Response to Anonymous Referees
Impact of Urban Emissions on a Biogenic Environment during the wet season: Explicit Modeling of the Manaus Plume Organic Chemistry with GECKO-A

Camille Mouchel-Vallon et al.

We thank both reviewers for their helpful comments. Below is our detailed answer to their specific comments, followed by the marked up manuscript.

Major Updates

- During the review process, we discovered an issue with the GECKO-A generated mechanism that cut off oxidation too early. As a result, later generations species were not accounted for in the explicit mechanism. A new mechanism has been generated after fixing this issue and was used to produce new model simulations. Although the contribution of missing species increased SOA mass by 50% in the clean scenario, the predicted values remain in the range of observed SOA mass at the T3 site, and our conclusions remain unchanged.

- In addition, as we explain in the answers to comment 8, we updated the anthropogenic emissions with emissions from the local power plants and oil refinery. The urban emissions of SO_2 and NO_x were significantly increased and this has an impact on the modeled urban plume SOA. The tables, figures and text have been updated in the paper to reflect changes due to updated emissions. The main effect is that SOA mass in the modeled plume is slightly lower than SOA mass in the background. This can be explained by the increase in OH and NO_x which favors the fragmentation of organic compounds, reducing the amount of material available to condense to form SOA. This somewhat modifies our conclusions as now we cannot say anymore that GECKO-A is capturing an SOA increase in the plume. The comparison with the VBS parameterization becomes even more crucial to investigate which processes may be missing to explain this. With the updated emissions, the VBS still captures the observed SOA mass increase and we are still able to show the importance of adding aging to SOA processes. We can still conclude that modeling SOA in the urban plume requires including these aging processes accounted for in the VBS parameterization.

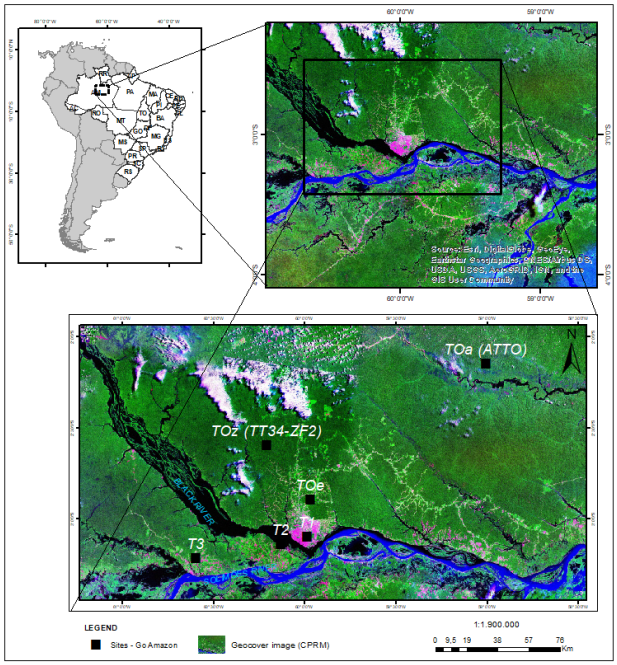
- Taking into account Comment 10, we propose a new title for the paper to better reflect its content:

> **Title:** Exploration of Oxidative Chemistry and Secondary Organic Aerosol Formation in the Amazon during the Wet Season: Explicit Modeling of the Manaus Urban Plume with GECKO-A

Response to Anonymous Referee 1

Comment 1: *Figure 1 - some readers with colorblindness may not be able to distinguish the shades of red, pink, and green used in the maps.*

Answer 1: We did not account for colorblindness when making this map. After checking it on an online colorblindness simulator (https://www.color-blindness.com/coblis-color-blindness-simulator/), we decided to change the color of the markers from red to black to give better contrast for colorblind people.



Comment 2: *Section 3.1.1 and Figure 2 - Line 112-113 state "The top box extends from the top of the planetary boundary layer to 1.5 km and represents the residual layer (RL). However the black dashed line in Fig 2 makes it look like the top of the second box is at* $\sim$ 900*m.*

Answer 2: This was a mistake in the text. It has now been fixed to read:

> **Sect. 3.1.1:** [. . .] The top box extends from the top of the planetary boundary layer to 850 m and represents the residual layer (RL) [. . .]

Comment 3: *First paragraph of 3.2.2 - the vehicle patterns and emissions in Manaus are assumed to be similar as in Sao Paolo. However, as noted in line 163-164, the fuel used in Manaus is very different than in Sao Paolo. There are many papers*

on how ethanol blends impact vehicle emissions; the authors should at least acknowledge this literature and comment on how their assumptions about vehicle emissions might impact their results.

Answer 3: We indeed assume that the VOC speciation of traffic emissions in Manaus is similar to that in São Paulo. However our emissions are scaled to the total emissions that were measured in Manaus and thus errors in emissions may come from a different distribution of vehicle types and different fuel blends between the two cities. This has an impact on the nature of emitted VOCs, which in turn would impact the SOA formation potential and oxidants lifetimes in the plume. This is now now acknowledge in the manuscript:

> **Sect. 3.2.2:** [...] The difference in the fuel blend used in São Paulo and Manaus can introduce errors in the traffic emissions VOC speciation. For instance, a recent study by Yang et al. (2019) showed that the combustion of fuels with higher ethanol content emits significantly less carbon monoxide and more acetaldehyde. Schifter et al. (2020) showed similar results, and also suggested that ethanol blends emit smaller amounts of simple aromatic compounds (*e.g.* benzene, toluene). This speciation uncertainty can especially have an impact on oxidants concentrations. Schifter et al. (2020) reported for instance that fuels containing ethanol would potentially produce less ozone after the oxidation of emitted organic species than fuels without ethanol. Moreover, the lifetime of OH is likely to change depending on the speciation of emitted VOCs due to varying reactivities with respect to OH. In the same way that the potential for ozone formation could depend on the use of ethanol fuel blends, it is also possible that the potential for SOA formation would depend on these fuel blends too. [...]

Comment 4: *Fig 5 - the biogenics in panel a and b show 5 hours where the box is over T3, but the anthropogenic panels c and d only show 4 hours over T3. Why is there this difference?*

Answer 4: As mentioned in Sect. 2, we separated T3 measurements between clean and polluted situations using the classification of T3 measurements established by de Sá et al. (2018). After hourly averaging this data, it appears that some data was positively identified as clean after 5 pm, whereas no air mass could be identified as polluted in that same time period.

Comment 5: *Figure 9 should be discussed more. The compositional differences between the clean and polluted cases seem very small. Is this consistent with AMS data?*

Answer 5: We agree this figure could be discussed more. The AMS atomic ratios data displayed on Fig. 10 also show very little difference between the clean and polluted cases. The following was added to the results discussion:

Sect. 4.2.3: [...] The change in overall modeled SOA composition between clean and polluted cases is quite small. AMS measurements give a similar impression of a small impact of polluted situations on atomic ratios (Fig. 10), with only a slight increase of O/C ratio (see Sect. 4.2.4). Other analyses of airborne and ground AMS data (de Sá et al., 2018; Shilling et al., 2018) similarly show that the relative contribution of hydrocarbon-like organic aerosol (HOA) slightly increases in the polluted plume at the expense of isoprene derived SOA. The model and the AMS data support the idea that the impact of anthropogenic emissions is mostly seen on the total organic aerosol mass, and that all constituents of the organic aerosol phase increase approximately in the same way.

Comment 6: *I'm unclear on what is plotted in Fig 12b. Is this the overall species diversity from the model? Or is it the number of species needed in the reduced model to reproduce 90% of the diversity from the full GECKO model run?*

Comment 7: *I'm also confused about Fig 12a. Was the number of species determined by adding up the number of species needed to capture 90% of the SOA mass (i.e., simply doing amass balance)? Or was the model run in some sort of reduced form?*

Answer 7: We grouped these two comments as they are related.

To be clear, no reduced version of GECKO-A was developed for this work. Figure 12a depicts the number of species needed to capture 90% of the SOA mass at each timestep in the explicit model. It is made by adding up species concentrations (in decreasing order) until we reach 90% of the total modeled SOA mass.

Independantly of the calculation shown on Fig. 12a, we also calculate the statistical diversity at each timestep in the explicit model, and this is depicted on Fig. 12b.

The text in Sect. 4.4 as well as the figure caption were rephrased to explain this better:

Sect. 4.4: [...] To answer this, two metrics are presented in Fig. 12. The first one $N_{90\%}$ is the smallest number of species needed in the explicit model to capture 90% of the total SOA mass at each timestep. After sorting species by decreasing concentration, this number is calculated by adding up these concentrations until 90% of the total modeled SOA mass is reached. The operation is repeated at each timestep. Calculated independantly, the second one is the particle diversity D in the explicit modeled SOA, as defined for instance in Riemer and West (2013): [...]

Figure 12: Smallest number of species needed to capture 90% of modeled SOA mass (left panel) with GECKO-A at each timestep ($N_{90\%}$, see text) and statistical diversity D in the GECKO-A modeled particle phase (right panel, see Eq. 3).

Comment 8: *The focus on this work claims to be on anthropogenic influence,but one major result is the relative lack of skill GECKO-A has in capturing the impact of the Manaus plume on SOA production. This suggests to me that a major result might actually be just that the box model is poorly including anthropogenic emissions. SOA formation is underpredicted, OH concentrations are significantly over predicted,so perhaps there are just more or more reactive VOCs in the city than being included in the model.*

Answer 8: We agree that the disagreement between the model and measurements could result from poorly constrained anthropogenic emissions. In the course of the review process, we noticed that we were missing a significant source of anthropogenic emissions in the model, namely the emissions from the Thermal Power Plants (TPP) that provide electricity to the Manaus area and an oil refinery. Using data from Medeiros et al. (2017), we included CO, NO_x and SO_2 additional emissions from these sources. The impact of these sources on particulate matter emissions was already intrinsically accounted for because we constrain these from the field measurements. Emissions of semi-volatile organic compounds from these sources were not included as there is not available speciation for these emissions. According to Abou Rafee et al. (2017), VOC total emissions from the power plants and the refinery are estimated to 1.6×10^3 tons yr^{-1} while total VOC emissions from traffic are estimated to 2.4×10^4 tons yr^{-1}. We therefore can neglect them in a first approach as they are an order of magnitude lower than traffic emissions.

We updated Fig. 4a to include these emissions and added the following description in the text:

> **Sect. 3.2.2:** [...] Additionally, emissions from 11 local thermal power plants (TPP) and one oil refinery located in the vicinity of Manaus were obtained from the data presented in Medeiros et al. (2017). Based on monthly statistics of fuel use in each of the TPP and the oil refinery, combined with emission factors of CO and NO_x for each type of fuel (diesel, fuel oil, natural gas), we calculated CO and NO_x emissions for February, March and April 2014. These total emissions were then averaged over the whole surface area of Manaus (377 km^2, Abou Rafee et al., 2017). Total SO2 emissions were taken from Abou Rafee et al. (2017) and added to the urban emissions for the considered Manaus area.

We ran the simulations with these new emissions. Their main impact is to increase OH and NOx concentrations in the plume. The impact of urban emissions on SOA mass is weaker with the new emissions. This is probably due to the combination of (i) higher NOx mixing ratios, which reduce biogenic SOA through enhanced fragmentation of alkoxy radicals in the gas phase, and (ii) the contribution of anthropogenic compounds. We have updated the corresponding figures and numbers to account for the new simulations, however our conclusions remain unchainged.

This shows the sensitivity of the model to urban emissions, but after our fix our emissions are quite similar to those used in other modeling studies (Abou Rafee et al., 2017; Shrivastava et al., 2019).

Comment 9: *This is further exacerbated by the issue that the surrogate composition of the fuel emissions are known to not really accurately represent the true composition. The ability of the model to capture clean conditions, which with MEGAN and*

the PTR have reasonably well constrained emissions, speaks well of GECKO-A, and suggests just that the issue may be in the emissions being fed to the model.

Answer 9: It is also true that the fuel composition is not accurately known, especially for IVOCs. We address this issue in our answer to Comment 3 from the first anonymous referee.

Comment 10: *This paper is primarily an exploration of the strengths and weakness of GECKO-A,it doesn't do much to discuss the "Impact of Urban Emissions" as the title suggests. There is some discussion, but mostly it is a comparison of GECKO-A to other models (VBS, Shrivastava) and not really providing new information beyond a further exploration of GECKO-A. I think there are interesting results regarding GECKO-A, in particular the discussion of reduced complexity, which seems like a real place for GECKO-A to provide generalizable scientific insight (how complex do we really need models to be?), and that the weaker part of the work is the attempts to understand the Manaus plume (for which the emissions may not be correct, and aging may not be incorporated,and other issues). Potential issues in modeling the plume end up conflated with potential issues in the model, and instead I think perhaps some of the discussions about the model could instead be bolstered and thought about more deeply (e.g., I don't think the explanations for H/C discrepancies are likely complete).*

Answer 10: After noticing the inability of GECKO-A to reproduce the observed urban SOA enhancement, we indeed focused more about finding reasons for this behavior than exploring the modeled impact of Manaus emissions, hence the comparison with the simpler model from Shrivastava et al. (2019). Nevertheless this investigation still gives some reasons to believe that the in-particle aging of organic aerosol is an important part of the interaction of urban emissions with clean biogenic air masses. In our answer to Comment 8, we highlight how we addressed a possible problem with urban emissions. We don't have yet the tools in GECKO-A to provide more information on the specific processes involved in aging of particles, but following this comment we extended the discussion of the H/C discrepancies with additional calculations to estimate the possible impact of dimerization and fragmentation:

Sect. 4.2.4: [...] As a test, we generalized this estimation to all C_{10} in the aerosol phase: we replaced each C_{10} by the corresponding C_{20} and halved its concentration. In this way, we can calculate what would H/C and O/C ratios be in the aerosol phase if aging processes only dimerized C_{10} compounds. The resulting modeled van Krevelen diagram is reported on Fig. 10 (labeled w/ dimer.). The impact of C_{10} dimerization is relatively strong on O/C ratio, ranging from 0.66 to 0.78 and remaining in the range of measured O/C ratios at T3 site and in the aircraft. H/C ratios are only reduced to 1.88–1.94, still 50% higher than measured H/C at the T3 site and 20% higher than airborne data.

[...]

As another test, we also estimated what would O/C and H/C ratios be if all C_{10} fragmented in the aerosol phase. The resulting modeled van Krevelen diagram is reported on Fig. 10 (labeled w/ frag.). In this case, modeled O/C ratios increase to a range of 0.88 to 0.96 and remain in the higher end of measured ratio at the T3 site. H/C are reduced further than in the dimerization test and sit at the higher end of airborne measured H/C ratios, but they still are 45% higher than H/C ratios measured at the T3 site.

Even if they apparently cannot account for the discrepancy between modeled and measured H/C ratios, the two tests presented here on C_{10} compounds in the aerosol phase show the potential importance of adding these missing processes in GECKO-A. These simple tests are however simplifications that overlook important factors in the potential impact on SOA composition: (i) not all C_{10} compounds would be affected by these processes, (ii) other compounds than C_{10} could react in a similar way, (iii) trimerization, tetramerization and other accretion processes could also occur in the aerosol phase, (iv) missing fragmentation processes could also happen in the gas phase.

Comment 11: *Line 1. Missing "of" : "investigation of the"*

Answer 11: Fixed

Comment 12: *Line 7. Not clear what it means for the model to "reproduce measured primary compounds" after tuning emissions. Aren't the primary compounds just the emissions, so you tune for this result, and it isn't really impacted by the skill of the model? Perhaps it will be more clear to me after a detailed reading.*

Answer 12: We agree this mention of primary compounds is of little interest to the abstract. The sentence was modified as follows:

Abstract: [...] The biogenic emissions estimated from existing emission inventories had to be reduced to match measurements. The model is able to reproduce ozone and NO_x for clean and polluted situations. [...]

Comment 13: *Line 15. "particularly intense all year long" is a bit odd, perhaps the authors mean "more photochemically active than other regions throughout most of the year"?*

Answer 13: We agree and used this suggestion in the text.

Comment 14: *Line 36-37. This mid-paragraph question is odd. "Would" in what case? Do the authors mean "Do"? Maybe just rephrase this without the use of this rhetorical question.*

Answer 14: We agree and rephrased the sentence as follows:

> **Introduction:** [...] Several studies have investigated how the biogenic nature of the SOA is affected by anthropogenic influence. [...]

Comment 15: *Line 50. The discussion of the "molecular view," while valuable, has some gaps or issues here. In particular, those citations of Koss et al. are an odd choice, as I believe they are just using a PTR, and if I'm not mistaken, there was a PTR run at T3 by the Martin group (some of the data of which is used in this work). Other groups (e.g.the Kroll group, and the CLOUD group) have tried to combine multiple spectrometers to actually capture the whole range of compounds, reaching more a molecular view. However, even still, this would be more reasonably considered a "formula view", as these instruments do not separate molecules out from their formulas. There were also several other instruments at T3 approaching a molecular view (I believe for instance the Goldstein group has collected and run GCxGC of filter samples, providing some molecular invormation), though perhaps not comprehensively. I would re-frame some of this discussion to more accurately capture the landscape and discuss what existing measurements can or can't provide (e.g., I agree the available instruments probably don't provide a comprehensive measurement to compare to models, particularly for gas-phase oxygenates).*

Answer 15: We agree that we should make the distinction between the "molecular" and "formula" views. We in fact mixed these and called them both "molecular" views. We reworked this part of the introduction to account for this comment, and introduced the term "formula" view:

> **Introduction:** [...] In a recent review, Heald and Kroll (2020) have reported on the recent progress in measurements of individual organic compounds, and how experimentalists are getting close to achieving closure on organic carbon in both gas and aerosol phases (*e.g.* Gentner et al., 2012; Isaacman-Vanwertz et al., 2018). As these measurements are able to capture elemental formulas, double bonds, some oxygenated functional groups and aromaticity (*e.g.* Yuan et al., 2017), they do not provide individual molecular identities. From this point of view, measurements are still restricted to a "formula view". For the GoAmazon field campaign, Yee et al. (2018) were able to sample and identify 30 sesquiterpenes and 40 of their oxidation products at the T3 site with a semi-volatile thermal desorption aerosol gas chromatograph (SV-TAG, Isaacman et al., 2014) but they do not achieve the coverage needed to approach the "molecular view". [...]

Comment 16: *Line 54. I'm a big fan of GECKO-A, but maybe not everyone will agree that it is "the ideal tool". "an excellent tool" perhaps?*

Answer 16: We are big fans of GECKO-A too, and we may have got carried away with our enthusiasm. The text now reads:

Introduction: The Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A, Aumont et al., 2005; Camredon et al., 2007) is an excellent tool to model atmospheric organic chemistry with a detailed molecular view.

Comment 17: *Line 171. The authors discuss the fact that n-alkanes are perhpas not a good surrogate for diesel fuel and gasoline IVOCs, which are mostly branched and cyclic. However,they do not explain why their estimates are less branched than other work has suggested, and importantly they do not disucss the impacts of these structural differences. They acknowledge that almost none of diesel fuel is comprised of the compounds they are using as surrogates, but do not further discuss this issue. Gentner et al. attempt to put estimates on the impact of branching and rings on SOA, so estimating this uncertainty shouldn't be too difficult.*

Answer 17: This IVOC alkane surrogate speciation was first established in Lee-Taylor et al. (2011) to obtain a good volatility distribution, because the identity of individually emitted species was not well known. We kept this simplified IVOCs emissions mostly because of mechanism size issues. Adding heavier molecular weight branched alkanes, cycloalkanes, linear and branched alkenes and aromatics matching the distribution displayed in Gentner et al. (2017) would increase the size of the mechanism by a factor of 5 or 6, to an unmanageable size even by GECKO-A standards. We agree that the possible impact of this simplification should be discussed and we added the following to the description of Manaus emissions:

Sect. 3.2.2: [...] Choosing alkanes as surrogates for emitted IVOCs is also likely to introduce uncertainties to SOA produced from their oxidation. Lim and Ziemann (2009) carried out multiple chamber experiments that investigated the impact of branching and rings on alkanes SOA yields. For instance they showed that SOA yields range from a few percent for branched alkanes with 12 carbon atoms to 80% for cyclododecane while n-dodecane has an SOA yield of $\approx 32\%$. La et al. (2016) simulated these experiments with GECKO-A and they were able to reproduce this experimentally observed behavior. This means that without a detailed inventory of emitted IVOCs, the uncertainty on the SOA yield from IVOCs is high in our version of the model. It should be noted that the range of measured SOA yields for structurally different compounds with the same number of carbon atoms seems to peak for C_{10}-C_{13} alkanes. The range of observed SOA yields in Lim and Ziemann (2009) decreases after this peak. For instance, SOA yields for C_{15} alkanes of various structures range from 45% to 90%. We can therefore expect the IVOCs SOA yield to be highly sensitive to the speciation of compounds ranging from C_{12} to C_{14}, but this sensitivity should decrease for heavier molecular weight species. [...]

