Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Wenhao Wang Date

Satellite-Based Daily Ground Ozone Estimates in California, Using Machine Learning Methods

By

Wenhao Wang
Master of Public Health

Gangarosa Department of Environmental Health

_____

Yang Liu, Ph.D.
Committee Chair

Satellite-Based Daily Ground Ozone Estimates in California, Using Machine Learning Methods

By

Wenhao Wang

Bachelor of Medicine
Huazhong University of Science and Technology
2018

Thesis Committee Chair: Yang Liu, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2020

Abstract

Satellite-Based Daily Ground Ozone Estimates in California, Using Machine Learning Methods

By Wenhao Wang

Exposure to the ground-level ozone can trigger a variety of health problems as well as ecological impacts. To estimate ground-level ozone concentration, seldom satellite-based machine learning models were used in the prediction for large spatial and temporal coverage due to the lack of adequate satellite products. Troposphere Monitoring Instrument (TROPOMI) on board of the Sentinel 5 Precursor can provide high quality and relatively high-resolution gas pollutants data for the model of prediction the ozone. We aim to develop a high-performance TROPOMI satellite-driving machine learning model to estimate the daily maximum 8-hour average ground-level ozone concentration at a spatial resolution of square 10 kilometers in the state of California from May 2018 to April 2019 combined with predictors including meteorological fields, land-use variables. All predictors data and ground measurement of ozone are re-gridded to the 10×10 kilometers grid we create to build a random forest model setting the daily ground concentration in each pixel of the grid as the outcome. Our model achieved overall 10-fold cross-validation (CV) R2 of 0.83 with random mean square error (RMSE) of 5.91 ppb, indicating a good fit between model prediction and observation. Our model achieved a good prediction on the ground-level ozone concentration in California, supporting the feasibility and advantage of application TROPOMI satellite product and machine learning method in the prediction of ground-level ozone concentration. The result of our model can be applied in future epidemiological studies as well as the strategies studies in the control of ground-level ozone pollution.

Satellite-Based Daily Ground Ozone Estimates in California, Using Machine Learning Methods

By

Wenhao Wang

Bachelor of Medicine
Huazhong University of Science and Technology
2018

Thesis Committee Chair: Yang Liu, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2020

Acknowledge

I would like to express my sincere appreciation to Dr. Yang Liu, Dr. Liuhua Shi and the members of The Environmental Remote Sensing Group at Emory for their kindness and help on the process of this thesis. And, I would like to express my most thanks to my love, family, and friends for the supporting I gained from them.

Table of Content

1. Introduction

The ground-level ozone (O_3) pollution is an emerging public health over the past decades to future that quantities of epidemiological studies have proved that the exposure to ozone is associated with adverse health outcomes (Nuvolone, Petri, & Voller, 2017; Zhao, Markevych, Romanos, Nowak, & Heinrich, 2018). As evidences showed the ozone is getting higher over the past decades, the ground level ozone pollution is predicted to increase company with the global warming (Orru, Ebi, & Forsberg, 2017). Previous ground-level ozone related epidemiology studies applied ground monitor stations to represent the ambient ozone concentration for the are each ground stations located, meaning the lack of the assigning accurate ozone exposure to individuals at different location with distinguish ozone exposure pattern (Vicedo-Cabrera, et al., 2020). A precision ozone exposure model with large spatial coverage and fine spatial resolution is needed for the future ozone-related epidemiology study in the addressing potential adverse health outcome of long-term exposure to ozone among common population. The understanding of the ozone reaction mechanism can be very important in the control of the ozone pollution as well as a modeling of the ozone exposure data. Ground-level ozone is a secondary air pollutant created by the chemical reaction between the nitrogen oxides (NOx) and violative organic compounds (VOCs) induced by the sunshine which is highly related to the human activities as well as the natural emission (Fenger, 2009; Levy, et al., 1997). As the formation of the ozone, human activities including industrial and vehicles emission act as the major source of the anthropogenic NOx and VOCs which are the precursors for the formation reaction of the ground-level ozone (Stowell, et al., 2017). Similarly, natural source of VOCs from biogenic emissions and abiotic emission take

important part of the reaction in the formation of ozone (Derwent, Eggleton, Williams, & Bell, 1978). However, the dose of NOx is not always positive relative to the ozone is involved in the reversible reaction cycle with NO_2 which means that higher dose in NO_2 may lead to decomposition of ozone (Sillman, Logan, & Wofsy, 1990). With the highly reactive and the oxidative of the ground level ozone, the ozone-related chemistry reactions can be highly active and complicated that many factors and chemicals are involved in reactions. The ozone reaction is also sensitive to the meteorological factors including solar radiation, wind speed, temperature and pressure (Austin, et al., 2014). Thus, the understanding of the relationship between the precursors of ozone can be difficult and extremely complicated.

Some of modeling approaches to estimate the concentration of ozone in order to better estimate the ozone exposure in the health issues was developed. Those attempts included: chemical transported models (Sun, Fu et al. 2015), statistical interpolation models (Jerrett et al. 2013), simple regression satellite model (Liang, et al. 2019), and machine learning models in short period (Di et al. 2017). For each kinds of these models, there are some degree of the misclassification in the estimate ground-level ozone concentration. Also, there is seldom ozone exposure models considering enough predictors with long study period, covering large study domain. Recent year, mount of studies applied satellite derived machine learning models in the prediction of air pollution, showing high effieciency and great spatial and temporal coverage (Ma, Hu, Huang, Bi, & Liu, 2014; Di, et al., 2019; Vu, et al., 2018). The advangtage of the satellite derived machine learning model is the combined of greate spatial coverage of the data as well as the advantage of machine learning model in the ability of model non-linear bariables (Hu, et

al., 2017). With the complex ozone related chemical reactions pattern which is hard to be quantified and highly correlation to several meteorological and land use factors, machine learning model can be a good fit in the prediction of ozone concentration to address those complicted factors related to the ozone. Previous study in a small study domain of the comparision of variety types of ground ozone models proved the advantage of machine learning models in the prediciton of ground ozone concentration compared to the chemical transportation models (Feng, et al., 2019).

