Spatiotemporal Estimates of Surface $PM_{2.5}$ Concentrations in the Western U.S. using NASA Retrievals, Deep Learning, and Data Assimilation Techniques

S. Marcela Loría-Salazar[1], Jeffrey Lee[1], and Heather A. Holmes[2]
[1] School of Meteorology, University of Oklahoma, Norman, Oklahoma, USA; Department of Chemical Engineering, The University of Utah, Salt Lake City, USA

Motivation

Air pollution has been recognized as one of the major concerns for human health and environmental preservation. Outdoor particulate matter (PM) and aerosols are major global causes of death and disease, having been found responsible for 3.2 million deaths per year. Many regions in North America have experienced increased acute exposure due to poor air quality (AQ), suffering the associated consequences (i.e., health, ecological, and economic effects). When AQ exposure information is extrapolated from monitoring networks, the accuracy of exposure models increases in higher population density areas because most monitors are located in urban areas. However, this can lead to poor information in areas with few or no sensors since fidelity is lost the further the point of interest is from monitors. Many satellite-based exposure models provide annual or monthly estimates; modeling acute, daily exposure can be highly biased, as discussed recently at the Virtual Workshop on Health Applications for Satellite-Derived Air Quality by the Health Effects Institute.

Research questions

Q1) How can we implement NASA Deep Blue aerosol retrievals to create daily $PM_{2.5}$ acute exposure estimates?

Q2) What is the associated change in surface $PM_{2.5}$ concentrations that impacts communities due to smoke, urban emissions, dust, etc.?

Our research goals are integrated into an overarching aim to improve acute daily exposure estimates of $PM_{2.5}$.

Research Goals

RG1) Create a **gap-filled spatial dataset of aerosol optical depth (AOD) from NASA heritage Deep Blue** aerosol retrievals from Terra/Aqua (MODIS) and Suomi-NPP (VIIRS) using Machine Learning (ML) UNet3+ architecture.

RG2) Utilize state-of-the-art **atmospheric models to simulate $PM_{2.5}$ concentrations** and estimate the $PM_{2.5}$ sources' contribution (e.g., dust, smoke) to acute elevated $PM_{2.5}$ exposure.

RG3) Create daily high-resolution **exposure estimates of $PM_{2.5}$** using gap-filled AOD **(RG1)**, atmospheric models **(RG2),** and statistical data fusion techniques.

Methods

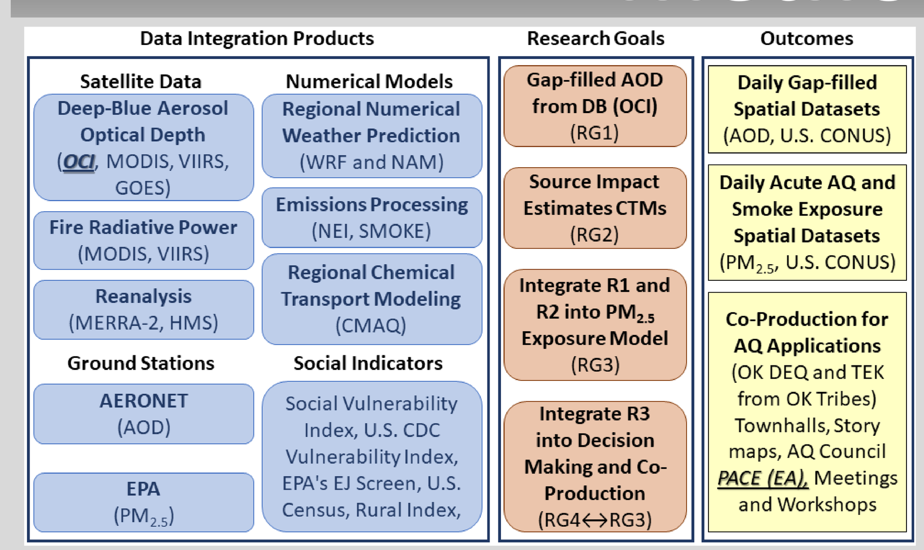


Diagram with the data we will implement. **RG1** intends to fill the AOD gaps in the Deep Blue (DB) algorithm. **RG2** will implement a chemical transport model (CTM) to estimate $PM_{2.5}$ source impacts. **RG1-2** will be spatial predictors and covariates for $PM_{2.5}$ (PM with an aerodynamic diameter smaller than 2.5 µm) in **RG3**.

