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

Marten in 't Veld[1,2], Roger Seco[1], Cristina Reche[1], Noemi Pérez[1], Andres Alastuey[1], Miguel Estrada-Portillo[3], Ivan A. Janssens[3], Josep Peñuelas[4,5], Marcos Fernandez-Martinez[3,4,5], Nicolas Marchand[6], Brice Temime-Roussel[6], Xavier Querol[1], Ana Maria Yáñez-Serrano[1,4,5]

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

1 Abstract

2   Volatile organic compounds (VOCs) play a crucial role in the formation of ozone (O_3) and secondary

3   organic aerosol (SOA). In NE Spain, elevated O_3 levels have been observed and attributed to factors

4   such as low summer precipitation, high insolation, elevated levels of NO_2 and anthropogenic VOC

5   pollution, substantial biogenic VOC emissions, and the vertical recirculation of air masses influenced

6   by the combination of orographic features and meteorological patterns. We conducted

7   measurements of VOC ambient mixing ratios during both summer and winter at two stations: a

8   Barcelona urban background station (BCN) and the Montseny rural background station (MSY).

9   Subsequently, we employed positive matrix factorization (PMF) to analyze the VOC mixing ratios and

10 identify their sources. Our analysis revealed five common sources: anthropogenic I (traffic &

11 industries); anthropogenic II (traffic & biomass burning); isoprene oxidation; monoterpenes; long-

12 lifetime VOCs. To assess the impact of these VOCs on the formation of secondary pollutants, we

13 calculated the ozone formation potential (OFP) and secondary organic aerosol formation potential

14 (SOAP) associated with each VOC. In conclusion, our study provides insights into the sources of VOCs

15 and their contributions to the formation of ozone and SOA in NE Spain. The OFP was primarily

16 influenced by anthropogenic aromatic compounds from the traffic & industries source at BCN (38–

17 49%) and during winter at MSY (34%). In contrast, the summer OFP at MSY was primarily driven by

18 biogenic contributions from monoterpenes and isoprene oxidation products (45%). Acetaldehyde (10–

19 35%) and methanol (13–14%) also made significant OFP contributions at both stations. Regarding SOA

20 formation, anthropogenic aromatic compounds originating from traffic, industries, and biomass

21 burning played a dominant role (88–93%). However, during summertime at MSY, monoterpenes

22 became the primary driver for SOA formation (41%). These findings emphasize the importance of

23 considering both anthropogenic and biogenic VOCs in air quality management strategies.

1

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_2SO_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

2

49 1.9% yr^{-1} at traffic sites. Rural background concentrations did not show a significant trend but were

50 higher in comparison to the other sites (urban, traffic, and industrial). Past studies have shown a

51 relative increase in SOA throughout the study area. For example, Via et al. (2021) compared organic

52 aerosol measurements using an aerosol chemical speciation monitor (ACSM) between May 2014 and

53 May 2015, as well as between Sep 2017 and Oct 2018, showing an increase in the relative SOA content

54 in PM_1 and a higher degree of oxidation during the latter period. The same observation was made by

55 in 't Veld et al. (2021), who estimated relative SOA increases of 12 and 8% in the chemical composition

56 of $PM_{2.5}$ in urban and rural background stations, respectively, between 2009 and 2018. Furthermore,

57 secondary pollutants are closely linked to one another since higher O_3 levels increase the readily

58 available hydroxyl- and nitrate-oxidizing radicals (OH$^{\cdot}$ and NO$_3^{\cdot}$) in the urban atmosphere. An example

59 of this was observed in downtown Madrid, where a 30–40% increase in O_3 was measured alongside

60 OH$^{\cdot}$ and NO$_3^{\cdot}$ increases of up to 70 and 90%, respectively, between 2007 and 2014 (Saiz-Lopez et al.,

61 2017). Additionally, a reduction in sulfur dioxide (SO_2) (EEA, 2015) and nitrogen oxides (NO_x) (EEA,

62 2018b) in Europe between 1990 and 2011 also led to a decrease in the consumption of the oxidizing

63 radicals, which increased their availability and might have resulted in an atmosphere more prone to

64 SOA production (Querol et al., 2018; Saiz-Lopez et al., 2017).

65 Increased knowledge regarding the relationships of VOC sources and their potential to form O_3 and

66 SOA is required to develop cost-effective abatement strategies in the western Mediterranean. VOCs

67 encompass a wide variety of organic compounds, which include over 1,000 chemical components

68 originating from a variety of sources. Although previous studies have investigated the VOCs occurring

69 in NE Spain (Filella and Peñuelas, 2006; Seco et al., 2013, 2011; Yáñez-Serrano et al., 2021a) and

70 contributed to a greater understanding of the VOCs in this region, source apportionment was only

71 performed in a rural background site during summer (Yáñez-Serrano et al., 2021a), while a comparison

72 between an urban background and rural background was only performed during winter (Seco et al.,

73 2013). Finally, most published studies have not investigated the potential of existing VOCs to form O_3

74 or SOA.

3

75   This study focuses on evaluating the levels and source contributions of VOCs in NE Spain by using

76   proton-transfer-reaction mass spectrometry (PTR-MS) to measure online VOC mixing ratios. This

77   allowed us to identify and quantify a selection of VOCs in both the Barcelona urban background site

78   and the Montseny rural background site. The differences between the stations allowed us to highlight

79   differences between an urban and rural environment and show which VOCs are emitted locally and

80   which are emitted regionally. Furthermore, the extensive measurements conducted during both the

81   summer and winter months at both stations allowed us to identify seasonal patterns. The use of both

82   the O_3 formation potential (OFP) and the SOA formation potential (SOAP) to estimate the amount of

83   O_3 and SOAs that VOCs can form under ideal situations, respectively, permits the identification of the

84   most potentially important precursors to both O_3 and SOA formation in the study area. Additionally,

85   to aid in the identification of VOC sources, a positive matrix factorization (PMF) model was applied on

86   the VOC dataset to obtain the sources.

87   2.0 Methodology

88   2.1 Sampling location

89   VOC measurements were performed using PTR-MS at two stations in the conurbation of Barcelona in

90   NE Spain. Both sites are included in the European Aerosols, Clouds, and Trace Gases Research

91   Infrastructure Network (ACTRIS) and the Air Quality Monitoring Network (AQMN) of the Catalan

92   Government, while MSY is part of the Global Atmosphere Watch (GAW) network of the World

93   Meteorological Organization (WMO) (Figure 1). A mountain in the Montseny Natural Park, located in

94   El Vilar de la Castanya (el Brull), is home to one of the two stations, which is a rural background station

95   (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

96   the Mediterranean coast and 50 km to the N-NE of Barcelona. It has been demonstrated to be

97   representative of the regional ambient background and is sufficiently elevated and distant from

98   specific urban anthropogenic emission sources (Cusack et al., 2012; in 't Veld et al., 2023, 2021;

99   Pandolfi et al., 2016, 2014; Pérez et al., 2008; Pey et al., 2009; Ripoll et al., 2015), although the station

4

100   may be impacted by emissions from urban and industrial areas during anticyclonic atmospheric

101   conditions due to wind circulation (Ealo et al., 2018; Gangoiti et al., 2001; Millán et al., 2002, 1997;

