Global atmospheric budget of simple monocyclic aromatic compounds

David Cabrera-Perez1, Domenico Taraborrelli1,a, Rolf Sander1, and Andrea Pozzer1

1Atmospheric Chemistry Department, Max-Planck Institute of Chemistry, Hahn-Meitner-Weg 1, 55128 Mainz, Germany
a now at: Institute of Energy and Climate Research (IEK-8), Forschungszentrum Jülich GmbH, Jülich, Germany

Correspondence to: David Cabrera-Perez (d.cabrera@mpic.de)

Received: 9 December 2015 – Published in Atmos. Chem. Phys. Discuss.: 23 February 2016
Revised: 18 May 2016 – Accepted: 24 May 2016 – Published: 7 June 2016

Abstract. The global atmospheric budget and distribution of monocyclic aromatic compounds is estimated, using an atmospheric chemistry general circulation model. Simulation results are evaluated with an ensemble of surface and aircraft observations with the goal of understanding emission, production and removal of these compounds.

Anthropogenic emissions provided by the RCP database represent the largest source of aromatics in the model (~23 TgC year⁻¹) and biomass burning from the GFAS inventory the second largest (~5 TgC year⁻¹). The simulated chemical production of aromatics accounts for ~5 TgC year⁻¹. The atmospheric burden of aromatics sums up to 0.3 TgC. The main removal process of aromatics is photochemical decomposition (~27 TgC year⁻¹), while wet and dry deposition are responsible for a removal of ~4 TgC year⁻¹.

Simulated mixing ratios at the surface and elsewhere in the troposphere show good spatial and temporal agreement with the observations for benzene, although the model generally underestimates mixing ratios. Toluene is generally well reproduced by the model at the surface, but mixing ratios in the free troposphere are underestimated. Finally, larger discrepancies are found for xylenes: surface mixing ratios are not only overestimated but also a low temporal correlation is found with respect to in situ observations.

1 Introduction

Volatile organic compounds (VOCs) play a significant role in the chemistry of the troposphere and in ozone formation (Atkinson, 2000; Seinfeld and Pandis, 2012). Within the VOCs class, aromatic compounds form a subgroup of special interest: in the troposphere of urban and semi-urban areas, aromatic hydrocarbons comprise a major fraction (up to 60%) of the VOCs (Lee et al., 2002; Ran et al., 2009). Consequently, they are highly relevant for ozone formation in these areas (Kansal, 2009; Barletta et al., 2005; Koppmann, 2008) as they can be responsible for up to 50% of the total ozone formation potential (Tan et al., 2012). Even in rural areas, high levels of aromatics have been reported, summing up to 35% of the total VOCs (Guo et al., 2006; You et al., 2008).

Typical benzene and toluene mixing ratios fall within the 0.1–10 pmol mol⁻¹ range. Estimated lifetimes are 2 days for toluene and 2 weeks for benzene (Koppmann, 2008). These lifetimes are long enough to allow the compounds to reach downwind areas far from sources, as for instance the Sahara desert (Yassaa et al., 2011).

Aromatic VOCs are emitted by a range of sources. They form a relevant fraction of fossil fuels, and they are released into the atmosphere by combustion (i.e., gasoline and diesel engines), gasoline evaporation, solvent usage, and spillage (Sack et al., 1992; Kim and Kim, 2002; Na et al., 2004; Baek and Jenkins, 2004). In urban air masses benzene, toluene, ethylbenzene, xylenes, styrene and trimethylbenzenes are highly present (Koppmann, 2008). After anthropogenic emissions, biofuel and biomass burning are the second largest sources of aromatics. They are important sources of benzene, toluene and phenol in tropical and boreal areas (Fu et al., 2008; Henze et al., 2008; Andreae and Merlet, 2001). Finally, only toluene is considered to be biogenically emitted as it has been shown by the MEGAN model in (Guenther et al., 2012), although a recent study pointed out...
that biogenic emissions of simple aromatics could be equal in importance to anthropogenic emissions (Misztal et al., 2015).

The primary atmospheric oxidation pathway of benzene and alkyl-substituted benzenes is via the reaction with OH, followed by the reaction with NO$_3$ (Atkinson, 2000, and references therein). The oxidation products of aromatic compounds contribute to ozone formation and production of secondary organic aerosol (SOA) (Odum et al., 1997; Butler et al., 2011).

In addition, there are a variety of chemical processes in the atmosphere that involve aromatic oxidation products, which can influence OH recycling in the atmosphere. For instance, ortho-nitrophenols are species of interest due to their HONO-production upon photolysis (Bejan et al., 2006; Chen et al., 2011). Moreover, nitrophenols are emitted directly into the atmosphere by traffic and biomass burning (Tremp et al., 1993; Mohr et al., 2013).

Many aromatic compounds can be dangerous for humans, animal life and plants (Ciarrocca et al., 2012; Snyder et al., 1993). For instance, benzene is known to be carcinogenic (Snyder et al., 1993); toluene and xylenes can have severe effects on the neural system (WMO, 2000; Sarigiannis and Gotti, 2008); and nitrophenols have acute toxicity for humans and plants (Natangelo et al., 1999; Michalowicz and Duda, 2007). Due to the high noxiousness and atmospheric impacts, aromatics have been the subject of monitoring and measurement campaigns (see Table 1), aiming at establishing control strategies for environmental and human health protection.

For these reasons, it is important to have a correct model description of aromatic compounds in the atmosphere, and to have a detailed knowledge of their budget, as this will improve our understanding of their photochemical production yields (Lewis et al., 2000).

However, so far only a few regional or global-scale studies focused on aromatics, e.g. (Henze et al., 2008; Hu et al., 2015; Lewis et al., 2013), while most of the global studies on VOCs only focused on aliphatic hydrocarbons (e.g. Millet et al., 2010; Pozzer et al., 2010; Fu et al., 2008; Paulot et al., 2011; Fischer et al., 2014). To our knowledge, this is the first comprehensive atmospheric budget study on major monocyclic aromatics in the gas phase.

