



HAL
open science

Probabilistic global maps of the CO₂ column at daily and monthly scales from sparse satellite measurements

Frederic Chevallier, Grégoire Broquet, Clémence Pierangelo, David Crisp

► To cite this version:

Frederic Chevallier, Grégoire Broquet, Clémence Pierangelo, David Crisp. Probabilistic global maps of the CO₂ column at daily and monthly scales from sparse satellite measurements. *Journal of Geophysical Research: Atmospheres*, 2017, 122 (14), pp.7614-7629. 10.1002/2017JD026453. hal-02957316

HAL Id: hal-02957316

<https://hal.science/hal-02957316>

Submitted on 8 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

10.1002/2017JD026453

Key Points:

- A satellite level 3 XCO₂ product at daily and 2° × 2° resolution has been generated with a Kalman filter for OCO-2
- The filter provides some preliminary information about the quality of the XCO₂ retrievals both in terms of random and systematic errors
- The error statistics of the filter are finer than the statistics of a climatological distribution of XCO₂ and are fairly predicted by the filter

Correspondence to:

F. Chevallier,
frederic.chevallier@lscce.ipsl.fr

Citation:

Chevallier, F., G. Broquet, C. Pierangelo, and D. Crisp (2017), Probabilistic global maps of the CO₂ column at daily and monthly scales from sparse satellite measurements, *J. Geophys. Res. Atmos.*, 122, 7614–7629, doi:10.1002/2017JD026453.



Received 4 JAN 2017

Accepted 22 JUN 2017

Accepted article online 24 JUN 2017

Published online 17 JUL 2017

Probabilistic global maps of the CO₂ column at daily and monthly scales from sparse satellite measurements

Frédéric Chevallier¹ , Grégoire Broquet¹, Clémence Pierangelo², and David Crisp³ 

¹Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, L'Orme des Merisiers, Gif-sur-Yvette, France, ²Centre National d'Études Spatiales, Toulouse, France, ³Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Abstract The column-average dry air-mole fraction of carbon dioxide in the atmosphere (XCO₂) is measured by scattered satellite measurements like those from the Orbiting Carbon Observatory (OCO-2). We show that global continuous maps of XCO₂ (corresponding to level 3 of the satellite data) at daily or coarser temporal resolution can be inferred from these data with a Kalman filter built on a model of persistence. Our application of this approach on 2 years of OCO-2 retrievals indicates that the filter provides better information than a climatology of XCO₂ at both daily and monthly scales. Provided that the assigned observation uncertainty statistics are tuned in each grid cell of the XCO₂ maps from an objective method (based on consistency diagnostics), the errors predicted by the filter at daily and monthly scales represent the true error statistics reasonably well, except for a bias in the high latitudes of the winter hemisphere and a lack of resolution (i.e., a too small discrimination skill) of the predicted error standard deviations. Due to the sparse satellite sampling, the broad-scale patterns of XCO₂ described by the filter seem to lag behind the real signals by a few weeks. Finally, the filter offers interesting insights into the quality of the retrievals, both in terms of random and systematic errors.

1. Introduction

Understanding the variability of the column-average dry air-mole fraction of carbon dioxide (XCO₂) in time and space is prerequisite to interpreting the scattered XCO₂ measurements made from polar-orbiting satellites [e.g., Alkhaled *et al.*, 2008]. The infrequent satellite visit of each location of the Earth is not correlated with XCO₂, which may justify simple gap-filling approaches, like a linear interpolation between soundings. However, other causes are likely correlated with XCO₂ values and encourage more sophisticated processing: (i) passive solar spectroscopy requires sunlight, which excludes nighttime observations and observations at high latitudes of the winter hemisphere; (ii) optically thick clouds or aerosols preclude XCO₂ retrievals; and (iii) all retrievals do not pass quality control procedures.

Since the Greenhouse Gases Observing SATellite (GOSAT) was launched in January 2009, various attempts have been made to fill the gaps between XCO₂ measurements without the help of a chemistry-transport model, either in order to compare these CO₂ estimates with distant (up to a few days and a few latitude-longitude degrees) reference ground-based measurements (for the purpose of validating the retrievals) or to infer the global distribution of XCO₂ (mostly for the purpose of visualizing the information brought by the retrievals in a convenient way). Some have neglected the variability of XCO₂ in a given space time domain [e.g., Wunch *et al.*, 2011b; Cogan *et al.*, 2012; Zhou *et al.*, 2016], while some others have modeled the space-time variability of XCO₂ with estimated parameters from either a geostatistical [e.g., Hammerling *et al.*, 2012; Nguyen *et al.*, 2014b; Zeng *et al.*, 2017] or a Bayesian approach [e.g., Katzfuss and Cressie, 2012; Nguyen *et al.*, 2014a]. All of these methods include some dimension reduction in order to facilitate the computation. They have so far been disconnected from other applications of the retrievals for data assimilation or atmospheric inversion.

The growing availability of realistic high-resolution simulations of XCO₂ at the global scale, like those made within the operational CO₂ forecasts of the Copernicus Atmosphere Monitoring Service (CAMS, <http://atmosphere.copernicus.eu/>) [Agustí-Panareda *et al.*, 2016a], provides new insight into the variability of XCO₂ and into the covariations between observed and unobserved XCO₂. These models may help address the

gap-filling problem differently and perhaps in a simpler way. Here we use such information in a Bayesian Kalman filter (KF) that assimilates X_{CO_2} retrievals in a model of persistence, i.e., a “model” where X_{CO_2} does not vary. Persistence has been chosen because it is more neutral (i.e., less informative) than, e.g., a 3-D transport model, while we want to highlight the information from the retrievals. We also define some additional hypotheses for our problem, for instance, to better constrain unobserved regions of the globe or to enforce the statistical optimality of the filter. Retrieval after retrieval, the filter updates a map of estimated X_{CO_2} , together with an associated error covariance matrix. Brunner *et al.* [2006] applied this principle to ozone retrievals before, but with simple parametric error statistics, while we use the Bayesian posterior error statistics of the retrievals and statistics of the CAMS CO_2 forecast day-to-day variability at the pixel scale to describe the uncertainty of the persistence model. We therefore focus our effort on the probabilistic aspect of the KF. We evaluate the realism of this approach when applied to retrievals from the second Orbiting Carbon Observatory (OCO-2), which was launched on 2 July 2014 [Eldering *et al.*, 2017]. We test our X_{CO_2} KF like any probabilistic prediction system, and we use some classical diagnostics [Talagrand *et al.*, 1999; Desroziers *et al.*, 2005; Talagrand, 2014] to statistically evaluate its skill. Our strategy is twofold. First, we verify the consistency of its internal statistics with its input hypotheses over a 2 year period (September 2014–August 2016). Second, we compare the KF daily-mean X_{CO_2} maps over a 16 month period (September 2014–December 2015) to independent observations from the Total Carbon Column Observing Network (TCCON) [Wunch *et al.*, 2011a] and with an atmospheric inversion that assimilated surface air sample measurements. To define a baseline skill, we also compare the performance of the KF with that of a climatology of X_{CO_2} .

The paper is organized as follows. Section 2 presents the observation and model data sets used in the filter and for its evaluation. The Kalman filter is described in section 3. Results are shown in section 4. Section 5 concludes the paper.

2. Retrievals and Model Simulation

2.1. ACOS-OCO-2 Retrievals

OCO-2 orbits around the Earth from pole to pole with a 16 day repeat cycle and a local crossing time at the equator around 1:30 P.M. It carries a 3-band spectrometer that measures the sunlight reflected by the Earth and its atmosphere in the near-infrared/shortwave infrared spectral regions with a narrow swath and with footprints of a few km^2 . Scientific data have been acquired since 5 September 2014 when the observatory is over the sunlit hemisphere and the instrument is oriented directly downward at the local nadir (nadir observations), or near the Sun-glint spot (glint observation), or toward specific surface targets. The satellite usually collects nadir or glint observations for complete orbits, but every 1 or 2 days, these observations are interrupted by target observations or transitions to or from the target observations. An updated version of NASA's Atmospheric CO_2 Observations from Space (ACOS) algorithm developed for retrieving X_{CO_2} from GOSAT measurements is used to analyze the radiance measurements in sufficiently cloud- and aerosol-free conditions, as described by O'Dell *et al.* [2012]. We will ignore retrievals from the target mode here and will keep all retrievals in nadir and glint modes that pass the quality control (variable `xco2_quality_flag` set to 0). In particular, very few X_{CO_2} retrievals are performed over the ocean in nadir mode because of low surface reflectivity, and our KF has to fill those gaps among others. The ACOS-OCO-2 X_{CO_2} retrievals are provided with a recommended bias correction that follows the parametric approach initiated by Wunch *et al.* [2011b] and O'Dell *et al.* [2012]. We use version v7r of the bias-corrected retrievals, but we ignore the associated averaging kernels and a priori profiles since our KF directly works in X_{CO_2} space. Further details of the OCO-2 mission are given in Eldering *et al.* [2017], in particular about the changing observing strategy over time that changes the gaps to be filled by the KF.

2.2. TCCON Retrievals

The TCCON provides the most accurate regular measurements of X_{CO_2} at a series of about 20 surface sites around the world. The technique is broadly similar to OCO-2, with the notable exception that the TCCON spectrometers directly view incoming solar radiation from the solar disk, rather than sunlight scattered from the surface and atmosphere. The TCCON X_{CO_2} retrievals have been calibrated to the WMO scale through an expanding series of aircraft measurements [Wunch *et al.*, 2011a]. For the present study, data were extracted

from the TCCON GGG2014 database on 5 December 2016. In 2015, 22 stations reported measurements in this version. We use data from 21 of these stations that are, from south to north, Lauder [Sherlock *et al.*, 2014], Wollongong [Griffith *et al.*, 2014a], Réunion Island [de Maziere *et al.*, 2014], Darwin [Griffith *et al.*, 2014b], Ascension Island [Feist *et al.*, 2014], Manaus [Dubey *et al.*, 2014], Saga [Kawakami *et al.*, 2014], Edwards [Iraci *et al.*, 2014], Tsukuba [Morino *et al.*, 2014a], Anmeyondo [Goo *et al.*, 2014], Lamont [Wennberg *et al.*, 2014a], Rikubetsu [Morino *et al.*, 2014b], Park Falls [Wennberg *et al.*, 2014b], Garmisch [Sussman and Rettinger, 2014], Orleans [Warneke *et al.*, 2014], Paris [Té *et al.*, 2014], Karlsruhe [Hase *et al.*, 2014], Bremen [Notholt *et al.*, 2014], Bialystok [Deutscher *et al.*, 2014], Sodankylä [Kivi *et al.*, 2014], and Eureka [Strong *et al.*, 2014]. Measurements from Pasadena have been excluded because they are too often contaminated (by several ppm) by local fossil fuel emissions from the Los Angeles basin. Each TCCON retrieval is provided with an averaging kernel and a prior profile, which are not used here. The impact of this simplification was estimated to have a standard deviation of $0.24 \mu\text{mol mol}^{-1}$ (ppm) by Nguyen *et al.* [2014b]. At each site and for each day, we select the TCCON observations that are closest from the usual OCO-2 daytime local crossing time at the equator (irrespective of the existence of actual OCO-2 soundings around that time and location) but exclude them if they are more than 2 h away from this crossing time. In total, the TCCON-based statistics that will be presented below include 3911 TCCON retrievals.

