ACTIVATE Science Team Meeting

Nov 13, 2023

'Machine Learning Shines a Laser Light on Murky Aerosol Picture'

Estimation of CCN Concentration from Lidar Observations Using Neural Networks

Lan Gao[1] and Jens Redemann[1]

Acknowledgement:

1. School of Meteorology, University of Oklahoma

   Emily Lenhardt, Benting Chen, Feng Xu

2. NASA Langley Research Center

   Richard Ferrare, Chris Hostetler, Marta Fenn, Johnathan Hair

CLouds · CLimatE · Aerosols · Radiation

Aerosol Cloud Meteorology Interactions Over the Western Atlantic Experiment

Motivation

REVIEW ARTICLE

10.1002/2013RG000441

Global observations of aerosol-cloud-precipitation-climate interactions

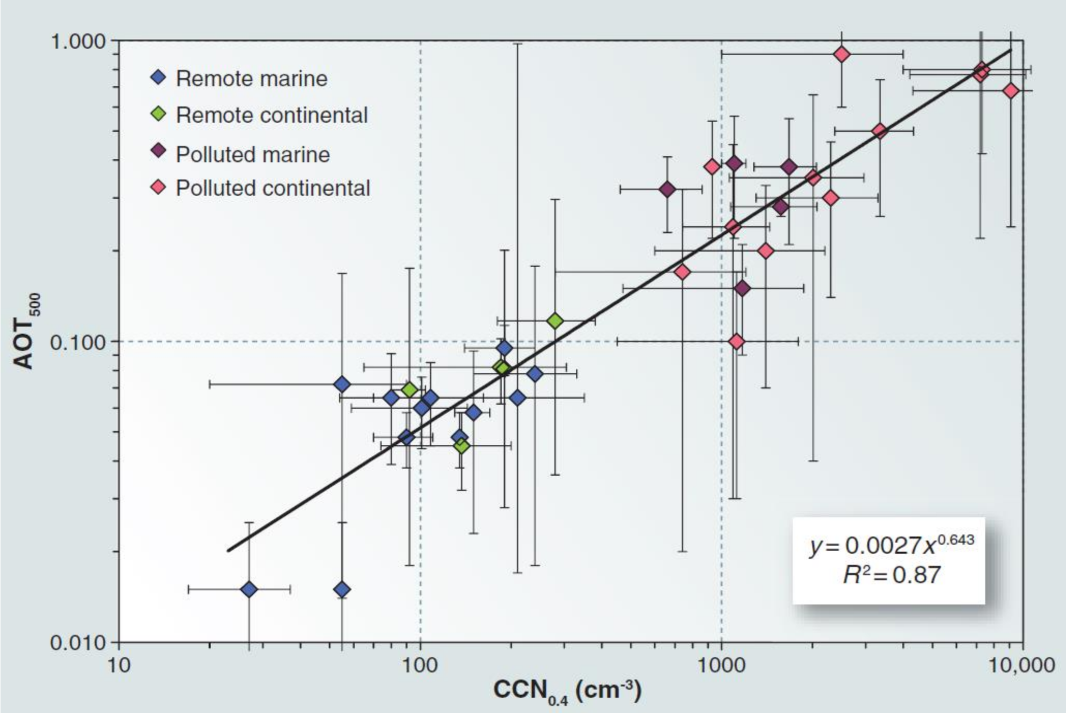
Key Points:
- Quantifying aerosol-cloud-climate interactions is a major challenge
- The science of existing and emerging new observational methods is reviewed
- A roadmap for in situ and remote sensing energy closure experiments is provided

Daniel Rosenfeld[1], Meinrat O. Andreae[2], Ari Asmi[3], Mian Chin[4], Gerrit de Leeuw[3,5], David P. Donovan[6], Ralph Kahn[4], Stefan Kinne[7], Niku Kivekäs[5,8], Markku Kulmala[3], William Lau[4], K. Sebastian Schmidt[9], Tanja Suni[3], Thomas Wagner[10], Martin Wild[11], and Johannes Quaas[12]

[1]Institute of Earth Sciences, The Hebrew University of Jerusalem, Israel, [2]Biogeochemistry Department, Max Planck Institute for Chemistry, Mainz, Germany, [3]Department of Physics, University of Helsinki, Helsinki, Finland, [4]Earth Science Division, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, [5]Atmospheric Composition Research Unit, Finnish

An urgent need for global observations of CCN(*S*) by remote sensing follows from these considerations. Because the microphysical and radiative effects of aerosols act simultaneously on a given cloud population and change the thermodynamic environment of cloud formation and the microphysical processes of the cloud development [*Rosenfeld et al.*, 2008a], the CCN(*S*) field should be observed simultaneously with aerosol light scattering and absorption properties. Since the effects of light scattering (cooling of the ground surface) and absorption (cooling at the ground combined with heating aloft) have different impacts on atmospheric stability, they must be observed independently. Here, quantitative measures of absorption are especially important.