This work focuses on the gas phase chemistry of simple aromatics, hence neglecting any SOA production. Other global studies as Henze et al. (2008) include SOA production as they were focused on the aerosol phase.
Table 2. Total annual anthropogenic emissions for the different emission scenarios in TgCyear$^{-1}$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Benzene</th>
<th>Toluene</th>
<th>Xylenes</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCP</td>
<td>6.35</td>
<td>6.68</td>
<td>5.66</td>
<td>Lamarque et al. (2010)</td>
</tr>
<tr>
<td>LIT</td>
<td>3.24</td>
<td>5.48</td>
<td>4.13</td>
<td>Fu et al. (2008)</td>
</tr>
</tbody>
</table>

In this work we focus only on the most abundant simple aromatics: benzene (C$_6$H$_6$), toluene (C$_8$H$_8$), xylenes (C$_8$H$_{10}$), phenol (C$_6$H$_5$O), styrene (C$_8$H$_8$), ethylbenzene (C$_8$H$_{10}$), trimethylbenzenes (C$_9$H$_{12}$), and benzaldehydes (C$_7$H$_{16}$O). The purpose of this study is to (1) simulate the aromatic compounds in the atmosphere with a global chemistry model; (2) evaluate the model results by comparing different emission scenarios with atmospheric observations from a number of surface and aircraft campaigns and monitoring stations; and (3) provide an estimate of the global atmospheric budget of the most abundant aromatic species in the atmosphere.

In Sect. 2 we present the model setup, including a detailed description of the chemical oxidation mechanism of the aromatic compounds, emissions and sinks. A detailed description of the observations used for the model evaluation is given in Sect. 3. A comparison of model results with the observations is shown in Sect. 4. Finally, we discuss the atmospheric budget of aromatic compounds in terms of sources, sinks and spatial distribution (Sect. 5), followed by the conclusions and outlook.

2 Model description and setup

We use the ECHAM/MESSy Atmospheric Chemistry (EMAC) model. EMAC is a numerical chemistry and climate simulation system which includes the 5th generation European Centre Hamburg general circulation model (ECHAM5, Roeckner et al. (2006)) as the core atmospheric model. Several submodels describing atmospheric processes are connected via the Modular Earth Submodel System (MESSy2.5.0) (Jöckel et al., 2010).

For this study, we use the T63L31ECMWF resolution, which corresponds to a horizontal resolution of T63 with spherical truncation (i.e., a Gaussian grid of approx. 1.9° × 1.9° in latitude and longitude). In this setup, the model has 31 vertical hybrid pressure layers up to 10 hPa. The simulation was nudged towards ECMWF analysis data for a realistic representation of tropospheric meteorology (Jeuken et al., 1996). In order to have the same atmospheric dynamics in all sensitivity simulations, the feedback between chemistry and dynamics is switched off (Chemical Transport Model mode, Deckert et al., 2011).

We performed a 24-month simulation from January 2004 to December 2005. The first 12 months are used as spin-up, and only the results for 2005 are used for the analysis.

Two different scenarios have been simulated, differing only with respect to anthropogenic emissions. A detailed description of the scenarios can be found in the following sections. A summary of the scenarios and the emissions of the different species can be found in Table 2.

In addition, box model simulations have been performed in order to better understand the chemical mechanism used in this work.

2.1 Emissions of aromatics

2.1.1 Anthropogenic

Emissions of aromatics are primarily anthropogenic, coming from numerous sources, including fuel evaporation and combustion, spillage, solvent use, refining of gasoline, landfill wastes and coal-fired stations (Kansal, 2009).

In our study, emissions due to human activities are taken from the Representative Concentration Pathways (RCP) inventory (van Vuuren et al., 2011). The RCP dataset was used in the IPCC’s Fifth Assessment Report and consists of a set of four emission scenarios, developed by four different modeling groups (van Vuuren et al., 2008). Each scenario has a specific radiative forcing for the year 2100 (2.6, 4.5, 6.0, and 8.5 W m$^{-2}$) (van Vuuren et al., 2011). We selected the RCP8.5 pathway (Riahi et al., 2007). Granier et al. (2011) indicate that this assumption is reasonable for the time span 2000–2010. The dataset has a yearly resolution and no seasonal variation. We adopted a vertical distribution of emissions based on the work of Pozzer et al. (2009).

Two simulations with identical meteorology and different anthropogenic emissions have been performed. One scenario has the default emissions developed by the IPCC (denoted as “RCP”). The second scenario, called “LIT”, has scaled RCP emissions, which are adapted to reproduce the total annual anthropogenic emissions reported by Fu et al. (2008). A summary of the scenarios and their total emissions for the different species can be found in Table 2. The scenarios are only different for benzene, toluene and xylenes, since no literature studies have been found for other species. Both scenarios are compared with surface and tropospheric observations, as described in Sect. 4.

In the RCP simulation, 23 TgC of aromatics are released into the atmosphere, which represents 18% of the total anthropogenic VOC emissions. In the LIT scenario, 16 TgC are emitted, and the aromatics represent 13% of the total anthropogenic VOC emissions. When looking into the sectors provided by the RCP, we found for benzene 49% of the emissions originate in the residential sector, followed by the energy sector (29%). In the case of toluene, emissions are evenly split for transportation, energy, solvents and residential. Xylenes emission are similarly distributed as for toluene, but with different magnitudes.
however solvents are the leading source with 30% of the emissions and residential only 7%. Trimethylbenzenes are abundantly emitted by the transportation sector (90%), as well as other aromatics (60%).

