Supplementary information to

Formation of organic aerosol in the Paris region during the MEGAPOLI
summer campaign: evaluation of the Volatility-Basis-Set approach within
the CHIMERE model

Q.J. Zhang (1, 13), M. Beekmann (1), F. Drewnick (2), F. Freutel (2), J. Schneider (2), M. Crippa (3),
A. Borbon (1), A. Colomb (1, 6), V. Michoud (1), J-F. Doussin (1), H. A. C. Denier van der Gon (7),
M. Haeffelin (8), J-C. Dupont (8), G. Siour (1, 9), H. Petetin (1, 12), B. Bessagnet (9), S.N. Pandis
(10), A. Hodzic (11), O. Sanchez (12), C. Honoré (12), O. Perrussel (12)

1) Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), Université Paris Est et 7,
IPSL, CNRS, Créteil, France
(2) Max Planck Institute for Chemistry, Particle Chemistry Department, Mainz, Germany
(3) Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, Villigen, Switzerland
(4) Leibniz Institute for Troposphärenforschung, Leipzig, Germany
(5) Laboratoire des sciences du climat et de l’environnement, IPSL, CEA et l’Université de Versailles,
Saint-Quentin, CNRS
(6) Laboratoire de Météorologie Physique, Clermont-Ferrand, France
(7) TNO, Dept. Climate, Air and Sustainability, Utrech, The Netherlands.
(8) IPSL, Ecole Polytechnique, INSU/CNRS l’Université de Versailles, Saint-Quentin., France
(9) Institut National de l’EnviRonnement industriel et des rISques, Verneuil en Halatte, France
(10) Institut of Chemical Engineering Sciences, Foundation for Research and Technology, Hellas,
Patras, Greece
(11) National Center for Atmospheric Research, Boulder, USA
(12) AIRPARIF, Agence de Surveillance de la qualité de l’air, Paris, France
(13) Aria technologies, Boulogne-Billancourt, France

Table S1. Parameters used to simulate the partitioning of POA emissions into different volatility bins;
the mass weight of each lumped species is 250 g mol\(^{-1}\). The enthalpies are derived from measurements
and theoretical estimation for specific low-volatile species in POA emissions by Dohanue et al.,

<table>
<thead>
<tr>
<th>Lumped species</th>
<th>10^{-2}</th>
<th>10^{-1}</th>
<th>10^{0}</th>
<th>10^{1}</th>
<th>10^{2}</th>
<th>10^{3}</th>
<th>10^{4}</th>
<th>10^{5}</th>
<th>10^{6}</th>
</tr>
</thead>
<tbody>
<tr>
<td>C* (µg m^{-3}) at 298K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔHv (kJ mol^{-1})</td>
<td>112</td>
<td>106</td>
<td>100</td>
<td>94</td>
<td>88</td>
<td>82</td>
<td>76</td>
<td>70</td>
<td>64</td>
</tr>
<tr>
<td>Emission factors</td>
<td>0.03</td>
<td>0.06</td>
<td>0.09</td>
<td>0.14</td>
<td>0.18</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Table S2. Lumped anthropogenic and biogenic VOCs and their major components for the 4 bins VBS approach for SOA formation.

<table>
<thead>
<tr>
<th>Anthropogenic origin</th>
<th>Biogenic origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumped species</td>
<td>Major components</td>
</tr>
<tr>
<td>ALK4</td>
<td>n-pentane, n-hexane, Branched C5-C6 Alkanes, Cyclopentane, Trimethyl Butane, Trimethyl Pentane, Isopropyl alcool, n-Propyl Alcool</td>
</tr>
<tr>
<td>ALK5</td>
<td>C7-C22 n-Alkanes, C6-C16 Cycloalkanes, branched/Unspeciated C8-C18 Alkanes</td>
</tr>
<tr>
<td>OLE1</td>
<td>Propene, C4-C15 terminal Alkanes</td>
</tr>
<tr>
<td>OLE2</td>
<td>Isobutene, C4-C15 Internal Alkenes, C6-C15 Cyclic or di-olefins, Styrenes</td>
</tr>
<tr>
<td>ARO1</td>
<td>Toluene, benzene, Ethyl benzene, C9-C13 Monosubstituted Benzenes</td>
</tr>
<tr>
<td>ARO2</td>
<td>Xylenes, Ethyl Toluenes, Dimethyl and Trimethyl Benzenes, Ethylbezenes, naphthalene, C8-C13 Di-, Tri-, Tetra-, Penta-, Hexa-substituted Benzenes, Unspeciated C10-C12 Aromatics</td>
</tr>
</tbody>
</table>

Table S3. Normalized SOA yields $\alpha_i$ in the 4 bins VBS approach with saturation concentrations from 1 to $10^3 \mu g m^{-3}$; $\Delta H$: enthalpy of products; Molar mass (MW) of products; the density of OA is assumed as 1.5 g cm$^-3$ (Murphy and Pandis, 2009).

