Atmospheric
Measurement
Techniques

Averaging bias correction for the future space-borne methane IPDA lidar mission MERLIN

Yoann Tellier[1], **Clémence Pierangelo**[2], **Martin Wirth**[3], **Fabien Gibert**[1], **and Fabien Marnas**[4]

[1]Laboratoire de Météorologie Dynamique (LMD/IPSL), CNRS, Ecole Polytechnique, Palaiseau CEDEX, France
[2]Centre National d'Etudes Spatiales (CNES), Toulouse CEDEX 9, France
[3]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Weßling, Germany
[4]Capgemini Technology Services (for CNES), Toulouse, France

Correspondence: Yoann Tellier (yoann.tellier@lmd.polytechnique.fr)

Abstract. The CNES (French Space Agency) and DLR (German Space Agency) project MERLIN is a future integrated path differential absorption (IPDA) lidar satellite mission that aims at measuring methane dry-air mixing ratio columns (X_{CH_4}) in order to improve surface flux estimates of this key greenhouse gas. To reach a 1 % relative random error on X_{CH_4} measurements, MERLIN signal processing performs an averaging of data over 50 km along the satellite trajectory. This article discusses how to process this horizontal averaging in order to avoid the bias caused by the non-linearity of the measurement equation and measurements affected by random noise and horizontal geophysical variability. Three averaging schemes are presented: averaging of columns of X_{CH_4}, averaging of columns of differential absorption optical depth (DAOD) and averaging of signals. The three schemes are affected both by statistical and geophysical biases that are discussed and compared, and correction algorithms are developed for the three schemes. These algorithms are tested and their biases are compared on modelled scenes from real satellite data. To achieve the accuracy requirements that are limited to 0.2 % relative systematic error (for a reference value of 1780 ppb), we recommend performing the averaging of signals corrected from the statistical bias due to the measurement noise and from the geophysical bias mainly due to variations of methane optical depth and surface reflectivity along the averaging track. The proposed method is compliant with the mission relative systematic error requirements dedicated to averaging algorithms of 0.06 % (± 1 ppb for $X_{CH_4} = 1780$ ppb) for all tested scenes and all tested ground reflectivity values.

1 Introduction

Methane (CH_4) is the second most important anthropogenic greenhouse gas after carbon dioxide (CO_2) (IPCC, 2013). Despite its key role in global warming, there are still uncertainties in the cause of the observed large fluctuations in the growth rate of atmospheric methane. Measuring atmospheric CH_4 concentration on a global scale with both high precision and accuracy is necessary to improve the surface flux estimate and thus develop the knowledge of the global methane cycle (Kirschke et al., 2013; Saunois et al., 2016).

The Methane Remote Sensing Lidar Mission (MERLIN – website: https://merlin.cnes.fr/, last access: 22 October 2018) is a joint French and German space mission with a launch scheduled for 2024 (Ehret et al., 2017). This mission is dedicated to the measurement of the integrated methane dry-air volume mixing ratio (X_{CH_4}). The German Space Agency (DLR) is responsible for the payload while the French Space Agency (CNES) is responsible for the platform (MYRIADE Evolution product line). The payload data processing centre is under CNES responsibility with significant contributions from DLR.

MERLIN's active measurement is based on a space-borne integrated path differential absorption (IPDA) lidar. Just like a differential absorption lidar (DIAL), MERLIN's IPDA lidar uses the difference in transmission between an online pulse with a frequency accurately set in the trough of several CH_4 absorption lines and an offline pulse whose wavelength has a negligible CH_4 absorption (Ehret et al., 2008). Furthermore, the two wavelengths are set close enough in such a way that the differential effects of any other interaction, excluding
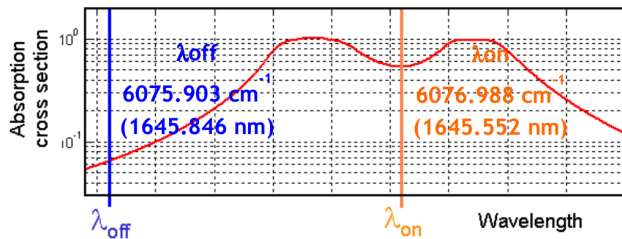
Figure 1. Laser frequency positioning of the online and offline laser beams. The online frequency is positioned in the trough of one of the methane absorption line multiplets. The offline frequency is positioned so that the methane absorption is negligible.

Figure 2. Principle schematics of the MERLIN IPDA lidar measurement. The lidar emits two laser beams with slightly different wavelengths (λ_{on} and λ_{off}). Every measurement corresponds to the small fraction of the two laser beams – called online and offline signals – that are reflected by a "hard" target (Earth's surface, top of dense clouds) to the satellite receiver telescope. For clarity, the three averaging windows are represented with four measurements instead of 150. On every averaging window, geophysical parameters such as altitude (or scattering surface elevation when there are clouds) or reflectivity vary.

CH_4 absorption, are minimized. Figure 1 shows the positioning of the two wavelengths. However, unlike a DIAL, MERLIN's IPDA lidar provides the column content of a specific trace gas along the line of sight rather than the range-resolved profile of CH_4. This column-integrated methane mixing ratio can be retrieved from the return signals after they are backscattered on a hard target such as the surface of the Earth or dense clouds. The much higher backscatter signal from these targets allows for a system with a relatively small power-aperture product as compared to a DIAL, which has to rely on atmospheric backscatter.

The MERLIN measurements require a well-defined processing chain that ensures the final performance of the mission. The processing chain is divided into four levels. Level 0 (L0) consists of raw data (backscattered signals and auxiliary data), and level 1 (L1) processes the vertically resolved products and the differential absorption optical depth (DAOD) values for both individual calibrated signal shot pairs and for a horizontal averaging window. Level 2 (L2) computes the X_{CH_4} for both individual calibrated signal shot pairs and for a horizontal averaging window, additionally using operational analyses from numerical weather prediction (NWP) centres. Note that in the presence of clouds, two products are provided; the first one computes the average for clear-sky shots only, and the other one averages all shots. Finally, level 3 (L3) produces X_{CH_4} maps using a Kalman filter approach (Chevallier et al., 2017).

To reach a usable precision, space-borne IPDA lidar missions often require an averaging of measurements along the orbit's ground track (Grant et al., 1988). This process of averaging data horizontally is a general concern for IPDA lidar missions. The data processing of the NASA Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (AS-CENDS) mission considers the averaging of multiple lidar measurements along track over 10 s (70 km with no gaps) to reduce the random error on the carbon dioxide mixing ratio: X_{CO_2} (Jucks et al., 2015). Likewise, MERLIN's averaging process is included into L1 and L2 algorithms in order to reduce the relative random error (RRE) of DAOD and X_{CH_4} (Fig. 2). For the MERLIN mission, measurements are aver-
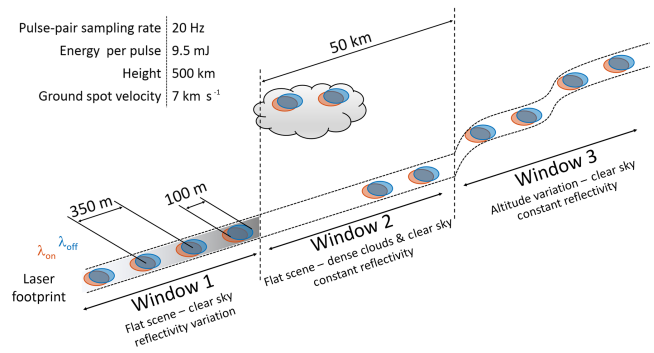
aged over a nominal window length of 50 km corresponding to about 150 shot pairs to reach an RRE of approximately 20 ppb.

The non-linearity of the equation relating calibrated signals and DAOD in combination with both the statistical noise inherent to any measurement and the varying geophysical quantities (altitude, pressure, reflectivity) of the sounded scene increases the relative systematic error (RSE or bias) and impairs measurement accuracy. Werle et al. (1993) describe RRE reduction when averaging signals using the concept of Allan variance. Up to an optimal integration time, measurement variance reduces because the measurement is dominated by white noise. For greater integration times, the estimation is biased due to drifts inherent to the measurement systems. The aim of the present article is not to correct biases caused by real system drift but to correct biases that are caused by the non-linearity of the IPDA lidar measurement equation.

MERLIN must reach an unprecedented precision and accuracy on X_{CH_4} with a targeted RRE of 1 % (18 ppb). The targeted RSE must remain under 0.2 % (± 3 ppb) in 68 % of cases; a limited budget of 0.06 % (± 1 ppb for a X_{CH_4} of 1780 ppb) is allocated to biases introduced by averaging algorithms with algorithms to correct these averaging biases. To reach the RRE target, levels 1 and 2 of MERLIN's signal processing requires a horizontal averaging of data over 50 km along track (Kiemle et al., 2011). Thus, the single shot online and offline random error is reduced by a factor of $\sqrt{150} \approx 12$. For instance, for the typical target reflectivity (0.1), the online and offline signal-to-noise ratios (SNRs) are of the order of 6.1 and 16.5, respectively, and the equivalent SNRs for

the averaged signals are, respectively, 79.6 and 197.2. This process greatly decreases the RRE of the X_{CH_4}.

Section 2 gives an overview of the IPDA measurement and MERLIN data processing. Section 3 defines and compares biases of several averaging schemes (described below) and suggests correction algorithms. Section 4 presents a comparative evaluation of these averaging schemes and associated bias correction procedures using modelled scenes based on real satellite data. And finally, in Sect. 5, the results of the simulation are described and a "best approach" algorithm (i.e. the least biased on tested scenes) is proposed for the MERLIN processing chain.

2 Overview of IPDA measurement and the MERLIN processing chain

2.1 IPDA measurement

MERLIN active measurement is based on a short-pulse IPDA lidar. The column content of methane between the satellite and a "hard" target (ground, vegetation, clouds, etc.) is retrieved by measuring the light that is reflected by the scattering surface, which is illuminated by two laser pulses with a slight wavelength difference. Figure 2 schematically shows the principle of the nadir-viewing space-borne lidar MERLIN. The pulse-pair repetition rate is 20 Hz, and the sampling distance is 350 m considering a ground spot velocity of about $7 \, km \, s^{-1}$. The online and offline ground spots are separated by about 2 m, which is negligible compared to the ground diameter of the spots of about 100 m (90 % encircled energy). Shot pairs will be averaged over a 50 km window (about 150 shots pairs). The online wavelength λ_{on} (1645.552 nm; 6076.998 cm^{-1}) is positioned in the trough of one of the methane absorption line multiplets, whereas the offline wavelength λ_{off} (1645.846 nm; 6075.903 cm^{-1}), which serves as reference, is positioned such that the methane absorption is negligible (Fig. 1). Both wavelengths are close enough so that interactions with the ground and the atmosphere and instrumental response can be considered identical, notably for reflectivity, which is defined as the ratio of the power reflected toward the satellite receiver to that incident on the hard target. The difference is thus mostly sensitive to the difference in methane absorption.

2.2 MERLIN processing chain

When the offline and online radiation reach the photodetector (Avalanche Photo Diode), it is converted to photoelectrons and to an electrical current. The measured raw signal obtained is the sum of the lidar signal and a background signal that is produced by background light, detector dark current and electronic offset. This background signal must be estimated to be removed from the raw signal. In the presence of measurement noise, when the SNR is low, this process of

background signal removal can lead to a negative estimated lidar signal.

