Using Machine Learning to Reconstruct Historical Surface $PM_{2.5}$ in Summer Alaska

Zhiwei Dong[1], Jingqiu Mao[1], Alqamah Sayeed[2], Pawan Gupta[2], Tianlang Zhao[1]
[1]University of Alaska Fairbanks
[2]Universities Space Research Association (USRA), NASA
Contact: zdong@alaska.edu

Introduction

Wildfires increase in sizes and frequencies in Alaska (Chapin et al., 2008, Woo et al., 2020)

Smoke from the wildfires endangers public's health (Hahn et al., 2021)

Lack of surface PM2.5 measurement data is a major challenge

Goal: We want to build good historical daily surface PM2.5 dataset for summer Alaska

Datasets

Time Period	Usage of the data
2021.05.01-2021.08.31	Model Training and Testing
2022.05.01-2022.08.31	Model Training and Testing
2019.06.01-2019.07.31	Independent Validation
2023.07.24-2023.08.10	Independent Validation

Table 1: Time Periods and Usage of Data

Predictor Variables

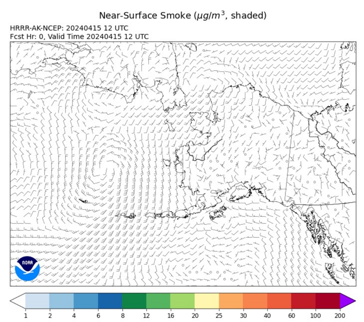
Smoke Variables from High-Resolution Rapid Refresh-Smoke (**HRRR-Smoke**) forecasts

