Overview of the retrieval algorithm

Hitoshi Irie (JAMSTEC)
Retrieval algorithm development

• The retrieval algorithm development is a key component for maximizing the MAX-DOAS capability.

• Quality of data products (DSCD, column, and profile) is subject to the retrieval method used.

• The lack of deep knowledge may waste your time, when you intend to go to a precise data analysis in the future.

• Here, the detail of the retrieval algorithm development, including the current status, caveats, and future plan, is presented.

• I expect this will assist you in proper interpretation of data products and bring new ideas to you.
<table>
<thead>
<tr>
<th>Retrieval procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dark count analysis</strong></td>
</tr>
<tr>
<td>- Estimate systematic and random components of noise at each pixel</td>
</tr>
<tr>
<td>- Use data at SZA&gt;100°, regarded as dark count observations</td>
</tr>
<tr>
<td><strong>Wavelength calibration</strong></td>
</tr>
<tr>
<td>- Estimate wavelength shift and spectral resolution as a function of wavelength</td>
</tr>
<tr>
<td>- Analyze spectra around noon by iterative nonlinear least squares method using the Kurucz et al. (1984) high-resolution solar spectrum</td>
</tr>
<tr>
<td><strong>DOAS analysis</strong></td>
</tr>
<tr>
<td>- Derive DSCDs of NO₂ and O₄</td>
</tr>
<tr>
<td>- Use DOAS method</td>
</tr>
<tr>
<td><strong>Aerosol retrieval</strong></td>
</tr>
<tr>
<td>- Retrieve vertical profile of aerosol extinction coefficient and box-AMF values</td>
</tr>
<tr>
<td>- Optimal estimation method, with special parameterization of profile</td>
</tr>
<tr>
<td><strong>NO₂ retrieval</strong></td>
</tr>
<tr>
<td>- Retrieve vertical profile of NO₂</td>
</tr>
<tr>
<td>- Optimal estimation method, with special parameterization of profile</td>
</tr>
</tbody>
</table>

Notes: Four steps of them use inversion → the need of careful treatment of data retrieved.
**DOAS analysis**

- **Fitting window**: 460-490 nm
- **Absorbers**:
  - NO₂ (294 K) : Vandaele et al. (1998)
  - O₄ : Hermans (unpublished) *replaced from Greenblatt (1990) in last summer
  - O₃ (273 K) : Bogumil et al. (2003)
  - H₂O (290 K): HITRAN2000 (Rothman et al., 2003)
  - Ring : Chance et al.(2005) model
    (I₀ effect is considered using the method of Aliwell et al. (2002))

- **Degrees of offset polynomial** : 2^{nd} (i.e., c(λ) = a₀ + a₁λ + a₂λ²)
- **Degrees of polynomial** : 3^{rd} (i.e., p(λ) = b₀ + b₁λ + b₂λ² + b₃λ³)

- **Reference**: Derived by interpolating two reference spectra measured at EL=90° (or 70°) within 30 min. before and after off axis measurement is made.

- **Levenberg-Marquardt method**:
  \[ x_{i+1} = x_i + (K^T S K + γ D)^{-1} K^T S [y - F(x_i)] \]

- **Forward model**:
  \[ \ln I(λ) = \ln(I_0(λ) - c(λ)) - \sum_{i=1}^{n} \sigma_i(λ) ΔSCD_i - p(λ) \]

- **Error estimate**:
  \[ \hat{S} = K^T S^{-1} K \]
Example of DOAS analysis

To increase the confidence, intercomparison with other groups is needed!
**CINDI campaign**

- **Cabauw Intercomparison Campaign of Nitrogen Dioxide measuring Instruments**

- **Place**
  Cabauw, the Netherlands (52.0°N, 4.9°E)
  *Meteorological observation site of KNMI, having a landmark of the 213m-high meteorological tower*

- **Period**
  June-July 2009
  *including setup and download periods*

  **June 15-28: semi-blind intercomparison for NO$_2$ and O$_4$ DSCDs**

- **Main objective**
  To intercompare the NO$_2$-measuring instruments that can be used for validation of tropospheric NO$_2$ from satellites.

