**Using a Machine Learning method to Make Projections of tropospheric Ozone Along Different** **Shared Socioeconomic Pathways**

**1. INTRODUCTION**

P1 对流层臭氧背景

Tropospheric ozone (O3) is one of the primary air pollutions that have adverse effect on public health and environment. Tropospheric ozone concentration is affected by anthropogenic emission and climate change. Chinese government has put enormous efforts into reducing anthropogenic sources emissions. China is committed to reaching its carbon emission peak by 2030 and fulfilling carbon neutrality target by 2060 to mitigate global warming. Under the implementation of a series of stringent emission reduction policies, the primary precursor emission will reduce substantially. Hence, future air pollution trends have been a great concern to both air pollution control and climate policy making.

P2 SSP情景

Intergovernmental Panel on Climate Change (IPCC) has developed a set of global emission and radiative forcing scenarios. These scenarios are produced by integrated assessment models (IAMs) to describe future changes in population, socioeconomics, science and technology, energy consumption and land use, etc., along with associated GHG and pollutant emissions (Moss et al., 2010). The new generation of global scenarios combining shared socio-economic pathways (SSPs) with climate forcing outcomes as described by the Representative Concentration Pathways (RCPs) can reflect plausible future emissions and climate conditions on a global scale (Rogelj et al., 2018). Principally, the SSPs represent a variety of levels of climate mitigation and adaptation policy strength to control emissions that include tropospheric O3, O3 precursors, and aerosols (O’Neill et al.,2014; van Vuuren et al., 2014). SSPs scenarios have been used in atmospheric chemistry and Earth system model simulations to examine future changes in air pollution (Rao et al., 2017). A few published works provided predictable air pollution data under different SSPs. Rao et al. (2017) estimated annual average PM2.5 concentration as well as six-month average ozone concentrations over the 21st century under SSP1 to SSP5. Liu et al., (2022) examined the tropospheric O3 and surface O3 sensitivity under present days (2004–2014) and future conditions (2045–2055) under a range of SSPs using a chemistry-climate model, United Kingdom Earth System Model (UKESM1). Considering the newly published regional emission control policies and updating combustion–production technologies in China, Tong et al. (2020) developed a Dynamic Projection model for Emissions in China (DPEC) to project the dynamic anthropogenic emission pass-ways from 2015 to 2050 connecting multiple scenarios. Current works focus on future atmospheric pollutant concentrations projection, lacking the quantitative contribution of meteorology to surface ozone in the absence of climate and radiative forcing under different strength of SSP scenarios. Moreover, most published studies relied on atmospheric chemical models, which demand considerable computer computing resources and are generally constrained by biased chemical transport mechanisms as well as the uncertainty of emission inventories.

P3 机器学习的方法

Numerous machine learning models have been widely used and well developed in simulating air pollutants (Geng et al., 2021; Ma et al., 2020; Sun et al., 2021). In the area of ambient ozone simulation, machine learning technologies even showed higher modelling performance compared to traditional chemistry in large-scale, long duration, and high spatial and temporal resolution research. Ma et al. (2020) applied an RF model to estimate ozone metrics, including daily O3-8hmax, O3-mean, and O3-1hmax, from 2010 to 2017 in Beijing-Tianjin-Hebei region in China. The RF model achieved high performance with the coefficient of determination value R2s for three indicators all higher than 0.80. Liu et al. (2020) used eXtreme Gradient Boosting (XGBoost) model to simulate nationwide daily maximum 8-hour average (MDA8) in China, and R2 were from 0.60 to 0.87 at the month level in different years. Similarly, the other popular machine learning model, Artificial Neural Network model (ANN), also been implemented for ambient ozone simulation. Di et al. (2017a) established a convolutional neural network model that considered multi-source dataset to simulate the ambient ozone exposure in the United States from 2000 to 2012. The Cross-validated R2 on the testing monitoring sites ranged from 0.74 to 0.80. Though ML models performed well in ozone simulation, few studies combine multi-source data to predict ozone in different SSPs scenarios. These research results indicate that ML models perform well in both historical surface ozone temporal variability and spatial interpolation projection. Therefore, applying ML models to predict ozone concentration under different SSPs and to quantify the contribution of meteorological factors might improve the ambient ozone spatiotemporal distribution forecast accuracy, helping to reduce exposure risks.

P5

This study explores the use of machine learning (ML) techniques to make accurate, computationally inexpensive projections of tropospheric ozone under Shared Socioeconomic Pathways scenarios (SSPs). Several ML models are investigated: Generalize Linear Model (GLM), Extreme gradient boosting model (XGBoost), Random Forest model (RF), and Artificial Neural Network model (ANN). Model features included meteorological variables, ozone monitoring data, emission inventory variables, and chemical transport model output variables. This study aims to provide statistical models to predictively quantify the impact of future climate and emissions on ozone concentrations under different Shared Socioeconomic Pathways.

**2. MATERIALS AND METHODS**

Figure S1. Illustrates the modeling framework of the projection.

**2.1. Multisource Input Data**

**Table 1.** documents all the input data set used in this study, including input data variables information, spatial and temporal resolution, and data sources.

Ozone observation data on 1034 sites from 2013 to 2017 was restrained from China National Monitoring Centre.

Meteorological field data under future scenarios is obtained from the ScenarioMIP multi-model simulations in CMIP6.

**2.2. Machine Learning Methods Description**

In this study the applicability of four different machine learning methods for making projections of tropical stratospheric column ozone is assessed. These methods are Random Forests, eXtreme Gradient Boosting, and ANN.

**3. RESULTS**

**3.1. Model Performance in ozone Prediction**

The year 2020 is used as a validation.

**3.2. Projection of Surface Ozone Variations in the Future**

**3.3 Climate-Driven Surface Ozone Variations**

**4. DISCUSSION**

**References**