



## Article

# Spectral Fingerprinting of Methane from Hyper-Spectral Sounder Measurements Using Machine Learning and Radiative Kernel-Based Inversion

Wan Wu <sup>1,\*</sup>, Xu Liu <sup>1</sup>, Xiaozhen Xiong <sup>1</sup>, Qiguang Yang <sup>2</sup>, Lihang Zhou <sup>3</sup>, Liqiao Lei <sup>2</sup>, Daniel K. Zhou <sup>1</sup> and Allen M. Larar <sup>1</sup>

<sup>1</sup> NASA Langley Research Center, Hampton, VA 23681, USA; xu.liu-1@nasa.gov (X.L.);

xiaozhen.xiong@nasa.gov (X.X.); daniel.k.zhou@nasa.gov (D.K.Z.); allen.m.larar@nasa.gov (A.M.L.)

<sup>2</sup> ADNET Systems Inc., Bethesda, MD 20817, USA; qiguang.yang@nasa.gov (Q.Y.); liqiao.lei@nasa.gov (L.L.)

<sup>3</sup> NOAA JPSS Program Office, Silver Spring, MD 20706, USA; lihang.zhou@noaa.gov

\* Correspondence: wan.wu@nasa.gov

**Abstract:** Satellite-based hyper-spectral infrared (IR) sensors such as the Atmospheric Infrared Sounder (AIRS), the Cross-track Infrared Sounder (CrIS), and the Infrared Atmospheric Sounding Interferometer (IASI) cover many methane (CH<sub>4</sub>) spectral features, including the ν<sub>1</sub> vibrational band near 1300 cm<sup>-1</sup> (7.7 μm); therefore, they can be used to monitor CH<sub>4</sub> concentrations in the atmosphere. However, retrieving CH<sub>4</sub> remains a challenge due to the limited spectral information provided by IR sounder measurements. The information required to resolve the weak absorption lines of CH<sub>4</sub> is often obscured by interferences from signals originating from other trace gases, clouds, and surface emissions within the overlapping spectral region. Consequently, currently available CH<sub>4</sub> data product derived from IR sounder measurements still have large errors and uncertainties that limit their application scope for high-accuracy climate and environment monitoring applications. In this paper, we describe the retrieval of atmospheric CH<sub>4</sub> profiles using a novel spectral fingerprinting methodology and our evaluation of performance using measurements from the CrIS sensor aboard the Suomi National Polar-orbiting Partnership (SNPP) satellite. The spectral fingerprinting methodology uses optimized CrIS radiances to enhance the CH<sub>4</sub> signal and a machine learning classifier to constrain the physical inversion scheme. We validated our results using the atmospheric composition reanalysis results and data from airborne in situ measurements. An inter-comparison study revealed that the spectral fingerprinting results can capture the vertical variation characteristics of CH<sub>4</sub> profiles that operational sounder products may not provide. The latitudinal variations in CH<sub>4</sub> concentration in these results appear more realistic than those shown in existing sounder products. The methodology presented herein could enhance the utilization of satellite data to comprehend methane's role as a greenhouse gas and facilitate the tracking of methane sources and sinks with increased reliability.

**Keywords:** methane; hyper-spectral infrared sounder measurements; CrIS; retrieval; spectral fingerprinting



**Citation:** Wu, W.; Liu, X.; Xiong, X.; Yang, Q.; Zhou, L.; Lei, L.; Zhou, D.K.; Larar, A.M. Spectral Fingerprinting of Methane from Hyper-Spectral Sounder Measurements Using Machine Learning and Radiative Kernel-Based Inversion. *Remote Sens.* **2024**, *16*, 578. <https://doi.org/10.3390/rs16030578>

Academic Editor: Carmine Serio

Received: 8 December 2023

Revised: 24 January 2024

Accepted: 1 February 2024

Published: 2 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As a significant greenhouse gas with 28 times greater global warming potential than carbon dioxide [1], the recognition of the importance of monitoring methane (CH<sub>4</sub>) in the atmosphere on a global scale has increased. Major absorption bands of CH<sub>4</sub> are located in the short-wave IR (centered at ~1.65 μm and ~3.3 μm) and mid-IR (centered ~7.7 μm) spectral regions, respectively. Spaceborne instruments like the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) aboard the European Space Agency (ESA)'s Environmental Satellite (ENVISAT), the Thermal and Near-infrared Sensor for Carbon Observation–Fourier Transform Spectrometer (TANSO-FTS) aboard the Greenhouse Gases Observing Satellite (GOSAT), and the Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 Precursor satellite (Sentinel-5P) measure the

absorption of solar radiation by CH<sub>4</sub> in the short-wave IR band and provide near-surface sensitivity [2–4]. Both the SCIAMACHY and GOSAT measure strong CH<sub>4</sub> absorption signals around 1.65 μm. The TROPOMI measures CH<sub>4</sub> mixing ratios using the absorption information from the Oxygen-A Band (760 nm) and the short-wave IR band (2.3–2.4 μm). GOSAT also provides measurements in the thermal infrared band between 5.5 and 14.3 μm, which is used to derive thermal IR CH<sub>4</sub> data [5,6]. Compared with short-wave IR observations, mid-wave IR spectral measurements in the 7.7 μm mid-IR band are more sensitive to mid to upper tropospheric CH<sub>4</sub>. IR sounders like the AIRS, IASI, and CrIS all measure the absorption of thermal IR radiation in the 7.7 μm band and therefore can provide important complementary information needed for global CH<sub>4</sub> monitoring.

Existing IR sounders have limited measurement sensitivity and instrument spectral resolution. Consequently, fully resolving weak CH<sub>4</sub> signals becomes difficult, impacting the accuracy of both total concentration and vertical profile distribution of the retrieved atmospheric CH<sub>4</sub>. CH<sub>4</sub> absorption lines in the mid-IR band overlap with those of other gases, such as water vapor and nitrous oxide (N<sub>2</sub>O), further introducing difficulties in isolating and accurately quantifying CH<sub>4</sub> concentrations. Retrieval studies using the AIRS and IASI have shown that the retrieved upper tropospheric CH<sub>4</sub> can easily have a bias error ranging from 1 to 4% [7,8]. Studies have shown inconsistencies between the CH<sub>4</sub> profiles retrieved from AIRS and GOSAT TIR measurements, as well as inconsistencies between the CH<sub>4</sub> profiles retrieved from AIRS and IASI measurements [9,10]. And the global spatial distribution of these CH<sub>4</sub> retrieval products remains to be validated.

CH<sub>4</sub> retrieval studies using hyper-spectral IR sounders measurements have been mostly based on the optimal estimation methodology (OEM) following Rodgers' formalism [11]. A priori knowledge of CH<sub>4</sub> is critically needed to complement information content from the IR sounder measurements. Xiong et al., used latitudinal-dependent CH<sub>4</sub> first-guess profiles derived from the monthly averaged results of in situ aircraft observations, ground-based flask network measurements, satellite observations, and the atmospheric transport model TM3 [7]. García et al., retrieved CH<sub>4</sub> from IASI measurements using mean profiles from WACCM (Whole 30 Atmosphere Community Climate Model-version 5, <https://www2.acom.ucar.edu/gcm/waccm>), averaged on a 1.9° × 2.5° grid for the 2004–2006 period, as the first-guess profiles [12]. An ad hoc Tikhonov–Philips slope constraint is used to maintain the vertical shape of the CH<sub>4</sub> profiles during the retrieval process. De Wachter et al. also used the a priori profile and the covariance matrix constraint derived from WACCM, but a single global climatological a priori was used in their study [8]. Siddans et al. used a fixed value of 1.75 ppmv in the troposphere as the a priori state and two years of zonal mean values from the TOMCAT chemical transport model to construct a priori error statistics for IASI CH<sub>4</sub> retrievals [13]. Razavi et al., used the a priori derived from the “Laboratoire de Météorologie Dynamique” global climate model that constraints north–south latitudinal gradients of CH<sub>4</sub> distribution [14].

OEM usually assumes a Gaussian distribution of the possible solution around its a priori state, as well as a linear relationship between the change in CH<sub>4</sub> concentration from the a priori state and the associated change in observed radiance. Considering the complexity of the inverse relationship to be established for CH<sub>4</sub> retrieval, Crevoisier et al. used a neural network-based non-linear inference method to retrieve CH<sub>4</sub> from IASI observations [15]. The neural network-based approach, in theory, can be used to represent a non-linear training-prediction relationship, but the prediction accuracy largely depends on the representativeness of the training samples. The neural network-based approach also lacks the error estimation that is provided by the OEM-based scheme. The approach was trained using simulated data. The difference between simulated radiances and the observations was simply addressed via a bias correction based on one year of data over the tropical region. Considering the scene-dependent nature of the simulation error, large uncertainty between the simulation and the observation likely remains and inevitably contributes to the retrieval error.

In this paper, we present a novel CH<sub>4</sub> retrieval methodology based on the spectral fingerprinting approach. Such an approach utilizes spectral information from IR sounder measurements to derive a scene-dependent a priori state and corresponding constraint for each individual CH<sub>4</sub> retrieval. Well-defined a priori information will greatly improve the linearity of the inverse relationship between the atmospheric CH<sub>4</sub> and the spectral radiances observed, as well as the Gaussian distribution characteristics of the solution around the a priori state. The spectral fingerprinting scheme is based on a pre-constructed database that comprises an ample set of representative reference states, along with their corresponding radiative kernels. In this approach, a clustering method based on machine learning is employed to stratify and identify the scene-dependent a priori state within the pre-constructed database. Spectral radiances serve as the predictors. The a priori state is identified via radiance spectral matching, and the corresponding radiative kernel is then used to find the fingerprinting solution via an OEM-based linear inversion scheme. Details about the spectral fingerprinting scheme will be introduced, with in-depth technical insights into the construction of the fingerprinting database also being provided. The application of the fingerprinting methodology on SNPP CrIS observations will be demonstrated with the derived CH<sub>4</sub> profiles validated using both CH<sub>4</sub> reanalysis data from the Copernicus Atmosphere Monitoring Service (CAMS) and in situ measurement data from the Atmospheric Tomography Mission (ATom) [16].

The CH<sub>4</sub> fingerprinting algorithm study presented herein contributes to the efforts to improve the single field-of-view sounder atmospheric product (SiFSAP) [17]. As a novel algorithm developed to complement other sounder products, such as CLIMCAPS and AIRS version 7 [18,19], the SiFSAP system produces hyper-spectral IR sounder Level 2 data at an instrument-native spatial resolution and provides high-accuracy spectral fitting to the top of the atmosphere (TOA) sounder observations by directly simulating cloud scattering. In order to maximize the information content from the measurements (i.e., minimized use of a priori information), the global climatology-based a priori constraint is used in the SiFSAP retrieval algorithm. We purposely designed the SiFSAP algorithm using relaxed a priori global climatology so that the retrieval results are more sensitive to the small climate signals caught by the TOA spectral radiance observations. High-quality SiFSAP Level 2 products of atmospheric temperature, water vapor, and other trace gases such as ozone (O<sub>3</sub>) have been used for various dynamics and climate studies [20]. However, the uncertainty in CH<sub>4</sub> data at a localized, instantaneous scale can be potentially large because the CH<sub>4</sub> information provided by the measurements can be very limited for a significant percentage of cases. The fingerprinting algorithm can be used to generate a high-quality first-guess and scene-dependent a priori covariance constraint to improve the CH<sub>4</sub> retrieval in the SiFSAP system for near-real-time applications. It will be implemented to produce the next version of SiFSAP.