<table>
<thead>
<tr>
<th>University/Institute</th>
<th>Group leader</th>
<th>Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IASB</td>
<td>Michel Van Roozendael</td>
<td>MAX-DOAS</td>
</tr>
<tr>
<td>JAMSTEC</td>
<td>Hitoshi Irie</td>
<td>MAX-DOAS</td>
</tr>
<tr>
<td>NIWA, Lauder</td>
<td>Karin Kreher</td>
<td>MAX-DOAS</td>
</tr>
<tr>
<td>INTA</td>
<td>Olga Puentedura</td>
<td>2 MAX-DOAS</td>
</tr>
<tr>
<td>IUP Heidelberg</td>
<td>Uni Friess</td>
<td>MAX-DOAS</td>
</tr>
<tr>
<td>IUP Bremen</td>
<td>Andreas Richter</td>
<td>MAX-DOAS</td>
</tr>
<tr>
<td>U. Leicester</td>
<td>Paul Monks</td>
<td>imaging DOAS</td>
</tr>
<tr>
<td>CNRS</td>
<td>Florence Goutail</td>
<td>SAOZ</td>
</tr>
<tr>
<td>WSU</td>
<td>George Mount</td>
<td>MF-DOAS</td>
</tr>
<tr>
<td>NASA</td>
<td>Jay Herman</td>
<td>3 Pandora instruments</td>
</tr>
<tr>
<td>NERC, Cambridge</td>
<td>Howard Roscoe</td>
<td>mini MAX-DOAS</td>
</tr>
<tr>
<td>MPI, Mainz</td>
<td>Thomas Wagner</td>
<td>mini MAX-DOAS</td>
</tr>
<tr>
<td>KNMI</td>
<td>Pieternel Levelt</td>
<td>2 mini MAX-DOAS</td>
</tr>
<tr>
<td>Gwangju IST</td>
<td>Young Joon Kim</td>
<td>mini MAX-DOAS and imaging DOAS</td>
</tr>
<tr>
<td>U Tronto</td>
<td>Kimberly Strong</td>
<td>UV-VIS spectrometer</td>
</tr>
<tr>
<td>EMPA</td>
<td>Dominik Brunner</td>
<td>2 TEI 42C in situ</td>
</tr>
<tr>
<td>RIVM</td>
<td>Rene van der Hoff</td>
<td>NO2 lidar</td>
</tr>
<tr>
<td>RIVM</td>
<td>Arnoud Apiturely</td>
<td>Aerosol lidar</td>
</tr>
</tbody>
</table>
The JAMSTEC's MAX-DOAS has been placed, next to the Bremen's MAX-DOAS, on the roof of the porte cabin #4.
Problems and solutions

• **Problem 1:** The fitting window 425-490 nm was too wide to finish the daily DOAS analysis in time.
  → **Solution:** We are now trying to reduce the number of inverse matrix calculations.

• **Problem 2:** NO$_2$ and O$_4$ DSCDs tended to be underestimated, particularly in the morning and evening.
  → **Solution:** The treatment of offsets in our DOAS analysis was inappropriate. We fixed this problem during CINDI.

---

Our DOAS equation

\[
\ln I(\lambda) = \ln(I_0(\lambda) - c(\lambda)) - \sum_{i=1}^{n} \sigma_i(\lambda) \Delta \text{SCD}_i - p(\lambda)
\]

The offset "c" is represented by the 2nd order polynomial.

---

**Original:** raw signals containing apparent offsets are used as $I$&$I_0$.

\[
\downarrow
\]

**Improved:** "c" is unchanged but signals after subtracting the dark count level (apparent offsets) are used as $I$&$I_0$. 
Example of results
(NO$_2$ DSCCDs at 2 degrees)
Preliminary comparison during CINDI

NO\textsubscript{2} differential slant columns for 02.07.2009 and 2° elevation angle

Statistics:
- Bremen – Jamstec
  - Slope: 0.96
  - Correlation: 0.998
- Bremen – Heidelberg
  - Slope: 1.06
  - Correlation: 0.993
- Bremen – NASA
  - Slope: 0.86
  - Correlation: 0.999
- Bremen – INTA
  - Slope: 0.96
  - Correlation: 0.997

Courtesy of F. Wittrock
Profile retrieval
Overview of the profile retrieval algorithm

Spectra → DOAS fit

Init. aerosol profile → \( \triangle \text{SCD}(O_4) \) → Agree? (Yes → Aerosol profile, No → Modify the profile)

Init. gas profile → Modeled \( \triangle \text{SCD}(\text{gas}) \) → Agree? (Yes → Gas profile, No → Modify the profile)

Modify the profile → As AMF
Ideas for the development

• I learned the concept of MAX-DOAS from the ACP paper of Hönninger et al. (2004), who was the pioneer of MAX-DOAS.

• I carried out the development without detailed simulations, but I referred much to ACP papers of Wagner et al. (2004), Frieß et al. (2006), PhD thesis of Wittrock (2006), and WinDOAS user manual (Fayt and van Roozendael, 2001). And, of course, many other publications were certainly helpful.

