remote sensing

MDPI

Article

Assimilation of GOSAT Methane in the Hemispheric CMAQ; Part I: Design of the Assimilation System

Sina Voshtani [1,2,*], Richard Ménard [2], Thomas W. Walker [1] and Amir Hakami [1]

[1] Department of Civil and Environmental Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada; ThomasWalker@cunet.carleton.ca (T.W.W.); AmirHakami@cunet.carleton.ca (A.H.)
[2] Air Quality Research Division, Environment and Climate Change Canada, Dorval, QC H9P 1J3, Canada; richard.menard@ec.gc.ca
* Correspondence: SinaVoshtani@cmail.carleton.ca or Seyyedsina.Voshtani@ec.gc.ca

Abstract: We present a parametric Kalman filter data assimilation system using GOSAT methane observations within the hemispheric CMAQ model. The assimilation system produces forecasts and analyses of concentrations and explicitly computes its evolving error variance while remaining computationally competitive with other data assimilation schemes such as 4-dimensional variational (4D-Var) and ensemble Kalman filter (EnKF). The error variance in this system is advected using the native advection scheme of the CMAQ model and updated at each analysis while the error correlations are kept fixed. We discuss extensions to the CMAQ model to include methane transport and emissions (both anthropogenic and natural) and perform a bias correction for the GOSAT observations. The results using synthetic observations show that the analysis error and analysis increments follow the advective flow while conserving the information content (i.e., total variance). We also demonstrate that the vertical error correlation contributes to the inference of variables down to the surface. In a companion paper, we use this assimilation system to obtain optimal assimilation of GOSAT observations.

Keywords: chemical data assimilation; methane; GOSAT; parametric Kalman filter

check for updates

Citation: Voshtani, S.; Ménard, R.; Walker, T.W.; Hakami, A. Assimilation of GOSAT Methane in the Hemispheric CMAQ; Part I: Design of the Assimilation System. *Remote Sens.* **2022**, *14*, 371. https://doi.org/10.3390/rs14020371

Academic Editor: Seon Ki Park

Received: 21 November 2021
Accepted: 8 January 2022
Published: 13 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

Methane is the second most important greenhouse gas (GHG) after CO_2, contributing to about 16% of the anthropogenic radiative forcing of all types of GHGs [1–3]. The globally averaged methane concentration has risen sharply since 2007, while the largest annual increase of methane (15.85 $\pm$ 0.47 ppb) was recorded in 2020 [4,5]. Owing to a stronger warming potential and a significantly shorter lifetime compared to CO_2, the reduction of methane has drawn much attention in GHG mitigation policy [6–9]. Furthermore, a higher methane concentration increases tropospheric ozone production, which is a critical problem, especially in populated areas in the Northern Hemisphere [10,11]. Methane is also identified as a major factor in the production of stratospheric water vapor, which indirectly affects the oxidation of other pollutants in the atmosphere [12].

Because of its impact on atmospheric chemistry and climate, better characterization of methane concentrations, and subsequently, inverse modelling of its emissions, has received significant attention over the past decade [1,13–18]. While most past studies have focused on the global inversion of methane sources at low spatial resolution, inversions over regional domains and at higher resolutions are important for constraining anthropogenic sources, such as fugitive and high-emitting sources from industry (e.g., oil and gas fracking, natural gas production) [19–22]. Inverse modelling of methane on a highly resolved regional domain requires an accurate estimate of both the initial concentrations within the domain and the inflows of methane at the lateral boundaries of the regional model and their uncertainties. In other words, the contribution from emissions on a short time scale (e.g., a week to a month) of a regional domain can be significantly smaller than the inflow

mass from the lateral boundary [23]. From an estimation point of view, the emissions signal (in inverse modelling) is masked by a much larger contribution from the state. Therefore, it is important to reduce the state uncertainty to a level comparable to the signal that we want to extract in inverse modelling. The scheme we have developed and verified by cross-validation in this two-part paper permits us to estimate the error variance of the state, which hopefully will be reduced enough by assimilation to extract the signal of the emissions.

Estimating the concentrations (i.e., in our case, the model state), whether for the purpose of constraining the initial conditions or boundary conditions or the whole domain in space and time, is a typical problem of data assimilation. We may contrast it with an inverse modelling problem where the variable to be estimated is not the model state, but rather a model parameter, such as the emissions. Furthermore, we should remark that the forward model in the data assimilation usually relates the estimated variable (i.e., a model concentration) with the observed variable (e.g., a column measurement). Hence, the forward model in the data assimilation context depends only on the measurement or instrument characteristics, such as the vertical averaging kernel or the radiative transfer problem when the radiance measurement is the observation (i.e., Level 1B data acquired by satellites [24]). In inverse modelling, the forward model relates, instead, the estimated parameter (i.e., emissions) with observations, and thus has to include the transport model (i.e., dynamic model of the atmosphere). See the tutorials in Kasibhatla et al. (2000) [25] for the commonalities and differences between inverse modelling and data assimilation.

Data assimilation has issues of its own, and several methods have been developed, such as the ensemble Kalman filter (EnKF), 4-dimensional variational (4D-Var), and hybrid ensemble-variational methods (see Asch et al. (2016) [26] for a detailed review). These methods require several model integrations (on the order of 30 to 100) for the state estimation. Ensemble methods are well-adapted for nonlinear atmospheric models, whereas variational methods are well-fitted for nonlinear observations operators (i.e., forward model in data assimilation context). However, regardless of nonlinearity, all data assimilation methods face the challenge of their applicability to large state-space models (e.g., atmospheric models state-space). For 3D- and 4D-Var, the applicability of the assimilation algorithm to large state-space is made possible with the introduction of the adjoint of operators combined with the use of an initial or background error correlation, assumed to be homogeneous and isotropic. In this case, the effect of multiplication of an (homogeneous and isotropic) error covariance with a state vector can be obtained as a sequence of operators without involving the storage of extremely large covariance matrices [27–30] (see Massart et al. [31] for 4D-Var assimilation of methane). For EnKF methods, applicability to large state-space arises from the use of a limited number of model integrations (e.g., 30–100 ensemble members) in combination with a localization performed with a Schur product of a tapering correlation function. The localization not only eliminates the spurious correlations at large distances but also increases the rank of the sample covariance to a value larger than the dimension of the state-space (as required by the assimilation algorithm), thus providing a full rank forecast error covariances to assimilate observations [32]. A comparison between EnKF and 4D-Var chemical data assimilation methods (i.e., to estimate the state, not the emissions) was conducted using a stratospheric chemistry transport model [33,34].

Another approach designed for the assimilation of long-lived chemical species is based on simplified Kalman filtering. This method is well-adapted to large state-space, linear problems, and exploits the properties of the continuity equation. In this scheme, the error variance is computed explicitly using only a single model integration. This assimilation method, historically also known as the sequential filter, was first used in the assimilation of long-lived chemical species in the stratosphere, both in case studies and over long time periods. In particular, it was used for the assimilation of MOPITT CO to perform a reanalysis of stratospheric ozone and to diagnose model error [35–40]. This scheme is, in fact, part of a larger class of algorithms under development called parametric Kalman filters (PKF), where additional covariance parameters, such as correlation lengths, are evolved

in time [41,42]. In its simplest form, where only the error variance is evolved, we will call this algorithm the PvKF for parametric variance Kalman filter. The PvKF requires only two model integrations, one for the state estimate and the other for its error variance (or uncertainty) and seems well-adapted for the assimilation of methane as a long-lived species. Considering a lifetime of about 10 years [43,44] and limited atmospheric chemistry [13], together with a linear observation mapping with smooth averaging kernels (such as those from GOSAT), renders the whole assimilation problem quasi-linear. In terms of algorithm computational efficiency, the applicability of the parametric Kalman filter (or PvKF) to large state-space is made from the correspondence between the continuous and discrete representations of the different operators, more specifically between spatial correlation functions and the corresponding correlation matrices. This idea was originally developed for the Optimal Interpolation (OI) method. Although OI is often used in conjunction with a data selection procedure to process a large number of observations by batches, we may not need such a procedure (data selection) if the number of profiles per time step is relatively small so that matrices in observation space can be directly inverted. The main idea underlying the applicability of OI to large state-space comes from computing the error correlations on the fly, as needed, using a continuous correlation function between pair of points without the need for the storage of a matrix in state-space. It also has the advantage that spatial correlation functions can be represented on any grid [45]. Note that the comparison of the computational cost mentioned above is for different assimilation schemes (not emissions inversion). The purpose of this paper is to design assimilation of GOSAT observations using the PvKF approach with the hemispheric version of the Community Multiscale Air Quality (CMAQ) model. In a companion paper, we use this assimilation system to obtain optimal assimilation of GOSAT observations and realistic error covariance parameters, including observation and model error and correlation lengths. We envision that the development of the current hemispheric assimilation system will be used in the future on a limited domain for joint data assimilation and emission inversion using the regional CMAQ model.

The organization of the current paper is as follows. In Section 2, first, we present the modifications made to CMAQ to include methane as an evolving species and, in particular, how anthropogenic and natural emissions of methane can be incorporated into the SMOKE processing system. Then, we explain how to use the observation averaging kernels and a priori provided by the GOSAT data in a form useful for the hemispheric CMAQ grid and for assimilation. In Section 3, we describe the PvKF and how the scheme can be adapted for a large state-space, such as the CMAQ model. Section 4 details other aspects of the assimilation system, specifically, the initial conditions, the observation bias correction, and the formulation of homogeneous isotropic correlation models in the uniform hemispheric CMAQ grid. We close this section with a description of the parameters used in the error covariance matrices. Finally, in Section 5, we conduct simulated observation experiments, in particular the one-observation experiment that permits the verification of the algorithm and the assumptions used for propagation of error variance. We conclude Section 5 with an estimation of the computational cost of such an algorithm.

2. Model and Observation Operator

2.1. Modifications of the CMAQ Model to Handle Methane Transport and Emissions

CMAQ is a limited-area model developed by the U.S. Environmental Protection Agency (EPA) [46] that is used as a regional model driven at the lateral boundaries by the hemispheric version of this model (referred to as H-CMAQ) [47]. H-CMAQ, based on CMAQ v5.3, is used here to simulate and assimilate methane in the Northern Hemisphere. For all practical purposes, a hemispheric model does not need to account for the inter-hemispheric exchange of mass for time scales up to several months [12]. The domain covers the Northern Hemisphere on a polar stereographic projection (see Appendix A), which includes 187 $\times$ 187 grid cells horizontally with a 108 km grid spacing and 44 vertical layers extended from the surface to the model top at 50 hPa. The modified vertical structure of H-

CMAQ maintains finer resolution than regional CMAQ above the boundary layer, aiming at a better representation of transport processes, also suitable for long-lived species [47]. We modified H-CMAQ to account for methane concentrations and emissions. Methane can be configured either as an inert trace gas or a reactive gas-phase species oxidized by hydroxyl radical (OH). Figure S2 shows an example of concentration loss due to chemical reactions after two weeks. We considered reactive methane with the gas-phase model within CMAQ v5.3 and based on the CB06 chemical mechanism [48]. The error variance is treated as a chemically inert tracer in CMAQ, subject only to an advection-only transport scheme.

Our data assimilation system includes methane emissions of both anthropogenic (~60%) and natural (~40%) sources. For the anthropogenic emissions, we obtained all available source sectors from the Emission Database for Global Atmospheric Research (EDGAR v6) [49,50]. These sources come with $0.1^\circ \times 0.1^\circ$ horizontal resolution and monthly temporal resolution emission grid maps. We processed them using Sparse Matrix Operator Kernel Emissions (SMOKE v4.5) [51] to provide hourly gridded methane emissions into the model. Wetlands are the primary source of natural emissions, accounting for about 80% of the total [13,52]. We used monthly wetlands emissions from WetCHARTS v1.0 with the full ensemble mean [53] and mapped it into our domain using a flat hourly/weekly temporal profile.

2.2. GOSAT Observation Operator for Data Assimilation

Greenhouse Gas Observing Satellite (GOSAT) was launched in January 2009 by the Japanese Space Agency (JAXA) [54]. It is in a Sun-synchronous orbit at an altitude of 666 km with a 3-day revisit time and an equator overpass time at about 13:00 local time. One main goal of GOSAT is monitoring the abundance of methane in Earth's atmosphere. Due to the high sensitivity of the shortwave infrared (SWIR) GOSAT retrieval at the surface, coupled with a suitable spatiotemporal resolution, the assimilation of atmospheric methane and inverse modelling of its sources and sinks are suitable [55–59].

The instrument, TANSO-FTS, onboard GOSAT, has a field of view with a 10.5 km diameter footprint operating in a cross-track scanning mode. It measures the abundance of methane by analyzing the backscattered solar radiance spectrum in the SWIR near 1.6 µm. A column-average dry-mole fraction of methane (XCH_4) represents the instrument retrieved observation, which corresponds to the methane average volume mixing ratio ($VMR \equiv y^o$) of a partial column atmosphere with a given surface,p_S, and top pressure, p_T. Two approaches are used to derive retrieval algorithms: Full Physics (FP) and Proxy (PR). The FP method integrates a sophisticated radiative transfer model and solely relies on methane modelling and its corresponding errors, while the PR algorithm provides more data points but relies on an accurate CO_2 model simulation and its retrieval (XCO_2) [56]. Both algorithms were developed at the Netherlands Institute for Space Research (SRON) and Karlsruhe Institute for Technology (KIT) [55], and their products are available through the ESA GHG-CCI initiative, https://climate.esa.int/en/projects/ghgs/ (accessed on 20 November 2021) [57]. We use both products, but mainly focus on PR due to higher density and/or better coverage [55,56].

For assimilation purposes, in addition to the retrieval data, we use supplementary products [57,60]. Thus, each retrieval also includes vectors of the normalized column-average kernel, $\mathbf{A}$, pressure levels, p_l, at which the average kernels are derived, and the corresponding vector of a priori, $\mathbf{y}^p$. PR and FP retrieval algorithms offer 5 and 13 levels of data in the vertical, respectively. A layer-based approach described in Bergamaschi et al. (2007) [61] is applied to compute the model partial-column value, y^m, equivalent to the retrieval, y^o

$$y^m = [(1 - \mathbf{A})\mathbf{y}^p + \mathbf{A}\mathbf{y}^m]\boldsymbol{\omega}^T, \quad (1)$$

$\mathbf{y}^m$ represents the mapped concentration of H-CMAQ on the pressure layers of the observation, and $\boldsymbol{\omega}$ is the vector of pressure layers weights, whose elements are expressed as

$$\omega_l = \frac{p_l - p_{l+1}}{p_S - p_T}, \tag{2}$$

where p_S is the surface pressure and p_T is the model top pressure. Note that Equations (1) and (2) are in the retrieval space so that all the vectors represent the vertical grids within a single retrieval. Now, let us consider the 3D model estimate of methane concentrations at time, t, in a vectorized form, $\mathbf{X}$, with N_x dimension. The observation vector, $\mathbf{Y}^o$, consists of a set of retrievals at approximately the same time but different locations from model grid points. It also provides the same type of quantity but with a dimension quite smaller than the model ($N_y \ll N_x$). Assuming a linear relationship between $\mathbf{Y}^o$ and $\mathbf{X}$, we have a linear observation operator, $\mathbf{H}$.

$$\mathbf{Y}^o = \mathbf{H}(\mathbf{X}) + \boldsymbol{\varepsilon}^o. \tag{3}$$

In fact, $\mathbf{H}$ is a combination of two linear operators, comprising a horizontal, $\mathbf{H}_h$, and a vertical, $\mathbf{H}_v$, operator,

$$\mathbf{H}(\mathbf{X}) = \mathbf{H}_v(\mathbf{H}_h(\mathbf{X})), \tag{4}$$

where $\mathbf{H}_h$ interpolates the components of $\mathbf{X}$ from the horizontal model grid points to observation locations using a bi-linear interpolation function, and $\mathbf{H}_v$ transforms $\mathbf{X}$ at the geographical location of observations to a vector equivalent to $\mathbf{Y}^o$ (using avergaing kernels). Thus, the assimilation problem consists of finding the best estimate of $\mathbf{X}$ with its error statistics, which together are called the analysis.

3. Data Assimilation System

3.1. Background of the Assimilation Scheme

First, let us define a correlation function on an arbitrary grid following Gaspari and Cohn 1999 [45]. It is sufficient to define a correlation function in $\mathbb{R}^3 \times \mathbb{R}^3$, so that any subspace (or manifold) of $\mathbb{R}^3$ (e.g., the surface of a sphere) also define a correlation on that subspace. This property was used in this study to define underlying homogeneous isotropic correlation functions (with periodicity on a sphere), which are then mapped onto the polar stereographic grid of H-CMAQ, which has a uniform grid spacing on the projected plane. PvKF uses these concepts, together with the property that for the advection equation, the error variance can be forecast without knowing the error correlation [62]. Thus, in a PvKF framework, the error variances are dynamically evolving according to the model's advection scheme, but the spatial error correlations are kept fixed and are computed using the same approach as taken by OI.

The algorithm of PvKF is decomposed into two steps; a forecast step and an analysis step, which accounts for the effect of observations. A covariance function, $P(\mathbf{x}, \mathbf{x}', t)$, is a function of a pair of points, $\mathbf{x} = (x, y, z)$ and $\mathbf{x}' = (x', y', z')$, at time t. It is related to the correlation function,$C(\mathbf{x}, \mathbf{x}', t)$, through the standard expression

$$P(\mathbf{x}, \mathbf{x}', t) = \sigma(\mathbf{x}, t)\, C(\mathbf{x}, \mathbf{x}', t)\, \sigma(\mathbf{x}', t), \tag{5}$$

where σ is the error standard deviation (i.e., $\sigma^2(\mathbf{x}, t)$ is the error variance at the point $\mathbf{x}$ and time t). In PvKF, the error correlation is generally time-invariant, homogeneous and isotropic in the horizontal (i.e., depends only on the horizontal distance). 3D spatial correlations are constructed using a horizontal/vertical separability assumption,

$$C(x, y, z, x', y', z') = C_h(x, y, x', y')\, C_v(z, z'), \tag{6}$$

and for the horizontal correlation

$$C_h(x, y, x', y') = C_h(\|(x, y) - (x', y')\|), \tag{7}$$

is assumed to be homogeneous and isotropic. Note that the separability assumption has been verified for long-lived species in the stratosphere by Menard et al. (2019) [63] (see Figure S3 therein).

