Harmonization of tropospheric ozone data records from satellites

A. Keppens, D. Hubert, O. Nath, S. Compernolle, and J.-C. Lambert
“Tropospheric ozone dataset validation and harmonization” in response to TOAR-II Satellite Ozone WG needs, addressing
- wide range of tropospheric ozone levels and trends from satellites
- differences in vertical sensitivity, sampling, and vertical range
- changes in bias over time
by application of
- satellite data harmonization
- harmonized validation process to identify and quantify remaining discrepancies

(TOAR-I, Gaudel, et al., 2018)
Transfer standard: CAMS Reanalysis EAC4

- CAMSRA (Inness et al., 2019)
  - 60 levels (1012-0.1 hPa) → lat-lon dependent, fixed in time
  - 0.7° x 0.7° (global) → regridded to 1° x 1°
  - 6 hours (2003-2021) → UTC interpolated to LST satellite overpass time
  - assimilated ozone: TC (SCIA, OMI, GOME-2A/B), PROF (MIPAS, Aura-MLS, SBUV/2)

- Used for
  - tropospheric column extension
  - new prior info
  - auxiliary data source for quantity matching
  - (assessment of horizontal and temporal sampling differences)

- Complementary choice to TCR-2 reanalysis (Miyazaki et al., 2020)
  transfer standard used by TOAR-II SOWG
VC-20-01 tropospheric ozone datasets

**Residual technique** (UV-VIS sensors)
- Convective Cloud Differential
- Nadir minus Limb
- Nadir minus Reanalysis

**Optimal Estimation profile retrieval**
- UV-VIS
- IR
- VIS+IR synergy
Harmonization of tropospheric column data

- Different data records have different Top of Troposphere definition $l$
  - CCD DLR: fixed pressure 270 hPa
  - CCD IUP: fixed pressure 200 hPa
  - SUNLIT: altitude WMO lapse-rate tropopause – 3 km (ERA5)
  - S5P-BASCOE: pressure dynamical tropopause (ERA5)
  - OMI-MLS: pressure WMO lapse-rate tropopause (NCEP)

- Harmonization equation (top level $l \rightarrow l'$):
  $$X(\text{sfc}, l') = X(\text{sfc}, l) + \int_{\text{sfc}}^{l'} x_{\text{CAMSRA}}(p) dp - \int_{\text{sfc}}^{l} x_{\text{CAMSRA}}(p) dp$$

- Challenges:
  - Not reported in data files: tropospheric top level, geolocation input data
  - Harmonized monthly mean data ≠ monthly mean of harmonized daily data
  - Correct for mean bias between $X(\text{sfc}, l)$ and $X_{\text{CAMSRA}}(\text{sfc}, l)$? How?
Illustration: native ToTr level $\rightarrow$ LRT$_{NCEP}$

Temporal mean (2019) harmonization term

Clear impact on spatial distribution due to varying top level definition.

- LRT$_{NCEP}$: OMI-MLS
- LRT-3km: SUNLIT
- dyn. TP: S5P-BASCOE
Illustration: native ToTr level $\rightarrow$ LRT\textsubscript{NCEP}

Spatial mean harmonization term
Some impact on long-term changes, due to varying ToTr level definition.
Illustration: native ToTr level $\rightarrow$ LRT$_{\text{NCEP}}$

Mean bias SAT – OMI-MLS [DU] (20S-20N)

- OMI
- SCIAMACHY
- GOME2A
- GOME2B
- TROPOMI
- GOME-SCIA-GOME2A
- OMI-SUNLIT
- TROPOMI-SUNLIT
- TROPOMI-BASCOE

Mean bias SAT – OMI-MLS [DU] (20S-20N)

- OMI
- SCIAMACHY
- GOME2A
- GOME2B
- TROPOMI
- GOME-SCIA-GOME2A
- OMI-SUNLIT
- TROPOMI-SUNLIT
- TROPOMI-BASCOE

LRT-3km

dyn TP

270 hPa

200 hPa

~2 PM

~10 AM
**VC-20-01 tropospheric ozone datasets**

<table>
<thead>
<tr>
<th>Year</th>
<th>TOMS N</th>
<th>TOMS EP</th>
<th>Merged GOME-type</th>
<th>SCIAMACHY</th>
<th>OMII</th>
<th>GOME-2A</th>
<th>GOME-2B</th>
<th>SSP</th>
<th>OMII-MLS</th>
<th>OMPS N°-LP</th>
<th>OMPS (L,limb)</th>
<th>SSI (L,limb)</th>
<th>OMPS-MERRA2</th>
<th>EPIC-MERRA2</th>
<th>SSI-BASCOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Residual technique (UV-VIS sensors)**
- Convective Cloud Differential
- Nadir minus Limb
- Nadir minus Reanalysis