Comment 18: *Figure 5. The measurement of benzene and toluene at T3 are fairly poor constraints on these species as they catch only the tail end of the decay (and even still don't really agree with the toluene model). Is there no aircraft data or VOC measurements at T1 or T2 near Manaus to better constrain these?*

Answer 18: There were measurements of benzene and toluene taken in the aircraft, but to our knowledge no data is available for these compounds in the sites in or close to Manaus. Airborne measurements were carried out in the plume and we added the corresponding points on the Fig. 5. The figure caption was updated:

Figure 5: Modeled (lines, second day) time evolution of primary species concentrations in the Lagrangian box-model described in Sect. 3.1, average experimental concentrations measured at the T3 site (dots) and in the airplane (triangles). The vertical range of the experimental data denotes the standard deviation of measured concentrations during events identified as clean (top, blue) and polluted (bottom, orange). The airborne data was measured during plume transects. For each transect, aircraft distance from Manaus was converted to a time separation from Manaus assuming the plume leaves the city at 8am and arrives above T3 at 2pm.

It should be noted that modeled benzene and toluene lines were switched on the original Fig. 5. We fixed that too and updated the text to account for this:

Sect. 4.1: The modeled mixing ratio of benzene matches the measurements, between 0.4 and 0.6 ppb, while modeled toluene is closer to the higher range of measurements, between 0.2 and 0.6 ppb during the afternoon. Figure 5 also displays the airborne measurements of the same anthropogenic compounds. The modeled mixing ratios of benzene and toluene decay in a similar way to the concentrations measured during the plume transects. The modeled peak is not seen by the aircraft measurements as the aircraft may not be flying close enough to the emission sources to capture it.

Comment 19: *Line 260. Again, are there T1 or T2 measurements that could help constrain VOCs in the city? Also, to what extent does model include residual biogenic VOCs present in the city? Presumably there are VOCs present in the city other than just the vehicle emissions, like biogenics from the surrounding forest (or volatile chemical products?), are those captured by the model? Is there a spin up time to allow the city emissions to have some equilibrium concentration of VOCs that would help suppress OH concentrations?*

Answer 19: We are not aware of measurements at T1 or T2 that could have helped constrain VOCs in the city itself. However the comparison to airborne measurements mentionned in Comment 18, provides an additional constraint. This two boxes boxmodel used in this study is designed to simulate an air mass traveling over the rainforest. The bottom box is then exposed to fresh Manaus emissions for 1 h, the approximate time it would take for an air mass traveling at 10 m s^{-1} to cross the urban area. This model design does not require the city chemistry to be at equilibrium before interaction with the box.

After the emission update presented in Comment 8, the two major sources of anthropogenic pollution are accounted for with traffic and power plants emissions. To our knowledge the potential contribution of personal care products VOCs emissinos to anthropogenic emissions has only been evaluated in North America (*e.g.* Coggon et al., 2018; McDonald et al., 2018; Shah et al., 2020). This contribution is likely to become relatively important in the future with the decrease of vehicle emissions in western developed countries, but is not likely to be important in Manaus in 2014 and 2015.

The interaction of biogenics from the surrounding forest with urban emissions is exactly what happens in the model as soon as it is exposed to Manaus emissions: biogenic emissions are replaced with urban emissions for a short time, but the urban emissions become mixed with the remaining background from biogenic chemistry.

Comment 20: *Figure 8. It would be helpful to add Glasius OS measurmenets to the figure as a dashed line.*

Answer 20: D one.

Comment 21: *Line 300-314. I'm not sure the explanations provided for the H/C disparity can really close the gap and it warrants further discussion. The examples the authors provide of oligomerization and fragmentation provide relatively modest decreases in H/C and also change the O/C. A huge fraction of the compounds would need to be oligomers or fragments for this to reconcile H/C, and this would likely shift the O/C. In the examples they provide, dimers still have H/C ratios well above the observed average. If, for instance, all of the C10 compounds were actually dimers, wouldn't this just bring the average down to 1.5 or 1.6? How do you get down to 1.3? What sorts of compounds can push H/C this far down? Trimers? Tetramers? And would the whole mass needto be comprised of these? The authors could play some games with their data to explore this, for instance assume all compounds are actually dimers or fragments and estimate the average H/C and O/C. This might further be scientifically interesting by putting some constraints on accretion products.*

Answer 21: As presented in Comment 10, we introduced the calculation of what would O/C and H/C ratios become after dimerization or fragmentation of C_{10} compounds in the aerosol phase. These processes seem to not be sufficient to bring modeled H/C ratios down to measured values, especially if we don't want to move modeled O/C ratios too far from measured values too.

Comment 22: *Line 316. How do the anthropogenic emissions used by Shrivastava compare to those used here?*

Answer 22: Shrivastava et al. (2019) followed an approach very similar to this work to estimate urban emissions. Their work is also based on combining data from Manaus and São Paulo for traffic and power plants emissions. We completed the description of their simulation to clarify this:

> **Sect. 4.3:** Shrivastava et al. (2019) modeled this same field campaign with WRF-Chem, a chemistry transport regional model (Grell et al., 2005) and similarly to this work they based their primary organic compounds emissions on the MEGAN inventory (Guenther et al., 2012) for biogenic compounds, and combining the methodology described in Andrade et al. (2015) with data from Medeiros et al. (2017) for anthropogenic emissions. [...]

Comment 23: *Line 322. I don't fully understand the aging parameterization, could the authors provide more detail?*

Answer 23: We added the following clarification to the text:

> **Sect. 4.3:** [...] This aging is parameterized as a reaction of each of the SOA surrogate with OH as follows:
>
> $$\mathrm{VBS_n} + \mathrm{OH} \rightarrow \alpha_{\mathrm{frag}}\mathrm{VBS_{n+1}} + (1-\alpha_{\mathrm{frag}})\mathrm{VBS_{n-1}} \quad \text{(R1)}$$
>
> The reaction rate is $k_{R1} = 2 \times 10^{-11}$ cm^3 molec^{-1} s^{-1}. The branching ratio for fragmentation α_{frag} is determined as the ratio of the reaction rate of peroxy radicals with NO to the sum of all peroxy radical reactions rates; it has an upper limit of 75%. [...]

Comment 24: *Line 362. The authors point out the importance of aging in capturing the polluted SOA,which the GECKO-A model does not really capture. Is this due to a lack of aging in the GECKO-A model? Maybe I missed it, but is the GECKO-A model only oxidizing the gas phase and not aging the particle? Considering the importance of aging on reproducing the SOA mass (Figure 7), could the authors include a parameterization of aging in the GECKO model? This might also address the H/C and O/C issues.*

Answer 24: There is no aging parameterization in GECKO-A, oxidation only happens in the gas phase. For technical reasons and lack of resources, we were not able to implement a parameterization of aging in GECKO-A for this paper. In our answer to Comment 10, we estimate the possible impact of dimerization and fragmentation of biogenic condensed species on H/C and O/C but we don't see how to easily estimate the potential impact on SOA mass.

Comment 25: *Line 367-368. Why is it unclear? It seems to capture polluted periods better than biogenic periods (Figure 7).*

Answer 25: The VBS parameterization is capturing both polluted and biogenic episodes relatively well, but in both situations SOA composition is dominated by biogenic oxidation products. In that sense, we don't know if the same parameterization would apply well in environments where the composition of SOA would be for instance dominated by species of anthropogenic or biomass burning origin.

Comment 26: *Line 400-402. The ideas of reducing complexity discussed in this section are very interesting. It's not completely clear to me that some of the conclusions aren't overextended.In particular, the conclusion that "diversity represents the number of species that would be needed to reproduce the same informational content regarding the composition of SOA." The parameter D is based solely on mass fraction, not their physicochemical properties, so it might not capture other properties such as oxygenation or volatility.Imagine a scenario where a small fraction of highly oxygenated compounds drive up O/C - this might still impact hygroscopicity but would this complexity be captured by D? Or a small fraction of more volatile components might partition between the gas-and particle-phase and drive oxidation of particle mass. While the exact mechanism of implementing such a reduction is out of scope of this mansucript, the overall point is that it's not to me that either of the diversity parameters really capture the complexity. Would it be possible to implement a variant of D for parameters of interest? For instance minimum number of components needed to describe O/C within 10% while capture some fraction of the mass? I'm not sure the best parameters, but the idea would be something that captures some of the properties beyond simply mass.*

Answer 26: This is a very interesting take on the reduction issues. We agree with the idea that this diversity approach should be extended to capture other properties than just mass and that we overextended the conclusion about the composition of SOA. We fixed the sentence you mention as follows:

> **Sect. 4.4:** [...] As this number is directly derived from informational entropy, we suggest that the diversity represents the number of species that would be needed to reproduce the same informational content regarding the time evolution of SOA mass in the explicit model. [...]

For this paper, we could not explore further this application of information theory to explicit mechanisms reduction. Hopefully this will be the subject of future work. We added the following to the discussion in that same section, raising the issues mentioned in this comment.

> **Sect. 4.4:** [...] Finally, we used in this section an entropy calculation for SOA mass: it is based only on mass fractions of the species composing the modeled organic particles. The effective number of species displayed on Fig. 12 is therefore only meaningful for SOA mass and properties directly linked to it. If the goal is to predict other properties, *e.g.* hygroscopicity, toxicity or optical properties, assuming we find a way to calculate these with GECKO-A, the diversity defined here would not necessarily be meaningful. For instance, hygroscopicity or toxicity could be driven by a handful of oxygenated species that do not matter for the informational content regarding SOA mass. We did not explore further down this path, as this is not the main subject of this paper, but it may be possible to generalize this definition of informational diversity to properties other than mass.

As an additional note, the idea is not to reduce complexity as written at the beginning of this comment. It is more about evaluating complexity, and showing how many species are needed to effectively produce the same complexity (regarding SOA mass in our example) as the explicit mechanism.

Comment 27: *Line 403-405. Isn't it not only possible but certain that an effective species is not an individual species but rather a combiniation of explicit species? Otherwise D would be more similar to the other metric, or equal to N. The very nature of the parameter D is to be a mathematical descriptor, not an individual species, correct?*

Answer 27: Thanks for this comment. Intuitively your reasoning makes sense and we added our understanding of it to that section (additionally, $N_{90\%}$ has been introduced to denote the "90%" metric):

> **Sect. 4.4:** [...] For instance, in the polluted scenario, D is a factor of 7 lower than $N_{90\%}$. This should mean that D cannot represent a subset of the individual species from the original mechanism, otherwise it would be expected to be equal or higher than $N_{90\%}$ if it is supposed to reproduce the informational content regarding SOA mass. It is therefore likely, and making this problem more complex, that each of these effective species is a (non) linear combination of explicit individual species. [...]

However we are trying to remain very cautious about what we write on the interpretation of this metric as we are not experts in the field of information theory.

References

Abou Rafee, S. A., Martins, L. D., Kawashima, A. B., Almeida, D. S., Morais, M. V. B., Souza, R. V. A., Oliveira, M. B. L., Souza, R. A. F., Medeiros, A. S. S., Urbina, V., Freitas, E. D., Martin, S. T., and Martins, J. A.: Contributions of mobile, stationary and biogenic sources to air pollution in the Amazon rainforest: a numerical study with the WRF-Chem model, Atmospheric Chemistry and Physics, 17, 7977–7995, https://doi.org/10.5194/acp-17-7977-2017, 2017.

Andrade, M. D. F., Ynoue, R. Y., Freitas, E. D., Todesco, E., Vara Vela, A., Ibarra, S., Martins, L. D., Martins, J. A., and Carvalho, V. S. B.: Air quality forecasting system for Southeastern Brazil, Frontiers in Environmental Science, 3, 6975, https://doi.org/10.3389/fenvs.2015.00009, 2015.

Aumont, B., Szopa, S., and Madronich, S.: Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach, Atmospheric Chemistry and Physics, 5, 2497–2517, https://doi.org/10.5194/acp-5-2497-2005, 2005.

Camredon, M., Aumont, B., Lee-Taylor, J., and Madronich, S.: The SOA/VOC/NOx system: an explicit model of secondary organic aerosol formation, Atmospheric Chemistry and Physics, 7, 5599–5610, https://doi.org/10.5194/acp-7-5599-2007, 2007.

Coggon, M. M., McDonald, B. C., Vlasenko, A., Veres, P. R., Bernard, F., Koss, A. R., Yuan, B., Gilman, J. B., Peischl, J., Aikin, K. C., DuRant, J., Warneke, C., Li, S.-M., and de Gouw, J. A.: Diurnal Variability and Emission Pattern of Decamethylcyclopentasiloxane (D 5) from the Application of Personal Care Products in Two North American Cities, Environmental Science & Technology, 52, 5610–5618, https://doi.org/10.1021/acs.est.8b00506, 2018.

de Sá, S. S., Palm, B. B., Campuzano-Jost, P., Day, D. A., Hu, W., Isaacman-VanWertz, G., Yee, L. D., Brito, J., Carbone, S., Ribeiro, I. O., Cirino, G. G., Liu, Y., Thalman, R., Sedlacek, A., Funk, A., Schumacher, C., Shilling, J. E., Schneider, J., Artaxo, P., Goldstein, A. H., Souza, R. A. F., Wang, J., McKinney, K. A., Barbosa, H., Alexander, M. L., Jimenez, J. L., and Martin, S. T.: Urban influence on the concentration and composition of submicron particulate matter in central Amazonia, Atmospheric Chemistry and Physics, 18, 12 185–12 206, https://doi.org/10.5194/acp-18-12185-2018, 2018.

Gentner, D. R., Isaacman, G., Worton, D. R., Chan, A. W. H., Dallmann, T. R., Davis, L., Liu, S., Day, D. A., Russell, L. M., Wilson, K. R., Weber, R., Guha, A., Harley, R. A., and Goldstein, A. H.: Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions, Proceedings of the National Academy of Sciences, 109, 18 318–18 323, https://doi.org/10.1073/pnas.1212272109, 2012.

Gentner, D. R., Jathar, S. H., Gordon, T. D., Bahreini, R., Day, D. A., El Haddad, I., Hayes, P. L., Pieber, S. M., Platt, S. M., De Gouw, J., Goldstein, A. H., Harley, R. A., Jimenez, J. L., Prévôt, A. S., and Robinson, A. L.: Review of Urban Secondary Organic Aerosol Formation from Gasoline and Diesel Motor Vehicle Emissions, https://doi.org/10.1021/acs.est.6b04509, 2017.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions, Geoscientific Model Development, 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Heald, C. L. and Kroll, J. H.: The fuel of atmospheric chemistry: Toward a complete description of reactive organic carbon, https://doi.org/10.1126/sciadv.aay8967, 2020.

Isaacman, G., Kreisberg, N. M., Yee, L. D., Worton, D. R., Chan, A. W., Moss, J. A., Hering, S. V., and Goldstein, A. H.: Online derivatization for hourly measurements of gas- and particle-phase semi-volatile oxygenated organic compounds by thermal desorption aerosol gas chromatography (SV-TAG), Atmospheric Measurement Techniques, 7, 4417–4429, https://doi.org/10.5194/amt-7-4417-2014, 2014.

Isaacman-Vanwertz, G., Massoli, P., O'Brien, R., Lim, C., Franklin, J. P., Moss, J. A., Hunter, J. F., Nowak, J. B., Canagaratna, M. R., Misztal, P. K., Arata, C., Roscioli, J. R., Herndon, S. T., Onasch, T. B., Lambe, A. T., Jayne, J. T., Su, L., Knopf, D. A., Goldstein, A. H., Worsnop, D. R., and Kroll, J. H.: Chemical evolution of atmospheric organic carbon over multiple generations of oxidation, Nature Chemistry, 10, 462–468, https://doi.org/10.1038/s41557-018-0002-2, 2018.

La, Y. S., Camredon, M., Ziemann, P. J., Valorso, R., Matsunaga, A., Lannuque, V., Lee-Taylor, J., Hodzic, A., Madronich, S., and Aumont, B.: Impact of chamber wall loss of gaseous organic compounds on secondary organic aerosol formation: explicit modeling of SOA formation from alkane and alkene oxidation, Atmospheric Chemistry and Physics, 16, 1417–1431, https://doi.org/10.5194/acp-16-1417-2016, https://www.atmos-chem-phys.net/16/1417/2016/, 2016.

Lee-Taylor, J., Madronich, S., Aumont, B., Baker, A., Camredon, M., Hodzic, A., Tyndall, G. S., Apel, E., and Zaveri, R. a.: Explicit modeling of organic chemistry and secondary organic aerosol partitioning for Mexico City and its outflow plume, Atmospheric Chemistry and Physics, 11, 13 219–13 241, https://doi.org/10.5194/acp-11-13219-2011, 2011.

Lim, Y. B. and Ziemann, P. J.: Effects of molecular structure on aerosol yields from OH radical-initiated reactions of linear, branched, and cyclic alkanes in the presence of NO x, Environmental Science and Technology, 43, 2328–2334, https://doi.org/10.1021/es803389s, 2009.

McDonald, B. C., de Gouw, J. A., Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., Jimenez, J. L., Lee-Taylor, J., Hayes, P. L., McKeen, S. A., Cui, Y. Y., Kim, S.-W., Gentner, D. R., Isaacman-VanWertz, G., Goldstein, A. H., Harley, R. A., Frost, G. J., Roberts, J. M., Ryerson, T. B., and Trainer, M.: Volatile chemical products emerging as largest petrochemical source of urban organic emissions, Science, 359, 760–764, https://doi.org/10.1126/science.aaq0524, 2018.

Medeiros, A. S., Calderaro, G., Guimarães, P. C., Magalhaes, M. R., Morais, M. V., Rafee, S. A., Ribeiro, I. O., Andreoli, R. V., Martins, J. A., Martins, L. D., Martin, S. T., and Souza, R. A.: Power plant fuel switching and air quality in a tropical, forested environment, Atmospheric Chemistry and Physics, 17, 8987–8998, https://doi.org/10.5194/acp-17-8987-2017, 2017.

Riemer, N. and West, M.: Quantifying aerosol mixing state with entropy and diversity measures, Atmospheric Chemistry and Physics, 13, 11 423–11 439, https://doi.org/10.5194/acp-13-11423-2013, 2013.

Schifter, I., Díaz, L., Sánchez-Reyna, G., González-Macías, C., González, U., and Rodríguez, R.: Influence of gasoline olefin and aromatic content on exhaust emissions of 15% ethanol blends, Fuel, https://doi.org/10.1016/j.fuel.2019.116950, 2020.

Shah, R. U., Coggon, M. M., Gkatzelis, G. I., McDonald, B. C., Tasoglou, A., Huber, H., Gilman, J., Warneke, C., Robinson, A. L., and Presto, A. A.: Urban Oxidation Flow Reactor Measurements Reveal Significant Secondary Organic Aerosol Contributions from Volatile Emissions of Emerging Importance, Environmental Science & Technology, 54, 714–725, https://doi.org/10.1021/acs.est.9b06531, 2020.

Shilling, J. E., Pekour, M. S., Fortner, E. C., Artaxo, P., de Sá, S., Hubbe, J. M., Longo, K. M., Machado, L. A. T., Martin, S. T., Springston, S. R., Tomlinson, J., and Wang, J.: Aircraft observations of the chemical composition and aging of aerosol in the Manaus urban plume during GoAmazon 2014/5, Atmospheric Chemistry and Physics, 18, 10 773–10 797, https://doi.org/10.5194/acp-18-10773-2018, 2018.