2.3. CAMS CO₂ Forecast

The CO₂ global forecasts of the CAMS operational service have been issued by the European Centre for Medium-Range Weather Forecasts (ECMWF) during our study period at a resolution of approximately 16 km in the horizontal and 137 vertical levels, without assimilating observations of atmospheric composition. The surface net ecosystem exchange of CO₂ is modeled online by the CTESSEL carbon model [Boussetta *et al.*, 2013] corrected from large-scale biases [Agustí-Panareda *et al.*, 2016b]. Compared to TCCON, these forecasts exhibit annual regional biases of about 1 ppm and remarkably reproduce the synoptic variability of XCO₂ on continental scales [Agustí-Panareda *et al.*, 2016a, 2016b]. We do not use the CAMS forecasts at their native 16 km resolution but upscale them on a $2^\circ \times 2^\circ$ longitude-latitude grid (section 3.2). However, the skill of the forecast at high resolution gives us confidence in the realism of the XCO₂ forecast at coarser resolution.

2.4. CAMS CO₂ Inversion

We use a second greenhouse gas product of the CAMS service: version 15r4 of the CO₂ inversion described by Chevallier *et al.* [2010], with updates from Chevallier [2016a]. The inversion method formulates optimal estimation in a variational framework. Version 15r4 covers all years from 1979 to 2015, at resolution $3.75^\circ \times 1.9^\circ$ (longitude-latitude) and 3-hourly, based on 133 CO₂ dry air mole fraction surface time series from four databases: the NOAA Earth System Research Laboratory archive (<http://www.esrl.noaa.gov/gmd/ccgg/index.html>), the World Data Centre for Greenhouse Gases archive (<http://ds.data.jma.go.jp/gmd/wdcgg/>), the Réseau Atmosphérique de Mesure des Composés à Effet de Serre database (<http://www.lscce.ipsl.fr/>), and the Integrated Carbon Observation System-Atmospheric Thematic Center (<https://icos-atc.lscce.ipsl.fr/>). The list of sites is given in Tables 1 and 2 of Chevallier [2016b].

The global inversion product includes an associated four-dimensional description of CO₂ at grid points and 39 vertical-layer resolution, which can be used to compute XCO₂, either by simply weighting the profile by the pressure width of the model layers or through a retrieval averaging kernel. Individual differences between the simulated XCO₂ from a previous version of the CAMS inversion product (v13r1) and TCCON (using the corresponding averaging kernels and prior profiles) or aircraft measurements were shown to be mostly within ± 1 ppm [Frankenberg *et al.*, 2016; Kulawik *et al.*, 2016]. In the present study, we use XCO₂ obtained from weighting the profile by the pressure width.

2.5. Climatology of XCO₂

We use the TCCON GGG2009 a priori model for CO₂ to provide a baseline skill for the estimation of XCO₂ (C. O'Dell, personal communication, 2015). This 1-D model only takes time (with a linear increase of 0.5%/yr since 1 January 2005), latitude, tropopause pressure, and boundary layer pressure as input and was designed based on surface, aircraft, and balloon air sample measurements acquired well before the OCO-2 period [Wunch *et al.*, 2011b]. In our application of this simple model, we use the ECMWF reanalysis data to get the tropopause and boundary layer pressures.

3. Kalman Filter

3.1. Generalities

In the following, we follow the convention of *Rayner et al.* [2016], who adapted previous ones for biogeochemistry estimation problems. Matrices are represented with uppercase letters in straight bold type (e.g., \mathbf{Q}). Vectors are represented with lowercase letters in straight bold type (e.g., \mathbf{x}). Scalars are in lower case italics (e.g., y). Time indices are in subscript. Superscript “ b ” denotes the input (*background*) of the analysis step of the KF. Superscript “ a ” denotes the output (*analysis*) of the analysis step of the KF. Superscript “ t ” denotes the (unknown) truth.

A standard KF estimates the time evolution of a series of prognostic variables. For convenience, we gather these target variables in a vector. We call \mathbf{x}_t its value at time index t and \mathbf{U}_t the uncertainty covariance matrix of a given estimate of \mathbf{x}_t . The prognostic nature of \mathbf{x}_t means that a prognostic model computes an estimate of \mathbf{x}_t based on \mathbf{x}_{t-1} . In the standard KF, the prognostic model is assumed to be linear: we call it \mathbf{M}_t (note that we will restrict ourselves to a time-independent operator, see section 3.2).

The time axis is split into discrete time steps. The KF works recursively from one time step $t-1$ to another t . At each step t , the KF estimates the best linear unbiased estimate (BLUE) of \mathbf{x}_t by merging its prediction of \mathbf{x}_t with a series of connected observations made close to t , jointly called \mathbf{y}_t . The prediction of \mathbf{x}_t is obtained by applying operator \mathbf{M}_t to \mathbf{x}_{t-1} . Each information piece (coming from the prediction of \mathbf{x}_t or from \mathbf{y}_t) is weighted by its uncertainty following Bayes’ theorem and assuming that it is Gaussian, unbiased, and that observation errors and prediction errors are uncorrelated with each other. The observations, \mathbf{y}_t , may not directly correspond to \mathbf{x}_t . In that case, a linear operator \mathbf{H}_t is used to link the two types of variables.

3.2. Implementation

We define the components of our KF as follows:

1. The target vector \mathbf{x}_t is made of the pixels of the 2-D global map of XCO₂ at a given time step t . The grid of the map, that we call G , is a regular $2^\circ \times 2^\circ$ longitude-latitude grid in our application: \mathbf{x}_t therefore includes $n = 16,200$ elements (grid cells). The OCO-2 spacecraft takes about 30 s to cross one grid box.
2. The model operator \mathbf{M}_t is the identity matrix.
3. The error covariance matrix (\mathbf{Q}_t) is defined as the integration of the error of \mathbf{M}_t over 3 h. It is made of monthly statistics from a model. In our application, they correspond to the day-to-day variability of the daily mean XCO₂ in the high-resolution CAMS CO₂ forecasts aggregated on G for each month of year 2015. We take the daily mean because the midday sampling time of OCO-2 approximates the 24 h average of XCO₂ in clear-sky conditions [*Olsen and Randerson, 2004*]. This day-to-day variability is reduced to 3 h by dividing the variances and the covariances by 8. The choice of inserting the variance every 3 h is a compromise between doing it after each assimilation cycle, one retrieval at a time (with less computational efficiency) and doing it once a day (with larger discontinuities). Our correlation matrix is empirically based on a relatively small statistical ensemble: its size at each grid point is the number of days in the month. The correlation matrix therefore likely includes spurious long-distance correlations. A localization technique is used to damp these contributions. It consists of an element-wise product (Hadamard product) of the raw covariance matrices with a correlation matrix \mathbf{C} . The elements in \mathbf{C} are calculated based on the distance over the globe between the elements of \mathbf{x}_t : we use an exponential function that decreases by e after 20,000 km. The choice of a long distance avoids changing the short-scale correlations of the original statistics much.
4. The observation vector \mathbf{y}_t is a single *super-observation* of XCO₂ made around t . The number of simultaneously assimilated observations is therefore 1 and \mathbf{y}_t is just a scalar y_t . We call *super-observations* the retrievals aggregated on G for each UTC day (note that the 16 km grid points of the CAMS forecasts have also been aggregated on G). These super-observations are built by averaging the individual observations in the grid boxes with an arithmetic mean. Typically, 200 retrievals are aggregated for a single $2^\circ \times 2^\circ$ super-observation. By comparison, the collocation criterion used by *Wunch et al.* [2017] between OCO-2 retrievals and TCCON measurements is much larger than our grid boxes. We also average the measurement time to compute the associated t . During most days, there are a few hundred super-observations. We assimilate them in chronological order.

5. The error covariance matrix of y_t (\mathbf{R}_t) is just a scalar r_t . It is assigned from the error statistics associated with the individual observations under the assumption that retrieval errors are fully correlated in each grid box of G (because of common hypotheses in the retrievals and of close sounding conditions). To do that, we explicitly form the covariance matrix of the individual XCO₂ observations in the grid box.
6. The observation operator \mathbf{H}_t is the sampling operator that selects the grid cell where y_t is located. \mathbf{H}_t is therefore just a vector \mathbf{h}_t filled with 0, except for the observation grid cell where the element is 1.

Our KF uses the following recurrence relations for the posterior mean and the posterior uncertainty:

$$\mathbf{x}_t^b = \mathbf{x}_{t-1}^a \quad (1)$$

$$\mathbf{U}_t^b = \mathbf{U}_{t-1}^a + \mathbf{Q}_t \quad (2)$$

Note that t was defined as an index (see section 3.1), so that $t - 1$ is still an index.

Due to the discrete nature of \mathbf{Q}_t (3-hourly for practical purpose, as explained above), equation (2) only applies when a new 3-hourly slot has passed (at 00, 03, 06, 12, 15, 18, or 21 UTC), rather than at each time step (defined as a new measurement time). If a new 3-hourly slot has not passed, equation (2) simplifies to

$$\mathbf{U}_t^b = \mathbf{U}_{t-1}^a \quad (3)$$

The analysis step of the KF can be written as

$$\mathbf{k}_t = \mathbf{U}_t^b \mathbf{h}_t^T (\mathbf{h}_t \mathbf{U}_t^b \mathbf{h}_t^T + r_t)^{-1} \quad (4)$$

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{k}_t (y_t - \mathbf{h}_t \mathbf{x}_t^b) \quad (5)$$

$$\mathbf{U}_t^a = (\mathbf{I}_n - \mathbf{k}_t \mathbf{h}_t) \mathbf{U}_t^b \quad (6)$$

\mathbf{I}_n is the ($n \times n$) identity matrix.