**Optimal Estimation profile retrieval**
- UV-VIS
- IR
- VIS+IR synergy
Harmonization and comparison of vertically resolved atmospheric state observations: methods, effects, and uncertainty budget

Arno Keppens, Steven Compernolle, Tijl Verhoest, Daan Hubert, and Jean-Christopher Lambert
Department of Atmospheric Composition, Royal Belgian Institute for Space Aeronomy (BIRA-IASB), 1180 Brussels, Belgium

Correspondence: Arno Keppens (arno.keppens@aeronomie.be)
Received: 21 March 2019 – Discussion started: 4 April 2019
Revised: 18 June 2019 – Accepted: 3 July 2019 – Published: 15 August 2019

Removing Prior Information from Remotely Sensed Atmospheric Profiles by Wiener Deconvolution Based on the Complete Data Fusion Framework

Arno Keppens *, Steven Compernolle , Daan Hubert , Tijl Verhoest , José Granville and Jean-Christopher Lambert

Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Avenue Circulaire 3, 1180 Brussels, Belgium;
steven.compernolle@aeronomie.be (S.C.); daan.hubert@aeronomie.be (D.H.); tijl.verhoest@aeronomie.be (T.V.);
josé.granville@aeronomie.be (J.G.); lambert@aeronomie.be (J.-C.L.)
* Correspondence: arno.keppens@aeronomie.be; Tel.: +32-2-373-0412
Harmonization for a single profile retrieval

<table>
<thead>
<tr>
<th>Matching operation</th>
<th>$x'$</th>
<th>$S'$</th>
<th>$A'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vert. quantity matching</td>
<td>$Mx$</td>
<td>$MSM^T$</td>
<td>$MAM^{-1}$</td>
</tr>
<tr>
<td>Vert. sampling matching</td>
<td>$Wx$</td>
<td>$WSW^T$</td>
<td>$WAW^*$</td>
</tr>
<tr>
<td>Vert. smoothing matching</td>
<td>$Vx$</td>
<td>$VSV^T$</td>
<td>$VA$</td>
</tr>
<tr>
<td>Meas. weight matching</td>
<td>$WM_x$</td>
<td>$S$</td>
<td>$WM_A$</td>
</tr>
<tr>
<td>Prior matching (PM)</td>
<td>$S'(S^{-1}x - R_a x_a + R'_a x'_a)$</td>
<td>$(S^{-1} - R_a + R'_a)^{-1}$</td>
<td>$A + SS_a^{-1} - S'S'_a^{-1}$</td>
</tr>
<tr>
<td>Re-optimized PM</td>
<td>$P[x - (I - A)x_a] + PS(A^T)^{-1}R'_a x'_a$</td>
<td>PSP$^T$</td>
<td>PA</td>
</tr>
<tr>
<td>AK smoothing (for $s$ on $r$)</td>
<td>$A_s x_r + (I - A_s)x_{a,s}$</td>
<td>$A_s S_r (A_s^T)^T$</td>
<td>$A_s$</td>
</tr>
<tr>
<td>Maximum likelihood repr.</td>
<td>$S'(S^{-1} - R_a x_a)$</td>
<td>$(S^{-1} - R_a)^{-1}$</td>
<td>$I$</td>
</tr>
<tr>
<td>Information-centered repr.</td>
<td>$W(S^{-1} - R_a)^{-1}(S^{-1}x - R_a x_a)$</td>
<td>$W(S^{-1} - R_a)^{-1}W^T$</td>
<td>$I'$</td>
</tr>
<tr>
<td>Co-location matching</td>
<td>$x - \Delta m$</td>
<td>$S + S_{\Delta m}$</td>
<td>$A - S_{\Delta m} S_a^{-1}$</td>
</tr>
</tbody>
</table>

(Keppens, et al., 2019)
Based on Complete Data Fusion (CDF) framework

CDF (Ceccherini et al., 2022):

\[
x'_{N} = \left( \sum_{i=1}^{N} S_{i}^{-1} A_{i} + S_{a}^{\prime -1} \right)^{-1} \left( \sum_{i=1}^{N} S_{i}^{-1} [x_{i} - (I - A_{i}) x_{a,i}] + S_{a}^{\prime -1} x'_{a} \right)
\]

where the sum is taken over \( N \) profiles within a predefined spatiotemporal domain

The corresponding averaging kernel matrix and covariance matrix are given by

\[
A'_{N} = \left( \sum_{i=1}^{N} S_{i}^{-1} A_{i} + S_{a}^{\prime -1} \right)^{-1} \left( \sum_{i=1}^{N} S_{i}^{-1} A_{i} \right)
\]

and

\[
S'_{N} = \left( \sum_{i=1}^{N} S_{i}^{-1} A_{i} + S_{a}^{\prime -1} \right)^{-1}
\]

which is mathematically equivalent to a joint retrieval (Ceccherini et al., 2015)
Choice of new prior info determines CDF output

