• Original Paper •

Optimization of the OCO-2 Cloud Screening Algorithm and Evaluation against MODIS and TCCON Measurements over Land Surfaces in Europe and Japan

Sijie CHEN[1], Shuaibo WANG[1], Lin SU[2], Changzhe DONG[3], Ju KE[1], Zhuofan ZHENG[1],
Chonghui CHENG[1], Bowen TONG[1], and Dong LIU[*1]

[1]*State Key Laboratory of Modern Optical Instrumentation*, *College of Optical Science and Engineering*, *Zhejiang University*, *Hangzhou, Zhejiang* 310027, *China*

[2]*Aerospace Information Research Institute*, *Chinese Academy of Sciences*, *Beijing* 100094, *China*

[3]*Shanghai Institute of Satellite Engineering*, *Shanghai*, 201109, *China*

ABSTRACT

A method to tighten the cloud screening thresholds based on local conditions is used to provide more stringent schemes for Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms. Cloud screening strategies are essential to remove scenes with significant cloud and/or aerosol contamination from OCO-2 observations, which helps to save on the data processing cost and ensure high quality retrievals of the column-averaged CO_2 dry air mole fraction (XCO_2). Based on the radiance measurements in the 0.76 μm O_2A band, 1.61 μm (weak), and 2.06 μm (strong) CO_2 bands, the current combination of the A-Band Preprocessor (ABP) algorithm and Iterative Maximum A Posteriori (IMAP) Differential Optical Absorption Spectroscopy (DOAS) Preprocessor (IDP) algorithm passes around 20%–25% of all soundings, which means that some contaminated scenes also pass the screening process. In this work, three independent pairs of threshold parameters used in the ABP and IDP algorithms are sufficiently tuned until the overall pass rate is close to the monthly clear-sky fraction from the MODIS cloud mask. The tightened thresholds are applied to observations over land surfaces in Europe and Japan in 2016. The results show improvement of agreement and positive predictive value compared to the collocated MODIS cloud mask, especially in summer and fall. In addition, analysis indicates that XCO_2 retrievals with more stringent thresholds are in closer agreement with measurements from collocated Total Carbon Column Observing Network (TCCON) sites.

Key words: cloud screening, CO_2 retrieval, OCO-2, MODIS, TCCON

Article Highlights:

• A method to tighten the cloud screening thresholds based on local conditions is used to provide more stringent schemes for OCO-2 cloud screening algorithms.
• The optimized scheme reduces the difference between TCCON XCO retrievals and OCO-2 measurements from 3.23 ± 2.25 ppm to 2.11 ± 1.76 ppm.
• Adjustment is applied according to average monthly clear-sky fractions, which helps incorporate seasonal variation in Europe and Japan.

---

1. Introduction

The Orbiting Carbon Observatory-2 (OCO-2) satellite provides quantification of global sources and sinks of atmospheric carbon dioxide (CO_2) based on high spatial and temporal sampling of reflected solar radiation (Crisp et al., 2004). The OCO-2 mission helps to solve the carbon budgets on sub-continental or regional spatial scales by delivering space-based column-averaged CO_2 dry air mole fraction (XCO_2) observations over land and ocean, which cannot be solved solely based on networks of in-situ CO_2 sensors (Gurney et al., 2002; Chevallier et al., 2007; Baker

---

* Corresponding author: Dong LIU
　Email: liudongopt@zju.edu.cn

et al., 2010). However, the utility of remotely sensed CO_2 concentrations from space has stringent requirements on the precision [in the range of 1–10 ppm (0.3%–3.0%) and without significant biases] (Rayner and O'Brien, 2001), which became the key consideration for developing new instruments and algorithms.

In the early stage of space-borne CO_2 observation, CO_2 measurements were retrieved using a thermal-infrared (IR) algorithm from instruments such as the High Resolution Infrared Sounder 2 (HIRS-2) aboard the NOAA 10 satellite, the Atmospheric Infrared Sounder (AIRS) on the Aqua satellite, and the Infrared Atmospheric Sounding Interferometer (IASI) on the Meteorological Operational Satellite (METOP) (Crevoisier et al., 2004, 2009; Engelen and McNally, 2005). These measurements contribute to the estimation of global CO_2 sources and sinks, but lack sensitivity to near-surface CO_2 due to the usage of thermal-IR data (Chevallier et al., 2005a, b). Shortwave infrared (SWIR) observations made by the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) instrument on the Environmental Satellite (ENVISAT), on the other hand, can retrieve CO_2 measurements with high near-surface sensitivity (Burrows et al., 1995; Bovensmann et al., 1999). However, it is also noted that there are systematic errors in SCIAMACHY-observed CO_2 due to aerosols (Houweling et al., 2005). Therefore, recent instruments have used combined spectral bands to improve the overall performance. For example, the Fourier transform spectrometer (FTS) operating on the Greenhouse-gases Observing Satellite (GOSAT) has three SWIR bands and one thermal-IR band; it also uses a cloud and aerosol imager to constrain scattering interferences (Kuze et al., 2009; Yoshida et al., 2011).

The original OCO instrument was designed to measure the SWIR absorption bands of CO_2 at 1.61 and 2.06 μm, as well as the O_2A-band at 0.765 μm. After the launch failure in 2009, a re-flight mission was initiated using the same instrument design, which became known as the OCO-2. The algorithm developed for the retrieval of XCO_2 from OCO, OCO-2 and new OCO-3 observations has been continuously refined, making it one of the most advanced and widely used algorithms for CO_2 study (Connor et al., 2008; O'Dell et al., 2012; Taylor et al., 2016). The algorithm can adapt to analyze observations from other instruments, including SCIAMACHY, GOSAT, and the ground-based FTS (Bösch et al., 2006). Therefore, further optimization and evaluation of the algorithm can benefit the refinement of current retrievals, and also help provide more information for testing and validation.

Collecting approximately 1 million soundings on a daily basis, reliable cloud screening algorithms are essential for both computational efficiency and quality assurance. Currently, the level 2 retrieval algorithm of OCO-2 utilizes a combination of results from two independent cloud screening methods—the A-Band Preprocessor (ABP) method and the Iterative Maximum A Posteriori (IMAP) Differential Optical Absorption Spectroscopy (DOAS) Preprocessor (IDP) method (Mandrake et al., 2013).

The ABP algorithm, developed by Colorado State University, is used for rapid screening of OCO-2 data for cloud or aerosol contamination. The ABP algorithm uses solar radiance spectral data in the 0.76 μm O_2A band and ECMWF model. It employs a fast Bayesian retrieval to estimate surface pressure and surface albedo, and minimizes the residual of a simulated and measured spectrum to get a goodness-of-fit (χ^2) statistic. The simulated spectra are calculated assuming clear-sky conditions with only molecular Rayleigh scattering present. The χ^2 of the residuals between the measured and simulated spectra is calculated as a cost function using Gauss–Newton iteration and five parameters: surface pressure, the offset to an assumed temperature profile, surface albedos at the band beginning and end points, and a wavelength (dispersion) multiplier (O'Dell and Taylor, 2014). Scenes that violate the condition, i.e., having clouds or aerosols, will generally yield large spectral residuals and thus can be identified easily.