102   Pandolfi et al., 2016, 2013). The second station is an urban background air quality station (hereafter

103   BCN; 41°23'14.5"N 2°06'55.6"E; 68 m asl) located at the Institute of Environmental Assessment and

104   Water Research (IDAEA-CSIC) in Barcelona. It is positioned adjacent to Diagonal Avenue, one of the

105   city's main roads, and is flanked by a variety of commercial activities (Amato et al., 2009; Cusack et al.,

106   2012; in 't Veld et al., 2023, 2021; Pandolfi et al., 2016, 2014; Pérez et al., 2008; Pey et al., 2009; Querol

107   et al., 2014, 2004a, 2004b; Ripoll et al., 2015). At both measurement sites, an aethalometer (Magee

108   Scientific, USA, model AE33) was used to quantify the aerosol light absorption coefficient at seven

109   different wavelengths (370, 470, 520, 590, 660, 880, and 950 nm) and the mass concentration of black

110   carbon (BC) (Drinovec et al., 2015). Additionally, the AQMN used instrumentation to measure hourly

111   concentrations of O_3, CO, NO, NO_2, and NO_x (Table S6). Finally, ambient temperature and solar

112   radiation were measured at the station in MSY, and at the Department of Physics building of the

113   Barcelona University for the BCN station. Wind speed and direction data were also usually included;

114   however, this was unavailable at MSY during the 2022 measurement period and is thus not discussed

115   in this paper.

116   ## 2.2 PTR-MS sampling

117   The VOC mixing ratios in ambient air were measured using PTR-MS. Different instruments were used

118   depending on the location and year. In 2022, a time-of-flight mass spectrometer (PTR-ToF-MS, Ionicon

119   PTR TOF 4000X2, Ionicon Analytik, Innsbruck, Austria) was used in Barcelona, while a PTR-MS with a

120   quadrupole mass spectrometer (PTR-MS, Ionicon PTR-MS Quad, Ionicon Analytik, Innsbruck, Austria)

121   was used at MSY. The measurements at MSY in 2017 used a PTR-ToF-MS (Ionicon PTR TOF 8000,

122   Ionicon Analytik, Innsbruck, Austria). A detailed description of the instrument is provided by Graus et

123   al. (2010). In summary, proton transfer is a form of soft chemical ionization from hydronium ions

124   (H_3O^+), with the distinct feature that H_3O^+ ionizes compounds with little energy excess, resulting in

125   minor fragmentation, thus making it an excellent method to measure ambient air (Dunne et al., 2018;

5

126   Gouw et al., 2003; Ionicon Analytik GmbH, 2014; Warneke et al., 2011). The parameters of each PTR-

127   MS are described in Table 1. For the 2022 measurements, a periodic calibration was made during the

128   measurements, with automatic blanks taken every hour for MSY and every 6 hours for BCN. For the

129   measurement in 2017, a calibration was performed at the start of the measurement. Since a blank

130   measurement was missing for the measurement period, an approximation of the blank was made

131   using blanks from the previous campaign using the same apparatus. Table 1 presents the

132   measurement periods at both stations, the apparatus used, and the measurement frequency. The

133   periods were selected as approximately 1 month in duration when data were available at both BCN

134   and MSY.

135   Two gravimetrically prepared multicomponent standards created by Apel-Riemer (Apel-Riemer

136   Environmental Inc., Miami, USA) and Restek (Restek Corporation, Bellefonte, PA, USA) contained all

137   VOC compounds discussed here, which were humidity-dependent calibrated at various dilution steps.

138   These calibrations used bubbled zero air to dilute the standards and were regulated as closely as

139   possible to the ambient humidity. A list of the compounds is presented in Table 2. Notably, since PTR-

140   MS measures the exact masses of compounds, it is not compound-specific since compounds with the

141   same masses may interfere with the measured signal. The compounds were assigned to the masses

142   based on their exact mass, the compounds included in the calibration canister, the GLOVOCs database

143   (Yáñez-Serrano et al., 2021b), and previous measurements in the study area (Peñuelas et al., 2009;

144   Seco et al., 2013, 2011; Yáñez-Serrano et al., 2021a). Nevertheless, some other compounds might

145   contribute to these masses during ambient measurement. While most masses had a single compound

146   assigned to them, a few masses had multiple compounds assigned to them. A mass-to-charge ratio

147   (m/z) of 59 is the same for both acetone and propanal. However, the ambient concentrations of

148   propanal are significantly lower when compared to acetone (Hellén et al., 2004; Mingxiao Li et al.,

149   2018). Therefore, in this study, most instances of m/z 59 will be assumed to be acetone. A similar

150   situation occurred with m/z 69, which was primarily attributed to isoprene but has possible

151   interference from furan and fragments of 2-methyl-3-butene-2-ol (MBO); however, these

6

152   concentrations are much lower when compared to isoprene. Therefore, m/z 69 will be assumed to be

153   isoprene in this study (Juráň et al., 2017; Karl et al., 2012; Kaser et al., 2013). Meanwhile, m/z 71 can

154   be both methyl vinyl ketone (MVK) and methacrolein (MACR) since both are isomers of C_4H_6O.

155   Therefore, this mass was considered to be a combination of these two VOCs. Moreover, m/z 107 can

156   be either ethylbenzene, *o*-, *p*-, or *m*-xylene, and is thus referred to as C_8 aromatics since they are all

157   isomers of C_8H_{10}. Another common VOC with m/z 107 is benzaldehyde (C_7H_6O). While benzaldehyde

158   has a different exact mass and can be distinguished using PTR-ToF-MS, it cannot be ruled out that it

159   interferes with the C_8 aromatics peak when using quadrupole-based PTR-MS instruments. Finally, m/z

160   137 contains various monoterpenes that are isomers of $C_{10}H_{16}$ and are thus referred to as

161   monoterpenes. The uncertainty of the measurements was calculated according to the error

162   propagation approach (Doerffel, 1984). This method calculates the total uncertainty of the PTR-MS

163   while considering the calibration uncertainty (which includes multicomponent gas standard and mass

164   flow controller errors), the PTR-MS instrument, and the background error. Table 2 also reports the

165   median uncertainty in % based on the uncertainty from the calibration gas, blanks, and measurements

166   combined.

167   ## 2.3 PMF source apportionment

168   The source apportionment analysis was performed on the VOC dataset via PMF using the U.S.

169   Environmental Protection Agency's PMF v5.0 software (Norris et al., 2014). PMF is a multivariate

170   factor analysis tool that decomposes time trends in chemical composition into factor contributions

171   and chemical profiles by performing chemical mass balance between measured species

172   concentrations and the sum of source contributions for those species:

173
$$X_{ij} = \sum_{k=1}^{p} g_{ik} * f_{kj} + e_{ij}$$

7

174   Here, X is the data matrix with i number of samples and j number of chemical species. p is the number

175   of sources, f is the chemical profile of each source with mass contribution g, and e_{ij} is the residual for

176   each sample. To obtain better factor profiles, certain VOCs were set as "weak" species, which means

177   that the uncertainty is tripled. These species were selected based on the signal-to-noise ratio (0.5 >

178   S/N > 1.5) and visual anomalies in the factor profiles. Strong species included acetone, isoprene,

179   benzene, toluene, and monoterpenes. Weak species included methanol, acetonitrile, acetaldehyde,

180   MVK+MACR, methyl ethyl ketone (MEK), and C_8 aromatics.

181   In this study, a multisite solution was used by aggregating the BCN and MSY data into a single dataset.

182   The main benefits of running a multisite PMF were the inclusion of a larger dataset when compared

183   to separate single-site PMF models. A multisite PMF produces more robust results and can obtain

184   source profiles common to both sites, thus allowing a direct comparison between the two stations.

