Technical Guidance Document

NASA Health and Air Quality Applied Sciences Team
2017–18 Tiger Team Project

Supporting the use of satellite data in State Implementation Plans

Comparison of CMAQ Simulation to Satellite Observations: NO_2 Column versus OMI NO_2

Prepared by:

Momei Qin and Talat Odman
School of Civil & Environmental Engineering
Georgia Institute of Technology
311 Ferst Drive, Atlanta, GA 30332-0512

August 28, 2018

PI:	**Co-PI:**
Ted Russell ted.russell@gatech.edu Phone: 404-894-3079 Fax: 404-894-8266	Talat Odman odman@gatech.edu Phone: 404-894-2783 Fax: 404-894-8266

This technical guidance document is a product of a 2017-2018 NASA Health and Air Quality Applied Sciences Team (HAQAST; www.haqast.org) Tiger Team project, “Supporting the use of satellite data in State Implementation Plans (SIPs)”. Team membership is listed below. We are grateful to colleagues who shared ideas and opinions that helped shape this document, and to NASA for hosting these products on the Air Quality From Space website at https://airquality.gsfc.nasa.gov/aq-managers.

HAQAST Tiger Team Participants

Grace Choi (gracesh27@gmail.com), University of Wisconsin—Madison
Bryan Duncan (bryan.n.duncan@nasa.gov), NASA GSFC
Arlene Fiore (amfiore@ldeo.columbia.edu), LDEO/Columbia University
Daven Henze (Daven.Henze@colorado.edu), University of Colorado – Boulder
Jeremy Hess (jjhess@uw.edu), University of Washington
Tracey Holloway (taholloway@wisc.edu), University of Wisconsin—Madison
Xiaomeng Jin (xjin@ldeo.columbia.edu), LDEO/Columbia University
Lok Lamsal (lok.lamsal@nasa.gov), NASA GSFC
George Milly (gpm2109@columbia.edu), LDEO/Columbia University
Jessica Neu (Jessica.L.Neu@jpl.nasa.gov), NASA JPL
Talat Odman (talat.odman@ce.gatech.edu), Georgia Institute of Technology
Jonathan Patz (patz@wisc.edu), University of Wisconsin—Madison
Ted Russell (ted.russell@gatech.edu), Georgia Institute of Technology
Daniel Tong (qtong@gmu.edu), George Mason University
Mark Zondlo (mzondlo@princeton.edu), Princeton University

Stakeholder Participants

Jim Boylan (James.boylan@dnr.ga.gov), GA EPD
Pat Dolwick (Dolwick.Pat@epa.gov), U.S. EPA
Mark Estes (mark.estes@tceq.texas.gov), TCEQ
Michael Geigert (Michael.Geigert@ct.gov), CT DEEP
Barron Henderson (henderson.barron@epa.gov), U.S. EPA
Kurt Kebschull (Kurt.Kebschull@ct.gov), CT DEEP
Sang-Mi Lee (slee@aqmd.gov), SCAQMD
Julie McDill (jmcdill@marama.org), MARAMA
Paul Miller (pmiller@nescaum.org), NESCAUM
Stephen Reid (sreid@baaqmd.gov), BAAQMD
Steve Soong (SSoong@baaqmd.gov), BAAQMD
Saffet Tanrikulu (STanrikulu@baaqmd.gov), BAAQMD
Di Tian (Di.Tian@dnr.ga.gov), GA EPD
Gail Tonnesen (Tonnesen.Gail@epa.gov), U.S. EPA
Luke Valin (Valin.Lukas@epa.gov), U.S. EPA
Susan Wierman (swierman@marama.org), MARAMA
Tao Zeng (Tao.Zeng@dnr.ga.gov), GA EPD

How to cite this technical guidance document:

Qin, M., T. Odman, T. Russell et al. (2018), Comparison of CMAQ Simulation to Satellite Observations: NO_2 Column versus OMI NO_2, HAQAST Tech. Guid. Doc. No. 4, doi: 10.7916/d8-cfwm-5x30, available at https://doi.org/10.7916/d8-cfwm-5x30.

DISCLAIMERS

The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official position, policy, or decision by any of the authors' institutions or agencies, nor may they be used for advertising or product endorsement. The guidance provided in this document was written with the intent to encourage "best practices" in using satellite data for air quality applications. Furthermore, neither the U.S. government nor any of the authors' agencies endorse or recommend any commercial products, processes, or services, including Web sites, maintained by public and/or private organizations, to which this document provides links. When users follow a link to an outside Web site, they are subject to the privacy and security policies of the owners/sponsors of the outside Web site(s). Neither the authors nor their agencies or institutions are responsible for the information collection practices of these sites. Reference to or appearance of any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise, in these materials does not constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any of the authors' institutions. Neither the authors nor their employers assume any liability resulting from the use of these materials.