Shrivastava, M., Andreae, M. O., Artaxo, P., Barbosa, H. M. J., Berg, L. K., Brito, J., Ching, J., Easter, R. C., Fan, J., Fast, J. D., Feng, Z., Fuentes, J. D., Glasius, M., Goldstein, A. H., Alves, E. G., Gomes, H., Gu, D., Guenther, A., Jathar, S. H., Kim, S., Liu, Y., Lou, S., Martin, S. T., McNeill, V. F., Medeiros, A., de Sá, S. S., Shilling, J. E., Springston, S. R., Souza, R. A. F., Thornton, J. A., Isaacman-VanWertz, G., Yee, L. D., Ynoue, R., Zaveri, R. A., Zelenyuk, A., and Zhao, C.: Urban pollution greatly enhances formation of natural aerosols over the Amazon rainforest, Nature Communications, 10, 1046, https://doi.org/10.1038/s41467-019-08909-4, 2019.

Yang, J., Roth, P., Durbin, T., and Karavalakis, G.: Impacts of gasoline aromatic and ethanol levels on the emissions from GDI vehicles: Part 1. Influence on regulated and gaseous toxic pollutants, Fuel, https://doi.org/10.1016/j.fuel.2019.04.143, 2019.

Yee, L. D., Isaacman-VanWertz, G., Wernis, R. A., Meng, M., Rivera, V., Kreisberg, N. M., Hering, S. V., Bering, M. S., Glasius, M., Upshur, M. A., Bé, A. G., Thomson, R. J., Geiger, F. M., Offenberg, J. H., Lewandowski, M., Kourtchev, I., Kalberer, M., de Sá, S., Martin, S. T., Alexander, M. L., Palm, B. B., Hu, W., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Liu, Y., McKinney, K. A., Artaxo, P., Viegas, J., Manzi, A., Oliveira, M. B., de Souza, R., Machado, L. A. T., Longo, K., and Goldstein, A. H.: Observations of sesquiterpenes and their oxidation products in central Amazonia during the wet and dry seasons, Atmospheric Chemistry and Physics Discussions, pp. 1–31, https://doi.org/10.5194/acp-2018-191, 2018.

Yuan, B., Koss, A. R., Warneke, C., Coggon, M., Sekimoto, K., and De Gouw, J. A.: Proton-Transfer-Reaction Mass Spectrometry: Applications in Atmospheric Sciences, Chemical Reviews, 117, 13 187–13 229, https://doi.org/10.1021/acs.chemrev.7b00325, 2017.

~~Impact~~ Exploration ~~of Urban Emissions on a Biogenic Environment~~ Oxidative Chemistry and Secondary Organic Aerosol Formation in the Amazon during the ~~wet season~~Wet Season: Explicit Modeling of the Manaus Urban Plume ~~Organic Chemistry~~ with GECKO-A

Camille Mouchel-Vallon[1,a], Julia Lee-Taylor[1,2], Alma Hodzic[1], Paulo Artaxo[3], Bernard Aumont[4], Marie Camredon[4], David Gurarie[5], Jose-Luis Jimenez[2,6], Donald H. Lenschow[7], Scot T. Martin[8,9], Janaina Nascimento[10,11], John J. Orlando[1], Brett B. Palm[2,6,b], John E. Shilling[12], Manish Shrivastava[12], and Sasha Madronich[1]

[1]Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, Boulder, CO 80301, USA
[2]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO 80309, USA
[3]University of Sao Paulo, Institute of Physics, Rua do Matao 1371, 05508-090, Sao Paulo, S.P., Brazil
[4]LISA, UMR CNRS 7583, Université Paris-Est-Créteil, Université de Paris, Institut Pierre Simon Laplace, Créteil, France
[5]Department of Mathematics and Center for Global Health and Diseases, Case Western Reserve University, Cleveland, OH 44106-7080, USA
[6]Department of Chemistry, University of Colorado, Boulder, CO 80309, USA
[7]Mesoscale and Microscale Meteorology Laboratory, National Center for Atmospheric Research, Boulder, CO 80301, USA
[8]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02318, USA
[9]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02318, USA
[10]Post-graduate Program in Climate and Environment, National Institute for Amazonian Research and Amazonas State University, Manaus, AM, Brazil
[11]Chemical Sciences Division, NOAA Earth System Research Laboratory, Boulder, CO 80305, USA
[12]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[a]Now at Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, Toulouse, France
[b]Now at Department of Atmospheric Sciences, University of Washington, Seattle, WA 91895, USA

Correspondence: C. Mouchel-Vallon (camille.mouchel-vallon@aero.obs-mip.fr)

Abstract. The GoAmazon 2014/5 field campaign took place in Manaus (Brazil) and allowed the investigation of the interaction between background level biogenic air masses and anthropogenic plumes. We present in this work a box model built to simulate the impact of urban chemistry on biogenic Secondary Organic Aerosol (SOA) formation and composition. An organic chemistry mechanism is generated with the Generator for Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A) to simulate the explicit oxidation of biogenic and anthropogenic compounds. A parameterization is also included to account for the reactive uptake of isoprene oxidation products on aqueous particles. ~~After some reductions of biogenic emissions relative to~~ The biogenic emissions estimated from existing emission inventories ~~, the~~ had to be reduced to match measurements. The model is able to reproduce ~~measured primary compounds, ozone and NOx for clean or~~ ozone and NO_x for clean and polluted situations. The explicit model is able to reproduce background case SOA mass concentrations but is ~~underestimating~~ not capturing the enhancement observed in the urban plume. Oxidation of biogenic compounds is the major contributor to SOA mass. A Volatility Basis Set parameterization (VBS) applied to the same cases obtains better results than GECKO-A for pre-

dicting SOA mass in the box model. The explicit mechanism may be missing SOA formation processes related to the oxidation of monoterpenes that could be implicitly accounted for in the VBS parameterization.

1 Introduction

The Amazonian rainforest is the largest emitter of biogenic primary hydrocarbons on Earth (*e.g.* Guenther et al., 2012). Photochemistry in this tropical region is ~~particularly intense all yearlong~~more photochemically active than other regions throughout most of the year, which stimulates the oxidation of the biogenic primary compounds by oxidants such as ozone and OH radicals. This part of the world is consequently a substantial source of Secondary Organic Aerosol (SOA) (Martin et al., 2010; Chen et al., 2015a) produced by condensation of oxygenated secondary organic species formed from gas and aqueous phase oxidation of biogenic compounds (Claeys, 2004; Carlton et al., 2009; Paulot et al., 2009). On the other hand, the city of Manaus (Brazil) is a source of anthropogenic pollution with 2.1 million inhabitants, ca. 600000 vehicles in circulation and 78 thermal power plants in its close surroundings (Abou Rafee et al., 2017). Manaus is situated at the confluence of the Rio Negro and Solimões rivers that subsequently form the Amazon River (Fig. 1). This metropolis is isolated from the rest of South American populated areas by over 1000 km of Amazonian tropical rainforest in every direction (*e.g.* Martin et al., 2016). Manaus is therefore a point source of urban pollution in a vast rainforest, making it an ideal place to study chemical interactions of biogenic and anthropogenic compounds. The Observations and Modeling of the Green Ocean Amazon (GoAmazon 2014/5) experiment was designed to characterize the anthropogenic perturbations to the clean air masses influenced by Amazonian natural emissions (Martin et al., 2016). The main instrumented site (T3) was situated approx. 70 km southwest of Manaus (see Fig. 1). In addition, the US Department of Energy (DOE) Gulfstream research aircraft (G-1) conducted 16 research flights to sample the Manaus plume as it was transported downwind and over the Amazon forest (Martin et al., 2016; Shilling et al., 2018). With varying meteorological conditions, this allowed sampling of clean background air from the Amazon basin and polluted air from Manaus (Martin et al., 2016).

Several studies have already shown that the overall composition of particulate matter (PM) in remote areas is distinctly different from urban areas, with anthropogenic PM being characterized by more sulfates and hydrocarbon-like compounds, whereas remote PM contains more oxidized organic matter (e.g. Xu et al., 2015; Budisulistiorini et al., 2016). In the Manaus environment, biogenic molecules would interact with the chemistry resulting from anthropogenic emissions. de Sá et al. (2018) have shown that the majority of sub micrometer particle mass at the T3 site is secondary. ~~Would SOA retain their biogenic nature and how would they exhibit anthropogenic influence?~~ Several studies have investigated how the biogenic nature of the SOA is affected by anthropogenic influence. For instance Aerosol Mass Spectrometer (AMS) measurements reported by de Sá et al. (2017) have shown that the contribution of epoxydiols derived from isoprene to SOA (IEPOX-SOA) amounts to 11 to 17% of the total organic mass when the Manaus plume is sampled, compared to 19 to 26% under background conditions. Using an Oxidation Flow reactor (OFR) and tracers for different source types, Palm et al. (2018) concluded that the Volatile Organic Compounds (VOC) and Intermediate Volatility Organic Compounds (IVOC) sampled during GoAmazon2014/5 could form SOA whose origin would be dominated by biogenic sources during the dry season, and by both biogenic and anthropogenic

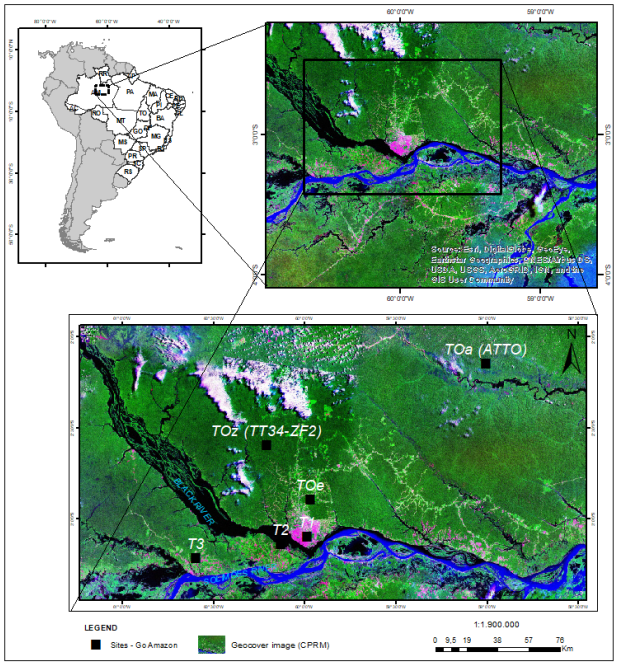


Figure 1. Map of the GoAmazon field campaign instrumented sites. Measurements used in this work came from the T3 site. Source: GeoCover, Instituto Brasileiro de Geografia e Estatistica (IBGE).

sources during the wet season. With a regional model study of the GoAmazon 2014/5 situation, Shrivastava et al. (2019) concluded that the higher oxidative capacity in the urban plume results in an enhancement of biogenic SOA production.

Models need to take into account the different nature of VOCs and SOA resulting from biogenic and anthropogenic chemistry to accurately represent their interactions. This can be done by looking at this problem with what Pankow et al. (2015) call a "molecular view", as opposed to the "anonymized view" followed by current 3D models. The molecular view attempts to predict SOA mass from the known and estimated properties of individually simulated organic compounds while the anonymized view uses hypothetical properties (*e.g.* volatility, solubility) of a small number of lumped compounds. ~~Even though a few instruments are getting closer to providing this molecular view of volatile organic compounds (e.g. Koss et al., 2017a, b), such measurements are not available for the GoAmazon 2014/5 campaign.~~ In a recent review, Heald and Kroll (2020) have reported on the recent progress in measurements of individual organic compounds, and how experimentalists are getting close to achieving closure on organic carbon in both gas and aerosol phases (*e.g.* Gentner et al., 2012; Isaacman-Vanwertz et al., 2018). As these measurements are now able to capture elemental formulas, double bonds, some oxygenated functional groups and aromaticity (*e.g.* Yuan et al., 2017), they still do not provide individual molecular identities. From this point of view, measurements are still restricted to a "formula view". For the GoAmazon field campaign, Yee et al. (2018) were able to sample and identify 30 sesquiterpenes and 40 of their oxidation products at the T3 site with a semi-volatile thermal desorption aerosol

gas chromatograph (SV-TAG, Isaacman et al., 2014) but they do not achieve the coverage needed to approach the "molecular view".

3D models that were run for the Manaus situation offer an anonymized view of SOA composition (Shrivastava et al., 2019) because they rely on a Volatility Basis Set parameterization (VBS, Donahue et al., 2006). The Generator for Explicit Chemistry and Kinetics of Organics in the Atmosphere (GECKO-A, Aumont et al., 2005; Camredon et al., 2007) is ~~the ideal~~ an excellent tool to model atmospheric organic chemistry with a detailed molecular view. GECKO-A is an automated chemical mechanism generator built to write the explicit chemistry of given precursors by following a prescribed set of systematic rules. This set of systematic rules relies on experimental data when available and Structure Activity Relationships (SAR) to determine unknown kinetic or thermodynamic constants. It has previously been run in box models to evaluate processes like secondary organic aerosol formation (Valorso et al., 2011; Aumont et al., 2012; Camredon and Aumont, 2006; Camredon et al., 2007) and dissolution of organic compounds (Mouchel-Vallon et al., 2013). It was also applied to simulate chamber experiments (Valorso et al., 2011; La et al., 2016) and urban and biogenic plumes (Lee-Taylor et al., 2011, 2015).

In this work, a box model is run to simulate the evolution of an Amazonian air mass intercepting Manaus emissions during the wet season. Emissions of anthropogenic and biogenic primary VOCs are estimated with available data. The chemical scheme describing the explicit oxidation of these primary compounds is generated with GECKO-A. The resulting detailed simulation is then used to explore the impact of Manaus emissions on the Amazonian biogenic chemistry. Comparisons with aerosol mass spectrometer data and the VBS parameterization are carried out to identify important processes involved in biogenic SOA formation that may not be accounted for in GECKO-A. Finally the potential for reduction of the explicit mechanism is estimated.

2 Experimental Data

The main instrumented site (referred to as T3 hereafter) of the GoAmazon 2014/5 field campaign was situated 70 km west of Manaus (Fig. 1). Two aircraft were also deployed, a G-159 Gulfstream I (G-I) (Schmid et al., 2014) that flew at low altitude and mostly sampled the boundary layer and a Gulfstream G550 (HALO) that flew higher altitudes and sampled the free troposphere (Wendisch et al., 2016). The flight tracks are depicted in Martin et al. (2016) and Wendisch et al. (2016). The G-1 airplane mainly flew daytime transects of the Manaus plume between the city and the T3 site.

The detailed instrumentation deployed at T3 and in the airplanes has been described elsewhere (Martin et al., 2016). For this study we mainly relied on ground deployed instruments briefly described here.

Ozone concentration measurements made with a Thermo Fisher Model 49i Ozone Analyzer were obtained from the Mobile Aerosol Observing System-Chemistry (MAOS-C).

Due to some issues with the NO_x analyzer deployed at T3 by the MAOS-C during the wet season, NO_x data reported here is weakly reliable. The values reported here are only qualitative indications of NO_x levels in the studied period.

OH radicals concentrations were provided by an OH chemical ionization mass spectrometer (Sinha et al., 2008, OH-CIMS).

Table 1. Box model constraints used in the clean and polluted setups

	Clean Background	Manaus
NO soil emission [molec cm^{-2} s^{-1}][(a)]	8.3×10^{9}	–
Aerosol number concentration [cm^{-3}][(b)]	5×10^{2}	1×10^{4}
Aerosol pH	3.0	1.5
Aerosol sulfate concentration [µg m^{-3}][(b)]	0.3	0.4
Aerosol nitrate concentration [µg m^{-3}][(b)]	0.05	0.1
Hygroscopicity Parameter (κ)[(c)]	0.15	0.15

[(a)]Shrivastava et al. (2019) [(b)]de Sá et al. (2018) [(c)]Thalman et al. (2017)

Organic compounds in the gas phase were measured with selected reagent ion proton transfer reaction time-of-flight mass spectrometer (SRI-PTR-ToFMS, Jordan et al., 2009a, b). Aerosol composition was monitored by a high-resolution time-of-flight aerosol mass spectrometer (HR-ToF-AMS) (DeCarlo et al., 2006; de Sá et al., 2018, 2019).

For the purpose of comparisons with the model, we need to be able to separate time periods representing clean and polluted episodes. Using a fuzzy c-means clustering algorithm (Bezdek, 1981; Bezdek et al., 1984) applied to T3 measurements, de Sá et al. (2018) were able to identify four different clusters corresponding to (i) fresh or (ii) aged (2+ days) biogenic production, and air masses influenced by the (iii) northern or (iv) southern parts of Manaus. Using the timeseries contribution of these clusters, we labeled as background air masses that were identified as being composed of at least 50% of any clean cluster (i or ii). Conversely, air masses that were identified by de Sá et al. (2018) as being composed of at least 50% of any polluted cluster (iii and iv) were labeled as polluted. The clustering methods constrained the classification to only include wet season afternoon air masses that were not exposed to rain in the previous day (see de Sá et al., 2018). These limitations match with our model restrictions which do not include cloud chemistry, nor fire emissions that would be important during the dry season. For comparison with the model, experimental data were hourly averaged for each cluster.

3 Model Setup

A Lagrangian box model was built to simulate chemistry in the planetary boundary layer and the residual layer for an air parcel traveling over the Amazonian forest and Manaus. Because experimental data compared to the model only contain air masses that were not exposed to rain in the previous day (see Sect. 2 and de Sá et al., 2018), the model simulates biogenic conditions for one day, assuming the air mass was washed out by rain prior to that day. After the one day spinup, biogenic emissions are replaced by urban emissions for one hour during the second day to represent the interaction of the air mass with the Manaus urban area. After the simulated encounter with Manaus, the model inputs return to biogenic emissions until the end of the second day. This simulation is defined hereafter as the "polluted" case. Another simulation is run where the box is only subjected to biogenic emissions for two days, without any exposure to urban emissions to simulate a background case.

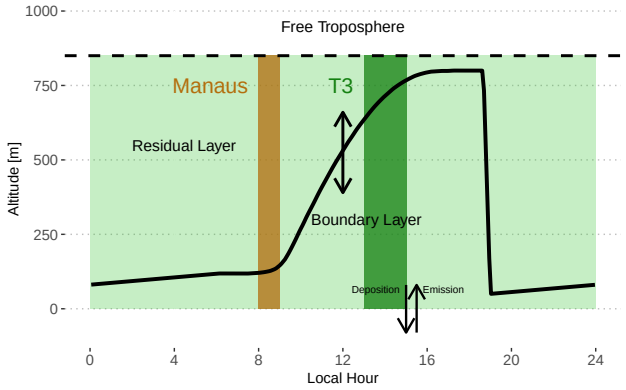


Figure 2. Schematic depiction of the box model setup used in this work. The continuous black line shows the time evolution of the PBL height. The dashed black line depicts the top of the residual layer box. The brown shaded area is the period when the box is subjected to Manaus emissions. For the rest of the time period, the box is subjected to biogenic emissions (light and dark green shaded areas). The dark green shaded area is approximately the period when the box would be over the main instrumented site T3, assuming a travel time of 4 to 6 hours.

This simulation is defined hereafter as the "clean" case. This section describes the box model setup, how the emissions were defined and the chemical mechanism used for this study.

3.1 Box model

This study relies on a box model described in this section. It includes emissions from the forest and the city, deposition and chemical evolution of the trace gases. Daytime growth of the planetary boundary layer is also simulated, with mixing with the residual layer.

