Global land mapping of satellite-observed CO2 total columns using spatio-temporal geostatistics — Source link ↗

Zhao-Cheng Zeng, Liping Lei, Kimberly Strong, Dylan B. A. Jones ...+18 more authors

Institutions: The Chinese University of Hong Kong, Chinese Academy of Sciences, University of Toronto, University of Wollongong ...+6 more institutions

Related papers:

- Global CO2 distributions over land from the Greenhouse Gases Observing Satellite (GOSAT)

- The ACOS CO 2 retrieval algorithm – Part 1: Description and validation against synthetic observations

- Orbiting Carbon Observatory: Inverse method and prospective error analysis

- Spatio-temporal data fusion for very large remote sensing datasets

- Mapping of satellite Earth observations using moving window block kriging

1-1-2017

Global land mapping of satellite-observed CO2 total columns using spatio-temporal geostatistics

Zhao-Cheng Zeng
Institute of Remote Sensing and Digital Earth, Chinese University of Hong Kong

Liping Lei
Institute of Remote Sensing and Digital Earth

Kimberly Strong
University of Toronto, strong@atmosp.physics.utoronto.ca

Dylan B. A Jones
University of Toronto

Lijie Guo
Institute of Remote Sensing and Digital Earth

See next page for additional authors

Recommended Citation

Global land mapping of satellite-observed CO2 total columns using spatio-temporal geostatistics

Abstract

This study presents an approach for generating a global land mapping dataset of the satellite measurements of CO_2 total column (XCO_2) using spatio-temporal geostatistics, which makes full use of the joint spatial and temporal dependencies between observations. The mapping approach considers the latitude-zonal seasonal cycles and spatio-temporal correlation structure of XCO_2, and obtains global land maps of XCO_2, with a spatial grid resolution of 1° latitude by 1° longitude and temporal resolution of 3 days. We evaluate the accuracy and uncertainty of the mapping dataset in the following three ways: (1) in cross-validation, the mapping approach results in a high correlation coefficient of 0.94 between the predictions and observations, (2) in comparison with ground truth provided by the Total Carbon Column Observing Network (TCCON), the predicted XCO_2 time series and those from TCCON sites are in good agreement, with an overall bias of 0.01 ppm and a standard deviation of the difference of 1.22 ppm and (3) in comparison with model simulations, the spatio-temporal variability of XCO_2 between the mapping dataset and simulations from the CT2013 and GEOS-Chem are generally consistent. The generated mapping XCO_2 data in this study provides a new global geospatial dataset in global understanding of greenhouse gases dynamics and global warming.

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Authors

Zhao-Cheng Zeng, Liping Lei, Kimberly Strong, Dylan B. A Jones, Lijie Guo, Min Liu, Feng Deng, Nicholas M. Deutscher, Manvendra K. Dubey, David W. T Griffith, and Voltaire A. Velazco

Global land mapping of satellite-observed CO2 total columns using spatio-temporal geostatistics

Zhao-Cheng Zeng[1,2], Liping Lei[2], Kimberly Strong[3], Dylan B. A. Jones[3], Lijie Guo[2], Min Liu[2], Feng Deng[3], Nicholas M. Deutscher[4,6], Manvendra K. Dubey[5], David W. T. Griffith[6], Frank Hase[7], Bradley Henderson[5], Rigel Kivi[8], Rodica Lindenmaier[5,*], Isamu Morino[9], Justus Notholt[4], Hirofumi Ohyama[10,**], Christof Petri[4], Ralf Sussmann[11], Voltaire Velazco[6], Paul O. Wennberg[12] and Hui Lin[1,13]

[1]{Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China }

[2]{Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China}

[3]{Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada}

[4]{Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany}

[5]{Los Alamos National Laboratory, Los Alamos, NM 87545, USA}

[6]{University of Wollongong, Wollongong, Australia}

[7]{Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research (IMK-ASF), Germany}

[8]{FMI Arctic Research Center, Tähteläntie 62, FIN-99600 Sodankylä, Finland}

[9]{National Institute for Environmental Studies (NIES), Tsukuba, Ibaraki, Japan}

[10]{Earth Observation Research Center, Japan Aerospace Exploration Agency (JAXA), Tsukuba-city, Ibaraki, Japan}

[11] {Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany}

[12]{California Institute of Technology, Pasadena, CA, USA}

[13]{Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China}

[*]{now at: PNNL, Richmond, WA 99352, USA}

[**]{now at: The Solar-Terrestrial Environment Laboratory, Nagoya University, Japan}

Correspondence to: Z.-C. Zeng (zczeng@link.cuhk.edu.hk) and H. Lin (huilin@cuhk.edu.hk)

Acknowledgements

Global land mapping of satellite-observed CO$_2$ total columns using spatio-temporal geostatistics

This study presents an approach for generating a global land mapping dataset of the satellite measurements of CO$_2$ total column (XCO$_2$) using spatio-temporal geostatistics, which makes full use of the joint spatial and temporal dependencies between observations. The mapping approach considers the latitude-zonal seasonal cycles and spatio-temporal correlation structure of XCO$_2$, and obtains global land maps of XCO$_2$, with a spatial grid resolution of 1° latitude by 1° longitude and temporal resolution of 3 days. We evaluate the accuracy and uncertainty of the mapping dataset in the following three ways, (1) in cross-validation, the mapping approach results in a high correlation coefficient of 0.94 between the predictions and observations, (2) in comparison with ground truth provided by the TCCON, the predicted XCO$_2$ time series and those from TCCON sites are in good agreement, with an overall bias of 0.01 ppm and a standard deviation of the difference of 1.22 ppm, and (3) in comparison with model simulations, the spatio-temporal variability of XCO$_2$ between the mapping dataset and simulations from the CT2013 and GEOS-Chem are generally consistent. The generated mapping XCO$_2$ data in this study provides a new global geospatial dataset in global understanding of greenhouse gases dynamics and global warming.

Keywords: XCO$_2$, ACOS-GOSAT, Spatio-temporal geostatistics, global mapping, geospatial dataset

1   Introduction

The concentration of atmospheric carbon dioxide (CO$_2$), the most important anthropogenic greenhouse gas, has increased by 40% since pre-industrial times as a result of the burning of fossil fuels, cement production and deforestation (IPCC 2013). While the surface carbon dioxide monitoring network (e.g., GLOBALVIEW-CO$_2$) provides *in situ* measurements of CO$_2$, the spatio-temporal variability and distribution of CO$_2$ sources and sinks are still not fully understood mainly due to the sparseness and spatial inhomogeneity of the network (Heimann, 2009; Ciais et al, 2014). Satellite observations of CO$_2$ columns, because of their global coverage and high measurement density, can complement the surface network and be used to advance our understandings of the carbon cycle and its changes (McKain et al., 2012; Duren and Miller, 2012). The Japanese Greenhouse gases Observation SATellite (GOSAT), the world's first spacecraft dedicated to quantify the atmospheric CO$_2$ and methane

concentrations, was successfully launched in 2009 (Yokota et al., 2009). The Atmospheric CO_2 Observations from Space (ACOS) project team from NASA's Orbiting Carbon Observatory (OCO), applied the OCO calibration, validation, and remote sensing retrieval algorithm to analyze the GOSAT Level 1B data products, and obtain the column-averaged CO_2 dry air mole fraction (X_{CO_2}) data products (Wunch et al., 2011a; O'Dell et al., 2012). These ACOS retrievals from GOSAT are hereafter referred to as ACOS-GOSAT. Other satellite instruments, including SCIAMACHY (Bovensmann et al., 1999) and OCO-2 (Crisp et al., 2004; Boesch et al., 2011), and the forthcoming CarbonSAT (Bovensmann et al., 2010) and TanSAT (Liu et al., 2013), are also aiming at measuring global greenhouse gases. As more data from these satellites are becoming available, global analysis of X_{CO_2} by applying methods for data modeling, validation and comparison, ranging from simple spatial and temporal averaging to sophisticated data assimilation approaches, have become possible (e.g., Buchwitz et al., 2005; Schneising et al., 2008, 2011, 2014; Morino et al., 2011; Wunch et al., 2011b; Katzfuss and Cressie, 2011; Reuter et al., 2013; Buchwitz et al., 2013).

However, due to constraints such as cloud coverage and GOSAT Fourier Transform Spectrometers (FTS) observation mode (NIES GOSAT Project, 2010), the number of available GOSAT X_{CO_2} retrievals are largely reduced and irregularly distributed in space and time, which make it difficult to directly interpret their scientific significance without further data analysis. Furthermore, the large gaps between satellite observations provide a challenge for study areas with insufficient data (Kort et al., 2012). One solution is the development of a statistical mapping method, which generates a regular distribution map of X_{CO_2} from the irregular observations, to improve the spatio-temporal coverage. This mapping dataset can potentially provide us a new way to evaluate the roles of anthropogenic emissions (sources) on atmospheric CO_2 enhancement and biopheric activities (sinks) on global CO2 seasonal cycle. As shown in Keppel-Aleks et al. (2011, 2012), the total column carbon dioxide X_{CO_2} retrieved from satellite has a larger footprint than conventional surface data because the column data is a combination of not only boundary layer CO_2 but also free troposphere CO_2. Therefore, variations in X_{CO_2} are primarily determined by large-scale flux patterns. As a result, significant correlation exits between X_{CO_2} data (Keppel-Aleks et al., 2011). Therefore, a geostatistical approach (Cressie, 1993; Cressie and Wikle, 2011), which uses the inherent autocorrelation between satellite-observed X_{CO_2} data to make optimal predictions and is flexible in handling irregular datasets, can be exploited to meet the requirements of filling gaps and mapping X_{CO_2} at high spatio-temporal resolution.

The mapping of satellite observations of X_{CO_2} has been investigated in several studies from different perspectives (Hammerling et al., 2012a; Nguyen et al., 2014; Zeng et al., 2014). Spatial-only kriging, a conventional geostatistical method, was widely adopted to generate the X_{CO_2} mapping data products (Tomosada et al., 2008, 2009; Liu et al., 2012; Watanabe, 2015). Hammerling et al. (2012a, 2012b) and Tadić et al. (2014) extended the spatial-only kriging method to a global scale using a moving window technique. This conventional spatial-only geostatistical method, which only makes use of spatial correlation, does not take into account the temporal correlation structure of the CO_2 data, and therefore the dynamic CO_2 temporal variations including the annual increase and seasonal cycles (WMO, 2014; Schneising *et al.*, 2014) are not fully considered. By incorporating the temporal variability of X_{CO_2}, Zeng et al. (2013, 2014) extended this method into the spatio-temporal domain to generate X_{CO_2} maps at high spatio-temporal resolution, and illustrated the effectiveness of the method by applying it to China as a study region. The advantages of this spatio-temporal geostatistical approach lie in (1) the utilization of the joint spatial and temporal dependences between observations to provide a probabilistic framework for data analysis and prediction, and (2) inclusion of a larger dataset both in space and time to support stable parameter estimation and prediction (Cressie and Wikle, 2011; De Iaco et al., 2012).

Building on the regional mapping method using spatio-temporal geostatistics developed by Zeng et al. (2014), this study aims to develop a global land mapping method for the ACOS-GOSAT X_{CO_2} dataset. However, the extension from the regional method to the global case is not straightforward, since the assumptions of a uniform spatio-temporal trend and spatio-temporal correlation structure in the regional dataset are often violated in the global case. As a solution, we use the long-term CarbonTracker CT2013 X_{CO_2} data and global annual CO_2 growth rate calculated from surface observation sites to constructe the spatio-temporal deterministic trend of X_{CO_2}, including latitude-zonal seasonal cycle and global annual increase, and then remove the trend from the ACOS-GOSAT data. The correlation structures of the X_{CO_2} data are then calculated and modeled for each latitude zone. Finally, the spatio-temporal geostatistical prediction approach is implemented to obtain the global land mapping of X_{CO_2}, and quantify the corresponding prediction uncertainties. The resulting dataset is assessed in three ways. Firstly, cross-validation technique (Arlot, 2010) for accuracy assessment of statistical models is applied to assess the prediction accuracy of this mapping approach. Secondly, a comparison with Total Carbon Column Observing Network (TCCON) data is performed to validate the mapping results at global TCCON sites. Finally, we investigate the

5

similarities and discrepancies between the mapping dataset and model X_{CO_2} calculations from both CarbonTracker CT2013 and GEOS-Chem by comparing their spatio-temporal variations. As suggested by Hammerling et al., (2012a), the mapping dataset will provide a useful complement to studies of carbon flux inversion with modeled CO_2 fields generated by carbon flux estimates coupled with an atmospheric transport model.

Production of global geospatial data plays a key role in supporting the development of digital earth (Guo et al., 2010), and application of these global geospatial data in global change research has been an important goal for "Digital Earth" (Shupeng and Genderen, 2008). Our study contributes to these two topics by demonstrating (1) how to effectively and precisely produce a global geospatial dataset of CO_2 column distribution at high spatio-temporal resolution, and (2) how these data can be used to analyze the spatio-temporal variations of CO_2. The mapping approach developed in this study can be further incorporated with "Digital Earth" framework for geospatial data production in diverse global change studies.

In Sect. 2, we provide a description of the data from satellite, ground observation and model simulations used in this paper. We then introduce the basic theories of spatio-temporal geostatistics, and the global land mapping approach in Sect. 3. The results of applying the approach are illustrated and discussed in Sect. 4, and conclusions follow in Sect. 5.

2   Data

2.1   *ACOS-GOSAT data product*

The v3.3 ACOS-GOSAT X_{CO_2} data, spanning from 01 June 2009 to 15 May 2013, are used in this study. Because of some potential deficiencies identified in land median-gain and ocean glint retrievals (Nguyen et al., 2014), only the land high-gain data are used. In comparison with the TCCON, the v3.3 data show a mean global bias of about 1.34 ppm and a standard deviation of 1.83 ppm (ACOS Data User's Guide, 2014). Following the data user guide, data-filtering and bias-correction are carried out before the following science analysis.

