Observing atmospheric formaldehyde from space: validation and implications

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:37944975

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
© 2016 Lei Zhu

All rights reserved.
Observing atmospheric formaldehyde from space: validation and implications

Abstract

Formaldehyde (HCHO) column data from satellites are widely used as a proxy for emissions of volatile organic compounds (VOCs), but validation of the data has been extremely limited. Here I use highly accurate HCHO aircraft observations from the NASA SEAC4RS campaign over the Southeast US in 2013 summer to validate and intercompare currently available HCHO retrievals. All retrievals feature a HCHO maximum over Arkansas and Louisiana, consistent with the aircraft observations and GEOS-Chem model results, and reflecting high emissions of biogenic isoprene. The retrievals are also broadly consistent in their spatial ($r = 0.4 – 0.8$ on a $0.5^\circ \times 0.5^\circ$ grid) and daily variability ($r = 0.5 – 0.8$) over the Southeast US in 2013 summer. Validation results show that HCHO column data provide a reliable proxy for isoprene emission variability but with a low mean bias (20–51%) due both to the spectral fitting and the scattering weights. I apply the corrected OMI data in the following studies.

First, I show that temporal oversampling of OMI HCHO column data for 2005–2008 enables detection of urban and industrial plumes in eastern Texas including Houston, Port Arthur, and Dallas/Fort Worth. By spatially integrating the HCHO enhancement in the Houston plume observed by OMI, I estimate an anthropogenic HCHO source of $250 \pm 140 \text{ kmol h}^{-1}$, implying that anthropogenic highly reactive VOC emissions in Houston are $4.8 \pm 2.7$ times higher than reported by the US Environmental Protection Agency inventory (EPA).

I then examine the interannual variability of temperature-independent HCHO columns over North America from 2005 to 2014. Significant trends are found to be likely driven by tightening emission controls over the Houston-Galveston-Brazoria area (-2.2% $\pm$ 0.7% a$^{-1}$), by oil/gas extraction activities over the Floyd shale in Alabama (-1.8% $\pm$ 0.7% a$^{-1}$) and the Athabasca oil
sands in Alberta (4.7% ± 1.1% a⁻¹), and by reforestation over the northwestern US (4.7% ± 1.0% a⁻¹). Variation in HCHO columns may also be associated with changing crop cover in the midwest US and with changing wildfire activity in California.

Finally, I use 2005–2015 OMI HCHO column data to map surface air HCHO concentrations across the contiguous US. Results are in good agreement with high-quality observations from urban sites and a factor of 2 lower than data from the EPA network. I estimate that up to 6600–13200 people in the US will develop cancer over their lifetimes by exposure to outdoor HCHO. I also find out that NO₃ emission controls to improve ozone air quality have a co-benefit in reducing HCHO-related cancer risks.
Table of Contents

Abstract ......................................................................................... iii

Table of Contents ......................................................................... iii

List of Tables ................................................................................ vi

List of Figures ................................................................................ vii

Acknowledgements ......................................................................... ix

Chapter 1 Overview ....................................................................... 1

1.1 Observing HCHO from space and applications ................................ 1

1.2 Research objectives and approach .................................................. 2

1.3 Summary of results ....................................................................... 3

References .................................................................................... 6

Chapter 2 Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC4RS aircraft observations over the Southeast US ................................................... 8

Abstract ....................................................................................... 8

2.1 Introduction ............................................................................. 9

2.2 Satellite data sets .................................................................... 11

2.3 Aircraft observations and GEOS-Chem model simulation ................. 15

2.4 Intercomparison and validation of satellite data sets over the Southeast US 23

2.5 Conclusions ........................................................................... 31

Acknowledgments ......................................................................... 32

References .................................................................................. 33
Chapter 3  Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns ____ 37

Abstract ____________________________________________________________ 37

3.1 Introduction_______________________________________________________ 38

3.2 Data and methods ________________________________________________ 39

3.3 Results and discussion ____________________________________________ 41

3.4 Conclusions_______________________________________________________ 50

Acknowledgements___________________________________________________ 50

References __________________________________________________________ 51

Chapter 4  Long-term trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing emissions of volatile organic compounds 55

Abstract ____________________________________________________________ 55

4.1 Introduction_______________________________________________________ 56

4.2 OMI HCHO column data and temperature correction _____________________ 57

4.3 Results and discussion ____________________________________________ 59

Acknowledgments___________________________________________________ 65

References __________________________________________________________ 66

Chapter 5  Formaldehyde (HCHO) as a Hazardous Air Pollutant: Mapping surface air concentrations from satellite and inferring cancer risks in the United States ___________ 70

Abstract ____________________________________________________________ 70

5.1 Introduction_______________________________________________________ 71

5.2 Methods________________________________________________________ 72

5.2.1 OMI observations______________________________________________ 72
5.2.2 Oversampling method ................................................................. 74
5.2.3 Deriving annual mean surface HCHO concentrations .................. 76
5.2.4 Estimating cancer risks ............................................................. 81
5.3 Results and discussions ................................................................ 82
  5.3.1 Annual mean surface HCHO concentrations .............................. 82
  5.3.2 National cancer risks from outdoor HCHO exposure ..................... 85
  5.3.3 Co-benefit of NOx emission controls for reducing HCHO-related cancer risks 85

Acknowledgement .............................................................................. 86

References .......................................................................................... 87

Supplemental Materials .......................................................................... 91
List of Tables

Table 2.1 Satellite HCHO products validated and intercompared in this work ____________ 13

Table 2.2 Spatial/temporal correlation coefficients between pairs of HCHO column products 25

Table 2.3 Satellite retrievals of HCHO columns over the southeast US ________________ 26

Table 3.1 NEI estimates of Houston AHRVOC emissions and HCHO yields ____________ 44
List of Figures

Figure 2.1 Formaldehyde concentrations along SEAC$^4$RS aircraft flight tracks___________ 16

Figure 2.2 Comparisons between HCHO measurements from the CAMS and ISAF instruments aboard the SEAC$^4$RS aircraft, and simulated by GEOS-Chem, for the Southeast US flight tracks_________________________________________ 17

Figure 2.3 Mean temperature, base isoprene emissions, and HCHO columns in the GEOS-Chem model for the SEAC$^4$RS period ___________________________________________ 19

Figure 2.4 Frequency distribution of mixed layer depths over the Southeast US during the SEAC$^4$RS period___________________________________________________________ 21

Figure 2.5 HCHO vertical column densities over the Southeast US averaged over the SEAC$^4$RS period ________________________________________________________________ 24

Figure 2.6 Air mass factor differences between retrievals____________________________________ 28

Figure 2.7 Daily variability of HCHO vertical column densities over the Southeast US during SEAC$^4$RS ________________________________________________________________ 30

Figure 3.1 OMI HCHO columns and HRVOC emission inventories for eastern Texas _______ 42

Figure 3.2 Relationship of OMI HCHO column with surface air temperature for the Houston urban core and Northeast Texas ____________________________ 45

Figure 3.3 Cross-section of the Houston HCHO plume along the direction of the prevailing SSE wind___________________________________________________________ 47

Figure 4.1 Absolute and relative mean temperature-corrected change in HCHO columns in May–September from 2005–2009 to 2010–2014 over the North America _____________ 60
Figure 4.2 Time series of temperature-corrected May–September HCHO columns and the potential explanatory variables driving HCHO variability from 2005 to 2014 in the five regions circled in Figure 4.1 62

Figure 5.1 Mean OMI HCHO columns over the contiguous US for June–August 2005–2015 with oversampling on a 0.05° × 0.05° grid 77

Figure 5.2 OMI-derived summer midday HCHO concentrations in surface air 78

Figure 5.3 Summer mean midday HCHO concentrations at urban sites 79

Figure 5.4 Diurnal variation of summertime HCHO concentrations in surface air 80

Figure 5.5 OMI-derived annual mean HCHO concentrations in surface air and associated cancer risks 83

Figure 5.6 Annual mean 2005–2015 HCHO concentrations from EPA surface network (SLTs) and comparison to OMI-derived surface air concentrations 84
Acknowledgements

First of all, I thank my advisors Daniel Jacob and Loretta Mickley. Thank you for your excellent guidance during my entire PhD life, for challenging me at every iteration of papers and presentations, for broadening my horizon, and for always treating me as your future colleague. I would also like to thank my thesis committee members Kelly Chance and Steven Wofsy for their insightful and helpful feedback on this work. I thank Yu Song, my former Master advisor, who opened the door of Atmospheric Science research for me and trained me as a scientist when I was at Peking University.

I want to thank NASA AQAST, Aura, and SEAC4RS science teams. This work can not be done without inputs from you. Special thanks to Gonzalo González Abad for insightful discussions. Thanks to NASA and Harvard University Center for the Environment for funding this work.

It has been a great honor working with the Harvard Atmospheric Chemistry Modeling Group. Thank Bob Yantosca and Melissa Suprizio for helping me with GEOS-Chem. I appreciate the help and friendship from past and current group members: Ploy Achakulwisut, Thomas Breider, Dan Cusworth, Chris Chan Miller, Jenny Fisher, Tzung-May Fu, Hannah Horowitz, Lu Hu, Christoph Keller, Patrick Kim, Shannon Koplitz, Jessica Kunke, Eric Leibensperger, Junhua Liu, Bram Maasakkers, Dylan Millet, Jonathan Moch, Justin Parrella, Fabien Paulot, Lu Shen, Jianxiong Sheng, Rachel Silvern, Amos Tai, Katie Travis, Alex Turner, Qiaoqiao Wang, Kevin Wecht, Xu Yue, Karen Yu, Lin Zhang, and Peter Zoogman. Special thanks to Eloïse Marais who helps me with my research as a reliable mentor. Thanks to Brenda Mathieu for fantastic logistics.

It is clearly true that life at graduate school is challenging, especially for an international student. I appreciate help I received from Harvard throughout the years. Thanks to the English Language Program for providing me a smooth transition in both language and culture when I first
came to Harvard five years ago. Thanks to the Bok Center for language and teaching training. Thanks to International Office for helping me deal with tedious paperwork.

I would like to thank all my Chinese friends I met at Harvard. Special thanks to Yang Chen, Peng Ding, Tianming Fu, Yang Li, Tong Qin, Lu Sun, and Chun Pong Tam. The life would be certainly less colorful without you. I also enjoy every match with my friends at Harvard Chinese Soccer Team.

I want to extend my deepest gratitude to my family for their continuous encouragement along this journey. To my parents Wumin Guo and Shuxia Jiang, my parents-in-law Hui Dai and Xia Wang, my sister Huihui Guo and brother-in-law Peng Wang, my brother Shoutao Guo and sister-in-law Qing Wu, and my brother-in-law Ruixiao Dai.

Lastly but perhaps most importantly, I want to thank my wife Ruizhi Dai. Thank you for supporting me at every turning point of my life. Thank you for always being there. The best is yet to come.
Chapter 1 Overview

1.1 Observing HCHO from space and applications

HCHO is a high-yield product from the atmospheric oxidation of volatile organic compounds (VOCs). Methane oxidation largely defines the tropospheric HCHO background. Higher HCHO concentrations over continents are due to short-lived non-methane VOCs (NMVOCs). Loss of HCHO is mainly by photolysis and oxidation by OH, resulting in an atmospheric lifetime of the order of a few hours.

HCHO is detectable from space by solar UV backscatter between 325 and 360 nm [Chance et al., 2000]. HCHO columns (molecules cm\(^{-2}\)) have been continuously observed since GOME (1996–2003; Chance et al. [2000]) followed by SCIAMACHY (2003–2012; Wittrock et al. [2006]). Observations are presently available from OMI (2004–), GOME2A (2006–), OMPS (2011–), and GOME2B (2012–). The satellite detects a slant column density of HCHO along the atmospheric path of the solar radiation back-scattered to the satellite from the surface and the atmosphere. Conversion to a vertical column is done with an air mass factor (AMF) that depends on the satellite viewing geometry, the surface albedo, the vertical HCHO profile, and the vertical distributions of clouds and aerosols [Palmer et al., 2001].

HCHO column data have been used widely as a proxy to constrain VOC emissions from biogenic sources [Palmer et al., 2003; Shim et al., 2005; Barkley et al., 2008; Millet et al., 2008; Dufour et al., 2009; Stavrakou et al., 2009; Curci et al., 2010; Marais et al., 2012; Barkley et al., 2013; Bauwens et al., 2016], anthropogenic sources [Fu et al., 2007; Marais et al., 2014] and open fires [Shim et al., 2005; Stavrakou et al., 2009; Gonzi et al., 2011]. This is done by assuming a local relationship between the two from a chemical transport model (CTM) or by applying a more
elaborate inversion method. Trends in HCHO columns also infer changes in the underlying VOC emissions [De Smedt et al., 2010, 2015; Khokhar et al., 2015; Mahajan et al., 2015].

1.2 Research objectives and approach

HCHO column data and its previous applications, however, have been limited in following aspects. (1) Satellite HCHO data have received little validation so far, leading to unclear bias in the retrievals and further applications. (2) Estimating anthropogenic emissions in the US by HCHO column data has been stymied by lack of a discernable HCHO signal. (3) No trend study to date has controlled for the effects of temperature on the biogenic emissions of HCHO precursors. (4) HCHO itself is the most important carcinogen in outdoor air among the 187 hazardous air pollutants (HAPs) identified by the US Environmental Protection Agency (EPA). The HAPs sampling network in the US is limited to urban/industrial sites and thus cannot properly quantify the health risk posed by ambient HCHO on a national scale. This thesis therefore addresses the following four specific objectives:

- Validate and intercompare presently available HCHO retrievals with aircraft observations from the NASA SEAC4RS campaign over the Southeast US in 2013 summer.
- Constrain anthropogenic highly reactive VOC emissions in eastern Texas by oversampling of the OMI HCHO column data.
- Examine the temperature-corrected variability in HCHO columns from 2005 to 2014 over North America and its possible driving factors.
- Map HCHO surface air concentrations over the contiguous US from space and infer cancer risks of outdoor HCHO on a national scale.
1.3 Summary of results

In Chapter 2, I use highly accurate HCHO aircraft observations from the NASA SEAC4RS campaign over the Southeast US in August–September 2013 to validate and intercompare six retrievals of HCHO columns from four different satellite instruments (OMI, GOME2A, GOME2B and OMPS) and three different research groups. The GEOS-Chem chemical transport model is used as a common intercomparison platform. All retrievals feature a HCHO maximum over Arkansas and Louisiana, consistent with the aircraft observations and reflecting high emissions of biogenic isoprene. The retrievals are also interconsistent in their spatial variability over the Southeast US ($r = 0.4–0.8$ on a $0.5^\circ \times 0.5^\circ$ grid) and in their day-to-day variability ($r = 0.5–0.8$). However, all retrievals are biased low in the mean by 20–51%, which would lead to corresponding bias in estimates of isoprene emissions from the satellite data. The smallest bias is for OMI-BIRA (Belgian Institute for Space Aeronomy), which has high corrected slant columns relative to the other retrievals and low scattering weights in its air mass factor (AMF) calculation. OMI-BIRA has systematic error in its assumed vertical HCHO shape profiles for the AMF calculation and correcting this would eliminate its bias relative to the SEAC4RS data. Our results support the use of satellite HCHO data as a quantitative proxy for isoprene emission after correction of the low mean bias. There is no evident pattern in the bias, suggesting that a uniform correction factor may be applied to the data until better understanding is achieved.

In Chapter 3, I show that temporal oversampling of HCHO data from the Ozone Monitoring Instrument (OMI) for 2005–2008 enables detection of urban and industrial plumes in eastern Texas including Houston, Port Arthur, and Dallas/Fort Worth. By spatially integrating the HCHO enhancement in the Houston plume observed by OMI, I estimate an anthropogenic HCHO source of $250 \pm 140$ kmol h$^{-1}$. This implies that anthropogenic HRVOC emissions in Houston are
4.8 ± 2.7 times higher than reported by the US Environmental Protection Agency inventory, and is consistent with field studies identifying large ethene and propene emissions from petrochemical industrial sources.