With this formulation, there is no matrix to be inverted since $(\mathbf{h}_t \mathbf{U}_t^b \mathbf{h}_t^T + r_t)$ in equation (4) is a scalar, which makes each KF analysis step computationally simple and fast. The filter is initialized by filling all elements of \mathbf{x}_0^b with the mean of all OCO-2 XCO₂ measurements during the first month. We also define \mathbf{U}_0^b by the combination of their variance (a single number for the whole globe) and of the correlations in \mathbf{Q}_0 . With this setup, the influence of \mathbf{x}_0^b and \mathbf{U}_0^b vanishes after a few weeks of KF analyses in most latitudes.

After each application of equation (6), we check that all error variances in \mathbf{U}_t^a are still positive (which is the case in our application). Negative values would reveal numerical instabilities or an error in the implementation.

3.3. Additional Choices

So far, the algorithm is a rather straight-forward application of the KF for a model of persistence. Each practical application of the theoretical KF has its limitations, and we have defined three pragmatic adaptations for our problem, as follows:

When some variances in \mathbf{U}_t^a become very small, future observations may not influence the filter much in the corresponding grid box, at least until matrix \mathbf{Q} has been applied enough to significantly raise the variances in \mathbf{U}_t^b again. In this case, we reset the variances in \mathbf{U}_t^a to a minimum value, which is 0.25 ppm². This choice makes the subsequent background uncertainty not less than the uncertainty of typical retrievals [Eldering *et al.*, 2017]. We update the covariances accordingly.

The long-term global trend of CO₂ biases our model of persistence \mathbf{M}_t , the uncertainty of which is therefore ill-described by covariance matrix \mathbf{Q}_t . In order to allow the use of the KF in near real time, we choose not to use the information provided by surface measurements to remove the bias. However, we avoid that some grid points of G diverge from the rest of the pixels, like those in the high latitudes of the winter hemisphere, by imposing the requirement that the less reliable pixels (defined as the pixels whose error variance in \mathbf{U}_t^a is larger than 4 ppm², consistent with the performance of the KF shown later in Figure 8a) shall not differ from

the mean of all pixels by more than 5 ppm (which usually represents about twice the standard deviation of the CAMS forecast over the globe for a given time step). This limit should also bind the errors of each pixel, and we put a ceiling value of 16 ppm² on it. By comparison, the variance of the global 16 km resolution CAMS XCO₂ forecasts for 2015 is less than this value at each time step for 2015.

In order to better satisfy the consistency diagnostics of *Desroziers et al.* [2005], we inflate the assigned super-observation error variances r_t by a time-independent factor in each grid box shown later in Figure 7b. The factor has been deduced from a preliminary 2 year run of the KF without any tuning, which will be explained in section 4.3.

3.4. Possible Temporal Aggregation of the Results

The main output of the KF is the time series $\{\mathbf{x}_t^a, \mathbf{U}_t^a\}$. In practice, $\{\mathbf{x}_t^a, \mathbf{U}_t^a\}$ is available after each measurement, a frequency that is unnecessarily high for practical use. The couple is therefore stored only after the last measurement of each day (in UTC). At this time step, the KF has assimilated all observations of that day, and \mathbf{x}_t^a therefore best represents the 2-D map of estimated XCO₂ for that day.

The series $\{\mathbf{x}_t^a, \mathbf{U}_t^a\}$ allows computing maps of XCO₂ at coarser temporal resolution, such as monthly time scales, by simply averaging the daily maps. The uncertainty of the monthly mean map can be estimated by first summing the corresponding terms in the diagonal of \mathbf{U}_t^a , divided by the number of days in the month. Temporal correlations of the errors of the daily maps can then be accounted for by noticing that the covariance between the errors of \mathbf{x}_t^a and of \mathbf{x}_t^b (and, by recurrence, of any past \mathbf{x}_t^b of the KF if we neglect the influence of \mathbf{Q}_t) is simply \mathbf{U}_t^a : the proof of this equality stems from the definition of this covariance combined with equations (5) and (6) and with the assumption of uncorrelated prediction and observation errors:

$$\begin{aligned}
 \mathbf{U}(\mathbf{x}_t^a, \mathbf{x}_t^b) &= \overline{(\mathbf{x}_t^a - \mathbf{x}_t^f)(\mathbf{x}_t^b - \mathbf{x}_t^f)} \\
 &= \overline{(\mathbf{x}_t^b + \mathbf{k}_t(y_t - \mathbf{h}_t \mathbf{x}_t^b) - \mathbf{x}_t^f)(\mathbf{x}_t^b - \mathbf{x}_t^f)} \\
 &= \overline{(\mathbf{x}_t^b - \mathbf{x}_t^f + \mathbf{k}_t(y_t - \mathbf{h}_t \mathbf{x}_t^f + \mathbf{h}_t \mathbf{x}_t^f - \mathbf{h}_t \mathbf{x}_t^b))(\mathbf{x}_t^b - \mathbf{x}_t^f)} \\
 &= \overline{(\mathbf{x}_t^b - \mathbf{x}_t^f - \mathbf{k}_t(\mathbf{h}_t \mathbf{x}_t^b - \mathbf{h}_t \mathbf{x}_t^f))(\mathbf{x}_t^b - \mathbf{x}_t^f)} \\
 &= \mathbf{U}_t^b - \mathbf{k}_t \mathbf{h}_t \mathbf{U}_t^b \\
 &= \mathbf{U}_t^a
 \end{aligned} \tag{7}$$

4. Results

4.1. Error Covariances of the Persistence Model

In the absence of a transport model in our KF, the error covariance matrix of the persistence model, \mathbf{Q}_t , is the only possibility for the pixels of G to exchange information. When an XCO₂ retrieval is assimilated, its information is spread in space further to the persistence uncertainty that has been accumulated so far (equation (2)). After the assimilation of a given retrieval, the KF uncertainty exhibits reduced spatial correlations around the location of that retrieval (equation (6)). \mathbf{Q}_t also controls the growth of the error in the KF and therefore plays a critical role in the reliability diagrams shown later in section 4.4.

Figure 1 illustrates the correlations of \mathbf{Q}_t . We take the example of the month of June 2015. We focus on the correlations along latitudes 40°N (Figure 1a) and 40°S (Figure 1b) and along longitudes between 0° and 16°E (Figure 1c). The correlations drop with the distance in the first 2000 km and even reach negative values (down to ~ -0.4 for latitude 40°N). Note that the correlation scales are still much larger than the grid-box size, which means that in our hypotheses, the retrievals are much less correlated than the persistence error, a property that will be exploited later in section 4.3. The negative values have not been noted in previous studies of satellite data gap filling. They reflect the movement of synoptic-scale CO₂ plumes, for instance, in low-pressure systems. Small correlations of either sign exist for longer distances, even after the ensemble localization (see section 3.2). The overall pattern varies between the two latitudes, the southern one exhibiting more negative correlations. Along longitudes between 0° and 16°E, correlations drop much more rapidly, implying that they are not isotropic. The correlation density is also flatter.

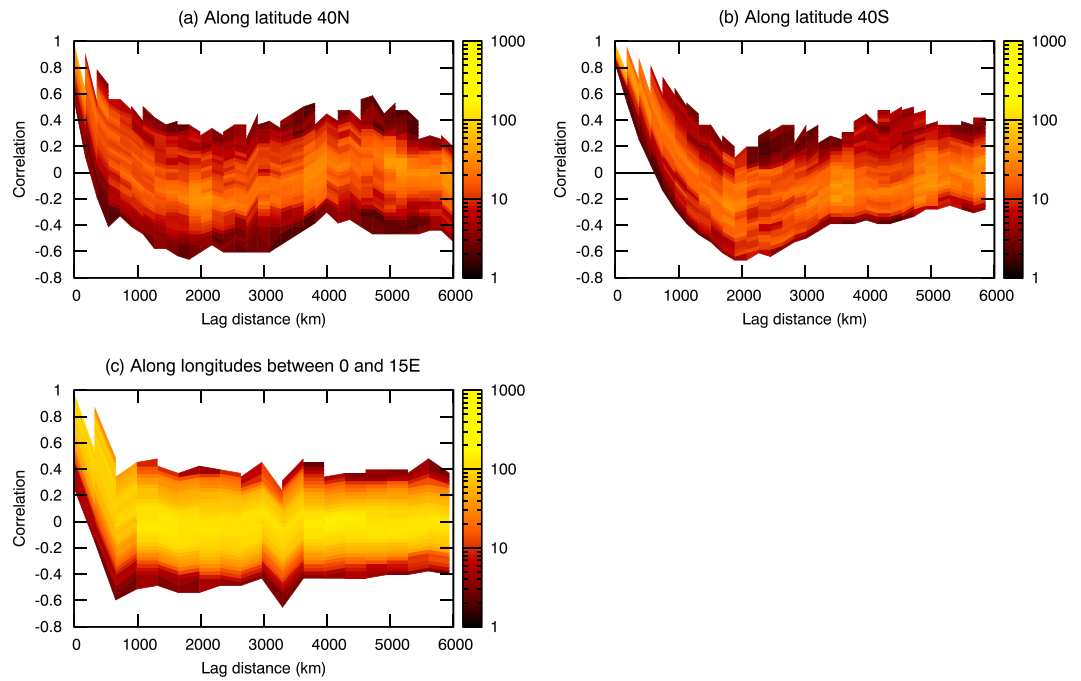


Figure 1. Density of the assigned correlations of the persistence errors (in matrix Q_t) for June 2015 between the grid points along latitudes (a) 40°N and (b) 40°S and along longitudes (c) between 0° and 15°E.

The geographical distribution of the variances (Figure 2) is heterogeneous, with most values around 0.2 ppm (per day) and a largest one of 2.8 ppm. This distribution reflects both intense surface fluxes (for instance, the biomass burning activity in Southern Africa or anthropogenic emissions on the East Coast of the USA and of China) and transport (for instance, in the storm tracks or the region around the Andes and the Mato Grosso plateau) during that month.

