TROPOMI ATBD of the total and tropospheric NO$_2$ data products
Document approval record

This document was prepared by J.H.G.M. van Geffen, H.J. Eskes, K.F. Boersma and J.P. Veefkind. It was checked by S. Beirle, A. Richter, B. Sanders.
## Document change record

<table>
<thead>
<tr>
<th>issue</th>
<th>date</th>
<th>item</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.1</td>
<td>2012-08-17</td>
<td>All</td>
<td>Initial draft version</td>
</tr>
<tr>
<td>0.0.2</td>
<td>2012-09-12</td>
<td>All</td>
<td>Major reordering, adding text and references throughout</td>
</tr>
<tr>
<td>0.0.3</td>
<td>2012-09-26</td>
<td>6.1–6.2</td>
<td>Several small corrections and additions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1–7.2</td>
<td></td>
</tr>
<tr>
<td>0.1.0</td>
<td>2012-09-26</td>
<td></td>
<td>First official release</td>
</tr>
<tr>
<td>0.2.0</td>
<td>2012-11-15</td>
<td>5.1–5.4</td>
<td>Large number of updates throughout the text after internal reviewing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.1–6.2</td>
<td>by the TROPOMI Level-2 Working Group</td>
</tr>
<tr>
<td>0.2.1</td>
<td>2013-04-10</td>
<td>5, 6, 7</td>
<td>Document number corrected, reorganisation of Sect. 6; removal of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>discussion on an alternative retrieval approach; various minor corrections,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>updates and additions</td>
</tr>
<tr>
<td>0.2.2</td>
<td>2013-06-03</td>
<td>5, 6, 7</td>
<td>Major updates and further reorganisation of Sect. 5–7 first versions of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.9, 10</td>
<td>Sect. 8–10</td>
</tr>
<tr>
<td>0.3.0</td>
<td>2013-06-04</td>
<td></td>
<td>Release for Level-2 Working Group review</td>
</tr>
<tr>
<td>0.3.1</td>
<td>2013-06-19</td>
<td>5.1–5.3</td>
<td>Correction and additions resulting from v0.3.0 internal Level-2 Working</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2–6.6</td>
<td>Group review, and other minor corrections and additions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.3–7.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.4, 9.2</td>
<td></td>
</tr>
<tr>
<td>0.5.0</td>
<td>2013-06-20</td>
<td></td>
<td>Release for external review</td>
</tr>
<tr>
<td>0.5.1</td>
<td>2013-09-05</td>
<td>5.1–5.2</td>
<td>First round of updates taking into account comments and suggestions of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.1–6.6</td>
<td>reviewers of version 0.5.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.2, 7.5</td>
<td></td>
</tr>
<tr>
<td>0.5.2</td>
<td>2013-11-21</td>
<td>5.5</td>
<td>Further updates in view of the reviewers comments, and providing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2–6.7</td>
<td>more details regarding the fit procedure and the processing chain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1, 7.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2, 8.3</td>
<td></td>
</tr>
<tr>
<td>0.9.0</td>
<td>2013-11-27</td>
<td></td>
<td>Release to ESA and external reviewers</td>
</tr>
<tr>
<td>0.9.1</td>
<td>2014-03-17</td>
<td>4.</td>
<td>Instrument overview section now in separate document; section numbering</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>of this document unchanged to maintain change record</td>
</tr>
<tr>
<td>0.9.2</td>
<td>2014-04-07</td>
<td>6.2</td>
<td>Some small typographic corrections and updates</td>
</tr>
<tr>
<td>0.10.0</td>
<td>2014-04-15</td>
<td></td>
<td>Release to ESA</td>
</tr>
<tr>
<td>0.10.1</td>
<td>2014-09-19</td>
<td>6.1–6.6</td>
<td>Minor updates and corrections in text and tables; old Sects. 6.5 and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1–7.4</td>
<td>6.7 combined; Sects. 6.6 updated; references updated.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Descriptions updated and tables of input and output data added.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Notation of variables improved or clarified in view of the IODD [RD1].</td>
</tr>
<tr>
<td>0.10.2</td>
<td>2014-09-25</td>
<td>6.6</td>
<td>Table added with overview user applications and data the users need</td>
</tr>
<tr>
<td>0.11.0</td>
<td>2014-10-02</td>
<td></td>
<td>Release to ESA</td>
</tr>
<tr>
<td>0.11.1</td>
<td>2015-08-27</td>
<td>6.4.1</td>
<td>Update of AMF look-up table entries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.4.1</td>
<td>Update text regarding using cloud fraction from NO₂ spectral window</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6, 7.4</td>
<td>Minor updates in product data set tables, incl. dataset units</td>
</tr>
<tr>
<td>0.13.0</td>
<td>2015-09-14</td>
<td></td>
<td>Limited release to SSP Validation Team</td>
</tr>
<tr>
<td>0.13.1</td>
<td>2015-12-08</td>
<td>6.2, 7.5</td>
<td>Minor textual corrections</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1</td>
<td>Correction of units of input datasets</td>
</tr>
<tr>
<td>0.14.0</td>
<td>2015-12-11</td>
<td></td>
<td>Release to ESA</td>
</tr>
<tr>
<td>0.14.1</td>
<td>2016-01-21</td>
<td>6.4.2</td>
<td>Temperature correction equation updated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.6</td>
<td>Description of de-striping implementation clarified</td>
</tr>
</tbody>
</table>
## Document change record – continued

<table>
<thead>
<tr>
<th>issue</th>
<th>date</th>
<th>item</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0.0</td>
<td>2016-02-05</td>
<td>—</td>
<td>Public release</td>
</tr>
<tr>
<td>1.0.1</td>
<td>2017-01-31</td>
<td>All</td>
<td>Corrections and additions in response to internal review comments</td>
</tr>
<tr>
<td>1.0.2</td>
<td>2017-06-13</td>
<td>A – D</td>
<td>Appendices added</td>
</tr>
<tr>
<td>1.0.3</td>
<td>2017-07-13</td>
<td>All</td>
<td>Finalising the corrections and additions</td>
</tr>
<tr>
<td>1.1.0</td>
<td>2017-08-16</td>
<td>—</td>
<td>Updated public release for commissioning phase (E1)</td>
</tr>
<tr>
<td>1.1.1</td>
<td>2018-02-02</td>
<td>6.2</td>
<td>Formulation of the Ring term to match operational implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2</td>
<td>Improved description of the wavelength calibration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.3</td>
<td>New section with text from main section to improve readability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.6</td>
<td>Expanded description of a possible de-striping algorithm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.4</td>
<td>Updated detailed product overview table to match output product</td>
</tr>
<tr>
<td>1.1.2</td>
<td>2018-05-31</td>
<td>6.2</td>
<td>Updated to match the operational implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, D</td>
<td>Updated to match the operational implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>Appendix on the qa_value definition added</td>
</tr>
<tr>
<td>1.1.3</td>
<td>2018-06-08</td>
<td>5 – 9</td>
<td>Many textual corrections and improvements to match the operational</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>implementation and to incorporate results based on the TROPOMI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>measurements from the commissioning phase</td>
</tr>
<tr>
<td>1.2.0</td>
<td>2018-06-11</td>
<td>—</td>
<td>Release for operational phase (E2)</td>
</tr>
<tr>
<td>1.2.1</td>
<td>2018-10-15</td>
<td>—</td>
<td>Version with main text changes v1.1.0 to v1.2.0 marked for reviewers</td>
</tr>
<tr>
<td>1.2.2</td>
<td>2018-11-08</td>
<td>—</td>
<td>Figures 1, 3, 5, 7, 27 redone using TROPOMI data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.4</td>
<td>Text and Table updated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.6</td>
<td>Text updated and Table 2 added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.3.1</td>
<td>Text on updates in TM5-MP and Figure 8 added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.6</td>
<td>Text on de-striping expanded and Figure 20 added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6</td>
<td>Improved treatment of snow/ice cases described and Figure 24 added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.4</td>
<td>Tables 11 and 12 updated to include stripe amplitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2</td>
<td>Text updated and superfluous figure removed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>Updated qa_value for treatment of snow/ice cases (cf. Sect. 6.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Further textual updates</td>
</tr>
<tr>
<td>1.3.0</td>
<td>2018-11-08</td>
<td>—</td>
<td>Release</td>
</tr>
<tr>
<td>1.3.1</td>
<td>2019-01-31</td>
<td>6.4.4.1</td>
<td>Text on treatment of cloud data improved; Figure 10 added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6</td>
<td>Product overview Table 5 and product usage Table 6 updated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.4</td>
<td>Detailed product overview Table 11 updated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>Updated qa_value for $M_{\text{top}}/M_{\text{geo}}$ threshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Further minor textual updates and corrections</td>
</tr>
<tr>
<td>1.4.0</td>
<td>2019-02-06</td>
<td>—</td>
<td>Release</td>
</tr>
<tr>
<td>1.4.1</td>
<td>2019-04-15</td>
<td>6.2 &amp; F</td>
<td>Added description of removal of outliers from the DOAS fit residual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>Description on limiting cloud fraction to $[0 : 1]$ added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td>Updated qa_value for solar_eclipse</td>
</tr>
<tr>
<td>1.4.2</td>
<td>2019-10-02</td>
<td>—</td>
<td>Remarks and suggestions from A. Richter &amp; S. Beirle addressed</td>
</tr>
<tr>
<td>1.4.3</td>
<td>2019-10-03</td>
<td>6.2</td>
<td>Section structure modified to improve readability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>Section on Validation restructured</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Further textual updates and corrections</td>
</tr>
<tr>
<td>1.5.0</td>
<td>2019-10-03</td>
<td>—</td>
<td>Release to ESA for review</td>
</tr>
<tr>
<td>2.0.0</td>
<td>—</td>
<td>—</td>
<td>Not released: no processor update implemented</td>
</tr>
</tbody>
</table>
## Document change record – continued

<table>
<thead>
<tr>
<th>issue</th>
<th>date</th>
<th>item</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0.1</td>
<td>2020-02-19</td>
<td>—</td>
<td>Textual updates in view of processor version changes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.4</td>
<td>Usage of ECMWF snow/ice flag &amp; impact of FRESCO update</td>
</tr>
<tr>
<td>2.0.2</td>
<td>2020-03-10</td>
<td>—</td>
<td>Textual updates addressing remarks from ESA on v1.5.0</td>
</tr>
<tr>
<td>2.0.3</td>
<td>2021-01-21</td>
<td>—</td>
<td>Minor updates and corrections &amp; digital signatures (page 2) removed</td>
</tr>
<tr>
<td>2.0.4</td>
<td>2021-02-24</td>
<td>5.1</td>
<td>General NO₂ description expanded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.6</td>
<td>Version history expanded, including note on data change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4</td>
<td>Several subsections finetuned and expanded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.4</td>
<td>Subsection on the FRESCO-wide update added</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6</td>
<td>Product description and usage updated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1–4</td>
<td>Feasibility expanded and updated where needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.1–4</td>
<td>Validation discussion updated where needed</td>
</tr>
<tr>
<td>2.1.0</td>
<td>2021-02-24</td>
<td>—</td>
<td>Release to ESA for review</td>
</tr>
<tr>
<td>2.1.1</td>
<td>2021-04-15</td>
<td>All</td>
<td>Minor updates after review and other textual improvements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.4</td>
<td>Section on cloud data reorganised and somewhat extended</td>
</tr>
<tr>
<td>2.1.2</td>
<td>2021-04-20</td>
<td>9.1–4</td>
<td>Validation sections reorganised and shortened</td>
</tr>
<tr>
<td>2.1.3</td>
<td>2021-06-15</td>
<td>—</td>
<td>Details finalised</td>
</tr>
<tr>
<td>2.2.0</td>
<td>2021-06-16</td>
<td>—</td>
<td>Release processor versions 1.4.0 &amp; 2.2.0</td>
</tr>
<tr>
<td>2.2.1</td>
<td>2022-01-27</td>
<td>All</td>
<td>Minor updates and textual corrections</td>
</tr>
<tr>
<td>2.2.2</td>
<td>2022-06-08</td>
<td>6.4.4.2</td>
<td>Description of new TROPOMI DLER v1 surface albedo usage in the FRESCO cloud and NO₂ data retrieval</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4.5</td>
<td></td>
</tr>
<tr>
<td>2.4.0</td>
<td>2022-07-11</td>
<td>—</td>
<td>Release processor versions 2.4.0</td>
</tr>
</tbody>
</table>
Contents

Document approval record .................................................................................................................. 2
Document change record ..................................................................................................................... 3
List of Tables ....................................................................................................................................... 7
List of Figures ..................................................................................................................................... 8

1 Introduction to the document ........................................................................................................ 9
  1.1 Identification ................................................................................................................................. 9
  1.2 Purpose and objective .................................................................................................................... 9
  1.3 Document overview ...................................................................................................................... 9
  1.4 Acknowledgements ....................................................................................................................... 9

2 Applicable and reference documents ............................................................................................. 10
  2.1 Applicable documents .................................................................................................................. 10
  2.2 Standard documents .................................................................................................................... 10
  2.3 Reference documents ................................................................................................................... 10
  2.4 Electronic references .................................................................................................................. 11

3 Terms, definitions and abbreviated terms ....................................................................................... 13
  3.1 Terms and definitions ................................................................................................................... 13
  3.2 Acronyms and abbreviations ....................................................................................................... 13

4 TROPOMI instrument description .................................................................................................. 14

5 Introduction to the TROPOMI NO₂ data products ...................................................................... 15
  5.1 Nitrogen dioxide in troposphere and stratosphere ...................................................................... 15
  5.2 NO₂ satellite retrieval heritage .................................................................................................. 17
  5.3 Separating stratospheric and tropospheric NO₂ with a data assimilation system ...................... 19
  5.4 NO₂ data product requirements ................................................................................................. 20
  5.5 NO₂ retrieval for TROPOMI ....................................................................................................... 20
  5.6 NO₂ data product: version history and access ........................................................................... 21

6 Algorithm description ....................................................................................................................... 23
  6.1 Overview of the NO₂ retrieval algorithm ..................................................................................... 23
  6.2 Spectral fitting ............................................................................................................................. 23
  6.2.1 Wavelength calibration & common wavelength grid ............................................................. 25
  6.2.2 Minimising the chi-squared merit function ............................................................................. 26
  6.2.3 Reference spectra .................................................................................................................... 27
  6.2.4 DOAS fit details for OMI and TROPOMI ............................................................................... 27
  6.2.5 Some notes regarding other DOAS implementations ........................................................... 28
  6.3 Separation of stratospheric and tropospheric NO₂ .................................................................... 29
  6.3.1 Stratospheric chemistry in the TM5-MP model .................................................................... 31
  6.4 Air-mass factor and vertical column calculations ........................................................................ 31
  6.4.1 Altitude dependent AMFs ....................................................................................................... 33
  6.4.2 Temperature correction ............................................................................................................ 34
  6.4.3 Correction for cloud cover ....................................................................................................... 36
  6.4.4 Cloud cover and cloud pressure data ....................................................................................... 36
  6.4.4.1 The FRESCO-S cloud pressure & NO₂ cloud fraction ..................................................... 36
  6.4.4.2 Cloud pressure: the FRESCO-wide update ..................................................................... 37
  6.4.4.3 The O₂–O₃ cloud pressure .................................................................................................. 41
  6.4.4.4 Other cloud data products ................................................................................................. 41
  6.4.5 Surface albedo ........................................................................................................................ 41
  6.4.5.1 Processor versions v1.0.2 up to v2.3.1 ................................................................................. 41
  6.4.5.2 Processor version v2.4.0 ..................................................................................................... 41
  6.4.6 Snow and ice cover ................................................................................................................ 43
  6.4.7 Surface pressure ....................................................................................................................... 44
  6.4.8 A-priori vertical NO₂ profiles .................................................................................................. 45
  6.4.9 Averaging kernels .................................................................................................................... 46
  6.4.10 De-striping the NO₂ data product ......................................................................................... 46
  6.5 Processing chain elements ........................................................................................................ 48
6.5.1 Off-line (re)processing ................................................................. 48
6.5.2 Near-real time processing ............................................................. 49
6.6 The NO₂ data product ................................................................. 50
7 Input-Output file description .......................................................... 54
  7.1 Required input ................................................................. 54
  7.1.1 Inputs at the PDGS for spectral fitting and air-mass factor calculation 54
  7.1.2 Inputs at the IDAF for the data assimilation ................................... 55
  7.2 Computational effort ................................................................. 55
  7.3 Near-real time timeliness ........................................................... 57
  7.4 NO₂ product description and size ................................................ 57
8 Error analysis ................................................................. 60
  8.1 Slant column errors ................................................................. 60
  8.2 Errors in the stratospheric (slant) columns ................................... 61
  8.3 Errors in the tropospheric air-mass factors ..................................... 62
  8.4 Total errors in the tropospheric NO₂ columns .................................. 63
9 Validation ................................................................. 66
  9.1 Routine validation & validation activities ....................................... 66
  9.2 Algorithm testing and verification ............................................... 66
  9.3 Stratospheric NO₂ validation ....................................................... 66
  9.4 Tropospheric NO₂ validation ....................................................... 67
10 Conclusion ................................................................. 69
A Wavelength calibration ............................................................... 70
A.1 Description of the problem ........................................................ 70
A.2 Non-linear model function and Jacobian ....................................... 70
A.2.1 Prior information for the optimal estimation fit ......................... 72
A.3 Application of the wavelength calibration in NO₂ ................................ 72
B High-sampling interpolation .......................................................... 73
C Effective cloud fraction in the NO₂ window ..................................... 74
C.1 Adjusting albedo to respect physical limits to the cloud fraction .......... 75
D Surface albedo correction using the snow/ice flag ............................. 76
E Data quality value: the qa_value flags ............................................. 77
F Spike removal in the DOAS fit ........................................................ 78
G References .................................................................................. 80

List of Tables

1 NO₂ data product requirements ....................................................... 20
2 NO₂ processor version overview .................................................... 21
3 Settings of DOAS retrieval of NO₂ .................................................. 26
4 AMF LUT ................................................................................. 35
5 Final NO₂ vertical column data product .......................................... 51
6 Data product user applications ....................................................... 52
7 Dynamic input data ....................................................................... 55
8 Static input data ............................................................................ 56
9 Computational effort: off-line processing ........................................ 56
10 Computational effort: NRT processing ........................................... 56
11 Data product list of main output file .............................................. 58
12 Data product list of support output file ........................................... 59
13 Estimate of AMF errors ............................................................... 63
14 Tropospheric AMF uncertainty estimates from OMI ....................... 65
15 A-priori values for the wavelength fit ............................................. 72
16 Look-up tables for reflectance calculations ...................................... 75
17 Snow/ice flags ............................................................................ 76
18 Data quality value determination ................................................... 77
List of Figures

1. Tropospheric NO₂ for April 2018 ................................................................. 15
2. COVID-19 lockdown impact on NO₂ in India ........................................... 16
3. Stratospheric NO₂ for 1 April 2018 ............................................................ 17
4. NO₂ data record UV/Vis satellite instruments ........................................... 18
5. DOAS fit ...................................................................................................... 25
6. Reference spectra ...................................................................................... 28
7. NO₂ forecast and analysis differences ...................................................... 32
8. High-latitude improvement ........................................................................ 33
9. Temperature correction factors ................................................................. 34
10. Example of improvements in cloud treatment ......................................... 38
11. Explanation of the FRESCO-wide approach .............................................. 38
12. FRESCO case study, 23 Jan 2019 ............................................................. 39
13. TROPOMI vs OMI NO₂ over China, Winter 2020/21 ............................... 39
14. FRESCO cloud fraction change due to S5P DLER ................................. 40
15. Cloud fraction method comparison ............................................................ 40
16. S5P DLER at 758 nm .............................................................................. 42
17. Change NO₂ column due to S5P DLER .................................................. 43
18. Albedo adjustment in version 2 compared to version 1 ............................ 44
19. Tropospheric NO₂ difference from resolution ......................................... 45
20. De-striping example ................................................................................ 47
21. Comparison of vertical columns ............................................................... 48
22. Scheme of the TROPOMI processing system .......................................... 49
23. Scheme of the TROPOMI processing system in NRT ............................... 50
24. Enhanced coverage over snow/ice ............................................................ 53
25. Error in slant column versus SNR ............................................................. 61
26. Comparison of slant column errors .......................................................... 61
27. Tropospheric column and error estimates from TROPOMI ..................... 64
28. Tropospheric column comparisons OMI-TROPOMI ............................... 67
29. High sampling interpolation on part of a solar observation ...................... 73
30. Examples of outliers in fit residual ............................................................ 78
31. Examples of outliers and saturation .......................................................... 79
1 Introduction to the document

1.1 Identification

This document, identified as S5P-KNMI-L2-0005-RP, is the Algorithm Theoretical Basis Document (ATBD) for the TROPOMI total and tropospheric NO$_2$ data products. It is part of a series of ATBDs describing the TROPOMI Level-2 data products. The latest public release version of the ATBD is available via [ER1].

This ATBD describes NO$_2$ processor version 2.4.0 and changes to the processor up to this version.

An overview of which NO$_2$ processor version is used for processing which TROPOMI orbits is given in Table 2.

Additional documents related to the TROPOMI NO$_2$ data products:

- Product User Manual (PUM), identified as S5P-KNMI-L2-0021-MA, available via [ER2].
- Product ReadMe File (PRF), identified as S5P-MPC-KNMI-PRF-NO2, available via [ER3].
- Product User Manual (PUM) for the TM5 NO$_2$, SO$_2$ and HCHO profile auxiliary support product, identified as S5P-KNMI-L2-0035-MA, available via [ER2].
- Quarterly Validation Report (ROCVR), identified as S5P-MPC-IASB-ROCVR, available via [ER4].

SSP/TROPOMI product and algorithm documents are also available via [ER5].

1.2 Purpose and objective

The purpose of this document is to describe the theoretical basis and the implementation of the NO$_2$ Level-2 algorithm for TROPOMI. The document is maintained during the development phase and the lifetime of the data products. Updates and new versions will be issued in case of changes of the algorithm.

1.3 Document overview

Sections 2 and 3 list the applicable and reference documents and the terms and abbreviations specific for this document; references to peer-reviewed papers and other scientific publications are listed in Appendix G. Section 4 provides a reference to a general description of the TROPOMI instrument, which is common to all ATBDs of the TROPOMI Level-2 data products. Section 5 provides an introduction to the NO$_2$ data products, their heritage, the set-up of their retrieval, the requirements of the products, and their availability. Section 6 gives an overview of the TROPOMI NO$_2$ data processing system and important aspects of the various steps in the processing. Section 7 lists some aspects regarding the feasibility of the NO$_2$ data products, such as the computational effort and the auxiliary information needed for the processing. Section 8 deals with an error analysis of the NO$_2$ data product. Section 9 gives a brief overview of validation issues and possibilities, such as campaigns and satellite intercomparisons. Section 10 formulates some conclusion regarding the NO$_2$ data products.

1.4 Acknowledgements

The authors would like to thank the following people for useful discussions, information, reviews of earlier versions of this document and other contributions: Andreas Hilboll, Andreas Richter, Angelika Dehn, Bram Maasakkers, Bram Sanders, Deborah Stein – Zweers, Dominique Brunner, Huan Yu, Isabelle De Smedt, Jason Williams Johan de Haan, Lidia Saavedra De Miguel, Maarten Sneep, Marina Zara, Mark ter Linden, Michel Van Roozendael, Piet Stammes, Pieter Valks, Ronald van der A, Steffen Beirle, Thomas Wagner.
2 Applicable and reference documents

2.1 Applicable documents

[AD1] Level 2 Processor Development – Statement of Work.
  source: ESA/ESTEC; ref: S5P-SW-ESA-GS-053; issue: 1.1; date: 2012-05-21.

  source: ESA/ESTEC; ref: S5p-RS-ESA-SY-0002; issue: 4.1; date: 2011-04-xx.

2.2 Standard documents

There are no standard documents.

2.3 Reference documents

[RD1] Sentinel 5 precursor/TROPOMI KNMI and SRON level 2 Input Output Data Definition.
  source: KNMI; ref: S5P-KNMI-L2-0009-SD; issue: 11.0.0; date: 2019-02-01.

[RD2] Terms, definitions and abbreviations for TROPOMI L01b data processor.
  source: KNMI; ref: S5P-KNMI-L01B-0004-LI; issue: 3.0.0; date: 2013-11-08.

[RD3] Terms and symbols in the TROPOMI Algorithm Team.
  source: KNMI; ref: S5P-KNMI-L2-0049-MA; issue: 2.0.0; date: 2016-05-17.

[RD4] TROPOMI ATBD of the UV aerosol index.
  source: KNMI; ref: S5P-KNMI-L2-0008-RP; issue: 1.1.0; date: 2018-06-15.

  source: ESA/ESTEC; ref: EOP-SMA/1507/JL-dr; issue: 3; date: 2011-09-21.

  source: KNMI; ref: EU-project 607405, SPA.2013.1.1-03; date: November 2012.

  source: KNMI, SRON; ref: RS-TROPOMI-KNMI-017; issue: 2.0.0; date: 2008-10-30.


  source: KNMI; ref: RP-CAM-KNMI-050; date: Nov. 2009.

  source: KNMI; ref: RP-ONTRAQ-KNMI-051; date: Jan. 2010.

  source: ESA; ref: ESA-EOPG-CSCOP-PL-0073; issue: 1; date: 2017-11-6.

[RD12] NO2 PGE Detailed Processing Model.
  source: Space Sytems Finland; ref: TN-NO2-0200-SSF-001; issue: 1.2; date: 2010-04-21.

  source: KNMI; ref: S5P-KNMI-L01B-0009-SD; issue: 8.0.0; date: 2017-06-01.

[RD14] S5P/TROPOMI Static input for Level 2 processors.
  source: KNMI/SRON/BIRA/DLR; ref: S5P-KNMI-L2CO-0004-SD; issue: 4.0.0; date: 2016-03-21.

  source: KNMI; ref: EU-project 607405, SPA.2013.1.1-03; date: April 2016.
[RD16] An improved temperature correction for OMI NO$_2$ slant column densities from the 405-465 nm fitting window.


source: EUMETSAT/KNMI; ref: EUM/CO/09/4600000655/RM; issue: 1.3; date: 2010-10-18.


source: KNMI; ref: KNMI-ESA-SSL2PP-ATBD-005; issue: 3.1; date: 2019-05-02.


source: DLR; ref: SSP-DLR-L2-ATBD-4001; issue: 2.2.0; date: 2020-06-15.


source: KNMI; ref: S5P-KNMI-L3-0301-RP; issue: 1.2.0; date: 2022-01-13.


source: KNMI; ref: OMI_NO2_HE5_2.0_2011; date: 18 August 2011.

[RD22] Product user manual for the TM5 NO$_2$, SO$_2$ and HCHO profile auxiliary support product.

source: KNMI; ref: S5P-KNMI-L2-0035-MA; issue: 1.0.0; date: 2021-02-04.

[RD23] Preparing elevation data for Sentinel 5 precursor.

source: KNMI; ref: S5P-KNMI-L2-0121-TN; issue: 2.0.0; date: 2015-09-11.


source: IUP; ref: S5P-IUP-L2-ScVR-RP; issue: 2.1; date: 2015-12-22.


source: KNMI; ref: S5P-KNMI-L2-0126-TN; issue: 1.0.0; date: 2015-09-11.

[RD26] Determine the effective cloud fraction for a specific wavelength.

source: KNMI; ref: S5P-KNMI-L2-0115-TN; issue: 2.0.0; date: 2019-04-10.

2.4 Electronic references


3 Terms, definitions and abbreviated terms

Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L0-1b data processor are described in [RD2]. Terms, definitions and abbreviated terms that are used in development program for the TROPOMI L2 data processors are described in [RD3]. Terms, definitions and abbreviated terms that are specific for this document can be found below.

3.1 Terms and definitions

The most important symbols related to the data product described in this document – some of which are not in [RD3] – are the following; see also the data product overview list in Table 11.

\[
\begin{align*}
M & \quad \text{total air-mass factor} \\
M_{\text{cl}} & \quad \text{cloudy air-mass factor} \\
M_{\text{clc}} & \quad \text{clear-sky air-mass factor} \\
M_{\text{trop}} & \quad \text{tropospheric air-mass factor} \\
M_{\text{strat}} & \quad \text{stratospheric air-mass factor} \\
N_{\text{s}} & \quad \text{total slant column density} \\
N_{\text{trop}} & \quad \text{tropospheric slant column density} \\
N_{\text{strat}} & \quad \text{stratospheric slant column density} \\
N_{\text{v}} & \quad \text{total vertical column density} \\
N_{\text{trop}} & \quad \text{tropospheric vertical column density} \\
N_{\text{strat}} & \quad \text{stratospheric vertical column density} \\
N_{\text{sum}} & \quad \text{sum of tropospheric and stratospheric vertical column density}
\end{align*}
\]

3.2 Acronyms and abbreviations

AAI Absorbing Aerosol Index
ACE Atmospheric Chemistry Experiment
AMF Air-mass factor
CAMS Copernicus Atmosphere Monitoring Service
CTM Chemistry Transport Model
DAK Doubling-Adding KNMI
DLER Directionally dependent Lambertian equivalent reflectivity
DEM Digital Elevation Map
DOAS Differential Optical Absorption Spectroscopy
DOMINO Dutch OMI NO\textsubscript{2} data products of KNMI for OMI
ECMWF European Centre for Medium-Range Weather Forecast
ENVISAT Environmental Satellite
EOS-Aura Earth Observing System (Chemistry & Climate Mission)
ERBS Earth Radiation Budget Satellite
ERS European Remote Sensing satellite
FRESCO Fast Retrieval Scheme for Clouds from the Oxygen A band
GOME Global Ozone Monitoring Experiment
HALOE Halogen Occultation Experiment
IDAF-L2 Instrument Data Analysis Facility, Level 2 (at KNMI)
IPA Independent pixel approximation
ISRF Instrument Spectral Response Function (aka slit function)
LER Lambertian equivalent reflectivity
LUT Look-up table
4 TROPOMI instrument description

A description of the TROPOMI instrument and performance, referred to from all ATBDs, can be found in [RD4]. See also the overview paper of Veefkind et al. [2012].
5 Introduction to the TROPOMI NO\textsubscript{2} data products

5.1 Nitrogen dioxide in troposphere and stratosphere

Nitrogen dioxide (NO\textsubscript{2}) and nitrogen oxide (NO) – together usually referred to as nitrogen oxides (NO\textsubscript{x} = NO + NO\textsubscript{2}) – are important trace gases in the Earth’s atmosphere, present in both the troposphere and the stratosphere. They enter the atmosphere as a result of anthropogenic activities (notably fossil fuel combustion and biomass burning) and natural processes (such as microbiological processes in soils, wildfires and lightning). Approximately 95% of the NO\textsubscript{x} emissions is in the form of NO. During daytime, i.e. in the presence of sunlight, a photochemical cycle involving ozone (O\textsubscript{3}) converts NO into NO\textsubscript{2} (and vice versa) on a timescale of minutes, so that NO\textsubscript{2} is a robust measure for concentrations of nitrogen oxides (Solomon [1999], Jacob [1999]).

In the troposphere NO\textsubscript{2} plays a key role in air quality issues, as it directly affects human health [World Health Organisation, 2003]. In addition nitrogen oxides are essential precursors for the formation of ozone in the troposphere (e.g. Sillman et al. [1990]) and they influence concentrations of OH and thereby (shorten) the lifetime of methane (CH\textsubscript{4}) (e.g. Fuglestvedt et al. [1999]). Although NO\textsubscript{2} is a minor greenhouse gas in itself, the indirect effects of NO\textsubscript{2} on global climate change are probably larger, with a presumed net cooling effect mostly driven by a growth in aerosol concentrations through nitrate formation from nitrogen oxides and enhanced levels of oxidants (e.g. Shindell et al. [2009]). Deposition of nitrogen is of great importance for eutrophication [Dentener et al., 2006], the response of the ecosystem to the addition of substances such as nitrates and phosphates – negative environmental effects include the depletion of oxygen in the water, which induces reductions in fish and other animal populations.

For typical levels of OH the lifetime of NO\textsubscript{x} in the lower troposphere is less than a day. For Riyadh, for example, Beirle et al. [2011] find a lifetime of about 4.0 ± 0.4 hours, while at higher latitudes (e.g. Moscow) the lifetime can be considerably longer, up to 8 hour in winter, because of a slower photochemistry in that season. For Switzerland Schaub et al. [2007] report lifetimes of 3.6 ± 0.8 hours in summer and 13.1 ± (3.8) hours in winter. With lifetimes in the troposphere of only a few hours, the NO\textsubscript{2} will remain relatively close to its source, making the NO\textsubscript{x} sources well detectable from space. As an example, Fig. 1 shows distinct hotspots of NO\textsubscript{2} pollution over the highly industrialised and urbanised regions of London, Rotterdam and the Ruhr area in the monthly average tropospheric NO\textsubscript{2} for April 2018 over Europe derived from TROPOMI data.