Introduction

Nitrogen dioxide (NO_2), a criteria air pollutant, is routinely monitored at major cities throughout the U.S.; however, in the remaining areas, it is measured only at a very limited number of sites. NO_2 measurements are mostly limited to ground level concentrations, and NO_2 concentrations aloft are largely unknown. There are rare exceptions such as intensive field studies (e.g., Discover AQ[1]) during which aircraft measurements have been carried out. When we evaluate CMAQ simulations, we typically compare modeled NO_2 with ground-based observations. Data for several pollutants including NO_2 are available from the EPA Air Quality System (AQS) website[2].

Compared to the sparse ground-based monitoring network, satellite observations have the advantage of coverage in unmonitored areas, particularly where monitoring is impossible such as over water and desert. For instance, the Ozone Monitoring Instrument (OMI), which is on board of NASA's EOS Aura satellite following a sun-synchronous obit, is able to provide NO_2 distributions at approximately 13:45 local time, with a global coverage and pixel size varying from 13km×26km near nadir (center of the swath) to 40km×250km at the swath edge. However, OMI provides the retrieval of the tropospheric NO_2 column instead of the concentration at certain altitude, with uncertainties tied to the slant column densities, the air mass factor, and the separation of the stratosphere and troposphere (Duncan et al., 2018).Since 25 June 2007, OMI has been affected by a number of so-called row anomalies, which resulted in low-quality radiance data for particular viewing directions of OMI[3]. On 1 January 2011, 29 of the 60 rows on the charge-coupled device (CCD) detectors were (partially) affected, with affected field-of-views discarded in the calculation of vertical column densities (VCDs) for the operational NO_2 retrieval, leading to the OMI NO_2 measurements covering the globe every two days (instead of every day). One should note that satellite observations from space measure full tropospheric columns rather than ground-level concentrations. Therefore, satellite retrievals are impacted less by some factors that affect surface measurement such as the planetary boundary layer (PBL) height. When discrepancies exist between ground-level measurements and simulations, it is difficult to judge whether uncertainties in meteorology, emissions or chemistry are the main cause for the discrepancies. Comparisons between model simulations and satellite retrievals can be more useful in evaluating model performance as some of the factors are ruled out.

This document describes a procedure for comparing CMAQ-simulated NO_2 columns to OMI retrievals and contains an example application in the Great Lakes Region. The NO_2 profile shapes from CMAQ are used to derive new tropospheric vertical column densities (VCD) of NO_2 for comparison, instead of operational retrievals of NO_2 data from the NASA OMNO2 v3.0 (OMNO2) product[4], which apply NO_2 profile shapes from the Global Model Initiative (GMI) at a coarser resolution (1.25° × 1°). Using profile shapes from the CMAQ model has been tested in a few studies and shown to result in better agreement of NO_2 satellite retrievals with measurements during the DISCOVER-AQ Maryland field campaign as it does not only ensure self-consistency but also leads to more accurate retrievals through improved spatial representation of NO_2 shape factors in the air mass factor (AMF; converting slant column to VCD) calculation (Lamsal et al., 2014; Travis et al., 2016; Goldberg et al., 2017). Tropospheric slant column density (SCD) NO_2 data

---

[1] https://discover-aq.larc.nasa.gov/data.html
[2] https://www.epa.gov/outdoor-air-quality-data
[3] http://projects.knmi.nl/omi/research/product/rowanomaly-background.php
[4] https://acdisc.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level3/OMNO2d.003/doc/README.OMNO2.pdf

from OMNO2 is used for the VCD calculation. The algorithm detailed in this document follows Goldberg et al. (2017). An implementation of the algorithm in **R** is appended to the document. In what follows, we offer instructions on how to download the data, use the code attached to recalculate NO_2 VCD, regrid the data with WHIPS (the Wisconsin Horizontal Interpolation Program for Satellites), and to perform the comparison.