In Chapter 4, I examine the long-term trends in HCHO columns observed by OMI from 2005 to 2014 across North America. Biogenic isoprene is the dominant source of HCHO and its emission has a large temperature dependence. After correcting for this dependence, I find significant long-term trends due to changes in anthropogenic emissions, land cover, and other factors. I find that HCHO columns over the Houston-Galveston-Brazoria industrial area decreased by 2.2% a\(^{-1}\), consistent with trends in anthropogenic VOCs. HCHO columns increased by 4.7% a\(^{-1}\) over the Athabasca oil sands in Alberta, reflecting increasing bitumen production, and decreased by 1.8% a\(^{-1}\) over the Floyd shale in Alabama, reflecting decreased production of natural gas. Other oil/gas production regions do not show significant HCHO enhancements and trends. An increase of 4.7% a\(^{-1}\) over the northwestern US may be due to reforestation. Variation in HCHO columns may also be associated with changing crop cover in the midwest US and with changing wildfire activity in California.

In Chapter 5, I use 2005–2015 HCHO column data from the OMI satellite instrument, validated with high-quality aircraft data and oversampled on a 5 × 5 km\(^2\) grid, to map surface air HCHO concentrations across the contiguous US. Results are in good agreement with high-quality observations from urban sites and a factor of 2 lower than data from the EPA network. I estimate that up to 6600–13200 people in the US will develop cancer over their lifetimes by exposure to outdoor HCHO. The main HCHO source in the US is atmospheric oxidation of biogenic isoprene. The HCHO yield from atmospheric oxidation of isoprene decreases as the concentration of
nitrogen oxides ($\text{NO}_x \equiv \text{NO} + \text{NO}_2$) decreases, so that NO$_x$ emission controls to improve ozone air quality have a co-benefit in reducing HCHO-related cancer risks.
References


Chapter 2 Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC4RS aircraft observations over the Southeast US

Abstract

Formaldehyde (HCHO) column data from satellites are widely used as a proxy for emissions of volatile organic compounds (VOCs) but validation of the data has been extremely limited. Here we use highly accurate HCHO aircraft observations from the NASA SEAC4RS campaign over the Southeast US in August–September 2013 to validate and intercompare six retrievals of HCHO columns from four different satellite instruments (OMI, GOME2A, GOME2B and OMPS; for clarification of these and other abbreviations used in the paper, please refer to the Supplemental Materials) and three different research groups. The GEOS-Chem chemical transport model is used as a common intercomparison platform. All retrievals feature a HCHO maximum over Arkansas and Louisiana, consistent with the aircraft observations and reflecting high emissions of biogenic isoprene. The retrievals are also interconsistent in their spatial variability over the Southeast US (r = 0.4–0.8 on a 0.5° × 0.5° grid) and in their day-to-day variability (r = 0.5–0.8). However, all retrievals are biased low in the mean by 20–51%, which would lead to corresponding bias in estimates of isoprene emissions from the satellite data. The smallest bias is for OMI-BIRA, which has high corrected slant columns relative to the other

---

retrievals and low scattering weights in its air mass factor (AMF) calculation. OMI-BIRA has systematic error in its assumed vertical HCHO shape profiles for the AMF calculation and correcting this would eliminate its bias relative to the SEAC^4RS data. Our results support the use of satellite HCHO data as a quantitative proxy for isoprene emission after correction of the low mean bias. There is no evident pattern in the bias, suggesting that a uniform correction factor may be applied to the data until better understanding is achieved.

2.1 Introduction

Formaldehyde (HCHO) is a high-yield product from the atmospheric oxidation of volatile organic compounds (VOCs). Methane oxidation largely defines the tropospheric HCHO background. Higher HCHO concentrations over continents are due to short-lived non-methane VOCs (NMVOCs). Loss of HCHO is mainly by photolysis and oxidation by OH, resulting in an atmospheric lifetime of the order of a few hours. HCHO is detectable from space by solar UV backscatter between 325 and 360 nm [Chance et al., 2000]. HCHO column data from satellites have been used in a number of studies as top-down constraints on NMVOC emissions from biogenic, anthropogenic, and open fire sources [Palmer et al., 2003; Shim et al., 2005; Stavrakou et al., 2009; Marais et al., 2012; Barkley et al., 2013; Zhu et al., 2014]. However, the satellite data have received little validation so far. Here we validate and intercompare six different HCHO retrievals from four satellites instruments (OMI, GOME2A, GOME2B, OMPS) and three different groups with aircraft observations from the NASA SEAC^4RS (Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys) campaign over the Southeast US in summer 2013 [Toon et al., 2015].
HCHO columns (molecules cm$^{-2}$) have been continuously observed from space since GOME (1996–2003; Chance et al. [2000]) and SCIAMACHY (2003–2012; Wittrock et al. [2006]). Observations are presently available from OMI (2004–), GOME2A (2006–), OMPS (2011–), and GOME2B (2012–). The satellite detects a slant column density of HCHO along the atmospheric path of the solar radiation back-scattered to the satellite from the surface and the atmosphere. Conversion to a vertical column is done with an air mass factor ($AMF$) that depends on the satellite viewing geometry, the surface albedo, the vertical HCHO profile, and the vertical distributions of clouds and aerosols [Palmer et al., 2001]. Scattering by air molecules causes the $AMF$ to be highly sensitive to the HCHO vertical distribution, which has to be independently specified [Hewson et al., 2015]. The resulting HCHO vertical column retrieved from the satellite includes errors from the slant column fitting and from the $AMF$ estimate [Marais et al., 2012].

Validation of HCHO satellite data sets has been extremely limited due to (1) the large noise in individual satellite retrievals, requiring extensive data averaging to enhance detection; and (2) the limited number of HCHO column measurements acquired from aircraft or from the ground. Martin et al. [2004] validated GOME HCHO columns with aircraft observations in eastern Texas averaged over two campaigns (June–July 1999 and August–September 2000), and found GOME to be too high by 16% on average. Comparison of SCIAMACHY data to ground-based measurements of HCHO columns found no significant mean bias [Wittrock et al., 2006; Vigouroux et al., 2009]. Barkley et al. [2013] found that OMI was 37% lower than aircraft measurements made in October 2005 over Guyana. Validation with ground-based remotely sensed vertical profiles indicates a 20–40% underestimate in OMI and GOME2 data [De Smedt et al., 2015].

The SEAC$^4$RS campaign offers an exceptional opportunity for validating satellite HCHO data. HCHO columns over the Southeast US in summer are among the highest in the world [Kurosu
et al., 2004], due to large emissions of biogenic isoprene from vegetation [Guenther et al., 2006]. Several studies have used HCHO data from space as constraints on isoprene emission in the Southeast US [Palmer et al., 2006; Millet et al., 2008; Valin et al., 2016]. The SEAC^{4}RS aircraft payload included two independently-calibrated in situ HCHO measurements: the Compact Atmospheric Multispecies Spectrometer (CAMS) [D. Richter et al., 2015] and the NASA GSFC In Situ Airborne Formaldehyde (ISAF) [Cazorla et al., 2015]. CAMS is a mid-IR laser-based spectrometer, which has 1 Hz detection sensitivity of ~ 40 ppt HCHO [D. Richter et al., 2015]. ISAF uses rotational-state specific laser for detection of HCHO with a 1 Hz detection limit of 36 ppt [Cazorla et al., 2015].

The SEAC^{4}RS aircraft did not conduct direct satellite validation profiles, hence we did not apply the direct validation method. Instead we use here an indirect validation method involving joint comparisons of satellite and in situ HCHO observations with the GEOS-Chem chemical transport model (CTM; Bey et al. [2001]). Satellite and in situ observations do not need to be concurrent, thus increasing considerably the range of data and conditions that can be used for validation.

### 2.2 Satellite data sets

**Table 2.1** lists the six different satellite retrievals of HCHO produced during the SEAC^{4}RS campaign. Additional details on the retrievals are in the **Supplemental Materials**. These are from four satellite instruments (OMI, GOME2A, GOME2B, OMPS) on different platforms, with retrievals produced by independent groups for OMI and OMPS. OMI, flown on the NASA Aura research satellite, has much higher spatial resolution than the other instruments. GOME2A and GOME2B are the first successive instruments of a long-term operational commitment by the
EUMETSAT European agency for observing atmospheric composition from space [Callies et al., 2000]. OMPS is the first instrument of a similar long-term operational commitment by NOAA in the US [Dittman et al., 2002].

All instruments in Table 2.1 provide dense data sets, with full coverage of the Earth’s surface in 1 day for OMI and OMPS, 3 days for GOME2A (since July 2013), and 1.5 days for GOME2B. The single-scene detection limit (0.5–1.0 × 10^{16} molecules cm^{-2}) is determined by uncertainty in fitting the backscattered solar spectra for SAO retrievals [González Abad et al., 2015, 2016] and is estimated as the standard deviation of HCHO column amounts over the remote Pacific for BIRA retrievals [De Smedt et al., 2012; 2015]. AMFs add another error of 30–100% for single-scene retrievals [González Abad et al., 2015]. OMPS-PCA has a single-scene detection limit of 1.2 × 10^{16} molecules cm^{-2} estimated as 4-times the standard deviation of HCHO column amounts over the Pacific Ocean [Li et al., 2015]. Uncertainties in HCHO columns can be reduced for monthly means, down to 20–40% for GOME-2A [De Smedt et al., 2008], 38% for OMI, 46% for GOME-2B and ~ 30% for OMPS, corresponding to 0.1–0.4 × 10^{16} molecules cm^{-2} over the Southeast US. Here and elsewhere, we use only satellite pixels with solar zenith angle less than 60°, cloud fraction less than 0.3, row anomalies (for OMI) screened, quality check passed (for SAO retrievals), and vertical column density within the −0.5 to 10 × 10^{16} molecules cm^{-2} range. The latter criterion excludes data that have passed quality tests but are nevertheless suspect as outliers. It excludes 5.8% of the data.
Table 2.1 Satellite HCHO products validated and intercompared in this work\(^a\)

<table>
<thead>
<tr>
<th>Retrieval</th>
<th>Nadir resolution (km(^2))</th>
<th>Local viewing time</th>
<th>Fitting windows (nm)</th>
<th>Chemical transport model(^b)</th>
<th>Detection limit (10(^{16}) molecules cm(^{-2}))</th>
<th>Reference(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMI-SAO (V003)</td>
<td>24 × 13</td>
<td>1330</td>
<td>328.5–356.5</td>
<td>GEOS-Chem v09-01-03</td>
<td>1.0</td>
<td>(1)</td>
</tr>
<tr>
<td>OMI-BIRA (V14)</td>
<td>24 × 13</td>
<td>1330</td>
<td>328.5–346.0</td>
<td>IMAGES v2</td>
<td>0.7</td>
<td>(2)</td>
</tr>
<tr>
<td>GOME2A-BIRA (V14)</td>
<td>40 × 40</td>
<td>0930</td>
<td>328.5–346.0</td>
<td>IMAGES v2</td>
<td>0.8</td>
<td>(3)</td>
</tr>
<tr>
<td>GOME2B-BIRA (V14)</td>
<td>80 × 40</td>
<td>0930</td>
<td>328.5–346.0</td>
<td>IMAGES v2</td>
<td>0.5</td>
<td>(2)</td>
</tr>
<tr>
<td>OMPS-SAO</td>
<td>50 × 50</td>
<td>1330</td>
<td>327.7–356.0</td>
<td>GEOS-Chem v09-01-03</td>
<td>0.75</td>
<td>(4)</td>
</tr>
<tr>
<td>OMPS-PCA</td>
<td>50 × 50</td>
<td>1330</td>
<td>328.5–356.5</td>
<td>GMI</td>
<td>1.2</td>
<td>(5)</td>
</tr>
</tbody>
</table>

\(^a\) Retrievals operational during the SEAC\(^4\)RS aircraft campaign (5 August–25 September 2013). These include four different sensors (OMI, GOME2A, GOME2B and OMPS), flown on different platforms, with different retrievals for OMI and OMPS produced by the Harvard Smithsonian Astrophysical Observatory (SAO), the Belgian Institute for Space Aeronomy (BIRA), and the NASA Goddard Space Flight Center by principal component analysis (PCA). Further retrieval details are in the Supplement. \(^b\) Chemical transport model (CTM) supplying the normalized mixing ratio vertical profiles (shape factors) and background correction (\(z_0\); see Sect. 2) used in the retrieval. References are ChanMiller et al. (2014) for GEOS-Chem v09-01-03, Stavrakou et al. (2009) for IMAGES v2 and Rodriguez (1996) for GMI. \(^c\) (1) González Abad et al. (2015); (2) De Smedt et al. (2015); (3) De Smedt et al. (2012); (4) González Abad et al. (2016); (5) Li et al. (2015). OMI-SAO data were downloaded from http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omhcho_v003.shtml. GOME2A-BIRA and GOME2B-BIRA data were downloaded from http://h2co.aeronomie.be. Other data were courtesy of the retrieval groups.
All retrievals (except OMPS-PCA) fit the slant column density (SCD) of HCHO from the backscattered solar radiance spectra and then subtract the SCD over the remote Pacific (known as reference sector correction) for the same latitude and observing time to remove offsets [Khokhar et al., 2005]. The resulting corrected SCD ($\Delta \Omega_S$) thus represents a HCHO enhancement over the Pacific background. Additional details on reference sector correction are in the Supplementary Material. $\Delta \Omega_S$ is converted to the HCHO vertical column density (VCD, $\Omega$) by applying an air mass factor ($AMF$) and a background correction ($\Omega_o$):

$$\Omega = \frac{\Delta \Omega_S}{AMF} + \Omega_o \quad (2.1)$$

The background correction, $\Omega_o$, is the HCHO vertical column simulated by a CTM (Table 2.1) for the remote Pacific at the corresponding latitude and observing time. OMPS-PCA derives the VCD in one step using spectrally varying Jacobians [Li et al., 2015].

The $AMF$ depends on the solar zenith angle ($\theta_Z$) and satellite viewing angle ($\theta_V$), on the scattering properties of the atmosphere and the surface, and on the vertical profile of HCHO concentration. It is computed following Palmer et al. [2001], as the product of a geometrical $AMF$ ($AMF_G$) describing the viewing geometry in a non-scattering atmosphere, and a correction with scattering weights $w$ applied to the vertical shape factors $S$:

$$AMF_G = \frac{1}{\cos \theta_Z} + \frac{1}{\cos \theta_V} \quad (2.2)$$

$$AMF = AMF_G \int_0^{p_S} w(p)S(p)dp \quad (2.3)$$

Here the integration is over the pressure ($p$) coordinate from the surface ($p_S$) to the top of the atmosphere. The shape factor is the normalized vertical profile of mixing ratio: $S(p) = C(p)\Omega_d/\Omega$
where $C$ is the HCHO mixing ratio and $\Omega_A$ is the total air column [Palmer et al., 2001]. The scattering weight measures the sensitivity of the backscattered radiation to the presence of HCHO at a given pressure. Impact of aerosols is not explicitly addressed in HCHO retrievals because it is considered to be implicitly included in the cloud correction scheme to the scattering weights [De Smedt et al., 2012, 2015].

All satellite data products (except OMPS-PCA) in Table 2.1 report for each retrieval $\Omega$, $AMF_G$, and $AMF$, and the scattering weights $w(p)$ or equivalent averaging kernels $A(p) = w(p)/AMF$. The BIRA retrievals report in addition the corrected SCD $\Delta\Omega_5$, and background correction $\Omega_0$. To be able to interpret differences between retrievals, we obtained the $\Omega_0$ values used by the SAO retrievals and applied Equation 2.1 to compute their values of $\Delta\Omega_5$. For OMPS-PCA, we computed the $AMF$ based on the reported $S(p)$, $w(p)$ and $AMF_G$ using Equation 2.3, computed $\Omega_0$ based on the reported uncorrected and corrected VCDs, and then obtained $\Delta\Omega_5$ by Equation 2.1.

