Identification of volatile organic compounds and their sources driving ozone and secondary organic aerosol formation in NE Spain

Marten in ’t Veld¹,², Roger Seco¹, Cristina Reche¹, Noemi Pérez¹, Andres Alastuey¹, Miguel Estrada-Portillo³, Ivan A. Janssens³, Josep Peñuelas⁴,⁵, Marcos Fernandez-Martinez³,⁴,⁵, Nicolas Marchand⁶, Brice Temime-Roussel⁶, Xavier Querol¹, Ana Maria Yáñez-Serrano¹,⁴,⁵

¹Institute of Environmental Assessment and Water Research, IDAEA-CSIC, 08034 Barcelona, Spain
²Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
³PLECO (Plants and Ecosystems), Department of Biology, University of Antwerp, Wilrijk, Belgium
⁴CREAF, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain
⁵CSIC, Global Ecology Unit, CREAF-CSIC-UAB, E08193 Bellaterra (Cerdanyola del Vallès), Catalonia, Spain
⁶Aix Marseille Univ., CNRS, Marseille, France
Abstract

Volatile organic compounds (VOCs) play a crucial role in the formation of ozone ($O_3$) and secondary organic aerosol (SOA). In NE Spain, elevated $O_3$ levels have been observed and attributed to factors such as low summer precipitation, high insolation, elevated levels of NO$_2$ and anthropogenic VOC pollution, substantial biogenic VOC emissions, and the vertical recirculation of air masses influenced by the combination of orographic features and meteorological patterns. We conducted measurements of VOC ambient mixing ratios during both summer and winter at two stations: a Barcelona urban background station (BCN) and the Montseny rural background station (MSY). Subsequently, we employed positive matrix factorization (PMF) to analyze the VOC mixing ratios and identify their sources. Our analysis revealed five common sources: anthropogenic I (traffic & industries); anthropogenic II (traffic & biomass burning); isoprene oxidation; monoterpenes; long-lifetime VOCs. To assess the impact of these VOCs on the formation of secondary pollutants, we calculated the ozone formation potential (OFP) and secondary organic aerosol formation potential (SOAP) associated with each VOC. In conclusion, our study provides insights into the sources of VOCs and their contributions to the formation of ozone and SOA in NE Spain. The OFP was primarily influenced by anthropogenic aromatic compounds from the traffic & industries source at BCN (38–49%) and during winter at MSY (34%). In contrast, the summer OFP at MSY was primarily driven by biogenic contributions from monoterpenes and isoprene oxidation products (45%). Acetaldehyde (10–35%) and methanol (13–14%) also made significant OFP contributions at both stations. Regarding SOA formation, anthropogenic aromatic compounds originating from traffic, industries, and biomass burning played a dominant role (88–93%). However, during summertime at MSY, monoterpenes became the primary driver for SOA formation (41%). These findings emphasize the importance of considering both anthropogenic and biogenic VOCs in air quality management strategies.
1.0 Introduction

Volatile organic compounds (VOCs) play an important role in the formation of ozone (O$_3$) and secondary organic aerosol (SOA). O$_3$ is formed in the troposphere by photochemical reactions involving the oxidation of VOCs, including methane (CH$_4$) and carbon monoxide (CO) (Jacob, 1999a, 1999b). It has been estimated that approximately 90% of global tropospheric O$_3$ is formed by the oxidation of VOCs (Jacob, 1999b, 1999a; Möller, 2004; Stevenson et al., 2006; Young et al., 2013). Meanwhile, SOAs are primarily formed via the gas-to-particle conversion mechanism from precursor gases, which include VOCs, NH$_3$, H$_2$SO$_4$, HNO$_3$, and HCl, among others (Jacob, 1999c; Seinfeld and Pandis, 2016).

After the success of the EU policy for abating air pollutants, which caused a reduction in the ambient air concentrations of several (especially primary) pollutants, interest in secondary air pollutants has increased in both the science and policy realms. Notably, the ambient concentrations of secondary pollutants, such as O$_3$ and SOA, did not follow a decreasing trend (EEA, 2019, 2018a; EMEP/CCC, 2016; in ’t Veld et al., 2021; Paoletti et al., 2014; Querol et al., 2016, 2014), which is detrimental to human health. The European Environmental Agency (EEA, 2021) reported 307,000 premature deaths in EU-28 countries attributable to fine particulate matter (PM$_{2.5}$) in 2019, 40,400 attributable to NO$_2$, and 16,800 attributable to O$_3$. Furthermore, O$_3$ also has a considerable negative effect on vegetation (Felzer et al., 2007; Krupa and Manning, 1988). It has been estimated that by 2030, there will be a global loss of €14–29 billion annually due to crop losses caused by O$_3$ (Avnery et al., 2011).

This is especially worrying for the western Mediterranean basin, which is prone to elevated O$_3$ levels due to low summer precipitation, high insolation, high NO$_2$ pollution, high biogenic VOC emissions, and the vertical recirculation of air masses caused by the combination of orographic features and meteorological patterns (Gangoiti et al., 2001; Millán, 2014; Millán et al., 2002, 1997; Pérez et al., 2004; Querol et al., 2017). Querol et al. (2016) reported that between 2000 and 2015, an O$_3$ level increase of 1.2% yr$^{-1}$ occurred at industrial and urban background sites in Spain, with an increase of
1.9% yr\(^{-1}\) at traffic sites. Rural background concentrations did not show a significant trend but were higher in comparison to the other sites (urban, traffic, and industrial). Past studies have shown a relative increase in SOA throughout the study area. For example, Via et al. (2021) compared organic aerosol measurements using an aerosol chemical speciation monitor (ACSM) between May 2014 and May 2015, as well as between Sep 2017 and Oct 2018, showing an increase in the relative SOA content in PM\(_{1}\) and a higher degree of oxidation during the latter period. The same observation was made by in ’t Veld et al. (2021), who estimated relative SOA increases of 12 and 8% in the chemical composition of PM\(_{2.5}\) in urban and rural background stations, respectively, between 2009 and 2018. Furthermore, secondary pollutants are closely linked to one another since higher O\(_3\) levels increase the readily available hydroxyl- and nitrate-oxidizing radicals (OH\(^{•}\) and NO\(_3^{•}\)) in the urban atmosphere. An example of this was observed in downtown Madrid, where a 30–40% increase in O\(_3\) was measured alongside OH\(^{•}\) and NO\(_3^{•}\) increases of up to 70 and 90%, respectively, between 2007 and 2014 (Saiz-Lopez et al., 2017). Additionally, a reduction in sulfur dioxide (SO\(_2\)) (EEA, 2015) and nitrogen oxides (NO\(_x\)) (EEA, 2018b) in Europe between 1990 and 2011 also led to a decrease in the consumption of the oxidizing radicals, which increased their availability and might have resulted in an atmosphere more prone to SOA production (Querol et al., 2018; Saiz-Lopez et al., 2017).

Increased knowledge regarding the relationships of VOC sources and their potential to form O\(_3\) and SOA is required to develop cost-effective abatement strategies in the western Mediterranean. VOCs encompass a wide variety of organic compounds, which include over 1,000 chemical components originating from a variety of sources. Although previous studies have investigated the VOCs occurring in NE Spain (Filella and Peñuelas, 2006; Seco et al., 2013, 2011; Yáñez-Serrano et al., 2021a) and contributed to a greater understanding of the VOCs in this region, source apportionment was only performed in a rural background site during summer (Yáñez-Serrano et al., 2021a), while a comparison between an urban background and rural background was only performed during winter (Seco et al., 2013). Finally, most published studies have not investigated the potential of existing VOCs to form O\(_3\) or SOA.
This study focuses on evaluating the levels and source contributions of VOCs in NE Spain by using proton-transfer-reaction mass spectrometry (PTR-MS) to measure online VOC mixing ratios. This allowed us to identify and quantify a selection of VOCs in both the Barcelona urban background site and the Montseny rural background site. The differences between the stations allowed us to highlight differences between an urban and rural environment and show which VOCs are emitted locally and which are emitted regionally. Furthermore, the extensive measurements conducted during both the summer and winter months at both stations allowed us to identify seasonal patterns. The use of both the O₃ formation potential (OFP) and the SOA formation potential (SOAP) to estimate the amount of O₃ and SOAs that VOCs can form under ideal situations, respectively, permits the identification of the most potentially important precursors to both O₃ and SOA formation in the study area. Additionally, to aid in the identification of VOC sources, a positive matrix factorization (PMF) model was applied on the VOC dataset to obtain the sources.

