INL/CON-24-77803-Revision-0

Verification, Validation, and Calibration Through a Causal Lens

May 2024

Ronald Louis Gonzales, Diego Mandelli, Congjian Wang, Mohammad Gamal M Mostafa Abdo, Paolo Balestra, Sunming Qin, Victor Petov, Zachary Welker, Annalisa Manera

INL/CON-24-77803-Revision-0

Verification, Validation, and Calibration Through a Causal Lens

Ronald Louis Gonzales, Diego Mandelli, Congjian Wang, Mohammad Gamal M Mostafa Abdo, Paolo Balestra, Sunming Qin, Victor Petov, Zachary Welker, Annalisa Manera

May 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

May 16, 2024

VVUQ

Ron Gonzales
Postdoctoral Researcher

Verification, Validation, and Calibration Through a Causal Lens

R. Gonzales, D. Mandelli, C. Wang, M. Abdo, P. Balestra, S. Qin

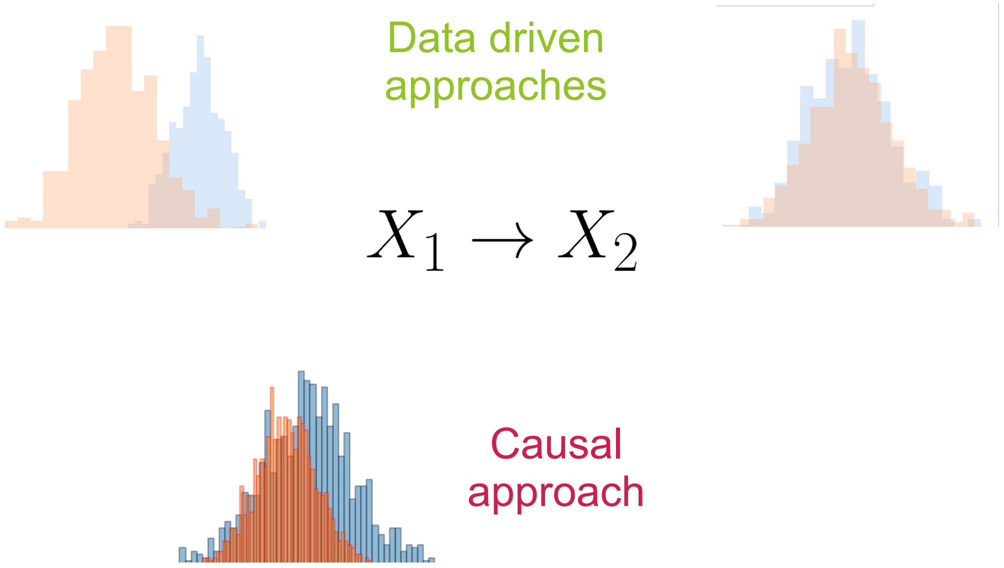
(*Idaho National Laboratory*)

Z. Welker, V. Petrov, A. Manera

(*University of Michigan*)

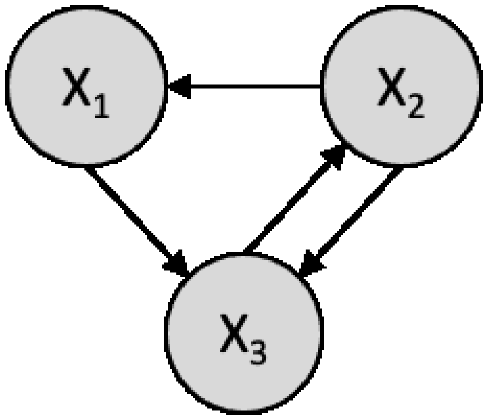
Introduction

- Simulation models are defined by a set of constituent and dynamic mathematical equations (e.g., conservation laws, equations of motion, thermodynamic laws)
  - Encoded cause-effect information
- Current methods for model validation: Data driven
  - Quantitative comparison between datasets (i.e., simulated vs. measured data)
  - No explicit consideration of the hypotheses behind them (e.g., boundary conditions) or the structure of the employed models
- This paper: Validation approach based on causal inference
  - Capture the causal relationships between data elements rather than looking at their associations