Since July 2018, with the first public release of the TROPOMI datasets including NO\textsubscript{2}, the number of TROPOMI users and publications has grown strongly. A review of these applications is beyond the scope of this ATBD. Topics addressed range from changes in global-scale NO\textsubscript{2} distributions and impacts on atmospheric

![TROPOMI NO2 tropospheric column, April 2018](image-url)

Figure 1: Monthly average distribution of tropospheric NO\textsubscript{2} columns for April 2018 over Europe based on TROPOMI data, derived with processor version 1.2.0.
Figure 2: Strong reduction of NO\textsubscript{2} over India as a result of the COVID-19 lockdown in March-April 2020 (lower panel) compared to 2019 (top panel). Concentrations were reduced strongly in cities like Delhi, Mumbai, Dhaka. In contrast, some of the coal-fired power plants continued the electricity production with only minor reductions. Source: https://www.esa.int/, news story 24 April 2020.

chemistry, data assimilation applications, validation of regional air quality models and NO\textsubscript{x} emission inversion studies. In particular the combination of the high spatial resolution, the large signal-to-noise ratio and daily global coverage makes TROPOMI unique. This has been and will be further used for the analysis of emissions and concentrations at the local scale, for individual cities, power plants, industrial complexes, road traffic and shipping lanes. The power of TROPOMI is nicely demonstrated by the observation of pollution plumes from individual ships [Georgoulias et al., 2020].

In 2020 the number of publications and attention in the media for TROPOMI NO\textsubscript{2} observations has exploded. As a result of the COVID-19 related lockdowns pollution levels have dropped dramatically as observed in real-time by the TROPOMI instrument, largely consistent with surface observations [Gkatzelis et al., 2021]. This clearly demonstrates the value of real-time global monitoring of concentrations from space. Fig. 2 shows, as example, COVID-19 lockdown impact on NO\textsubscript{2} in India.

In the stratosphere NO\textsubscript{2} is involved in some photochemical reactions with ozone and thus affects the ozone layer (Crutzen et al. [1970]; Seinfeld and Pandis [2006]). NO\textsubscript{2} in the stratosphere originates mainly from oxidation of N\textsubscript{2}O in the middle stratosphere, which leads to NO\textsubscript{x}, which in turn acts as a catalyst for ozone destruction (Crutzen et al. [1970]; Hendrick et al. [2012]). But NO\textsubscript{x} can also suppress ozone depletion by converting reactive chlorine and hydrogen compounds into unreactive reservoir species (such as ClONO\textsubscript{2} and HNO\textsubscript{3}; Murphy et al. [1993]).

Fig. 3 shows, as an example, the stratospheric NO\textsubscript{2} distribution derived from TROPOMI measurements on 1 April 2018 at the 13:30 overpass local time. The image shows variability related to atmospheric transport and diurnal variability in the stratosphere. In a study into the record ozone loss, triggered by enhanced NO\textsubscript{2} levels, in the exceptionally strong Arctic polar vortex in Spring 2011, Adams et al. [2013] showed the usefulness of such data when investigating the anomalous dynamics and chemistry in the stratosphere. With its high spatial resolution and signal-to-noise ratio, TROPOMI is clearly well-suited to help understand the stratospheric NO\textsubscript{2} content and its implications for the ozone distribution.
Figure 3: Distribution of stratospheric NO$_2$ on 1 April 2018 along the individual TROPOMI orbits, derived with processor version 1.2.0. The image shows that atmospheric dynamics creates variability in the stratospheric columns, mainly at mid-latitudes. Furthermore we can see the effect of the increase of NO$_2$ in the stratosphere during daytime leading to small jumps from one orbit to the next. Note that the colour scale range is different from the range in Fig. 1.

From observed trends in N$_2$O emissions one would expect a trend in stratospheric NO$_2$ with potential implications for persistent ozone depletion well into the 21st century [Ravishankara et al., 2009]. There have been some reports of such trends in stratospheric NO$_2$, for instance from New Zealand [Liley et al., 2000] and northern Russia [Gruzdev and Elokhov, 2009]. On the other hand, Hendrick et al. [2012] report that changes in the NO$_x$ partitioning in favour of NO may well conceal the effect of trends in N$_2$O. TROPOMI continues the important record of stratospheric NO$_2$ observations that started with GOME in 1995, and improves the detectability of trends.

Over unpolluted regions most NO$_2$ is located in the stratosphere (typically more than 90%). For polluted regions 50–90% of the NO$_2$ is located in the troposphere, depending on the degree of pollution. Over polluted regions, most of the tropospheric NO$_2$ is found in the planetary boundary layer, as has been shown among others in campaigns using measurements made from aeroplanes, such as INTEX (e.g. Hains et al. [2010]). In areas with strong convection, enhanced NO$_2$ concentrations are observed at higher altitudes due to production of NO$_x$ by lightning (e.g. Ott et al. [2010]; Allen et al. [2021]).

The important role of NO$_2$ in both troposphere and stratosphere implies that it is not only important to know the total column density of NO$_2$, but rather the tropospheric NO$_2$ and stratospheric NO$_2$ concentrations separately. A proper separation between the two is therefore important, in particular for areas with low pollution, where the stratospheric concentration forms a significant part of the total column.

### 5.2 NO$_2$ satellite retrieval heritage

Tropospheric concentrations of NO$_2$ are monitored all over the world by a variety of remote sensing instruments – ground-based, in-situ (balloon, aircraft) or satellite-based – each with its own specific advantages, and to some extent still under development.

Stratospheric NO$_2$ has been measured by a number of satellite instruments since the 1980s, such as the spectrometer aboard SME (1981-1989; Mount et al. [1984]), SAGE-II/III (ERBS/Meteor-3M, 1984-2005;
Figure 4: Overview of the European UV/Vis polar orbiting and geostationary backscatter satellite instruments capable of retrieving tropospheric and stratospheric NO$_2$ column data since the launch of GOME aboard ERS-2, including missions to be launched in the near future.


Over the past 22 years tropospheric NO$_2$ has been measured from UV/Vis backscatter satellite instruments such as GOME (ERS-2, 1995–2011; Burrows et al. [1999]), SCIAMACHY (ENVISAT, 2002–2012; Bovensmann et al. [1999]), OMI (EOS-Aura, 2004–present; Levelt et al. [2006]), the GOME-2 instruments [Munro et al., 2006] aboard MetOp-A (2007–present), MetOp-B (2012–present) and MetOp-C (2019–present), the OMPS instrument [Yang et al., 2014] on the Suomi NPP platform (2011–present) and the NOAA-20 satellite (2017–present). TROPOMI (see [RD4]; Veefkind et al. [2012]) extends the records of these observations, and in turn will be followed up by several forthcoming instruments including Sentinel 5 and the geostationary platforms GEMS (Bak [2013], Kim [2020]; launched in 2020), TEMPO [Zoogman et al., 2017] and Sentinel 4 [Ingmann et al., 2012], [RD5]. Fig. 4 shows the timelines of the NO$_2$ data records of some of these instruments. Note that TROPOMI, OMI, the GOME-2 instruments and Sentinel-5 provide (near-)global coverage in one day, and that Sentinel-4 is a geostationary instrument.

For the UV/Vis backscatter instruments that observe NO$_2$ down into the troposphere, KNMI has operated – in close collaboration with BIRA-IASB, NASA and DLR – a real-time data processing system, the results of which are freely available via the TEMIS website [ER6]. The data has been used for a variety of studies in areas like validation (see e.g. Boersma et al. [2009], Hains et al. [2010], Lamsal et al. [2010]), trends (see e.g. Van der A et al. [2008], Stavrakou et al. [2008], Dirksen et al. [2011], Castellanos and Boersma [2012], DeRuyter et al. [2012]), and NO$_x$ emission and lifetime estimates (see e.g. Lin et al. [2010], Beirle et al. [2011], Mijling and Van der A [2012], Wang et al. [2012]).

The DOMINO approach for OMI (and the similar approach called TM4NO2A for GOME, SCIAMACHY and GOME-2) is based on a DOAS retrieval, a pre-calculated air-mass factor (AMF) look-up table and a data assimilation / chemistry transport model for the separation of the stratospheric and tropospheric contributions to the NO$_2$ column (see Sect. 6 for details). The differences between the processing systems for the different instruments are small and related to instrument issues, such as available spectral coverage and wavelength calibration, other absorbing trace gases fitted along, and details of the cloud cover data retrieval.

The European Quality Assurance for Essential Climate Variables (QA4ECV) project ([RD6], [ER7], Boersma et al. [2018]) has led to a homogeneous reprocessing dataset of NO$_2$ for the sensors GOME, SCIAMACHY, OMI and GOME-2A. This project has investigated and improved all the individual steps/modules in the NO$_2$ retrieval. The new NO$_2$ datasets are available via the QA4ECV project website at [ER8]. This new release replaces the DOMINO-v2 OMI NO$_2$ dataset and TM4NO2A datasets for the other sensors. Due to IT equipment issues the OMI/QA4ECV dataset ends on 29 March 2021; a follow-up dataset, based on new collection-4 OMI data, with reprocessing of the full mission, is currently being set up.

The TROPOMI NO$_2$ processor includes many of the developments from the QA4ECV project, including improvements in the TM5-MP/DOMINO chemistry modelling-retrieval-assimilation approach, DOAS optimisations (cf. Van Geffen et al. [2020]) and air-mass factor lookup table. On top of that, several further improvements...
5.3 Separating stratospheric and tropospheric NO\textsubscript{2} with a data assimilation system

The NO\textsubscript{2} processing system starts with a DOAS retrieval step that determines the NO\textsubscript{2} slant column density, which represents the total amount of NO\textsubscript{2} along the line of sight, i.e. from sun via earth’s atmosphere to satellite. To determine the tropospheric NO\textsubscript{2} slant column density, the stratospheric NO\textsubscript{2} slant column density is subtracted from the total slant column, after which the tropospheric sub-column is converted to the tropospheric vertical NO\textsubscript{2} column.

Several approaches to estimate the stratospheric NO\textsubscript{2} amount have been introduced in the past. The TM5-MP/DOMINO approach uses information from a chemistry transport model by way of data assimilation to simulate the instantaneous stratospheric NO\textsubscript{2} distribution and to force consistency between the stratospheric NO\textsubscript{2} column and the satellite measurement [Boersma et al., 2004]. Other methods applied elsewhere include the following (in arbitrary order).

\begin{itemize}
\item[a)] The \textit{wave analysis method} uses subsets of satellite measurements over unpolluted areas to remove known areas of pollution, i.e. areas with potentially large amounts of tropospheric NO\textsubscript{2} from a 24-hour composite of the satellite measured NO\textsubscript{2} and expands the remainder with a planetary wave analysis across the whole stratosphere, followed where necessary by a second step to mask pollution events (e.g. Bucsela et al. [2006]). This approach has been used between 2004 and 2012 for the OMI NO\textsubscript{2} Standard Product (SP) of NASA/KNMI.
\item[b)] The \textit{reference sector method} method uses a north-to-south region over the Pacific Ocean that is assumed to be free of tropospheric NO\textsubscript{2}, as there are no (surface) sources of NO\textsubscript{2}, so that all NO\textsubscript{2} measured is assumed to be in the stratosphere (e.g. Richter and Burrows [2002], Martin et al. [2002]). This stratospheric NO\textsubscript{2} is then assumed to be valid in latitudinal bands for all longitudes. In some implementations this method is extended with a spatial filtering to include other relatively clean areas across the world (e.g. Bucsela et al. [2006], Valks et al. [2011]).
\item[c)] \textit{Image processing techniques} assume that the stratospheric NO\textsubscript{2} shows only smooth and low-amplitude latitudinal and longitudinal variations (e.g. Leue et al. [2001], Wenig et al. [2003]). This approach will probably miss the finer details in the stratospheric NO\textsubscript{2} distribution (as is the case for methods a and b above). The next version of NASA’s OMI NO\textsubscript{2} SP will use a similar approach [Bucsela et al., 2013].
\item[d)] \textit{Independent stratospheric NO\textsubscript{2} data}, such as collocated limb measurements (e.g. Beirle et al. [2010], Hilboll et al. [2013b]) or data taken from a chemistry transport model (e.g. Hilboll et al. [2013a]), can be subtracted from the total (slant) column measurements to find the tropospheric NO\textsubscript{2} concentrations. Unfortunately, limb collocated stratospheric measurements are not available for satellite retrievals from the GOME(-2), OMI, and TROPOMI sensors. Nevertheless this approach is potentially very useful for comparison and validation studies. Possible cross-calibration problems between the stratospheric and the total measurements would complicate the approach.
\item[e)] The \textit{STRatospheric Estimation Algorithm from Mainz} (STREAM; Beirle et al. [2016]). The STREAM approach is based on the total column measurements over clean, remote regions as well as over clouded scenes where the tropospheric column is effectively shielded. STREAM is a flexible and robust interpolation algorithm and does not require input from chemical transport models. It was developed as a verification algorithm for the then upcoming satellite instrument TROPOMI, as a complement to the operational stratospheric correction based on data assimilation. STREAM was successfully applied to the UV/vis satellite instruments GOME 1/2, SCIAMACHY, and OMI. It overcomes some of the artifacts of previous algorithms, as it is capable of reproducing some of the gradients of stratospheric NO\textsubscript{2}, e.g., related to the polar vortex, and reduces interpolation errors over continents.
\item[f)] The \textit{Standard Product 2} (SP2) includes a new stratosphere-troposphere separation approach (Bucsela et al. [2013]). This approach has aspects in common with STREAM. It is based on the measurements only and uses tropospheric pollution masking and subsequent interpolation over the masked areas.
\end{itemize}

These ways of treating the stratospheric NO\textsubscript{2} field may not be accurate enough to capture the variability of the stratospheric NO\textsubscript{2} in latitudinal and longitudinal direction, as well as in time. At the same time it is not certain whether these methods do actually separate stratospheric NO\textsubscript{2}; some of the NO\textsubscript{2} interpreted as "stratospheric" may be in the (upper) troposphere.

Also the assimilation approach suffers from these uncertainties, but in a different way since actual meteorological fields are used to model the dynamical and chemical variability of NO\textsubscript{2} in the stratosphere and free troposphere. The assimilation analyses the retrieved total slant column with a strong forcing to the observations.
Table 1: NO$_2$ data product requirements for the TROPOMI NO$_2$ data products, where accuracies are split in the systematic and random components. The numbers are taken from [RD11]; see also the Product ReadMe File (PRF; available via [ER3]).

<table>
<thead>
<tr>
<th>NO$_2$ data product</th>
<th>Vertical resolution</th>
<th>Bias</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stratospheric NO$_2$</td>
<td>Stratospheric column</td>
<td>$&lt; 10%$</td>
<td>$0.5 \times 10^{15}$ molec/cm$^2$</td>
</tr>
<tr>
<td>Tropospheric NO$_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tropospheric column</td>
<td>$25 - 50%$</td>
<td>$0.7 \times 10^{15}$ molec/cm$^2$</td>
</tr>
</tbody>
</table>

over clean regions (regions with small tropospheric column amounts). The data assimilation ensures that the model simulations of the stratospheric NO$_2$ column agrees closely with the satellite measurements. The modelled stratospheric NO$_2$ (slant column) amount is subtracted from the full column observation to derive the tropospheric column.

The use of a data assimilation system to provide stratospheric NO$_2$ concentrations has been shown to provide realistic results, as indicated by validation studies. For example, Hendrick et al. [2012] found very good agreement between satellite retrievals using data assimilation to estimate the stratospheric NO$_2$ column (GOME, SCIAMACHY and GOME-2) and ground-based measurements at the station of Jungfraujoch.

The advantages of the use of stratospheric chemistry transport modelling in combination with data assimilation are:

- The system models the chemistry (diurnal cycle) and dynamics of the stratosphere based on meteorological analyses.
- Data assimilation provides a realistic error estimate of the stratospheric NO$_2$ column [Dirksen et al., 2011].
- The height of the tropopause, obtained from the meteorological data, provides a point of separation of the stratospheric from the tropospheric NO$_2$ column.
- The result of the data assimilation is a comprehensive understanding of 3-D NO$_2$ distributions that covers the whole world, taking into account the spatial and temporal variability of the NO$_2$ profiles.

5.4 NO$_2$ data product requirements

SSP/TROPOMI mission requirements have been discussed in several documents, including the GMES Sentinels-4, -5 and -5Precursor Mission Requirements Document [RD5] and the Science Requirements Document for TROPOMI [RD7]. These requirements are based on the findings of the CAPACITY [RD8], CAMELOT [RD9] and TRAQ [RD10] studies. For the TROPOMI NO$_2$ column data products the set of requirements which are used as baseline in the routine validation work are the NO$_2$ data product requirement listed in Table 1; these are given in the "Sentinel-5P Calibration and Validation Plan for the Operational Phase" document [RD11] and also given in the NO$_2$ Product ReadMe File (PRF; available via [ER3]).

The uncertainties stated in Table 1 include retrieval errors as well as instrument errors. Over polluted areas retrieval errors will dominate the uncertainties; these relate to the presence of clouds and aerosols and to the surface albedo. Over rural areas, with low NO$_2$ concentrations, errors in tropospheric NO$_2$ are mostly driven by random noise related to the instrument’s Signal-to-Noise Ratio (SNR), to estimates of the stratospheric NO$_2$ column, and to uncertainties in the NO$_2$ profile.

5.5 NO$_2$ retrieval for TROPOMI

The TROPOMI retrieval of total and tropospheric NO$_2$ is based on the TM5-MP/DOMINO system (see Sect. 6.1), thus extending the long-term record of NO$_2$ data, produced using a reliable, well-established and well-described processing system (see Boersma et al. [2004], Boersma et al. [2007] and Boersma et al. [2011]). In particular, the inclusion of many of the retrieval developments of the QA4ECV project ([RD6], [ER7]) in the TROPOMI NO$_2$ retrieval will ensure a good continuity from the QA4ECV OMI and GOME-2 NO$_2$ records to TROPOMI. For the OMI NO$_2$ retrieval a number of improvements are related to spectral fitting [Van Geffen et al., 2015] and to the chemistry modelling, stratosphere-troposphere separation and the air-mass factor [Maasakkers et al., 2013]. The TROPOMI NO$_2$ processing chain is described in Sect. 6.5.

In order to comply with the SI unit definitions, the TROPOMI NO$_2$ data product file (described further in Sect. 6.6) provides the trace gas columns in mol/m$^2$, rather then in the commonly used unit molec/cm$^2$. For convenience sake, most of the text and figures of this document will remain in the latter unit; only the tables listing the input (Sect. 7.1) and output (Sect. 7.4) dataset use the SI based units.
Table 2: Overview of periods of operation of the operational NO\textsubscript{2} processor versions in the near-real time (NRTI) and the off-line (OFFL) data streams, as well as for officially reprocessed (RPRO) data; see also the latest PRF [ER3]. Concerning NO\textsubscript{2}, there is no difference between the different minor release versions v1.2.x, and the same holds for the versions v1.3.x. The differences between version 1.2.2 and 1.3.2 are relatively small and involve only a small fraction of the observations. Version v1.4.0 involves a major upgrade in the Level-2 cloud product FRESCO used by the NO\textsubscript{2} processor (viz. Sect. 6.4.4), leading to a significant increase of the tropospheric NO\textsubscript{2} over polluted scenes. Major release version v2.2.0 of the NO\textsubscript{2} processor marks the switch to v2.0.0 of the Level-1b spectral data as well as a significant change to the surface albedo treatment. A full mission reprocessing is planned to take place in 2022. Note that on 6 August 2019, as of orbit 9388, the nadir ground pixel dimensions reduced from approximately \(7.0 \times 3.5 \text{ km}^2\) to approximately \(5.5 \times 3.5 \text{ km}^2\) without a change in the processor.

<table>
<thead>
<tr>
<th>Processor version</th>
<th>ATBD version</th>
<th>Data stream</th>
<th>In operation from</th>
<th>In operation until</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.00.00</td>
<td>1.2.0</td>
<td>NRTI</td>
<td>03745</td>
<td>2018-07-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>03661</td>
<td>2018-07-18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>03946</td>
<td>2018-07-11</td>
</tr>
<tr>
<td>01.01.00</td>
<td>1.2.0</td>
<td>NRTI</td>
<td>03947</td>
<td>2018-07-18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>03848</td>
<td>2018-07-11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05333</td>
<td>2018-10-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05235</td>
<td>2018-10-17</td>
</tr>
<tr>
<td>01.02.00</td>
<td>1.3.0</td>
<td>NRTI</td>
<td>05336</td>
<td>2018-10-24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>05236</td>
<td>2018-10-17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05929</td>
<td>2018-12-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05832</td>
<td>2018-11-28</td>
</tr>
<tr>
<td>01.02.02</td>
<td>1.3.0</td>
<td>NRTI</td>
<td>05931</td>
<td>2018-12-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>05833</td>
<td>2018-11-28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05833</td>
<td>2018-11-28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05235</td>
<td>2018-10-17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>05832</td>
<td>2018-11-28</td>
</tr>
<tr>
<td>01.03.00</td>
<td>1.4.0</td>
<td>NRTI</td>
<td>07519</td>
<td>2019-03-27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>07425</td>
<td>2019-03-20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>07999</td>
<td>2019-04-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>07907</td>
<td>2019-04-23</td>
</tr>
<tr>
<td>01.03.01</td>
<td>1.4.0</td>
<td>NRTI</td>
<td>08000</td>
<td>2019-04-30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>07907</td>
<td>2019-04-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>08906</td>
<td>2019-07-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>08814</td>
<td>2019-06-26</td>
</tr>
<tr>
<td>01.03.02</td>
<td>1.4.0</td>
<td>NRTI</td>
<td>08906</td>
<td>2019-07-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>08815</td>
<td>2019-06-26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16256</td>
<td>2020-12-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16212</td>
<td>2020-11-29</td>
</tr>
<tr>
<td>01.04.00</td>
<td>2.2.0</td>
<td>NRTI</td>
<td>16259</td>
<td>2020-12-02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>16213</td>
<td>2020-11-29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>19306</td>
<td>2021-07-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>19257</td>
<td>2021-07-01</td>
</tr>
<tr>
<td>02.02.00</td>
<td>2.2.0</td>
<td>NRTI</td>
<td>19308</td>
<td>2021-07-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>18258</td>
<td>2021-07-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21222</td>
<td>2021-11-17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21187</td>
<td>2021-11-14</td>
</tr>
<tr>
<td>02.03.01</td>
<td>2.2.0</td>
<td>NRTI</td>
<td>21223</td>
<td>2021-11-17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>21188</td>
<td>2021-11-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[ TBD ]</td>
<td>2022-07-...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[ TBD ]</td>
<td>2022-07-...</td>
</tr>
<tr>
<td>02.04.00</td>
<td>2.4.0</td>
<td>NRTI</td>
<td>[ TBD ]</td>
<td>2022-07-...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OFFL</td>
<td>[ TBD ]</td>
<td>2022-07-...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>02818</td>
<td>2022-04-30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[ TBD ]</td>
<td>2022-07-...</td>
</tr>
</tbody>
</table>

5.6 NO\textsubscript{2} data product: version history and access

The NO\textsubscript{2} processing has started directly after "first light", providing data for initial checks and validations. The near-real time (NRTI) data product is released from 4 July 2018 onwards; the reprocessed (RPRO) and off-line (OFFL) datasets contains data starting from 30 April 2018. TROPOMI NO\textsubscript{2} data processed in near-real time (NRT) is available within 3 hours after measurement; this data stream uses a forecast of the TM5-MP data
A few days later, the data is processed in off-line mode (OFFL), using TM5-MP analysis data. Table 2 provides an overview of the operational TROPOMI NO$_2$ data processor versions; see also the latest NO$_2$ PRF [ER3]. The table also lists for each processor version the version number of the ATBD that describes the data product. Older ATBD versions are usually not publicly available, though ATBD v1.4.0 will remain available as long as v1.2.x and v1.3.x data will be available.

It is important to realise that a change of processor version may imply significant changes in NO$_2$ and as a general rule different version cannot be used together for time-series and trend studies. The change from version 1.2.x to 1.3.x, however, was a relatively minor update [ER3], affecting only a minority of pixels, and these versions have been combined in past studies. The update from version 1.3.x to 1.4.0 [Van Geffen et al., 2022] on 2 December 2020 has led to significant changes in tropospheric NO$_2$, in particular over scenes with low but non-zero cloud fractions (cloud radiance fractions between 0.1 and 0.5) and high pollution levels where NO$_2$ tropospheric columns have increased. The updates from version 1.4.0 to 2.2.0 [Van Geffen et al., 2022] have mainly affected cloud-free scenes and lead to a further increase in tropospheric NO$_2$ in polluted regions. Furthermore, for snow/ice conditions and coastal areas the impact on tropospheric NO$_2$ may be larger. Datasets should therefore not be combined without carefully addressing these changes.

All data (near-real time, offline and reprocessed) is freely accessible via the Copernicus Open Access Data Hub [ER9]. Note that NRT data is available for the last 30 or so days; older NRT data is removed from the Hub. See the TROPOMI website [ER10] for more information on data availability and dissemination.
6 Algorithm description

6.1 Overview of the NO2 retrieval algorithm

The TROPOMI NO2 processing system is based on the DOMINO and QA4ECV processing systems, with improvements related to specific TROPOMI aspects and new scientific insights. The basis for the processing is a retrieval-assimilation-modelling system which uses the 3-dimensional global TM5 chemistry transport model as an essential element. The retrieval consists of a three-step procedure, performed on each measured Level-1b spectrum:

1. the retrieval of a total NO2 slant column density \( N_s \) from the Level-1b radiance and irradiance spectra measured by TROPOMI using a DOAS (Differential Optical Absorption Spectroscopy) method,
2. the separation of the \( N_s \) into a stratospheric \( N_s^{\text{strat}} = N_s^{\text{strat}} \cdot M^{\text{strat}} \) and a tropospheric \( N_s^{\text{trop}} \) part on the basis of information coming from a data assimilation system, and
3. the conversion of the tropospheric slant column density into a tropospheric vertical column density \( N_v^{\text{trop}} = N_s^{\text{trop}} / M^{\text{trop}} \).

where \( M^{\text{trop}} \) and \( M^{\text{strat}} \) are the tropospheric and stratospheric air-mass factor (AMFs), which are derived from a look-up table of altitude-dependent AMFs and actual, daily information on the vertical distribution of NO2 from the TM5-MP model on a \( 1^\circ \times 1^\circ \) grid; the altitude-dependent AMF depends on the satellite geometry, terrain height, cloud fraction and height and surface albedo.

The retrieval process is described in detail in the sections below.

6.2 Spectral fitting

The baseline method to determine NO2 total slant columns is DOAS (see Platt [1994], Platt and Stutz [2008]). The DOAS fitting function for TROPOMI follows the current non-linear fitting approach for OMI (Boersma et al. [2011], Van Geffen et al. [2015], Van Geffen et al. [2020], [RD12]).

The reflectance spectrum observed by the satellite instrument, \( R_{\text{meas}}(\lambda) \), is the ratio of the radiance at the top of the atmosphere, \( I(\lambda) \), and the extraterrestrial solar irradiance, \( E_0(\lambda) \) (where \( I \) also depends on the viewing geometry, but those arguments are left out for brevity):

\[
R_{\text{meas}}(\lambda) = \pi I(\lambda) / \mu_0 E_0(\lambda)
\]  

(1)

where \( E_0 \) and \( I \) are recorded at the same detector row and given on the same wavelength grid (see below), and \( \mu_0 = \cos(\theta_0) \) is the cosine of the solar zenith angle. The \( E_0 \) is measured once a day, when TROPOMI crosses the terminator along a given orbit, and the TROPOMI data processing uses \( E_0 \) measured closest in time to \( I \).

Since in Eq. (1) the factor \( \pi / \mu_0 \) becomes very large at high solar zenith angle, the algorithm internally actually uses the sun-normalised radiance \( I(\lambda) / E_0(\lambda) \); note that this ratio is sometimes also called reflectance.

In space-borne DOAS, \( R_{\text{meas}} \) is related to the extinction of light by scattering and absorbing species along the average photon path between sun and satellite instrument. The effective, integrated absorption of NO2 along the average photon path is represented by the total NO2 slant column density \( N_s \). The DOAS spectral fitting is performed for all satellite ground pixels with \( \theta_0 < 88^\circ \), so that there is no potential danger from the division by \( \mu_0 \) in Eq. (1). The FRESCO cloud data product (Sect. 6.4.4) uses this \( \theta_0 \) cut-off as well.

The DOAS spectral fitting attempts to find the optimal modelled reflectance spectrum, \( R_{\text{mod}}(\lambda) \), by minimising the chi-squared merit function, i.e. the smallest possible differences between the observed and modelled reflectance spectrum:

\[
\chi^2 = \sum_{\lambda_i} \left( \frac{R_{\text{meas}}(\lambda_i) - R_{\text{mod}}(\lambda_i)}{\Delta R_{\text{meas}}(\lambda_i)} \right)^2
\]  

(2)

with \( n_{\lambda} \) the number of wavelengths in the fit window and \( \Delta R_{\text{meas}}(\lambda_i) \) the noise on the reflectance, which depends on the radiance and irradiance noise given in the Level-1b product:

\[
\Delta R_{\text{meas}}(\lambda_i) = \frac{1}{E_0(\lambda_i)} \sqrt{\left( \Delta I(\lambda_i) \right)^2 + \left( \Delta E_0(\lambda_i) \right)^2 \cdot \left( R_{\text{meas}}(\lambda_i) \right)^2}
\]  

(3)

i.e. on the signal-to-noise (SNR) of the measurements. For numerical reasons a maximum of 2500 is set to the SNR on the reflectance and \( \Delta R_{\text{meas}} \) is adjusted upwards when this SNR limit is exceeded. In the NO2
wavelength range the SNR is typically 1500, but it may be larger over very bright scenes (notably clouds), potentially leading to saturation effects in the spectra, hence limiting the SNR does not pose a significant limit on the lowest SCD error estimates (the lower $\Delta R_{\text{meas}}$ the lower the SCD error is). Radiance spectral pixels flagged in the Level-1b data as bad or as suffering from saturation are filtered out before doing any further processing step.

The magnitude of $\chi^2$ is a measure for how good the fit is. Another measure for the goodness of the fit is the so-called root-mean-square (RMS) error, which is defined as follows:

$$R_{\text{RMS}} = \sqrt{\frac{1}{n_k} \sum_{i=1}^{n_k} \left( R_{\text{meas}}(\lambda_i) - R_{\text{mod}}(\lambda_i) \right)^2}$$

where the difference $R_{\text{meas}}(\lambda) - R_{\text{mod}}(\lambda)$ is usually referred to as the residual of the fit.

The baseline model function for TROPOMI follows the approach for OMI and reads as follows:

$$R_{\text{mod}}(\lambda) = P(\lambda) \cdot \exp \left[ - \sum_{k=1}^{n_k} \sigma_k(\lambda) \cdot N_{s,k} \right] \cdot \left( 1 + C_{\text{ring}} \frac{I_{\text{ring}}(\lambda)}{E_0(\lambda)} \right)$$

with $\sigma_k(\lambda)$ the absolute cross section and $N_{s,k}$ the slant column amount of molecule $k = 1, \ldots, n_k$ taken into account in the fit ($\text{NO}_2$, $\text{O}_3$, etc.), $C_{\text{ring}}$ the Ring fitting coefficient and $I_{\text{ring}}(\lambda)$ the synthetic Ring spectrum (generated from an $E_{\text{ref}}(\lambda)$ reference irradiance) and $E_0(\lambda)$ the measured irradiance. The Ring spectrum describes the differential spectral signature arising from inelastic Raman scattering of incoming sunlight by $\text{N}_2$ and $\text{O}_2$ molecules. The last term in Eq. (5) describes both the contribution of elastic scattering to the differential absorption signatures (i.e. the 1), and the modification of these differential structures by inelastic scattering (the $+ C_{\text{ring}} I_{\text{ring}}(\lambda)/E_0(\lambda)$ term) to the reflectance spectrum. The sources of the reference spectra used are discussed in Sect. 6.2.3.

In the modelled spectrum of Eq. (5) a polynomial of order $n_p$ with coefficients $a_m$:

$$P(\lambda) = \sum_{m=0}^{n_p} a_m \lambda^m$$

is introduced to account for spectrally smooth structures resulting from molecular (single and multiple) scattering and absorption, aerosol scattering and absorption, and surface albedo effects. Because of the polynomial term, only the highly structured differential absorption features contribute to the fit of the slant column densities. In order to prevent the numerical value of the polynomial components in Eq. (6) to become very large or very small (for the 405 – 465 nm fit window, for example, usually $n_p = 5$), the wavelengths are scaled to the range $[-1 : +1]$ over the fit window in the processor.