185   This has not been previously done for VOC measurements in the study area but has been successfully

186   applied to the chemical speciation of particulate matter (in 't Veld et al., 2023, 2021; Pandolfi et al.,

187   2016). However, we also acknowledge the limitations of this analysis stemming from the physical

188   distance between the two stations, possible differences in local source types and their magnitude, and

189   different chemical profiles for the same types of sources (Escrig et al., 2009). The results were

190   bootstrapped 100 times, and a displacement model was used to determine the uncertainties in the

191   PMF results.

2.4 Ozone formation potential

193   Ozone formation potential (OFP) has frequently been used to assess the maximal possible contribution

194   to photochemical ozone formation from particular VOCs under ideal reaction circumstances (Carter,

195   2010, 2009, 1994; Venecek et al., 2018). This is known as the maximum incremental reactivity (MIR).

196   The MIR values in this study were obtained from Venecek et al. (2018), which was based on the

197   California statewide air pollution research center (SAPRC) box model that uses a representative urban

198   atmosphere (Carter, 1994; Venecek et al., 2018).

8

199
$$OFP_i = C_i * MIR_i$$

200 Here, the OFP of VOC i is calculated by multiplying the ambient concentration C_i with the MIR of VOC

201 i.

2.5 Secondary organic aerosol potential

203 The SOAP was calculated in a similar manner, where the SOAP of VOC i was calculated by multiplying

204 the ambient concentration C_i with the SOA yield $Y_{SOA,i}$.

205
$$SOAP_i = C_i * Y_{SOA,i}$$

206 The SOA yields were obtained from Gu et al. (2021), who experimentally determined the SOA yields

207 from smog chamber experiments for a broad inventory of VOCs based on the two-product model (Gu

208 et al., 2021; Odum et al., 1996; Wu and Xie, 2018). In the case of no VOC data being available, they

209 obtained the SOA yields from Wu et al. (2017) or used the SOA yield of a VOC with a similar structure.

210 The SOAP was converted to μgm^{-3} in order to obtain the SI units often used for particle concentrations.

211
$$SOAP_{\mu gm^{-3}} = \frac{P * MW * SOAP_{ppb}}{62.4 * (273.2 + T)}$$

212 Here, P is the atmospheric pressure in mmHg, which was set at 706 mmHg (equivalent to 1 atm). MW

213 is the molecular weight of the VOC, while $SOAP_{ppb}$ is the SOAP calculated in ppb. This was divided by

214 the ideal gas constant of 62.4 torr/mol, multiplied by the ambient temperature in K.

215

2.6 Correlations

217 Statistical analysis was performed using the SciPy package for Python (Virtanen et al., 2020). Using this

218 package, Pearson's correlation coefficient (r) was calculated between the VOCs, gases, BC, PMF

219 sources, and meteorological data (Figure S1-4). The statistical significance of each correlation was also

220 calculated and denoted accordingly when $p < 0.05$.

9

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 $O^{\bullet}$. In turn, the $O^{\bullet}$ 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

10

246 radiation (Monks et al., 2015), but also to the specific atmospheric dynamics of the western

247 Mediterranean during this season, resulting in the vertical recirculation of air masses along the coast

248 and throughout the basins (Millán et al., 2002, 1997). O_3 levels were also generally higher at MSY (76.7

249 ± 23.5 ppbv) when compared to BCN (58.4 ± 25.7 μgm^{-3}). MSY typically has higher levels of O_3 than

250 the urban site of BCN due to frequent urban pollution plume transport from BCN to MSY and the less

251 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

253 Table 3 presents the 11 compounds, their masses, and their average (taken over the whole campaign,

254 Section 1 of the SI) mixing ratios separated by season and station. In all cases (both at BCN and MSY

255 during both seasons), methanol was the most abundant VOC, which mostly originated from biogenic

256 emissions and secondary formation (Holzinger et al., 2005, 1999; Lewis et al., 2005; Seco et al., 2007;

257 Sjostedt et al., 2012). Acetone was the second-most abundant VOC and, like methanol, mostly

258 originated from biogenic emissions and secondary formation (Arnold et al., 2004; Holzinger et al.,

259 2005, 1999; Lewis et al., 2005; Seco et al., 2007; Singh et al., 1994; Sjostedt et al., 2012). The high

260 levels of these short-chain oxygenated VOCs can be attributed to their low reactivity with atmospheric

261 oxidants, especially when compared to the other set of VOCs (Table 6) (Arnold et al., 2004; Filella and

262 Peñuelas, 2006; Lewis et al., 2005; Seco et al., 2013), which will be discussed in detail in subsequent

263 sections.

264 The average mixing ratios of this study were compared to previous studies conducted in the study

265 area between 2009 and 2019 (Section 2 of the SI). The comparison between these studies showed

266 that the mixing ratios at MSY were stable between 2009 and 2019, indicating consistency in the

267 regional air quality regarding VOCs. At BCN, differences were observed, where the different

268 measurement periods and protocols used could have contributed to this. Additionally, since BCN was

269 located closer to the sources in comparison to MSY, the measurements might be more prone to the

11

270   variations of local sources from one campaign to another, especially regarding the mean values of

271   short-period measurement campaigns.

272   In summer, all VOCs at MSY show a small peak in the evening (Figure 3). This is a currently unknown

273   anomaly of the MSY station, which was also observed in the temperature and radiation data. This

274   phenomenon was also observed by Yáñez-Serrano et al. (2021a), who sampled VOCs a few kilometers

275   from the measurement station, which indicates that this is not an anomaly of the measurement

276   station.

277   ## 3.3 Source apportionment

278   To determine the sources of all VOCs, we performed a multisite PMF source apportionment. The

279   model was applied to an aggregated dataset from both the summer and winter periods of both

280   stations. This made it possible to compare source profiles between the different seasons and stations.

281   The source apportionment analysis identified five factors (Table 4), with the profiles presented in

282   Figure S5. For ease of interpretation, all factor names are written in *italics* in this paper. The optimal

283   number of sources was selected by inspecting the Q values, residuals, G space plots, and physical

284   meanings of the factors. To confirm the optimal factor profiles of the PMF model, the data were also

285   bootstrapped. Each dataset was bootstrapped 100 times with a minimum correlation R-value of 0.6.

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

287   - *Factor 1: Anthropogenic I: traffic & industry factor*
288   - This source originated from traffic and industry emissions at BCN and was mostly long-
289   distance transport at MSY.
290   - *Factor 2: Anthropogenic II: traffic & biomass burning factor*
291   - Additional anthropogenic source that accounted for a second traffic source and
292   biomass burning.
293   - *Factor 3: Isoprene oxidation factor*

12

294   o Contains isoprene and its oxidation products, which are both anthropogenic and

295   biogenic at BCN, and mostly biogenic at MSY.