Summary

Smoke

1. $PM_{2.5}$ from the data fusion model with **UNet 3+ AOD Gap-Filled** input (R~0.65) can capture **local and long-range transport smoke.**
2. $PM_{2.5}$ from the data fusion model with **CMAQ $PM_{2.5}$** input (R~0.65) is not able to recreate up/downwind transport

Temperature Inversions

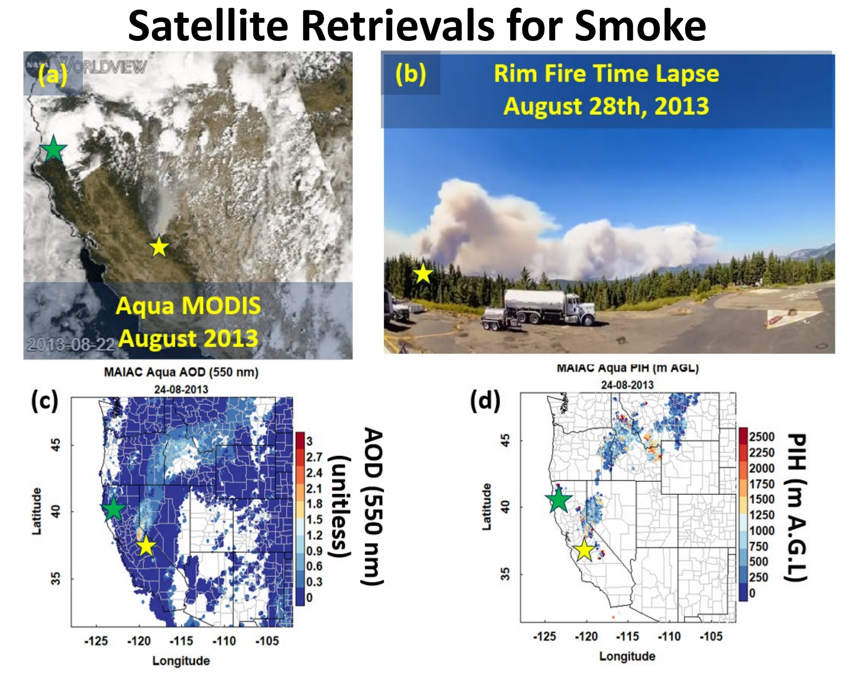
1. $PM_{2.5}$ from the data fusion model with **UNet 3+ AOD Gap-Filled** input (R~0.65) can capture **temperature inversions in California.** However, satellite AOD presents significant limitations in estimating $PM_{2.5}$ during winter in the north-western U.S. (WA, OR, NV, UT, ID, WY, MT, CO) due to cloud cover and snowpack.
2. $PM_{2.5}$ from the data fusion model with **CMAQ $PM_{2.5}$** input (R~0.65) can reproduce the effects on air quality due to temperature inversions despite cloud cover and snowpack challenges.

Acknowledgements

The MODIS, VIIRS, and MERRA-2 (https://earthdata.nasa.gov) data used in this study are freely available from NASA. The NAM (https://www.ncei.noaa.gov/products/weather-climate-models/north-american-mesoscale) data used in this study are freely available from NOAA. Computational resources were provided by the Center for High-Performance Computing (CHPC) at the University of Utah.