Spatial Resolution: **3km x 3km**

Coverage for Alaska started since Nov 2020

Emissions: Parameterized based upon the measured MODIS/VIIRS FRP

Meteorology Variables from HRRR forecasts



HRRR Smoke Alaska Coverage Area

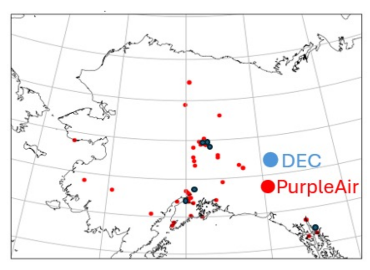
Target Variable

PM2.5 from **PurpleAir**, a low-cost air quality sensor network

120 PurpleAir sensors in Alaska as of year 2022

Calibration Equation according to Barkjohn et al., 2021

PM2.5 = 0.524*PurpleAir_cf1-0.0862*RH+5.75

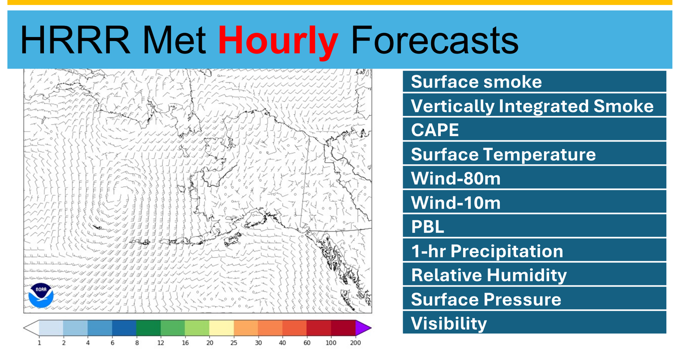


PurpleAir and DEC sensors in Alaska, 2022

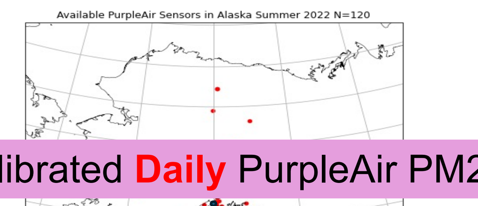
Develop Machine Learning Methods

HRRR-Smoke **Hourly** Forecasts

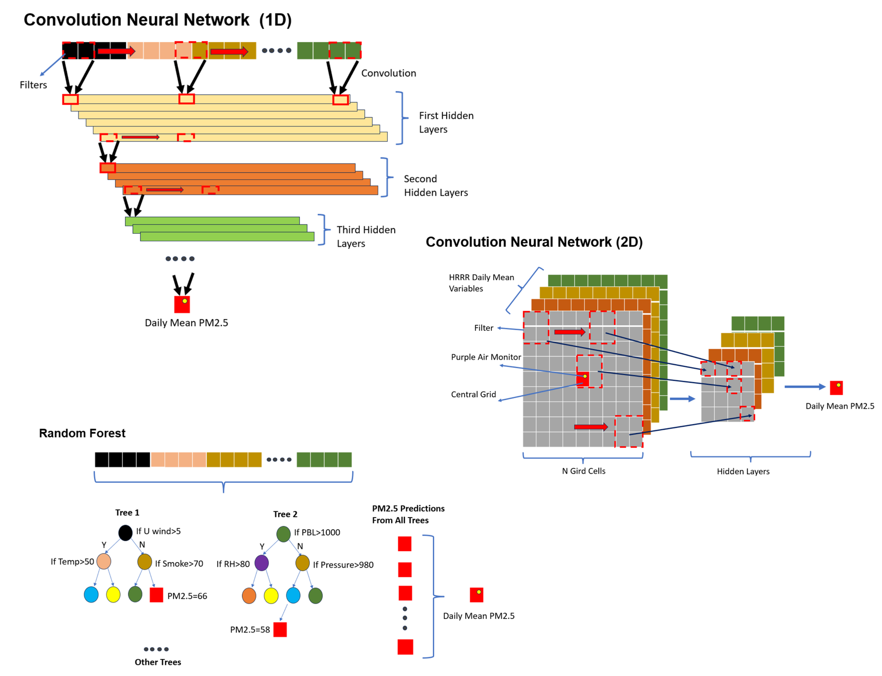
HRRR Met **Hourly** Forecasts



Step1: Use machine learning to build relationship/model between collocated HRRR variables and PurpleAir PM2.5



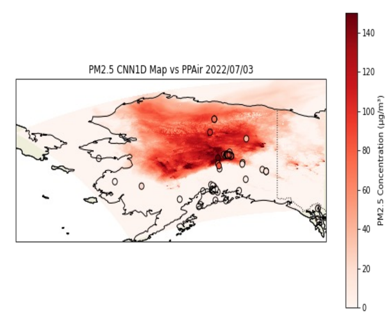
Calibrated **Daily** PurpleAir PM2.5



Step2: Apply the trained Machine Learning model to the entire domain to generate the PM2.5 map

Popular Machine Learning Models:

Random Forest (RF)
Convolutional Neural Network (CNN)

