A Machine Learning Approach for Ozone and PM2.5 Forecasting in Pacific Northwest

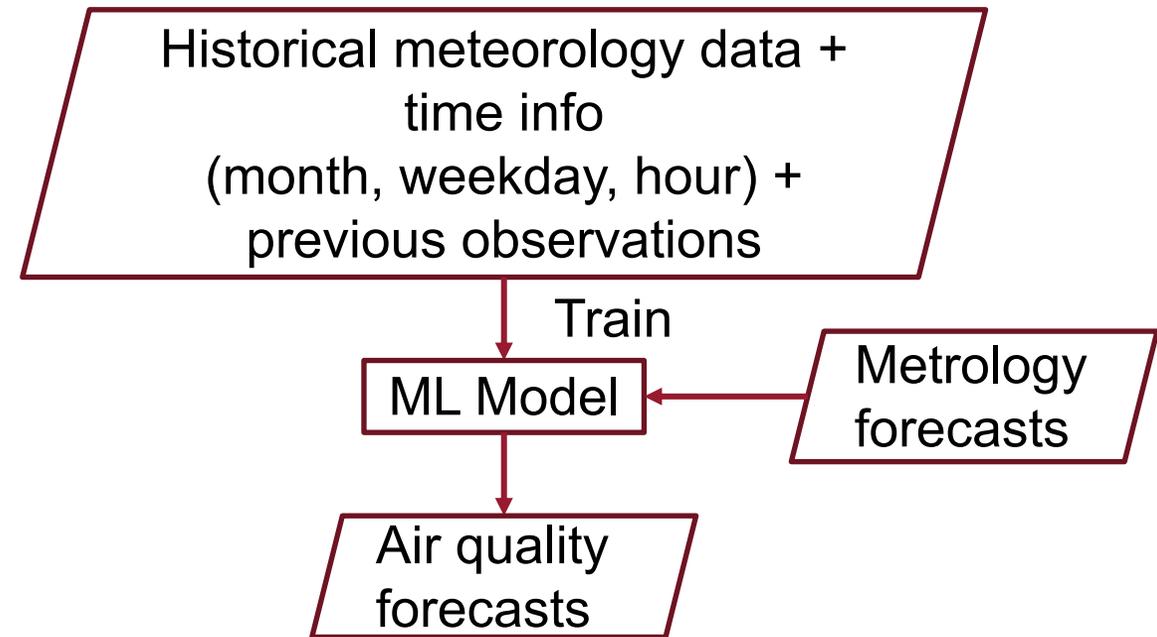Kai Fan[1], Brian Lamb[1], Ranil Dhammapala[2], and Yunha Lee[1]

[1]Laboratory for Atmospheric Research,
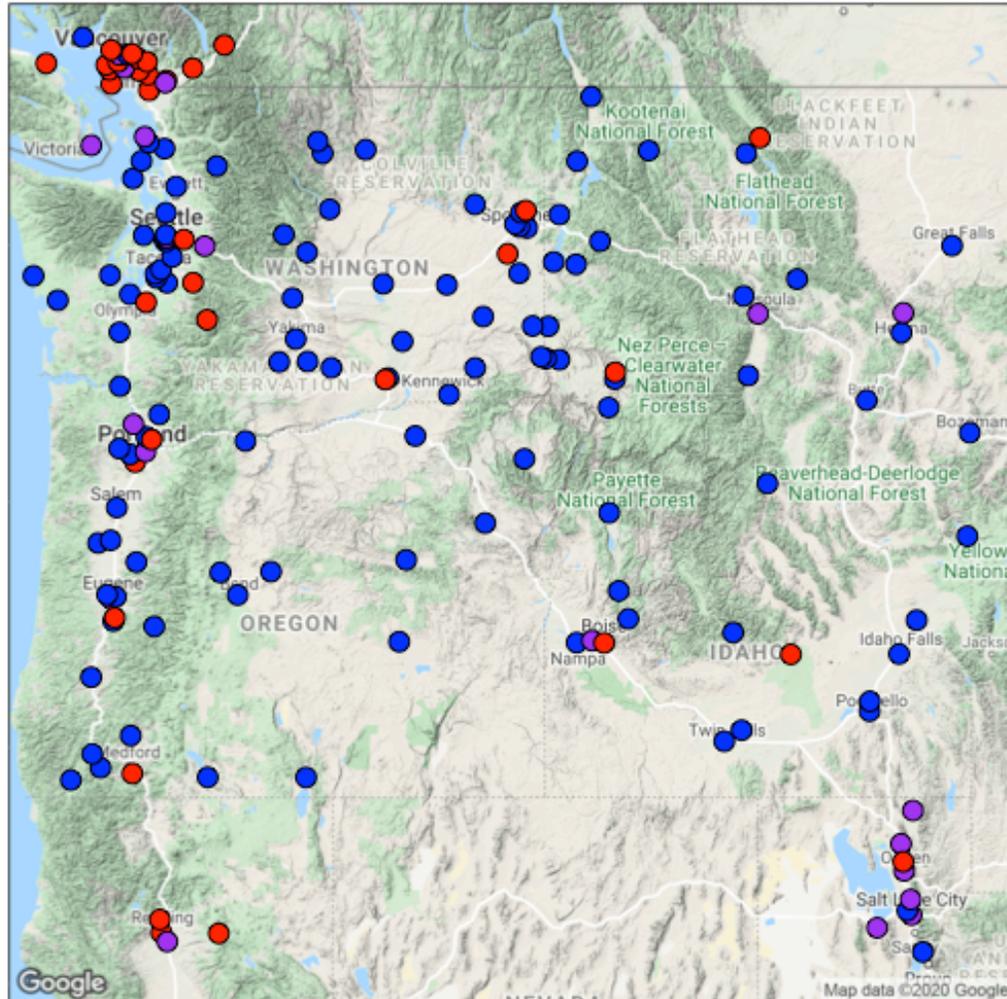Civil and Environmental Engineering, Washington State University
[2]Washington State Department of Ecology

Machine Learning Models

- Machine Learning is an application of artificial intelligence that lets the model learn from historical data and then make future forecasts

- Our approach uses multiple linear regression (MLR) model and random forest (RF) model

- Multiple linear regression fits a line between multiple independent variables and one dependent variable

- Random forest is the average result of many decision trees

Historical meteorology data +
time info
(month, weekday, hour) +
previous observations

Train

ML Model ← Metrology forecasts

Air quality forecasts
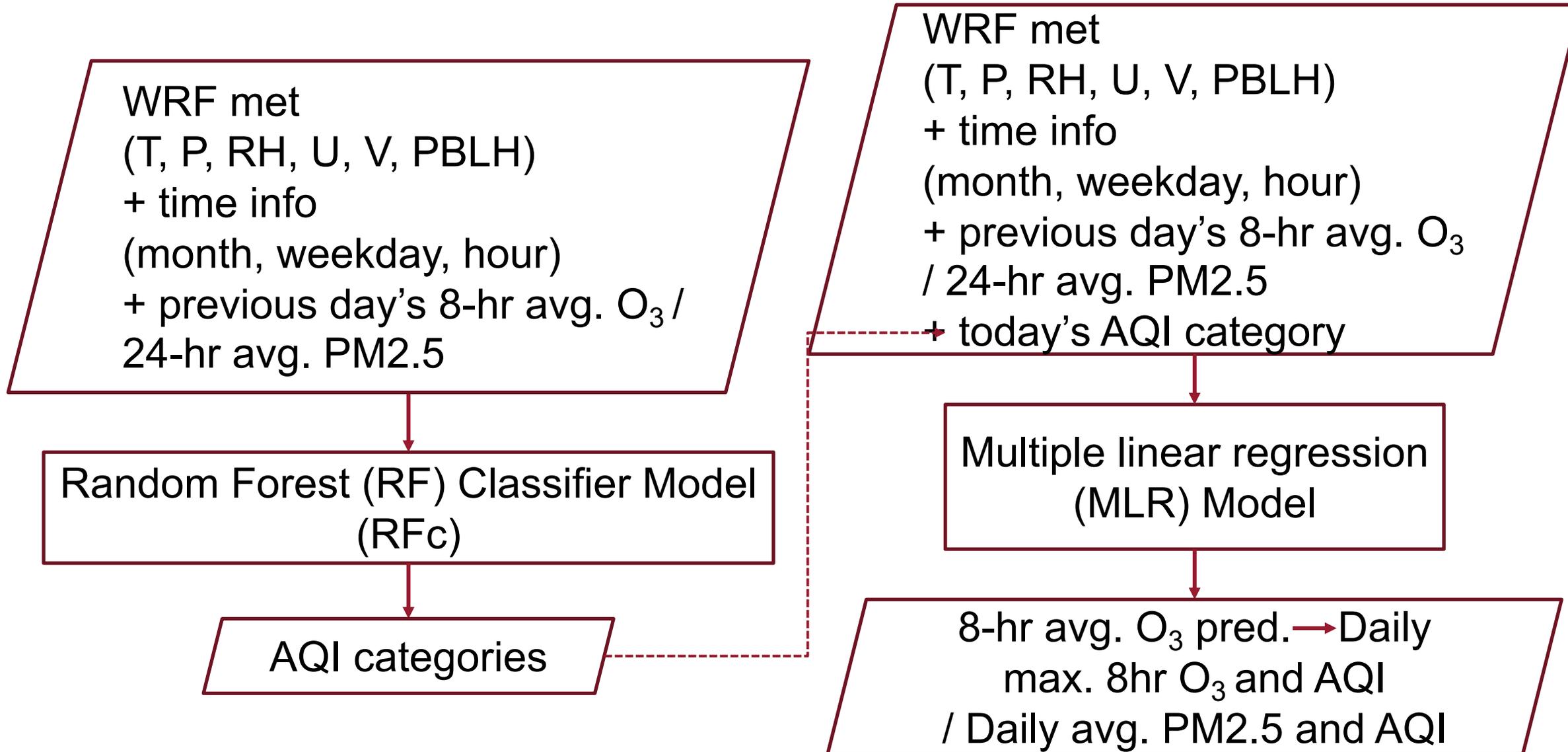
Observation sites in PNW

