



# The STRatospheric Estimation Algorithm from Mainz (STREAM): estimating stratospheric NO<sub>2</sub> from nadir-viewing satellites by weighted convolution

Steffen Beirle<sup>1</sup>, Christoph Hörmann<sup>1</sup>, Patrick Jöckel<sup>2</sup>, Song Liu<sup>3</sup>, Marloes Penning de Vries<sup>1</sup>, Andrea Pozzer<sup>1</sup>,  
Holger Sihler<sup>1</sup>, Pieter Valks<sup>3</sup>, and Thomas Wagner<sup>1</sup>

<sup>1</sup>Max Planck Institute for Chemistry, Mainz, Germany

<sup>2</sup>Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre,  
Oberpfaffenhofen, Germany

<sup>3</sup>Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Methodik der Fernerkundung (IMF),  
Oberpfaffenhofen, Germany

*Correspondence to:* Steffen Beirle (steffen.beirle@mpic.de)

Received: 23 December 2015 – Published in Atmos. Meas. Tech. Discuss.: 18 January 2016

Revised: 20 May 2016 – Accepted: 1 June 2016 – Published: 4 July 2016

**Abstract.** The STRatospheric Estimation Algorithm from Mainz (STREAM) determines stratospheric columns of NO<sub>2</sub> which are needed for the retrieval of tropospheric columns from satellite observations. It is based on the total column measurements over clean, remote regions as well as over clouded scenes where the tropospheric column is effectively shielded. The contribution of individual satellite measurements to the stratospheric estimate is controlled by various weighting factors. STREAM is a flexible and robust algorithm and does not require input from chemical transport models. It was developed as a verification algorithm for the 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 gradients of stratospheric NO<sub>2</sub>, e.g., related to the polar vortex, and reduces interpolation errors over continents. Based on synthetic input data, the uncertainty of STREAM was quantified as about  $0.1\text{--}0.2 \times 10^{15}$  molecules cm<sup>-2</sup>, in accordance with the typical deviations between stratospheric estimates from different algorithms compared in this study.

## 1 Introduction

Beginning with the launch of the Global Ozone Monitoring Experiment (GOME) on the ERS-2 satellite in 1995 (Burrows et al., 1999), several instruments (SCIAMACHY, OMI, GOME-2; see Table 1 for acronyms and references) perform spectrally resolved measurements of sunlight reflected by the Earth's surface and atmosphere. With differential absorption spectroscopy (DOAS) (Platt and Stutz, 2008), the column densities (denoted as “columns” henceforth) of numerous important atmospheric absorbers can be determined by their characteristic spectral “fingerprints”, amongst others nitrogen dioxide (NO<sub>2</sub>).

Nitrogen oxides (NO<sub>x</sub> = NO<sub>2</sub> + NO) play a key role in the chemistry of both the stratosphere and the troposphere. Stratospheric NO<sub>x</sub> has been a research topic for several decades due in particular to its role in ozone and halogen chemistry.

Satellite measurements provide long-term global information on spatiotemporal patterns of stratospheric NO<sub>2</sub> (e.g., Wenig et al., 2004; Dirksen et al., 2011). During the last decades, the analysis of tropospheric trace gases from nadir-viewing satellite instruments moved more and more into focus, supported by the availability of longer time series and improved spatial resolution. Tropospheric NO<sub>2</sub> columns derived from satellite are nowadays widely used by the scientific community to deduce spatial patterns, source type

**Table 1.** UV/vis satellite instruments compared or discussed in this study

Acronym	Instrument	Launch	Footprint (km <sup>2</sup> )	Earth coverage per day	Instrument reference	Data product used in this study	Data reference
GOME	Global Ozone Monitoring Experiment	1995	40 × 320	1/3	Burrows et al. (1999)	TEMIS	Boersma and Eskes (2004)
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartography	2002	30 × 60	1/6	Bovensmann et al. (1999)	MPI-C Mainz	Beirle and Wagner (2012)
GOME-2	Global Ozone Monitoring Experiment-2	2006 <sup>a</sup>	40 × 80 <sup>b</sup>	2/3	Callies et al. (2000)	O3M SAF	Valks et al. (2015)
OMI	Ozone Monitoring Instrument	2004	13 × 26 <sup>c</sup>	1 <sup>d</sup>	Levelt et al. (2006)	NASA v003 / SP2 DOMINO v2	Bucsela et al. (2013) Boersma et al. (2011)
TROPOMI	TROPospheric Monitoring Instrument	2016	7 × 7 <sup>c</sup>	1	Veefkind et al. (2012)		
Sentinel 4		2021	7 × 7	– <sup>e</sup>	Ingmann et al. (2012)		

<sup>a</sup> On Metop-A. A second GOME-2 instrument was launched 2012 on Metop-B, and a third is planned to be launched on Metop-C in 2018.

<sup>b</sup> Switched to 40 × 40 km<sup>2</sup> for GOME-2/Metop-A in Metop-A and Metop-B tandem operation.

<sup>c</sup> At nadir.

<sup>d</sup> Reduced coverage after row anomaly in 2007; see <http://projects.knmi.nl/omi/research/product/rowanomaly-background.php>.

<sup>e</sup> Geostationary orbit: hourly coverage over Europe.

and strength, and trends of NO<sub>x</sub> emissions from fossil fuel combustion, biomass burning, soil emissions, and lightning. Overviews over the wide range of scientific applications of satellite-based tropospheric NO<sub>2</sub> products are given in, e.g., Martin (2008) or Monks and Beirle (2011).

The retrieval of tropospheric NO<sub>2</sub> columns from total column measurements requires the estimation and removal of the stratospheric column, a procedure we refer to as “stratosphere–troposphere separation” (STS) as in Bucsela et al. (2006).

One of the first STS algorithms is the reference sector method (RSM), which estimates the global stratospheric NO<sub>2</sub> fields from measurements over the remote Pacific (Richter and Burrows, 2002; Martin et al., 2002; Beirle et al., 2003), based on the assumptions of (a) longitudinal homogeneity of stratospheric NO<sub>2</sub> and (b) negligible tropospheric contribution over the reference region in the Pacific. This procedure is quite simple, transparent, and robust. A further side effect is that any systematic bias in the NO<sub>2</sub> columns, which might be introduced by the instrument (e.g., degradation or spectral interference caused by the diffusor plate used for measurements of the solar reference; Richter and Burrows, 2002) or sub-optimal spectral analysis (van Geffen et al., 2015; Marchenko et al., 2015), is classified as stratospheric signal and thereby removed from the tropospheric column.

The RSM was applied by different groups to different satellite instruments and generally performs well. However, the resulting tropospheric NO<sub>2</sub> columns are affected by systematic biases caused by the following simplifying assumptions.

- The tropospheric background column in the Pacific is very low (compared to columns over regions exposed to significant NO<sub>x</sub> sources) but not 0. Neglecting the tropospheric background results in tropospheric columns that are biased low by about some 10<sup>14</sup> molec cm<sup>−2</sup> (Martin et al., 2002; Valks et al., 2011; Hilboll et al., 2013). Some algorithms explicitly correct for this tropospheric background: Martin et al. (2002) perform a correction based on GEOS-CHEM, while Valks et al. (2011) assume a constant background of 0.1 × 10<sup>15</sup> molec cm<sup>−2</sup>. Other algorithms prefer to stick to the tropospheric “excess” columns, which are slightly biased low but do not need any model input (Richter and Burrows, 2002; Bucsela et al., 2006).
- The assumption of longitudinal homogeneity is generally reasonable, at least in temporal means when small-scale stratospheric dynamic features cancel out. However, large longitudinal variations can occur in particular close to the polar vortex, as already discussed by Richter and Burrows (2002), Martin et al. (2002), and Boersma et al. (2004). Thus, tropospheric columns derived by RSM can be off by more than 10<sup>15</sup> molec cm<sup>−2</sup> in winter at latitudes from 50° polewards, thereby affecting scientific interpretations of tropospheric columns over North America or northern Europe. Note that also at low latitudes, systematic artifacts show up in tropospheric columns resulting from RSM, in particular over the Indian ocean, which are related to longitudinal inhomogeneities.

To overcome the artifacts caused by the assumption of longitudinal homogeneity, several modifications of the RSM have been proposed in recent years, while the basic approach of

using nadir measurements over clean regions for STS has been retained. We refer to this group of algorithms as “modified RSM” (MRSM). MRSMs generally define a “pollution mask” of regions with potentially non-negligible tropospheric columns. Measurements over these regions are skipped within the stratospheric estimation. Thus, in order to define stratospheric columns over the masked areas, interpolation is required. For this purpose, Leue et al. (2001) and Wenig et al. (2004) applied “normalized convolution” (Knutsson and Westin, 1993), an efficient algorithm which combines interpolation and smoothing. Bucsela et al. (2006) realized interpolation by fitting harmonics (wave-2) over the “clean” areas. Valks et al. (2011) applied a zonal boxcar filter of 30° width.

All of these algorithms applied a rather conservative masking approach for potentially polluted pixels. Continents were masked out almost completely. At northern midlatitudes, the masked area is often even larger than the area used for the stratospheric estimation, and over the Eurasian continent the MRSM algorithms miss any supporting measurement points over about 10 000 km. This can lead to significant errors during interpolation. In particular the wave fitting approach can lead to large biases (Dirksen et al., 2011).

Leue et al. (2001) estimated the stratospheric fields based on clouded measurements over the ocean and subsequent interpolation. The focus on clouded observations provides a direct stratospheric measurement, as the tropospheric column is mostly shielded; thus, no further correction of the tropospheric background should be needed. However, clouded pixels possibly contain NO<sub>x</sub> produced by lightning (e.g., Beirle et al., 2006). Therefore, Wenig et al. (2004) changed the Heidelberg STS algorithm (Leue et al., 2001) by switching from clouded to cloud-free observations as input for the stratospheric estimate<sup>1</sup>.

Recently, Bucsela et al. (2013) proposed an MRSM which defines “unpolluted” pixels not with a fixed mask but according to the a priori expected tropospheric contribution to the total column for each individual satellite observation. This is determined from radiative transfer calculations based on a monthly mean NO<sub>2</sub> profile from a chemical transport model (CTM) and the actual cloud conditions. This procedure results in additional supporting points over continents in cases of clouds shielding the tropospheric column and thereby largely reduces potential interpolation artifacts.

Apart from (modified) reference sector methods, there are two further completely different approaches used for STS, which are based on (a) independent measurements or (b) CTMs.

- a. Coincident, but independent, stratospheric measurements are available for SCIAMACHY (Bovensmann et al., 1999). It was operated in alternating nadir/limb geometry, such that the stratospheric air masses sensed in

nadir were scanned in limb shortly before (“limb–nadir matching”, LNM). This unique instrumental setup allowed for a direct stratospheric correction, although systematic offsets between limb and nadir measurements still had to be corrected empirically. STS by LNM was successfully applied for NO<sub>2</sub> (Sioris et al., 2004; Sierk et al., 2006; Beirle et al., 2010a; Hilboll et al., 2013) and ozone (Ebojje et al., 2014). However, such direct coincident measurements of total columns (nadir) and stratospheric concentration profiles (limb) are not available for other satellite instruments, and merging measurements from different sensors always faces the problem of spatiotemporal mismatching, requiring interpolation and photochemical corrections (compare Belmonte Rivas et al., 2014), and thus cannot be easily used for consistent long-term operational retrievals.

- b. Stratospheric NO<sub>2</sub> concentrations provided by CTMs can be used directly for STS after empirical correction of systematic offsets between satellite and model columns, e.g., by matching both over the Pacific (Richter et al., 2005; Hilboll et al., 2013). A more sophisticated way to incorporate CTMs in STS is data assimilation (Eskes et al., 2003; Dirksen et al., 2011), in which modeled 3-D distributions of NO<sub>2</sub> are regularly updated such that the modeled stratospheric column is in close agreement with the satellite measurement when the tropospheric contribution (as forecasted by the CTM) is low.

In 2016, the ESA’s Sentinel 5 precursor (S5p) satellite (Ingmann et al., 2012) will be launched, carrying the Tropospheric Monitoring Instrument (TROPOMI) (Veeffkind et al., 2012). The operational (“prototype”) tropospheric column product of NO<sub>2</sub> from TROPOMI will be derived by a STS using data assimilation (van Geffen et al., 2014), based on the expertise of the Koninklijk Nederlands Meteorologisch Instituut (KNMI) as demonstrated by a 20-year record of tropospheric columns from different satellite sensors provided by the Tropospheric Emission Monitoring Internet Service (TEMIS, [www.temis.nl](http://www.temis.nl); Boersma and Eskes, 2004; Boersma et al., 2011).

Within the S5p level 2 project, for each prototype product a “verification” product was developed in order to verify the prototype algorithms, detect possible shortcomings, and reveal potential improvements. The TROPOMI verification algorithm for NO<sub>2</sub> STS, the STRatospheric Estimation Algorithm from Mainz (STREAM), was developed at the Max Planck Institute for Chemistry (MPI-C), Mainz. It is an MRSM, requiring no further model input, and can thus be considered as a complementary approach to data assimilation.

STREAM does not apply a strict discrimination of “clean” vs. “polluted” satellite pixels. Instead, weighting factors are defined for each satellite pixel determining its impact on the

<sup>1</sup>This aspect will be discussed in detail in Sect. 5.4.

stratospheric estimate (similar to data assimilation). In particular, clouded observations are weighted high, as they provide direct measurements of the stratospheric field. This approach dampens the small but systematically high bias of stratospheric columns estimated from total column measurements and the resulting low bias of tropospheric columns.

The paper is organized in the following way. In Sect. 2, the STREAM algorithm is described in detail. Section 3 provides information on the satellite and model data sets used in this study. Section 4 analyses the performance of STREAM and its sensitivity to input parameters based on both actual satellite measurements and synthetic data. In Sect. 5, the STREAM results are discussed in comparison to other STS algorithms, including the TROPOMI prototype algorithm. A general discussion on the challenges and uncertainties of STREAM in particular, and STS in general, is given, followed by conclusions (Sect. 6). Several additional images and tables are provided in the Supplement and referenced by a prefix “S”.

## 2 Methods

STREAM is in the tradition of MRSM algorithms that estimate the stratospheric field directly from satellite measurements for which the tropospheric contribution is considered to be negligible. For this purpose, measurements over remote regions with negligible tropospheric sources, as well as cloudy measurements, are used. In contrast to other MRSMs, however, no strict pollution mask is applied. Instead, weighting factors are used.

STREAM consists basically of two steps:

1. A set of weighting factors is calculated for each satellite pixel: a “pollution weight” that reduces the contribution of potentially polluted pixels, a “cloud weight” that increases the contribution of cloudy observations, and the “tropospheric residue (TR) weight” that adjusts the total weight in case of exceptionally large or negative TRs. The product of these weighting factors determines to what extent the associated NO<sub>2</sub> total columns contribute to the estimated stratospheric field (Sect. 2.2).
2. Global maps of stratospheric NO<sub>2</sub> are determined by applying weighted convolution (Sect. 2.3).

Before describing the details of the STREAM algorithm, however, we first define the investigated quantities and abbreviations used hereafter, as summarized in Table 2.

### 2.1 Terminology

#### 2.1.1 NO<sub>2</sub> column densities and units

With differential optical absorption spectroscopy (DOAS; Platt and Stutz, 2008), so-called slant column densities (SCDs)  $S$ , i.e., concentrations integrated along the mean light

path, are derived. SCDs are converted into VCDs (vertical column densities, i.e., vertically integrated concentrations)  $V$  via the air-mass factor (AMF)  $A$ :  $V = S/A$ . The AMF  $A$  depends on radiative transfer (determined by wavelength, atmospheric absorbers, viewing geometry, surface albedo, clouds, and aerosols) and the trace gas profile. For the stratospheric column of NO<sub>2</sub>,  $A$  is basically determined by viewing geometry.

In this study, all column densities are given in column density units (CDU) to increase readability:

$$1 \text{ CDU} := 1 \times 10^{15} \text{ molecules cm}^{-2}. \quad (1)$$

#### 2.1.2 Total vertical column $V^*$

We define  $V^*$  as “total” vertical column, given by the SCD  $S$  divided by the stratospheric AMF  $A_{\text{strat}}$ :

$$V^* = S/A_{\text{strat}}. \quad (2)$$

The application of the stratospheric AMF basically removes the dependencies of  $S$  on viewing angles. Over clean regions with negligible tropospheric columns,  $V^*$  represents the actual total VCD and can be used for the estimation of stratospheric fields. In case of tropospheric pollution, however,  $V^*$  underestimates the actual total VCD, as the AMF is generally smaller in the troposphere than in the stratosphere (see also next section). These situations are, to the best of our knowledge, excluded from the stratospheric estimate by the definition of appropriate weighting factors (see Sect. 2.2).

#### 2.1.3 Stratospheric vertical column and tropospheric residue

STREAM yields an estimate for the stratospheric VCD  $V_{\text{strat}}$  based on the assumption that  $V^*$  can be considered as proxy for  $V_{\text{strat}}$  in “clean” regions and over cloudy scenes.

