A Roadmap for ML-assisted Ice Microphysics Parameterizations in PUMAS

Credit: pynsrps

Tuesday, Feb 4, 2025

2025 CESM Winter Working Group Meeting

Joseph Ko[1], Jerry Harrington[2], Kamal Chandrakar[3], Jasmine Remillard[4], Marcus van Lier-Walqui[4, 1], Kara Sulia[5], Vanessa Przybylo[5], Trude Eidhammer[3], Hugh Morrison[3], Kara Lamb[1]

[1]Columbia University, [2]Penn State University,
[3]NCAR, [4]NASA GISS, [5]University at Albany

Outline

1. Motivation
2. Lab + NODE
3. In situ + ML
4. LES + Bayesian + ROM
5. Future steps + timeline

Outline

1. Motivation
2. Lab + NODE
3. In situ + ML
4. LES + Bayesian + ROM
5. Future steps + timeline

Outline

1. Motivation
2. Lab + NODE
3. In situ + ML
4. LES + Bayesian + ROM
5. Future steps + timeline

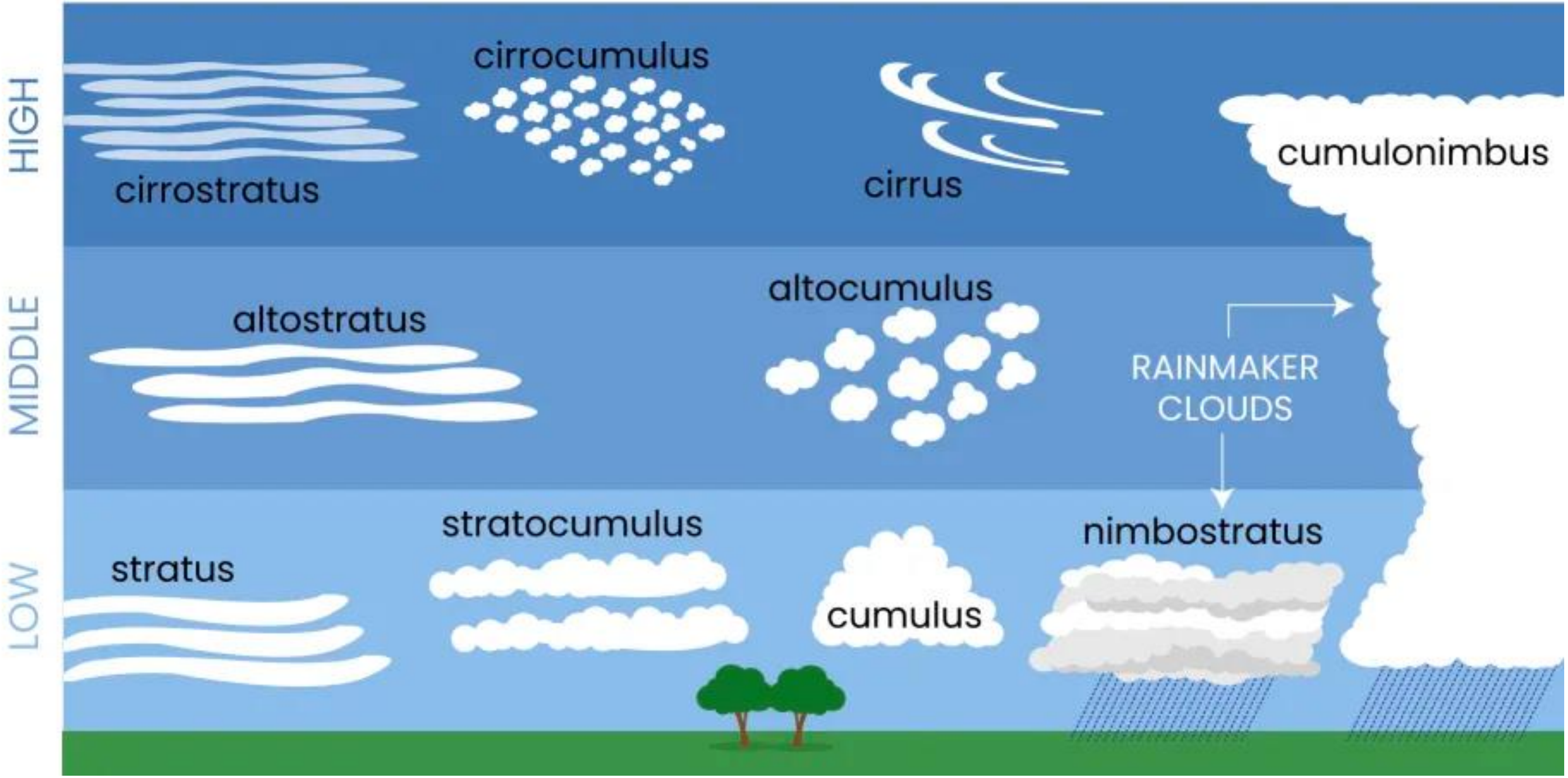
Outline

1. Motivation
2. Lab + NODE
3. In situ + ML
4. LES + Bayesian + ROM