2.0 Methodology

2.1 Sampling location

VOC measurements were performed using PTR-MS at two stations in the conurbation of Barcelona in NE Spain. Both sites are included in the European Aerosols, Clouds, and Trace Gases Research Infrastructure Network (ACTRIS) and the Air Quality Monitoring Network (AQMN) of the Catalan Government, while MSY is part of the Global Atmosphere Watch (GAW) network of the World Meteorological Organization (WMO) (Figure 1). A mountain in the Montseny Natural Park, located in El Vilar de la Castanya (el Brull), is home to one of the two stations, which is a rural background station (hereafter MSY; 41°46'45.63"N, 02°21'28.92"E; 720 m a.s.l.). The MSY station is located 40 km from the Mediterranean coast and 50 km to the N-NE of Barcelona. It has been demonstrated to be representative of the regional ambient background and is sufficiently elevated and distant from specific urban anthropogenic emission sources (Cusack et al., 2012; in ’t Veld et al., 2023, 2021; Pandolfi et al., 2016, 2014; Pérez et al., 2008; Pey et al., 2009; Ripoll et al., 2015), although the station
may be impacted by emissions from urban and industrial areas during anticyclonic atmospheric
conditions due to wind circulation (Ealo et al., 2018; Gangoiti et al., 2001; Millán et al., 2002, 1997;
Pandolfi et al., 2016, 2013). The second station is an urban background air quality station (hereafter
BCN; 41°23′14.5″N 2°06′55.6″E; 68 m asl) located at the Institute of Environmental Assessment and
Water Research (IDAEA-CSIC) in Barcelona. It is positioned adjacent to Diagonal Avenue, one of the
city’s main roads, and is flanked by a variety of commercial activities (Amato et al., 2009; Cusack et al.,
2012; in’t Veld et al., 2023, 2021; Pandolfi et al., 2016, 2014; Pérez et al., 2008; Pey et al., 2009; Querol
et al., 2014, 2004a, 2004b; Ripoll et al., 2015). At both measurement sites, an aethalometer (Magee
Scientific, USA, model AE33) was used to quantify the aerosol light absorption coefficient at seven
different wavelengths (370, 470, 520, 590, 660, 880, and 950 nm) and the mass concentration of black
carbon (BC) (Drinovec et al., 2015). Additionally, the AQMN used instrumentation to measure hourly
concentrations of O$_3$, CO, NO, NO$_2$, and NO$_x$ (Table S6). Finally, ambient temperature and solar
radiation were measured at the station in MSY, and at the Department of Physics building of the
Barcelona University for the BCN station. Wind speed and direction data were also usually included;
however, this was unavailable at MSY during the 2022 measurement period and is thus not discussed
in this paper.

2.2 PTR-MS sampling

The VOC mixing ratios in ambient air were measured using PTR-MS. Different instruments were used
depending on the location and year. In 2022, a time-of-flight mass spectrometer (PTR-ToF-MS, Ionicon
PTR TOF 4000X2, Ionicon Analytik, Innsbruck, Austria) was used in Barcelona, while a PTR-MS with a
quadrupole mass spectrometer (PTR-MS, Ionicon PTR-MS Quad, Ionicon Analytik, Innsbruck, Austria)
was used at MSY. The measurements at MSY in 2017 used a PTR-ToF-MS (Ionicon PTR TOF 8000,
Ionicon Analytik, Innsbruck, Austria). A detailed description of the instrument is provided by Graus et
al. (2010). In summary, proton transfer is a form of soft chemical ionization from hydronium ions
(H$_3$O$^+$), with the distinct feature that H$_3$O$^+$ ionizes compounds with little energy excess, resulting in
minor fragmentation, thus making it an excellent method to measure ambient air (Dunne et al., 2018;
The parameters of each PTR-MS are described in Table 1. For the 2022 measurements, a periodic calibration was made during the measurements, with automatic blanks taken every hour for MSY and every 6 hours for BCN. For the measurement in 2017, a calibration was performed at the start of the measurement. Since a blank measurement was missing for the measurement period, an approximation of the blank was made using blanks from the previous campaign using the same apparatus. Table 1 presents the measurement periods at both stations, the apparatus used, and the measurement frequency. The periods were selected as approximately 1 month in duration when data were available at both BCN and MSY.

Two gravimetrically prepared multicomponent standards created by Apel-Riemer (Apel-Riemer Environmental Inc., Miami, USA) and Restek (Restek Corporation, Bellefonte, PA, USA) contained all VOC compounds discussed here, which were humidity-dependent calibrated at various dilution steps. These calibrations used bubbled zero air to dilute the standards and were regulated as closely as possible to the ambient humidity. A list of the compounds is presented in Table 2. Notably, since PTR-MS measures the exact masses of compounds, it is not compound-specific since compounds with the same masses may interfere with the measured signal. The compounds were assigned to the masses based on their exact mass, the compounds included in the calibration canister, the GLOVOCs database (Yáñez-Serrano et al., 2021b), and previous measurements in the study area (Peñuelas et al., 2009; Seco et al., 2013, 2011; Yáñez-Serrano et al., 2021a). Nevertheless, some other compounds might contribute to these masses during ambient measurement. While most masses had a single compound assigned to them, a few masses had multiple compounds assigned to them. A mass-to-charge ratio (m/z) of 59 is the same for both acetone and propanal. However, the ambient concentrations of propanal are significantly lower when compared to acetone (Hellén et al., 2004; Mingxiao Li et al., 2018). Therefore, in this study, most instances of m/z 59 will be assumed to be acetone. A similar situation occurred with m/z 69, which was primarily attributed to isoprene but has possible interference from furan and fragments of 2-methyl-3-butene-2-ol (MBO); however, these
concentrations are much lower when compared to isoprene. Therefore, m/z 69 will be assumed to be isoprene in this study (Juráň et al., 2017; Karl et al., 2012; Kaser et al., 2013). Meanwhile, m/z 71 can be both methyl vinyl ketone (MVK) and methacrolein (MACR) since both are isomers of C4H6O. Therefore, this mass was considered to be a combination of these two VOCs. Moreover, m/z 107 can be either ethylbenzene, o-, p-, or m-xylene, and is thus referred to as C8 aromatics since they are all isomers of C8H10. Another common VOC with m/z 107 is benzaldehyde (C7H6O). While benzaldehyde has a different exact mass and can be distinguished using PTR-ToF-MS, it cannot be ruled out that it interferes with the C8 aromatics peak when using quadrupole-based PTR-MS instruments. Finally, m/z 137 contains various monoterpenes that are isomers of C10H16 and are thus referred to as monoterpenes. The uncertainty of the measurements was calculated according to the error propagation approach (Doerffel, 1984). This method calculates the total uncertainty of the PTR-MS while considering the calibration uncertainty (which includes multicomponent gas standard and mass flow controller errors), the PTR-MS instrument, and the background error. Table 2 also reports the median uncertainty in % based on the uncertainty from the calibration gas, blanks, and measurements combined.

