

# Machine learning-based air quality simulations over the United States under multiple climate change scenarios

Monday 6 December 2021 13:15 (15 minutes)

Air quality regulations have reduced emissions of pollutants in the U.S., but many prognostic studies suggest that future air quality might be degraded by global climate change. The simulated climate by various climate models shows a large variation in the future decades, and it is important to account for such variations to study future air quality. A typical approach to study future air quality projections uses three-dimensional (3D) Eulerian models, but these models are computationally too expensive to perform an ensemble of long-term simulations for various climate projections. Therefore, we have developed a machine learning (ML) based air quality model to study, in an efficient way, how future air quality might be influenced by climate change. Our ML model uses two-phase random forest to predict the O<sub>3</sub> and PM<sub>2.5</sub> concentrations with training datasets of key meteorological information and air quality pollutant emissions. To evaluate the model performance, we used the input datasets for the U.S. Environmental Protection Agent (EPA) the Community Multiscale Air Quality Modeling System (CMAQ) simulations and compared our model predictions against the CMAQ output as a benchmark. The 1995–1997 data were used to train the ML model; 2025–2035 data were used to evaluate it. The ML model is well performed for hourly O<sub>3</sub> predictions over the whole domain in four selected months (January, February, July, and August), and the R<sup>2</sup> values are in 0.5–0.7, the normalized mean bias (NMB) values are within  $\pm 3\%$ , the overall normalized mean error (NME) values are below 20%. Compared to CMAQ, our ML model tends to overpredict the O<sub>3</sub> in the Southeast U.S and California, and underpredict in the Central U.S, and the NMB values computed for each grid are generally within  $\pm 10\%$ . Predicting PM<sub>2.5</sub> is more challenging than predicting O<sub>3</sub>, but our ML model performance is still acceptable. The overall R<sup>2</sup> values of PM<sub>2.5</sub> predictions are in 0.4–0.6, and the NMB values are within  $\pm 6\%$ , but the NME can be up to 60%. The NMB in each grid is within  $\pm 30\%$ . There is no clear trend for the regional variation of ML model performance for PM<sub>2.5</sub>. Our ML model performs better for summer PM<sub>2.5</sub> (July and August) than winter (January and February): NME is 10%–20% lower in summer. While the model performs better in winter than summer with about 10% lower NME for O<sub>3</sub>. Our ML model with GPU acceleration runs less than one hour using a single GPU processor to predict 11-year one-month (total 11 months) simulations. It uses significantly less computing resources compared to the 3D models, like CMAQ, while it results in comparable predictability to CMAQ. It shows that our ML model a reliable and efficient tool to assess the air quality under various climate change scenarios.

## Physical Presentation

I am not so sure at this point if I would present physically.

**Primary authors:** FAN, Kai; Dr LEE, Yunha

**Presenter:** FAN, Kai

**Session Classification:** Parallel Session