

CARBON ATMOSPHERIC TRACER RESEARCH TO IMPROVE NUMERICAL SCHEMES AND EVALUATION



CATRINE

Carbon Atmospheric Tracer
Research to Improve
Numerics and Evaluation

D7.2 Evaluation of preliminary global model intercomparison

Due date of deliverable	30/06/2025
Submission date	25/08/2025
File Name	CATRINE-D7.2-V1
Work Package /Task	WP 7 / Task 7.2
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Frédéric Chevallier, Anna Agustí-Panareda, Matthieu Gautherot, Joram Hooghiem, Maarten Krol, Wouter Peters and Stefan Versick
Revision number	3
Status	Final
Dissemination Level / location	Public



Funded by the
European Union

The CATRINE project (grant agreement No 101135000) is funded by the European Union.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

While Deliverable D7.1 “Design of protocol for preliminary global model intercomparisons” defined a framework for a high-resolution atmospheric transport model intercomparison exercise, the present deliverable D7.2 describes its implementation among the CATRINE partners and beyond, and presents an end-to-end intercomparison dataflow. Seven models have been used to generate the simulations for each required tracer, representing a total volume of about 10 terabytes (TB) of data. More models outside the CATRINE consortium will likely join this ambitious effort as a few teams have expressed their interest, in particular after milestone M8 “Workshop to discuss TransCom intercomparison protocol to assess transport from anthropogenic emission hotspots in global models”. A simple quality control procedure has been implemented that highlighted a series of errors in some submitted model simulations: some of the simulations have already been rerun while others are still being investigated. The quality-controlled data have been transferred to a unique facility for exploitation. The large volume of data involved in this exercise has represented a major technical challenge, from simulating the tracer with a sophisticated model on some high-performance computing facility to mining into the resulting data on another infrastructure. We are still working on the optimization of the transfer and of the post-processing steps, while first time series of the ensemble of model simulations vs. tracer measurements are available.

Table of Contents

1	Executive Summary	2
2	Introduction	4
2.1	Background	4
2.2	Scope of this deliverable	4
2.2.1	Objectives of this deliverables	4
2.2.2	Work performed in this deliverable	4
2.2.3	Deviations and counter measures	4
2.3	Project partners	5
3	Requested outputs	6
3.1	3D variables	6
3.2	2D variables	6
4	Additional diagnostics	6
5	Additional sensitivity tests	8
6	Models	11
6.1	Current participant list	11
6.2	Developments of the LMDz models	11
7	Quality control	14
7.1	Tracer global mass change	14
7.2	Zonal averages	14
8	Comparison to observations	17
9	Next steps	23
10	Conclusion	24
	References	25

2 Introduction

2.1 Background

Deliverable D7.1 defined an intercomparison exercise dedicated to global high-resolution tracer transport modelling. By “high-resolution”, we refer here to the equivalent of a grid spacing of 100 km in the horizontal, or a minimum of 51,000 horizontal grid points (about 1 degree × 1 degree resolution in the Tropics) to cover the globe. On the vertical, a lower limit of 50 layers between the surface and the top of the atmosphere is defined. The European project Carbon Atmospheric Tracer Research to Improve Numerics and Evaluation (CATRINE), initiated the intercomparison, funds its coordination, as well as some contributors, but it is hoped that other partners in the world will join the momentum and contribute to the study with their respective model and expertise.

The intercomparison considers four tracers: carbon dioxide (CO₂), sulfur hexafluoride (SF₆) and the radon isotope ²²²Rn, as well as a fictitious passive tracer fed by fossil CO₂ emissions. Within CATRINE, the tracers simulated by the contributors under the protocol are compared with each other, including vertically integrated mass fluxes for the total column. Diverse reference observations are used: ground-based or satellite tracer retrievals, air-sample measurements, as well as meteorological observations. We will define metrics to score the model performance. We will investigate how model configuration, like the vertical resolution, influences the model skill. The diagnostics will be applied to all tracers on weekly to monthly timescales, focusing on the impact of transport errors on the seasonal cycle at various layers and the total column. The correlation between the errors in the different layers with the error in the total column will support the attribution of the transport error.

The high-resolution simulation database currently under construction is the first of its kind and may serve various applications beyond the planned model evaluation, from, e.g., model uncertainty quantification to AI training.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverables

After the protocol design of D7.1, this deliverable D7.2 aims at providing a first ensemble of corresponding simulations. It also gathers observations and provides a first evaluation of the model simulations.

2.2.2 Work performed in this deliverable

We performed the various simulations required by the protocol with the expected four global models from the CATRINE partners and collected the simulations kindly prepared by an external contributor as well. Two off-line versions of the LMDz Atmospheric General Circulation model benefitted from significant developments in order to be ready for this exercise. A quality control procedure has been designed and measurements for the three real tracers have been gathered. First comparisons of model simulations and measurements have been prepared.

2.2.3 Deviations and counter measures

Running the high-resolution global simulations, transferring them to a unique place, giving feedback to the contributors in order to identify problems and potentially rerun their simulations, and finally comparing the models together or with the observations has represented significant technical challenges. They were expected to some extent, but we were still slowed down by a number of issues:

CATRINE

- Some of the models were found not mature enough at the start of the project to follow the protocol and resources were allocated to upgrade them so they can contribute to the simulation ensemble.
- Given the complexity of tracer transport models at high resolution, the partners could only finalize their first simulation of the four tracers and with all the requested outputs between November 2024 for the earliest, and May 2025 for the latest. In the meantime, some simulations had to be redone because errors were identified in the corresponding model configuration, for instance after the submitted data failed the quality control. A stable ensemble was therefore ready very late in the first half of CATRINE, leaving much less time than expected for the evaluation of the set.
- Some contributors made their simulations available via web interfaces that prevent a batch download (using the Linux *wget* command). The corresponding terabytes (TBs) worth of data had to be manually transferred to a laptop, one file after another, before being uploaded on the CEA/LSCE cluster.
- All data were submitted in the form of large compressed archives, which were sometimes found corrupted or incomplete after the tedious transfer process.
- a 120-TB disk mounted on the CEA/LSCE cluster had been prepared to host all simulations, but both the hardware and the software could not cope with the unusual disk capacity. After weeks of slow access and discussions with the vendors, the data were transferred to a smaller 55 TB disk, the behaviour of which is now nominal.

In addition, given the impact of the above deviations from the initial plan on partner CEA (need of developments of the LMDz-Dispersion model, late simulation deliveries from the other partners, need of more technical work to manage the data transfer and the database) and given hiring difficulties, the CEA work has been taken in charge by permanent staff rather than by a postdoc. A postdoc will start in October 2025.

2.3 Project partners

Partners	
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	ECMWF
COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	CEA
METEO-FRANCE	METEO-FRANCE
WAGENINGEN UNIVERSITY	WU
KARLSRUHER INSTITUT FUER TECHNOLOGIE	KIT
HELSINGIN YLIOPISTO	UH
UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE	URCA
ALBERT-LUDWIGS-UNIVERSITAET FREIBURG	UFR

3 Requested outputs

3.1 3D variables

The protocol requests the following 3D instantaneous variables every 3 hours for the full target period (2022-2023) from the all participating models:

- CO₂, fossil CO₂, SF₆ and ²²²Rn dry air mole fractions
- Height above the surface (m)

3.2 2D variables

The protocol requests the following 2D variables every 3 hours for the full target period:

- The instantaneous surface pressure (Pa),
- The instantaneous planetary boundary layer height (m),
- The instantaneous tropopause height (m).
- The instantaneous vertically integrated tracer mass over the whole atmospheric column (kg/m²):

$$m_c = \frac{1}{g} \int_0^{p_s} q dp \quad (1)$$

- Time integrated (over 3 hours), vertically integrated northward (F_N) and eastward (F_E) tracer flux (kg/m) for the total column:

$$F_N = \frac{1}{g} \int_0^{p_s} v q dp \quad \text{and} \quad F_E = \frac{1}{g} \int_0^{p_s} u q dp, \quad (2)$$

where q is the tracer mixing ratio (kg/kg), u and v are the zonal and meridional wind components (m/s), p (Pa) is atmospheric pressure, p_s is the surface pressure, g (m/s²) is the gravitational acceleration

- Time integrated (over 3 hours), vertically integrated sink of ²²²Rn (kg/m²) (new addition to the protocol)

