



Towards monitoring CO₂ source-sink distribution over India via inverse modelling: Quantifying the fine-scale spatiotemporal variability of atmospheric CO₂ mole fraction

5 Vishnu Thilakan^{1,4}, Dhanyalekshmi Pillai^{1,4}, Christoph Gerbig², Michal Galkowski^{2,3},
Aparna Ravi^{1,4}, and Thara Anna Mathew¹

¹Indian Institute of Science Education and Research Bhopal (IISERB), Bhopal, India

²Max Planck Institute for Biogeochemistry, Jena, Germany

³AGH University of Science and Technology, Kraków, Poland

10 ⁴Max Planck Partner Group (IISERB), Max Planck Society, Munich, Germany

Correspondence to: Dhanyalekshmi Pillai (dhanya@iiserb.ac.in, kdhanya@bgc-jena.mpg.de)

Abstract

15 Improving the estimates of CO₂ sources and sinks over India through inverse methods calls for a comprehensive
atmospheric monitoring system involving atmospheric transport models that realistically account for
atmospheric CO₂ variability along with good coverage of ground-based monitoring stations. This study
investigates the importance of representing fine-scale variability of atmospheric CO₂ in models for the optimal
use of observations through inverse modelling. The unresolved variability of atmospheric CO₂ in coarse models
20 is quantified by using WRF-Chem simulations at a spatial resolution of 10 km × 10 km. We show that the
representation errors due to unresolved variability in the coarse model with a horizontal resolution of one degree
(~ 100 km) are considerable (median values of 1.5 ppm and 0.4 ppm for the surface and column CO₂,
respectively) compared to the measurement errors. The monthly averaged surface representation error reaches
up to ~5 ppm, which is comparable to a quarter to half of the magnitude of seasonal variability. Representation
error shows a strong dependence on multiple factors such as time of the day, season, terrain heterogeneity, and
25 changes in meteorology and surface fluxes. By employing a first-order inverse modelling scheme using pseudo
observations from nine tall tower sites over India, we show that the Net Ecosystem Exchange (NEE) flux
uncertainty solely due to unresolved variability is in the range of 3.1 to 10.3% of the total NEE of the region. By
estimating the representation error and its impact on flux estimations during different seasons, we emphasize the
need for taking account of fine-scale CO₂ variability in models over the Indian subcontinent to better understand
30 processes regulating CO₂ sources and sinks. The efficacy of a simple parameterization scheme is further
demonstrated to capture these unresolved variations in coarse models.

1 Introduction

Accurate assessment of sources and sinks of CO₂ is essential in planning and implementing mitigation strategies
for greenhouse gas emissions and associated climate change. However, estimations of CO₂ fluxes contain
35 significant uncertainties, which increase even more with finer spatial scales such as those required for the
climate change mitigation policies at regional and national levels (e.g., Ciais et al., 2014; Li et al., 2016;
Cervarich et al., 2016). By using atmospheric CO₂ concentration measurements, the CO₂ fluxes can be estimated
by a multi-constrained observation-modelling approach, often referred to as top-down approach or inverse



40 modelling (Enting, 2002). For about two decades, these top-down approaches have been widely used to
understand the modifications in the carbon cycle through natural and anthropogenic induced environmental
changes (Bousquet, 2000; Schimel et al., 2001; Rödenbeck et al., 2003; Patra et al., 2005). In addition to the
observations, the inverse modelling system makes use of an atmospheric transport model (forward model),
which determines the distribution of CO₂ concentration. Thereby, the inverse optimization approach derives the
surface fluxes that are consistent with measured concentration. The United Nations Framework Convention on
45 Climate Change (UNFCCC) has acknowledged the increasing capability of inverse modelling to systematically
monitor greenhouse gas (GHG) concentrations (Bergamaschi et al., 2018).

Most of the inverse modelling systems rely on global atmospheric transport models with coarse horizontal
resolution (often greater than one degree) (Rödenbeck et al., 2003; Peters et al., 2007; Rödenbeck et al., 2018a,
b; Inness et al., 2019). These global data assimilation systems play an important role in studying continental or
50 sub-continental fluxes at annual or sub-annual scales. However, regional estimation of fluxes using global
models is hindered by the inability of these transport models to represent the observed CO₂ variability. The
observed variability, as seen from the spatial and temporal distribution of atmospheric CO₂, is highly correlated
with the space and time scales of weather systems (Parazoo et al., 2011). This explains the presence of large
model-data mismatches in regions where mesoscale circulation is predominant (Ahmadov et al., 2007). Wind
55 speed, wind direction, and height of the planetary boundary layer (PBL) are the critical variables that determine
the atmospheric CO₂ variability. Strong wind normalizes other small-scale variations in observed concentration
due to mixing, and the predictability can be higher during these conditions (Sarrat et al., 2007). The height of the
PBL is an essential variable since the atmospheric CO₂ is subjected to rapid mixing up to this altitude. Hence,
for a given location with a negative gradient in CO₂ vertical distribution, an overestimation of PBL height leads
60 to an underestimation of CO₂ concentration and vice-versa (Gerbig et al., 2008).

Another important variable that impacts the CO₂ variability is the heterogeneous topography. Variations in
topography influence the transport of the tracers. When the small-scale orographic details are not adequately
represented in the models, they can lead to representation errors in CO₂ simulations as large as 3 ppm at scales
of 100 km (Tolk et al., 2008; Pillai et al., 2010). Horizontal gradients in CO₂ concentrations can go up to values
65 of 30 ppm within a spatial scale of 200 km, depending on the land surface heterogeneity (van der Molen and
Dolman, 2007). Further, variations in land use patterns between neighbouring regions can cause considerable
variability in the CO₂ surface fluxes. Thus, a proper representation of land use patterns is also important in terms
of simulating CO₂ variability. Previous studies based on airborne measurements reported that transport models
need a spatial resolution smaller than 30 km to be able to represent CO₂ spatial variability in the continental
70 boundary layer (Gerbig et al., 2003). Significant efforts have been invested in deriving fluxes by taking into
account these fine-scale variations (e.g., Gerbig et al., 2003; Lauvaux et al., 2009a; Carouge et al., 2010; Pillai et
al., 2011, 2012; Broquet et al., 2013) over North American and Eurasian domains in the past decade. However,
there still exists lower confidence in estimates over the regions, where there is a lack of both advanced
modelling systems at relevant spatio-temporal resolutions and good coverage of ground-based monitoring
75 stations.

In the context of the Indian sub-continent, the inverse-based estimation of fluxes at fine scales is essentially
new; hence many questions remain. A number of monitoring sites measuring atmospheric greenhouse gases



80 have become available in India during the last decade (Tiwari et al., 2011; Lin et al., 2015, Nomura et al., 2021).
Aside from the ongoing progress in augmenting observational data streams, it remains challenging to assimilate
these data for deducing process-specific information effectively (e.g., McKain et al., 2012; Bréon et al., 2015;
Pillai et al., 2016). The limitation of coarse global models in representing observations over the Indian
subcontinent is reflected in the analysis made by Patra et al. (2011).

85 The seasonally reversing South Asian monsoon system is a prominent meteorological phenomenon affecting the
Indian subcontinent, which is also expected to influence the terrestrial-atmosphere flux exchanges. Various
studies have demonstrated the role of Indian monsoon circulations on regional atmospheric transport by strong
south westerly winds during the summer monsoon (June to September) and by north easterly winds during the
winter monsoon (October to November) (e.g., Goswami and Xavier, 2005; Krishnamurthy and Shukla, 2007).
Monsoon convection transports the boundary layer air into the upper troposphere. Subsequently, air parcels are
slowly uplifted by diabatic heating to higher altitudes (e.g., Vogel et al., 2019). An accurate representation of
90 convective vertical transport is very challenging and an important source of uncertainty in current transport
models (Willets et al., 2016). Note that the Asian summer monsoon anticyclone (ASMA) active during the
Indian summer monsoon period plays a key role in uplifting trace gases to the upper troposphere and lower
stratosphere (e.g., Park et al., 2007). Moreover, a significant component of flux variations can arise from
biospheric fluxes (Schimel et al., 2014), which is influenced by variables such as rainfall, availability of
95 radiation, and temperature (Chen et al., 2019). Several studies showed that the monsoon system substantially
impacts vegetation growth, generating distinct spatio-temporal patterns of the biogenic fluxes (e.g., Gadgil,
2003; Valsala and Maksyutov, 2013, Ravi Kumar et al., 2016). It is noteworthy that the cropping patterns over
India have a strong dependence on seasons and are mainly determined by dry and wet seasons for nearly 65 to
100 70 % of the country's area except over north-eastern and south-western (Western Ghats) regions of India. In
India, wet season crops (Kharif crops cultivated from June to November) including Rice, Millets, and Maize
mainly depend on monsoon rain. Dry season crops (Rabi crops, e.g. Wheat, Barley, and Mustard, cultivated
from November to April,) are less water-dependent and primarily rely on irrigation (DAC/MA 2015). Therefore,
employing a higher resolution modelling over the Indian subcontinent is desirable to better account for fine-
scale variations generated by both mesoscale transport processes and surface flux patterns.

105 This study focuses on accounting for unresolved sub-grid scale variability when employing current generation
global models. Assimilation of observations in an inverse framework requires the characterization of these error
structures at relevant scales that can be utilized to retrieve source-sink distribution over India. The main
objectives of this paper are to describe and quantify the expected spatiotemporal variability of atmospheric CO₂
that is not resolved by the current generation global models, quantify to what extent these variations cause
110 uncertainty in flux estimations, and assess how these uncertainties can be minimized by modelling the sub-grid
variations in the global models. Specifically, we address the following questions: 1) how good is the level of
agreement among global transport models that are used in current generation inversion systems for predicting
atmospheric CO₂ concentrations over the Indian subcontinent? 2) how large are the variations of atmospheric
CO₂ that are unresolved by global and regional models, which operate at different spatial scales from 4° × 4° to
115 0.5° × 0.5° ? 3) what is the role of seasonal changes on generating different patterns in these sub-grid variations
of CO₂? 4) how much is the uncertainty in the inverse-based flux estimation caused by these unresolved



120 variations in the coarse models when utilizing a given network of surface observations over the domain? 5) how effectively can we capture the key aspects of the variability and account for it in flux estimations? Information from observations can be better utilized if we improve the atmospheric transport models to resolve the observed variability as accurately as possible. As a result, the data assimilation system gains significantly (e.g. with increasing weights on observations and performing minimal data filtering), from this for improving the flux estimates.

125 In this article, we present results based on the analyses of high-resolution simulations at a spatial resolution of 10 km × 10 km for the months of July and November 2017. The year 2017 was characterized by neutral Indian Ocean Dipole conditions over the Indian Ocean with the beginning of a mild La Nina over the Pacific by the end of the year (NOAA/ESRL, 2022a, b). The month of July represents a monsoon period when the biospheric activity is significant together with atmospheric convection activities. July is also characterized by strong low-pressure system activity over the Bay of Bengal, which results in large rainfall over central India (Krishnamurthy and Ajayamohan, 2010). On the other hand, the month of November is more representative of
130 post-monsoon wintertime over the Indian subcontinent. We quantify the sub-grid variability using these high-resolution simulations. By designing a pseudo surface observation network over the domain, we investigate the impact of these unresolved variations on the regional flux estimations and assess how a simple parameterization scheme can help in reducing these errors in the global model. To our knowledge, there is no comprehensive published study of this kind over the Indian subcontinent until now assessing the magnitude and impact of
135 temporal and spatial variability exhibited by atmospheric CO₂.

The outline of the paper (see Supplementary Fig. S1) is as follows: Section 2 describes our modelling system, data and methods used for estimating the sub-grid scale variability of CO₂. In Sect. 3, we present the global model comparisons and spatial variability analysis, highlighting potential modelling difficulties for estimating the CO₂ budget over India. We provide a quantification of the expected sub-grid scale variability based on our
140 high-resolution simulations, as well as its impact on regional flux estimations. Finally, we discuss the implications of our findings in Sect. 4, suggesting the ways forward to yield an improved estimation of CO₂ budgets over India.

2 Data & Methodology

145 We have performed a series of analyses using the simulations generated by our high-resolution modelling system, which is described in Sect. 2.1. Additionally, we have utilized optimized CO₂ products at global scales to provide a more comprehensive overview of the typical mismatch between the existing model simulations over the Indian subcontinent at monthly and annual scales (Sect. 2.2). These global model outputs are derived from inverse model simulations, which estimate the source-sink distributions of CO₂ and then generate three-dimensional CO₂ concentration fields that are consistent with the optimized posterior fluxes. In this study, the
150 high-resolution simulations are used to quantify the sub-grid scale variability of CO₂ that cannot be captured by the global models due to their coarse resolution. For this quantification of the spatial variability, we use the representation error approach described in Sect. 2.3. An observation system simulation experiment (OSSE)



155 using high-resolution CO₂ simulations has been carried out to estimate the impact of the derived sub-grid scale variations on flux estimations over India via inverse optimization (see Sect. 2.4).

2.1 WRF-Chem GHG Modelling System

160 We use the modelling system WRF-Chem GHG in which the Weather Research and Forecasting model (WRF) version 3.9.1.1 (Skamarock et al., 2008) is coupled with the greenhouse gas module (WRF-Chem-GHG, Beck et al., 2011), implemented as part of the WRF-Chem distribution (WRF-Chem, Grell et al., 2005). For simulating the atmospheric transport, the model uses fully compressible Eulerian non-hydrostatic equations on Arakawa C-staggered grid, conserving mass, momentum and scalars (Skamarock et al., 2008). In the WRF-Chem GHG (hereafter referred as WRF-GHG), we use the passive tracer chemistry option to simulate changes in CO₂ mixing ratios associated with surface fluxes and atmospheric transport. We utilize a biospheric model and emission inventory data to simulate atmospheric CO₂ enhancements associated with biogenic and emission
165 fluxes as described in Sect. 2.1.1 and 2.1.2. Table. 1 summarizes the model configuration, including physics parameterizations and input data used in this study.

The model domain covers a region spanning from 65°E to 100°E and 5°N to 40°N, configured in a Lambert conformal conic (LCC) projection with 307 × 407 grid points. The spatial resolution of the grid is 10 km × 10 km, and the model time-step is 60 s. We have used model output with a temporal resolution of 1 hour for this
170 study. The simulations are performed using 39 vertical levels with the model top at 50 hPa and 10 levels within the lowest 2 km. WRF-GHG simulations are performed for the entire July and November 2017. Implementation of the WRF-GHG system over the Indian subcontinent enables us to customize it according to the domain features and build a state-of-the-art modelling system, which eventually estimates CO₂ fluxes through regional inverse systems. The potential of the WRF-GHG model in simulating fine-scale spatial variability was also
175 established in previous studies (Ahmadov et al., 2009; Pillai et al., 2011; Park et al., 2018).

