Atmos. Meas. Tech., 18, 241–264, 2025
https://doi.org/10.5194/amt-18-241-2025
© Author(s) 2025. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Measurement Techniques
Open Access
EGU

Retrieving the atmospheric concentrations of carbon dioxide and methane from the European Copernicus CO2M satellite mission using artificial neural networks

Maximilian Reuter[1], Michael Hilker[1], Stefan Noël[1], Antonio Di Noia[1], Michael Weimer[1], Oliver Schneising[1], Michael Buchwitz[1], Heinrich Bovensmann[1], John P. Burrows[1], Hartmut Bösch[1], and Ruediger Lang[2]

[1]Institute of Environmental Physics, University of Bremen, FB 1, P.O. Box 330440, 28334 Bremen, Germany
[2]EUMETSAT, Eumetsat Allee 1, 64295 Darmstadt, Germany

Correspondence: Maximilian Reuter (mail@maxreuter.org)

Received: 25 July 2024 – Discussion started: 2 August 2024
Revised: 15 October 2024 – Accepted: 9 November 2024 – Published: 15 January 2025

Abstract. Carbon dioxide (CO_2) and methane (CH_4) are the most important anthropogenic greenhouse gases and the main drivers of climate change. Monitoring their concentrations from space helps detect and quantify anthropogenic emissions, supporting the mitigation efforts urgently needed to meet the primary objective of the Paris Agreement, adopted at the 21st Conference of the Parties to the United Nations Framework Convention on Climate Change (UNFCCC) in 2015, to limit the global average temperature increase to well below 2 °C above pre-industrial levels. In addition, satellite observations can be used to quantify natural sources and sinks, improving our understanding of the carbon cycle. Advancing these goals is one key motivation for the European Copernicus CO_2 monitoring mission CO2M. The necessary accuracy and precision requirements for the measured quantities XCO_2 and XCH_4 (the column-averaged dry-air mole fractions of CO_2 and CH_4) are demanding. According to the CO2M mission requirements, the spatial and temporal variability of the systematic errors (or spatio-temporal systematic errors) of XCO_2 and XCH_4 must not exceed 0.5 ppm and 5 ppb, respectively. The stochastic errors due to instrument noise must not exceed 0.7 ppm for XCO_2 and 10 ppb for XCH_4. Conventional so-called full-physics algorithms for retrieving XCO_2 and/or XCH_4 from satellite-based measurements of reflected solar radiation are typically computationally intensive and still usually require empirical bias corrections based on supervised machine learning methods. Here we present the retrieval algorithm Neural networks for Remote sensing of Greenhouse gases from CO2M (NRG-CO2M), which derives XCO_2 and XCH_4 from CO2M radiance measurements with minimal computational effort using artificial neural networks (ANNs). In addition, NRG-CO2M also provides estimates of both the noise-driven uncertainties and the averaging kernels of XCO_2 and XCH_4 for each sounding. Since CO2M will not be launched until 2026, our study exploits simulated measurements over land surfaces from a comprehensive observing system simulation experiment (OSSE) that includes realistic meteorology, aerosols, surface bidirectional reflectance distribution function (BRDF), solar-induced chlorophyll fluorescence (SIF), and CO_2 and CH_4 concentrations. We created a novel hybrid learning approach that combines advantages of simulation-based and measurement-based training data to ensure coverage of a wide range of XCO_2 and XCH_4 values, making the training data representative of future concentrations as well. The algorithm's postprocessing is designed to achieve a high data yield of about 80 % of all cloud-free soundings. The spatio-temporal systematic errors of XCO_2 and XCH_4 are 0.44 ppm and 2.45 ppb, respectively. The average single sounding precision is 0.41 ppm for XCO_2 and 2.74 ppb for XCH_4. Therefore, the presented retrieval method has the potential to meet the demanding CO2M mission requirements for XCO_2 and XCH_4. While the presented results are a solid proof of concept, the actual achievable quality can only be determined once NRG-CO2M is trained on real data, where it is confronted, e.g., with unknown instrument effects and systematic errors in the training truth.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

Carbon dioxide (CO_2) and methane (CH_4) are the most important anthropogenic greenhouse gases because they are the main drivers of climate change. Monitoring their concentrations from space is essential to identify and quantify anthropogenic emissions, thereby supporting the mitigation efforts needed to achieve the primary objective of the Paris Agreement of the United Nations Framework Convention on Climate Change (UNFCCC) to limit the global average temperature increase to well below 2 °C above pre-industrial levels (UNFCCC, 2015). In addition, satellite observations can be used to study natural sources and sinks of these gases, contributing to a better understanding of the carbon cycle and thus improving climate predictions.

Advancing these goals is the motivation for the European Copernicus CO_2 monitoring mission CO2M (Meijer et al., 2020; Lespinas et al., 2020; Sierk et al., 2021), which will serve as a central element of the monitoring and verification support (MVS) service capacity currently being developed as an integral part of the Copernicus Atmosphere Monitoring Service (CAMS). The mission involves the deployment of a constellation of three satellites, with the launch of the first CO2M satellite planned for 2026. CO2M builds on the heritage of the CarbonSat concept (Bovensmann et al., 2010; Velazco et al., 2011; Buchwitz et al., 2013; Broquet et al., 2018).

The accuracy and precision requirements for the measured quantities XCO_2 and XCH_4 (the column-averaged dry-air mole fractions of CO_2 and CH_4) are demanding, and achieving them is a major challenge. Specifically, the mission requirement document (MRD; Meijer et al., 2020) states that the systematic errors of XCO_2 and XCH_4 should not exceed a maximum spatial and temporal variability of 0.5 ppm and 5 ppb, respectively. The stochastic errors due to instrument noise should not exceed 0.7 ppm for XCO_2 and 10 ppb for XCH_4 for a reference scenario over vegetation. This is why CO2M is equipped not only with the main CO_2 imager (CO2I) instrument, comprising four imaging spectrometers, but also with the Multi-Angle Polarimeter (MAP), which helps to better account for light scattering on aerosols and the surface bidirectional reflectance distribution function (BRDF), and the Cloud Imager (CLIM), which helps to identify clouds in the field of view.

Conventional so-called full-physics algorithms for retrieving XCO_2 and/or XCH_4 (XGAS) from satellite-based measurements of reflected solar radiation in the near-infrared (NIR) and shortwave-infrared (SWIR) spectral region require accurate radiative transfer (RT) and instrument simulations, which are typically computationally expensive. Examples of such retrieval methods are described in the publications of Reuter et al. (2010, 2011, 2017a, b), Boesch and Di Noia (2023), Noël et al. (2021, 2022), Kiel et al. (2019), Guerlet et al. (2013), and Cogan et al. (2012). Three full-physics algorithms for the analysis of CO2M data are currently also being implemented in the EUMETSAT ground segment. One of these methods is the Fast atmOspheric traCe gAs retrievaL (FOCAL; Noël et al., 2024). The others are RemoTAP (Lu et al., 2022) and FUSIONAL-P, a further development of the algorithm described by Boesch and Di Noia (2023). It is anticipated that continuous analysis of the data stream from a single CO2M satellite using these three methods will require the computing power of several thousand CPU cores, and reprocessing the data from two or more CO2M satellites will require several times that amount.

Despite the high computing power required, there are still a number of reasons that can lead to more or less large systematic inaccuracies in the retrieved XGAS quantities. Examples are simplifications of the RT (e.g., neglect of Raman scattering, neglect of polarization, a suboptimal number of streams, reduced accuracy of scattering phase functions, 3D effects), which are necessary to keep the computation time within acceptable limits. In addition, insufficiently characterized geophysical input parameters (e.g., spectroscopic parameters, aerosol and cloud microphysical properties, surface BRDF, all kinds of subpixel inhomogeneities) and insufficiently characterized instrument properties (e.g., inadequate stray light correction, crosstalk or sensor nonlinearity) can lead to biases.

For these reasons, currently existing full-physics retrievals typically exploit more or less complex empirical bias corrections in order to meet the demanding accuracy requirements. This applies to established methods for instruments/satellites, such as the Orbiting Carbon Observatory-2 (OCO-2; Crisp et al., 2004), the Greenhouse Gases Observing Satellite (GOSAT; Kuze et al., 2009), GOSAT-2 (Suto et al., 2021), and the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY; Burrows et al., 1995; Bovensmann et al., 1999) (Reuter et al., 2017b, a; Kiel et al., 2019; Noël et al., 2021, 2022; Boesch and Di Noia, 2023; Guerlet et al., 2013; Cogan et al., 2012; Schneising et al., 2013, 2014), and it is not unlikely that the same will apply to the CO2M XGAS retrieval algorithms currently being implemented by EUMETSAT, once they are confronted with actual measurements.

The variance of the bias correction can be of the same order of magnitude as the retrieval increment, i.e., the difference between a priori knowledge and the result (Reuter et al., 2017a; Kiel et al., 2019), implying that the bias correction contributes a non-negligible fraction of the information of the result.

Most bias correction methods are empirical and usually exploit supervised machine learning techniques. These include simple multidimensional linear regressions (Kiel et al., 2019) or more complex methods based on, e.g., random forest regressors (Noël et al., 2022; Schneising et al., 2019, 2023). For this reason, they also face the issues associated with data-driven methods, such as the need for a representative training data set including ground truth.

Consequently one motivation for this study is to try to avoid the complicated and computationally intensive step of full-physics algorithms and instead analyze the measured spectra from the outset using a data-driven method. Multilayer perceptrons (MLPs) are artificial neural networks (ANNs) that are well suited for this task and, once trained, can analyze large amounts of data with minimal computational effort. An MLP is a nonlinear function whose parameters are adjusted during training to best map the input features (e.g., spectra, meteorological profiles, observation angles) to the output target (e.g., XCO_2, XCH_4). This is called supervised learning, and it requires a representative set of input features for which one or more known output target variables exist. The principle of the method is analogous to that of linear regression, which is one of the simplest forms of supervised learning. A general introduction to MLPs can be found, e.g., in the textbook of Rojas (1996).

As is known from other regressors with many free fit parameters, MLPs tend to be good interpolators but poor extrapolators (Krasnopolsky and Schiller, 2003). This is particularly relevant because CO_2 and CH_4 increase over time, and a training data set consisting of today's measurements is not representative of the future. Furthermore, MLPs can learn from spurious correlations just as efficiently as from actual physical relationships; i.e., they can give significant weight to correlations between input and target that are not directly caused by a physical relationship but by factors such as similar seasonality (e.g., XCO_2 and solar zenith angle). However, generalized learning occurs only in the latter case, and applying the MLP to unknown scenarios leads to accurate data products only in this case. Another potential hurdle is that MLPs can be affected by uncertainties in the training target. Consequently, it is necessary to ensure that the training data set is representative of current and future conditions and that the training target is not too far from the truth.

One possible solution to obtain representative training data is to generate the training data set from simulated measurements. This simulation-based approach is followed with the nonlinear inference scheme (NLIS) algorithm developed by Crevoisier (2023) for the retrieval of midtropospheric CO_2 and CH_4 columns from Infrared Atmospheric Sounding Interferometer (IASI) and Atmospheric InfraRed Sounder (AIRS) measurements in the thermal infrared spectral region. It is also applied by Xie et al. (2024) to retrieve XCO_2 from OCO-2 measurements over east Asia. However, building the training data set from simulations does not only have advantages. For the reasons discussed above, there are usually differences between simulated and measured spectra that cannot be explained by instrument noise. As with full-physics methods, these can affect the quality of the data products and again may require empirical bias correction.

A different approach was taken by David et al. (2021), who trained an MLP to retrieve XCO_2 using actual measured OCO-2 data. This measurement-based learning has the potential advantage of virtually eliminating many of the sources of systematic errors discussed above. However, it turned out that their ANN also appeared to have learned from spurious correlations, as it was unable to detect known local increases. After modifying the ANN and its input, Bréon et al. (2022) were able to show that their ANN was now able to detect local enhancements that were not part of the training data set. However, the authors also state that their ANN is not suitable for analyzing future data due to increasing CO_2 concentrations. In addition, they emphasize that despite the promising results, it is difficult to ensure that their ANN does not learn from a spurious correlation again, especially since the reasons for the previous failure could not be fully determined.

In this paper we present the Neural networks for Remote sensing of Greenhouse gases from CO2M (NRG-CO2M) algorithm, which allows the use of actual measured spectra for training, but they are modified to cover a much larger range of XCO_2 and XCH_4 values. This type of hybrid learning combines the advantages of simulation-based and measurement-based learning. The characteristics of the actual measured spectra, including potential instrument effects, are preserved, almost any meaningful CO_2 and CH_4 concentration can be trained, and the variability of the training truth is dominated by prescribed artificial variations which can suppress learning from spurious correlations.

Nevertheless, our method also requires estimates of the true atmospheric concentrations to provide a representative training data set. These could be obtained in the same way as for empirical bias corrections (e.g., Noël et al., 2022; Schneising et al., 2023; Kiel et al., 2019) or as for the training data sets of other measurement-based ML methods (Bréon et al., 2022). Since CO2M will not be launched until 2026, for the time being our study uses simulated measurements from an extensive observing system simulation experiment (OSSE), which is a refinement of the OSSE described by Noël et al. (2024). As we are dealing with simulations, the true concentrations are known, and, similar to Noël et al. (2024), we assume that there are no systematic errors in the training truth. Obviously, such errors would have the potential to reduce the accuracy of the prediction, but a realistic estimate of the to-be-expected error patterns of the training truth is difficult and beyond the scope of this study. We do, however, allow for stochastic deviations of the training data from the true concentrations.

Section 2 describes the data sets and methods used, including the OSSE; the hybrid learning method; the transformation of the input data using principal component analysis (PCA); the method to modify the spectra; and the setup and training of the MLPs to determine XCO_2, XCH_4, and the corresponding uncertainties. Section 3 presents the results of the study, and Sect. 4 provides a summary and conclusions.

2 Data sets and methods

2.1 Observing system simulation experiment

A comprehensive observing system simulation experiment (OSSE) was performed as part of a EUMETSAT study to develop the FOCAL CO2M retrieval algorithm (Noël et al., 2024). It contains simulated CO2M radiance data for nadir-mode measurements over land, generated with the SCIATRAN RT model (Rozanov et al., 2017), taking into account realistic meteorology, surface BRDF, solar-induced chlorophyll fluorescence (SIF), aerosols, clouds, and vertical profiles of CO_2 and CH_4. The data set includes 2 years (2015 and 2020) of simulated CO2M orbit data with reduced sampling, hereafter referred to as subset data, as well as a high-resolution (HR) scene simulated with the full CO2M sampling. It is an updated and extended version of the data set of simulated CO2I measurements used by Noël et al. (2024) and is therefore only briefly described here. The data set includes 2 years of subset data instead of one, and in addition to the simulated CO2I measurements, it has been extended to also contain simulated measurements from the MAP and CLIM instruments. In addition, the spectral variation in the surface BRDF within the instrument bands is now more realistic and no longer constant.

The exact instrument characteristics of CO2M were not fully defined at the time of our study, so we used the MRD as a guide. The simulated main instrument CO2I consists of four imaging spectrometers for the wavelength ranges 405–490 nm (VIS, NO_2), 747–773 nm (NIR, O_2), 1590–1675 nm (SWIR-1, CO_2, and CH_4), and 1990–2095 nm (SWIR-2, CO_2), having spectral resolutions of 0.6, 0.12, 0.3, and 0.35 nm, respectively. In line with currently available information about CO2I, the instrument line shape functions are assumed to be Gaussian with full width at half maximum, corresponding to the respective spectral resolution.