## 2. The Spectral Fingerprinting Methodology

The spectral fingerprinting methodology has been widely used for the characterization and quantification of biological materials, chemical components, mineral analysis, and remote sensing [21–24]. The concept is based on the fact that the spectral feature of a measured signal can be used for the constitutive component analysis of target samples. The analysis usually involves the classification of spectral features associated with known constituents so that the constitutive component can be identified by characterizing the similarity between the measured signal and the spectral features of a prescribed constituent. When the measured spectral signal  $r$  of a target sample matches a known reference spectrum, the constituents of the target sample  $x$  are instantly identified using the reference database. In broad terms, spectral fingerprinting integrates both spectral classification and spectral matching procedures. Recent corresponding research in the field of remote sensing has predominantly centered on the application of spectral fingerprinting in the analysis of hyperspectral images [25–27].

If the target signal is composed of multiple constituents, the spectral fingerprinting also involves the decomposition of the total measured signal into its different spectral components of individual constituents. The technique has been used for the optimal detection and attribution of climate change signals in the outgoing spectra of TOA radiances [28–32]. In those studies, the spectral fingerprints are the anomalies of outgoing spectral radiances, and the constituents are the changes in the TOA spectra that are associated with different feedbacks and forcings. The attribution of those spectral fingerprints to the change in essential climate variables involves the use of radiative kernels and a linear inversion scheme. The spectral fingerprinting scheme can be expressed as:

$$\Delta r = \mathbf{K}\Delta x + \varepsilon, \quad (1)$$

where  $\Delta r$  is the spectral fingerprints,  $\mathbf{K}$  is the radiative kernel,  $\Delta x$  is the change in the geophysical variables, and  $\varepsilon$  is the fingerprinting residual term.  $\Delta x$  can be derived from the spectral fingerprints as follows:

$$\Delta x = \left( \mathbf{K}^T \boldsymbol{\Sigma}^{-1} \mathbf{K} + \mathbf{S}_a^{-1} \right)^{-1} \mathbf{K}^T \boldsymbol{\Sigma}^{-1} \Delta r, \quad (2)$$

where  $\boldsymbol{\Sigma}$  is the covariance of the residual term  $\varepsilon$ , and  $\mathbf{S}_a$  is the covariance constraint of  $\Delta x$ . In the satellite-based remote sensing of trace gases, spectral fingerprints are established as variations in the observed spectral radiances  $r$  with respect to the reference spectrum  $r_0$ . These variations are exclusively attributed to the changes in the trace gas profile  $x$  from the reference profile  $x_0$ . The radiative kernel  $\mathbf{K}$  defines the linear relationship between  $\Delta r$  and  $\Delta x$ .

The spectral fingerprinting of CH<sub>4</sub> using the IR sounder measurements includes both classification and one-step linear inversion procedures. The classification analysis is carried out to build a predictive model based on a reference database that includes the representative radiance spectra, the collocated CH<sub>4</sub> profile data, and corresponding radiative kernels. The classification of IR spectral measurements is based on their spectral characteristics associated with the CH<sub>4</sub> absorption. A given IR sounder observation can therefore be automatically assigned to one of the predefined classes. The CH<sub>4</sub> profile  $x_0$ , the corresponding spectrum  $r_0$ , the radiative kernel  $\mathbf{K}$ , and the fingerprinting residual covariance  $\boldsymbol{\Sigma}$  of the assigned class are then used in the subsequent inversion procedure. The final solution  $x$  is given by adjusting  $x_0$  to account for the small spectral difference  $\Delta r$  between the measurements  $r$  and  $r_0$ :

$$x = x_0 + \left( \mathbf{K}^T \boldsymbol{\Sigma}^{-1} \mathbf{K} + \mathbf{S}_a^{-1} \right)^{-1} \mathbf{K}^T \boldsymbol{\Sigma}^{-1} (r - r_0), \quad (3)$$

where the radiative kernel  $\mathbf{K}$  is derived from the radiative transfer calculation as the Jacobian, i.e., the derivative of the TOA radiance with respect to the changes in  $x$ , for the assigned class.

$$\mathbf{K} = \left. \frac{dr}{dx} \right|_{x=x_0}. \quad (4)$$

The formulas shown in Equations (2) and (3) follow the standard OEM scheme. Further details regarding the establishment of  $x_0$ ,  $r_0$ , and  $\mathbf{K}$ , along with the construction of covariance matrices  $\boldsymbol{\Sigma}$  and  $\mathbf{S}_a$ , will be elaborated upon in Section 3.2.2.

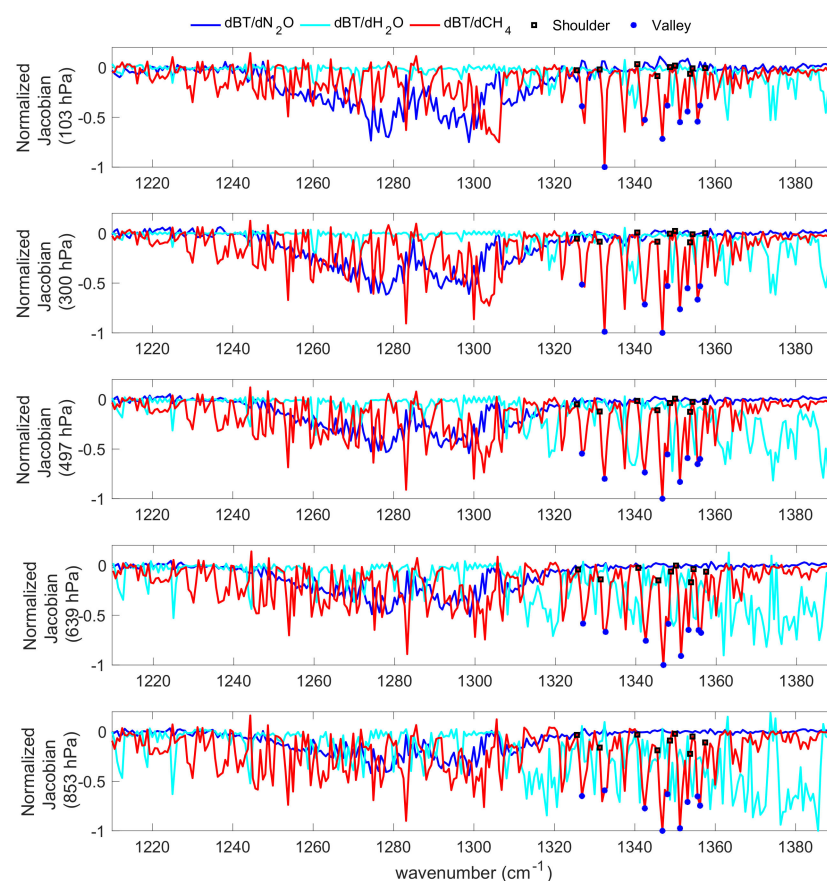
### 3. Implementation of the Spectral Fingerprinting Scheme

#### 3.1. Optimization of Spectral Information

Strong absorption lines of H<sub>2</sub>O and N<sub>2</sub>O exist in the spectral region where CH<sub>4</sub> absorption lines are located. In order to reduce the interference from those two species to the CH<sub>4</sub> retrieval, we select IR sounder spectral channels that are relatively less sensitive to H<sub>2</sub>O and N<sub>2</sub>O and more sensitive to CH<sub>4</sub>. Figure 1 shows sample Jacobians of N<sub>2</sub>O, H<sub>2</sub>O, and CH<sub>4</sub> in the 1210–1390 cm<sup>-1</sup> spectral region which illustrate the brightness temperature (BT) responses of unapodized CrIS spectra introduced by the changes in three trace gas species at different pressure levels, respectively. Nine CrIS channels (marked by blue dots

in Figure 1) that are sensitive to CH<sub>4</sub> but not to N<sub>2</sub>O are chosen. Those channels are selected to avoid the interference in the 1240–1320 cm<sup>-1</sup> spectral region, where N<sub>2</sub>O interference is strong. In order to mitigate the impact of broad spectral features such as the water continuum absorption, cloud emission and scattering, and surface emissivity variation in the selected channels, we construct differential spectral signals by pairing those nine ‘valley’ channels that show strong responses to CH<sub>4</sub> absorption with adjacent ‘shoulder’ channels (marked by ‘□’ in Figure 1). Table 1 lists the ‘valley’ and the ‘shoulder’ channels selected for CH<sub>4</sub> retrieval. Using the difference between the two sets of channels enhances the spectral signal contrast between CH<sub>4</sub> and H<sub>2</sub>O and therefore facilitates the identification of CH<sub>4</sub> fingerprints. To further enhance the information due to trace gas absorption, we normalize the CH<sub>4</sub> channel radiances by using the measurement from the atmospheric ‘window’ region. Using a radiance ratio from between the CH<sub>4</sub> absorption channels and a ‘window’ channel (e.g., the CrIS channel at 900.625 cm<sup>-1</sup>) can reduce the contributions from surface properties (emissivity and skin temperature) and clouds, which helps to better catch the spatial–temporal difference among sounder measurements mainly due to the change in CH<sub>4</sub> concentration. The spectral radiance  $r$  (in Equation (3)) to be used for the fingerprinting can be formulated as

$$r = \frac{r_{\text{valley}} - r_{\text{shoulder}}}{r_{\text{window}}} \quad (5)$$



**Figure 1.** CrIS (unapodized) Jacobians of CH<sub>4</sub>, N<sub>2</sub>O, and H<sub>2</sub>O at different pressure levels simulated by PCRTM. The demonstrated Jacobians are normalized by dividing the values by the maximum value for each dataset to better highlight the spectral response difference. The red squares and blue dots mark the ‘shoulder’ and ‘valley’ channels selected for the spectral fingerprinting of CH<sub>4</sub>, respectively.