2.3 PMF source apportionment

The source apportionment analysis was performed on the VOC dataset via PMF using the U.S. Environmental Protection Agency’s PMF v5.0 software (Norris et al., 2014). PMF is a multivariate factor analysis tool that decomposes time trends in chemical composition into factor contributions and chemical profiles by performing chemical mass balance between measured species concentrations and the sum of source contributions for those species:

\[ X_{ij} = \sum_{k=1}^{p} g_{ik} \cdot f_{kj} + e_{ij} \]
Here, $X$ is the data matrix with $i$ number of samples and $j$ number of chemical species. $p$ is the number of sources, $f$ is the chemical profile of each source with mass contribution $g$, and $e_{ij}$ is the residual for each sample. To obtain better factor profiles, certain VOCs were set as “weak” species, which means that the uncertainty is tripled. These species were selected based on the signal-to-noise ratio ($0.5 < S/N < 1.5$) and visual anomalies in the factor profiles. Strong species included acetone, isoprene, benzene, toluene, and monoterpenes. Weak species included methanol, acetonitrile, acetaldehyde, MVK+MACR, methyl ethyl ketone (MEK), and C$_8$ aromatics.

In this study, a multisite solution was used by aggregating the BCN and MSY data into a single dataset. The main benefits of running a multisite PMF were the inclusion of a larger dataset when compared to separate single-site PMF models. A multisite PMF produces more robust results and can obtain source profiles common to both sites, thus allowing a direct comparison between the two stations. This has not been previously done for VOC measurements in the study area but has been successfully applied to the chemical speciation of particulate matter (in ’t Veld et al., 2023, 2021; Pandolfi et al., 2016). However, we also acknowledge the limitations of this analysis stemming from the physical distance between the two stations, possible differences in local source types and their magnitude, and different chemical profiles for the same types of sources (Escrig et al., 2009). The results were bootstrapped 100 times, and a displacement model was used to determine the uncertainties in the PMF results.

2.4 Ozone formation potential

Ozone formation potential (OFP) has frequently been used to assess the maximal possible contribution to photochemical ozone formation from particular VOCs under ideal reaction circumstances (Carter, 2010, 2009, 1994; Venecek et al., 2018). This is known as the maximum incremental reactivity (MIR). The MIR values in this study were obtained from Venecek et al. (2018), which was based on the California statewide air pollution research center (SAPRC) box model that uses a representative urban atmosphere (Carter, 1994; Venecek et al., 2018).
Here, the OFP of VOC \( i \) is calculated by multiplying the ambient concentration \( C_i \) with the MIR of VOC \( i \).

### 2.5 Secondary organic aerosol potential

The SOAP was calculated in a similar manner, where the SOAP of VOC \( i \) was calculated by multiplying the ambient concentration \( C_i \) with the SOA yield \( Y_{SOA,i} \).

\[
SOA_i = C_i \ast Y_{SOA,i}
\]

The SOA yields were obtained from Gu et al. (2021), who experimentally determined the SOA yields from smog chamber experiments for a broad inventory of VOCs based on the two-product model (Gu et al., 2021; Odum et al., 1996; Wu and Xie, 2018). In the case of no VOC data being available, they obtained the SOA yields from Wu et al. (2017) or used the SOA yield of a VOC with a similar structure.

The SOAP was converted to \( \mu g m^{-3} \) in order to obtain the SI units often used for particle concentrations.

\[
SOA_{\mu g m^{-3}} = P \ast MW \ast SOAP_{ppb} \div 62.4 \ast (273.2 + T)
\]

Here, \( P \) is the atmospheric pressure in mmHg, which was set at 706 mmHg (equivalent to 1 atm). \( MW \) is the molecular weight of the VOC, while \( SOAP_{ppb} \) is the SOAP calculated in ppb. This was divided by the ideal gas constant of 62.4 torr/mol, multiplied by the ambient temperature in K.

### 2.6 Correlations

Statistical analysis was performed using the SciPy package for Python (Virtanen et al., 2020). Using this package, Pearson’s correlation coefficient \( (r) \) was calculated between the VOCs, gases, BC, PMF sources, and meteorological data (Figure S1-4). The statistical significance of each correlation was also calculated and denoted accordingly when \( p < 0.05 \).
3.0 Results and Discussion

3.1 Regulated gaseous pollutants

The gas data from the two stations showed that anthropogenic gaseous pollutants, such as NO$_2$ and CO, were much higher at BCN when compared to MSY (Figure 2). At BCN, NO$_2$ levels had two distinct peaks. One peak was in the morning, and another was during the evening rush hour, which was also observed for CO. Furthermore, the concentrations during weekends were much lower when compared to the weekly average due to the lower traffic levels during weekends (in ’t Veld et al., 2021; Massagué et al., 2019). At MSY, peaks during working days were missing. Nonetheless, NO$_2$ showed an increase in the evening that is considerably lower during the weekend, which suggests the penitential entrance of pollutants from nearby areas (in ’t Veld et al., 2023, 2021; Pandolfi et al., 2016; Querol et al., 2013; Seco et al., 2011; Yáñez-Serrano et al., 2021a). The CO concentrations did not show a diel cycle due to the low accuracy of our measurements in combination with the low variability in CO concentrations; however, they did show higher levels during winter when compared to summer. This seasonal trend was also observed at BCN due to the increase in anthropogenic emissions during winter, such as the increased use of residential heating (Reche et al., 2012; Viana et al., 2013) and a higher boundary layer height (Pérez et al., 2008).

The O$_3$ diel cycle showed a clear photochemical pattern at both stations with increasing levels during the day, which peaks around midday, followed by a decrease (Figure 3). Tropospheric O$_3$ is formed in the atmosphere due to the photolysis of NO$_2$ to NO and O$. In turn, the O$ can react with O$_2$ to form O$_3$ (Jacob, 1999b, 1999a; Möller, 2004; Stevenson et al., 2006; Young et al., 2013). Since this process requires sunlight, it only occurs during daylight, showing an increase during daytime and a decrease after the midday peak (Figure 3). Furthermore, NO$_2$ concentrations also increase during the day at BCN due to the oxidation of NO by O$_3$, as well as anthropogenic emissions (Gligorovski et al., 2015; Jacob, 1999c, 1999b; Kim et al., 2013; Levy, 1971; Mengze Li et al., 2018; McElroy, 2002). As expected, O$_3$ concentrations were higher during the summer due to higher temperatures and stronger solar
radiation (Monks et al., 2015), but also to the specific atmospheric dynamics of the western Mediterranean during this season, resulting in the vertical recirculation of air masses along the coast and throughout the basins (Millán et al., 2002, 1997). \(O_3\) levels were also generally higher at MSY (76.7 ± 23.5 ppbv) when compared to BCN (58.4 ± 25.7 \(\mu\)gm\(^{-3}\)). MSY typically has higher levels of \(O_3\) than the urban site of BCN due to frequent urban pollution plume transport from BCN to MSY and the less effective NO titration of \(O_3\) at MSY when compared to BCN (Massagué et al., 2019; Querol et al., 2016).

### 3.2 VOC mixing ratios

Table 3 presents the 11 compounds, their masses, and their average (taken over the whole campaign, Section 1 of the SI) mixing ratios separated by season and station. In all cases (both at BCN and MSY during both seasons), methanol was the most abundant VOC, which mostly originated from biogenic emissions and secondary formation (Holzinger et al., 2005, 1999; Lewis et al., 2005; Seco et al., 2007; Sjostedt et al., 2012). Acetone was the second-most abundant VOC and, like methanol, mostly originated from biogenic emissions and secondary formation (Arnold et al., 2004; Holzinger et al., 2005, 1999; Lewis et al., 2005; Seco et al., 2007; Singh et al., 1994; Sjostedt et al., 2012). The high levels of these short-chain oxygenated VOCs can be attributed to their low reactivity with atmospheric oxidants, especially when compared to the other set of VOCs (Table 6) (Arnold et al., 2004; Filella and Peñuelas, 2006; Lewis et al., 2005; Seco et al., 2013), which will be discussed in detail in subsequent sections.