The mapping dataset has a temporal resolution of 3 days determined by the GOSAT 3-day orbiting period. Therefore, in this study we set the basic time unit to be 3-day. In general, there are nearly 122 time-units in 1 year ($365/3 \approx 122$), and from 01 June 2009 to 15 May 2013, 483 time-units of ACOS-GOSAT v3.3 data are used in this study. The basic statistics

for the $X\text{CO}_2$ data used each year are listed in Table I. The standard deviations for the four years are almost the same, while the annual mean value increases steadily, with an average annual increase of 2.03 ppm, which is in good agreement with the global mean annual increase of about 2 ppm from the GLOBALVIEW surface *in situ* flask network (WMO, 2014; Wunch et al., 2011b). Figure 1 shows the monthly global distribution of the ACOS-GOSAT $X\text{CO}_2$ retrievals, taking four months (July 2009, October 2009, January 2010 and April 2010) in four different seasons for the first year of ACOS-GOSAT measurements as examples. The study area of global land region for mapping covers land regions between 40°S and 70°N, as shown in grey in Figure 3. It can be inferred from Figure 1 that the satellite retrievals are irregularly distributed globally, and the available retrievals also change with observation time. Large gaps between satellite observations can be found in some key regions, such as over the forests in southern China, where net ecosystem production is high but there are few satellite observations in summer, mainly due to the frequent cloudy weather. Figure 2 shows the spatial distribution and temporal variation of the number of available satellite observations after data filtering. The numbers of retrievals are irregularly distributed both in space and time. As shown in Figure 2 (a), it can be seen that the number of retrievals is greater over the southern regions of South America, North America and Africa, as well as over Central Eurasia and Australia. However, data over northern Africa (around the Sahara desert) and parts of the central Australia, where data are measured with median-gain (Crisp et al., 2012), are excluded by the data filtering approach. From Figure 2 (b), the largest number of measurements is available around August and smallest around January, with an obvious annual cycle.

Table 1. Annual statistics for the ACOS-GOSAT v3.3 $X\text{CO}_2$ global land data used in this study after filtering and bias-correction.

	Number of Retrievals	Minimum (ppm)	Maximum (ppm)	Mean (ppm)	Standard Deviation (ppm)
June 2009 to May 2010	58437	355.89	403.65	386.73	3.11
June 2010 to May 2011	67840	375.72	406.60	389.01	2.92
June 2011 to May 2012	67987	357.58	404.66	390.73	2.94
June 2012 to May 2013	65360	378.92	408.14	392.82	2.95

(a) July 2009   (b) October 2009

(c) January 2010   (d) April 2010

373   383   392   402 ppm
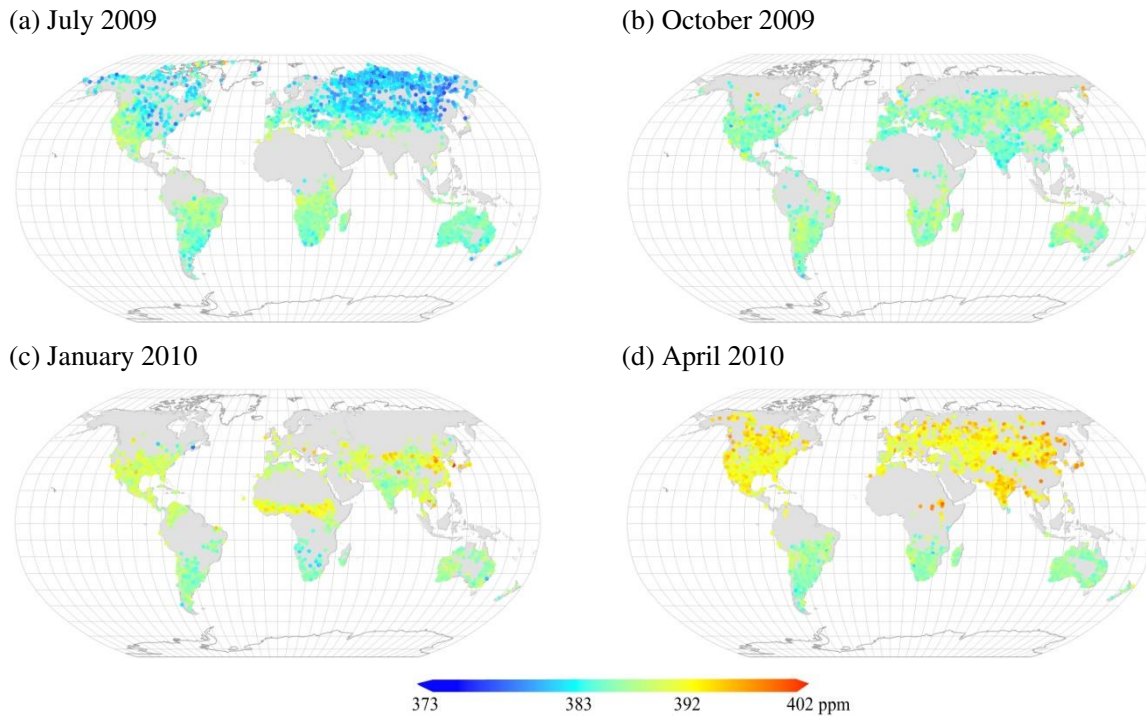
Figure 1. Example of the spatial distribution and variation of the used ACOS-GOSAT XCO$_2$ retrievals for four months of (a) July 2009, (b) October 2009, (c) January 2010 and (d) April 2010 over global land.

(a)

(b)

1   300   700   >1000
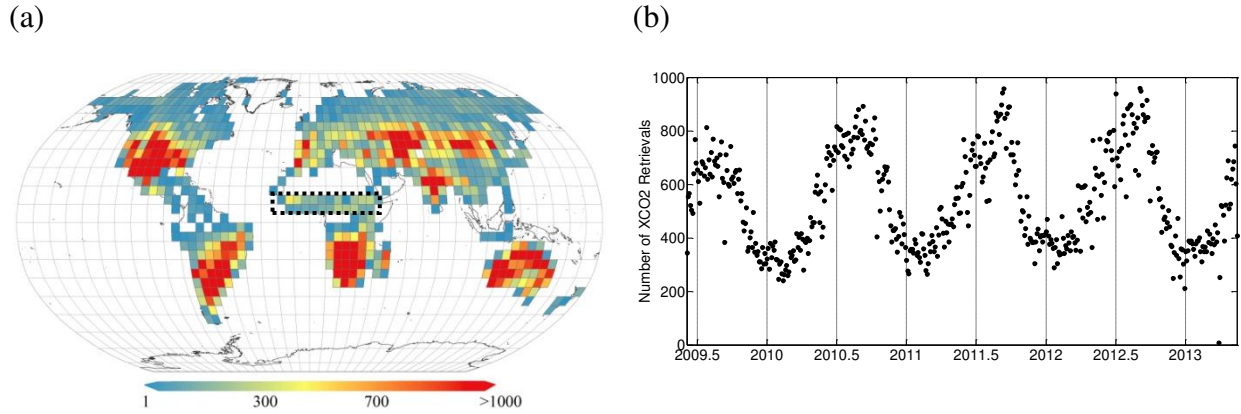
Figure 2. (a) Global spatial distribution density of the used ACOS-GOSAT XCO$_2$ retrievals (after filtering) from June 2009 to May 2013 in terms of available data number in grids of 5°longitude by 5°latitude, and (b) temporal variation of available ACOS-GOSAT retrievals (after filtering) in each 3-day interval. The dashed rectangle area in (a) from 5 to 15°N and 20°W to 40°E covers the tropical region in central Africa, which will be used in Sections 4.5 and 4.6.
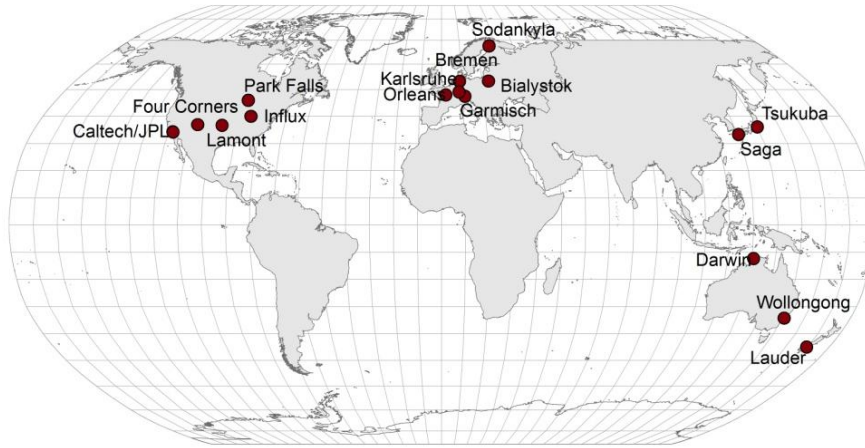
Figure 3. Global distribution of the 16 TCCON sites used for the comparison with the mapping dataset, and the global land study area for mapping in grey.

2.2 The Total Carbon Column Observing Network

The Total Carbon Column Observing Network (TCCON), a global network of ground-based FTS established for the validation of near-infrared total-column measurements from satellite observations (Wunch et al., 2011b; see TCCON Data Access in the references for detail), is collected for the validation of the global land mapping dataset. Based on comparisons with integrated aircraft profiles, TCCON has a high accuracy of approximately 0.25% in X_{CO_2} (Wunch et al., 2010). The TCCON data has been used to calibrate the ACOS-GOSAT retrievals (Wunch et al., 2011a). In this study, the 2012 release version of the TCCON data ("GGG2012") for 16 sites is used. For each TCCON site, we use all available data within the period from June 2009 to May 2013. Figure 3 shows the spatial distribution of the TCCON sites used in this study. More detailed information about the TCCON sites is given in Table 4 in Section 4.4. TCCON sites at high north latitudes (>70° N) and the ocean are not considered in this study.

2.3 CarbonTracker CT2013 and GEOS-Chem model calculations of X_{CO_2}

Both the X_{CO_2} data from CarbonTracker CT2013 and GEOS-Chem model are used in this study, in order to investigate the similarities and differences between the global land mapping dataset and the modeled X_{CO_2} and provide potential evidences for further improvement of the satellite retrievals and model simulations.

2.3.1 CarbonTracker CT2013 X_{CO_2} dataset

CarbonTracker is a modeling system for atmospheric CO_2 developed by the National Oceanic and Atmospheric Administration (NOAA). Coupled with an atmospheric transport model, CarbonTracker assimilates global atmosphere CO_2 observations from ground surface air samples, tall tower and aircrafts to simulate the global CO_2 distribution and to track global CO_2 sources and sinks (Peters et al., 2007). The 2013 release of CarbonTracker, CT2013, combines observations and flux estimates through the end of 2012 to produce estimates of global atmospheric CO_2 and surface-atmosphere fluxes from January 2000 to December 2012 (see CarbonTracker CT2013 Data Access in the references for detail). The model produces regularly gridded CO_2 at a spatial resolution of $2°$ in latitude by $3°$ in longitude with 25 vertical levels before 2006 and 34 vertical levels after 2006, and with a high temporal resolution of 3 hours. CarbonTracker is widely used in analyzing global CO_2 with satellite observations (e.g., Schneising et al., 2008, 2014; Nguyen et al., 2014). In this study, we use the long-term CT2013 data from January 2000 to December 2012 to model the deterministic spatio-temporal trend of X_{CO_2}. Besides, we use the CT2013 data from June 2009 to December 2012 for comparison with the original satellite retrievals and the corresponding mapping dataset. The CO_2 profile data from the model are transformed to CO_2 dry air mole fractions, X_{CO_2}, by using the pressure-averaged method described by Connor et al. (2008).

2.3.2 GEOS-Chem X_{CO_2} dataset

The GEOS-Chem model (http://geos-chem.org) is a global 3-D chemical transport model for simulating atmospheric composition using assimilated meteorological inputs from the Goddard Earth Observing System (GEOS-5) of the NASA Global Modeling and Assimilation Office. A recent update of the atmospheric CO_2 simulations using GEOS-Chem was developed by Nassar et al. (2010). GEOS-Chem is widely used in analyzing global CO_2 with satellite observations (e.g., Feng et al., 2009; Cogan et al., 2012). Deng et al. (2014) employed the GEOS-Chem four-dimensional variational (4D-Var) data assimilation system to assimilate ACOS-GOSAT X_{CO_2} from June 2009 to December 2010 for quantifying monthly and regional estimates of CO_2 fluxes for the year 2010. In this study, we used the a posteriori atmospheric CO_2 distributions from Deng et al. (2014), which were produced by assimilating ACOS-GOSAT v2.10 data at a horizontal resolution of $4°$ by $5°$. Only the data for 2010 are available in this study. The CO_2 profiles from the a posteriori GEOS-Chem simulation were transformed to X_{CO_2} using the pressure-averaged method described by Connor et al. (2008).

3   Methodologies

As an extension of spatial geostatistics, the application of spatio-temporal geostatistics in data analysis has become common in recent years (Cressie and Wikle, 2011; De Iaco et al., 2012). Zeng et al. (2014) proposed the use of spatio-temporal geostatistics in generating regional satellite-observed $X\text{CO}_2$ maps at high spatio-temporal resolution, and illustrated the method by applying it to the study area of China. However, extending this regional spatio-temporal geostatistical approach to the global land case is not straightforward, since the regional approach depends on uniform spatio-temporal trends and spatio-temporal correlation structures in the regional dataset. This assumption for global $X\text{CO}_2$ is often violated, as shown by Alkhaled et al. (2008). Instead of assuming a uniform trend as in the regional study, in this study we construct the spatio-temporal trend of $X\text{CO}_2$, including latitude-zonal seasonal cycle and global annual increase, by incorporating precise ground-based observations and model simulated data, which are shown to have well reproduced large scale features of the atmospheric CO_2 distribution (Cogan, 2012). We use the NOAA Earth System Research Laboratory (ESRL) global annual CO_2 growth rate (Conway and Tans, 2013) to determine the annual increase of $X\text{CO}_2$ as in Wunch et al. (2013), and fit the CT2013 $X\text{CO}_2$ time series in each latitude band using a set of annual harmonic functions to construct the latitude-zonal seasonal cycle. After detrending the data by excluding the modeled trend to get the residual dataset, we then calculate the spatio-temporal correlation structures between $X\text{CO}_2$ data in each 10° latitude zone since the variations in $X\text{CO}_2$ are similar zonally (Wunch et al., 2011a; Schneising et al., 2014). Finally the space-time kriging with a moving flexible kriging neighborhood (Haas, 1995) is implemented to obtain the global land maps of $X\text{CO}_2$. A brief introduction of spatio-temporal geostatistics is furnished in this section, followed by a introduction of the global land mapping approach for the ACOS-GOSAT $X\text{CO}_2$ data. The spatial-only method, as a comparision with spatio-temporal method, is also briefly introduced.

3.1   Spatio-temporal geostatistics for global land mapping

3.1.1   Spatio-temporal random field model

In spatio-temporal geostatistics, $X\text{CO}_2$ data (Z) can be modeled as a partial realization of spatial (**s**) and temporal (t) random functions $Z = \{Z(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{R}^2 \times \mathbb{R}\}$, where $\mathbb{R}$ is the real set. The spatio-temporal variation of global land $X\text{CO}_2$ data can be further modeled by

decomposing it into an inherent and deterministic trend component (m) and the residual component (R), given by,

$$Z(\boldsymbol{s}, t) = m(\boldsymbol{s}, t) + R(\boldsymbol{s}, t), \tag{1}$$

where $m(\boldsymbol{s}, t)$ is a deterministic space-time mean component that models the spatial trend and temporal trend including the seasonal cycle and annual increase, and $R(\boldsymbol{s}, t)$ is a stochastic residual component that represents an intrinsically stationary space-time error process (De Iaco et al., 2012), which will be used in the following geostatistical analysis.