3.2. Forecast Step

The forecast step consists of two model integrations, one for the state and the other for the error variance,

$$\begin{aligned} X(\mathbf{x},t) &= \mathbf{M}(X(\mathbf{x},t-\delta t)) \\ \sigma^2(\mathbf{x},t) &= \mathbf{M}^*\left(\sigma^2(\mathbf{x},t-\delta t)\right) \end{aligned} \tag{8}$$

where $\mathbf{M}$ represents the chemical transport model based on the atmospheric diffusion equation [64,65] of CMAQ, and $\mathbf{M}^*$ denotes the advection model of error variances based on the CMAQ native advection scheme. $X(\mathbf{x},t)$ represents the chemical concentration as a function of 3-dimensional model space $\mathbf{x} = (x,y,z)$ and time t. Note that the error covariance forecast step of the PvKF does not follow the standard Kalman filter equation, but uses the continuous formulation of propagation of error covariances, which can be solved by using the method of characteristic as discussed in Cohn (1993) [62]. This approach was applied in several studies of assimilation of long-lived species [35–37,40]. Using the advection of error variance also has two advantages. First, it avoids the loss of error variance in standard Kalman filter formulation, such as in EnKF (see Section 2 in Menard et al. (2021) [66]). Secondly, it significantly reduces the computational cost compared to other methods (e.g., EnKF).

The $\mathbf{M}^*$ operator is, in fact, the integration from $t-\delta t$ to t of

$$\frac{\partial \sigma^2}{\partial t} + \mathbf{V}\cdot\nabla\sigma^2 \;=\; q, \tag{9}$$

where $\mathbf{V}$ is the 3-dimensional advection wind, and $q(\mathbf{x},t)$ accounts for the model error variance growth (e.g., any other processes that are not advection; we discuss further in Section 4.4 and in Part II). It is important to note that in the context of data assimilation, the effect of the observations usually has the largest impact on error covariances, so that it is common to use a simpler model to forecast the error covariances or to run the adjoint model (e.g., incremental 4D-Var [67] used operationally for weather forecasting using a lower resolution model with simplified physics). Here, as in the incremental approach, we use only advection in the forecast of error variances.

After the integration time, δt, we obtain a forecast concentration, $X^f(\mathbf{x},t)$, and a forecast error variance, $(\sigma^f)^2$, from which the covariance function can be reconstructed by applying the correlation function from Equation (5). Therefore, a forecast error covariance function, $P^f(\mathbf{x},\mathbf{x}',t)$, is obtained, considering that the correlation function is time-invariant (i.e., stationary). Thus,

$$\begin{aligned} P^f(\mathbf{x},\mathbf{x}',t) &= \sigma^f(\mathbf{x},t)\; C(\mathbf{x},\mathbf{x}')\, \sigma^f(\mathbf{x}',t) \\ C(\mathbf{x},\mathbf{x}') &= C_h(\|(x,y)-(x',y')\|)\; C_v(z,z') \end{aligned} \tag{10}$$

and equivalently in matrix notation

$$\begin{aligned} \mathbf{P}^f(t) &= \Sigma^f(t)\; \mathbf{C}\; \Sigma^f(t) \\ \mathbf{C} &= \mathbf{C}_h \otimes \mathbf{C}_v \end{aligned} \tag{11}$$

where Σ^f is a diagonal matrix of forecast error standard deviations and $\otimes$ is the Kronecker product of matrices. The matrix form is useful to write out the analysis step.

3.3. Analysis Step

Let us start as if $\mathbf{H}$ is only a horizontal operator (we will discuss the vertical aspects in the following Section 3.4). A computational simplification in the analysis step arises when

observations are considered horizontally as point measurements. This allows simplification of the Kalman gain matrix and the analysis error covariance and variance. Horizontally, a point measurement observation operator can be modelled as a delta function at the observation location,$\mathbf{x}_o$. Specifically, if $f(\mathbf{x})$ is a continuous function of space, $\mathbf{x}$, and the horizontal observation operator, H_h (as previously defined in Section 2.2), applied on f ($H_h[f]$), gives the value of the function at the observation location, $f(\mathbf{x}_o)$, we can write

$$H_h[f] \;= f(\mathbf{x}_o) = \int f(\mathbf{x})\;\delta(\mathbf{x}-\mathbf{x}_o)\,d\mathbf{x}\,. \tag{12}$$

The $H_h[\cdot]$ operator is thus associated with a delta function, $\delta(\mathbf{x}-\mathbf{x}_o)$. Although we never use this representation of the observation operator on the state fields themselves since they are represented discretely on a grid, we do apply this definition when we consider the application of the observation operator on error covariance functions. Thus, we have the following,

$$[\mathbf{HP}^f]_{o,j} \Leftrightarrow H_h[P^f]_{\,o,j} \;=\; \int P^f(\mathbf{x},\,\mathbf{x}'_j)\;\delta(\mathbf{x}-\mathbf{x}_o)\,d\mathbf{x} \;=\; P^f(\mathbf{x}_o,\mathbf{x}'_j)\,, \tag{13}$$

$$[(\mathbf{HP}^f)^T]_{i,o} = [\mathbf{P}^f\mathbf{H}^T]_{i,o} \Leftrightarrow H_h^T[P^f]_{i,o} \;=\; \int P^f(\mathbf{x}_i\,,\,\mathbf{x}')\;\delta(\mathbf{x}'-\mathbf{x}'_o)\,d\mathbf{x}' \;=\; P^f(\mathbf{x}_i,\mathbf{x}'_o), \tag{14}$$

where subscript i and j refer to model grid points, whereas subscript o refers to the position of the observations.

Considering the covariance function between any pair of observation locations, $\mathbf{x}_{o(1)}$, $\mathbf{x}_{o(2)}$, we get an element of the matrix $\mathbf{HP}^f\mathbf{H}^T$. That is

$$\begin{aligned}[\mathbf{HP}^f\mathbf{H}^T]_{o(1),o(2)} \;&\Leftrightarrow H_{h,o(2)}[H_{h,o(1)}[P^f]]_{o(1)\,,o(2)}\\ &= \textstyle\int P^f(\mathbf{x},\,\mathbf{x}')\;\delta(\mathbf{x}-\mathbf{x}_{o(1)})\,\delta(\mathbf{x}'-\mathbf{x}'_{o(2)})\,d\mathbf{x}\,d\mathbf{x}' \;=\; P^f(\mathbf{x}_{o(1)},\mathbf{x}_{o(2)})\,.\end{aligned} \tag{15}$$

Suppose now we have K observations (with different locations) that are used in an analysis. The error covariance matrix in observation space (i.e., between each pair of observations) is $\mathbf{HP}^f\mathbf{H}^T$ (a $K\times K$ matrix), where each element of this matrix is given by Equation (15). With this preamble, we are now able to formulate the analysis step for both the state estimate and the error variance. The analysis state is written as usual and requires the computation of the Kalman gain matrix $\mathbf{K}$,

$$\mathbf{X}^a = \;\mathbf{X}^f + \;\mathbf{K}(\mathbf{Y}^o - \mathbf{H}\mathbf{X}^f)\,, \tag{16}$$

where

$$\mathbf{K} \;=\; (\mathbf{HP}^f)^T(\mathbf{HP}^f\mathbf{H}^T+\mathbf{R})^{-1}. \tag{17}$$

$\mathbf{R}$ represents the observation error covariance matrix. We should remark that in data assimilation, the observation error is not simply the instrument error, but should also contain an error of representativeness since, ultimately, observations are used to update the model state (see Section 4.4).

All state quantities (i.e., the forecast and the analysis) are written as vectors of dimension N_x, where N_x is the number of model grid points. We recall that $\mathbf{Y}^o$ is a vector of dimension K that contains all observations processed in data assimilation. The matrix $\mathbf{HP}^f$ is sometimes called the matrix of representors, where each observation location comes with all the model grid points. We can think of it as a vector of *impact covariance functions*. $\mathbf{HP}^f$ can be denoted in column form as

$$\mathbf{HP}^f = \;\begin{pmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_N\end{pmatrix}, \tag{18}$$

$$\mathbf{p}_i = \begin{pmatrix} P^f(\mathbf{x}_{o(1)}, \mathbf{x}_i) \\ P^f(\mathbf{x}_{o(2)}, \mathbf{x}_i) \\ \cdots \\ P^f(\mathbf{x}_{o(K)}, \mathbf{x}_i) \end{pmatrix}, \tag{19}$$

where each $\mathbf{p}_i$ is a column vector ($K \times 1$) that represents the forecast error between all observations and a single model grid point that we call the *sensitivity error covariance function*— in matrix form, $(\mathbf{HP}^f)^T$. The analysis update of the error covariance, called the analysis error covariance in data assimilation, takes the form,

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f = \mathbf{P}^f - (\mathbf{HP}^f)^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{HP}^f), \tag{20}$$

is derived using Equation (17).

In particular, an element (i, j) of the analysis error covariance matrix, $\mathbf{P}^a$, can be conveniently written as

$$\mathbf{P}^a_{ij} = \mathbf{P}^f_{ij} - \mathbf{p}_i^T\boldsymbol{\Gamma}^{-1}\mathbf{p}_j, \tag{21}$$

where

$$\boldsymbol{\Gamma} = \mathbf{HP}^f\mathbf{H}^T + \mathbf{R}, \tag{22}$$

represents the modelled innovation covariance matrix. The analysis error variance, i.e., when $j = i$, is simply computed as

$$(\sigma_i^a)^2 = (\sigma_i^f)^2 - \mathbf{p}_i^T\boldsymbol{\Gamma}^{-1}\mathbf{p}_i. \tag{23}$$

The analysis error variance is then used as an initial condition to integrate the variance through Equation (9). This completes the algorithm.

The next section deals with the three-dimensional observation operator (Equation (4)), which includes the vertical structure of the averaging kernel, rendering the analysis update more complex.

3.4. Analysis Step with 3D Observation Operator Using Averaging Kernels

The three-dimensional observation operator is given in Equation (4). First, it consists of interpolating the 3D field horizontally, as in Equation (12), to get vertical profiles at the observation location and then applying the vertical observation operator, $\mathbf{H}_v$, using the instrument averaging kernel to get the observation equivalent quantity. The vertical observation operator in function form is denoted as $H_{v(z)}$. To apply it, first, we use a mapping function (i.e., vertical interpolation,V_z, on z) to convert the model vertical concentrations to the vertical layers of the observation retrieval, so that $\mathbf{y}^m$ (see Equation (1)), the a priori $\mathbf{y}^p$, and the averaging kernel $\mathbf{A}$ are in the same vertical coordinates. Thus, in an observation location, we get

$$V_z[f_o(z,t)] \Rightarrow \mathbf{y}^m(z^*,t), \tag{24}$$

where f_o denotes the vertical model concentrations at the observation location and z^* is the vertical coordinate as in averaging kernels.

In the second step, $H_{v(z)}$ carries out $\mathbf{y}^m(z^*,t)$ in Equation (1) and computes a single value equivalent to the retrieval,y^m (See Section 2.2 for an explanation to Equation (1)).

$$H_{v(z)}[V_z[f_o(z,t)], \mathbf{A}, \mathbf{y}^p, \boldsymbol{\omega}] \Rightarrow y^m. \tag{25}$$

Note that the quantities in bold ($\mathbf{A}$, $\mathbf{y}^p$, $\boldsymbol{\omega}$, $\mathbf{y}^m$) are defined on the vertical layers of the averaging kernel, and the non-bold (y^m) is in fact a scalar quantity representing the column-averaged quantity. Therefore, $H_{v(z)}$ maps a vector in the model vertical space to a scalar in the observation space.

The observation operator H, which is applied on functions, is the composition of $H_{v(z)}$ with H_h,

$$H = H_{v(z)} \circ H_h \Leftrightarrow \mathbf{H}_v(\mathbf{H}_h(\ )), \tag{26}$$

which is equivalent to the matrix form of the $\mathbf{H}$ in Equation (4).

Before applying the observation operator on a covariance function, we need to specify on which variable it is applied. If it is applied on the spatial variable, $\mathbf{x}$, it is equivalent, in matrix form, to a left multiplication of the observation operator on the covariance matrix,

$$H_{\mathbf{x}}[P^f(\mathbf{x},\mathbf{x}^{'},t)] \ \Leftrightarrow \ \mathbf{H}\,\mathbf{P}^f. \tag{27}$$

However, if the observation operator is applied on the spatial variable, $\mathbf{x}^{'}$, it is then equivalent to a right multiplication of the transpose of the observation operator on the covariance matrix,

$$H_{\mathbf{x}^{'}}[P^f(\mathbf{x},\mathbf{x}^{'},t)] \ \Leftrightarrow \ \mathbf{P}^f\mathbf{H}^T. \tag{28}$$

Using the separable form of the covariance function in Equation (10), we thus get

$$\begin{aligned} H_{\mathbf{x}}[P^f(\mathbf{x},\mathbf{x}^{'},t)] &= H_{v(z)}\Big\{\sigma^f(x_o,y_o,z,t)\,C_h(x_o,y_o,x',y')\,C_v(z,z')\,\sigma^f(x',y',z',t)\Big\} \\ &= H_{v(z)}[\sigma_o^f(z)\ C_v(z,z')]\ \times\ C_h(x_o,y_o,x',y')\,\sigma^f(x',y',z',t) \end{aligned} \tag{29}$$

where $H_{v(z)}$ is to denote, in functional form, the vertical observation operator operating on z (and not z'). However, to use the vertical observation operator, the expression in square brackets in Equation (29) has to be written in vector/matrix form. Note that in matrix form, $\sigma_o^f(z,t)\,C_v(z,z')$ is an element of an $N_{lev} \times N_{lev}$ matrix, where N_{lev} is the number of vertical levels in the model. This matrix can actually be written as $\mathbf{diag}(\sigma_o^f)\mathbf{C}_v$, where $\mathbf{diag}(\sigma_o^f)$ is an $N_{lev} \times N_{lev}$ diagonal matrix where the elements on the diagonal are $\sigma_o^f(z,t)$. Therefore, the application

$$H_{v(z)}[\sigma_o^f(z,t)\,C_v(z,z')] \ = \ f_o(z',t) \ \ \Rightarrow \ \ \mathbf{f}_o^T \text{ a row vector } 1 \times N_{lev}\,, \tag{30}$$

is a row vector that depends on the (second) vertical coordinate, z'. Overall, the application of the observation operator in Equation (29) produces a 3D spatial field of elements,

$$H_{\mathbf{x}}[P^f(\mathbf{x},\mathbf{x}^{'},t)] \ = \ C_h(x_o,y_o,x',y')\,f_o(z',t)\ \sigma^f(x',y',z',t)\,. \tag{31}$$

Based on Equation (28), a similar expression is obtained,

$$H_{\mathbf{x}^{'}}[P^f(\mathbf{x},\mathbf{x}^{'},t)] \ = \ C_h(x,y,{x'}_{o(2)},{y'}_{o(2)})\,f_{o(2)}(z,t)\ \sigma^f(x,y,z,t)\,. \tag{32}$$

where we let the reference observation location be $o(2)$. To compute the (simulated) innovation covariance matrix (i.e., $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$) we combine Equation (27) with Equation (28),

$$H_{\mathbf{x}}\Big[H_{\mathbf{x}^{'}}\Big(P^f(\mathbf{x},\mathbf{x}^{'},t)\Big)\Big] \ \ \Leftrightarrow \ \ \mathbf{H}\mathbf{P}^f\mathbf{H}^T, \tag{33}$$

and account for a pair of non-coincident observations, $o(1)$, $o(2)$, we get

$$\begin{aligned} &H_{\mathbf{x}}[C_h(x,y,{x'}_{o(2)},{y'}_{o(2)})\,f_{o(2)}(z,t)\ \sigma^f(x,y,z,t)\,] \\ &\quad= \ H_{v(z)}\Big\{C_h(x_{o(1)},y_{o(1)},\ {x'}_{o(2)},\ {y'}_{o(2)})\ f_{o(2)}(z,t)\,\sigma^f_{o(1)}(z,t)\Big\} \\ &\quad= \ C_h(x_{o(1)},y_{o(1)},\ {x'}_{o(2)},\ {y'}_{o(2)})\ H_{v(z)}[f_{o(2)}(z,t)\,\sigma^f_{o(1)}(z,t)] \\ &\quad= \ C_h(x_{o(1)},y_{o(1)},\ {x'}_{o(2)},\ {y'}_{o(2)})\ \times\ \beta_{o(1),o(2)}(t)\,, \end{aligned} \tag{34}$$

where $\beta_{o(1),o(2)}$ is an element of a $N_{lev} \times N_{lev}$ covariance matrix,

$$\beta_{o(1),o(2)} = H_{v(z)}\left\{H_{v(z')}\left[\sigma^{f}_{o(1)}(z,t)\,\sigma^{f}_{o(2)}(z',t)\;C_v(z,z')\right]\right\}. \quad (35)$$

3.5. An Overview of the Assimilation Algorithm

Figure 1 presents a flowchart of the forecast and analysis steps of the PvKF algorithm described above. The algorithm's forecast step (left side) involves two parallel model simulations, one for the state's transport with the atmospheric diffusion equation [64,65] and the other for the advection of the error variance (Equation (9)). In the first part of the analysis step (right side), the covariance function is convolved with the observation operator on the left for a series of available observations at a given time (Equation (27)). This is followed by a second convolution with the observation operator on the right for the same observations to obtain the innovation covariance matrix (Equations (28) and (33)–(35)). The second part of the analysis step (Equation (22)) performs a Cholesky decomposition and inversion of the modelled innovation covariance matrix. Finally, the analysis and analysis error variance increment are computed for each set of observations at a time t (Equations (16) and (23)). A complete PvKF assimilation algorithm in a matrix form is presented in Appendix B.