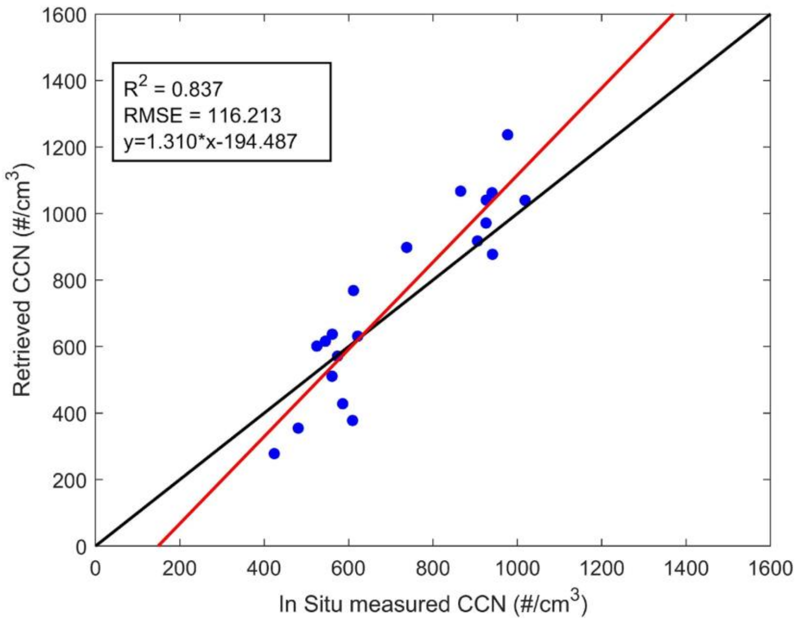
- **Column-effective aerosol quantities may not be relevant to aerosol-cloud interaction.**
- **The uncertainty of CCN-AOD parameterization can be very large depending on:**
  - **Types of aerosol**
  - **Vertical distribution**
  - **Humidity response of light scattering**
  - **Spatial-temporal variability**

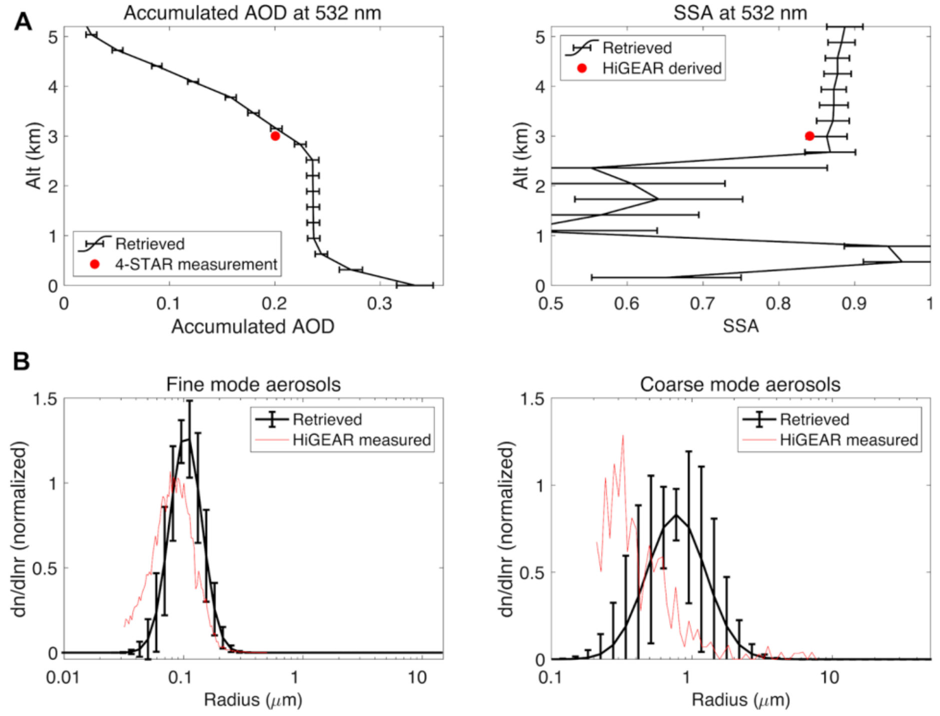


Stier, 2016: "…71 % of the area of the globe shows correlation coefficients between CCN0.2 % at cloud base and aerosol optical depth (AOD) below 0.5, i.e. AOD variability explains only 25 % of the CCN variance" – model-based, self-consistent.

Physics-based retrieval of CCN using lidar and polarimeter observations (Gao et al, AGU 2021)

- Developed an optimization approach to retrieve the speciated aerosol profiles from lidar and polarimeter observations.
- Used aerosol reanalysis product to estimate the bulk hygroscopicity parameter for aerosol mixture.
- Applied κ-Köhler theory to calculate CCN concentration.