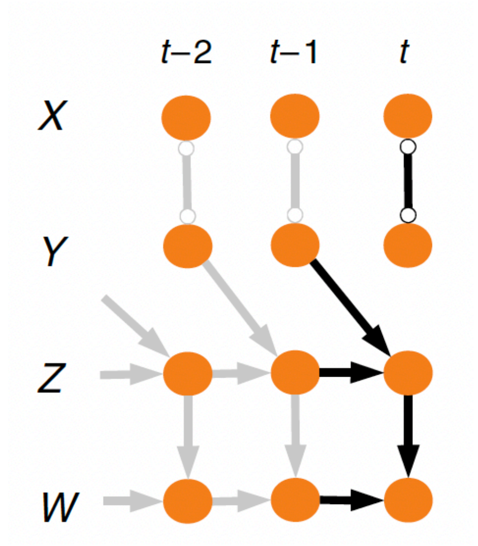
Causal Inference and Discovery

- Probabilistic inference (classical statistics or machine learning methods) is designed to
  - Model/construct a probabilistic distribution from historic observations
  - Identify patterns out of this distribution (e.g., to predict the outcome of future observations)
- Causal inference aims to identify the possible mechanisms that might have generated historic observations
- Structural causal models (SCMs): Causal models that can model cycles, interventions, and counterfactuals
  - Given a set of N variables , an SCM consists of a set of structural equations of the form
  - SCM can be visualized in a graphical form: Directed acyclic graph (DAG)
- SCMs and dynamic equations

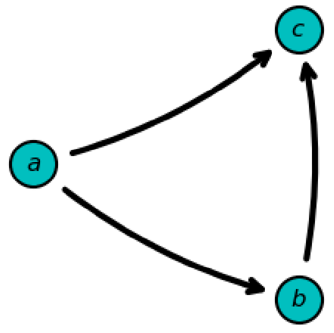


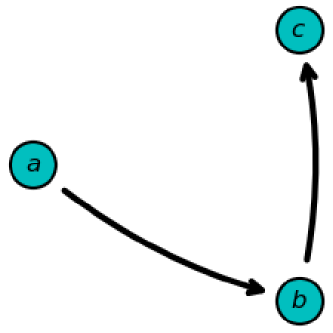
Discovery Methods

- **Input:** Temporal profile of variables
- **PC:** Perform independence tests on all possible connections in the DAG given sufficient conditioning set
  - Classical formulation:
- **PCMCI:** Check for statistical independence
  - Algorithm steps
    1. PC1: (modified PC) Conditional independency testing among variables in a specific timestamp
    2. MCI (Momentary conditional independence): determine causal relationship between variables in different timestamps



Source: Runge, J., et al. Inferring causation from time series in Earth system sciences. Nature Communications, vol. 10, 2553 (2019)





Discovery Methods

- **SINDy (sparse identification of nonlinear dynamics):** Sparse regression to find a linear combination of basis functions that best capture the dynamic behavior of the physical system. That is, estimating in , where

  Here is the matrix of time series data, is a set of basis functions and is   a presumably sparse coefficient matrix.

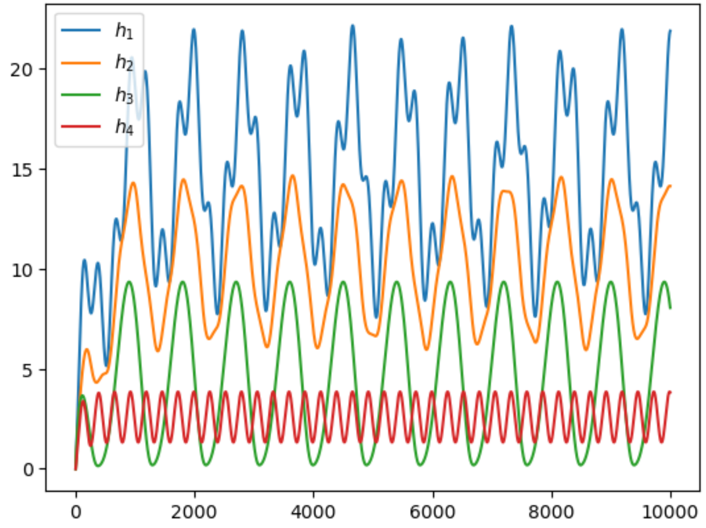
- **MMD:** An intervention-based Maximum Mean Discrepancy (MMD) testing method to automatically identify causal structure of a control system. Tests the physical influence when manipulating the input variables. It is a kernel-based two-sample test method to measure the distance between the kernel mean embeddings of two random variables in the reproducing kernel Hilbert space (RKHS).