2.1.2 Biomass burning

Biomass burning presents a large source of VOCs for the atmosphere (Lamarque et al., 2010). This contribution is represented by the BIOBURN submodel. BIOBURN calculates the emission fluxes, based on the Global Fire Assimilation System (GFAS) datasets (Kaiser et al., 2012). GFAS uses satellite retrievals of fire radiative power and transforms these into dry matter combustion rates. GFAS has a daily time resolution, and therefore seasonal variations can be observed. The dry matter combustion rates are used in the model in combination with biomass burning emission factors to estimate the biomass burning emissions of specific compounds.

For each aromatic species, we applied emission factors retrieved from literature (Yokelson et al., 2013; Andreea and Merlet, 2001; Akagi et al., 2011). The emission factors used in this work are listed in Table 3. For other VOCs, we selected evaluated factors as in Pozzer et al. (2010). Akagi et al. (2011) showed that at least 400 Tg year\(^{-1}\) of VOCs are emitted into the atmosphere from biomass burning. Approximately 5 TgC year\(^{-1}\) are aromatics, which consequently represent less than 2% of the total biomass burning VOC emissions. It is worth mentioning the study of Johnson et al. (2013), who estimated an emission of 10.19 TgC year\(^{-1}\) for phenol. These emissions are dominated by open cooking, although it remains unclear how calculations were done. In the present study open cooking emissions are included within anthropogenic sources but the RCP database does not present such large phenol emissions.

2.1.3 Biogenic

Biogenic emissions have been reported for more than 25 aromatic species (Misztal et al., 2015), although at low amounts. Moreover, most compounds have complex structures (polycyclic aromatics) or more than eight carbon atoms. These compounds are out of the scope of this paper, since only emissions of simple monoaromatic compounds, such as toluene, have been considered here. Toluene fluxes from plants have been measured. The production mechanism is not clear (Heiden et al., 1999) but a considerable source of toluene from vegetation of more than 1 TgC year\(^{-1}\) has been reported (Sindelarova et al., 2014). In this study, biogenic emissions are calculated online by the MEGAN model (v2.04, Guenther et al., 2012). For toluene, the model yields an emission rate of \(\simeq 0.32\) TgC year\(^{-1}\).

2.2 Chemistry

Reaction with OH is the major removal process of aromatic compounds in the troposphere, followed by a small contribution via reaction with NO\(_3\) radicals (Atkinson, 2000). For benzene and the alkyl-substituted benzenes there are two possible pathways concerning the OH reaction, the first and most prominent is the OH radical addition, which amounts to about 90% of the reactions, and the other 10% correspond to H-atom abstraction (Atkinson, 2000). Only for styrene, which contains a non-aromatic double bond, the reaction with O\(_3\) is relevant. Phenol undergoes mostly OH-addition, while benzaldehyde reacts almost exclusively via H-abstraction (Clifford et al., 2005).

In our model, chemical kinetics calculations are done with the MECCA submodel (Sander et al., 2011) which uses the Kinetic PreProcessor (KPP, Sandu and Sander (2006)). For this study, a new reaction mechanism for aromatics has been developed and added to MECCA (Taraborrelli et al, manuscript in preparation). It describes the chemistry of benzene, toluene, xylene, phenol, styrene, ethylbenzene, trimethylbenzenes, benzaldehydes and higher aromatics (lumped alkyl-substituted benzenes with 10 or more carbon atoms). The new scheme is based on the Master Chemical Mechanism (MCMv3.2), the most detailed oxidation mechanism available for aromatics with 3788 reactions and 1271 species (Bloss et al., 2005b). Such detailed representation of the chemistry allows further studies related to the impacts of aromatics in the atmosphere. However, since the MCM is too computationally expensive for global models, it had to be reduced to 666 reactions and 229 species. A complete list of chemical equations and species involved can be found in the Supplement. The mechanism reduction has been performed according to the following procedure:

1. The oxidation mechanisms for benzene and toluene were taken from MCM, because of their relatively high abundance in the atmosphere in comparison with the other aromatics. Therefore, these two species are described with the highest available accuracy.

2. For the other aromatics, the first oxidation step is taken from the MCM, and the second oxidation step is linked to that of toluene, because of the similar chemical structure.

3. Intermediates having a lifetime always below 1 s are replaced by their products with the corresponding reactions being removed.

4. Xylenes and trimethylbenzenes have been lumped, assuming equal proportions of single isomers.

Moreover, the photolysis rate of benzaldehyde is updated following IUPAC recommendations (http://iupac.pole-ether.fr) and ortho-nitrophenols have a new photolysis channel leading to HONO formation. Although shorter or more simplified mechanisms of aromatic decomposition already exist (CRI, MOZART-4) (Jenkin et al., 2008; Emmons et al., 2010), the chemical
Table 3. Biomass burning emission factors for the BIOBURN submodel. Emission factors are given in units of g (species)/kg (dry matter burned). Last column present the global biomass burning emissions for the year 2005.

<table>
<thead>
<tr>
<th>Specie</th>
<th>Savanna</th>
<th>Tropical forest</th>
<th>Boreal forest</th>
<th>Agriculture</th>
<th>Peat</th>
<th>Reference</th>
<th>Global emissions (TgCyear(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>0.20</td>
<td>0.39</td>
<td>1.11</td>
<td>0.15</td>
<td>1.21</td>
<td>Akagi et al. (2011)</td>
<td>1.49</td>
</tr>
<tr>
<td>Toluene</td>
<td>0.08</td>
<td>0.26</td>
<td>0.48</td>
<td>0.19</td>
<td>2.46</td>
<td>Akagi et al. (2011)</td>
<td>0.91</td>
</tr>
<tr>
<td>Xylenes</td>
<td>0.05</td>
<td>0.11</td>
<td>0.18</td>
<td>0.01</td>
<td>0.00</td>
<td>Andreae and Merlet (2001), Akagi et al. (2011)</td>
<td>0.29</td>
</tr>
<tr>
<td>Phenol</td>
<td>0.52</td>
<td>0.45</td>
<td>0.48</td>
<td>0.52</td>
<td>4.36</td>
<td>Akagi et al. (2011)</td>
<td>1.86</td>
</tr>
<tr>
<td>Styrene</td>
<td>0.02</td>
<td>0.03</td>
<td>0.13</td>
<td>0.03</td>
<td>0.00</td>
<td>Andreae and Merlet (2001)</td>
<td>0.14</td>
</tr>
<tr>
<td>Ethylbenzene</td>
<td>0.13</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
<td>0.00</td>
<td>Andreae and Merlet (2001)</td>
<td>0.08</td>
</tr>
<tr>
<td>Trimethylbenzenes</td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>Yokelson et al. (2013)</td>
<td>0.05</td>
</tr>
<tr>
<td>Benzaldehyde</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>Andreae and Merlet (2001), Yokelson et al. (2013)</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Figure 1. Mixing ratios from box model simulation. In red, mixing ratios from the mechanism used in this work. In black, mixing ratios from the same mechanism without the updated photolysis rates for benzaldehyde and the new photolysis channels for nitrophenols.