**Table 1.** List of ‘valley’ and ‘shoulder’ channels selected for the spectral fingerprinting of CH<sub>4</sub> (highlighted in Figure 1).

|   | ‘Valley’ Channel $\nu$<br>(cm <sup>-1</sup> ) | ‘Shoulder’ Channel $\nu$<br>(cm <sup>-1</sup> ) |
|---|---|---|
| 1 | 1326.875                                      | 1325.625  |
| 2 | 1332.500                                      | 1331.250  |
| 3 | 1342.500                                      | 1340.625  |
| 4 | 1346.875                                      | 1345.625  |
| 5 | 1348.125                                      | 1348.750  |
| 6 | 1351.250                                      | 1350.000  |
| 7 | 1353.125                                      | 1353.750  |
| 8 | 1355.625                                      | 1354.375  |
| 9 | 1356.250                                      | 1357.500  |

### 3.2. Classification Using the Reference Database

#### 3.2.1. Data Sources of the Reference Database

Spectral data are obtained from the CrIS Level 1B Full Spectral Resolution radiance data product [33]. CrIS is a Fourier transform infrared spectrometer with three measurement bands that cover long-wave IR (650–1095 cm<sup>-1</sup>), mid-wave IR (1210–1750 cm<sup>-1</sup>), and short-wave IR (2155–2550 cm<sup>-1</sup>) spectral regions, respectively. Its spectral measurements from 1210 to 1400 cm<sup>-1</sup> provide the information that can be used for CH<sub>4</sub> retrieval. We used the results from the Carbon Tracker-CH<sub>4</sub> (CT-CH<sub>4</sub>) system as the CH<sub>4</sub> reference profiles [34]. CT-CH<sub>4</sub> provides global-scale 3 h interval CH<sub>4</sub> mixing ratio profile data for each 3.0° × 2.0° (longitude × latitude) grid cell. CT-CH<sub>4</sub> data are spatio-temporally interpolated to match the selected CrIS observations and saved in the reference database. The radiative kernels ( $\mathbf{K}$  in Equation (4)) are simulated using the Principal Component-based Radiative Transfer Model (PCRTM) [35]. Except for CH<sub>4</sub>, the other input parameters (e.g., temperature, water vapor, cloud, and surface properties) needed for the radiative kernel simulation are extracted from the SNPP CrIS SiFSAP [18].

The data in the reference database are matched to CrIS observations spanning six years (2016–2021). The objective is to establish a database with low redundancy that effectively captures the vertical and geographic distribution of CH<sub>4</sub> profiles from the original CT-CH<sub>4</sub> dataset, along with the corresponding spectral information from CrIS measurements. A stratified random sampling strategy is implemented for data selection. The CrIS observations are initially divided into 15 groups based on their relative across-track scan positions from nadir. Within each group, observations with a defined scan angle range are further organized by five-degree latitude zones. The data within each latitude zone are stratified based on surface pressure and CH<sub>4</sub> surface concentration values. Observations are then selected via random sampling, adhering to the joint probability distribution of the surface pressure and CH<sub>4</sub> surface concentration values for each latitude zone. To effectively capture the seasonal variations in CH<sub>4</sub> profiles, the selected samples are evenly distributed across each month. For each scan angle group, more than 400,000 observations are chosen and utilized to construct distinct reference databases. Each reference database includes spectral radiances (as defined in Equation (5)), CT-CH<sub>4</sub> profiles matched to the corresponding observations, radiative kernels, and associated surface pressure and latitude information.

#### 3.2.2. Classification Based on Optimized Spectral Radiance Information

Building the inversion scheme defined by Equation (3) requires the use of a classification procedure to find the reference state and therefore its corresponding components  $r_0$ ,  $x_0$ , and  $\mathbf{K}$  for a given observation with a normalized differential spectral radiance vector  $r$  (defined in Equation (5)). The procedure involves selecting a group of samples from the reference databases using nearest neighbor search. The Euclidean distance between  $r$  and the sample values from the database is used as the distance metric. The scope of the search is limited to reference samples within a 20-degree latitude zone of the observation and with close surface pressure values (within ±100 hPa). Considering the search is carried out in a low dimension (nine channel pairs), we employ a  $K$  Dimensional-tree (KD-tree)

structure to partition the spectral radiance dataset, enabling highly efficient searches. The nearest neighbor search selects a dataset of  $N$  ( $N < 30$ ) samples that includes  $\text{CH}_4$  profiles  $\{x_1 \cdots x_N\}$ , normalized differential spectra data  $\{r_1 \cdots r_N\}$ , and corresponding radiative kernels  $\{\mathbf{K}^1 \cdots \mathbf{K}^N\}$ .  $\{x_1 \cdots x_N\}$  are expected to be distributed around the true  $\text{CH}_4$  profile of the given observation. With  $r_0$ ,  $x_0$ , and  $\mathbf{K}$  being established using the mean values of the dataset, the residual dataset  $\{\varepsilon_1 \cdots \varepsilon_N\}$  can be calculated as follows:

$$\varepsilon_i = r_i - \mathbf{K}(x_i - x_0). \quad (6)$$

The covariance terms ( $\Sigma$  and  $S_a$  in Equation (3)) can therefore be derived as the covariance matrix of  $\varepsilon$  and  $x - x_0$ , respectively.

### 3.3. Constraints on $\text{CH}_4$ Profiles via Parameterization

Similar to other retrieval algorithms based on the OEM scheme, as depicted in Equations (2) and (3), the  $\text{CH}_4$  retrieval sensitivity can be quantified by the averaging kernel matrix, which is formularized as

$$\mathbf{A} = \left( \mathbf{K}^T \Sigma^{-1} \mathbf{K} + S_a^{-1} \right)^{-1} \mathbf{K}^T \Sigma^{-1} \mathbf{K}. \quad (7)$$

The dimension of  $\mathbf{A}$  is determined by the number of variables used to represent the  $\text{CH}_4$  profile. The number of independent observations provided by the OEM scheme can be estimated by the degrees of freedom (DOF), which is quantified as the trace of  $\mathbf{A}$ . In an ideal retrieval system, vertical profiles of  $\text{CH}_4$  can be perfectly represented as the combination of independent components targeted by corresponding observations. However, information content from CrIS, AIRS, and IASI measurements is limited. The DOF value for an IR sounder-based  $\text{CH}_4$  retrieval system is typically around 1.0 or even smaller [7,14]. Therefore, the number of independent components that can be obtained from  $\text{CH}_4$  retrieval is expected to be very low. A well-designed parameterization scheme should optimize the choice of parameters in order to better represent the vertical profiles using a minimal number of parameters.

We have selected three representative parameters to represent the  $\text{CH}_4$  profiles, taking into account the fundamental physical factors that influence the vertical distribution characteristics of  $\text{CH}_4$  in the atmosphere.  $\text{CH}_4$  undergoes oxidation by the hydroxyl radical (OH) in the upper troposphere. Above the tropopause, its mixing ratios decrease significantly due to additional chemical reactions with atomic oxygen and chlorine. Photolysis also contributes to  $\text{CH}_4$  decomposition in the higher-altitude stratosphere. The vertical profiles of  $\text{CH}_4$ , represented by symbol  $f$ , can be approximated as a sigmoid-like function:

$$f(h) = \frac{S}{1 + e^{-\left(\frac{h-P}{n}\right)}}, \quad (8)$$

where  $S$  defines the near-surface mixing ratio,  $P$  defines the pressure level of the turning point where the  $\text{CH}_4$  concentration goes from an accelerating decline to a decelerating decline, and  $n$  determines the rate of decrement in the stratosphere. The fingerprinting solution  $x$  defined in Equation (2) can therefore be represented as the state vector  $[S, P, n]$ . In this way, the change in  $\text{CH}_4$  profiles is constrained to a solution defined by three parameters:

$$\Delta f(h) = \frac{\partial f}{\partial S} \Delta S + \frac{\partial f}{\partial P} \Delta P + \frac{\partial f}{\partial n} \Delta n. \quad (9)$$

with

$$\frac{\partial f}{\partial S} = \frac{1}{1 + e^{-\left(\frac{h-P_0}{n_0}\right)}}, \quad (10)$$

$$\frac{\partial f}{\partial P} = \frac{S_0}{n_0} \frac{e^{-\left(\frac{h-P_0}{n_0}\right)}}{\left(1 + e^{-\left(\frac{h-P_0}{n_0}\right)}\right)^2}, \quad (11)$$

$$\frac{\partial f}{\partial n} = \frac{S_0 P_0}{n_0^2} \frac{e^{-\left(\frac{h-P_0}{n_0}\right)}}{\left(1 + e^{-\left(\frac{h-P_0}{n_0}\right)}\right)^2}, \quad (12)$$

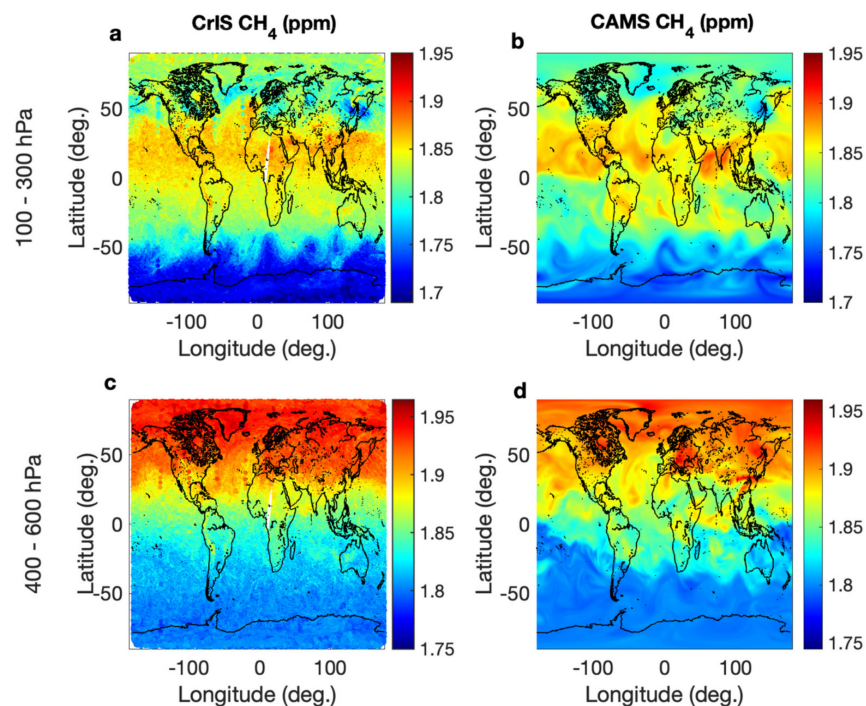
and  $S_0$ ,  $P_0$ , and  $n_0$  are the parameters associated with the reference state vector  $x_0$ . The Jacobian  $\mathbf{K}$  in Equation (2) can therefore be transformed from the profile domain to the parameter domain as  $\tilde{\mathbf{K}}(\mathbf{x}) = \tilde{\mathbf{K}}(\mathbf{f})\mathbf{T}$ . The dimension of the transformation matrix  $\mathbf{T}$  is  $m \times 3$ , where  $m$  denotes the number of level (layers) quantities used to represent a  $\text{CH}_4$  profile. The components of the  $i$ th row of the  $\mathbf{T}$  matrix are  $\left[\frac{\partial f(h_i)}{\partial S}, \frac{\partial f(h_i)}{\partial P}, \frac{\partial f(h_i)}{\partial n}\right]$ . The use of a sigmoid-like function (Equation (8)) to represent vertical profile serves as a smoothing constraint. Consequently,  $\Delta f$  must follow the vertical distribution patterns defined by Equations (10)–(12). This effectively prevents  $f$  from deviating from a sigmoid-like vertical profile to a solution with a drastically different and unrealistic vertical structure in the linear inversion step (Equation (2)).

