OMBRO—OMHCHO—OMOCLO De-Striping README FILE

Date of this Document: 1 February 2007

Overview

This document describes algorithm implementations of destriping corrections for the L2 products of the SAO PGEs OMBRO, OMHCHO, and OMOCLO. The corrections consist of: (a) A ground pixel by ground pixel determination (and subsequent exclusion) of "outliers", *i.e.*, abnormally large values, in the fitting residuals; (b) the use of a composite solar spectrum, derived from a number of individual irradiance measurements, in lieu of the daily solar observations; and (c) cross-track smoothing of the retrieved columns by fitting a polynomial and an average cross-track stripe pattern derived from the average of a range of swath lines to the columns across a single swath line and replacing the originally retrieved columns with the smoothed values. While options *(a)* and *(b)* can be considered an improvement of screening for and eliminating pixels with larger than average dark current in the radiance and irradiance measurements, option *(c)* is highly experimental and almost certainly introduces an as yet unquantified bias to the fitted columns that the user of the data should be aware of. However, we encourage users of L2 data to provide feedback on the quality of the smoothed data. Each SAO data product provides two complete sets of columns in the data fields: *ColumnAmount*, which contains the columns derived from on-line destriping procedures (methods *(a)* and *(b)* only), and *ColumnAmountDestriped*, which is derived from the former by applying the cross-track smoothing (option *(c)*).

Release-Specific Information

OMBRO, OMHCHO, and OMOCLO data products, Software Versions 0.9.50 (validation release) and 1.0 (public release).

Origin of Cross-Track Striping

To a higher or lesser degree, any L2 data product inherits cross-track striping from the OMI L1b data product. Weaker absorbers like BrO, HCHO, and OClO are more strongly affected than ozone or nitrogen dioxide because the peak-to-peak magnitude of the striping becomes comparable to the column amounts. The main cause of the stripes in the L1b data are hot and transient pixels in the CCD as well as elevated dark current signals, all of which lead to increased noise in the affected wavelength positions and, in turn, manifest as abnormally large (compared to a statistical average over a large number of wavelength positions) fitting residuals. Panel (a) in Figure 1 shows the original column retrieval from a BrO fit: Stripes are clearly visible and can reach magnitudes several times of what may be considered reasonable column values.

Subsequent versions of the OMI L1b data product have seen improvements in the identification of hot and transient pixels as well as the correction for dark current, which has reduced the magnitude of the cross-track striping. It is expected that the soon to be implemented dynamic update of the dark current measurements in the level 0-1b processing will reduce the striping considerably, although not necessarily eliminate it. However, for weak absorbers like BrO, HCHO, and OClO the spectral features that lead to striping are on the 0.5% level or less and hence extremely difficult to remove from the L1b product without possibly compromising information that originates from the absorbers themselves. Additional reduction of stripe effects has therefore to be implemented in the level 1b-2 processing.

Destriping Corrections

(a) Removal of outliers in fitting residuals

Figure 2 shows fitting residuals from a BrO fit for three selected cross-track positions. The black line in each panel (mostly covered by the orange line) is the fitting residual for a single radiance fit. Typical residuals for BrO are of magnitude ±0.003 (arbitrary units; roughly equivalent to ±0.3%), but a few outliers of up to ±0.015 are present. Different cross-track positions may contain such outliers at different wavelength positions, but the same cross-track position often shows the same outlier at the same wavelength position along track, correlated with the presence of a stripe.

Outlier identification in the fitting residuals uses thresholds that are automatically determined from the actual fit: Any wavelength position in the fitting window that contains a residual larger than 3 standard deviations around the median of the residual will be flagged, and the radiance is fitted again with flagged positions excluded from the fit. The orange lines in Figure 2 show the fitting residual of this subsequent fit - largely identical to the original residual but with marked differences at positions where an outlier was identified. Panel (b) in Figure 1 shows the resultant improvement in the BrO columns from Panel (a). A reduction in the cross-track striping is clearly discernible, indicating that the spectral features observed in the fitting residuals are indeed connected to the striping.

