Machine learning based time series forecasting of chemical levels from the East Palestine, Ohio train derailment site

Vikas Ramachandra* & Mohit Sethi
*Associate Professor and CTO
vikas.ramachandra@jacobs.ucsd.edu

Abstract
The paper presents a time series analysis of the air, water and soil data from the East Palestine, Ohio, train derailment site. The goal of the analysis is to investigate the temporal pattern of chemical pollutant levels in air, water and soil and to build machine learning and statistical models for forecasting various chemical concentrations over time.

Introduction
The release of hazardous materials into the environment due to industrial accidents, transportation mishaps, and natural disasters poses a significant risk to public health and safety. Such incidents require prompt and effective response to mitigate the impact on the affected population and the surrounding environment. In the case of the train derailment that occurred in East Palestine, Ohio on 3 February 2023, 38 rail cars carrying hazardous material derailed, resulting in the release of hazardous material into the air, water, and soil.

The United States Environmental Protection Agency (EPA) and other agencies conducted air sampling in the area to monitor the concentration of the hazardous material and assess the potential risk to human health. The air sampling data collected at various locations in the vicinity of the incident site is a valuable source of information for evaluating the extent of contamination and predicting the future trend of the pollutant concentration. Time series analysis of the air sampling data can provide insights into the temporal behavior of the pollutant concentration and help identify any underlying patterns or trends.

In this paper, we perform a time series analysis of the air sampling data collected by the EPA following the East Palestine train derailment. Our objective is to identify any temporal patterns or trends in the pollutant concentration and evaluate the effectiveness of the remediation efforts. We focus on the "Result_Final_Txt" variable, which represents the concentration of the hazardous material in the air at different sampling locations over time. We then present our methodology for time series analysis and discuss the results of our analysis. Finally, we conclude with a discussion of the implications of our findings and the potential for future research in this area.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Dataset

EPA Lab Results: Air (through 02/22/23) - East Palestine, OH Response (csv) [1]
(https://www.epa.gov/oh/air-sampling-data-east-palestine-ohio-train-derailment)
The dataset contains 1553 records starting from 04/02/2023 to 22/02/2023. The dataset contains following features: ['Location', 'Samp_No', 'SampleDate_txt', 'SampleTime', 'Matrix', 'SampleMedia', 'Activity', 'Analytical_Method', 'CAS_NO', 'Analyte', 'Result_Units', 'Reporting_Limit', 'Validation_Level', 'Result_Final_Txt', 'Result_Qualifier_Final', 'RL_Comparison'].
Our analysis is focussed on the ‘Results_Final.Txt’ which represents the concentration level of the hazardous chemical. The ‘Analyte’ and ‘CAS_NO’ (CAS Registry Number) [4] help identify the chemicals. There are 49 unique chemicals. The records before 17/02/2023 are used for the training set(1036) and the remaining are used for the validation set(517).

EPA Lab Results: Water (through 02/14/23) - East Palestine, OH Response (csv) [2]
The dataset contains 1822 records starting from 04/02/2023 to 17/02/2023. The dataset contains following features: ['Location', 'Samp_No', 'SampleDate_txt', 'SampleTime', 'Matrix', 'SampleMedia', 'Activity', 'Analytical_Method', 'CAS_NO', 'Analyte', 'Result_Units', 'Reporting_Limit', 'Validation_Level', 'Result_Final_Txt', 'Result_Qualifier_Final', 'RL_Comparison'].
Our analysis is focussed on the ‘Results_Final.Txt’ which represents the concentration level of the hazardous chemical. The ‘Analyte’ and ‘CAS_NO’ (CAS Registry Number) [4] help identify the chemicals. There are 134 unique chemicals. The records before 10/02/2023 are used for the training set(1166) and the remaining are used for the validation set(656).

EPA Lab Results: Soil and Sediment (through 02/10/23) - East Palestine, OH Response (csv) [3]
The dataset contains 1082 records starting from 08/02/2023 to 10/02/2023. The dataset contains following features: ['Location', 'Samp_No', 'SampleDate_txt', 'SampleTime', 'Matrix', 'SampleMedia', 'Activity', 'Analytical_Method', 'CAS_NO', 'Analyte', 'Result_Units', 'Reporting_Limit', 'Validation_Level', 'Result_Final_Txt', 'Result_Qualifier_Final', 'RL_Comparison'].
Our analysis is focussed on the ‘Results_Final.Txt’ which represents the concentration level of the hazardous chemical. The ‘Analyte’ and ‘CAS_NO’ (CAS Registry Number) [4] help identify the chemicals. There are 128 unique chemicals. The records before 10/02/2023 are used for the training set(512) and the remaining are used for the validation set(570).

We combined the ‘SampleDate_txt’ and ‘SampleTime’ features into a single feature called ‘Date-Time’
Model & Results

Air Sample

We use the ‘Analyte’ feature to group the records for each chemical. There are 49 unique hazardous chemicals. We trained separate models for every chemical. We selected the following features for the prediction of our target variable 'Result_Final_Txt'. ‘Location’, ‘Date_Time’ 'SampleMedia', ‘Analytical_Method’, 'Reporting_Limit', 'Validation_Level', ', 'RL_Comparison'. We use the Augmented Dickey–Fuller test whether the time series is stationary or not. After experimentation with several forecasting models, ARIMA, SARIMA, LSTM Forecasting. We finally selected the FBProphet forecasting model as it best fitted our training data. FBProphet is a time series forecasting model developed by Facebook's Data Science team. It is designed to make accurate predictions of future values based on historical time series data, while also handling missing values, outliers, and trend changes. FBProphet uses a decomposable time series model with three main components: trend, seasonality, and holidays. It employs Bayesian inference to fit the model and provides uncertainty intervals for the predictions. We also use the metrics mean square error and mean absolute percentage error to evaluate the performance of the model in predicting the target variable 'Result_Final_Txt', which tells the concentration level of the hazardous chemical in the air. Fig.1, Fig.2, Fig.3, Fig.4 show the prediction of our model against the test data for 4 different chemicals.