- 54 O_3 observation sites in 2017-2019

- 139 PM2.5 observations sites
  - 21 sites with both O_3 and PM2.5

- The ML models are trained individually in each site with WRF meteorology and observations.
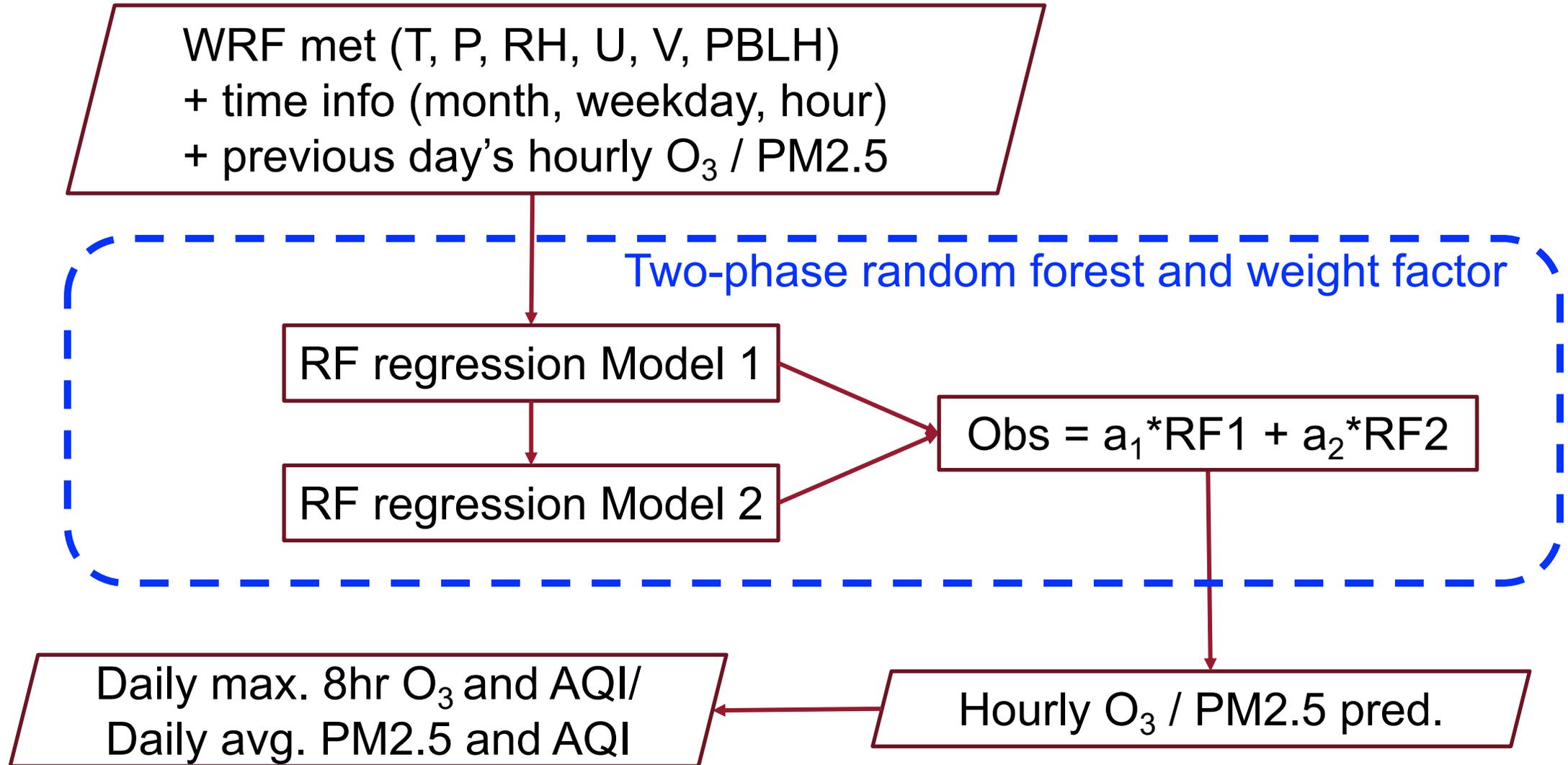
- O3
- PM2.5
- O3 and PM2.5

Machine Learning Model Framework 1: ML1
## Combining Random Forest and Multiple Linear Regression methods

WRF met
(T, P, RH, U, V, PBLH)
+ time info
(month, weekday, hour)
+ previous day's 8-hr avg. O_3 /
24-hr avg. PM2.5

WRF met
(T, P, RH, U, V, PBLH)
+ time info
(month, weekday, hour)
+ previous day's 8-hr avg. O_3
/ 24-hr avg. PM2.5
+ today's AQI category

Random Forest (RF) Classifier Model
(RFc)

Multiple linear regression
(MLR) Model

AQI categories

8-hr avg. O_3 pred. → Daily
max. 8hr O_3 and AQI
/ Daily avg. PM2.5 and AQI

Machine Learning Model Framework 2: ML2
Two RF models weighted for optimal results

WRF met (T, P, RH, U, V, PBLH)
+ time info (month, weekday, hour)
+ previous day's hourly O_3 / PM2.5

Two-phase random forest and weight factor

RF regression Model 1

RF regression Model 2

$Obs = a_1 * RF1 + a_2 * RF2$

Daily max. 8hr O_3 and AQI/
Daily avg. PM2.5 and AQI

Hourly O_3 / PM2.5 pred.

* Jiang, N., & Riley, M. L. (2015). Exploring the utility of the random forest method for forecasting ozone pollution in SYDNEY. *Journal of Environment Protection and Sustainable Development*, *1*(5), 245-254.

Walk-forward cross validation
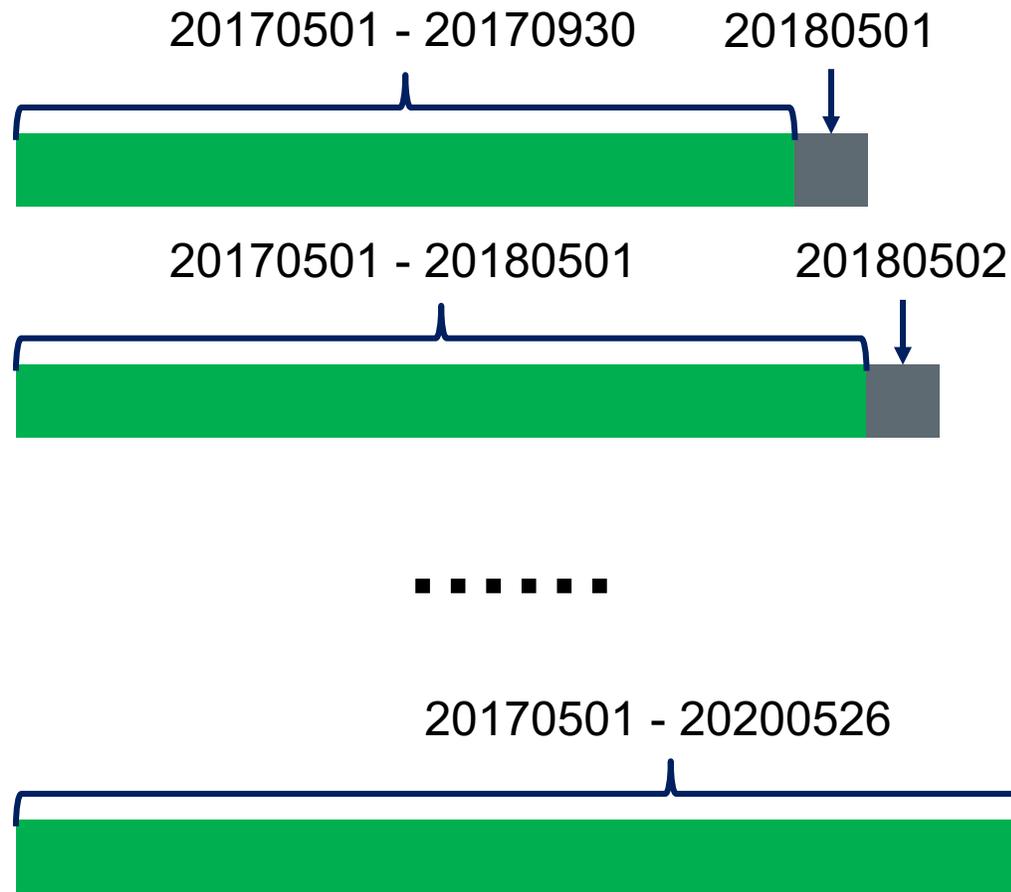
- The training data for O_3 cover May – September

- The training data for PM2.5 cover two period, October – February and May – September.

- The walk-forward cross validation is a tool for time-series data, which is used to evaluate the ML models.

Forecast evaluation parameters

Heidke Skill Score (HSS)

- What is the accuracy of the forecast in predicting the correct category, relative to that of random forecasts?
- Range $-\infty$ to 1
- Perfect score = 1

Hanssen-Kuiper Skill Score (KSS)

- How well does the model separate different categories?
- Range -1 to 1
- Perfect score = 1
- KSS = Hit rate – False alarm rate

O$_3$ evaluation during May - Sep in 2018 - 2020
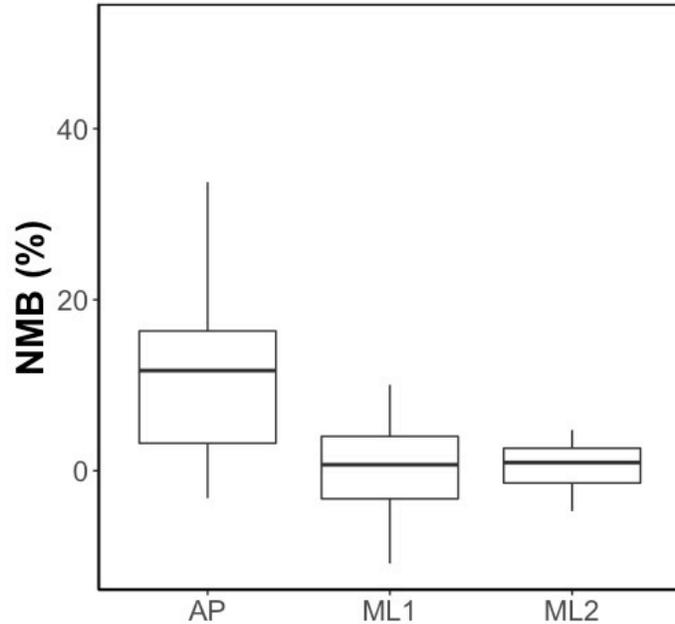
- The normalized mean bias (NMB) of AIRPACT is higher than ML models; ML1 and ML2 are close.