<table>
<thead>
<tr>
<th>VOC precursors</th>
<th>Mass yields of each bin</th>
<th>$\Delta H$ (kJ mol$^{-1}$)</th>
<th>MW (g mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C* (µg m$^{-3}$)</td>
<td>1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>ALK4</td>
<td>0.0</td>
<td>0.075</td>
<td>0.0</td>
</tr>
<tr>
<td>ALK5</td>
<td>0.0</td>
<td>0.300</td>
<td>0.0</td>
</tr>
<tr>
<td>OLE1</td>
<td>0.0045</td>
<td>0.009</td>
<td>0.060</td>
</tr>
<tr>
<td>OLE2</td>
<td>0.0225</td>
<td>0.435</td>
<td>0.129</td>
</tr>
<tr>
<td>ARO1</td>
<td>0.075</td>
<td>0.225</td>
<td>0.375</td>
</tr>
<tr>
<td>ARO2</td>
<td>0.075</td>
<td>0.300</td>
<td>0.375</td>
</tr>
<tr>
<td>TERP</td>
<td>0.1073</td>
<td>0.0918</td>
<td>0.3587</td>
</tr>
<tr>
<td>ISOP</td>
<td>0.009</td>
<td>0.03</td>
<td>0.015</td>
</tr>
</tbody>
</table>
Table S4. Statistics for comparisons between MM5 results and meteorological measurements at SIRTA, based on 30 days of comparisons with hourly data (corresponding to the frequency of the model output). R is the correlation coefficient; RMSE is root mean square error.

<table>
<thead>
<tr>
<th>Statistics at SIRTA</th>
<th>Wind speed (m s(^{-1}))</th>
<th>Temperature (K)</th>
<th>PBL height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias (relative)</td>
<td>+1.07 (+42%)</td>
<td>-0.86</td>
<td>-268 (-28%)</td>
</tr>
<tr>
<td>RMSE (relative RMSE)</td>
<td>1.64 (+64%)</td>
<td>1.93</td>
<td>764 (+79%)</td>
</tr>
<tr>
<td>R</td>
<td>0.73</td>
<td>0.92</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table S5. Statistics for comparison of modelled and measured NO\(_x\), BC, O\(_3\), inorganic aerosol, and organic aerosol from three configurations (CSS, VBS-T1 and VBS-T2) by using the LA emission inventory at GOLF, LHVP and SIRTA, based on 30 days of comparisons with hourly data (corresponding to the frequency of the model output). R is the correlation coefficient; RMSE is root mean square error.

<table>
<thead>
<tr>
<th>GOLF</th>
<th>LHVP</th>
<th>SIRTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO(_x) (ppb)</td>
<td>+6.1 (+58%)</td>
<td>+2.7 (+16%)</td>
</tr>
<tr>
<td>O(_3) (ppb)</td>
<td>+0.3 (+1.0%)</td>
<td>-1.2 (-4.1%)</td>
</tr>
<tr>
<td>BC (µg m(^{-3}))</td>
<td>+2.07 (+180%)</td>
<td>+2.08 (+167%)</td>
</tr>
<tr>
<td>SO(_4)(^{2-}) (µg m(^{-3}))</td>
<td>+0.97 (+84%)</td>
<td>+0.83 (+65%)</td>
</tr>
<tr>
<td>NO(_2) (µg m(^{-3}))</td>
<td>+0.44 (+137%)</td>
<td>+0.23 (+59%)</td>
</tr>
<tr>
<td>NH(_4) (µg m(^{-3}))</td>
<td>+0.40 (+73%)</td>
<td>+0.45 (+97%)</td>
</tr>
<tr>
<td>OA- CCS (µg m(^{-3}))</td>
<td>4.02</td>
<td>3.58</td>
</tr>
<tr>
<td>OA- VBS-T1 (µg m(^{-3}))</td>
<td>3.13</td>
<td>2.71</td>
</tr>
<tr>
<td>OA- VBS-T2 (µg m(^{-3}))</td>
<td>2.04</td>
<td>2.14</td>
</tr>
</tbody>
</table>
Table S6. Statistics for comparison of modelled and measured NO\textsubscript{x}, BC, O\textsubscript{3}, inorganic aerosol, and organic aerosol from the configuration VBS-MPOLI by using the TNO-MEGAPOLI emission inventory at GOLF, LHVP and SIRTA. R is the correlation coefficient; RMSE is root mean square error.