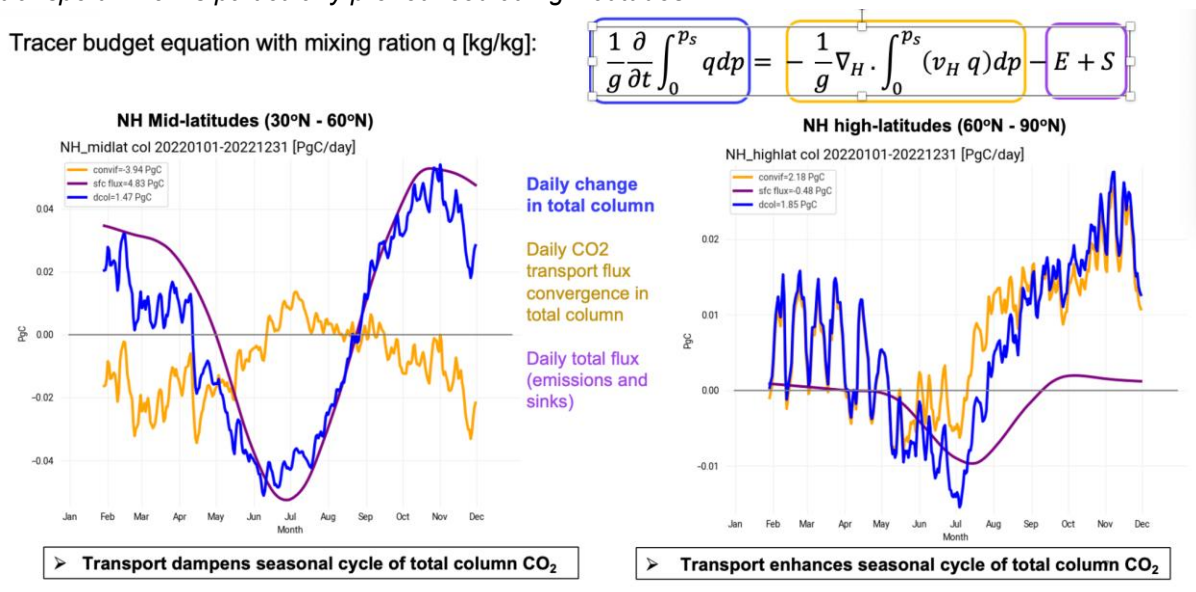
4 Additional diagnostics

New diagnostics for the budget equation of the total column have been implemented and tested with the IFS to assess their usefulness for understanding and quantifying the role of atmospheric transport in the variability of CO₂ in the atmospheric column observed by satellites. These diagnostics also aim to support the development of new skill scores to rank tracer transport models and provide feedback for atmospheric transport model developments. **Figure 1** below shows an example of vertically integrated diagnostics based on the tracer budget equation to assess the impact of transport on the seasonal cycle of the total column CO₂ in mid-latitudes and high latitudes. The impact of atmospheric transport is to enhance the seasonal cycle of CO₂ at high latitudes with a large influx from mid-latitudes of high-CO₂ air in winter and autumn, and low-CO₂ influx at the end of spring and early summer. At mid-latitudes, the seasonal cycle of the CO₂ column follows quite closely the budget of the total surface flux at the surface, but it is dampened by the transport, as the convergence of the atmospheric transport flux is anticorrelated with the total surface flux at the surface. This type of diagnostics

CATRINE

are important when it comes to interpreting variations in the total column observations from satellite data and understanding discrepancies in atmospheric inversion products. Indeed, atmospheric inversion models have the largest divergence in their estimates at high latitudes (Hugelius et al., 2024), where there are fewer satellite observations and the influence of modelled atmospheric transport is largest.

Figure 1. Time series of daily values for the different components of the vertically-integrated mass budget equation of CO₂ in units of PgC/day (described in the middle of the plot). These budget components are integrated over 2 latitudinal bands in the Northern Hemisphere. Left: mid-latitudes (between 30°N and 60°N). Right: high-latitudes (between 60°N and 90°N). The terms in the mass budget equation: q is the tracer concentration of CO₂ in this case; v_H is the horizontal wind vector, p is pressure; g is the gravitational constant; E and S are the emissions and sinks in the atmospheric column respectively. A 30-day running mean has been applied to reduce the synoptic variability associated with transport which is particularly pronounced at high latitudes.



In addition to the total column diagnostics, the online computation of the mass budget diagnostics for the partial columns of the atmospheric boundary layer and the troposphere has also been proposed as useful diagnostics to assess the vertical gradients and vertical transfer of mass across specified boundaries. These boundaries have been selected to be:

- The height of the highest model level that is less than or equal to 2 km above the surface. It represents the top of the boundary layer, or the lower boundary of the free troposphere.
- The model level closest to the tropopause. Here the tropopause represents the top of the free troposphere or the lower boundary of the tropopause transition layer. The tropopause definition is based on a stability and humidity criteria by searching from top of atmosphere downwards between 50 hPa and 700 hPa for model level L that fulfills the following criteria:

$$q(z) > 5ppm \ \& \ q(z - 600m) > 8ppm \ \& \ \frac{p}{T} \frac{dT}{dp} + \frac{R_d T}{g^2} 2.5 \times 10^{-4} < \kappa \quad (3)$$

where $\kappa = R_d/c_p$ is the ratio of the gas constant for dry air to the specific heat of dry air at constant pressure, and it's used in the computation of potential temperature.

Using pressure vertical coordinates, the mass budget diagnostics follow the vertically-integrated continuity equation over a layer from the surface (p_s) to the model level of the specific boundary (p_k):

$$\frac{1}{g} \frac{\partial}{\partial t} \int_{p_k}^{p_s} q dp = -\frac{1}{g} \nabla_H \cdot \int_{p_k}^{p_s} (v_H q_c) dp + F_k + E - S \quad (4)$$

All the terms need to be computed online as they are integrated with time and output daily from 00 UTC, except for F_k (the vertical exchange flux at the top boundary of the partial column) which will be computed as a residual from the other terms.

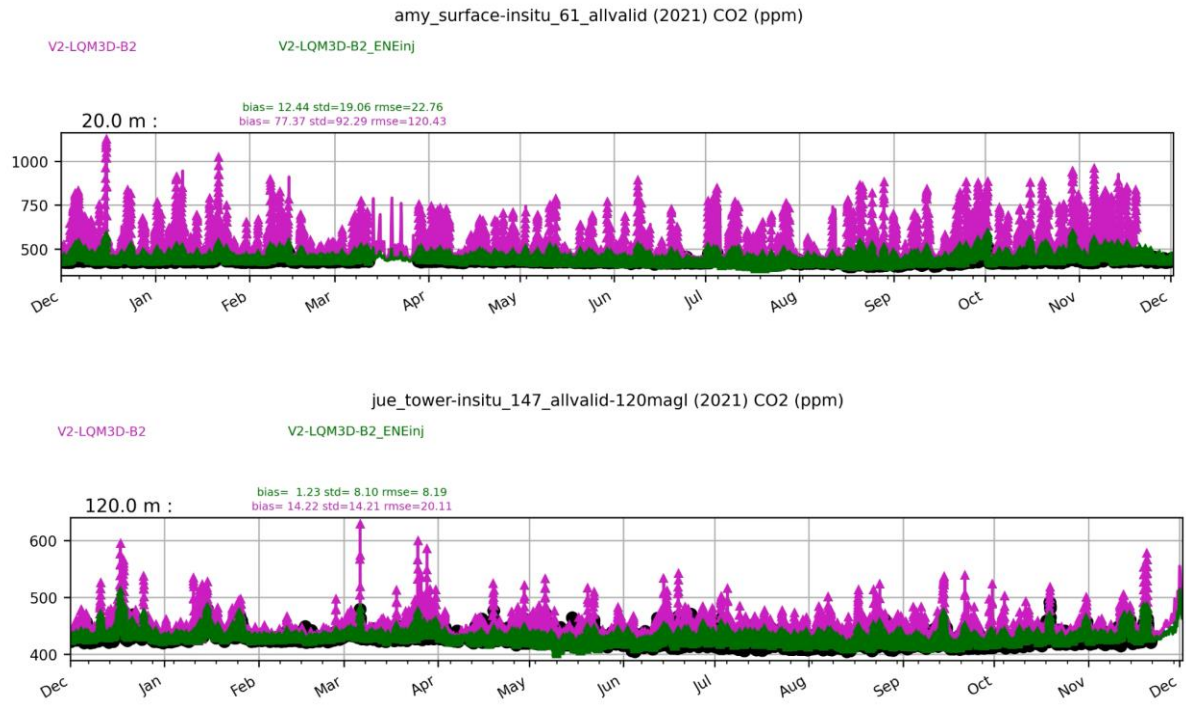
These additional diagnostics have been added to the revised D7.1 protocol. Because of the complexity of implementing such diagnostics, and the additional requirements for data storage and data transfer, they will be optional for the TransCom model intercomparison. In WP8, the atmospheric tracer mass budget diagnostics will be further developed and linked to the atmospheric transport evaluation scores and the WP5/6 test beds.

5 Additional sensitivity tests

The fossil fuel emission inventory GridFEDv2024.0 used in the protocol does not provide any indication on emission height and users usually assume that they correspond to the surface, while in reality emissions from aviation should have a 3D distribution and the power stations emit from chimney stacks. As CATRINE focuses on atmospheric transport of anthropogenic emissions, it is important to assess the realism of the model transport from emission hotspots. The energy sector is particularly important because it constitutes around 40% of total anthropogenic emissions. A procedure has been devised to extract the energy sector from the GridFEDv2024.0 total emissions using information from EDGAR v432 which is also used by GridFEDv2024.0. An injection height between 200 and 800 m above the surface is proposed following what is done by the operational CAMS global atmospheric composition forecast (<https://www.ecmwf.int/en/elibrary/81374-ifs-documentation-cy48r1-part-viii-atmospheric-composition>, Table 3.2, section 3.1.1d). The impact of using this injection height for the extracted energy sector has been tested with the IFS. The results show that there is a positive impact on sites near power stations at the surface and for the total column (see **Figures 2 and 3** below). Overall, CO₂ at the observing sites near power stations is reduced and the fit to observations is improved. This is also the case for the total column, which means that the transport is also sensitive to the injection height. Away from anthropogenic emissions the impact is small. Given the importance of the energy sector and the positive impact of the injection height on the model evaluation at the sites near emission hotspots, the sensitivity of the transport to the emission injection heights has been included in the update of the CATRINE TransCom protocol as a strongly recommended option.