5. Future steps + timeline

Outline

1. Motivation
2. Lab + NODE
3. In situ + ML
4. LES + Bayesian + ROM
5. Future steps + timeline

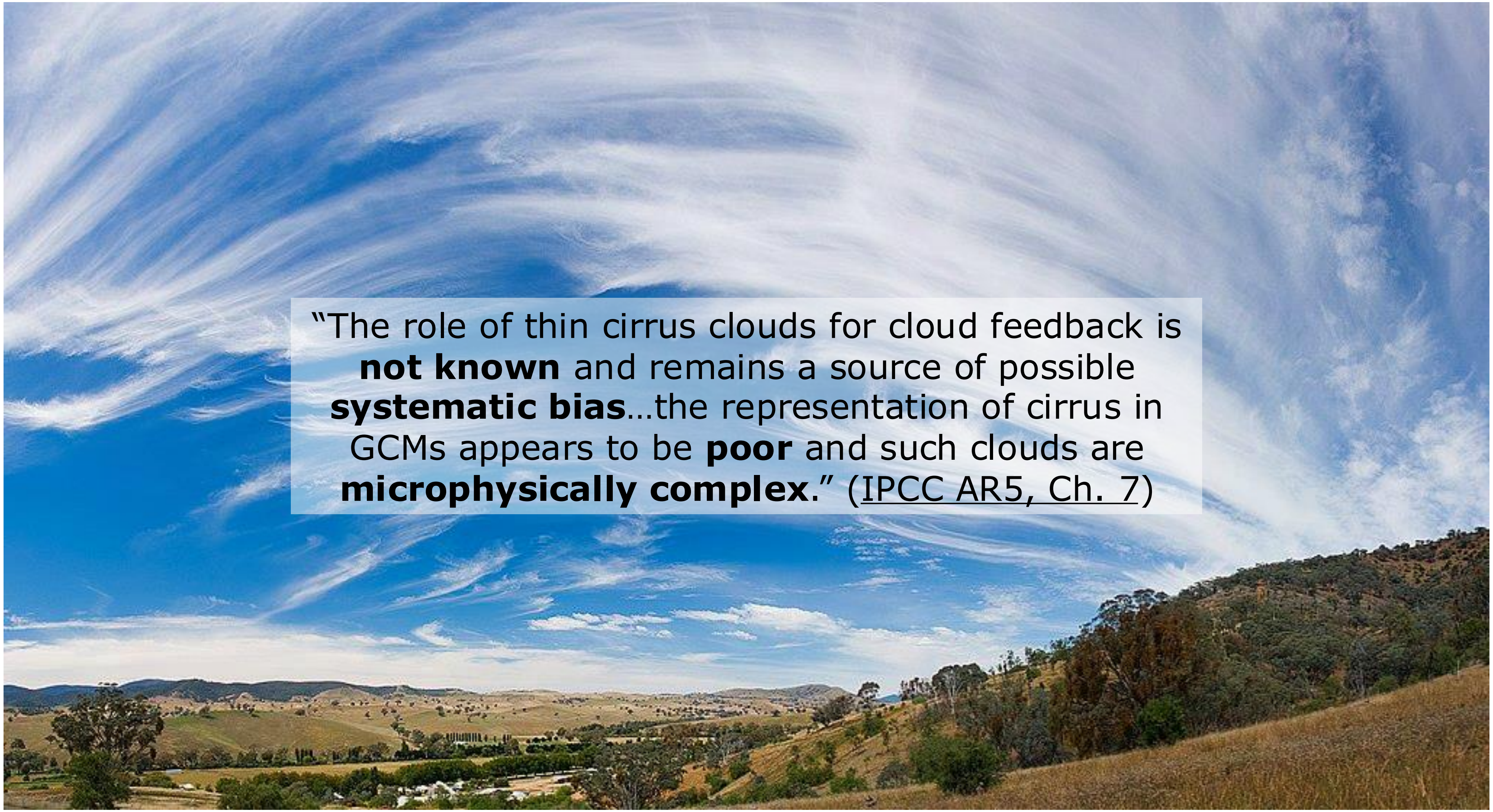
Clouds strongly impact the climate



Source: UCAR

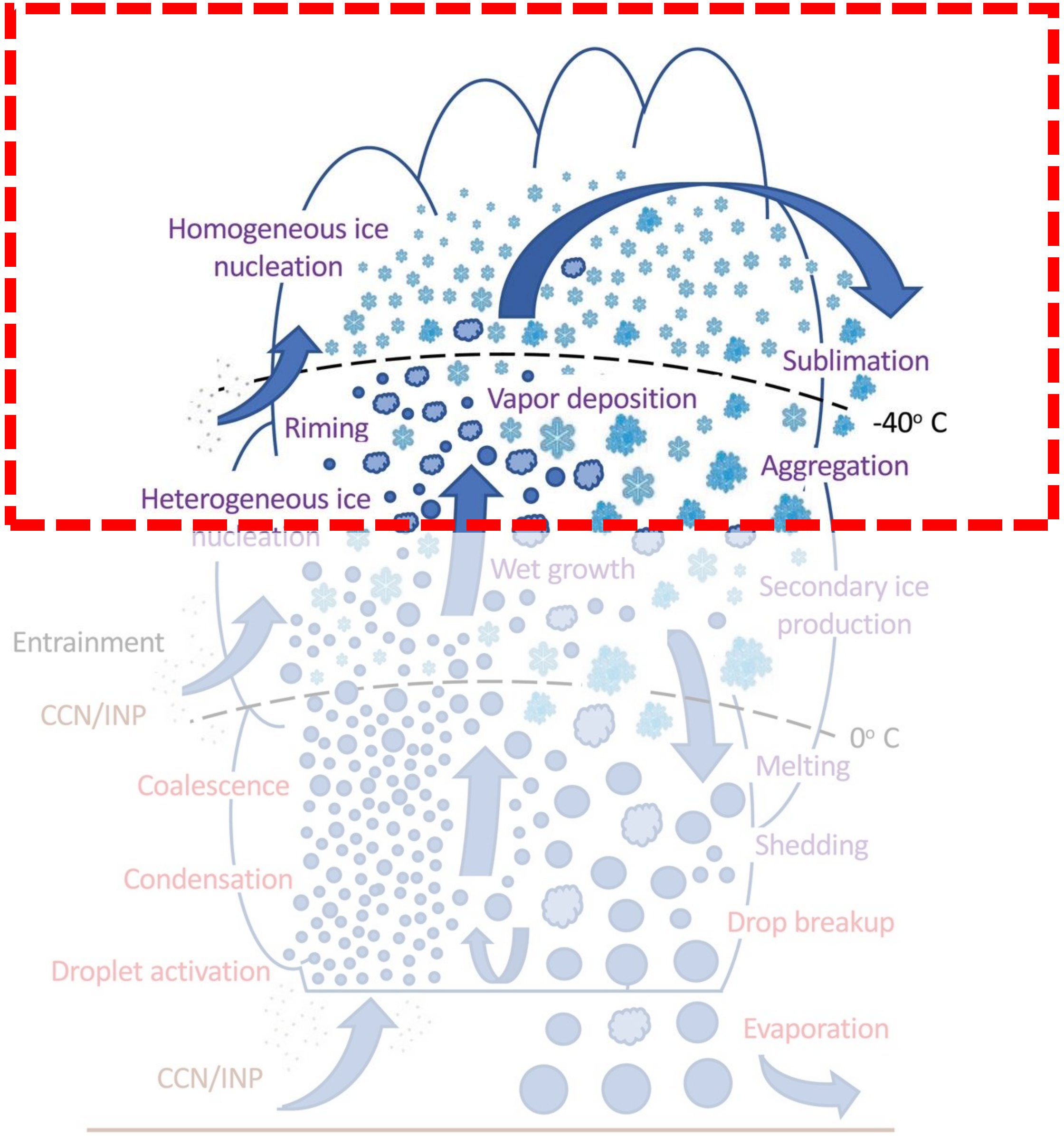
Clouds impact Earth's **energy balance** and **hydrologic cycle**

Ice clouds are poorly understood

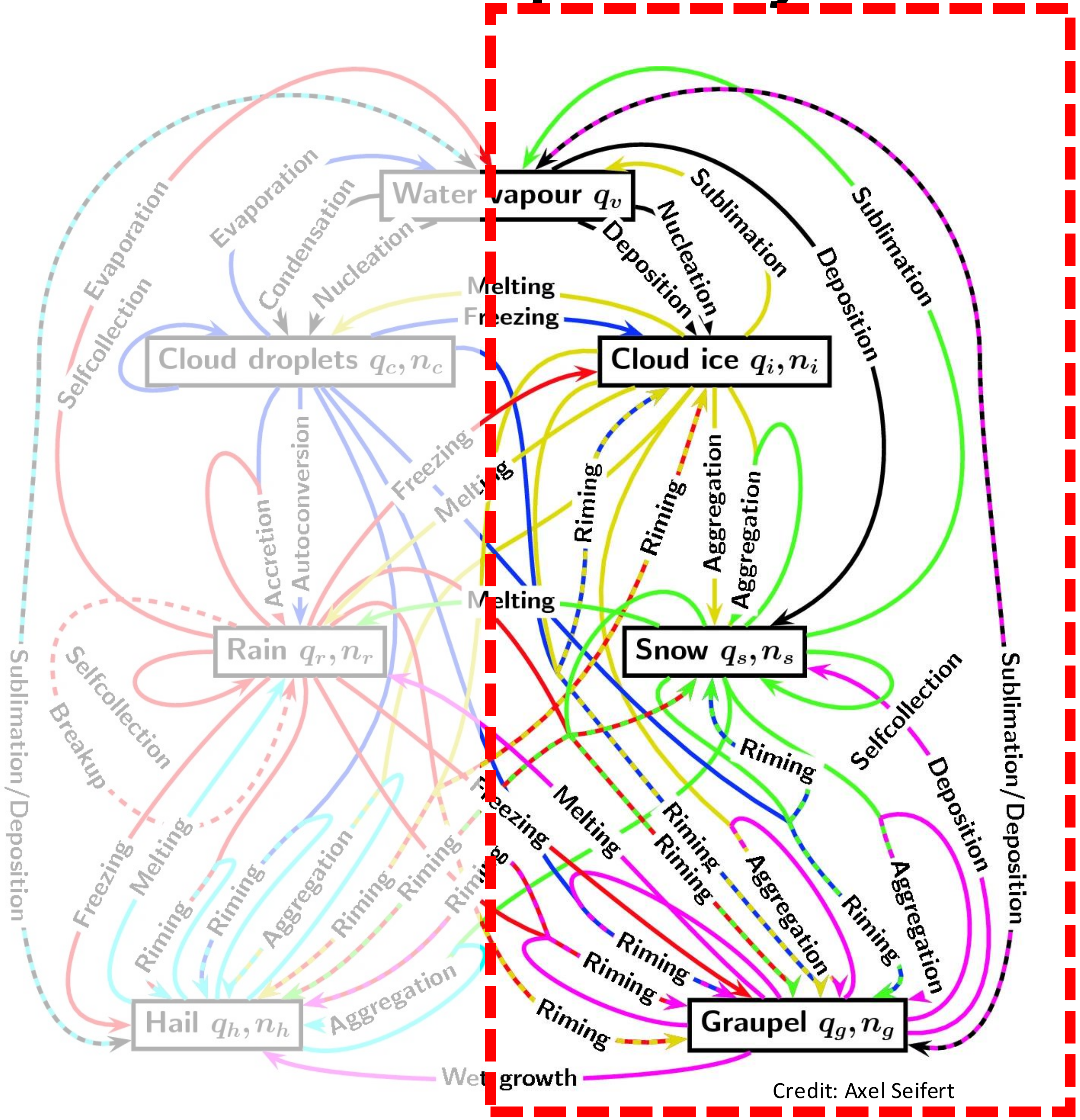


Source: Fir0002/Flagstaffotos

Cloud Microphysics: small-scale processes that describe the formation, evolution, and interaction of droplets/crystals



Credit: Morrison et al. 2020



Credit: Axel Seifert

2. Lab + NODE

Learning ice growth rates with lab measurements and neural ODE's (Kara Lamb, Jerry Harrington)

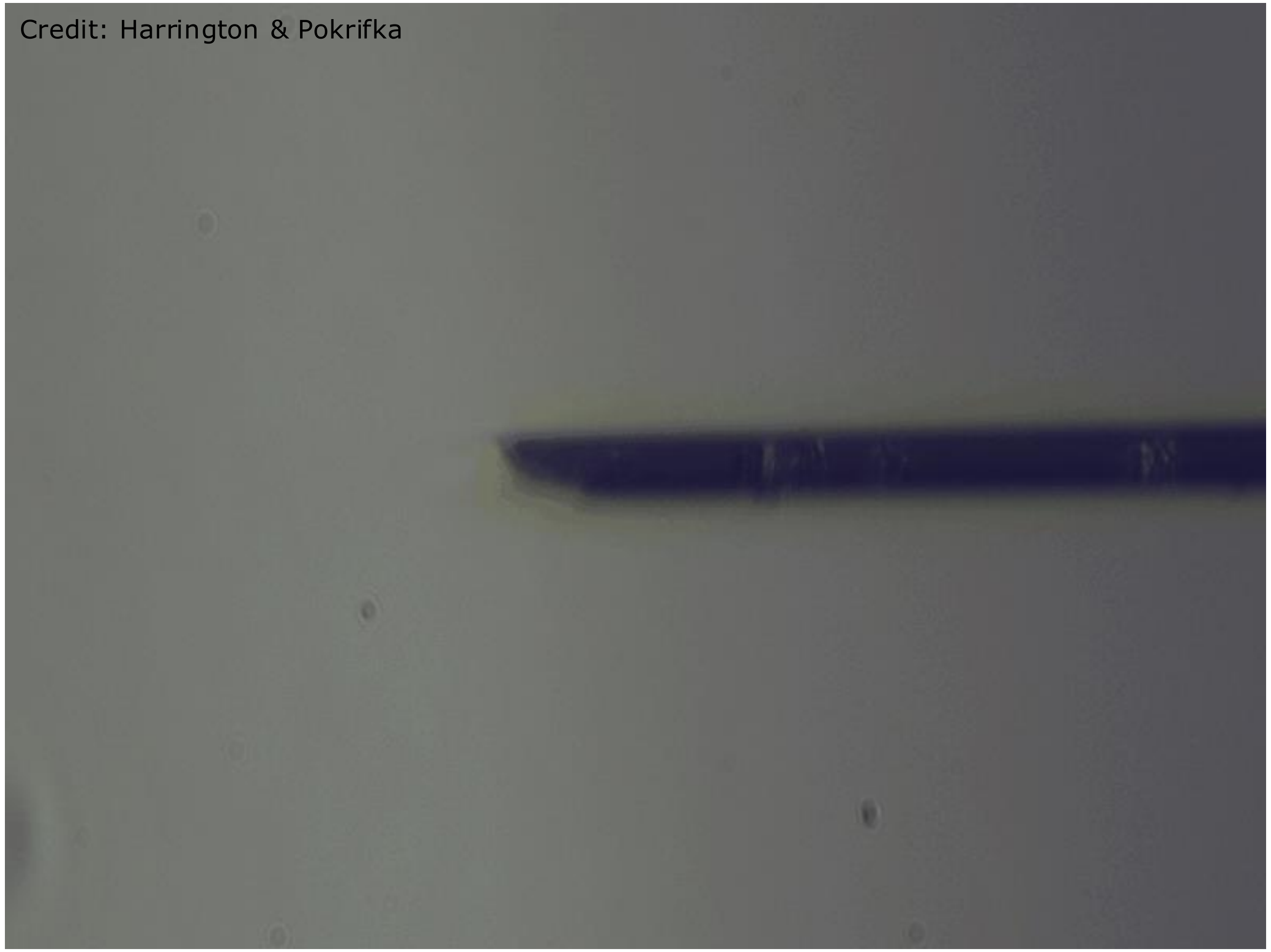
Can we use lab measurements to improve representation of vapor depositional growth?