Mean Square Error (MSE) is a commonly used statistical metric that measures the average squared difference between the predicted values and the actual values.

\[
MSE = \frac{1}{n} \sum (y_i - \hat{y})^2
\]

Mean Absolute Percentage Error (MAPE) is a statistical metric that measures the average absolute percentage difference between the predicted values and the actual values.

\[
MAPE = \frac{1}{n} \sum \frac{|y_i - \hat{y}_i|}{|y_i|} * 100\%
\]
Table 1 shows the performance metrics for each chemical in the Air Sample Data.

Table 1
<table>
<thead>
<tr>
<th>Chemical</th>
<th>MS/PEA Data</th>
<th>MAPEA Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethylbenzene</td>
<td>100.074</td>
<td>32.536</td>
</tr>
<tr>
<td>trans-1,2-Dichloroethene</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>1,1-Dichloroethane</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>Methyl tert-butyl ether</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>cis-1,2-Dichloroethene</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>Chloroform</td>
<td>0.0003</td>
<td>0.0286</td>
</tr>
<tr>
<td>Tetrachloroethene</td>
<td>0.1053</td>
<td>4.804</td>
</tr>
<tr>
<td>Chlorobenzene</td>
<td>0.0</td>
<td>0.0008</td>
</tr>
<tr>
<td>1,3,5-Trimethylbenzene</td>
<td>0.0006</td>
<td>0.1302</td>
</tr>
<tr>
<td>Styrene</td>
<td>0.0162</td>
<td>0.4133</td>
</tr>
<tr>
<td>o-Xylene</td>
<td>73.4363</td>
<td>17.6687</td>
</tr>
<tr>
<td>1,1,2,2-Tetrachloroethane</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>1,2-Dichloroethane</td>
<td>0.0055</td>
<td>0.1287</td>
</tr>
<tr>
<td>1,2,4-Trimethylbenzene</td>
<td>0.0706</td>
<td>2.2503</td>
</tr>
<tr>
<td>1,2-Dichlorobenzene</td>
<td>0.0</td>
<td>0.901</td>
</tr>
<tr>
<td>1,4-Dichlorobenzene</td>
<td>2.2509</td>
<td>8.1583</td>
</tr>
<tr>
<td>1,2-Dichlorobenzene</td>
<td>0.0</td>
<td>0.904</td>
</tr>
<tr>
<td>1,2-Dibromo 3,3-Dimethoxypropane</td>
<td>0.0</td>
<td>0.0916</td>
</tr>
<tr>
<td>1,2,4-Trichlorobenzene</td>
<td>0.0</td>
<td>0.0917</td>
</tr>
<tr>
<td>n-p-Xylene</td>
<td>2821.5551</td>
<td>35.8002</td>
</tr>
<tr>
<td>Methylisocyanide</td>
<td>135.6987</td>
<td>26.3974</td>
</tr>
<tr>
<td>cis-1,3-Dichloropropene</td>
<td>0.0</td>
<td>0.0001</td>
</tr>
<tr>
<td>Dibromochloromethane</td>
<td>0.0</td>
<td>0.001</td>
</tr>
<tr>
<td>Dichlorodifluoromethane (CFCl_2)</td>
<td>0.3424</td>
<td>0.2919</td>
</tr>
<tr>
<td>Chloromethane</td>
<td>0.0352</td>
<td>0.4955</td>
</tr>
<tr>
<td>Vinyl Chloride</td>
<td>28.423</td>
<td>49.5516</td>
</tr>
<tr>
<td>1,3 Butadiene</td>
<td>0.057</td>
<td>2.0566</td>
</tr>
<tr>
<td>Bromomethane</td>
<td>0.0005</td>
<td>0.4346</td>
</tr>
<tr>
<td>Chlorothene</td>
<td>0.0002</td>
<td>0.2781</td>
</tr>
<tr>
<td>Acrolein</td>
<td>0.2012</td>
<td>0.8925</td>
</tr>
<tr>
<td>Acetone</td>
<td>6.7423</td>
<td>0.64</td>
</tr>
<tr>
<td>Trichlorofluoromethane</td>
<td>56.5654</td>
<td>4.0665</td>
</tr>
<tr>
<td>1,2-Dibromoethane</td>
<td>0.0</td>
<td>0.901</td>
</tr>
<tr>
<td>1,1-Dichloroethene</td>
<td>0.0</td>
<td>0.901</td>
</tr>
<tr>
<td>Benzene</td>
<td>16.8874</td>
<td>3.1617</td>
</tr>
<tr>
<td>Carbon Tetrachloride</td>
<td>0.0106</td>
<td>0.198</td>
</tr>
<tr>
<td>1,2-Dichloropropene</td>
<td>0.0294</td>
<td>0.2135</td>
</tr>
<tr>
<td>Bromodichloromethane</td>
<td>0.0</td>
<td>0.901</td>
</tr>
<tr>
<td>Trichloroethane</td>
<td>0.0206</td>
<td>3.7841</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>0.0</td>
<td>0.9032</td>
</tr>
<tr>
<td>trans-1,3-Dichloropropene</td>
<td>0.0</td>
<td>0.0001</td>
</tr>
<tr>
<td>1,1,2-Trichloroethane</td>
<td>0.0</td>
<td>0.9008</td>
</tr>
<tr>
<td>Toluene</td>
<td>121.2752</td>
<td>1.6103</td>
</tr>
<tr>
<td>1,2-Dichloroethane</td>
<td>0.0911</td>
<td>2.2759</td>
</tr>
<tr>
<td>1,1,1-Trichloroethane</td>
<td>0.0</td>
<td>0.904</td>
</tr>
<tr>
<td>Hexachlorobutadiene</td>
<td>0.0</td>
<td>0.9016</td>
</tr>
<tr>
<td>Naphthalene</td>
<td>0.6466</td>
<td>3.1154</td>
</tr>
</tbody>
</table>
Water Sample