For the previous studies in the developing model with satellite measured ozone value, the OMI satellite data was used seldom studies in the prediction of the tropospheric ozone concentration (Kajino, et al., 2019). The the lack of satellite derived ground-level ozone concentration can be concluded in two major reasons. First, the spatical resolution of the ozone measurement satellite instrument is too coarse to achevie modeling in reflecting spatial variartion of ozone. Second, the ground-level ozone is hard to measure by the satellite instrument since most of ozone in the atomsphere are in the stratosphere (Tang, Wilson, Solomon, Shao, & Madronich, 2011). The TROPOMI instrument is onboard the sentinel-5p satellite launched on 13 October 2017 at 11:27 CEST/09:27 UTC focusing on the atmosphere gas pollutant measurement operated by the Royal Netherland Meteorological Institute (RNMI). Compared to other ozone measurement instrument like OMI with resolution of 13 km × 24 km, the TROPOMI support significantly finer resolution of 3.5 km × 7 km in the measurement of ozone and tropospheric nitrogen dioxide column. Previous studies applying the TROPOMI in the prediction of nitrogen dioxide showed great result in the prediction of ground nitrogen dioxide concentration

which is one of the most important precursors of ozone (Lorente, et al., 2019; Sandhiya, Kolandaivel, & Senthilkumar, 2014).

In this study, we would like to develop a random forest model with TROPOMI satellite data, metrological data, and multiple land use variables to estimate daily maximum 8-hour average ground-level concentration in California following one year from the release date of TROPOMI satellite data started on May 2018 to April 2019.

2. Data and Method

California was choosing as our study domain due to the relative high ozone concentration and great coverage of ground air pollution monitors. Our study domain was defined as the California as the state of the United States at the Pacific Coast hosting near 40 million residence. The study area of square meter covers the land area of the California and several small island near the coast to the Pacific Ocean. The study period is from May 2018 to April 2019, covering entire calendar year starting from the release date of the TROPOMI satellite product. In figure 1, the range of the study domain and all EPA ozone monitoring stations are marked in the map.

We developed a grid with 4206 pixels with size of ~ 10 kilometer by the fishnet tool in the ArcGIS Pro to cover the study area as the primary spatial unit for the ozone modeling as been shown in figure 2. Previous studies prove the strong link between several factors to the ozone chemical reaction, we selected some of parameters with high correlation to the ozone concentration in the modeling (Di, Rowland, Koutrakis, & Schwartz, 2016).

2.1 Ground-Level Ozone measurements.

The daily maximum 8-hour average ozone concentration for the study domain of the year of 2018 and 2019 was obtained from the 176 ground monitoring stations under the United States Environmental Protection Agency reference method. The raw pre-generated data was download from the Air Data platform of the EPA website (https://www.epa.gov/outdoor-air-quality-data). We assign our ground monitoring ozone data to the grid we created. In pixels with one ground monitoring station, the value from the stations represent the grid in pixels with more than one ground monitors, average data was obtained from all of ground monitoring stations within pixels.

2.2 TROPOMI satellite data

We applied satellite measured ozone total column which reflect the ozone concentration and the tropospheric nitrogen dioxide column which measure the nitrogen dioxide as an important precursor in the ozone formation. We obtained the level 2 TROPOMI data with the original spatial resolution of 3.5 km × 7 km from the GES DISC platform of NASA. A resampling of the TROPOMI satellite data need to be done for the modelling since the location of the measurement pixel will change in location and angle for every overpass orbit of the Sentinel-5p satellite which the TROPOMI onboard. For each day with valid TROPOMI observation, we average the value of ozone and nitrogen dioxide measurement from each TROPOMI pixel whose central point locate in each pixel we created as the modeling unit.

2.3 Meteorological Field

We used the meteorological data from the High-Resolution Rapid Refresh (HRRR) dataset from the Earth System Research Laboratory of the National Oceanic & Atmospheric Administration (NOAA) affiliated the United State Department of Commerce. The HRRR is a real-time 3-km resolution, hourly updated, cloud-resoling, convection-allowing atmospheric models by the radar simulation. We filter 19 HRRR meteorological parameters may highly related to the modeling of the daily ozone concentration including: wind speed (gust), U-component (eat-west) and V-component of wind (north-south) at 10 meter and 250 mb (ugrd_10m,vgrd_10m,vgrd_250,ugrd_250), temperature (tmp), pressure(pres), surface geopotential height (hgt), Moisture Availability (mstav), specific humidity (spfh), relative humidity (rh), surface roughness (sfcr), ground heat flux (gflux), vegetation type (vgtvp), convective available potential energy (cape), convective inhibition (cin), medium cloud cover (mcdc), visible diffuse downward solar flux (vddsf), and planetary boundary layer height (hpbl). Considering the reaction time of the production of ozone, the meteorological measurements for the period of study is from 10 a.m. to 4 p.m. local standard time were averaged to generate daytime meteorological measurements. The averaged meteorological fields also represent the average weather condition at the Sentinel-5p satellite overpass time which is 13:30 Mean Local Solar time (https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-5p/orbit) and lower the influence created by the extreme meteorological conditions during the measuring time on the association between the ozone concentration and other prediction variables.

