Atmos. Chem. Phys., 13, 3205–3225, 2013
www.atmos-chem-phys.net/13/3205/2013/
doi:10.5194/acp-13-3205-2013

Atmospheric Chemistry and Physics Open Access

Comparison of improved Aura Tropospheric Emission Spectrometer CO_2 with HIPPO and SGP aircraft profile measurements

S. S. Kulawik[1], J. R. Worden[1], S. C. Wofsy[2], S. C. Biraud[3], R. Nassar[4], D. B. A. Jones[5], E. T. Olsen[1], R. Jimenez[6], S. Park[7], G. W. Santoni[2], B. C. Daube[2], J. V. Pittman[2], B. B. Stephens[8], E. A. Kort[1], G. B. Osterman[1], and TES team

[1]Jet Propulsion Laboratory, California Institute of Technology, CA, USA
[2]Abbott Lawrence Rotch Professor of Atmospheric and Environmental Chemistry School of Engineering and Applied Science and Department of Earth and Planetary Science Harvard University, Cambridge, MA, USA
[3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[4]Environment Canada, Toronto, Ontario, Canada
[5]Department of Physics, University of Toronto, Tonronto, Canada
[6]Air Quality Research Group, Department of Chemical and Environmental Engineering, Universidad Nacional de Colombia, Bogota, DC 111321, Colombia
[7]Department of Oceanography, College of Ecology and Environmental Science, Kyungpook National University, Daegu, South Korea
[8]National Center for Atmospheric Research, Boulder, CO, USA

Correspondence to: S. S. Kulawik (susan.kulawik@jpl.nasa.gov)

Received: 18 January 2012 – Published in Atmos. Chem. Phys. Discuss.: 29 February 2012
Revised: 23 February 2013 – Accepted: 25 February 2013 – Published: 18 March 2013

Abstract. Thermal infrared radiances from the Tropospheric Emission Spectrometer (TES) between 10 and 15 µm contain significant carbon dioxide (CO_2) information, however the CO_2 signal must be separated from radiative interference from temperature, surface and cloud parameters, water, and other trace gases. Validation requires data sources spanning the range of TES CO_2 sensitivity, which is approximately 2.5 to 12 km with peak sensitivity at about 5 km and the range of TES observations in latitude (40° S to 40° N) and time (2005–2011). We therefore characterize Tropospheric Emission Spectrometer (TES) CO_2 version 5 biases and errors through comparisons to ocean and land-based aircraft profiles and to the CarbonTracker assimilation system. We compare to ocean profiles from the first three Hiaper Pole-to-Pole Observations (HIPPO) campaigns between 40° S and 40° N with measurements between the surface and 14 km and find that TES CO_2 estimates capture the seasonal and latitudinal gradients observed by HIPPO CO_2 measurements. Actual errors range from 0.8–1.8 ppm, depending on the campaign and pressure level, and are approximately 1.6–2 times larger than the predicted errors. The bias of TES versus HIPPO is within 1 ppm for all pressures and datasets; however, several of the sub-tropical TES CO_2 estimates are lower than expected based on the calculated errors. Comparisons to land aircraft profiles from the United States Southern Great Plains (SGP) Atmospheric Radiation Measurement (ARM) between 2005 and 2011 measured from the surface to 5 km to TES CO_2 show good agreement with an overall bias of −0.3 ppm to 0.1 ppm and standard deviations of 0.8 to 1.0 ppm at different pressure levels. Extending the SGP aircraft profiles above 5 km using AIRS or CONTRAIL measurements improves comparisons with TES. Comparisons to CarbonTracker (version CT2011) show a persistent spatially dependent bias pattern and comparisons to SGP show a time-dependent bias of $-0.2\,\mathrm{ppm\,yr^{-1}}$. We also find that the predicted sensitivity of the TES CO_2 estimates is too high, which results from using a multi-step retrieval for CO_2 and temperature. We find that the averaging kernel in the TES product corrected by a pressure-dependent factor accurately reflects the sensitivity of the TES CO_2 product.

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

1 Introduction

Over the past decade, measurements of carbon dioxide (CO_2) from space have become increasingly prevalent, with CO_2 measurements from SCIAMACHY, AIRS, TES, IASI, ACE, and GOSAT (e.g. Reuter et al., 2011; Chahine et al., 2008; Kulawik et al., 2010; Crevosier et al., 2009; Foucher et al., 2011; Yoshida et al., 2011; Crisp et al., 2012; Butz et al., 2011). Robust calculation of errors in the CO_2 estimates is critical because errors in interferences can be larger than the expected variability. There is also a need to understand and validate biases and errors with great accuracy for the data to be useful for estimating CO_2 sources and sinks. Consistent validation and intercomparisons for satellite data, necessary for combining or utilizing multiple satellite results, are challenging since the different products have different coverage, vertical sensitivity, and averaging strategies (as summarized in Table 1). In this paper, we present comparisons of TES CO_2 to aircraft profile data from the HIPPO campaigns and from the Southern Great Plains ARM site to quantify errors, biases, and correlations between TES and the validation data. The techniques and methods shown in this paper are applicable to validation of other instruments with coincident aircraft profiles.

Multiple studies have estimated the precision and bias required to utilize atmospheric CO_2 measurements for source and sink estimates. Using simulated observations, Rayner and O'Brien (2001) showed that satellite measurements of CO_2 total column abundances with a precision of 2.5 ppm, averaged monthly on spatial scales of $8^\circ \times 10^\circ$, would offer more information on CO_2 fluxes than can be obtained from the existing surface network. Houweling et al. (2004) also carried out simulations suggesting that latitude-dependent biases of less than 0.3 ppm are necessary for upper tropospheric CO_2 data to be useful for estimating sources and sinks. Nassar et al. (2011) showed that $5^\circ \times 5^\circ$ monthly-averaged TES observations at 500 hPa (about 5.5 km altitude) over ocean with mean errors of 4.7 ppm between 40° S and 40° N provided information that was complementary to flask data and especially helped constrain tropical land regions. Nassar et al. (2011) mitigated latitude and seasonally dependent biases of 1–2 ppm using 3 different correction methods to estimate sources and sinks from combined TES mid-tropospheric CO_2 and surface flask CO_2. Although the exact magnitude of regional fluxes differed based on the bias correction approach used, key results are generally robust within the predicted errors. Thus assessing the robustness of flux estimates with spatially or temporally varying biases is possible; however, smaller biases are of course preferable.

Kulawik et al. (2010) showed that the TES CO_2 prototype results compared well to aircraft data over Northern Hemisphere ocean sites but showed less reliable results over Southern Hemisphere ocean sites in some months and over land. The peak sensitivity of TES CO_2 was seen to be near 500 hPa with sensitivity between approximately 40° S and 45° N. Based on the findings of Kulawik et al. (2010), updates were made to the retrieval strategy which significantly improved the accuracy of the TES CO_2 retrieval over land and changed the overall bias of TES CO_2 from a 1.8 % to a 0.3 % low bias. Results with the new version, processed with the TES v5 production code, are shown in this paper. The TES CO_2 netcdf "lite" products, on 14 pressure levels, were used for this analysis, available through links from the TES website, at http://tesweb.jpl.nasa.gov/data/. Special runs, e.g. processed with a constant initial guess and prior, were run using the TES prototype code which has minor differences from the v5 TES production code. These runs are used to assess the linearity of the retrieval system and validate the vertical sensitivity of the CO_2 estimates.

2 Measurements

2.1 The TES instrument

TES is on the Earth Observing System Aura (EOS-Aura) satellite and makes high spectral resolution nadir measurements of thermal infrared emission (660 cm^{-1} to 2260 cm^{-1}, with unapodized resolution of 0.06 cm^{-1}, apodized resolution of 0.1 cm^{-1}). TES was launched in July 2004 in a sun-synchronous orbit at an altitude of 705 km with an equatorial crossing time of 13:38 (local mean solar time) and with a repeat cycle of 16 days. In standard "global survey" mode, 2000–3000 observations are taken every other day (Beer, 2006). CO_2 is estimated for TES observations between 40° S and 45° N. In 2006, TES averaged 1570 "global survey" observations per day. Of these, 743 per day are between 40° S and 45° N, and 505 per day have cloud < 0.5 optical depth (OD) and are of good quality. There are additional targeted "special observations", which are not used in this analysis as they are less spatially and temporally uniform. TES global survey observations were consistently taken from late 2004 through June, 2011. For details on the TES instrument, see Beer (2006), and for information on the retrieval methods see Bowman et al. (2006) and Kulawik et al. (2006, 2010).

2.2 HIPPO aircraft measurements

For validation of observations over oceans, we compare to the HIPPO-1, HIPPO-2, and HIPPO-3 campaigns (Wofsy, 2011; Daube et al., 2002; Kort et al., 2011) over the Pacific from 85° N to 67° S for January, 2009, November, 2009, and April 2010, respectively. The profiles are measured between 0.3 km and 9 km (∼307 hPa) with some extending up to 14 km (∼151 hPa), covering a large fraction of the TES vertical sensitivity with data traceable to World Meteorological Organization (WMO) standards with a comparability of approximately 0.1 ppm (Kort et al., 2011). For comparison, the TES mid-Tropospheric averaging kernel, which describes the sensitivity of the CO_2 estimate to variations in CO_2, has a full-width-half-maximum range of 2.5 to 12 km. We select

Table 1. Comparisons between CO_2 datasets.

	launch	spectral region	Peak sens.	day/night	land/ocean	latitude	cloud OD	obs/day	averaging	precis (ppm)
AIRS	2002	TIR	6–9 km	both	both	60° S–90° N	all	∼15 000	≥ 9 observations	2
SCIAMACHY	2002	UV-VIS-IR	column (col)	day	land	80° S–80° N	∼0	< 10 000	5° × 2 month	∼1.4
TES	2004	TIR	5 km	both	both	40° S–40° N	< 0.5	∼500	15° × 1 month	∼1.2
IASI	2006	TIR	11–13 km	both	both	20° S–20° N	clear		5° × 1 month	2.0
GOSAT	2009	near IR TIR	col. 5–7 km	day both	both both	80° S–80° N	< 0.2	∼2000	none	2
						80° S–80° N	∼0	∼2000	none	10
OCO-2	2014	near IR	col.	day	both	80° S–80° N	< 0.2	∼200 000	none	< 2

Summary of coverage, sensitivity, averaging strategies, and errors for several different CO_2 products. The averaging and precision are somewhat subjective estimates, with information provided through communication with Ed Olsen (AIRS), Max Reuter (SCIAMACHY) algorithm (Reuter et al., 2011)), Greg Osterman (GOSAT and OCO-2), Crevosier et al., 2009 (IASI), and Naoko Saitoh with Saitoh et al. (2009). The obs/day are the approximate number of CO_2 estimates which pass quality screening.

all HIPPO measurements for each campaign within a 10° latitude band of a TES observation. If the HIPPO measurements are separated by more than 30 days or 20° longitude, they are split into two groups and each group is averaged. TES measurements for the same latitude range, ± 10° longitude from the HIPPO average longitude, and ± 15 days from the HIPPO mean time are averaged for comparison. The impact of varying the coincidence criteria for time, latitude, and longitude is discussed in Sect. 4.3. We use the profiles identified by the HIPPO team and the CO2.X field, based on 1s data median-filtered to 10s. The CO2.X field is primarily derived from the quantum cascade laser spectrometer (CO2-QCLS) measurement with calibration gaps filled by measurements from the Observations of the Middle Stratosphere (CO2-OMS) instrument. For description of these sensors, see Wofsy (2011) and documentation online (http://www.eol.ucar.edu/projects/hippo and http://hippo.ornl.gov). Note we do not use CO_2 profiles from HIPPO-1 flights 8-11, when these 2 CO_2 instruments received a small fraction of air contaminated by the aircraft cabin. The contaminated measurements showed more than 2 ppm altitude-dependent differences from flask data and a third in situ measurement. Flight 7 CO2.X data have been altitude-adjusted to match the flask data and correct for a small contamination effect of less than 1 ppm. Changes to the aircraft sampling system were made after HIPPO-1 and no contamination was detected thereafter in the reported data.

2.3 SGP aircraft measurements

For validation of observations over land, we compare TES CO_2 to flask sample observations collected bi-weekly over the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site (Ackerman et al., 2004), as part of the ACME project. This site is located in the southern United States at 36.8° N, 97.5°W, and has data starting in 2002. Flask samples are collected, using a small aircraft (Cessna 206), at 12 levels at standard altitudes between 0.3 and 5.3 km altitude with a precision of ± 0.2 ppm (Biraud et al., 2013). Only flask sample measurements with good quality are used in this study (flag = "...*"). Starting in late 2010, coincident aircraft measurements have been coordinated with TES stare observations consisting of up to 32 observations at the same ground location; the stare observations will be analyzed in a future paper.

2.4 CONTRAIL aircraft and AIRS satellite measurements

Because the SGP aircraft measurements cover only part of the altitude range of TES sensitivity to CO_2, we test extending these aircraft measurements with measurements from the Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL) aircraft (Matsueda et al., 2002, 2008; Machida et al., 2008) or co-located Atmospheric Infrared Sounder (AIRS) CO_2 measurements, which have peak vertical sensitivity at ∼9 km (about 325 hPa). The CONTRAIL measurements are between 9 and 11 km (325–250 hPa) and are located over the western Pacific Ocean (between Japan and Australia); these are matched by latitude to the SGP site. For AIRS, the Level 3 calendar monthly v5 product was used with spatial averaging to match the TES spatial averaging.

2.5 CarbonTracker CO_2 model estimates

CO_2 profiles from the CarbonTracker 2011 release (Peters et al., 2007, http://carbontracker.noaa.gov, henceforth CT2011) are used to put TES comparisons to aircraft profiles into spatial and temporal context. The NOAA CarbonTracker CO_2 data assimilation system uses atmospheric CO_2 observations, flux inventories, and an atmospheric transport model to derive optimized estimates of CO_2 fluxes and atmospheric CO_2 distributions. We compare TES and HIPPO results to CT2011 to put the TES comparisons to validation data in perspective within time series and spatial patterns.

