

Uncertainties in OCO-2 satellite retrievals of XCO² limit diagnosis of transport model simulation uncertainty

4 Chiranjit Das¹, Ravi Kumar Kunchala¹, Prabir K. Patra^{2,3}, Naveen Chandra², Kentaro Ishijima⁴, 5 Toshinobu Machida⁵

¹ Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India

² Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama 236-0001, Japan

³ Research Institute for Humanity and Nature, Kyoto 6038047, Japan

⁴Meteorological Research Institute, Tsukuba, Japan

⁵National Institute for Environmental Studies, Tsukuba, Japan

Correspondence to: Prabir K. Patra (prabir@jamstec.go.jp) and Chiranjit Das (Chiranjit.Das@cas.iitd.ac.in)

13 **Abstract.** Estimating regional CO₂ sources and sinks is challenging due to limited data and uncertainties in 14 transport models. Orbiting Carbon Observatory-2 (OCO-2) overcomes measurement limits, providing CO2 15 variations beyond in-situ networks. This study analyses altitude-wise model-observation CO₂ differences from surface to upper troposphere using aircraft observations from ATom, Amazon, and CONTRAIL campaigns over OCO-2 total column CO² (XCO2) sampling location to characterise sources of uncertainty in MIROC4-ACTM. 18 We show model aligns better with ATom tropospheric columns $(0.03 \pm 0.03 \text{ ppm})$ than OCO-2 XCO₂ (0.2 \pm 0.5 ppm), especially over oceans, highlighting the need for expanded profile measurements to characterise errors robustly. Altitude-wise comparisons reveal this differences primarily occur in the lower troposphere (0-2 km), 21 likely due to ACTM's near-surface land $CO₂$ flux errors. In contrast, ACTM better matches aircraft $CO₂$ in the 22 middle (2-5 km) and upper (5-8 km) troposphere, likely due to accurate large-scale transport representation. Over 23 the Amazon, CO₂ differences with aircraft and OCO-2 differ, likely due to a lack of regional surface sites for 24 inversion and insufficient high-altitude profile (~4km) not representative of XCO2. Over Asian megacity airports, which are significant emission hotspots, the model shows a large negative difference with CONTRAIL than OCO-26 2. This discrepancy likely hints that MIROC4-ACTM is unable to capture urban fossil CO₂ emission signals at airports due to coarse resolution (~2.8° x 2.8°) and higher resolution of OCO-2 limits ability to fully capture actual emission footprints.

Keywords: Carbon dioxide; Aircraft; Transport Model; OCO-2

-
-
-
-
-
-

1. Introduction

 Atmospheric $CO₂$ is the most significant anthropogenic greenhouse gas (GHG) present in the Earth's atmosphere, responsible for a major global warming and climate change since the preindustrial era, circa 1750 (Canadell et al., 2022). Therefore, recognizing its importance in direct impact on climate, monitoring of highly accurate surface CO² measurements were first started from the South Pole (SPO) and Mauna Loa (MLO), Hawaii (Keeling, 1960) 47 and later expanded across the globe. These in-situ measurements are widely used for estimating surface CO₂ fluxes using Bayesian-based "top-down" chemistry transport models due to their long-term record and high 49 measurement accuracy (Chandra et al., 2022; Chevallier et al., 2010; Peylin et al., 2013). However, in-situ CO₂ measurement sites around globe is sparse, mostly situated in mid-latitude north America and Europe, with less coverage over tropical land (Patra et al., 2011; Schimel et al., 2015) and open oceans, which increase difficulties in inferring surface CO² fluxes from inverse models in data void regions (Chevallier et al., 2010, 2011).

 To increase spatiotemporal monitoring of CO2, spaced-based measurements such as SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY (SCIAMACHY), Greenhouse Gases Observing Satellite "IBUKI" (GOSAT), and the Orbiting Carbon Observatory-2 (OCO-2) were launched to provide column-57 average dry-air mole fraction or mixing ratio of CO₂, termed XCO₂ (Bovensmann et al., 1999; Crisp, 2008; Kuze et al., 2009). NASA's OCO-2 satellite launched in 2014 to achieve finer spatial resolution and better precision as compared to previous satellites. This advancement has proved beneficial for understanding global and regional carbon cycle science in various satellite based studies (Crisp, 2015; Das et al., 2023; Liang et al., 2017; Liu et al., 2017; Chatterjee et al., 2027). OCO-2 version 10 XCO2 measurements has shown retrieval error with mean bias (RMSE) of 0.24 (0.81) ppm over land and 0.43 (0.84) ppm over ocean globally, compared against more accurate 63 WMO scale maintained XCO₂ from surface-based Total Carbon Column Observation Network (TCCON) sites 64 (Taylor et al., 2023; Wunch et al., 2017). Studies reported that assimilation of OCO-2 XCO₂ available at greater 65 spatial density $($ \sim 100 times of GOSAT) into an inversion requires the data to be extremely precise, stable and 66 regionally unbiased to effectively estimate surface $CO₂$ fluxes in regional scale (Byrne et al., 2017, 2023; Crowell et al., 2019; Philip et al., 2022; Rastogi et al., 2021). Also, Miller et al. (2007) reported that satellite-retrieved XCO² needs regional precision of 1-2 ppm to reduce uncertainty in inversion-derived flux estimates from in-situ 69 networks. Because, XCO_2 retrievals having many sources of uncertainty hinder their fidelity to utilize inversion 70 approach to accurately estimate surface CO₂ flux (Chevallier et al., 2014; Villalobos et al., 2020). These retrieval errors in OCO-2 include cloud effects (Massie et al., 2021;Merrelli et al., 2015), instrumental errors, retrieved surface pressure, and then aerosol, the largest source of systematic error can be approximately 2 ppm over land 73 regions (Connor et al., 2016). Therefore, to enhance the accuracy of surface $CO₂$ flux estimations, studies are focussing on improving retrieval algorithm by correcting for cloud effects and incorporating a digital elevation model (DEM) to correct surface pressure (Jacobs et al., 2023; Mauceri et al., 2023). Apart from retrieval errors, misrepresentation of transport and uncertainty in prior fluxes can further reduce reliability in top-down model inferred surface CO² fluxes (Chandra et al., 2022; Fu et al., 2021; Schuh et al., 2019). To address and assess such kind of errors impact on top-down CO² budgets, OCO-2 model intercomparison project (MIP) is formed with different CO² inverse modelling groups assimilating OCO-2, in-situ and combination of both (80 https://gml.noaa.gov/ccgg/OCO2_v10mip/index.php).

81 Previous studies have attempted to reconstruct observation based $CO₂$ profiles combining ship, aircraft and model 82 simulation to compare with GOSAT XCO₂, but are limited to characterise inversion errors (Müller et al., 2021; 83 Inoue et al., 2013; Wofsy, 2011). Frankenberg et al. (2016), showed using HIPPO aircraft CO₂ vertical profiles, 84 that the retrievals of GOSAT, TES, AIRS satellites and inversion simulation can have large difference of \sim 4 ppm due to inaccurate vertical transport in higher latitude during vegetation growing or decaying periods. These studies 86 often lack in providing uncertainties linked with inversion through altitude-based $CO₂$ comparison from near surface to different tropospheric layers between inversion-based model simulations, surface, and aircraft data. This gap is particularly evident in regions with in-situ sparse data coverage, such as vast oceanic areas, as well as 89 in emission or sink hotspots over land while OCO-2 dense measurements have not helped to overcome precision issues providing global coverage. However, no studies attempted to understand how OCO-2 retrieval errors or accuracy hampers its full potential to uncover the uncertainties associated with inverse models. The present study 92 aims to understand sources of error associated with the MIROC4-ACTM model through altitude-based CO₂ comparison among MIROC4-ACTM, OCO-2 and aircraft observation across different tropospheric layers. To 94 accomplish this we leverage highly accurate and precise aircraft vertical CO₂ measurements from ATom campaigns over northern America, Pacific, Atlantic, Southern Ocean regions, CONTRAIL over airports in Asia and four specific sites in Amazon. Before analysing, we first validated the MIROC4-ACTM model simulated 97 tropospheric column CO_2 (XCO₂) with highly precise CO_2 vertical profiles from independent aircraft measurements over open oceanic regions, Amazon and local urban hotspot over Asia.

2. Data and Methodology

 102 **2.1 Aircraft and surface CO² measurements**

 We have used individual independent aircraft and surface in-situ $CO₂$ measurements around the globes from 105 NOAA's obspack co2_1_GLOBALVIEWplus_v8.0 data product (Schuldt et al., 2022) and WDCGG (World Data Centre for Greenhouse Gases) respectively. We have selected a few individual campaigns such as ATom, Amazon aircraft campaigns, CONTRAIL because of their extensive latitudinal/longitudinal spatial coverage over ocean and land regions, multiple vertical CO² profile measurements and extended period of measurements. A brief description of each aircraft measurement is discussed in the next paragraphs.

