CAMS Service Evolution

D2.3 Report on assimilation of HCHO data

Due date of deliverable	30.09.2024
Submission date	30.09.2024
File Name	CAMEO-D2-3-V1.2
Work Package /Task	WP2/ T2.2
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Flora Kluge, Vincent Huijnen, Antje Inness, Glenn-Michael Oomen, Jean-François Müller, Trissevgeni Stavrakou, Miró van der Worp
Revision number	1.2
Status	Issued
Dissemination Level	Public

Funded by the European Union

The CAMEO project (grant agreement No 101082125) 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

In this report, we present first steps towards a biogenic volatile organic compound emission inversion system in IFS-COMPO. As part of this work, operational TROPOMI S5P formaldehyde data has been included in the CAMS system. Complemented by aircraft measurements, the satellite observations were used for a comprehensive assessment of IFS-COMPO formaldehyde observations in different global regions, seasons, altitudes, and air masses. A linearised and simplified formaldehyde (HCHO) chemistry scheme has been developed and optimised for low NOx conditions dominated by biogenic volatile organic compound emissions, as they are usually found over tropical rainforests. Based on this simple chemistry scheme, a tangent linear and adjoint code for HCHO has been derived. The assimilation of HCHO observations in IFS-COMPO has been implemented and was successfully tested in a first step without tangent linear and adjoint or biogenic volatile organic compound emission inversion.

Table of Contents

2 Introduction

2.1 Background

Monitoring the composition of the atmosphere is a key objective of the European Union's flagship Space programme Copernicus, with the Copernicus Atmosphere Monitoring Service (CAMS) providing free and continuous data and information on atmospheric composition.

The CAMS Service Evolution (CAMEO) project will enhance the quality and efficiency of the CAMS service and help CAMS to better respond to policy needs such as air pollution and greenhouse gases monitoring, the fulfilment of sustainable development goals, and sustainable and clean energy.

CAMEO will help prepare CAMS for the uptake of forthcoming satellite data, including Sentinel-4, -5 and 3MI, and advance the aerosol and trace gas data assimilation methods and inversion capacity of the global and regional CAMS production systems.

CAMEO will develop methods to provide uncertainty information about CAMS products, in particular for emissions, policy, solar radiation and deposition products in response to prominent requests from current CAMS users.

CAMEO will contribute to the medium- to long-term evolution of the CAMS production systems and products.

The transfer of developments from CAMEO into subsequent improvements of CAMS operational service elements is a main driver for the project and is the main pathway to impact for CAMEO.

The CAMEO consortium, led by ECMWF, the entity entrusted to operate CAMS, includes several CAMS partners thus allowing CAMEO developments to be carried out directly within the CAMS production systems and facilitating the transition of CAMEO results to future upgrades of the CAMS service.

This will maximise the impact and outcomes of CAMEO as it can make full use of the existing CAMS infrastructure for data sharing, data delivery and communication, thus supporting policymakers, business and citizens with enhanced atmospheric environmental information.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

This deliverable reports on the ongoing work of the assimilation of HCHO in IFS-COMPO. Adding formaldehyde to the assimilation system requires the preparation of HCHO observational data, the development of a linearised and simplified formaldehyde chemistry scheme in IFS-COMPO, and the derivation of a corresponding tangent linear and adjoint. This combined work is an essential step towards developing a biogenic volatile organic compound (BVOC) emission inversion system in the IFS.

2.2.2 Work performed in this deliverable

In this deliverable, the work as planned in the Description of Action (DoA, WP2.2 T2.3.1, T2.3.2, and T2.3.3) was performed:

Task 2.3.1: Preparation of HCHO observational data for use in CAMS system. Provide bias correction based on validation studies. Provide guidance on observational errors, filtering requirements. Provide evaluation of CAMS model against HCHO retrievals: assess and interpret model biases.

Task 2.3.2: Develop and assess simplified HCHO chemistry. Design simplified chemistry mechanism applicable for biogenic hydrocarbons under tropical (low-NOx) conditions, using box-model calculations, tuned with respect to IFS(CB05) chemistry. Implement simplified chemistry in IFS and compare performance against the full-chemistry configuration. Adjust simplified chemistry for high NOx regimes and assess its performance for biomass burning and anthropogenic sources.

Task 2.3.3: Assimilation of HCHO data in 4DVar context, derivation of TL/AD of simplified HCHO chemistry. Perform assessment of impact of HCHO assimilation in CAMS system: Compare assimilated HCHO values against direct retrievals and provide interpretation of differences. Assess dependence on emissions and deposition treatment. Design TL and adjoint code of linearized chemistry and carry out corresponding HCHO assimilation tests. Assess the performance of the HCHO assimilation, including its impact on other trace gases and aerosol.

2.2.3 Deviations and counter measures

No deviations have been encountered.

2.2.4 CAMEO Project Partners:

ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
Met Norway	METEOROLOGISK INSTITUTT
BSC	BARCELONA SUPERCOMPUTING CENTER-CENTRO NACIONAL DE SUPERCOMPUTACION
KNMI	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMi
SMHI	SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT
BIRA-IASB	INSTITUT ROYAL D'AERONOMIE SPATIALEDE BELGIQUE
HYGEOS	HYGEOS SARL
FMI	ILMATIETEEN LAITOS
DLR	DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV
ARMINES	ASSOCIATION POUR LA RECHERCHE ET LE DEVELOPPEMENT DES METHODES ET PROCESSUS INDUSTRIELS
CNRS	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS
GRASP-SAS	GENERALIZED RETRIEVAL OF ATMOSPHERE AND SURFACE PROPERTIES EN ABREGE GRASP

CU	UNIVERZITA KARLOVA
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
MF	METEO-FRANCE
TNO	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO
INERIS	INSTITUT NATIONAL DE L ENVIRONNEMENT INDUSTRIEL ET DES RISQUES - INERIS
IOS-PIB	INSTYTUT OCHRONY SRODOWISKA - PANSTWOWY INSTYTUT BADAWCZY
FZJ	FORSCHUNGSZENTRUM JULICH GMBH
AU	AARHUS UNIVERSITET
ENEA	AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

3 Preparation of HCHO observational data for use in the CAMS system

The assimilation of formaldehyde (HCHO) in the CAMS system requires daily and globally available HCHO observations. The Tropospheric Monitoring Instrument (TROPOMI), onboard the S5P satellite platform, daily provides operational near-real time and offline level 2 HCHO (https://doi.org/10.5270/S5P-tjlxfd2) at a spatial resolution of 5.5 km × 3.5 km. The following chapter reports on the preparation of TROPOMI HCHO retrievals for use in the CAMS system as well as an extensive validation of IFS-COMPO HCHO using satellite and airborne observations.

3.1 HCHO super-obbing using negative HCHO satellite retrievals

Considering that the horizontal resolution of TROPOMI HCHO data (5.5 km × 3.5 km) is higher than the model resolution of T511 (about 40 km × 40 km, https://confluence.ecmwf.int/display/CKB/ERA5%3A+What+is+the+spatial+reference), TROPOMI observations are not spatially representative for the model grid boxes. To overcome this representativeness error, the data are converted into so-called "super-observations" before they are included in the CAMS system (e.g. Rijsdijk et al., 2024). This is achieved by averaging the data onto the T511 resolution of the model. The method currently used by CAMS to create super-observations simply calculates the averages of the observations in a grid box. This method averages the data separately for different surface types (e.g. land, ocean, ice, etc.). The same applies for the observational errors and averaging kernels of the data, which are averaged in the same way. This assumes uncorrelated observation errors. Averaging the data in a grid box reduces the random errors and also the representativeness errors due to unresolved small-scale features that are seen in TROPOMI data but are not resolved in the model.

CAMS uses a wide range of atmospheric composition observations that come in various data formats (e.g. netcdf, HDF, BUFR, ASCII, etc.). These data are converted into BUFR format in ECMWF's Scalable Acquisition and Pre-Processing (SAPP) system using WMO BUFR templates. The CAMS super-obbing method works on the 6-hourly BUFR files instead of the original data files and is therefore limited to using the number of parameters available in the BUFR files. This has the advantage of being easily applicable to a wide range of retrieval products from various instruments and for various species. Further, the SAPP processing removes any duplicates, which may be a result of orbital overlaps. The ECMWF super-observation software provides the flexibility to change the resolution of the created super-observations in an experiment by simply changing a parameter on script level, without having to reprocess the input data.

Currently the TROPOMI HCHO observations are super - obbed to the model resolution of T511, which corresponds to a reduced gaussian grid of N256 (https://confluence.ecmwf.int/display/EMOS/N256). In a pre-screening step the data are sorted by surface type and subsequently, super-observations are created separately for each surface type. This ensures that the averaging kernels of the HCHO data are similar for the averaged observations. Super-observations are generated if more than 6 observations are contained in a grid box. For spectral retrievals close to the instrument's detection limit, negative measurements occur out of statistical reasons. The deterministic model does not reproduce these negative observations. As such, negative satellite retrievals have previously been discarded in the super-obbing procedure. However, this may cause a positive offset. Hence, the BUFR files have been modified to also include negative satellite retrievals when generating super-observations for a grid box.

The latter is of particular relevance when analysing trace gases with low optical densities, such as formaldehyde. However, it has to be noted that this improved procedure is currently still being tested for TROPOMI NO2, as discussed in CAMEO deliverable D2.2. As a result, the super-obbing procedure applied in the presented analysis does not yet include these updates.

3.2 Bias correction of S5P HCHO retrievals based on validation studies

Comparisons of satellite HCHO vertical columns with ground-based data reveal biases in the satellite HCHO data; more precisely, low HCHO columns are found to be generally too high whereas high HCHO columns are most often too low (Vigouroux et al. 2020; De Smedt et al. 2021). Validations of TROPOMI HCHO vertical column densities with aircraft observations corroborate these results (Zhu et al. 2020). The definite source of this bias is unknown, but it is generally assumed to be related to the retrieval procedure of the satellite HCHO column densities. Several potential causes have been highlighted in literature (e.g., Vigouroux et al. 2020) and include aerosol effects, spectral interferences with O_3 and BrO at high solar zenith angles, and a coarse albedo climatology. The use of averaging kernels in the validation studies is expected to mitigate the impact of the a priori profile on the bias.

