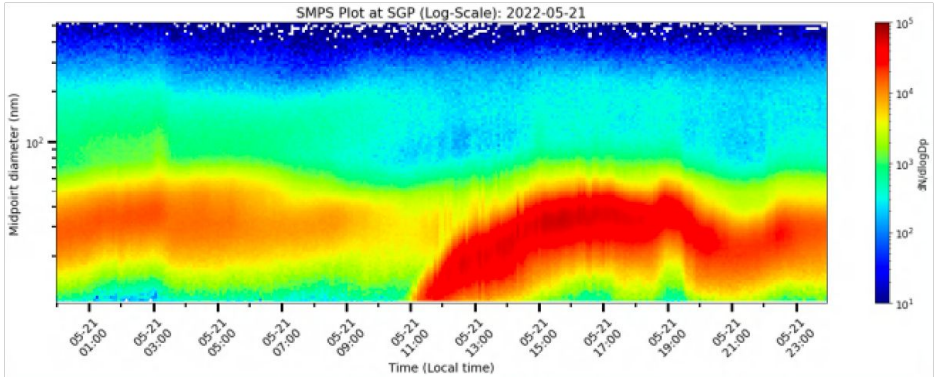
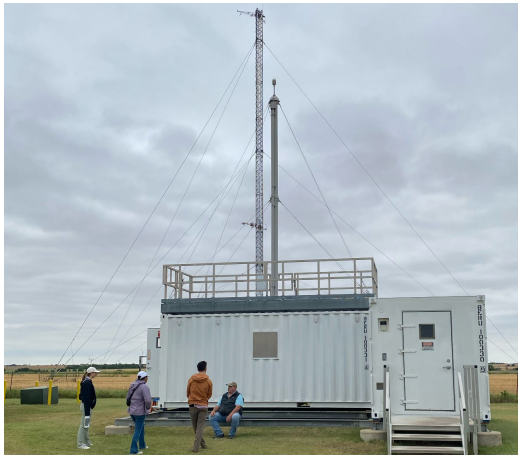
# **A**erosol Event **I**dentification using **M**achine Learning at the **S**outhern **G**reat **P**lains Site (AIM-SGP)

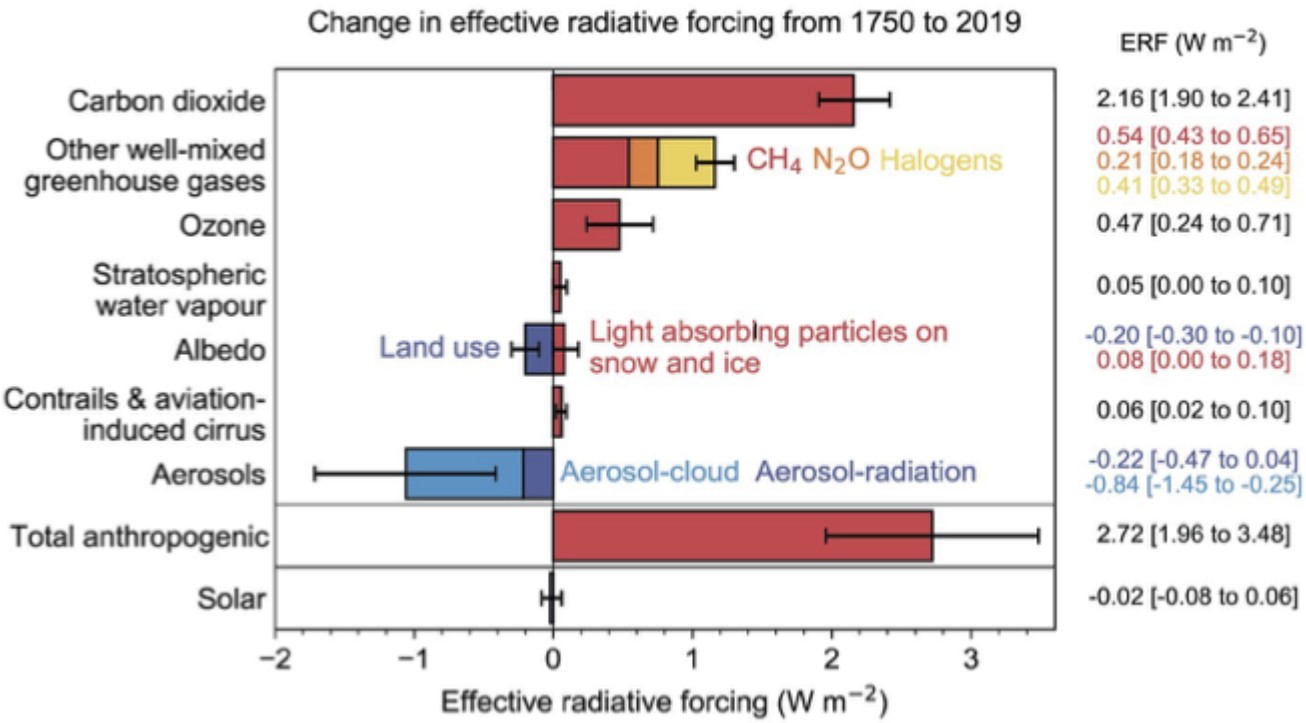
Team Members: Ashlynne Gary, Gerardo Carrillo-Cardenas, Rory Zhang
Mentors: Connor Flynn



# Introduction - New Particle Formation

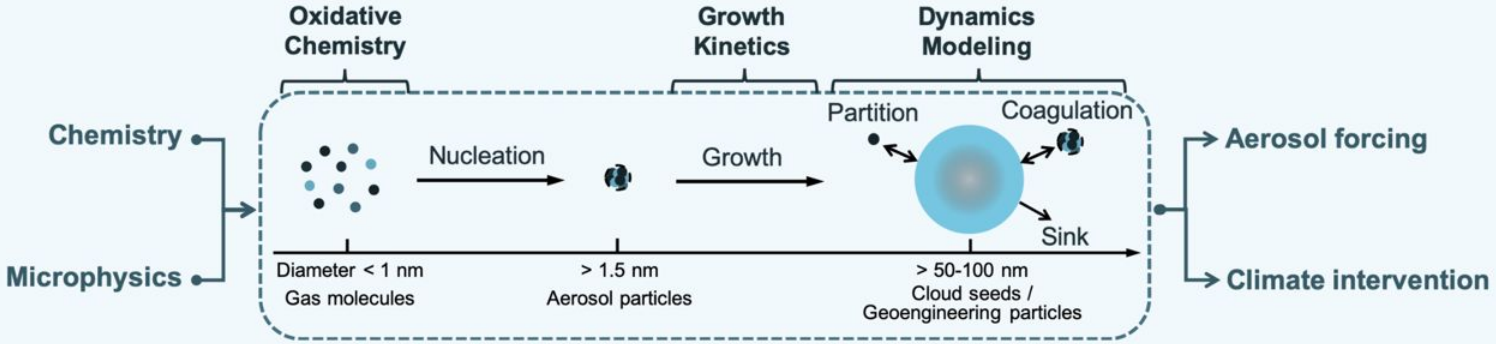
# Aerosols: Why do we care?

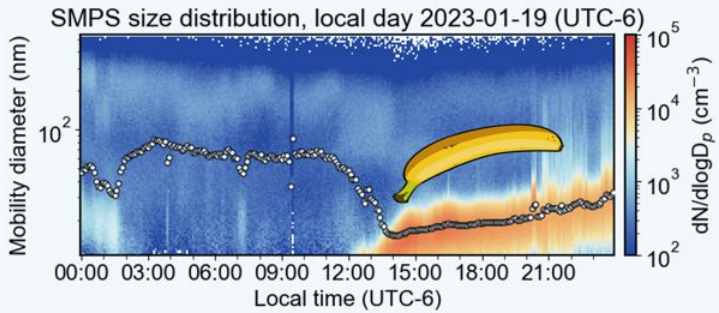



ARM

ipcc
INTERGOVERNMENTAL PANEL ON climate change

# Introduction




- **Definition:** gas-phase molecules undergo chemical and microphysical transformations to form new aerosol particles through nucleation and subsequent growth.
- **Signature:** "Banana curve" is commonly recognized as a signature of a new particle formation event.



Wang, M.; et.al . Nature. 2020. Curtius, J.: Nucleation of atmospheric particles, EPJ Web of Conferences, 2009.

# New Particle Formation Classification: AI & Machine Learning

Image Based Classification

- Convoluted Neural Networks (Su et al, 2022)

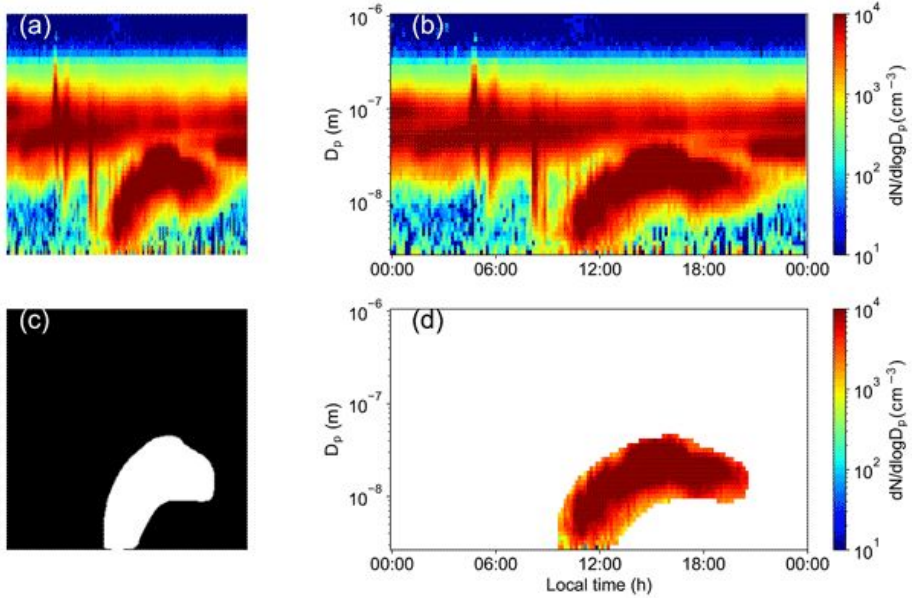



Image credit: Su et al, 2022

# New Particle Formation Classification: AI & Machine Learning

Image Based Classification

- Convoluted Neural Networks (Su et al, 2022)

