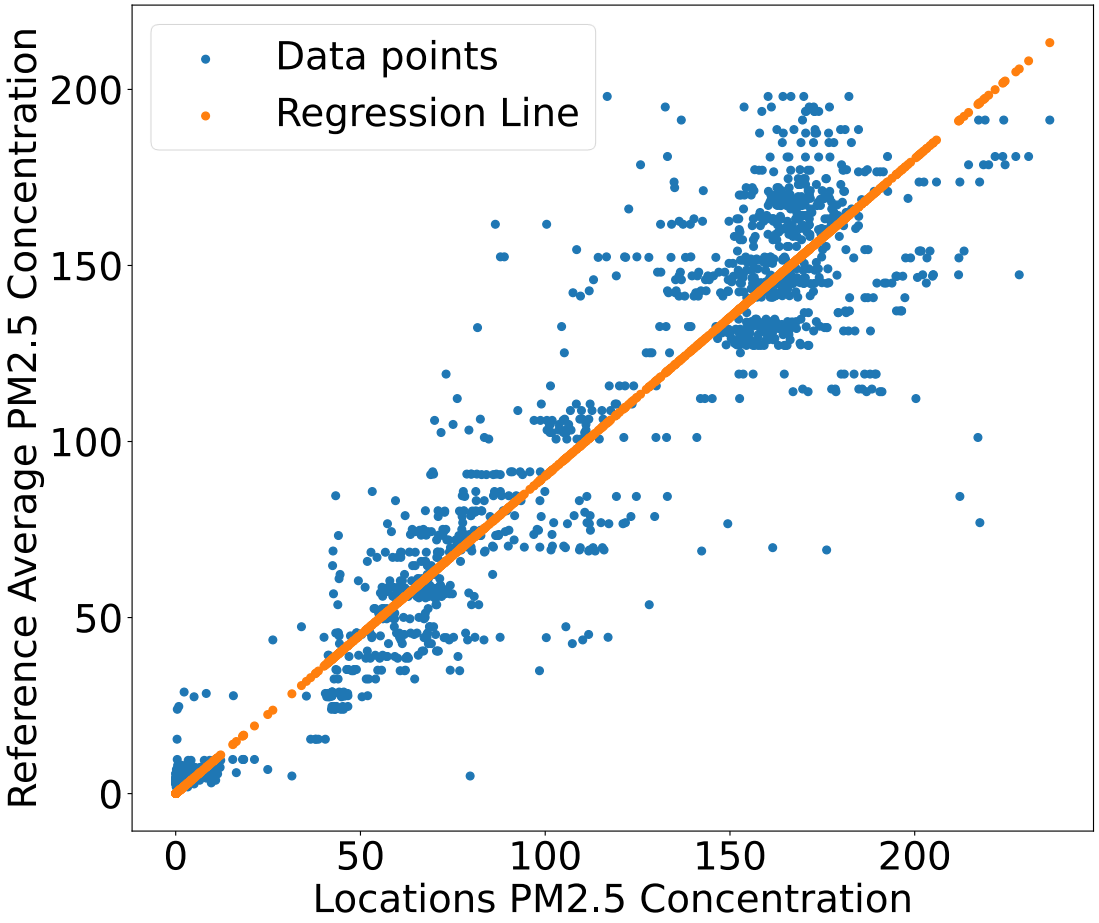
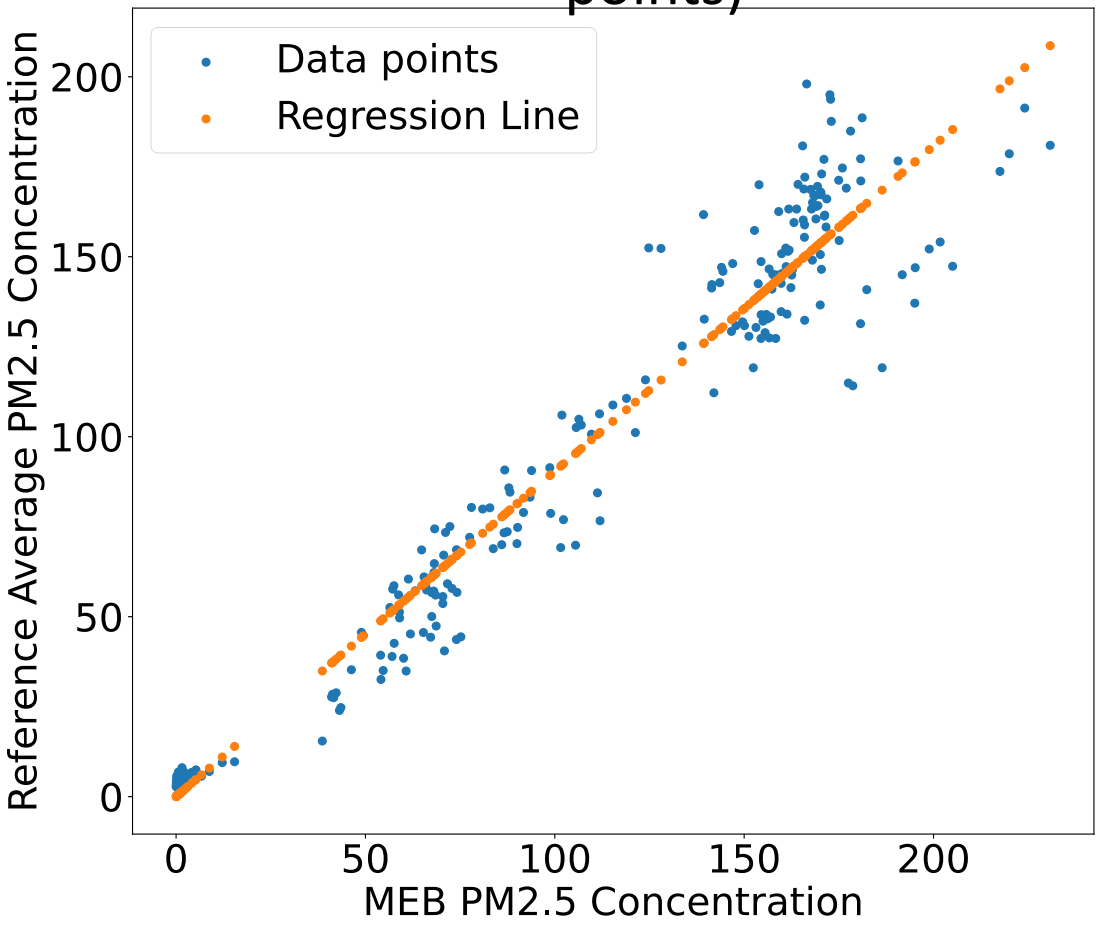
# Calibration with Linear Regression On Outdoor Sensor Data