296   • *Factor 4: Monoterpene factor*

297   o A biogenic source at both stations that contained only monoterpenes.

298   • *Factor 5: Long-lifetime VOC factor*

299   o A source combining the VOCs with long atmospheric lifetimes. These were mostly

300   anthropogenic at BCN, while biogenic and formed due to photo-oxidation at MSY. This

301   source also contained VOCs that could not be attributed to the other four factors.

302 To confirm the findings of the multisite PMF, separate PMF analyses were performed for both the BCN

303 and MSY stations, combining both seasons. All sources were identified in separate PMFs with the

304 following notes. At MSY, the biogenic sources were identified as a single source combining isoprene,

305 MVK+MACR, and monoterpenes. At BCN, the biogenic sources were separated into two different

306 sources, as observed in the multisite solution. As will be described later in this section, this difference

307 is due to the following factors. In the BCN area, isoprene can be emitted from biogenic and

308 anthropogenic sources; however, such an anthropogenic source was negligible at MSY. Furthermore,

309 the *long-lifetime VOC* factor was separated into two distinct groups at MSY, with one being

310 characterized by photo-oxidized VOCs, which was lacking in the BCN PMF—and thus in the multisite

311 PMF. All other sources were identified at both stations with the same tracers.

312

313

13

3.3.1 Factor 1: Anthropogenic I: Traffic & Industry

Factor 1 was mostly traced by toluene (81% of the total toluene) and C_8 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 NO_2 (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 NO_2 (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 C_8 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

14

339   increase in benzene emissions from residential heating, which will be discussed further in *Factor 2*.

340   During both seasons, a decrease in the T/B ratio could be observed due to the substantial decrease in

341   traffic emissions during the weekend (Jiménez et al., 2005). The decrease between summer and winter

342   was also observed at MSY, where the T/B ratio was 3.58 ± 3.50 during summer compared to 2.44 ±

343   2.91 in winter due to the same phenomena. The lower values at MSY compared to BCN were a result

344   of photochemical aging since toluene is scavenged more rapidly than benzene (Table 6) (Gelencsér et

345   al., 1997; Seco et al., 2013) since MSY is located further from the sources of traffic and industry, which

346   will most likely involve transported air masses and result in a lower T/B ratio.

347   Notably, anthropogenic emissions are generally higher during winter than summer at BCN, as

348   observed for NO_2 and CO (Figure 2) (in 't Veld et al., 2021; Pandolfi et al., 2016; Pérez et al., 2016;

349   Querol et al., 2014, 2001; Viana et al., 2013) and previous studies of benzene and toluene at BCN

350   (Filella and Peñuelas, 2006). Contrary to this, in our study, the mixing ratios of toluene and the C_8

351   aromatics in the atmosphere were higher during summer at BCN, while the winter levels were higher

352   at MSY (Table 3). Currently, it remains unknown why this increase occurred during summer.

353   *Factor 1* showed a high correlation with other anthropogenic contaminants in the study area, such as

354   BC concentrations (summer: r = 0.73; winter: r = 0.83), NO_2 concentrations (summer: r = 0.46; winter:

355   r = 0.66), and other anthropogenic factors, namely the *traffic & biomass burning factor* (*Factor 2*)

356   (summer: r = 0.55; winter: r = 0.89) and the *isoprene oxidation factor* (*Factor 3*) (summer: r = 0.42;

357   winter: r = 0.72), which will be discussed in their respective sections. Overall, *Factor 1* represents

358   emissions from traffic and industry at BCN.

359   At MSY, the mixing ratios of *Factor 1* were considerably lower when compared to BCN due to its

360   distance from anthropogenic sources. During summer, levels were negligible and increased in winter

361   due to greater anthropogenic emissions. During summer at MSY, the *traffic & industry factor* showed

362   some correlation with solar radiation (r = 0.42), with increasing concentrations during daylight hours

363   when compared to the evening (Figure 5), while the two peaks that coincided with rush hours at BCN

15

364   were absent. All other correlations with anthropogenic contaminants were not statistically significant.

365   This indicates that instead of an anthropogenic origin, the factor might have a minor biogenic

366   contribution to MSY. This should be considered an option since toluene could be emitted by

367   Mediterranean holm oaks and pines (Heiden et al., 2009; Holzinger et al., 2000; Misztal et al., 2015).

368   Figure S6 presents the diel cycle of toluene and temperature for each day. Each diel cycle shows a

369   peak during midday on most measured days, which was absent for BCN. This correlation was only

370   present for toluene and missing for the C_8 aromatics. Although the diel cycle showed a similar pattern

371   for both the toluene and the C_8 aromatics, the correlation between these two compounds at MSY was

372   not as high (r = 0.52, compared to r = 0.94 at BCN for the same season). Both VOCs still had a significant

373   correlation with other anthropogenic tracers (Figure S3-4), which would indicate an anthropogenic

374   origin for this factor with a minor biogenic contribution from toluene. Yáñez-Serrano et al. (2021)

375   identified a similar factor at a site located nearby the MSY station and attributed it to traffic emissions,

376   which was the lowest contributing factor during summer, as we observed in our study. Tan et al. (2021)

377   also found an anthropogenic source in Hong Kong with the same tracers, attributing it to industrial

378   and traffic emissions.

379   ### 3.3.2 Factor 2: Anthropogenic II: Traffic & Biomass Burning

380   *Factor 2* was traced by benzene (69%), acetaldehyde (67%), and acetonitrile (36%), as well as various

381   oxidized VOCs (OVOCs), such as MEK (27%), methanol (19%), and C_8 aromatics (19%). Both benzene

382   and acetonitrile are typical markers of biomass burning, with studies also showing that acetaldehyde,

383   MEK, and methanol can originate from biomass burning (Cerqueira et al., 2013; Holzinger et al., 1999;

384   Li et al., 2014; Tan et al., 2021).

385   The mixing ratio of this factor was much higher during winter when compared to summer at both

386   stations. While biomass burning can have various sources, such as forest fires and the burning of

387   agricultural waste, the increase during winter indicated that this increase was due to residential

388   heating, which was previously detected in the study area (Reche et al., 2012; Viana et al., 2013). The

16

389   T/B ratio also confirmed this, with lower ratios during the winter season when compared to the

390   summer at both stations, indicating a higher influence of biomass burning (Figure 7).

391   Figure 7 presents a scatterplot of the benzene and toluene mixing ratios. While they are mostly

392   grouped together, showing a high T/B ratio, two periods deviated from this (marked in red). The first

393   was at BCN on the 14th of February 2022 between 07:00 and 10:00 and the second was at MSY on the

394   13th of June 2022 between 17:00 and 21:30. During these events, the T/B ratio differentiated from the

395   group to a ratio below 1, indicating biomass burning as the dominant source (Seco et al., 2013; Tan et

396   al., 2021). However, the peaks coincided with increases in various other VOCs and pollutants;

397   therefore, the origin of this factor cannot be completely attributed to biomass-burning events.