For the sake of conciseness, we introduce for any variable X the notation $X^{on,off}$ that interchangeably represents the online or offline variables X^{on} or X^{off}. By measuring the online and offline pulse energies denoted $P^{on,off}$ (P^{on} or P^{off}, respectively), it is possible to compute the DAOD of methane and then retrieve X_{CH_4} for the sounded column. We denote $Q^{on,off}$ the measurements after normalization by the laser pulse energies, denoted $E^{on,off}$, and range r, which is the distance from the satellite to the reflective target:

$$Q^{on,off} = \frac{P^{on,off} \cdot r^2}{E^{on,off}}. \tag{1}$$

The quantity $Q^{on,off}$ will be referred to as calibrated signals in the following sections of the present article. The DAOD used in this study, in which the contributions of other gases are neglected, is denoted δ and is computed as Eq. (2):

$$\delta = \frac{1}{2} \cdot \ln\left(\frac{Q^{off}}{Q^{on}}\right) = -\frac{1}{2} \cdot \ln\left(\tau^2\right), \tag{2}$$

where τ^2 is the relative two-way transmission. From δ, we can derive X_{CH_4} from Eq. (3) (Ehret et al., 2008; Kiemle et al., 2011):

$$X_{CH_4} = \frac{\delta}{IWF} = \frac{\int_{p_{surf}}^{0} vmr_{CH_4}(p) \cdot WF(p,T) \cdot dp}{\int_{p_{surf}}^{0} WF(p,T) \cdot dp}, \tag{3}$$

where p_{surf} denotes the target pressure where the laser beam hits the ground, p and T are the pressure and temperature profiles, and $vmr_{CH_4}(p)$ is the dry-air volume mixing ratio profile of methane. The weighting function $WF(pT)$ describes the measurement sensitivity of X_{CH_4} along the vertical, and IWF is the integrated weighting function of the column. These quantities are computed from meteorological and spectroscopic data, and the WF is given by the following equation:

$$WF(p,T) = \frac{\sigma_{on}(p,T) - \sigma_{off}(p,T)}{g(p) \cdot (M_{air} + M_{H_2O} \cdot \rho_{H_2O}(p,T))}. \tag{4}$$

M_X denotes the molecular masses of the chemical species X, ρ_{H_2O} is the dry-air volume mixing ratio of water vapour, $g(p)$ stands for the acceleration of gravity (treated as altitude and hence p dependent), and here, $\sigma_{on,off}$ are the cross sections for the online or offline wavelengths (not to be confused with the standard deviation notation σ used elsewhere in this article).

As previously mentioned, in order to reach the targeted 1 % relative random error on X_{CH_4} measurements, the signal processing of MERLIN requires a horizontal averaging of the data. However, we will show in next section that the non-linearity of Eq. (2) in combination with the measurement noise and the variability of the observed scene (surface elevation, reflectivity, meteorology) along the averaging window induces biases on the average X_{CH_4}.

2.3 MERLIN measurement noise

As will be seen in the following sections, the noise that affects the measurement is one of the factors that induce the averaging bias on the retrieved methane mixing ratio. The noise originates from the detector noise, shot noise and speckle noise. In the case of MERLIN system, the dominant noise is the detector noise which is considered to be normal as it is mainly thermal noise. Then, due to the high number of photons within the signal (approximately 10^3 for the dark current and lidar signal), the Poisson statistics approximates a shifted Gaussian distribution very well (central limit theorem). Furthermore, according to Kiemle et al. (2011), the laser speckle is not the dominant source of the statistical fluctuation and is even negligible thanks to the relatively large field of view and surface spot size. The normality of the noise on calibrated signals $Q^{\text{on,off}}$ is also justified by real measurements (out of the scope of this paper). The noise model used to generate the simulated signals is based on MERLIN system parameters and is presented in the Appendix A.

3 Averaging schemes and bias correction: a theoretical approach

3.1 Definitions

In the following sections, we will use triangular bracket notation to denote the arithmetic sample mean $\langle Y \rangle = \frac{1}{N} \sum_{i=1}^{N_s} Y_i$ of the quantity Y, and $\Delta Y_i = Y_i - \langle Y \rangle$ will represent the deviation of the ith quantity to this arithmetic mean. By extension, when we use a weighted sample mean of the quantity Y, weighted by a quantity Z, we will denote it $\langle Y \rangle_{w[Z]} = \sum_{i=1}^{N_s} w_i[Z] \cdot Y_i$, where $w_i[Z] = Z_i / \sum_{k=1}^{N_s} Z_k$ are the normalized weights used. The expected value of a random variable X will be denoted $E[X]$, and the fact that X follows a normal distribution of mean value μ and variance σ^2 will be denoted $X \sim N\left(\mu, \sigma^2\right)$.

We are interested in the retrieval of the column-integrated methane concentration on a 50 km horizontal section along the satellite track. This quantity will be hereafter denoted $\overline{X_{\text{CH}_4}}^{\text{T}}$ (where T stands for target). The information that we can compute using the satellite measurements is the shot-by-shot $X_{\text{CH}_4,i}$ (i is the shot index), which is related to the shot-by-shot volume mixing ratio of methane $\text{vmr}_{\text{CH}_4,i}(p)$ and the shot-by-shot weighting function $\text{WF}_i(p)$ by Eq. (3). For the purpose of building the data processing chains, all the quantities must be described on a gridded model (vertical and horizontal discretization) of the atmosphere. This grid is composed of $(N_1 \cdot N_s)$ cells where N_1 is the number of vertical layers of the model and N_s is the number of shots that we want to average along the satellite path. To model the atmosphere, the pressure at the interface of each layer (at each $N_1 + 1$ levels) uses a hybrid sigma coordinate system and is denoted $P_{i,j}$. Note that the standard notation for in-

dices will be kept consistent throughout this article. The first index (often denoted i) will represent the shot index and the second index (often denoted j) will represent the layer index (or level index). The term "level" stands for the vertical level in pressure units. The pressure thickness of every layer, denoted $\Delta P_{i,j}$, is then derived from the pressure at every level.

The discrete form of Eq. (3) is

$$X_{\text{CH}_4,i} = \frac{\sum_{j=1}^{N_1} \text{vmr}_{\text{CH}_4,i,j} \cdot \text{WF}_{i,j} \cdot \Delta \text{P}_{i,j}}{\sum_{j=1}^{N_1} \text{WF}_{i,j} \cdot \Delta \text{P}_{i,j}}. \tag{5}$$

In order to define the average value $\overline{X_{\text{CH}_4}}^{\text{T}}$, we must define average values for the volume mixing ratio of methane and the weighting function. As the two quantities are intensive properties, it is necessary to multiply them by the pressure thickness to get the corresponding additive quantity. The average volume mixing ratio and the average weighting function of the jth layer are thus given by

$$\overline{\text{vmr}_{\text{CH}_4}}_j = \sum_{i=1}^{N_s} \pi_{i,j} \cdot \text{vmr}_{\text{CH}_4,i,j}, \tag{6}$$

$$\overline{\text{WF}}_j = \sum_{i=1}^{N_s} \pi_{i,j} \cdot \text{WF}_{i,j}, \tag{7}$$

where the weights are defined as

$$\pi_{i,j} = \frac{\Delta P_{i,j}}{\sum_{k=1}^{N_s} \Delta P_{k,j}}. \tag{8}$$

The pressure thickness, as an extensive property, is averaged arithmetically, and the average value is denoted $\overline{\Delta P}_j$. Then, we can define the average column-integrated methane concentration as

$$\overline{X_{\text{CH}_4}}^{\text{T}} = \frac{\sum_{j=1}^{N_1} \overline{\text{vmr}}_{\text{CH}_4,j} \cdot \overline{\text{WF}}_j \cdot \overline{\Delta P}_j}{\sum_{j=1}^{N_1} \overline{\text{WF}}_j \cdot \overline{\Delta P}_j}. \tag{9}$$

3.2 Averaging schemes and types of biases

There are several ways to average the X_{CH_4} provided the shot-by-shot calibrated signals $Q_i^{\text{on,off}}$. Table 1 presents four different averaging schemes: averaging of columns of X_{CH_4} (AVX – first line of Table 1), averaging of columns of DAOD and IWF (AVD – second line of Table 1), averaging of signals (AVS – third line of Table 1) and averaging of quotients (AVQ – fourth line of Table 1). Since these four averaging schemes do not average the same physical quantity, they are differently biased.

There are two main causes of bias on the retrieved X_{CH_4}: the statistical bias and geophysical biases. The statistical bias, which affects every shot individually, is not produced by the averaging process and must be taken into account for shot-by-shot measurement. It is induced by the random nature of the measurement of online and offline signals into non-linear equations. Figure 3 illustrates the statistical bias on the DAOD, when online and offline signals follow normal

Table 1. Averaging schemes and characteristics of their biases.

Averaging scheme	Abbreviation	Definition	Bias characteristics
Averaging of columns of X_{CH_4}	AVX	$\overline{X_{CH_4}}^{avx} = \langle \frac{\delta}{IWF} \rangle_{w[IWF]}$	– Statistical bias due to measurement noise on every shot – Type 1 geophysical bias from averaging concentrations instead of molecular content
Averaging of columns of DAOD and IWF	AVD	$\overline{X_{CH_4}}^{avd} = \frac{\langle \delta \rangle}{\langle IWF \rangle}$	– Statistical bias due to measurement noise on every shot
Averaging of signals	AVS	$\overline{X_{CH_4}}^{avs} = \frac{\frac{1}{2} \cdot \ln\left(\frac{\langle Q^{off} \rangle}{\langle Q^{on} \rangle}\right)}{\langle IWF \rangle_{w[Q^{off}]}}$	– Statistical bias due to measurement noise of the resulting signals on the averaging window – Type 2 geophysical bias due to linearization of the DAOD variations and correlation between DAOD and reflectivity variations – Type 3 geophysical bias due to the higher sensitivity to measurements with high offline signal strength
Averaging of quotients (not detailed in this paper due to bad performances)	AVQ	$\overline{X_{CH_4}}^{avq} = \frac{\frac{1}{2} \cdot \ln\langle \frac{Q^{off}}{Q^{on}} \rangle}{\langle IWF \rangle}$	– Statistical bias due to measurement noise mixed with geophysical biases into the nonlinear equation (cf. Appendix B)

distributions. It highlights that, in this case, the DAOD derived from these signals is no longer normally distributed, and it indicates a bias and a skewness. The second main sources of bias are called geophysical biases. These biases are induced by the process of averaging. The successive averaged shots do not sound the same portion of atmosphere (surface pressure and gas concentrations vary), they are not reflected on the same surface (reflectivity varies) and the elevation of the scattering surface is not constant in general (altitude and hard-target surface pressure vary). All these variations of geophysical quantities induce several biases on the average values.

The first scheme, AVX, directly averages the column mixing ratios of methane. Every shot is impacted by the statistical bias developed in Sect. 3.3.1. Furthermore, since a column with a high total molecular content and another with fewer molecules would count the same in the averaged mixing ratio, the uniform weighting of methane concentrations leads to the creation of a bias that is called geophysical bias of type 1, described in Sect. 3.4.1.

The second scheme, AVD, computes the ratio of the mean DAOD and the mean IWF. It is also impacted by the statistical bias (cf. Sect. 3.3.1). However, this scheme takes into account the fact that every column does not present the same molecular content as DAOD and IWF are averaged separately. Thus, it is not impacted by geophysical bias of type 1.

The third scheme, AVS, averages signals before computing relative transmissions, DAOD and X_{CH_4}. The statisti-
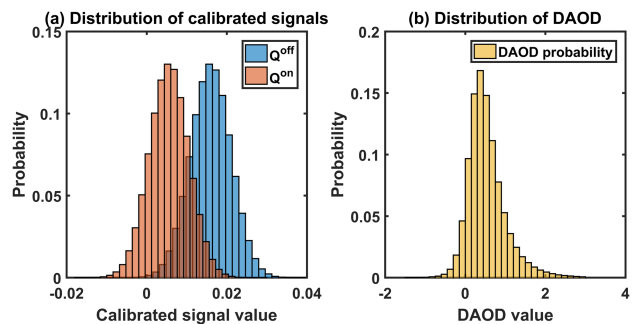

Figure 3. Effect of the non-linearity on the DAOD distribution for a low reflectivity (0.016 ice and snow cover). Panel **(a)** shows that the online and offline calibrated signals are normally distributed. A significant part of online calibrated signals (orange) is negative, which makes the corresponding double shots unusable (undefined logarithm). Panel **(b)** shows that the DAODs corresponding to the usable calibrated signals are not normally distributed, and they present a bias and are skewed. The true DAOD is 0.53 whereas the mean of the distribution is about 0.54, which leads to a bias on the X_{CH_4} of approximately +34 ppb here.

cal bias is only applicable to the resulting averaged signals such that this bias is highly reduced compared to AVX and AVD (cf. Sect. 3.3.2). However, geophysical biases are increased. First, when the DAOD varies from shot to shot – due to altitude (or surface pressure) variations or methane concentration variations for instance – the DAOD computed

from averaged signals is not representative of the true mean DAOD. This is called geophysical bias of type 2, presented in Sect. 3.4.2. Secondly, for the AVS scheme, the average DAOD is weighted by the offline signal strength. Consequently, the variance of the average quantities is reduced. However, a correlation between methane concentration and reflectivity adds a bias to the retrieved quantities. This bias is called geophysical bias of type 3 and will also be discussed in Sect. 3.4.2.