In this study, we use CO2I data from the entire NIR band (1930 spectral features) and from the same wavelength ranges as used by Noël et al. (2024) in the SWIR-1 band (1590–1670 nm, 931 spectral features) and SWIR-2 band (1990–2090 nm, 953 spectral features). The VIS band is mainly intended for the determination of NO_2 atmospheric columns and was therefore not simulated in this study.

For the hypothetical MAP instrument, we assumed that it has seven broadband channels (MAP1–7) with center wavelengths of 410, 443, 490, 555, 670, 760, and 865 nm, within which it determines the Stokes parameters I, Q, and U for each CO2I ground pixel at 45 equidistantly distributed along-track observation angles. In reality, MAP will have a higher spatial resolution, which will be aggregated to the CO2I measurements, and its MAP6 channel will only measure intensity, but this is not taken into account in this study.

The simulated CLIM instrument has three broadband channels (CLIM1–3), the first two of which spectrally coincide with MAP5 and MAP6. The central wavelength of CLIM3 is at 1370 nm in a strong absorption band of water vapor, which makes this channel suitable for the identification of cirrus clouds. In reality, CLIM will have a much higher spatial resolution than CO2I, but this is also not taken into account in this study, so only CLIM3 provides additional information here.

An example of the complete simulated radiance measurements of a CO2I sounding, including co-located MAP and CLIM radiance measurements, is shown in Fig. 1.

2.1.1 Subset data

CO2I will have ground pixels with a spatial resolution of approximately 2 km × 2 km and 110 ground pixels per scan line across track, and each orbit will comprise approximately 9200 daytime scan lines along track. In order to create representative training, test, and evaluation data sets, a minimum of 2 years of simulated CO2M data were desired. However, accurate RT simulations are computationally expensive, so it was not possible to simulate that many soundings in a reasonable amount of time. Consequently, we adopted a strategy of subsetting the data set by simulating only every 15th ground pixel across track and every 20th ground pixel along track. This approach reduced the computational cost by a factor of 300 while largely maintaining the spatial and temporal coverage. For the SCIATRAN RT simulations, we used pressure, temperature, specific humidity, cloud ice content, cloud water content, and cloud fraction from the ECMWF ERA5 reanalysis (Hersbach et al., 2020). Since we focus mainly on cloud-free conditions, we used static cloud microphysical properties for convenience, representing spherical water droplets with a gamma particle size distribution with an effective radius of 12 µm and fractal ice particles with an effective radius of 50 µm (Fig. 3 of Reuter et al., 2010, shows the corresponding volume scattering functions). Aerosol data were derived from CAMS' ECMWF atmospheric composition reanalysis EAC4 (Inness et al., 2019). CO_2 profiles were derived from the CAMS global CO_2 atmospheric inversion v20r2 (Chevallier et al., 2005, 2010; Chevallier, 2013), and CH_4 profiles were obtained from the CAMS global CH_4 atmospheric inversion v20r1 (Segers, 2022). Surface reflectivity was modeled using the Moderate Resolution Imaging Spectroradiometer (MODIS) BRDF and albedo model parameter data set MCD43C1 version 6.1 (Schaaf and Wang, 2021). SIF was modeled using the MODIS Normalized Difference Vegetation Index (NDVI) MYD13C version 6.12 (Didan, 2021) as a proxy, following the approach outlined by RAL (2022). The resulting data set includes approximately 2.13 million cloud-free soundings over land in 2015 and 2.15 million in 2020. In addition, the data set also includes cloudy scenes that are sampled less densely depending on the cloud optical depth (COD), in order to emulate an imperfect cloud-masking algorithm. Specifically, cloudy scenes are computed with a probability of $P_{\mathrm{COD}} = 1 - \mathrm{COD}$ but at least 0.05. This means that optically thin clouds are likely to

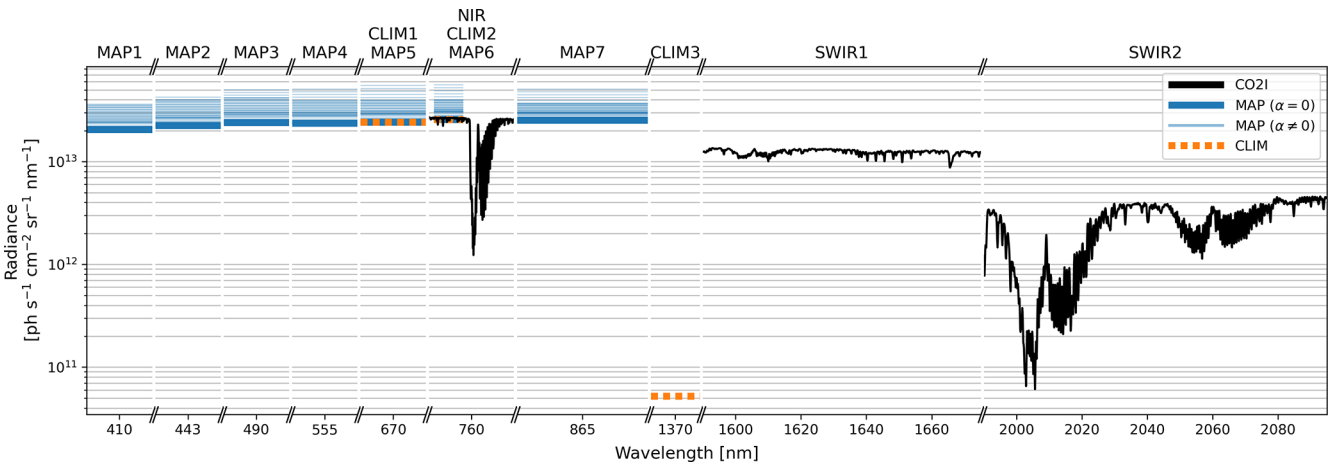

Figure 1. Simulated radiance measurements of a random CO2I sounding, including co-located MAP and CLIM radiance measurements. For MAP, the Stokes parameter I (total intensity) is shown for all simulated along-track observation angles α.

make it into the data set, while the probability of an optically thick cloud is only 5 %. This results in nearly half a million cloud-contaminated land scenes per year.

2.1.2 Berlin high-resolution scene

In addition to the subset data, we simulated a scene with the full CO2M sampling. It is a 3 min orbit granule with geophysical conditions of 3 July 2015, and since it includes Berlin (Germany), it is referred to as the Berlin HR scene. This scene is also used in EUMETSAT's CO2M preparation activities, and HR model data are available for it. Our SCIATRAN input for this scene is the same as for the subset data, except for pressure, temperature, specific humidity, CO_2, and CH_4, which have been provided by EUMETSAT and which are based on the CAMS "nature run" model data with a spatial resolution of about 9 km (Agustí-Panareda et al., 2022). In particular, this means that the scene includes HR CO_2 and CH_4 signals, such as XCO_2 plumes from power plants in eastern Germany, which are not resolved in the subset data. For the Berlin HR scene we simulated 43 671 soundings over land, of which 42 398 are cloud-free.

2.2 Noise

For key parts of our study (e.g., PCA and ANN training), we need data scattering within realistic uncertainties, i.e., with noise distributions reflecting the expected statistical variability. In the case of the radiometric CO2M measurements, we used the same noise models as Noël et al. (2024) and Meijer et al. (2020). Based on the study by Salstein et al. (2008), we assume that the uncertainty in the dry-air column is 2.5 ‰. We further assume that the atmospheric temperature is uncertain by 1 K, achieved by a shift in the entire profile. The atmospheric humidity is assumed to be uncertain by 10 %, achieved by profile scaling. For all observation angles, we define the uncertainty as 0.1°. The target quantities XCO_2 and XCH_4 used as training truths are assumed to have uncertainties of 1 ppm and 5 ppb, respectively, which are somewhat larger than the differences between models and ground-based measurements found by Knapp et al. (2021) and Kulawik et al. (2016) and somewhat smaller than those found by Tu et al. (2020). The uncertainty in the CO_2 and CH_4 a priori profiles is accounted for by multivariate noise computed with the same a priori error covariance matrices used by Noël et al. (2024), scaled so that the a priori XCO_2 and XCH_4 scatter around the truth with a standard deviation of 4 ppm and 20 ppb, respectively. All uncertainty specifications in this section represent 1σ values of normally distributed random variables. It should be noted that the input data for the RT simulations of the OSSE are free of noise. The main use of noise in our analyses is to generate realistically noisy training data (Sect. 2.5.2).

2.3 Modification of spectra

As discussed in Sect. 1, learning from simulated spectra can lead to biases for the same reasons as for conventional full-physics retrieval methods, namely because of inaccuracies in the RT and/or instrument simulation. This is why we prefer to learn from measured spectra. However, this approach also has some potential disadvantages: XCO_2 and XCH_4 increase over time so that today's concentrations are not representative of the future, XCO_2 and XCH_4 may have correlations to quantities such as albedo or observation geometry from which an ANN can learn as efficiently as from spectral features, and uncertainties in the training truth may exist. For these reasons, we use a method to modify measured spectra as if they include more or less of the target gases. Since CO2M was not yet operational at the time of our study, these measured spectra are simulations, i.e., the measurements of our OSSE simulated with SCIATRAN (Sect. 2.1).

An obvious way to modify a spectrum would be to use a synthetic Jacobian to simulate linear changes with respect to the geophysical state. However, due to nonlinearities of the RT, a more accurate alternative is to use the ratio of two synthetic spectra, i.e., a reference spectrum and a perturbed spectrum, for the modification. Since both the Jacobian and the ratio of the synthetic spectra depend on the geophysical state, it is necessary to first estimate it from the measurement.

In the following, we describe how we estimate the state from the measurement and generate the synthetic reference spectrum; how we perturb the state to make it representative of a wider range of conditions; how we generate the modified synthetic spectrum from it; and, finally, how we compute the modified measurement using the ratio of the synthetic spectra.

Let $I_m(x)$ be the measured CO2I intensity, i.e., a SCIATRAN-simulated measurement of our OSSE (Sect. 2.1). It is a function of the true state x including the true atmospheric concentration profiles of CO_2 and CH_4. In reality, when working with real measurements instead of simulations, the true state is of course not known. We fit this measurement using the FOCAL retrieval as described by Noël et al. (2024) but with some adaptations, guaranteeing that the vast majority of soundings converge. Specifically, we enlarge the measurement error covariance by assuming an unrealistically large forward model uncertainty of 1 % of the continuum radiance in all four fit windows and by allowing up to 40 iterations. The fitted radiance, i.e., the synthetic reference spectrum, is $I_f(\hat{x})$, where $\hat{x}$ is the retrieved state containing the retrieved concentration profiles $\hat{CO_2}$ and $\hat{CH_4}$. These profiles consist of five layers, each representing the same number of dry-air particles.

It is important to note that FOCAL's RT is much simpler than the RT of SCIATRAN used to simulate the measurements so that a perfect spectral fit is usually not possible, which is likewise the case when applying FOCAL to actually measured satellite data (Noël et al., 2021, 2022; Reuter et al., 2017a). As a result, the retrieved concentrations can significantly vary from the true atmospheric state, especially in scenes with enhanced scattering due to aerosols or clouds. This is more likely to be the case than in the earlier studies by Noël et al. (2021, 2022, 2024) and Reuter et al. (2017a) because we here forced FOCAL to almost always converge, and we applied no filtering or bias correction. However, this is not an issue for our study, since we are mainly interested in relative spectral changes, and we show that it is even sufficient to use a simple non-scattering RT model that considers only gaseous absorption.

In the next step, we compute the perturbed concentration profiles $\tilde{CO_2}$ and $\tilde{CH_4}$ by adding delta profiles, which we calculate as explained as follows using the example of CO_2.

- We randomly select two five-layer CO_2 profiles of the year 2015 from the Simple cLImatological Model (SLIM) for atmospheric CO_2 or CH_4 (Noël et al., 2022) and compute the difference concentration profile ΔCO_2.
- We randomly increase or decrease ΔCO_2 in the lowermost layer according to a normal distribution with a standard deviation of 10 ppm, emulating the signal of a local source or sink.
- We compute the profile anomaly; i.e., we subtract the column average of ΔCO_2 from ΔCO_2.
- We randomly shift the entire ΔCO_2 profile according to a uniform distribution between -40 and $+40$ ppm.

In this way, the shape of the delta profile ΔCO_2 has large but not unrealistic variations with height, and the variation in its column average ΔXCO_2 is large enough to be representative of atmospheric growth of many years.

ΔCH_4 and ΔXCH_4 are calculated using the same method but with all variations multiplied by a factor of 5×10^{-3}. This means that the standard deviation of the random CH_4 perturbation in the lowermost layer becomes 50 ppb instead of 10 ppm, and the range of the uniform distribution for the random shift in the profile in the last step becomes [−200 ppb, +200 ppb] instead of [−40 ppm, +40 ppm]. Note that the perturbations of CH_4 are independent of those of CO_2.

As discussed above, the FOCAL-retrieved dry-air column averages $\hat{XCO_2}$ and $\hat{XCH_4}$ of our study may be significantly biased, and we here consider them to be representative of only the apparent light path. However, the corresponding climatological values XCO_{2SLIM} and XCH_{4SLIM} are relatively close to reality (Noël et al., 2022). Therefore, we scale the delta profiles ΔCO_2 and ΔCH_4 by a factor of $\hat{XCO_2}/XCO_{2SLIM}$ and $\hat{XCH_4}/XCH_{4SLIM}$, respectively, before performing the FOCAL forward run. This primarily affects scenarios with large deviations between retrieved and true concentrations.

$$\tilde{CO_2} = \hat{CO_2} + \Delta CO_2 \frac{\hat{XCO_2}}{XCO_{2SLIM}} \quad (1)$$

$$\tilde{CH_4} = \hat{CH_4} + \Delta CH_4 \frac{\hat{XCH_4}}{XCH_{4SLIM}} \quad (2)$$

These modified concentration profiles are part of the perturbed state $\tilde{x}$, which we use to perform an additional FOCAL forward run, i.e., RT and instrument simulation, in order to compute the modified synthetic spectrum $I_f(\tilde{x})$. This is then used to approximate what the measured radiance would look like if ΔCO_2 and ΔCH_4 were added to the true CO_2 and CH_4 profiles.

$$I_m(x + \Delta x) \approx I_m(x) \frac{I_f(\tilde{x})}{I_f(\hat{x})} \quad (3)$$