398

399   At BCN, *Factor 2* has a high correlation with the *traffic & industry* factor (*Factor 1*) (summer: r = 0.80;

400   winter: r = 0.87) and BC concentrations (summer: r = 0.82; winter: r = 0.79), which indicates the

401   anthropogenic origin of *Factor 2*. This is further confirmed by the two peaks in the morning and

402   evening, as observed for the *traffic & industry factor* (*Factor 1*), which is common for anthropogenic

403   contaminants at BCN. Notably, the levels of *Factor 2* were much higher during winter when compared

404   to the summer period due to increased anthropogenic contamination, which can also be observed

405   with the higher concentrations of the weekly averages when compared to weekends. During summer,

406   this difference was also much smaller (**Error! Reference source not found.**).

407   Although biomass burning had a major impact on this factor at BCN, it is not likely its sole source.

408   Benzene is emitted from traffic emissions, which cannot be ignored in an urban area such as BCN

409   (Fernández-Iriarte et al., 2020; Gelencsér et al., 1997; Heeb et al., 2000). Furthermore, the diel cycles

410   of acetaldehyde and acetonitrile showed different patterns when compared to this factor (**Error!**

411   **Reference source not found.**). Acetaldehyde can have various origins, such as fossil fuel combustion

412   (Nogueira et al., 2015; Sinharoy et al., 2019; Xu et al., 2022), biomass burning (Cerqueira et al., 2013;

413   Holzinger et al., 2005, 1999), hydrocarbon oxidation (Luecken et al., 2012), and even vegetation

17

414   emissions (Seco et al., 2007). At BCN during winter, acetaldehyde showed the highest mixing ratios

415   during midday, with an additional increase in the evening. However, there were no weekly patterns,

416   with similar concentrations observed during weekly averages and weekend averages. The levels

417   during summer were much lower, showing peaks in the morning and evening. The lower

418   concentration during the day could be caused by the fact that acetaldehyde has an atmospheric

419   lifetime of only a few hours (Possanzini et al., 2002; Seco et al., 2013) due to photochemical

420   destruction (Atkinson and Arey, 2003), which resulted in lower mixing ratios being observed during

421   the day in the summer season. This effect was not observed in the winter due to less sunlight and

422   fewer daylight hours, which reduces the ability of acetaldehyde to be removed from the atmosphere

423   when compared to summer (Filella and Peñuelas, 2006; Viskari et al., 2000).

424   Acetonitrile showed higher levels during the day when compared to the evening during both summer

425   and winter at BCN, with higher levels in the winter (Figure 3). Compared to acetaldehyde, the levels

426   during weekends were much lower, indicating an anthropogenic source. Acetonitrile is generally

427   considered a biomass burning marker with a minor source from fossil fuel combustion, which might

428   be the main source at BCN (Holzinger et al., 2005, 1999). Given the stark difference between weekdays

429   and weekends, this is most likely the case.

430   An investigation of the three biggest tracers of *Factor 2* indicates that while this factor partially

431   included biomass burning at BCN, it also was likely a second traffic source in addition to *Factor 1*.

432   Previous studies assessing sources of biomass burning at BCN stated that after the switch to natural

433   gas, the contribution from residential heating in the particulate matter was relatively low (AIRUSE,

434   2016; Amato et al., 2016; Viana et al., 2013). Notably, its contribution remains significant since a study

435   conducted during winter 2011 at BCN attributed 8% of the $PM_{2.5}$ mass to biomass burning (Reche et

436   al., 2012), while a recent study attributed 15% of the total organic aerosols in PM_{10} to biomass burning

437   (in 't Veld et al. To be Published), indicating a considerable contribution to atmospheric pollution.

18

438 At MSY, the peaks associated with anthropogenic traffic emissions were absent due to the distance

439 from their source, with the mixing ratios of this factor being mostly even throughout the day. During

440 summer, the diel cycle of *Factor 2* showed a peak at approximately 18:00. (Figure 6); however, this

441 was due to a biomass-burning event observed on the 13th of June. The correlation with BC further

442 indicates that this peak was caused by a biomass burning event, with a correlation of r = 0.66 for the

443 entire dataset of *Factor 2* during summer, while the *traffic & industry factor* (*Factor 1*) did not correlate

444 with either this factor or the BC levels. In winter, this factor likely incorporated pollutants from

445 residential heating and other sources transported from BCN and surrounding industrial areas. This was

446 also confirmed by a correlation with the *traffic & industry factor* (r = 0.65), as well as the appearance

447 of a peak at 07:00 for all tracers of *Factor 2*.

448 The diel cycle of benzene at MSY showed similar patterns to the *traffic & biomass factor* (*Factor 2*);

449 however, just like BCN, acetaldehyde and acetonitrile had different diel cycles when compared to

450 *Factor 2*. Both compounds showed higher levels in summer than in winter, increasing during the day

451 and declining in the afternoon. In the case of acetonitrile, this could include long-range transport from

452 forest fires in the Mediterranean area (Seco et al., 2013). Notably, acetonitrile levels were significantly

453 lower when compared to BCN, suggesting that acetonitrile at BCN was of local origin.

454 Acetaldehyde also has a biogenic origin, which was more prominent at MSY (Lewis et al., 2005;

455 Luecken et al., 2012; Possanzini et al., 2002). Furthermore, it can also be formed in the atmosphere

456 by photochemical production from akenes (including biogenic alkenes such as terpenes) (Luecken et

457 al., 2012; Seco et al., 2007). Higher levels in the summer compared to winter could be a result of the

458 increased physiological activity of vegetation and increased secondary production due to greater

459 photochemical activity, as well as higher local acetaldehyde emissions from forest fires (Seco et al.,

460 2011).

461 A previous study by Yáñez-Serrano et al. (2021) near MSY did not identify a biomass burning source

462 during summer due to its low contributions from biomass burning during this season and the lack of

19

463   confirmed wildfires during their study. Tan et al. (2021) also identified a biomass burning source in

464   Hong Kong using the same tracers as this study.

3.3.3 Factor 3: Isoprene Oxidation Source

466

467   *Factor 3* was traced by isoprene (81%) and MVK+MACR (65%), both of which are oxidation products

468   of isoprene (Ling et al., 2019; Liu et al., 2013). This factor also had lower levels of OVOCs attributed to

469   it, such as MEK (15%), acetone (10%), and monoterpenes (10%). The diel cycles of MSY showed a clear

470   increase during daylight in summer, which was less pronounced at BCN (Figure 8). Both stations

471   confirmed this trend with a significant correlation of *Factor 3* with solar radiation (BCN: r = 0.48; MSY:

472   r = 0.52) at both stations (Figures S1 and S3).