The bias in TROPOMI HCHO column densities has a large impact on emission inversion studies (Muller et al. 2024). Because the satellite measures too low HCHO columns over source regions, top-down VOC emissions derived from an emission inversion are also too low. The satellite bias can be mitigated by applying a linear correction derived from a comparison with ground-based HCHO vertical columns. For this purpose, the FTIR NDACC network can be used for the derivation of the bias correction, since FTIR measurements have a similar vertical measurement sensitivity compared to TROPOMI observations. Therefore, FTIR data are well suited for a TROPOMI validation. For the CAMEO project, we derive a bias correction based on an assessment of TROPOMI data similar to Vigouroux et al. (2020) except that it has been applied to FTIR data from the operational version of the TROPOMI HCHO product (v2.04) over an extended time range (2018-2023). The data are co-located, i.e. only TROPOMI scenes within 20 km of the station location are averaged and only FTIR data taken within 3 hours of the TROPOMI overpass time (i.e. near 13:30 local time) are selected. Additionally, the FTIR data are smoothed by substituting the a priori profile and subsequently applying the TROPOMI averaging kernel to the modified column as described in Vigouroux et al. (2020). The results of this validation study are described in the quarterly validation reports of the S5p operational data products (https://mpc-vdaf.tropomi.eu/ProjectDir/reports//pdf/S5P-MPC-IASB-ROCVR-23.00.00_FINAL_signed.pdf, date of last access: 25/07/2024).

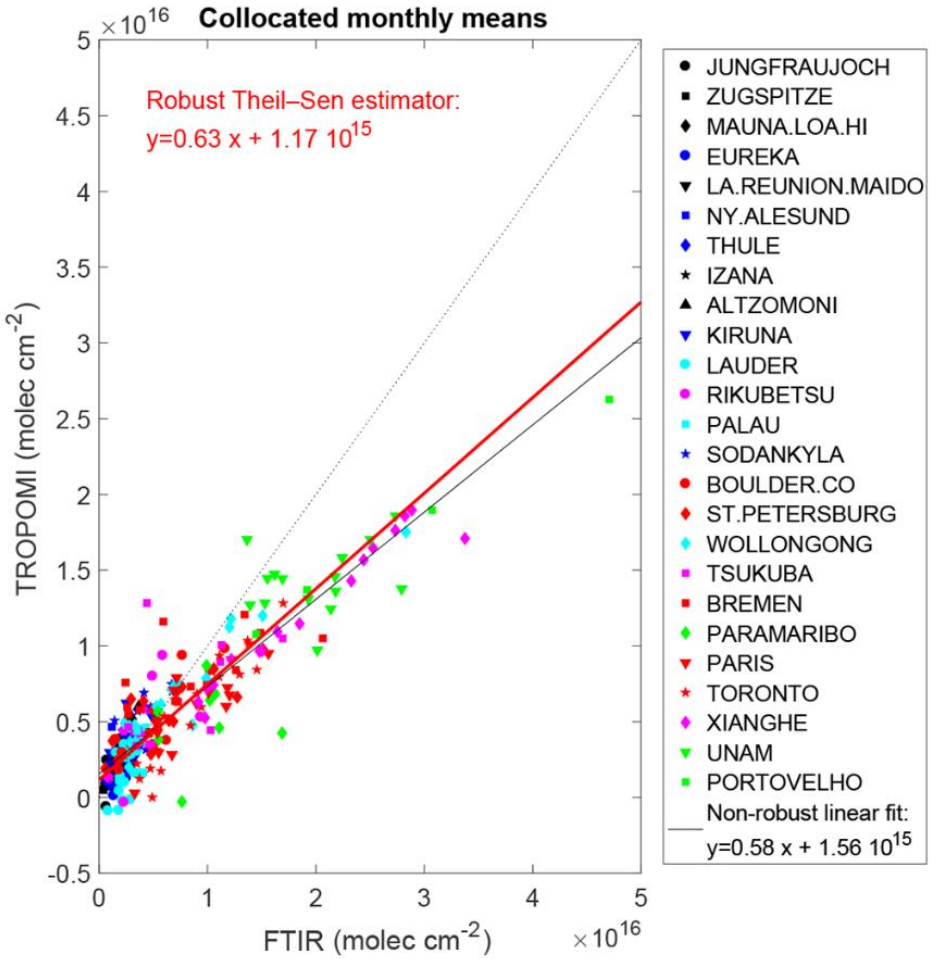


Figure 1: Scatter plot of TROPOMI HCHO vertical columns versus FTIR HCHO vertical columns from the FTIR network. The linear fit is calculated using a Theil-Sen estimator. This figure is adopted from Vigouroux et al. (2020) using TROPOMI v1.1.5 and FTIR data up to 2020. An updated linear fit using TROPOMI v2.04 dataset and FTIR measurements up to 2023 yields the regression y = 0.62 x+1.14×10^{15} molec cm^{-2}.

Figure 1 shows a scatter plot of TROPOMI and FTIR HCHO columns. High TROPOMI HCHO columns are lower than high FTIR columns by 40% on average. A linear fit of the scatter plot (more precisely, the inverse relationship of this fit) provides the following equation for the bias correction:

$\Omega_{BC} = 1.61 \times \Omega - 1.84 \times 10^{15}$ molec cm^{-2},

in which Ω and Ω_{BC} are the retrieved and bias-corrected TROPOMI HCHO columns in molec cm^{-2}, respectively. The application of this correction leads to strongly increased columns over source regions, and a reduction of the columns over remote, clean regions. This is shown in Fig. 2, which shows a map of the TROPOMI HCHO columns before and after application of the bias correction.

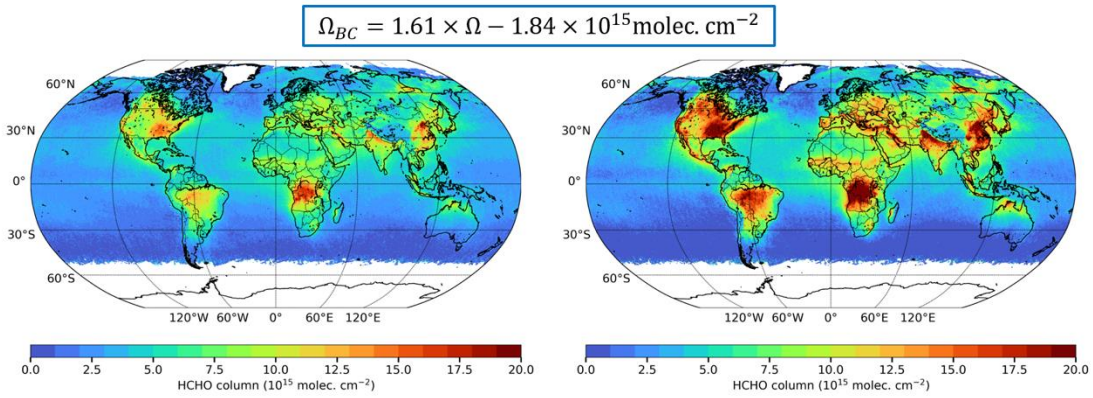


Figure 2: TROPOMI HCHO columns in July 2021 before (left panel) and after (right panel) application of the bias correction, shown in yellow above the figure. The bias correction leads to strongly increased HCHO columns over source regions (Tropical forests, North America, Eastern China, …). The bias correction leads to a small decrease of the columns over remote, pristine regions.

The impact of this bias correction on the comparison with IFS-COMPO HCHO simulations is briefly discussed below. The IFS-COMPO HCHO simulations (experiment IDs *icir*, *ieg6*) were performed based on CAMS cycle 48R1 at T511 horizontal truncation (spatial resolution of ~40km x 40km) for 137 model levels using year-specific emissions based on CAMS_GLOB_BIO v3.1 and CAMS_GLOB_ANT v5.3 and biomass burning emissions from the Global Fire Assimilation System (GFASv1.4). Simulations and satellite observations were averaged over a 1-month period. The analysis omits negative satellite retrievals to prevent a false positive model bias.

A direct evaluation of IFS-COMPO HCHO against not bias corrected and bias corrected TROPOMI HCHO shows a significant degradation of the data agreement when applying the bias correction (fig. 3). This appears true for both the background atmosphere (e.g. marine regions) as well as formaldehyde source regions (e.g. the tropics), with the only exception being the South American rain forest region between -5° and 10°North (fig. 4). For regions with a negative model bias relative to uncorrected TROPOMI HCHO (i.e. lower simulated than observed HCHO), the bias correction appears to increase TROPOMI HCHO further. This leads to an even larger underestimation of HCHO by IFS-COMPO (fig. 3, blue regions). Towards the poles, HCHO has no major surface sources, such that atmospheric VCDs decrease well below 3x10^{15} molec cm^{-2}. As a result, in these regions the bias correction mainly decreases TROPOMI HCHO. This leads to a positive model bias, i.e. larger simulated than observed HCHO, as compared to not bias corrected TROPOMI HCHO (fig. 3, red areas). An exception to above findings is the rainforest region north of the Amazon river delta, where IFS-COMPO overestimates HCHO as compared to the satellite retrievals (fig. 4). This overestimation is reduced by the TROPOMI HCHO bias correction, since the latter increases satellite retrieved HCHO over most terrestrial regions. As a result, between 10.5°South and 11.5°North and 58° to 78°West, the TROPOMI bias correction reduces the positive bias of modelled HCHO from 37% to 3% (fig. 4, rectangle 1). However, the negative model bias in surrounding areas increases due to the bias correction from 13% to 36% (-28° and -4.5°North and 41° to 58°West; fig. 4, rectangle 2).

Potential reasons for the biases between model and satellite observations are currently unknown and deserve more study. A predominance of model underestimations against bias-corrected satellite HCHO observations over most continental regions was also found in a recent model study (Müller et al., 2024, see Fig. 1). Over sources regions, the model discrepancies (under- and overestimations) with respect to bias-corrected satellite HCHO data may be improved through inverse modelling of emissions. The question whether the

assimilation of HCHO in the CAMS system will make use of the TROPOMI HCHO bias correction in the future is currently under discussion.