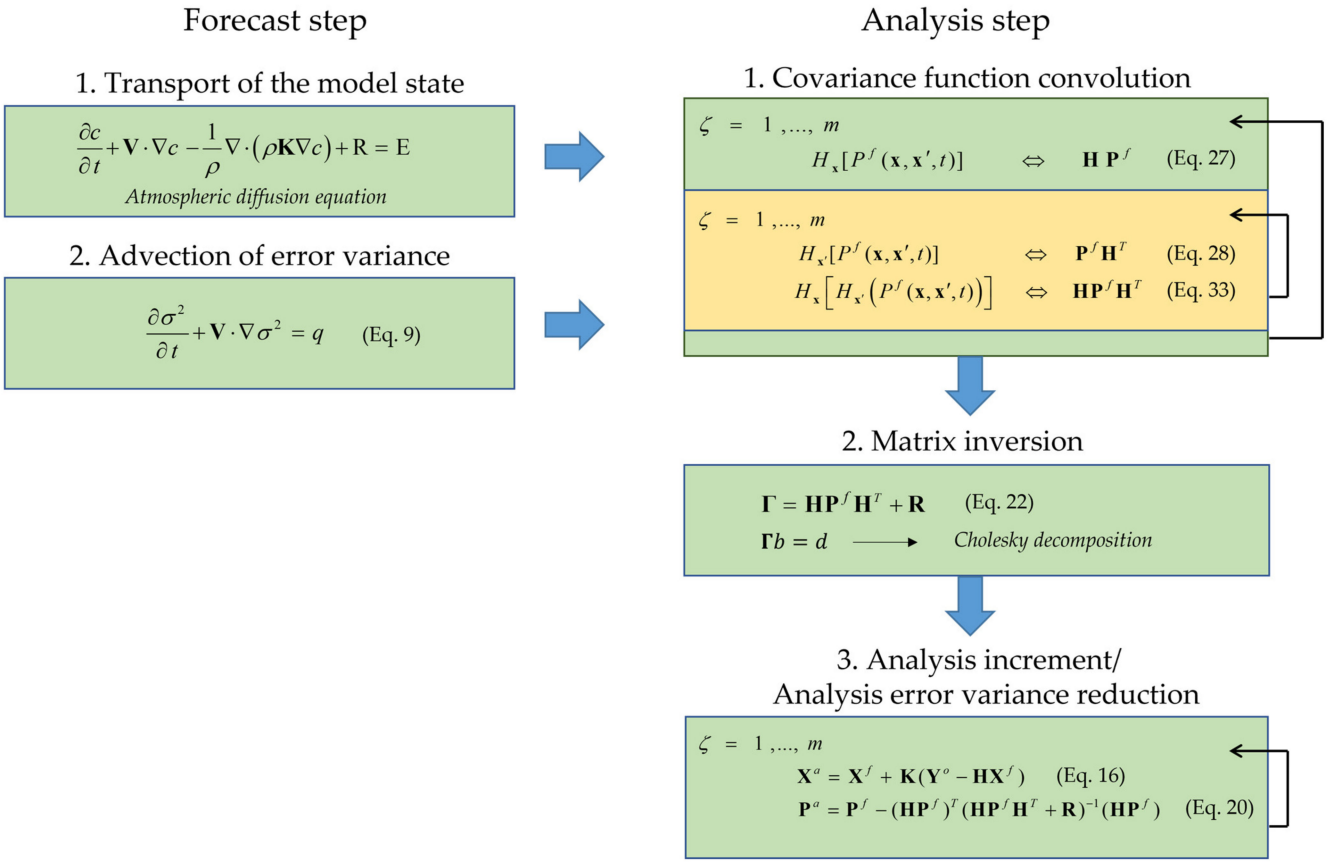


Figure 1. Flowchart of the analysis and forecast steps of the PvKF assimilation. In the forecast step, the transport of the model state uses the atmospheric diffusion equation [64,65].

As mentioned earlier, the PvKF assimilation relies on the continuous properties of operators, which makes it a suitable scheme for evolving the analysis and its error variance. An example of the error variance evolution with synthetic GOSAT observations over land is shown in Figure 2, emphasizing the combination of the 3rd part of the analysis step and the 2nd part of the forecast step of the flowchart (Figure 1) in creating the analysis error variance evolution. In Figure 2, the error variance is initiated with a constant field

using a 5% error (standard deviation) of the globally averaged methane concentration (Figure 2a). The reduction of error variance occurs in the presence of every observation at a particular time and location based on the correlation function (Figure 1–analysis step) while the updated error variance field is propagated with the flow (Figure 1–forecast step). The reduction gradually grows over land (Figure 2b) by assimilating a new batch of observations over the same regions according to GOSAT revisit time (i.e., 3 days).

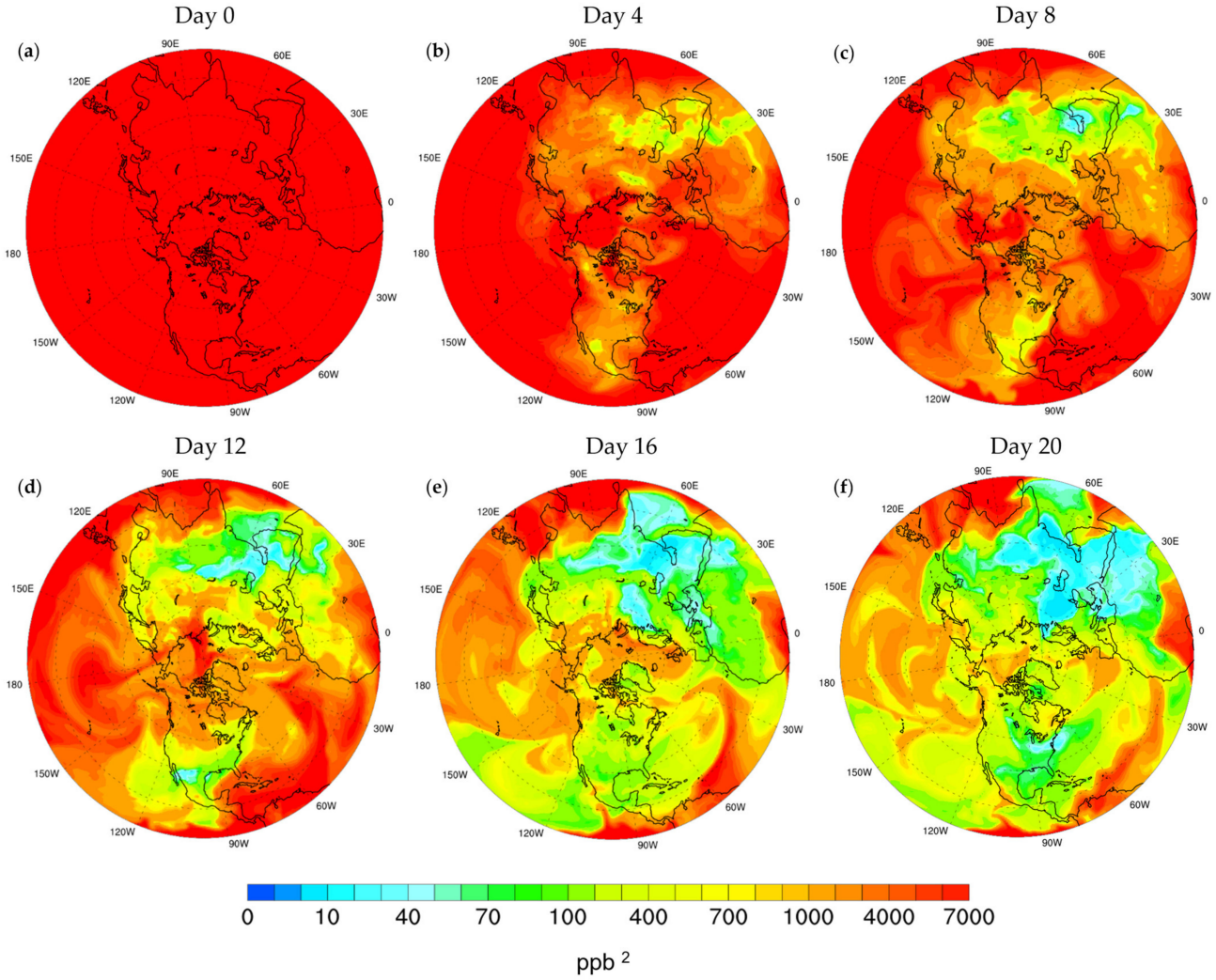


Figure 2. Evolution of the analysis error variance on (**a**) Day 0, (**b**) Day 4, (**c**) Day 8, (**d**) Day 12, (**e**) Day 16, and (**f**) Day 20.

After 8 days, the analysis error variance field (Figure 2c) shows a noticeable reduction over Western Asia, followed by Central Asia, Eastern Africa, and the southern/eastern part of North America, mainly due to denser GOSAT observations and higher innovations at those regions. Then, owing to the advection, the reduction of error variance spread out with the flow over other regions with fewer or no observations, such as oceans and higher latitudes. On day 12 (Figure 2d), most of the lands, except polar regions and near the Equator, are subjected to a significant error reduction, suggesting a higher impact of the GOSAT observations due to their density over those areas. The maximum reduction (light blue) indicates a 20–40 times reduction in the initial error, while assimilating more observations within PvKF after 16 days does not provide a tangible reduction in the analysis error (Figure 2e). Nonetheless, over time, the variance reduction significantly spreads out over the entire domain, particularly over the oceans with no observations (Figure 2f). This

highlights the behaviour of the assimilation system that provides an error estimation over the whole domain.

4. System Setup

4.1. Initial Conditions

The impact of the initial condition can persist long in the model. For different species, it is recommended that H-CMAQ be initialized to the clean background and driven based on its emissions and chemical transport [47]. However, due to an almost well-known background field of methane, the long lifetime of atmospheric methane (~10 years) [43], and the high level of uncertainties in its source and sink [68], conducting a lengthy model integration might be unnecessary. Furthermore, in the context of data assimilation, the sensitivity to the initial conditions disappears rapidly with time [35].

In this study, we initialize the model using prescribed vertical profiles as well as a 2D concentration field obtained from previous global analyses. Our model setup uses the approach of Olsen et al. (2013) [69], which is derived from a nonlinear polynomial fit to global models and various types of measurements from 2003 to 2006. This initial field, taken as a first guess, is a function of latitude and altitude varying smoothly from the North Pole to the Equator with no temporal and longitudinal variation. The details of this preparation can be found in Xiong et al. (2008) [70]. Maintaining the shape of the vertical profiles, we rescale the initial field to our simulation period, April 2010, based on the annual/monthly increase in globally-averaged atmospheric methane [71].

With the aim to remove the potential biases in the model, we obtain accurate surface observations during the assimilation period from GLOBALVIEWplus CH4-ObsPack v3.0 [72] compiled by National Oceanic and Atmospheric Administration (NOAA) (Figure 3a). In our bias correction, we assume that a significant part of bias arises from the emissions at the surface, and the rate of change of concentration with height remains the same (shape of the vertical profile). In fact, we multiply the vertical profiles by a constant that depends on the discrepancy at the surface (we assume that vertical mixing in the troposphere is unchanged). Note that the model vertical coordinate is extended up to the upper troposphere/lower stratosphere, and the profiles are adapted from the analysis of global methane studies [69,70]. This assumption is made mainly due to the fact that other accurate observations such as Total Carbon Column Observing Network (TCCON) are used for the validations of our assimilation results (see Section 5 of Part II).

The bias removal involves adjusting the initial concentration field (i.e., initial guess) using a latitudinal linear regression model to match the surface observations (i.e., flasks and towers). Accordingly, first, the monthly averaged linear fit of Observation—Model against latitude using the first guess model initial conditions is obtained (Figure 3b–green dots). Next, the achieved fit is rescaled (with respect to globally averaged concentration from NOAA [71]) to our simulation period in April 2010 (Figure 3c–red dots), and eventually, the fit is employed to remove the bias of the entire space of the initial conditions (Figure 3d–blue dots). It indicates that the latitude-based bias in the initial conditions is subtracted, resulting in a better (unbiased) agreement of the model and the surface observations. This could imply that the polynomial function [69] deriving our first guess initial conditions is too smooth; thus, it tends to overpredict the lower latitude concentrations compared to the surface in-situ observations. In addition, the hemispheric absolute mean bias is decreased from 17.1 ppb for the rescaled initial to 4.3 ppb for the bias-corrected initial. Note that the negative values in Figure 3b–d correspond to the positively biased model forecasts (i.e., model overestimation). We consider the model that is initialized with the bias-corrected field as our nearly unbiased model and rely on this to remove the potential bias between GOSAT and H-CMAQ. An illustration of the bias correction of GOSAT with respect to H-CMAQ is provided in the following subsection. Note that the bias correction here is only applied for one month (i.e., April 2010) and may not be representative for another month or a longer period; thus, one needs to take into account the same type of bias correction for the time of assimilation.

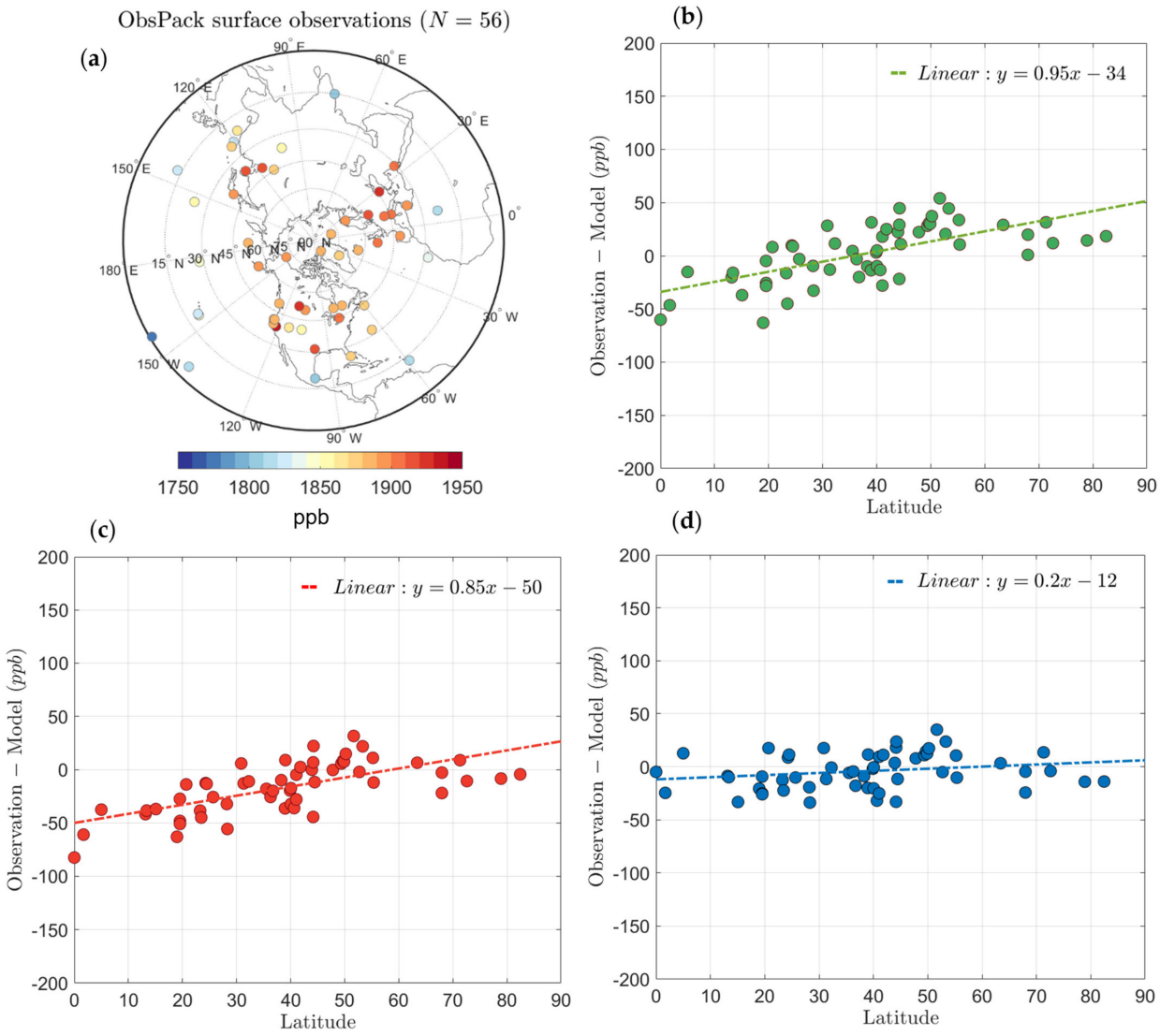


Figure 3. Difference between (**a**) methane surface observations (GLOBALVIEWplus surface flask and tower, shown top left) and model equivalent values of concentration for (**b**) the initial guess (green dots/line), (**c**) rescaled initial to April 2010 but before bias correction (red dots/line), and (**d**) after bias correction (blue dots/line).

4.2. Observation Bias Correction

GOSAT provides high accuracy retrieval data (~0.7% precision) with reasonable near-surface sensitivity and global coverage, making it a strong candidate for methane assimilation analysis [55,73]. The potential bias in GOSAT XCH_4 retrieval has been addressed previously by evaluating the data against other types of measurements [74–77]. Both PR and FP versions of GOSAT data used here are post-processed and validated against surface-based Fourier transform infrared (FTRS) methane column abundance from TCCON. It was shown that the difference between XCH_4 retrieval and TCCON correlates with the albedo, α, at 1.6 nm in band 2 [57].