# All data vs MEB data as Training Data

All Data Points (1512 data points)




MEB Outdoor Data Points (252 data points)



# Linear Regression: Offset vs No Offset

- Method: 10-fold cross validation used for 2 linear regression models: one with offset (ax+b) and one without offset (ax)
- 2 sets of training data for training: All data points (PM2.5 concentration data points from all 6 locations, excluding Saad's), MEB data points

| | **average MSE** | **average R squared** |
|---|---|---|
| **All data points as training with no offset** | 285.584 | 0.9233 |
| **All data points as training with offset** | 280.7394 | 0.9270 |
| **MEB as training with no offset** | 195.7051 | 0.9462 |
| **MEB as training with offset** | 208.1803 | 0.9441 |

Conclusion: Miniscule difference. Either one can work.

# Linear Regression: Offset vs No Offset

- Our linear regression model: Y = aX + b
- Training data: All data points
- Use each location's PM2.5 concentration data as test data and test it with our model

| **ALL DATA POINT TRAIN a = 0.9046 b = 0.5310** | **MSE** | **r squared** |
|---|---|---|
| MEB | 206.14 | 0.9454 |
| Igor | 315.80 | 0.9163 |
| Stephanie | 281.59 | 0.9254 |
| Alex | 280.27 | 0.9258 |
| Edmund | 234.31 | 0.9379 |
| Brad | 373.05 | 0. 9012 |

| **ALL DATA POINT TRAIN a = 0.9046 b = 0** | **MSE** | **r squared** |
|---|---|---|
| MEB | 205.87 | 0.9455 |
| Igor | 309.66 | 0.9179 |
| Stephanie | 280.22 | 0.9258 |
| Alex | 282.97 | 0.9250 |
| Edmund | 236.08 | 0.9375 |
| Brad | 378.01 | 0.8999 |

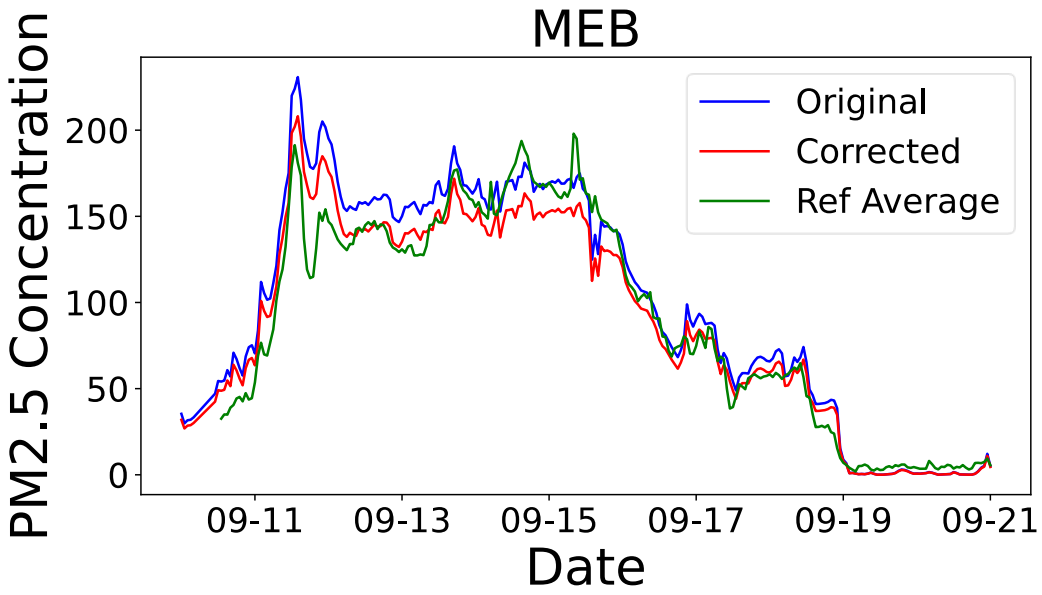
# Linear Regression: All data points vs MEB w/o offset