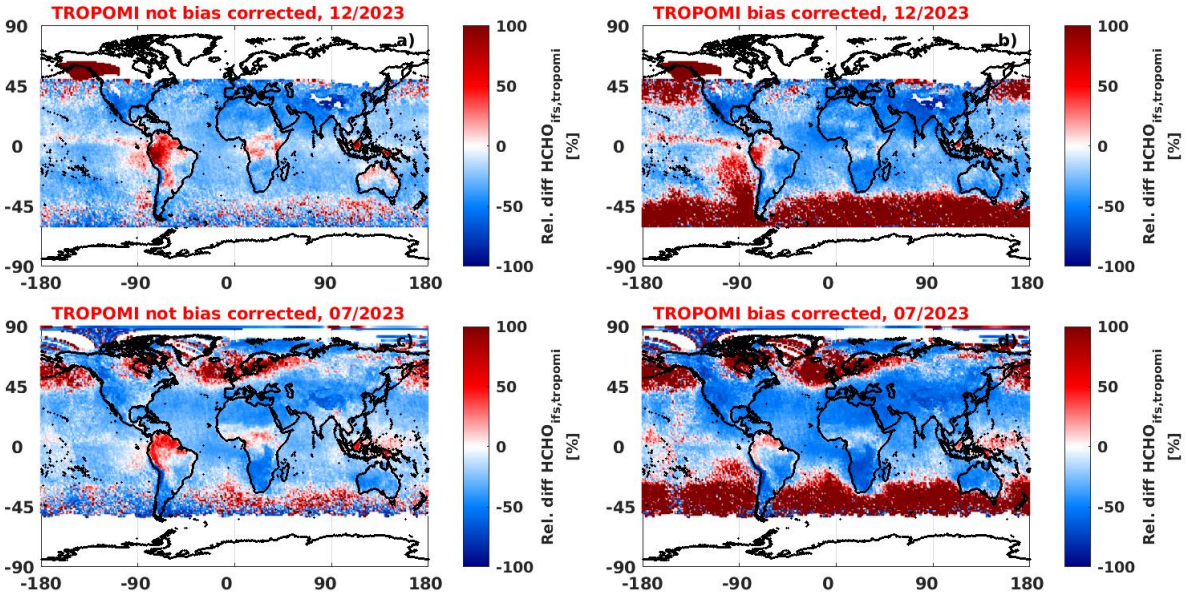


Figure 3: Relative difference of satellite retrieved and simulated HCHO based on CAMS 48R1 for December (panels a, b) and July (panels c, d) 2023 without bias correction of TROPOMI HCHO (panels a, c) and when applying the bias correction (panels b, d). Negative satellite retrievals are omitted to avoid a false positive model bias. Difference in % expressed relative to TROPOMI HCHO with red indicating larger simulated HCHO as compared to the retrievals and blue indicating smaller simulated HCHO as compared to the satellite observations. Clearly visible is the diverging pattern over large BVOC source regions as compared to most other terrestrial regions (e.g. north of the Amazon River Delta, fig. 4).

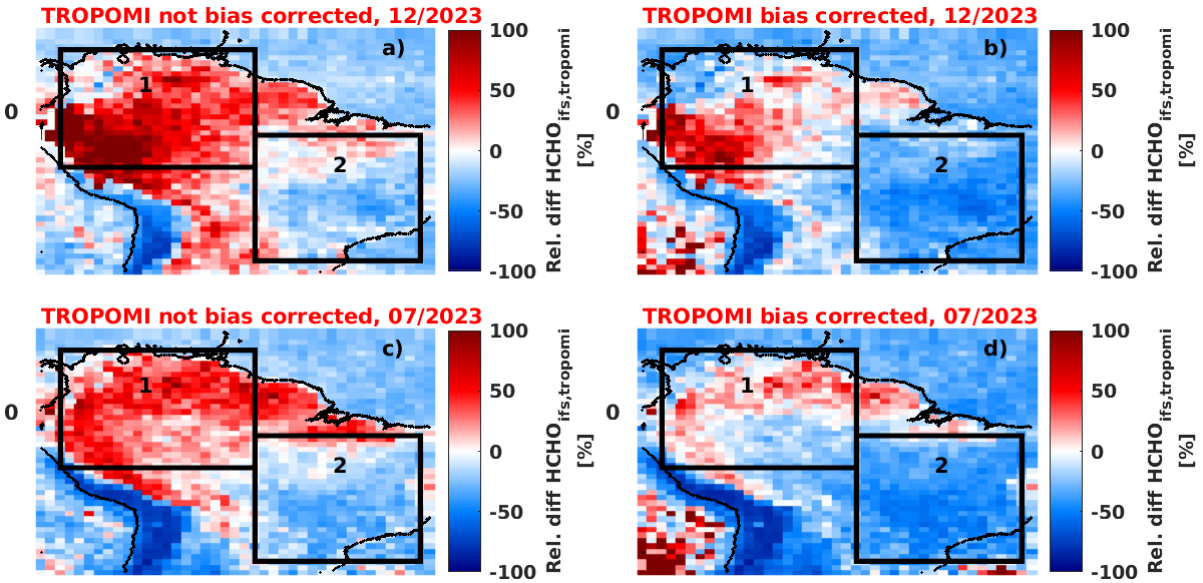


Figure 4: Zoom onto northern South America for December (top row) and July (lower row) 2023. Figure legend according to fig. 3.

3.3 Observational errors and filtering requirements

HCHO is a weak absorber in the ultraviolet (UV) and its absorption features are weaker than those of other compounds like NO_2. Furthermore, due to its short lifetime of several hours only, the region of interest for HCHO inversion studies is the planetary boundary layer, which is where most reactive VOC emissions and oxidation processes take place. However, in particular in the UV spectral range, scattering processes limit the amount of boundary layer photons that reach the satellite. This causes a decreased sensitivity of satellite instruments towards absorption processes in this atmospheric region. Additionally, the absorption features of HCHO are heavily blended with other UV-absorbing compounds like ozone and BrO. Consequently, the satellite retrieval has a relatively low sensitivity to the altitude range that is most relevant to top-down emission estimation. Due to the resulting substantial contribution of the HCHO background (from methane oxidation) to the spaceborne HCHO columns, large uncertainties in the HCHO-derived VOC emissions are expected, especially when using daily data.

The TROPOMI HCHO product described in De Smedt et al. (2018) provides random and systematic uncertainties related to numerous uncertainty sources in the retrieval. The random error is associated to measurement noise and is assumed to be on the order of 10^{16} molec cm^{-2} on the slant column fit, whereas other error contributions from absorption cross-sections, albedo, clouds, aerosols, etc., are attributed to be systematic errors. The systematic observational errors can be partly mitigated by the bias correction described in sect. 3.2. However, the derivation of the bias correction also introduces an uncertainty, as the validation dataset is limited in size and has a large scatter (fig. 1).

In terms of filtering requirements, we follow the recommendations provided in the TROPOMI ATBD (https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-ATBD-HCHO-TROPOMI, date of last access: 25/07/2024). We impose a quality flag larger than 0.5. The latest version of the TROPOMI HCHO dataset (v2.04) includes an updated cloud treatment, hence no additional cloud filter is required, as opposed to previous versions of the product (e.g. De Smedt et al. 2021).

3.4 Evaluation of IFS-COMPO model simulations against HCHO retrievals: assessment and interpretation of model biases

This section evaluates IFS-COMPO formaldehyde simulations using observational HCHO data from space (TROPOMI) and from aircraft observations (mini-DOAS instrument). All simulations were performed using year-specific emissions based on CAMS_GLOB_BIO v3.1 and CAMS_GLOB_ANT v5.3 .

3.4.1 Validation based on aircraft measurements

The following section analyses IFS-COMPO formaldehyde simulations in different global regions, seasons, altitudes, and air masses based on aircraft observations in the years 2014 to 2019. The airborne formaldehyde measurements were performed by the remote sensing instrument mini-DOAS (Heidelberg Institute for Environmental Physics) from aboard the German research aircraft DLR-HALO operated by the Deutsches Zentrum für Luft- und Raumfahrt in Oberpfaffenhofen (Germany). The mini-DOAS instrument simultaneously performs Nadir and Limb measurements of atmospheric straylight. While the Nadir retrievals are used to infer vertical column densities of the target trace gases in the atmosphere below the aircraft, the Limb soundings additionally provide altitude resolved information on the vertical distribution of the gases.

The presented measurements were performed during over 70 research flights as part of the scientific missions (1) ACRIDICON-CHUVA (Aerosol, Cloud, Precipitation, and Radiation

Interactions and Dynamics of Convective Cloud Systems-Cloud Processes of the Main Precipitation Systems in Brazil) over the Amazon in fall 2014 (Wendisch et al., 2016), (2) the Effect of Megacities on the Transport and Transformation of Pollutants on the Regional to Global Scales in Europe (EMeRGe-EU) in summer 2017 (Andrés Hernández et al., 2022), (3) Wave-driven Isentropic Exchange (WISE) over the North Atlantic and Europe in fall 2017, (4) the Effect of Megacities on the Transport and Transformation of Pollutants on the Regional to Global Scales in Asia (EMeRGe-Asia) in early spring 2018, (5) the Carbon Dioxide and Methane Mission (CoMet) over Europe in late spring 2018 (Fix et al., 2018), (6) Chemistry of the Atmosphere: Field Experiment in Africa (CAFE-Africa) over the tropical Atlantic and West Africa in summer 2018, and (7) Transport and Composition of the Southern Hemisphere UTLS (SouthTRAC; where UTLS is the upper troposphere and lower stratosphere) in Patagonia/western Antarctica in fall 2019 (Rapp et al., 2021).

Detailed descriptions of each mission, the deployed instruments, and research objectives can be found in the respective mission publications. More details of the mini-DOAS instrument design, its major features, and the deployment on the HALO aircraft, the measurement method, the spectral retrieval, and the data processing can be found e.g. in Hüneke et al. (2017). The discussed IFS-COMPO experiments are i0o6 (ACRIDION-CHUVA, Southtrac), i2h3 (EMeRGe-Asia, CoMet, CAFE), i538 (EMeRGe-EU), and i53a (WISE).

The large geographical coverage of the research missions provides airborne formaldehyde observations from altitudes close to the surface up to the lower stratosphere in pristine terrestrial environments (southern part of South America), anthropogenically polluted air masses (Central Europe and Taiwan), regions dominated by strong biogenic emissions (Amazon rainforest) as well as the remote and coastal marine atmosphere over the North, Tropical and South Atlantic, the Mediterranean Sea and the East China Sea, and finally also extensive probing of biomass burning plumes (West African coast and Amazon rainforest, fig. 5).

