Coarse-Graining of Aerosol Mixing State Metrics Empowered by Machine Learning

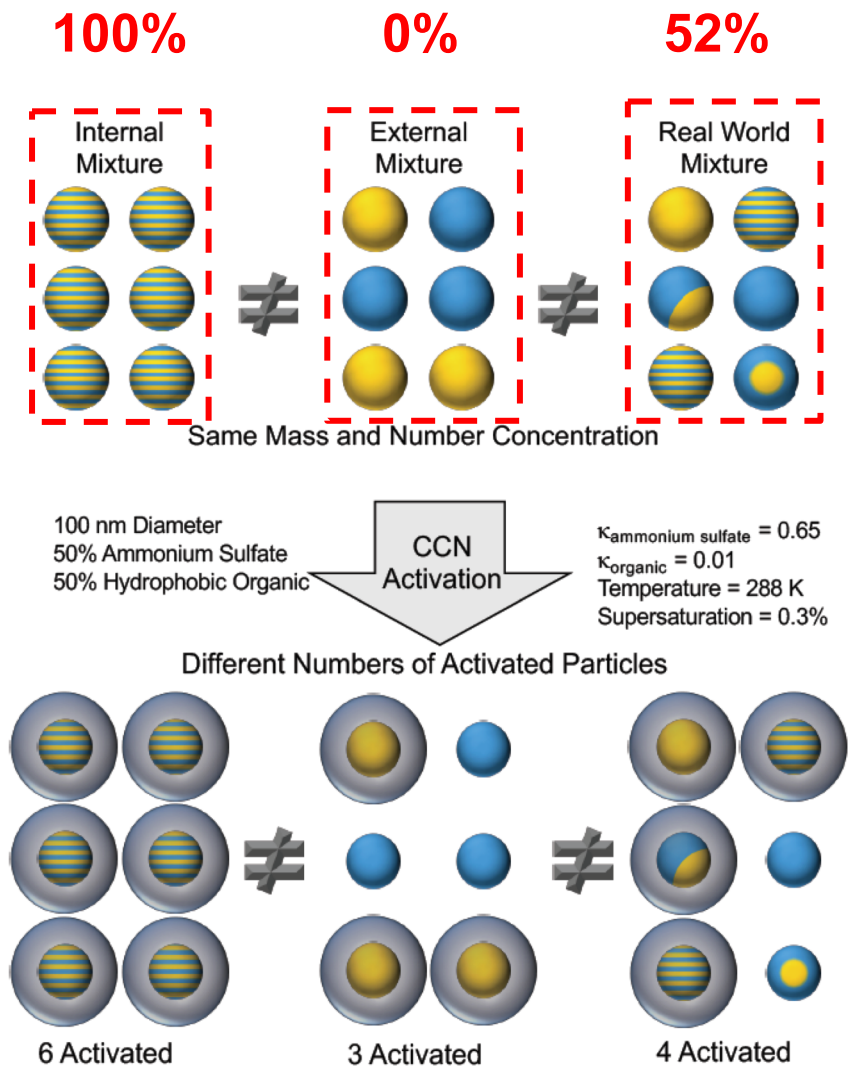
Zhonghua Zheng
Advisor: Prof. Nicole Riemer
12/05/2019

Aerosol Mixing State & Metric

- **Aerosol mixing state**
  - describes the distribution of aerosol chemical species among particles in a population

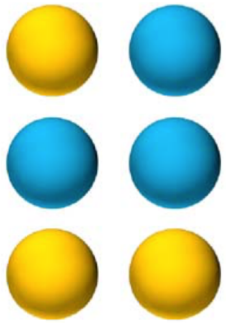
- **Importance of aerosol mixing state**

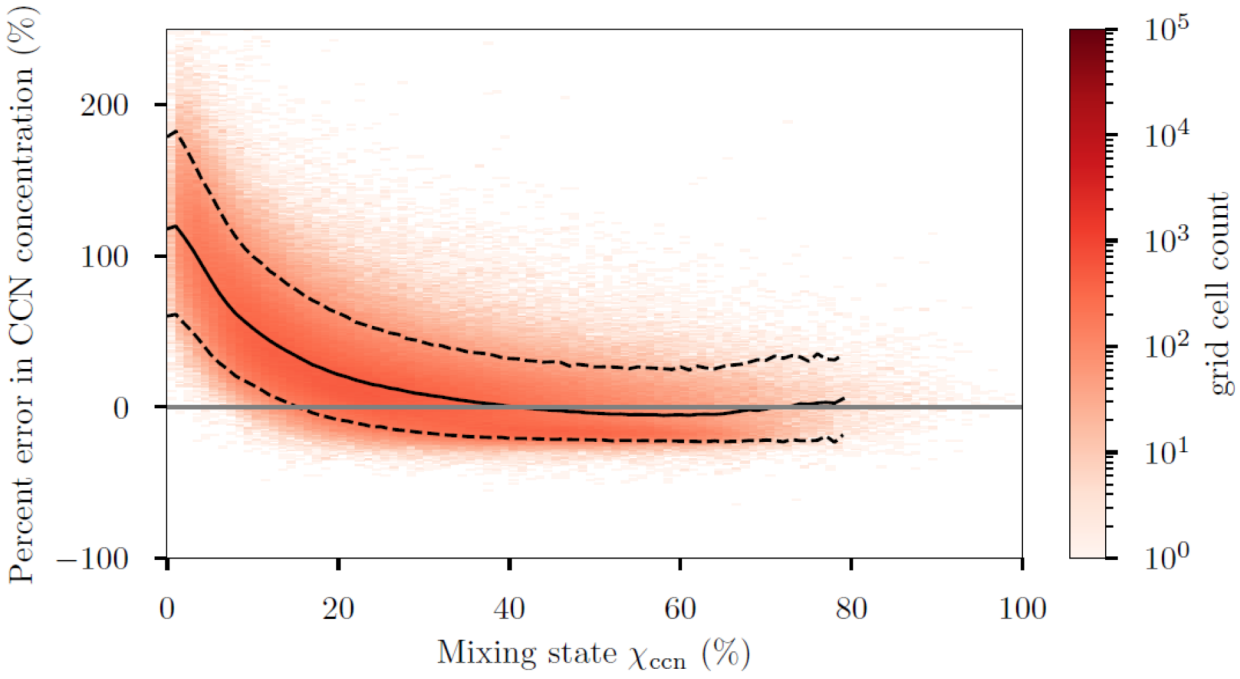
- **Aerosol mixing state index** (Riemer & West, ACP, 2013)
  - Range: [0%, 100%]
  - 100% for internally mixed
  - 0% for externally mixed