Fig. 5 shows an example of a reflectance spectrum observed by TROPOMI on 4 July 2018 during orbit 03747, along with the modelled spectrum obtained from the DOAS fit using Eq. (5). The (almost cloud-free) ground pixels lies over the industrial area of Rotterdam (scanline 2012, row 323, $\theta_0 = 27.82^\circ$, $\theta = 31.47^\circ$). Fit results: $N_{s,\text{NO}_2} = 2.73 \times 10^{16}$ molec/cm$^2$, $N_{s,\text{NO}_3} = 1.60 \times 10^{16}$ molec/cm$^2$, $N_{\text{trop}}^{\text{NO}_2} = 1.52 \times 10^{16}$ molec/cm$^2$, RMS = $1.59 \times 10^{-4}$. The residual (shown in the bottom panel of Fig. 5) is of the order of $10^{-4}$, corresponding to an unexplained residual reflectance (which is about equal to the differential optical depth times the average reflectance in the fit window; Van Geffen et al. [2020]) of that magnitude.

In order to remove strong outliers in the DOAS fit residual (caused by, e.g., high-energy particles hitting the CCD detector, variations in the dark current, or bad pixels not correctly flagged in the Level-1b data), a "spike removal" algorithm was implemented in v2.2.0 of the DOAS processor to detect outliers in the fit residual; after removal of the spectral pixels of such outliers from the measured reflectance, the $\text{NO}_2$ DOAS fit is redone to provide the final fit parameters. See Appendix F and Van Geffen et al. [2022] for a description and examples of the spike removal.

Table 3 provides an overview of the operational DOAS fit settings used in the TROPOMI processor and those used for some current and past UV/Vis backscatter satellite instruments: the fit window, the reference spectra used in the fit (see Sect. 6.2.3) and the degree of the DOAS polynomial. Note that for the processing of GOME(-1) data it was necessary to include a correction for the undersampling of the spectra, i.e. the fact that the spectral sampling is of the same order as the FWHM of the instrument slit function. For the instruments listed in Table 3 this correction is not necessary: their spectral resolution, i.e. the FWHM of the slit function, is 2–3 times as large as their spectral sampling. For TROPOMI, for example, the spectral sampling is about 0.2 nm and the FWHM is about 0.55 nm [RD4].
Figure 5: The top panel shows an example of a reflectance spectrum (black solid line) obtained by TROPOMI on 4 July 2018 during orbit 03747 and the spectrum modelled in the DOAS fit procedure (dashed red line); the inset shows an enlargement of a 10 nm wide part of the fit window. The bottom panel shows the residual of the DOAS fit, i.e. the measured minus the modelled reflectance spectrum; note that the vertical scale is a factor of 100 smaller than the scale in the top panel.

6.2.1 Wavelength calibration & common wavelength grid

Both the irradiance and radiance spectra are wavelength calibrated prior to the DOAS fit, using the same wavelength calibration approach, in the \([\lambda_b : \lambda_e]\) wavelength window. Prior to NO\(_2\) data version v2.2.0 this window equalled the NO\(_2\) fit window, with \(\lambda_b = 405\)nm and \(\lambda_e = 465\)nm. As of v2.2.0 the NO\(_2\) processor includes the O\(_2\)--O\(_2\) slant column retrieval for cloud properties (see Sect. 6.4.4.3) and to accommodate this fit the wavelength calibration windows is extended to \(\lambda_e = 495\)nm.

Using the subscripts ‘nom’ and ‘cal’ to denote nominal (i.e. from the Level-1b data product [RD13]) and calibrated wavelengths, respectively, the calibrated irradiance and radiance to be used in Eq. (1) are then given by:

\[
E_0(\lambda^\text{cal}) = E_0(\lambda^\text{nom} + w^E_0)
\]

\[
I(\lambda^\text{cal}) = I(\lambda^\text{nom} + w_s + w_q(\lambda^\text{nom} - \lambda_0))
\]

(7)

where \(w_s\) represents a wavelength shift and \(w_q\) a wavelength stretch (\(w_q > 0\)) or squeeze (\(w_q < 0\)), with \(w_q\) defined w.r.t. the central wavelength of the fit window \(\lambda_0\). Since in view of numerical stability, the wavelengths
are scaled to the range $[-1 : +1]$ over the fit window, computationally $\lambda_0 = 0$. Each wavelength calibration of Eq. (7) comes with its own $\chi^2$ as a goodness-of-fit. Turning on the $w_q$ fit parameter in the calibration of the radiance of a given TROPOMI orbit resulted in a very small stretch with a precision larger than the stretch itself, and the effect on the retrieval results is negligible [Van Geffen et al., 2020], and hence the $w_q$ fit parameter will remain turned off. Note that for the irradiance calibration we only consider a shift.

In order to avoid possible extrapolations, both these steps are performed on a wavelength range that is 1 nm wider than the fit window, i.e. the measured reflectance is formed on the common wavelength grid and then cut to the fit window. Up to v1.4.0 of the NO$_2$ processor the spectral pixel index selection is based on the first pixel after $\lambda_0$ and the last before $\lambda_n$. As of v2.2.0 the selection is based on nearest neighbours, providing a more consistent pixel index selection along-track.

The wavelength calibration of Eq. (7) is performed on the irradiance at the start of the processing of a given granule, and per radiance spectrum prior to forming the measured reflectance of Eq. (1). In order to form this reflectance, both (calibrated) spectra $I(\lambda)$ and $E_0(\lambda)$ need to be given on the same wavelength grid. In our approach $E_0(\lambda)$ is converted to the radiance wavelength grid by way of a high-sampling interpolation, taking advantage of the fact that we have additional information from a high-resolution solar reference spectrum $E_{sl}(\lambda)$. Details of the wavelength calibration and the high-sampling interpolation implemented for TROPOMI are given in Appendices A and B, respectively.

### 6.2.2 Minimising the chi-squared merit function

Slant column densities $N_{a,k}$, the Ring coefficient $C_{\text{ring}}$, and the polynomial coefficients $a_m$ are obtained from a minimisation of the $\chi^2$ of Eq. (2), i.e. the differences between the observed and modelled reflectances. In the initial TROPOMI NO$_2$ DOAS, we implemented a version of the OMI NO$_2$ DOAS processor, called OMNO2A, which uses a non-linear least squares fitting based on routines available in the SLATEC mathematical library [Vandevender and Haskell, 1982]. During the commissioning phase, however, we discovered that this implementation suffered from some issues (in particular the $\chi^2$ and/or the slant column error estimates were scaled incorrectly) that could not be solved due to inflexibility of the OMNO2A code. To solve this issue, we chose to use the optimal estimation (OE) routine based on Rodgers [2000] already available in the processor, since it was implemented for the wavelength calibration; see Appendix A.

For the $\chi^2$-minimisation using the OE solver suitable a-priori values of the fit parameters were selected and the a-priori errors are set very large, so as to limit the solution of the fit, while for numerical stability reasons a pre-whitening of the data is performed. (Whitening transforms a vector of random variables with a known covariance matrix into a set of new variables whose covariance is the identity matrix, meaning that they are uncorrelated and each have variance 1; cf. Rodgers [2000], Ch. 2.)

A number of fitting diagnostics is provided by the fitting procedure. Estimated slant column and fitting coefficient uncertainties are obtained from the covariance matrix of the standard errors, which is given as
a standard output of the OE procedure. The SCD error estimates are scaled with the square-root of the normalised $\chi^2$, where $\chi^2$ is normalised by $(n_k - D)$, with $n_k$ the number of wavelengths in the fit window and $D$ the degrees of freedom of the fit, which is almost equal to the number of fit parameters. All fitting coefficients are provided in the NO$_2$ output data file as diagnostic data.

### 6.2.3 Reference spectra

The selection of the reference spectra for the trace gas cross sections in Eq. (5) is driven by whether a species shows substantial absorption in the wavelength range relevant for NO$_2$ retrieval, and exploits the best available sources prior to commissioning phase. Experience with OMI has shown that NO$_2$, ozone, water vapour, and Rotational Raman Scattering (RRS), i.e. the inelastic part of the Rayleigh scattering (the so-called "Ring effect"), are most relevant in the wavelength interval relevant to NO$_2$. Van Geffen et al. [2015] (cf. Sect. 6.2.4) showed that including also absorption in liquid water and by the O$_2$--O$_2$ collision complex improves the fit, hence these are included for TROPOMI.

High-resolution laboratory measured absorption cross sections are convolved with the TROPOMI slit function (or: instrument spectral response function, ISRF; available via [ER11]), and sampled at a resolution of 0.01 nm to create the necessary reference spectra in the data files used by the processor. Since the ISRF is (slightly) different for different detector rows, the convolved reference spectra are determined per detector row. Given the relative smoothness of these convolved cross sections, interpolation to the radiance wavelength grid in Eq. (5) is performed by way of a 4th degree spline interpolation. The final set of reference spectra (see also [RD14] and Fig. 6) is:

- trace gas cross sections $\sigma_k(\lambda)$ in Eq. (5):
  - NO$_2$ from Vandaeele et al. [1998] at 220 K; see [ER12]
  - O$_3$ from Gorshlev et al. [2014] and Serdyuchenko et al. [2014] at 243 K
  - Water vapour (H$_2$O$_{\text{vap}}$) based on HITRAN 2012 data (see Van Geffen et al. [2015] and Sect. 4.1 of [RD14])
  - O$_2$--O$_2$ from Thalman and Volkmann [2013] at 293 K
  - Liquid water (H$_2$O$_{\text{liq}}$) from Pope and Frey [1997], resampled at 0.01 nm with a cubic spline interpolation
- an effective Ring spectrum $I_{\text{ring}}(\lambda)$ following Chance and Spurr [1997] (see Van Geffen et al. [2015] and Sect. 4.2 of [RD14])
- a high-resolution solar reference spectrum $E_{\text{ref}}(\lambda)$ from Chance and Kurucz [2010]

The inclusion of absorption by soil (as discussed by e.g. Richter et al. [2011]; Merlaud et al. [2012]) is not considered for TROPOMI, as its potential absorption signal lies well above 465 nm, the upper limit of the NO$_2$ fit window. Also currently not being considered for inclusion in the fit is the vibrational Raman scattering in clear ocean waters (e.g. Vasilkov et al. [2002], Vountas et al. [2003]), as its potential effect on the fit is currently poorly understood; cf. Sect. 6.2.5.

The temperature for the O$_3$, H$_2$O$_{\text{vap}}$ and O$_2$--O$_2$ cross section spectra is fixed. Variation of these cross section temperatures has little effect on the fit residual in the retrieval of NO$_2$ slant columns, since the shape of the differential NO$_2$ cross section is in good approximation invariant of temperature. In the case of TROPOMI, the baseline is to use an NO$_2$ cross section that has been measured for 220 K.

Note that the amplitude of the differential cross section features has a significant temperature dependence which is important to account for. The resulting NO$_2$ slant column are corrected for deviations from 220 K at later retrieval steps, as described in Sect. 6.4.2.

### 6.2.4 DOAS fit details for OMI and TROPOMI

Comparisons of OMI NO$_2$ data from the DOMINO v2 processing system to independent data from other instruments have shown that OMI slant NO$_2$ columns are higher than columns derived from GOME-2 and SCIAMACHY (as first stated by N. Krotkov at the OMI Science Meeting in Sept. 2012), as well as columns derived from ground-based measurements. Due to the separation between stratospheric and tropospheric NO$_2$, which proceeds in the same way for the three satellite instruments, the high bias in the NO$_2$ slant columns is propagated to the stratospheric column [Belmonte Rivas et al., 2014].

Van Geffen et al. [2015] showed that improving the OMI wavelength calibration of the Level-1b spectra in the OMNO2A data processing of the NO$_2$ slant columns used by DOMINO v2 reduces both the total NO$_2$ slant column values and the RMS of the DOAS fit. Van Geffen et al. [2015] further showed that including both O$_2$--O$_2$
Figure 6: Absorption cross sections $\sigma_j(\lambda)$ for NO$_2$, O$_3$, water vapour, O$_2$–O$_2$ and liquid water, as well as the Ring spectrum $I_{ring}(\lambda)$, the pseudo-absorber which accounts for the Ring effect, in Eq. (5) for the 405 – 465 nm wavelength range used in the TROPOMI data processor. The reference spectra have been multiplied by the factors given in the plot legend to make the spectral signatures visible in one plot.

and H$_2$O$_{liq}$ (discussed by e.g. Richter et al. [2011], Lerot et al. [2010]) in the fit improves the OMI NO$_2$ fit results and ensures that fit coefficients for O$_3$ and O$_2$–O$_2$ have realistic values. Criteria for establishing what are the best settings for the fit can be summarised as follows: (a) a low error on the NO$_2$ slant column, (b) a low RMS error value, (c) inclusion of secondary trace gases that clearly improve the fit, e.g. by removing specific features in the fit residual, (d) physically realistic values for the slant column values of these secondary trace gases.

The improvements described by Van Geffen et al. [2015] for OMNO2A have been used for the processing of OMI NO$_2$ data within the QA4ECV-project ([RD6], [ER7]), which demonstrated that the different spectral fitting approaches lead to consistent results, but with better precision for the QA4ECV spectral fitting algorithm [Zara et al., 2018]. The improvements are incorporated in the TROPOMI NO$_2$ slant column processing. For comparisons between TROPOMI and OMI slant column retrieval results, see Van Geffen et al. [2020].

6.2.5 Some notes regarding other DOAS implementations

Many implementations of DOAS deploy a linearised version of Eq. (5), with the Ring effect included as a pseudo-absorber, giving the equation in terms of optical depth rather than in terms of reflectances:

$$\ln|R_{mod}(\lambda)| = P^*(\lambda) - \sum_{k=1}^{n_k} \sigma_k(\lambda) \cdot N_{s,k} - \sigma_{ring}(\lambda) \cdot C_{ring}$$

(8)

where the Ring coefficient $C_{ring}^*$ and the polynomial $P^*(\lambda)$ are essentially different from $C_{ring}$ and $P(\lambda)$ in Eq. (5). In this approach the Ring cross section $\sigma_{ring}(\lambda)$ is constructed from the Ring radiance spectrum $I_{ring}(\lambda)$ divided by a reference solar spectrum minus a low-order polynomial (so that $\sigma_{ring}(\lambda)$ varies around zero). The relation between $C_{ring}$ and $C_{ring}^*$ is discussed briefly by Van Geffen et al. [2020].

The linearisation leading to Eq. (8) allows then for the use of a linear least squares fitting routine, which is computationally faster than a non-linear solver needed when using Eq. (5). We feel, however, that the Ring effect is physically described better by the non-linear approach of Eq. (5) and we therefore use that in the NO$_2$ data processing for TROPOMI. Apart from dropping the physical description of the Ring effect, a disadvantage
of the linearised approach is that error propagation is no longer straightforward, because taking the logarithm of the observed spectra implies that the error no longer has a Gaussian distribution.

Several DOAS applications include an intensity offset correction, a constant or linear in wavelength, to improve the retrievals in some spectral ranges. The precise physical origin of such an intensity offset is not known, but it is thought to be related to instrumental issues (e.g., incomplete removal of stray light or dark current in Level-1b spectra) and/or atmospheric issues (e.g., incomplete removal of Ring spectrum structures, vibrational Raman scattering in clear ocean waters); see, for example, Platt and Stutz [2008], Richter et al. [2011], Lampel et al. [2015], [RD15].

In Eq. (5) such an intensity offset correction would be represented by an additional term on the right hand side:

$$\ldots + \frac{S_{\text{off}}}{E_0(\lambda)} \sum_{m=0}^{n_{\text{off}}} c_m \lambda^m$$

with fit parameters $c_m$ and $S_{\text{off}}$ a suitable scaling factor which has the unit of the irradiance $E_0$; in most applications $n_{\text{off}} = 0$ or 1 in case an intensity offset is included. (Note that the NO$_2$ slant column retrieval algorithm of OMI (OMNO2A) is not able to handle such an intensity offset correction.)

The possibility of an intensity offset correction has been implemented in the TROPOMI NO$_2$ slant column processor, but this option is currently turned off as (i) it has not been fully tested yet, (ii) we would first like to understand the physical meaning and implications of such a correction term, and (iii) we need to investigate whether it might be relevant for TROPOMI NO$_2$ retrievals. For a first discussion see Van Geffen et al. [2020], based on which an intensity offset correction will not be included in the regular TROPOMI NO$_2$, also because instrumental effects such as straylight and dark current are corrected for in the spectral calibration in the Level 0-to-1b processor (Kleipool et al. [2018]; Ludewig et al. [2020]).

### 6.3 Separation of stratospheric and tropospheric NO$_2$

The baseline method for the TROPOMI NO$_2$ algorithm to separate the stratospheric and tropospheric contribution to the NO$_2$ total slant columns is by data assimilation of slant columns in the TM5-MP CTM [Williams et al., 2017]. KNMI has considerable experience with this method [Dirksen et al., 2011], and in the absence of collocated independent (e.g., limb) information on stratospheric NO$_2$, we consider this to be the most viable method to distinguish stratospheric from tropospheric NO$_2$. This approach explicitly accounts for chemistry and dynamics in the stratosphere.

The central idea of the data assimilation is to regularly update a CTM simulation of the three-dimensional, coupled troposphere-stratosphere NO$_2$ distribution with available measurement data in such a way that the CTM simulation of the stratospheric NO$_2$ column achieves close agreement with the TROPOMI slant columns over areas known to have little or no tropospheric NO$_2$. The assimilation effectively relies on slant columns observed over regions where the model predicts the NO$_2$ column to be dominated by stratospheric NO$_2$ (e.g., over the remote oceans). For those regions and times, the modeled slant column, i.e., the inner product of the observation operator $H$ and the simulated vertical distribution $\vec{x}$, is effectively forced to the observed state. For regions and times where the model predicts large tropospheric contributions, the slant column is not a good proxy for stratospheric NO$_2$, and the analysis adjustment is only very small. Because total reactive nitrogen (NO$_y$) is a well-conserved quantity in the stratosphere, with relatively small source and sink contributions, the information from the observations can be stored in the model over long time periods. The stratospheric wind will transport the stratospheric analysis results from the oceans and remote regions to the polluted areas.

The assimilation scheme is based on the Kalman filter technique, with a prescribed parameterisation of the horizontal correlations between forecast errors. The assimilation time step in the model is 30 minutes. A full orbit of TROPOMI observations is analysed simultaneously by the Kalman filter. This is done in order to avoid discontinuities in the analysis that may occur at the location where the orbit is divided. The mid-time of the orbit is used to determine the model time step of the analysis. The analysed profile field $\vec{x}_a$ is the 3D model field of NO$_2$ including both troposphere and stratosphere, and is calculated from the forecast $\vec{x}_f$ and the satellite superobservations $\vec{y}$ by:

$$\vec{x}_a = \vec{x}_f + \mathbf{P}^T (\mathbf{H} \mathbf{P}^T + \mathbf{R})^{-1} (\vec{y} - \mathbf{y}_f)$$

with matrix $H$ the observation operator, $P$ the forecast error covariance matrix, and $R$ the combined observation and representativeness error covariance (Eskes et al. [2003]; Dirksen et al. [2011]). The superobservations are constructed by averaging the observations and averaging kernels over each $1^\circ \times 1^\circ$ model grid cell [Boersma et al., 2016]. The term $\mathbf{P}^T (\mathbf{H} \mathbf{P}^T + \mathbf{R})^{-1}$ determines the weight given to the observations depending on the uncertainty of the observation versus the uncertainty of the model forecast. The departure $(\vec{y} - \mathbf{y}_f)$ is the
difference between observed and forecasted column (observation minus forecast). The ratio between analysis and forecast following from the Kalman equation is also applied to species that are chemically closely related to NO$_2$ in the stratosphere, i.e. NO, NO$_3$, N$_2$O$_5$ and HNO$_3$ [Dirksen et al., 2011].

A simplified modelling of the observation error is introduced [Dirksen et al., 2011], with a fixed small error attached to the stratospheric part of the slant column (where the AMF is well known) and a large error attached to the tropospheric contribution to the slant column (where the AMF is uncertain). These uncertainties (0.2 and 6.0 $\times$ 10$^{15}$ molec/cm$^2$ for the stratosphere and troposphere, respectively; both numbers refer to the total slant column divided by the geometric AMF, defined by Eq. (17)) are fixed and have been optimised with sensitivity runs in such a way that the impact of major source regions on the analysis is minimal, and the forcing over clean regions is strong and consistent with observation-minus-forecast statistics.

The observation operator $H$ is a combination of a horizontal interpolation and the application of the averaging kernel [Eskes and Boersma, 2003], an $n_l$-element vector that contains the sensitivity of TROPOMI to NO$_2$ in each model layer. The scalar product of the observation operator vector and the model NO$_2$ profile at the location of the individual TROPOMI observations yields the slant column that would be observed by TROPOMI. To further speed up the analysis step by about a factor of 8, we assimilate only half of the superobservations, those grid cells with $i+j$ = even, where $i$ and $j$ are the longitude and latitude indices of the TMS-MP grid (the "checkerboard" approach). $\tilde{y}_f$ is the model forecast of the superobservations, given by $Hx_f$.

We use the TM5-MP CTM (Williams et al. [2017]; see also Huijnen et al. [2010a]; Huijnen et al. [2010b]; [ER13]) for the assimilation of TROPOMI NO$_2$ slant columns. This is a major improvement over the DOMINO v2 data assimilation systems operated at KNMI for GOME, SCIAMACHY, OMI, and GOME-2, which use an older version of the TM CTM (TM4; e.g. Dentener et al. [2003]). The main advantage of the transition to TM5-MP is the better spatial resolution ($1^\circ \times 1^\circ$), updated information on (NO$_x$) emissions, and an improved description of relevant physical (photolysis rate constants) and chemical (reaction rate constants) processes in that model [Williams et al., 2017]. The assimilation system operates at a resolution of $1^\circ \times 1^\circ$ (longitude $\times$ latitude), with $n_l$ sigma pressure layers up to 0.1 hPa in the vertical direction. TM5-MP uses forecast and analysed 3-hourly meteorological fields from the European Centre for Medium Range Weather Forecast (ECMWF) operational model. These fields include global distributions of wind, temperature, surface pressure, humidity, (liquid and ice) water content, precipitation and surface parameters.

Once the TROPOMI slant columns have been assimilated, the integral from the layer above the tropopause to the upper TM5-MP layer provides the stratospheric slant column that can be isolated from the total slant column, giving the tropospheric slant column (cf. Sect. 6.4):

$$N_{s}^{\text{strat}} = N_{s} - N_{s}^{\text{strat}}$$

For the tropopause definition the WMO-1985 temperature gradient criterion is followed, but other definitions would not lead to significantly different results (e.g. Bucsela et al. [2013]). NO$_x$ has a C-shape profile and the air around the tropopause has only a small contribution to the total column. Compared to the QA4ECV processing of OMI a new routine has been introduced for TROPOMI as of data version 1.2.0 which reduces the fine-scale jumps in tropopause level.

The TM5-MP model provides the following information, necessary for the subsequent processing in the calculation of the AMF (see Sect. 6.4) needed for the conversion of the tropospheric slant column to the tropospheric vertical column and the final NO$_2$ data product (see Sect. 6.6):

- the stratospheric slant and vertical columns: $N_{s}^{\text{strat}}$ and $N_{v}^{\text{strat}}$
- an estimate of the error on the stratospheric vertical column: $\Delta N_{s}^{\text{strat}}$
- the NO$_2$ profile: $v_{i,NO_2}$, with $i$ the index for layer number 1, 2, ..., $n_l$ — this is represented by $\bar{x}_f$ in Eq. (10)
- the temperature profile at the layers: $T_{l}^{\text{TM5}}$, for $l = 1, 2, ..., n_l$
- the pressure level coefficients: $A_{l}^{\text{TM5}}$, $B_{l}^{\text{TM5}}$, for $l = 0, 1, ..., n_l$
- the index of the pressure level of the tropopause: $l_{TBP}^{\text{TM5}}$
- the surface elevation and pressure: $z_{s}^{\text{TM5}}$ and $p_{s}^{\text{TM5}}$, at the $1^\circ \times 1^\circ$ model resolution

Note that the model divides the atmosphere in $n_l$ layers. The pressure level coefficients determine the pressure at the $n_l + 1$ layers separating the layers: $p_{l} = A_{l}^{\text{TM5}} + B_{l}^{\text{TM5}} \cdot p_{s}$, for $l = 0, 1, ..., n_l$, with $p_{s}$ the surface pressure for the given TROPOMI ground pixel. The pressure for the layer $l$, for which the concentration (volume mixing ratio) $v_{i,NO_2}$ and the temperature $T_{l}^{\text{TM5}}$ are given, is then midway between the level pressures $p_{l-1}^{\text{TM5}}$ and $p_{l}^{\text{TM5}}$. The layer with index $l_{TBP}^{\text{TM5}}$ contains the tropopause.
6.3.1 Stratospheric chemistry in the TM5-MP model

TM5-MP is primarily a tropospheric chemistry model [Williams et al., 2017]. NO$_x$-O$_y$-HO$_y$ chemical processes are implemented according to the Carbon Bond 05 (CB05) chemistry scheme, which includes non-methane hydrocarbons to account for loss by reactions with OH [Williams et al., 2017]. Because the chemistry version of TM5-MP does not simulate N$_2$O, the actual source of NO$_x$ to the stratosphere, NO$_x$ is derived from simulated HNO$_3$ concentrations, which follow climatological HNO$_3$:O$_3$ ratios observed by ODIN between 2 hPa and 60 hPa [Maasakkers et al., 2013] and the multi-sensor reanalysis of stratospheric O$_3$ columns [Van der A et al., 2015] with climatological ozone profile shapes. In this way the model partly compensates for the biases that occur due to the missing N$_2$O source globally, and the missing reactions involving halogens which are important in the polar vortex. During the QA4ECV project, the representation of stratospheric NO$_x$ in the model has been improved by nudging ODIN HNO3:O3 ratios, leading to more realistic NO$_x$ concentrations in the free-running mode. These improvements are applied to TROPOMI as well.

Processes included in the TM5-MP tracer evolution are advection, convection, diffusion, photolysis and deposition. Rapid changes in stratospheric NO$_2$ due to e.g. sudden stratospheric warmings or changes in the vortex edge location are largely accounted for through the use of the ECMWF analysis. Solar proton events are not included in the model, but the related biases are largely removed by the assimilation. NO$_x$ emissions are based on the RETRO-REAS emission inventories for 2006. For more details, the reader is referred to Dirksen et al. [2011].

For QA4ECV and for TROPOMI the original implementation of the stratospheric climatologies has been improved by a better interpolation to the TM5-MP vertical levels and by adding an extra nudging to NOx observations from HALOE in the upper stratosphere above 1 hPa which is not well constrained by HNO$_3$ [Groß and Russell, 2005].

The data assimilation provides a regular update of the TM5-MP simulation, with a time step of 30 minutes, of the NO$_2$ distribution in the atmosphere on the basis of available observations: if NO$_2$ slant columns are available with a measurement time within 15 minutes of the model time, the model field is updated, i.e. the forecast TM5-MP state is adjusted towards the observations. The stratospheric error estimate is based on "observation minus forecast" statistics (over relatively unpolluted areas) in the assimilation. Our experience with NO$_2$ data assimilation using GOME, SCIAMACHY, OMI, and GOME-2 in TM has shown that the model chemistry responds smoothly to the updates forced by the satellite measurements.

Fig. 7 provides an example of the "observation minus forecast" (O–F) and the model forcing ("analysis minus forecast", A–F) for TROPOMI data of 1 April 2018. The difference between the two panels of Fig. 7 illustrates the effect of the assimilation: considerable O–F differences, resulting mostly from (anthropogenic) tropospheric NO$_2$ sources, have only a minor influence on the analysis. On the other hand, synoptic-scale structures in O–F persist in the A–F differences. That the A–F differences are much smaller (generally less than ±0.15 × 10$^{15}$ molec/cm$^2$) than the O–F differences (up to ±1.0 × 10$^{15}$ molec/cm$^2$) in particular over polluted regions like China, Europe and the USA, demonstrates that most tropospheric contributions are effectively discounted by the assimilation procedure.

As of TROPOMI NO$_2$ product v1.2.0 several improvements have been included. The TM5-MP model was upgraded to the latest version, including some bug fixes. In the TM5-MP model the photolysis for SZA > 85° was improved, impacting in particular the stratospheric NO$_2$ columns at high latitudes. Furthermore the assimilation of NO$_2$ observations is now restricted to the ascending part of the orbit, which is especially important during the spring-summer months (June-July). These changes have improved the retrieval for high SZA and in the polar regions (see Fig. 8).

6.4 Air-mass factor and vertical column calculations

The TROPOMI NO$_2$ algorithm uses as default pre-calculated air-mass factor look-up tables to convert the slant columns into meaningful vertical columns [Lorente et al., 2017]. The AMF, denoted by the symbol $M$, is the ratio of the slant column density of the absorbing trace gas along the (slant) optical path from sun to satellite, and the vertical column density above the point at the surface area the satellite is viewing. The total vertical column density then follows from the retrieved total slant column density:

$$N_v = N_s / M$$  \hspace{1cm} (12)

The AMF depends on the vertical profile shape of the trace gas and can be written as (Palmer et al. [2001]; Eskes and Boersma [2003]):

$$M = \frac{\sum_l m_l v_l c_l}{\sum_l v_l}, \quad m_l \equiv \delta N_s / \delta v_l$$  \hspace{1cm} (13)
Figure 7: Observation-minus-forecast (O–F, top panel) and analysis-minus-forecast (A–F, bottom panel) differences in NO$_2$ slant columns divided by the geometric AMF (Eq. (17)), for 1 April 2018 processed with version 1.2.0. The observations are averaged to “superobservations” on the 1° x 1° grid of the TM5-MP model. The model forecast is simulating the observations using the kernels and air-mass factors. The O–F demonstrates clear differences (dark-blue and bright-red spots) between the model forecast and TROPOMI concerning the fine-scale distribution of tropospheric pollution. The A–F plot shows that the assimilation hardly changes the tropospheric distribution, but efficiently updates the stratospheric fields over the more unpolluted regions like the oceans.

with $m_l$ the altitude-dependent AMFs or box AMF (see Sect. 6.4.1) that describe the vertically resolved sensitivity to NO$_2$, $v_l$ the column density in layer $l$, and $c_l$ the temperature correction term discussed below (see Sect. 6.4.2) for layer $l = 1, 2, \ldots, n_l$ [Boersma et al., 2004]. The altitude-dependent AMFs depend on retrieval (forward model) parameters, including the satellite viewing geometry, as well as surface albedo and surface pressure, cloud fraction, and cloud pressure.

The TM5-MP assimilation of the total slant columns ($N_s$) leads to an estimate for the stratospheric vertical
Figure 8: NO2 tropospheric column retrievals for the descending part of TROPOMI orbit 3623, 25 June 2018, 19h UTC, over Siberia. Version 1.0.2 is shown on the left, and version 1.2.0 on the right. Prominent unrealistic positive biases are observed in v1.0.2 and v1.1.0 for the highest solar zenith angles on the left side of the orbit, while v1.2.0 has much more realistic values close to zero with a tendency towards a weak negative bias.

profile shape with a stratospheric column amount \( N_{\text{strat}} \) in close agreement to TROPOMI as shown in Fig. 7. Summation over the layers above the tropopause level \( (l > l_{TPM}) \) to top-of-atmosphere \( (l = n_l) \) and multiplication with the box AMF provides the stratospheric AMF, from which the stratospheric slant column \( (N_{s\text{strat}}^{v\text{strat}}) \) can then be calculated:

\[
N_{s\text{strat}}^{v\text{strat}} = \sum_{l=l_{TPM}+1}^{n_l} m_l v_l c_l \quad (14)
\]

Note that there is a fundamental difference between \( N_v \) and \( N_{strat}^{vstrat} \). The total column \( N_v \) is a satellite-observed quantity, related to the true profile shapes through the averaging kernel. In contrast, the stratospheric column \( N_{strat}^{vstrat} \) is a model quantity, the direct sum of the model layer subcolumns from the tropopause to the top of the atmosphere. A comparision of \( N_{strat}^{vstrat} \) with an other model or a profile measurement should therefore not make use of the averaging kernels!

Subtracting \( N_{s\text{strat}}^{v\text{strat}} \) from the total slant column and using the tropospheric AMF, determined by adding up the layers from the surface \( (l = 1) \) up to and including the tropopause level \( (l = l_{TPM}) \) in Eq. (13), then gives the tropospheric vertical column:

\[
N_{trop}^{v} = N_{s} - N_{s\text{strat}}^{v\text{strat}} \quad \Rightarrow \quad N_{trop}^{v} = N_{s} / M_{trop}^{v} \quad (15)
\]

Note that the total vertical column \( N_v \) in Eq. (12) is not the same as sum of the partial vertical columns:

\[
N_{v\text{sum}} = N_{trop}^{v} + N_{strat}^{vstrat} \quad \neq \quad N_{v} \quad (16)
\]

Our best physical estimate of the NO2 vertical column at any given place is the sum \( N_{v\text{sum}} \). Users who, for example, wish to assimilate NO2 total columns should, however, use the total column \( N_v \) for this. The total column \( N_v \) depends strongly on the modelled ratio of the stratospheric and tropospheric sub-columns, a dependency which is partly removed in the summed product. For data assimilation use is made of the averaging kernels, and in this way the resulting analyses are not dependent on the a-priori (including the ratio of the model tropospheric and stratospheric column).