We use the exact same procedure for the Water Sample Data. We use the ‘Analyte’ feature to group the records for each chemical. There are 134 unique hazardous chemicals. We trained separate models for every chemical. Fig. 5, Fig. 6, Fig. 7, Fig. 8 show the prediction of our model against the test data for 4 different chemicals.

Table 2, 3, 4, 5 shows the performance metrics for each chemical in the Water Sample Data.
### Table 2

<table>
<thead>
<tr>
<th>Chemical</th>
<th>MSE(Water Data)</th>
<th>MAPE(Water Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24.66300000</td>
<td>4.3203</td>
</tr>
<tr>
<td>1</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>2</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>3</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>4</td>
<td>24.66300000</td>
<td>4.3203</td>
</tr>
<tr>
<td>5</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>6</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>7</td>
<td>7.03750000</td>
<td>2.238</td>
</tr>
<tr>
<td>8</td>
<td>2-Nitroaniline</td>
<td>5.7653</td>
</tr>
<tr>
<td>9</td>
<td>2-Methyphenol</td>
<td>5.7653</td>
</tr>
<tr>
<td>10</td>
<td>2-Methylnaphthalene</td>
<td>74.8085</td>
</tr>
<tr>
<td>11</td>
<td>2-Chlorophenol</td>
<td>5.7653</td>
</tr>
<tr>
<td>12</td>
<td>2-Chloronaphthalene</td>
<td>9.5397</td>
</tr>
<tr>
<td>13</td>
<td>2,6-Dinitrotoluene</td>
<td>5.7653</td>
</tr>
<tr>
<td>14</td>
<td>2,4-Dinitrotoluene</td>
<td>5.7653</td>
</tr>
<tr>
<td>15</td>
<td>1-Methylnaphthalene</td>
<td>14.0021</td>
</tr>
<tr>
<td>16</td>
<td>1,4-Dioxane</td>
<td>3.4628</td>
</tr>
<tr>
<td>17</td>
<td>1,2,4,5-Tetrachlorobenzene</td>
<td>3.9184</td>
</tr>
<tr>
<td>18</td>
<td>Pyridine</td>
<td>1.4139</td>
</tr>
<tr>
<td>19</td>
<td>1,1,1-Trichloroethane</td>
<td>2.238</td>
</tr>
<tr>
<td>20</td>
<td>1,1,2,2-Tetrachloroethane</td>
<td>2.238</td>
</tr>
<tr>
<td>21</td>
<td>1,1,2-Trichloroethane</td>
<td>2.238</td>
</tr>
<tr>
<td>22</td>
<td>1,1,2-Trichlorotrifluoroethane</td>
<td>2.238</td>
</tr>
<tr>
<td>23</td>
<td>1,1-Dichloroethane</td>
<td>2.238</td>
</tr>
<tr>
<td>24</td>
<td>Bromomethane</td>
<td>4.3203</td>
</tr>
<tr>
<td>25</td>
<td>1,1-Dichloroethene</td>
<td>2.238</td>
</tr>
<tr>
<td>26</td>
<td>1,2,3-Trichlorobenzene</td>
<td>4.3203</td>
</tr>
<tr>
<td>27</td>
<td>1,2,4-Trichlorobenzene</td>
<td>4.3203</td>
</tr>
<tr>
<td>28</td>
<td>1,2,4-Trichloroethene</td>
<td>2.238</td>
</tr>
<tr>
<td>29</td>
<td>1,2-Dibromo 3-Chloropropane</td>
<td>4.3203</td>
</tr>
<tr>
<td>30</td>
<td>GRO (C6-C10)</td>
<td>288.7863</td>
</tr>
</tbody>
</table>