The average mixing ratios of this study were compared to previous studies conducted in the study area between 2009 and 2019 (Section 2 of the SI). The comparison between these studies showed that the mixing ratios at MSY were stable between 2009 and 2019, indicating consistency in the regional air quality regarding VOCs. At BCN, differences were observed, where the different measurement periods and protocols used could have contributed to this. Additionally, since BCN was located closer to the sources in comparison to MSY, the measurements might be more prone to the
variations of local sources from one campaign to another, especially regarding the mean values of short-period measurement campaigns.

In summer, all VOCs at MSY show a small peak in the evening (Figure 3). This is a currently unknown anomaly of the MSY station, which was also observed in the temperature and radiation data. This phenomenon was also observed by Yáñez-Serrano et al. (2021a), who sampled VOCs a few kilometers from the measurement station, which indicates that this is not an anomaly of the measurement station.

3.3 Source apportionment

To determine the sources of all VOCs, we performed a multisite PMF source apportionment. The model was applied to an aggregated dataset from both the summer and winter periods of both stations. This made it possible to compare source profiles between the different seasons and stations. The source apportionment analysis identified five factors (Table 4), with the profiles presented in Figure S5. For ease of interpretation, all factor names are written in italics in this paper. The optimal number of sources was selected by inspecting the Q values, residuals, G space plots, and physical meanings of the factors. To confirm the optimal factor profiles of the PMF model, the data were also bootstrapped. Each dataset was bootstrapped 100 times with a minimum correlation R-value of 0.6.

Finally, the model error was estimated using the base model displacement method.

- **Factor 1: Anthropogenic I: traffic & industry factor**
  - This source originated from traffic and industry emissions at BCN and was mostly long-distance transport at MSY.

- **Factor 2: Anthropogenic II: traffic & biomass burning factor**
  - Additional anthropogenic source that accounted for a second traffic source and biomass burning.

- **Factor 3: Isoprene oxidation factor**
Contains isoprene and its oxidation products, which are both anthropogenic and biogenic at BCN, and mostly biogenic at MSY.

- **Factor 4: Monoterpene factor**
  - A biogenic source at both stations that contained only monoterpenes.

- **Factor 5: Long-lifetime VOC factor**
  - A source combining the VOCs with long atmospheric lifetimes. These were mostly anthropogenic at BCN, while biogenic and formed due to photo-oxidation at MSY. This source also contained VOCs that could not be attributed to the other four factors.

To confirm the findings of the multisite PMF, separate PMF analyses were performed for both the BCN and MSY stations, combining both seasons. All sources were identified in separate PMFs with the following notes. At MSY, the biogenic sources were identified as a single source combining isoprene, MVK+MACR, and monoterpenes. At BCN, the biogenic sources were separated into two different sources, as observed in the multisite solution. As will be described later in this section, this difference is due to the following factors. In the BCN area, isoprene can be emitted from biogenic and anthropogenic sources; however, such an anthropogenic source was negligible at MSY. Furthermore, the long-lifetime VOC factor was separated into two distinct groups at MSY, with one being characterized by photo-oxidized VOCs, which was lacking in the BCN PMF—and thus in the multisite PMF. All other sources were identified at both stations with the same tracers.
3.3.1 Factor 1: Anthropogenic I: Traffic & Industry

Factor 1 was mostly traced by toluene (81% of the total toluene) and C₈ aromatics (67%), with tracers of benzene (25%) and isoprene (19%, discussed in more detail in Factor 3). These VOCs are mostly emitted anthropogenic VOCs. The high content of aromatic hydrocarbons and oxygenated compounds without biogenic VOCs further indicates the anthropogenic origin of this factor.

The diel cycle of BCN (Figure 4) showed two clear traffic peaks during both seasons (8:00 and 20:00), coinciding with those of CO and NO₂ (Figure 2). Previous studies at BCN determined that anthropogenic VOC emissions are primarily caused by road traffic (Filella and Peñuelas, 2006; in ’t Veld et al., 2021; Jiménez et al., 2005). The lower concentrations during weekends further confirm this since anthropogenic emissions from traffic and industry are significantly lower during weekends when compared to weekdays. This phenomenon was also observed for CO and NO₂ (Filella and Peñuelas, 2006; in ’t Veld et al., 2021). Moreover, these phenomena were also observed for the major constituents of this factor, toluene and C₈ aromatics (Error! Reference source not found. Figure 3), indicating that they both drive this factor.

These observations point to the anthropogenic origin of this factor, which is further confirmed by the toluene to benzene (T/B) ratio. Generally, high T/B ratios are observed in industrial areas (Seco et al., 2013; Tan et al., 2021), while lower ratios would be observed as fresh vehicle emissions increase (Heeb et al., 2000; Khoder, 2007; Langford et al., 2009; Pinthong et al., 2022). If the T/B ratio fell below 1, the main source of emissions would come from biomass burning (Liu et al., 2008; Tan et al., 2021). In both stations, the T/B ratio was lower in the winter when compared to the summer (Figure 5). At BCN, the T/B ratio was 9.73 ± 5.16 during summer, which indicates that this factor is dominated by industrial emissions mixed with traffic emissions since industrial emissions are characterized by low levels of benzene and high levels of toluene (Filella and Peñuelas, 2006; Seco et al., 2013). During winter, the T/B ratio was lower with 2.92 ± 2.28, which is in line with previous studies conducted in the study area (Filella and Peñuelas, 2006; Seco et al., 2013). The lower values during winter were also due to an
increase in benzene emissions from residential heating, which will be discussed further in Factor 2. During both seasons, a decrease in the T/B ratio could be observed due to the substantial decrease in traffic emissions during the weekend (Jiménez et al., 2005). The decrease between summer and winter was also observed at MSY, where the T/B ratio was 3.58 ± 3.50 during summer compared to 2.44 ± 2.91 in winter due to the same phenomena. The lower values at MSY compared to BCN were a result of photochemical aging since toluene is scavenged more rapidly than benzene (Table 6) (Gelencsér et al., 1997; Seco et al., 2013) since MSY is located further from the sources of traffic and industry, which will most likely involve transported air masses and result in a lower T/B ratio.

Notably, anthropogenic emissions are generally higher during winter than summer at BCN, as observed for NO\textsubscript{2} and CO (Figure 2) (in ‘t Veld et al., 2021; Pandolfi et al., 2016; Pérez et al., 2016; Querol et al., 2014, 2001; Viana et al., 2013) and previous studies of benzene and toluene at BCN (Filella and Peñuelas, 2006). Contrary to this, in our study, the mixing ratios of toluene and the C\textsubscript{8} aromatics in the atmosphere were higher during summer at BCN, while the winter levels were higher at MSY (Table 3). Currently, it remains unknown why this increase occurred during summer.

Factor 1 showed a high correlation with other anthropogenic contaminants in the study area, such as BC concentrations (summer: r = 0.73; winter: r = 0.83), NO\textsubscript{2} concentrations (summer: r = 0.46; winter: r = 0.66), and other anthropogenic factors, namely the traffic & biomass burning factor (Factor 2) (summer: r = 0.55; winter: r = 0.89) and the isoprene oxidation factor (Factor 3) (summer: r = 0.42; winter: r = 0.72), which will be discussed in their respective sections. Overall, Factor 1 represents emissions from traffic and industry at BCN.

At MSY, the mixing ratios of Factor 1 were considerably lower when compared to BCN due to its distance from anthropogenic sources. During summer, levels were negligible and increased in winter due to greater anthropogenic emissions. During summer at MSY, the traffic & industry factor showed some correlation with solar radiation (r = 0.42), with increasing concentrations during daylight hours when compared to the evening (Figure 5), while the two peaks that coincided with rush hours at BCN
were absent. All other correlations with anthropogenic contaminants were not statistically significant. This indicates that instead of an anthropogenic origin, the factor might have a minor biogenic contribution to MSY. This should be considered an option since toluene could be emitted by Mediterranean holm oaks and pines (Heiden et al., 2009; Holzinger et al., 2000; Misztal et al., 2015).