The fourth scheme, AVQ, averages transmissions before computing average DAOD (and average X_{CH_4}). The transmissions for every column are averaged with a uniform weighting. Note that the major drawback of this scheme is that it mixes several bias sources that cannot be easily corrected. Indeed with the averaging being made inside the logarithm, it is not possible to separate the bias into two terms due to the measurement noise and the variation of geophysical parameters of the scene (cf. Appendix B). This scheme will not be considered in the next sections.

In the following sections, for each averaging scheme of Table 1 (except averaging of quotients), we will quantify separately the statistical bias and the geophysical biases and will in the end combine them in order to determine the total bias for various scenarios.

3.3 Statistical bias

3.3.1 Statistical bias on AVX and AVD

The averaging of columns (either X_{CH_4} or DAOD) needs DAODs to be computed for every column of calibrated signals (Q^{off}, Q^{on}). However, as the measurements are affected by random noise and the IPDA lidar equation (Eq. 2) is not linear, a bias appears when computing the DAOD (Fig. 3). Let us define the estimator of the DAOD $\hat{\delta}$ as follows:

$$\hat{\delta} = \frac{1}{2} \cdot \ln\left(\frac{Q^{off}}{Q^{on}}\right). \qquad (10)$$

The total noise contributions affecting offline and online signals are statistically independent. Thus, for each single shot, the calibrated signals Q^{on} and Q^{off} can be considered as independent random variables. Furthermore, due to the relatively high number of photons in a single pulse, we can assume that these random variables are normally distributed around a mean value $\mu^{on,off}$ and with a standard deviation $\sigma^{on,off}$.

Under the normality assumption, Eq. (10) can be decomposed into three terms:

$$\hat{\delta} = \frac{1}{2} \cdot \ln\left(\frac{\mu^{off} + \sigma^{off} \cdot X^{off}}{\mu^{on} + \sigma^{on} \cdot X^{on}}\right) = \frac{1}{2} \cdot \ln\left(\frac{\mu^{off}}{\mu^{on}}\right) + \frac{1}{2}$$
$$\cdot \ln\left(1 + \frac{X^{off}}{SNR^{off}}\right) - \frac{1}{2} \cdot \ln\left(1 + \frac{X^{on}}{SNR^{on}}\right), \qquad (11)$$

where $X^{on,off}$ follows standard normal distributions. And the signal-to-noise ratios are defined as

$$SNR^{on,off} = \frac{\mu^{on,off}}{\sigma^{on,off}}. \qquad (12)$$

The first term of Eq. (11) is the parameter that needs to be estimated (i.e. the unbiased DAOD), and the two last terms are error terms that correspond to the bias of $\hat{\delta}$ due to the non-linearity of the function:

$$Bias_{stat}\left(\hat{\delta}\right) = \frac{1}{2} \cdot E\left[\ln\left(1 + \frac{X^{off}}{SNR^{off}}\right)\right] - \frac{1}{2}$$
$$\cdot E\left[\ln\left(1 + \frac{X^{on}}{SNR^{on}}\right)\right]. \qquad (13)$$

The task is now to evaluate this bias to remove or at least reduce it. Analytically, under the normal distribution hypothesis, the expected values are defined by:

$$E\left[\ln\left(1 + \frac{X^{on,off}}{SNR^{on,off}}\right)\right] = \frac{1}{\sqrt{2\pi}} \int\limits_{-SNR^{on,off}}^{+\infty}$$
$$\ln\left(1 + \frac{x}{SNR^{on,off}}\right) \cdot e^{-\frac{x^2}{2}} dx. \qquad (14)$$

Providing that $SNR^{on,off}$ is high enough, we can use a Taylor expansion of the logarithm around zero so that the bias can be approximated by the following formula (Bösenberg, 1998):

$$Bias_{stat}\left(\hat{\delta}\right) \approx \frac{1}{4}\left[\frac{1}{(SNR^{on})^2} - \frac{1}{(SNR^{off})^2}\right]. \qquad (15)$$

The assumption that the calibrated signals follow a normal distribution does not rigorously hold when the DAOD is computed. Indeed, over dark surfaces (low reflectivity), the SNR may happen to be so low that either one or both calibrated signals Q^{off} and Q^{on} takes negative values; hence, DAOD is undefined. This can actually happen as the calibrated signal $Q^{on,off}$ is computed from the raw signal that corresponds to a photon count (positive quantity) from which the estimated background energy has been subtracted. Whenever one of the calibrated signals is negative, the corresponding couple ($Q^{off}Q^{on}$) must be discarded. And as the lowest calibrated signals are systematically discarded from the averaging set, the average measurement is biased. This bias can be corrected by considering Eqs. (11) and (13) with $X^{on,off}$

as a left-truncated normal distribution with a mean value of zero, a variance of one and a left-truncation at $-\mathrm{SNR}^{\mathrm{on,off}}$ (Johnson et al., 1994). When done so, it becomes

$$
E\left[\ln\left(1+\frac{X^{\mathrm{on,off}}}{\mathrm{SNR}^{\mathrm{on,off}}}\right)\right] = \frac{1}{\sqrt{2\pi}\left[1-\Phi\left(-\mathrm{SNR}^{\mathrm{on,off}}\right)\right]}
$$
$$
\int\limits_{-\mathrm{SNR}^{\mathrm{on,off}}}^{+\infty} \ln\left(1+\frac{x}{\mathrm{SNR}^{\mathrm{on,off}}}\right) \cdot e^{-\frac{x^2}{2}} \,\mathrm{d}x, \qquad (16)
$$

where Φ is the standard normal cumulative distribution function.

To correct the bias due to the non-linearity of the IPDA lidar equation, the SNR must be estimated. Once done, the bias correction scheme would either need to estimate the bias directly from the approximate Taylor expansion formula of Eq. (15) or estimate the bias using Eq. (13) and a numerical computation of Eq. (16). Typically, for MERLIN observations, the error made by using the Taylor expansion of Eq. (15) instead of Eq. (16) is lower than 1 ppb on the $X_{\mathrm{CH_4}}$ for a surface reflectivity value greater than 0.1 ($\mathrm{SNR}^{\mathrm{off}} \approx 16$ and $\mathrm{SNR}^{\mathrm{on}} \approx 7$) as shown on Fig. 4. Table 2 shows the error made by using the Taylor expansion instead of computing the truncated normal integral. For values of reflectivity smaller than 0.1, it would be preferable to use the exact formula for the bias presented in Eqs. (13) and (16). Further study (not presented here) shows that for very low reflectivity, the estimation of the noise induced bias is really sensitive to an error on the SNR, and this correction is no longer applicable in practice. The way statistical bias on the DAOD is translated to bias on $X_{\mathrm{CH_4}}$ will be treated in Sect. 3.5.

3.3.2 Statistical bias on AVS

The third averaging scheme defined on Table 1, AVS, averages online and offline calibrated signals separately. The corresponding estimator of the average DAOD is written

$$
\hat{\delta}^{\mathrm{avs}} = \frac{1}{2} \cdot \ln\left(\frac{\langle Q^{\mathrm{off}}\rangle}{\langle Q^{\mathrm{on}}\rangle}\right). \qquad (17)
$$

Consistently with Sect. 3.3.1, we consider the individual calibrated signals to be normal random variables of mean $\mu_i^{\mathrm{on,off}}$ and standard deviation $\sigma_i^{\mathrm{on,off}}$. The parameters of the distributions depend on shot i since each shot is considered as the realization of a different distribution, depending on the geophysical parameters of the scene (reflectivity, atmospheric transmission, surface pressure). The successive measurements are considered independent and, as the sum of independent normal random variables, is a normal random variable. We introduce $S^{\mathrm{on,off}}$, the average random variable:

$$
S^{\mathrm{on,off}} = \langle Q^{\mathrm{on,off}}\rangle \sim N\left(m^{\mathrm{on,off}}, \left(\epsilon^{\mathrm{on,off}}\right)^2\right), \qquad (18)
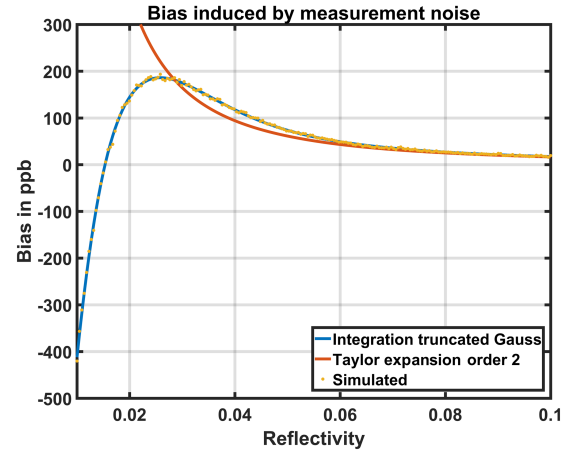$$

Figure 4. Statistical bias induced by measurement noise. The online and offline SNRs drive the value of the statistical bias. The blue line is derived from the integration of the truncated normal distribution (Eqs. 16 and 13). The orange line is the Taylor development (Eq. 15), only valid when reflectivity is high enough (i.e. high SNR). The expected bias computed from a simple Monte Carlo simulation (yellow dots) shows that the integration approach is the most accurate. For reflectivity values of 0.1 (vegetation cover), integration (blue) and Taylor development (orange), it differs by about 1 ppb (cf. Table 2 for some values).

where the mean and variance of $S^{\mathrm{on,off}}$ are

$$
m^{\mathrm{on,off}} = \langle \mu^{\mathrm{on,off}}\rangle, \qquad (19)
$$
$$
\left(\epsilon^{\mathrm{on,off}}\right)^2 = \frac{1}{N^2}\sum_{i=1}^{N}\left(\sigma_i^{\mathrm{on,off}}\right)^2. \qquad (20)
$$

The empirical estimate of the SNR of the equivalent measurement $S^{\mathrm{on,off}}$ on the whole averaging window can be written

$$
\mathrm{SNR}_{\mathrm{eq}}^{\mathrm{on,off}} = \frac{m^{\mathrm{on,off}}}{\epsilon^{\mathrm{on,off}}} = \left(\sum_{i=1}^{N}\mu_i^{\mathrm{on,off}}\right)
$$
$$
\left(\sum_{i=1}^{N}\left(\frac{\mu_i^{\mathrm{on,off}}}{\mathrm{SNR}_i^{\mathrm{on,off}}}\right)^2\right)^{-\frac{1}{2}}. \qquad (21)
$$

Given these definitions, we can write the bias due to shot random variations as in Eq. (13):

$$
\mathrm{Bias}_{\mathrm{stat}}\left(\hat{\delta}^{\mathrm{avs}}\right) = E\left[\hat{\delta}^{\mathrm{avs}}\right] - \frac{1}{2}\cdot\ln\left(\frac{m^{\mathrm{off}}}{m^{\mathrm{on}}}\right) = \frac{1}{2}\cdot E
$$
$$
\left[\ln\left(1+\frac{X^{\mathrm{off}}}{\mathrm{SNR}_{\mathrm{eq}}^{\mathrm{off}}}\right)\right] - \frac{1}{2}\cdot E\left[\ln\left(1+\frac{X^{\mathrm{on}}}{\mathrm{SNR}_{\mathrm{eq}}^{\mathrm{on}}}\right)\right], \quad (22)
$$

Provided an estimation of the shot-by-shot SNR, we can estimate the bias of Eq. (22) with the same methods as in Sect. 3.3.1, either considering the simplified Taylor expansion approximation (Eq. 15) or the more accurate integral of truncated normal distribution (Eqs. 13 and 16). Compared to

Table 2. Error on the statistical bias estimation by using the Taylor expansion instead of using a truncated normal distribution (cf. Fig. 4).