In order to evaluate the performance of the stratospheric estimation, we define the TR as the difference of total and stratospheric VCDs (based on a stratospheric AMF):

$$T^* = V^* - V_{\text{strat}}. \quad (3)$$

Tropospheric VCDs (TVCDs), which are the final product of NO<sub>2</sub> retrievals used for further tropospheric research, are connected to  $T^*$  via the ratio of stratospheric and tropospheric AMF:

$$V_{\text{trop}} = T^* \times \frac{A_{\text{strat}}}{A_{\text{trop}}}. \quad (4)$$

For cloud-free satellite pixels, the ratio  $A_{\text{strat}}/A_{\text{trop}}$  typically ranges from about 1 above clean oceans at low and midlatitudes to  $\approx 2$ –3 above moderately polluted regions and up to  $> 4$  at high latitudes and over strong NO<sub>x</sub> sources, where NO<sub>2</sub> profiles peak close to the ground, causing low  $A_{\text{trop}}$ . Figure S1 in the Supplement displays monthly mean ratios

**Table 2.** Terms and abbreviations related to STREAM used in this study.

Symbol	Abbrev.	Term	Description
<i>A</i>	AMF	Air-mass factor	Factor relating vertical to slant column density
	CDU	Column density unit	Unit of column densities: $1 \times 10^{15}$ molecules $\text{cm}^{-2}$
<i>G</i>	CK	Convolution kernel	Kernel used for weighted convolution; here 2-D Gaussian
$p_{\text{cld}}$	CP	Cloud pressure	
<i>c</i>	CRF	Cloud radiance fraction	
	LNM	Limb–nadir matching	Stratospheric correction based on coincident stratospheric measurements in limb geometry (SCIAMACHY).
<i>P</i>	MRSM	Modified reference sector method	An STS estimating stratospheric columns based on the total columns over “clean” regions, but allowing for longitudinal variations see Sect. S2.2
	RSM	Reference sector method	An STS estimating stratospheric columns based on the total columns over a reference sector (here 180 to 140° W), assuming longitudinal homogeneity
<i>S</i>	SCD	Slant column density	Concentration integrated along mean light path
	STS	Stratosphere–troposphere separation	The procedure of separating the total column into stratospheric and tropospheric fractions
<i>T*</i>	TR	Tropospheric residue	Difference of total and stratospheric column*, Eq. (3)
<i>w</i>		Weighting factor	
$V_{\text{trop}}$	TVCD	Tropospheric vertical column density	See Eq. (4)
<i>V</i>	VCD	Vertical column density	Vertically integrated concentration
$V^*$			Total VCD*
$V_{\text{strat}}$			Stratospheric VCD
$\vartheta$	lat	Latitude	
$\varphi$	long	Longitude	

\* based on stratospheric AMF

$A_{\text{strat}}/A_{\text{trop}}$  for cloud-free scenes based on AMFs provided in the NASA OMNO2 product.

In this study, we focus on the tropospheric residue  $T^*$  instead of  $V_{\text{trop}}$  for several reasons.

1. As only stratospheric AMFs are applied, biases in the stratospheric estimation can directly be related (factor  $-1$ ) to the respective biases in  $T^*$ .
2. The comparison of TRs among different algorithms instead of TVCDs isolates the effect of the different STS and excludes differences in tropospheric AMFs (which are beyond the scope of this study).
3.  $T^*$  can be determined and is of high interest for the evaluation of STS performance also for clouded scenes with very low tropospheric AMFs.

#### 2.1.4 Version

The description given in this paper and the definition of a priori settings refer to STREAM version v0.92.

## 2.2 Definition of weighting factors

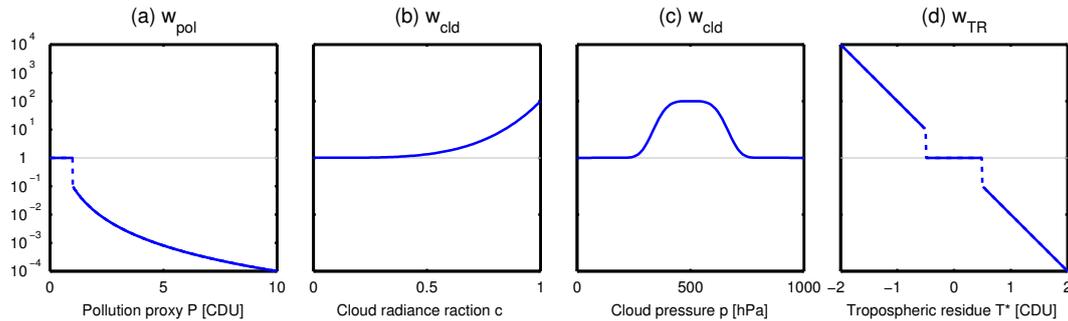
MRSMs usually flag satellite pixels as either clean or (potentially) polluted and skip the latter for the stratospheric esti-

mation. In STREAM, instead, weighting factors for individual satellite pixels determine how strongly they are considered in the stratospheric estimation. Satellite measurements which are expected to have low/high tropospheric contribution are assigned a high/low weighting factor, respectively.

#### 2.2.1 Pollution weight

In order to estimate the stratospheric NO<sub>2</sub> field from total column measurements, only “clean” measurements where the tropospheric column is negligible should be considered. In cases of very high total columns ( $V^* > 10$  CDU), which clearly exceed the domain of stratospheric columns, a tropospheric contribution is obvious, and these measurements are excluded by assigning them a weighting factor of 0.

In most cases, however, the tropospheric contribution to the total column is not that easy to determine. We thus define a pollution weight  $w_{\text{pol}}$  based on our a priori knowledge about the mean spatial distribution of tropospheric NO<sub>2</sub>, reflecting potentially polluted regions. For this purpose, we make use of the multiannual mean tropospheric NO<sub>2</sub> column as derived from SCIAMACHY (Beirle and Wagner, 2012). Based on this climatology, a “pollution proxy”  $P$  is defined as function of latitude  $\vartheta$  and longitude  $\varphi$ .  $P$  indicates the regions affected by tropospheric pollution plus a “safety margin” in order to account for possible advection, while it is



**Figure 1.** Definition of weighting factors (a)  $w_{\text{pol}}$  as a function of the pollution proxy  $P$  (Eq. 5), (b)  $w_{\text{cld}}$  as a function of the cloud radiance fraction (Eq. 6) for a cloud pressure of 500 hPa, (c)  $w_{\text{cld}}$  as a function of the cloud pressure (Eq. 6) for a cloud radiance fraction of 1, and (d)  $w_{\text{TR}}$  as a function of the tropospheric residue (Eq. 7).

undefined for remote unpolluted regions. Details on the definition of  $P$  are given in the Supplement (Sect. S2.2.1), and  $P$  is displayed in Fig. S2d.

The pollution weight  $w_{\text{pol}}$  is then defined as

$$w_{\text{pol}} = 0.1/P(\vartheta, \varphi)^3, \quad (5)$$

where  $P$  is defined, and  $w_{\text{pol}} = 1$  elsewhere. Hence, the higher the pollution proxy  $P$ , the lower the weighting factor and the less the measurement contributes to the stratospheric estimate. Equation (5) is displayed in Fig. 1a, and the resulting map for  $w_{\text{pol}}$  is shown in Fig. 2a. Large continental regions are assigned with a weight  $\leq 0.1$ . Strongly polluted regions like the USA, Europe, or China have weights of 0.01 down to below 0.001. Note that the additional application of the tropospheric residue weight (Sect. 2.2.3) further decreases the weight of satellite measurements containing high tropospheric pollution.

## 2.2.2 Cloud weight

In addition to measurements over remote regions free of tropospheric sources, clouded satellite measurements, where the tropospheric column is shielded, also provide a good proxy for the stratospheric column. Thus, the factor  $w_{\text{cld}}$  is used to increase the weight of clouded satellite pixels. This is achieved by the following definition:

$$w_{\text{cld}} := 10^{2 \times w_c \times w_p} \quad (6a)$$

$$\text{with } w_c := c^4 \quad (6b)$$

$$\text{and } w_p := e^{-\frac{1}{2} \left( \frac{p_{\text{cld}} - p_{\text{ref}}}{\zeta_p} \right)^4}. \quad (6c)$$

$w_c$  reflects the dependency on the cloud radiance fraction (CRF)  $c$ . Due to the exponent of 4, only pixels with large cloud radiance fraction obtain a high weighting factor and contribute strongly to the stratospheric estimation.

$w_p$  describes the dependency on cloud pressure (CP)  $p_{\text{cld}}$ . It is defined as a modified Gaussian (with exponent 4 instead of 2, making it flat-topped) centered at  $p_{\text{ref}} = 500$  hPa

with a width  $\zeta = 150$  hPa; i.e., only cloudy measurements at medium altitudes are assigned a high weighting factor, while high clouds (potentially contaminated by lightning NO<sub>x</sub>) as well as low clouds (where tropospheric pollution might still be visible) are excluded.

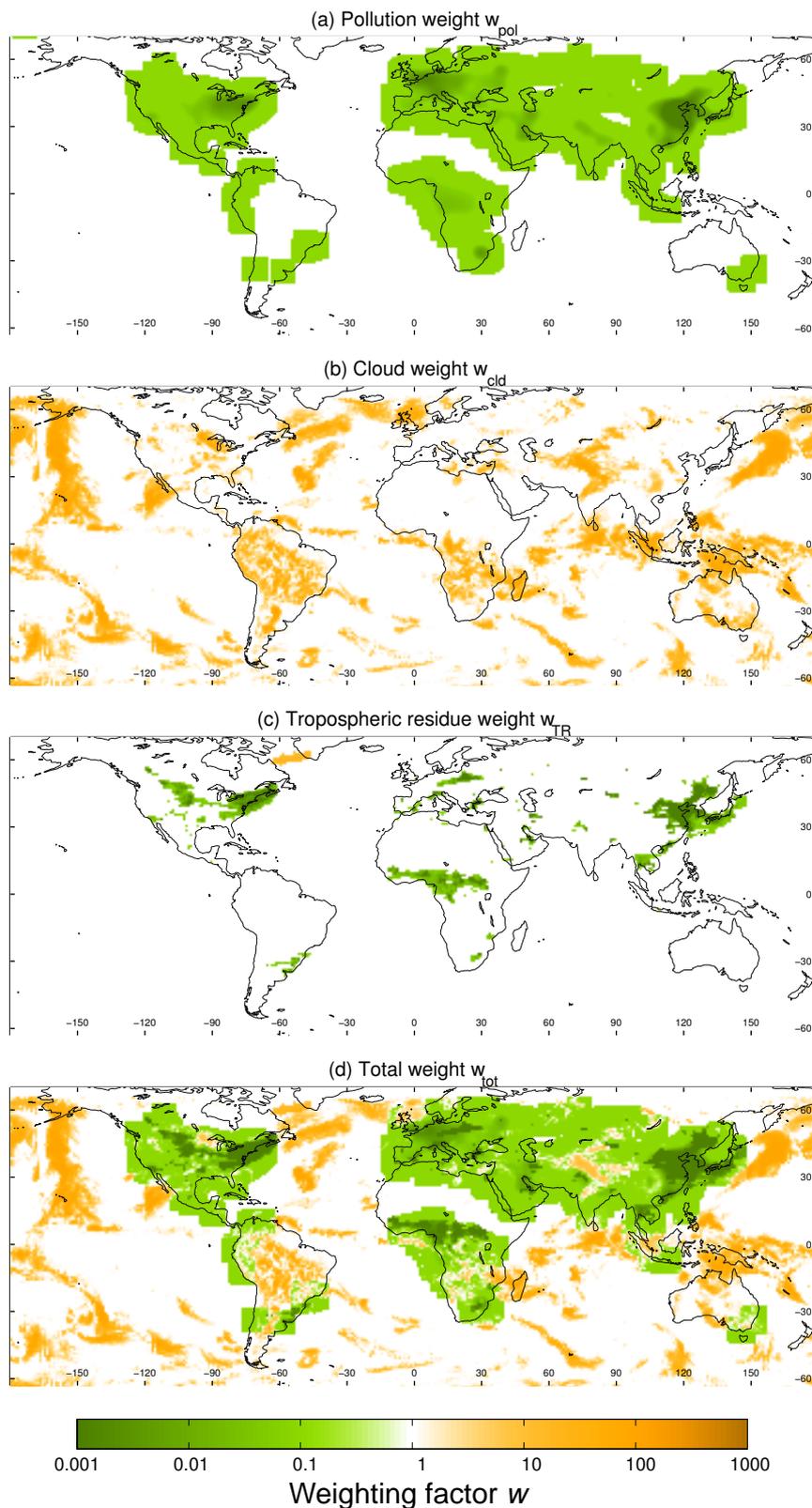
As both  $w_c$  and  $w_p$  yield values in the range from 0 to 1, the factor of 2 in the exponent of Eq. (6a) sets the maximum value of  $w_{\text{cld}}$  to  $10^2$ , which would compensate for pollution weights down to  $10^{-2}$ .

The dependencies of  $w_{\text{cld}}$  on CRF and CP, as defined in Eq. (6), are displayed in Fig. 1b and c, respectively. The spatial pattern of  $w_{\text{cld}}$  is shown exemplarily for OMI CP and CRF on 1 January 2005 in Fig. 2b.  $w_{\text{cld}}$  reaches values up to 100 in several parts of the world, including regions which were pre-classified as potentially polluted, thus competing with a low  $w_{\text{pol}}$  (Fig. 2a).

## 2.2.3 Tropospheric residue weight

STREAM yields global fields of stratospheric VCDs  $V_{\text{strat}}$ , explained in detail below (Sect. 2.3), which allow us to calculate tropospheric residues  $T^*$  according to Eq. (3). While the “true” tropospheric fields are not known, the resulting  $T^*$  can still be used in order to evaluate the STS performance and improve the stratospheric estimate in a second iteration, whenever  $T^*$  clearly indicates an under- or overestimation of  $V_{\text{strat}}$ .

- A high value of  $T^*$  likely indicates tropospheric pollution, in particular over potentially polluted regions. The respective satellite pixels should not be used for the stratospheric estimation.
- As negative columns are nonphysical,  $T^* < 0$  indicates that the stratospheric field has been overestimated. This happens when the weighted convolution with neighboring pixels with high total columns causes the estimated stratosphere to be even higher than the local total columns. In order to avoid this effect, consequently, the respective local total columns should be assigned a



**Figure 2.** Maps of the weighting factors for 1 January 2005 for OMI: (a) pollution weight  $w_{pol}$ , (b) cloud weight  $w_{cld}$ , (c) tropospheric residue weight  $w_{TR}$ , and (d) product of all weighting factors (Eq. 8).

higher weighting factor such that they contribute more strongly to the stratospheric estimate.

We thus define a further weighting factor  $w_{\text{TR}}$ , which weights down/up the pixels associated with a large positive/negative TR, respectively. It turned out, however, that the stratospheric estimate is very sensitive to the definition of  $w_{\text{TR}}$ , and a simple definition based on the TR of individual satellite pixels can easily result in systematic artifacts. This results from  $T^*$  being defined as the difference of  $V^*$  and  $V_{\text{strat}}$  (Eq. 3), i.e., two quantities of the same order of magnitude with non-negligible errors. Thus, the resulting statistical distribution of  $T^*$  inevitably includes negative values. These negative values caused by statistical fluctuations must not be excluded from the probability density function in order to keep the mean unbiased, but they should also not be used as a trigger for weighting up the respective measurement within the stratospheric estimation. Thus,  $w_{\text{TR}}$  should be only applied to significant and systematic deviations of  $T^*$  from 0. This is achieved by the following settings.

1. In contrast to  $w_{\text{cld}}$ , which is defined for each individual satellite measurement,  $w_{\text{TR}}$  is defined based on the TRs averaged over  $1^\circ \times 1^\circ$  grid pixels; i.e., first the values of  $T^*$  within one grid pixel are averaged, reducing statistical noise, before  $w_{\text{TR}}$  is calculated, and the resulting weight is then assigned to all satellite measurements within the grid pixel.
2.  $w_{\text{TR}}$  is only applied when the absolute value of the mean grid box  $T^*$  exceeds a threshold of 0.5 CDU, which is typically larger than the spectral fitting error:

$$w_{\text{TR}} := \begin{cases} 10^{-2 \times T^*} & \text{if } |T^*| > 0.5 \text{ CDU} \\ 1, & \text{else.} \end{cases} \quad (7)$$

3.  $w_{\text{TR}}$  is only applied for grid pixels where the adjacent grid pixels exceed the threshold as well. By this additional condition it is guaranteed that a single outlier in the satellite measurements cannot trigger  $w_{\text{TR}}$ , as every satellite measurement is assigned to exactly one grid pixel (see Sect. 2.3).
4.  $w_{\text{TR}} < 1$ , which is meant to decrease the weight of polluted pixels, is only applied over potentially polluted regions with  $w_{\text{pol}} < 1$ . Without this additional condition, patterns of erroneously enhanced TR caused by stratospheric dynamics would even be amplified by  $w_{\text{TR}}$ .

$w_{\text{TR}}$  could in principle be tuned in multiple iterations. In STREAM v0.92, only one iteration is performed, as a second iteration turned out to have marginal effect (see Sect. S4.2.5).

The dependency of  $w_{\text{TR}}$  on TR (grid pixel average), as defined in Eq. 7, is displayed in Fig. 1d, and the resulting map for  $w_{\text{TR}}$  on 1 January 2005 is shown in Fig. 2c. After the initial stratospheric estimate, STREAM yields high values for

$T^*$  over parts of the USA, Europe, central Africa, and China, resulting in low  $w_{\text{TR}}$ . Observations over these regions are already associated with a low pollution weight. However, due to the additional application of  $w_{\text{TR}}$ , the net weight is lowered further by orders of magnitude, and the respective satellite pixels will hardly contribute to the stratospheric estimate in the next iteration, even in the case of high  $w_{\text{cld}}$ .