 ATom is an aircraft field campaign, providing airborne measurements of remote tropospheric and lower stratospheric CO² from Northern America, Arctic, Pacific, Southern and Atlantic Oceans (Thompson et al., 2022). ATom has four campaigns conducted using NASA DC-8 aircraft, taking vertical profile of CO² from near surface (0.15 km) to13 km altitude range over four seasons from 2016 to 2018. Four campaigns each lasted around 28 days, namely, ATom–1, 28 July-22 August 2016; ATom–2, 26 January–22 February 2017; ATom–3, 28 September–26 October 2017; ATom–4, 24 April to 21 May 2018 respectively (Wofsy et al., 2021). These vertical 117 CO₂ measurements enable the validation of XCO₂ measurements from the satellites because most of XCO₂ variability is constrained in the troposphere, therefore, these vertical measurements effectively serve as a reference 119 for satellite-retrieved CO₂ validation (Frankenberg et al., 2016). We have utilised these CO₂ measurements freely 120 available at https://gml.noaa.gov/ccgg/obspack/data.php.

122 Additionally, aircraft vertical campaign $CO₂$ measurements from four sites such as TEF (3.39° S, 65.6° W), SAN 123 (2.86° S, 54.95° W), RBA (9.38° S, 67.62° W) and ALF (8.80° S, 56.75° W) in the Amazon region are also 124 considered for the present study and more details about these measurements can be found in Gatti et al. (2021a). 125 In ALF, RBA, SAN, $CO₂$ measurements are available from 2010 to 2018, whereas TEF has measurements from 126 2013-2018. Bi-monthly vertical CO² profile measurements were taken during 12:00 and 13:00 local time at all 127 these sites covering altitude from 0.3 to 4 km when the daytime boundary layer is well mixed. These measurements 128 were taken using an automatic sampler onboarded in the light aircraft, which underwent accuracy and precision 129 testing at greenhouse gas laboratory at National Institute of Space Research (LaGEE/INPE), Brazil (Gatti et al. 130 2014). We have used a set of vertical CO₂ profiles during September, 2014 till 2018, freely available at: 131 https://doi.pangaea.de/10.1594/PANGAEA.926834 (Gatti et al., 2021b).

132

 Further, CONTRAIL aircraft program CO² measurements over Asian regions are also considered for the current analysis (Ishijima et al., 2021; Machida et al., 2008; Matsueda et al., 2008). In this program, several regular 135 passenger aircraft operated by Japan Airlines (JAL) are installed with instruments like CME (Continuous CO₂ 136 Measuring Equipment) to provide extensive spatial CO₂ data coverage in the upper troposphere and lower stratosphere (UT/LS) region. We have only considered measurements during OCO-2 measurements period at four 138 representative zones in Asia, specifically around airports, to retrieve vertical $CO₂$ profiles during ascent or descent of the aircraft, following Niwa et al., 2011. The measurements dataset is freely accessible at https://www.cger.nies.go.jp/contrail/.

141

142 **2.2 OCO-2**

143

144 OCO-2 is a sun-synchronous satellite, retrieves XCO₂ to understand the carbon source-sink activity throughout 145 the globe (Eldering et al., 2017). The satellite uses three high-resolution grating spectrometers to retrieve the 146 reflected sunlight spectral signature of weak CO_2 (1.61 µm), strong CO_2 (2.06 µm) and O_2 A (0.76 µm) which are 147 later analysed through Atmospheric Carbon Observations from Space (ACOS) algorithm to estimate global 148 spatiotemporal XCO² distribution (Crisp, 2015; Crisp et al., 2017; Eldering et al., 2017). It has a spatial resolution 149 of 1.29 km \times 2.25 km (nadir mode), and a temporal periodicity of sixteen days. We have utilised OCO-2 version 150 10 which is an update from previous version v8/v9 in terms of important changes in spectroscopy, aerosol, CO² 151 prior source, and solar continuum model which reduced RMSE validated against XCO₂ measured at TCCON sites 152 for land and ocean-glint measurement (Taylor et al., 2023). In this study, we have used level–2 OCO-2 version 153 10r data product (available at: 154 https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Standard_10r/summary?keywords=OCO-2) and have considered 155 only bias-corrected quality checked soundings (XCO₂ quality flag = 0 or good data) for analysis.

156

157 **2.3 Transport model**

158

159 The Model for Interdisciplinary Research on Climate, version 4, based Atmospheric Chemistry Transport Model 160 (MIROC4–ACTM) chemistry transport model is used, which is run at T42 spectral spatial resolution (∼2.8° × 161 2.8° latitude-longitude grid) with 67 vertical hybrid-pressure levels from the surface to 90 km to simulate the $CO₂$

 concentration and fluxes (Patra et al., 2018). For realistic representation of the transport in the model, model meteorology, horizontal winds (zonal and meridional) and temperature is nudged to Japanese 55-year Reanalysis or JRA-55 data (Kobayashi et al., 2015). The MIROC4-ACTM model conducted two distinct simulations: one utilizing bottom-up model a prior or "FG" fluxes and the other employing a posteriori or "InvFG" fluxes derived from top-down inversion using 50 inversion sites (Figure S1). To derive total concentration in the simulation 167 different bottom-up tracers are utilised, gridded GridFED fossil fuel; $CO_{2\text{ ff}}$ (Jones et al., 2021), annually balanced 168 CASA biospheric flux; $CO_{2,ind}$ (Randerson et al., 1997) and ocean exchange i.e., sea-air $CO₂$ flux; $CO_{2,ocn}$ (Takahashi et al., 2009). Then, prior CO² simulation case of MIROC4-ACTM is prepared by adding the prior 170 tracers as follows: $CO_{2,ff}$ (GridFED) + $CO_{2,hd}$ (CASA–3hr) + $CO_{2,con}$ (Taka–Ocn). A detailed discussion on this is given in Chandra et al. 2022. Further, to minimise the edge effect on the simulated dataset, we discarded the 172 first two years and last one year of our simulation period (2012-2022), only analysed $CO₂$ of both simulation and observation during 2014-2021. Model performance evaluated by comparing with each vertical profile of 174 independent observations of ATom CO₂ not used in the inversion as well as at two surface sites, MLO (19.53 °N, 175 155.57 °W) and SYO (69.01 °S, 35.59 °W) representative of northern and southern hemisphere CO₂ variability 176 (Fig. S2, 3). To do that firstly, model simulated $CO₂$ is resampled to the nearest grid of the aircraft and surface sampling locations, considering linear interpolations at spatial grid and time. We have not considered any other co-location criteria unless it is mentioned, e.g., geometric and dynamic for comparison; therefore, estimated CO² difference is essentially uncertainty either in observation or inversion (Kulawik et al., 2016, 2019). The result shows good performance with lesser bias with InvFG over prior at different latitudes, showing an overall good match of InvFG CO² and ATom at different latitudes (Fig. S2, 3). Similarly at surface sites, InvFG showed better performance over prior with correlation of 0.99 with observation (p < 0.05) (Fig. S4). Then, to compare 183 OCO-2 XCO₂ and ACTM-XCO₂, we formulated XCO₂^{ACTM} following the Patra et al., 2017. Here, ACTM 184 simulated CO_2 profile or CO_2^{ACTM} resampled at each OCO-2 retrieval location (latitude, longitude) with further use of corresponding OCO-2 priori and column average kernel sensitivity (Ai) represents instrumental sensitivity 186 for 20 vertical levels from top of atmosphere (TOA) to surface to produce XCO_2^{ACTM} using the following equation.

188 $XCO_2{}^{ACTM} = \sum_i (CO_2{}^{priori} \cdot dP_i) + \sum_i A_i \cdot dP_i \left(\sum_i CO_2{}^{ACTM} - \sum_i CO_2{}^{priori} \right)$ (1)

189 $CO_2^{priori} = OCO-2$ priori; $A_i = OCO-2$ column averaging kernel; $dP_i =$ thickness of each pressure layer.

191 We have also resampled XCO_2^{ACTM} at 21 TCCON sites around the globe to validate model performance with 192 more accurate XCO₂ measurements at surface-based TCCON sites (Wunch et al., 2011). To perform that we first filtered data points considering only good quality retrieval, then curve fitted the remaining retrieval to remove outliers and finally considered retrievals with solar zenith angle < 60° following methodology mentioned in Appendix C of Crowell et al. 2019. Figure S5 shows a good agreement between ACTM and TCCON XCO² for majority of TCCON sites considering the fact that TCCON has its own bias due to topography, surface brightness and aerosols as well as latitudinal varying bias (Wunch et al., 2017).

