Global nature run data with realistic high-resolution carbon weather for the year of the Paris agreement

Anna Agustí-Panareda, Joe Mcnorton, Gianpaolo Balsamo, Bianca Baier, Nicolas Bousserez, Souhail Boussetta, Dominik Brunner, Frederic Chevallier, Margarita Choulga, Michail Diamantakis, et al.

scientific data

OPEN

DATA DESCRIPTOR

Global nature run data with realistic high-resolution carbon weather for the year of the Paris Agreement

Anna Agustí-Panareda[1]✉, Joe McNorton[1], Gianpaolo Balsamo[1], Bianca C. Baier[2,3], Nicolas Bousserez[1], Souhail Boussetta[1], Dominik Brunner[4], Frédéric Chevallier[5], Margarita Choulga[1], Michail Diamantakis[1], Richard Engelen[1], Johannes Flemming[1], Claire Granier[6,2,14], Marc Guevara[7], Hugo Denier van der Gon[8], Nellie Elguindi[6], Jean-Matthieu Haussaire[4], Martin Jung[9], Greet Janssens-Maenhout[10], Rigel Kivi[11], Sébastien Massart[1], Dario Papale[12], Mark Parrington[1], Miha Razinger[1], Colm Sweeney[3], Alex Vermeulen[13] & Sophia Walther[9]

The CO_2 Human Emissions project has generated realistic high-resolution 9 km global simulations for atmospheric carbon tracers referred to as nature runs to foster carbon-cycle research applications with current and planned satellite missions, as well as the surge of *in situ* observations. Realistic atmospheric CO_2, CH_4 and CO fields can provide a reference for assessing the impact of proposed designs of new satellites and *in situ* networks and to study atmospheric variability of the tracers modulated by the weather. The simulations spanning 2015 are based on the Copernicus Atmosphere Monitoring Service forecasts at the European Centre for Medium Range Weather Forecasts, with improvements in various model components and input data such as anthropogenic emissions, in preparation of a CO_2 Monitoring and Verification Support system. The relative contribution of different emissions and natural fluxes towards observed atmospheric variability is diagnosed by additional tagged tracers in the simulations. The evaluation of such high-resolution model simulations can be used to identify model deficiencies and guide further model improvements.

Background & Summary

Reducing human-made emissions of CO_2 is at the heart of the climate change mitigation efforts in the Paris Agreement. In support of such efforts, the CO_2 Human Emission (CHE) project (www.che-project.eu) has designed a prototype system to monitor CO_2 fossil fuel emissions at the global scale. This challenging task requires the capability to detect and quantify the localised and relatively small signals of fossil fuel emissions in the atmosphere compared to the large variability of background CO_2 concentrations not directly affected by local sources, and to distinguish anthropogenic sources from vegetation fluxes[1–3]. Using observations of atmospheric constituents to estimate emissions[4,5] relies on a good understanding and accurate modelling of their atmospheric variability, which is largely determined by the weather-driven atmospheric transport together with surface biogenic fluxes and anthropogenic emissions. In the CHE project a library of nature runs of CO_2 and

[1]European Centre for Medium Range Weather Forecasts, Reading, RG2 9AX, UK. [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA. [3]NOAA, Global Monitoring Laboratory, Boulder, CO, USA. [4]EMPA, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, Dübendorf, Switzerland. [5]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, Université Paris Saclay, 91191, Gif-sur-Yvette CEDEX, France. [6]Laboratoire d'Aérologie, CNRS-Université de Toulouse, Toulouse, France. [7]Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain. [8]TNO, Department of Climate, Air and Sustainability, Utrecht, the Netherlands. [9]Max Planck Institute for Biogeochemistry (MPI-BGC), Jena, Germany. [10]European Commission, Joint Research Centre (JRC), Directorate for Energy, Transport and Climate, Air and Climate Unit, Via E. Fermi 2749, I-21027, Ispra, VA, Italy. [11]Finnish Meteorological Institute, Sodankylä, Finland. [12]Dipartimento per la Innovazione nei Sistemi Biologici, Agroalimentari e Forestali, Università degli Studi della Tuscia, Largo dell'Università, Viterbo, Italy. [13]ICOS ERIC Carbon Portal, Sölvegatan 12, 22362, Lund, Sweden. [14]NOAA, Chemical Sciences Laboratory, Boulder, CO, USA. ✉e-mail: Anna.Agusti-Panareda@ecmwf.int

species co-emitted with CO_2 has been produced at different scales and with varying degrees of complexity[6] which complements previous nature runs[7].

Nature runs are very high-resolution simulations that mimic nature, in that they provide a realistic representation of processes of interest, in this case those modulating atmospheric CO_2 variability. These simulations provide a reference for Observation System Simulation Experiments (OSSEs)[8] Quantitative Network Design (QND)[9]. In OSSEs and QND studies, synthetic observations extracted from nature runs are used to assess the impact of different observing system configurations[10]. It is envisaged that such a monitoring system will rely on the use of a large variety of measurements including species co-emitted with CO_2 that can help to isolate the fossil fuel emissions[3,11]. The future CO2M (Copernicus CO2 Monitoring) satellite mission is purposely designed to provide a high-resolution imaging capability to detect CO_2 emission hotspots with high-precision observations of atmospheric CO_2 concentrations[2,3,12]. CO2M will complement a constellation of satellites[4] and a global *in situ* network[5] to quantify the atmospheric CO_2 variability from which emissions will be derived with atmospheric inversion systems.

Simulating a realistic distribution of CO_2 and co-emitters depends on the representation of the surface fluxes, chemical sources/sinks, and atmospheric transport. Here we use the Copernicus Atmosphere Monitoring Service (CAMS) high-resolution forecast of CO_2, CH_4 and CO (https://atmosphere.copernicus.eu/charts/cams/carbon-dioxide-forecasts) which has been demonstrated to produce realistic and accurate variability of carbon weather[13–15]. The configuration of the nature run is shown in Fig. 1. Note that the CHE nature run is a free-running tracer simulation unlike the CAMS high-solution forecast which is initialised daily from an atmospheric composition analysis.

The CHE nature run aims to support scientific studies that will shed light on the challenges of estimating CO_2 emissions with the goal to build a CO_2 monitoring and verification support capacity[3]. These challenges span a wide range of aspects from sparse observing systems, consistency between ocean/land observations from different satellite-view modes[16], large variability in the biogenic signal[17], large representativity errors in anthropogenic emissions[13], transport errors[18] and stringent requirements of high accuracy observations to estimate small signal with respect to large background values[16,19]. This global high-resolution dataset can provide a reference for testing different approaches to address those challenges.

Methods

Modeling framework. The CAMS high resolution forecasting system at the European Centre for Medium Range Weather Forecasts (ECMWF)[13,14,20] has been used to produce the nature run dataset which includes simulations of CO_2, CH_4 and CO as illustrated in Fig. 1. It is based on the Integrated Forecasting System (IFS) model cycle 46R1 used to produce the operational weather forecast from June 2019 to June 2020[21]. The model has a reduced octahedral Gaussian grid[22] with a resolution of Tco1279 (corresponding to approximately 9 km) and 137 model levels. The simulations have been produced by running a sequence of 1-day IFS forecasts of the carbon tracers and weather. The weather forecasts are initialized with state-of-the-art re-analysis of meteorological fields (ERA5)[23]. The atmospheric tracers start from the CAMS re-analysis[24,25] initial conditions at the initial date of the dataset and from then onwards they are cycled from one forecast to the next in a free-running style. The different model components for the carbon weather forecast in the IFS, including the representation of the emissions for the different tracers, are listed in Table 1. All the emissions at the surface are prescribed except for the CO_2 biogenic fluxes which are modelled online[26,27], providing consistency between the response of fluxes to atmospheric conditions and tracer transport[28]. There are various differences with respect to the CAMS operational high-resolution forecast in 2015: improved anthropogenic emissions[29–31] and natural CO_2 ocean fluxes[32]; as well as an improved IFS model version[21] and initial conditions[23–25]. The configuration of the simulations with daily re-initialisation of the weather forecast and free-running tracers ensures consistency of the tracer evolution throughout the simulation by avoiding jumps in their concentrations brought by the assimilation of observations in the analysis, while maintaining a realistic and accurate simulation of their atmospheric transport and variability of the underlying biogenic fluxes from the model[26,27].