- ML1 captured the most high O$_3$ days, and ML2 provided the fewest false alarms.

		Observation				Observation				Observation	
		AQI<=2	AQI>2			AQI<=2	AQI>2			AQI<=2	AQI>2
AQI<=2	AP	15400	126	ML1		15248	92	ML2		15540	157
AQI>2		167	44			319	78			27	13

Models vs. observed daily maximum 8-hr averaged O$_3$ (MDA8)

AIRPACT
R2 : 0.54
NMB: 11%
NME: 19%

ML1
R2 : 0.66
NMB: 1.4%
NME: 16%

ML2
R2 : 0.76
NMB: 1.2%
NME: 12%
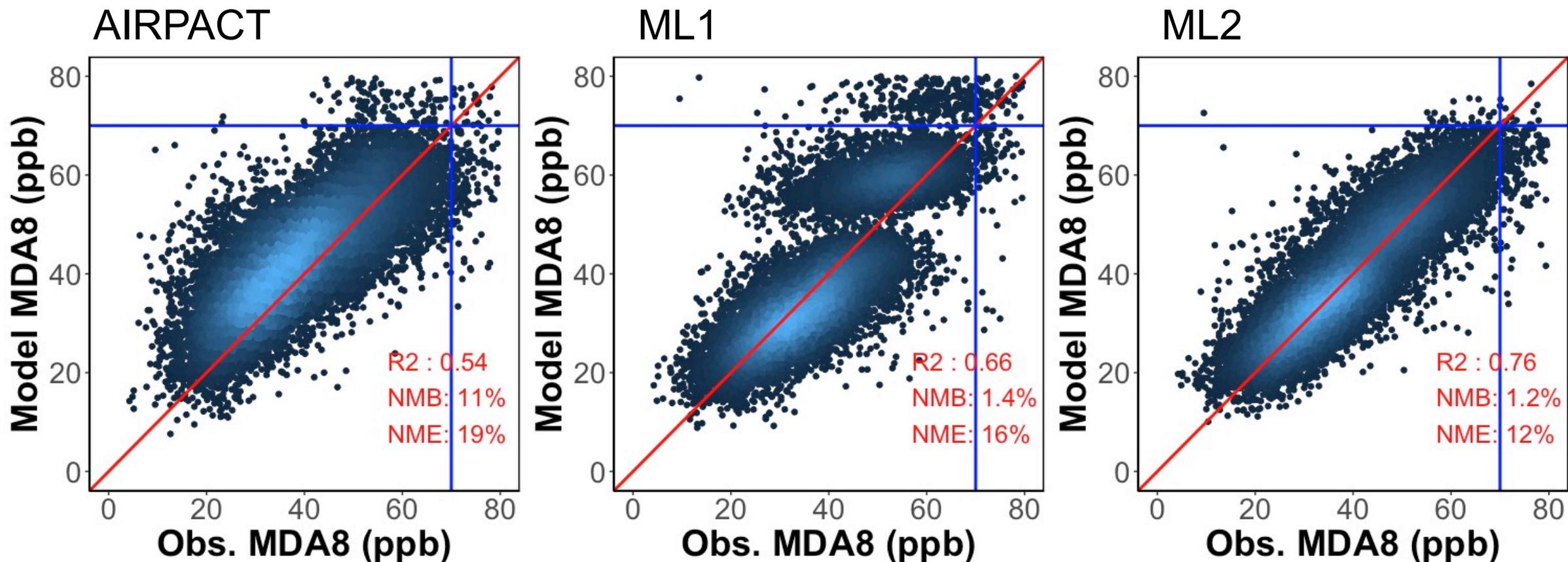
- Compared to AIRPACT, ML1 and ML2 improved the model performance based on the statistical variables (R^2, NMB and NME).
- The scatter plot of ML1 has three clouds, and ML2 shape is close to the 1:1 line.

MDA8 NMB at each observation site

AIRPACT
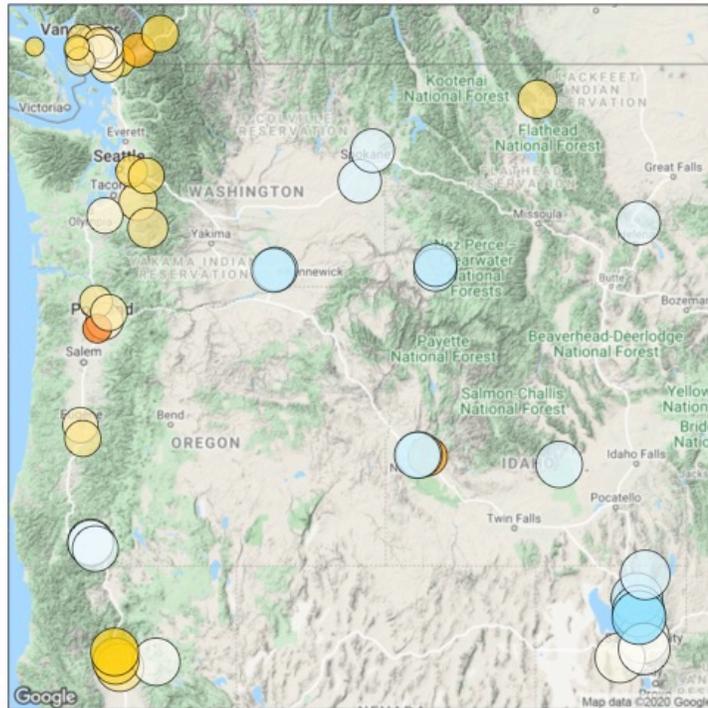
ML1
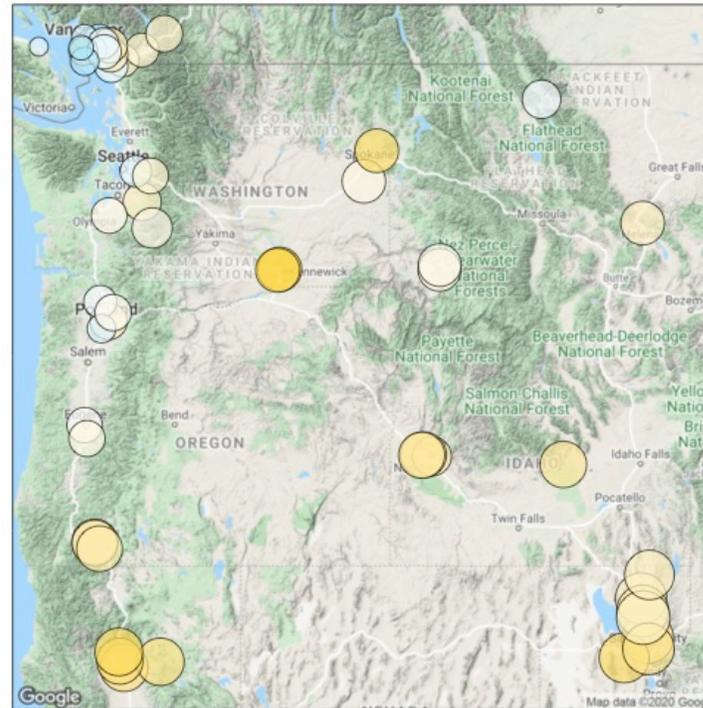
ML2
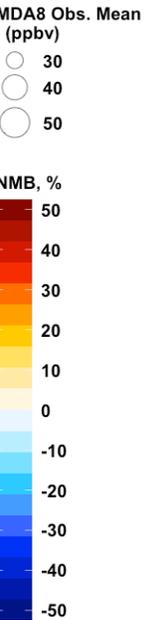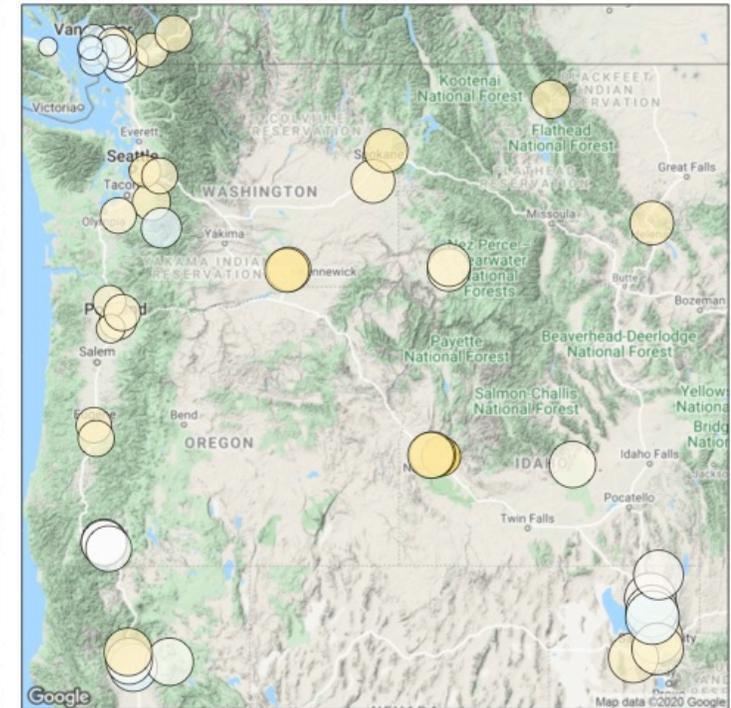
- Overprediction of AIRPACT mostly along the coast
- Underprediction of AIRPACT over the inland sites
- Improved predictions by ML models