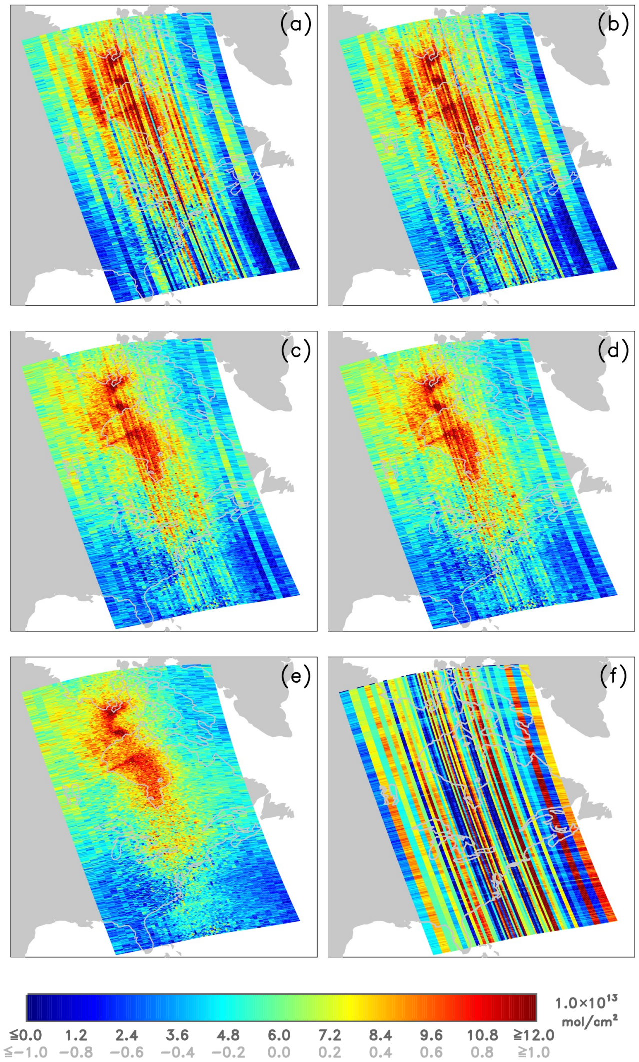

Figure 1: Destriping of BrO columns. (a): individual irradiance measurement, no outlier removal; (b): like "(a)" with outlier removal; (c): solar composite spectrum, no outlier removal; (d): like "(c)" with outlier removal; (e): post-processing destriping of "(d)"; (f): correction factor "(d) -> (e)". Black numbers on legend are for panels (a) to (e), gray numbers for the correction factor (f).

Great care has to be taken when selecting acceptable thresholds for the fitting residuals, as too small a threshold may lead to the discard of spectral information on the retrieval target. An important check for the validity of the selected threshold is that, on average, the retrieved column amounts do not vary significantly. Changes in the column amounts of cross-track positions affected by striping are expected (actually, desired) but if all 60 cross-track pixels are considered, the average column should remain fairly constant. Too great a change between first and second fit indicate the removal of important information from the spectrum. By defining thresholds in terms of standard deviations of the actual fitting residual, variations between spectra are automatically taken into account.

An added benefit from the removal of outliers is the reduction in fitting uncertainties. Since wavelength positions associated with outliers are excluded from the fit, the RMS error is reduced, and the resulting fitting uncertainties are improved.

It should be noted that, if used with appropriate care, there is nothing questionable about the approach to remove outliers from the residual. The method is on par with the flagging of hot or transient pixels in the L1b process.

Implementation: On-line; default. Column amounts are provided in the data field *ColumnAmount*

(b) Use of a composite solar spectrum

The irradiance measurements have been identified as the major source for stripes in the L2 data products. Improvements in the level 0-1b processing, particularly the soon to be implemented dynamic update of dark current measurements, are expected to greatly reduce the striping. Yet, at this point in time, the retrieval of weakly absorbing gases like BrO, HCHO, and OClO benefits from a work-around solution to reduce the impact of the affected solar measurements. To this end, a

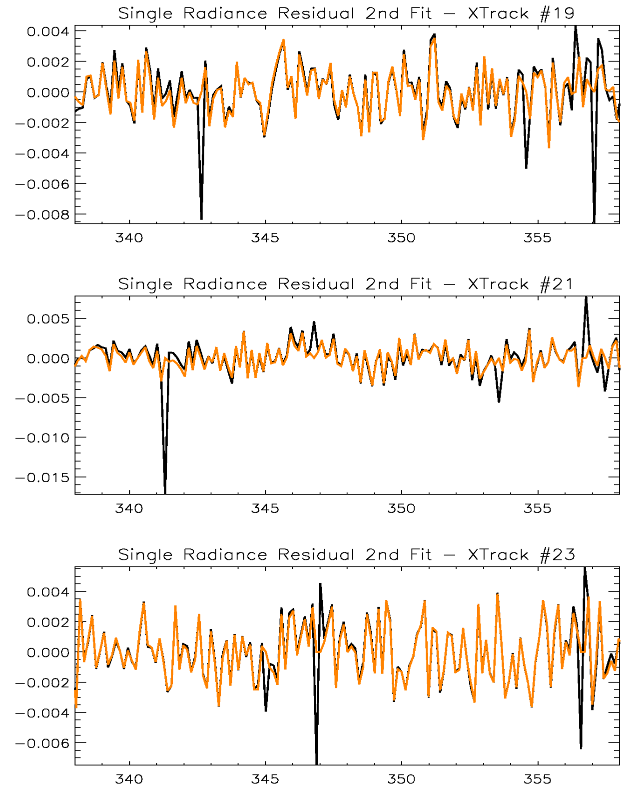

Figure 2: BrO fitting residuals for three selected cross-track positions (19, 21, 23) showing the original residual (black line) and the residual after exclusion of outliers (orange line).

solar composite spectrum – derived from a large number of individual OMI solar measurements - is used in the retrieval in lieu of the daily irradiance spectra.