Feature Based Classification

- Random Forest Model (Hao et al, 2026)

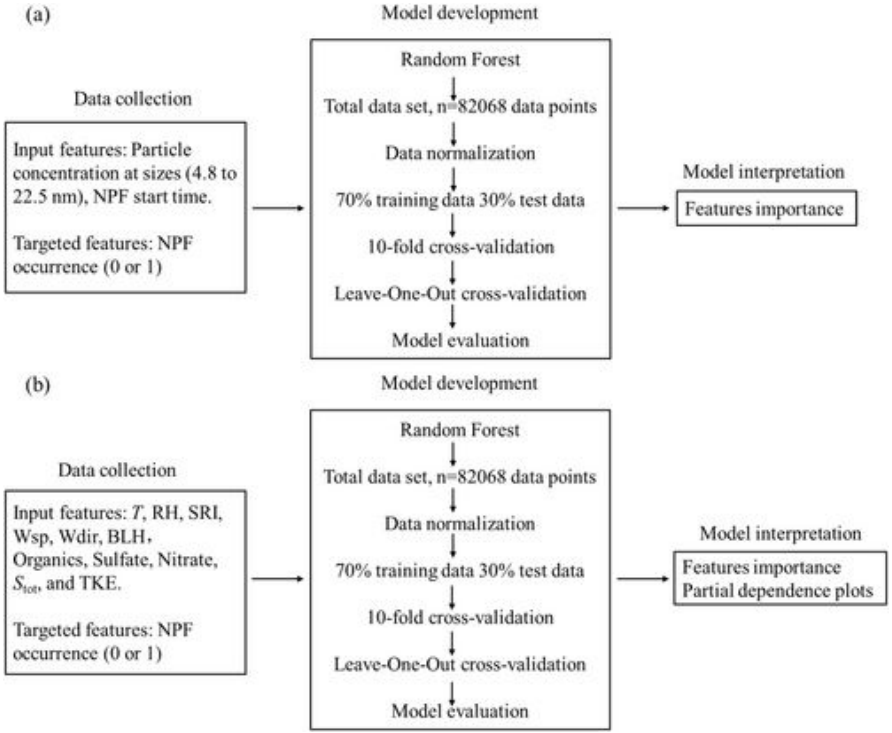



Image credit: Hao et. al., 2026

# Methodology - Instrumentation and ML Framework

Methodology

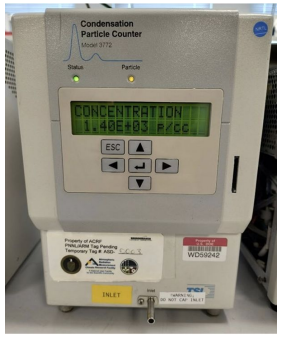
# Instrumentation and Datastreams

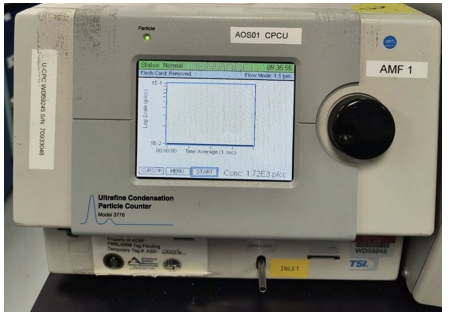
**Location**: Southern Great Plains Site, Lamont, OK

**Time Range** : January 1st 2023 - December 31st 2023

**Instruments**:

- Scanning Mobility Particle Sizer (SMPS)
  - Aerosol size distribution (10 - 500 nm)
  - sgpaossmpsE13.b1
- NanoSMPS
  - Aerosol size distribution (2 - 150 nm)
  - sgpaosnanosmpsE13.b1
- Condensation Particle Counter (CPCf)
  - Total Aerosol Concentration (10 - 3000 nm)
  - sgpaoscpcfE13.b1
- Ultrafine Condensation Particle Counter (CPCuf)
  - Total Aerosol Concentration (3 - 3000 nm)
  - sgpaoscpcufE13.b1




ARM

Methodology

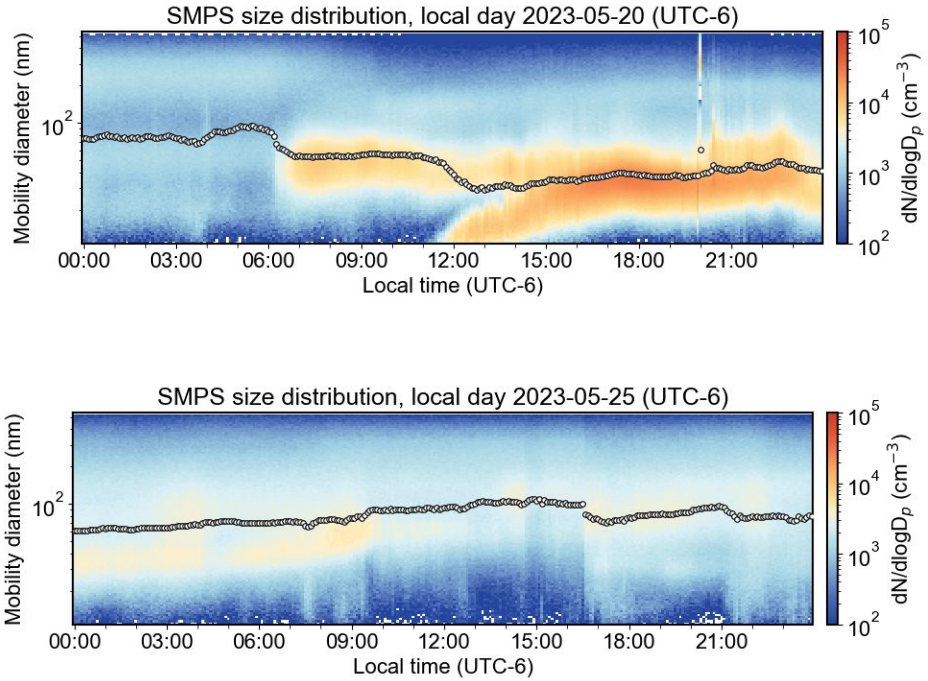
# Visual Classification of New Particle Formation Events

Visual Classification for NPF events consists of the following characteristics (Dal Maso et al., 2005) :

- A new mode of particles must appear in the size distribution.
- The mode must start in the nucleation mode size range
- The mode must persist over a time span of hours and show signs of growth

Usually classified into the following categories:

- Class I
- Class II
- Undefined
- Non-Event



Methodology

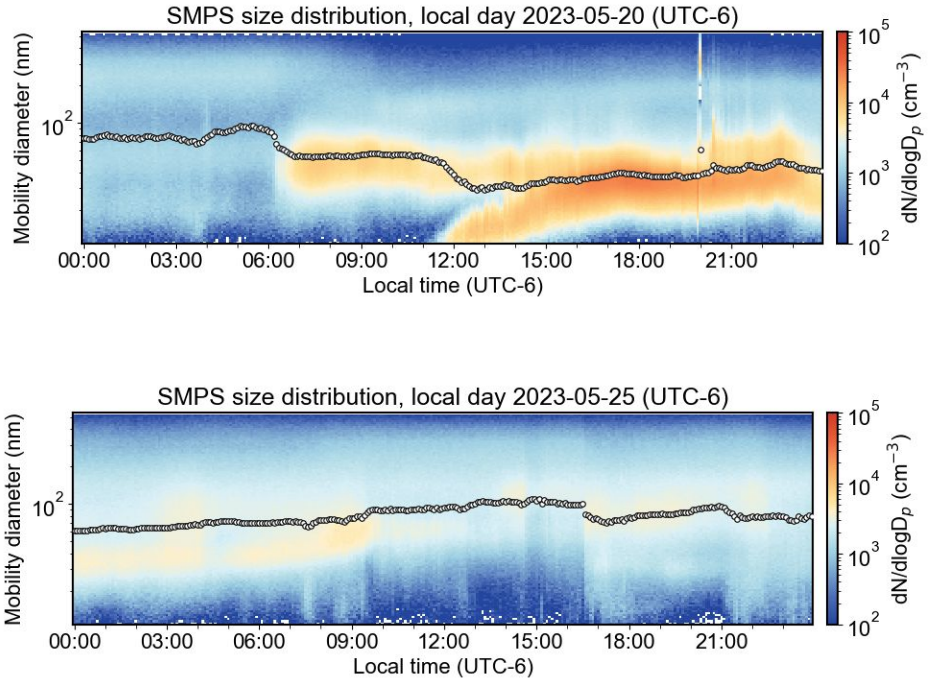
# Visual Classification of New Particle Formation Events

Visual Classification for NPF events consists of the following characteristics (Dal Maso et al., 2005) :

- A new mode of particles must appear in the size distribution.
- The mode must start in the nucleation mode size range
- The mode must persist over a time span of hours and show signs of growth

Usually classified into the following categories:

- Class I
- Class II
- Undefined
- Non-Event



Methodology

# Example Framework: Random Forest Model to Detect New Particle Formation

*Objective: Classify time periods as NPF event (1) or non-event (0) using CPC and SMPS observations*

## 1. Data Input

**Input Data Sources**

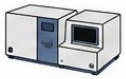
**Ultrafine CPC**
(UCPC)

Particle concentration
(e.g., $D_p$ = **(3 - 3000 nm)**)

**Fine CPC**
(FCPC)

Particle concentration
(e.g., $D_p$ = **(10 - 3000 nm)**)

**SMPS**



Size distribution
(e.g., 10–500 nm)

## 2. Feature Engineering

**Engineered Features**

$\Delta N$ ($cm^{-3}$)
Difference in aerosol concentration between ultrafine and fine CPCs