3.1.2 Modelling of global land temporal trend of XCO_2

The deterministic spatio-temporal trend of X_{CO_2} consists of the inherent seasonal CO_2 cycle mainly affected by the biosphere activities (Wunch, 2013; Schneising, 2014) and an annual CO_2 increase mainly due to the anthropogenic fossil fuel emissions (IPCC, 2013). We use the long-term CT2013 data from the year 2000 to 2012 in each 2° latitude band to determine the latitude-dependent seasonal cycle. Following Wunch et al. (2013), we use the ESRL global annual CO_2 growth rate (Figure 4; Conway and Tans, 2013) to determine the annual increase of X_{CO_2} in the trend. As a result, the temporal trend in each latitude band can then be modeled by a simple linear function plus a set of annual harmonic functions (Kyriakidis and Journel, 1999; Tsutsumi et al., 2009). The spatial trend within each latitude zone is not considered here. The trend component m in Equation (1) can therefore be described as follows,

$$m(\boldsymbol{s}, t) = a_0(\boldsymbol{s}) + a_t(\boldsymbol{s}) + \sum_{i=1}^{4}(\beta_i(\boldsymbol{s})\sin(i\omega t) + \gamma_i(\boldsymbol{s})\cos(i\omega t)). \tag{2}$$

where $\omega = \frac{2 \cdot \pi}{T}$ and T is period of 122 time-units, t is time in time-unit, and a_0, β_{1-4} and γ_{1-4} are parameters to be estimated. a_t is the cummulative annual increase for each time-unit determined by the ESRL global annual CO_2 growth rate. i is the order of the harmonic functions, and we use 4 orders here to specially fit the annual cycle (Wunch et al., 2013), semi-annual oscillation (Jiang et al., 2012), seasonal variation and monthly variation of X_{CO_2} (Keppel-Aleks et al., 2011). The estimated trend component is then subtracted from the full dataset to yield the spatio-temporal residual component $R(\boldsymbol{s}, t)$.
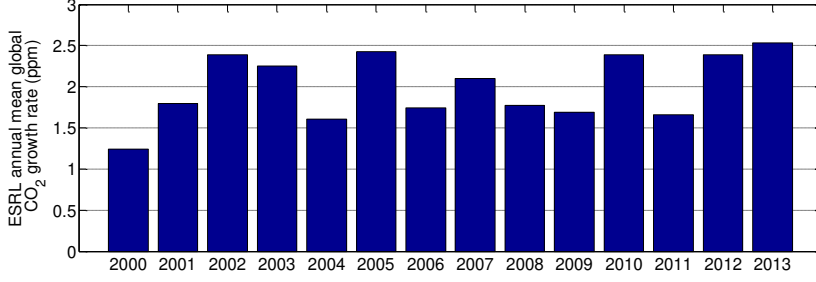
Figure 4. NOAA Earth System Research Laboratory (ESRL) annual mean global CO_2 growth rate for the year 2000 to 2013, published in ESRL-NOAA webpage: http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html.

3.1.3 Modelling of spatio-temporal correlation structure

In spatio-temporal geostatistical analysis, the optimal kriging prediction of $Z(\mathbf{s}_0, t_0)$ at an unobserved position $(\mathbf{s}_0, t_0)$ can be calculated as the linear weighted sum of the ACOS-GOSAT XCO$_2$ values that minimizes the mean squared prediction error. The weights for the observations are determined by the geometry of observations and the spatio-temporal covariance or variogram model, which characterizes the spatio-temporal correlation structure of the data. Therefore, estimating and modeling the covariance function or variogram of XCO$_2$ are crucial steps in kriging prediction for the global land mapping. Due to the non-uniform correlation structure of global XCO$_2$ variation (Alkhaled, 2008), in this study we divide the global land region into 10° bins, producing 11 latitude zones from 40°S to 70°N, and assume that the spatio-temporal correlation structures of the ACOS-GOSAT XCO$_2$ data within each zone are homogeneous and locally stationary (Haas, 1990, Schabenberger and Gotway, 2004) for the mapping method. The spatio-temporal empirical variogram (Cressie and Wikle, 2011; Zeng et al., 2014) of ACOS-GOSAT XCO$_2$ data is calculated from the residual component $R(\mathbf{s}, t)$ in each 10° latitude zone. The empirical variogram value $\hat{\gamma}_{ST}(r_s, r_t)$ at the lag (r_s, r_t) is given by

$$\hat{\gamma}_{ST}(\mathbf{r}_s, r_t) = \frac{1}{2|I(r_s, r_t)|} \sum_{I(r_s, r_t)} [R(\mathbf{s} + \mathbf{h_s}, t + h_t) - R(\mathbf{s}, t)]^2 \tag{3}$$

where $\mathbf{h_s} \in \mathbb{R}^2$, $h_t \in \mathbb{R}$, and $|I(\mathbf{r}_s, r_t)|$ is the cardinality of the set: $|I(\mathbf{r}_s, r_t)| = \{(\mathbf{s} + \mathbf{h_s}, t + h_t) \in H, (\mathbf{s}, t) \in H$ such that $\|\mathbf{r_s} - \mathbf{h_s}\| < \delta_s \wedge \|r_t - h_t\| < \delta_t\}$. (δ_s, δ_t) are the spatial tolerance and temporal tolerance. Once the empirical variogram has been constructed, a spatiotemporal variogram model, $\gamma_{ST}(\mathbf{r}_s, r_t)$, is fitted to it. As in Zeng et al. (2014), the spatio-temporal variogram model adopted here is a combination of the product-sum model

13

(De Iaco *et al.*, 2001 and 2010) and an extra global nugget model to capture the nugget effect (Cressie 1993; Cressie and Wikle, 2011), given by

$$\gamma_{ST}(\boldsymbol{h}_s, h_t;\ \theta_s, \theta_t, \kappa, N_{ST}) =$$

$$\gamma_S(\boldsymbol{h}_s; \theta_s) + \gamma_T(h_t; \theta_t) - \kappa \cdot \gamma_S(\boldsymbol{h}_s; \theta_s) \cdot \gamma_T(h_t; \theta_t) + N_{ST} \cdot \gamma_0(\boldsymbol{h}_s, h_t). \tag{4}$$

The following exponential model for the marginal variogram model $\gamma_S(\boldsymbol{h}_s)$ and $\gamma_T(h_t)$ is adopted,

$$\gamma(h; \theta = [C, a]) = \begin{cases} 0, & h = 0, \\ C \cdot \left(1 - \exp\left(-\frac{\|h\|}{a}\right)\right), & h \neq 0, \end{cases} \tag{5}$$

where $\gamma_S(\boldsymbol{h}_s; \theta_s)$ and $\gamma_T(h_t; \theta_t)$ are the marginal spatial and temporal variograms, $[\theta_s, \theta_t, \kappa, N_{ST}]=[C_s, a_s, C_t, a_t, \kappa, N_{ST}]$ are parameters to be estimated, and $N_{ST} \geq 0$, $C \geq 0$, $a \geq 0$. In the exponential model, a is called the range, where the variogram value stabilized at a value, C, the partial sill. N_{ST} is called nugget effect, which is the semi-variance value when spatial and temporal lag are closest to 0. The nugget effect is an important variogram parameter characterizing both the measurement error and micro-scale variability of the data (Cressie, 1993; Chatterjee et al., 2010). A larger nugget effect value typically indicates larger overall variability of microstructure in the data. For each latitude zone, the empirical variogram is calculated from Equation (3) and modeled using the variogram model provided by Equations (4) and (5). To avoid biased result in the edges of the study area, a 5° buffer zone is added in each latitude zone in both the south and north directions. All parameters are estimated simultaneously to overcome the limitation of conventional product-sum fitting (Josh et al., 2005). The nonlinear, weighted least square estimation technique (Cressie, 1985; Zeng et al., 2014) was used for parameter estimation in this study.

3.1.4 Prediction using space-time kriging with moving cylinder kriging neighbourhood

Based on the spatio-temporal variogram model, space-time kriging predictes the value $R(s_0, t_0)$ at unobserved location (s_0, t_0), from the stochastic residual component $R = \{R(\boldsymbol{s}, t), (s, t) \in D \times T\}$. Suppose that $R^*(s_0, t_0)$ is a prediction of $R(s_0, t_0)$, the kriging method predicts $R^*(s_0, t_0)$ as a linear weighted sum of residual data within a kriging neighborhood (Cressie, 1993; De Iaco *et al.*, 2012) in space and time relative to the prediction location $(\boldsymbol{s}_0, t_0)$. Assume that the used data number is $n(\boldsymbol{s}_0, t_0)$, then

$$R^*(\boldsymbol{s}_0, t_0) = \sum_{i=1}^{n(s_0,t_0)} \lambda_i(\boldsymbol{s}_0, t_0) R(\boldsymbol{s}_i, t_i), \text{ with } \sum_{i=1}^{n(s_0,t_0)} \lambda_i(\boldsymbol{s}_0, t_0) = 1. \tag{6}$$

where $\lambda_i(\boldsymbol{s}_0, t_0)$ is the weight assigned to a known observation $R(\boldsymbol{s}_i, t_i)$ so as to minimize the prediction error variance while maintaining unbiasedness of the prediction. The prediction error variance, which is a measurement of prediction uncertainty, is given by

$$\sigma^2 = \gamma_0^T \Gamma^{-1} \gamma_0 - \frac{\left(\mathbf{1}^T \Gamma^{-1} \gamma_0 - 1\right)^2}{\mathbf{1}^T \Gamma^{-1} \mathbf{1}}, \tag{7}$$

where $\Gamma(i,j) = \gamma\big(|\boldsymbol{s}_i - \boldsymbol{s}_j|, |t_i - t_j|\big)$, $\gamma_0(i,1) = \gamma(|\boldsymbol{s}_i - \boldsymbol{s}_0|, |t_i - t_0|)$, and $\mathbf{1}$ is $n \times 1$ unit vector. To reduce computational complexity and preserve local variability, data used in the prediction were searched within an appropriate spatio-temporal neighborhood centered on the predicting point, called the kriging neighborhood (Cressie, 1993; De Iaco et al., 2012; Chiles and Delfiner, 2012). Following the prediction approach in Zeng et al. (2014), the kriging neighborhood used will be a cylinder in space and time as described in Hass (1995). The radii of the initial search range are set to 300 km in space and 20 time-units, the increment lags are 10 km and 1 time-unit for each search process if the number of observations in the cylinder neighborhood is less than 20, and the search range radii limits are set to 500 km and 40 time-units. The position to be predicted will be left as missing data if the available data number within the search range of limit radius is less than 20 points, a threshold value adopted by NIES GOSAT Project (2011), which has been shown to be an appropriate number for $X\text{CO}_2$ mapping by Zeng et al. (2014). In kriging, extremely-high or -low data values may have a significant impact on the prediction if there are no other data nearby (NIES GOSAT Project, 2011) in space and time. In this study, we used the bias removal method recommended by Tukey (1977) and Hoaglin et al. (1986), based on the quartile of the data, to remove the extreme values of the residual component in each latitude zone, before implementing the space-time kriging prediction approach.

3.2 *Spatial-only geostatistics for global land mapping*

The spatial-only method implemented here for comparison is similar to that for generating monthly GOSAT Level 3 data (Watanabe et al., 2015). The whole $X\text{CO}_2$ dataset is firstly grouped into monthly datasets (4 years = 48 months), and then the spatial variogram for each month is calculated and modeled by the exponential variogram model (as in equation (5)). Finally, based on the monthly modeled variograms, spatial-only kriging is applied for mapping global land $X\text{CO}_2$ in each month.

3.3   Cross-validation based on Monte Carlo sampling

Cross-validation is a widely used method for assessing prediction accuracy of statistical models (Arlot, 2010), and it can therefore be used to assess the prediction accuracy of the global land mapping approach based on spatio-temporal geostatistics in this study. As above, we use the ACOS-GOSAT X_{CO_2} dataset $\{Z(\mathbf{s_i}, t_i), (\mathbf{s}, t) \in \mathbb{R}^2 \times \mathbb{R}, i = 1, \dots, n\}$. The Cross-validation is implemented by first removing a a small ratio of observation data $\{Z(\mathbf{s}_j, t_j), j = 1 \dots m\}$ and then making its prediction $\{\hat{Z}(\mathbf{s}_j, t_j), j = 1 \dots m\}$ using the remaining dataset. Finally, two data sets, the predicted data set $\{\hat{Z}(\mathbf{s}_j, t_j), j = 1 \dots m.\}$ and the corresponding original data set $\{ Z(\mathbf{s}_j, t_j), j = 1 \dots m. \}$ are obtained. In this study, the following four summary statistics from cross-validation were derived to assess prediction precision using the above two data sets: the correlation coefficient (r^2), the mean absolute prediction error (*MAPE*), the root mean square error (*RMSE*), and the percentage of estimation error (PEE) less than 1 ppm. MAPE and RMSE are defined by

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|Z(\mathbf{s}_j, t_j) - \hat{Z}(\mathbf{s}_j, t_j)\right|. \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}[Z(\mathbf{s}_j, t_j) - \hat{Z}(\mathbf{s}_j, t_j)]^2}. \tag{9}$$

A larger r^2 means a stronger linear association between the predictions and observations. Both *MAPE* and *RMSE* measure model accuracy, and conceptually *MAPE* provides a measurement of average bias for an individual prediction. Smaller values for these two variables indicate better performance. Moreover, the percentage of prediction error (PPE) less than 1 ppm provides a detailed description of the prediction errors. In this study, we carry out a Monte Carlo study by randomly sampling 5% of the data that will be left out in cross-validation, and repeating this process for 100 times. With the ensemble outputs from this Monte Carlo study, we can further assess the uncertainties in these four statistics from cross-validation and uncertainties in parameter estimation for variograms. This study is carried out for both spatial-only and spatial-temporal method, and we further conduct a comparison between them based on all the statistics from cross-validation.