### Table 3
<table>
<thead>
<tr>
<th></th>
<th>Substance</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>1,1'-Biphenyl</td>
<td>49.500200000</td>
<td>5.7653</td>
</tr>
<tr>
<td>32</td>
<td>1,1-Dichloropropene</td>
<td>58.702100000</td>
<td>5.8189</td>
</tr>
<tr>
<td>33</td>
<td>Carbon disulfide</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>34</td>
<td>Carbon Tetrachloride</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>35</td>
<td>Chlorobenzene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>36</td>
<td>2-Butanone</td>
<td>4.3682000000</td>
<td>0.4098</td>
</tr>
<tr>
<td>37</td>
<td>2-Hexanone</td>
<td>0.8190000000</td>
<td>0.1531</td>
</tr>
<tr>
<td>38</td>
<td>4-Methyl-2-pentanone</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>39</td>
<td>Acetone</td>
<td>921.818600000</td>
<td>2.3764</td>
</tr>
<tr>
<td>40</td>
<td>Benzene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>41</td>
<td>Bromochloromethane</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>42</td>
<td>Xylenes, Total</td>
<td>5.9787000000</td>
<td>1.1224</td>
</tr>
<tr>
<td>43</td>
<td>Bromodichloromethane</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>44</td>
<td>m,p-Xylenes</td>
<td>0.8952000000</td>
<td>0.3526</td>
</tr>
<tr>
<td>45</td>
<td>Methyl acetate</td>
<td>9.2621000000</td>
<td>1.1911</td>
</tr>
<tr>
<td>46</td>
<td>Methyl tert-butyl ether</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>47</td>
<td>Methylcyclohexane</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>48</td>
<td>Methylene chloride</td>
<td>115332.826400000</td>
<td>51.2957</td>
</tr>
<tr>
<td>49</td>
<td>o-Xylene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>50</td>
<td>Isopropylbenzene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>51</td>
<td>Pyrene</td>
<td>29.667200000</td>
<td>48.1067</td>
</tr>
<tr>
<td>52</td>
<td>Vinyl Chloride</td>
<td>10895673.570300000</td>
<td>1897.9536</td>
</tr>
<tr>
<td>53</td>
<td>Trichloroethene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>54</td>
<td>Chloroethane</td>
<td>24.663000000</td>
<td>4.3203</td>
</tr>
<tr>
<td>55</td>
<td>Chloroform</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>56</td>
<td>Chloromethane</td>
<td>24.663000000</td>
<td>4.3203</td>
</tr>
<tr>
<td>57</td>
<td>cis-1,2-Dichloroethene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>58</td>
<td>cis-1,3-Dichloropropene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>59</td>
<td>Cyclohexane</td>
<td>4.4239000000</td>
<td>0.8032</td>
</tr>
<tr>
<td>60</td>
<td>Trichlorofluoromethane</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>61</td>
<td>Dibromochloromethane</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>62</td>
<td>Ethylbenzene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
<tr>
<td>63</td>
<td>Styrene</td>
<td>7.037500000</td>
<td>2.238</td>
</tr>
</tbody>
</table>

Table 4
<table>
<thead>
<tr>
<th></th>
<th>Chemical Name</th>
<th>LogP</th>
<th>EDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>Tetrachloroethene</td>
<td>7.0375</td>
<td>2.238</td>
</tr>
<tr>
<td>65</td>
<td>Toluene</td>
<td>7.0375</td>
<td>2.238</td>
</tr>
<tr>
<td>66</td>
<td>trans-1,2-Dichloroethene</td>
<td>7.0375</td>
<td>2.238</td>
</tr>
<tr>
<td>67</td>
<td>trans-1,3-Dichloropropene</td>
<td>7.0375</td>
<td>2.238</td>
</tr>
<tr>
<td>68</td>
<td>Dibromodifluoromethane</td>
<td>58.7021</td>
<td>5.8189</td>
</tr>
<tr>
<td>69</td>
<td>Phenol</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>70</td>
<td>1,2,3-Trichloropropane</td>
<td>7.0375</td>
<td>2.238</td>
</tr>
<tr>
<td>71</td>
<td>Pentachlorophenol</td>
<td>485.3865</td>
<td>3.9184</td>
</tr>
<tr>
<td>72</td>
<td>Caprolactam</td>
<td>485.3865</td>
<td>3.9184</td>
</tr>
<tr>
<td>73</td>
<td>Butyl benzyl phthalate</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>74</td>
<td>Bis(2-ethylhexyl)phthalate</td>
<td>108.2798</td>
<td>4.4252</td>
</tr>
<tr>
<td>75</td>
<td>Bis(2-chloroethyl)ether</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>76</td>
<td>Bis(2-chloroethoxy)methane</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>77</td>
<td>Benzo(k)fluoranthene</td>
<td>0.4691</td>
<td>6.9062</td>
</tr>
<tr>
<td>78</td>
<td>Benzo(g,h,i)perylene</td>
<td>2.6049</td>
<td>16.0966</td>
</tr>
<tr>
<td>79</td>
<td>Benzo(b)fluoranthene</td>
<td>20.0234</td>
<td>36.5454</td>
</tr>
<tr>
<td>80</td>
<td>Benzo(a)pyrene</td>
<td>0.6199</td>
<td>7.198</td>
</tr>
<tr>
<td>81</td>
<td>Benzo(a)anthracene</td>
<td>137.3715</td>
<td>111.9364</td>
</tr>
<tr>
<td>82</td>
<td>Benzaldehyde</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>83</td>
<td>Atrazine</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>84</td>
<td>Anthracene</td>
<td>49.5100</td>
<td>64.9035</td>
</tr>
<tr>
<td>85</td>
<td>Phenanthrene</td>
<td>820.7950</td>
<td>172.379</td>
</tr>
<tr>
<td>86</td>
<td>Acenaphthylene</td>
<td>14.1423</td>
<td>32.1705</td>
</tr>
<tr>
<td>87</td>
<td>Acenaphthene</td>
<td>0.8281</td>
<td>7.9685</td>
</tr>
<tr>
<td>88</td>
<td>4-Nitrophenol</td>
<td>485.3865</td>
<td>3.9184</td>
</tr>
<tr>
<td>89</td>
<td>4-Nitroaniline</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>90</td>
<td>4-Chlorophenyl phenyl ether</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>91</td>
<td>4-Chloroaniline</td>
<td>28.4845</td>
<td>4.9227</td>
</tr>
<tr>
<td>92</td>
<td>4-Chloro-3-methylphenol</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>93</td>
<td>4-Bromophenyl phenyl ether</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>94</td>
<td>4,6-Dinitro-2-methylphenol</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>95</td>
<td>3-Nitroaniline</td>
<td>49.5002</td>
<td>5.7653</td>
</tr>
<tr>
<td>96</td>
<td>3,3'-Dichlorobenzidine</td>
<td>485.3865</td>
<td>3.9184</td>
</tr>
</tbody>
</table>