In addition to the retrieval bias, GOSAT can still have biases relative to atmospheric chemical transport models. This bias most likely originates from the emissions as well as the limitation of global models to realistically simulate the methane in the stratosphere, particularly at higher latitudes [78–80]. For example, Turner et al. (2015) [58] showed that their GEOS-Chem simulation of GOSAT features a positive latitudinal bias, where the mean Model—Observation is adjusted based on a fit to a quadratic regression function.

Cressot et al. (2014) [52] obtained a linear regression fit as a function of air mass factor (AMF) to correct for the biases in GOSAT with their chemistry-transport model, LMDz-SACS. Following them, we parametrize the bias in GOSAT with respect to our unbiased CMAQ simulation. First, the model is mapped to observation space using our observation operator (**H**) described in Section 2.2. Next, GOSAT—CMAQ is obtained as a function of retrieval parameters. Accordingly, a separate linear regression fit of GOSAT—CMAQ is computed for each of latitude, air mass factor (AMF), solar zenith angle θ_s, and satellite viewing zenith angle, θ_v (Figure 4). Except for the θ_v with no specific correlation pattern, the largest correlation of GOSAT—CMAQ is found with respect to latitude $(r = 0.43)$, followed by $\theta_s (r = 0.40)$ and AMF $(r = 0.33)$. All three fits show similar behaviour, indicating that the discrepancy is smoothly increasing toward the larger values of those parameters (Figure 4a–c). Our latitude dependency of the bias agrees with Turner et al. (2015) [58], but the AMF dependency disagrees with Cressot et al. (2014) [52]. Perhaps it is related to the fact that LMDz is a low-resolution general circulation model (GCM), whereas CMAQ and GEOS-Chem appear more as chemical transport models with higher resolution. The positive bias of CMAQ (Model > Observation) was found previously in other atmospheric chemistry models such as GEOS-Chem and addressed mainly as extratropical stratospheric bias due to the extra meridional stratospheric transport [80–82] as well as polar vortices [83].

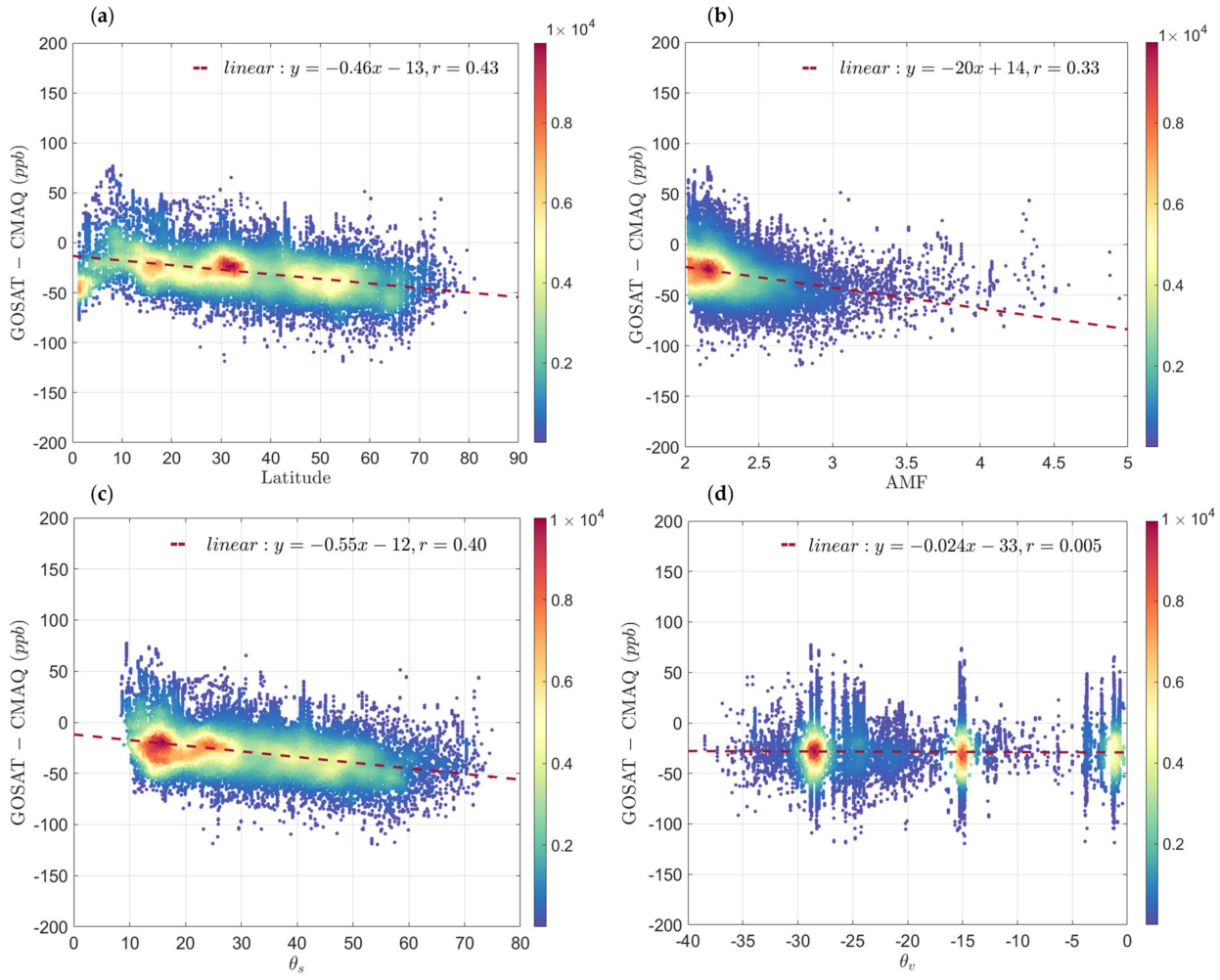


Figure 4. Linear regression fit of the difference between GOSAT observation and H-CMAQ simulated XCH4 as a function of (**a**) latitude, (**b**) air mass factor (AMF), (**c**) solar zenith angle, θ_s, and (**d**) satellite viewing zenith angle, θ_v. The colorbar indicates the number of GOSAT observations, with a higher observations density in red.

Figure 5 compares the Observation—Model before bias correction (GOSAT^0—CMAQ) and after bias correction with respect to latitude (GOSAT$^{bias(lat)}$—CMAQ). The hemispheric absolute mean bias (*MB*) significantly decreased from 32.9 ppb to 0.9 ppb, while the residual standard deviation (σ) turned out to be slightly smaller (20.1 ppb to 18.1 ppb) after bias correction. As expected from the correlation values in Figure 4, using AMF or θ_s instead of latitude to remove the potential bias of GOSAT resulted in a less significant bias reduction because of a larger hemispheric σ and *MB*. Figure S1 in the Supplementary Materials indicates a similar comparison using bias-corrected GOSAT data with respect to AMF and θ_s.

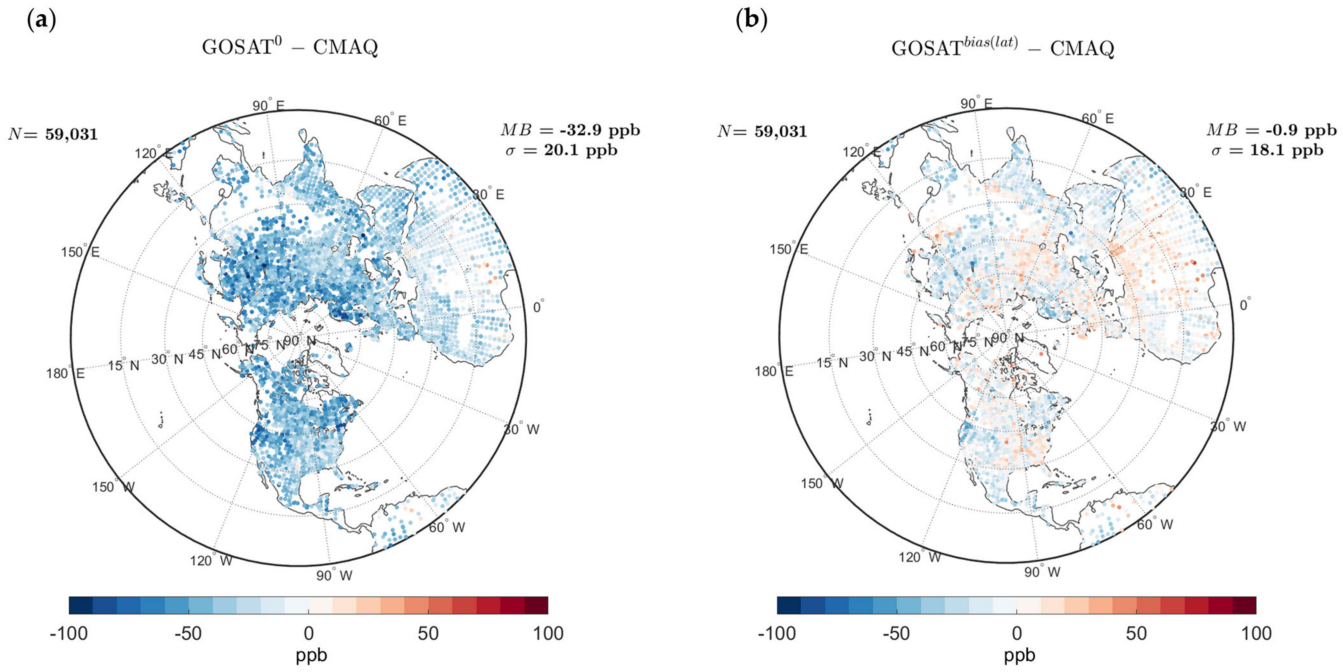


Figure 5. (**a**) Difference between observations and model before bias correction (GOSAT^0—CMAQ) and (**b**) after bias correction with respect to latitude (GOSAT$^{bias(lat)}$—CMAQ) over a month with the number of observations N = 59,031. A bias-corrected field of residuals calculated from the regression with respect to AMF and θ_s is provided in Figure S1.

In addition, in an attempt to maximize the information in the regression model, we adopted a simple multivariate linear regression [84]. However, we found that such an algorithm cannot considerably improve our prediction model (see Tables S1–S3 in Supplementary Materials). This suggests that all three parameters are dependent and can be highly correlated to each other. Therefore, a bias-corrected GOSAT—CMAQ with only latitudinal adjustment could provide a reasonable bias correction to our data. After bias correction, the model discrepancy with observations can be primarily random and could be attributed to the random emission errors, the correlation lengths (L_h, L_v), and random errors in the observations (ε^o), initial (ε^i), and forward model (ε^q), for which we provide a description in Section 4.4.

4.3. Construction of Spatial Correlation Functions on the H-CMAQ Grid

The H-CMAQ grid is uniform on the projected plane. Nearly all spatial correlation models rely on an underlying uniform and isotropic grid representation. This is what most variational assimilation systems take into account, and further, assume either an infinite or periodic domain. Here, we will explain how to construct spatial error correlations on the H-CMAQ grid while representing an underlying homogeneous isotropic correlation function. The procedure consists in constructing a homogeneous isotropic and periodic

correlation function on the surface of a sphere that is independent of the grid, and then mapping it onto the H-CMAQ grid.

The construction of homogeneous isotropic correlation functions on the surface of a sphere has been developed in Gaspari and Cohn (1999) [45]. Background information on the method is also given in Ménard (2000) [35]. A correlation function is a function of a pair of points $(i\,,\ j)$. Consider a sphere of radius a, with position vectors (from the centre of the sphere) $\mathbf{r}_i$, $\mathbf{r}_j$ for the points i and j. We define a Cartesian coordinate system in $\mathbb{R}^3$,(x^C, y^C, z^C), where we use the superscript C to recall the Cartesian space. A point i (or j) on the surface of the sphere that has (in a spherical coordinate system) a longitude φ_i and colatitude (or inclination) θ_i is related to the Cartesian coordinates, as follows

$$\begin{aligned} x_i^C &= a\ \cos\varphi_i\ \sin\theta_i \\ y_i^C &= a\ \sin\varphi_i\ \sin\theta_i\,, \\ z_i^C &= a\cos\theta_i \end{aligned} \tag{36}$$

and similarly for the grid point j. The chordal distance D_{ij} between the position vectors $\mathbf{r}_i = (x_i^C, y_i^C, z_i^C)$ and $\mathbf{r}_j = (x_j^C, y_j^C, z_j^C)$ is given by

$$D_{ij} = \left\|\mathbf{r}_i - \mathbf{r}_j\right\| = \sqrt{2}\,a\left\{1 - \left[\sin\theta_i\,\sin\theta_j\,\cos(\varphi_i - \varphi_j) + \cos\theta_i\cos\theta_j\right]\right\}^{1/2}, \tag{37}$$

Using a chordal distance to define a "distance" for a correlation function on a sphere has the double advantage that a correlation model originally defined in $\mathbb{R}$ can be used in $\mathbb{R}^3$ to define a correlation on a sphere and has the property of being periodic [45].

Three models have been considered in this study: the Gaussian (or double exponential) model,

$$C_G(i,j) = \exp\left(-\frac{D_{ij}^2}{2\,L_c^2}\right), \tag{38}$$

the First-Order-Auto-Regressive (FOAR), or exponential model,

$$C_{FOAR}(i,j) = \exp\left(-\frac{D_{i\ j}}{L_c^*}\right) \quad \text{where } L_c^* = L_c/(0.5005), \tag{39}$$

and the Second-Order-Auto-Regressive (SOAR) model

$$C_{SOAR}(i,j) = \left(1 + \frac{D_{ij}}{L_c^{**}}\right)\exp\left(-\frac{D_{i\ j}}{L_c^{**}}\right) \quad \text{where } L_c^{**} = L_c/(1.3494). \tag{40}$$

Note that L_c denotes a correlation length, defined from the curvature at the origin, while L_c^* or L_c^{**} is a model parameter adapted from Ménard et al. (2016) [85]. We recommend always using L_c, as it is a physically meaningful quantity.

The second step consists of mapping the H-CMAQ grid onto the corresponding point on the surface of the sphere (see Appendix A for information on the polar stereographic projection used in H-CMAQ). As mentioned in Gaspari and Cohn [45] (1999), if there exists a one-to-one transformation,

$$T(\varphi,\theta) = (x^p, y^p), \tag{41}$$

(see Equation (A1)) then we can define a correlation function with respect to the (x^p, y^p) coordinates. The computation then goes as follows

$$\left\{\begin{array}{c}(x_i^p, y_i^p)\\(x_j^p, y_j^p)\end{array}\right\} \overset{T^{-1}}{\rightarrow} \left\{\begin{array}{c}(\varphi_i, \theta_i)\\(\varphi_j, \theta_j)\end{array}\right\} \rightarrow D_{ij} \rightarrow\ C(D_{ij})\,, \tag{42}$$

where C is one of the correlation models above (Equations (38)–(40)).

4.4. Observation, Model and Initial Error Covariance Modelling

As in any assimilation system, the input error covariances need to be specified. In this section, we describe the modelling and assumptions of those error covariances. It is generally assumed that observation errors, model errors, and initial errors are uncorrelated, and furthermore that the observation and model errors are serially (temporally) uncorrelated. These are the standard assumptions used to derive the Kalman filter. We should note that because of the dynamics and the interplay with the analysis, as the forecast error becomes realistic, it also becomes correlated in space and time and with past observation errors (this is a standard result in Kalman filtering theory; see, for example, Jazwinski (1970) [86]).

Let us assume that the observation error **R** and the model error **Q** are both spatially uncorrelated. The observation error has two components: the measurement error (ε^m) provided by the instrument team and the representativeness error (ε^r) arising from the mismatch between the subgrid-scale represented in the observation and the gridded model quantity [87,88]. We assume that the observation error covariance is diagonal and takes the form of $\mathbf{R} = (\varepsilon^o)^2\mathbf{I}$.

Many studies estimate the representativeness error as an additive source of error to the measurement error [35,36,89–91]. However, we tune the overall observation error by considering a multiplicative correction factor applied to the measurement error as denoted in Equation (43). A multiplicative factor is consistent with other types of errors (i.e., modelling and initial error) used in this study. In fact, it is used in the majority of data assimilation papers [35,92]. Therefore, the representativeness error is included as part of the correction factor f^o.

$$\mathbf{R}^{'} = (f^o\varepsilon^m)^2\mathbf{I}, \tag{43}$$

where $\mathbf{R}^{'}$ denotes the observation error covariance after tuning. Note that initially, no representativeness error is accounted for $\mathbf{R}^{'}$.

The model error (ε^q) is assumed to be proportional to the analysis [35]. We simply assume a uniform accumulation of model error in time, where it becomes almost equivalent to a particular time average of the analysis, $\overline{X}^a(\mathbf{x}, t)$, (i.e., monthly averaged analysis is used in this study). Model error covariance is considered as a diagonal matrix, $\mathbf{Q} = (\varepsilon^q)^2\mathbf{I}$ [93,94], and a relative model error standard deviation,f^q, is defined for tuning **Q**, which has the form

$$\mathbf{Q}^{'} = (f^q X^a(\mathbf{x}, t))^2\mathbf{I}, \tag{44}$$

where $\mathbf{Q}^{'}$ denotes the model error covariance after tuning. Note that model error variance, q, in Equation (9) is equivalent to the diagonal elements of $\mathbf{Q}^{'}$.