3 Description of the TES CO_2 product

3.1 Retrieval strategy

3.1.1 Updates from the previous version

The retrieval strategy for the TES CO_2 estimates was updated from the strategy discussed in Kulawik et al. (2010) to address issues found through validation of the prototype CO_2 data. The previous version compared well to validation data over the Northern Hemisphere ocean, but less well to observations over land and the Southern Hemisphere ocean. Observations over land showed a high bias and higher than expected standard deviation differences compared with aircraft data, and observations over ocean in the Southern Hemisphere showed some latitudinal and seasonal biases (see Kulawik et al., 2010, Figs. 9, 10, and 12). One known issue in the TES retrieval is the spectroscopic inconsistency between the $CO_2\,\nu_2$ and laser bands used for the CO_2 retrieval (Kulawik et al., 2010); consequently a retrieval using both bands simultaneously will have inconsistent biases depending on the relative weights of the two bands.

The laser bands are located between 900 and 1100 cm^{-1}, in a relatively transparent region of the spectrum. We use the two bands centered at 960 cm^{-1} and 1080 cm^{-1}. The laser bands yield the best results when temperature and water profiles are known, and the ν_2 band is essential for constraining temperature and water. So, to address the need for the ν_2 band, but to mitigate the effects of the inconsistent spectroscopy, a 2-step retrieval is used. In the first step, atmospheric temperature, water, ozone, carbon dioxide, surface temperature, cloud optical depth and height, and emissivity (over land) are retrieved for windows covering both the ν_2 and laser bands. This uses the 5-level CO_2 retrieval grid (surface, 511 hPa, 133 hPa, 10 hPa, 0.1 hPa). The 511 hPa result (at about 5.5 km) is biased low by about 6 ppm, with the surface result tending to be biased more than 6 ppm and the 133 hPa result tending to be biased less than 6 ppm. Adding more retrieval levels to this step resulted in increased altitude-dependent biases. The second step retrieves only CO_2 and surface temperature in the 980 cm^{-1} laser band keeping atmospheric temperature, water, etc. from Step 1 and using a 14-level retrieval vector for CO_2 (surface, 909, 681, 511, 383, 287, 215, 161, 121, 91, 51, 29, 4.6, 0.1 hPa).

We found that ozone is a significant interferent in this spectral band and so we now jointly estimate ozone with CO_2. We also found that radiances measured at the 1080 cm^{-1} laser band is significantly affected by a large silicate emissivity feature; we therefore do not use this spectral region in our final retrieval step.2. We also found that extending the window used for the ν_2 band from 671–725 cm^{-1} to 660–775 cm^{-1} improved results because of sensitivity to CO_2 and other jointly retrieved parameters at these augmented wavelengths. Finally, we removed some spectral regions contaminated by minor interferent species, such as formic acid and formaldehyde, as well as spectral regions with unidentified but persistent radiance residual features. Formic acid and formaldehyde typically exist at very low levels in the atmosphere, but appear at significant concentrations in biomass burning plumes, which could lead to spatially dependent biases in CO_2 if their spectral regions are included in the retrieval. The spectral ranges used for the two steps are shown in Table 2. The resulting strategy is implemented in the TES products for v5 data.

Table 2. Spectral windows.

Step 1 TES filter	2 Start (cm^{-1})	End (cm^{-1})
2B1	660.04	775.00
1B2	968.06	1003.28
1B2	1070.000	1100.00
1B2	1110.00	1117.40
Step 2		
1B2	968.06	989.66

The spectral ranges used for TES CO_2 with the filter name characteristic of the TES instrument. These spectral ranges have many narrow spectral regions removed to avoid minor interferent species and persistent spectral residuals. The species included in the forward model were H_2O, CO_2, O_3, HNO_3 for the 2B1 filter and H_2O, CO_2, O_3, CFC-11, CFC-12, NH_3 for the 1B2 filter.

3.1.2 A priori and values and assumptions

We use optimal estimation to infer CO_2 and interfering trace gasses from the TES measured radiances and to provide a robust calculation of the errors (Rodgers 2000; Bowman et al., 2006) and vertical resolution, critical components for using these data for scientific analysis. Because the problem of estimating tropospheric concentrations of CO_2 is ill-posed regularization must be used to distinguish likely estimates from unphysical estimates. This regularization comes in the form of a priori covariances and constructed constraints as well as a priori states around which the solution is regularized.

The a priori covariance and the constraint for the 5-level CO_2 retrieval in Step 1 are described in Kulawik et al. (2010). The constraint for the 14-level CO_2 retrieval in Step 2 was created with a similar process as the 5-level constraint described in Kulawik et al. (2010). This constraint is constructed such that greater variability and uncertainty, but also increased sensitivity, is allowed in the final estimate. An assumption when utilizing this constraint is that an average of multiple solutions is un-biased; this assumption is tested with comparison of the TES CO_2 estimates to the aircraft data. The TES radiative transfer forward model and spectroscopic parameters are the same as in Kulawik et al. (2010).

The TES initial guess and a priori states are taken from the chemical transport model MATCH (Nevison et al., 2008) used in conjunction with a variety of other models to provide

CO_2 surface fluxes based on 2004 (D. Baker, private communication, 2008). The surface CO_2 fluxes are derived from models including the Carnegie-Ames-Stanford Approach (CASA) land biosphere model (Olsen and Randerson, 2004), ocean fluxes from the Wood's Hole Oceanographic Institute (WHOI) model (Moore et al., 2004) and a realistic, annually varying fossil fuel source scheme (Nevison et al., 2008). The CO_2 fields generated by the model compared well to GLOBALVIEW atmospheric CO_2 data (Osterman, TES Design File Memo). The initial guess and a priori are binned averages of the model for every 10° latitude and 180° longitude (i.e. 18 latitude bins and 2 longitude bins (0–180° E, 180° E–360° E)). This binned monthly mean climatology for 2004 was then scaled upward yearly (by 1.0055) to best match the annual increase in CO_2. The initial guess is new for this version; Kulawik et al. (2010) used a constant initial guess.

3.2 Characterizing and validating TES errors and sensitivity

Predicted errors and sensitivity are important to characterize for application of the data to science applications, particularly when errors and sensitivity vary because of variability of clouds and surface properties. The following error analysis (Eq. 1 through 4) is a shortened version from Kulawik et al. (2010). For error analysis and sensitivity characterization, the iterative, non-linear retrieval process is assumed to be represented by the linear estimate (e.g. Rodgers, 2000; Connor et al., 2008):

$$\boldsymbol{x}_{\text{est}} = \boldsymbol{x}_{\text{a}} + \mathbf{A}(\boldsymbol{x}_{\text{true}} - \boldsymbol{x}_{\text{a}}) + \mathbf{G}\boldsymbol{n} + \mathbf{G}\mathbf{K}_b\boldsymbol{\Delta b} \quad (1)$$

where $\boldsymbol{x}_{\text{est}}$ is the logarithm of the estimate, $\boldsymbol{x}_{\text{est}}$, $\boldsymbol{x}_{\text{a}}$, and $\boldsymbol{x}_{\text{true}}$, and are the logarithm of the estimate, a priori constraint vector, and true state, respectively, $\mathbf{A}$ is the averaging kernel (sensitivity of the retrieved state to the true state), $\mathbf{G}$ is the gain matrix (sensitivity of the measurement to radiance errors), $\boldsymbol{n}$ is the radiance error vector, $\mathbf{K}_b$ is the interferent Jacobian (sensitivity of the radiance to each interferent parameter), and $\boldsymbol{\Delta b}$ are the errors in the interferent parameters.

Note that for TES, all parameters besides temperature and emissivity are retrieved in log( ), so that the retrieved parameter, $\boldsymbol{x}$, is the logarithm of the gas volume mixing ratio relative to dry air (VMR). In TES processing, all trace gases are retrieved in log(VMR) because many of the trace gases measured vary logarithmically. TES uses 65 pressures for the radiative transfer pressure grid. The retrieved parameters for CO_2 are on a reduced set of pressures, e.g instead of retrieving 65 CO_2 values, we retrieve 5 in Step 1 (see pressure list in Sect. 3.1.1) and 14 pressures in Step 2 (see list in Table 3). Mapping between pressure grids is discussed in Bowman et al. (2006). Connor et al. (2008) further separates the retrieval vector, $\boldsymbol{x}$, into retrieved CO_2 parameters (here denoted $\boldsymbol{x}$) and all other jointly retrieved parameters (here denoted $\mathbf{y}$).

$$\boldsymbol{x}_{\text{est}} = \boldsymbol{x}_{\text{a}} + \mathbf{A}_{xx}(\boldsymbol{x}_{\text{true}} - \boldsymbol{x}_{\text{a}}) + \mathbf{A}_{xy}(\mathbf{y}_{\text{true}} - \mathbf{y}_{\text{a}}) + \mathbf{G}\boldsymbol{n} + \mathbf{G}\mathbf{K}_b\boldsymbol{\Delta b} \quad (2)$$

Table 3. Sensitivity factor to multiply the averaging kernel row on the CO_2 retrieval pressure grid.

Pressure (hPa)	Ratio
1000.00	0.351038
908.514	0.513463
681.291	0.635048
510.898	0.616426
383.117	0.649254
287.298	0.787116
215.444	1.15804
161.561	1.69716
121.152	2.34417
90.8518	1.99004
51.0896	0.753712
28.7299	0.745675
4.6416	0.365056
0.1000	1.000000

where $\mathbf{A}_{xx}$ is the sub-block of the averaging kernel corresponding to the impact of CO_2 on the retrieved CO_2 parameters, and the $\mathbf{A}_{xy}$ is the sub-block of the averaging kernel corresponding to the impact of non-CO_2 parameters on retrieved CO_2.

Subtracting $\boldsymbol{x}_{\text{true}}$ from the left and right side of Eq. (2) and taking the covariance gives the predicted error covariance:

$$\mathbf{S}_{\text{err}} = \underbrace{\mathbf{G}\mathbf{S}_m\mathbf{G}^T}_{\text{Measurement}} + \underbrace{\mathbf{G}\mathbf{K}_\mathbf{b}\mathbf{S}_{b\,\text{err}}(\mathbf{G}\mathbf{K})^T}_{\text{Interferent}} + \quad (3)$$
$$\underbrace{(\mathbf{I}-\mathbf{A}_{xx})\mathbf{S}_{\text{a},xx}(\mathbf{I}-\mathbf{A}_{xx})^T}_{\text{Smoothing}} + \underbrace{\mathbf{A}_{xy}\mathbf{S}_{\text{a},yy}(\mathbf{I}-\mathbf{A}_{xy})^T}_{\text{Cross state}}$$

where $\mathbf{S}_{\text{err}}$ is the total error covariance, $\mathbf{S}_m$ is the covariance of the radiance error, and $\mathbf{S}_{\text{a}}$ is the a priori covariance. The cross state error (described in Worden et al., 2004; Connor et al., 2008) is the CO_2 error resulting from jointly retrieved species, and the smoothing error results from the effects of the constraint matrix. For more details on the derivation and terms in Eq. (3), see Connor et al. (2008) Sect. 4.1 or Kulawik et al. (2010), Sect. 3.3.

The cross-state component is due to the propagation of error from jointly retrieved species into CO_2; in this case, surface temperature. This error should decrease with averaging over regional and monthly scales, as the surface temperature error will likely vary in sign and magnitude. Similarly, interferent and measurement errors should also decrease with averaging over regional, monthly scales. However, averaging observations with the same CO_2 true state results in a bias for the smoothing term which does not decrease with averaging. The predicted total error covariance for an n observation average is:

$$\mathbf{S}_{\text{err}} = (\mathbf{S}_{\text{meas}} + \mathbf{S}_{\text{int}} + \mathbf{S}_{\text{cross-state}})/n + \mathbf{S}_{\text{smooth}} \quad (4)$$
$$\mathbf{S}_{\text{err}} = \mathbf{S}_{\text{obs}}/n + \mathbf{S}_{\text{smooth}}$$

The observation error ($\mathbf{S}_{\rm obs}$) and smoothing error covariances in Eq. (4) are included in the TES products (Osterman et al., 2009). The predicted error for a particular level is the square-root diagonal of the predicted error covariance at that level, and the off-diagonal terms describe correlated errors between levels. Spectroscopic and calibration errors, which may contribute an additional bias and/or varying error, are not included in Eq. (2), but could be added in, if known, as the gain matrix multiplied by the radiance error.

3.3 Comparisons to aircraft profile data

We validate with aircraft profile data, where the true state, $\boldsymbol{x}_{\rm true}$, is known for at least portions of the atmosphere. To construct $\boldsymbol{x}_{\rm true}$ on the TES pressure levels, the following steps are taken: (1) interpolate/extrapolate the aircraft profile to the 65-level TES pressure grid; (2) replace values below all aircraft measurements with the lowest altitude aircraft measurement value; (3) replace values above all aircraft measurements with the highest altitude aircraft measurement value. We then apply the "Observation operator" to this profile to assess the effects of TES sensitivity (Boxe et al., 2010 Eq. 11):

$$\boldsymbol{x}_{\rm pred} = \boldsymbol{x}_{\rm a} + \mathbf{A}_{\boldsymbol{xx}}(\boldsymbol{x}_{\rm true} - \boldsymbol{x}_{\rm a}) \quad (5)$$

$\boldsymbol{x}_{\rm pred}$ is what TES would see if it observed the air mass described by the aircraft profile in the absence of any other errors due to the vertical resolution and sensitivity of the TES instrument. Since we have applied the TES sensitivity to the aircraft profile, there is no smoothing error term when comparing $\boldsymbol{x}_{\rm pred}$ and $\boldsymbol{x}_{\rm tes}$. The predicted error for $\boldsymbol{x}_{\rm tes}$ compared to $\boldsymbol{x}_{\rm pred}$ is the observation error, which is significantly smaller than the smoothing error when averaging over ∼40 profiles.