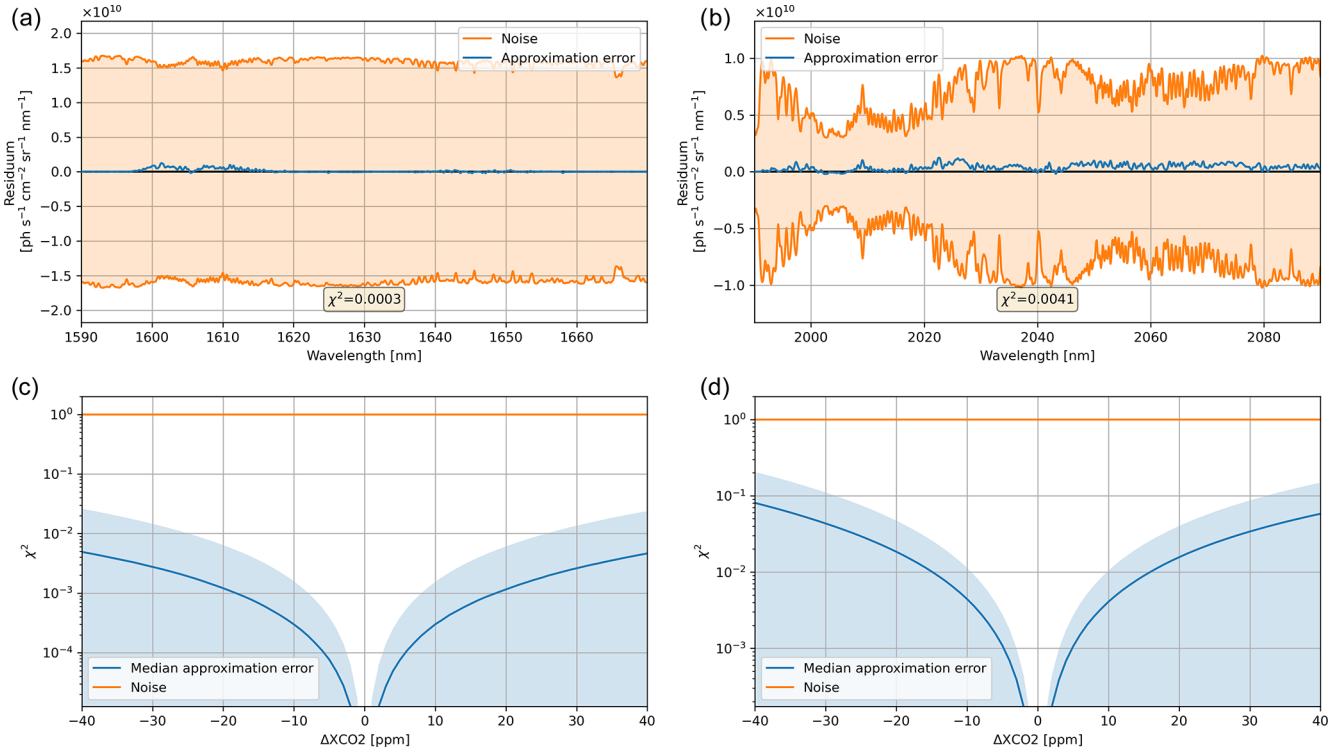

Figure 2. Spectrum of the approximation error (approximation minus SCIATRAN simulation) and instrumental noise for a typical scene and a 10 ppm shift in the CO_2 profile in the SWIR-1 **(a)** and SWIR-2 **(b)**. Median (blue line) $\pm 1\sigma$ (light blue area) of the approximation error χ^2 statistics for all cloud-free soundings of one orbit of subset data from 3 July 2015 as a function of the profile shift for the SWIR-1 **(c)** and SWIR-2 **(d)**.

The quality of this approximation can be determined by comparing radiances approximated by Eq. (3) with corresponding SCIATRAN simulations. For this purpose, we selected one orbit of subset data of 3 July 2015, including many cloud-free scenes above Europe and Africa, and shifted the entire CO_2 profile from -40 to $+40$ ppm in steps of 1 ppm. Figure 2 shows an example spectrum of the approximation error in the SWIR-1 (Fig. 2a) and SWIR-2 (Fig. 2b) bands for a 10 ppm shift in the CO_2 profile. The figure shows that the approximation error, i.e., the difference between the approximation and the SCIATRAN simulation, is much smaller than the instrumental noise. As can be seen in Fig. 2c and d, the approximation error disappears for small profile shifts and steadily increases towards larger profile shifts. It is usually 1 order of magnitude larger in the SWIR-2 than in the SWIR-1. However, it is always significantly smaller than the instrumental noise. As an example, for a 10 ppm shift, the median χ^2 amounts to 0.0003 in the SWIR-1 and 0.0041 in the SWIR-2. This means the approximation is valid within a range much larger than the current annual growth rate, thus allowing us to generate a training data set from modified measured spectra that is also representative of atmospheric conditions several years in the future.

2.4 Principal component analysis

Atmospheric spectra, such as those measured by CO2I, contain a large amount of redundant information. In such cases, PCA is an efficient tool for dimensionality reduction without losing important information (e.g., Liu et al., 2006). It can significantly reduce the size of the training data set and improve the learning efficiency of ANNs.

We used every seventh sounding of all even weeks in the 2015 subset data set and performed a PCA on different input data sets: the NIR band, the combination of both SWIR bands, the combination of all three bands, the MAP data, and the meteorological profiles of temperature and humidity.

The choice of the number of principal components used is not trivial and is somewhat subjective. Using a large number of components ensures that no information is lost, but the dimensionality reduction is small. If only a few components are used, the dimensionality reduction is high, but important information may be lost. We found that 25 components are sufficient for the NIR band, 90 for the combined SWIR bands, 100 for the combination of all three bands, 100 for the MAP data, and 5 for the temperature and humidity profiles.

We based our choice of the number of components on calculations of the unexplained variance, the χ^2 of the reconstruction error, and the number of components that lead

to a minimization of the denoising error. The unexplained variance is equal to 1 minus the explained variance, which is commonly used in the context of PCA (e.g., Jolliffe and Cadima, 2016). The χ^2 of the reconstruction error is calculated from the residual of the reconstructed and original measurements relative to the noise estimate of the measurements. The denoising error analyzes the residual between the reconstructed noisy data and the noise-free original data. It depends on the number of components used and reaches a minimum when the use of additional components would predominantly lead to fitting noise but not signal (Aires et al., 2002; Di Noia et al., 2015). The χ^2 of the reconstruction error and the denoising error were only determined for the radiation measurements where the noise estimates are known and reliable.

As an example, for the combination of the NIR and both SWIR bands, we find that when using 100 components, the fraction of unexplained variance amounts to 4.2×10^{-9}. The χ^2 of the reconstruction error is 1.1×10^{-3}, which means that the instrumental noise can be expected to be about 1000 times larger than the reconstruction error. The denoising error becomes minimal when using 75 components. Selecting a significantly larger number of components can result in fitting noise, while selecting a significantly smaller number can result in loss of information. The results of all PCA studies are summarized in Table 1.

2.5 Artificial neural networks

2.5.1 Setup

In our study, we examined four different input compositions. The *baseline* setup is the standard setup used in this study. All other setups differ only in their details in order to study their influence on the ANN's prediction quality separately. The baseline setup exists in a variant for XCO_2 and a variant for XCH_4. For simplicity, all other input setups exist only for XCO_2.

The baseline input consists of the scores of the 100 most significant principal components (PCs) of the combined NIR, SWIR-1, and SWIR-2 spectra; the scores of the 100 most significant PCs of the MAP data; the CLIM3 radiance; the scores of the five most significant PCs of the meteorological temperature and humidity profiles; the number of dry-air particles in the atmospheric column; the solar zenith angle; the satellite zenith angle; and the azimuth difference. As for conventional retrievals based on optimal estimation, the input also contains a noisy/uncertain a priori CO_2 or CH_4 profile (Sect. 2.2), which in our case consists of five atmospheric layers, each containing the same number of dry-air particles. Figure 3 illustrates the baseline ANN training setup on the example of XCO_2.

The *no-MAP* input differs from the baseline input only in that it does not contain MAP and CLIM data. In addition to the missing MAP data, the *no-NIR* input also lacks data from the NIR band. The *non-scat.* setup is the same as the baseline setup, except that the modified spectra used for the training data set were generated by a FOCAL variant that only considers absorption but not scattering in the atmosphere.

All results were generated using MLP regressors with three hidden layers of 150, 30, and 150 neurons. The idea behind this ANN architecture is to improve the generalization capabilities of the network by adding a so-called information bottleneck in the middle layer, which holds the information of intermediate meta-parameters. Conceptually, there are parallels to first performing a conventional retrieval and then using the set of output parameters as input to a bias correction. We used the logistic, i.e., sigmoid, activation function and trained the MLPs with the Adam stochastic optimization method (Kingma and Ba, 2014) of the scikit-learn Python library (Pedregosa et al., 2011).

2.5.2 Training and test data set

To construct a representative and realistic training data set, we use noisy input and target data (see Sect. 2.2), which we construct from the data of all even weeks of the 2015 subset data set (Sect. 2.1.1). The data of the odd weeks are mainly reserved for testing. Separating the data sets on a weekly basis ensures that seasonal variations are sampled finely enough while avoiding strong correlations between the two data sets that could occur with random sampling. It is important that the training data set contains noise, as all input and target features will of course be subject to inherent uncertainties during later training with real CO2M data. In addition, the noise supports generalized learning and suppresses overfitting. The subset data contain a small fraction of cloudy scenes (Sect. 2.1.1), which we also expect to be the case in real data due to imperfect cloud masking. In order to create a realistic data set and to make the prediction less sensitive to residual cloud contamination, we filter out only clouds with an optical thickness greater than 0.05. From each remaining sounding, we generate 10 soundings whose SWIR spectra have been modified as described in Sect. 2.3. Only these modified soundings, which have artificially increased XGAS variabilities, are the basis of our training data set.

2.5.3 Prediction of uncertainties

Interpretation of XCO_2 or XCH_4 satellite data requires appropriate uncertainty estimates. There are a number of ways to estimate the uncertainty in an ANN's prediction from the uncertainty in its input. The simplest approach is to present an existing ANN with multiple realizations of the input, modified according to its error characteristics, and then statistically analyze the predictions. However, there are more sophisticated methods, such as the use of probabilistic ANNs (Mohebali et al., 2020). Here, we use a simple but efficient method by training MLPs to predict the XGAS uncertainties σXCO_2 and σXCH_4 from the same inputs used to pre-

Table 1. PCA results for various input data sets: NIR band, combination of both SWIR bands, combination of all three bands, MAP data, and meteorological profiles of temperature and humidity. Input data sets used for the baseline configuration of the ANNs (see later sections) are highlighted in bold. The table lists the number of components used, the corresponding unexplained variance and the χ^2 of the reconstruction error, and the number of components for which the denoising error is minimal. Note: NA – not available.

	NIR	SWIR-1 + 2	NIR + SWIR-1 + 2	MAP	Temperature	Specific humidity
Components used	25	90	**100**	**100**	**5**	**5**
Unexplained variance	2.9×10^{-9}	3.5×10^{-9}	$\mathbf{4.2 \times 10^{-9}}$	$\mathbf{2.4 \times 10^{-5}}$	$\mathbf{1.3 \times 10^{-2}}$	$\mathbf{8.8 \times 10^{-3}}$
Reconstruction error χ^2	2.4×10^{-4}	9.9×10^{-4}	$\mathbf{1.1 \times 10^{-3}}$	$\mathbf{9.5 \times 10^{-2}}$	**NA**	**NA**
Components with minimal denoising error	17	55	**75**	**120**	**NA**	**NA**

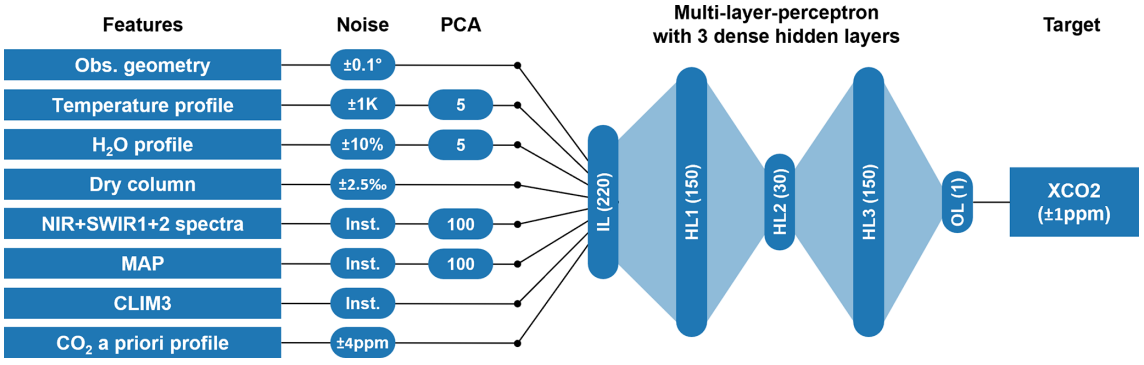

Figure 3. Baseline ANN training setup on the example of XCO_2, including the amount of noise added to the training features and to the target variable (Sect. 2.2) and the PCA components used (Sect. 2.4). When training with actual measured data in the future, the addition of noise will be omitted. Inst – noise of the instrument model, IL – input layer, HL – hidden layer, OL – output layer.

dict XCO_2 and XCH_4, except that no a priori information is used. More specifically, we apply the XGAS MLPs to the test data set (Sect. 2.5.2) and compute the squared prediction mismatches (prediction minus training truth) $\Delta XGAS^2$ and use them as training targets for additional MLPs that predict the XGAS variances $\sigma XGAS^2$, as suggested by Bishop (1996). The rationale behind this is that the expected value of $\Delta XGAS$ is small, enabling the variance VAR(XGAS) to be approximated by the expected value of $\Delta XGAS^2$. We use data from the test period instead of the training period because the prediction mismatches $\Delta XGAS$ can be considered more realistic.

2.5.4 Column averaging kernel

In addition to reliable uncertainty estimates, the interpretation of XCO_2 or XCH_4 satellite data also requires information about the column averaging kernel (AK). The AK quantifies the retrieval's sensitivity to changes in the target gas concentration profile and is defined by

$$\mathrm{AK}_i = \frac{1}{w_i} \frac{\partial \hat{\mathrm{XGAS}}}{\partial \mathrm{GAS}_i^{\mathrm{t}}}, \tag{4}$$

where $\hat{\mathrm{XGAS}}$ is the retrieved XGAS, $\mathrm{GAS}_i^{\mathrm{t}}$ the true gas concentration in the height layer i, and w_i the relative dry-air weight of that layer; i.e., the number of dry-air particles in sub-column i divided by the total number of dry-air particles in the atmospheric column. In the context of retrieval comparison studies or surface flux inversions (e.g., Reuter et al., 2011; Bergamaschi et al., 2007; Wunch et al., 2011), $1 - \mathrm{AK}$ can be interpreted as the influence of or the dependence on the a priori used. While we do not have direct access to the column averaging kernel, the influence of the a priori $\partial \hat{\mathrm{XGAS}} / \partial \mathrm{GAS}_i^{\mathrm{a}}$ can be easily determined numerically by predicting XGAS for perturbed a priori profiles $\mathrm{GAS}^{\mathrm{a}}$ and approximating

$$\mathrm{AK}_i \approx 1 - \frac{1}{w_i} \frac{\partial \hat{\mathrm{XGAS}}}{\partial \mathrm{GAS}_i^{\mathrm{a}}}. \tag{5}$$

2.5.5 Postprocessing

As with conventional greenhouse gas retrieval algorithms, we filter out the least promising scenes during postprocessing. To do this, we analyzed the 2015 evaluation data set (Sect. 2.5.6) and computed a threshold for the maximum allowed predicted uncertainty (Sect. 2.5.3) that filters out 10 % of the cloud-free 2015 evaluation data. From the remaining data, we computed a threshold for the maximum allowed dependence on the a priori, which filters out another 11.11 %. In this way, the most promising 80 % of all cloud-

free soundings remain after both filters. The thresholds are setup-specific and are listed in Table 2.

Additionally, we used the 2015 evaluation data set to compute a setup-specific overall offset (Table 2), which we subtract from the prediction during postprocessing.

For each sounding, the a priori dependence is computed from the profile average sensitivity of the prediction to the a priori (Sect. 2.5.4). For example, if the dependence on the a priori was 5 %, then adding 1 ppm to the CO_2 a priori would increase the XCO_2 prediction by 0.05 ppm.