Model output. The standard parameters available from the CHE nature run dataset are listed in Table 2 and Table S1 in Supplementary Information file 1. Additional experimental tagged tracers are provided to characterize the atmospheric enhancement associated with the natural surface fluxes and anthropogenic emissions (Table 3). The enhancement can be computed by subtracting the concentrations of the background tracer without the specific emission/flux from the tracer concentration with the flux/emission. This assumes that the transport is linear. It is worth noting that artificial negative enhancements can occur in the vicinity of plumes due to numerical oscillations associated with the cubic interpolation of the advection scheme around very steep gradients. This can be considered a numerical error in the simulation. The CO_2 tagged tracers are simulated without applying any mass fixer in order to ensure the signal comes only from the flux. The tagged tracers provide the enhancement during each 1-day simulation as they are re-initialised every day at 00UTC in order to avoid growing errors associated with the mass conservation[33,34]. This means the flux enhancement is reset to zero at 00 UTC. Detailed information on those tracers is provided in Table 4.

Figure 1b provides an overview of the different types of model output from the CHE nature run dataset and how these can be compared to other datasets including various types of observations[5,35,36] as well as atmospheric inversions/simulations of carbon tracers[9]. Such a comparison can shed some light on the different components of the uncertainty in the simulations of carbon tracers coming from the surface fluxes, the atmospheric transport and the representativity error associated with the limited model resolution[14]. A complementary lower resolution ensemble of simulations[18] (25 km in the horizontal) has been also produced using the same model setup which provides information on emission uncertainty[30], transport uncertainty and impact of meteorological uncertainty on biogenic fluxes. Two other major sources of uncertainty stem from the initial conditions of the carbon

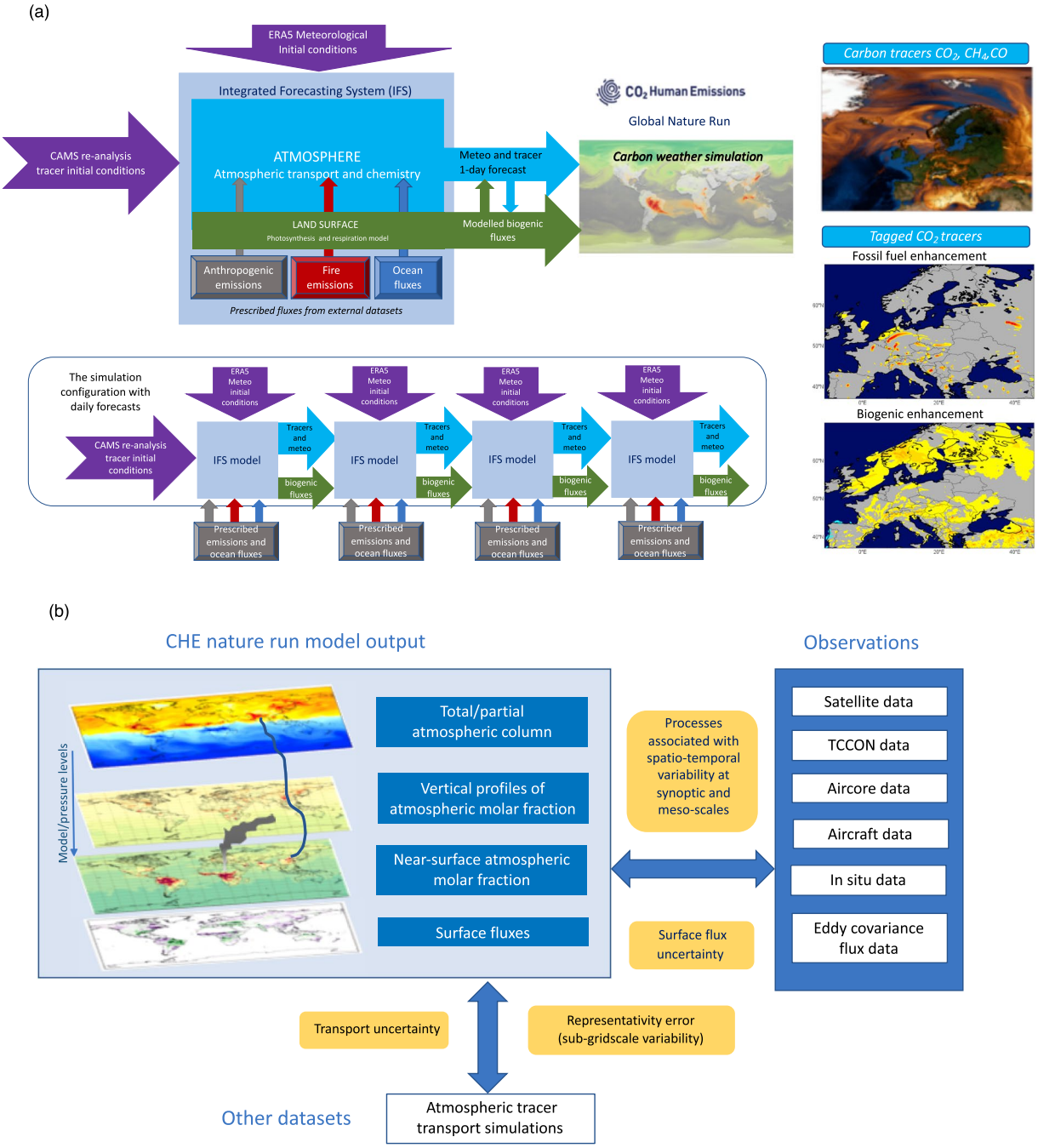

Fig. 1 (**a**) Schematic of production framework for CHE nature run dataset (details of different components of the simulation in the text); (**b**) Overview of CHE nature run model output and strategy for comparison with different types of observations of carbon tracers and other relevant datasets such as lower resolution simulations. The differences between the CHE nature run and the various observations can be used to estimate and shed light into the different sources of uncertainty (orange boxes).

tracers at the beginning of the simulation[24,25] and the biogenic flux model[26,27]. An estimation of these uncertainties is provided in the Technical Validation section.

Example: Using tagged tracers to characterise anthropogenic plumes over land and ocean. In order to monitor anthropogenic CO_2 emissions, it is crucial to observe the CO_2 plumes emanating from the emission sources. These observations need to be based either targeted field campaign observations[13] or on high resolution imaging satellites[10]. As satellites have different viewing geometries over land and ocean[16], it is very important to understand how many of these plumes are located over land, ocean and coastal regions. Moreover, satellite observations only provide total column CO_2 over cloud-free regions. Table 4 provides an example of statistics on the proportion of anthropogenic plumes accumulated over a 24-hour period over land/ocean and the proportion

Model components and emission datasets	Source	Horizontal and temporal resolution	References
Tracer advection	*IFS semi-Lagrangian scheme*	*Model resolution and time step*	33,34,91,92
Tracer convective transport	*IFS Tiedtke scheme*	*Model resolution and time step*	93,94
tracer turbulent mixing	*IFS boundary layer scheme*	*Model resolution and time step*	95–97
CO_2 biogenic fluxes	*IFS CTESSEL A-gs with bias correction*	*Model resolution and time step*	26,27
CO_2 anthropogenic emissions	*EDGARv4.5FT2015 annual emissions and EDGARv4.2FT2010 monthly scaling factors CAMS-GLOB-TEMPO daily scaling factors for residential heating, CAMS-GLOB-AIR monthly emissions from aviation*	*$0.1^o \times 0.1^o$, monthly, daily (residential sector)*	29–31
CO_2 ocean fluxes	*Jena CarboScope v16*	*$2.0^o \times 2.5^o$, daily fluxes averaged to monthly mean fluxes*	32
CO_2, CH_4, CO fire emissions	*GFAS v1.2*	*$0.1^o \times 0.1^o$, daily*	63
CH_4 wetland fluxes	*LPJ-HYMN climatology (1990-2008)*	*$1^o \times 1^o$,monthly*	98
CH_4 anthropogenic emissions	*CAMS-GLOB-ANT v2.1 (based on EDGAR4.3.2 in 2012 and EDGARv4.2FT2010 seasonal cycle for 2010).*	*$0.1^o \times 0.1^o$, monthly*	99,100
CH_4 other fluxes	*Termites, wild animals, ocean fluxes and soil sink*	*$1^o \times 1^o$,monthly*	101–104
CO emissions	*CAMS-GLOB-ANT v2.1*	*$0.1^o \times 0.1^o$, monthly*	99,100
CO chemistry	*Linear CO chemistry scheme*	*Model resolution and time step*	105
CH_4 chemical loss rate	*Climatological loss rate*	*$6^o \times 4^o$, monthly*	106

Table 1. Model components with emission datasets used as boundary conditions in the nature run simulation and prescribed atmospheric chemical sources/sinks. Model resolution is around 9 km and model time step is 7.5 minutes.