3.1.1 Boundary Layer

The model includes two boxes on top of each other separated by a moving boundary representing the height of the boundary layer. The bottom box extends from the surface to the top of the planetary boundary layer (PBL). The top box extends from the top of the planetary boundary layer to ~~1.5 km~~ 850 m and represents the residual layer (RL) (see Fig. 2). The daytime PBL height evolution is parameterized according to the Tennekes (1973) approach and was calculated using the Second-Order Model for Conserved and Reactive Unsteady Scalars (SOMCRUS, Lenschow et al., 2016) (see Fig. 2). At sunset, stratification is assumed to quickly shrink the PBL to 50 m which results in the contents of the PBL being reallocated to the RL. During the night, the PBL is constrained to linearly grow to reach the next morning level. The PBL height evolution is the same for each of the two simulated days. During the day, the PBL is therefore slowly incorporating residual chemicals resulting from the previous day and night chemistry. Thalman et al. (2017) report PBL heights estimated from ceilometer measurements during the wet season in the central Amazonian Forest, for polluted and background conditions. The measurements reach a maximum of 800 m at

around 3pm local time. This value was used to further constrain the PBL height evolution by scaling the SOMCRUS output to reach this measured PBL height maximum. The growth and shrinking of the PBL dilute the expanding box and transfer gases from the shrinking box to the expanding box. This is parameterized according to Eqs. 1 and 2:

$$\frac{dC_b}{dt} = \begin{cases} \frac{1}{h}\frac{dh}{dt}C_t - \frac{1}{h}\frac{dh}{dt}C_b & \text{if } \frac{dh}{dt} > 0 \\ 0 & \text{if } \frac{dh}{dt} \leq 0 \end{cases} \quad (1)$$

$$\frac{dC_t}{dt} = \begin{cases} 0 & \text{if } \frac{dh}{dt} \geq 0 \\ -\frac{1}{H-h}\frac{dh}{dt}C_b + \frac{1}{H-h}\frac{dh}{dt}C_t & \text{if } \frac{dh}{dt} < 0 \end{cases} \quad (2)$$

C_b and C_t [molec cm^{-3}] are chemical species concentrations in the PBL (bottom) and RL (top) boxes respectively. h [m] is the variable height of the PBL and H [m] is the fixed altitude of the RL top. The first term in each equation describes the addition of material coming from the shrinking box and the second term describes the dilution of the growing box. Following these equations, mixing happens in two stages: (i) the long RL entrainment into the PBL over day time and (ii) the rapid transfer of the PBL to the RL at sunset. The box model approach assumes rapid mixing in both layers and that chemistry is applied to well-mixed concentrations. The residual layer is also slowly mixed with the free troposphere. The free troposphere is assumed to be a fixed reservoir of CO (80 ppb) and ozone (15 ppb, *e.g.* Browell et al., 1990; Gregory et al., 1990; Kirchhoff et al., 1990). The subsidence velocity is constant and fixed at 0.1 m s^{-1} (*e.g.* Raes, 1995).

Temperature is assumed to follow a sinusoidal daily variation, with an average of 27 °C, an amplitude of 4 °C and a maximum at 6 pm local time. Relative humidity is initially set at 75% at 6 am (23 °C) and is free to evolve with temperature changes assuming water vapor concentration is constant.

3.2 Emissions

3.2.1 Biogenic Emissions

VOC emissions from the rainforest were estimated with the Model of Emissions of Gases and Aerosols from Nature (MEGAN v2.1, Guenther et al., 2012). Biogenic emissions on March, 13$^{\text{th}}$ 2014 (the golden day of the GoAmazon field campaign, see de Sá et al., 2017) in a domain situated in the forest around Manaus were averaged to obtain total isoprene and monoterpene hourly averaged emissions for a day typical of the wet season, without any recorded rain event. Monoterpenes were then speciated to match concentrations measured by Jardine et al. (2015) at the top of an Amazonian rainforest canopy. Based on this emission inventory, we then simultaneously optimized isoprene and monoterpenes emissions to match the model with T3 isoprene and total monoterpenes under clean conditions. This resulted in the need to reduce isoprene emissions by a factor of 7. Using measurements from a similar site in Amazonia, Alves et al. (2016) reported that MEGAN 2.1 overestimated isoprene emissions by a factor of 5 on average during the dry season. They assumed that the T3 site configuration (a clearing in the forest, near a road) could affect local isoprene concentrations compared to average Amazonian emissions. For instance measurements in the Amazon rainforest by Batista et al. (2019) indicate that biogenic emissions exhibit large intermediate scale heterogeneity,

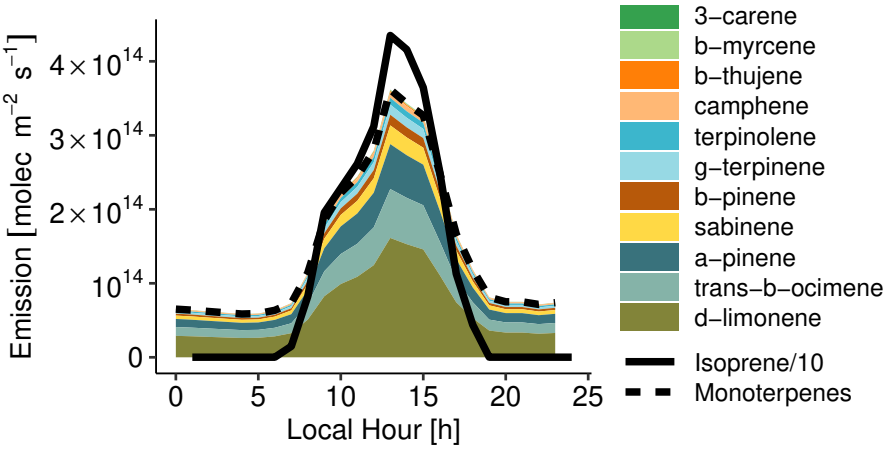


Figure 3. Hourly biogenic emissions estimated with MEGAN and scaled to match measured concentrations (see 3.2.1). The lines depict isoprene (continuous line) and total monoterpenes (dashed line). The colored areas depict the contribution of each individual species to total monoterpenes. Please note that isoprene emissions are divided by 10 to fit on the plot.

with estimated emission variations of 220% to 330%. Recent satellite based estimates of biogenic emissions also reported that MEGAN overestimates isoprene emissions in Amazonia by 40% (Worden et al., 2019). In a similar way, monoterpenes emissions had to be reduced by a factor of 8 compared to the MEGAN values. Figure 3 depicts the resulting daily biogenic emission cycle. Isoprene emissions dominate monoterpene emissions by approximately an order of magnitude. δ-limonene is the most emitted monoterpene (45%), followed by trans-β-ocimene (18%) and α-pinene (17%). NO soil emissions are also accounted for with a constant flux of 8.3×10^{9} molec cm^{-2} s^{-1} following Shrivastava et al. (2019).

3.2.2 Manaus Emissions

The emissions used to represent the influence of Manaus are shown in Fig. 4a and were calculated following the methodology described in Abou Rafee et al. (2017) and Medeiros et al. (2017). Traffic emissions have been estimated from vehicle use intensity and emission factors for CO, NO_x, SO_2 and VOCs depending on type of fuel use in Manaus (Abou Rafee et al., 2017). VOC speciation is assumed to be similar to the average speciation of the vehicle fleet emissions of São Paulo, Brazil in 2004 (Martins et al., 2006). Hourly distribution of the traffic emissions is considered to be similar to the hourly traffic distribution in São Paulo (Andrade et al., 2015). In the past decades, Brazil has become known for pioneering the large scale use of ethanol based biofuels. However, due to its isolation and being distant from south Brazilian biofuel producing regions, Manaus traffic doesn't involve consumption of significant amounts of ethanol-based fuel.

~~This~~ The difference in the fuel blend used in São Paulo and Manaus can introduce errors in the traffic emissions VOC speciation. For instance, a recent study by Yang et al. (2019) showed that the combustion of fuels with higher ethanol content emits significantly less carbon monoxide and more acetaldehyde. Schifter et al. (2020) showed similar results, and also suggested that ethanol blends emit smaller amounts of simple aromatic compounds (*e.g.* benzene, toluene). This speciation uncertainty can

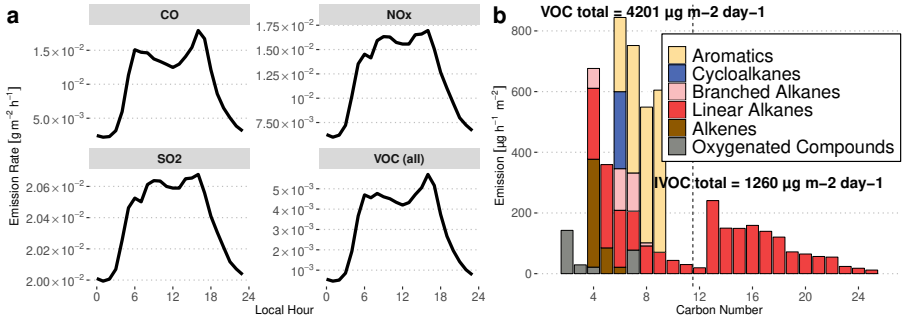


Figure 4. Diurnal evolution of simulated traffic emissions in Manaus deduced from inventories in Manaus and Saõ Paulo. (a) NO_x, SO_2, CO and total VOC daily emissions. (b) Carbon number distribution of Manaus emissions at noon. Total daily emissions are indicated for lighter organic compounds (VOC) and less volatile compounds (IVOC). The dashed line denotes the separation between VOCs (left) and IVOCs (right).

especially have an impact on oxidants concentrations. Schifter et al. (2020) reported for instance that fuels containing ethanol would potentially produce less ozone after the oxidation of emitted organic species than fuels without ethanol. Moreover, the lifetime of OH is likely to change depending on the speciation of emitted VOCs due to varying reactivities with respect to OH. In the same way that the potential for ozone formation could depend on the use of ethanol fuel blends, it is also possible that the potential for SOA formation would depend on these fuel blends too.

This traffic emission estimate does not include Intermediate Volatile Organic Compounds (IVOC) which would mainly be produced by diesel vehicle emissions (Gentner et al., 2012, 2017). Zhao et al. (2015, 2016) showed that the IVOC to VOC emissions ratio lies between 4% for gasoline vehicles and 65% for diesel vehicles. Knowing that diesel vehicles account for ca. 45% of the total driven distance in Manaus (Abou Rafee et al., 2017), we therefore assume that IVOC total emissions are approximately equal to 30% of total VOC emissions. To estimate the distribution of species resulting from IVOC emissions, we assumed that the distribution in volatility is similar to the distribution used to simulate traffic emissions in Mexico City in Lee-Taylor et al. (2011), with n-alkanes from C_{12} to C_{25} acting as surrogates for these heavier emitted organic compounds.

The resulting distribution of urban organic emissions at noon as a function of the number of carbon atoms is presented in Fig. 4b. As reported in the Gentner et al. (2017) review, gasoline emissions have a maximum for C_8 species, with no emission of importance above C_{12}, whereas diesel vehicles emit species from C_{10} to C_{25}, with a peak at C_{12}. These features are present in the emissions estimated in this work, with the gasoline peak around $C_{6\text{-}7}$ and the diesel maximum at C_{13}. Gentner et al. (2017) also report that half of the gasoline VOC emissions are composed of linear and branched alkanes, the other half consisting of aromatics and cycloalkanes. In our estimates of gasoline emissions ($C_{<12}$) the proportion of branched alkanes is smaller, alkenes constitute a more important fraction of emitted $C_{4\text{-}6}$ species, branched cycloalkanes are missing, and aromatics constitute the majority of emissions of $C_{7\text{-}10}$ compounds. These differences could represent differing sources of fuels or different distributions of vehicle brands and ages. In the case of diesel emissions, Gentner et al. (2017) report that they are approximately equally

distributed between aromatics, branched cycloalkanes, bicycloalkanes and branched alkanes whereas our method leads to diesel emissions being only constituted of n-alkanes, which are used here as surrogate species for the entire mixture.

Choosing alkanes as surrogates for emitted IVOCs is likely to introduce uncertainties to SOA produced from their oxidation. Lim and Ziemann (2009) carried out multiple chamber experiments that investigated the impact of branching and rings on alkanes SOA yields. For instance they showed that SOA yields range from a few percent for branched alkanes with 12 carbon atoms to 80% for cyclododecane while n-dodecane has an SOA yield of $\approx 32\%$. La et al. (2016) simulated these experiments with GECKO-A and they were able to reproduce this experimentally observed behavior. This means that without a detailed inventory of emitted IVOCs, the uncertainty on the SOA yield from IVOCs is high in our version of the model. It should be noted that the range of measured SOA yields for structurally different compounds with the same number of carbon atoms seems to peak for C_{10} to C_{13} alkanes. The range of observed SOA yields in Lim and Ziemann (2009) decreases after this peak. For instance, SOA yields for C_{15} alkanes of various structures range from 45% to 90%. We can therefore expect the IVOCs SOA yield to be highly sensitive to the speciation of compounds ranging from C_{12} to C_{14}, but this sensitivity should decrease for heavier molecular weight species.

Additionally, emissions from 11 local thermal power plants (TPP) and one oil refinery located in the vicinity of Manaus were obtained from the data presented in Medeiros et al. (2017). Based on monthly statistics of fuel use in each of the TPP and the oil refinery, combined with emission factors of CO and NO_x for each type of fuel (diesel, fuel oil, natural gas), we calculated CO and NO_x emissions for February, March and April 2014. These total emissions were then averaged over the whole surface area of Manaus (377 km^2, Abou Rafee et al., 2017). Total SO2 emissions were taken from Abou Rafee et al. (2017) and added to the urban emissions for the considered Manaus area.

3.3 Chemical Mechanism

3.3.1 GECKO-A

All emitted organic compounds were used as inputs to GECKO-A to automatically generate the chemical scheme used in this study. The GECKO-A protocol has been described in detail in Aumont et al. (2005) and updated in Camredon et al. (2007), Valorso et al. (2011), Aumont et al. (2013), and La et al. (2016). Partitioning of low volatility compounds to the aerosol phase is described dynamically as in La et al. (2016). Vapor pressures are estimated with the Nannoolal et al. (2008) structure activity relationship. As isoprene first oxidations steps have been widely studied in the literature, there is no need to automatically generate them with GECKO-A. Isoprene chemistry first two generations of oxidation were therefore taken from the Master Chemical Mechanism 3.3.1 (Jenkin et al., 1997; Saunders et al., 2003; Jenkin et al., 2015, MCM, *e.g.*). With 12 biogenic and 53 anthropogenic precursors ranging from C_2 to C_{25}, some reductions are carried out to reduce the size of the generated mechanisms. Species with an estimated vapor pressure below 10^{-13} atm are assumed to entirely partition to the aerosol phase so quickly that a description of their gas phase oxidation is not needed (Valorso et al., 2011). Furthermore, lower yield, longer chain species are lumped with chemically similar compounds according to a a hierarchical decision tree based on molecular structure (Valorso et al., 2011). The resulting chemical scheme contains ~~16~~ 23 million reactions involving ~~4~~ 4.4

million species of which ~~470000~~ 780000 can partition into the aerosol phase. The time integration in the two-box model setup takes approximately 0.5 computing hour per simulated hour on 16 cores (Computational and Information Systems Laboratory, 2017).

3.3.2 Isoprene SOA formation

GECKO-A treats SOA formation through a dynamic approach that converges towards the equilibrium defined by the Pankow formulation of Raoult's Law (Pankow, 1994). However it is likely that isoprene SOA (ISOPSOA) formation is not only controlled by vapor pressure (Paulot et al., 2009). Among factors that have been identified to play a role in ISOPSOA are: aqueous phase oxidation in deliquescent aerosol (*e.g.* Blando and Turpin, 2000; Ervens et al., 2011; Daumit et al., 2016), organic sulfate/nitrate formation via interaction with the inorganic component of the aerosol (*e.g.* McNeill et al., 2012; Pratt et al., 2013; Wang et al., 2018; Glasius et al., 2018; Jo et al., 2019), and accretion reactions in the bulk aerosol (*e.g.* oligomerization, dimerization, Altieri et al., 2006; Liu et al., 2012; Renard et al., 2015). None of these processes is currently implemented in the GECKO-A framework. For this study we use a simplified approach based on Marais et al. (2016) allowing the representation of ISOPSOA formation depending on the assumed composition of the inorganic aerosol. This parameterization describes the heterogeneous reactive uptake of important isoprene oxidation products. This accounts for the diffusion of the gases to the surface of the wet aerosol particle, their accomodation to the surface and their dissolution. The relevant parameters used here are listed in Marais et al. (2016). Isoprene epoxides (epoxydiols and hydroxyepoxides) react in the aqueous phase to open their epoxide ring via acid-catalized reactions. These reactions are followed by either the nucleophilic addition of (i) H_2O to form methyltetrols or (ii) sulfate and nitrate ions to form organosulfates and organonitrates. The uptake of epoxides therefore depends on the acidity of particles, as well as their sulfate and nitrate content. These parameters had to be constrained in the model and were deduced from the T3 AMS measurements and literature data (see Table 1). On the other hand, isoprene oxidation products containing nitrate moieties (dihydroxydinitrates and isoprene nitrate) hydrolyze and form polyols and nitric acid.

3.4 Dry Deposition

Dry deposition is treated following the Wesely (1989) parameterization. This parameterization is a resistance model that allows calculating dry deposition velocities based on multiple resistances defined as properties of the surfaces. The city and the forest were respectively attributed the properties of surfaces defined as urban and deciduous forest in the Wesely (1989) paper. The dry deposition velocity of a given species depends on its solubility expressed by its Henry's law coefficient. Because the solubility of most organic compounds generated with GECKO-A is unknown, they are here estimated using the GROupcontribution Method for Henry's law Estimate SAR (Raventos-Duran et al., 2010).

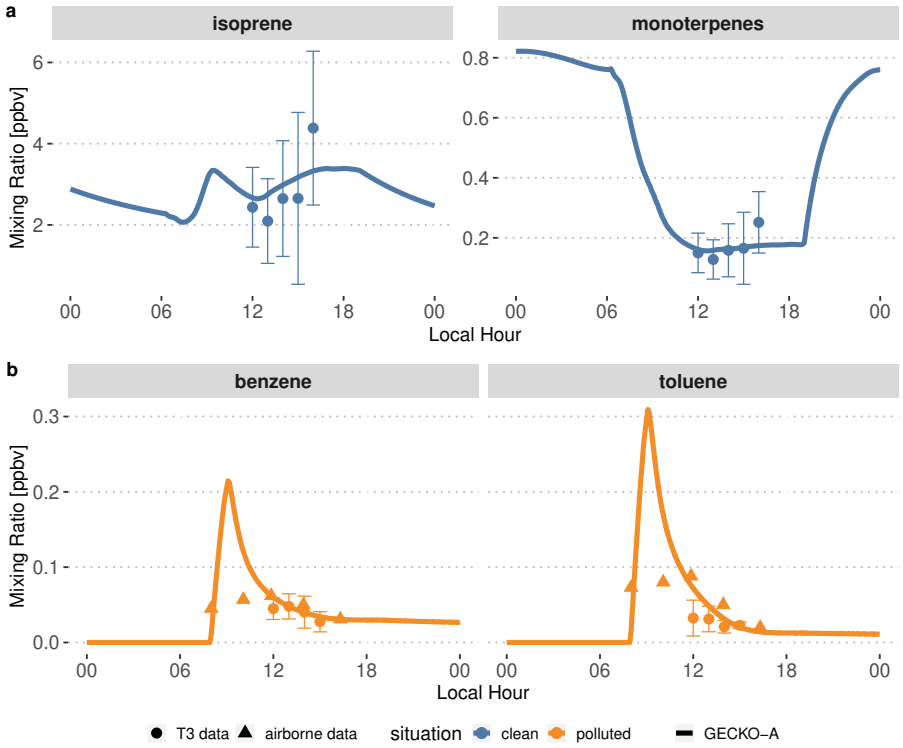


Figure 5. Modeled (lines, second day) time evolution of primary species concentrations in the Lagrangian box-model described in Sect. 3.1~~and~~ , average experimental ~~(dots)~~ concentrations measured at the T3 site (dots) and in the airplane (triangles). The vertical range of the experimental data denotes the standard deviation of measured concentrations during events identified as clean (top, blue) and polluted (bottom, orange). The airborne data was measured during plume transects. For each transect, aircraft distance from Manaus was converted to a time separation from Manaus assuming the plume leaves the city at 8am and arrives above T3 at 2pm.