4.2. Overall Behavior

The KF has been run in the above-described configuration for the period of September 2014–August 2016. The mean uncertainty reduction at the grid point scale (defined as 1 minus the ratio of the analysis uncertainty standard deviation on the background one) is about 0.4, with a corresponding mean degree of

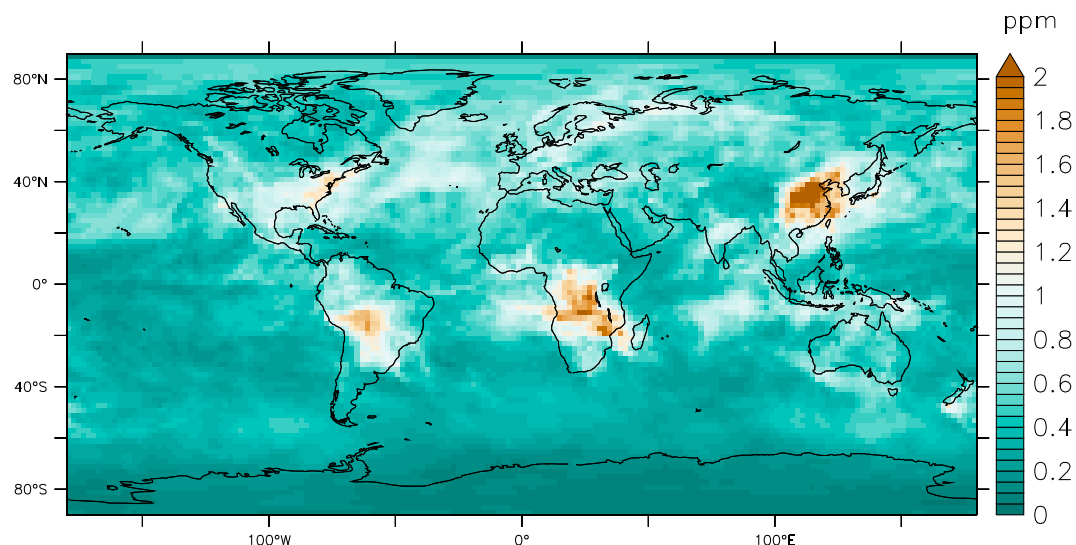


Figure 2. Standard deviations of Q_t per 24 h for June 2015, in ppm.

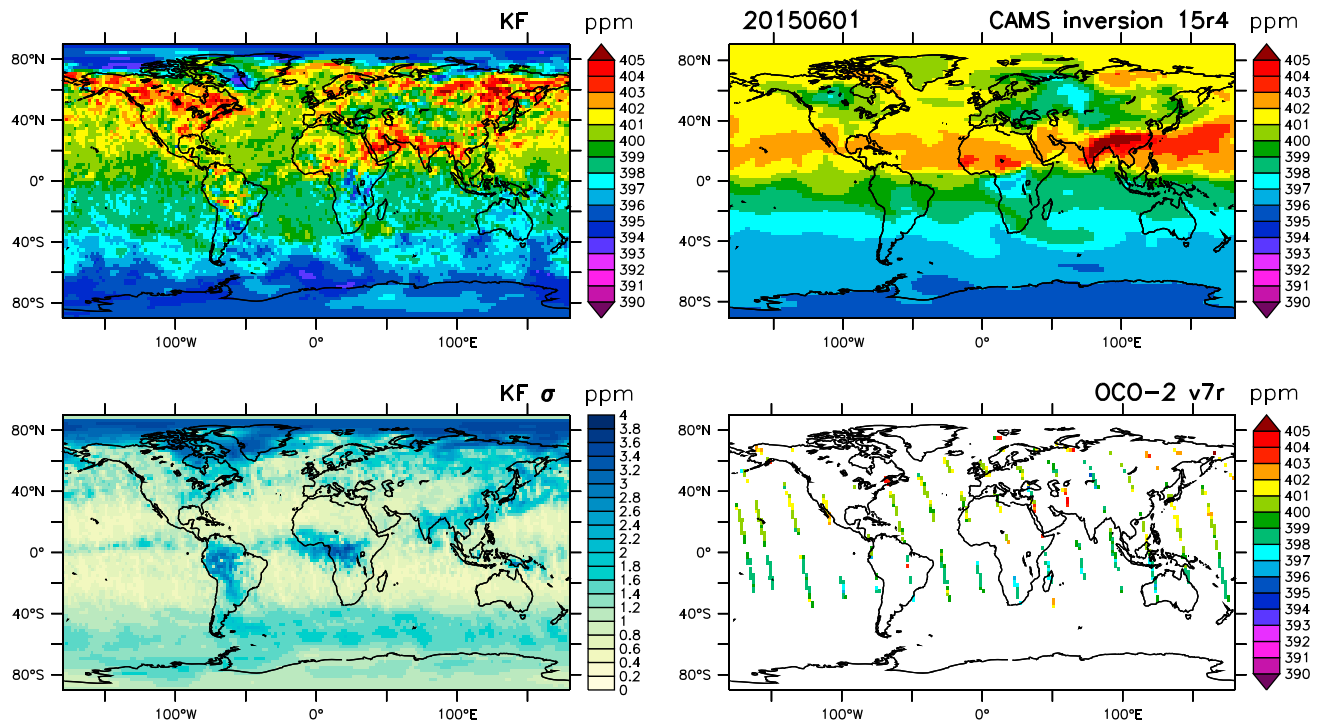


Figure 3. The KF analysis for 1 June 2015 is shown in the top left map, together with its uncertainty standard deviation in the bottom left map. The assimilated OCO-2 super-observations on that day are in the bottom right map. The daily-mean CAMS CO₂ inversion is shown for comparison in the top right map.

freedom for signal of 0.6. The medium uncertainty reduction reflects some redundancy between the spatial information brought by each retrieval. It implies that the KF does not perfectly fit the super-observation values, but nearly all of them are still fitted within their assigned uncertainty standard deviation (not shown).

Figure 3 displays the KF state for 1 June 2015. The 475 assimilated super-observations for the same day are shown in the bottom right corner: OCO-2 was operated in glint mode, allowing XCO₂ to be retrieved on both land and ocean, in 15 orbit tracks across the sunlit hemisphere. The KF provides values for all grid points (upper left corner) but with varying uncertainty (bottom left corner). Some areas where Q_t has large variances (Figure 2) or where there are no recent retrievals, like the high latitudes, show larger uncertainty ($\sigma > 2$ ppm), while pixels in the subtropical highs, which are well observed by OCO-2, have more reliable XCO₂ values ($\sigma \approx 0.5$ ppm). The KF diagnoses a large north-south XCO₂ gradient of 10 ppm between 60°N and South Pole. Values are comparable in the high latitudes of both hemispheres, but the northern ones are much more uncertain than the southern ones. The northern ones beyond 70°N have hardly been constrained by the assimilation of the retrievals, and their values do not differ much from their initialization values of 1 September 2014. Large values (~9 ppm larger than the initialization value) are seen over the boreal forests in both North America and Eurasia. They echo the large unrealistic XCO₂ retrievals found in the ACOS-GOSAT retrievals for the same month of the year (but for different years) [Chevallier, 2015]. In particular, they show the same discontinuity with the neighboring regions, like the tundra vegetation north of Siberia and the grassland/cropland regions south of them (see Figure 3b of Chevallier [2015] and the corresponding discussion). For comparison, the CAMS CO₂ inversion for the same day is shown in the upper right corner of Figure 3. The overall distribution of XCO₂ in the northern hemisphere is quite different, with a maximum around the 20°N and much lower values further north over land. The distribution in the southern hemisphere is more comparable, even though the latitudinal gradient appears to be larger with the KF.

Figure 4 repeats Figure 3 for 1 September 2015. This time, the KF has larger values than the CAMS inversion in the middle and high northern latitudes, while values are smaller south of 40°S. The KF shows a band of high values around 40°S, including over the ocean, which do not reflect known CO₂ sources. Elderling et al. [2017] discussed this feature as a bias that is partly linked to the lack of stratospheric aerosols in the current version of the retrieval algorithm.

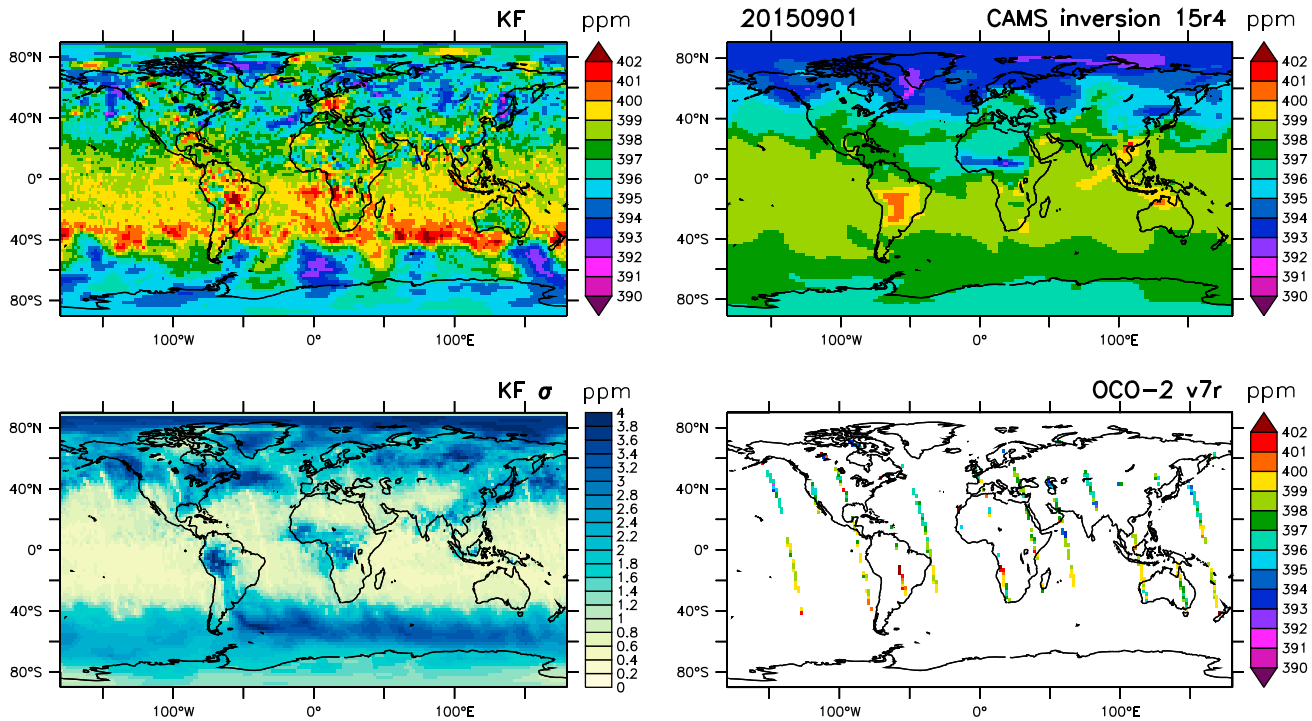


Figure 4. Same as Figure 3 for 1 September 2015.

