

Number distribution differences among multiple satellite-based X_{CO_2} datasets and their impact on monthly CO_2 flux estimation

Hiroshi Takagi¹, Sander Houweling^{2,3}, Robert J. Andres⁴, Dmitry Belikov^{1,5,6}, Andrey Bril^{1,*}, Hartmut Boesch⁷, Andre Butz⁸, Otto Hasekamp², Makoto Inoue^{1,**}, Isamu Morino¹, Tomohiro Oda^{9,10}, Christopher W. O'Dell¹¹, Sergey Oshchepkov^{1,*}, Robert Parker⁷, Makoto Saito¹, Osamu Uchino¹, Vinu Valsala¹², Tatsuya Yokota¹, Yukio Yoshida¹, and Shamil Maksyutov¹

1 National Institute for Environmental Studies, Tsukuba, Japan

2 Netherland Institute for Space Research, Utrecht, The Netherlands

3 Institute of Marine and Atmospheric Research Utrecht, Utrecht, The Netherlands

4 Oak Ridge National Laboratory, TN, USA

5 National Institute of Polar Research, Tokyo, Japan

6 Tomsk State University, Tomsk, Russian Federation

7 Department of Physics and Astronomy, University of Leicester, Leicester, UK

8 Karlsruhe Institute of Technology, Leopoldshafen, Germany

9 Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, MD, USA

10 Goddard Earth Sciences Technology and Research, Universities Space Research Association, MD, USA

11 Colorado State University, CO, USA

12 Indian Institute for Tropical Meteorology, Pune, India

* Now at National Academy of Science of Belarus, Minsk, Belarus

** Now at Department of Biological Environment, Akita Prefectural University, Akita, Japan

Corresponding author: H. Takagi, National Institute for Environmental Studies, Tsukuba,
Japan (takagi.hiroshi@nies.go.jp)

Key points

- Impact of distribution differences among five GOSAT CO₂ data on flux estimation was assessed
- Distribution differences can cause flux spreads as large as half the flux annual amplitude
- Impact on global annual flux was found to be rather small

Abstract

We assessed to what extent differences in the number distribution of satellite-based CO₂ data can impact the estimation of surface CO₂ fluxes on monthly and regional scales. Study motivation comes from the fact that CO₂ data number yields by existing five algorithms for retrieving CO₂ data from GOSAT spectral soundings differ largely from one another over some parts of the globe. For this assessment, we used a single inversion system and five synthetic CO₂ datasets whose geolocations are identical to those of the actual datasets. The synthetic data were used to eliminate confusion coming from possible differences in concentration and bias adjustment among the five actual datasets. We found flux spreads as large as 2.0 gC m⁻² day⁻¹ in a month in some regions, and some spreads were found to exceed half the flux annual amplitude. The impact of these large spreads on the annual global flux was rather small.

Keywords: GOSAT, column-averaged CO₂ concentration, surface CO₂ flux estimation

50

51 **1. Introduction**

52 Atmospheric inversion is a technique that systematically searches for spatiotemporal
53 distributions of trace gas fluxes that yield modeled atmospheric concentrations close to
54 observations. This technique, also known as Bayesian optimization, has been commonly used
55 for the inference of surface CO₂ flux distributions. As the flux inference relies heavily upon
56 observational constraints, several sensitivity studies were conducted in the past [e.g. Law et
57 al., 2001; Maksyutov et al., 2003; Yuen et al., 2005; Gurney et al., 2008; Bruhwiler et al.,
58 2011; Saeki et al., 2013]. These studies showed that the CO₂ flux estimation is highly
59 sensitive to changes in the spatial distribution of surface CO₂ observation sites. Bruhwiler et
60 al. [2011], in particular, demonstrated that additional constraints introduced to the flux
61 estimation can significantly reduce the uncertainties of fluxes especially for under-sampled
62 regions, but they can also cause sudden changes in the fluxes that sometimes exceed the
63 amplitudes of flux inter-annual variability. Changes in observational constraints both in time
64 and space can be problematic when trends of the estimated fluxes are studied.

65 Now, with the advent of the Japanese Greenhouse gases Observing SATellite (GOSAT)
66 in 2009 and more recently NASA's Orbiting Carbon Observatory-2 in 2014, the satellite
67 retrievals of column-averaged CO₂ concentration (X_{CO_2}), which can complement the surface-
68 based measurements, are available for constraining surface CO₂ fluxes. These satellite-based
69 X_{CO_2} retrievals are, however, quite different from the surface-based measurements in
70 precision, frequency, and spatial distribution. In the case of GOSAT retrievals, their precision
71 is reported to be ~2 ppm [Oshchepkov et al., 2013a], which is about one-order-of-magnitude
72 larger than that of typical surface measurements [e.g. Tans and Thoning, 2008]. About one
73 hundred thousand GOSAT X_{CO_2} retrievals over land are available in a year, but unlike the
74 surface measurements at fixed locations, success in X_{CO_2} retrieving is highly dependent on

local cloudiness and aerosol loading and thus repeat retrievals at the same location on subsequent satellite orbits are not guaranteed. Further, for obtaining better results, existing five algorithms for retrieving GOSAT X_{CO_2} differ in a number of aspects including input spectral data pre-screening, cloud and aerosol modeling, radiative transfer modeling, and post-retrieval screening, all of which impact the X_{CO_2} number yield. The X_{CO_2} number yields by the five algorithms were found to differ by tens of thousands in a year [Takagi et al., 2014]. Variations in X_{CO_2} number distribution are equivalent to introducing or withdrawing observational constraints in flux inference, as in the study by Bruhwiler et al. [2011].

We thus assessed to what extent the X_{CO_2} number distribution variations, owing to the X_{CO_2} retrieval algorithm differences, can impact CO_2 flux estimation. We did so by inter-comparing regional fluxes estimated from five different synthetic X_{CO_2} datasets whose geographical distribution patterns are identical to those of the five actual GOSAT X_{CO_2} datasets. A single inversion system was used for this estimation. The use of synthetic X_{CO_2} data eliminates confusion coming from possible differences in concentration and bias adjustment among the five actual X_{CO_2} datasets; this way only the contribution of the number distribution differences to the flux estimation can be evaluated.