We attempt to tune the initial concentration error (ε^i) separately, although it is often integrated as part of the modelling error. The initial error covariance matrix is also assumed to be diagonal. We assume that the initial error before tuning is about 5% of the initial concentration (i.e., $\varepsilon^i = 0.05\ X_0^f$) with the same spatial distribution. By conducting a similar analysis as in Section 4.2 on the first day of the simulation, we found that the mean standard deviation of the residual (Observation—Model) is relatively smaller (~20 ppb ≡ 1.2%). The parametric form of the tuned initial error covariance is

$$\mathbf{P}_0{}' = \left(f^i\varepsilon^i\right)^2\mathbf{I}, \tag{45}$$

where $\mathbf{P}_0{}'$ represents the initial error covariance after tuning.

5. Verification of the Basic Properties of the Assimilation System

5.1. One-Observation Experiment

We conduct a one-observation experiment using a single simulated observation. It is a standard experiment to verify that the mechanics of the assimilation system is working and can also be used to verify the validity of some assumptions in the formulation [27,95]. The one-observation experiment can also be used to compare different assimilation systems [96].

When a single observation is assimilated, the analysis increment has the same spread in space as the spatial correlation function—a property that can be derived directly from the analysis equation and Kalman gain.

With the one-observation experiment, first, we will verify that our formulation of error covariances (in $\mathbb{R}^3 \times \mathbb{R}^3$ and projected onto the polar stereographic grid of H-CMAQ) does yield a homogenous isotropic correlation. Secondly, we will show that our assumption to use only advection to propagate the error variance is adequate for assimilation on a time scale relevant to GOSAT assimilation. Errors in the wind, emissions, and any errors not accounted for our simple advection of error variance such as horizontal and vertical diffusion effect on error variance will contribute to the model error.

The assimilation system is examined with a single simulated observation and arbitrary error parameters, including $f^i = 0.05$, $f^q = 0.015$, $f^o = 0.5$. Note that the initial concentration is equivalent to the bias-corrected concentration field (see Section 4.1), and initial variance is obtained from Equation (45) where $\varepsilon^i = 0.05\ X_0^f$. These quantities are kept the same for all the cases in this experiment. In this experiment, we consider the SOAR correlation model with a horizontal correlation length of 500 km and 10 model vertical layers (starting from the surface) of the sigma-pressure coordinate (referred to $10\sigma_l$ length scale). Figures 6 and 7 show two single-observation experiments—one with an observation in the high latitude (71° N, 155° E) and the other closer to the Equator (27° N, 105° W). In both experiments, the assimilation started at Day 0 with a single synthetic observation. The synthetic observations represent column-averaged concentrations (Equation (1)) generated with the model while accounting for the GOSAT averaging kernels and a priori. To define the magnitude of the synthetic observations, we assume a multiplicative factor of 1.4 (40% higher) to the model column-averaged concentrations. The assimilation consists of propagating the analysis and its error variance for a duration of 3 days after the single observation is taken into account. We only examine 3 days, which corresponds to the GOSAT revisit time, that is, the time after which a new batch of observations becomes available for the same region.

For the observation in the high latitude (Figure 6a–d), the analysis increment and the error variance reduction are shown near the surface, whereas for the observation in the lower latitude (Figure 7a–d), they are demonstrated at about 600 hPa (~layer 23). By comparing Figure 6a with Figure 7a, we note that the error correlation, captured by the analysis increment, is indeed homogeneous and isotropic. The apparent increase in radius of the spatial spread in lower latitudes is due to the polar stereographic projection (we verified that the length scale is, in fact, the same). By comparing the propagated analysis increment with the propagated error variance reduction (i.e., Figures 6b,d and 7b,d), we found that the transport of the analysis increment and the reduction of error variance are quite similar in terms of their spread and spatial distribution. This suggests that the propagation of the error variance using advection only is a reasonable assumption to make, since it maintains a similar structure as the methane fields. Furthermore, we found a comparable weight of the analysis error variance at the start day (Day 0) and after 3 days; in particular, we note that the maximum reduction of error variance has not changed more than 5% after 3 days (i.e., $\frac{\left(\left(\sigma^a_{(Day\ 3)}\right)^2 - \left(\sigma^f_{(Day\ 3)}\right)^2\right)_{\max}}{\left(\left(\sigma^a_{(Day\ 0)}\right)^2 - \left(\sigma^f_{(Day\ 0)}\right)^2\right)_{\max}} \sim 1$). This result contrasts with the variance losses that are known to occur in Kalman filter and ensemble Kalman filter systems [42,66,97].

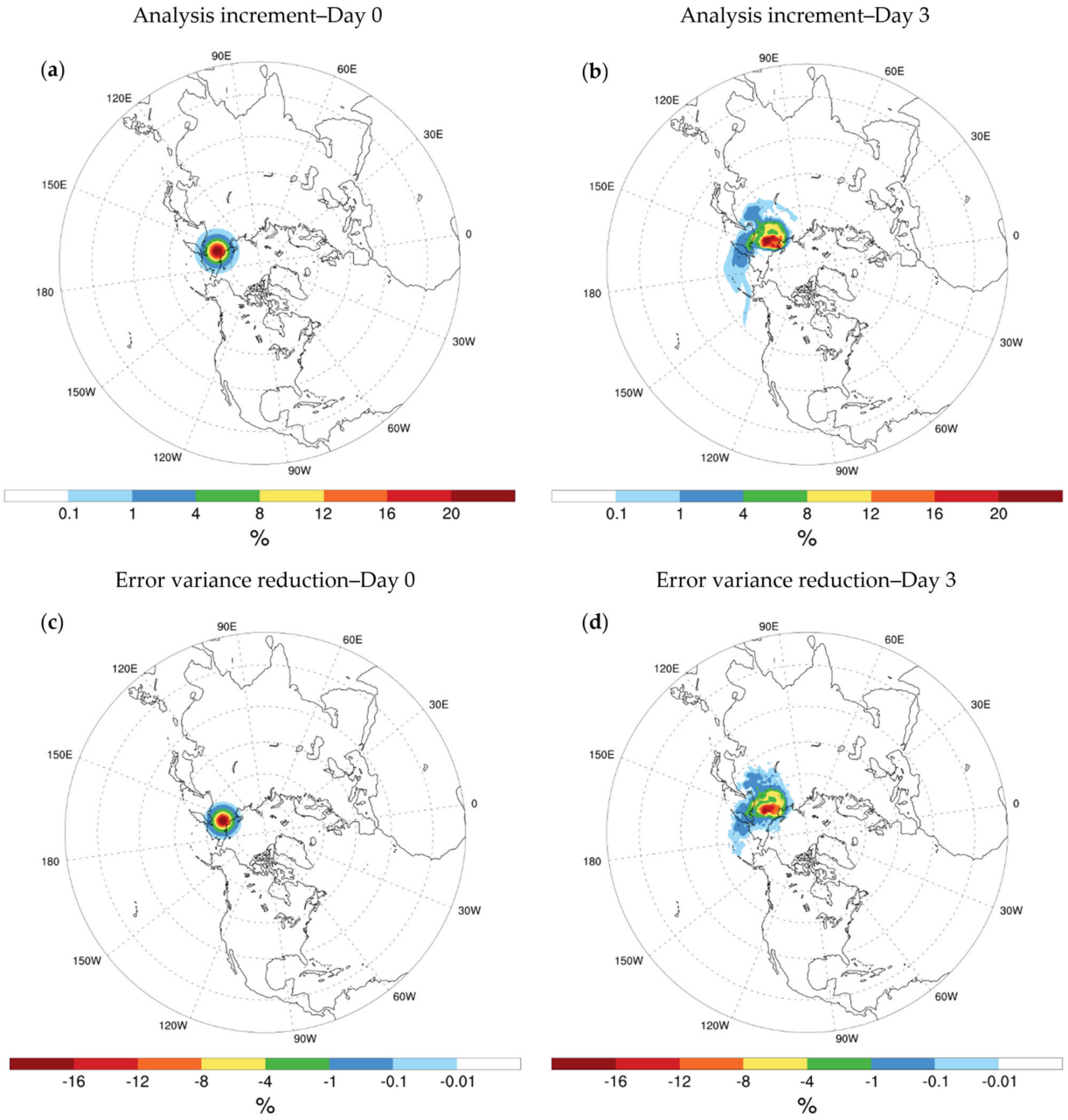


Figure 6. The first one-observation experiment near the surface at higher latitude (71° N, 155° E) to show (**a**) the analysis increment at Day 0 and (**b**) Day 3; and (**c**) the error variance reduction at Day 0, and (**d**) Day 3.

In another series of experiments, we examine the contribution of error correlation to address the vertical estimation properties, especially the quantities at the surface. To illustrate the propagation in the vertical direction, we obtain a vertical cross-section (Figure 7a–d) passing through the observation at Day 0, starting from (28° N, 120° W) to (15° N, 80° W). The observation is at the center of the analysis increment and center of the error variance reduction (Figure 7a,c), also called increment/reduction. Figure 8a,c illustrates the increment/reduction at the moment when a single observation is assimilated. Several factors such as the averaging kernel, the model layer pressure weight (ω), and the correlation length scale based on the SOAR correlation model define the increment/reduction patterns of the analysis and error variance. A relatively large correlation length ($L_v = 10\sigma_l$) results in a stretched pattern of increment/reduction along the vertical direction. The vertical distribution also indicates that they are mainly influenced by ω (Figure 9c), resulting in a

maximum increment/reduction at mid-troposphere (~600 hPa, or 23rd layer). On Day 3, these quantities are shown on the same plane. The largest increments/reductions are shifted toward the southeast of the initial point (Figure 7d), mainly due to advective transport. The large distension is mainly influenced by the wind pattern, while the diffusion improves the smoothness of the field, especially in lower levels (Figure 6b,d). We should note that the higher increment/reduction to the lower troposphere and near-surface enhances the ability of the assimilation in constraining the surface quantities such as emissions within a short period of assimilation integration (i.e., 3 days).

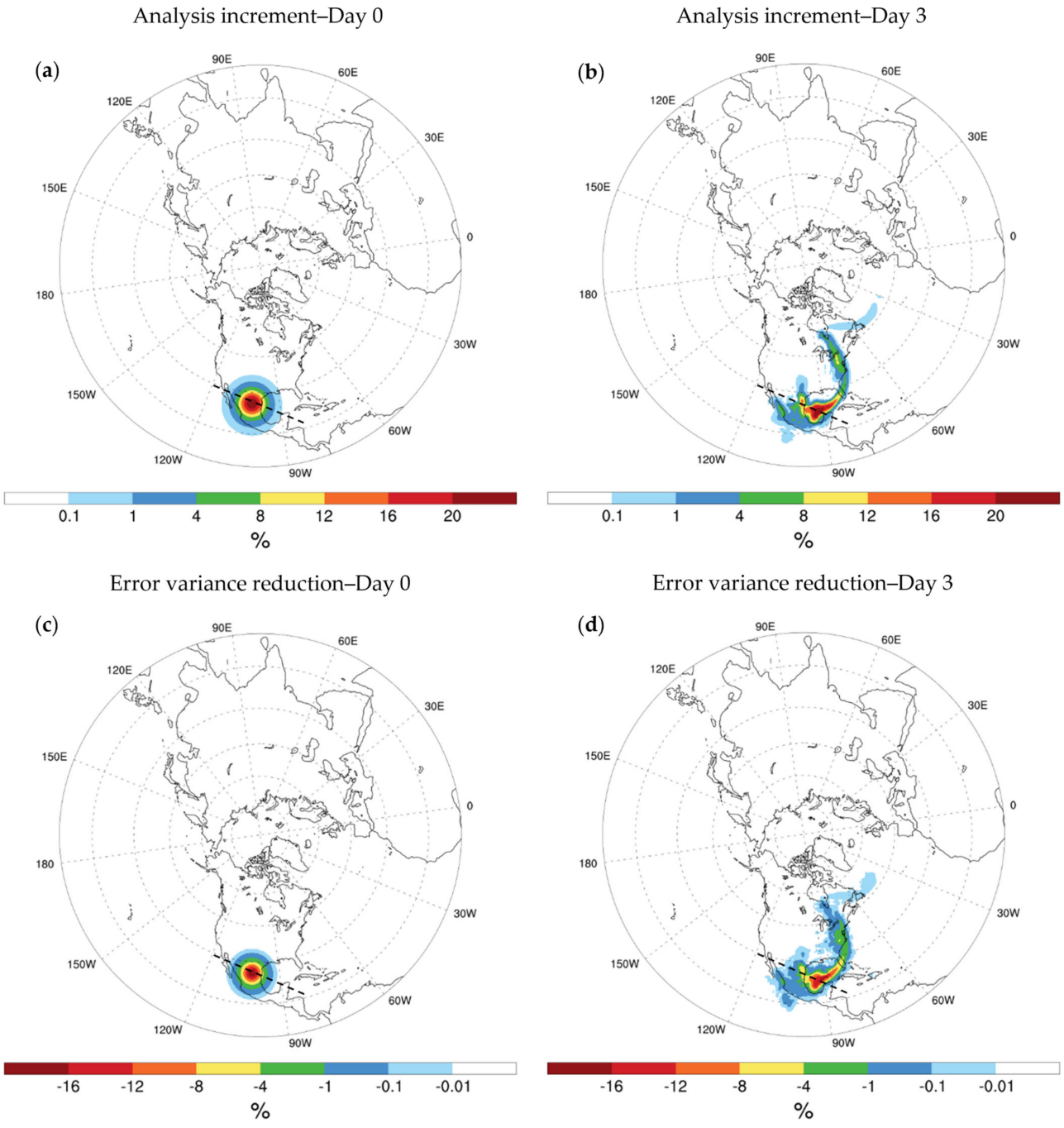


Figure 7. The second one-observation experiment at about 600 hPa and lower latitude (27° N, 105° W) to show (**a**) the analysis increment at Day 0 and (**b**) Day 3; and (**c**) the error variance reduction at Day 0, and (**d**) Day 3. The dashed (black) line represents a cross-section along the vertical direction, where the vertical distribution of the analysis and its error variance are demonstrated in Figure 8.

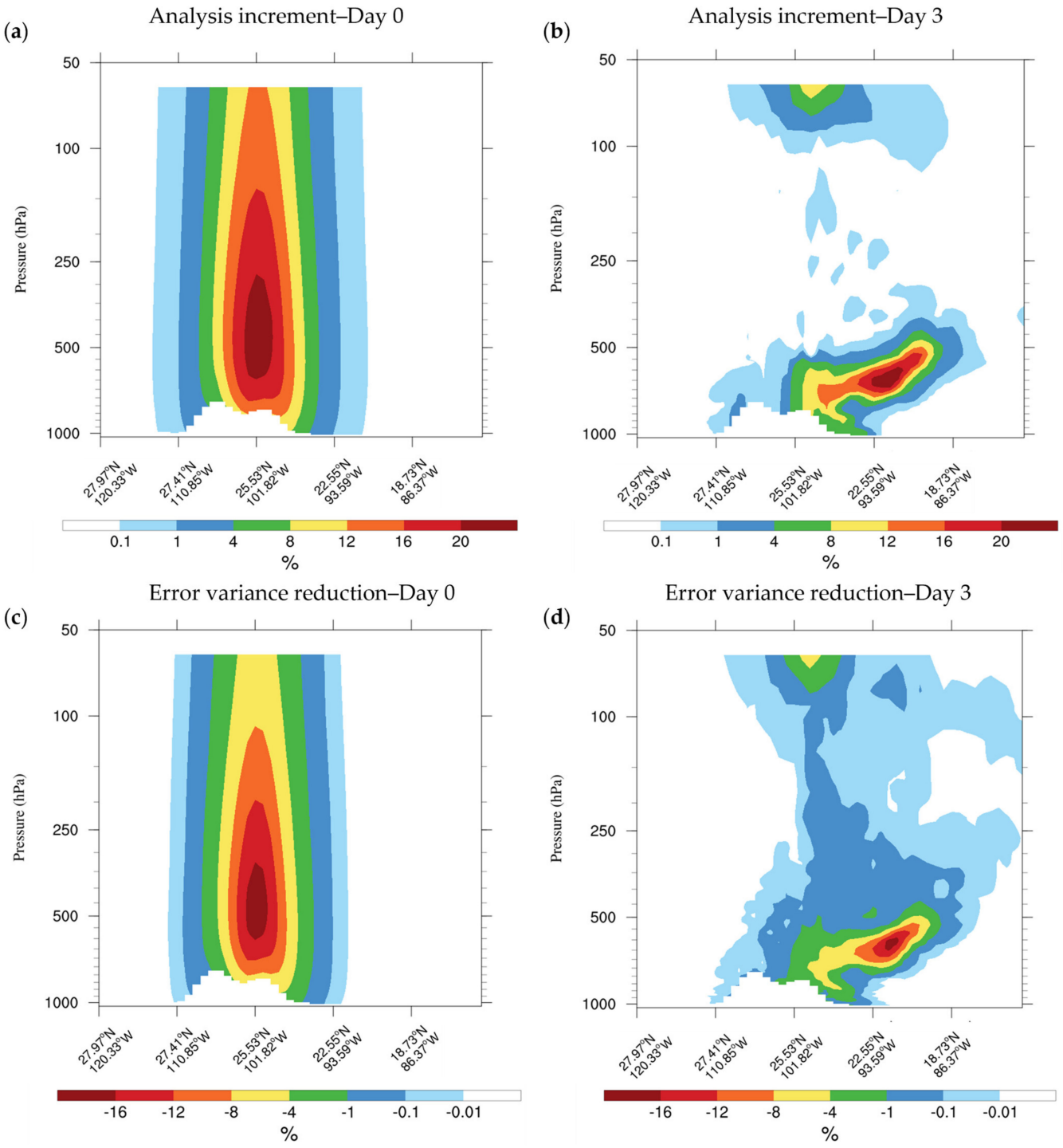


Figure 8. Vertical distribution of (**a**) analysis increment at Day 0 and (**b**) Day 3 along the cross-section (black dashed line) shown in Figure 7a,b; Vertical distribution of (**c**) error variance reduction at Day 0 and (**d**) Day 3 along the cross-section (black dashed line) shown in Figure 7c,d. The cross-section starts from (28° N, 120° W) to (15° N, 80° W).