In the absence of atmospheric scattering and in a plane-parallel atmosphere, a so-called geometric AMF, denoted by \( M_{geo}^{v} \), can be defined by way of a simple function of the solar zenith angle \( \theta_0 \) and of the viewing zenith angle \( \theta \): 

\[
M_{geo}^{v} = \frac{1}{\cos \theta_0} + \frac{1}{\cos \theta} \quad (17)
\]

This quantity is used in the criteria for the qa_value (see Appendix E) but not written to the output data product. The ratio \( N_{geo}^{v} = N_v / M_{geo}^{v} \) could be called the geometric column density, to distinguish it from the vertical column densities computed using AMFs that contain model information [Van Geffen et al., 2020].

6.4.1 Altitude dependent AMFs

The altitude-dependent AMFs, or vertical sensitivities, have been calculated with a radiative transfer model by adding a small, optically thin amount of NO2 to the model atmosphere layer \( l \) for an atmosphere that is
otherwise devoid of NO\textsubscript{2}, and subsequently ratioing the excess NO\textsubscript{2} slant column (simulated with a radiative transfer model) to the vertical column added to that layer (\(m_l = \delta N_s / \delta v_l\)) [Lorente et al., 2017]. The model atmosphere does not include aerosols and describes the Earth’s surface as a Lambertian reflector.

As radiative transfer model we use the Doubling-Adding KNMI (DAK) radiative transfer model (De Haan et al. [1987]; Stammes et al. [2001]), version 3.2, which has the possibility to include a pseudo-sphericity correction. The radiative transfer calculations take the sphericity of the atmosphere into account, with Rayleigh scattering (including multiple scattering effects) and polarisation correction included (see Boersma et al. [2011] and references therein); this includes a simple sphericity correction based on detailed comparisons between DAK and McArtim as described in Lorente et al. [2017]. The DAK model atmosphere consists of a Lambertian surface albedo, and an adjustable number of atmospheric layers. Atmospheric data are from the standard AFGL midlatitude summer profile. We calculate the AMF at 437.5 nm, near the middle of the spectral fitting window, for the corresponding TROPOMI NO\textsubscript{2} slant column retrievals; this is a suitable choice for both the small (425 – 450 nm) and wide (405 – 465 nm) fit windows, as demonstrated in the QA4ECV-project ([RD6], [ER7], see document [RD15]).

The altitude-dependent AMFs are stored in a look-up table (LUT) as a function of solar zenith angle (\(\theta_0\)), viewing zenith angle (\(\theta\)), relative azimuth angle (\(\phi_{rel}\)), Lambertian surface albedo (\(A_s\)), surface pressure (\(p_s\)), and (mid-level) atmospheric pressure (\(p_l\)). This 6-dimensional LUT is extended with more reference points compared to earlier versions in order to respect the increase in variability of TROPOMI retrieval parameters (coarser OMI pixels have less variability in spatially smeared surface albedo and surface pressure values than TROPOMI) and to minimise interpolation errors when looking up the appropriate altitude-dependent AMF. Pixel-specific altitude-dependent AMFs are obtained by using the best estimates for forward model parameters and a 6-D linear interpolation scheme.

Table 4 gives an overview of the reference points for the quantities that make up the 6 dimensions. The dimensions for the LUT are chosen to balance sufficiently accurate 6-dimensional linear interpolation with computational efficiency and resource economy. For out-of-bounds values (there is a slight chance that this occurs for surface pressure or atmospheric pressure) we use the point nearest to the LUT reference point. In the current OMI NO\textsubscript{2} data product only ground pixels with \(\theta_0 < 80^\circ\) (\(\cos(\theta_0) = 0.174\)) are used in the conversion to vertical columns. For TROPOMI and future OMI NO\textsubscript{2} data products the slant to vertical column conversion will not be limited in terms of \(\theta_0\); in practice this means the range will be the same as for the FRESCO cloud retrieval: \(\theta_0 < 88^\circ\) (i.e. \(\cos(\theta_0) = 0.035\)), hence the lower limit of \(\cos(\theta_0)\) of 0.03 in Table 4. (For practical implementation purposes the LUT contains the altitude-dependent air-mass factor scaled with the geometric AMF, \(v_l/M_{geo}\), rather than \(v_l\) directly.)

The TROPOMI qa_value (see Appendix E) indicates that observations with \(\theta_0 > 81.2\) should not be used. Experience has shown that observation-minus-forecast differences increase rapidly above this point.

6.4.2 Temperature correction

For the TROPOMI NO\textsubscript{2} retrieval, a temperature correction is applied in the air-mass factor step (see Eq. (13)). The NO\textsubscript{2} cross-sections used in the DOAS retrieval, taken from Vandaele et al. [1998] [ER12], are valid for NO\textsubscript{2} at a temperature of 220 K. The temperature at which the NO\textsubscript{2} cross-section is evaluated does significantly

![Figure 9: Temperature correction factors used in OMI NO\textsubscript{2} algorithms from [RD16] (green line), Bucsela et al. [2013] (yellow line), and from Boersma et al. [2004] (red line) as a function of temperature. Source: [RD16].](image_url)
Table 4: Quantities and their reference points in the AMF look-up table to be used in the TROPOMI NO$_2$ data processing to convert the tropospheric slant column into the tropospheric vertical column. The lower limit of $\cos(\theta)$ in the list is related to the maximum value of $\theta$ for TROPOMI, which is $72^\circ$ (as for OMI).

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Number of reference points</th>
<th>Values at reference points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar zenith angle $\cos(\theta_0)$</td>
<td>17</td>
<td>1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05, 0.03</td>
</tr>
<tr>
<td>Viewing zenith angle $\cos(\theta)$</td>
<td>11</td>
<td>1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.45, 0.40, 0.35, 0.30</td>
</tr>
<tr>
<td>Relative azimuth angle $180^\circ -</td>
<td>\phi - \phi_0</td>
<td>$</td>
</tr>
<tr>
<td>Surface albedo $A_s$</td>
<td>26</td>
<td>0.00, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00</td>
</tr>
<tr>
<td>Surface pressure $p_s$ [hPa]</td>
<td>14</td>
<td>1048, 1036, 1024, 1013, 978, 923, 840, 754, 667, 554, 455, 372, 281, 130</td>
</tr>
<tr>
<td>Atmospheric pressure $p_l$ [hPa]</td>
<td>174</td>
<td>1054.995, 1042.82, 1030.78, 1018.89, 1007.13, 995.51, 984.0309, 972.67, 961.45, 950.35, 939.39, 928.55, 917.84, 907.24, 896.71, 886.24, 875.88, 865.65, 855.54, 845.54, 835.67, 825.90, 816.26, 806.72, 797.12, 787.47, 777.93, 766.51, 759.21, 750.01, 740.93, 731.96, 723.09, 714.33, 705.65, 697.04, 688.54, 680.14, 671.85, 663.65, 655.56, 647.56, 639.66, 631.86, 624.07, 616.30, 608.62, 601.03, 593.54, 586.15, 578.85, 571.63, 564.51, 557.48, 550.44, 543.39, 536.43, 529.56, 522.77, 515.08, 509.47, 502.97, 496.50, 490.14, 483.75, 477.32, 470.97, 464.71, 458.53, 452.44, 446.42, 440.49, 434.63, 428.86, 423.12, 417.42, 411.80, 406.26, 400.79, 395.39, 390.07, 384.82, 379.64, 374.52, 369.43, 364.37, 359.37, 354.44, 349.57, 344.78, 340.05, 335.38, 330.78, 326.24, 321.70, 317.15, 312.66, 308.24, 303.89, 299.59, 295.35, 291.18, 287.06, 283.00, 261.31, 255.35, 193.41, 165.49, 141.03, 120.12, 102.68, 87.82, 75.12, 64.30, 55.08, 47.20, 40.535, 34.79, 29.86, 25.70, 22.14, 19.08, 16.46, 14.20, 12.30, 10.69, 9.29, 8.06, 6.70, 6.11, 5.37, 4.70, 4.10, 3.57, 3.12, 2.74, 2.41, 2.12, 1.87, 1.65, 1.46, 1.29, 1.141, 1.01, 0.89, 0.79, 0.69, 0.61, 0.54, 0.48, 0.42, 0.37, 0.33, 0.29, 0.23, 0.18, 0.13, 0.10, 0.07, 0.05, 0.04, 0.03, 0.02, 0.014, 0.0099, 0.0066, 0.00471, 0.002997, 0.002005, 0.001352, 0.0009193, 0.0006330, 0.0004387, 0.000307</td>
</tr>
</tbody>
</table>

Influence the fit: amplitudes of the differential NO$_2$ absorption features decrease with increasing temperature, while the overall shape of the differential cross-section is in good approximation independent of temperature.

To account for the temperature sensitivity, a correction factor has been determined for the difference between the effective temperature of the NO$_2$ (which is derived from the ECMWF temperature profile and the modelled profiles in the data assimilation system) and the temperature of the cross-section, where the temperature dependence is assumed to be linear. For layer $l$ of the NO$_2$ profile the correction factor $c_l$ is [RD16]:

$$c_l = 1 - 0.00316(T_l - T_{\sigma}) + 3.39 \times 10^{-6}(T_l - T_{\sigma})^2$$ (18)

with $T_l$ and $T_{\sigma}$ the temperature of the profile layer and cross-section, respectively. The function in Eq. (18) is
an update w.r.t. the correction used for the OMI NO\textsubscript{2} data in DOMINO v2 (Boersma et al. [2002], Boersma et al. [2004], Bucsela et al. [2013]) – see Fig. 9. Note that the temperature sensitivity given in the above equation is determined for the default wavelength window 405 – 465 nm used for the fit; depending on the fit window and on TROPOMI’s spectral resolution details, the function may need to be adapted.

6.4.3 Correction for cloud cover

The AMF formulation accounts for cloud-contaminated pixels. Following Martin et al. [2002] and Boersma et al. [2002], the independent pixel approximation (IPA) is used to express the AMF as a linear combination of a cloudy AMF ($M_{\text{clid}}$) and a clear-sky AMF ($M_{\text{clt}}$), both for the total column and the tropospheric column:

\[ M = w M_{\text{clid}} + (1 - w) M_{\text{clt}}, \quad M_{\text{trop}} = w M_{\text{clid}}^{\text{trop}} + (1 - w) M_{\text{clt}}^{\text{trop}} \]  

(19)

with $w$ the radiance weighted cloud fraction, which depends on the effective cloud fraction ($f_{\text{eff}}$):

\[ w = \frac{f_{\text{eff}} I_{\text{cl}}}{R} = \frac{f_{\text{eff}} I_{\text{cl}}}{f_{\text{eff}} I_{\text{cl}} + (1 - f_{\text{eff}}) I_{\text{cr}}} \]  

(20)

where $I_{\text{cl}}$ is the radiance from the cloudy part of the pixel, $I_{\text{cr}}$ the radiance from the clear part of the pixel, and $R$ the total scene radiance. The NO\textsubscript{2} column below the clouds, i.e. the TM5-MP NO\textsubscript{2} profile integrated from the surface to the cloud pressure level, is called the ghost column ($N_{\text{cl}}^{\text{hos}}$). Both $I_{\text{cl}}$ and $I_{\text{cr}}$ depend on the viewing geometry, the assumed (cloud) albedo, the surface pressure and the cloud pressure, following from the FRESCO algorithm (Sect. 6.4.4). In the DOMINO v2 and TM4NO2A processing of data from OMI, GOME-2 and their predecessors, these radiances were calculated following the analytical approach of Vermote and Tanré [1992], using $f_{\text{eff}}$ from the cloud retrieval process for the same instrument. For TROPOMI, the cloud (radiance) fraction is determined from the radiance in the NO\textsubscript{2} fit window using LUTs, as detailed below (Sect. 6.4.4).

6.4.4 Cloud cover and cloud pressure data

The large spectral range covered by TROPOMI offers multiple wavelength windows where information on the cloud height may be derived, including the O\textsubscript{2} A-band [Wang et al., 2008], the O\textsubscript{2} B-band [Desmons et al., 2019] and the O\textsubscript{2}–O\textsubscript{3} absorption feature at 477 nm [Veefkind et al., 2016]. The OMI spectral range does not cover the A- and B-band, and most retrievals make use of the O\textsubscript{2}–O\textsubscript{3} absorption. Because of the strong absorption with enhanced sensitivity for high clouds, and long experience with cloud retrievals for the SCIAMACHY and GOME-2 instruments, it was decided to use the O\textsubscript{2} A-band as default for the cloud pressure retrieval for TROPOMI, using the FRESCO algorithm. This choice is a major difference between OMI and TROPOMI and has a significant impact on the retrieved NO\textsubscript{2} tropospheric columns.

The FRESCO\textsuperscript{+} algorithm (Wang et al. [2008]; [RD17]) retrieves cloud information from the O\textsubscript{2} A-band around 758 nm: the cloud fraction and the cloud pressure, for all satellite ground pixels with solar zenith angle $\theta_{s} < 88^\circ$. The surface albedo database that is used by the FRESCO\textsuperscript{+} algorithm is based on GOME-2 observations [Tilstra et al., 2017] at 758 and 772 nm; see Sect. 6.4.5.

Due to the high spectral resolution of TROPOMI compared to GOME-2, the FRESCO\textsuperscript{+} algorithm needed to be re-written and the corresponding lookup tables have been generated once more. The result, called FRESCO-S (short for FRESCO-Sentinel), is used for the TROPOMI NO\textsubscript{2} product and will be used for other Sentinels as well. The FRESCO-S algorithm was used up to processor version 1.3.x. As of version 1.4.0 a further improvement, nicknamed FRESCO-wide and desribed in Sect. 6.4.4.2, was implemented.

The general name “FRESCO” is used below, unless a specific version is referred to. Note that FRESCO-S is described in the Sentinel-5 ATBD [RD18], which is not yet publicly available, while the FRESCO-wide approach is discussed in Sect. 6.4.4.2 below.

6.4.4.1 The FRESCO-S cloud pressure & NO\textsubscript{2} cloud fraction

FRESCO does not provide the geometric cloud fraction but rather a radiometric equivalent cloud fraction: an effective cloud fraction, $f_{\text{eff}}$, that results in the same top-of-atmosphere (TOA) radiance as the real cloud, assuming an optically thick Lambertian cloud with a fixed albedo of $A_{c} = 0.8$ (which may be adapted in case of very bright scenes) at the cloud pressure level, $p_{c}$. This approach has proven to be useful for trace gas retrievals; see Wang et al. [2008], who evaluated this for ozone and NO\textsubscript{2}.
Because of the large difference in wavelength between the O$_2$ A-band and the NO$_2$ retrieval window, the cloud fraction retrieved by FRESCO in the O$_2$ A-band may not be exactly representative for the cloud fraction in the NO$_2$ window, although Van Diedenhoven et al. [2007] found that cloud parameters retrieved from UV and O$_2$ A-band measurements showed good consistency for cloud fractions $> 0.2$: for mostly clear skies, FRESCO provides somewhat higher cloud fractions than UV-based retrievals. For small cloud fractions the O$_2$ A-band retrieval becomes sensitive to errors in the surface albedo climatology used, especially over the bright vegetation. In addition, a misalignment between ground pixel field-of-view of the VIS and NIR bands [RD4], containing the NO$_2$ retrieval window and the O$_2$ A-band, respectively, exists for the TROPOMI measurements.

For these reasons, the baseline option for the TROPOMI NO$_2$ retrieval is to (i) use the cloud pressure $p_c$ from FRESCO and (ii) retrieve the cloud fraction and cloud radiance fraction from the NO$_2$ spectral window itself at 440 nm. The latter is done by fitting the continuum reflectance at 440 nm to a simulated reflectance constructed with the independent pixel approximation and radiative transfer calculations for the clear-sky and cloudy-sky part of the pixel, using the appropriate surface albedo in that spectral window, $A_{\lambda, NO2w}$, as forward model parameter. Here $A_{\lambda, NO2w}$ is taken from the OMI albedo climatology (cf. Sect. 6.4.5) at 440 nm, interpolated linearly in time, and using nearest neighbour sampling in latitude and longitude.

The continuum reflectance at 440 nm could be determined from the observed spectrum, averaged over a small wavelength interval, but that may lead to unexpected values, e.g. in case of spikes in the measurement or missing wavelength pixels. Instead, as of v1.4.0 [Van Geffen et al., 2022], the modelled reflectance of Eq. (5) is evaluated at $\lambda = 440$ nm, without taking the absorbing trace gases into account: $RT_\text{OA}(\lambda) = P(\lambda) \cdot (1 + C_{\text{ring}})$, where the $C_{\text{ring}}$ term is included because Rayleigh scattering is a combination of elastic Cabannes scattering and inelastic Raman scattering without the spectral structures of the latter. The trace gas absorption is not taken into account here because the reflectance LUT used for the determination of the cloud fraction does not include trace gas absorption either. This reflectance-based approach to determine the cloud fraction in the NO$_2$ window is very similar to FRESCO and explicitly accounts for Rayleigh scattering and involves the calculation of LUTs with the TOA reflectance at 440 nm as a function of viewing geometry, surface/cloud albedo, and surface/cloud pressure. See Appendix C for details on the cloud fraction retrieval.

With processor version 1.3.0 some improvements were made in the FRESCO-S algorithm and in the way the FRESCO results are treated in the NO$_2$ algorithm. Previously in FRESCO the surface albedo from the database was used without modifications other than snow or ice at the surface (see Sect. 6.4.5 and Appendix D). In the updated version the surface albedo is reduced to match the top-of-atmosphere reflectance in case the top of atmosphere reflectance is lower than expected for the climatological surface albedo (see Sect. 6.4.5 and App. C.1). Before the update these cases would lead to negative cloud fractions and unrealistic cloud pressures. Another change in FRESCO is the treatment of very high cloud fractions. If the scene albedo indicates an elevated scene height and a scene albedo higher than 0.8, the parameters from the scene retrieval are used: the cloud fraction is set to 1, the cloud albedo is set to the scene albedo and the cloud pressure is set to the scene pressure. This will prevent odd behaviour for scenes with cloud fraction $f_{\text{eff}} > 1$. Care has been taken to not use this for snow scenes. Fig. 10 gives an example of the results of these improvements, implemented for version 1.3.0, on the cloud pressure (top panels) and on the tropospheric NO$_2$ column (lower panels).

### 6.4.4.2 Cloud pressure: the FRESCO-wide update

Early studies of the TROPOMI NO$_2$ retrieval indicated that the cloud pressures produced by FRESCO-S show a positive bias (of order 50 hPa) in cases where either low clouds are expected (e.g. low clouds and fog over ocean near the coast) or locations with thick aerosol layers, like often observed over Eastern China. The NO$_2$ retrieval is very sensitive to a positive bias in the cloud pressure in combination with cloud radiance fractions close to 0.5. This will result in large $M$ values and underestimates of the tropospheric column. Comparisons between TROPOMI and OMI-QA4ECV NO$_2$ tropospheric columns have shown that NO$_2$ data versions 1.2.x and 1.3.x are substantially lower than OMI-QA4ECV, while the slant columns are in good agreement. The main difference between these retrievals is the cloud pressure retrieval.

For processor version 1.4.0 an update of FRESCO-S was introduced [Van Geffen et al., 2022]. The main change is indicated in Fig. 11. The new scheme, called FRESCO-wide, makes use of the longer wavelength part of the O$_2$ A-band. The wavelength ranges in FRESCO-wide are 758 – 759 nm, 760 – 761 nm and 765 – 770 nm. The inclusion of the weaker absorption lines in the latter window mainly impacts the lower clouds, generally increasing the height and decreasing the cloud pressure. In those cases the differences are typically in the order of 50 hPa. For high clouds both FRESCO versions deliver very similar cloud heights on average.

Fig. 12 shows a worst-case example of wintertime China with a thick aerosol layer. The FRESCO-S and
Figure 10: The retrieval of tropospheric NO$_2$ over France on 24 February 2018, for processor version 1.2.0 or older (left) and version 1.3.0 (right). The old FRESCO-S implementation was producing unrealistic low cloud top pressures for negative cloud fractions, the blue spots in the top-left panel. These have been removed in the new implementation (top-right). In the (relatively rare) case that the cloud fraction from the NO$_2$ spectral window is positive, but FRESCO-S (v1.2.0) retrieves a negative cloud fraction with unrealistic cloud pressures, this was resulting in high, noisy NO$_2$ spots around Paris (left panel, v1.2.0). The new treatment in FRESCO-S results in more realistic cloud pressures, the spots disappear, and we observe a well-defined pollution plume from Paris transported by the wind from the north-east (right panel, v1.3.0). Note that the NO$_2$ data has been filtered for cloud-radiance fractions larger than 0.5 (white area in the lower panels).

Figure 11: The original FRESCO-S cloud retrieval (left panel) makes use of three narrow wavelength ranges in the O$_2$ A-band. The new approach (FRESCO-wide, right panel), introduced in version 1.4.0, extends the third wavelength band and makes use of the weaker absorption lines.

NO$_2$ spectral window derived cloud radiance fractions are consistent and are around 0.5 as a result of the enhanced scattering of sunlight by the aerosols. The FRESCO-S effective pressure, however, does not pick up the aerosol scattering taking place at elevated levels, and the effective cloud pressure is close to or equal to the surface pressure. FRESCO-wide does produce a uniform and elevated layer, with $p_s - p_e$ values between 40 and 70 hPa. This level is in good agreement with roughly the middle of the aerosol layer as modelled by CAMS.

As shown in Fig. 13, the replacement of the FRESCO-S by FRESCO-wide in version 1.4.0 results in a large
Figure 12: Surface pressure minus the original FRESCO-S cloud pressure (left panel) or the FRESCO-wide cloud pressure (right panel) for the overpass over Eastern China on 23 January 2019. The VIIRS image for the same day (centre panel) indicates a thick aerosol layer over the populated part of China, and the TROPOMI-retrieved effective cloud-radiance fraction is close to 0.5.

Figure 13: Comparison of TROPOMI (vertical axis) and OMI-QA4ECV (horizontal axis) for Eastern China for v1.3.x December/January 2019/2020 with FRESCO-S (top panels) and for v1.4.0 December/January 2020/2021 with FRESCO-wide (bottom panels). Both datasets have been time-averaged and gridded to a common $0.8^\circ \times 0.4^\circ$ grid. Individual grid boxes are shown in the plot. Slopes are 0.52 and 0.66 for v1.3.x, 0.98 and 1.11 for v1.4.0.

increase of tropospheric NO$_2$, especially over Eastern China in winter. Slopes in the plots are now close to 1, indicating a much better agreement between the OMI and TROPOMI results. Similar results are found for version 2.x.x. In other regions like Europe or the USA we also see increases and better mean agreements with OMI-QA4ECV. In summer the differences are smaller than in winter.

In processor version 2.4.0 the NIR albedo LER datasets, based on GOME-2 observations up to v2.3.1, was replaced by an albedo DLER datasets derived from the TROPOMI observations, see Sect. 6.4.5. This replacement, which also accounts for the (viewing-angle dependent) directionality of the LER, leads to significant changes in the FRESCO-S cloud retrievals.

Fig. 14 shows the impact of the replacement of the GOME-2 LER by the TROPOMI DLER.
Figure 14: FRESCO-wide cloud fraction for processor version 2.3.1 (left panel) and version 2.4.0 (middle panel). The difference is shown in the right panel. Data is shown for orbit 21896 of 3 January 2022, passing over South America.

Figure 15: Comparison between the TROPOMI cloud fraction retrieved in the NO$_2$ fit window ($f_{\text{eff}, \text{NO}_2}$; x-axis) and in the O$_2$ A-band by the FRESCO algorithm ($f_{\text{eff}}$, y-axis) for the v2.3.1 S5P-PAL reprocessing data ([ER14]), which uses the OMI and GOME-2 LER (left panel) and the v2.4.0 test reprocessing data (named DDS6), which uses the new TROPOMI DLER v1 (right panel). The comparison is shown for ground pixels over South America (latitude range: $[-30^\circ: +5^\circ]$) of orbit 20236 (8 Sep. 2021).

For internal consistency of the NO$_2$ column retrieval the cloud fractions derived in the O$_2$ A-band and in the NO$_2$ fitting window should be similar. As shown above, in Fig. 15 this was not the case over high-albedo vegetated scenes in processor versions up to v2.3.1. For instance, when the NO$_2$ fitting window indicates a significant cloud fraction, while the NIR has a cloud fraction close to zero, the cloud pressure is ill defined and we expect a large error in the tropospheric column retrieval. With the new TROPOMI DLER in the NIR this situation has improved substantially and cloud fractions have become much more equivalent.
The changes in the cloud fraction and LER also impact the cloud pressure. On average we find an increase of the FRESCO cloud pressure by about 15 hPa, but there are strong regional differences. The histogram (not shown) indicates that this is partly related to the disappearance of an unexpected maximum in the number of FRESCO retrieved pressures around 550 hPa.

### 6.4.4.3 The O$_2$–O$_2$ cloud pressure

The O$_2$–O$_2$ cloud pressure retrieval used for the OMI NO$_2$ retrievals [Veefkind et al., 2016] has been ported to TROPOMI and is included in the operational software since version 2.2.0. The Level-2 NO$_2$ product files are extended with two separate groups for the cloud retrievals, one containing the FRESCO results taken from the separate FRESCO support product files, and one containing the O$_2$–O$_2$ retrieval datasets. The cloud data is derived from the O$_2$–O$_2$ slant column density retrieved in the wavelength window 460–490 nm with a DOAS approach similar to the NO$_2$ retrieval (Sect. 6.2), with the model function formulated in Eq. (5) and using cross-sections $\sigma_k(\lambda)$ for O$_2$–O$_2$, NO$_2$ and ozone.

The air mass factors in version 2.2.0 are computed solely using the FRESCO-wide cloud pressures (Sect. 6.4.4.2). For a later NO$_2$ data release we will study the quality of both cloud products and rules will be developed to use one or the other cloud product depending on under which circumstances they work best, which will be then described in a future update of this ATBD.

### 6.4.4.4 Other cloud data products

Apart from the FRESCO and O$_2$–O$_2$ support cloud products, TROPOMI cloud parameters are provided by an algorithm developed at DLR ([RD19], available via [ER1]). Once the validity and reliability of this cloud data product is established, its cloud parameters will be tested in the NO$_2$ processor and the results will be compared to the results found with FRESCO, O$_2$–O$_2$ and other cloud data.

Cloud parameter retrieval similar to FRESCO in the O$_2$ A-band [Wang et al., 2008] can also be done in the the O$_2$ B-band [Desmons et al., 2019], which has the advantage of a smaller albedo over vegetation. We plan to also test whether these cloud products may be useful for the NO$_2$ retrieval.

### 6.4.5 Surface albedo

#### 6.4.5.1 Processor versions v1.0.2 up to v2.3.1

The baseline surface albedo climatology for TROPOMI NO$_2$ retrievals for processor versions v1.0.2 up to v2.3.1 is the OMI database, aggregated to a grid of 0.5° × 0.5°; see Kleipool et al. [2008], which describes a climatology made from 3 years of OMI data. Meanwhile the climatology has been improved by using 5 years of data, based on the the same method [ER15]. This 5 years based climatology (version 3) has been used for the DOMINO v2 and QA4ECV OMI NO$_2$ retrievals, and is also used for the TROPOMI NO$_2$ retrievals, where the "mode LER" is used. For the surface albedo in the NO$_2$ window, $A_{\text{NO}_2}$, the 440 nm data is used. The climatological value of the surface albedo are adapted in case the snow/ice flag (cf. Sect. 6.4.6) indicates there may be substantial differences in albedo; see Appendix D for some details on this correction.

The OMI albedo climatology was chosen because of its spectral coverage in the NO$_2$ fit region, its relatively high spatial resolution, and the seamless transition between land and sea. The OMI albedo guarantees a consistency between TROPOMI and OMI NO$_2$, since it is also used in the QA4ECV data product Boersma et al. [2018]. An additional advantage is that the OMI climatology has been derived from observations taken at similar local times and under similar viewing conditions as for the TROPOMI observations.

The Kleipool surface albedo climatology is based on OMI data, which does not cover the near-infrared wavelengths in use by the FRESCO algorithm to derive cloud properties (Sect. 6.4.4). Instead, the surface albedo database that is used by the FRESCO algorithm up to version 2.3.1 is based on GOME-2 observations [Tilstra et al., 2017] at 758 and 772 nm, given at 0.25° × 0.25° resolution. The relatively coarse spatial resolution of GOME-2 measurements underlying the climatology and the fact that the overpass time of the OMI measurements used for the Kleipool climatology is quite similar to the overpass times of TROPOMI (i.e. the measurements are taken under similar viewing geometries), while the overpass time of GOME-2 is several hours earlier, are in favour of our choice for the Kleipool surface albedo climatology for the NO$_2$ retrieval, and for our choice to determine the cloud fraction in the NO$_2$ window.

Taking the anisotropic properties of the surface reflectance (so-called BRDF effects) into account was not implemented in the TROPOMI NO$_2$ retrieval algorithm up to version 2.3.1. Accounting for BRDF in OMI NO$_2$ retrievals has a generally small effect (< 5%) with substantial effects only occurring at extreme viewing angles
Figure 16: An example of the Sentinel-5P DLER computed for January (orbit 21896, 3 Jan. 2022) at 758 nm in the near infrared, relevant for the FRESCO cloud retrieval. The central panel shows the TROPOMI DLER used in v2.4.0, and is compared to the GOME-2 LER used in v2.3.1 and older versions of the NO$_2$ retrieval. The difference is shown in the right panel.

at high solar zenith angles [Zhou et al., 2010]. Developments in generating improved surface albedo data are described for instance from the ADAM ([RD14], Sect. 6.1) and QA4ECV ([RD6], [ER7]) projects, and the recent work by Vasilkov et al. [2017].

An alternative approach is to make use of the MODIS geometry-dependent surface Lambertian equivalent reflectivity. This was used to generate version V4.0 of the NASA standard NO$_2$ product from OMI [Lamsal et al., 2021]. Recently, it was shown by Loyola et al. [2020] how an effective geometry-dependent Lambertian equivalent reflectivity may be derived from the TROPOMI data.

### 6.4.5.2 Processor version v2.4.0

A directional surface albedo climatology for GOME-2 was developed by Tilstra et al. [2021]. Following the same approach, the author also developed a DLER based on TROPOMI Level-1b v1.0 measurements. This v1.0 dataset, along with its ATBD [RD20], PUM and validation report is available from the TEMIS website at [ER16]. The directional LER, or DLER, is available for wavelengths in all bands, and can therefore be applied consistently to both the FRESCO cloud retrieval and the NO$_2$ retrieval in the 405 – 465 nm window. NO$_2$ processor version 2.4.0, which is operational as of mid-2022, will make use of this viewing-angle dependent DLER product; this version will also be used for a full mission reprocessing.

This Sentinel-5P LER/DLER monthly product is based on TROPOMI v1 L1B data from 2018 onwards, has 21 selected wavelength bands in the range 328 – 2314 nm. The resolution of the dataset is 0.125$\degree$ × 0.125$\degree$ globally, which is a major improvement compared to previous LER products for OMI or GOME-2. The viewing-angle dependence is parameterised using a third order polynomial, and the four coefficients, together with the mean LER determine the DLER (see [RD20]).

An example of the implemented DLER is shown in Fig. 16. Most striking are the high albedo values at the East side over the Amazon region in Brazil. This East-West asymmetry reflects a real property of the vegetation and is missing in the GOME-2 LER. As shown in Fig. 14 above, these higher albedo values remove the biases in the cloud fraction which are systematically observed in FRESCO at the East side of the swath over forest areas. Due to the small pixel size and short revisit time of TROPOMI we can see that over Canada and the USA the albedo map is less influenced by snow at the surface. The big improvement in spatial resolution is also clearly demonstrated by these maps, for instance over mountain ranges.

The impact of replacing the OMI and GOME-2 LER by the TROPOMI DLER is shown in Fig. 17. For September 2020 we find relatively minor differences over Europe and the US. More major changes are found over the rain forests in Brazil and Central Africa, as well as in the South of China. These are regions where the directionality of the LER is large in the NIR spectral range. Especially in Brazil we observe major increases in the column up to $10^{15}$ molec/cm$^2$ related to changes in the FRESCO cloud pressure. Note that changes in albedo are partly compensated for in the NO$_2$ retrieval through the cloud (fraction) retrieval and by applying albedo adjustments (see below), which may explain the smaller impacts over Europe and the US.
Figure 17: The difference in tropospheric NO\textsubscript{2} column between the test dataset (named DDS6) for processor version 2.4.0, which uses the TROPOMI DLER database, and the S5P-PAL reprocessed datasets [ER14] based on version 2.3.1, which uses the OMI LER in the NO\textsubscript{2} fit window and the GOME-2 LER for the FRESCO cloud retrieval in the NIR. Note that 10^{15} molec/cm\textsuperscript{2} corresponds to about 16 \mu mol/m\textsuperscript{2}. Data is gridded and averaged for September 2020.