**Table 5**
<table>
<thead>
<tr>
<th></th>
<th>Compound</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>3,4-Methylphenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>98</td>
<td>2-Nitrophenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>99</td>
<td>Carbazole</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>100</td>
<td>Chrysene</td>
<td>1.81720000</td>
<td>12.239</td>
</tr>
<tr>
<td>101</td>
<td>Acetophenone</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>102</td>
<td>Hexachloroethane</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>103</td>
<td>Dibenzo(a,h)anthracene</td>
<td>0.91470000</td>
<td>9.5397</td>
</tr>
<tr>
<td>104</td>
<td>N-Nitrosodiphenylamine</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>105</td>
<td>N-Nitrosodi-n-propylamine</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>106</td>
<td>Nitrobenzene</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>107</td>
<td>Naphthalene</td>
<td>4.17720000</td>
<td>10.2575</td>
</tr>
<tr>
<td>108</td>
<td>Isophorone</td>
<td>380.15200000</td>
<td>3.4628</td>
</tr>
<tr>
<td>109</td>
<td>Indeno(1,2,3-cd)pyrene</td>
<td>4.31440000</td>
<td>18.5967</td>
</tr>
<tr>
<td>110</td>
<td>Hexachlorocyclopentadiene</td>
<td>380.15200000</td>
<td>3.4628</td>
</tr>
<tr>
<td>111</td>
<td>Hexachlorobutadiene</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>112</td>
<td>Hexachlorobenzene</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>113</td>
<td>Fluoranthene</td>
<td>224.06530000</td>
<td>112.7274</td>
</tr>
<tr>
<td>114</td>
<td>2,4-Dinitrophenol</td>
<td>380.15200000</td>
<td>3.4628</td>
</tr>
<tr>
<td>115</td>
<td>Fluorene</td>
<td>14.19970000</td>
<td>31.5878</td>
</tr>
<tr>
<td>116</td>
<td>2,4-Dichlorophenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>117</td>
<td>Dibenzofuran</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>118</td>
<td>Diethyl phthalate</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>119</td>
<td>Dimethyl phthalate</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>120</td>
<td>Di-n-butyl phthalate</td>
<td>39.39040000</td>
<td>5.9788</td>
</tr>
<tr>
<td>121</td>
<td>Di-n-octyl phthalate</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>122</td>
<td>2,4-Dimethylphenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>123</td>
<td>ORO (C28-C40)</td>
<td>42558591339820.28906250</td>
<td>20595.075</td>
</tr>
<tr>
<td>124</td>
<td>DRO (C10-C28)</td>
<td>78432168272756.54687500</td>
<td>36134.811</td>
</tr>
<tr>
<td>125</td>
<td>2,2'-Oxybis(1-chloropropane)</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>126</td>
<td>2,3,4,6-Tetrachlorophenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>127</td>
<td>2,4,5-Trichlorophenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
<tr>
<td>128</td>
<td>2,4,6-Trichlorophenol</td>
<td>49.50020000</td>
<td>5.7653</td>
</tr>
</tbody>
</table>

**Soil Sample**

We use the exact same procedure for the Soil Sample Data. We use the ‘Analyte’ feature to group the records for each chemical. There are 128 unique hazardous chemicals. We trained separate models for every chemical. Fig.9, Fig.10, Fig.11, Fig.12 show the prediction of our model against the test data for 4 different chemicals.
Table 6, 7, 8, 9 shows the performance metrics for each chemical in the Soil Sample Data.
### Table 6

<table>
<thead>
<tr>
<th>Chemical</th>
<th>MSE(Soil Data)</th>
<th>MAPE(Soil Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 2-Butanone</td>
<td>43.2635</td>
<td>13.2777</td>
</tr>
<tr>
<td>1 cis,1,2-Dichloroethene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>2 Chloromethane</td>
<td>0.0114</td>
<td>6.4255</td>
</tr>
<tr>
<td>3 Bromomethane</td>
<td>0.0114</td>
<td>6.4255</td>
</tr>
<tr>
<td>4 Fluoranthene</td>
<td>2621.26</td>
<td>101.3533</td>
</tr>
<tr>
<td>5 Hexachlorobutadiene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>6 Hexachlorobenzene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>7 Dichlorodifluoromethane (CFC 12)</td>
<td>0.0114</td>
<td>6.4255</td>
</tr>
<tr>
<td>8 Hexachlorocyclopentadiene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>9 Hexachloroethane</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>10 Fluorene</td>
<td>14.3053</td>
<td>29.7113</td>
</tr>
<tr>
<td>11 Ethylbenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>12 Methylcyclohexane</td>
<td>0.0107</td>
<td>0.7653</td>
</tr>
<tr>
<td>13 Methyl tert-butyl ether</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>14 Indeno(1,2,3-cd)pyrene</td>
<td>5018.5557</td>
<td>203.2561</td>
</tr>
<tr>
<td>15 Methylene chloride</td>
<td>10.5566</td>
<td>4.9879</td>
</tr>
<tr>
<td>16 o-Xylene</td>
<td>1.3968</td>
<td>12.7102</td>
</tr>
<tr>
<td>17 Styrene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>18 trans-1,3-Dichloropropene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>19 Trichloroethene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>20 m,p-Xylenes</td>
<td>0.0</td>
<td>0.0142</td>
</tr>
<tr>
<td>21 Methyl acetate</td>
<td>10.5566</td>
<td>4.9879</td>
</tr>
<tr>
<td>22 Tetrachloroethene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>23 Isopropylbenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>24 Isophorone</td>
<td>5021.6528</td>
<td>78.2781</td>
</tr>
<tr>
<td>25 Naphthalene</td>
<td>14.2888</td>
<td>21.8916</td>
</tr>
<tr>
<td>26 Nitrobenzene</td>
<td>5021.8528</td>
<td>78.2781</td>
</tr>
<tr>
<td>27 Trichlorofluoromethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>28 Vinyl Chloride</td>
<td>128895.3877</td>
<td>1488.9461</td>
</tr>
<tr>
<td>29 Xylenes, Total</td>
<td>75.8403</td>
<td>33.6355</td>
</tr>
<tr>
<td>30 1,4-Dichlorobenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
</tbody>
</table>