2.4 Data analysis

201 To conduct the spatial discrepancy analysis of XCO₂ difference between ACTM and OCO-2, bias corrected and 202 quality checked good soundings from OCO-2 XCO₂ retrieval or XCO_2 ^{OCO-2} and ACTM simulated XCO₂ or 203 XCO_2^{ACTM} is re-gridded into $0.5^{\circ} \times 0.5^{\circ}$ latitude-longitude grid boxes. Further, to assess this XCO_2^{ACTM} -204 $XCO_2^{\rm OCO-2}$ difference, we used aircraft vertical CO_2 measurement available at different vertical tropospheric 205 layers, essentially to conduct an altitude-wise comparison of CO₂ among ACTM, OCO-2 and aircraft. Therefore, 206 we have employed a methodology outlined using a schematic in Figure 1, specifically designed to provide $CO₂$ 207 profile from aircraft, OCO-2 and ACTM.

208

209 It shows typical CO₂ concentration vertical profiles and relative altitude range captured in OCO-2, aircraft 210 measurements, and corresponding MIROC4-ACTM simulations. OCO-2 measures XCO₂ concentration from 211 space, representing $CO₂$ profile from top of the atmosphere to surface so as the ACTM simulation at those pressure 212 levels is represented by a blue colour double-headed arrow. On the other hand, aircraft tropospheric columns of 213 CO² typically capture concentration variability up to an altitude of 15 km and ACTM resampled concentration 214 values at those measurement locations represented by orange colour double headed arrows. Since the main 215 purpose is to compare the aircraft tropospheric column $CO₂$ against $OCO₂$ XCO₂ and ACTM simulations, we 216 subdivided the tropospheric CO₂ column into three different vertical tropospheric layers, namely, the lower 217 troposphere: lowest level to 2 km, middle troposphere: 2–5 km, and upper troposphere: 5–8 km to understand the 218 model performance in each of the vertical layers. In addition, a total tropospheric vertical $CO₂$ column of aircraft 219 or aircraft $XCO₂$ is calculated only when the vertical measurements reach at least 8 km altitude; otherwise, any 220 profile not reaching 8 km is discarded from the analysis unless otherwise mentioned specifically. For each vertical 221 tropospheric layer, pressure-weighted partial column $CO₂$ is calculated to consider air mass variation between 222 pressure levels for each considered tropospheric layer. Moreover, to become more robust on the analysis, each 223 vertical depth layer is subdivided into 200 bins unless otherwise mentioned and 80% of vertical bins having 224 measurements for the specific tropospheric layer only considered for analysis. For instance, middle troposphere 225 (upper troposphere), i.e., 2–5 (5–8) km is divided into 15 vertical bins each of 200 meters, then in a specific 226 latitude or longitude while calculating a partial column of $CO₂$, we only considered profiles that encompass a 227 minimum 12 vertical bins in them.

228
229

232

3. Results 230

231 **3.1. MIROC4-ACTM intercomparison with OCO-2 and surface measurements**

233 Figure 2a shows a monthly mean spatial XCO_2 difference $(XCO_2^{\text{ACTM}} - XCO_2^{\text{OCO-2}})$ during January, 2015 – 234 December, 2021. It reveals a heterogeneous signature of spatial XCO₂ difference or mismatch across the globe 235 with maximum negative XCO₂ difference of approximately 2 ppm over Amazon, Africa, south-east Asia, China, 236 primarily in the global tropical land regions. This is likely due to the lack of long-term surface $CO₂$ measurements 237 (Figure S1) available for inversion particularly over the global tropics to constrain the prior $CO₂$ flux (also 238 discussed in Chandra et al., 2022). Further, humid tropics is also the region of lesser valid OCO-2 retrievals due 239 to the persistent shallow cumulus cloud blocking the infrared signals, makes it challenging to validate the transport 240 model and studies has shown sparse sampling over land increases chances of the error almost two times 241 (Frankenberg et al., 2024; Kulawik et al., 2019). On the other hand, high-latitude land regions, North America

242 and Russia, exhibit a positive $XCO₂$ difference of nearly 2 ppm, however, $XCO₂$ differences over ocean regions 243 are generally within ± 1 ppm, possibly because of lesser variability in ocean CO₂ flux compared to land CO₂ flux 244 and oceanic regions are minimally affected by land air mass. Further, we observed a negative XCO₂ difference 245 especially over northern extra-tropics between 30° N to 60° N of nearly -0.6 ppm in agreement with Byrne et al. 246 2023 likely attributed to OCO-2 ocean glint (OG) retrieval biases that adds up a layer of complexity of diagnosing 247 the transport model against OCO-2 retrievals over this vast ocean. Further, studies showed that sampling 248 variance between land and ocean could also lead significant uncertainty (Basu et al., 2018). Additionally, we 249 observed a bias in the Southern Hemisphere, the underlying causes of which are still unknown and need further 250 research (Byrne et al., 2023). Overall, our results show some regions under/over-estimated by ACTM, however, 251 it is challenging to comprehend quantitatively and qualitatively about sources of error across diverse regions 252 (source-sink dynamics and transport mechanism) as it could result either due to inaccuracies in the inversion (prior 253 fluxes, transport) or errors in satellite retrievals (Chandra et al., 2022; Chevallier et al., 2014). Furthermore, to 254 check the time variation of these $XCO₂$ differences, we analysed the time versus latitude distribution of $XCO₂$ 255 difference taking the average across global longitude from 180° W to 180° E (Fig. 2b). We observed that XCO₂ 256 difference has a seasonal and spatially varying repeating signature, with maximum (minimum) difference during 257 February-March-April (September-October-November) consistent across study period. A prominent positive 258 (negative) systematic XCO₂ difference is observed over the southern hemisphere tropic to mid-latitude from 10° 259 S–40° S and northern latitude around 30° N (northern tropic to mid latitude) this is in agreement with Kulawik et 260 al. (2019). However, after separating time vs latitude distribution of $XCO₂$ difference into land and ocean, we 261 observed that this systematic XCO₂ difference mainly originates from the southern ocean part, which matches 262 well with overall (land and ocean) time vs latitude XCO₂ difference distribution (Fig. S6). However, it is 263 challenging to explain the difference since ocean glint has biases (Byr ne et al., 2023) and a study by Kulawik 264 et al. (2019) also reported a systematic error of 0.6 ± 0.1 ppm could arise over the ocean and land in OCO-2 265 satellite $XCO₂$ retrievals.

266

267 To further examine whether this difference comes from inversion or $XCO₂$ retrieval because both have 268 uncertainties. We similarly analysed time versus latitude distribution of surface $CO₂$ concentration difference with **269** respect to ACTM simulation or CO_2 ^{ACTM} - CO_2 ^{In-situ} considering accurate surface CO_2 concentration data from 53 270 measuring sites around the globe. Most of these sites are situated in the northern hemisphere having at least 90% 271 data during study period (all sites geographical location can be visualised in Figure S7 in supporting information). 272 MIROC4-ACTM simulated CO₂ near the surface is resampled to the nearest grid of surface sites (latitude, 273 longitude, altitude) and measurement time from hourly interval model output. For each of the 53 sites, CO₂ 274 concentration difference between ACTM and surface $CO₂$ concentration is calculated first, and then we linearly 275 interpolate it spatially as presented in Figure 2(c). Results show no such annually and spatially systematic 276 signature of CO₂ difference near the surface between equator to 45° S considering six stations (SEY: 4.7° S, ASC: 277 8° S, SMO: 14.2° S, NMB: 23.6° S, CPT: 34.4° S, CGO: 40.7° S) situated at different latitudes within this latitude 278 band (Fig. S7). Considering the fact that inferred CO₂ difference may arise due to much lower data density for the 279 in situ measurements within mentioned latitude bands, the analysis with the available sites suggests that systematic 280 signatures of difference exist when we compared with OCO-2 XCO₂ (Fig. 2b). It hints at uncertainties in OCO-2 281 retrieval or systematic vertical transport error in the model (Schuh et al., 2019), given the relatively lower

282 uncertainty in in-situ $CO₂$ measurements compared to OCO-2 XCO₂. This vast part of the region remained 283 challenging for a model to understand its error characteristics due to OCO-2 retrieval error. Further, we also 284 compared the latitudinal average time series of $CO₂ (XCO₂)$ difference of ACTM with surface (OCO-2) in Fig. 285 2(d). It shows an overall agreement of $CO₂$ difference with a correlation coefficient of 0.68 at 99% significance 286 level with total time series variability (1-σ STDEV) of 0.28 and 0.19 ppm in CO² difference in surface as compared 287 to OCO-2. Overall, this $CO₂$ space-time variability analysis clearly demonstrates that systematic signature in 288 XCO₂ difference, primarily concentrated in southern mid and northern high latitudes, previous studies also 289 indicated towards potential uncertainties may arise in retrieval over the ocean and or misrepresentation of vertical 290 transport in inversion (Byrne et al., 2023; Frankenberg et al., 2016; Schuh et al., 2019). Since systematic errors in 291 transport could result in inaccurate $CO₂$ flux estimates and, consequently, posterior simulated concentration (Deng 292 et al., 2015; Stephens et al., 2007). In an inversion estimation, transport, surface $CO₂$ flux and $CO₂$ spatial gradient 293 are closely interconnected so any misrepresentation in vertical or horizontal mixing consequently affects the 294 estimated flux. This highlights the complexity of interpreting $CO₂$ differences across diverse regions, 295 measurement platforms and error quantification of the optimised flux of inversions from surface and satellite 296 measurements. Therefore, to better understand the consistency of CO₂ differences across different global regions 297 and identify regions of major uncertainty which will enable us to address them effectively, we analysed $CO₂$ 298 variation in different vertical tropospheric layers using vertical $CO₂$ profile datasets from aircraft measurements, 299 discussed in section 3.2.