*mloria@ou.edu

Challenges of Estimating Acute $PM_{2.5}$ Exposure Using Satellite Retrievals for Smoke



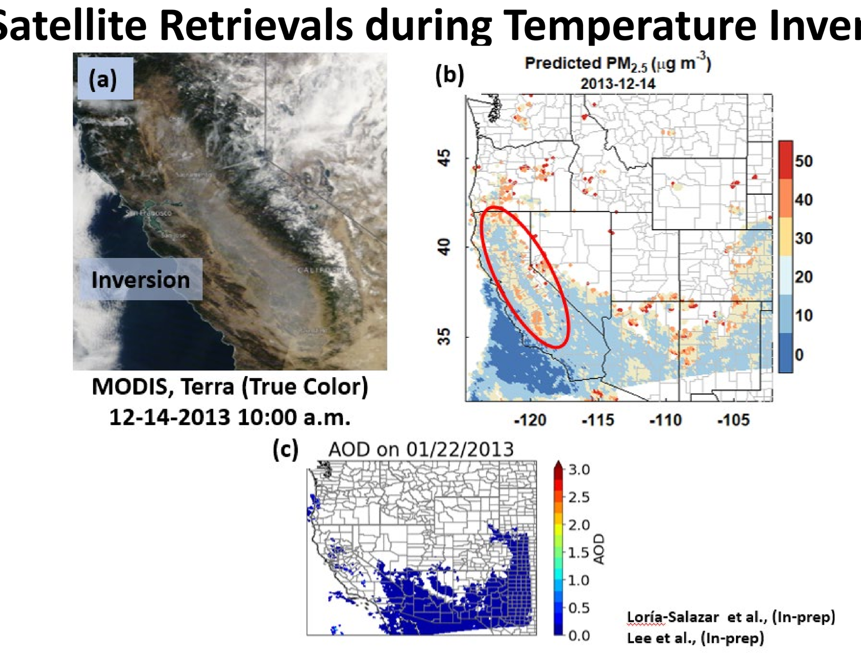
Advantages

1) **Improved smoke detection**, and higher horizontal resolution (a and c).
2) Plume injection height **(PIH)** products provide a **smoke vertical top** estimate (d).

Challenges

1) **Large spatial gaps** due to cloud cover and high surface reflectance (a and c).
2) PIH products still have **high up/downwind biases** (after 100 km) (d).
3) PIH products are **less available** than AOD products or **unfriendly to the user**.

Challenges of Estimating Acute $PM_{2.5}$ Exposure Using Satellite Retrievals during Temperature Inversions