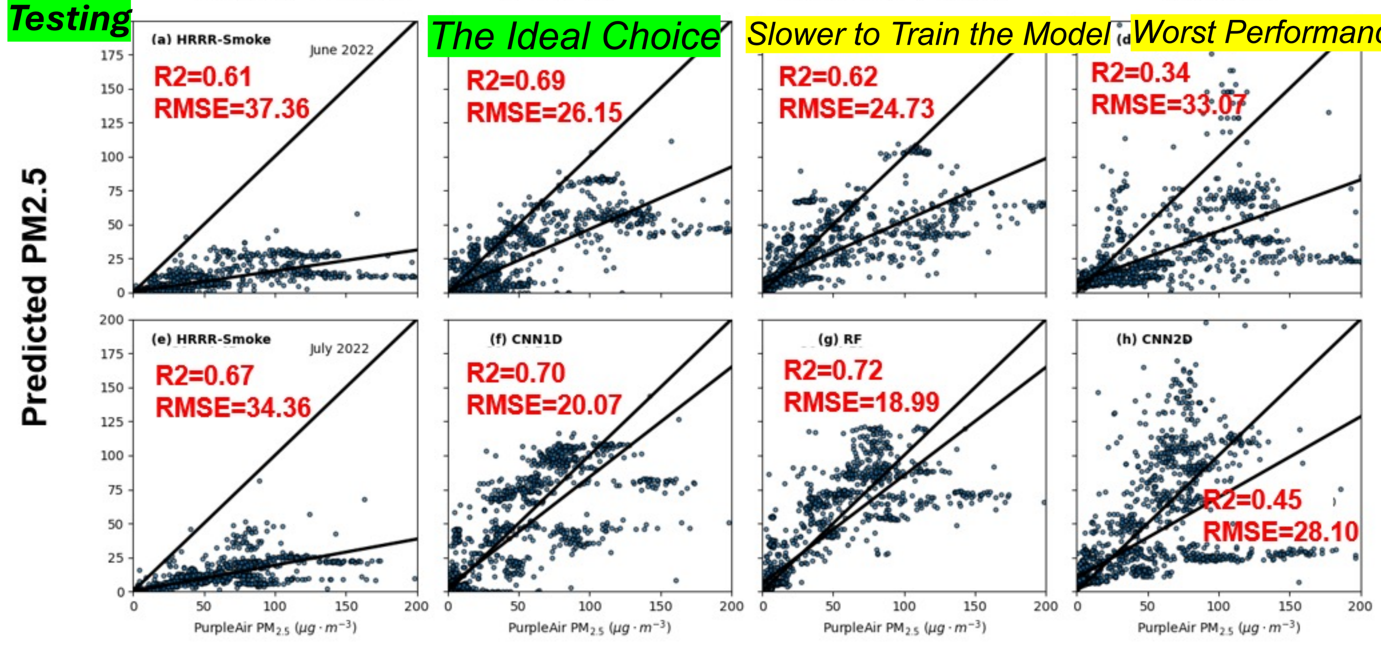


Step3: Post-process with Satellite **Daily** AOD

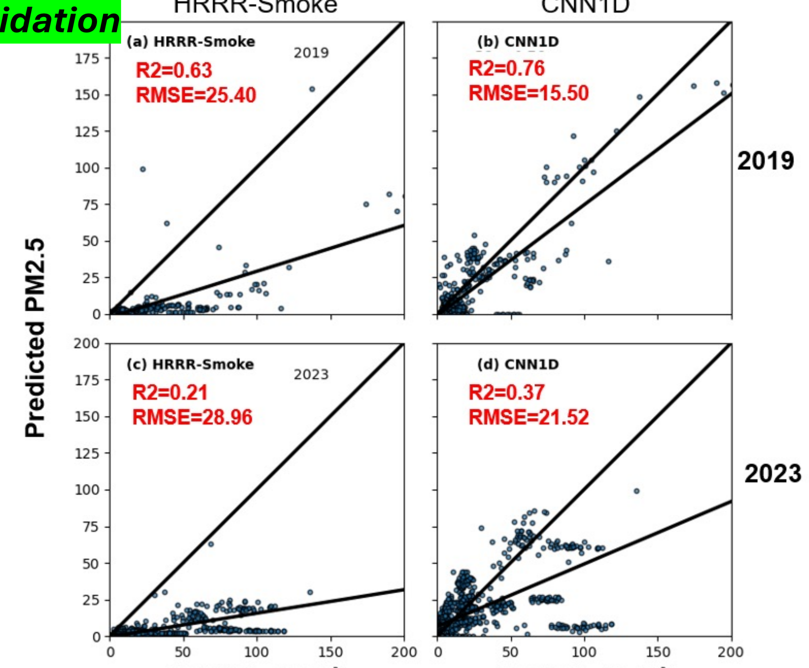
Historical **Daily** Surface PM2.5 Maps for Alaska