2. Method

The five X_{CO_2} retrieval algorithms considered here were developed independently by the National Institute for Environmental Studies (NIES), Japan (NIES [Yoshida et al., 2013] and PPDF-S [Oshchepkov et al., 2013b] algorithms), the NASA Atmospheric CO_2 Observation from Space (ACOS) team [O'Dell et al., 2012] (ACOS algorithm), the Netherlands Institute for Space Research / Karlsruhe Institute for Technology [Butz et al., 2011] (RemoTeC algorithm), Germany, and the University of Leicester (UoL), UK [Boesch et al., 2011] (UoL-FP algorithm). Descriptions of these five algorithms can be found in an inter-

comparison study by Oshchepkov et al. [2013a]. Key differences among these algorithms are summarized in Table 1 of a previous GOSAT flux inter-comparison report by Takagi et al. [2014] (T14). The versions of the retrieval datasets used here are as follows: NIES v02.11, PPDF-S v02.11, ACOS B3.4, RemoTeC v2.11, and UoL-FP v4. For creating the synthetic X_{CO_2} datasets, information on X_{CO_2} geolocation and column averaging kernel was taken from each of the five datasets and used to sample X_{CO_2} in a forward concentration simulation in five different ways. The forward simulation was performed with version 08.1i of the NIES atmospheric tracer transport model (NIES-TM) [Belikov et al., 2013] (resolution: $2.5^\circ \times 2.5^\circ$ horizontal and 32 vertical) driven by the Japan Meteorological Agency (JMA)'s JCDAS (JMA Climate Data Assimilation System) meteorological analysis data [Onogi et al., 2007]. We used a surface-data-optimized CO_2 flux dataset by Chevalier et al. [2010] (ver.13.1; available at <http://apps.ecmwf.int/datasets/data/macc-ghg-inversions/>) as “truth” in this forward simulation.

The flux inference by the five synthetic X_{CO_2} datasets was performed with an inverse modeling system used in T14 and described in detail by Maksyutov et al. [2013]. The system consists of NIES-TM and a fixed-lag Kalman Smoother optimization scheme [Bruhwiler et al., 2005] that estimates monthly fluxes of 64 global regions on sub-continental and ocean-basin scales (42 terrestrial and 22 oceanic) [Patra et al., 2005]. The a priori flux data used here consist of the following four components: anthropogenic emissions by ODIAC (Open source Data Inventory of Anthropogenic CO_2 emission) high-resolution dataset [Oda and Maksyutov, 2011] combined with the Carbon Dioxide Information Analysis Center's monthly $1^\circ \times 1^\circ$ resolution dataset [Andres et al., 2011]; monthly biomass burning emissions by GFED (Global Fire Emissions Database) version 3.1 inventory [van der Werf et al. 2010]; daily net ecosystem exchange (NEE) predicted by VISIT (Vegetation Integrative Simulator for Trace

gases) terrestrial biosphere process model [Ito 2010; Saito et al. 2011]; monthly ocean-atmosphere CO₂ fluxes generated with an ocean pCO₂ data assimilation system [Valsala and Maksyutov 2010].

Only the land synthetic X_{CO2} values (oceanic data not included in UoL-FP v4 dataset) over a period between June 2009 and March 2011 were used for this flux inference. Surface-based CO₂ observations were not involved here. A model-observation mismatch uncertainty of 3 ppm (above-mentioned X_{CO2} precision of 2 ppm + forward modeling error of ~1 ppm [Belikov et al., 2013]) was assigned to all individual synthetic X_{CO2} values. Unlike T14, the individual X_{CO2} retrievals were not aggregated nor monthly-averaged in the inversion so that the full extent of the impact of the spatial distribution variations can be evaluated. Along with the synthetic X_{CO2} inversion, we also performed inversion with the actual X_{CO2} retrievals (biases were corrected by individual data providers) for a comparison purpose. In this analysis, we focused on flux estimates for 12 months in 2010 out of the 25-months calculation period.

3. Results

Presented in Figure 1 is the spread of the five regional fluxes estimated from the five synthetic X_{CO2} datasets. Values shown here are the standard deviations (SD) of five regional fluxes in gC m⁻² day⁻¹ over the one-year analyzed period. One characteristic found in this time series is that many of SDs of temperate and boreal regions in the Northern Hemisphere (NH) peak out between May and September of the analyzed year and gradually diminish toward December, implying X_{CO2} distribution variations change with time. On one hand, SDs of Amazonia (Regions 9-12; region IDs are found in upper left panel of Figure 1) and tropical Asia (33) remain large throughout the year. Those of southwestern North America (5), southern tips of South America and Africa (13, 21, and 22), and southwestern Australia (35)

stay very small ($<0.15 \text{ gC m}^{-2} \text{ day}^{-1}$) over the analyzed period, suggesting small X_{CO_2} distribution variations. We take a close look at the flux time series and X_{CO_2} spatial distributions of three selected regions that are representative of the above-mentioned characteristics: Regions 6 (NH temperate), 9 (large SD year-round), and 13 (small SD year-round).

Panel A of Figure 2 shows the one-year time series of five estimated fluxes for Region 6, the southeastern quadrant of the contiguous US. The fluxes shown are net but without anthropogenic emissions. The a posteriori uncertainties of the estimated fluxes are shown with the error bars. The five fluxes are in good agreement from January through April (SD in January = $0.1 \text{ gC m}^{-2} \text{ day}^{-1}$), but after that period the spread among them becomes large toward June and July (SD in June = $0.7 \text{ gC m}^{-2} \text{ day}^{-1}$) and diminishes toward the end of the year. Panel A of Figure 3 displays the spatial distributions of the five X_{CO_2} retrievals over this region (circles in red) and the surroundings (in pink). Distributions for January and June 2010 are contrasted here. The total number of X_{CO_2} retrievals found over the region in each month is also indicated in the figure. We note here that the circles are drawn over the others thus some may not be visible in the figure. The January distributions shown here resemble those of spring, fall, and winter seasons; the number of X_{CO_2} retrievals found over this region ($2.23 \text{ million km}^2$) in a month during these seasons is similar among the five X_{CO_2} datasets (January population range: 183 – 273 per region). Although some fine differences exist, X_{CO_2} retrievals by the five algorithms are distributed nearly similarly over the region. During the summer months (June to August), however, the X_{CO_2} populations drop significantly and become quite different from one to another (June population range: 15 – 122 per region). The population difference causes variations in the spatial coverage, as shown in the lower part of Panel A (June 2010). The overall population drop can be attributed to increased local cloudiness and/or aerosol loading, and the population differences among the five datasets suggest

differences in the current X_{CO_2} retrieval approaches. A separate, detailed investigation is needed to understand which processes and approaches in the five retrieval algorithms contribute most to these population differences.