2.1.1 Representation of biospheric fluxes

We use the Vegetation Photosynthesis and Respiration Model (VPRM) in the modelling system to calculate Net Ecosystem Exchange (NEE) representing the biospheric fluxes (Mahadevan et al., 2008). VPRM is a diagnostic biosphere model, which utilizes remote sensing products: Enhanced Vegetation Index (EVI) and Land Surface
180 Water Index (LSWI) derived from reflectance data of the Moderate resolution Imaging Spectroradiometer (MODIS) as well as meteorological data: solar radiation and air temperature. In this study, these hourly NEE calculations are performed within WRF-GHG, simultaneously with the meteorology simulations in which NEE is calculated as a sum of gross ecosystem exchange (GEE) and ecosystem respiration (R_{eco}). VPRM, in this case, uses the meteorological data provided by WRF-GHG. VPRM uses the SYNMAP vegetation classification (using the tile approach) (Jung et al., 2006) as well as EVI and LSWI from MODIS surface reflectance data at a
185 resolution of 1 km and 8 days. We aggregate these indices specific for different vegetation types onto the LCC projection for the entire domain at the model's spatial resolution. A number of studies have used VPRM for other regions around the world in which derived NEE shows good prediction skills for hourly to monthly timescales (Ahmadov et al., 2009; Pillai et al., 2011; Liu et al., 2018; Park et al., 2018).

190 2.1.2 Representation of emission fluxes



Anthropogenic CO₂ emission fluxes are prescribed from the Emission Database for Global Atmospheric Research (EDGAR) dataset, version 6.0, provided at a horizontal resolution of 0.1° × 0.1° (Crippa et al., 2021). We disaggregate the available annual emission data into hourly emissions using the temporal distribution CO₂ profiles (Steinbach et al., 2011; Kretschmer et al., 2014). To represent biomass burning emission, we have used data from the Global Fire Assimilation System (GFAS) with a spatial resolution of 0.1° × 0.1° and a temporal resolution of one day. GFAS is based on satellite data, which provides the fire emission by assimilating fire radiative power (FRP) observations from MODIS instruments (Kaiser et al., 2012). All these flux data are gridded and projected to WRF-GHG's model domain.

2.1.3 Initial and boundary conditions

Meteorological and chemical initial and boundary conditions are required in WRF-GHG to account for the initial state and inflow or background flow. The initial and lateral boundary conditions for the meteorological variables, including horizontal wind components, pressure, specific humidity, sea surface temperature (SST), and the necessary surface initialization fields are obtained from the ERA5 reanalysis dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF), extracted at a horizontal resolution of 25 km and a temporal resolution of 1 hour (Hersbach et al., 2020). The initial and lateral boundary conditions of CO₂ tracers are obtained from the Copernicus Atmosphere Monitoring Service (CAMS, 2.2.4) products (Massart et al., 2016; Agustí-Panareda et al., 2019). We have used the dry air mole fractions of CO₂ from the CAMS-GHG, which has a temporal resolution of 6 hour and horizontal resolution of 0.5° × 0.5° with 137 vertical levels. Note that there exists a CAMS product at 9 km × 9 km resolution, which is in the developmental phase and not yet available to the general public (personal contact: Anna.Agusti-Panareda@ecmwf.int).

We have utilized a simulation strategy to update the initial meteorological conditions for taking advantage of assimilated meteorological fields from ECMWF. The model is reinitialized each day with ECMWF assimilated data at the model starting time of 12:00 UTC (day+0) and runs for 30 hours until 18:00 UTC of the next day (day+1). The first six hours are considered for meteorological spin-up, and the remaining 24 hours (from day+0, 18:00 UTC to day+1, 18:00 UTC) are used for the analysis. The initialization of CO₂ is done at the beginning of the first hour of model simulation, which is 00:00 UTC (e.g., Ahmadov et al., 2012; Pillai et al., 2011).

2.2 Global model products

We have used optimized products at global scales to examine the differences in the representation of CO₂ variability over the Indian subcontinent at monthly and annual scales. Four global inverse modelling products - CarbonTracker, CarboScope, LSCE v18r3 and LSCE FT18r1- available during the year 2017 are used for our analysis (See Table. 2 for more details). The LSCE model version v18r3 (hereafter LSCE) utilizes surface observations for the optimization, and the model version FT18r1 (hereafter LSCE FT) uses satellite retrievals from the Orbiting Carbon Observatory (OCO-2) for the optimization of CO₂ fluxes (Chevallier et al., 2005; Chevallier et al., 2010; Chevallier, 2013). All these above models differ in terms of the model formulations and configuration (e.g., transport and the employed inversion methodology), observational datasets that were assimilated (e.g., data from surface monitoring stations, aircraft missions, ship cruises, AirCore balloon soundings, and satellite's total column retrievals), prior datasets, and spatiotemporal resolutions. None of these



230 products used ground-based observations from the Indian subcontinent for their optimization, which raises
concerns about the reliability of the optimized flux estimations over the region. Hence, it can be assumed that a
part of the inter-model differences in predicting the variability is related to the paucity of CO₂ observations over
the region. To represent the daytime, we have used the concentration fields for the local time ranging from
11:30 to 16:30 from all these models for the analysis.

2.3 Quantification of spatial variability

235 For quantifying the spatial variability due to sub-grid scale processes that cannot be resolved by the coarse
resolution models, we follow the approach as described in Pillai et al. (2010). The term ‘representation error’
indicates the mismatch between the scales of model simulations and observations collected (Pillai et al., 2010;
Janjić et al., 2017). In other words, the representation errors arise due to unresolved scales, which could not be
captured by the model. Here we calculate the representation errors in the coarse resolution models, which can be
240 resolved by implementing a high-resolution model at 10 km resolution. It is assumed that the high-resolution
simulation captures the majority of the sub-grid scale variability even though it cannot be expected to resolve all
observed variability. Most of the current global model simulations are performed at coarse resolutions of several
degrees. But with the recent advancement in computational capacity and numerical techniques, a horizontal
resolution of 1° × 1° is quite likely achievable for the global data assimilation systems. For estimating the
245 representation error in a coarse model with a typical spatial resolution of 1° × 1°, we have calculated the
standard deviation of CO₂ dry air mole fraction simulated by the WRF-GHG model within the coarse grid boxes
of 1° × 1° as follows:

$$\sigma_{CO_2(\text{tot})} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (m_j - \bar{m})^2} \quad (1)$$

$$\text{where } \bar{m} = \frac{1}{n} \sum_{j=1}^n m_j$$

250 n is the number of 10 km boxes inside the coarser grid cell of 1° × 1°; m is the CO₂ dry air mole fraction
corresponding to 10 km boxes; and \bar{m} is the average within the coarser grid cell. So, the estimated values
represent the sub-grid scale variability within the coarse model grid cell with a horizontal resolution of 1° × 1°.
The representation errors are calculated at corresponding vertical model levels to represent the impact of surface
influence and mesoscale transport adequately as predicted by the high-resolution model. As mentioned before,
255 we assume that the high-resolution simulations represent the realistic distribution of CO₂. Further, we assume
that the coarse resolution model also has a terrain-following vertical coordinate system and also has the same
vertical grid spacing of high-resolution model. As the space-borne instruments can also make the mixing ratio
measurements, we extend the analysis to column-averaged dry air mole fraction (XCO₂) as measured by the
satellite instrument. i.e., m represents either CO₂ at a given model level or XCO₂. In order to assess the
260 dependence of representation error on the horizontal resolution of the employed model, we have computed
representation error for multiple resolutions ranging from 0.5° × 0.5° to 4° × 4°, in addition to 1° × 1°, which
would encompass the resolutions of both present and near-future global inverse modelling systems.

The surface representation errors are calculated using the model simulations from the second model level (mean
height is ~200 m from sea level) to avoid the inconsistency that can be generated from inputting emission fluxes



265 at the first model level. Representation errors are calculated separately for daytime (11:30 to 16:30 local time) and nighttime (23:30 to 4:30 local time) to account for the difference in the sub-grid scale process during these times. The representation error presented in Eq. (1) varies from one model time step to the next. In order to obtain a typical (average) representation error, we compute the monthly average representation error (σ_{CO_2}) using Eq. (2).

$$270 \quad \sigma_{CO_2} = \frac{1}{T} \sum_{t=1}^T \sigma_{CO_2(tot)} \quad (2)$$

where T is the total number of simulations in a month during daytime or nighttime. Further, we have calculated the representation error ($\sigma_{\overline{CO_2}(mon)}$) using Eq. (3), which only contain systematic component of representation error that can provide important constraints for inversions using both ground-based and satellite observations over India.

$$275 \quad \sigma_{\overline{CO_2}(mon)} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (M_j - \overline{M})^2} \quad (3)$$

$$\text{where } \overline{M} = \frac{1}{n} \sum_{j=1}^n M_j$$

n is the number of 10 km boxes inside the coarser grid cell of $1^\circ \times 1^\circ$; M_j is the monthly averaged CO_2 dry air mole fraction at a 10 km spatial scale; and \overline{M} is the corresponding average within the coarser grid cell of 1° . The difference between Eq. (1) and Eq. (3) is that we use monthly averaged CO_2 concentration values in Eq. (3) instead of hourly values as in Eq. (1). Both July and November are used to understand the differences in the variability during summer and winter.

280 Due to the paucity of adequate ground-level observations over India, satellite observations play an essential role in the estimation of CO_2 fluxes. Satellite observations can provide column average CO_2 (XCO_2) concentration with a precision of 1 to 1.5 ppm (O'Dell et al., 2012; Wunch et al., 2017; Liang et al., 2017). In order to utilize these satellite observations, the transport models being used in the inverse estimation must be highly accurate. Since satellite footprints are smaller ($\sim 2 - 20 \text{ km}^2$) than the current model grid size ($> 100 \text{ km}$), using these measurements for optimization via inverse modelling introduces spatial representation errors and associated uncertainties in the inferred fluxes. Note that the spatial biases of a few tenths of a ppm in column-averaged CO_2 can potentially alter even the annual sub-continental fluxes in the range of tenths of a gigaton of carbon fluxes (Chevallier et al., 2007, Miller et al., 2007 and Chevallier et al., 2010). To quantify these systematic transport errors when representing satellite measurements in inverse models, we calculate the spatial representation errors for XCO_2 that coarse inverse modelling would suffer from using highly precise and accurate satellite measurements.

295 We have selected monsoon (July) and post-monsoon (November) periods for our analysis to identify the seasonal changes in the sub-grid variability over India. In July, many low-pressure systems were active in the monsoon trough region (IMD weather reports, <https://mausam.imd.gov.in>). In general, tropical cyclones in the Asian monsoon region can cause fast uplift of air masses into the upper troposphere and lower stratosphere (e.g.



Li et al., 2021), which may increase the modelling error due to the misrepresentation of the associated mesoscale activity. The presence of enhanced biospheric activity during July can reduce the CO₂ concentration in the lower troposphere. Also, the strong vertical and horizontal mixing due to the monsoon circulation dilutes the CO₂ concentration in the atmosphere during July compared to November. The convective activity associated with the Indian summer monsoon was absent during November, however the convection caused by synoptic systems such as tropical cyclones was still present. Such a low-pressure system activity was found over the Bay of Bengal and over the Lakshadweep area ($\approx 8^\circ \text{ N}$, 74° E) from 22nd November onwards. One of these low-pressure systems in the Bay of Bengal further developed and intensified as a deep depression and moved to the southeast Arabian Sea and evolved into a severe cyclonic storm (Ockhi) by 30th November.

2.4 Estimation of representation error induced flux uncertainty using pseudo surface measurements

In order to quantify the impact of representation errors on flux estimations when utilizing surface measurements, we have devised the following strategy. We used nine CO₂ surface monitoring sites representing various geographical regions in India (Fig. 1). Not all these observation stations are currently fully operational or have continuous measurements. We have performed an observation system simulation experiment (OSSE) using high-resolution CO₂ simulations generated by the WRF-GHG model for each of these stations. We focus on the biospheric flux component, NEE. The simulated values of coarse models to compare with the observations are obtained from the nine grid cells of the coarse model covering these sites. The pseudo observations for these sites correspond to the values simulated by the WRF-GHG model at one of the fine grid cells contained in one cell of the coarse model. Since there are 100 fine grid cells per coarse grid cell, 100 different time series are generated and 100 corresponding inversions are performed to obtain robust results. For deducing the contribution of the representation error to the biospheric flux uncertainty, we have taken the following assumptions: 1) the hourly WRF-GHG simulations at 10 km ($\sim 0.1^\circ$) spatial scale represents actual variations in CO₂ mixing ratios of the measurement site, 2) there are no model or observation errors other than representation error, 3) the model captures the spatial and temporal patterns of fluxes correctly, and 4) the contribution from other surface fluxes and background mixing ratio (in ppm) are known. As a first-order simplification for the inversion, we assume that the footprints of each observation site span a radius of 200 km around the site based on our analysis using the Stochastic Time-Inverted Lagrangian Model (STILT, Lin et al., 2003). STILT footprints indicate that 50% of the sensitivity of a site to fluxes over India is located in a region that has about the same area as a circle with a radius of 200 km. For nine stations, this footprint area covers around 35 % of the total area of India. The STILT is driven with ECMWF IFS (Integrated Forecasting System) meteorological fields and the trajectories are calculated based on 100 virtual particles that are released for each time interval and location. The residence time of particles in the surface layer is weighted by the atmospheric density to derive the footprints of each location.

In our inversion set-up, we have used the hourly biospheric contribution of the atmospheric CO₂ mixing ratios simulated by WRF-GHG over the coarse grid cell of $1^\circ \times 1^\circ$ surrounding the location of each measurement site as OSSE observations ($m_{i,j}(t)$).

$$y_{i,j}(t) \equiv m_{i,j}(t) = \mathbf{H}_{i,j}(t) \cdot \mathbf{F}(\lambda) \quad (4)$$



335 where \mathbf{H} is the transport operator and $\mathbf{F}(\boldsymbol{\lambda})$ is the flux model in which a subset of parameters $\boldsymbol{\lambda}$ out of total model parameters \boldsymbol{p} will be optimized in the inversion. Here, i ($i = 1$ to 9) represents the nine observation sites and j ($j = 1$ to 100) is the number of WRF-GHG pixels inside the coarser grid cell of $1^\circ \times 1^\circ$.