2.3 Aircraft observations and GEOS-Chem model simulation

The SEAC4RS DC-8 aircraft flew 21 flights over the Southeast US between 5 August and 25 September 2013, providing extensive mapping of the mixed layer and vertical profiling from the mixed layer to the upper troposphere (Figure 2.1). The mixed layer is defined here as the convectively unstable region of the atmosphere in contact with the ground, as measured from the aircraft by aerosol lidar [Browell et al., 1989; Hair et al., 2008; DIAL-HSRL Mixed Layer Heights README, 2014; Scarino et al., 2014]. It typically extended to 1–3 km altitude (~ 700–900 hPa) during the afternoon. The mixed layer was often capped by a convective cloud layer of fair-weather cumuli extending to about 3 km, with the free troposphere above [Kim et al., 2015]. 95% of flight hours were between 0930 and 1800 local time for the data in Figure 2.1, and 78% in the afternoon.
Figure 2.1 Formaldehyde (HCHO) concentrations along SEAC4RS aircraft flight tracks (5 August–25 September 2013). The top left panel shows the DC-8 flight tracks (in grey) and the CAMS measurements aboard the aircraft in the mixed layer. The mixed layer is the convectively unstable region of the atmosphere in contact with the surface, diagnosed locally from aerosol lidar observations aboard the aircraft [DIAL-HSRL Mixed Layer Heights README, 2014] and typically extending to 1–3 km altitude. The states of Missouri (MO), Arkansas (AR), and Louisiana (LA) are indicated. The right panel shows the mean vertical profiles observed by the CAMS and ISAF instruments, and simulated by GEOS-Chem, for the Southeast US domain (30.5°–39°N, 95°–81.5°W) defined by the black rectangle in the top left panel. Horizontal bars represent observed standard deviations. GEOS-Chem is sampled along the flight tracks at the time of the measurements. The dashed black line shows the mean vertical CAMS profile in marine air over the Gulf of Mexico (22°–28°N, 96.5°–88.5°W), which is used in determining background HCHO column (0.40 × 10^{16} molecules cm^{-2}, see Section 2.3). The bottom left panel shows the mean HCHO columns on a 0.5° × 0.5° grid derived from the CAMS measurements after normalizing for temperature, for mixed layer depth, and for the contribution from HCHO aloft (see text in Section 2.3).
Diurnal variability of the HCHO columns is expected from models to be less than 10%, assuming a correctly simulated diurnal photochemical cycle, since photochemistry is both a source and a sink [Millet et al., 2008; Valin et al., 2016].

**Figure 2.2** (left panel) shows a point-to-point comparison of 1-minute averaged ISAF and CAMS HCHO observations (R3 version) aboard the aircraft. There is excellent correlation in the mixed layer \( r = 0.96 \) and above \( r = 0.99 \). Reduced major axis (RMA) regression of the two data sets yields a slope of \( 1.10 \pm 0.00 \), with ISAF 10% higher than CAMS due to the fact that the two instruments are independently calibrated. This difference is generally within the mutual stated accuracy for both instruments. The strong correlation between CAMS and ISAF provides confidence that they can be used for satellite validation purposes, and suggests that they can be used as equivalent data sets after 10% bias correction. We use CAMS in what follows.

**Figure 2.2** Comparisons between HCHO measurements from the CAMS and ISAF instruments aboard the SEAC4RS aircraft, and simulated by GEOS-Chem, for the Southeast US flight tracks (box in **Figure 2.1**). The left panel compares 1-minute measurements from CAMS and ISAF. The right panel compares GEOS-Chem and CAMS HCHO. Here and elsewhere for model-observation comparisons, HCHO observations along the flight tracks are averaged onto the GEOS-Chem grids \((0.25^\circ \times 0.3125^\circ, 47 \text{ vertical layers})\) and time steps (10 minutes). HCHO data points are colored by atmospheric pressure. Slopes and intercepts of reduced major axis (RMA) regressions are shown along with the correlation coefficient \( r \), sample size \( N \), RMA regression line (in blue), and 1:1 line.
The aircraft data show high concentrations in the mixed layer due to biogenic isoprene emission, and a sharp drop above the mixed layer because of the short lifetimes of isoprene (~1 h) and of HCHO itself (~2 h). Horizontal variability in the mixed layer reflects the density of isoprene-emitting vegetation but also surface air temperature (affecting isoprene emission) and mixing depth (affecting vertical mixing) at the time of the flights. We wish to convert the data to mean HCHO columns for the SEAC4RS period (5 August–25 September) in order to compare to the satellite data averaged over the same period. This requires time-averaging of the local surface air temperature and mixing depth, and conversion of the mixed layer concentration to a total column. We convert the aircraft HCHO mixing ratios in Figure 2.1 to HCHO columns by assuming uniform HCHO mixing ratios from the surface up through the local mixing depth measured from the aircraft [DIAL-HSRL Mixed Layer Heights README, 2014], an exponential decay from the top of the mixed layer to 650 hPa with a scale height of 1.9 km, and a fixed background of $0.40 \times 10^{16}$ molecules cm$^{-2}$ above, based on the mean vertical profiles information in Figure 2.1. Day-to-day variability in HCHO columns in the Southeast US is mainly driven by the temperature dependence of isoprene emission and can be fitted well by $\ln \Omega = 0.11T + 2.62$ ($r^2 = 0.64$), where $\Omega$ is in units of $10^{15}$ molecules cm$^{-2}$ and $T$[K] is the surface air temperature [Palmer et al., 2006; Zhu et al., 2014]. We applied this temperature dependence to the local HCHO columns inferred from the aircraft mixed layer data converted HCHO columns in order to correct for the difference between the local surface air temperature at the time of the flight and the local mean midday (1200–1300 local time) surface air temperature for the SEAC4RS period. Temperatures were taken from the Goddard Earth Observing System-Forward Processing (version 5.11.0, GEOS-FP hereafter) assimilated meteorological data product of the NASA Global Modeling and Assimilation Office (GMAO) [Molod et al., 2012].
Figure 2.3 Mean temperature, base isoprene emissions, and HCHO columns in the GEOS-Chem model for the SEAC4RS period (5 August–25 September 2013). The top panel shows the midday (1200–1300 local time) surface air temperature from the GEOS-FP assimilated meteorological data. The middle panel shows the MEGAN 2.1 base isoprene emissions from Guenther et al. [2012] for standard conditions (air temperature = 303 K, photosynthetic photon flux density = 200 µmol m$^{-2}$ s$^{-1}$ for sunlit leaves and 50 µmol m$^{-2}$ s$^{-1}$ for shaded leaves.) The bottom panel shows the GEOS-Chem HCHO columns computed with MEGAN 2.1 isoprene emissions and sampled at 1330 local time, under OMI-SA0 schedule, and filtered by OMI-SA0 quality flags and cloud conditions.
The bottom left panel of Figure 2.1 shows the resulting mean HCHO columns for the SEAC\textsuperscript{4}RS period as inferred from the CAMS measurements. We estimate the error in this mean HCHO columns is ~15%, which is mainly from the mixing depths, assumed background, scale height and temperature dependence. The spatial distribution is markedly different and smoother than for the original mixed layer data (top left panel), reflecting in large part the temperature normalization. Figure 2.3 shows the spatial distribution of midday temperatures for the SEAC\textsuperscript{4}RS period, along with base isoprene emissions at 303 K from the MEGAN 2.1 model [Guenther et al., 2012]. The base isoprene emissions feature a hotspot in the Ozarks region of Southeast Missouri, where there is dense oak cover. This region was repeatedly sampled by the aircraft on hot days. The HCHO aircraft observations are particularly high there but this feature is muted after correction for the mean August–September temperatures, which are much cooler in Missouri than further south. Inferred HCHO columns in Figure 2.1 are instead highest over Arkansas and Louisiana, where August–September temperatures are high.

We simulated the SEAC\textsuperscript{4}RS period using the GEOS-Chem v9-02 CTM (http://geos-chem.org) with 0.25° × 0.3125° horizontal resolution over North America driven by NASA GEOS-FP assimilated meteorological fields. The model has 47 vertical levels including 18 below 3 km. As can be seen in Figure 2.4, initial simulations of the SEAC\textsuperscript{4}RS data with GEOS-Chem pointed to a positive bias in the daytime GEOS-FP diagnostic for the height of the mixed layer (mixing depth), used in GEOS-Chem for surface-driven vertical mixing. Previous comparisons of GEOS-FP mixing depths to lidar and ceilometer data for other field studies in the Southeast US found a 30–50% high bias [Scarino et al., 2014; Millet et al., 2015]. For the SEAC\textsuperscript{4}RS simulation we decreased the GEOS-FP mixing depths by 40%, and comparison to the aircraft lidar measurements along the DC-8 flight tracks shows that this corrects the bias (red line in Figure 2.4). Corrected
afternoon (1200–1700 local time) GEOS-FP mixing depths along the flight tracks in the Southeast US average 1530 ± 330 m, compared to 1690 ± 440 m in the lidar data.

**Figure 2.4** Frequency distribution of mixed layer depths over the Southeast US during the SEAC⁴RS period (5 August–25 September 2013). Observations by aerosol lidar aboard the aircraft [DIAL-HSRL Mixed Layer Heights README, 2014] are compared to the local GEOS-FP data used to drive GEOS-Chem, before and after the 40% downward correction. The frequency distributions are constructed from 1-minute average data along the aircraft flight tracks over the Southeast US (box in Figure 2.1) for the 1200–1700 local time window.

Formaldehyde production in GEOS-Chem over the Southeast US in summer is mainly from isoprene. Companion papers by Fisher et al. [2016], Marais et al. [2016], and Travis et al. [2016] describe the GEOS-Chem simulation of isoprene chemistry in SEAC⁴RS and comparisons to aircraft and surface observations. Biogenic VOC emissions are from the MEGAN 2.1 model as implemented in GEOS-Chem by Hu et al. [2015] and with a 15% decrease applied to isoprene [Wolfe et al., 2015]. Surface-driven vertical mixing up to the mixing depth uses the non-local
mixing scheme of [Holtslag and Boville, 1993], as implemented in GEOS-Chem by Lin and McElroy [2010].

Figure 2.2 (right panel) compares simulated and observed HCHO mixing ratios along the SEAC4RS flight tracks, averaged over the GEOS-Chem grid and time step. Comparison for the ensemble of data shows high correlation \((r = 0.80)\) and no significant bias. Part of the correlation reflects the dependence on altitude, which is well captured by GEOS-Chem (Figure 2.1, right panel). After removing this dependence on altitude (by only examining observations within the mixed layer), the correlation between model and observations remains high \((r = 0.64)\) with only a small bias \((-3 \pm 2\%)\) indicated by the RMA linear regression. GEOS-Chem is less successful in reproducing the HCHO concentrations in the free troposphere (3–12 km, ~ 700–200 hPa), with a -41% normalized mean bias. This may be due to insufficient deep convection in the model.

Integration of the mean vertical profiles in Figure 2.1 indicates a mean GEOS-Chem HCHO column of \(1.46 \times 10^{16}\) molecules cm\(^{-2}\) over the Southeast US during the SEAC4RS period, which is 10% lower than observed by CAMS \((1.63 \times 10^{16}\) molecules cm\(^{-2}\)), and 23% lower than observed by ISAF \((1.90 \times 10^{16}\) molecules cm\(^{-2}\)). Here we assume there is no uncertainty associated with the CAMS mean column, and further assume that the 10% low bias in GEOS-Chem is valid for the SE US during the SEAC4RS period. The spatial correlation between GEOS-Chem mean HCHO columns (Figure 2.3, bottom panel) and the HCHO columns inferred from the CAMS data is 0.44 (0.47 for ISAF) on the 0.5° × 0.5° grid, with GEOS-Chem capturing the region of maximum HCHO in Arkansas and Louisiana.
2.4 Intercomparison and validation of satellite data sets over the Southeast US

**Figure 2.5** shows the spatial distribution of mean HCHO columns over the SEAC4RS period taken from the six satellite retrievals of **Table 2.1**, along with values from GEOS-Chem and columns inferred from the CAMS aircraft observations. All retrievals feature high values over the Southeast US due to isoprene emission and maximum values over and around Arkansas and Louisiana, consistent with GEOS-Chem and CAMS (**Figure 2.3**).

Spatial correlation coefficients between HCHO columns for different pairs of satellite retrieval data in **Figure 2.5** are given in **Table 2.2**. The correlation coefficients are computed for the temporally averaged (5 August–25 September 2013) data on the $0.5^\circ \times 0.5^\circ$ grid of **Figure 2.5** for the Southeast US domain (box in **Figure 2.1** and 2.5). Correlation coefficients for the different satellite retrievals are only 0.24–0.44 with CAMS but 0.38–0.85 with GEOS-Chem and typically 0.4–0.8 between pairs of retrievals. We conclude that there is consistency between retrievals in the spatial information even at the $0.5^\circ$ scale. The GOME2A-BIRA retrieval is noisier than the others and we attribute this to degradation of the instrument after 7 years of operations rather than its reduced swath mode operated since July 2013 [De Smedt et al., 2015], because the noise of GOME2A columns is significantly the same before and after the swath mode reduction.

We see from **Figure 2.5** that all retrievals are biased low relative to CAMS and GEOS-Chem. **Table 2.3** gives statistics for these biases as spatial averages for the Southeast US. GEOS-Chem columns are sampled on the same schedule and scenes as the individual retrievals, and are increased by 10% to correct for the bias with CAMS. Satellite retrieval biases relative to the corrected GEOS-Chem values range from $-20\%$ (OMI-BIRA) to $-51\%$ (OMPS-PCA). The GOME2A and GOME2B observations are made at 0930 local time, while the OMI and OMPS
observations are made at 1330 local time. GEOS-Chem columns increase by 6% from 0930 to 1330 and this is accounted for in the GEOS-Chem comparisons of Table 2.3.

![Map of HCHO vertical column densities over the Southeast US](image)

**Figure 2.5** HCHO vertical column densities over the Southeast US averaged over the SEAC4RS period (5 August–25 September 2013). The bottom panels show six retrievals from four satellites (OMI, GOME2A, GOME2B and OMPS) and three different groups (Table 2.1). The top panels show (1) GEOS-Chem model results sampled on the OMI schedule with filtering by OMI-SA0 quality flags and cloud conditions, and increased by 10% to correct for the bias relative to CAMS aircraft measurements; and (2) columns derived from the CAMS aircraft measurements (same as bottom left panel of Figure 2.1 but on a different color scale). The black rectangle represents the Southeast US domain (same as in Figure 2.1). Colorbar is a logarithmic scale.
Table 2.2 Spatial/temporal correlation coefficients (r) between pairs of HCHO column products\(^a\)

<table>
<thead>
<tr>
<th>HCHO product</th>
<th>OMI-SA0 (V003)</th>
<th>OMI-BIRA</th>
<th>GOME2A-BIRA (V14)</th>
<th>GOME2B-BIRA (V14)</th>
<th>OMPS-SA0</th>
<th>OMPS-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMI-SA0 (V003)</td>
<td>1/1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMI-BIRA</td>
<td>0.55/0.67</td>
<td>1/1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOME2A-BIRA (V14)</td>
<td>0.28/0.48</td>
<td>0.38/0.50</td>
<td>1/1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOME2B-BIRA (V14)</td>
<td>0.50/0.76</td>
<td>0.65/0.60</td>
<td>0.49/0.26</td>
<td>1/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMPS-SA0</td>
<td>0.48/0.77</td>
<td>0.70/0.50</td>
<td>0.45/0.55</td>
<td>0.72/0.76</td>
<td>1/1</td>
<td></td>
</tr>
<tr>
<td>OMPS-PCA</td>
<td>0.40/0.70</td>
<td>0.60/0.51</td>
<td>0.53/0.63</td>
<td>0.71/0.68</td>
<td>0.85/0.84</td>
<td>1/1</td>
</tr>
<tr>
<td>GEOS-Chem(^b)</td>
<td>0.38/0.88</td>
<td>0.50/0.65</td>
<td>0.68/0.82</td>
<td>0.85/0.88</td>
<td>0.74/0.86</td>
<td>0.82/0.75</td>
</tr>
<tr>
<td>Aircraft (CAMS)(^c)</td>
<td>0.24/–</td>
<td>0.44/–</td>
<td>0.26/–</td>
<td>0.35/–</td>
<td>0.43/–</td>
<td>0.37/–</td>
</tr>
</tbody>
</table>

\(^a\) Correlation coefficients between HCHO columns for different pairs of satellite retrievals, GEOS-Chem and CAMS aircraft observations. Values are for the southeast US domain (box in Figs. 1 and 5) during SEAC\(^4\)RS (5 August–25 September 2013). Spatial correlation coefficients are computed for the temporally averaged data on the 0.5° × 0.5° grid of Fig. 5. Temporal correlation coefficients are computed from daily averages of each retrieval over the southeast US domain.  

\(^b\) GEOS-Chem CTM columns sampled for the same scenes as the individual retrievals.  