Results

Step 1 Machine Learning Model Build-up

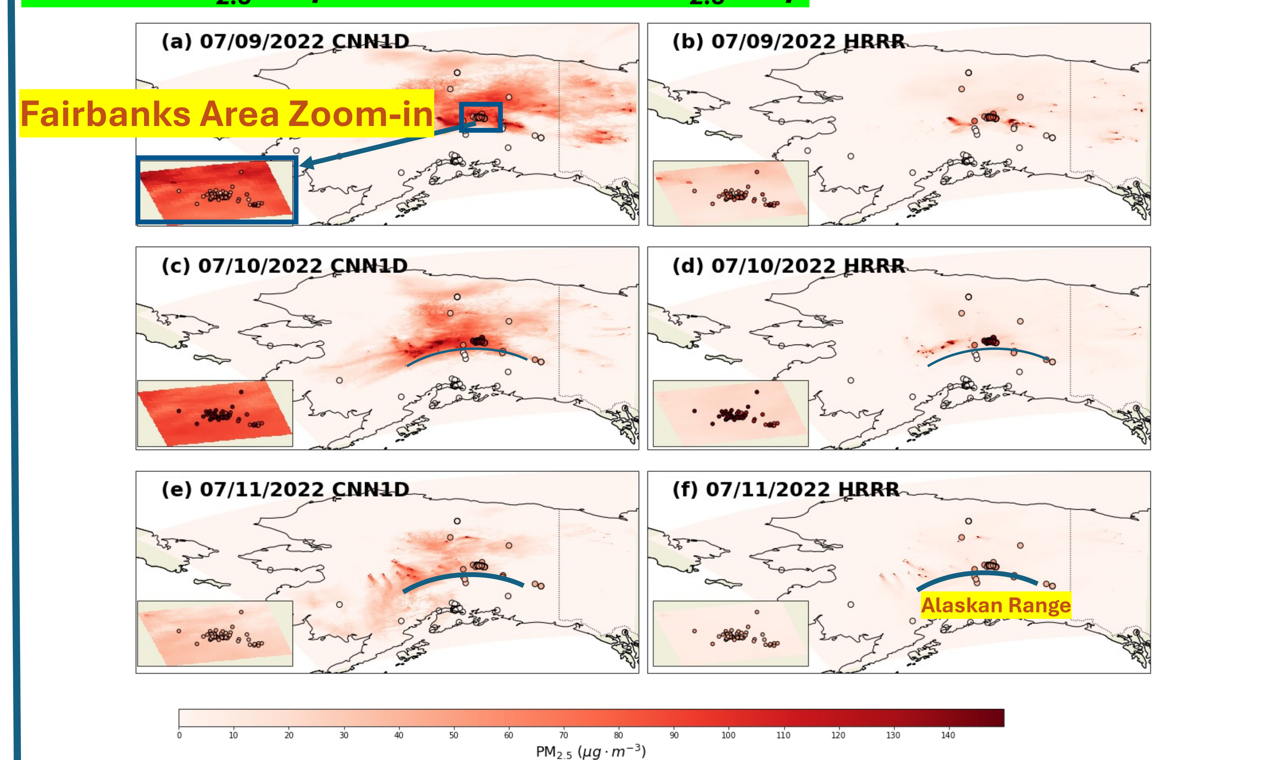


- CNN1D and Random Forest improved the original HRRR-Smoke $PM_{2.5}$
- CNN2D did not improve possibly due to redundant information from finer resolution
- CNN1D is much quicker to train a model than Random Forest