HSS/KSS vs. NMB of O₃ models

AIRPACT

ML1

ML2
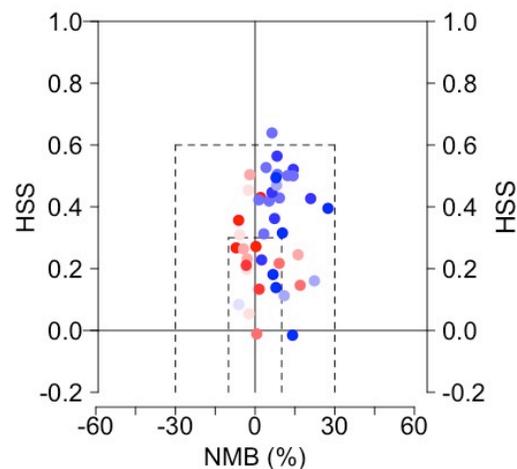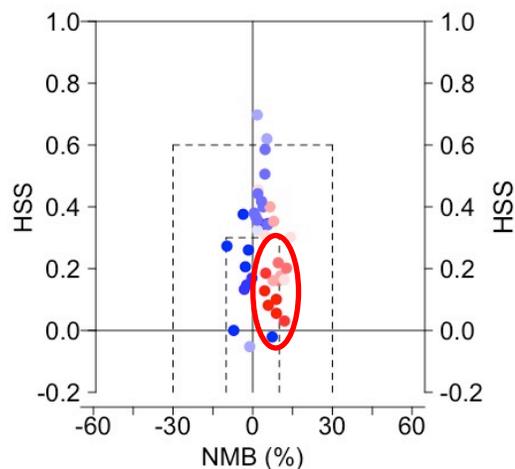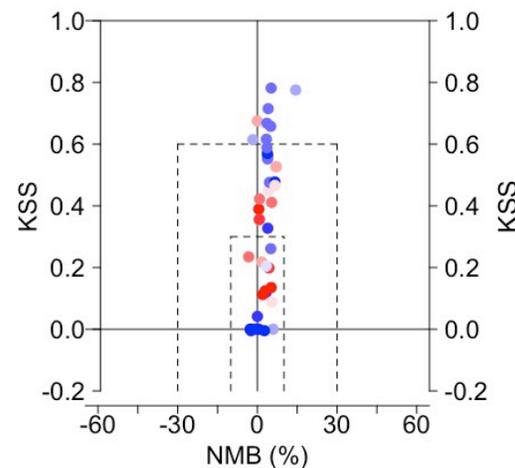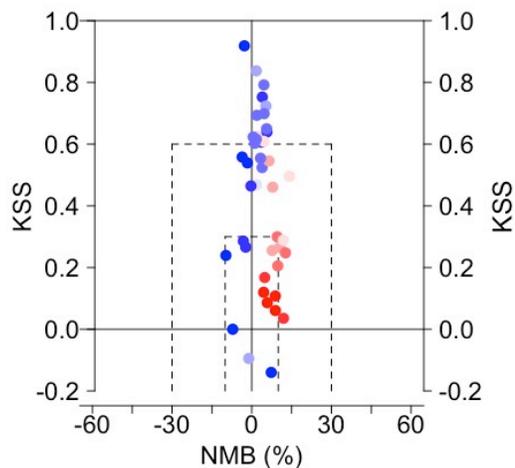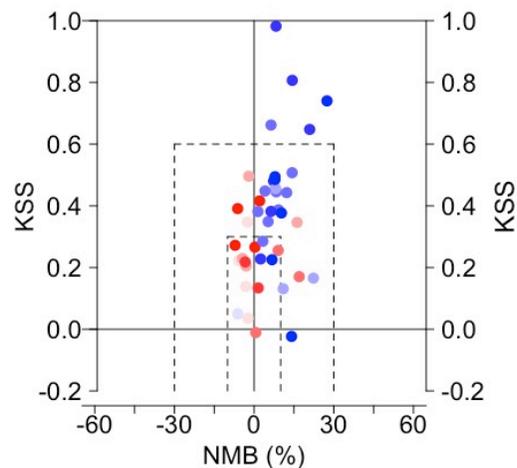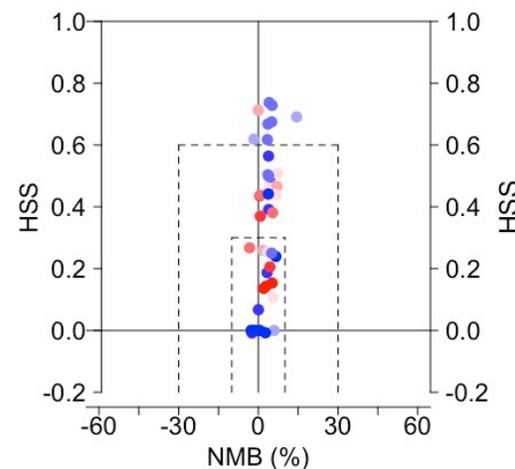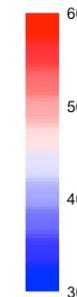
- The sites with lower MDA8 show higher HSS and KSS for all models.

- ML1 tends to overpredict MDA8 at the sites with higher MDA8.
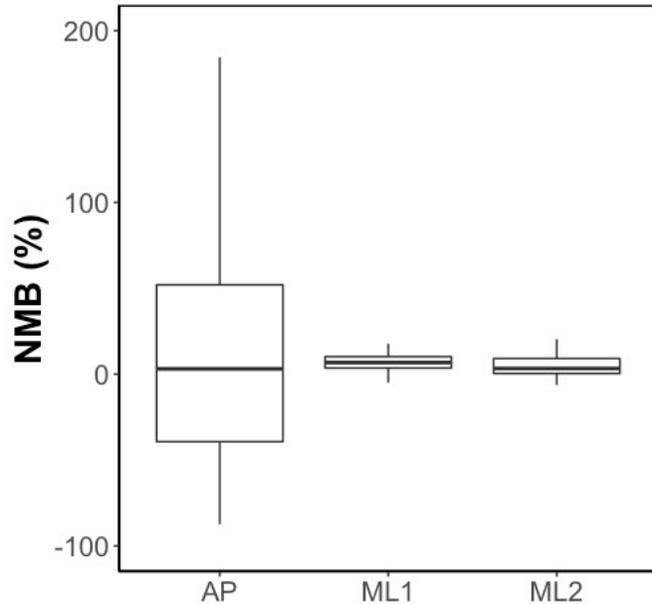
- ML2 shows narrow range of NMB.

Q-Q plots of MDA8 O₃ models
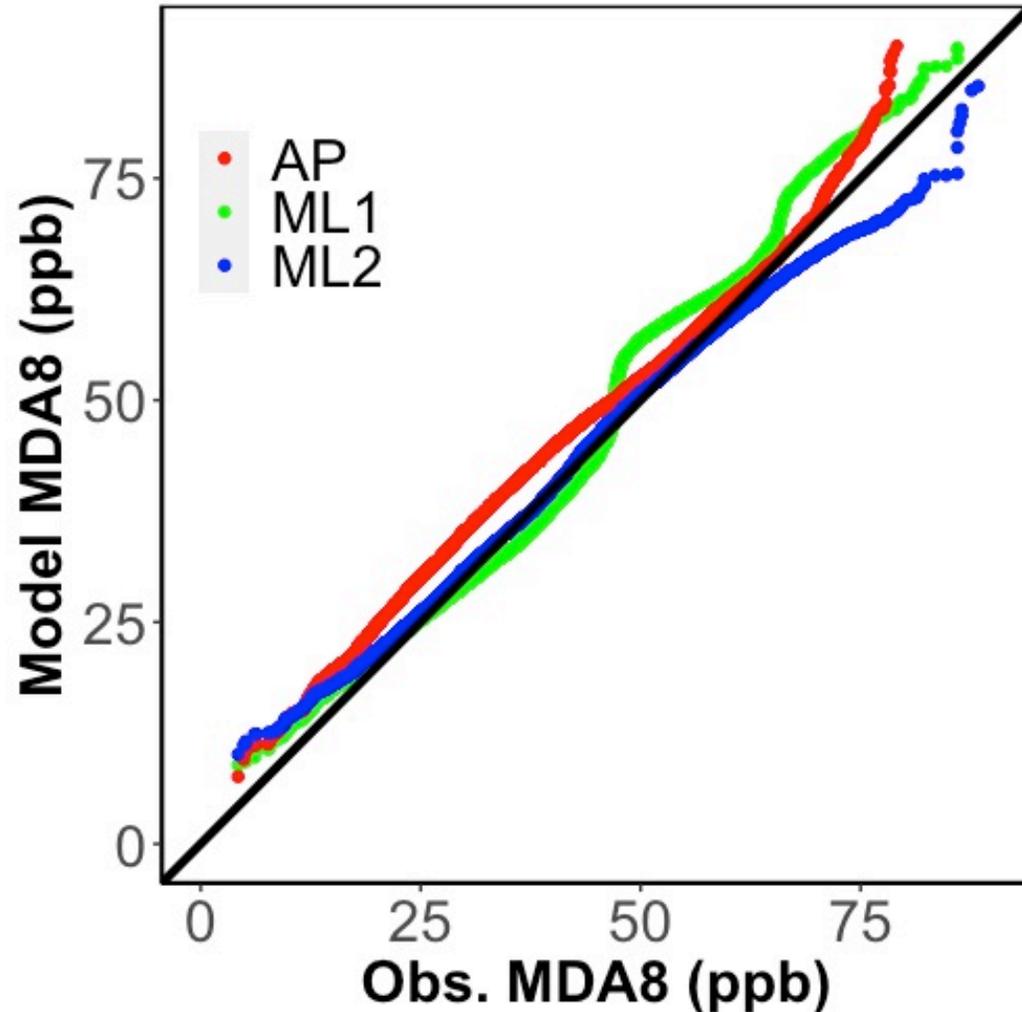
- AIRPACT may provide extremely high MDA8. This also happens for ML models, but is not as high as AIRPACT

- ML2 prediction is close to 1:1 line below 70 ppb but gets lower for high MDA8.

- ML1 may overpredict low MDA8 but gets close to 1:1 line when the MDA8 is higher.

PM2.5 evaluation during Nov – Feb in 2018 - 2020

- Similar medians of the NMB of three models

- AIRPACT with a larger variation

- AIRPACT captured the most high PM2.5 days but too many false alarms.

- The ML models for high PM2.5 days performed similarly, but ML2 reduced the false alarm number.

