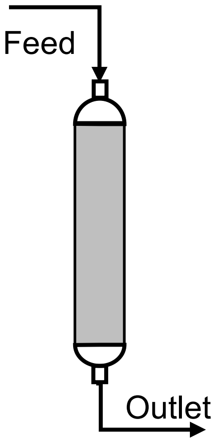
`stochastic_program`

Dynamic optimization



SBDoE Facilitates Collaborations

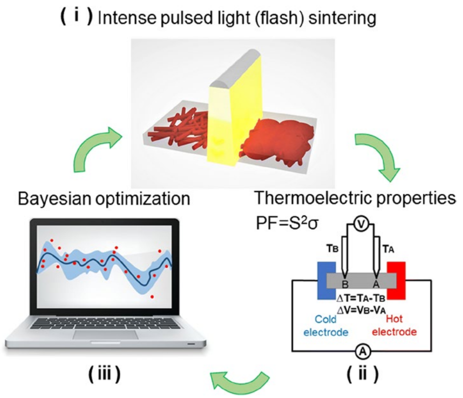
CO_2 Capture



Jialu Wang

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

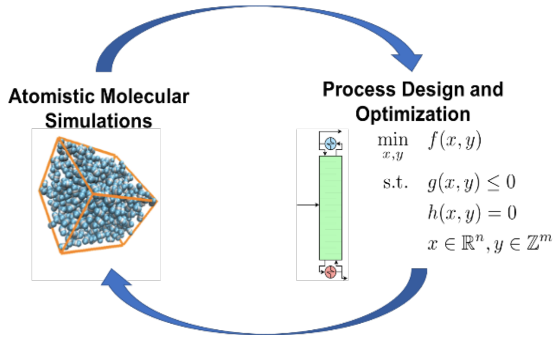
Additive Manufacturing of Thermoelectric Devices



Ke Wang

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

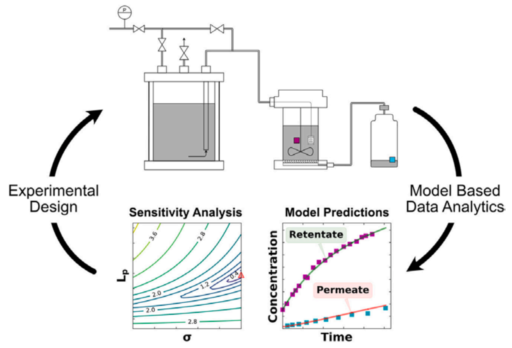
Molecular Design (Ionic Liquids)