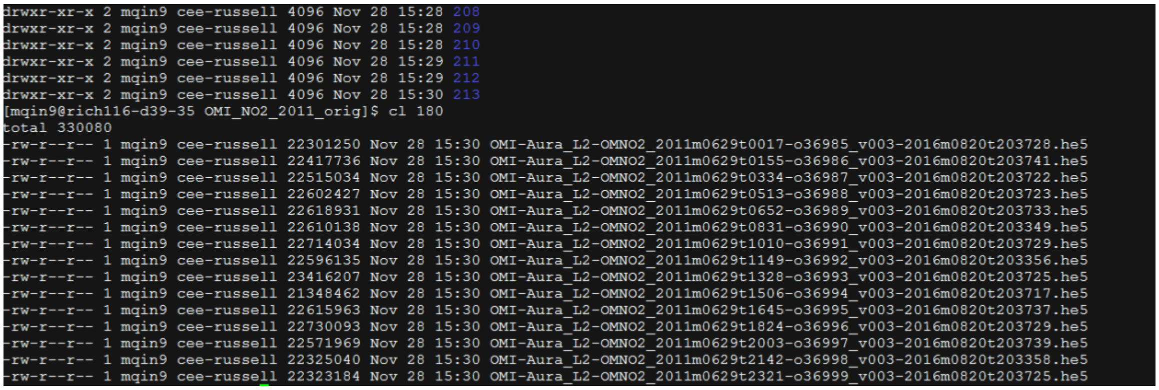
Procedure

1. **Download the OMNO2 product and auxiliary corner files**

   The OMNO2 data product contains slant column NO_2, total NO_2 VCD, stratospheric and tropospheric VCDs, scattering weights, cloud radiative fraction and optical centroid pressure, and other ancillary data. The files are available from the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) data archive website[5]. There, you can also find instructions on how to download the data in bulk with the scripting method.

   Corner files (the OMPIXCOR product) contain ground locations of the OMI pixel corners in the global scanning mode. They are required for regridding of the OMNO2 data. The OMI corner files are available from a different page of the GES DISC data archive website[6].

   Both the OMNO2 product and corner files are stored in the version 5 of Earth Observing System Hierarchical Data Format (HDF-EOS5). Each file contains data from the day lit portion of an orbit (~53 minutes). There are approximately 14 orbits per day, with the data stored in approximately 14 files (Figure 1).



```
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:28 208
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:28 209
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:28 210
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:29 211
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:29 212
drwxr-xr-x 2 mqin9 cee-russell 4096 Nov 28 15:30 213
[mqin9@rich116-d39-35 OMI_NO2_2011_orig]$ cl 180
total 330080
-rw-r--r-- 1 mqin9 cee-russell 22301250 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0017-o36985_v003-2016m0820t203728.he5
-rw-r--r-- 1 mqin9 cee-russell 22417736 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0155-o36986_v003-2016m0820t203741.he5
-rw-r--r-- 1 mqin9 cee-russell 22515034 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0334-o36987_v003-2016m0820t203722.he5
-rw-r--r-- 1 mqin9 cee-russell 22602427 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0513-o36988_v003-2016m0820t203723.he5
-rw-r--r-- 1 mqin9 cee-russell 22618931 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0652-o36989_v003-2016m0820t203733.he5
-rw-r--r-- 1 mqin9 cee-russell 22610138 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t0831-o36990_v003-2016m0820t203349.he5
-rw-r--r-- 1 mqin9 cee-russell 22714034 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1010-o36991_v003-2016m0820t203729.he5
-rw-r--r-- 1 mqin9 cee-russell 22596135 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1149-o36992_v003-2016m0820t203356.he5
-rw-r--r-- 1 mqin9 cee-russell 23416207 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1328-o36993_v003-2016m0820t203725.he5
-rw-r--r-- 1 mqin9 cee-russell 21348462 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1506-o36994_v003-2016m0820t203717.he5
-rw-r--r-- 1 mqin9 cee-russell 22615963 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1645-o36995_v003-2016m0820t203737.he5
-rw-r--r-- 1 mqin9 cee-russell 22730093 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t1824-o36996_v003-2016m0820t203729.he5
-rw-r--r-- 1 mqin9 cee-russell 22571969 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t2003-o36997_v003-2016m0820t203739.he5
-rw-r--r-- 1 mqin9 cee-russell 22325040 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t2142-o36998_v003-2016m0820t203358.he5
-rw-r--r-- 1 mqin9 cee-russell 22323184 Nov 28 15:30 OMI-Aura_L2-OMNO2_2011m0629t2321-o36999_v003-2016m0820t203725.he5
```

Figure 1. The OMNO2 files for 29 June 2011.