Single crystal ice mass growth rate (Pruppacher & Klett, 1997)

$$\frac{dm_p}{dt} = \frac{4\pi C(S_{ice} - 1)}{\frac{RT_g}{\hat{e}_{ice}(T_g) D_w^* M_w} + LH}$$

$$D_w^* = \frac{D_w}{\frac{r}{(r+\Delta_v)} + \frac{D_w}{r\boxed{\alpha_D}}\left(\frac{w\pi M_w}{RT_a}\right)^{1/2}}$$

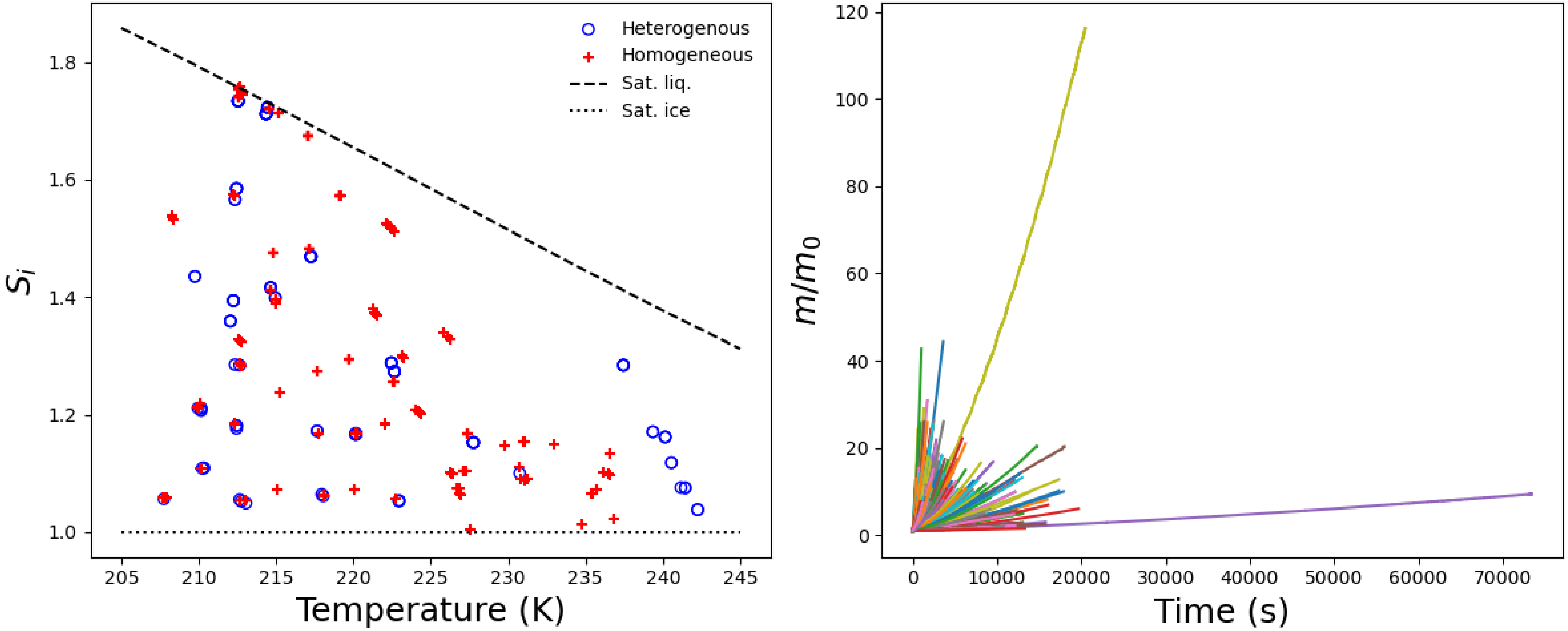
The functional dependence of α_D is uncertain (typically assumed to be a constant value)



Lamb & Harrington, 2024

Mass growth time series from levitation diffusion chamber experiments (Harrison et al. 2016; Pokrifka et al. 2020, 2023)

N = 290 experiments



Lamb & Harrington, 2024

Neural ODE (NODE) to learn growth rate

a) Ice growth rate from theory

$$\frac{dm}{dt} = 4\pi r(S_i - 1)\boxed{G(r, T, S_i, \boxed{\alpha})}$$

Strong constraint

$$\boxed{\alpha = f_\alpha(S, T|\theta_\alpha)}$$

Weak constraint

$$\boxed{\frac{G}{G_c} = f_G(S, T, m|\theta_G)}$$

b) NODE model to predict mass ratio

$$ODESolve(m_0, \frac{dm}{dt}, f, \theta_f, t_0, ..., t_M)$$

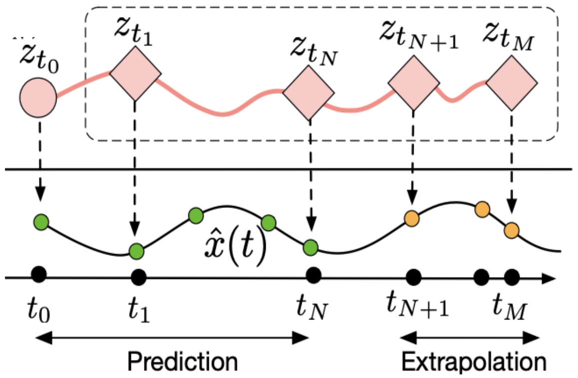
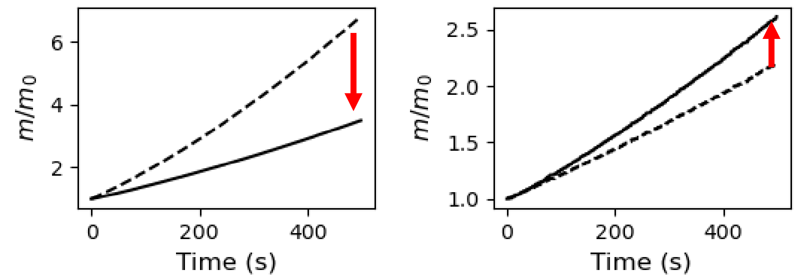


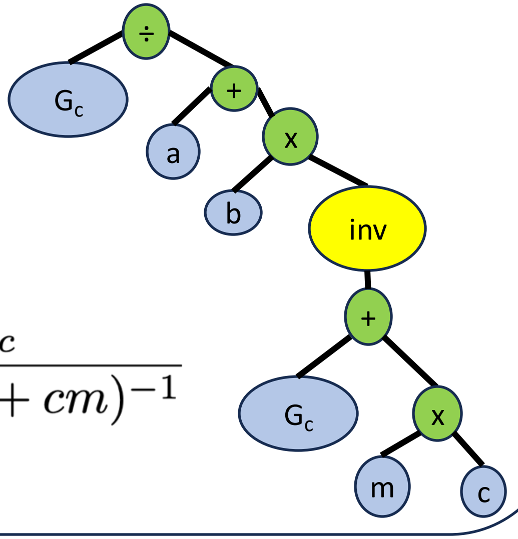
Figure adapted from Chen et al. 2018

c) Minimize distance between obs. & model by optimizing neural network (NN) weights using stochastic gradient descent