\(^c\) Aircraft column data are temporal averages for the SEAC\(^4\)RS period as shown in Fig. 1 (bottom left panel) and Fig. 5 (top right panel).
Table 2.3 Satellite retrievals of HCHO columns over the southeast US

<table>
<thead>
<tr>
<th>Retrieval</th>
<th>Mean values&lt;sup&gt;b&lt;/sup&gt;</th>
<th>With CAMS shape factors</th>
<th>GEOS-Chem + 10%&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ω</td>
<td>ΔΩ&lt;sub&gt;S&lt;/sub&gt;</td>
<td>AMF&lt;sub&gt;G&lt;/sub&gt;</td>
</tr>
<tr>
<td>OMI-SA0 (V003)</td>
<td>1.06</td>
<td>0.65</td>
<td>2.66</td>
</tr>
<tr>
<td>OMI-BIRA</td>
<td>1.33</td>
<td>0.87</td>
<td>2.62</td>
</tr>
<tr>
<td>GOME2A-BIRA (V14)</td>
<td>0.89</td>
<td>0.62</td>
<td>2.37</td>
</tr>
<tr>
<td>GOME2B-BIRA (V14)</td>
<td>1.09</td>
<td>0.86</td>
<td>2.56</td>
</tr>
<tr>
<td>OMPS-SAO</td>
<td>1.09</td>
<td>0.72</td>
<td>2.54</td>
</tr>
<tr>
<td>OMPS-PCA</td>
<td>0.80</td>
<td>0.49</td>
<td>2.53</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mean values over the southeast US domain (box in Figs. 1 and 5) for the data in Fig. 5 collected during the SEAC<sup>d</sup>RS period (5 August–25 September 2013).<sup>b</sup> Mean values provided as part of the retrieval product including vertical HCHO columns (Ω), corrected slant columns (ΔΩ<sub>S</sub>), geometrical and scatter-corrected AMFs, and background correction (Ω<sub,o</sub>), following Eq. (1). Columns are in units of 10<sup>16</sup> molecules cm<sup>−2</sup>, and AMFs are dimensionless. The corrected slant columns and background correction are not reported in the SAO and OMPS-PCA retrievals and are reconstructed here to enable comparison with the other retrievals (see Sect. 2).<sup>c</sup> GEOS-Chem columns sampled for the same scenes as the individual retrievals and increased by 10% to correct for the bias relative to the SEAC<sup>d</sup>RS CAMS aircraft measurements (Fig. 1). Mean GEOS-Chem columns increase with time of day by 6.0% from 09:30 LT (GOME2A and GOME2B) to 13:30 LT (OMI and OMPS).<sup>d</sup> Normalized mean bias relative to the corrected GEOS-Chem values (last column in the table).<sup>e</sup> AMFs recalculated using the mean HCHO vertical shape factor from the CAMS aircraft instrument (Figs. 1 and 6) and the scattering weights or averaging kernels provided as part of the satellite product (Fig. 6).<sup>f</sup> Columns recomputed using AMFs constrained by the CAMS aircraft measurements.
Retrieval biases in the vertical column $\Omega$ could be contributed by the corrected slant column ($\Delta \Omega_S$), the $AMF$, and the background correction $\Omega_0$ (Equation 2.1). Table 2.3 gives mean values for these different terms. We see that the OMI-BIRA column is the highest because it has the highest $\Delta \Omega_S$ and lowest $AMF$, while the OMPS-PCA column is the lowest because its $\Delta \Omega_S$ is the lowest. OMPS-SAO and OMPS-PCA use the same OMPS spectra but the OMPS-SAO $\Delta \Omega_S$ are much higher and more consistent with the other retrievals. One caveat is that the derived $\Delta \Omega_S$ of OMPS-PCA may not be the best measure for its algorithm sensitivity, since OMPS-PCA doesn’t retrieve a slant column nor does it subtract the Pacific SCD to remove offsets, as described in Section 2.2.

GOME2A-BIRA columns average 18% lower than GOME2B-BIRA despite sharing the same retrieval algorithm and overpass time. This reflects instrument degradation as pointed out above. GOME2A performed much better during its first five years of operation (2007–2011) [De Smedt et al., 2012, 2015].

The OMI-BIRA retrieval has the smallest bias relative to the GEOS-Chem and CAMS HCHO columns, and this is due in part to its low $AMF$ (0.88). Figure 2.6 shows the mean reported scattering weights and shape factors for that retrieval (Equation 2.3), in comparison to other retrievals and to the CAMS aircraft observations. OMI-BIRA has lower scattering weights than the other retrievals, contributing to the lower $AMF$, and we discuss that below. The shape factors in the SAO (from GEOS-Chem CTM with horizontal resolution of $2^\circ \times 2.5^\circ$) and BIRA retrievals (from the IMAGES CTM with horizontal resolution of $2^\circ \times 2.5^\circ$) underestimate HCHO in the boundary layer and overestimate it in the free troposphere. With the correct shape factor from CAMS the OMI-BIRA retrieval has an even lower $AMF$ (0.74), as shown in Table 2.3, making it
even better in comparison to GEOS-Chem and to the aircraft data. The shape factor from ISAF is consistent with that from CAMS (Figure 2.1).

![Figure 2.6 Air mass factor differences between retrievals. The left panel shows mean scattering weights (w) and shape factors (S) for HCHO retrievals over the Southeast US during the SEAC4RS period, and the right panel shows the product of the two from which the AMF is derived by vertical integration using Equation 2.3. Values are shown for the OMI-SA0, OMI-BIRA, GOME2B-BIRA, and OMPS-PCA retrievals. Mean AMF values are given in legend. Also shown is the observed HCHO shape factor (black) from the mean CAMS profile in Figure 2.1. Table 2.3 also gives the AMFs for the other retrievals re-computed using CAMS shape factors. The differences with the original AMFs are less than 6% except for OMI-BIRA (14%). Although the results for OMI-BIRA illustrate how sensitive the AMF calculation is to the specification of shape factor, we find that this is not a significant source of bias in the other retrievals. This may reflect compensating errors in the vertical profile, as illustrated in Figure 2.6 with the OMI-SA0 shape factors in comparison with CAMS. We also compute AMFs using GEOS-Chem shape factors for each retrieval (Figure S1 in the Supplemental Materials), but we find this makes no difference to the results. 

| Table 2.3 also gives the AMFs for the other retrievals re-computed using CAMS shape factors. The differences with the original AMFs are less than 6% except for OMI-BIRA (14%). Although the results for OMI-BIRA illustrate how sensitive the AMF calculation is to the specification of shape factor, we find that this is not a significant source of bias in the other retrievals. This may reflect compensating errors in the vertical profile, as illustrated in Figure 2.6 with the OMI-SA0 shape factors in comparison with CAMS. We also compute AMFs using GEOS-Chem shape factors for each retrieval (Figure S1 in the Supplemental Materials), but we find this makes no difference to the results. |
When the AMFs for all retrievals are re-computed with common CAMS shape factors, as shown in Table 2.3, the remaining differences in AMFs are driven by viewing angles (as described by AMF<sub>G</sub> in Table 2.3), scattering weights and cloud parameters. Figure 2.6 shows that scattering weights are 10–30% higher in the OMI-SAO retrieval (AMF = 1.02) than in the OMI-BIRA retrieval (AMF = 0.85). The difference remains for cloud-free satellite pixels (cloud fraction < 0.01) and so is not due to different treatments of cloud effects. Surface reflectivity averages 0.048 in OMI-SAO and 0.037 in OMI-BIRA. Although both use the OMI surface reflectance climatology of Kleipool et al. [2008], OMI-SAO applies monthly mean reflectivities while OMI-BIRA applies monthly minimum reflectivities. This can explain some though not all of the difference in scattering weights. De Smedt et al. [2008] found that the HCHO AMF increases from 0.4 to 4.0 when the surface albedo changes from 0 to 1.

The background corrections (Ω<sub>o</sub> = 0.30–0.38 × 10<sup>16</sup> molecules cm<sup>-2</sup>) in the different retrievals are all consistent and amount to about 30% of the mean Ω over the Southeast US. They agree with background HCHO columns measured by aircraft over the remote North Pacific (0.37 ± 0.09 × 10<sup>16</sup> molecules cm<sup>-2</sup>, Table 8 in Singh et al. [2009]).

Previous studies have shown that variability in HCHO columns seen from space over the Southeast US in summer is mainly driven by the temperature dependence of isoprene emission [Palmer et al., 2006; Millet et al., 2008; Duncan et al., 2009; Zhu et al., 2014]. Figure 2.7 shows time series of daily HCHO columns averaged spatially over the Southeast US for the OMI-SAO and OMI-BIRA retrievals. All retrievals have day-to-day temporal coherence consistent with the temperature dependence of isoprene emission. Temporal correlation between the daily HCHO column and midday temperature is 0.52 for GOME2A-BIRA, 0.59 for OMPS-PCA, 0.59 for OMI-BIRA, 0.69 for GOME2A-BIRA, 0.71 for OMPS-SAO and 0.75 for OMI-SAO. GOME2A-BIRA
shows the lowest correlation with temperature, again likely due to noise from instrument degradation.

HCHO over the Southeast US in summer is mainly from oxidation of isoprene [Millet et al., 2006, 2008]. Satellite retrievals validated in this study show consistency in capturing both spatial and daily variations in HCHO columns, as demonstrated by the indirect validation between SEAC³RS observations and satellite retrievals. This supports their use as a quantitative proxy for isoprene emissions. However, the systematic low bias (20–51%) in the HCHO retrievals needs to be corrected. Our results show no indication of a pattern in the biases, suggesting that these could be removed as a uniform correction until better understanding is achieved.

**Figure 2.7** Daily variability of HCHO vertical column densities over the Southeast US during SEAC³RS. The top panel shows daily HCHO columns averaged over the Southeast US (box in Figure 2.5) for the OMI-SAO and OMI-BIRA retrievals. GEOS-Chem columns (black) are sampled following the OMI viewing geometry with filtering by OMI-SAO quality flags and cloud conditions, and scaled up by 10% on the basis of comparison with CAMS aircraft columns. The bottom panel shows the local midday (1200–1300 local time) surface air temperature over the Southeast US domain from the GEOS-FP assimilated meteorological data. Also shown for each data set is the temporal correlation coefficient ($r$) with temperature.
2.5 Conclusions

We have used SEAC^4RS aircraft observations of formaldehyde (HCHO) from two redundant in situ instruments over the Southeast US for 5 August–25 September 2013, together with a GEOS-Chem chemical transport model simulation at 0.25° × 0.3125° horizontal resolution, to validate and intercompare six HCHO retrievals from four different satellite instruments operational during that period. The combination of aircraft data and GEOS-Chem model fields provides strong constraints on the mean HCHO columns and their variability over the Southeast US, where high column amounts are driven by biogenic isoprene emission.

We find that the different retrievals show a large degree of consistency in their simulation of spatial and temporal variability. All retrievals capture the HCHO maximum over Arkansas and Louisiana seen in the aircraft data and in GEOS-Chem, and corresponding to the region of highest isoprene emission. Spatial correlation coefficients between retrievals are moderate to relatively high (0.4–0.8) even on a 0.5° × 0.5° grid. All retrievals are also consistent in their simulation of day-to-day variability correlated with temperature. This supports the use of HCHO columns observed from space as a proxy for isoprene emission. GOME2A-BIRA (launched in 2006) is noisier than other retrievals. We attribute this to instrument degradation.

Despite this success and consistency in observing HCHO variability from space, we find that all satellite retrievals are biased low in the mean, by 20% to 51% depending on the retrieval. This would cause a corresponding bias in estimates of isoprene emission made from the satellite data. The bias is smallest for OMI-BIRA and could be further reduced by correcting the assumed HCHO vertical profiles (shape factors) assumed in the AMF calculation. Other retrievals have larger biases that appear to reflect a combination of (1) spectral fitting affecting the corrected slant columns, and (2) scattering weights in the radiative transfer model affecting the AMF. Aside from
OMI-BIRA, the shape factors used in the retrievals are not a significant source of error in determining the $AMF$.

Our work points to the need for improvement in satellite HCHO retrievals to correct the mean low bias. We find no evident spatial or temporal pattern in the bias, at least for the Southeast US in summer, that would compromise the interpretation of the satellite data to estimate patterns of isoprene emission. The biases may be removed by applying uniform correction factors until better understanding is achieved.

**Acknowledgments**

We acknowledge contributions from the NASA SEAC$^4$RS Science Team. We would also like to thank the SEAC$^4$RS flight crews and support staff for their outstanding efforts in the field. This work was funded by the US National Aeronautics and Space Administration. We thank Michel Van Roozendael for helpful discussions. Jenny A. Fisher acknowledges support from a University of Wollongong Vice Chancellor’s Postdoctoral Fellowship. We thank three anonymous reviewers who provided thorough and thoughtful comments.
References


Chapter 3 Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns

Abstract

Satellite observations of formaldehyde (HCHO) columns provide top-down constraints on emissions of highly reactive volatile organic compounds (HRVOCs). This approach has been used previously in the US to estimate isoprene emissions from vegetation, but application to anthropogenic emissions has been stymied by lack of a discernable HCHO signal. Here we show that temporal oversampling of HCHO data from the Ozone Monitoring Instrument (OMI) for 2005–2008 enables detection of urban and industrial plumes in eastern Texas including Houston, Port Arthur, and Dallas/Fort Worth. By spatially integrating the HCHO enhancement in the Houston plume observed by OMI we estimate an anthropogenic HCHO source of 250 ± 140 kmol h\(^{-1}\). This implies that anthropogenic HRVOC emissions in Houston are 4.8 ± 2.7 times higher than reported by the US Environmental Protection Agency inventory, and is consistent with field studies identifying large ethene and propene emissions from petrochemical industrial sources.

3.1 Introduction

Anthropogenic highly reactive volatile organic compounds (AHRVOCs) with atmospheric lifetimes of less than a day are important precursors of ozone and organic aerosols in urban air and industrial plumes. Their sources are poorly quantified in emission inventories, as shown by air quality studies in eastern Texas [Ryerson et al., 2003; Parrish et al., 2012] and in an oil/gas field of northern Colorado [Gilman et al., 2013]. Satellite column measurements of formaldehyde (HCHO), a high-yield product from atmospheric oxidation of VOCs, have been used to constrain AHRVOC emissions in East Asia [Fu et al., 2007] and Nigeria [Marais et al., 2014a]. However, detection of AHRVOC emissions in the US from satellite HCHO data has been elusive [Martin et al., 2004; Millet et al., 2008]. The highest-resolution data are from the Ozone Monitoring Instrument (OMI), which provides daily global coverage of HCHO columns by cross-track scanning with $13 \times 24$ km$^2$ nadir pixel resolution [Levelt et al., 2006]. An analysis of OMI urban data in the US by Boeke et al. [2011] found only weak HCHO enhancements in the New York and Los Angeles urban cores in summer, and in the Houston urban core in spring and fall.

The difficulty of observing US AHRVOC emissions from space likely reflects their relatively small magnitude. The single-retrieval detection limit for OMI HCHO is $2 \times 10^{16}$ molecules cm$^{-2}$ [Millet et al., 2008], which corresponds to $\sim 4$ ppb HCHO in a 2-km deep boundary layer. HCHO concentrations of $\sim 10$ ppb are commonly observed in urban air and industrial plumes [Wert et al., 2003; Buzcu Guven and Olaguer et al., 2011; Lin et al., 2012; Zheng et al., 2013] but would be diluted on the scale sampled by the satellite pixels. Temporal averaging of the satellite data greatly improves the detection limit [Boeke et al., 2011], though quantifying this improvement is difficult as it depends on the random vs. systematic character of the retrieval error. The urban
signal can also be masked by large regional emissions of isoprene, the dominant biogenic HRVOC contributing to HCHO [Palmer et al., 2003; Martin et al., 2004; Boeke et al., 2011].

Here we demonstrate that quantitative detection of AHRVOC emissions in eastern Texas can be achieved by oversampling of the OMI HCHO data. “Oversampling” refers to temporal averaging of the satellite data on a spatial grid finer than the pixel resolution on the instrument. The technique achieves high signal-to-noise ratio at high spatial resolution by sacrificing temporal resolution, i.e., averaging over a long time period. It takes advantage of the spatial offset and changing geometry (from off-track viewing) of the satellite pixels from day to day. Oversampling of OMI data has been applied previously with success to detection of \( \text{SO}_2 \) and \( \text{NO}_2 \) from urban and point sources [de Foy et al., 2009; Fioletov et al., 2011; McLinden et al., 2012; Lu et al., 2013]. We demonstrate here its application to HCHO.