In the initial STREAM run, the resulting TR is systematically  $< 0$  over east Canada and Greenland, caused by the asymmetric polar vortex. Over the Labrador Sea, initial values for  $T^*$  are systematically below the threshold of 0.5 CDU and thus trigger a high  $w_{\text{TR}}$ , and the respective observations of low total VCDs contribute strongly to the stratospheric estimate in the next iteration.

Note that, due to the threshold of 0.5 CDU (criterion 2),  $w_{\text{TR}}$  cannot correct small biases such as the expected low bias in TR caused by estimating the stratospheric column from total column measurements.

#### 2.2.4 Total weight

The total weight of each satellite pixel is defined as the product of the individual weighting factors:

$$w_{\text{tot}} := w_{\text{pol}} \times w_{\text{cld}} \times w_{\text{TR}} \quad (8)$$

(i.e., the logarithms as shown in Fig. 2a–c are simply added, resulting in Fig. 2d). The a priori pollution weight can still be recognized in the global pattern but is significantly modified by  $w_{\text{TR}}$  (further reducing the overall weight over, e.g., the USA and China) and  $w_{\text{cld}}$ , which competes with the pollution weights  $< 1$ . In some regions (e.g., west of the Great Lakes, Scotland, or the Himalayas) the cloud weight shifts the initially low  $w_{\text{pol}}$  to a net weight  $> 1$ .

The concept of the combination of different weighting factors is easily extendible by further weights, e.g., based on fire or flash counts in order to account for NO<sub>x</sub> emissions from biomass burning or lightning.

### 2.3 Weighted convolution

Global daily maps of the stratospheric column are derived by applying “weighted convolution”, i.e., a spatial convolution which takes the individual weights for each satellite pixel into account. This approach is an extension of the “normalized convolution” presented in Knutsson and Westin (1993). Weighted convolution at the same time smoothes and interpolates the stratospheric field. A similar approach was used by Leue et al. (2001), who applied the fitting errors of NO<sub>2</sub> SCDs as single weights.

The algorithm is implemented as follows.

- A lat/long grid is defined with  $1^\circ$  resolution. Each satellite pixel is sorted into the matching grid pixel according to its center coordinates. At the  $j$ th latitudinal/ $i$ th longitudinal grid position, there are  $K$  satellite pixels

with the total columns  $V_{ijk}$  ( $k = 1 \dots K$ ) and the weights  $w_{ijk}$ . We define

$$C_{ij} := \sum w_{ijk} \times V_{ijk} \quad (9)$$

and

$$W_{ij} := \sum w_{ijk}. \quad (10)$$

In the case of measurement gaps (i.e.,  $K = 0$ ), both  $C_{ij}$  and  $W_{ij}$  are set to 0.

The weighted mean VCD for each grid pixel is then given as

$$V_{ij} = \frac{C_{ij}}{W_{ij}} \quad (11)$$

for  $K > 0$  and undefined for  $K = 0$  (gaps).

- A convolution kernel (CK)  $G$  is defined (see below). Spatial convolution is applied to both  $C$  and  $W$  (taking the dateline into account appropriately, i.e.,  $i = 1$  and  $i = 360$  are adjacent grid pixels):

$$\overline{C} := G \otimes C, \quad (12)$$

$$\overline{W} := G \otimes W. \quad (13)$$

- The smoothed stratospheric VCD for each grid pixel as derived from weighted convolution is then given as

$$\overline{V}_{ij} := \frac{\overline{C}_{ij}}{\overline{W}_{ij}}. \quad (14)$$

We illustrate this procedure for a simple 1-D example in the Supplement (Sect. S2.3 and Fig. S3).

The degree of smoothing is determined by the definition of the CK  $G$ , which is defined as a 2-D Gaussian in STREAM v0.92 with the longitudinal/latitudinal variances  $\sigma_\varphi^2$  and  $\sigma_\vartheta^2$ , respectively. Generally, information on the stratospheric column over polluted regions should be taken from clean measurements at the same latitude. Thus,  $\sigma_\varphi$  has to be sufficiently large, while  $\sigma_\vartheta$  has to be low as gradients in latitudinal dimension should be mostly conserved. For high latitudes, however, the longitudinal extent of the CK has to be small enough as well in order to be able to resolve the strong gradients caused by the polar vortex.

In order to meet these requirements, we implement the convolution in the following way:

1. Two CKs are defined in order to meet the different requirements for polar vs. equatorial regions (see Fig. S4):

$$\begin{aligned} G^{\text{pol}} &:= G(\sigma_\varphi = 10^\circ, \sigma_\vartheta = 5^\circ) \\ G^{\text{eq}} &:= G(\sigma_\varphi = 50^\circ, \sigma_\vartheta = 10^\circ). \end{aligned} \quad (15)$$

Note that the difference of the CKs, which are defined on a regular degree grid, is even more drastic in kilometer space.

2. Stratospheric VCDs  $V_{\text{strat}}^{\text{eq}}$  and  $V_{\text{strat}}^{\text{pol}}$  are derived for both CK according to Eqs. (12)–(14).

3. The final stratospheric VCD is defined as the weighted mean of both, depending on latitude  $\vartheta$ :

$$V_{\text{strat}} := \cos^2(\vartheta) V_{\text{strat}}^{\text{eq}} + \sin^2(\vartheta) V_{\text{strat}}^{\text{pol}}. \quad (16)$$

By this method, spatial smoothing is wide enough at the equator (needed to interpolate, e.g., the stratosphere over central Africa) but small enough at the polar vortex.

In latitudinal direction, this procedure can cause small, but systematic, biases when stratospheric NO<sub>2</sub> shows significant latitudinal gradients on scales of  $\sigma_{\text{lat}}$  or smaller. To overcome this, STREAM provides the (default) option to run the weighted convolution on “latitude-corrected” VCDs; i.e., the mean dependency of  $V^*$  on latitude is (1) determined (again over the Pacific), (2) subtracted from all individual  $V_{ijk}$ , and (3) added again to the stratospheric estimate from weighted convolution. By this procedure, latitudinal gradients are largely removed for the convolution (but not from the final stratospheric fields), and the systematic biases vanish (as shown in Sect. S2.3).

## 2.4 Data processing

STREAM estimates stratospheric fields and tropospheric residues for individual orbits, using NO<sub>2</sub> measurements of the dayside of the orbit. Note that the effect of changes of local time on stratospheric NO<sub>2</sub> across orbit is generally low (see Sect. S2.4) and is thus neglected within STREAM. For each orbit under investigation, the orbit itself plus the seven previous and subsequent orbits (corresponding to about  $\pm 12$  h in time, or  $\pm 180^\circ$  in space (longitude), for the investigated satellite instruments in polar sun-synchronous orbits) are used for the calculation of  $V^*$ , weighting factors, and thus  $V_{\text{strat}}$  via weighted convolution. For the daily means presented in this study, all orbits where the orbit start date matches the day of interest are averaged.

Alternatively, STREAM can be run in “near-real time” (NRT) mode, in which the 14 past, but no future, orbits are included in the weighted convolution. We discuss the performance of STREAM NRT for the example of GOME-2 in Sect. 5.2.

STREAM v0.92, implemented as a MATLAB script at MPI-C, requires about 10 s for processing one orbit of OMI data on a normal desktop computer (3.4 GHz). Time-consuming steps are, at about equal parts, the sorting of the satellite pixels on the global grid  $V_{ijk}$  and the convolution process, while the time needed for the calculation of weighting factors is negligible.

### 3 Data sets

#### 3.1 Satellite data sets

Several UV/vis satellite instruments provide column measurements of atmospheric NO<sub>2</sub>. Table 1 summarizes the characteristics of the instruments and provides references to the data products used in this study, from which the total NO<sub>2</sub> SCD, the stratospheric AMF, and the cloud fraction/cloud top height are taken as input for STREAM. Below we provide details on the satellite characteristics and the data sets used in this study, starting with OMI (as STREAM was optimized for OMI within TROPOMI verification) and GOME-2, followed by older instruments with particular challenges such as poor spatial coverage (SCIAMACHY) or resolution (GOME).

##### 3.1.1 OMI

In this study we mainly focus on OMI for two reasons.

1. OMI provides daily coverage with small ground pixels. While this already results in a high number of available satellite pixels per day ( $> 10^6$ ), the number of clouded pixels matching the requirements to cause a high  $w_{\text{cld}}$  is also high (more than  $10^5$  pixels have a  $w_{\text{cld}} > 5$ ).
2. STREAM is the STS verification algorithm for TROPOMI. Algorithm testing within TROPOMI verification and comparisons to the TROPOMI prototype algorithm are performed based on actual OMI measurements.

STREAM basically requires  $V^*$  ( $=S/A_{\text{strat}}$ ) as input. For OMI, we use the level 2 “OMNO2” data product (version 3) provided by NASA (Bucsela et al., 2013) and labeled as “Standard Product 2” (SP2) therein, which provides de-striped NO<sub>2</sub> SCDs and stratospheric AMFs<sup>2</sup>. In addition, quality proxies are used to exclude dubious measurements (like those affected by the “row anomaly”<sup>3</sup>). Also information on CRF and CP, which is needed for the calculation of  $w_{\text{cld}}$ , is provided by the OMNO2 v003 hdf files, based on the “improved OMI O<sub>2</sub>–O<sub>2</sub> cloud algorithm” (Bucsela et al., 2013) OMCLD02.

The NASA v003 product involves a STS algorithm based on an MRSM as well. The resulting tropospheric residues of STREAM and NASA v003 are compared and discussed in detail in Sect. 5.1.2.

In addition to the NASA product, we also extract the DOMINO (version 2) level 2 data as provided by TEMIS, for two purposes.

<sup>2</sup>In the DOMINO v2 product, total SCDs are not de-striped, and stratospheric AMFs are only provided up to a solar zenith angle (SZA) of 80°

<sup>3</sup><http://projects.knmi.nl/omi/research/product/rowanomaly-background.php#overview>

1. The TROPOMI “prototype algorithm” (van Geffen et al., 2014) is developed by KNMI based on model assimilation similar to the DOMINO v2 algorithm. Due to the high computational effort of data assimilation, no dedicated TROPOMI verification data set is available for verification. Instead, we compare the results of STREAM directly to DOMINO v2 (Sect. 5.1.1).
2. DOMINO provides TM4 model profiles of NO<sub>2</sub> (needed for the calculation of DOMINO tropospheric AMFs). Here, we use the TM4 data in order to construct synthetic total columns of NO<sub>2</sub> for performance tests of STREAM (see Sect. 3.3).

Both OMI products are based on the same spectroscopic analysis; i.e., both start with the same NO<sub>2</sub> SCD. Note that this SCD is biased high by about 1 CDU due to shortcomings in the spectral retrieval (see van Geffen et al. (2015) and references therein). Recent algorithm refinements have removed this bias (van Geffen et al., 2015; Marchenko et al., 2015), but updated NASA or TEMIS products are not available yet. However, such an overall bias will be interpreted as stratospheric feature by STREAM and thus does not affect its performance (the same holds for the operational NASA and TEMIS STS algorithms). Still, the resulting TRs are expected to decrease slightly as the bias decreases for larger SCDs (Marchenko et al., 2015, Fig. 3 therein).

##### 3.1.2 GOME-2

The GOME-2 instruments on the Metop-A and B satellites provide a time series of almost 10 years with the perspective of continuation until 2025 due to the upcoming instrument on Metop-C. GOME-2 provides a good spatial coverage with moderate satellite ground pixel size.

We applied STREAM to total NO<sub>2</sub> columns from the operational product (GDP 4.7), as provided by DLR in the framework of the Ozone Satellite Application Facilities (O3M SAF), for Metop-A.

The operational product uses an MRSM for STS (Valks et al., 2011, 2015) as well. We compare the results of STREAM and the GDP 4.7 algorithm in Sect. 5.2.

##### 3.1.3 SCIAMACHY

STREAM was applied to the SCIAMACHY VCDs retrieved at MPIC Mainz (Beirle et al., 2010a; Beirle and Wagner, 2012). While OMI provides daily global coverage, the coverage of SCIAMACHY is rather poor (only about one-sixth of the Earth per day), and ground pixels are larger than for OMI (except for swath edges). Consequently, also the number of total (about 60 000) and cloudy (about 4000) pixels per day is much lower than for OMI. Thus, SCIAMACHY can be considered as extreme test case for the performance of STREAM.

One reason for the poor spatial coverage of SCIAMACHY is the measurement mode alternating between nadir and limb geometry. This, however, provides the unique SCIAMACHY feature of a direct measurement of the stratospheric column. We thus compare the TR resulting from STREAM to the MPI-C SCIAMACHY product based on LNM (Beirle et al., 2010a), using the MPI-C retrieval scheme for NO<sub>2</sub> concentration profiles from limb measurements (Kühl et al., 2008) (Sect. 5.3).

### 3.1.4 GOME

GOME was the first nadir-viewing spectrometer in the UV/vis spectral range with a spectral resolution enabling DOAS analyses. Due to large ground pixel size (320 km across track), only a low number of (total as well as clouded) satellite pixels per day is available. We nevertheless included GOME in this analysis in order to investigate to what extent STREAM can be applied within homogenized retrievals for multiple satellite instruments, as planned within the QA4ECV (Quality Assurance for Essential Climate Variables) project<sup>4</sup>. We apply STREAM to the VCDs provided by TEMIS (Boersma and Eskes, 2004) and compare the resulting TRs to a simple RSM (Sect. 5.4).

## 3.2 Model data

For comparisons, and for the calculation of synthetic total columns for performance tests of STREAM, we make use of stratospheric NO<sub>2</sub> as provided by the ECHAM5/MESy Atmospheric Chemistry (EMAC) model, which is a modular global climate and chemistry simulation system (Jöckel et al., 2006, 2010, 2016).

We use the results from simulation *RCISD-base-10a* of the ESCiMo (Earth System Chemistry integrated Modelling) project as detailed by Jöckel et al. (2016). Here, only basic information on this specific simulation is summarized.

The model results were obtained with ECHAM5 version 5.3.02 (Roeckner et al., 2006) and MESSy version 2.51 at T42L90MA resolution, i.e., with a spherical truncation of T42, corresponding to a quadratic Gaussian grid of approx. 2.8° by 2.8° in latitude and longitude, and 90 vertical hybrid pressure levels up to 0.01 hPa. The dynamics of the general circulation model was nudged by Newtonian relaxation towards ERA-Interim reanalysis data (Dee et al., 2011).

Simulation *RCISD-base-10a* was selected from among the various ESCiMo simulations for several reasons:

- a. it has been nudged to reproduce the “observed” synoptic situations;
- b. its stratospheric resolution is, with  $\simeq 65$  levels, finer compared to other simulations from the ESCiMo project;

- c. the simulated total column and tropospheric partial column ozone compare well with observations (Jöckel et al., 2016); and
- d. the precursor emissions from the land transport sector are most realistic in comparison to other simulations.

In conclusion, this simulation represents the state-of-the-art in terms of numerical simulation of the atmospheric chemistry. Moreover, the applied nudging technique allows a direct comparison with observational data, since the simulated meteorological situation corresponds to the observed.

Specifically for this study, the submodel SORBIT (Jöckel et al., 2010) was used to extract NO<sub>2</sub> mixing ratios along the sun-synchronous orbit of the Aura satellite, thus matching the local time of OMI observations. Stratospheric VCDs were calculated by vertical integration of the modeled NO<sub>2</sub> mixing ratios between the tropopause height (as diagnosed according to the WMO definition based on lapse rate equatorwards of 30° north/south and as iso-surface of 3.5 PVU potential vorticity poleward of 30° latitude) and the top of the atmosphere.

In this study, we make use of the modeled stratospheric columns for two purposes.

1. We perform a simple model-based STS for comparison. To remove systematic biases between satellite measurements and EMAC, a latitude dependent offset is determined in the Pacific and corrected for globally, similar as in Richter et al. (2005) and Hilboll et al. (2013). We refer to this EMAC-based STS as STS<sub>EMAC</sub> and applied it to OMI data (Sect. 5.1.3).
2. Stratospheric VCDs from EMAC are used to construct a synthetic data set of total NO<sub>2</sub> VCDs for performance tests of STREAM (see next section).