degradation scheme introduced here allows for the introduction of new features, for instance the photolysis of nitrophenols and updated rate constants. In contrast, CRI and MOZART-4 are less explicit and more difficult to extend because of the high number of lumped species that they contain.

In order to better understand which are the atmospheric implications for the global simulations due to the developed mechanism, we run two simulations with the CAABA/MECCA box model (Sander et al., 2011). The simulations are representative for the summer period in urban equatorial areas and include emissions for benzene, toluene, phenol and NO (the details of the setup can be found in the Supplement). Figure 1 shows the differences in mixing ratios between two box model simulations: one uses the mechanism employed in the global simulations (i.e. the MCM mechanism reduced and updated as explained in this section). The second simulation uses the same mechanism without any of the photolysis updates mentioned above. Benzaldehyde mixing ratios for the updated version of the mechanism are lower than the non-updated version because the latest photolysis rate from IUPAC is faster than the previous one, leading to an increase in the production rate of nitrophenols. Despite this increase in the production, nitrophenols are almost depleted because the new photolysis channel is included, revealing the strong influence of the photolysis channel as a sink. HONO is enhanced in the updated version due to formation via photolysis of nitrophenols. Consequently, HO\(_x\) (OH + HO\(_2\)) production has also increased due to HONO recycling and OH formation.

Atmospheric oxidation of some aromatic compounds can result in a major production (or sink) of other (aromatic) compounds. For instance, when benzene reacts with OH, more than 50 % is transformed to phenol. Due to the large amount of benzene that is released into the atmosphere, up to 3 TgCyear\(^{-1}\) of phenol is produced by this pathway, which represents more than 50 % of the global phenol source (see Table 5). Additionally, benzaldehyde is produced from several oxidation pathways, which together constitute more than 50 % of the total benzaldehyde source.

2.3 Scavenging and dry deposition

Because of the hydrophobic nature of aromatic hydrocarbons, scavenging is a minor sink in the atmosphere. However, most oxidation products of aromatics have a strong hydrophilic character. Therefore, wet deposition is a removal process of minor importance for aromatics but essential for its oxidation products.

In the model, dry deposition velocities are calculated using an algorithm based on the big leaf approach (Wesely, 1989; Ganzeveld and Lelieveld, 1995). To account for scavenging, the aqueous and gas-phase chemistry is coupled with physical processes related to clouds and precipitation, which represents the wet deposition (Tost et al., 2006).

Scavenging and dry deposition are calculated in the MESSy submodels SCAV (Tost et al., 2006) and DDEP.
from monitoring stations. Table 1 summarizes the locations and periods of the different campaigns.

Observations include data from the following:

- **CARIBIC**: the CARIBIC project (Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container) is a long-term monitoring program, based on atmospheric measurements on board of a passenger aircraft (Lufthansa A340-600) (Brenninkmeijer et al., 2007; Baker et al., 2010). Cruising altitudes are 10–12 km and, on average, 50% in the upper troposphere and 50% in the lowermost stratosphere. The data span from 2005 to 2012. CARIBIC flights take off from Frankfurt (Germany) on routes to India, East Asia, South America and North America.

- **EMEP**: the European Monitoring and Evaluation Programme (EMEP) (Tørseth et al., 2012) is a network of monitoring sites over Europe and includes measurements of a wide number of species. One of its principal aims is to quantify the long-range transmission of air pollutants, and their fluxes across boundaries. The Chemical Coordinating Centre at NILU (Norwegian Institute for Air Research) is responsible for the data harmonisation after the data have successfully passed quality controls. EMEP sites are located in such a way that local influences are minimal, and consequently the observations are representative of large regional areas.

- **EEA**: data provided by the European Environmental Agency (EEA) are based on the public air quality database AirBase (EEA, 2014). EEA gathers information from a large network of monitoring stations in urban, semi-urban and background areas. However, only rural background stations are used for the comparison because the simulation horizontal scale is not representative for traffic or industrial influenced stations. Moreover, for the comparison, annual averages for each station have been used. We selected observations from the year 2005. However, the number of stations that is feasible for this study is small.

- **Literature**: A compilation of measurements from the literature, covering multiple parts of the globe in multiple campaigns, is summarized in Table 1. The table provides detailed information on the location and the time span of the observational dataset. The data cover the years 1995–2012 for 111 surface sites located in rural, semi-urban and urban areas. Each observation represents a different period, ranging from months to years (e.g., Barletta et al., 2005), which can be a source of error in the comparison.

## 4 Evaluation with observations

In this section the model results for benzene, toluene and xylene are evaluated by comparison with observations. Comparisons for other aromatic compounds cannot be made due to lack of a consistent set of global measurements. The full set of figures can be found in the Supplement.

Model results for the year 2005 were chosen for the comparison with observations, assuming that interannual variability is not a significant source of error and that emissions of aromatics were rather constant over the period 1995–2010 for the RCP dataset with a relative increase of 3%.