The modelled biospheric CO_2 signal (\bar{m}_i) for the inversion is given by:

$$\bar{m}_i(t) = m_{i,j}(t) + \boldsymbol{\varepsilon}_{i,j}(t) \quad (5)$$

340 The modelled values deviate from the observations by a representation error $\boldsymbol{\varepsilon}_{i,j}(t)$. Since the modelled values (\bar{m}_i) correspond to the mean of the 100 fine grid cells, the simulated values at site i are given as:

$$\bar{m}_i(t) = \frac{1}{100} \sum_1^{100} m_{i,j}(t) \quad (6)$$

Here, $\mathbf{F}(\boldsymbol{\lambda})$ is taken as linearly dependent on $\boldsymbol{\lambda}$; hence can be expressed as

345 $\mathbf{F}(\boldsymbol{\lambda}) = \boldsymbol{\Phi} \cdot \boldsymbol{\lambda} \quad (7)$

where $\boldsymbol{\Phi}$ is the biospheric flux (NEE) distribution over the region.

In the inversion, we retrieve monthly NEE by utilizing hourly $m_{i,j}(t)$ and $\bar{m}_i(t)$ over a month. For OSSE and uncertainty flux estimation, we use the VPRM-derived NEE fluxes as the “true” fluxes (see Sect. 2.1.1). By this inverse modelling design, we require to perform 100 inversions per site, each of which uses a realization of
350 representation error to estimate the corresponding realization of the resulting uncertainty in the retrieved fluxes.

Both the observation and simulation vector have 6480 ($=9 \times 30 \times 24$) elements for a month having 30 days, and the state vector has 9 elements corresponding to scaling factors of fluxes for that month over regions around the 9 sites (see Fig. 1). In other words, each site has been assigned with one scaling factor for NEE, and there is a total of 9 scaling factors for a given month. We use a unit vector $\boldsymbol{\lambda}$ as prior scaling factors. The prior uncertainty
355 is neglected here, as the expected impact of the representation error on the retrieved fluxes is significantly smaller than typical prior uncertainties assumed in Bayesian inversions (on the order of 50% – 100% for biospheric fluxes). Hence neglecting this prior uncertainty does not have a large impact on our results. The inversion retrieves optimized scaling factors $\boldsymbol{\lambda}_{retr}$.

We have performed 100 inversions per site, and the scaling factors are retrieved by minimising the cost function
360 for each observation station:

$$J(\boldsymbol{\lambda}_{i,j}) = \frac{1}{T} \sum_{t=1}^T (m_{i,j}(t) - \bar{m}_i(t)\boldsymbol{\lambda}_{i,j})^2 \quad (8)$$

where T is the number of observations for a month. Minimizing these cost functions results in an optimized estimate of scaling factors $\boldsymbol{\lambda}_{retr}$, which is a vector of scaling factors with nine elements ($\boldsymbol{\lambda}_{retr,i}$) for each of the 100 inversion cases.

365 By this inverse design, the deviation of posterior fluxes from the true fluxes over India is thus the uncertainty in retrieved fluxes, \mathbf{S}_{rep} , that arises solely due to the contribution from the representation error. Standard deviation



of the scaling factors from these 100 inversions ($\sigma_{\lambda_{retr}}$) are used to retrieve flux uncertainty. S_{rep} is obtained as follows:

$$S_{rep} = \sqrt{\sum_{k=1}^K (S_{\lambda_{retr,k}} \Phi_{true,k})^2} \quad (9)$$

370 where Φ_{true} is the monthly VPRM biospheric flux (NEE) over the Indian region and k is the number of pixels
(33141 pixels) over the Indian region. Here, $S_{\lambda_{retr}}$ has the dimension of Indian region at a 10 km spatial
resolution and is defined in such a way that all the grids (at 10 km spatial resolution) other than the grids within
the influence region (200 km radius around the station) of each station is given with zero values (21335 pixels)
375 and the grids in the influence region of each station (11806 pixels) is given with the corresponding values of
 $\sigma_{\lambda_{retr,i}}$. This way, the approach doesn't depend on Eq. (1) to Eq. (3), but shows the impact of difference between
 m_j and \bar{m} on retrieved fluxes.

Any temporal correlations in the representation error are not considered for this experiment. We have performed
the inversion separately for daytime and nighttime values to identify the impact of diurnal variations of
representation errors on flux uncertainty. Note that by following the above inversion design and assumptions,
380 there is a high likelihood of underestimating the impact of the modelling error on flux estimations since we have
not considered other sources of uncertainties such as model transport uncertainty and inappropriate prior
assumptions. Thus, the quantification of flux uncertainty using this approach can be inferred as the lower bound
of the uncertainty (i.e., the minimum flux uncertainty one may expect while estimating fluxes using a model
with a grid cell of $1^\circ \times 1^\circ$ and 9 stations with the representativeness of 200 km).

385 3 Results and Discussions

3.1 Agreement among global models

We first analyse the level of agreement among current-generation global transport models in simulating CO_2
concentration over the Indian subcontinent. Note that a mere agreement among the coarse models is not
sufficient to justify the models' performance over the region due to their plausibly large model errors in
390 common and interdependency in terms of data sources. We restrict this analysis to daytime-only values since
different processes control the variability of CO_2 concentration at daytime and nighttime, and simulating
nighttime variability is more complicated than the daytime (Lauvaux et al., 2009a). For a consistent comparison
among global models, all the products are sampled at the same time for the region spanning from 67°E to 98°E
and 7°N to 38°N . Figure 2a depicts the annual vertical profiles of CO_2 concentration, showing models'
395 discrepancy in simulating the vertical gradients in concentration values including the boundary layer and the
free troposphere. A notable difference is observed in the simulation of the gradient within the boundary layer.
The magnitude and the height up to which this positive gradient is observed are different for these models.
LSCE (both versions) has the largest positive gradient among these models ($\sim 1 \text{ppm}$). It shows the maximum
concentration at around 700 m height and then a decrease in concentration. CarbonTracker also shows this
400 positive gradient in the surface layers up to a height of 900 m. But the gradient is much smaller compared to the
other two models. Among these four models, CarboScope does not exhibit this tendency in the lower
atmosphere. Its concentration decreases linearly from the surface as the height increases.



The seasonal variability of CO₂ uptake through photosynthesis, release through ecosystem respiration, and vertical transport is seen while analysing the monthly averaged CO₂ concentration profiles over the Indian subcontinent (Figs. 2b and 3). Comparatively lower surface CO₂ concentrations are found during months with an active biosphere (June to October) than the rest of the period, owing to the higher ecosystem productivity over the northern hemisphere and particularly over the Indian subcontinent in response to the availability of monsoon rainfall. Also, the presence of strong southwest monsoon winds from June to September may result in bringing CO₂ depleted air from the southern hemisphere and thereby lowering the CO₂ concentration over the domain. While comparing the seasonal maximum (May) and minimum (September) of CO₂ concentrations measured at the Mauna Loa observatory (MLO) located in Hawaii, Fig. 2b shows a temporal shift of around one month for exhibiting seasonal maximum (April) and minimum (August) CO₂ concentrations. This temporal shift is attributed to the differential impacts of anthropogenic and terrestrial ecosystem activities on atmospheric concentration as well as the long-distance transfer of atmospheric carbon dioxide to remote location (Nomura et al., 2021). MLO observations are generally representative of global mean CO₂ due to the minimal influence of terrestrial ecosystems and anthropogenic activities at remote location. The seasonal variation of monthly averaged CO₂ seen over the Indian subcontinent is mostly dominated by terrestrial carbon fluxes, i.e., net ecosystem exchange (NEE) as seen from the VPRM simulations (see Supplementary Fig. S2).

Further, we see a CO₂ vertical profile with a small vertical gradient (~0.5 ppm within an altitude range of ~500 m to 4000 m) from June to October (Fig. 3). This is likely linked to the increased convective activities associated with the monsoon. The strong vertical gradient in the surface levels as simulated by the LSCE model during the monsoon period is little plausible given the strong vertical mixing expected for this convective period. The considerable inter-model variation in monthly averaged CO₂ concentration profiles as predicted by different global models is problematic as it indicates significant uncertainties in flux estimations over India. A part of this discrepancy can come from the coarse resolution global model's inability to represent transport processes like convection and vertical mixing, strength and distribution of anthropogenic sources and ecosystem activities that operate at fine scales. The extent of this unresolved variability in global models is further explored in Sect. 3.2. The spatial distribution of CO₂ concentration shows structural differences among these models (see Supplementary Fig. S3), indicating a substantial knowledge gap in representing atmospheric CO₂ variability over the Indian subcontinent, which can have severe implications for the country's carbon budget estimations.

3.2 Representation errors in global transport models

The spatio-temporal variability of representation error and the influence of various factors in creating this variability are examined here. The larger the variations that are caused due to sub-grid processes within the grid box of 1° × 1°, the larger the representation error. The derived seasonal differences in structural patterns of the sub-grid variability facilitate to 1) quantify what would be typical representation errors associated with incorporating seasonally varying observations into atmospheric models 2) determine what drives the seasonality in sub-grid variability and ultimately 3) design a possible parameterization of representation error with a seasonal component in the inverse modelling framework as well as identify periods or seasons where the use of this parameterization would be valid to improve our estimations of CO₂ fluxes. Further, the seasonal spatial variability analysis of column averages can provide useful information for the satellite community to gap-fill the



satellite soundings over India when large data gaps and low sounding precision on daily or monthly time scales are present, especially the case for monsoon periods in India.

3.2.1 Spatio-temporal patterns

Representation errors in the surface CO₂ concentrations of a global model at a spatial resolution of 1° × 1° for July and November are shown in Fig. 4. The representation error at 1° × 1° spatial scale reaches values ranging from 0.5 ppm to 5 ppm, which are comparable to the magnitude of variability at hotspot emission regions or half of the seasonal variability of CO₂ over the region (see Fig. 2b). The median representation error is 1.2 ppm at the surface, which is considerably larger than the measurement errors. In the case of high accuracy in situ measurements, the typical uncertainty for CO₂ measurements is less than 0.1 ppm (Andrews et al., 2014). A remarkable feature is the presence of very high representation error over North-East and Western Ghats regions, where the biosphere activity is very prominent. The heterogeneous distribution of biosphere fluxes generates significant sub-grid scale variability that leads to high representation error. Also, we can find high representation error along the foothills of the Himalayas. In addition to the complex terrain, the region over the Ganges basin is characterized by increased anthropogenic activity, which contributes to a larger representation error surrounding this region. High representation error is also found in the coastal regions, ranging from 2 ppm to 5 ppm (median of 4 ppm) due to the temporal covariance between the coastal meteorology and exchange fluxes. The CO₂ fluxes from coastal regions can be transported over the ocean and accumulated in the shallow boundary layer over the ocean. The shallow boundary layer is a characteristic of the marine atmosphere due to the less vertical mixing compared to land regions. Horizontal CO₂ gradients can also be generated by the influence of highly varying biospheric fluxes under different advection patterns over the land and ocean boundary. A similar mechanism is applicable to mountain regions where temporal covariance of mountain-valley circulation and respired CO₂ fluxes are regulated by atmospheric radiation. The terrain-following coordinates as used in the model may also result in spurious tracer concentration gradients over the steep mountain terrain (Beck et al., 2020; Skamarock et al., 2021; Park et al., 2019). Though the mesoscale models are expected to perform better in simulating CO₂ variations over the complex terrain than the coarse models (e.g. Engelen et al., 2002; Gerbig et al., 2003; Ahmadov et al., 2007; Corbin et al., 2008; Lauvaux et al., 2009b; Pillai et al., 2011; Uebel et al., 2017; Agustí-Panareda et al., 2019), they may also suffer from the inadequate representation of complex weather features and associated variability. We can also find individual cells with high representation errors associated with point emission sources such as cities, mining sites, and coal-fired power plants at different parts of the domain. The daily variations in surface representation errors are small within a month, although there exists a clear distinction between daytime and nighttime values (Figure not shown). The nighttime representation error is higher (e.g. a median value of 1.5 ppm for surface during November) compared to the daytime representation error (e.g. a median value of 1.1 ppm for surface during November) throughout the analysed domain. This is expected due to the coupling between nocturnal shallow transport and different flux processes accentuating local effects. During the nighttime, photosynthesis is absent, and respiration is the major biospheric activity, leading to an increase in CO₂ concentration in the atmosphere. The large heterogeneity in flux distribution that is mostly from respired CO₂ fluxes, the shallow boundary layer processes and the weak nocturnal turbulence cause CO₂ to be accumulated locally near the surface with large variations. Compared to July, we find higher representation



480 error in November owing to the wintertime transport with decreased vertical mixing and heterogeneous
biospheric uptake (see Fig. 4).

In the case of XCO₂, the magnitude of sub-grid scale variability is much smaller than that of surface CO₂ (Fig. 5), but it follows a similar spatial pattern. This confirms the dominance of surface-level processes in causing sub-grid variability of column averages. The sub-grid scale variability in XCO₂ reaches up to 2 ppm in some parts of the region, especially where there are high variations in topographic features or point emission sources.
485 The estimated column representation error is thus capable of causing significant biases in the satellite inferred CO₂ fluxes over these regions. Also, the representation error for a large part of the domain is found to be above 0.5 ppm, which is around half of the typical precision of current satellite measurements. Note that the representation error reported here is different from satellite measurement errors (e.g. spectroscopic retrieval error or sampling biases) and tends to be systematic in nature.

490 Figure 6 shows the statistical distribution of the representation error (σ_{CO_2}) sampled over India, during July and November, separated by daytime and nighttime. July shows a median surface representation error of 0.9 ppm and 1.1 ppm during daytime and nighttime respectively, while November shows a median value of 1.1 ppm and 1.4 ppm for daytime and nighttime respectively. In July, 95 % of the representation error is less than 2.1 ppm for daytime (3.9 ppm for nighttime) while it is 3 ppm for daytime (4.2 ppm for nighttime) for November. For
495 column average, median values for representation error are 0.3 ppm and 0.4 ppm for July daytime and November daytime respectively.