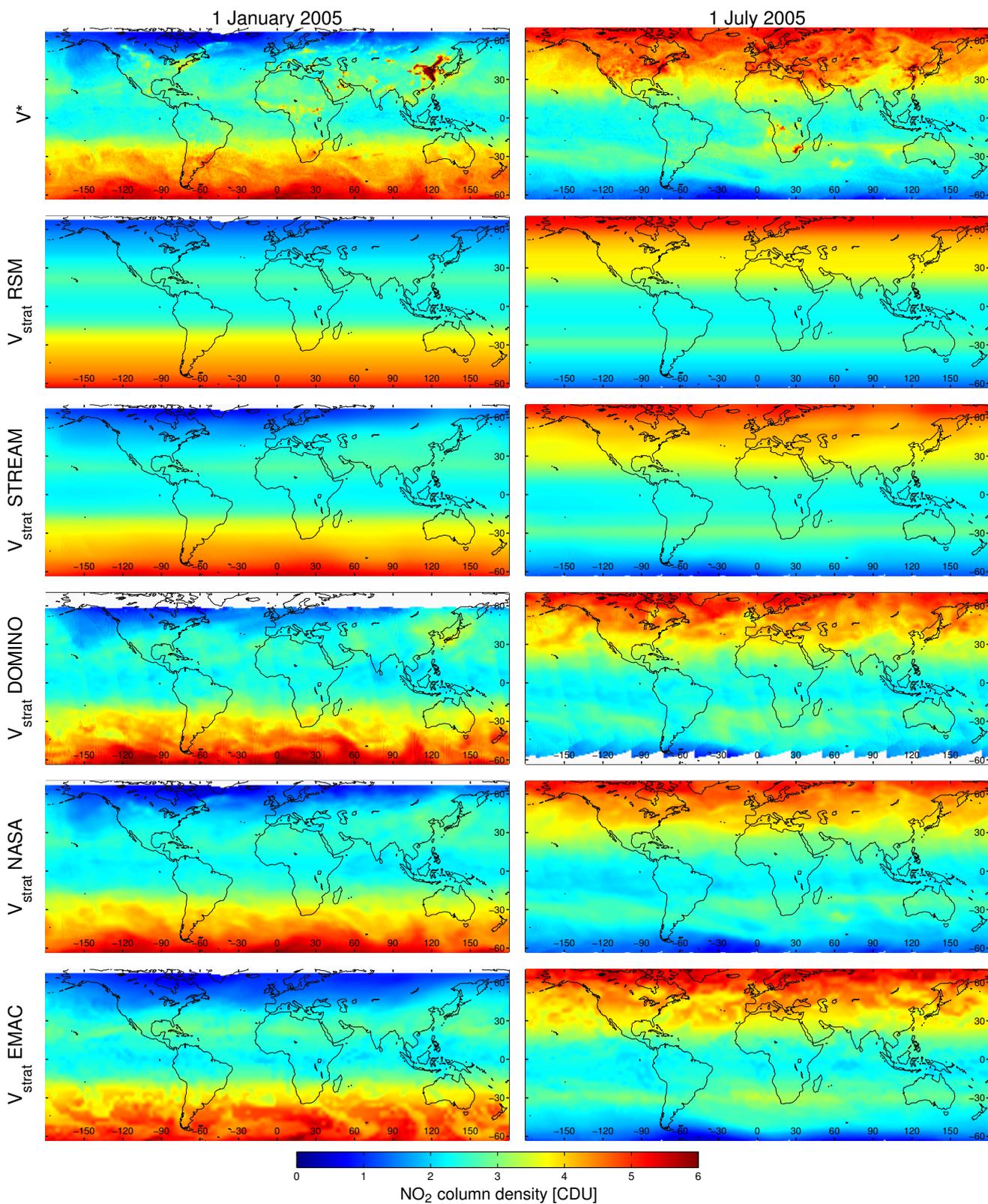
## 3.3 Synthetic VCD

We test the performance of STREAM on synthetic VCDs, which allows a quantitative comparison of the estimated TR to the a priori “truth”. The input to STREAM, i.e., synthetic total columns of NO<sub>2</sub>, should realistically represent (a) stratospheric chemistry and dynamics, (b) tropospheric emissions, transport, and chemistry, (c) cloud fields, and (d) the satellite sampling. For these purposes, we construct synthetic NO<sub>2</sub> column densities  $V^*$  based on

- a. stratospheric VCDs from EMAC<sup>5</sup> at AURA overpass time (Sect. 3.2),
- b. modeled tropospheric VCDs from TM4 (Sect. 3.1.1), and

<sup>5</sup>Stratospheric columns are taken from EMAC rather than TM4, as the latter does not represent a free model run of stratospheric chemistry and dynamics but uses the satellite measurements for assimilation.

<sup>4</sup><http://www.qa4ecv.eu/>



**Figure 3.** Total OMI VCD  $V^*$  (top) and the resulting stratospheric estimate  $V_{\text{strat}}$  from RSM (second row) and STREAM (third row) for 1 January (left) and 1 July (right) 2005. Resulting  $V_{\text{strat}}$  from other algorithms are included as well for comparison (see Sect. 5).

- c. measured cloud properties and the respective tropospheric AMFs from OMI as provided in the DOMINO NO<sub>2</sub> product.

Synthetic TRs are given as  $T^* = V_{\text{trop}} \times A_{\text{trop}} / A_{\text{strat}}$  (compare Eq. 4). Synthetic total columns  $V^*$  are then calculated as  $V_{\text{strat}} + T^*$  (Eq. 3) and fed into STREAM. The resulting fields of stratospheric VCDs and the respective TRs can then be compared to the a priori “truth”. Synthetic  $V_{\text{strat}}$ , TVCD, and  $T^*$  are displayed in Fig. S7 for 2 selected days.

#### 4 Algorithm performance

In this section we analyze the performance of STREAM. As the true stratospheric VCD is not known, the error of any STS algorithm is not easily accessible. Still, the STS performance can be evaluated based on the properties of the resulting TR: in remote regions without substantial NO<sub>x</sub> emissions,  $T^*$  should generally be low but still positive (about 0.1 CDU; Valks et al., 2011). Also the variability of  $T^*$  over both space and time should be low in regions free of tropospheric sources.

Below, we investigate the characteristics of  $T^*$  from STREAM (Sect. 4.1) and its dependency on a priori settings (Sect. 4.2) for OMI measurements. In addition, the error of  $T^*$  is quantified based on synthetic data (Sect. 4.3). Application of STREAM to other satellite instruments and the comparisons between STREAM and other STS algorithms are provided in Sect. 5.

##### 4.1 Performance of STREAM for OMI compared to RSM

Figure 3 displays the OMI daily mean VCD  $V^*$  (top) as well as the respective stratospheric field from RSM (second row) and STREAM (third row) for 1 January (left) and 1 July 2005 (right), respectively. The overall latitudinal as well as longitudinal dependencies are clearly reflected in the stratospheric fields, while small-scale stratospheric features are lost by the spatial convolution. Figures 4 and 5 display the resulting TRs, respectively, for both daily (top) and monthly (bottom) means. Figure 6 summarizes the daily and monthly statistical properties of TR, i.e., the median as well as 10th/90th and 25th/75th percentiles (light/dark bars) for different regions (see Fig. S8 for an illustrative sketch of the meaning of the percentile bars, as well as the definition of regions).

Overall, spatial patterns of TR are similar for RSM and STREAM, in particular the enhanced values reflecting tropospheric pollution over, e.g., the USA, central Africa, or China. However, RSM reveals several artifacts of both enhanced as well as systematically negative TR as a consequence of the simple assumption of zonal invariability of stratospheric NO<sub>2</sub>. For instance, on 1 January 2005, VCDs over northern Canada are lowered due to the polar vortex (Fig. 3 top left). Consequently, the simple RSM results in

negative  $T_{\text{RSM}}^*$  down to  $-0.7$  CDU (Fig. 4). In contrast,  $T_{\text{RSM}}^*$  over northeastern Russia is quite high ( $> 0.5$  CDU). This pattern is slightly reduced but still present in the monthly mean (see the statistics of  $T_{\text{RSM}}^*$  for high latitudes in Fig. 6).

This artifact is largely reduced by STREAM (Fig. 5 top left). The spread of  $T^*$  at high latitudes is more than 3 times lower than that of  $T_{\text{RSM}}^*$  (Fig. 6). Also for July, systematic structures showing up in  $T_{\text{RSM}}^*$  (in polar regions, but also in the Indian ocean at 30–60° S) are largely reduced in STREAM.

Over the Pacific,  $T_{\text{RSM}}^*$  is, by construction, 0 on average.  $T^*$  is systematically higher by about 0.1 CDU (Fig. 6). This results from the emphasis of clouded pixels used for STREAM, which directly reflect the stratospheric rather than the total VCD. This additional advantage of STREAM over RSM is further discussed below (Sect. 5.6).

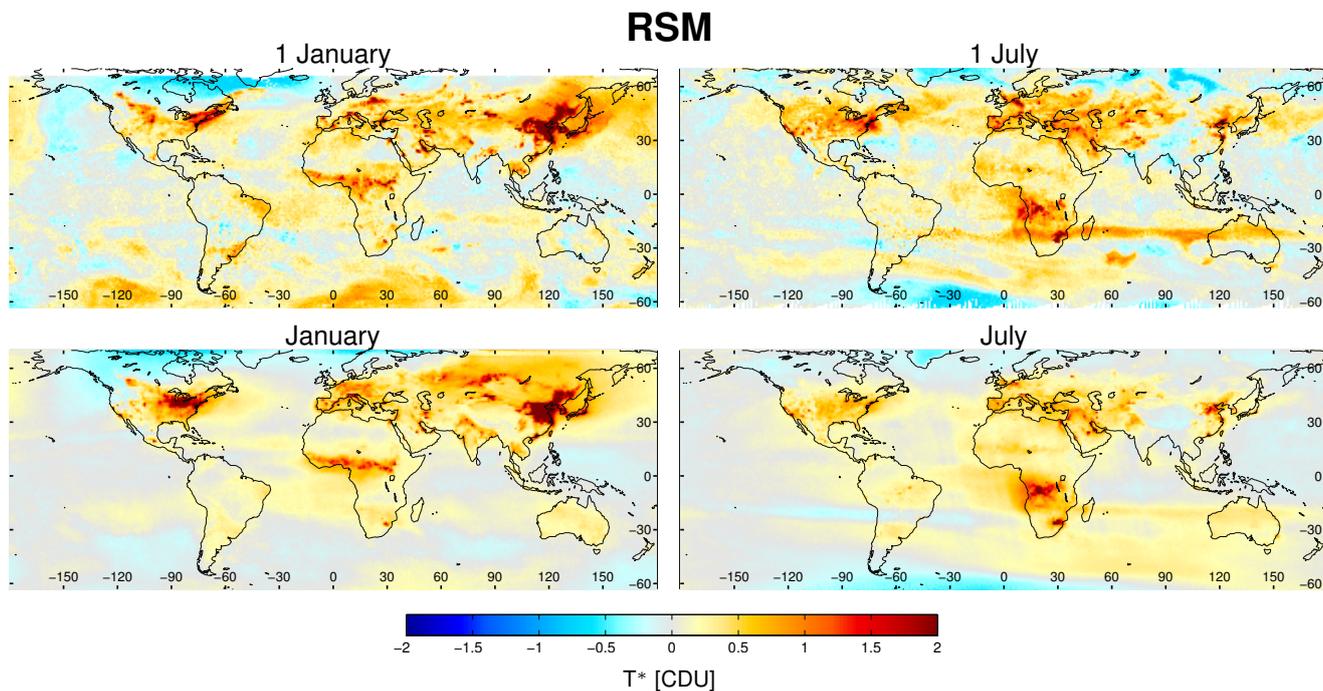
As both RSM and STREAM generally assume stratospheric patterns of NO<sub>2</sub> to be smooth, i.e., do not resolve longitudinal variations at all (RSM) or on scales  $< \sigma_{\varphi}$  (STREAM), the small-scale variations in daily total VCDs (Fig. 3 top) are transferred to the TR, resulting in “patchy” daily TRs ranging from about  $-0.1$  up to  $+0.4$  CDU in remote regions (10th–90th percentiles). In the monthly means, however, these patchy structures have mostly vanished (both for RSM and STREAM), as the spatial patterns of different days at variable locations cancel each other out. The remaining systematic patterns in monthly mean  $T^*$  have generally larger spatial scales and are within 0 up to  $+0.25$  CDU in remote regions.

On 1 July, a band of enhanced  $V^*$  shows up around 20–30° S, where (a)  $V^*$  is higher in the Indian Ocean compared to the Pacific and (b) the structure of enhanced  $V^*$  is “tilted” in the Pacific (see Fig. 3 top right); i.e., the RSM assumption of zonal invariance is not fulfilled. Consequently, the RSM results in extended horizontal structures (“stripes”) of low/high-biased  $T^*$  over South America and the Indian Ocean, respectively, ranging from  $-0.5$  up to almost 1.0 CDU (Fig. 4 top right). Again, temporal averaging reduces the amplitude, but systematic patterns of about  $\pm 0.4$  CDU remain in the monthly mean (Fig. 4 bottom right). As STREAM also assumes a weak variation of  $V_{\text{strat}}$  with longitude, in particular at low latitudes, the artifacts in  $T^*$  are very similar to those of  $T_{\text{RSM}}^*$  at 20–30° S. Note that this artifact is particularly strong in July 2005 (as compared to 2010; see Sect. 5.1.4).

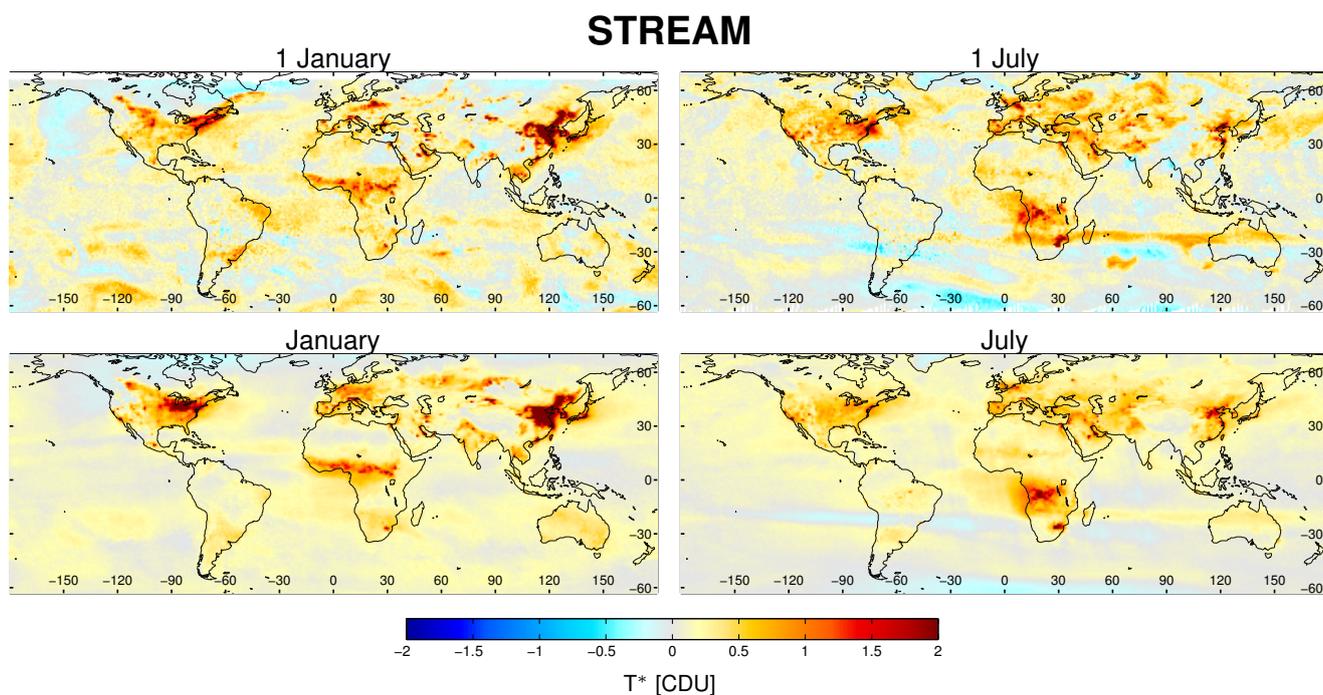
In Sect. 5, TRs from STREAM are investigated for other satellite instruments and compared to other STS algorithms, and the advantages and limitations of STREAM are discussed further.

##### 4.2 Impact of a priori settings

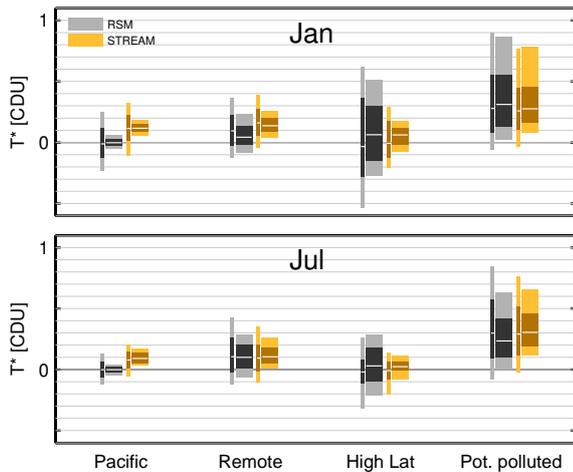
STREAM determines the stratospheric NO<sub>2</sub> VCD  $V_{\text{strat}}$  based on weighting factors as described in Sect. 2. The resulting TRs thus depend on the weighting factor



**Figure 4.** OMI tropospheric residues  $T^*$  based on RSM for January (left) and July (right) 2005 for the first day of the month (top) and the monthly mean (bottom).



**Figure 5.** OMI tropospheric residues  $T^*$  based on STREAM for January (left) and July (right) 2005 for the first day of the month (top) and the monthly mean (bottom).



**Figure 6.** Regional statistics of OMI tropospheric residues  $T^*$  from RSM and STREAM for January (top) and July (bottom) 2005. Light and dark bars reflect the 10–90th and 25–75th percentiles, respectively. The median is indicated in white. Narrow bars show the statistics for the first day of the month, wide bars those of the monthly means (see also Fig. S8 left for illustration). The regions are defined in Fig. S8 right. “High latitudes” refer to the respective hemispheric winter only.

definition and convolution settings. We performed runs of STREAM with one-by-one modifications of each parameter and compared the results to the baseline setting. Overall, the effects of a priori settings on  $T^*$  have been found to be rather small (of the order of 0.1 CDU), and the STREAM results are thus robust with respect to the parameters chosen in v0.92.

Below, we summarize the main findings of the performed sensitivity studies. Figures and details are provided in the Supplement.

#### 4.2.1 Impact of cloud weight

The cloud weight  $w_{\text{cld}}$  was varied (a) by setting it to 1 (i.e., not accounting for cloud properties at all), (b) increasing  $w_{\text{cld}}$  by a factor of 10 for clouded pixels, (c) including high-altitude clouds in the calculation of  $w_{\text{cld}}$ , and (d) including low-altitude clouds in the calculation of  $w_{\text{cld}}$ .