Similar to the dependence on the a priori, we compute the relative dependence of the prediction on the dry column. This quantity specifies how dry-column errors propagate to XGAS prediction errors. For example, if the dependence on the dry column was 5 %, then a 1 % error in the dry column would result in a 0.05 % error in the predicted XGAS. This quantity is not used directly during postprocessing but is analyzed when interpreting the results.

2.5.6 Evaluation data sets

We quantified the ANNs' prediction quality by applying them to three evaluation data sets that were not used for training: (i) the unmodified 2015 subset data set (Sect. 2.1.1), which we divided into a training and test period because it served as the basis for computing the training and test data sets (Sect. 2.3); (ii) the unmodified 2020 subset data set (Sect. 2.1.1) with geophysical conditions and greenhouse gas concentrations not seen during the training; and (iii) the Berlin HR scene (Sect. 2.1.2), also with geophysical conditions and greenhouse gas plumes that were not part of the training data set.

3 Results

For the input setups described in Sect. 2.5.1, MLPs with the properties described in the same section were trained to predict XCO_2 and the associated uncertainty. In the case of the baseline setup, MLPs were also trained to predict XCH_4 and its uncertainty. In order to analyze the prediction quality, the MLPs were applied to the evaluation data described in Sect. 2.5.6, and the prediction was compared with the truth.

Since the CO2M mission requirements are defined for cloud-free conditions, we filtered the evaluation data accordingly. Additionally, we applied the postprocessing filters described in Sect. 2.5.5. Most of the analyses were performed with noise-free input data, so the prediction errors can be considered purely systematic.

The results for the 2020 subset evaluation data and the Berlin HR scene are the most conclusive because their input is the most independent of the training data set. In the following, we focus on the results for these data sets obtained with the baseline setup. However, Table 2 summarizes the main results for the analysis of all evaluation data sets and input configurations.

3.1 Column averaging kernels

We analyzed the XCO_2 and XCH_4 AKs of the 2020 subset evaluation data set. Figure 4a shows that the XCO_2 AKs are close to optimal, i.e., close to unity, in large parts of the atmosphere. Significantly lower values are observed only in the stratosphere. The XCH_4 AKs also decrease in the stratosphere but show a slight overestimation of departures from the a priori in other layers (Fig. 4b).

3.2 Stochastic errors

In order to determine the overall retrieval precision due to instrumental noise, we predicted XCO_2 and XCH_4 from input with and without instrumental noise and calculated the standard deviation of the difference. For the postprocessed 2020 evaluation data set and the baseline setup, it amounts to 0.41 ppm for XCO_2 and 2.74 ppb for XCH_4.

As can be seen in Table 2, these values are basically identical to those obtained for the training and test periods of the 2015 evaluation data set and similar to those obtained for the Berlin HR scene.

The stochastic XCO_2 error does not change for the *non-scat.* setup (Sect. 2.5.1) but increases slightly to 0.45 ppm when the MAP instrument is not used. We see a more significant increase to 0.66 ppm when also not using the NIR band.

In addition to the analysis of the overall precision, we validated the MLPs predicting the retrieval uncertainty in the individual soundings (Sect. 2.5.3). For this purpose, we defined 15 bins, each containing the same number of soundings, for the predicted XCO_2 or XCH_4 uncertainty. For each bin, we determined the average predicted uncertainty, which we compared to the actual precision in that bin.

Figure 5 shows that the XCO_2 and XCH_4 uncertainties are well predicted by the MLPs. The predicted XCH_4 uncertainties are almost accurate. The predicted XCO_2 uncertainties behave similarly but with a small offset of about 0.03 ppm.

3.3 Systematic errors

3.3.1 Overall statistics

We compute systematic errors by comparing postprocessed predicted XCO_2 and XCH_4 values with corresponding true values for noise-free input data. Figure 6 shows such a comparison for the 2020 subset data and the baseline setup.

With 0.04 ppm for XCO_2 and 0.20 ppb for XCH_4, the mean bias (prediction minus truth) for the 2020 subset data is negligible. It is not surprising that this is also the case for the 2015 subset data, as this data set has been used to derive the postprocessing offset correction (Sect. 2.5.5). The mean

Table 2. Algorithm setup, postprocessing parameters, and main results generated from the evaluation subset data sets of 2015 and 2020 and from the Berlin HR scene for the baseline (bold), no-MAP, no-NIR, and non-scat. configuration.

	Baseline		No-MAP	No-NIR	Non-scat.
Setup					
Target	XCO_2	XCH_4	XCO_2	XCO_2	XCO_2
NIR	**yes**	**yes**	yes	no	yes
SWIR-1+2	**yes**	**yes**	yes	yes	yes
MAP+CLIM	**yes**	**yes**	no	no	yes
Modification method	**scat.**	**scat.**	scat.	scat.	non-scat.
Postprocessing					
Max $\sigma XCO_2/\sigma XCH_4$ [ppm/ppb]	**0.71**	**5.15**	0.71	0.74	0.69
Max a priori dependence [%]	**16.1**	**0.4**	17.0	32.2	15.6
Subtracted offset [ppm/ppb]	**−0.11**	**0.64**	0.17	0.04	−0.00
Evaluation results 2015 subset data					
Soundings [no.]	**1 704 695**	**1 699 842**	1 704 595	1 704 181	1 702 525
Throughput [%]	**80**	**80**	80	80	80
Precision train/test [ppm/ppb]	**0.41/0.41**	**2.72/2.72**	0.46/0.46	0.65/0.65	0.41/0.41
Accuracy train/test [ppm/ppb]	**0.39/0.42**	**2.20/2.37**	0.42/0.46	0.38/0.42	0.39/0.43
Mean bias [ppm/ppb]	**−0.00**	**0.00**	0.00	−0.00	−0.00
Mean a priori dependence [%]	**9.2**	**−4.7**	9.2	13.5	9.1
Mean dry-column dependence [%]	**−6.2**	**−4.7**	−16.5	−60.6	−5.6
Evaluation results 2020 subset data					
Soundings [no.]	**1 704 349**	**1 724 657**	1 691 721	1 685 554	1 679 922
Throughput [%]	**79**	**80**	79	78	78
Precision [ppm/ppb]	**0.41**	**2.74**	0.45	0.66	0.41
Accuracy [ppm/ppb]	**0.44**	**2.45**	0.48	0.44	0.44
Mean bias [ppm/ppb]	**0.04**	**0.20**	0.02	−0.04	0.00
Mean a priori dependence [%]	**9.6**	**−4.8**	9.3	14.2	8.7
Mean dry-column dependence [%]	**−5.9**	**−4.7**	−16.0	−60.1	−5.2
Evaluation results Berlin HR scene					
Soundings [no.]	**41 757**	**41 685**	41 390	41 189	41 888
Throughput [%]	**98**	**98**	98	97	99
Precision [ppm/ppb]	**0.44**	**3.12**	0.47	0.70	0.43
Accuracy [ppm/ppb]	**0.31**	**1.72**	0.40	0.39	0.34
Mean bias [ppm/ppb]	**−0.18**	**−2.13**	−0.29	0.22	−0.36
Mean a priori dependency [%]	**11.4**	**−5.2**	12.3	18.7	11.0
Mean dry-column dependency [%]	**−3.9**	**0.8**	−17.3	−70.8	−2.4

bias for the Berlin HR scene is −0.18 ppm for XCO_2 and −2.13 ppb for XCH_4 (baseline setup).

Surface flux inverse modeling and emission estimation results are much more sensitive to spatially and/or temporally varying biases than to constant offsets. Therefore, we consider the standard deviation of the difference between the predicted and true values of XGAS as a measure of accuracy. For the 2020 subset data and the baseline setup it amounts to 0.44 ppm and 2.45 ppb for XCO_2 and XCH_4, respectively.

The accuracy values determined from the 2015 subset data are slightly smaller. The modification of the spectra used for the training can introduce small spectral errors (Sect. 2.3). These can erode the prediction quality the further we depart from the concentrations of the training year 2015. Additionally, we observe that the prediction accuracy is about 10 % better for the training period than for the test period.

For the Berlin HR scene and the baseline setup, we obtain an accuracy of 0.28 ppm and 1.49 ppb for XCO_2 and XCH_4, respectively.

As can be seen in Table 2, the XCO_2 accuracy depends only slightly on the setup, particularly for the subset evaluation data. At first glance, this appears to be surprising because it would imply that the NIR band and the MAP instrument have only little influence on the systematic errors, which is

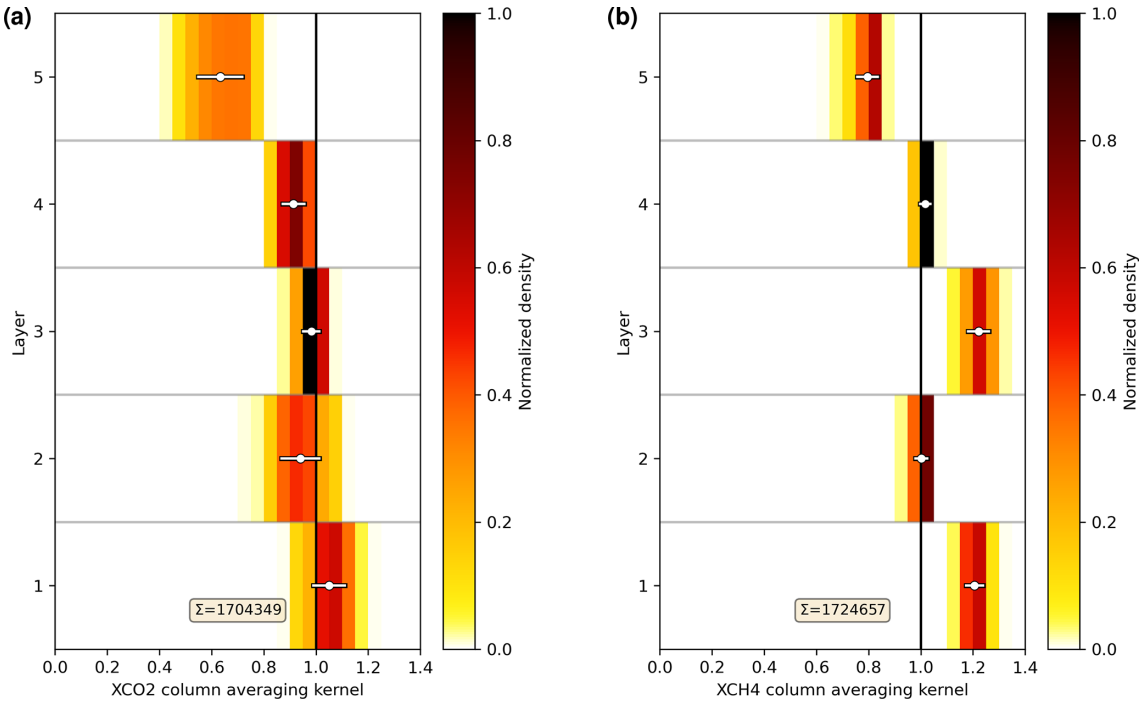

Figure 4. Normalized density distribution of the XCO_2 **(a)** and XCH_4 **(b)** column averaging kernels of all postprocessed soundings of the 2020 subset evaluation data set. Mean values and standard deviations are overlaid. The profiling splits the atmospheric column into five layers, each containing the same number of dry-air particles. Layer 1 is the closest to the surface and includes the boundary layer, and the stratosphere extends into layer 5. Σ represents the total number of soundings.

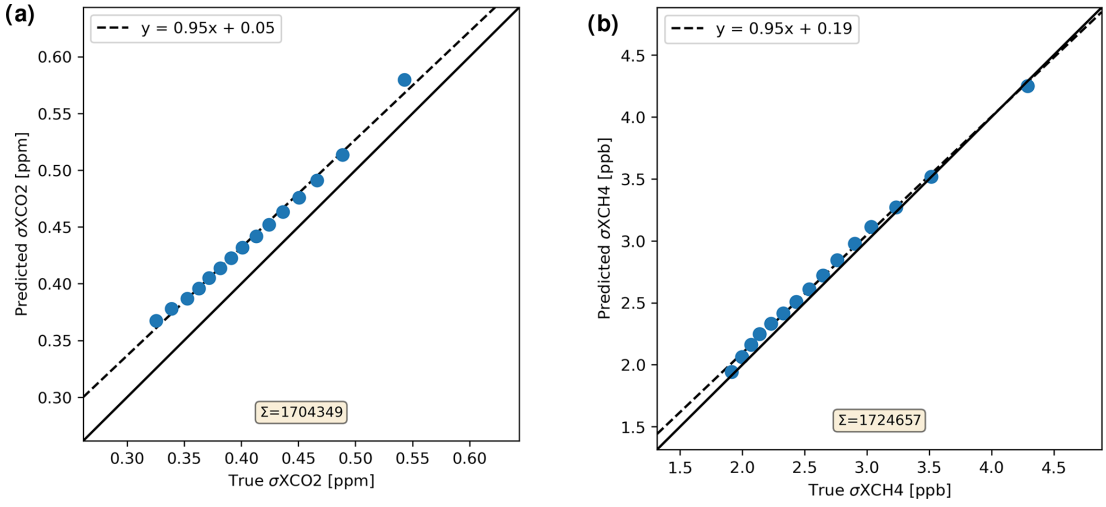

Figure 5. Comparison of the predicted and true XCO_2 **(a)** and XCH_4 **(b)** retrieval uncertainties due to instrumental noise for the postprocessed 2020 subset data. Σ represents the total number of soundings. The figure also contains the results of a linear regression.

not necessarily the case. Our analyses of systematic errors do not consider systematic errors in the input, such as the dry column or the a priori, which will exist in reality. When removing MAP from the input, the average dependence of the XCO_2 prediction on the dry column increases from −5.9 % to −16.0 %. Additionally, removing the NIR band further increases the dry-column dependence to −60.1 % and also increases the mean a priori dependence from 9.6 % to 14.2 %. For comparison, the dry-column dependence of the FOCAL CO2M XCO_2 retrieval is 100 % by design (Noël et al., 2024), and the dry-column dependence of the operational OCO-2 XCO_2 retrieval (v11.1) is approximately 85 % (Jacobs et al., 2024).