Parameters types and levels	Archived time step	Range of data volume per parameter in GBytes
Model levels parameters from level 1 (model top) to level 137 (model bottom)*	3-hourly	9–**7,373**
Pressure level parameters: 1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100 to 300 by 50, 400 to 700 by 100, 850, 925, 950, 1000 hPa	3-hourly	219–**1,241**
Parameters on surface layers	3-hourly	~146
1: 0–7 cm, 2: 7–21 cm, 3: 21–72 cm, 4: 72 cm–1.82 m		
2D surface fields	3-hourly	~55
2D prescribed daily emissions	daily	~7
2D prescribed monthly emissions	monthly	~0.24
3D prescribed monthly aviation emissions on model levels from level 1 (model top) to level 137 (model bottom)	monthly	~33

Table 2. Content of CHE nature run dataset with different parameter types and their associated data volume for the full year. *Model levels can be converted to pressure levels p with the following equation $p_i = p_{sfc}B_i + A_i$ [in Pa] where p_{sfc} is surface pressure and A_i and B_i are static coefficients defined for each model level i (https://confluence.ecmwf.int/display/UDOC/L137+model+level+definitions). The volume of atmospheric-tracer parameters has been highlighted in bold. The individual parameters are listed in Supplementary Information file 1 (Table S1).

of plumes under cloudy conditions for January and July 2015. These fossil fuel tagged tracers and other tagged tracers associated with the biogenic fluxes, ocean fluxes and biomass burning emissions are all included in the CHE nature run dataset (see Table 3).

Example: Insights into total column variability. The CO_2, CH_4 and CO observing system is based on *in situ* observations, at the surface or from tall towers, and remote sensing observations from ground-based stations or satellites providing partial/total column observations. There are currently very few vertical profile observations from aircrafts[37,38] and Aircore measurements[36,39] that can be used to link the two observation types. For low-resolution transport models assimilating both surface and total column observations in an atmospheric inversions framework, it can sometimes be challenging to combine the surface and total column variability for various reasons. These include errors in the remote sensing observations[16], representation errors near the surface and model transport errors associated with vertical mixing[40], atmospheric chemistry[41], as well as long-range transport[42] and the impact of stratospheric intrusions[43]. The global nature run can be useful to characterize the column variability of carbon tracers[44] associated with transport. Figure 2 illustrates the potential use of the CHE nature run to explain the variability of XCO_2, XCH_4 and XCO at 24 TCCON sites (https://tccondata.org). The coefficient of determination shows that the variance of the total column can be explained by the different layers in the column in the nature run. When the column is well mixed, the contribution from the different layers is similar. At the sites where the influence of local emissions or natural fluxes is strong, the layers near the surface dominate the variability. Long-range transport in the free troposphere and upper troposphere/lower stratosphere

Tagged tracers	Parameter type	Parameter ID	Units	Enhancement processing (parameter IDs)
CO_2 tracer	3D (model and pressure levels)	12.212	kg kg^{-1}	
CO_2 tracer without fire emissions	3D (model and pressure levels)	13.212	kg kg^{-1}	3D Biomass burning (12.212-13.212)
CO_2 tracer without anthropogenic emissions	3D (model and pressure levels)	14.212	kg kg^{-1}	3D Anthropogenic (12.212-14.212)
CO_2 tracer without biogenic fluxes	3D (model and pressure levels)	15.212	kg kg^{-1}	3D Biogenic (12.212-15.212)
CO_2 tracer without ocean fluxes	3D (model and pressure levels)	16.212	kg kg^{-1}	3D Ocean (12.212-16.212)
Total-column CO_2 tracer	2D (surface level)	112.212	kg m^{-2}	
Total-column CO_2 tracer without fire emissions	2D (surface level)	113.212	kg m^{-2}	Column Biomass burning (112.212-113.212)***
Total-column CO_2 tracer without anthropogenic emissions	2D (surface level)	114.212	kg m^{-2}	Column Anthropogenic (112.212-114.212)***
Total-column CO_2 tracer without biogenic emissions	2D (surface level)	115.212	kg m^{-2}	Column Biogenic (112.212-115.212)***
Total-column CO_2 tracer without ocean fluxes	2D (surface level)	116.212	kg m^{-2}	Column Ocean (112.212-116.212)***

Table 3. List of experimental CO_2 tagged tracers from the CHE nature run dataset. Each tracer is identified with a given experimental parameter ID. ***Note that the units of tagged tracers for the total column need to be converted from kg m^{-2} to ppm as described in 2D Atmospheric Composition parameters.

	XCO_2_FF > 0.25 ppm Number model cells % model cells (Number clear-sky model cells) (% clear-sky model cells)			**XCO_2_FF > 0.50 ppm Number model cells % model cells (Number clear-sky model cells) (% clear-sky model cells)**		
	Land	**Coast**	**Ocean**	**Land**	**Coast**	**Ocean**
January	36,533 +/−2,458	41,018 +/−1924	15,689 +/−3,031	14,933 +/−1,312	18,178 +/−1,243	5,444 +/−1475
	0.55 +/−0.04	**0.62 +/−0.03**	**0.24 +/−0.05**	**0.23 +/−0.02**	**0.28 +/−0.02**	**0.08 +/−0.02**
	(11,194 +/−2,468)	(10,809 +/−3,101)	(3745 +/−1,466)	(4,995 +/−1,599)	(5,096 +/−1,915)	(1,471 +/−853)
	(0.17 +/−0.04)	**(0.16 +/−0.05)**	**(0.06 +/−0.02)**	**(0.08 +/−0.02)**	**(0.08 +/−0.03)**	**(0.02 +/−0.01)**
July	24,352 +/−1,235	28,203 +/−867	9,107 +/−2212	8,603 +/−583	11,325 +/−610	2,901 +/−954
	0.37 +/−0.02	**0.43 +/−0.01**	**0.14 +/−0.03**	**0.13 +/−0.01**	**0.17 +/−0.01**	**0.04 +/−0.01**
	(6,314 +/−1,288)	(5,746 +/−1,556)	(2,181 +/−1169)	(2,238 +/−613)	(2,289 +/−768)	(732 +/−554)
	(0.10 +/−0.02)	**(0.09 +/−0.02)**	**(0.03 +/−0.02)**	**(0.03 +/−0.01)**	**(0.03 +/−0.01)**	**(0.01 +/−0.01)**

Table 4. Distribution of XCO_2 anthropogenic enhancement (XCO_2_FF) accumulated over a 24-hour period from the CHE global nature run as mean number (and percentage in bold) of model cells with XCO_2_FF > 0.25 ppm (left columns) and XCO_2_FF > 0.50ppm (right columns). The variability with respect to the mean number is shown by the +/− standard deviation. The statistics are also provided for clear-sky conditions, land, ocean and coastal regions, as these considerations are all relevant for satellite observations. Clear-sky model cells are defined with a cloud fraction threshold less than 10% over the 9 km × 9 km model cell; land cells have more than 99% land; ocean cells have less than 1% land and model cells over the coast have land between 1 and 99%.

also plays an important role, as depicted by the green/orange bars with higher r^2 values than the near-surface layers in purple/red. The dataset can also be used to assess the important contribution of the stratosphere in the variability of XCH_4[45].

Data Records

The CHE nature run dataset can be accessed through the ECMWF API following the examples provided in[46]. The data can be extracted on the native octahedral grid with the original resolution (tco1279, corresponding to approximately 9 km) or on a regular latitude/longitude grid at the required resolution of the user. Both grib and NetCDF formats are available. The dataset extends from 26 December 2014 to 31 December 2015. The list of contents is provided in Table 2. All meteorological and tracer fields and surface fluxes have been archived with 3-hourly time steps with respect to the 00 UTC initialization of the weather forecast. Step 0 of all the meteorological parameters represents the initial conditions taken from ERA5[23]. Atmospheric species (CO_2, CH_4 and CO) at step 0 are equivalent to tracers from the previous day at step 24, because they are free-running from one 1-day forecast to the next as illustrated in Fig. 1. Note that the emissions of CO and the CO_2 emissions from aviation are not stored in the CHE nature run dataset, but they can be obtained from the Copernicus Atmosphere Data Store (https://ads.atmosphere.copernicus.eu).