Reflectivity value	0.093	0.077	0.062	0.53	0.025
Offline SNR	15.1	13.1	10.9	9.5	4.8
Online SNR	6.1	5.2	4.2	3.6	1.8
Error made by Taylor expansion (Eq. 15)	$-1\,\mathrm{ppb}$	$-2\,\mathrm{ppb}$	$-5\,\mathrm{ppb}$	$-10\,\mathrm{ppb}$	$+50\,\mathrm{ppb}$

AVX and AVD schemes, the equivalent SNRs, when averaged horizontally on 150 shot pairs, are considerably larger and as a consequence, the bias is considerably smaller. The Taylor expansion approximation holds really well and an error on the estimation of SNR has a negligible impact on the bias estimation.

3.4 Geophysical bias

3.4.1 Geophysical bias of type 1 on AVX

Considering an arithmetic averaging for both AVX and AVD schemes yields different results, since the former scheme averages concentrations and the latter averages quantities that are proportional to number of molecules of methane. Whereas the AVX scheme computes the arithmetic mean of X_{CH_4} (Eq. 23), the AVD scheme computes average X_{CH_4} weighted by the IWF (Eq. 24):

$$\overline{X_{CH_4}}^{\mathrm{avx}} = \left\langle \frac{\delta}{\mathrm{IWF}} \right\rangle = \langle X_{CH_4} \rangle \qquad (23)$$

and

$$\overline{X_{CH_4}}^{\mathrm{avd}} = \frac{\langle \delta \rangle}{\langle \mathrm{IWF} \rangle} = \langle X_{CH_4} \rangle_{w[\mathrm{IWF}]}, \qquad (24)$$

where

$$w_i[\mathrm{IWF}] = \frac{\mathrm{IWF}_i}{\sum_{k=1}^{N_s} \mathrm{IWF}_k}. \qquad (25)$$

For the AVX scheme, the quantity that is averaged is the column concentration which is an intensive property. If a uniform weighting is considered, there is the same contribution from columns with many molecules as from ones with less molecules. For this scheme, a variation of IWF from shot to shot (i.e. variation of altitude and/or surface pressure) leads to an overestimation of the methane content of columns that contain fewer molecules in the average X_{CH_4}. This bias will be called geophysical bias of type 1 and is simply corrected by introducing the weighted average by the IWF. This has to be taken into account when computing the statistical bias for this scheme, as will be introduced in Sect. 3.5.

On the contrary, the AVD scheme averages the extensive properties of DAOD and IWF separately. Thus, when the DAODs are averaged, the molecule amount is preserved such that the AVD scheme is not affected by a type 1 geophysical bias.

3.4.2 Geophysical bias of type 2 and 3 on AVS

Once the bias induced by the random nature of the measurement has been subtracted, the resulting estimator is still biased by the effects of horizontal variations of geophysical quantities. Indeed, using Eq. (22), we are left with

$$\overline{\delta^{\mathrm{avs}}} = \hat{\delta}^{\mathrm{avs}} - \mathrm{Bias}_{\mathrm{stat}}\left(\hat{\delta}^{\mathrm{avs}}\right) = \frac{1}{2} \cdot \ln\left(\frac{m^{\mathrm{off}}}{m^{\mathrm{on}}}\right), \qquad (26)$$

where m^{off} and m^{on}, as defined by Eq. (19), are the average of signal expected values. Successive shot pairs are not measuring the same column of atmosphere, such that altitude, reflectivity and CH_4 concentration vary horizontally. Unlike for the AVX and AVC schemes where the ratio is computed separately, for the AVS scheme, the changing reflectivity or atmospheric transmission does not cancel out directly when computing the ratio of signals. Although measurement random noise is significantly reduced, a geophysical noise appears. We can rewrite Eq. (26) as

$$\overline{\delta^{\mathrm{avs}}} = -\frac{1}{2} \cdot \ln\left(\sum_{i=1}^{N_s} \frac{\mu_i^{\mathrm{off}}}{\sum_{k=1}^{N} \mu_k^{\mathrm{off}}} \cdot \exp\left(-2 \cdot \delta_i\right)\right) =$$
$$-\frac{1}{2} \cdot \ln\left(\sum_{i=1}^{N_s} w_i\left[\mu^{\mathrm{off}}\right] \cdot \tau_i^2\right). \qquad (27)$$

Using a Taylor expansion of Eq. (27), it is possible to show that $\overline{\delta^{\mathrm{avs}}}$ approximately equals the mean of the single-shot DAODs weighted by the $w_i[\mu^{\mathrm{off}}]$:

$$\overline{\delta^{\mathrm{avs}}} = \sum_{i=1}^{N_s} w_i\left[\mu^{\mathrm{off}}\right] \cdot \delta_i + R_{\mathrm{res}}, \qquad (28)$$

where R_{res} is the residual error of the linear approximations when averaging DAODs instead of transmissions. This term will be called type 2 geophysical bias. Equations (27) and (28) lead to

$$R_{\mathrm{res}} = -\frac{1}{2} \cdot \ln\left(\sum_{i=1}^{N_s} w_i\left[\mu^{\mathrm{off}}\right] \cdot \exp\left(-2 \cdot \delta_i\right)\right)$$
$$- \sum_{i=1}^{N} w_i\left[\mu^{\mathrm{off}}\right] \cdot \delta_i. \qquad (29)$$

Note that when the DAOD is constant all along the averaging scene, R_{res} is exactly zero. Furthermore, when μ^{off} is horizontally constant, R_{res} is approximately the variance of the DAOD. In fact, the term R_{res} is twofold: on the one hand, it is linked to DAOD fluctuations and, on the other hand, to

the correlation between DAOD and reflectivity fluctuations. These correlations might occur, for instance, if there are co-variations between topography (and thus DAOD) and surface type (e.g. snow on mountain tops and vegetation in valleys). In the general case, R_{res} is not zero and can be estimated using $\overline{\delta^{avs}}$ from Eq. (26), corrected for the statistical bias only, to compute a first-order estimate for X_{CH_4}:

$$X_{CH_4}^{(1)} = \frac{\overline{\delta^{avs}}}{\langle IWF \rangle_{w[Q^{off}]}}. \tag{30}$$

Using Q^{off} instead of μ^{off} which is unknown, we can estimate R_{res} as

$$R_{res}^{(1)} \approx -\frac{1}{2} \cdot \ln \left(\sum_{i=1}^{N_s} w_i \left[Q^{off} \right] \cdot \exp \left(-2 \cdot X_{CH_4}^{(1)} \right. \right.$$
$$\left. \left. \cdot IWF_i \right) \right) - \overline{\delta^{avs}}. \tag{31}$$

This process could be turned into an iterative correction. However, the first-order estimate is sufficiently accurate in all cases (not shown).

According to Eq. (28), we notice that the AVS scheme, corrected for type 2 geophysical bias, computes an average DAOD weighted by the off-signal strength. Since the main cause of variation of the offline received power is the variation of surface or hard-target reflectivity, the transmissions associated to brighter scenes count more in the average than the transmissions of darker scenes. The AVS scheme averages the measurements in such a way that a greater weight is given to high SNR signals. Consequently, this DAOD estimate is more precise (lower standard deviation) but also biased. This bias is called type 3 geophysical bias and will be defined in Sect. 3.5.

3.5   From biases on DAOD to biases on X_{CH_4}

In Sect. 3.3 and 3.4, the statistical and geophysical biases on DAOD have been derived. Here we are interested in translating biases on DAOD to biases on X_{CH_4} that we want to estimate. As shown by Eq. (3), X_{CH_4} is obtained by dividing the DAOD by the IWF. This needs the IWF to be averaged horizontally, consistent with the DAOD averaging scheme. Not only is the computation of the average IWF with consistent weights important to compute X_{CH_4}, but it is also needed by the data users for the assimilation to transport models.

For the AVD scheme, the DAODs are arithmetically averaged with a uniform weight. Hence, the IWF must be averaged in the same fashion. A shot-by-shot DAOD bias according to Eq. (13) translates into a statistical bias on X_{CH_4} as follows:

$$Bias_{stat} \left(\overline{X_{CH_4}}^{avd} \right) = \frac{\langle Bias_{stat} (\hat{\delta}) \rangle}{\langle IWF \rangle}. \tag{32}$$

For the AVX scheme, X_{CH_4} is computed for every shot. The statistical bias on every shot is the quotient of the bias on

the shot DAOD over the shot IWF. However, when horizontally averaging the statistical bias on X_{CH_4}, the type 1 geophysical bias must be taken into account (Sect. 3.4.1). The average bias should be weighted by the shot-by-shot IWF as in Eq. (24):

$$Bias_{stat} \left(\overline{X_{CH_4}}^{avx} \right) = \langle \frac{Bias_{stat} (\hat{\delta})}{IWF} \rangle_{w[IWF]}. \tag{33}$$

For the AVS scheme, the IWF must be weighted consistent with the averaging scheme. Equation (28) shows that the average DAOD is weighted by the offline signal strength. As presented in the third line of Table 1, in order to keep the mixing ratio of methane consistent, the averaging of the IWF must also be weighted by $w_i \left[Q^{off} \right]$. Consistently, the translation of bias on the DAOD to bias on the X_{CH_4} considers the same weighting for IWF. The statistical bias translates from Eq. (22) to

$$Bias_{stat} \left(\overline{X_{CH_4}}^{avs} \right) = \frac{Bias_{stat} (\hat{\delta}^{avs})}{\langle IWF \rangle_{w[Q^{off}]}}. \tag{34}$$

Concerning geophysical biases, a type 2 geophysical bias (due to the linearization of DAOD variations and the correlation of signal and transmission fluctuations) described by Eq. (31) becomes

$$Bias_{geo2} \left(\overline{X_{CH_4}}^{avs} \right) = \frac{R_{res}^{(1)}}{\langle IWF \rangle_{w[Q^{off}]}}. \tag{35}$$

The geophysical bias of type 3, caused by the higher sensitivity to the spots with higher reflectivity, could be written as

$$Bias_{geo3} \left(\overline{X_{CH_4}}^{avs} \right) = \frac{\langle \delta^{true} \rangle_{w[Q^{off}]}}{\langle IWF \rangle_{w[Q^{off}]}} - \frac{\langle \delta^{true} \rangle}{\langle IWF \rangle}. \tag{36}$$

Indeed, the AVS scheme does not measure the true concentration of CH_4 on the 50 km window. The weighting by $w_i \left[Q^{off} \right]$ implies that greater weight is granted to shots measuring brighter targets. This could be detrimental to the measurement if there were a strong correlation between reflectivity and CH_4 concentration on a global scale, which should not be the case. For assimilation or inverse modelling to models with a higher resolution than 50 km, the weighting could also be taken into account in the forward model for the X_{CH_4}.