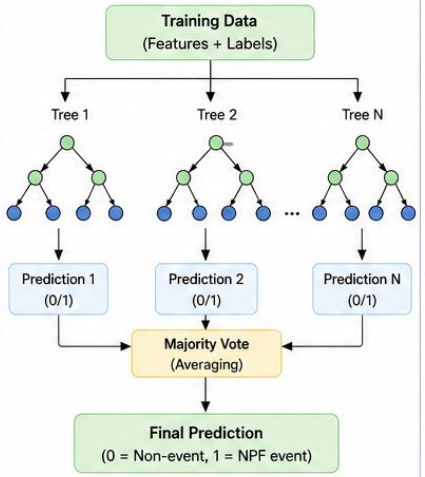
$\Delta N = N_{UCPC} - N_{FCPC}$

$D_{g,SMPS}$ (nm)
Geometric mean diameter from SMPS
(e.g., 10–500 nm)

Features computed for each time window (e.g., 1 hour)

## 3. Random Forest Model

**Random Forest Classifier**




**Model Training**

- Bootstrap sampling
- Random feature selection at each split
- Hyperparameter tuning (e.g., n_estimators, max_depth, min_samples_leaf)
- Cross-validation for robust performance

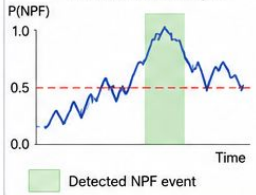
## 4. Prediction Output

**Model Output**

**Probability of NPF event**
**P(NPF) = 0.82**

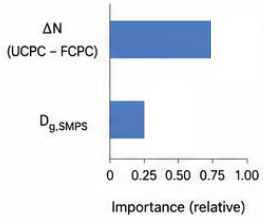

**Predicted Class**
**NPF event (1)**

**Time Series Example**




## 5. Interpretation

**Feature Importance (Example)**



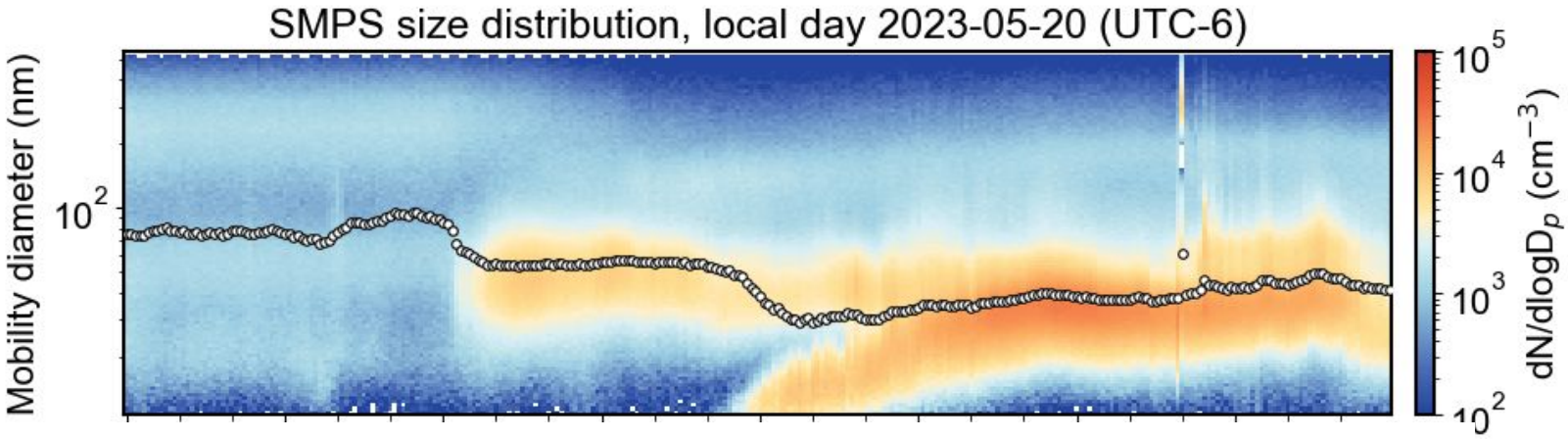

**Applications**

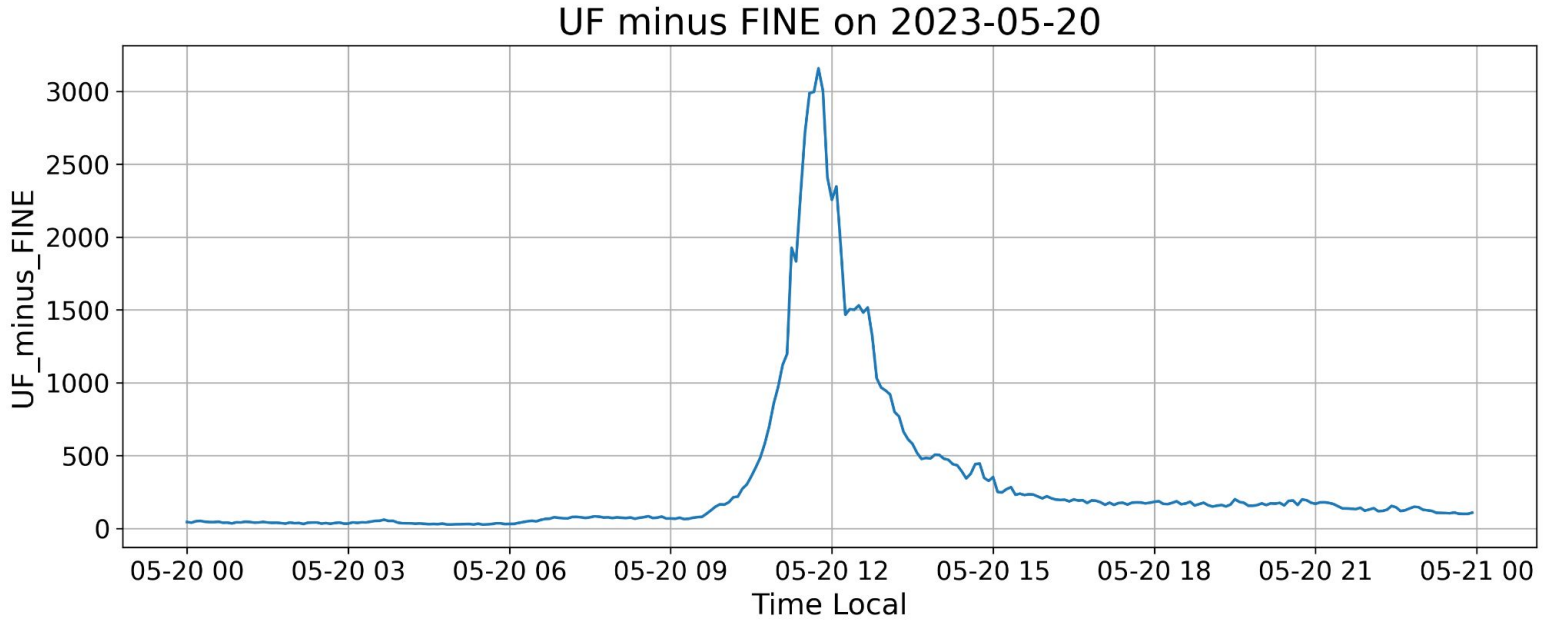
- ✓ Real-time NPF detection
- ✓ Event statistics & climatology
- ✓ Model evaluation
- ✓ Process understanding

**Workflow Summary**

Collect data from CPCs and SMPS → Compute features ($\Delta N$, $D_{g,SMPS}$) → Train Random Forest model → Predict NPF probability for new data → Interpret results & make decisions