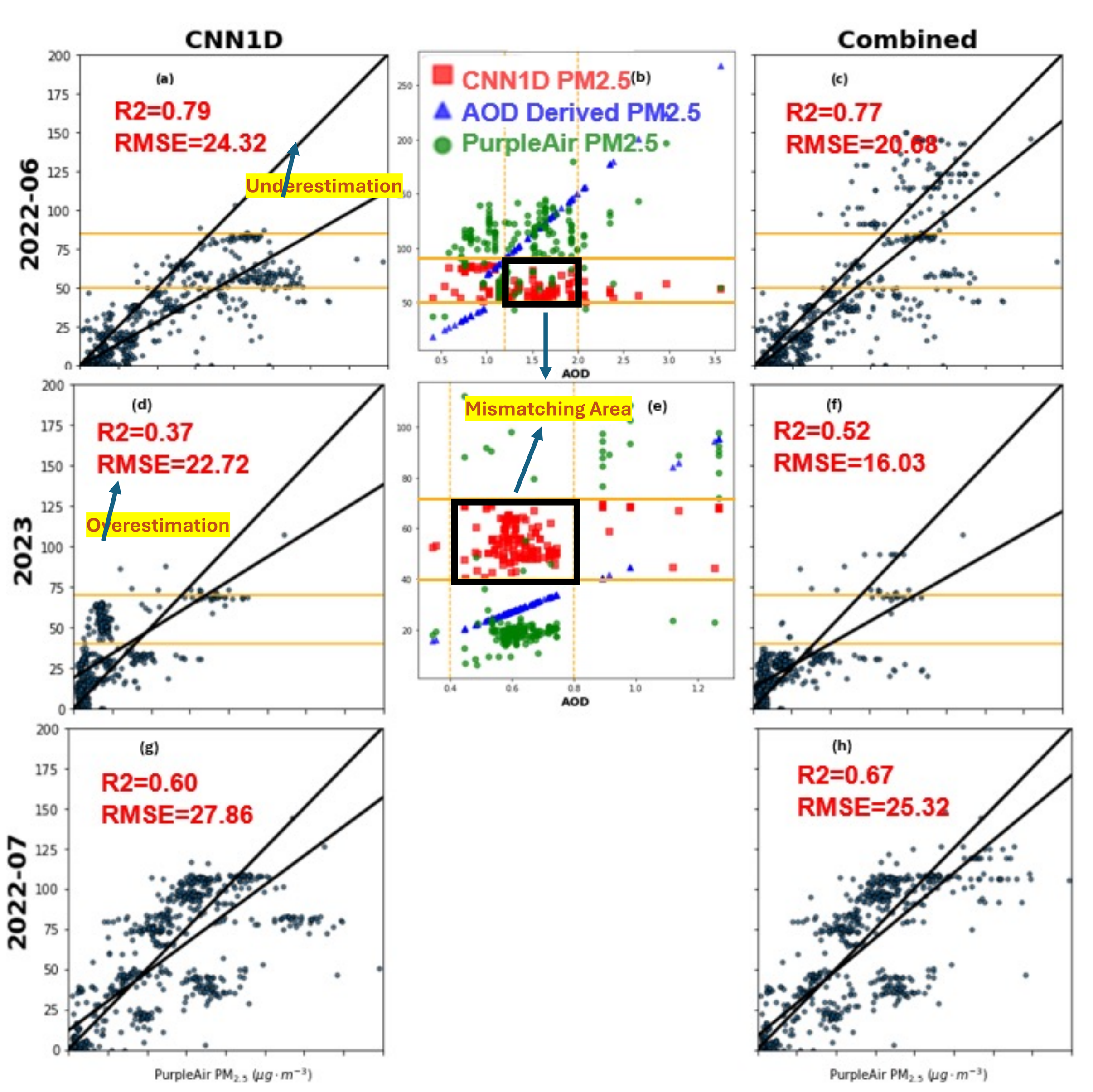
- CNN1D model independently validated using 2019 and 2023 data
- CNN1D model proved to improve HRRR-Smoke $PM_{2.5}$, though with some over- and underestimations