300

303

301 **3.2 CO² difference in tropospheric layers**

302 **3.2.1 Over Globe**

 Figure 3 represents mean CO_2 difference or CO_2 ^{ACTM}-CO₂ aircraft across different latitudes using individual aircraft observations for different tropospheric layers LT (light red), MT (orange), UT (dodger blue) and tropospheric Total Column (teal), color coded to represent different altitude ranges. Aircraft measurements are generally available in two modes: continuous measurements from the same site over a long period, and campaign measurements that cover extensive vertical and horizontal distances with high data density over a limited period. Therefore, we have subdivided aircraft measurements into two subcategories for our analysis: specific site aircrafts having latitude coverage maximum 5°(Fig. 3a) and campaign aircrafts having latitude coverage 311 maximum 30° (Fig. 3b). Only those aircrafts having measurements during OCO-2 period are selected, each aircraft sampling location, number of data points at different latitude bins of 30° and altitude bins of 1000 meters is provided in the supplementary material (Fig. S8 and S9 gif for each aircraft category). Then, we calculated 314 model-observation CO₂ difference for each aircraft measurement category. Therefore, estimated CO₂ difference serves as a model and observation mismatch for specific latitude (entire latitude range) for specific sites (campaign) aircraft with latitude information mentioned inside parenthesis of first x-axis tick marks in Fig. 3a (3b). Here, the second x-axis shows the number of data points in the corresponding aircraft campaign. Here, the 318 number of data points or samples is critical when comparing CO₂ differences among aircraft. A higher number of 319 samples provides better confidence to the calculated CO₂ difference while aircraft with fewer samples are considered less weightage.

321 Mean (variability) of $CO₂$ difference for LT, MT, UT and tropospheric column for specific sites aircraft are -0.45 322 (± 0.49) , -0.32 (± 0.48) , -0.34 (± 0.5) , and -0.2 (± 0.41) respectively (Fig. 3a). It shows the highest mismatch in 323 terms of mean exist in LT as compared to other tropospheric layers MT, UT and total tropospheric column, likely 324 due to uncertainty in prior $CO₂$ flux or transport in LT. Studies have shown that in the LT, concentration changes 325 are mainly regulated by surface CO₂ fluxes and diurnal-synoptic mixing patterns (Law et al., 2008; Patra et al., 326 2008). However, $CO₂$ change in UT is mainly dominated by changes in large scale dynamical transport, where 327 surface emission has subdued influence. Hence, studies have found that coarse spatial resolution transport models 328 adequately simulated CO₂ in the MT to UT regions (Baier et al., 2020; Niwa et al., 2011). Additionally, it is also 329 noted that there is a systematic underestimation by the model in terms of magnitude in all tropospheric layers. 330 Result shows minimum or maximum (near zero) model-observation CO₂ difference (ppm) observed for "esp" or 331 "rta" (cma) aircraft at 49.48° N or 21.19° S (38.83° N) in the LT for specific sites aircraft. In the tropospheric 332 column, maximum (minimum) CO² difference is observed in "cma" ("car") aircraft at 38.83° N (40.66° N) 333 respectively. Further, it has been observed that the tropospheric column $CO₂$ difference matches pattern of $CO₂$ 334 difference at LT at most aircraft sites; LT apparently contributes more to the total tropospheric column than MT 335 and UT. Further, it could be seen that the overall mean model-observation $CO₂$ difference is highest in the northern 336 mid-latitudes compared to the tropical latitudes of the northern and southern hemispheres.

337

338 Further, we have calculated XCO_2^{ACTM} - $XCO_2^{\text{OCO-2}}$ at those specific sites aircraft location considering $5^\circ \times 5^\circ$ grid 339 box surrounding it to check the difference similarity between $XCO₂$ and tropospheric column $CO₂$ selecting only 340 specific times from OCO-2 of aircraft measurement. Results show XCO₂ difference mean (variability) of - 0.37 (+0.38) , highlighting that the model is underestimating and also has minimum variability as compared to any individual layers. On the other hand, $CO₂$ difference with campaign aircraft showed similar results; overall highest mean and variability exists in LT observed. Further, the overall mean (variability) of the tropospheric column is - 344 0.39 (\pm 0.1) controlled largely by CO₂ difference in LT. Here, we also check XCO₂ difference at those campaign aircraft considering their covered tracks and then taking the average of all XCO₂ differences to calculate a mean 346 XCO₂ difference. Results show lesser variability in XCO₂ as compared to other layers and also negative mean XCO² represents overall underestimation by the model. It has been observed in both aircraft categories that the 348 model has underestimated overall CO₂ concentration in all tropospheric layers and total columns, also the maximum mean and variability of CO² difference are in LT. These differences are attributed possibly due to underestimation by prior flux in the inversion or misrepresentation of transport in the model. To further investigate CO² differences altitude-wise, we considered individual vertical CO² profiles from different campaigns for different regions North America, Pacific, Southern Ocean, Atlantic, Amazon and Asia discussed in detail in subsequent sections.

354

355 **3.2.2 North America, Pacific, Southern and Atlantic ocean**

356
357 Figure 4(a) illustrates integrated tracks traversed by the aircraft during ATom campaign (ATom-1, ATom-2, ATom-3, ATom-4) across oceanic and land parts, subdivided into four segmented track categories corresponding to specific geographical regions delineated with different colours. Represented segments are North America and neighbours; east to west aircraft campaigning (magenta), Pacific; north to south aircraft campaigning (yellow), Southern Ocean; west to east aircraft campaigning (red), and Atlantic; south to north aircraft campaigning (green).

362 Figure 4(b) shows the mean XCO₂ difference taken considering a collocation criteria of $5^\circ \times 5^\circ$ latitude-longitude 363 grid box around sampling location and $XCO₂$ retrievals during corresponding ATom campaign period. We 364 observed a maximum XCO₂ difference of nearly 2 ppm over 120° W and 90° W in North America and neighbouring land regions whereas oceanic regions, particularly Pacific and Atlantic are mostly confined within ±1 ppm at any specific latitude. We also checked the latitudinal bias MIROC4-ACTM against TCCON XCO² across the latitude during the ATom campaign period shown in Figure S11. This comparison also showed higher bias over this latitude location against TCCON sites at Park Falls, JPL, Lamont and East Trout Lake. Therefore, model bias is consistent both against OCO-2 and TCCON. Then, to understand the altitude-wise variation of CO_2^{ACTM} - CO_2^{aircraft} difference at different tropospheric depths (LT, MT, UT, tropospheric column), we used 371 ATom vertical CO_2 dataset. We also compared these CO_2 differences with OCO-2 across segmented ATom tracks (Fig. 4c-f). Result shows largest CO² difference in terms of mean ± STDEV (calculated taking 1-σ standard 373 deviation) of CO₂ differences across longitude range is -0.41 ± 0.94 ppm exist in LT as compared to the other 374 tropospheric layers likely due to uncertainty and large variability in prior land $CO₂$ flux near surface (Fig. 4c). 375 When we compare CO₂ differences from other layers of different track segments, North America and neighbour's CO² difference at LT appears to be the highest, mainly occurring during the ATom-1 period (Figure S10). 377 Moreover, OCO-2 $XCO₂$ difference also showed large variability with longitudinal mean CO₂ difference of -0.34 \pm 1.07 ppm compared to aircraft tropospheric column CO₂ of -0.01 \pm 0.48 ppm. This essentially reflects the model's overall good performance against ATom as compared to OCO-2 XCO2. Similarly, ACTM and ATom $CO₂$ discrepancy was also evident in vertical cross-section, highest approximately \sim 2 ppm appeared at high latitude land regions during vegetation growing (respiration) period of northern hemisphere July-August, 2016 in ATom-1 (April-May, 2018 in ATom-4) (Figure S10). This large difference occurs across the vertical altitude range prominent above 8000 meters likely arises due to the coarse resolution of the ACTM model unable to represent the vertical transport. Needs further research on improving convective transport parameterization in forward model to improve vertical mixing (Patra et al., 2018). These results are also in line with the study Frankenberg et al., 2016. We have also validated model simulation with TCCON measurements during the ATom period, results also shown differences up to 1 ppm at the sites over northern America (Figure S11).