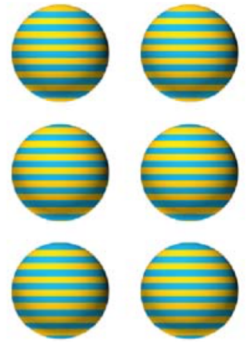


(Riemer et al., 2019, *Reviews of Geophysics*)

Impacts of Aerosol Mixing State Index (Quantitative Analysis)







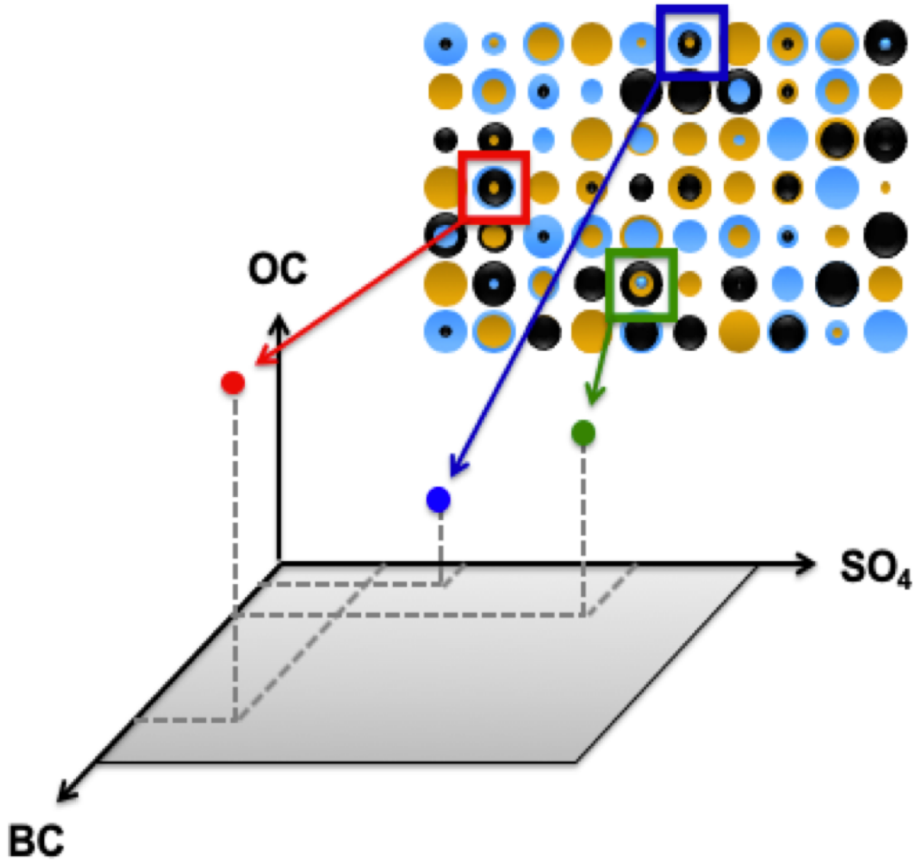
Externally Mixed

Internally Mixed

Error in CCN concentration when assuming average composition

(Curtis, 2019)

How to Get Aerosol Mixing State Index by Modeling? (PartMC as a benchmark)