2.4 Land-Use Variables

The elevation data in the model is downloaded from the Nation Elevation Dataset (NED) (http://ned.usgs.gov) with a spatial resolution of square 30 m which was average to the square 10 km pixels in the grid we created as the elevation in each pixel. We obtained the land cover data at the spatial resolution of square 30 m from the 2011 Landsat-derived land cover map downloaded from the National Land Cover Database (NLCD) (http://www.mrlc.gov). The land cover data was calculated by the 10 km grid showing the area percentage for each type of land cover including: water, developed, barren and forest. The population data is processed from the Gridded Population of the World (GPW) v4 from the Socioeconomic Data and Application Center (SEDAC) (https://sedac.ciesin.columbia.edu/data/collection/gpw-v4). Population count data v4.11 with spatial resolution of ~1 km at years of 2015 and 2020 were obtained to create a simple linear model in the calculation of the population at the study period which was then re-gridded to the grid we created.

2.5 Data Process

All the data were processed under the projection coordination system of USA Contiguous Albers Equal Area Conic system. For the training set of the model, we used all the 10 km gridded prediction data into the pixels including the ground monitoring stations. For the prediction data set, we used all the ~ 10 km pixels in the California with all the prediction variables. For the entire dataset, we standardize the value of each variable before the stage of modeling.

2.6 Random Forest Model

A random forest machine learning model is a set of decision trees. The model average decision tree sets to get the best predictions based on subset of predictors. There are two parameters of the model: m_{try} as the number of predictors sampled for splitting at each node, n_{tree} as the number of trees grows. We will train the model by each subset of the combination of the m_{try} and n_{tree} to get the model with best prediction accuracy (Breiman, 2001). There are 27 variables used in the random forest model, including 2 TROPOMI satellite measurements, 19 meteorological variables in the format of daytime average, and 6 land-use variables.

To validate the prediction result, we applied a 10-fold cross-validation (CV) technique. We randomly split the full tanning set into 10 subsets which contain about 10 percent amount of all training data. For each validation process, we used 9 subsets of the training set to do the model training which set the rest 1 subset as the testing for the model result from the model of 9 subsets. We repeat the validation process by 10 times with 10 different choice of the validation training set. We conducted a spatial CV and temporal CV, based on partitioning the training data set by ground stations and day of year. We get the statistical indicators of R-square, root mean squared error (RMSE) of the 10-fold CV prediction and ground measurement to assess the prediction accuracy.

We calculated the importance value for each variable to assess which variables are having more impacts on the prediction results. The measure is estimated using out-of-bag samples as a result of each predictor variable being permuted which is the increase of mean square errors (IncMSE%) of predictions. Higher number means higher importance in the prediction (Breiman, 2001).

All the modeling process and data analysis were done in R software, version 3.6.0. The projection process and the fishnet gridding process were done in ArcGIS Pro.

3. Result

3.1 Descriptive Statistics

Daily prediction of maximum 8-hour average ozone concentration started from 1 May 2018 to 30 April 2019. There are 50,648 observations in total applied to in the data training. And, 1,493,153 daily concentrations were estimated by the random forest. The annually mean daily maximum 8-hour average ozone concentration is 45.00 ppb for the entire California in the study period while the concentration of the ground monitoring ozone is 43.65 ppb. The maximum value of the estimate ozone concentration is 108.93 ppb while the minimum value is 5.16 ppb with. Over the entire study period and the study domain, estimate from 6510 observation-days are above the standard of USPEA in the As the definition of the ozone season, the mean value for the ozone season from May 2018 to October 2018 is 50.72 ppb, 39.18 ppb for the non-ozone season of November 2018 to April 2019. Figure 3 shows the comparison in daily ozone concentration among the entire California between ground monitors and modeling estimate. The modeling data and the ground monitors achieved a correlation of 0.93 ($p < 0.01$) over the time series of the study period in the mean daily ozone concentration over the entire state.

3.2 Result of Model Validation

The random, spatial, and temporal cross validation results, including R_2, RMSE as well as the linear slope between the predicted value and the ground monitors observed value

for the entire study area and period are showed in the table 1 and Figure 4. The overall random 10-fold cross validation R_2 result reached high value of 0.83. The overall RMSE is 5.91 ppb compared to the mean observation value of 43.57 ppb. The result show that there is a good agreement on the CV estimate ozone value and the observation ozone value in the California over the study domain.

3.3 Importance of Variables

Figure 5 shows the importance rank of the predictors in our random forest model. The importance rank of the random forest model indicates the most importance of the TROPOMI total ozone column, TROPOMI tropospheric nitrogen dioxide column, Forest land cover percentage, surface temperature and wind speed at 10 meter high.

3.4 Model Estimate of Ozone Concentration

The Figure 6 shows the spatial pattern of the ozone concentration at different time estimated by the random forest model. Clear spatial patterns can be observed over different period of estimate. The results showed that the ozone concentration is generally higher in the Southern California compared to Northern California. Compared to the coastal area, land area has higher concentration of ozone at both ozone and non-ozone season. The ozone concentration is much higher in the ozone season compared to the none ozone season. The highest zone of ozone concentration is suburb area east to the city of Los Angeles. In the ozone season, there are clustering of ozone pollution in the Napa Valley and east of it.

4. Discussion

4.1 Model Analysis

Following the launch of the TROPOMI satellite product, our model is the first TROPOMI satellite data derived machine learning model in the estimating ground level ozone concentration.

Although our study covers a large area of the state of California, our model shows great accuracy in the prediction of the ozone concentration. Our model has great temporal R_2 of 0.80, even though the study covered the entire calendar year with different seasons expressing distinct climate feature and characteristics. In the correlation test between the monitor data and the modeling data of the figure 3, a correlation of 0.93 was observed which is acceptable as we compare the mean from the monitoring stations to the estimate ozone from the modeling of the entire state. As for the yearly average we calculated for the ground monitors and the grids with ground monitors, a correlation of 0.94 was achieved, meaning a good efficiency of the model in the prediction of long-term average dose. For the strength of the random forest in this study, the included prediction parameters are in fairish amount of 27 without plenty of parameters at low importance, reducing the complexity of the model and degree of overfitting. Compared to previous studies in the application of OMI satellite data in the prediction of ground ozone concentration without machine learning model, our data is accurate and well validated (Liang, et al, 2019). Compared to ozone composition analysis of GEOS-CF from NASA's Goddard Earth Observing System (GEOS) in spatial resolution of ~ 25 km_2, we have better spatial resolution of ~ 10 km_2 (Knowland, et al., 2018). The previous country level CMAQ model (Liu, et al., 2010) provide an estimate of ground ozone at a spatial

resolution of square 36 km which is much coarser than ours. The previous chemical related models are also lack of adequate statistically validation. But we prove the efficiency of our model in 10-fold cross-validation. With the join of the TROPOMI satellite data, we extend the application of ozone concentration estimate from wildfire event to entire year (Watson, Telesca, Reid, Pfister, & Jerrett, 2019). Their model gave an R_2 of 0.66, which is much lower than ours (0.83). Compared to the previous satellite-based ground ozone concentration estimation model in the continuous United State (Di, et al., 2017), our model has better cross-validated R_2 (0.83 better than 0.76) and lower RMSE (5.91 lower than 7.36). Also, the Di's result applied complicated machine learning method with huge amount of predication. Compared to their model, we didn't include chemical transportation model result, ozone profile, CO, methane, VOCs, and convolutional layers. The random forest model we applied are much easier than the neuro network they did. It means that we achieve better modelling estimate with simpler model. Without the adding of the convolutional layer which will rely on the ground monitoring data as prediction to the modeling estimate, our model can be applied to areas without enough ground monitoring stations. Also, we didn't include the chemical transported model, meaning that the random forest model proves the advantage of applying machine learning method in the prediction of ozone concentration since it can achieve good fit of estimation to the ground monitors. The reason for the advantage is probably from the cumbersome chemical mechanism the ozone and its precursors involved which may difficult to quantify by the chemical transform model in the ambient air (Sillman, Logan, & Wofsy, 1990). And, the machine learning takes the advantage of the flexibility in the

non-linear relationship come with complicate interaction among prediction parameters (Bishop, 2016).

4.2 Importance Rank Analysis

In terms of the importance rank, the two TROPOMI measurement ranked at first which is reasonable since they are direct measuring of the ozone and nitrogen dioxide as an important precursor of the ozone. Admittedly, we applied the total ozone column data product from the TROPOMI in the prediction, containing the stratospheric ozone concentration in the measurement. Since 90 percent of ozone in the atmosphere is in the stratosphere, the 90 percent of the total ozone column measurement is contributed to stratosphere ozone instead of tropospheric ozone which is the ozone may relate to the ground-level ozone (Wilson, Madronich, Longstreth, & Solomon, 2019). The possible assumption could be the stratospheric ozone may be static in certain conditions of meteorological factors. So that, the variation of the total ozone column at the moment may be from the differences in ground-level ozone the other predictors share similar value or patterns. The TROPOMI tropospheric nitrogen dioxide column ranks second in the importance ranking that the nitrogen dioxide is an important precursor of the ozone related chemistry mechanism. So, it is rational that the tropospheric nitrogen dioxide column ranks high among predictors. Forest coverage for each pixel ranked third in the importance of predictors. The possible explanation may be the forest will release VOCs as the precursor of the ozone formation (Hu, et al., 2018). Since there is complicated relationship between the NOx and ozone in the rural and urban area that the relationship between the NOx and the ozone formation may bias as the increasing of NO_2 (Domínguez-López, et al., 2014). As been discussed in many previous studies, the main

source of NO_2 is from human activities, meaning the NO_2 emission will be low in the forest area (Bosson, Mudway, & Sandström, 2019). So, forest area may have related low concentration of ozone due to the lack of NO_2 source. Secondly, the ozone concentration in the forest area may highly positive related to the NO_2 concentration measured by the tropospheric NO_2 column since the local dose of NO_2 is low and saturated.

4.3 Prediction Analysis

From the figure 6, we can clearly identify the spatial distribution of the ozone concentration in California that the Southern California has higher concentration than the north. The Southern California is with higher population and more developed land meaning more industrial and transported related emission of the precursors of the nitrogen dioxide. Consequently, the ozone concentration in the urban area of the Southern California is higher. In another hand, Southern California is under stronger heat and more drought weather which is proved to be related to the high concentration of the ozone (Niu, et al., 2018). We can also observe the ozone concentration is lower in the coastal area. With high rank of the u-component wind speed as well as water land use type in the importance rank, we can postpone that the low concentration in coastal areas may be related to the climate characteristics brought by wind and seawater. As the u-component wind which is east-west direction is vertical to the coastline, the possible low concentration in coastal area may be from the accelerated diffusion of ozone brought by the sea breeze.

4.4 Limitation and Future Plan

Although we achieved great efficiency in the ground ozone estimate in California, Using TROPOMI satellite product and machine learning methods, there still are some limitation

of this study. First, the spatial resolution of the model is still high. The difficulties of the improvement are from the coarse resolution of TROPOMI product. To improve the spatial resolution, we may apply multiple statistical and machine learning methods to do the interpolation or resampling. Second, the current study domain is high in ozone concentration. This means the model efficiency is not proved in areas with low ground ozone concentration. In future of our studies, we may extend our study domain to national level to validate the model in various kind of ground ozone pollution scenario. Third, the feasibility of the application of TROPOMI in high attitude area is not clear. The TROPOMI satellite measurement ozone have great efficiency and data quality in the lower attitude areas as we got from the data while we haven't developed the model in high attitude areas. We may continue to expand our study domain and study period to generalize our model application. Last, our current TROPOMI data in the modeling is still the ozone column data for the entire atmosphere. This means the ozone measurement will be bias by the stratospheric ozone. We are looking forward the coming ozone profile product of TROPOMI which will be released in 2020.

5. Conclusion

In general, our model achieved great accuracy with overall R_2 of 0.83 under large spatial coverage and long time period in acceptable spatial resolution. The result of our method showed strong efficiency of the model in the prediction of ozone concentration in California. We also validated the great feasibility of the application of the TROPOMI satellite product in the prediction of ground ozone pollution. Following the going development of the TROPOMI products, we will keep trying the new product and

improve our model to achieve more accurate ozone estimate covering large area with finer spatial resolution.

6. Tables and Figures

Figure 1. Study area and the ground monitors with data availability.

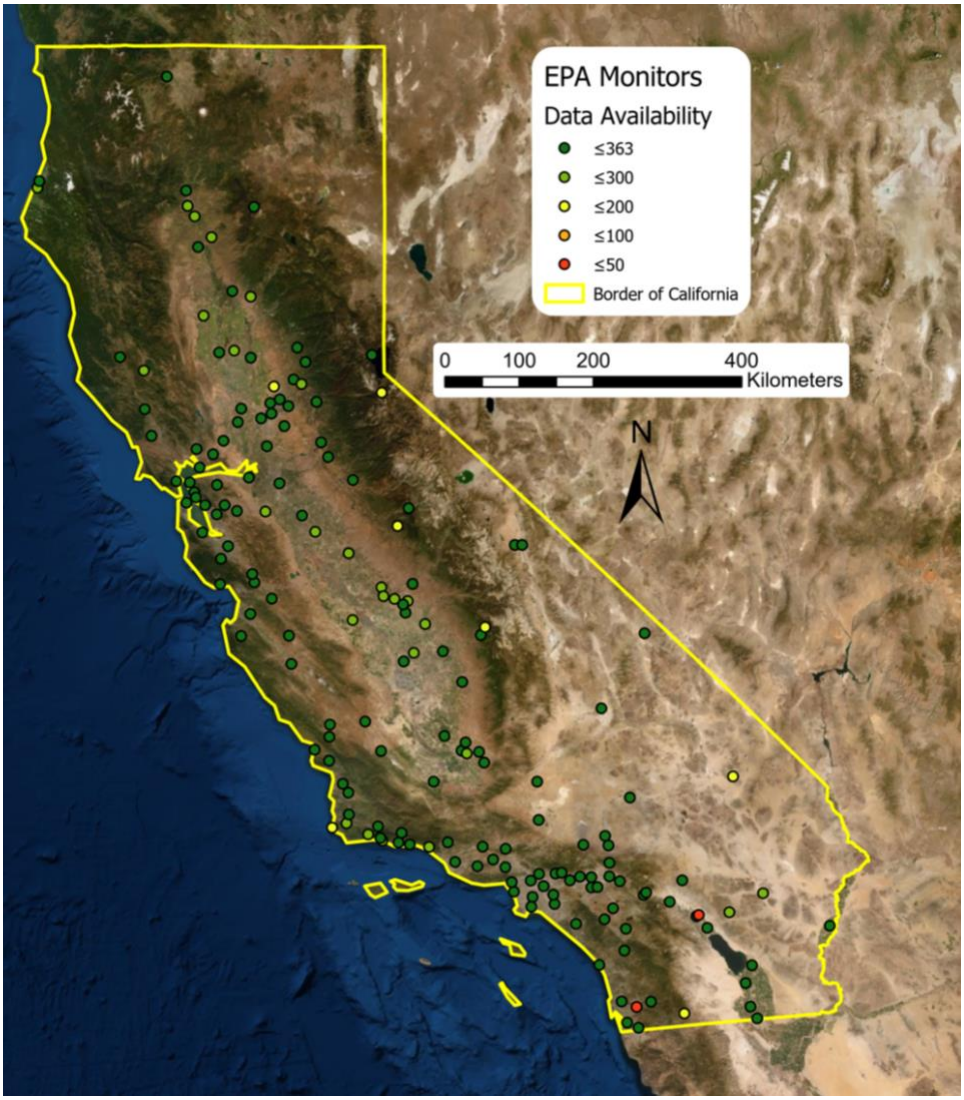


Figure 2. Grid in spatial resolution of 10 km × 10 km.

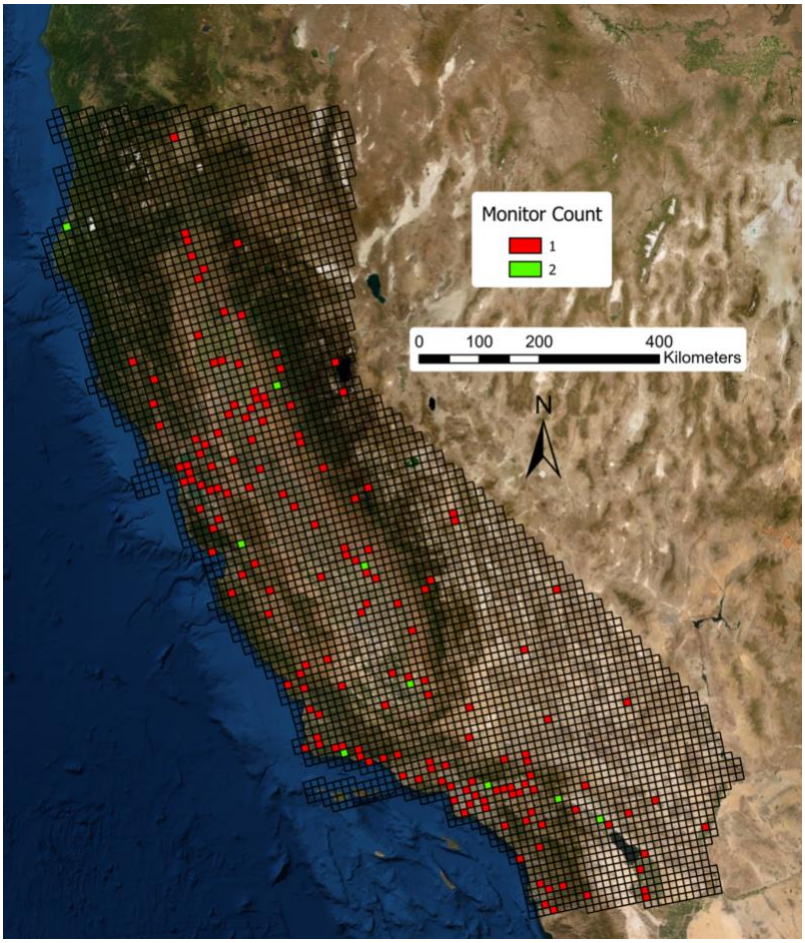


Figure 3. Comparison between ground monitors and modelling estimate of daily average ozone concentration of California over study period from May 2018 to April 2019.

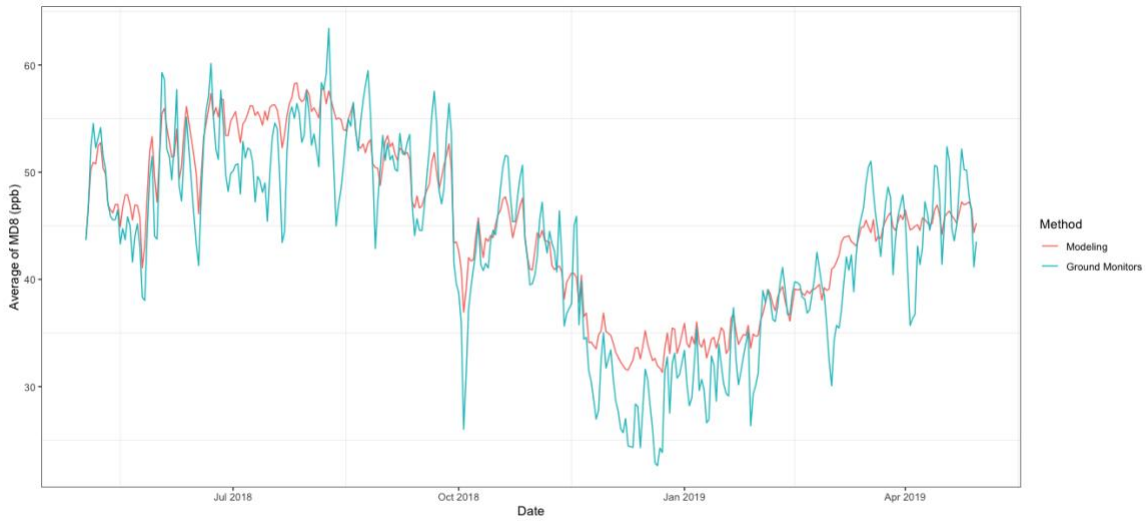


Table 1. Cross-Validation results for the study area and study period.

CV	R_2	RMSE (ppb)	slope
random	0.83	5.91	1.07
spatial	0.70	8.02	1.08
temporal	0.80	6.43	1.06

Figure 4. Scatter plots of the cross-validation.

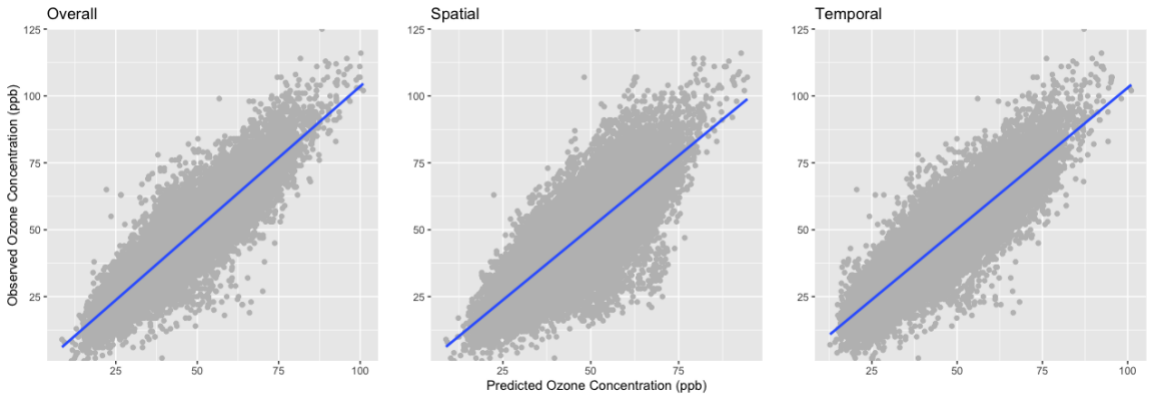


Figure 5. Importance rank for the predictors applied in the model, the importance is expressed as the percent increase mean square prediction error for each predictors in the model.