The IDP algorithm, on the other hand, uses solar spectra to retrieve the vertical column density (VCD) of the gas molecules (CO_2 and H_2O) in the 1.61 μm (weak) and 2.06 μm (strong) CO_2 bands (Frankenberg et al., 2005). The ratio of the VCD between the weak band and the strong band should approach unity in the absence of clouds or aerosols, which in reality is centered around 0.99 due to imperfect spectroscopy (Mandrake et al., 2013). Scenes that apparently diverge from unity are affected by clouds or aerosols, which have the ability to significantly alter the distribution of the light paths, thus biasing the retrieved column properties, surface reflectance and other parameters by up to a few percent (Rayner and O'Brien, 2001).

As preprocessors, both methods aim to efficiently identify scenes, or soundings, containing significant amounts of clouds and/or aerosols. However, some contaminated scenes might pass the filter due to the loosely set thresholds (Taylor et al., 2016). Cloud or aerosol contamination can lead to failure in accurately retrieving XCO_2 and, consequently, influence the identification of regional sources and sinks (Miller et al., 2007; Connor et al., 2016; Wunch et al., 2017). If the measured radiances are significantly impacted by scattering due to clouds and aerosols, XCO_2 retrievals are unlikely to converge; even if the impacts are rather modest, i.e., with a total optical depth (TOD) ≤ 0.3, they can still introduce biases in the XCO_2 retrievals (Butz et al., 2011; O'Dell et al., 2012; Guerlet et al., 2013). Previous studies suggest that both the ABP and IDP methods are reasonably effective at identifying high clouds (O'Dell et al., 2012; Taylor et al., 2012, 2016). In addition, the combination of the two methods could reduce the misidentification of low clouds by one third (O'Dell et al., 2012; Taylor et al., 2012, 2016). However, comparison with collocated profiles measured by the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) shows that more than 40% of the low clouds with TOD > 1 at 532 nm might pass both screeners (Taylor et al., 2016).

In this work, a method to tighten the cloud screening thresholds based on local conditions is used to provide more stringent schemes for OCO-2 cloud screening algorithms. The method aims to provide fast adaptation to regional conditions, reducing the inflation in the operational algorithms caused by loose thresholds setting and unified application over the globe. The same inputs as the ABP and IDP algorithms are utilized (listed in the data section) to provide an independent set of cloud detection results. One year of measurements in 2016 (from 1 December 2015 to 30 November 2016) collected over Europe are used for the study. OCO-2 cloud screening results are compared directly to the cloud mask from collocated Moderate Resolution Imaging Spectrometer (MODIS) products. The threshold parameters currently used by the OCO-2 algorithm are tuned until the overall passing rate is decreased from 20%–25% of all soundings to a close match of the clear-sky fraction observed by MODIS. By design, the narrowed range of thresholds increases the agreement with MODIS cloud screening results, thereby increasing the quality assurance of XCO_2 retrievals, at a cost of a certain decrease in the number of soundings passing the screeners.

This paper is organized as follows. Section 2 provides a brief background on the cloud screening algorithm of OCO-2, and a list of datasets from different satellite sensors and ground-based sites used in this work. Section 3 describes the procedure for comparing collocated OCO-2 and MODIS measurements, optimizing the threshold parameters, and adapting to monthly variations. Section 4 presents the results and analysis of re-screened OCO-2 measurements, including comparison with the MODIS cloud mask, as well as XCO_2 retrievals from Total Carbon Column Observing Network (TCCON) sites. The optimization scheme is also confirmed with one year of data in Japan. Section 5 concludes the paper.

2. Sensors and data

Figure 1 shows a conceptual diagram of this work. The flowchart is detailed in the following sections. In summary, a one-year period of OCO-2 data from December 2015 to November 2016 is used in this study. The dataset is generally referred to as a 2016 dataset, as the December from the previous year is only included for a complete winter season. The data cover a selected area in Europe, consisting of land within 10°–50°W longitude and 30°–60°N latitude. The area was chosen for its relatively high density of TCCON sites, making it more convenient for collocated comparison. Observations over water are not considered in this work, since validation against measurements from TCCON sites would be difficult. We combined observations made in Nadir mode and Glint mode because they use the same thresholds over land. Focused analysis is made based on two monthly datasets, in January and July, which lead to configuration of trends between threshold parameters and cloud screening results, further discussed in section 3.

2.1. OCO-2

The OCO-2 satellite carries a single instrument incorporating three co-bore sighted, long-slit imaging grating spectro-
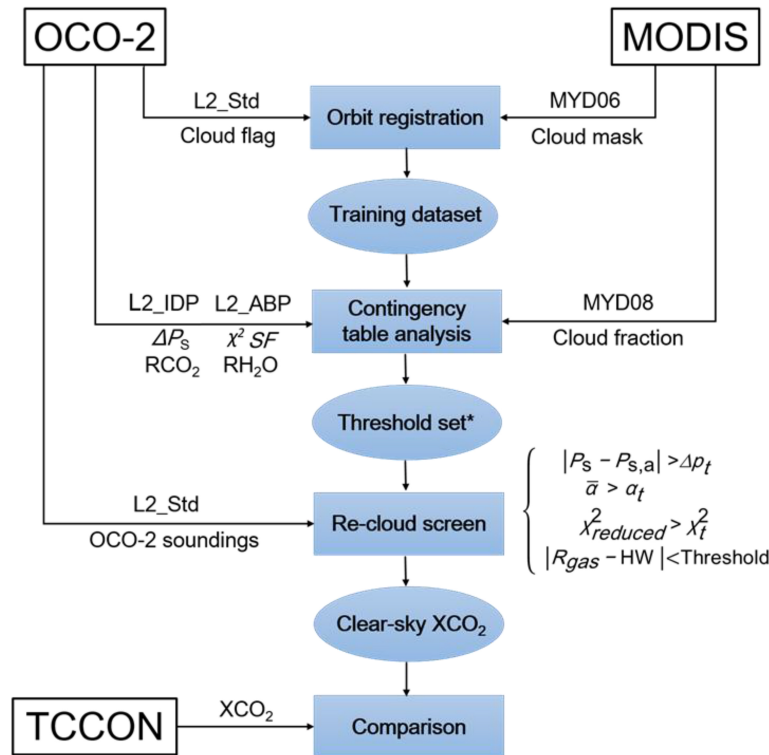

Fig. 1. Conceptual diagram of the proposed optimization scheme of the OCO-2 cloud screening algorithm.

meters optimized for the O_2A band at 0.765 μm and the CO_2 bands at 1.61 and 2.06 μm (Crisp et al., 2004). Flying in a 705-km sun-synchronous polar orbit at a repeat cycle of 16 days, the footprint of OCO-2 has a resolution of about 1.3 km across-track and 2.3 km along-track. OCO-2 is a member of the A-Train constellation, which allows it to collect synergistic measurements with other members in close proximity.

The OCO-2 instrument functions on the daylight side of the orbit, operating in either Nadir, Glint or Target mode. In Nadir mode, the instrument provides the highest spatial resolution by looking straight down to Earth and collecting data along the ground track. In Glint mode, the instrument is pointed toward the bright glint spot where solar radiation is specularly reflected off the surface. The primary purpose of Glint mode is to provide higher SNR over the ocean, which could be 100 times higher than observations in Nadir mode at high latitudes. The Target mode is mainly used for calibration over specific ground sites, and therefore it is not discussed further in this work.

The OCO-2 products used here include Level 2 spatially ordered geolocated retrievals screened using the A-band Preprocessor (OCO2_L2_ABand), Level 2 spatially ordered geolocated retrievals screened using the IMAP-DOAS Preprocessor (OCO2_L2_IMAPDOAS), and Level 2 geolocated XCO2 retrievals results (OCO2_L2_Standard) (Gunson and Eldering, 2014a, b, c). All the products are from Version 8, retrospective processing (8r). The data from OCO-2 are available at https://daac.gsfc.nasa.gov/.