Technical Validation

The dataset is based on the state-of-the-art operational NWP and CAMS forecasting system[21,47] which has been proven to produce reliable and accurate atmospheric CO_2, CH_4 and CO variability[13–15]. The CHE nature run focuses on 2015, a year characterised by a pronounced decrease in the terrestrial carbon sink associated with the

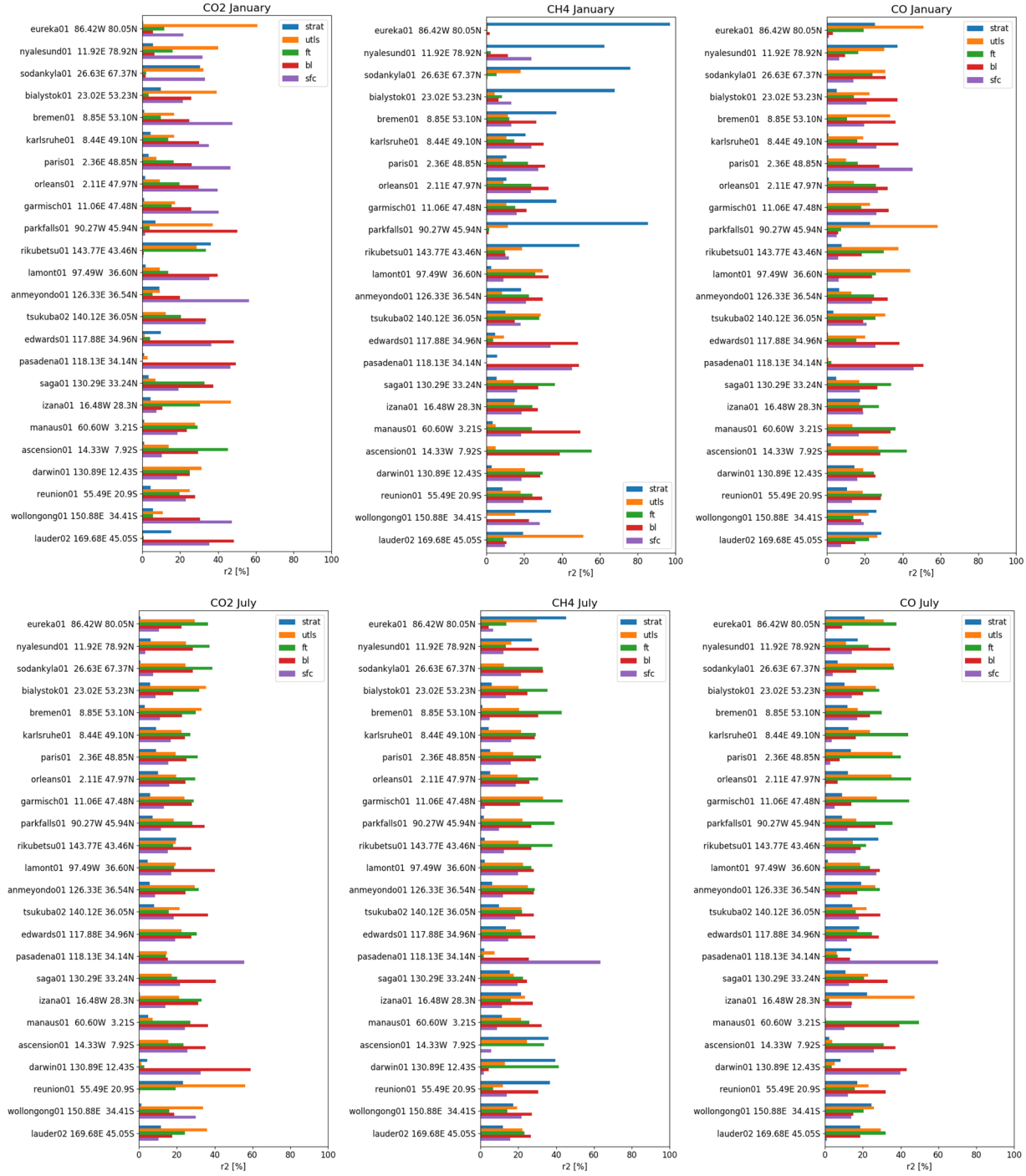

Fig. 2 Coefficient of determination (r^2) [%] of CO_2, CH_4 and CO total column with different partial layers in the atmospheric column in January and July 2015 at 24 TCCON sites (tccon.org). The atmospheric layers are defined as follows: from surface to 400 m (SFC), from 400 to 2 km (BL), from 2 km to 5 km (FT), from 5 km to 10 km (UTLS), from 10 km to the top of atmosphere (STRAT). All the column and partial column data have been detrended before calculating the coefficient of determination. All r^2 values shown are statistically significant with p-value < 0.01 except when the $r^2 < 0.001$.

strong El Niño Southern Oscillation (ENSO) of 2015-2016[48] with droughts[49], as well as fires in several regions, particularly over the tropics[50]. The larger than normal CO_2 atmospheric growth rate in 2015[48,51] and anomalously high fire emissions are well captured by the CHE nature run with a total global annual flux of 6.60 GtC (equivalent to 3.16 ppm/year from 1 January 2015 to 31 December 2015), which is close to the NOAA estimate of $2.99 +/-0.07$ ppm/year (https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html). The CO_2 components of the budget include 9.29 GtC of anthropogenic emissions, 2.09 GtC of fire emissions, 2.10 GtC ocean sink and 2.69 GtC sink from land ecosystems. These values are consistent with the global carbon budget estimates[52].

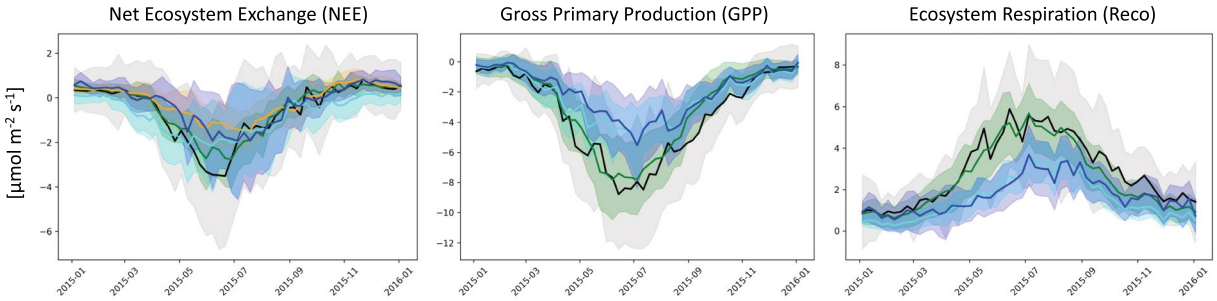

Fig. 3 Mean seasonal cycle of CO_2 biogenic fluxes [µmol m^{-2} s^{-1}] at the 25 Eddy Covariance sites. FLUXNET2015[58] observations [ICOS 2018 drought dataset[53]] are shown in black; the IFS modelled fluxes in cyan and the bias corrected fluxes used in the CHE nature run in blue; the CAMS inversion product[56,57,65] (total flux –anthropogenic emissions) based on surface observations is shown in orange; and the CHE FLUXCOM product[54,55] in green. The shading depicts the standard deviation across the 25 sites.

Example: Evaluation of CO_2 sources/sink by vegetation. Biogenic CO_2 fluxes associated with vegetation over land can dominate atmospheric CO_2 variability on a wide range of time scales from diurnal, synoptic, seasonal to inter-annual[28]. They are a crucial component for the estimation of the background CO_2 underlying the fossil fuel plumes from emission hotspots. This background CO_2 has not been directly influenced by the plumes emanating from local anthropogenic sources, but it results from the larger-scale fluxes associated with biogenic sources and sinks over land. The European Eddy Covariance (EC) ecosystem flux data collected and processed by the Integrated Carbon Observation System (ICOS)[53] are used to evaluate the uncertainty of modelled biogenic fluxes in the IFS (Fig. 3) which are bias-corrected[27] in the CHE nature run. These modelled fluxes are also compared to other flux products, such as FLUXCOM[54,55] (extended to include varying diurnal meteorology from ERA5) and the CAMS CO_2 inversion (v18r3) product[56,57]. The EC data were processed and the Gross Primary Production (GPP) and ecosystem respiration (Reco) estimated using the standard methods applied in FLUXNET[58] using the observed Net Ecosystem Exchange (NEE). Fig. 3 shows an overall underestimation of the seasonal cycle of NEE, GPP and Reco at the EC sites with typical errors of around 2 µmol m^{-2} s^{-1}. Synoptic-scale errors are smaller while the diurnal cycle has larger errors of around 4 µmol m^{-2} s^{-1} (not shown in Fig. 3). This underestimation is exacerbated by the anomalously high NEE and Reco observed during the European drought in 2015 (Fig. SB7.3[49]). This type of evaluation can be used to understand the source of biogenic flux errors and improve the underlying biogenic models, as well as to quantify the uncertainty of prior fluxes for atmospheric inversions[59].

