Environmental Pollution 316 (2023) 120685

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

A novel machine learning method for evaluating the impact of emission sources on ozone formation

Yong Cheng, Xiao-Feng Huang *, Yan Peng, Meng-Xue Tang, Bo Zhu, Shi-Yong Xia, Ling-Yan He

Laboratory of Atmospheric Observation Supersite, School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, 518055, China

ARTICLE INFO

Keywords:
Ozone
VOCs
Emission sources
Machine learning
SHAP

ABSTRACT

Ambient ozone air pollution is one of the most important environmental challenges in China today, and it is particularly significant to identify pollution sources and formulate control strategies. In present study, we proposed a novel method of positive matrix factorization-SHapley Additive explanation (PMF-SHAP) for evaluating the impact of emission sources on ozone formation, which can quantify the main emission sources of ozone pollution. In this method, we first used the PMF model to identify the source of volatile organic compounds (VOCs), and then quantified various emission sources using a combination of machine learning (ML) models and the SHAP algorithm. The R^2 of the optimal ML model in this method was as high as 0.96, indicating that the prediction performance was excellent. Furthermore, we explored the impact of different emission sources on ozone formation, and found that ozone formation in Shenzhen was more affected by VOCs, of which vehicle emission sources may have the greatest impact. Our results suggest that the appropriate combination of traditional models with ML models can well address environmental pollution problems. Moreover, the conclusions obtained based on the PMF-SHAP method were different from the traditional ozone formation potential (OFP) results, providing valuable clues for related mechanism studies.

1. Introduction

High concentrations of ozone are very harmful to human health. In recent years, long-term emissions of air pollutants in the Pearl River Delta (PRD) region of China have resulted in complex air pollution with ozone as the main pollutant (Shao et al., 2009; Ou et al., 2016). Additionally, this trend is gradually increasing in other regions of China (Shu et al., 2017; Huang et al., 2021). Studies have shown that the main factors affecting the formation of ozone are volatile organic compounds (VOCs) and nitrogen oxides (NOx), among which the pollution characteristics and sources of VOCs are relatively more complex (Zhu et al., 2021). Currently, the technologies for source identification of VOCs mainly include source emission inventories (Zheng et al., 2009; Wu and Xie, 2017), chemical transport models (CTMs) (Fang et al., 2021), and receptor models (Huang et al., 2021; Ding et al., 2022). However, it is difficult to make an inventory of source emissions, and the latest emission inventories usually lag current conditions by a year or more (Xing et al., 2020). Additionally, the limitation of the source emission inventory further constrains the performance of the CTM model and the results are often lower than expected.

Receptor models tend to be more popular due to their fewer hardware requirements, high accuracy, and ease of configuration, especially the positive matrix factorization (PMF) model (Song et al., 2007; Liu et al., 2020; Huang et al., 2021). The researchers combined the PMF model and the VOC species maximum incremental reactivity (MIR) value (Carter, 2010) to calculate the ozone formation potential (OFP) to evaluate the relationship between VOCs and ozone concentrations (Yan et al., 2017; Huang et al., 2020; Huang and Hsieh, 2020). Although this method is efficient, there are still some problems. The key parameter MIR value used in this method was often calculated based on 39 cities in the United States where ozone exceeds the standard. However, the actual situation of the atmospheric environment varies from region to region. Whether the MIR values can fully reflect the contribution of VOCs to ozone under the condition of complex air pollution in China is controversial (Qiu et al., 2020; Zhang et al., 2021), and the OFP calculation has some other limitations, such as the contribution of already depleted VOCs in the atmosphere to ozone formation cannot be evaluated. Therefore, it is of great significance to develop a novel method that is rapid, accurate, comprehensive and widely applicable to evaluate the relationship between pollutant precursors and ozone concentrations.

* Corresponding author.
E-mail address: huangxf@pku.edu.cn (X.-F. Huang).

https://doi.org/10.1016/j.envpol.2022.120685
Received 28 September 2022; Received in revised form 6 November 2022; Accepted 14 November 2022
Available online 15 November 2022

Currently, machine learning (ML) models are increasingly used in scientific research (Zahrt et al., 2019; Ogata et al., 2021; Rybarczyk and Zalakeviciute, 2021; Zhong et al., 2021). When faced with complex nonlinear problems, ML models show strong performance and application potential (Cheng et al., 2019; Shrock et al., 2020; Cheng et al., 2021; Nair et al., 2021; Bland et al., 2022; Zhu et al., 2022). However, because ML model operations are often viewed as a "black box", it is difficult to determine how predictors produce results (Liu et al., 2022), which often leads to ML being questioned. With the continuous development of the field of ML, the SHapley Additive explanation (SHAP) algorithm developed by a scientific team based on the classic Shapley theory (Shapley, 1953) can better quantify and explain the importance of each factor in some ML models (Lundberg et al., 2020). Compared with the indicators of traditional models, SHAP value is based on statistical theoretical knowledge and it is mainly affected by the performance of ML model. Since ML models are suitable for dealing with complex nonlinear relationship problems, the proper application of SHAP algorithm may provide new clues for some mechanism studies.

In present study, we proposed a novel method to evaluate the impact of emission sources on ozone formation. In this method, we innovatively combined the PMF model as well as multiple types of ML models (XGBoost, LightGBM, CatBoost, and RF) to systematically train and test the data. Further combined with the SHAP algorithm, the actual impact of different emission sources on ozone formation can be well quantified, which can provide valuable information support for formulating ozone control strategies in different regions and novel ideas for addressing environmental pollution problems.

2. Materials and methods

2.1. Study area and data source

Fig. 1 shows that during the observation period (September 6, 2019 to October 31, 2019), the average daily ozone concentrations in Guangdong Province ranked first among all provinces in China, with an average concentration of approximately 58 ppb. The air pollution datasets are obtained from the Ministry of Environmental Protection of China and the China Air Quality Online Monitoring and Analysis Platform (https://www.aqistudy.cn/).

Our laboratory of atmospheric observation supersite is located at the Peking University Shenzhen Graduate School (PKU-AOSS). The daytime (9:00–17:00) average ozone concentration of PKU-AOSS during the observation period is approximately 60 ppb, which is at the average level of Shenzhen and has good regional representation. In present study, the instrument used for VOC observation was the TH-300 B VOC fast online monitoring system, which is equipped with an ultralow temperature capture system and an analysis system (Agilent GCMS 7820A/5977 E). We selected 36 representative VOCs as our research objects, such as C2–C5 hydrocarbon species, alkanes with branched chains, aromatic hydrocarbons and isoprene, which in turn can reflect the sources of fossil fuels, combustion, solvent and natural emissions. In addition, air pollutants (NOx and O_3) were monitored simultaneously at a distance of approximately 50 m from TH-300 B (the data are shown in Fig. S1). All monitoring instruments were installed in the atmospheric supersite building, and the sampling entrance was set on the roof.

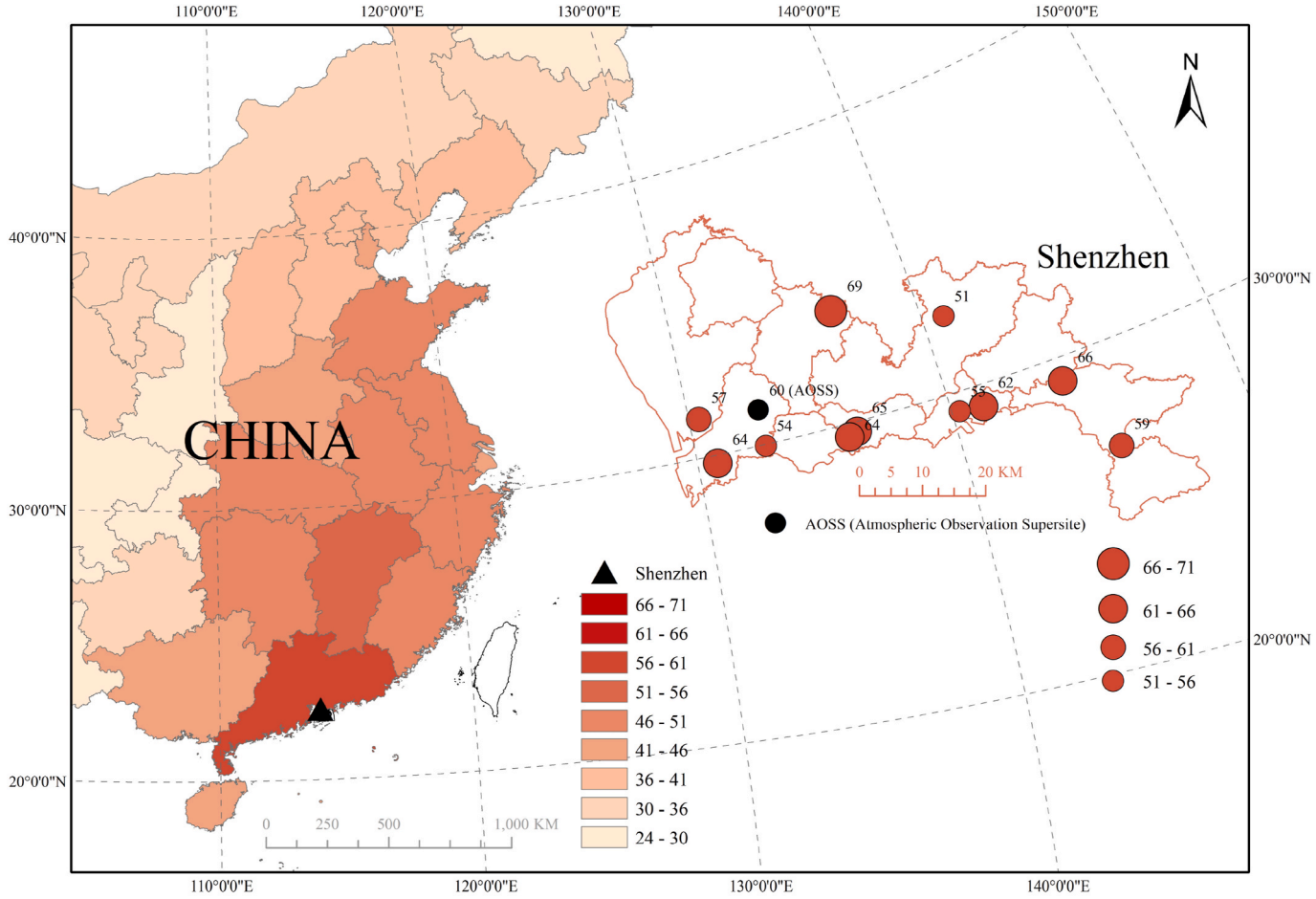

Fig. 1. The national ozone distribution and observation station map during the observation period. The left part of the figure is the distribution of the average concentration of ozone during the day in the eastern provinces of China (calculated based on the data of the national control stations in each province); The right part of the figure shows the daily average ozone concentration of the national control stations in Shenzhen.

2.2. Novel method

This study proposed PMF-SHAP method for evaluating the impact of emission sources on ozone formation. The framework of the PMF-SHAP is shown in Fig. 2. In PMF-SHAP, first, we performed data preprocessing on the VOC and NOx data collected during the daytime, and then the source of characteristic VOC data was identified by the PMF model (Section S1.2 for model details). In this study, we identified the anthropogenic emission sources of VOCs into five sources: vehicle emissions, gasoline leakage, solvent usage, biomass burning, and regional transmission. The specific PMF results are shown in Fig. S2. Then, the processed anthropogenic emission sources, natural sources (with isoprene as a tracer), and NOx data were "fed" to the ML models (including XGBoost, LightGBM, CatBoost, and RF Model; Section S1.3 for model details) for systematic training and testing, which compared with complex deep learning models, the four ML models more easily captured the changes in the internal parameters of the model and variable interpretation (Liu et al., 2022). Finally, the optimal model for ozone prediction was selected through continuous adjustment of the training strategy as well as multiple random experiments. Based on the prediction result of the optimal model, the SHAP algorithm (Section S1.4 for model details) was used to calculate the impact of different emission sources, and NOx on ozone formation. Furthermore, we used the observation based model (OBM) (Section S1.5 for model details) for comparison and verification.

Due to the complex nonlinear relationship between ozone and precursors, the photochemical reactions will require certain conditions and time. Therefore, to more accurately mine the influence of different sources on ozone formation during active photochemical reactions (9:00–17:00), we integrated the 1 h-ahead data of each driving factor and ozone data to obtain the final dataset. To ensure the accuracy of the experiments, each ML model uniformly used 80% of the dataset as the training set and 20% as the test set. Additionally, to reduce the influence of the randomness of the model on the test set results, all the involved ML models were subjected to 5 independent random repeated experiments. In this study, the ML models and SHAP algorithm were mainly implemented on the Python 3.7 and Anaconda 5.2 platforms, and the PMF model used the US EPA PMF 5.0 version.

3. Results and discussions

3.1. Comparing the performance of different ML models

It is very important to compare and test multiple models when dealing with different data, which will directly affect the accuracy of the results (Liu et al., 2022). In this study, we established four ML models, including the XGBoost, LightGBM, CatBoost, and RF models, to train and test ozone data with different driver data. The test results are shown in Fig. 3(a). The XGBoost model performs the best, and its MAE and RMSE values were the smallest among the four models, which were 8.37 and 10.79, respectively. The R^2 value was the largest, which was 0.84. In addition, it can be seen from the scatter distribution in this figure that the deviation between the predicted value and the actual value of the XGBoost model was smaller than that of the other three models, indicating that the predicted result is also more accurate. Therefore, we chose the XGBoost model as the optimal model to predict all ozone data during the day. The model prediction results were shown in Fig. 3(b). From the linear relationship in the Fig., it can be seen that the model has high prediction accuracy, and the R^2 was 0.96, the MAE was 2.44, and the RMSE was 4.98, indicating that the XGBoost model can accurately capture and predict the nonlinear relationship between ozone and different driving factors. Furthermore, the model had high accuracy and can well support the analysis of the SHAP algorithm.

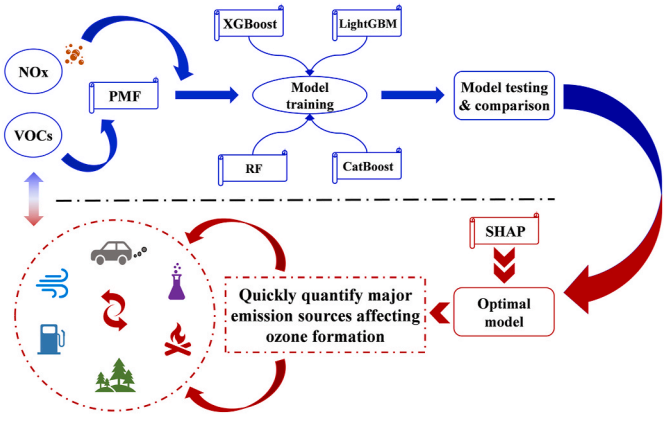

Fig. 2. PMF-SHAP framework structure.

3.2. Impacts of different driving factors in daily situations

As seen from Fig. S3, the hourly mean ozone concentrations exhibit a single-peak diurnal variation, the maximum concentration appears at approximately 1:00 p.m., and the lowest concentration occurs at approximately 7:00 a.m.; Factors similar to the diurnal characteristics of ozone are wind speed, JNO_2, temperature and natural sources of VOCs, while the opposite factors to the diurnal characteristics of ozone are NOx, RH, vehicle emission source and gasoline leakage source. From Fig. S3, the effective information we can obtain is very limited. To further explore the relationship between ozone and emission sources and NOx, we explored the results of the PMF-SHAP method.

In Fig. 4(a), we can measure the impact of different sources on ozone formation by the absolute value of SHAP. The larger the absolute value of SHAP is, the greater the impact on ozone formation. Fig. 4(b) mainly shows the relationship between the ozone SHAP value and the concentrations of different driving factors. When the concentration of the driving factor is larger, the SHAP value is also larger, indicating that the positive correlation between ozone and driving factor is more obvious; When the concentration of the driving factor is larger, the SHAP value is smaller, indicating that the negative correlation between ozone and driving factor is more obvious.

From the SHAP value of each driving factor in Fig. 4(a), it can be seen that three driving factors have a dominant impact on ozone formation, which include the source of vehicle emissions, regional transmission and natural sources. The SHAP value corresponding to the vehicle emission source was 11.64, the regional transmission was 6.39, and the natural sources was 4.96. Fig. 4(b) shows that there is an obvious negative correlation between the vehicle emission source and ozone. When the value of the vehicle emission source is larger, the ozone value is smaller. In general, a large number of primary VOCs emitted by most emission sources can directly react with substances such as OH radicals, Cl radicals and other substances, breaking the stable state of NOx-O_3 (Luecken et al., 2018), generating HO_2 radicals and RO_2 radicals, and then rapidly reacting with NO to convert to NO_2, which eventually leads to the accumulation of O_3. In particular, the reactivity of VOCs emitted by vehicles could be very high, and on 1-h resolution data scale, it was observed that these VOCs were quickly consumed, which was supported by the diurnal variations of vehicle emission sources (Figs. S3 and S6): the concentration of VOCs from vehicles is very high at night when the environmental conditions were unfavorable, however, it can be quickly consumed to very low levels during the day. This could be closely related to the highly reactive substances in the exhaust gas emitted by vehicles, such as formaldehyde and HONO (Kurtenbach et al., 2001). These species can rapidly be photolyzed to generate OH radicals, then accelerating the loss of VOCs.

The second most influential source is the regional transmission source, which includes VOC species that are relatively more stable, and tend to carry out regional transmission synchronously with ozone. Fig. 4 (b) also shows that there is a relatively obvious positive correlation

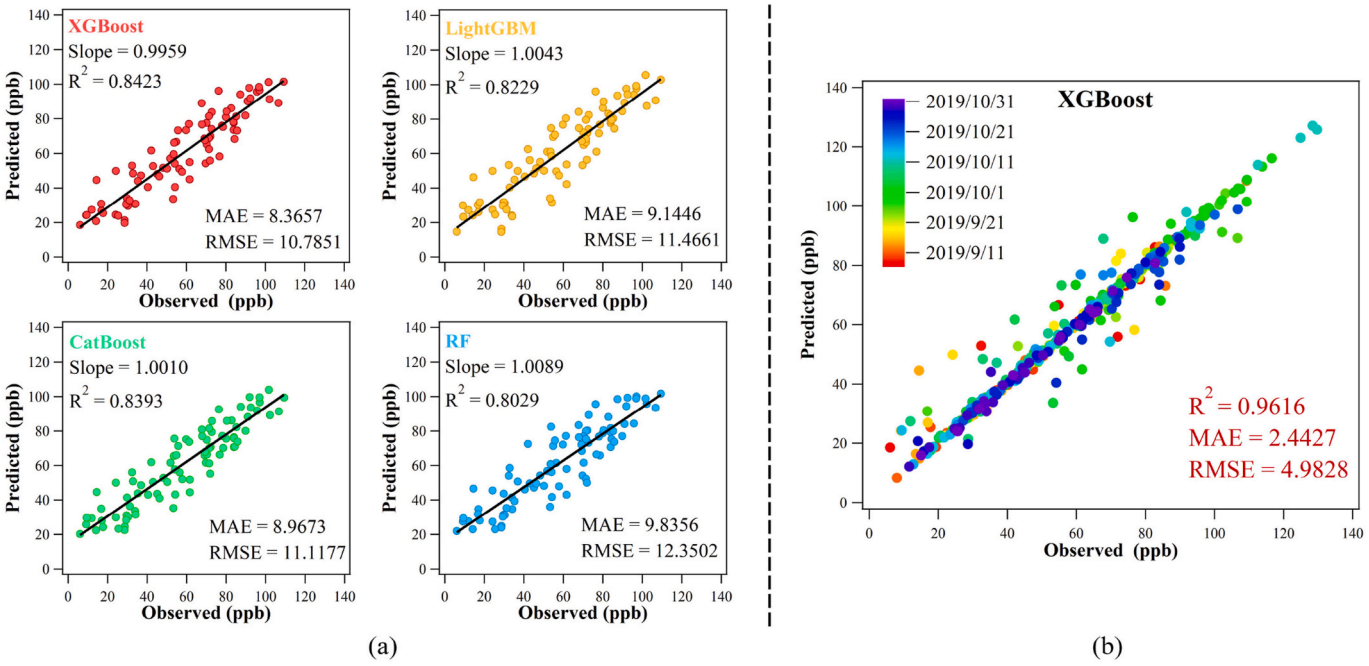

Fig. 3. Performance of different ML models. (a) The modeling results of different models on the test dataset; (b) The modeling results of the XGBoost model on all dataset.

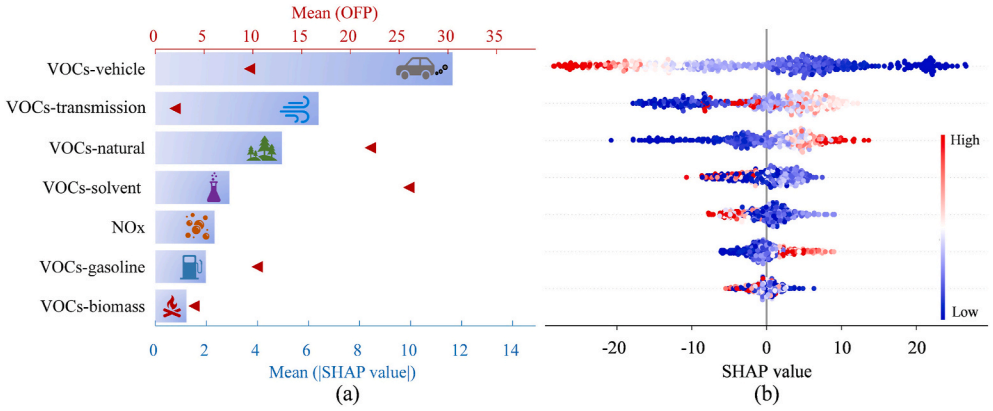

Fig. 4. Relationship between driving factors and ozone over the observation period. (a) The impact of different driving factors on ozone formation in daily situations; different factors are represented on the Y-axis in descending order of the absolute value of SHAP, the value is represented by columns, and the red triangles represent the OFP value of different factors; (b) Relationship between SHAP value and the values of different driving factors. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

between the two. The third driving factor is natural sources, which have a significant positive correlation with ozone, indicating that natural sources promote ozone formation, but the efficiency may not be as high as that of vehicle emissions (e.g., the total reaction time to form ozone may exceed 1 h). Compared with anthropogenic sources, natural sources have less impact on ozone formation. Additionally, we found that NOx ranks fifth, indicating that the local atmosphere was more affected by VOCs than NOx during autumn, which is also consistent with the conclusion obtained based on the OBM model analysis (Fig. S5).

In addition, we calculated the OFP of different VOC sources, and the results are shown in Fig. 4(a). There is a certain difference between the results of OFP and SHAP. In the SHAP results, vehicle emission sources had the greatest impact on ozone formation, whereas in the OFP results they had a lower impact. Since vehicles can directly emit formaldehyde and HONO (Kurtenbach et al., 2001), the loss of VOCs can be further accelerated. Furthermore, considering that the Shenzhen area is close to the sea, Cl radicals have a significant impact on local atmospheric oxidation (Niu et al., 2022). In this regard, we calculated the loss rate of Cl radicals from different emission sources according to the study of Atkinson et al. (2006), and found that vehicle emissions were the fastest anthropogenic source of Cl radical loss (Fig. S4). The above evidence shows that VOCs from vehicle emission sources in Shenzhen easily react to promote ozone formation. However, OFP calculates the maximum amount of ozone that can theoretically be produced by VOCs from different sources, while VOC species that are rapidly consumed during active photochemical reactions are difficult to quantify using OFP (Zhang et al., 2022), especially from vehicle emission sources. Considering that there are still some limitations in MIR values (Qiu et al., 2020), we believe that OFP calculation may be more applicable in some other specific scenarios.

In this study, we obtained SHAP results by mining the measured air pollution data, which can largely avoid problems caused by time scale or regional differences. In some ways, it can make up for the shortcomings of OFP calculation, providing clues and ideas for related mechanism studies.

3.3. *Impacts of different driving factors on ozone pollution events*

We used the PMF-SHAP method to analyze the main driving factors of three ozone pollution events (September 27, 28 and October 11,

2019). In this study, an ozone pollution event was defined as an episode with hour ozone concentration continuously exceeding the standard level of 101.8 ppb for more than 3 h. From Fig. 5, we found that there were five main factors affecting the formation of ozone, followed by sources of vehicle emissions, regional transmission, gasoline leakage, natural emissions, and NOx. Among them, vehicle emission sources have the greatest impact, and the SHAP value was as high as 46.80, indicating that the local vehicle emission source plays a key role in ozone pollution events.

For the remaining four driving factors, the impact on ozone formation was quite similar. The SHAP values were 7.70, 7.15, 7.08 and 6.28 in descending order. However, compared with the daily situations, the value of SHAP has different degrees of change, of which the largest changes were gasoline leakage sources and NOx, and the relative increases are 261% and 171% respectively. Especially the source of gasoline leakage, the relative increase rate was significantly higher than others. From Figs. S3 and S6, the gasoline leakage source generally peaks at 9:00 during the day and reaches the lowest value at 16:00. The average concentration at 9:00 in daily situations was approximately 8.4 ppb, while the concentration at 9:00 on ozone pollution days was approximately 21 ppb, which is significantly higher than the daily situations, and the loss is also much larger. It shows that during ozone pollution days, human activities related to gasoline use will increase. Combined with unfavorable meteorological factors such as wind speed reduction and temperature increase, the accumulation of related VOCs in the atmospheric environment will further increase (Huang et al., 2022), resulting in a great increase in the probability of ozone pollution. It can be seen that the gasoline leakage source was one of the key factors for the formation of local ozone pollution.

Furthermore, we found that regardless of whether it is in daily situations or ozone pollution events, the impact of regional transmission sources on ozone was always at the forefront, indicating that the local ozone concentrations could be greatly affected by the surrounding areas.

4. Conclusions

This study innovatively proposed the PMF-SHAP method, which can well quantify the main emission sources that affect ozone formation. Based on the air pollution data of Shenzhen, we found that vehicle emission sources may have the greatest impact on the formation of ozone in Shenzhen. Meanwhile, in ozone pollution events, gasoline leakage sources may have the largest increase in the impact of ozone formation compared to other emission sources. Due to the PMF-SHAP method is more related to statistical theoretical knowledge, the conclusions of this study were somewhat different from traditional method. In some ways, it can better make up and address the shortcomings of traditional method. In addition, although the method mentioned in the study have high application potential in the environmental field, the interpretability of some factors was still limited. The PMF model involved in this study only mainly considered the primary emission of VOCs and didn't explore the secondary generated OVOCs and other species, which could limit the identification of VOCs sources. In the follow-up research, we are considering adding more data to explore the drivers of ozone formation in greater depth to provide more support for addressing ozone pollution problem.

Credit author statement

Yong Cheng: Methodology, Software, Formal analysis and Writing – Original draft. Xiao-Feng Huang: Conceptualization, Formal analysis and Writing – Review & Editing. Yan Peng: Data curation, Visualization. Meng-Xue Tang: Investigation. Bo Zhu: Investigation, Data curation. Shi-Yong Xia: Investigation. Ling-Yan He: Supervision, Writing – review & editing.

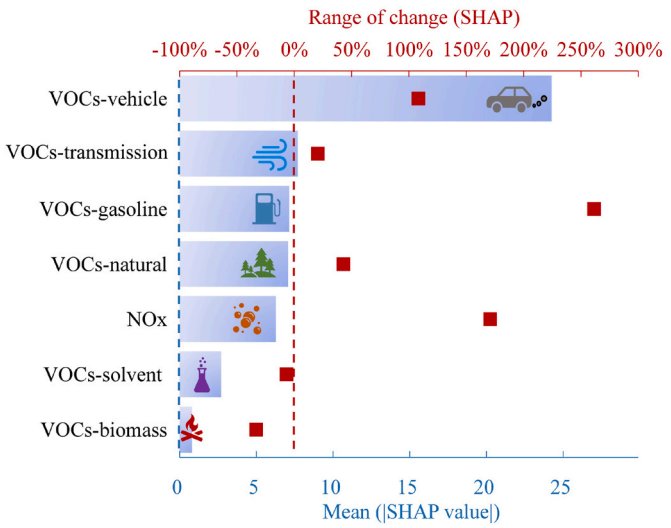

Fig. 5. Impacts of different driving factors on ozone formation in ozone pollution events. The red squares represent the SHAP changes for different drivers under ozone pollution events compared to the daily situations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by Key-Area Research and Development Program of Guangdong Province (2020B1111360003) and Science and Technology Plan of Shenzhen Municipality (JCYJ20200109120401943).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2022.120685.

References

Atkinson, R., Baulch, D.L., Cox, R.A., Crowley, J.N., Hampson, R.F., Hynes, R.G., Jenkin, M.E., Rossi, M.J., Troe, J., Subcommittee, I., 2006. Evaluated kinetic and photochemical data for atmospheric chemistry: volume II - gas phase reactions of organic species. Atmos. Chem. Phys. 6 (11), 3625–4055. https://doi.org/10.5194/acp-6-3625-2006.

Bland, G.D., Battifarano, M., Pradas del Real, A.E., Sarret, G., Lowry, G.V., 2022. Distinguishing engineered TiO2 nanomaterials from natural Ti nanomaterials in soil using spICP-TOFMS and machine learning. Environ. Sci. Technol. 56 (5), 2990–3001. https://doi.org/10.1021/acs.est.1c02950.

Carter, W.P.L., 2010. Development of the SAPRC-07 chemical mechanism. Atmos. Environ. 44 (40), 5324–5335. https://doi.org/10.1016/j.atmosenv.2010.01.026.

Cheng, Y., Zhang, H., Liu, Z., Chen, L., Wang, P., 2019. Hybrid algorithm for short-term forecasting of PM2.5 in China. Atmos. Environ. 200, 264–279. https://doi.org/10.1016/j.atmosenv.2018.12.025.

Cheng, Y., Zhu, Q., Peng, Y., Huang, X.-F., He, L.-Y., 2021. Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models. J. Clean. Prod. 326 https://doi.org/10.1016/j.jclepro.2021.129451.

Ding, D., Xing, J., Wang, S., Dong, Z., Zhang, F., Liu, S., Hao, J., 2022. Optimization of a NOx and VOC cooperative control strategy based on clean air benefits. Environ. Sci. Technol. 56 (2), 739–749. https://doi.org/10.1021/acs.est.1c04201.

Fang, T., Zhu, Y., Wang, S., Xing, J., Zhao, B., Fan, S., Li, M., Yang, W., Chen, Y., Huang, R., 2021. Source impact and contribution analysis of ambient ozone using

multi-modeling approaches over the Pearl River Delta region, China. Environ. Pollut. 289, 117860 https://doi.org/10.1016/j.envpol.2021.117860.
Huang, Y.S., Hsieh, C.C., 2020. VOC characteristics and sources at nine photochemical assessment monitoring stations in western Taiwan. Atmos. Environ. 240, 117741 https://doi.org/10.1016/j.atmosenv.2020.117741.
Huang, X.-F., Zhang, B., Xia, S.-Y., Han, Y., Wang, C., Yu, G.-H., Feng, N., 2020. Sources of oxygenated volatile organic compounds (OVOCs) in urban atmospheres in North and South China. Environ. Pollut. 261, 114152 https://doi.org/10.1016/j.envpol.2020.114152.
Huang, X.-F., Cao, L.-M., Tian, X.-D., Zhu, Q., Saikawa, E., Lin, L.-L., Cheng, Y., He, L.-Y., Hu, M., Zhang, Y.-H., Lu, K.-D., Liu, Y.-H., Daellenbach, K., Slowik, J.G., Tang, Q., Zou, Q.-L., Sun, X., Xu, B.-Y., Jiang, L., Shen, Y.-M., Ng, N.L., Prévôt, A.S.H., 2021. Critical role of simultaneous reduction of atmospheric odd oxygen for winter haze mitigation. Environ. Sci. Technol. 55 (17), 11557–11567. https://doi.org/10.1021/acs.est.1c03421.
Huang, J., Yuan, Z., Duan, Y., Liu, D., Fu, Q., Liang, G., Li, F., Huang, X., 2022. Quantification of temperature dependence of vehicle evaporative volatile organic compound emissions from different fuel types in China. Sci. Total Environ. 813, 152661 https://doi.org/10.1016/j.scitotenv.2021.152661.
Kurtenbach, R., Becker, K.H., Gomes, J.A.G., Kleffmann, J., Lörzer, J.C., Spittler, M., Wiesen, P., Ackermann, R., Geyer, A., Platt, U., 2001. Investigations of emissions and heterogeneous formation of HONO in a road traffic tunnel. Atmos. Environ. 35 (20), 3385–3394. https://doi.org/10.1016/S1352-2310(01)00138-8.
Liu, Y., Song, M., Liu, X., Zhang, Y., Hui, L., Kong, L., Zhang, Y., Zhang, C., Qu, Y., An, J., Ma, D., Tan, Q., Feng, M., 2020. Characterization and sources of volatile organic compounds (VOCs) and their related changes during ozone pollution days in 2016 in Beijing, China. Environ. Pollut. 257, 113599 https://doi.org/10.1016/j.envpol.2019.113599.
Liu, X., Lu, D., Zhang, A., Liu, Q., Jiang, G., 2022. Data-Driven machine learning in environmental pollution: gains and problems. Environ. Sci. Technol. 56 (4), 2124–2133. https://doi.org/10.1021/acs.est.1c06157.
Luecken, D.J., Napelenok, S.L., Strum, M., Scheffe, R., Phillips, S., 2018. Sensitivity of ambient atmospheric formaldehyde and ozone to precursor species and source types across the United States. Environ. Sci. Technol. 52 (8), 4668–4675. https://doi.org/10.1021/acs.est.7b05509.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. https://doi.org/10.1038/s42256-019-0138-9.
Nair, A.A., Yu, F., Campuzano-Jost, P., DeMott, P.J., Levin, E.J.T., Jimenez, J.L., Peischl, J., Pollack, I.B., Fredrickson, C.D., Beyersdorf, A.J., Nault, B.A., Park, M., Yum, S.S., Palm, B.B., Xu, L., Bourgeois, I., Anderson, B.E., Nenes, A., Ziemba, L.D., Moore, R.H., Lee, T., Park, T., Thompson, C.R., Flocke, F., Huey, L.G., Kim, M.J., Peng, Q., 2021. Machine learning uncovers aerosol size information from chemistry and meteorology to quantify potential cloud-forming particles. Geophys. Res. Lett. 48 (21) https://doi.org/10.1029/2021gl094133.
Niu, Y., Huang, X., Wang, H., Wang, S., Lin, X., Chen, Y., Zhu, B., Zhu, Q., He, L., 2022. Effects of nighttime heterogeneous reactions on the formation of secondary aerosols and ozone in the Pearl River Delta. Chin. Sci. Bull. 67, 2060. https://doi.org/10.1360/TB-2021-0638, 0023-074X.
Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., Nakaoku, Y., Suzuki, K., Hagihara, A., Noguchi, T., Iihara, K., Kitazume, K., Morioka, T., Yamazaki, S., Yoshida, T., Yamagata, Y., Nishimura, K., 2021. Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. Nat. Commun. 12 (1), 4575. https://doi.org/10.1038/s41467-021-24823-0.
Ou, J., Yuan, Z., Zheng, J., Huang, Z., Shao, M., Li, Z., Huang, X., Guo, H., Louie, P.K., 2016. Ambient ozone control in a photochemically active region: short-term despiking or long-term attainment? Environ. Sci. Technol. 50 (11), 5720–5728. https://doi.org/10.1021/acs.est.6b00345.
Qiu, W., Liu, Y., Tan, Z., Chen, X., Lu, K., Zhang, Y., 2020. Calculation of maximum incremental reactivity scales based on typical megacities in China. Chin. Sci. Bull. 65 (7), 610–621. https://doi.org/10.1360/tb-2019-0598.
Rybarczyk, Y., Zalakeviciute, R., 2021. Assessing the COVID-19 impact on air quality: a machine learning approach. Geophys. Res. Lett. 48 (4), e2020GL091202 https://doi.org/10.1029/2020GL091202.
Shao, M., Zhang, Y., Zeng, L., Tang, X., Zhang, J., Zhong, L., Wang, B., 2009. Ground-level ozone in the Pearl River Delta and the roles of VOC and NO(x) in its production. J. Environ. Manag. 90 (1), 512–518. https://doi.org/10.1016/j.jenvman.2007.12.008.
Shapley, L.S., 1953. A value for n-person games. Contributions to the Theory of Games II. Princeton University Press, Princeton, pp. 307–317. doi:10.1515/9781400881970-018.
Shu, L., Xie, M., Gao, D., Wang, T., Fang, D., Liu, Q., Huang, A., Peng, L., 2017. Regional severe particle pollution and its association with synoptic weather patterns in the Yangtze River Delta region, China. Atmos. Chem. Phys. 17 (21), 12871–12891. https://doi.org/10.5194/acp-17-12871-2017.
Song, Y., Shao, M., Liu, Y., Lu, S., Kuster, W., Goldan, P., Xie, S., 2007. Source apportionment of ambient volatile organic compounds in beijing. Environ. Sci. Technol. 41 (12), 4348–4353. https://doi.org/10.1021/es0625982.
Wu, R., Xie, S., 2017. Spatial distribution of ozone formation in China derived from emissions of speciated volatile organic compounds. Environ. Sci. Technol. 51 (5), 2574–2583. https://doi.org/10.1021/acs.est.6b03634.
Xing, J., Li, S., Jiang, Y., Wang, S., Ding, D., Dong, Z., Zhu, Y., Hao, J., 2020. Quantifying the emission changes and associated air quality impacts during the COVID-19 pandemic on the North China Plain: a response modeling study. Atmos. Chem. Phys. 20 (22), 14347–14359. https://doi.org/10.5194/acp-20-14347-2020.
Yan, Y., Peng, L., Li, R., Li, Y., Li, L., Bai, H., 2017. Concentration, ozone formation potential and source analysis of volatile organic compounds (VOCs) in a thermal power station centralized area: a study in Shuozhou, China. Environ. Pollut. 223, 295–304. https://doi.org/10.1016/j.envpol.2017.01.026.
Zahrt, A.F., Henle, J.J., Rose, B.T., Wang, Y., Darrow, W.T., Denmark, S.E., 2019. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science 363 (6424). https://doi.org/10.1126/science.aau5631 eaau5631.
Zhang, Y., Xue, L., Carter, W.P.L., Pei, C., Chen, T., Mu, J., Wang, Y., Zhang, Q., Wang, W., 2021. Development of ozone reactivity scales for volatile organic compounds in a Chinese megacity. Atmos. Chem. Phys. 21 (14), 11053–11068. https://doi.org/10.5194/acp-21-11053-2021.
Zhang, C., Song, Y., Wang, H., Zeng, L., Hu, M., Lu, K., Xie, S., Carter, W.P.L., 2022. Observation-based estimations of relative ozone impacts by using volatile organic compounds reactivities. Environ. Sci. Technol. Lett. 9 (1), 10–15. https://doi.org/10.1021/acs.estlett.1c00835.
Zheng, J., Shao, M., Che, W., Zhang, L., Zhong, L., Zhang, Y., Streets, D., 2009. Speciated VOC emission inventory and spatial patterns of ozone formation potential in the Pearl River Delta, China. Environ. Sci. Technol. 43 (22), 8580–8586. https://doi.org/10.1021/es901688e.
Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine learning: new ideas and tools in environmental science and engineering. Environ. Sci. Technol. 55 (19), 12741–12754. https://doi.org/10.1021/acs.est.1c01339.
Zhu, B., Huang, X.-F., Xia, S.-Y., Lin, L.-L., Cheng, Y., He, L.-Y., 2021. Biomass-burning emissions could significantly enhance the atmospheric oxidizing capacity in continental air pollution. Environ. Pollut. 285 https://doi.org/10.1016/j.envpol.2021.117523, 117523-117523.
Zhu, J.-J., Sima, N.Q., Lu, T., Menniti, A., Schauer, P., Ren, Z.J., 2022. Adaptive soft sensing of river flow prediction for wastewater treatment operation and risk management. Water Res. 220, 118714 https://doi.org/10.1016/j.watres.2022.118714.
Shrock, E., Fujimura, E., Kula, T., Timms, R.T., Lee, I.-H., Leng, Y., Robinson, M.L., Sie, B.M., Li, M.Z., Chen, Y., Logue, J., Zuiani, A., McCulloch, D., Lelis, F.J.N., Henson, S., Monaco, D.R., Travers, M., Habibi, S., Clarke, W.A., Caturegli, P., Laeyendecker, O., Piechocka-Trocha, A., Li, J.Z., Khatri, A., Chu, H.Y., Villani, A.-C., Kays, K., Goldberg, M.B., Hacohen, N., Filbin, M.R., Yu, X.G., Walker, B.D., Wesemann, D.R., Larman, H.B., Lederer, J.A., Elledge, S.J., Lavin-Parsons, K., Parry, B., Lilley, B., Lodenstein, C., McKaig, B., Charland, N., Khanna, H., Margolin, J., Gonye, A., Gushterova, I., Lasalle, T., Sharma, N., Russo, B.C., Rojas-Lopez, M., Sade-Feldman, M., Manakongtreecheep, K., Tantivit, J., Thomas, M.F., Abayneh, B.A., Allen, P., Antille, D., Armstrong, K., Boyce, S., Braley, J., Branch, K., Broderick, K., Carney, J., Chan, A., Davidson, S., Dougan, M., Drew, D., Elliman, A., Flaherty, K., Flannery, J., Forde, P., Gettings, E., Griffin, A., Grimmel, S., Grinke, K., Hall, K., Healy, M., Henault, D., Holland, G., Kayitesi, C., LaValle, V., Lu, Y., Luthern, S., Marchewka, J., Martino, B., McNamara, R., Nambu, C., Nelson, S., Noone, M., Ommerborn, C., Pacheco, L.C., Phan, N., Porto, F.A., Ryan, E., Selleck, K., Slaughenhaupt, S., Sheppard, K.S., Suschana, E., Wilson, V., Alter, G., Balazs, A., Bals, J., Barbash, M., Bartsch, Y., Boucau, J., Chevalier, J., Chowdhury, F., Einkauf, K., Fallon, J., Fedirko, L., Finn, K., Garcia-Broncano, P., Hartana, C., Jiang, C., Kaplonek, P., Karpell, M., Lam, E.C., Lefteri, K., Lian, X., Lichterfeld, M., Lingwood, D., Liu, H., Liu, J., Ly, N., Michell, A., Millstrom, I., Miranda, N., O'Callaghan, C., Osborn, M., Pillai, S., Rassadkina, Y., Reissis, A., Ruzicka, F., Seiger, K., Sessa, L., Sharr, C., Shin, S., Singh, N., Sun, W., Sun, X., Ticheli, H., Trocha-Piechocka, A., Worrall, D., Zhu, A., Daley, G., Golan, D., Heller, H., Sharpe, A., Jilg, N., Rosenthal, A., Wong, C., 2020. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. Science. 370(6520), eabd4250. doi:10.1126/science.abd4250.