3.2 Data and methods

OMI is a UV/Vis nadir solar backscatter spectrometer launched in 2004 on the Aura satellite [Levelt et al., 2006]. It achieves daily global coverage with an equator crossing time of 1338 local time. HCHO slant column densities (SCD) along the solar backscatter optical path are fitted in the spectral window 327.5–356.5 nm [Chance et al., 2000]. We use OMI HCHO Version 2.0 (Collection 3) SCD retrievals for 2005–2008 (http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omhcho_v003.shtml) that (1) pass all fitting and statistical quality checks, (2) have a cloud fraction less than 0.3 and solar zenith angle less than 60°, and (3) are not affected by the “OMI row anomaly” (http://www.knmi.nl/omi/research/product/rowanomaly-background.php). Drift from instrument aging [Marais et al., 2012] is removed with a linear temporal regression of background SCD over the North Pacific (130°–125°W, 35°–40°N).
The air mass factor (AMF = SCD / VCD) to convert SCD to vertical column density (VCD, column hereafter) is computed following Palmer et al., [2001] with the LIDORT radiative transfer model [Spurr et al., 2001]. Satellite viewing geometry, cloud fraction, and cloud centroid pressure are from the OMI data. The AMF calculation requires information on HCHO and aerosol vertical distributions, and these are specified locally from the GEOS-Chem model with 0.5° × 0.667° horizontal resolution over North America [Zhang et al., 2011]. Oversampling uses a higher horizontal resolution for the OMI data than 0.5° × 0.667°, but we expect that the error from subgrid variability in the HCHO and aerosol vertical distributions is small relative to other sources of error.

Wintertime observations would avoid biogenic interference on HCHO but we find that OMI HCHO columns are then indistinguishable from noise over the US including over Houston. This does not reflect loss of measurement sensitivity as the mean AMF over Houston in winter (December–February; AMF = 0.99 ± 0.17) is actually higher than in summer (May–August; AMF = 0.78 ± 0.13). The higher AMF in winter is due to longer light path and lower cloud cover, more than compensating for the effects of reduced UV light penetration and shallower planetary boundary layer (PBL). We attribute the lack of detectable HCHO in winter to low OH concentrations, delaying the oxidation of AHRVOCs to HCHO and thus smearing the HCHO signal. Average 1200–1500 (local time) OH concentrations in the GEOS-Chem model over Houston are a factor of 5 lower in December–February than in May–August. Mean surface wind speeds are 30% higher in December–February than May–August (http://www.wunderground.com), also contributing to the smearing. We limit our attention here to May–August HCHO columns.

We oversample the OMI HCHO data over eastern Texas (99.5°–92.5°W, 28°–34°N) for May–August 2005–2008 by averaging individual pixels onto a 0.02° × 0.02° (~ 2 × 2 km²) grid. In this procedure, the column measurement for a given pixel is assumed to apply to a circle defined
by the center point of the pixel and an “averaging radius” of 24 km. Previous OMI oversampling studies have used the same strategy. Fioletov et al. [2012] chose an averaging radius of 12 km for oversampling OMI SO$_2$ pixels, and McLinden et al. [2012] used 20 km and 24 km for OMI NO$_2$ and SO$_2$ pixels, respectively. We find that a 12-km or 20-km averaging radius for HCHO pixels leads to excessive noise. Our oversampling approach leads to ~ 800 OMI measurements being averaged in each 0.02° × 0.02° grid square. We develop a new “oversampling” method based on overlapped areas and retrieval uncertainty. More details can be found in Chapter 4.

3.3 Results and discussion

Figure 3.1 (left) shows the oversampled OMI HCHO concentrations for May–August 2005–2008. Urban/industrial sources in and near Houston, Port Arthur, and Dallas/Fort Worth are clearly detected. Urban areas of Austin and San Antonio are less industrial and only marginally detected. The Houston and Port Arthur plumes are transported northward by the prevailing SSE wind. The enhancement west of Dallas/Fort Worth can be attributed to AHRVOC emissions from the Barnett Shale, the largest onshore natural gas field in the US. The high values over Northeast Texas are due to isoprene emission as discussed below.

HCHO columns over the Houston urban area peak at 1.4 × 10$^{16}$ molecules cm$^{-2}$ near the Houston ship channel where major refineries and petrochemical industries emit large amounts of HRVOCs. Johansson et al. [2013] previously reported HCHO columns in the channel of up to 2.4 × 10$^{16}$ molecules cm$^{-2}$ from ground-based remote sensing in May 2009. Our mean column of 9.4 × 10$^{15}$ molecules cm$^{-2}$ in the Houston-Galveston-Brazoria urban metropolitan area (HGB; Figure 3.1, thin blue line) corresponds to a mean HCHO mixing ratio of 2.4 ppb for a 1.7-km deep PBL [Haman et al., 2012]. This result agrees with the mean HCHO concentration of 2.4 ppb measured
in the HGB during the summer 2006 Texas Air Quality Study (TexAQS) [Gilman et al., 2009]. HGB concentrations measured during that study ranged from 1 to 20 ppb [Zhang et al., 2013].

Figure 3.1 OMI HCHO columns and HRVOC emission inventories for eastern Texas. The left panel shows OMI HCHO columns averaged over May–August 2005–2008 and oversampled to a $0.02^\circ \times 0.02^\circ$ resolution using an averaging radius of 24 km. Crosses indicate city centers. The right panel shows the May–August 2008 biogenic isoprene emissions and the major anthropogenic HRVOC (AHRVOC, Table 3.1) point sources (dots) with emission larger than 3 kg C h$^{-1}$ binned on a $0.02^\circ \times 0.02^\circ$ grid. Isoprene emissions are computed with MEGAN v2.1 [Guenther et al., 2012]. AHRVOC emissions are from the 2005 National Emissions Inventory (NEI05) of the US Environmental Protection Agency (EPA) as implemented by Stuart McKeen [Brioude et al., 2011; Kim et al., 2011]. The Houston plume outline (gray) is used in the text to estimate AHRVOC emissions from the HCHO data. The boundary of the Houston-Galveston-Brazoria urban metropolitan area (HGB) is shown as the thin blue outline.

The right panel of Figure 3.1 shows HRVOC emissions for eastern Texas estimated from current inventories. These include MEGAN v2.1 for biogenic isoprene [Guenther et al., 2012] and the 2005 National Emission Inventory (NEI05) of the US Environmental Protection Agency (EPA) as implemented by Stuart McKeen (see Brioude et al., [2011]; Kim et al., [2011]). Here AHRVOCs are defined as having atmospheric lifetimes of less than 1 day against oxidation by OH. Table 3.1
lists the main NEI05 AHRVOCs emitted in the Houston plume area defined in Figure 3.1 (gray line). Ethene and propene are the most important HCHO precursors, as also observed in the TexAQS campaigns [Wert et al., 2003; Parrish et al., 2012]. Wert et al. [2003] found from speciated VOC samples that 78% of the HCHO production potential was from terminal alkenes including 30% from ethene, 22% from propene, 14% from isoprene, and 12% from other alkenes.

Biogenic isoprene makes a large background contribution to HCHO over eastern Texas, as seen in Figure 3.1 and previously noted by Martin et al. [2004]. Isoprene emissions in MEGAN v2.1 are particularly high over forested Northeast Texas, explaining the high OMI HCHO columns there. Some distinction between biogenic and anthropogenic contributions to OMI HCHO can be made on the basis of correlation with surface air temperature. Isoprene emission increases exponentially with temperature [Guenther et al., 2006], and this dependence is apparent in regional HCHO satellite data over the Southeast US [Palmer et al., 2006]. Figure 3.2 shows the relationships of OMI HCHO with surface air temperature over Northeast Texas and the Houston core for May–September 2006–2008. The data over Northeast Texas show a strong exponential relationship with temperature ($R^2 = 0.64$) with an argument of 0.11 K$^{-1}$, consistent with that expected for isoprene emission [Guenther et al., 2006; Palmer et al., 2006]. By contrast, OMI HCHO columns over the Houston core show no significant relationship with temperature ($R^2 = 0.03$), supporting the dominant anthropogenic influence. The lack of correlation in the Houston data partly reflects a cluster of three points in Figure 3.2 with T > 300 K and HCHO column < 2 $\times$ 10$^{15}$ molecules cm$^{-2}$, but even excluding these points the relationship with temperature yields only $R^2 = 0.34$. Some correlation of HCHO with temperature would be expected even over Houston due to (1) the regional HCHO background contributed by isoprene [Wert et al., 2003], (2) the
temperature dependence of AHRVOCs oxidation, and (3) the association of high temperature with stagnation.

Table 3.1 NEI estimates of Houston AHRVOC emissions and HCHO yields

<table>
<thead>
<tr>
<th>Species</th>
<th>Lifetime b (h)</th>
<th>Emission a (kmol h⁻¹)</th>
<th>Molar HCHO yield c</th>
<th>HCHO production (kmol h⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethene</td>
<td>2.9</td>
<td>16</td>
<td>1.6</td>
<td>27</td>
</tr>
<tr>
<td>Propene</td>
<td>0.8</td>
<td>6.3</td>
<td>1.8</td>
<td>12</td>
</tr>
<tr>
<td>Higher Alkenes</td>
<td>0.6 d</td>
<td>5.3</td>
<td>0.6 d</td>
<td>3.2</td>
</tr>
<tr>
<td>Formaldehyde</td>
<td>1.6 e</td>
<td>9.4</td>
<td>1.0</td>
<td>9.4</td>
</tr>
<tr>
<td>Acetaldehyde</td>
<td>1.6</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Other f</td>
<td>0.4 g</td>
<td>0.7</td>
<td>0.6 g</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>–</td>
<td>39</td>
<td>–</td>
<td>52</td>
</tr>
</tbody>
</table>

a Mean emissions for May–August 2005 from the US EPA National Emission Inventory (NEI05) implemented by Stuart McKeen [Brioude et al., 2011; Kim et al., 2011], ftp://aftp.fsl.noaa.gov/divisions/taq) over the area of the Houston plume defined by the OMI HCHO data (Figure 3.1, gray outline).

b Lifetime against oxidation by OH computed using a mean OH concentration of 1.1 × 10⁷ molecules cm⁻³ for 0930–1330 local time taken from Mao et al. [2009]. Kinetic data are from the Master Chemical Mechanism (MCM) v3.2 [Jenkin et al., 1997; Saunders et al., 2003] (http://mcm.leeds.ac.uk/MCM).

c Prompt yield of HCHO realized within 1 day of initial oxidation, as computed using MCM v3.2 for oxidation by OH in the high-NOₓ regime.

d Using 1-butane as representative species.

e Also includes photolysis, using a mean 0930–1330 photolysis frequency of 8.1 × 10⁻⁵ s⁻¹ computed with the Tropospheric Ultraviolet and Visible (TUV) Radiation Model (http://cprm.acd.ucar.edu/Models/TUV). HCHO loss is 46% from photolysis and 54% from oxidation by OH.

f Including dienes, glyoxal, and styrene.

g Using 1,3-butadiene as representative species.
Figure 3.2 Relationship of OMI HCHO column with surface air temperature for the Houston urban core (95.5°–95°W, 29.5°–30°N) and Northeast Texas (95°–93.5°W, 32°–33.5°N). Temperatures are 1300–1400 local time values from the NASA Modern-era Retrospective Analysis for Research and Application (MERRA). Individual points are 10-day averages for May–September 2006–2008. Coefficients of determination ($R^2$) are shown inset. The green solid line is an exponential fit of the HCHO column ($\Omega$) to the surface air temperature ($T$) over Northeast Texas as $\ln \Omega = 2.62 + 0.11T$.

A number of previous studies have used satellite HCHO data to constrain isoprene emissions by assuming a local relationship between the two from a chemical transport model (CTM) [Palmer et al., 2003; Fu et al., 2007; Millet et al., 2008; Barkley et al., 2008; Marais et al., 2012] or by applying a more elaborate inversion method [Shim et al., 2005; Dufour et al., 2009; Stavrakou et al., 2009; Curci et al., 2010]. In regions of the world where anthropogenic HCHO is readily discernable, these approaches have also been used to constrain AHRVOC emissions [Shim et al., 2005; Fu et al., 2007; Marais et al., 2014b]. In our case, the AHRVOC enhancement is on top of a large regional background (Figure 3.1). We constrain AHRVOC emissions for the Houston plume area by integrating the OMI HCHO enhancement over the Houston plume as the difference between the observed HCHO column ($\Omega$) and the regional background ($\Omega_0$) contributed...
by biogenic and long-lived anthropogenic emissions. From the HCHO lifetime ($\tau_{HCHO}$) we deduce the corresponding HCHO source $S$ per unit time as

$$S = \frac{1}{\tau_{HCHO}} \iint (\Omega - \Omega_0) dA \tag{3.1}$$

where the integral is over the area $A = 1.9 \times 10^4$ km$^2$ of the plume as defined in Figure 3.1. This represents the total emission of AHRVOCs within the plume area weighted by their prompt yield of HCHO. We can convert this quantity to a total AHRVOC emission ($E$) by applying independent estimates of the fraction $f_i$ of the total emission contributed by species $i$ and the corresponding HCHO yield $Y_i$ (Table 3.1):

$$E = \frac{S}{\sum_i f_i Y_i} \tag{3.2}$$

Here we estimate the regional background ($\Omega_0$) as the HCHO column downwind of the discernible Houston plume, corresponding roughly to the green color in Figure 3.1 (8–10 $\times 10^{15}$ molecules cm$^{-2}$).

Figure 3.3 shows HCHO columns averaged across the prevailing wind as a function of the distance from Houston city center. The plume decays to a regional background value at about 110 km downwind of the city center. For a mean wind speed of 3.3 m s$^{-1}$ (http://www.wunderground.com) this corresponds to an aging time of 9.3 hours, long relative to the lifetimes of ethene and propene (Table 3.1). From the downwind asymptote of the plume we derive a best estimate for the regional background of $\Omega_0 = 9.6 \pm 0.3 \times 10^{15}$ molecules cm$^{-2}$ (Figure 3.3). The actual uncertainty in the regional background is likely larger considering the fine-scale heterogeneity in isoprene emissions [Gulden and Yang, 2006]. We therefore adopt $\Omega_0 = 9.6 \pm 0.5 \times 10^{15}$ molecules cm$^{-2}$ as a more conservative estimate of the uncertainty. More accurate $\Omega_0$ can be obtained by examining the mean downwind structure of HCHO columns at different wind
directions, but this is beyond the scope of this work. The uncertainty in the Houston plume enhancement ($\Omega - \Omega_0$) is largely defined by the uncertainty in $\Omega_0$, considering that any systematic errors in the retrieval would likely be canceled in computing the difference between $\Omega$ and $\Omega_0$. We thus obtain a total HCHO column enhancement integrated over the plume of $400 \pm 180$ kmol (red area in Figure 3.3).

![Figure 3.3](image)

**Figure 3.3** Cross-section of the Houston HCHO plume along the direction of the prevailing SSE wind. The inset map is an excerpt from Figure 3.1 delineating the Houston plume (black line, same as the gray line in the right panel of Figure 3.1). The Figure shows the mean HCHO columns for May–August 2005–2008 averaged across the plume width as a function of downwind distance from the Houston city center (open circle on the map). Green dashed lines bracket the regional background ($\Omega_0$) defined by the HCHO columns beyond the extent of the plume. The HCHO enhancement in the plume is shown as red fill.
Loss of HCHO is by photolysis and oxidation by OH. From Table 3.1, we estimate $\tau_{\text{HCHO}}$ to be 1.6 h. Photolysis accounts for half of total HCHO loss and is relatively well constrained. Loss by reaction with OH is not as well constrained because of uncertainty in OH concentrations. Here we adopted a mean OH concentration of $1.1 \times 10^7$ molecules cm$^{-3}$ at 0930–1330 local time from measurements in downtown Houston [Mao et al., 2009] and estimate the overall uncertainty in $\tau_{\text{HCHO}}$ to be 30%. The resulting anthropogenic HCHO source $S$ over the area of the plume is $250 \pm 140$ kmol HCHO h$^{-1}$, propagating in quadrature our estimated uncertainty in $\tau_{\text{HCHO}}$.

Combining this result with the data on $f_i$ and $Y_i$ in Table 3.1, we deduce a total AHRVOC emission $E$ for the HGB of $190 \pm 100$ kmol h$^{-1}$, which can be compared to the NEI05 estimate of 39 kmol h$^{-1}$ from Table 3.1. The OMI observations thus suggest that AHRVOC emissions in the Houston plume area are underestimated by a factor of $4.8 \pm 2.7$ in the NEI05 inventory for 2005–2008. This is consistent with previous studies pointing to a large underestimate of alkene emissions in the HGB [e.g., Wert et al. 2003; de Gouw et al., 2009; Parrish et al., 2009; Mellqvist et al., 2010]. Our estimate of the HCHO source in the Houston plume is consistent with the estimate of $240 \pm 90$ kmol HCHO h$^{-1}$ from Parrish et al. [2012] computed using an improved alkene inventory [Kim et al., 2011] with updated ethene and propene emission factors from petrochemical facilities [Mellqvist et al., 2010].