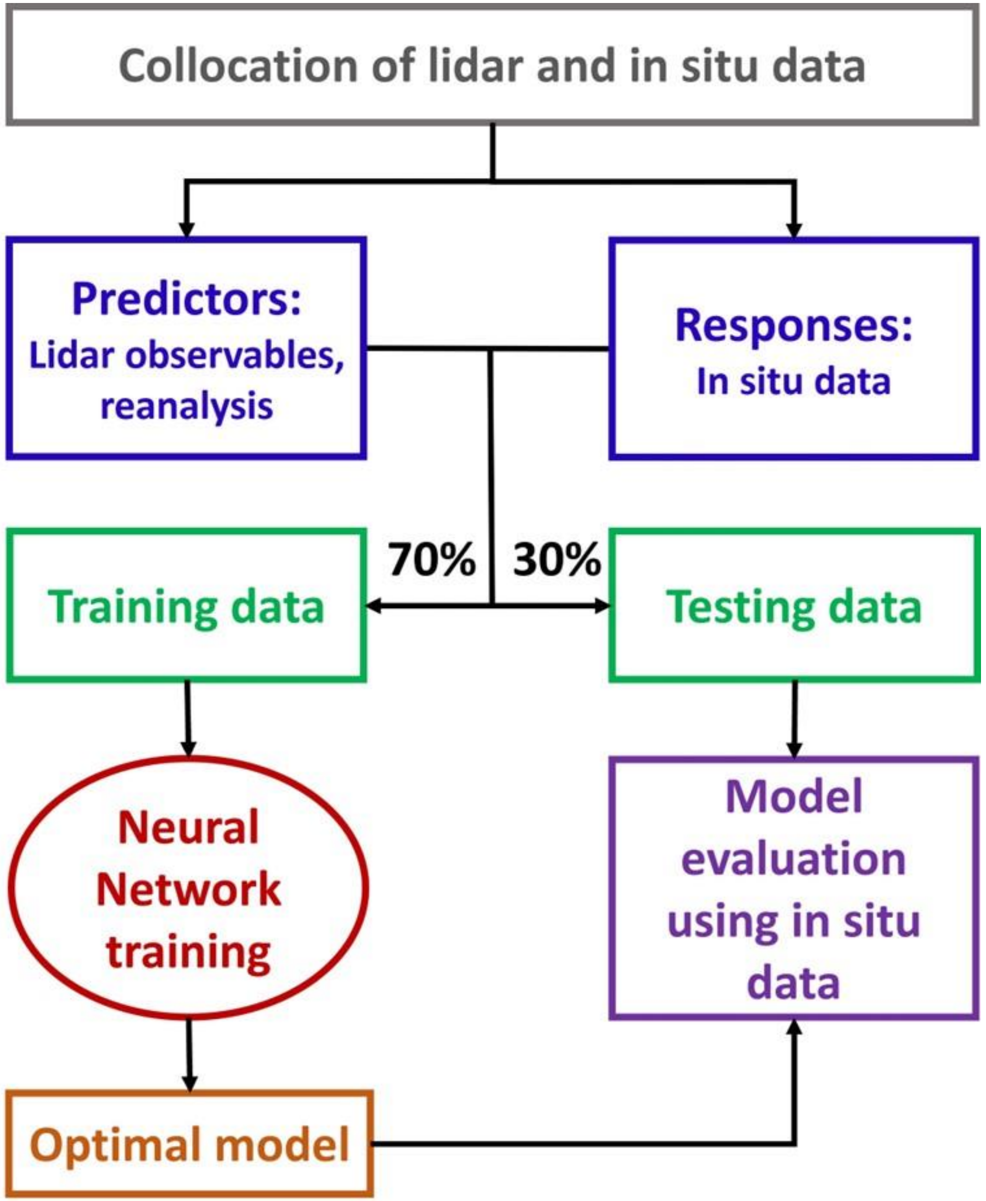
(Xu, F., Gao, L., Redemann, J., et al., 2021)

Limitations – AOS retrieval simulation "lessons learned"

- Great dependence on a priori information (aerosol size distribution and chemical composition) to retrieve CCN
- Computationally very expensive.

Methodology

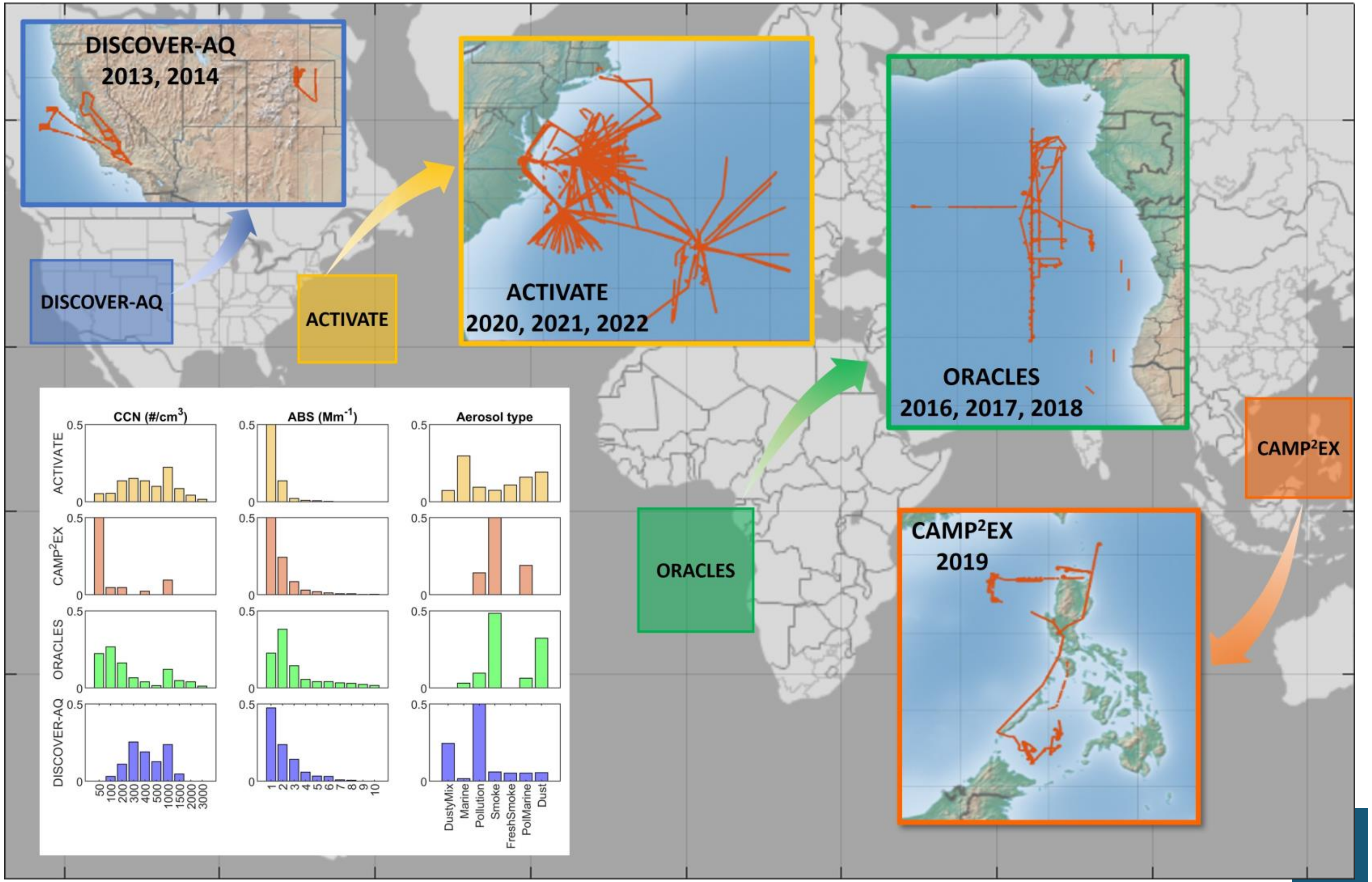
- Collocate HSRL-2 and in-situ measured CCN from multiple campaigns.
  - ✓ ACTIVATE, CAMP^{2}EX, DISCOVER-AQ, ORACLES
- Train neural networks for different sets of lidar observables (e.g., NASA AOS, EarthCARE).
  - ✓ HSRL-2: 3β + 2α + 3δ
  - ✓ HSRL-1: 2β + 1α + 2δ
  - ✓ EarthCARE: 1β + 1α + 1δ
  - ✓ Simulated-Elastic-Backscatter (SEBL): 2β + 2δ
- Evaluate model prediction using in-situ measured CCN.
  - ✓ Correlation coefficient (R)
  - ✓ Mean absolute error (MAE)
  - ✓ Mean relative error (MRE)