388

389 Other three segments of CO_2^{ACTM} - CO_2^{aircraft} difference are primarily focused over oceanic regions (Southern, 390 Pacific, Atlantic) where magnitude of ocean $CO₂$ flux variability is less as compared to land regions as much as 391 10 times less. Land and ocean $CO₂$ flux variability over the longitude and latitude band around ATom tracks for 392 different campaigns is shown in Figure S12. Figure 4d shows that over the Southern Ocean, the model-observation 393 CO₂ difference is lowest within ± 0.2 ppm for each vertical tropospheric layer and 0.06 ppm for the aircraft 394 tropospheric column, with minimal variability compared to other layers. When compared with other ATom 395 segments, the Southern Ocean shows the lowest CO₂ difference in both mean and variability. This is because the 396 aircraft sampling is at the background troposphere and is farthest from land having little influence from land CO₂ 397 air mass. This reflects that the optimized MIROC4-ACTM model, considering 50 ground-based sites, simulates 398 fairly well the aircraft background concentrations, however, it is unable to match a similar level of reproducibility 399 for OCO-2 XCO₂. Further, the latitudinal CO₂ difference variability against the aircraft tropospheric column 400 (lowest 8 km) is 0.15, compared to 0.77 with OCO-2 XCO₂. In most latitudes, OCO-2 XCO₂ differences are larger 401 than aircraft CO₂, indicating likely retrieval errors in OCO-2 given the lower uncertainty in aircraft CO₂

402 measurements. This $XCO₂$ difference with OCO-2 over the southern ocean is mainly during the ATom-2 period (Fig. S14); however large XCO² difference pixels near 50° W and 70° W are during ATom-3 (September-October 2017; Fig. S15). Previous studies reported that OCO-2 ocean glint retrieval over southern ocean has more residual biases while comparing against individual measurements (TCCON, In-situ, OCO-2 land) in top-down inversion (Byrne et al. 2023; O'Dell et al., 2018). In addition, this is also location of the southern hemisphere zone of stratosphere-troposphere exchange (STE) vary greatly spatiotemporally due to significant vertical mixing which strongly changes with season, less constrained by model transport, which may also result in an error in estimated posterior concentration. Next, we analysed CO² differences over the southbound Pacific segments of the ATom 410 campaign (Fig. 4e), it shows latitudinal $CO₂$ difference mean (variability) is highest of about -0.16 (\pm 0.53) ppm in LT, compared to MT and UT. On the other hand, aircraft tropospheric columns showed a mean (variability) 412 difference of approximately -0.04 (\pm 0.38) ppm whereas OCO-2 XCO₂ with a value of 0.27 (\pm 0.42) ppm. This 413 shows although the mean is significantly different however the variability in both $CO₂$ differences is close to each 414 other, this is reflected in overall matching of both aircraft columns and OCO-2 XCO₂ (Fig. 4e). Lastly, we analysed CO² difference over the Atlantic i.e., longest northbound part of ATom campaign shown in Figure 4(f). 416 Aircraft tropospheric column $CO₂$ difference of aircraft shows value of 0.03 ppm as compared to $XCO₂$ difference 417 with OCO-2 showing value of 0.26 ppm. Although, it has been observed that the latitudinal CO₂ difference in the aircraft tropospheric column closely matches pattern in OCO-2 XCO2. Another important point is that overall CO² differences variability in the Atlantic is observed higher as compared to the pacific segment especially over tropics within 30° S–30° N, more in-situ aircraft measurements are required to better understand the underlying error. Individual ATom campaigns results are presented in supplementary Figure S13-16.

422

423 **3.2.3 Amazon**

424
425 The climate-sensitive global tropic is a crucial part of the global carbon cycle due to the threats posed by climate 426 change, especially the Amazon region, which holds the largest above-ground biomass (AGB) pool of 427 approximately $123 \pm \text{PgC}$ (Malhi et al., 2006; Santoro et al., 2010). Inversion based estimate showed Amazon 428 was a carbon source of 0.3 ± 0.2 PgC/yr in agreement with bottom-up calculation (Alden et al., 2016; Beienen et 429 al., 2015; Gatti et al., 2014, 2021a,c) during 2010-2019, though significant uncertainty remain. This is also the 430 region under-sampled by OCO-2 retrievals due to clouds and high spatial resolution satellite monitoring is needed 431 in the future (Frankenberg et al., 2024). Prevalent uncertainty in flux estimation in modelling approach and low 432 sampling of satellites highlight the need for more research in understanding error better way and improving both 433 inversion and retrievals methods over Amazon. In section 3.1, we observed large model-observation XCO₂ 434 differences exist over South America, especially over Amazon. To investigate it further, we have utilised the 435 vertical profile (VP) of CO² measurements from vertical aircraft campaigns across Brazilian Amazon sites, SAN, 436 ALF, RBA, and TEF presented in Figure 5(a). The aircraft measurement has an accuracy of ~0.03 ppm (Gatti et 437 al., 2023) and a detailed description of measurements can be found in Gatti et al., 2021a. Studies have shown that 438 although these VPs are taken up to an altitude range of 4 km, they provide important insights into $CO₂$ variability 439 near the surface (Gatti et al., 2023; Tejada et al., 2023). An important point to note here, since Amazon aircraft 440 campaigns measure VPs of CO₂ approximately at an altitude up to 4 km, therefore, we considered the tropospheric 441 column as 0 to 4 km and only those VPs having measurements at least 4 km are chosen for calculation. We kept

442 the same criteria for the data availability of 80% vertical bin filter for 500-meter bin resolution as mentioned 443 previously in the methodology section.

444

445 Before analysing the ACTM bias against the Amazon aircraft CO₂, we validated ACTM simulated CO₂ with the 446 aircraft CO₂ at these sites which are not used in the inversion, showing a good correlation (r) of ~0.8 at 95% 447 significance level shown in Figure S17. We also checked that model able to capture TCCON XCO₂ at Manaus, 448 Brazil with time series XCO² difference mean of 0.05 ppm, but the model shows higher differences (-1 ppm) with 449 OCO-2 XCO² in Figure 2. It hints, OCO-2 retrieval likely have error could arise from Amazon dense vegetation 450 cover, cloud cover-aerosols and high humid conditions which can block sunlight spectra, reduce the signal 451 strength, limit valid sampling and increase retrieval error (Frankenberg et al., 2024; Taylor et al., 2016; Yu et al, 452 2019). These retrieval challenges precludes robust understanding of inversion error across the broader Amazonian 453 region. To further check this error altitude wise, a monthly mean time series of ACTM-aircraft CO₂ difference 454 considering all vertical profiles within a month is calculated for three vertical tropospheric layers, LT (lowest–2 455 km), MT (2–4 km) and tropospheric column (lowest-4 km) during OCO-2 measurement periods is presented in 456 Figure 5c-f. Figure 5c represents CO² difference at SAN aircraft campaign sites having data gaps from mid-2015 457 to early-2017 because of no measurements conducted during this period. Maximum model-observation 458 differences in terms of mean(variability) of 0.93(± 3.36) ppm observed in LT as compared to MT and tropospheric 459 column. This mismatch is comparable with previous study by Basso et al. 2023. Further, OCO-2 XCO₂ difference 460 showed overall negative mean of -0.83 ppm with variability of \pm 1.04 ppm as compared to aircraft VPs profile 461 with aircraft tropospheric column shown better constrained having value of 0.76 ppm. Further, we analysed VPs 462 at the ALF site presented in Figure 5(d) shows overall that aircraft model-observation CO₂ difference matches 463 well with XCO₂. CO₂ differences at LT, MT, tropospheric column, XCO₂ shows mean (STDEV) are -0.9 (\pm 4.24), 464 0.08 (\pm 2.03), -0.13 (\pm 2.48), -0.65 (\pm 1.03) respectively. Basso et al. 2023 has shown that some of this difference 465 between inversion and aircraft CO² could be significantly improved (57% below 1.5 km and 49% above 3.5 km) 466 when using regional aircraft CO_2 data in the inversions. . In RBA, CO_2 difference at LT, MT, tropospheric 467 column, XCO₂ shows mean (STDEV) are -0.61 (\pm 4.33), -0.03 (\pm 2.52), 0.27 (\pm 2.95), -0.69 (\pm 1.04) respectively 468 shown in Figure 5e. Therefore, it shows CO₂ difference with the aircraft tropospheric column (OCO-2 XCO₂) has 469 opposite signature; it represents ACTM over (under) estimates considering the whole time window. In TEF, CO² 470 difference at LT, MT, tropospheric column, XCO₂ shows mean (STDEV) are -0.4 (\pm 4.29), 0.64 (\pm 3.02), 0.19 (\pm 471 2.89), -1.37 ± 0.99 respectively presented in Figure 5f. Except for SAN, at all other sites, we observed that the 472 ACTM matches in total column better with aircraft than OCO-2, and but this profile is still insufficient to match 473 with XCO₂ needs further high profile measurement over this location. It is worth noting that the large discrepancy 474 or bias in LT in RBA, TEF (SAN, ALF) during January-March (August-December) in west-central (south-east) 475 Amazon regions may potentially arise due to fire CO₂ emission is reported in Basso et al. 2023. Since our inverse 476 simulations using CASA biospheric flux lack observation-based biomass burning data, this could also affect the 477 overall simulated concentration as well. We also checked monthly land $CO₂$ flux anomaly, calculated by taking 478 area mean around campaign sites within $5^\circ \times 5^\circ$ degree and then removing seasonal cycle from actual time series. 479 We noticed no such anomalous flux change during the anomalous $CO₂$ difference period, likely due to the coarse 480 resolution of the MIROC4-ACTM and also because no regional CO₂ data from Amazon is used in our inversion 481 which could potentially capture Amazon land CO₂ flux changes better way (Fig. 5b; Fig. S1). Basso et al. 2023