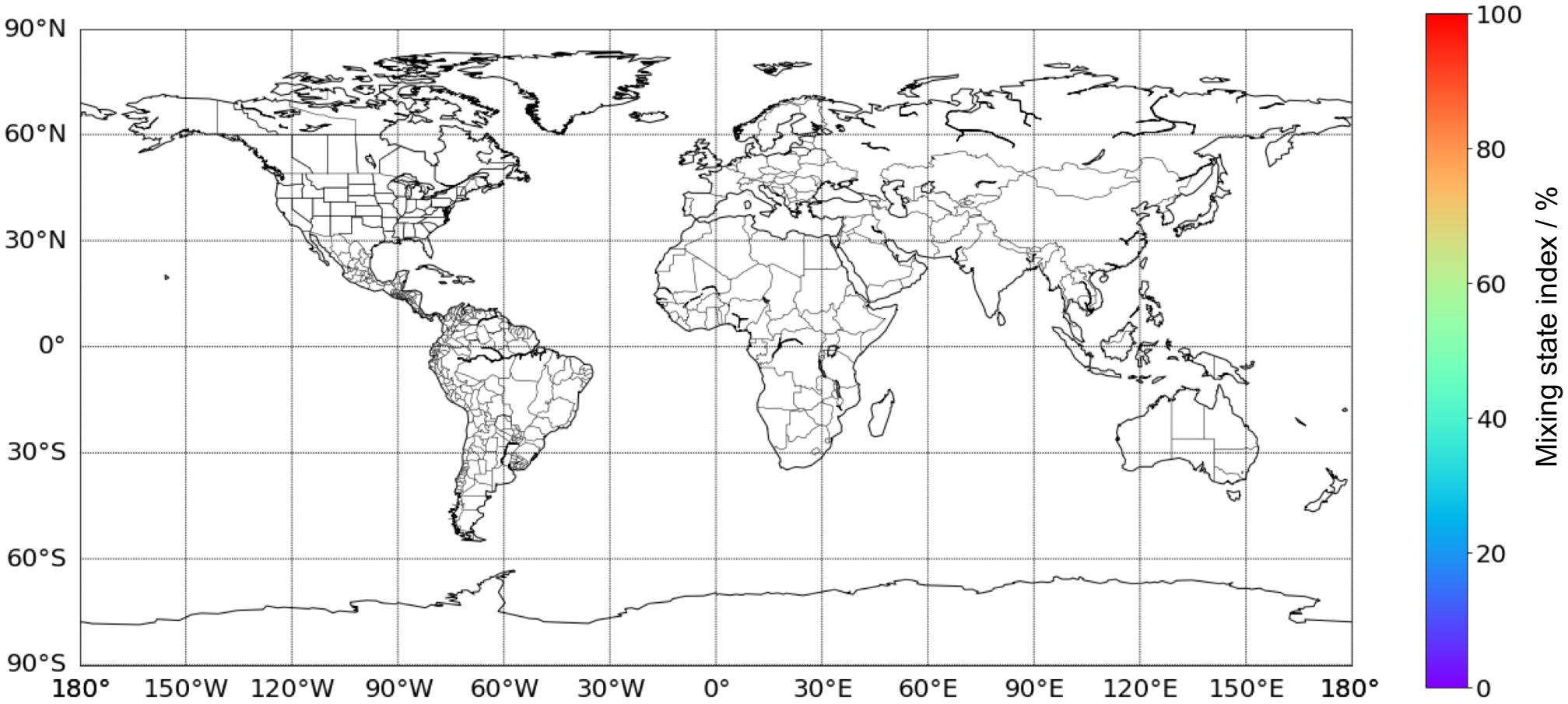


- Particle-Resolved Model
  - CBM-Z (Gas-Phase)
  - MOSAIC (Aerosol-Phase)

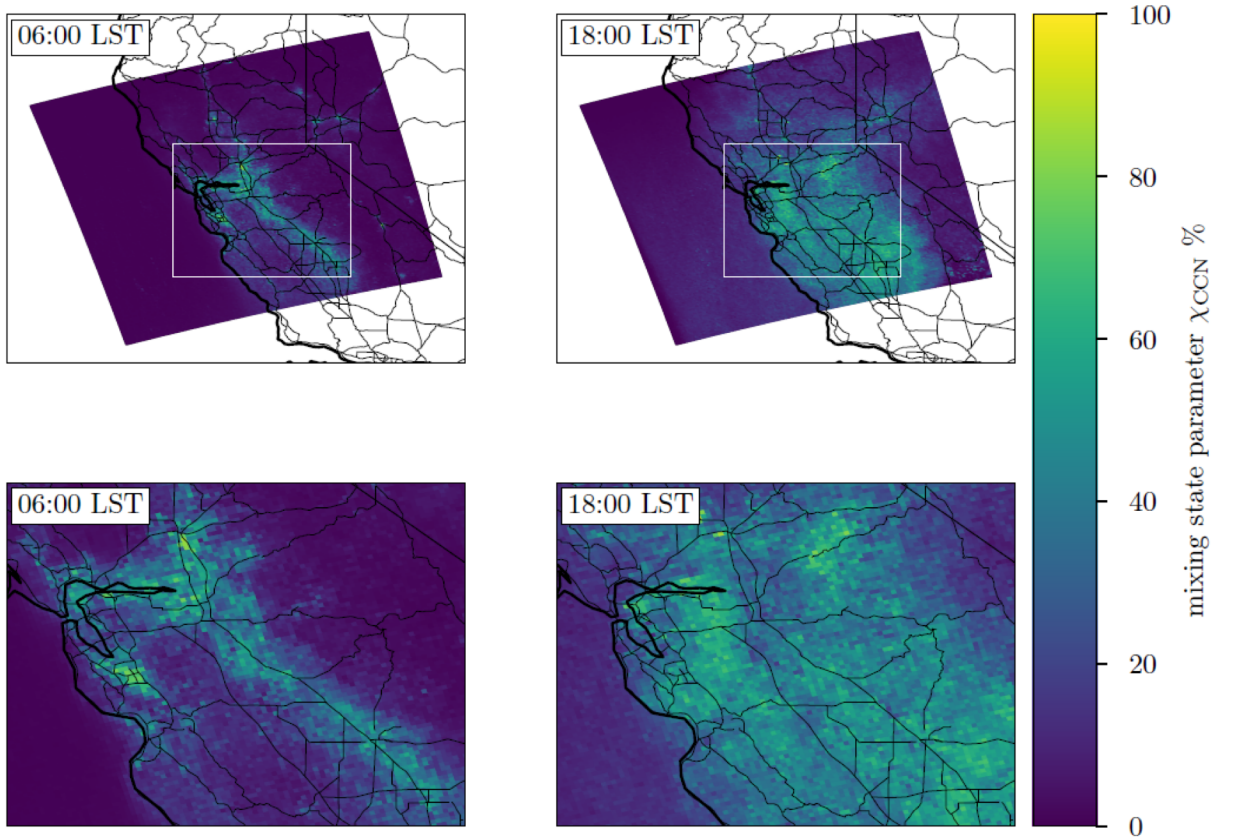
	Particle 1	Particle 2	Particle 3
BC	10	1	3
SO_4	2	4	8
OC	9	2	5

(Riemer et al., JGR, 2009, and MOSAIC: Zaveri et al., 2008)

Question: How to Scale up the Mixing State Index at Global Scale?



