**INTRODUCTION and MOTIVATION**

- Three-dimensional Eulerian chemical transport models such as CMAQ often report a significant model-measurement error due to uncertainties in the treatment of physical processes and also require higher run-time.
- Machine models are more computationally efficient and are currently used widely for forecasting purposes. Deep Neural network (DNN) techniques comprise a popular class of machine learning methods.
- Predicting hourly air quality, especially ozone, is challenging due to its highly varying and complex behavior in the atmosphere. Here, we used modeled meteorological parameters (by MCIP) along with selected gaseous species to improve the CMAQ model’s input for predicting future ozone concentrations.

**MATERIALS and METHODS**

**Deep Learning Algorithm**

- A deep convolutional neural networks (CNN) was evaluated for predicting hourly ozone concentration. Inspired by biological processes, CNN is a class of deep, feed-forward artificial neural networks.
- CNN uses relatively little pre-processing compared to other machine learning techniques. This means that the network learns the features that in traditional algorithms were hand-engineered.
- In CNN (Figure 1), the convolutional layer applies a convolution operation to the input, passing the result to the next layer. The fully connected layer connects every neuron in the last convolutional layer to every neuron in the output layer, similar to the traditional multi-layer perception neural network (MLP).

**Figure 1: Schematic for the regressive 1D convolutional neural networks (ConvNet - CNN) model.**

<table>
<thead>
<tr>
<th>meteorological variables</th>
<th>chemical species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature/ground</td>
<td>Nitrogen dioxide</td>
</tr>
<tr>
<td>humidity</td>
<td>Formaldehyde</td>
</tr>
<tr>
<td>Pressure</td>
<td>Hydroxide</td>
</tr>
<tr>
<td>solar radiation</td>
<td>Hydroxyde</td>
</tr>
<tr>
<td>wind speed/10 m</td>
<td>PBL height</td>
</tr>
<tr>
<td>latent heat flux</td>
<td>Total Cloud Fraction</td>
</tr>
<tr>
<td>sensible heat flux</td>
<td>Ozone</td>
</tr>
<tr>
<td>liquid water content</td>
<td>Precipitation</td>
</tr>
<tr>
<td>relative humidity</td>
<td>Aerosol</td>
</tr>
</tbody>
</table>

Since the CMAQ is used in this model, some limitations of the CNN model can be mitigated, especially those related to interpreting the physical variables and their relationship with ozone concentrations, and availability of important meteorological predictors such as PBL height and chemical predictors such as OH.

**Column 1:** Meteorological variables (left) and chemical species (right) used as input in CMAQ-CNN model.

**RESULTS**

- ConvNet modeling time period: 2011 – 2013
- Training data: 2011 – 2013
- Real-time prediction: 2014
- We trained/tested the model only on ozone season (April-October).
- The focus is to improve CMAQ output concentration, particularly in predicting ozone peaks.

**Figure 2: Monthly mean of maximum ozone for 7 months of ozone season (April-October) in 2014.**

**Table 1:** Meteorological variables (left) and chemical species (right) used as input in CMAQ-CNN model.

**RESULTS (cont.)**

We developed two different CMAQ-CNN models:
- **CNN generalization:** we trained/tested the model by combining all input samples from the entire domain (United States).
- **CNN Standalone:** we trained/tested the model for each station individually.

**Figure 3:** Scatter plots of CMAQ and CMAQ-CNN models for 2014 ozone season (April-October).

**Figure 4:** Taylor diagrams showing the performance of all models in different months of 2014 averaged over 1084 AQS stations in the continental US.

**Figure 5:** Monthly mean ozone concentration of all models 1084 AQS stations in the continental US.

**Key message:**
- The CMAQ model generally overpredicted ozone peaks during the ozone season (Figure 2). This overprediction was more pronounced in central and southeast United States.
- The CMAQ-CNN model recovered the CMAQ model’s overprediction (Figure 2). However, the CMAQ-CNN generally underpredicted the high ozone peaks.
- The CMAQ-CNN model predicts next day’s hourly ozone concentration with promising improvement in accuracy (IOA) (Figure 3). The monthly median of the daily IOAs was mostly above 0.8 threshold indicating reasonable prediction level.
- The number of low accuracy days was also decreased significantly for the CMAQ-CNN model.

**ACKNOWLEDGEMENTS**

This study was partially supported by funding from the Department of Earth and Atmospheric Sciences (EAS Research Grant) of the University of Houston.

**corresponding author:** Dr. Yunsoo Choi, PI of UH EAS ICAS air quality forecasting/modeling group, email: ychoi6@uh.edu