4 Results and Discussion

4.1 Gas Phase Organics: Primary Organic Compounds and Oxidants

Figure 5 depicts the time evolution of selected primary organic species, and compares the model with available measurements. In the clean situations, measured isoprene mixing ratios range from 2–3 ppb at noon to 5–6 ppb at the end of the afternoon. The sum of all monoterpenes follows a similar increasing trend in the afternoon, from 0.1 to 0.3 ppb. After adjusting biogenic emissions rates (see Sect. 3.2.1), the model is able to reproduce these mixing ratios, with isoprene and monoterpenes being simulated ~~in the higher range~~ to the average of experimental values. In polluted situations, the model shows a peak of anthropogenic organic compounds when the plume encounters Manaus emissions between 8 and 9 am. This peak reaches ~~0.3 ppb and~~ ~~0.2~~ 0.2 ppb and 0.3 ppb respectively for benzene and toluene (Fig. 5). Their levels decay for the remainder of the day. Because the T3 measurement site is situated 4 to 6 hours downwind of Manaus, measurements of benzene and toluene can be compared

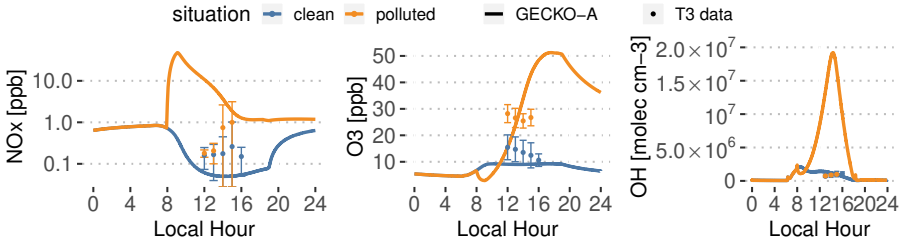


Figure 6. Experimental (dots, T3 site) and modeled (lines, second day) time evolution of ~~radicals concentrations~~ NO_x (left, note log scale), ozone (middle) mixing ratios and OH radicals concentrations (right~~, note log scale~~)~~mixing ratios~~. The vertical range of the experimental data denotes the standard deviation of measured concentrations at T3 during events identified as clean (blue) and polluted (orange).

to decayed modeled levels after that time span. The modeled mixing ratio of benzene matches the ~~lower range of~~ measurements, between ~~0.2~~ 0.4 and 0.6 ppb, while modeled toluene is closer to the higher range of measurements, between ~~0.4~~ 0.2 and 0.6 ppb during the afternoon. Figure 5 also displays airborne measurements of the same anthropogenic compounds during plume transects. The modeled mixing ratios of benzene and toluene decay in a similar way to the concentrations measured at each plume transects. The modeled peak is not seen by the aircraft measurements as the aircraft may not be flying close enough to the emission sources to capture it.

Pristine forest conditions are characterized in the model by low NO_x emissions from the soil (8.3×10^9 molec cm^{-2} s^{-1} $\approx$ 1.5×10^{-5}g m^{-2} h^{-1}, see Table 1). The model predicts NO_x mixing ratios around 50 ppt in the afternoon. In the polluted case, the background air mass is exposed to a complex mixture of anthropogenic compounds emissions as well as three orders of magnitude higher NO_x emissions ($\approx 1 \times 10^{-2}$g m^{-2} h^{-1}, see Fig. 4). This leads to modeled NO_x ~~between 500 ppt and 2~~ around 1 ppb in the afternoon, after a ~~25~~ 48 ppb peak in the city in the morning. The increase in NO_x is not as important in the experimental data, but these NO_x measurements are highly uncertain, which could explain the modeled discrepancies.

~~Modeled daytime~~ Daytime ozone mixing ratios ~~of 7 to 8~~ are modeled around 9 ppb in the clean situation, in the lower range of measured values. The higher NO_x levels result in strong ozone production in the polluted plume, characterized by mixing ratios of ~~16~~ 15 ppb at noon and up to ~~33~~ 51 ppb at the end of the afternoon. ~~These values are in the upper range of measured ozone at~~ During this increase of ozone production, the model matches T3 ~~during polluted events~~measurements around around 23 ppb at 1pm. On average, measured ozone in the polluted case is a factor of 2 higher than the clean case while the model sees an increase by a factor of ~~more than 3.~~ 2 to 4 between noon and 6pm. It should also be noticed that the model completely separates clean and polluted situation, which increases the contrast for all variables compared to the classification of the measurements that always includes some degree of mixing (see 2). ~~The~~ It should also be noted that the nighttime decay of ozone can be explained by dry deposition to the forest surface.

Furthermore, VOCs in the plume are exposed to high OH concentrations, with modeled concentration reaching ~~1.6~~1.9$\times 10^7$ molec cm^{-3} in the afternoon. In the clean background, OH concentrations only reach ~~1~~2$\times 10^6$ molec cm^{-3}. These clean values are in the lower range of reported measurements at T3 ~~.~~ Unlike the model, OH measurements averaged at T3 and identified

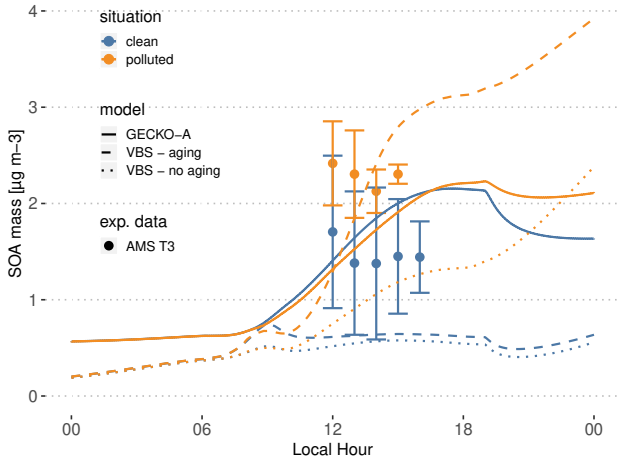


Figure 7. Experimental (circles, T3 site) and modeled (lines, second day) time evolution of SOA mass concentration. The vertical range of the experimental data denotes the standard deviation of measured concentrations. Cases are identified as clean (blue) and polluted (orange). The continuous lines depict the GECKO-A model run and the dashed lines depict the modeled SOA mass predicted with the VBS approach from Shrivastava et al. (2019). The dotted lines depict modeled SOA mass predicted with the VBS approach without including aging processes (see Sect. 4.3).

as clean and polluted did not exhibit any difference between both situations ~~.~~ (Fig. 6). In that case, there could be issues with the OH measurements at T3. Indirect constraints have shown differences between clean and polluted situations. Liu et al. (2018) derived OH concentrations from isoprene and its oxidation products measurement. They showed that noontime OH concentrations vary between 5×10^5 molec cm^{-3} in clean situations to 1.5×10^6 molec cm^{-3} in polluted events. The Shrivastava et al. (2019) 3D model exhibits a similar OH behavior to this work with concentrations at T3 ranging from $2\sim5\times10^5$ molec cm^{-3} (clean) to more than 4×10^6 molec cm^{-3} (polluted). The GECKO-A model is therefore likely to be overestimating OH concentrations in the urban plume by a factor of 5 to 10. This could stem from either overestimating NO or underestimating VOCs emissions in the city.

4.2 Modeled Urban Impact on SOA Mass and Composition

4.2.1 Modeled vs Measured SOA Mass Concentrations

At the measurement site, SOA mass concentrations measured by AMS range from 0.6 to 2.5 µg m^{-3} in clean conditions. In polluted conditions, SOA mass concentrations range from 1.9 to 2.9 µg m^{-3} (Fig. 7). In the clean case, the modeled SOA mass is within the ~~lower~~ range of T3 measurements, increasing from ~~0.2~~ 0.6 µg m^{-3} at sunrise to ~~1.25~~ 2.16 µg m^{-3} at the end of the afternoon. In the polluted situation, ~~SOA mass production rate is higher after encountering Manaus at 8am~~modeled SOA mass concentration is very similar to the clean simulation, with only a 20 minutes delay in the start of SOA production. The maximum concentration is ~~1.5~~ 2.23 µg m^{-3}, ~~a 20~~only a 3.5% increase compared to the clean simulation~~. Even if~~, while

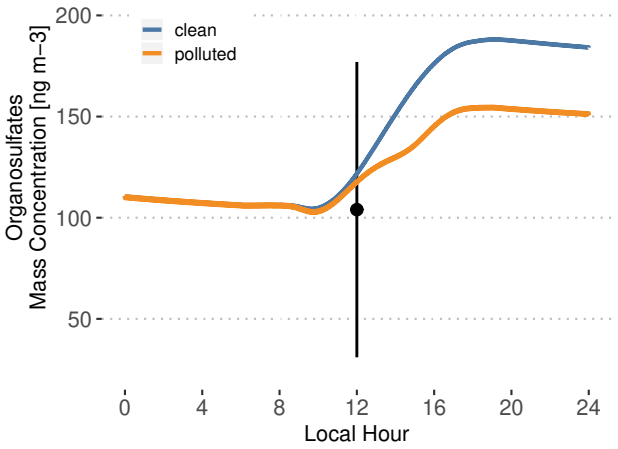


Figure 8. Modeled time evolution of particle phase organosulfates mass concentration. Cases are identified as clean (blue) and polluted (orange). The point and vertical line depicts the average and standard deviation of measurements reported in Glasius et al. (2018) for the wet season.

experimentally this increase averaged around 56%. Because the model is ~~able to qualitatively~~ unable reproduce the observed urban SOA enhancement, in the polluted situation the model ~~still~~ underestimates SOA mass by ~~30 to 50~~10 to 45%.

4.2.2 Organosulfates

Figure 8 depicts modeled particle phase organosulfates, with mass concentrations ranging from ~~105~~ 104 ng m^{-3} in the morning to ~~205~~ 188 ng m^{-3} in the evening in the clean case scenario. The polluted situation ~~increases~~ decreases late afternoon concentrations to ~~220~~ 155 ng m^{-3}. These values are in the higher range of the reported measured range of 104 ± 73 ng m^{-3} in Glasius et al. (2018). This is consistent with Glasius et al. (2018) who reported that the main source of the measured organosulfates is IEPOX heterogeneous uptake, which is the only pathway represented in this model. Furthermore, this shows that the combination of the MCM 3.3.1 isoprene oxidation mechanism to produce IEPOX and the reactive uptake parameterization from Marais et al. (2016) is able to predict realistic levels of organosulfates, assuming that aerosol properties are also realistic (hygroscopicity, inorganic sulfates and pH).

4.2.3 Modeled Organic Functional Groups

Figure 9 depicts the distribution of organic functional groups in the particle phase. In the clean case scenario, total functionalization, defined as the number of functional groups per carbon atom, is constant around approximately 0.5. As expected for a low-NO_x ~~situations~~situation, approximately 40% of these functional groups are hydroxy moieties and ~~20~~30% of the organic functional groups are hydroperoxides. The remaining functional groups are dominated by carbonyls and nitrates to a lower

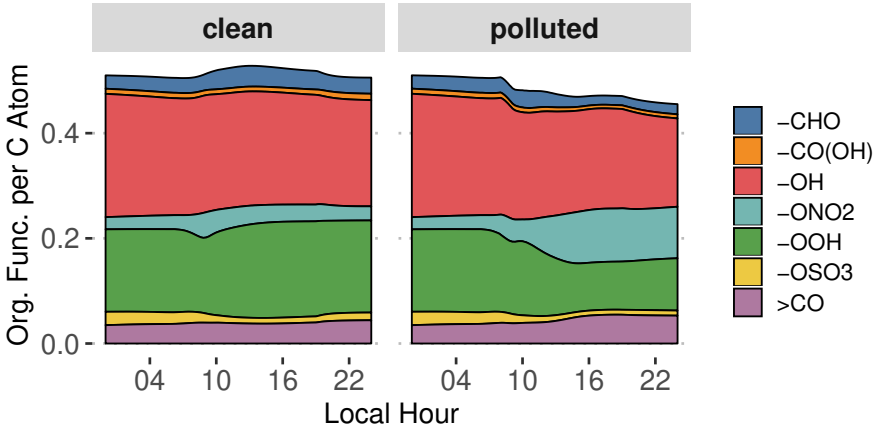


Figure 9. GECKO-A modeled time evolution of particle phase organic functionalization for the clean (left panel) and the polluted (right panel) cases. The y-axis is read as the number of a given organic function per carbon atom, i.e. in the clean case there is in total approximately one organic function for every two carbon atom.

extent. Manaus pollution has the direct effect of reducing total functionalization by 10% because of the contribution of long-chain primary hydrocarbons to SOA formation in the plume. Oxidation of organics in the higher NO_x environment also leads to an increase of nitrate moieties contribution at the expense of hydroxy and hydroperoxide moieties.

The change in overall modeled SOA composition between clean and polluted cases is quite small. AMS measurements give a similar impression of a small impact of polluted situations on atomic ratios (Fig. 10), with only a slight increase of O/C ratio (see Sect. 4.2.4). Other analyses of airborne and ground AMS data (de Sá et al., 2018; Shilling et al., 2018) similarly show that the relative contribution of hydrocarbon-like organic aerosol (HOA) slightly increases in the polluted plume at the expense of isoprene derived SOA. The model and the AMS data support the idea that the impact of anthropogenic emissions is mostly seen on the total organic aerosol mass, and that all constituents of the organic aerosol phase increase approximately in the same way.

4.2.4 Modeled vs Measured Atomic Ratios

Figure 10 depicts simulated, ground measurements and airborne measurements of O/C and H/C atomic ratios in aerosol particles on a van Krevelen diagram. At the T3 site, experimental O/C ratios range from 0.7 to 1 in both clean and polluted conditions while H/C ratios range from 1.2 to 1.4. Additionally airborne measurements above the T3 site report O/C ratios ranging from 0.35 to 0.9 and H/C ratios ranging from 1.5 to 1.9. Compiling multiple field campaigns AMS measurements, Chen et al. (2015b) reported van Krevelen diagrams slopes (H/C vs O/C) ranging from -1 to -0.7. A linear regression over the data points from both airborne and ground measurements (dotted line on Fig. 10) gives a slope of -1.3, close to values reported in Chen et al. (2015b). This means that T3 air masses ~~are~~ may be sampled at a later stage of oxidation than the airborne samples, possibly because they were exposed to higher levels of oxidants than the higher altitude air masses.

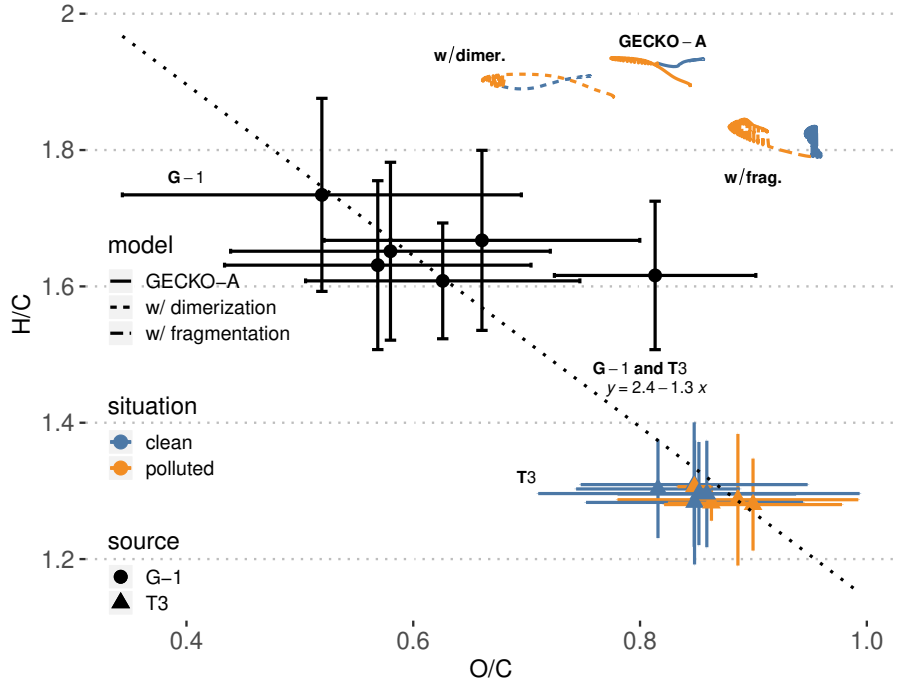


Figure 10. T3 site (colored triangles), airborne (black dots) and modeled (lines, afternoon of second day) van Krevelen diagrams of H/C (y-axis) vs O/C (x-axis) average ratios in SOA. The vertical and horizontal range of the experimental data denotes the standard deviation of measured concentrations. Cases are identified as clean (blue) and polluted (orange). Airborne data were filtered to only include measurements taken within 20 km of the T3 site. The dotted line and the associated equation depict the linear regression obtained with all experimental points (T3 and G-1). Modeled lines depict three different calculations (see Sect. 4.2.4): the reference calculation (continuous lines, labeled GECKO-A), a calculation where all C_{10} are supposed to be dimerized (short dashes, labeled w/ dimer.) and a calcualtion where all C_{10} are supposed to fragment (long dashes, labeled w/ frag.)

The modeled average particle phase O/C ratios range from ~~0.78 to 0.88~~0.77 to 0.86, within the ratios measured at the T3 site. Modeled H/C ratios are however overestimated compared to T3 site measurements, ranging from ~~1.93 to 1.96~~1.89 to 1.94. Claflin and Ziemann (2018) reported experimental evidence that the reaction of β-pinene with NO_3 produces oligomers derived from β-pinene C_{10} oxidation products. For instance one of the proposed mechanisms for dimerization of a $C_{10}H_{17}O_5$ (H/C = 1.7) produces a $C_{20}H_{30}O_9$ (H/C = 1.5). In the GECKO-A modeled aerosol phase, after organosulfate and nitrates derived from isoprene, C_{10} compounds dominate OA composition. As examples, a $C_{10}H_{~~18~~20}O_{~~7~~6}$ (H/C = ~~1.8~~2; O/C = ~~0.7~~0.6) and a $C_{10}H_{~~20~~18}O_{~~6~~7}$ (H/C = ~~2~~1.8; O/C = ~~0.6~~0.7) derived from limonene are the ~~third and fourth~~ second and third most important organic species in the aerosol phase on a molecule basis. Following the dimerization pathways suggested by Claflin and Ziemann (2018), these compounds could potentially form $C_{20}H_{~~32~~36}O_{~~13~~11}$ (H/C = ~~1.6~~1.8; O/C = ~~0.65~~0.55) and $C_{20}H_{~~36~~32}O_{~~11~~13}$ (H/C = ~~1.8~~1.6; O/C = ~~0.55~~0.65) dimers respectively. Dimerization, or similar oligomerization processes, would then possibly move the modeled van Krevelen diagram towards lower H/C ratios, closer to AMS measurements.

As a test, we generalized this estimation to all C_{10} in the aerosol phase: we replaced each C_{10} by the corresponding C_{20} and halved its concentration. In this way, we can calculate what would H/C and O/C ratios be in the aerosol phase if aging processes only dimerized C_{10} compounds. The resulting modeled van Krevelen diagram is reported on Fig. 10 (labeled w/ dimer.). The impact of C_{10} dimerization is relatively strong on O/C ratio, ranging from 0.66 to 0.78 and remaining in the range of measured O/C ratios at T3 site and in the aircraft. H/C ratios are only reduced to 1.88–1.94, still 50% higher than measured H/C at the T3 site and 20% higher than airborne data.

Oppositely, GECKO-A could be missing processes that would fragment ~~these~~ the aforementioned two C_{10} compounds. Fragmenting $C_{10}H_{18}O_7$ into a $C_4H_6O_4$ (H/C = 1.5; O/C = 1) and a $C_6H_{10}O_5$ (H/C = 1.7; O/C = 0.8) species would bring the average H/C ratio down from 1.8 to 1.6. This possibility of missing fragmentation processing means that either the modeled gas phase chemistry doesn't compete enough with condensation to fragment these species, or these C_{10} species should be fragmented by heterogeneous or condensed phase processes in the particles themselves, which are not accounted for by the model. It should be noted that because the fragmented compounds are lighter, they would exhibit higher volatility. However this does not necessarily mean that the SOA mass would decrease because these shorter species are still oxygenated, maybe enough to contribute to SOA mass through solubility controlled processes in the same fashion as what is known about isoprene oxidation products.