The availability of the TROPOMI version 2 Level-1b radiance/irradiance product with degradation correction will be an important next step to generate an improved DLER surface albedo database. This use of the Level-1b v2 product (available since summer 2021) is important especially for the shorter wavelengths. A version-2 full mission reprocessing of TROPOMI is expected to become available in the second half of 2022.

One more major change in processor version 2.x.y compared to version 1.x.y is in the albedo treatment. As mentioned above, the cloud fraction and cloud radiance fraction are retrieved from the NO\textsubscript{2} fit window. In version 1.x the cloud fractions are clipped to the [0, 1] interval before the AMF is computed. As of version 2.x the albedo is adjusted if the cloud retrieval would give a cloud fraction outside the [0, 1] interval to ensure radiative closure [Van Geffen et al., 2022].

The process is the following: when the reflectivity is lower than expected based on the surface albedo climatology (the OMI LER) then the albedo is adjusted to match the observations, see App. C.1. The adjusted albedo is written to the Level-2 output file, replacing the climatology. This albedo is subsequently used to compute the (tropospheric) AMF. Because the albedo is lowered by this process, the AMF is lower and the approach leads to increased tropospheric NO\textsubscript{2} columns.

The same approach is also applied for very bright scenes, but now the cloud albedo is adjusted. In the cloud fraction retrieval the cloud albedo is fixed to 0.8. When the reflectivity is large than expected for a fully clouded scene, the cloud albedo is increased to match the observation. The new cloud albedo is reported in the Level-2 NO\textsubscript{2} file.

Fig. 18 shows the impact of this procedure for one overpass over Europe. Note that pixels with a cloud cover > 0 are not affected (white parts in the figure). A lowering of the albedo is observed over most of the cloud-free pixels (white areas correspond largely to cloud fractions > 0). A lowering of the albedo by 1\% may already have a significant impact on the retrieved tropospheric NO\textsubscript{2} column.

6.4.6 Snow and ice cover

Substantial errors are introduced if the real albedo differs considerably from what is expected from the albedo climatology, for example in the case of the sudden snowfall or ice cover. Correcting the surface albedo from
Figure 18: Difference in the surface albedo in the NO$_2$ fit window between v1.2.2 and v2.1.0 for orbit 3704 on 1 July 2018 over Europe. Blue colors indicate smaller albedo values in v2.1.0. The white parts correspond largely with retrieved cloud fractions > 0. The red spots over the Alps correspond to snow- or cloud-covered pixels where the albedo was increased.

the climatology (which contains a climatological snow cover) using knowledge of actual snow/ice cover will therefore improve the final data product, in terms of the retrieval itself and for flagging such cases. For the $A_{SN}^{NO_2}$, this correction follows the approach included in the OMI cloud data product OMCLDO2 [Veefkind et al., 2016] to adapt the surface albedo in the O$_2$–O$_2$ fit window (i.e. at 471 nm); Appendix D provides some details on this correction of the surface albedo reported in the Level-2 file.

The baseline for snow/ice cover information as of processor version v2.2.0 is the daily data provided by the ECMWF [Van Geffen et al., 2022]; up to v1.4.0 the NISE snow/ice data [ER17] was used (see also [RD1]). A feature of the NISE data is that the flag is set to 252 when the scene contains both land and water ("mixed pixels at coastlines"), which occurs frequently. In these cases there is no information on snow or ice, and (for high latitudes) the qa_value is set to a small value to reflect this additional uncertainty. The snow/ice information from ECMWF does not have this coastline problem. In addition, the spatial resolution of the ECMWF data is higher than that of NISE. The ECMWF snow data [DeRosnay et al., 2015] is derived from synoptic data and from the Interactive Multi-sensor Snow and Ice Mapping System (IMS); Cooper et al. [2018] show that IMS has better agreement with in situ observations over North America and that NISE misses a significant number of snow-covered pixels. An alternative for the snow/ice cover might come from OSISAF [ER18], but implementation is not straightforward as this are currently separate products for land and for sea.

The FRESCO algorithm (Sect. 6.4.4) provides two sets of data (Wang et al. [2008]; [RD17]): (i) the effective cloud fraction $f_{eff}$ and cloud pressure $p_c$ using a cloud albedo $A_c = 0.8$ (this cloud albedo may adapted by FRESCO over bright scenes), and (ii) the scene albedo $A_{sc}$ and the scene pressure $p_{sc}$ assuming a cloud fraction $f_{eff} = 0.0$. With the snow/ice flag the NO$_2$ processing will select which of these two sets is used for the determination of the AMFs and subsequent vertical NO$_2$ columns. When the snow/ice flag indicates that there is more than a 1% snow/ice coverage, the retrieval will move to scene mode by setting the cloud radiance fraction Eq. (20) equal to 1.0. The $A_{sc}$ and $p_{sc}$ from FRESCO are then used to determine the effective albedo and pressure of this (ficticious) cloud. Which mode is used can be found via the selection criteria of the qa_value definition, listed in Appendix E; see also the NO$_2$ Product User Manual (PUM; available via [ER2]).

6.4.7 Surface pressure

The (altitude-dependent) AMFs in Eq. (13) depend on the surface pressure, $p_s$. For the TROPOMI NO$_2$ retrieval this information is obtained from the TM5-MP model ($1^\circ \times 1^\circ$) driven by ECMWF meteorological data. Because
Figure 19: Tropospheric NO\textsubscript{2} from OMI retrieved with TM5-MP at a resolution of $1^\circ \times 1^\circ$ minus retrieved with TM5-MP at a resolution of $3^\circ \times 2^\circ$ for 20–30 October 2004 over Europe.

The TM5-MP information is representative for spatially coarse pressures, the TM5-MP results are corrected based on the method described in Zhou et al. [2009] and Boersma et al. [2011]. This correction computes a new surface pressure based on the difference between the corresponding spatially coarse terrain height and the actual, pixel-averaged terrain height based on a 3-km resolution digital elevation map [Maasakkers et al., 2013].

6.4.8 A-priori vertical NO\textsubscript{2} profiles

A CTM is considered to be the best source of information for a-priori NO\textsubscript{2} vertical profiles. The baseline for the TROPOMI NO\textsubscript{2} retrieval algorithm is to use TM5-MP [Williams et al., 2017] vertical NO\textsubscript{2} profile shape simulated at a $1^\circ \times 1^\circ$ (longitude $\times$ latitude) spatial resolution for $n_l = 34$ layers covering troposphere and stratosphere. In future updates this layer choice may be further optimised. The a-priori profile shapes are calculated at the centre of the TROPOMI ground pixel via weighted linear interpolation based on the four nearest neighbour TM5-MP cell centres. Using TM5-MP instead of TM4 constitutes a significant improvement in itself: TM5 v3 is a benchmarked recent model version (Huijnen et al. [2010a]; Huijnen et al. [2010b]; [ER13]; Williams et al. [2017]), with more up-to-date NO\textsubscript{x} emissions (from the MACCity inventory), chemistry, and ongoing improvements of ship, soil and lightning NO\textsubscript{x} emission descriptions.

Using TM5-MP with a global $1^\circ \times 1^\circ$ resolution is an important improvement over previous global satellite NO\textsubscript{2} retrievals that used vertical profile shape computed at spatial resolutions of $2^\circ \times 2.5^\circ$ or $3^\circ \times 2^\circ$ (e.g. Lamsal et al. [2010], Boersma et al. [2011]). Obviously, there are spatial gradients in NO\textsubscript{2} concentrations over scales smaller than a degree, but a resolution of $1^\circ \times 1^\circ$ should capture the most relevant gradients much better than a resolution of $3^\circ \times 2^\circ$. Using higher resolution models in combination with the TROPOMI averaging kernels will in effect further improve the spatial resolution in the a-priori NO\textsubscript{2} fields for advanced users interested in regionally focused investigations (e.g. Huijnen et al. [2010b]).

The effect of the improved spatial resolution is illustrated by Fig. 19, which shows the difference between averaged tropospheric NO\textsubscript{2} columns from the OMI sensor from 20–30 October 2004 retrieved with TM5 at $3^\circ \times 2^\circ$ and at $1^\circ \times 1^\circ$. The retrieval with the higher resolution profile shape leads to more pronounced contrasts between the sources of pollution and background (ventilated) pollution. To better capture the sources of air pollution is an important target of the TROPOMI mission.
6.4.9 Averaging kernels

For each ground pixel, the TROPOMI data product provides the corresponding total NO$_2$ column averaging kernel. The averaging kernel for DOAS retrievals is equal to the altitude-dependent AMF ratioed by the total air-mass factor [Esikes and Boersma, 2003]. Furthermore, the height-dependent air-mass factors include a term that corrects for the difference between the temperature of the cross section used in the DOAS fit and the actual temperature in a given layer [RD16] (cf. Sect. 6.4.2).

The averaging kernel as provided in the file is linked to the total column NO$_2$ product. The tropospheric averaging kernel is obtained by scaling the total-column kernel by $M/M^{\text{top}}$ (see [RD21]) and setting all elements of the kernel to zero above the tropopause layer, i.e. for $l > l^{\text{top}}$. This step is explained in the Product User Manual (PUM; available via [ER2]) and has to be implemented by the user. Note that the stratospheric NO$_2$ column reported in the product is derived from the model after assimilation of the TROPOMI measurements. Therefore this quantity does not have a corresponding averaging kernel.

Using the averaging kernel is important for data users who wish to minimise the discrepancies between the assumptions in the TROPOMI retrieval and their application of interest, for example for validation, data assimilation, or comparison to a model (e.g. Silver et al. [2013]; Boersma et al. [2016]). In particular, comparisons that make use of the averaging kernel are no longer depending on the a-priori TM5-MP profile shape [Esikes and Boersma, 2003].

The averaging kernel should be used in validation exercises, model evaluations, and assimilation or inversion attempts with TROPOMI NO$_2$ columns whenever possible (i.e. whenever independent profile information is available). The recipe for using the averaging kernel $A$ for the purpose of obtaining a model estimate of the tropospheric NO$_2$ column ($N_{\text{trop}}^{\text{mod}}$) that can be compared to TROPOMI is as follows:

$$N_{\text{trop}}^{\text{mod}} = A \bar{x}_m = \sum_{l=1}^{n_l} A_l S_l x_{m,l}$$

where $S_l$ are the components at the $l$-th vertical layer of an operator that executes a mass-conserving vertical interpolation, followed by a conversion to sub-columns (molec/cm$^2$) in case the model vertical distribution $x_{m,l}$ is not given in those units.

Alternatively, the kernels may be used to replace the global TM5-MP a-priori profile used in the retrieval by an alternative modelled NO$_2$ profile shape, e.g. from a high-resolution regional chemistry-transport model (Griffin et al. [2019], Lin et al. [2014]). The recipe for this replacement is provided in the PUM [ER2].

6.4.10 De-striping the NO$_2$ data product

The OMI measurements show across-track biases (stripes) in NO$_2$ resulting from viewing zenith angle dependent calibration errors in the OMI backscatter reflectances. For the DOMINO v2 NO$_2$ data product, Boersma et al. [2011] developed an empirical post-hoc de-striping correction based on the daily mean across-track dependency of the NO$_2$ slant columns. This correction is applied in the final step of the NO$_2$ processing, i.e. after the conversion to vertical columns. A new de-striping approach was developed within the QA4ECV project, which is now fully integrated in the retrieval approach, avoiding the extra post-processing step.

Given that TROPOMI is measuring with a CCD detector similar to the one used by OMI, calibration related across-track biases are likely also present in the TROPOMI NO$_2$ data. For this reason an option has been included in the Level-2 processor that allows for a de-striping correction on the NO$_2$ slant column data, similar to the approach implemented for QA4ECV.

Because the stripping amplitude was found to be much smaller than for OMI, the stripe correction was switched off in the first release of the TROPOMI NO$_2$ product, v1.0.2 and v1.1.0. of July 2018. In the first update to v1.2.0 (24 October 2018) it was however decided to turn on the de-striping to remove small but systematic across-track features and further improve the product quality in this way.

The de-striping is determined from on orbits over the Pacific Ocean (longitude between 150°W – 180°W), in order to avoid interference by tropospheric pollution hotspots. Observations are averaged over a 30° latitude range in the tropical belt, which moves north-south with the seasons.

A slant column stripe amplitude is determined for each viewing angle. The row-dependent slant column difference $N_{\text{str}}^{\text{NO}_2}$ is defined as the difference between the measured total slant columns ($N_{\text{s}}^{\text{NO}_2}$) and total slant columns derived from the model profiles ($N_{\text{s}}^{\text{mod}}$) and the averaging kernels ($A$), averaged along-track over the 30° latitude range. The stripe amplitude ($N_{\text{str}}^{\text{NO}_2}$) is then computed as $N_{\text{str}}^{\text{NO}_2} = \text{mean}(N_{\text{str}}^{\text{NO}_2})$, where the mean is taken across the across-track rows. The slant columns are stripe corrected by subtracting this stripe amplitude from the individual slant column observations: $N_{\text{corr}}^{\text{NO}_2} = N_{\text{s}}^{\text{NO}_2} - N_{\text{str}}^{\text{NO}_2}$. The NO$_2$ data product file contains $N_{\text{s}}^{\text{NO}_2}$ and $N_{\text{str}}^{\text{NO}_2}$, so that a user of the slant column data can/must apply the stripe correction.
In order to retain only features which are slowly varying over time, and in order to reduce the sensitivity to features observed during a single overpass, the stripe correction factors are averaged over a time period of 7 days, or about 7 Pacific orbits.

The slant column stripe amplitudes, one for each viewing angle, are stored in three places:

- For off-line (OFFL) processing, in a separate daily data file which contains the stripe amplitudes determined on the previous day. This file is read during every restart of the TM5-MP/DOMINO system in order to initialise the stripe correction. These files are written during the TM5-MP/DOMINO run after every update of the stripe amplitudes, when processing a Pacific orbit.
- The stripe amplitudes are written to the NO\textsubscript{2} Level-2 datafiles.
- The stripe amplitudes are written to the TM5-MP output files, which are used by the near-real time (NRT) NO\textsubscript{2} Level-2 processor.

Fig. 20 shows an example for TROPOMI of the impact of the de-striping of the total NO\textsubscript{2} column (i.e. stratosphere plus troposphere), averaged over the tropical Pacific Ocean on 15 July 2018 (orbit 03747) as a function of row index. The red curve is the v1.1.0 result without de-striping, the blue curve is the v1.2.0 result with the de-striping (the stripe correction is averaged over a week), and the black shows a 50-row running mean of the blue curve. The red curve shows single-row spikes, as well as correlated structures, such as the high values around row 200 and the low values around rows 40, 320 or 420. These correlated features are partly removed with the update of the Level 0-to-1b data to v2.0.0, in use as of NO\textsubscript{2} v2.2.0, but part of the structures remain unexplained. The plot shows that the stripe filtering removes the major part of both the high and low frequency variability. Note that the amplitude of the structures in the red curve is small, generally within 5% of the column over the clean Pacific Ocean. Also note that we expect an increase of the total column in the stratosphere from left to right, as indicated by the black curve, due to the diurnal cycle of stratospheric NO\textsubscript{X} chemistry.

Fig. 21 shows a direct comparison of the NO\textsubscript{2} retrieval – in the form of the “geometric” vertical columns, defined as $N_s/M_{geo}$ – of OMI and TROPOMI as reported by the DOAS fits on actual observations from both instruments. The figure demonstrates that TROPOMI and OMI retrieve roughly the same absolute retrieved NO\textsubscript{2} slant columns, but that TROPOMI has a much smaller across-track variation (“stripiness”). A more detailed comparison of TROPOMI and OMI slant column retrieval results is presented by Van Geffen et al. [2020].
Figure 21: NO$_2$ "geometric" vertical column densities, defined as $N_s/M_{geo}$, for the Pacific Ocean orbit on 20 Feb. 2018 of TROPOMI are compared to those for the almost overlapping OMI orbit (with rows affected by the row anomaly removed) and to those of a similar OMI orbit from 20 Feb. 2005. Same data as in Fig. 26.

6.5 Processing chain elements

6.5.1 Off-line (re)processing

The off-line (re)processing of the TROPOMI NO$_2$ retrieval algorithm, schematically displayed in Fig. 22, takes place at two locations (for more details, see [RD1]):

(1) The first step of the NO$_2$ processing system, illustrated in the top left part of Fig. 22, the DOAS retrieval, ingests the Level-1b spectra and is running in the PDGS TROPOMI processing system at DLR. Also performed in the PDGS, in a separate processing chain (not shown), is the FRESCO cloud retrieval (Sect. 6.4.4), needed by several Level-2 data products. The processor uses the slant column and cloud cover data to assemble a "backup" NO$_2$ vertical column product, based on the TM5-MP NO$_2$ vertical profile forecast produced for the NRT processing at the observation date (cf. Sect. 6.5.2).

(2) The NO$_2$ backup data product is then transferred to the IDAF at KNMI, where once a day the data of all orbits is ingested in the data assimilation / chemistry transport model TM5-MP, as illustrated in the right part of Fig. 22, to compute the off-line NO$_2$ product using the CTM model profiles computed with the latest ECMWF meteorological fields.

(3) The off-line or nominal NO$_2$ data product is then transferred back to the PDGS (bottom left part of Fig. 22), where it is made available for the users.

The motivation for this set-up is to take full advantage of the available processing elements at DLR and KNMI, and at the same time keep the number of data transfers limited. DLR will operate in the PDGS a suite of processors geared to handling large amounts of TROPOMI spectra, including the processing of NO$_2$ column data from TROPOMI spectra. The IDAF at KNMI hosts a complete data assimilation system based on the TM5-MP model, and has considerable experience in both the off-line and on-line retrieval of NO$_2$ from the GOME, SCIAMACHY, OMI, and GOME-2 instruments. The essential inputs for the processing of TROPOMI NO$_2$ data are (1) the Level-1b spectra measured by TROPOMI at the PDGS, and (2) the ECMWF meteo data at the IDAF.

As illustrated by Fig. 22, the data assimilation system not only provides vertical profiles for the processing of NO$_2$ data, but also for other TROPOMI data products: formaldehyde (HCHO) and sulphur dioxide (SO$_2$). Unlike NO$_2$, HCHO and SO$_2$ are not assimilated in the TM5-MP model: their profiles are output of the TM5-MP model, based on the chemistry involving these species.
6.5.2 Near-real time processing

The NRT processing of TROPOMI NO$_2$ is based on the same principles as the off-line processing, described in Sect. 6.5.1. The main difference between the NRT processing, depicted in Fig. 23, and the off-line processing (Fig. 22) is the timing of the data assimilation step and the use of ECMWF meteorological forecasts rather than analysed ECMWF meteorological fields. For the NRT processing of TROPOMI data, the TM5-MP model is run once per day in the IDAF at KNMI, and ingests the NO$_2$ slant columns from the orbits that have been observed thus far. Based on the assimilated “state” of day $i$, the TM5-MP model provides a forecast of the NO$_2$ vertical distribution for days $i+1$ until $i+4$. This information is then transferred to the PDGS, as illustrated in Fig. 23, for the NO$_2$ NRT data product.

This procedure ensures that as soon as new TROPOMI measurements are available in NRT, all necessary information from the TM5-MP model is ready to be processed in the PDGS to provide an NO$_2$ vertical column data product, without the need for a (time consuming) model run first. With NO$_2$ profile data available in a 5-day forecast, an interruption of the data stream from the IDAF system is not an immediate problem for the NRT processing system. In case the interruption lasts longer than 5 days, the PDGS processing system will use the latest available NO$_2$ profile as a fall-back to be able to continue providing NO$_2$ data in NRT. As fall-back the latest available NO$_2$ profile is used, rather than NO$_2$ profile data from a climatology, because a switch to climatology data would constitute an evident discontinuity in the NO$_2$ data.

In the NRT processing, the TM5-MP data assimilation run is started just after midnight, as soon as the ECMWF meteo data has arrived. In that run, the system incorporates all the NO$_2$ slant column data that has been processed since the previous data assimilation run (from 24 hours before). Since the (ECMWF) forecast is provided up to 5 days ahead, the NRT processing is capable of providing tropospheric NO$_2$ data, even after a period of missing data. Previous analysis [Boersma et al., 2007] has shown that the forecast is accurate.
Figure 23: Schematic representation of the TROPOMI processing of tropospheric NO\textsubscript{2} data from a Level-1b spectrum received in the PDGS in the NRT mode. The dotted line marks the division of the processing locations: the Payload Data Ground Segment (PDGS) at DLR on the left and the Instrument Data Analysis Facility (IDAF-L2) at KNMI on the right. (Source of the figure: [RD1].)

enough to provide reliable NO\textsubscript{2} tropospheric columns for a few days ahead. The differences between the off-line and NRT NO\textsubscript{2} product are found to be small.

6.6 The NO\textsubscript{2} data product

The NO\textsubscript{2} vertical column data product contains the data sets listed in Table 5. The main product is the tropospheric NO\textsubscript{2} column, but the file also contains all intermediate steps such as the results from the DOAS retrieval, output from the data assimilation, cloud information, input database information, flags, uncertainties and the AMF calculation results. The attributes in the file provide full traceability of the data product (including information on processor version, settings, inputs).

Table 6 provides a list of seven main classes of possible TROPOMI NO\textsubscript{2} data usage, and lists the data sets that these users will need for their applications. For notes on applying the averaging kernel, see Sect. 6.4.9. More information on the content and usage of the data product can be found in the NO\textsubscript{2} Product User Manual (PUM; available via [ER2]).

In order to comply with the SI unit definitions, the TROPOMI NO\textsubscript{2} data product file gives trace gas concentrations in mol/m\textsuperscript{2}, rather than in the commonly used unit molec/cm\textsuperscript{2}. The following multiplication factors – also provided as attributes to the data sets – enabling the user to easily make the conversions, if needed:

- The multiplication factor to convert mol/m\textsuperscript{2} to molec/cm\textsuperscript{2} is 6.02214 \times 10\textsuperscript{19}.
- The multiplication factor to convert mol/m\textsuperscript{2} to DU is 2241.15.
- The O\textsubscript{2}–O\textsubscript{2} concentration is given in mol\textsuperscript{2}/m\textsuperscript{5}; the multiplication factor to convert this to the commonly used unit molec\textsuperscript{2}/cm\textsuperscript{5} is 3.62662 \times 10\textsuperscript{17}.
Table 5: Overview of data sets for each ground pixel in the final NO$_2$ data product assembled for dissemination via the TROPOMI website, the Sentinel-5P Core Service. Where relevant, the precision of a data set is provided as well. Data sets marked with * are not part of the official Level-2 data product, but will be provided in a separate support data file. A more detailed overview can be found in Tables 11 and 12.

<table>
<thead>
<tr>
<th>origin of data set</th>
<th>for each ground pixel</th>
<th>symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-1b spectrum</td>
<td>measurement time</td>
<td>$t$</td>
</tr>
<tr>
<td></td>
<td>ground pixel centre and corner coordinates</td>
<td>$\vartheta_{\text{geo}}$, $\delta_{\text{geo}}$</td>
</tr>
<tr>
<td></td>
<td>viewing geometry data</td>
<td>$\theta_0$, $\theta$, $\phi_0$, $\phi$</td>
</tr>
<tr>
<td>Databases</td>
<td>surface albedo in the NO$_2$ window</td>
<td>$A_{s,\text{NO}_2}$</td>
</tr>
<tr>
<td></td>
<td>surface albedo used for the cloud retrieval</td>
<td>$A_s$</td>
</tr>
<tr>
<td></td>
<td>surface elevation and pressure</td>
<td>$z_s$, $p_s$</td>
</tr>
<tr>
<td>Cloud retrieval</td>
<td>cloud fraction and cloud pressure FRESCO</td>
<td>$f_{\text{eff}}$, $p_c$</td>
</tr>
<tr>
<td></td>
<td>scene pressure and scene albedo FRESCO</td>
<td>$p_{sc}$, $A_{sc}$</td>
</tr>
<tr>
<td></td>
<td>cloud fraction in the NO$_2$ window</td>
<td>$f_{\text{eff},\text{NO}_2}$</td>
</tr>
<tr>
<td></td>
<td>cloud radiance fraction in the NO$_2$ window</td>
<td>$w_{\text{NO}_2}$</td>
</tr>
<tr>
<td>DOAS retrieval</td>
<td>NO$_2$ slant column</td>
<td>$N_{s,\text{NO}_2}$</td>
</tr>
<tr>
<td></td>
<td>slant columns of secondary trace gases</td>
<td>$N_{s,\text{O}<em>3}$, $N</em>{s,\text{H}_2\text{O}<em>2}$, $N</em>{s,\text{...}}$</td>
</tr>
<tr>
<td></td>
<td>Ring effect coefficient</td>
<td>$C_{\text{ring}}$</td>
</tr>
<tr>
<td></td>
<td>polynomial coefficients</td>
<td>$a_m$ [$m = 0, 1, ..., n_P$]</td>
</tr>
<tr>
<td></td>
<td>intensity offset coefficients</td>
<td>$c_m$ [$m = 0, 1, ..., n_{\text{off}}$]</td>
</tr>
<tr>
<td></td>
<td>number of spectral points</td>
<td>$n_\lambda$</td>
</tr>
<tr>
<td></td>
<td>degrees of freedom of the fit</td>
<td>$D$</td>
</tr>
<tr>
<td></td>
<td>RMS error and $\chi^2$ of the fit</td>
<td>$R_{\text{RMS}}$, $\chi^2$</td>
</tr>
<tr>
<td></td>
<td>wavelength calibration coefficients</td>
<td>$w_{E0}^s$, $w_s^q$, $w_q^s$</td>
</tr>
<tr>
<td>Data assimilation &amp; AMF calculation</td>
<td>NO$_2$ tropospheric vertical column</td>
<td>$N_{v}^{\text{top}}$</td>
</tr>
<tr>
<td></td>
<td>NO$_2$ stratospheric vertical column</td>
<td>$N_{v}^{\text{strat}}$</td>
</tr>
<tr>
<td></td>
<td>NO$_2$ total vertical columns</td>
<td>$N_{v} \equiv N_{v}^{\text{top}} + N_{v}^{\text{strat}}$</td>
</tr>
<tr>
<td></td>
<td>NO$_2$ slant column stripe amplitude</td>
<td>$N_{v,\text{NO}_2}^{\text{str}}$</td>
</tr>
<tr>
<td></td>
<td>NO$_2$ ghost column</td>
<td>$N_{v,\text{ghost}}$</td>
</tr>
<tr>
<td></td>
<td>tropospheric AMF</td>
<td>$M_{v}^{\text{top}}$, $M_{v,\text{clrd}}^{\text{top}}$</td>
</tr>
<tr>
<td></td>
<td>stratospheric and total AMF</td>
<td>$M_{v}^{\text{strat}}$, $M_{v,\text{clrd}}^{\text{strat}}$</td>
</tr>
<tr>
<td></td>
<td>averaging kernel</td>
<td>$A$</td>
</tr>
<tr>
<td></td>
<td>TM5 tropopause layer index</td>
<td>$l_{\text{TM5}}$, $l_{\text{TM5}}^{\text{tp}}$</td>
</tr>
<tr>
<td></td>
<td>TM5 pressure level coefficients</td>
<td>$A_l^{\text{TM5}}$, $B_l^{\text{TM5}}$</td>
</tr>
<tr>
<td></td>
<td>* NO$_2$ profile for stratosphere and troposphere</td>
<td>$v_{l,\text{NO}_2}$</td>
</tr>
<tr>
<td></td>
<td>* TM5 temperature profile</td>
<td>$T_l^{\text{TM5}}$</td>
</tr>
<tr>
<td></td>
<td>* TM5 surface elevation and pressure</td>
<td>$z_{l}^{\text{TM5}}$, $p_{l}^{\text{TM5}}$</td>
</tr>
<tr>
<td>Flags</td>
<td>quality assurance value (qa_value)</td>
<td>$f_{\text{QA}}$</td>
</tr>
<tr>
<td></td>
<td>processing quality flags</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>absorbing aerosol index</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>snow/ice flag and land/water classification</td>
<td>—</td>
</tr>
</tbody>
</table>

The output for each ground pixel is accompanied by two flags indicating the status of the results of the processing and the retrieval. The "quality assurance value" (qa_value or $f_{\text{QA}}$) is a continuous variable, ranging from 0 (no output) to 1 (all is well). Warnings that occur during processing or results of the processing can be reasons to decrease the flag value. The qa_value is the main flag for data usage:

- qa_value > 0.75.

For most users this is the recommended pixel filter. This removes clouds (cloud radiance fraction > 0.5),
Table 6: Overview of different user applications of NO\textsubscript{2} data and the data sets from the TROPOMI NO\textsubscript{2} data product the users will need. In addition all users may need pixel related data, such as measurement time, geolocation, viewing geometry, etc., as well as the processing and data quality flags.

<table>
<thead>
<tr>
<th>user application</th>
<th>data sets needed</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1 Tropospheric chemistry / air quality model evaluation</td>
<td>(N_{\text{trop}}), (\Delta N_{\text{trop}}), \(M_{\text{trop}}^\dagger, M, A, A_{\text{TM5}}^\dagger, R_{\text{TM5}}^\dagger, i_{\text{TM5}}, t_{\text{TM5}}, p_s, f_{\text{QA}})</td>
</tr>
<tr>
<td>and data assimilation</td>
<td></td>
</tr>
<tr>
<td>Validation with tropospheric NO\textsubscript{2} profile measurements</td>
<td></td>
</tr>
<tr>
<td>(aircraft, balloon, MAX-DOAS)</td>
<td></td>
</tr>
<tr>
<td># 2 Tropospheric column comparisons, e.g. with other NO\textsubscript{2} column retrievals</td>
<td>(N_{\text{trop}}), (\Delta N_{\text{trop}}), (f_{\text{QA}})</td>
</tr>
<tr>
<td># 3 Stratospheric chemistry model evaluation and data assimilation</td>
<td>(N_{\text{strat}}), (\Delta N_{\text{strat}}), (f_{\text{QA}})</td>
</tr>
<tr>
<td>Validation with stratospheric NO\textsubscript{2} profile measurements</td>
<td>(N_{\text{strat}}), (\Delta N_{\text{strat}}), (f_{\text{QA}})</td>
</tr>
<tr>
<td>(limb/occultation satellite observations)</td>
<td></td>
</tr>
<tr>
<td># 4 Stratospheric column comparisons, e.g. with ground-based remote sensors</td>
<td>(N_{\text{strat}}), (\Delta N_{\text{strat}}), (f_{\text{QA}})</td>
</tr>
<tr>
<td># 5 Whole atmosphere (troposphere + stratosphere) data assimilation systems</td>
<td>(N_{\text{v}}), (\Delta N_{\text{v}}), (A^\dagger, A_{\text{TM5}}^\dagger, R_{\text{TM5}}^\dagger, i_{\text{TM5}}, t_{\text{TM5}}, p_s, f_{\text{QA}})</td>
</tr>
<tr>
<td># 6 Whole atmosphere (troposphere + stratosphere) comparisons with ground-based remote sensing (e.g. Pandora)</td>
<td>(N_{\text{v}}), (\Delta N_{\text{v}}), (f_{\text{QA}})</td>
</tr>
<tr>
<td># 7 Visualisation of the NO\textsubscript{2} product, as well as generation of Level-3 gridded and time averaged NO\textsubscript{2} fields</td>
<td>(N_{\text{v}}), (\Delta N_{\text{v}}), (f_{\text{QA}})</td>
</tr>
</tbody>
</table>

\(\dagger\) The tropospheric kernel \(A_{\text{trop}}\) is derived from the total kernel \(A\) and the air-mass factors \(M\) and \(M_{\text{trop}}^\dagger\).
\(\dagger\) The stratospheric column is a model (data assimilation analysis) quantity and therefore does not have an averaging kernel. The tropopause pressure can be computed using \(i_{\text{TM5}}, t_{\text{TM5}}\) and the hybrid level specification.
\(\dagger\) Note that the total NO\textsubscript{2} vertical column \(N_{\text{v}}\) is \(\equiv N_{\text{sum}}/M\) is not the same as the sum \(N_{\text{sum}}\) that is \(\equiv N_{\text{trop}} + N_{\text{strat}}\).

scenes covered by snow/ice, errors and problematic retrievals. Default choice for user applications 1, 2, 5, 6, and 7 (viz. Table 6).

