**Using Machine Learning to Quantitatively Predict Climate Impacts on Surface Ozone Along Different** **Shared Socioeconomic Pathways**

# Introduction

China has implemented a set of strict emission control policies (Clean Air Action Plan) since 2013 to fight against severe air pollution. Consequently, the level of fine particulate matter (PM2.5), inhalable particles (PM10), nitrogen oxides and sulfur dioxide in major urban clusters has decreased dramatically. However, ambient ozone (O3) concentrations show an upward trend in recent years, and its health hazards have also been recognized. Chinese government has put enormous efforts into reducing anthropogenic sources emissions. Meanwhile, 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 the community. What is unrevealed is the quantified impact of climate and emission under different policy interventions. Will global warming and climate change offset the effort of ozone precursors emission reduction?

The simulated absolute values of chemical transport models remain biased due to uncertainties in emission inventories and the lack of certain physical and chemical processes in the models. 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 will be 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.

# Literature Review

High-level surface ozone pollution has harmful effects on public health, such as damaging human lungs and skins (Day et al., 2017). Both short-term and long-term exposure to ambient O3 link to total mortality, cardiovascular diseases, and respiratory diseases (Bell and Michelle, 2004; Yin et al., 2017; Wong et al., 2008; Di et al., 2017). High-level ozone also destroys crops and forest vegetation (Sawada and Kohno, 2010; Avnery and Mauzerall, 2011). Though the background ozone concentration in the northern hemisphere remains unchanged (David et al., 2020), the average ozone concentration in China has increased in the past decade, and urban areas have undergone even more robust ozone pollution (CNEMC, 2020). Surface ozone pollution has become one of the most crucial pollutants influencing air quality compliance rate in densely populated cities of China.

Some studies proved that stratospheric flux (Ordonez et al., 2007; Hess and Zbinden, 2013) and changes in long-distance transport patterns (Pausata et al., 2012) are partial factors influencing the concentration variation of tropospheric ozone. However, plenty of evidence supports that the trend variation of ozone concentration is mainly attributed to the co-effect of ozone precursors variation and meteorological effect (Li and Jacob et al., 2019). Previous studies evidenced that the inappropriate VOCs-to-NOx reduction ratio caused by stringent anthropogenic emission reduction measures is one of the most significant factors leading to recent years’ ambient ozone concentration rapid increase in China (Zheng et al., 2018; Yu et al., 2019). Meanwhile, meteorological conditions affect the photochemical reaction and regional transmission process of surface ozone (Lu and Zhang, 2019), which may either promote or dampen the increase of ozone concentration. In the short-term, specific meteorological conditions are favorable to the photochemical reaction and persistence of O3 such as hot, dry, stagnant weather (Gong and Liao, 2019; Han et al., 2020). Moreover, meteorological factors affect the dissipation and long-distance transmission process of tropospheric ozone (Gao et al., 2020; Ni et al., 2018; Yang et al., 2014). In the long-term, with global warming, meteorological conditions may become increasingly favorable to the photochemical generation of surface ozone (Delcloo et al., 2016; Fu et al., 2015). Under the influence of climate change, atmosphere conditions in some regions such as North China Plain are becoming more stable, possibly increasing the risk of severe air pollution events (Cai et al., 2017; Chen et al., 2018; Han et al., 2017). Thus, complex meteorological conditions and global climate change both interfere with diagnosing the causes of ozone pollution and make it more difficult to formulate appropriate long-term ozone pollution control strategies. Improving the effectiveness of ozone control policies requires quantitatively evaluating the contribution of precursors emission changes and meteorological or climate changes to the trend of ozone concentration.