The SGP aircraft data go up to ∼5 km, covering only part of the range of TES sensitivity and so the choice of the value for $\boldsymbol{x}_{\rm true}$ above 5 km could have an impact on $\boldsymbol{x}_{\rm pred}$. We set $\boldsymbol{x}_{\rm true}$ above 5 km to carbon dioxide values either from AIRS, CONTRAIL, or the highest altitude aircraft measurement; the differences in these results characterize the size of the uncertainty introduced from uncertainty in the true profile.

3.4 Predicted errors for TES CO_2

Figure 1 shows the predicted errors for a single observation and for a 40-observation average for land and ocean scenes. At 500 hPa (about 5.5 km), the dominant error source for a single observation is interferent error, at about 4–7 ppm, due to, in order of importance, temperature, cloud parameters, water, and ozone. Measurement error is also significant, contributing nearly 4 ppm, followed by the smoothing error, which contributes about 1.5 ppm. Errors from the jointly retrieved surface temperature are small. The total error is about 6–7 ppm for a single observation. However, when 40 observations are averaged, the interferent and measurement errors are taken to be quasi-random, and are reduced by the factor square root of 40. The dominant error for the 40-observation average is the smoothing error, resulting from imperfect sensitivity. Land observations in general have higher interferent error due to the uncertainty in emissivity. In most cases averaging over 1 month, $\pm 5^\circ$ latitude, and $\pm 10^\circ$ longitude gives enough variability in the errors to result in quasi-random errors, which reduces the predicted and actual errors by the square root of the number of observations. We find that measurements taken close together, such as a "stare" special observation, tend to have correlated errors, e.g. from temperature, and averaging does not improve the error. From comparisons to HIPPO data in particular, the quasi-random error assumption is not always valid.

3.5 Predicted sensitivity for TES CO_2

The predicted sensitivity and retrieval non-linearity can be validated, as described in Kulawik et al. (2008, 2010), by running non-linear retrievals using two different a priori vectors, $\boldsymbol{x}_{\rm a}$ and $\boldsymbol{x}'_{\rm a}$, resulting in the iterative, non-linear retrievals, $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{x}}'$, respectively. $\hat{\boldsymbol{x}}'$ is then converted via a linear transformation to use $\boldsymbol{x}_{\rm a}$ using the following linear equation:

$$\hat{\boldsymbol{x}}_{\rm est} = \hat{\boldsymbol{x}}' + \mathbf{A}(\boldsymbol{x}_{\rm a} - \boldsymbol{x}'_{\rm a}) \quad (6)$$

where $\boldsymbol{x}$ is log(VMR). $\hat{\boldsymbol{x}}_{\rm est}$ from Eq. (6) is compared to $\hat{\boldsymbol{x}}$. If they compare within the predicted errors, it validates both the predicted sensitivity and the non-linearity of the system. The comparison between $\hat{\boldsymbol{x}}_{\rm est}$ and $\hat{\boldsymbol{x}}$ answers two questions: (1) how sensitive are the results to the starting point of the retrieval? and (2) can we use the sensitivity to predict the results we expect to see?

The calculated averaging kernel ($\mathbf{A}$) is shown in Fig. 2. The left panel shows $\mathbf{A}$ for all levels for step 1, which includes the joint retrieval of all interferents in both the ν_2 and the laser band spectral regions. $\mathbf{A}$ shows the potential for partially resolving CO_2 at different pressure levels. If the spectroscopy were addressed, the step 1 results and averaging kernel could be used for CO_2, rather than needing to do a final CO_2 step with restricted windows. The middle panel shows the predicted $\mathbf{A}$ for the final CO_2 step. Note that all levels have very similar sensitivity but with more predicted sensitivity than the first step, mainly because the second step only retrieves CO_2 and surface temperature in a narrow spectral range. The right panel compares the averaging kernel row at 511 hPa (about 5.5 km) for TES observations matching the HIPPO campaigns and observations near SGP. Note that the TES observations at SGP, over land, show more variability in the sensitivity because of seasonal and day/night variations in surface temperature. The averaging kernel on the far right panel of Fig. 2 has been corrected by a pressure dependent factor, shown in Table 3, to reflect the actual sensitivity (see Appendix for a description of how the averaging kernel was validated and the pressure dependent factor was calculated.). We find that this ratio is very similar for all pressure levels (results not shown), so that this ratio can be used for all

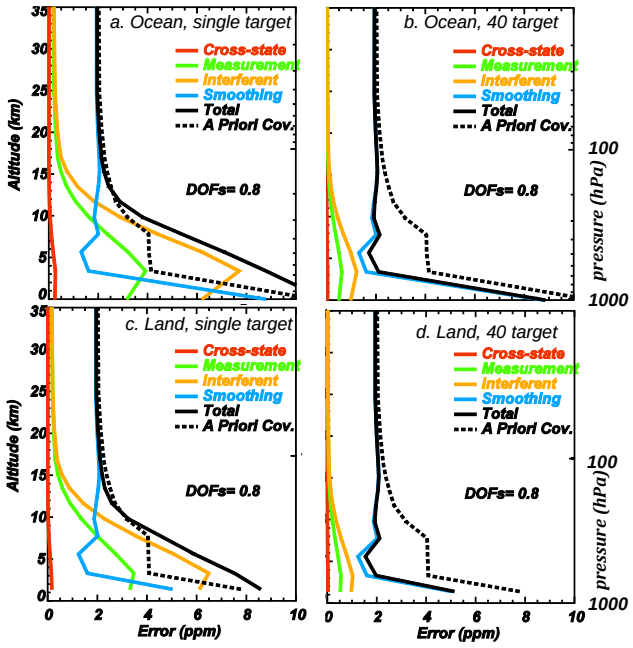

Fig. 1. Errors for an ocean scene (top) and land scene (bottom). Left panels show single observation errors and right panels show errors for 40-observation averages assuming a random distribution of measurement, interferent, and cross-state errors.

retrieval pressures. All remaining results in this paper have this factor applied to the predicted averaging kernel.

3.5.1 Characterizing sensitivity through assimilation

Previous flux estimates using TES CO2 utilized observations at 511 hPa over the ocean (Nassar et al., 2011). We use an observing system simulation experiment (OSSE) to assess the information added by the full profile and by land observations. Using a similar OSSE to that which was used in Kulawik et al. (2010) we found an increase in 1.4 DOF when including TES land results and 0.1 or less DOF change when including all levels versus just the 511 hPa level (at about 5.5 km). The small increase when including all levels is because the averaging kernel row is very similar for all TES pressure levels as seen in Fig. 2. Even though little information seems to be added by using the full profile, 3-D var assimilation of TES profiles (described in Kuai et al., 2013) compared better to validation data than a single level assimilation (unpublished work).

3.6 Bias correction for TES CO_2

Biases are difficult to estimate because of the uncertainty and variability introduced by errors and quality flag choices. A global bias is corrected using the equation $\boldsymbol{x}_{\mathrm{corr},i} = \boldsymbol{x}_{\mathrm{raw},i} + \mathbf{A}_{ij}\, \boldsymbol{bias}_j \boldsymbol{x}_{\mathrm{raw},i}$, as discussed in Kulawik et al. (2010), where $\boldsymbol{bias}_j = -0.0013$ for the prototype results for all j and a bias correction is not applied for v5 production results. Here, $\boldsymbol{x}_{\mathrm{raw}}$ and $\boldsymbol{x}_{\mathrm{corr}}$ are the retrieved and corrected VMR values, respectively (not log(VMR)). The presence of a time-dependent bias was checked using the NOAA CCGCRV fitting software (Thoning et al., 1989; see also http://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html). Fit of years 2005–2009 of monthly averages of TES or SGP aircraft data with the TES observation operator find a difference in the fitted yearly increase of $-0.20\,\mathrm{ppm\,yr^{-1}}$ in the mid-troposphere. A linear fit of the difference of TES and SGP with the TES observation operator found the same trend of $-0.19 \pm 0.09\,\mathrm{ppm\,yr^{-1}}$. A comparison of TES and CT2011 for Southern Hemisphere ocean observations between 20S and 40S found a similar trend in the difference of $-0.27 \pm 0.06\,\mathrm{ppm\,yr^{-1}}$. We correct TES values with a time-dependent bias: $\boldsymbol{x}_{\mathrm{corr},i} = \boldsymbol{x}_{\mathrm{raw},i} + \mathbf{A}_{ij}\, \boldsymbol{bias}_j$, with $\boldsymbol{bias}_j = 0.3^*$ (year-2008). This bias value is the more conservative fit of -0.20 divided by the total averaging kernel row in the mid Troposphere of 0.65. Since after 2010, TES data calibration changed to preserve instrument lifetime (by taking fewer cold space observations), a separate bias correction of 0.0025 (+0.25 %) is applied after 2010. The TES data corrected by the above parameters was re-checked and the trends in the troposphere range from $+0.04\,\mathrm{ppm\,yr^{-1}}$ at the surface to $-0.02\,\mathrm{ppm\,yr^{-1}}$ in the upper Troposphere, which are within the error. The results of these improvements are seen in Fig. 3, with all subsequent analyses using these corrections. The change of CO_2 over time could result from a drift in some aspect of TES calibration or input parameters (e.g. temperature inputs or laser frequency). A drift of $-0.2\,\mathrm{ppm\,yr^{-1}}$ could result from a drift on the order of a $10\,\mathrm{mK\,yr^{-1}}$ drift in brightness temperature. Connor et al. (2011) found no trend in TES brightness temperature, within $5\text{–}10\,\mathrm{mK\,yr^{-1}}$ (Thomas Connor, personal communication). The prototype shows the low bias after 2010, but not the drift in 2005–2009 (note that the prototype runs used TES version 4 inputs for radiance and initial guess values). In Sect. 4.1 and Fig. 9, we see biases which vary by location and are persistent in time. An average of TES data over one year versus the CT2011 model shows a spatial pattern (as seen in Fig. 8) which is persistent from year to year. The difference modified by the averaging kernel can also be used for a location-dependent bias correction, with improvements for all comparisons except for the TES bias versus HIPPO-3. The above bias correction factors (time-dependent, post-2010, and spatially-dependent) will be included for each observation in upcoming TES Lite products.

4 Actual and predicted errors compared with HIPPO and SGP

Figure 4 shows a plot of the matching locations for TES and HIPPO-1, HIPPO-2, HIPPO-3, and SGP. For HIPPO coincidences, both HIPPO and TES are averaged within a box centered around HIPPO locations and times. The mean time for the TES observations must be within 7 days of the HIPPO

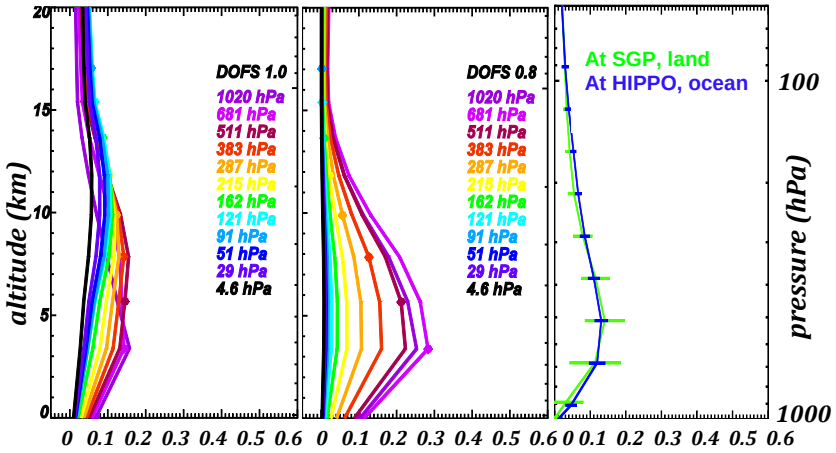

Fig. 2. Averaging kernel for the initial CO_2 step (left), the final CO_2 step (center), and the corrected Averaging Kernel row for 511 hPa for the final CO_2 step (right) for SGP and HIPPO cases. Note that the predicted sensitivity in the lower troposphere is less for the initial step because temperature, water, and cloud properties are jointly retrieved. The FWHM pressures, where the averaging kernel has half the peak value, occur at 750 and 215 hPa (2.5 and 12 km).

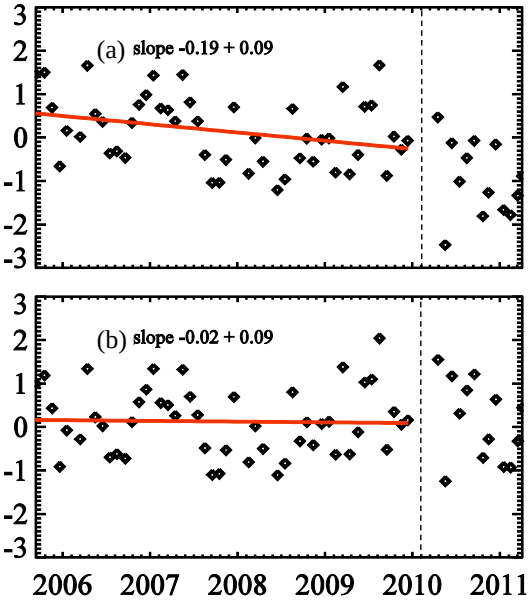

Fig. 3. TES time dependent bias, shown by the difference between TES and SGP with the TES observation operator applied. **(a)** original TES data show a year-dependent trend as well as a bias after 2010 when TES calibration changed. **(b)** corrected TES-SGP, with TES corrected by 0.3 ppm yr^{-1} (multiplied by the TES averaging kernel) and a 0.25 % bias correction after 2010.

average time, and the mean longitude and latitude differences must be less than half of the box width. The time criteria only affects HIPPO-3, as TES was not taking measurements before the first half of this campaign. For SGP comparisons, TES and SGP data are both averaged within each month, and TES is averaged within 10° longitude and 5° latitude of the SGP observations.