Figure 8b,d indicate that there is reasonable consistency between the propagation of analysis increment and variance reduction in the vertical direction. It shows that the vertical effects from a single observation can persist for at least 3 days. Furthermore, a significant part of analysis increment and error variance reduction remains in the lower elevation with a downward shift of the maximum values. These increments/reductions are quite small between the mid to upper troposphere, except for another slight increase at

the upper layers, which occurs likely due to a zero-flux assumption across the model top boundary [46,98]. Hence, similar to the surface, part of the increments/reductions lingers near the upper layers.

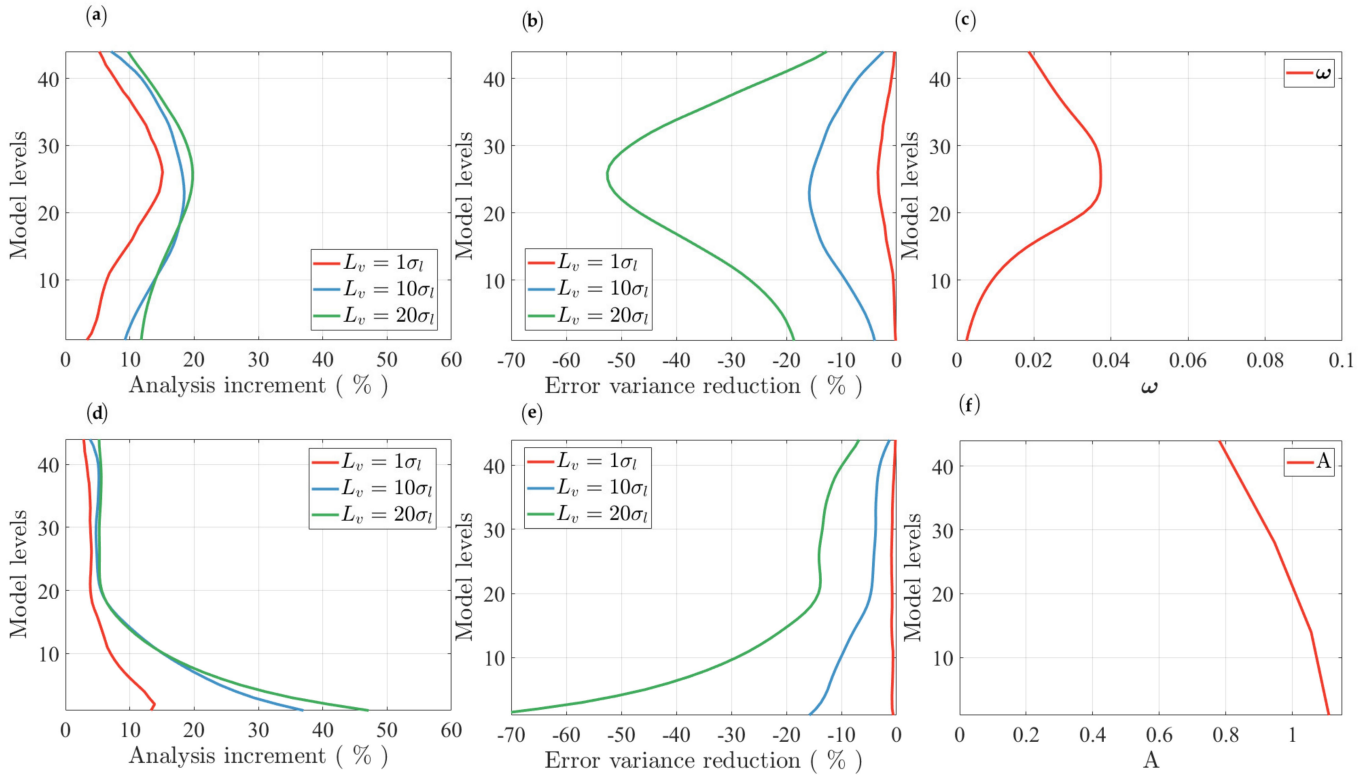


Figure 9. (**a**) Analysis increment and (**b**) error variance reduction during assimilation step for different vertical correlation length scales (red: $L_v = 1\sigma_l$, blue: $L_v = 10\sigma_l$, and green: $L_v = 20\sigma_l$); (**c**) Pressure weight (ω) at each model layer; (**d**) analysis increment and (**e**) error variance reduction where the pressure weight is considered uniform; (**f**) GOSAT column averaging kernel.

We examine (Figure 9) the impact on the analysis increments and error variance reductions for different vertical correlation lengths ($L_v = 1\sigma_l$, $10\sigma_l$, $20\sigma_l$). It indicates that the error variance reduction is significantly more sensitive than the analysis increment to the vertical correlation length scale. As mentioned earlier, the layer pressure weights normalize the adjustments and allocate more impact on the mid-layers. We remove this effect by considering uniform layer pressure weights, which results in emphasizing the influence of the averaging kernels on the vertical profile of the analysis increment and error variance reduction during the analysis step (Figure 9a,b,d,e). Thus, higher sensitivity of GOSAT retrieval to the surface associated with its average kernel (Figure 9f) results in a similar effect on the analysis (Figure 9d). Likewise, the error variance, whose vertical sensitivity is more substantial than the analysis, exhibits a larger influence from the surface (Figure 9e). It implies that for observations with the highest retrieval sensitivity to the near-surface (e.g., GOSAT), integrating the reduction of error variance with the assimilation scheme enhances our ability to retrieve information from the surface. For a single-observation problem, one can theoretically show that the error variance reduction maintains a quadratic relationship with the analysis increments (Equation (23)). Therefore, besides the continuous estimate of the analysis and its uncertainty, the distinctive feature of PvKF in explicitly computing the error variance can be encouraging to infer surface quantities (i.e., emission inversion).

5.2. Timing (Computational Efficiency)

The computational cost of advanced assimilation schemes is often compared against the cost of one model integration. The cost of the forecast step in PvKF involves two model simulations: the forecast of concentration and the forecast of the error variance. Thus, the forecast step of PvKF requires a little less than twice the computational time of the H-CMAQ forecast simulation, since the chemistry and diffusion are turned off for the propagation of error variance. When observations and analyses are included, the total computational cost depends on how many observations are assimilated per time step. The computational time required to perform the analysis step increases rapidly as the number of observations increases, due to the inversion of the innovation covariance matrix (or solving the equivalent linear problem) and the computation of the analysis error variance. Note that, contrary to 4D-Var, the storage space of PvKF always remains the same for any length of integration.

The computational cost of assimilation with real GOSAT data (developed by SRON/KIT) should also take into account the number of actual quality-controlled observations. Obtaining quality-controlled observations consists of several steps. In addition to using the quality-controlled observations provided by the instrument team, we apply two additional filters on real GOSAT data before including them in the assimilation. First, we remove outliers whose departure from the global mean of the observations is three times larger than its standard deviation, and second, we conduct a thinning process on the observations with the aim of providing uncorrelated observation errors [99] (see Section 3 of Part II for more details). Accordingly, the filtered GOSAT data holds below 100 retrievals per hour, which is about one-third of observations before filtering (i.e., full GOSAT data). Note that the quality control steps are performed as pre-processing steps; hence, they do not impact the computational cost of assimilation.

In order to make an assessment of the computational cost of the PvKF algorithm, we investigate the assimilation of an arbitrary number of observations per hour. For this, we have an arbitrary number of simulated observations that imitate the behaviour and influence of real GOSAT observations. All the simulated observations are retrieval products and have averaging kernels and a priori columns similar to the retrieved XCH_4, and are basically proxy retrievals (i.e., PR) of GOSAT observations developed by SRON/KIT. The advantage of this approach is that we can generate an arbitrary number of "retrieved" observations. Therefore, the simulated observations could be generated at other times than for the retrievals. Both model (H-CMAQ) and analysis are carried out for a one-hour simulation of size $187 \times 187 \times 44$ in the computational domain with 108 km resolution. Figure 10 compares the computational time of the analysis step (dashed red line) as normalized to one model integration (solid black line). Contrary to the PvKF forecast (solid blue line), the analysis cost largely varies depending on the number of assimilated observations, attaining an equivalent cost to one model integration at about 850 observations. The total assimilation time of PvKF (dashed green line) is obtained by adding the analysis time with approximately twice the model running time. For the number of observations equivalent to the real GOSAT data (both filtered and full data), the analysis step is much shorter than the model (i.e., ~0.1 per model integration). In this case, matrix inversion of the innovation covariance matrix (i.e., $\mathbf{\Gamma}^{-1} = \left(\mathbf{H}\mathbf{P}^{f}\mathbf{H}^{T} + \mathbf{R}\right)^{-1}$ in Equations (17) and (20) uses Cholesky decomposition [100] to solve either the inversion of the innovation covariance matrix or a system of linear equations $\mathbf{\Gamma} b = d$. Thus, our hourly assimilation scheme with real GOSAT data maintains approximately twice the H-CMAQ computational time. Nonetheless, for the higher number of measurements (e.g., >500) per hour generated synthetically with GOSAT, our computational time estimate increases exponentially to near $O(p^3)$, which is comparable to the order of a Cholesky decomposition to compute the matrix inversion (i.e., it is dominated by the analysis step computation).

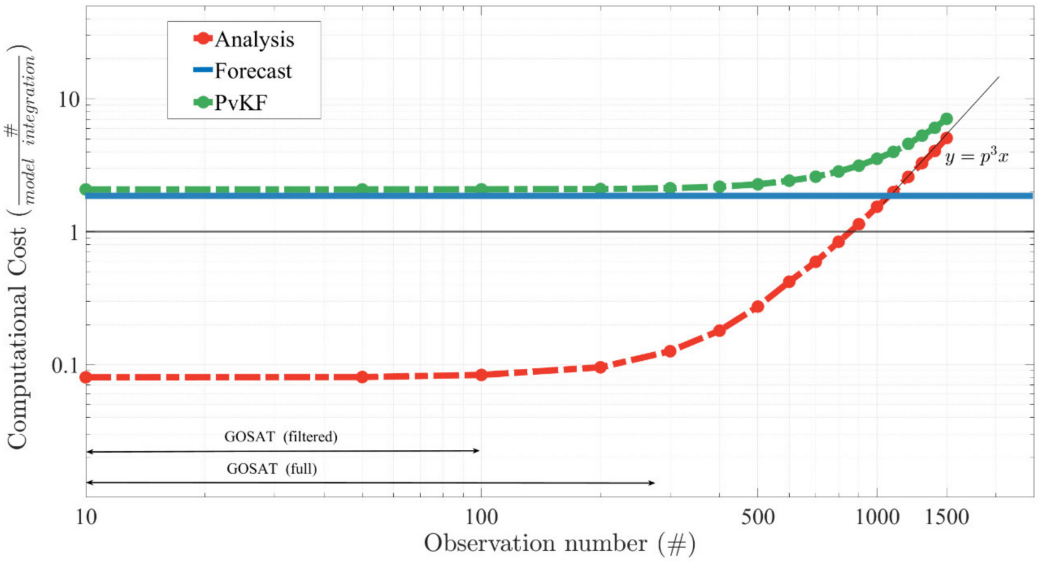


Figure 10. 1-h computational time of the analysis (dashed red line) with respect to the number of observations within the PvKF assimilation. The solid black line shows the computational time of a 1-h H-CMAQ simulation with a known computational configuration. The solid blue line denotes the forecast step of PvKF that includes two forecasts, one for the methane transport and the other for the advection of error variance. The dashed green line represents the total computational time expected for the PvKF.

As a remark, we can roughly compare the computational efficiency of the PvKF assimilation described above with other popular schemes such as 4D-Var and EnKF. For example, 4D-Var entails the adjoint of CMAQ where the computation of the backward mode requires at least twice (and up to three times) the time as the forward simulation [101]. In addition, several forward-backward iterations are required until the algorithm's convergence. The adjoint computation also requires a relatively large (~10 times the model) amount of storage space [102]. The computational time of EnKF assimilation also highly depends on the number of ensembles carried out with the model that adopts it. In general, dozens of ensembles are expected to maintain proper assimilation [32]; for example, Peng et al. (2015) [103] used 48 ensembles in their CFI-CMAQ. In summary, PvKF assimilation can achieve a high level of computational efficiency compared to the other typical assimilation schemes. We demonstrate the performance of this framework using real GOSAT observations in the second part of this paper.

6. Summary and Conclusions

We have designed an assimilation system based on the parametric (variance only) Kalman filter, or PvKF, with the hemispheric CMAQ model and GOSAT methane observations (the assimilation scheme can be used for any long-lived species). The scheme is capable of providing analysis together with its uncertainty (i.e., error variance) while being computationally cheaper than other popular data assimilation schemes such as 4D-Var and EnKF. The analysis is derived sequentially with a minimum hourly batch of observations that could maintain real-time assimilation. The uncertainty is obtained by advection of error variance using the H-CMAQ model with a predefined error correlation model. The assimilation system was tested with a single simulated observation experiment to verify basic properties, as well as the conservation of information and the appropriateness of using advection of error variance. This scheme does not assume a perfect model, nor does it require the development of adjoint or ensemble simulations.

To conduct PvKF methane assimilation, we modified H-CMAQ to include methane transport, chemical reaction with OH, and emissions from both anthropogenic (EDGAR v6) and natural (WetCHARTS v1.0) source categories. Preparation of an unbiased initial

field and addressing the bias in GOSAT with respect to independent surface measurements (GLOBALVIEWplus CH4-ObsPack v3.0 compiled by NOAA) is demonstrated. We found that a latitudinal correction provides a reasonable bias correction to our data. Thus, the model discrepancy with observations can be primarily attributed to the random errors in the model, observations, and emissions.

Results using simulated observations indicate that the expected behaviour of the analysis error variance and analysis increment, which is derived from the model, is fairly consistent, while the information content (i.e., total variance) is conserved. Note that this may not be the case for EnKF if inflation is not added. We also demonstrate that the effect of a single observation can persist within a period of the GOSAT revisiting time, which is about 3 days. In addition, it is shown that the vertical error correlation could assist in deducing quantities at or close to the surface.

We also discussed the computational cost of the PvKF assimilation against an arbitrary number of observations per time step. We found that the assimilation scheme with GOSAT maintains approximately twice the H-CMAQ computational time, while a larger number of observations per hour (e.g., >500), which is not the case of GOSAT, suggests an exponential increase in the computational time. Nonetheless, with a couple of thousands of observations per hour, the method still performs acceptably but is slower. We emphasize that this assimilation system does not assume a perfect model, in agreement with the fact that the wind and emissions are not entirely known. Yet, PvKF is computationally advantageous compared to 4D-Var and EnKF.

The main limitation of this method is related to the lifetime of the species. PvKF is well-adapted to long-lived species, such as methane, and still applies to shorter lifetime chemical species, but with the caveat that a smaller fraction of the total forecast error variance is explained by the advection of error variance. In this case, the residual error variance (i.e., unexplained error variance) is captured by the stationary model error. Thus, for chemical species with a shorter lifetime, the PvKF more resembles an OI (Optimal Interpolation) scheme. Another limitation is that the framework's feasibility depends on the observation characteristics (e.g., observation number and density). The larger number of observations we assimilate, the more accurate analysis we may obtain. However, the assimilation scheme is limited to a certain number of observations due to the computational capacity of PvKF, as explained in Section 5.2. In addition, increasing the number of observations can result in a higher spatial density, which increases the error correlation in observation space. This contradicts the necessary condition to obtain an optimal PvKF analysis (see Part II, Sections 2 and 3). Therefore, the number of observations may limit both the efficiency and the quality of the assimilation.

In general, we found that the PvKF algorithm is sufficiently adaptable as a lightweight scheme for carrying long-lived tracers inside a chemistry-enabled atmospheric model. In a companion paper, we will discuss this framework with the objective of obtaining optimal assimilation of GOSAT and deducing realistic statistics for transport, observations, and model parameters, including correlation lengths.

A potential application of this algorithm, which has not yet been pursued in this study, is to improve the inverse modelling of methane emissions on a highly resolved regional domain by integrating an accurate initial and the inflows of methane concentrations at the lateral boundaries of the regional model and their uncertainties.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs14020371/s1, Figure S1: (a) GOSAT$^{\text{bias(AMF)}}$—CMAQ with air mass factor bias correction, (b) GOSAT$^{\text{bias}(\theta_s)}$—CMAQ with solar zenith angle bias correction. *MB* and σ denote domain-wide mean bias and standard deviation of the residual. Table S1: Evaluating the multiple (iterative) regression algorithm based on two measures: Mean Square Error (*MSE*) and the correlation coefficient r. Latitude, θ_s, and AMF represent the order of parameters in the multiple regression algorithm, respectively. Table S2: Evaluating the multiple (iterative) regression algorithm based on two measures: Mean Square Error (*MSE*) and the correlation coefficient r. θ_s, latitude, and AMF represent the order of parameters in the multiple regression algorithm, respectively. Table S3:

Evaluating the multiple (iterative) regression algorithm based on two measures: Mean Square Error (*MSE*) and the correlation coefficient *r*. AMF, θ_s, and latitude represent the order of parameters in the multiple regression algorithm, respectively. Figure S2: Hemispheric spatial distribution of the relative methane concentration loss due to chemical reactions. It shows daily average values after two weeks of the model forecast.