Modeling Assumptions

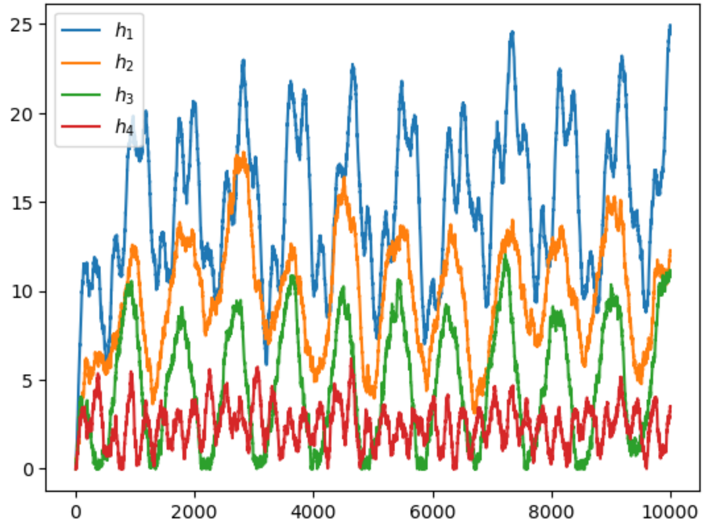
- Causal sufficiency: Assume that the set of variables contains all common causes of any subset of variables
  - Normally guaranteed in a simulation model
  - Experimental data might contain a subset of variables that might have a causal role in the experiment dynamics and are not part of the actual dataset
- Constant causal structure: Causal relations between variables do not change with time
  - E.g., sudden activation of a pump during a cooling transient would cause a change in the SCM
  - If these system configuration changes are recorded, then the SCM generation/quantification should be redone after each configuration change
- Data synchronicity: Set of time instants equally spaced in time they are identical for each variable
  - If these conditions are not satisfied, then proper data synchronization techniques (e.g., time series resampling, time series interpolation) should be applied
- Noisy data: Most methods are able to process the presence of noise as part of the time series
  - Noise terms associated with each variable are assumed to be independent