A brief introduction is given below for the specific dataset utilized in the work.

2.1.1. *OCO2_L2_ABand*

From A-band products, primary screening parameters such as surface pressure, surface albedo, and reduced χ^2 values near 0.765 μm are used for comparison with collocated MODIS products and optimization of threshold parameters. The surface pressure parameter gives the difference between the surface pressure estimated by the ECMWF and that retrieved from satellite observation. It is calculated as

$$\Delta p_s = p_s - p_{s,a} , \qquad (1)$$

where s indicates the surface and a indicates a model priori value. More specifically, the ECMWF estimate is a linear interpolation in time and space of modeled surface pressure with a 0.25° spatial and 3-h temporal resolution. Correction for an offset from path length dependence due to imperfect spectroscopy used in the retrieval algorithm is also taken into account. The surface albedo parameter is the average of retrieved surface albedos at 0.755 μm and 0.785 μm. The reduced χ^2 value is a goodness-of-fit parameter of the fast retrievals. The results of filtering with these three parameters are summarized into a cloud flag product.

2.1.2. *OCO2_L2_IMAPDOAS*

From this dataset, independently retrieved VCDs of CO_2 and H_2O in the strong CO_2 band and weak CO_2 band are used. The ratios of measured values in the weak and strong band are the primary parameters to categorize scenes as cloudy or clear.

2.1.3. *OCO2_L2_Standard*

The retrieved XCO_2 and associated uncertainties are used for comparison with measurements at TCCON stations.

2.2. **MODIS**

The MODIS instrument provides calibrated radiances in 36 spectral bands ranging in wavelength from 0.4 μm to 14.4 μm, which are used to infer many key properties of clouds and aerosols (Kaufman et al., 2002; Minnis et al., 2008). The instrument aboard the Aqua satellite, also a member of the A-Train, provides collocated measurements with OCO-2. The MODIS products used include MYD03, MYD06_L2, and MYD08. MYD03 provides latitude and longitude at a 1 km resolution for collocating with OCO-2. MYD06 provides the cloud mask and other cloud properties used for analysis. MYD08 provides the monthly averaged cloud fraction at a 0.5° × 0.5° resolution. It is noted that the comparisons in this study are in reference to MODIS as truth. This assumption could be affected by the uncertainty of MODIS cloud screening products, which is a function of instrument noise in the channels and the magnitude of the correction that is necessary due to surface spectral radiative properties, as well as atmospheric moisture and/or aerosol reflection contributions (Minnis et al., 2008). The data from MODIS are available at https://ladsweb.modaps.eosdis.nasa.gov/.

2.3. **TCCON**

TCCON is a network of ground-based FTSs recording near-IR direct solar spectra (Wunch et al., 2011, 2017). TCCON data are widely used as the most accurate and precise retrieval of column-averaged abundance of CO_2, CH_4, H_2O and other trace gases, providing a validation resource for the OCO, SCIAMACHY, and GOSAT projects (Morino et al., 2011; Reuter et al., 2011; Thompson et al., 2012).

Data from six TCCON sites in Europe, including Bialystok (53.23°N, 23.025°E), Bremen (53.10°N, 8.85°E), Garmisch (47.48°N, 11.06°E), Karlsruhe (49.10°N, 8.44°E), Orleans (47.97°N, 2.113°E), and Paris (48.846°N, 2.356°E), are utilized in this work. In 2016, these sites collected 68 590 measurements in total, covering 324 days in the year. Most of these measurements were made in summer (June–July–August), while fewest measurements were made during winter (December–January–February), likely due to the high cloud fraction, surface snow cover, and limited sunlight hours.

To compare original and re-screened OCO-2 XCO_2 retrievals with TCCON records, temporal and spatial averaging is necessary (Fig. 2). The average of OCO-2 XCO_2 retrievals within a 200 km radius from the specific site was compared to the daily average of XCO_2 measured by that
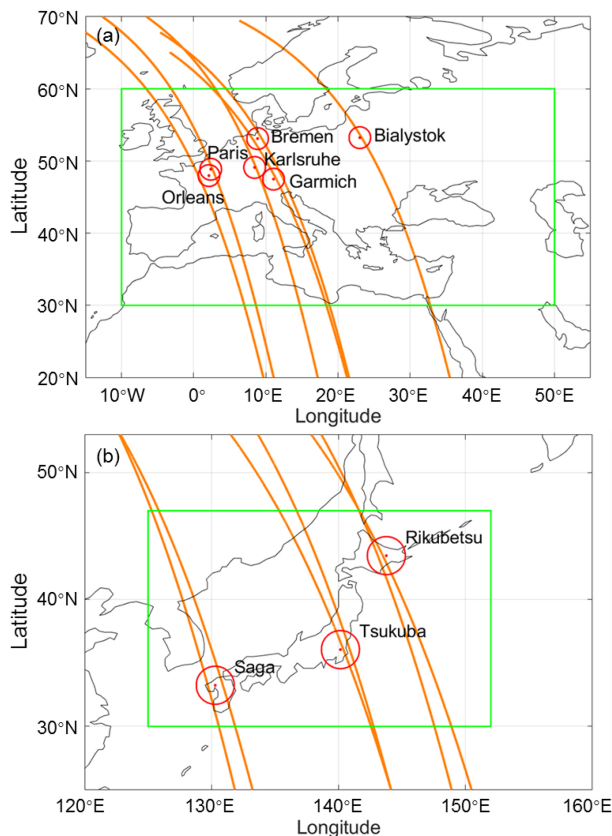
Fig. 2. Selected area in Europe and Japan (land inside green box) for comparison between OCO-2 and the MODIS cloud mask, as well as collocating ranges (red circles) for OCO-2 and TCCON XCO_2 retrievals. The orange lines show examples of OCO-2 orbits passing over nearby TCCON sites.

TCCON site. The range is chosen to provide sufficient measurements for comparison, yet can be assumed to have relatively constant XCO_2 within the range. A similar range was applied in Liang et al. (2017). In total, 143 comparison pairs were made, covering 111 days in the one-year period.

In addition, three sites in Japan—Rikubetsu (43.46°N, 143.77°E), Saga (33.24°N, 130.29°E), and Tsukuba (36.05°N, 140.12°E)—are used for validation of the optimized cloud screening scheme.

3. Method

3.1. *OCO-2 cloud screening method*

The ABP and IDP methods of OCO-2 provide two sets of independent cloud screening. Details of the methods are stated in the algorithm theoretical basis documents (Frankenberg, 2014; O'Dell and Taylor, 2014). The principles of the two methods are similar: since the presence of clouds and aerosols causes scattering and modifies the optical path length, there should be apparent differences between modeled and measured results assuming clear-sky conditions (with no scattering, or molecular Rayleigh scattering only) (Frankenberg et al., 2005). Following this principle, the cloud screening criteria for each method are reiterated briefly here for the convenience of readers.

For the ABP method, three thresholds are set to test if the scene meets the clear-sky conditions: (1) the threshold for surface pressure as explained in section 2; (2) the threshold for the average of retrieved surface albedos at 0.755 μm and 0.785 μm; (3) the threshold for the reduced χ^2 value. If either one of the tested parameters is above the threshold value, the scene is classified as cloudy. These thresholds could be set to different values for different operation modes, surface types and other observational conditions.