Data

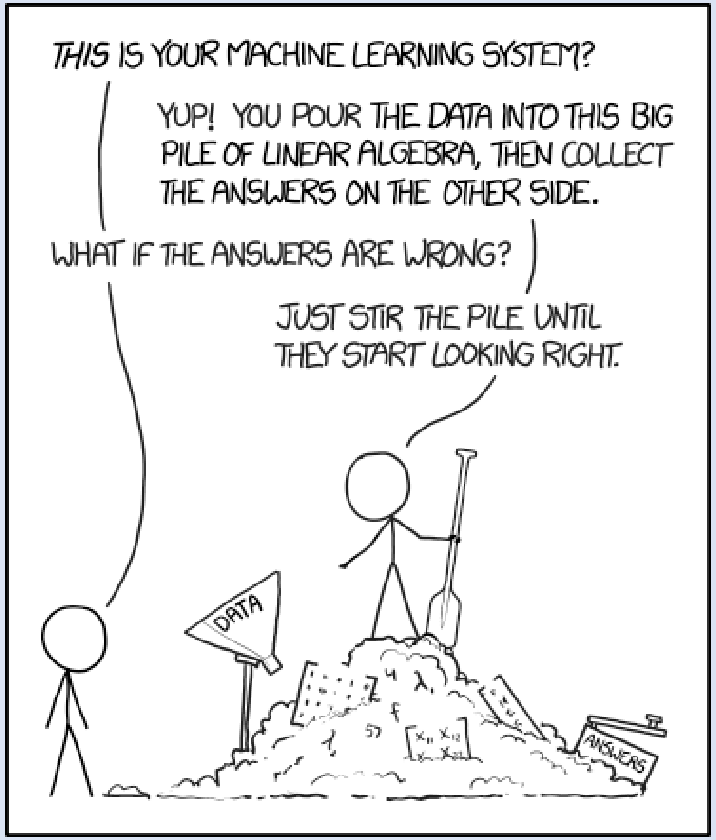
- Aircraft observations (lidar and in situ) from multiple campaigns

Lidar observables	EXT_{355}, EXT_{532}, BSC_{355}, BSC_{532}, BSC_{1064}, $DEPO_{355}$, $DEPO_{532}$, $DEPO_{1064}$
Ancillary	Altitude (ALT), Relative humidity (RH), Temperature (T)
In situ	CCN concentration (N_{ccn}) at 0.4% SS



How to Deal with the black box…

> ‘Garbage in, garbage out’
> – computer science



(https://xkcd.com/1838/)

Before machine learning:

- Remove data that has large uncertainties
  - Lidar
    - ✓ Negative lidar observables
    - ✓ Aerosol depolarization ratio greater than 1
  - In situ
    - ✓ CCN below 10 cm^{-3}
    - ✓ ABS below 0.1 Mm^{-1}