To further reduce the effect of random error that might be introduced by short-term weather phenomena, the representation errors ($\sigma_{CO_2(mon)}$) are calculated from the monthly averaged CO₂ field and are denoted as a systematic error (Fig. 6). Uncorrelated errors are expected to decrease when averaging over a sufficiently long
500 period. As expected, the median values of the systematic representation errors are smaller for all cases, showing the effect of random errors. Especially for November when the cyclonic event was present, the values of the systematic errors (in the 95% percentile) for the surface CO₂ are considerably lower than total errors, reducing from 3 ppm (daytime) and 4.2 ppm (nighttime) to 2.2 ppm (daytime) and 3 ppm (nighttime). In the case of column CO₂, this reduction is from 1.1 ppm (daytime) and 0.9 ppm (nighttime) to 0.8 ppm (daytime) and 0.7
505 ppm (nighttime) in the 95% percentile. In contrast to surface representation error (Fig. 6a), median values of nighttime representation errors are found to be slightly lower than daytime representation error for column average (Fig. 6b). To assess the dependence of representation error on possible horizontal resolutions of the global models, we have further derived the representation errors for different spatial resolutions between 0.5° and 4°. As expected, we see reductions in representation errors for both surface and column averaged CO₂ with
510 increasing horizontal resolution of the model (See Fig. 7 & supplementary Fig. S4). During July, the median surface representation error reduced from 1.6 ppm (2 ppm) to 0.6 ppm (0.7 ppm) during daytime (nighttime) while increasing horizontal resolution from 4° to 0.5°. This increment in spatial resolution has also resulted in similar error reductions in November during which the median of surface representation error shows a reduction from 2.4 ppm (2.8 ppm) to 0.7 ppm (0.9 ppm) during daytime (nighttime). In the case of column-averaged
515 values, the median representation error decreased from 0.7 ppm (0.6 ppm) to 0.25 ppm (0.2 ppm) during July daytime (nighttime) and from 0.95 ppm (0.9 ppm) to 0.25 ppm (0.2 ppm) during November daytime



(nighttime). The spatial distribution of representation errors for a model with a horizontal grid resolution of $0.5^\circ \times 0.5^\circ$ (e.g. regional models) is provided in Supplementary Figs. S5 and S6. On average, we find ~33 to 36 % of decrease in daytime representation errors for both months when increasing model grid resolution from 1° to 0.5° . There exists a similar spatial pattern of representation errors for both resolutions of 0.5° and 1° . Though our results indicate a reduction of representation error for regional models with a typical resolution of 0.5° compared to global models with 1° spatial resolution, the emission hotspots and point sources are still pronounced with high sub-grid scale variability, especially during nighttime. The above analyses indicate that the sub-grid variability alone can produce significantly higher errors compared to the measurement errors (e.g., 0.1 ppm as per WMO standards for surface measurements), which necessitates a proper treatment of these errors in models for the optimal estimation of CO_2 fluxes.

3.2.2 Vertical distribution

Figure 8 shows the vertical profile of representation error distribution within different altitude bins. We find that the maximum representation error is in the surface layer, and most of the higher values are found to be within the lowest 4-6 km bins. Also, sub-grid scale variability decreases sharply with increasing altitude. This dominance of variability in surface concentration can be explained by surface flux heterogeneity influencing mole fractions in lower atmospheric layers (PBL) as described in van der Molen and Dolman (2007) and Pillai et al. (2010). There is a slight increase in representation error in the upper tropospheric levels near 12 to 14 km altitude range. This may be associated with the presence of strong circulations in the upper troposphere and lower stratosphere, such as subtropical westerly jets.

3.3. Influence of terrain heterogeneity and flux variability on representation errors

Here we explore the factors influencing the size and patterns of the representation error in coarse models. For this, statistical relationships between representation error and possible explanatory variables are examined for both surface and column-averaged CO_2 . Identifying these factors influencing representation errors and quantifying their local effects facilitate us to further investigate on how these biases in retrieved fluxes can be minimized in global models (see Sect. 3.5).

We find a significant influence of terrain heterogeneity on representation error, which is evident from the spatial maps in Figs. 4 and 5, where the largest sub-grid scale variations are found in the Himalayan regions. Spatial variations in topography produce mesoscale circulation patterns and corresponding variations in atmospheric CO_2 at fine scales. At the same time, there is a plausible additional error in global model simulations related to the insufficient resolution of vertical grids necessary to account for different surface influences (e.g. mountain vs valley). This effect of coarse vertical resolution is excluded in our representation error estimates by preserving the vertical grids used for the high-resolution simulations. To further explore the importance of using the high-resolution topography data on representing the CO_2 variability, we analyse the dependence of terrain variations (as derived from the standard deviation of terrain height) on the distribution of the representation error. We have estimated the statistical dependence of topographic variability within the global climate models' grids on corresponding representation error to estimate the relation between them. Topographic variability within $1^\circ \times 1^\circ$ spatial box is estimated as the standard deviation of topography (m) for all $10 \text{ km} \times 10 \text{ km}$ boxes within the larger grid, and is denoted as σ_{topo} . Bins are created based on the values of this topographic



555 variability, in which different points from different parts of the domain are binned together on the basis of their
standard deviation of topography. Each bin is created with a size of 50 m variation in terrain height. The linear
fit is estimated between the average value of topographic variability within a bin and the average value of
560 representation error of the corresponding points in the bin. Our results show that the terrain heterogeneity alone
can explain about 20-48% of the surface representation errors over the domain. In a similar way, we have
estimated the influence of topographic variability on representation error in the column-averaged model
simulations. It is found that topography alone can explain 45-52 % of representation errors in the column-
averaged simulations.

Further, we estimate the statistical relationship between the surface flux heterogeneity and representation error.
The surface representation error is strongly linked to the biosphere flux variability, and the relationship between
565 heterogeneity in biospheric surface flux (as derived from the standard deviation of VPRM-derived NEE fluxes,
denoted as σ_{bio}) and representation errors depends on the time of the day and season. During daytime when
there is strong ecosystem activity, the dependence of representation error (σ_{CO_2}) on σ_{bio} of surface and column
 CO_2 is found to be ~75-80 % and ~66-74 % respectively. σ_{bio} explains about 62% for the surface CO_2
variability and 48 % for the column variability during July nighttime. However, σ_{CO_2} and σ_{bio} are less correlated
570 (23 % for surface and 19 % for column) during November nighttime. The diurnal difference in the dependence
of σ_{bio} on representation error can be explained by the increased magnitude and spatial variability of daytime
biospheric fluxes in the growing season (primarily due to photosynthesis activities) compared to nighttime
fluxes. Moreover, poor vertical mixing under the stable nocturnal atmospheric conditions with more advection
and drainage flow reduces the influence of surface fluxes on spatial variability of mixing ratios. The dependence
575 of representation error on the anthropogenic flux heterogeneity (as derived from the standard deviation of
EDGAR fluxes, denoted as σ_{ant}) is found to be negligible except for nighttime (13–30 %). We find less
influence of seasonality on the relationship between anthropogenic surface flux heterogeneity and representation
errors (see Supplementary Table S1). Similar to the above analysis with σ_{bio} , the combined effect of
atmospheric stability and flux heterogeneity can explain the diurnal differences of the relationship between σ_{ant}
580 and σ_{CO_2} .

In case of the variability of monthly averages, we see that $\sigma_{\overline{CO_2}(mon)}$ is well explained by σ_{bio} during daytime
(see Supplementary Table S2), as expected. A similar strong correlation can be seen between $\sigma_{\overline{CO_2}(mon)}$ and
 σ_{bio} (23–69 %) during nighttime for surface variability of CO_2 , while there exists only less dependence of
nocturnal column CO_2 variability on local fluxes. This shows the decoupling of the mixing ratios in other parts
585 of the column from the surface during the night due to less vertical mixing, combined with more drainage flow
in the nocturnal boundary layer, which reduces the effect of surface flux variability on the column CO_2
variability.



In general, the above analysis underlines the need for using accurate Digital Elevation Models (DEMs) in the atmospheric transport models as one of the most critical datasets for determining the mesoscale atmospheric flows adequately. Further, the results also indicate the importance of utilising surface fluxes at high resolution.

3.4 Estimation of NEE flux uncertainty due to representation error

By following the assumptions and approach as given in Sect. 2.4, we have estimated the NEE flux uncertainty resulting from the representation errors. The results based on the OSSEs for nine observation sites are given in Table. 3. The scaling factors, which are calculated separately for each site by adjusting the prior fluxes using pseudo-observations, are applied to the VPRM monthly fluxes. The total NEE flux for India estimated by VPRM for July and November are -373.3 MtCO₂ per month and -417.1 MtCO₂ per month, respectively. The flux uncertainties over India that arise solely due to the contribution from the representation error are estimated to be 38.59 (daytime observations) to 30.14 (nighttime observations) MtCO₂ per month (10.33% to 8.07%) for July and 18.42 (daytime observations) to 13.34 (nighttime observations) MtCO₂ per month (4.4% to 3.1%) for November while utilizing data from nine observation stations. The maximum flux uncertainty was found for July due to the enhanced biosphere activity and unresolved convection activities. The estimated uncertainties are considerable for the carbon budget assessment especially given that these errors arise solely from the global models' representation error. Note that calculated representation error does not include other transport error sources such as advection, convection or vertical mixing.

3.5 Possible treatment of representation error in the global model

The simplest possible way to minimize the uncertainty in flux estimation using a coarse model is to construct a parameterization model that can account for the representation error using explanatory variables. For this, we create a multivariate model to capture spatial patterns in the representation error. Employing this parameterization in a global model would thus redefine the likelihood of better estimates (improving the state of knowledge) with variance greater than that of the measurement error in the inverse framework by minimizing the modelling error. The multivariate linear model with explanatory variables that include sub-grid variations of terrain (σ_{topo}), biospheric (σ_{bio}) and anthropogenic (σ_{ant}) fluxes remarkably captures the derived column representation error all over the Indian region during July daytime with a R^2 value of 0.96 (Fig. 9). The difference between the modelled and derived representation error is found to be well below 0.5 ppm in most parts of the domain. Similarly, we have modelled the surface representation error using the linear model with these three explanatory variables and found that the proposed model could capture the derived surface representation error well with a deviation less than 1 ppm in most of the regions (see Supplementary Fig. S7 and Supplementary Table S1 and S2). More work is needed to demonstrate the extent of applicability of this method to minimize the flux uncertainties while utilizing actual observations. Nevertheless, the above finding provides a possibility for a parameterization that can be further developed in inverse models or data assimilation systems, which defines the degrees of freedom for describing the posterior states. Applying this parameterization scheme to the specific problem requires a high-resolution map of the terrain and prior information on anthropogenic and biogenic fluxes. The uncertainties in the topography can significantly impact flux estimation, and the likely reduction of flux uncertainty depends on the accuracy of the DEM employed. The caveat of this linear model is that the uncorrelated spatial variability in the prior and true states of the fluxes is ignored in the present form,



which cannot be the case for the real inverse problems. This assumption obviously hampers the system in achieving the maximum reduction in uncertainty, and further study is needed to refine this model from a practical perspective. We emphasise, however, that the above parameterization does not require a high-resolution simulation of transport, which has high computational costs.

630 4. Conclusion

Given the upcoming availability of atmospheric observations over India, significant effort is required to critically enhance the modelling capabilities to derive carbon budgets over India within the definite confidence intervals and at scales relevant to the ecosystem and countrywide policy-making. The misrepresentation of mesoscale transport phenomena and unresolved flux variations in modelling systems operating on coarse grids
635 hinders the optimal utilization of observations. In this context, the present study quantifies the spatial variability of atmospheric CO₂ mixing ratio over India that is not resolved by the coarse models and assesses their impact on flux estimations. We demonstrate the potential of a simple parameterization scheme to model these unresolved variations in the coarse models for minimizing the uncertainty in retrieved fluxes.

A large spread among global model simulations in representing monthly averaged CO₂ concentration profiles
640 indicates a considerable knowledge gap in the estimations of fluxes even at a monthly scale. It can be argued that a significant part of these differences arises due to the lack of observational constraints over India, which leads to a possible compensatory model artefact over this region in order to match the global mass constraint. At the same time, it is also expected that the spatial variability of the observed atmospheric CO₂ mole fractions can be large so that these coarse models fail to represent them adequately. For instance, we find that the unresolved
645 variations (representation error) of global models with a spatial resolution of 1° × 1° can be ~1.5 ppm on average for the surface CO₂ that is even larger than the currently reported inter-global model differences. Similarly, the average representation error estimated for the column-averaged CO₂ is ~1.1 ppm. These estimated values are larger than the corresponding measurement errors, which cause the inverse optimization to infer a state that is not close to the truth as is required in the regional CO₂ budget for various applications.

650 Coastal areas and mountains have particularly high representation errors (≈4 ppm for surface CO₂). Emission hotspots can also lead to significant CO₂ variability near the surface as large as ≈8 ppm. Larger values are typically associated with the nocturnal shallow boundary layer dynamics and the stronger respiration signal with considerable flux variability. These findings are consistent with Pillai et al. (2010), which show that there exist spatial differences in the sub-grid variability for both surface and column CO₂. Although the magnitude of the
655 sub-grid variability of the total column is an order of magnitude smaller than the variability at the surface, the spatial pattern remains similar for both, owing to the dominance of surface heterogeneity in topography and fluxes. With the underlying assumptions, the total uncertainty in optimized fluxes solely due to the unresolved sub-grid variations is estimated to be 3.1 to 10.3% of the total NEE while utilizing pseudo-data from nine observation stations over India. Increasing the spatial and temporal resolutions of the transport models can
660 generally capture the mesoscale features and associated CO₂ gradients, thereby reducing the representation error. Increasing the model's resolution from 1° to 0.5° has shown an improvement in capturing variability with representation error reduction of 33% and 36% for summer time and winter time, respectively. By showing the existence of unresolved variability in 0.5° resolution with a similar spatial pattern of error as of 1° spatial



665 resolution, we demonstrate the need for a much finer resolution than 0.5° for representing the atmospheric CO_2
variability over India. However, merely increasing the resolution without having a realistic representation of
terrain heterogeneity and flux (both natural and anthropogenic) variability would not be beneficial. The
uncertainties in the high-resolution fluxes can worsen the model's skills, whose effect would not be more
pronounced at coarser resolutions due to the diffusive nature of fluxes, as seen in Agustí-Panareda et al. (2019).

670 A parameterization scheme with explanatory variables of sub-grid variations of terrain, biospheric and
anthropogenic fluxes is shown to capture a considerable fraction of expected representation error in the global
model. The proposed method is easy to implement in the coarse models as it does not require computationally
expensive transport simulations at high resolution. As we see a significant dependence of the distribution of sub-
grid variability on terrain variations, our results reinforce the requirement for using accurate DEMs in the
atmospheric transport models. The biosphere flux variability explains about 62 to 80% of the surface
675 representation errors over the Indian region, indicating the need for using precise high-resolution surface fluxes.