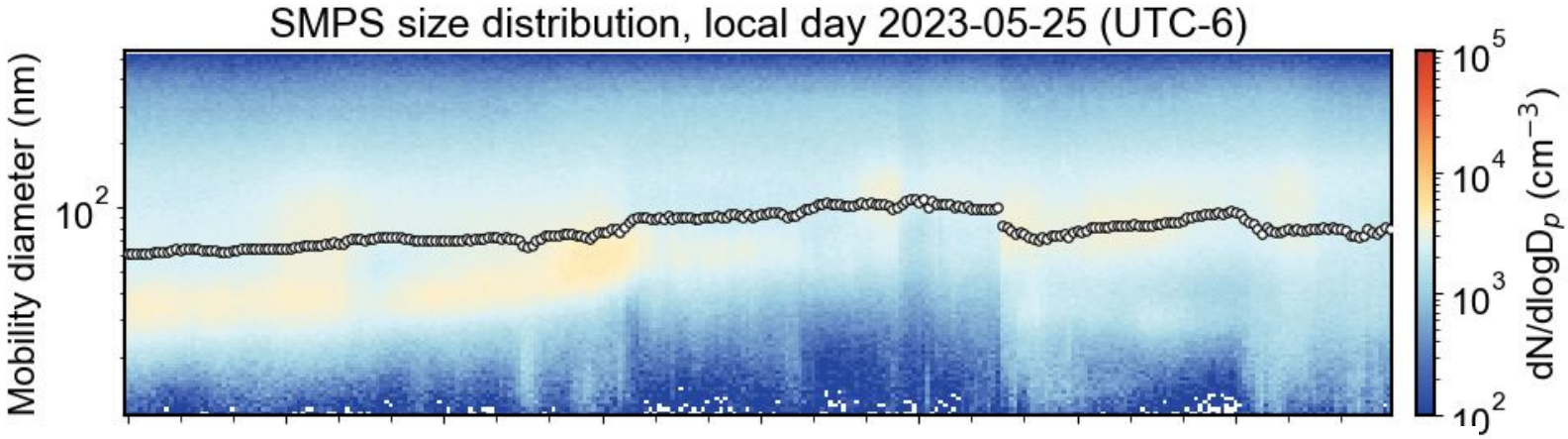
Parameter extract

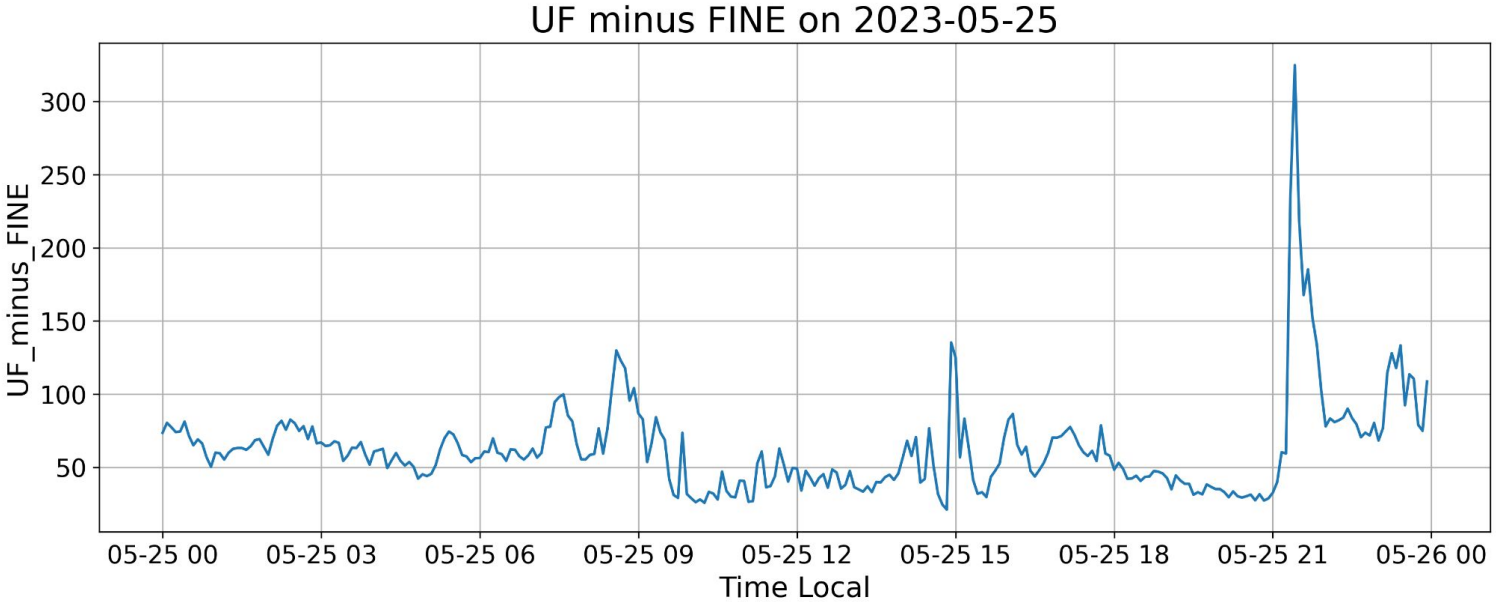



ARM

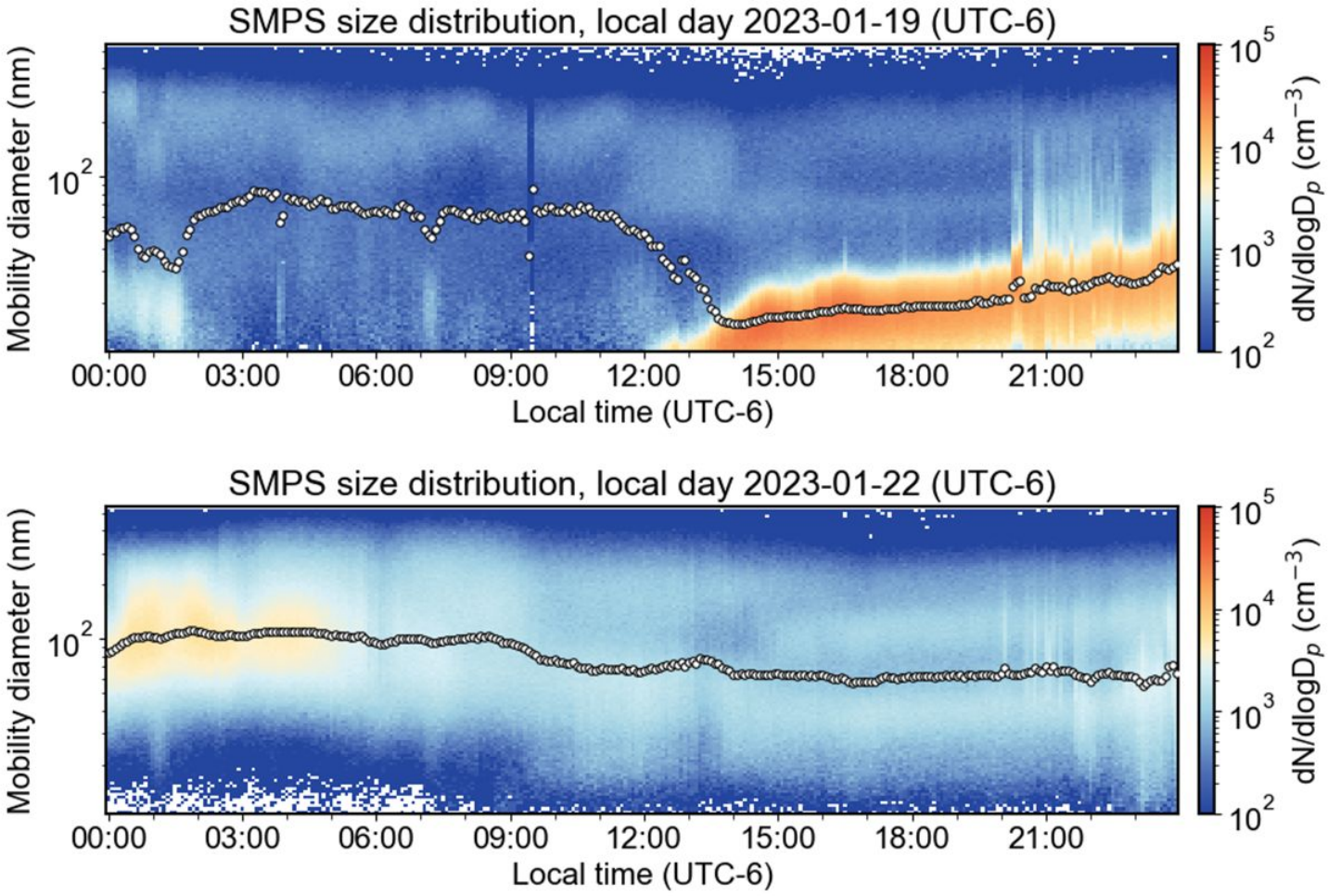
Parameter extract



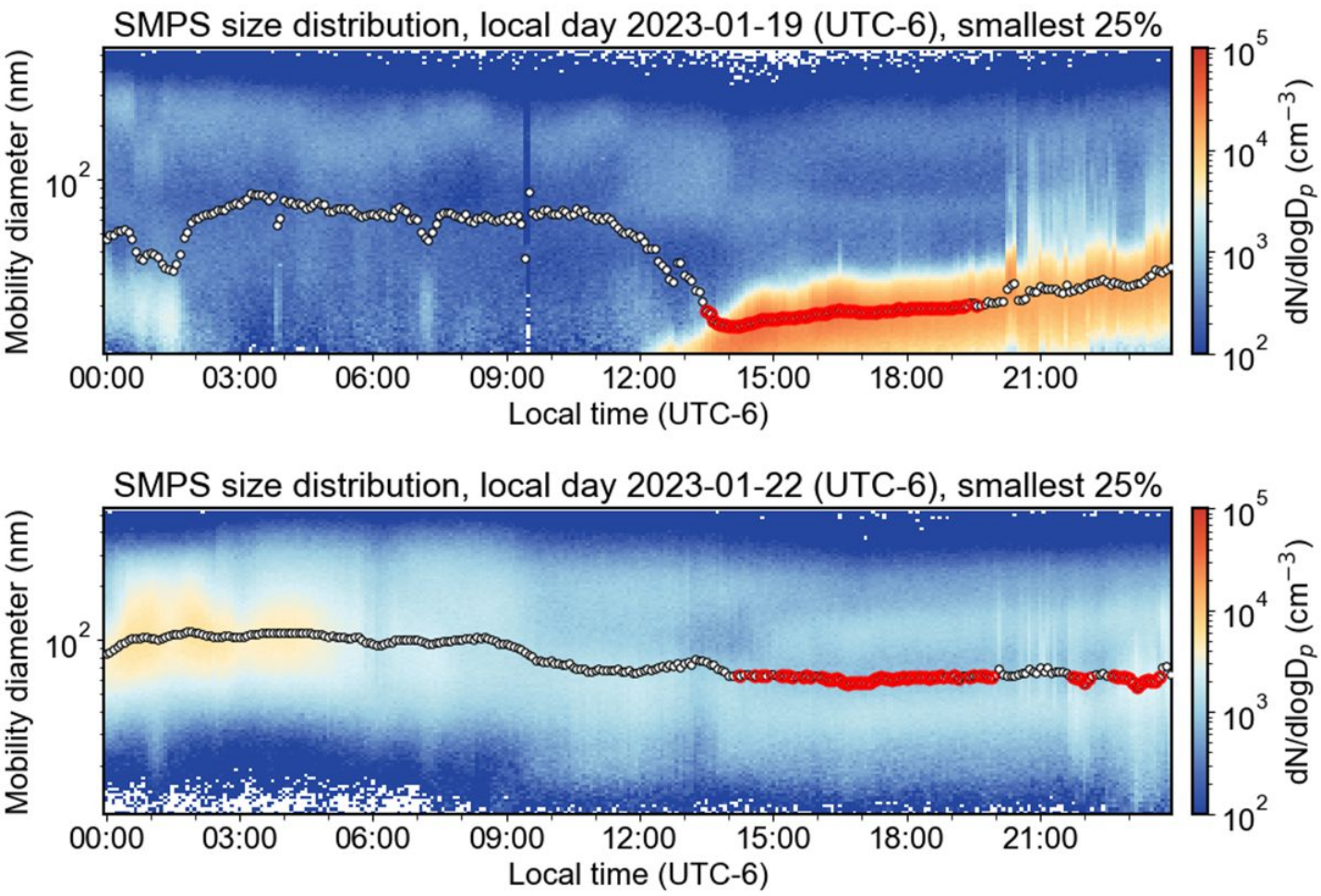

ARM

Parameter extract



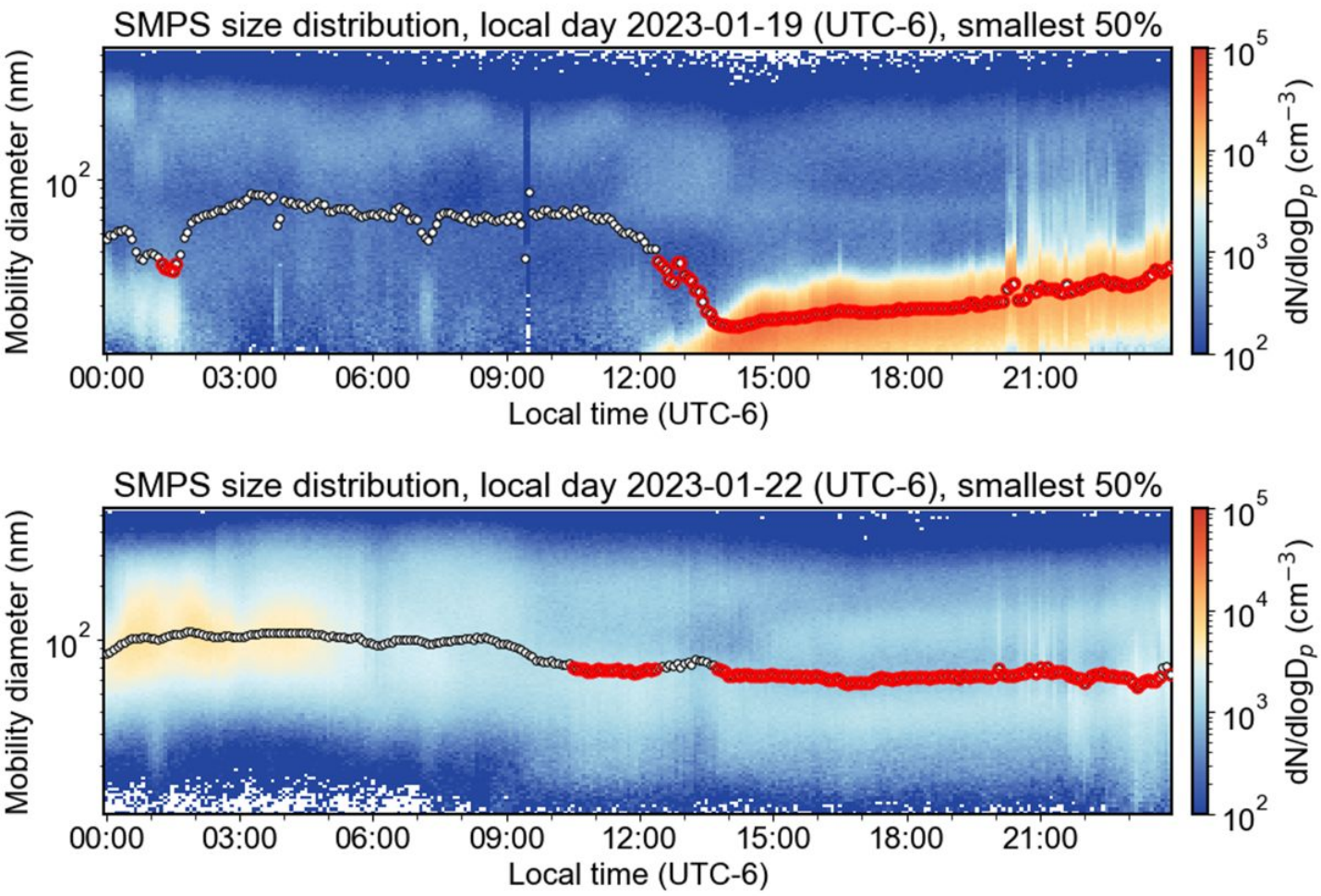

ARM

Parameter extract



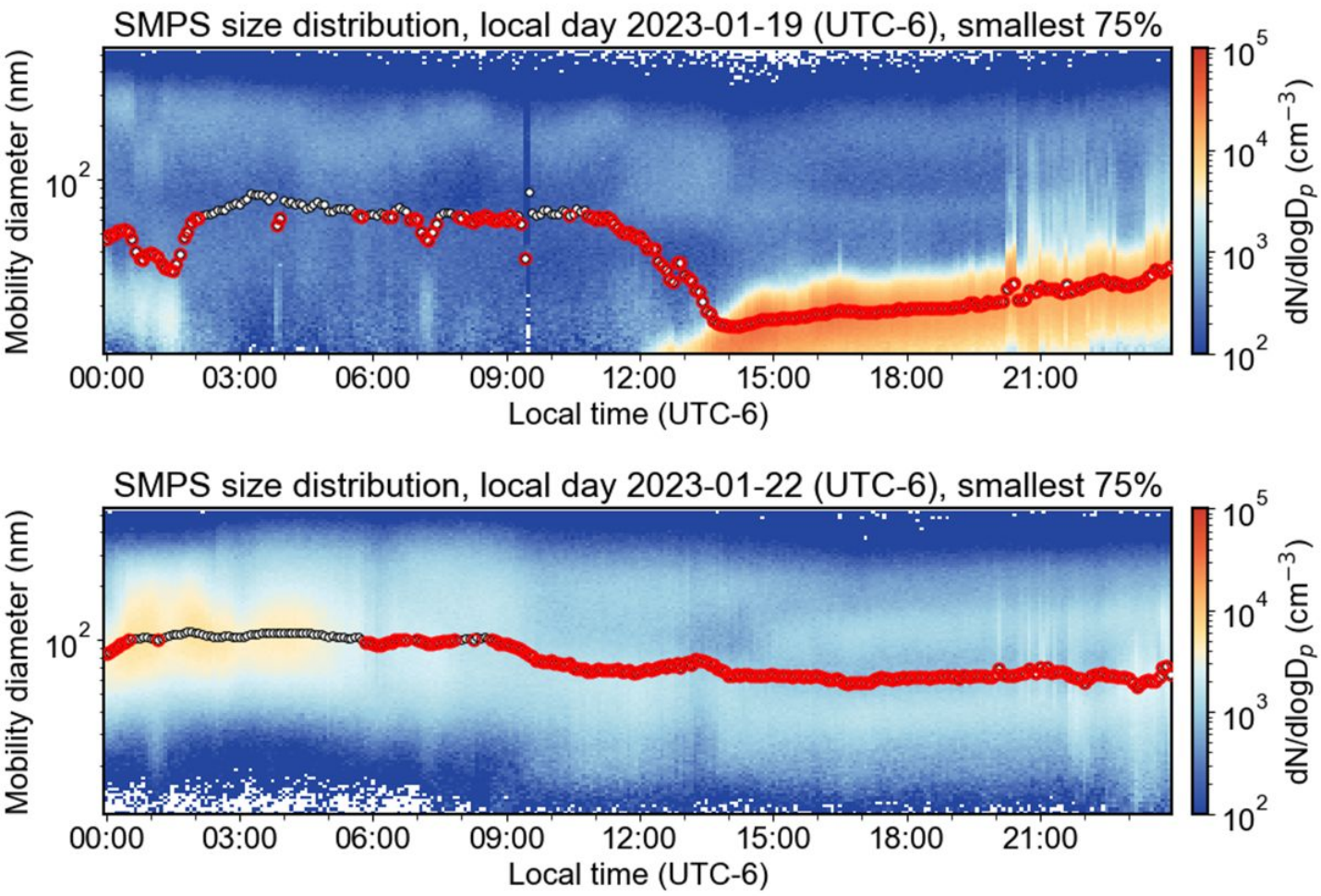

ARM

Parameter extract