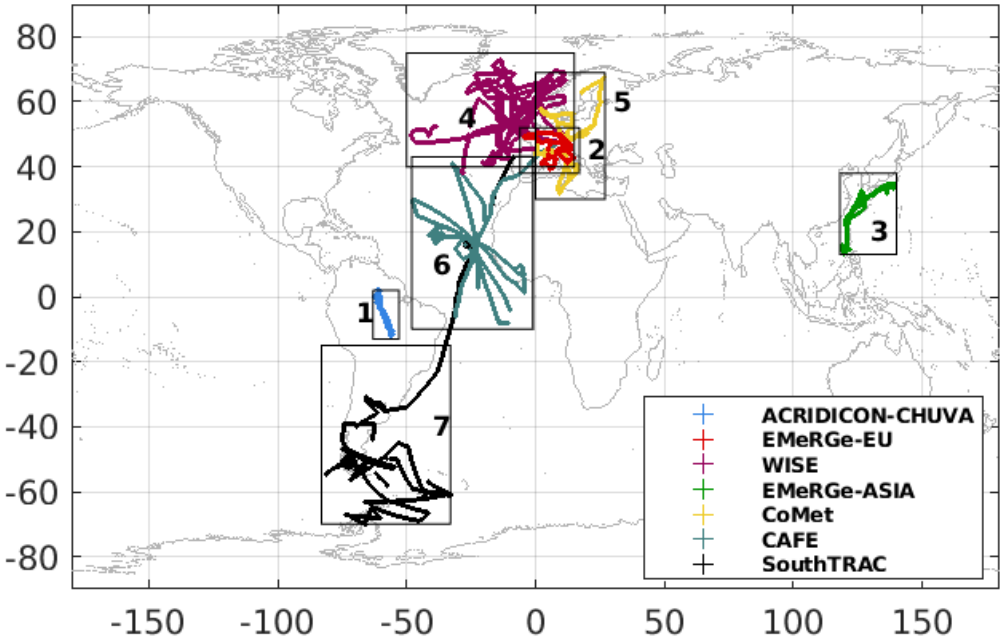


Figure 5: mini-DOAS formaldehyde measurements performed with the German research aircraft DLR-HALO during the seven research missions ACRDICON-CHUVA (2014), EMeRGe-EU (2017), WISE (2017), EMeRGe-Asia (2018), CoMet (2018), CAFE-Africa (2018), and SouthTRAC (2019). The operational bases for the individual missions were Manaus

(Brazil) for ACRIDICON-CHUVA, Oberpfaffenhofen (Germany) for EMeRGe-Europe and CoMet, Shannon (Ireland) for WISE, Tainan (Taiwan) for EMeRGe-Asia, and Rio Grande (Argentina) for SouthTRAC.

Figure 6 provides an overview of tropospheric formaldehyde profiles in the eight investigated regions. The IFS-COMPO simulations were performed based on CAMS cycle 48R1 at T511 horizontal truncation (spatial resolution of ~40km x 40km) for 137 model levels using year-specific emissions based on CAMS_GLOB_BIO v3.1 and CAMS_GLOB_ANT v5.3 and biomass burning emissions from the Global Fire Assimilation System (GFASv1.4). In a second step, the model output was spatially and temporally interpolated onto the aircraft flight tracks and subsequently all data within a specified region were averaged in 100m altitude steps. For the direct comparison of model and observations, it should be noted that the mini-DOAS Limb telescopes probe the atmosphere to the right of the aircraft with varying field of view length between a few km to over 100km, depending on the cloud cover during the spectral retrieval (Kluge et al., 2020). As a result, the observations provide an averaged concentration of formaldehyde over a slightly different atmospheric scenery as compared to the model simulations, which are interpolated onto the aircraft flight position.

Over the different global regions, air masses of varying origins and compositions, and thus largely different formaldehyde sources and concentrations, were probed. This includes (1) pristine marine air, (2) pristine continental air, (3) biogenic air masses, (4) air affected by fresh or aged anthropogenic emissions, and (5) biomass-burning affected air of different ages. The following section highlights the main findings for each air mass type.

1. Pristine marine air

   The most pristine marine air masses were observed in the troposphere over the South Atlantic and northern Antarctica. The comparison of formaldehyde vertical profiles over such remote marine regions shows a good agreement between simulations and observations (figs. 6 and 7, panel e). In comparison to terrestrial regions, the marine profiles are significantly more constant, without significant enhancements towards lower altitudes due to the lack of surface sources. For such air masses, methane oxidation presents the main formaldehyde source. This creates a nearly constant formaldehyde background throughout the troposphere of 18±18ppt (observations) and 16±3ppt (IFS-COMPO), respectively, with a minor decrease towards higher altitudes mainly caused by the temperature dependence of methane oxidation. Within the marine boundary layer over the other marine regions, formaldehyde mixing ratios increase towards the surface. This is mainly a consequence of the performed flight pattern. Over the Mediterranean Sea and the North Atlantic, lower altitudes were mainly probed in close vicinity to the coasts and when flying from and towards local airports. As a consequence, the probed air masses were affected by continental outflow and are not characteristic for pristine marine air (see point 4 for a detailed discussion of this data). Formaldehyde in the upper troposphere over the Mediterranean Sea and the Tropical and North Atlantic is found to be comparable to the observations over the South Atlantic, with median mixing ratios between 8 and 23ppt (observations) and 15 to 63ppt (IFS-COMPO). This suggests methane as the main formaldehyde source in the free and upper troposphere over all probed marine regions. Over the East China Sea, most flights were performed within the boundary layer. Only two research flights reached higher altitudes up to ~10km. There, median formaldehyde of 191±91ppt (obs.) and 171±15ppt (IFS-COMPO) suggest additional formaldehyde sources beside methane, most likely transported pollution plumes from China mainland or Taiwan. This is notable since such enhanced upper tropospheric formaldehyde was not observed in any of the other marine regions, where formaldehyde was roughly 150ppt lower above 8km altitude. In general, one can

conclude that airborne and simulated formaldehyde agree well in pristine marine air, with deviations mostly within the uncertainty range.

Noteworthy are the observations within the tropical marine boundary layer (fig. 6, panel f). These were partly derived by flight sections towards and from local islands, but also from dives of the aircraft in large distance from the coasts. Lower altitude measurements above the North Atlantic and Mediterranean Sea yield comparable formaldehyde mixing ratios, but were mostly performed in close vicinity to the coasts and hence more strongly affected by fresh anthropogenic outflow. Elevated concentrations of VOCs in the tropical marine boundary layer distant from the coasts have previously been observed by numerous studies (e.g. Weller et al., 2000) and also translate into enhanced VCDs in that region, which are observable by satellite instruments like TROPOMI (e.g. De Smedt et al., 2021). Potential sources for the observed formaldehyde in the tropical marine boundary layer range from sea surface emissions of precursor carbons to (more importantly) long-range transport processes mainly originating from West Africa mainland. IFS-COMPO underestimates this enhancement of formaldehyde in the tropical boundary layer (fig. 7, panel f). Instead, the vertical profiles of simulated formaldehyde over North and Tropical Atlantic are similar. This causes a comparably larger deviation of simulations and observations over the Tropical Atlantic as compared to the other remote marine regions.

2. Pristine continental air
   Continental air masses over regions without strong biogenic or anthropogenic formaldehyde sources were mostly found in southern South America. There, the lack of major population centres or large vegetational coverage reduces direct or secondary formaldehyde emissions to a minimum. As a result, the vertical profile over this region is similar to the South Atlantic with only minor enhancements towards lower altitudes (fig. 6, panel a). Above 8km altitude, median formaldehyde mixing ratios are found to be similar to those in the marine upper troposphere, with 5±13ppt (obs.) and 16±3ppt (IFS-COMPO), respectively. Towards the surface, average formaldehyde increases to 183±107ppt (obs.) and 325±142ppt (IFS-COMPO). Due to the performed flight pattern, this formaldehyde enhancement can mainly be attributed to anthropogenic emissions of formaldehyde and/or formaldehyde precursors. In fact, the largest deviation of observations and IFS-COMPO is found during a flight section below 2km altitude in vicinity to Buenos Aires. There, IFS-COMPO simulates almost three times more formaldehyde than observed (289ppt versus 109ppt). This discrepancy causes the resulting higher median boundary layer formaldehyde from IFS-COMPO. However, the range of observed formaldehyde mixing ratios during this flight section is quite large (98ppt), such that observations and simulations still agree within twice the uncertainty range.

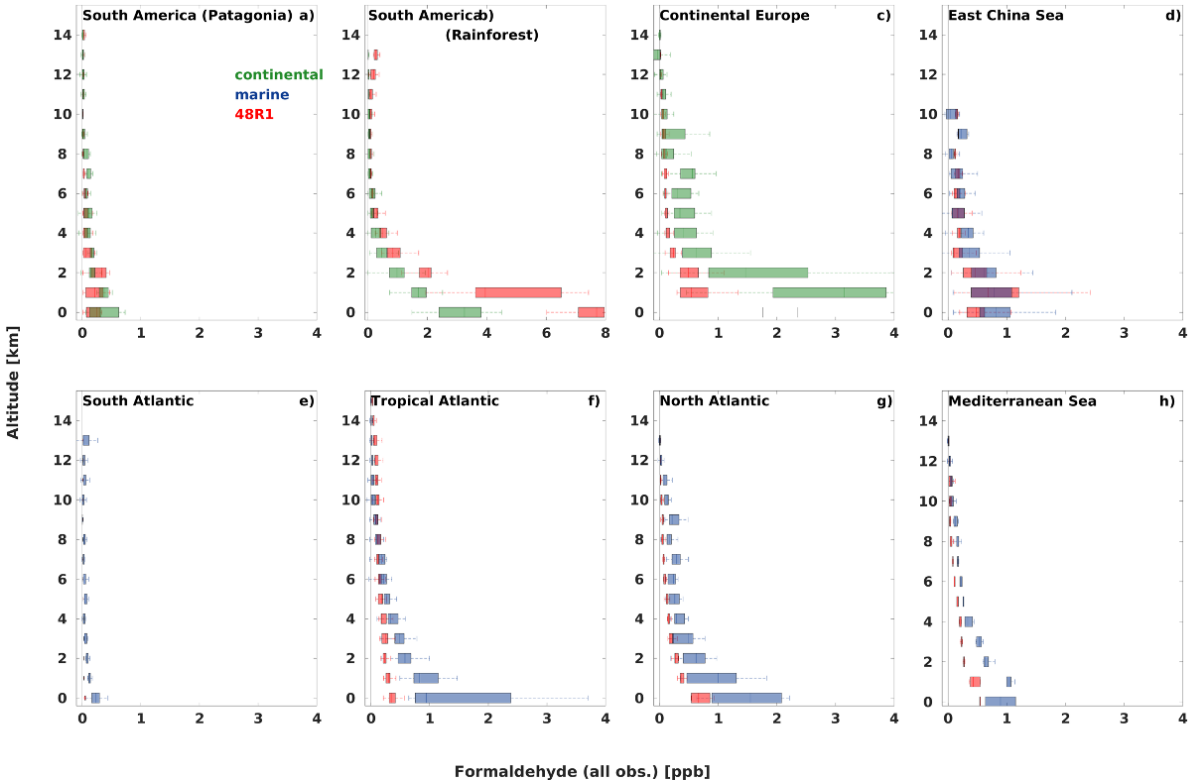


Figure 6: Tropospheric profiles of formaldehyde in the eight global regions based on IFS-COMPO formaldehyde simulations based on CAMS cycle 48R1 (red) and airborne observations (green, blue).