The composite solar spectrum is derived from a set of individual irradiance measurements in the following manner:

1. Wavelength calibration of the individual irradiances.
2. Interpolation of the spectra to a common wavelength grid.
3. Computation of the composite spectrum either from the median of all irradiance spectra or principal component analysis.

Just like in the original measurements, a separate spectrum is derived for each of the 60 cross-track positions. An example of a composite spectrum is shown in Figure 3. Further examples as well as actual data can be found here.

Panel (c) in Figure 1 shows the effect of the solar composite spectrum on the BrO retrievals from Panel (a). The columns in Panel (d) are further improved by the use of outlier identification as described above. Note that the effect "(c) -> (d)" is less pronounced than "(a) -> (b)" but an additional reduction in the stripes is still discernible.

Implementation: On-line; default. Column amounts are provided in the data field *ColumnAmount*

(c) Post-processing cross-track smoothing

In addition to outlier removal and the use of a composite solar spectrum, a cross-track smoothing procedure has been implemented in all SAO PGEs. This smoothing is considered experimental, and the smoothed data product should be used with great caution. It is provided primarily as a diagnostic tool for stripe correction. Feedback is encouraged from users who may want to compare the smoothed product to the standard L2 columns.

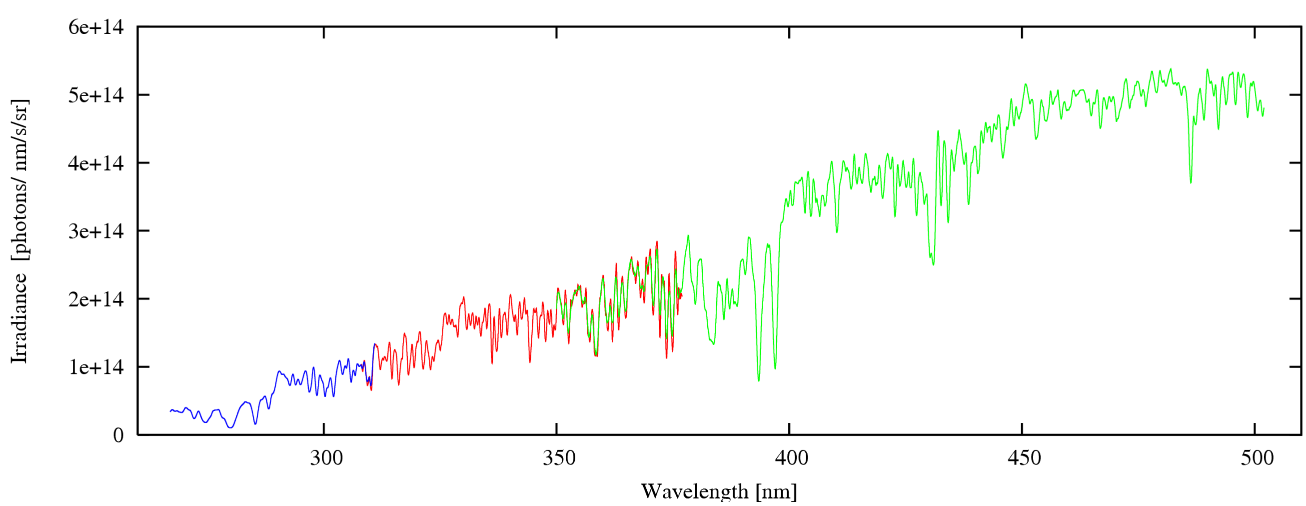

Figure 3: Example of an OMI composite solar spectrum for UV-1 (blue), UV-2 (red), and VIS (green) channels.