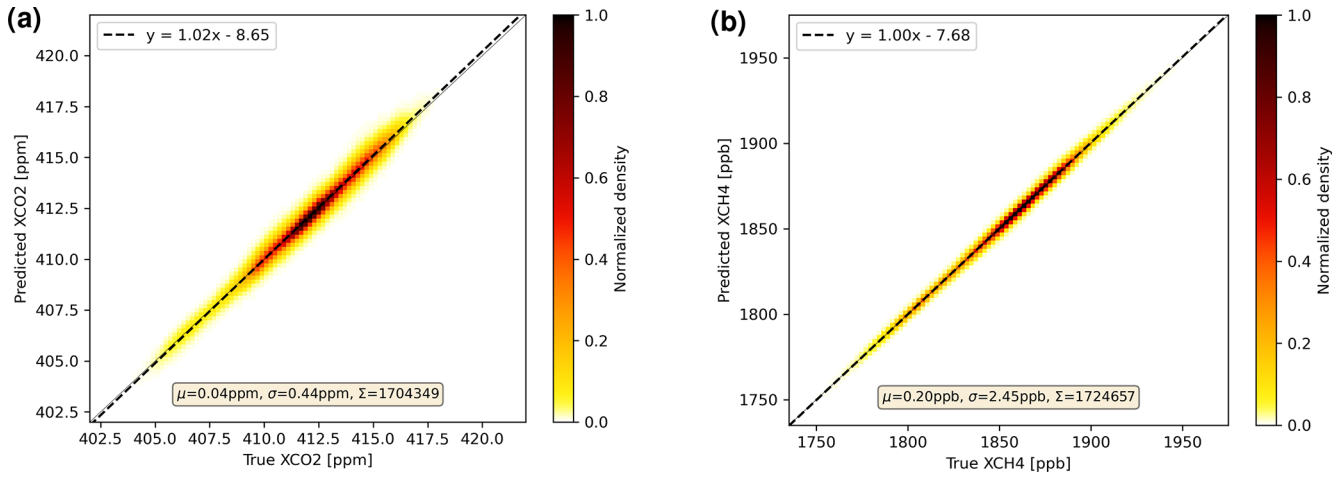

Figure 6. Comparison of postprocessed predicted XCO_2 **(a)** and XCH_4 **(b)** with corresponding true values for noise-free 2020 subset input data. Δ represents the average prediction error (prediction minus true), σ the standard deviation of the prediction error, and Σ the total number of soundings. The figure also contains the results of a linear regression.

3.3.2 Large-scale features

In order to investigate the spatial structures of the systematic errors, we generated global maps for XCO_2 (Fig. 7) and XCH_4 (Fig. 8), showing the postprocessed predicted and corresponding true values, as well as their difference for the noise-free subset input data of April and August 2020. First, the maps show a dense sampling because the postprocessing filters are designed to have a high throughput of about 80 % for all cloud-free soundings (Sect. 2.5.5).

The maps of the predicted and true XGAS show expected large-scale features, like low XCO_2 values in northern middle and high latitudes in August at the end of the growing season or relatively high XCH_4 values in the tropics. The differences between predicted and true XGAS values are generally much smaller than those of the large-scale features. However, the differences are not purely random and exhibit some country- to continental-scale systematic features, such as the small XCO_2 and XCH_4 high bias in Greenland in April or the small XCH_4 high bias in northern Africa in August.

There are some similarities between the XCO_2 and XCH_4 bias patterns, which may indicate that some systematic errors could cancel out in a proxy product when using, e.g., the ratio XCH_4 / XCO_2 as a training target.

The global monthly average biases are small, and the corresponding standard deviations are similar to the annual averages.

3.3.3 Seasonal cycle

Systematic errors may also have a seasonal component, e.g., due to seasonal variations in illumination conditions, albedo, or aerosols. Figure 9 shows the XGAS prediction error as a function of the week in the year 2020. According to this figure, the average systematic XCO_2 prediction error slowly drifts around zero, with the largest values of about 0.2 ppm in late (northern hemispheric) spring and the smallest values of about -0.1 ppm in autumn. The standard deviation of the XCO_2 error is larger in spring and summer (up to about 0.55 ppm) compared to autumn and winter (down to about 0.40 ppm). Various reasons can cause this behavior; e.g., sampling in summer covers a wider latitude range and, therefore, also more surface types and observation angles than in winter, and the CO_2 profiles vary more during the (northern hemispheric) growing season. The weekly average prediction error in XCH_4 has no clear seasonal cycle and is always smaller than ± 1 ppb. Its standard deviation varies between about 2.2 and 3.3 ppb.

3.3.4 Aerosols

Aerosols modify the light path and can for this reason be an important source of XGAS retrieval errors. Figure 10 shows the XGAS prediction error as a function of aerosol optical depth (AOD) for noise-free 2020 subset data. As can be seen, the XCO_2 average prediction error stays close to zero up to an AOD of 0.2. For larger AODs up to 0.5, the average prediction error steadily increases to values of about 0.1 ppm. The standard deviation of the prediction error increases with AOD from about 0.35 to about 0.60 ppm. The average XCH_4 prediction error is usually below ± 0.5 ppb, and its standard deviation increases from about 2.0 ppb for basically aerosol-free scenarios to about 3.2 ppb for scenarios with an AOD of up to 0.5.

3.3.5 Berlin HR scene

Since the AKs are close to unity in large parts of the atmosphere (Fig. 4), the prediction can be considered to be dominated by the measurement but not the a priori. In order to

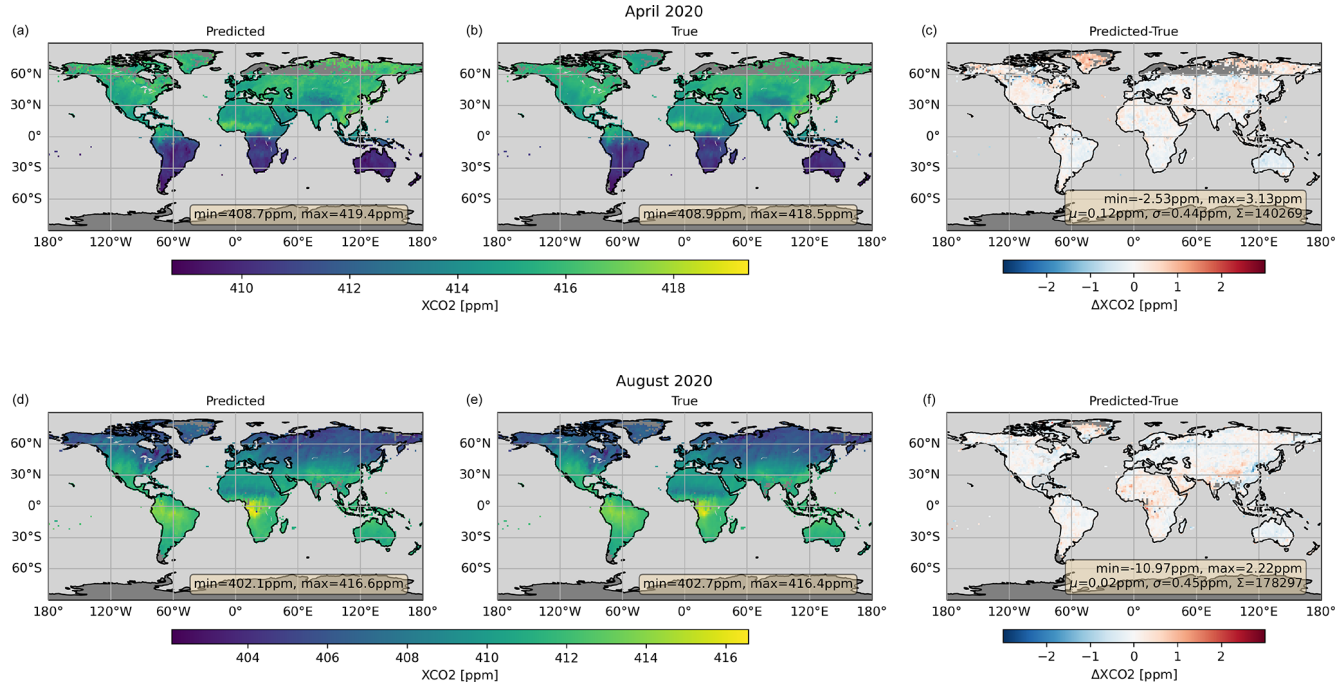

Figure 7. Global maps of postprocessed predicted XCO_2 **(a, d)** and the corresponding true values **(b, e)**, as well as their difference **(c, f)** for noise-free subset input data of April **(a–c)** and August **(d–f)** 2020. μ represents the average prediction error (prediction minus true), σ the standard deviation of the prediction error, and Σ the total number of soundings.

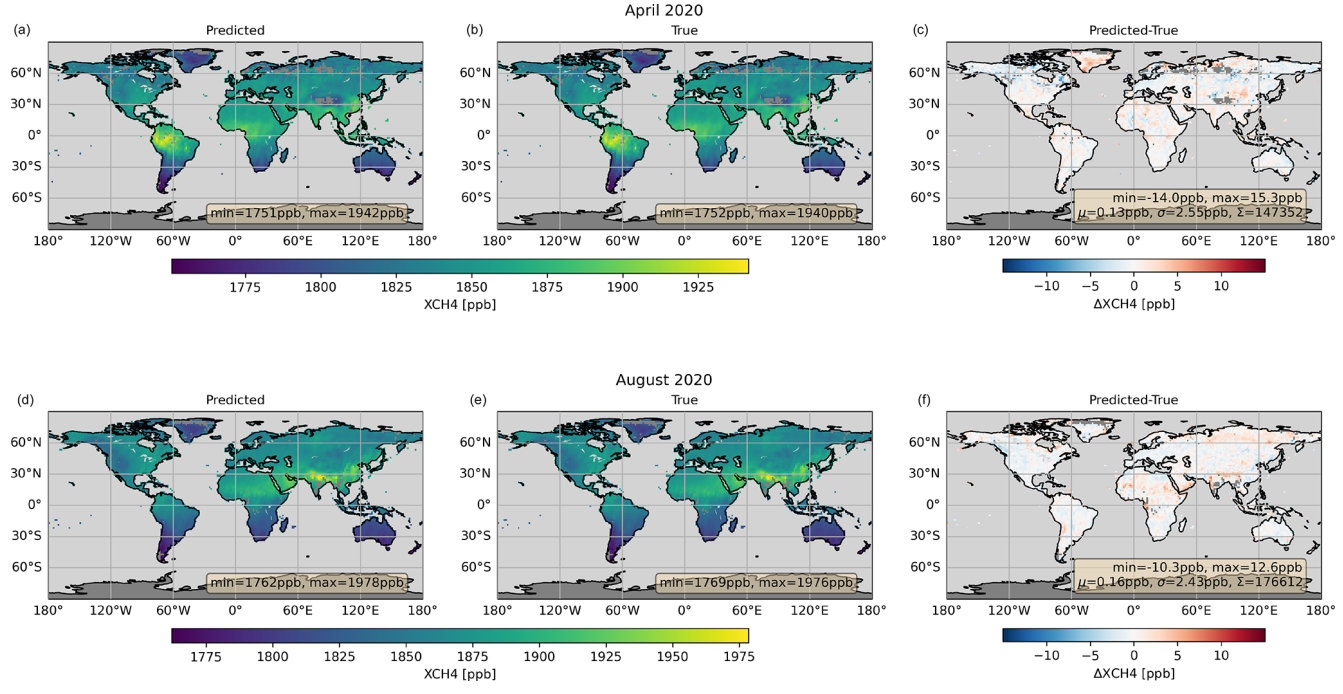

Figure 8. Global maps of postprocessed predicted XCH_4 **(a, d)** and the corresponding true values **(b, e)**, as well as their difference **(c, f)** for noise-free subset input data of April **(a–c)** and August **(d–f)** 2020. μ represents the average prediction error (prediction minus true), σ the standard deviation of the prediction error, and Σ the total number of soundings.

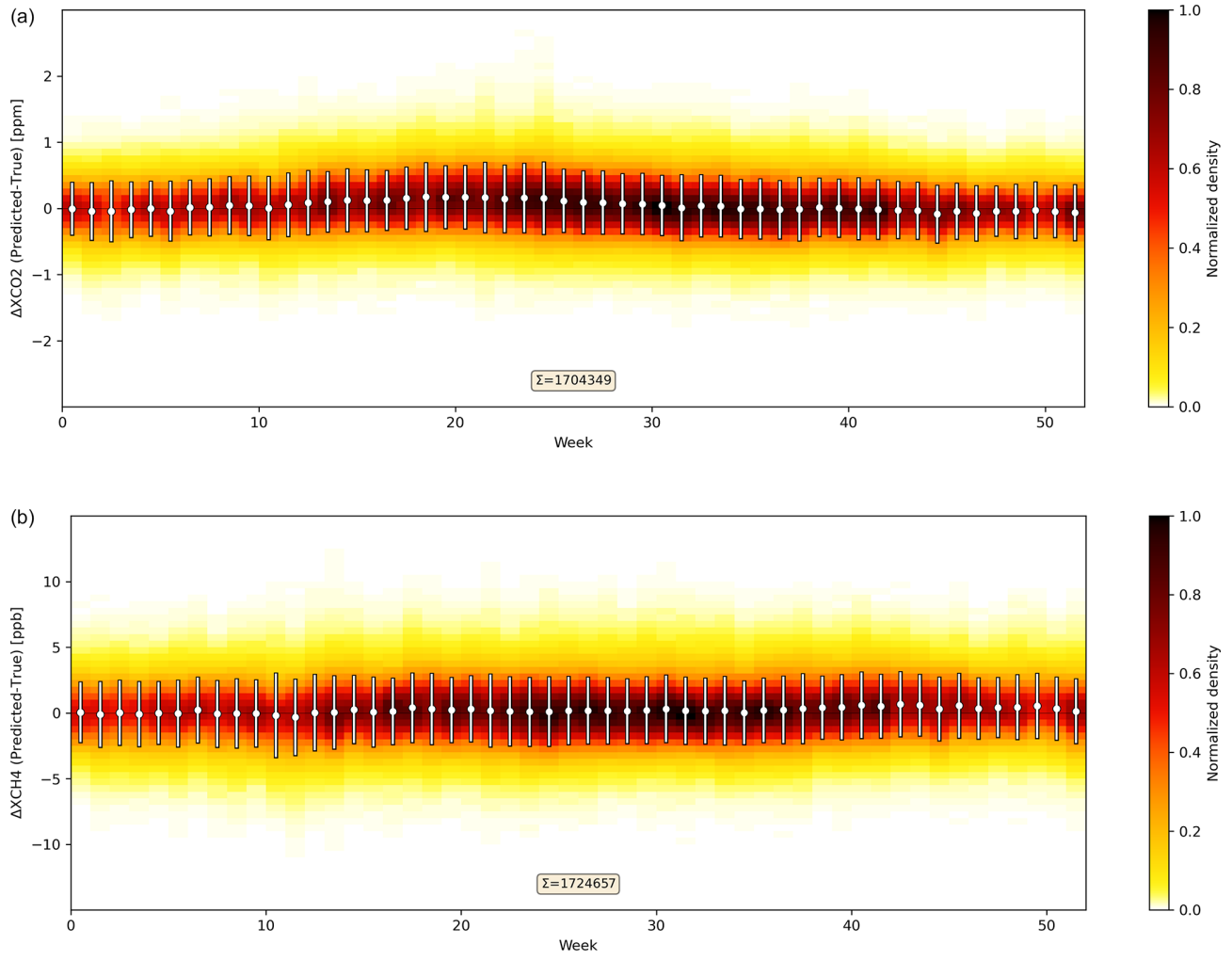

Figure 9. XCO_2 **(a)** and XCH_4 **(b)** prediction error as a function of the week for noise-free 2020 subset input data. The dots and bars represent the mean and standard deviation, respectively. Σ represents the total number of soundings.

illustrate this, we used scene-wide constant a priori profiles instead of the true concentration profiles to analyze the Berlin HR scene. Specifically, we used the scene-wide average true CO_2 and CH_4 concentration profiles as a priori.

Figures 11 and 12 show that the predictions reproduce the true concentrations well, even though the meteorological conditions and gas concentrations, including plumes from strong CO_2 and CH_4 sources, were not part of the training data or the a priori.

The variability of the difference structures is much smaller than the variability of the atmospheric signals. The XCO_2 prediction error has a standard deviation of 0.31 ppm and an average of −0.18 ppm. It shows no obvious correlations with the XCO_2 pattern, especially the CO_2 plumes from the coal-fired power plants Jänschwalde, Schwarze Pumpe, and Boxberg in eastern Germany.