- When no  $w_{\text{cld}}$  is applied, the tropospheric estimate over the Pacific is  $\approx 0$ , as for the classical RSM, instead of about 0.1 CDU for the baseline. This difference corresponds to the order of the tropospheric background of NO<sub>2</sub>. Over potentially polluted regions, however, the difference to the baseline is larger (0.2 CDU). Here, the stratospheric estimate is additionally biased high due to missing supporting points over continents.
- The “high  $w_{\text{cld}}$ ” scenario is achieved by modifying Eq. (6a) from  $w_{\text{cld}} := 10^{2 \times w_c \times w_p}$  to  $w_{\text{cld}} := 10^{3 \times w_c \times w_p}$ ; i.e.,  $w_{\text{cld}}$  is increased by a factor of 10

for cloudy pixels of mid-altitude but stays unchanged for cloud-free pixels. In this scenario, measurements over clouds by far dominate the stratospheric estimate, yielding lower  $V_{\text{strat}}$ , and thus higher  $T^*$ , compared to the baseline. However, the difference is very small ( $< 0.05$  CDU). In addition, the variability of  $T^*$  is generally slightly higher in case of a 10 fold increased  $w_{\text{cld}}$ .

- When high-altitude clouds are included in the calculation of  $w_{\text{cld}}$ , the resulting TR hardly changes at all, indicating that the impact of lightning NO<sub>x</sub> on NO<sub>2</sub> satellite observations is generally small.
- The inclusion of low-altitude clouds has almost no effect as well, as expected over clean regions. Over potentially polluted regions, however, it is expected that low-altitude clouds result in increased total columns  $V^*$  as soon as there is significant NO<sub>2</sub> above or within the cloud, causing high tropospheric AMFs. Consequently,  $V_{\text{strat}}$  is expected to be biased high, and  $T^*$  biased low over potentially polluted regions, when low clouds are included in the calculation of  $w_{\text{cld}}$ . This effect was indeed found, but the absolute change is rather small ( $< 0.1$  CDU in winter, almost 0 in summer). This weak dependency on the inclusion of low-altitude clouds probably results from the conservative definition of  $w_{\text{pol}}$ , which is already very low over regularly polluted regions.

Following the argument that cloudy observations provide a direct measurement of the stratospheric column, a higher cloud weight would be expected to be more favorable and to result in higher tropospheric background over the Pacific. This is indeed observed for OMI. For other satellite instruments, however, results are somewhat contradictory (see Sects. 5.3, 5.4, and 6). Thus, the definition of  $w_{\text{cld}}$  in Eq. (6) is kept as a compromise in order to have common algorithm settings across different satellite platforms.

#### 4.2.2 Impact of convolution

In STREAM, two different CK are applied, yielding two stratospheric estimates, and the final  $V_{\text{strat}}$  is calculated as weighted mean of both (see Sect. 2.3 and Eqs. 15 and 16). We tested the impact of the choice for CK by applying both the polar (“narrow”) and equatorial (“wide”) CK globally. The narrow CK, and thus the potential range of influence of satellite pixels with high weights, is limited to about  $2 \times \sigma_\varphi = 20^\circ$  in longitude. This potentially leads to biases over continents caused by spatial interpolation. Thus, the resulting  $T^*$  is (too) low over central Africa. Overall, median  $T^*$  over potentially polluted regions is lower compared to the baseline settings by about 0.1 CDU.

For wide CK, however, the longitudinal gradients at high latitudes are not resolved anymore. Consequently, the spatial variability of daily  $T^*$  at high latitudes is increased by a factor of 2. We conclude that our choice of the combined CK

for high and low latitudes is a good compromise for realizing weighted convolution.

#### 4.2.3 Impact of latitude correction

When the initial correction of the latitudinal dependency of  $V^*$  over the Pacific is omitted, the resulting TR reveals global stripes with negative values around the equator and maxima ( $\approx 0.5$  CDU) at about  $30^\circ$  N/S, both in winter and summer.

#### 4.2.4 Impact of the number of considered orbits

In STREAM baseline settings, for each orbit, stratospheric estimation is based on the previous and subsequent seven orbits, corresponding to full global coverage for OMI. Switching this parameter to either  $\pm 14$  or  $\pm 3$  orbits has almost no impact on the resulting TR.

In case of NRT application of STREAM, no subsequent orbits are available, and the previous 14 orbits have to be considered. This setup also results in essentially the same  $T^*$  statistics (compare Sect. 5.2).

#### 4.2.5 Impact of tropospheric residue weight

In STREAM v0.92, one iteration for  $w_{\text{TR}}$  is applied. When  $w_{\text{TR}}$  is omitted, the spread of  $T^*$  slightly increases for high latitudes. A second iteration does not yield a further improvement. Lowering the threshold in Eq. (7) from 0.5 to 0.3 CDU results in a slightly lower spread of  $T^*$  at high latitudes in summer.

#### 4.2.6 Impact of pollution weight

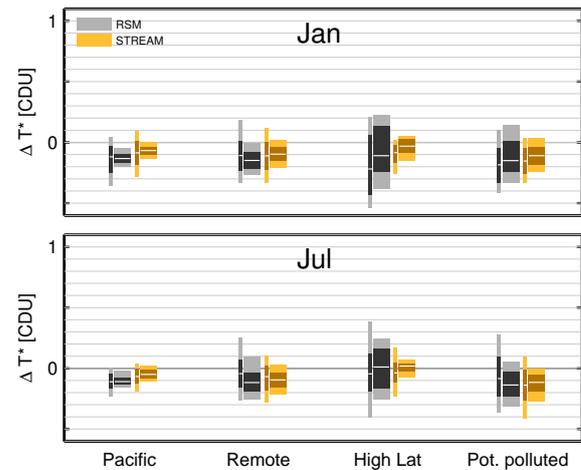
The impact of pollution weight is investigated by multiplying  $w_{\text{pol}}$  (where different from 1, compare Fig. 2a) by 0.1 (“low  $w_{\text{pol}}$ ”) or 10 (“high  $w_{\text{pol}}$ ”). In the first case, the resulting pollution weight over most continents is below 0.01, while in the second case it is increased to 1 (meaning that  $w_{\text{pol}}$  is switched off) except for industrialized pollution hotspots.

In remote regions, the change of pollution weight has almost no impact. In potentially polluted regions, the impact is only moderate as well. Low  $w_{\text{pol}}$  does not differ much from the baseline, as the latter already assigns rather low weighting factors to potentially polluted pixels; a further decrease by factor 0.1 thus does not change much.

Only for high  $w_{\text{pol}}$  can a significant change of TR be seen; in this case, the inclusion of more partly polluted observations causes a high bias in the stratospheric estimate and the resulting TRs are biased low by almost 0.1 CDU in winter.

### 4.3 Performance for synthetic data

In order to estimate the uncertainties of the STREAM stratospheric estimate (and thus tropospheric residues), we apply the algorithm to synthetic input data, as defined in Sect. 3.3, for which the “true” stratospheric fields and TR are known. Again, a simple RSM is applied as well for comparison.



**Figure 7.** Regional statistics of the error of  $T^*$  from STREAM, i.e., the difference of estimated and a priori TR (based on synthetic total columns as defined in Sect. 3.3).

Figure 7 displays the statistics of the error of  $T^*$ , i.e., the difference  $\Delta T^*$  of estimated and a priori TR, which equals the difference between the true and the estimated stratospheric VCD, for different regions. The spatial patterns of  $\Delta T^*$  are shown in the Supplement (Fig. S20).

Over the Pacific, RSM results in TR biased low by 0.1 CDU. With STREAM, the bias is reduced (0.05 CDU) but not completely removed. On 1 January 2005,  $\Delta T^*$  shows a variability of almost 0.4 CDU (10th to 90th percentile) for both algorithms. This is mainly caused by the small-scale structures of stratospheric NO<sub>2</sub> in EMAC over the Pacific, in particular at southern latitudes (see Fig. S7), which are resolved by neither STREAM nor RSM. The respective spatial variability of the monthly mean, however, is much lower (about 0.1 CDU).

Again, the simple RSM results in large biases and high variability of  $\Delta T^*$  at high latitudes, which are largely reduced by STREAM.

Overall, the agreement of a priori and estimated  $T^*$  from STREAM is very good, in particular for monthly means. Remaining systematic biases are about  $-0.1$  CDU over potentially polluted regions; i.e., resulting TRs are slightly underestimated, as expected due to the general approach of using total column measurements as proxy for the stratospheric estimation.

The application of STREAM to synthetic data thus provides a valuable estimate of the algorithm’s accuracy. One might think that using the synthetic data for optimizing the definition of weighting factors is the next step forward. However, we refrain from doing so due to some contradictory results for different instruments. Concretely, the remaining bias in TR for synthetic data of about 0.1 CDU could be further reduced by increasing  $w_{\text{cld}}$ . This, however, has adverse ef-

fects on SCIAMACHY and GOME results (see Sects. 5.3 and 5.4).

## 5 Comparison to other algorithms and discussion

In this section, we apply STREAM to different satellite instruments, compare the results to various existing STS algorithms, and discuss the challenges, limitations, and uncertainties of STS in general and STREAM in detail.

### 5.1 OMI

As shown in the previous section, STREAM as applied to OMI data generally shows a good performance (Figs. 5 and 6). The systematic artifacts of a simple RSM, such as the large variability of  $T^*$  at high latitudes, are largely removed by STREAM. In addition, the application of  $w_{\text{cld}}$  emphasizes cloudy observations which directly reflect the stratospheric column. Mean  $T^*$  over the Pacific is thus not 0 anymore as in RSM, and an additional correction for the tropospheric background is not required in STREAM.

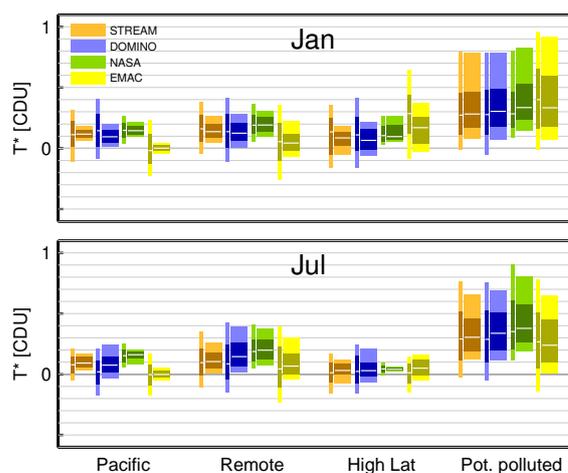
The sensitivity of STREAM on a priori parameters has been found to be small. Remaining monthly mean TRs in clean regions and their variability are of the order of 0.1 CDU.

Below, we compare the OMI results for 2005 to other algorithms, i.e., the operational DOMINO (Sect. 5.1.1) and NASA (Sect. 5.1.2) data products as well as a simple model-based correction using EMAC (Sect. 5.1.3). Figure 8 summarizes the statistics of regional  $T^*$  from the different algorithms. Note that only coincident measurements where all four data products exist are included in Fig. 8 in order to allow for a meaningful comparison; in particular, high latitudes in hemispheric winter are skipped, as DOMINO data are not provided for  $\text{SZA} > 80^\circ$ . Thus, the statistics for STREAM slightly differ from those shown in Fig. 6.

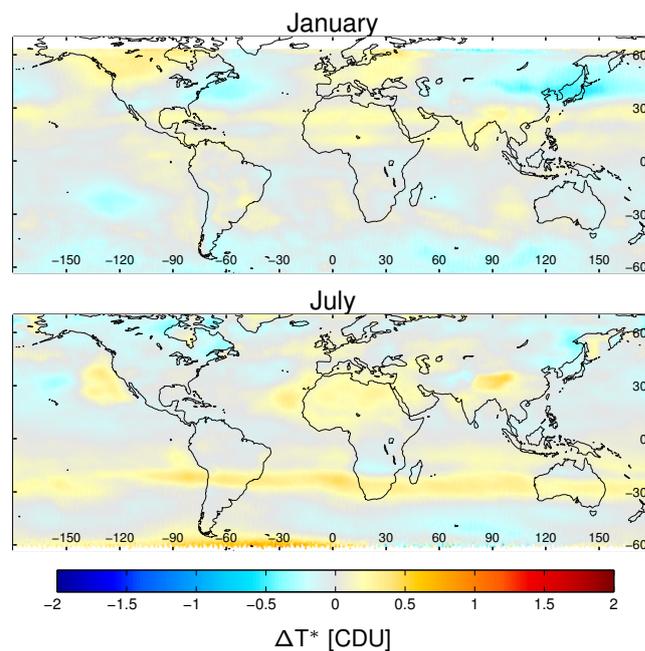
#### 5.1.1 Comparison to DOMINO

STREAM is part of the TROPOMI verification activities. The operational TROPOMI (“prototype”) algorithm for STS of NO<sub>2</sub> (van Geffen et al., 2014) was developed by KNMI, based on the DOMINO data processor for OMI (Boersma et al., 2007, 2011) (Boersma et al., 2011). The STS therein is done by assimilating the satellite measurements in the CTM TM4 (Dirksen et al., 2011).

For TROPOMI verification, we compare STREAM results for OMI to the respective DOMINO product as shown in Fig. S21. On a daily basis, “patchy” patterns of enhanced as well as negative TR show up over remote regions (Fig. S21), which result from the dynamical features already present in total VCDs (Fig. 3) combined with the respective dynamics prognosed by the model; spatial mismatch of these patterns can easily cause biases of the estimated TRs in both directions. Interestingly, some patterns look even reversed as



**Figure 8.** Regional statistics of OMI tropospheric residues  $T^*$  from different STS algorithms for January (top) and July (bottom) 2005. Note that the values for STREAM slightly differ from Fig. 6, as here only coincident satellite pixels of STREAM, DOMINO, and NASA are included.



**Figure 9.** Monthly mean difference of tropospheric residues  $T^*$  from DOMINO and STREAM for OMI measurements in January (top) and July (bottom) 2005.

compared to STREAM (Fig. 5), for instance southeast from South Africa (around  $50^\circ\text{S}$ ,  $50^\circ\text{E}$ ). In the monthly means, these patches again are mostly canceled out.

Mean regional TRs (Fig. 8) are very similar between STREAM and DOMINO. However, the variability of  $T^*$  is slightly higher for DOMINO, in particular at high latitudes, as well as in the Pacific and in remote regions in July.

Figure 9 displays the differences of the monthly mean TR between STS\_EMAC and STREAM for January and July 2005. Overall, the differences are quite small (below  $\pm 0.1$  CDU for 65 % of the world between 60° S and 60° N). Nonetheless, the monthly means reveal systematic regional deviations of more than  $\pm 0.3$  CDU (for less than 3 % of the world).

In January, TRs over East Asia at high latitudes are systematically higher for STREAM. This is probably related to an underestimation by DOMINO, as the DOMINO TRs are very low and partly negative in this region. Over North America, TRs from STREAM are higher than from DOMINO at the east coast but vice versa over western Canada. In both cases, the lower TR is slightly negative, indicating an overestimation of  $V_{\text{strat}}$  from DOMINO/STREAM at the east/west coast, respectively.

In July, the TR reveals a “stripe”-like structure at about 30° S, as already discussed in Sect. 4.1. In DOMINO, similar bands of enhanced tropospheric residue are found around 30° S, in particular in the Indian ocean. As the amplitude and width of these bands is different for STREAM and DOMINO, this feature is most striking in the difference map; TRs around 30° S are generally higher for DOMINO.

DOMINO reveals some patches of systematically enhanced TRs that are not observed by STREAM and thus show up in the difference map as well (west of the USA, west of the Sahara, Himalaya). Reasons for these regionally enhanced TRs (and thus low-biased stratosphere) have to be investigated in future studies.

### 5.1.2 Comparison to NASA

The official OMI NO<sub>2</sub> product provided by NASA uses an MRSM for STS as well, as described in Bucsele et al. (2013). Daily and monthly maps of TR from NASA (OMNO2 v003/SP2) are shown in Fig. S22.

The NASA STS corrects for the tropospheric background based on a “fixed model estimate” (Bucsele et al., 2013). Consequently, TRs are about 0.1 to 0.3 CDU over clean regions throughout the world.

TRs from NASA are impressively smooth even on a daily basis. This results from the STS algorithm which, over clean regions, interprets the difference between the total column and the (small) modeled tropospheric column as stratospheric column whenever the quotient of the modeled tropospheric slant column and stratospheric AMF (matching our definition of  $T^*$ ) goes below a threshold of 0.3 CDU. Thus, at southern high latitudes in July (completely classified as unpolluted in the NASA algorithm), the TR is almost  $0 \pm 0$ , i.e., shows no variability at all (compare Fig. 8) just by construction, as all the variability present in the total column was assigned to the stratospheric column (compare Fig. 3).

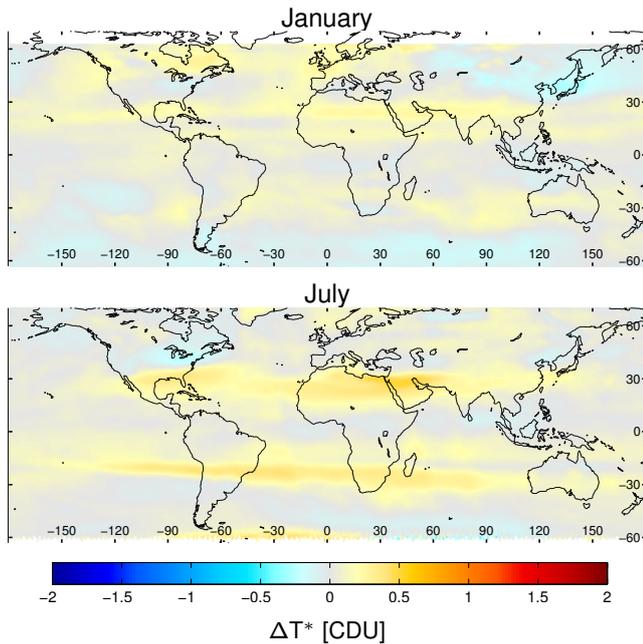
While this is probably a reasonable procedure over completely clean regions, we would like to point out the following.