The largest difference in population and spatial distribution can be seen between NIES and UoL-FP datasets, and the difference resulted in the largest flux spread of $1.6 \text{ gC m}^{-2} \text{ day}^{-1}$ in June 2010 (between NIES and UoL-FP estimates), which is 40 times larger than that in January ($0.04 \text{ gC m}^{-2} \text{ day}^{-1}$). The June spread is also 56% of the mean peak-to-peak amplitude of the five one-year flux time series ($2.8 \text{ gC m}^{-2} \text{ day}^{-1}$). The X_{CO_2} population difference is reflected in the magnitude of the a posteriori flux uncertainty: NIES = $0.8 \text{ gC m}^{-2} \text{ day}^{-1}$ (smaller population and less constrained); UoL-FP = $0.4 \text{ gC m}^{-2} \text{ day}^{-1}$ (larger population thus better constrained). In terms of a priori flux uncertainty reduction (pUR), which denotes the degree of how well a monthly regional flux is constrained by observations (given as $\text{pUR} = 100 \times (1 - \text{a posteriori uncertainty} / \text{a priori uncertainty})$), the UoL-FP retrievals attain a 74% reduction while the NIES retrievals only achieve 42% (a 32% difference).

The time series of the five flux estimates for Region 9, the southwestern part of Amazonia, are shown in Panel B of Figure 2. The time series indicate much larger spreads among the five estimates than those of Region 6 almost throughout the year. These larger spreads are attributable to considerably scant X_{CO_2} retrievals over the region (2.59 million km^2 , similar to Region 6) and those scattered around the region. The X_{CO_2} spatial distributions for March 2010 are shown in Panel B of Figure 3 (population range: 0 – 6 per region), which are representative of those in early and late 2010 (some X_{CO_2} retrievals exist between May and August). The fluxes of these months are not well constrained by X_{CO_2} retrievals within the region thus influenced by those found in the neighboring regions that are variable in spatial pattern (see distributions of pink circles in the figure). The largest flux spread (between PPDF-S and RemoTeC estimates) in March was found to be $2.0 \text{ gC m}^{-2} \text{ day}^{-1}$, which is the

largest among all the monthly regional cases in 2010 (also indicated in Figure 1 with dark red; SD = 0.8 gC m⁻² day⁻¹). The average pUR for this month is 37%, which is close to that of the Region 6 NIES case in June 2010.

Regions associated with very small flux SDs year-round (5, 13, 21, 22, and 35) are mostly semi-arid; owing to the frequent occurrence of clear sky, X_{CO2} retrievals by the five algorithms are constantly available throughout the year over these regions. Population differences among the five X_{CO2} datasets are found to be small throughout the year, and that is reflected in small SDs for these regions (Figure 1). Panel C of Figure 3 shows the five X_{CO2} distributions over Region 21, the southern tip of Africa (2.21 million km²), for June 2010. The five X_{CO2} populations and the spatial distributions over this region and the surroundings are similar to one another throughout the year, and the five fluxes estimated for this region are found to agree well year round (Panel C of Figure 2). The range of flux spread over the one-year is 0.07 – 0.35 gC m⁻² day⁻¹ (annual mean SD = 0.04 gC m⁻² day⁻¹).

SD of Region 20, the eastern half of the Sahara Desert, is found to be the smallest in 2010 (0.0 gC m⁻² day⁻¹; Figure 1), but this case may be considered as an exception because of its near-zero regional NEE predicted throughout the year and a very small a priori uncertainty assigned that leave nearly no room for being optimized by X_{CO2} retrievals.

The impact that these X_{CO2} spatial distribution variations have on the regional and global fluxes on an annual time scale was also checked. The five global annual flux estimates, obtained by aggregating the individual 64 monthly regional estimates over the one-year period, differ as much as 0.2 GtC yr⁻¹ (SD = 0.1 GtC yr⁻¹), or 4% of the mean of the five net annual global flux values (5.0 GtC yr⁻¹). The fluxes of the three focused regions (Regions 6, 9 and 21) on an annual scale were found to differ as much as 0.2, 0.6, and 0.1 GtC region⁻¹ yr⁻¹, respectively. The largest annual-scale difference was 0.8 GtC region⁻¹ yr⁻¹ (Region 30), which is 16% of the net annual global flux. Other regions with particularly large flux differences

($>0.3 \text{ GtC region}^{-1} \text{ yr}^{-1}$) include Regions 15, 26, 32, 33, and 34. These large regional spreads, however, do not surface as much in the global annual estimates ($\text{SD} = 0.1 \text{ GtC yr}^{-1}$) after the aggregation of the 64 regional values.

Comparing the five fluxes estimated from the synthetic X_{CO_2} retrievals with those estimated from the actual X_{CO_2} retrievals may indicate the extent that the X_{CO_2} spatial distribution variations account for the spreads among the actual flux estimates, in which contributions from both the spatial distribution variations and possible concentration/bias differences are reflected. We calculated SDs of five fluxes estimated from the five actual X_{CO_2} datasets, and the values are shown in Figure S1 in the supplementary information. SDs of the actual flux estimates (SD_{act}) are overall larger than SDs of the synthetic flux estimates (SD_{syn}), although a few SD_{act} are found to be smaller than the corresponding SD_{syn} (e.g. Region 21 in January 2010: $\text{SD}_{\text{act}} = 0.1$ and $\text{SD}_{\text{syn}} = 0.2$), suggesting that the concentration contributions may have worked to bring the flux estimates closer to one another. On average, the ratio of SD_{act} to SD_{syn} turned out to be 1:0.38 (cases with $\text{SD}_{\text{act}} < \text{SD}_{\text{syn}}$ were excluded). The exact contribution of concentration and/or bias differences among the five X_{CO_2} datasets to SD_{act} is not yet known, but our finding here implies that a large contribution to spreads found in the actual flux estimates comes from concentration/bias differences among the five X_{CO_2} datasets, although the contribution from X_{CO_2} distribution variations is not trivial.