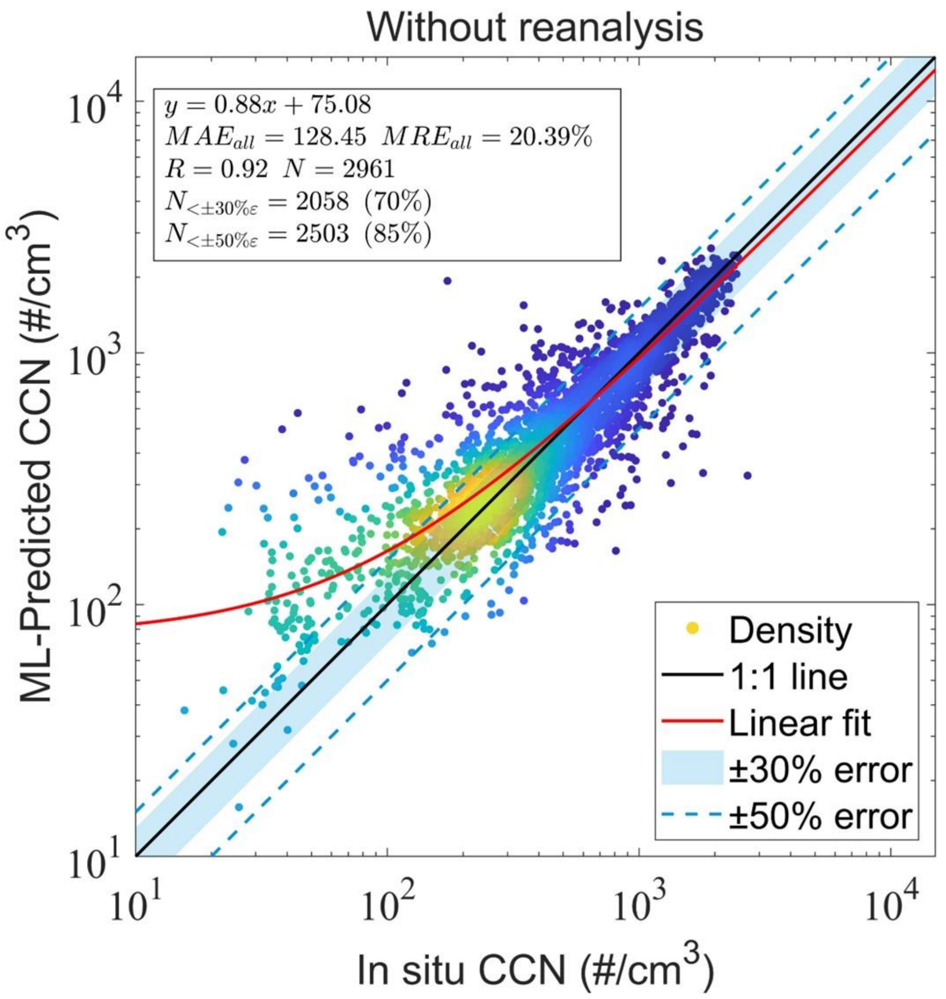
Algorithm selection:

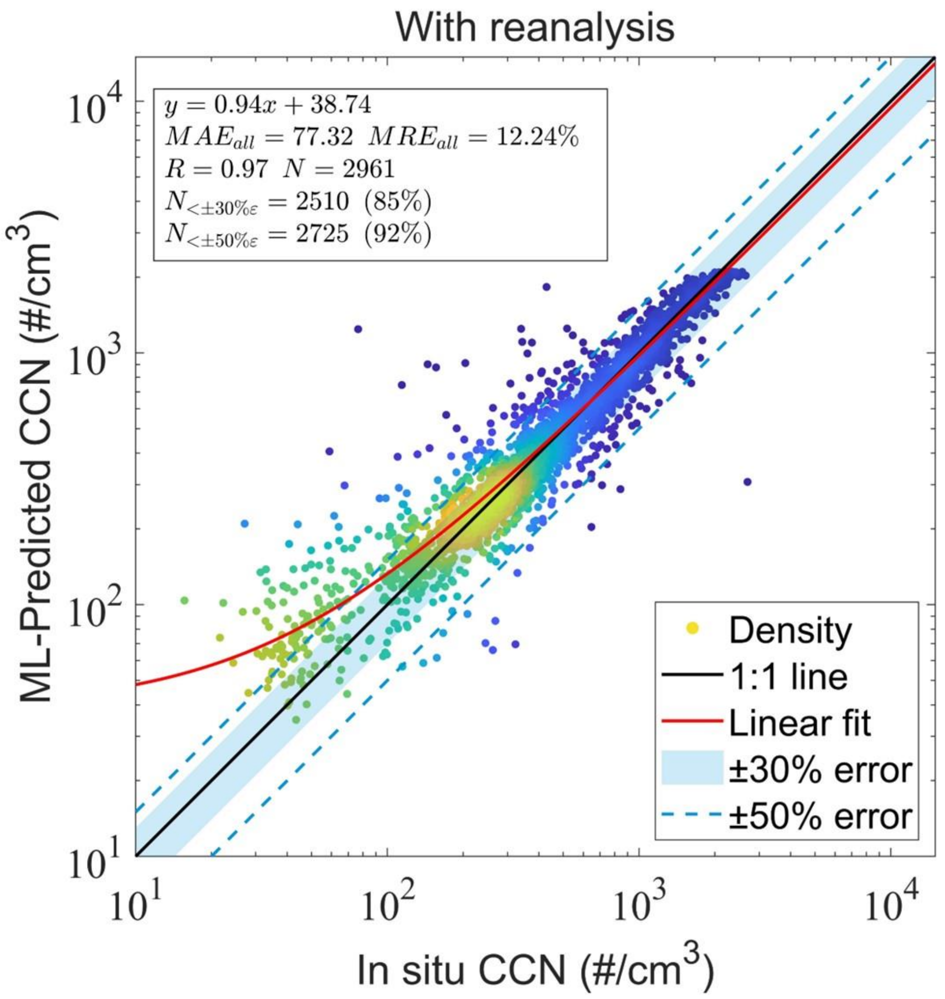
- Supervised regression learning problem with large number of numerical features.
- Fully-Connected Neural Network (FCNN) regression model

Architecture setup:

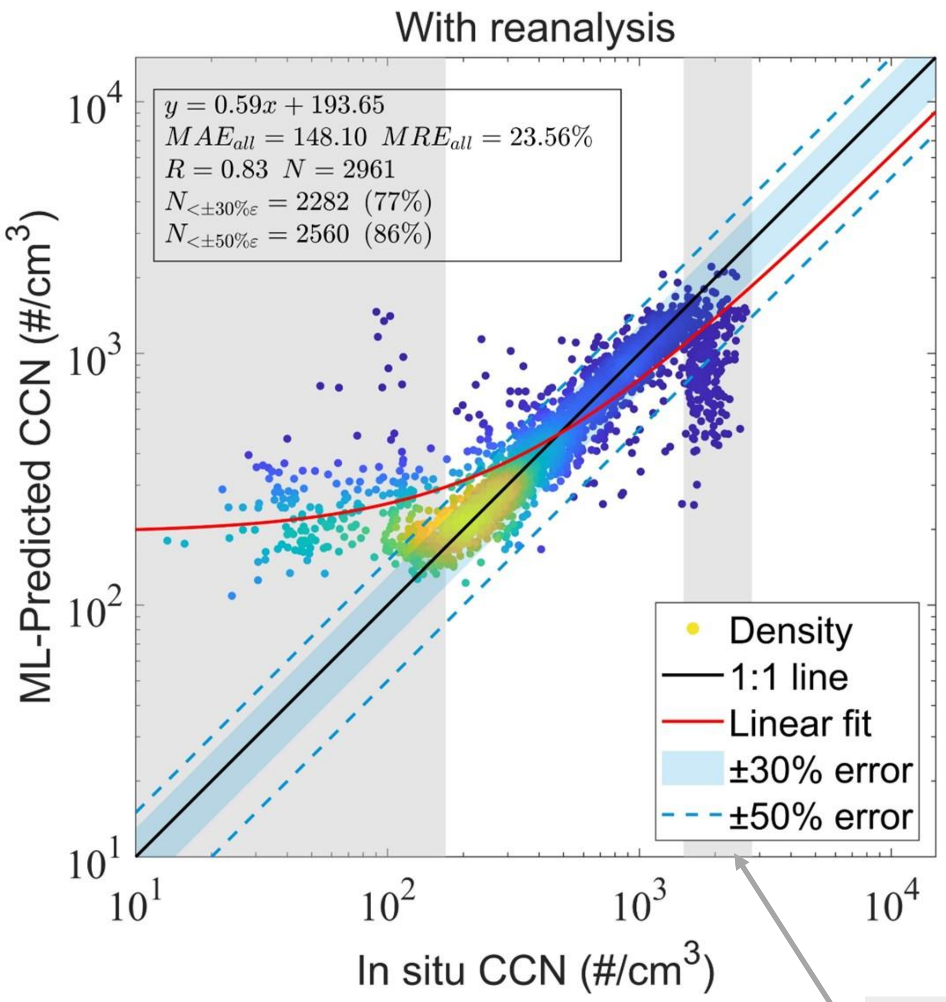
- Training data: 70%, Testing data: 30%
- 10-fold cross validation
- Hyperparameters are tuned iteratively during the training using Bayesian optimization

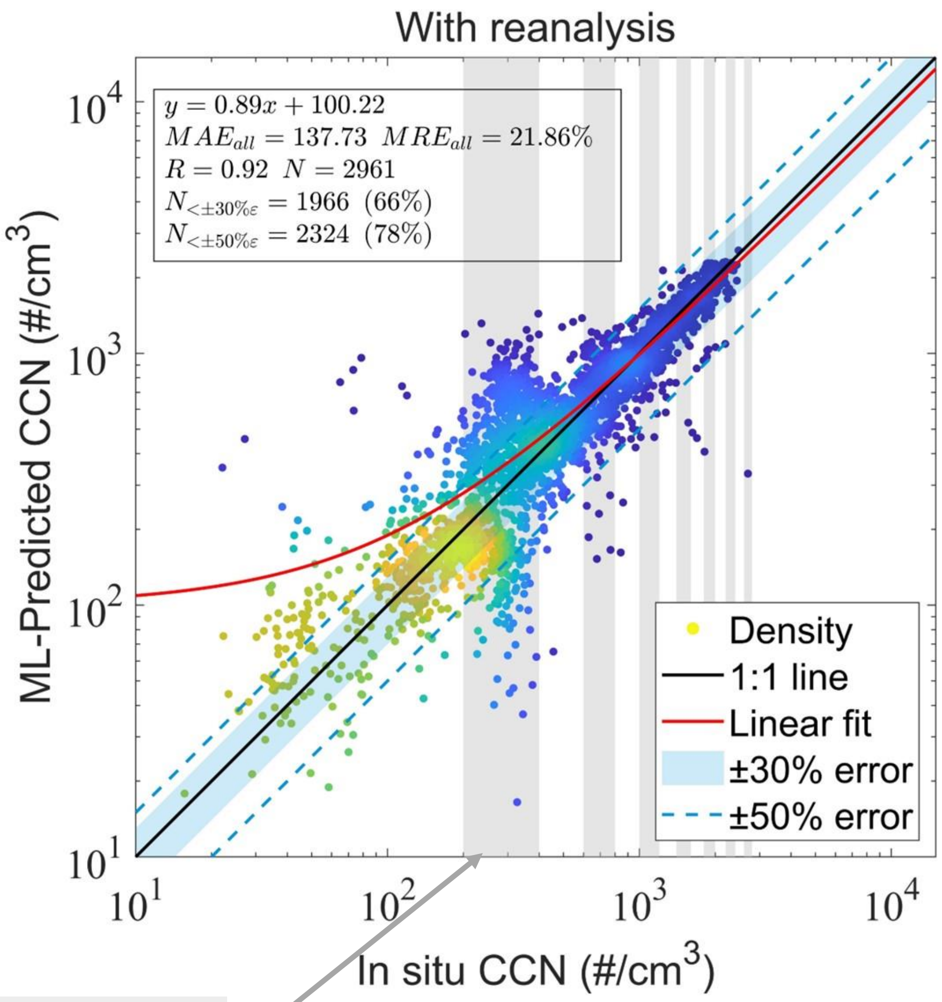
ML-predicted CCN concentrations from HSRL-2 without and with reanalysis





