## Surl et al., 2017, ACP, Which processes drive observed variations of HCHO columns over India?

### General Description of manuscript:

The authors use total column formaldehyde measurements from the Ozone Monitoring Instrument to assess spatial and temporal variability of volatile organic compounds over India. They identify that isoprene oxidation leading to the formation of formaldehyde is the largest local source contributor to formaldehyde over India and that anthropogenic emissions of non-volatile organic compounds can only be spatially resolved over Delhi with oversampling. The content of this manuscript is appropriate for ACP. The authors have taken careful consideration of the unique seasonality and spatial features over India, but there are quite a few editorial errors that can be eliminated with a careful read-through. Some of these are identified below, along with general comments and specific minor changes.

### General Comments:

Indicate the locations of geographic features in India on the Figures provided so that it is easy for someone not familiar with the country to follow along. For example, locations of the Thar Desert, the IGP, and the five megacities should be provided in Figure 1 (Lines 50-52), the features referred to in lines 310-315 should be indicated in Figure 5, and the features referred to in lines 292-293 should be located on the plot (in particular Rann of Kutch).

In the list of methods applied to deriving isoprene emissions (Lines 85-90), the local $NO_x$-dependent relationship used in Marais et al. (2012) in missing. The authors point out that the correlations and slopes obtained in each season between HCHO and isoprene emissions in GEOS-Chem are highly variable. What drives this variability? Is it $NO_x$? Would it be more effective then to use a $NO_x$-dependent HCHO yield to derive isoprene emissions from OMI HCHO?

Why when comparing the multiyear average and year-to-year variability of OMI HCHO to the 2014 values is OMI HCHO for 2008-2015 used? Why not use the full OMI HCHO record starting from the first complete year (i.e. 2005-2015)?

What drives the year-to-year variability in HCHO (Figure 7, Line 22)?

Some information regarding the isoprene oxidation chemistry in each model is needed so that the reader can assess whether the 2 models (CAABA and GEOS-Chem) have the same isoprene oxidation chemistry, or whether CAABA, due to lower computational needs than a 3D model, has a more detailed representation of isoprene oxidation than GEOS-Chem. Have both models kept pace with new findings of isoprene oxidation mechanisms reported in the literature? Are $NO_x$-dependent HCHO yields from isoprene oxidation similar for the 2 models?

The manuscript is missing a Data Availability section (https://www.atmospheric-chemistry-and-physics.net/about/data_policy.html).

### Specific Comments:

Line 11: "comparable (slower)" is contradictory. Which is it?

Lines 69-71: Please specify whether these trends in HCHO are for large urban areas, rural regions or across the country.

Line 73: "By" is missing in "supported … detailed modelling studies".

Line 86: Close bracket missing.

Line 139: "detailed" should be "detail".

Lines 163-164: Incorrect in-text referencing style for Sander et al., 2014.

Line 186: Should "signal" be "sigma"?

Line 186-187: "30 in the troposphere" is appropriate for a fixed tropopause, but doesn't the model have a dynamic tropopause that leads to variability in the number of levels in the troposphere?

Line 194: Remove redundant "of".

Line 195: Is MIX a mosaic of regional inventories? As written this isn't clear.

Line 201: Provide the version of MEGAN that is used.

Lines 206-207: Say where the drivers of MEGAN are from. Is LAI from the model fields or from MODIS?

Line 297: The meaning of the 11% is not clear. Is 11% the percent increase in vertical columns relative to slant columns after applying the AMF? Or is this an increase in the spatial correlation after applying the AMF?

Lines 347-348: Reiterate that the slant columns are used for 2008-2015.

Line 351: "that" should be removed in "shows that".

Line 352: Should "Median" be mean? That's what's shown in Figure 8.

Line 384: Don't mature leaves emit more isoprene than young and senescing leaves (Guenther et al., 2006)? This sentence needs to be revised to reflect that.

Line 404: "thes" is a typo.

Lines 424-425: Why does the bottom-up emission inventory overestimate emissions from the transport sector when the inventory is for 2010 and the simulation year is 2014? Wouldn't this misrepresent the growth in this sector and so underestimate emissions from the transport sector? The reasoning for an overestimate is not immediately clear.

Line 456: Are the "sample mean" and "sample standard deviation" the regional mean and standard deviation of column HCHO for the whole country? If so, please clarify this in the text.

Line 461: Rather than stating that further work is "outside the scope of this current study", more impactful to point out that this is an early demonstration of the efficacy of future geostationary satellites that will provide constraints on the temporal variability of HCHO.

Line 477: Incorrect in-text citation style for Barkley et al., 2013.

Figure 1:
- City names (perhaps capital cities?) and the legend in the population map are not legible.
- The population map has a panel header ("India Population Map"). The other panels would benefit from headers too.
- The red outline could be made clearer in the top right map.

Figure 2:
- Label "C" is missing.
- Consider choosing a colour bar with a more dynamic colour range. As is, JJAS emissions seem only marginally higher than emissions in other months.
- What are the isoprene emission units? "g C", "molecules C", or "mg C"?

Figure 3:
- Consider including initial $NO_x$ (or NO) concentrations for each isoprene HCHO yield line shown in the plot.

Figure 7:
- "Left" should be "top" and "right" should be "bottom" in the figure caption.

Figure 9:
- Consider choosing a colour bar with a more dynamic colour range.

Figure 12:
- Labels on the map are not legible.
- Consider replacing "map imagery" with "road network".

Table 1:
- Include a footnote in the table to indicate the source of variables (T, RH, BLH, $O_3$, CO, $NO_2$) provided.

**References:**
Guenther et al., 2006, Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, 2006.
Marais et al., 2012, Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, Atmos. Chem. Phys., 12, 6219-6235, 2012.