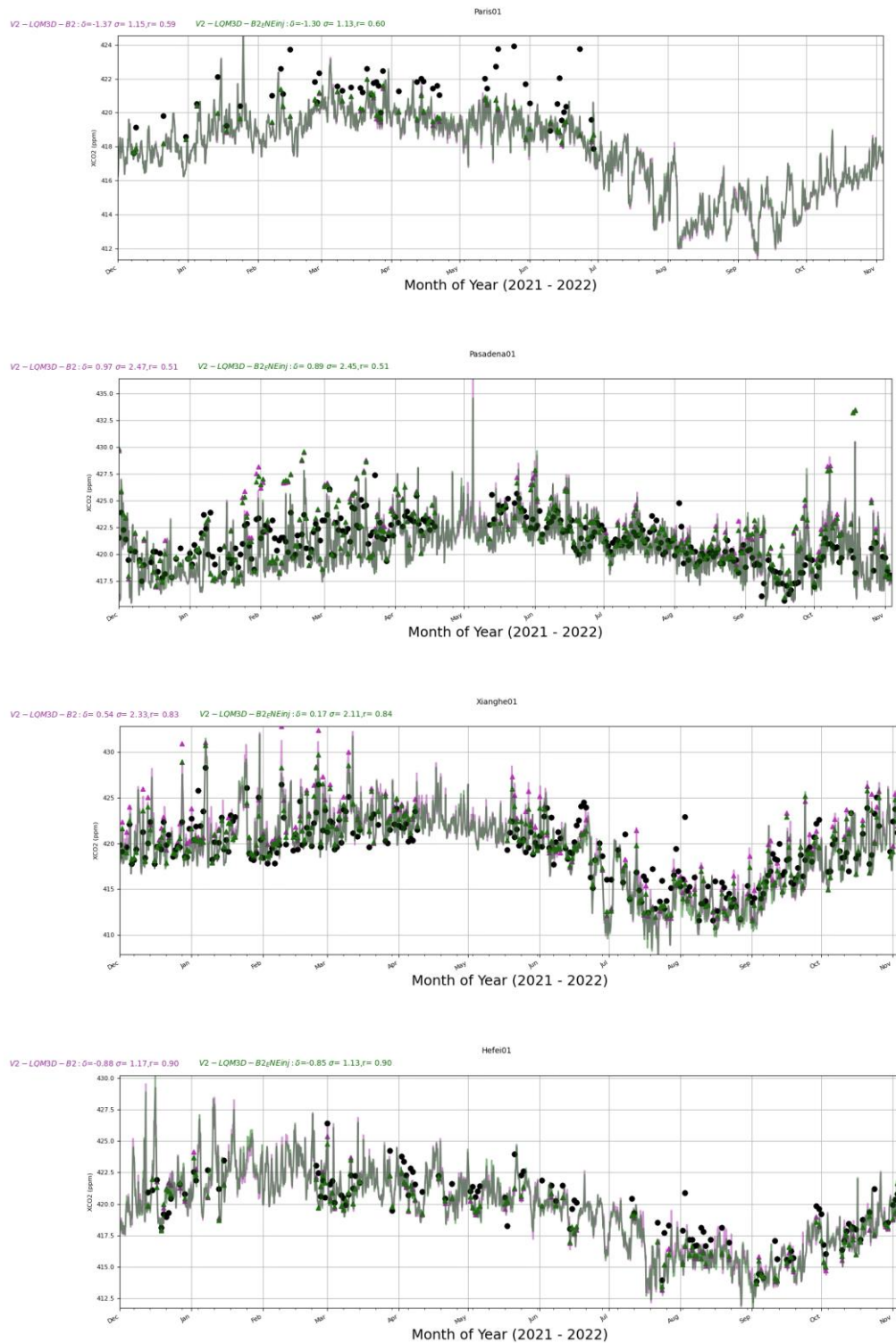
CATRINE

Figure 2. Evaluation of impact of injection height at two sites near power stations: Juelich (JUE, Germany) and Anmyeon-do (AMY, South Korea). The IFS simulations without and with injection height are depicted by the triangles in magenta and green respectively. The observations are shown as black circles. The bias, standard error (STD) and Root Mean Square Error (RMSE) are listed at the top of the plot.



CATRINE

Figure 3. Evaluation of impact of injection of injection height at 4 TCCON sites near strong anthropogenic emissions: Paris (France), Pasadena (CA, USA), Xianghe (China) and Hefei (China). The XCO₂ IFS simulations without and with injection height are depicted by the triangles in magenta and green respectively. The observations are shown as black circles. The bias, standard error (STD) and Pearson correlation coefficient (r) are listed at the top of the plot.



6 Models

6.1 Current participant list

As of now, seven models have been used to perform simulations for the protocol. They are listed in **Table 1**. Three of them correspond to different variations of the same model core (LMDz) and benefitted from significant developments in order to participate in this exercise as detailed below. Not all of them have been fully transferred to a central storage facility yet: only one month from ICON-ART has been archived on the CEA/LSCE cluster yet since it does not pass the quality control, as explained below.

Table 1. List of participating model configurations.

Model	Institute	Online/ Offline	horizontal	vertical	Driving meteorology	Contact person
ICON-ART	KIT	Online	79 km	120 levels	ERA5	S. Versick
IFS	ECMWF	Online	28 km	137 levels	ERA5	A. Agustí- Panareda
LMDz- degrees	CEA	Offline	1.4° × 0.7°	79 levels	ERA5 (nudging)	F. Chevallier and A. Martinez
LMDz-km	CEA	Offline	90 km	79 levels	ERA5 (nudging)	F. Chevallier and A. Martinez
LMDz-era5	CEA	Offline	90 km	137 levels	ERA5 (direct use)	F. Chevallier
TM5	WU	Offline	1° × 1°	68 levels	ERA5	J. Hooghiem
TOMCAT	Leeds Univ.	Offline	1.125° × 1.125°	60 levels	ERA5	C. Wilson

6.2 Developments of the LMDz models

The offline transport model LMDz has been the operational model for CO₂ inversions in the CAMS service since the origin. At the start of CATRINE, only one version was maintained, called LMDz-degrees in **Table 1** because its horizontal resolution is regular in terms of longitude and latitude degrees. This type of grid is not very numerically efficient at high latitudes. As a consequence, a new version based on a Goldberg polyhedron was under

CATRINE

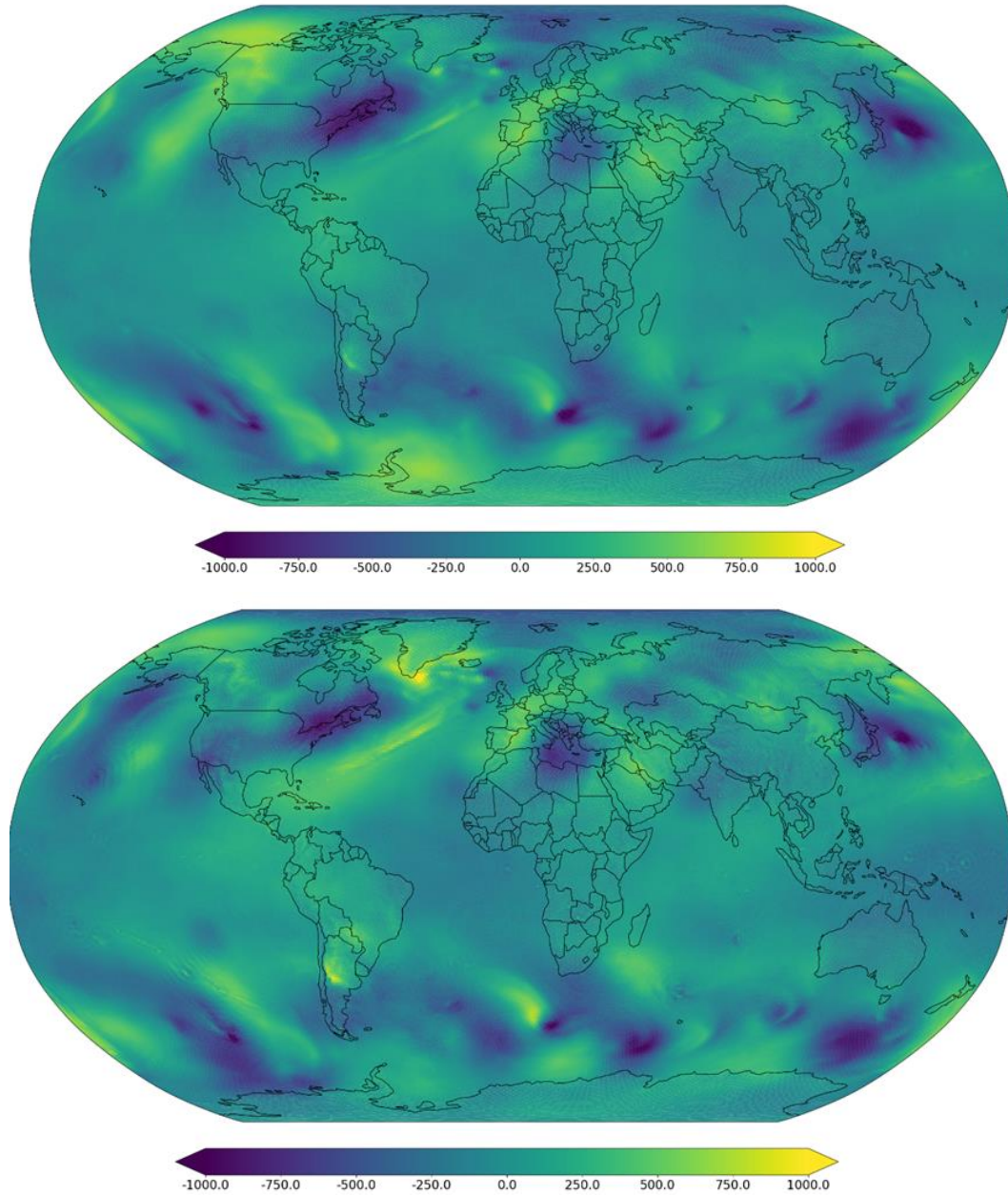
development, which avoids the computationally expensive convergence of the longitude-latitude grid's longitudinal dimension toward the poles. This development came about a few months after CATRINE was launched and the participation of the model in the intercomparison became obvious, in addition to the previous version. The model is called LMDz-km in **Table 1** because its horizontal resolution is regular in terms of kilometers. However, the implementation of the protocol diagnostics revealed unphysical inconsistencies between the surface pressure changes and the vertically-integrated air mass convergence, mostly as a result of a series of interpolations of the wind field: back-and-forth from cell edges to cell centroids, between the LMDz parent model and the off-line model. This situation arose because of technical constraints in the parent LMDz model, in which variables at the edges of cells could not be extracted. By joining resources from CAMS and from CATRINE, we could find a solution to extract the air mass fluxes at the edges of the cells, record them in a lossless compressed format to minimise the volume of data, and exploit them in the offline version. **Figure 4** illustrates the consistency of air mass convergence in the new version. The paper that documents the new LMDz-km model acknowledges the CATRINE resources (Chevallier et al., 2025).