<table>
<thead>
<tr>
<th></th>
<th>VBS-MPOLI</th>
<th>GOLF</th>
<th>LHVP</th>
<th>SIRTA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (relative)</td>
<td>RMSE (relative)</td>
<td>R</td>
<td>Bias (relative)</td>
</tr>
<tr>
<td>NO\textsubscript{x} (ppb)</td>
<td>+10.0 (+95%)</td>
<td>15.6 (148%)</td>
<td>0.57</td>
<td>+7.2 (+42%)</td>
</tr>
<tr>
<td>O\textsubscript{3} (ppb)</td>
<td>-1.97 (-6.7%)</td>
<td>10.0 (34%)</td>
<td>0.78</td>
<td>-3.7 (-13%)</td>
</tr>
<tr>
<td>BC (µg m\textsuperscript{-3})</td>
<td>+0.33 (+29%)</td>
<td>0.97 (84%)</td>
<td>0.55</td>
<td>+0.52 (+41%)</td>
</tr>
<tr>
<td>SO\textsubscript{4}\textsuperscript{2-} (µg m\textsuperscript{-3})</td>
<td>+0.87 (+76%)</td>
<td>1.31 (114%)</td>
<td>0.73</td>
<td>+0.75 (+59%)</td>
</tr>
<tr>
<td>NO\textsubscript{3} (µg m\textsuperscript{-3})</td>
<td>+0.01 (+2.1%)</td>
<td>0.48 (148%)</td>
<td>0.90</td>
<td>-0.11 (-27%)</td>
</tr>
<tr>
<td>NH\textsubscript{4} (µg m\textsuperscript{-3})</td>
<td>+0.24 (+43%)</td>
<td>0.45 (80%)</td>
<td>0.82</td>
<td>+0.32 (+69%)</td>
</tr>
<tr>
<td>OA (µg m\textsuperscript{-3})</td>
<td>-1.15 (-32%)</td>
<td>2.16 (61%)</td>
<td>0.80</td>
<td>-0.95 (-29%)</td>
</tr>
</tbody>
</table>

Table S7. Statistic for comparison of modelled POA and SOA with measured HOA and OOA derived from the 3 factor PMF analysis from the four configurations (CSS, VBS-T1, VBS-T2 and VBS-MPOLI) at LHVP. R is the correlation coefficient; RMSE is root mean square error.

<table>
<thead>
<tr>
<th></th>
<th>POA (µg m\textsuperscript{-3})</th>
<th>SOA (µg m\textsuperscript{-3})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (relative)</td>
<td>RMSE (relative)</td>
</tr>
<tr>
<td>CCS (+555%)</td>
<td>4.36 (708%)</td>
<td>0.53 (48%)</td>
</tr>
<tr>
<td>VBS-T1 (+1.8%)</td>
<td>0.61 (99%)</td>
<td>0.45 (40%)</td>
</tr>
<tr>
<td>VBS-T2 (+17%)</td>
<td>0.61 (100%)</td>
<td>0.49 (52%)</td>
</tr>
<tr>
<td>VBS-MPOLI (-75%)</td>
<td>0.67 (99%)</td>
<td>0.55 (37%)</td>
</tr>
</tbody>
</table>
Fig. S1. $\log_{10}(OM)$ emissions (unit: molecule cm$^{-2}$ of a typical weekday in July) centred for the Paris region from LA inventory (left) and the MEGAPOLI database (right), for fine PM (PM$_{2.5}$).

Fig. S2. Comparison of simulated (blue line) and observed (black dots) July 2009 average daily variation of the PBL height at SIRTA.

Fig. S3. BC vs. NO$_x$ at LHVP from measurements (black), from VBS-T2 for the LA-EMEP inventory (green) and from VBS-MPOLI for the MEGAPOLI inventory (purple).
Fig. S4. Comparisons of simulated and observed isoprene, terpenes and aromatics (sum of benzene, toluene, xylene ethyltoluene and 1,2,4-TMB from GC-FID; sum of m/z 79, 93, 107 and 121 from PTRMS which are corresponding to the sum of benzene, toluene, aromatic C8 and C9) at LHVP.
Fig. S5  Comparison of simulated and observed inorganic aerosol species at GOLF

Fig. S6  Comparison of simulated and observed inorganic aerosol species at SIRTA.
Fig. S7 Comparison of simulated and observed primary and secondary organic aerosol when simulated OPOA is considered as part of HOA, on the left side time series, on the right side daily average for July 2009. The color code in the diurnal variations (right side) is the same as in the time series (left side).

Fig S8. Comparison of simulated and observed OH at SIRTA.

Predicted OH concentrations are correlated ($R=0.62$) with the measurement at SIRTA, the driving force for correlation being the strong diurnal cycle with a noon maximum due to enhanced UV radiation. Daily maxima are mostly overestimated with a factor up to three. This overestimation could be due to a variety of reasons: OH and HO$_x$ radical sources ($O_3$, H$_2$O, HONO formed by heterogeneous reactions, photolysable VOC, photolysis frequencies) and sinks of HO$_x$ radicals (NO$_2$, ...) and in NO and VOC levels responsible for radical inter-conversion reactions.