4   Methodology to test averaging algorithms and their bias corrections

4.1   Data sets (latitude, longitude, altitude, surface pressure, and relative reflectivity)

The three averaging schemes and their associated biases will be tested on scenes modelled from real satellite data in terms

of geophysical properties. For this purpose, we are interested in simulating the calibrated signals $Q_i^{\mathrm{on,off}}$ and the integrated weighting function IWF_i, both on a 50 km scale. To be computed, the signals require the weighting functions for every shot ($\mathrm{WF}_{i,j}$), the volume mixing ratio of methane ($\mathrm{vmr}_{\mathrm{CH}_4,i,j}$), both defined on the pressure grid ($P_{i,j}$), and the target reflectivity (ρ_i) for every shot. The integrated weighting function is computed from $\mathrm{WF}_{i,j}$ (and $P_{i,j}$). The data sets are built from satellite data provided by the SPOT-5 satellite for latitude, longitude and relative reflectivity; the Shuttle Radar Topography Mission (SRTM) digital elevation map data for topography; and the European Centre for Medium-Range Weather Forecasts (ECMWF) analyses for surface pressure, from which we deduce the pressure grid on 150 shots and 19 levels.

SPOT-5 was a CNES satellite launched in 2002 and operated until 2015 (Gleyzes et al., 2003). Amongst the five spectral bands of the High Resolution Geometric (HRG) instrument, it has a spectral band in the short-wave infrared domain (1.55 to 1.7 µm) with a spatial resolution of 20 m. This band includes the MERLIN laser wavelength and, as we expect spectral variations of surface albedo to have rather low spectral variations, we use the Spot SYSTEM SCENE level 1A product (images using radiometric corrections, equivalent radiance in $\mathrm{W\,m^{-2}\,Sr^{-1}\,\mu m^{-1}}$) as a proxy of surface reflectivity. Indeed, as we were careful to select images with no clouds, we neglect the effect of atmospheric extinction on the SPOT-5 measurements. Note that we are interested here in a description of the reflectivity variations in the 50 km averaging window, not by the absolute value of reflectivity. This is why we consider this SPOT-5 product suitable, and we will scale it to any prescribed value of surface reflectivity in the simulations described hereafter anyway. The topography is taken from the SRTM digital elevation model (Jarvis et al., 2008), which has a spatial resolution of about 90 m. Surface pressure is taken from ECMWF 4D variational analyses from the long window daily archive and interpolated at SRTM grid points. A correction from difference between ECMWF Integrated Forecasting System (IFS) model topography and SRTM altitude is applied. In order to make both SRTM and SPOT-5 data consistent, the three selected SPOT-5 images are first processed by a low pass convolution to obtain a 90 m spatial resolution and then projected into the SRTM geometry. Note that the spatial resolution thus obtained is also close to MERLIN single shot footprint. Table 3 summarizes the data sets' content and resolutions.

Three sites have been selected to be representative of topographic variability; they are located in the neighbourhood of three French cities: Toulouse, Millau and Chamonix. The different characteristics of the three samples are described in Table 4. Figures 5 and 6 show the variation of surface pressure and relative variations of reflectivity along the averaging scheme. Toulouse presents a medium variation of geophysical parameters (altitude and thus surface pressure), Millau presents a high variation and Chamonix a very high variation.
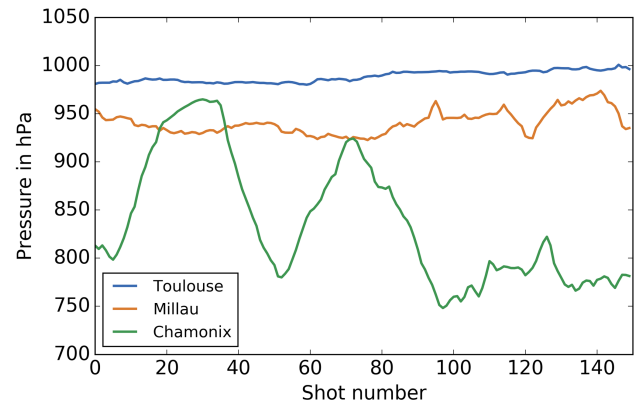

Figure 5. Surface pressure of the three scenes from the data sets. Toulouse, Millau and Chamonix present medium, high and very high variability, respectively (cf. Table 4).
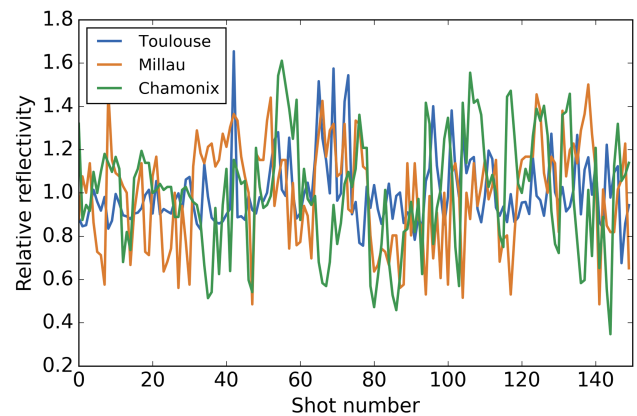

Figure 6. Relative variations of reflectivity of the three scenes from the data set. Toulouse, Millau and Chamonix present medium, high and high) variability, respectively (cf. Table 4).

Figure 7 shows the global cumulative distribution of standard deviations of altitude of SRTM database worldwide. We notice that 67 % and 97 % of the scenes present a lower altitude standard deviation than the one considered on the Millau and Chamonix data, respectively.

For sensitivity study purposes, the reflectivity relative variations from the SPOT-5 data set are multiplied by a reference mean reflectivity that can be chosen to obtain the usable scene reflectivity. Four mean reflectivity values will be considered: 0.1 (vegetation), 0.05 (mixed water and vegetation), 0.025 (sea and ocean) and 0.016 (ice and snow).

The pressure grid $P_{i,j}$ and the pressure thickness grid $\Delta P_{i,j}$ are obtained from surface pressure P_i^{surf} from the ECMWF analyses data set using a hybrid sigma coordinate system.

The methane volume mixing ratio, $\mathrm{vmr}_{\mathrm{CH}_4,i,j}$, is arbitrarily set to values assumed to be realistic. For every shot i and

Table 3. Data sets' resolution characteristics.

Geophysical parameter	Origin	Original grid resolution	Interpolated grid resolution
Coordinates (lat, long)	SPOT-5	20 m	
Relative reflectivity		90 m	(surface pressure is corrected to take into account SRTM small-scale variations of topography)
Altitude	SRTM	90 m	
Surface pressure	ECMWF	∼ 10 km	

Table 4. Characteristics of the data used for the simulation.

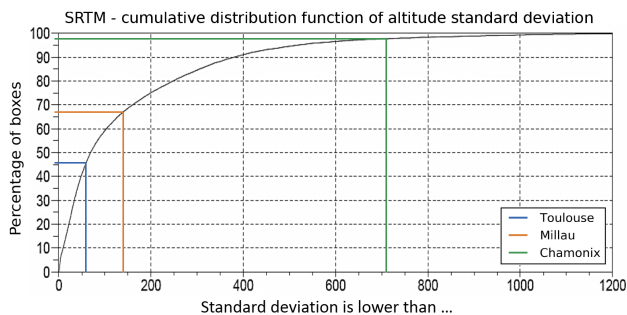|	Toulouse	Millau	Chamonix
Latitude range	43.56–43.93° N	43.56–43.93° N	45.75–46.12° N
Longitude	1.62° E	3.06° E	7.22° E
Altitude range (m)	108–321	359–902	473–2967
Altitude mean (m)	223.1	697.4	1753.7
Altitude standard deviation (m)	57.5	141.8	711.1
Surface pressure range (hPa)	980.2–1000.9	922.7–973.9	748.2–965.0
Surface pressure mean (hPa)	988.7	940.9	837.3
Surface pressure standard deviation (hPa)	5.7	12.2	64.5
Relative reflectivity range	0.68–1.65	0.49–1.50	0.35–1.61
Relative reflectivity standard deviation	0.16	0.24	0.27

Figure 7. Global cumulative distribution of the standard deviation of altitude obtained on SRTM. A total of 46 %, 67 % and 97 % of SRTM boxes present a lower standard deviation than the Toulouse scene, Millau scene and Chamonix scene, respectively. The three scenes are representative of medium, high and very high variations of altitude.

layer j

$$\begin{cases} \text{vmr}_{CH_4,i,j} = 1780\,\text{ppb} & \text{if } P_{i,j} < (\max(p_i^{\text{surf}}) - \min(p_i^{\text{surf}}))/2 \\ \text{vmr}_{CH_4,i,j} = 1880\,\text{ppb} & \text{otherwise} \end{cases} \quad (37)$$

This replicates the possible correlation between methane concentration and altitude (more methane in valleys and less over mountain tops).

Finally, the weighting functions are calculated, as described in Eq. (4), from methane absorption cross sections and meteorological data ($\Delta P_{i,j}$, temperature, humidity). They are computed using CH_4 absorption cross sections from the 4A radiative transfer model (Scott and Chédin,

1981; Chéruy et al., 1995) on a reference winter mid-latitude atmosphere from the Thermodynamic Initial Guess Retrieval (TIGR) data set (Chevallier et al., 2000). The sensitivity to the thermodynamic condition of the atmosphere has been tested and is negligible here (not shown).

4.2 Overall test framework

The aim of the simulation is to compare the biases of the estimated X_{CH_4} for several averaging schemes and to evaluate the accuracy of the bias correction. A global description of the simulation is presented on Fig. 8. Each simulation case considers a typical number of $N_s = 150$ double shots per averaging window, approximately corresponding to 50 km along the satellite ground track. It relies on a description of the geophysical scene in terms of surface pressure P_i^{surf}, reflectivity ρ_i, an arbitrary CH_4 concentration field $\text{vmr}_{CH_4,i,j}$ and weighting functions $WF_{i,j}$ (cf. Sect. 4.1). Then, the online and offline calibrated signals are computed from surface reflectivity and a random noise simulation, and the weighting functions are integrated (cf. Sect. 4.3). Next, we proceed to the computation of the average X_{CH_4} on 50 km resolution with the different averaging schemes (AVQ not simulated) and the correction algorithms, presented in Sect. 4.4 and Table 5.

In order to estimate the bias, the computation of an average column-integrated methane concentration X_{CH_4} from the shot-by-shot volume mixing ratio profiles, $\text{vmr}_{CH_4,i}(p)$, is needed and will be computed as $\overline{X_{CH_4}}^T$ in Eq. (9).

Table 5. Computational details about averaging schemes and bias evaluation.

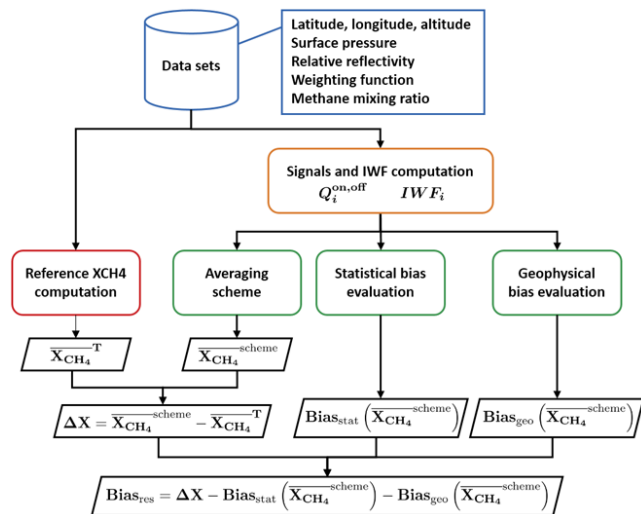|	AVX scheme	AVD scheme	AVS scheme
Averaging scheme	– Discard negative signals – Table 1, line 1	– Discard negative signals – Table 1, line 2	– Table 1, line 3
Statistical bias evaluation	– Discard negative signals – SNR estimation – Bias evaluation: Option 1 Eqs. (33) and (15) Option 2 Eqs. (33), (16) and (13)	– Discard negative signals – SNR estimation – Bias evaluation: Option 1 Eqs. (32) and (15) Option 2 Eqs. (32), (16) and (13)	– SNR estimation – Equivalent window SNR by Eq. (21) – Bias evaluation: Option 1 Eqs. (34) and (15) Option 2 Eqs. (34), (16) and (13)
Geophysical bias evaluation	– None (Type 1 geophysical bias built-in w[IWF] weights of Table 1, line 1 and Eq. 33)	– None	– Type 2 geophysical bias of Eq. (35) – Type 3 geophysical bias of Eq. (36) not estimated

Figure 8. Global description of the simulation. Data sets (blue) are described in Sect. 4.1. Signals and the IWF computation (orange) are described in Sect. 4.3. Averaging strategies performed and their related bias corrections (green) are described in Sect. 4.4 and Table 4. Target X_{CH_4} computation (red) is described in Sect. 3.1. ΔX is the scheme bias which is the difference between scheme and target X_{CH_4}, and Bias$_{res}$ is the residual bias when evaluated biases have been subtracted from the scheme bias.