Previous analyses of HCHO data in Houston have reached contradictory conclusions on whether most of the HCHO is primary, i.e., directly emitted [Rappenglück et al., 2010; Buzcu Guven and Olaguer, 2011; Johansson et al., 2013; Olaguer, 2013; Olaguer et al., 2013] or secondary, i.e., produced within the plume from alkene oxidation [Friedfeld et al., 2002; Wert et al., 2003; Parrish et al., 2012; Zhang et al., 2013]. The distinction is important because primary HCHO would accumulate at night and photolyze in early morning, providing a source of radicals.
to initiate ozone formation. Our inability to detect the Houston urban plume in winter from the OMI data (see above) argues against a major primary source of HCHO. We attempted to constrain the speciation of AHRVOCs by using the shape of the OMI plume in Figure 3.3 and a simple constant-wind model, as primary HCHO would decay closer to the core. We were unsuccessful, partly because of the complexity arising from primary emissions at night and in early morning when the HCHO lifetime is long.

We see from Figures 3.1 and 3.3 that the anthropogenic HCHO enhancements in the Houston plume and elsewhere are on top of a larger regional background ($\Omega_0$). This background is dominantly from biogenic isoprene. For the Houston plume area in Figure 3.3, the 24-h average isoprene emission calculated by MEGAN in May–August is 95 kmol h$^{-1}$. Assuming a HCHO molar yield of 2.3 from isoprene oxidation [Millet et al., 2006], this yields a HCHO production rate of 220 kmol h$^{-1}$, comparable in magnitude to the anthropogenic source.

Observations from the TexAQS aircraft campaigns in 2000 and 2006 documented a decrease of AHRVOC emissions from the HGB over that period. Gilman et al. [2009] reported 56% and 51% decreases in ethene and propene median concentrations, respectively; Washenfelder et al. [2010] found emission decreases in the Houston ship channel of 41% for ethene and 8% for propene from 1999 to 2006. Presently the useful OMI HCHO record is limited to 2005–2008; data after 2008 are too noisy for trend analysis because of the row anomaly. De Smedt et al. [2010] used data from the GOME and SCIAMACHY satellite instruments to infer global 1997–2009 trends in HCHO, but we find that the pixel resolution of these instruments is too coarse for detection of the Houston plume. Post-2008 OMI data are expected to be corrected in a future product [González Abad et al., 2015], which will then allow analysis of AHRVOC trends as well
as examination of AHRVOC emissions associated with the large increase in oil/gas exploration across the US over the past five years.

3.4 Conclusions

We have shown that multi-year oversampling of summertime OMI HCHO satellite data enables detection of HCHO enhancements from large US urban/industrial sources of anthropogenic highly reactive volatile organic compounds (AHRVOCs). The enhancement in the Houston urban plume is sufficiently extensive to allow quantitative interpretation in terms of AHRVOC emissions. Our resulting estimate of AHRVOC emissions for Houston is $4.8 \pm 2.7$ times higher than the US EPA inventory and consistent with previous field estimates that identified large ethene and propene emissions from the petrochemical industry. The lack of detectable OMI HCHO enhancements in winter suggests that anthropogenic HCHO is mainly produced by photochemical oxidation of alkenes rather than directly emitted.

Acknowledgements

This work was supported by the NASA Aura Science Team and Air Quality Applied Sciences Team. The authors thank Barry Lefer and James Flynn at the University of Houston and Xiong Liu at the Harvard-Smithsonian Center for Astrophysics for their help. We thank two anonymous reviewers who provided thorough and thoughtful comments.
References


Parrish, D. D., et al. (2009), Overview of the Second Texas Air Quality Study (TexAQS II) and the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS), J. Geophys. Res., 114, D00F13, doi:10.1029/2009JD011842.


Chapter 4 Long-term trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing emissions of volatile organic compounds

Abstract

Satellite observations of formaldehyde (HCHO) columns provide top-down information on emissions of highly reactive volatile organic compounds (VOCs). We examine the long-term trends in HCHO columns observed by the Ozone Monitoring Instrument (OMI) from 2005 to 2014 across North America. Biogenic isoprene is the dominant source of HCHO and its emission has a large temperature dependence. After correcting for this dependence, we find significant long-term trends due to changes in anthropogenic emissions, land cover, agriculture, and open fires. We find that HCHO columns over the Houston-Galveston-Brazoria industrial area decreased by 2.2% a⁻¹, consistent with trends in anthropogenic VOCs. HCHO columns increased by 4.7% a⁻¹ over the Athabasca oil sands in Alberta, reflecting increasing bitumen production, and decreased by 1.8% a⁻¹ over the Floyd shale in Alabama, reflecting decreased production of natural gas. Other oil/gas production regions do not show significant HCHO enhancements and trends. An increase of 4.7% a⁻¹ over the northwestern US may be due to reforestation. Variation in HCHO columns may also be associated with changing crop cover in the midwest US and with changing wildfire activity in California. Except in the southeast US, the impact of declining NOₓ emission over the US on HCHO columns is small (< 10%).
4.1 Introduction

Formaldehyde (HCHO), a high-yield product from the atmospheric oxidation of volatile organic compounds (VOCs), is detectable from space as a total column by solar UV backscatter [Chance et al., 2000]. Remotely sensed HCHO columns have been used to constrain VOC emissions from biogenic sources [Palmer et al., 2003; Millet et al., 2008; Stavrakou et al., 2009; Marais et al., 2012; Barkley et al., 2013], anthropogenic sources [Fu et al., 2007; Marais et al., 2014; Zhu et al., 2014] and open fires [Shim et al., 2005; Gonzi et al., 2011]. Previous research has diagnosed long-term trends in HCHO columns from the satellite data to infer VOC emission trends [De Smedt et al., 2010, 2015; Khokhar et al., 2015; Mahajan et al., 2015]. De Smedt et al. [2010, 2015] found significant downward trends in US cities. Interpretation is complicated by the strong temperature dependence of the emission of biogenic isoprene, which dominates the interannual variability of HCHO columns [Abbot et al., 2003; Palmer et al., 2006; Duncan et al., 2009]. Here we use a 10-year record (2005–2014) of HCHO columns over North America from the Ozone Monitoring Instrument (OMI) to infer long-term trends, correcting for the temperature dependence of isoprene emission. We make a first attempt to attribute these trends to changes in anthropogenic emissions, land cover, and other factors.

Most of the HCHO column over North America during the growing season is from isoprene [Miller et al., 2006]. Outside of the growing season HCHO is undetectable [Abbot et al., 2003], which reflects both the lack of isoprene and the slow rate of photochemical oxidation of other VOCs [Zhu et al., 2014]. Isoprene emission doubles for every 7 K temperature increase [Guenther et al., 2012]. This temperature dependence drives about 80% of the variability in HCHO columns over the eastern US on a day-to-day or month-to-month basis [Palmer et al., 2006; Zhu et al., 2016a]. It is thus important to account for in trend analyses.
4.2 OMI HCHO column data and temperature correction

OMI is a UV/Vis nadir solar backscatter spectrometer launched in 2004 on the Aura satellite in a polar sun-synchronous orbit [Levett et al., 2006]. It observes the whole globe daily at 1330 local time. We use OMI HCHO Version 2.0 (Collection 3) retrievals [González Abad et al., 2015] from the Smithsonian Astrophysical Observatory (OMI-SAO). The single-scene precision is $1 \times 10^{16}$ molecules cm$^{-2}$, with a relative uncertainty of 45–105% [González Abad et al., 2015]. The precision can be improved by spatial and temporal averaging [De Smedt et al., 2008; Boeke et al., 2011; Zhu et al., 2016b].

Zhu et al. [2016a] validated the OMI-SAO HCHO columns with high-quality aircraft measurements from the SEAC4RS flight campaign over the southeast US in August–September 2013 [Toon et al., 2016]. The OMI-SAO data showed strong spatial correlation with the aircraft, and day-to-day temporal variability consistent with that expected from isoprene emission. They were biased low by 37% relative to the aircraft data, and here we apply a uniform correction factor of 1.6 to account for this. We use data for May–September when the HCHO data are generally well above the detection limit. We select only those data that (1) pass all fitting and statistical quality checks (MainDataQualityFlag = 0), (2) have cloud fraction less than 0.3 and solar zenith angle less than 60°, (3) are from rows 1–20 and 55–60 of the OMI detector, and (4) have a vertical column density within the range of $-0.5 \times 10^{16}$ molecules cm$^{-2}$ to $1.0 \times 10^{17}$ molecules cm$^{-2}$. The third criterion excludes those data points affected by the growing problem of OMI “row anomalies” (http://projects.knmi.nl/omi/research/product/rowanomaly-background). The last criterion excludes 5.8% of data that pass the other quality tests.

OMI-SAO HCHO columns show significant drift caused by instrument aging [Marais et al., 2012; Zhu et al., 2014]. We correct for the drift by temporal regression of the zonal mean
monthly averaged HCHO columns in 0.5-degree latitudinal increments over the remote Pacific (160°–140°W, 20°–60°N), where no trend in HCHO is expected. The drift shows strong latitudinal dependence, varying from \( \sim 3.2 \times 10^{12} \) molecules cm\(^{-2}\) per month at 30° N to \( \sim 1.2 \times 10^{13} \) molecules cm\(^{-2}\) per month at 60° N.

We compute the temperature-corrected changes in HCHO columns from 2005 to 2014 as follows. (1) We first assign a temperature to each satellite pixel based on its location and passing time, using the hourly surface skin temperature data (0.5° × 0.667°) from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). (2) We then allocate the satellite pixels into 120 temperature bins in increments of 0.25 K over the range of 288 K to 318 K. (3) For each temperature bin \( i \), we divide the array of satellite pixels into two periods, 2005–2009 and 2010–2014, denoted as periods \( 1 \) and \( 2 \). (4) We then follow the averaging method of Zhu et al. [2016b] to map the HCHO column amounts for the two periods onto a 0.5° × 0.5° horizontal grid, with \( N_{i,j,1} \) and \( N_{i,j,2} \) overlapping pixels collected in each gridcell \( j \) for the two periods. (5) The change in mean HCHO column from 2005–2009 to 2010–2014 for temperature increment \( i \) at gridcell \( j \) can then be calculated as

\[
\Delta \bar{N}_{i,j} = \bar{N}_{i,j,2} - \bar{N}_{i,j,1}
\]  

(4.1)

where \( \bar{N}_{i,j,1} \) and \( \bar{N}_{i,j,2} \) represent the mean column HCHO amounts for each period. For quality control, we consider valid only those changes in gridcells with at least 25 measurements in both periods in that temperature bin. (6) Finally, we compute the temperature-independent change in column HCHO for gridcell \( j \) as the mean of the changes across all the temperature bins, weighted by the sum of pixels in each bin:
\[
\Delta \overline{\Omega}_j = \frac{\sum_l \Delta \overline{\Omega}_{l,j} N_{l,j}}{\sum_l N_{l,j}} \tag{4.2}
\]
\[
N_{l,j} = N_{l,j,1} + N_{l,j,2} \tag{4.3}
\]

Here we consider only those gridcells with valid changes in at least 20 temperature bins. We use the \( t \)-test to check the robustness of \( \Delta \overline{\Omega}_j \), with \( p < 0.05 \) as the threshold for statistical significance. By propagating a random retrieval error of 105\% [González Abad et al., 2015], we estimate an error in \( \Delta \overline{\Omega}_j \) of less than 3\%.

### 4.3 Results and discussion

Figure 4.1 shows the absolute and relative temperature-corrected changes in HCHO columns from 2005–2009 to 2010–2014. HCHO columns increase by 20–30\% in areas scattered over central southern Canada and the northwestern and central regions of the US, but decrease elsewhere, especially in the southeastern US. We focus here on five regions, marked by circles in Figure 4.1, where HCHO changes significantly and where we can most confidently identify the drivers of these changes: (a) the Houston-Galveston-Brazoria area (HGB; \(-14.6\% \pm 2.7\%\)), (b) Athabasca oil sands in Alberta and Saskatchewan (34.9\% \pm 10.7\%), (c) the Floyd Shale in Alabama (\(-11.8\% \pm 1.7\%\)), (d) the northwestern US (25.4\% \pm 5.9\%), and (e) Illinois and Missouri (19.0\% \pm 4.8\%).
Figure 4.1 Absolute (top) and relative (bottom) mean temperature-corrected change in HCHO columns in May–September from 2005–2009 to 2010–2014 over the North America. Mean change is computed as the average across all temperature bins (see text). Black dots in top panel represent a significant change ($p < 0.05$). Bottom panel shows only statistically significant relative changes. Circled regions represent the areas further discussed in the text: (a) Houston-Galveston-Brazoria (HGB) in Texas, (b) Athabasca oil sands, (c) Floyd Shale in Alabama, (d) northwestern US, and (e) Illinois and Missouri.
We constructed time series of monthly mean HCHO columns spatially averaged over the regions enclosed by circles in Figure 4.1. In the averaging, we considered only those pixels where the HCHO change is statistically significant. To remove the temperature effect, we first determine the exponential dependence of the monthly mean HCHO column on monthly surface temperature \cite{Palmer2006, Zhu2014} in each gridcell over the 2005–2014 May–September time period. We can then normalize the observed HCHO columns in each cell to the mean May–September temperature for that cell. Averaging the resulting columns in each of the five regions yields the temperature-corrected variability of HCHO over the 2005–2014 time period. We report only statistically significant trends, with $p < 0.05$ as the threshold for significance unless otherwise specified.

Figure 4.2 shows the 2005–2014 time series of mean May–September temperature-corrected HCHO columns in the five regions, together with time series of hypothesized explanatory variables. In HGB, HCHO columns decline by $-2.2 \pm 0.7\%$ a$^{-1}$ in a trend that strongly correlates ($r = 0.75$) with the concurrent decrease in anthropogenic VOC emissions of $-5.2\% \pm 0.6\%$ a$^{-1}$ \cite{TCEQ2015}. The HCHO trend is also consistent with measurements from TexAQS aircraft campaigns in 1999 and 2006, which reveal decreases in median concentrations of 56% for ethene and 51% for propene over HGB between those two years \cite{Gilman2009}. Zhu et al. \cite{Zhu2014} previously identified the HGB as the area in the US where HCHO columns show the largest enhancement from anthropogenic VOCs, on top of a large isoprene-driven background that explains why HCHO trends are weaker than anthropogenic VOC trends. We see from the HGB example how HCHO columns from space can help monitor anthropogenic VOC emissions in areas where these are particularly high.
Figure 4.2 Time series of temperature-corrected May–September HCHO columns (black) and the potential explanatory variables driving HCHO variability (colors) from 2005 to 2014 in the five regions circled in Figure 4.1. Observed monthly mean HCHO columns are normalized to the 2005–2014 May–September mean temperature using the local dependence of HCHO on surface temperature. The shading represents 1 standard deviation of the May–September monthly temperature-corrected HCHO columns. Explanatory variables depend on the region. The panels show (a) anthropogenic VOC emissions from point sources in the Houston-Galveston-Brazoria (HGB) area [TCEQ, 2015], (b) bitumen productions in Alberta [AER, 2015], (c) natural gas
production in Alabama [EIA, 2015], (d) mean areal extent of five types of evergreen needleleaf forest in the northwestern US [Friedl et al., 2010], and (e) production of corn silage in Illinois and Missouri from the USDA (http://quickstats.nass.usda.gov). Also shown are correlation coefficients between the temperature-correct HCHO columns and the explanatory variables, as well as linear regression trends if statistically significant ($p < 0.05$).

We next examine two other regions where oil/gas extraction is important and where HCHO columns show significant trends over the 2005–2014 time period. Over the Athabasca oil sands in Alberta, HCHO columns increase by $4.7\% \pm 1.1\% a^{-1}$ (Figure 4.2), concurrent with an increase in bitumen production in this region ($12.4\% \pm 0.8\% a^{-1}, r = 0.82$) [AER, 2015]. Trends in bitumen production have been previously linked to 2005–2011 increases in remotely sensed SO$_2$ and NO$_2$ columns in the same region [McLinden et al., 2012]. In contrast, HCHO columns over the Floyd Shale in Alabama decline by $1.8\% \pm 0.7\% a^{-1}$ from 2005 to 2014 (Figure 4.2), consistent with declining statewide gas production ($-4.4\% \pm 0.4\% a^{-1}, r = 0.63$) [EIA, 2015]. Oil/gas operations can result in elevated HCHO levels through (1) primary emissions at sites where HCHO is used to scavenge H$_2$S from the gas [Amosa et al., 2010; Kenreck, 2014], (2) secondary formation from VOCs [Gilman et al., 2013; Edwards et al., 2014; Stoeckenius, 2015], and (3) flaring [Pikelnaya et al., 2013; Marais et al., 2014]. However, we find no HCHO enhancements in other oil/gas production regions (e.g., Pennsylvania, North Dakota, Oklahoma, and eastern Colorado), suggesting that HCHO emission may depend on local practices.