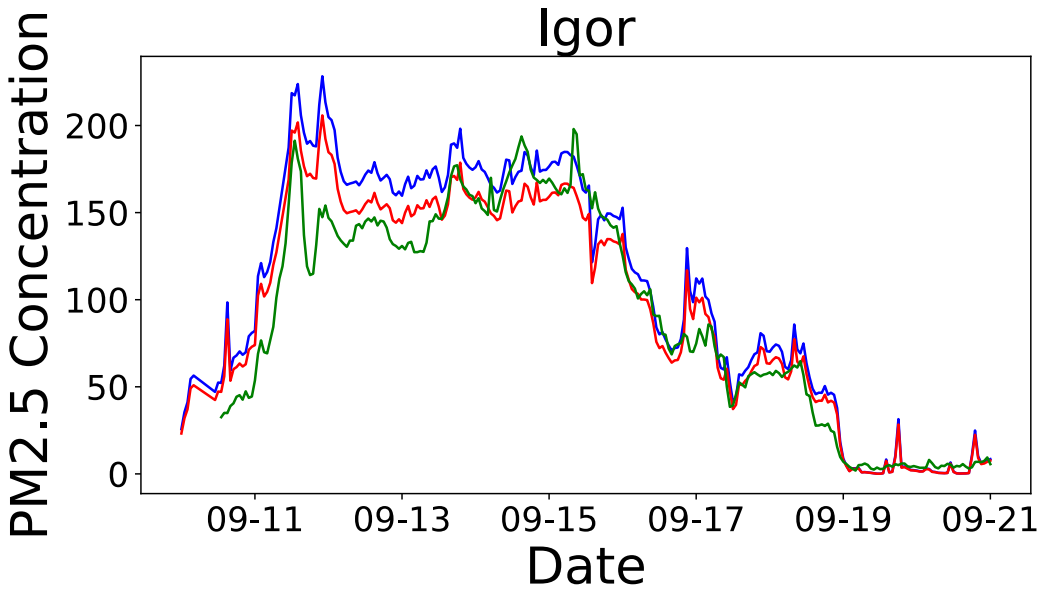
- Compare using all data points as training data vs using MEB as training data
- Use linear regression model Y = aX

| ALL DATA POINT TRAIN a = 0.9046 | MSE | r squared |
|---|---|---|
| MEB | 205.87 | 0.9455 |
| Igor | 309.66 | 0.9179 |
| Stephanie | 280.22 | 0.9258 |
| Alex | 282.97 | 0.9250 |
| Edmund | 236.08 | 0.9375 |
| Brad | 378.01 | 0.8999 |

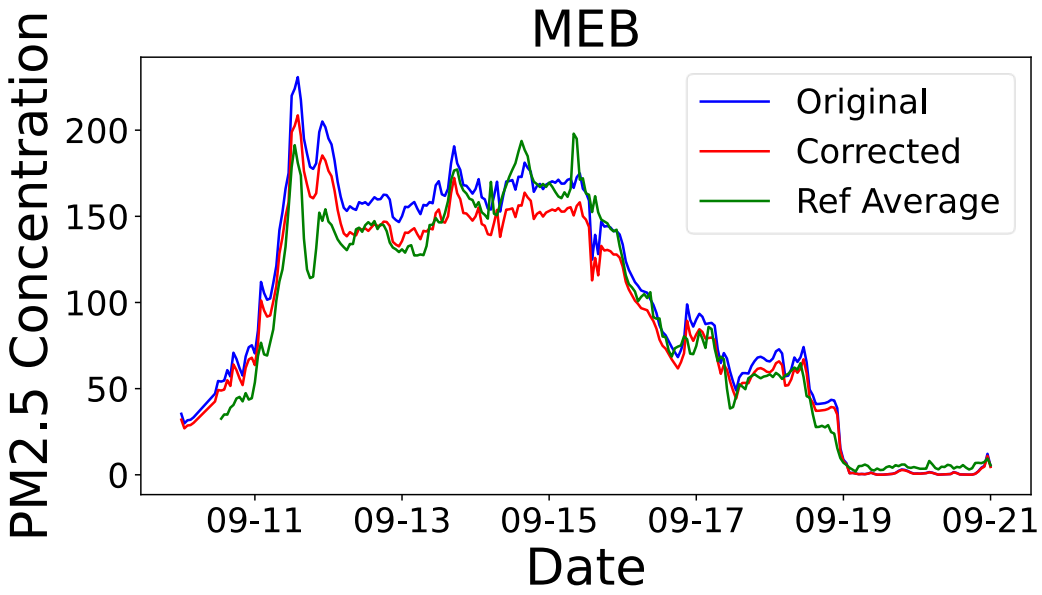
| MEB Train Data a = 0.9039 | MSE | r squared |
|---|---|---|
| MEB | 205.87 | 0.9455 |
| Igor | 308.51 | 0.9183 |
| Stephanie | 279.98 | 0.9258 |
| Alex | 283.35 | 0.9249 |
| Edmund | 236.48 | 0.9374 |
| Brad | 378.69 | 0.8997 |