Overall, we show that the mesoscale transport mechanisms and flux variability contribute to fine-scale CO_2
variations that the current-generation models cannot resolve. Our findings indicate that the models need to be
critically improved to capture mesoscale variations associated with horizontal and vertical transport and fine-
scale flux variability to maximize the potential of highly precise and accurate measurements. Our results provide
680 a baseline for overcoming the shortcomings mentioned above and accounting for the realistic distribution of
atmospheric CO_2 to improve the estimation of surface fluxes through inverse modelling.

Code/Data availability

The WRF-Chem source code is publicly available at <https://ruc.noaa.gov/wrf/wrf-chem/> (last access: 10 August
685 2019). The CarbonTracker (CT-2019B) products are available online at <http://carbontracker.noaa.gov> (last
access: 21 July 2020, Jacobson et al., 2020). The data from the CarboScope inversion system are available
online at <http://www.bgc-jena.mpg.de/CarboScope/> (last access: 20 July 2020, Rödenbeck et al., 2003). The data
from the LSCE modelling system used in this study are available at <http://atmosphere.copernicus.eu> (last access:
22 July 2020, Chevallier et al., 2019). The sub-grid variability products based on the WRF-GHG model
690 simulations can be accessed from <https://zenodo.org/record/6616466> (last access: 23 May 2022, Thilakan and
Pillai, 2022). The WRF-GHG model CO_2 simulations used for this study are available upon request to the
corresponding author, Dhanyalekshmi Pillai (dhanya@iiserb.ac.in, kdhanya@bgc-jena.mpg.de). The EDGAR
data used in this study are publicly available at <https://edgar.jrc.ec.europa.eu/> (last access: 15 March 2020,
Crippa et al., 2018). The GFAS data are publicly available at <http://apps.ecmwf.int/datasets/data/cams-gfas/> (last
695 access: 15 March 2020, Kaiser et al., 2012). The ERA5 data are available at
<https://cds.climate.copernicus.eu/cdsapp#!home> (last access: 18 March 2020, Hersbach et al., (2020)).

Author Contribution

DP designed the study and performed the model simulations. VT performed the analysis and wrote the paper.
700 VT and DP interpreted the results. CG, MG, AR and TAM provided significant input to the interpretation, and
the improvement of the paper. All authors discussed the results and commented on the paper.



Competing interests

The authors declare they have no conflict of interest.

Acknowledgements

705 This study is supported by the funding from the Max Planck Society allocated to the Max Planck Partner Group
at IISERB and the Science and Engineering Research Board (SERB) through Early Career Research Award
(ECR/2018/001111) to DP. TAM acknowledges the financial support provided by SERB grant (Junior Research
Fellowship). We acknowledge the support of IISERB's high performance cluster system for computations, data
analysis and visualisation. The WRF-Chem simulations were done on the high-performance cluster Mistral of
710 the Deutsches Klimarechenzentrum GmbH (DKRZ). We thank the Editor and both referees for their
constructive comments.

References

- 715 Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., Curcoll, R.,
Engelen, R., Langerock, B., Law, R. M., Loh, Z., Morguí, J. A., Parrington, M., Peuch, V.-H.,
Ramonet, M., Roehl, C., Vermeulen, A. T., Warneke, T., and Wunch, D.: Modelling CO₂ weather – why
horizontal resolution matters, *Atmospheric Chemistry and Physics*, 19, 7347–7376,
<https://doi.org/10.5194/acp-19-7347-2019>, 2019.
- 720 Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Neining, B., Dolman, A. J., and Sarrat, C.: Mesoscale
covariance of transport and CO₂ fluxes: Evidence from observations and simulations using the WRF-
VPRM coupled atmosphere-biosphere model, *Journal of Geophysical Research*, 112, D22107,
<https://doi.org/10.1029/2007jd008552>, 2007.
- 725 Ahmadov, R., Gerbig, C., Kretschmer, R., Körner, S., Rödenbeck, C., Bousquet, P., and Ramonet, M.:
Comparing high resolution WRF-VPRM simulations and two global CO₂ transport models with coastal
tower measurements of CO₂, *Biogeosciences*, 6, 807–817, <https://doi.org/10.5194/bg-6-807-2009>,
2009.
- 730 Ahmadov, R., McKeen, S. A., Robinson, A. L., Bahreini, R., Middlebrook, A. M., de Gouw, J. A., Meagher, J.,
E.-Y., Edgerton, E., Shaw, S., and Trainer, M.: A volatility basis set model for summertime secondary
organic aerosols over the eastern United States in 2006, *Journal of Geophysical Research:
Atmospheres*, 117, D06301, <https://doi.org/10.1029/2011JD016831>, 2012.
- 735 Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis,
D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crowell, M. J., Fischer, M. L.,
Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D.
E., Munger, J. W., and Tans, P. P.: CO₂, CO, and CH₄ measurements from tall towers in the NOAA
Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation,
uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts,
Atmospheric Measurement Techniques, 7, 647–687, <https://doi.org/10.5194/amt-7-647-2014>, 2014.
- 740 Beck, V., Koch, T., Kretschmer, R., Marshall, J., Ahmadov, R., Gerbig, C., Pillai, D., and Heimann, M.: The
WRF Greenhouse Gas Model (WRF-GHG), Tech. Rep. 25, Max Planck Institute for Biogeochemistry,
Jena, Germany, 2011.



- Beck, J., Brown, J., Dudhia, J., Gill, D., Hertneck, T., Klemp, J., Wang, W., Williams, C., Hu, M., James, E., Kenyon, J., Smirnova, T., & Kim, J.: An Evaluation of a Hybrid, Terrain-Following Vertical Coordinate in the WRF-Based RAP and HRRR Models, *Weather and Forecasting*, 35(3), 1081–1096, <https://doi.org/10.1175/WAF-D-19-0146.1>, 2020.
- 745 Bergamaschi, P., Danila, A., Weiss, R. F., Ciais, P., Thompson, R. L., Brunner, D., Levin, I., Meijer, Y., Chevallier, F., Bovensmann, H., Crisp, D., Basu, S., Dlugokencky, E., Engelen, R., Gerbig, C., Günther, D., Hammer, S., Henne, S., Houweling, S., Peylin, P., Pinty, B., Ramonet, M., Reimann, S., Röckmann, T., Schmidt, M., Strogies, M., Sussams, J., Tarasova, O., van Aardenne, J., Vermeulen, A. T., and Vogel, F.: Atmospheric monitoring and inverse modelling for verification of greenhouse gas inventories, Publications Office of the European Union, 2018.
- 750 Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P., and Tans, P. r. P.: Regional Changes in Carbon Dioxide Fluxes of Land and Oceans Since 1980, *Science*, 290, 1342–1346, <https://doi.org/10.1126/science.290.5495.1342>, 2000.
- Bréon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I., Ramonet, M., Dieudonné, E., Lopez, M. and Schmidt, M., Perrussel, O., and Ciais, P.: An attempt at estimating Paris area CO₂ emissions from atmospheric concentration measurements, *Atmospheric Chemistry and Physics*, 15, 1707–1724, <https://doi.org/10.5194/acp-15-1707-2015>, 2015.
- 755 Broquet, G., Chevallier, F., Bréon, F.-M., Kadyrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO₂ ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, *Atmospheric Chemistry and Physics*, 13, 9039–9056, <https://doi.org/10.5194/acp-13-9039-2013>, 2013.
- 760 Carouge, C., Bousquet, P., Peylin, P., Rayner, P. J., and Ciais, P.: What can we learn from European continuous atmospheric CO₂ measurements to quantify regional fluxes – Part 1: Potential of the 2001 network, *Atmospheric Chemistry and Physics*, 10, 3107–3117, <https://doi.org/10.5194/acp-10-3107-2010>, 2010.
- 765 Cervarich, M., Shu, S., Jain, A. K., Arneth, A., Canadell, J., Friedlingstein, P., Houghton, R. A., Kato, E., Koven, C., Patra, P., Poulter, B., Sitch, S., Stocker, B., Viovy, N., Wiltshire, A., and Zeng, N.: The terrestrial carbon budget of South and Southeast Asia, *Environmental Research Letters*, 11, 105006, <https://doi.org/10.1088/1748-9326/11/10/105006>, 2016.
- 770 Chen, H. W., Zhang, L. N., Zhang, F., Davis, K. J., Lauvaux, T., Pal, S., Gaudet, B., and DiGangi, J. P.: Evaluation of Regional CO₂ Mole Fractions in the ECMWF CAMS Real-Time Atmospheric Analysis and NOAA CarbonTracker Near-Real-Time Reanalysis with Airborne Observations from ACT-America Field Campaigns, *Journal of Geophysical Research: Atmospheres*, 124, 8119–8133, <https://doi.org/10.1029/2018jd029992>, 2019.
- 775 Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research*, 110, D24309, <https://doi.org/10.1029/2005jd006390>, 2005.
- Chevallier, F., Bréon, F. M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO₂ sources and sinks: Theoretical study in a variational data assimilation



- 780 framework, *Journal of Geophysical Research*, 112, D09307, <https://doi.org/10.1029/2006JD007375>, 2007.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *Journal of Geophysical Research*, 115, D21307, <https://doi.org/10.1029/2010jd013887>, 2010.
- 785
- Chevallier, F.: On the parallelization of atmospheric inversions of CO₂ surface fluxes within a variational framework, *Geoscientific Model Development*, 6, 783–790, <https://doi.org/10.5194/gmd-6-783-2013>, 2013.
- 790
- Chevallier, F., Remaud, M., O’Dell, C. W., Baker, D., Peylin, P., and Cozic, A.: Objective evaluation of surface- and satellite-driven carbon dioxide atmospheric inversions, *Atmospheric Chemistry and Physics*, 19, 14233–14251, <https://doi.org/10.5194/acp-19-14233-2019>, 2019.
- 795
- Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregón, A., Rayner, P. J., Miller, C., Gobron, N., Kinerman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R. J., Balsamo, G., Bopp, L., Bréon, F.-M., Broquet, G., Dargaville, R., Battin, T. J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J. G., Cook, R. B., DeFries, R., Engelen, R., Gurney, K. R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luysaert, S., Miller, J., Moriyama, T., Moulin, C., Myneni, R. B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J.-D., Piao, S. L., Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., van der Werf, G., Wickland, D., Williams, M., and Zehner, C.: Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system, *Biogeosciences*, 11, 3547–3602, <https://doi.org/10.5194/bg-11-3547-2014>, 2014.
- 800
- 805
- Corbin, K. D., Denning, A. S., Lu, L., Wang, J.-W., and Baker, I. T.: Possible representation errors in inversions of satellite CO₂ retrievals, *Journal of Geophysical Research*, 113, D02301, doi:10.1029/2007JD008716, 2008.
- Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U., Olivier, J. G. J., Pagliari, V., and Janssens-Maenhout, G.: Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2, *Earth System Science Data*, 10, 1987–2013, <https://doi.org/10.5194/essd-10-1987-2018>, 2018.
- 810
- Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J., Vignati, E.: EDGAR v6.0 Greenhouse Gas Emissions. European Commission, Joint Research Centre (JRC) [Dataset] PID: <http://data.europa.eu/89h/97a67d67-c62e-4826-b873-9d972c4f670b>, 2021.
- 815
- Department of Agriculture & Cooperation, Ministry of Agriculture (DAC/MA): Agricultural Statistics at a Glance 2014, Directorate of Economics and Statistics, Department of Agriculture and Cooperation (DAC), Ministry of Agriculture (MA), Government of India, OUP, New Delhi, India, ISBN 0-19-945965-7, available at: <https://eands.dacnet.nic.in/PDF/Agricultural-Statistics-At-Glance2014.pdf> (last access: 11 February 2021), 2015.



- 820 Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCO₂ over the global ocean, *Geoscientific Model Development*, 12, 2091–2105, <https://doi.org/10.5194/gmd-12-2091-2019>, 2019.
- Engelen, R. J., Denning, A. S., Gurney, K. R., and TransCom3 modelers: On error estimation in atmospheric CO₂ inversions, *Journal of Geophysical Research*, 107, 4635, doi:10.1029/2002JD002195, 2002.
- 825 Enting, I. G.: *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, <https://doi.org/10.1017/CBO9780511535741>, 2002.
- Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Site h, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bas trikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lau vset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Mun ro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G. ., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B. ., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, *Earth System Science Data*, 11, 1783–1838, <https://doi.org/10.5194/essd-11-1783-2019>, 2019.
- 835 Gadgil, S.: The Indian Monsoon and its Variability, *Annual Review of Earth and Planetary Sciences*, 31, 429–467, <https://doi.org/10.1146/annurev.earth.31.100901.141251>, 2003.
- Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO₂ with atmospheric observations over a continent: I. Observed spatial variability from airborne platforms, *Journal of Geophysical Research: Atmospheres*, 108, D24-4756, <https://doi.org/10.1029/2002jd003018>, 2003.
- 845 Gerbig, C., Körner, S., and Lin, J. C.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, *Atmospheric Chemistry and Physics*, 8, 591–602, <https://doi.org/10.5194/acp-8-591-2008>, 2008.
- Goswami, B. N. and Xavier, P. K.: Dynamics of “internal” interannual variability of the Indian summer monsoon in a GCM, *Journal of Geophysical Research*, 110, D24104, <https://doi.org/10.1029/2005jd006042>, 2005.
- 850 Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled online chemistry within the WRF model, *Atmospheric Environment*, 39, 6957-6975, <https://doi.org/10.1016/j.atmosenv.2005.04.027>, 2005.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, Julien an d Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G. o., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diaman takis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G. a., de Rosnay, P.,