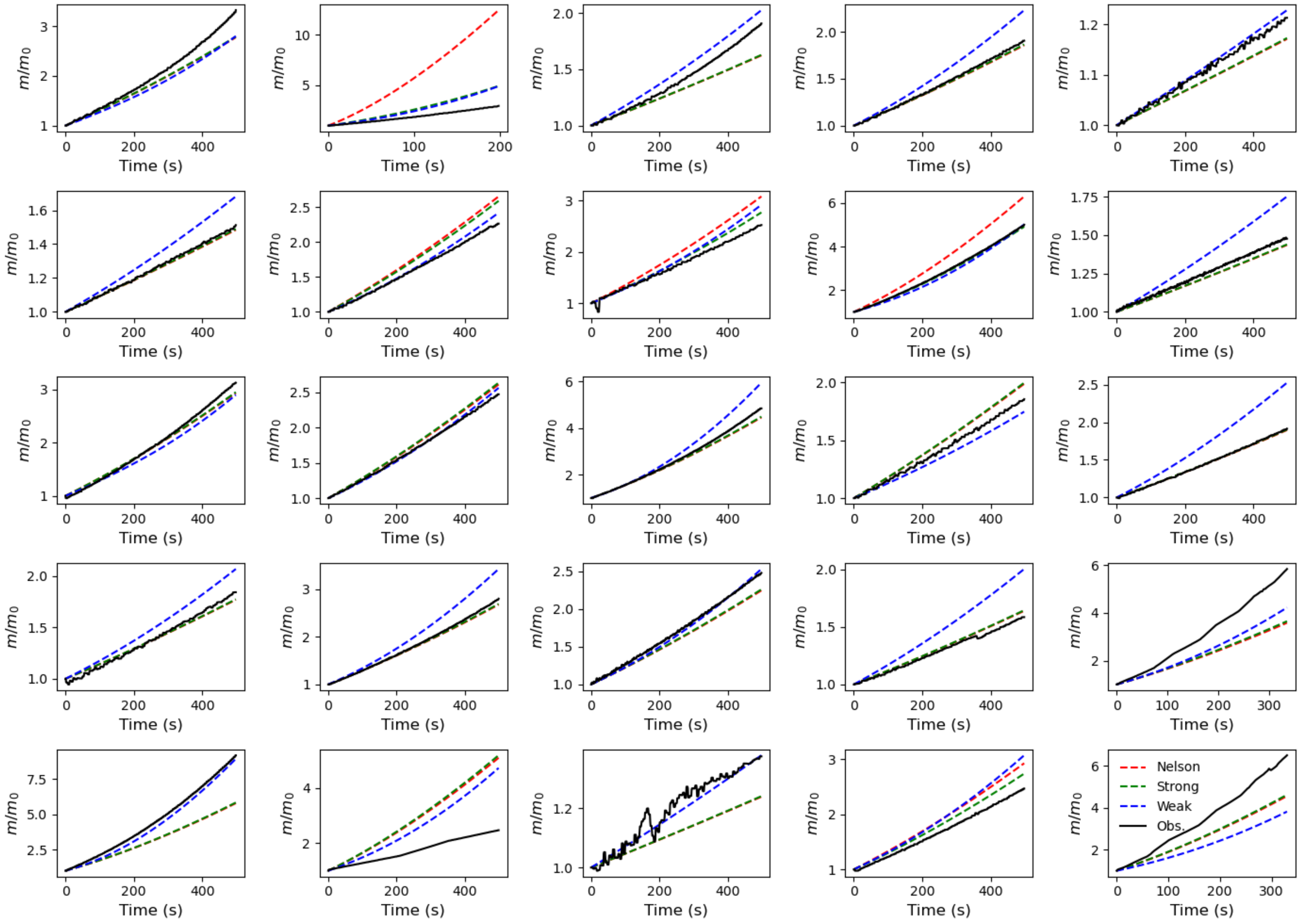


d) Symbolic regression on trained NN

$$G = \frac{G_c}{a + b(G_c + cm)^{-1}}$$



Lamb & Harrington, 2024



$$\frac{dm}{dt} = 4\pi r(S_i - 1)\boxed{G(r, T, S_i, \boxed{\alpha})}$$

Strong constraint

$$\boxed{\alpha = f_\alpha(S, T|\theta_\alpha)}$$

Weak constraint

$$\boxed{\frac{G}{G_c} = f_G(S, T, m|\theta_G)}$$

Weakly constrained NODE model performed best:

- Lowest error (MSE)
- Best fit for 154 of 290 experiments

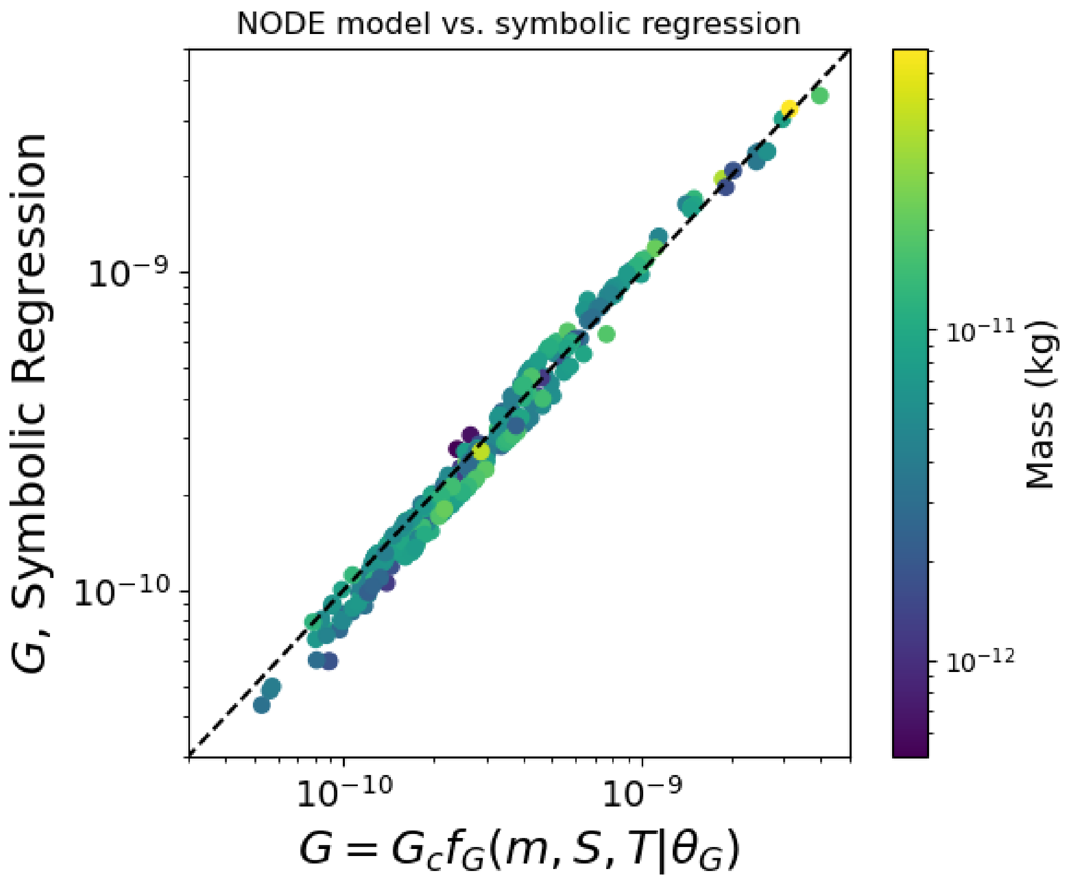
Functional form learned with symbolic regression (PySR)

$$\boxed{G = \frac{G_c}{a + b(G_c + cm)^{-1}}}$$

$$a = 0.6517688, b = 2.98707 \times 10^{-9}, c = 1000$$

Can be directly implemented into growth parameterization

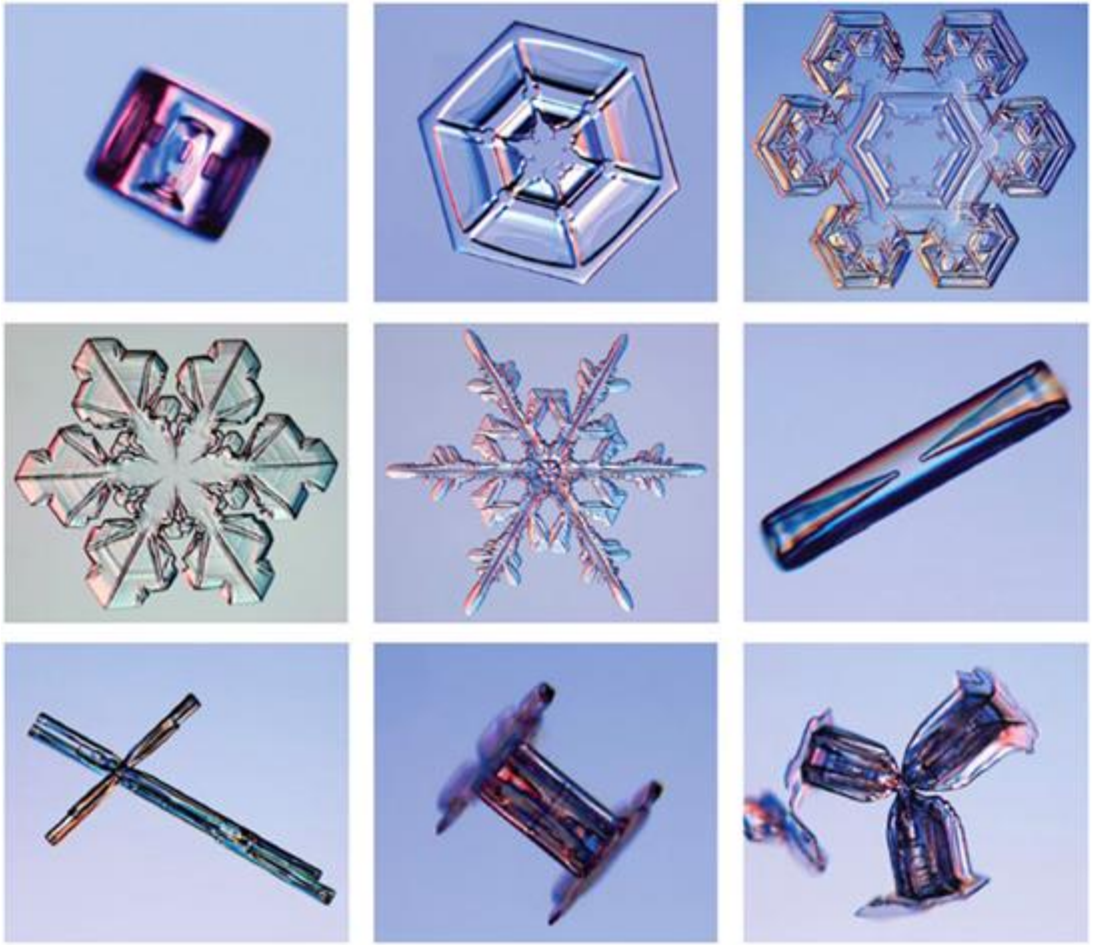
(i.e., don't need Python to Fortran bridge!)