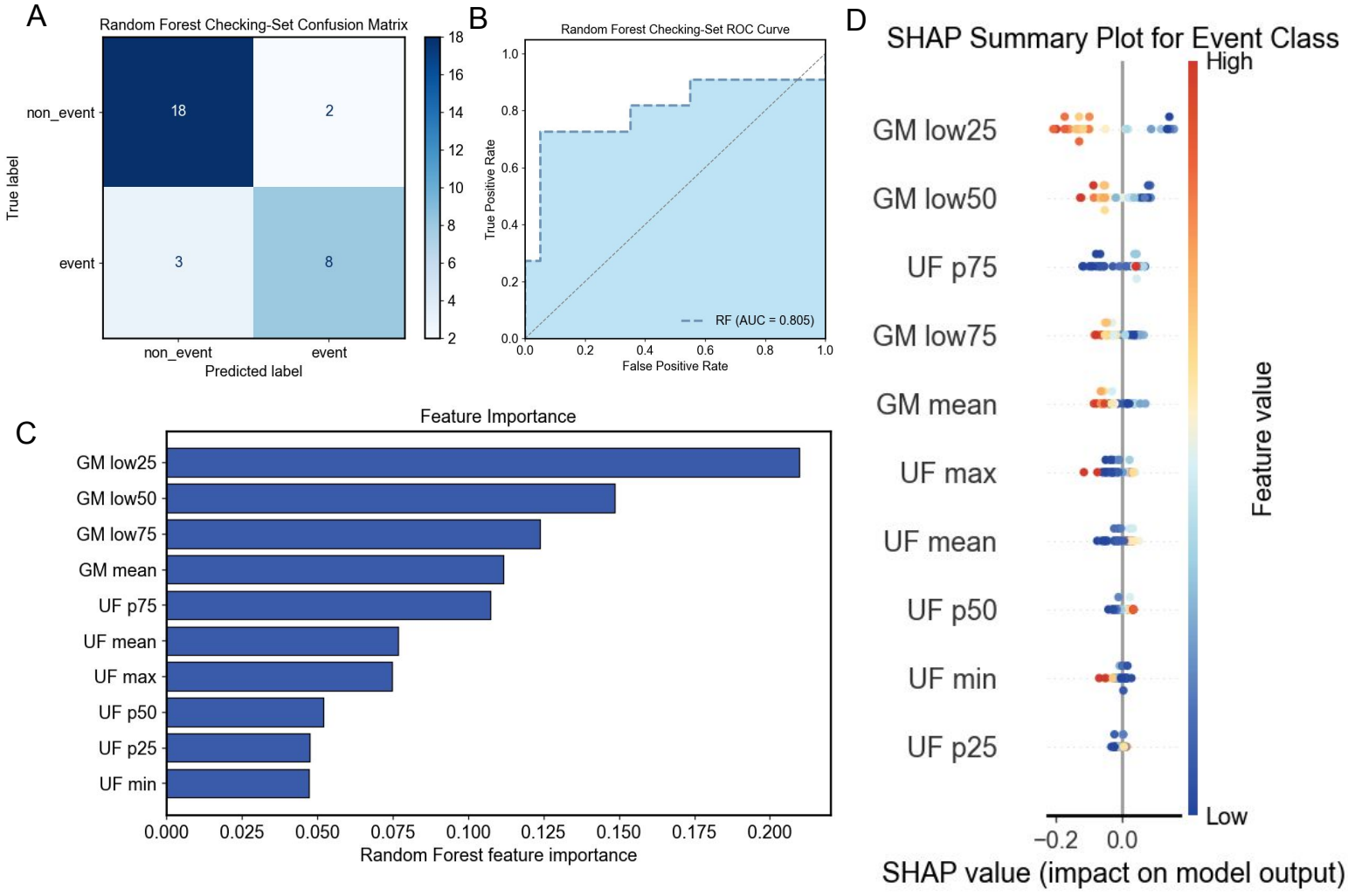

ARM

Parameter extract

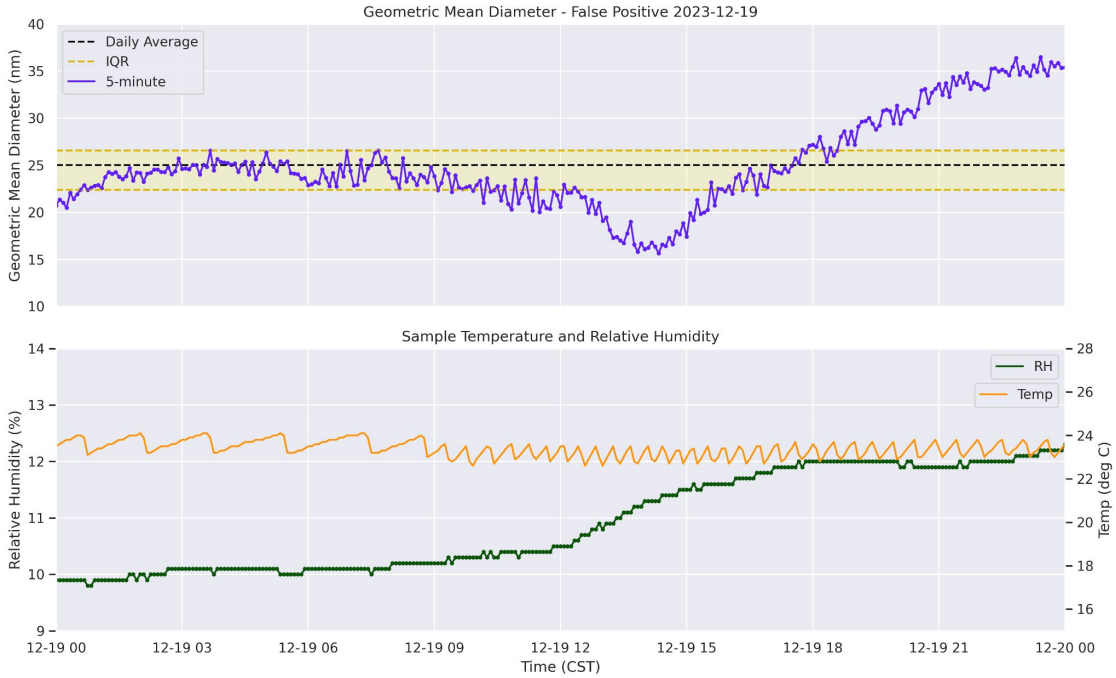



ARM

# Preliminary Results


A
Random Forest Checking-Set Confusion Matrix
non_event
event
True label
18
2
3
8
non_event
event
Predicted label
B
Random Forest Checking-Set ROC Curve
True Positive Rate
False Positive Rate
RF (AUC = 0.805)
C
Feature Importance