The prediction error in XCH_4 is on average −2.13 ppb and has a standard deviation of 1.72 ppb. However, the map of the XCH_4 prediction error (Fig. 12) shows an interesting feature at about 50.53° N, 13.61° E in the Czech Republic. There is a strong CH_4 plume at this position, the strength of which is obviously overestimated by the prediction.

As a reminder, the AKs describe the behavior of the retrieval in over- or underestimating differences between the true and the a priori concentrations. The plume stands out from the scene average concentrations, i.e., the a priori, at roughly 90 ppb. The XCH_4 AKs in the lowermost layer can have values of up to 1.3, which would result in an overestimation of the departure from the a priori by 30 %, i.e., 27 ppb in this case. It should be noted that this would not result in an overestimation of the emission strength, if AKs are considered appropriately.

When using the true CO_2 and CH_4 profiles as a priori, the difference maps look similar, except that there is no such overestimation of the CH_4 plume because the departure from the a priori becomes much smaller (Figs. A1 and A2).

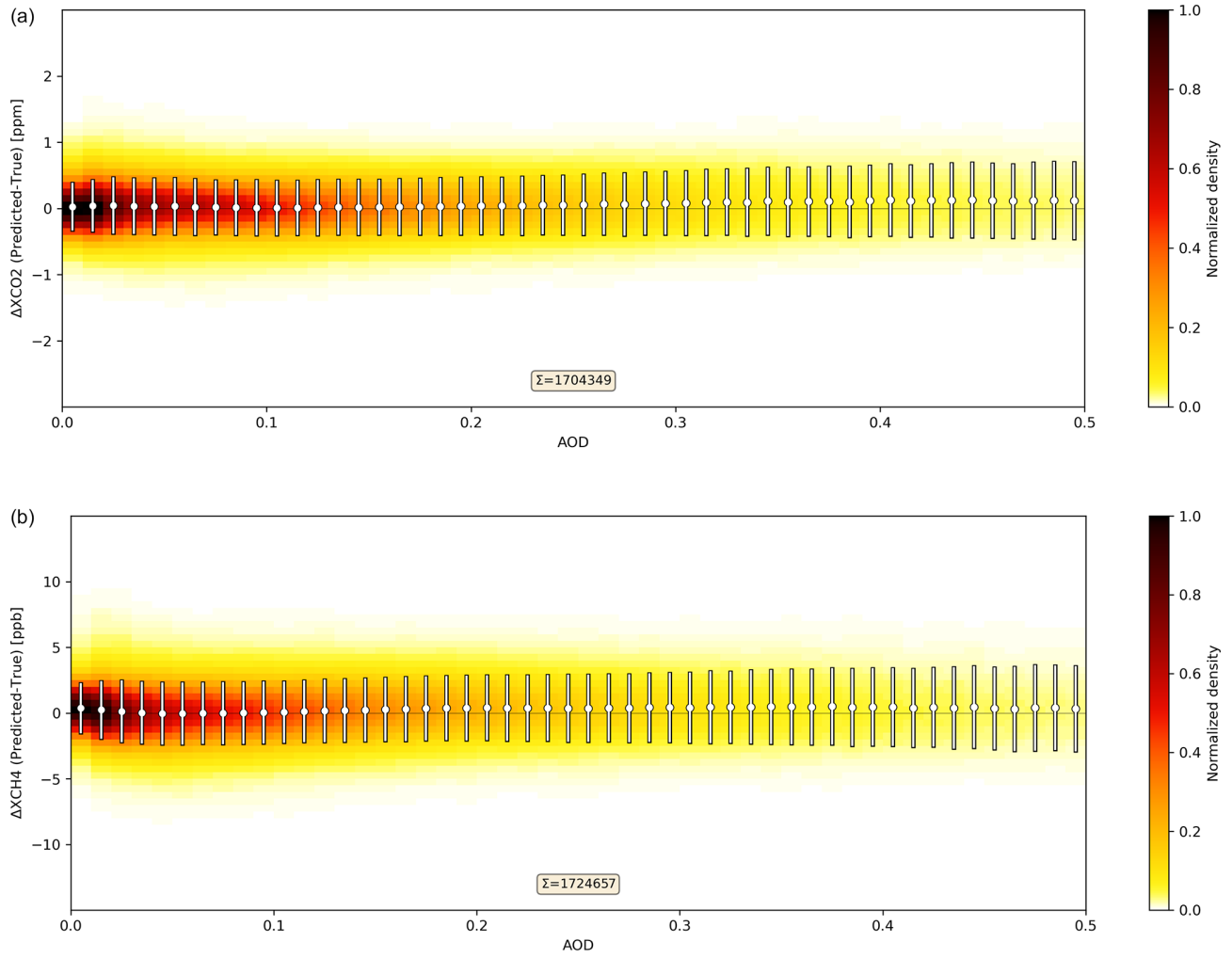

Figure 10. XCO_2 **(a)** and XCH_4 **(b)** prediction error as a function of AOD for noise-free 2020 subset input data. The dots and bars represent the mean and standard deviation, respectively. Σ represents the total number of soundings.

4 Summary and conclusion

In preparation for the analysis of the large number of radiance measurements from the CO2M satellite mission, we developed the computationally efficient ANN-based algorithm NRG-CO2M to retrieve XCO_2 and XCH_4 with high accuracy and precision and high data yield.

It adapts a novel hybrid learning method that is designed to use measured spectra modified to represent a wider range of XCO_2 and XCH_4 values. The approach combines the advantages of simulation-based and measurement-based learning, preserving the characteristics of the real spectra, including instrumental effects, while allowing learning over a wide range of CO_2 and CH_4 concentrations.

It minimizes learning from spurious correlations by dominating the variability of the training data with prescribed artificial variations. However, the method still requires estimates of the true atmospheric concentrations for a representative training data set, which can be obtained similarly to methods used for empirical bias corrections.

It should be noted that the method could be applied to other instruments and applications. In addition to generating representative training data, spectra could also be modified, e.g., to study the ability of a machine learning model to predict changes in its target variable.

Since the CO2M mission will not be launched until 2026, our study is based on simulated measurements from an OSSE. These simulations assume no systematic errors in the training truth, although they do account for stochastic deviations from true concentrations.

Due to the design of the OSSE used, we have focused in this study only on soundings over land in nadir geometry, but the methods presented should be equally applicable to measurements over water surfaces and under glint conditions.

From our analyses of the 2020 subset data, we find that the systematic XCO_2 and XCH_4 errors scatter with a stan-

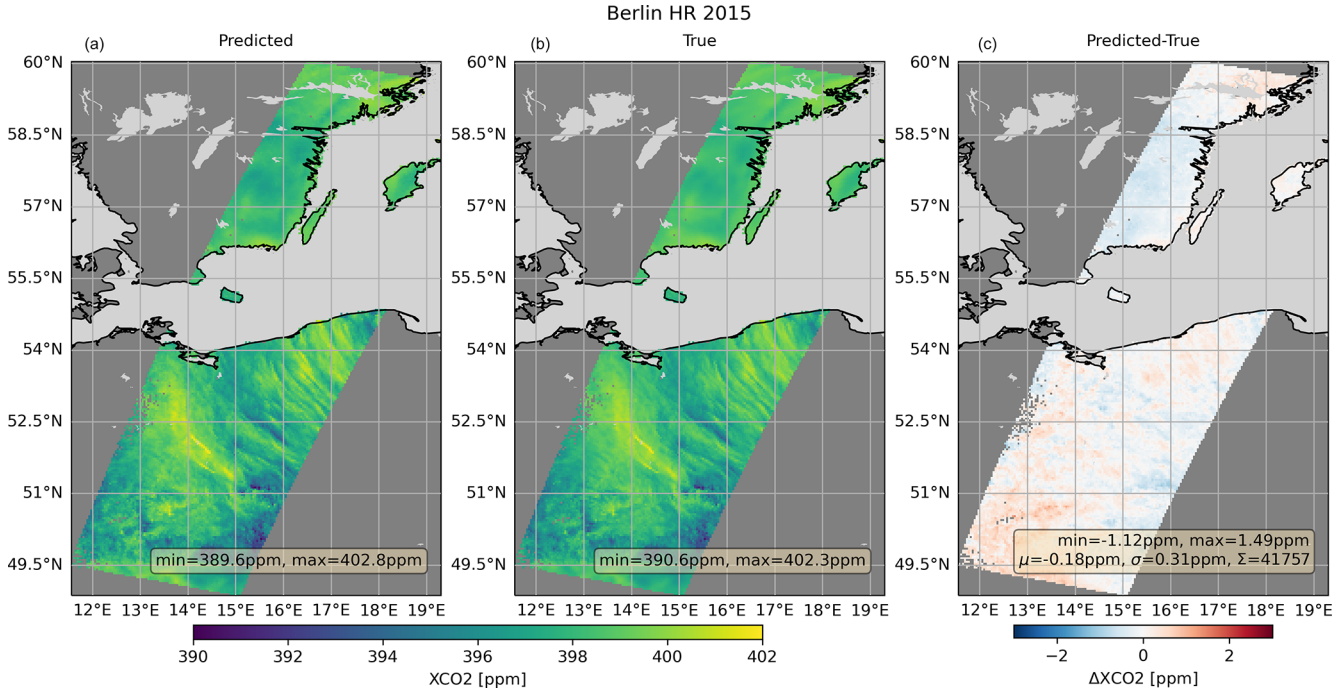

Figure 11. Postprocessed predicted XCO_2 **(a)** and the corresponding true values **(b)**, as well as their difference **(c)** for noise-free Berlin HR input data. μ represents the average prediction error (prediction minus true), σ the standard deviation of the prediction error, and Σ the total number of soundings.

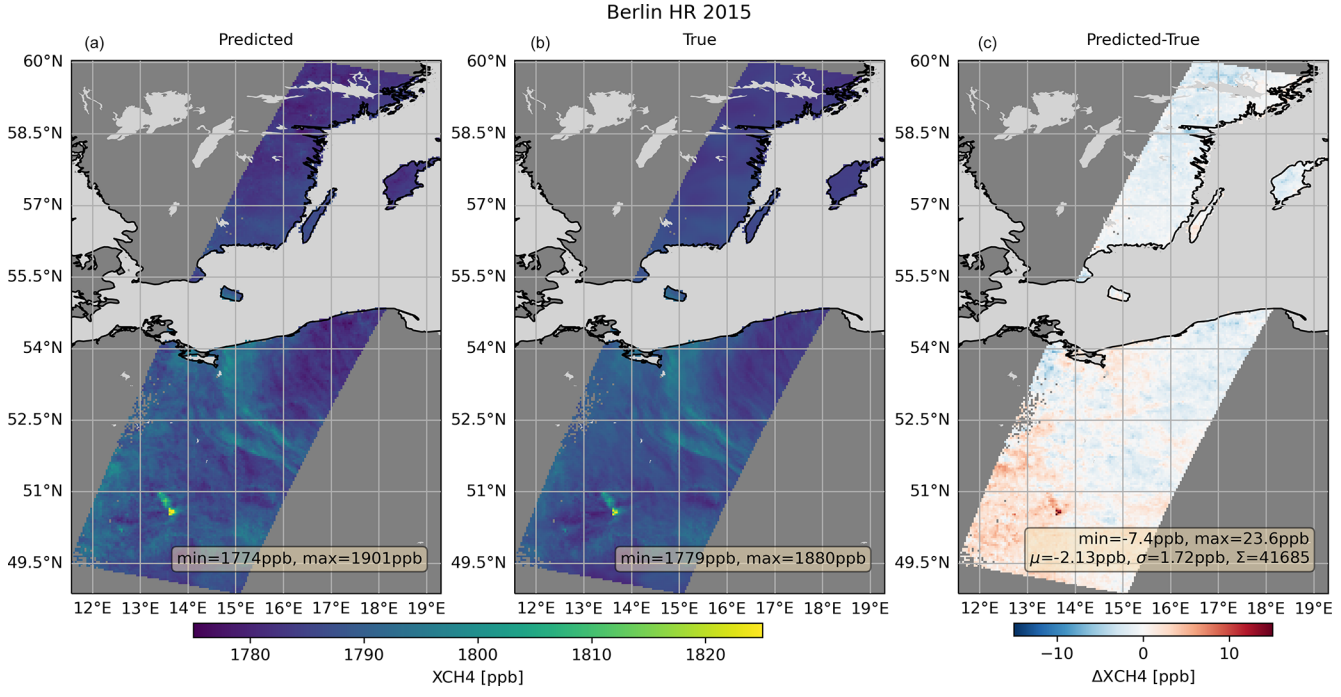

Figure 12. As Fig. 11 but for XCH_4.

dard deviation of 0.44 ppm and 2.45 ppb, respectively. This compares to mission requirements for spatio-temporal systematic errors below 0.5 ppm for XCO_2 and 5 ppb for XCH_4 (MRD; Meijer et al., 2020). The average single sounding precision is 0.41 ppm for XCO_2 and 2.74 ppb for XCH_4 compared to mission requirements for stochastic errors due to instrumental noise of less than 0.7 ppm for XCO_2 and 10 ppb for XCH_4 defined for a specific vegetation scenario (MRD; Meijer et al., 2020). Therefore, we conclude that the proposed retrieval method has the potential to meet the demanding CO2M mission requirements for systematic and stochastic XCO_2 and XCH_4 errors.

Our results are qualitatively similar to those of Noël et al. (2024). They estimated the spatio-temporal systematic errors in their FOCAL setup to be 0.5 ppm for XCO_2 and 3.7 ppb for XCH_4 and the stochastic errors to be 0.5 ppm for XCO_2 and 5.0 ppb for XCH_4. However, unlike Noël et al. (2024), we did not divide the systematic error into long correlation length parts, which are relevant for the application of large-scale surface flux inversions, or short correlation length parts, which are relevant for the application of small-scale (e.g., point source) emission estimation. Our results for the Berlin HR scene illustrate how this affects estimates of the relevant systematic errors. The total systematic error in this scene comprises a variable-part scattering with a standard deviation of 0.28 ppm for XCO_2 and 1.49 ppb for XCH_4 and a scene-wide bias of -0.18 ppm for XCO_2 and -2.12 ppb for XCH_4. However, only the variable part of the systematic error is relevant for the application of small-scale (e.g., point source) emission estimation. It should also be noted that our postprocessing is designed to globally reject about 20 % of the least promising soundings compared to a rejection rate of about 37 % used by Noël et al. (2024).

We trained the ANNs with (modified) spectra from the year 2015. Consequently, it can be expected that the modification error becomes more important the further we deviate from the training year. Nevertheless, we observe that the quality of the prediction erodes only slowly because compared to 2020, the accuracy is only slightly better during the test period in 2015 (0.02 ppm for XCO_2, 0.08 ppb for XCH_4), and the precision is the same. This shows that the introduced spectrum modification method is able to efficiently improve the representativeness of the training data for future concentration years ahead.

We used a conventional XCO_2 and XCH_4 retrieval to modify the spectra used for the training data set. It is a variant of the FOCAL algorithm described by Noël et al. (2024), which takes into account scattering in the atmosphere. However, our results show that using an absorption-only retrieval for this task leads to results with essentially the same accuracy and precision.