4. Discussion and concluding remarks

A few things may be worth noting on what we found out in this assessment. First, X_{CO_2} population differences found among the five X_{CO_2} datasets have potentials to cause significant differences in fluxes estimated on monthly and regional scales. As demonstrated in the case of Region 6 (SE part of US) in June 2010, a X_{CO_2} population difference of this extent can lead to flux spreads greater than a half of the amplitude of the flux seasonal cycle, which

can sum up to 0.6 GtC difference in a year (12% of the global annual flux). This is clearly disadvantageous to studying both short- and long-term CO₂ flux trends. X_{CO2} population differences during the growing season of plants over temperate and boreal regions, for instance, can confuse the inference of NEE for those regions. Again, in the case of Region 6 in June, the a priori flux (NEE + biomass burning emission) was prescribed at -1.4 gC m⁻² day⁻¹, and the NIES and UoL a posteriori flux estimates were -0.3 and -1.9 gC m⁻² day⁻¹, respectively; the NIES X_{CO2} retrievals adjust the a priori flux toward the source side by 1.1 gC m⁻² day⁻¹, whereas the UoL retrievals force it toward the sink side by 0.5 gC m⁻² day⁻¹. Notice here that the departure from the a priori value in the NIES case (1.1 gC m⁻² day⁻¹) is nearly equivalent to the absolute value of the prescribed a priori flux itself. In the long run, such population differences can also complicate the process of distinguishing anomalies in flux time series. In view of diminishing current differences in X_{CO2} number yield, further X_{CO2} retrieval algorithm inter-comparison studies are desired, particularly on differences in the approaches of pre-filtering input spectral data and post-screening low quality X_{CO2} retrievals.

Second, it is necessary to evaluate the impact of X_{CO2} distribution variations on the flux estimation on smaller spatial and temporal scales than what we adopted in this study (~3000 km mesh and monthly), as the population and spatial distribution differences among the datasets can become more pronounced on smaller spatiotemporal scales. As shown in Panel A of Figure 3, despite the population similarities found among the five X_{CO2} datasets over Region 6 in January 2010, their distributions are not identical to one another; fine differences exist in a mesh size of several degrees (e.g. distribution differences near the NW corner of Region 6). Additional synthetic inversion studies on smaller scales are desired. Also needed for characterizing satellite-based inversion is an inter-comparison of current flux inversion systems to understand how differences among those systems can impact flux estimates, using a common input X_{CO2} retrieval dataset.

Third, as implied in the case of Region 9 (SW Amazonia), there is a clear need to improve the current X_{CO_2} spatial coverage based on GOSAT spectral measurement. Shown in Figure 4 are the one-year population distributions of the five X_{CO_2} datasets on a 2.5-degree mesh used for the inference of the 2010 fluxes. Although the GOSAT X_{CO_2} retrievals fill out many blanks in the existing networks of ground-based monitoring stations (Figure 4 top panel), the populations are mostly concentrated over the mid-latitudinal regions in the both hemispheres, such as the US, temperate Eurasia, Australia, and southern parts of South America and Africa. The populations over Amazonia, tropical Africa, southeastern and tropical Asia, and NH boreal and arctic regions, on the other hand, are one to two orders-of-magnitude smaller than those of the mid-latitudinal bands (only a few to several tens per grid cell per year; light green to blue color in the log scale). The reasons behind this include the fact that the NH boreal regions see most X_{CO_2} retrievals only between May and August, and over the rest of the year these regions become undersampled due to low local solar zenith angles. Also, retrieving X_{CO_2} over tropical Asia, tropical Africa, and Amazonia is heavily hindered by frequent cloud coverage around local noon when GOSAT measurement is made. To better cover these regions and seasons that are currently undersampled by GOSAT, some future missions are being considered that utilize observing platforms placed in a quasi-geostationary orbit to observe the Arctic and NH boreal regions [Nassar et al., 2014] and in several geostationary orbits to scan specific parts of the globe [Polonsky et al., 2014; Butz et al., 2015]. These platforms may allow for more contiguous and frequent spectral sounding. Combining the outcomes of measurement by platforms such as these may yield spatiotemporally more seamless CO_2 distributions favorable for inferring fluxes globally.

Until such future space-based CO_2 data have become available and the impact of differences among the existing X_{CO_2} retrieval algorithms and flux inversion systems on the estimates of surface CO_2 fluxes have been well characterized and understood, it may be sound

to take into account the above-mentioned potential uncertainties in translating the results of GOSAT-based flux estimates.

Acknowledgements

The GOSAT Project is a joint undertaking of three organizations: the Japan Aerospace Exploration Agency, the National Institute for Environmental Studies (NIES), and the Japanese Ministry of the Environment. The authors would like to thank the members of the GOSAT Project for their contribution to this work. The computational resources (SGI UV20) were provided by NIES. RJA was sponsored by U.S. Department of Energy, Office of Science, Biological and Environmental Research programs and performed at Oak Ridge National Laboratory under U.S. Department of Energy contract DE-AC05-00OR22725. AB was supported through the Emmy-Noether programme of Deutsche Forschungsgemeinschaft, grant BU2599/1-1 (RemoTeC). HB and RP acknowledge funding by the NERC National Centre for Earth Observation and the ESA Climate Change Initiative. A request for data that support the figures may be sent to H. Takagi (takagi.hiroshi@nies.go.jp). The meteorological data used in the forward transport modeling were provided by the cooperative research project of the JRA-25 long-term reanalysis by Japan Meteorological Agency and Central Research Institute of Electric Power Industry and obtained from <http://jra.kishou.go.jp/JRA-25/> (data provision terminated in February 2014). NIES v02.11 dataset was obtained from <https://data.gosat.nies.go.jp/>. ACOS B3.4 dataset was obtained from <https://co2.jpl.nasa.gov/>. PPDF-S v2.11 dataset can be obtained from the NIES GOSAT Project upon request (gosat-support@nies.go.jp). RemoTeC v2.11 dataset was downloaded from imkwww1.imk.kit.edu. UoL-FP v4 dataset was obtained from <http://www.leos.le.ac.uk/GHG/data/>. GFED v3.1 dataset was downloaded from <http://www.globalfiredata.org/>. ODIAC dataset was obtained via request (<http://odiac.org/>). VISIT and ocean-atmosphere exchange data used in this study

are found in v02.02 of GOSAT Level 4A data product downloadable at
<https://data.gosat.nies.go.jp/>.