GM low25
GM low50
GM low75
GM mean
UF p75
UF mean
UF max
UF p50
UF p25
UF min
Random Forest feature importance
D
SHAP Summary Plot for Event Class
GM low25
GM low50
UF p75
GM low75
GM mean
UF max
UF mean
UF p50
UF min
UF p25
High
Low
Feature value
SHAP value (impact on model output)

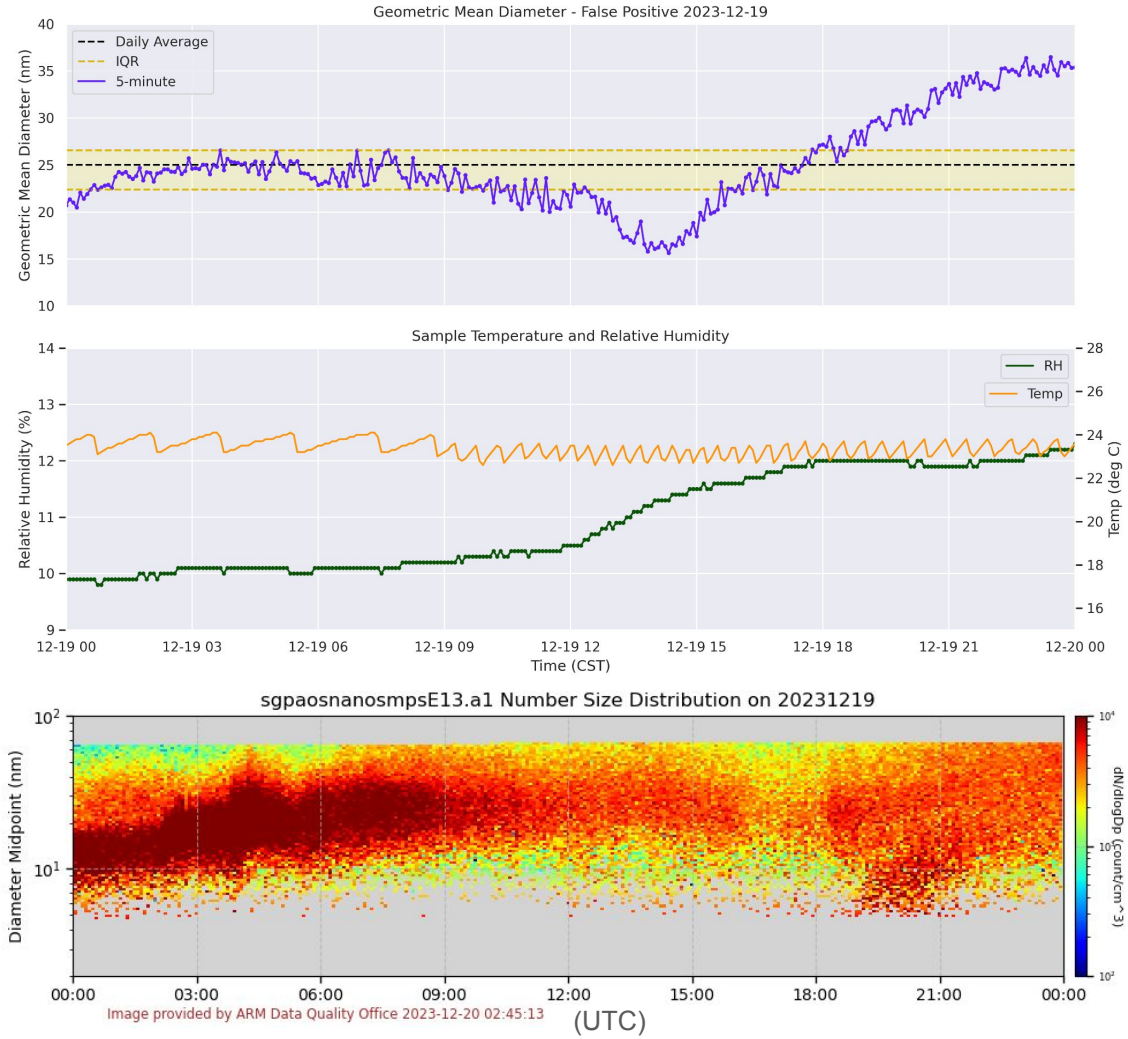
# Using Nano SMPS to Check Model Failures