Figure S6 presents the diel cycle of toluene and temperature for each day. Each diel cycle shows a peak during midday on most measured days, which was absent for BCN. This correlation was only present for toluene and missing for the C₈ aromatics. Although the diel cycle showed a similar pattern for both the toluene and the C₈ aromatics, the correlation between these two compounds at MSY was not as high (r = 0.52, compared to r = 0.94 at BCN for the same season). Both VOCs still had a significant correlation with other anthropogenic tracers (Figure S3-4), which would indicate an anthropogenic origin for this factor with a minor biogenic contribution from toluene. Yáñez-Serrano et al. (2021) identified a similar factor at a site located nearby the MSY station and attributed it to traffic emissions, which was the lowest contributing factor during summer, as we observed in our study. Tan et al. (2021) also found an anthropogenic source in Hong Kong with the same tracers, attributing it to industrial and traffic emissions.

3.3.2 Factor 2: Anthropogenic II: Traffic & Biomass Burning

Factor 2 was traced by benzene (69%), acetaldehyde (67%), and acetonitrile (36%), as well as various oxidized VOCs (OVOCs), such as MEK (27%), methanol (19%), and C₈ aromatics (19%). Both benzene and acetonitrile are typical markers of biomass burning, with studies also showing that acetaldehyde, MEK, and methanol can originate from biomass burning (Cerqueira et al., 2013; Holzinger et al., 1999; Li et al., 2014; Tan et al., 2021).

The mixing ratio of this factor was much higher during winter when compared to summer at both stations. While biomass burning can have various sources, such as forest fires and the burning of agricultural waste, the increase during winter indicated that this increase was due to residential heating, which was previously detected in the study area (Reche et al., 2012; Viana et al., 2013). The
T/B ratio also confirmed this, with lower ratios during the winter season when compared to the summer at both stations, indicating a higher influence of biomass burning (Figure 7).

Figure 7 presents a scatterplot of the benzene and toluene mixing ratios. While they are mostly grouped together, showing a high T/B ratio, two periods deviated from this (marked in red). The first was at BCN on the 14th of February 2022 between 07:00 and 10:00 and the second was at MSY on the 13th of June 2022 between 17:00 and 21:30. During these events, the T/B ratio differentiated from the group to a ratio below 1, indicating biomass burning as the dominant source (Seco et al., 2013; Tan et al., 2021). However, the peaks coincided with increases in various other VOCs and pollutants; therefore, the origin of this factor cannot be completely attributed to biomass-burning events.

At BCN, Factor 2 has a high correlation with the traffic & industry factor (Factor 1) (summer: $r = 0.80$; winter: $r = 0.87$) and BC concentrations (summer: $r = 0.82$; winter: $r = 0.79$), which indicates the anthropogenic origin of Factor 2. This is further confirmed by the two peaks in the morning and evening, as observed for the traffic & industry factor (Factor 1), which is common for anthropogenic contaminants at BCN. Notably, the levels of Factor 2 were much higher during winter when compared to the summer period due to increased anthropogenic contamination, which can also be observed with the higher concentrations of the weekly averages when compared to weekends. During summer, this difference was also much smaller (Error! Reference source not found.).

Although biomass burning had a major impact on this factor at BCN, it is not likely its sole source. Benzene is emitted from traffic emissions, which cannot be ignored in an urban area such as BCN (Fernández-Iriarte et al., 2020; Gelencsér et al., 1997; Heeb et al., 2000). Furthermore, the diel cycles of acetaldehyde and acetonitrile showed different patterns when compared to this factor (Error! Reference source not found.). Acetaldehyde can have various origins, such as fossil fuel combustion (Nogueira et al., 2015; Sinharoy et al., 2019; Xu et al., 2022), biomass burning (Cerqueira et al., 2013; Holzinger et al., 2005, 1999), hydrocarbon oxidation (Luecken et al., 2012), and even vegetation...
emissions (Seco et al., 2007). At BCN during winter, acetaldehyde showed the highest mixing ratios during midday, with an additional increase in the evening. However, there were no weekly patterns, with similar concentrations observed during weekly averages and weekend averages. The levels during summer were much lower, showing peaks in the morning and evening. The lower concentration during the day could be caused by the fact that acetaldehyde has an atmospheric lifetime of only a few hours (Possanzini et al., 2002; Seco et al., 2013) due to photochemical destruction (Atkinson and Arey, 2003), which resulted in lower mixing ratios being observed during the day in the summer season. This effect was not observed in the winter due to less sunlight and fewer daylight hours, which reduces the ability of acetaldehyde to be removed from the atmosphere when compared to summer (Filella and Peñuelas, 2006; Viskari et al., 2000).

Acetonitrile showed higher levels during the day when compared to the evening during both summer and winter at BCN, with higher levels in the winter (Figure 3). Compared to acetaldehyde, the levels during weekends were much lower, indicating an anthropogenic source. Acetonitrile is generally considered a biomass burning marker with a minor source from fossil fuel combustion, which might be the main source at BCN (Holzinger et al., 2005, 1999). Given the stark difference between weekdays and weekends, this is most likely the case.

An investigation of the three biggest tracers of Factor 2 indicates that while this factor partially included biomass burning at BCN, it also was likely a second traffic source in addition to Factor 1. Previous studies assessing sources of biomass burning at BCN stated that after the switch to natural gas, the contribution from residential heating in the particulate matter was relatively low (AIRUSE, 2016; Amato et al., 2016; Viana et al., 2013). Notably, its contribution remains significant since a study conducted during winter 2011 at BCN attributed 8% of the PM$_{2.5}$ mass to biomass burning (Reche et al., 2012), while a recent study attributed 15% of the total organic aerosols in PM$_{10}$ to biomass burning (in ‘t Veld et al. To be Published), indicating a considerable contribution to atmospheric pollution.
At MSY, the peaks associated with anthropogenic traffic emissions were absent due to the distance from their source, with the mixing ratios of this factor being mostly even throughout the day. During summer, the diel cycle of Factor 2 showed a peak at approximately 18:00. (Figure 6); however, this was due to a biomass-burning event observed on the 13th of June. The correlation with BC further indicates that this peak was caused by a biomass burning event, with a correlation of r = 0.66 for the entire dataset of Factor 2 during summer, while the traffic & industry factor (Factor 1) did not correlate with either this factor or the BC levels. In winter, this factor likely incorporated pollutants from residential heating and other sources transported from BCN and surrounding industrial areas. This was also confirmed by a correlation with the traffic & industry factor (r = 0.65), as well as the appearance of a peak at 07:00 for all tracers of Factor 2.

The diel cycle of benzene at MSY showed similar patterns to the traffic & biomass factor (Factor 2); however, just like BCN, acetaldehyde and acetonitrile had different diel cycles when compared to Factor 2. Both compounds showed higher levels in summer than in winter, increasing during the day and declining in the afternoon. In the case of acetonitrile, this could include long-range transport from forest fires in the Mediterranean area (Seco et al., 2013). Notably, acetonitrile levels were significantly lower when compared to BCN, suggesting that acetonitrile at BCN was of local origin.

Acetaldehyde also has a biogenic origin, which was more prominent at MSY (Lewis et al., 2005; Luecken et al., 2012; Possanzini et al., 2002). Furthermore, it can also be formed in the atmosphere by photochemical production from akenes (including biogenic alkenes such as terpenes) (Luecken et al., 2012; Seco et al., 2007). Higher levels in the summer compared to winter could be a result of the increased physiological activity of vegetation and increased secondary production due to greater photochemical activity, as well as higher local acetaldehyde emissions from forest fires (Seco et al., 2011).