As a test, we also trained ANNs without MAP data. This had an apparently small effect on accuracy and precision, which is not consistent with the results of Lu et al. (2022), whose retrieval method became significantly less accurate under these conditions. We can only speculate about possible reasons for this. (i) We use a different aerosol microphysical model, which is consistent with the MACC aerosol model but is less complex than the one used by Lu et al. (2022). (ii) Their CO2I-only retrieval method is fundamentally different from ours and also from FOCAL, which may result in different sensitivities to aerosol-induced biases. In this context, it should be noted that our CO2I-only results are in good agreement with those of Noël et al. (2024), suggesting that it may be possible to meet the CO2M mission requirements without using MAP. (iii) The statistics computed by Lu et al. (2022) to quantify the systematic and stochastic errors differ from those computed by us.

However, we observe that the dependence of the XCO_2 prediction on the dry column increases when MAP is not used, which may introduce systematic errors of the order of 0.1 ppm in reality when perfect knowledge of the dry column cannot be expected. Additionally, removing the NIR band further increased the dependence on the dry column but also the dependence on the a priori, making it less likely to meet the CO2M mission requirements.

It is expected that several thousand CPU cores will be required to continuously analyze the data stream from a single CO2M satellite using conventional full-physics algorithms, which are currently being implemented by EUMETSAT. In comparison, the computational requirements of the presented ANN retrievals, once trained, are negligible and can be considered to be driven by pre- and postprocessing as well as input and output operations.

However, the development of neural networks for retrieval of greenhouse gases from satellite-based measurements of reflected solar radiation in the NIR and SWIR is still in its early stages, while there is much experience with full-physics methods. This includes how they respond to instrumental effects, such as spectral artifacts or temporal changes, and machine learning strategies for bias correction. In addition, it should be noted that the results of Noël et al. (2024) suggest that 1 month of training data may be sufficient for a machine-learning-based bias correction. In contrast, we used 1 year of training data for our ANN-based approach, which can make a difference in the early phase of a satellite mission when little data are available.

In order to use NRG-CO2M to retrieve XCO_2 and XCH_4 as well as the associated uncertainties and averaging kernels from real CO2M radiance measurements, once available, the PCAs and the training of the ANNs would have to be repeated with real data. In this case, the training truth could, e.g., consist of model data confirmed by an ensemble of models, as done for NASA's OCO-2 XCO_2 bias correction (O'Dell et al., 2018), or by corresponding TCCON measurements, as done for FOCAL's GOSAT and GOSAT-2 XCO_2 bias correction (Noël et al., 2021). We expect that at least 1 full year should be used for training, although the modification of the training spectra makes them representative of a wider range of atmospheric conditions.

In the analysis of real data, several effects, detailed investigation of which is beyond the scope of this paper, may lead to somewhat degraded retrieval quality. These include unknown systematic errors in the training truth, a priori, and met profiles; non-ideal sampling of the training data set; and potential instrument or RT features that are not well approximated by our spectrum modification method. Therefore, the actual retrieval quality achievable can only be determined after NRG-CO2M has been trained on and applied to real data.

However, due to the quality achieved in the analysis of synthetic CO2M data, the proposed retrieval algorithm NRG-CO2M can be considered a promising candidate for meeting the high accuracy and precision mission requirements of CO2M while providing high data yield and negligible computational requirements, making it a valuable addition to the ensemble of conventional algorithms.

Appendix A

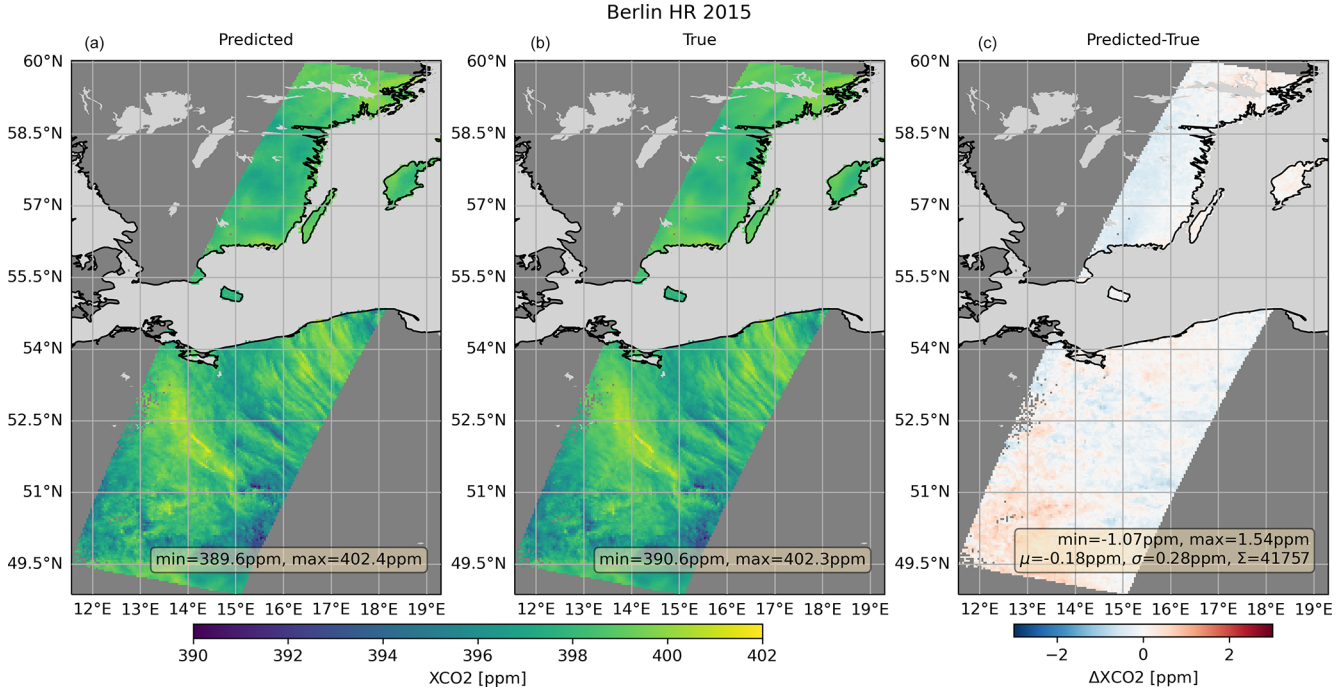

Figure A1. As Fig. 11 but using the true CO_2 concentration profiles as a priori instead of their scene-wide average.

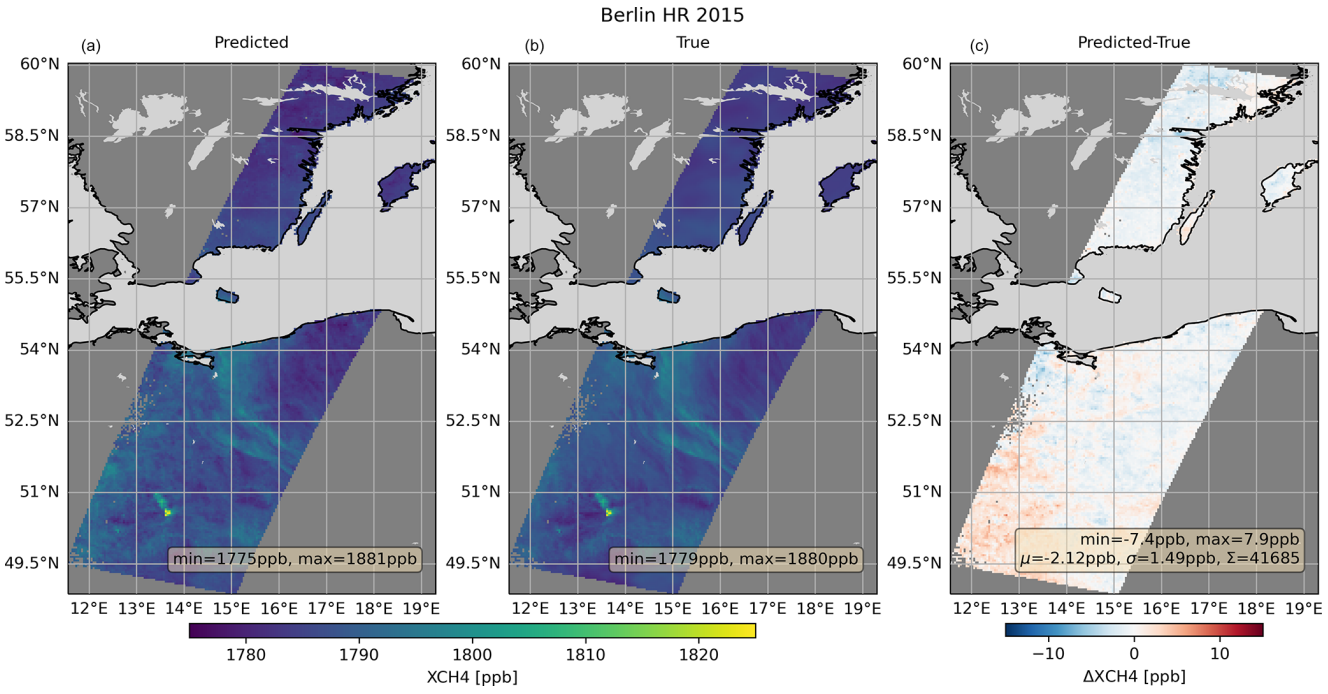

Figure A2. As Fig. 12 but using the true CH_4 concentration profiles as a priori instead of their scene-wide average.

Data availability. We used meteorological data from the ECMWF ERA5 reanalysis (https://doi.org/10.24381/cds.143582cf, Hersbach et al., 2017), aerosol data from CAMS' ECMWF EAC4 atmospheric composition reanalysis (https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4, Inness et al., 2019), CO_2 concentrations from the CAMS global CO_2 atmospheric inversion v20r2 (https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion, Chevallier et al., 2005, 2010; Chevallier, 2013), CH_4 concentrations from the CAMS global CH_4 atmospheric inversion v20r1 (https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion, Segers, 2022), NASA's MODIS surface BRDF and albedo model parameter data set MCD43C1 v6.1 (https://doi.org/10.5067/MODIS/MCD43C1.061, Schaaf and Wang, 2021), and the NASA NDVI data set MYD13C v6.12 (https://doi.org/10.5067/MODIS/MYD13C2.061, Didan, 2021). The CO_2 and CH_4 data sets generated and analyzed in this study are available upon request from the corresponding author.

Author contributions. Experimental setup, data analysis, design and operation of the ANN retrieval, writing of the paper: MR. Design and creation of the OSSE database, design and operation of the FOCAL retrieval: MR, SN, MH. Provision of the CO2M geolocation information and CO2M instrument expertise: RL. Interpretation of the results, improvement of the paper: all authors.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This research was supported by the European Union Copernicus program through EUMETSAT, the Deutsche Forschungsgemeinschaft, and the state, and the University of Bremen. We used meteorological and aerosol data from ECMWF reanalyses, CO_2 and CH_4 concentrations from CAMS global atmospheric inversions, and NASA MODIS surface BRDF, albedo, and NDVI data sets. DeepL (http://www.deepl.com, last access: 9 January 2025) was used for linguistic improvements in an earlier version of the paper. The paper received copy-editing for English from Copernicus Publications.

Financial support. This research has been supported by the European Union Copernicus program through EUMETSAT (grant no. EUM/CO/19/4600002372/RL), the state, and the University of Bremen. Parts of the calculations reported here were performed on HPC facilities of the IUP, University of Bremen, funded by the Deutsche Forschungsgemeinschaft (grant nos. INST 144/379-1 and INST 144/493-1).

The article processing charges for this open-access publication were covered by the University of Bremen.

Review statement. This paper was edited by Peer Nowack and reviewed by Christopher O'Dell and one anonymous referee.

References

Agustí-Panareda, A., McNorton, J., Balsamo, G., Baier, B. C., Bousserez, N., Boussetta, S., Brunner, D., Chevallier, F., Choulga, M., Diamantakis, M., Engelen, R., Flemming, J., Granier, C., Guevara, M., Denier van der Gon, H., Elguindi, N., Haussaire, J.-M., Jung, M., Janssens-Maenhout, G., Kivi, R., Massart, S., Papale, D., Parrington, M., Razinger, M., Sweeney, C., Vermeulen, A., and Walther, S.: Global nature run data with realistic high-resolution carbon weather for the year of the Paris Agreement, Scientific Data, 9, 160, https://doi.org/10.1038/s41597-022-01228-2, 2022.

Aires, F., Rossow, W. B., Scott, N. A., and Chédin, A.: Remote sensing from the infrared atmospheric sounding interferometer instrument 1. Compression, denoising, and first-guess retrieval algorithms, J. Geophys. Res.-Atmos., 107, 4619, https://doi.org/10.1029/2001jd000955, 2002.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Dentener, F., Wagner, T., Platt, U., Kaplan, J. O., Körner, S., Heimann, M., Dlugokencky, E. J., and Goede, A.: Satellite chartography of atmospheric methane from SCIAMACHY onboard ENVISAT: 2. Evaluation based on inverse model simulations, J. Geophys. Res., 112, D02304, https://doi.org/10.1029/2006JD007268, 2007.

Bishop, C. M.: Neural networks for pattern recognition, Clarendon Press, Oxford, reprinted edition, ISBN 0198538499, 1996.

Boesch, H. and Di Noia, A.: Algorithm Theoretical Basis Document(ATBD) – ANNEX A for productsCO2_GOS_OCFP (v7.3), CH4_GOS_OCFP(v7.3) & CH4_GOS_OCPR (v9.0) (CDR6,2009-2021), C3S2_312a_Lot2_DLR – Atmosphere, Tech. rep., Copernicus Climate Change Service, http://wdc.dlr.de/C3S_312b_Lot2/Documentation/GHG/C3S2_312a_Lot2_ATBD_GHG_A_latest.pdf (last access: 10 January 2025), 2023.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A.: 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, 1999.

Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications, Atmos. Meas. Tech., 3, 781–811, https://doi.org/10.5194/amt-3-781-2010, 2010.

Bréon, F.-M., David, L., Chatelanaz, P., and Chevallier, F.: On the potential of a neural-network-based approach for estimating XCO_2 from OCO-2 measurements, Atmos. Meas. Tech., 15, 5219–5234, https://doi.org/10.5194/amt-15-5219-2022, 2022.

Broquet, G., Bréon, F.-M., Renault, E., Buchwitz, M., Reuter, M., Bovensmann, H., Chevallier, F., Wu, L., and Ciais, P.: The potential of satellite spectro-imagery for monitoring CO_2 emissions from large cities, Atmos. Meas. Tech., 11, 681–708, https://doi.org/10.5194/amt-11-681-2018, 2018.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Löscher, A.: Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric CO_2 and CH_4 retrieval errors by error parameterization, Atmos. Meas. Tech., 6, 3477–3500, https://doi.org/10.5194/amt-6-3477-2013, 2013.

Burrows, J. P., Hölzle, E., Goede, A. P. H., Visser, H., and Fricke, W.: SCIAMACHY – Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, Acta Astronaut., 35, 445–451, 1995.

Chevallier, F.: On the parallelization of atmospheric inversions of CO_2 surface fluxes within a variational framework, Geosci. Model Dev., 6, 783–790, https://doi.org/10.5194/gmd-6-783-2013, 2013.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F., Chédin, A., and Ciais, P.: 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, 2005 (data available at: https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion, last access: 9 January 2025).

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, D21307, https://doi.org/10.1029/2010jd013887, 2010.

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J. F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, J. Geophys. Res.-Atmos., 117, D21301, https://doi.org/10.1029/2012JD018087, 2012.