The successful development of LMDz-km stimulated the creation of a third version of the LMDz off-line model in which the subgrid parameterizations (convection and boundary layer turbulence) are from the IFS rather than from the parent LMDz model. The model has been initially developed in order to prefigure some of the CO2MVS system and it was therefore interesting to make it participate in the intercomparison as well. The model is called LMDz-era5 in **Table 1**, because it directly uses the ERA5 archive without running the LMDz parent model to generate an air mass flux archive. CATRINE resources were used specifically to develop and test different options for the generation of the air mass flux archive from ERA5 (in terms of variable list and in terms of interpolation strategy) and to efficiently read the air mass flux archive in Dispersion-era5 (using MPI). They were also used to implement the protocol diagnostics at different spatial resolutions and correspondingly adjust the model.

Last, we plan to run the new LMDz-km model at 22 km horizontal resolution rather than at the standard 90-km one. This implies running the parent LMDz model at the same resolution in order to generate the air mass flux archive which only a few people have managed to do. We now have found an HPC configuration that supports the model at 22 km resolution. It gathers 2820 CPU cores, while the standard 90-km version only needs 640 ones. However, the model still diverges and we are investigating the problem.

CATRINE

Figure 4. Change in surface pressure, in Pa per 3 hours, on 1 Jan 2015, 00:00-06:00 UTC. Top: in the parent LMDz model. Bottom: recomputed by the air-mass convergence within LMDz-km.



7 Quality control

7.1 Tracer global mass change

Given prescribed surface fluxes for all tracers, a prescribed e-folding lifetime of 5.5 days for ^{222}Rn and neither atmospheric sink nor atmospheric source for the other tracer simulations, we expect consistent global mass changes for each tracer among the simulation ensemble: for CO_2 , and SF_6 , they should be close to the prescribed surface fluxes, while for ^{222}Rn , they should be close to zero. Checking this feature allows validating the surface flux interpolation that each contributor had to perform, together with the implementation of the ^{222}Rn lifetime. In the case of December 2022, **Table 2** shows that this is indeed the case, except for one model simulation (in grey). The corresponding contributor is investigating the problem.

Table 2. Tracer global mass change in the atmosphere for December 2022. The global budget of the prescribed surface fluxes over the same period is reported in the second column. ICON-ART is an outlier for all tracers and has been excluded from the computation of the lowest and largest values, written in black: its values appear in grey.

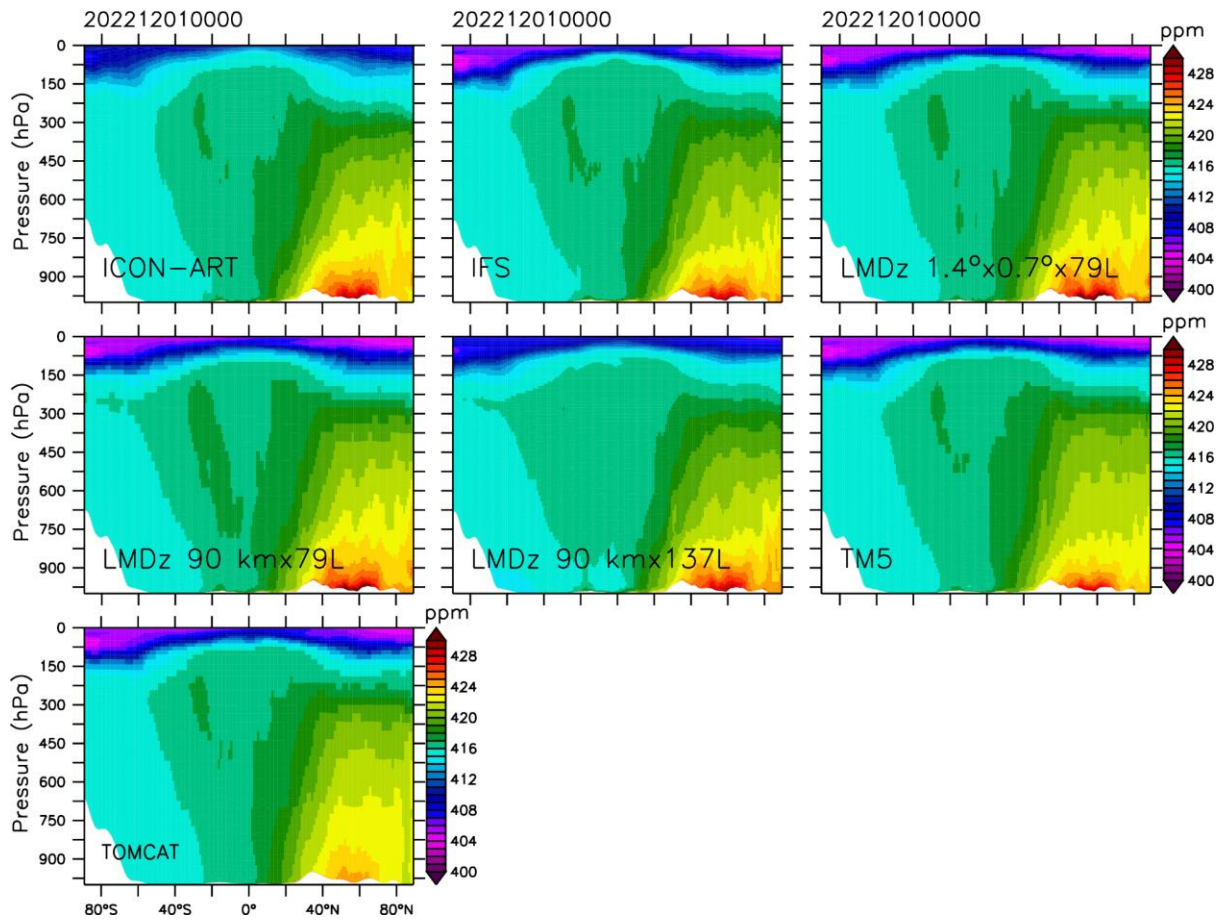
Molecule (unit)	Surface flux	Lowest tracer mass change	Largest tracer mass change
CO_2 (GtC)	1.16	1.14 / 1.06	1.17
SF_6 (Gg)	0.810	0.808 / 0.787	0.810
^{222}Rn (kg)	1.18533	-0.00386	-0.00471 / -0.00027

7.2 Zonal averages

To complement the tracer mass-change results, zonal averages of the tracer vertical profiles have been checked visually for all tracers, as illustrated in **Figures 5-7**. They reflect the different inter-hemispheric transport times of each tracer, as simulated by each model. They also show different vertical distributions. Higher-resolution models (prominently IFS) allow more details than the lower-resolution ones (TOMCAT has the coarsest resolution), both in the latitudinal and in the vertical direction, that we will aim at evaluating. The ^{222}Rn simulation by ICON-ART shows much smaller values than with the other models and this has been traced back to a specified e-folding lifetime of ~ 3.8 days rather than of 5.5 days imposed by the protocol. This simulation will be rerun. Similarly, the TM5 ^{222}Rn simulation shows very large values during the first months (not shown) likely as a result of an inappropriate initial state and will be rerun as well.

CATRINE

Figure 5. Zonal averages of the CO₂ simulations on 1 December 2022 at 00:00.