Figure 5 displays the time series of the mean daily change of XCO_2 in the tropics ($23.4^{\circ}S-23.4^{\circ}N$) over 30 day rolling periods for both the CAMS CO_2 inversion (until December 2015 only) and the KF. The change of XCO_2 is a quantity of particular interest because it can be related to the change of CO_2 mass in the atmosphere when combined with dry surface pressure fields (for instance from numerical weather prediction centers); it is therefore one of the terms needed to estimate the surface fluxes with a direct carbon budgeting approach (the second term is the lateral transport of CO_2 [see Crevoisier *et al.*, 2006]). We focus here on the tropics because this part of the globe is relatively well constrained in the KF throughout the year (bottom left corners of Figures 3 and 4). The variations of KF and of the CAMS inversion are broadly similar, with peaks late in the year and in Spring, likely related to fire activity in Tropical Africa or Asia. The disagreement between the series during the first few weeks can be attributed to the KF spin up. We also notice that the KF tends to lag behind CAMS by a few weeks, consistent with the low revisit time of the retrievals. The CAMS inversion likely better picks up the underlying flux signal at the correct time, as can be seen later in Figure 8a.

4.3. Consistency Diagnostics

Our KF uses specific error models to describe uncertainty. They are fully expressed by covariance matrices U_t^a , U_t^b , R_t , and Q_t that are linked together in the KF equations (2)–(6). The validity of most assumptions can be cross-checked from these equations as follows.

U_t^b is too large to be inverted for each retrieval, and we therefore cannot monitor the Bayesian cost function of the KF, which is used for instance in χ^2 tests [e.g., Talagrand, 2014]. However, other diagnostics can still be used.

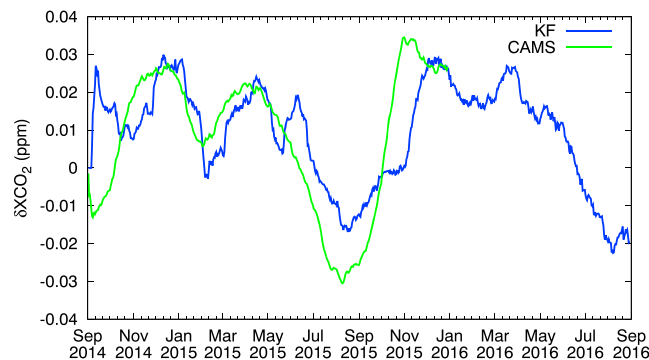


Figure 5. Time series of the 30 day rolling means of the daily change of XCO_2 in the tropics from the CAMS inversion and from the KF.

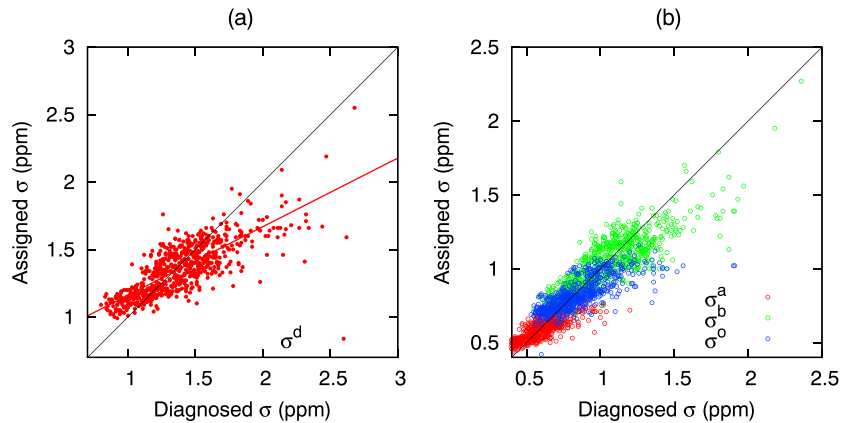


Figure 6. Scatter plot of standard deviations, either assigned in the KF or diagnosed from the consistency tests over 2 years. (a) Daily statistics of the innovation d , in red, and associated regression line. (b) Daily statistics of the analysis, background, and observations, in red, green, and blue, respectively. The bisector is also shown in black.

The KF assumes unbiased normally distributed \mathbf{x}_t^b and y_t , which should imply that the background misfits $y_t - \mathbf{h}_t \mathbf{x}_t^b$ (also called *innovations*) are also unbiased and normally distributed. In our case, the mean, skewness, and kurtosis of the innovation are, respectively, 0.1 ppm, 0.0, and 9, which characterizes a nearly unbiased and symmetric distribution, but with many more outliers than a Gaussian one with the same standard deviation. Additionally, innovations should not be serially correlated for the KF to be optimal [Talagrand, 2014]. Our assimilation of super-observations individually makes it difficult to check this property at any lag (given the size of the super-observations and the characteristics of the OCO-2 orbit, correlations may exist at any lag between 30 s and 1 month) or even to define lags for the autocorrelations (because the soundings are irregularly spread in time). However, we checked the decorrelation property from one KF analysis step to the next each day (usually corresponding to nearby super-observations): the distribution of autocorrelations at lag 1 within each day has a mean and standard deviation of 0.1 ± 0.1 only.

The statistics of the innovation ($y_t - \mathbf{h}_t \mathbf{x}_t^b$) should be consistent with the sum of background and observation error covariance matrices [see, e.g., Chevallier and O'Dell, 2013]. Figure 6a presents the scatterplot of assigned versus diagnosed standard deviations of the daily innovations (one dot represents 1 day, irrespective of the number of super-observations or of where they are each day). The mean slope is 0.5 ppm/ppm, which shows some skill in the assigned error statistics despite an overestimation of the smaller ones and an underestimation of the larger ones. The offset of the regression line is 0.6 ppm. Complementary diagnostics provide some insight into this underestimation. As shown by Desroziers *et al.* [2005], in a well-tuned KF, the statistics of various residuals in the observation space (described hereafter) equal $\mathbf{H}\mathbf{U}_t^a\mathbf{H}^T$, $\mathbf{H}\mathbf{U}_t^b\mathbf{H}^T$, and \mathbf{R}_t . A mean r_t can be obtained from the statistics of the covariances between corresponding innovations ($y_t - \mathbf{h}_t \mathbf{x}_t^b$) and observation-minus-analysis residuals ($y_t - \mathbf{h}_t \mathbf{x}_t^a$). Similarly, the mean background error variance in the observation space should equal the covariance between corresponding analysis-minus-background residuals ($\mathbf{h}_t \mathbf{x}_t^a - \mathbf{h}_t \mathbf{x}_t^b$) and innovation ($y_t - \mathbf{h}_t \mathbf{x}_t^b$), while the mean analysis error variance in the observation space should equal the covariance between ($\mathbf{h}_t \mathbf{x}_t^a - \mathbf{h}_t \mathbf{x}_t^b$) and ($y_t - \mathbf{h}_t \mathbf{x}_t^a$). These equivalences allow tuning the error statistics iteratively provided that background and observation errors have different structures in reality and in the assigned statistics, so that background and observation errors predominate in different scales [Desroziers *et al.*, 2005; Cressot *et al.*, 2014; Todling, 2015]. This advantageous configuration is also ours, since background errors, driven by \mathbf{Q}_t , likely have much larger scales than the retrievals (section 4.1). Figure 6b shows the scatterplot of assigned versus diagnosed statistics of daily $\mathbf{H}\mathbf{U}_t^a\mathbf{H}^T$, $\mathbf{H}\mathbf{U}_t^b\mathbf{H}^T$, and \mathbf{R}_t (one dot represents 1 day, again irrespective of the number of super-observations or of where they are each day). This diagnostic attributes the lack of resolution of the assigned innovation statistics mostly to the background error statistics: the slope and offset are similar (0.6 ppm/ppm and 0.4 ppm, respectively). The diagnosed observation error statistics agree with the assigned ones below 1 ppm, but the assigned values appear to saturate for the larger diagnosed values. The assigned analysis error statistics behave like the observation ones with a fair behavior below 0.7 ppm (the correlation is 0.7) and saturation about that value.