### Table 7
<table>
<thead>
<tr>
<th></th>
<th>Chemical Name</th>
<th>MW</th>
<th>LogD</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>1,3-Dichlorobenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>32</td>
<td>1,3,5-Trimethylbenzene</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>33</td>
<td>1,2-Dichloropropane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>34</td>
<td>1,2-Dichloroethane</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>35</td>
<td>1,2-Dichlorobenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>36</td>
<td>1,2-Dibromoethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>37</td>
<td>1,2-Dibromo 3-Chloropropane</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>38</td>
<td>1,2,4-Trimethylbenzene</td>
<td>10.3035</td>
<td>37.7154</td>
</tr>
<tr>
<td>39</td>
<td>1,2,4-Trichlorobenzene</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>40</td>
<td>2-Hexanone</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>41</td>
<td>1,2,3-Trichloropropane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>42</td>
<td>1,1-Dichloroethene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>43</td>
<td>Bromoform</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>44</td>
<td>1,1,2-Trichlorotrifluoroethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>45</td>
<td>1,1,2-Trichloroethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>46</td>
<td>1,1,2,2-Tetrachloroethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>47</td>
<td>1,1,1-Trichloroethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>48</td>
<td>Pyridine</td>
<td>5021.6528</td>
<td>78.2701</td>
</tr>
<tr>
<td>49</td>
<td>Pyrene</td>
<td>13.9359</td>
<td>3.1967</td>
</tr>
<tr>
<td>50</td>
<td>Phenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>51</td>
<td>Phenanthrene</td>
<td>2623.3958</td>
<td>150.8965</td>
</tr>
<tr>
<td>52</td>
<td>Pentachlorophenol</td>
<td>585.7008</td>
<td>104.9471</td>
</tr>
<tr>
<td>53</td>
<td>N-Nitrosodiphenylamine</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>54</td>
<td>N-Nitrosodi-n-propylamine</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>55</td>
<td>1,2,3-Trichlorobenzene</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>56</td>
<td>1,1-Dichloroethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>57</td>
<td>3,4-Methylenephene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>58</td>
<td>Di-n-butyl phthalate</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>59</td>
<td>Acetophenone</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>60</td>
<td>Acenaphthylene</td>
<td>14.3036</td>
<td>28.6866</td>
</tr>
<tr>
<td>61</td>
<td>Acenaphthene</td>
<td>14.3102</td>
<td>33.277</td>
</tr>
<tr>
<td>62</td>
<td>4-Chlorophenyl phenyl ether</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
</tbody>
</table>

Table 8
Table 9

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound</th>
<th>m/z</th>
<th>RT/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>4-Chloro-3-methylphenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>65</td>
<td>3-Nitroaniline</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>66</td>
<td>3,3'-Dichlorobenzidine</td>
<td>584.5026</td>
<td>30.4496</td>
</tr>
<tr>
<td>67</td>
<td>2-Nitrophenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>68</td>
<td>2-Nitroaniline</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>69</td>
<td>2-Methylnaphthalene</td>
<td>14.2888</td>
<td>21.8916</td>
</tr>
<tr>
<td>70</td>
<td>2-Chloronaphthalene</td>
<td>14.3369</td>
<td>93.4808</td>
</tr>
<tr>
<td>71</td>
<td>Anthracene</td>
<td>14.2806</td>
<td>19.3457</td>
</tr>
<tr>
<td>72</td>
<td>2,6-Dinitrotoluene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>73</td>
<td>2,4-Dinitrophenol</td>
<td>5023.3764</td>
<td>20.0552</td>
</tr>
<tr>
<td>74</td>
<td>2,4-Dimethylphenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>75</td>
<td>2,4-Dichlorophenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>76</td>
<td>2,4,6-Trichlorophenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>77</td>
<td>2,4,5-Trichlorophenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>78</td>
<td>2,3,4,6-Tetrachlorophenol</td>
<td>5021.3183</td>
<td>189.9926</td>
</tr>
<tr>
<td>79</td>
<td>2,2'-Oxybis(1-chloropropane)</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>80</td>
<td>1-Methylnaphthalene</td>
<td>5020.4261</td>
<td>587.4744</td>
</tr>
<tr>
<td>81</td>
<td>1,4-Dioxane</td>
<td>5021.6528</td>
<td>78.2781</td>
</tr>
<tr>
<td>82</td>
<td>1,2,4,5-Tetrachlorobenzene</td>
<td>584.3026</td>
<td>30.4496</td>
</tr>
<tr>
<td>83</td>
<td>1,1'-Biphenyl</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>84</td>
<td>GRO (C8-C10)</td>
<td>657261399.7043</td>
<td>522.2637</td>
</tr>
<tr>
<td>85</td>
<td>ORO (C28-C40)</td>
<td>2289338511484775.5</td>
<td>60217.3525</td>
</tr>
<tr>
<td>86</td>
<td>DRO (C10-C28)</td>
<td>8647921735.1756</td>
<td>223.7186</td>
</tr>
<tr>
<td>87</td>
<td>2,4-Dinitrotoluene</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>88</td>
<td>Atrazine</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>89</td>
<td>Benza[al]anthracene</td>
<td>14.1396</td>
<td>6.3967</td>
</tr>
<tr>
<td>90</td>
<td>Dimethyl phthalate</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>91</td>
<td>Diethyl phthalate</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>92</td>
<td>Dibenzofuran</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
</tbody>
</table>
Air Sample with Water Sample