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>$x'_a$</th>
<th>$S'_{a,i^{-1}}$</th>
<th>$x'_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>$x_m$</td>
<td>$\alpha S_m^{-1}$</td>
<td>$(\sum S_i^{-1} A_i + \alpha S_m^{-1})^{-1} \left( \sum S_i^{-1} [x_i - (I - A_i)x_{a,i}] + \alpha S_m^{-1} x_m \right)$</td>
</tr>
<tr>
<td>APR</td>
<td>$x_m$</td>
<td>$\sum S_{a,i}^{-1}$</td>
<td>$(\sum S_i^{-1})^{-1} \left( \sum S_i^{-1} [x_i - (I - A_i)(x_{a,i} - x_m)] \right)$</td>
</tr>
<tr>
<td>WAV</td>
<td>$x_{a,i}$</td>
<td>$\sum S_{a,i}^{-1}$</td>
<td>$(\sum S_i^{-1})^{-1} \left( \sum S_i^{-1} x_i \right)$</td>
</tr>
<tr>
<td>MLR</td>
<td>$x_{a,i}$</td>
<td>$0$</td>
<td>$(\sum S_i^{-1} A_i)^{-1} \left( \sum S_i^{-1} [x_i - (I - A_i)x_{a,i}] \right)$</td>
</tr>
<tr>
<td>ICR*</td>
<td>$0$</td>
<td>$0$</td>
<td>$(\sum W_i^* S_i^{-1} A_i W_i^<em>)^{-1} \left( \sum W_i^</em> S_i^{-1} [x_i - (I - A_i)x_{a,i}] \right)$</td>
</tr>
</tbody>
</table>

~ deconv. on original grid

~ deconv. on DFS grid

~ MLR if $\alpha = 0$
Choice of new prior info determines CDF output

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>$x'_a$</th>
<th>$S'_{a^{-1}}$</th>
<th>$x'_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>$x_m$</td>
<td>$\alpha S^{-1}_m$</td>
<td>$(\sum S^{-1}_i A_i + \alpha S^{-1}_m)^{-1} (\sum S^{-1}_i)$</td>
</tr>
<tr>
<td>APR</td>
<td>$x_m$</td>
<td>$\sum S^{-1}_{a,i}$</td>
<td>$(\sum S^{-1}_i)^{-1} (\sum S^{-1}_i [x_i])$</td>
</tr>
<tr>
<td>WAV</td>
<td>$x_{a,i}$</td>
<td>$\sum S^{-1}_{a,i}$</td>
<td>$(\sum S^{-1}_i)^{-1} (\sum S^{-1}_i [x_i])$</td>
</tr>
<tr>
<td>MLR</td>
<td>$x_{a,i}$</td>
<td>0</td>
<td>$(\sum S^{-1}_i A_i)^{-1} (\sum S^{-1}_i [x_i])$</td>
</tr>
<tr>
<td>ICR*</td>
<td>0</td>
<td>0</td>
<td>$W'_i W'^*_i (\forall i)$</td>
</tr>
</tbody>
</table>

---

**FTIR spectrometer (208 profs.) on layers**

**IASI-A FORLI v2015 (514 profs.) on layers**

**GOME–2A RAL v2.14 (347 profs.) on layers**

**MIPAS ORM v8.22 (160 profs.) on layers**
Harmonization of TOAR-II profile datasets

- L2 profiles → monthly-gridded (1° x 1°) & harmonized (prior info) satellite L3 data
- Illustration: 1° x 1° box containing Brussels, for January 2020
- Comparison with both CAMSRA and Uccle ozonesondes
- Sampling example: March 24, 2021 (courtesy of N. Zoppetti, IFAC)
Illustration: RAL OMI v2.14
Illustration: RAL OMI v2.14

WAV

APR

MLR

CDF-1

CDF-10

CDF-100

Pressure [hPa]

Ozone vmr [ppmv]

Pressure [hPa]

Ozone vmr [ppmv]

Pressure [hPa]

Ozone vmr [ppmv]

Pressure [hPa]

Ozone vmr [ppmv]

Pressure [hPa]

Ozone vmr [ppmv]

Legend:
- sat. prior
- CAMS
- sonde (12)
- retr. (8)
Dispersion reduction upon harmonization? (9 sats)
CEOS VC-20-01: harmonization progress

- (TOAR) tropospheric ozone assessment needs identified
- CAMSRA transfer standard selection / manipulation
- Satellite tropospheric ozone harmonization:
  - Ozone tropospheric columns from vertical correction
  - Ozone profiles from Complete Data Fusion framework:
    APR, MLR, and CDF (APR method // TOAR-II SOWG approach)
- Next:
  - Harmonization refinements / sensitivity studies / screening
  - Full harmonized time series processing
  - Ground-based quantification of residual discrepancies
- Eventually:
  - Calculating ozone burden (filling of gaps?)
  - Tropospheric ozone assessment studies