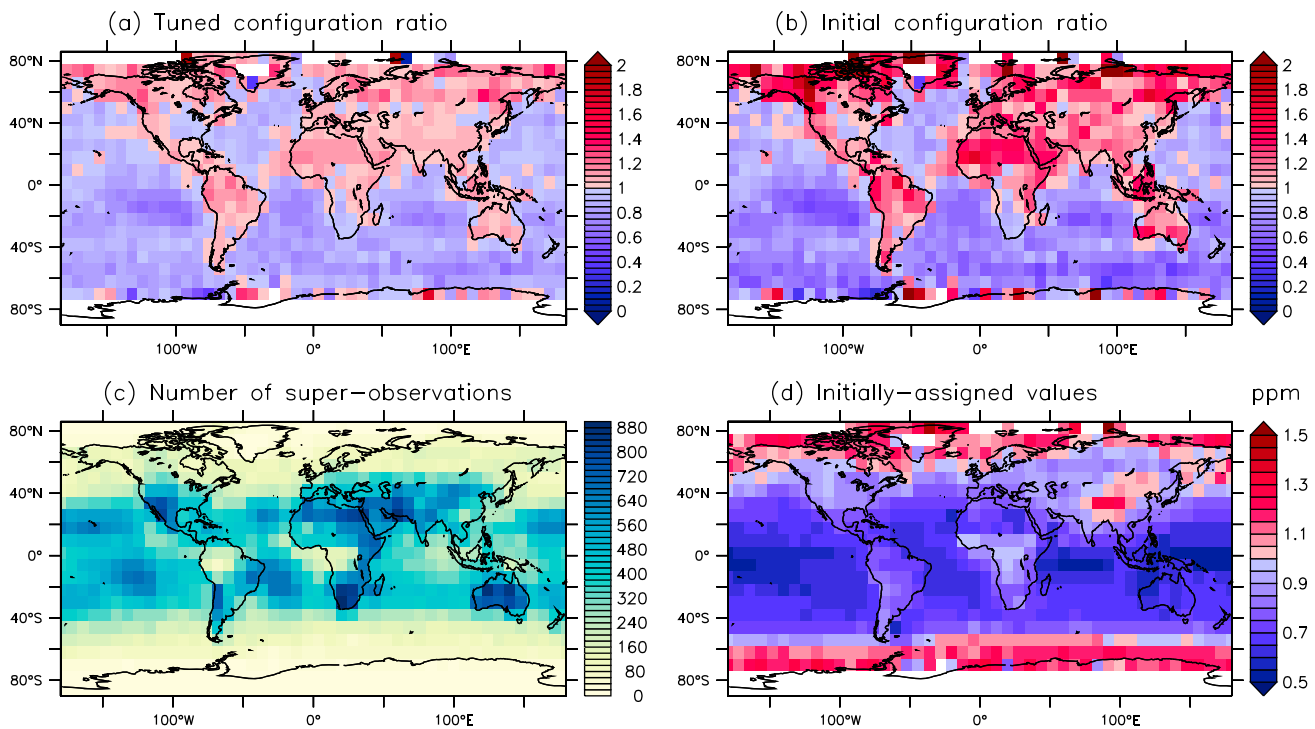


Figure 7. $8^\circ \times 8^\circ$ maps of the ratio of observation error standard deviations diagnosed from the consistency tests over the ones assigned in the KF in the (a) tuned configuration used in this paper and for the (b) configuration without tuning over 2 years. (c) Number of data in the statistics. (d) Quadratic mean of the observation error standard deviations assigned before the tuning.

The statistics can also be computed on the full time series in each grid box. Figure 7a displays the map of the ratio of diagnosed over assigned observation error statistics. The resolution has been coarsened to $8^\circ \times 8^\circ$ to enhance the robustness of the statistics. Diagnosed standard deviations are mostly larger than the assigned ones by up to 10% over land (up to 30% in the high latitudes), while being smaller by up to 20% over the ocean. The land versus ocean distinction would be much enhanced if we had not tuned our observation error statistics based on this diagnostics. Indeed, as mentioned in section 3.3, we inflate the observation error statistics in each KF analysis step with the ratio map of Figure 7b obtained from a preliminary 2 year run of the KF without any tuning, in order to enhance the internal consistency of the KF (following Desroziers *et al.* [2005] again). Comparing Figures 7a and 7b, we see that the tuning reduces the relative differences by about three-fold. The initial ratio does not correlate well with the data density (Figure 7c), which seems to exclude an artifact of the ensemble size in the residual statistics. It also shows distinct behavior compared to the initially assigned error statistics (Figure 7d), for instance, over the Sahara and around the Saharan air layer in the tropical North Atlantic Ocean, where the assigned error standard deviations are small (0.7 ppm) further to the bright surface reflectance (in glint mode over the ocean, in both nadir and glint modes over the land), while the diagnostics bring them about 1 ppm. Note that the overall patterns of Figure 7b are stable over time: they are very similar when the diagnostics are computed over either 1 or 2 years (not shown). We do not try to iterate further on the observation error statistics, even though Figure 7a suggests further increasing the land-sea contrast. Similar tuning of the background error statistics is possible but is hampered by its dependency on the observation error statistics of the previous KF analysis steps. Hence, we do not attempt this here.

4.4. Prediction Skill

Following Talagrand *et al.* [1999], we use a reliability diagram to visualize the agreement between the predicted probabilities ($\mathbf{x}_t^a, \mathbf{U}_t^a$) at the end of each day and verifying individual TCCON observations. Predictions \mathbf{x}_t^a are sampled at the location of TCCON measurements and binned as a function of the predicted uncertainty. The mean and the standard deviation of the prediction-minus-TCCON misfits are

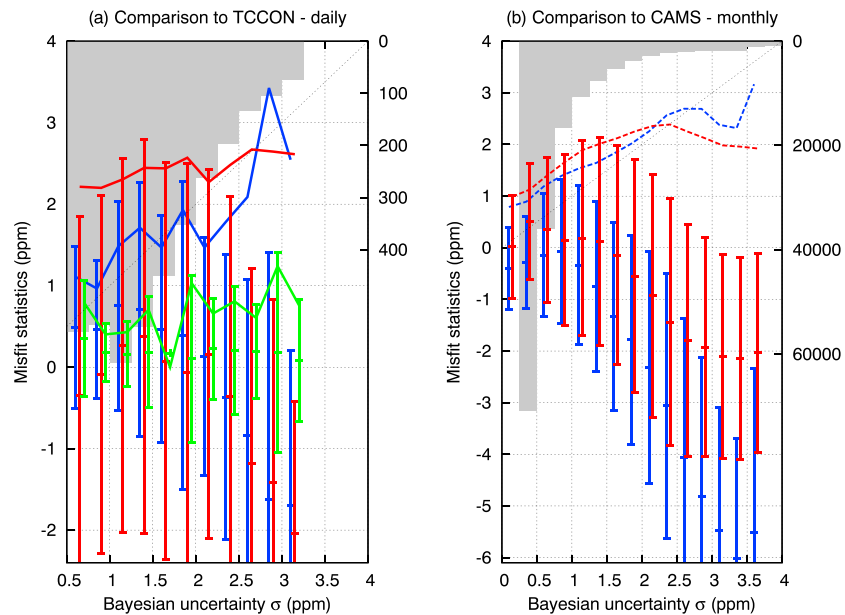


Figure 8. Reliability diagram of the KF (blue) at the (a) daily and (b) monthly scales for the period from September 2014 until December 2015. The vertical bars denote the mean \pm the standard deviation for the misfits (KF minus reference) in each KF uncertainty bin. Standard deviations are corrected from TCCON errors and from neglecting the retrieval averaging kernels (see text). For comparison, the statistics for the XCO₂ climatology (red) and for the CAMS atmospheric inversion (green, at the daily scale only) are also shown in each KF bin. The colored continuous lines in Figure 8a across the bins correspond to the RMSD. The colored dotted line in Figure 8b across the bins corresponds to the standard deviation. The bisector is shown as a thin black line in Figures 8a and 8b. The number of data in each bin appears as an upside-down gray-shaded histogram, with the axis on the right. Bins with less than 30 data have been discarded and statistics bar has been slightly shifted from each other for their visibility. The reference for the misfits is provided by TCCON and the CAMS inversion for Figures 8a and 8b, respectively.

displayed in each uncertainty bin in Figure 8a. With a reliable KF, the mean curve and the standard deviation curve align with the abscissa and the bisector, respectively. We remove the TCCON uncertainty standard deviation and a term that accounts for neglecting the retrieval averaging kernels (0.24 ppm, from *Nguyen et al.* [2014b]) from the misfit standard deviations in quadrature. The root-mean-square difference curve (RMSD) is also shown in Figure 8a. The performance of the KF appears in blue, while the same metric applied to the climatology of XCO₂ (see section 2.5) and to the CAMS atmospheric inversion (see section 2.4) appears in red and green, respectively. The KF has usually positive biases around 0.5 ppm, consistent with the validation of the v7r retrievals (Table 3 of *Wunch et al.* [2017]), but they become much negative for the larger uncertainty bins, again reflecting the bias of the persistence model. The KF RMSD curve reasonably aligns with the bisector indicating fair reliability. Its oscillations are related to its geographical heterogeneity, as illustrated by the bottom left corners of Figures 3 and 4: small uncertainty bins are rather from the TCCON stations located around the subtropical highs (Ascension, La Réunion, Darwin, Wollongong), while large uncertainty bins are mostly from the middle and high latitude ones or from Manaus in Brazil. The underestimation of the small uncertainty bins was already seen in the internal diagnostics of section 4.3 and Figure 6b (the red dots), but these examples did not cover the larger uncertainty bins because these bins mostly correspond to areas far from the retrieval locations. The climatology of XCO₂ also has a rather stable RMSD around 2.5 ppm. This means that the KF does not simply predict the climatological distribution of XCO₂ (which would make it reliable but practically useless [see *Candille and Talagrand*, 2005]): the KF shows skill when its predicted uncertainty is less than 2.5 ppm. Consistent with previous validation exercises [e.g., *Kulawik et al.*, 2016], the error statistics of the CAMS inversion are mostly well within 1 ppm. If we select the KF uncertainty bins less than, e.g., 1 ppm for the period from September 2014 until April 2016 that was used by *Wunch et al.* [2017] to validate the retrievals, we find 1330 misfits with TCCON, a mean bias of 0.3 ppm, and an RMSD of 1.5 ppm. By comparison, *Wunch et al.* [2017], based on specific collocation criteria, found 1618 comparison points, a mean bias of 0.1 ppm, and an RMSD of 1.4 ppm.

Statistics at the site levels are less comparable because the number of selected TCCON data varies much between the two approaches.

The good performance of the CAMS inversion with respect to TCCON gives us confidence to take it as the reference for monthly means. The advantage of using it rather than TCCON is the full continuity of its time series within each month. The performance of each grid points can be accounted for, rather than just TCCON points, but we thin the grid point density toward the pole in order to have equal density in km^2 along the latitudes. The corresponding reliability diagram is displayed in Figure 8b. Note that we have not removed the uncertainty of the CAMS monthly XCO_2 values from the statistics because we have not estimated it at this temporal scale. The better representation of high latitudes in these statistics highlights the bias of the persistence model, with negative biases of several ppm for the larger KF uncertainty bins. We therefore look at the KF standard deviation curve (blue dotted line) rather than at the RMSD curve (not shown). It also shows smaller slope than the bisector and beats the climatology curve (red dotted line) by usually several tenths ppm for KF uncertainty bins less than 2.5 ppm. Note that the proportion of uncertainty bins less than 2.5 ppm is much larger at the monthly than at the daily scale, as shown by the two bin histograms.

