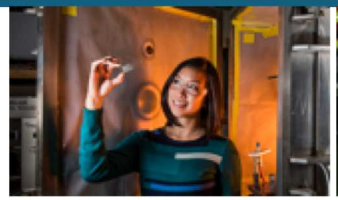
SAND2019-4424PE

Parameter Estimation and Process Models



Andrea Staid and Bill Hart

April 4, 2019

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.

SAND2018-12122 PE

Parameter Estimation using Process Models

Objectives: Develop process models of proliferation-related activities, and, given a process model, employ parameter estimation and pattern-matching techniques to detect and characterize activities of interest within a data stream of observables.

Approach: Apply parameter estimation and data reconciliation techniques to "physics based" process models that describe manufacturing processes, chemical processes, construction, and related activities.

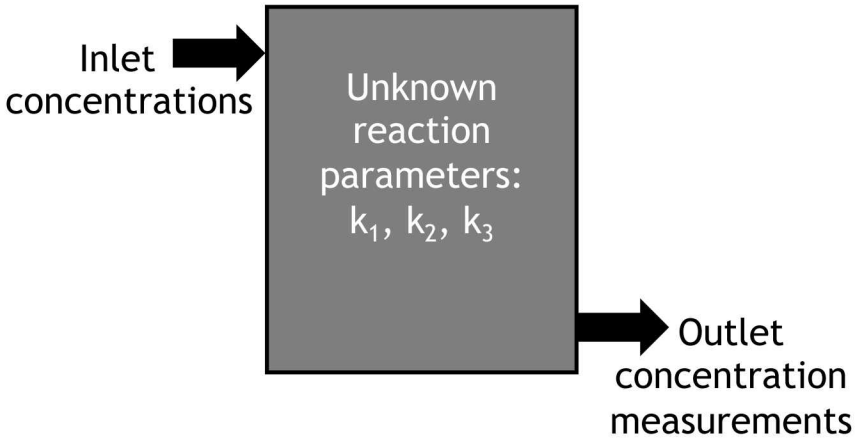
SNL Focus: Optimization-based process models and parameter estimation

- Complements previous work on the Facility Modeling Toolkit
- Can address issues of multiple hypotheses
- Can be used for data reconciliation
- Can incorporate uncertainty for confidence assessments

Capabilities of Parameter Estimation

Example: In a facility with a known chemical reactor in operation, we can measure (with uncertainty) the concentrations of outflow from the facility.

- What can we say about the underlying manufacturing process?
- We can estimate the values of unknown, internal parameters – where these values lie may provide critical information about the exact nature of the process
- Identifying which one (of multiple) hypotheses is most likely using both models and data



By estimating k_1, k_2, and k_3 we can:

- Assess our confidence in our hypothesis
- Identify the possibility of multiple hypotheses
- Determine whether the data available is sufficient for high-confidence estimates
- Follow changes in operation over time

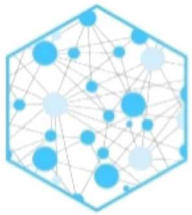
Quantifying Uncertainty

Methods to estimate unknown parameters are common; what is more critical in many problems is also estimating the uncertainty of those estimates or understanding what it means to get a bad estimate

- Confidence regions around your estimates enable trust in both the model and the data; this is critical when the results are used in sensitive decision-making contexts

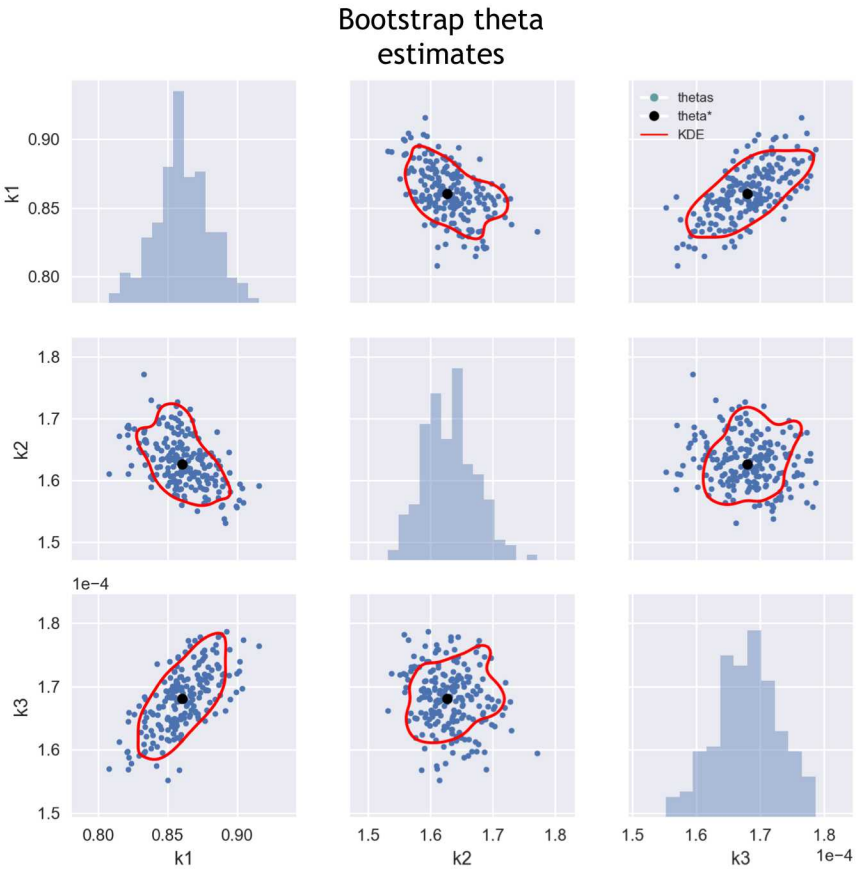
Existing work on the IDAES project has been focused on building up capabilities for dealing with this uncertainty