For the IDP method, thresholds for center and half-width (HW) values are set for the ratio calculated as the VCD of CO_2 in the weak CO_2 band divided by the VCD in the strong CO_2 band. This ratio is denoted as RCO_2 in the following section. Similarly, the ratio between the VCD of H_2O in the two bands is denoted as RH_2O.

3.2. *Hybrid MODIS mask and collocation method*

To assign a MODIS reference state for each OCO-2 sounding, considering OCO-2 products are provided at a 1.3 km × 2.3 km resolution, and MODIS products at 1 km × 1 km, they need to be merged first for further analysis. Here, we adopt the procedure first described in Taylor et al. (2012) for comparing OCO-2 ABP and IDP cloud screening results to the MODIS cloud mask hybrid with cirrus reflectance. For each OCO-2 sounding, the reference state is determined by averaging MODIS pixels with a center latitude and longitude less than 2 km away from the center of that sounding. If none of the selected MODIS pixels for an OCO-2 sounding is marked as confident or probably cloudy, and all cirrus reflectance $R \leqslant 0.01$, the scene is classified as clear. Otherwise, the scene is classified cloudy. For simplicity, flags indicating confident or probably cloudy are interpreted as cloudy.

3.3. *Contingency table analysis*

A contingency table analysis provides compact summary statistics for comparing large predictive datasets. It is performed to compare the cloud screening results from the OCO-2 ABP and IDP methods and hybrid MODIS cloud mask following the terms and procedure given in Taylor et al. (2012), which provides compact summary statistics for comparing large predictive datasets. Like in Taylor's work, the results from MODIS are referred to as truth in this study. Therefore, the comparison at each scene can be classified as one of four categories: true positive (TP) for both OCO-2 and MODIS indicates a clear scene; true negative (TN) for both indicates a cloudy scene; false positive (FP) for MODIS indicates clear but OCO-2 indicates cloudy; and false negative (FN) for MODIS indicates cloudy but OCO-2 indicates clear. The rate of each category is denoted as true positive rate (TPR), false negative rate (FNR), false positive rate (FPR), and true negative rate (TNR), respectively. According to the contingency table analysis, summary statistics for each category are calculated as follows:

$$TPR = N_{TP}/N_{clear} \, ,$$
$$FNR = N_{FN}/N_{clear} \, ,$$
$$FPR = N_{FP}/N_{cloud} \, ,$$
$$TNR = N_{TN}/N_{cloud} \, , \qquad (2)$$

where N is the total number of collocated soundings for each category, and also for clear scenes and cloud scenes.

Based on these summary statistics, three diagnostic variables are calculated as follows:

$$THR = (N_{TP} + N_{FP})/N_{total} \, ,$$
$$AGR = (N_{TP} + N_{TN})/N_{total} \, ,$$
$$PPV = N_{TP}/(N_{TP} + N_{FP}) \, , \qquad (3)$$

where N_{total} is the total number of all collocated soundings. By design, the throughput (THR) gives the fraction of scenes that pass the OCO-2 cloud screening algorithms, i.e., identified as clear. The agreement (AGR) gives the fraction of scenes that are correctly classified by the algorithms, either as clear or cloudy, relative to the collocated MODIS results. The positive predictive value (PPV) gives the fraction of clear scenes predicted by MODIS that are also predicted clear by the OCO-2 algorithms.

4. Results

4.1. *Tightening of the cloud screening algorithm thresholds*

To optimize the threshold parameters for cloud screening with results from the contingency table analysis, we analyzed the variation of summary statistics and diagnostic variables independently with the OCO-2 measurements and MODIS reference state in January and July 2016. The collocated cloud screening dataset in January is composed of 71 orbits, with 311 500 soundings passing over the selected region in Europe. The collocated dataset in July is composed of 72 orbits, and a total of 304 438 soundings.

Figure 3 shows the changes of diagnostic values, throughput, agreement, and PPV, in response to altering cloud screening thresholds in a chosen range. Based on the July dataset, the trend of these changes gives a way to evaluate the influence of each threshold value, including the surface pressure difference and χ^2 scale factor (SF) for the ABP method, and center value and HW of RCO$_2$ and RH$_2$O for the IDP method. Because the limit of χ^2 is dynamically calculated for each sounding, a multiplicative SF is used to evaluate all soundings. For retrieved surface albedos, the threshold is adopted from the current OCO-2 parameter and therefore its influence is not examined in this study.

Based on the trend shown in Fig. 3, we first determined Δp_s and HW since the figure indicates they have a stronger influence over the changes of the outcome, and then we determined the other parameter for each pair (χ^2 SF or center value, respectively), noting that the first determined parameter would be more crucial. The six major threshold parameters are adjusted back and forth, until the throughput of the combined results from the ABP and IDP

methods are closely matched with the average monthly clear-sky fraction in the region from MYD08.

In general, a set of tight thresholds, i.e., lower limits of Δp_s, and χ^2 SF, and a narrower HW range, as well as some shifts in the center value for acceptable RCO$_2$ and RH$_2$O, creates a more stringent cloud screening scheme, which leads to lower throughput, but a higher agreement and PPV. In other words, stringent thresholds, compared to loose ones, help to select scenes that are more "confidently clear". The fewer scenes remaining have better agreement with the MODIS reference state and are supposed to have less influence from clouds or aerosol contamination, thus giving better quality assurance.

Similar trends are also observed in the contour plot created with the January dataset; although, compared to the July dataset, the limit for Δp_s in January is more than doubled to allow a reasonable throughput from the ABP method. This could be explained by the high snow cover in winter, which is known to increase errors in cloud screening and XCO$_2$ retrievals.

A sensitivity test is performed to evaluate the rate of change of each diagnostic variable relative to different threshold values (Table 1). Five major threshold parameters are tested one at a time, while others stay the same. The χ^2 SF is not tested, because significant change in the ABP results is not observed unless the SF is set to be extremely small.

4.2. *Seasonal variation*

Based on trends in the January and July contour plots, we set the values of seasonal thresholds according to the average monthly clear-sky fraction in the selected area (Table 2). The current thresholds used in OCO-2 algorithms are designed to have 25%–30% throughput globally, which means 5%–10% more than the clear-sky fraction observed by MODIS (Taylor et al., 2016). In contrast, the narrowed thresholds aim to have throughput close to the observed local clear-sky fraction in each month. The reduction of inflation over the MODIS clear-sky fraction helps to minimize the chance that some cloud- or aerosol-contaminated scenes also pass the screening. It is also worth noting that cloud coverage varies greatly throughout the year. The highest clear-sky fraction occurs in summer, which is 55.1%, while the lowest occurs in winter, which is 29.5%. The spring and fall have close values, which are 39.3% and 40.0%, respectively. Therefore, custom thresholds for each season is important to fit the regional conditions.

A summary of the statistical values for the re-screened dataset in each season is given in Table 3. For scenes with a clear reference state, the TPR ranges from 0.69 to 0.84, which is lowest in spring and highest in summer. For scenes with a cloudy reference state, the TNR ranges from 0.91 to 0.94. The results suggest that, compared to the global results in winter and spring given in Taylor et al. (2016), the correctly predicted clear scenes increased about 10%, and the correctly predicted cloudy scenes increased about 5%.