5. Conclusions

We have presented a Kalman filter (KF) that assimilates the individual OCO-2 CO_2 column retrievals in a model of persistence and, cycle after cycle, provides a global map of that column with associated uncertainties in the form of a covariance matrix. It does not imply any dimension reduction beyond a horizontal discretization ($2^\circ \times 2^\circ$ in our case) and therefore does not impose any smoothness or isotropy constraint on the XCO_2 maps. Error autocovariances of the KF are also estimated, which allows documenting the uncertainty of the maps when aggregated a posteriori at coarser temporal resolution, like the month.

The global maps generated by the KF fulfill the role of a “level 3” satellite product, i.e., retrievals that are mapped on a uniform space-time grid with some completeness and consistency properties [NASA, 1986]. To maximize its utility in this role, we have excluded any information from outside the retrievals themselves (like meteorological variables, advection models, or surface global growth rate measurements), apart from frozen statistics of the day-to-day variability of daily-mean XCO_2 taken from high-resolution model simulations. In each data assimilation cycle, we have also ignored measurements made after the cycle date (i.e., we have not designed a Kalman smoother), so that the level 3 product can be generated in real time together with the retrievals, provided that all input statistics are properly assigned and do not need tuning after a training period. At global horizontal resolution $2^\circ \times 2^\circ$, a whole year worth of OCO-2 retrievals could be processed in 2 days on a single Intel Xeon E5-2650 v2 processor at 2.6 GHz and with 11 GB allocated memory. Running the KF at higher spatial resolutions (like $1^\circ \times 1^\circ$) may be interesting if the retrieval errors have much smaller correlation lengths than the size of the $2^\circ \times 2^\circ$ grid boxes. In that case, to gain computational efficiency, the retrievals can be processed in batches with parallel algorithms, rather than one at a time on a single processor.

The KF uses frozen statistics of the XCO_2 variability within a month. The statistics have been obtained from the CAMS high-resolution forecasts for the year 2015, but we have not noticed any degradation of the KF behavior for the months of 2014 and 2016. Alternatively, it is possible to use a rolling library of forecast statistics that is generated in real time for the KF. The filter statistics quality actually degrades from the analysis to the later background (Figure 6b), which suggests that those statistics do not perfectly fulfill their role: improving them is left for future work, but one could think of the KF as a test bed for the realism of variability statistics of XCO_2 , as it is for the realism of error statistics of the retrievals.

Due to the scattered satellite sampling, the broad scale patterns of XCO_2 described by the KF seem to lag behind the real signals by a few weeks. The KF predicted uncertainties are usually smaller where recent retrievals have been assimilated and larger elsewhere. If the filter works well, they are statistically consistent with the misfits between the KF XCO_2 maps and independent estimates of the column, given their respective uncertainty. We found KF biases up to 6 ppm in the high latitudes of the winter hemisphere and a lack of resolution of the predicted standard deviations. However, our KF shows some skill in predicting its error statistics at both daily and monthly scales, while being overall more accurate than a climatology of XCO_2 . Reliability is critical for users of the KF maps because these maps are heterogeneous in quality. In particular, this property

can be used to rigorously compare the retrievals with distant validation data, for instance, for satellite missions that have a narrow swath without any target mode, like the Methane Remote Sensing Lidar Mission (MERLIN) [Kiemle et al., 2011; Pierangelo et al., 2016] scheduled for launch in 2021. Note that in that case, the satellite retrievals and the validation data have to have similar weighting function, since our KF ignores the retrieval averaging kernels.

We used classical consistency diagnostics to evaluate the KF behavior. Fair consistency is achieved after tuning the error variances of the super-observations (i.e., of spatially averaged retrievals) to reduce some of their lack of resolution. Spatial averaging has likely damped the uncertainty of each super-observation (see section 3.4 of Kulawik et al. [2016]), while we assumed that it did not (section 3.2). The tuning partly accounts for this overestimation, and our diagnostics consistently suggests deflating the assigned observation uncertainty over the ocean. Over land in both hemispheres, however, the diagnostics suggests the opposite, which points to a feature from the retrievals rather than from the aggregation process. Such feature is consistent with the known neglect of some error sources in the computation of the OCO-2 v7r retrieval uncertainty [Eldering et al., 2017, Connor et al., 2016]. We have also seen that the overall KF-TCCON statistics for the most precise areas of the KF maps are similar to the overall statistics shown in the retrieval validation study of Wunch et al. [2017]. Last, we have shown that a simple visual inspection of the maps already revealed some OCO-2 retrieval biases, which would be less straight-forward by looking at the individual retrievals. This KF can therefore help constrain the quality of the retrievals to some extent, with direct benefits to other users of the retrievals like inverse modelers. Data assimilation systems built on chemistry-transport models [e.g., Massart et al., 2014] may also provide such a feedback on retrieval random and systematic errors, and they may even be more exhaustive, but the simplicity and the relative computational economy of the Kalman filter allows running it closer to the retrieval production for faster monitoring. Additionally, using a chemistry-transport model would further damp the specific contribution of the retrievals themselves in the data assimilation results, while we saw that the mean degree of freedom for signal of the retrievals in our Kalman filter is only 0.6.

The algorithm described here is mostly generic and contains few adaptations that are specific to OCO-2 retrievals (control of the extreme values of the maps and of its uncertainty, see section 3.3). It could easily be adapted to other CO₂ satellite missions, or even to retrievals of the column of other long-lives species, like methane or carbon monoxide, provided that statistics of their variability can be estimated.

Acknowledgments

Some of this work was performed using HPC resources of DSM-CCRT and of CCRT under allocation t2016012201 made by GENCI (Grand Équipement National de Calcul Intensif). The ACOS OCO-2 data can be obtained from <http://co2.jpl.nasa.gov>. They were produced by the ACOS/OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology. The CAMS products can be obtained from <http://atmosphere.copernicus.eu/>. TCCON data were obtained from the TCCON Data Archive, hosted by the Carbon Dioxide Information Analysis Center (CDIAC) at Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, <http://tcccon.ornl.gov>. The authors are very grateful to the many people involved in the surface and satellite CO₂ measurements and in the archiving of these data that were kindly made available to them. They received funding from the French space agency, CNES, as part of the preparation for the MERLIN satellite, and from the Copernicus Atmosphere Monitoring Service, implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission. They also thank Christopher O'Dell (CSU) for many inspiring discussions about the ACOS retrievals and François-Marie Bréon (LSCE) for his careful reading of an earlier version of the paper.

References

- Agustí-Panareda, A., M. Diamantakis, V. Bayona, F. Klappenbach, and A. Butz (2016a), Improving the inter-hemispheric gradient of total column atmospheric CO₂ and CH₄ in simulations with the ECMWF semi-Lagrangian atmospheric global model, *Geosci. Model Dev.*, doi:10.5194/gmd-10-1-2017.
- Agustí-Panareda, A., S. Massart, F. Chevallier, G. Balsamo, S. Boussetta, E. Dutra, and A. Beljaars (2016b), A biogenic CO₂ flux adjustment scheme for the mitigation of large-scale biases in global atmospheric CO₂ analyses and forecasts, *Atmos. Chem. Phys.*, *16*, 10399–10418, doi:10.5194/acp-16-10399-2016.
- Alkhaled, A. A., A. M. Michalak, and S. R. Kawa (2008), Using CO₂ spatial variability to quantify representation errors of satellite CO₂ retrievals, *Geophys. Res. Lett.*, *35*, L16813, doi:10.1029/2008GL034528.
- Boussetta, S., et al. (2013), Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: Implementation and offline validation, *J. Geophys. Res. Atmos.*, *118*, 5923–5946, doi:10.1002/jgrd.50488.
- Brunner, D., J. Staehelin, H.-R. Künsch, and G. E. Bodeker (2006), A Kalman filter reconstruction of the vertical ozone distribution in an equivalent latitude—Potential temperature framework from TOMS/GOME/SBUV total ozone observations, *J. Geophys. Res.*, *111*, D12308, doi:10.1029/2005JD006279.
- Candille, G., and O. Talagrand (2005), Evaluation of probabilistic prediction systems for a scalar variable, *Q. J. R. Meteorol. Soc.*, *131*, 2131–2150, doi:10.1256/qj.04.71.
- Chevallier, F. (2015), On the statistical optimality of CO₂ atmospheric inversions assimilating CO₂ column retrievals, *Atmos. Chem. Phys. Discuss.*, *11*, 889–11,923, doi:10.5194/acpd-15-11889-2015.
- Chevallier, F. (2016a), Description of the CO₂ inversion production chain. CAMS report CAMS73_2015S1_D73.1.1_201603, 13 pp. [Available at https://atmosphere.copernicus.eu/sites/default/files/FileRepository/Resources/Documentation/Fluxes/CAMS73_2015SC1_D73.1.3_201603_CO2_Inversion_Production_Chain.pdf].
- Chevallier, F. (2016b), Validation report for the inverted CO₂ fluxes, v15r2. CAMS report CAMS73_2015S1_D73.1.2_201606, 47 pp. [Available at http://atmosphere.copernicus.eu/sites/default/files/FileRepository/Resources/Validation-reports/Fluxes/CAMS73_2015SC1_D73.1.2_201606_Val_Inverted_CO2_Fluxes_v3.pdf].
- Chevallier, F., and C. W. O'Dell (2013), Error statistics of Bayesian CO₂ flux inversion schemes as seen from GOSAT, *Geophys. Res. Lett.*, *40*, 1252–1256, doi:10.1002/grl.50228.
- Chevallier, F., et al. (2010), CO₂ surface fluxes at grid point scale estimated from a global 21-year reanalysis of atmospheric measurements, *J. Geophys. Res.*, *115*, D21307, doi:10.1029/2010JD013887.
- Cogan, A., et al. (2012), Atmospheric carbon dioxide retrieved from the Greenhouse Gases Observing Satellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, *J. Geophys. Res.*, *117*, D21301, doi:10.1029/2012JD018087.