- \(qa_value > 0.50\).
  This adds the good quality retrievals over clouds and over scenes covered by snow/ice. Errors and problematic retrievals are still filtered out. In particular this may be useful for assimilation and model comparison studies. Default choice for user applications 3 and 4, and may be considered for user applications 1 and 5 (viz. Table 6).

The determination of the \(qa_value\) is described in detail in Appendix E. The \(qa_value\) indicates whether the footprint is cloud covered or not, and whether there is snow or ice on the surface. It is set to 0 if anywhere in the processing an error occurred, as indicated by the \(\text{processing_quality_flags}\). Warnings related to the South Atlantic Anomaly, sun glint, or missing non-critical input data lower the \(qa_value\). The \(qa_value\) depends on the solar zenith angle, tropospheric air-mass factor and quality of the DOAS fit, and filters unrealistic albedo values.

As of TROPOMI NO\textsubscript{2} product version 1.2 the rather strict filtering of snow/ice covered scenes was abandoned and retrievals receive a \(qa_value > 0.75\) when the scene pressure from the cloud retrieval is close to the surface pressure. This significantly enhances the coverage over high latitudes, as shown in Figure 24.

The \"processing quality flags\" (\(\text{processing_quality_flags}\)) contains the individual event that led to processing failure, and/or a precise record of the warnings that occurred during processing. The definitions and
Figure 24: In TROPOMI NO₂ product version 1.2 (operational since 24 October 2018) the filtering over snow and ice is relaxed, greatly improving the coverage. Left panel: monthly mean NO₂ concentration for April 2018, v1.2; right panel: same but for v1.0.

usage of these two flags is harmonised between the Level-2 data products of TROPOMI and is documented in the NO₂ Product User Manual (PUM; available via [ER2]).

The NO₂ data product provides the Absorbing Aerosol Index (AAI; the ATBD of the AAI is available via [ER1] and [ER5]) and a snow/ice flag (see Sect. 6.4.6) as additional information for the NO₂ data users, both in the off-line and the NRT processing mode. The AAI is not yet used in the flags discussed above, but this may be added in an upcoming update.

The data product consists of two files: one with the main NO₂ retrieval results and a separate TM5-MP model data file with vertical information on atmospheric NO₂, SO₂ and HCHO profile and temperature at the 1° × 1° grid of TM5-MP on a half-hourly basis. The additional model datafile (described in [RD22] and available via [ER5]), is large and is not relevant for most NO₂ data users, but for some advanced users the model profiles have shown to be useful.
7 Input-Output file description

7.1 Required input

The processing of TROPOMI NO$_2$ data poses different demands for different retrieval steps, as described in Sect. 6.5. The dynamic and static input data needed in the PDGS for the off-line and NRT processing of the NO$_2$ data product are listed below and summarised in Tables 7 and 8, respectively. Table 7 also mentions the fall-back in the processing of a given ground pixel in case the dynamical data is not available.

The FRESCO/KNMI cloud product is a TROPOMI Level-2 support product, provided by KNMI software running in the PDGS as separate process with its own Level-2 data files. As of NO$_2$ v2.2.0 an additional cloud product, based on the KNMI O$_2$–O$_2$ cloud algorithm developed for OMI, is produced during the processing for NO$_2$ (cf. Sect. 6.4.4). Both the FRESCO and the O$_2$–O$_2$ cloud data are given in separate data groups in the NO$_2$ product files. The NO$_2$ v2.2.0 data product uses the FRESCO cloud product. For a later release we will study the quality of both cloud products and develop rules to use one or the other cloud product depending on under which circumstances they work best. The S5P/DLR cloud product is optional; the actual use of this product will be investigated.

For the snow/ice cover data, NISE [ER17] and ECMWF assimilated data are requested, at least one is required, with daily updates near the polar region, less frequent updates closer to the equator. (See also the general TROPOMI documents [RD14] and [RD1].)

7.1.1 Inputs at the PDGS for spectral fitting and air-mass factor calculation

In the PDGS at DLR the following input is required, making a distinction between: (a) static (constant) input data and dynamic input data, which changes every orbit, and (b) data needed for the spectral fitting, and information needed in the subsequent processing step. After the DOAS NO$_2$ retrieval, the PDGS assembles the NO$_2$ vertical column data product, using information from the data assimilation system from the IDAF at KNMI for further processing, as illustrated in Fig. 22.

Spectral fitting input data for the DOAS fit

- Dynamic input:
  - Level-1b Earthshine and Solar spectra
- Static input:
  - Reference spectra (convolved with the TROPOMI slit function; [ER11]) for NO$_2$, O$_3$, H$_2$O$_{vap}$, O$_2$–O$_2$, H$_2$O$_{liq}$, Ring effect, irradiance

Note that since the TROPOMI slit function differs for each of the viewing directions, i.e. for each of the detector rows, there is one set of reference spectra for each viewing direction.

NO$_2$ data product input data

- Dynamic input:
  - NO$_2$ slant column density & errors from the DOAS fit
  - NO$_2$ profile shape from the TM5-MP data assimilation system
  - Temperature and pressure profile, orography and tropopause level from TM5-MP / ECMWF
  - Geolocation data (incl. pixel corner coordinates)
  - Viewing geometry
  - Effective cloud fraction and cloud pressure
  - Scene albedo and scene pressure
  - Snow and ice cover data
  - Surface pressure data
- Static input:
  - Pixel-average representative (interpolated) surface albedo at 440 nm (representative for the NO$_2$ window)
  - Pixel-average representative (interpolated) terrain height from a digital elevation map, including a land/water classification
  - Altitude-dependent AMF look-up table
  - Cloud fraction and cloud radiance fraction look-up table
Table 7: Overview of the dynamic input data needed for both the off-line and the NRT NO\textsubscript{2} data processing in the PDGS. The table does not list the meteorological input needed by the data assimilation system in the IDAF. See Sect. 7.1 for further remarks.

<table>
<thead>
<tr>
<th>name/data</th>
<th>symbol</th>
<th>unit</th>
<th>source</th>
<th>pre-process needs</th>
<th>backup if not available</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5P Level-1b Earth radiance VIS band</td>
<td>(l(\lambda))</td>
<td>mol/s/m\textsuperscript{2}/nm/sr</td>
<td>S5P Level-1b product</td>
<td>per pixel</td>
<td>no retrieval</td>
<td>—</td>
</tr>
<tr>
<td>S5P Level-1b Solar irradiance VIS band</td>
<td>(E_{0}(\lambda))</td>
<td>mol/s/m\textsuperscript{2}/nm</td>
<td>S5P Level-1b product</td>
<td>per pixel</td>
<td>use previous</td>
<td>—</td>
</tr>
<tr>
<td>NO\textsubscript{2} profile</td>
<td>(\nu_{\text{NO}_2})</td>
<td>mixing ratio</td>
<td>TM5-MP model</td>
<td>per pixel</td>
<td>latest available (^\dagger)</td>
<td>NRT off-line</td>
</tr>
<tr>
<td>KNMI cloud products</td>
<td>(f_{\text{sat}})</td>
<td>1</td>
<td>S5P Level-2 support product</td>
<td>—</td>
<td>no VCD product</td>
<td>—</td>
</tr>
<tr>
<td>FRESCO &amp; O\textsubscript{2}–O\textsubscript{2} (\dagger)</td>
<td>(p_{c})</td>
<td>Pa</td>
<td>S5P Level-2 support product</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>S5P/DLR cloud product</td>
<td>(f_{\text{sat}})</td>
<td>1</td>
<td>S5P Level-2 cloud product</td>
<td>—</td>
<td>no VCD product</td>
<td>optional</td>
</tr>
<tr>
<td>surface pressure</td>
<td>(p_{s})</td>
<td>m</td>
<td>ECMWF</td>
<td>per pixel</td>
<td>previous day</td>
<td>#</td>
</tr>
<tr>
<td>aerosol absorbing index</td>
<td>AAI</td>
<td>1</td>
<td>S5P Level-2 AAI</td>
<td>per pixel</td>
<td>from DEM (Table 8)</td>
<td>$</td>
</tr>
</tbody>
</table>

\(^{\dagger}\) Latest available forecast NO\textsubscript{2} profile for that day.
\(^{\dagger}\) The FRESCO and O\textsubscript{2}–O\textsubscript{2} cloud product data are both included in the NO\textsubscript{2} data files in data sub-groups of DETAILED_RESULTS; the cloud product selected from these two for a given ground pixels is copied to the data group INPUT_DATA.

\(\#\) If the ECMWF value for the day is not available, the value previous day is used; if that value is unavailable, the snow/ice flag from NISE is used, with ultimate fall-back to a climatological value.

\(\$\) When determining the surface pressure from the DEM surface altitude, ECMWF data (if available) is used correct the DEM data.

7.1.2 Inputs at the IDAF for the data assimilation

In the IDAF at KNMI the NO\textsubscript{2} slant column data received from the PDGS is used in the data assimilation system to determine the NO\textsubscript{2} profile shape needed for the conversion of the NO\textsubscript{2} slant columns from the DOAS fit into the stratospheric and tropospheric NO\textsubscript{2} columns at the PDGS. For this step the following input is required for the data assimilation system.

Data assimilation input data

- Dynamic input:
  - ECMWF meteorological fields (pressure, temperature, wind, \ldots)
  - The "NO\textsubscript{2} data product input data" listed above
  - TM5-MP start field from the previous day
  - (In case de-striping is turned on:) Destriping coefficients from the previous day

- Static input:
  - TM5-MP static input data: NO\textsubscript{x} (and other) emission inventories, climatologies, \ldots

7.2 Computational effort

Table 9 contains an overview of the processing time needed for the NO\textsubscript{2} product; for one pixel the spectral fitting and AMF calculation takes about 0.006 seconds (excluding overhead). The code was developed at KNMI, mainly in C++, and was transferred to and tested at DLR.

Compared to OMI, TROPOMI has about 10 times more observations, which implies a factor 10 extra computing time for the DOAS NO\textsubscript{2} SCD retrieval; the inclusion of the O\textsubscript{2}–O\textsubscript{2} cloud data retrieval as of NO\textsubscript{2} processor v2.2.0 increased the processing time by a factor of about 2.7. Especially the TM5-MP processing of TROPOMI was taking quite some CPU time. In April-May 2018 the AMF retrieval loop inside TM5-MP was
### Table 8: Overview of the static input data needed for both the off-line and the NRT NO\(_2\) data processing in the PDGS. The reference spectra convolved with the TROPOMI slit function (see column 5) are given each of the detector rows. The table does not list the input needed by the data assimilation system in the IDAF. See Sect. 7.1 for further remarks.

<table>
<thead>
<tr>
<th>name/data</th>
<th>symbol</th>
<th>unit</th>
<th>source</th>
<th>pre-process needs</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>absorption cross sections</td>
<td>(\sigma_{\text{NO}_2}(\lambda))</td>
<td>m(^2)/mol</td>
<td>Vandaele et al. [1998]</td>
<td>convolution</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(\sigma_{\text{O}_3}(\lambda))</td>
<td>m(^2)/mol</td>
<td>Gorshelev et al. [2014] &amp; Serdyuchenko et al. [2014]</td>
<td>convolution</td>
<td>—</td>
</tr>
<tr>
<td>(\text{O}_2–\text{O}_2)</td>
<td>(\sigma_{\text{O}_2–\text{O}_2}(\lambda))</td>
<td>m(^2)/mol(^2)</td>
<td>Thalman and Volkamer [2013]</td>
<td>convolution</td>
<td>—</td>
</tr>
<tr>
<td>(\text{H}<em>2\text{O}</em>{\text{vap}})</td>
<td>(\sigma_{\text{H}<em>2\text{O}</em>{\text{vap}}}(\lambda))</td>
<td>m(^2)/mol</td>
<td>HITRAN 2012 data</td>
<td>convolution</td>
<td>†</td>
</tr>
<tr>
<td>(\text{H}<em>2\text{O}</em>{\text{liq}})</td>
<td>(\sigma_{\text{H}<em>2\text{O}</em>{\text{liq}}}(\lambda))</td>
<td>1/m</td>
<td>Pope and Frey [1997]</td>
<td>convolution</td>
<td>†</td>
</tr>
<tr>
<td>Ring reference spectrum</td>
<td>(I_{\text{ring}}(\lambda))</td>
<td>mol/s/m(^2)/nm</td>
<td>Chance and Spurr [1997]</td>
<td>convolution</td>
<td>†</td>
</tr>
<tr>
<td>Irradiance reference spectrum</td>
<td>(E_{\text{ref}}(\lambda))</td>
<td>mol/s/m(^2)/nm</td>
<td>Chance and Kurucz [2010]</td>
<td>convolution</td>
<td>†</td>
</tr>
<tr>
<td>retrieval input settings</td>
<td>—</td>
<td>—</td>
<td>KNMI</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>air-mass factor lookup table</td>
<td>—</td>
<td>—</td>
<td>KNMI</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>cloud fraction lookup table</td>
<td>—</td>
<td>—</td>
<td>KNMI</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>digital elevation map (DEM)</td>
<td>(z_s)</td>
<td>m</td>
<td>GMTED2010 [ER19]</td>
<td>per pixel</td>
<td>§</td>
</tr>
<tr>
<td>surface albedo database</td>
<td>(A_{\text{S,NO}_2})</td>
<td>1</td>
<td>Kleipool et al. [2008]; [ER15]</td>
<td>per pixel</td>
<td>#</td>
</tr>
</tbody>
</table>

† Created e.g. as in Van Geffen et al. [2015]; see also [RD14].
‡ For the cloud fraction retrieval in the NO\(_2\) fit window and for the cloud radiance fraction.
§ Climatological value may be adjusted based on the dynamical snow/ice flag; cf. Sect. 6.4.5 & App. D.
¶ The elevation data is constructed as an average over a 10-km circular region (with standard deviation and extrema), which is representative for the TROPOMI pixel, sampled per 3 km [RD23].

### Table 9: Approximate computational effort for the off-line TROPOMI NO\(_2\) processing for processing ground pixels after the along-track pixel size reduction on 6 Aug. 2019 and including the \(\text{O}_2–\text{O}_2\) cloud retrieval. Any delays introduced by the different processing steps having to wait for data to be available are not included.

<table>
<thead>
<tr>
<th></th>
<th>Time needed for processing one TROPOMI orbit</th>
<th>Time needed for processing one day of TROPOMI data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral fitting &amp; (\text{O}_2–\text{O}_2) clouds &amp; AMF</td>
<td>27 min (9 cores)</td>
<td>6.5 hours (9 cores)</td>
</tr>
<tr>
<td>Data transfer DLR (\rightarrow) KNMI</td>
<td>&lt; 1 min</td>
<td>&lt; 15 min</td>
</tr>
<tr>
<td>AMF/assimilation/modelling (TM5-MP)</td>
<td>4 mins (20 cores)</td>
<td>1 hour (20 cores)</td>
</tr>
<tr>
<td>Data transfer KNMI (\rightarrow) DLR</td>
<td>&lt; 1 min</td>
<td>&lt; 15 min</td>
</tr>
<tr>
<td>Total processing time</td>
<td>35 min</td>
<td>8 hours</td>
</tr>
</tbody>
</table>

### Table 10: Estimate of the computational effort for the near-real time TROPOMI NO\(_2\) processing for processing ground pixels after the along-track pixel size reduction on 6 Aug. 2019 and including the \(\text{O}_2–\text{O}_2\) cloud retrieval. Any delays introduced by the different processing steps having to wait for data to be available are not included.

<table>
<thead>
<tr>
<th></th>
<th>Time needed for processing one TROPOMI orbit in NRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral fitting &amp; (\text{O}_2–\text{O}_2) clouds &amp; AMF</td>
<td>27 min (16 cores)</td>
</tr>
<tr>
<td>Data transfer DLR (\rightarrow) KNMI</td>
<td>&lt; 1 min</td>
</tr>
<tr>
<td>Data assimilation with TM5-MP</td>
<td>N/A</td>
</tr>
<tr>
<td>Data transfer KNMI (\rightarrow) DLR</td>
<td>&lt; 15 min (once a day)</td>
</tr>
<tr>
<td>Total processing time</td>
<td>45 min</td>
</tr>
</tbody>
</table>

parallelised over observations. Together with other optimisations this has resulted in a speed-up of a factor 5, to 1h processing time on 20 processors for one day of TROPOMI retrievals.

The assimilation of 15 million observations per day does not lead to a slowdown of the analysis step
compared to OMI. The TROPOMI pixels are binned to so-called superobservations at $1^\circ \times 1^\circ$ (Sect. 6.3). The number of TROPOMI superobservations is comparable to the number for OMI (before the row anomaly occurred). The number of superobservations to be assimilated is thinned out by a factor 2 (checkerboard approach) to further reduce the computational burden.

### 7.3 Near-real time timeliness

For the NRT Level-2 data to be available within the required 3 hours after measurement, it is required that the processing of Level-2 data does not take more than about 40 minutes per orbit. Table 10 shows that the actual processing time is about 27 minutes (using 16 cores) and is therefore within the NRT constraints.

The data assimilation run is done at KNMI once a day (just after midnight) to provide a forecast of the NO$_2$ profile for the coming 5 days, based on assimilation of TROPOMI slant columns observed over the previous day. These forecast runs have recently been accelerated and take $\sim$3h for one forecast. Note that the NRT chain does not need to wait for this, as mentioned in Sect. 6.5.2.

### 7.4 NO$_2$ product description and size

The TROPOMI NO$_2$ data output product consists of the retrieved tropospheric and stratospheric NO$_2$ columns, along with error estimates and the (total) averaging kernel. A general overview of the data product contents is given in Sect. 6.6 and Table 5. Table 11 provides a more detailed overview of the data sets, their unit, type, etc. in the main output data product.

As of NO$_2$ v2.2.0 the processor includes the O$_2$–O$_2$ cloud retrieval; see Sect. 6.4.4.3. With this processor change two data sub-groups of DETAILED_RESULTS are introduced to hold the FRESCO and the O$_2$–O$_2$ cloud & scene data, named FRESCO and O22CLD, respectively. The cloud data selected from these two for a given NO$_2$ ground pixels are copied to the data group INPUT_DATA. For readability Table 11 lists the cloud & scene data fields only once.

Given the number of data per ground pixel listed in Table 11, the Level-2 NO$_2$ output file of one orbit has the following size: about 340 MB for v1.2.x and v1.3.x prior to the along-track pixel size reduction on 6 Aug. 2019, and after that about 440 MB for v1.3.x and v1.4.0, while v2.2.0 is about 580 MB including the FRESCO and O$_2$–O$_2$ data variables. Before (after) the pixel size reduction, an orbit has about 1.5 (1.9) million observations.

The averaging kernel describes how the retrieved NO$_2$ columns relate to the true NO$_2$ profile shape [Eskes and Boersma, 2003]. The averaging kernel should be used in validation exercises, model evaluations, and assimilation or inverse modelling attempts with TROPOMI NO$_2$ data (cf. Sect 6.4.9). The output product also contains the necessary information (surface pressure and TM5-MP sigma coordinates) to construct the pressure grid to which the averaging kernel values correspond.

For advanced users, a separate support file is made available that contains the temperature and NO$_2$ (SO$_2$, HCHO) vertical profile. This data is given at the TM5-MP grid resolution of $1^\circ \times 1^\circ$, rather than on TROPOMI pixel basis, on a half-hourly basis for one day per file (i.e. 48 time steps); each file is about 1.6GB. The temperature and NO$_2$ profiles are not included in the standard Level-2 product, because most users will not need these and because vertical profiles will drastically increase the size of the TROPOMI Level-2 retrieval files. Table 12 provides an overview of the data sets in the support output data product; a Product User Manual (PUM) for this TM5 NO$_2$, SO$_2$ and HCHO profile auxiliary support product, identified as SSP-KNMI-L2-0035-MA [RD22], is available via [ER2].
Table 11: Overview of the data sets, their units, types and sizes, in the main data output product file, listed alphabetically; cf. Table 5. All quantities followed by a * in the “symbol” column consist of the value and the associated precision (for these the number of data per pixel is doubled in the 6th column); for the vertical column densities the precisions are listed explicitly to clearly show the different types of precisions. See Sect. 7.4 for some remarks on the data from the cloud retrieval. In the last column “PV” denotes the processor version when this variable was introduced. The data sets in the support data file are listed in Table 12.

<table>
<thead>
<tr>
<th>name/data</th>
<th>symbol</th>
<th>unit</th>
<th>description</th>
<th>type</th>
<th>data per pixel</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>aerosol absorbing index</td>
<td></td>
<td></td>
<td>L2 AAI 354/388 nm wavel. pair</td>
<td>float</td>
<td>1</td>
<td>added as flag</td>
</tr>
<tr>
<td>air-mass factor</td>
<td></td>
<td></td>
<td>tropospheric AMF</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>clear-sky tropospheric AMF</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>cloudy tropospheric AMF</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>stratospheric AMF</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>total AMF</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>averaging kernel</td>
<td></td>
<td></td>
<td>float</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>chi-squared</td>
<td></td>
<td></td>
<td>χ²</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>cloud albedo</td>
<td></td>
<td></td>
<td>used in the cloud retrieval</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>cloud fraction</td>
<td></td>
<td></td>
<td>from the cloud retrieval</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>cloud fraction NO₂</td>
<td></td>
<td></td>
<td>for the NO₂ VCD</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>cloud pressure</td>
<td></td>
<td></td>
<td>from the cloud retrieval</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>cloud radiance fraction</td>
<td></td>
<td></td>
<td>for the NO₂ VCD</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>DOAS fit results</td>
<td></td>
<td></td>
<td>of the slant column fit</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>ghost column</td>
<td></td>
<td></td>
<td>NO₂ column below the clouds</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>ground pixel coordinates</td>
<td></td>
<td></td>
<td>δs,NO₂ × 10⁻⁶ mol/m² per mp of NO₂</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>intensity off. coefficients</td>
<td></td>
<td></td>
<td>cₜ × 10⁻⁶</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>land/water classification</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>measurement time</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>number of wavelengths</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>number of iterations</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>polynomial coefficients</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>qa value</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>root-mean-square error</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>satellite coordinates</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>scanline index</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>scene albedo</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>scene pressure</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>snow-ice flag</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>stripe amplitude</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>surface NO₂</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>surface pressure</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>TMS pressure level</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>layer index</td>
<td></td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td>—</td>
</tr>
</tbody>
</table>

Table continues on next page
Table 11: — continued.

<table>
<thead>
<tr>
<th>name/data</th>
<th>symbol</th>
<th>unit</th>
<th>description</th>
<th>type</th>
<th>data per pixel</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertical column density</td>
<td>N_{\text{NO}_2}</td>
<td>mol/m²</td>
<td>tropospheric NO₂ VCD</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>ΔN_{\text{NO}_2}</td>
<td>mol/m²</td>
<td>id. precision, kernel not applied</td>
<td>float</td>
<td>1</td>
<td>cf. Sect. 8.4</td>
</tr>
<tr>
<td></td>
<td>ΔN_{\text{NO}_2}</td>
<td>mol/m²</td>
<td>id. precision, kernel applied</td>
<td>float</td>
<td>1</td>
<td>cf. Sect. 8.4</td>
</tr>
<tr>
<td></td>
<td>N_{\text{NO}_2}</td>
<td>mol/m²</td>
<td>stratospheric NO₂ VCD</td>
<td>float</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>viewing geometry data</td>
<td>θ</td>
<td>°</td>
<td>solar zenith angle</td>
<td>float</td>
<td>1</td>
<td>at surface</td>
</tr>
<tr>
<td></td>
<td>φ</td>
<td>°</td>
<td>solar azimuth angle</td>
<td>float</td>
<td>1</td>
<td>at surface</td>
</tr>
<tr>
<td>wavelength calibration</td>
<td>w_{\text{ns}}</td>
<td>nm</td>
<td>wavelength shift</td>
<td>float</td>
<td>1 × 2</td>
<td>cf. Eq. (7)</td>
</tr>
<tr>
<td>radiance</td>
<td>w_{\text{ns}}</td>
<td>1</td>
<td>wavelength stretch</td>
<td>float</td>
<td>1 × 2</td>
<td>cf. Eq. (7)</td>
</tr>
<tr>
<td>wavelength calibration</td>
<td>\lambda_{\text{ns}}</td>
<td>nm</td>
<td>wavelength shift</td>
<td>float</td>
<td>0 × 2</td>
<td>cf. Eq. (7)</td>
</tr>
<tr>
<td>irradiance</td>
<td>(\lambda_{\text{ns}}^2)</td>
<td>1</td>
<td>\chi^2 of the calibration</td>
<td>float</td>
<td>0</td>
<td>cf. Eq. (7)</td>
</tr>
</tbody>
</table>

1 The number of TM5-MP layers is \( n_l = 34 \).
2 The NO₂ ghost column is the TM5-MP NO₂ profile integrated from the surface to the cloud pressure level.
3 The actual number of wavelengths \( n_s \) used in the fit (cf. Eq. (2)), i.e. after removal of, for example, bad pixels within the fit window.
4 The degree of the DOAS polynomial is \( n_p = 5 \).
5 One set of \( n_l + 1 \) (see note 1) TM5-MP pressure level coefficients per data granule.
6 One set per detector row.

Table 12: Overview of the data set units, types and sizes in the support output product file; this file is also used to store the profiles of HCHO and SO₂, delivered along with the NO₂ profile by the TM5-MP model (see [RD22], available via [ER5]). The data is provided on the TM5-MP grid resolution of \( 1° \times 1° \) on a half-hourly basis, rather than on TROPOMI pixel basis. The data sets in the main data file are listed in Table 11.

<table>
<thead>
<tr>
<th>name/data</th>
<th>symbol</th>
<th>unit</th>
<th>description</th>
<th>type</th>
<th>data per grid cell</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCHO profile</td>
<td>v_1,\text{HCHO}</td>
<td>1</td>
<td>volume mixing ratio</td>
<td>float</td>
<td>( n_l )</td>
<td>?</td>
</tr>
<tr>
<td>NO₂ profile</td>
<td>v_1,\text{NO}_2</td>
<td>1</td>
<td>volume mixing ratio</td>
<td>float</td>
<td>( n_l )</td>
<td>?</td>
</tr>
<tr>
<td>SO₂ profile</td>
<td>v_1,\text{SO}_2</td>
<td>1</td>
<td>volume mixing ratio</td>
<td>float</td>
<td>( n_l )</td>
<td>?</td>
</tr>
<tr>
<td>TM5 temperature profile</td>
<td>\mathcal{T}_{\text{TM5}}</td>
<td>K</td>
<td>—</td>
<td>float</td>
<td>( n_l )</td>
<td>?</td>
</tr>
<tr>
<td>TM5 pressure level coefficients</td>
<td>A_{\text{TM5}}</td>
<td>Pa</td>
<td>—</td>
<td>float</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>B_{\text{TM5}}</td>
<td>Pa</td>
<td>—</td>
<td>float</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>TM5 surface elevation</td>
<td>\mathcal{T}_{\text{TM5}}</td>
<td>m</td>
<td>—</td>
<td>float</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>TM5 surface pressure</td>
<td>p_{\text{TM5}}</td>
<td>Pa</td>
<td>—</td>
<td>float</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>TM5 tropopause layer index</td>
<td>\mathcal{r}_{\text{TM5}}</td>
<td>m</td>
<td>—</td>
<td>int</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>stripe amplitude</td>
<td>N_{\text{NO}_2}^\text{VCD}</td>
<td>mol/m²</td>
<td>NO₂ SCD stripe amplitude</td>
<td>float</td>
<td>0</td>
<td>cf. Sect. 6.4.10</td>
</tr>
<tr>
<td>date &amp; time</td>
<td>—</td>
<td>year, month, day, hour, min, sec</td>
<td>int</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>d</td>
<td>days</td>
<td>int</td>
<td>0</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

1 The number of TM5 layers is \( n_l = 34 \).
2 One set of \( n_l + 1 \) (see note 1) TM5-MP pressure level coefficients per data file.
3 This data set is provided via a separate static TROPOMI digital elevation map file.
4 One set per data file.
5 One set per detector row.
8 Error analysis

The TROPOMI NO$_2$ retrieval algorithm generates stratospheric and tropospheric vertical column densities for all pixels. Since assumptions differ considerably for stratospheric and tropospheric retrievals, the error budget for each case is treated separately below.

The overall error for the retrieved tropospheric columns is determined through propagation of the three main error sources: (a) measurement noise and spectral fitting affecting the slant columns, (b) errors related to the separation of stratospheric and tropospheric NO$_2$, and (c) systematic errors due to uncertainties in model parameters such as clouds, surface albedo, and a-priori profile shape, affecting the tropospheric air-mass factor. For the stratospheric NO$_2$ column, the errors are driven by slant column errors, errors in the estimate of the stratospheric contribution to the slant column, and stratospheric AMF (observation operator) errors.

For NO$_2$, the overall error budget thus consists of several different error source terms. Errors in the slant columns are driven in part by instrumental noise (random errors), and in part by necessary choices on the physical model and reference spectra used (systematic errors). Errors in the AMF are mostly systematic (e.g. assumptions on albedo) but will also have random contributions (e.g. from observed cloud parameters, or sampling / interpolation errors). It is thus not possible to make a clear distinction between these error types in the total error reported in the TROPOMI NO$_2$ data product. This implies that by averaging TROPOMI pixels over time or over a larger area, the random part of the overall error can be largely eliminated, but systematic effects may still persist in averaged retrievals.

Experience with errors in OMI NO$_2$ over polluted regions, largely stemming from theoretical error analysis and practical validation studies, indicates that overall errors on the order of 25% for individual tropospheric NO$_2$ column retrievals may be expected. Validation studies show that the systematic part of this error is on the order of 10-15% (e.g. Hains et al. [2010]; Irie et al. [2012]; Ma et al. [2013]). For stratospheric NO$_2$ columns, the errors are considerably smaller and depend mostly on the accuracy of the slant columns, and on the separation of the stratospheric and tropospheric contributions. The stratospheric NO$_2$ column error is expected to have errors on the order of 5-10% (e.g. Hendrick et al. [2012]) or 0.15 – 0.2 × 10$^{15}$ molec/cm$^2$.

8.1 Slant column errors

Instrument noise is the main source of errors in the spectral fitting of TROPOMI Level-1b spectra. The radiometric signal-to-noise ratios (SNR) of TROPOMI in the 400 – 500 nm range turns out to be 1400 – 1500 for a individual Level-1b spectra [RD4]. Experience with OMI spectral fitting in the 405 – 465 nm spectral domain showed that the uncertainty in OMI NO$_2$ slant column densities of about 0.75×10$^{15}$ molec/cm$^2$ in 2005 (when the SNR of OMI was 900 – 1000) to about 0.90×10$^{15}$ molec/cm$^2$ in 2015 (Boersma et al. [2007], Zara et al. [2018]). Van Geffen et al. [2020] show a comparison between TROPOMI and OMI slant column errors and statistical uncertainties. Other, potentially systematic, errors include inaccuracies in the NO$_2$ cross-section, but these have been shown to be of little concern for the slant column errors [Boersma et al., 2002].

Fig. 25 shows as function of the SNR an estimate of the uncertainty of the retrieved slant column density determined by a DOAS fit in the wavelength window 405 – 465 nm with polynomial degree 5. Spectra were simulated with a radiative transfer code using an atmosphere with two NO$_2$ profiles, taken from the CAMELOT study [RD9], with the same profile shape in the stratosphere:

(a) European background profile, simulated with a total vertical column $N_v = 2.5 × 10^{15}$ molec/cm$^2$

(b) European polluted profile, simulated with a total vertical column $N_v = 7.5 × 10^{15}$ molec/cm$^2$

The simulations are performed with surface albedo $A_s = 0.05$, no clouds, solar zenith angle $\theta_s = 50^\circ$, and looking down in nadir. The legend of Fig. 25 gives the total slant column $N_s$ in 10$^{15}$ molec/cm$^2$. The retrieved $N_s$ varies very little with the SNR: about 3 × 10$^{12}$ molec/cm$^2$ between SNR = 700 and 1100. For profile (a) the retrieved $N_s$ is within 5% of the initial $N_v$ and for profile (b) it is within 3%. Given this, a good accuracy of the DOAS fits can be expected, with uncertainties in the range of 10 – 15% for background NO$_2$ cases and 5 – 10% for polluted cases.

Fig. 26 shows a direct comparison of the NO$_2$ slant column error of OMI and TROPOMI as reported by the DOAS fits on actual observations from both instruments. The figure demonstrates that TROPOMI has an SCD error of about 8 – 10 μmol/m$^2$, or 0.5 – 0.6 × 10$^{15}$ molec/cm$^2$. The OMI noise level in 2005 is about 40% higher than what is observed with TROPOMI. This is in agreement with the theoretical dependence on SNR described above. Some further comparisons are reported by Van Geffen et al. [2020].