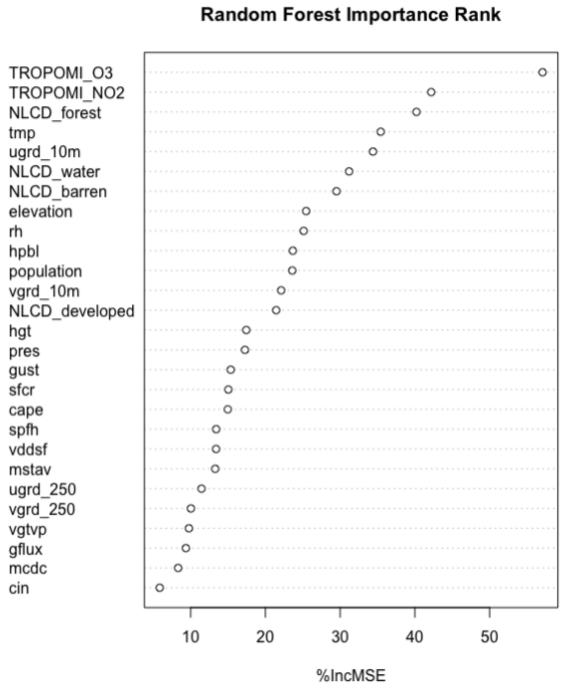
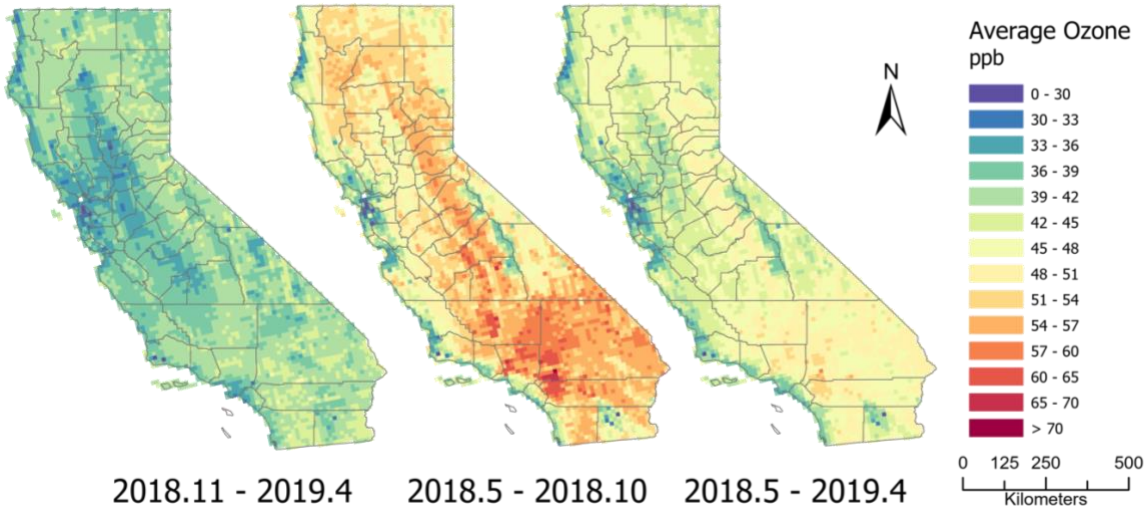


Figure 6. Mean ozone concentration prediction over the state of California in different period. The November to April is defined as non-ozone season while the May to October was define as the ozone season.



7. Reference

Austin, E., Zanobetti, A., Coull, B., Schwartz, J., Gold, D. R., & Koutrakis, P. (2014). Ozone trends and their relationship to characteristic weather patterns. Journal of Exposure Science & Environmental Epidemiology, 25(5), 532–542. doi: 10.1038/jes.2014.45

Bishop, C. M. (2016). Pattern Recognition and Machine Learning. New York, NY: Springer New York.

Bosson, J. A., Mudway, I. S., & Sandström, T. (2019). Traffic-related Air Pollution, Health, and Allergy: The Role of Nitrogen Dioxide. American Journal of Respiratory and Critical Care Medicine, 200(5), 523–524. doi: 10.1164/rccm.201904-0834ed

Breiman, L. (2001, October 1). Random Forests. Retrieved from https://dl.acm.org/doi/10.1023/A:1010933404324#

Derwent, R., Eggleton, A., Williams, M., & Bell, C. (1978). Elevated ozone levels from natural sources. Atmospheric Environment (1967), 12(11), 2173–2177. doi: 10.1016/0004-6981(78)90172-5

Di, Q., Rowland, S., Koutrakis, P., & Schwartz, J. (2016). A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. Journal of the Air & Waste Management Association, 67(1), 39–52. doi: 10.1080/10962247.2016.1200159

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM2.5Exposures with High Spatiotemporal Resolution across the

Continental United States. Environmental Science & Technology, 50(9), 4712–4721. doi: 10.1021/acs.est.5b06121

Di, Q., Dai, L., Wang, Y., Zanobetti, A., Choirat, C., Schwartz, J. D., & Dominici, F. (2017). Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. Jama, 318(24), 2446. doi: 10.1001/jama.2017.17923

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., … Schwartz, J. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environment International, 130, 104909. doi: 10.1016/j.envint.2019.104909

Domínguez-López, D. A., Adame, J. A., Hernández-Ceballos, M. A. D. L., Vaca, F. P., Morena, B. undefined, & Bolívar, J. undefined. (2014). Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environmental Monitoring and Assessment, 186(9), 5337–5351. doi: 10.1007/s10661-014-3783-9

Feng, R., Zheng, H. J., Zhang, A. R., Huang, C., Gao, H., & Ma, Y. C. (2019). Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison:A case study in hangzhou, China. Environmental pollution (Barking, Essex : 1987), 252(Pt A), 366–378. https://doi.org/10.1016/j.envpol.2019.05.101

Fenger, J. (2009). Air pollution in the last 50 years – From local to global. Atmospheric Environment, 43(1), 13–22. doi: 10.1016/j.atmosenv.2008.09.061