Author Contributions: Conceptualization, S.V. and R.M.; methodology, S.V. and R.M.; software, S.V.; formal analysis, S.V.; investigation, S.V.; writing—original draft preparation, S.V. and R.M.; writing—review and editing, R.M., T.W.W. and A.H.; visualization, S.V.; supervision, R.M., T.W.W. and A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The GOSAT satellite data are available from the European Space Agency Greenhouse Gases Climate Change Initiative at http://cci.esa.int/ghg (last access: 5 November 2021). The GLOBALVIEWplus CH4 ObsPack v3.0 data product is available at www.esrl.noaa.gov/gmd/ccgg/obspack/data.php?id=obspack_ch4_1_GLOBALVIEWplus_v3.0_2021-05-07, accessed on 20 November 2021 (Cooperative Global Atmospheric Data Integration Project, 2019). EDGAR v6 anthropogenic emission inventories are available at https://edgar.jrc.ec.europa.eu/dataset_ghg60, accessed on 20 November 2021 (European Commission, 2021). Modelling data and the source code for the PvKF assimilation have been deposited in GitHub, https://github.com/Sinavo/PvKF_reg_methane.git, accessed on 20 November 2021. Additional information related to this paper may be requested from the corresponding author: Sina Voshtani (Seyyedsina.Voshtani@ec.gc.ca).

Acknowledgments: The authors of this paper acknowledge all of the data providers and laboratories (https://search.datacite.org/works/10.25925/20210401, last access: 4 November 2021) that contributed to the GLOBALVIEWplus CH4 ObsPack v3.0 data product compiled by NOAA Global Monitoring Laboratory. We also appreciate the free use of CH_4 emissions data from Emissions Database for Global Atmospheric Research (EDGAR) (https://edgar.jrc.ec.europa.eu/index.php/dataset_ghg60, last access: 20 November 2021). Simulations for this study were performed on computational resources of Compute Canada (http://www.computecanada.ca, accessed on 20 November 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Polar Stereographic Projection

H-CMAQ adapts the horizontal coordinates and the projections, including the polar stereographic projection, specified in MCIP [104]. For the Northern Hemisphere, the projection starts from the South Pole and runs through each point on the Earth's surface to the projection plane. Considering the reference projection parameters specified in H-CMAQ (see Figure A1, i.e., the true latitude θ_0 and longitude φ_0), the transformation equations from the Earth's surface in S to the projection plane in $\mathbb{R}^2\left(x_i^p, y_i^p\right)$, where x_i^p and y_i^p denote the coordinates of the projected plane, are

$$\begin{aligned} x^p &= r\sin(\varphi-\varphi_0) \\ y^p &= -r\cos(\varphi-\varphi_0) \\ r &= \gamma r_e \cos\theta \\ \gamma &= \tfrac{1+\sin(\theta_0)}{1+\sin(\theta)} \end{aligned} \tag{A1}$$

where φ and θ are the corresponding longitude and the latitude of an arbitrary point (e.g., observation point G in Figure A1) and r_e is the radius of the Earth. Note that the y-axis in the projection plane, as shown in Figure A1, describes the meridian and γ denotes the polar stereographic map scale factor.

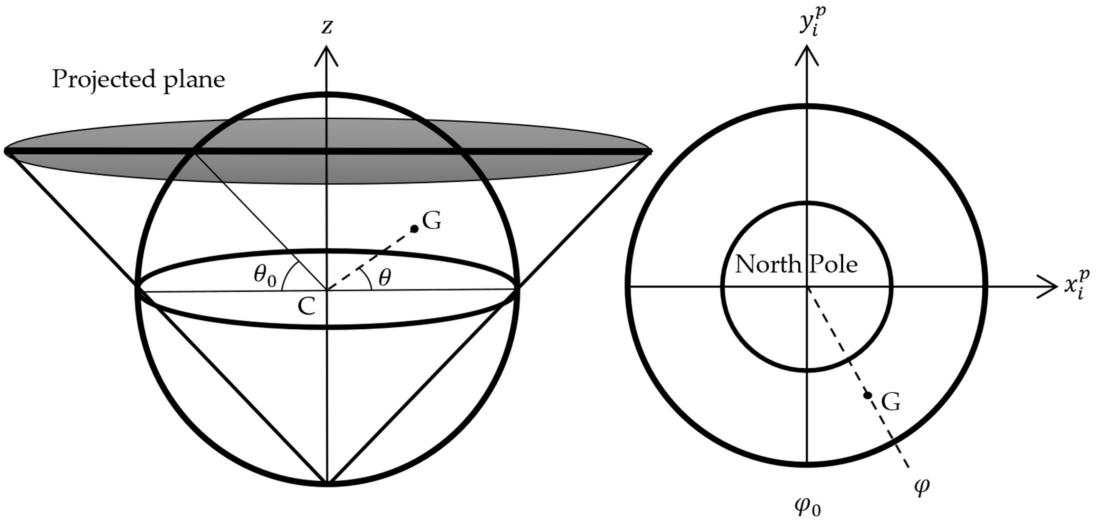


Figure A1. Polar stereographic projection geometry.

For the transformation into the horizontal gridded model, we use the latitude and longitude of the center of each gridcell. According to CMAQ documentation [98], we account for the Earth radius of 6,370,000 m, the true latitude of 45°, and the meridian taken as the true longitude. Note that the polar stereographic projection is integrated into the observation operator, so that each observation point is transformed using the same set of parameters and equations described above.

Appendix B. PvKF Data Assimilation Algorithm

Algorithms A1. Parametric variance Kalman filter (PvKF) assimilation.

Pre-processing of observations (quality control, bias correction, etc.)
Initialization of state vector, $\mathbf{X}_0^f$, and covariance matrix, $\mathbf{P}_0^f$
for $k = 0, \ldots, K$ **do**
——————————**Analysis step**——————————
  if $\mathbf{Y}_k^o \neq \varnothing$ **then**
$\mathbf{K}_k = (\mathbf{H}_k\mathbf{P}_k^f)^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}$
$\mathbf{X}_k^a = \mathbf{X}_k^f + \mathbf{K}_k(\mathbf{Y}_k^o - \mathbf{H}_k\mathbf{X}_k^f)$
$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f = \mathbf{P}_k^f - (\mathbf{H}_k\mathbf{P}_k^f)^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}(\mathbf{H}_k\mathbf{P}_k^f)$
  else
$\mathbf{X}_k^a \leftarrow \mathbf{X}_k^f$
$\mathbf{P}_k^a \leftarrow \mathbf{P}_k^f$
  end if
——————————**Forecast step**——————————
    $\mathbf{X}_{k+1}^f = \mathbf{M}_{k:k+1}\,\mathbf{X}_k^a$
    $\mathbf{P}_{k+1}^f(i,i) = \mathbf{M}_{k:k+1}^*\,\mathbf{P}_k^a(i,i)$
    Note: $\mathbf{M}_{k:k+1}$: H-CMAQ $\leftarrow$ CH_4
        $\mathbf{M}_{k:k+1}^*$: H-CMAQ (Advection-only) $\leftarrow$ Inert tracer
——————————————————————————
end for

References

1. Saunois, M.; Stavert, A.R.; Poulter, B.; Bousquet, P.; Canadell, J.G.; Jackson, R.B.; Raymond, P.A.; Dlugokencky, E.J.; Houweling, S.; Patra, P.K.; et al. The Global Methane Budget 2000–2017. *Earth Syst. Sci. Data* **2020**, *12*, 1561–1623. [CrossRef]
2. Myhre, G.; Shindell, D.; Bréon, F.; Collins, W.; Fuglestvedt, J.; Huang, J.; Koch, D.; Lamarque, J.; Lee, D.; Mendoza, B.; et al. Anthropogenic and natural radiative forcing. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2013; pp. 659–740.
3. Etminan, M.; Myhre, G.; Highwood, E.J.; Shine, K.P. Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.* **2016**, *43*, 12614–12623. [CrossRef]
4. Butler, J.H.; Montzka, S.A. The NOAA Annual Greenhouse Gas Index (AGGI). Available online: https://gml.noaa.gov/aggi/aggi.html (accessed on 5 October 2021).
5. WMO Greenhouse Gas Bulletin (GHG Bulletin), No.17: The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2020. Available online: https://library.wmo.int/doc_num.php?explnum_id=10904 (accessed on 2 November 2021).
6. Dlugokencky, E.J.; Nisbet, E.G.; Fisher, R.; Lowry, D. Global atmospheric methane: Budget, changes and dangers. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **2011**, *369*, 2058–2072. [CrossRef]
7. Fletcher, S.E.M.; Schaefer, H. Rising methane: A new climate challenge. *Science* **2019**, *364*, 932–933. [CrossRef] [PubMed]
8. National Academies of Sciences, Engineering, and Medicine. *Improving Characterization of Anthropogenic Methane Emissions in the United States*; National Academies Press: Washington, DC, USA, 2018. [CrossRef]
9. Nisbet, E.G.; Fisher, R.E.; Lowry, D.; France, J.L.; Allen, G.; Bakkaloglu, S.; Broderick, T.J.; Cain, M.; Coleman, M.; Fernandez, J.; et al. Methane Mitigation: Methods to Reduce Emissions, on the Path to the Paris Agreement. *Rev. Geophys.* **2020**, *58*, 51. [CrossRef]
10. Fiore, A.M.; Jacob, D.J.; Field, B.D.; Streets, D.G.; Fernandes, S.D.; Jang, C. Linking ozone pollution and climate change: The case for controlling methane. *Geophys. Res. Lett.* **2002**, *29*, 25-1–25-4. [CrossRef]
11. Forster, P.; Ramaswamy, V.; Artaxo, P.; Berntsen, T.; Betts, R.; Fahey, D.W.; Haywood, J.; Lean, J.; Lowe, D.C.; Myhre, G. Changes in atmospheric constituents and in radiative forcing. Chapter 2. In *Climate Change 2007: The Physical Science Basis*; IPCC: Geneva, Switzerland, 2007.
12. Brasseur, G.P.; Jacob, D.J. *Modeling of Atmospheric Chemistry*; Cambridge University Press: New York, NY, USA, 2017; p. 606.
13. Jacob, D.J.; Turner, A.J.; Maasakkers, J.D.; Sheng, J.X.; Sun, K.; Liu, X.; Chance, K.; Aben, I.; McKeever, J.; Frankenberg, C. Satellite observations of atmospheric methane and their value for quantifying methane emissions. *Atmos. Chem. Phys.* **2016**, *16*, 14371–14396. [CrossRef]
14. Turner, A.J.; Frankenberg, C.; Kort, E.A. Interpreting contemporary trends in atmospheric methane. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2805–2813. [CrossRef] [PubMed]
15. Qu, Z.; Jacob, D.J.; Shen, L.; Lu, X.; Zhang, Y.Z.; Scarpelli, T.R.; Nesser, H.; Sulprizio, M.P.; Maasakkers, J.D.; Bloom, A.A.; et al. Global distribution of methane emissions: A comparative inverse analysis of observations from the TROPOMI and GOSAT satellite instruments. *Atmos. Chem. Phys.* **2021**, *21*, 14159–14175. [CrossRef]
16. Lu, X.; Jacob, D.J.; Zhang, Y.Z.; Maasakkers, J.D.; Sulprizio, M.P.; Shen, L.; Qu, Z.; Scarpelli, T.R.; Nesser, H.; Yantosca, R.M.; et al. Global methane budget and trend, 2010–2017: Complementarity of inverse analyses using in situ (GLOBALVIEWplus CH4 ObsPack) and satellite (GOSAT) observations. *Atmos. Chem. Phys.* **2021**, *21*, 4637–4657. [CrossRef]
17. Kuze, A.; Kikuchi, N.; Kataoka, F.; Suto, H.; Shiomi, K.; Kondo, Y. Detection of Methane Emission from a Local Source Using GOSAT Target Observations. *Remote Sens.* **2020**, *12*, 267. [CrossRef]
18. Wang, F.J.; Maksyutov, S.; Tsuruta, A.; Janardanan, R.; Ito, A.; Sasakawa, M.; Machida, T.; Morino, I.; Yoshida, Y.; Kaiser, J.W.; et al. Methane Emission Estimates by the Global High-Resolution Inverse Model Using National Inventories. *Remote Sens.* **2019**, *11*, 2489. [CrossRef]
19. Johnson, M.R.; Tyner, D.R.; Conley, S.; Schwietzke, S.; Zavala-Araiza, D. Comparisons of Airborne Measurements and Inventory Estimates of Methane Emissions in the Alberta Upstream Oil and Gas Sector. *Environ. Sci. Technol.* **2017**, *51*, 13008–13017. [CrossRef] [PubMed]
20. Zavala-Araiza, D.; Alvarez, R.A.; Lyon, D.R.; Allen, D.T.; Marchese, A.J.; Zimmerle, D.J.; Hamburg, S.P. Super-emitters in natural gas infrastructure are caused by abnormal process conditions. *Nat. Commun.* **2017**, *8*, 10. [CrossRef] [PubMed]
21. Zavala-Araiza, D.; Lyon, D.R.; Alvarez, R.A.; Davis, K.J.; Harriss, R.; Herndon, S.C.; Karion, A.; Kort, E.A.; Lamb, B.K.; Lan, X.; et al. Reconciling divergent estimates of oil and gas methane emissions. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15597–15602. [CrossRef]
22. Fox, T.A.; Barchyn, T.E.; Risk, D.; Ravikumar, A.P.; Hugenholtz, C.H. A review of close-range and screening technologies for mitigating fugitive methane emissions in upstream oil and gas. *Environ. Res. Lett.* **2019**, *14*, 18. [CrossRef]
23. Wecht, K.J.; Jacob, D.J.; Frankenberg, C.; Jiang, Z.; Blake, D.R. Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data. *J. Geophys. Res.-Atmos.* **2014**, *119*, 7741–7756. [CrossRef]
24. NASA's Earth Observing System Data and Information System (EOSDIS), Data Processing Levels. Available online: https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels (accessed on 1 November 2021).

25. Kasibhatla, P.S.; Heimann, M.; Rayner, P.J.; Mahowald, N.M.; Prinn, R.G.; Hartley, D.E. *Inverse Methods in Global Biogeochemical Cycles*; American Geophysical Union: Washington, DC, USA, 2000; Volume 114, p. 324.
26. Asch, M.; Bocquet, M.; Nodet, M. *Data Assimilation: Methods, Algorithms, and Applications*; SIAM: Philadelphia, PA, USA, 2016.
27. Gauthier, P.; Charette, C.; Fillion, L.; Koclas, P.; Laroche, S. Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmosphere-Ocean* **1999**, *37*, 103–156. [CrossRef]
28. Gauthier, P.; Tanguay, M.; Laroche, S.; Pellerin, S.; Morneau, J. Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.* **2007**, *135*, 2339–2354. [CrossRef]
29. Errera, Q.; Daerden, F.; Chabrillat, S.; Lambert, J.C.; Lahoz, W.A.; Viscardy, S.; Bonjean, S.; Fonteyn, D. 4D-Var assimilation of MIPAS chemical observations: Ozone and nitrogen dioxide analyses. *Atmos. Chem. Phys.* **2008**, *8*, 6169–6187. [CrossRef]
30. Errera, Q.; Menard, R. Technical Note: Spectral representation of spatial correlations in variational assimilation with grid point models and application to the Belgian Assimilation System for Chemical Observations (BASCOE). *Atmos. Chem. Phys.* **2012**, *12*, 10015–10031. [CrossRef]
31. 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]
32. Houtekamer, P.L.; Zhang, F.Q. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Weather Rev.* **2016**, *144*, 4489–4532. [CrossRef]
33. Skachko, S.; Errera, Q.; Menard, R.; Christophe, Y.; Chabrillat, S. Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model. *Geosci. Model Dev.* **2014**, *7*, 1451–1465. [CrossRef]
34. Skachko, S.; Menard, R.; Errera, Q.; Christophe, Y.; Chabrillat, S. EnKF and 4D-Var data assimilation with chemical transport model BASCOE (version 05.06). *Geosci. Model Dev.* **2016**, *9*, 2893–2908. [CrossRef]
35. Menard, R.; Cohn, S.E.; Chang, L.P.; Lyster, P.M. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation. *Mon. Weather Rev.* **2000**, *128*, 2654–2671. [CrossRef]
36. Menard, R.; Changs, L.P. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part II: Chi(2)-validated results and analysis of variance and correlation dynamics. *Mon. Weather Rev.* **2000**, *128*, 2672–2686. [CrossRef]
37. Khattatov, B.V.; Lamarque, J.F.; Lyjak, L.V.; Menard, R.; Levelt, P.; Tie, X.X.; Brasseur, G.P.; Gille, J.C. Assimilation of satellite observations of long-lived chemical species in global chemistry transport models. *J. Geophys. Res.-Atmos.* **2000**, *105*, 29135–29144. [CrossRef]
38. Eskes, H.J.; Van Velthoven, P.F.J.; Valks, P.J.M.; Kelder, H.M. Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 1663–1681. [CrossRef]
39. Rosevall, J.D.; Murtagh, D.P.; Urban, J.; Jones, A.K. A study of polar ozone depletion based on sequential assimilation of satellite data from the ENVISAT/MIPAS and Odin/SMR instruments. *Atmos. Chem. Phys.* **2007**, *7*, 899–911. [CrossRef]
40. Van der A, R.J.; Allaart, M.A.F.; Eskes, H.J. Multi sensor reanalysis of total ozone. *Atmos. Chem. Phys.* **2010**, *10*, 11277–11294. [CrossRef]
41. Pannekoucke, O.; Ricci, S.; Barthelemy, S.; Menard, R.; Thual, O. Parametric Kalman filter for chemical transport models. *Tellus Ser. A-Dyn. Meteorol. Oceanogr.* **2016**, *68*, 31547; Erratum in *Tellus Ser. A-Dyn. Meteorol. Oceanogr.* **2018**, *70*, 1472954. [CrossRef]
42. Pannekoucke, O.; Menard, R.; El Aabaribaoune, M.; Plu, M. A methodology to obtain model-error covariances due to the discretization scheme from the parametric Kalman filter perspective. *Nonlinear Processes Geophys.* **2021**, *28*, 1–22. [CrossRef]
43. Prinn, R.G.; Huang, J.; Weiss, R.F.; Cunnold, D.M.; Fraser, P.J.; Simmonds, P.G.; McCulloch, A.; Harth, C.; Reimann, S.; Salameh, P.; et al. Evidence for variability of atmospheric hydroxyl radicals over the past quarter century. *Geophys. Res. Lett.* **2005**, *32*, 4. [CrossRef]
44. Prather, M.J.; Holmes, C.D.; Hsu, J. Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry. *Geophys. Res. Lett.* **2012**, *39*, 5. [CrossRef]
45. Gaspari, G.; Cohn, S.E. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **1999**, *125*, 723–757. [CrossRef]
46. Byun, D.; Schere, K.L. Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* **2006**, *59*, 51–77. [CrossRef]
47. Mathur, R.; Xing, J.; Gilliam, R.; Sarwar, G.; Hogrefe, C.; Pleim, J.; Pouliot, G.; Roselle, S.; Spero, T.L.; Wong, D.C.; et al. Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: Overview of process considerations and initial applications. *Atmos. Chem. Phys.* **2017**, *17*, 12449–12474. [CrossRef] [PubMed]
48. CMAQ Tutorials. Available online: https://www.epa.gov/cmaq/cmaq-documentation (accessed on 14 October 2021).
49. Crippa, M.; Solazzo, E.; Huang, G.L.; Guizzardi, D.; Koffi, E.; Muntean, M.; Schieberle, C.; Friedrich, R.; Janssens-Maenhout, G. High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. *Sci. Data* **2020**, *7*, 121. [CrossRef]
50. Crippa, M.; Guizzardi, D.; Muntean, M.; Schaaf, E.; Lo Vullo, E.; Solazzo, E.; Monforti-Ferrario, F.; Olivier, J.; Vignati, E. EDGAR v6.0 Greenhouse Gas Emissions. Available online: https://data.jrc.ec.europa.eu/dataset/97a67d67-c62e-4826-b873-9d972c4f670b (accessed on 5 November 2021).
51. UNC. Community Modeling and Analysis System CMAS [WWW Document]. SMOKE v4.5 User's Man. Available online: https://www.cmascenter.org/smoke/ (accessed on 2 November 2021).