Example: Simulation and observation mismatch in the total column of CO_2, CH_4 and CO. The TCCON data[60] which is widely used as a reference to evaluate biases in global measurement of CO_2, CH_4 and CO total column averages–referred to as XCO_2, XCH_4 and XCO–from space[16] is used here to assess the inter-hemispheric gradient, seasonal cycle and synoptic day-to-day variability in the nature run dataset (Fig. 4). The large-scale patterns of variability on a monthly scale are generally well represented for the three species. The amplitude of the XCO_2 seasonal cycle is underestimated at most TCCON sites, with the summer trough being 1 to 3 ppm higher than observed. This is consistent with the general underestimation of the biogenic sink during the growing season shown in Fig. 3. XCH_4 is overestimated in spring/summer and underestimated in autumn/winter, due to errors in the seasonality of the chemical sink and emissions (e.g. wetlands, agriculture and biomass burning). XCO is underestimated in winter which is a common feature in many models and emission data sets[61] and overestimated in summer/autumn, often caused by the biogenic emissions of isoprene, which have a large impact on southern hemisphere and global background values[62] of CO. Other sources of error are associated with the chemical sources/sinks[61] and fire emissions[63], as 2015 was an extreme year for CO because of Indonesian fires in autumn[64]. Part of the bias shown in Fig. 4 also comes from the CO_2, CH_4 and CO initial conditions at the start of the nature run extracted from the CAMS re-analysis[24,25]. The random error in the sub-monthly variability (STDE in Fig. 4) - associated with surface fluxes/emissions and atmospheric transport - is generally below 1.5 ppm for XCO_2, 10 ppb for XCH_4 and 10 ppb for XCO, except at urban sites near emission hotspots such as Pasadena, Tsukuba and Paris.

Example: Fine-scale structure in vertical profiles. The vertical profiles of CO_2, CH_4 and CO are illustrated in Fig. 5 with a comparison to AirCore observations[36,39] from the National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory and the lower-resolution CAMS surface *in situ* inversion dataset[57,65,66]. While most global transport models used in atmospheric inversion systems have too coarse horizontal and vertical resolution to be able to represent the fine-scale vertical structure, the CHE nature run is able to capture the small-scale anomalies along the atmospheric column from the surface up to the lower stratosphere (50 hPa). The profiles on three different consecutive days show the large variability associated with day-to-day synoptic transport, particularly for CO_2. Capturing this type of vertical variability is important because it reflects the ability of atmospheric transport models to represent vertical mixing and long-range transport. Both need

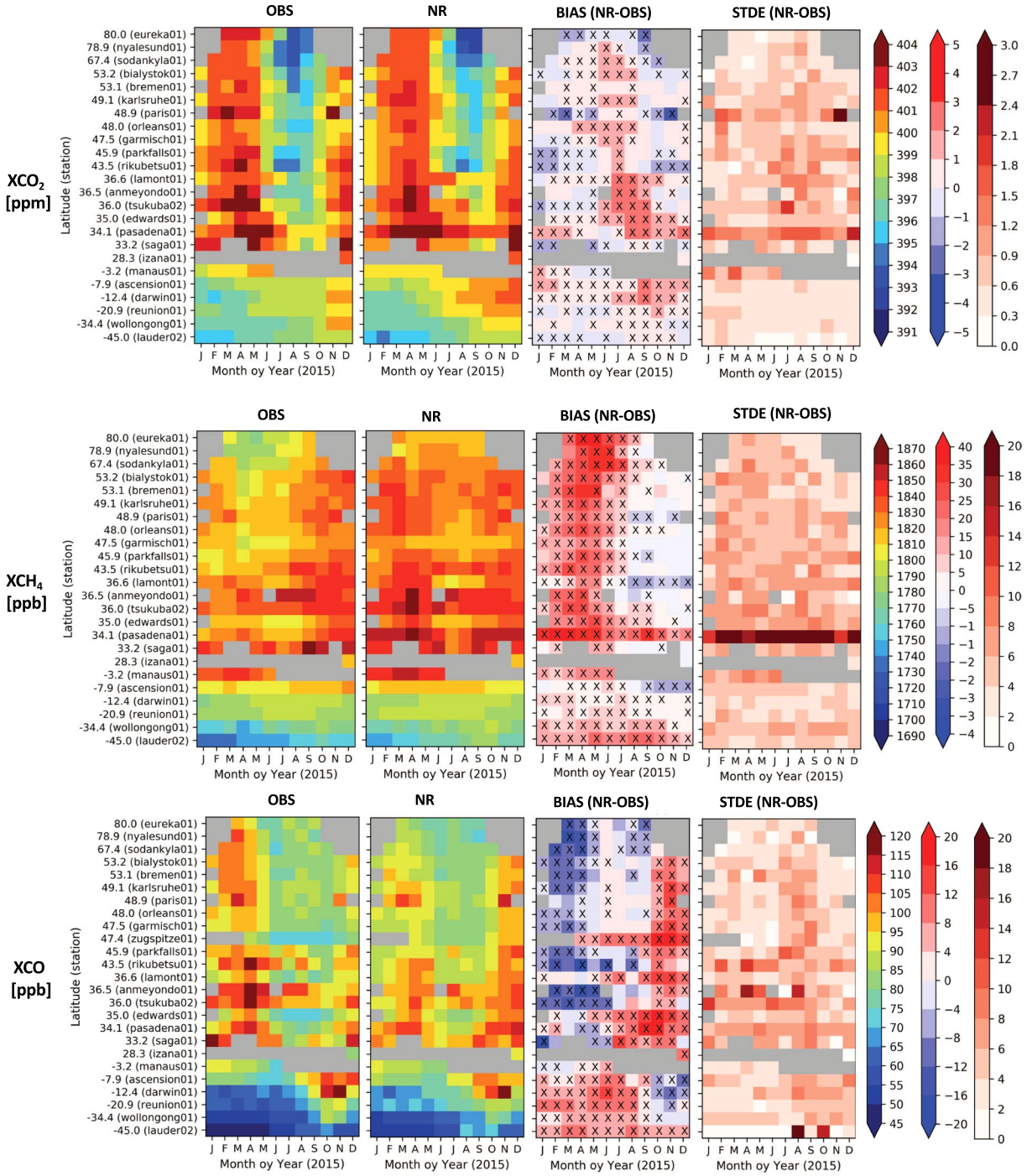

Fig. 4 Evaluation of XCO_2, XCH_4 and XCO from the CHE nature run (NR). The nature run is compared to total column FTIR observation[35,60] at the TCCON stations[67–90] (OBS). The crosses indicate that the bias is statistically significant (p-value < 0.01).

to be accurately represented in atmospheric inversions in order to accurately infer surface fluxes. Examples of anticorrelation between the near-surface CO_2 and XCO_2 are also shown in Fig. 5j (e.g. 7, 9, 15, 20, 21 and 24 June) which are associated with the advection of anomalously high/low CO_2 air in the free troposphere (above 700 hPa) and the opposite decrease/increase of CO_2 near the surface. This emphasizes the importance of tracer transport above the planetary boundary layer in explaining the variability of XCO_2 also shown in Fig. 2.

Code availability

The IFS forecast model and the Meteorological Archival and Retrieval System (MARS) software are not available for public use as the ECMWF Member States are the proprietary owners. However, the CHE global nature run