Many previous studies have applied statistical models or chemical transportation models to investigate the quantitative contribution of meteorological conditions and anthropogenic precursors to the ozone concentration and understand how both factors influence surface ozone pollution in China. For example, the changes in meteorological parameters led to summer surface ozone concentration varied 2-5% in central China from 1986 to 2006 (Yang et al., 2014). From 2013 to 2017, the anthropogenic emissions contributed 1-3 ppbv·a-1 surface ozone increment of the eastern urban area (Li and Jacob, 2019). In contrast, the meteorological factors change daily ozone by -3.5 to 8.5 ppbv (Ding et al., 2019). Li and Jacob (2020) revealed that meteorology contributed 0.7 ppbv·a-1 across China and 1.4 ppbv·a-1 over the North China Plain from 2013 to 2019. From 2007 to 2017, meteorology mitigates the increase of surface ozone by 15% in the Pearl River Delta (PRD) region (Yang et al., 2019). These studies revealed notable spatiotemporal heterogeneity in the predominant factors affecting ground-level ozone. What is unrevealed is whether the effort of anthropogenic emissions control policies will be offset by climate change and abnormal meteorological conditions. Conversely, different levels of economic development and strengths of anthropogenic emission control policies will lead to different climate warming outcomes, which in turn have other potential effects on surface ozone concentration trends.

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)](https://www.nature.com/articles/s41558-018-0091-3" \t "_blank). 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. However, 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.

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 O3 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 O3-8h max, O3-mean, and O3-1h max, from 2010 to 2017 in Beijing-Tianjin-Hebei region in China. The RF model achieved high performance with R2s for three indicators all higher than 0.80. Liu et al. (2020) used XGBoost model to simulate nationwide daily maximum 8-hour average (MDA8) in China, and the R2 value were from 0.60 to 0.87 at the month level in different years. Similarly, the other popular machine learning model, ANN, also been implemented for ambient ozone simulation. Di et al. (2017a) established a convolutional neural network model that considered multi-source datasets 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, including both emission inventories and reanalyzed meteorological variables to predict surface ozone in different SSPs scenarios. These research results indicate that ML models perform well in both past times ambient ozone spatiotemporal variability simulation and spatial interpolation simulation. 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.

# Research questions

1. Could ML models effectively predict ozone concentrations in different SSP scenarios? Could ML models perform better than traditional chemical transport models?
2. How will future meteorological conditions and emissions contribute to surface ozone trend under different SSPs?

# Research aims

1. To predict daily surface ozone concentration under different SSPs scenarios from 2020 to 2060.

2. To quantify the impact of future climate and meteorology to surface ozone along with different SSPs emission scenarios.

# Data and Methods

The feature variables included in ML models are determined based on previous simulation studies as well as data availability (Ma, et al., 2020). Figure. 1 illustrates the modeling framework of the ozone simulation, including input data obtained from multiple data sources, processing of the input data, fusing the multi-source data and generating ozone prediction.