4.1 Comparison of TES and HIPPO measurements

Figures 5–7 show the comparisons between TES and HIPPO. Figure 5 shows curtain plots of the comparisons, with the y-axis showing altitude and the x-axis showing latitude. In HIPPO-1, in January 2009, the TES prior is fairly constant with latitude and too high in the Southern Hemisphere, seen in Fig. 6 for the mid-troposphere. The TES results show an improved gradient versus latitude compared with the TES prior, but show a low bias. The correlation in the mid-troposphere is 0.86 ppm with a standard deviation of 0.6 ppm. HIPPO-2, in November 2009, is overall fairly constant within this latitude range with higher values in the Southern Hemisphere and a spread of ∼2 ppm. TES captures the overall pattern but shows anomalously low values at 15S and 10N. The correlation in the mid-troposphere is 0.46 with standard deviation of 1.3 ppm. HIPPO-3, in March–April 2010, has the strongest latitudinal gradient. The TES a priori gradient is again too small with the TES results improving the gradient. However, similar to HIPPO-2, TES has anomalously low values at about 10° N. The TES instrument was not operating during the first half of the campaign resulting in higher errors due to fewer averaged observations. In particular, the TES averages between 20–40° N, where high values are seen, are primarily from observations within a 6 day period, whereas monthly averages are needed to produce uncorrelated errors. The correlation in the mid-troposphere is 0.77 with a standard deviation of 1.8 ppm. For the three HIPPO campaigns, the actual errors are larger than the predicted, by an average factor of 2, likely because the interferent errors are at least somewhat correlated, rather than random (e.g. see Boxe et al., 2010). Consistency between predicted and actual errors is critical for the scientific use of the data, especially data assimilation or CO_2 flux estimates.

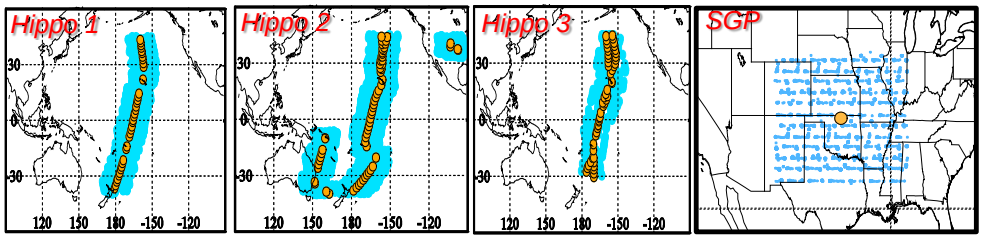

Fig. 4. HIPPO-1, HIPPO-2, and HIPPO-3, SGP and TES coincident observation locations. For HIPPO, each orange dot shows a CO_2 profile location. The blue values show the TES observations which are averaged for comparisons. Note that for plots versus latitude, there can be multiple longitudes or times as seen on the above plots. All TES observations shown are within 10 degrees longitude and 5 degrees latitude of the validation data.

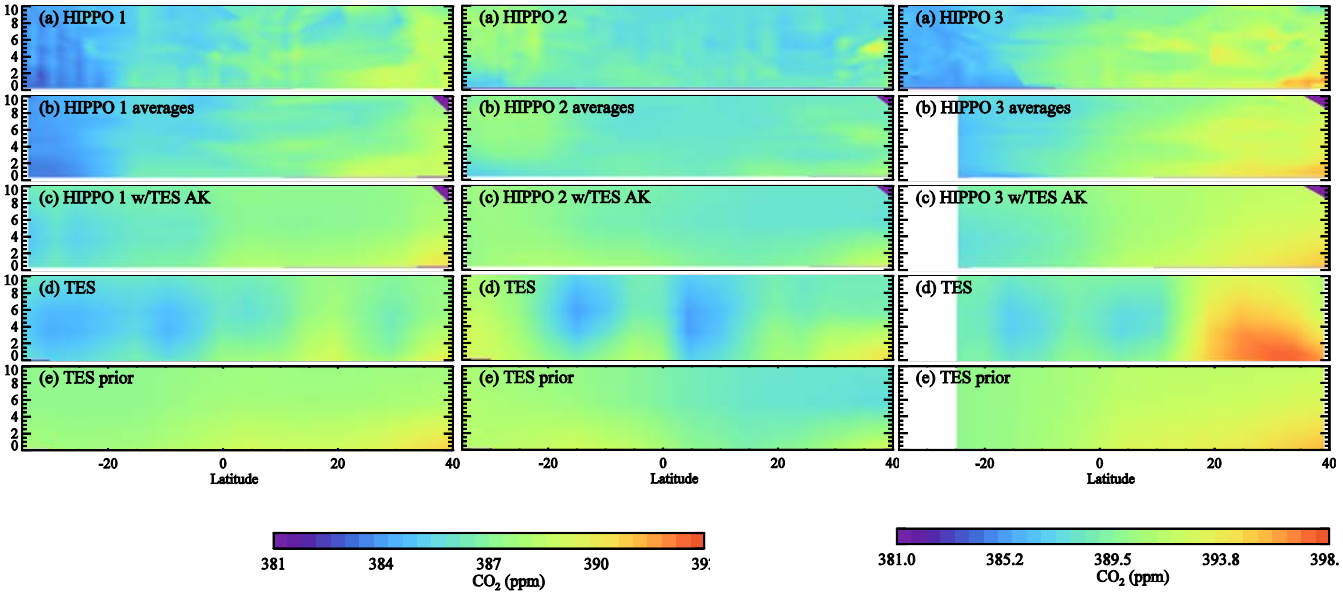

Fig. 5. Curtain plots comparing TES and HIPPO-1 (left), HIPPO-2 (middle) and HIPPO-3 (right) versus latitude. **(a)** shows the HIPPO measurements, **(b)** shows HIPPO measurements averaged over latitude and longitude bins matching TES observations, **(c)** shows HIPPO measurements with the TES observation operator applied, **(d)** shows TES measurements, averaged over the same latitude and longitude bins, and **(e)** shows the TES prior. Data gaps in HIPPO or TES can cause the latitudes to be slightly mismatched. These plots show persistent low features in the TES observations at 15S and 10N and improvements in the CO_2 values (relative to the prior) for HIPPO-1 in particular.

For the HIPPO comparisons, the TES a priori latitudinal gradient is too small with values that are too high in the Southern Hemisphere. The TES retrieved values are generally closer to HIPPO values but with persistent errors larger than the predicted errors seen at ∼15° S and ∼10° N in HIPPO-2 and HIPPO-3. A histogram of the values composing the TES averages for the TES points (not shown) shows that the entire distribution of points is shifted, rather than a few outliers causing the anomalous values. The correlation of errors in a particular region and preliminary analysis of the TES "Stare" observations at SGP indicates that likely these outliers result from a bias in the interferent errors, rather than the assumed quasi-random distribution of interferent error. Since averaging does not reduce a biased error, the error for the averaged product would be comparable to the single-observation interferent error of 4–6 ppm, which is consistent with the errors seen.

Figure 7 shows the TES/HIPPO comparisons in the context of the overall patterns seen by TES monthly averages. In Fig. 7b, the low TES values at ∼10° S and ∼15° N can be seen as part of a larger spatial pattern seen by TES and can be seen in the spatially-dependent bias pattern in Fig. 7d. Looking at the other TES retrieved values, a similar pattern can be seen in TES ozone, water, and HDO at 681 hPa for November 2009 (http://tes.jpl.nasa.gov/visualization/SCIENCE_PLOTS/TES_L3_Monthly.htm). As this pattern is persistent in TES CO_2 from year to year (data not shown) but is not seen with the HIPPO data, it most likely indicates a problem in the retrieved TES CO_2 at these locations. Note that the anomalously high values in HIPPO-3

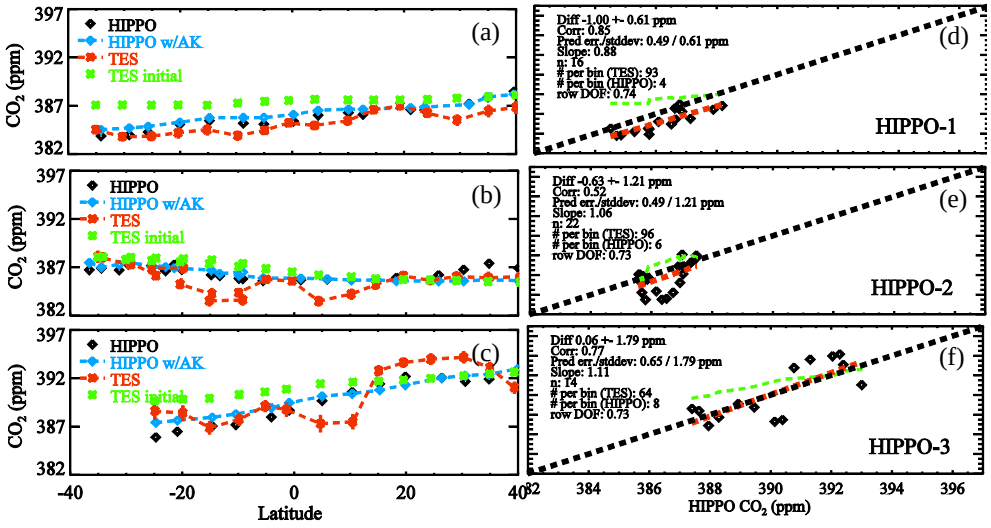

Fig. 6. (a)–(c) Plots versus latitude for the 511 hPa pressure level showing the TES value with error bars (red), HIPPO at the same pressure level (black), HIPPO with the TES observation operator applied (blue), and TES prior and initial guess (green). Results are averaged 5 degrees latitude, and 10 degrees longitude, with TES results within 15 days of each of the HIPPO campaigns. The HIPPO results have the TES observation operator applied to account for TES sensitivity. **(d)–(f)** Correlations between TES and HIPPO observations: the green dashed line is the linear fit for the TES prior, the red line is the fit for the TES results, and the black dashed line shows the ideal 1:1 correlation. The statistical information for **(d)–(f)** is listed in Table 6.

seen in the TES-HIPPO comparisons of Figs. 5–6 are not seen in the complete monthly average (Fig. 7c) for TES for April, 2010 or in the spatially-dependent bias pattern.

Figure 8 shows monthly comparisons between TES and CT2011 for ocean observations in the Pacific averaged between 180° E and 120° E for 10° latitude bands at near-surface (at 908 hPa, about 1 km altitude) and mid-troposphere (at 511 hPa, about 5.5 km altitude). The three HIPPO campaigns are shown as dotted lines showing the best matches from Fig. 5 (from panel c, with TES observation operator). TES results in the Southern Hemisphere deviate from the TES prior, aligning better with CT2011 from 10–30° S and showing seasonal features 30–40° S. Note that TES collected data only to 30° S starting in 2010. Figure 8 shows that locations where TES has a low bias, e.g. 0–10° N or high bias, e.g. 30–40° S, show a constant persistent bias versus HIPPO or CT2011 versus time. This indicates that the spatial pattern seen in Fig. 7d could be corrected in the TES data.

To validate the predicted sensitivity, runs were also performed for HIPPO comparisons using the prototype production code with a uniform 385 ppm a priori and initial guess. We compare the difference between the results obtained with the fixed 385 ppm prior to the variable prior results (also run with the prototype), which are then linearly converted to a uniform 385 ppm prior via Eq. (6). When the differences are smaller or comparable to the observation error, the sensitivity, as described by the averaging kernel, is validated. For HIPPO-1, the TES-TES comparisons (TES results with a variable prior converted to a fixed prior via Eq. (6), versus TES results with a fixed prior) have a 0.02 ppm bias and 0.16 ppm standard deviation compared to observation error of 0.8 ppm. For HIPPO-2, the TES-TES comparisons have a −0.03 ppm bias and 0.34 ppm standard deviation compared to an observation error of 0.6 ppm. For HIPPO-3, the TES-TES comparisons have a −0.45 ppm bias and 1.3 ppm standard deviation compared to an observation error of 0.9 ppm. The bias for all cases is less than the observation error, and in 2 of the 3 cases, the standard deviation difference is less than the observation error. In HIPPO-3, the standard deviation is 0.4 ppm larger than the observation error. This comparison validates the predicted sensitivity and linearity of the retrieval system for ocean observations.

4.1.1 Correlations between TES and HIPPO

Correlations between HIPPO and TES are summarized in Table 4. Because the coincidences are subject to the natural variability of the atmosphere within the coincidence region and time, a correlation of 1.00 is not to be expected. This section also shows how the correlation degrades when the error is comparable to the variability. The correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ (where $\boldsymbol{x}$ and $\boldsymbol{y}$ have mean of 0) is defined as:

$$c_o = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}\sqrt{\boldsymbol{y} \cdot \boldsymbol{y}}} \tag{7}$$

Adding in errors for $\boldsymbol{x}$ and assuming that the errors are uncorrelated with $\boldsymbol{x}$ or $\boldsymbol{y}$, the correlation c is:

Table 4. TES correlations and errors versus all validation data.