During the TROPOMI commissioning phase it became clear that over bright scenes, e.g. high thick clouds in the tropics, the measurements at the visible wavelengths may become saturated. One very sensitive parameter
Figure 25: DOAS retrieval slant column uncertainty estimate [in $10^{15}$ molec/cm$^2$] as function of the SNR for two NO$_2$ profiles. The plot legend gives the retrieved slant column in $10^{15}$ molec/cm$^2$. At SNR equal 800 and 1000 the relative slant column uncertainty is indicated. For further details see the text.

Figure 26: NO$_2$ slant column error estimates for the Pacific Ocean orbit on 20 Feb. 2018 of TROPOMI are compared to those for the almost overlapping OMI orbit (with rows affected by the row anomaly removed) and to those of a similar OMI orbit from 20 Feb. 2005. Data of scanlines with nadir latitude in the range $[−20°, +20°]$ is averaged along-track; OMI data is processed within the QA4ECV project [RD6], [ER7].

to detect saturation is the NO$_2$ slant column uncertainty, and therefore this is used as one of the quality criteria (qa_value, see Appendix E): when the slant column uncertainty exceeds 33 µmol/m$^2$, or $2 \times 10^{15}$ molec/cm$^2$, the pixel is flagged as bad quality.

8.2 Errors in the stratospheric (slant) columns

Data assimilation of TROPOMI NO$_2$ slant columns in TM5-MP provides the estimate of the stratospheric contribution to the NO$_2$ slant columns. The accuracy of these estimates is largely determined by the accuracy of the slant columns, as the TM5-MP stratospheric NO$_2$ distributions are scaled to become consistent with the
retrieved slant columns. Random error estimates are derived from the assimilation approach: a considerable advantage of the assimilation scheme is that it provides a statistical estimate of the uncertainties in the stratospheric (slant) columns through the standard deviation of the differences between the TM5-MP model analysis and forecast stratospheric NO\textsubscript{2} ("A–F"). Generally, the uncertainty for the stratospheric NO\textsubscript{2} columns is of the order of $0.1 - 0.2 \times 10^{15} \text{molec/cm}^2$, similar to OMI [Dirksen et al., 2011]. This similarity with OMI is partly the result of using superobservations, which reduces the random contribution to the errors in the stratospheric slant column estimates. Fig. 7, bottom panel, shows the average A–F difference for 1 April 2018 in the data assimilation system based on TM5-MP. The A–F differences are on average $0.15 \times 10^{15} \text{molec/cm}^2$, and O–F over unpolluted scenes are about $0.2 \times 10^{15} \text{molec/cm}^2$. The latter is used as estimate of the uncertainties of the stratospheric NO\textsubscript{2} columns.

Forward (radiative transfer) model calculations are important for, but contribute little to errors in the assimilation procedure. The observation operator $H$ (see Eq. (10)) is proportional to the averaging kernel [Eskes and Boersma, 2003], the vector that contains the vertical sensitivity of TROPOMI to NO\textsubscript{2} in each layer. The scalar product of the observation operator vector and the TMS-MP NO\textsubscript{2} profile at the location of the individual TROPOMI observations yields the slant column that would be observed by TROPOMI given the modeled profile. Stratospheric radiative transfer calculations around 435 nm are relatively straightforward compared to those for the troposphere, where multiple scattering occurs, and the effects of clouds and aerosols interact with the vertical distribution of NO\textsubscript{2}. The main forward model parameter influencing errors in the stratospheric estimate is the a-priori stratospheric NO\textsubscript{2} profile shape (and associated temperature correction), but sensitivity tests suggest that uncertainties in the exact shape of this profile are of little influence to the overall error of the stratospheric NO\textsubscript{2} column.

One potential source of error is the sphericity correction in the radiative transfer model. These errors are negligible for most viewing geometries, but need to be considered for far off-nadir viewing angles and high solar zenith angles. Lorente et al. [2017] investigated the differences between stratospheric NO\textsubscript{2} AMFs calculated with a model simulating radiative transfer for an atmosphere spherical for incoming, single-scattered, and multiple-scattered light (McArtim), and a model with an atmosphere that is spherical for incoming light, but plane-parallel for scattered sunlight. When solar and viewing zenith angles are both large, the DAK model overestimates the stratospheric AMFs by 5-10%. For TROPOMI, we therefore use an AMF LUT that is based on DAK radiative transfer simulations, but whose values for extreme viewing geometries have been made consistent with the McArtim simulations. This is the same AMF LUT that is being used in the QA4ECV retrievals of NO\textsubscript{2} from OMI and GOME-2A ([RD6], [ER7]).

### 8.3 Errors in the tropospheric air-mass factors

The tropospheric AMF is calculated with a forward model (here version 3.2 of the DAK radiative transfer model) and depends on the a-priori assumed profile shape and forward model parameters (cloud fraction, cloud pressure, surface albedo, surface pressure and aerosol properties). The AMF also depends on the solar zenith, viewing zenith and relative azimuth angles, but the measurement geometry is known with high accuracy and therefore does not contribute significantly to the AMF errors. The forward model itself is assumed to represent the physics of the measurement accurately, so that forward model errors can be characterised in terms of model parameters only.

The most important AMF errors are cloud fraction, surface albedo, and a-priori profile shape. Cloud parameters are obtained from TROPOMI observations, and these have random as well as systematic components. Surface albedo and NO\textsubscript{2} profile shape are obtained from a-priori assumptions (i.e. a pre-calculated climatology and CTM simulations, respectively), and much depends on the accuracy of these assumptions that are different for different retrieval situations (e.g. season, surface type etc.). Because the retrieved cloud fraction depend on similar (if not the same) surface albedo assumptions as the NO\textsubscript{2} air-mass factors, errors will be dampened to some extent [Boersma et al., 2004].

In Table 13 the most probable uncertainties of the forward model parameters to provide a cautious error prediction for TROPOMI NO\textsubscript{2} AMFs are listed. For this the theoretical error propagation framework used in Boersma et al. [2004] is followed. This approach takes into account the sensitivity of the AMF to uncertainties around the actual value of a particular forward model parameter (e.g. the AMF is much more sensitive to albedo errors for dark surfaces than for brighter surfaces).

Aerosol-related errors are intimately coupled to cloud parameter errors. The O\textsubscript{2} A-band cloud algorithm currently does not correct for the presence of aerosols, so that an effective cloud fraction and cloud pressure are retrieved; the same holds for the cloud fraction in the NO\textsubscript{2} window, which is computed in the same manner as the O\textsubscript{2} A-band cloud fraction (Sect. 6.4.4). It is a matter of ongoing research whether or not the
Table 13: Estimate of the contributions to the error in the AMF due to individual error sources (‘BL’ stands for Boundary Layer.) The estimated AMF errors are considered to be representative of ‘typical’ retrieval scenarios over regions of interest, i.e. with substantial NO₂ pollution for mostly clear-sky situations, and non-extreme boundary conditions for surface albedo and pressure. Note that the uncertainties can be substantially larger for specific condition, e.g. for very small albedo and large SZA.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Estimated error</th>
<th>Corresponding AMF error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud fraction</td>
<td>±0.025</td>
<td>±10%</td>
</tr>
<tr>
<td>Cloud pressure</td>
<td>±0.50 hPa</td>
<td>±[0 - 10]%</td>
</tr>
<tr>
<td>Surface albedo</td>
<td>±0.015</td>
<td>±[0 - 10]%</td>
</tr>
<tr>
<td>Surface pressure</td>
<td>±0.20 hPa</td>
<td>±[0 - 5]%</td>
</tr>
<tr>
<td>A-priori NO₂ profile shape</td>
<td>BL height &amp; mixing schemes &amp; free troposphere &amp; emissions</td>
<td>±0 - 25]%</td>
</tr>
<tr>
<td>A-priori NO₂ emissions</td>
<td>±0.015</td>
<td>±[0 - 10]%</td>
</tr>
<tr>
<td>Aerosol-related errors</td>
<td></td>
<td>±[0 - 10]%</td>
</tr>
<tr>
<td>Overall error</td>
<td></td>
<td>±[15 - 25]%</td>
</tr>
</tbody>
</table>

The overall error variance for each pixel is written as in Boersma et al. [2004]:

\[
\langle \varepsilon^2 \rangle = \left( \frac{\sigma(N_s)}{M_{\text{trop}}} \right)^2 + \left( \frac{\sigma(N_{s,\text{strat}})}{M_{\text{trop}}} \right)^2 + \left( \frac{(N_s - N_{s,\text{strat}}) \cdot \sigma(M_{\text{trop}})}{(M_{\text{trop}})} \right)^2
\]  

(22)

with \(\sigma(N_s)\) the slant column error, \(\sigma(N_{s,\text{strat}})\) the stratospheric slant column error and \(\sigma(M_{\text{trop}})\) the estimated error in the tropospheric air-mass factor (±25%). The total error depends on details in the retrieval and therefore differs from one pixel to the next. For small tropospheric excess slant columns, the overall retrieval uncertainty is dominated by the random errors in spectral fitting, whereas for large tropospheric slant columns, the retrieval uncertainty is dominated by air-mass factor uncertainties (the last term in Eq. (22)).

Fig. 27 shows the absolute and relative error in the tropospheric NO₂ column retrieved for clear-sky scenes from TROPOMI data on 17 April 2018. We see that over the oceans and the remote continental regions, the overall tropospheric retrieval uncertainty is dominated by errors in the spectral fitting and the stratospheric column estimate and is typically more than 100% (indicated by purple colours in the bottom panel of Fig. 27).
Figure 27: TROPOMI tropospheric NO\textsubscript{2} vertical column values (top panel; in 10\textsuperscript{15} molec/cm\textsuperscript{2}), the corresponding absolute error estimate (middle panel; in 10\textsuperscript{15} molec/cm\textsuperscript{2}; note that the scale range is reduced by a factor 4), and the relative error (bottom panel; in \%) for 17 April 2018. Large relative errors are seen mostly over areas with small NO\textsubscript{2} column values: oceans and remote continental regions. These errors reflect uncertainties in the slant and stratospheric column. Over the very polluted hotspots typical errors are in the 25 – 40\% range, reflecting uncertainties in the air-mass factor.
Table 14: Relative tropospheric NO$_2$ vertical column per pixel uncertainty due to the tropospheric AMF uncertainty only. Estimates based on QA4ECV OMI NO$_2$ data for selected regions for the year 2005, taken from Boersma et al. [2018].

<table>
<thead>
<tr>
<th>Region</th>
<th>Average AMF uncertainty</th>
<th>box size ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>longitude</td>
</tr>
<tr>
<td>China</td>
<td>17 - 22 %</td>
<td>110 : 140</td>
</tr>
<tr>
<td>USA</td>
<td>17 - 27 %</td>
<td>−100 : −75</td>
</tr>
<tr>
<td>Europe</td>
<td>18 - 26 %</td>
<td>−10 : 15</td>
</tr>
<tr>
<td>Johannesburg</td>
<td>15 - 20 %</td>
<td>26 : 30</td>
</tr>
</tbody>
</table>

For larger columns over continental areas, the relative uncertainty in the retrieved column reduces to 15 – 50%, and is dominated by the uncertainty in the tropospheric air-mass factor. Retrieval results are generally best for regions with strong NO$_2$ sources and/or high surface albedos.

Based on the instrumental performance for TROPOMI, and our experience with OMI tropospheric NO$_2$ retrievals (see Fig. 27 and Table 14), the overall error budget for individual TROPOMI tropospheric NO$_2$ retrievals can tentatively be approximated as $\varepsilon = 0.5 \times 10^{15}$ molec/cm$^2$ $+ [0.2$ to $0.50] \cdot N_{\text{trop}}$. This is a more complete and realistic error statement than the requirements from [RD5] ($\varepsilon = 1.3 \times 10^{15}$ molec/cm$^2$ $+ 0.1 \cdot N_{\text{trop}}$ for a horizontal resolution of 5 – 20 km; cf. Table 1).

The error components can be split in two classes: input parameter plus DOAS related uncertainties (cloud, albedo, aerosol, stratosphere, slant column) and a-priori related uncertainties (profile shape). In Rodgers optimal estimation formalism [Rodgers, 2000] the latter may be called the smoothing error. It depends on the use of the data which uncertainty should be used. When the NO$_2$ vertical columns are used without knowledge of the NO$_2$ profiles, then the uncertainty, $\Delta N_{\text{trop, kernel}}$, is the sum of input parameter, DOAS and smoothing. When profile information is available (e.g. when comparisons with models are performed) and the kernels are used, the uncertainty, $\Delta N_{\text{trop}}$, is the sum of input parameter and DOAS only, without the smoothing error contribution. Both uncertainty estimates for the tropospheric vertical column are made available in the product: one for applications with the kernel, one for applications without.

The individual components of the total uncertainty of the tropospheric column are available in the code and provided in the NO$_2$ data files of the QA4ECV project ([RD6], [ER7]). In the current TROPOMI NO$_2$ processor (mid-2017) only the total error is made available in the data product. In the next upgrade we may consider to add the tropospheric column error components due to the slant column uncertainties, stratospheric estimate and the air-mass factor, and contributions of this AMF uncertainty due to cloud fraction, cloud pressure, albedo and profile shape uncertainties.
9 Validation

9.1 Routine validation & validation activities

The routine validation of TROPOMI is organised through the S5P Mission Performance Centre (MPC), see [ER21]. Since November 2018 the MPC generates routine validation results for the TROPOMI Level-2 products in the form of up-to-date validation results and consolidated quarterly validation reports (ROCVR, available via [ER4]). These validation activities are coordinated through the MPC VDAF website [ER22]. Further validation is performed by the S5P Validation Team (S5PVT) members.

The MPC validation activities include comparisons with MAX-DOAS and PANDORA observations (to evaluate the tropospheric column), SAOZ observations (to evaluate the stratospheric column), and satellite observations (comparisons with OMI in particular); see, for example, Griffin et al. [2019], Ialongo et al. [2020], Van Geffen et al. [2020], Verhoelst et al. [2021], Zhao et al. [2020], Van Geffen et al. [2022]. On top of the routine MPC activities, the TROPOMI data has been and will be compared to any campaign data organised by Europe and partners outside Europe.

9.2 Algorithm testing and verification

Algorithm testing and verification by the S5PVT before launch provided confidence in the retrieval algorithms, including forward and inverse models, based on simulations, and comparisons between different techniques and software programs, as described in [RD24]. That activity also included reviews and updates of the TROPOMI NO\textsubscript{2} ATBD.

Verification covers a wide range of activities, including:

- Testing of all the individual input datasets (albedo, cloud parameters, surface properties, snow and ice data) and comparisons with alternative input datasets and measurements when available.
- Testing of the air mass factor calculations by comparing with alternative radiative transfer calculations [Lorente et al., 2017].
- Comparisons with alternative retrieval approaches, such as the scientific retrievals performed by the university of Bremen, or approaches that start from the operational TROPOMI data and aim to improve the air-mass factors (Liu et al. [2020]; Griffin et al. [2019]).
- Study the impact of alternative (high-resolution) a-priori profile shapes from for instance regional air quality modelling systems (Laughner et al. [2019]; Marecal et al. [2015]; Liu et al. [2020]; Griffin et al. [2019]).

9.3 Stratospheric NO\textsubscript{2} validation

For stratospheric NO\textsubscript{2} columns, correlative (column and profile) measurements are needed in regions that are representative for a complete zonal band, and hence need to be relatively unpolluted. The currently operational NDACC ZSL-DOAS/SAOZ instruments are a key dataset to evaluate the stratospheric column (e.g. Verhoelst et al. [2021], Dirksen et al. [2011] and the S5P MPC validation reports [ER4]). Additional routine information comes from the Pandonia Global Network (PGN) remote and/or high altitude stations.

In our view a priority of the stratospheric validation efforts should be a better characterisation of the spatial and seasonal variations of the vertical profile of stratospheric NO\textsubscript{2}. These profile shapes are an essential input to the data assimilation system in use for the separation between tropospheric and stratospheric NO\textsubscript{2} columns from the TROPOMI measurements. Useful sources of stratospheric NO\textsubscript{2} profile data are satellite instruments that measure in limb view, SCIAMACHY (Beirle et al. [2010], Hilboll et al. [2013b]), HIRDLS and MLS [Belmonte et al., 2014], OSIRIS [Adams et al., 2016]. Note that there are difficulties in using these for direct validation as they are often only sparsely validated themselves.

Stratospheric NO\textsubscript{2} measurements near the Arctic vortex in late winter and early spring would be useful to better test the capability of the data assimilation scheme (and other stratosphere-troposphere separation schemes) in capturing the influence of stratospheric air masses low in NO\textsubscript{x} on stratospheric NO\textsubscript{2} at lower latitudes. Such excursions are known to occur and may lead to systematic errors in the separation scheme (e.g. Dirksen et al. [2011]; Bucsela et al. [2013]). Independent measurements may provide further important information on how to improve these issues in the future.
Tropospheric NO\textsubscript{2} validation

Information on tropospheric NO\textsubscript{2} concentrations – with the NO\textsubscript{2} in the planetary boundary layer and/or in the free troposphere – comes from in-situ instruments (at the ground, in masts, on aircraft or on low-flying balloons) and from remote-sensing instruments at the ground, on balloons or aircraft.

For the validation of TROPOMI tropospheric NO\textsubscript{2} columns, correlative (column and profile) measurements are needed in the highly populated polluted regions at mid-latitudes, and also in regions with natural sources of nitrogen oxides, e.g. from biomass burning, microbial soil activity and lightning. The expanding networks of MAX-DOAS and PANDORA (PGN) instruments are used as backbone for tropospheric column validation in the MPC VDAF validation reports ([ER4]; Verhoelst et al. [2021]).

Aircraft remote sensing tropospheric column observations with high spatial resolution mapping instruments is also a very valuable source of validation data for TROPOMI, (e.g. Nowlan et al. [2016], Judd et al. [2020], Tack et al. [2021]). In this way the impact of the fine-scale horizontal variability in NO\textsubscript{2} can be quantified.

The local overpass time of OMI and TROPOMI are nearly identical. This implies that a direct comparison of OMI and TROPOMI for (nearly) overlapping orbits is a key element of the TROPOMI validation and is also reported in the MPC VDAF validation reports [ER4]. An early example of this is shown in Fig. 28. The figure demonstrates a good quantitative agreement between OMI and TROPOMI measurements, but also demonstrates the good agreement between the OMI-QA4ECV and TROPOMI retrieval software.

Important for the validation as well as for the data assimilation system in use for the separation between tropospheric and stratospheric NO\textsubscript{2} columns from the TROPOMI measurements is a good understanding of the vertical profile of the tropospheric NO\textsubscript{2}. The best source of information on vertical profiles of NO\textsubscript{2} is
still from incidental aircraft campaigns. Apart from this, routine profile measurements with aircraft, such as provided by a recent extension of the IAGOS programme [ER23] are very valuable. Alternatively, experimental NO₂ profiles from (tethered) balloon sondes and measurement towers, will provide valuable information on the vertical distribution of NO₂.

Since tropospheric retrievals depend on the concept of the air-mass factor, which has to rely on a-priori information, it is important to also validate the inputs and assumptions that go into the air-mass factor calculation. This mostly concerns cloud parameters – cloud fraction and cloud pressure – that should be well characterised. Another critical issue, about which very little is known as yet, is the effect of the presence of aerosols on the NO₂ retrieval. Collocated information on the aerosol profile – e.g. coming from the TROPOMI Aerosol Layer Height data product – could be useful for this. There is also a need for correlative surface albedo data to investigate the accuracy of the surface albedo climatology.

Despite the routine validation activities, based on a large number of measurement instruments and sites, the quantification of the biases in the TROPOMI retrievals remains a difficult task. This difficulty is related to several factors, e.g. the uncertainty of the independent surface remote sensing NO₂ data used for the validation, large differences in the sensitivity profiles between e.g. MAXDOAS and TROPOMI, and the issue of representativity of the local ground-based and in-situ measurements w.r.t. the finite-sized satellite ground pixels (of order 5 × 5 km²), given the large fine-scale variability of NO₂ close to the surface.


### 10 Conclusion

We have presented the baseline approach for the retrieval of the operational tropospheric and stratospheric NO\textsubscript{2} column products from the TROPOMI sensor. The NO\textsubscript{2} data are delivered both as an off-line product for the NO\textsubscript{2} data record and as a near-real time product, with the NO\textsubscript{2} data delivered within 3 hours after observation. The TROPOMI NO\textsubscript{2} data products pose an improvement over previous NO\textsubscript{2} data sets, particularly in their unprecedented spatial resolution (approximately 7.0 × 3.5 km\textsuperscript{2} or, since the along-track pixel size reduction on 6 Aug. 2019, 5.5 × 3.5 km\textsuperscript{2} at nadir), high signal-to-noise and daily global coverage, but also in the separation of the stratospheric and tropospheric contributions of the retrieved slant columns, and in the calculation of the air-mass factors used to convert slant to total columns.

The backbone of the retrieval system is the TM5-MP chemistry transport model, that is operated at a global resolution of 1° × 1°. The assimilation of NO\textsubscript{2} slant columns in TM5-MP ensures that the modelled stratospheric state becomes consistent with the TROPOMI slant columns over regions with small tropospheric NO\textsubscript{2} amounts. The information from the data assimilation system is used to separate the slant column into its stratospheric and tropospheric components and to provide the a-priori NO\textsubscript{2} vertical profile required by the air-mass factor calculation.

For each TROPOMI pixel an air-mass factor (AMF) is calculated, using altitude-dependent AMFs from a look-up table calculated with the DAK radiative transfer model, in combination with the vertical distribution of NO\textsubscript{2} provided by the TM5-MP chemistry transport model (in assimilation mode) at a spatial resolution of 1° × 1°. The AMF calculation uses local surface albedos from the OMI surface reflectance climatology that is based on 5 years of OMI measurements. It accounts for cloud scattering using information on effective cloud fraction and cloud pressure retrieved for every TROPOMI pixel from the reflectance at the AMF wavelength and from the FRESCO retrieval algorithm, respectively.

Several additional algorithm improvements w.r.t. the OMI / DOMINO v2 processing have been implemented, such as the inclusion of additional reference spectra in the DOAS spectral fit to improve the accuracy of the retrieved NO\textsubscript{2} slant columns, major updates to the data assimilation / chemistry transport model used to determine the vertical column densities, and a more careful quality filtering of the measurements, reflected in the qa_value. Part of these improvements were developed during the European QAECV project. Residuals resulting from tests with the TROPOMI prototype fitting algorithm on OMI spectra suggest a need to include absorption by liquid water, in any case over cloud-free ocean scenes without substantial oceanic chlorophyll. Revisiting the OMI spectra also re-emphasised the importance of an appropriate spectral calibration that is representative for the complete fitting window. Using TM5-MP at a spatial resolution of 1° × 1° (instead of a lower spatial resolution) has been shown to provide more accurate estimates of the NO\textsubscript{2} profiles. The conversion of the slant to vertical columns has been improved by using an air-mass factor look-up table with more nodes, in order to reduces interpolation errors.

The TROPOMI NO\textsubscript{2} processing chain enables us to provide a realistic error budget. The retrieval error is dominated by the spectral fitting error over oceans and regions with low tropospheric NO\textsubscript{2} amounts. Over the polluted regions, air-mass factor errors contribute substantially to the overall error, which can be generally approximated as 0.5 × 10\textsuperscript{15} molec/cm\textsuperscript{2} + 25% for an individual pixel.

Besides a complete error analysis, the TROPOMI data product also provides the averaging kernel, which describes the sensitivity of TROPOMI to NO\textsubscript{2} in each model layer, for every pixel. The averaging kernel is especially relevant for data users who wish to minimise the discrepancies between the assumptions in the TROPOMI retrieval and their own application of interest, e.g. for data assimilation, validation, or comparison studies.

TROPOMI’s high spatial resolution enables monitoring NO\textsubscript{2} columns with an unprecedented accuracy, both in the troposphere and the stratosphere. From these measurements we learn more about the distribution of NO\textsubscript{2}, its sources and sinks, its transport through the atmosphere, its role in stratospheric and tropospheric chemistry, as well as in climate issues, notably through the important role that nitrogen oxides play in the formation of secondary pollutants ozone and aerosol. The early-afternoon NO\textsubscript{2} data record, which started with OMI, is extended by TROPOMI, alongside the mid-morning measurements of the GOME-2 instruments, and Sentinel-5 in the near future, thus providing essential information on the diurnal cycle of NO\textsubscript{2}. Over the past 20-odd years various UV/Vis backscatter instruments have been used to monitor NO\textsubscript{2} on a global scale. The operational TROPOMI NO\textsubscript{2} data processing is consistent with the NO\textsubscript{2} retrieval record generated at KNMI, and will continue and improve that record.
A  Wavelength calibration

The S5P/TROPOMI radiance and irradiance spectra in the Level-1b input data are not wavelength calibrated, because most Level-2 processors perform a wavelength calibration on the fitting window specific to the algorithm. If the calibrations had been done in the Level 0-to-1b processor, they would have been done for the whole spectral band, and this may or may not have met the science requirements for the wavelength calibration for the Level-2 trace gas retrievals. For the wavelength calibration of the radiances an atmosphere model is needed, especially at the shorter wavelengths where ozone absorption is significant, but also the Ring effect modifies the radiance spectra in ways that have to be taken into account when calibrating the wavelength. In the unlikely case the wavelength calibration of radiance or irradiance fails, the retrieval will be performed using the respective nominal wavelength grids.

For the calibration of a complete band or a complete detector, the calibration is split up in micro-windows, and a polynomial is drawn through the micro-windows to cover the whole band. When fitting for a specific retrieval window, a single fit covering the retrieval window is more appropriate. The model function that is used for the radiance wavelength calibration is a modified version of a DOAS fit. Sections A.1 and A.2 describe the generic wavelength fit used in most retrieval algorithms for S5P/TROPOMI, in section A.3 the actual application to NO₂ retrieval is discussed. For the irradiance calibration the same procedure is used, except that atmosphere related effects should be disabled, specifically the Ring effect should not be included in this fit.

The polynomial order \( N \) is set to 1 for the irradiance fit.

### A.1 Description of the problem

The S5P/TROPOMI Level-1b radiance spectra have a nominal wavelength scale \( (\lambda_{\text{nom}}) \), but this wavelength grid is not corrected for inhomogeneous slit illumination [RD13, section 28]. The measurements are also not temperature corrected, but because the instrument itself is temperature stabilized it is expected that this effect can be ignored. The Level-2 processors must correct the nominal wavelength scale of the radiance measurements for inhomogeneous slit illumination due to the presence of clouds in the field of view.

One would like to follow the calibration of the irradiance spectra, for a short wavelength interval. The range \( \lambda_{\text{fit}} = [\lambda_-, \lambda_+ ] \) is the approximate range on which to do the wavelength calibration. To avoid non-linearities this wavelength range is tailored to the specific Level-2 algorithm. For each detector row the nominal wavelength \( \lambda_{\text{nom}} \) is adjusted with a wavelength offset (or: shift) \( w_s \) and a wavelength stretch \( w_q \) to find the calibrated wavelength \( \lambda_{\text{cal}} \):

\[
\lambda_{\text{cal}} = \lambda_{\text{nom}} + w_s + w_q \left( \frac{2 \lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \right) + \ldots \tag{23}
\]

with \( \lambda_0 \) the center of the fit window, \( \lambda_- \) the beginning of the fit window and \( \lambda_+ \) the end of the fit window. In the third term the factor 2 is used to ensure that the wavelength factor of the stretch lies in the range \([-1 : +1]\). The higher order terms in Eq. (23) are ignored, even fitting \( w_q \) is optional.

### A.2 Non-linear model function and Jacobian

The model function in the fit is similar to a non-linear DOAS equation. Instead of fitting the reflectance \( R \), we fit the radiance \( I \) directly, bringing the (model) irradiance \( I_{\text{mod}} \) to the other side of the equation. The model function \( \mathcal{M} \) is given by:

\[
\mathcal{M}(\lambda_{\text{nom}}; a_0, \ldots, a_N, C_{\text{ring}}, w_s, w_q, N_{s,0}, \ldots, N_{s,M}) = P_N(\lambda^{\ast}) \cdot \exp \left( \sum_{k=0}^{a_1} -N_k \cdot \sigma_k(\lambda_{\text{cal}}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal}}) + C_{\text{ring}} \cdot l_{\text{ring}}(\lambda_{\text{cal}}) \right) \tag{24}
\]

with \( \lambda_{\text{cal}} \) the calibrated wavelength as given by the first three terms in Eq. (23),

\[
P_N(\lambda^{\ast}) = \sum_{k=0}^{N} a_j(\lambda^{\ast})^j, \quad \lambda^{\ast} = 2 \frac{\lambda_{\text{nom}} - \lambda_0}{\lambda_+ - \lambda_-} \tag{25}
\]

a polynomial of order \( N \), \( E_{\text{mod}} \) the reference irradiance spectrum, and \( l_{\text{ring}} \) the Ring spectrum; both \( E_{\text{mod}} \) and \( l_{\text{ring}} \) are convolved with the instrument slit function (or: instrument spectral response function; ISRF; available via [ER11]). The spectra \( \sigma_k \) (\( k = 0, \ldots, M \)) are optional absorption spectra that have a relevant impact on the radiance, for instance the \( O_3 \) absorption cross section. These additional reference spectra have also been
convolved with the ISRF, but note that the DOAS assumption still applies: this merit function is not applicable to line absorbers such as H₂O, CH₄, CO or O₂, and will fail at wavelengths below \( \sim 320 \) nm because the profile shape of O₃ is relevant at those wavelengths. The order of the polynomial is \( 1 \leq N \leq 5 \), depending on the length of the fit window.

The wavelength calibration fit adjusts the parameters \( a_0, \ldots, a_N, C_{\text{ring}}, w_q, w_q, N_{s,0}, \ldots, N_{s,M} \) to minimize \( \chi^2 \):

\[
\chi^2 = \frac{1}{m-n} \sum_{i=0}^{m-1} \left( \frac{I_i - \mathcal{M}(a_0, \ldots, a_N, C_{\text{ring}}, w_q, w_q, N_{s,0}, \ldots, N_{s,M})}{\Delta I_i} \right)^2
\]

with \( I_i \) the measured radiance at detector pixel index \( i \), \( \Delta I_i \) the precision of this radiance, and \( m \) the number of spectral points between \( \lambda_- \) and \( \lambda_+ \). The number of degrees of freedom is \( m \) minus the number of fit parameters:

\[
n = N + 1 + M + 1 + 3
\]

The additional 3 here is when fitting \( C_{\text{ring}}, w_q \) and \( w_q \): if \( C_{\text{ring}} \) and/or \( w_q \) are not fitted, the number of degrees of freedom increases.

To minimize the number of function calls in the optimisation routine derivatives with respect to the fit parameters as a Jacobian matrix need to be supplied, with \( i \) the detector pixel index: The components of the Jacobian are given by Eqs. (28–32) below.

\[
\frac{\partial \mathcal{M}_i}{\partial a_j} = (\lambda^*_i)^j \cdot \exp \left( \sum_{k=0}^{M} -N_{s,k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot [E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}}I_{\text{ring}}(\lambda_{\text{cal},i})]
\]

\[
\frac{\partial \mathcal{M}_i}{\partial C_{\text{ring}}} = P_N(\lambda^*_i) \cdot \exp \left( \sum_{k=0}^{M} -N_{s,k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}}I_{\text{ring}}(\lambda_{\text{cal},i}) \right)
\]

\[
\frac{\partial \mathcal{M}_i}{\partial w_q} = P_N(\lambda^*_i) \cdot \exp \left( \sum_{k=0}^{M} -N_{s,k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}}I_{\text{ring}}(\lambda_{\text{cal},i}) \right)
\]

\[
\frac{\partial \mathcal{M}_i}{\partial w_q} = P_N(\lambda^*_i) \cdot \exp \left( \sum_{k=0}^{M} -N_{s,k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot \left( E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}}I_{\text{ring}}(\lambda_{\text{cal},i}) \right)
\]

\[
\frac{\partial \mathcal{M}_i}{\partial N_{s,k}} = -P_N(\lambda^*_i) \cdot \sigma_k(\lambda_{\text{cal},i}) \cdot \exp \left( \sum_{k=0}^{M} -N_{s,k} \sigma_k(\lambda_{\text{cal},i}) \right) \cdot [E_{\text{mod}}(\lambda_{\text{cal},i}) + C_{\text{ring}}I_{\text{ring}}(\lambda_{\text{cal},i})]
\]

The reference spectra \( E_{\text{mod}}(\lambda), I_{\text{ring}}(\lambda) \) and \( \sigma_k(\lambda) \) are pre-convolved with the ISRF. During the fitting 4th degree splines are used to represent these spectra. An interesting feature is that a spline of the derivative with respect to the independent variable can be calculated from the parameters of the original spline (given that the derivatives are w.r.t. the wavelength, the resulting spline for these derivatives is 3rd degree).