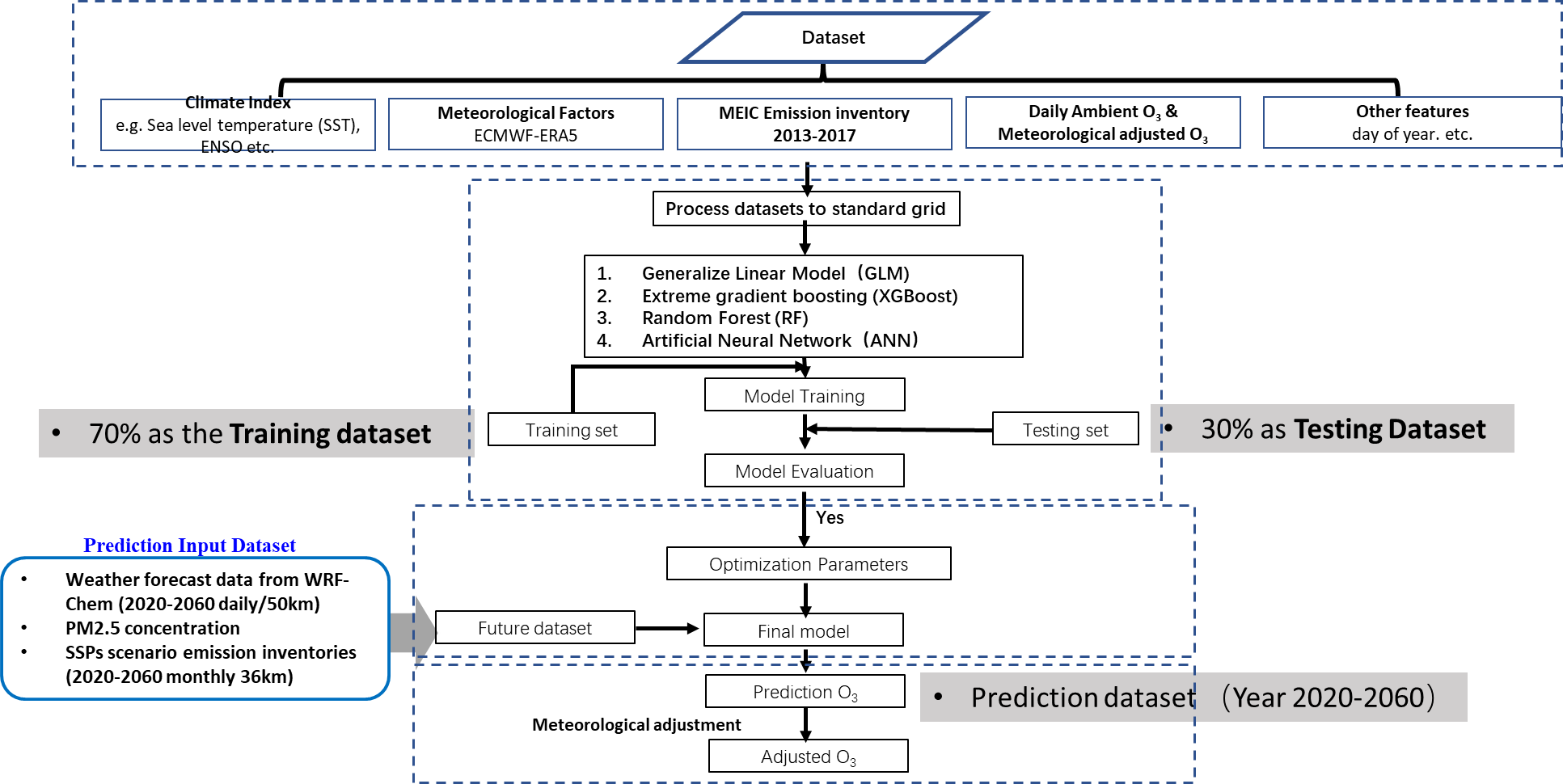


Figure 1 Operational process of the ozone projection.

## Ozone monitoring data

Hourly ozone monitoring data from 2013 to 2021 is collected from the national air quality monitoring network (http://www.cnemc.cn/) in China, which includes more than 1600 air pollution monitoring stations over China. And then the MDA8 ozone concentration is calculated. Gridded O3 and PM2.5 concentration data are retrieved from Tracking Air Pollution in China (TAP, http://tapdata.org.cn/) database as well to facilitate matching with gridded inventory dataset. Detailed information on the gridded ozone data generation method can be found in Geng et al. (2021).

## Historical and future emission inventories under different SSPs

Anthropogenic emission inventories are taken from the Dynamic Projection model for Emissions in China (DPEC, http://meicmodel.org/), which dynamically project and assess China’s future emissions (including both greenhouse gases and air pollutants) in the context of socio-economic development, global climate adaption, national carbon peak and carbon neutrality target, synergy pollution mitigation and carbon reduction pathways. DPEC published six emission scenario datasets that connecting five SSP scenarios (SSP1–5), five RCP scenarios (RCP8.5, 7.0, 6.0, 4.5, and 2.6), and three pollution control. The monthly DPEC emission inventory at a spatial resolution of 36km, covering the period from 2015 to 2060. Detailed information of DPEC emission inventories’ development and basis can be found in Tong et al., (2020) and Cheng et al., (2021).