Source	Campaign	Pressure (hPa)	Variability (ppm)	Bias (ppm)	Pred. Error (ppm)	Actual Error (ppm)	c	c_o
Prod. code	HIPPO-1	Surf.	1.67	−1.0	0.58	0.73	0.90	0.95
Prod. code	HIPPO-2	Surf.	0.64	−0.1	0.57	1.14	0.57	0.76
Prod. code	HIPPO-3	Surf.	2.65	0.6	0.74	1.78	0.85	0.88
Prod. code	SGP	Surf.	4.75	−0.3	0.54	0.96	0.98	0.99
Prod. code	HIPPO-1	Mid Trop.	1.47	−1.0	0.49	0.61	0.85	0.90
Prod. code	HIPPO-2	Mid Trop.	0.51	−0.6	0.49	1.22	0.50	0.69
Prod. code	HIPPO-3	Mid Trop.	2.43	0.1	0.65	1.79	0.82	0.85
Prod. code	SGP	Mid Trop.	4.38	0.1	0.52	0.80	0.98	0.99
Prototype	SGP – var prior	Surf.	3.54	0.9	0.59	1.28	0.95	0.96
Prototype	SGP – conv. const prior	Surf.	2.26	0.1	0.59	1.26	0.84	0.86
Prototype	SGP – const prior	Surf.	2.26	−0.1	0.55	1.14	0.87	0.90
Prototype	SGP – var prior	Mid Trop.	3.37	0.5	0.57	0.79	0.98	0.99
Prototype	SGP – conv. const prior	Mid Trop.	1.58	0.5	0.57	1.00	0.93	0.99
Prototype	SGP – const prior	Mid Trop.	1.57	0.3	0.54	0.98	0.92	0.97

The calculated correlations, c, between TES, HIPPO, and SGP are shown, as well as the correlations corrected by the degrading effects of errors, c_o, calculated with Eq. (8) for TES near the surface (900 hPa for HIPPO and 880 for SGP) and in the mid-troposphere (511 hPa). "Variability" is the standard deviation of the aircraft data with the TES observation operator and "Bias" is TES – validation data. All results have error-corrected correlations greater than 0.85 except for HIPPO-2, which has the least variability combined with issues in the TES subtropic values. The 6 "prototype" entries, processed with the prototype, compare results when run with variable and a constant prior. The time period for the prototype is mid-2005 to mid-2008 rather than mid-2005 to mid-2011.

$$c = \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\sqrt{\boldsymbol{x}\cdot\boldsymbol{x}+\boldsymbol{\varepsilon}_x\cdot\boldsymbol{\varepsilon}_x}\sqrt{\boldsymbol{y}\cdot\boldsymbol{y}}} = c_o\frac{1}{\sqrt{1+\boldsymbol{\varepsilon}_x^2/\sigma_x^2}} \tag{8a}$$

$$c_o = c\sqrt{1+\boldsymbol{\varepsilon}_x^2/\sigma_x^2} \tag{8b}$$

where the variability of $\boldsymbol{x}$ is denoted σ_x and the error in $\boldsymbol{x}$ is denoted $\boldsymbol{\varepsilon}_x$. From Eq. (8), it is apparent that errors that are equal to or larger than the variability will significantly degrade the observed correlations; for a more detailed discussion of how errors affect correlations, see Zhang et al. (2008). Using the predicted errors, variability, and observed correlations, and using Eq. (8b), we can calculate the underlying correlation in the absence of error, with results shown in Table 4. The raw correlations range from 0.46–0.92, and the correlations corrected for error range from 0.65–0.98 when the predicted error is used in Eq. (8). We find that the observed correlation is lower, as expected, for the HIPPO comparisons which have a lower variability/error ratio, but also highlights that the TES-HIPPO comparisons are not good for HIPPO-2 even considering the predicted error.

4.2 Comparison to aircraft data from the Southern Great Plains (SGP) ARM site

For comparisons between TES and SGP aircraft profile data, both datasets are monthly averaged, and TES is also averaged within 5° latitude, and 10° longitude of the SGP site (see Fig. 3). On the plots, the average of all aircraft data above 2 km is shown in orange labeled "ave SGP" (e.g. Fig. 9a and b). Aircraft profiles with the TES observation operator applied are shown in green labeled "SGP w/obs".

4.2.1 Effects of the validation profile above 5 km

As shown in Fig. 2, TES has significant sensitivity from 1–10 km. Since the aircraft profiles range between the surface to ∼5 km, we test three methods for extending the aircraft profiles to the upper range of TES sensitivity (1) extend the top aircraft value upwards indefinitely, (2) interpolate from the top SGP value to the AIRS value at 9 km, (3) interpolate from the top SGP value to CONTRAIL value at ∼10 km. Fig. 10 shows results for the first two methods. Extension of the SGP aircraft data with AIRS CO_2 values changes the bias from 0.21 to 0.13 ppm, and improves the standard deviation from 1.10 to 0.80 ppm in the mid-troposphere and changes the bias from −0.1 to −0.3 ppm and improves the standard deviation from 1.00 to 0.96 ppm near the surface. The use of CONTRAIL aircraft data to extend the SGP profile improves the standard deviation from 1.10 to 0.82 and increases the bias from 0.21 to 0.45 ppm in the mid-troposphere. Note that the CONTRAIL data are flask measurements taken in the west Pacific matched by latitude to the SGP latitude and are not co-located with the SGP observations. As seen in Fig. 10, all datasets show similar seasonal cycles and yearly increases, with the amplitude on the AIRS cycle somewhat less and with the CONTRAIL data averaging somewhat lower (again note that the CONTRAIL data are at a different longitude). The difference between extending the SGP profile versus AIRS is -0.2 ± 1.4 ppm (AIRS higher) and versus CONTRAIL is 0.5 ± 1.5 ppm at 10 km (CONTRAIL lower).

Since extending SGP with AIRS above the SGP measurements gives somewhat better results, this method will be used to extend the SGP data for the remainder of this paper. This study shows that missing validation data above 5 km results

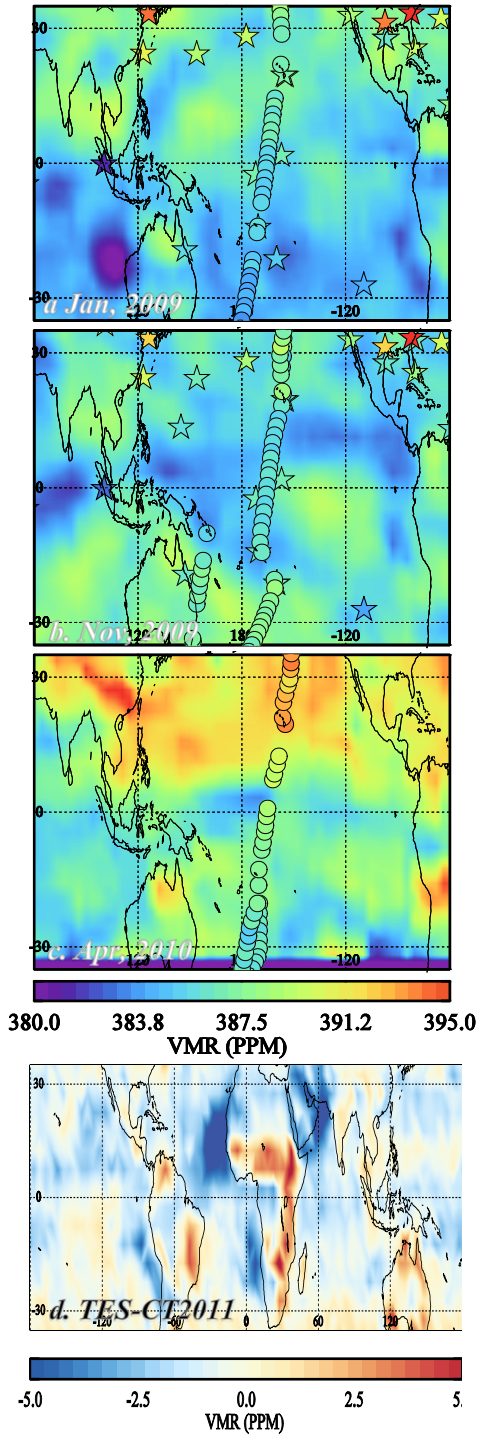

Fig. 7. TES monthly averaged results with a moving box ±5 degrees latitude, ±10 degrees longitude, at 511 hPa. HIPPO values at 5 km measured in each month shown as circles. **(a)** corresponds to HIPPO 1, **(b)** to HIPPO 2, and **(c)** to HIPPO 3. The monthly averaged GLOBALVIEW station values (stars) are shown for context; the surface measurements are not necessarily expected to agree with mid-Tropospheric values. **(d)** shows yearly averages of TES-CT2011 (with TES observation operator) averaged over 5×5 degrees. The pattern seen is persistent over the TES record.

in at least a 0.3 ppm standard deviation difference and a bias uncertainty on the order of 0.3 ppm in the mid-troposphere for the validation data.

4.2.2 Results for different a priori and initial guesses

In this next section we evaluate TES CO_2 using both the standard TES prior, and using a constant prior, i.e. without any a priori knowledge of the CO_2 values. We compare results using these two different priors, linearly transforming the variable prior results to use the constant prior using Eq. (6), to determine whether the TES CO_2 retrieval strategy is linear and that the predicted sensitivity is correct, as well as verify that TES can capture the seasonal and yearly trends in the absence of a priori knowledge of CO_2.

Figure 9 shows time trend comparisons between TES and SGP aircraft measurements. The top two plots show the monthly averages of SGP data above 2 km and the SGP data with the TES observation operator with two a priori choices. The constant a priori choice dampens the expected results in a predictable manner. Since the sum of the row of the 511 hPa averaging kernel for SGP averages about 0.65, about 2/3 of the variability should be captured when using a fixed prior.

Figure 9c to e show results when TES is started at the standard TES initial guess and prior in Fig. 8c, when TES is converted to a fixed prior after retrievals using Eq. (6) in Fig. 8d, and when TES is started at a uniform initial guess and prior in Fig. 9e. To validate sensitivity and retrieval non-linearity, it is important that the results in Fig. 9d and Fig. 9e agree, as discussed in Sect. 3.5. Comparing TES and the validation data, TES shows expected seasonal and yearly patterns over the 4 yr of comparisons, both when TES is started at a "good" initial guess and prior, and when TES uses a uniform initial guess and prior for CO_2 which gives the TES retrieval system no a priori knowledge of CO_2. As seen in Table 4, the correlation between TES and the aircraft is 0.95 at the surface and 0.98 in the free troposphere when a variable prior is used, 0.84 and 0.87 at the surface, and 0.93 and 0.92 in the free troposphere when the results are linearly converted to a constant prior using Eq. (6) or run with a constant prior, respectively. The similarity between the results indicates that the TES CO_2 retrieval strategy is predictably linear and that the predicted sensitivity is correct.

A significant improvement is found over the previous data version (Kulawik et al., 2010), which used a 385 ppm prior and showed a correlation of 0.7, actual rms errors of 1.5 ppm, and a bias of 2.50 ppm. The current results (for the same uniform prior and same analysis period of 2005.5–2008.5) show a correlation of 0.95, actual rms errors of 0.80, and a bias of 0.13 ppm. From this analysis, we expect that the land data in this version are well-characterized with respect to the validation data and are therefore sufficiently reliable for use in scientific analyses.

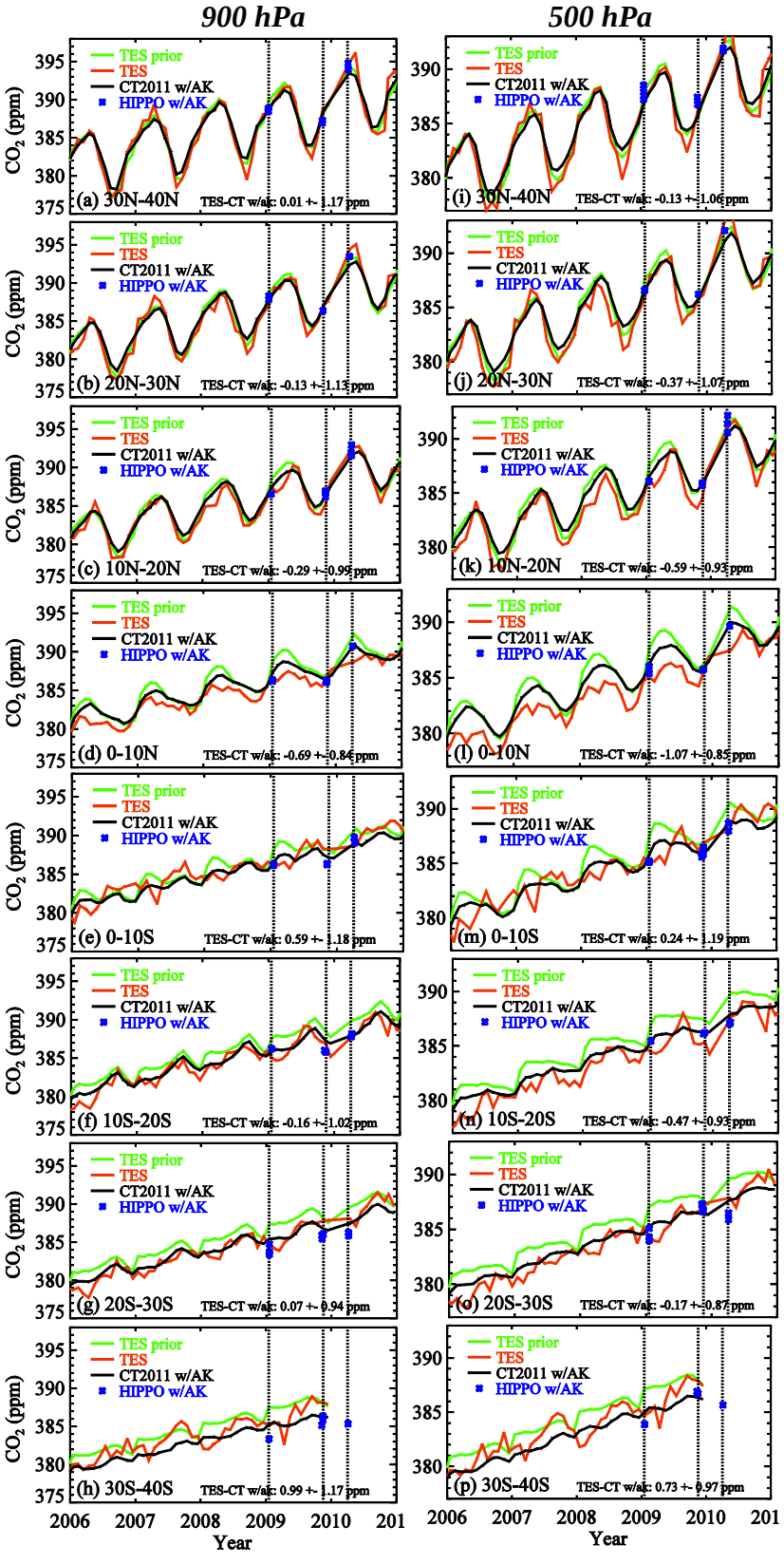

Fig. 8. TES (red) compared to CT2011 with TES observation operator applied (black) with the 3 HIPPO campaign results shown in blue. Latitude bands where TES shows a bias versus CT2011 and HIPPO, e.g. 0–10° N in **(d)** and **(l)**, show persistent low offsets versus time.