These equations can be solved with various optimization routines, for instance Levenberg-Marquardt or Gauss-Newton, with or without constraints or regularization methods. After thorough testing the optimal estimation method as implemented in DISAMAR, which is based on Rodgers [2000] and uses an unmodified Gauss-Newton to find the state vector for the next iteration, was selected for the S5P/TROPOMI wavelength calibration. For this usage, the a-priori error estimates are set very large (see Sect A.2.1), so that these do not limit the solution, and a pre-whitening of the data is performed to improve numerical stability.
A.2.1 Prior information for the optimal estimation fit

Optimal estimation needs prior information for the regularisation process during the fitting procedure, both a starting value and a covariance value. For input only the diagonal elements of the covariance matrix are specified, on output a full posterior error covariance matrix is available. The polynomial coefficients are not important, the values and variance were estimated from a large number of retrievals. The Ring coefficient was taken from the same data set. The value for $w_s$ is taken from the spacing of the nominal grid. A 1-$\sigma$ error of a third of the spacing of the wavelength grid seems reasonable: $\sigma_{\text{prior}}(w_s) = \Delta \lambda / 3$. This value will mostly prevent fitting a shift $w_s$ that is larger than half of the grid spacing, which basically means the wavelength is not known at all. The prior value for $w_q$ is 0 (zero), i.e. no stretch or squeeze. The range depends on the size of the fitting window, a consequence of the use of $\lambda^*$, as defined in Eq. (25). The current value is a deliberate overestimation. The slant column of O$_3$ is typically 0.18 mol/m$^2$ (about 600 DU); other trace gases are not included. An overview of the prior information used for S5P/TROPOMI is given in Table 15.

Table 15: A-priori values and a-priori error for the optimal estimation wavelength fit for S5P/TROPOMI. The ozone slant column is expressed in mol/m$^2$; the other quantities are dimensionless.

<table>
<thead>
<tr>
<th>Names</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$a_2,...,N$</th>
<th>$C_{\text{ring}}$</th>
<th>$w_s$</th>
<th>$w_q$</th>
<th>$N_{a, O_3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>1</td>
<td>−0.5</td>
<td>0.01</td>
<td>$6 \times 10^{-2}$</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>Covariance</td>
<td>(1)$^2$</td>
<td>(0.5)$^2$</td>
<td>(0.1)$^2$</td>
<td>(6 $\times 10^{-2}$)$^2$</td>
<td>yes</td>
<td>no</td>
<td>(0.1)$^2$</td>
</tr>
<tr>
<td>Optional</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>(0.18)$^2$</td>
</tr>
</tbody>
</table>

A.3 Application of the wavelength calibration in NO$_2$

For the retrieval of NO$_2$ the $N_{a, O_3}$ is not fitted, as O$_3$ shows little structure and is a weak absorber in band 4, where the slant columns of NO$_2$ (window 405 – 465 nm) and O$_2$–O$_2$ (window 460 – 490 nm) are fitted.

Testing with OMI [RD25] has shown that there is no significant amount of stretch in the wavelength of the spectra of that instrument in the 405 – 465 nm range and given the similarities of the OMI and S5P/TROPOMI detectors, no significant stretch was expected for S5P/TROPOMI. This has been confirmed using retrieval results for a S5P/TROPOMI orbit, which resulted in a very small stretch with a precision larger than the stretch itself, and a negligible effect on the retrieval results [Van Geffen et al., 2020], and hence the $w_q$ fit parameter will remain turned off.

The order of the polynomial in Eq. (25) is set to 2 and the Ring effect is included in the fit. The a-priori error of $w_s$ is set to 0.07 nm.
B High-sampling interpolation

After the wavelength calibration of the radiancespectrum, discussed in Appendix A, the irradiance and radiance observations need to be brought to the same wavelength grid in order to be able to compute the reflectance in Eq. (1). Because of the geometry of the solar observations, these measurements are shifted with respect to the radiance observations due to the Doppler shift caused by the motion of the satellite relative to the sun. Given that the irradiance spectrum is known better than the radiance spectrum, the irradiance spectrum is shifted to the radiance grid and the radiance observations are left without modification:

\[ E_0(\lambda_{\text{earth}}) = \frac{E_{\text{high}}(\lambda_{\text{earth}})}{E_{\text{high}}(\lambda_{\text{solar}})} E_0(\lambda_{\text{solar}}) \]  

(33)

with \( E_0 \) the observed irradiance, \( E_{\text{high}} \) a high resolution solar reference spectrum, convolved with the instrument spectral response function, \( \lambda_{\text{earth}} \) the wavelength of the earth radiance spectrum for pixel \( i \), and \( \lambda_{\text{solar}} \) the wavelength of the solar irradiance spectrum for pixel \( i \). The index \( i \) is synchronized between the radiance and irradiance observations, such that they refer to the same physical pixel on the detector. On \( E_{\text{high}} \), 4th degree spline interpolation is used to find the value at the indicated wavelengths. The input data for the splines have sufficient spectral resolution to allow for this.

Fig. 29 shows the procedure graphically. Panel (d) shows the effect of spline interpolation on the irradiance data to find the values at the earth radiance wavelength grid. Errors are small but systematic. Note that these errors appear directly in the reflectance data. The reflectance in Eq. (1) is then be calculated at the radiance wavelength grid: \( R_{\text{meas}}(\lambda_{\text{earth}}) = \pi I(\lambda_{\text{earth}})/\mu_0 E_0(\lambda_{\text{earth}}) \).

Figure 29: High sampling interpolation on part of a solar observation. (a) The red dots show the actual observation (taken from GOME-2A). The blue vertical lines indicate the wavelength grid of the radiance observation. The solid line shows a high resolution solar reference spectrum that has been convolved with the instrument spectral response function of the instrument (in this case GOME-2A). (b) The ratio \( E_{\text{high}}(\lambda_{\text{earth}})/E_{\text{high}}(\lambda_{\text{solar}}) \). (c) In red \( E_0(\lambda_{\text{solar}}) \), in blue \( E_0(\lambda_{\text{earth}}) \). (d) The solid line is the solar reference spectrum. The dash-dotted line is a high resolution irradiance spectrum created by spline interpolation directly on the observed irradiances, brought to the same average level of the window shown here to ease comparisons. The black dots indicate the error in % that are caused by using spline interpolation directly on the irradiance observations. Clear artifacts are caused by this, especially because noise on the observations becomes correlated between nearby points in the spectrum.
C Effective cloud fraction in the NO$_2$ window

The cloud radiance fraction, $w_{\text{NO}_2}$ (Sect 6.4.3), and the effective cloud fraction, $f_{\text{eff,NO}_2}$ (Sect. 6.4.4), in the NO$_2$ fit window, can be computed from a look-up table (LUT) with the top-of-atmosphere (TOA) reflectance at $\lambda_{\text{NO}_2} = 440$ nm as a function of viewing geometry, surface & cloud albedo, and surface & cloud pressure, based on the continuum reflectance at 440 nm of the measurement. The continuum reflectance at 440 nm could be determined from the observed spectrum, averaged over a small wavelength interval, but that may lead to unexpected values, e.g. in case of spikes in the measurement. Instead, we have opted for using the modelled reflectance of Eq. (5) evaluated at 440 nm. The approach is very similar to FRESCO [RD18] and explicitly accounts for Rayleigh scattering. The following description is adapted from [RD26].

The LUT assumes that the measured reflectance at TOA is defined as (cf. Eq. (1)):

$$R_{\text{TOA}}(\lambda) = \frac{\pi I(\lambda)}{\mu_0 E_0(\lambda)}$$  \hspace{1cm} (34)

In the independent pixel approximation the cloud fraction, $f_c$, for a given wavelength is given by:

$$f_c = \frac{R_{\text{TOA}} - R_s}{R_c - R_s}$$  \hspace{1cm} (35)

and the cloud radiance fraction, the fraction of the total radiation that comes from the clouds, is given by:

$$w_{\text{NO}_2} = \frac{f_c R_c}{R_{\text{TOA}}} = \frac{f_c R_c}{f_c R_c + (1 - f_c) R_s}$$  \hspace{1cm} (36)

where $R_s$ and $R_c$ are the reflectances at surface and cloud, respectively. These are computed from a limited LUT, based on Chandrasekhar (Chandrasekhar et al. [1950], Sect. 72). For bounding surface ‘b’, i.e. either surface (‘s’) or cloud (‘c’):

$$R_b(\lambda, A_b(\lambda)) = R_b(\lambda) + \frac{A_b(\lambda)}{1 - A_b(\lambda)} T(\lambda) s(\lambda)$$  \hspace{1cm} (37)

where:

$$R_b(\lambda, A_b(\lambda)) = \text{The reflectance of the combined atmosphere-surface system related to the light coming from the boundary ‘b’, i.e. either surface (‘s’) or cloud (‘c’).}$$

$$R_b(\lambda) = \text{The reflectance of the atmosphere if the surface is perfectly black: } A_b = 0.$$  \hspace{1cm}

$$A_b(\lambda) = \text{The albedo at the bounding surface, either cloud (} A_c \text{) or surface (} A_s \text{).}$$

$$T(\lambda) = \text{The transmittance of the atmosphere, a measure for the probability that photons travel through the atmosphere, are reflected by a surface with unit albedo, and travel back to the sensor (reflections by the atmosphere back towards the surface are ignored here).}$$

$$s(\lambda) = \text{The spherical albedo of the atmosphere for illumination at its lower boundary: } 1/[1 - A_b(\lambda) s(\lambda)] \text{ is the sum of a geometrical series accounting for the reflections between the atmosphere and the surface.}$$

The transmittance of the atmosphere $T(\lambda)$ is a product of two terms depending on the viewing and solar zenith angles:

$$T(\lambda) = t(\lambda; \mu) t(\lambda; \mu_0)$$  \hspace{1cm} (38)

where $\mu = \cos(\theta)$ and $\mu_0 = \cos(\theta_0)$ and:

$$t(\lambda; \mu) = \exp(-\frac{\tau(\lambda)}{\mu}) + \int_0^1 2\mu^\prime T_0(\lambda; \mu, \mu^\prime) d\mu^\prime$$  \hspace{1cm} (39)

In Eq. (39) we assume a plane parallel atmosphere; for a spherical shell atmosphere the factor $1/\mu$ in $\exp(-\tau/\mu)$ has to be replaced by a different expression.

The TOA reflectance related to the light coming from the boundary ‘b’, i.e. either surface (‘s’) or cloud (‘c’), is a function of solar and viewing geometries and surface properties: $R_b(\lambda, A_b(\lambda)) = R_b(\lambda; \theta_0, \phi - \phi_0; p_0, A_b(\lambda))$, where $p_0$ is the pressure at the boundary ‘b’. In addition extra dependencies may be needed to account for absorbing species, in particular at shorter wavelengths where absorption by ozone (O$_3$) is significant. A more
which simplifies Eq. (40) to

\[ A_s(\lambda) = \frac{R_{\text{TOA}}(\lambda) - R_0(\lambda, p_b)}{T(\lambda, p_b) + s(\lambda, p_b) [R_{\text{TOA}}(\lambda) - R_0(\lambda, p_b)]} \]  

(40)

In a similar fashion it is clear from Eq. (35) that a cloud fraction larger than 1 results when \( R_{\text{TOA}} > R_c \). Rewriting Eq. (37) to set \( R_c = R_{\text{TOA}} \) provides an adjusted value for \( A_c \):

\[ A_c(\lambda) = \frac{R_{\text{TOA}}(\lambda) - R_0(\lambda, p_c)}{T(\lambda, p_c) + s(\lambda, p_c) [R_{\text{TOA}}(\lambda) - R_0(\lambda, p_c)]} \]  

(41)

Note that in the FRESCO cloud retrieval (Sect. 6.4.4) the surface albedo is adjusted ignoring Rayleigh scattering, which simplifies Eq. (40) to \( A_s(\lambda) = R_{\text{TOA}}(\lambda) \), and Eq. (41) to \( A_c(\lambda) = R_{\text{TOA}}(\lambda) \).

### Table 16: Look-up tables and dimensions for reflectance calculations; no trace gas column entries included.

<table>
<thead>
<tr>
<th>( R_0 )</th>
<th>Reflectance of the black surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>For all wavelengths where a cloud fraction must be computed [1, ..., ( n )]</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>For ( \mu_0 = [0.0012141231 : 1.0] ), i.e. ( \theta_0 = [89.93° : 0°] ), in 42 steps of ( 2 – 5° )</td>
</tr>
<tr>
<td>( \mu )</td>
<td>For ( \mu = [0.0012141231 : 1.0] ), i.e. ( \theta = [89.93° : 0°] ), in 42 steps of ( 2 – 5° )</td>
</tr>
<tr>
<td>( \phi - \phi_0 )</td>
<td>Dependency stores in three Fourier terms</td>
</tr>
<tr>
<td>( p_b )</td>
<td>Pressure of the bounding surface (cloud or surface) for ( p_b = [1075 \text{ hPa : 95 \text{ hPa}] ) in 68 steps *</td>
</tr>
</tbody>
</table>

| \( T \) | Transmittance of the atmosphere |
| \( \lambda \) | For all wavelengths where a cloud fraction must be computed [1, ..., \( n \)] |
| \( \mu_0 \) | For \( \mu_0 = [0.0012141231 : 1.0] \), i.e. \( \theta_0 = [89.93° : 0°] \), in 42 steps of \( 2 – 5° \) |
| \( \mu \) | For \( \mu = [0.0012141231 : 1.0] \), i.e. \( \theta = [89.93° : 0°] \), in 42 steps of \( 2 – 5° \) |
| \( p_b \) | Pressure of the bounding surface (cloud or surface) for \( p_b = [1075 \text{ hPa : 95 \text{ hPa}] \) in 68 steps * |

| \( s \) | Spherical albedo of the atmosphere |
| \( \lambda \) | For all wavelengths where a cloud fraction must be computed [1, ..., \( n \)] |
| \( p_b \) | Pressure of the bounding surface (cloud or surface) for \( p_b = [1075 \text{ hPa : 95 \text{ hPa}] \) in 68 steps * |

*) Through a fixed scale height \( p_b \) is linked to the elevation of the bounding surface: \( z_b = [-55 \text{ m : 16250 \text{ m}]} \).
D Surface albedo correction using the snow/ice flag

The retrieval process uses the surface albedo in the NO\textsubscript{2} fit window, \(A_{s,NO_2}\), as one of the input parameters for the air-mass factor and vertical column calculations (Sect. 6.4). This surface albedo is taken from a surface albedo climatology. For TROPOMI that is the 440 nm data from the 5-year average OMI database [Kleipool et al., 2008]; see Sect. 6.4.5. Substantial errors are introduced in the retrieval results if the real surface albedo, \(A_s\), differs considerably from what is expected, for example in the case of the sudden snowfall or ice cover. Correcting the surface albedo from the climatology, \(A_{\text{clim}}\), using knowledge of actual snow/ice cover (Sect. 6.4.6) will therefore improve the final data product, in terms of the retrieval itself and for flagging such cases. For the \(A_{s,NO_2}\), this correction follows the approach included in the OMI cloud data product OMCLDO2 [Veefkind et al., 2016] to adapt the surface albedo in the O\textsubscript{2}-O\textsubscript{2} fit window (i.e. at 471 nm).

The basis for the correction are the snow/ice flag values, sampled at the ground pixel centre coordinate, as used in the processing (Sect. 6.4.6). Table 17 provides an overview of these flags, where an asterisk marks snow/ice flag values that lead to an adjustment of the surface albedo in case a certain threshold is exceeded.

Table 17: Overview of the snow/ice flag values \(f_{\text{NISE}}\) used by both the NISE and ECMWF snow/ice data sets, where nrs. 252, 253 and 254 do not occur in the ECMWF data. Flag values marked with an asterisk may lead to adjustment of the climatological surface albedo as described in the text.

<table>
<thead>
<tr>
<th>(f_{\text{NISE}})</th>
<th>meaning</th>
<th>remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>000 *</td>
<td>snow-free land</td>
<td></td>
</tr>
<tr>
<td>001-100 *</td>
<td>sea ice concentration (percent)</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>permanent ice</td>
<td></td>
</tr>
<tr>
<td>103 *</td>
<td>dry &amp; wet snow</td>
<td></td>
</tr>
<tr>
<td>252</td>
<td>mixed pixels at coastlines</td>
<td>land-ocean or snow/ice-ocean boundaries</td>
</tr>
<tr>
<td>253</td>
<td>suspect ice value</td>
<td>considered to represent an error</td>
</tr>
<tr>
<td>254</td>
<td>error value</td>
<td></td>
</tr>
<tr>
<td>255 *</td>
<td>ocean</td>
<td></td>
</tr>
</tbody>
</table>

The rules for modifying the climatological surface albedo \(A_{\text{clim}}\) are as follows:

- In case of snow-free land or open ocean (flags 0 and 255) adjust \(A_s\) if the difference between \(A_{\text{clim}}\) and default value \(A_{\text{def}} = 0.04\) is larger than a given threshold \(A_{\text{thrs}} = 0.1\), where the albedo is decreased only if \(A_{\text{clim}} > A_{\text{snow}} = 0.6\).

  \[
  \text{if } \left( (A_{\text{clim}} - A_{\text{def}}) > A_{\text{thrs}} \text{ and } A_{\text{clim}} > A_{\text{snow}} \right) \text{ then } A_s = A_{\text{def}} \text{ else } A_s = A_{\text{clim}}
  \]

- In case of dry or wet snow (flag \(f_{\text{NISE}} = 103\)) adjust \(A_s\) if the difference between \(A_{\text{clim}}\) and \(A_{\text{snow}} = 0.6\) is larger than a given threshold \(A_{\text{thrs}} = 0.1\).

  \[
  \text{if } (A_{\text{clim}} - A_{\text{snow}}) > A_{\text{thrs}} \text{ then } A_s = A_{\text{snow}} \text{ else } A_s = A_{\text{clim}}
  \]

- In case of a non-zero sea ice concentration (flags \(f_{\text{NISE}} = 1 - 100\)) adjust \(A_s\) if the difference between \(A_{\text{clim}}\) and a default value \(A_{\text{def}} = 0.4\) is larger than a given threshold \(A_{\text{thrs}} = 0.1\).

  \[
  \text{if } |A_{\text{clim}} - A_{\text{def}}| > A_{\text{thrs}} \text{ then } A_s = A_{\text{def}} \text{ else } A_s = A_{\text{clim}}
  \]

where \(A_{\text{def}}\) depends on the month of the year \(M\) and on which hemisphere \(H\) the ice is found:

\[
A_{\text{def}} = (1.0 - 0.01 \cdot f_{\text{NISE}}) \cdot 0.065 + 0.01 \cdot f_{\text{NISE}} \cdot A_{\text{ice}}(M, H)
\]

with \(A_{\text{ice}}(M, H)\) following from the LUT used to determine the OMI surface albedo database [Kleipool et al., 2008]:

\[
A_{\text{ice}}(M, 'north') = 0.70, 0.73, 0.76, 0.80, 0.84, 0.78, 0.61, 0.61, 0.62, 0.68, 0.67, 0.71
A_{\text{ice}}(M, 'south') = 0.53, 0.50, 0.44, 0.60, 0.61, 0.64, 0.68, 0.76, 0.80, 0.83, 0.78, 0.66
\]
E  Data quality value: the qa_value flags

To make the use of the TROPOMI data products easier, a so-called qa_value (where ‘qa’ stands for ‘quality assurance’) is assigned to each ground pixel. The qa_value is intended to serve as an easy filter of the observations (dividing the dataset in useful versus not useful observations), depending on how the data is used.

The data files have for each ground pixel the so-called processing_quality_flags, which provides the user information on processing issues, such as errors that were encountered in the processing, as well as a number of warnings. Some of these warnings have been included in the qa_value. The meaning of the processing_quality_flags values is detailed in Appendix A of the NO2 Product User Manual (PUM; available via [ER2]).

The following differentiation of the qa_value, fQA, for usage of the NO2 data product has been made:

\[
\frac{0.75}{0.50} \leq f_{QA} < 1.00 \quad \begin{align*}
\text{The ground pixel is recommended for all applications, including column comparisons, visualisation, trends, monthly/seasonal averages. The data is restricted to cloud-free observations (cloud radiance fraction < 0.5), and snow-ice free observations.} \\
\text{The ground pixel is recommended for use in data assimilation and comparisons against models or vertical profile observations, given that the averaging kernel is used to specify the sensitivity profile in cloudy situations; this includes good quality retrievals over clouds and snow/ice.} \\
\text{The ground pixel is not recommended for use due to serious retrieval issues.} \\
\text{A processing error occurred so that the ground pixel cannot be used at all, or the solar zenith angle exceeds the limit set in the data assimilation.}
\end{align*}
\]

The determination of the qa_value is done as follows. Starting from the initial value \( f_{QA} = 1 \), \( f_{QA} \) is multiplied by the modification factor \( f_i \) of each of the criteria \( i \) listed in Table 18 that have been met (i.e. if criterion \( i \) is not met then \( f_{QA} = 1 \)).

Table 18: Overview of the selection criteria for the qa_value, fQA, for the version v2.2.0 (viz. Table 2); previous versions may have different settings. Some quantities have a minimum or maximum value; these values are configuration parameters in the processing. In this table \( f_{NISE} \) stands for the snow/ice flags listed in Table 17 and used by both the NISE and ECMWF snow/ice data sets. And \( f_{AAI} \) represents the aerosol index 354/388 nm pair, which is passed on to the NO2 data product file as added flag. Warning flags in processing_quality_flags not used for fQA are not listed.

<table>
<thead>
<tr>
<th>( i )</th>
<th>criterion</th>
<th>( f_{QA} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>fatal error encountered according to processing_quality_flags</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>south_atlantic_anomaly_warning set in processing_quality_flags</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>sun_glint_warning set in processing_quality_flags</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>pixel_level_input_data_missing_warning set in processing_quality_flags</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>interpolation_warning set in processing_quality_flags</td>
<td>0.90</td>
</tr>
<tr>
<td>5</td>
<td>solar_eclipse set in geolocation_flags</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>( \theta_0 &gt; \theta_{0,1}^{\text{max}} = 81.2\degree )</td>
<td>0.30</td>
</tr>
<tr>
<td>7</td>
<td>( \theta_0 &gt; \theta_{0,2}^{\text{max}} = 84.5\degree )</td>
<td>0.10</td>
</tr>
<tr>
<td>8</td>
<td>( M_{\text{rad}}/M_{\text{geo}} &lt; M_{\text{prop}}^{\text{min}} = 0.1 )</td>
<td>0.45</td>
</tr>
<tr>
<td>9</td>
<td>( \Delta N_v &gt; (\Delta N_v)^{\text{max}} = 33.0 \times 10^{-6} ) mol/m² (≈ 2 \times 10^{15} molec/cm²) ( )</td>
<td>0.15</td>
</tr>
<tr>
<td>10</td>
<td>( f_{\text{NISE}} &lt; \frac{f_{\text{NISE}}}{f_{\text{NISE}}^{\text{min}}} = 1 ) ( ) or ( f_{\text{NISE}} = 252 ) or ( f_{\text{NISE}} = 255 ) then ( ) [no snow or ice] ( )</td>
<td>0.20</td>
</tr>
<tr>
<td>11</td>
<td>( \text{AAI}<em>{\text{NO}<em>2} &gt; A</em>{\text{AAI}}^{\text{max}} = 0.3 ) ( ) or ( \text{AAI}</em>{\text{NO}<em>2} &gt; A</em>{\text{AAI}}^{\text{max}} = 0.5 ) ( )</td>
<td>0.74</td>
</tr>
<tr>
<td>else-if</td>
<td>( f_{\text{NISE}} \neq 253 ) and ( f_{\text{NISE}} \neq 254 ) ( )</td>
<td>[snow/ice case]</td>
</tr>
<tr>
<td>12</td>
<td>( f_{\text{NISE}} &gt; 80 ) and ( f_{\text{NISE}} &lt; 104 ) and ( p_{\text{sc}} &gt; 0.96 ) ( \text{ps}^{\dagger} ) ( )</td>
<td>0.88</td>
</tr>
<tr>
<td>else</td>
<td>( )</td>
<td>[cloud-free snow/ice]</td>
</tr>
<tr>
<td>13</td>
<td>( )</td>
<td>0.73</td>
</tr>
<tr>
<td>14</td>
<td>( p_{\text{sc}} &lt; p_{\text{sc}}^{\text{min}} = 3.0 \times 10^4 ) Pa ( )</td>
<td>0.25</td>
</tr>
<tr>
<td>else</td>
<td>( )</td>
<td>[snow/ice error]</td>
</tr>
<tr>
<td>15</td>
<td>( f_{\text{AAI}} &gt; f_{\text{AAI}}^{\text{max}} = 1.0 \times 10^{10} ) ( )</td>
<td>0.00</td>
</tr>
<tr>
<td>16</td>
<td>( )</td>
<td>[for future use]</td>
</tr>
</tbody>
</table>

\(^1\) Note that this criterion means that the system switches to the scene mode if there is 1% or more snow/ice.

\(^{\dagger}\) In NO2 data versions prior to v1.4.0 this threshold was 0.98 · ps.
**F Spike removal in the DOAS fit**

In version 2.2.0 of the TROPOMI NO$_2$ processor a "spike removal" was implemented into the DOAS fit (Sect. 6.2), to remove strong outliers in the fit residual. Such outliers may be caused by, e.g., high-energy particles hitting the CCD detector (so-called transients), variations in the dark current, or bad pixels not correctly flagged in the Level-1b data. After removal of an outlier from the measured reflectance, the NO$_2$ DOAS fit is redone. To avoid ending up in a cycle, the new fit residual is not checked again for outliers.

To detect outliers the so-called box-plot method used by Veefkind et al. [2016] for the OMI O$_3$–O$_2$ cloud algorithm is implemented. This method [ER24] determines lower and upper quartiles based on the first and third quartiles, $Q_1$ and $Q_3$, i.e. the 25th and 75th percentile of a distribution (the second quartile, $Q_2$, is the median).

For data with a Gaussian distribution $Q_1 = -0.67$ and $Q_2 = +0.67$, which means that for normally distributed data, one-half of the data is within 2/3-rd of a standard deviation unit of the mean [ER25].

If a certain value is larger than $Q_1 + Q_2 \cdot Q_{1-1}$ or lower than $Q_1 - Q_2 \cdot Q_{1-1}$, with $Q_{1-1} = Q_3 - Q_1$ the inter-quartile range and $Q_f$ a suitable multiplication factor, it is termed an outlier. If $Q_f$ is set too low, valid observations are unjustly removed from the fit and the NO$_2$ SCD error will be underestimated.

For the so-called inner fences, defined by $Q_f = 1.5$ [ER24] and used by Veefkind et al. [2016], 0.74% of normally distributed data is termed an outlier [ER25]. This means that on average for each TROPOMI ground pixel 2 or 3 wavelength pixels will be designated as outlier, if the NO$_2$ fit residual on the 304 or 305 wavelength pixels within the NO$_2$ fit window would be normally distributed. In reality the fit residual is not normally distributed but drops off less steep, hence $Q_f = 1.5$ is clearly too low a factor to use, as this would inevitably lead to the removal of valid observations.

For the so-called outer fences, defined by $Q_f = 3.0$ [ER24], about 0.002% of normally distributed data is termed an outlier [ER26], or in terms of the NO$_2$ fit window about 1/100-th of a wavelength pixel (i.e. roughly one wavelength pixel per 100 ground pixels), so that the chance that valid observations are removed is small. Based on this evaluation, the TROPOMI NO$_2$ processor is configured to use $Q_f = 3.0$.

Figure 30 shows two examples of outlier detection. The fit residual in the left panel has one modest outlier, at a level of $Q_f = 3.51$. Removing this one spike, the NO$_2$ SCD changes from $14.9445 \times 10^{15} \pm 0.6752 \times 10^{15}$ molec/cm$^2$ to $14.8626 \times 10^{15} \pm 0.6599 \times 10^{15}$ molec/cm$^2$ ($2.4680 \pm 0.196 \times 10^{-4}$ mol/m$^2$), while the RMS error decreases by 3.7% from $5.1082 \times 10^{-4}$ to $4.8928 \times 10^{-4}$. The right panel shows the effect on the fit residual of a high-energy particle in the South Atlantic Anomaly, which appears to be so massive that it affects two neighbouring wavelength pixels as well, while the three ground pixels to the east of this one also have outliers at the same wavelength pixels, showing that the high-energy particle hit the detector under a slant angle. The main outlier has a level of $Q_f = 69.04$, while the two neighbouring pixels have levels of $Q_f = 18.64$ and $Q_f = 14.28$. For this pixel removing these three wavelength pixels changes the NO$_2$ SCD from an unrealistically low value with large error ($4.8324 \pm 4.5284 \times 10^{15}$ molec/cm$^2$) to a more realistic $6.8079 \pm 0.6510 \times 10^{15}$ molec/cm$^2$ ($1.1305 \pm 0.1081 \times 10^{-4}$ mol/m$^2$), while the RMS error decreases from $33.475 \times 10^{-4}$ to $4.4143 \times 10^{-4}$.

![Figure 30](image_url): Examples of outliers in the TROPOMI NO$_2$ fit residual. **Left panel**: fit residual for a ground pixel over the northern Pacific Ocean, where the quartiles and the inner and outer fences are indicated by dashed lines. **Right panel**: fit residual for a ground pixel over Brazil, where the impact of a high-energy particle caused by the South Atlantic Anomaly created a massive spike, involving three wavelength pixels; the inset shows a zoom-in of the residual before and after spike removal. For more details see the text.
Figure 31: Examples of outliers in the TROPOMI NO\textsubscript{2} SCD retrieval using Level-1b version 1 data of orbit 03658 (28 June 2018) over the Americas. Left panel: white spots mark ground pixels with 2 or more outliers, where the South Atlantic Anomaly is clearly visible. Middle panel: idem, but 8 or more outliers. Right panel: fit residual for a fully clouded ground pixel, where two wavelength pixels are flagged as saturated and thus removed from the fit (at 450.8 and 451.0 nm), the spike removal then finds 12 outliers and after removal of these there are 3 outliers left (these are not removed, because there is no second round in the spike removal); due to the spike removal the RMS error decreases from $21.173 \times 10^{-4}$ to $5.613 \times 10^{-4}$.

Most ground pixels with outliers have one or two outliers, even those over the South Atlantic Anomaly, where high-energy particle impacts on the CCD detector occur more often. Ground pixels suffering from saturation effects, however, may have a much larger number of outliers. Due to the high SNR of TROPOMI, bright scenes, notably over clouds in the tropics, the CCD may be over-exposed. As a result of this a number of wavelength pixels may get saturated and the electronic effects of this may spread to neighbouring wavelength pixels and to ground pixels in neighbouring rows, an effect called "blooming". Saturation effects occur most in the detector parts of band 4 ($400 - 496$ nm) and in band 6 ($720 - 785$ nm). In band 4 saturation is most prominent at high wavelengths but may spread far down into the NO\textsubscript{2} fit window.

Version 1 Level-1b data contains flags on wavelength pixels affected by saturation (such wavelength pixels are removed from the spectrum before the DOAS fit), but the criteria for this flagging are set rather conservative and there is no flagging of blooming. This left a number of wavelength pixels unflagged but affected, leading to multiple strong outliers (up to more than 50 outliers were observed in some ground pixels). Even after the spike removal, the new fit may show outliers and these fit results cannot always be trusted.

Version 2 Level-1b data has a better flagging for saturation and flagging for blooming, drastically reducing the number of outliers and spectra repaired by the spike removal give in most cases reliable results. (Flagging for saturation and blooming uses the same Level-1b error code.)

Figure 31 shows the location of ground pixels with outliers detected in an orbit over the Americas, i.e. including part of the South Atlantic Anomaly, with some high and bright cloud complexes over the eastern part of the USA and the western part of the Atlantic Ocean; one of the pixels in the latter region is used to give an example of the fit residual with and without spike removal in Fig. 31.

The treatment of saturation and outliers on wavelength pixels in the NO\textsubscript{2} SCD retrieval is governed by configuration parameters, listed in Table 19: the settings for NO\textsubscript{2}-v2.2 are made for use with version-2 Level-1b data in mind; prior NO\textsubscript{2} processor versions accepted very few saturation flags.

Table 19: Configuration parameters in the NO\textsubscript{2} processing related to saturation in the Level-1b spectra and outliers in the NO\textsubscript{2} retrieval residual; NO\textsubscript{2} data version coverages are listed in Table 2.

<table>
<thead>
<tr>
<th>configuration parameter</th>
<th>NO\textsubscript{2}-v1.x L1B-v1.0</th>
<th>NO\textsubscript{2}-v2.2 L1B-v2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>The maximum fraction of the radiance spectrum that is allowed to be flagged as saturated before the ground pixel is skipped</td>
<td>0.01</td>
<td>0.25</td>
</tr>
<tr>
<td>The maximum number of outliers that is allowed to be in a spectrum before the ground pixel is skipped</td>
<td>N/A</td>
<td>10</td>
</tr>
</tbody>
</table>
G References