• Participation in a radiative transfer model (RTM) inter-comparison study (Wagner et al., 2007) was a great chance for me to learn.
Radiative transfer model

- Monte Carlo Atmospheric Radiative Transfer Simulator (Iwabuchi, JAS, 2006)

- A parallelized 3D radiative transfer model utilizing the forward-propagating Monte Carlo photon transport algorithm.

- Box-AMF and the radiance intensity calculations have been validated through inter-comparison with other radiative transfer models (Wagner et al., 2007)

- Box-AMF varies much depending on aerosols!

- A lookup table of box-AMF was created as functions of EL, sun position, and aerosol scenario.

How has the aerosol scenario been represented?
Special parameterization of profile

- In my intensive development period 2005-2007, just before network observations started, Frieß et al. (2006) suggested that the aerosol retrieval can strongly depend on the a priori, if the absolute value of AEC is used.

- In fact, lidar data at Tsukuba showed rapid temporal variations, but most aerosols were, in general, near the surface (70% of AOD in 0-1 km).

- At that time, some groups (incl. Heidelberg group) was using a simplified profile, represented by AOD and the mixing height.

- I decided to use a parameterization consisting of:
  - AOD (total column)
  - profile shape (e.g., $F_1$ is the fraction of AOD in 0-1 km)

How do we set a-priori values?
General characteristic of the retrieval, in terms of the a priori parameters used

• Aerosol retrieval
  – A-priori is based on lidar data at Tsukuba for 2006-2007.
  – The retrieval is not subject to the a priori of AOD, because the area* for AOD is almost unity. (*a rough measure of the fraction of the retrieval that comes from the measurements)
  – The retrieval of AEC at 0-1 km is less subject to the a priori of $F_1$, but the AECs above 1 km could be influenced to some extent.

• NO$_2$ retrieval
  – Profile shape factors are based on lidar data at Tsukuba for 2006-2007.
  – The a priori NO$_2$ column used is the 20% of the largest value in a set of NO$_2$ DSCDs for each retrieval. → The a priori varies with time.
  – The retrieval is not subject to the a priori of NO$_2$ column, because the area is almost unity.
  – The retrieval of NO$_2$ at 0-1 km is less subject to the a priori of $F_1$, but the NO$_2$ above 1 km could be influenced to some extent.

Profile is retrieved by scaling the given a-priori profile first, followed by changing the profile shape if enough information is contained.
Identical a priori setups have been used for all the network sites.
Advantages/disadvantages

• Advantages
  – The same a-priori assumptions for different sites are valid as the retrieval is not much dependent on the choice of a-priori.
    → better at development stage for network sites!
  – A thick vertical layer for the lowest layer would reduce the variation of $F_1$, making the use of constant $F_1$ valid for many cases.

• Disadvantages
  – Profile above ~2 km may not properly be represented.
  – Vertical resolution and averaging kernels (on altitude grid) are unavailable for each profile.
Initial data screening

• Data screening is performed,
  – if retrieved AOD is greater than 3 (the largest value in the LUTs)
    • Cloud screening to some extent
  – If the measured and fitted DSCDs do not agree to within 20%
    • Screening out poorly-converged cases
  – If the degrees of freedom for signal (DOFS) is less than 1.0
    • Screening out cases containing insufficient information

• Obviously, we need an additional cloud screening method for better statistics of data analysis.
  – Color index and water vapor, both of which can be derived from MAX-DOAS, were found to be very useful (Takashima et al., 2009).

We are now making comparison of our profile/column retrieval with those of other groups using CINDI data.
Evaluation of the profile retrieval is ongoing using CINDI data

- Our NO$_2$ VCDs are in good agreement with those from the other groups.
- Differences could reach $\sim$30%.

Courtesy of Folkard Wittrock
More on vertical profile of NO₂

- Our NO₂ VMR profile is in good agreement with those from the other groups.

What about the future plan?
Toward 8-component retrievals!

- **O\textsubscript{3}**
- **SO\textsubscript{2}**
- **HCHO**
- **CHOCHO**
- **H\textsubscript{2}O**

**FWHM = 0.4 nm**

**Under development**
CINDI → Japanese sites → Other sites
Summary

• Major revisions were made in last summer during the CINDI campaign. All the data have been reprocessed with the new version and provided to you.

• In brief, the profile is retrieved by scaling the given a-priori profile first, followed by changing the profile shape if enough information is contained.

• Our retrieval algorithm has disadvantages and advantages.

• Needs your feedbacks to characterize the retrieval more!