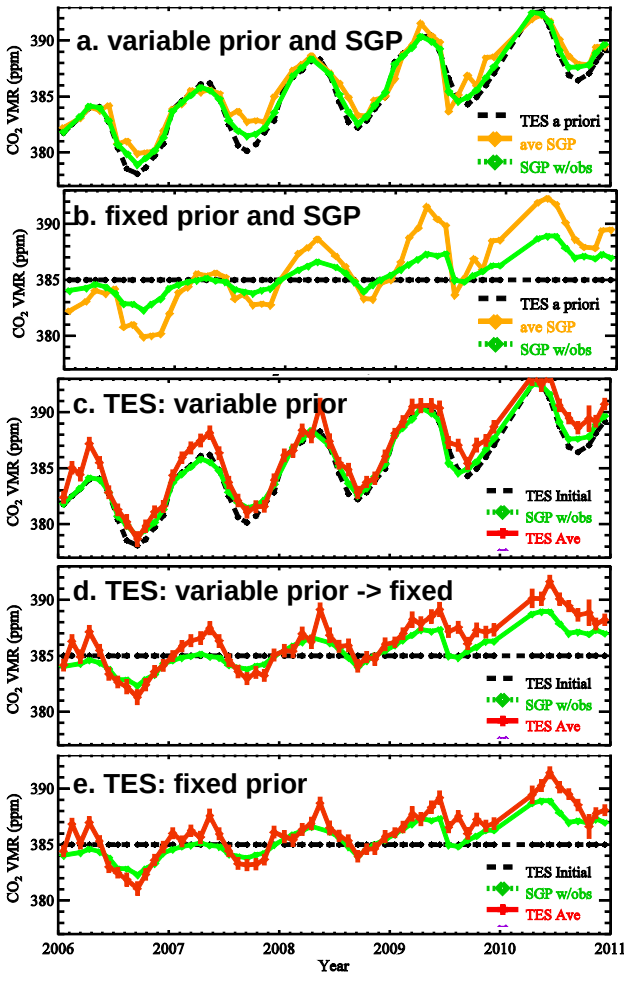

Fig. 9. Time series comparisons at SGP. **(a)** and **(b)** show aircraft data with and without the TES averaging kernel applied, in green and orange respectively. The orange values show all aircraft measurements averaged above 2 km. The green line shows what TES should measure, given the aircraft observations (orange), the TES constraint vector (dashed line), and the TES averaging kernel. The seasonal variability is blunted in a predictable way with the constant constraint vector (**b**). **(c)**, **(d)**, and **(e)** show TES actual measurements (red) versus the aircraft data with the TES averaging kernel applied (green). **(c)** is when TES uses a variable constraint vector. **(d)** is when the TES results from **(c)** are transformed to a constant constraint vector value of 385 ppm using Eq. (6). **(e)** is when the constant constraint vector value of 385 ppm is used for TES nonlinear retrievals. The agreement of **(d)** and **(e)** (with more details shown in Fig. 4) validates the TES averaging kernel. Note: for the first few months of 2010, TES did not collect data.

4.2.3 SGP results for all pressures

Figure 11 shows curtain plots of TES versus SGP between the surface and 10 km. The aircraft measurement at SGP with AIRS observations shown at 8 km are seen in Fig. 11a. The monthly averaged "true" atmosphere, shown in Fig. 11b, is constructed by combining the aircraft measurements (which primarily go up to 5 km) with AIRS at ∼9 km. Figure 11c shows the true state after applying the TES observation operator. Since the different TES pressure levels have similar sensitivity, the variation in altitude in Fig. 11c by pressure is markedly reduced from Fig. 11b. The TES results are seen in Fig. 11d, which agree with the true accounting for TES sensitivity shown in Fig. 11c with 1 ppm standard deviation near the surface and 0.8 ppm standard deviation at 500 hPa (about 5.5 km altitude).

4.3 Coincidence criteria and effect on errors

Differences between the air parcels measured by aircraft and those measured by TES will impart an error in the comparison between these two data sets, but we can reduce this error through averaging. We look at different coincidence criteria to determine the effects on the comparisons. The TES data shown in Fig. 4 are within 5° latitude, 10° longitude, and 15 days of the HIPPO or SGP observations. Because of the range of spatio-temporal locations, we expect that most interferent errors contribute quasi-randomly to the total error budget and scale as the inverse square root of the number of observations. Table 5a shows the effects of averaging within 5°, 10°, or 20° longitude (keeping the latitude and time coincidence specified as above) for near-surface (at 900 hPa for HIPPO and 875 hPa for SGP) and mid-troposphere (at 511 hPa, about 5.5 km above sea level). The actual errors are approximately a factor of 1.6–2 times larger than predicted with the actual errors scaling approximately with the inverse square root of the number of observations. The correlations, predicted errors, and actual errors show consistent improvement between 5°, 10° and 20° averaging indicating that at this scale TES observations are dominated by quasi-random error.

Table 5b shows the results for averaging of the TES data within 15, 30, and 45 days; 2.5°, 5°, and 10°latitude; and cloud cutoffs of 0.1, 0.5, or 1.0 OD. Results shown are averages of the 3 HIPPO results for pressure at 900 hPa and 500 hPa, and SGP results near the surface and mid-troposphere. The predicted errors scale according to the inverse of the square root of observations. Similar to the longitudinal conclusions, there is improvement with increasing numbers of observations, with the exception of cloud OD > 0.5, which resulted in worse comparisons.

4.3.1 Displacement in longitude

TES observes significantly more variability in longitude than simulated by models. To test whether the variability represents variability from the true state versus error, TES coincidences are offset in longitude by 15° east or west before comparing to SGP and HIPPO observations. The longitudinal shift improves results for some cases and worsens the results for other cases. For example, shifting TES coincidences to SGP by 15° east (so that TES observations are

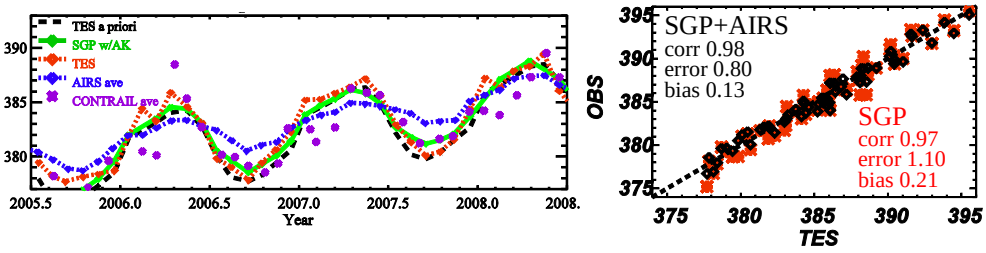

Fig. 10. TES compared to SGP aircraft profile data either extending the aircraft data with the top value or transitioning to AIRS CO_2 measurements at 9 km. Left panel: A time series showing monthly averages for TES (red), AIRS (blue), SGP with the TES observation operator (green), and CONTRAIL aircraft measurements (purple x) for a 3-yr period. Right panel: statistics for SGP and SGP + AIRS results. Adding AIRS in the upper troposphere results in an improvement in the comparison.

Table 5a. Detailed effects of longitude coincidence criteria on results.

	Longitude ± 5					Longitude ± 10					Longitude ± 20				
	corr	*pred*	*actl*	*bias*	*n*	*corr*	*pred*	*actl*	*bias*	*n*	*corr*	*pred*	*actl*	*bias*	*n*
HIPPO-1surf	0.77	0.80	1.16	−0.37	46	0.90	0.58	0.73	−1.04	92	0.86	0.40	0.86	−0.86	196
HIPPO-2surf	0.41	0.82	1.36	−0.08	47	0.59	0.57	1.13	−0.14	97	0.71	0.40	0.92	0.29	202
HIPPO-3surf	0.82	0.99	2.22	0.64	38	0.85	0.74	1.79	0.55	64	0.88	0.56	1.63	0.33	108
SGP-surf	0.96	0.77	1.39	−0.25	49	0.98	0.54	0.96	−0.28	94	0.99	0.30	0.62	−0.62	285
HIPPO-1trop	0.68	0.69	1.07	−0.50	46	0.85	0.49	0.61	−1.00	93	0.78	0.34	0.75	−1.15	197
HIPPO-2trop	0.37	0.71	1.35	−0.58	47	0.50	0.49	1.22	−0.62	96	0.74	0.34	0.92	−0.09	200
HIPPO-3trop	0.74	0.87	2.11	0.19	14	0.77	0.65	1.79	0.06	64	0.81	0.48	1.65	−0.14	108
SGP-trop	0.96	0.72	1.23	0.06	49	0.98	0.52	0.80	0.13	94	0.99	0.29	0.62	−0.23	285
Mean	0.71	0.80	1.49	−0.11	42	0.80	0.57	1.13	−0.29	87	0.85	0.39	1.00	−0.31	198

Calculated correlations ("*corr*"), predicted ("*pred*") and actual ("*actl*") errors and biases (in ppm) when averaging within 5°, 10°, and 20° longitude for each of the datasets at 900 hPa near the surface ("surf") and 500 hPa in the mid-troposphere ("trop"). *n* is the number of TES observations averaged per comparison (important because the error should scale as the inverse of the square root of the number of observations per comparison if errors are uncorrelated and the measurements have the same true). The correlation and errors improve with increasing box size indicating that quasi-random, rather than systematic, errors dominate.

moved towards the east coast) improved comparisons near the surface (from 0.96 to 0.81 ppm standard deviation) and in the mid-troposphere (from 0.80 to 0.75 ppm standard deviation), whereas shifting TES coincidence to SGP by 15° west degraded the comparisons (from 0.96 to 1.09 ppm and from 0.80 to 1.01 ppm standard deviation at the surface and troposphere, respectively). Previous results (Kulawik et al., 2010) did show a statistical improvement with no longitude shift over results with a longitude shift whereas current results do not show a statistical improvement.

4.3.2 Spatial correction

Table 6 shows comparisons for standard coincidence criteria with and without spatial correction (as discussed in Sect. 3.6). Spatial correlation improves actual error for most cases, particularly for SGP comparisons near the surface and in HIPPO-2 comparisons. The bias improved for HIPPO-1 and HIPPO-2, but not for HIPPO-3. The spatial correction as implemented is promising but does not uniformly improve comparisons with validation data and should therefore be used with caution.

4.3.3 Co-location using temperature

A different co-location scheme was tried using atmospheric temperature to define coincidences (Keppel-Aleks et al., 2010) using the criteria developed for ACOS-GOSAT in Wunch et al. (2011). TES observations within 5 days of an SGP observation were selected when satisfying:

$$\left(\left(\frac{\Delta\text{latitude}}{10}\right)^2+\left(\frac{\Delta\text{longitude}}{30}\right)^2+\left(\frac{\Delta\text{Temperature}}{2}\right)^2\right)<1 \quad (9)$$

SGP and TES observations satisfying the above criteria were averaged by month and compared, with results shown in Table 7. Atmospheric temperature coincidences at 500 hPa, rather than the 700 hPa used in Wunch et al. (2011), were used to more closely match TES sensitivity. The coincidence scheme seems marginally better than the standard coincidence criteria shown in Table 6, particularly improving HIPPO correlations and reducing variability in the bias.

Table 5b. Average effects of all coincidence criteria (averaged over all datasets).

	tight criteria					medium					loose criteria				
	corr	*pred*	*actl*	*bias*	*n*	*corr*	*pred*	*actl*	*bias*	*n*	*corr*	*pred*	*actl*	*bias*	*n*
longitude	0.71	0.80	1.49	−0.11	42	0.80	0.57	1.13	−0.29	87	0.85	0.39	1.00	−0.31	198
time	0.75	0.75	1.38	−0.34	50	0.80	0.57	1.13	−0.29	87	0.81	0.48	1.04	−0.30	125
latitude	0.77	0.80	1.45	−0.39	45	0.80	0.57	1.13	−0.29	87	0.84	0.43	0.85	−0.21	161
clouds	0.80	0.72	1.35	−0.14	55	0.80	0.57	1.13	−0.29	87	0.78	0.54	1.23	−0.51	98

Calculated correlations ("*corr*"), predicted ("*pred*") and actual ("*actl*") errors and biases (in ppm) averaged over each of the datasets near the surface at ∼900 hPa and 500 hPa in the mid-Troposphere. *n* is the number of TES observations averaged per comparison (important because the error should scale as the inverse of the square root of the number of observations per comparison if errors are uncorrelated and the measurements have the same true). Results improve with increasing number of observations with the exception of worse results when adding in cases with OD > 0.5.

Table 6. Detailed effects of spatial correction on results.

	TES v5 standard coincidence					TES v5 with spatial bias correction				
	corr	*pred*	*actl*	*bias*	*n*	*corr*	*pred*	*actl*	*bias*	*n*
HIPPO-1surf	0.85	0.49	0.61	−1.00	92	0.88	0.58	0.85	−0.25	92
HIPPO-2surf	0.50	0.49	1.22	−0.62	97	0.51	0.57	1.01	0.43	97
HIPPO-3surf	0.77	0.65	1.79	0.06	64	0.85	0.75	1.78	1.61	64
SGP-surf	0.98	0.54	0.96	−0.28	95	0.98	0.54	0.84	−0.38	95
HIPPO-1trop	0.85	0.49	0.61	−1.00	93	0.87	0.50	0.69	−0.28	93
HIPPO-2trop	0.50	0.49	1.22	−0.62	96	0.24	0.49	0.94	−0.05	96
HIPPO-3trop	0.77	0.65	1.79	0.06	64	0.76	0.65	1.71	1.05	64
SGP-trop	0.98	0.52	0.80	0.13	95	0.99	0.52	0.82	0.01	95
mean	0.78	0.54	1.13	−0.41	87	0.76	0.58	1.08	0.27	87

Similar to Table 5a, comparisons in ppm between TES and validation data with and without spatial bias correction (see Sect. 3.6). Spatial bias correlation improves actual error for SGP comparisons near the surface and in HIPPO-2 comparisons. The bias improved for HIPPO-1 and HIPPO-2, but not for HIPPO-3. Spatial correction is promising but does not uniformly improve comparisons.