In order to assess the performance of averaging schemes and bias correction algorithms, the standard deviation and mean of the difference ΔX between the $\overline{X_{CH_4}}^{scheme}$ estimated from one of the studied averaging schemes and the target value $\overline{X_{CH_4}}^{T}$ must be computed over a set of M simulations. The number of simulation M has to be high enough to compute the residual bias (empirical mean of ΔX) with sufficient accuracy. Let us denote σ the standard deviation of

the distribution of the variable ΔX, $S_M = \langle \Delta X \rangle$ the empirical mean over M samples (i.e. the empirical estimate of the bias of the averaging scheme). To get an estimate of the expected value of ΔX with an accuracy of 0.1 ppb with 90 % confidence, it requires approximately $M = 300\,000$ samples according to the central limit theorem.

The typical standard deviation can also be evaluated from the sample and is approximately 22 ppb for the typical case (mean reflectivity of 0.1).

4.3 Simulation of online and offline lidar calibrated signals and IWF

Once the scene parameters are defined on the 50 km averaging window and the atmosphere is modelled, the online and offline calibrated signals must be simulated. We first have to compute the deterministic values of the calibrated signals without noise and simulate the random noise that affects them. The values of the signals are determined by the scene reflectivity (for both online and offline signals) and by the atmospheric transmission (online signals only). From the weighting functions, the methane field and the pressure field, we compute the reference DAOD, denoted δ_i^{true}, as the numerator of Eq. (5). Then we compare the transmission for each double shot according to

$$\left(\tau_i^{true}\right)^2 = \exp\left(-2 \cdot \delta_i^{true}\right). \tag{38}$$

From them, considering the reflectivity, we are able to determine the relative value of the online and offline mean cali-
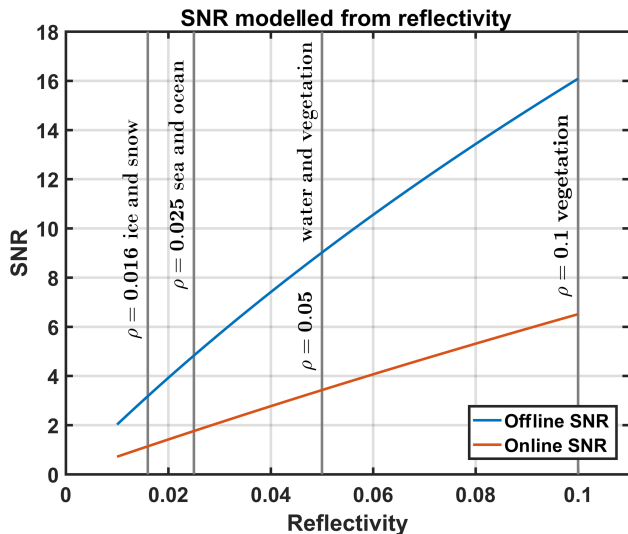
Figure 9. Online and offline SNR computed from reflectivity according to instrument characteristics.

brated signals:

$$\mu_i^{\text{off}} = \rho_i,\tag{39}$$

$$\mu_i^{\text{on}} = \rho_i \cdot \left(\tau_i^{\text{true}}\right)^2,\tag{40}$$

where i is the shot index, τ_i is the transmission and ρ_i is the reflectivity. Note that any constant affecting both online and offline signals can be disregarded here.

Then, Gaussian random noise has to be added to the values of the signals. It is computed from the SNR that depends on the number of photons reaching the detector (i.e. $\mu_i^{\text{on,off}}$). Figure 9 shows the theoretical dependence of the SNR to the reflectivity according to instrument characteristics and Appendix A presents the noise model used.

The IWF_i are simply computed by integrating the $\text{WF}_{i,j}$ on all pressure layers as the denominator of Eq. (5).

4.4 Tested averaging algorithms and bias corrections

The simulation tested the three averaging schemes described in Sect. 3.2: AVX (Table 1, line 1), AVD (Table 1, line 2) and AVS (Table 1, line 3). Table 5 details the computational steps used for averaging, statistical bias evaluation and geophysical bias evaluation for the three schemes. For the AVX and AVD schemes, as explained in Sect. 3.3.1, signal couples with at least one negative calibrated signal must be discarded to compute the shot DAOD. However, since signals are averaged first for the AVS scheme, the probability that one of the averaged signals is negative is extremely small. Thus, no negative calibrated signal discarding is needed for the AVS scheme.

Concerning statistical bias evaluation, an SNR estimation is needed. It is directly estimated from instrument parameters and online and offline calibrated signal strength. Once

the SNR is estimated, as described in Sect. 3.3, there are two options to evaluate the statistical bias either using the Taylor expansion approximation or the numerical integral of a truncated normal distribution. Contrary to AVS, where Taylor expansion and the numerical integral make a negligible difference, for AVX and AVD, it is better to use the numerical integral as it is more accurate, and this is what is done here.

Type 1 geophysical bias, that affects the AVX scheme, is already compensated by weighting the average X_{CH_4} and the average statistical bias by the IWF. The AVD scheme is not affected by geophysical biases. However, type 2 and type 3 geophysical biases affect the AVS scheme. Type 2 geophysical bias is evaluated by Eq. (35) using the first-order X_{CH_4} of Eq. (30). The type 3 geophysical bias is not evaluated and will not be corrected because a correlation between reflectivity and CH_4 concentration is unlikely to occur. Indeed, on a smaller than 50 km scale the typical atmospheric transport should smear out the CH_4 concentration very effectively over areas larger than the small-scale reflectivity jumps even if this is not true for narrow valleys.

5 Results

5.1 Comparison of averaging schemes

5.1.1 Bias of averaging schemes without bias correction

The first results presented here are the respective biases of each averaging scheme without any bias correction. Figure 10 shows the bias on the average X_{CH_4} on the three scenes (Toulouse, Millau and Chamonix) for the three averaging schemes that have been studied (AVS, AVD and AVX) without any correction. The bias due to measurement noise and due to geophysical variation appears on the results, as it is not yet subtracted. For the AVS scheme we compare the results with the uniformly weighted average IWF and the average of the IWF weighted by the offline calibrated signal strength: $w_i \left[Q^{\text{off}} \right]$. For the AVD scheme, a uniform weight is considered. And, for the AVX scheme, we compare the uniformly weighted average X_{CH_4} (Eq. 23) and the average X_{CH_4} weighted by the IWF: $w_i \, [\text{IWF}]$ (Eq. 24). The mean reflectivity is set to the typical value of 0.1.

For the AVS scheme on Toulouse and Millau scenes, where there are medium to high variations of geophysical quantities, the bias is contained in the ± 1 ppb range. However, it is higher on the Chamonix scene, where there are very high variations of geophysical parameters. As expected, the bias for the AVS scheme is mainly impacted by the variations of the geophysical parameters over the scene. Note that on the Chamonix scene, the weighting of the average IWF by the offline calibrated signal strength reduces this bias.
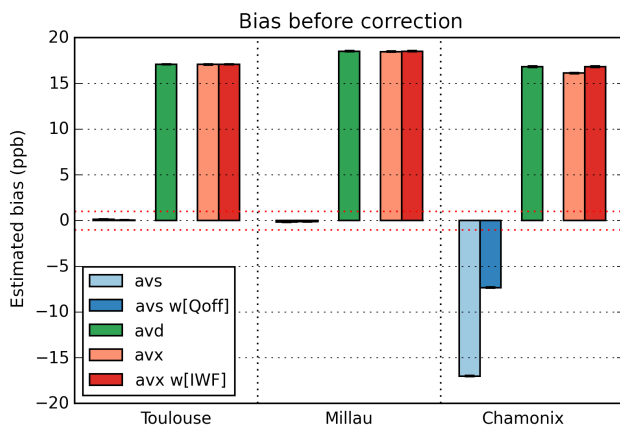
On the contrary, the bias of the AVD and AVX schemes is not affected by the geophysical variations but is mainly

Table 6. Resulting bias (in ppb) for the AVD scheme after noise induced bias correction.

	Taylor bias correction (not usable)				Integral bias correction			
Reflectivity	0.1	0.05	0.025	0.016	0.1	0.05	0.025	0.016
Offline SNR	16.1	9.0	4.8	3.2	16.1	9.0	4.8	3.2
Online SNR	6.5	3.4	1.8	1.1	6.5	3.4	1.8	1.1
Toulouse (ppb)	-6.70×10^{-1}	-9.02×10^{4}	-9.70×10^{10}	-5.11×10^{9}	-1.79	-4.28	207	416
Millau (ppb)	-1.52×10^{2}	-8.94×10^{5}	-6.27×10^{7}	-4.62×10^{8}	-2.43	9.69	204	442
Chamonix (ppb)	-1.19×10^{1}	-1.59×10^{6}	-4.35×10^{9}	-2.44×10^{8}	-2.05	8.24	197	498
Uncertainty (ppb)	$\pm 6.7 \times 10^{1}$	$\pm 8.8 \times 10^{5}$	$\pm 5.5 \times 10^{10}$	$\pm 2.0 \times 10^{9}$	± 0.10	± 0.24	± 0.61	± 0.89

Table 7. Resulting bias (in ppb) for the AVS scheme after noise induced bias correction and geophysical induced bias correction.

	Taylor bias correction				Integral bias correction			
Reflectivity	0.1	0.05	0.025	0.016	0.1	0.05	0.025	0.016
Offline SNR	16.1	9.0	4.8	3.2	16.1	9.0	4.8	3.2
Online SNR	6.5	3.4	1.8	1.1	6.5	3.4	1.8	1.1
Toulouse (ppb)	-0.01	-0.02	-0.03	-0.05	-0.03	-0.07	-0.03	-0.08
Millau (ppb)	-0.03	-0.03	-0.03	-0.05	-0.04	-0.08	-0.04	-0.08
Chamonix (ppb)	-0.51	-0.51	-0.50	-0.48	-0.53	-0.57	-0.50	-0.49
Uncertainty (ppb)	± 0.09	± 0.17	± 0.33	± 0.51	± 0.09	± 0.17	± 0.33	± 0.51

Figure 10. Bias before correction for the three studied averaging schemes (red dotted lines: targeted bias ± 1 ppb).

driven by the measurement noise, which essentially depends on the scene's mean reflectivity. As shown in Sect. 3.4.1, the AVD scheme with uniform weighting and the AVX scheme weighted by the integrated weighting function (w_i[IWF] weights) show the same bias. Although the comparison between the AVX scheme weighted uniformly (light red on Fig. 10) and the AVD scheme (green on Fig. 10) shows that their biases are close when variations of surface pressure (main cause of variations of IWF) are low (Toulouse, Millau), they become significant when variations are higher (Chamonix).

Without any correction and for the typical reflectivity, the AVS scheme is less biased than the AVD and AVX schemes. However, as we have seen in previous sections, there are ways to estimate the biases and to correct them. The following section will show the results after estimation and correction of the bias induced by the measurement noise.

5.1.2 After correction of statistical bias

As explained in Sect. 3.4, the random nature of the measurement associated with the non-linearity of the measurement equation implies that the estimation of the X_{CH_4} is biased. The statistical bias corrections for AVS, AVD and AVX are based on an estimation of the online and offline SNRs for the measured calibrated signals (cf. Sect. 4.4 and Table 5). Figure 11 shows the residual biases, after subtraction of estimated statistical bias, for the three averaging schemes, with and without relevant weightings, on the three studied scenes and for the typical reflectivity of 0.1. The chosen estimation of the bias is done by numerically computing the integral of the truncated Gaussian distribution (Sect. 3.3).