3.4 Implementation workflow

In this study, multi-source datasets and different data processing techniches have been used to generate and evaluate the global land mapping dataset of XCO_2 from satellite observations using spatio-temporal geostatistics. As a conclusion of all the used methodologies, figure 5 shows a workflow chart for the spatio-temporal mapping process and the evaluation of the mapping data products.
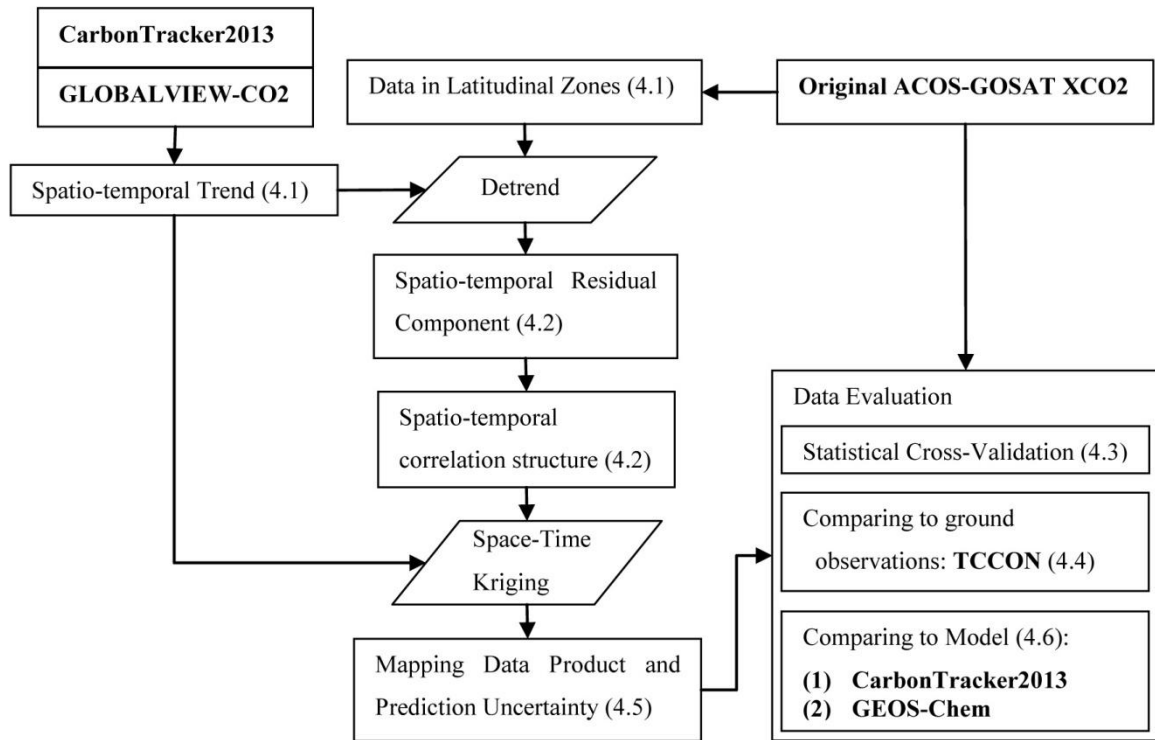
Figure 5. Workflow chart for the spatio-temporal mapping process and the evaluation of the mapping data products in this study. The input datasets are displayed in bold and the numbers in the boxes denote the section numbers in the manuscript.

4 Results and discussions

4.1 Spatio-temporal trend of XCO_2

As shown in Equation (1), the spatio-temporal trend of $X\text{CO}_2$ represents the deterministic mean component of the $X\text{CO}_2$ data, and can be obtained by fitting the model in Equation (2) to the CT2013 time series data in each latitude band and then combined to form the spatio-temporal trend component. The parameters in Equation (2) are estimated using CT2013 data from January 2000 to December 2012 and then used to calculate the trend component from June 2009 to May 2013. Figure 6(b) shows the estimated spatio-temporal trend from CT2013 $X\text{CO}_2$

data as a function of latitude and time. As a comparison, figure 6(a) shows the corresponding spatio-temporal distribution of X_{CO_2} from original ACOS-GOSAT data aggregated with a grid resolution of 5° in latitude and one month in time. From the figure we find that the two figures are generally in agreement, indicating the estimated trend from the CT2013 data well reproduce the spatial and temporal variation of X_{CO_2} from satellite observations. The results are also consistent with that from Schneising et al. (2014) based on long-term SCIAMACHY X_{CO_2} data and from WMO (2014) based on ground-based CO_2 observations. Stronger seasonal cycles can be identified in the Northern Hemisphere than in the Southern Hemisphere, indicating a stronger effect of the temporally varying imbalance between photosynthesis and respiration of vegetation in the North Hemisphere (Keppel-Aleks et al., 2011). In the Southern Hemisphere, the seasonal variation is relatively weak. An overall global annual increase can be clearly observed at a rate of about 2 ppm from both the CT2013 trend data and ACOS-GOSAT data, which is primarily caused by the world-wide consumption of fossil fuels (IPCC, 2013).

(a)   (b)

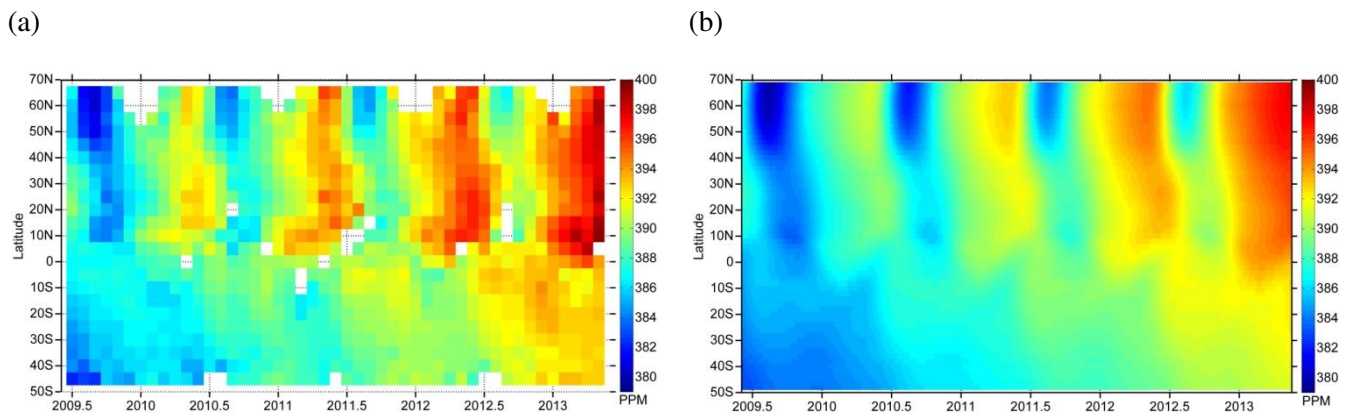Figure 6. (a) Overview of the global spatio-temporal distribution of X_{CO_2} as a function of latitude and time from original ACOS-GOSAT data aggregated with a grid resolution of 5° in latitude and one month in time, and (b) the corresponding deterministic spatio-temporal trend calculated from model simulation of CarbonTracker CT2013 X_{CO_2} data with a grid resolution of 2° in latitude and one time-unit (3-day) in time.

Table 2. Zonal statistics for the 11 bands, including the central latitude of each 10° latitude zone, the number of available data, the median and standard deviation of the ACOS-GOSAT XCO$_2$ data from June 2009 to May 2013 in the corresponding zone, the parameters of the spatio-temporal variogram model of the ACOS-GOSAT data, and the corresponding nugget/sill ratios in space and time, respectively. Uncertianties of the variogram related parameters, quantified by one standard deviation, are derived from the Monte Carlo study described in Section 3.3.

	Zone Center Latitude	Number of Data	Median (ppm)	Standard Deviation (ppm)	Estimated Parameters for Variogram Model $[\hat{C}_s, \hat{a}_s, \hat{C}_t, \hat{a}_t, \hat{\kappa}, \hat{N}_{ST}]$	Spatial Ratio Nugget/Sill	Temporal Ratio Nugget/Sill
Zone 1	65° N	10462	387.76	4.73	[1.84±0.02, 170.73±4.15, 3.82±0.04, 13.08±0.22, 0.26±0.00, 1.22±0.02]	0.40±0.01	0.24±0.00
Zone 2	55° N	16461	389.02	4.59	[1.43±0.01, 238.22±3.49, 2.44±0.02, 14.70±0.14, 0.27±0.00, 1.28±0.01]	0.47±0.00	0.34±0.00
Zone 3	45° N	49821	389.88	4.20	[1.35±0.01, 319.44±4.91, 1.84±0.01, 13.26±0.13, 0.34±0.00, 1.30±0.01]	0.49±0.00	0.41±0.00
Zone 4	35° N	44095	390.65	3.82	[1.31±0.01, 261.64±6.97, 2.04±0.01, 16.31±0.19, 0.41±0.00, 1.21±0.01]	0.48±0.00	0.37±0.00
Zone 5	25° N	18374	391.23	3.58	[1.13±0.01, 167.18±4.65, 3.16±0.05, 42.15±0.92, 0.32±0.00, 1.03±0.01]	0.48±0.00	0.25±0.00
Zone 6	15° N	8571	391.82	3.68	[1.44±0.01, 172.63±4.04, 2.75±0.14, 20.82±1.53, 0.32±0.00, 0.87±0.01]	0.38±0.00	0.24±0.01
Zone 7	5° N	2774	391.81	3.57	[2.38±0.07, 384.00±19.16, 2.36±0.21, 17.02±2.31, 0.38±0.01, 0.62±0.02]	0.21±0.01	0.21±0.01
Zone 8	5° S	6678	390.22	2.67	[1.24±0.01, 397.80±5.41, 1.37±0.04, 23.60±0.97, 0.63±0.00, 0.74±0.00]	0.37±0.00	0.35±0.01
Zone 9	15° S	31771	389.77	2.69	[0.91±0.00, 235.32±2.58, 1.14±0.02, 15.57±0.46, 0.79±0.00, 0.70±0.00]	0.43±0.00	0.38±0.00
Zone 10	25° S	44436	389.38	2.73	[0.90±0.00, 184.73±1.74, 1.21±0.03, 17.04±0.56, 0.78±0.00, 0.66±0.00]	0.43±0.00	0.35±0.00
Zone 11	35° S	20754	389.05	2.78	[0.98±0.01, 129.41±2.17, 1.34±0.02, 18.35±0.53, 0.69±0.00, 0.58±0.01]	0.37±0.00	0.30±0.00

Table 2 summarizes the zonal statistics of all the 10° latitude zones for modelling the spatio-temporal X_{CO_2} correlation structure, including the zonal central latitudes and the medians and standard deviations of the ACOS-GOSAT data in each latitude zone from June 2009 to May 2013. These description statistics are derived only from ACOS-GOSAT X_{CO_2} data over land since the ocean glint data are not recommended for science analysis (Nguyen et al., 2014). It is clear from Table 2 that the standard deviations gradually decrease from north to south, indicating the ACOS-GOSAT X_{CO_2} data have larger variability in the north, mainly due to a stronger seasonal cycle than that in the south (Miller et al., 2007). The median values in the Northern Hemisphere are generally larger than those in the Southern Hemisphere. However, in the high latitude of the Northern Hemisphere the median values are underestimated because fewer ACOS-GOSAT X_{CO_2} retrievals are available in this region during winter when X_{CO_2} is high. The uncertainties in the estimated variogram parameters are quantified by one standard deviation ($\pm 1\sigma$), and uncertainty value of 0.00 in Table 2 means that the uncertianty is acturally less than 0.005. From Tabel 2, we find that most of uncertianties are around 1% and all of them are smaller than 5%, indicating a robust variogram modelling for all the latitudinal zones.

4.2   *Spatio-temporal empirical variogram and variogram modelling*

Figure 7 shows the spatio-temporal empirical variograms calculated from full ACOS-GOSAT dataset for the 11 latitude zones, and the corresponding spatio-temporal variogram models fitted by the product-sum model as described in Equations (4) and (5). The statistics for the 11 latitude zones are given in Table 2. As shown in Cambardella et al. (1994) and Fu et al. (2014), the ratio of nugget to sill, expressed as percentage, can be used as an indicator to classify data dependence. In general, a ratio of less than 25% indicates a strong spatial or temporal dependence, between 25% and 75% indicates a moderate spatial or temporal dependence, and larger than 75% indicates a weak spatial or temporal dependence. The ratios of nugget to sill of both space and time for the 11 latitude zones are listed in Table 2. All the ratio values are less than 0.5 indicating all the zonal X_{CO_2} data show either moderate or strong spatial and temporal dependences. Moreover, the temporal ratio values are all lower than the spatial ratio values in the same latitude zone, showing stronger dependences in time between X_{CO_2} data than in space, justifying the use of spatio-temporal geostatistics in this study, which makes full use of the joint spatial and temporal dependencies between observations.

Zone 1

Zone 2

Zone 3

Zone 4

Zone 5

Zone 6

Zone 7

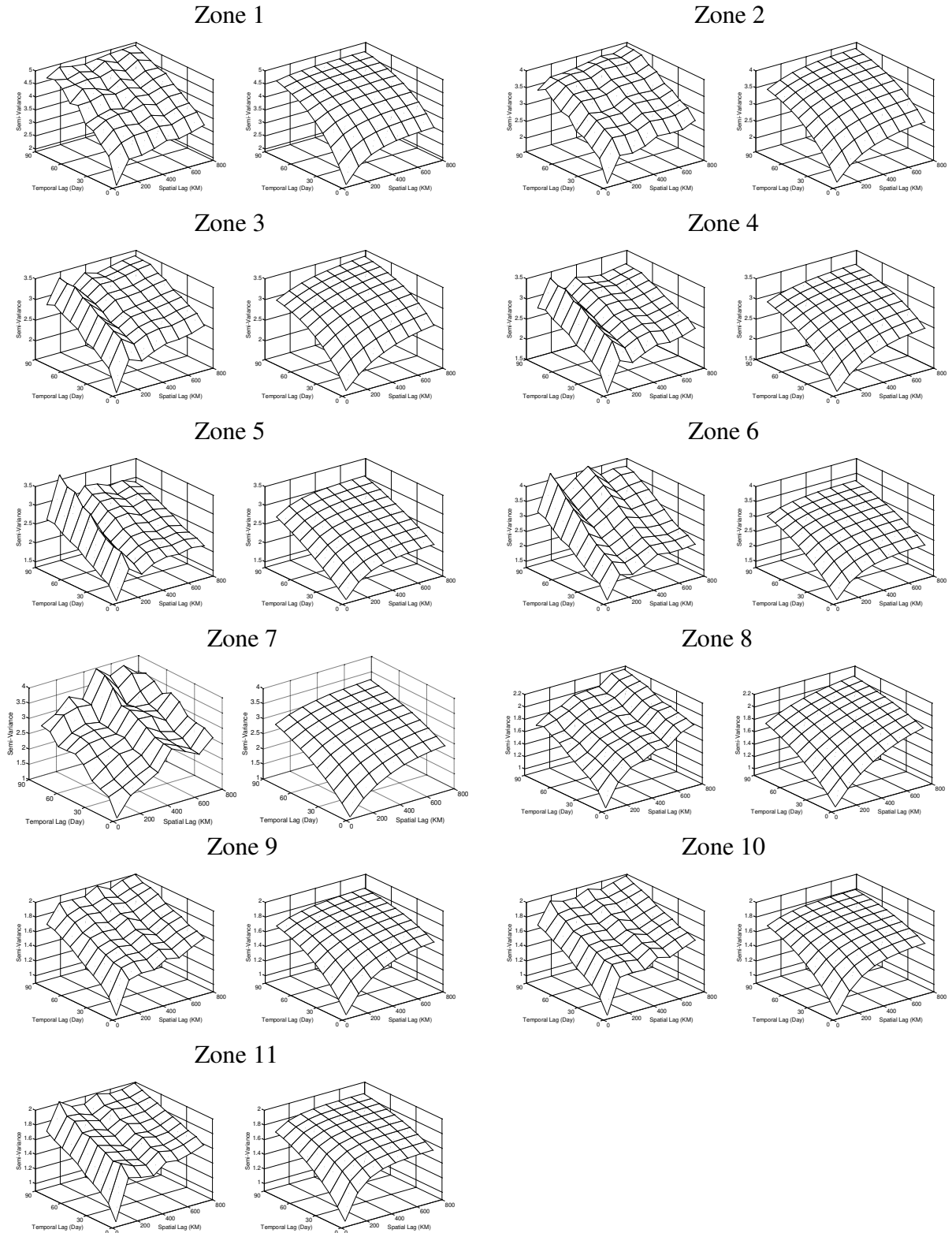Zone 8

Zone 9

Zone 10

Zone 11

Figure 7. For each latitude zone, the left plot shows the spatio-temporal empirical variogram of the ACOS-GOSAT XCO$_2$ residual data after the spatio-temporal trend is excluded, and the right plot shows its fitted variogram model using product-sum model in Equation (4).