- 860 Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19, 3515–3556, <https://doi.org/10.5194/acp-19-3515-2019>, 2019.
- 865 Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Arlyn Andrews, Mund, J., Ott, L., Collatz, G. J., Aalto, T., Afshar, S., Aikin, K., Aoki, S., Apadula, F., Baier, B., Bergamaschi, P., Beyersdorf, A., Biraud, S. C., Bollenbacher, A., Bowling, D., Brailsford, G., Abshire, J. B., Chen, G., Huilin Chen, Lukasz Chmura, Sites Climadat, Colomb, A., Conil, S., Cox, A., Cristofanelli, P., Cuevas, E., Curcoll, R., Sloop, C. D., Davis, K., Wekker, S. D., Delmotte, M., DiGangi, J. P., Dlugokencky, E., Ehleringer, J., Elkins, J. W., Emmenegger, L., Fischer, M. L., Forster, G., Frumau, A., Galkowski, M., Gatti, L. V., Gloor, E., Griffis, T., Hammer, S., Haszpra, L., Hatakka, J., Heliasz, M., Hensen, A., Hermansen, O., Hintsa, E., Holst, J., Jaffe, D., Karion, A., Kawa, S. R., Keeling, R., Keronen, P., Kolari, P. a., Kominkova, K., Kort, E., Krummel, P., Kubistin, D., Labuschagne, C., Langenfelds, R., Laurent, O., 875 Laurila, T., Lauvaux, T., Law, B., Lee, J., Lehner, I., Leuenberger, M., Levin, I., Levula, J., Lin, J., Lindauer, M., Loh, Z., Lopez, M., Lujikx, I. T., Myhre, C. L., Machida, T., Mammarella, I., Manca, G., Manning, A., Manning, A., Marek, M. V., Marklund, P., Martin, M. Y., Matsueda, H., McKain, K., Meijer, H., Meinhardt, F., Miles, N., Miller, C. E., Mölder, M., Montzka, S., Moore, F., Josep-Anton Morgui, Morimoto, S., Munger, B., Jaroslaw Necki, Newman, S., Nichol, S., Niwa, Y., O'Doherty, S., 880 Mikael Ottosson-Löfvenius, Paplawsky, B., Peischl, J., Peltola, O., Jean-Marc Pichon, Piper, S., Plass-Dömler, Christian and Ramonet, M., Reyes-Sanchez, E., Richardson, S., Riris, H., Ryerson, T., Saito, K., Sargent, M., Sasaki, M., Sawa, Y., Say, D., Scheeren, B., Schmidt, M., Schmidt, A., Schumacher, M., Shepson, P., Shook, M., Stanley, K., Steinbacher, M., Stephens, B., Sweeney, C., Thoning, K., Torn, M., Turnbull, J. n., Tørseth, K., Bulk, P. V. D., Dinter, D. V., Vermeulen, A., 885 Viner, B., Vitkova, G., Walker, S., Weirauch, D., Wofsy, S., Worthy, D., Dickon Young, and Mirosław Zimnoch: CarbonTracker CT2019B, <https://doi.org/10.25925/20201008>, 2020.
- Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P.: On the representation error in data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 144, 1257–1278, <https://doi.org/10.1002/qj.3130>, 2017.
- 890 Jung, M., Henkel, K., Herold, M., Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modelling, *Remote Sensing of Environment*, 101(4), 534–553, <https://doi.org/10.1016/j.rse.2006.01.020>, 2006.
- Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, *Biogeosciences*, 9, 527–554, <https://doi.org/10.5194/bg-9-527-2012>, 2012.



- Kretschmer, R., Gerbig, C., Karstens, U., Biavati, G., Vermeulen, A., Vogel, F., Hammer, S., and Totsche, K. U.: Impact of optimized mixing heights on simulated regional atmospheric transport of CO₂, *Atmospheric Chemistry and Physics*, 14, 7149–7172, <https://doi.org/10.5194/acp-14-7149-2014>, 2014.
- 900 Krishnamurthy, V. and Shukla, J.: Seasonal persistence and propagation of intraseasonal patterns over the Indian monsoon region, *Climate Dynamics*, 30, 353–369, <https://doi.org/10.1007/s00382-007-0300-7>, 2007.
- Krishnamurthy, V. and Ajayamohan, R. S.: Composite Structure of Monsoon Low Pressure Systems and Its Relation to Indian Rainfall, *Journal of Climate*, 23, 4285–4305, <https://doi.org/10.1175/2010JCLI2953.1>, 2010.
- 905 Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J., and Rayner, P. J.: Structure of the transport uncertainty in mesoscale inversions of CO₂ sources and sinks using ensemble model simulations, *Biogeosciences*, 6, 1089–1102, <https://doi.org/10.5194/bg-6-1089-2009>, 2009a.
- Lauvaux, T., B. Gioli, C. Sarrat, P. J. Rayner, P. Ciais, F. Chevallier, J. Noilhan, F. Miglietta, Y. Brunet, E. Ceschia, H. Dolman, J. A. Elbers, C. Gerbig, R. Hutjes, N. Jarosz, D. Legain, M. Uliasz: Bridging the gap between atmospheric concentrations and local ecosystem measurements, *Geophysical Research Letters*, 36, L19809, <https://doi.org/10.1029/2009GL039574>, 2009b.
- 910 Landschützer, P., Gruber, N., and Bakker, D. C. E.: An observation-based global monthly gridded sea surface pCO₂ product from 1982 onward and its monthly climatology (NCEI Accession 0160558), <https://doi.org/10.7289/v5z899n6>, 2017.
- 915 Li, D., Vogel, B., Müller, R., Bian, J., Günther, G., & Riese, M.: Tropical cyclones reduce ozone in the tropopause region over the western Pacific: An analysis of 18 years ozonesonde profiles, *Earth's Future*, 9, e2020EF001635, <https://doi.org/10.1029/2020EF001635>, 2021.
- Li, W., Ciais, P., Wang, Y., Peng, S., Broquet, G., Ballantyne, A. P., Cana dell, J. G., Cooper, L., Friedlingstein, P., Le Quéré, C., Myneni, R. B., Peters, G. P., Piao, S., and Pongratz, J.: Reducing uncertainties in decadal variability of the global carbon budget with multiple datasets, *Proceedings of the National Academy of Sciences*, 113, 13104–13108, <https://doi.org/10.1073/pnas.1603956113>, 2016.
- 920 Liang, A., Gong, W., Han, G., and Xiang, C.: Comparison of Satellite-Observed XCO₂ from GOSAT, OCO-2, and Ground-Based TCCON, *Remote Sensing*, 9, 1033, <https://doi.org/10.3390/rs9101033>, 2017.
- 925 Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research: Atmospheres*, 108, 4493, <https://doi.org/10.1029/2002JD003161>, 2003.
- 930 Lin, X., Indira, N. K., Ramonet, M., Delmotte, M., Ciais, P., Bhatt, B. C., Reddy, M. V., Angchuk, D., Balakrishnan, S., Jorphail, S., Dorjai, T., Mahey, T. T., Patnaik, S., Begum, M., Brenninkmeijer, C., Durairaj, S., Kirubakaran, R., Schmidt, M., Swathi, P. S., Vinithkumar, N. V., Yver Kwok, C., and Gaur, V. K.: Long-lived atmospheric trace gases measurements in flask samples from three stations in India, *Atmospheric Chemistry and Physics*, 15, 9819–9849, <https://doi.org/10.5194/acp-15-9819-2015>, 2015.



- 935 Liu, Y., Yue, T., Zhang, L., Zhao, N., Zhao, M., and Liu, Y.: Simulation and analysis of XCO₂ in North China based on high accuracy surface modeling, *Environmental Science and Pollution Research*, 25, 27378–27392, <https://doi.org/10.1007/s11356-018-2683-x>, 2018.
- Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochemical Cycles*, 22, GB2005, <https://doi.org/10.1029/2006gb002735>, 2008.
- 940 Massart, S., Agustí-Panareda, A., Heymann, J., Buchwitz, M., Chevallier, F., Reuter, M., Hilker, M., Burrows, J. P., Deutscher, N. M., Feist, D. G., Hase, F., Sussmann, R., Desmet, F., Dubey, M. K., Griffith, D. W. T., Kivi, R., Petri, C., Schneider, M., and Velasco, V. A.: Ability of the 4-D-Var analysis of the GOSAT BESD XCO₂ retrievals to characterize atmospheric CO₂ at large and synoptic scales, *Atmospheric Chemistry and Physics*, 16, 1653–1671, <https://doi.org/10.5194/acp-16-1653-2016>, 2016.
- 945 McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, *Proceedings of the National Academy of Sciences*, 109, 8423–8428, <https://doi.org/10.1073/pnas.1116645109>, 2012.
- 950 Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Law, R. M.: Precision requirements for space-based XCO₂ data, *Journal of Geophysical Research*, 112, D10314, <https://doi.org/10.1029/2006JD007659>, 2007.
- 955 NOAA/ESRL: Dipole Mode Index, available at: https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/DMI/, last access: 11 February 2022a.
- NOAA/ESRL: Multivariate ENSO Index, available at: <https://www.esrl.noaa.gov/psd/enso/mei/>, last access: 11 February 2022b.
- 960 Nomura, S., Naja, M., Ahmed, M. K., Mukai, H., Terao, Y., Machida, T., Sasakawa, M., and Patra, P. K.: Measurement report: Regional characteristics of seasonal and long-term variations in greenhouse gases at Nainital, India, and Comilla, Bangladesh, *Atmos. Chem. Phys.*, 21, 16427–16452, <https://doi.org/10.5194/acp-21-16427-2021>, 2021.
- 965 O'Dell, Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmospheric Measurement Techniques*, 5, 99–121, <https://doi.org/10.5194/amt-5-99-2012>, 2012.
- 970 Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., 625 Randall, D. A., Kawa, S. R., Pauluis, O., and Doney, S. C.: Moist synoptic transport of CO₂ along the mid-latitude storm track, *Geophysical Research Letters*, 38, L09804, <https://doi.org/10.1029/2011gl047238>, 2011.



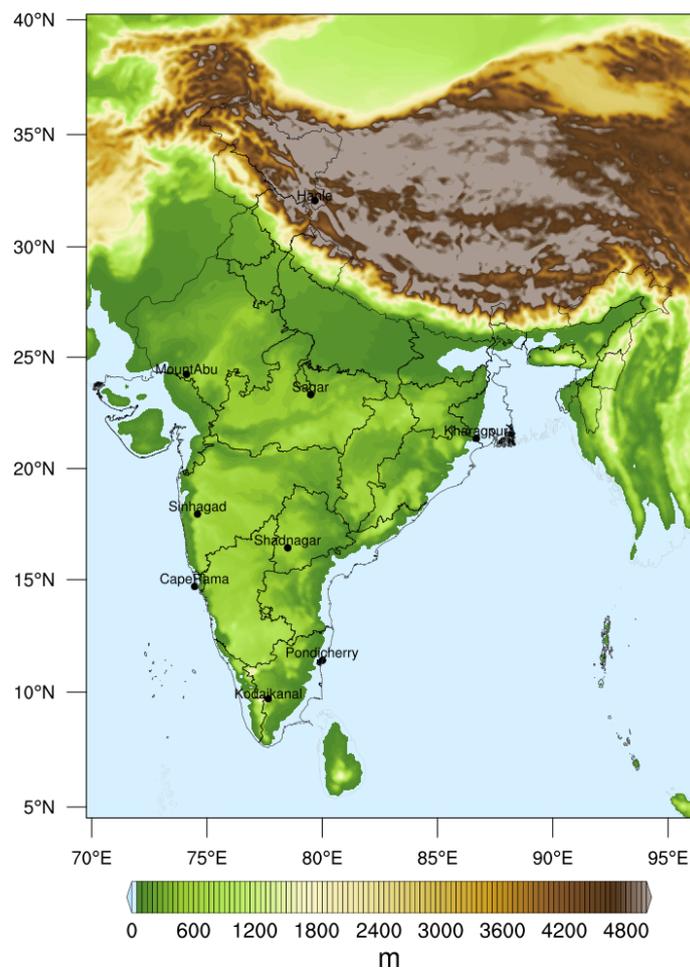
- 975 Park, M., Randel, W. J., Gettelman, A., Massie, S. T., and Jiang, J. H.: Transport above the Asian summer monsoon anticyclone inferred from Aura Microwave Limb Sounder tracers, *Journal of Geophysical Research*, 112, D16309, <https://doi.org/10.1029/2006JD008294>, 2007.
- 980 Park, C., Gerbig, C., Newman, S., Ahmadov, R., Feng, S., Gurney, K. R., Carmichael, G. R., Park, S.-Y., Lee, H.-W., Goulden, M., Stutz, J., Peischl, J., and Ryerson, T.: CO₂ Transport Variability, and Budget over the Southern California Air Basin Using the High-Resolution WRF-VPRM Model during the CalNex 2010 Campaign, *Journal of Applied Meteorology and Climatology*, 57, 1337–1352, <https://doi.org/10.1175/jamc-d-17-0358.1>, 2018.
- Park, S., Klemp, J. B., & Kim, J.: Hybrid Mass Coordinate in WRF-ARW and Its Impact on Upper-Level Turbulence Forecasting, *Monthly Weather Review*, 147(3), 971–985, <https://doi.org/10.1175/MWR-D-18-0334.1>, 2019.
- 985 Patra, P. K., Ishizawa, M., Maksyutov, S., Nakazawa, T., and Inoue, G.: Role of biomass burning and climate anomalies for land-atmosphere carbon fluxes based on inverse modeling of atmospheric CO₂, *Global Biogeochemical Cycles*, 19, GB3005, <https://doi.org/10.1029/2004gb002258>, 2005.
- Patra, P. K., Niwa, Y., Schuck, T. J., Brenninkmeijer, C. A. M., Machida, T., Matsueda, H., and Sawa, Y.: Carbon balance of South Asia constrained by passenger aircraft CO₂ measurements, *Atmospheric Chemistry and Physics*, 11, 4163–4175, <https://doi.org/10.5194/acp-11-4163-2011>, 2011.
- 990 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of Sciences*, 104, 18925–18930, <https://doi.org/10.1073/pnas.0708986104>, 2007.
- 995 Pillai, D., Gerbig, C., Marshall, J., Ahmadov, R., Kretschmer, R., Koch, T., and Karstens, U.: High resolution modeling of CO₂ over Europe: implications for representation errors of satellite retrievals, *Atmospheric Chemistry and Physics*, 10, 83–94, <https://doi.org/10.5194/acp-10-83-2010>, 2010.
- 1000 Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., and Lavrié, J. V.: High-resolution simulations of atmospheric CO₂ over complex terrain – representing the Ochsenkopf mountain tall tower, *Atmospheric Chemistry and Physics*, 11, 7445–7464, <https://doi.org/10.5194/acp-11-7445-2011>, 2011.
- 1005 Pillai, D., Gerbig, C., Kretschmer, R., Beck, V., Karstens, U., Neininger, B., and Heimann, M.: Comparing Lagrangian and Eulerian models for CO₂ transport – a step towards Bayesian inverse modeling using WRF/STILT-VPRM, *Atmospheric Chemistry and Physics*, 12, 8979–8991, <https://doi.org/10.5194/acp-12-8979-2012>, 2012.
- Pillai, D., Buchwitz, M., Gerbig, C., Koch, T., Reuter, M., Bovensmann, H., Marshall, J., and Burrows, J. P.: Tracking city CO₂ emissions from space using a high-resolution inverse modelling approach: a case study for Berlin, Germany, *Atmospheric Chemistry and Physics*, 16, 9591–9610, <https://doi.org/10.5194/acp-16-9591-2016>, 2016.
- 1010 Ravi Kumar, K., Valsala, V., Tiwari, Y. K., Revadekar, J. V., Pillai, P., Chakraborty, S., and Murtugudde, R.: Intra-seasonal variability of atmospheric CO₂ concentrations over India during summer monsoons, *Atmospheric Environment*, 142, 229–237, <https://doi.org/10.1016/j.atmosenv.2016.07.023>, 2016.