- 482 highlighted the importance of assimilating Amazon aircraft measurements in deriving regional land $CO₂$ flux. In 483 all Amazon aircraft sites, an increase in land CO₂ flux during 2015-16 was observed due to strong ENSO events 484 occurred during this period also reported in Das et al. (2022).
- 485

486 **3.2.4 Asia**

487 In Asia there are very few aircraft campaigns for CO₂ measurements compared to Northern America and Europe 488 (Crevoisiera et al., 2010; Xueref-Remy et al., 2011). Although, efforts have been made to measure CO₂ vertical profile over monsoon-dominated Indian subcontinents for a shorter time period (Vogel et al., 2023). Therefore, 490 available long-term $CO₂$ measurements like CONTRAIL is very important to provide unprecedented insights into long-term CO² variability in UT/LS and model evaluations over these regions (Bisht et al., 2021; Das et al., 2022; Niwa et al., 2011). Therefore, we have utilised these measurements to compare and understand model- observations CO² difference for OCO-2 and CONTRAIL aircraft in different regions across Asia. Figure 6a depicts the spatial distribution of the CONTRAIL campaign CO² sampling location from January 2015-December, 2021, covering altitudes ranging up to ~12 to 14 km with topographic altitudes information (topography elevation 496 data is downloaded from https://www.ncei.noaa.gov/products/etopo-global-relief-model). Here, we have selected 497 four separate regions around airport locations delineated through deep green colors having $CO₂$ vertical profiles resulting from aircraft ascent or descent near airports. The four regions are namely Far East Asia, Southeast China, northern Southeast Asia and Equatorial Southeast Asia, based on the locations of airports. In Far East Asia, two airports are considered: Tokyo International Airport, Japan (site code: HND) (35.6° N, 139.8° E) and Narita International Airport (site code: NRT) (35.8° N, 140.4° E) are considered together, named TYO (35.7° N, 140.8° E); in Southeast China, Hong Kong International Airport (site code: HKG) (22.2° N, 113.6° E); in northern Southeast Asia, Suvarnabhumi International Airport, Thailand (site code: BKK) (13.7° N, 100.7° E); and in southern Southeast Asia, Singapore Changi International Airport, Singapore (site code: SIN) (1.4° N, 104.0° E), all airports are marked with a small square box in Figure 6a ,b. During the aforementioned period, no vertical sampling was performed over the Indian subcontinent and other two airports highlighted on map Incheon International Airport (site code: ICN) and Shanghai Pudong International Airport (site code: PVG), were not 508 considered due to less number of sampling dataset. Figure 6b presents mean model-OCO-2 XCO₂ differences over 509 sampling locations of CONTRAIL, showing mainly negative CO₂ difference ranging -0.5 to -1 ppm over boxed 510 airports location highlighting likely reason is underestimation overall fossil emission of urban CO₂ signature in 511 the model. There is very limited TCCON sites over city scale that validates OCO-2 XCO₂, however, Rißmann et 512 al. (2022) using Munich Urban Carbon Column network (MUCCnet) XCO₂ across three sites over Germany found out OCO-2 has a RMSE of 0.6 ppm in urban site. Since OCO-2 has retrieval error over city scale it makes it challenging to discuss the sources of error could come from the model.

515 To understand this we analysed more robust CONTRAIL aircraft $CO₂$ (~0.2 ppm for each CONTRAIL data point), 516 figure 6c-f represents a time series of model-observation CO₂ differences over each airport for different vertical 517 depths of troposphere and XCO2. Here, we have considered all CO² vertical profiles, selecting aircraft ascent and 518 descent flight modes over airports within a month and done a monthly average for 200-meter vertical bins to 519 calculate partial column $CO₂$ for aircraft and similarly for model simulations resampled at aircraft measurement 520 location considering methodology described in section 2.4. For OCO-2, we computed the mean over designated

521 airports to calculate model-observation difference for $XCO₂$ for the specific months. Results show in far east Asia, 522 TYO location CO₂ difference in LT, MT, UT, tropospheric column, XCO₂ shows mean of -2.0, -0.88, -0.73, -523 1.02, and -0.3 respectively. In HKG airport, the number of samples in LT was very less therefore ignored in the 524 analysis however CO_2 difference in MT, UT, tropospheric column, XCO₂ shows values of -1.2, -0.99, -1.13 and 525 -0.02 respectively. For northern Southeast Asia, BKK airport CO² difference at LT, MT, UT, tropospheric 526 column, XCO₂ shows mean (STDEV) of -2.71 (\pm 1.67), -0.83 (\pm 0.74), -0.6 (\pm 0.59), -1.06 (\pm 0.71), and -0.1 (\pm 527 0.71) respectively. Further, in equatorial Southeast Asia, SIN airport difference at LT, MT, UT, tropospheric 528 column, XCO₂ shows mean (STDEV) of -1.89 (\pm 1.26), -1.03 (\pm 0.54), -0.81 (\pm 0.59), -1.05 (\pm 0.56), and -0.25 529 (± 0.42) respectively.

 Result indicates that in northern Southeast Asia and southern Southeast Asia, mean and variability of model-531 observation CO₂ difference is higher in LT as compared to UT, MT and weighs more. In all regions, model- observation difference for OCO-2 showed better constrained compared to aircraft measurements and readily observable that it closely matches the tropospheric column pattern. A notable fact in all regions is that the total time series mean of model-observation difference is negative for both aircraft (-1.02 to -1.13 ppm) and OCO-2 (- 535 0.02 to -0.3 ppm), which would imply an underestimation of model simulated CO₂. While the OCO-2 XCO₂ vs 536 MIROC4-ACTM differences are not statistically significant but the large and systematic CONTRAIL CO₂ vs MIROC4-ACTM differences may suggest that actual emission footprints captured by satellite observations are 538 greater than the measurement resolution $(-1.29 \times 2.25 \text{ km}^2 \text{ for OCO-2})$. It suggests OCO-2 capturing emissions 539 from broader urban areas than its nominal resolution, possibly either due to the well-mixed nature of $CO₂$ and OCO-2 measuring total column or the spatial extent of urban footprint. Further, the large variability and significant 541 differences between the aircraft $CO₂$ column and XCO₂ are evident in all regions. This is likely attributable to the selection of a specific box area, which surrounds airport locations situated in urban areas, one of the significant sources for fossil CO² emission. This inference is discussed in earlier studies (Patra et al., 2011; Umezawa et al. 2020), wherein they reported an urban emission footprint in CONTRAIL aircraft measurements conducted over 545 airport megacities. The inversion process, utilized in this context, exclusively optimizes total CO₂ fluxes, for biosphere and ocean regions considering background sites, whereas this CONTRAIL measurement over airports having signature of urban interiors. Consequently, noteworthy disparities may emerge due to uncertainties 548 associated with fossil fuel CO₂ emissions and also coarse horizontal resolution of MIROC4-ACTM (T42, \sim 2.8° $549 \times 2.8^\circ$) unable to reproduce the sub-grid-scale variations. This limitation does not influence optimized flux for the large area studies but affect our ability to simulate posterior concentrations, leading to underprediction of concentration near the surface over the emissions or sinks hotspots, e.g., anthropogenic emissions at the megacity areas or plumes of intense biomass burning. Note that the location of ascent and descent of the aircraft may change by seasons following the meteorological conditions, and thus the location of measurements less strictly follows 554 year around. Previous studies underscore the critical role of fossil fuels in shaping simulated CO₂ dynamics, emphasizing their potential to introduce systematic errors in optimized surface fluxes (Suntharalingam et al., 2005; Wang et al., 2020).