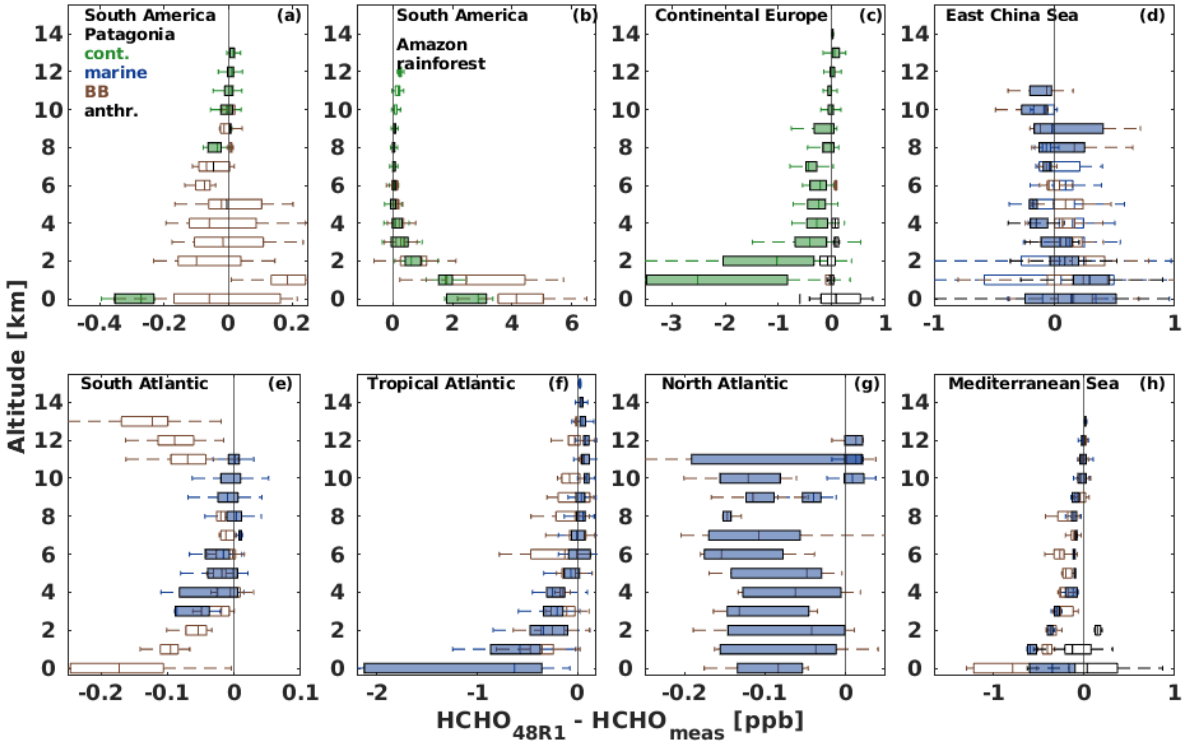


Figure 7: Difference of simulations and observations in each region for background air (full bars) and for identified pollution plumes (empty bars, biomass burning in brown and anthropogenic plumes in black). Plumes were identified based on visual imagery and collocated measurements of respective markers.

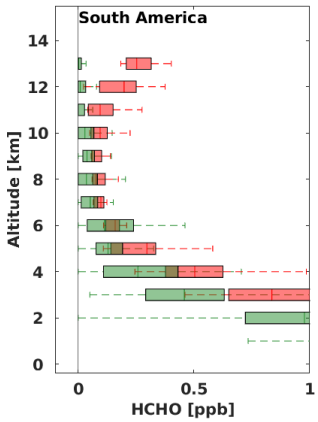


Figure 8: Zoom into fig. 6, panel b, with airborne observations (green) and IFS-COMPO simulations (red). Clearly visible is the increase in simulated formaldehyde above 10km altitude.

3. Biogenic air masses

On a global scale, most atmospheric formaldehyde is formed as an intermediate oxidation species from the degradation of both methane and non-methane volatile organic compounds (e.g. Seinfeld and Pandis, 2013). Among the latter, isoprene is the predominant atmospheric precursor, with formaldehyde being formed both directly and secondarily during the oxidation of intermediate isoprene oxidation products. As a result, large formaldehyde concentrations are expected to be found above isoprene emission hotspots like tropical rainforests and such regions, e.g. the Amazon rainforest, are of particular interest when studying the emissions and atmospheric chemistry of VOCs.

As expected, aircraft observations and IFS-COMPO simulations both show strong formaldehyde enhancements towards the surface over the Amazon rainforest (fig. 6, panel b). Above 4km altitude, the mixing ratios quickly decrease to a median background of 27±63 ppt (obs.) and 131±88 ppt (IFS-COMPO), respectively. In the upper troposphere, IFS-COMPO formaldehyde displays enhanced mixing ratios above 10km altitude, possibly as a result of deep convection of NMVOCs to the upper troposphere (fig. 7, panel b and fig. 8). This enhancement is not seen in the observations.

Within the planetary boundary layer, formaldehyde mixing ratios reach maxima of 4.5 ppb (obs.) and 8.1 ppb (IFS-COMPO). The significant overestimation of formaldehyde by IFS-COMPO (CAMS cycle 48R1) leads to three times higher median mixing ratios of 6.5±1.7 ppb (IFS) as compared to the aircraft observations (2.0±0.8 ppb). The apparent even larger difference with respect to the observations when probing biomass burning affected air masses (fig. 7, panel b) is caused by the period chosen for the IFS-COMPO simulations. While individual model experiments were run for each of the other research missions (2017 – 2019), a climatological comparison is performed for ACRIDICON-CHUVA (2014) based on simulations for the year 2019. As a result, we do not expect IFS-COMPO to reproduce the wildfire situation in September 2014. The apparent mismatch of simulations and observations in the tropical air undisturbed by fires, however, is of interest. The overestimation of formaldehyde by CAMS cycle 48R1 in the northern part of the Amazon region can also be observed when comparing the simulations to respective satellite observations. This is discussed in more detail in sections 3.4.2 and chapter 5. We note that the isoprene oxidation chemistry of earlier CAMS cycles was replaced by a more explicit scheme in CAMS cycle 48R1 (Williams et al., 2022). It includes more intermediate oxidation products (e.g. glyoxal) based on Stavrakou et al. (2010), Lamarque et al. (2012), and Myriokefalitakis et al. (2020).

Overall, the new scheme leads to a higher formaldehyde production. This has led to a reduced negative bias of IFS-COMPO in regions where formaldehyde was previously underestimated. At the same time, we find an increased positive bias over the main isoprene source regions, i.e. the tropics. There, the increase in formaldehyde causes a large positive model bias of ~60% relative to not bias corrected TROPOMI observations (Eskes et al., 2024). While the satellite evaluation analyses total VCDs, the presented aircraft comparison of boundary layer formaldehyde concentrations shows an even larger positive bias of a factor of three. This is plausible, since most isoprene chemistry takes place in the lowest atmospheric altitudes, such that the impact of the updated isoprene chemistry scheme can be expected to be largest close to the surface. However, while the more explicit chemistry scheme may enlarge an already existent model bias, it does not fully explain the observed deviations.

4. Air masses affected by fresh or aged anthropogenic emissions

   Air masses affected by anthropogenic activities, such as large population centres or extended industrial activities were probed in varying extent during all research missions and in all discussed regions. Fresh pollution plumes were mostly encountered when flying above large cities (e.g. Munich, Marseille, Buenos Aires) or industrial areas (North Italian Po valley and Upper Silesian Coal Valley). Aged pollution plumes that had been transported downwind were frequently observed over the East China Sea and when probing the outflow of large coastal pollution centres along the marine coastlines.

   Formaldehyde enhancements due to anthropogenic emissions are found to be comparable and even larger than formaldehyde from biogenic sources, however much more locally confined. Over Continental Europe, a mixture of anthropogenic and biogenic direct and secondary formaldehyde production causes high formaldehyde background in the planetary boundary layer of 3.1±1.5 ppb (obs.) and 0.5±0.3 ppb (IFS-COMPO), respectively. Evidently, there is a significant bias of airborne and simulated formaldehyde in this air mass type. The bias decreases when directly probing anthropogenic pollution plumes (fig. 7, panel d) and is more pronounced in the mixed polluted background air. Interestingly, IFS-COMPO not only underestimates absolute formaldehyde, but also the range of observed mixing ratios at a given altitude. This reflects in much shorter bar sizes of IFS-COMPO formaldehyde at a given altitude range in fig. 6, panel c). The mixture of probed air masses over the East China Sea is expected to be similar to those over Europe. However, simulations and observations agree much better in that region (fig. 6, panel d and fig. 7, panel c). At least to a part, the large bias over continental Europe may be a result of extensive probing of the Upper Silesian Coal Valley (USCV, Poland). Multiple research flights were directed to this region, where over 50 coal power plants are located in close vicinity, causing significant anthropogenic emissions. The aircraft instrument measured strongly elevated formaldehyde mixing ratios up to 4.5ppb in the lowest altitudes above the USCV (fig. 9, panel c). IFS-COMPO simulations are more than four times lower (1.1ppb). The comparably large quantity of performed measurements over the USCV (where many of the available European low-altitude measurements were performed there) reflects in the average Continental European profile (fig. 6) and consequently also the related strong negative bias of IFS-COMPO.

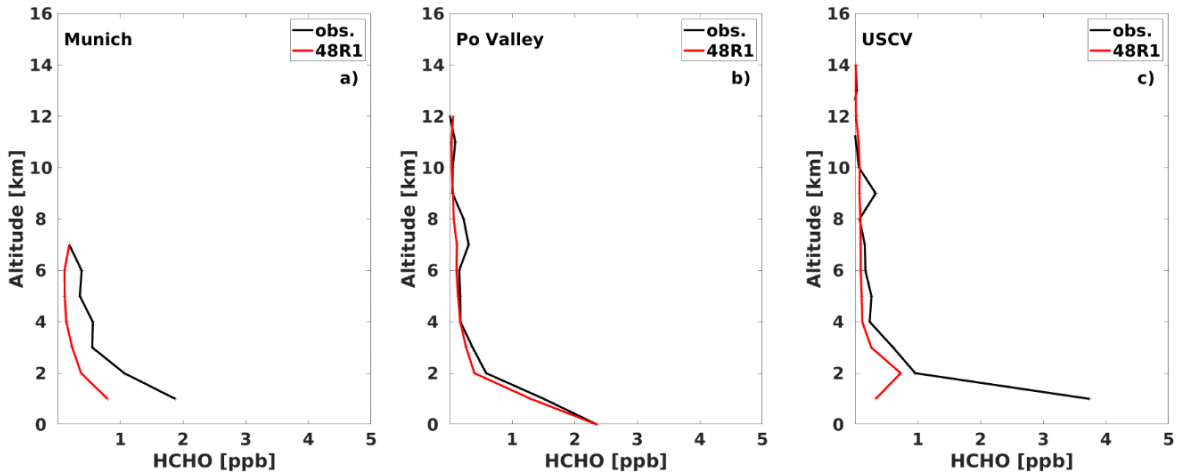


Figure 9: Averaged vertical formaldehyde profiles above Munich (Germany, panel a), the north Italian Po Valley (panel b) and the Upper Silesian Coal Valley (panel c) based on aircraft observations (black) and model simulations from IFS-COMPO (CAMS cycle 48R1, red).