As another test, we also estimated what would O/C and H/C ratios be if all C_{10} fragmented in the aerosol phase. The resulting modeled van Krevelen diagram is reported on Fig. 10 (labeled w/ frag.). In this case, modeled O/C ratios increase to a range of 0.88 to 0.96 and remain in the higher end of measured ratio at the T3 site. H/C are reduced further than in the dimerization test and sit at the higher end of airborne measured H/C ratios, but they still are 45% higher than H/C ratios measured at the T3 site.

Even if they apparently cannot account for the discrepancy between modeled and measured H/C ratios, the two tests presented here on C_{10} compounds in the aerosol phase show the potential importance of adding these missing processes in GECKO-A. These simple tests are however simplifications that overlook important factors in the potential impact on SOA composition: (i) not all C_{10} compounds would be affected by these processes, (ii) other compounds than C_{10} could react in a similar way, (iii) trimerization, tetramerization and other accretion processes could also occur in the aerosol phase, (iv) missing fragmentation processes could also happen in the gas phase.

4.3 Comparison with VBS approach

Shrivastava et al. (2019) modeled this same field campaign with WRF-Chem, a chemistry transport regional model (Grell et al., 2005) and similarly to this work they based their primary organic compounds emissions on the MEGAN inventory (Guenther et al., 2012) for biogenic compounds, and the methodology described in Andrade et al. (2015) and data from Medeiros et al. (201 for anthropogenic emissions. Using a Volatility Basis Set approach (VBS) to account for condensation of low volatility species, and considering ISOPSOA separately with an approach similar to this work, they modeled airborne SOA mass to within 15% of airborne measurements. The VBS parameterization described in Shrivastava et al. (2019) represents the formation of SOA as four surrogate species differing by their volatility ($C^\star$ = 0.1, 1, 10 and 100 µg m^{-3}). For biogenic SOA, isoprene and monoter-

penes produce these four surrogates from the oxidation by OH, ozone and NO_3, with yields depending on NO_x. Moreover multigenerational aging is accounted for the surrogate species assigning fragmentation (*i.e.* increasing volatility) and functionalization (*i.e.* decreasing volatility). This aging is parameterized as a reaction of each of the SOA surrogate species VBS_n with OH as follows:

$$VBS_n + OH \rightarrow \alpha_{frag} VBS_{n+1} + (1-\alpha_{frag}) VBS_{n-1} \quad \text{(R1)}$$

The reaction rate is $k_{R1} = 2 \times 10^{-11}$ cm^3 molec^{-1} s^{-1}. The branching ratio for fragmentation α_{frag} is determined as the ratio of the reaction rate of peroxy radicals with NO to the sum of all peroxy radical reactions rates; it has an upper limit of 75%. The yields used in this VBS approach were fitted over a variety of low OA loading atmospheric chamber studies of biogenics oxidation under high and low NO_x concentrations (Shrivastava et al., 2019). More details about this VBS approach can be found in Shrivastava et al. (2013, 2015, 2019).

In order to compare the GECKO-A model results with the VBS approach used in Shrivastava et al. (2019), additional simulations were run where the explicit condensation of low volatility biogenic species was replaced by the formation of the four surrogate species used in Shrivastava et al. (2019). Fig. 7 shows the time evolution of predicted SOA mass with GECKO-A, after replacing the original condensation of low volatility biogenic species by the VBS approach used in Shrivastava et al. (2019) (dashed lines). In this test, the VBS modeled SOA mass is well within the range of measured values in the afternoon for the polluted case scenario. The VBS version of the box-model is however underestimating SOA mass concentrations in the clean situation, with only 0.5 µg m^{-3} during daytime compared to the measured 0.6 to 2.5 µg m^{-3} range. Like in Shrivastava et al. (2019), exposure of the background air mass to the urban increased oxidative capacity increases VBS predicted SOA mass by almost 400%, which explains how the VBS can reach the higher polluted case SOA mass. Figure 7 also depicts the predicted SOA mass if SOA aging is not included in the VBS model (dotted lines). Shrivastava et al. (2019) reported that SOA aging does not have a strong effect in their simulations, which is not the case when applied in the box-model. In our simulation without aging processes, the polluted case SOA mass concentration drops below ~~1~~1.3 µg m^{-3} in the afternoon. However in the clean case scenario, the SOA mass concentration only decreases by approximately 10% when SOA aging is removed. This means that SOA aging becomes more important in the ground case scenario when the air mass is exposed to high OH concentrations that were not seen by the model run by Shrivastava et al. (2019): their maximum OH concentrations reach 2×10^6 molec cm^{-3} while our maximum OH concentration reach 1.6×10^7 molec cm^{-3}.

Figure 11 and Table 2 attribute sources of SOA according to the GECKO-A explicit simulation and the VBS approach. In the clean case scenario, GECKO-A attributes most of SOA mass to monoterpene oxidation products (~~60~~65% at 2pm). The remainder is attributed to isoprene SOA, with condensation of low volatility compounds contributing in the same proportion as reactive uptake (~~19% and 21~~17% and 18% respectively). In Shrivastava et al. (2019), monoterpene oxidation products account for 45% of SOA sources in the airborne plume. With their VBS applied to the ground situation, ~~31~~28% of SOA is attributed to monoterpenes at 2pm, approximately half of the proportion predicted by the GECKO-A explicit approach. Like in the 3D model calculation, the VBS in the box model attributes the remainder of background SOA mass mostly to reactive uptake of isoprene oxidation products (~~50~~53% of total SOA).

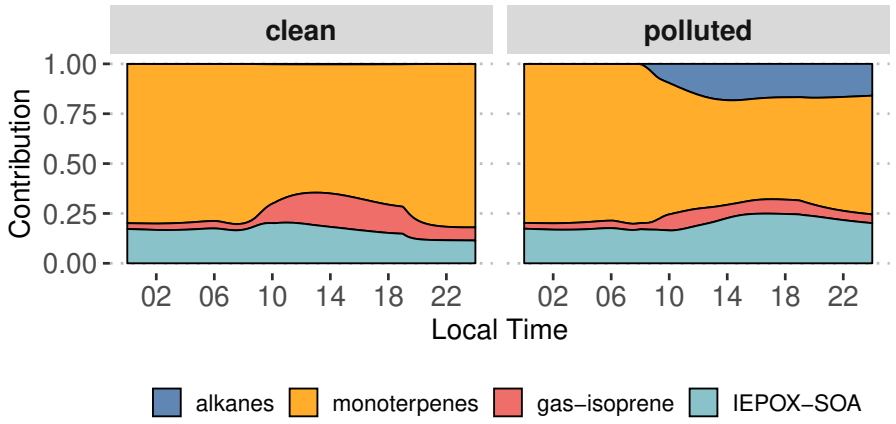


Figure 11. Contribution of primary hydrocarbons categories to GECKO-A modeled SOA mass for the clean (left panel) and polluted case (right panel).

Table 2. Contribution of primary hydrocarbons categories to modeled SOA mass at 2pm. Percentages in parentheses indicate the relative contribution to total SOA mass.

	GECKO-A		VBS - aging		VBS - no aging	
SOA mass [μg m^{-3}]	clean	polluted	clean	polluted	clean	pollute
Monoterpenes	~~0.70 (60~~1.19 (65%)	~~0.60 (44~~0.91 (53%)	~~0.15 (31~~0.18 (28%)	~~0.73 (32~~0.71 (30%)	~~0.12 (28~~0.14 (24%)	0.17 (~~181~~
Isoprene (gas)	~~0.23 (19~~0.31 (17%)	~~0.17 (13~~0.11 (6%)	~~0.09~~0.12 (19%)	1.00 (~~43~~41%)	~~0.07~~0.09 (16%)	~~0.17 (18~~0.18
IEPOX-SOA	~~0.24 (21~~0.34 (18%)	~~0.36 (27~~0.39 (23%)	~~0.24 (50~~0.34 (53%)	~~0.36~~0.39 (16%)	~~0.24 (56~~0.34 (60%)	~~0.36 (40~~0.39
biogenics	~~1.17~~1.84 (100%)	~~1.13 (84~~1.41 (82%)	~~0.48~~0.64 (100%)	~~2.09 (91~~2.1 (87%)	~~0.43~~0.57 (100%)	~~0.70(76~~0.74
anthropogenics	0 (0%)	~~0.22 (16~~0.32 (18%)	0 (0%)	~~0.22 (9~~0.32 (13%)	0 (0%)	~~0.22 (24~~0.32
total	~~1.17~~ 1.84	~~1.35~~ 1.73	~~0.48~~ 0.64	~~2.31~~ 2.42	~~0.43~~ 0.57	~~0.92~~ 1.0

[(a)] de Sá et al. (2018)

In the polluted case, the explicit model predicts ~~an increase of 15~~a slight decrease of 6% in total SOA at 2pm while measurements exhibit an increase of 33% on average. The urban effect is ~~stronger~~ strong in the VBS case than the explicit approach with a 380% increase in mass. In the comparison with airborne measurements, the Shrivastava et al. (2019) model predicts that the city oxidants cause the same large increase in biogenic SOA formation (up to 400%), and that this increase is due to enhanced monoterpene oxidation. With GECKO-A at the ground site, ~~most of the SOA increase is due to~~ SOA mass remains constant because of the contribution of anthropogenics ~~. In fact biogenic SOA remains constant (3% decrease) with respect to the clean case. In particular biogenic SOA produced by condensation of low volatility oxidation products of monoterpenes and isoprene~~

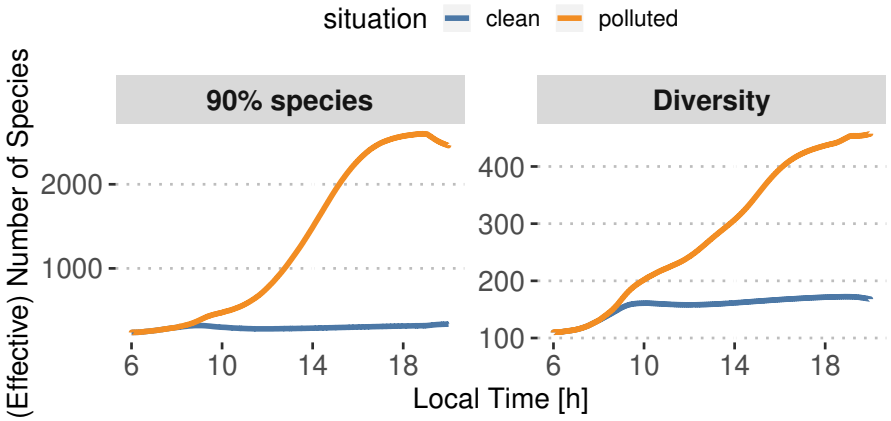


Figure 12. ~~Minimum~~ Smallest number of species needed to ~~model~~ capture 90% of modeled SOA mass (left panel) with GECKO-A ~~.~~ ~~Statistical~~ at each timestep ($N_{90\%}$, see text) and statistical diversity D in the GECKO-A modeled particle phase (right panel, see Eq. 3~~(right panel)~~).

~~decreases by 17~~which compensates the decrease in the contribution from the condensation of isoprene and monoterpenes oxidation products by 32%. This loss is slightly compensated for by an increase in the production of SOA via reactive uptake of isoprene oxidation products (~~50~~15% increase) because the plume favors these processes with higher sulfate load and lower pH (see Table 1). Overall biogenic SOA decreases by 23% with respect to the clean case. In the VBS test case, ~~the~~ SOA mass formed from condensation of low volatility oxidation products of isoprene and monoterpenes is enhanced in the polluted case respectively by a factor of ~~11 and 4.9.~~ 7 and 3. This enhancement is notably inhibited when the aging parameterization is removed from the VBS approach with a mass increase due to condensation of low volatility products of isoprene and monoterpenes of respectively ~~42% and 143~~100% and 21%. This highlights the importance of modeling aging of low volatility oxidation products to explain the enhanced production of SOA in the urban plume.

4.4 Potential for Reduction of the Explicit GECKO-A mechanism

It is obvious that the chemical mechanisms generated with GECKO-A are too large to be implemented in 3D models. The GECKO-A mechanisms need to be reduced to sizes manageable by 3D models, typically a few hundred species and reactions. The VBS parameterization used for comparison in this work is fit for low OA loadings, biogenic dominated situations but it is unclear that it should be applied to other situations.

In this section, we are *not* proposing a much needed new approach to reducing explicit mechanisms with the goal of predicting SOA mass concentrations, but we explore here the potential for reduction of the chemical mechanism that was generated for this study. In other words, what is the theoretical lower limit to the number of species that should be used in a reduced scheme to still be able to model the same SOA mass concentration time profile as the explicit model?

To answer this, two metrics are presented in Fig. 12. The first one ~~is the minimum~~ $N_{90\%}$ is the smallest number of species needed in the explicit model to ~~represent at least~~ capture 90% of the total SOA mass at each timestep. ~~The~~ After sorting species by decreasing concentration, this number is calculated by adding up these concentrations until 90% of the total modeled SOA mass is reached. The operation is repeated at each timestep. Calculated independantly, the second one is the particle diversity D in the explicitly modeled SOA, as defined for instance in Riemer and West (2013):

$$D = \exp S \tag{3}$$

where S is the first order generalized entropy (also known as Shannon entropy):

$$S = \sum_{i=1}^{N} -p_i \ln p_i \tag{4}$$

where p_i is the mass fraction of species i in the organic particle phase and N is the total number of species in the organic particle phase. As stated in Riemer and West (2013), the diversity is a measure of the effective number of species with the same concentration in the organic fraction of the aerosol phase. If $D = 1$, the organic fraction is pure as it is composed of a single species. Therefore, a value $D \ll N$ means that of all the species contributing to modeled organic aerosol, only a few significantly contribute to its composition. Oppositely, $D = N$ is the maximum value reachable by D and is obtained when the organic fraction is composed of N equally distributed species. In the case where D is close to N, only a few species are negligible. For more details and better explanations, we refer the reader to Riemer and West (2013, esp. Fig. 1). We make the hypothesis here that D can be interpreted as an effective number of species derived from the informational entropy of the modeled particle phase.

In the clean situation both metrics behave similarly, with a morning increase of the number of species until 10 am, after which the number remains relatively constant until sunset. During daytime, on average ~~310~~ $N_{90\%} = 292$ species are needed to represent 90% of the SOA mass. The calculated diversity is around ~~170~~ 153 effective species. For the polluted situation, ~~the number of species needed to represent 90% of the SOA mass during daytime increases~~ $N_{90\%}$ increases during daytime by about a factor of ~~7~~9, reaching about ~~2300.~~ 2500. The calculated diversity only increases up to approximately ~~410~~ 260 effective species. These increases in the species numbers for the polluted case are logical as the variety of precursors, and hence secondary species that could potentially contribute to SOA, is increased by urban emissions.

The number of species needed to represent most of the modeled SOA mass in all cases seems too high to be used in 3D models applications. Furthermore there is no guarantee that the most important species at a given timestep would be the same most important species at the following timestep. This suggests that reductions should not come from simply selecting species identified as important to represent the variety of species that could arise in the interaction of biogenic air and an urban plume.

The statistical diversity calculation seems like a better approach to estimate the minimum number of species needed to model SOA mass. As this number is directly derived from informational entropy, we suggest that the diversity represents the number of species that would be needed to reproduce the same informational content regarding the ~~composition of SOA~~ time evolution of SOA mass in the explicit model. Even if the effective species numbers fall in the higher range of what would be acceptable in a 3D model chemical mechanism, the practical construction of the mechanism remains to be explored. For instance, ~~nothing~~

here says that the effective species would be individual species already present in the original explicit mechanismin the polluted scenario, D is a factor of 9 lower than $N_{90\%}$. This should mean that D cannot represent a subset of the individual species from the original mechanism, otherwise it would be expected to be equal or higher than $N_{90\%}$ if it is supposed to reproduce the informational content regarding SOA mass. It is possibletherefore likely, and making this problem more complex, that each of these effective species is a (non) linear combination of explicit individual species.

Finally, we used in this section an entropy calculation for SOA mass: it is based only on mass fractions of the species composing the modeled organic particles. The effective number of species displayed on Fig. 12 is therefore only meaningful for SOA mass and properties directly linked to it. If the goal is to predict other properties, *e.g.* hygroscopicity, toxicity or optical properties, assuming we find a way to calculate these with GECKO-A, the diversity defined here would not necessarily be meaningful. For instance, hygroscopicity or toxicity could be driven by a handful of oxygenated species that do not matter for the informational content regarding SOA mass. We did not explore further down this path, as this is not the subject of this paper, but it may be possible to generalize this definition of informational diversity to properties other than mass.

5 Conclusions

An explicit chemical mechanism generated with GECKO-A was used in a box model to simulate a situation similar to the situation studied in Manaus during the GoAmazon 2014/5 field campaign. After scaling down the emissions generated from the MEGAN biogenic emissions model and estimating urban emissions in Manaus, the model was able to reproduce realistic primary organic compounds mixing ratios as well as NO_x, ozone and OH concentrations.

The model is able to reproduce background SOA mass concentrations but is not able to reproduce the observed enhancement in the polluted plume. When running a Volatility Basis Set approach that was previously applied to the Manaus case (Shrivastava et al., 2019), modeled SOA mass matches measurements which suggests that the incorrect explicit model prediction is not caused by incorrect primary organic compound emissions or oxidant levels. Modeled particle phase organosulfates are within the range of previous measurements (Glasius et al., 2018) which suggests that isoprene oxidation and SOA formation in the model are reasonably well simulated. In both polluted and clean situations, biogenics are identified as the main contributors to SOA by both GECKO-A and the VBS parameterization. In both approaches, the majority of SOA production is attributed to monoterpenes oxidation and condensation of lower volatility products. Yee et al. (2018) measured and described sesquiterpenes during GoAmazon 2014/5 for the same situations and suggested that these species may be important for modeling studies. However the modeling study of Shrivastava et al. (2019) estimated that the contribution of sesquiterpenes to SOA production is less than 10%. It is more likely that physico-chemical processes involved in monoterpene SOA formation are either unknown or missing in the explicit model. Comparison of modeled and measured elemental ratios (H/C and O/C) indicates that fragmentation of monoterpenes oxidation products and their condensation or reactive uptake to the condensed phase may play an important role in understanding the sources of biogenic SOA mass. This reactive uptake may in turn involve oligomerization and fragmentation processes. However, simple sensitivity tests show that these processes alone may not explain the discrepancies between the explicit model and measurements. Because the VBS parameterization is based on multiple chamber

experiments, it could implicitly be accounting for these missing processes. Of the high diversity of monoterpenes identified in Amazonia (Jardine et al., 2015), only a handful of monoterpenes have been studied to the extent that we can be as confident in model predictions of SOA formation from monoterpenes as from isoprene. Detailed mechanistic studies of monoterpene oxidation are therefore needed for further incorporation in explicit models to better understand the nature and the magnitude of the contribution of monoterpenes to SOA formation, as well as their response to the interaction with urban pollution (*e.g.* Claflin and Ziemann, 2018).

Even if a parameterization was implemented in GECKO-A to properly address the formation of isoprene SOA via aqueous phase processes (Marais et al., 2016), to explicitly treat these in a more general way, future GECKO-A developments for mechanism generation will need to include the following: (i) aerosol thermodynamics, for instance via coupling with a model like MOSAIC (Zaveri et al., 2008) or ISORROPIA (Nenes et al., 1998), (ii) aqueous phase processes including explicit dissolution (*e.g.* Mouchel-Vallon et al., 2013), oxidation (*e.g.* Mouchel-Vallon et al., 2017), accretion reactions (*e.g.* Renard et al., 2015), and interaction with dissolved inorganic ions, (iii) explicit treatment of the fate of newly formed species like dimers or organo-sulfates.