2. **Re-calculate OMI NO_2 tropospheric VCD (VCD_{trop}) using the NO_2 vertical profile from CMAQ based on the method in Goldberg et al. (2017)**

   a) Calculate tropospheric air mass factor (AMF_{trop}) for each CMAQ grid cell as follows:

---

[5] https://aura.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level2/OMNO2.003/

[6] https://aura.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level2/OMPIXCOR.003/

$$AMF_{trop} = \frac{\sum_{surface}^{tropopause} SW \times \chi_a}{\sum_{surface}^{tropopause} \chi_a}. \tag{1}$$

Here SW is the scattering weight that represents optical atmospheric/surface properties and is provided by OMNO2, and χ_a is the partial column NO_2 provided by CMAQ.

SW values, which are altitude-dependent, are available at 35 fixed pressure levels (refer to the variable "SWPRES" in OMNO2). SW at a CMAQ grid cell is obtained from the SW values for the OMI pixel that overlaps it, using linear interpolation in the vertical direction (with pressure as the vertical coordinate).

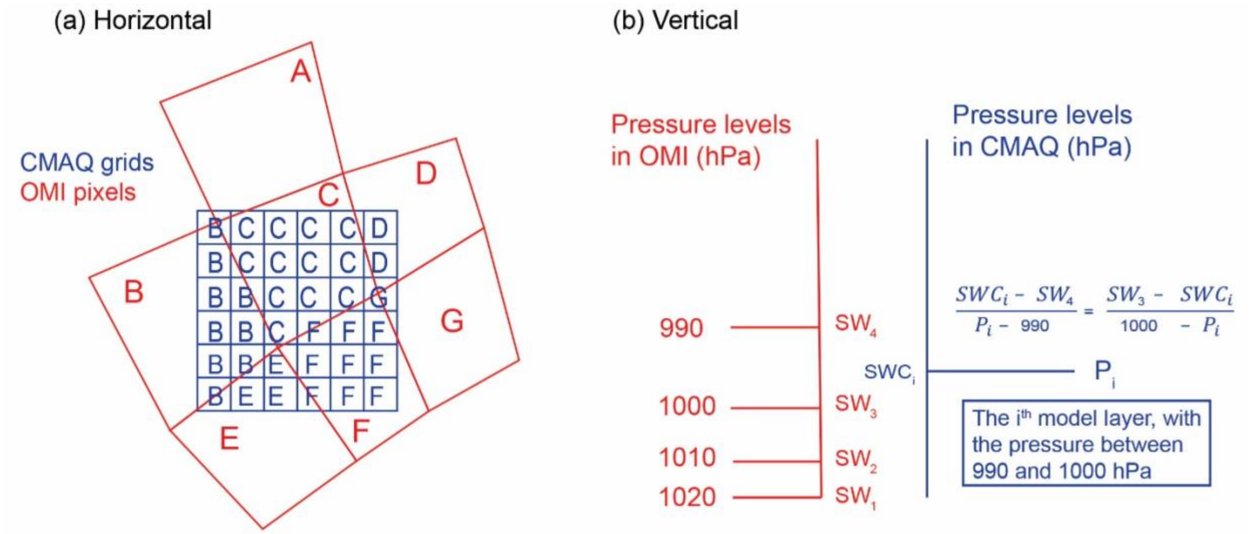


Figure 2. In the **R** code appended, we use the function **over** to assign each CMAQ grid the *SW* values of the OMI pixel where the CMAQ grid center sits as shown in Figure (a). The function **approx** is used for linear interpolation of *SW* values in the vertical direction. An example is given to illustrate how the *SW* value at a CMAQ model layer C_i with pressure between 990 and 1000 hPa is derived in Figure (b).

χ_a is computed as follows:

$$\chi_a = \rho_{air} \times \frac{N_A}{MW_{air}} \times Z \times C_{NO_2} \times 10^{-6}. \tag{2}$$

Here ρ_{air} is air density in $g \cdot cm^{-3}$, N_A is Avogadro's number in mol^{-1} (6.0221367×10^{23}), MW_{air} is mean molecular weight for dry air in $g \cdot mol^{-1}$ (28.9628), Z is the full layer height in cm and C_{NO_2} is NO_2 volume mixing ratio in ppm. All the above variables can be obtained from CMAQ inputs/outputs with some unit conversions. Additionally, averaged NO_2 concentration from 13 to 15 EST is applied in the calculation (OMI has an Equator overpass time at 13:45 local time). For simplicity, we computed the integral (i.e., sum) from the surface layer to Layer 31 of CMAQ (the pressure at the top of Layer 31 is ~200 hPa in our simulation) instead of surface to tropopause.

b) Average all interim air mass factors within an OMI pixel to generate a single AMF_{trop} for each OMI pixel.

c) Calculate VCD_{trop} for each OMI pixel as follows:

$$VCD_{trop} = \frac{SCD_{total} - SCD_{strat}}{AMF_{trop}}. \tag{3}$$

Here SCD_{total} and SCD_{strat} are total slant column density and stratospheric slant column density, respectively, which are both provided by OMNO2. We only calculate VCD_{trop} for valid pixels. The criteria for a "valid" pixel include solar zenith angle $< 80°$, cloud radiance fraction < 0.5 and surface albedo < 0.3. We also remove the five largest pixels at the swath edge or any pixel with a "VcdQualityFlag" of an odd integer, indicating bad data.