#### 4. Results and Validation

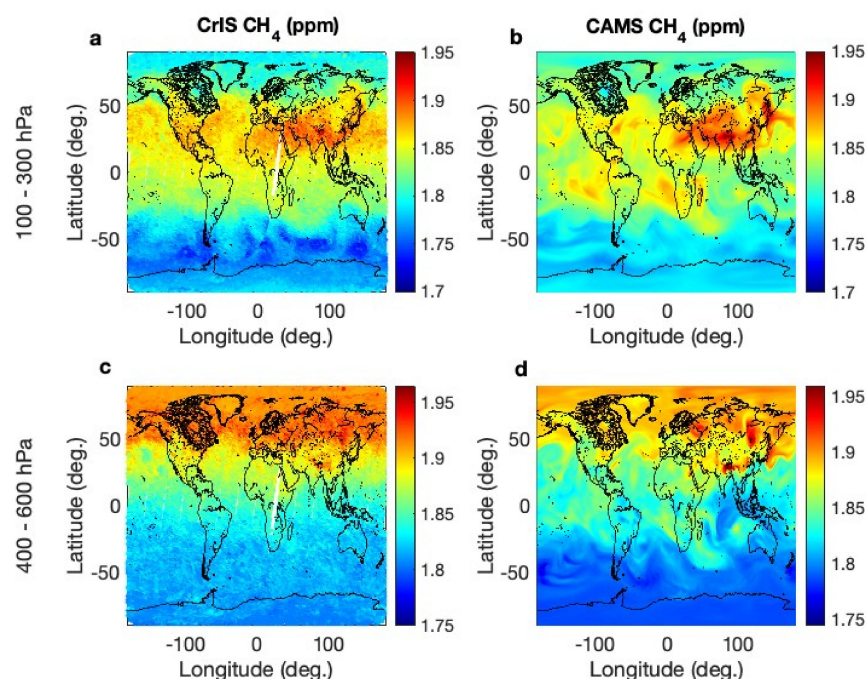
The left panels in Figures 2 and 3 show two days' worth of SNPP CrIS  $\text{CH}_4$  fingerprinting results (only descending orbital results are presented for simplicity). The fingerprinting results are compared with the global reanalysis dataset of atmospheric composition produced by CAMS. CAMS reanalysis provides sub-daily data interpolated to a regular  $0.75^\circ \times 0.75^\circ$  lat/lon grid. The CAMS reanalysis is produced using 4-DVar data assimilation in ECMWF's Integrated Forecasting System (IFS) using a comprehensive inventory, climatological, and chemical modeling dataset to initialize and constrain  $\text{CH}_4$  emissions, natural sources/sinks, and chemical sinks [36]. The accuracy of the CAMS  $\text{CH}_4$  results have been assessed in the CAMS greenhouse gas technical note [37]. When comparing the global mean values of CAMS  $\text{CH}_4$  with observations, the bias in the difference is generally small, with an uncertainty value of 0.01 ppm. Specifically, both the bias and the standard deviation values of the difference between the  $\text{CH}_4$  tropospheric profiles and the NOAA AirCore data are below 0.05 ppm. The largest difference in surface and tropospheric column  $\text{CH}_4$  between the CAMS results and the Total Carbon Column Observing Network (TCCON) data, with an averaged magnitude of up to 2.5%, is observed at mid- and high-latitude TCCON sites. The difference between CAMS results and data from the Network for the Detection of Atmospheric Composition Change (NDACC) is about 0.4% across all NDACC sites. CAMS reanalysis data are not recommended for investigating a local emission change or quantifying the changes in the atmospheric  $\text{CH}_4$  growth rate, but these data can be used to characterize synoptic spatial variability and the seasonal cycle of  $\text{CH}_4$  [37].

We compared the global distribution of the upper to middle tropospheric  $\text{CH}_4$  volume mixing ratio (VMR) characterized by the CrIS fingerprinting results with the corresponding daily mean values from the CAMS reanalysis. The side-by-side comparisons illustrated in Figures 2 and 3 show good correlations between two results concerning the latitudinal gradient and several large-scale thermodynamic characteristics. Satellite-based CrIS observations for  $\text{CH}_4$  are inevitably affected by cloud contamination. As a result, certain areas with high concentrations of  $\text{CH}_4$  in the tropical region, as indicated in the CAMS reanalysis, are not accurately reflected in the fingerprinting results. The systematic differences between the two sets of results, most notably in the southern polar region, fall within the range of 1–3%. Nevertheless, the changes in  $\text{CH}_4$  concentration from winter to summer in 2017, depicted through the difference between Figures 2 and 3, are consistent in both datasets. Figure 4 demonstrates the latitudinal change in  $\text{CH}_4$  concentration at different altitudes. In Figure 4, the latitudinal variations in  $\text{CH}_4$  concentration at different altitudes are illustrated. Consistent latitudinal patterns spanning from the Southern Hemisphere to the Arctic region can be observed in the daily results of the CAM reanalysis and CrIS fingerprinting. Both results also capture similar seasonal changes in  $\text{CH}_4$  profiles. Generally, there is a positive tropospheric  $\text{CH}_4$  increment in the Southern Hemisphere, contrasting with a negative increment in the Northern Hemisphere. The contrast between the  $\text{CH}_4$  profiles on January 11 and July 9 reveals a significant increase in  $\text{CH}_4$  concentration within the 100–200 hPa vertical region over Antarctica ( $90$ – $60^\circ\text{S}$ ) and the northern mid-latitude zone ( $30$ – $60^\circ\text{N}$ ).

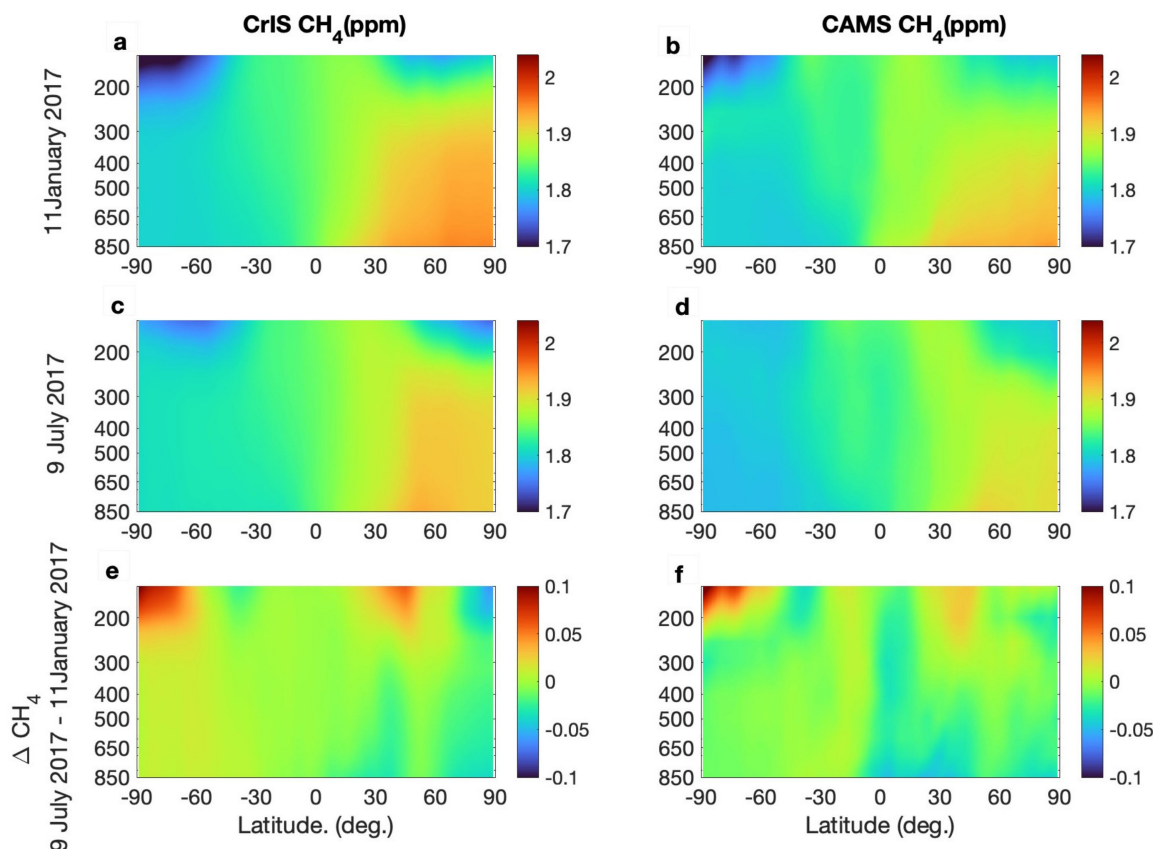




**Figure 2.** Comparison between daily CrIS CH<sub>4</sub> profile VMRs and those from CAMS reanalysis for 11 January 2017 in both the upper tropospheric region (100–300 hPa) and the middle tropospheric region (400–600 hPa). (a,b): averaged CH<sub>4</sub> volume mixing ratio in the 100–300 hPa latitudinal region derived from spectral fingerprinting results and from CAMS reanalysis). (c,d): averaged CH<sub>4</sub> volume mixing ratio in the 400–600 hPa latitudinal region derived from the spectral fingerprinting results and from the CAMS reanalysis.



**Figure 3.** Comparison between daily CrIS CH<sub>4</sub> profile VMRs and those from CAMS reanalysis for 9 July 2017 in both the upper tropospheric region (100–300 hPa) and the middle tropospheric region (400–600 hPa). (a,b): averaged CH<sub>4</sub> volume mixing ratio in the 100–300 hPa latitudinal region derived from spectral fingerprinting results and from CAMS reanalysis). (c,d): averaged CH<sub>4</sub> volume mixing ratio in the 400–600 hPa latitudinal region derived from the spectral fingerprinting results and from the CAMS reanalysis.