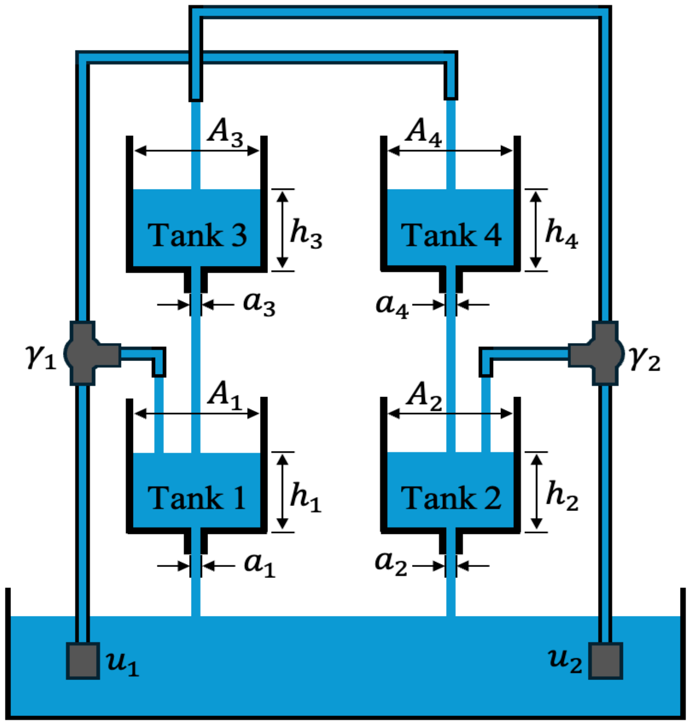
Quadruple Tank Problem

Deterministic

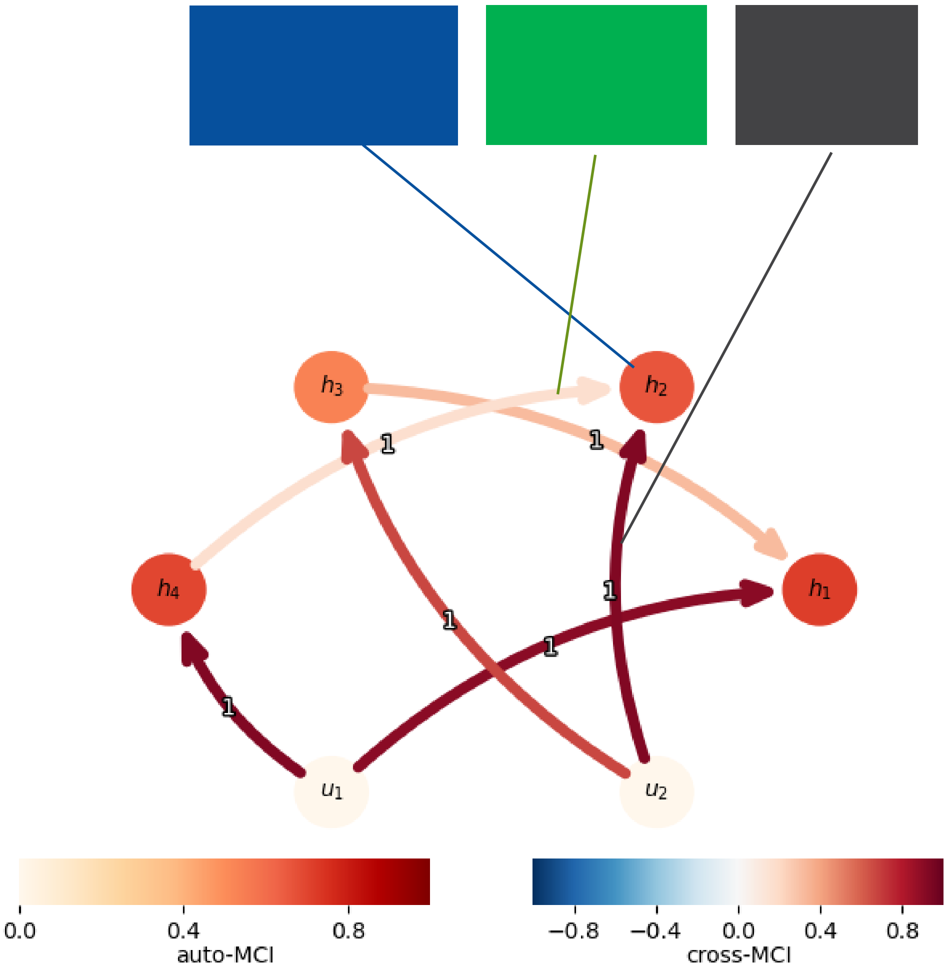


Stochastic



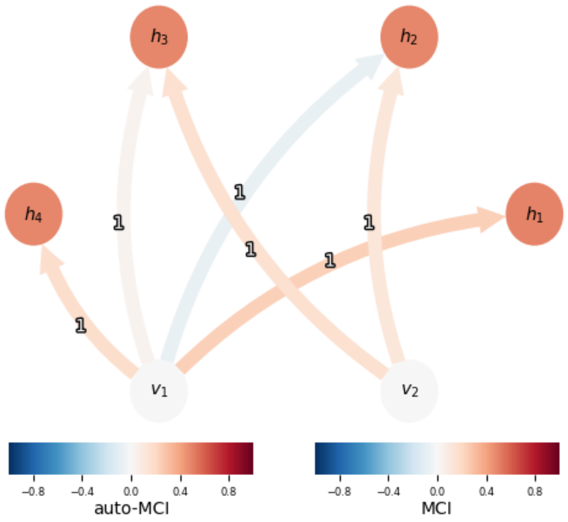


Quadruple Tank Analysis PC/PCMCI

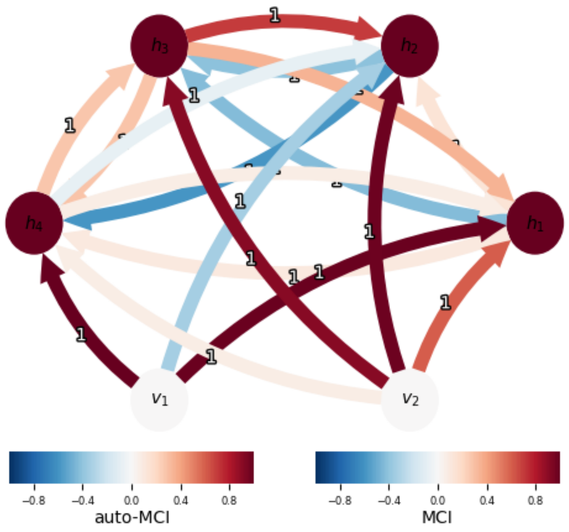


Internal Noise

External Noise



Deterministic



Quadruple Tank Analysis SINDy

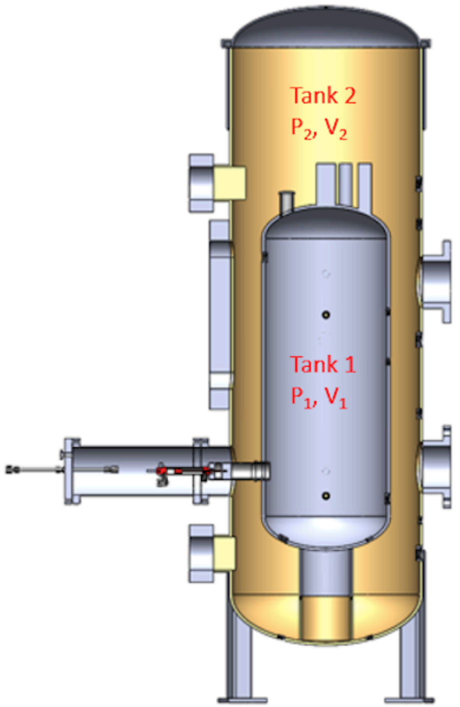
Modeled ODE

External Noise

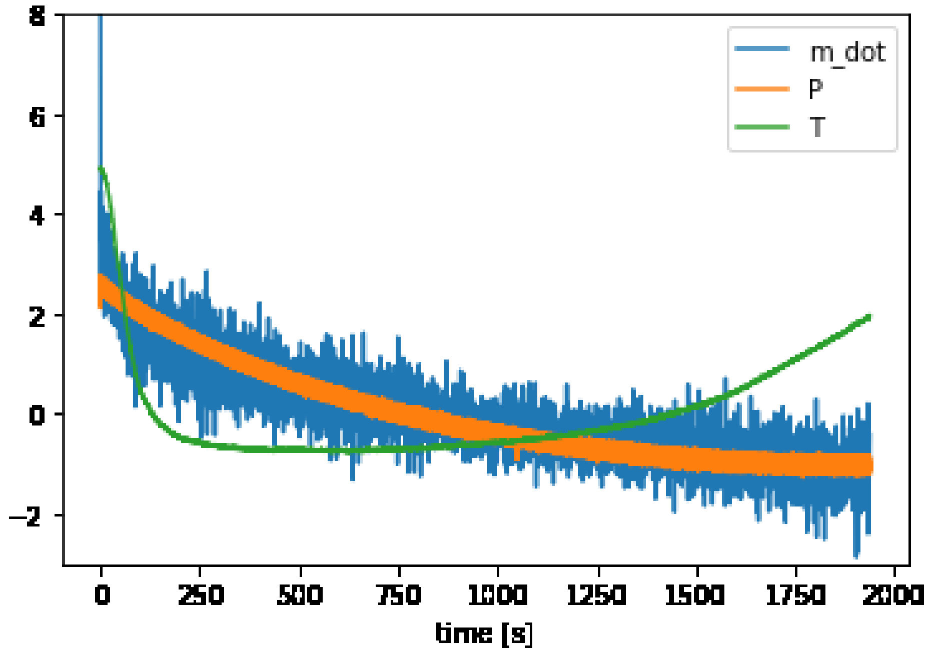
Deterministic

Internal Noise

Analysis Of HAIRE Experimental Setup



- Validation test: Helium Air Ingress gas Reactor Experimental (HAIRE) facility at University of Michigan
- Available measurements: Pressure *P* and temperature *T*
  - Mass flow rate is calculated from perfect gas state equation
- Depressurization experiments
  - Reactor vessel is vacuumed and pressurized with the gas of interest
  - The initiation mechanism is released which allows the depressurization to start
  - The gaseous depressurization continues until an equilibrium pressure is reached between atmosphere and the pressure vessel
- Experiments were performed with a variety of initial pressures, gases, and crack diameters



Analysis Of HAIRE Experimental Setup: Results

- Simulation model: Temporal evolution of tank pressure and mass flow rate (iso-thermal conditions)
  - Mass flow rate formulation changes when critical pressure is reached
  - SCMs are generated separately before and after critical pressure is reached