A previous study by Yáñez-Serrano et al. (2021) near MSY did not identify a biomass burning source during summer due to its low contributions from biomass burning during this season and the lack of...
confirmed wildfires during their study. Tan et al. (2021) also identified a biomass burning source in Hong Kong using the same tracers as this study.

### 3.3.3 Factor 3: Isoprene Oxidation Source

Factor 3 was traced by isoprene (81%) and MVK+MACR (65%), both of which are oxidation products of isoprene (Ling et al., 2019; Liu et al., 2013). This factor also had lower levels of OVOCs attributed to it, such as MEK (15%), acetone (10%), and monoterpenes (10%). The diel cycles of MSY showed a clear increase during daylight in summer, which was less pronounced at BCN (Figure 8). Both stations confirmed this trend with a significant correlation of Factor 3 with solar radiation (BCN: r = 0.48; MSY: r = 0.52) at both stations (Figures S1 and S3).

At MSY, the diel cycle indicated a biogenic origin for isoprene (Figure 3), which is known to be emitted by vegetation (Pacifico et al., 2011; Sanadze, 2004). This was further confirmed by the high correlation between this factor and the monoterpenes factor (Factor 4) (r = 0.91) during summer, which was also emitted by biogenic activity (see description for Factor 4). This confirms the findings of Yáñez-Serrano et al. (2021a), who also measured high levels of isoprene and its oxidation products in Montseny Natural Park (a few kilometers from our station) during summertime and found a diel cycle similar to the solar radiation cycle, which was also observed in this study (r = 0.52). This could be surprising since the forest at MSY is dominated by holm oak, which typically emits isoprene at a lower rate (only 5% compared to monoterpenes) (Peñuelas et al., 2009). Furthermore, several Quercus species, such as Quercus ilex (a dominant species in the MSY region), also emit isoprene (Fernández-Martínez et al., 2018). However, a previous report for the same MSY site by Seco et al. (2011) found summertime isoprene mixing ratios amounting to approximately two-thirds of those of monoterpenes and with very similar diel patterns, which is in agreement with the present study and suggests that other vegetation present in those forests and the surrounding areas are isoprene emitters. While the exact origin of isoprene cannot be confirmed yet, it was speculated these come from Erica arborea, a shrub.
located near MSY and known to emit a considerable amount of isoprene (Pagès et al., 2020; Yáñez-Serrano et al., 2021a). Yáñez-Serrano et al. (2021a) also hypothesized that isoprene could originate from local vehicle exhaust emissions (Borbon et al., 2001), but no correlation was found between isoprene and any anthropogenic contaminant (Figure S3). Meanwhile, concentrations during winter were considerably lower due to the smaller biogenic emissions as a result of reduced solar radiation and temperature. Moreover, the diel pattern of the isoprene to MVK+MACR ratio (Error! Reference source not found.) showed that during summer at MSY, the ratios decrease during daylight hours until 18:00, with a maximum at approximately 09:00. This indicates that biogenic isoprene becomes oxidized into secondary MVK and MACR.

At BCN, a different situation occurred. Factor 3 showed an increase during daylight hours and decreased in the afternoon, with an additional small peak in the morning and evening (Error! Reference source not found.). The morning and evening peaks coincide with those from other anthropogenic emissions, which implies there was also an anthropogenic source for this factor (Error! Reference source not found.). The anthropogenic origin was further confirmed with the positive correlation with BC (r = 0.50), the traffic & biomass burning factor (Factor 2) (r = 0.56), and the traffic & industries factor (Factor 1) (r = 0.44) (Figure S1) during summer. The correlation was even higher during winter, with a positive correlation with BC (r = 0.46), traffic & biomass burning factor (Factor 2) (r = 0.72), and the traffic & industries factor (Factor 1) (r = 0.60). Isoprene is known to be emitted by anthropogenic activities, particularly traffic (Borbon et al., 2002, 2001; Filella and Peñuelas, 2006; Wagner and Kuttler, 2014; Yee et al., 2020). However, other anthropogenic VOCs can also contribute to the signal of m/z 69 since it is a common fragment of cycloalkenes (Gueneron et al., 2015).

Nevertheless, the maximum isoprene mixing ratios in summer occurred at 13:00, when temperature and solar radiation were also at their maximum. This observation suggests that a biogenic source remains present at BCN, which is logical given the forests surrounding the area.
The isoprene oxidation factor was previously identified by Yáñez-Serrano et al. (2021) in Montseny as a biogenic source along with monoterpenes. While monoterpenes are mostly of biogenic origin, isoprene is not necessarily biogenic. Combining the datasets of the two stations (i.e., BCN and MSY) separated this biogenic source between the isoprene oxidation factor (Factor 4) and the monoterpe factor (Factor 5). The isoprene oxidation factor is mostly biogenic at MSY, with high levels in the summer and a clear correlation with temperature (Figure S3). Meanwhile, at BCN, an additional anthropogenic source of isoprene exists, especially during winter, as shown by the diel cycle and correlation with other anthropogenic contaminants (Figure 9; Figure S1).

3.3.4 Factor 4: Monoterpenes

The monoterpane factor (Factor 4) was mostly traced by monoterpenes (84%), with minor contributions from methanol (10%) and C8 aromatics (10%). The monoterpane factor was emitted from biogenic sources at MSY, as seen in the diel pattern (Figure 10). Just like for the isoprene oxidation factor, levels were extremely high during the day while being low at night and in the evening. Furthermore, a clear distinction between seasons is apparent since this factor exhibited its minimum mixing ratios during winter without any diel cycle, whereas it had a clear diel cycle correlating with sunlight and temperature in summer (Figure S3-4). The high levels measured are due to the MSY station being embedded among holm oak trees, which emit these compounds in relevant quantities (Peñuelas et al., 2009; Seco et al., 2013, 2011). This factor was previously identified by Yáñez-Serrano et al. (2021) at Montseny, where it was combined with the isoprene oxidation factor into a single biogenic source.

This was not the case for BCN, where the diel cycle did not resemble one increasing with sunlight. This factor did correlate with the anthropogenic factors (traffic & biomass burning and traffic & industries) during both seasons, with a correlation factor between 0.61 and 0.72 during summer and 0.71 and 0.78 during winter (Figure S1-2). However, the peaks in the morning and evening did not coincide with
the peaks of the anthropogenic contaminants (Figures 2 and 10). Hellén et al. (2012) proposed a traffic-related origin of monoterpenes since the diel cycle of monoterpenes resembles those of anthropogenic aromatics; however, as previously mentioned, the peaks at BCN do not coincide with said anthropogenic aromatics. It is more likely that the diel cycle shape was caused by the rapid reaction of monoterpenes with daytime generated OH$^\cdot$ radicals and dilution by atmospheric mixing, as was reported previously in the study area (Filella and Peñuelas, 2006; Harrison et al., 2001). Although lower levels were observed during weekends, those values might still indicate an anthropogenic origin. The reaction rates with OH$^\cdot$, presented in Table 6, indicate that the reactivity of monoterpenes is at least 10- to 100-fold faster when compared to the other VOCs.

3.3.5 Factor 5: Long-lifetime VOCs

The final factor was a mixture of the remaining VOCs in the study area. These were traced by various OVOCs, including acetone (71%), methanol (69%), acetonitrile (61%), MEK (40%), MVK+MACR (31%), and acetaldehyde (23%). Methanol and acetone had the highest mixing ratios at both stations during both seasons, which can be attributed to their long lifetime in the atmosphere and the fact that they are formed from the oxidation of other VOCs (Seco et al., 2007) (Table 6). This resulted in high background levels, as observed in previous studies (Filella and Peñuelas, 2006; Lewis et al., 2005; Seco et al., 2013; Yáñez-Serrano et al., 2021a). Acetonitrile had much lower mixing ratios but also has a long atmospheric lifetime ranging from months to years (Sanhueza et al., 2004). Therefore, this factor was called the long-lifetime VOCs factor.