Here we try to predict the target for Air Sample data by combining both Air Sample and Water Sample Data and using Water Sample data as an auxiliary variable. Table 10, 11 shows the performance metrics for each chemical.

Table 10

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Air Sample</th>
<th>Water Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butyl benzyl phthalate</td>
<td>5021.3185</td>
<td>189.9926</td>
</tr>
<tr>
<td>Bis(2-ethylhexyl)phthalate</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>Bis(2-chloroethyl)ether</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>Bis(2-chloroethoxy)methane</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>Benzo[k]fluoranthene</td>
<td>14.2381</td>
<td>12.6546</td>
</tr>
<tr>
<td>Benzo(g,h,i)perylene</td>
<td>14.2397</td>
<td>12.2321</td>
</tr>
<tr>
<td>Benzo(b)fluoranthene</td>
<td>14.0466</td>
<td>4.3756</td>
</tr>
<tr>
<td>Benzo(a)pyrene</td>
<td>14.1717</td>
<td>7.5604</td>
</tr>
<tr>
<td>Dibromochloromethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Cyclohexane</td>
<td>0.0114</td>
<td>0.4255</td>
</tr>
<tr>
<td>cis-1,3-Dichloropropane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Chloroform</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Chloroethane</td>
<td>436.4504</td>
<td>81.438</td>
</tr>
<tr>
<td>Chlorobenzene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>4,6-Dinitro-2-methylphenol</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>4-Bromophenyl phenyl ether</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>4-Nitroaniline</td>
<td>5021.6528</td>
<td>78.2781</td>
</tr>
<tr>
<td>4-Nitrophenol</td>
<td>584.5026</td>
<td>30.4496</td>
</tr>
<tr>
<td>Carbazole</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>Chrysene</td>
<td>14.1556</td>
<td>6.9301</td>
</tr>
<tr>
<td>Di-n-octyl phthalate</td>
<td>5021.2002</td>
<td>387.6414</td>
</tr>
<tr>
<td>Dibenz(a,h)anthracene</td>
<td>14.3136</td>
<td>36.171</td>
</tr>
<tr>
<td>Acetone</td>
<td>58181.8618</td>
<td>550.6706</td>
</tr>
<tr>
<td>Benzene</td>
<td>2610992.6339</td>
<td>3135.9938</td>
</tr>
<tr>
<td>Bromochloromethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Bromodichloromethane</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Carbon disulfide</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Carbon tetrachloride</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>4-Methyl-2-pentanone</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Caprolactam</td>
<td>584.4983</td>
<td>74.3885</td>
</tr>
<tr>
<td>trans-1,2-Dichloroethene</td>
<td>10.2881</td>
<td>40.352</td>
</tr>
<tr>
<td>Toluene</td>
<td>10.3493</td>
<td>32.64</td>
</tr>
<tr>
<td>Chemical</td>
<td>MSE(Air Data)</td>
<td>MAPE(Air Data)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1,1-Dichloroethane</td>
<td>0.0</td>
<td>0.0130</td>
</tr>
<tr>
<td>1,2-Dichloroethane</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>Trichloroethane</td>
<td>5.9247</td>
<td>139.3039</td>
</tr>
<tr>
<td>Chloroform</td>
<td>1.3211</td>
<td>0.9742</td>
</tr>
<tr>
<td>Bromoform</td>
<td>0.0809</td>
<td>0.4147</td>
</tr>
<tr>
<td>Benzene</td>
<td>11.369</td>
<td>64.5203</td>
</tr>
<tr>
<td>Acetone</td>
<td>29.4871</td>
<td>5.9773</td>
</tr>
<tr>
<td>Carbon Tetrachloride</td>
<td>13.5854</td>
<td>0.9008</td>
</tr>
<tr>
<td>1,1,2,2-Tetrachloroethane</td>
<td>0.0886</td>
<td>0.6756</td>
</tr>
<tr>
<td>Bromoformaldehyde</td>
<td>0.0913</td>
<td>0.7401</td>
</tr>
<tr>
<td>Trichloroethane</td>
<td>0.0224</td>
<td>2.221</td>
</tr>
<tr>
<td>1,4-Dichloroethane</td>
<td>0.0494</td>
<td>2.6689</td>
</tr>
<tr>
<td>trans-1,3-Dichloropropene</td>
<td>0.205</td>
<td>0.4983</td>
</tr>
<tr>
<td>Trichloroethylene</td>
<td>0.0096</td>
<td>30.2187</td>
</tr>
<tr>
<td>1,2-Dichloroethane</td>
<td>1.29319</td>
<td>0.0799</td>
</tr>
<tr>
<td>1,1,1-Trichloroethane</td>