## Meteorology reanalysis data

The historical meteorological data from 2013 to 2021 is retrieved from the European Center for Medium-range Weather Forecast (ECMWF) ERA5 6/12 h reanalysis dataset. The ERA5 reanalysis dataset is available each hour with a horizontal resolution of 0.125°× 0.125°. We utilize the following parameters consisting of our previous study (Yang et al., 2019). Meteorological variables are calculated as daily 10m mean u-component of wind (U, m·s-1), daily 10m mean v-component of wind (V, m·s-1), daily maximum 2m temperature (T, ℃), daily minimum relative humidity at 1000 hPa (RH, %), daily total sky direct solar radiation at the surface (SSR, J·m-2), daily total precipitation (TP, mm), etc.

## The chemical transport model output

Meteorological projection data and atmospheric pollutant data under different SSP scenarios came from outputs of the WRF-Chem chemical transport model. The spatial and temporal resolution of the model was 36km and 1 h, respectively. To simulate and verify the prediction performance of ozone concentration in different future scenarios, PM2.5 and meteorological parameters under different SSPs from 2020 to 2060 will be used in the prediction dataset. Ambient O3 concentrations generated by WRF-Chem will be used as validation and comparison of ML ozone simulation results.

## Dataset preprocessing

Before conducting the spatiotemporal simulation of ambient O3 concentration, all variables will be converted to gridded data and assigned to grid cells based on sites’ coordinates by inverse distance weighted interpolation methods (Ma et al., 2021). Each day is assigned a corresponding month or year level value for datasets’ temporal resolution matching. A processed dataset containing all O3 monitoring data and environmental variables are prepared for the next step in model development. Low contribution variables might be excluded in the final simulation because of the low spatiotemporal resolution and high missing value.

## Machine Learning Models application and validation

Both random forest and neural network models are black-box models that enable fast computation of large-scale pollution simulations when exploring the relationship between ozone concentration and different variables. The RF model was developed by including multiple decision trees that generated by the bagging ensemble method (Breiman et al., 2001). The algorithm usage of artificial neural network and XGboost refer to Di et al. (2017) and Liu et al. (2020), respectively. The advantages and limitations of above models are described by Ma et al. (2020). All modeling work will built by R and Python software.

The sample-based division method randomly divides data into a training dataset and test dataset, where the training dataset includes 90% of the total data. The training dataset will be used to build models. The test set included 10% of the data, and test subsets will validate the model performance.

ML validation will use the cross-validation (CV) method, which can be divided into two types. The first type is leave-out-one cross-validation (Wolf et al., 2017), which is the most common CV method, referred as Geisser (1974). The other is V-fold CV method (Son et al., 2018).Both two types of cross-validation methods will be used for result validation.

## XX

# Expected Results

This research will simulate daily tropospheric ozone concentration along with different SSP scenarios and contributions of meteorological conditions and emissions to ozone under different SSP scenarios. The research attempts to provide a set of machine learning approaches combined with chemical transport model simulation results to predict future links between ozone and climate change.

# Research Planning

|  |  |
| --- | --- |
| **Timescale** | **Experiment schedule** |
| May 2022 | Data preprocessing, ML models evaluation and input feature variables screening. |
| June 2022 | Assessing the performance of the above-mentioned ML models. Simulating ozone concentrations under 5 SSPs scenarios. |
| July 2022 | Developing the method to quantify the impact of climate on ozone change. |
| August 2022 | Validating machine learning model prediction results. |
| October 2022 | Collating experimental results and data analysis. |
| November 2022 | Writing the thesis. |

# Bibliography

BELL & MICHELLE, L. (2004), "Ozone and short-term mortality in 95 US urban communities, 1987-2000.", Jama, Vol. 292 No. 19, pp. 2372-2378.

CAI, W., LI, K., LIAO, H., WANG, H. & WU, L. (2017), "Weather conditions conducive to Beijing severe haze more frequent under climate change", Nature Climate Change, Vol. 7 No. 4, pp. 257–262.