5 Conclusions

Table 7. Temperature-based coincident criteria.

	TES v5 temperature-based coincidence				
	corr	*pred*	*actl*	*bias*	*n*
HIPPO-1surf	0.91	0.57	1.26	-0.42	106
HIPPO-2surf	0.50	0.71	1.13	0.19	82
HIPPO-3surf	0.86	0.81	1.59	0.12	75
SGP-surf	0.97	0.46	1.02	-0.43	177
HIPPO-1trop	0.85	0.50	1.24	−0.55	106
HIPPO-2trop	0.71	0.60	1.24	−0.38	83
HIPPO-3trop	0.79	0.68	1.55	−0.27	75
SGP-trop	0.98	0.45	0.92	−0.08	177
mean	0.82	0.60	1.24	−0.23	110

Similar to Tables 5a and 6, comparisons in ppm between TES and validation data using temperature-based coincidence criteria (see Sect. 4.3.3). Results in Table 7 should be compared with the "TES v5 standard coincidence" entries in Table 6, which use a latitude-longitude-time box for coincidence. Temperature-based coincidence improves correlations and reduces the bias and variability of the bias from -0.41 ± 0.47 ppm to -0.23 ± 0.27 ppm.

The improved TES CO_2 estimates described in this work capture the latitudinal gradients and seasonal patterns found in the HIPPO and SGP aircraft data. The comparison with HIPPO and SGP data show biases ≤ 1.0 ppm and errors for monthly-averaged data on the order of 0.8–1.2 ppm. Comparison of HIPPO-3, which averages TES over a partial month, had errors of ∼1.8 ppm. Improvements from the previous TES CO_2 product are remarkable over land, and both land and ocean data for all pressure levels in this version of TES CO_2 can be used for scientific analyses, although sensitivity of all levels is similar and peaks near 500–600 hPa (about 4–5 km). Comparisons of averaged TES to both HIPPO and SGP aircraft profile data show the actual errors averaging ∼1.6–2.0 times the predicted errors for monthly averages. HIPPO and CT2011 comparisons to TES show persistent spatially-dependent biases which can be on the order of the single-observation predicted errors. A correction based on the persistent spatial pattern has been developed and included in the TES Lite CO_2 product, which overall improves comparisons but needs testing in the context of assimilation. We also find a time-dependent bias on the order of

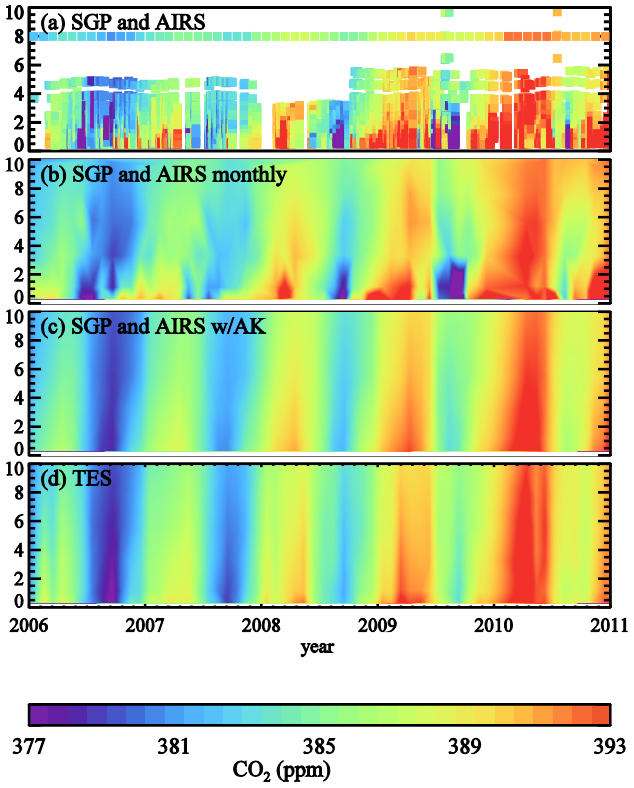

Fig. 11. Curtain plots comparing TES and SGP versus time. **(a)** shows the SGP measurements up to ∼5 km with AIRS at ∼9 km, **(b)** shows monthly averages of **(a)**, **(c)** shows SGP measurements with the TES observation operator applied, **(d)** shows TES measurements. As seen in Fig. 2, the TES results at all pressures have similar sensitivity.

-0.2 ppm yr^{-1} and bias of -0.25 % after 2010, when TES calibration changed. For the SGP dataset, the bias over a 5-yr period is -0.3 ppm at the surface and 0.1 in the troposphere. Because the SGP aircraft data only go up to 5 km, we find that using AIRS values above 5 km improves comparisons. The correlations between TES and validation data are 0.9, 0.5–0.6, and 0.8–0.9 for HIPPO-1, -2, and -3 and 1.0 for SGP comparisons at different TES pressures. We show that lower correlations can partially be explained by a predictable degradation of the correlation due to the error/variability ratio and the observed correlations are consistent with underlying correlations of 0.7–1.0.

In the Appendix, we find that the sensitivity reported in the TES products over-predicts the actual sensitivity because of the 2-step retrieval strategy that is used for the CO_2 estimates. For the current product release, we find that a pressure-dependent multiplicative factor applied to the sensitivity in the TES product results in an accurate prediction for the TES sensitivity to the vertical distribution of CO_2. The TES sensitivity peaks at ∼500 hPa with some sensitivity in the upper troposphere. On average, ocean observations show greater sensitivity than land observations, but both capture seasonal and yearly cycles in CO_2. We validate the sensitivity by comparing results with a very good initial guess and prior compared with a fixed initial guess and prior. Both show the same seasonal and yearly patterns and agree within the observation error when converted to use the same prior using a linear transform.

Averaging TES within 5° latitude, 10° longitude, and 15 days gives the good results when comparing to validation data, considering correlations, errors, and biases. Averaging over larger times or distances improves comparisons and averaging over smaller times or distances results in higher predicted and actual errors as expected. For assimilation, averaging over smaller areas, or no averaging, is fine, however, because the assimilation scheme accounts for the trade-off between number of observations and error (e.g. see Kulawik et al., 2010). When averaging over the above times and distances, the actual errors for comparisons to validation data are 0.8–1.2 ppm, with the predicted errors about a factor of 1.6–2.0 too small for averaged data. TES CO_2 Lite products (available through links from the TES website, at http://tesweb.jpl.nasa.gov/data/) have corrected averaging kernels and errors, with corrections included for spatial and time-dependent biases.

Appendix A

Accurate sensitivity calculation for the multi-step retrieval

Careful characterization of the uncertainties and sensitivity of the CO_2 estimates are critical for using these data in scientific analysis. We find through comparisons of the TES CO_2 retrievals using different a priori vectors that the calculated sensitivity is inconsistent with the actual sensitivity. We calculate the averaging kernel theoretically by propagating perturbations in the true state at every level through the retrieval system and show how effects from retrieved parameters from previous steps need to be included in the calculated averaging kernel.

Considering a single-step retrieval, the effect of a perturbation in the true state on the retrieved state is captured by the following steps:

1. CO_2 is changed by a small amount at a particular pressure level, $\Delta \boldsymbol{x}$
2. This results in a change in the radiance, which is the Jacobian multiplied by the state change: $\Delta \boldsymbol{L} = \mathbf{K}^x \Delta \boldsymbol{x}$
3. When the perturbed radiance is used in a retrieval, it results in a change in the retrieved CO_2: $\Delta \hat{\boldsymbol{x}} = \mathbf{G}^x \Delta \boldsymbol{L}$
4. The averaging kernel, $\mathbf{A}$, is defined as $\Delta \hat{\boldsymbol{x}} / \Delta \boldsymbol{x}$, and this is:

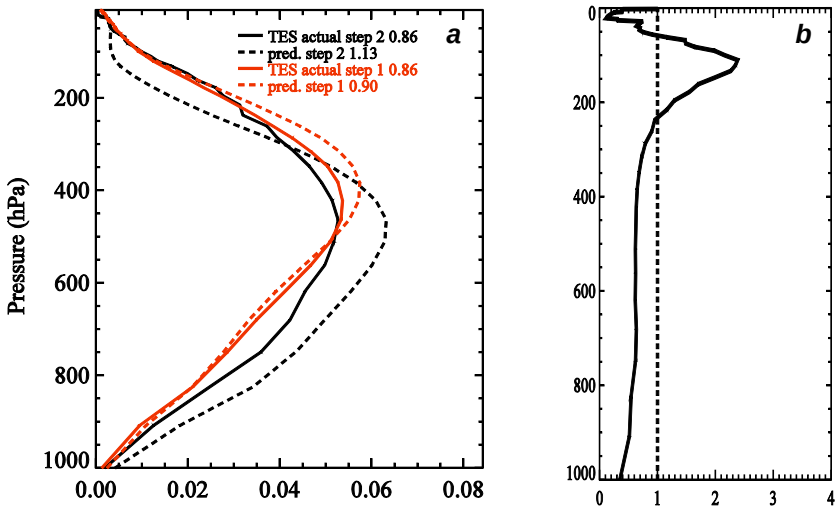

Fig. A1. (a) Step 1 predicted and actual averaging kernel row for 511 hPa. Step 1 actual and predicted are very similar (red), however the step 2 actual averaging kernel (black solid) is significantly different than predicted (black dashed). The peak sensitivity is ∼500 hPa, similar to predicted, however the sensitivity is less, and the sensitivity above 200 hPa follows the Step 1 sensitivity. **(b)** The ratio of actual divided by predicted sensitivity for 511 hPa. This ratio is used to correct the averaging kernel provided in the TES product.

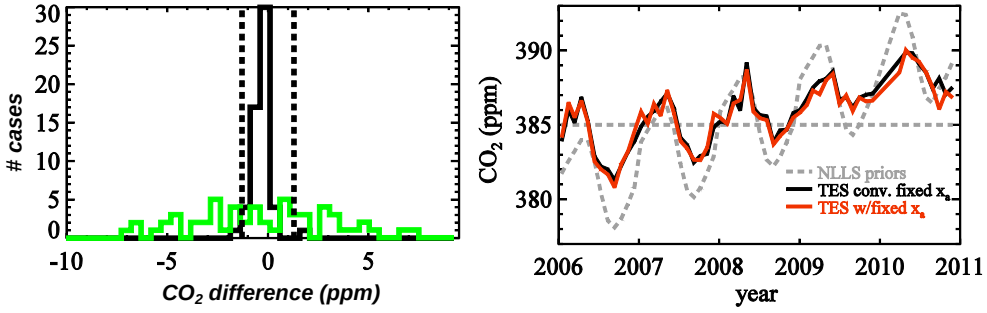

Fig. A2. Validation of the predicted sensitivity and non-linearity using a constant prior versus a variable prior converted to a constant prior via a linear transform following the nonlinear estimate. (left panel) green shows a histogram of the differences of the prior values and black shows a histogram of the result values for TES monthly averages near the SGP site. The dashed lines are the predicted total errors. (right panel) shows a time series, with the fixed prior shown in red and the variable prior converted to a fixed prior shown in black. The gray dashed lines show the two prior values. The red and black lines show excellent agreement, validating the predicted sensitivity and non-linearity.

$$\Delta\hat{\boldsymbol{x}}/\Delta\boldsymbol{x} = \mathbf{G}^x\mathbf{K}_x = \mathbf{A} \quad \text{(A1)}$$

A in Eq. (A1) is the standard equation for the averaging kernel. However, since in our case CO_2 is retrieved in two steps, the radiance perturbation first passes through step 1 of the retrieval. The theoretical averaging kernel resulting from the two-step retrieval, using a similar analysis to the above, but propagating through both steps, is:

$$\Delta\hat{\boldsymbol{x}}/\Delta\boldsymbol{x} = \mathbf{G}_2^x\mathbf{K}_2^x - \mathbf{G}_2^x\mathbf{K}_2^y\mathbf{G}_1^y\mathbf{K}_1^x \quad \text{(A2)}$$

where $\mathbf{G}_2^x$ is the gain matrix for CO_2 for step 2, $\mathbf{K}_2^x$ is the Jacobian for CO_2 for Step 2, $\mathbf{K}_2^y$ is the Jacobian for all non-retrieved and interferent parameters in the step 2 windows, $\mathbf{G}_1^y$ is the gain matrix for all non-CO_2 parameters for step 1, and $\mathbf{K}_1^x$ is the Jacobian for CO_2 parameters for step 1. The first term is the standard averaging kernel predicted for step 2, as seen above, and is included in the TES products, and the second term is a result of both non-CO_2 parameters and CO_2 parameters being active in both step windows. If CO_2 were not active in the step 1 windows, then $\mathbf{K}_1^x$ would be zero, or if temperature, for example, had no influence in the step 2 windows, then $\mathbf{K}_2^y$ would be zero.

We calculate the averaging kernel both from Eq. (A2) and from using perturbations in the CO_2 state vector propagated through the full non-linear least squares (NLLS) retrieval system. For the latter, a retrieval is performed on noise-free simulated radiances with the initial and prior state both set to the true state. Following this, the true state CO_2 at a single pressure is increased by 0.08 %, a new radiance is calculated, and a new retrieval is performed. The change in the retrieved state at 511 hPa divided by the change in the true state at each pressure gives the actual 511 hPa row of the averaging

kernel. Different perturbation strengths were tried, resulting in the same averaging kernel results. Both Eq. (7) and the perturbation approach show enhanced CO_2 sensitivity higher in the atmosphere than predicted in the step 2-only averaging kernel (Eq. A1). Figure A1 shows the TES products **A** and actual **A** (using retrievals on perturbed data) for both steps. Since the TES products do not contain the information to calculate the full averaging kernel using Eq. (A2), the ratio between the actual and predicted sensitivity is calculated as a function of pressure. This scale factor, shown in Fig. A1b and Table 3, is applied to correct the calculated averaging kernel. This corrected averaging kernel is used and validated in this paper.