Table 4 summarizes the statistics of the model–measurement comparison for the three species mentioned above and for the monitoring networks described in Sect. 3. To calculate the statistics, model results were sampled within the geographical locations of the observations.

We follow the criteria of Barna and Lamb (2000) and Pozzer et al. (2012) for the analysis of the model performance. A ratio of RMS (root mean square error) to SD (standard deviation of observations) below 1 is taken as the criterion to establish good modelling quality. In general, the ratio RMS/SD gives a better agreement for the RCP simulation than for the LIT simulation, but both simulations give ratios close to one, meaning relatively good agreement. As a final note, comparisons with station observations must be taken cautiously, as they could be influenced by local emissions and consequently not fully representative for background air, which would be best suited for comparison with large-scale models as the one used in this work.

### 4.1 Benzene

**EEA**: this set of 22 stations with observations for 2005 shows annually averaged mixing ratios of 194 pmol mol$^{-1}$ (see Table 4). In general, the model underestimates observations by a factor of 3.1 in the LIT simulation and by 1.7 in the RCP simulation. As expected due to the coarse model resolution, the simulated spatial variability of simulations (with standard deviations of 15 and 29 pmol mol$^{-1}$ for LIT and RCP, respectively) is lower than that of the observations (118 pmol mol$^{-1}$). The RMS shows better agreement for the RCP than for the LIT simulation. In addition, RCP and LIT show good spatial agreement for all stations, except for one station in central Europe (figures can be found in the Supplement). In conclusion, this comparison suggests that emissions from the RCP scenario give better results in Europe.
Table 4. Summary of the statistical comparison of observed and simulated mixing ratios. Arithmetic means and standard deviations are shown in pmol mol$^{-1}$. MLIT, MRCP, and MObs represent the mean values for the LIT simulation, the RCP simulation, and the observations, respectively. SLIT, SRCP, and SObs are the standard deviations of the previously mentioned cases.

<table>
<thead>
<tr>
<th>Species</th>
<th>Network</th>
<th>Number of locations</th>
<th>Time resolution</th>
<th>MLIT</th>
<th>SLIT</th>
<th>MRCP</th>
<th>MObs</th>
<th>SLIT/MLIT</th>
<th>MRCP/MLIT</th>
<th>MObs/MLIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>CARIBIC</td>
<td>1241</td>
<td>Instantaneous</td>
<td>6.6</td>
<td>2.1</td>
<td>15.8</td>
<td>18.0</td>
<td>0.41</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>EEA</td>
<td>22</td>
<td>Monthly</td>
<td>62.2</td>
<td>14.9</td>
<td>117.1</td>
<td>118.4</td>
<td>0.32</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>EMEP</td>
<td>14</td>
<td>Monthly</td>
<td>50.1</td>
<td>22.4</td>
<td>95.3</td>
<td>104.6</td>
<td>0.56</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Toluene</td>
<td>CARIBIC</td>
<td>789</td>
<td>Instantaneous</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
<td>0.6</td>
<td>0.19</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>EEA</td>
<td>6</td>
<td>Annual mean</td>
<td>118.5</td>
<td>65.0</td>
<td>144.0</td>
<td>144.4</td>
<td>0.49</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>EMEP</td>
<td>12</td>
<td>Monthly</td>
<td>91.8</td>
<td>37.0</td>
<td>111.5</td>
<td>104.7</td>
<td>0.69</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Xylenes</td>
<td>LITERATURE</td>
<td>105</td>
<td>Campaign dependent</td>
<td>290.3</td>
<td>200.3</td>
<td>321.2</td>
<td>245.4</td>
<td>0.34</td>
<td>1.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 2. EMEP observations of toluene for six different locations for the year 2005 (monthly average) (in black) and the simulated toluene mixing ratios for the RCP simulation (in red), both in pmol mol$^{-1}$. Error bars show standard deviation of the observations and red dashes show the standard deviation of the model simulation.

**EMEP:** this dataset has a daily resolution for 14 stations located in Europe. In this work only monthly values estimated from the database are used. In Fig. 2, the RCP simulation results for benzene are compared with observations from six stations. It can be noticed that mixing ratios are better captured by the model in summer than in winter, which is a feature that has been previously observed in EMAC for other simpler VOCs (Pozzer et al., 2007). The RCP simulation yields an amplitude of the annual cycle that is closer to the observations than that of the LIT simulation (see Table 4). In addition, the observations show larger standard deviations than the simulations, with the ratio between the observed standard deviation and the RCP standard deviation being 1.8. The ratio RMS(RCP)/SD(OBS) is above 1, but the temporal correlation between the observations and the RCP scenario is very good (above 0.8 in most cases; Fig. 3). This supports the good representation of the observations by the RCP simulation. On the other hand, the RCP simulation underestimates the annual average systematically (44 %) compared to the EMEP dataset, which is consistent with underestimation compared to the EEA dataset (40 %).

**Literature:** statistics cannot be calculated for these data, because in general the measurements were not performed in...
Figure 3. Top: annually averaged background mixing ratios of benzene for the RCP simulation, with the squares depicting annually averaged EMEP observations, both in pmol mol\(^{-1}\). Bottom: temporal correlation between observations and simulations for benzene.