3. In Situ + ML

Constraining ice properties with in situ imagery and ML

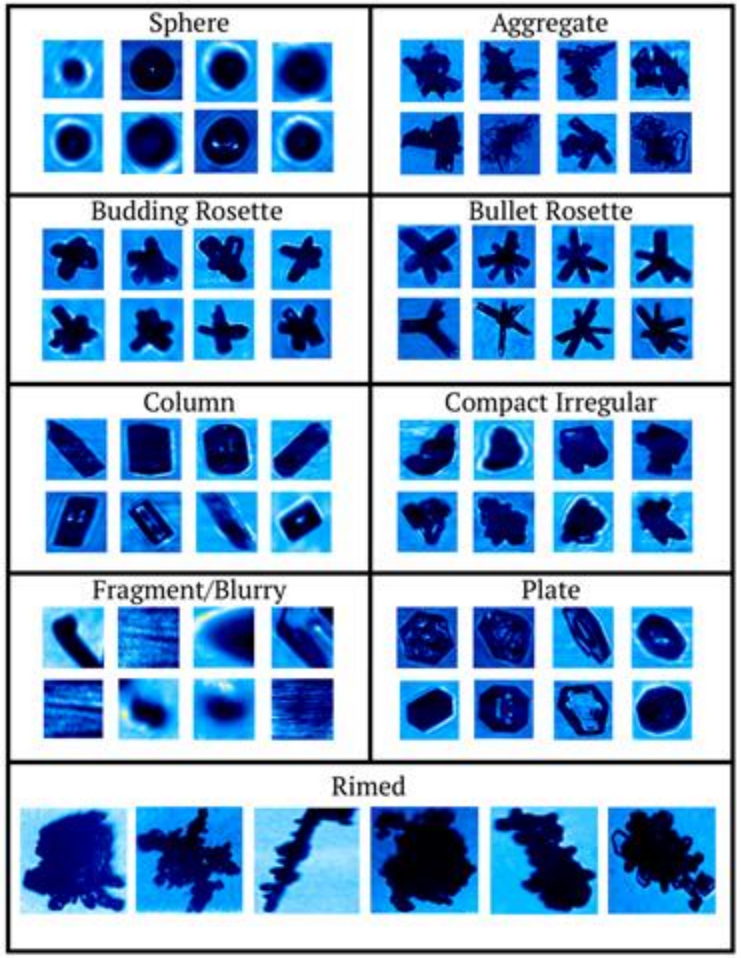
Ice habit (i.e., shape) matters



Source: Kenneth Libbrecht, snowcrystals.com

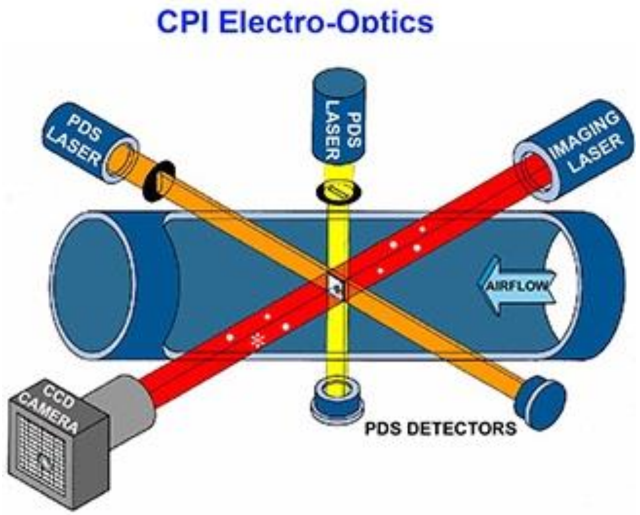
- Habit = Shape
- Habit ~ function of *temperature* and *supersaturation* (i.e., humidity)
- Habit influences:
  - **microphysical process rates**
  - **fall speeds**
  - **optical properties**
- E.g. Ice complexity may induce additional cooling effect of -1.1 W m^{-2} (Jarvinen et al. 2018)
  - For reference: CO_2 forcing is ~2 W m^{-2}

In situ measurements are crucial to understanding habit in real clouds

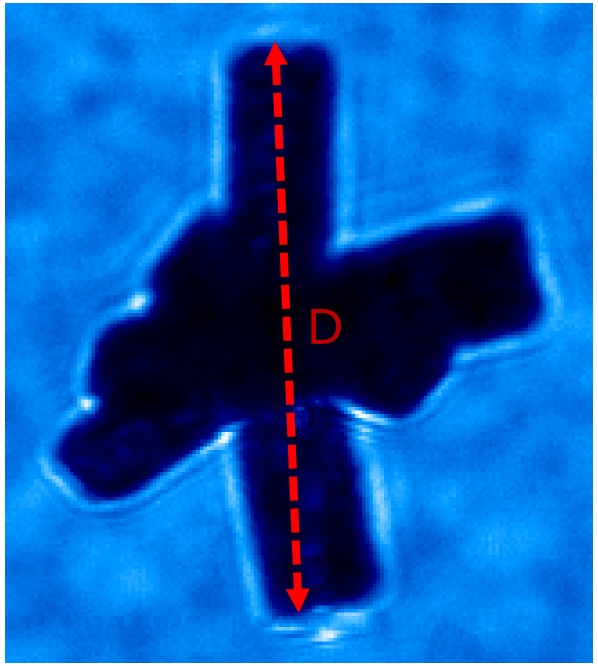


Source: [Przybylo et al. (2022)]

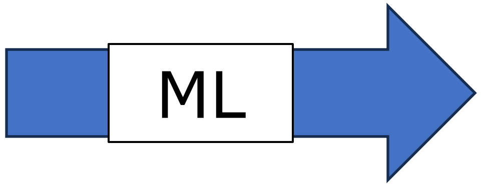
Source: [Xiao et al. 2019]



Mass-Size (m-D) relationships are important for ice microphysics



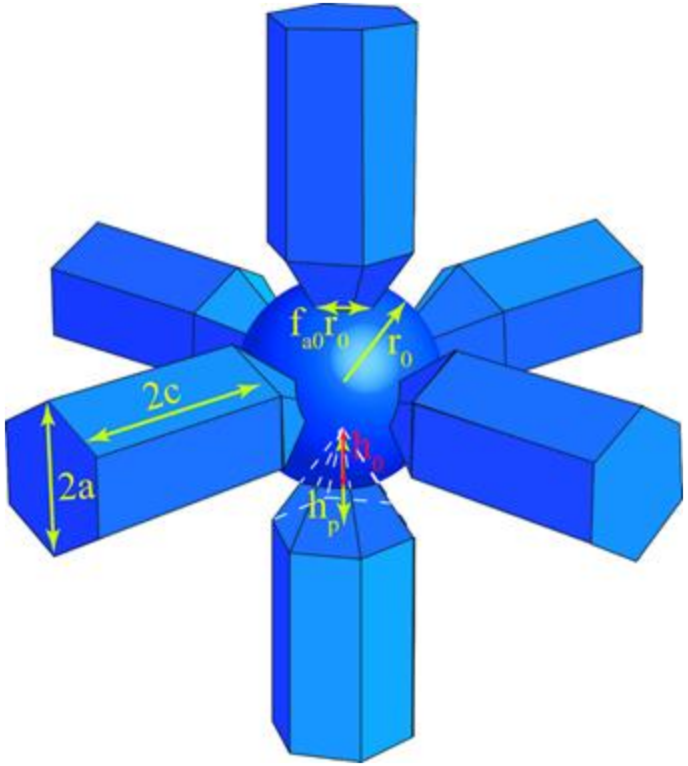
CPI image of
bullet rosette



Mass?
Surface Area?
Etc.

No ground truth → Use synthetic data

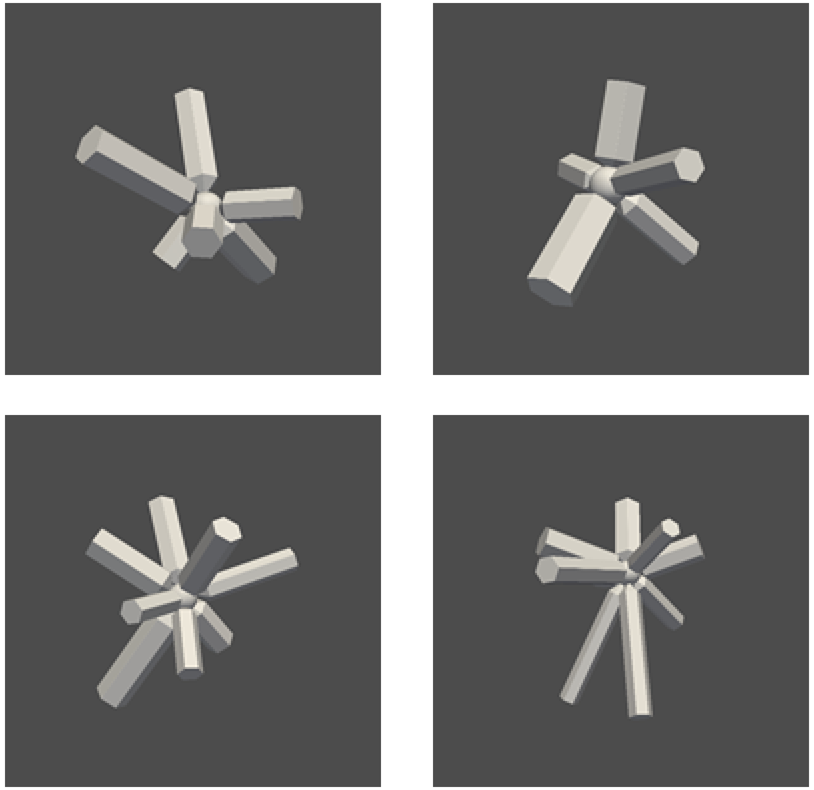
A priori geometric model of
bullet rosette