At BCN, the long-lifetime VOCs factor showed a similar diel pattern to those of methanol and acetone, which was similar to other anthropogenic contaminants during summertime (Figure 11). This resulted in positive correlations with said factors (traffic & industries: 0.49; traffic & biomass burning: 0.52), as well as the observation of two peaks—one early in the morning and one in the evening—with much lower mixing ratios during weekends. At BCN, Filella & and Peñuelas (2006, and references therein)
found that both methanol and acetone were emitted by road traffic, which resulted in peaks during rush hours. The observed decrease of factor 5 during the day would be a result of photochemical destruction and dilution effects. The anthropogenic origin of these VOCs is further confirmed by the lower mixing ratios over weekends.

During winter at BCN, the highest mixing ratios of Factor 5 occurred at midday, with a small peak in the evening (Figure 11). During summer, the disparity between the weekday and weekend averages was highly accentuated (Figure 11); however, this difference was only observed in the middle of the day during winter. This suggests that at BCN, this factor has a combination of sources—mostly anthropogenic and from atmospheric oxidation—and that this combination can change with the season. The higher levels observed during winter were most likely the result of less sunlight and the consequently diminished photochemical destruction of methanol and acetone.

During summer at MSY, the long-lifetime VOCs factor mixing ratio does not represent the trends of any VOCs. This is most likely a result of the different trends that all the VOCs have at this station, which averages to an overall neutral trend. Methanol showed three peaks (Figure 3), with one in the morning (08:00) possibly coinciding with stomatal opening (Fall and Benson, 1996; Hüve et al., 2007), and two more around midday and in the evening. This suggests both the biogenic and photo-oxidation origins of methanol at the MSY station. On the other hand, acetone only displayed a peak during the day that was highly correlated with temperature \((r = 0.75)\) and the biogenic sources (isoprene oxidation: \(r = 0.69\); monoterpenes \(r = 0.56\)). This would indicate that, like methanol, acetone at MSY can be both formed due to photo-oxidation and biogenic emissions at MSY. The only difference was that acetone did not have a stomatal release in the morning, hence the lack of a morning peak.

On the other hand, during the winter season, all three long-lifetime VOCs showed no diel cycle due to lower biogenic emissions and sunlight. This indicates that they are mostly long-range transported VOCs due to their lengthy atmospheric lifetimes.
3.4 Ozone formation potential

The contribution of VOCs to the formation of O\textsubscript{3} was estimated using the MIR method to calculate and assess the contribution of each VOC to ozone formation. Table 5 presents the MIR of each VOC. All MIRs were obtained from Venecek et al. (2018), except for acetonitrile, which has a currently unknown MIR. However, it is important to consider that the MIR values used in this comparison are based on an urban environment, which limits the accuracy of the estimated OFP for MSY (Venecek et al., 2018), although the MIR has been applied in rural environments to estimate the OFP (Gómez et al., 2020; Kumar et al., 2018; Luo et al., 2020). The results of the OFP are presented in Table 5 and Figure 12.

The OFP of the sum of the VOCs was similar during summer at both stations, with 18.51 ppbv at BCN and 19.07 ppbv at MSY. During winter, a clear difference was observed. The winter OFP was 16.50 ppbv at BCN, only slightly lower compared to the summer season, while the OFP was considerably lower at MSY during winter at 8.42 ppbv. The differences in OFP between the stations and seasons were due to low concentrations of high-MIR VOCs during winter in the rural area. There are two groups with high MIR values among our measured VOCs. These were primarily the aromatics, which are mostly anthropogenic and thus quite low at MSY, with the second-largest group being biogenic VOCs such as isoprene, its oxidation products, and monoterpenes, whose emissions considerably decrease during winter due to a decrease in sunlight. The OFP does not completely account for the O\textsubscript{3} concentrations measured simultaneously. At BCN during winter, the OFP was 36% of the total O\textsubscript{3}, while it was 13% at MSY during winter. Meanwhile, during summer, BCN exhibited 28% of the total O\textsubscript{3}, while MSY exhibited 22%. This dispersion can be due to the limited VOC set measured, as well as various other sources of O\textsubscript{3}.

Comparing the hydrocarbon VOCs (HVOCs) and OVOCs showed a similar situation during winter at both stations, where OVOCs drove the OFP slightly higher, with 7.22 ppbv (51%) at BCN and 4.89 ppbv (58%) at MSY, compared to 6.93 ppbv (49%) and 3.54 (42%) for HVOC at BCN and MSY, respectively. The situation was quite different during the summer months for the two stations. At BCN, the OFP was mostly driven by HVOCs, with 11.7 ppbv (63%) compared to 6.81 ppbv (37%) of OVOCs. At MSY,
the situation was reversed, where the OFP was primarily driven by OVOCs, with 13.06 ppbv (68%) compared to 6.01 ppbv (32%) of HVOCs. This difference in contribution is a commonly observed phenomenon, where OVOCs contribute significantly more to rural areas when compared to urban areas (Chen et al., 2022; Louie et al., 2013; Luo et al., 2011).

At BCN, the OFP primarily originates from the traffic & industry factor (Factor 1) since both toluene and C₈ aromatics, which were the main tracers of this factor, together made up 50% of the estimated OFP during summer, with 9.13 ppbv. On the other hand, due to the relatively lower MIR and mixing ratios of benzene, it contributed very little toward OFP. The total OFP from the aromatics (benzene, toluene, and C₈ aromatics) was nearly 1.5 times lower during winter (6.28 ppbv) due to the lower mixing ratios of toluene and C₈ aromatics, accounting only for 39% of the total OFP. At MSY, the contribution of the traffic & industries factor was significantly lower, with the relative contribution of toluene and C₈ aromatics to OFP being 35% (2.85 ppbv) during winter, which is nearly 2.2-fold lower when compared to BCN during the same season. During summer, the contribution of these aromatics to OFP was only 7% with 1.37 ppbv, representing one of the smallest contributing sources and 6.6 times lower when compared to summer at BCN. Although this is worth keeping in mind, since MSY is a rural area, most toluene and C₈ aromatics will be transported from industrial and urban sources and are thus already photo oxidized to O₃ and SOAs when the air masses reach the station (Seco et al., 2013).

The biogenic sources, which are the monoterpenes factor (Factor 4) and the isoprene oxidation products (Factor 3), had much higher mixing ratios during summer due to increased biogenic emissions at both stations. The combination of isoprene, MVK+MACR, and monoterpenes accounted for 20% of the OFP during summer at BCN. This significantly decreased during winter to a combined 7%, with an OFP of 1.16 ppbv compared to 3.71 ppbv during summer. Notably, most of the OFP during winter originated from isoprene and its oxidation products since the monoterpene’s contribution at BCN was low. As previously established, this also includes anthropogenic isoprene. At MSY, these sources were
mainly biogenic and contributed much more to the total OFP. During summer, the biogenic sources made up 45% of the total OFP with 8.59 ppbv, which was much lower during winter when they only comprised 11% of the total OFP with 0.91 ppbv. This made biogenic VOCs the main driver of OFP at MSY during summer.

The final significant contribution to OFP was from acetaldehyde, which, with its high MIR and mixing ratios, contributed significantly to the OFP at both stations. During winter, it contributed 34% of the total OFP at BCN with 5.60 ppbv, which was much higher than during summer, when it only contributed 10% of the total OFP with 1.89 ppbv. During winter at MSY, acetaldehyde also contributed 35% of the total OFP, with 2.94 ppbv, while during summer it only contributed 27% of the total OFP but it contributed more (5.07 ppbv) in absolute terms. Ultimately, acetaldehyde contributed relatively the most during the winter season at both stations, even though the actual OFP was higher during the summer season at MSY. Methanol followed closely with 12 to 14% of the total OFP at all stations.