From the estimated parameters for spatio-temporal variogram models shown in Table 2 and the experimental variograms and the corresponding fitted models in Figure 7, we find that the spatial and temporal marginal variogram models shown in Equation (5) vary considerably between latitude zones, and all the joint spatio-temporal variograms present different shapes, indicating differences of spatio-temporal correlation structure for different latitude zones. These differences illustrate the spatial heterogeneity of the observed space–time correlation structure in global ACOS-GOSAT X_{CO_2} data. From Table 2, furthermore, we find that the nugget effects, a measure of both the measurement error and micro-scale variability of the X_{CO_2} data, in the Northern Hemisphere are generally greater than that in the Southern Hemisphere, which indicates greater overall data variability in the Northern Hemisphere.

4.3 Cross-validation and comparison with spatial-only method

Results from cross-validation based on Monte Carlo sampling are shown in Table 3. For our developed spatio-temporal method, the r^2 is 0.94 ± 0.00, indicating a very strong correlation between the original ACOS-GOSAT data and the predictions. Moreover, the MAPE is 0.93±0.01 ppm, indicating the averaged absolute error of each prediction is less than 1 ppm. From the prediction errors, we find that more than 65.53±0.42% of the prediction errors are less than 1 ppm. These summary statistics from cross-validation for the global land with four years of ACOS-GOSAT data are better than those found for China as a study region with only two years of ACOS-GOSAT data by Zeng et al. (2014). The improvement may be due to the facts that in this study more satellite data are available, and the zonal characteristics of X_{CO_2} are fully considered, while Zeng et al. (2014) simply assumed a uniform spatio-temporal trend. All the statistics from this cross-validation study indicate that the global land mapping method based on spatio-temporal geostatistics is effective, and we can therefore generate precise maps of X_{CO_2} using this method from the full dataset of ACOS-GOSAT observations.

Moreover, we compare prediction precision between the spatio-temporal and spatial-only method in terms of cross-validation statistics. For spatial-only method, the four statistcs are shown Table 3, and these statistics indicate less precision and larger prediction error for the spatial-only mehtod than the spatio-temporal method. Therefore, these statistics justify the improvement of the developed spatio-temporal method over the traditional spatial-only method in terms of cross-validation.

Table 3. Summary statistics for the developed spatio-temporal method and spatial-only method from global land cross-validation based on Monte Carlo samping technique, including correlation coefficient (r^2), Mean Absolute Prediction Errors (MAPE), Root Mean Square Error (RMSE), and Percentage of Prediction Error within 1 ppm (PPE1) for the global mapping approach. Both mean and one standard deviation of the ensemble outputs from cross-valiation are shown.

r^2	MAPE (ppm)	RMSE (ppm)	PPE1 (%)
Spatio-temproal method			
0.94±0.00	0.93±0.01	1.28±0.01	65.53±0.42
Spatial-only method			
0.92±0.00	1.06±0.01	1.41±0.01	58.62±0.43

4.4 Comparison with the Total Carbon Column Observing Network (TCCON)

TCCON provides an essential data source for validating satellite X_{CO_2} retrievals and the data have been widely used for validation of satellite X_{CO_2} retrievals using different retrieval algorithms in several previous studies (e.g., Morino et al., 2011; Wunch et al., 2011a; Butz et al., 2011; Cogan et al., 2012; Nguyen et al., 2014; Dils et al., 2014). In this study, we perform comparison between TCCON X_{CO_2} data and the X_{CO_2} time series reconstructed using the developed global land mapping approach from ACOS-GOSAT X_{CO_2} observations. As suggested by Rodgers and Connor (2003), when comparing two observations from different instruments, the retrievals should be calculated using a common a priori profile, and the smoothing effect of the retrievals should be considered by applying the averaging kernels. A detailed description of the use of the averaging kernel for comparing ACOS-GOSAT and TCCON data can be found in Wunch et al. (2011a) and Nguyen et al. (2014). Following the instructions in Sect. 2.1 of Nguyen et al. (2014), we apply the ACOS-GOSAT averaging kernel equation to TCCON data from the 16 chosen sites to obtain what ACOS-GOSAT would have retrieved at the TCCON sites assuming the TCCON profile as "truth". The TCCON data are chosen using coincidence criteria of within ± 2 hours of GOSAT overpass time, which is about 13:00 local time for most sites, and 3-day (one time-unit) median is calculated for the comparison if the observation number within the time-unit is at least 3. Since the X_{CO_2} estimates at each TCCON site are obtained from a kriging neighborhood within 500 km in space, as described in Section 3.1, the original ACOS-GOSAT X_{CO_2} data within 500 km of each TCCON site will be chosen, and similarly, the 3-day median will be

calculated for comparison. As a result, Figure 8 shows the comparison of the X_{CO_2} estimates from this study, the coincident TCCON X_{CO_2} 3-day median data, and the coincident ACOS-GOSAT X_{CO_2} observations and their 3-day medians.

As the global land mapping method is a data-driven prediction approach, the resulting predictions will rely heavily on the original data. From Figure 8, it can be observed that the predictions go through the original satellite observations and agree well with the median data. Table 4 gives a more detailed summary statistics in terms of number of coincident data pairs, averaged bias, averaged absolute bias and standard deviation of the bias between the mapping dataset and TCCON (hereafter referred to as Mapping–TCCON), as well as between ACOS− GOSAT 3-day median and TCCON (hereafter referred to as ACOS–TCCON). From Table 4, we find that the numbers of coincident data pairs for Mapping–TCCON range from 32 at Influx to 416 at Lamont, while the numbers for all sites are smaller for ACOS–TCCON which is based on the 500 km geographical coincident criteria. The difference of numbers of coincident data pairs can be obviously seen in Figure 9, which shows plots of the statistics in Table 4. For Mapping–TCCON the averaged biases, except Four Corners (1.62 ppm), are within 1 ppm, and standard deviations for all the sites are close, ranging from 0.66 to 1.58 ppm. In terms of averaged absolute biases, Mapping–TCCON are smaller than ACOS– TCCON for 13 out of 16 stations. Overall, the X_{CO_2} predictions at all TCCON sites are in good agreement with the TCCON data, with overall averaged bias of 0.01 ppm, overall averaged absolute bias of 0.94 ppm and standard deviation of 1.22 ppm. For ACOS-TCCON, the overall averaged bias is -0.35 ppm and the corresponding standard deviation is 1.38 ppm, which is consistent with result from the ACOS-GOSAT X_{CO_2} data after filtering and bias correction as described in ACOS data user's guide (pp. 13). Unfortunately, both TCCON and ACOS-GOSAT observation are affected by clouds, so when few ACOS-GOSAT data are available, few TCCON retrievals are available either. However, with all possible plots and statistics given, we believe Figure 8 and Table 4 have provided a robust assessment of the mapping dataset compared to TCCON.
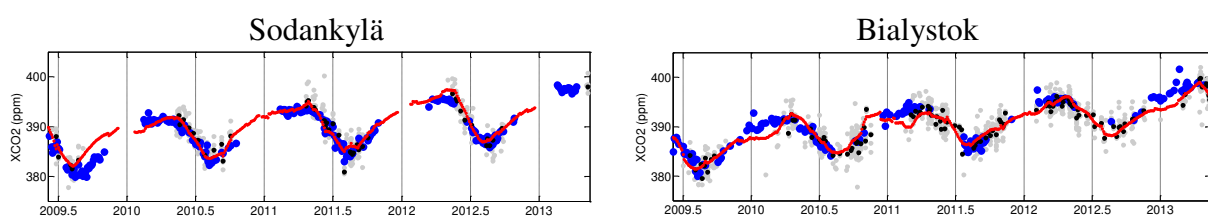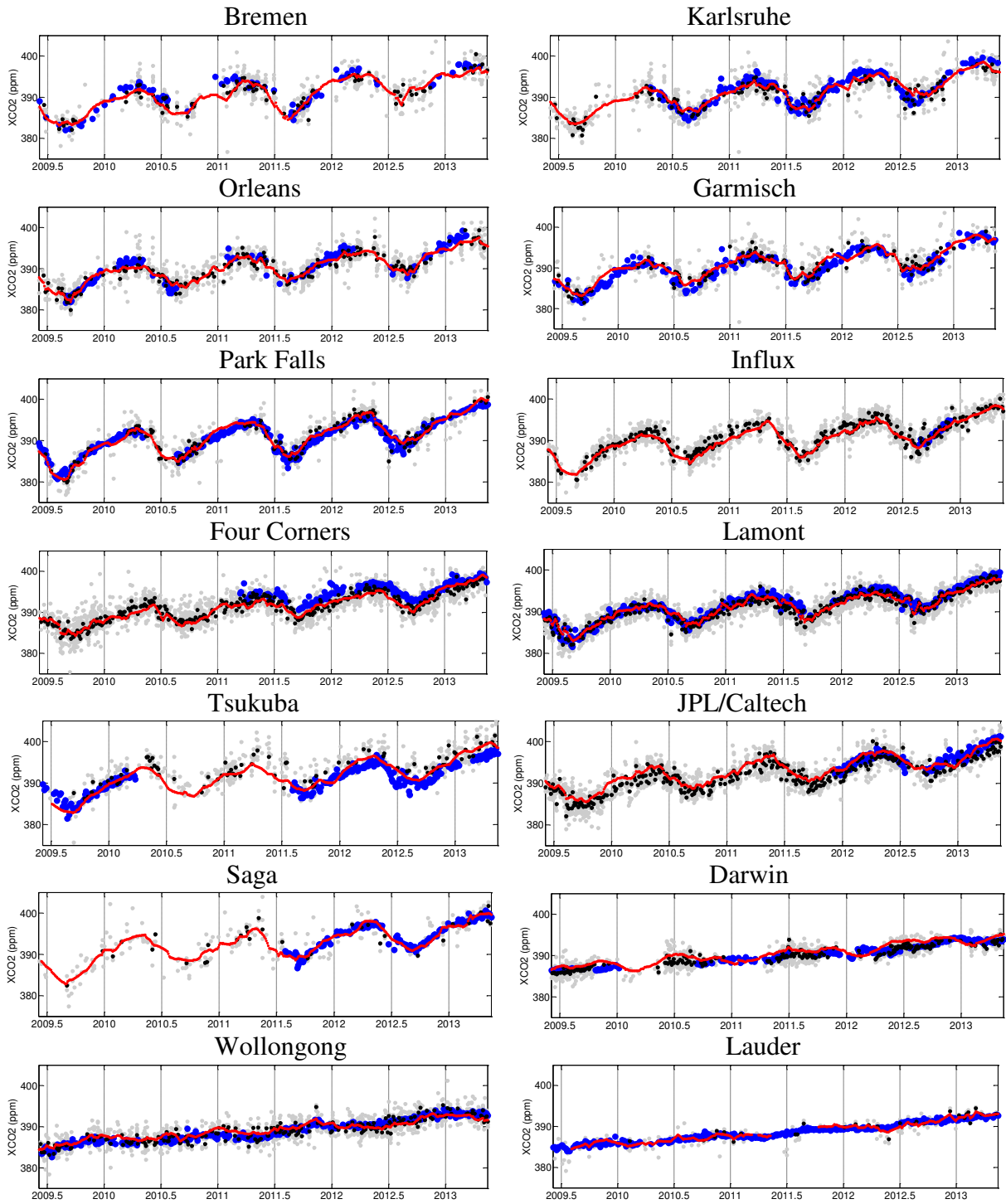
Figure 8. Temporal variation comparison for the 16 TCCON sites. As shown in these panels, the original ACOS-GOSAT X_{CO_2} retrievals within 500 km of the TCCON site are in grey dots, the corresponding medians are in black dots when at least 3 data points are available within the time-unit. The TCCON data, smoothed by applying the ACOS-GOSAT averaging kernel, are indicated by blue circles. The data are chosen using coincidence criteria of within ± 2 hours of GOSAT overpass time, and a 3-day (one time-unit) median is calculated for the comparison if the number of data points within the time-unit is at least 3. The predicted TCCON site X_{CO_2} time series using the global land mapping approach are indicated by the red line.
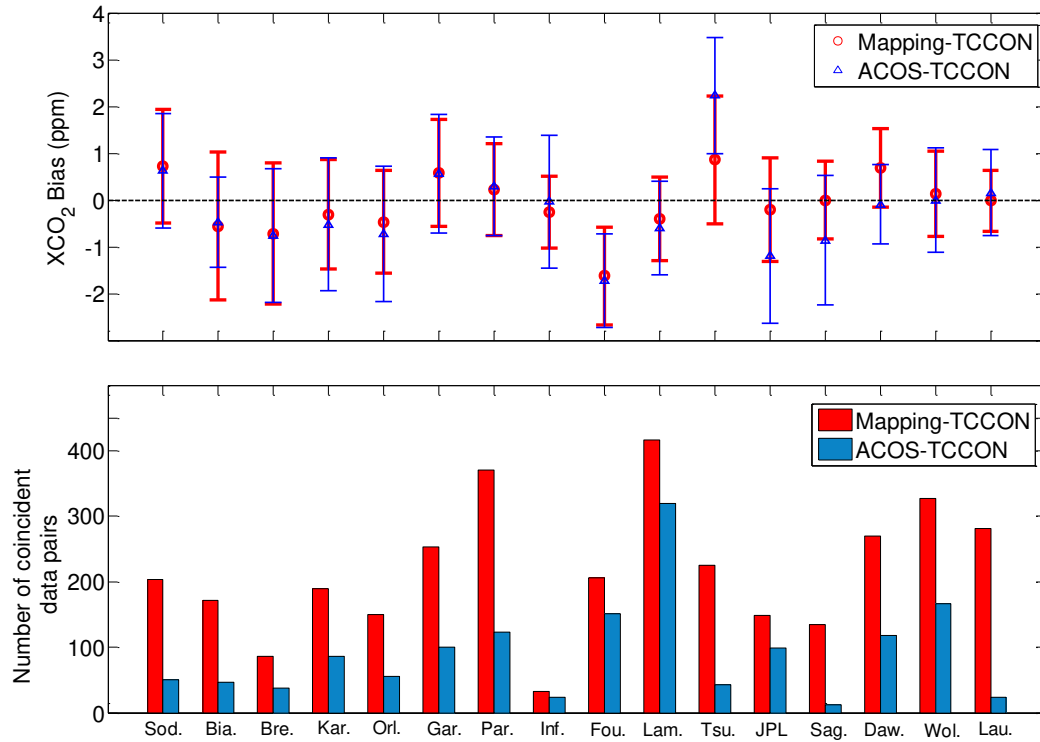
Figure 9. Statistics of comparison between the geostatistical mapping results and the TCCON data (smoothed by applying the ACOS-GOSAT averaging kernel), and between the ACOS-GOSAT 3-day median X_{CO_2} data and TCCON data, hereafter referred to as Mapping−TCCON and ACOS−TCCON, respectively. The statistics are also shown in Table 4. The top panel shows the comparison of X_{CO_2} biases between Mapping−TCCON and ACOS−TCCON, while the down panel shows the comparison of numbers of coincident data pairs between them.