Over the northwestern US, HCHO columns increase by $4.7\% \pm 1.0\% a^{-1}$ (Figure 4.2), a trend highly correlated with growth in the areal extent of evergreen needleleaf forests in this region ($4.3\% \pm 0.5\% a^{-1}, r = 0.83$). Such forests are an important source of isoprene [Gunther et al., 2012] and thus of HCHO. Here the areal extent of evergreen needleleaf forests is obtained by averaging five types of such forests detected by the Moderate Resolution Imaging Spectrometer (MODIS) MCD12Q1 data [Friedl et al., 2010]. We find another potential link between changing HCHO
columns and land cover in Illinois and Missouri, where the columns correlate moderately with corn silage production \((r = 0.57, p < 0.10, \text{Figure 4.2})\). Elsewhere in the Corn Belt, however, we fail to find significant correlation between HCHO columns and production of either corn silage or grain, even in areas that show significant increases in HCHO over time (e.g., Nebraska, South Dakota, and North Dakota in \text{Figure 4.1}).

In addition to oil/gas operations and vegetation, open fires can also serve as source of HCHO through primary emission or oxidation of NMVOCs in the plume [Alvarado et al., 2010; Marais et al., 2012; Barkley et al., 2013]. We find the decline in HCHO columns over the southwestern California \((-22.5\% \pm 4.7\%, \text{Figure 4.1})\) may be traced to open fires. Seasonal mean HCHO columns in this area over 2005–2014 correlate with HCHO emissions from the Quick Fire Emission Dataset (QFED, \(r = 0.71\)) [Darmenov and da Silva, 2015]. This correlation, however, drops to 0.14 when we remove 2007, a high-fire year, from the timeseries. We fail to find significant correlation between the timeseries of HCHO columns and fire emissions in other areas.

Decreasing NO\(_x\) emissions might also be expected to drive trends in HCHO. Oxidation of VOCs produces organic peroxy radicals that can either react with NO (high-NO\(_x\) pathway) or undergo other fates (low-NO\(_x\) pathways). The high-NO\(_x\) pathway has a higher HCHO yield [Marais et al., 2012; Wolfe et al., 2016; Chan Miller et al., 2016]. In response to air quality regulations, anthropogenic NO\(_x\) emissions in the US have decreased by 4.5% a\(^{-1}\) over 2005–2014 [EPA, 2015]. High- and low-NO\(_x\) pathways are now of comparable importance for isoprene oxidation in the southeastern US [Travis et al., 2016]. We conduct two GEOS-Chem simulations: (1) a control simulation at 2007 May–September intended to represent the middle of the 2005–2009 period, and (2) a sensitivity simulation the same as the control but with a 20% reduction of US anthropogenic NO\(_x\) emissions, intended to represent the middle of the 2010–2014 period. As
expected, HCHO columns generally decline by $0.1-0.7 \times 10^{15}$ molecules cm$^{-2}$ across North America in the sensitivity simulation, with more decrease seen over the southeast US. Over some urban areas with high NO$_x$ emissions, however, HCHO columns slightly increase by $0.1-0.2 \times 10^{15}$ molecules cm$^{-2}$, likely due to an increase in OH as NO$_x$ emissions decrease under NO$_x$-saturated conditions. These responses in HCHO columns to the 20% cut in NO$_x$ emission are small relative the changes in column amount seen by OMI (Figure 4.1). In general, NO$_x$ effect could partly account for the observed decreases over the southeast US, but only has slight ($<10\%$) in other areas.

In summary, we have used OMI satellite observations to diagnose long-term trends in HCHO columns across North America from 2005 to 2014. Controlled for temperature, HCHO increased largely over much of North America, with decreases in the southeastern US and elsewhere from 2005–2009 to 2010–2014. We find significant regional trends that can be related to trends in urban anthropogenic VOC emissions, oil/gas production, land cover, agriculture, and open fires. Except in the southeast US, the impact of declining NO$_x$ emission over the US on HCHO columns is likely small ($<10\%$). Our approach has utility since the effect of temperature on the HCHO precursor isoprene dominates the interannual variability of HCHO column [e.g., Abbot et al., 2003; Palmer et al., 2006; Duncan, et al., 2009]. By controlling for temperature, we can with greater confidence identify trends in HCHO columns over the North America driven by changing emissions from oil/gas operations and other sources.

Acknowledgments

This work was funded by the NASA Earth Science Division as part of the Aura Science Team.
References


Chapter 5 Formaldehyde (HCHO) as a Hazardous Air Pollutant: Mapping surface air concentrations from satellite and inferring cancer risks in the United States

Abstract

Formaldehyde (HCHO) is the most important carcinogen in outdoor air among the 187 hazardous air pollutants (HAPs) identified by the US Environmental Protection Agency (EPA), not including ozone and particulate matter. However, surface observations of HCHO are sparse and the EPA monitoring network could be prone to positive interferences. Here we use 2005–2015 HCHO column data from the OMI satellite instrument, validated with high-quality aircraft data and oversampled on a 5 × 5 km² grid, to map surface air HCHO concentrations across the contiguous US. Results are in good agreement with high-quality observations from urban sites and a factor of 2 lower than data from the EPA network. We estimate that up to 6600–13200 people in the US will develop cancer over their lifetimes by exposure to outdoor HCHO. The main HCHO source in the US is atmospheric oxidation of biogenic isoprene. The HCHO yield from atmospheric oxidation of isoprene decreases as the concentration of nitrogen oxides ($\text{NO}_x \equiv \text{NO} + \text{NO}_2$) decreases, so that NO$_x$ emission controls to improve ozone air quality have a co-benefit in reducing HCHO-related cancer risks.

---

5.1 Introduction

Formaldehyde (HCHO) in outdoor air is a known carcinogen. Exposure to a mean HCHO concentration of 1 μg m$^{-3}$ (about 0.7 ppb at STP) over one’s lifetime will cause up to 13 people in a million to develop lung and nasopharyngeal cancer according to the US Environmental Protection Agency (EPA) [EPA, 2015a]. HCHO is one of 187 hazardous air pollutants (HAPs) identified by the EPA [EPA, 2015a] to cause cancer or other serious health impacts in ambient outside air. It is by far the most important HAP in terms of health risks, accounting for over 50% of the total HAPs-related cancer risks in the US [Strum and Scheffe, 2016; Scheffe et al., 2016]. The second most important HAP is benzene (~ 10%). Unlike most other HAPs, HCHO is not mainly associated with local anthropogenic hotspots but instead is widely present across the US as a product of the oxidation of volatile organic compounds (VOCs) including in particular biogenic isoprene [Palmer et al., 2003]. The HAPs sampling network in the US is limited to urban/industrial sites [Strum and Scheffe, 2016] and thus cannot properly quantify the health risk posed by ambient HCHO. Here we use 11 years of HCHO observations from the OMI satellite instrument [González Abad et al., 2015] with 5×5 km$^2$ spatial resolution enabled by an oversampling technique [Zhu et al., 2014] to map HCHO surface air concentrations over the contiguous US and infer cancer risks on a national scale.

EPA reports HCHO as an ambient air toxic using data from 300–400 sites operated by states, local agencies, and tribes (SLTs network), including ~ 50 national air toxics trends sites (NATTS network). HCHO is collected by 2,4-dinitrophenylhydrazine (DNPH) coated cartridges and then analyzed by high-performance liquid chromatography (HPLC), known as the EPA compendium method TO-11A [EPA, 1999] HCHO measured using this method has potential interferences by ozone [Achatz et al., 1999] and NO$_2$ [Karst et al., 1993; Tang et al., 2004]. High-
quality HCHO measurements in surface air are available only from occasional field campaigns [Lin et al., 2012; Ryerson et al., 2013; Millet et al., 2016]. Satellites provide a continuously operating high spatial resolution data set. HCHO satellite data over the US were recently validated using aircraft observations [Zhu et al., 2016]. Although satellites only measure total HCHO columns (molecules per cm$^2$ of surface), the bulk of that column is in the boundary layer [Martin et al., 2004; Millet et al., 2006; Zhu et al., 2016] and surface concentrations can therefore be inferred.

HCHO columns have been observed continuously from space since GOME (1996–2003) [Chance et al., 2000] and SCIAMACHY (2003–2012) [Wittrock et al., 2006]. Observations are presently available from OMI (2004–) [González Abad et al., 2015; De Smedt et al., 2015], GOME2A (2006–) [De Smedt et al., 2012], OMPS (2011–) [Li et al., 2015; González Abad et al., 2016] and GOME2B (2012–) [De Smedt et al., 2015]. Among those sensors, OMI provides the most suitable data for HCHO mapping due to its daily global coverage, long data record, and fine pixel resolution (13 × 24 km$^2$ at nadir). Its spatial resolution can be further refined by oversampling, as described below.

5.2 Methods

5.2.1 OMI observations

OMI is a UV/Vis nadir solar backscatter spectrometer launched in 2004 on the Aura satellite in a polar sun-synchronous orbit [Levelt et al., 2006]. It observes the whole globe daily at 1330 local time (LT). We use OMI HCHO Version 2.0 (Collection 3) retrievals from the Smithsonian Astrophysical Observatory (OMI-SAO) [González Abad et al., 2015], available at
The data archive extends from 2005 to present. We select data for June–August 2005–2015 that (1) pass all the fitting and statistical quality checks (MainDataQualityFlag = 0), (2) have cloud fraction less than 0.3 and solar zenith angle less than 60°, and (3) are not affected by the instrumental “row anomaly” (http://projects.knmi.nl/omi/research/product/rowanomaly-background). The single-scene precision is 1 × 10^{16} molecules cm^{-2} [González Abad et al., 2015], which corresponds to about 2 ppb in a 2-km deep well-mixed boundary layer. The precision can be improved by multi-scene averaging [De Smedt et al., 2008; Boeke et al., 2011]. We only use the summertime data when HCHO columns are highest and detectable from space. HCHO columns in winter are generally below the detection limit [Abbot et al., 2003].

Zhu et al. [2016] validated the OMI-SA0 product with high-quality HCHO aircraft measurements from the SEAC4RS flight campaign [Toon et al., 2016] over the Southeast US in August–September 2013. Aircraft measurements during SEAC4RS were made in situ from 0.3 to 12 km altitude by two independent laser instruments: CAMS [Richter et al., 2015] and ISAF [Cazorla et al., 2015]. The two instruments were extremely consistent throughout the campaign (r = 0.99) with ISAF 10% higher than CAMS. The horizontal patterns from the satellite retrievals were highly correlated with the aircraft and consistent with a dominant source from biogenic isoprene [Millet et al., 2006]. However, the retrievals were biased low by 37% relative to the CAMS aircraft data. Here we apply a uniform correction factor of 1.59 to the OMI-SA0 retrieval, thus making it unbiased relative to CAMS and 10% lower than ISAF.
5.2.2 Oversampling method

Relating concentrations to population exposure requires the highest spatial resolution possible. Temporal resolution is less critical since the HCHO cancer risk is based on a lifetime-averaged exposure. Here we oversample the OMI HCHO data to increase spatial resolution to $5 \times 5$ km$^2$ through temporal averaging. Oversampling takes advantage of shifting pixel locations and sizes in day-to-day observations [McLinden et al., 2012; Streets et al., 2013] to achieve a spatial resolution finer than pixel size as a temporal average. Oversampling of OMI observations to achieve an effective spatial resolution of a few km has been used previously on urban/regional scales for HCHO [Zhu et al., 2014], SO$_2$ [de Foy et al., 2009; Fioletov et al., 2011; McLinden et al., 2012; Lu et al., 2013] and NO$_2$ [McLinden et al., 2012; Lu et al., 2013]. The common assumption in all these studies has been to view individual satellite observations as uniformly representative of a circle around the pixel center, with the circle radius optimized to balance smoothing and noise. This approach is somewhat arbitrary and computationally demanding.

Here we developed an improved and faster oversampling method enabling application over the entire contiguous US. Consider a satellite pixel $p$ with HCHO column $\Omega(p)$. The overlap area between the pixel $p$ and oversampling grid cell $i$ is $A(p,i)$. Grid cell $i$ collects $N(i)$ overlapping satellite pixel data points over the oversampling period, from which an average column for that grid cell is calculated. We assume that the averaging weight for each individual satellite observation is proportional to the ratio of the overlap area $A(p,i)$ to the pixel area $S(p)$ and inversely proportional to the error standard deviation $\sigma(p)$ of that observation as reported in the OMI-SAO product. $S(p)$ varies by a factor of 10 from ~300 km$^2$ at nadir to ~3000 km$^2$ at the outermost swath-angle [Levelt et al., 2006; de Graaf et al., 2016]. $\sigma(p)$ can vary by a factor of 6 from $0.2 \times 10^{16}$ molecules cm$^{-2}$ for background conditions to $1.2 \times 10^{16}$ molecules cm$^{-2}$ in high-concentration
regions [González Abad et al., 2015]. The area- and error-weighted average column for grid cell $i$ is then derived as:

$$\bar{\Omega}(i) = \frac{\sum_{p=1}^{N(i)} \frac{A(p, i)}{S(p)\sigma(p)} \Omega(p)}{\sum_{p=1}^{N(i)} \frac{A(p, i)}{S(p)\sigma(p)}} \quad (4.1)$$

Besides being computationally fast, this method has the advantage that it fully uses and appropriately weighs the information from all individual satellite observations.

We applied our oversampling method to the OMI observations to produce an 11-year (June–August, 2005–2015) summer average map of HCHO columns with $0.05^\circ \times 0.05^\circ \approx 5 \times 5$ km$^2$ grid resolution. Figure 5.1 shows the result. Values are highest over the Southeast US and are due to oxidation of biogenic isoprene [Palmer et al., 2003; Millet et al., 2008]. The Southeast data were previously validated with the SEAC$^4$RS aircraft observations described above. The HCHO column peaks in the urban areas of Atlanta, Birmingham, and Houston, which could reflect industrial and vehicle sources [Rappenglück et al., 2010; 2013; Boeke et al., 2011; Johansson et al., 2014; Zhu et al., 2014]. However, these peaks are relatively modest on top of the biogenic enhancement. More detailed inspection of the Atlanta maximum (bottom panel of Figure 5.1) suggests a source from ring road traffic. HCHO hotspots in the western US are mostly due to fires as biogenic emissions in that part of the country are much lower than in the eastern US. Satellite retrievals of HCHO columns in fire plumes are highly uncertain because of strong sensitivity to plume rise and to light extinction by the smoke particles [Fu et al., 2007].
5.2.3 Deriving annual mean surface HCHO concentrations

We use the summer mean $0.05^\circ \times 0.05^\circ$ satellite data for HCHO columns in combination with Equation 5.2 to derive the annual mean surface concentrations required for cancer risk assessments:

$$ \bar{C}(i) = \bar{\Omega}(i)\gamma_1(i)\gamma_2(i)\gamma_3(i) \quad (5.2) $$

Here $\bar{C}(i)$ is the annual mean surface air concentration in $0.05^\circ \times 0.05^\circ$ grid cell $i$, $\bar{\Omega}(i)$ is the summer mean oversampled OMI column in that grid cell (Figure 5.1), $\gamma_1(i)$ is the ratio of midday surface air to column concentrations in summer, $\gamma_2(i)$ is the ratio of 24-h average to midday concentrations in summer, and $\gamma_3(i)$ is the ratio of annual to summer mean concentrations. We use the GEOS-Chem chemical transport model to infer $\gamma_1$ and $\gamma_3$, and surface observations to infer $\gamma_2$. GEOS-Chem has been used previously to simulate HCHO over the US including comparisons to satellite and in situ observations [Palmer et al., 2003; Millet et al., 2006; Palmer et al., 2006]. Zhu et al. [2016] and Chan Miller et al. [2016] find that GEOS-Chem provides an unbiased simulation of SEAC4RS and SENEX aircraft observations over the Southeast US in summer, including boundary layer horizontal patterns and mean vertical profiles.