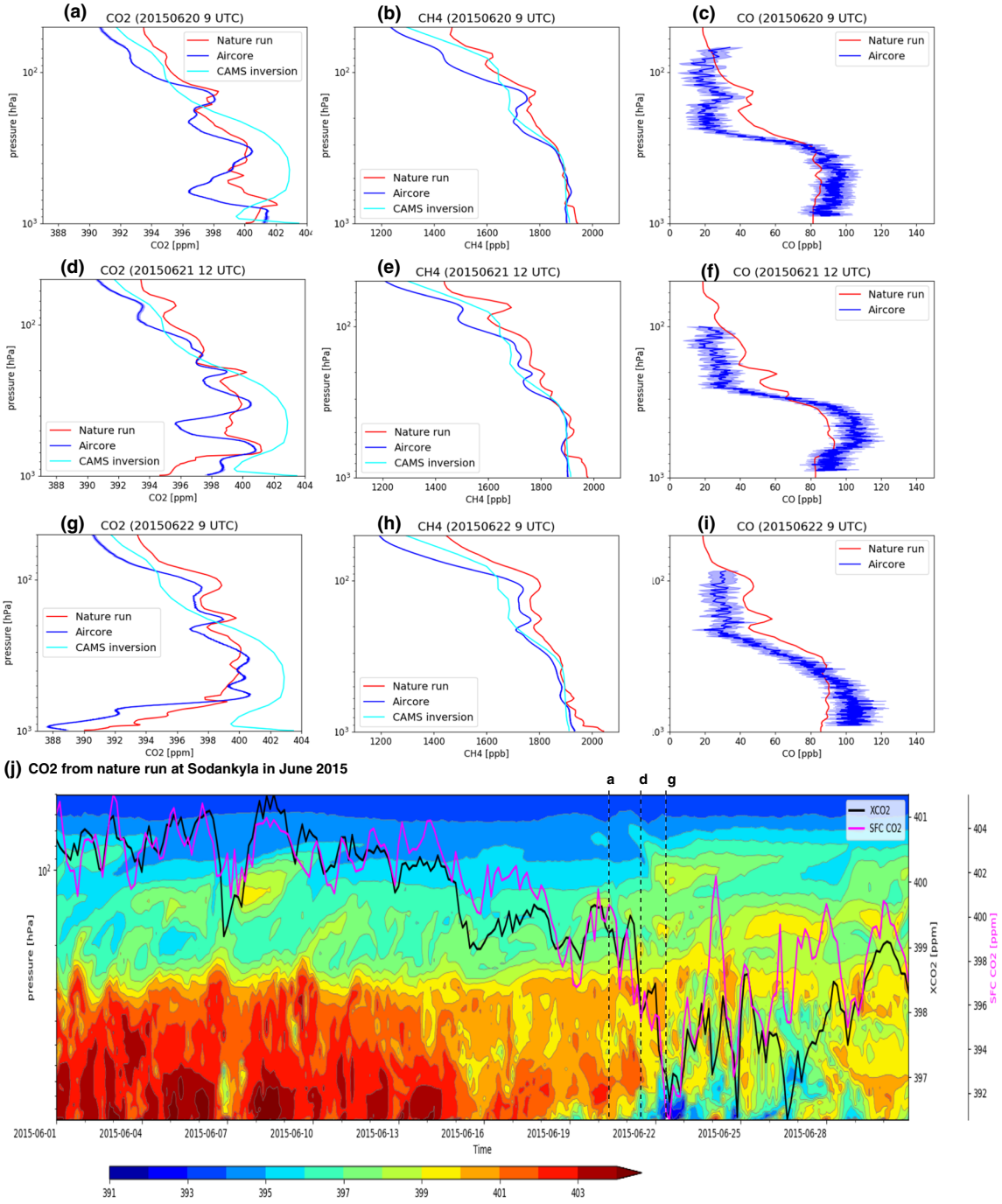

Fig. 5 Examples of CO_2 CH_4 and CO vertical profiles from the CHE nature run at Sodankylä (67.37°N, 26.63°E). The nature run is compared to NOAA AirCore (v20201223) observations[36,39] and the CAMS CO_2 and CH_4 inversion[57,65,66] (**a**–**i**) during three days in June, depicted by the dashed lines in (**j**) where the nature run hovmöller plot for CO_2 shows the temporal variability of the vertical profile at Sodankylä over the whole month of June. The solid black and magenta lines show the time series of XCO_2 and near-surface CO_2 averaged over the model levels from the surface to 400 m above the surface (SFC CO2) respectively.

dataset and the MARS data extraction features are freely available through ECMWF API (https://www.ecmwf.int/en/forecasts/access-forecasts/ecmwf-web-api) following a registration step (https://apps.ecmwf.int/registration/). The data can be accessed using python (https://www.python.org). The commands and steps required are detailed in the Supplementary Information file 1 (S2).

Received: 26 March 2021; Accepted: 21 February 2022;
Published online: 11 April 2022

References

1. European Commission, Joint Research Centre, Ciais, P. *et al*. Towards a European Operational Observing System to Monitor Fossil CO2 emission. *European Commission Joint Research Centre Publication Office*, 65 pp., https://doi.org/10.2788/350433 (2016).
2. Pinty, B. *et al*. An operational anthropogenic CO2 emissions monitoring & verification system: baseline requirements, model components and functional architecture. *European Commission Joint Research Centre Publications Office*, EUR 28736 EN, 98 pp., https://doi.org/10.2760/08644 (2017).
3. Janssens-Maenhout, G. *et al*. Towards an operational anthropogenic CO 2 emissions monitoring and verification support capacity. *Bull. Am. Meteorol. Soc.* **101**(8), E1439–E1451, https://doi.org/10.1175/BAMS-D-19-0017.1 (2020).
4. Crisp, D. & CEOS Atmospheric Composition Virtual Constellation Greenhouse Gas Team. *A constellation architecture for monitoring carbon dioxide and methane from space*. https://ceos.org/document_management/Meetings/Plenary/32/documents/CEOS_AC-VC_White_Paper_Version_1_20181009.pdf (2018).
5. Pinty, B. *et al*. An operational anthropogenic CO_2 emissions monitoring & verification support capacity: needs and high level requirements for *in situ* measurements. *European Commission Joint Research Centre*, EUR 29817 EN, 72 pp., https://doi.org/10.2760/182790 (2019).
6. Haussaire, J.-M. *Model systems and simulation configurations*. (2018).
7. Gelaro, R. *et al*. Evaluation of the 7-km GEOS-5 Nature Run. *Technical Report Series on Global Modeling and Data Assimilation* **36**, 20150011486 https://ntrs.nasa.gov/api/citations/20150011486/downloads/20150011486.pdf (2015).
8. Liu, J. *et al*. Carbon monitoring system flux estimation and attribution: Impact of ACOS-GOSAT XCO2 sampling on the inference of terrestrial biospheric sources and sinks. *Tellus Ser. B Chem. Phys. Meteorol.* **66**, 1 (2014).
9. Kaminski, T. *et al*. Assimilation of atmospheric CO 2 observations from space can support national CO 2 emission inventories. *Environ. Res. Lett.* **17**, 1, https://doi.org/10.1088/1748-9326/ac3cea (2021).
10. Lespinas, F. *et al*. The potential of a constellation of low earth orbit satellite imagers to monitor worldwide fossil fuel CO2emissions from large cities and point sources. *Carbon Balance Manag.* **15**, 18 (2020).
11. Balsamo, G. *et al*. The CO2 Human Emissions (CHE) Project: First Steps Towards a European Operational Capacity to Monitor Anthropogenic CO2 Emissions. *Front. Remote Sens.* **2**, https://doi.org/10.3389/frsen.2021.707247 (2021).
12. Kuhlmann, G., Brunner, D., Broquet, G. & Meijer, Y. Quantifying CO2 emissions of a city with the Copernicus Anthropogenic CO2 Monitoring satellite mission. *Atmospheric Meas. Tech.* **13**, 6733–6754 (2020).
13. Tang, W. *et al*. Evaluating high-resolution forecasts of atmospheric CO and CO2 from a global prediction system during KORUS-AQ field campaign. *Atmospheric Chem. Phys.* **18**, 11007–11030 (2018).
14. Agustí-Panareda, A. *et al*. Modelling CO2 weather-why horizontal resolution matters. **19**, 7347–7376 https://doi.org/10.5194/acp-19-7347-2019 (2019).
15. Gałkowski, M. *et al*. *In situ* observations of greenhouse gases over Europe during the CoMet 1.0 campaign aboard the HALO aircraft. *Atmospheric Meas. Tech.* **14**, 1525–1544 (2021).
16. O'Dell, C. W. *et al*. Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmospheric Meas. Tech.* **11**, 6539–6576 (2018).
17. Feng, S. *et al*. Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in Atmospheric CO_2 Simulations Over North America. *J. Geophys. Res. Atmospheres* **124**, 14325–14346 (2019).
18. McNorton, J. R. *et al*. Representing model uncertainty for global atmospheric CO2 flux inversions using ECMWF-IFS-46R1. *Geosci. Model Dev.* **13**, 2297–2313 (2020).
19. Chevallier, F. *et al*. Local anomalies in the column-averaged dry air mole fractions of carbon dioxide 1 across the globe during the first months of the coronavirus recession 2. *Geophysical Research Letters* **47**, 1–9 (2020).
20. Barré, J. *et al*. Systematic detection of local CH4 anomalies by combining satellite measurements with high-resolution forecasts. *Atmospheric Chem. Phys.* **21**, 5117–5136 (2021).
21. Haiden, T. *et al*. Evaluation of ECMWF forecasts, including the 2019 upgrade. *ECMWF Technical Memorandum*, **853**, https://www.ecmwf.int/en/elibrary/19277-evaluation-ecmwf-forecasts-including-2019-upgrade (2019).
22. Malardel, S. *et al*. A new grid for the IFS. *ECMWF Newsletter* **146**, 23–28, https://www.ecmwf.int/en/elibrary/17262-new-grid-ifs (2016).
23. Hersbach, H. *et al*. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
24. Inness, A. *et al*. The CAMS reanalysis of atmospheric composition. *Atmospheric Chem. Phys.* **19**, 3515–3556 (2019).
25. Ramonet, M., Langerock, B., Warneke, T. & Eskes, H. J. Validation report for the CAMS greenhouse gas global reanalysis, years 2003 - 2016. *Copernicus Atmosphere Monitoring Service (CAMS) report*, https://doi.org/10.24380/Y034-7672 (2021)
26. Boussetta, S. *et al*. Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: Implementation and offline validation. *J. Geophys. Res. Atmospheres* **118**, 5923–5946 (2013).
27. Agustí-Panareda, A. *et al*. A biogenic CO2 flux adjustment scheme for the mitigation of large-scale biases in global atmospheric CO2 analyses and forecasts. *Atmospheric Chem. Phys.* **16**, 10399–10418, https://doi.org/10.5194/acp-16-10399-2016 (2016).
28. Agustí-Panareda, A. *et al*. Forecasting global atmospheric CO2. *Atmospheric Chem. Phys.* **14**, 11959–11983 (2014).
29. Janssens-Maenhout, G. *et al*. EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970-2012. *Earth Syst. Sci. Data Discuss*. 1–55 https://doi.org/10.5194/essd-2017-79 (2017).
30. Choulga, M. *et al*. Global anthropogenic CO2 emissions and uncertainties as a prior for Earth system modelling and data assimilation. *Earth Syst. Sci. Data* **13**, 5311–5335 (2021).
31. Guevara, M. *et al*. Copernicus Atmosphere Monitoring Service TEMPOral profiles (CAMS-TEMPO): global and European emission temporal profile maps for atmospheric chemistry modelling. *Earth Syst. Sci. Data* **13**, 367–404 (2021).
32. Rödenbeck, C. *et al*. Global surface-ocean pCO2 and sea-Air CO2 flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Sci.* **9**, 193–216 (2013).
33. Agusti-Panareda, A., Diamantakis, M., Bayona, V., Klappenbach, F. & Butz, A. Improving the inter-hemispheric gradient of total column atmospheric CO 2 and CH 4 in simulations with the ECMWF semi-Lagrangian atmospheric global model. *Geosci. Model Dev.* **10**, 1–18 (2017).
34. Diamantakis, M. & Agusti-Panareda, A. *819 A positive definite tracer mass fixer for high resolution weather and atmospheric composition forecasts*. http://www.ecmwf.int/en/research/publications (2017).
35. Wunch, D. *et al*. The total carbon column observing network. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **369**, 2087–2112 (2011).
36. Tans, P. P. System and method for providing vertical profile measurements of atmospheric gases. (2009).
37. Filges, A. *et al*. The IAGOS-core greenhouse gas package: a measurement system for continuous airborne observations of CO_2, CH_4, H_2 O and CO. *Tellus B Chem. Phys. Meteorol.* **67**, 27989 (2015).
38. Umezawa, T. *et al*. Statistical characterization of urban CO2 emission signals observed by commercial airliner measurements. *Sci. Rep.* **10**, 7963 (2020).
39. Karion, A., Sweeney, C., Tans, P. & Newberger, T. AirCore: An Innovative Atmospheric Sampling System. 1839–1853 (2010).
40. Yang, Z. *et al*. New constraints on Northern Hemisphere growing season net flux. *Geophys. Res. Lett.* **34**, L12807 (2007).
41. Rigby, M. *et al*. Role of atmospheric oxidation in recent methane growth. *Proc. Natl. Acad. Sci. USA* **114**, 5373–5377 (2017).
42. Eastham, S. D. & Jacob, D. J. Limits on the ability of global Eulerian models to resolve intercontinental transport of chemical plumes. *Atmospheric Chem. Phys.* **17**, 2543–2553, https://doi.org/10.5194/acp-17-2543-2017 (2017).