Solution 1: Numerical Simulation -- WRF-PartMC



Regional Benchmark Simulation, WRF-PartMC

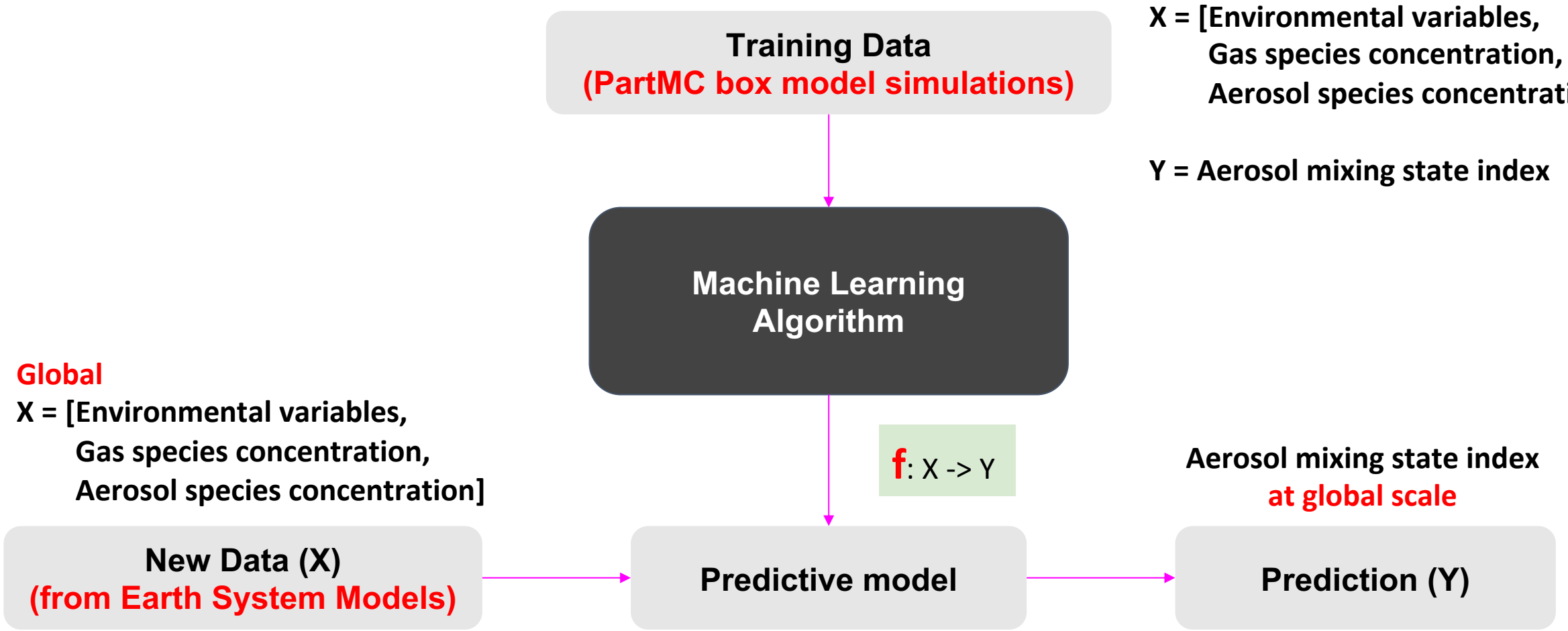
computationally expensive!

(Curtis, 2019)