Test for model performance with incomplete in situ data

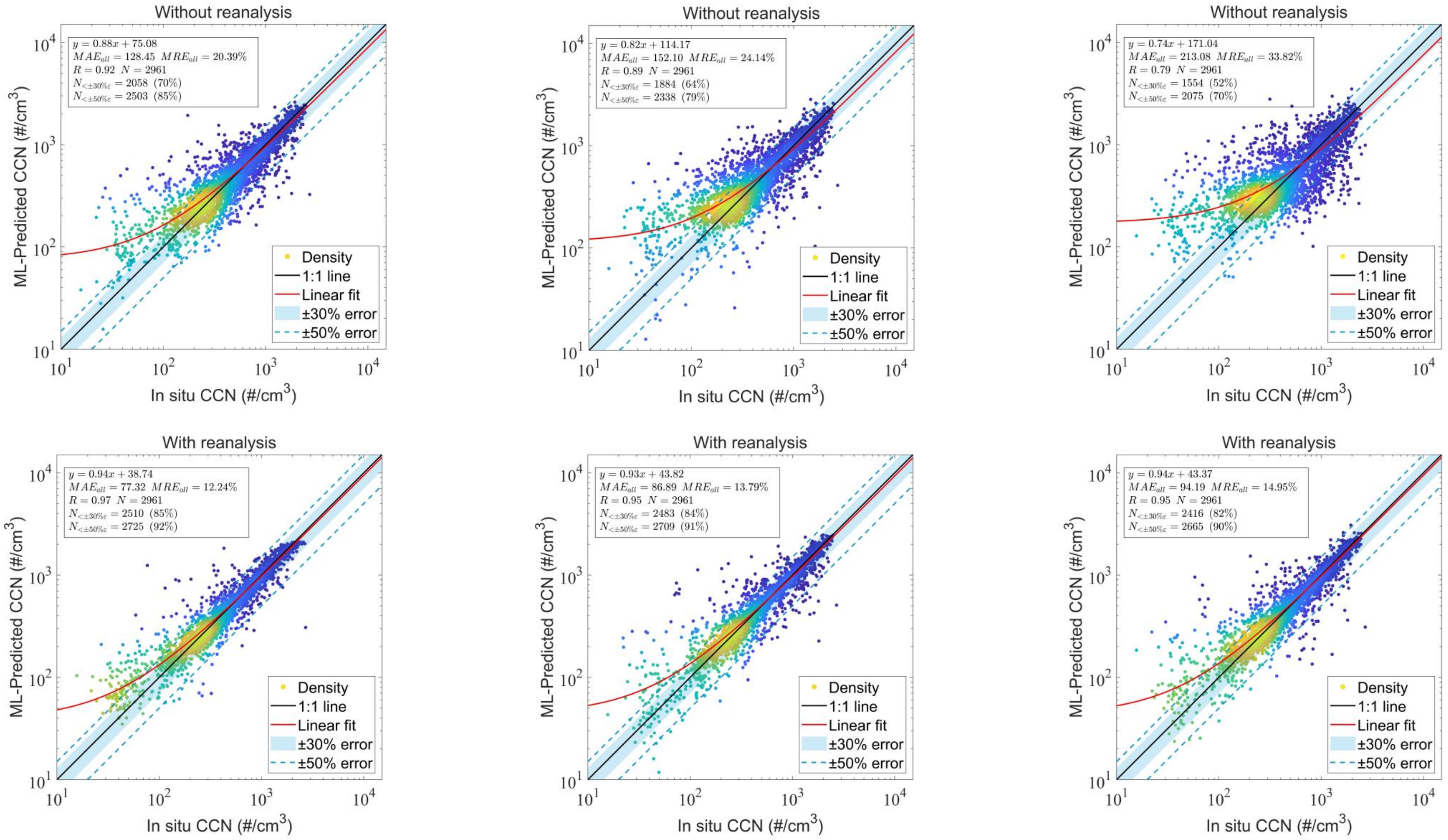




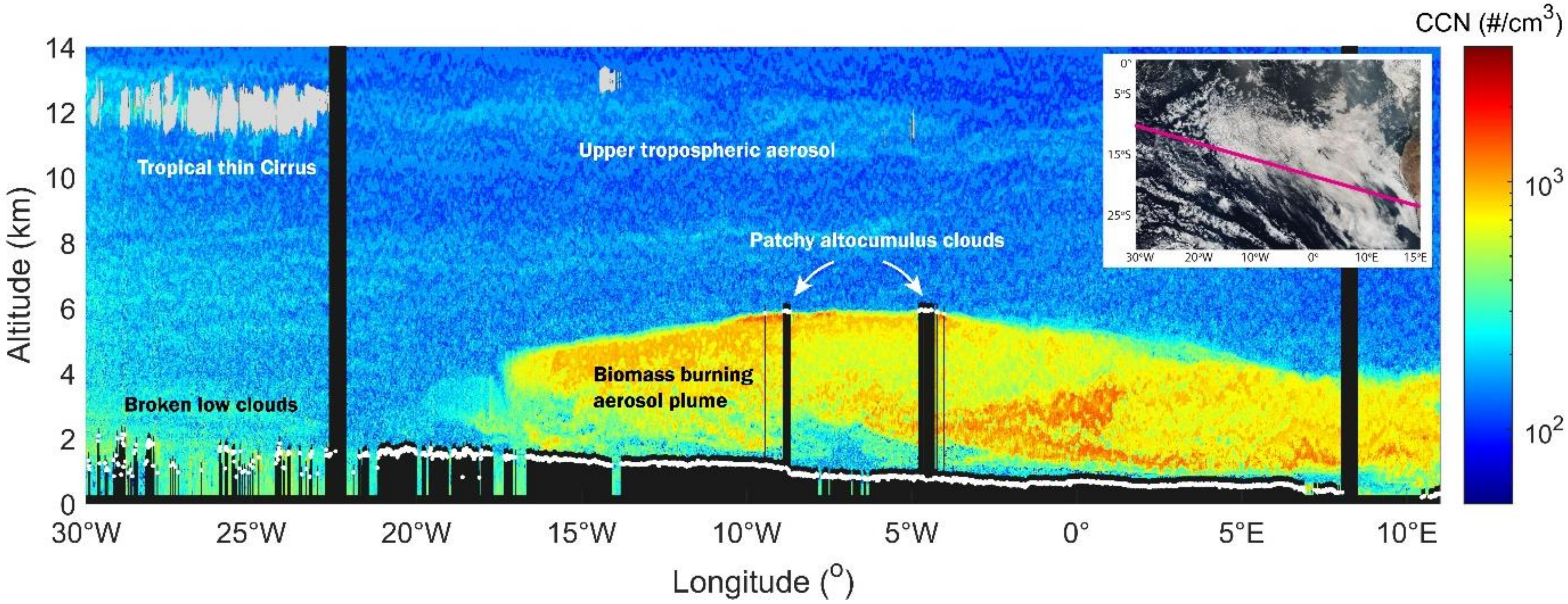
Grey-shaded areas excluded in training

CCN for HSRL-2, HSRL-1 and EarthCARE observables

3β + 2α + 3δ | 2β + 1α + 2δ | 1β + 1α + 1δ



CCN predicted by ML-model for ER-2 flight across Southeast Atlantic - Aug. 26, 2016

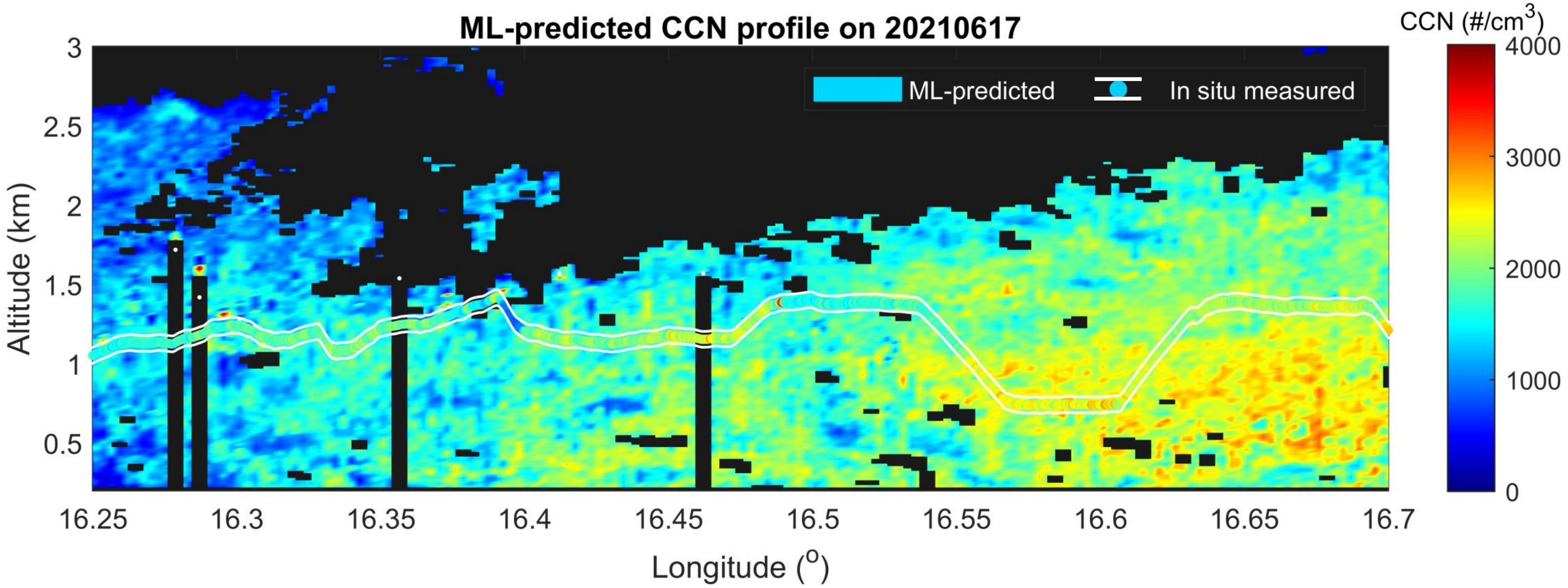


Mean Absolute (Relative) Error of CCN and ABS predictions for all and pristine conditions

Predictor Data set →	HSRL-2 observables		HSRL-2 observables + Reanalysis Data	
Predictor Indicator →	Mean Absolute Error (Relative)			
	Full range	Pristine 0<CCN<100	Full range	Pristine 0<CCN<100
CCN [1/cm^3]	128.45 (20.39%)	101.11 (181.03%)	77.32 (12.24%)	55.48 (100.09%)

Potentially deliverable product

- CCN profile at lidar product grid when lidar observables are available.



Summary

- We successfully trained machine learning models to perform CCN predictions using HSRL-2 lidar data from four airborne campaigns.
- Our algorithm shows its capability to predict vertically-resolved CCN concentration with an approximate relative uncertainty of 20% with lidar observables, and uncertainty can be further reduced by adding reanalysis constraints.
- For pristine conditions, CCN retrieval errors are much higher (partly due to sparse training data).
- We expect that the machine learning-based algorithm can substantially enhance the campaign data product and benefit future satellite mission.

Thank you!

- The computation in this paper was performed at the Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU).