toxics

MDPI

Article

Decoding the Primacy of Transportation Emissions of Formaldehyde Pollution in an Urban Atmosphere

Shi-Qi Liu [1,†], Hao-Nan Ma [1,†], Meng-Xue Tang [1,*], Yu-Ming Shao [1], Ting-Ting Yao [2], Ling-Yan He [1] and Xiao-Feng Huang [1]

[1] Key Laboratory for Urban Habitat Environmental Science and Technology, Peking University Shenzhen Graduate School, Shenzhen 518055, China; liushiqi67@stu.pku.edu.cn (S.-Q.L.); mahaonan@stu.pku.edu.cn (H.-N.M.)
[2] Shenzhen Academy of Metrology and Quality Inspection, Shenzhen 518107, China
* Correspondence: tangmx@pku.edu.cn
† These authors contributed equally to this work.

Abstract

Understanding the differential impacts of emission sources of volatile organic compounds (VOCs) on formaldehyde (HCHO) levels is pivotal to effectively mitigating key photochemical radical precursors, thereby enhancing the regulation of atmospheric oxidation capacity (AOC) and ozone formation. This investigation systematically selected and analyzed year-long VOC measurements across three urban zones in Shenzhen, China. Photochemical age correction methods were implemented to develop the initial concentrations of VOCs before source apportionment; then Positive Matrix Factorization (PMF) modeling resolved six primary sources: solvent usage (28.6–47.9%), vehicle exhaust (24.2–31.2%), biogenic emission (13.8–18.1%), natural gas (8.5–16.3%), gasoline evaporation (3.2–8.9%), and biomass burning (0.3–2.4%). A machine learning (ML) framework incorporating Shapley Additive Explanations (SHAP) was subsequently applied to evaluate the influence of six emission sources on HCHO concentrations while accounting for reaction time adjustments. This machine learning-driven nonlinear analysis demonstrated that vehicle exhaust nearly always emerged as the primary anthropogenic contributor in diverse functional zones and different seasons, with gasoline evaporation as another key contributor, while the traditional reactivity metric method, ozone formation potential (OFP), tended to underestimate the role of the two sources. This study highlights the primacy of strengthening emission reduction of transportation sectors to mitigate HCHO pollution in megacities.

Keywords: HCHO; PMF; machine learning; transportation emission; photochemical pollution

check for updates

Academic Editors: Xinlei Ge and Ming Wang

Received: 28 June 2025
Revised: 25 July 2025
Accepted: 26 July 2025
Published: 30 July 2025

Citation: Liu, S.-Q.; Ma, H.-N.; Tang, M.-X.; Shao, Y.-M.; Yao, T.-T.; He, L.-Y.; Huang, X.-F. Decoding the Primacy of Transportation Emissions of Formaldehyde Pollution in an Urban Atmosphere. *Toxics* **2025**, *13*, 643. https://doi.org/10.3390/toxics13080643

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Formaldehyde (HCHO) is among the most abundant reactive carbonyls in the troposphere and serves as a key precursor of HO_x radicals, thereby enhancing the atmospheric oxidative capacity (AOC) and driving tropospheric ozone (O_3) production [1–3]. Primary HCHO emissions arise from anthropogenic activities, including vehicular and industrial sources, and from natural sources such as biogenic emissions and biomass burning [4,5]. Furthermore, the involvement of volatile organic compounds (VOCs) in the photochemical oxidation processes mediated by $OH/NO_3/O_3$ in the atmosphere constitutes an indispensable secondary mechanism for HCHO formation [6,7]. The photolytic decomposition of HCHO predominantly yields OH· radicals, while its oxidation indirectly fosters HO_2 radical production through NO-to-NO_2 conversion, both processes critically strengthening the

AOC [8,9]. Given the severe ozone pollution in the Pearl River Delta (PRD) and the critical role of HCHO as both a significant photochemical byproduct and an active participant in radical-driven processes, it is imperative to investigate HCHO in this region [10–12].

Moreover, HCHO has been designated as a Group I human carcinogen by the International Agency for Research on Cancer (IARC) [13] and is widely recognized for its well-documented carcinogenicity and mutagenicity [7], posing a substantial threat to public health. Relevant studies have demonstrated that outdoor sources of HCHO (e.g., motor vehicle emissions and photochemical reactions) constitute a pivotal factor in shaping indoor HCHO pollution and its consequent health hazards [14,15]. Hence, establishing effective evidence-based strategies for HCHO control is essential to improving atmospheric photochemical conditions and enhancing public health in both indoor and outdoor settings.

The commonly used methods for source apportionment of HCHO include Multiple Linear Regression (MLR) [16], Photochemical Age-Based Parameterization (PCAP) method [3,10–12], and the Positive Matrix Factorization (PMF) model [17]. Cui et al. [18] pioneered a concurrent application of MLR, PCAP, and PMF models for HCHO source apportionment, revealing method-specific discrepancies in secondary contributions while highlighting the inherent limitations of individual approaches. Notably, the complex nonlinear relationships between HCHO and its emission sources challenge traditional source apportionment methods. It warrants further emphasis that VOC sources may have undergone photochemical consumption in the atmosphere prior to observation, rendering them unrepresentative of initial states [19,20]. Neglecting uncorrected photochemical losses systematically biases source apportionment and can exacerbate discrepancies between inferred emission sources and actual emissions [21–23].

Currently, advancements in AI have transcended traditional computational algorithms, enabling computers to progressively develop autonomous learning capabilities, which were the quintessential human traits [24]. Trained AI models have achieved subseasonal forecasting accuracy of complex events, such as global weather prediction, which was regarded as a significant breakthrough in the scientific community [25,26]. Machine learning (ML), as a transformative subfield of AI methodology, is increasingly providing novel analytical frameworks in environmental sciences for predicting and simulating pollutant behaviors [24,27,28]. ML has unique advantages through adaptive weight optimization in addressing complex nonlinear problems and identifying latent variables in emission matrices, demonstrating robust performance in terms of temporal and cost efficiency [29,30]. Moreover, with the advent of the Shapley Additive Explanations (SHAP) algorithm, the model interpretability of typically "black-box" ML has been substantially enhanced. The SHAP algorithm quantifies feature contributions of each factor by elucidating the interactions and dependencies among driving variables within ML models according to coalitional game theory [31]. The integrated application of these technologies holds the potential to elucidate the nonlinear dynamic transformations of HCHO in atmospheric photochemical pollution.

This study introduced an innovative PMF-ML framework to assess the impact of primary VOC emissions on HCHO concentrations. In contrast to conventional approaches that typically perform direct source apportionment of HCHO, this study first independently conducted PMF-based source apportionment of photochemically adjusted VOCs representing primary emissions. Subsequently, machine learning techniques were employed to establish a nonlinear correlation between VOCs and HCHO. This novel methodology offered a new perspective for evaluating the influence of emissions on HCHO levels, thereby providing valuable insights to inform the development of targeted photochemical pollution control strategies.

2. Materials and Methods

2.1. Study Areas and Photochemical Age Correction

Three monitoring sites were meticulously selected to epitomize quintessential industrial and urban typologies characterized by intricate traffic networks and substantial population densities. These archetypal stations conducted synchronized and real-time measurements of 116 VOC species and meteorological parameters in Shenzhen (Figure 1). Comprehensive geospatial characteristics, including coordinates and contextual information for each site, are presented in Table S1. The observational period spanned from 1 July 2023 to 30 June 2024, which ensured a full annual cycle of data collection. In addition, the comprehensive descriptions of analytical instruments and the principles of quality control are listed in Texts S1 and S2, respectively. Meanwhile, the daytime period (07:00–18:00) was captured when photolytic effects were pronounced by the atmospheric photolysis rate analyzer (PFS-100, Hangzhou Juguang, China) across the three Shenzhen sites, and this period was selected for photochemical age corrections of all VOC species included in the analysis of the source apportionment (detailed methods are illustrated in Text S3 and Figure S1) [20,21,32–35].

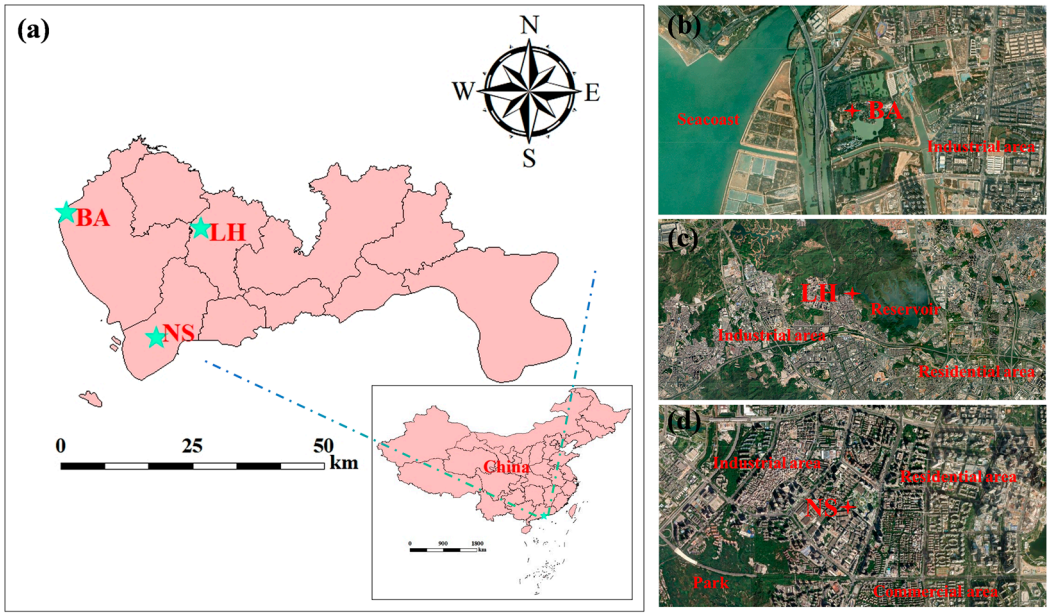

Figure 1. (**a**) The location conditions of the study areas in the Shenzhen region of China. Satellite maps of monitoring sites from Google Earth in (**b**) Bao'an district (BA), (**c**) Longhua district (LH), and (**d**) Nanshan district (NS).

2.2. Positive Matrix Factorization Model

Source apportionment and component analysis of VOC species were conducted utilizing the Positive Matrix Factorization (PMF v5.0) model [36,37]. PMF utilizes the concentrations of representative species combined with associated uncertainties (Unc) as model inputs. Species selection followed three stringent criteria: (1) prolonged photochemical lifetime, (2) source-specific tracer characteristics, and (3) high detection frequency [38]. Ultimately, 25 species were identified (Tables S2 and S3), and concentrations were normalized using photochemical age correction. The factor evaluations and model fitting assessments were meticulously executed in accordance with the PMF v5.0 User Guide (US EPA, 2014), and comprehensive methodologies and results are detailed in Text S4 [38–40]. Furthermore, the ozone formation potential (OFP) metric [41,42] was applied to evaluate the differential contributions of VOCs with varying reactivities to O_3 formation, thereby quantifying the

contributions of both VOC species and PMF-resolved sources to photochemical processes. Further methodological details are provided in Texts S5 and S6.

2.3. Machine Learning Model

This study developed machine learning models using Python 3.11 (Anaconda v2.1.1) with key packages including scikit-learn (v1.2.2) and XGBoost (v1.7.3). The Extreme Gradient Boosting (XGBoost) regression model [28,29] was developed to predict ambient HCHO concentrations using six PMF-derived source contribution variables and four meteorological parameters (temperature [T], relative humidity [RH], photolysis rate [JV], and wind speed [WS]) as independent predictors (number of decision trees = 300–1000, learning rate = 0.001–0.03, max depth = 3–10). The PMF results were based on hourly VOC data apportioned by principal sources, with photochemical aging taken into account. For robustness validation, Light Gradient Boosting Machine (LightGBM) and Random Forest (RF) models were established with identical predictor–response configurations [30]. Following the removal of incomplete data records (<10% missing values), the dataset was stratified into training (80%) and testing (20%) sets, with performance evaluation through R^2, root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) metrics. Hyperparameter optimization was conducted through Bayesian optimization using the BayesSearchCV algorithm from the scikit-optimize package (v0.10.1), implementing 50 iterations with five-fold cross-validation [43]. This approach replaced conventional grid search for more efficient hyperparameter space exploration. The framework and mathematical principles of machine learning and Bayesian optimization are comprehensively documented in Text S7. All above models exhibited robust predictive performance (testing set R^2: 0.79–0.83; RMSE: 0.79–0.92), with XGBoost consistently achieving the superior performance bounds (R^2: 0.80–0.83; RMSE: 0.79–0.85). To assess the relative contributions of predictors to the dependent variable HCHO, feature importance was analyzed using SHAP values computed with the Python SHAP package (v0.42.1) [31]. All model-related Python packages are permanently archived on the Python Package Index (PyPI) at https://pypi.org, accessed on 1 March 2025, ensuring computational reproducibility.

3. Results

3.1. Overview of VOC Components and Activities

The concentration levels of atmospheric pollutants are predominantly governed by the intensity of emission sources, prevailing meteorological conditions, and photochemical reaction dynamics [44]. Herein, the annual mean concentrations of total measured VOCs (TVOCs) at three representative Shenzhen sites were measured as 43.0 ± 45.0 ppb (BA), 39.6 ± 33.8 ppb (LH), and 32.2 ± 23.8 ppb (NS), respectively (Figure 2a). The annual component time series (Figures S2–S4) exhibited a uniform trend across all three sites. Specifically, oxygenated VOCs (OVOCs) predominated in traditional spring and summer compositions, likely attributable to conducive photochemical conditions, including elevated temperatures and increased solar radiation, which facilitated the secondary formation of OVOCs [10,12]. Conversely, alkanes and aromatics exhibited higher proportions during autumn and winter when photochemical activities were subdued, causing high-reactive aromatics to further accumulate, while the presence of low-reactive alkanes indicated heightened vehicular emissions [22]. These findings offered a nuanced understanding of VOC concentration fluctuations throughout Shenzhen. Synchronously, an overview of key meteorological parameters is described in Text S8.

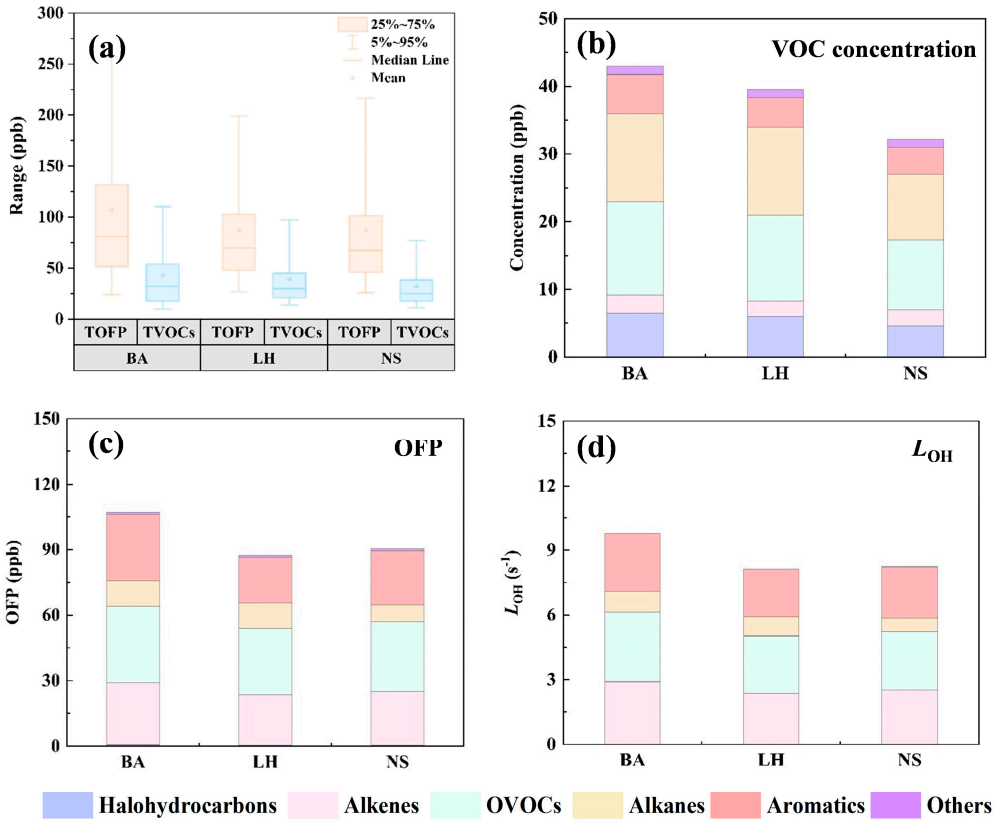

Figure 2. (**a**) Box plots of TVOCs and TOFP. The compositions of (**b**) concentration, (**c**) OFP, and (**d**) LOH of VOC species.

As illustrated in Figure 2b, OVOCs were the dominant VOC species at the BA site, followed by alkanes and halohydrocarbons, with solvent emissions from nearby industrial activities being a major influence [45]. In contrast, at the LH site, alkanes and OVOCs exhibited comparable dominance, reflecting the intricate emissions from vehicular activities and industrial solvents prevalent in the southern and western regions. At the NS site, OVOCs and alkanes again dominated, with acetone (9.20%) and acetaldehyde (7.37%) being the predominant species at NS (Figure S4). Acetaldehyde and acetone are primary OVOCs emitted from diesel and gasoline vehicle exhaust, respectively [3,46], implying that vehicle exhaust significantly shapes the VOC profile at NS [30]. Comprehensive details on VOC concentration levels across the three sites are provided in Table S2.

Annual mean total OFP (TOFP) reached 107.1 ± 103.5 ppb (BA), 87.3 ± 69.7 ppb (LH), and 90.3 ± 76.8 ppb (NS), respectively (Figure 2a). VOC groups such as OVOCs, alkenes, and aromatics consistently dominated the contributions to TOFP at three sites (Figure 2c). The most influential species included acetaldehyde, n-hexanal, ethylene, propylene, isoprene, toluene, and m,p-xylene, exhibiting similar composition patterns at all locations (Figure S5). In terms of OH· reactivity, OVOCs were the predominant contributors, followed by alkenes, aromatics, and alkanes in Figure 2d and Figure S5. At all three sites, the substantial proportions of styrene and isoprene underscored the significant roles of vehicular emissions and biogenic releases in modulating atmospheric oxidative capacity through OH· consumption [11,47].

3.2. Analysis Characteristics of VOC Contribution Apportionment

After species fingerprint screening (Table S3), factor diagnostics and model robustness validation (Table S4), the PMF model identified five anthropogenic and one biogenic emission sources after photochemical age correction in the daytime (7:00–18:00) across

three monitoring sites. The anthropogenic sources included biomass burning (BB), gasoline evaporation (GE), solvent usage (SU), natural gas (NG), and vehicle exhaust (VE), while the biogenic source (BE) was characterized by isoprene as the representative tracer. Source-specific profiles are depicted in Figure S6.

The first source was distinguished by substantial contributions from chloromethane, acetonitrile, and dichloromethane, which were established tracers of biomass burning [48,49]. Additionally, minor quantities of ethane (secondary combustion indicators) and ethylene that were associated with combustion emissions were detected [46]. Consequently, this source was classified as biomass burning (BB). The second source was predominantly composed of isopentane, n-pentane, and methyl tert-butyl ether (MTBE), where MTBE is considered a common gasoline additive, while isopentane and n-pentane are principal constituents of gasoline [50,51] and supplemented by minor contributions from propane and butane related to petroleum products [45]. Complementarily, the ratio of isopentane/n-pentane at the three sites was 2.45, 2.83, and 2.26 (Table S5), falling within the typical range for gasoline evaporation (1.8–4.0) [22], thus leading to the identification of this source as gasoline evaporation (GE). The third source featured representative aromatic hydrocarbons commonly associated with solvent usage, such as benzene, toluene, ethylbenzene, and m,p-xylene [23,49]. This source was designated as solvent usage (SU), which mainly includes solvent usage from industries, especially in the industrial areas of Shenzhen. The fourth source encompassed ethane, propane, ethylene, propylene, and acetylene, along with minor amounts of isobutane, n-butane, and dichloromethane. Ethylene and propylene are combustion byproducts [46], ethane is a significant emission from natural gas combustion [52,53], and acetylene is closely linked to combustion processes [54]. Meanwhile, the ratio of isobutane/n-butane exceeding 0.6 (Table S5) across all sites indicated the predominance of natural gas combustion [22]. Thus, this source was identified as natural gas (NG), aligning with the clean energy infrastructure of Shenzhen, which primarily utilizes natural gas [45], and this identification is corroborated by similar profiles in other natural gas source studies [22,55]. Further validation of the fourth source was achieved through offline VOC measurements under simulated residential natural gas combustion using a GC-MS system (GCMS 8890-5977B, Agilent, Santa Clara, CA, USA). The VOC source profile obtained from three independent experiments post-background correction (Figure S7) was congruent with the PMF-resolved source profile. The fifth source was dominated by 2-methylpentane, 3-methylpentane, styrene, and various alkanes, where 2-methylpentane and 3-methylpentane are major components of vehicle emissions [56], while C_2-C_3 alkanes and alkenes are pivotal exhaust species from gasoline vehicles [3,22]. Long-chain alkanes of C_8-C_{10} are primarily linked to diesel engine emissions [19], and styrene serves as a tracer for vehicle exhaust [47]. The ratio of propane/ethane below 3.0 across all sites substantiated the influence of vehicle emissions [22]. Therefore, this source was classified as vehicle exhaust (VE).

Figure 3a depicts the source contributions to VOC concentrations and OFP at the three sites. The concentrations in source composition of all three sites were dominated by VE, accounting for 37.9% at BA, 34.6% at LH, and 33.8% at NS, respectively. However, the potential photochemical pollution contribution from these sources could not be fully assessed solely based on their concentrations due to the differences in reactivity among emission sources. An examination of OFP contributions across emission sources, as presented in Figure 3b, reveals that SU was the predominant contributor at BA and NS, accounting for 47.9% and 30.9% of TOFP, respectively, while VE was the leading source at LH, contributing 31.2%. Among the three monitoring sites, SU and VE consistently ranked among the most significant contributors. Notably, SU emerged as the dominant anthropogenic source at BA, likely due to the high density of industrial parks and related facilities in the area. The

proportions of biogenic emissions were 13.8%, 18.1%, and 17.9% at three sites, although their contribution was proportionally trivial with respect to anthropogenic sources, despite the dominance of BE, which was further accentuated post-photochemical correction [22,57].

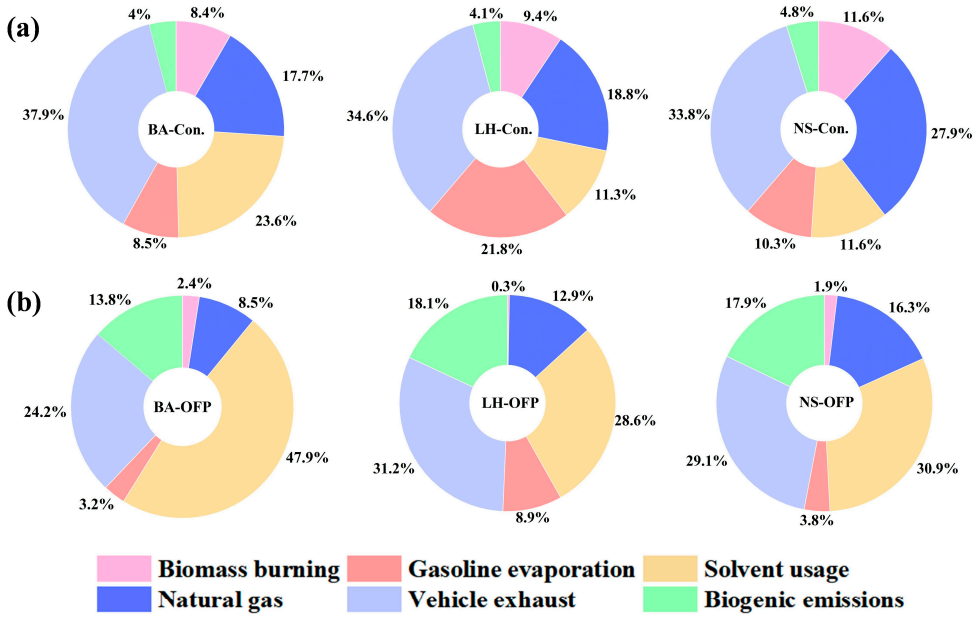

Figure 3. The daytime (7:00–18:00) emission source composition of (**a**) VOCs concentration and (**b**) OFP.

Overall, while conventional PMF modeling offered preliminary insights into emission source structures, especially following photochemical age correction, PMF highlighted limitations in resolving the nuanced differences and nonlinear relationships of source contributions to photochemical reactions. Additionally, the potential influence of meteorological factors, regional pollutant transport from other areas, and possible overestimation of actual pre-photochemical correction contributions requires further mechanistic investigation and exploration.

3.3. Identification of the Critical Impact of Factors by Machine Learning

Upon being released into the atmosphere, VOCs participate in a cascade of complex photochemical reactions, producing a series of photo-oxidative radicals at successive stages, which eventually culminate in ozone formation. Among the photochemical products, HCHO plays a pivotal intermediary role by amplifying HO_2 radical formation through its photolysis and subsequent oxidation pathways. Model simulations and RIR-based sensitivity analyses further demonstrate that secondary HCHO, predominantly derived from CH_3CHO, substantially accelerates O_3 production, accounting for up to 45.7% of HO_2 generation within the CH_3CHO degradation cascade [17]. However, atmospheric photochemical reactions evolve through a highly intricate network of pathways, influenced not only by local precursor intensities but also by meteorological forcing such as elevated temperatures, low humidity, and intense solar radiation, which collectively enhance the production of photo-oxidants including O_3 [58]. Utilizing HCHO can provide a more precise and effective means of elucidating the mechanisms of indigenous photochemical pollution, especially in megacities with highly complex meteorological conditions.

Particularly, traditional PMF-derived source apportionment lacks the capacity to quantify the specific contributions of various meteorological drivers. Failure to account for meteorological influences may compromise the authenticity of source apportionment assessments in a complex atmospheric environment. In this study, a machine learning

approach was employed to discern the nonlinear contributions of intricate meteorological features, biogenic and anthropogenic VOC emission sources to photochemical pollution. Considering that HCHO possesses complex nonlinear relationships with VOC precursors, and photochemical reactions entail certain temporal lags [30]. To reduce uncertainties caused by temporal–dynamic differences in the continuity of photochemical reactions and to resolve characteristic timescales, the reaction time correction methods were developed through systematic evaluating the influence of the factors with time lags ranging from 0 to 3 h during daytime hours (7:00–18:00). Summarized results in Table S6 indicated that the optimal fits for HCHO were achieved with a 1-h and 1.5-h lag at all three sites ($R^2 > 0.80$), which also implied the photochemical reaction timescale may be between 60 and 90 min. Furthermore, the SHAP analysis was introduced to ascertain the driving contributions of the ten factors to HCHO generation through summary plots at the three sites (Figures S9–S11). Simultaneously, multi-model validation (XGBoost, LightGBM, RF) under identical 1-hour lag optimal conditions demonstrated consistent driving factor results, confirming XGBoost-derived HCHO output robustness while eliminating single-model bias (Figure S11). Among the models, XGBoost outperformed LightGBM and RF, achieving higher R2 values and lower RMSE, MSE, and MAE, as listed in Table S7. Furthermore, all three models demonstrated excellent fits across the three sites ($R^2 > 0.90$), and the observed and predicted values were in satisfactory agreement under the optimal parameters, indicating that the models were sufficiently trained (Figure S12). In summary, HCHO demonstrated superior representativeness, and the XGBoost model effectively simulated the reaction mechanisms of the ten driving factors under photochemical conditions of the 1-hour time lag. All models were comprehensively trained and validated, providing reliable insights into the contributions of VOCs and meteorological factors to photochemical processes.

Figure 4 presents the SHAP values of driving factors for HCHO generation under optimal test conditions, alongside comparisons of OFP. Meanwhile, all driving factors passed the variance inflation factor (VIF) test with high variable tolerance and no multicollinearity problem (VIF < 10, indicating negligible multicollinearity), further excluding the misleading effects of collinearity (Figure S13). As illustrated in Figure 4a–c, SHAP analysis revealed that RH and T played pivotal roles in HCHO generation, where lower humidity (blue dots) and higher temperatures (red dots) may have a positive correlation with HCHO production. This outcome is consistent with the frequent occurrence of high photochemical pollution under high-temperature and low-humidity meteorological conditions [51,54]. The JV exhibited a moderate contribution, suggesting a positive influence of solar radiation on the photochemical reactions driving atmospheric oxidant formation [59]. Conversely, the relatively low contribution of WS across the three sites reflected the limited effect of regional transport and lateral dispersion of pollution plumes, thus supporting the predominance of local emission sources. Among emission sources of VOCs, the concentrations of all factors were positively correlated with SHAP-derived HCHO values, reaffirming the significance of biogenic and anthropogenic emissions in driving HCHO generation, where higher concentrations of emissions were associated with greater impacts on HCHO production. In Figure 4d–f, the nonlinear perspective offered by the machine learning approach revealed a result that differs from the traditional linear view based on the OFP method. This contrast underscores the scientific validity and enhanced interpretative capacity of nonlinear approaches in elucidating the intricacies of atmospheric chemical processes. From this nonlinear perspective, VE emerged as the dominant anthropogenic source across all three sites during daytime hours, aligning with the high daytime activity of VE emissions in megacities, particularly near major roadways. Furthermore, VE emissions contribute to the production of active species such as HONO and OVOCs, which generate OH· and disrupt

the NO_x-O3 balance, enhancing HCHO formation and photochemical pollution [47,60]. This observation is consistent with previous regional studies [30], emphasizing the critical role of VE in driving urban photochemical processes during daytime. At the BA site, the heightened influence of GE and NG relative to SU, contrasting with the SU dominance indicated by PMF-derived OFP results, suggested the potential impact of GE and NG on HCHO generation [46,53]. From the perspective of LH, SU and GE were the principal active sources after VE due to the presence of numerous industrial parks and gas stations in the vicinity. VE and GE dominated the contributions at NS, which was a representative urban commercial district with high traffic density along major surrounding roads. Notably, the elevated contribution of GE at three sites highlighted the potential influence of gasoline evaporation, characterized by species such as alkanes, on HCHO generation. Moreover, the integrated emissions from VE and GE, which collectively constitute transportation-related sources in megacities, predominantly account for anthropogenic emissions as revealed through machine learning-based comprehensive assessments, which appear to substantiate the pivotal role of traffic-derived pollution in metropolitan atmospheric environments. While some features may influence HCHO indirectly through their effects on intermediate variables such as VOCs, the machine learning framework employed in this study is capable of capturing both direct and indirect relationships, which are reflected comprehensively in the model results.

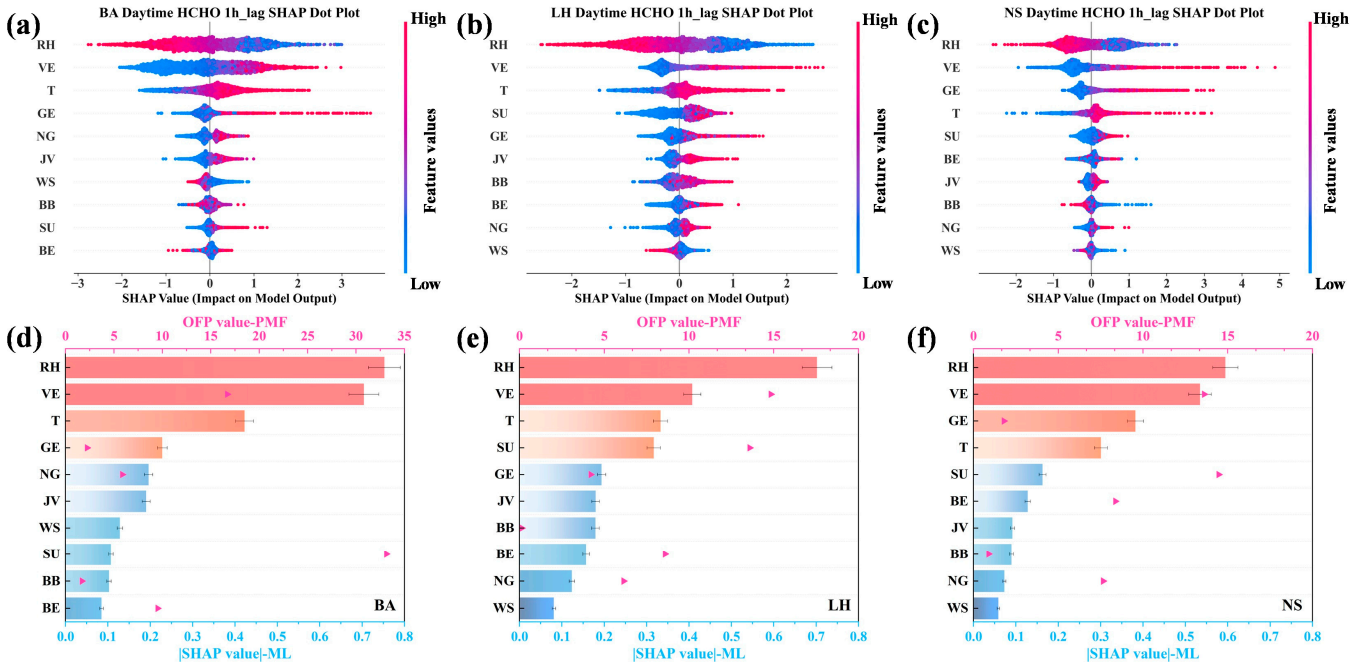

Figure 4. (**a**–**c**) The SHAP dot plots of HCHO. (**d**–**f**) The SHAP values and OFP from emission sources (RH = relative humidity; T = temperature; JV = photolysis rate; WS = wind speed; BB = biomass burning; VE = vehicle exhaust; GE = gasoline evaporation; NG = natural gas; SU = solvent usage; BE = background emissions).

To further reveal the mechanistic insights of the impact of driving factors under varying climatic conditions, the SHAP values were meticulously stratified into dry (spanning from November to April) and wet seasons (encompassing May to October), grounded in the recognition that Shenzhen is situated within a subtropical monsoon climate zone [61]. As depicted in Figure 5a–c, RH exerted a dominant role during the wet season, whereas T predominantly drove pollutant variations in the dry season, suggesting a climate sensitivity-dependent mechanism of meteorological factors. Supplementally, the atmospheric dynamics are governed by three distinct air masses in Shenzhen, including continental (northerly),

marine (southerly), and coastal mixed air masses [62]. Seasonal shifts in air mass dominance are evident, where clean marine air masses (southern origin) prevail in the wet season, while polluted continental air masses from the north dominate the dry season [61]. Notably, although the spatial–seasonal disparities were continually underscored in regional pollution dynamics, anthropogenic emissions, especially the transportation sources, showed unapparent changes in the wet and dry seasons from nonlinear views by machine learning, which further highlighted the urgency and criticality of traffic atmospheric pollution in megacities. Meanwhile, the persistent VE dominance in megacities with high-density traffic and heterogeneous fleet profiles underscores unresolved challenges in electrification and emission standard enforcement. Contrarily, the traditional assessment method based on parameters may overlook or underestimate the impact of transportation sources while disproportionately amplifying contributions from short-lived photochemical sources (e.g., solvent usage and biogenic emissions).

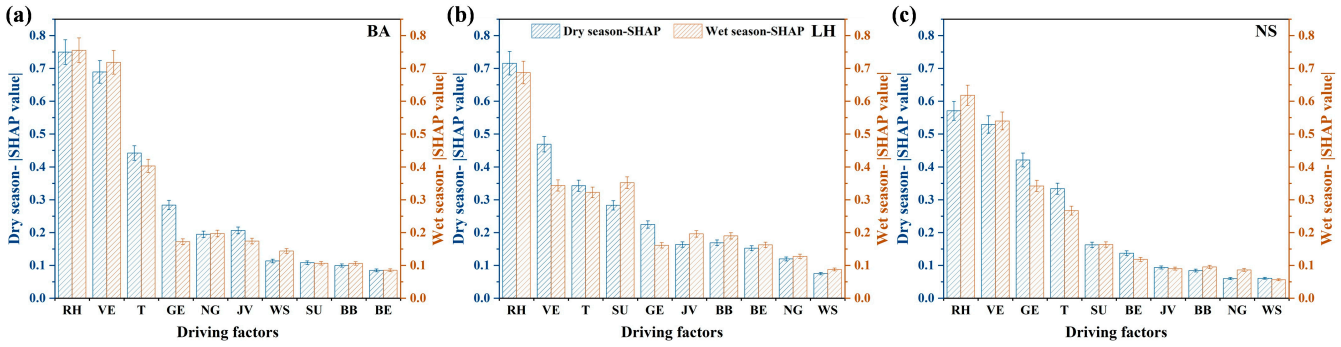

Figure 5. (**a**–**c**) The SHAP histograms of HCHO in wet and dry seasons at three sites (feature abbreviations are the same as in Figure 4).

4. Conclusions

Based on year-long VOC observations at three urban sites in Shenzhen, China, this study proposed an innovative diagnostic framework that integrates PMF with machine learning (XGBoost) and SHAP analysis to directly quantify the nonlinear contributions of VOC emissions to HCHO. Comprehensive observations revealed that OVOCs were the dominant VOCs. OVOCs, alkenes, and aromatics consistently contributed most to both OFP and OH· reactivity, with acetaldehyde, isoprene, and styrene identified as key reactive species influencing AOC. Using photochemical age-corrected VOCs data, PMF resolved six major sources, including solvent usage, vehicle exhaust, biogenic emission, natural gas, gasoline evaporation, and biomass burning. A novel reaction time correction framework was also established to address uncertainties in photochemical reaction timescales. SHAP analysis revealed that vehicle exhaust was the primary driver of HCHO formation, in contrast to traditional reactivity-based indicators (OFP), which tended to emphasize solvent use. This approach overcomes limitations of traditional indicators of VOC reactivity, such as OFP and L_{OH}. Moreover, the increased importance of vehicle exhaust in HCHO formation revealed by the machine learning framework may be attributed to potential mechanisms such as the co-occurrence of HCHO with other oxidation-enhancing compounds near vehicular sources, and the Cl·-induced acceleration of alkane activation from vehicular emissions in coastal megacities. Overall, our findings underscore the dominant role of vehicular emissions in urban photochemical pollution, providing scientific evidence to support the integration of traffic planning with photochemical control strategies. These insights also contribute to mitigating public health risks associated with HCHO exposure in residential environments. While the SHAP-based analysis effectively captures nonlinear

source–product relationships, future efforts will focus on incorporating causal inference to more robustly resolve the underlying chemical mechanisms.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/toxics13080643/s1. Text S1: Instrument and Analysis. Text S2: Instrument Principles and Quality Control. Text S3: Introduction of Photochemical Age Correction [19–22,32–35,63–66]. Text S4: Principles and Parameter Selection for PMF Analysis [19,36,37,39,40,67]. Text S5: Calculation of Ozone Formation Potential [41,42,45]. Text S6: Estimation of OH· Consumption Rate [41,63,64]. Text S7: The Framework of Machine Learning and Bayesian Optimization [24,27–31,43]. Text S8: The Overview of Meteorological Factors. Figure S1: (a–c) The relationship between the concentration of m,p-Xylene and ethylbenzene at three sites; (d–f) the diurnal of the X/E at three sites. Figure S2: The time series of (a) concentrations of total measured VOCs (TVOCs), (b) temperature (°C), (c) relative humidity (%), and (d) wind speed (m/s) at three sites. Figure S3: Annual time series of VOC components percentage at (a) BA, (b) LH, and (c) NS. Figure S4: Solar maps of 116 VOC species concentrations and components at BA, LH, and NS. Figure S5: Ozone formation potential ranking and proportion of OH· consumption of VOC species. Figure S6: The PMF profiles of (a) BA, (b) LH, and (c) NS. Figure S7: The natural gas source profiles of (a) offline samples, (b) BA, (c) LH, and (d) NS. Figure S8: The SHAP model of HCHO output at different time lags (0–3 h) under the influence of multiple factors by XGBoost at BA. Figure S9: The SHAP model of HCHO output at different time lags (0–3 h) under the influence of multiple factors by XGBoost at LH. Figure S10: The SHAP model of HCHO output at different time lags (0–3 h) under the influence of multiple factors by XGBoost at NS. Figure S11: The SHAP model of HCHO output at 1 h lag under the influence of multiple factors by XGBoost, LightGBM, and Random Forest. Figure S12: Scatter plots of observation values from PMF and prediction values from ML at three sites under (a–c) XGBoost, (d–f) LightGBM, (g–i) RF, and (j–l) the time series of prediction and observation by XGBoost and PMF at three sites. Figure S13: Heatmap and bar charts of variance inflation factor (VIF) between multiple factors output from the SHAP values. Table S1: The information of the monitoring sites. Table S2: The concentration levels of monitoring species. Table S3: The list of representative species for PMF. Table S4: The test results of PMF. Table S5: Ratio results of key species using the characteristic ratio method. Table S6: Machine learning outputs test results at various intervals for HCHO by XGBoost at three sites. Table S7: Machine learning outputs test results under optimal conditions by various models at three sites.

Author Contributions: Conceptualization, M.-X.T., T.-T.Y., L.-Y.H., and X.-F.H.; methodology, M.-X.T. and X.-F.H.; software, S.-Q.L. and H.-N.M.; validation, M.-X.T.; formal analysis, S.-Q.L. and H.-N.M.; data curation, S.-Q.L., H.-N.M., and Y.-M.S.; writing—original draft preparation, S.-Q.L. and H.-N.M.; writing—review and editing, M.-X.T. and X.-F.H.; visualization, S.-Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2023YFC3709201).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are available in the main text.

Conflicts of 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.

References

1. de Blas, M.; Ibáñez, P.; García, J.A.; Gómez, M.C.; Navazo, M.; Alonso, L.; Durana, N.; Iza, J.; Gangoiti, G.; de Cámara, E.S. Summertime high resolution variability of atmospheric formaldehyde and non-methane volatile organic compounds in a rural background area. *Sci. Total Environ.* **2019**, *647*, 862–877. [CrossRef] [PubMed]
2. Qian, H.; Xu, B.; Xu, Z.; Zou, Q.; Zi, Q.; Zuo, H.; Zhang, F.; Wei, J.; Pei, X.; Zhou, W.; et al. Anthropogenic Oxygenated Volatile Organic Compounds Dominate Atmospheric Oxidation Capacity and Ozone Production via Secondary Formation of Formaldehyde in the Urban Atmosphere. *ACS ES&T Air* **2025**, *2*, 1033–1041. [CrossRef]
3. Wang, C.; Huang, X.F.; Han, Y.; Zhu, B.; He, L.Y. Sources and Potential Photochemical Roles of Formaldehyde in an Urban Atmosphere in South China. *J. Geophys. Res. Atmos.* **2017**, *122*, 11–934. [CrossRef]
4. Lowe, D.C.; Schmidt, U. Formaldehyde (HCHO) measurements in the nonurban atmosphere. *J. Geophys. Res. Ocean.* **2012**, *88*, 10844–10858. [CrossRef]
5. Luecken, D.J.; Napelenok, S.L.; Strum, M.; Scheffe, R.; Phillips, S. Sensitivity of Ambient Atmospheric Formaldehyde and Ozone to Precursor Species and Source Types Across the United States. *Environ. Sci. Technol.* **2018**, *52*, 4668–4675. [CrossRef]
6. Liu, T.; Lin, Y.; Chen, J.; Chen, G.; Yang, C.; Xu, L.; Li, M.; Fan, X.; Zhang, F.; Hong, Y. Pollution mechanisms and photochemical effects of atmospheric HCHO in a coastal city of southeast China. *Sci. Total Environ.* **2023**, *859*, 160210. [CrossRef]
7. Yang, X.; Zhang, G.; Sun, Y.; Zhu, L.; Wei, X.; Li, Z.; Zhong, X. Explicit modeling of background HCHO formation in southern China. *Atmos. Res.* **2020**, *240*, 104941. [CrossRef]
8. Ling, Z.H.; Zhao, J.; Fan, S.J.; Wang, X.M. Sources of formaldehyde and their contributions to photochemical O_3 formation at an urban site in the Pearl River Delta, southern China. *Chemosphere* **2017**, *168*, 1293–1301. [CrossRef] [PubMed]
9. Zhang, K.; Duan, Y.; Huo, J.; Huang, L.; Wang, Y.; Fu, Q.; Wang, Y.; Li, L. Formation mechanism of HCHO pollution in the suburban Yangtze River Delta region, China: A box model study and policy implementations. *Atmos. Environ.* **2021**, *267*, 118755. [CrossRef]
10. Huang, X.-F.; Wang, C.; Zhu, B.; Lin, L.-L.; He, L.-Y. Exploration of sources of OVOCs in various atmospheres in southern China. *Environ. Pollut.* **2019**, *249*, 831–842. [CrossRef]
11. Wei, C.-B.; Yu, G.-H.; Cao, L.-M.; Han, H.-X.; Xia, S.-Y.; Huang, X.-F. Tempo-spacial variation and source apportionment of atmospheric formaldehyde in the Pearl River Delta, China. *Atmos. Environ.* **2023**, *312*, 120016. [CrossRef]
12. Zhu, B.; Han, Y.; Wang, C.; Huang, X.; Xia, S.; Niu, Y.; Yin, Z.; He, L. Understanding primary and secondary sources of ambient oxygenated volatile organic compounds in Shenzhen utilizing photochemical age-based parameterization method. *J. Environ. Sci.* **2019**, *75*, 105–114. [CrossRef]
13. Gatlin, C.G. IARC Classifies Formaldehyde as Carcinogenic. *Oncol. Times* **2004**, *26*, 72.
14. Zhang, H.; Zheng, Z.; Yu, T.; Liu, C.; Qian, H.; Li, J. Seasonal and diurnal patterns of outdoor formaldehyde and impacts on indoor environments and health. *Environ. Res.* **2022**, *205*, 112550. [CrossRef]
15. Zhu, L.; Jacob, D.J.; Keutsch, F.N.; Mickley, L.J.; Scheffe, R.; Strum, M.; González Abad, G.; Chance, K.; Yang, K.; Rappenglück, B.; et al. Formaldehyde (HCHO) As a Hazardous Air Pollutant: Mapping Surface Air Concentrations from Satellite and Inferring Cancer Risks in the United States. *Environ. Sci. Technol.* **2017**, *51*, 5650–5657. [CrossRef] [PubMed]
16. Bao, J.; Li, H.; Wu, Z.; Zhang, X.; Zhang, H.; Li, Y.; Qian, J.; Chen, J.; Deng, L. Atmospheric carbonyls in a heavy ozone pollution episode at a metropolis in Southwest China: Characteristics, health risk assessment, sources analysis. *J. Environ. Sci.* **2022**, *113*, 40–54. [CrossRef]
17. Qian, X.; Shen, H.; Chen, Z. Characterizing summer and winter carbonyl compounds in Beijing atmosphere. *Atmos. Environ.* **2019**, *214*, 116845. [CrossRef]
18. Cui, Y.; Hua, J.; He, Q.; Guo, L.; Wang, Y.; Wang, X. Comparison of three source apportionment methods based on observed and initial HCHO in Taiyuan, China. *Sci. Total Environ.* **2024**, *926*, 171828. [CrossRef]
19. Tong, M.; Zhang, Y.; Zhang, H.; Chen, D.; Pei, C.; Guo, H.; Song, W.; Yang, X.; Wang, X. Contribution of Ship Emission to Volatile Organic Compounds Based on One-Year Monitoring at a Coastal Site in the Pearl River Delta Region. *J. Geophys. Res. Atmos.* **2024**, *129*, e2023JD039999. [CrossRef]
20. Wang, Z.; Shi, Z.; Wang, F.; Liang, W.; Shi, G.; Wang, W.; Chen, D.; Liang, D.; Feng, Y.; Russell, A.G. Implications for ozone control by understanding the survivor bias in observed ozone-volatile organic compounds system. *NPJ Clim. Atmos. Sci.* **2022**, *5*, 39. [CrossRef]
21. He, Z.; Wang, X.; Ling, Z.; Zhao, J.; Guo, H.; Shao, M.; Wang, Z. Contributions of different anthropogenic volatile organic compound sources to ozone formation at a receptor site in the Pearl River Delta region and its policy implications. *Atmos. Chem. Phys.* **2019**, *19*, 8801–8816. [CrossRef]
22. Wu, Y.; Fan, X.; Liu, Y.; Zhang, J.; Wang, H.; Sun, L.; Fang, T.; Mao, H.; Hu, J.; Wu, L.; et al. Source apportionment of VOCs based on photochemical loss in summer at a suburban site in Beijing. *Atmos. Environ.* **2023**, *293*, 119459. [CrossRef]
23. Zheng, H.; Kong, S.; Chen, N.; Niu, Z.; Zhang, Y.; Jiang, S.; Yan, Y.; Qi, S. Source apportionment of volatile organic compounds: Implications to reactivity, ozone formation, and secondary organic aerosol potential. *Atmos. Res.* **2021**, *249*, 105344. [CrossRef]

24. Liu, X.; Lu, D.; Zhang, A.; Liu, Q.; Jiang, G. Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. *Environ. Sci. Technol.* **2022**, *56*, 2124–2133. [CrossRef] [PubMed]
25. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [CrossRef] [PubMed]
26. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef]
27. Chen, X.; Ma, W.; Zheng, F.; Wang, Z.; Hua, C.; Li, Y.; Wu, J.; Li, B.; Jiang, J.; Yan, C.; et al. Identifying Driving Factors of Atmospheric N2O5 with Machine Learning. *Environ. Sci. Technol.* **2024**, *58*, 11568–11577. [CrossRef] [PubMed]
28. Ogata, S.; Takegami, M.; Ozaki, T.; Nakashima, T.; Onozuka, D.; Murata, S.; Nakaoku, Y.; Suzuki, K.; Hagihara, A.; Noguchi, T.; et al. Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nat. Commun.* **2021**, *12*, 4575. [CrossRef]
29. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
30. Cheng, Y.; Huang, X.-F.; Peng, Y.; Tang, M.-X.; Zhu, B.; Xia, S.-Y.; He, L.-Y. A novel machine learning method for evaluating the impact of emission sources on ozone formation. *Environ. Pollut.* **2023**, *316*, 120685. [CrossRef]
31. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4768–4777.
32. Fang, H.; Huang, X.; Xiao, S.; Lowther, S.; Fu, X.; Zhang, Y.; Wu, T.; Hu, W.; Zhang, G.; Ding, X.; et al. Intermediate-Volatility Organic Compounds Observed in a Coastal Megacity: Importance of Non-Road Source Emissions. *J. Geophys. Res. Atmos.* **2022**, *127*, e2022JD037301. [CrossRef]
33. McKeen, S.A.; Liu, S.C. Hydrocarbon ratios and photochemical history of air masses. *Geophys. Res. Lett.* **1993**, *20*, 2363–2366. [CrossRef]
34. Stroud, C.A.; Roberts, J.M.; Goldan, P.D.; Kuster, W.C.; Murphy, P.C.; Williams, E.J.; Hereid, D.; Parrish, D.; Sueper, D.; Trainer, M.; et al. Isoprene and its oxidation products, methacrolein and methylvinyl ketone, at an urban forested site during the 1999 Southern Oxidants Study. *J. Geophys. Res. Atmos.* **2001**, *106*, 8035–8046. [CrossRef]
35. Zheng, H.; Kong, S.; Yan, Y.; Chen, N.; Yao, L.; Liu, X.; Wu, F.; Cheng, Y.; Niu, Z.; Zheng, S.; et al. Compositions, sources and health risks of ambient volatile organic compounds (VOCs) at a petrochemical industrial park along the Yangtze River. *Sci. Total Environ.* **2020**, *703*, 135505. [CrossRef]
36. Paatero, P.; Hopke, P.K.; Song, X.-H.; Ramadan, Z. Understanding and controlling rotations in factor analytic models. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 253–264. [CrossRef]
37. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126. [CrossRef]
38. Shao, M.; Lu, S.; Liu, Y.; Xie, X.; Chang, C.; Huang, S.; Chen, Z. Volatile organic compounds measured in summer in Beijing and their role in ground-level ozone formation. *J. Geophys. Res. Atmos.* **2009**, *114*, D00G06. [CrossRef]
39. Brown, S.G.; Eberly, S.; Paatero, P.; Norris, G.A. Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and guidance on reporting PMF results. *Sci. Total Environ.* **2015**, *518*, 626–635. [CrossRef]
40. Peng, X.; Huang, X.-F.; Wei, F.-H.; Yan, R.-H.; Tang, M.-X.; Ji, J.-P.; He, L.-Y. Identifying the key drivers in retrieving blue sky during rapid urbanization in Shenzhen, China. *J. Clean. Prod.* **2022**, *356*, 131829. [CrossRef]
41. Carter, W.P.L. Development of the SAPRC-07 chemical mechanism. *Atmos. Environ.* **2010**, *44*, 5324–5335. [CrossRef]
42. Carter, W.P.L. Development of Ozone Reactivity Scales for Volatile Organic Compounds. *Air Waste* **2012**, *44*, 881–899. [CrossRef]
43. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian optimization of machine learning algorithms. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 2, pp. 2951–2959.
44. Liu, J.; Mauzerall, D.L.; Chen, Q.; Zhang, Q.; Song, Y.; Peng, W.; Klimont, Z.; Qiu, X.; Zhang, S.; Hu, M.; et al. Air pollutant emissions from Chinese households: A major and underappreciated ambient pollution source. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7756–7761. [CrossRef]
45. Zhu, B.; Cao, L.-M.; Xia, S.-Y.; Niu, Y.-B.; Man, H.-Y.; Du, K.; Yu, K.; Huang, X.-F. Identifying the airport as a key urban VOC source in the Pearl River Delta, China. *Atmos. Environ.* **2023**, *301*, 119721. [CrossRef]
46. Sha, Q.; Zhu, M.; Huang, H.; Wang, Y.; Huang, Z.; Zhang, X.; Tang, M.; Lu, M.; Chen, C.; Shi, B.; et al. A newly integrated dataset of volatile organic compounds (VOCs) source profiles and implications for the future development of VOCs profiles in China. *Sci. Total Environ.* **2021**, *793*, 148348. [CrossRef] [PubMed]
47. Gentner, D.R.; Worton, D.R.; Isaacman, G.; Davis, L.C.; Dallmann, T.R.; Wood, E.C.; Herndon, S.C.; Goldstein, A.H.; Harley, R.A. Chemical Composition of Gas-Phase Organic Carbon Emissions from Motor Vehicles and Implications for Ozone Production. *Environ. Sci. Technol.* **2013**, *47*, 11837–11848. [CrossRef]

48. Shi, X.; Zheng, G.; Shao, Z.; Gao, D. Effect of source-classified and mixed collection from residential household waste bins on the emission characteristics of volatile organic compounds. *Sci. Total Environ.* **2020**, *707*, 135478. [CrossRef] [PubMed]
49. Zheng, J.; Yu, Y.; Mo, Z.; Zhang, Z.; Wang, X.; Yin, S.; Peng, K.; Yang, Y.; Feng, X.; Cai, H. Industrial sector-based volatile organic compound (VOC) source profiles measured in manufacturing facilities in the Pearl River Delta, China. *Sci. Total Environ.* **2013**, *456*, 127–136. [CrossRef]
50. Tsai, W.Y.; Chan, L.Y.; Blake, D.R.; Chu, K.W. Vehicular fuel composition and atmospheric emissions in South China: Hong Kong, Macau, Guangzhou, and Zhuhai. *Atmos. Chem. Phys.* **2006**, *6*, 3281–3288. [CrossRef]
51. Zhao, Q.; Bi, J.; Liu, Q.; Ling, Z.; Shen, G.; Chen, F.; Qiao, Y.; Li, C.; Ma, Z. Sources of volatile organic compounds and policy implications for regional ozone pollution control in an urban location of Nanjing, East China. *Atmos. Chem. Phys.* **2020**, *20*, 3905–3919. [CrossRef]
52. Brewer, J.F.; Millet, D.B.; Wells, K.C.; Payne, V.H.; Kulawik, S.; Vigouroux, C.; Cady-Pereira, K.E.; Pernak, R.; Zhou, M. Space-based observations of tropospheric ethane map emissions from fossil fuel extraction. *Nat. Commun.* **2024**, *15*, 7829. [CrossRef]
53. Kanai, M. Photocatalytic upgrading of natural gas. *Science* **2018**, *361*, 647–648. [CrossRef] [PubMed]
54. Liu, Y.; Song, M.; Liu, X.; Zhang, Y.; Hui, L.; Kong, L.; Zhang, Y.; Zhang, C.; Qu, Y.; An, J.; et al. Characterization and sources of volatile organic compounds (VOCs) and their related changes during ozone pollution days in 2016 in Beijing, China. *Environ. Pollut.* **2020**, *257*, 113599. [CrossRef]
55. Cui, Y.; Liu, B.; Yang, Y.; Kang, S.; Wang, F.; Xu, M.; Wang, W.; Feng, Y.; Hopke, P.K. Primary and oxidative source analyses of consumed VOCs in the atmosphere. *J. Hazard. Mater.* **2024**, *476*, 134894. [CrossRef] [PubMed]
56. Chan, L.Y.; Chu, K.W.; Zou, S.C.; Chan, C.Y.; Wang, X.M.; Barletta, B.; Blake, D.R.; Guo, H.; Tsai, W.Y. Characteristics of nonmethane hydrocarbons (NMHCs) in industrial, industrial-urban, and industrial-suburban atmospheres of the Pearl River Delta (PRD) region of south China. *J. Geophys. Res. Atmos.* **2006**, *111*, D11304. [CrossRef]
57. Zhang, J.; Liu, Z.; Wu, Y.; Zhu, Y.; Cao, T.; Ling, D.; Wang, H.; Wang, S. The impacts of photochemical loss on the source apportionment of ambient volatile organic compounds: A case study in Northern China. *Atmos. Environ.* **2024**, *333*, 120671. [CrossRef]
58. Hu, F.; Xie, P.; Tian, X.; Xu, J.; Li, A.; Lupaşcu, A.; Butler, T.; Hu, Z.; Lv, Y.; Zhang, Z.; et al. Integrated analysis of the transport process and source attribution of an extreme ozone pollution event in Hefei at different vertical heights: A case of study. *Sci. Total Environ.* **2024**, *906*, 167237. [CrossRef] [PubMed]
59. Lyu, X.; Li, H.; Lee, S.-C.; Xiong, E.; Guo, H.; Wang, T.; de Gouw, J. Significant Biogenic Source of Oxygenated Volatile Organic Compounds and the Impacts on Photochemistry at a Regional Background Site in South China. *Environ. Sci. Technol.* **2024**, *58*, 20081–20090. [CrossRef]
60. Kurtenbach, R.; Becker, K.H.; Gomes, J.A.G.; Kleffmann, J.; LÃ¶rzer, J.C.; Spittler, M.; Wiesen, P.; Ackermann, R.; Geyer, A.; Platt, U. Investigations of emissions and heterogeneous formation of HONO in a road traffic tunnel. *Atmos. Environ.* **2001**, *35*, 3385–3394. [CrossRef]
61. Xia, S.-Y.; Wang, C.; Zhu, B.; Chen, X.; Feng, N.; Yu, G.-H.; Huang, X.-F. Long-term observations of oxygenated volatile organic compounds (OVOCs) in an urban atmosphere in southern China, 2014–2019. *Environ. Pollut.* **2021**, *270*, 116301. [CrossRef]
62. Huang, X.-F.; Zhang, B.; Xia, S.-Y.; Han, Y.; Wang, C.; Yu, G.-H.; Feng, N. Sources of oxygenated volatile organic compounds (OVOCs) in urban atmospheres in North and South China. *Environ. Pollut.* **2020**, *261*, 114152. [CrossRef]
63. Atkinson, R.; Arey, J. Atmospheric Degradation of Volatile Organic Compounds. *Chem. Rev.* **2003**, *103*, 4605–4638. [CrossRef]
64. 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. Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II – gas phase reactions of organic species. *Atmos. Chem. Phys.* **2006**, *6*, 3625–4055. [CrossRef]
65. Parrish, D.D.; Stohl, A.; Forster, C.; Atlas, E.L.; Blake, D.R.; Goldan, P.D.; Kuster, W.C.; de Gouw, J.A. Effects of mixing on evolution of hydrocarbon ratios in the troposphere. *J. Geophys. Res. Atmos.* **2007**, *112*. [CrossRef]
66. Wei, D.; Cao, C.; Karambelas, A.; Mak, J.; Reinmann, A.; Commane, R. High-Resolution Modeling of Summertime Biogenic Isoprene Emissions in New York City. *Environ. Sci. Technol.* **2024**, *58*, 13783–13794. [CrossRef] [PubMed]
67. U.S. Environmental Protection Agency (EPA). *EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide*; EPA/600/R-14/108; Office of Research and Development, U.S. Environmental Protection Agency: Washington, DC, USA, 2014.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.