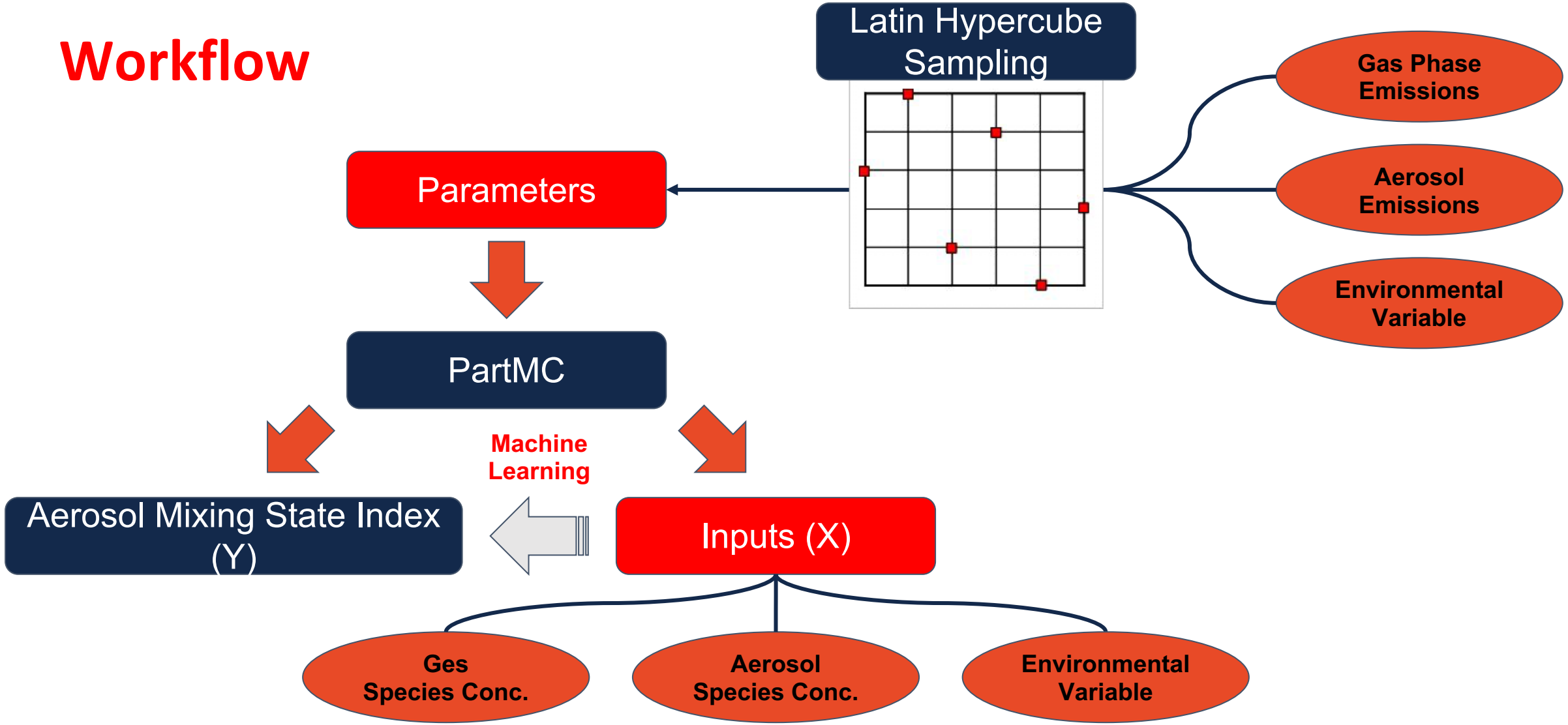
- 170 x 160 grid cells
- spatial resolution: 4 km x 4 km
- 2-days simulation
- temporal resolution: 10~20 s
- full particle store every 2 hrs
- 10,000 particles each population

- 10,000 node hours
- 320,000 core hours

Solution 2: Machine Learning enabled coarse-grained modeling



Workflow


Latin Hypercube Sampling
Gas Phase Emissions
Aerosol Emissions
Environmental Variable
Parameters
PartMC
Machine Learning
Aerosol Mixing State Index (Y)
Inputs (X)
Ges Species Conc.
Aerosol Species Conc.
Environmental Variable

PartMC simulations as training data and testing data

Parameters
Environmental Variable
Relative humidity (RH)
Latitude
Day of Year
Temperature
Gas Phase Emissions Scaling Factor
SO_2, NO_2, NO, NH_3, CO, ALD2
HCHO, ETH, OLEI, OLET, TOL, XYL
AONE, PAR, ISOP, CH_3OH, ANOL, DMS
Carbonaceous Aerosol Emissions (one mode)
D_g
σ_g
BC/OC mass ratio
E_a
Sea Salt Emissions (two modes) (1)
$D_{g,1}$
$\sigma_{g,1}$
$E_{a,1}$
$D_{g,2}$
$\sigma_{g,2}$
$E_{a,2}$
OC fraction
Dust Emissions (two modes) (1)
$D_{g,1}$
$\sigma_{g,1}$
$E_{a,1}$
$D_{g,2}$
$\sigma_{g,2}$
$E_{a,2}$
Restart Timestamp (3)
Timestamp

More details

- One third of scenarios without sea salt
- One third of scenarios without dust
- "Restart": use simulation to create simulation
- Overall training samples (800+400+400+200)*25=45,000
- Additional 120*25=3,200 samples for testing the robustness of the ML models