One could be tempted to think that since the VBS parameterization is behaving particularly well in this GoAmazon 2014/5 case, it could be the answer to predict SOA mass in larger scale 3D models. However this approach is limited by the fact that it was fitted for low biogenic OA loading situations and was run in a limited domain regional model (Shrivastava et al., 2019). One possible way of building reduced mechanisms is to reduce existing detailed chemical mechanisms to sizes manageable by 3D models (e.g. Szopa et al., 2005; Kaduwela et al., 2015). Using an information theory based approach, we provide here a lower limit to the size of these reduced mechanisms, assuming the goal is to produce the same informational content as the explicit mechanism. This lower limit of a few hundred species is four orders of magnitudes lower than the actual number of species that are actually accounted for in the explicit mechanism (4×10^6) and shows the potential for progress in future mechanism reduction endeavors. Even if a direct application of this statistical approach to create a reduced mechanism would likely require some atmospheric chemistry breakthrough, it could at least currently be used as a statistical indicator for comparing reduced mechanisms with reference explicit mechanisms.

Competing interests. The authors declare no competing interests.

Acknowledgements. The National Center for Atmospheric Research is sponsored by the National Science Foundation. We gratefully acknowledge support from U.S. Department of Energy (DOE) ASR grant DE-SC0016331. JLJ and BBP were supported by NSF AGS-1822664 and EPA 83587701-0. This manuscript has not been reviewed by EPA, and thus no endorsement should be inferred. Dr. Shrivastava was also supported by the U.S. DOE, Office of Science, Office of Biological and Environmental Research through the Early Career Research Program. Data were obtained from the Atmospheric Radiation Measurement (ARM) User Facility, a U.S. DOE Office of Science user facility managed by the Office of Biological and Environmental Research. The research was conducted under scientific license 001030/2012-4 of the Brazilian

National Council for Scientific and Technological Development (CNPq). We are grateful to Louisa K. Emmons for providing the MEGAN emissions data; and Suzane S. de Sà for providing the clustering analysis results. We are thanking Siyuan Wang for helpful comments.

References

Abou Rafee, S. A., Martins, L. D., Kawashima, A. B., Almeida, D. S., Morais, M. V. B., Souza, R. V. A., Oliveira, M. B. L., Souza, R. A. F., Medeiros, A. S. S., Urbina, V., Freitas, E. D., Martin, S. T., and Martins, J. A.: Contributions of mobile, stationary and biogenic sources to air pollution in the Amazon rainforest: a numerical study with the WRF-Chem model, Atmospheric Chemistry and Physics, 17, 7977–7995, https://doi.org/10.5194/acp-17-7977-2017, 2017.

Altieri, K. E., Carlton, A. G., Lim, H.-J., Turpin, B. J., and Seitzinger, S. P.: Evidence for oligomer formation in clouds: reactions of isoprene oxidation products., Environmental science & technology, 40, 4956–60, 2006.

Alves, E. G., Jardine, K., Tota, J., Jardine, A., Yãnez-Serrano, A. M., Karl, T., Tavares, J., Nelson, B., Gu, D., Stavrakou, T., Martin, S., Artaxo, P., Manzi, A., and Guenther, A.: Seasonality of isoprenoid emissions from a primary rainforest in central Amazonia, Atmospheric Chemistry and Physics, 16, 3903–3925, https://doi.org/10.5194/acp-16-3903-2016, 2016.

Andrade, M. D. F., Ynoue, R. Y., Freitas, E. D., Todesco, E., Vara Vela, A., Ibarra, S., Martins, L. D., Martins, J. A., and Carvalho, V. S. B.: Air quality forecasting system for Southeastern Brazil, Frontiers in Environmental Science, 3, 6975, https://doi.org/10.3389/fenvs.2015.00009, 2015.

Aumont, B., Szopa, S., and Madronich, S.: Modelling the evolution of organic carbon during its gas-phase tropospheric oxidation: development of an explicit model based on a self generating approach, Atmospheric Chemistry and Physics, 5, 2497–2517, https://doi.org/10.5194/acp-5-2497-2005, 2005.

Aumont, B., Valorso, R., Mouchel-Vallon, C., Camredon, M., Lee-Taylor, J., and Madronich, S.: Modeling SOA formation from the oxidation of intermediate volatility n-alkanes, Atmospheric Chemistry and Physics, 12, 7577–7589, https://doi.org/10.5194/acp-12-7577-2012, https://www.atmos-chem-phys.net/12/7577/2012/, 2012.

Aumont, B., Camredon, M., Mouchel-Vallon, C., La, S., Ouzebidour, F., Valorso, R., Lee-Taylor, J., and Madronich, S.: Modeling the influence of alkane molecular structure on secondary organic aerosol formation, Faraday Discussions, 165, 105, https://doi.org/10.1039/c3fd00029j, 2013.

Batista, C. E., Ye, J., Ribeiro, I. O., Guimarães, P. C., Medeiros, A. S. S., Barbosa, R. G., Oliveira, R. L., Duvoisin, S., Jardine, K. J., Gu, D., Guenther, A. B., McKinney, K. A., Martins, L. D., Souza, R. A. F., and Martin, S. T.: Intermediate-scale horizontal isoprene concentrations in the near-canopy forest atmosphere and implications for emission heterogeneity, Proceedings of the National Academy of Sciences, p. 201904154, https://doi.org/10.1073/pnas.1904154116, 2019.

Bezdek, J. C.: Pattern recognition with fuzzy objective function algorithms, Plenum, New York, 1981.

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences, 10, 191–203, https://doi.org/10.1016/0098-3004(84)90020-7, 1984.

Blando, J. D. and Turpin, B. J.: Secondary organic aerosol formation in cloud and fog droplets: a literature evaluation of plausibility, Atmospheric Environment, 34, 1623–1632, 2000.

Browell, E. V., Gregory, G. L., Harriss, R. C., and Kirchhoff, V. W. J. H.: Ozone and aerosol distributions over the Amazon Basin during the wet season, Journal of Geophysical Research, 95, 16 887, https://doi.org/10.1029/JD095iD10p16887, 1990.

Budisulistiorini, S. H., Baumann, K., Edgerton, E. S., Bairai, S. T., Mueller, S., Shaw, S. L., Knipping, E. M., Gold, A., and Surratt, J. D.: Seasonal characterization of submicron aerosol chemical composition and organic aerosol sources in the southeastern United States: Atlanta, Georgia, and Look Rock, Tennessee, Atmospheric Chemistry and Physics, 16, 5171–5189, https://doi.org/10.5194/acp-16-5171-2016, 2016.

Camredon, M. and Aumont, B.: Assessment of vapor pressure estimation methods for secondary organic aerosol modeling, Atmospheric Environment, 40, 2105–2116, https://doi.org/10.1016/j.atmosenv.2005.11.051, 2006.

Camredon, M., Aumont, B., Lee-Taylor, J., and Madronich, S.: The SOA/VOC/NOx system: an explicit model of secondary organic aerosol formation, Atmospheric Chemistry and Physics, 7, 5599–5610, https://doi.org/10.5194/acp-7-5599-2007, 2007.

Carlton, A. G., Wiedinmyer, C., and Kroll, J. H.: A review of Secondary Organic Aerosol (SOA) formation from isoprene, Atmospheric Chemistry and Physics, 9, 4987–5005, https://doi.org/10.5194/acp-9-4987-2009, 2009.

Chen, Q., Farmer, D. K., Rizzo, L. V., Pauliquevis, T., Kuwata, M., Karl, T. G., Guenther, A., Allan, J. D., Coe, H., Andreae, M. O., Pöschl, U., Jimenez, J. L., Artaxo, P., and Martin, S. T.: Submicron particle mass concentrations and sources in the Amazonian wet season (AMAZE-08), Atmospheric Chemistry and Physics, 15, 3687–3701, https://doi.org/10.5194/acp-15-3687-2015, https://www.atmos-chem-phys.net/15/3687/2015/, 2015a.

Chen, Q., Heald, C. L., Jimenez, J. L., Canagaratna, M. R., Zhang, Q., He, L.-y., Huang, X.-f., Campuzano-jost, P., Palm, B. B., Poulain, L., Kuwata, M., Martin, S. T., Abbatt, J. P. D., Lee, A. K. Y., and Liggio, J.: Elemental composition of organic aerosol: The gap between ambient and laboratory measurements, Geophys. Res. Lett., 42, 4182–4189, https://doi.org/10.1002/2015GL063693.Received, 2015b.

Claeys, M.: Formation of Secondary Organic Aerosols Through Photooxidation of Isoprene, Science, 303, 1173–1176, https://doi.org/10.1126/science.1092805, 2004.

Claflin, M. S. and Ziemann, P. J.: Identification and Quantitation of Aerosol Products of the Reaction of β-Pinene with NO 3 Radicals and Implications for Gas- and Particle-Phase Reaction Mechanisms, The Journal of Physical Chemistry A, p. acs.jpca.8b00692, https://doi.org/10.1021/acs.jpca.8b00692, 2018.

Computational and Information Systems Laboratory: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing), https://doi.org/10.5065/D6RX99HX, 2017.

Daumit, K. E., Carrasquillo, A. J., Sugrue, R. A., and Kroll, J. H.: Effects of Condensed-Phase Oxidants on Secondary Organic Aerosol Formation, The Journal of Physical Chemistry A, 120, 1386–1394, https://doi.org/10.1021/acs.jpca.5b06160, 2016.

de Sá, S. S., Palm, B. B., Campuzano-jost, P., Day, D. A., Newburn, M. K., Hu, W., Isaacman-VanWertz, G., Yee, L. D., Thalman, R., Brito, J., Carbone, S., Artaxo, P., Goldstein, A. H., Manzi, A. O., Souza, R. A. F., Mei, F., Shilling, J. E., Springston, S. R., Wang, J., Surratt, J. D., Alexander, M. L., Jimenez, J. L., and Martin, S. T.: Influence of urban pollution on the production of organic particulate matter from isoprene epoxydiols in central Amazonia, Atmospheric Chemistry and Physics, 17, 6611–6629, https://doi.org/10.5194/acp-17-6611-2017, 2017.

de Sá, S. S., Palm, B. B., Campuzano-Jost, P., Day, D. A., Hu, W., Isaacman-VanWertz, G., Yee, L. D., Brito, J., Carbone, S., Ribeiro, I. O., Cirino, G. G., Liu, Y., Thalman, R., Sedlacek, A., Funk, A., Schumacher, C., Shilling, J. E., Schneider, J., Artaxo, P., Goldstein, A. H., Souza, R. A. F., Wang, J., McKinney, K. A., Barbosa, H., Alexander, M. L., Jimenez, J. L., and Martin, S. T.: Urban influence on the concentration and composition of submicron particulate matter in central Amazonia, Atmospheric Chemistry and Physics, 18, 12 185–12 206, https://doi.org/10.5194/acp-18-12185-2018, 2018.

de Sá, S. S., Rizzo, L. V., Palm, B. B., Campuzano-Jost, P., Day, D. A., Yee, L. D., Wernis, R., Isaacman-VanWertz, G., Brito, J., Carbone, S., Liu, Y. J., Sedlacek, A., Springston, S., Goldstein, A. H., Barbosa, H. M. J., Alexander, M. L., Artaxo, P., Jimenez, J. L., and Martin, S. T.: Contributions of biomass-burning, urban, and biogenic emissions to the concentrations and light-absorbing properties of particulate matter in central Amazonia during the dry season, Atmospheric Chemistry and Physics, 19, 7973–8001, https://doi.org/10.5194/acp-19-7973-2019, 2019.

DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K. S., Worsnop, D. R., and Jimenez, J. L.: Field-Deployable, High-Resolution, Time-of-Flight Aerosol Mass Spectrometer, Analytical Chemistry, 78, 8281–8289, https://doi.org/10.1021/ac061249n, 2006.

Donahue, N. M., Robinson, A. L., Stanier, C. O., and Pandis, S. N.: Coupled Partitioning, Dilution, and Chemical Aging of Semivolatile Organics, Environmental Science & Technology, 40, 2635–2643, https://doi.org/10.1021/es052297c, 2006.

Ervens, B., Turpin, B. J., and Weber, R. J.: Secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): a review of laboratory, field and model studies, Atmospheric Chemistry and Physics, 11, 11 069–11 102, 2011.

Gentner, D. R., Isaacman, G., Worton, D. R., Chan, A. W. H., Dallmann, T. R., Davis, L., Liu, S., Day, D. A., Russell, L. M., Wilson, K. R., Weber, R., Guha, A., Harley, R. A., and Goldstein, A. H.: Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions, Proceedings of the National Academy of Sciences, 109, 18 318–18 323, https://doi.org/10.1073/pnas.1212272109, 2012.

Gentner, D. R., Jathar, S. H., Gordon, T. D., Bahreini, R., Day, D. A., El Haddad, I., Hayes, P. L., Pieber, S. M., Platt, S. M., De Gouw, J., Goldstein, A. H., Harley, R. A., Jimenez, J. L., Prévôt, A. S., and Robinson, A. L.: Review of Urban Secondary Organic Aerosol Formation from Gasoline and Diesel Motor Vehicle Emissions, Environmental Science and Technology, 51, 1074–1093, https://doi.org/10.1021/acs.est.6b04509, 2017.

Glasius, M., Bering, M. S., Yee, L. D., De Sá, S. S., Isaacman-VanWertz, G., Wernis, R. A., Barbosa, H. M., Alexander, M. L., Palm, B. B., Hu, W., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Shrivastava, M., Martin, S. T., and Goldstein, A. H.: Organosulfates in aerosols downwind of an urban region in central Amazon, Environmental Science: Processes and Impacts, 20, 1546–1558, https://doi.org/10.1039/c8em00413g, 2018.

Gregory, G. L., Browell, E. V., Warren, L. S., and Hudgins, C. H.: Amazon Basin ozone and aerosol: Wet season observations, Journal of Geophysical Research, 95, 16 903, https://doi.org/10.1029/JD095iD10p16903, 1990.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957–6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions, Geoscientific Model Development, 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Heald, C. L. and Kroll, J. H.: The fuel of atmospheric chemistry: Toward a complete description of reactive organic carbon, https://doi.org/10.1126/sciadv.aay8967, 2020.

Isaacman, G., Kreisberg, N. M., Yee, L. D., Worton, D. R., Chan, A. W., Moss, J. A., Hering, S. V., and Goldstein, A. H.: Online derivatization for hourly measurements of gas- and particle-phase semi-volatile oxygenated organic compounds by thermal desorption aerosol gas chromatography (SV-TAG), Atmospheric Measurement Techniques, 7, 4417–4429, https://doi.org/10.5194/amt-7-4417-2014, 2014.

Isaacman-Vanwertz, G., Massoli, P., O'Brien, R., Lim, C., Franklin, J. P., Moss, J. A., Hunter, J. F., Nowak, J. B., Canagaratna, M. R., Misztal, P. K., Arata, C., Roscioli, J. R., Herndon, S. T., Onasch, T. B., Lambe, A. T., Jayne, J. T., Su, L., Knopf, D. A., Goldstein, A. H., Worsnop, D. R., and Kroll, J. H.: Chemical evolution of atmospheric organic carbon over multiple generations of oxidation, Nature Chemistry, 10, 462–468, https://doi.org/10.1038/s41557-018-0002-2, 2018.

Jardine, A. B., Jardine, K. J., Fuentes, J. D., Martin, S. T., Martins, G., Durgante, F., Carneiro, V., Higuchi, N., Manzi, A. O., and Chambers, J. Q.: Highly reactive light-dependent monoterpenes in the Amazon, Geophysical Research Letters, 42, 1576–1583, https://doi.org/10.1002/2014GL062573, 2015.

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: A protocol for mechanism development, Atmospheric Environment, 31, 81–104, 1997.

Jenkin, M. E., Young, J. C., and Rickard, a. R.: The MCM v3.3.1 degradation scheme for isoprene, Atmospheric Chemistry and Physics, 15, 11 433–11 459, https://doi.org/10.5194/acp-15-11433-2015, 2015.

Jo, D. S., Hodzic, A., Emmons, L. K., Marais, E. A., Peng, Z., Nault, B. A., Hu, W., Campuzano-Jost, P., and Jimenez, J. L.: A simplified parameterization of isoprene-epoxydiol-derived secondary organic aerosol (IEPOX-SOA) for global chemistry and climate models: a case study with GEOS-Chem v11-02-rc, Geoscientific Model Development, 12, 2983–3000, https://doi.org/10.5194/gmd-12-2983-2019, 2019.

Jordan, A., Haidacher, S., Hanel, G., Hartungen, E., Herbig, J., Märk, L., Schottkowsky, R., Seehauser, H., Sulzer, P., and Märk, T.: An online ultra-high sensitivity Proton-transfer-reaction mass-spectrometer combined with switchable reagent ion capability (PTR+SRI-MS), International Journal of Mass Spectrometry, 286, 32–38, https://doi.org/10.1016/j.ijms.2009.06.006, 2009a.

Jordan, A., Haidacher, S., Hanel, G., Hartungen, E., Märk, L., Seehauser, H., Schottkowsky, R., Sulzer, P., and Märk, T.: A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS), International Journal of Mass Spectrometry, 286, 122–128, https://doi.org/10.1016/j.ijms.2009.07.005, 2009b.

Kaduwela, A., Luecken, D., Carter, W., and Derwent, R.: New directions: Atmospheric chemical mechanisms for the future, Atmospheric Environment, 122, 609–610, https://doi.org/10.1016/j.atmosenv.2015.10.031, 2015.

Kirchhoff, V. W. J. H., da Silva, I. M. O., and Browell, E. V.: Ozone measurements in Amazonia: Dry season versus wet season, Journal of Geophysical Research, 95, 16 913, https://doi.org/10.1029/JD095iD10p16913, 1990.

Koss, A., Yuan, B., Warneke, C., Gilman, J. B., Lerner, B. M., Veres, P. R., Peischl, J., Eilerman, S., Wild, R., Brown, S. S., Thompson, C. R., Ryerson, T., Hanisco, T., Wolfe, G. M., St Clair, J. M., Thayer, M., Keutsch, F. N., Murphy, S., and De Gouw, J.: Observations of VOC emissions and photochemical products over US oil- and gas-producing regions using high-resolution H3O+CIMS (PTR-ToF-MS), Atmospheric Measurement Techniques, 10, 2941–2968, https://doi.org/10.5194/amt-10-2941-2017, 2017a.

Koss, A. R., Sekimoto, K., Gilman, J. B., Selimovic, V., Coggon, M. M., Zarzana, K. J., Yuan, B., Lerner, B. M., Brown, S. S., Jimenez, J. L., Krechmer, J., Roberts, J. M., Warneke, C., Yokelson, R. J., and de Gouw, J.: Non-methane organic gas emissions from biomass burning: identification, quantification, and emission factors from PTR-ToF during the FIREX 2016 laboratory experiment, Atmospheric Chemistry and Physics Discussions, pp. 1–44, https://doi.org/10.5194/acp-2017-924, 2017b.

La, Y. S., Camredon, M., Ziemann, P. J., Valorso, R., Matsunaga, A., Lannuque, V., Lee-Taylor, J., Hodzic, A., Madronich, S., and Aumont, B.: Impact of chamber wall loss of gaseous organic compounds on secondary organic aerosol formation: explicit modeling of SOA formation from alkane and alkene oxidation, Atmospheric Chemistry and Physics, 16, 1417–1431, https://doi.org/10.5194/acp-16-1417-2016, https://www.atmos-chem-phys.net/16/1417/2016/, 2016.

Lee-Taylor, J., Madronich, S., Aumont, B., Baker, A., Camredon, M., Hodzic, A., Tyndall, G. S., Apel, E., and Zaveri, R. a.: Explicit modeling of organic chemistry and secondary organic aerosol partitioning for Mexico City and its outflow plume, Atmospheric Chemistry and Physics, 11, 13 219–13 241, https://doi.org/10.5194/acp-11-13219-2011, 2011.

Lee-Taylor, J., Hodzic, A., Madronich, S., Aumont, B., Camredon, M., and Valorso, R.: Multiday production of condensing organic aerosol mass in urban and forest outflow, Atmospheric Chemistry and Physics, 15, 595–615, https://doi.org/10.5194/acp-15-595-2015, 2015.