We see that the biases of the AVD and AVX schemes are significantly reduced (absolute value decrease by 85 % to 90 %) on every scene. The residual bias is caused by the fact that the SNRs are estimated from the noisy calibrated signals so that the estimation of the bias is not perfectly accurate. This implies that the calibrated signal outcomes from the lower part of the distribution lead to a high error on the estimated bias. This effect could be slightly compensated if, instead of discarding all the negative or null calibrated signals (extremely rare for a reflectivity value of 0.1 over 150),
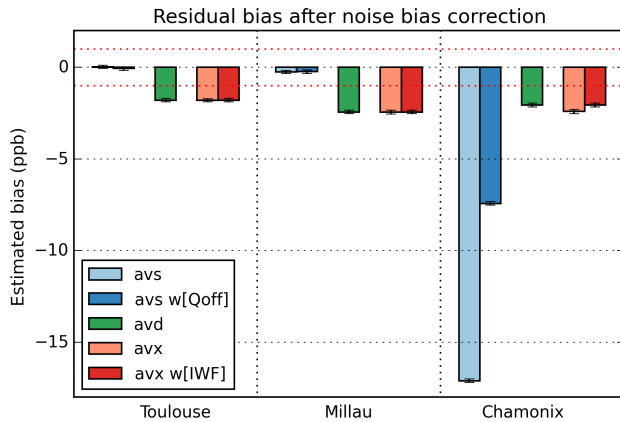
Figure 11. Residual bias after statistical bias correction for the three studied averaging schemes (red dotted lines: targeted bias ±1 ppb).

Figure 12. Residual bias after noise induced bias and geophysical variation induced bias corrections for the three studied averaging schemes (red dotted lines: targeted bias ±1 ppb).

we discarded calibrated signals higher than a strictly positive threshold (e.g. 0.01, not shown). This would lead to a better correction and thus a lower bias, but at the cost of discarding more single-shot observations.

For the AVS scheme, as the signals are averaged first, the equivalent SNR is very high ($SNR_{eq}^{off} \approx 190$ and $SNR_{eq}^{on} \approx 90$) on the scenes with a mean reflectivity of 0.1. Consequently, the bias due to the equivalent measurement noise is really low (about 0.1 ppb), and this bias correction has only a small effect on the residual bias.

Taking into account the correction of the bias induced by the measurement noise, the AVS scheme still presents a lower bias on Toulouse and Millau scenes than the bias of the AVX and AVD schemes. However, on the Chamonix scene, where the geophysical variations are very high, the AVX and AVD schemes are less biased than AVS.

5.1.3 After correction of geophysical biases

The biases induced by the variation of the geophysical parameters (cf. Sect. 3.4) does not affect the AVD scheme, as the additive properties of DAOD and IWF are averaged separately. The variation of the IWF affects the bias of the AVX scheme and has already been corrected by introducing the w_i [IWF] weights when directly averaging mixing ratios. The AVS scheme is the one most affected by the variations of the geophysical variations, as seen in Sect. 5.1.1.

Figure 12 shows the residual bias after the corrections of the statistical bias induced by the measurement noise and the variations of geophysical parameters (cf. Sect. 4.4 and Table 5). We notice that the residual bias for the AVS scheme is considerably reduced when the average weighting function is weighted by the offline calibrated signal strength. Furthermore, the iterative estimation of the bias converges at the first iteration of Eqs. (26) to (31).

Once geophysical biases are subtracted, the three scenes present a low bias. The mean residual bias on the three
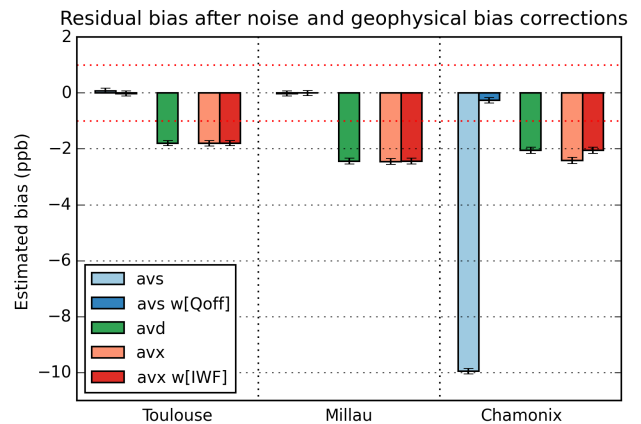
scenes for the AVD and AVX schemes is approximately -2.1 ± 0.1 ppb, whereas for the AVS scheme, it is approximately -0.09 ± 0.09 ppb. After all corrections, even on highly structured scenes, AVS is the least biased scheme of the three studied schemes. When the average IWF is consistently weighted with the $w_i [Q^{off}]$ weights, the geophysical induced bias is almost completely removed.

5.2 Impact of the mean reflectivity on the residual bias

All results presented above are computed for scenes with a mean reflectivity of 0.1, which roughly corresponds to vegetation cover. For the purpose of choosing the least biased algorithm to compute average X_{CH_4}, it is interesting to test the robustness to reflectivity. Indeed, reflectivity is the main driver for the expected value of SNR; low reflectivity scenes lead to lower SNR and consequently higher bias. Tables 6 and 7 show the residual bias comparing four different reflectivity values: 0.1 (vegetation), 0.05 (mixed sea and vegetation), 0.025 (sea and ocean) and 0.016 (ice and snow). Table 6 gives the residual bias for the AVD scheme (the AVX scheme presents similar results), and Table 7 shows the residual bias for the AVS scheme, where the average IWF is weighted by the offline calibrated signal strength ($w_i [Q^{off}]$ weights) and corrected of the bias due to geophysical variations from shot to shot. For both tables, the Taylor expansion bias correction (Eq. 15) and numerical computation of the expectation (so-called integral truncated normal distribution Eqs. 16 and 13) are compared.

First, as seen in Table 6 (AVD scheme), the Taylor bias correction does not succeed in quantifying the bias on any of the four mean reflectivity values. The uncertainties are too high and prevent quantitative analysis of the results. This is due to the fact that there are some calibrated signals that are really close to zero and for which the SNR is underestimated; thus the bias (and standard deviation) is overestimated. This

Table 8. Bias (in ppb) of several averaging biases before any bias correction schemes are applied on the three scenes for four reflectivity values.

	AVX	AVD	AVS	AVQ (on–off)	AVQ (off–on)
Toulouse ($\rho = 0.1$)	17.09	17.09	0.15	−5.25	40.71
Millau ($\rho = 0.1$)	18.50	18.54	−0.14	−6.40	46.53
Chamonix ($\rho = 0.1$)	16.65	17.36	−17.00	−14.74	53.41
Toulouse ($\rho = 0.05$)	71.30	71.30	0.42	−17.18	219.92
Millau ($\rho = 0.05$)	72.63	72.67	0.10	−23.60	268.72
Chamonix ($\rho = 0.05$)	65.77	67.28	−16.80	−31.93	259.48
Toulouse ($\rho = 0.025$)	173.34	173.34	1.46	−122.17	1055.46
Millau ($\rho = 0.025$)	173.31	173.29	1.05	−115.22	1022.00
Chamonix ($\rho = 0.025$)	189.46	189.83	−16.03	−92.89	1025.43
Toulouse ($\rho = 0.016$)	144.79	144.77	3.44	−273.53	1418.72
Millau ($\rho = 0.016$)	184.98	184.84	2.87	−232.10	1457.83
Chamonix ($\rho = 0.016$)	274.46	269.75	−14.56	−155.45	1596.10

could be mitigated by the choice of a higher threshold of the usable calibrated signal before the computation of the DAOD (not shown). The results when using the integral bias correction on AVD are more physical. However, they also show an over estimation of the bias, especially for low reflectivity values. In every case for the AVD scheme, the bias threshold is exceeded.

Table 7 gives the results of the robustness of the AVS scheme to decreasing reflectivity. Unlike the AVD scheme, the AVS scheme, when all corrections are made, presents satisfying results for all reflectivity values, and in every scene the biases remain contained into the threshold interval of ±1 ppb. The effect of the decreasing reflectivity has a very small impact on the residual bias.

To summarize, the best algorithm to limit the bias for MERLIN processing algorithms is clearly the AVS scheme, with an average IWF weighted by the offline calibrated signal strength and both corrections of the geophysical bias and the bias induced by the measurement noise (either Taylor or integral bias correction). On every scene and for all expected reflectivity values, this algorithm is compliant with the averaging bias specifications of the MERLIN mission. Note that this conclusion holds in the case where all the 150 shots are considered; in the case of a partially cloudy window where only a subsample of clear sky shots are averaged, the AVS will still be the best averaging scheme, but the performance will be decreased.

6   Conclusions

The French–German space-borne IPDA lidar mission MER-LIN will measure the average integrated column dry-air mixing ratio of methane (X_{CH_4}) on a 50 km scale. The processing algorithms must limit both the relative random error (RRE) and the relative systematic error (RSE) on the X_{CH_4}. As the IPDA technique relates the signal measurements to the X_{CH_4}

by a non-linear equation, a simple and naive averaging can lead to high biases.

Three averaging schemes have been studied: averaging of X_{CH_4} (AVX), averaging of DAOD (AVD) and averaging of signals (AVS). For these averaging schemes, possible sources of bias can either be due to the measurement noise, the variation of the geophysical parameters on the averaging scene or both.

The three schemes are sensitive to the bias induced by the measurement noise even if AVS is far less impacted for the typical reflectivity. This bias can be corrected by a formula introducing the estimated SNR on the measured signals if the SNR is high enough. The bias due to the variation of geophysical parameters does not affect the AVD scheme because it directly averages the desired additive quantities. On the contrary, the AVX scheme must average the concentration weighted by the integrated weighting function (IWF) in order to average a molecule number instead of averaging concentrations. The third scheme AVS measures the average X_{CH_4} weighted by the offline signal strength, which means that more weight to the measurements with a high SNR is given when averaging. The bias of this scheme is sensitive to the variation of geophysical parameters (surface pressure and surface reflectivity). This bias is corrected using an iterative process with the uncorrected X_{CH_4} as first guess.

These averaging schemes and their bias corrections have been tested on scenes modelled from real satellite data in terms of altitude, surface pressure, weighting functions and relative variations of reflectivity. The three scenes present interesting characteristics, as they show different geophysical variations that could impact averaging biases. Besides, the signals and random noise are simulated from geophysical parameters and instrument parameters.

The simulation shows that the lowest biases are obtained for the AVS scheme using appropriate bias corrections and averaging weights. Furthermore, this scheme is robust to low reflectivity values unlike the AVX and AVD schemes, which

are highly sensitive to the accuracy of the SNR estimation. The best scheme, AVS, is compliant with the allocated averaging bias requirements (RSE) of 0.06 % (1 ppb for a X_{CH_4} of 1780 ppb) for the whole range of expected reflectivity values (from 0.1 down to 0.016).

A continuation of this study could evaluate the sensitivity of a poor (unprecise or biased) estimation of the SNR on the estimation of the bias due to measurement noise for low reflectivity values. Furthermore, the use of the lidar simulator and processor suites, currently in development at the LMD, could be beneficial to the evaluation of the biases, and more specifically of the averaging biases, on a wider scale (many scenes, atmosphere types, etc.).

Data availability. SPOT-5 data can be accessed at https://earth.esa.int/web/guest/data-access (last access: 22 October 2018). SRTM data can be accessed at http://srtm.csi.cgiar.org/ (last access: 22 October 2018). ECMWF data can be accessed at https://www.ecmwf.int/en/forecasts/datasets (last access: 22 October 2018).

Appendix A: Signal generation and noise model

The simulation of calibrated signals requires a noise model. The signal distribution is considered to be Gaussian first, because the number of photons that reaches the photodetector is high enough for the Poisson distribution to be considered as Gaussian. Secondly, the system is limited by the detector noise that is mainly thermal noise, which is normally distributed.