		Observation				Observation				Observation	
		AQI<=2	AQI>2			AQI<=2	AQI>2			AQI<=2	AQI>2
AQI<=2	AP	31942	78	ML1		32630	90	ML2		32689	94
AQI>2		768	29			80	17			21	13

Models vs. observed daily mean PM2.5 in winter
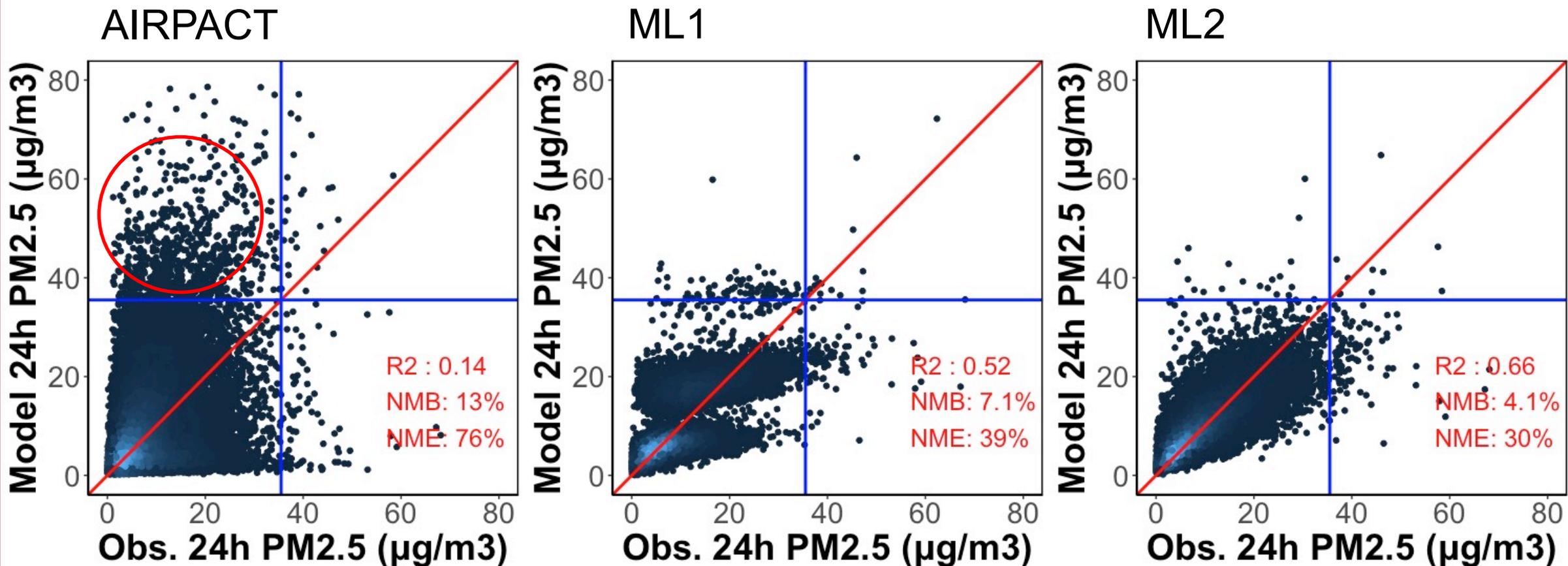
- The AIRPACT overpredicted the low PM2.5 cases.
- Like O_3 predictions, there are 3 groups of scatters for ML1 predictions.
- ML2 shows highest R^2 and lowest NMB and NME.

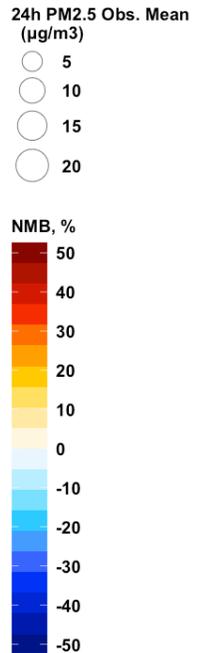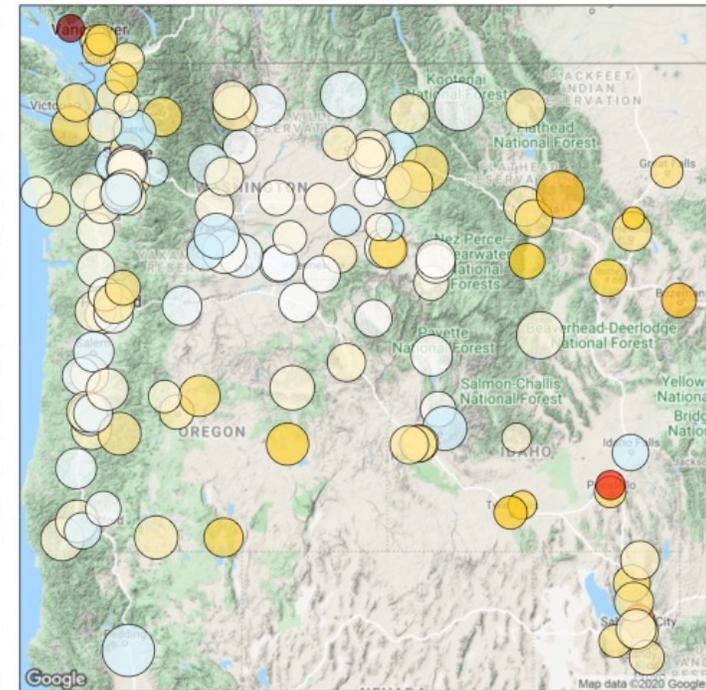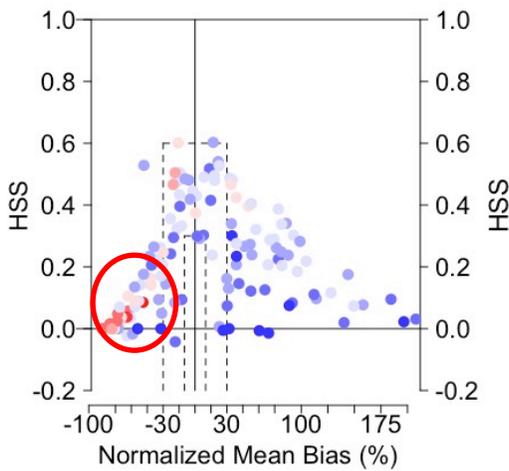# Daily mean PM2.5 NMB at each observation site in winter
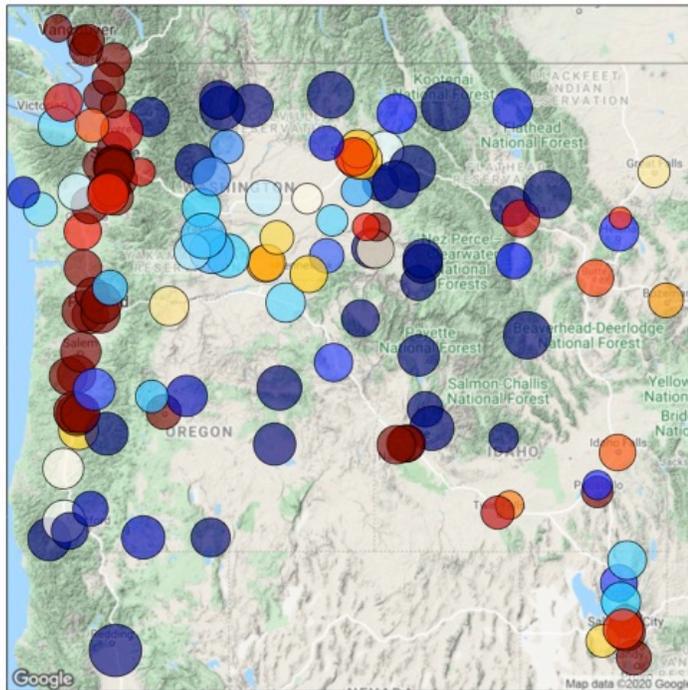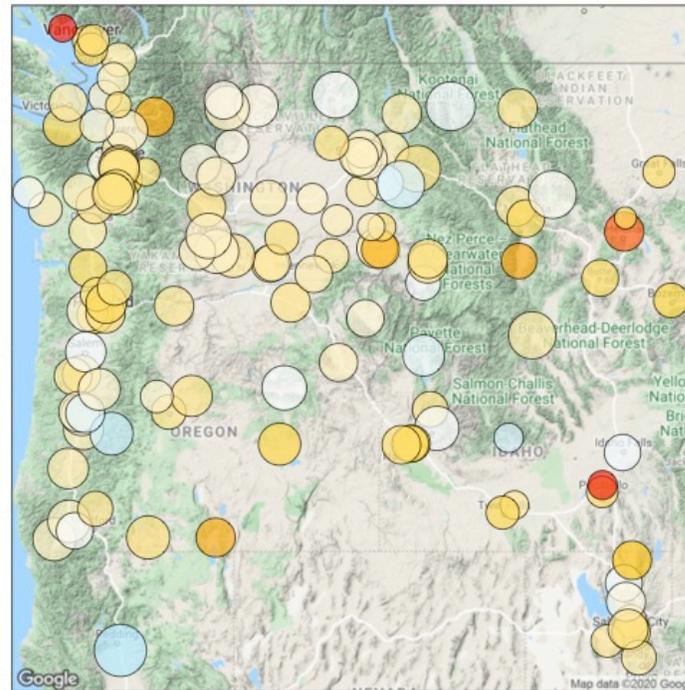
AIRPACT   ML1   ML2