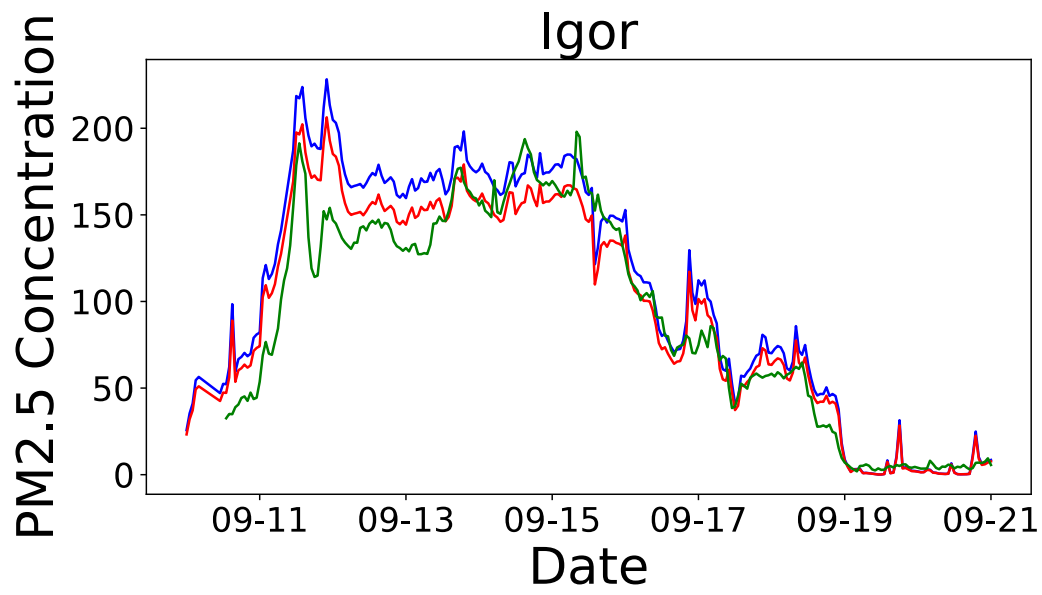
## All Data as train data



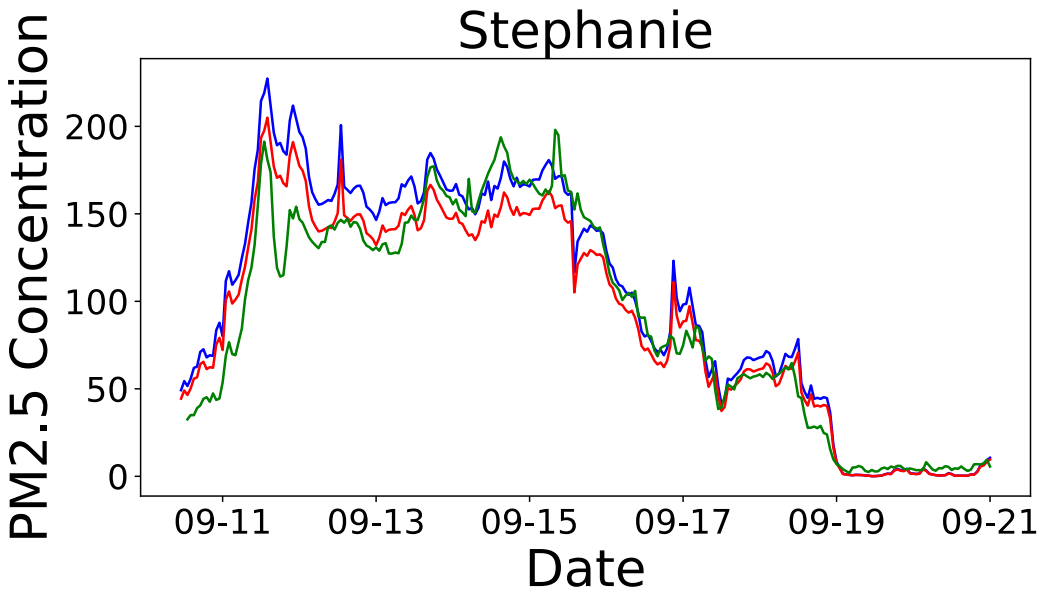



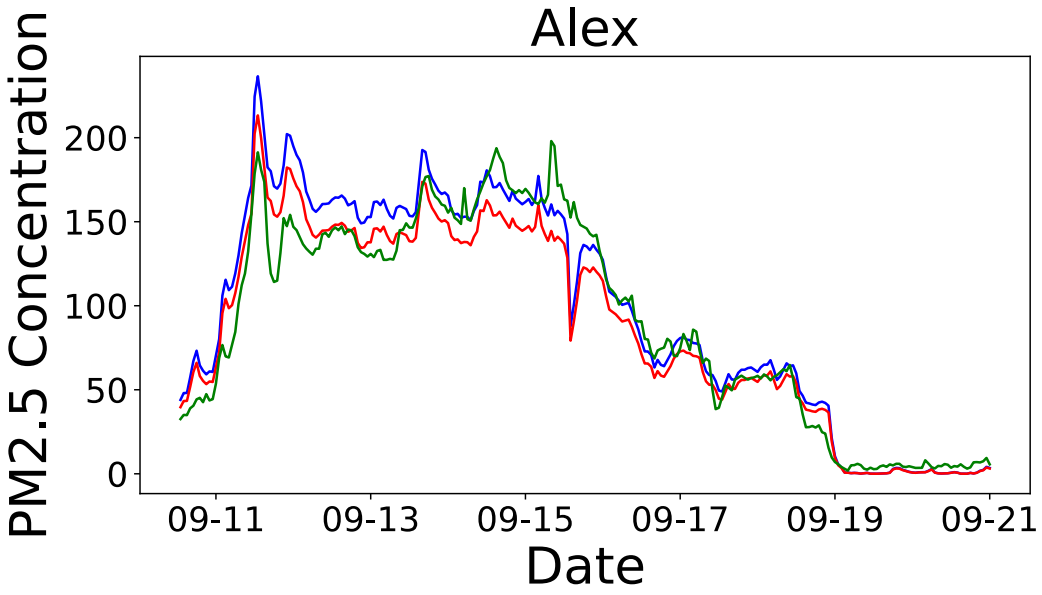

## MEB as train data



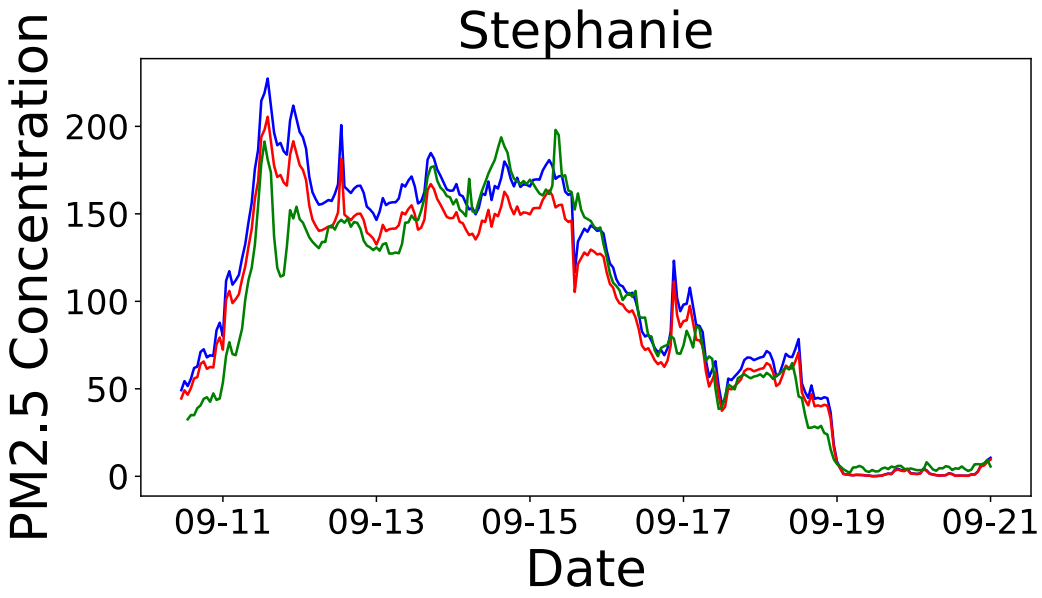