52. Cressot, C.; Chevallier, F.; Bousquet, P.; Crevoisier, C.; Dlugokencky, E.J.; Fortems-Cheiney, A.; Frankenberg, C.; Parker, R.; Pison, I.; Scheepmaker, R.A.; et al. On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements. *Atmos. Chem. Phys.* **2014**, *14*, 577–592. [CrossRef]
53. Bloom, A.A.; Bowman, K.W.; Lee, M.; Turner, A.J.; Schroeder, R.; Worden, J.R.; Weidner, R.; McDonald, K.C.; Jacob, D.J. A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models (WetCHARTs version 1.0). *Geosci. Model Dev.* **2017**, *10*, 2141–2156. [CrossRef]
54. 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] [PubMed]
55. Butz, A.; Guerlet, S.; Hasekamp, O.; Schepers, D.; Galli, A.; Aben, I.; Frankenberg, C.; Hartmann, J.M.; Tran, H.; Kuze, A.; et al. Toward accurate CO2 and CH4 observations from GOSAT. *Geophys. Res. Lett.* **2011**, *38*, 1–6. [CrossRef]
56. Schepers, D.; Guerlet, S.; Butz, A.; Landgraf, J.; Frankenberg, C.; Hasekamp, O.; Blavier, J.F.; Deutscher, N.M.; Griffith, D.W.T.; Hase, F.; et al. Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms. *J. Geophys. Res.-Atmos.* **2012**, *117*, 17549. [CrossRef]
57. Buchwitz, M.; Reuter, M.; Schneising, O.; Hewson, W.; Detmers, R.G.; Boesch, H.; Hasekamp, O.P.; Aben, I.; Bovensmann, H.; Burrows, J.P.; et al. Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO2 and XCH4 CRDP3 data set. *Remote Sens. Environ.* **2017**, *203*, 276–295. [CrossRef]
58. Turner, A.J.; Jacob, D.J.; Wecht, K.J.; Maasakkers, J.D.; Lundgren, E.; Andrews, A.E.; Biraud, S.C.; Boesch, H.; Bowman, K.W.; Deutscher, N.M.; et al. Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data. *Atmos. Chem. Phys.* **2015**, *15*, 7049–7069. [CrossRef]
59. Parker, R.J.; Boesch, H.; Byckling, K.; Webb, A.J.; Palmer, P.I.; Feng, L.; Bergamaschi, P.; Chevallier, F.; Notholt, J.; Deutscher, N.; et al. Assessing 5 years of GOSAT Proxy XCH4 data and associated uncertainties. *Atmos. Meas. Tech.* **2015**, *8*, 4785–4801. [CrossRef]
60. Bovensmann, H.; Burrows, J.; Buchwitz, M.; Frerick, J.; Noël, S.; Rozanov, V.; Chance, K.; Goede, A. SCIAMACHY: Mission objectives and measurement modes. *J. Atmos. Sci.* **1999**, *56*, 127–150. [CrossRef]
61. Bergamaschi, P.; Frankenberg, C.; Meirink, J.F.; Krol, M.; Dentener, F.; Wagner, T.; Platt, U.; Kaplan, J.O.; Korner, S.; Heimann, M.; et al. Satellite chartography of atmospheric methane from SCIAMACHYon board ENVISAT: 2. Evaluation based on inverse model simulations. *J. Geophys. Res.-Atmos.* **2007**, *112*, 07268. [CrossRef]
62. Cohn, S.E. Dynamics of short-term univariate forecast error covariances. *Mon. Weather Rev.* **1993**, *121*, 3123–3149. [CrossRef]
63. Menard, R.; Gauthier, P.; Rochon, Y.; Robichaud, A.; de Grandpre, J.; Yang, Y.; Charrette, C.; Chabrillat, S. Coupled Stratospheric Chemistry-Meteorology Data Assimilation. Part II: Weak and Strong Coupling. *Atmosphere* **2019**, *10*, 798. [CrossRef]
64. Seinfeld, J.I. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
65. Jacobson, M.Z.C.U.P. *Fundamentals of Atmospheric Modeling*; Cambridge University Press: Cambridge, UK, 2020.
66. Menard, R.; Skachko, S.; Pannekoucke, O. Numerical discretization causing error variance loss and the need for inflation. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 3498–3520. [CrossRef]
67. Courtier, P.; Thepaut, J.N.; Hollingsworth, A. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Q. J. R. Meteorol. Soc.* **1994**, *120*, 1367–1387. [CrossRef]
68. Kirschke, S.; Bousquet, P.; Ciais, P.; Saunois, M.; Canadell, J.G.; Dlugokencky, E.J.; Bergamaschi, P.; Bergmann, D.; Blake, D.R.; Bruhwiler, L.; et al. Three decades of global methane sources and sinks. *Nat. Geosci.* **2013**, *6*, 813–823. [CrossRef]
69. Olsen, E.; Fetzer, E.; Hulley, G.; Manning, E.; Blaisdell, J.; Iredell, L.; Susskind, J.; Warner, J.; Wei, Z.; Blackwell, W. *AIRS/AMSU/HSB Version 6 Level 2 Product User Guide*; JPL Technical Report; NASA: Washington, DC, USA, 2013.
70. Xiong, X.Z.; Barnet, C.; Maddy, E.; Sweeney, C.; Liu, X.P.; Zhou, L.H.; Goldberg, M. Characterization and validation of methane products from the Atmospheric Infrared Sounder (AIRS). *J. Geophys. Res.-Biogeosci.* **2008**, *113*, 500. [CrossRef]
71. Dlugokencky. NOAA/GML. 2020; pp. 17021–17043. Available online: www.esrl.noaa.gov/gmd/ccgg/trends_ch4/ (accessed on 20 November 2021).
72. Schuldt, K.N.; Aalto, T.; Andrews, A.; Aoki, S.; Arduini, J.; Baier, B.; Bergamaschi, P.; Biermann, T.; Biraud, S.C.; Boenisch, H.; et al. *Multi-Laboratory Compilation of Atmospheric Methane Data for the Period 1983–2020; obspack_ch4_1_GLOBALVIEWplus_v3.0_2021–05–07*; NOAA Earth System Research Laboratory, Global Monitoring Laboratory: Boulder, CO, USA, 2021. [CrossRef]
73. Buchwitz, M.; Reuter, M.; Schneising, O.; Boesch, H.; Guerlet, S.; Dils, B.; Aben, I.; Armante, R.; Bergamaschi, P.; Blumenstock, T.; et al. The Greenhouse Gas Climate Change Initiative (GHG-CCI): Comparison and quality assessment of near-surface-sensitive satellite-derived CO2 and CH4 global data sets. *Remote Sens. Environ.* **2015**, *162*, 344–362. [CrossRef]
74. Oshio, H.; Yoshida, Y.; Matsunaga, T.; Deutscher, N.M.; Dubey, M.; Griffith, D.W.T.; Hase, F.; Iraci, L.T.; Kivi, R.; Liu, C.; et al. Bias Correction of the Ratio of Total Column CH_4 to CO_2 Retrieved from GOSAT Spectra. *Remote Sens.* **2020**, *12*, 3155. [CrossRef]
75. Inoue, M.; Morino, I.; Uchino, O.; Nakatsuru, T.; Yoshida, Y.; Yokota, T.; Wunch, D.; Wennberg, P.O.; Roehl, C.M.; Griffith, D.W.T.; et al. Bias corrections of GOSAT SWIR XCO_2 and XCH_4 with TCCON data and their evaluation using aircraft measurement data. *Atmos. Meas. Tech.* **2016**, *9*, 3491–3512. [CrossRef]
76. Zhou, M.Q.; Dils, B.; Wang, P.C.; Detmers, R.; Yoshida, Y.; O'Dell, C.W.; Feist, D.G.; Velazco, V.A.; Schneider, M.; De Maziere, M. Validation of TANSO-FTS/GOSAT XCO_2 and XCH_4 glint mode retrievals using TCCON data from near-ocean sites. *Atmos. Meas. Tech.* **2016**, *9*, 1415–1430. [CrossRef]

77. Wunch, D.; Toon, G.C.; Blavier, J.F.L.; Washenfelder, R.A.; Notholt, J.; Connor, B.J.; Griffith, D.W.T.; Sherlock, V.; Wennberg, P.O. The Total Carbon Column Observing Network. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **2011**, *369*, 2087–2112. [CrossRef]
78. Maasakkers, J.D.; Jacob, D.J.; Sulprizio, M.P.; Scarpelli, T.R.; Nesser, H.; Sheng, J.X.; Zhang, Y.Z.; Hersher, M.; Bloom, A.A.; Bowman, K.W.; et al. Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015. *Atmos. Chem. Phys.* **2019**, *19*, 7859–7881. [CrossRef]
79. Alexe, M.; Bergamaschi, P.; Segers, A.; Detmers, R.; Butz, A.; Hasekamp, O.; Guerlet, S.; Parker, R.; Boesch, H.; Frankenberg, C.; et al. Inverse modelling of CH4 emissions for 2010–2011 using different satellite retrieval products from GOSAT and SCIAMACHY. *Atmos. Chem. Phys.* **2015**, *15*, 113–133. [CrossRef]
80. Saad, K.M.; Wunch, D.; Deutscher, N.M.; Griffith, D.W.T.; Hase, F.; De Maziere, M.; Notholt, J.; Pollard, D.F.; Roehl, C.M.; Schneider, M.; et al. Seasonal variability of stratospheric methane: Implications for constraining tropospheric methane budgets using total column observations. *Atmos. Chem. Phys.* **2016**, *16*, 14003–14024. [CrossRef]
81. Patra, P.K.; Houweling, S.; Krol, M.; Bousquet, P.; Belikov, D.; Bergmann, D.; Bian, H.; Cameron-Smith, P.; Chipperfield, M.P.; Corbin, K.; et al. TransCom model simulations of CH4 and related species: Linking transport, surface flux and chemical loss with CH4 variability in the troposphere and lower stratosphere. *Atmos. Chem. Phys.* **2011**, *11*, 12813–12837. [CrossRef]
82. Bader, W.; Bovy, B.; Conway, S.; Strong, K.; Smale, D.; Turner, A.J.; Blumenstock, T.; Boone, C.; Coen, M.C.; Coulon, A.; et al. The recent increase of atmospheric methane from 10 years of ground-based NDACC FTIR observations since 2005. *Atmos. Chem. Phys.* **2017**, *17*, 2255–2277. [CrossRef]
83. Zhang, Y.Z.; Jacob, D.J.; Lu, X.; Maasakkers, J.D.; Scarpelli, T.R.; Sheng, J.X.; Shen, L.; Qu, Z.; Sulprizio, M.P.; Chang, J.F.; et al. Attribution of the accelerating increase in atmospheric methane during 2010–2018 by inverse analysis of GOSAT observations. *Atmos. Chem. Phys.* **2021**, *21*, 3643–3666. [CrossRef]
84. Seber, G.A.; Lee, A.J. *Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012; Volume 329.
85. Menard, R.; Deshaies-Jacques, M.; Gasset, N. A comparison of correlation-length estimation methods for the objective analysis of surface pollutants at Environment and Climate Change Canada. *J. Air Waste Manag. Assoc.* **2016**, *66*, 874–895. [CrossRef]
86. Jazwinski, A.H. *Stochastic Processes and Filtering Theory*; Academic Press: New York, NY, USA, 1970.
87. Janjic, T.; Bormann, N.; Bocquet, M.; Carton, J.A.; Cohn, S.E.; Dance, S.L.; Losa, S.N.; Nichols, N.K.; Potthast, R.; Waller, J.A.; et al. On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 1257–1278. [CrossRef]
88. Cohn, S.E. An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **1997**, *75*, 257–288. [CrossRef]
89. Szenasi, B.; Berchet, A.; Broquet, G.; Segers, A.; van der Gon, H.D.; Krol, M.; Hullegie, J.J.S.; Kiesow, A.; Gunther, D.; Petrescu, A.M.R.; et al. A pragmatic protocol for characterising errors in atmospheric inversions of methane emissions over Europe. *Tellus Ser. B-Chem. Phys. Meteorol.* **2021**, *73*, 1–23. [CrossRef]
90. Berchet, A.; Pison, I.; Chevallier, F.; Bousquet, P.; Conil, S.; Geever, M.; Laurila, T.; Lavric, J.; Lopez, M.; Moncrieff, J.; et al. Towards better error statistics for atmospheric inversions of methane surface fluxes. *Atmos. Chem. Phys.* **2013**, *13*, 7115–7132. [CrossRef]
91. Heald, C.L.; Jacob, D.J.; Jones, D.B.A.; Palmer, P.I.; Logan, J.A.; Streets, D.G.; Sachse, G.W.; Gille, J.C.; Hoffman, R.N.; Nehrkorn, T. Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide. *J. Geophys. Res.-Atmos.* **2004**, *109*, 5185. [CrossRef]
92. Segers, A.J.; Eskes, H.J.; Van der A, R.J.; Van Oss, F.; Van Velthoven, P.F.J. Assimilation of GOME ozone profiles and a global chemistry-transport model using a Kalman filter with anisotropic covariance. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 477–502. [CrossRef]
93. Tremolet, Y. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 3127. [CrossRef]
94. Stanevich, I.; Jones, D.; Strong, K.; Parker, R.J.; Boesch, H.; Wunch, D.; Notholt, J.; Petri, C.; Warneke, T.; Sussmann, R. Characterizing model errors in chemical transport modeling of methane: Impact of model resolution in versions v9–02 of GEOS-Chem and v35j of its adjoint model. *Geosci. Model Dev.* **2020**, *13*, 3839–3862. [CrossRef]
95. Lahoz, W.A.; Schneider, P. Data assimilation: Making sense of Earth Observation. *Front. Environ. Sci.* **2014**, *2*, 16. [CrossRef]
96. Buehner, M.; Houtekamer, P.L.; Charette, C.; Mitchell, H.L.; He, B. Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Weather Rev.* **2010**, *138*, 1550–1566. [CrossRef]
97. Gilpin, S.; Matsuo, T.; Cohn, S.E. Continuum covariance propagation for understanding variance loss in advective systems. *arXiv* **2021**, arXiv:2109.02530.
98. CMAQv5.3 User's Guide. Available online: https://github.com/USEPA/CMAQ/blob/5.3/DOCS/Users_Guide/README.md (accessed on 14 October 2021).
99. Menard, R.; Deshaies-Jacques, M. Evaluation of Analysis by Cross-Validation, Part II: Diagnostic and Optimization of Analysis Error Covariance. *Atmosphere* **2018**, *9*, 70. [CrossRef]
100. Krishnamoorthy, A.; Menon, D. Matrix inversion using Cholesky decomposition. In Proceedings of the 2013 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 26–28 September 2013.
101. Hakami, A.; Henze, D.K.; Seinfeld, J.H.; Singh, K.; Sandu, A.; Kim, S.T.; Byun, D.W.; Li, Q.B. The adjoint of CMAQ. *Environ. Sci. Technol.* **2007**, *41*, 7807–7817. [CrossRef] [PubMed]
102. Zhao, S.L.; Russell, M.G.; Hakami, A.; Capps, S.L.; Turner, M.D.; Henze, D.K.; Percell, P.B.; Resler, J.; Shen, H.; Russell, A.G.; et al. A multiphase CMAQ version 5.0 adjoint. *Geosci. Model Dev.* **2020**, *13*, 2925–2944. [CrossRef] [PubMed]

103. Peng, Z.; Zhang, M.; Kou, X.; Tian, X.; Ma, X. A regional carbon data assimilation system and its preliminary evaluation in East Asia. *Atmos. Chem. Phys.* **2015**, *15*, 1087–1104. [CrossRef]
104. Otte, T.L.; Pleim, J.E. The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: Updates through MCIPv3.4.1. *Geosci. Model Dev.* **2010**, *3*, 243–256. [CrossRef]