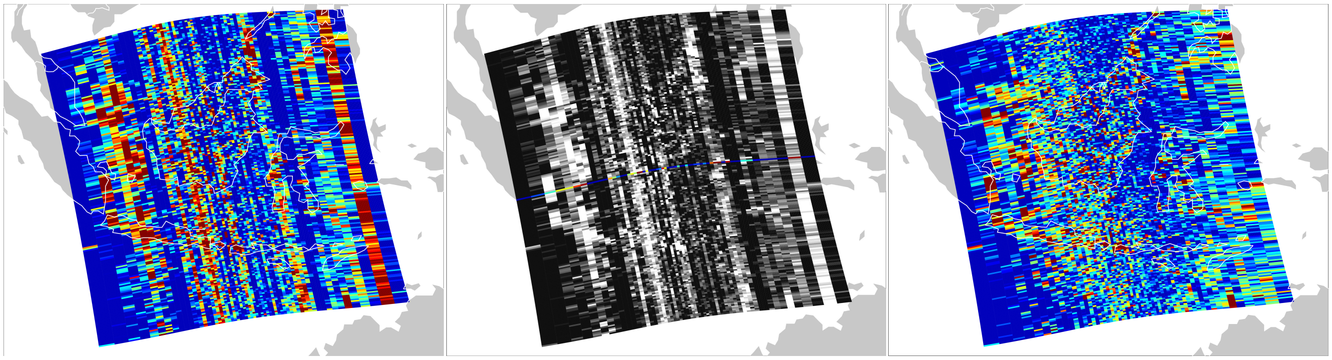

Figure 4: Cross-track smoothing of HCHO columns. Left: Columns as provided in the *ColumnAmount* data field. Center: The to-be-destriped center swath line (color) and the 200 surrounding swath lines (b&w) that are averaged; see Figure 5 for details on the smoothing. Right: smoothed columns as provided in the *ColumnAmountDestriped* data field.

For each individual swath-line, cross-track smoothing proceeds along the following steps:

1. For each swath line *n*, a range of surrounding swath lines is selected; the current default is 200, *i.e.*, [*n*-100, *n*+100] for a block of 201 swath lines. See middle panel in Figure 4.
2. The average cross-track column pattern C_{avg} is computed from the along-track average for each cross-track position - Panel (a) in Figure 5.
3. A low-order polynomial (current default is 5th order) *P* is fitted to the average cross-track columns - Panel (b) in Figure 5.
4. The average cross-track stripe correction $\Sigma = C_{avg} - P$ is computed (Panel (c) in Figure 5).
5. The loading of Σ in the current swath line *n* is determined through a cross-track fit to the columns of the current swath line *n* – Panel (d) in Figure 5).
6. Finally, the loading of Σ is subtracted from the columns of swath line *n* – Panel (e) in Figure 5.
7. Steps 1-6 are repeated for all lines in the swath.

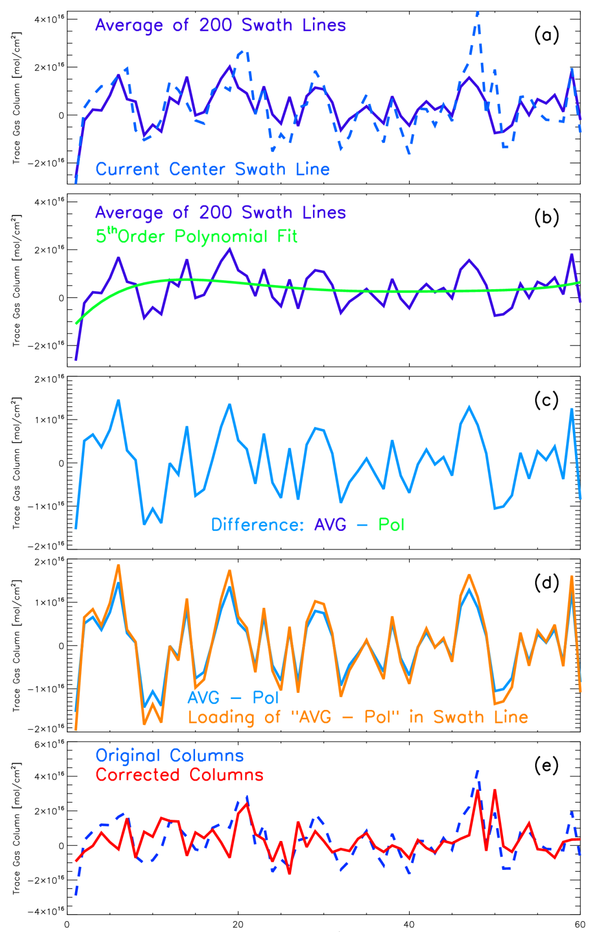

Figure 5: Cross-track smoothing of an individual HCHO swath line. (a) current swath line column values and 200 swath line average (compare Figure 3); (b) polynomial fit; (c) cross-track stripe pattern (difference between average and polynomial); (d) fitted loading of cross-track stripe pattern in current swath line; (e) original and corrected HCHO columns. X-axes in all plots are cross-track position 1 to 60.

Figure 4 shows the cross-track smoothing for a section of a HCHO swath: The left panel contains the original HCHO columns (retrieval including outlier removal and solar composite spectrum), while the right panel shows the smoothed columns. The center panel visualizes the ±100 swath lines centered around swath line *n*. For the first and last 100 lines in a swath, the first/last block of 201 swath lines is used for the computation of the column average.

<u>Implementation:</u> Post-processing. Smoothed columns are provided in the data field *ColumnAmountDestriped*.