Bridgette Befort

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

Rapid/Automated Membrane Characterization

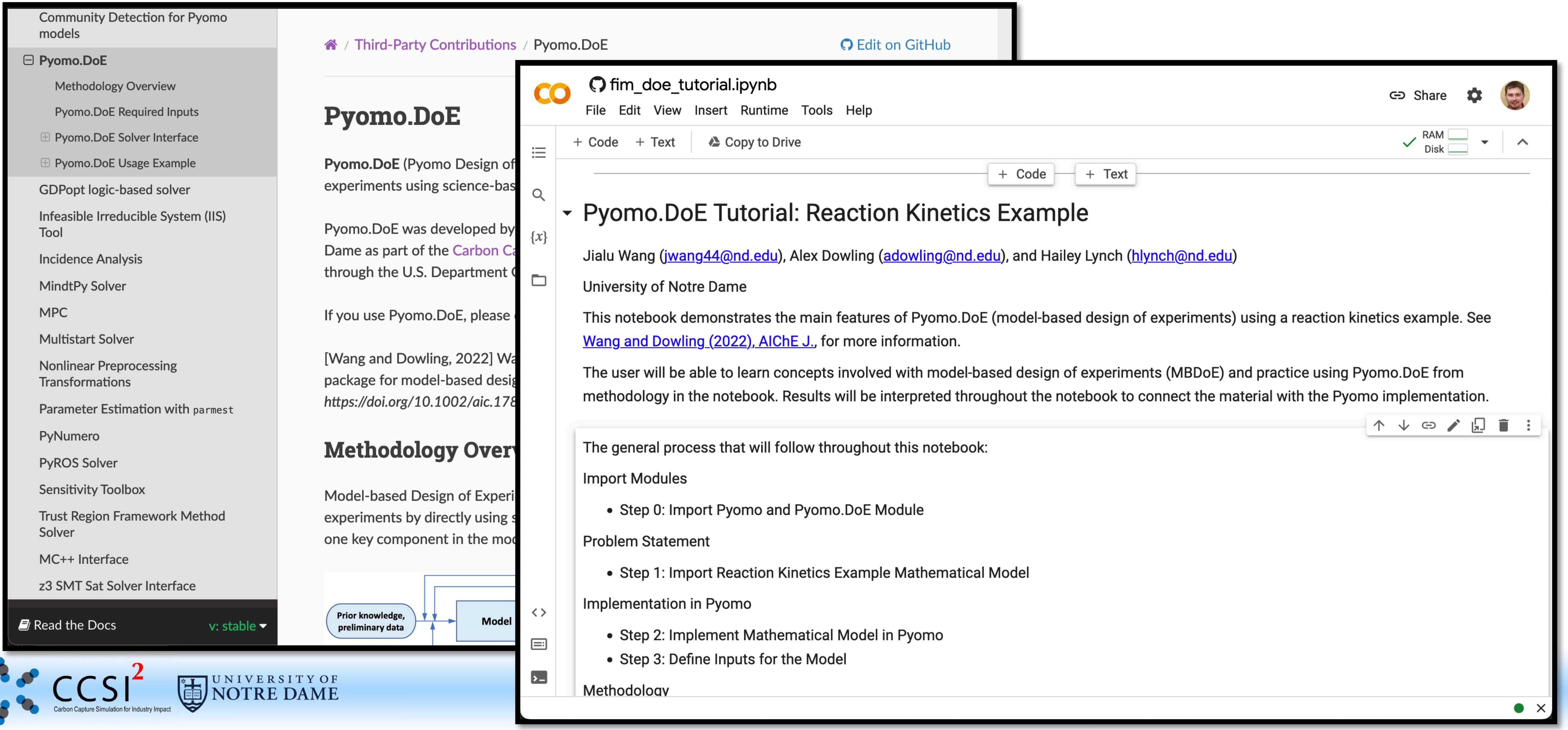


Xinhong Liu

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

Getting Started with Pyomo.DoE

Documentation: https://pyomo.readthedocs.io/en/stable/contributed_packages/doe/doe.html
Tutorial: https://colab.research.google.com/github/Pyomo/pyomo/blob/main/pyomo/contrib/doe/examples/fim_doe_tutorial.ipynb



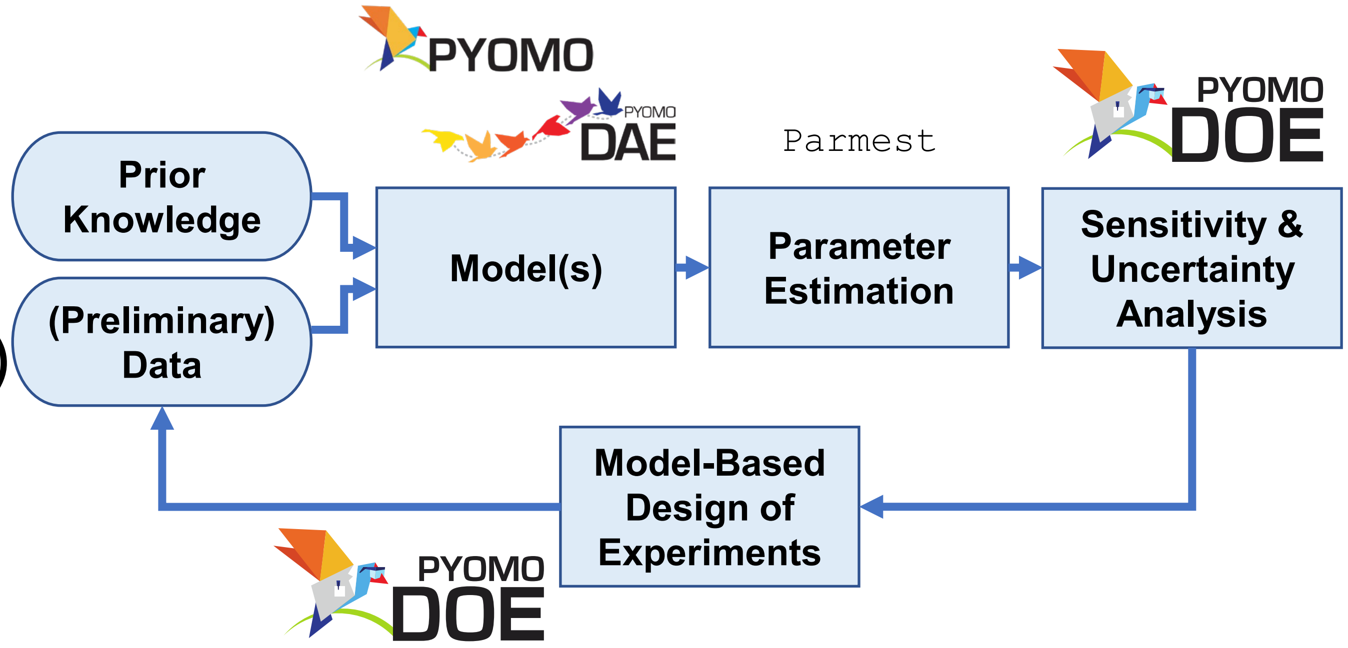
Next Steps for `Pyomo.DoE` Development

Explore NLP decomposition algorithms (exploit 2-stage structure)