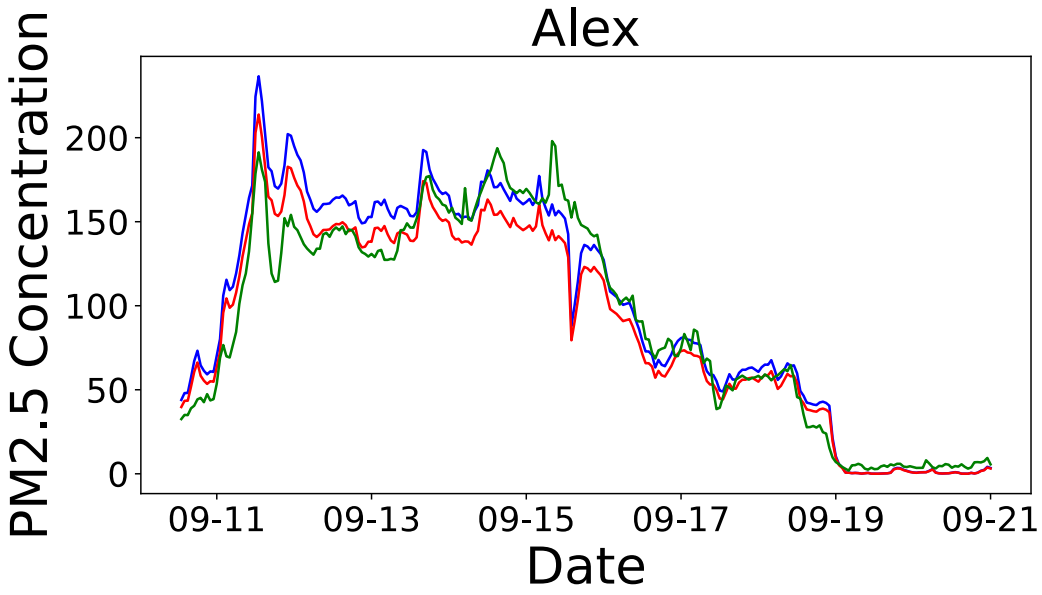
## All Data as train data



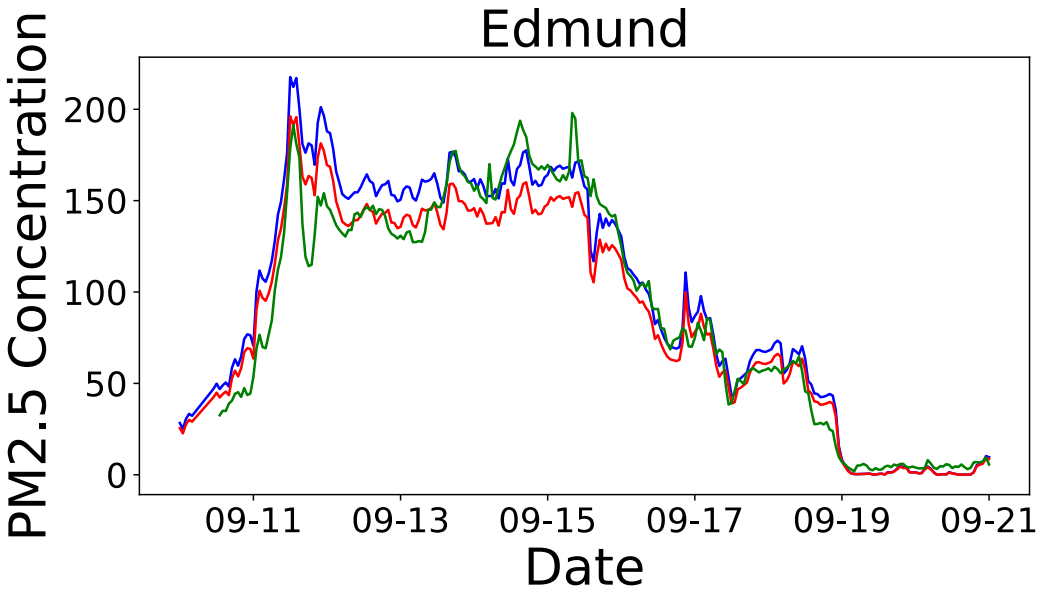



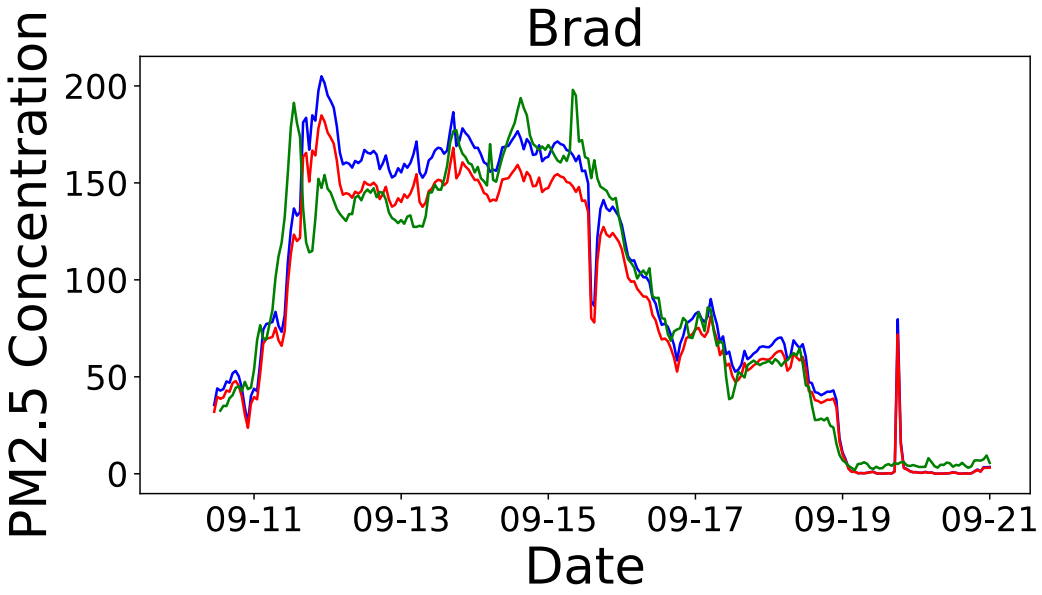

## MEB as train data



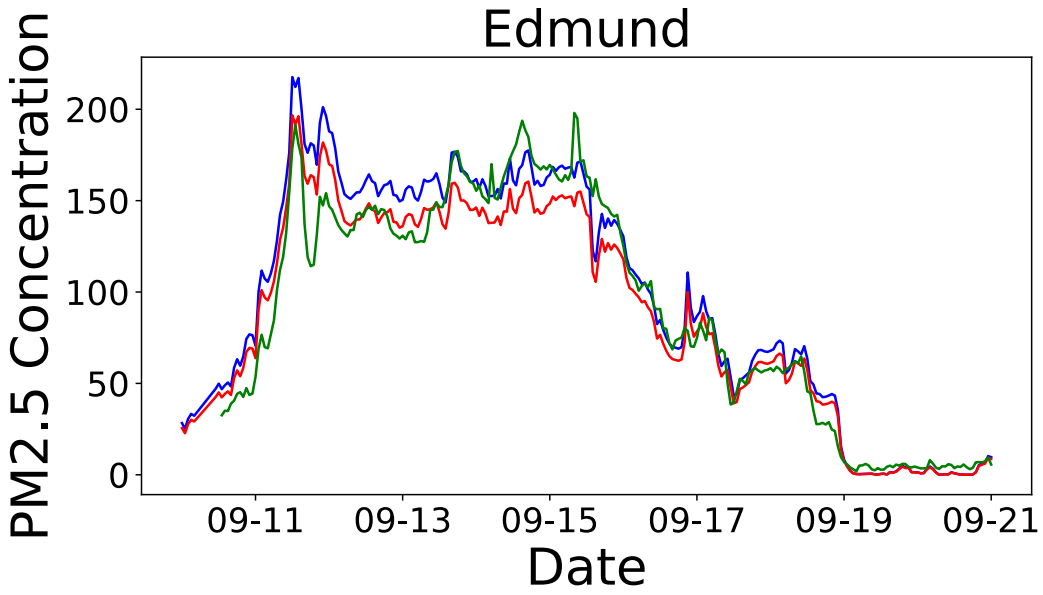



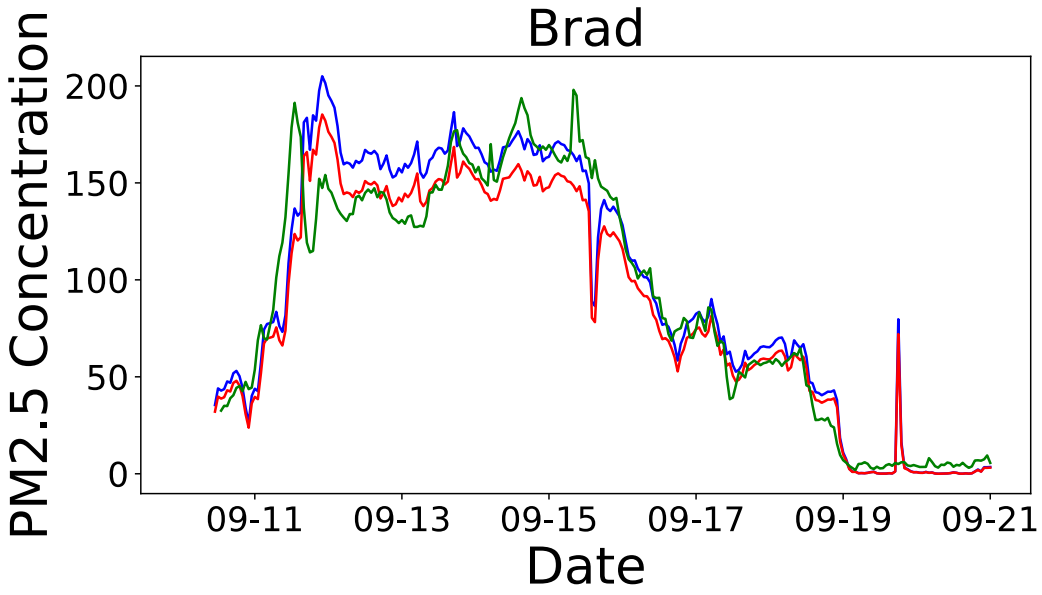