473   At MSY, the diel cycle indicated a biogenic origin for isoprene (Figure 3), which is known to be emitted

474   by vegetation (Pacifico et al., 2011; Sanadze, 2004). This was further confirmed by the high correlation

475   between this factor and the *monoterpene factor* (*Factor 4*) (r = 0.91) during summer, which was also

476   emitted by biogenic activity (see description for Factor 4). This confirms the findings of Yáñez-Serrano

477   et al. (2021a), who also measured high levels of isoprene and its oxidation products in Montseny

478   Natural Park (a few kilometers from our station) during summertime and found a diel cycle similar to

479   the solar radiation cycle, which was also observed in this study (r = 0.52). This could be surprising since

480   the forest at MSY is dominated by holm oak, which typically emits isoprene at a lower rate (only 5%

481   compared to monoterpenes) (Peñuelas et al., 2009). Furthermore, several *Quercus* species, such as

482   *Quercus ilex* (a dominant species in the MSY region), also emit isoprene (Fernández-Martínez et al.,

483   2018). However, a previous report for the same MSY site by Seco et al. (2011) found summertime

484   isoprene mixing ratios amounting to approximately two-thirds of those of monoterpenes and with

485   very similar diel patterns, which is in agreement with the present study and suggests that other

486   vegetation present in those forests and the surrounding areas are isoprene emitters. While the exact

487   origin of isoprene cannot be confirmed yet, it was speculated these come from *Erica arborea*, a shrub

20

488   located near MSY and known to emit a considerable amount of isoprene (Pagès et al., 2020; Yáñez-

489   Serrano et al., 2021a). Yáñez-Serrano et al. (2021a) also hypothesized that isoprene could originate

490   from local vehicle exhaust emissions (Borbon et al., 2001), but no correlation was found between

491   isoprene and any anthropogenic contaminant (Figure S3). Meanwhile, concentrations during winter

492   were considerably lower due to the smaller biogenic emissions as a result of reduced solar radiation

493   and temperature. Moreover, the diel pattern of the isoprene to MVK+MACR ratio (**Error! Reference**

494   **source not found.**) showed that during summer at MSY, the ratios decrease during daylight hours until

495   18:00, with a maximum at approximately 09:00. This indicates that biogenic isoprene becomes

496   oxidized into secondary MVK and MACR.

497   At BCN, a different situation occurred. *Factor 3* showed an increase during daylight hours and

498   decreased in the afternoon, with an additional small peak in the morning and evening (**Error!**

499   **Reference source not found.**). The morning and evening peaks coincide with those from other

500   anthropogenic emissions, which implies there was also an anthropogenic source for this factor (**Error!**

501   **Reference source not found.**). The anthropogenic origin was further confirmed with the positive

502   correlation with BC (r = 0.50), the *traffic & biomass burning factor* (*Factor 2*) (r = 0.56), and the *traffic*

503   *& industries factor* (*Factor 1*) (r = 0.44) (Figure S1) during summer. The correlation was even higher

504   during winter, with a positive correlation with BC (r = 0.46), *traffic & biomass burning factor* (*Factor*

505   *2*) (r = 0.72), and the *traffic & industries factor* (*Factor 1*) (r = 0.60). Isoprene is known to be emitted

506   by anthropogenic activities, particularly traffic (Borbon et al., 2002, 2001; Filella and Peñuelas, 2006;

507   Wagner and Kuttler, 2014; Yee et al., 2020). However, other anthropogenic VOCs can also contribute

508   to the signal of m/z 69 since it is a common fragment of cycloalkenes (Gueneron et al., 2015).

509   Nevertheless, the maximum isoprene mixing ratios in summer occurred at 13:00, when temperature

510   and solar radiation were also at their maximum. This observation suggests that a biogenic source

511   remains present at BCN, which is logical given the forests surrounding the area.

21

512 The isoprene oxidation factor was previously identified by Yáñez-Serrano et al. (2021) in Montseny as

513 a biogenic source along with monoterpenes. While monoterpenes are mostly of biogenic origin,

514 isoprene is not necessarily biogenic. Combining the datasets of the two stations (i.e., BCN and MSY)

515 separated this biogenic source between the *isoprene oxidation factor* (*Factor 4*) and the *monoterpene*

516 *factor* (*Factor 5*). The *isoprene oxidation factor* is mostly biogenic at MSY, with high levels in the

517 summer and a clear correlation with temperature (Figure S3). Meanwhile, at BCN, an additional

518 anthropogenic source of isoprene exists, especially during winter, as shown by the diel cycle and

519 correlation with other anthropogenic contaminants (Figure 9; Figure S1).

520

521 ### 3.3.4 Factor 4: Monoterpenes

522 The monoterpene factor (Factor 4) was mostly traced by monoterpenes (84%), with minor

523 contributions from methanol (10%) and C8 aromatics (10%). The monoterpene factor was emitted

524 from biogenic sources at MSY, as seen in the diel pattern (Figure 10). Just like for the isoprene

525 oxidation factor, levels were extremely high during the day while being low at night and in the

526 evening. Furthermore, a clear distinction between seasons is apparent since this factor exhibited its

527 minimum mixing ratios during winter without any diel cycle, whereas it had a clear diel cycle

528 correlating with sunlight and temperature in summer (Figure S3-4). The high levels measured are

529 due to the MSY station being embedded among holm oak trees, which emit these compounds in

530 relevant quantities (Peñuelas et al., 2009; Seco et al., 2013, 2011). This factor was previously

531 identified by Yáñez-Serrano et al. (2021) at Montseny, where it was combined with the isoprene

532 oxidation factor into a single biogenic source.

533 This was not the case for BCN, where the diel cycle did not resemble one increasing with sunlight. This

534 factor did correlate with the anthropogenic factors (*traffic & biomass burning and traffic & industries*)

535 during both seasons, with a correlation factor between 0.61 and 0.72 during summer and 0.71 and

536 0.78 during winter (Figure S1-2). However, the peaks in the morning and evening did not coincide with

22

537   the peaks of the anthropogenic contaminants (Figures 2 and 10). Hellén et al. (2012) proposed a

538   traffic-related origin of monoterpenes since the diel cycle of monoterpenes resembles those of

539   anthropogenic aromatics; however, as previously mentioned, the peaks at BCN do not coincide with

540   said anthropogenic aromatics. It is more likely that the diel cycle shape was caused by the rapid

541   reaction of monoterpenes with daytime generated OH$^{\bullet}$ radicals and dilution by atmospheric mixing,

542   as was reported previously in the study area (Filella and Peñuelas, 2006; Harrison et al., 2001).

543   Although lower levels were observed during weekends, those values might still indicate an

544   anthropogenic origin. The reaction rates with OH$^{\bullet}$, presented in Table 6, indicate that the reactivity of

545   monoterpenes is at least 10- to 100-fold faster when compared to the other VOCs.

546   ### 3.3.5 Factor 5: Long-lifetime VOCs

547   The final factor was a mixture of the remaining VOCs in the study area. These were traced by various

548   OVOCs, including acetone (71%), methanol (69%), acetonitrile (61%), MEK (40%), MVK+MACR (31%),

549   and acetaldehyde (23%).

550   Methanol and acetone had the highest mixing ratios at both stations during both seasons, which can

551   be attributed to their long lifetime in the atmosphere and the fact that they are formed from the

552   oxidation of other VOCs (Seco et al., 2007) (Table 6). This resulted in high background levels, as

553   observed in previous studies (Filella and Peñuelas, 2006; Lewis et al., 2005; Seco et al., 2013; Yáñez-

554   Serrano et al., 2021a). Acetonitrile had much lower mixing ratios but also has a long atmospheric

555   lifetime ranging from months to years (Sanhueza et al., 2004). Therefore, this factor was called the

556   *long-lifetime VOCs factor*.