Here we conducted a GEOS-Chem simulation at $2.0^\circ \times 2.5^\circ$ horizontal resolution for the entire year of 2010 using GEOS-5 assimilated meteorological fields [Molod et al., 2012] produced at $0.25^\circ \times 0.3125^\circ$ resolution by the NASA Global Modeling and Assimilation Office (GMAO).
Figure 5.1 Mean OMI HCHO columns over the contiguous US for June–August 2005–2015 with oversampling on a 0.05° × 0.05° (∼5 × 5 km²) grid. The bottom panel zooms in on the ∼100 × 100 km² Atlanta area. The white circle indicates the Atlanta city center.

In order to convert HCHO columns to surface concentrations (scaling factor $\gamma_1$), we sample daily surface HCHO concentrations and total columns from the June–August model output at the
OMI overpass time (1300–1400 local time; midday here and elsewhere). Figure 5.2 shows the resulting summer midday mean surface concentrations. The spatial patterns in the OMI data are retained because the GEOS-Chem scaling factors are fairly uniform. Figure 5.3 compares this product with local measurements from research field campaigns at several urban sites [Dasgupta et al., 2005; Lin et al., 2012] and at Houston (B. Rappenglück, unpublished data). There is good agreement with no significant bias averaged across all sites (−2.8 ± 18%).

**Figure 5.2** OMI-derived summer midday HCHO concentrations in surface air. Values are 2005–2015 averages for June–August at 1300–1400 local time.
Figure 5.3 Summer mean midday HCHO concentrations at urban sites. OMI-derived values for 2005–2015 are compared to local measurements in different years. Measurements for Houston are from the Moody Tower in August 2006 and 2010 (B. Rappenglück, unpublished data). Measurements for New York City are from Lin et al. [2012]. Measurements for Atlanta, Philadelphia, and Nashville are from Dasgupta et al. [2005]. Error bars represent ± 1 standard deviation in the measurements.

To convert midday to 24-h averaged surface air HCHO concentrations (scaling factor $\gamma_2$), we use ground-site HCHO measurements from three field campaigns including (1) CalNex (May–June 2010, Pasadena, California) (2) SOAS (June–July 2013, Brent, Alabama) and (3) SLAQRS (August–September 2013, East St. Louis, Illinois) [Ryerson et al., 2013; Millet et al., 2016; Sareen et al., 2016]. Figure 5.4 shows the diurnal variations in HCHO concentrations measured at those three sites. HCHO is depleted during the night because of dry deposition [Sumner et al., 2001]. The 24-h average to midday value ratio is 0.64 at the California site, 0.77 at the Alabama site, and 0.79 at the Illinois site. It is not clear that these differences reflect geographical specificity, therefore we apply a single scaling factor $\gamma_2 = 0.73$ throughout the US to convert midday to 24-h average summer concentrations.
Figure 5.4 Diurnal variation of summertime HCHO concentrations in surface air. Observations are from three field campaigns including CalNex (May–June 2010) [Ryerson et al., 2013] in Pasadena, California; SOAS (June–July 2013) [Sareen et al., 2016] in Brent, Alabama; and SLAQRS (August–September 2013) [Millet et al., 2016] in East St. Louis, Illinois. Error bars are standard deviations in the hourly averaged data. Red lines are 24-h averages.

Finally, we use GEOS-Chem to convert these summertime 24-h averages to annual 24-h averages (scaling factor $\gamma_3$; Figure 5.5). Surface HCHO concentrations in GEOS-Chem have strong seasonal variations driven mostly by biogenic emissions, with annual to summertime average ratios of 0.4–0.5 in the Southeast US and 0.6–0.7 in the West. Fires also contribute locally to seasonality in the West but the resulting inferences are highly uncertain as noted above. Summertime HCHO values as given in Figure 5.2 contribute 60% of the annual total in the
Southeast and 40% in the West. Comparison of GEOS-Chem to HCHO aircraft observations from the WINTER campaign [Jaeglé et al., 2015] over the Northeast US in February–March 2015 indicates a 39% low bias in the model, but the winter season contributes only 5% of the annual mean concentration in the eastern US and 15% in the west, so the impact of this underestimate would be slight.

![Annual-to-summer HCHO ratios](image)

**Figure 5.5** Ratios of annual to summer (JJA) mean HCHO concentrations in surface air. Values are from the GEOS-Chem model.

### 5.2.4 Estimating cancer risks

EPA uses the inhalation unit risk estimate (URE) to quantify the cancer risks of HCHO and other HAPs [EPA, 2015a]. The URE represents the upper bound for the increased cancer risk from
inhalation exposure to an air concentration of 1 µg m\(^{-3}\) over an individual’s lifetime. Based on the upper confidence limit of the fitted dose-response curve, the inhalation URE for HCHO is estimated to be 1.3 × 10\(^{-5}\) (µg m\(^{-3}\))\(^{-1}\) by the Agency’s Integrated Risk Information System (IRIS) [EP\(A\), 2015a; Strum and Scheffe, 2016]. This means that individuals exposed to a mean HCHO concentration of 1 µg m\(^{-3}\) (about 0.7 ppb) have a chance of up to 13 in a million to develop cancer over their lifetime from this HCHO exposure. Risk is assumed to increase linearly with HCHO concentration.

5.3 Results and discussions

5.3.1 Annual mean surface HCHO concentrations

Figure 5.6 shows the annual mean HCHO concentrations in surface air across the US as derived from Equation 5.2. The patterns largely reflect those of the original OMI column data in Figure 5.1.

Figure 5.7 compares these values with 2005–2015 annual averages from the EPA SLTs network. SLTs sites report 24-h average concentrations every 6 days. Annual averages are computed here for sites with full yearly coverage (at least 12 samples per quarter) for at least 9 years of the 2005–2015 period. 50 SLTs sites meet the above criteria and their locations are shown in the top panel of Figure 5.7.
Figure 5.5 OMI-derived annual mean HCHO concentrations in surface air and associated cancer risks. Cancer risk is related to surface HCHO concentration using the EPA inhalation unit risk estimate (URE) of $1.3 \times 10^{-5} \text{ (µg m}^{-3})^{-1}$. We find from Figure 5.7 that OMI-derived surface HCHO is a factor of 2 lower than the EPA SLTs data, despite being unbiased relative to high-quality urban summer measurements (Figure 5.3). The SLTs HCHO data could have positive interferences, as noted in the Introduction. On the other hand, GEOS-Chem may underestimate HCHO during winter. There is moderate correlation between the OMI-derived and SLTs data ($r = 0.56$). Here we will view the factor of 2 difference as representing a range of uncertainty for annual mean surface HCHO concentrations and hence cancer risks.
Figure 5.6 Annual mean 2005–2015 HCHO concentrations from EPA surface network (SLTs) and comparison to OMI-derived surface air concentrations. Top panel shows the network site locations and the annual mean data. Only sites with at least 9 years of complete data for the 2005–2015 period are shown (see text for details). Bottom panel compares the OMI-derived and EPA data for individual sites. Slopes (95% confidence interval) of reduced major axis (RMA) regressions are shown along with the normalized mean bias (NMB), sample size (N) and correlation coefficient (r).
5.3.2 National cancer risks from outdoor HCHO exposure

Figure 5.6 shows the distribution of cancer risks in the US inferred from the OMI-derived mean surface HCHO concentrations, based on a conversion factor of 1.23 µg m⁻³ ppb⁻¹ at 298 K and 1 atm. We estimate the total national cancer risk from HCHO exposure by convolving the OMI-derived distribution of cancer risks in Figure 5.6 with gridded 0.05° × 0.05° population data for 2015 [Gridded Population of the World, 2015]. We infer in this manner that 6600 people in the US will develop cancer at some point in their lives due to exposure to outdoor HCHO. As pointed out above, OMI-derived HCHO concentrations are a factor of 2 lower than measured at the EPA SLTs sites. Thus we increase the OMI-derived cancer risk by a factor of 2 to get an upper estimate. The resulting upper bound number of lifelong cancer occurrences due to exposure to outdoor HCHO over the US is in the range 6600–13200, i.e., one person in 2–4 × 10⁵ based on a US population of 319 million. This cancer risk is well above the 1 in 10⁶ level generally considered as the threshold of tolerable risk [EPA, 1994; European Chemicals Agency, 2012]. Based on a US life expectancy of 78.7 years (2012), it translates into 84–168 cancer cases per year in the US caused by exposure to outdoor HCHO.

5.3.3 Co-benefit of NOₓ emission controls for reducing HCHO-related cancer risks

HCHO in the US originates mainly from the oxidation of biogenic isoprene emitted by vegetation, and thus would seem largely uncontrollable. However, the HCHO yield from isoprene oxidation is higher in the presence of nitrogen oxide radicals (NOₓ) [Marais et al., 2012; Wolfe et al., 2016] and NOₓ in the US is mainly anthropogenic. We conducted a sensitivity GEOS-Chem simulation with no anthropogenic NOₓ emissions and found that HCHO annual mean surface concentrations in the US decrease by 10–30%. Convolving this decrease with the US population
map we find that cancer risks from outdoor HCHO decrease by 20%. Thus the increment in HCHO cancer risks due to anthropogenic NO$_x$ is one person in $4\times 8 \times 10^6$, which by itself is above the tolerable threshold of 1 in $10^6$. Anthropogenic NO$_x$ emissions in the US have decreased by 2.1% a$^{-1}$ over 1991–2013 in response to regulations to improve ozone air quality [EPA, 2015b]. Such efforts to mitigate NO$_x$ emissions thus have a significant co-benefit in also reducing HCHO-related cancer risks.

Acknowledgement

This work was supported by the NASA Earth Science Division. Funding for the SLAQRS measurements was provided by NSF (Grant #1148951). We thank Noelle Selin for valuable discussions on HCHO cancer risks.
References


Chan Miller, C., et al. (2011), Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1042.


Supplemental Materials

Abbreviations and acronyms used in Chapter 2

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMF</td>
<td>Air Mass Factor</td>
</tr>
<tr>
<td>BIRA</td>
<td>Belgian Institute for Space Aeronomy</td>
</tr>
<tr>
<td>CAMS</td>
<td>Compact Atmospheric Multispecies Spectrometer</td>
</tr>
<tr>
<td>CTM</td>
<td>Chemical Transport Model</td>
</tr>
<tr>
<td>GEOS-FP</td>
<td>Goddard Earth Observing System-Forward Processing</td>
</tr>
<tr>
<td>GMAO</td>
<td>Global Modeling and Assimilation Office</td>
</tr>
<tr>
<td>GMI</td>
<td>Global Modeling Initiative</td>
</tr>
<tr>
<td>GOME2</td>
<td>Global Ozone Monitoring Experiment-2</td>
</tr>
<tr>
<td>IMAGES</td>
<td>Intermediate Model of Global Evolution of Species</td>
</tr>
<tr>
<td>ISAF</td>
<td>In Situ Airborne Formaldehyde</td>
</tr>
<tr>
<td>MEGAN</td>
<td>Model of Emissions of Gases and Aerosols from Nature</td>
</tr>
<tr>
<td>OMI</td>
<td>Ozone Monitoring Instrument</td>
</tr>
<tr>
<td>OMPS</td>
<td>Ozone Mapping and Profiler Suite</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>RMA</td>
<td>Reduced Major Axis</td>
</tr>
<tr>
<td>SAO</td>
<td>(Harvard) Smithsonian Astrophysical Observatory</td>
</tr>
<tr>
<td>SCD</td>
<td>Slant Column Density</td>
</tr>
<tr>
<td>SCIAMACHY</td>
<td>Scanning Imaging Absorption spectrometer for Atmospheric Chartography</td>
</tr>
<tr>
<td>SEAC^4RS</td>
<td>Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys</td>
</tr>
<tr>
<td>VCD</td>
<td>Vertical Column Density</td>
</tr>
<tr>
<td>VOCs</td>
<td>Volatile Organic Compounds</td>
</tr>
</tbody>
</table>
Table S1 Summary of different HCHO retrievals validated and intercompared in this work

<table>
<thead>
<tr>
<th></th>
<th>OMI-SA0 (V003)</th>
<th>OMI-BIRA</th>
<th>GOME2A-BIRA (V14)</th>
<th>GOME2B-BIRA (V14)</th>
<th>OMPS-SA0</th>
<th>OMPS-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitting window (nm)</td>
<td>328.5–356.5</td>
<td>328.5–346.0</td>
<td>328.5–346.0</td>
<td>328.5–346.0</td>
<td>327.7–356.0</td>
<td>328.5–356.5</td>
</tr>
<tr>
<td>Spectral resolution (nm, at 340 nm)</td>
<td>0.42</td>
<td>0.42</td>
<td>0.27</td>
<td>0.28</td>
<td>~1</td>
<td>~1</td>
</tr>
<tr>
<td>Fitted species</td>
<td>O₃, NO₂, BrO, O₂⁺O₂, Ring</td>
<td>O₃, NO₂, BrO, O₂⁺O₂, Ring</td>
<td>O₃, NO₂, BrO, O₂⁺O₂, Ring</td>
<td>O₃, NO₂, BrO, O₂⁺O₂, Ring</td>
<td>O₃, NO₂, BrO, O₂⁺O₂, Ring</td>
<td>BrO (O₃, Ring, NO₂ features are from principal components)</td>
</tr>
<tr>
<td>HCHO cross-section</td>
<td>Chance and Orphal [2011], 300K</td>
<td>Meller and Moortgat [2000], 298K</td>
<td>Meller and Moortgat [2000], 298K</td>
<td>Meller and Moortgat [2000], 298K</td>
<td>Chance and Orphal [2011], 300K</td>
<td>-</td>
</tr>
<tr>
<td>Closure polynomial</td>
<td>3rd order</td>
<td>5th order</td>
<td>5th order</td>
<td>5th order</td>
<td>3rd order</td>
<td>-</td>
</tr>
<tr>
<td>Underdamping correction</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>Additional fit parameters</td>
<td>Residual common mode spectrum</td>
<td>Non-linear O₃ absorption effect</td>
<td>Non-linear O₃ absorption effect; Polarisation cross-section</td>
<td>Non-linear O₃ absorption effect; Polarisation cross-section</td>
<td>Residual common mode spectrum</td>
<td>Principal components derived from radiances</td>
</tr>
<tr>
<td>Vertical layers of the CTM</td>
<td>47</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>47</td>
<td>72</td>
</tr>
<tr>
<td>Horizontal resolutions of the CTM</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
</tr>
<tr>
<td>A prior HCHO profiles</td>
<td>1300 to 1400 Monthly average in 2007</td>
<td>Sampled at the satellite overpass times</td>
<td>Sampled at the satellite overpass times</td>
<td>Sampled at the satellite overpass times</td>
<td>1300 to 1400 Monthly average in 2007</td>
<td>Monthly average, 2005–2007</td>
</tr>
<tr>
<td>Resolution of scattering weight grids</td>
<td>2° × 2.5°</td>
<td>Interpolated for each satellite pixel (surface albedo, surface)</td>
<td>Interpolated for each satellite pixel (surface albedo, surface)</td>
<td>Interpolated for each satellite pixel (surface albedo, surface)</td>
<td>2° × 2.5°</td>
<td>2° × 2.5°</td>
</tr>
<tr>
<td></td>
<td>altitude, observation angles</td>
<td>altitude, observation angles</td>
<td>altitude, observation angles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrain corrections</td>
<td>Yes</td>
<td>Yes, based on Zhou et al. [2009]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment of aerosols</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reference sector correction</td>
<td>Pixel-by-pixel correction</td>
<td>Modelled by a polynomial</td>
<td>Modelled by a polynomial</td>
<td>Modelled by a polynomial</td>
<td>Pixel-by-pixel correction</td>
<td>Model climatology</td>
</tr>
</tbody>
</table>

*SAO retrievals applied the “pixel-by-pixel correction” method, which accounts for the difference between retrieved and GEOS-Chem modelled slant columns over the remote Pacific Ocean. For BIRA retrievals, the reference sector correction refers to the latitudinal dependency (modelled by a polynomial) of the HCHO slant columns in the reference sector. For OMPS PCA retrievals, the latitude-dependent GMI monthly climatology HCHO column amount over the remote Pacific is added to all pixels, without pixel-specific correction.
Figure S1. GEOS-Chem based air mass factors (AMF) over the Southeast US averaged over the SEAC4RS period (5 August–25 September 2013). Each panel shows the AMF computed using GEOS-Chem HCHO shape factors with scattering weights from the six retrievals (Table 2.1). For each retrieval, GEOS-Chem shape factors are sampled under its schedule, and filtered by its quality flags and cloud conditions.
References