- Includes methods that work well in sparse-data situations (i.e., bootstrap sampling)
- This software is now being used for broader applications, and has great potential for ADAPD

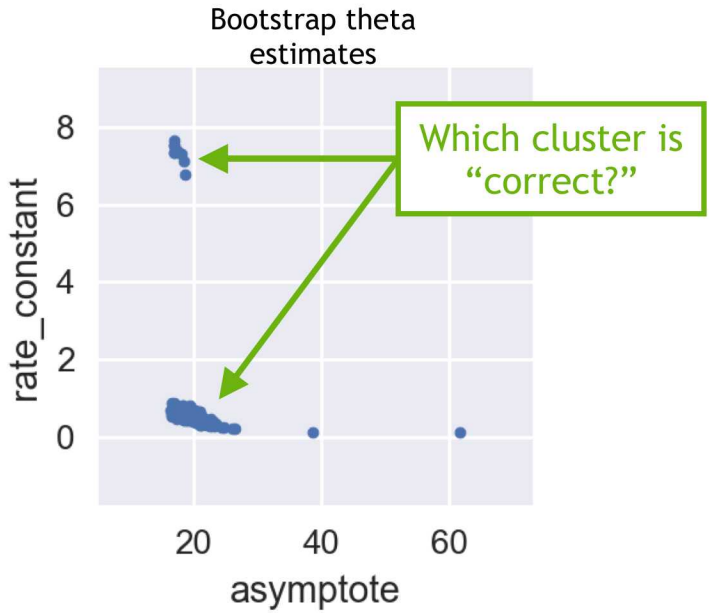


IDAES
Institute for the Design of Advanced Energy Systems

What can this look like?



Reactor design, estimating values of (and relationships among) k_1, k_2, and k_3



An example of multiple hypotheses

- What does it mean to end up with two clusters? And a long tail?
- Is your data sufficient? Do you trust the solution?
- In many cases – an issue of too little data, but could also indicate critical operational transition points

Our Project Plan

The availability of appropriate data and models has forced us to re-evaluate our plan in this space…

- Do we have continuous process models?
  - Many HP2 applications have potential, but we don't yet have models
  - Extensions to our existing parameter estimation methods should be motivated by ADAPD relevance
- Will we need discrete-event process models?
  - Mixed-integer parameter estimation would be a challenging research area
  - Discontinuous dynamics in ADAPD applications may motivate this in FY20

We have explored methods for evaluating robustness of parameter estimation solutions

- Are there multiple optima? This is especially important in sparse-data situations
- Cluster-based global optimization applied to parameter estimation

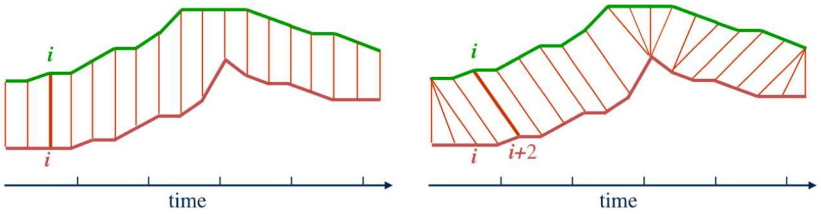
Without models, what *can* we do?

If We Don't Have a Process Model

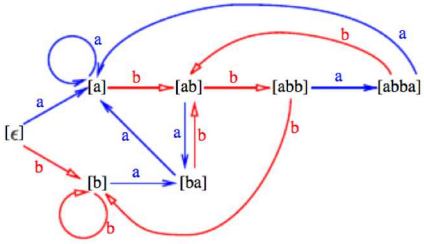
We may not have a trusted process model but we have trusted data exemplars

Idea: Perform intelligent pattern matching

Dynamic time warping[1]



Multiple pattern matching using Markov Chains[2]



Status:

- Explored application to HERTZ (missing time series data)
- Will explore applications to OM and new HERTZ data (FY19)
- Will explore synergies with HP3 (w/D Stracuzzi)

Research challenge: rigorous confidence in dynamic warping with missing/uncertain data

1. Elena Tsiporkova, Dynamic Time Warping Algorithm for Gene Expression Time Series. Universiteit Gent
2. Lladser, Manuel E., Meredith D. Betterton, and Rob Knight. "Multiple pattern matching: A Markov chain approach." *Journal of mathematical biology* 56.1-2 (2008): 51-92.

Future Plans

Will explore applications for parameter estimation using process models to OM and new HERTZ data (FY19)

Will explore synergies with HP3 (w/Dave Stracuzzi)

Research challenge: can we achieve confidence in dynamic time warping with missing or uncertain data

Near Term

Will continue to explore opportunities for process model applications in FY19, especially w.r.t. HP2

As we explore opportunities for process model applications, we will assess whether this is a key element of the research and whether new methods are needed to augment current capabilities

Will explore applications of robust parameter estimation to machine learning in ADAPD

TBD,
If/when models become available