<td>0.2014</td>
<td>4.0909</td>
</tr>
<tr>
<td>Hexachlorobutadiene</td>
<td>0.0813</td>
<td>2.0608</td>
</tr>
<tr>
<td>Naphthalene</td>
<td>0.6803</td>
<td>3.139</td>
</tr>
<tr>
<td>Vinyl Chloride</td>
<td>10.9839</td>
<td>13.8984</td>
</tr>
<tr>
<td>1,2-Dichloropropane</td>
<td>13.2683</td>
<td>0.8394</td>
</tr>
<tr>
<td>Cresylophloroglucin</td>
<td>0.3291</td>
<td>7.4835</td>
</tr>
<tr>
<td>1,2-Dibromo 3-Chloropropane</td>
<td>0.0289</td>
<td>0.5887</td>
</tr>
<tr>
<td>Name</td>
<td>ppm</td>
<td>ppm</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Dichloromethane</td>
<td>0.0113</td>
<td>0.2489</td>
</tr>
<tr>
<td>Dichlorofluoromethane (CCl2F2)</td>
<td>3.8515</td>
<td>54.3685</td>
</tr>
<tr>
<td>trans,1,2-Dichlorobenzene</td>
<td>5.9989</td>
<td>0.9850</td>
</tr>
<tr>
<td>1,3-Dichlorobenzene</td>
<td>0.0</td>
<td>0.001</td>
</tr>
<tr>
<td>Chlorobenzene</td>
<td>0.0</td>
<td>0.001</td>
</tr>
<tr>
<td>cis-1,2-Dichlorobenzene</td>
<td>0.0</td>
<td>0.004</td>
</tr>
<tr>
<td>Chloroform</td>
<td>0.0123</td>
<td>2.0694</td>
</tr>
<tr>
<td>Tetrahydrofuran</td>
<td>0.1106</td>
<td>2.8817</td>
</tr>
<tr>
<td>Chloroform</td>
<td>0.0221</td>
<td>1.2799</td>
</tr>
<tr>
<td>1,3,5-Trimethylbenzene</td>
<td>0.0090</td>
<td>0.3451</td>
</tr>
<tr>
<td>Styrene</td>
<td>0.0339</td>
<td>0.5798</td>
</tr>
<tr>
<td>p-Xylene</td>
<td>74.9000</td>
<td>29.5812</td>
</tr>
<tr>
<td>Isobenzene</td>
<td>137.6334</td>
<td>22.0647</td>
</tr>
<tr>
<td>1,1,2-Trichloro/nitrobenzene</td>
<td>0.0171</td>
<td>1.9922</td>
</tr>
<tr>
<td>1,2,4-Trichlorobenzene</td>
<td>0.0543</td>
<td>0.9239</td>
</tr>
<tr>
<td>1,3-Trichlorobenzene</td>
<td>0.2598</td>
<td>0.0988</td>
</tr>
<tr>
<td>1,4-Trichlorobenzene</td>
<td>2.6223</td>
<td>32.0237</td>
</tr>
<tr>
<td>1,3-Trichlorobenzene</td>
<td>0.5702</td>
<td>0.7569</td>
</tr>
<tr>
<td>1,4-Trichlorobenzene</td>
<td>0.0013</td>
<td>0.0829</td>
</tr>
<tr>
<td>p,p’-Dichlorobenzene</td>
<td>2391.5644</td>
<td>347.5865</td>
</tr>
<tr>
<td>Methylene chloride</td>
<td>1.39.063</td>
<td>14.556</td>
</tr>
<tr>
<td>1,1,2,2-Tetrachloroethane</td>
<td>0.1770</td>
<td>0.930</td>
</tr>
<tr>
<td>cis-1,2-Dichloropropane</td>
<td>0.0013</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
The mean and median for MSE for Air Sample Data is 73.4789 and 0.0006 respectively.

The mean and median for MAPE for Air Sample Data is 4.2328 and 0.198 respectively.

The mean and median for MSE for Water Sample Data is 938175511251.79 and 28.4845 respectively.

The mean and median for MAPE for Water Sample Data is 466.434 and 5.7653 respectively.

The mean and median for MSE for Soil Sample Data is 17885529841550.645 and 436.4504 respectively.

The mean and median for MAPE for Soil Sample Data is 663.1171 and 57.37025 respectively.

The mean and median for MSE for combined Air and Water Sample Data is 267.11765 and 0.0014 respectively.

The mean and median for MAPE for Soil Sample Data is 10.8287 and 0.1722 respectively.

**Conclusion**

In conclusion the time series forecasting of Ohio train derailment data has provided valuable insights into the patterns of the hazardous chemical present in air, water and soil.

The time series forecasting models used in this research have demonstrated strong predictive power when using Air Sample Data and Air Sample Data with Water Sample Data used as an auxiliary variable.

The relatively poor performance of the soil and water model with respect to the air model can be alluded to the fact that there is a lack of temporal data for both soil and water. While Air Sample Data has a total data collection of 18 days, data samples for water and soil have only been collected for 11 and 2 days respectively.

We hope that this research comes to use and the models can be used for forecasting hazardous chemical concentration over time.
References

7. https://github.com/facebook/prophet