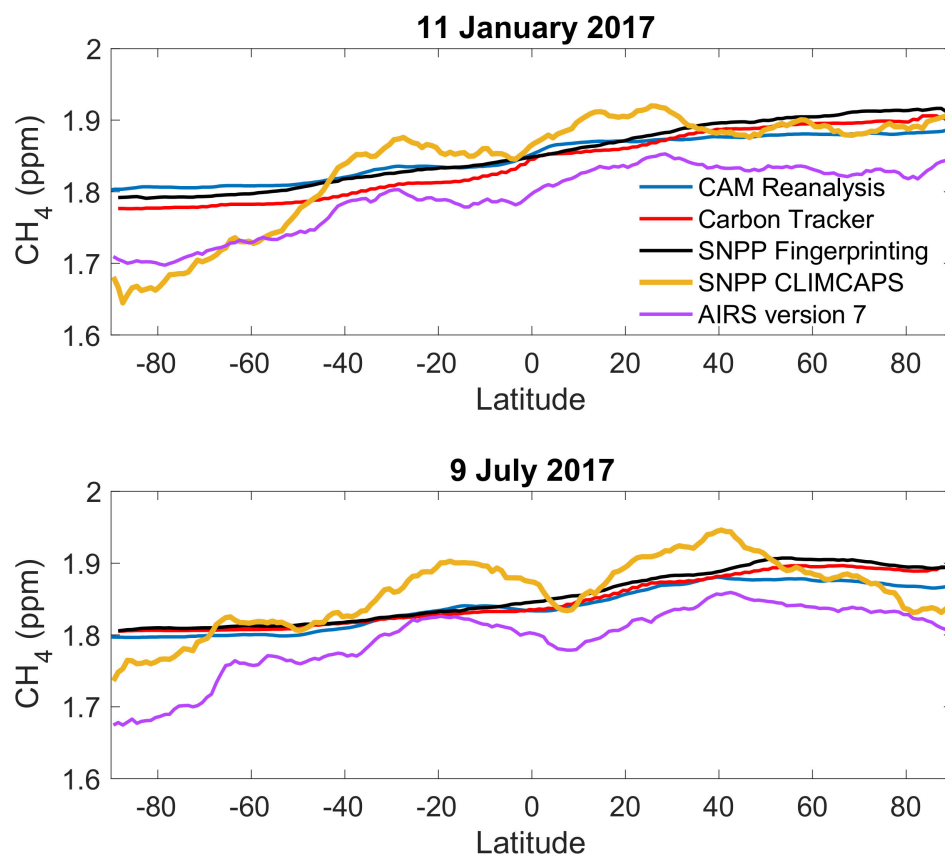


**Figure 4.** Latitudinal distribution of daily  $\text{CH}_4$  profile (CAMS reanalysis versus CrIS fingerprinting). (a,b):  $\text{CH}_4$  profiles of 11 January 2017 averaged to two-degree latitudinal zones; (c,d):  $\text{CH}_4$  profiles of 9 July 2017 averaged to two-degree latitudinal zones. (e,f): Difference between  $\text{CH}_4$  zonal mean profiles of 11 January 2017 and 9 July 2017. The daily zonal average has been computed for CrIS  $\text{CH}_4$  data spanning all longitudes.

Hyperspectral sensors like the AIRS, IASI, and CrIS only have measurement sensitivity to  $\text{CH}_4$  profiles within a limited vertical region. The AIRS is known to be sensitive to the upper troposphere (200–300 hPa) in the tropics and the middle troposphere (400–500 hPa) in the polar regions [7]. Similarly, studies on the IASI indicate its sensitivity to mid-upper tropospheric  $\text{CH}_4$  [13–15]. CLIMCAPS exclusively provides  $\text{CH}_4$  mass mixing ratios at 400 hPa, situated near the sensitivity peak defined by the algorithm and the spectral characteristics of the CrIS instrument. Despite the limited information from the measurements, the latitudinal distribution of the  $\text{CH}_4$  profiles derived by fingerprinting SNPP CrIS measurements, as illustrated in Figures 2–4, generally agrees with the CAM reanalysis. This underscores the benefit of a scene-dependent a priori scheme that can be precisely constructed using the spectral fingerprinting methodology.

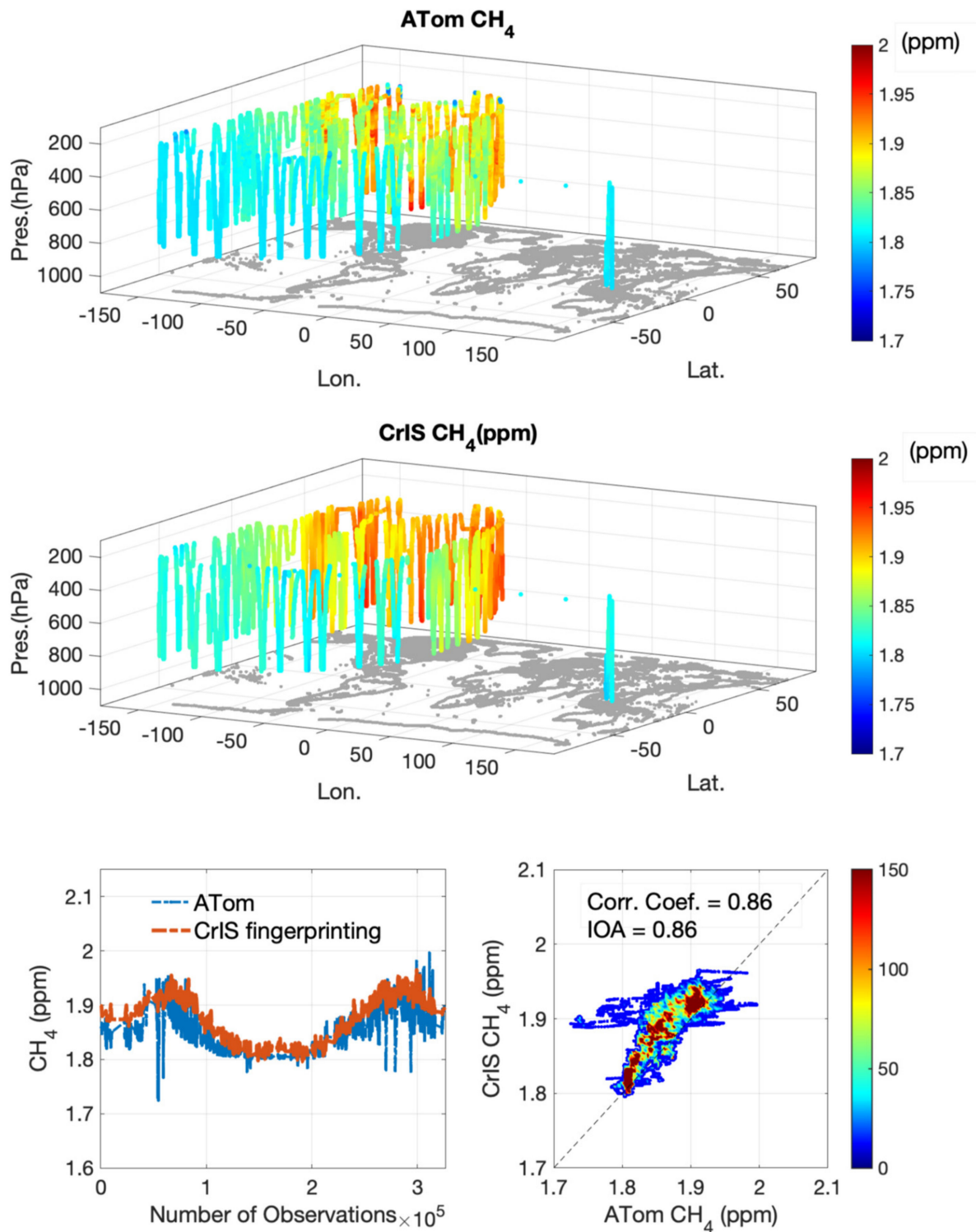
Under the fingerprinting scheme, the scene-dependent a priori information obtained through machine learning not only supplements the vertical information but also enhances the retrieval accuracy, particularly in areas where  $\text{CH}_4$  signals are relatively weak. Physical retrieval algorithms like CLIMCAPS and AIRS version 7 are susceptible to errors in both the forward model and information from measurements. Addressing those errors becomes more challenging in regions covered by cloud and lacking thermal contrast. The fingerprinting methodology uses an a priori scheme obtained via machine learning to effectively constrain the impact of measurement errors in those regions. Figure 5 illustrates the latitudinal distribution of mid-tropospheric  $\text{CH}_4$  (200–500 hPa) from the SNPP CrIS fingerprinting results of two days (11 January and 9 July 2017), along with the results from SNPP CLIMCAPS and AIRS version 7. Using CAM reanalysis and CT data as the reference, we can see that the daily latitudinal variations in  $\text{CH}_4$  from AIRS version 7 and CLIMCAPS are

unrealistically large. The global-scale  $\text{CH}_4$  concentration is underestimated in AIRS version 7. CLIMCAPS, on the other hand, overestimates  $\text{CH}_4$  concentration over the subtropical region and underestimates it over the Antarctic region during the winter. In comparison with the operational sounder retrieval results, the fingerprinting results provide more reasonable estimates of both the magnitude of the  $\text{CH}_4$  concentration and the latitudinal distribution on a global scale. These improvements in the accuracy of  $\text{CH}_4$  data, which could significantly benefit studies that emphasize daily-scale geographical distributions.



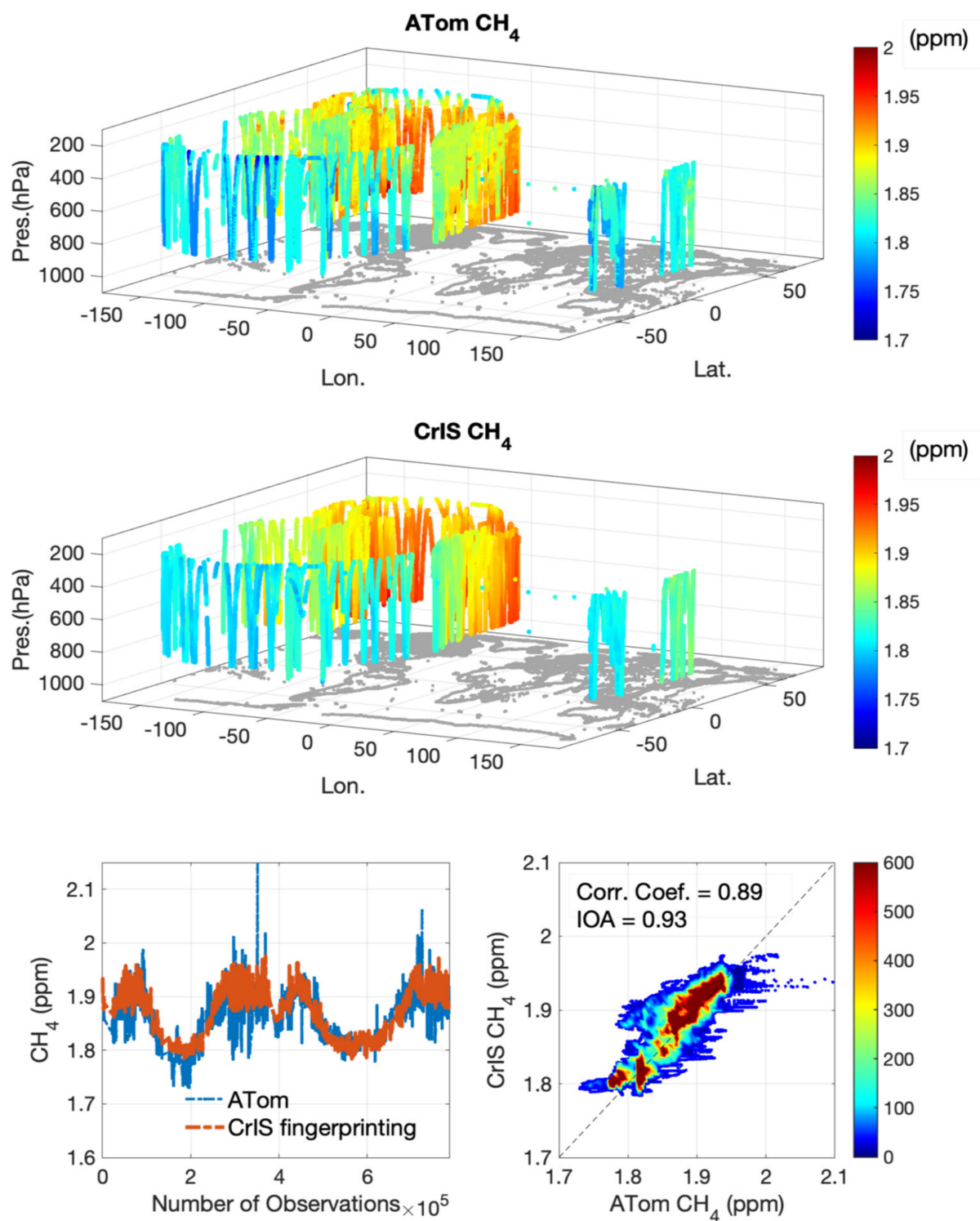
**Figure 5.** Inter-comparison between daily  $\text{CH}_4$  concentration (mean values of 200–500 hPa) averaged over a one-degree latitudinal region obtained via different methods.