Observations for 28 US cities (Baker et al., 2008) were taken during summer months in background locations (thus excluding New York, Philadelphia and Salt Lake City). Chicago and Detroit are strong industrialised cities, and therefore the model is biased low in both simulations. Despite the low resolution of the model, in the rest of US cities the simulation is clearly able to capture spatial gradients towards the urban areas. When comparing with the RCP simulation for the month of July, we find an underestimation of 48\% on average. In China, benzene was measured in different areas (mainly residential, commercial, and industrial) in 48 cities during the winter season (Barletta et al., 2005). Both simulations reproduce the observed spatial gradient well, but strongly underestimate the mixing ratios, probably because the instruments were located close to sources. In general the RCP simulation is closer to the observations. At six different locations in the Sahara desert region, observations show mixing ratios on the order of tens of pmol mol\(^{-1}\) for the winter months in background remote locations (Yassaa et al., 2011). The RCP simulation consistently reproduces those mixing ratios during the winter. Compared to the observations at a regional background station in South Africa (Jaars et al., 2014), the model shows again a low bias (88\% lower than the observations) but within a reasonable range. The model is able to represent the mixing ratios peak in Rio de Janeiro (Brazil) (Martins et al., 2007).

**CARIBIC**: model results have been sampled along the flight tracks, and observations within the 200–300 hPa levels are compared to the annual mean of the simulation in the 250 hPa level. The LIT simulation shows a stronger underestimation than RCP, with the RCP simulation underestimating tropospheric benzene mixing ratios by 20\% and the LIT simulation by more than 100\%. For the RCP scenario, the underestimations appear to be lower in the free troposphere than at the surface (RCP simulation underestimates observations of EMEP by 34\% and EEA by 35\%). Despite the large annual variability in benzene mixing ratios, the model is able to capture the gradients along the Africa and Europe-Brasil paths. For the North America-Europe-Asia tracks, the high variability of the measurements makes it hard to compare them with the simulations. In general, the model shows smaller spatial variability than the measurements and an RMS / SD ratio slightly above one, as for the EEA and Literature data.

4.2 Toluene

**EEA**: toluene mixing ratios for the European stations used in this study show an average annual value of 240 pmol mol\(^{-1}\). Compared to observations, model results from the LIT simulation show an underestimation of 51\%, while the model results from the RCP simulation have a low bias of 40\%. Contrary to the case of benzene, simulated mixing ratios have a larger standard deviation than observations, meaning larger simulated spatial variability.

**EMEP**: similar to the comparison with the EEA database, model results from the LIT simulation are closer to the EMEP observations than those from the RCP simulation, with annual average underestimations of 31 and 16\%, respectively. Additionally, the RMS / STD ratio is for both simulations lower than in the comparison with the EEA database, which means that there is a better spatio-temporal correlation. The temporal correlation for some stations is weaker than for benzene (see Supplement). Model results from both simulations reproduce the annual cycle; in the case of the RCP simulation, the agreement is in general higher (see RCP case in Fig. 5). As opposed to benzene, the variability of the model results from the RCP simulation is 33\% lower than that of the observations.

**Literature**: generally, the performance of the model for the observations from the literature of toluene is similar to the one of benzene, and spatial gradients and large urban areas are correctly simulated. Nevertheless, for the same reasons as for benzene, the model is showing a general underestimation for both simulations; RCP results are biased low by more than 3 times for benzene and almost 8 times for toluene. Both
simulations capture the spatial distributions reasonably well, compared to the observations during summer for the US and during winter for the Sahara and China. The stronger discrepancies for toluene compared to benzene can be explained by the short lifetime of toluene in combination with the short distance from sources of the observation sites. In addition, we also compared the toluene to benzene ratios for the RCP scenario with the Literature observations. We found that the model captures well the areas where the observed ratios are higher or lower than one (Supplement). The agreement is in general good in America, Europe and East Asia regions, although the ratios are underestimated.

**CARIBIC**: in contrast to benzene, toluene annual average mixing ratios are underestimated greatly in both simulations, by more than 4 times (3.6 pmol mol$^{-1}$ for observations and 0.8 pmol mol$^{-1}$ for the RCP scenario). In contrast to the surface observations, the RCP simulation is closer to observations in the troposphere than the LIT simulation. Spatial gradients are best captured in the European-African and European-Asian tracks. Geographical variability is larger than for benzene, due to the shorter lifetime of toluene. In the simulations toluene is almost depleted above the planetary boundary layer, which suggests too strong model sinks in those regions. However, as pointed out by Helsel (1990), the underestimation due to the large number of measurements under the instrumental detection limit (1 pmol mol$^{-1}$) is a source of error, since it artificially causes too high values in the observations. In this case, 46% of the CARIBIC observations for toluene is below detection limit, which partially explains the bias. We only use the other 54% of the data for the calculations in Table 4. As for benzene, the ratio RMS / STD is somewhat above one for both simulations.

**4.3 Xylenes**

**EEA**: due to the low number of stations available for this dataset (only 2), the results may be not representative and therefore we did not include them in Table 4. However, for the two stations, mixing ratios from the LIT and RCP simu-
Figure 6. Mixing ratios of xylenes (pmol mol\(^{-1}\)) from EMEP observations in the year 2005 (monthly average) (in black), and from the LIT simulation (in red). Error bars show the standard deviation of the observations, red dashes show the standard deviation of the model simulation.

Simulations are 66 and 100\% higher than the observations, respectively.

**EMEP**: a comparison with model results of this set of eight stations shows a similar result as the comparison for toluene. Figure 6 shows observations and model results from the LIT simulation. Results from both simulations are poorly correlated with observations in terms of time and space (see Supplement). Model results from the LIT simulation are closer to the measurements than those from RCP, but in both cases the RMS / STD ratio is relatively high, which points at a low consistency in reproducing spatio-temporal features. EMEP observations are overestimated by 31 and 79\% by results from the LIT and RCP simulations, respectively.

**Literature**: observations of xylenes are only available for the US and for one location in China. As for other species, xylenes are well represented in the US, with the exception of some cities. In China, the model reproduces the increase in mixing ratios towards the Hong Kong area. In the Southern Hemisphere, the model reproduces the polluted spots in South Africa and Rio de Janeiro. We also compared the ratio of xylenes to benzene. The simulation agrees better for xylenes/benzene ratio than for the toluene/benzene, this can observed clearly observed in US regions (see Supplement).