5. Biomass-burning affected air

   Most research missions took place during local summer periods, such that biomass burning plumes were frequently encountered. Over the Amazon rain forest and neighbouring agricultural areas, the aircraft probed multiple plumes of varying ages and extents. Formaldehyde is both directly emitted by incomplete combustion processes and secondarily formed by precursor trace gases while the plumes are transported downwind. All plume encounters are characterized by significant enhancements of formaldehyde mixing ratios relative to the local background. Over the Amazon, simulated formaldehyde usually exceeds the observations (fig. 7, panel b), in agreement with the general HCHO overestimation found in this region. Biomass burning plumes in other global regions show much better agreement. Extended wildfires in the Portuguese Algarve in August 2018 caused large biomass burning plumes that were crossed by the aircraft on 7 August 2018 (fig. 10, panel a). IFS-COMPO simulations for the same day show comparable enhancements in formaldehyde up to 8×10^{15} molec cm^{-2}, while TROPOMI observes slightly higher VCDs of 16×10^{15} molec cm^{-2} during its local overpass three hours earlier.

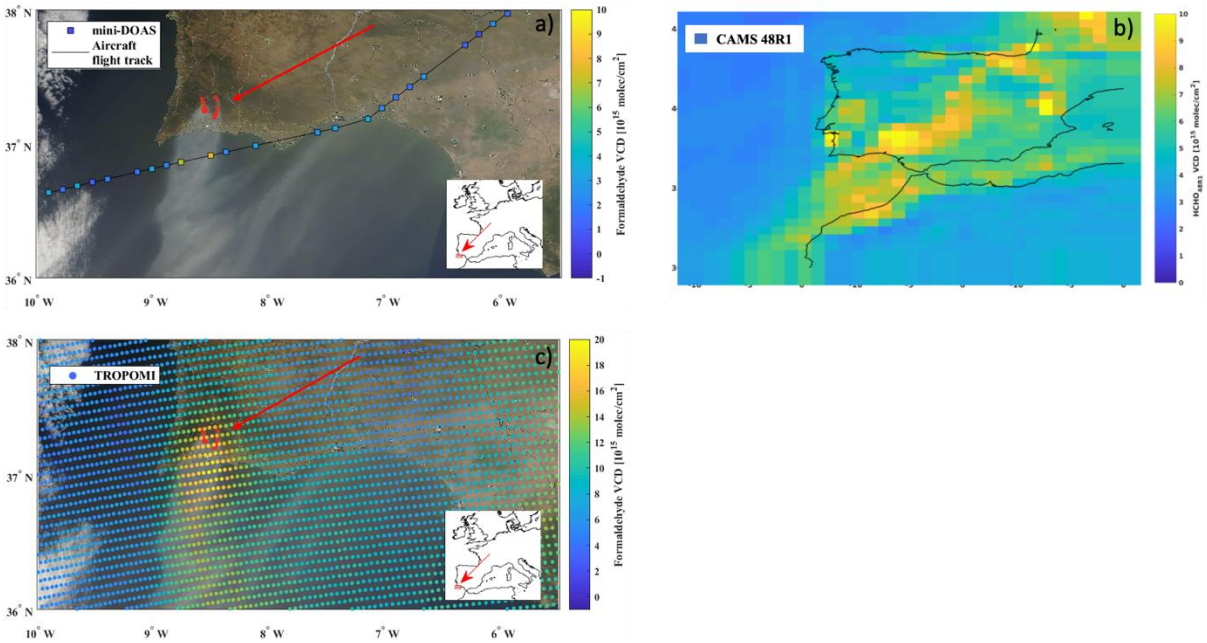


Figure 10: Probing of a large biomass burning plume from wildfires in the Serra de Monchique (Algarve, Portugal) detected by the mini-DOAS instrument (panel a, at 13 km flight altitude) and TROPOMI (panel c) over the Gulf of Cádiz at 11:00 and 13:50 UTC, respectively, on 7 August 2018. IFS-COMPO simulations for the same day show multiple biomass burning events in the surrounding areas as well as extended transport of their plumes over the Gulf (panel b).

3.4.2 Validation based on satellite measurements

Evaluation procedure

To effectively compare IFS-COMPO model output against TROPOMI-retrievals, KNMI developed the METRO-tool (Model Evaluation against TROpomi; Douros et al., 2023). This is a python/shell package to

1. Import model and TROPOMI data from ECMWF's MARS-archive,
2. Post-process and compare the data.

The METRO-package has been expanded to handle HCHO, along with NO_2, CO and glyoxal (CHOCHO). The package is freely available to all who are interested (miro.van.der.worp@knmi.nl).

In this study, the METRO-tool was used to evaluate CAMS CY49R1 simulations against TROPOMI formaldehyde vertical column densities (VCDs), using the reprocessed data version 2.1. The TROPOMI bias correction as suggested by BIRA is used (section 2.3.1b). The TROPOMI averaging kernels for HCHO are applied to the simulated mixing ratios. This minimises the dependency on the a priori profile assumptions based on the TM5-model that is used in the TROPOMI data processing. Model and retrieval data are collocated in time and on a common 0.5x0.5° output grid, using only observational data with a quality assurance threshold larger than 0.5 (i.e. cloud-free pixels). Furthermore, to ensure sufficient sampling, a grid box is used only if more than 50% of the output grid box is covered by observations and discharded otherwise. The local overpass time of TROPOMI is 13:30LT. Consequently, the presented analysis focuses on local midday.

IFS-COMPO model evaluation

The IFS-COMPO model has been evaluated for the year 2019. The experiment uses the standard chemistry as planned for the operational version of CY49R1. This is based on a modified version of the CB05 scheme in the troposphere and BASCOE chemistry in the stratosphere (Williams et al., 2022). Emissions are taken from CAMS-GLOB-ANT v6.1 for the anthropogenic emissions, CAMS-GLOB-BIO v3.1 for biogenic emissions (both specific for the year 2019) and GFASv1.4 for biomass burning emissions. The model experiment has expid *b2j1*. This experiment is evaluated against the reprocessed TROPOMI-measurements (version 2.04.01). The discussion will focus on three main global regions, which are a) the Amazon rainforest, where BVOC emissions are strong and dominate over anthropogenic emissions, b) Europe, where anthropogenic contributions to formaldehyde concentrations are significant, and c) Australia, which is both very remote, making it a clean case study, and contains some high BVOC-emitting areas like the north coast.

1. Amazonia

Over the Amazon, formaldehyde is predominantly produced by the oxidation of biogenic volatile organic compounds (BVOCs). The Amazon has two dominant seasons, which are the rain season from November to March and the dry season from April to October. The fire season in Amazonia roughly goes from July to November.

Figures 11 and 12 show a direct comparison of retrieved and simulated HCHO vertical column densities as well as their bias averaged over the months January, November and December 2019.

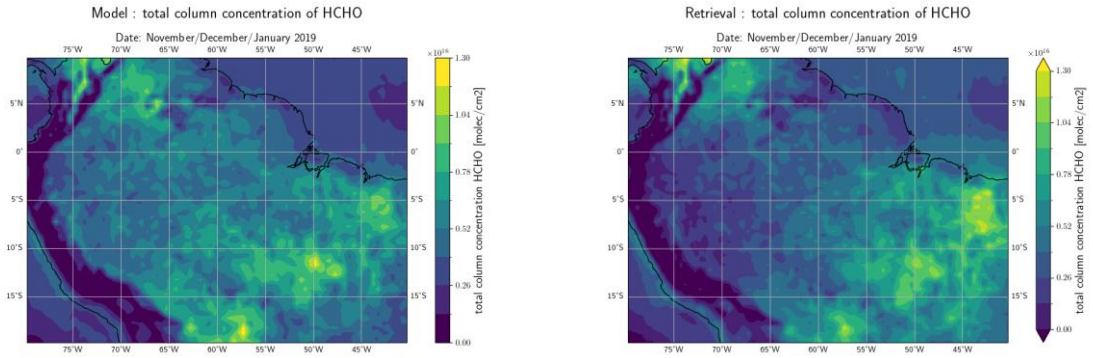


Figure 11. Collocated averaged HCHO VCDs [molecules cm^{-2}*] from IFS-COMPO (left) and TROPOMI (right) over the Amazon for the period November/December/January 2019 at the local satellite overpass time (~13:30LT).*

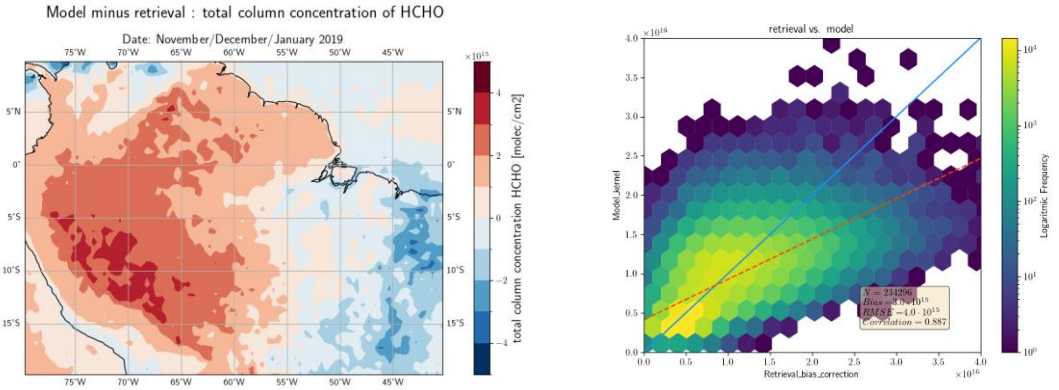


Figure 12: The mean model bias (model – retrieval) of the total column HCHO over the Amazon for the period November/December/January 2019 in molecules cm^{-2} (left panel) and scatterplot of the same data (right panel). A linear fit to the data is shown as red line.