We compare retrievals using the methodology described near Eq. (6) of the main paper. For our comparisons, $\boldsymbol{x}'_{\mathbf{a}}$ was set to the TES operational a priori value, and $\boldsymbol{x}_{\mathrm{a}}$ was set to a constant 385 ppm value. Note that the initial guess is also set to the same a priori value – the initial guess should not matter but, depending on the non-linearity, can influence the answer. So the constant 385 ppm retrievals contain no prior knowledge of CO_2 in either the initial state or the prior.

The left panel of Fig. A2 shows a histogram of the difference between $\boldsymbol{x}'_{\mathrm{a}}$ and $\boldsymbol{x}_{\mathrm{a}}$ (green), the difference between $\hat{\boldsymbol{x}}_{\mathrm{est}}$ and $\hat{\boldsymbol{x}}$ (black), and the predicted total error (dashed) for monthly average TES values near SGP. The prior distribution has a standard deviation of 3.6 ppm and a bias of -0.86 ppm, and the final difference has a standard deviation of 0.49 ppm and a bias of $+0.11$ ppm. For comparison, the observation error is 0.5 ppm for these monthly averages. The time series panel (Fig. A2) shows that even when the initial guess and prior are set to a constant value, with no a priori knowledge of the CO_2 concentrations, the correct seasonal and yearly cycles are seen with the predicted sensitivity. When this same test was performed with the *original* averaging kernels in the TES products, the analysis showed that the sensitivity was over-predicted because differences between the red and black lines were correlated with the differences between the variable prior and the fixed prior. This correlation indicated that the predicted averaging kernels were not correct, leading to the correction discussed in this section.

Acknowledgements. CONTRAIL data were obtained from Toshinobu Machida at the National Institute for Environmental Studies, Tsukuba, Japan.

SGP aircraft measurements were supported by the Office of Biological and Environmental Research of the US Department of Energy under contract No. DE-AC02-288 05CH11231 as part of the Atmospheric Radiation Measurement Program (ARM), ARM Aerial Facility, and Terrestrial Ecosystem Science Program.

CarbonTracker 2011 results provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov

Work at the Jet Propulsion Laboratory, California Institute of Technology, was performed under a contract with the National Aeronautics and Space Administration and funded through NASA ACMAP, 2007.

The HIPPO campaign was funded by National Science Foundation (NSF) grants to Harvard University, Scripps Institution of Oceanography, and NCAR (ATM-0628575, ATM-0628519, and ATM-0628388) and by several offices and programs of the National Oceanic and Atmospheric Administration: the Atmospheric Composition and Climate Program, the Office of Oceanic and Atmospheric Research, and the Environmental Research Laboratory. The National Center for Atmospheric Research is sponsored by the NSF.

Edited by: C. Gerbig

References

Ackerman, T. P., Genio, A. D. 5 D., Ellingson, R. G., Ferrare, R. A., Klein, S. A., McFarquhar, G. M., Lamb, P. J., Long, C. N., and Verlinde, J.: Atmospheric radiation measurement program science plan: current status and future direcitons of the ARM science program, US Department of Energy, Office of Biological and Environmental Research, Washington, DC, USA, 2004.

Beer, R.: TES on the Aura mission: Scientific objectives, measurements, and analysis overview, IEEE Trans. Geosci. Remote Sens., 44, 1102–1105, 2006.

Biraud, S. C., Torn, M. S., Smith, J. R., Sweeney, C., Riley, W. J., and Tans, P. P.: A Multi-Year Record of Airborne CO_2 observations in the U.S. Southern Great Plains. Atmos. Meas. Tech., 6, 751–763, doi:10.5194/amt-6-751-2013, 2013.

Bowman, K. W., Rodgers, C. D., Kulawik, S. S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Lou, M., Eldering, A., Shephard, M., Worden, H., Lampel, M., Clough, S., Brown, P., Rinsland, C., Gunson, M. and Beer, R.. Tropospheric emission spectrometer: Retrieval method and error analysis, IEEE Trans. Geosci. Remote Sens., 44, 1297–1307, 2006.

Boxe, C. S., Worden, J. R., Bowman, K. W., Kulawik, S. S., Neu, J. L., Ford, W. C., Osterman, G. B., Herman, R. L., Eldering, A., Tarasick, D. W., Thompson, A. M., Doughty, D. C., Hoffmann, M. R., and Oltmans, S. J.: Validation of northern latitude Tropospheric Emission Spectrometer stare ozone profiles with ARC-IONS sondes during ARCTAS: sensitivity, bias and error analysis, Atmos. Chem. Phys., 10, 9901–9914, doi:10.5194/acp-10-9901-2010, 2010.

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.: Toward accurate CO_2 and CH_4 observations from GOSAT, Geophys. Res. Lett., 38, L14812, doi:10.1029/2011GL047888, 2011.

Chahine, M. T., Chen, L., Dimotakis, P., Jiang, X., Li, Q. B., Olsen, E. T., Pagano, T., Randerson, J., and Yung, Y. L.: Satellite remote sounding of mid-tropospheric CO_2, Geophys. Res. Lett., 35, L17807, doi:10.1029/2008GL035022, 2008.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336, 2008.

Connor, T. C., Shephard, M. W., Payne, V. H., Cady-Pereira, K. E., Kulawik, S. S., Luo, M., Osterman, G., and Lampel, M.: Long-term stability of TES satellite radiance measurements, At-

mos. Meas. Tech., 4, 1481–1490, doi:10.5194/amt-4-1481-2011, 2011.

Crevoisier, C., Chédin, A., Matsueda, H., Machida, T., Armante, R., and Scott, N. A.: First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations, Atmos. Chem. Phys., 9, 4797–4810, doi:10.5194/acp-9-4797-2009, 2009.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization, Atmos. Meas. Tech., 5, 687–707, doi:10.5194/amt-5-687-2012, 2012.

Daube, B. C., Boering, K. A., Andrews, A. E., and Wofsy, S. C.: A High-Precision Fast-Response Airborne CO_2 Analyzer for In Situ Sampling from the Surface to the Middle Stratosphere. J. Atmos. Ocean. Technol., 19, 1532–1543, doi:10.1175/1520-0426(2002)019<1532:AHPFRA>2.0.CO;2, 2002.

Foucher, P. Y., Chedin, A., Armante, R., Boone, C., Crevoisier, C., and Bernath, P.. Carbon dioxide atmospheric vertical profiles retrieved from space observation using ACE-FTS solar occultation instrument, Atmos. Chem. Phys., 11, 2455–2470, doi:10.5194/acp-11-2455-2011, 2011.

Houweling, S., Breon, F. M., Aben, I., Rodenbeck, C., Gloor, M., Heimann, M., and Ciais, P.: Inverse modeling of CO_2 sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their performance as a function of space and time, Atmos. Chem. Phys., 4, 523–538, doi:10.5194/amt-4-523-2004, 2004.

Keppel-Aleks, G., Wennberg, P. O., and Schneider, T.: Sources of variations in total column carbon dioxide, Atmos. Chem. Phys., 11, 3581–3593, doi:10.5194/acp-11-3581-2011, 2011.

Kort, E. A., Patra, P. K., Ishijima, K., Daube, B. C., Jiménez, R., Elkins, J., Hurst, D., Moore, F. L., Sweeney, C., and Wofsy, S. C.: Tropospheric distribution and variability of N_2O: Evidence for strong tropical emissions, Geophys. Res. Lett. 38, L15806, doi:10.1029/2011GL047612, 2011.

Kuai, L., Worden, J., Kulawik, S., Bowman, K., Lee, M., Biraud, S. C., Abshire, J. B., Wofsy, S. C., Natraj, V., Frankenberg, C., Wunch, D., Connor, B., Miller, C., Roehl, C., Shia, R.-L., and Yung, Y.: Profiling tropospheric CO_2 using Aura TES and TCCON instruments, Atmos. Meas. Tech., 6, 63–79, doi:10.5194/amt-6-63-2013, 2013.

Kulawik, S. S., Osterman, G., Jones, D. B. A., and Bowman, K. W.: Calculation of altitude-dependent Tikhonov constraints for TES nadir retrievals, IEEE Trans. Geosci. Remote Sens., 44, 1334–1342, 2006.

Kulawik, S. S., Bowman, K. W., Luo, M., Rodgers, C. D., and Jourdain, L.: Impact of nonlinearity on changing the a priori of trace gas profile estimates from the Tropospheric Emission Spectrometer (TES), Atmos. Chem. Phys., 8, 3081–3092, doi:10.5194/acp-8-3081-2008, 2008.

Kulawik, S. S., Jones, D. B. A., Nassar, R., Irion, F. W., Worden, J. R., Bowman, K. W., Machida, T., Matsueda, H., Sawa, Y., Biraud, S. C., Fischer, M. L., and Jacobson, A. R.. Characterization of Tropospheric Emission Spectrometer (TES) CO_2 for carbon cycle science, Atmos. Chem. Phys., 10, 5601–5623, doi:10.5194/acp-11-3581-2011, 2010.

Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.. Worldwide Measurements of Atmospheric CO_2 and Other Trace Gas Species Using Commercial Airlines, J. Atmos. Ocean. Technol., 25, 1744–1754, 2008.

Matsueda, H., Inoue, H. Y., and Ishii, M.: Aircraft observation of carbon dioxide at 8-13 km altitude over the western Pacific from 1993 to 1999, Tellus B – Chem. Phys. Meteorol., 54, 1–21, 2002.

Matsueda, H., Machida, T., Sawa, Y., Nakagawa, Y., Hirotani, K., Ikeda, H., Kondo, N., and Goto, K.: Evaluation of atmospheric CO_2 measurements from new flask air sampling of JAL airliner observation, Pap. Meteorolo. Geophys., 59, 1–17, 2008.

Moore, J. K., Doney, S. C., and Lindsay, K.: Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, Global Biogeochem. Cy., 18, GB4028, doi:10.1029/2004GB002220, 2004.

Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W., Andres, R. J., Suntharalingam, P., Chen, J. M., Brenninkmeijer, C. A. M., Schuck, T. J., Conway, T. J., and Worthy, D. E.: Inverse modeling of CO_2 sources and sinks using satellite observations of CO_2 from TES and surface flask measurements, Atmos. Chem. Phys., 11, 6029–6047, doi:10.5194/acp-11-6029-2011, 2011.

Nevison, C., Mahowald, N., Doney, S., Lima, I., van der Werf, G., Randerson, J., Baker, D., Kasibhatla, P., and McKinley, G.: Contribution of ocean, fossil fuel, land biosphere and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO_2, J. Geophys. Res., 113, G01010, doi:10.1029/2007JG000408, 2008.

Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO_2 and implications for carbon cycle research, J. Geophys. Res.-Atmos., 109, D0230, doi:10.1029/2003JD003968, 2004.

Osterman, G. B. (ed.) Contributors: Bowman, K., Eldering, A., Fisher, B., Herman, R., Jacob, D., Jourdain, L., Kulawik, S., Luo, M., Monarrez, R., Osterman, G., Paradise, S., Payne, V., Poosti, S., Richards, N., D. Rider, Shepard, D., Shepard, M., Vilnrotter, F., Worden, H., Worden, J., Yun, H., and Zhang, L.: Tropospheric Emission Spectrometer TES L2 Data User's Guide, Version 4.0, Pasadena, Jet Propulsion Laboratory/California Institute of Technology, 2009.

Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Helta, D., Meinhardt, F., di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Strom, J., Haszpra, L., Meijer, H. A. J., van der Laan, S., Neubert, R. E. M., Jordan, A., Rodo, X., Morgui, J. A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C. Dolman, A. J., Ciais, P., and Heimann, M.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Global Change Biol. 16, 1317–1337, 2007.

Rayner, P. J. and O'Brien, D. M.: The utility of remotely sensed CO_2 concentration data in surface source inversions, Geophys. Res. Lett., 28, 175–178, 2001.

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

Rodgers, C.: Inverse Methods for Atmospheric Sounding: Theory and Practice, chapter 3, Singapore, World Scientific Publishing Co., 2000.

Saitoh, N., Imasu, R., Ota, Y., and Niwa, Y.: CO_2 retrieval algorithm for the thermal infrared spectra of the Greenhouse Gases Observing Satellite: Potential of retrieving CO2 vertical profile from high-resolution FTS sensor, J. Geophys. Res.-Atmos., 114, D17305, doi:10.1029/2008JD011500, 2009.

Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory, 2. Analysis of the NOAA/GMCC data, J. Geophys. Res., 94, 8549–8565, 1989.

Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Phil. Trans. Roy. Soc. A: Math. Phys. Eng. Sci., 369, 2073–2086, doi:10.1098/rsta.2010.0313, 2011.

Worden, J., Kulawik, S. S., Shephard, M. W., Clough, S. A., Worden, H., Bowman, K., and Goldman, A.: Predicted errors of tropospheric emission spectrometer nadir retrievals from spectral window selection, J. Geophys. Res.-Atmos., 109, D09308, doi:10.1029/2004JD004522, 2004.

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

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., Morino, I., and Yokota, T.: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite, Atmos. Meas. Tech., 4, 717–734, doi:10.5194/amt-4-717-2011, 2011.

Zhang, L., Jacob, D. J., Boersma, K. F., Jaffe, D. A., Olson, J. R., Bowman, K. W., Worden, J. R., Thompson, A. M., Avery, M. A., Cohen, R. C., Dibb, J. E., Flock, F. M., Fuelberg, H. E., Huey, L. G., McMillan, W. W., Singh, H. B., and Weinheimer, A. J.: Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations, Atmos. Chem. Phys., 8, 6117–6136, doi:10.5194/acp-8-6117-2008, 2008.