From Figure 9, the averaged biases of Mapping−TCCON for most sites are generally closer to zeros than ACOS−TCCON, with overall averaged bias of 0.01 ppm for Mapping−TCCON and -0.35 ppm for ACOS−TCCON. It indicates that the mapping dataset is more accurate than ACOS-GOSAT median data when compared to TCCON. However, the standard deviations of the bias are almost the same between Mapping−TCCON and ACOS−TCCON, with overall standard deviation of 1.22 ppm and 1.38 ppm, respectively. Moreover, the dependence of the developed mapping method on the original satellite data can be clearly observed in Figure 9, where the averaged bias of Mapping−TCCON is positively correlated to that of ACOS−TCCON, indicating that the prediction precision of the global land mapping will increase as X_{CO_2} retrieval precision of ACOS-GOSAT improves.

Furthermore, the global land mapping method developed in this study also provides an effective spatio-temporal geostatistical method for collocating satellite X_{CO_2} data with the ground observations, as investigated in Nguyen *et al*. (2014). More ground observations should be used to validate this mapping dataset as more and more TCCON sites are available

in the future, particularly for the vast region in Africa and Asia where TCCON data are still unavailable. At the same time, the quality of TCCON retrievals is also improving and therefore provide more accurate data for validation of the satellite retrievals. For example, as noted in the data description in TCCON website (https://tccon-wiki.caltech.edu/), all Four Corners TCCON GGG2012 version X_{CO_2} retrievals are high biased by ~0.28% or ~1.1ppm due to an error in the surface pressure used in the retrievals. The error will be fixed in the new version and this will lead to a ~1.1 ppm offset from the absolute bias (1.68 ppm) compared with satellite data as shown in Table 4.

Table 4. Statistics of comparison between the geostatistical mapping results and the TCCON data (smoothed by applying the ACOS-GOSAT averaging kernel), and between the ACOS-GOSAT 3-day median X_{CO_2} data and TCCON data, hereafter referred to as Mapping−TCCON and ACOS−TCCON, respectively. The statistics include the location information, and statisics of the differences computed by subtracting the TCCON median X_{CO_2} from the mapping predictions and ACOS-GOSAT X_{CO_2} 3-day medians, respectively, at the corresponding TCCON sites.

Sites	Location [longitude, Latitude]	Mapping−TCCON				ACOS−TCCON			
		coincident data pairs	Averaged biases (ppm)	Averaged Absolute bias (ppm)	Standard deviation (ppm)	coincident data pairs	Averaged biases (ppm)	Averaged Absolute bias (ppm)	Standard deviation (ppm)
Sodankylä	[26.63, 67.37]	204	0.72	1.05	1.22	50	0.62	1.18	1.22
Bialystok	[23.03, 53.23]	171	-0.56	1.30	1.58	47	-0.47	0.77	0.96
Bremen	[8.85, 53.10]	86	-0.72	1.28	1.51	38	-0.76	1.17	1.43
Karlsruhe	[8.44, 49.10]	190	-0.30	0.97	1.17	86	-0.52	1.09	1.42
Orleans	[2.11, 47.97]	150	-0.47	0.97	1.10	56	-0.72	1.28	1.45
Garmisch	[11.06, 47.48]	253	0.57	1.02	1.14	100	0.56	1.09	1.26
Park Falls	[-90.27, 45.94]	371	0.22	0.78	0.97	123	0.30	0.82	1.05
Influx	[-86.00, 39.86]	32	-0.26	0.63	0.76	23	-0.04	1.02	1.42
FourCorners	[-108.48, 36.80]	206	-1.62	1.68	1.05	151	-1.72	1.72	1.00
Lamont	[-97.49, 36.60]	416	-0.41	0.79	0.90	320	-0.60	0.91	1.00
Tsukuba	[140.12, 36.05]	225	0.86	1.29	1.37	43	2.24	2.24	1.24
JPL/Caltech	[-118.18, 34.20]	148	-0.21	0.90	1.11	99	-1.19	1.58	1.44
Saga	[130.29, 33.24]	135	-0.01	0.67	0.83	12	-0.86	1.21	1.39
Darwin	[130.89, -12.43]	270	0.69	0.91	0.84	118	-0.10	0.66	0.85
Wollongong	[150.88, -34.41]	327	0.14	0.76	0.91	167	-0.00	0.88	1.11
Lauder	[169.68, -45.05]	281	-0.01	0.53	0.66	23	0.16	0.71	0.92
Overall		3465	-0.01	0.94	1.22	1456	-0.35	1.10	1.38

4.5 Assessment of mapping results

4.5.1 Generated mapping dataset

Examples of spatial distribution of the mapping X_{CO_2} data from 2010 to 2012 are shown in Figure 10 by taking their seasonally-mean values. In summer, globally, a relatively lower X_{CO_2} value can be observed in the high latitude region of the Northern Hemisphere than that in the low latitude region, mainly due to the large CO_2 uptake by the growing forest in North Hemisphere. In other seasons, the X_{CO_2} value is higher in the Northern Hemisphere than that in the Southern Hemisphere. The geostatistical mapping data show wider coverage and more detailed spatial distributions of X_{CO_2}, especially in some key regions, such as south China and Central Asia, than the original satellite retrievals with sparse coverage as shown in Figure 1.

4.5.2 Prediction uncertainty

For each geostatistical prediction, the corresponding kriging variance, defined in Equation (7), can be obtained at the same time as a measure of prediction uncertainty. Kriging variance is calculated without the knowledge of the true X_{CO_2} distribution, and it is determined by both the density of ACOS-GOSAT X_{CO_2} observations and the data variability in the neighborhood of the prediction location. Generally speaking, areas with more homogeneous X_{CO_2} variation and denser observations surrounding the prediction location will have lower prediction uncertainty (Hammerling *et al*., 2012a; Chiles and Delfiner, 2012). Figure 11 shows the spatial and temporal variations of the prediction uncertainty from the global land mapping. The prediction uncertainty is unevenly distributed in space, which is generally in negative correlation with the spatial distribution of observation number density as shown in Figure 2(a). Predictions in the Southern Hemisphere, where large number of observations are available with lower variability, present a relatively smaller uncertainty, which can clearly be seen in Figure 11(b). However, regions with fewer observations and larger variability present a larger prediction uncertainty, as in the high latitude Northern Hemisphere. Temporally, the time series of spatial averaged prediction uncertainty vary from 1.0 to 2.6 ppm and we find them negatively correlated with time series of available ACOS-GOSAT X_{CO_2} data number in all latitudinal zones (not shown).

To assess how well the prediction uncertainties describe the prediction errors, we use the ensemble outputs from cross-validation based on Monte Carlo sampling technique, and calculate the coverage percentage of prediction errors by the prediction uncertainties for our

developed spatio-temporal method. This coverage percentage is 70.35±0.39%, which indicates more than 70% of the prediction errors can be constrained by the prediction uncertainty map shown in Figure 11. For the spatial-only method, it is 69.04±0.42%, which slightly smaller but no significat difference from that for the spatio-temporal method.
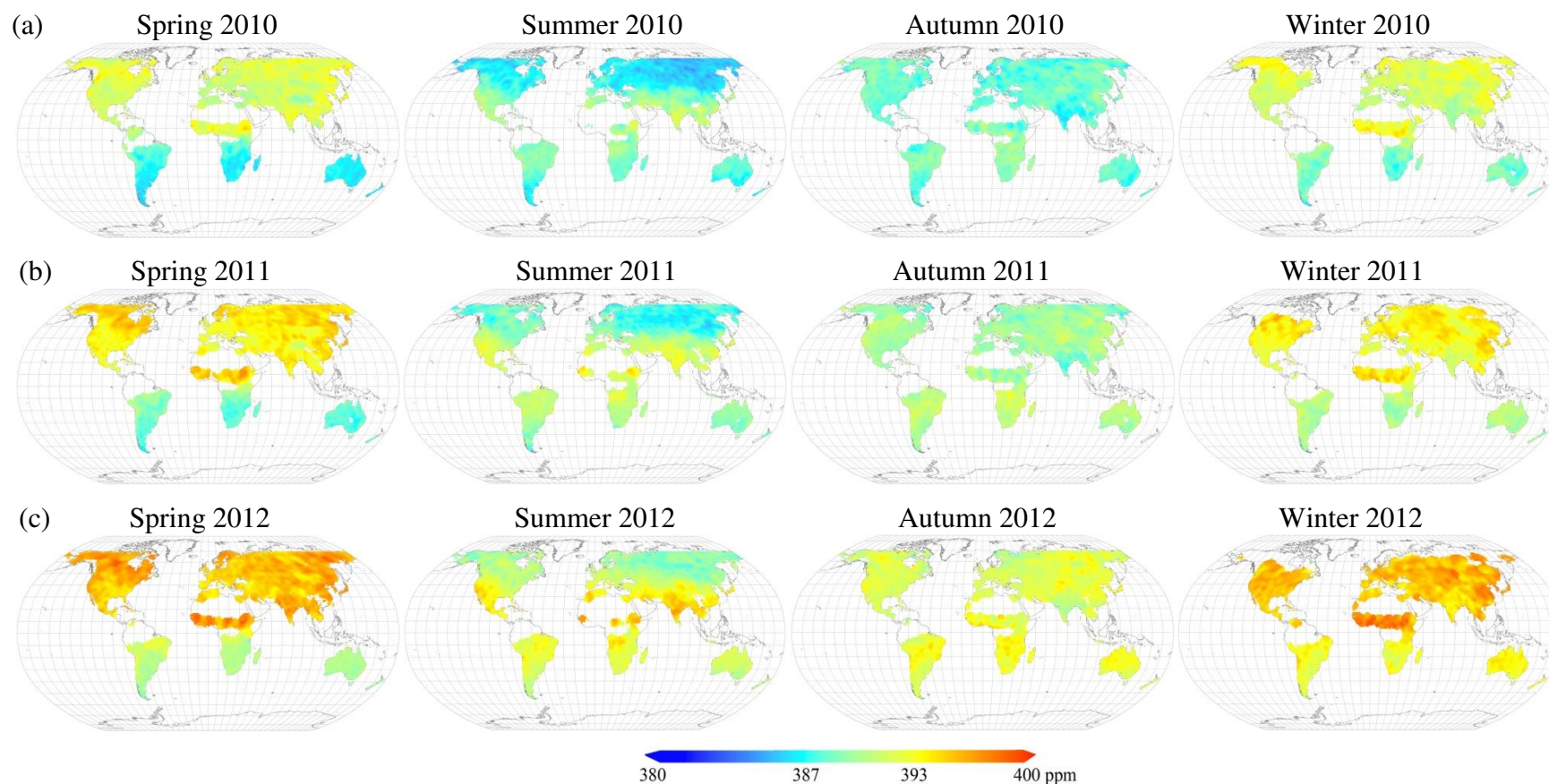
Figure 10. Spatial distribution of seasonally-mean X_{CO_2} from global land mapping for three years of 2010 in panel (a), 2011 in panel (b) and 2012 in panel (c). Four images in each panel corresponds to spring, summer, autumn and winter from left to right, respecitvely. These global land mapping seasonally averaged results in 1° by 1° grid are obtained by calculating the seasonal mean from the geostatistical mapping results when at least one data is available for each of the three months in that season. As usual, the season spring includes three months of March, April and May, summer includes June, July and August, autumn includes September, October and November, and winter includes December, next year January and February.
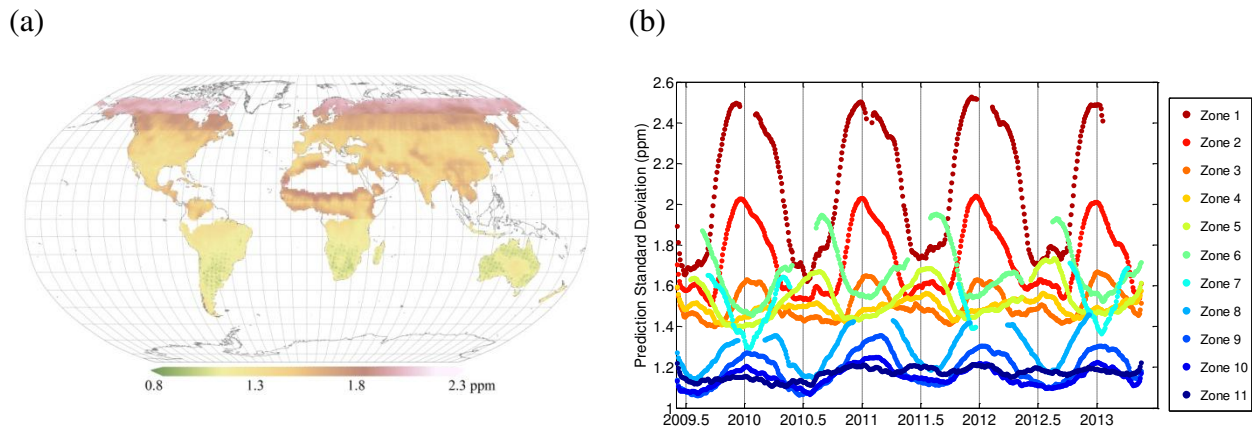
30

Figure 11. Kriging standard deviations of prediction, a measurement of prediction uncertainty. (a) spatial distribution, which is obtained by averaging the kriging standard deviations over all time-units in the global land. (b) Temporal distribution for the 11 zones, which are averaged for each time-unit of each zone from June 2009 to May 2013. For each time-unit and each zone, the averages are calculated only when more than 50% of the data are available.