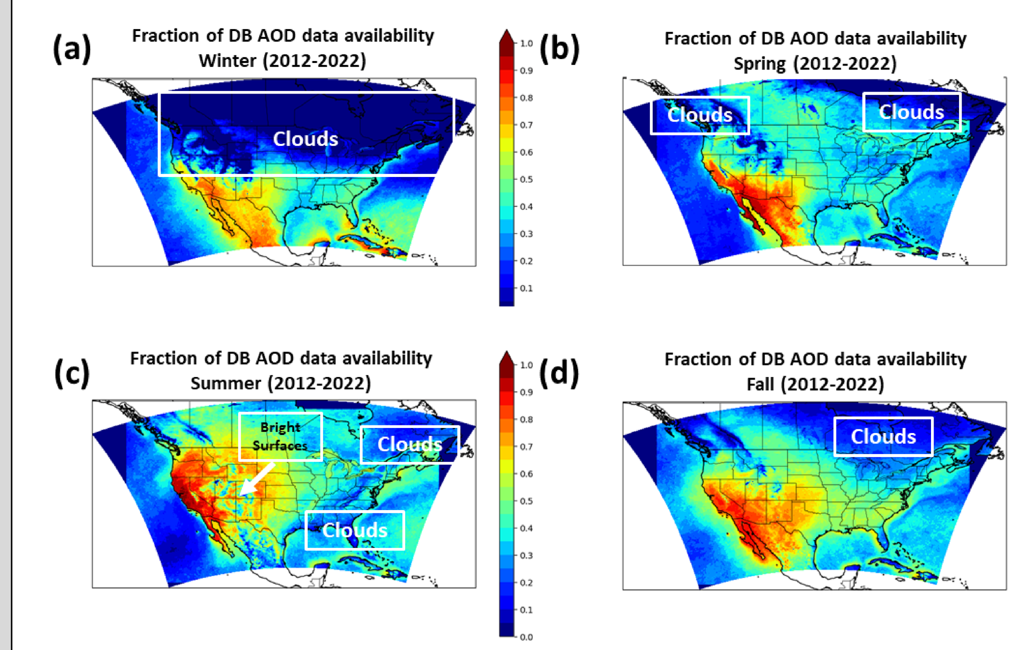
Advantages

1) **Able to retrieve** some temperature inversion **episodes in California** (a and b).

Challenges

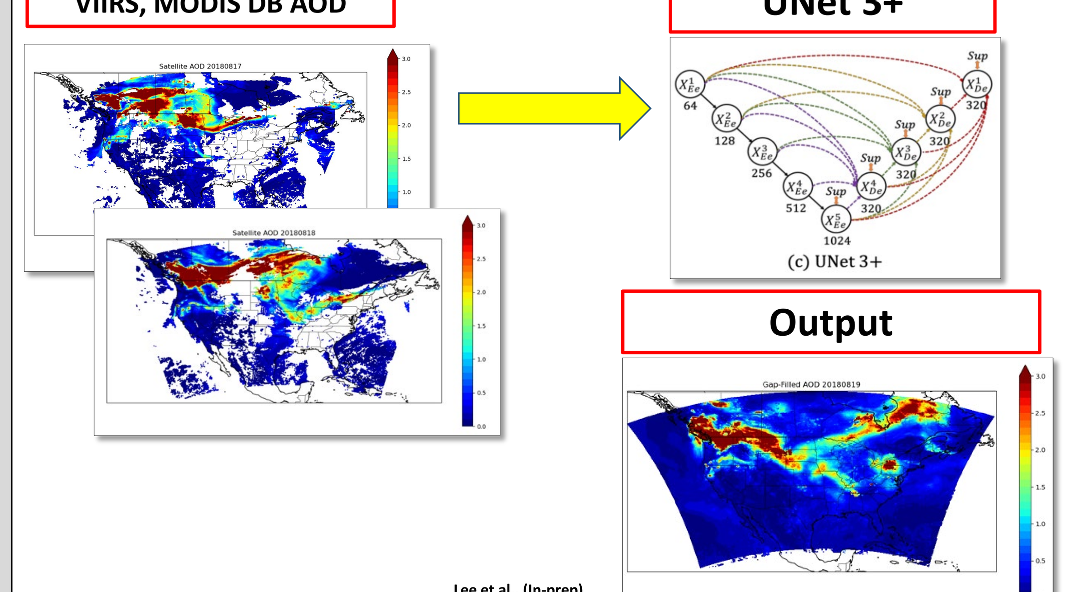
1) **Large spatial gaps** due to cloud cover and snowpack (b and c).
2) **AOD gaps** can be found all winter **(Nov-Apr)** (b and c).
3) States such as UT, NV, OR, WA, ID, WY, MT, and CO often **would not have any AOD retrievals for this period**, limiting AQ risk assessments (c).

Challenges of Estimating Acute $PM_{2.5}$ Exposure Using Satellite Retrievals due to Cloud Cover