557 **3.3 Discussion and conclusions**

- 563 We demonstrated that MIROC4-ACTM, using only 50 surface-based CO₂ sites globally, accurately 564 simulates tropospheric column and OCO-2 XCO₂, showing strong agreement with aircraft and TCCON 565 data (correlation = 0.9 , $p < 0.0001$) at most sampling sites.
- 566 Our analysis highlighted that the regional hemispheric MIROC4-ACTM CO₂ difference with OCO-2 567 and in-situ measurements has heterogeneous signatures of CO₂ differences, particularly over land. However, Kulawik et al. 2019, noting that OCO-2 retrievals over lands have more random errors especially over Amazon which is less sampled by OCO-2 makes retrieval less reliable for comparison (Frankenberg et al., 2024). Additionally, comparison against in-situ indicates that OCO-2 likely has a systematic retrieval error over the southern hemisphere oceanic region. Both random error over land, less sampling of OCO-2 over global tropics and systematic error over ocean makes it difficult to detect and understand the uncertainties in inverse models in a global perspective. We need more vertical aircraft profile measurements to more robustly understand this error especially over the global tropics.
- 575 Altitude wise comparison of CO₂ difference from categorical specific and campaign aircraft measurements around the globe consistently highlight the model's highest mismatch in LT as compared to MT, UT, and tropospheric columns. Additionally, LT contributes more to the mean and variability 578 to the total tropospheric column than the MT, UT. This maximum uncertainty in the LT likely arises 579 from the uncertainties in prior fluxes near the surface. In contrast, the MT and UT, where large-scale dynamical mixing predominates, show better model performance, likely due to realistic transport of the 581 forward model. Further, aircraft tropospheric $CO₂$ columns are better reproduced by MIROC4-ACTM compared to individual tropospheric layers and OCO-2 XCO2. Further, studies have shown OCO-2 583 XCO₂ is more prone to erroneous retrieval due to near surface aerosol and cloud contamination in LT which makes it challenging for total column comparison with model (Connor et al., 2016; Massie et al., 2021).
- 586 Results from ATom show large CO₂ difference variability over North America regions as compared to 587 other integrated tracks over ocean, likely because of the influence of land air mass having large variability 588 in land $CO₂$ flux. Similarly, large $CO₂$ differences in aircraft sites over Amazon may likely arise due to uncertainty in prior flux and coarse resolution of the model unable to represent small scale variation requires more regional measurements in inversion, however, comparison against OCO-2 highlights robust requirement of good amount of valid retrievals to diagnose the inversion from large region perspective as well as insufficient high-altitude profile measurements (~4 km) demands more high profile measurements. Aircraft measurements over the remote background troposphere in the Pacific, Southern Ocean, and Atlantic showed the best match within 0.03 ppm when compared to OCO-2 with 0.2 ppm,

Code availability

- Data and figure processing codes prepared for this study will be made available upon reasonable request from the
- corresponding author. Inversion code used here is available on https://github.com/prabirp/co2l2r84.

Data availability

 The ObsPack data product is available at https://gml.noaa.gov/ccgg/obspack/, CONTRAIL at https://www.cger.nies.go.jp/contrail/, Amazon aircraft campaign at https://doi.pangaea.de/10.1594/PANGAEA.926834. ACTM model outputs at aircraft sampling and OCO-2 sounding locations will be made accessible upon requests from corresponding authors.

Author contributions

- CD, PKP and RKK developed the idea and methodology of the study. CD ran the entire analysis and prepared the
- main manuscript and supplementary. PKP, RKK and NC helped in review and editing. KI, and TM are part of the
- CONTRAIL CO² measurement group. All co-authors actively engaged in scientific discussions.

Competing interests

-
- The authors declare that they have no conflict of interest.

Acknowledgments

 All co-authors sincerely thanks to National Oceanic and Atmospheric Administration (NOAA), Japan Meteorological Agency (JMA), Commonwealth Scientific and Industrial Research Organisation (CSIRO), Laboratoire des sciences du climat et de l'environnement/Institut Pierre Simon Laplace (LSCE/IPSL), Scripps Institution of Oceanography (SIO), South African Weather Service (SAWS), Environment and Climate Change Canada (ECCC), Mt. Waliguan/ China Meteorological Administration (WLG/CMA), Tae-ahn Peninsula/Korean Meteorological Administration (TAP/KMA) for site observations, and the CONTRAIL, Amazon, ATom campaign teams for the aircraft observations. Surface in-situ CO² observations from Jungfraujoch are supported by the Swiss Federal Office for the Envrionment and ICOS Switzerland (ICOS-CH). This study would not have been possible without these $CO₂$ concentration measurements datasets around the globe.

Financial support

CD thanks to the Director of IIT Delhi, for providing PhD fellowship to carry out this work. This research is partly

supported by the Environment Research and Technology Development Fund (JPMEERF24S12205) of the

- Environmental Restoration and Conservation Agency of Japan, and the Arctic Challenge for Sustainability phase
- II (ArCS-II; JPMXD1420318865) Projects of the Ministry of Education, Culture, Sports, Science and Technology
- (MEXT).

Reference

921 emission uncertainties and accounting for 3-D chemical CO₂ production on inverse natural carbon flux

922 estimates from satellite and in situ data. Environmental Research Letters, 15(8).
923 https://doi.org/10.1088/1748-9326/ab9795, 2020. 923 https://doi.org/10.1088/1748-9326/ab9795, 2020.
924 WDCGG (World Data Centre for Greenhouse Gases

924 WDCGG (World Data Centre for Greenhouse Gases): World Data Centre (WDC) operated by the Japan 925 Meteorological Agency (JMA) under the Global Atmosphere Watch (GAW) programme of the World 925 Meteorological Agency (JMA) under the Global Atmosphere Watch (GAW) programme of the World
926 Meteorological Organization (WMO) [data set], https://gaw.kishou.go.jp/ 926 Meteorological Organization (WMO) [data set], https://gaw.kishou.go.jp/
927 Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained

927 Wofsy, S. C.: HIAPER Pole-to-Pole Observations (HIPPO): fine- grained, global-scale measurements of 928 climatically important atmospheric gases and aerosols. Philos. T. R. Soc. A., 369, 2073– 2086. 928 climatically important atmospheric gases and aerosols, Philos. T. R. Soc. A., 369, 2073– 2086, 929 https://doi.org/10.1098/rsta.2010.0313, 2011. 929 https://doi.org/10.1098/rsta.2010.0313, 2011.
930 Wofsy, S.C., S. Afshar, H.M. Allen, E.C. Apel, E.C.