Hu, B., Jarosch, A.-M., Gauder, M., Graeff-Hönninger, S., Schnitzler, J.-P., Grote, R., … Kreuzwieser, J. (2018). VOC emissions and carbon balance of two bioenergy

plantations in response to nitrogen fertilization: A comparison of Miscanthus and Salix. Environmental Pollution, 237, 205–217. doi: 10.1016/j.envpol.2018.02.034

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM2.5Concentrations in the Conterminous United States Using the Random Forest Approach. Environmental Science & Technology, 51(12), 6936–6944. doi: 10.1021/acs.est.7b01210

Kajino, M., Hayashida, S., Sekiyama, T. T., Deushi, M., Ito, K., & Liu, X. (2019). Detectability assessment of a satellite sensor for lower tropospheric ozone responses to its precursors emission changes in East Asian summer. Scientific Reports, 9(1). doi: 10.1038/s41598-019-55759-7

Knowland, K., E., K., Keller, K., Ott, K., Duncan, K., & K., K. (2018). Application of NASA's new global high-resolution air quality forecasts: Stratospheric intrusion-influenced ozone exceedance events in the USA. Retrieved from https://ui.adsabs.harvard.edu/abs/2018AGUFM.A44B..07K/abstract

Levy, H., Kasibhatla, P. S., Moxim, W. J., Klonecki, A. A., Hirsch, A. I., Oltmans, S. J., & Chameides, W. L. (1997). The global impact of human activity on tropospheric ozone. Geophysical Research Letters, 24(7), 791–794. doi: 10.1029/97gl00599

Liang, S., Li, X., Teng, Y., Fu, H., Chen, L., Mao, J., … Azzi, M. (2019). Estimation of health and economic benefits based on ozone exposure level with high spatial-temporal resolution by fusing satellite and station observations. Environmental Pollution, 255, 113267. doi: 10.1016/j.envpol.2019.113267

Liu, X. H., Zhang, Y., Xing, J., Zhang, Q., Wang, K., Streets, D. G., ... & Hao, J. M. (2010). Understanding of regional air pollution over China using CMAQ, part II.

Process analysis and sensitivity of ozone and particulate matter to precursor emissions. Atmospheric Environment, 44(30), 3719-3727.

Lorente, A., Boersma, K. F., Eskes, H. J., Veefkind, J. P., J. H. G. M. Van Geffen, Zeeuw, M. B. D., … Krol, M. C. (2019). Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI. Scientific Reports, 9(1). doi: 10.1038/s41598-019-56428-5

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating Ground-Level PM2.5 in China Using Satellite Remote Sensing. Environmental Science & Technology, 48(13), 7436–7444. doi: 10.1021/es5009399

Niu, Y., Cai, J., Xia, Y., Yu, H., Chen, R., Lin, Z., … Kan, H. (2018). Estimation of personal ozone exposure using ambient concentrations and influencing factors. Environment International, 117, 237–242. doi: 10.1016/j.envint.2018.05.017

Nuvolone, D., Petri, D., & Voller, F. (2017). The effects of ozone on human health. Environmental Science and Pollution Research, 25(9), 8074–8088. doi: 10.1007/s11356-017-9239-3

Orru, H., Ebi, K. L., & Forsberg, B. (2017). The Interplay of Climate Change and Air Pollution on Health. Current Environmental Health Reports, 4(4), 504–513. doi: 10.1007/s40572-017-0168-6

Sandhiya, L., Kolandaivel, P., & Senthilkumar, K. (2014). Oxidation and Nitration of Tyrosine by Ozone and Nitrogen Dioxide: Reaction Mechanisms and Biological and Atmospheric Implications. The Journal of Physical Chemistry B, 118(13), 3479–3490. doi: 10.1021/jp4106037

Sillman, S., Logan, J. A., & Wofsy, S. C. (1990). The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes. Journal of Geophysical Research, 95(D2), 1837. doi: 10.1029/jd095id02p01837

Stowell, J. D., Kim, Y.-M., Gao, Y., Fu, J. S., Chang, H. H., & Liu, Y. (2017). The impact of climate change and emissions control on future ozone levels: Implications for human health. Environment International, 108, 41–50. doi: 10.1016/j.envint.2017.08.001

Tang, X., Wilson, S. R., Solomon, K. R., Shao, M., & Madronich, S. (2011). Changes in air quality and tropospheric composition due to depletion of stratospheric ozone and interactions with climate. Photochemical & Photobiological Sciences, 10(2), 280. doi: 10.1039/c0pp90039g

Vicedo-Cabrera, A. M., Sera, F., Liu, C., Armstrong, B., Milojevic, A., Guo, Y., … Gasparrini, A. (2020). Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries. Bmj, m108. doi: 10.1136/bmj.m108

Vu, B. N., Bi, J., Sánchez, O., Gonzales, G. F., Steenland, K., & Liu, Y. (2018). Developing Advanced PM2.5 Exposure Models in Lima, Peru. ISEE Conference Abstracts, 2018(1). doi: 10.1289/isesisee.2018.o01.04.09

Watson, G. L., Telesca, D., Reid, C. E., Pfister, G. G., & Jerrett, M. (2019). Machine learning models accurately predict ozone exposure during wildfire events. Environmental Pollution, 254, 112792. doi: 10.1016/j.envpol.2019.06.088

Wilson, S. R., Madronich, S., Longstreth, J. D., & Solomon, K. R. (2019). Interactive effects of changing stratospheric ozone and climate on tropospheric composition and

air quality, and the consequences for human and ecosystem health. Photochemical & Photobiological Sciences, 18(3), 775–803. doi: 10.1039/c8pp90064g

Zhao, T., Markevych, I., Romanos, M., Nowak, D., & Heinrich, J. (2018). Ambient ozone exposure and mental health: A systematic review of epidemiological studies. Environmental Research, 165, 459–472. doi: 10.1016/j.envres.2018.04.015