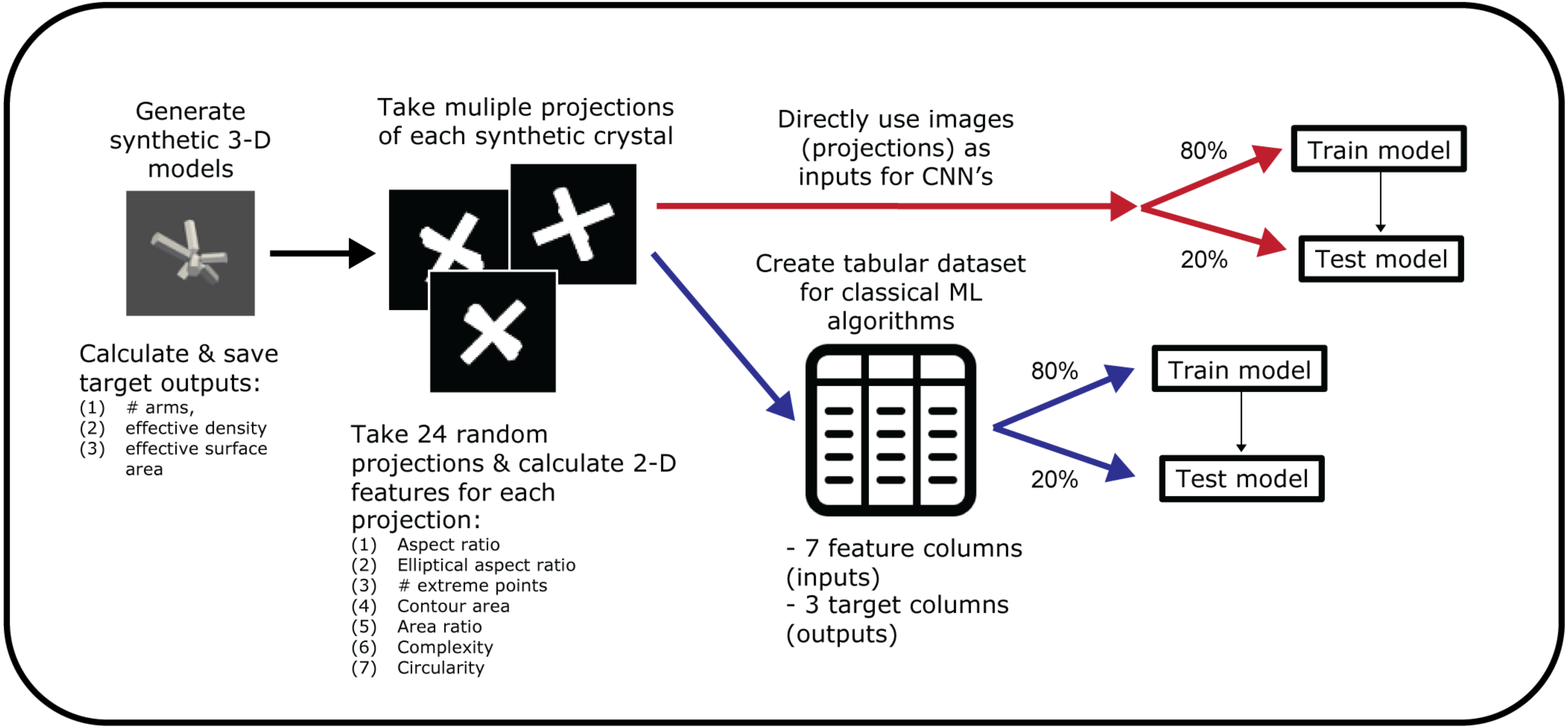
Source: [Pokrifka et al., 2023]

Computationally generate
random variations

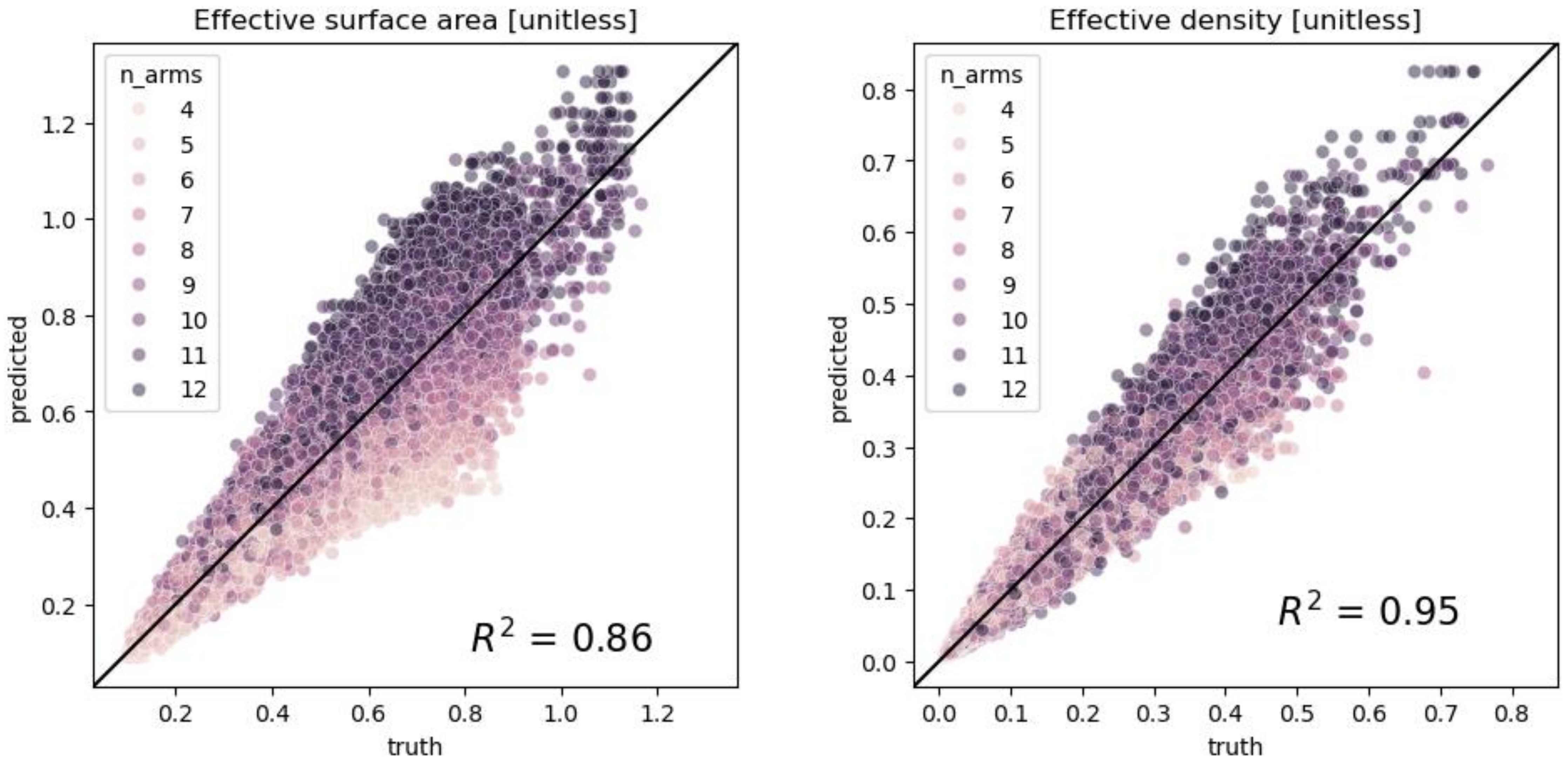
Synthetic 3-D models



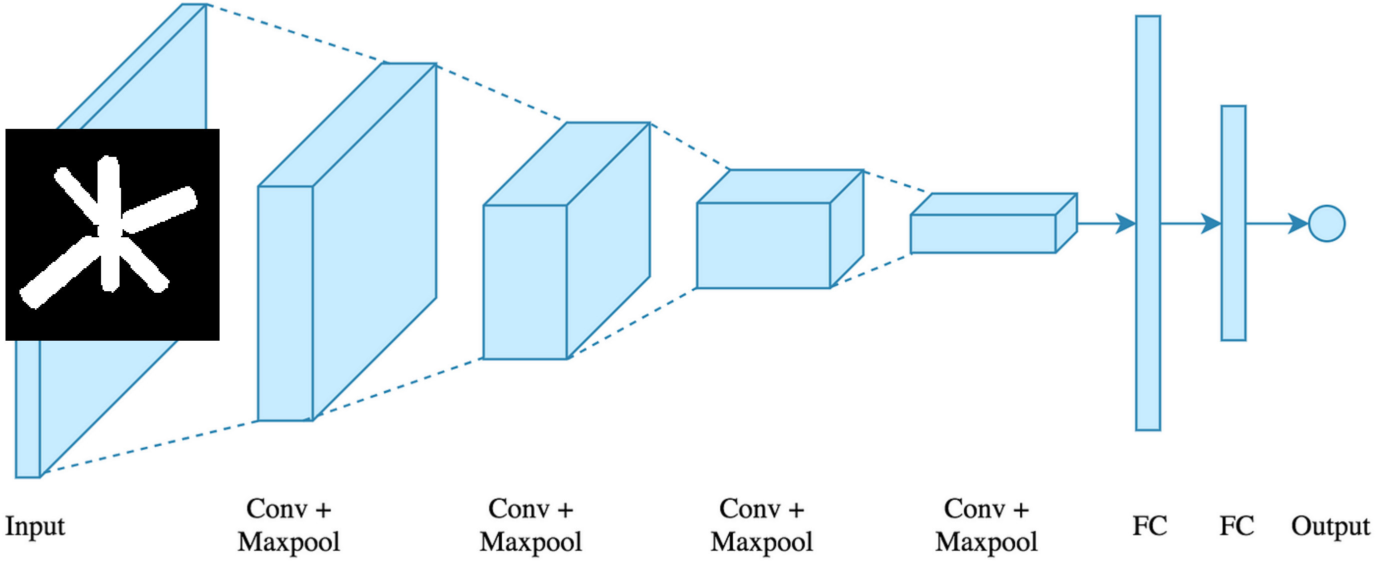
Predicting 3-D properties from 2-D imagery