Step 2 Apply CNN1D to the Entire Alaska Domain



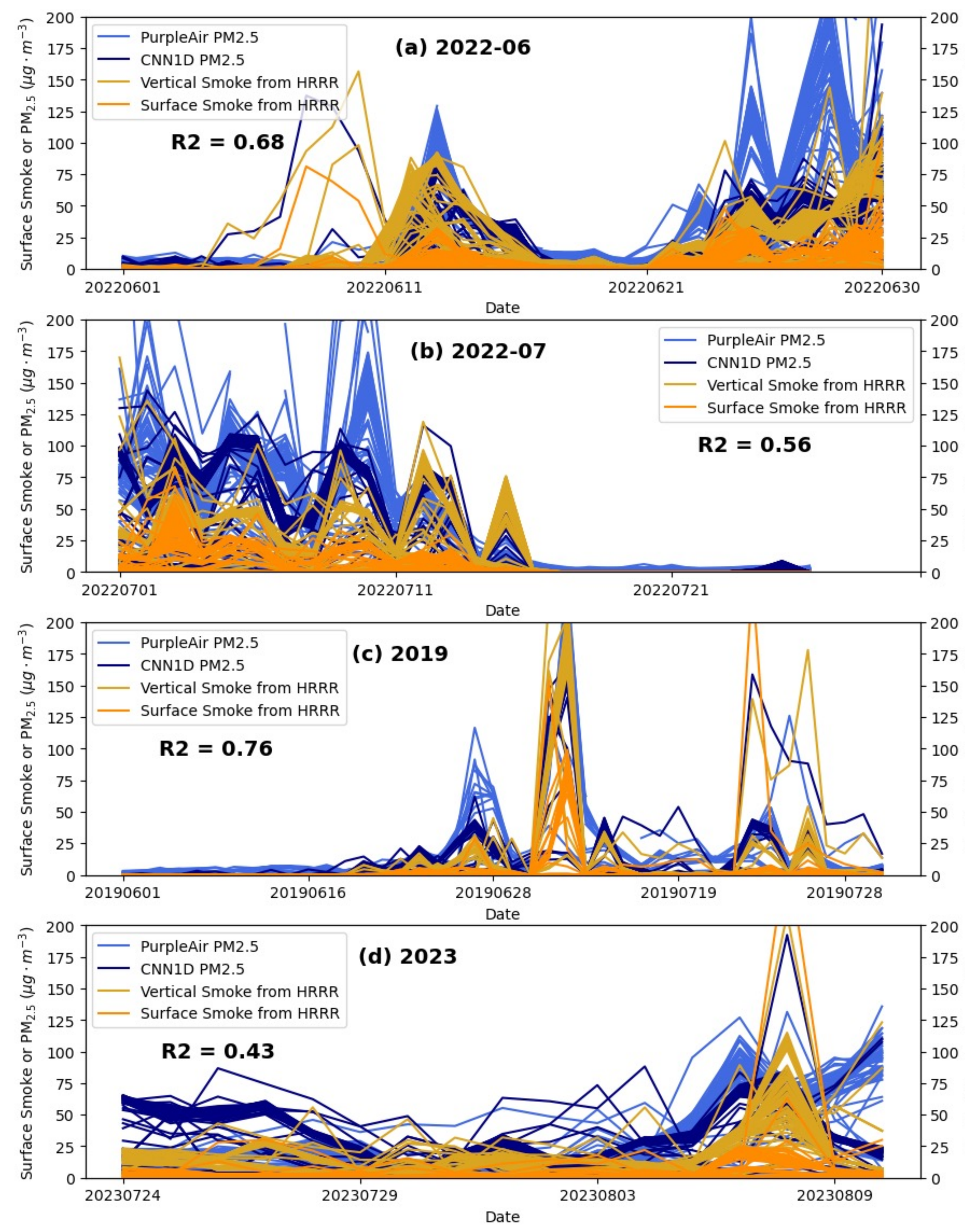
- CNN1D $PM_{2.5}$ map shows that wildfire smoke can transport much further than HRRR-Smoke demonstrates.
- CNN1D $PM_{2.5}$ map has much more details than HRRR-Smoke $PM_{2.5}$, such as the blockade effect of Alaska Range