4.6 Comparison with model simulations of X_{CO_2}

We compare the original ACOS-GOSAT X_{CO_2} data, the global land mapping X_{CO_2} dataset, and the model simulations of X_{CO_2} from CT2013 and GEOS-Chem by illustrating and analyzing their spatio-temporal variation as a function of latitude and time. The X_{CO_2} value for each specific latitude bin and time bin is obtained by calculating the zonal mean of X_{CO_2} within the time lag. Figure 12(a) shows the result for the global land mapping data, with latitude bins of 1 degree and temporal bins of 3-day. The zonal mean data in Figure 12(a) in the subtropical region are excluded when the available data number within the zonal band are less than 15, which is the number of pixels of equatorial Africa, to avoid bias in calculation especially around the equatorial region where few mapping data are available. Figure 12(b) shows the result from original ACOS-GOSAT X_{CO_2} observations, with latitude bins of 5 degree and temporal bins of one month, which is the same with Figure 6(a). Figure 12(c) and (d) are the results for CT2013 from June 2009 to December 2012 and GEOS-Chem for 2010, respectively. Strong seasonal X_{CO_2} variations in northern high and mid-latitudes can be observed from these panels, but in the Southern Hemisphere they are indistinct. In addition, the steady annual increase of global X_{CO_2} mainly due to the burning of fossil fuels can be easily identified in both the Northern and Southern Hemisphere.
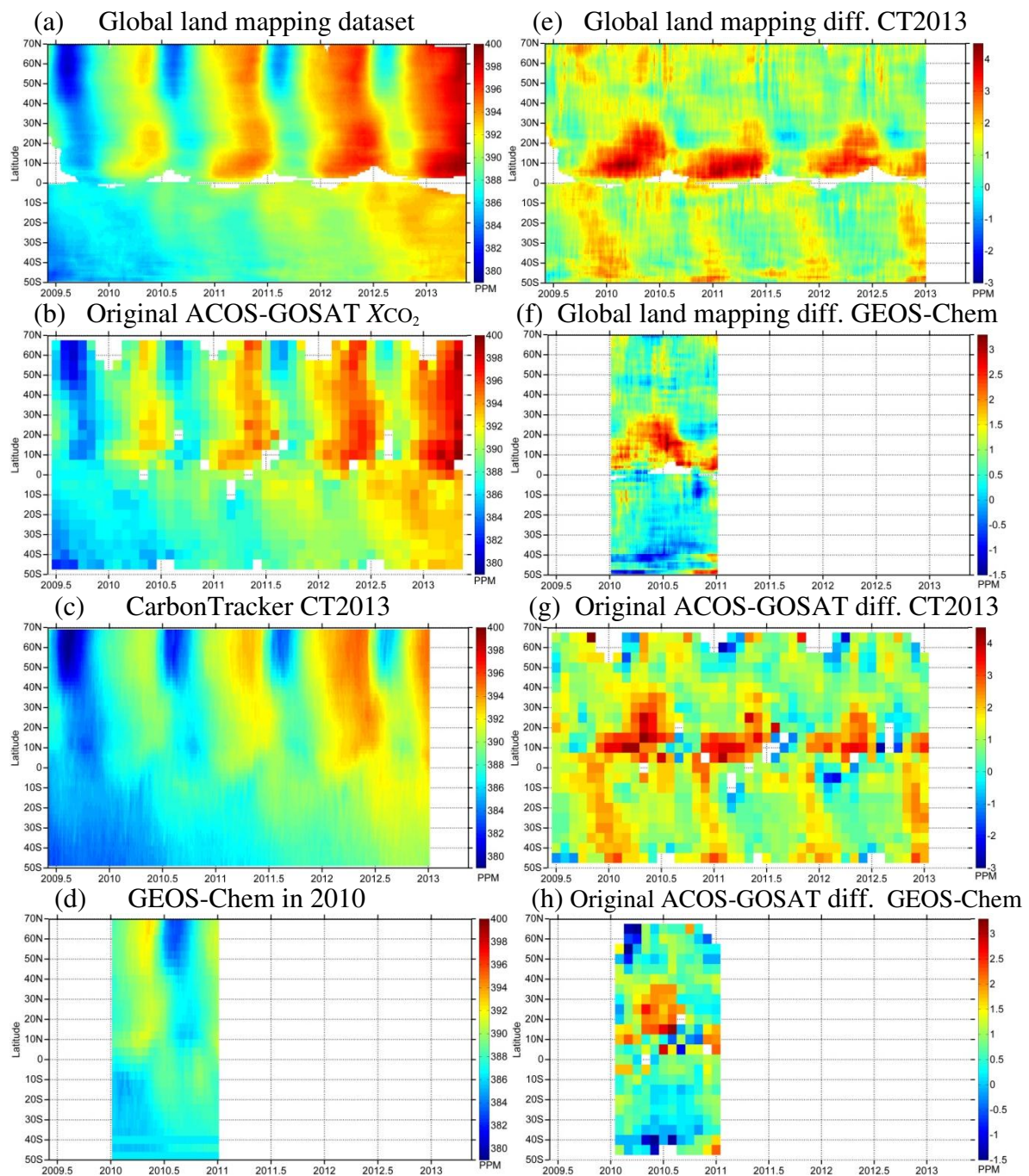
Figure 12. Overview of the global spatio-temporal distribution of X_{CO_2} as a function of latitude and time, from (a) global land mapping dataset with a grid resolution of 1° in latitude and one time-unit in time, (b) original ACOS-GOSAT data with a grid resolution of 5° in latitude and one month in time, and model simulations from both (c) CarbonTracker CT2013 data with a grid resolution of 2° in latitude and one time-unit in time and (d) GEOS-Chem data in 2010 with a grid resolution of 4° in latitude and one time-unit in time, and an overview of the differences of global spatio-temporal distribution of X_{CO_2} between (e) global land mapping dataset and CarbonTracker2013 data, (f) global land mapping dataset and GEOS-Chem data, (g) original ACOS-GOSAT data and CarbonTracker CT2013 data, and (h) original ACOS-GOSAT data and GEOS-Chem data.
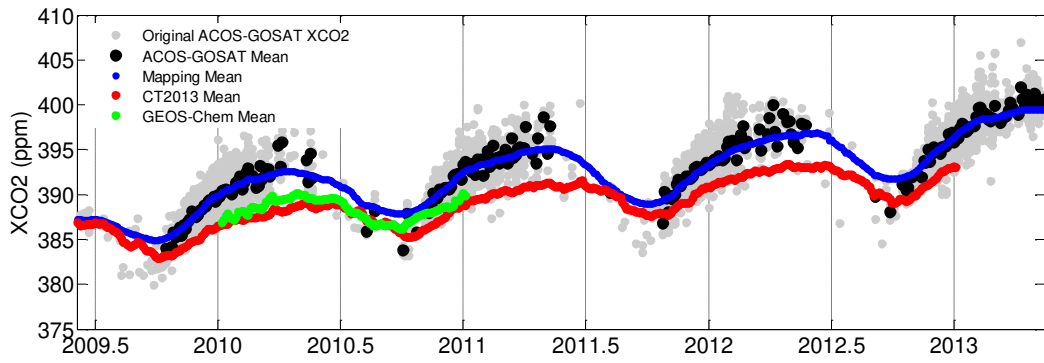
32

Figure 13. The panel shows the $X\mathrm{CO_2}$ time series in central Africa within 5°N to 15°N in latitude and 20°W to 40°E in longitude. The original ACOS-GOSAT data indicated by grey dots and its mean in black dots calculated when at least 3 data points are available within the time-unit, The mapping dataset mean is indicated by the blue line, CarbonTracker CT2013 data mean by the red line and GEOS-Chem data mean by the green line.

| (a) ACOS-GOSAT $X\mathrm{CO_2}$ in 1°x1° Grids | (b) Deterministic Trend Component from CT2013 | (c) Stochastic Predicted Residual Component |

Spring 2010
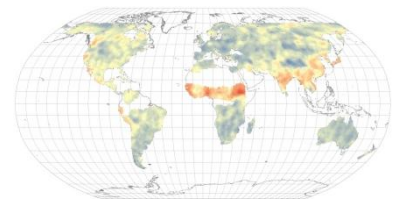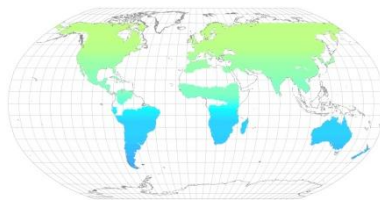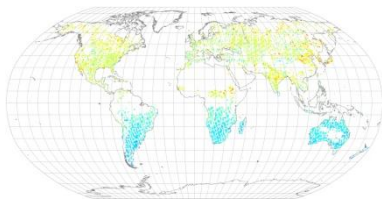
Autumn 2010
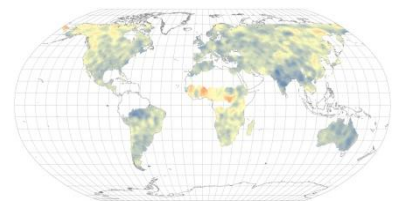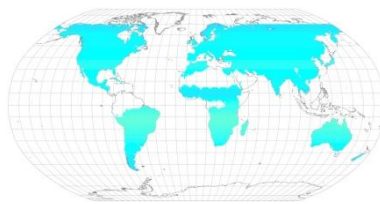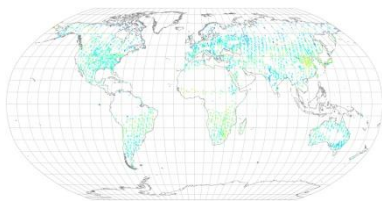
Figure 14. (a) Original ACOS-GOSAT $X\mathrm{CO_2}$ distributions averaged over 1°x1° grids, and the corresponding mapping decompositions, including (b) averaged deterministic trend component from CT2013 and (c) averaged stochastic residual component for Spring 2010 and Autumn 2010, respectively.

Figure 12(e) and (f) show the differences by subtracting CT2013 in Figure 12(c) and GEOS-Chem in Figure 12(d) respectively from the mapping dataset in Figure 12(a). For most regions, the differences are generally within 2 ppm, indicating the spatio-temporal variation of

the mapping dataset and the model simulations are in agreement for most of the region. However, large difference of 3-4 ppm can be identified in the region of 5-15° N during late winter (January and February) and early spring (March) each year. This latitude zone mainly covers the central African tropical region. A similar result can also be concluded from Figure 12(g) and (h), which show the difference by subtracting the result of both models from the original ACOS-GOSAT dataset in Figure 12(b). In Figure 13, we specifically analyze the data in this central African land region within 5 to 15° N in latitude and 20°W to 40°E in longitude as specified in Figure 2(a), by comparing the time series of the original satellite observations, mapping dataset, CT2013 data and the GEOS-Chem data. As expected, the predictions are consistent with the original satellite observations and their mean values, indicating the data-driven feature of the geostatistical prediction method although the trend component is derived from the CT2013 model simulation data. Moreover, the satellite data and model simulations generally agree during summer time, while larger difference can be observed from January to March, which is consistent with the conclusions from Figure 12. This difference may be due to the deficiency of CarbonTracker in constraining tropical biomass burning CO_2 fluxes because of the lack of atmospheric observations due to the sparse observational network in this region (Peter *et al*., 2007). Moreover, the thin clouds that often occur in the tropics may have an impact on the retrievals of satellite CO_2 and contribute to the X_{CO_2} difference in the tropical region, as described in the study comparing X_{CO_2} data from SCIAMACHY and CarbonTracker by Heymann *et al*. (2012).

From Figure 13, we can see that there is a big difference in the amplitude of seasonal cycle between CT2013 data and the mapping data eventhough the temporal trend in the geostatistical prediction is derived from CT2013. To understand more on the trend data, Figure 14 shows the decomposition of the mapping data into trend component and residual component for two seasons of Spring 2010 and Autumn 2010, when the differences are the largest and smallerst, respectively. We can see that the trend component (panel (b)) generally represents the global meridian trend of the original data (panel(a)), and the predicted residual component, on the other hand, well presents the local variation of X_{CO_2}. In Spring 2010, large residual value can be seen in central Africa and southern and eastern Asia because the trend components are not well capturing the small scale variations in these two regions where strong natural emission and anthropogenic emission domitate, respectively. While in Autumn 2010, the residuals are small, indicating consistency between the model simulation and satellite observations, which can also be seen from Figure 13. Figure 14 indicates that the

34

contribution of predicted residual component depends on the performance of the trend componen. If the trend component do not performs well then the residual component contribution will be large so that the final prediciton results will always be consistent with the observations.

5 Conclusions

Global geostatistical mapping of satellite-observed X_{CO_2} data is a useful gap-filling method for dealing with satellite data that are sparse and irregularly distributed in space and time. In this study, we have demonstrated a valid method based on spatio-temporal geostatistics for global land mapping of satellite X_{CO_2} observations from ACOS-GOSAT. We find that the method presents a significant high correlation coefficient of 0.94 between the predictions and observations in cross-validation, and moreover, the mapping results are in good agreement with the data from TCCON sites with overall bias of less than 0.01 ppm and standard deviation of the difference of 1.22 ppm, which show the effectiveness of the global land mapping method developed in this study. Furthermore, we demonstrate the difference and similarity between the geostatistical mapping of satellite data and model simulations of X_{CO_2} data from CarbonTracker2013 and GEOS-Chem by analyzing their spatio-temporal distribution. From the comparison, we find that the model outputs and the satellite data show consistent spatial patterns in most regions, except in tropical central Africa where there is a large discrepancy during late winter and early spring. This discrepancy may result from two possible causes related to model simulations and satellite retrievals. First, the lack of atmospheric observations in this region leads to the deficiency of CarbonTracker in constraining tropical biomass burning CO_2 fluxes (Peter *et al*., 2007). Second, the retrievals of XCO_2 in this region may be affected by thin clouds that often occur in the tropics (Heymann *et al*., 2012). Further investigation and validation are needed to explain the underlying causes of this discrepancy. We can see that the relatively high spatio-temporal resolution of the global land mapping dataset enables us to identify the locations and times of the discrepancies between the satellite retrievals and model simulations. Such detailed comparisons can provide potential evidences for further improvement of both satellite data and model data. Moreover, in this study, the spatio-temporal geostatistical prediction method has been used to reconstruct time series of X_{CO_2} at TCCON sites, making it an effective method for collocating satellite X_{CO_2} data with ground-based data (e.g. Nguyen *et al*., 2014).