The analysis reveals a positive bias of IFS-COMPO over Amazonia during the wet season, with an average bias of 2.7 x 10^{15} molecules cm^{-2}. Notably, low TROPOMI VCDs (<0.7 x 10^{16}) are overestimated by IFS-COMPO, with the largest HCHO overestimation observed over Western Amazonia. There, the observed total column is around 0.4 x 10^{16} molecules cm^{-2}. The highest mean VCDs during the wet season are found not over the remote Amazonian rainforest, but in its southeastern part and along its northern border. For instance, in the Brazilian state of Piauí, where the rainforest transitions to a mix of tropical forest, tropical savanna, and agriculture, the mean total column is approximately 1.2 x 10^{16} molecules cm^{-2}. This observed pattern is well captured by IFS-COMPO but is underestimated in absolute magnitude (fig. 12).

The research period during the Amazonian dry season has a small overlap with the start of the fire season (fig. 13). The maximum fire emissions over the Amazon take place during August and September, which are therefore excluded from the analysis. While biogenic sources still prevail, part of the observed formaldehyde enhancements may be related to forest fires instead of BVOC oxidation.

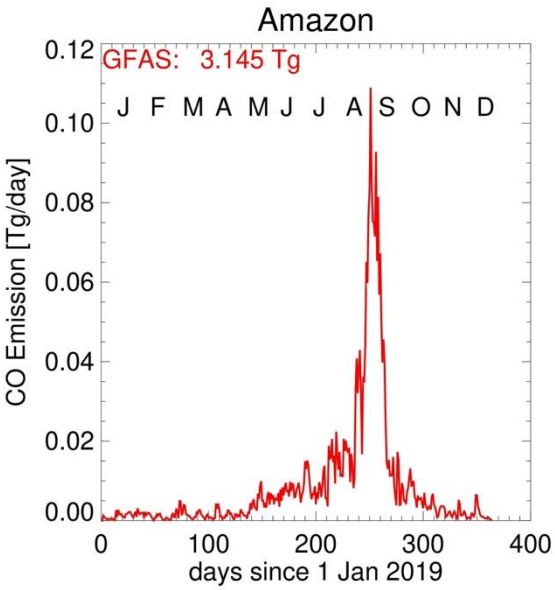


Figure 13: Time series of daily GFAS fire emissions of CO over the Amazon in 2019.

Figures 14 and 15 show a direct comparison of retrieved and simulated HCHO vertical column densities as well as their bias averaged over the months May, June and July 2019.

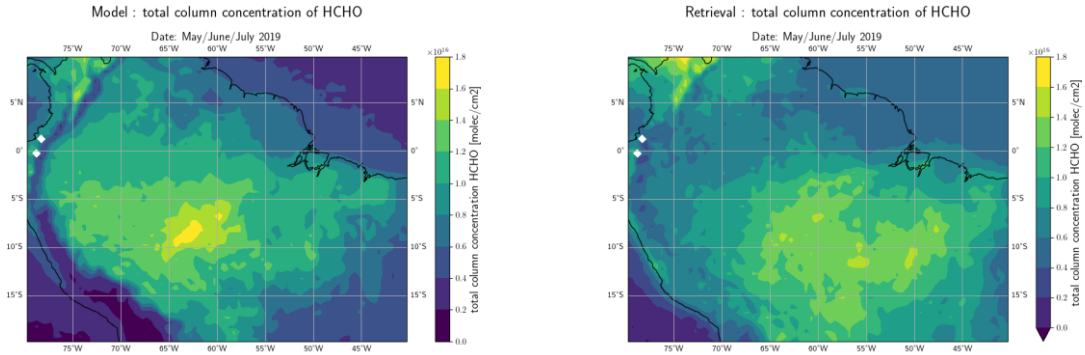


Figure 14. Collocated averaged HCHO VCDs [molecules cm^{-2}*] from IFS-COMPO (left) and TROPOMI (right) over the Amazon for the period May/June/July 2019 at the local satellite overpass time (~13:30LT).*

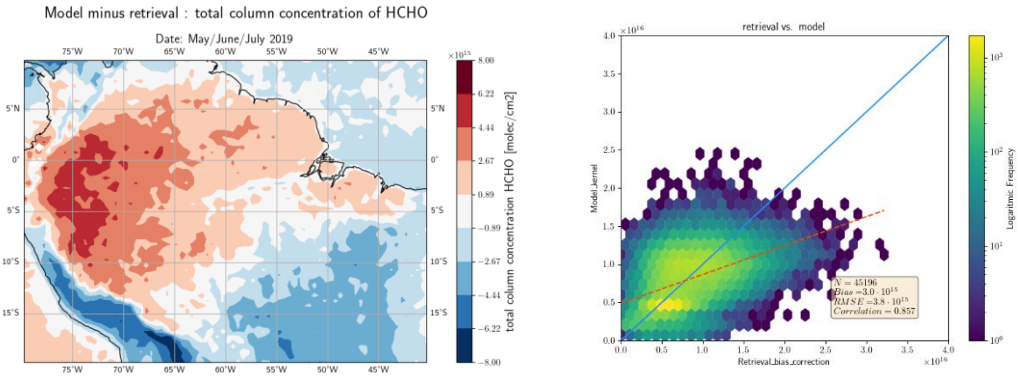


Figure 15: The mean model bias (model – retrieval) of the total column HCHO over the Amazon for the period May - July 2019 in molecules cm^{-2} (left panel) and scatterplot of the same data (right panel). A linear fit to the data is shown as red line.

The formaldehyde columns during the dry season are significantly higher than during the wet season with an average VCD of 1.8 x 10^{16} molecules cm^{-2} compared to ~1.0 x 10^{16} molecules cm^{-2} during the wet season. The highest average VCDs are now observed in the southern Amazon rainforest, while VCDs are much lower over the Brazilian tropical savannah to the south. IFS-COMPO captures this distribution well. It does, however, still significantly underestimate the VCDs above the Amazon rainforest, especially at the most western edge of the rainforest. There, the bias reaches 0.5 x 10^{16} molecules cm^{-2}. The mean bias of all land-based pixels is 0.3 x 10^{16} molecules cm^{-2}. This is slightly higher, but of the same order, as the mean bias during the wet season (0.37 x 10^{16} molecules cm^{-2}).

2. Europe

In comparison to the tropical rainforest, formaldehyde over Europe has significant contributions from anthropogenic sources for both the summer and winter season.
Figures 16 and 17 show a direct comparison of retrieved and simulated HCHO vertical column densities as well as their bias during the European winter. All VCDs are averaged over the months January, November, December of 2019 and are plotted at the local satellite overpass time.

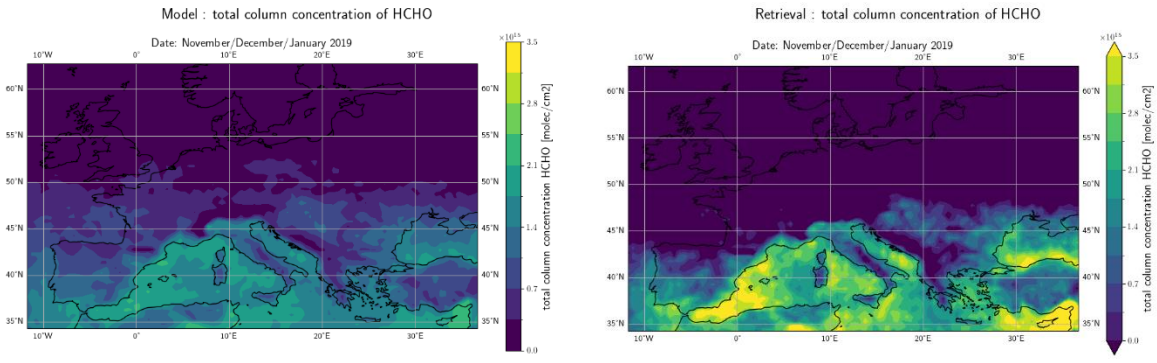


Figure 16. Collocated averaged HCHO VCDs [molecules cm^{-2}*] from IFS-COMPO (left) and TROPOMI (right) over Europe for the period November/December/January 2019 (winter) at the local satellite overpass time (~13:30LT).*

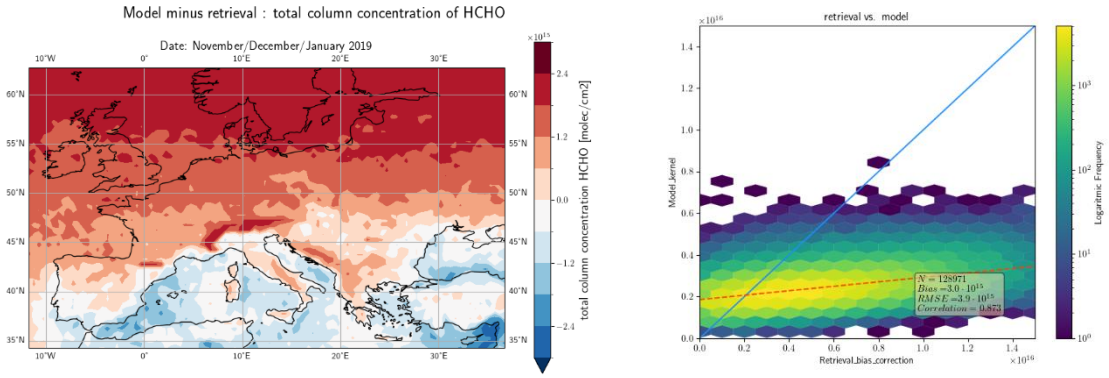


Figure 17: The mean model bias (model – retrieval) of the total column HCHO over the Europe for the period November - January 2019 in molecules cm^{-2} (left panel) and scatterplot of the same data (right panel). A linear fit to the data is shown as red line.

Observed formaldehyde VCDs during the European winter are of the order 0.3 x 10^{16} molec cm^{-2} over the Mediterranean Sea and quickly fall off with latitude. This is roughly five times lower than VCDs over the Amazon. IFS-COMPO shows a similar distribution to the observations but overestimates the magnitude across the European mainland. Only over the Mediterranean Sea IFS-COMPO underestimates formaldehyde VCDs. Highly polluted areas cannot be distinguished based on the observed formaldehyde pattern. The absolute mean bias of IFS-COMPO over Europe during the winter is 0.3 x 10^{16} molec cm^{-2} (sea-tiles excluded). This is similar to the bias over the Amazon.