3. **Regrid derived VCD_{trop} for OMI pixels into CMAQ grids using the Wisconsin Horizontal Interpolation Program for Satellites (WHIPS) version 2**[7,8]
   a) Install dependencies including Python version 2.X (WHIPS is not compatible with Python 3.X), HDF4, HDF5, netCDF, and the GEOS framework.
   b) Install Anaconda (Python version 2.7) instead of installing multiple python modules to spare the troubles with installation due to the environment. Follow the guide for installation on the Conda webpage[9]. For example, if you are a Linux user, edit ~/.bash_profile to add the Anaconda directory to your PATH environment variable using the following line:

   ```
   export PATH=/…/…/anaconda2/bin:$PATH
   ```

   Then activate the default environment *root*.

   ```
   > source activate root
   ```

   c) Install packages that WHIPS depends upon, including numpy, shapely, numexpr, pyproj, cython, h5py, netcdf4, pytables and pyhdf using the following commands:

   ```
   > conda install numpy
   ```

   ```
   > conda install -c conda-forge pyhdf (pyhdf is not available from the Anaconda repository but available from a different repository developed by conda-forge)
   ```

   The packages installed are put into the environment root by default.
   d) Install easy_install

   ```
   > pip install easy_install
   ```

   e) Install WHIPS using the easy_install command

   ```
   > easy_install whips2
   ```

   f) Run WHIPS
   Prepare a text file to specify all the options and locations of input files that WHIPS needs. Please refer to the example of the text file given in the instruction of WHIPS. The input files of WHIPS include new OMNO2 data and auxiliary corner files. The new OMNO2 data is almost the same as the operational retrieval (OMNO2 product), with the only change made to VCD_{trop}, which has been replaced with the recalculated VCD_{trop} using NO_2 vertical profiles from CMAQ. Then, the command line call consists only of specifying the location of the text file WHIPS should read from.

   ```
   > Whips.py –inFromFile /…/…/input.txt
   ```

   Finally, you can get daily gridded OMI NO_2 retrievals written in netCDF format with the grids desired and specified in the text file.

---

[7] https://nelson.wisc.edu/sage/data-and-models/software.php
[8] https://nelson.wisc.edu/sage/docs/WHIPS_User_Guide_V2.pdf
[9] https://conda.io/docs/user-guide/getting-started.html

If you see error messages like *'module' object have no attribute 'isHDF5Fille'* when you run WHIPS, revise the code of parse_geo.py (/anaconda2/lib/python2.7/site-packages/WHIPS2-2.0.0-py2.7.egg/process_sat) using the following command:

```
> Pt2to3 parse_geo.py > parse_geo.py.new
```

WHIPS is not compatible with pytables 3.This issue happens when pytable3 is installed but functions in pytable2 are used in the source code.

4. **Compute CMAQ tropospheric column NO_2 using the function *vertintegral* in I/O API**[10]

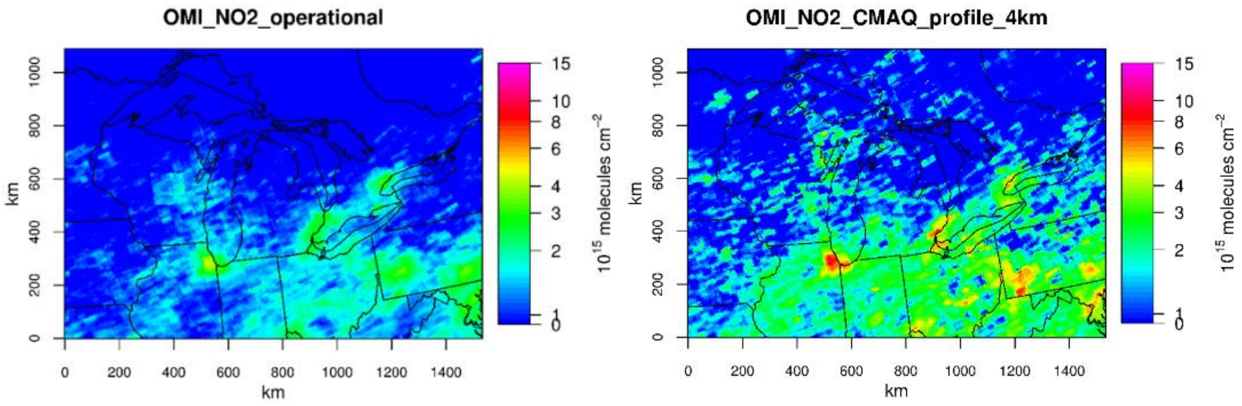
Before running *vertintegral*, set environment variables as follows:

```
> csh  # change to C shell if bash is the login shell
> setenv INFILE /…/the input file that must contain 3-D ozone in a unit of ppm # CMAQ outputs
> setenv METCRO3D /…/the input file that must contain 3-D full-layer height and air density # MCIP outputs
> setenv OUTFILE /…/the output file
> setenv LAY_LO 1  # Lower bound of integration-interval (Layer 1 in this case)
> setenv LAY_HI 31 # Upper bound of integration-interval (Layer 31, in line with the definition of tropopause in satellite data processing)
```