The SNPP CrIS  $\text{CH}_4$  profiles were further validated using airborne measurement data. Three years' (2016–2018) worth of ATom data were used as the reference dataset. ATom observations provide continuous in situ measurements of  $\text{CH}_4$  at various altitudes ranging from 1.2 km to 12 km. We began by selecting SNPP CrIS measurements that fall within a 12 h window of individual ATom observations. Subsequently, we generated collocated CrIS results through a two-dimensional space interpolation process. We excluded samples where the horizontal distance between an ATom observation and the nearest CrIS footprint exceeds 1 degree ( $\sim 100$  km). Each collocated sample's fingerprinting result, representing an individual  $\text{CH}_4$  vertical profile, was then aligned with individual in situ results measured at different altitudes using vertical interpolation. ATom deployed flights over several months during a campaign year. The total number of collocated samples used for each year falls within the range of several hundred thousand. Figures 6–8 show the CrIS  $\text{CH}_4$  results along with the collocated ATom in situ observations. It is evident that the fingerprinting results effectively capture the fluctuations in  $\text{CH}_4$  concentration as the aircraft traversed different regions, collecting measurements at diverse altitudes.



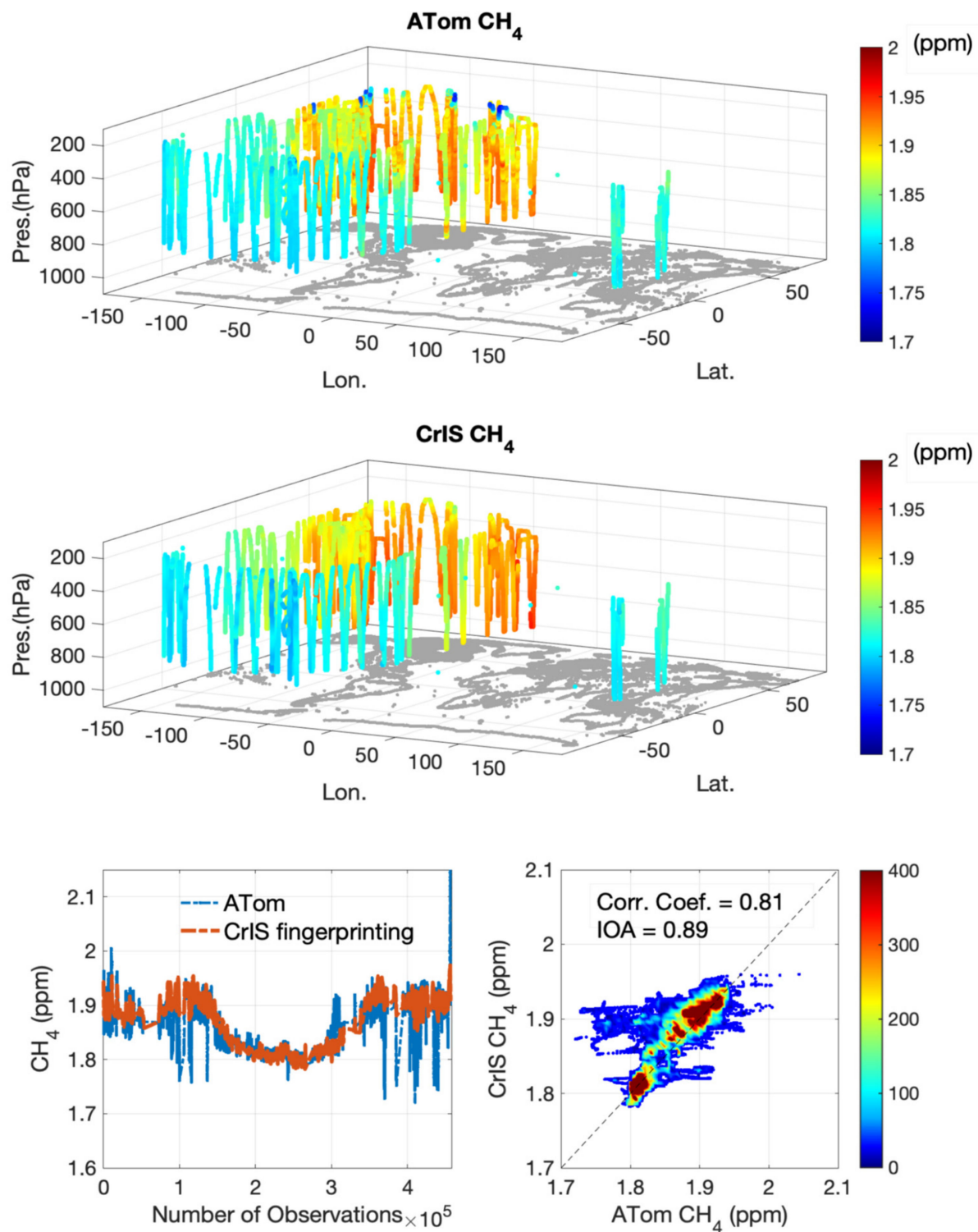
**Figure 6.** Upper panel: 3D illustration of CH<sub>4</sub> VMRs obtained via in situ ATom observations from 2016; middle panel: CH<sub>4</sub> VMRs retrieved from collocated SNPP CrIS observations; lower left panel: inter-comparison between two sets of results via an overlay plot; lower right panel: scatter plot of CrIS versus ATom CH<sub>4</sub> VMRs for all samples measured at a wide range of height and geolocations. The correlation coefficient between the fingerprinting results and ATom observations based on linear regression is displayed, along with the index of agreement (IOA) value. The color bar illustrates the number density of the collocated observations.





**Figure 7.** Upper panel: 3D illustration of  $\text{CH}_4$  VMRs obtained via in situ ATom observations from 2017; middle panel:  $\text{CH}_4$  VMRs retrieved from collocated SNPP CrIS observations; lower left panel: inter-comparison between two sets of results via an overlay plot; lower right panel: scatter plot of CrIS versus ATom  $\text{CH}_4$  VMRs for all samples, with the number density of the observations being illustrated by the color scale. Similar to Figure 5, the agreement between the fingerprinting results and observations is illustrated using the linear correlation coefficient and IOA. It should be noted here that the total number of collocated samples for 2017 is much larger than that for 2016, leading to the difference in color bar scales between Figures 5 and 6.





**Figure 8.** Upper panel: 3D illustration of CH<sub>4</sub> VMRs obtained via in situ ATom observations from 2018; middle panel: CH<sub>4</sub> VMRs retrieved from collocated SNPP CrIS observations; lower left panel: inter-comparison between two sets of results via an overlay plot; lower right panel: scatter plot of CrIS versus ATom CH<sub>4</sub> VMRs for all samples measured at a wide range of height and geolocations. The number density distribution of the observations, the correlation, and the index of agreement between CrIS and ATom are also presented.

Statistics regarding the difference between the CrIS fingerprinting and in situ measurement results are detailed in Table 2. Both the bias and the standard deviation values below

300 hPa are around 1% or smaller. The bias of the CrIS—ATom difference is comparable to what was reported in another CrIS—ATom inter-comparison study which used CH<sub>4</sub> results from the NOAA-Unique Combined Atmospheric Processing System (NUCAPS) [38]. However, it is noted that our temporal collocation criterion of  $\pm 12$  h is much more relaxed compared with the  $\pm 1.5$  h criterion used for the NUCAPS CH<sub>4</sub> study. Such a choice is based on the consideration that the change in atmospheric CH<sub>4</sub> concentration in the troposphere is predominantly less than 1% (root mean square difference) within a 12 h timeframe, as evidenced by statistics derived from CT-CH<sub>4</sub> and CAM reanalysis data. Importantly, our study benefits from a significantly larger sample size that is two orders of magnitude higher than that used in the NUCAPS CH<sub>4</sub> study.

**Table 2.** Statistical difference between the CrIS fingerprinting and the ATom CH<sub>4</sub>. The mean  $\pm$  standard deviation values of CrIS—Tom (in relative error %) have been computed for different layers (separated by upper and lower pressure levels). They are also split into regions: Arctic above 60°N, northern mid-latitude between 30°N and 60°N (N-Mid), tropics between 30°S and 30°N, southern mid-latitude between 60°S and 30°S (S-Mid), and Antarctic below 60°S.

| Pressure (hpa) | Arctic           | N-Mid           | Tropics         | S-Mid           | Antarctic       | Global          |
|----------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| above 250      | 5.34 $\pm$ 3.50  | 1.90 $\pm$ 2.83 | 0.21 $\pm$ 1.18 | 0.89 $\pm$ 2.16 | 0.58 $\pm$ 1.42 | 1.77 $\pm$ 3.02 |
| 250–350        | 2.39 $\pm$ 2.66  | 1.01 $\pm$ 1.50 | 0.32 $\pm$ 1.02 | 0.24 $\pm$ 0.92 | 0.49 $\pm$ 1.35 | 0.87 $\pm$ 1.71 |
| 350–450        | 1.13 $\pm$ 1.01  | 0.93 $\pm$ 0.95 | 0.46 $\pm$ 1.09 | 0.37 $\pm$ 0.81 | 0.07 $\pm$ 0.93 | 0.71 $\pm$ 1.04 |
| 450–550        | 0.69 $\pm$ 0.77  | 0.99 $\pm$ 1.00 | 0.58 $\pm$ 1.08 | 0.26 $\pm$ 0.77 | 0.22 $\pm$ 0.95 | 0.65 $\pm$ 0.98 |
| 550–650        | 0.40 $\pm$ 0.91  | 1.05 $\pm$ 1.07 | 0.68 $\pm$ 1.09 | 0.22 $\pm$ 0.71 | 0.41 $\pm$ 0.85 | 0.65 $\pm$ 1.03 |
| 650–750        | 0.45 $\pm$ 0.80  | 0.91 $\pm$ 1.27 | 0.67 $\pm$ 1.20 | 0.33 $\pm$ 0.73 | 0.42 $\pm$ 0.89 | 0.63 $\pm$ 1.10 |
| 750–850        | 0.51 $\pm$ 0.82  | 0.80 $\pm$ 1.19 | 0.57 $\pm$ 1.30 | 0.42 $\pm$ 0.78 | 0.36 $\pm$ 0.88 | 0.59 $\pm$ 1.10 |
| 850–950        | 0.29 $\pm$ 0.80  | 0.52 $\pm$ 1.52 | 0.40 $\pm$ 1.34 | 0.42 $\pm$ 0.90 | 0.52 $\pm$ 0.96 | 0.42 $\pm$ 1.22 |
| below 950      | −0.04 $\pm$ 1.40 | 1.33 $\pm$ 1.82 | 0.27 $\pm$ 1.42 | 0.40 $\pm$ 1.06 | 0.30 $\pm$ 0.85 | 0.61 $\pm$ 1.58 |