References

Andres, R. J., J. S. Gregg, L. Losey, G. Marland, and T. A. Boden (2011), Monthly, global emissions of carbon dioxide from fossil fuel combustion. *Tellus*, **63B**, 309–327.

Belikov, D.A., et al. (2013), Simulations of column-averaged CO₂ and CH₄ using the NIES TM with a hybrid sigma-isentropic (σ - θ) vertical coordinate, *Atmos. Chem. Phys.*, **13**, 1713-1732, doi:10.5194/acp-13-1713-2013.

Boesch, H., D. Baker, B. Connor, D. Crisp and C. Miller (2011), Global characterization of CO₂ column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 Mission, *Remote Sens.*, **3**, 270-304; doi:10.3390/rs3020270.

Bruhwyler, L. M. P., A. M. Michalak, W. Peters, D. F. Baker, and P. Tans (2005), An improved Kalman Smoother for atmospheric inversions, *Atmos. Chem. Phys.*, **5**, 2691–2702, doi:10.5194/acp-5-2691-2005.

Bruhwyler, L. M. P., A. M. Michalak, and P. P Tans, (2011), Spatial and temporal resolution of carbon flux estimates for 1983–2002, *Biogeosciences*, **8**, 1309-1331, doi:10.5194/bg-8-1309-2011.

Butz, A., et al. (2011), Toward accurate CO₂ and CH₄ observations from GOSAT, *Geophys. Res. Lett.*, **38**, L14812, doi:10.1029/2011GL047888.

Butz, A., et al. (2015), Geostationary Emission Explorer for Europe (G3E): mission concept and initial performance assessment, *Atmos. Meas. Tech. Discuss.*, **8**, 6949–6991, doi:10.5194/amtd-8-6949-2015

348 Chevallier, F., et al. (2010), CO₂ surface fluxes at grid point scale estimated from a global 21
 349 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, **115**, D21307,
 350 doi.org/10.1029/2010JD013887

351 Cogan, A. J., et al. (2012), Atmospheric carbon dioxide retrieved from the Greenhouse gases
 352 Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and
 353 GEOS-Chem model calculations, *J. Geophys. Res.*, **117**, D21301,
 354 doi:10.1029/2012JD018087.

355 Gurney, K. R., D. Baker, P. Rayner, and S. Denning (2008), Interannual variations in
 356 continental-scale net carbon exchange and sensitivity to observing networks estimated
 357 from atmospheric CO₂ inversions for the period 1980 to 2005, *Global Biogeochem. Cycles*,
 358 **22**, GB3025, doi:10.1029/2007GB003082.

359 Ito, A. (2010), Changing ecophysiological processes and carbon budget in East Asian
 360 ecosystems under near-future changes in climate: implications for long-term monitoring
 361 from a process-based model, *J. Plant Res.*, **123**, 577–588, doi:10.1007/s10265-009-0305-x.

362 Law, R. M., Y.-H. Chen, K. R. Gurney and TransCom 3 Modellers (2003), TransCom 3 CO₂
 363 inversion intercomparison: 2. Sensitivity of annual mean results to data choices, *Tellus*,
 364 **55B**, 580–595. doi: 10.1034/j.1600-0889.2003.00053.x.

365 Maksyutov, S., et al. (2013), Regional CO₂ flux estimates for 2009–2010 based on GOSAT
 366 and ground-based CO₂ observations, *Atmos. Chem. Phys.*, **13**, 9351–9373, doi:10.5194/acp-
 367 13-9351-2013.

368 Nassar, R., C. E. Sioris, D. B. A. Jones, and J. C. McConnell (2014), Satellite observations of
 369 CO₂ from a highly elliptical orbit for studies of the Arctic and boreal carbon cycle, *J.*
 370 *Geophys. Res. Atmos.*, **119**, 2654–2673, doi:10.1002/2013JD020337.

Oda, T., and S. Maksyutov (2011), A very high-resolution ($1\text{ km} \times 1\text{ km}$) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, *Atmos. Chem. Phys.*, **11**, 543–556, doi:10.5194/acp-11-543-2011.

O'Dell, C. W., et al. (2012), The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, **5**, 99–121, doi:10.5194/amt-5-99-2012.

Onogi, K., et al. (2007), The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.

Oshchepkov, S., et al. (2013a), Simultaneous retrieval of atmospheric CO_2 and light path modification from space-based spectroscopic observations of greenhouse gases: methodology and application to GOSAT measurements over TCCON sites, *Applied Optics*, **52**, 1339–1350.

Oshchepkov, S., et al. (2013b), Effects of atmospheric light scattering on spectroscopic observations of greenhouse gases from space. Part 2: Algorithm intercomparison in the GOSAT data processing for CO_2 retrievals over TCCON sites, *J. Geophys. Res. Atmos.*, **118**, 1493–1512, doi:10.1002/jgrd.50146.

Patra, P. K., M. Ishizawa, S. Maksyutov, T. Nakazawa, and G. Inoue (2005), Role of biomass burning and climate anomalies on land atmosphere carbon fluxes based on inverse modeling of atmospheric CO_2 , *Global Biogeochem. Cycles*, **19**, GB3005, doi:10.1029/2004GB002258.

Polonsky, I. N., D. M. O'Brien, J. B. Kumer, C. W. O'Dell, and the geoCARB Team (2014), Performance of a geostationary mission, geoCARB, to measure CO_2 , CH_4 and CO column-averaged concentration, *Atmos. Meas. Tech.*, **7**, 959–981, doi:10.5194/amt-7-959-2014.

Saeki, T., et al. (2013), Carbon flux estimation for Siberia by inverse modeling constrained by aircraft and tower CO_2 measurements, *J. Geophys. Res.*, **118**, 1100–1122, doi:10.1002/jgrd.50127.