## All Data as train data



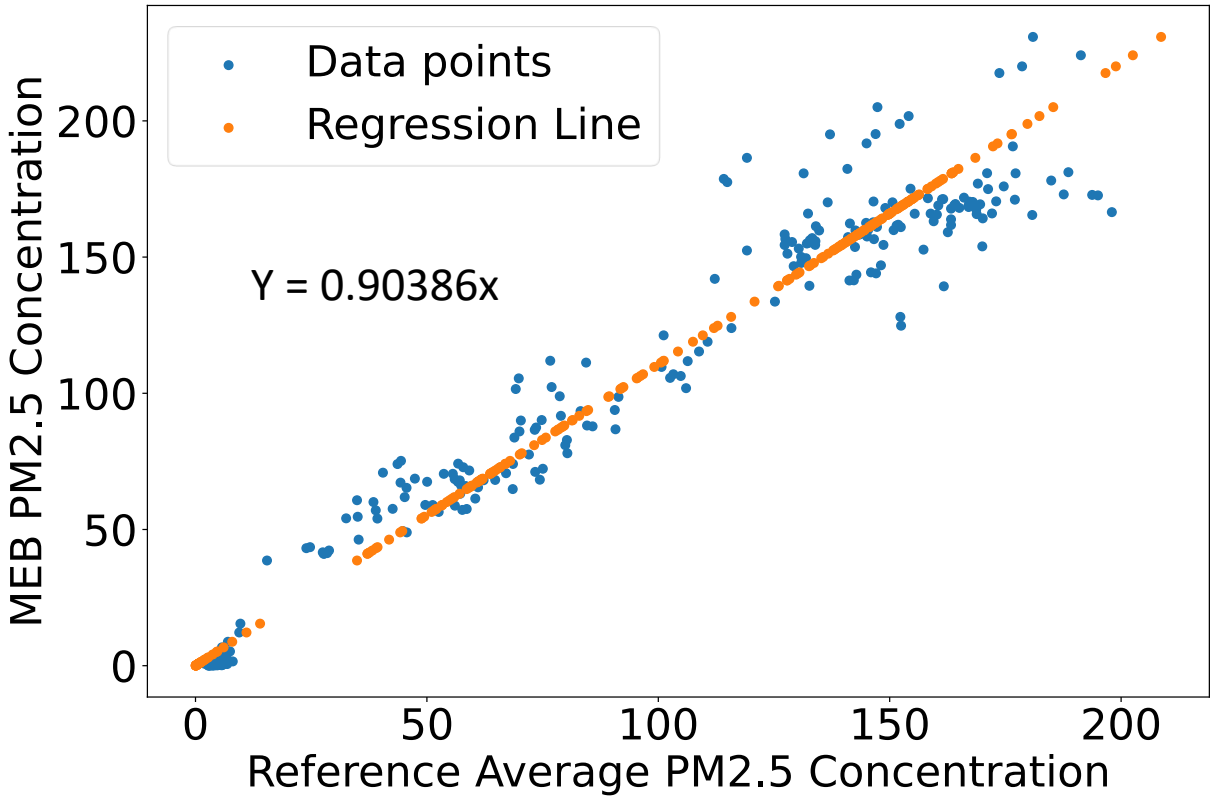

## MEB as train data

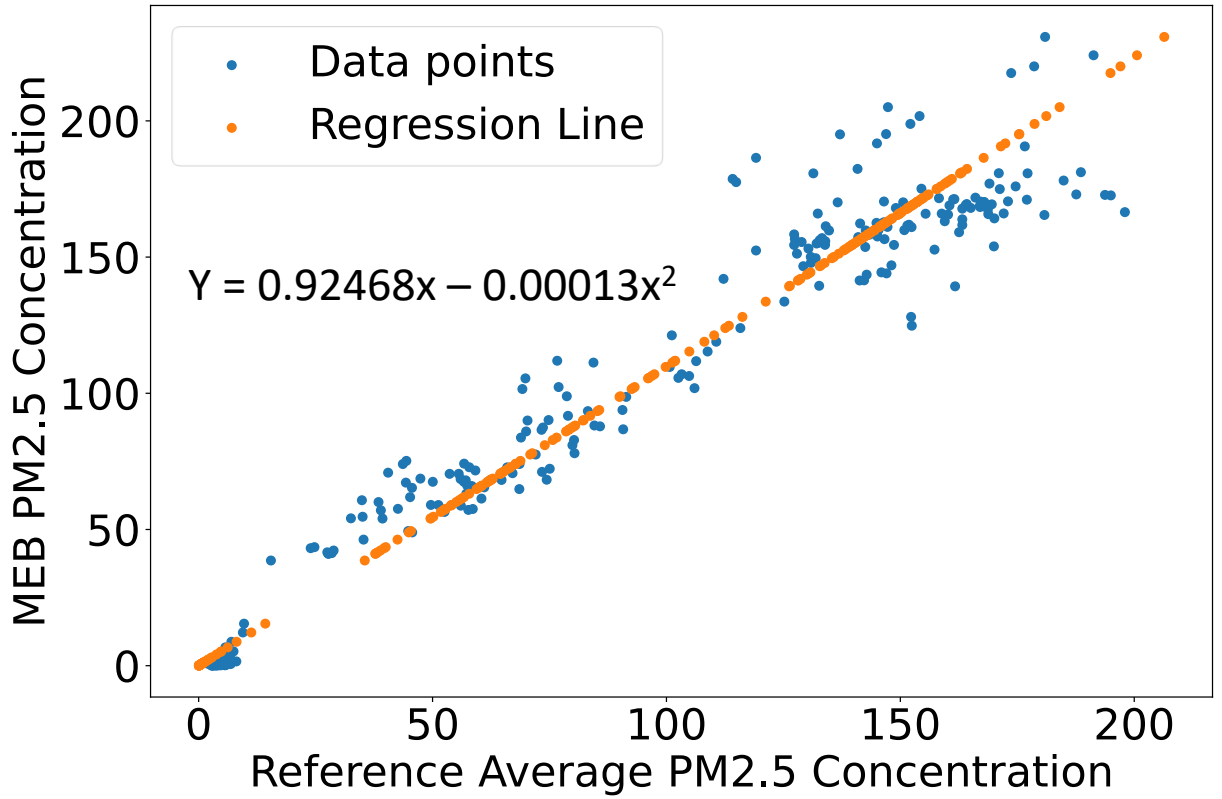


# Linear Regression vs Polynomial Regression

- MEB data as training data; other locations as test data



# Linear Regression vs Polynomial Regression

| Linear Model | MSE | r squared |
|---|---|---|
| Igor | 308.51 | 0.9183 |
| Stephanie | 279.98 | 0.9258 |
| Alex | 283.35 | 0.9249 |
| Roosevelt | 236.48 | 0.9374 |
| Brad | 378.69 | 0.8997 |

BIC: 1347.99

| Polynomial Model | MSE | r squared |
|---|---|---|
| Igor | 307.39 | 0.9186 |
| Stephanie | 279.03 | 0.9249 |
| Alex | 279.37 | 0.926 |
| Roosevelt | 236.12 | 0.9375 |
| Brad | 372.74 | 0.9013 |

BIC: 1353.12