- 930 Wofsy, S.C., S. Afshar, H.M. Allen, E.C. Apel, E.C. Asher, B. Barletta, J. Bent, H. Bian, B.C. Biggs, D.R. Blake, 931 N. Blake, I. Bourgeois, C.A. Brock, W.H. Brune, J.W. Budney, T.P. Bui, A. Butler, P. Campuzano-Jost, 931 N. Blake, I. Bourgeois, C.A. Brock, W.H. Brune, J.W. Budney, T.P. Bui, A. Butler, P. Campuzano-Jost, C.S. Chang, M. Chin, R. Commane, G. Correa, J.D. Crounse, P. D. Cullis, B.C. Daube, D.A. Day, J.M. 932 C.S. Chang, M. Chin, R. Commane, G. Correa, J.D. Crounse, P. D. Cullis, B.C. Daube, D.A. Day, J.M. 933 Dean-Day, J.E. Dibb, J.P. DiGangi, G.S. Diskin, M. Dollner, J.W. Elkins, F. Erdesz, A.M. Fiore, C.M. 934 Flynn, K.D. Froyd, D.W. Gesler, S.R. Hall, T.F. Hanisco, R.A. Hannun, A.J. Hills, E.J. Hintsa, A. Hoffman, 934 Flynn, K.D. Froyd, D.W. Gesler, S.R. Hall, T.F. Hanisco, R.A. Hannun, A.J. Hills, E.J. Hintsa, A. Hoffman,
935 R.S. Hornbrook. L.G. Huev. S. Hughes. J.L. Jimenez. B.J. Johnson. J.M. Katich. R.F. Keeling. M.J. Kim. 935 R.S. Hornbrook, L.G. Huey, S. Hughes, J.L. Jimenez, B.J. Johnson, J.M. Katich, R.F. Keeling, M.J. Kim, 936 A. Kupc, L.R. Lait, K. McKain, R.J. Mclaughlin, S. Meinardi, D.O. Miller, S.A. Montzka, F.L. Moore, E.J.
937 Morgan, D.M. Murphy, L.T. Murray, B.A. Nault, J.A. Neuman, P.A. Newman, J.M. Nicely, X. Pan, W. 937 Morgan, D.M. Murphy, L.T. Murray, B.A. Nault, J.A. Neuman, P.A. Newman, J.M. Nicely, X. Pan, W.
938 Paplawsky, J. Peischl, M.J. Prather, D.J. Price, E.A. Ray, J.M. Reeves, M. Richardson, A.W. Rollins, K.H. 938 Paplawsky, J. Peischl, M.J. Prather, D.J. Price, E.A. Ray, J.M. Reeves, M. Richardson, A.W. Rollins, K.H. 939 Rosenlof, T.B. Ryerson, E. Scheuer, G.P. Schill, J.C. Schroder, J.P. Schwarz, J.M. St.Clair, S.D. Steenrod,
940 B.B. Stephens, S.A. Strode, C. Sweeney, D. Tanner, A.P. Teng, A.B. Thames, C.R. Thompson, K. Ullmann, 940 B.B. Stephens, S.A. Strode, C. Sweeney, D. Tanner, A.P. Teng, A.B. Thames, C.R. Thompson, K. Ullmann, 941 P.R. Veres. N.L. Wagner, A. Watt, R. Weber, B.B. Weinzierl, P.O. Wennberg, C.J. Williamson, J.C. 941 P.R. Veres, N.L. Wagner, A. Watt, R. Weber, B.B. Weinzierl, P.O. Wennberg, C.J. Williamson, J.C.
942 Wilson, G.M. Wolfe, C.T. Woods, L.H. Zeng, and N. Vieznor.: ATom: Merged Atmospheric Chemistry. 942 Wilson, G.M. Wolfe, C.T. Woods, L.H. Zeng, and N. Vieznor.: ATom: Merged Atmospheric Chemistry,
943 Trace Gases, and Aerosols, Version 2. ORNL DAAC, Oak Ridge, Tennessee, 943 Trace Gases, and Aerosols, Version 2. ORNL DAAC, Oak Ridge, Tennessee, 944 USA. https://doi.org/10.3334/ORNLDAAC/1925, 2021. 944 USA. https://doi.org/10.3334/ORNLDAAC/1925, 2021.
- 945 Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, M. C., O'Dell, C., Mandrake, L., Viatte, 946 C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, 946 C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De
947 Maziere, M., Sha, M. K., Sussmann, R., ... Eldering, A. (2017). Comparisons of the Orbiting Carbon 947 Maziere, M., Sha, M. K., Sussmann, R., ... Eldering, A. (2017). Comparisons of the Orbiting Carbon
948 Observatory-2 (OCO-2) XCO₂ measurements with TCCON. In Atmospheric Measurement Techniques 948 Observatory-2 (OCO-2) XCO₂ measurements with TCCON. In Atmospheric Measurement Techniques
949 (Vol. 10, Issue 6). https://doi.org/10.5194/amt-10-2209-2017, 2017. 949 (Vol. 10, Issue 6). https://doi.org/10.5194/amt-10-2209-2017, 2017.
950 Xueref-Remy, I., Bousquet, P., Carouge, C., Rivier, L., & Ciais, P.: Vari
- 950 Xueref-Remy, I., Bousquet, P., Carouge, C., Rivier, L., & Ciais, P.: Variability and budget of CO_2 in Europe:
951 Analysis of the CAATER airborne campaigns-Part 2: Comparison of CO_2 vertical variability and fluxes 951 Analysis of the CAATER airborne campaigns-Part 2: Comparison of CO₂ vertical variability and fluxes between observations and a odelling framework. Atmospheric Chemistry and Physics, 11(12). 952 between observations and a odelling framework. Atmospheric Chemistry and Physics, 11(12).
953 https://doi.org/10.5194/acp-11-5673-2011. 2011. 953 https://doi.org/10.5194/acp-11-5673-2011, 2011.
954 Yu, L., Wen, J., Chang, C. Y., Frankenberg, C., & Si
- 954 Yu, L., Wen, J., Chang, C. Y., Frankenberg, C., & Sun, Y.: High-Resolution Global Contiguous SIF of OCO-
955 2. Geophysical Research Letters, 46, 1449–1458. https://doi.org/10.1029/2018GL081109. 2019. 955 2. *Geophysical Research Letters*, 46, 1449–1458. https://doi.org/10.1029/2018GL081109, 2019.
- 956 957 958
- 959
-
- 960

- 962
- 963 964
- 965
- 966
- 967
- 968
- 969
- 970
- 971 972
- 973
- 974
- 975
- 976

984 **Figure 1:** Schematic of CO₂ concentration vertical profile (dark blue line) by satellite, ACTM, and aircraft CO₂ 985 (orange, golden) Arrowheads represent different layers of the atmosphere, specifically LT (lowest–2 km), MT 986 (2–5 km), UT (5–8 km), and the tropospheric column (lowest–8 km) corresponds to the aircraft CO₂ 987 measurement. Blue arrow represents total variation captured by satellite covers from surface to top of the 988 atmosphere. 989

991

992 **Figure 2:** CO₂ space-time variability with (a) spatial mean XCO₂ difference map between InvFG and OCO-2 993 during January,2015-December,2021. (b) Time vs latitude distribution of XCO₂ difference between InvFG and 994 OCO-2 considering mean across global longitude. (c) Time vs latitude cross-section of CO₂ concentration 995 difference between InvFG and in-situ CO₂ measurement, considering CO₂ from 53 surface sites. (d) Latitude

996 averaged time series of CO_2 (XCO₂) concentration difference between InvFG and Surface (OCO-2) respectively

represented by black (red) colours. "r" value in panel-d represents correlation between time series of surface and

998 OCO-2 difference at 99% significance level.

 Figure 3. Mean model-observation CO² difference (ppm) at different vertical tropospheric depths LT (light red), 1004 MT (orange), UT (dodger blue), total column (teal) and XCO₂ for specific sites aircraft (panel–a) and campaign aircraft measurements having latitudinal coverage maximum 30° (panel–b). Aircrafts names are organized based on aircraft observations location, progressing from high latitudes in the Northern Hemisphere, through the equator, to southern latitudes. The second x-axis represents a number of data points for specific aircraft observation. The first and second number inside the panel represents mean and 1-σ standard deviation (STDEV) of model- observation difference across latitude for each tropospheric layer.

LT(lowest-2 km) $MT(2-5 km)$ UT(5-8 km) - Tropospheric Column(lowest-8 km) - OCO-2 XCO2

1011 1012

1013 **Figure 4:** (a) Integrated tracks traversed during ATom campaigns (ATom-1, ATom-2, ATom-3 and ATom-4). 1014 (b) Spatial model-observation XCO_2 difference against OCO-2 over ATom integrated track during campaign (c), 1015 (d), (e), and (f) shows model-observation $CO₂$ difference over different tropospheric layers from vertical $CO₂$ 1016 profile measurements of ATom and XCO₂ from OCO-2 for North America and neighbours, Southern Ocean, 1017 Pacific, and Atlantic segments respectively. Tropospheric layers are LT (light red), MT (orange), UT (dodger 1018 blue) and Total Column (teal), and OCO-2 XCO₂ (black) representation for difference against OCO-2. The first 1019 and second number on the right side of each middle and bottom panel represents the mean and 1-σ standard 1020 deviation (STDEV) of model-observation difference across latitude or longitude, respectively.

 Figure 5: (a) Amazon aircraft vertical CO² profile campaign sites, SAN, ALF, RBA, and TEF. (b) Time series of land carbon flux anomalies at the campaign sites. (c), (d), (e) and (f) represent a time series of model-observation CO2 differences for LT, MT, and tropospheric column, and XCO2 during OCO-2 measurement periods for SAN, ALF, RBA and TEF, respectively. The numbers inside the middle and bottom panels represent the mean and 1-σ 1029 standard deviation (STDEV) of model-observation CO₂ difference over the time period.

 Figure 6: (a) CONTRAIL aircraft sampling locations with associated colours represent sampling altitude in km over Asia regions with surface elevation in km. There are four defined regions, each associated with specific airports, covering various vertical zones of carbon dioxide (CO2) profiles. Far East Asia consisting of two airports, HND, NRT merged to prepare TYO, Southeast China with one airport HKG, northern Southeast Asia with one airport BKK, and equatorial Southeast Asia encompassing one airport SIN. (b) mean model-OCO2 XCO² differences over sampling locations during the mentioned period. (c), (d) (e), and (f) are time series of model- observation CO2 differences over representative airports at different vertical depths of troposphere MT, LT and 1039 tropospheric column and XCO₂ utilising aircraft measurement and OCO-2. The numbers inside the (c), (d), (e), and (f) panels represent the mean and 1-σ standard deviation (STDEV) of model-observation difference over the period.