- Saito, M., A. Ito, and S. Maksyutov (2011), Evaluation of biases in JRA-25/JCDAS precipitation and their impact on the global terrestrial carbon balance, *J. Climate*, **24**, 4109–4125.
- Takagi, H. et al. (2014), Influence of differences in current GOSAT X_{CO2} retrievals on surface flux estimation, *Geophys. Res. Lett.*, **41**, 2598–2605, doi:10.1002/2013GL059174
- Tans and Thoning (2008), How we measure background CO₂ levels on Mauna Loa, web resource viewable at http://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html.
- Valsala, V., and S. Maksyutov (2010), Simulation and assimilation of global ocean *p*CO₂ and air-sea CO₂ fluxes using ship observations of surface ocean *p*CO₂ in a simplified biogeochemical offline model, *Tellus*, **62B**, 821–840, doi:10.1111/j.1600-0889.2010.00495.x.
- van der Werf, G. R., et al. (2010), Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, **10**, 11707–11735, doi:10.5194/acp-10-11707-2010.
- Yuen, C.-W., K. Higuchi, and TransCom-3 Modellers (2005), Impact of Fraserdale CO₂ observations on annual flux inversion of the North American boreal region, *Tellus*, **57B**, 203–209.
- Yoshida, Y., et al. (2013), Improvement of the retrieval algorithm for GOSAT SWIR XCO₂ and XCH₄ and their validation using TCCON data, *Atmos. Meas. Tech.*, **6**, 1533–1547, doi:10.5194/amt-6-1533-2013.

Figures

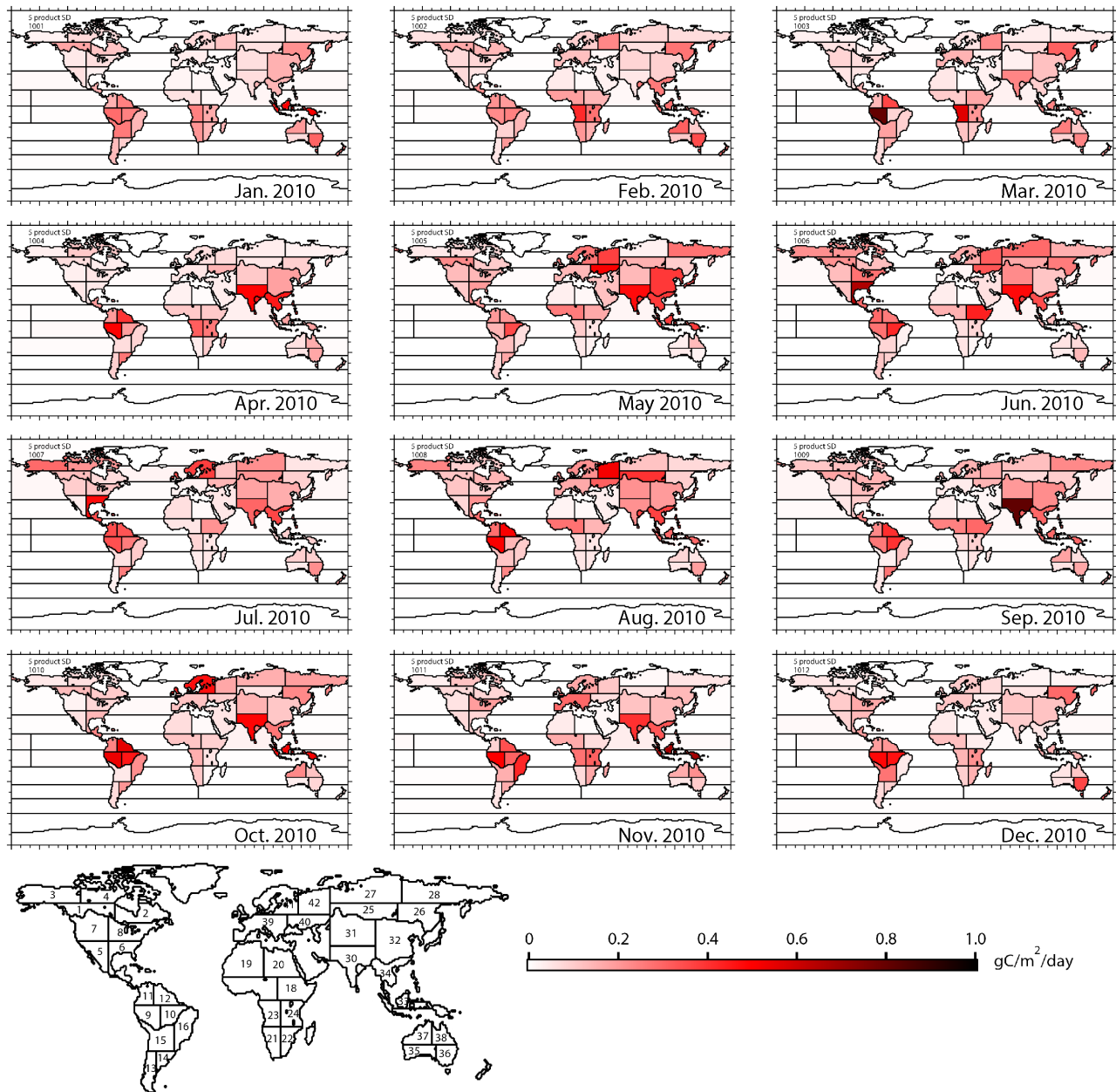
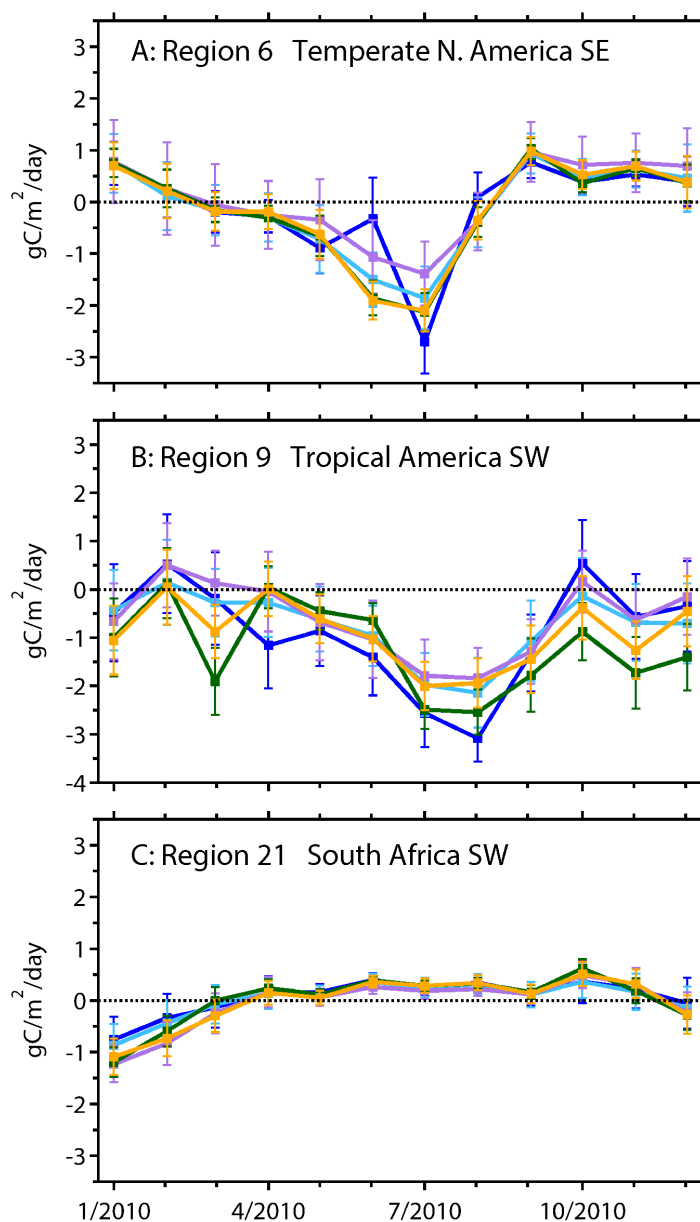