1. The smoothness of NASA TR over oceans is not surprising, as it is reached by construction. In particular, the smooth patterns of TR over oceans allow no conclusion on the NASA STS performance over polluted continental regions, where TRs are based on interpolated stratospheric fields, just as in STREAM.
2. The NASA procedure of assigning the total column variability in clean regions completely to the stratospheric estimate also removes any cloud dependency of the TR, which affects applications such as profile retrievals by cloud slicing (e.g., Belmonte Rivas et al., 2014).
3. The NASA procedure runs the risk of labeling episodic NO<sub>2</sub> transport events over oceans (Zien et al., 2014) as stratospheric pattern. Bucsele et al. (2013) perform an automatic “hotspot” identification and elimination scheme to avoid this. Nonetheless, on 1 January, a NO<sub>2</sub> transport event can be seen in the total VCD east of Canada (Fig. 3 top left) which is similar to the “meteorological bomb” described in Stohl et al. (2003). This event is clearly visible in  $T^*$  from STREAM (Fig. 5 top left) but only weakly in  $T^*$  from NASA (Fig. S22 top left). The reason for this discrepancy is that the local enhancement of NO<sub>2</sub> is partly classified as a stratospheric feature in the NASA product, as illustrated in Fig. S23 (left).

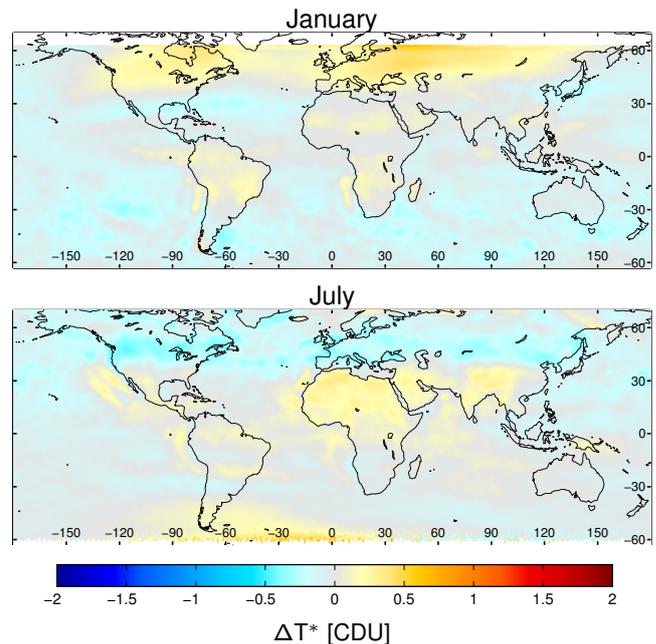
Figure 10 displays the differences of the monthly mean TR for January and July 2005. Again, overall agreement is very good: in January, both products agree within 0.1 CDU for 69 % of the Earth and within 0.3 CDU anywhere. In July, agreement within 0.1/0.3 CDU is found for 64 %/94 % of the Earth (for latitudes below 60°), respectively. Again, the band at 30° S sticks out in the difference map as discussed above. Highest deviations of up to 0.5 CDU, however, are observed over the Sahara. Within the NASA STS, the Sahara is masked out completely, as the high albedo and low cloud fractions result in high tropospheric AMFs, such that even low tropospheric VCDs could contribute significantly to the total column. In STREAM, however, large parts of the Sahara are treated as unpolluted and are assigned with  $w = 1$ . A close check of the stratospheric estimates from STREAM and NASA over the Sahara reveals that the large deviation probably results from both a high-biased  $V_{\text{strat}}$  by STREAM and a low-biased  $V_{\text{strat}}$  by NASA (see Fig. S23 right).

### 5.1.3 Comparison to STS<sub>EMAC</sub>

We have used the stratospheric 3-D mixing ratios provided by EMAC in order to perform a simple model-based STS, similar to Hilboll et al. (2013). First, the latitude-dependent offset between EMAC and OMI VCDs is estimated over the Pacific (when a multiplicative adjustment is performed,



**Figure 10.** Monthly mean difference of tropospheric residues  $T^*$  from NASA and STREAM for OMI measurements in January (top) and July (bottom) 2005.



**Figure 11.** Monthly mean difference of tropospheric residues  $T^*$  from STSEMAC and STREAM for OMI measurements in January (top) and July (bottom) 2005.

results hardly change). Second, the offset-corrected stratospheric NO<sub>2</sub> VCDs is used for global STS. No additional correction for the tropospheric background is performed, such that the mean TR over the Pacific is 0 by construction.

Daily and monthly maps of TR from STSEMAC are shown in Fig. S24. Daily maps reveal patches of TR from  $-0.3$  CDU up to  $0.4$  CDU resulting from mismatches in actual and modeled stratospheric dynamics. In the monthly mean, these fluctuations largely cancel out. Overall, variability (10th–90th percentiles) of  $T^*$  in remote regions was found to be about  $0.3$ – $0.4$ , similar to that for DOMINO.

Figure 11 displays the differences of the monthly mean TR for January and July 2005. The overall negative values over ocean are a result of the neglect of the tropospheric background in STSEMAC. Besides this, the most striking features are

1. positive deviations (i.e., TR from STSEMAC being higher than from STREAM) over North America and Eurasia in January (up to  $0.45$  CDU, north from  $35^\circ$  N),
2. negative deviations over North America and Eurasia in July (down to  $-0.45$  CDU, north from  $35^\circ$  N), and
3. positive deviations over the Sahara, Middle East, India, and western China in July (up to  $0.38$  CDU).

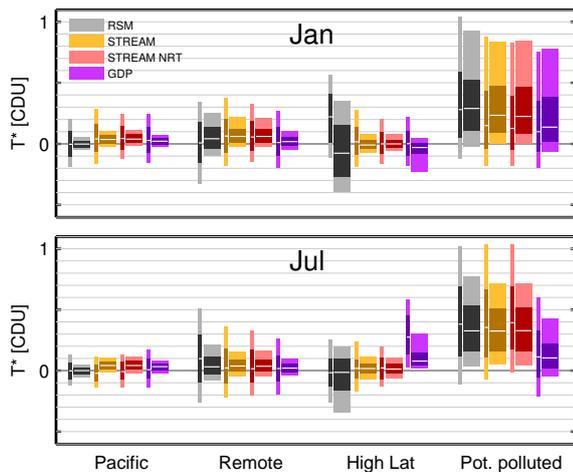
The systematic deviations north from  $35^\circ$  N (1 and 2) are caused by the longitudinal dependency of stratospheric NO<sub>2</sub> from EMAC which differs from the pattern in total column (see Fig. 3). In detail, stratospheric NO<sub>2</sub> over Siberia is quite

low in EMAC, resulting in high-biased TR (similar as for RSM) and indicating that the mean longitudinal dependency of stratospheric NO<sub>2</sub> is not fully reproduced by EMAC. Deviations in July over Sahara and southern Asia (3), however, are at least partly caused by a low bias of  $T^*$  from STREAM as discussed in the previous section.

Overall, deviations are moderate, and STSEMAC still improves the statistics of TR for high latitudes as compared to a simple RSM. It thus might be considered as a simple alternative STS with the advantage that it can be expected to work with the same performance for any satellite instrument, independent of spatiotemporal coverage.

#### 5.1.4 OMI after row anomaly

In 2005, OMI measurements were performed with good instrumental performance, providing daily global coverage. This has changed since summer 2007, when radiance measurements of poor quality regularly occurred at particular cross-track positions (“row anomaly”). We thus also tested STREAM on OMI data after the onset of the row anomaly: Figs. S25 and S26 show  $T^*$  for 2010. While the daily maps reveal gaps due to the exclusion of measurements affected by the row anomaly, the monthly mean patterns as well as the statistical properties are comparable to the results for 2005. The row anomaly thus does not impact the performance of STREAM (or DOMINO or NASA retrievals).



**Figure 12.** Regional statistics of GOME-2 tropospheric residues  $T^*$  from different algorithms for January (top) and July (bottom) 2010. Conventions as in Fig. 6.

## 5.2 GOME-2

STREAM has been applied to GOME-2 (Metop-A) data for the year 2010. The resulting daily and monthly mean maps are shown in Fig. S27. Again, statistical properties are summarized in Fig. 12.

The overall performance of STREAM, i.e., median and variability range of TR, is generally similar to that found of OMI. However, while OMI TRs are about 0.1 CDU over the Pacific, lower values (0.05 CDU) are found for GOME-2. This might be related to differences of cloud statistics due to pixel size, in particular a lower number of fully clouded pixels for GOME-2, as well as differences in local time, cloud products, or systematic spectral interferences caused by clouds in either retrieval algorithm.

On 1 July 2010, GOME-2 is operated in narrow swath mode, causing poor global coverage. This, however, does not affect STREAM performance.

On 15 January 2010, STREAM results in extraordinarily high TR over the ocean (Fig. S28), which turned out to be caused by a solar eclipse (Espenak and Anderson, 2008). Removing the affected orbit results in normal performance for this day. We recommend that screening of solar eclipses be done automatically (as done for OMI) before running any STS algorithm.

### 5.2.1 Comparison to NRT mode

STREAM is foreseen to be implemented in an update of the operational GOME-2 data processor as operated in the framework of the O3M SAF. This requires a slight modification of STREAM in order to work on NRT data.

In STREAM v0.92, the stratospheric fields are estimated for each orbit based on the total column measurements, including seven previous and seven subsequent orbits. In

NRT, however, no subsequent orbits are available. Thus, STREAM has to be operated on the current plus 14 previous orbits instead.

We ran STREAM in NRT mode. The resulting maps are shown in Fig. S28, and the statistics of TR are included in Fig. 12. The deviations between baseline and NRT are marginal. Thus, STREAM can be operated in NRT with stable performance.

### 5.2.2 Comparison to operational product (GDP 4.7)

In the current operational data processor (GDP 4.7), STS for NO<sub>2</sub> is done by an MRS as described in Valks et al. (2011, 2015). Basically, polluted regions (defined by monthly mean TVCDs from the MOZART-2 model being larger than 1 CDU) are masked out. Global stratospheric fields are derived by low-pass filtering in zonal direction by a 30° boxcar filter.

Figure S30 displays daily and monthly mean maps of  $T^*$  in January and July 2010. The respective regional statistics are included in Fig. 12.

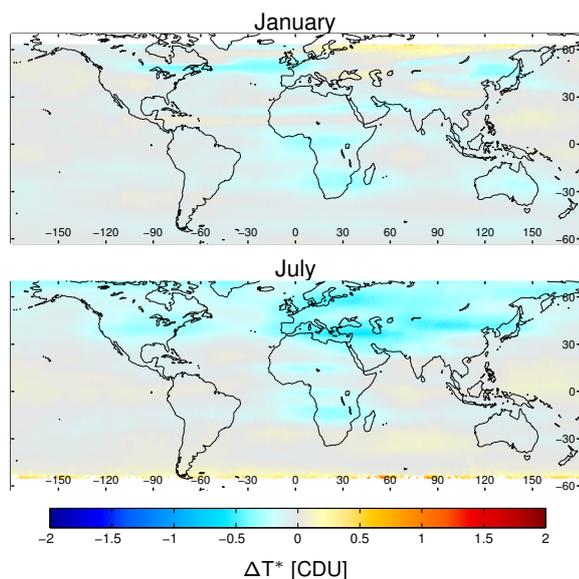
Overall, TRs from GDP are relatively low. Over the Pacific, mean  $T^*$  is close to 0 in January, despite the applied tropospheric background correction of 0.1 CDU. Over potentially polluted regions, median TR from GDP is systematically lower (by 0.2 CDU in July) than from STREAM, and almost a quarter of all TRs are even negative.

Figure 13 displays the differences of the monthly mean TR from GDP 4.7 and STREAM for January and July 2010, again pointing out the systematically lower values of GDP TR over continents in July. The systematic low bias of GDP TR probably results from moderately polluted pixels over regions labeled as “unpolluted”, which still might imply MOZART-2 TVCDs of up to 1 CDU. These measurements cause a high bias of the estimated stratospheric field around polluted regions; by the subsequent low-pass filtering, this high bias is passed over the polluted regions and results in low-biased TR. Further investigations are needed to find out why this effect is stronger in July than in January.

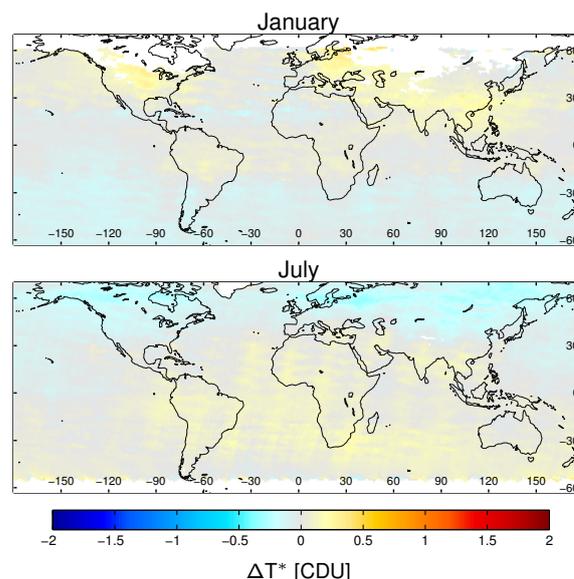
## 5.3 SCIAMACHY

We have applied STREAM to SCIAMACHY VCDs from the MPI-C NO<sub>2</sub> retrieval (Beirle et al., 2010a). The resulting daily and monthly mean maps for 2010 are shown in Fig. S31. Regional statistics are provided in Fig. 14, compared again to the simple RSM and, additionally, to the results of LNM.

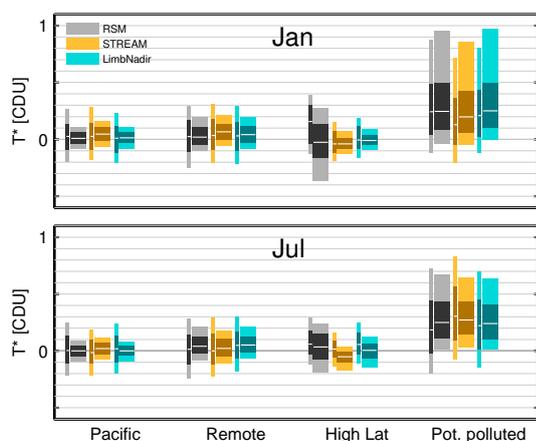
Though SCIAMACHY provides poorer daily spatial coverage, STREAM overall still works well. Again, a clear reduction of the variability of  $T^*$  is found at high latitudes as compared to RSM. Over the Pacific, mean TR from STREAM is higher than for the RSM (= 0) but, similar to GOME-2, not as high as for OMI. Again, this could be related to the low number of cloudy satellite pixels and spec-



**Figure 13.** Monthly mean difference of tropospheric residues  $T^*$  from GDP 4.7 and STREAM for GOME-2 measurements in January (top) and July (bottom) 2010.



**Figure 15.** Monthly mean difference of tropospheric residues  $T^*$  from LNM and STREAM for SCIAMACHY measurements in January (top) and July (bottom) 2010. Gaps at high latitudes in January are caused by respective gaps in the FRESCO cloud product.



**Figure 14.** Regional statistics of SCIAMACHY tropospheric residues  $T^*$  from different algorithms for January (top) and July (bottom) 2010. Conventions as in Fig. 6.

tral interferences, affecting the DOAS analysis, related to clouds. Overall, regional statistics of  $T^*$  are similar to OMI or GOME-2. However, a systematic latitudinal dependency of  $T^*$  remains, showing positive values in hemispheric summer and negative values in hemispheric winter. This results from the latitudinal dependencies of  $V^*$  being different for clouded and cloud-free observations, as shown in Fig. S33, for reasons not yet understood.

### 5.3.1 Comparison to LNM

Within the SCIAMACHY MPIC NO<sub>2</sub> retrieval (Beirle et al., 2010a), STS is performed based on LNM. This can be

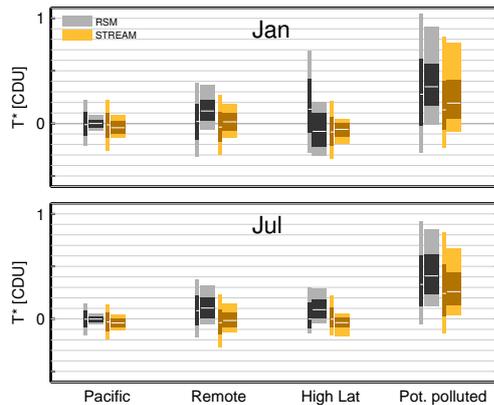
considered as a completely different STS approach, based on actual measurements but not involving CTMs or large-scale interpolation, and thus provide a valuable information on STREAM performance. Figure S32 displays the daily and monthly mean TR from LNM in 2010.