The seasonal throughput, agreement and PPV for the
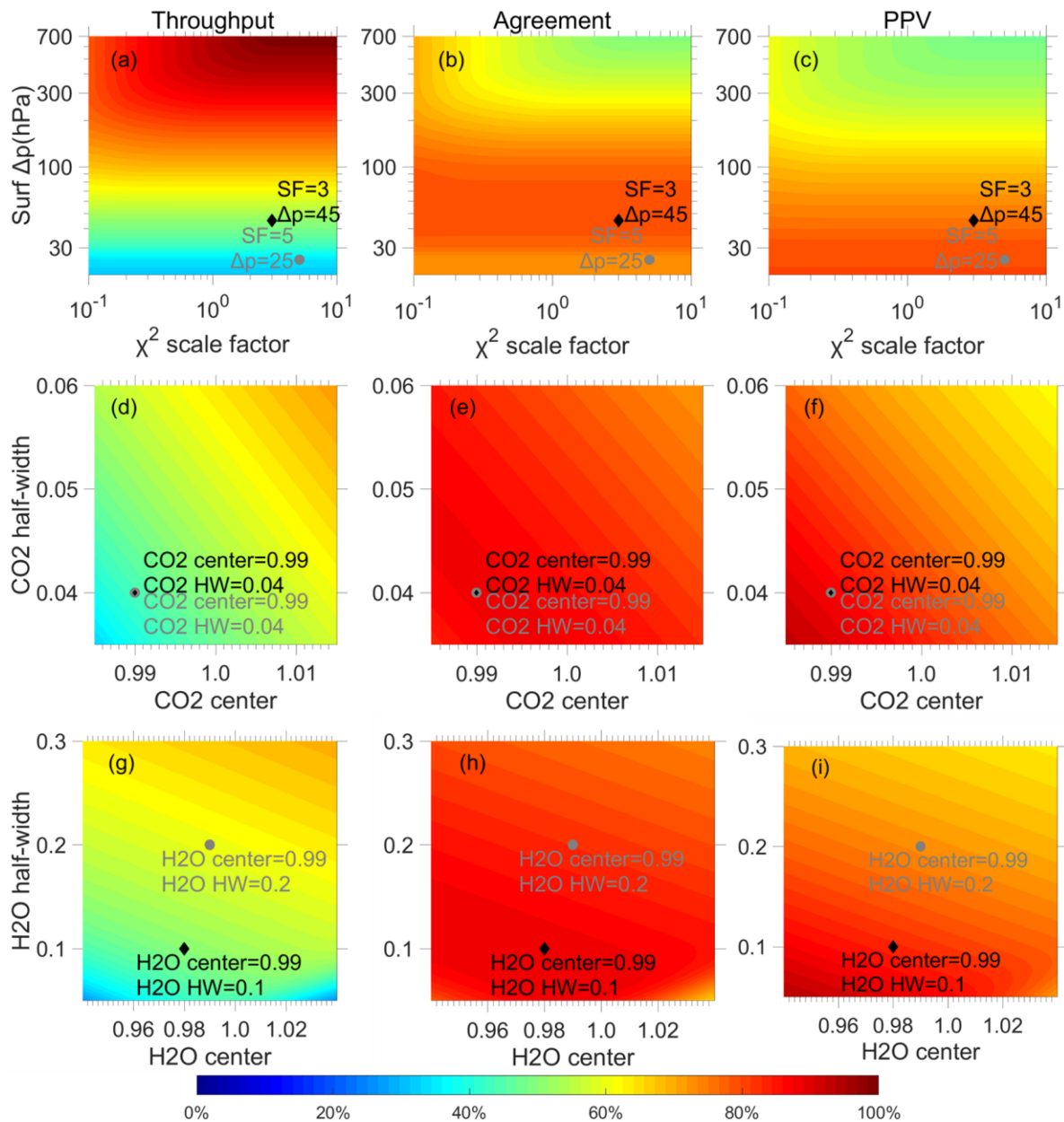
Fig. 3. Changes of the throughput (left-hand column), agreement (middle column) and positive predictive value (PPV, right-hand column) for variations in the ABP surface pressure and scale factor thresholds (a–c) and the IDP RCO_2 (d–f) and RH_2O (g–i) thresholds based on OCO-2 and MODIS data in Europe in July 2016. The numbers in black indicate the tightened thresholds in this work, while the numbers in white indicate the original OCO-2 thresholds.

ABP method, the IDP method and combined outcomes are given in Fig. 4. The total throughput is 0.18 in spring, 0.42 in summer, 0.38 in fall, and 0.20 in winter. The numbers in spring and winter are close to the values from the Glint-land viewing scenario in Taylor's work (~0.19), but lower than values from the Nadir-land (~0.26). The overall agreement with MODIS is 0.88 on average, and is relatively consistent throughout year. There is a constant improvement compared to the current OCO-2 results (~0.83). The overall PPV is 0.77, and the average PPV of spring and winter is 0.63, which is higher than the 0.58 from Taylor's results.

In general, the statistics in the summer and fall seasons are much better than in winter and spring. This indicates that the remaining data might still contain influence from snow-covered surfaces. A close examination of the results from the ABP and IDP methods shows that significant improvement of the ABP method is mainly in summer and fall, wherein the FNR can be reduced to about 0.02; on the other hand, improvement of the IDP method is mainly shown in the same seasons, with the FPR reduced to less than 0.1.

4.3. Comparison with TCCON in Europe and Japan

After determining the new thresholds and re-screening the OCO-2 measurements, the remaining retrievals were com-

Table 1. Sensitivity test for each threshold value.

Tested term*	Selected value	Test value	Change (%)	Result change (%) (for either ABP or IDP)		
				THR	AGR	PPV
Δp_s (hPa)	45	22.50	−50.00%	−9.33%	3.70%	7.83%
		33.75	−25.00%	−3.08%	2.14%	3.12%
		67.50	50.00%	4.59%	−3.86%	−4.61%
		90.00	100.00%	8.40%	−7.34%	−8.17%
RCO_2 center	0.99	0.97	−2.02%	−32.90%	−22.84%	8.67%
		0.98	−1.01%	−11.09%	−4.78%	5.46%
		1.00	1.01%	7.73%	−1.43%	−7.07%
		1.01	2.02%	14.33%	−5.12%	−13.73%
RCO_2 HW	0.04	0.030	−25.00%	−11.09%	−4.78%	5.46%
		0.035	−12.50%	−4.74%	−1.22%	3.02%
		0.045	12.50%	4.01%	−0.29%	−3.47%
		0.050	25.00%	7.73%	−1.43%	−7.07%
RH_2O center	0.99	0.96	−2.04%	−2.72%	0.25%	2.18%
		0.97	−1.02%	−1.27%	0.22%	1.08%
		0.99	1.02%	1.10%	−0.33%	−1.04%
		1.00	2.04%	2.12%	−0.74%	−2.06%
RH_2O HW	0.1	0.050	−50.00%	−8.86%	−0.96%	5.98%
		0.075	−25.00%	−3.60%	0.24%	2.83%
		0.15	50.00%	5.11%	−2.35%	−5.20%
		0.20	100.00%	9.19%	−5.34%	−9.66%

*χ^2 scale factor is not tested, because significant change in the ABP results is not observed unless the scale factor is set to be extremely small.

Table 2. Settings of the ABP and IDP cloud screening thresholds used for the seasonal OCO-2 measurements discussed in section 2.1, including the differences between the retrieved and priori surface pressure (Δp_s), χ^2 scale factor (SF), and center and half-width (HW) range for RCO_2 and RH_2O.