- The NMB of AIRPACT shows a large variation.
- Like O_3, AIRPACT overpredicted the coast PM2.5 and underpredicted the inland PM2.5.
- ML1 and ML2 reduced the NMB, and ML2 gave lower NMB than ML1 in WA.
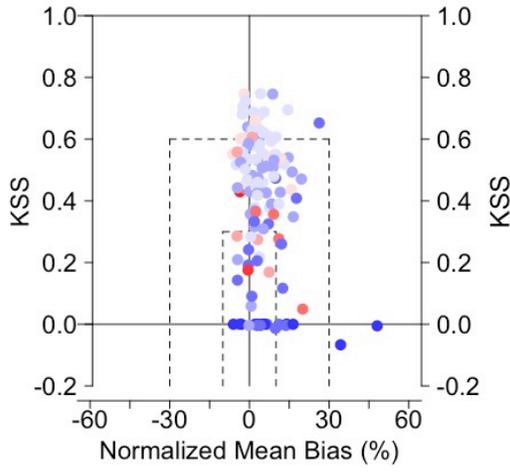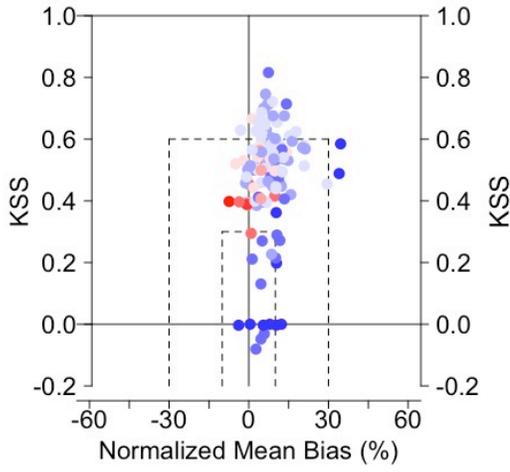
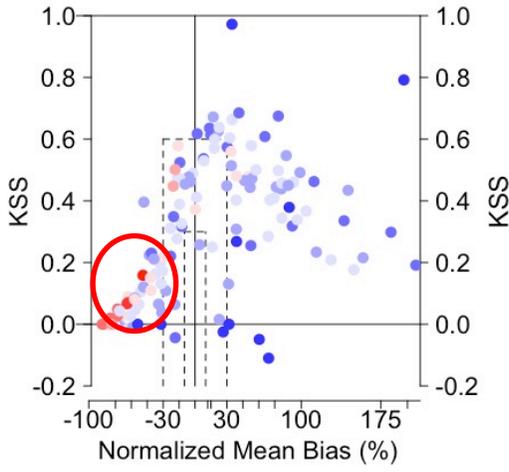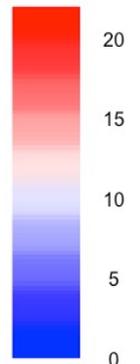# HSS/KSS vs. NMB of winter PM2.5 models

AIRPACT   ML1   ML2
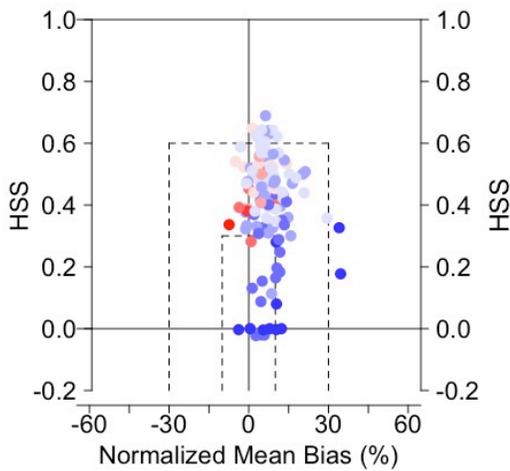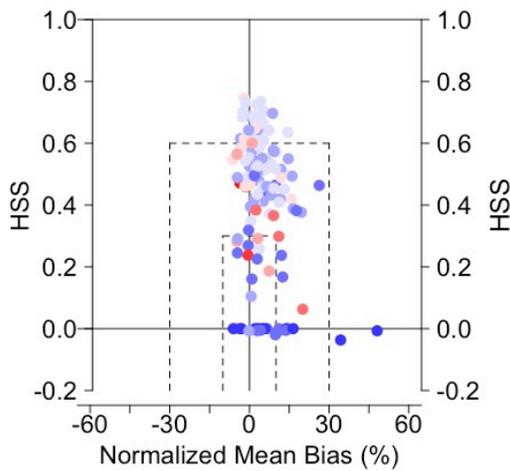
- AIRPACT underpredicted the high PM2.5 sites with poor HSS and KSS values

- No clear trend for ML1 and ML2, but the narrow range of NMB
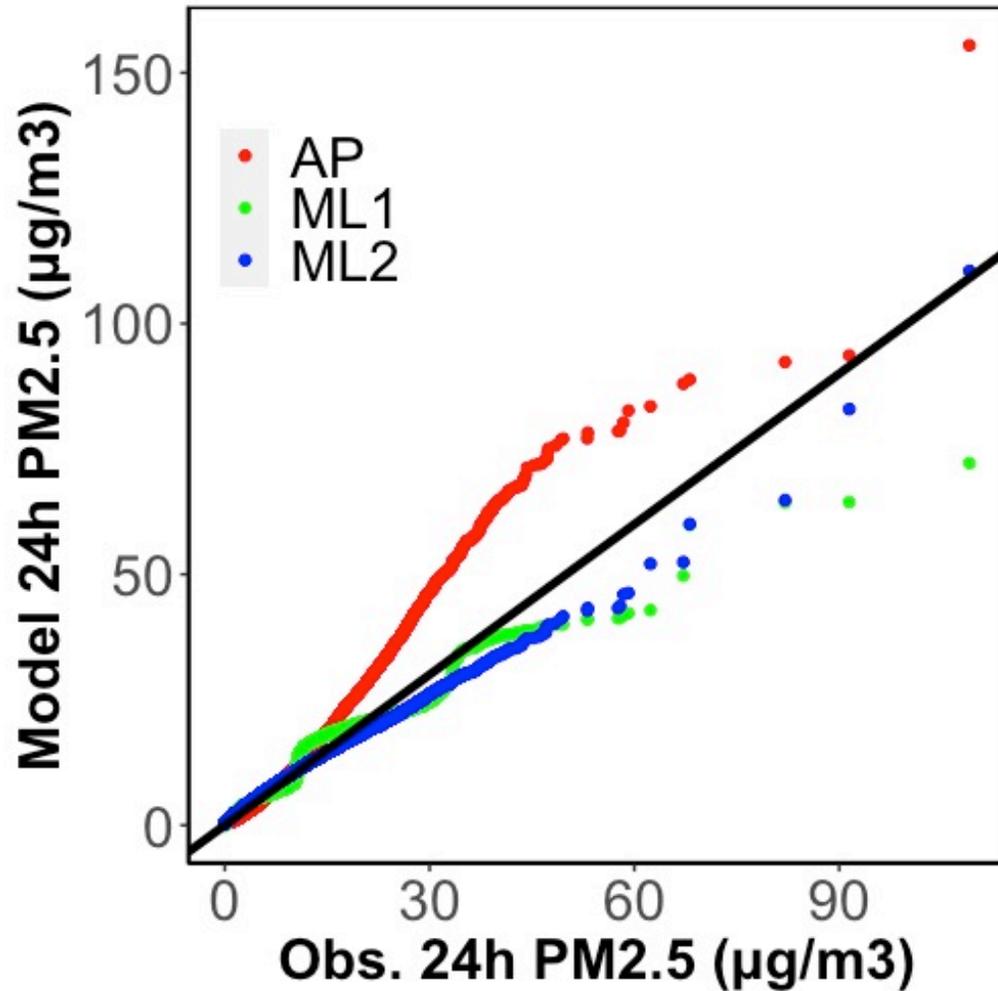
- ML models shows higher HSS and KSS than AIRPACT.
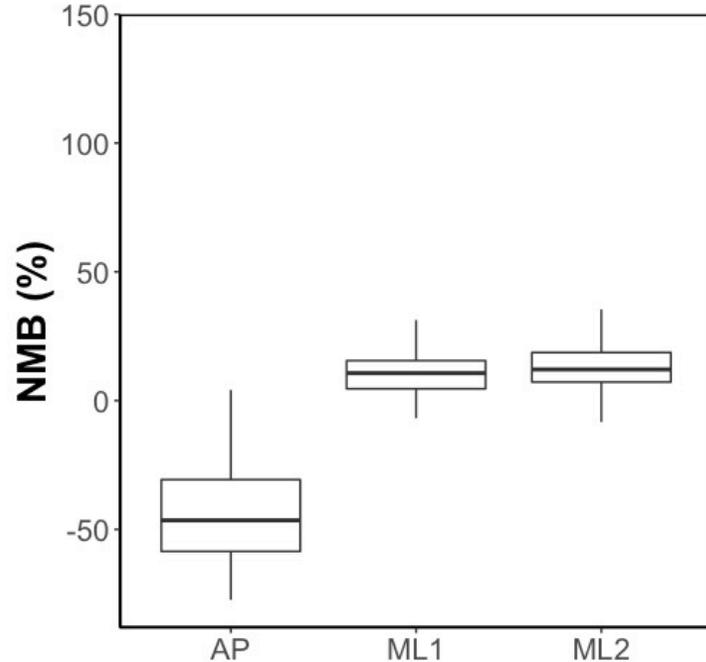
Q-Q plots of winter PM2.5 models

- AIRPACT overpredicts PM2.5 when PM2.5 is above 20 µg/m^3.

- ML1 and ML2 is close to 1:1 line.

# PM2.5 evaluation during May – Sep in 2018 - 2020

- The NMB of AIRPACT is lower than ML models; ML1 and ML2 are close.

- AIRPACT provided fewer false alarms but missed the most high PM2.5 days.

- The model performance of ML1 and ML2 for high PM2.5 days is close, but ML2 reduced the false alarm number.