A similar satellite–airborne inter-comparison study validated CH<sub>4</sub> results retrieved from the Aura Tropospheric Emission Spectrometer (TES) and AIRS radiances using ATom data [39]. This study implemented a coincident criteria of 9 h and 50 km, closer to the criteria used in our study. Additionally, specific quality control was applied to screen out low-sensitivity and cloudy cases, leading to a low yield rate ( $\sim 1/4$ ) of samples [39]. The difference between AIRS and in situ results exhibited a bias of  $\sim 3\%$  over the ocean and  $\sim 4\%$  over land before bias correction, with a standard deviation less than 2% [39]. Both the NUCAPS and the TES-AIRS CH<sub>4</sub> validation studies involved the use of averaging kernel correction to take into account the relative difference introduced by averaging kernel ‘smoothing’ [40]. However, applying averaging kernel correction on ATom data necessitates extending ATom CH<sub>4</sub> measurements within the limited vertical region to match the complete profile measured by the sounders. It can be difficult to assess errors introduced by extending aircraft measurements using the assumed ‘true’ profiles typically obtained from global atmospheric chemistry models [7,10,40]. Nalli et al. [38] cautioned that applying averaging kernel correction for an inter-comparison study between satellite-based and airborne measurements can be misleading under certain conditions when the retrieval system has little to no measurement sensitivity to CH<sub>4</sub>. Based on these considerations, we did not implement averaging kernel correction in our study. Therefore, the difference between the spectral fingerprinting results and ATom results include the null-space errors.

The error statistics listed in Table 2 reflect the accuracy of the CH<sub>4</sub> derived under all-sky conditions without the imposition of a carefully designed quality control scheme to filter out low-sensitivity samples with potentially large errors. Despite the absence of a quality control scheme and the retention of null-space errors, the precision and accuracy of our results are still comparable to, or even better than, the NUCAPS and TES-AIRS results.

This suggests the potential of using spectral fingerprinting methodology to enhance the CH<sub>4</sub> data products from existing hyper-spectral sounder missions.

Table 2 shows that the errors in the lower troposphere are most pronounced in the northern mid-latitude region as opposed to the other regions. Compared with the ATom observations, the fingerprinting results for individual CrIS footprints demonstrate a consistently positive bias globally. The most substantial errors, both in terms of bias and standard deviation, can be observed in the upper troposphere to lower stratosphere region (above 250 hPa) over the Canadian wetlands and Greenland regions. These discrepancies can be attributed to the limitations in the vertical resolution of CrIS spectral measurements, preventing the precise sensing of the rapid decrease in methane concentrations in the upper-troposphere to low-stratosphere altitudes in the Arctic region. This challenge is particularly notable in regions where the tropopause height can be as low as 300 hPa.

## 5. Conclusions and Future Work

Compared to ground-based measurements, which serve as anchor observations for long-term CH<sub>4</sub> trends and inter-annual variations in the atmosphere, satellite-based measurements play a crucial role in assessing the geographical distribution of CH<sub>4</sub> concentrations globally. Accurate information about the sources and sinks of CH<sub>4</sub> is essential for climate models to predict atmospheric CH<sub>4</sub> concentrations. However, existing CH<sub>4</sub> data products based on infrared (IR) sounder observations face limitations in accurately capturing the global geographical distribution characteristics of CH<sub>4</sub>. A major challenge arises in areas where satellite-based measurements lack information content due to factors like insufficient thermal contrast or cloud blockage. To address this challenge, scene-dependent a priori information is required to enhance CH<sub>4</sub> retrieval in these areas. While high-quality CH<sub>4</sub> information can be obtained from CH<sub>4</sub> data assimilation systems like CT and CAM reanalysis, spectral radiance information from IR sounders has yet to be adequately assimilated in those systems. Additionally, considerations for data latency, especially in applications for environmental monitoring, hinder the direct use of CH<sub>4</sub> data from these systems to provide a priori constraints for physical retrievals.

We have developed a spectral fingerprinting scheme to tackle the challenges of retrieving CH<sub>4</sub> from satellite-based IR hyper-spectral sounder measurements. The scheme has a lot of similarities to a data assimilation system, but it differs from this type of system because it uses a machine learning-based model to initialize the a priori background and an optimized scheme to enhance the spectral fingerprints of CH<sub>4</sub>. The fingerprinting algorithm follows a 'lazy learning' methodology to efficiently identify a group of matched CH<sub>4</sub> profiles using the optimized CrIS spectral radiances as the predictors. The a priori information that largely retains the accuracy and characteristics of CT can be provided in a real-time (near-real-time) manner via a classification scheme based solely on the spectral radiance measured by sounders. A final solution for the CH<sub>4</sub> profile can be obtained by using a radiative kernel-based optimal inversion procedure to fit the optimized spectral radiance signals from CrIS measurements. This combination of machine learning and radiative kernel-based inversion has the potential to offer advantages in terms of accuracy and computational efficiency in the context of sounder-based CH<sub>4</sub> retrieval.

We have demonstrated that the CH<sub>4</sub> retrieved from SNPP CrIS observations via the fingerprinting method can generally catch the vertical and spatial distribution characteristics at a global scale and at different seasons, using the CAMS CH<sub>4</sub> reanalysis data as the reference. A validation study carried out using in situ ATom data demonstrated that both the systematic error and the uncertainty associated with the derived CH<sub>4</sub> profiles at various altitudes in the tropospheric region range from less than 1% to no more than 2%.

The fingerprinting scheme leverages SiFSAP to generate radiative kernels under all-sky conditions. The results and error statistics presented herein are associated with individual CrIS observations under both clear- and cloudy-sky conditions. Quality temperature, water vapor, surface, and cloud properties from the SiFSAP used in the offline training ensure the

accuracy of the radiative kernels at the individual footprint scale of the CrIS observations, thereby safeguarding the accuracy of the CH<sub>4</sub> fingerprinting at the fine spatial–temporal scale.

This paper has demonstrated the potential of using satellite-based spectral observations to facilitate the instantaneous monitoring of height-resolved methane distribution. This study showcases the effectiveness of employing machine learning to overcome the challenges posed by modeling and measurement errors in a standard retrieval scheme. Our future work will focus on enhancing the accuracy of the a priori background state by exploring more sophisticated machine learning models. Additionally, efforts will be made to integrate fingerprinting results as the a priori information for the iterative physical retrieval procedure used for the production of SiFSAP CH<sub>4</sub>.

**Author Contributions:** Conceptualization, W.W. and X.L.; methodology, W.W., X.L. and Q.Y.; software, W.W.; validation, W.W.; formal analysis, W.W.; investigation, W.W. and X.L.; resources, X.L. and L.Z.; writing—original draft preparation, W.W.; writing—review and editing, W.W., X.L., X.X., Q.Y., L.L., L.Z., D.K.Z. and A.M.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the NASA 2017 Research Opportunities in Space and Earth Sciences (ROSES) solicitation (grant number NNH17ZDA001N-TASNPP) and the NASA 2020 ROSES solicitation (grant number NNH20ZDA001N-SNPPSP).

**Data Availability Statement:** Data from the ATom campaign are located at <https://doi.org/10.3334/ORNLDAAAC/1925> (accessed on 2 September 2023) [41]. The Carbon Tracker CT-CH<sub>4</sub>-2023 results were provided by NOAA Global Monitory Laboratory (GML), Boulder, CO, USA, from the following website: <https://gml.noaa.gov/ccgg/carbontracker-ch4/> (accessed on 11 October 2023) [42]. The SiFSAP data used to generate radiative kernels can be obtained from the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) at [https://disc.gsfc.nasa.gov/datasets/SNDRSNIML2SFSPSUP\\_2/summary?keywords=SiFSAP](https://disc.gsfc.nasa.gov/datasets/SNDRSNIML2SFSPSUP_2/summary?keywords=SiFSAP) (accessed on 1 December 2023).

**Acknowledgments:** We really appreciate Youmi Oh of NOAA GML for providing the CT-CH<sub>4</sub>-2023 data.

**Conflicts of Interest:** Authors Qiguang Yang and Liqiao Lei are employed by the ADNET Systems Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Myhre, G.; Shindell, D.; Bréon, F.-M.; Collins, W.; Fuglestedt, J.; Huang, J.; Koch, D.; Lamarque, J.-F.; Lee, D.; Mendoza, B.; et al. Anthropogenic and natural radiative forcing. In *IPCC Climate Change 2014: Impacts, Adaptation, and Vulnerability*; Jacob, D., Ravishankara, A.R., Shine, K., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2014; pp. 661–684.
2. Frankenberg, C.; Meirink, J.F.; van Weele, M.; Platt, U.; Wagner, T. Assessing methane emissions from global space-borne observations. *Science* **2005**, *308*, 1010–1014. [[CrossRef](#)]
3. Kuze, A.; Suto, H.; Nakajima, M.; Hamazaki, T. Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Appl. Opt.* **2009**, *48*, 6716–6733. [[CrossRef](#)]
4. Butz, A.; Galli, A.; Hasekamp, O.; Landgraf, J.; Tol, P.; Aben, I. TROPOMI aboard Sentinel-5 Precursor: Prospective performance of CH<sub>4</sub> retrievals for aerosol and cirrus loaded atmospheres. *Remote Sens. Environ.* **2012**, *120*, 267–276. [[CrossRef](#)]
5. Saitoh, N.; Touno, M.; Hayashida, S.; Imasu, R.; Shiomi, K.; Yokota, T.; Yoshida, Y.; Machida, T.; Matsueda, H.; Sawa, Y. Comparisons between XCH<sub>4</sub> from GOSAT Shortwave and Thermal Infrared Spectra and Aircraft CH<sub>4</sub> Measurements over Guam. *SOLA* **2012**, *8*, 145–149. [[CrossRef](#)]
6. Holl, G.; Walker, K.A.; Conway, S.; Saitoh, N.; Boone, C.D.; Strong, K.; Drummond, J.R. Methane cross-validation between three Fourier transform spectrometers: SCISAT ACEFTS, GOSAT TANSO-FTS, and ground-based FTS measurements in the Canadian high Arctic. *Atmos. Meas. Tech.* **2016**, *9*, 1961–1980. [[CrossRef](#)]
7. Xiong, X.; Barnet, C.; Maddy, E.; Sweeney, C.; Liu, X.; Zhou, L.; Goldberg, M. Characterization and validation of methane products from the Atmospheric Infrared Sounder (AIRS). *J. Geophys. Res.* **2008**, *113*, 2005–2012. [[CrossRef](#)]
8. De Wachter, E.; Kumps, N.; Vandaele, A.C.; Langerock, B.; De Mazière, M. Retrieval and validation of MetOp/IASI methane. *Atmos. Meas. Tech.* **2017**, *10*, 4623–4638. [[CrossRef](#)]
9. Zou, M.; Xiong, X.; Saitoh, N.; Warner, J.; Zhang, Y.; Chen, L.; Weng, F.; Fan, M. Satellite observation of atmospheric methane: Intercomparison between AIRS and GOSAT TANSO-FTS retrievals. *Atmos. Meas. Tech.* **2016**, *9*, 3567–3576. [[CrossRef](#)]
10. Xiong, X.; Han, Y.; Liu, Q.; Weng, F. Comparison of Atmospheric Methane Retrievals from AIRS and IASI. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3297–3303. [[CrossRef](#)]



11. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding*; WorldSci: Hackensack, NJ, USA, 2000.
12. Garcia, O.E.; Sepúlveda, E.; Schneider, M.; Wiegeler, A.; Borger, C.; Hase, F.; Barthlott, S.; Blumenstock, T.; de Frutos, Á.M. Upper tropospheric CH<sub>4</sub> and N<sub>2</sub>O retrievals from MetOp/IASI within the project MUSICA. *Atmos. Meas. Tech.* **2018**, *11*, 4171–4215. [[CrossRef](#)]
13. Siddans, R.; Knappett, D.; Kerridge, B.; Waterfall, A.; Hurley, J.; Latter, B.; Boesch, H.; Parker, R. Global height-resolved methane retrievals from the Infrared Atmospheric Sounding Interferometer (IASI) on MetOp, *Atmos. Meas. Tech.* **2017**, *10*, 4135–4164. [[CrossRef](#)]
14. Razavi, A.; Clerbaux, C.; Wespes, C.; Clarisse, L.; Hurtmans, D.; Payan, S.; Camy-Peyret, C.; Coheur, P.F. Characterization of methane retrievals from the IASI space-borne sounder. *Atmos. Chem. Phys.* **2009**, *9*, 7889–7899. [[CrossRef](#)]
15. Crevoisier, C.; Nobileau, D.; Fiore, A.M.; Armante, R.; Chédin, A.; Scott, N.A. Tropospheric methane in the tropics—First year from IASI hyperspectral infrared observations. *Atmos. Chem. Phys.* **2009**, *9*, 6337–6350. [[CrossRef](#)]
16. Thompson, C.R.; Wofsy, S.C.; Prather, M.J.; Newman, P.A.; Hanisco, T.F.; Ryerson, T.B.; Fahey, D.W.; Apel, E.C.; Brock, C.A.; Brune, W.H.; et al. The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere. *B. Am. Meteorol. Soc.* **2022**, *103*, E761–E790. [[CrossRef](#)]
17. Wu, W.; Liu, X.; Lei, L.; Xiong, X.; Yang, Q.; Yue, Q.; Zhou, D.K.; Larar, A.M. Single field-of-view sounder atmospheric product retrieval algorithm: Establishing radiometric consistency for hyper-spectral sounder retrievals. *Atmos. Meas. Tech.* **2023**, *16*, 4807–4832. [[CrossRef](#)]
18. Smith, N.; Barnet, C.D. Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS). *Remote Sens.* **2019**, *11*, 1227. [[CrossRef](#)]
19. AIRS-Team Retrieval for Core Products and Geophysical Parameters: Versions 6 and 7 Level 2. Available online: [https://docsserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/L2\\_ATBD.pdf](https://docsserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/L2_ATBD.pdf) (accessed on 30 December 2017).
20. Xiong, X.; Liu, X.; Wu, W.; Knowland, K.E.; Yang, Q.; Welsh, J.; Zhou, D.K. Satellite observation of stratospheric intrusions and ozone transport using CrIS on SNPP. *Atmos. Environ.* **2022**, *273*, 118956. [[CrossRef](#)]
21. Mata-Miranda, M.M.; Guerrero-Ruiz, M.; Gonzalez-Fuentes, J.R.; Hernandez-Toscano, C.M.; Garcia-Andino, J.F.; Sanchez-Brito, M.; Vazquez-Zapien, G.J. Characterization of the Biological Fingerprint and Identification of Associated Parameters in Stress Fractures by FTIR Spectroscopy. *BioMed Res. Int.* **2019**, *2019*, 1241452. [[CrossRef](#)]
22. Mihrin, D.; Li, M.; Sánchez, M.A.; Hoeck, C.; Larsen, R.W.; Feilberg, K.L. Spectroscopic fingerprinting of organic material extracted from tight chalk core samples of the North Sea. *ACS Omega* **2020**, *5*, 31753–31764. [[CrossRef](#)]
23. Van Duijvenbode, J.R.; Buxton, M.W.N.; Shishvan, M.S. Performance Improvements during Mineral Processing Using Material Fingerprints Derived from Machine Learning—A Conceptual Framework. *Minerals* **2020**, *10*, 366. [[CrossRef](#)]
24. Brosinsky, A.; Foerster, S.; Segl, K.; Kaufmann, H. Spectral fingerprinting: Sediment source discrimination and contribution modelling of artificial mixtures based on VNIR-SWIR spectral properties. *J. Soils Sediments* **2014**, *14*, 1949–1964. [[CrossRef](#)]
25. Yadav, P.; Shetty, A.; Raghavendra, B.; Narasimhadhan, A. Gradient Based Spectral Similarity Measure for Hyperspectral Image Analysis. In Proceedings of the 2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS), Ahmedabad, India, 6–10 December 2021; pp. 242–245. [[CrossRef](#)]
26. Ren, Z.; Zhai, Q.; Sun, L. A Novel Method for Hyperspectral Mineral Mapping Based on Clustering-Matching and Nonnegative Matrix Factorization. *Remote Sens.* **2022**, *14*, 1042. [[CrossRef](#)]
27. Qu, S.; Li, X.; Gan, Z. A new hyperspectral image classification method based on spatial-spectral features. *Sci. Rep.* **2022**, *12*, 1541.
28. Leroy, S.; Anderson, J.; Dykema, J.; Goody, R. Testing climate models using thermal infrared spectra. *J. Clim.* **2008**, *21*, 1863–1875. [[CrossRef](#)]
29. Huang, Y.; Leroy, S.; Gero, P.J.; Dykema, J.; Anderson, J. Separation of longwave climate feedbacks from spectral observations. *J. Geophys. Res.* **2010**, *115*, D07104. [[CrossRef](#)]
30. Jin, Z.; Sun, M. An initial study on climate change fingerprinting using the reflected solar spectra. *J. Clim.* **2016**, *29*, 2781–2796. [[CrossRef](#)]
31. Pan, F.; Huang, X.; Leroy, S.S.; Lin, P.; Strow, L.L.; Ming, Y.; Ramaswamy, V. The Stratospheric Changes Inferred from 10 Years of AIRS and AMSU-A Radiances. *J. Clim.* **2017**, *30*, 6005–6016. [[CrossRef](#)]
32. Wu, W.; Liu, X.; Yang, Q.; Zhou, D.K.; Larar, A.M. Radiometrically consistent climate fingerprinting using CrIS and AIRS hyperspectral observations. *Remote Sens.* **2020**, *12*, 1291. [[CrossRef](#)]
33. Suomi NPP CrIS Level 1B Full Spectral Resolution V2. Available online: [https://disc.gsfc.nasa.gov/datasets/SNPPCrISL1B\\_2/summary](https://disc.gsfc.nasa.gov/datasets/SNPPCrISL1B_2/summary) (accessed on 13 October 2020).
34. Bruhwiler, L.; Dlugokencky, E.; Masarie, K.; Ishizawa, M.; Andrews, A.; Miller, J.; Sweeney, C.; Tans, P.; Worthy, D. CarbonTracker-CH<sub>4</sub>: An assimilation system for estimating emissions of atmospheric methane. *Atmos. Chem. Phys.* **2014**, *14*, 8269–8293. [[CrossRef](#)]
35. Liu, X.; Smith, W.L.; Zhou, D.K.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Appl. Opt.* **2006**, *45*, 201–209. [[CrossRef](#)]
36. Massart, S.; Agusti-Panareda, A.; Aben, I.; Butz, A.; Chevallier, F.; Crevoisier, C.; Engelen, R.; Frankenberg, C.; Hasekamp, O. Assimilation of atmospheric methane products into the MACC-II system: From SCIAMACHY to TANSO and IASI. *Atmos. Chem. Phys.* **2014**, *14*, 6139–6158. [[CrossRef](#)]



37. Agustí-Panareda, A.; Barré, J.; Massart, S.; Inness, A.; Aben, I.; Ades, M.; Baier, B.C.; Balsamo, G.; Borsdorff, T.; Bousserez, N.; et al. The CAMS greenhouse gas reanalysis from 2003 to 2020. *Atmos. Chem. Phys.* **2023**, *23*, 3829–3859. [[CrossRef](#)]
38. Nalli, N.R.; Tan, C.; Warner, J.; Divakarla, M.; Gambacorta, A.; Wilson, M.; Zhu, T.; Wang, T.; Wei, Z.; Pryor, K.; et al. Validation of Carbon Trace Gas Profile Retrievals from the NOAA-Unique Combined Atmospheric Processing System for the Cross-Track Infrared Sounder. *Remote Sens.* **2020**, *12*, 3245. [[CrossRef](#)]
39. Rodgers, C.D.; Connor, B.J. Intercomparison of remote sounding instruments. *J. Geophys. Res.* **2003**, *108*, 4116. [[CrossRef](#)]
40. Kulawik, S.S.; Worden, J.R.; Payne, V.H.; Fu, D.; Wofsy, S.C.; McKain, K.; Sweeney, C.; Daube, B.C., Jr.; Lipton, A.; Polonsky, I.; et al. Evaluation of single-footprint AIRS CH<sub>4</sub> profile retrieval uncertainties using aircraft profile measurements. *Atmos. Meas. Tech.* **2021**, *14*, 335–354. [[CrossRef](#)]
41. Wofsy, S.C.; Afshar, S.; Allen, H.M.; Apel, E.C.; Asher, E.C.; Barletta, B.; Bent, J.; Bian, H.; Biggs, B.C.; Blake, D.R.; et al. *ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2*; ORNL DAAC: Oak Ridge, TN, USA. Available online: [https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds\\_id=1925](https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1925) (accessed on 9 June 2023).
42. Oh, Y.; Bruhwiler, L.; Lan, X.; Basu, S.; Schuldt, K.; Thoning, K.; Michel, S.E.; Clark, R.; Miller, J.B.; Andrews, A.; et al. CarbonTracker CH<sub>4</sub> 2023. Available online: <https://gml.noaa.gov/ccgg/carbontracker-ch4/carbontracker-ch4-2023/> (accessed on 4 November 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.