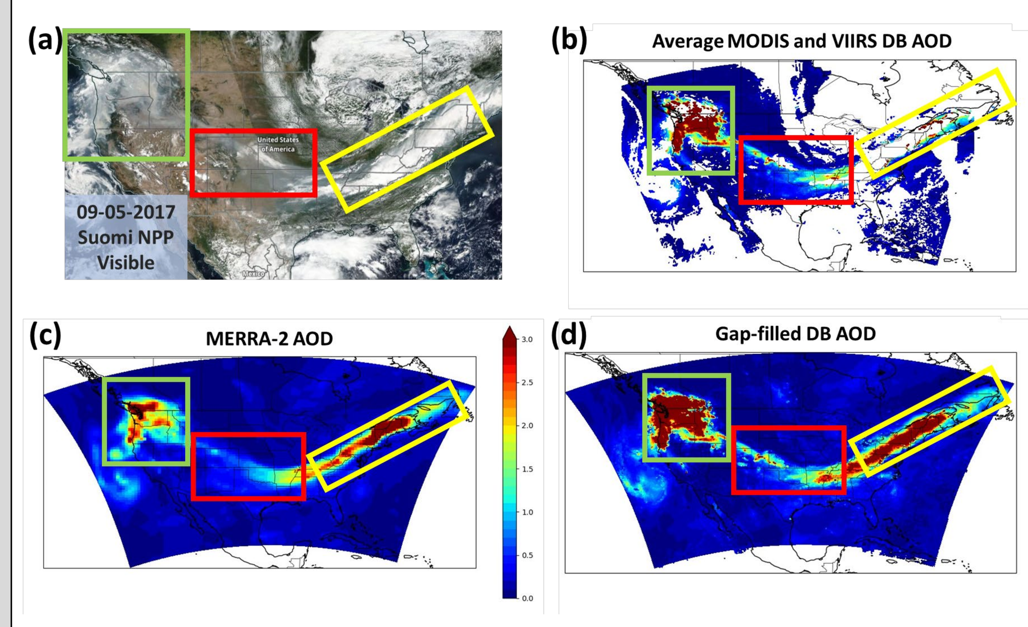
Multiple investigations have shown that tracking aerosol pollution using satellite AOD is challenging because of heterogeneous vertical transport, clouds impeding AOD retrieval algorithms (Fig. a, b, c, and d), bright surfaces (Fig. c), and temperature inversion events.

ML-Derived Gap-filled AOD



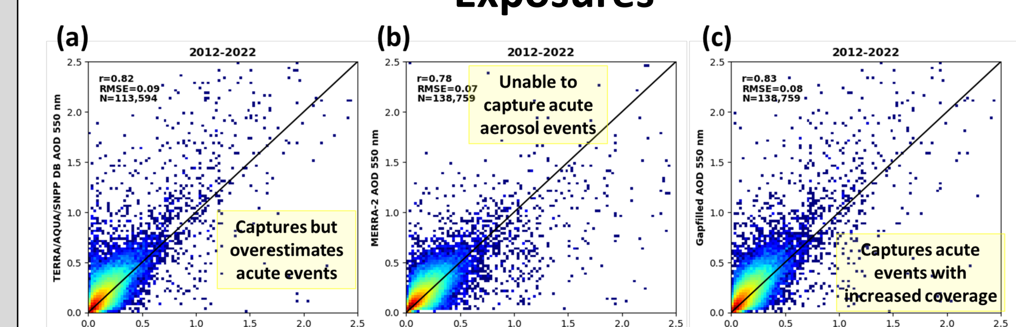
Input Variable	Resolution (km^2)	Lag
NASA MODIS and VIIRS DB Best Estimate AOD	10 X 10	2d, 1d
NASA VIIRS DB aerosol type, land and ocean algorithm flags	6 X 6	2d, 1d
NAM-12km reanalysis – P, T, U, V, orography, vegetation, PBLH	12 X 12	2d, 1d, 0d
NASA MERRA-2 reanalysis AOD	50 X 62.5	2d, 1d, 0d
NOAA HMS Smoke Product	~1 X 1	2d, 1d
NASA MODIS FRP	~1 X 1	2d, 1d