Experimental

	P	$\dot{m}$	T
P	1	-3.4E-7	3.6E-8
$\dot{m}$	8.7E-6	1	9.5E-7
T	-1.7E-5	1.9E-5	1

	P	$\dot{m}$	T
P	1	-2.9E-7	4.2E-7
$\dot{m}$	9.5E-8	1	2.7E-7
T	-2.5E-7	9.92E-9	1

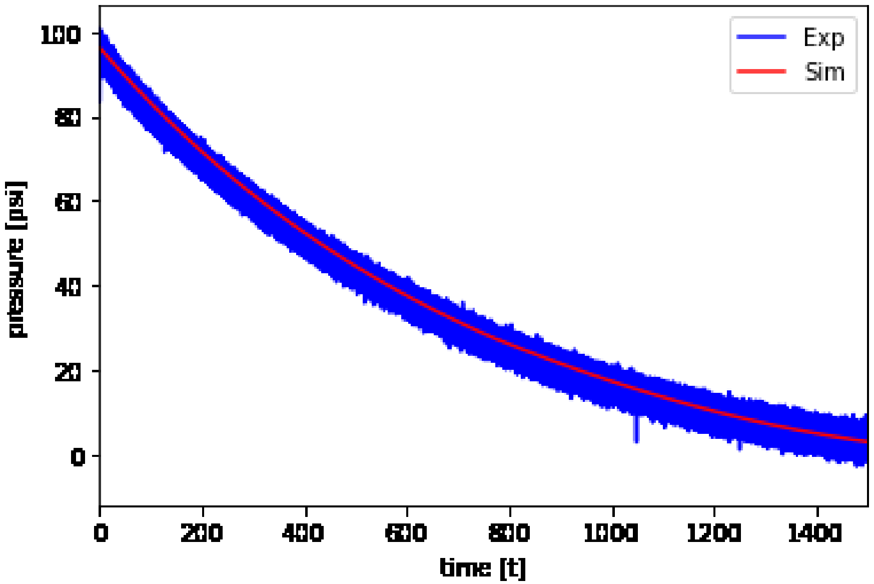
	P	$\dot{m}$	T
P	1	-2.5E-7	4.4E-7
$\dot{m}$	2.6E-7	1	3.2E-7
T	-6.4E-7	4.7E-7	1

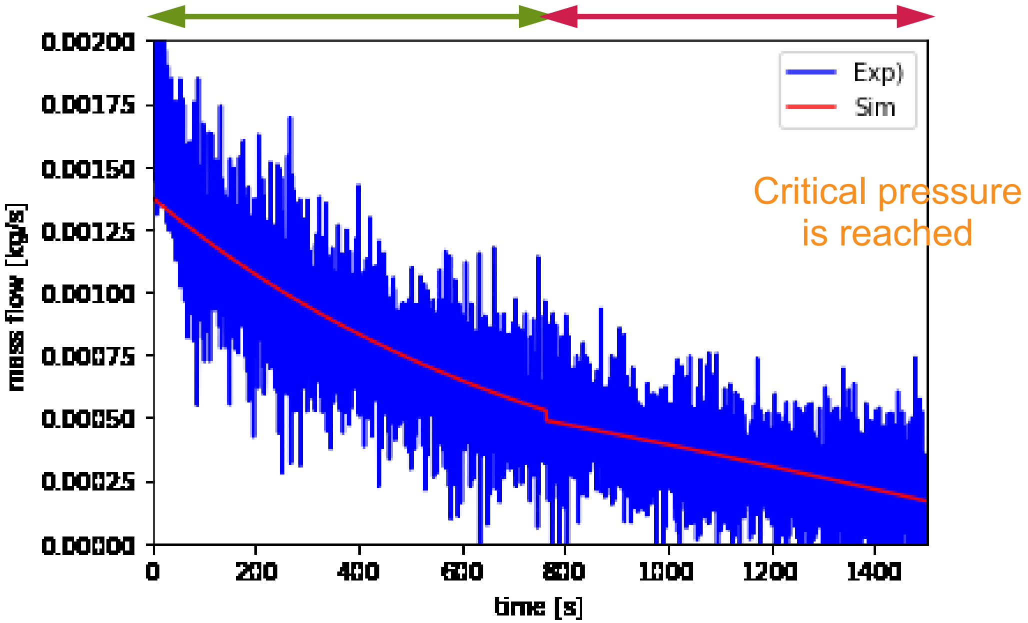
	P	$\dot{m}$	T
P	1	-2.8E-7	6.1E-8
$\dot{m}$	3.7E-7	1	5.1E-8
T	4.1E-8	-6.1E-7	1