- Connor, B., et al. (2016), Quantification of uncertainties in OCO-2 measurements of XCO₂: Simulations and linear error analysis, *Atmos. Meas. Tech.*, *9*, 5227–5238, doi:10.5194/amt-9-5227-2016.
- Cressot, C., et al. (2014), On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, *Atmos. Chem. Phys.*, *14*, 577–592, doi:10.5194/acp-14-577-2014
- Crevoisier, C., M. N. Gloor, E. Gloaguen, L. W. Horowitz, Jorge Sarmiento, C. Sweeney, and P. P. Tans (2006), A direct carbon budgeting approach to infer carbon sources and sinks. Design and synthetic application to complement the NACP observation network, *Tellus*, *58*(5), doi:10.1111/j.1600-0889.2006.00214.x366-375.
- de Maziere, M., M. K. Sha, F. Desmet, C. Hermans, F. Scolas, N. Kumps, J.-M. Metzger, V. Duflot, and J.-P. Cammas (2014), TCCON data from Réunion Island (RE), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.reunion01.R0/1149288.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. R. Meteorol. Soc.*, *131*, 3385–3396, doi:10.1256/qj.05.108.
- Deutscher, N. M., J. Notholt, J. Messerschmidt, C. Weinzierl, T. Warneke, C. Petri, P. Gripe, and K. Katrynski (2014), TCCON data from Bialystok (PL), Release GGG2014R1. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.bialystok01.R1/1183984.
- Dubey, M., et al. (2014), TCCON data from Manaus (BR), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.manaus01.R0/1149274.
- Eldering, A., et al. (2017), The Orbiting Carbon Observatory-2: First 18 months of science data products, *Atmos. Meas. Tech.*, *10*, 549–563, doi:10.5194/amt-10-549-2017.
- Feist, D. G., S. G. Arnold, N. John, and M. C. Geibel (2014), TCCON data from Ascension Island (SH), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.ascension01.R0/1149285.
- Frankenberg, C., S. S. Kulawik, S. Wofsy, F. Chevallier, B. Daube, E. A. Kort, C. O'Dell, E. T. Olsen, and G. Osterman (2016), Using airborne HIAPER Pole-to-Pole Observations (HIPPO) to evaluate model and remote sensing estimates of atmospheric carbon dioxide, *Atmos. Chem. Phys.*, *16*, 7867–7878, doi:10.5194/acp-16-7867-2016.
- Goo, T.-Y., Y.-S. Oh, and V. A. Velazco (2014), TCCON data from Anmeyondo, South Korea, release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.anmeyondo01.R0/1149284.
- Griffith, D. W. T., et al. (2014a), TCCON data from Wollongong (AU), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.wollongong01.R0/1149291.
- Griffith, D. W. T., et al. (2014b), TCCON data from Darwin (AU), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.darwin01.R0/1149290.
- Hammerling, D. M., A. M. Michalak, C. O'Dell, and S. R. Kawa (2012), Global CO₂ distributions over land from the Greenhouse Gases Observing Satellite (GOSAT), *Geophys. Res. Lett.*, *39*, L08804, doi:10.1029/2012GL051203.
- Hase, F., T. Blumenstock, S. Dohe, J. Gross and M. Kiel (2014), TCCON data from Karlsruhe (DE), release GGG2014R1. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.karlsruhe01.R1/1182416.
- Iraci, L., et al. (2014), TCCON data from Edwards (US), release GGG2014R1. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.edwards01.R1/1255068.
- Katzfuss, M., and N. Cressie (2012), Bayesian hierarchical spatio-temporal smoothing for very large datasets, *Environmetrics*, *23*, 94–107, doi:10.1002/env.1147.
- Kawakami, S., H. Ohyama, K. Arai, H. Okumura, C. Taura, T. Fukamachi, and M. Sakashita (2014), TCCON data from Saga (JP), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.saga01.R0/1149283.
- Kiemle, C., M. Quatrevalet, G. Ehret, A. Amediek, A. Fix, and M. Wirth (2011), Sensitivity studies for a space-based methane lidar mission, *Atmos. Meas. Tech.*, *4*, 2195–2211, doi:10.5194/amt-4-2195-2011.
- Kivi, R., P. Heikkinen, and E. Kyrö (2014), TCCON data from Sodankylä (FI), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.sodankyla01.R0/1149280.
- Kulawik, S. S., et al. (2016), Consistent evaluation of GOSAT, SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON. *Atmos. Meas. Tech.*, *9*, 683–709, doi:10.5194/amt-9-683-2016.
- Massart, S., A. Agustí-Panareda, I. Aben, A. Butz, F. Chevallier, C. Crevoisier, R. Engelen, C. Frankenberg, and O. Hasekamp (2014), Assimilation of atmospheric methane products in the MACC-II system: From SCIAMACHY to TANSO and IASI, *Atmos. Chem. Phys.*, *14*, 6139–6158, doi:10.5194/acp-14-6139-2014.
- Morino, I., T. Matsuzaki, and A. Shishime (2014a), TCCON data from Tsukuba (JP), 125HR, release GGG2014R1. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.tsukuba02.R1/1241486.
- Morino, I., N. Yokozeki, T. Matzuzaki, and M. Horikawa (2014b), TCCON data from Rikubetsu (JP), release GGG2014R1. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.rikubetsu01.R1/1242265.
- NASA (1986), Report of the EOS data panel, Earth Observing System, data and information system, data panel report, vol. Ila., NASA Technical Memorandum 87777, June 1986, 62 pp. [Available at <http://hdl.handle.net/2060/19860021622>.]
- Nguyen, H., M. Katzfuss, N. Cressie, and A. Braverman (2014a), Spatio-temporal data fusion for very large remote sensing datasets, *Technometrics*, *56*(2), 174–185, doi:10.1080/00401706.2013.831774.
- Nguyen, H., G. Osterman, D. Wunch, C. O'Dell, L. Mandrake, P. Wennberg, B. Fisher, and R. Castano (2014b), A method for collocating satellite XCO₂ data to ground-based data and its application to ACOS-GOSAT and TCCON, *Atmos. Meas. Tech.*, *7*, 2631–2644, doi:10.5194/amt-7-2631-2014.
- Notholt, J., C. Petri, T. Warneke, N. Deutscher, M. Buschmann, C. Weinzierl, R. Macatangay, and P. Gripe (2014), TCCON data from Bremen (DE), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.bremen01.R0/1149275.
- O'Dell, C. W., et al. (2012), The ACOS CO₂ retrieval algorithm—Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, *5*, 99–121, doi:10.5194/amt-5-99-2012.
- Olsen, S. C., and J. T. Randerson (2004), Differences between surface and column atmospheric CO₂ and implications for carbon cycle research, *J. Geophys. Res.*, *109*, D02301, doi:10.1029/2003JD003968.
- Pierangelo, C., et al. (2016), MERLIN (Methane Remote Sensing Lidar Mission): An overview, *EPJ Web Conf.*, *119*(2016), 26001, doi:10.1051/epjconf/201611926001.
- Rayner, P., A. M. Michalak, and F. Chevallier (2016), Fundamentals of data assimilation, *Geosci. Model Dev. Discuss.*, doi:10.5194/gmd-2016-148.
- Sherlock, V., B. Connor, J. Robinson, H. Shiona, D. Smale, and D. Pollard (2014), TCCON data from Lauder (NZ), 125HR, release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.lauder02l.R0/1149298.
- Strong, K., J. Mendonca, D. Weaver, P. Fogal, J. R. Drummond, R. Batchelor, and R. Lindenmaier (2014), TCCON data from Eureka (CA), release GGG2014R0. TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge Natl. Lab., Oak Ridge, Tenn., doi:10.14291/tcon.ggg2014.eureka01.R0/1149271.

- Sussmann, R., and M. Rettinger (2014), TCCON data from Garmisch (DE), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.garmisch01.R0/1149299.
- Talagrand, O., R. Vautard, and B. Strauss (1999), Evaluation of probabilistic prediction systems, in *Proceeding of Workshop on Predictability*, pp. 1–25, October 1997. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, U. K. [Available at <http://www.ecmwf.int/sites/default/files/elibrary/1997/12555-evaluation-probabilistic-prediction-systems.pdf>]
- Talagrand, O. (2014), Errors. A posteriori diagnostics, in *Advanced Data Assimilation for Geosciences Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012*, edited by É. Blayo et al., pp. 229–254, Oxford Univ. Press, Oxford.
- Té, Y., P. Jeseck, C. Janssen (2014), TCCON data from Paris (FR), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.paris01.R0/1149279.
- Todling, R. (2015), A complementary note to 'A lag-1 smoother approach to system-error estimation': The intrinsic limitations of residual diagnostics, *Q. J. R. Meteorol. Soc.*, *141*, 2917–2922, doi:10.1002/qj.2546.
- Warneke, T., et al. (2014), TCCON data from Orleans (FR), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.orleans01.R0/1149276.
- Wennberg, P. O., C. Roehl, D. Wunch, G. C. Toon, J.-F. Blavier, R. Washenfelder, G. Keppel-Aleks, N. Allen, and J. Ayers (2014a), TCCON data from Park Falls (US), release GGG2014R0. TCCON data archive, hosted by CDIAC, doi:10.14291/tcon.ggg2014.parkfalls01.R0/1149161.
- Wennberg, P. O., D. Wunch, C. Roehl, J.-F. Blavier, G. C. Toon, N. Allen, P. Dowell, K. Teske, C. Martin, and J. Martin (2014b), TCCON data from Lamont (US), release GGG2014R1, TCCON data archive, hosted by the Carbon Dioxide Information Analysis Center, Oak Ridge Natl. Lab., Oak Ridge, Tenn., doi:10.14291/tcon.ggg2014.lamont01.R1/1255070.
- Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, P. O. Wennberg (2011a), The total carbon column observing network. *Phil. Trans. R. Soc. A*, *369*, doi:10.1098/rsta.2010.0240.
- Wunch, D., et al. (2011b), A method for evaluating bias in global measurements of CO₂ total columns from space, *Atmos. Chem. Phys.*, *11*, 12,317–12,337, doi:10.5194/acp-11-12317-2011.
- Wunch, D., et al. (2017), Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO₂} measurements with TCCON, *Atmos. Meas. Tech.*, *10*, 2209–2238, doi:10.5194/amt-10-2209-2017.
- Zeng, Z. C., et al. (2017), Global land mapping of satellite-observed CO₂ total columns using spatio-temporal geostatistics, *Int. J. Digital Earth*, *10*(4), 426–456, doi:10.1080/17538947.2016.1156777.
- Zhou, M., B. Dils, P. Wang, R. Detmers, Y. Yoshida, C. W. O'Dell, D. G. Feist, V. A. Velazco, M. Schneider, and M. De Mazière (2016), Validation of TANSO-FTS/GOSAT XCO₂ and XCH₄ glint mode retrievals using TCCON data from near-ocean sites, *Atmos. Meas. Tech.*, *9*, 1415–1430, doi:10.5194/amt-9-1415-2016.