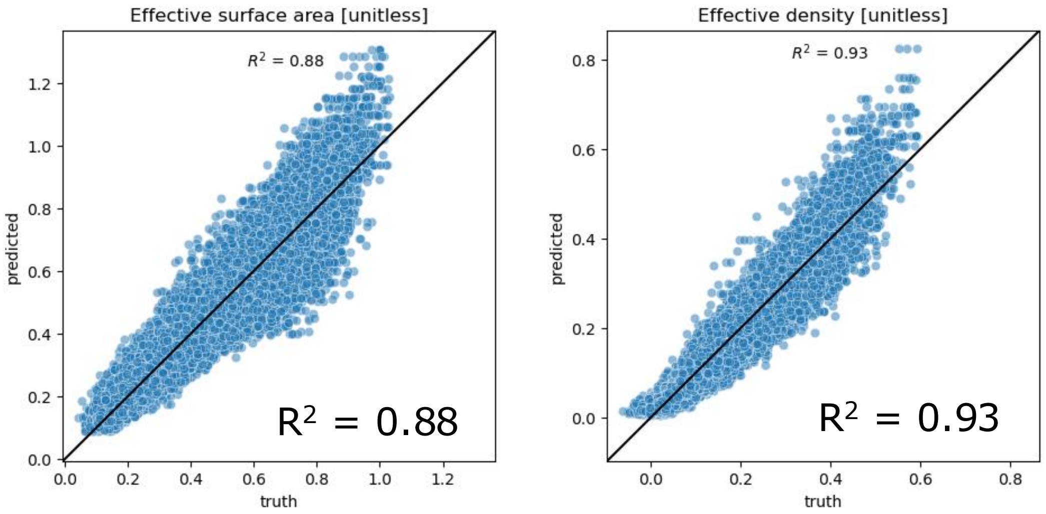
Random forest results: surface area and mass



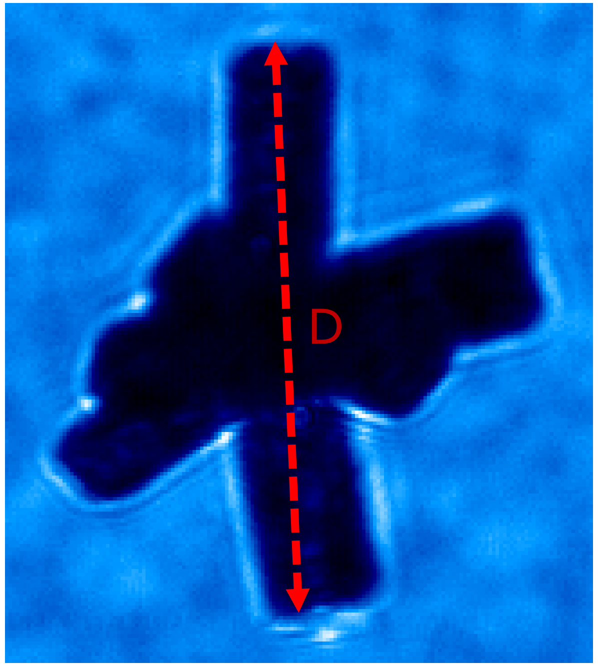
CNN's can be used to circumvent feature engineering



Credit: Arden Dertat



End goal: Revisit m-D relationships using millions of CPI images



CPI images

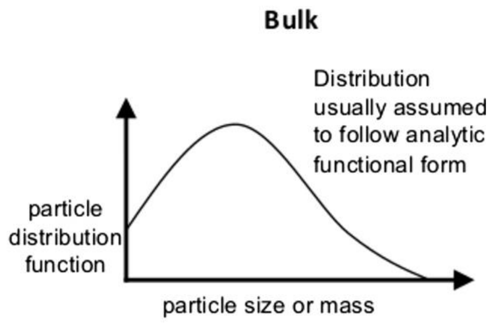
ML → m-D → Compare to existing m-D lookup tables.

Update as necessary.

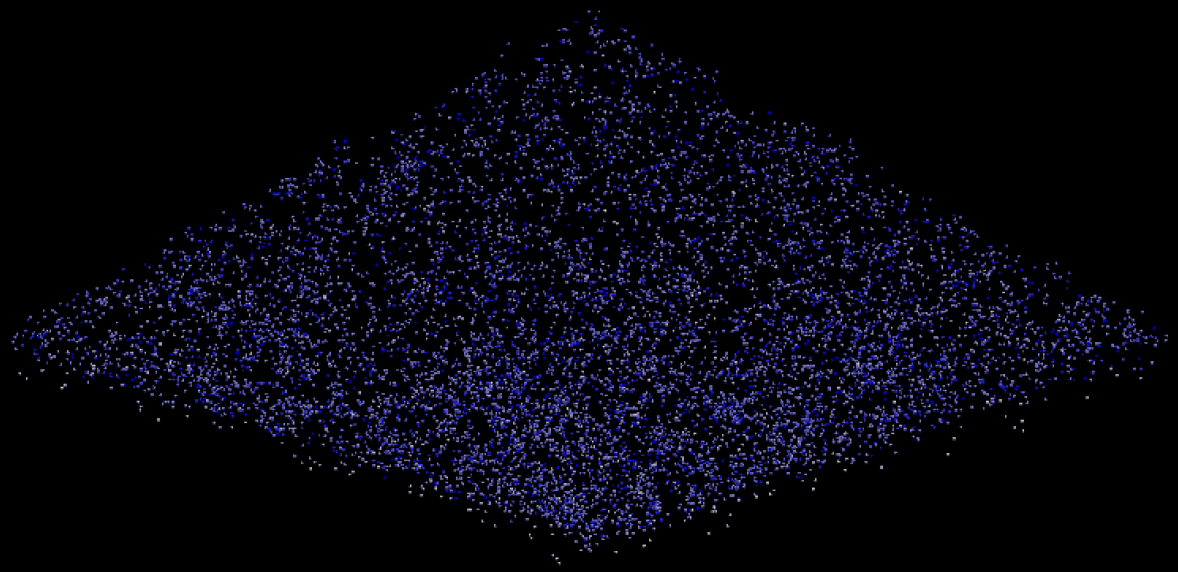
4. LES + Bayesian + ROM

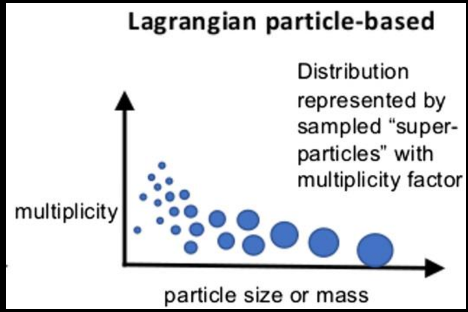
Using high-fidelity models to reduce parametric and structural uncertainties.

Bulk LES



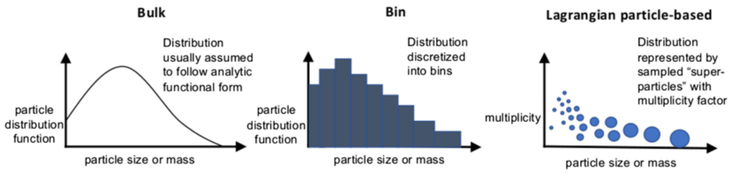
Lagrangian (SDM) LES





Structural vs. *Parametric*

Microphysics framework



Different process parameterizations

e.g.,

$$y = \theta_1 x + \theta_2$$

$$y = \phi_1 x^2 + \phi_2 x + \phi_3$$

Generally, referring to coefficients in equations

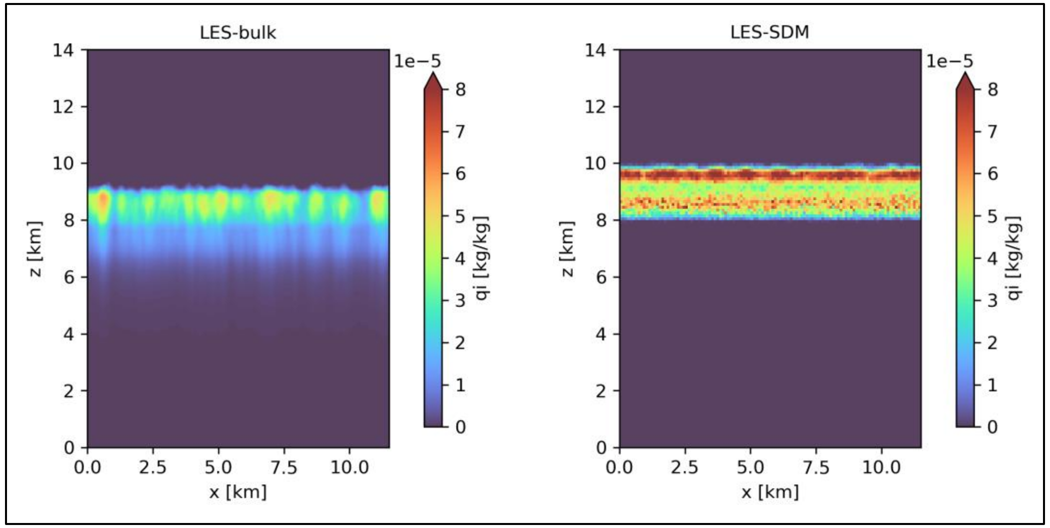
e.g.,

$$y = \boxed{\theta_1} x + \boxed{\theta_2}$$

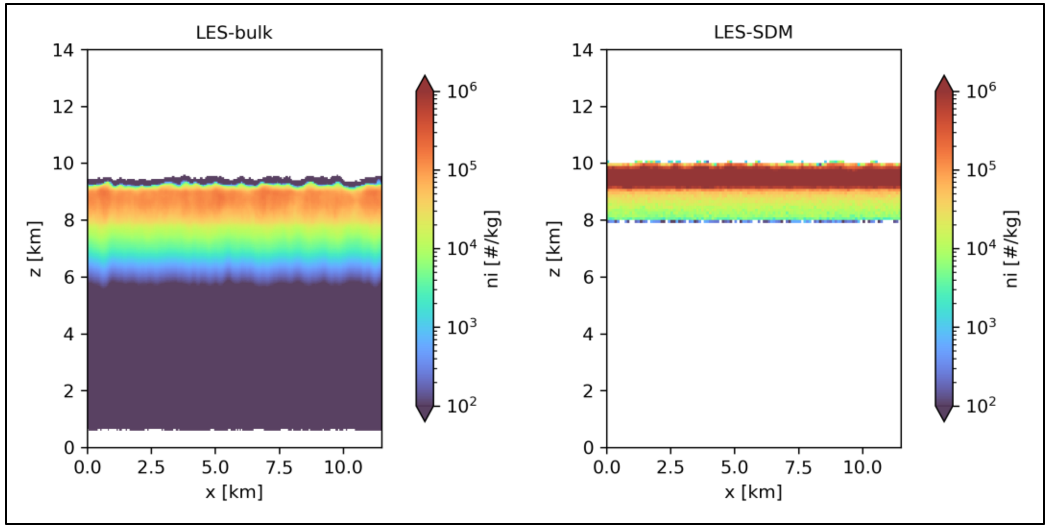
$$y = \boxed{\phi_1} x^2 + \boxed{\phi_2} x + \boxed{\phi_3}$$