CATRINE

Figure 6. Zonal averages of the SF₆ simulations on 1 December 2022 at 00:00.

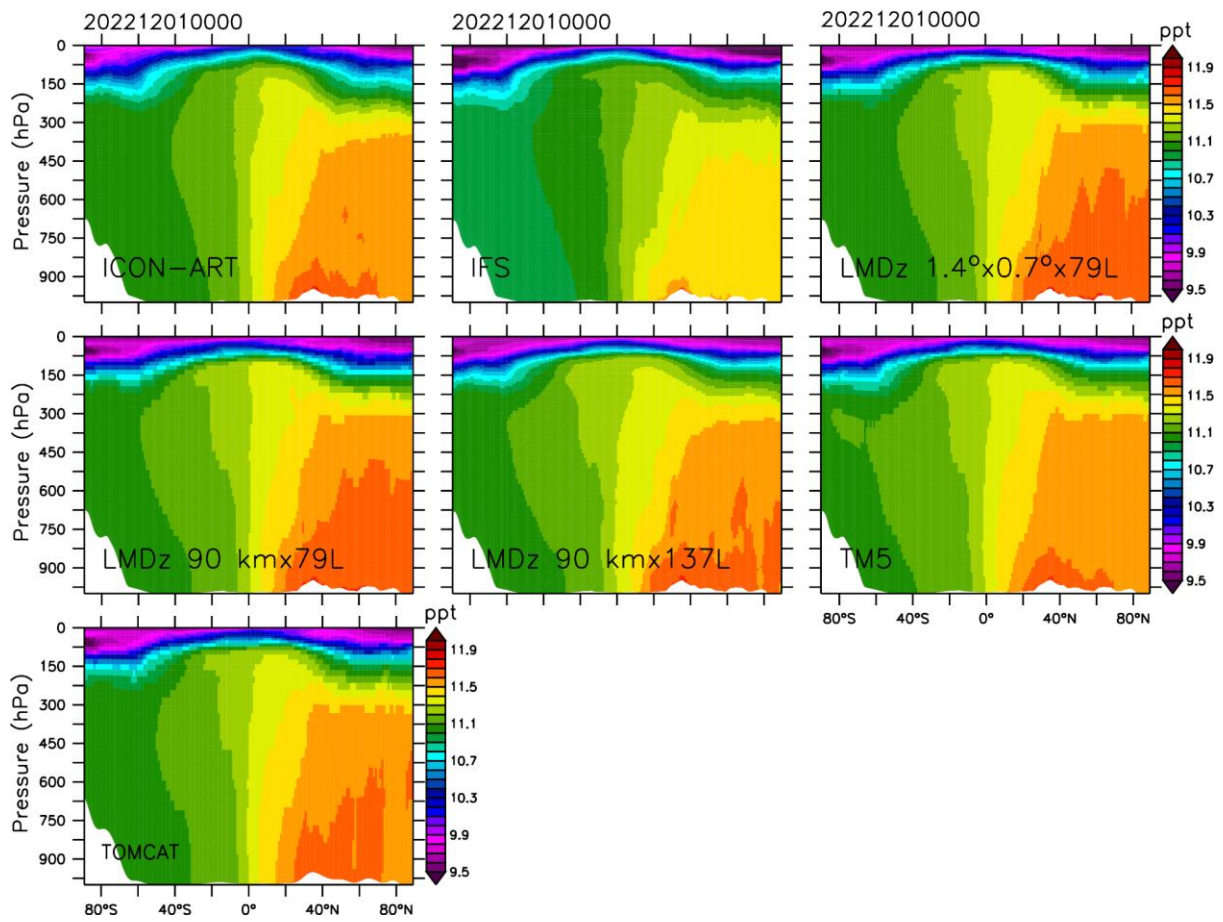
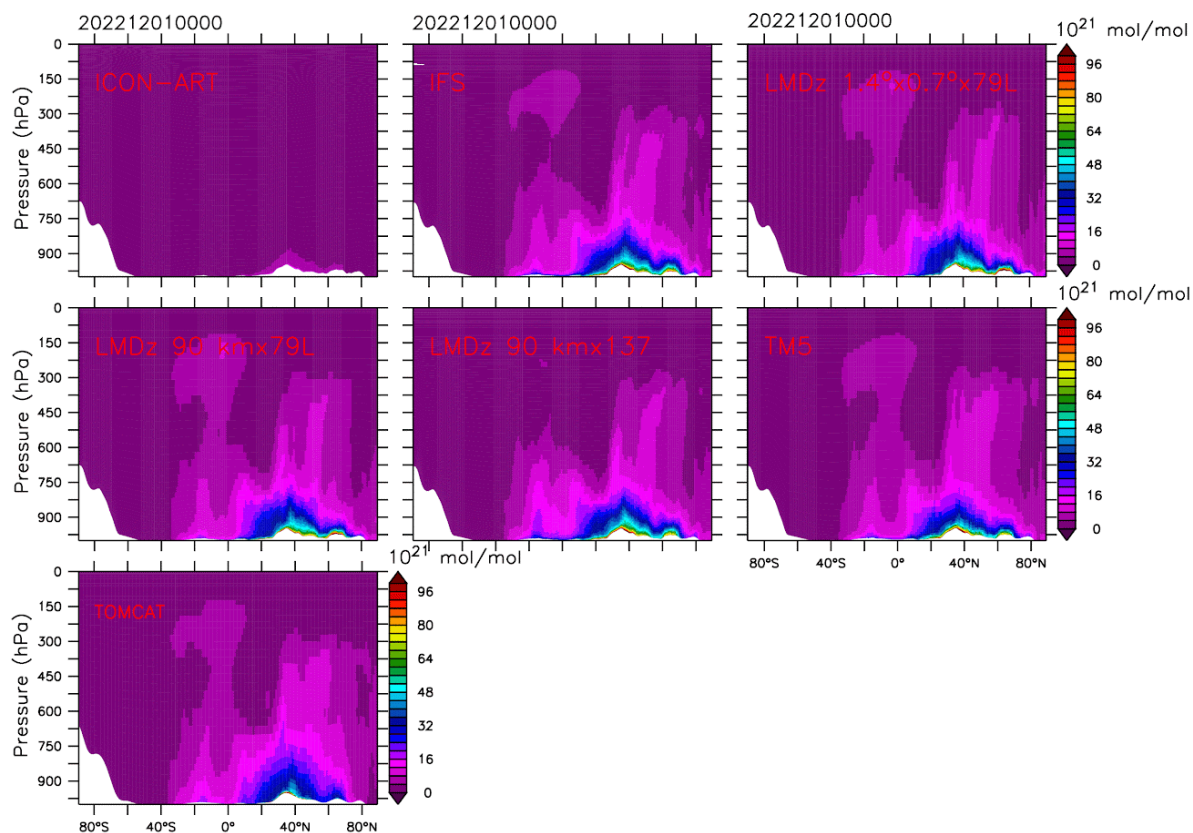


Figure 7. Zonal averages of the ^{222}Rn simulations on 1 December 2022 at 00:00.

8 Comparison to observations

The simulated data have been compared to CO_2 measurements from ObsPack Globalview+_{v10.1} (Schuldt et al., 2024), SF_6 measurements from Obspack sf6_1_v3.1 (Schuldt et al., 2025) and ^{222}Rn measurements from the World Data Centre for Greenhouse Gases (WDCGG, <https://gaw.kishou.go.jp/>).

The air pressure around the observations is usually not reported while protocol D7.1 has minimised the list of required variables in order to minimise file size: the list includes the vertical pressure grid but not the geometric heights, which appear only in the files of the three LMDz models (at the start of each month) and of TM5 (every three hours). To compensate for this absence for the moment, we associate each measurement to the pressure value at the middle of the LMDz-era5 vertical layer where the measurement was taken, given the layer height at the start of the corresponding month. This information in LMDz-era5 corresponds to that of the ERA5 reanalysis where it was taken, aggregated at the 90-km resolution of the LMDz-era5 model. The protocol will be adjusted to collect the height information directly from the contributors in the future.

Figure 8 and **Figure 9** compare the model simulations with NOAA's Aircore CO_2 profiles close to Boulder, Co, USA, in 2022 and 2023. There is usually one profile per month. The models spread much, in particular around and above the tropopause and in the boundary layer, despite the same initial CO_2 state on 1 December 2021 and the same surface CO_2 fluxes. Even the version of LMDz that directly exploits ERA5 air mass fluxes (light blue curve, called *LMDz-era5* in the key) and IFS (red curve) yield very different profiles while they use the same subgrid parameterization. TM5 (purple curve) never appears as an outlier while the other

CATRINE

models may stick out in one month or another, particularly the lower resolution model, TOMCAT (orange curve).

Figure 8. Comparison of six of the models with NOAA's Aircore CO₂ profiles in 2022 close to Boulder, CO, USA. The key lists the models in the same order as Table 1. The month of each profile is indicated in the title of the corresponding subplot in YYYYMM format. The measurements are from file `co2_aircorenoaa_aircore_1_allvalid.nc` of Schuldt et al. (2024).