557   At BCN, the *long-lifetime VOCs factor* showed a similar diel pattern to those of methanol and acetone,

558   which was similar to other anthropogenic contaminants during summertime (Figure 11). This resulted

559   in positive correlations with said factors (*traffic & industries*: 0.49; *traffic & biomass burning*: 0.52), as

560   well as the observation of two peaks—one early in the morning and one in the evening—with much

561   lower mixing ratios during weekends. At BCN, Filella & and Peñuelas (2006, and references therein)

23

562   found that both methanol and acetone were emitted by road traffic, which resulted in peaks during

563   rush hours. The observed decrease of *factor 5* during the day would be a result of photochemical

564   destruction and dilution effects. The anthropogenic origin of these VOCs is further confirmed by the

565   lower mixing ratios over weekends.

566   During winter at BCN, the highest mixing ratios of *Factor 5* occurred at midday, with a small peak in

567   the evening (Figure 11). During summer, the disparity between the weekday and weekend averages

568   was highly accentuated (Figure 11); however, this difference was only observed in the middle of the

569   day during winter. This suggests that at BCN, this factor has a combination of sources—mostly

570   anthropogenic and from atmospheric oxidation—and that this combination can change with the

571   season. The higher levels observed during winter were most likely the result of less sunlight and the

572   consequently diminished photochemical destruction of methanol and acetone.

573   During summer at MSY, the long-lifetime VOCs factor mixing ratio does not represent the trends of

574   any VOCs. This is most likely a result of the different trends that all the VOCs have at this station, which

575   averages to an overall neutral trend. Methanol showed three peaks (Figure 3), with one in the morning

576   (08:00) possibly coinciding with stomatal opening (Fall and Benson, 1996; Hüve et al., 2007), and two

577   more around midday and in the evening. This suggests both the biogenic and photo-oxidation origins

578   of methanol at the MSY station. On the other hand, acetone only displayed a peak during the day that

579   was highly correlated with temperature (r = 0.75) and the biogenic sources (isoprene oxidation: r =

580   0.69; monoterpenes r = 0.56). This would indicate that, like methanol, acetone at MSY can be both

581   formed due to photo-oxidation and biogenic emissions at MSY. The only difference was that acetone

582   did not have a stomatal release in the morning, hence the lack of a morning peak.

583   On the other hand, during the winter season, all three long-lifetime VOCs showed no diel cycle due to

584   lower biogenic emissions and sunlight. This indicates that they are mostly long-range transported

585   VOCs due to their lengthy atmospheric lifetimes.

24

3.4 Ozone formation potential

The contribution of VOCs to the formation of O_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_3 concentrations measured simultaneously. At BCN during winter, the OFP was 36% of the total O_3, while it was 13% at MSY during winter. Meanwhile, during summer, BCN exhibited 28% of the total O_3, while MSY exhibited 22%. This dispersion can be due to the limited VOC set measured, as well as various other sources of O_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,

25

612   the situation was reversed, where the OFP was primarily driven by OVOCs, with 13.06 ppbv (68%)

613   compared to 6.01 ppbv (32%) of HVOCs. This difference in contribution is a commonly observed

614   phenomenon, where OVOCs contribute significantly more to rural areas when compared to urban

615   areas (Chen et al., 2022; Louie et al., 2013; Luo et al., 2011).

616   At BCN, the OFP primarily originates from the *traffic & industry factor* (*Factor 1*) since both toluene

617   and C_8 aromatics, which were the main tracers of this factor, together made up 50% of the estimated

618   OFP during summer, with 9.13 ppbv. On the other hand, due to the relatively lower MIR and mixing

619   ratios of benzene, it contributed very little toward OFP. The total OFP from the aromatics (benzene,

620   toluene, and C_8 aromatics) was nearly 1.5 times lower during winter (6.28 ppbv) due to the lower

621   mixing ratios of toluene and C_8 aromatics, accounting only for 39% of the total OFP. At MSY, the

622   contribution of the *traffic & industries factor* was significantly lower, with the relative contribution of

623   toluene and C_8 aromatics to OFP being 35% (2.85 ppbv) during winter, which is nearly 2.2-fold lower

624   when compared to BCN during the same season. During summer, the contribution of these aromatics

625   to OFP was only 7% with 1.37 ppbv, representing one of the smallest contributing sources and 6.6

626   times lower when compared to summer at BCN. Although this is worth keeping in mind, since MSY is

627   a rural area, most toluene and C_8 aromatics will be transported from industrial and urban sources and

628   are thus already photo oxidized to O_3 and SOAs when the air masses reach the station (Seco et al.,

629   2013).

630   The biogenic sources, which are the *monoterpenes factor* (*Factor 4*) and the *isoprene oxidation*

631   *products* (*Factor 3*), had much higher mixing ratios during summer due to increased biogenic emissions

632   at both stations. The combination of isoprene, MVK+MACR, and monoterpenes accounted for 20% of

633   the OFP during summer at BCN. This significantly decreased during winter to a combined 7%, with an

634   OFP of 1.16 ppbv compared to 3.71 ppbv during summer. Notably, most of the OFP during winter

635   originated from isoprene and its oxidation products since the monoterpene's contribution at BCN was

636   low. As previously established, this also includes anthropogenic isoprene. At MSY, these sources were

26

637   mainly biogenic and contributed much more to the total OFP. During summer, the biogenic sources

638   made up 45% of the total OFP with 8.59 ppbv, which was much lower during winter when they only

639   comprised 11% of the total OFP with 0.91 ppbv. This made biogenic VOCs the main driver of OFP at

640   MSY during summer.

641   The final significant contribution to OFP was from acetaldehyde, which, with its high MIR and mixing

642   ratios, contributed significantly to the OFP at both stations. During winter, it contributed 34% of the

643   total OFP at BCN with 5.60 ppbv, which was much higher than during summer, when it only

644   contributed 10% of the total OFP with 1.89 ppbv. During winter at MSY, acetaldehyde also contributed

645   35% of the total OFP, with 2.94 ppbv, while during summer it only contributed 27% of the total OFP

646   but it contributed more (5.07 ppbv) in absolute terms. Ultimately, acetaldehyde contributed relatively

647   the most during the winter season at both stations, even though the actual OFP was higher during the

648   summer season at MSY. Methanol followed closely with 12 to 14% of the total OFP at all stations.

3.5 SOA formation potential

650   Most of the existing literature has estimated the SOAP using the relative SOAP to toluene ratio

651   obtained from Derwent et al. (2010). While useful, the list presented there does exclude some

652   essential VOCs that we measured in our study. Therefore, we obtained the SOAP from Gu et al. (2021),

653   who reported the SOA yield of various VOCs included in this study. It is crucial to state that this

654   method, like the OFP, has the limitation that the SOA yield was calculated based on an urban

655   environment with a high NO_x atmosphere, which influences SOA formation (Camredon et al., 2007).