The generated global land mapping $X\text{CO}_2$ data in this study, with continous data distribution in space and time, well present the spatio-temporal variations of $X\text{CO}_2$ that are hard to directly inferred from the sparse coverage of original satellite data retrievals. Therefore, it provides a new global geospatial dataset in global understanding of greenhouse gase dynamics and global warming. Specifically, the mapping dataset will play potential roles in evaluating the modelled CO_2 fields from carbon flux estimates coupled with an atmospheric transport model, as discussed in Hammerling et al. (2012b), detection of atmospheric CO_2 enhancement due to anthropogenic emissions, as discussed in Keppel-Aleks et al. (2013) and assessment of biospheric activities in the global scale, as discussed in Schneising et al. (2014). Recently, Liu et al. (2015) has shown an promising application of the mapping data in viewing the spatial patterns of CO_2 sources and sinks.

Some gaps still exist in the mapping dataset, because (1) there are areas where no observations are available, such as the Sahara desert area where data observed in median-gain mode were filtered before further analysis, and (2) we limit the kriging neighborhood to a spatial range of 500 km in order to make predictions efficiently using the most related surrounding dataset to the predictors. As a potential solution, the strength of the model data may be further incorporated in a data assimilation way to map some areas where data are extremely sparse (*e.g.* van de Kassteele *et al*., 2009; Chatterjee *et al*., 2010) since the model can well reproduce the large-scale features of atmospheric CO_2. Furthermore, the global land mapping method developed in this study depends on the assumption of uniform spatio-temporal correlation structures in the same latitudinal zone. As more satellite $X\text{CO}_2$ data become available in the future, we will be able to investigate the characteristics of correlation structures within each zone over different time interval, such as every season. Moreover, in comparison between spatio-temporal and spatial-only approcah using cross-validation, leaving out larger regions would conceivably better highlight the advantage of a spatio-temporal approach where temporal correlation can provide large benefits given the lack of nearby data. This will also be our future work.

References

ACOS-GOSAT Data Access: Goddard Earth Sciences Data and Information Services Center, available at: http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS.shtml, last access: June 2014.

ACOS Data User's Guide, ACOS Level 2 Standard Product Data User's Guide, v3.3, Goddard Earth Sciences Data and Information Services Center, available at: http://disc.sci.gsfc.nasa.gov/acdisc/documentation /ACOS_v3.3_DataUsersGuide.pdf, last access: June 2014.

Alkhaled, A., et al., 2008. A global evaluation of the regional spatial variability of column integrated CO_2 distributions. J. Geophys. Res., 113, D20303.

Arlot, S., 2010. A survey of cross-validation procedures for model selection. Statist Surv, 4, 40–79.

Boesch, H., 2011. Global characterization of CO_2 column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 mission. Remote Sens., 3, 270-304.

Bovensmann, H., et al., 1999. SCIAMACHY-Mission objectives and measurement modes, J. Atmos. Sci., 56, 127-150.

Bovensmann, H., et al., 2010. A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications. Atmos. Meas. Tech., 3, 781-811.

Buchwitz, M., et al., 2005. Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: initial comparison with chemistry and transport models. Atmos. Chem. Phys., 5, 941-962.

Buchwitz, M., et al., 2013. The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO_2 and CH_4 global data sets. Remote Sens. Environ., 162, 344-362.

Butz, A., et al., 2011. Toward accurate CO_2 and CH_4 observations from GOSAT. Geophys. Res. Lett., 38, L14812.

Cambardella, C. A., et al., 1994. Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America Journal, 58, 1501-1511.

CarbonTracker CT2013 Data Assess: National Oceanic and Atmospheric Administration Earth System Research Laboratory Global Monitoring Division, available at: http://www.esrl.noaa.gov/gmd/ccgg/carbontracker, last assess: May 2013.

Chatterjee, A., et al., 2010. A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness. J. Geophys. Res.,115, D20207.

Chiles, J. P. and Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty, 2nd Edition, New York, NY, USA: Wiley, pp. 204–215.

Ciais, P., *et al.*, 2014, Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. Biogeosciences, 11, 3547-3602.

Cogan, A. J., et al., 2012. Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations. J. Geophys. Res. Atmos., 117, D21301.

Connor, B. J., et al., 2008. Orbiting Carbon Observatory: Inverse method and prospective error analysis. J. Geophys. Res. Atmos., 113, D05305.

Conway, T. and Tans, P., 2015. Annual mean global carbon dioxide growth rates, available at: http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html. NOAA/ESRL, last access: 15 February 2015.

Cressie, N., 1985. Fitting variogram models by weighted least squares. Math. Geol., 17, 563–586.

Cressie, N., 1993. Statistics for spatial data. New York: John Wiley & Sons, 29-210.

Cressie, N. and Wikle, C. K., 2011. Statistics for spatiotemporal data. New York: John Wiley & Sons, 297-360.

Crisp, D., et al., 2004. The Orbiting Carbon Observatory (OCO) Mission. Adv. Space. Res., 34, 700–709.

Crisp, D., et al., 2012. The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization. Atmos. Meas. Tech., 5, 687-707.

De Iaco, S., Myers, D. E., and Posa, D., 2001. Space–time analysis using a general product–sum model. Statist. Probab. Lett., 52, 21–28.

De Iaco, S., 2010. Space–time correlation analysis: A comparative study. J. Appl. Stat., 37, 1027–1041.

De Iaco, S. and Posa, D., 2012. Predicting spatio-temporal random fields: some computational aspects. Comput. Geosci., 41, 12–24.

Deng, F., et al., 2014. Inferring regional sources and sinks of atmospheric CO_2 from GOSAT XCO_2 data. Atmos. Chem. Phys., 14, 3703-3727.

Dils, B., et al., 2014. The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI SCIAMACHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON. Atmos. Meas. Tech., 7, 1723-1744.

Duren, R. M. and Miller, C. E., 2012. Measuring the carbon emissions of megacities, Nature Climate Change, 2, 560-562.

Feng, L., et al., 2009. Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter. Atmos. Chem. Phys., 9, 2619-2633.

Fu, W. J., et al., 2014. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. Biogeosciences, 11, 2401-2409.

Guo, H.D., Z. Liu and L.W. Zhu, 2010. Digital Earth: decadal experiences and some thoughts. International Journal of Digital Earth, 3:1, 31-46.

Gething, P. W., et al., 2007. A local space-time Kriging approach applied to a national outpatient malaria data set. Comput. Geosci., 33, 1337–1350.

Haas, T. C., 1990. Lognormal and moving window methods of estimating acid deposition. J. Am. Stat. Assoc., 85, 950-963.

Haas, T. C., 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. J. Am. Stat. Assoc., 90, 1189-1199.

Hammerling, D. M., Michalak, A. M., and Kawa, S. R., 2012a. Mapping of CO_2 at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2. J. Geophys. Res., 117, 1-10.

Hammerling, D. M., et al., 2012b. Global CO_2 distributions over land from the Greenhouse Gases Observing Satellite (GOSAT). Geophys. Res. Lett., 39(8).

Heimann, M., 2009. Searching out the sinks. Nature Geoscience, 2, 3-4.

Heymann, J., et al., 2012. SCIAMACHY WFM-DOAS XCO_2: comparison with CarbonTracker XCO_2 focusing on aerosols and thin clouds. Atmos. Meas. Tech., 5, 1935-1952.

Hoaglin, D. C., Iglewicz B., and Tukey, J. W., 1986. Performance of some resistant rules for outlier labeling, J. Am. Stat. Assoc., 81, 991-999.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate

Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jost, G., Heuvelink, G. B. M., and Papritz, A., 2005. Analyzing the space-time distribution of soil water storage of a forest ecosystem using spatio-temporal Kriging. Geoderma, 128, 258–273.

Katzfuss, M. and Cressie, N., 2011. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. J. Time Ser. Anal., 32, 430–446.

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

Keppel-Aleks, G., et al., 2012. The imprint of surface fluxes and transport on variations in total column carbon dioxide. Biogeosciences, 9, 875-891.

Keppel-Aleks, G., et al., 2013. Towards constraints on fossil fuel emissions from total column carbon dioxide, Atmos. Chem. Phys., 13, 4349–4357.

Kort, E. A., et al., 2012. Space-based observations of megacity carbon dioxide. Geophys. Res. Lett., 39, L17806, 1-5.

Kyriakidis, P. C. and Journel, A. G., 1999. Geostatistical space-time models: A review. Math. Geol., 31, 651–684.

Liu, D., Lei, L., Guo, L., and Zeng, Z., 2015. A Cluster of CO_2 Change Characteristics with GOSAT Observations for Viewing the Spatial Pattern of CO_2 Emission and Absorption. *Atmophere*, 6, 1695-1713.

Liu, Y., Yang, D., and Cai, Z., 2013. A retrieval algorithm for TanSat XCO_2 observation: retrieval experiments using GOSAT data, Chin. Sci. Bull., 58, 1520-1523.

Liu, Y., et al., 2012. H.: Mapping the FTS SWIR L2 product of XCO_2 and XCH_4 data from the GOSAT by the Kriging method-a case study in East Asia. Int. J. Remote Sens., 33, 3004-3025.

Maksyutov, S., et al., 2013. Regional CO_2 flux estimates for 2009–2010 based on GOSAT and ground-based CO_2 observations. Atmos. Chem. Phys., 13, 9351-9373.

McKain, K., et al., 2012. Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region. Proc. Natl. Acad. Sci. USA, 109, 8423-8428.

Miller, C. E., et al., 2007. Precision requirements for space-based XCO2 data. J. Geophys. Res., 112, D10314.

Morino, I., et al., 2011. Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra, Atmos. Meas. Tech., 4, 1061-1076.

Nassar, R., et al. 2010. Modeling global atmospheric CO_2 with improved emission inventories and CO_2 production from the oxidation of other carbon species, Geosci. Model Dev., 3, 689-716.

Nguyen, H., et al., 2014. A method for colocating satellite XCO_2 data to ground-based data and its application to ACOS-GOSAT and TCCON, Atmos. Meas. Tech., 7, 2631-2644.

Nguyen, H., et al., 2014. Spatio-temporal data fusion for remote-sensing applications. Technometrics, 56(2), 174-185.

NIES GOSAT Project, 2010. Algorithm theoretical basis document for CO_2 and CH_4 column amounts retrieval from GOSAT TANSO-FTS SWIR. NIES-GOSAT-PO-017, V1.0.O'Dell, et al., 2012. The ACOS CO_2 retrieval algorithm − Part 1: Description and validation against synthetic observations. Atmos. Meas. Tech., 5, 99–121.

Peters, W., et al., 2007. An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. Proc. Natl. Acad. Sci. USA, 104, 18925–18930.

Piao, S., et al., 2009, The carbon balance of terrestrial ecosystems in China. Nature, 458, 7241, 1009–1013.

Reuter, M., et al., 2013, A joint effort to deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: the ensemble median algorithm EMMA. Atmos. Chem. Phys., 13, 1771-1780.

Rodgers, C. and Connor, B., 2003. Intercomparison of remote sounding instruments. J. Geophys. Res, 108, 4116–4229.

Schabenberger, O. and Gotway, C. A., 2004. Statistical methods for spatial data analysis. CRC Press. 474-476.

Schneising, O., et al., 2008. Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite − Part 1: Carbon dioxide. Atmos. Chem. Phys., 8, 3827–3853.

Schneising, O., et al., 2011. Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY. Atmos. Chem. Phys., 11, 2863–2880.

Schneising, O., et al., 2014. Terrestrial carbon sink observed from space: variation of growth rates and seasonal cycle amplitudes in response to interannual surface temperature variability. Atmos. Chem. Phys., 14, 133-141.

Shupeng C. and J. van Genderen, 2008. Digital Earth in support of global change research. International Journal of Digital Earth, 1(1), 43-65.

Tadić, J. M., et al., 2014. Mapping of satellite earth observations using moving window block kriging, Geosci. Model Dev. Discuss., 7, 5381-5405.

TCCON Data Access. TCCON Data Archive, available at: http://tccon.ipac.caltech.edu/, last access: August 2014.

Tomosada, M., et a., 2008. Application of the spatial statistics to the retrieved CO_2 column abundances derived from GOSAT data, in Proceedings of the 4th WSEAS International Conference on Remote Sensing, Venice, Italy, 67-73.

Tomosada, M., et al., 2009. A prediction method of the global distribution map of CO_2 column abundance retrieved from GOSAT observation derived from ordinary kriging", ICROS-SICE International Joint Conference 2009, Japan, 4869-4873.

Tsutsumi, Y., et al., 2009. Technical report of global analysis method for major greenhouse gases by the World Data Center for Greenhouse Gases, Global Atmosphere Watch Research and Monitoring Report, No. 184 (WMO/TD No. 1473), 1–23.

Tukey, J. W., 1977. Exploratory data analysis, Addison-Wesley.

van de Kassteele, J., et al., 2009. External drift kriging of NOx concentrations with dispersion model output in a reduced air quality monitoring network. Environ. Ecol. Stat., 16, 321-339.

Watanabe, H., et al., 2015. Global mapping of greenhouse gases retrieved from GOSAT Level 2 products by using a kriging method. International Journal of Remote Sensing, 36(6), 1509-1528.

World Meteorological Organization, 2014. WMO WDCGG data summary, WMO WDCGG No. 38.

Wunch, D., et al., 2011a. A method for evaluating bias in global measurements of CO_2 total columns from space. Atmos. Chem. Phys., 11, 12317-12337.

Wunch, D., et al., 2011b. The Total Carbon Column Observing Network. Philos. T. Roy. Soc. A, 369, 2087–2112.

Wunch, D., et al., 2013. The covariation of Northern Hemisphere summertime CO_2 with surface temperature in boreal regions, Atmos. Chem. Phys., 13, 9447-9459.

Yokota, T., et al., 2009. Global Concentrations of CO_2 and CH_4 Retrieved from GOSAT: First Preliminary Results. SOLA, 5, 160–163.

Zeng, Z., et al., 2013. Incorporating temporal variability to improve geostatistical analysis of satellite-observed CO_2 in China. Chin. Sci. Bull., 58, 1948-1954.

Zeng, Z., et al., 2014. A Regional Gap-Filling Method Based on Spatiotemporal Variogram Model of CO_2 Columns. IEEE Trans. Geosci. Remote Sens., 52, 3594–3603.