ML-Derived Gap-filled AOD

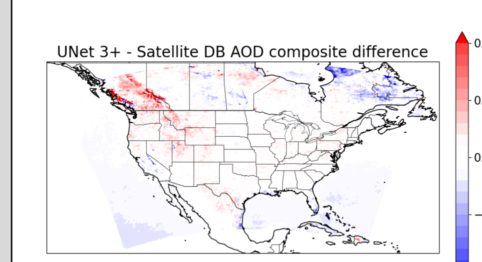


The large smoke plume in the green box is mainly missing from the satellite retrievals in (Fig. b). **It is reproduced by UNet 3+ in (Fig. d). MERRA-2 AOD shows a similar smoke plume shape in a courser resolution but underestimates intensities.** *These underestimations can impact acute exposure assessments of $PM_{2.5}$.* The red boxes show areas of smoke where MERRA-2 vastly underestimated AOD and missed some portions of the smoke plume. The visible image (Fig. a) supports the thickness of the smoke plume. The yellow boxes show aerosol transport in a cloudy region (a-d).

ML-Derived Gap-filled AOD Evaluation for Acute Exposures

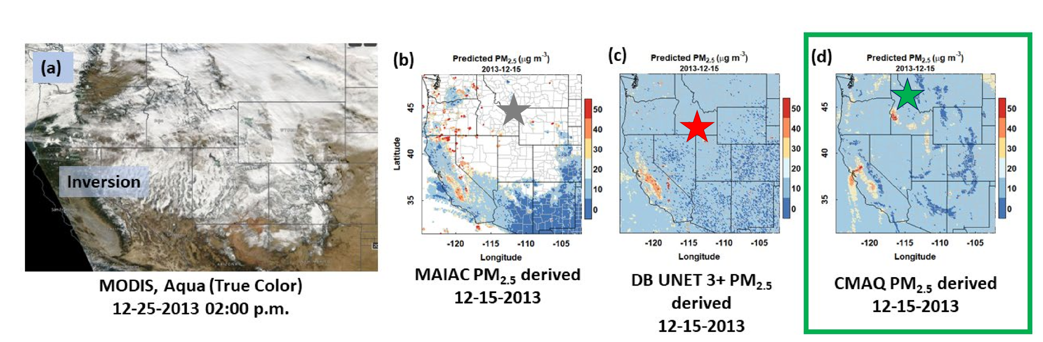


We evaluated our results using 68 AERONET stations across the U.S. The **UNet 3+ DB AOD overall evaluation achieved a correlation R~0.83, root mean square error RMSE~0.08, and normalized mean bias NMB~0.07** with an additional 25,000 paired points than the MODIS/VIIRS DB AOD (113,594 for DB vs. 138,759 for UNet 3+ DB). A significant difference between the UNet 3+ DB model and MERRA-2 is that MERRA-2 underestimates AOD (Fig. b) and cannot capture acute aerosol events for $PM_{2.5}$ exposure assessments.



Mean composite difference between DB AOD and UNet 3+ DB AOD (2012-2022).

Modeling $PM_{2.5}$ Exposure Using Satellite Retrievals during Temperature Inversions with Statistical Data Fusion Techniques (SDF)