Crevoisier, C.: Algorithm Theoretical Basis Document (ATBD) – ANNEX E for IASI CO_2 and CH_4 (v9.1) and AIRS CO_2 mid-tropospheric products, C3S2_312a_Lot2_DLR – Atmosphere, Tech. rep., Copernicus Climate Change Service, http://wdc.dlr.de/C3S_312b_Lot2/Documentation/GHG/C3S2_312a_Lot2_ATBD_GHG_E_latest.pdf (last access: 10 January 2025), 2023.

Crisp, D., Atlas, R. M., Bréon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. S., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, P., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Adv. Space Res., 34, 700–709, 2004.

David, L., Bréon, F.-M., and Chevallier, F.: XCO_2 estimates from the OCO-2 measurements using a neural network approach, At-

mos. Meas. Tech., 14, 117–132, https://doi.org/10.5194/amt-14-117-2021, 2021.

Didan, K.: MODIS/Aqua Vegetation Indices Monthly L3 Global 0.05Deg CMG V061, NASA EOSDIS Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MYD13C2.061, 2021.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmos. Meas. Tech., 8, 281–299, https://doi.org/10.5194/amt-8-281-2015, 2015.

Guerlet, S., Butz, A., Schepers, D., Basu, S., Hasekamp, O. P., Kuze, A., Yokota, T., Blavier, J.-F., Deutscher, N. M., Griffith, D. W. T., Hase, F., Kyro, E., Morino, I., Sherlock, V., Sussmann, R., Galli, A., and Aben, I.: Impact of aerosol and thin cirrus on retrieving and validating XCO_2 from GOSAT shortwave infrared measurements, J. Geophys. Res., 118, 4887–4905, https://doi.org/10.1002/jgrd.50332, 2013.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service (C3S) Data Store (CDS) [data set], https://doi.org/10.24381/cds.143582cf, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019 (data available at: https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4, last access: 9 January 2025).

Jacobs, N., O'Dell, C. W., Taylor, T. E., Logan, T. L., Byrne, B., Kiel, M., Kivi, R., Heikkinen, P., Merrelli, A., Payne, V. H., and Chatterjee, A.: The importance of digital elevation model accuracy in X_{CO_2} retrievals: improving the Orbiting Carbon Observatory 2 Atmospheric Carbon Observations from Space version 11 retrieval product, Atmos. Meas. Tech., 17, 1375–1401, https://doi.org/10.5194/amt-17-1375-2024, 2024.

Jolliffe, I. T. and Cadima, J.: Principal component analysis: a review and recent developments, Philos. T. Roy. Soc. A, 374, 20150202, https://doi.org/10.1098/rsta.2015.0202, 2016.

Kiel, M., O'Dell, C. W., Fisher, B., Eldering, A., Nassar, R., MacDonald, C. G., and Wennberg, P. O.: How bias correction goes wrong: measurement of X_{CO_2} affected by erroneous surface pressure estimates, Atmos. Meas. Tech., 12, 2241–2259, https://doi.org/10.5194/amt-12-2241-2019, 2019.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv [preprint], https://doi.org/10.48550/arxiv.1412.6980, 22 December 2014.

Knapp, M., Kleinschek, R., Hase, F., Agustí-Panareda, A., Inness, A., Barré, J., Landgraf, J., Borsdorff, T., Kinne, S., and Butz, A.: Shipborne measurements of XCO_2, XCH_4, and XCO above the Pacific Ocean and comparison to CAMS atmospheric analyses and S5P/TROPOMI, Earth Syst. Sci. Data, 13, 199–211, https://doi.org/10.5194/essd-13-199-2021, 2021.

Krasnopolsky, V. M. and Schiller, H.: Some neural network applications in environmental sciences. Part I: forward and inverse problems in geophysical remote measurements, Neural Networks, 16, 321–334, https://doi.org/10.1016/S0893-6080(03)00027-3, 2003.

Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., Chevallier, F., Sherlock, V., Buchwitz, M., Osterman, G., Miller, C. E., Wennberg, P. O., Griffith, D., Morino, I., Dubey, M. K., Deutscher, N. M., Notholt, J., Hase, F., Warneke, T., Sussmann, R., Robinson, J., Strong, K., Schneider, M., De Mazière, M., Shiomi, K., Feist, D. G., Iraci, L. T., and Wolf, J.: Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON, Atmos. Meas. Tech., 9, 683–709, https://doi.org/10.5194/amt-9-683-2016, 2016.

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

Lespinas, F., Wang, Y., Broquet, G., Bréon, F.-M., Buchwitz, M., Reuter, M., Meijer, Y., Loescher, A., Janssens-Maenhout, G., Zheng, B., and Ciais, P.: The potential of a constellation of low earth orbit satellite imagers to monitor worldwide fossil fuel CO_2 emissions from large cities and point sources, Carbon Balance and Management, 15, 18, https://doi.org/10.1186/s13021-020-00153-4, 2020.

Liu, X., Smith, W. L., Zhou, D. K., and Larar, A.: Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept, Appl. Optics, 45, 201–209, https://doi.org/10.1364/AO.45.000201, 2006.

Lu, S., Landgraf, J., Fu, G., van Diedenhoven, B., Wu, L., Rusli, S. P., and Hasekamp, O. P.: Simultaneous Retrieval of Trace Gases, Aerosols, and Cirrus Using RemoTAP – The Global Orbit Ensemble Study for the CO2M Mission, Front. Remote Sens., 3, 914378, https://doi.org/10.3389/frsen.2022.914378, 2022.

Meijer, Y., Bösch, H., Bombelli, A., Brunner, D., Buchwitz, M., Ciais, P., Crisp, D., Engelen, R., Holmund, K., Houweling, S., Janssen-Maenhout, G., Marshall, J., Nakajima, M., Pinty, B., Scholze, M., Bezy, J., Drinkwater, M., Fehr, T., Fernandez, V., Löscher, A., Nett, H., Sierk, B., Dubovik, O., Landgraf, J., Lang, R., Lindqvist, H., Tamminen, J., and Veefkind, P.:

Copernicus CO_2 Monitoring Mission Requirements Document, Tech. rep., ESA Earth and Mission Science Division, https://esamultimedia.esa.int/docs/EarthObservation/CO2M_MRD_v3.0_20201001_Issued.pdf (last access: 10 January 2025), 2020.

Mohebali, B., Tahmassebi, A., Meyer-Baese, A., and Gandomi, A. H.: Probabilistic neural networks: a brief overview of theory, implementation, and application, in: Handbook of Probabilistic Models, edited by: Samui, P., Tien Bui, D., Chakraborty, S., and Deo, R. C., Butterworth-Heinemann, ISBN 978-0-12-816514-0, 347–367, https://doi.org/10.1016/B978-0-12-816514-0.00014-X, 2020.

Noël, S., Reuter, M., Buchwitz, M., Borchardt, J., Hilker, M., Bovensmann, H., Burrows, J. P., Di Noia, A., Suto, H., Yoshida, Y., Buschmann, M., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Podolske, J. R., Pollard, D. F., Sha, M. K., Shiomi, K., Sussmann, R., Té, Y., Velazco, V. A., and Warneke, T.: XCO_2 retrieval for GOSAT and GOSAT-2 based on the FOCAL algorithm, Atmos. Meas. Tech., 14, 3837–3869, https://doi.org/10.5194/amt-14-3837-2021, 2021.

Noël, S., Reuter, M., Buchwitz, M., Borchardt, J., Hilker, M., Schneising, O., Bovensmann, H., Burrows, J. P., Di Noia, A., Parker, R. J., Suto, H., Yoshida, Y., Buschmann, M., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Kivi, R., Liu, C., Morino, I., Notholt, J., Oh, Y.-S., Ohyama, H., Petri, C., Pollard, D. F., Rettinger, M., Roehl, C., Rousogenous, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Vrekoussis, M., and Warneke, T.: Retrieval of greenhouse gases from GOSAT and GOSAT-2 using the FOCAL algorithm, Atmos. Meas. Tech., 15, 3401–3437, https://doi.org/10.5194/amt-15-3401-2022, 2022.

Noël, S., Buchwitz, M., Hilker, M., Reuter, M., Weimer, M., Bovensmann, H., Burrows, J. P., Bösch, H., and Lang, R.: Greenhouse gas retrievals for the CO2M mission using the FOCAL method: first performance estimates, Atmos. Meas. Tech., 17, 2317–2334, https://doi.org/10.5194/amt-17-2317-2024, 2024.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

RAL: Provision of Top-Of-Atmosphere simulations for the evaluation of data processing for the CO_2 monitoring mission: Task 1, 2, and 3 Report, Tech. rep., RAL Space Remote Sensing Group, https://www-cdn.eumetsat.int/files/2023-01/RAL_EUM_CO2Msims_TN123_v1p1.pdf (last access: 10 January 2025), 2022.

Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H., and Burrows, J. P.: A method for improved SCIAMACHY CO_2 retrieval in the presence of optically thin clouds, Atmos. Meas. Tech., 3, 209–232, https://doi.org/10.5194/amt-3-209-2010, 2010.

Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M., Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri, C., Robinson, J., Schneising, O., Sherlock, V., Velazco, V., Warneke, T., Wennberg, P. O., and Wunch, D.: Retrieval of atmospheric CO_2 with enhanced accuracy and precision from SCIAMACHY: Validation with FTS measurements and comparison with model results, J. Geophys. Res., 116, D04301, https://doi.org/10.1029/2010JD015047, 2011.

Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Bovensmann, H., and Burrows, J. P.: A fast atmospheric trace gas retrieval for hyperspectral instruments approximating multiple scattering – Part 2: application to XCO_2 retrievals from OCO-2, Remote Sens., 9, 1102, https://doi.org/10.3390/rs9111102, 2017a.

Reuter, M., Buchwitz, M., Schneising, O., Noël, S., Rozanov, V., Bovensmann, H., and Burrows, J. P.: A fast atmospheric trace gas retrieval for hyperspectral instruments approximating multiple scattering – Part 1: radiative transfer and a potential OCO-2 XCO_2 retrieval setup, Remote Sens., 9, 1159, https://doi.org/10.3390/rs9111159, 2017b.

Rojas, R.: Neural Networks – A Systematic Introduction, Springer, Berlin/Heidelberg, 1st edn., ISBN 9783642610684, 1996.

Rozanov, V., Dinter, T., Rozanov, A., Wolanin, A., Bracher, A., and Burrows, J.: Radiative transfer modeling through terrestrial atmosphere and ocean accounting for inelastic processes: Software package SCIATRAN, J. Quant. Spectrosc. Ra., 194, 65–85, https://doi.org/10.1016/j.jqsrt.2017.03.009, 2017.

Salstein, D. A., Ponte, R. M., and Cady-Pereira, K.: Uncertainties in atmospheric surface pressure fields from global analyses, J. Geophys. Res.-Atmos., 113, D14107, https://doi.org/10.1029/2007jd009531, 2008.

Schaaf, C. and Wang, Z.: MODIS/Terra+Aqua BRDF/Albedo-Model Parameters Daily L3 Global 0.05Deg CMG V061, NASA EOSDIS Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MCD43C1.061, 2021.

Schneising, O., Heymann, J., Buchwitz, M., Reuter, M., Bovensmann, H., and Burrows, J. P.: Anthropogenic carbon dioxide source areas observed from space: assessment of regional enhancements and trends, Atmos. Chem. Phys., 13, 2445–2454, https://doi.org/10.5194/acp-13-2445-2013, 2013.

Schneising, O., Reuter, M., Buchwitz, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Terrestrial carbon sink observed from space: variation of growth rates and seasonal cycle amplitudes in response to interannual surface temperature variability, Atmos. Chem. Phys., 14, 133–141, https://doi.org/10.5194/acp-14-133-2014, 2014.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simulta-

neously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech., 12, 6771–6802, https://doi.org/10.5194/amt-12-6771-2019, 2019.

Schneising, O., Buchwitz, M., Hachmeister, J., Vanselow, S., Reuter, M., Buschmann, M., Bovensmann, H., and Burrows, J. P.: Advances in retrieving XCH_4 and XCO from Sentinel-5 Precursor: improvements in the scientific TROPOMI/WFMD algorithm, Atmos. Meas. Tech., 16, 669–694, https://doi.org/10.5194/amt-16-669-2023, 2023.

Segers, A.: Description of the CH_4 Inversion Production Chain, Copernicus Atmosphere Monitoring Service, https://atmosphere.copernicus.eu/sites/default/files/2022-10/CAMS255_2021SC1_D55.5.2.1-2021CH4_202206_production_chain_CH4_v1.pdf (last access: 10 January 2025), 2022 (data available: https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion, last access: 9 January 2025).

Sierk, B., Fernandez, V., Bézy, J.-L., Meijer, Y., Durand, Y., Bazalgette Courrèges-Lacoste, G., Pachot, C., Löscher, A., Nett, H., Minoglou, K., Boucher, L., Windpassinger, R., Pasquet, A., Serre, D., and te Hennepe, F.: The Copernicus CO2M mission for monitoring anthropogenic carbon dioxide emissions from space, in: International Conference on Space Optics – ICSO 2020, 30 March–2 April 2021, online only, edited by: Sodnik, Z., Cugny, B., and Karafolas, N., SPIE, https://doi.org/10.1117/12.2599613, 2021.

Suto, H., Kataoka, F., Kikuchi, N., Knuteson, R. O., Butz, A., Haun, M., Buijs, H., Shiomi, K., Imai, H., and Kuze, A.: Thermal and near-infrared sensor for carbon observation Fourier transform spectrometer-2 (TANSO-FTS-2) on the Greenhouse gases Observing SATellite-2 (GOSAT-2) during its first year in orbit, Atmos. Meas. Tech., 14, 2013–2039, https://doi.org/10.5194/amt-14-2013-2021, 2021.

Tu, Q., Hase, F., Blumenstock, T., Kivi, R., Heikkinen, P., Sha, M. K., Raffalski, U., Landgraf, J., Lorente, A., Borsdorff, T., Chen, H., Dietrich, F., and Chen, J.: Intercomparison of atmospheric CO_2 and CH_4 abundances on regional scales in boreal areas using Copernicus Atmosphere Monitoring Service (CAMS) analysis, COllaborative Carbon Column Observing Network (COCCON) spectrometers, and Sentinel-5 Precursor satellite observations, Atmos. Meas. Tech., 13, 4751–4771, https://doi.org/10.5194/amt-13-4751-2020, 2020.

UNFCCC: Paris Agreement, United Nations Framework Convention on Climate Change, https://unfccc.int/sites/default/files/english_paris_agreement.pdf (last access: 10 January 2025), 2015.

Velazco, V. A., Buchwitz, M., Bovensmann, H., Reuter, M., Schneising, O., Heymann, J., Krings, T., Gerilowski, K., and Burrows, J. P.: Towards space based verification of CO_2 emissions from strong localized sources: fossil fuel power plant emissions as seen by a CarbonSat constellation, Atmos. Meas. Tech., 4, 2809–2822, https://doi.org/10.5194/amt-4-2809-2011, 2011.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO_2 total columns from space, Atmos. Chem. Phys., 11, 12317–12337, https://doi.org/10.5194/acp-11-12317-2011, 2011.

Xie, F., Ren, T., Zhao, C., Wen, Y., Gu, Y., Zhou, M., Wang, P., Shiomi, K., and Morino, I.: Fast retrieval of XCO_2 over east Asia based on Orbiting Carbon Observatory-2 (OCO-2) spectral measurements, Atmos. Meas. Tech., 17, 3949–3967, https://doi.org/10.5194/amt-17-3949-2024, 2024.