3.5 SOA formation potential

Most of the existing literature has estimated the SOAP using the relative SOAP to toluene ratio obtained from Derwent et al. (2010). While useful, the list presented there does exclude some essential VOCs that we measured in our study. Therefore, we obtained the SOAP from Gu et al. (2021), who reported the SOA yield of various VOCs included in this study. It is crucial to state that this method, like the OFP, has the limitation that the SOA yield was calculated based on an urban environment with a high NOx atmosphere, which influences SOA formation (Camredon et al., 2007). While this situation applies to BCN, it is more inaccurate in the case of MSY. Additionally, it is worth mentioning that the SOA yields used OH as the primary oxidant to form SOA (Gu et al., 2021). O3 is a secondary oxidant that can also form SOA, and with O3 levels being generally higher in the MSY rural area, its capability to oxidize will increase accordingly. However, the reaction rates of VOCs with O3 are 105 to 1011 times lower (Atkinson and Arey, 2003; Atkinson and Carter, 1984) than those with OH,
thereby diminishing the impact of this difference. The results of the SOAP are presented in Table 7 and Figure 12.

Overall, aromatics contributed the most due to their high SOA yields at both stations during both seasons (Table 7, Figure 12). Since the mixing ratios of these aromatic compounds were much higher at BCN, a much higher SOAP was observed in this study. Due to the aromatic VOCs' high SOA yields, they contributed the most to the SOAP at both stations during both seasons (Table 7). Since the mixing ratios of these aromatic compounds were much higher at BCN, the SOAP was considerably higher compared to MSY. At MSY, the SOAP was similar between the two seasons but was dominated by anthropogenic aromatic compounds (benzene, toluene, C8 aromatics) during winter, while the monoterpenes were a considerable source of SOAP in summer.

This would indicate that traffic & industry (Factor 1) was the main contributor to SOA formation from the measured VOCs at BCN since none of the remaining OVOCs and isoprene oxidation products have a significant SOA yield. Although monoterpenes have a significant SOA yield, the levels at BCN were much lower, resulting in minimal influence. However, the diel cycle of the monoterpene factor (Factor 4) suggests that during summer at BCN, monoterpenes were rapidly reacting with daytime-generated OH radicals, which would result in the formation of both O3 and SOA; therefore, the actual contribution of monoterpenes would be much higher than the values given in Table 7.

At MSY during winter, the traffic & industry factor (Factor 1) seemed to be the greatest source of SOA in the area. However, as was the case with the OFP, the actual SOAP might be higher as the transported toluene and C8 aromatics deplete during transport. During summer, the monoterpenes were the main contributors to the SOAP, while other biogenic compounds, such as isoprene, did not contribute as much due to their low SOA yields.
4.0 Conclusions

In this study, we quantified 11 VOC mixing ratios during summer and winter at both the BCN urban background site and the MSY rural background site (Table 5). Applying PMF to the combined dataset revealed five factors (Table 6). The first factor, traffic & industries (Factor 1), originated from anthropogenic sources such as traffic and industry in and around Barcelona, mostly consisting of aged air masses at MSY. The second factor, traffic & biomass burning (Factor 2), represented a secondary source combining traffic emissions and biomass burning. The third factor, isoprene oxidation (Factor 3), was associated with isoprene and its oxidation products, which had both anthropogenic and biogenic origins at BCN and were mostly biogenic at MSY. The fourth factor, monoterpenes (Factor 4), was predominantly biogenic at both stations, with lower levels at BCN due to rapid oxidation by OH* and lower biogenic emissions in the area. The fifth factor (Factor 5) represented a mixture of VOCs with longer atmospheric lifetimes and included methanol, acetone, and acetonitrile, which had longer lifetimes when compared to other VOCs.

Analyzing the sources of the VOCs in relation to OFP revealed that at BCN, anthropogenic VOCs—particularly toluene and C₈ aromatics—were the primary contributors to O₃ formation, accounting for 38 and 49% of the measured OFP during winter and summer, respectively (Table 7). BCN also had significant contributions from isoprene oxidation products (6 and 19% during winter and summer, respectively) and acetaldehyde (34 and 10% during winter and summer, respectively). Methanol contributed 14 and 13% during winter and summer, respectively. At MSY, biogenic sources—and specifically isoprene oxidation products and monoterpenes—played a larger role in the total OFP, contributing 11 and 45% during winter and summer, respectively. The aged anthropogenic VOCs from the traffic & industries factor made a considerable contribution of 34% during winter and 7% during summer. Acetaldehyde contributed 35 and 27%, while methanol contributed 12 and 14%, during winter and summer, respectively.
At BCN, the results demonstrate that both $O_3$ and SOA formation were predominantly driven by anthropogenic emissions, particularly those originating from the traffic & industries factor. The aromatic compounds had a significant impact on the SOAP due to their higher SOA yields (Table 9). The seasonality showed a significant increase during summer, with concentrations of 0.766 $\mu$gm$^{-3}$ compared to 0.603 $\mu$gm$^{-3}$ during winter, primarily due to higher levels of toluene and C$_8$ aromatics during summer (Table 9). The OFP also showed seasonality with higher levels during the summer, but the difference in total OFP was small, with 16.5 ppbv during winter and 18.51 ppbv during summer (Table 7). The MSY rural background site exhibited a different pattern, where both OFP and SOAP were driven by aged anthropogenic emissions, predominantly from the traffic & industries factor, during winter. In contrast, during summer, biogenic sources—specifically the isoprene oxidation products and monoterpenes factors—played a more prominent role. In the case of SOAP, the monoterpenes factor contributed 53% of the overall SOAP during summer, while aromatic compounds contributed 29%.

Comparing the total OFP between the two stations, BCN and MSY revealed similar values during the summer season, with 18.51 ppbv at BCN and 19.07 ppbv at MSY. However, during winter, the OFP was considerably lower at MSY, with 8.42 ppbv compared to 16.50 ppbv at BCN. On the other hand, the SOAP exhibited higher concentrations at BCN when compared to MSY, with 1.163 $\mu$gm$^{-3}$ and 0.603 $\mu$gm$^{-3}$ during winter at BCN and MSY, respectively, and 1.833 $\mu$gm$^{-3}$ and 0.766 $\mu$gm$^{-3}$ during summer at BCN and MSY, respectively.

While this study provides valuable insights into the sources of VOCs and their impacts on OFP and SOAP, some knowledge gaps should be acknowledged. First, both OFP and SOAP estimations are based on high nitrogen oxide (NO$_x$) environments, which may not accurately represent the low NOx conditions at the MSY rural background station. Secondly, the biogenic contribution of monoterpenes is likely underestimated at BCN during summer due to their rapid oxidation by OH$^-$ in an urban environment. Lastly, this study focused on a selection of 11 VOCs, and although they make substantial contributions to atmospheric concentrations, the inclusion of additional VOCs—particularly alkanes.
(~5–10% contribution to OFP and SOAP) and alkenes (>30% contribution to OFP)—would provide a more comprehensive understanding (Jookjantra et al., 2022; Wang et al., 2023; Wu et al., 2017; Yuan et al., 2009; Zhan et al., 2021).

5.0 Acknowledgments

This study was supported by the European Union’s Horizon 2020 research and innovation program under grant agreement 101036245 (RI-URBANS), the “Agencia Estatal de Investigación” from the Spanish Ministry of Science and Innovation, FEDER funds under the projects CAIAC (PID2019-108990RB-I00), the Generalitat de Catalunya (AGAUR 2021 SGR 00447), and the Direcció General de Territori. AMYS acknowledges a Ramón y Cajal grant (RYC2021-032519-I) and the La Caixa Foundation Junior Leader retaining fellowship. RS acknowledges a Ramón y Cajal grant (RYC2020-029216-I) funded by MCIN/AEI/ 10.13039/501100011033 and by “ESF Investing in your future”. IDAEA-CSIC is a Severo Ochoa Centre of Research Excellence (MCIN/AEI, Project CEX2018-000794-S).