Then, call *vertintegral* and type in the contents (e.g. names of input and output files) that *vertintegral* prompts.

Example Comparison

Error! Reference source not found. contains an example showing CMAQ-simulated tropospheric column NO_2 at a resolution of 4-km compared to OMI NO_2 retrieval over the Great Lakes Region for July 2011.

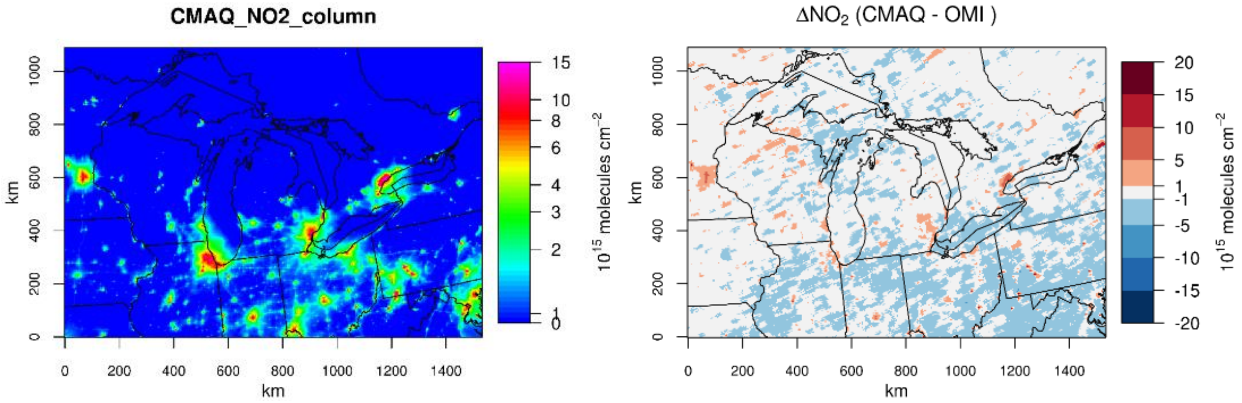


(a)

(b)

---

[10] https://www.cmascenter.org/ioapi/



(c) (d)

Figure 3. (a) OMI NO_2 operational retrieval, (b) Retrieved NO_2 VCD_{trop} re-calculated by using NO_2 profiles from CMAQ, (c) CMAQ-simulated tropospheric column NO_2 and (d) Difference between (c) and (b).

References

1. Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z., and Streets, D. G.: A high-resolution and observationally constrained OMI NO_2 satellite retrieval, Atmospheric Chemistry and Physics, 17, 11403-11421, 2017.

2. Lamsal, L., Krotkov, N., Celarier, E., Swartz, W., Pickering, K., Bucsela, E., Gleason, J., Martin, R., Philip, S., and Irie, H.: Evaluation of OMI operational standard NO_2 column retrievals using in situ and surface-based NO 2 observations, Atmospheric Chemistry and Physics, 14, 11587-11609, 2014.

3. Duncan, B. N., Geigert, and M., Lamsal, L.: A brief tutorial on using the Ozone Monitoring Instrument (OMI) nitrogen dioxide (NO_2) data product for SIPS preparation, HAQAST Tech. Guid. Doc. No. 3, 2018. Doi:10.7916/D80K3S3W, available at https://doi.org/10.7916/D80K3S3W.

4. Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., and Sulprizio, M. P.: Why do models overestimate surface ozone in the Southeast United States? Atmospheric Chemistry and Physics**,** *16*, 13561-13577, 2016.