In the LNM STS, the latitude-dependent offset between nadir and limb is determined over the Pacific and corrected for globally; i.e., mean TR in the Pacific is 0 by construction. Overall, regional statistics of  $T^*$  from LNM are very similar to those from STREAM. Figure 15 displays the monthly mean difference of both algorithms. The deviation is dominated by the latitudinal dependency of  $T^*$  from STREAM (see above); when this is removed, both algorithms agree within 0.2 CDU for most parts of the globe.

## 5.4 GOME

GOME was the first instrument of the investigated series of UV/vis spectrometers suited for DOAS analyses of tropospheric trace gases. The comparably small number of GOME pixels and the large across-track footprint (320 km) required a modification of STREAM: we have switched the resolution of the global grid used for weighted convolution from 1 to 5°, i.e., wider than the GOME across-track width at moderate latitudes. Thereby it is ensured that a grid pixel usually contains multiple satellite pixels and that also adjacent grid pixels are not empty (as would be the case for an 1° grid), which is the prerequisite for the calculation of  $w_{TR}$ .

We have applied STREAM to GOME data as provided by TEMIS. The resulting TRs are again compared to the sim-



**Figure 16.** Regional statistics of GOME tropospheric residues  $T^*$  from different algorithms for January (top) and July (bottom) 1999. Conventions as in Fig. 6.

ple RSM. Figure 16 displays the regional TR statistics for GOME in January and July 1999. The respective maps are provided in the Supplement (Fig. S34).

Overall, STREAM yields reasonable results for GOME as well. However, some systematic biases are observed:

- over the Pacific, TRs from STREAM were found to be negative, which can only be explained when the measured columns for cloudy pixels are higher than for cloud-free pixels;
- over potentially polluted regions,  $T^*$  from STREAM is systematically lower than from RSM (by 0.2 CDU in July). This might be a consequence of the applied cloud weight, which has obviously different effects on GOME than on OMI.

This explanation would be consistent with previous findings: while Leue et al. (2001) base the STS on cloudy pixels, Wenig et al. (2004) switched the Heidelberg STS to cloud-free pixels after noticing that GOME columns are higher instead of lower over clouds. Wenig et al. (2004) relate this to the contribution of lightning NO<sub>x</sub>. However, as (a) the impact of lightning NO<sub>x</sub> on satellite observations is generally small (Beirle et al., 2010b) and (b) lightning activity over the remote Pacific used for the RSM is very weak, we rather suspect that a different effect is responsible for this finding, most probably related to the specific instrumental features of GOME (Burrows et al., 1999), in particular the dichroic mirrors causing polarization dependent spectral structures. It might thus be worth re-checking the DOAS analysis of NO<sub>2</sub> for GOME for spectral interferences related to clouds. A second possible effect, which might in particular contribute to the large discrepancy over polluted regions, is that cloud properties are averaged over the large GOME ground pixel; i.e., in an extreme case, low and high cloud layers, which would both be skipped in  $w_{\text{clid}}$  if resolved by the satellite pixel, might yield, on average, an effective cloud height

with a high  $w_{\text{clid}}$ . Any tropospheric pollution within (or directly above) the low cloud layer would then bias high the stratospheric estimate and bias low the TR.

## 5.5 Future instruments

### 5.5.1 TROPOMI

TROPOMI (Veeffkind et al., 2012) on S5p will be launched in 2016. Instrumental setup and spatial coverage are similar to OMI, but TROPOMI will provide a better spatial resolution of  $7 \times 7 \text{ km}^2$  at nadir.

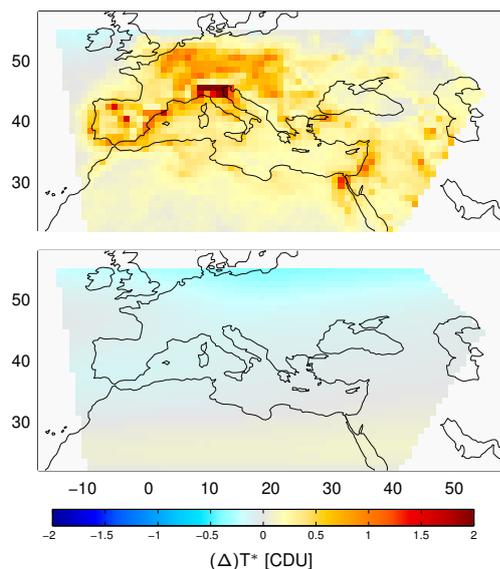
STREAM was developed as a verification algorithm for TROPOMI STS and was tested and compared to the TROPOMI prototype algorithm based on OMI measurements (see above). Though no TROPOMI measurements are available yet, it can be expected that the performance of STREAM on TROPOMI will be even better than for OMI, because, due to the better spatial resolution, more individual satellite pixels are available and among them a higher fraction of clouded pixels. Thus, more sampling points over potentially polluted regions will be available, further decreasing interpolation errors.

### 5.5.2 Sentinel 4 (S4)

The satellite instruments investigated so far are all operated in low, sun-synchronous orbits, providing global coverage at fixed local time. In the near future, a new generation of spectrometers on geostationary orbits will be launched by different space agencies. Over Europe, S4 (Ingmann et al., 2012) will be the first mission providing a spectral resolving UV/vis instrument on a geostationary satellite. The spatial coverage is focussed on Europe. Thus, no “clean” reference regions are regularly available. STREAM might overcome this problem by using clouded observations where the tropospheric pollution is effectively shielded.

We simply evaluate the expected performance of STREAM on S4 measurements by clipping OMI measurements to the area covered by S4 (as given in Courrèges-Lacoste et al., 2010). The STREAM settings are identical to v0.92, except for the a priori removal of the overall latitude dependency in the reference sector, as no Pacific measurements are available for S4. Figure 17 displays the resulting TR (top) and the difference of TR between clipped and global OMI data (bottom) for January 2005.

Though tropospheric pollution over Europe and the Middle East is evident, i.e., an extended clean reference region is actually not available, STREAM is still capable of yielding an accurate stratospheric estimate. Only at the northern and southern borders are systematic biases observed, which can be caused by the overall latitudinal dependency of the stratospheric VCD and border effects of the weighted convolution and can probably be reduced by dedicated optimization of the algorithm for S4.



**Figure 17.** Performance of STREAM on “S4 data” (i.e., OMI measurements clipped to the area covered by S4) for January 2005. The top panel displays the resulting TR, the bottom panel shows the difference to the TR resulting from full OMI data as shown in Fig. 5. The area covered by S4 in winter has been taken from Courrèges-Lacoste et al. (2010).

Situation will probably be improved for real S4 measurements due to the higher number of clouded pixels in S4 compared to OMI. Thus, this first check is highly encouraging to further investigate the applicability of STREAM to S4 and possible improvements.

## 5.6 Advantages and limitations of STREAM

STREAM was successfully applied to various satellite measurements with a wide range of spatial resolution and coverage. STREAM is an MRSM and does not need any model input. It can thus be considered as a complementary approach to data assimilation, as chosen for the TROPOMI prototype algorithm.

As (M)RSMs usually estimate the stratospheric column based on total column measurements over clean regions, they generally miss the (small) tropospheric background of the order of some 0.1 CDU. Several (M)RSMs explicitly correct for this effect based on a priori tropospheric background columns (Martin et al., 2002; Valks et al., 2011; Bucselá et al., 2013). In case of STREAM, however, cloudy pixels, which allow a direct measurement of the actual stratospheric column (except for the small tropospheric column above the cloud), are emphasized. Thus, an additional tropospheric background correction should be unnecessary. Accordingly, in case of OMI, TRs from STREAM are about 0.1 CDU over clean regions, similar as for TRs from DOMINO and NASA. This is close to the a priori value chosen by Valks et al. (2011) but below the values given in Martin et al. (2002)

(about 0.15–0.3 CDU, assuming a tropospheric AMF of 2) and Hilboll et al. (2013) (0.1 up to >0.6 CDU<sup>6</sup>).

In case of other satellite instruments, however, the TR over the Pacific was found to be lower (GOME-2 and SCIAMACHY) or even negative (GOME-1). The latter can only be explained by cloudy measurements being systematically higher than cloud-free measurements. Further investigations are needed to infer this discrepancy between OMI and GOME-1/2/SCIAMACHY and find how it is related to differences in the cloud products and/or the spectral analysis of NO<sub>2</sub>.

STREAM assumes stratospheric NO<sub>2</sub> fields having low zonal variability, in particular at low latitudes. This is reflected by the choice of a wide convolution kernel at the equator. STREAM is thus not capable of resolving diurnal small-scale patterns caused by stratospheric dynamics. These patterns, however, largely cancel out in monthly means.

Whenever actual stratospheric fields do not match the a priori assumption of zonal smoothness, e.g., in case of “tilted” structures or actual large-scale zonal gradients like differences in the stratospheric column over Pacific and Indian ocean, the TR resulting from STREAM can show artificial “stripes”. Further investigations might lead to additional sophisticated algorithm steps to remove these artifacts. However, it has to be taken care that the benefit really outbalances the drawbacks (added complexity) and that no other artifacts/biases are introduced.

The dependencies of TR on STREAM parameter settings have been found to be low ( $\lesssim 0.1$  CDU). The application of STREAM on synthetic data results in deviations to the a priori truth of the same order. These deviations are systematic, i.e., the stratospheric patterns estimated by STREAM are slightly biased high, which can be expected, as they are based on total column measurements, which are always higher than the stratospheric column.

Overall, STREAM uncertainty is well within the general uncertainties of STS (see next section). Note that systematic changes of the NO<sub>2</sub> columns of the same order of 0.1 CDU can also result from changes of the settings for the DOAS analysis, like fit interval, inclusion of additional absorbers in the analysis, or the treatment of rotational and vibrational Raman scattering, creating overall biases as well as spatial patterns, e.g., over oligotrophic oceans (E. Peters, personal communication, 2016).

## 5.7 General uncertainties and challenges of stratosphere–troposphere separation

The uncertainty of STS can often not be directly quantified, as the “true” stratospheric 4-D concentration fields are not known. One approach to assess the STS performance is the

<sup>6</sup>Note, however, that the high values are only reported at higher latitudes in winter, when the ratio  $A_{\text{strat}}/A_{\text{trop}}$  is almost 2 (Fig. S1); thus the large discrepancy is at least partly resolved when the TR is transferred in a TVCD via Eq. (4).

usage of synthetic data, as in Sect. 4.3. In addition, the TR can be used to evaluate the plausibility of the stratospheric estimate and to derive realistic uncertainties:

- Negative TRs are nonphysical. Thus, the occurrence of on average negative  $T^*$  (exceeding the values/frequencies explainable by noise) clearly indicates a positive bias of the estimated stratosphere.
- Tropospheric background columns over regions free of NO<sub>x</sub> sources are expected to have low spatiotemporal variability. Thus, the observed variability of  $T^*$  over clean regions serves as proxy of the uncertainty (precision) of the STS. From different algorithms (MRSMs as well as model-based methods), typical variabilities of  $T^*$  over remote regions are about 0.5 CDU for daily means and about 0.2–0.3 CDU for monthly means. For a simple RSM, much higher values ( $\approx 1$  CDU) are found at high latitudes.

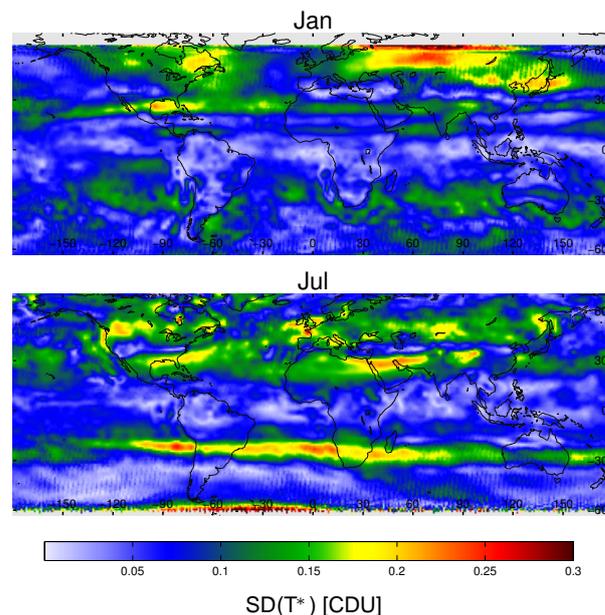
Systematic biases (accuracy) of STS can be estimated from the intercomparison of TRs from different algorithms. Figure 18 displays the standard deviation of monthly mean TR from the algorithms shown in Fig. 8 and discussed in Sect. 5.1, i.e., two different, independent MRSM approaches (STREAM and NASA) as well as two STS based on models, a simple one (STS<sub>EMAC</sub>) and a complex data assimilation setup (DOMINO). Note that the upper range of the color bar was lowered to 0.3 CDU.

Overall, the standard deviation of TR from different STS is low (typically  $<0.1$  CDU and below  $<0.2$  CDU for most parts of the world). It is thus consistent with the uncertainty estimates of stratospheric columns given in literature (Boersma et al. (2011): 0.15–0.25 CDU (SCD); Valks et al. (2011): 0.15–0.3 CDU (VCD); Bucsela et al. (2013): 0.2 CDU (VCD)) and with the magnitude of systematic deviations found in the study on synthetic data (Sect. 4.3).

With respect to the final NO<sub>2</sub> TVCD product, which is higher than TR by the ratio of stratospheric and tropospheric AMFs (Eq. 4), uncertainties of this order are completely negligible over polluted regions such as the US east coast, central Europe, or eastern China. Nonetheless, a regional bias  $>0.2$  CDU (e.g., over Russia in January) can contribute significantly to the relative uncertainty of TVCDs aside the pollution hotspots. Thus, the uncertainty of STS has to be kept in mind in studies focusing on NO<sub>x</sub> emissions from, e.g., biomass burning or soil emissions over regions like Siberia, the Sahel, or Australia.

### 5.8 Other trace gases

STREAM was developed as STS algorithm for NO<sub>2</sub>. However, several other trace gas satellite retrievals face problems which are similar to STS from an algorithmic point of view, i.e., that a small-scale tropospheric signal has to be separated from a smooth background (e.g., caused by stratospheric



**Figure 18.** Standard deviation of monthly mean  $T^*$  from different algorithms (STREAM, DOMINO, NASA, and STS<sub>EMAC</sub>) for January (top) and July (bottom) 2005 (OMI).

columns or, in particular in case of trace gases with low optical depth, shortcomings of the spectral analysis, introducing artificial dependencies on, for example, SZA or ozone columns). Thus, the concept of weighted convolution could be used within the satellite retrievals of, for example, SO<sub>2</sub>, BrO, HCHO, or CHOCHO, with appropriately chosen and optimized weighting factors.

## 6 Conclusions

The separation of the stratospheric and tropospheric column is a key step in the retrieval of NO<sub>2</sub> TVCDs from total column satellite measurements. As coincident direct measurements of the stratospheric column are usually not available (except for SCIAMACHY), current STS algorithms either use CTMs (directly or via data assimilation) or follow a modified reference sector method (MRSM) approach, where the stratospheric columns are basically estimated from total column measurements over clean regions.

We have developed the MRSM STREAM. Weighting factors determine how far individual satellite pixels contribute to the stratospheric estimate. Over potentially polluted regions (according to an NO<sub>2</sub> climatology), weights are lowered, whereas measurements over mid-altitude clouds are assigned with a high weighting factor. Global stratospheric fields are derived by weighted convolution and subtracted from total columns to yield tropospheric residues (TRs). In a second iteration, weighting factors are modified based on the TR: high TR indicates tropospheric pollution, and the respective satel-

lite pixels are assigned with a lower weight. For systematically negative TR, however, weighting factors are increased. The concept of multiplicative weights can easily be extended by additional factors, e.g., based on fire counts in order to explicitly exclude biomass burning events.

STREAM results are robust with respect to variations of the algorithm settings and parameters. With the baseline settings, the errors of STREAM on a synthetic data set have been found to be below 0.1 CDU on average.

STREAM was successfully applied to satellite measurements from GOME 1/2, SCIAMACHY, and OMI. The resulting TRs over clean regions and their variability have been found to be low. However, systematic “stripes” can appear in STREAM TR when the basic assumption that the stratospheric column varies smoothly with longitude is not fulfilled, e.g., in case of “tilted” stratospheric patterns.

The emphasis of clouded observations, which provide a direct measurement of the stratospheric rather than the total column, should supersede an additional correction for the tropospheric background, which successfully worked for OMI but less so for GOME and SCIAMACHY. This might be related to differences in pixel size or local overpass time, both potentially affecting cloud statistics, or differences in the cloud algorithms. However, the detailed reasons are not yet fully understood and require further investigations.

STREAM, which was developed as TROPOMI verification algorithm, was optimized for OMI measurements. Within an O3M SAF visiting scientist project, it was also applied to GOME-2, and STREAM is foreseen to be implemented in an upcoming GDP update.

Results from STREAM were compared to the TROPOMI prototype algorithm, as represented by the DOMINO v2 product, in which STS is implemented by data assimilation. Differences between monthly mean TRs from STREAM and DOMINO are found to be low (almost 0 on average with regional patterns up to about  $\pm 0.1$ – $0.2$  CDU). A comparison to other state-of-the-art STS schemes yields deviations of similar order.

The impact of STS is thus generally negligible for TVCDs over heavily polluted regions. However, the remaining uncertainties still contribute significantly to the total error of TVCDs over moderately polluted regions and have to be kept in mind for emission estimates of area sources of NO<sub>x</sub> such as soil emissions or biomass burning.