The calibrated signals are produced using a pseudorandom number generator. The expected values of the calibrated signal distributions, $\mu^{on,off}$, depend on the atmospheric transmission and the surface reflectivity as presented in Eqs. (39) and (40). Then the standard deviation $\sigma^{on,off}$ is deduced from the SNR, which is defined as

$$\text{SNR}^{on,off} = \frac{\mu^{on,off}}{\sigma^{on,off}}. \tag{A1}$$

And the SNR model is described by

$$\text{SNR}^{on,off} = \left(\frac{N^2}{a + b \cdot N + c \cdot N^2} \right)^{\frac{1}{2}}, \tag{A2}$$

where N is the number of photoelectrons, and a, b and c are parameters computed from the MERLIN system parameters. This is illustrated on Fig. 9.

The number of photons is computed from the reflectivity and the atmospheric transmission. In the standard case (reflectivity of $0.1\,\text{sr}^{-1}$), its values are approximately 3000 for the offline pulse and 1000 for the online pulse.

The first term of the denominator corresponds to the detector noise, the second to the shot noise and the third to the speckle. Note that the speckle term has been neglected in this article, whereas both detector noise and shot noise have been considered, as they are dominant compared to speckle noise.

Appendix B: Averaging of quotients

The averaging of quotients estimates the average of the shot-by-shot two-way transmissions τ_i^2. Due to the measurement noise, we will suppose that, for the shot i, the online and offline measured calibrated signals, Q_i^{on} and Q_i^{off}, respectively, are outcomes of normal distributions with mean values denoted μ_i^{on} and μ_i^{off}, respectively, and standard deviations denoted σ_i^{on} and σ_i^{off}, respectively. The X_{CH_4} computed from averaging quotients can be defined as

$$\overline{X_{CH_4}}^{avq} = \frac{-\frac{1}{2} \cdot \ln \langle \tau^2 \rangle}{\langle \text{IWF} \rangle}, \tag{B1}$$

with the transmission defined as

$$\langle \tau^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{Q_i^{on}}{Q_i^{off}}. \tag{B2}$$

If we define the standardized signals corresponding to Q_i^{on} and Q_i^{off} as X_i^{on} and X_i^{off}, the average transmission can be written as

$$\langle \tau^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{\mu_i^{on} + \sigma_i^{on} \cdot X_i^{on}}{\mu_i^{off} + \sigma_i^{off} \cdot X_i^{off}}. \tag{B3}$$

Then we can further separate the random part due to measurement noise and the deterministic part due to varying geophysical parameters as follows:

$$\langle \tau^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\mu_i^{on}}{\mu_i^{off}} \left(1 + \frac{\sigma_i^{on}}{\mu_i^{on}} X_i^{on} \right) \right. $$
$$\left. \left(1 + \frac{\sigma_i^{off}}{\mu_i^{off}} X_i^{off} \right)^{-1} \right], \tag{B4}$$

$$\langle \tau^2 \rangle = e^{-2\langle \delta^{true} \rangle} \cdot \frac{1}{N} \sum_{i=1}^{N} \left[e^{-2\Delta \delta_i} \left(1 + \frac{\sigma_i^{on}}{\mu_i^{on}} X_i^{on} \right) \right. $$
$$\left. \left(1 + \frac{\sigma_i^{off}}{\mu_i^{off}} X_i^{off} \right)^{-1} \right], \tag{B5}$$

where $\langle \delta^{true} \rangle$ is the average DAOD computed from noiseless mean signals, and $\Delta \delta_i$ the difference compared to the shot-by-shot DAOD. Then we can deduce the error of AVQ scheme as follows:

$$\overline{X_{CH_4}}^{avq} = X_{CH_4}^{true} - \frac{1}{2\langle \text{IWF} \rangle} \cdot \ln \left(\frac{1}{N} \sum_{i=1}^{N} \left[e^{-2\Delta \delta_i} \right. \right. $$
$$\left. \left. \left(1 + \frac{\sigma_i^{on}}{\mu_i^{on}} X_i^{on} \right) \left(1 + \frac{\sigma_i^{off}}{\mu_i^{off}} X_i^{off} \right)^{-1} \right] \right). \tag{B6}$$

The corresponding bias is the expected value of the error term:

$$\text{Bias} \left(\overline{X_{CH_4}}^{avq} \right) = -\frac{1}{2\langle \text{IWF} \rangle} \cdot E \left[\ln \left(\frac{1}{N} \sum_{i=1}^{N} \right. \right. $$
$$\left. \left. \left[e^{-2\Delta \delta_i} \left(1 + \frac{\sigma_i^{on}}{\mu_i^{on}} X_i^{on} \right) \left(1 + \frac{\sigma_i^{off}}{\mu_i^{off}} X_i^{off} \right)^{-1} \right] \right) \right]. \tag{B7}$$

In Eq. (B5), the empirical average transmission is decomposed into two factors. The first is the transmission corresponding to the average DAOD from noiseless signals. The second factor is the average of multiplicative errors that is the deterministic error from geophysical variations on the averaging scene and the random factors due to the presence of measurement noise. As shown in Eq. (B7), the error sources are mixed into the non-linear function, which makes them difficult to evaluate. It is possible to derive a suitable bias correction based on Eq. (B7) for AVQ, but in the end it is not expected to be better than the other ones. Table 8 presents the biases for every averaging scheme presented before any bias correction is applied.

References

Bösenberg, J.: Ground-based differential absorption lidar for water-vapor and temperature profiling: methodology, Appl. Optics, 37, 3845–3860, https://doi.org/10.1364/AO.37.003845, 1998.

Chéruy, F., Scott, N. A., Armante, R., Tournier, B., and Chedin, A.: Contribution to the development of radiative transfer models for high spectral resolution observations in the infrared, J. Quant. Spectrosc. Ra., 53, 597–611, https://doi.org/10.1016/0022-4073(95)00026-H, 1995.

Chevallier, F., Chédin, A., Chéruy, F., and Morcrette, J.-J.: TIGR-like atmospheric-profile databases for accurate radiative-flux computation, Q. J. Roy. Meteor. Soc., 126, 777–785, https://doi.org/10.1002/qj.49712656319, 2000.

Chevallier, F., Broquet, G., Pierangelo, C., and Crisp, D.: Probabilistic global maps of the CO_2 column at daily and monthly scales from sparse satellite measurements, J. Geophys. Res., 122, 7614–7629, https://doi.org/10.1002/2017JD026453, 2017.

Ehret, G., Kiemle, C., Wirth, M., Amediek, A., Fix, A., and Houweling, S.: Space-borne remote sensing of $CO_2, CH_4,$ and N_2O by integrated path differential absorption lidar: a sensitivity analysis, Appl. Phys. B.-Lasers O., 90, 593–608, https://doi.org/10.1007/s00340-007-2892-3, 2008.

Ehret, G., Bousquet, P., Pierangelo, C., Alpers, M., Millet, B., Abshire, J. B., Bovensmann, H., Burrows, J. P., Chevallier, F., Ciais, P., Crevoisier, C., Fix, A., Flamant, P., Frankenberg, C., Gibert, F., Heim, B., Heimann, M., Houweling, S., Hubberten, H. W., Jöckel, P., Law, K., Löw, A., Marshall, J., Agusti-Panareda, A., Payan, S., Prigent, C., Rairoux, P., Sachs, T., Scholze, M., and Wirth, M.: MERLIN: A French-German Space Lidar Mission Dedicated to Atmospheric Methane, Remote Sens.-Basel, 9, 1052, https://doi.org/10.3390/rs9101052, 2017.

Gleyzes, J.-P., Meygret, A., Fratter, C., Panem, C., Ballarin, S., and Valorge, C.: SPOT5: System overview and image ground

segment, Proceedings of IGARSS 2003, Toulouse, France, https://doi.org/10.1109/IGARSS.2003.1293756, 21–25 July, 2003.

Grant, W. B., Brothers, A. M., and Bogan, J. R.: Differential absorption lidar signal averaging, Appl. Optics, 27, 1934–1938, https://doi.org/10.1364/AO.27.001934, 1988.

IPCC: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., https://doi.org/10.1017/CBO9781107415324, 2013

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, CGIAR Consortium for Spatial Information (CGIAR-CSI), available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org, last access: 22 October 2018),CGIAR, Washington, United States, 2008.

Johnson, N. L., Kotz, S., and Balakrishnan, N.: Continuous Univariate Distributions, 1, 2nd edn., 156–162, 1994.

Jucks, K. W., Neeck, S., Abshire, J. B., Baker, D. F., Browell, E. V., Chatterjee, A., Crisp, D., Crowell, S. M., Denning, S., Hammerling, D., Harrison, F., Hyon, J. J., Kawa, S. R., Lin, B., Meadows, B. L., Menzies, R. T., Michalak, A., Moore, B., Murray, K. E., Ott, L. E., Rayner, P., Rodriguez, O. I., Schuh, A., Shiga, Y., Spiers, G. D., Shih Wang, J., and Zaccheo, T. S.: Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (ASCENDS) Mission (white paper), ASCENDS 2015 Workshop, available at: https://cce.nasa.gov/ascends_2015/ASCENDS_FinalDraft_4_27_15.pdf (last access: 22 October 2018), 2015.

Kiemle, C., Quatrevalet, M., Ehret, G., Amediek, A., Fix, A., and Wirth, M.: Sensitivity studies for a space-based methane lidar mission, Atmos. Meas. Tech., 4, 2195–2211, https://doi.org/10.5194/amt-4-2195-2011, 2011.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith P., Castaldi, S., Chevallier F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quéré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G.: Three decades of global methane sources and sinks, Nat. Geosci., 6, 813–823, https://doi.org/10.1038/ngeo1955, 2013.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., Janssens-Maenhout, G., Tubiello, F. N., Castaldi, S., Jackson, R. B., Alexe, M., Arora, V. K., Beerling, D. J., Bergamaschi, P., Blake, D. R., Brailsford, G., Brovkin, V., Bruhwiler, L., Crevoisier, C., Crill, P., Covey, K., Curry, C., Frankenberg, C., Gedney, N., Höglund-Isaksson, L., Ishizawa, M., Ito, A., Joos, F., Kim, H.-S., Kleinen, T., Krummel, P., Lamarque, J.-F., Langenfelds, R., Locatelli, R., Machida, T., Maksyutov, S., McDonald, K. C., Marshall, J., Melton, J. R., Morino, I., Naik, V., O'Doherty, S., Parmentier, F.-J. W., Patra, P. K., Peng, C., Peng, S., Peters, G. P., Pison, I., Prigent, C., Prinn, R., Ramonet, M., Riley, W.

J., Saito, M., Santini, M., Schroeder, R., Simpson, I. J., Spahni, R., Steele, P., Takizawa, A., Thornton, B. F., Tian, H., Tohjima, Y., Viovy, N., Voulgarakis, A., van Weele, M., van der Werf, G. R., Weiss, R., Wiedinmyer, C., Wilton, D. J., Wiltshire, A., Worthy, D., Wunch, D., Xu, X., Yoshida, Y., Zhang, B., Zhang, Z., and Zhu, Q.: The global methane budget 2000–2012, Earth Syst. Sci. Data, 8, 697–751, https://doi.org/10.5194/essd-8-697-2016, 2016.

Scott, N. A. and Chédin, A.: A fast line-by-line method for atmospheric absorption computations: The Automatized Atmospheric Absorption Atlas, J. Appl. Meteorol., 20, 802–812, https://doi.org/10.1175/1520-0450(1981)020<0802:AFLBLM>2.0.CO;2, 1981.

Werle, P., Miicke, R., and Slemr, F.: The Limits of Signal Averaging in Atmospheric Trace-Gas Monitoring by Tunable Diode-Laser Absorption Spectroscopy (TDLAS), Appl. Phys. B-Lasers O., 57, 131–139, https://doi.org/10.1007/BF00425997, 1993.