Structural uncertainty dominate over parametric uncertainty

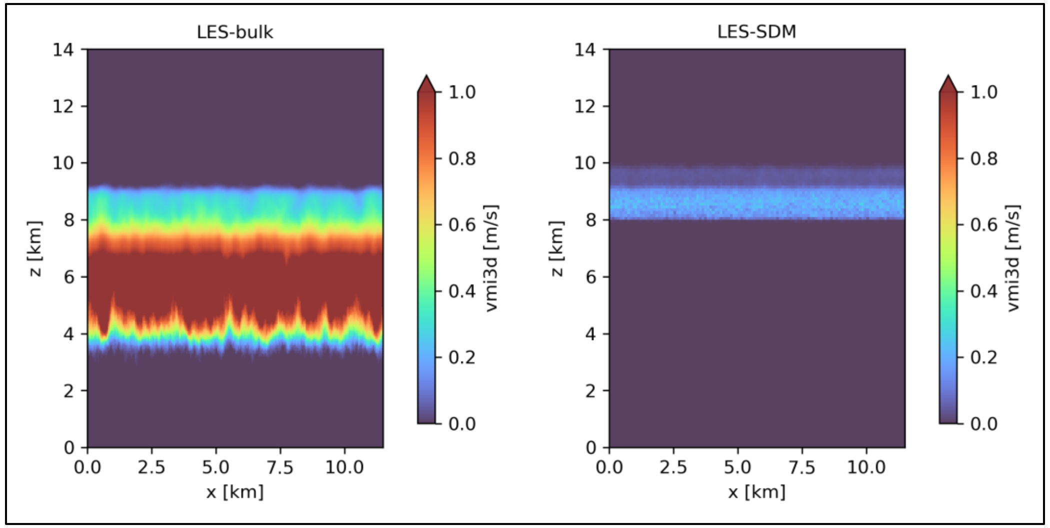
Mass mixing ratio



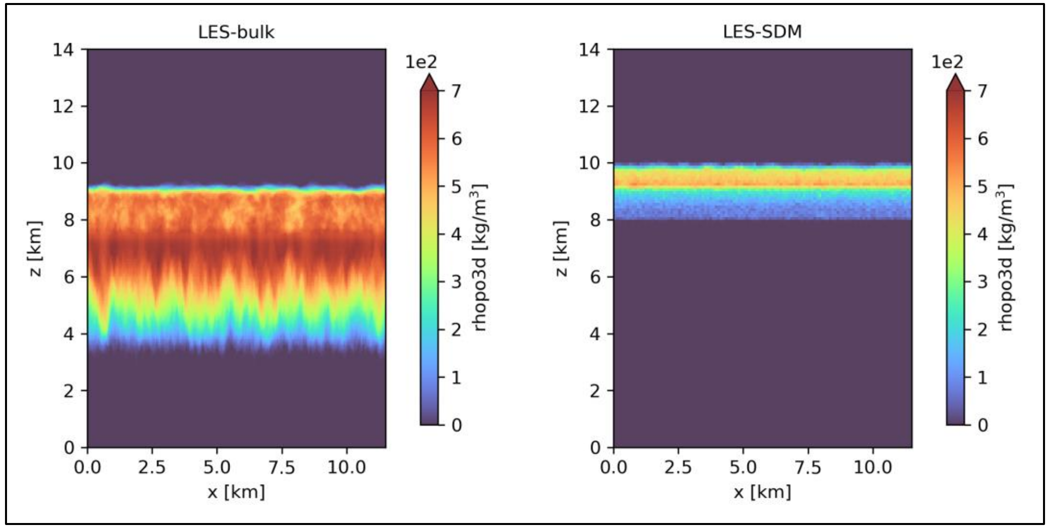
Number mixing ratio



Mass-weighted fall speed



Mass-weighted particle density



End goal: reduce parametric + structural uncertainty and move towards implementation in CAM

Large ensembles of ice cloud LES

1. Optimize parameters within existing (inadequate) structure:
- Approximate Bayesian Computation for estimating posteriors without likelihoods

2. Improve the structure:
- Reduced order modeling with encoders and decoders (e.g., Lamb et al. 2024)

JAMES Journal of Advances in Modeling Earth Systems®

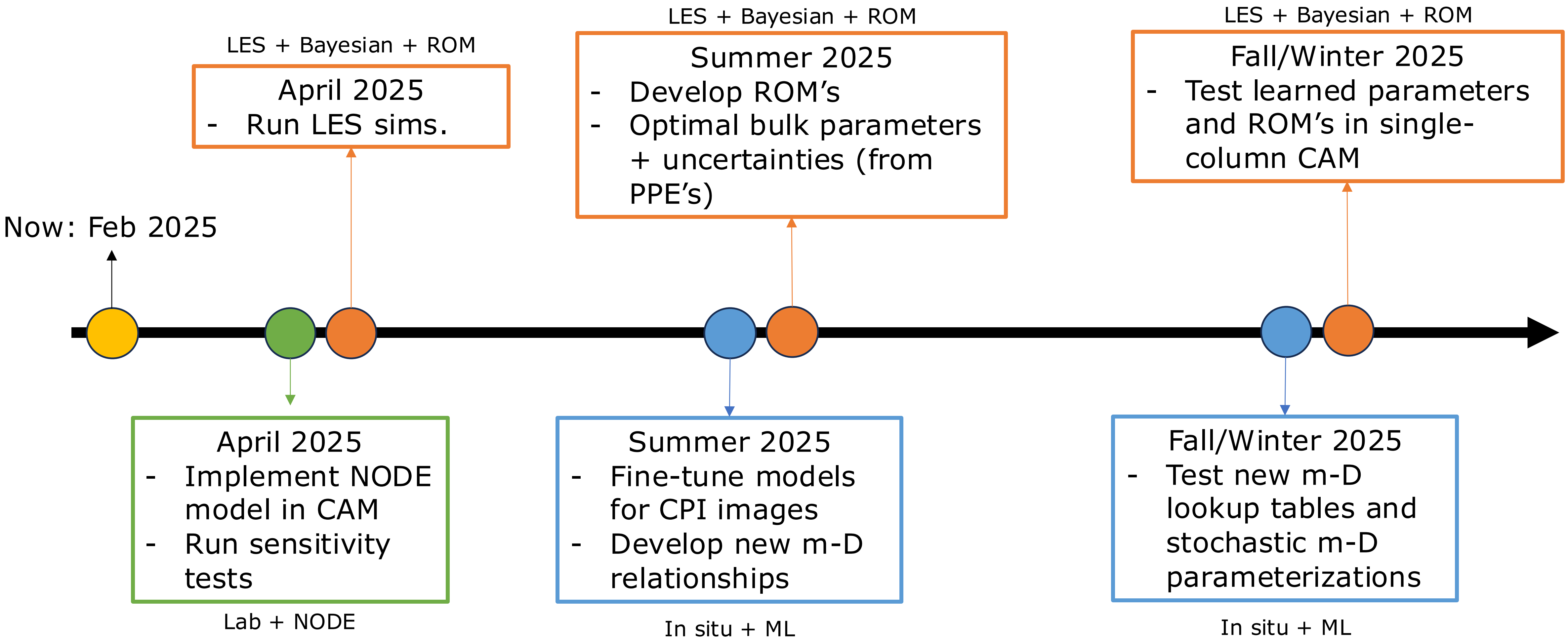
Research Article | Open Access

Reduced-Order Modeling for Linearized Representations of Microphysical Process Rates

K. D. Lamb, M. van Lier-Walqui, S. Santos, H. Morrison

First published: 28 June 2024 | https://doi.org/10.1029/2023MS003918

Timeline of progress & CAM integration



Conclusion

1. NODE’s applied to single-particle lab measurements to develop physics-informed ML models for vapor depositional growth

Conclusion

1. NODE’s applied to single-particle lab measurements to develop physics-informed ML models for vapor depositional growth
2. Framework for predicting 3-D properties from *in situ* imagery was developed → goal to improve m-D relationships

Conclusion

1. NODE’s applied to single-particle lab measurements to develop physics-informed ML models for vapor depositional growth
2. Framework for predicting 3-D properties from *in situ* imagery was developed → goal to improve m-D relationships
3. LES + Bayesian methods → optimal parameters in existing schemes. LES + ROM → Improved structure?

Contact: jk4730@columbia.edu

Thanks!

Collaborators & Help:

- Kara Lamb (Columbia)
- Marcus van Lier-Walqui (Columbia, NASA GISS)
- Jerry Harrington (Penn State)
- Gwenore Pokrifka (Penn State)
- Kamal Chandrakar (NCAR)
- Jasmine Remillard (NASA GISS)
- Hugh Morrison (NCAR)
- Kaitlyn Loftus (Columbia)
- Nathan Magee (TCNJ)
- And more…

NSF LEAP U.S. DEPARTMENT OF ENERGY

Conclusion

Contact: jk4730@columbia.edu

1. NODE's applied to single-particle lab measurements to develop physics-informed ML models for vapor depositional growth
2. Framework for predicting 3-D properties from *in situ* imagery was developed → goal to improve m-D relationships
3. LES + Bayesian methods → optimal parameters in existing schemes. LES + ROM → Improved structure?