Optimization under uncertainty (e.g., MBDoE with uncertainty parameters)

External model function evaluations (linear algebra) to speed up Pyomo.DoE, solve measurement optimization MINLP

Refactor `Pyomo.DoE` and `parmest` interfaces to improve user experience



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.*

And more to come!

Acknowledgements

We graciously acknowledge funding from the U.S. Department of Energy, Office of Fossil Energy and Carbon Management, through the Carbon Capture Program.

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):

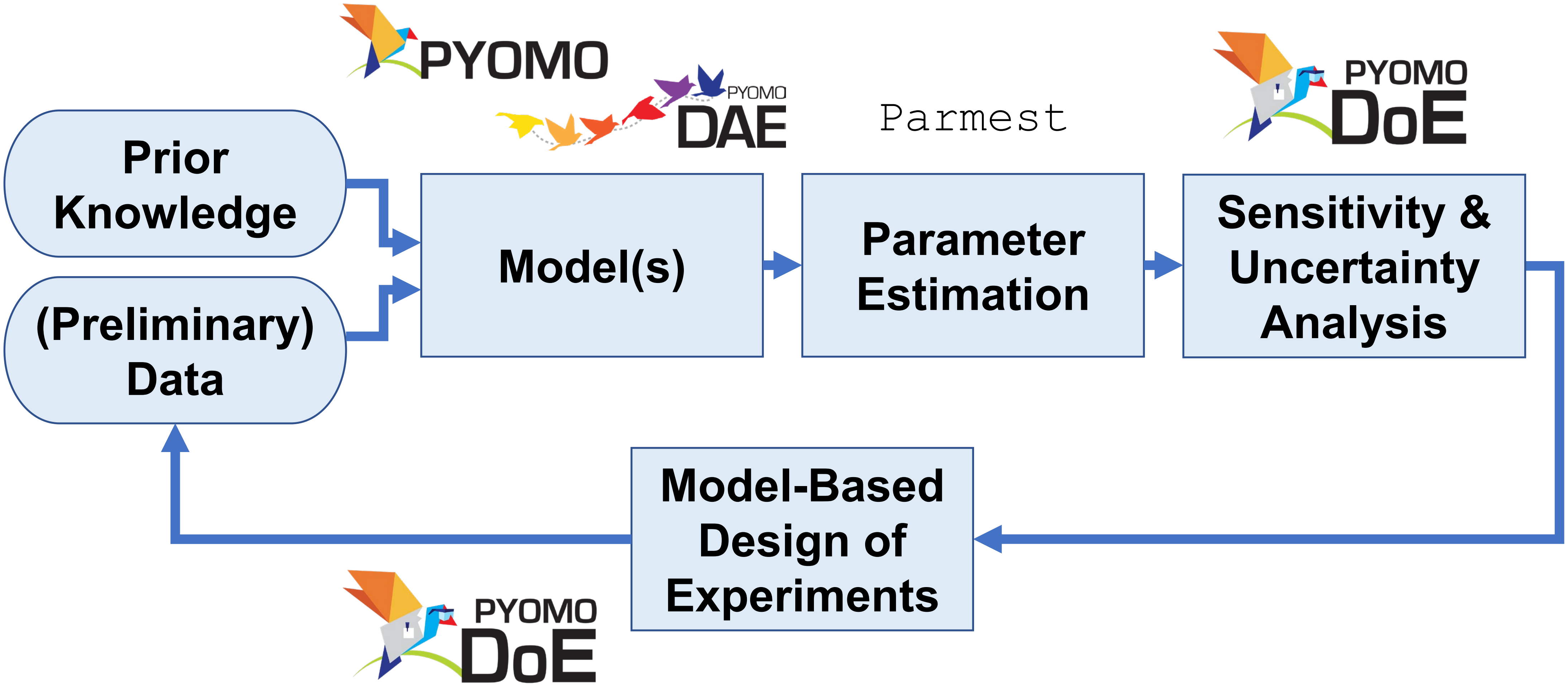
- Bethany Nicholson
- Miranda Mundt
- John Siirola

Membrane Case Study (ND):

- William Phillip
- Jonathan Ouimet

Take Away Message

While there is some "art" to building science-based models…



…statistical tools can maximize the usefulness of data