Figure 1. Standard deviation of five monthly regional fluxes estimated from the five synthetic X_{CO_2} datasets. IDs of terrestrial regions are indicated at the bottom.

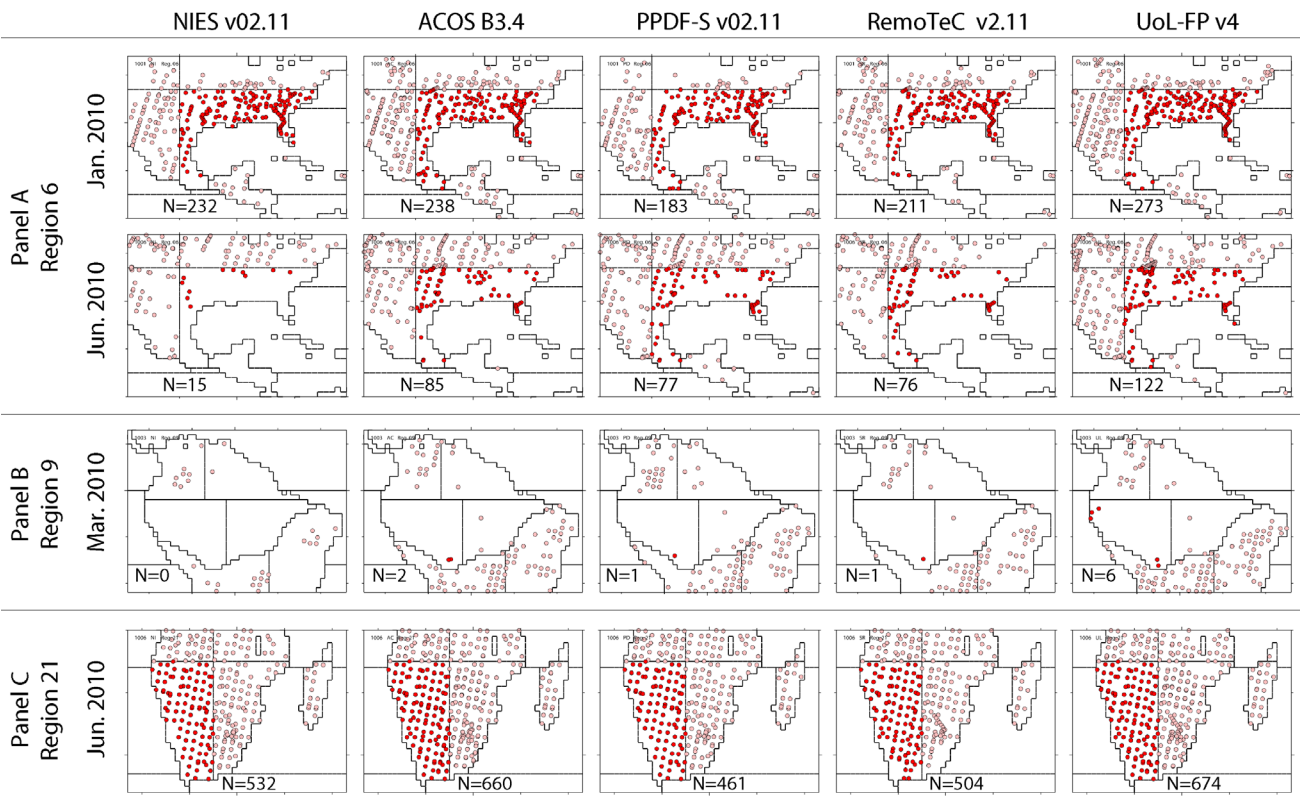


428

429 **Figure 2.** Time series of five estimated fluxes for Region 6 (top), 9 (middle), and 21 (bottom)
 430 over the year 2010. Values shown are net without the anthropogenic emissions. Blue: NIES
 431 v02.11; Light blue: ACOS B3.4; Purple: PPDF-S v02.11; Green: RemoTeC v2.11; Yellow:
 432 UoL-FP v4. Error bars indicate the a posteriori flux uncertainties.

