A Machine Learning Approach for Ozone Forecasting and its Application for Kennewick, WA

Kai Fan¹, Brian Lamb¹, Ranil Dhammapala², Ryan Lamastro³, and Yunha Lee¹

¹Laboratory for Atmospheric Research, Civil and Environmental Engineering, Washington State University
²Washington State Department of Ecology
³State University of New York at New Paltz
Motivation

• Kennewick, WA lies 32 km (20 mi) north of Washington's southern border, where high \( O_3 \) events occur during summer and fall.

• AIRPACT is a state-of-the-science CMAQ-based air quality forecasting system for Pacific Northwest. However, AIRPACT struggles to predict high \( O_3 \) concentrations in this area.

• The goal of our study is to provide a reliable forecast for high \( O_3 \) events using the machine learning (ML) models, which can learn from the historical data to make future forecasts.
Machine Learning (ML) Model Approach for the Kennewick Monitoring Site

2015-2017 WRF met (T, P, RH, U, V, PBLH) + time info (month, weekday, hour) + previous day’s observed O₃

Training

ML Model

Evaluation

2018 observed O₃

2018 WRF met + time info + previous day’s observed O₃
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. O₃

Random Forest (RF) Classifier Model (RFc)

AQI categories

Multiple linear regression (MLR) Model

8-hr avg. O₃ pred.

Daily max. 8hr O₃ and AQI
Random Forest (RF) classifier

- RF classifier is the consensus of many decision trees, which we use to predict the AQI categories.

Multiple linear regression (MLR)

\[ Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \ldots \]

- MLR approach is used to predict the 8-h average $O_3$, which shows good performance to predict high $O_3$ days.
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 O₃

Two-phase random forest and weight factor

RF regression Model 1

Obs = a₁*RF1 + a₂*RF2

RF regression Model 2

Daily max. 8hr O₃ and AQI

Hourly O₃ pred.

Two-phase random forest (RF)

- The first RF model can usually make right prediction for low O$_3$ events, and the second phase isolates the events incorrectly predicted to form a second training dataset.

- We separate the initial predicted mixing ratios to three categories and give three sets of weight to two phases. The weight of two models are based on a simple linear regression equation.

\[
\text{Obs} = a_1 \times \text{RF1} + a_2 \times \text{RF2}
\]

- RF regression Model 1
  - Correctly predicted
  - Not correctly predicted

- RF regression Model 2

- RF 1 & 2 prediction
  - low
  - med
  - high

Weight factor calculation
# Forecast evaluation metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td>True positive/negative</td>
</tr>
<tr>
<td>False Alarms</td>
<td>False positive</td>
</tr>
<tr>
<td>Misses</td>
<td>False negative</td>
</tr>
</tbody>
</table>
| FAR (False Alarm Ratio)       | \[
|                               | \# of false alarms / \total \# of events forecast \] |
| POD (Probability of Detection)| \[
|                               | \# of hits / \total \# of events forecast \] |

*Image from www.deq.ok.gov*
Historical data summary

<table>
<thead>
<tr>
<th>Year</th>
<th>Simulated days</th>
<th># of days for each AQI</th>
<th>AQI &gt; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2015</td>
<td>106</td>
<td>75</td>
<td>27</td>
</tr>
<tr>
<td>2016</td>
<td>143</td>
<td>125</td>
<td>16</td>
</tr>
<tr>
<td>2017</td>
<td>114</td>
<td>71</td>
<td>35</td>
</tr>
<tr>
<td>2018</td>
<td>152</td>
<td>120</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>515</td>
<td>391</td>
<td>104</td>
</tr>
</tbody>
</table>

More fires
ML1 Evaluation

Leave one out cross validation

<table>
<thead>
<tr>
<th>Metric</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td>94 (100)</td>
<td>127 (130)</td>
<td>99 (92)</td>
<td>138 (140)</td>
</tr>
<tr>
<td>False Alarms</td>
<td>8 (1)</td>
<td>4 (0)</td>
<td>4 (6)</td>
<td>5 (0)</td>
</tr>
<tr>
<td>Misses</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>2 (7)</td>
<td>2 (5)</td>
</tr>
<tr>
<td>FAR</td>
<td>8% (1%)</td>
<td>3% (0%)</td>
<td>4% (5%)</td>
<td>3% (0%)</td>
</tr>
<tr>
<td>POD</td>
<td>91% (97%)</td>
<td>96% (98%)</td>
<td>94% (88%)</td>
<td>95% (97%)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis are the AIRPACT forecast performance.

- ML1 predicts more false alarms but fewer misses.
- For high O₃ year 2017, ML1 performs better than AIRPACT.
ML2 Evaluation

Leave one out cross validation

<table>
<thead>
<tr>
<th>Metric</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits</td>
<td>99 (100)</td>
<td>130 (130)</td>
<td>97 (91)</td>
<td>140 (141)</td>
</tr>
<tr>
<td>False Alarms</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>0 (6)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Misses</td>
<td>3 (2)</td>
<td>2 (2)</td>
<td>7 (7)</td>
<td>5 (5)</td>
</tr>
<tr>
<td>FAR</td>
<td>1% (1%)</td>
<td>0 (0)</td>
<td>0 (6%)</td>
<td>1% (0)</td>
</tr>
<tr>
<td>POD</td>
<td>96% (97%)</td>
<td>98% (98%)</td>
<td>93% (88%)</td>
<td>96% (97%)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis are the AIRPACT forecast performance.

- ML2 predicts much fewer false alarms but similar miss number as AIRPACT.
- Both AIRPACT and ML2 fail to predict the high ozone days in 2017.
Tri-Cities Ozone “Ensemble” Forecast in 2019

To get more data to train the model, we retrain our model everyday including previous day’s measurements.

Model Uncertainty

<table>
<thead>
<tr>
<th>Model</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML1</td>
<td>1.58%</td>
</tr>
<tr>
<td>ML2</td>
<td>1.82%</td>
</tr>
</tbody>
</table>

Time series of daily max. $O_3$

2015-2018 WRF met + 2019 WRF ensemble + time info + previous day’s observed $O_3$
ML2 performs the best to reduce false AQI2 days (in red cells). Thus we chose ML2 to run our operational daily ozone forecasting for Kennewick.
Our Machine Learning $O_3$ forecasts go public everyday!

http://ozonematters.com/
Summary

• The ML1 model raised more false alarms than AIRPACT, but performed better in the high ozone year.

• Both ML2 and AIRPACT missed some high ozone events, but ML2 raised fewer false alarms than AIRPACT.

• Our training dataset contains only a few high O₃ days, which makes it difficult to predict a high O₃ day using a ML approach. To overcome that issue, we updated the training dataset each day.

• We plan to apply our ML models to other cities that has a well-distributed AQI (Air Quality Index) values.
Thank you!