Step 3 Satellite AOD Post-Process

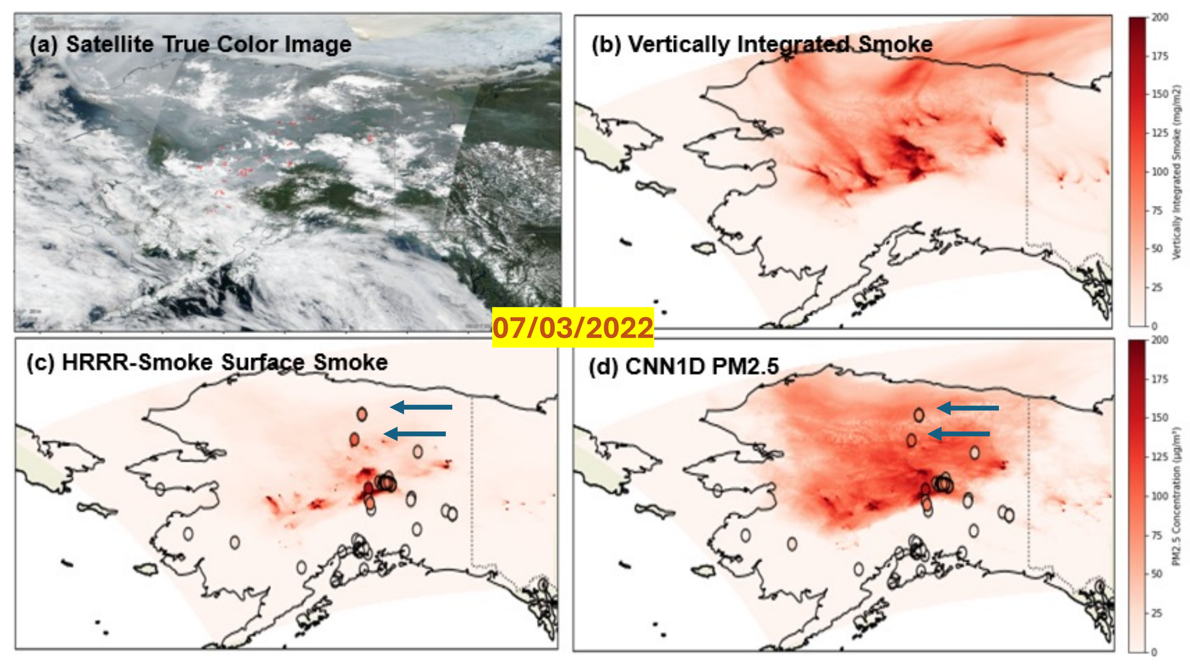


- Using Tianlang Zhao's AOD-PM2.5 deriving functions, we can get the AOD-derived $PM_{2.5}$ (Zhao et al., 2024)
- At certain mismatching areas, where CNN1D $PM_{2.5}$ **underestimates** but AOD is relatively **high**, or CNN1D $PM_{2.5}$ **overestimates** but AOD is relatively **low**, the AOD-derived $PM_{2.5}$ agrees better with PurpleAir $PM_{2.5}$. We replace CNN1D $PM_{2.5}$ in these mismatching areas with AOD-derived $PM_{2.5}$ to generate the combined PM2.5 and it further improved the results.

CNN1D Performances Analysis



- Importance rank shows that the **most important predictors** are **Surface Smoke** and **Vertically Integrated Smoke** from HRRR-Smoke.
- When Surface Smoke is low while PurpleAir PM2.5 is higher, Vertically Integrated Smoke is usually high, possibly due to misplacement of smoke in height in HRRR-Smoke model.
- CNN1D learns such relationship and redistributes some smoke aloft to surface to increase and improve the surface PM2.5 results.



- Satellite True Color Image agrees with the Vertically Integrated Smoke, confirming the long-distance transportation of smoke from wildfire spots. (Seen in the Northern Alaska)
- CNN1D PM2.5 combines Surface Smoke and Vertically Integrated Smoke and gives better surface PM2.5 results.
- Satellite images also show that there are misdetections in Fire Radiative Power (FRP) due to heavy smoke and clouds, which HRRR-Smoke uses as emission. This also contributes to the underestimation of Surface Smoke from HRRR-Smoke model. (Not shown)

Conclusions

- CNN1D can help to reconstruct high resolution historical surface PM2.5 data in Alaska using HRRR-Smoke forecast and PurpleAir PM2.5 measurements
- HRRR-Smoke underestimates surface PM2.5 in Alaska due to missing FRP detection and misdistributions of Surface and Vertical Smoke
- CNN1D can help improving the Surface/Vertical Smoke issue but can not cope with missing FRP

References:
Chapin, F. S., Trainor, S. F., Huntington, O., Lovecraft, A. L., Zavaleta, E., Natcher, D. C., McGuire, A. D., Nelson, J. L., Ray, L., Calef, M., Fresco, N., Huntington, H., Rupp, T. S., DeWilde, L., and Naylor, R. L.: Increasing Wildfire in Alaska's Boreal Forest: Pathways to Potential Solutions of a Wicked Problem, BioScience, 58, 531–540, https://doi.org/10.1641/B580609, 2008.
Woo, S. H. L., Liu, J. C., Yue, X., Mickley, L. J., and Bell, M. L.: Air pollution from wildfires and human health vulnerability in Alaskan communities under climate change, Environ. Res. Lett., 15, 094019, https://doi.org/10.1088/1748-9326/ab9270, 2020.
Hahn, M. B., Kuiper, G., O'Dell, K., Fischer, E. V., and Magzamen, S.: Wildfire Smoke Is Associated With an Increased Risk of Cardiorespiratory Emergency Department Visits in Alaska, GeoHealth, 5, e2020GH000349, https://doi.org/10.1029/2020GH000349, 2021.
Barkjohn, K. K., Gantt, B., and Clements, A. L.: Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor, Atmospheric Meas. Tech., 14, 4617–4637, https://doi.org/10.5194/amt-14-4617-2021, 2021