Simulated

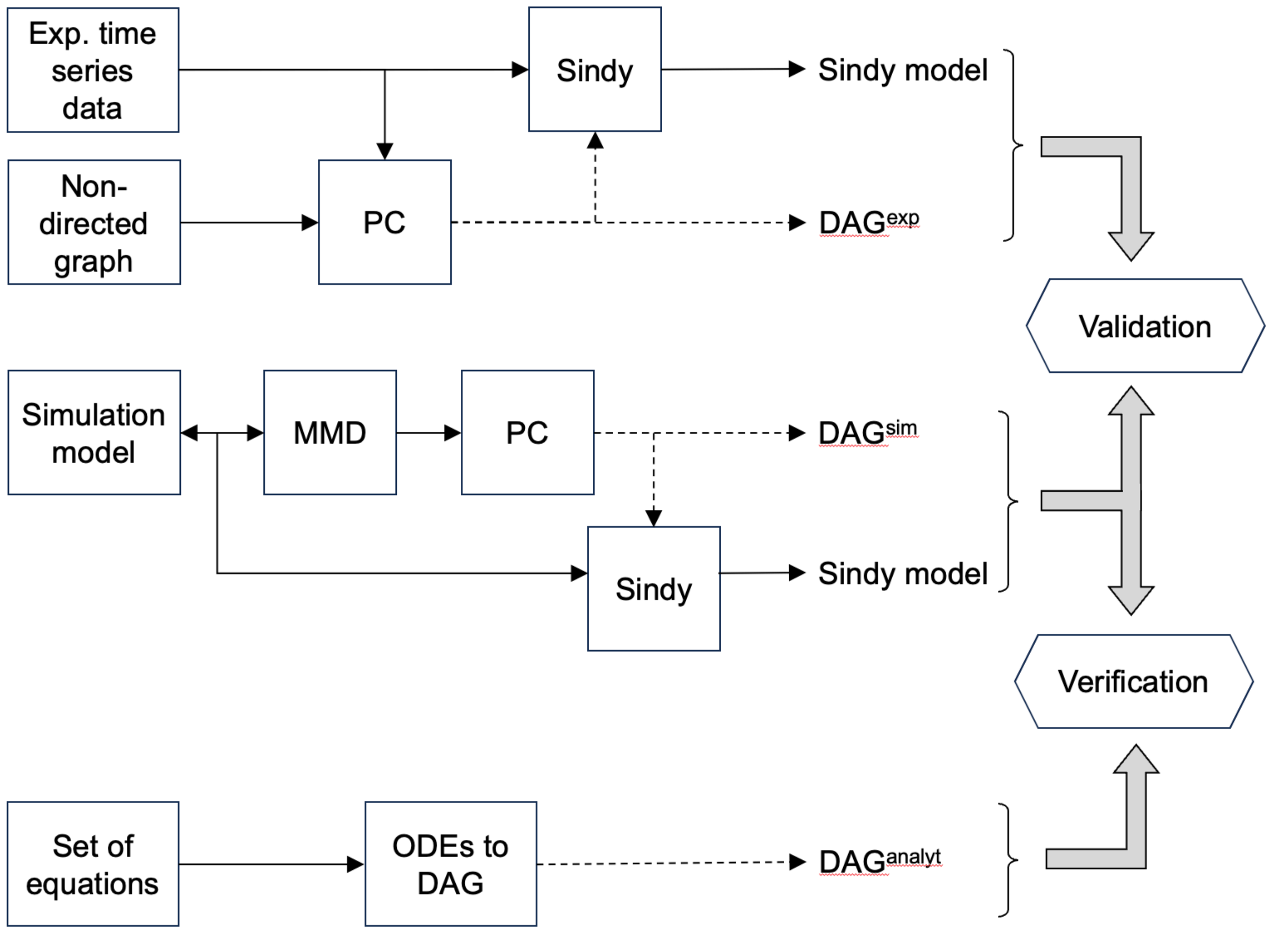
	P	$\dot{m}$
P	1	-2E-7
$\dot{m}$	7.6E-8	1

	P	$\dot{m}$
P	1	-3E-7
$\dot{m}$	4.3E-8	1





Current direction



Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy. INL is the nation's center for nuclear energy research and development, and also performs research in each of DOE's strategic goal areas: energy, national security, science and the environment.

WWW.INL.GOV