433

434



436

437

438

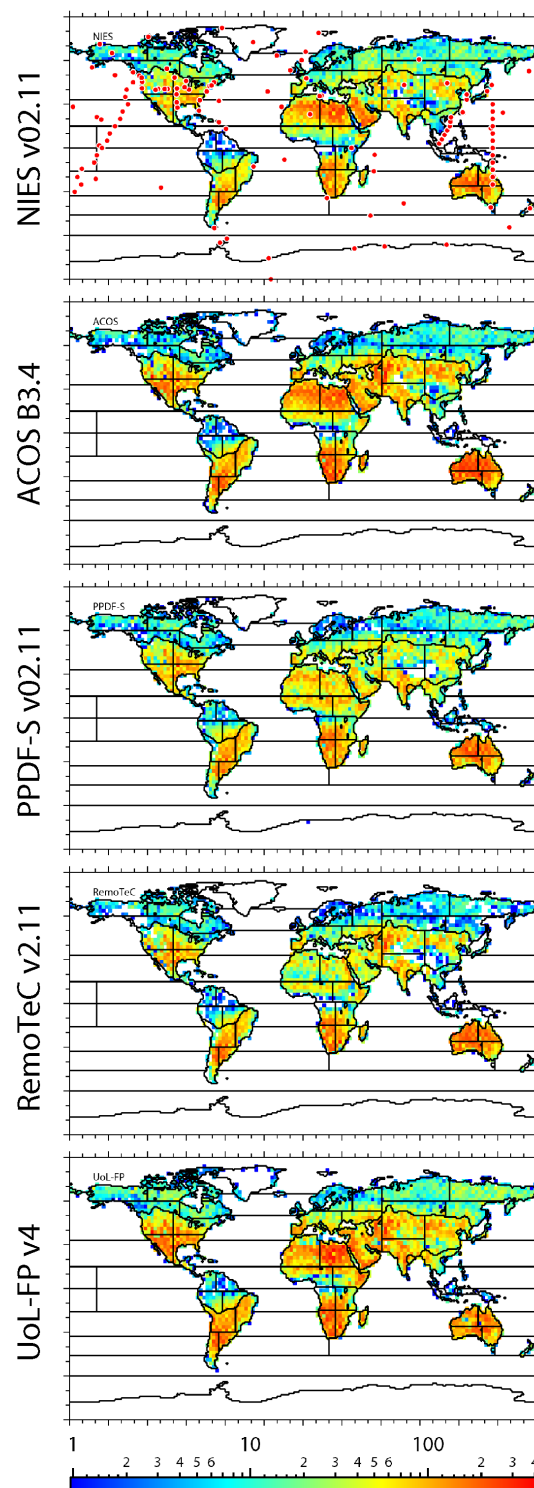
439

440

441

442

Figure 3. Spatial distributions of the five X_{CO_2} retrievals over Regions 6 (top; southeastern quadrant of contiguous US), 9 (middle; southwestern part of Amazonia), and 21 (bottom; southern tip of Africa). Circles indicate the locations of X_{CO_2} retrievals over the regions of focus (in red) and the surrounding areas (in pink). The circles may be overlaid onto one another. The total number of X_{CO_2} retrievals over the region (N) is shown in each panel.



443

444

Figure 4. One-year population distributions of the five X_{CO_2} datasets on a 2.5-degree mesh used for the inference of the 2010 fluxes. Red circles in the top panel indicate the location of surface-based CO_2 monitoring stations (220 as used in Takagi et al. [2014]).

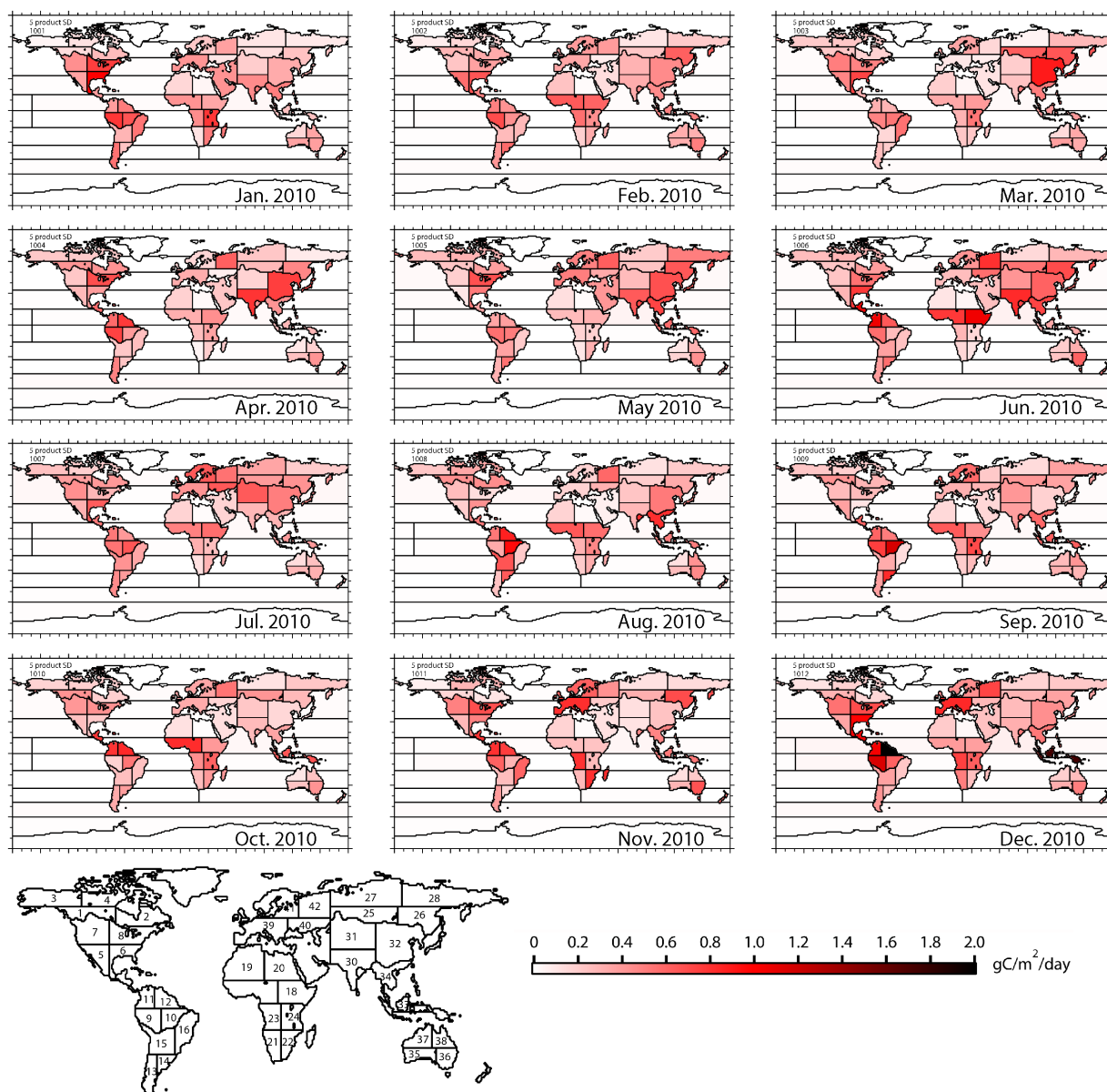


Figure S1. Standard deviation of five monthly regional fluxes estimated from the five actual X_{CO_2} datasets. Note that the scale range shown here is wider than that in Figure 1 (range: 0 – 1 $gC\ m^{-2}\ day^{-1}$).