5 Global budget

Table 5 summarizes the global budget of aromatic compounds for the RCP simulation. This simulation has been selected because it reproduces benzene and toluene observations in terms of annual average mixing ratios, yearly cycle and spatial variations better than the LIT simulation. Moreover, the differences for xylenes are not significant between the two simulations.

For benzene, the total global primary emission of 7.8 TgC year\(^{-1}\) is composed of anthropogenic emissions (81\%), biomass burning emissions (19\%) and chemical production (<1\%). The sources are balanced by the sinks due to OH oxidation (87\%), and dry deposition (13\%). As expected due to the strong hydrophobicity of aromatic compounds, wet deposition has been found to be a negligible process for the budget.

Figure 7 shows modelled annually averaged mixing ratios of benzene, toluene and xylenes. For benzene (upper left panel), the surface mixing ratios are as high as 300–400 pmol mol\(^{-1}\) in highly urbanised and industrialised areas in the Northern Hemisphere (US, Europa, China). Western Asia shows similar mixing ratios, probably due to the large petrol industry. The highest modelled mixing ratios can be found in India and China, due to large anthropogenic emissions. Central Africa and Northern Asia mixing ratios are mainly driven by biomass burning emissions. In general, areas with high mixing ratios of benzene are located close to sources, due to its relatively short lifetime. Over the oceans, mixing ratios vary between 20 and 70 pmol mol\(^{-1}\) (due to ship emissions). In southern hemispheric continental areas, we find mixing ratios in the 100–300 pmol mol\(^{-1}\) range. The highest mixing ratios are found in Africa, due to the strong biomass burning season. Continental background areas show mixing ratios between 10–50 pmol mol\(^{-1}\). Lowest mixing ratios (1–5 pmol mol\(^{-1}\)) are found in oceanic areas due to the far distances to sources.

Figure 7 (upper right panel) shows the modelled annual zonal mean benzene mixing ratios. Note the strong north–south gradient and the averaged mixing ratios of 60–100 pmol mol\(^{-1}\) for the free troposphere. The highest values are found at the surface in the Northern Hemisphere.

Toluene emissions sum up to 7.9 TgC year\(^{-1}\), a number similar to that of benzene emissions. Anthropogenic emissions (84\% of the total) play a larger role for this compound than the biomass burning emissions (11\% of the total). Additionally, the model estimates 4\% of emissions from biogenic sources. Sinks are dominated by OH oxidation (85\%), and the remaining 15\% is removed by dry deposition.

Mixing ratios at the surface are in the order of 20–200 pmol mol\(^{-1}\) in continental areas (Fig. 7, middle left
Table 5. Global atmospheric budget of aromatic compounds from the RCP simulation. Units are TgC year\(^{-1}\) in all cases, unless noted otherwise.

<table>
<thead>
<tr>
<th>Source (Tg year(^{-1}))</th>
<th>Benzene</th>
<th>Toluene</th>
<th>Xylene</th>
<th>Phenol</th>
<th>Styrene</th>
<th>Ethylbenzene</th>
<th>TMB</th>
<th>Benzaldehyde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass burning</td>
<td>1.5</td>
<td>0.9</td>
<td>0.3</td>
<td>1.9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Anthropogenic</td>
<td>6.3</td>
<td>6.7</td>
<td>5.7</td>
<td>0.5</td>
<td>0.7</td>
<td>1.4</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Biogenic</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical production</td>
<td>0.01</td>
<td></td>
<td></td>
<td>3.4</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sources</td>
<td>7.8</td>
<td>7.9</td>
<td>5.9</td>
<td>5.8</td>
<td>0.8</td>
<td>1.4</td>
<td>2.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sink (Tg year(^{-1}))</th>
<th>Benzene</th>
<th>Toluene</th>
<th>Xylene</th>
<th>Phenol</th>
<th>Styrene</th>
<th>Ethylbenzene</th>
<th>TMB</th>
<th>Benzaldehyde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry deposition</td>
<td>1.4</td>
<td>1.1</td>
<td>0.6</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Oxidation by OH</td>
<td>6.4</td>
<td>6.7</td>
<td>5.1</td>
<td>3.2</td>
<td>0.2</td>
<td>0.6</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Oxidation by NO(_3)</td>
<td>0.1</td>
<td></td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Photolysis</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>Total sinks</td>
<td>8.3</td>
<td>7.8</td>
<td>5.8</td>
<td>5.2</td>
<td>0.7</td>
<td>0.8</td>
<td>1.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

| Burden (GgC)             | 213     | 57      | 20     | 2      | 0       | 5            | 3   | 2           |
| Lifetime (days)          | 8.3     | 1.0     | 0.5    | 0.1    | 0.1     | 1.2          | 0.2 | 0.1         |

panel), which are larger than for benzene for specific urban regions urban (Europe), due to large anthropogenic emissions. However, in the free troposphere we find low mixing ratios (a few pmol mol\(^{-1}\)), due to the short lifetime of toluene. Background areas (oceans, deserts) show surface mixing ratios below pmol mol\(^{-1}\) levels for the same reason.

More than 95% of xylenes are emitted by anthropogenic sources. Total emissions sum up 5.9 TgC year\(^{-1}\), from which 88% is removed by OH, 11% by dry deposition and the remainder (<1%) by reaction with NO\(_3\). Very low mixing ratios are present at the surface in the Southern Hemisphere (Fig. 7, bottom left panel), except for a few specific locations (i.e. Indonesia, Nigeria, São Paulo in Brazil). In the free troposphere, mixing ratios are below pmol mol\(^{-1}\) levels in the Southern Hemisphere, even in the lowermost levels.

Phenol has a different distribution of sources compared to other aromatics. The main is source is the atmospheric oxidation of benzene with OH, which produces 3.4 TgC year\(^{-1}\) (59% of total sources). The second important source of phenol is the primary emission from biomass burning, which represents 32% of the total emissions. Anthropogenic emissions are only 9% of the total. Nevertheless, mixing ratios of phenol in the atmosphere are low because of its short lifetime.