43. Ott, L. E. *et al*. Frequency and impact of summertime stratospheric intrusions over Maryland during DISCOVER-AQ (2011): New evidence from NASA's GEOS-5 simulations. *J. Geophys. Res.* **121**, 3687–3706 (2016).
44. Keppel-Aleks, G., Wennberg, P. O. & Schneider, T. Sources of variations in total column carbon dioxide. *Atmospheric Chem. Phys.* **11**, 3581–3593 (2011).
45. Verma, S. *et al*. Extending methane profiles from aircraft into the stratosphere for satellite total column validation using the ECMWF C-IFS and TOMCAT/SLIMCAT 3-D model. *Atmospheric Chem. Phys.* **17**, 6663–6678, https://doi.org/10.5194/acp-17-6663-2017 (2017).
46. The CO2 Human Emissions (CHE) global nature run. *ECMWF* https://doi.org/10.21957/w4wq-sd03 (2021).
47. Wagner, A. *et al*. Validation report of the CAMS near-real-time global atmospheric composition service: Period September–November 2019, *Copernicus Atmosphere Monitoring Service (CAMS) report*, https://doi.org/10.24380/XZKK-BZ05 (2019).
48. Bastos, A. *et al*. Impact of the 2015/2016 El Niño on the terrestrial carbon cycle constrained by bottom-up and top-down approaches. *Philos. Trans. R. Soc. B Biol. Sci.* **373** (2018).
49. Dong, B., Sutton, R., Shaffrey, L. & Wilcox, L. The 2015 European heat wave. *Bull. Am. Meteorol. Soc.* **97**, S57–S62 (2016).
50. Huijnen, V. *et al*. Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. *Sci. Rep.* **6**, 26886 (2016).
51. Patra, P. K. *et al*. The Orbiting Carbon Observatory (OCO-2) tracks 2-3 peta-gram increase in carbon release to the atmosphere during the 2014-2016 El Niño. *Sci. Rep.* **7**, 13567, https://doi.org/10.1038/s41598-017-13459-0 (2017).
52. Friedlingstein, P. *et al*. Global Carbon Budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
53. Drought 2018 Team. Drought-2018 ecosystem eddy covariance flux product in FLUXNET-Archive format - release 2019-1 (Version 1.0). *ICOS Carbon Portal* https://doi.org/10.18160/YVR0-4898 (2019).
54. Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D. & Reichstein, M. Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product. *Earth Syst. Sci. Data* **10**, 1327–1365 (2018).
55. Jung, M. *et al*. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* **17**, 1343–1365 (2020).
56. Chevallier, F. *et al*. CO2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements. *J. Geophys. Res. Atmospheres* **115**, 1–17 (2010).
57. CAMS global inversion-optimised greenhouse gas fluxes and concentrations. *Copernicus Atmosphere Data Store* https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-greenhouse-gas-inversion?tab=overview (2020).
58. Pastorello, G., Trotta, C., Canfora, E., Housen, C. & Christianson, D. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Nat. Sci. Data* **7**, 27 (2020).
59. Chevallier, F. *et al*. What eddy-covariance measurements tell us about prior land flux errors in CO 2 -flux inversion schemes. *Glob. Biogeochem. Cycles* **26** (1), GB1021, JRC68617 (2012).
60. Total Carbon Column Observing Network (TCCON) Team. 2014 TCCON Data Release (Version GGG2014). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014 (2017).
61. Stein, O. *et al*. On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations. *Atmospheric Chem. Phys.* **14**, 9295–9316 (2014).
62. Flemming, J. *et al*. Tropospheric chemistry in the integrated forecasting system of ECMWF. *Geosci. Model Dev.* **8**, 975–1003 (2015).
63. Kaiser, J. W. *et al*. Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences* **9**, 527–554 (2012).
64. Blunden, J. & Arndt, D. S. State of the Climate in 2015. *Bull. Am. Meteorol. Soc.* **97**, Si–S275 (2016).
65. Chevallier, F. *Description of the CO2 inversion production chain 2020, Copernicus Atmosphere Monitoring Service*. https://atmosphere.copernicus.eu/sites/default/files/2020-06/CAMS73_2018SC2_%20D5.2.12020_202004_%20CO2%20inversion%20production%20chain_v1.pdf (2020).
66. Segers, A., Tokaya, J. & Houweling, S. *Description of the CH4 Inversion Production Chain*. https://atmosphere.copernicus.eu/sites/default/files/2021-01/CAMS73_2018SC3_D73.5.2.2-2020_202012_production_chain_Ver1.pdf (2020).
67. Strong, K. *et al*. TCCON data from Eureka (CA), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.EUREKA01.R0/1149271 (2014).
68. Notholt, J. *et al*. TCCON data from Ny Ålesund, Spitsbergen (NO), Release GGG2014.R1(Version R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.NYALESUND01.R1 (2019).
69. Kivi, R., Heikkinen, P. & Kyro, E. TCCON data from Sodankylä (FI), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.SODANKYLA01.R0/1149280 (2014).
70. Deutscher, N. M. *et al*. TCCON data from Bialystok (PL), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.BIALYSTOK01.R1/1183984 (2015).
71. Notholt, J. *et al*. TCCON data from Bremen (DE), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.BREMEN01.R0/1149275 (2014).
72. Hase, F., Blumenstock, T., Dohe, S., Gros, J. & Kiel, M. TCCON data from Karlsruhe (DE), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.KARLSRUHE01.R1/1182416 (2015).
73. Te, Y., Jeseck, P. & Janssen, C. TCCON data from Paris (FR), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.PARIS01.R0/1149279 (2014).
74. Wennberg, P. O. *et al*. TCCON data from Park Falls (US), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.PARKFALLS01.R1 (2017).
75. Warneke, T. *et al*. TCCON data from Orléans (FR), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.ORLEANS01.R0/1149276 (2014).
76. Sussmann, R. & Rettinger, M. TCCON data from Garmisch (DE), Release GGG2014.R2 (Version R2). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.GARMISCH01.R2 (2018).
77. Morino, I., Yokozeki, N., Matsuzaki, T., & Horikawa. TCCON data from Rikubetsu (JP), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.RIKUBETSU01.R1/1242265 (2016).
78. Wennberg, P. O. *et al*. TCCON data from Lamont (US), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.LAMONT01.R1/1255070 (2016).
79. Goo, T.-Y., Oh, Y.-S. & Velazco, V. A. TCCON data from Anmeyondo (KR), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.ANMEYONDO01.R0/1149284 (2014).
80. Morino, I., Matsuzaki, T. & Horikawa, M. TCCON data from Tsukuba (JP), 125HR, Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.TSUKUBA02.R1/1241486 (2016).
81. Iraci, L. T. *et al*. TCCON data from Edwards (US), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.EDWARDS01.R1/1255068 (2016).
82. Wennberg, P. O. *et al*. TCCON data from Caltech (US), Release GGG2014.R1 (Version GGG2014.R1). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.PASADENA01.R1/1182415 (2015).
83. Kawakami, S. *et al*. TCCON data from Saga (JP), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.SAGA01.R0/1149283 (2014).
84. Blumenstock, T., Hase, F., Schneider, M., Garcia, O. & Sepulveda, E. TCCON data from Izana (ES), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.IZANA01.R0/1149295 (2014).