	Clear-sky fraction	Δp_s (hPa)	χ^2 SF	RCO_2 center	RCO_2 +/−HW	RH_2O center	RH_2O +/−HW
OCO-2(in operation)	–	20	5	0.99	0.04	0.99	0.2
spring	0.30	70	3	0.99	0.04	0.97	0.1
	0.28						
	0.31						
summer	0.34	45	2	0.99	0.04	0.98	0.1
	0.41						
	0.43						
fall	0.50	40	3	0.99	0.035	0.99	0.08
	0.56						
	0.60						
winter	0.50	100	3	0.99	0.04	0.98	0.08
	0.39						
	0.31						

pared with collocated TCCON measurements, as discussed in section 2.3. Figure 5 shows scatterplots of the seasonal daily average XCO_2 from OCO-2 versus collocated TCCON observations, in the order of time (winter, spring, summer and fall). It is rather obvious from the figure that the re-screened data [(e–h) on the right-hand side], compared to the original data [(a–d) on the left-hand side], show improvement, especially for measurements deviating from the one-to-one ratio line.

For a total of 143 days, or 143 pairs of data, 31 pairs are removed after re-screening. For the remaining 112 pairs, 97 pairs have a smaller difference compared to the original dataset. Overall, the difference between the average XCO_2 from the six TCCON sites and from the OCO-2 measurements passing nearby regions reduced 34.7%, decreasing from 3.23 ppm to 2.11 ppm. The average OCO-2 XCO_2 before re-screening is 398.19 ppm, the average uncertainty is 3.85 ppm, and the standard deviation is 0.76 ppm. After rescreening, the average XCO_2 increases slightly to 399.71 ppm, the average uncertainty decreases to 2.52 ppm, and the

Table 3. Contingency tables for the comparison of the OCO-2 cloud screening results to MODIS cloud mask for each season in 2016 in Europe. Results are from the combination of the ABP and IDP methods.

		Reference clear				Reference cloudy			
Total									
	Season	N_{TP}	TPR	N_{FN}	FNR	N_{FP}	FPR	N_{TN}	TNR
	Spring	49588	0.69	22633	0.31	27483	0.076	332775	0.92
	Summer	188747	0.84	34628	0.16	22193	0.081	253491	0.92
	Fall	173842	0.78	49533	0.22	16127	0.058	259557	0.94
	Winter	52054	0.72	20167	0.28	32380	0.090	327878	0.91
ABP									
	Season	N_{TP}	TPR	N_{FN}	FNR	N_{FP}	FPR	N_{TN}	TNR
	Spring	52097	0.72	20375	0.28	54808	0.12	388478	0.88
	Summer	220793	0.98	4182	0.019	85834	0.29	207364	0.71
	Fall	219865	0.98	5110	0.023	80070	0.27	213128	0.73
	Winter	57545	0.79	14927	0.21	80848	0.18	362438	0.82
IDP									
	Season	N_{TP}	TPR	N_{FN}	FNR	N_{FP}	FPR	N_{TN}	TNR
	spring	56648	0.78	15573	0.22	64347	0.18	296080	0.82
	summer	189336	0.85	34504	0.15	23668	0.086	252247	0.91
	fall	174424	0.78	49416	0.22	16732	0.061	259183	0.94
	winter	55735	0.77	16486	0.23	57270	0.16	303157	0.84
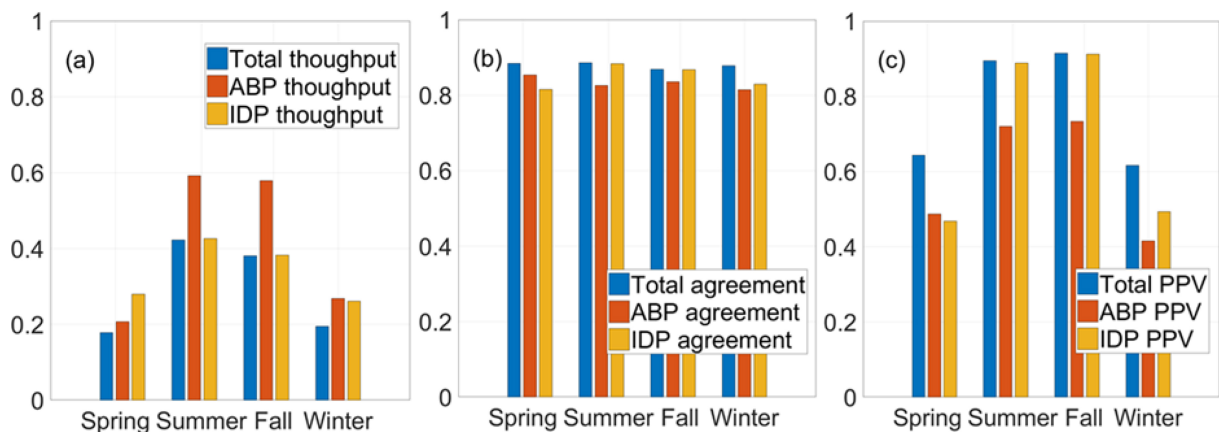
Fig. 4. Seasonal throughput (a), agreement (b) and positive predictive value (PPV, c) for the ABP method, IDP method, and combined outcomes.

standard deviation decreases to 0.71 ppm.

Among the six sites, the Garmich site shows the greatest improvement: the difference of XCO_2 between the TCCON and OCO-2 measurements decreases by 59.4%. The Karlsruhe site shows the second greatest improvement with a decrease of 42.6%. Next, the difference at the Orleans site decreases by 31.0%; the difference at the Bialystok site decreases by 28.7%; and the difference at the Paris site decreases by 24.7%. The Bremen site shows the least improvement with a decrease of 15.7%. There appears to be no relation between the position of these TCCON sites and the degree of improvement they have.

Based on the similarity of trends found in this work and Taylor's work, we believe that the same optimizing scheme can be applied to worldwide locations. We applied the same procedure to OCO-2 measurements over the land area of Japan, and compared the re-screened data with three local

TCCON sites (Table 4). However, there are far fewer data collected by these TCCON sites, resulting in fewer possible comparisons during the same period. Significant improvement of agreement between the TCCON and re-screened OCO-2 XCO_2 is shown in summer and winter, though the latter has a very small sample size.