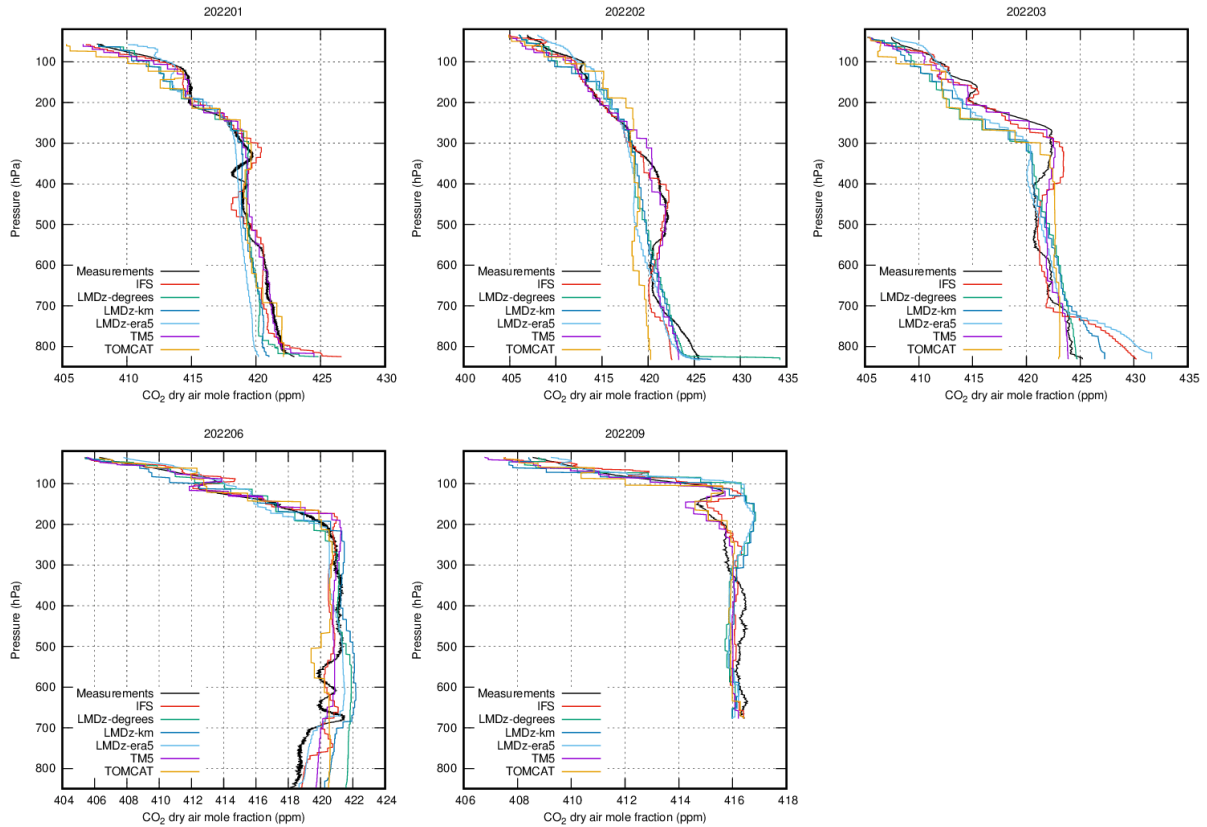
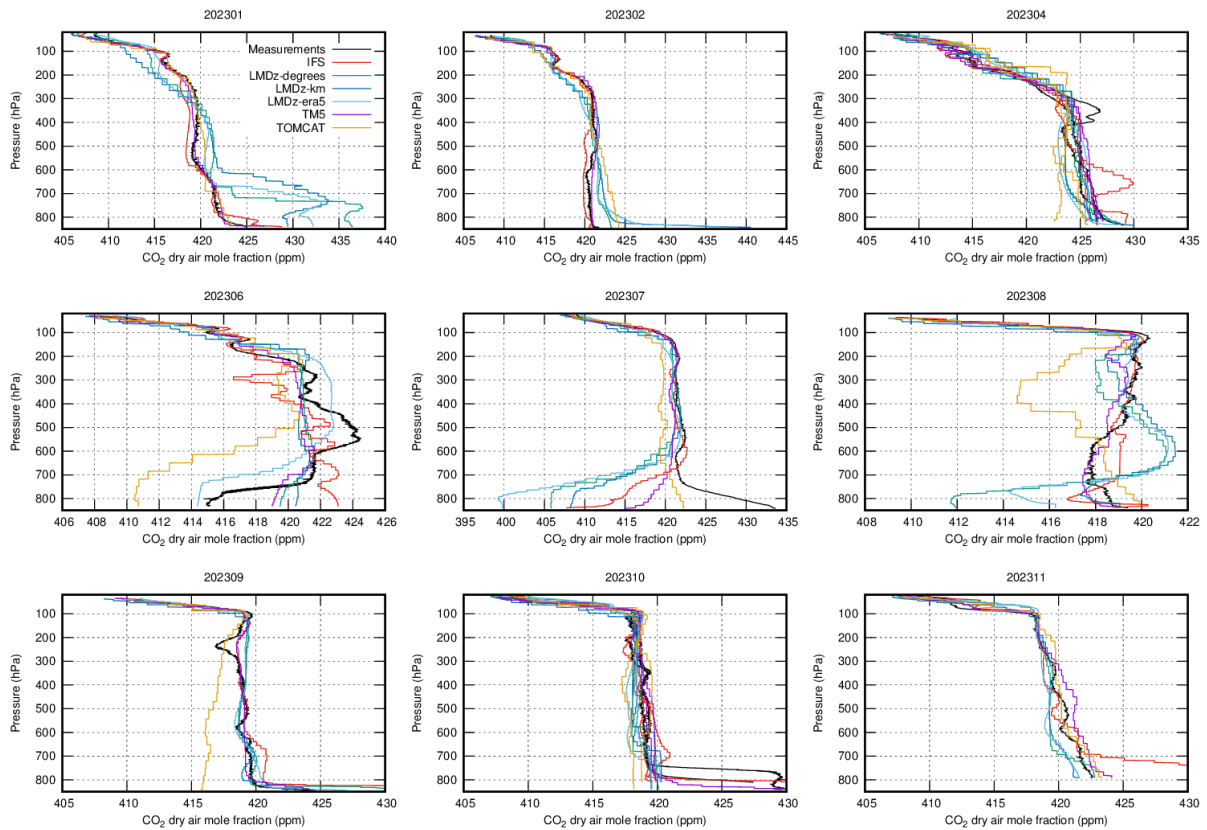


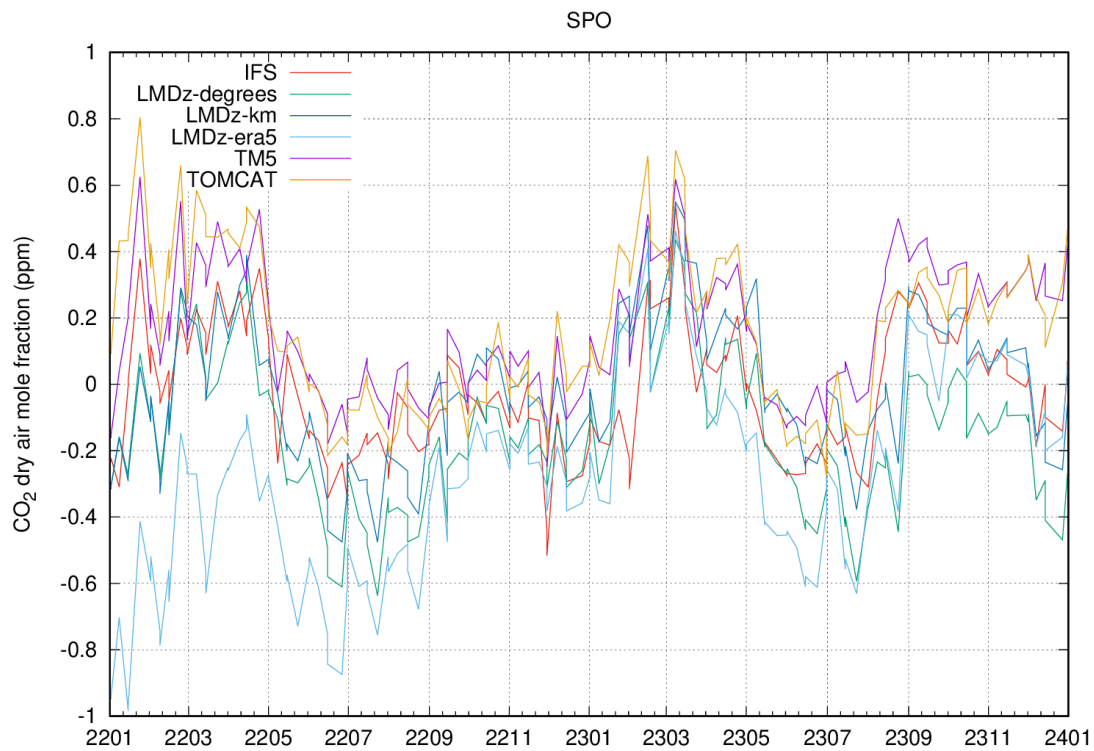
Figure 9. Same as Figure 8 for the year 2023.



The CO₂ simulations are compared with flasks and in situ measurements at the South Pole and on the West coast of Ireland (Mace Head) in **Figure 10** and **Figure 11**. LMDz-er5 is an outlier at South Pole during the first six months suggesting that it is still in spin-up phase and has not stabilized after its start from the common initial CO₂ state on 1 December 2021.