- ***False positive***:
  Model says yes event
  Data says no event

- Increase in the GMD, IQR in lower 25th

The 25th percentile is *1.13 nm farther* from the mean than the 75th percentile



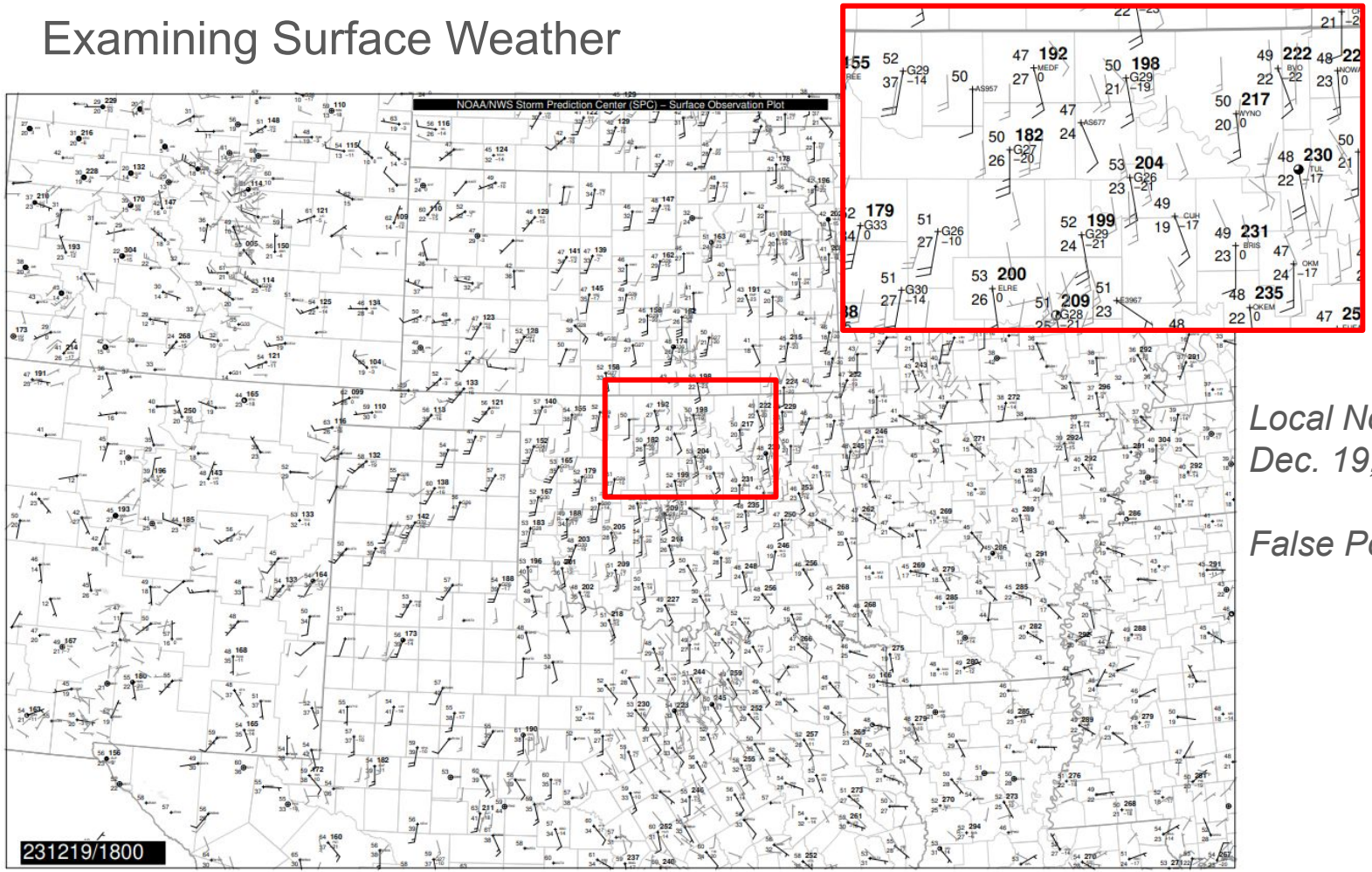

# Using Nano SMPS to Check Model Failures

- ***False positive***:
  Model says yes event
  Data says no event

- Increase in the GMD, IQR in lower 25th

- Size distribution reveals a low concentration

# Examining Surface Weather

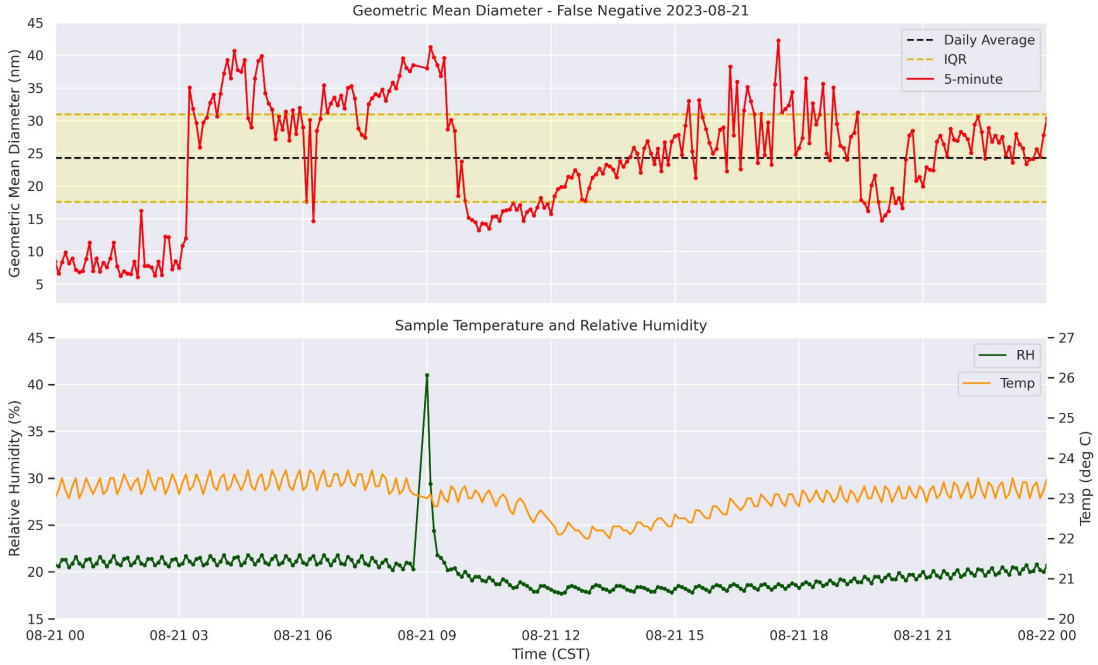



*Local Noon, Dec. 19, 2023*

*False Positive*

Archived Surface Maps - NOAA SPC

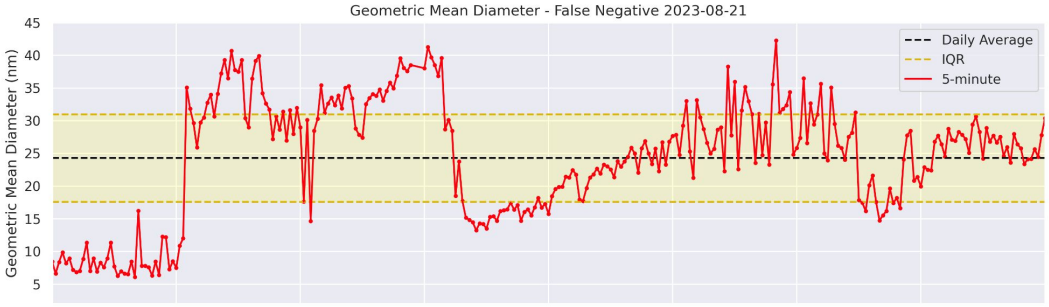
# Using Nano SMPS to Check Model Failures