5. Conclusion

In this work, individual sets of threshold parameters were sufficiently tuned to seasonal data from Europe in 2016. Based on two months of collocated OCO-2 and MODIS data in January and July 2016, trends of diagnostic variables as a function of threshold parameters in the ABP and IDP cloud screening methods were studied, including the fraction of scenes that pass the OCO-2 cloud screening algorithms (throughput), the fraction of scenes that are cor-
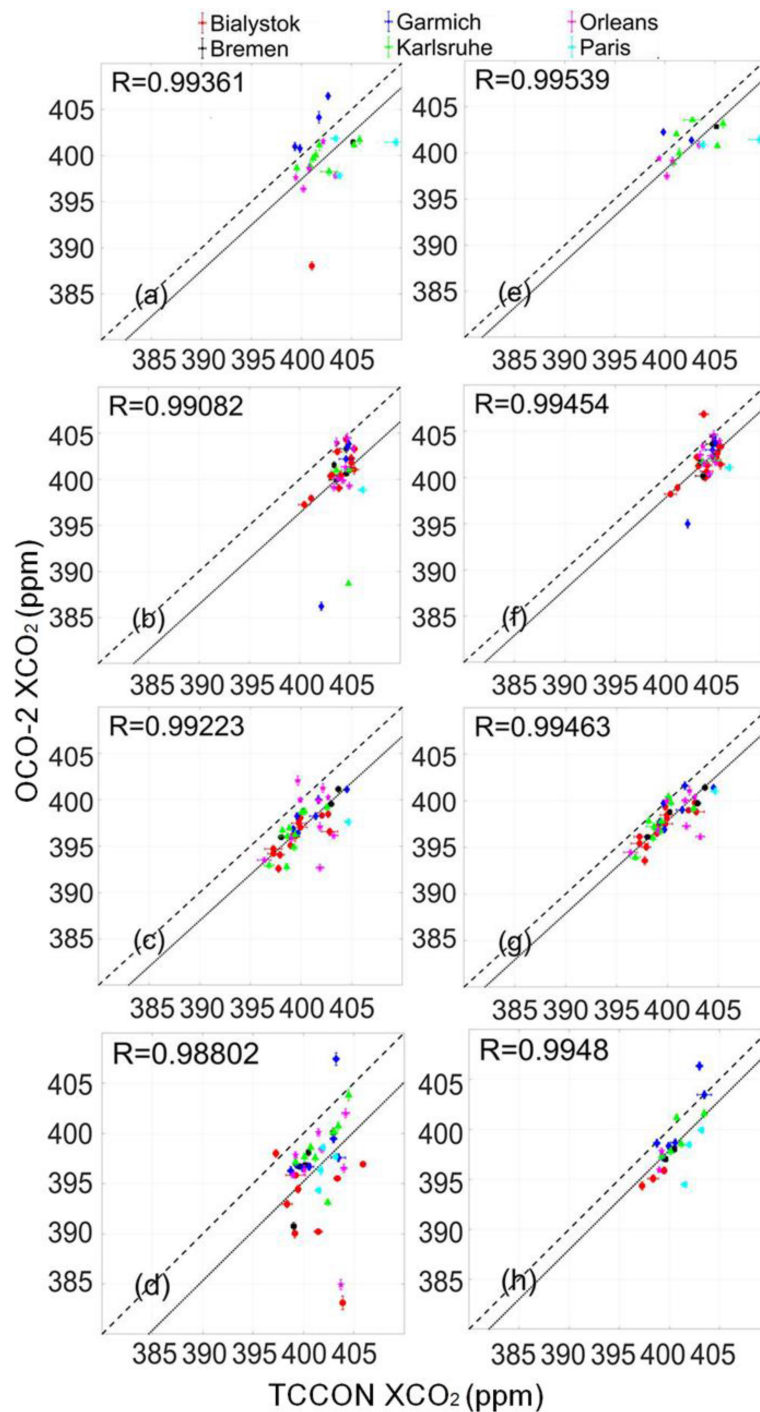
Fig. 5. Scatterplots of seasonal daily average XCO_2 from OCO-2 versus collocated TCCON observations. The dotted line is the one-to-one ratio line, while the solid line is the regression line.

rectly classified by the algorithms (agreement), and the fraction of clear scenes predicted by MODIS that are also predicted clear by the OCO-2 algorithms (PPV). By lowering the limit of throughput to the actual clear-sky fraction for each season, the possibility that contaminated scenes might pass the filter due to the loosely set thresholds was minimized. The agreement and PPV compared to a hybrid MODIS cloud mask improved steadily, at the cost of reducing total throughput.

The analysis of re-screened OCO-2 measurements confirms that stringent thresholds lead to steady improvement of agreement and PPV compared to the collocated MODIS cloud mask. In addition, comparisons with six European TCCON sites showed that the difference between the average XCO_2 from TCCON sites and from OCO-2 measurements passing nearby regions decreased from 3.23 ± 2.25 ppm to 2.11 ± 1.76 ppm. In another case study, the same optimizing procedure was applied to the Japan area, and the agreement

Table 4. Summary of the comparisons among TCCON, OCO-2, and re-screened OCO-2 measurements of XCO_2 for the Japan area during 2016.

Season	TCCON XCO_2 (ppm)	OCO-2 Data				Re-screened Data			
		Count	XCO_2 (ppm)	Diff* (ppm)	R	Count	XCO_2 (ppm)	Diff (ppm)	R
spring	407.15	1111	402.57	4.58	0.9887	869	403.32	3.83	0.9906
summer	398.63	1056	391.37	7.26	0.9817	821	399.03	−0.40	0.9919
fall	403.54	418	401.41	2.13	0.9947	46	400.41	3.13	0.9970
winter	403.92	86	399.87	4.05	0.9899	16	403.57	0.35	0.9996

*Diff refers to the difference between average TCCON measurements and OCO-2 measurements of XCO_2.

between the TCCON and OCO-2 results improved from 4.83 ± 5.25 ppm to 3.11 ± 4.82 ppm.

It is worth noting that certain assumptions in this work might lead to biases. First, we obtained the clear-sky fraction based on the cloud fraction estimated by MODIS, which does not account for the influence of aerosol, even though by the design of OCO-2 cloud screening, the days with high aerosol loadings should also be eliminated. However, the aerosol loading in Europe is generally low, so the impact from aerosols is expected to be small. Second, the rather low number of TCCON sites could lead to a representative bias. The low measuring frequency of MODIS and OCO-2 could lead to the same issue. Therefore, the results in this work focus only on the collocated measurements in time and space. For future work, the procedure described in this work could be generally applied to any other place with TCCON sites around the world. It would be very helpful to develop regional specialized cloud screening thresholds, and to provide better quality assurance for XCO_2 retrievals.

REFERENCES

Baker, D. F., H. Bösch, S. C. Doney, D. O'Brien, and D. S. Schimel, 2010: Carbon source/sink information provided by column CO_2 measurements from the Orbiting Carbon Observatory. *Atmospheric Chemistry and Physics*, **10**, 4145−4165, https://doi.org/10.5194/acp-10-4145-2010.

Bösch, H., and Coauthors, 2006: Space-based near-infrared CO_2 measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin. *J. Geophys. Res. Atmos.*, **111**, D23302, https://doi.org/10.1029/2006JD007080.

Bovensmann, H., J. P. Burrowsa, M. Buchwitza, J. Frericka, S. Noëla, V. V. Rozanova, K. V. Chanceb, and A. P. H. Goedec, 1999: SCIAMACHY: Mission objectives and measurement modes. *J. Atmos. Sci.*, **56**, 127−150, https://doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.

Burrows, J. P., E. Hölzle, A. P. H. Goede, H. Visser, and W. Fricke, 1995: SCIAMACHY - scanning imaging absorption spectrometer for atmospheric chartography. *Acta Astronautica*, **35**, 445−451, https://doi.org/10.1016/0094-5765(94)00278-T.

Butz, A., and Coauthors, 2011: Toward accurate CO_2 and CH_4 observations from GOSAT. *Geophys. Res. Lett.*, **38**, L14812, https://doi.org/10.1029/2011GL047888.

Chevallier, F., M. Fisher, P. Peylin, S. Serrar, P. Bousquet, F. M. Bréon, A. Chédin, and P. Ciais, 2005b: Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data. *J. Geophys. Res. Atmos.*, **110**, D24309, https://doi.org/10.1029/2005JD006390.

Chevallier, F., R. J. Engelen, and P. Peylin, 2005a: The contribution of AIRS data to the estimation of CO_2 sources and sinks. *Geophys. Res. Lett.*, **32**, L23801, https://doi.org/10.1029/2005GL024229.

Chevallier, F., F. M. Bréon, and P. J. Rayner, 2007: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework. *J. Geophys. Res. Atmos.*, **112**, D09307, https://doi.org/10.1029/2006JD007375.