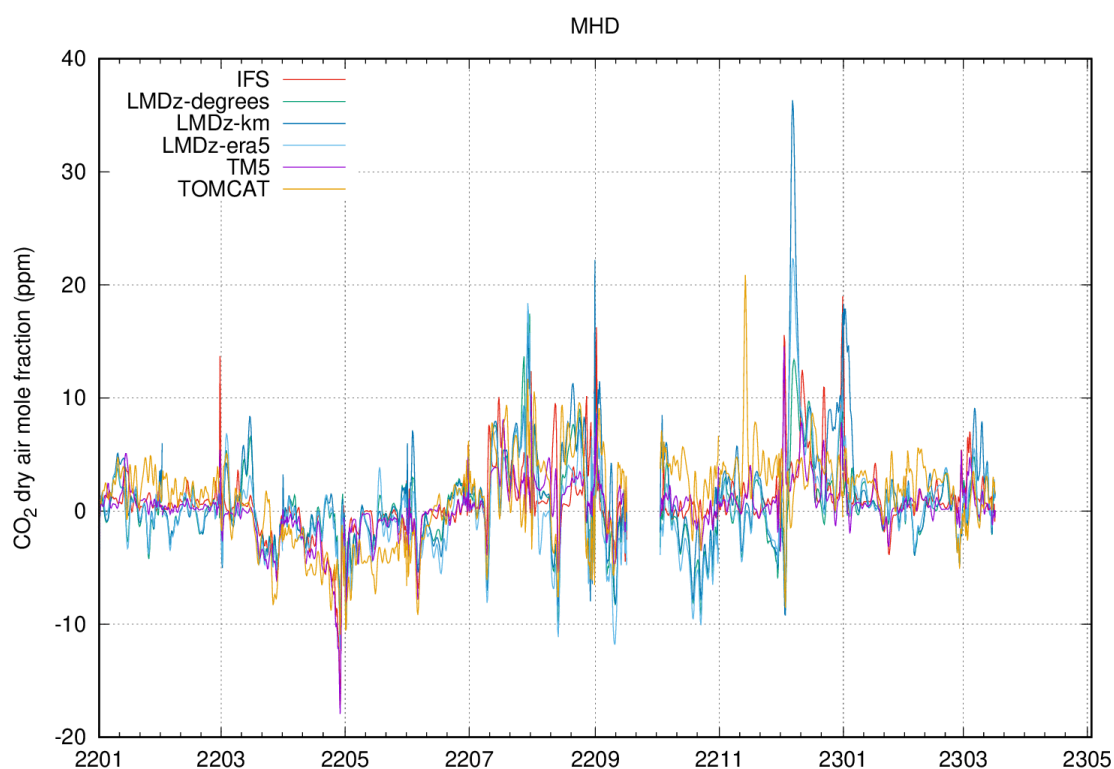
CATRINE

Figure 10. Difference between each one of the six models with NOAA's CO₂ flask measurements at station South Pole Observatory (SPO). The date appears in YYMM format on the x-axis. The measurements are from file `co2_spo_surface-flask_1_representative.nc` of Schuldt et al. (2024).



CATRINE

Figure 11. Same as Figure 10 with ICOS's CO₂ in situ measurements at station Mace Head (MHD) from file `co2_mhd_tower-insitu_11_allvalid-24magl.nc` of Schuldt et al. (2024). A smoothing has been applied to each curve per month for better visibility.



The model simulations are compared with SF₆ measurements at SPO in **Figure 12**. This time IFS is the outlier, starting close from the other models but then diverging to stabilize after a year closer to the measurements. Over the two year period, IFS shows the best latitudinal distribution which suggests the best interhemispheric gradient (**Figure 13**). It is followed in quality by TM5. **Figure 14** illustrates the work on ²²²Rn with model-minus-measurement difference at Minamitori-shima, Japan. As mentioned before, the TM5 simulation starts very high further to an inappropriate initial state, but its spin-down finishes, its differences align close to IFS, with smaller amplitudes than the three LMDz versions. The three LMDz models show much larger difference amplitudes than the others, despite different vertical resolution and physics among themselves. They also show larger ²²²Rn surface concentration variations at the South Pole than the others (not shown). This feature suggests that, for instance, the time stepping in the LMDz off-line model is not optimal, a point that is being investigated.

CATRINE

Figure 12. Same as Figure 11 with NOAA's SF₆ flask measurements at station South Pole Observatory (SPO from file co2_spo_surface-flask_1_representative.nc of Schuld et al. (2024).

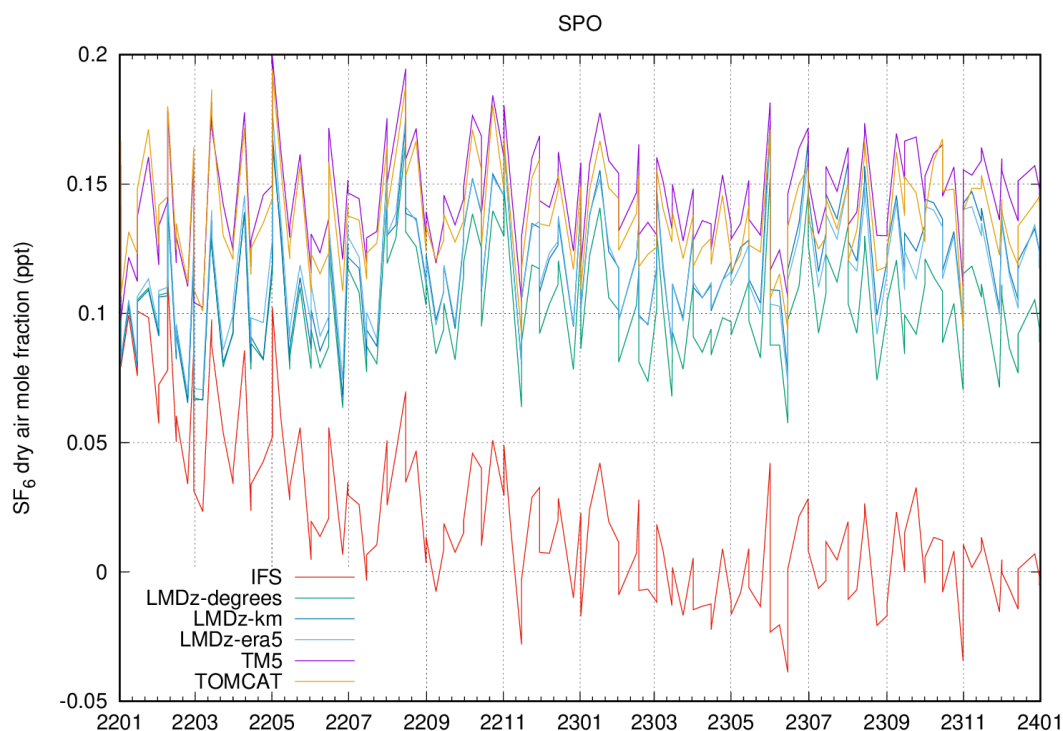


Figure 13. Mean difference between each one of the six models with NOAA's SF₆ flask measurements at stations SPO, PSA, CGO, SMO, RPB, MLO, MHD, BRW, SUM and ALT, represented here by their latitude along the x-axis, over the full period 2022-2023. The date appears in YYMM format on the x-axis. The measurements are from files sf6*_surface-flask_1_representative.nc in Schuld et al. (2025).