Simulation

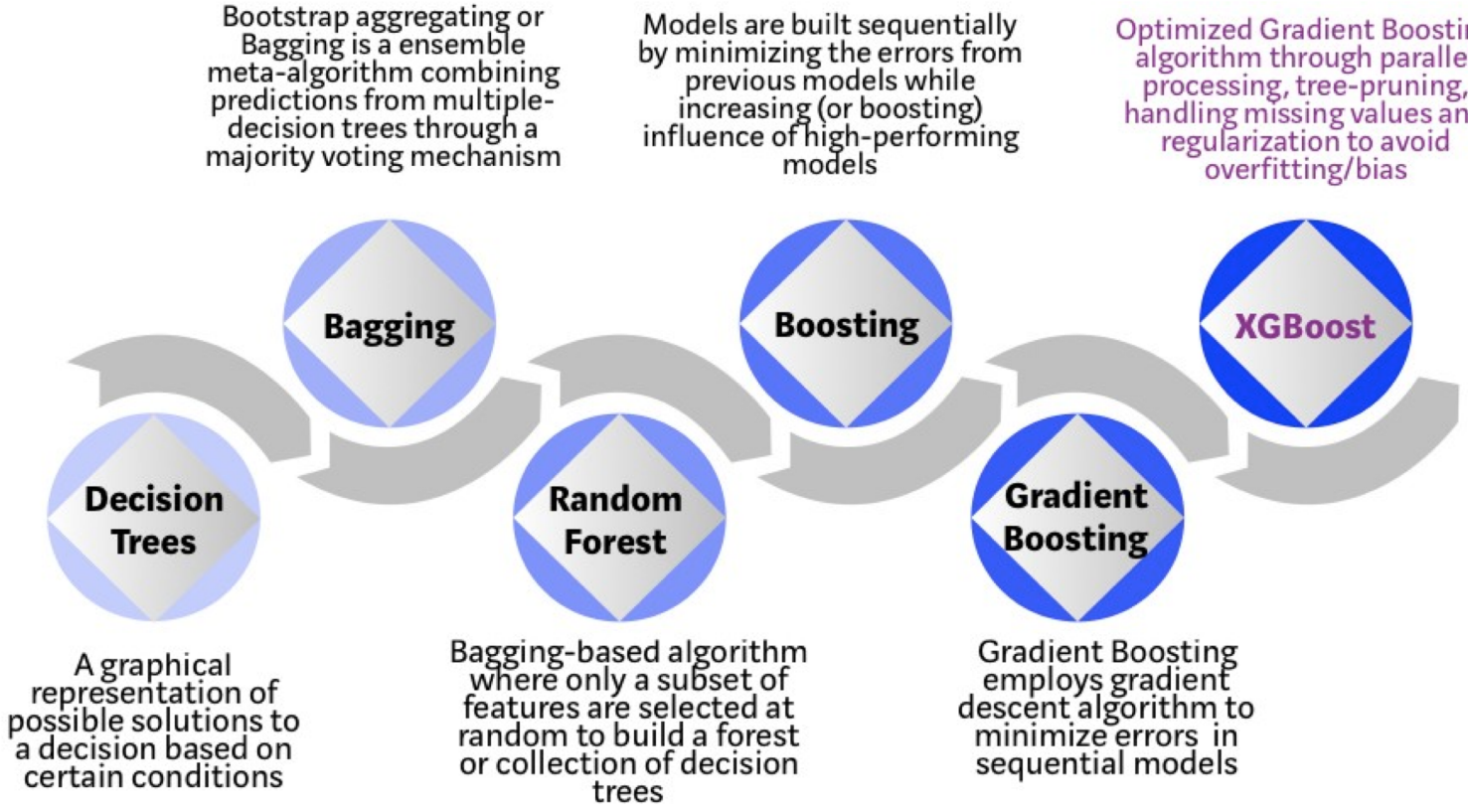
- Parallelized and scalable

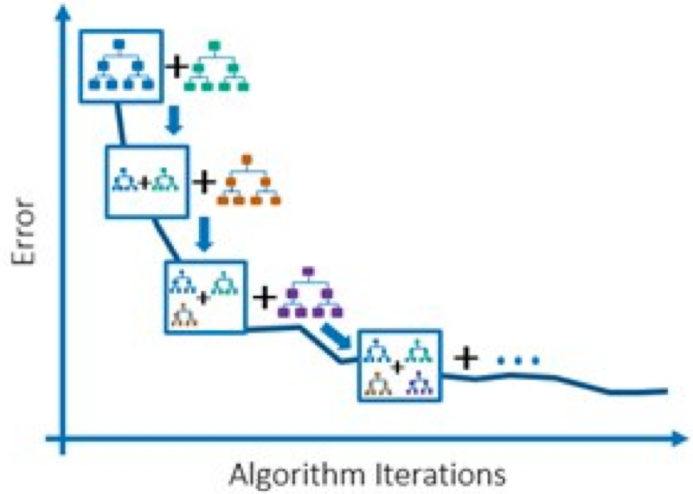
Objective

- Mixing State Index of submicron aerosols (d < 1um)

(Hughes et al., 2018)