Connor, B., and Coauthors, 2016: Quantification of uncertainties in OCO-2 measurements of XCO_2: Simulations and linear error analysis. *Atmospheric Measurement Techniques*, **9**, 5227−5238, https://doi.org/10.5194/amt-9-5227-2016.

Connor, B. J., H. Boesch, G. Toon, B. Sen, C. Miller, and D. Crisp, 2008: Orbiting carbon observatory: Inverse method and prospective error analysis. *J. Geophys. Res. Atmos.*, **113**, D05305, https://doi.org/10.1029/2006JD008336.

Crevoisier, C., S. Heilliette, A. Chédin, S. Serrar, R. Armante, and N. A. Scott, 2004: Midtropospheric CO_2 concentration retrieval from AIRS observations in the tropics. *Geophys. Res. Lett.*, **31**, L17106, https://doi.org/10.1029/2004GL020141.

Crevoisier, C., A. Chédin, H. Matsueda, T. Machida, R. Armante, and N. A. Scott, 2009: First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations. *Atmospheric Chemistry and Physics*, **9**, 4797−4810, https://doi.org/10.5194/acp-9-4797-2009.

Crisp, D., and Coauthors, 2004: The orbiting carbon observatory (OCO) mission. *Advances in Space Research*, **34**, 700−709, https://doi.org/10.1016/j.asr.2003.08.062.

Engelen, R. J., and A. P. McNally, 2005: Estimating atmospheric CO_2 from advanced infrared satellite radiances within an operational four-dimensional variational (4D-Var) data assimilation system: Results and validation. *J. Geophys. Res. Atmos.*,

110, D18305, https://doi.org/10.1029/2005JD005982.

Frankenberg, C., 2014: D-81519 OCO-2 Algorithm Theoretical Basis Document IMAP-DOAS preprocessor.

Frankenberg, C., U. Platt, and T. Wagner, 2005: Iterative maximum a posteriori (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH_4 and CO_2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT. *Atmospheric Chemistry and Physics*, **5**, 9−22, https://doi.org/10.5194/acp-5-9-2005.

Guerlet, S., and Coauthors, 2013: Impact of aerosol and thin cirrus on retrieving and validating XCO_2 from GOSAT shortwave infrared measurements. *J. Geophys. Res. Atmos.*, **118**, 4887−4905, https://doi.org/10.1002/jgrd.50332.

Gunson, M., and A. Eldering, 2014a: OCO-2 Level 2 spatially ordered geolocated retrievals screened using the A-band Preprocessor, Retrospective Processing V8r.

Gunson, M., and A. Eldering, 2014b: OCO-2 Level 2 spatially ordered geolocated retrievals screened using the IMAP-DOAS Preprocessor (IDP), Retrospective Processing V8r.

Gunson, M., and A. Eldering, 2014c: OCO-2 Level 2 geolocated XCO_2 retrievals results, physical model, Retrospective Processing V8r.

Gurney, K. R., and Coauthors, 2002: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, **415**, 626−630, https://doi.org/10.1038/415626a.

Houweling, S., W. Hartmann, I. Aben, H. Schrijver, J. Skidmore, G. J. Roelofs, and F. M. Breon, 2005: Evidence of systematic errors in SCIAMACHY-observed CO_2 due to aerosols. *Atmospheric Chemistry and Physics*, **5**, 3003−3013, https://doi.org/10.5194/acp-5-3003-2005.

Kaufman, Y. J., D. Tanré, and O. Boucher, 2002: A satellite view of aerosols in the climate system. *Nature*, **419**, 215−223, https://doi.org/10.1038/nature01091.

Ku ze, A., H. Suto, M. Nakajima, and T. Hamazaki, 2009: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Appl. Opt.*, **48**, 6716−6733, https://doi.org/10.1364/AO.48.006716.

Liang, A. L., W. Gong, G. Han, and C. Z. Xiang, 2017: Comparison of satellite-observed XCO_2 from GOSAT, OCO-2, and Ground-Based TCCON. *Remote Sensing*, **9**, 1033, https://doi.org/10.3390/rs9101033.

Mandrake, L., C. Frankenberg, C. W. O'Dell, G. Osterman, P. Wennberg, and D. Wunch, 2013: Semi-autonomous sounding selection for OCO-2. *Atmospheric Measurement Techniques*, **6**, 2851−2864, https://doi.org/10.5194/amt-6-2851-2013.

Miller, C. E., and Coauthors, 2007: Precision requirements for space-based XCO_2 data. *J. Geophys. Res. Atmos.*, **112**, D10314, https://doi.org/10.1029/2006JD007659.

Minnis, P., and Coauthors, 2008: Cloud detection in nonpolar regions for CERES using TRMM VIRS and Terra and Aqua MODIS data. *IEEE Trans. Geosci. Remote Sens.*, **46**, 3857−3884, https://doi.org/10.1109/TGRS.2008.2001351.

Morino, I., and Coauthors, 2011: Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra. *Atmospheric Measurement Techniques*, **4**, 1061−1076, https://doi.org/10.5194/amt-4-1061-2011.

O'Dell, C. W., and Coauthors, 2012: The ACOS CO_2 retrieval algorithm - Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, **5**, 99−121, https://doi.org/10.5194/amt-5-99-2012.

O'Dell, C. W., and T. E. Taylor, 2014: D-81520 OCO-2 Algorithm Theoretical Basis Document Oxygen-A Band Cloud Screening Algorithm (ABO2).

Rayner, P. J., and D. M. O'Brien, 2001: The utility of remotely sensed CO_2 concentration data in surface source inversions. *Geophys. Res. Lett.*, **28**, 175−178, https://doi.org/10.1029/2000GL011912.

Reuter, M., and Coauthors, 2011: Retrieval of atmospheric CO_2 with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results. *J. Geophys. Res. Atmos.*, **116**, D04301, https://doi.org/10.1029/2010JD015047.

Taylor, T. E., and Coauthors, 2012: Comparison of cloud-screening methods applied to GOSAT near-infrared spectra. *IEEE Trans. Geosci. Remote Sens.*, **50**, 295−309, https://doi.org/10.1109/TGRS.2011.2160270.

Taylor, T. E., and Coauthors, 2016: Orbiting Carbon Observatory-2 (OCO-2) cloud screening algorithms: Validation against collocated MODIS and CALIOP data. *Atmospheric Measurement Techniques*, **9**, 973−989, https://doi.org/10.5194/amt-9-973-2016.

Thompson, D. R., and Coauthors, 2012: Atmospheric validation of high accuracy CO_2 absorption coefficients for the OCO-2 mission. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **113**, 2265−2276, https://doi.org/10.1016/j.jqsrt.2012.05.021.

Wunch, D., and Coauthors, 2011: The total carbon column observing network. *Philos. Trans. Roy. Soc. A-Math. Phys. Eng. Sci.*, **369**, 2087−2112, https://doi.org/10.1098/rsta.2010.0240.

Wunch, D., and Coauthors, 2017: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON. *Atmospheric Measurement Techniques*, **10**, 2209−2238, https://doi.org/10.5194/amt-10-2209-2017.

Yoshida, Y., Y. Ota, N. Eguchi, N. Kikuchi, K. Nobuta, H. Tran, I. Morino, and T. Yokota, 2011: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite. *Atmospheric Measurement Techniques*, **4**, 717−734, https://doi.org/10.5194/amt-4-717-2011.