Appendix (R Code)

```
library(rhdf5)
library(ncdf)
library(sp)
library(raster)
library(rgdal)

#This program will overwrite the variable ColumnAmountNO2Trop in the OMNO2 files. Back up the files before any changes are made
#OMIDIR/JULDDD is where the OMINO2 files are put
setwd("OMIDIR/JULDDD")

#################################################################
# Read NO2 conc, full layer height and air density from CMAQ output
#################################################################

# NO2
 NO2file<-open.ncdf("RDIR/TODAY/NO2file", write=F)
 NO2conc_tmp<-get.var.ncdf(NO2file,"NO2",start=c(1,1,1,19),count=c(-1,-1,-1,3))  #18:00UTC or 13:00EST

# Layer height of each model layer in CMAQ, obtained from full layer height (ZF)
 Metfile<-open.ncdf("RDIR/TODAY/Metfile", write=F)
 ZF_tmp<-get.var.ncdf(Metfile,"ZF",start=c(1,1,1,19),count=c(-1,-1,-1,3))
 ZFbot_tmp<-array(0,dim = c(dim(ZF_tmp)))
 ZFbot_tmp[,,2:35,]<-ZF_tmp[,,1:34,]   # Bottom of a layer
 ZH_tmp<- ZF_tmp - ZFbot_tmp  # Layer height

# Air density
 DENS_tmp<-get.var.ncdf(Metfile,"DENS",start=c(1,1,1,19),count=c(-1,-1,-1,3))

# Pressure
 PRES<-get.var.ncdf(Metfile,"PRES",start=c(1,1,1,19),count=c(-1,-1,-1,3))
 dim(PRES)<-c(dim(PRES)[1]*dim(PRES)[2],dim(PRES)[3],dim(PRES)[4]) #PRES(COL*ROW,LAYER,TIME)
```

```
 PRES2<-apply(PRES,1:2,mean)  # 3hr average; PRES2(COL*ROW,LAYER)

# Longitude & latitude of each grid
 Latlonfile<-open.ncdf("RDIR/TODAY/Latlonfile", write=F)
 LONCMAQ<-get.var.ncdf(Latlonfile,"LON",start=c(1,1,1,1),count=c(-1,-1,-1,-1))
 LATCMAQ<-get.var.ncdf(Latlonfile,"LAT",start=c(1,1,1,1),count=c(-1,-1,-1,-1))

 close.ncdf(NO2file)
 close.ncdf(Metfile)
 close.ncdf(Latlonfile)

# Compute NO2 partial column for CMAQ grids
 M2CM <- 1.0e2    # m to cm
 PPM_MCM3 <- 1.0e-06 # mol(species)/mol(air) = ppm * 1.0e-06
 AVO <- 6.0221367e+23  # Avogadro's constant [number/mol]
 MWAIR <- 28.9628  # Mean molecular weight for dry air [g/mol]
 DES_CONV <- (1.0e3 * AVO / MWAIR) * 1.0e-06  # convert from kg/m**3 to g/cm**3
 NO2_PC_tmp <- PPM_MCM3 * DENS_tmp * DES_CONV * ZH_tmp *M2CM* NO2conc_tmp
 NO2_PC <- apply(NO2_PC_tmp, 1:3, mean) # 3hr average
 dim(NO2_PC)<-c(dim(NO2_PC)[1]*dim(NO2_PC)[2],dim(NO2_PC)[3])

# Create CMAQ grids
 LCC <- c("+proj=lcc +datum=WGS84 +lat_1=33 +lat_2=45 +lat_0=40 +lon_0=-97 +x_0=0 +y_0=0")
 LATLON <- c("+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs")
 CMAQCOR<-data.frame( c(LONCMAQ),c(LATCMAQ));colnames(CMAQCOR)<-c("x","y")
 coordinates(CMAQCOR)<-~x+y
 proj4string(CMAQCOR)<-CRS(LATLON)

###################################
# Start to process OMI Level2 data
####################################
 files<-system("ls -lh OMI-Aura_L2-OMNO2_2011m????t????-??????_v???-????????????????.he5",intern=T)
 nf<-length(files)
```

```r
for(f in 1:nf) {
 files[f]
 parts<-unlist(strsplit(files[f], " "))
 file<-parts[length(parts)]
 writeLines(paste("input file:",file))

# Read variables from NASA L2 data
 AmfTrop<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfTrop")
 AmfTrop[AmfTrop>(-1.267652e+30) & AmfTrop<(-1.267650e+30)]<-NA  #fill-value 1.267251e+30
 AmfTrop[c(1:5,56:60),]<-NA # Remove pixels at the swath edges