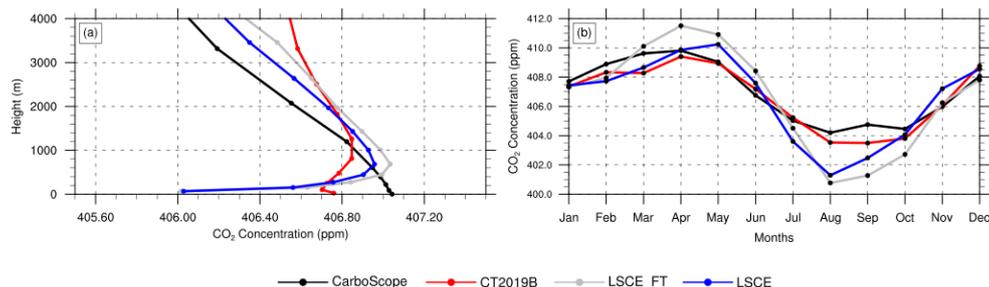
- 1015 Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmospheric Chemistry and Physics*, 3, 1919–1964, <https://doi.org/10.5194/acp-3-1919-2003>, 2003.
- Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to inter-annual climatic variations? A quantification based on atmospheric CO₂ data, *Biogeosciences*, 15, 2481–2498, <https://doi.org/10.5194/bg-15-2481-2018>, 2018a.
- 1020 Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: History of El Niño impacts on the global carbon cycle 1957–2017: a quantification from atmospheric CO₂ data, *Phil. Trans. R. Soc. B*, 373, 20170303, <https://doi.org/10.1098/rstb.2017.0303>, 2018b.
- Sarrat, C., Noilhan, J., Dolman, A. J., Gerbig, C., Ahmadov, R., Tolk, L. F., Meesters, A. G. C. A., Hutjes, R. W. A. and Ter Maat, H. W., Pérez-Landa, G., and Donier, S.: Atmospheric CO₂ modeling at the regional scale: an intercomparison of 5 meso-scale atmospheric models, *Biogeosciences*, 4, 1115–1126, <https://doi.org/10.5194/bg-4-1115-2007>, 2007.
- 1025 Schimel, D. S., House, J. I., Hibbard, K. A., Bousquet, P., 665 Ciais, P., Peylin, P., Braswell, B. H., Apps, M. J., Baker, D., Bondeau, A., Canadell, J., Churkina, G., Cramer, W., Denning, A. S., Field, C. B., Friedlingstein, P., Goodale, C., Heimann, M., Houghton, R. A., Melillo, J. M., Moore, B., Murdiyarso, D., Noble, I., Pacala, S. W., Prentice, I. C., Raupach, M. R., Rayner, P. J., Scholes, R. J., Steffen, W.
- 1030 L., and Wirth, C.: Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems, *Nature*, 414, 169–172, <https://doi.org/10.1038/35102500>, 2001.
- Schimel, D., Stephens, B. B., and Fisher, J. B.: Effect of increasing CO₂ on the terrestrial carbon cycle, *Proceedings of the National Academy of Sciences*, 112, 436–441, <https://doi.org/10.1073/pnas.1407302112>, 2014.
- 1035 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Duda, D. M. B. M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 3, NCAR Tech. Note, NCAR/TN-468+STR, Tech. rep., National Center for Atmosphere Research, Boulder, Colorado, USA, <https://doi.org/doi:10.5065/D68S4MVH>, 2008.
- 1040 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker D. M., Huang, X. -yu.: A Description of the Advanced Research WRF Model Version 4.3., <https://doi:10.5065/1dfh-6p97>, 2021.
- Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H.: The CO₂ release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, *Atmospheric Chemistry and Physics*, 11, 6855–6870, <https://doi.org/10.5194/acp-11-6855-2011>, 2011.
- 1045 Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., and Gernot Friederich, B. H., Chavez, F., Sabine, C., Watson, A., Bakker, D. C., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., and Jon Olafsson, M. H., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C., Ile, B. D., Bates, N., and de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO₂, and net sea–air CO₂ flux over the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, 56, 554 – 577, <https://doi.org/10.1016/j.dsr2.2008.12.009>, 2009.
- 1050



- 1055 Tiwari Y. K., Patra P. K., Chevallier F., Francey R. J., Krummel P. B., Allison C. E., Revadekar J. V., Chakraborty S., Langenfelds R. L., Bhattacharya S. K., Borole D. V., Kumar K. R. and Steele L. P.: CO₂ observations at Cape Rama, India for the period of 1993–2002: implications for constraining Indian emissions. *Current Science, Indian Academy of Sciences*, 101, 1562–1568, 2011
- Thilakan, V. and Pillai, D.: Representation error in global model CO₂ simulations over India [Data set], Zenodo, <https://doi.org/10.5281/zenodo.6616466>, 2022.
- 1060 Tolk, L. F., Meesters, A. G. C. A., Dolman, A. J., and Peters, W.: Modelling representation errors of atmospheric CO₂ mixing ratios at a regional scale, *Atmospheric Chemistry and Physics*, 8, 6587–6596, <https://doi.org/10.5194/acp-8-6587-2008>, 2008.
- Uebel, M., Herbst, M. and Bott, A.: Mesoscale simulations of atmospheric CO₂ variations using a high-resolution model system with process-based CO₂ fluxes. *Q.J.R. Meteorol. Soc.*, 143: 1860–1876. <https://doi.org/10.1002/qj.3047>, 2017.
- 1065 Valsala, V. and Maksyutov, S.: Interannual variability of the air–sea CO₂ flux in the north Indian Ocean, *Ocean Dynamics*, 63, 165–178, <https://doi.org/10.1007/s10236-012-0588-7>, 2013.
- van der Molen, M. K. and Dolman, A. J.: Regional carbon fluxes and the effect of topography on the variability of atmospheric CO₂, *Journal of Geophysical Research*, 112, D01104, <https://doi.org/10.1029/2006jd007649>, 2007.
- 1070 Vogel, B., Müller, R., Günther, G., Spang, R., Hanumanthu, S., Li, D., Riese, M., and Stiller, G. P.: Lagrangian simulations of the transport of young air masses to the top of the Asian monsoon anticyclone and into the tropical pipe, *Atmos. Chem. Phys.*, 19, 6007–6034, <https://doi.org/10.5194/acp-19-6007-2019>, 2019.
- 1075 Willetts, P.D., Marsham, J.H., Birch, C.E., Parker, D.J., Webster, S. and Petch, J.: Moist convection and its upscale effects in simulations of the Indian monsoon with explicit and parametrized convection. *Q.J.R. Meteorol. Soc.*, 143, 1073–1085, <https://doi.org/10.1002/qj.2991>, 2017.
- 1080 Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO₂ measurements with TCCON, *Atmospheric Measurement Techniques*, 10, 2209–2238, <https://doi.org/10.5194/amt-10-2209-2017>, 2017.
- 1085

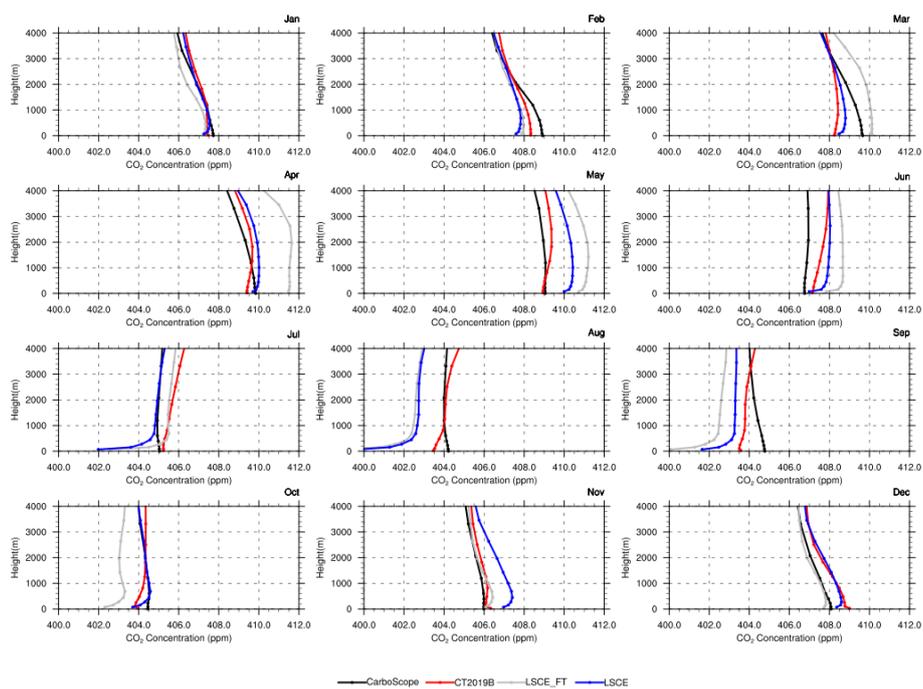


1090 **Figure 1: The WRF-GHG model domain used in this study, showing topography. The CO₂ monitoring sites over India used for the OSSE experiments are marked. Not all these observation stations are currently fully operational. The colour scale is restricted to 5000 m for the better visualization of terrain details over the Indian subcontinent.**



1095

Figure 2: Comparison of global models over the model domain during daytime (11:30 to 16:30 local time) in 2017. a) Annually averaged vertical profiles of CO₂ concentration in the lower troposphere b) Time series of monthly averaged CO₂ concentration at surface (~100 m above surface).



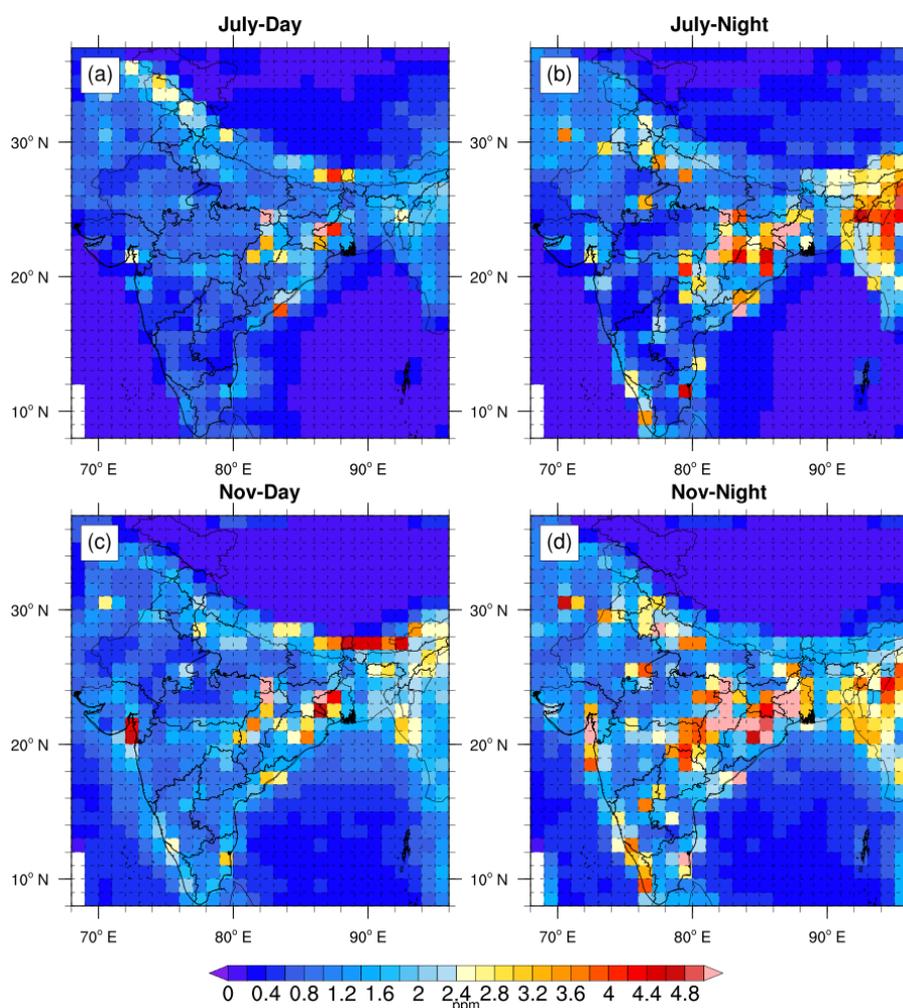
1100

Figure 3: Comparison of average monthly vertical profiles of CO₂ concentration from global atmospheric transport models over the model domain during daytime (11:30 to 16:30 local time) in 2017. Panels show data for respective months as indicated on the top of each panel.

1105



1110



1115

Figure 4: Monthly averaged values of representation error estimated for surface CO₂ concentration (second model level, mean height is ~200 m from sea level) over the region 8° N to 37° N and 68° E to 96° E during 2017. a) July daytime (11:30 to 16:30 local time) b) July nighttime (23:30 to 4:30 local time). c) November daytime. d) November nighttime.