## 7 Data availability

STREAM has been tested on NO<sub>2</sub> retrievals from different satellite instruments as listed in Table 1. The input data sets are publicly accessible; the respective links to the data sets are included in the references provided in Table 1.

## Information about the Supplement

Additional images, tables, and text are provided in the Supplement. All references to tables and figures in the Supplement are indicated by a prefix “S”. For readability, the Supplement is structured analogously to the paper; i.e., additional material to Sect. 2.3 can be found in Sect. S2.3 of the Supplement.

**The Supplement related to this article is available online at doi:10.5194/amt-9-2753-2016-supplement.**

*Acknowledgements.* We would like to thank the agencies providing the satellite data: The OMI OMNO<sub>2</sub> product is archived and distributed from the Goddard Earth Sciences Data & Information Services center (NASA). We acknowledge the free use of GOME and OMI (DOMINO) level 2 data from [www.temis.nl](http://www.temis.nl). GOME-2 GDP 4.7 level 2 data is provided by DLR within the framework of EUMETSAT’s O3M SAF. SCIAMACHY level1 data is provided by ESA.

This study was supported by the following projects: Within TROPOMI verification, funding was provided by DLR Bonn under contract 50EE1247 for the development of STREAM and the comparison to the TROPOMI prototype algorithm. Within an O3M SAF visiting scientist project, STREAM was adopted to GOME-2 and compared to the current GDP 4.7 data product. Steffen Beirle acknowledges funding from the FP7 Project Quality Assurance for Essential Climate Variables (QA4ECV), no. 607405.

We acknowledge fruitful discussions with and valuable feedback from many colleagues, in particular during the TROPOMI verification and QA4ECV meetings, especially Andreas Richter, Andreas Hilboll (both IUP Bremen, Germany), Folkert Boersma, and Henk Eskes (both KNMI De Bilt, the Netherlands). Andreas Stohl (NILU, Kjeller, Norway) is acknowledged for discussions on NO<sub>2</sub> transport.

The article processing charges for this open-access publication were covered by the Max Planck Society.

Edited by: B. Veihelmann

## References

- Beirle, S., Platt, U., Wenig, M., and Wagner, T.: Weekly cycle of NO<sub>2</sub> by GOME measurements: a signature of anthropogenic sources, *Atmos. Chem. Phys.*, 3, 2225–2232, doi:10.5194/acp-3-2225-2003, 2003.
- Beirle, S., Spichtinger, N., Stohl, A., Cummins, K. L., Turner, T., Boccippio, D., Cooper, O. R., Wenig, M., Grzegorski, M., Platt, U., and Wagner, T.: Estimating the NO<sub>x</sub> produced by lightning from GOME and NLDN data: a case study in the Gulf of Mexico, *Atmos. Chem. Phys.*, 6, 1075–1089, doi:10.5194/acp-6-1075-2006, 2006.

- Beirle, S., Köhl, S., Pukite, J., and Wagner, T.: Retrieval of tropospheric column densities of NO<sub>2</sub> from combined SCIAMACHY nadir/limb measurements, *Atmos. Meas. Tech.*, 3, 283–299, doi:10.5194/amt-3-283-2010, 2010a.
- Beirle, S., Huntrieser, H., and Wagner, T.: Direct satellite observation of lightning-produced NO<sub>x</sub>, *Atmos. Chem. Phys.*, 10, 10965–10986, doi:10.5194/acp-10-10965-2010, 2010b.
- Beirle, S. and Wagner, T.: Tropospheric vertical column densities of NO<sub>2</sub> from SCIAMACHY, available at: [http://www.sciamachy.org/products/NO2/NO2tc\\_v1\\_0\\_MPI\\_AD.pdf](http://www.sciamachy.org/products/NO2/NO2tc_v1_0_MPI_AD.pdf) (last access: 17 June 2016), 2012 (data available at: <http://www.sciamachy.org/products/index.php?species=NO2>).
- Belmonte Rivas, M., Veeffkind, P., Boersma, F., Levelt, P., Eskes, H., and Gille, J.: Intercomparison of daytime stratospheric NO<sub>2</sub> satellite retrievals and model simulations, *Atmos. Meas. Tech.*, 7, 2203–2225, doi:10.5194/amt-7-2203-2014, 2014.
- Boersma, K., Eskes, H., and Brinksma, E.: Error analysis for tropospheric NO<sub>2</sub> retrieval from space, *J. Geophys. Res.-Atmos.*, 109, D04311, doi:10.1029/2003JD003962, 2004.
- Boersma, K. and Eskes, H.: Assimilated SCIAMACHY NO<sub>2</sub> fields: HDF data file user manual, available at: [http://www.temis.nl/airpollution/no2col/papers/no2col\\_hdf\\_docu.pdf](http://www.temis.nl/airpollution/no2col/papers/no2col_hdf_docu.pdf) (last access: 17 June 2016), 2004 (data available at: <http://www.temis.nl/airpollution/no2col/no2colgome.php>).
- Boersma, K. F., Eskes, H. J., Veeffkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO<sub>2</sub> from OMI, *Atmos. Chem. Phys.*, 7, 2103–2118, doi:10.5194/acp-7-2103-2007, 2007.
- Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veeffkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO<sub>2</sub> column retrieval algorithm for the Ozone Monitoring Instrument, *Atmos. Meas. Tech.*, 4, 1905–1928, doi:10.5194/amt-4-1905-2011, 2011 (data available at: [http://www.temis.nl/airpollution/no2col/no2regioomi\\_v2.php](http://www.temis.nl/airpollution/no2col/no2regioomi_v2.php)).
- Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V., Chance, K., and Goede, A.: SCIAMACHY: Mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 127–150, 1999.
- Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veeffkind, J. P., Boersma, K. F., and Brinksma, E. J.: Algorithm for NO<sub>2</sub> vertical column retrieval from the ozone monitoring instrument, *IEEE T. Geosci. Remote*, 44(5), 1245–1258, doi:10.1109/TGRS.2005.863715, 2006.
- Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veeffkind, J. P., Gleason, J. F., and Pickering, K. E.: A new stratospheric and tropospheric NO<sub>2</sub> retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, *Atmos. Meas. Tech.*, 6, 2607–2626, doi:10.5194/amt-6-2607-2013, 2013 (data available at: [http://mirador.gsfc.nasa.gov/collections/OMNO2\\_003.shtml](http://mirador.gsfc.nasa.gov/collections/OMNO2_003.shtml)).
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K. U., and Eisinger, M.: The global ozone monitoring experiment (GOME): Mission concept and first scientific results, *J. Atmos. Sci.*, 56, 151–175, doi:10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2, 1999.
- Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 – Metop’s second-generation sensor for operational ozone monitoring, *ESA Bull.-Eur. Space Agency*, 102, 28–36, 2000.
- Courrèges-Lacoste, G. B., Ahlers, B., Guldimann, B., Short, A., Veihelmann, B., Stark, H.: The Sentinel-4/UVN instrument on-board MTG-S, available at: [https://www.eumetsat.int/cs/idcplg?IdcService=GET\\_FILE&dDocName=pdf\\_conf\\_p59\\_s1\\_01\\_bazalget\\_p&allowInterrupt=1&noSaveAs=1&RevisionSelectionMethod=LatestReleased](https://www.eumetsat.int/cs/idcplg?IdcService=GET_FILE&dDocName=pdf_conf_p59_s1_01_bazalget_p&allowInterrupt=1&noSaveAs=1&RevisionSelectionMethod=LatestReleased) (last access: 17 June 2016), 2010.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Dirksen, R. J., Boersma, K. F., Eskes, H. J., Ionov, D. V., Bucsela, E. J., Levelt, P. F., and Kelder, H. M.: Evaluation of stratospheric NO<sub>2</sub> retrieved from the Ozone Monitoring Instrument: Intercomparison, diurnal cycle, and trending, *J. Geophys. Res.-Atmos.*, 116, D08305, doi:10.1029/2010JD014943, 2011.
- Ebojje, F., von Savigny, C., Ladstätter-Weißmayer, A., Rozanov, A., Weber, M., Eichmann, K.-U., Bötzel, S., Rahpoe, N., Bovensmann, H., and Burrows, J. P.: Tropospheric column amount of ozone retrieved from SCIAMACHY limb-nadir-matching observations, *Atmos. Meas. Tech.*, 7, 2073–2096, doi:10.5194/amt-7-2073-2014, 2014.
- Eskes, H. J., Velthoven, P. F. J. V., Valks, P. J. M., and Kelder, H. M.: Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model, *Q. J. Roy. Meteor. Soc.*, 129, 1663–1681, doi:10.1256/qj.02.14, 2003.
- Espenak, F., and Anderson, J.: Annular and Total Solar Eclipses of 2010, NASA/TP–2008–214171, available at: <http://eclipse.gsfc.nasa.gov/SEpubs/2010/TP214171a.pdf> (last access: 17 June 2016), 2008.
- Hilboll, A., Richter, A., Rozanov, A., Hodnebrog, Ø., Heckel, A., Solberg, S., Stordal, F., and Burrows, J. P.: Improvements to the retrieval of tropospheric NO<sub>2</sub> from satellite – stratospheric correction using SCIAMACHY limb/nadir matching and comparison to Oslo CTM2 simulations, *Atmos. Meas. Tech.*, 6, 565–584, doi:10.5194/amt-6-565-2013, 2013.
- Ingmann, P., B. Veihelmann, A.G. Straume, Y. Meijer, The status of implementation of the atmospheric composition related GMES missions Sentinel-4/Sentinel-5 and Sentinel-5p, Proceedings of the 2012 EUMETSAT Meteorological Satellite Conference, Sopot, Poland, 3–7 September 2012.
- Jöckel, P., Tost, H., Pozzer, A., Brühl, C., Buchholz, J., Ganzeveld, L., Hoor, P., Kerkweg, A., Lawrence, M. G., Sander, R., Steil, B., Stiller, G., Tanarhte, M., Taraborrelli, D., van Aardenne, J., and Lelieveld, J.: The atmospheric chemistry general circulation model ECHAM5/MESy1: consistent simulation of ozone from the surface to the mesosphere, *Atmos. Chem. Phys.*, 6, 5067–5104, doi:10.5194/acp-6-5067-2006, 2006.

- Jöckel, P., Kerkweg, A., Pozzer, A., Sander, R., Tost, H., Riede, H., Baumgaertner, A., Gromov, S., and Kern, B.: Development cycle 2 of the Modular Earth Submodel System (MESSy2), *Geosci. Model Dev.*, 3, 717–752, doi:10.5194/gmd-3-717-2010, 2010.
- Jöckel, P., Tost, H., Pozzer, A., Kunze, M., Kirner, O., Brenninkmeijer, C. A. M., Brinkop, S., Cai, D. S., Dyroff, C., Eckstein, J., Frank, F., Garny, H., Gottschaldt, K.-D., Graf, P., Grewe, V., Kerkweg, A., Kern, B., Matthes, S., Mertens, M., Meul, S., Neu-maier, M., Nützel, M., Oberländer-Hayn, S., Ruhnke, R., Runde, T., Sander, R., Scharffe, D., and Zahn, A.: Earth System Chem-istry integrated Modelling (ESCI-Mo) with the Modular Earth Submodel System (MESSy) version 2.51, *Geosci. Model Dev.*, 9, 1153–1200, doi:10.5194/gmd-9-1153-2016, 2016.
- Knutsson, H., and Westin, C.-F.: Normalized and differential convo-lution, in Proceedings of Computer Society Conference on Com-puter Vision and Pattern Recognition (CVPR), IEEE, 515–523, doi:10.1109/CVPR.1993.341081, 1993.
- Kühl, S., J. Pukite, T. Deutschmann, U. Platt, and T. Wagner: SCIA-MACHY limb measurements of NO<sub>2</sub>, BrO and OClO. Retrieval of vertical profiles: Algorithm, first results, sensitivity and com-parison studies, *Adv. Space Res.*, 42, 1747–1764, 2008.
- Leue, C., Wenig, M., Wagner, T., Klimm, O., Platt, U., and Jähne, B.: Quantitative analysis of NO<sub>x</sub> emissions from Global Ozone Monitoring Experiment satellite image sequences, *J. Geophys. Res.*, 106, 5493–5505, 2001.
- Levelt, P., Van den Oord, G., Dobber, M., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., and Saari, H.: The Ozone Monitoring Instrument, *IEEE T. Geosci. Remote*, 44, 1093–1101, doi:10.1109/TGRS.2006.872333, 2006.
- Marchenko, S., Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., and Bucsel, E. J.: Revising the slant column den-sity retrieval of nitrogen dioxide observed by the Ozone Mon-itoring Instrument, *J. Geophys. Res.-Atmos.*, 120, 5670–5692, doi:10.1002/2014JD022913, 2015.
- Martin, R., Chance, K., Jacob, D., Kurosu, T., Spurr, R., Bucsel, E., Gleason, J., Palmer, P., Bey, I., Fiore, A., Li, Q., Yantosca, R., and Koelemeijer, R.: An improved retrieval of tropospheric nitro-gen dioxide from GOME, *J. Geophys. Res.-Atmos.*, 107, 4437, doi:10.1029/2001JD001027, 2002.
- Martin, R. V.: Satellite remote sensing of surface air quality, *Atmos. Environ.*, 42, 7823–7843, 2008.
- Monks, P. S. and Beirle, S.: Applications of Satellite Observations of Tropospheric Composition, in *The Remote Sensing of Tropo-spheric Composition from Space*, 365–449, Springer-Verlag Berlin Heidelberg, 2011.
- Platt, U. and Stutz, J.: *Differential Optical Absorption Spec-troscopy*, Springer-Verlag Berlin Heidelberg, ISBN-13: 978-3-540-21193-8, 2008.
- Richter, A. and Burrows, J. P.: Tropospheric NO<sub>2</sub> from GOME Measurements, *Adv. Space Res.*, 29, 1673–1683, 2002.
- Richter, A., Burrows, J. P., Nusz, H., Granier, C., and Niemeier, U.: Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 437, 129–132, doi:10.1038/nature04092, 2005.
- Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornbluh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of simulated climate to horizontal and vertical reso-lution in the ECHAM5 atmosphere model, *J. Clim.*, 19, 3771–3791, doi:10.1175/JCLI3824.1, 2006.
- Sierk, B., Richter, A., Rozanov, A., von Savigny, Ch., Schmolt-ner, A. M., Buchwitz, M., Bovensmann, H., and Burrows, J. P.: Retrieval and monitoring of atmospheric trace gas con-centrations in nadir and limb geometry using the space-borne SCIAMACHY instrument, *Environ. Monit. Assess.*, 120, 65–77, doi:10.1007/s10661-005-9049-9, 2006.
- Sioris, C. E., Kurosu, T. P., Martin, R. V., and Chance, K.: Strato-spheric and tropospheric NO<sub>2</sub> observed by SCIAMACHY: first results, *Adv. Space Res.*, 34, 780–785, 2004.
- Stohl, A., Huntrieser, H., Richter, A., Beirle, S., Cooper, O. R., Eckhardt, S., Forster, C., James, P., Spichtinger, N., Wenig, M., Wagner, T., Burrows, J. P., and Platt, U.: Rapid intercontinental air pollution transport associated with a meteorological bomb, *Atmos. Chem. Phys.*, 3, 969–985, doi:10.5194/acp-3-969-2003, 2003.
- Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M., and Emmadi, S.: Operational total and tropospheric NO<sub>2</sub> column retrieval for GOME-2, *Atmos. Meas. Tech.*, 4, 1491–1514, doi:10.5194/amt-4-1491-2011, 2011.
- Valks, P., Loyola, D., Hao, N., Hedelt, P., Slijkhuys, S., and Grossi, M.: Algorithm Theoretical Basis Document for GOME-2 Total Column Products of Ozone, NO<sub>2</sub>, BrO, SO<sub>2</sub>, H<sub>2</sub>O, HCHO, and Cloud Properties, DLR/GOME-2/ATBD/01, Issue 3/A, available at: <https://atmos.eoc.dlr.de/gome2/documentation.html> (last access: 17 June 2016), 2015 (data available at: <http://atmos.eoc.dlr.de/gome/product.html>).
- van Geffen, J. H. G. M., Boersma K. F., Eskes, H. J., Maasackers, J. D., and Veefkind, J.P.: TROPOMI ATBD of the total and tropo-spheric NO<sub>2</sub> data products, Reports S5P-KNMI-L2-0005-RP version 0.11.0, 2 October 2014, KNMI, De Bilt, the Netherlands, 2014.
- van Geffen, J. H. G. M., Boersma, K. F., Van Roozendael, M., Hen-drick, F., Mahieu, E., De Smedt, I., Sneep, M., and Veefkind, J. P.: Improved spectral fitting of nitrogen dioxide from OMI in the 405–465 nm window, *Atmos. Meas. Tech.*, 8, 1685–1699, doi:10.5194/amt-8-1685-2015, 2015.
- Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sens. Environ.*, 120, 70–83, doi:10.1016/j.rse.2011.09.027, 2012.
- Wenig, M., Kühl, S., Beirle, S., Bucsel, E., Jähne, B., Platt, U., Gleason, J., and Wagner, T.: Retrieval and analysis of strato-spheric NO<sub>2</sub> from the Global Ozone Monitoring Experiment, *J. Geophys. Res.*, 109, D04315, doi:10.1029/2003JD003652, 2004.
- Zien, A. W., Richter, A., Hilboll, A., Blechschmidt, A.-M., and Burrows, J. P.: Systematic analysis of tropospheric NO<sub>2</sub> long-range transport events detected in GOME-2 satellite data, *Atmos. Chem. Phys.*, 14, 7367–7396, doi:10.5194/acp-14-7367-2014, 2014.