Machine Learning Algorithm: XGBoost



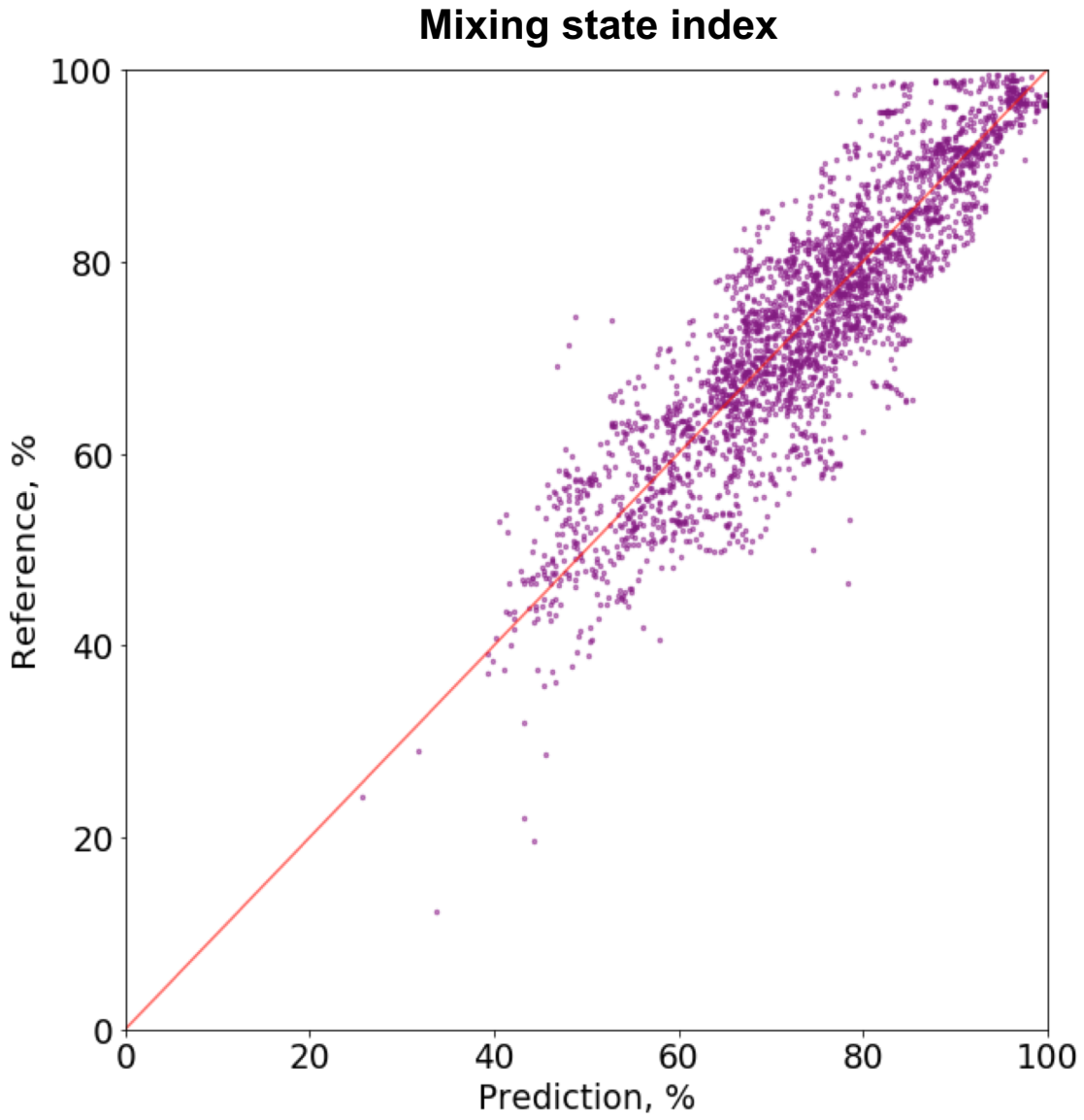


https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d
https://www.kdnuggets.com/2017/10/understanding-machine-learning-algorithms.html

Model evaluation

Our ML emulator are robust for the testing dataset!

Evaluation metrics	Value
RMSE	0.062
Mean Absolute Error	0.048
Median Absolute Error	0.037
Index of Agreement	0.943
Correlation Coefficient	0.893
R2	0.795

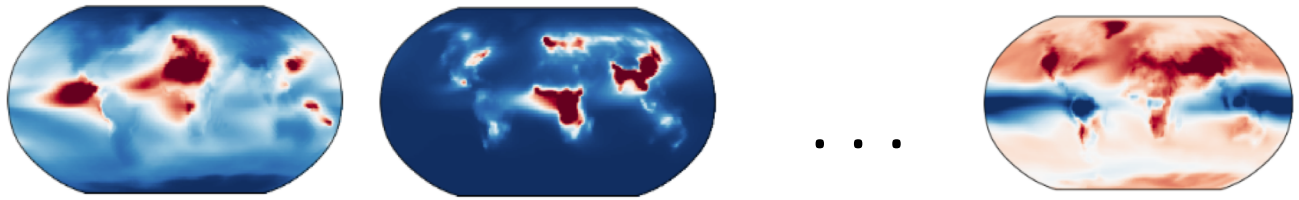


Apply ML emulator

ML emulator

+

X
(outputs from CESM FHIST compset)

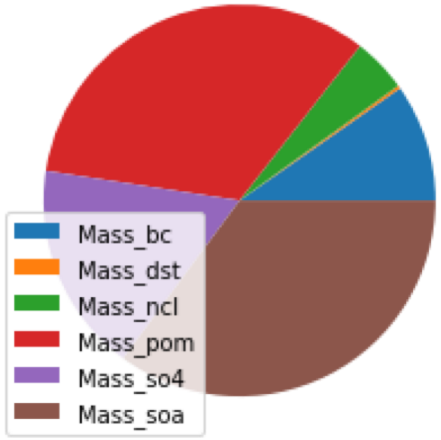


. . .

=

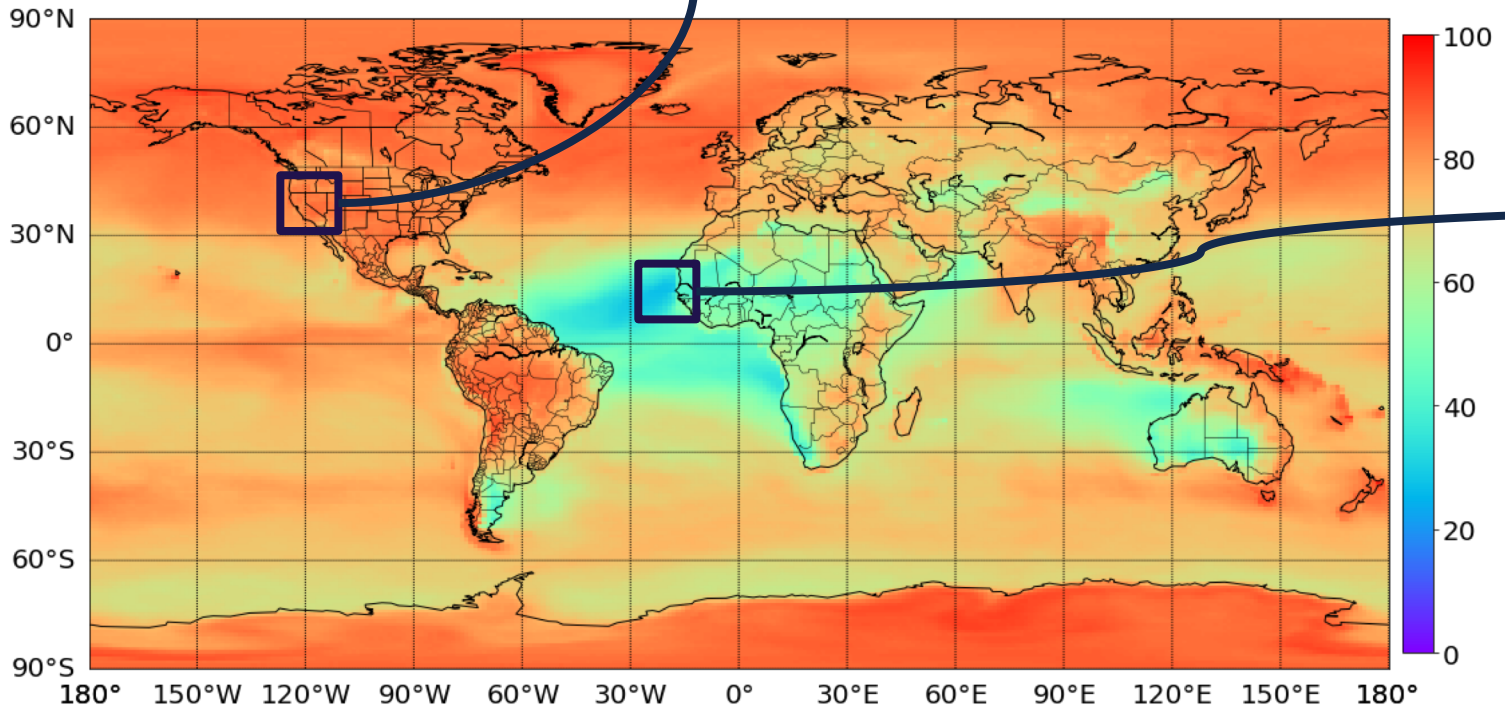
?