CHEN, S., GUO, J., SONG, L., LI, J., LIU, L. & BLAKE, C. J. (2018), "Inter-annual variation of the spring haze pollution over the North China Plain: Roles of atmospheric circulation and sea surface temperature", International Journal of Climatology,.

CHENG, G. & HONG, L. (2019), "A typical weather pattern for ozone pollution events in North China", Atmospheric Chemistry and Physics, Vol. 19 No. 22, pp. 13725-13740.

CHENG, J., TONG, D., ZHANG, Q., LIU, Y., LEI, Y., YAN, G., YAN, L., YU, S., CUI, R. Y. & LEON, C. (2021), "Pathways of China's PM2.5 air quality 2015–2060 in the context of carbon neutrality", National Science Review, No. 12, pp. 12.

DELCLOO, A. W., DUCHÊNE, F., HAMDI, R., BERCKMANS, J., DECKMYN, A. & TERMONIA, P. (2016), "The Impact of Heat Waves and Urban Heat Island on the Production of Ozone Concentrations Under Present and Future Climate Conditions for the Belgian Domain", in International Technical Meeting on Air Pollution Modelling & Its Application, pp.

DI, Q., WANG, Y., ZANOBETTI, A., WANG, Y. & SCHWARTZ, J. D. (2017), "Air Pollution and Mortality in the Medicare Population", New England Journal of Medicine, Vol. 376 No. 26, pp. 2513-2522.

DING, D., XING, J., WANG, S., CHANG, X. & HAO, J. (2019), "Impacts of emissions and meteorological changes on China's ozone pollution in the warm seasons of 2013 and 2017", Frontiers in Environmental Science and Engineering (English), Vol. 13 No. 5, pp. 9.

FU, T. M., ZHENG, Y., PAULOT, F., MAO, J. & YANTOSCA, R. M. (2015), "Positive but variable sensitivity of August surface ozone to large-scale warming in the southeast United States", Nature Climate Change, Vol. 5 No. 5, pp. 454-458.

GAO, M., GAO, J., ZHU, B., KUMAR, R. & MCELROY, M. B. (2019), "Ozone Pollution over China and India: Seasonality and Sources".

GEISSER, S. (1974), "A Predictive Approach to the Random Effect Model", Biometrika, Vol. 61 No. 1, pp. 101-107.

GENG, G., XIAO, Q., LIU, S., LIU, X., CHENG, J., ZHENG, Y., XUE, T., TONG, D., ZHENG, B. & PENG, Y. (2021), "Tracking Air Pollution in China: Near Real-Time PM2.5 Retrievals from Multisource Data Fusion", Environmental Science and Technology, Vol. 55.

HAN, Z., ZHOU, B., XU, Y., WU, J. & SHI, Y. (2017), "Projected changes in haze pollution potential in China: an ensemble of regional climate model simulations", Atmospheric Chemistry and Physics, Vol. 2017 No. 16, pp. 1-46.

HESS, P. G. & ZBINDEN, R. (2013), "Stratospheric impact on tropospheric ozone variability and trends : 1990-2009.", ATMOSPHERIC CHEMISTRY AND PHYSICS, Vol. 13 No. 2, pp.

LI, K., JACOB, D. J., LIAO, H., SHEN, L., ZHANG, Q. & BATES, K. H. (2019), "Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China", Proceedings of the National Academy of Sciences, Vol. 116 No. 2, pp. 422-427.

LIU, R., MA, Z., LIU, Y., SHAO, Y., ZHAO, W. & BI, J. (2020), "Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach", Environment International, Vol. 142105823.

LIU, Z., DOHERTY, R. M., WILD, O., O'CONNOR, F. M. & TURNOCK, S. T. (2022), "Tropospheric ozone changes and ozone sensitivity from the present day to the future under shared socio-economic pathways", Atmos. Chem. Phys., Vol. 22 No. 2, pp. 1209-1227.

LU, X., ZHANG, L. & SHEN, L. (2019), "Meteorology and Climate Influences on Tropospheric Ozone: a Review of Natural Sources, Chemistry, and Transport Patterns", Current Pollution Reports, No. 7398, pp.

MA, R., BAN, J., WANG, Q. & LI, T. (2020), "Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review", Science of The Total Environment, Vol. 701134463.

MA, R., BAN, J., WANG, Q., ZHANG, Y. & LI, T. (2021), "Random Forest Model based Fine Scale Spatiotemporal O3 Trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017", Environmental Pollution, No. 10159, pp. 116635.

MOSS, R. H., EDMONDS, J. A., HIBBARD, K. A., MANNING, M. R., ROSE, S. K., VUUREN, D. V., CARTER, T. R., EMORI, S., KAINUMA, M. & KRAM, T. (2010), "The next generation of scenarios for climate change research and assessment", Nature, Vol. 463 No. 7282, pp. 747-56.

NI, R., LIN, J., YAN, Y. & LIN, W. (2018), "foreign and domestic contributions to springtime ozone over china 1".

ORDÓ EZ, C., BRUNNER, D., STAEHELIN, J., HADJINICOLAOU, P., PYLE, J. A., JONAS, M., WERNLI, H. & PRÉV T, A. S. H. (2007), "Strong influence of lowermost stratospheric ozone on lower tropospheric background ozone changes over Europe", Geophysical Research Letters, Vol. 34 No. 7, pp. L07805.

PAUSATA, F., POZZOLI, L., VIGNATI, E. & DENTENER, F. J. (2012), "North Atlantic Oscillation and tropospheric ozone variability in Europe: model analysis and measurements intercomparison", Atmospheric Chemistry & Physics, Vol. 12 No. 14, pp. 6357-6376.

ROGELJ, J., POPP, A., CALVIN, K. V., LUDERER, G., EMMERLING, J., T, D. G. A., FUJIMORI, S., STREFLER, J., HASEGAWA, T. & MARANGONI, G. (2018), "Scenarios towards limiting global mean temperature increase below 1.5 °C", Nature Climate Change.

SON, Y., R, Á. O., MARIE, S. O., HYSTAD, P. & JOSÉ, L. T. (2018), "Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters.", The Science of the total environment.

WOLF, K., CYRYS, J., HARCINÍKOVÁ, T., GU, J., KUSCH, T., HAMPEL, R., SCHNEIDER, A. & PETERS, A. (2016), "Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany", Science of the Total Environment, 1531-1540.

WONG, C., VADAKAN, N. V., KAN, H. & QIAN, Z. (2007), "Public Health and Air Pollution in Asia (PAPA): A Multicity Study of Short-Term Effects of Air Pollution on Mortality", Environmental Health Perspectives, Vol. 116 No. 9, pp. 1195-1202.

YANG, L., LUO, H., YUAN, Z., ZHENG, J. & CHEN, D. (2019), "Quantitative impacts of meteorology and precursor emission changes on the long-term trend of ambient ozone over the Pearl River Delta, China and implications for ozone control strategy", Atmospheric Chemistry and Physics, 1-26.

YANG, Y., LIAO, H. & LI, J. (2014), "Impacts of the East Asian summer monsoon on interannual variations of summertime surface-layer ozone concentrations over China", Atmos. Chem. Phys., Vol. 14 No. 13, pp. 6867-6879.

YIN, P., CHEN, R., WANG, L., MENG, X., LIU, C., NIU, Y., LIN, Z., LIU, Y., LIU, J. & QI, J. (2017), "Ambient Ozone Pollution and Daily Mortality: A Nationwide Study in 272 Chinese Cities", Environmental Health Perspectives, Vol. 125.

YU, Y., WANG, Z., HE, T., MENG, X., XIE, S. & YU, H. (2019), "Driving factors of the significant increase in surface ozone in the Yangtze River Delta, China, during 2013–2017", Atmospheric Pollution Research, Vol. 10 No. 4, pp. 1357-1364.

ZHENG, B., DAN, T., MENG, L., LIU, F., HONG, C., GENG, G., LI, H., LI, X., PENG, L. & QI, J. (2018), "Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions", Atmospheric Chemistry & Physics Discussions, 1-27.