(b) DF model with MAIAC AOD

1. R~0.65
2. Able to capture inversions in CA
3. Large spatial gaps

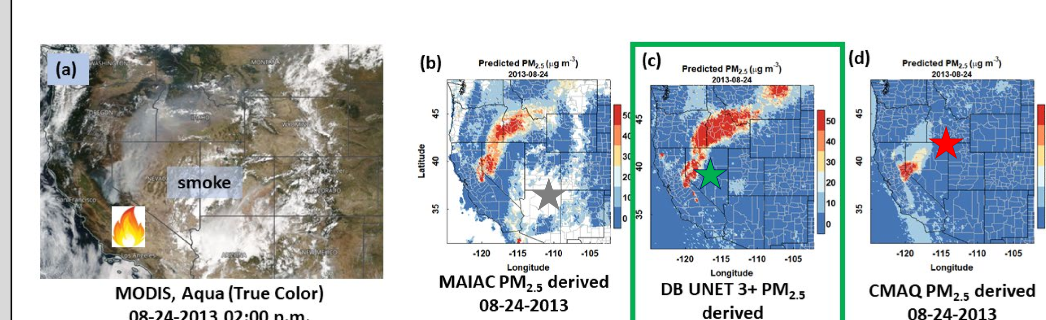
(c) DF model with VIIRS UNET 3+ AOD

1. R~0.65
2. Able to capture inversions in CA
3. No spatial Gaps
4. Issues estimating in northern states' inversions

(d) DF Model with CMAQ

1. R~0.66
2. No spatial gaps
3. Capture Inversions beyond CA
4. Underestimate acute $PM_{2.5}$

Modeling $PM_{2.5}$ Exposure Using Satellite Retrievals from Smoke with Statistical Data Fusion Techniques (SDF)



(b) DF model with MAIAC AOD

1. R~0.65
2. Able to capture smoke plumes
3. Large spatial gaps

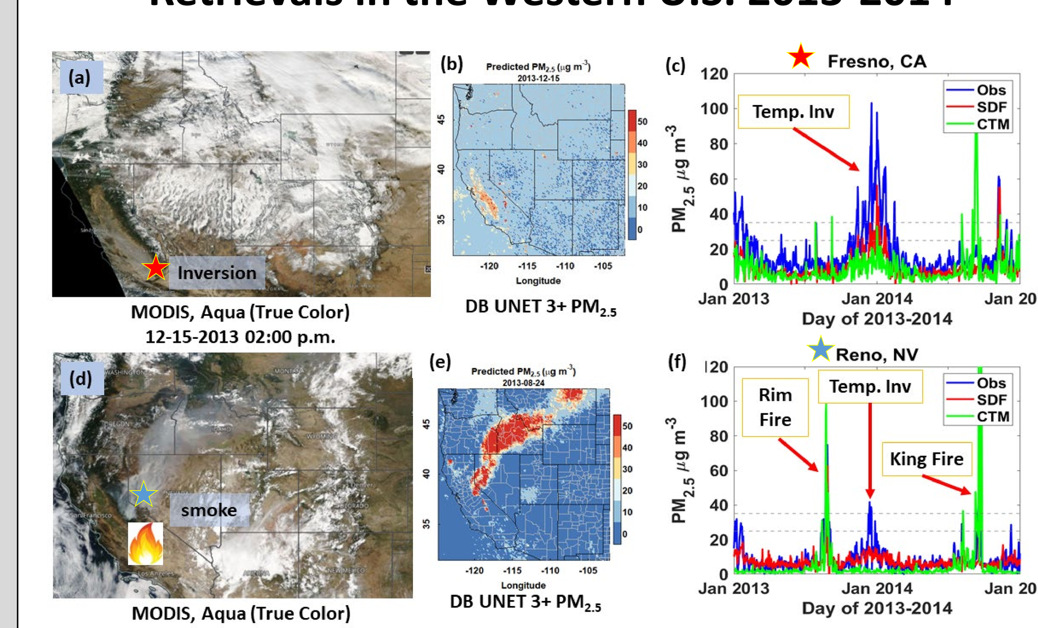
(d) DF Model with CMAQ

1. R~0.66
2. No spatial gaps
3. No able to capture up/downwind smoke

(c) DF model with VIIRS UNET 3+ AOD

1. R~0.65
2. Able to capture local and long-range smoke
3. No spatial gaps

Estimating Acute $PM_{2.5}$ Exposure Using Satellite Retrievals in the Western U.S. 2013-2014



We ran the SDF model from 2013-2014 in the western U.S. with multiple wildfires (e.g., Rim and King) and episodes of strong temperature inversions increasing $PM_{2.5}$ concentrations. We compared CTM outputs and observations. We selected this region because the western U.S. presents atmospheric phenomena that challenge AQ exposure models. **We performed 10-fold cross-validation using 446 EPA stations from UNet 3+ DB AOD as input. Our results show a correlation between an R~0.65 and a RMSE~7.8 µg m⁻³.**