1120



1125

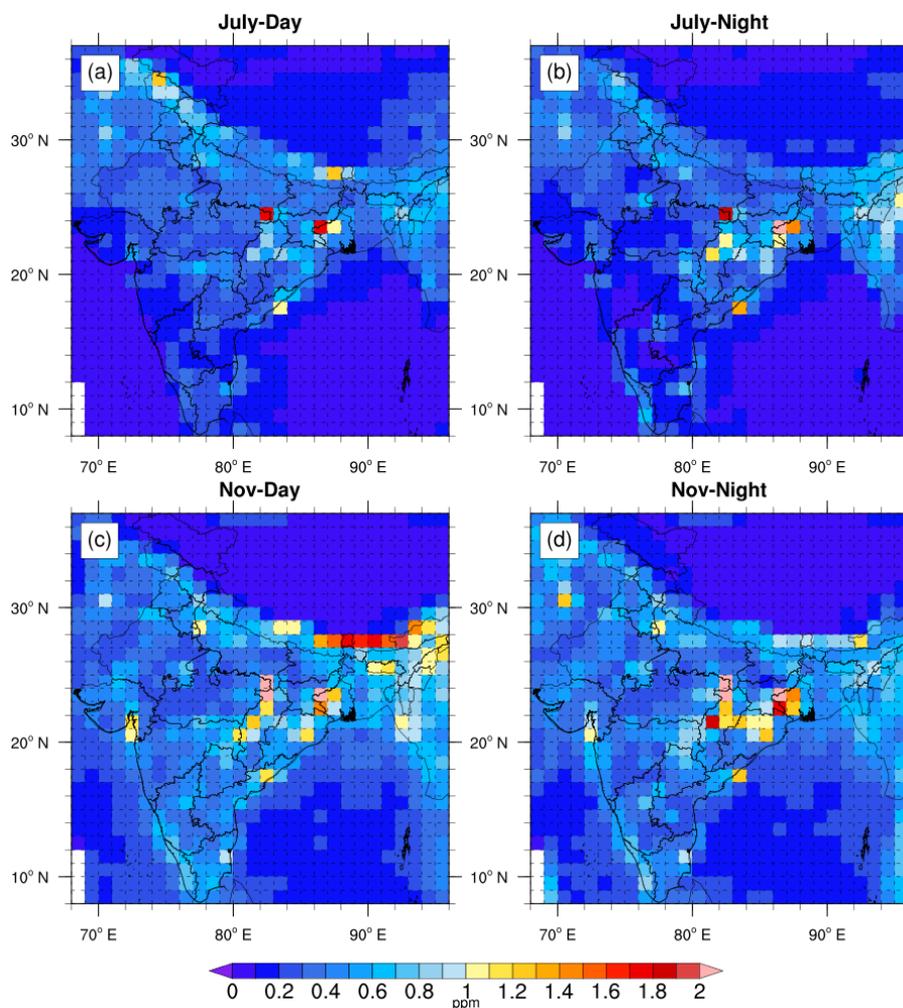


Figure 5: Monthly averaged values of representation error estimated for column averaged CO₂ concentration over the region 8° N to 37° N and 68° E to 96° E during 2017. a) July daytime (11:30 to 16:30 local time) b) July nighttime (23:30 to 4:30 local time). c) November daytime. d) November nighttime.

1130



1135

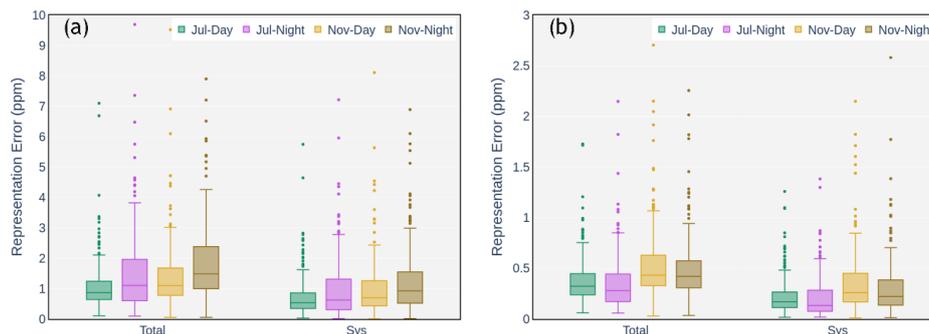
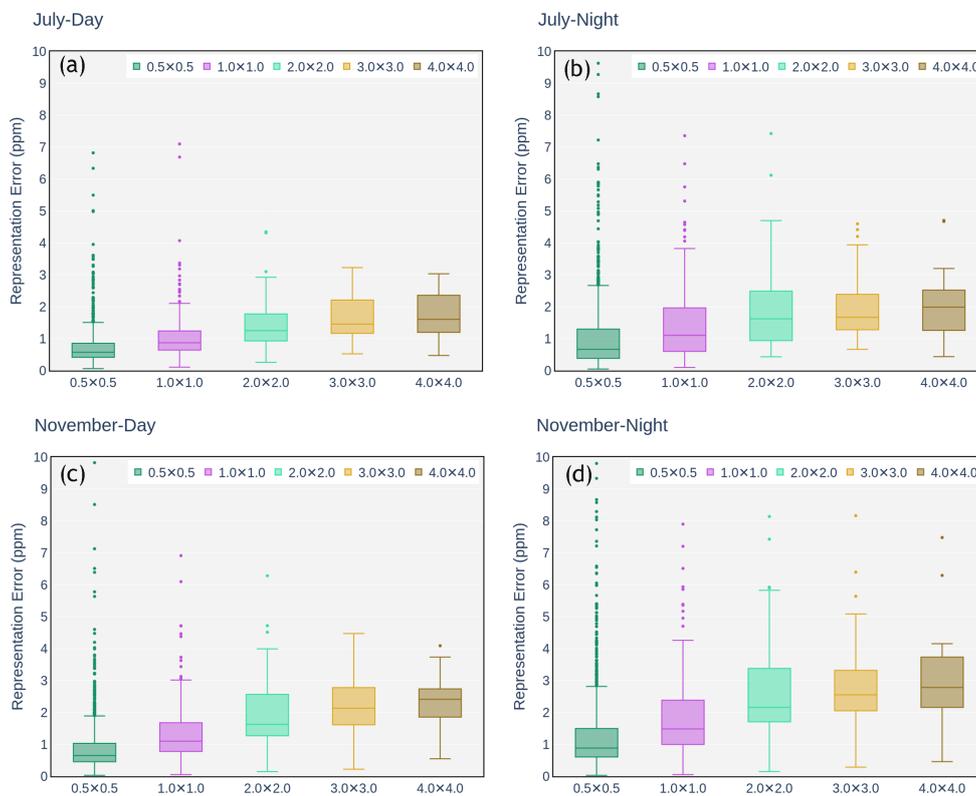


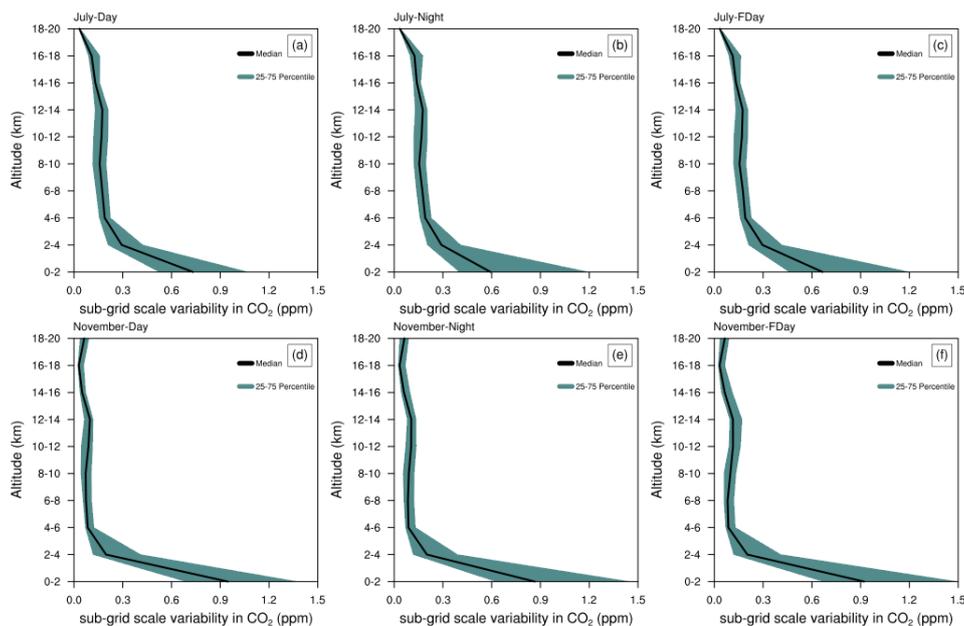
Figure 6: Variability of derived representation error over India in July and November 2017 (both during daytime and nighttime). Boxes indicate the central 50%, the bar across the box is the median value, and the whiskers indicate the values between 5 and 95 percentiles. Individual data points shown are the outliers. a) Representation error estimated for the surface CO₂. b) Representation error estimated for the column averaged CO₂.

1140



1145

Figure 7: Variability of derived surface representation error over India for different horizontal resolutions. Boxes indicate the central 50%, the bar across the box is median value, and the whiskers indicate the value between 5 and 95 percentiles. Individual data points shown are the outliers. a) Representation error estimated for July daytime. b) July nighttime. c) November daytime. d) November nighttime.



1150

Figure 8: Variability of representation error over India with altitude for July and November 2017. a) July daytime, b) July nighttime, c) July full time, d) November daytime, e) November nighttime, and f) November full time. Median values are plotted with black curves and the shaded region indicates 25 to 75 percentiles of data.

1155

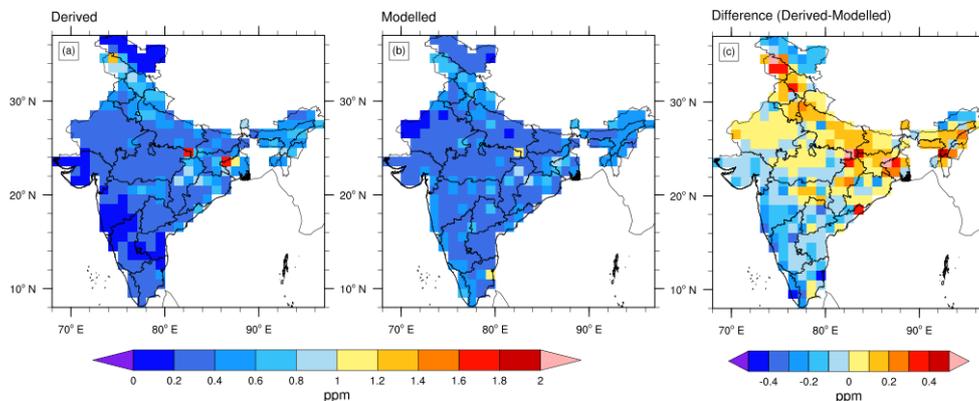


Figure 9: Monthly averaged values of representation error estimated for column averaged CO₂ concentration during July daytime (11:30 to 16:30 local time) in 2017. a) Representation error derived from WRF-GHG simulations as explained in Sect. 2.3. b) Representation error calculated from the multivariate linear model as described in Sect. 3.5. c) Difference between (a) and (b).

1160



Table 1: WRF-GHG Model setup

Domain						
Configuration	Single domain with horizontal resolution of 10 km; 39 vertical levels; 307 × 407 grid points					
Vertical coordinates	Terrain-following hydrostatic pressure vertical coordinates					
Basic equations	non-hydrostatic; compressible					
Grid type	Arakawa-C grid					
Time integration	3rd order Runge-Kutta split-explicit					
Spatial integration	3rd and 5th order differencing for vertical and horizontal advection respectively; both for momentum and scalars					
Timestep	60 s					
Physics schemes						
Radiation	Rapid Radiative Transfer Model (RRTM) for Longwave & Dudhia for shortwave					
Microphysics	WSM 3-classic simple ice scheme					
PBL	YSU					
Surface layer	Monin-Obukhov					
Land-surface	NOAH LSM					
Cumulus	Grell-Freitas ensemble scheme					
Emission fields						
Flux type	Product	Version	Spatial resolution	Temporal resolution	Source/website	Reference
Anthropogenic Biomass burning Biospheric	EDGAR	v4.3	10km	Annual	https://edgar.jrc.ec.europa.eu/ http://apps.ecmwf.int/datasets/data/cams-gfas/	Crippa et al., (2018) Kaiser et al., (2012) Mahadevan et al., (2008)
	GFAS	v1.2	10km	Daily		
	VPRM		Adapted to model	Adapted to model		
Initial and Lateral Boundary conditions						
Field	Product	Version	Spatial resolution	Temporal resolution	Source/website	Reference
Meteorology Tracer	ERA5	n/a	25km	1hour	https://cds.climate.copernicus.eu/cdsapp#!/ http://atmosphere.copernicus.eu	Hersbach et al., (2020) Agusti-Panareda et al., (2019)
	ECMWF /CAMS	gqi	50km	6hour		



Table 2: Specifications of different global model products used in this study

Data availability								
Product	Version	Spatial resolution	Vertical levels	Temporal resolution	Source/website	Reference		
Carbon Tracker	CT2019B	3 × 2	25	3 hours	http://carbontracker.noaa.gov	Jacobson et al., (2020)		
CarboScope	s10oc_v2020	5 × 3.8	19	6 hours	http://www.bgc-jena.mpg.de/CarboScope/	Rödenbeck et al., (2003)		
LSCE	v18r3	3.7 × 1.8	39	3 hours	http://atmosphere.copernicus.eu	Chevallier et al., (2019)		
LSCE	FT18r1	3.7 × 1.8	39	3 hours	http://atmosphere.copernicus.eu	Chevallier et al., (2019)		
Data used in the inverse model simulations								
Product	Version	Forward Model	Meteorology	Observation data	Anthropogenic emission fields	Biospheric emission	Fire emission	Oceanic emission
Carbon Tracker	CT2019B	TM5	ECMWF	Ground based	Miller and ODIAC	CASA	GFED and GFED CMS	OIF and Takahashi et al., (2009)
CarboScope	s10oc_v2020	TM3	NCEP	Ground based	EDGAR	LPJ Biosphere Model	CDIAC	SOCAT
LSCE/PyVar	v18r3	LMDz6A	ECMWF	Ground based	EDGAR, CDIAC and GCP	ORCHID EE 4.6.9.5	GFED and GFAS	Denvil-Sommer et al., (2019) with updates described in Friedling et al., (2019)
LSCE/PyVar	FT18r1	LMDz6A	ECMWF	Satellite (OCO-2 NASA)	EDGAR, CDIAC and GCP	ORCHID EE 1.9.5.2	GFED and GFAS	Landschutzer et al., (2018)



1175

Table 3: Flux uncertainty over India calculated from the OSSE experiments using pseudo-observation network of surface observations. The time filter indicates the time of the data sampled for estimation of the scaling factors. Full day – 24 hours in each day; Daytime – 11:30 to 16:30 local time; Nighttime – 23:30 to 4:30 local time. * The fraction of uncertainty to the total NEE.

Month	Time filter	True flux, aggregated over India. $\sum_{k=1}^K \Phi_{true}$ (MtCO ₂ per month)	Flux uncertainty S_{rep} (MtCO ₂ per month) In brackets: fraction of uncertainty* (%)
July	Daytime observations	-373.31	38.59 (10.33)
July	Nighttime observations		30.14 (8.07)
July	Full day observations		23.20 (6.21)
November	Daytime observations	-417.12	18.42 (4.4)
November	Nighttime observations		13.34 (3.1)
November	Full Day observations		13.48 (3.2)

1180

1185