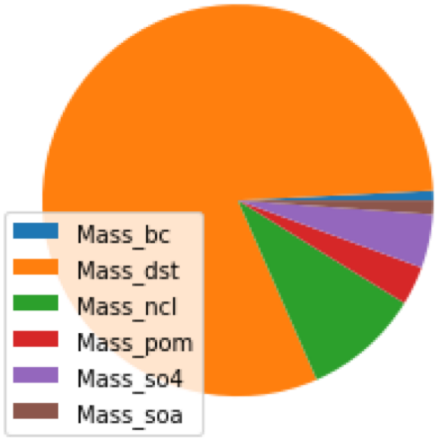
Mixing State Index (2011 DJF)



χ = 81%

Davis, CA (38.54° N, 121.76° W)



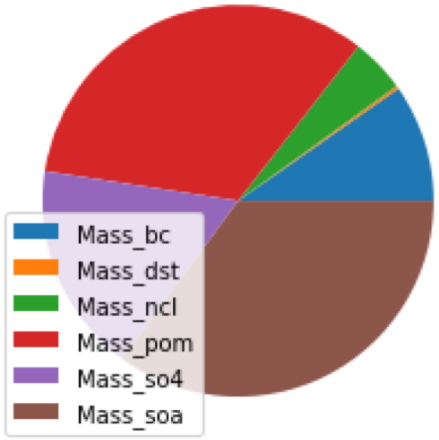


χ = 27%

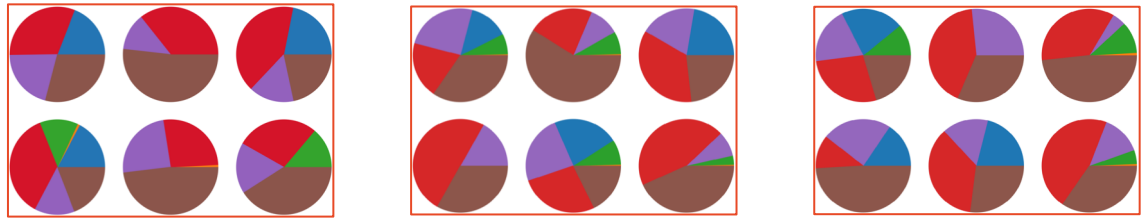
Close to Dakar (15.55° N, 21.2° W)

Infer the possible aerosol populations given the mixing state index

Davis, CA
(38.54° N, 121.76° W)

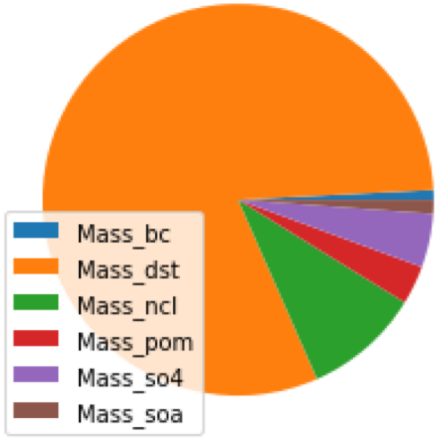


χ = 81%

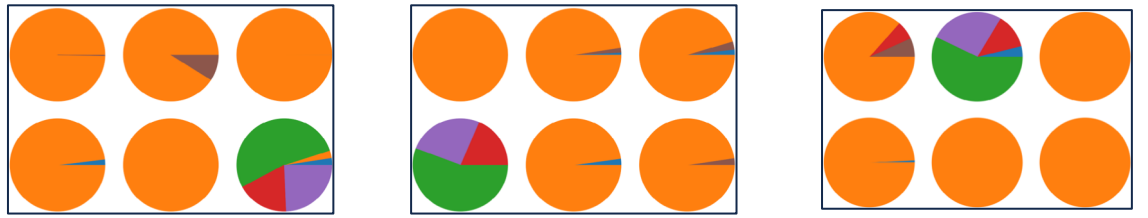


Possible population | Possible population | Possible population

Close to Dakar
(15.55° N, 21.2° W)



χ = 27%



Summary

1. **Designed** a pipeline to conduct 1800 box-model simulations according to the latin hypercube sampling approach

2. **Developed** Machine Learning-enabled emulator to estimate aerosol mixing state index at a global scale for CESM

3. Tools for verifying the aerosol assumptions at global scale

4. This Machine Learning-enabled workflow can be **leveraged** to other aerosol research