656   While this situation applies to BCN, it is more inaccurate in the case of MSY. Additionally, it is worth

657   mentioning that the SOA yields used $OH^{\bullet}$ as the primary oxidant to form SOA (Gu et al., 2021). O_3 is a

658   secondary oxidant that can also form SOA, and with O_3 levels being generally higher in the MSY rural

659   area, its capability to oxidize will increase accordingly. However, the reaction rates of VOCs with O_3

660   are 10^5 to 10^{11} times lower (Atkinson and Arey, 2003; Atkinson and Carter, 1984) than those with $OH^{\bullet}$,

27

661   thereby diminishing the impact of this difference. The results of the SOAP are presented in Table 7

662   and Figure 12.

663   Overall, aromatics contributed the most due to their high SOA yields at both stations during both

664   seasons (Table 7, Figure 12). Since the mixing ratios of these aromatic compounds were much higher

665   at BCN, a much higher SOAP was observed in this study. Due to the aromatic VOCs' high SOA yields,

666   they contributed the most to the SOAP at both stations during both seasons (Table 7). Since the mixing

667   ratios of these aromatic compounds were much higher at BCN, the SOAP was considerably higher

668   compared to MSY. At MSY, the SOAP was similar between the two seasons but was dominated by

669   anthropogenic aromatic compounds (benzene, toluene, C_8 aromatics) during winter, while the

670   monoterpenes were a considerable source of SOAP in summer.

671   This would indicate that *traffic & industry* (*Factor 1*) was the main contributor to SOA formation from

672   the measured VOCs at BCN since none of the remaining OVOCs and isoprene oxidation products have

673   a significant SOA yield. Although monoterpenes have a significant SOA yield, the levels at BCN were

674   much lower, resulting in minimal influence. However, the diel cycle of the *monoterpene factor* (*Factor

675   4*) suggests that during summer at BCN, monoterpenes were rapidly reacting with daytime- generated

676   OH· radicals, which would result in the formation of both O_3 and SOA; therefore, the actual

677   contribution of monoterpenes would be much higher than the values given in Table 7.

678   At MSY during winter, the *traffic & industry* factor (*Factor 1*) seemed to be the greatest source of SOA

679   in the area. However, as was the case with the OFP, the actual SOAP might be higher as the

680   transported toluene and C_8 aromatics deplete during transport. During summer, the monoterpenes

681   were the main contributors to the SOAP, while other biogenic compounds, such as isoprene, did not

682   contribute as much due to their low SOA yields.

28

4.0 Conclusions

684   In this study, we quantified 11 VOC mixing ratios during summer and winter at both the BCN urban

685   background site and the MSY rural background site (Table 5). Applying PMF to the combined dataset

686   revealed five factors (Table 6). The first factor, *traffic & industries* (*Factor 1*), originated from

687   anthropogenic sources such as traffic and industry in and around Barcelona, mostly consisting of aged

688   air masses at MSY. The second factor, *traffic & biomass burning* (*Factor 2*), represented a secondary

689   source combining traffic emissions and biomass burning. The third factor, *isoprene oxidation* (*Factor

690   3*), was associated with isoprene and its oxidation products, which had both anthropogenic and

691   biogenic origins at BCN and were mostly biogenic at MSY. The fourth factor, *monoterpenes* (*Factor 4*),

692   was predominantly biogenic at both stations, with lower levels at BCN due to rapid oxidation by $OH^\bullet$

693   and lower biogenic emissions in the area. The fifth factor (*Factor 5*) represented a mixture of VOCs

694   with longer atmospheric lifetimes and included methanol, acetone, and acetonitrile, which had longer

695   lifetimes when compared to other VOCs.

696   Analyzing the sources of the VOCs in relation to OFP revealed that at BCN, anthropogenic VOCs—

697   particularly toluene and C_8 aromatics—were the primary contributors to O_3 formation, accounting for

698   38 and 49% of the measured OFP during winter and summer, respectively (Table 7). BCN also had

699   significant contributions from *isoprene oxidation products* (6 and 19% during winter and summer,

700   respectively) and acetaldehyde (34 and 10% during winter and summer, respectively). Methanol

701   contributed 14 and 13% during winter and summer, respectively. At MSY, biogenic sources—and

702   specifically *isoprene oxidation products* and *monoterpenes*—played a larger role in the total OFP,

703   contributing 11 and 45% during winter and summer, respectively. The aged anthropogenic VOCs from

704   the *traffic & industries factor* made a considerable contribution of 34% during winter and 7% during

705   summer. Acetaldehyde contributed 35 and 27%, while methanol contributed 12 and 14%, during

706   winter and summer, respectively.

707   At BCN, the results demonstrate that both O_3 and SOA formation were predominantly driven by

708   anthropogenic emissions, particularly those originating from the *traffic & industries factor*. The

709   aromatic compounds had a significant impact on the SOAP due to their higher SOA yields (Table 9).

710   The seasonality showed a significant increase during summer, with concentrations of 0.766 μgm^{-3}

711   compared to 0.603 μgm^{-3} during winter, primarily due to higher levels of toluene and C_8 aromatics

712   during summer (Table 9). The OFP also showed seasonality with higher levels during the summer, but

713   the difference in total OFP was small, with 16.5 ppbv during winter and 18.51 ppbv during summer

714   (Table 7). The MSY rural background site exhibited a different pattern, where both OFP and SOAP were

715   driven by aged anthropogenic emissions, predominantly from the *traffic & industries factor*, during

716   winter. In contrast, during summer, biogenic sources—specifically the *isoprene oxidation products* and

717   *monoterpenes* factors—played a more prominent role. In the case of SOAP, the *monoterpenes factor*

718   contributed 53% of the overall SOAP during summer, while aromatic compounds contributed 29%.

719   Comparing the total OFP between the two stations, BCN and MSY revealed similar values during the

720   summer season, with 18.51 ppbv at BCN and 19.07 ppbv at MSY. However, during winter, the OFP was

721   considerably lower at MSY, with 8.42 ppbv compared to 16.50 ppbv at BCN. On the other hand, the

722   SOAP exhibited higher concentrations at BCN when compared to MSY, with 1.163 μgm^{-3} and 0.603

723   μgm^{-3} during winter at BCN and MSY, respectively, and 1.833 μgm^{-3} and 0.766 μgm^{-3} during summer

724   at BCN and MSY, respectively.

725   While this study provides valuable insights into the sources of VOCs and their impacts on OFP and

726   SOAP, some knowledge gaps should be acknowledged. First, both OFP and SOAP estimations are

727   based on high nitrogen oxide (NO_x) environments, which may not accurately represent the low NOx

728   conditions at the MSY rural background station. Secondly, the biogenic contribution of monoterpenes

729   is likely underestimated at BCN during summer due to their rapid oxidation by $OH^{\cdot}$ in an urban

730   environment. Lastly, this study focused on a selection of 11 VOCs, and although they make substantial

731   contributions to atmospheric concentrations, the inclusion of additional VOCs—particularly alkanes

30

732 (~5–10% contribution to OFP and SOAP) and alkenes (>30% contribution to OFP)—would provide a

733 more comprehensive understanding (Jookjantra et al., 2022; Wang et al., 2023; Wu et al., 2017; Yuan

734 et al., 2009; Zhan et al., 2021).

5.0 Acknowledgments

31