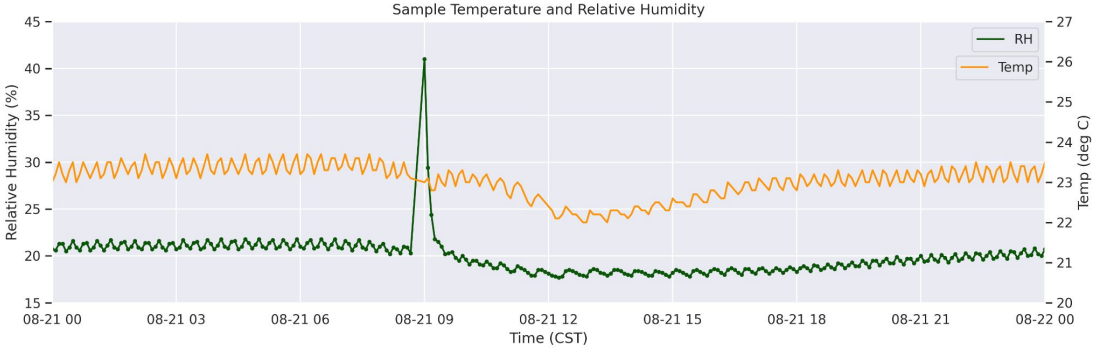


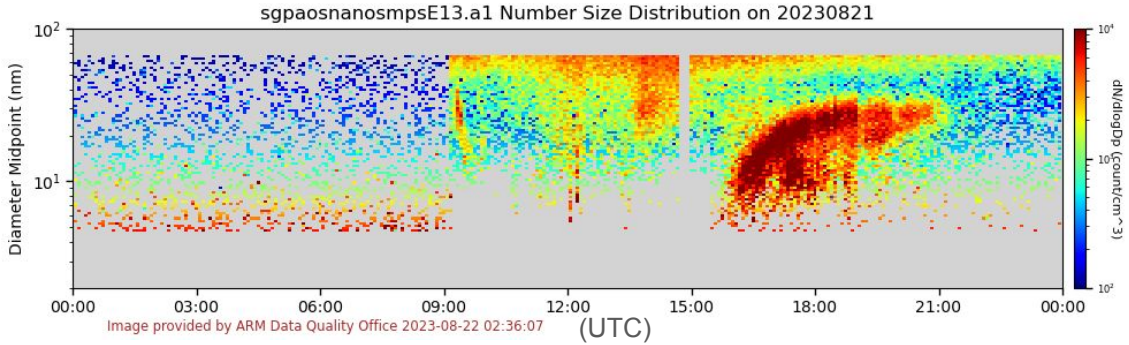

- ***False negative***:
  Model says no event
  Data says yes event

- Increase in the GMD, equidistant IQR

The 25th percentile is *0.04 nm farther* from the mean than the 75th percentile



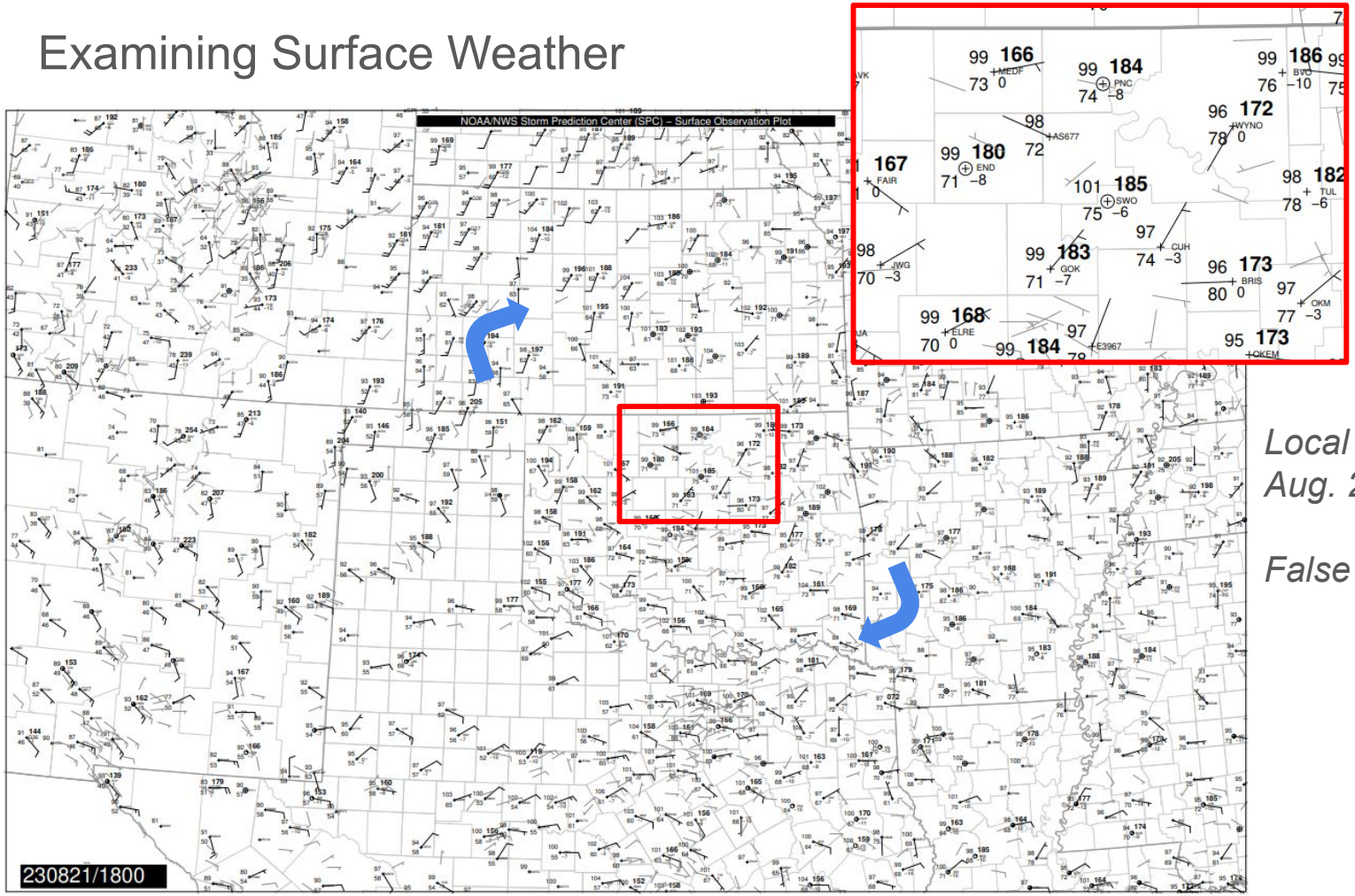

# Using Nano SMPS to Check Model Failures

- ***False negative***:
  Model says no event
  Data says yes event

- Increase in the GMD, equidistant IQR

- Size distribution reveals a clear NPF event

# Examining Surface Weather




*Local Noon, Aug. 21, 2023*

*False Negative*

Archived Surface Maps - NOAA SPC

# Summary and Future work

## Summary

- Built a lightweight model for automatic daily NPF event classification.
- Used a simple dataset and limited labeled data. Achieved not bad results
- The method can be applied to long-term datasets.
- Since it relies on commonly available SMPS/CPC data, it repeatable at other sites.

## Future Directions

- Explore better predictive features from size-distribution data.
- Improve event/non-event classification with more labeled days.
- Combine labeled NPF results with chemical measurements.
- Investigate how chemical species and atmospheric conditions affect NPF occurrence.

ARM

# Works Cited


1. Hao, Weixing, et al. "Employing machine learning for new particle formation identification and mechanistic analysis: Insights from a six-year observational study in the Southern Great Plains." Journal of Geophysical Research: Atmospheres 131.1 (2026): e2024JD043116.
2. Kuang, C., Ermold, B., Singh, A., Salwen, C., & Hayes, C. Scanning mobility particle sizer (AOSNANOSMPS), 2023-01-01 to 2024-01-01, Southern Great Plains (SGP), Lamont, OK (Extended and Co-located with C1) (E13). Atmospheric Radiation Measurement (ARM) User Facility. https://doi.org/10.5439/1635016
3. Atmospheric Radiation Measurement (ARM) user facility. 2016. Scanning mobility particle sizer (AOSSMPS), 2023-01-01 to 2024-12-31, Southern Great Plains (SGP) Supplemental 1, Lamont, OK (S01). Compiled by A. Singh, C. Kuang, J. Howie, C. Salwen and C. Hayes. ARM Data Center. Data set accessed 2026-05-21 at http://dx.doi.org/10.5439/1476898.
4. Su, P., Joutsensaari, J., Dada, L., Zaidan, M. A., Nieminen, T., Li, X., Wu, Y., Decesari, S., Tarkoma, S., Petäjä, T., Kulmala, M., and Pellikka, P.: New particle formation event detection with Mask R-CNN, Atmos. Chem. Phys., 22, 1293–1309, https://doi.org/10.5194/acp-22-1293-2022, 2022.
5. Koontz, Annette, et al. “Condensation Particle Counter (AOSCPCUF), 2017-06-29 to 2026-05-20, Southern Great Plains (SGP), Lamont, OK (Extended and Co-Located with C1) (E13).” Atmospheric Radiation Measurement (ARM) User Facility, doi:10.5439/1378933. Accessed 21 May 2026.
6. Koontz, Annette, et al. “Condensation Particle Counter (AOSCPCF), 2016-11-14 to 2026-05-20, Southern Great Plains (SGP), Lamont, OK (Extended and Co-Located with C1) (E13).” Atmospheric Radiation Measurement (ARM) User Facility, doi:10.5439/1352536. Accessed 21 May 2026.
7. Singh, Ashish, et al. “Scanning Mobility Particle Sizer (AOSSMPS), 2016-11-15 to 2026-05-20, Southern Great Plains (SGP), Lamont, OK (Extended and Co-Located with C1) (E13).” Atmospheric Radiation Measurement (ARM) User Facility, doi:10.5439/1476898. Accessed 21 May 2026.
8. Kuang, Chongai, et al. “Scanning Mobility Particle Sizer (AOSNANOSMPS), 2016-11-15 to 2026-05-20, Southern Great Plains (SGP), Lamont, OK (Extended and Co-Located with C1) (E13).” Atmospheric Radiation Measurement (ARM) User Facility, doi:10.5439/1635016. Accessed 21 May 2026.

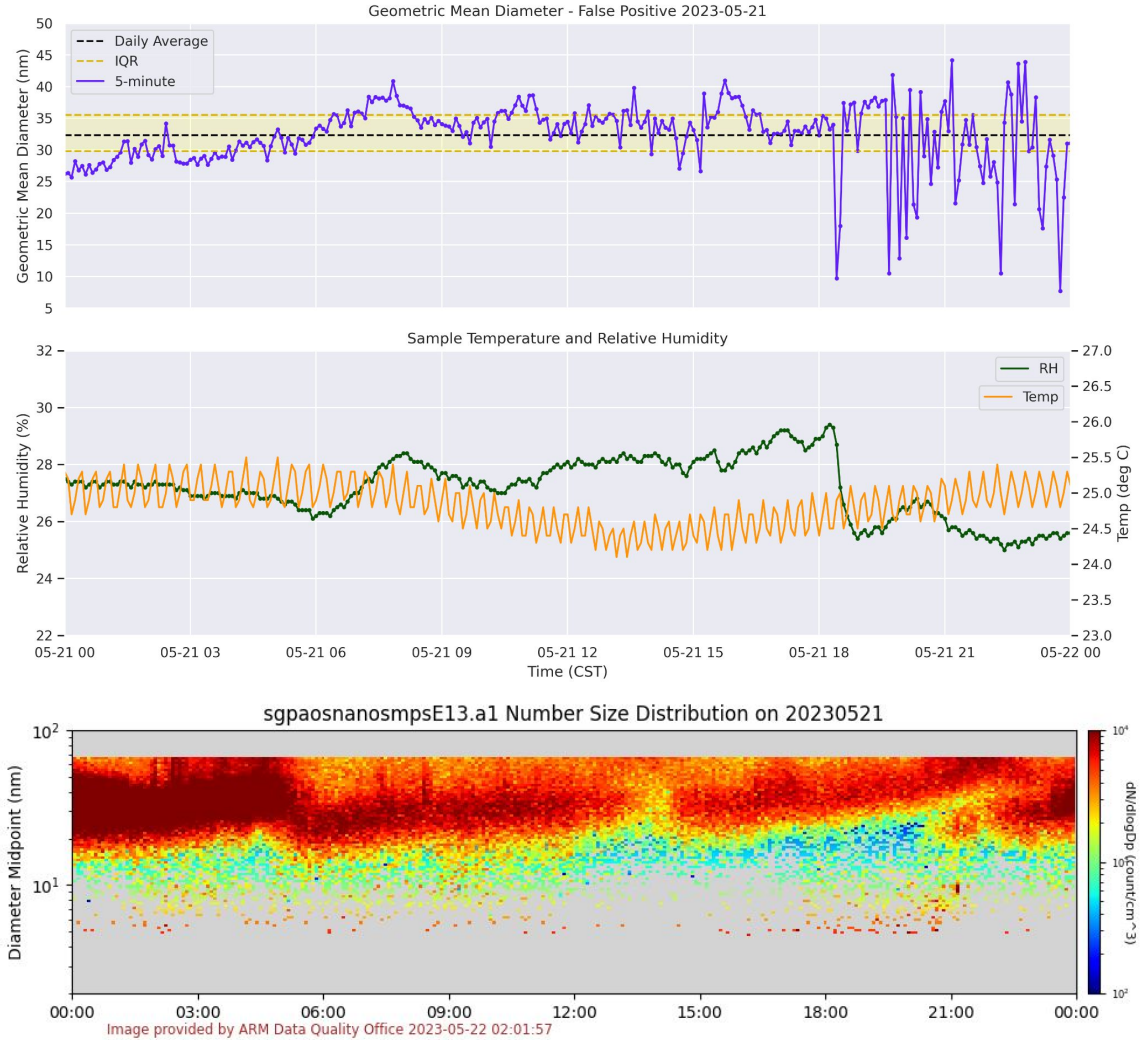

ARM

# Extra:
# False Positive

Difference b/w 75th percentile and mean:
= 3.17 nm

Difference b/w 25th percentile and mean:
= 2.52 nm

Total difference:
= 0.65 nm



# Extra:
# False Negative

Difference b/w 75th percentile and mean:
= 5.25 nm

Difference b/w 25th percentile and mean:
= 5.09 nm

Total difference:
= 0.16 nm