Lenschow, D. H., Gurarie, D., and Patton, E. G.: Modeling the diurnal cycle of conserved and reactive species in the convective boundary layer using SOMCRUS, Geoscientific Model Development, 9, 979–996, https://doi.org/10.5194/gmd-9-979-2016, 2016.

Lim, Y. B. and Ziemann, P. J.: Effects of molecular structure on aerosol yields from OH radical-initiated reactions of linear, branched, and cyclic alkanes in the presence of NO x, Environmental Science and Technology, 43, 2328–2334, https://doi.org/10.1021/es803389s, 2009.

Liu, Y., Siekmann, F., Renard, P., El Zein, A., Salque, G., El Haddad, I., Temime-Roussel, B., Voisin, D., Thissen, R., and Monod, A.: Oligomer and SOA formation through aqueous phase photooxidation of methacrolein and methyl vinyl ketone, Atmospheric Environment, 49, 123–129, https://doi.org/10.1016/j.atmosenv.2011.12.012, 2012.

Liu, Y., Seco, R., Kim, S., Guenther, A. B., Goldstein, A. H., Keutsch, F. N., Springston, S. R., Watson, T. B., Artaxo, P., Souza, R. A., McKinney, K. A., and Martin, S. T.: Isoprene photo-oxidation products quantify the effect of pollution on hydroxyl radicals over Amazonia, Science Advances, 4, 1–9, https://doi.org/10.1126/sciadv.aar2547, 2018.

Marais, E. A., Jacob, D. J., Jimenez, J. L., Campuzano-Jost, P., Day, D. A., Hu, W., Krechmer, J., Zhu, L., Kim, P. S., Miller, C. C., Fisher, J. A., Travis, K., Yu, K., Hanisco, T. F., Wolfe, G. M., Arkinson, H. L., Pye, H. O. T., Froyd, K. D., Liao, J., and McNeill, V. F.: Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: application to the southeast United States and co-benefit of SO2 emission controls, Atmospheric Chemistry and Physics, 16, 1603–1618, https://doi.org/10.5194/acp-16-1603-2016, 2016.

Martin, S. T., Andreae, M. O., Artaxo, P., Baumgardner, D., Chen, Q., Goldstein, A. H., Guenther, A., Heald, C. L., Mayol-Bracero, O. L., McMurry, P. H., Pauliquevis, T., Pöschl, U., Prather, K. A., Roberts, G. C., Saleska, S. R., Silva Dias, M. A., Spracklen, D. V., Swietlicki, E., and Trebs, I.: Sources and properties of Amazonian aerosol particles, Reviews of Geophysics, 48, RG2002, https://doi.org/10.1029/2008RG000280, 2010.

Martin, S. T., Artaxo, P., Machado, L. A. T., Manzi, A. O., Souza, R. A. F., Schumacher, C., Wang, J., Andreae, M. O., Barbosa, H. M. J., Fan, J., Fisch, G., Goldstein, A. H., Guenther, A., Jimenez, J. L., Pöschl, U., Silva Dias, M. A., Smith, J. N., and Wendisch, M.: Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5), Atmospheric Chemistry and Physics, 16, 4785–4797, https://doi.org/10.5194/acp-16-4785-2016, 2016.

Martins, L. D., Andrade, M. F., Freitas, E. D., Pretto, A., Gatti, L. V., Albuquerque, É. L., Tomaz, E., Guardani, M. L., Martins, M. H. R. B., and Junior, O. M. A.: Emission factors for gas-powered vehicles traveling through road tunnels in São Paulo, Brazil, Environmental Science and Technology, 40, 6722–6729, https://doi.org/10.1021/es052441u, 2006.

McNeill, V. F., Woo, J. L., Kim, D. D., Schwier, A. N., Wannell, N. J., Sumner, A. J., and Barakat, J. M.: Aqueous-Phase Secondary Organic Aerosol and Organosulfate Formation in Atmospheric Aerosols: A Modeling Study, Environmental Science & Technology, 46, 8075–8081, https://doi.org/10.1021/es3002986, 2012.

Medeiros, A. S., Calderaro, G., Guimarães, P. C., Magalhaes, M. R., Morais, M. V., Rafee, S. A., Ribeiro, I. O., Andreoli, R. V., Martins, J. A., Martins, L. D., Martin, S. T., and Souza, R. A.: Power plant fuel switching and air quality in a tropical, forested environment, Atmospheric Chemistry and Physics, 17, 8987–8998, https://doi.org/10.5194/acp-17-8987-2017, 2017.

Mouchel-Vallon, C., Bräuer, P., Camredon, M., Valorso, R., Madronich, S., Herrmann, H., and Aumont, B.: Explicit modeling of volatile organic compounds partitioning in the atmospheric aqueous phase, Atmospheric Chemistry and Physics, 13, 1023–1037, https://doi.org/10.5194/acp-13-1023-2013, 2013.

Mouchel-Vallon, C., Deguillaume, L., Monod, A., Perroux, H., Rose, C., Ghigo, G., Long, Y., Leriche, M., Aumont, B., Patryl, L., Armand, P., and Chaumerliac, N.: CLEPS 1.0: A new protocol for cloud aqueous phase oxidation of VOC mechanisms, Geoscientific Model Development, 10, 1339–1362, https://doi.org/10.5194/gmd-10-1339-2017, 2017.

Nannoolal, Y., Rarey, J., and Ramjugernath, D.: Estimation of pure component properties, Fluid Phase Equilibria, 269, 117–133, https://doi.org/10.1016/j.fluid.2008.04.020, 2008.

Nenes, A., Pilinis, C., and Pandis., S. N.: ISORROPIA: A New Thermodynamic Model for Multiphase Multicomponent Inorganic Aerosols., Aquatic Geochemistry, 4, 123–152, 1998.

Palm, B. B., de Sá, S. S., Day, D. A., Campuzano-Jost, P., Hu, W., Seco, R., Sjostedt, S. J., Park, J.-H., Guenther, A. B., Kim, S., Brito, J., Wurm, F., Artaxo, P., Thalman, R., Wang, J., Yee, L. D., Wernis, R., Isaacman-VanWertz, G., Goldstein, A. H., Liu, Y., Springston, S. R., Souza, R., Newburn, M. K., Alexander, M. L., Martin, S. T., and Jimenez, J. L.: Secondary organic aerosol formation from ambient air in an oxidation flow reactor in central Amazonia, Atmospheric Chemistry and Physics, 18, 467–493, https://doi.org/10.5194/acp-18-467-2018, 2018.

Pankow, J. F.: An absorption model of gas/particle partitioning of organic compounds in the atmosphere, Atmospheric Environment, 28, 185–188, https://doi.org/10.1016/1352-2310(94)90093-0, 1994.

Pankow, J. F., Marks, M. C., Barsanti, K. C., Mahmud, A., Asher, W. E., Li, J., Ying, Q., Jathar, S. H., and Kleeman, M. J.: Molecular view modeling of atmospheric organic particulate matter: Incorporating molecular structure and co-condensation of water, Atmospheric Environment, 122, 400–408, https://doi.org/10.1016/j.atmosenv.2015.10.001, 2015.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kürten, A., St Clair, J. M., Seinfeld, J. H., and Wennberg, P. O.: Unexpected epoxide formation in the gas-phase photooxidation of isoprene., Science, 325, 730–3, https://doi.org/10.1126/science.1172910, http://www.ncbi.nlm.nih.gov/pubmed/19661425, 2009.

Pratt, K. A., Fiddler, M. N., Shepson, P. B., Carlton, A. G., and Surratt, J. D.: Organosulfates in cloud water above the Ozarks' isoprene source region, Atmospheric Environment, 77, 231–238, https://doi.org/10.1016/j.atmosenv.2013.05.011, 2013.

Raes, F.: Entrainment of free tropospheric aerosols as a regulating mechanism for cloud condensation nuclei in the remote marine boundary layer, Journal of Geophysical Research, 100, 2893, https://doi.org/10.1029/94JD02832, 1995.

Raventos-Duran, T., Camredon, M., Valorso, R., Mouchel-Vallon, C., and Aumont, B.: Structure-activity relationships to estimate the effective Henry's law constants of organics of atmospheric interest, Atmospheric Chemistry and Physics, 10, 7643–7654, https://doi.org/10.5194/acp-10-7643-2010, 2010.

Renard, P., Siekmann, F., Salque, G., Demelas, C., Coulomb, B., Vassalo, L., Ravier, S., Temime-Roussel, B., Voisin, D., and Monod, A.: Aqueous-phase oligomerization of methyl vinyl ketone through photooxidation - Part 1: Aging processes of oligomers, Atmospheric Chemistry and Physics, 15, 21–35, https://doi.org/10.5194/acp-15-21-2015, 2015.

Riemer, N. and West, M.: Quantifying aerosol mixing state with entropy and diversity measures, Atmospheric Chemistry and Physics, 13, 11 423–11 439, https://doi.org/10.5194/acp-13-11423-2013, 2013.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmospheric Chemistry and Physics, 3, 161–180, https://doi.org/10.5194/acp-3-161-2003, http://www.atmos-chem-phys.net/3/161/2003/, 2003.

Schifter, I., Díaz, L., Sánchez-Reyna, G., González-Macías, C., González, U., and Rodríguez, R.: Influence of gasoline olefin and aromatic content on exhaust emissions of 15% ethanol blends, Fuel, https://doi.org/10.1016/j.fuel.2019.116950, 2020.

Schmid, B., Tomlinson, J. M., Hubbe, J. M., Comstock, J. M., Mei, F., Chand, D., Pekour, M. S., Kluzek, C. D., Andrews, E., Biraud, S. C., and McFarquhar, G. M.: The DOE arm aerial facility, Bulletin of the American Meteorological Society, 95, 723–742, https://doi.org/10.1175/BAMS-D-13-00040.1, 2014.

Shilling, J. E., Pekour, M. S., Fortner, E. C., Artaxo, P., de Sá, S., Hubbe, J. M., Longo, K. M., Machado, L. A. T., Martin, S. T., Springston, S. R., Tomlinson, J., and Wang, J.: Aircraft observations of the chemical composition and aging of aerosol in the Manaus urban plume during GoAmazon 2014/5, Atmospheric Chemistry and Physics, 18, 10 773–10 797, https://doi.org/10.5194/acp-18-10773-2018, 2018.

Shrivastava, M., Zelenyuk, A., Imre, D., Easter, R., Beranek, J., Zaveri, R. A., and Fast, J.: Implications of low volatility SOA and gas-phase fragmentation reactions on SOA loadings and their spatial and temporal evolution in the atmosphere, Journal of Geophysical Research: Atmospheres, 118, 3328–3342, https://doi.org/10.1002/jgrd.50160, 2013.

Shrivastava, M., Easter, R. C., Liu, X., Zelenyuk, A., Singh, B., Zhang, K., Ma, P.-L., Chand, D., Ghan, S., Jimenez, J. L., Zhang, Q., Fast, J., Rasch, P. J., and Tiitta, P.: Global transformation and fate of SOA: Implications of low-volatility SOA and gas-phase fragmentation reactions, Journal of Geophysical Research: Atmospheres, 120, 4169–4195, https://doi.org/10.1002/2014JD022563, 2015.

Shrivastava, M., Andreae, M. O., Artaxo, P., Barbosa, H. M. J., Berg, L. K., Brito, J., Ching, J., Easter, R. C., Fan, J., Fast, J. D., Feng, Z., Fuentes, J. D., Glasius, M., Goldstein, A. H., Alves, E. G., Gomes, H., Gu, D., Guenther, A., Jathar, S. H., Kim, S., Liu, Y., Lou, S., Martin, S. T., McNeill, V. F., Medeiros, A., de Sá, S. S., Shilling, J. E., Springston, S. R., Souza, R. A. F., Thornton, J. A., Isaacman-VanWertz, G., Yee, L. D., Ynoue, R., Zaveri, R. A., Zelenyuk, A., and Zhao, C.: Urban pollution greatly enhances formation of natural aerosols over the Amazon rainforest, Nature Communications, 10, 1046, https://doi.org/10.1038/s41467-019-08909-4, 2019.

Sinha, V., Williams, J., Crowley, J. N., and Lelieveld, J.: The comparative reactivity method - A new tool to measure total OH Reactivity in ambient air, Atmospheric Chemistry and Physics, 8, 2213–2227, https://doi.org/10.5194/acp-8-2213-2008, 2008.

Szopa, S., Aumont, B., and Madronich, S.: Assessment of the reduction methods used to develop chemical schemes: building of a new chemical scheme for VOC oxidation suited to three-dimensional multiscale HOx-NOx-VOC chemistry simulations, Atmospheric Chemistry and Physics, 5, 2519–2538, https://doi.org/10.5194/acp-5-2519-2005, 2005.

Tennekes, H.: A Model for the Dynamics of the Inversion Above a Convective Boundary Layer, Journal of the Atmospheric Sciences, 30, 558–567, https://doi.org/10.1175/1520-0469(1973)030<0558:AMFTDO>2.0.CO;2, 1973.

Thalman, R., de Sá, S. S., Palm, B. B., Barbosa, H. M. J., Pöhlker, M. L., Alexander, M. L., Brito, J., Carbone, S., Castillo, P., Day, D. A., Kuang, C., Manzi, A., Ng, N. L., Sedlacek III, A. J., Souza, R., Springston, S., Watson, T., Pöhlker, C., Pöschl, U., Andreae, M. O., Artaxo, P., Jimenez, J. L., Martin, S. T., and Wang, J.: CCN activity and organic hygroscopicity of aerosols downwind of an urban region in central Amazonia: seasonal and diel variations and impact of anthropogenic emissions, Atmospheric Chemistry and Physics, 17, 11 779–11 801, https://doi.org/10.5194/acp-17-11779-2017, 2017.

Valorso, R., Aumont, B., Camredon, M., Raventos-Duran, T., Mouchel-Vallon, C., Ng, N. L., Seinfeld, J. H., Lee-Taylor, J., and Madronich, S.: Explicit modelling of SOA formation from α-pinene photooxidation: sensitivity to vapour pressure estimation, Atmospheric Chemistry and Physics, 11, 6895–6910, https://doi.org/10.5194/acp-11-6895-2011, 2011.

Wang, Y., Hu, M., Guo, S., Wang, Y., Zheng, J., Yang, Y., Zhu, W., Tang, R., Li, X., Liu, Y., Le Breton, M., Du, Z., Shang, D., Wu, Y., Wu, Z., Song, Y., Lou, S., Hallquist, M., and Yu, J.: The secondary formation of organosulfates under interactions between biogenic emissions and anthropogenic pollutants in summer in Beijing, Atmospheric Chemistry and Physics, 18, 10 693–10 713, https://doi.org/10.5194/acp-18-10693-2018, 2018.

Wendisch, M., Poschl, U., Andreae, M. O., MacHado, L. A., Albrecht, R., Schlager, H., Rosenfeld, D., Martin, S. T., Abdelmonem, A., Afchine, A., Araujo, A. C., Artaxo, P., Aufmhoff, H., Barbosa, H. M., Borrmann, S., Braga, R., Buchholz, B., Cecchini, M. A., Costa, A., Curtius, J., Dollner, M., Dorf, M., Dreiling, V., Ebert, V., Ehrlich, A., Ewald, F., Fisch, G., Fix, A., Frank, F., Futterer, D., Heckl, C., Heidelberg, F., Huneke, T., Jakel, E., Jarvinen, E., Jurkat, T., Kanter, S., Kastner, U., Kenntner, M., Kesselmeier, J., Klimach, T., Knecht, M., Kohl, R., Kolling, T., Kramer, M., Kruger, M., Krisna, T. C., Lavric, J. V., Longo, K., Mahnke, C., Manzi, A. O., Mayer, B., Mertes, S., Minikin, A., Molleker, S., Munch, S., Nillius, B., Pfeilsticker, K., Pohlker, C., Roiger, A., Rose, D., Rosenow, D., Sauer, D., Schnaiter, M., Schneider, J., Schulz, C., De Souza, R. A., Spanu, A., Stock, P., Vila, D., Voigt, C., Walser, A., Walter, D., Weigel, R., Weinzierl, B., Werner, F., Yamasoe, M. A., Ziereis, H., Zinner, T., and Zoger, M.: Acridicon-chuva campaign: Studying tropical deep convective clouds

and precipitation over amazonia using the New German research aircraft HALO, Bulletin of the American Meteorological Society, 97, 1885–1908, https://doi.org/10.1175/BAMS-D-14-00255.1, 2016.

Wesely, M. L.: Parametrization of surface resistance to gaseous dry deposition in regional-scale numerical model, Atmospheric Environment, 23, 1293–1304, 1989.

Worden, H. M., Bloom, A. A., Worden, J. R., Jiang, Z., Marais, E., Stavrakou, T., Gaubert, B., and Lacey, F.: New Constraints on Biogenic Emissions using Satellite-Based Estimates of Carbon Monoxide Fluxes, Atmospheric Chemistry and Physics Discussions, 30, 1–19, https://doi.org/10.5194/acp-2019-377, 2019.

Xu, L., Guo, H., Boyd, C. M., Klein, M., Bougiatioti, A., Cerully, K. M., Hite, J. R., Isaacman-VanWertz, G., Kreisberg, N. M., Knote, C., Olson, K., Koss, A., Goldstein, A. H., Hering, S. V., de Gouw, J., Baumann, K., Lee, S.-h., Nenes, A., Weber, R. J., and Ng, N. L.: Effects of anthropogenic emissions on aerosol formation from isoprene and monoterpenes in the southeastern United States, Proceedings of the National Academy of Sciences, 112, 37–42, https://doi.org/10.1073/pnas.1417609112, 2015.

Yang, J., Roth, P., Durbin, T., and Karavalakis, G.: Impacts of gasoline aromatic and ethanol levels on the emissions from GDI vehicles: Part 1. Influence on regulated and gaseous toxic pollutants, Fuel, https://doi.org/10.1016/j.fuel.2019.04.143, 2019.

Yee, L. D., Isaacman-VanWertz, G., Wernis, R. A., Meng, M., Rivera, V., Kreisberg, N. M., Hering, S. V., Bering, M. S., Glasius, M., Upshur, M. A., Bé, A. G., Thomson, R. J., Geiger, F. M., Offenberg, J. H., Lewandowski, M., Kourtchev, I., Kalberer, M., de Sá, S., Martin, S. T., Alexander, M. L., Palm, B. B., Hu, W., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Liu, Y., McKinney, K. A., Artaxo, P., Viegas, J., Manzi, A., Oliveira, M. B., de Souza, R., Machado, L. A. T., Longo, K., and Goldstein, A. H.: Observations of sesquiterpenes and their oxidation products in central Amazonia during the wet and dry seasons, Atmospheric Chemistry and Physics Discussions, pp. 1–31, https://doi.org/10.5194/acp-2018-191, 2018.

Yuan, B., Koss, A. R., Warneke, C., Coggon, M., Sekimoto, K., and De Gouw, J. A.: Proton-Transfer-Reaction Mass Spectrometry: Applications in Atmospheric Sciences, Chemical Reviews, 117, 13 187–13 229, https://doi.org/10.1021/acs.chemrev.7b00325, 2017.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), Journal of Geophysical Research, 113, 1–29, https://doi.org/10.1029/2007JD008782, 2008.

Zhao, Y., Nguyen, N. T., Presto, A. A., Hennigan, C. J., May, A. A., and Robinson, A. L.: Intermediate Volatility Organic Compound Emissions from On-Road Diesel Vehicles: Chemical Composition, Emission Factors, and Estimated Secondary Organic Aerosol Production, Environmental Science and Technology, 49, 11 516–11 526, https://doi.org/10.1021/acs.est.5b02841, 2015.

Zhao, Y., Nguyen, N. T., Presto, A. A., Hennigan, C. J., May, A. A., and Robinson, A. L.: Intermediate Volatility Organic Compound Emissions from On-Road Gasoline Vehicles and Small Off-Road Gasoline Engines, Environmental Science and Technology, 50, 4554–4563, https://doi.org/10.1021/acs.est.5b06247, 2016.