Emissions of benzaldehyde, styrene and trimethylbenzenes sum up to 3.2 TgC year\(^{-1}\). Their spatial patterns resemble those of toluene, with most emissions, and hence higher mixing ratios, located in the Northern Hemisphere. Nevertheless, their mixing ratios are almost 1 order of magnitude lower than those of toluene at the surface and in the free troposphere. For this reason, they have only been measured in few campaigns (Baker et al., 2008; Yurdakul et al., 2013).

Consistent with their main sink (i.e. reaction with OH), which is weaker during winter (NDJ months) and the main region of their emission (i.e. the Northern Hemisphere), the burden of most species shows a clear annual cycle, with higher mixing ratios in winter than in summer. As an exception, phenol and styrene have an annual cycle with a small amplitude. The specific pattern in their atmospheric burden is caused by their very short lifetimes and the relative high strength of the biomass burning emissions during autumn.

The atmospheric aromatic burden totals 0.3 TgC. Benzene contributes 70% to the total mass, toluene and xylenes 25% and the remaining species 5%.

Estimated lifetimes of aromatics are for most species on the order of a day or less (except for benzene, toluene and ethylbenzene), and can be found in Table 5. Estimated lifetimes are in line with values of the literature Atkinson (2000).

6 Conclusions

The 3-D atmospheric chemistry general circulation model EMAC and an ensemble of airborne and surface observations were used to evaluate our current understanding of the global atmospheric budget of monoaromatic compounds, including benzene, toluene, xylenes, phenol, styrene, ethylbenzene, trimethylbenzenes and benzaldehyde.

We extended the chemical mechanism of MECCA in order to accurately describe the chemical reactions of aromatic compounds. Emissions of simple aromatics were included in the model, considering biomass burning, anthropogenic ac-
Figure 7. The left column shows annually averaged surface mixing ratios (pmol mol$^{-1}$) for the RCP simulation. The right column shows the annual zonal average. Upper plots show benzene, the middle toluene and the bottom plots show xylenes. Figures for other species can be found in the Supplement.

Activities and natural sources. As sinks, wet and dry deposition were included. Simulations with two different sets of anthropogenic emissions were evaluated against observations. The comparison with surface and aircraft observations shows that for benzene, the model seems to underestimate mixing ratios consistently at the surface and in the free troposphere, while the spatial distributions and seasonal cycles are well reproduced. The model captures the spatial variability and averaged mixing ratios at the surface of toluene well, but it does not accurately reproduce the seasonal cycle and considerably underestimates mixing ratios in the free troposphere. This suggests an overestimation of the efficiency of the chemical removal processes, of which the chemical reaction with OH is the most important. The uncertainty of the rate constant for the reaction of toluene + OH is about $5.6 \pm 0.9 \times$
10−12 cm3 molecule−1 s−1 at 298 K, which implies a 30 % error on the chemical sink estimate. Additionally, the relative bias of the observations, due to the large number of observations below the instrumental detection limit of the aircraft measurements, can partially explain the disagreement in the upper troposphere. The model shows large temporal discrepancies for mixing ratios of xylenes, although they remain within an acceptable range. We can conclude that the RCP scenario captures better total amounts of aromatics released to the atmosphere than the LIT. Nevertheless, in both scenarios we observed underestimation of the observations which could indicate an underestimation on the emission ratios.

Because of the low mixing ratios of some species (phenol, styrene, ethylbenzene, trimethylbenzenes and benzaldehydes) in the atmosphere, and the limitations of the present instrumentation, a model-measurement comparison was not possible for all species. Therefore, a wider array of samples would be helpful to assess the model’s accuracy in remote regions and constrain the respective global atmospheric budgets.

The budget of aromatic compounds is characterized by a total emission rate of 32 TgC year−1. For most species, with the exception of phenol, anthropogenic emissions are the main source. Large emissions are located in industrialised and heavily populated areas, such as Asia and Europe. Emissions from biomass burning play a secondary role on the global scale, although they can be the strongest source of aromatics in specific areas such as Central Africa, South America and boreal areas.

The chemical production generates 4.7 TgC year−1 of aromatics (mainly phenol), making them nearly ubiquitous. Biogenic emissions form only a small fraction of the total toluene source (4 %), although other studies suggest that this fraction could be larger (Sindelarova et al., 2014). Photochemical reaction with OH is the most important removal process of aromatics from the atmosphere, followed by dry deposition. As an exception, styrene and benzaldehyde also react with O3 and NO3, respectively, as their primary sink.

Further studies focused radical production, ozone formation (Bloss et al., 2005a; Nehr et al., 2014) and general impact of aromatics on atmospheric chemistry will be performed based on the mechanism developed in this study.

7 Data availability


The Supplement related to this article is available online at doi:10.5194/acp-16-6931-2016-supplement.

Acknowledgements. We want to thank Ruud Janssen for useful discussion. We also want to thank Angela Baker for providing the CARIBIC data. The authors wish also to acknowledge the use of the Ferret program for analysis and graphics in this paper. Ferret is a product of NOAA’s Pacific Marine Environmental Laboratory (information is available at http://www.ferret.noaa.gov).

The article processing charges for this open-access publication were covered by the Max Planck Society.

References


D. Cabrera-Perez et al.: Global budget of aromatic compounds in the atmosphere 6943


Acknowledgements. We want to thank Ruud Janssen for useful discussion. We also want to thank Angela Baker for providing the CARIBIC data. The authors wish also to acknowledge the use of the Ferret program for analysis and graphics in this paper. Ferret is a product of NOAA’s Pacific Marine Environmental Laboratory (information is available at http://www.ferret.noaa.gov).

The article processing charges for this open-access publication were covered by the Max Planck Society.

References