 VCDTrop<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ColumnAmountNO2Trop")
 VCDTrop[VCDTrop>(-1.267652e+30) & VCDTrop<(-1.267650e+30)]<-NA
 VCDTrop[c(1:5,56:60),]<-NA
 SolarZA<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Geolocation Fields/SolarZenithAngle")
 CloudRF<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/CloudRadianceFraction")*0.001  # scale factor=0.001
 SurfAb<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/TerrainReflectivity")*0.001
 VcdFlag<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/VcdQualityFlags")

 data<-array(c(AmfTrop, VCDTrop, SolarZA, CloudRF, SurfAb, VcdFlag), dim = c(dim(VCDTrop),6))
 data[data[,,3] >= 80      # Solar zenith angles >= 80 degree
     |data[,,4] >= 0.5     # Cloud radiance fractions >= 0.5
     |data[,,5] >=0.3      # Surface albedo >= 0.3
     |data[,,6]%%2 != 0 ]<-NA  # VcdFlag is an odd integer

 SW<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ScatteringWeight")
 SW[SW>(-1.267652e+30) & SW<(-1.267650e+30)]<-NA
 SWPRES<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ScatteringWtPressure")*100
 LONCOR<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Geolocation Fields/FoV75CornerLongitude")
 LATCOR<-h5read(file,"/HDFEOS/SWATHS/ColumnAmountNO2/Geolocation Fields/FoV75CornerLatitude")
```

```
# Create OMI pixels
 COOR <-cbind(c(LONCOR[,,1]),  c(LATCOR[,,1]), c(LONCOR[,,4]), c(LATCOR[,,4]),
c(LONCOR[,,3]), c(LATCOR[,,3]), c(LONCOR[,,2]), c(LATCOR[,,2]), c(LONCOR[,,1]),
c(LATCOR[,,1]))   # clockwise meaning island, and counter-clockwise meaning hole
 ID <- paste0("omi",seq_len(nrow(COOR)))


 OMIPIX <- SpatialPolygons(mapply(
          function(poly,id){
          xy<-matrix(poly,ncol=2,byrow=TRUE)
          Polygons(list(Polygon(xy)),ID=id)},
          split(COOR,row(COOR)),ID),proj4string=CRS(LATLON))


# Assign SW to CMAQ grids (horizontal&vertical)
 SWROW <- aperm(SW,c(2,3,1))
# dim(SWROW) <- c(98640,35)
 dim(SWROW)<-c(dim(SWROW)[1]*dim(SWROW)[2],dim(SWROW)[3])
 OMIPIX_SW <- SpatialPolygonsDataFrame(OMIPIX,data.frame(row.names=ID, code=ID,
sw=SWROW))
 CMAQ_SW<-over(CMAQCOR, OMIPIX_SW)  #data.frame


 if(sum(is.na(CMAQ_SW[,2]))== length(CMAQ_SW[,2])){
#    writeLines("No overlaps between CMAQ grids and OMI pixels. Skip to the next iteration")
    next}


 CMAQ_SW_tmp<-mapply(function(sw_old,pres_cmaq){

ifelse(is.na(sw_old[1]),NA,approx(x=SWPRES,y=sw_old,xout=pres_cmaq,rule=2)$y)},split(CMAQ_SW
[,2:36],row(CMAQ_SW[,2:36])),PRES2[,])
 dim(CMAQ_SW_tmp)<-dim(PRES2)


# Recalculate air mass factor
 AMF_tmp<-NO2_PC*CMAQ_SW_tmp
 AMF<-apply(AMF_tmp[,1:31],1,sum)/apply(NO2_PC[,1:31],1,sum)   # AMFTrop = sum
(SW*NO2_PC)/sum(NO2_PC) (from surface to tropopause, Layer 1-31, ~200hPa)
 AMF_SP <- SpatialPointsDataFrame(CMAQCOR,data.frame(amf=AMF))
 OMI_AMF <- over(OMIPIX,AMF_SP,fun=mean,na.rm=T)
```

```
 VCDTrop_new<-c(data[,,1]*data[,,2])/OMI_AMF[,1]  # SCDTrop = AmfTrop*VCDTrop
 dim(VCDTrop_new)<-c(dim(VCDTrop))
# VCDTrop_new[is.na(VCDTrop_new)] <- -1.267651e30
 h5write(VCDTrop_new,file,"/HDFEOS/SWATHS/ColumnAmountNO2/Data
Fields/ColumnAmountNO2Trop")


}
```