85. Dubey, M. K. *et al*. TCCON data from Manaus (BR), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.MANAUS01.R0/1149274 (2014).
86. Feist, D. G., Arnold, J, & Geibel. TCCON data from Ascension Island (SH), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.ASCENSION01.R0/1149285 (2014).
87. Griffith, D. W. T. *et al*. TCCON data from Darwin (AU), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.DARWIN01.R0/1149290 (2014).
88. De Maziere, M. *et al*. TCCON data from Réunion Island (RE), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/tccon.ggg2014.reunion01.R0/1149288 (2014).
89. Griffith, D. W. T. *et al*. TCCON data from Wollongong (AU), Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/tccon.ggg2014.wollongong01.R0/1149291 (2014).
90. Sherlock, V. *et al*. TCCON data from Lauder (NZ), 125HR, Release GGG2014.R0 (Version GGG2014.R0). *CaltechData* https://doi.org/10.14291/TCCON.GGG2014.LAUDER02.R0/1149298 (2014).
91. Temperton, C., Hortal, M. & Simmons, A. A two-time-level semi-Lagrangian global spectral model. *Q. J. R. Meteorol. Soc.* **127**, 111–127 (2001).
92. Diamantakis, M. & Magnusson, L. Sensitivity of the ECMWF model to semi-Lagrangian departure point iterations. *Mon. Weather Rev.* **144**, 3233–3250 (2016).
93. Tiedtke, M. A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Mon. Weather Rev.* **117**, 1779–1800 (1989).
94. Bechtold, P. *et al*. Representing Equilibrium and Nonequilibrium Convection in Large-Scale Models. *J. Atmospheric Sci.* **71**, 734–753 (2013).
95. Beljaars, A. & Viterbo, P. *The role of the boundary layer in a numerical weather prediction model, in Clear and cloudy boundary layers*. (Royal Netherlands Academy of Arts and Sciences, North Holland Publishers, Amsterdam, 1998).
96. Koehler, M., Ahlgrimm, M. & Beljaars, A. Unified treatment of dry convective and stratocumulus-topped boundary layers in the ECMWF model. *Q. J. R. Meteorol. Soc.* **137**, 43–57 (2011).
97. Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T. & Balsamo, G. Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *J. Adv. Model. Earth Syst.* **5**, 117–133 (2013).
98. Spahni, R. *et al*. Constraining global methane emissions and uptake by ecosystems. *Biogeosciences* **8**, 1643–1665 (2011).
99. Granier, C., Elguindi, N. & Darras, S. *D81.2.2.3: CAMS emissions for all species for years 2000–2018, including documentation, CAMS_81–Global and Regional emissions*. https://atmosphere.copernicus.eu/sites/default/files/2019-11/05_CAMS81_2017SC1_D81.2.2.3-201808_v2_APPROVED_Ver2.pdf (2018).
100. Crippa, M. *et al*. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth Syst. Sci. Data* **10**, 1987–2013 (2018).
101. Sanderson, M. S. Biomass of termites and their emissions of methane and carbon dioxide: A global database. *Glob. Biogeochem. Cycles* **10**, 543–557 (1996).
102. Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J. & Heimann, M. Inverse modeling of methane sources and sinks using the adjoint of a global transport model. *J. Geophys. Res. Atmospheres* **104**, 26137–26160 (1999).
103. Lambert, G. R. & Schmidt, S. Reevaluation of the oceanic flux of methane: uncertainties and long term variations. *Chemosphere* **26**, 579–589 (1993).
104. Ridgwell, A. J., Marshall, S. J. & Gregson, K. Consumption of atmospheric methane by soils: A process-based model. *Glob. Biogeochem. Cycles* **13**, 59–70 (1999).
105. Claeyman, M. *et al*. A linear CO chemistry parameterization in a chemistry-transport model: Evaluation and application to data assimilation. *Atmospheric Chem. Phys.* **10**, 6097–6115 (2010).
106. Bergamaschi, P. *et al*. Inverse modeling of global and regional CH4 emissions using SCIAMACHY satellite retrievals. *J. Geophys. Res.* **114**, D22301 (2009).

Acknowledgements

The Copernicus Atmosphere Monitoring Service is operated by the European Centre for Medium-Range Weather Forecasts on behalf of the European Commission as part of the Copernicus Programme (http://copernicus.eu). The CHE and CoCO2 projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776186 and No 958927. We also thank the FLUXNET and TCCON PIs for providing the data used for the validation of the nature run dataset.

Author contributions

Anna Agustí-Panareda performed the model simulations, the data analysis and evaluation and wrote the manuscript. Joe MCNorton and Mark Parrington provided support with the processing of anthropogenic emissions. Michail Diamantakis provided support for the numerical aspects of the atmospheric transport in the model and insights on the interpretation of the tagged tracers. Margarita Choulga, Greet Janssens-Maenhout,Claire Granier, Marc Guevara, Hugo Denier van der Gon and Nellie Elguindi provided the CHE, CAMS-GLOB-ANT and CAMS-TEMPO anthropogenic emissions. Martin Jung and Sophia Walther provided the CHE FLUXCOM dataset and support with the interpretation of the evaluation of the modelled biogenic fluxes. Dario Papale and Alex Vermeulen provided advice on the FLUXNET data used to evaluate the modelled biogenic fluxes. Bianca C. Baier and Colm Sweeney provide the data and guidance for the use of NOAA Aircore profiles. Rigel Kivi contributed with launch of the Aircore data at Sodankylä. Johannes Flemming provided support for the interpretation of the CO simulations and their evaluation. Sébastien Massart contributed with the linear chemistry CO scheme used in the model simulations. Dominik Brunner and Jean-Matthieu Haussaire contributed towards the design of the simulations and the configuration of the model output. Gianpaolo Balsamo, Richard Engelen, Nicolas Bousserez, Souhail Boussetta and Frédéric Chevallier provided support with the writing of the manuscript. Miha Razinger provided technical support and advice for the publication of the dataset. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01228-2.

Correspondence and requests for materials should be addressed to A.A.-P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.