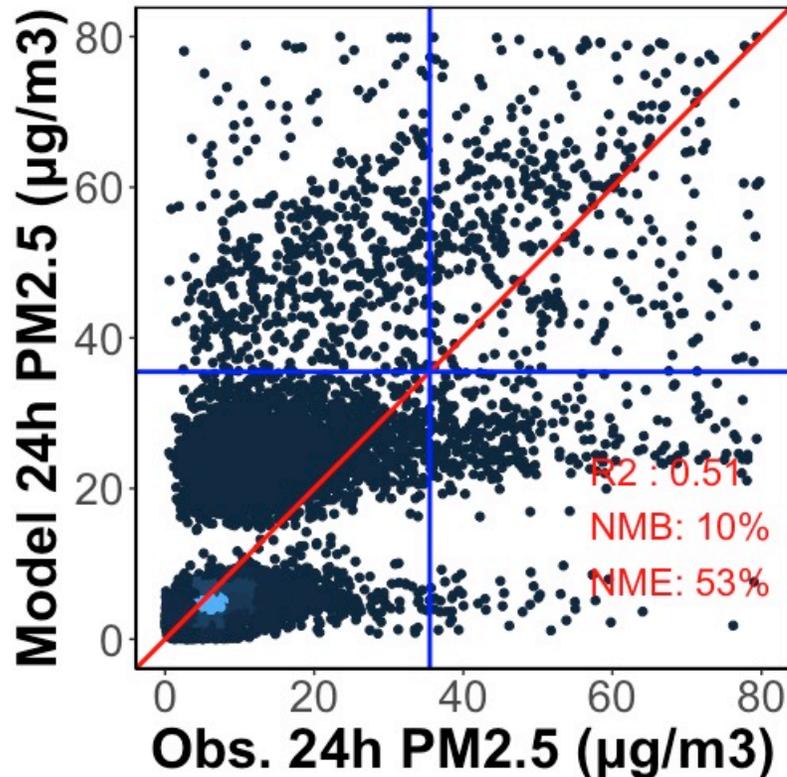|		Observation								
		AQI<=2	AQI>2			AQI<=2	AQI>2		AQI<=2	AQI>2
AQI<=2	AP	38917	925	ML1		38433	371	ML2	38682	429
AQI>2		68	97			552	651		303	593

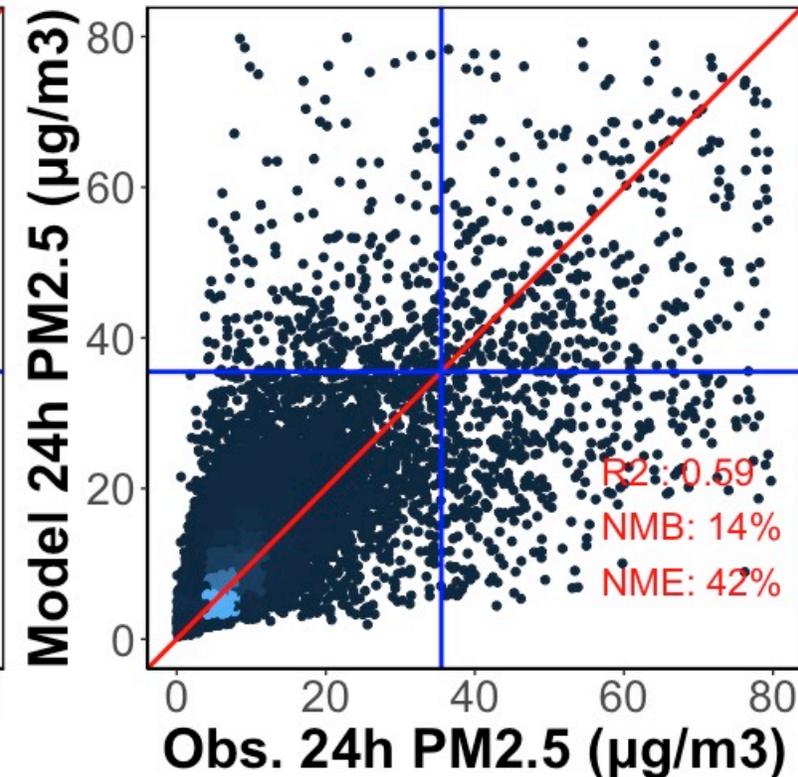# Models vs. observed daily mean PM2.5 in summer
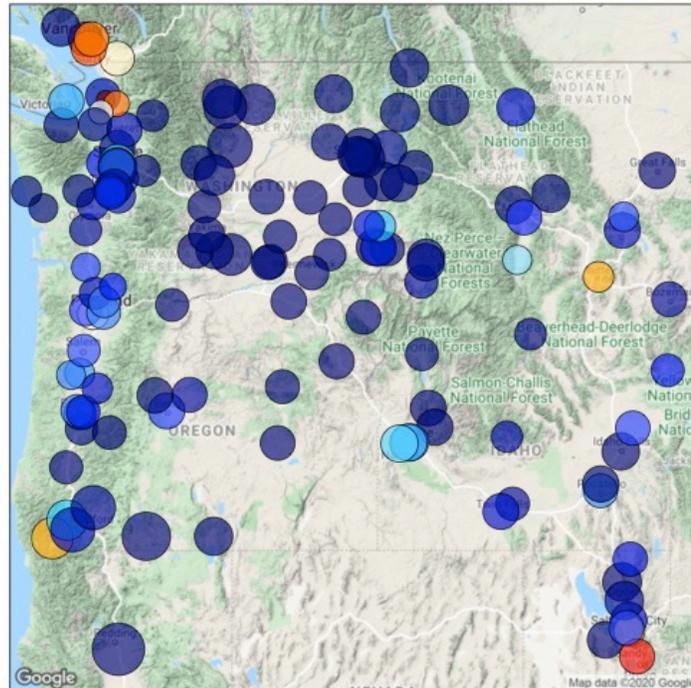
AIRPACT
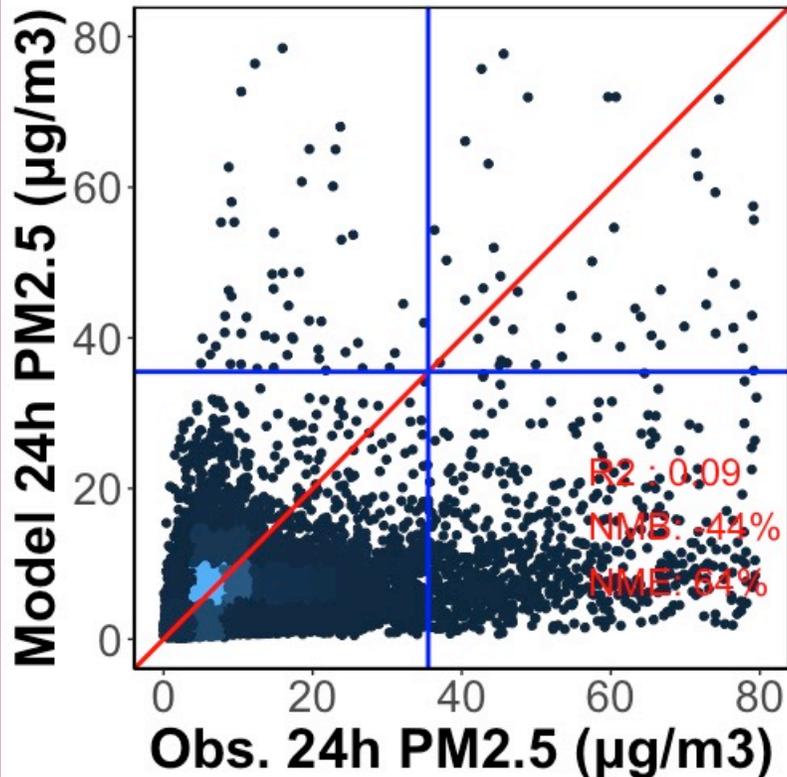
ML1

ML2

- The AIRPACT underpredicted the high PM2.5 cases
- Like previous results, there are 3 groups of scatters for ML1 predictions.
- ML2 shows highest R^2 and lowest and NME, but more scattered than winter ML2.
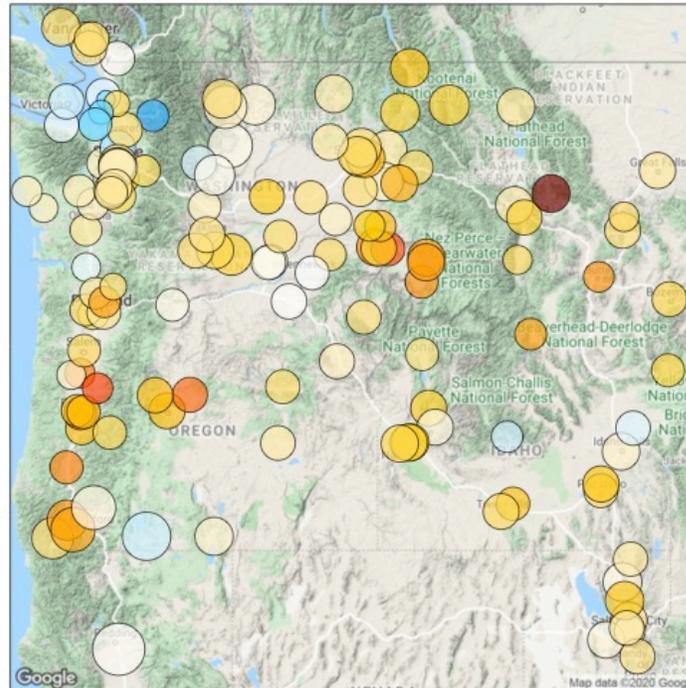
Daily mean PM2.5 NMB at each observation site in summer
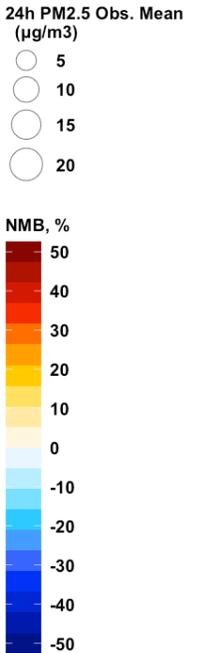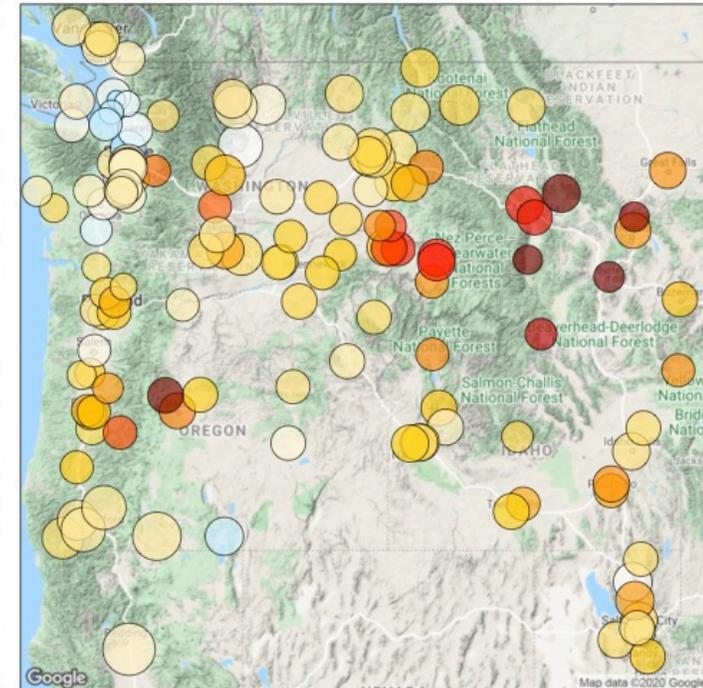
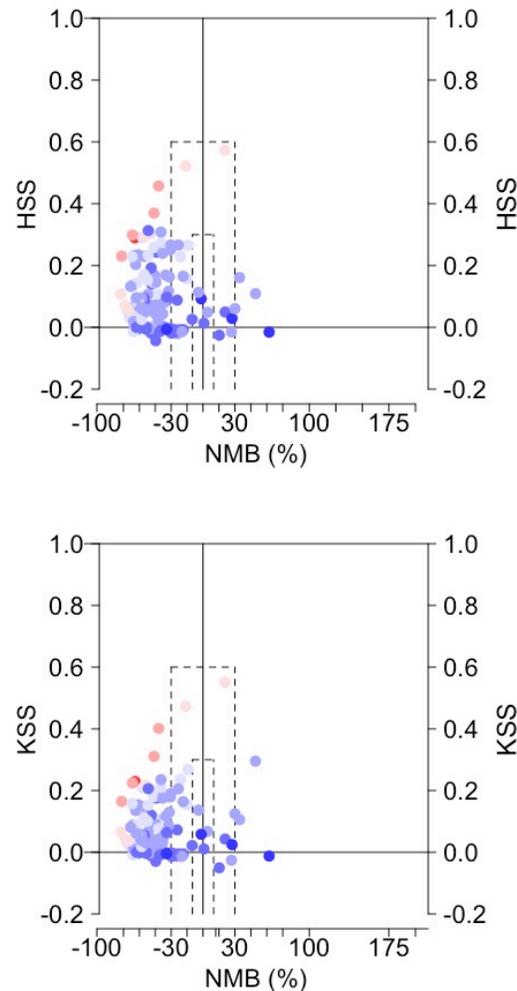AIRPACT   ML1   ML2