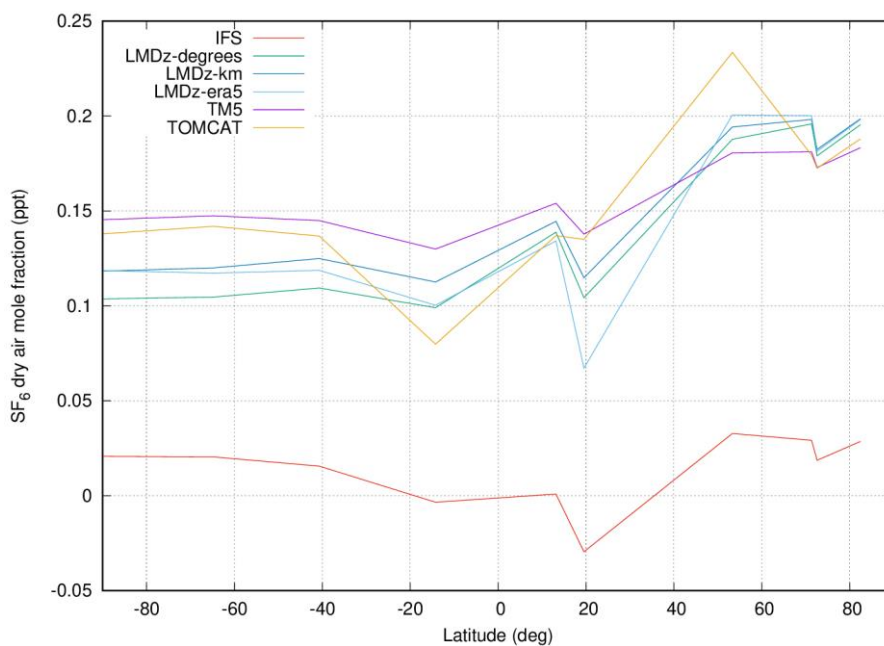
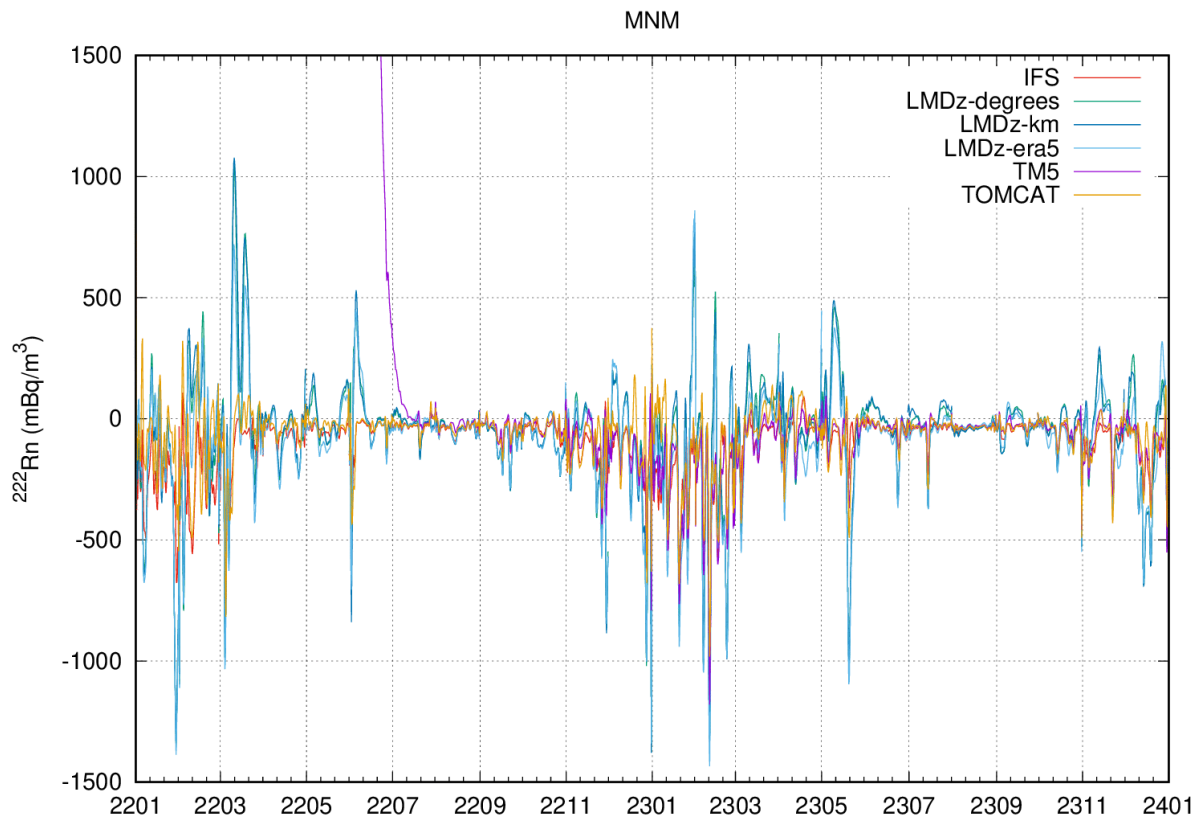


Figure 14. Difference between each one of the six models with MRI's ²²²Rn continuous measurements at station Minamitori-shima, Japan. The date appears in YYMM format on the x-axis. The

CATRINE

measurements are from file `222rn_mnm_surface-insitu_48_9999-9999_hourly.nc` of WDCGG. The TM5 curve starts outside the figure further to an inappropriate initial state.



9 Next steps

The models show distinct differences when compared towards the various observations with a large spread between model simulations associated with transport in the boundary layer and the upper-troposphere. The present results raise interesting questions regarding the model behaviour and reruns are likely after further analysis. More models will also likely join the ensemble. Beyond correcting model configuration errors and bugs, our goal in the coming WP8 is to develop common metrics and test cases allowing comparisons between models and providing guidelines for the development of the CO2MVS modelling systems. Our strategy is summarized in **Table 3**. It covers spatial scales from local to global and temporal scales from sub-daily to seasonal, each time exploiting the diversity of existing measurements for the three real tracers which are simulated.

Table 3. Multiscale evaluation objectives and corresponding observational support.

Spatial scales	Temporal scales	Tracer indices	Observations
Local to city (gridpoint scale)	3-hourly (diurnal cycle)	- Vertical gradients - Amplitude and phase of diurnal cycle	- Tower vertical profiles - AirCore and aircraft vertical profiles - Radon surface measurements
Region/ country	Daily to weekly (synoptic episodes)	- Vertical variability associated with plumes in the free troposphere - Day-to-day variability (sigma and correlation)	AirCore and aircraft vertical profiles
Continental/ Sea-land	Weekly to monthly (inter-continental transport)	Surface and total column gradients between continents	- Obspack surface network - TCCON and OCO-2 XCO ₂
Global latitudinal bands in NH and SH: high-latitudes, mid-latitudes and tropics	Monthly (seasonal cycle)	- Inter-hemispheric gradient (total column, boundary layer and free troposphere) - Amplitude and phase of seasonal cycle (total column, boundary layer and free troposphere)	- AirCore, aircraft vertical profiles - TCCON and OCO-2 XCO ₂ - Obspack background sites

10 Conclusion

The protocol defined in Deliverable D7.1 has been implemented among the CATRINE partners and beyond: seven models have already been used to generate the simulations for each required tracer. Most of the data has been transferred to a unique facility which gathers 10 TB worth of data at present. A simple quality control procedure has been implemented that already highlighted a series of errors in some model simulations: some of the simulations have already been rerun while others are still being investigated. The large volume of data generated by this exercise represents a major technical challenge, from simulating the tracers with a sophisticated model on a high-performance computing facility to exploring the result on another infrastructure. Comparison scripts have been prepared to

CATRINE

extract the model-equivalent to large measurement databases (NOAA's ObsPack and WDCGG). They work in chunks of one month and have to be run on large memory computation nodes. An evaluation of CO₂ profiles, SF₆ inter-hemispheric gradient and near-surface ²²²Rn variability has been performed. Although the fit to the observations at background sites is generally good, systematic differences between models in the inter-hemispheric gradient, as well as the large spread near the surface and in the upper-troposphere, reveal significant differences in the transport models. In WP8 we still plan to process several tens of measurement files per species in order to expand the evaluation of the model simulations. They will benefit from the scripts prepared by the Copernicus Atmosphere Monitoring Service for the Evaluation and Quality Control of its global service products, in particular those prepared at CEA/LSCE for the Greenhouse gas variables. The additional proposed diagnostics will also be instrumental at elucidating the systematic errors and differences detected in the transport models during the TransCom intercomparison exercise in WP8.

References

- Chevallier, F., Martinez, A., Lloret, Z., Takache, S., & Cozic, A. (2025). Offline atmospheric transport on a global mesh of hexagons. *Journal of Geophysical Research: Atmospheres*, 130, e2025JD043579. <https://doi.org/10.1029/2025JD043579>
- Hugelius, G., Ramage, J., Burke, E., Chatterjee, A., Smallman, T. L., Aalto, T., et al. (2024). Permafrost region greenhouse gas budgets suggest a weak CO₂ sink and CH₄ and N₂O sources, but magnitudes differ between top-down and bottom-up methods. *Global Biogeochemical Cycles*, 38, e2023GB007969. <https://doi.org/10.1029/2023GB007969>
- Schuldt, K.N., J. Mund, et al. (2024); Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2023; obspack_co2_1_GLOBALVIEWplus_v10.1_2024-11-13; NOAA Earth System Research Laboratory, Global Monitoring Laboratory. <http://doi.org/10.25925/20241101>.
- Schuldt, A. R. Jacobson, et al. (2025); Multi-laboratory compilation of atmospheric sulfur hexafluoride data for the period 1983-2024; obspack_sf6_1_v3.1_2025-07-01; NOAA Earth System Research Laboratory, Global Monitoring Laboratory. <http://doi.org/10.25925/20250701>.

Document History

Version	Author(s)	Date	Changes
0 (Initial document created)	Frédéric Chevallier, Anna Agustí-Panareda, Matthieu Gautherot, Joram Hooghiem, Marteen Krol, Wouter Peters and Stefan Versick	25/06/2025	
0.1	Frédéric Chevallier, Anna Agustí-Panareda, Matthieu Gautherot, Joram Hooghiem, Marteen Krol, Wouter Peters and Stefan Versick	27/06/2025	update further to the review
0.2	Frédéric Chevallier, Anna Agustí-Panareda, Matthieu Gautherot, Joram Hooghiem, Marteen Krol, Wouter Peters and Stefan Versick	30/07/2025	Expanded with evaluation of SF6 and 222Rn observations and inclusion on sections with optional diagnostics, sensitivity tests and model development
0.3	Frédéric Chevallier, Anna Agustí-Panareda, Matthieu Gautherot, Joram Hooghiem, Marteen Krol, Wouter Peters and Stefan Versick	25/08/2025	Revised version after internal review

Internal Review History

Internal Reviewers	Date	Comments
Michail Diamantakis	27/06/2025	minor
Leena Järvi	21/08/2025	minor
Tanya Warnaars	25/08/2025	Final editorial edits

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.