- AIRPACT underpredicted the PM2.5 at most sites.
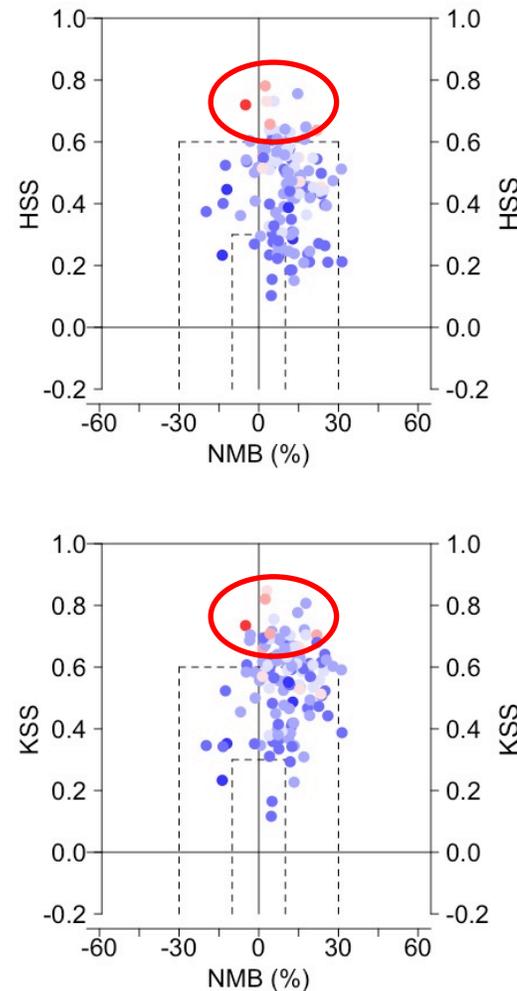- ML models overpredicted PM2.5, but the NMB is lower than 30% mostly.
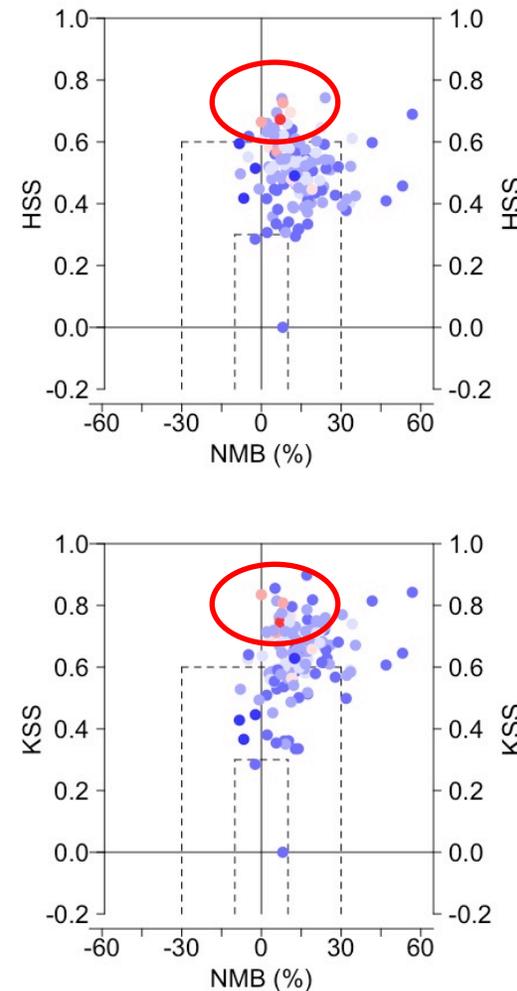
HSS/KSS vs. NMB of summer PM2.5 models
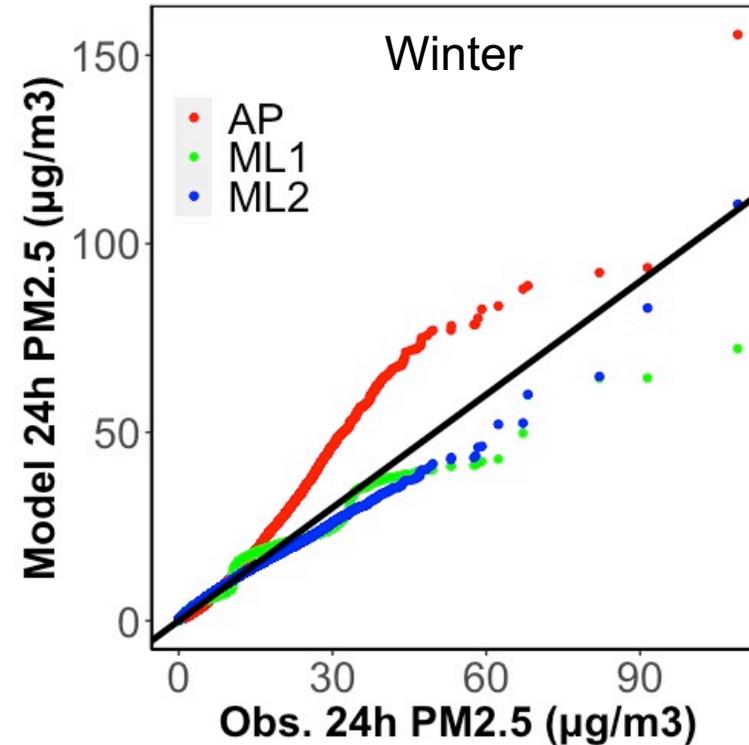
AIRPACT

ML1
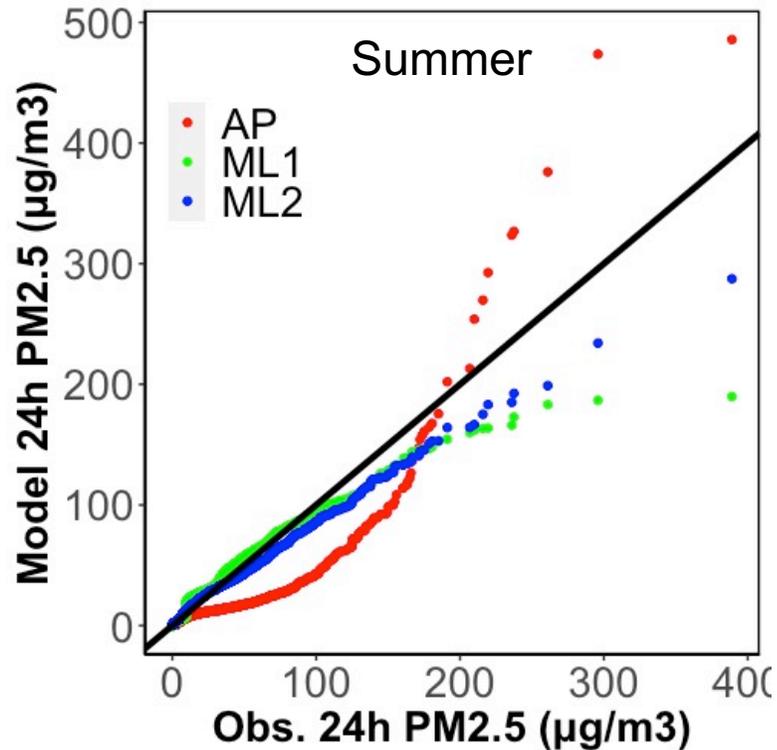
ML2

- AIRPACT underpredicted PM2.5.

- The HSS and KSS from ML models are higher for high PM2.5 sites.

- ML models get higher HSS and KSS than AIRPACT.

Q-Q plots of summer vs. winter PM2.5 models

- AIRPACT underpredicted PM2.5 in summer and overpredicted in winter. It provided some extremely high concentrations in both seasons.

- The model performance of ML models is similar in two season. ML1 and ML2 are closer to 1:1 line, but underpredicted the high PM2.5. ML2 is better than ML1.

Summary

- Machine learning models were used to predict O_3 and PM2.5 in the PNW and performed well for O_3 and PM2.5 during time periods considered here.

- For summer O_3, the ML models performed better than AIRPACT. ML1 captured more high O_3 events but had higher false alarms than ML2. ML2 performed better for low O_3 events.

- For winter PM2.5, the ML models showed lower false alarms and narrower NMB variations than AIRPACT. In summer, ML models slightly overpredicted but showed significantly better performance for high concentrations, esp. ML2.

- For future O_3 forecasts, ML1 will be used for high O_3 events and ML2 for low O_3.

- For future PM2.5 forecasts, ML2 will be used.

- All ML models will be re-trained every two weeks to include the latest data.

Thank you