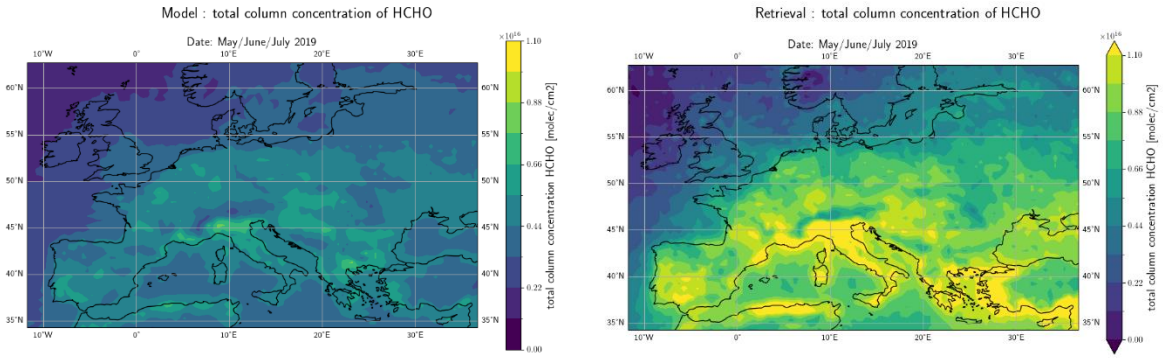


Figure 18. Collocated averaged HCHO VCDs [molecules cm^{-2}*] from IFS-COMPO (left) and TROPOMI (right) over Europe for the period May/June/July 2019 (summer) at the local satellite overpass time (~13:30LT).*

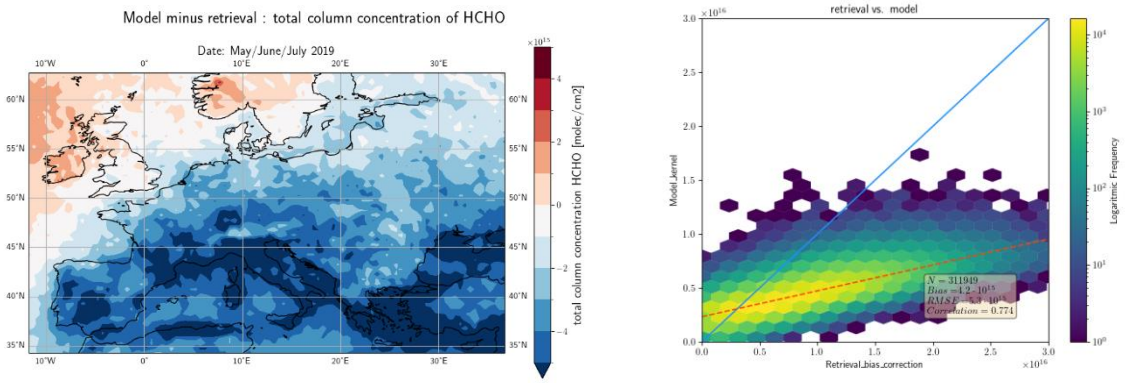


Figure 19. Left: The mean model bias (model – retrieval) of total column HCHO over the Amazon for the period May/June/July 2019 (summer) [molecules cm^{-2}*]. Right: scatterplot of model columns versus TROPOMI retrievals [molecules* cm^{-2}*]. The line: model equals retrieval is plotted in blue; the red-striped line is a linear fit to the data.*

During European summer (May – July 2019), formaldehyde levels over Europe are significantly higher as compared to winter. Observed levels are of the order of 0.5 x 10^{16} molecules cm^{-2} compared to 0.1 x 10^{16} molecules cm^{-2} during the winter. IFS-COMPO strongly underestimates the mean total column concentrations over Europe –especially along the Mediterranean coast. Observed values over northern Italy for example reach 1.0 x 10^{16} molecules cm^{-2} compared to model values of the order 0.7 x 10^{16}molecules cm^{-2}. The model bias gets smaller with latitude until it turns positive for southern Scandinavia, Ireland and Scotland. The measured formaldehyde VCD is distributed equally over the European mainland. Areas of strong anthropogenic activity such as Paris, Cologne and Milan show small increases in measured VCD's of the order of 0.1 x 10^{16} molecules cm^{-2} compared to the surrounding areas.

3. Australia

The formaldehyde concentrations over the Australia are predominantly produced by the oxidation of BVOCs on the tropical north coast and the temperate east coast. The Australian outback contains relatively few biogenic and anthropogenic sources.

Figures 20 and 21 show a direct comparison of retrieved and simulated HCHO vertical column densities as well as their bias during the Austral summer. All VCDs are averaged over the months January, November, December of 2019 and are plotted at the local satellite overpass time (13:30 LT).

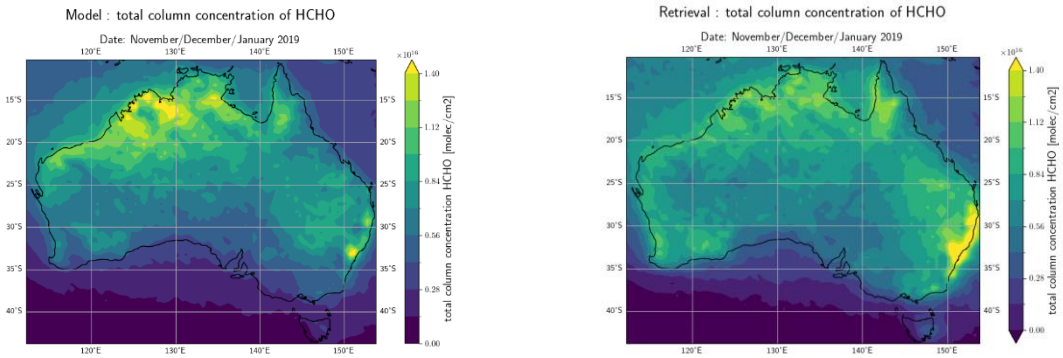


Figure 20. Collocated averaged HCHO VCDs [molecules cm^{-2}*] from IFS-COMPO (left) and TROPOMI (right) over Australia for the period November/December/January 2019 (austral summer) at the local satellite overpass time (~13:30LT).*

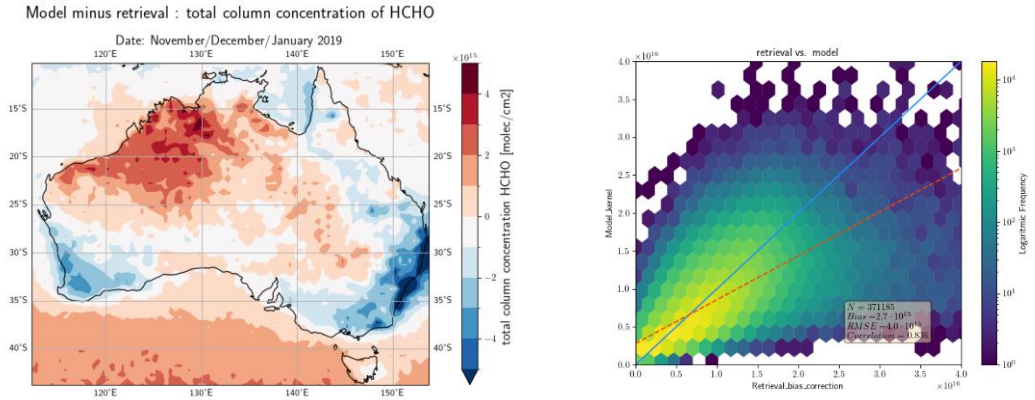


Figure 21. Left: The mean model bias (model – retrieval) of total column HCHO over Australia for the period November/December/January 2019 (austral summer) [molecules cm^{-2}*]. Right: scatterplot of model columns versus TROPOMI retrievals [molecules* cm^{-2}*]. The line: model equals retrieval is plotted in blue; the red-striped line is a linear fit to the data.*

During the Australian summer, formaldehyde VCD's peak on the tropical northern coast and on the south-eastern coast between Sydney and Brisbane. IFS-COMPO is successful in reproducing the observed distribution. Although Sydney and Brisbane are clearly distinguishable in the formaldehyde-profile, it is still strongly underestimated compared to the observed VCD's by over ~0.5 x 10^{16} molecules cm^{-2}. The north-west of Australia, dominated by biogenic sources, is overestimated by ~0.3 x 10^{16} molecules cm^{-2}.

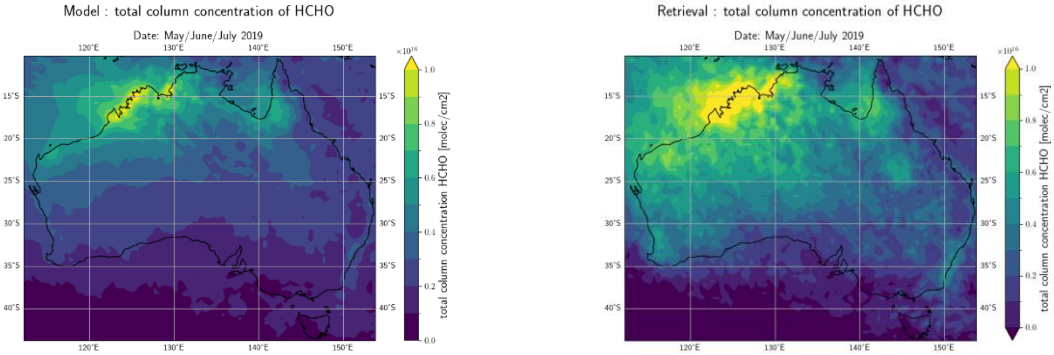


Figure 22. The total column formaldehyde concentrations of the model (left) and the TROPOMI retrieval (right) over Australia. The concentrations have been averaged over the months May/June/July 2019 (austral winter) for the time of satellite overpass.

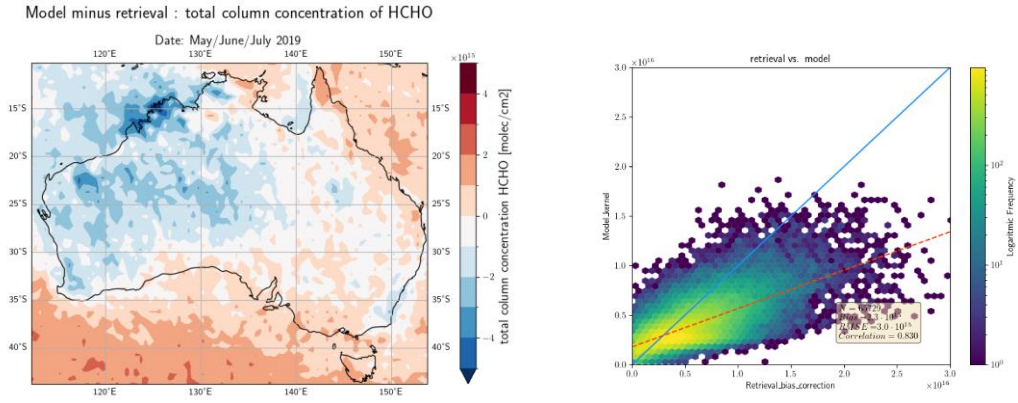


Figure 23. The model bias (model – retrieval) of total column HCHO concentrations over Australia. The concentrations have been averaged over the months May/June/July 2019 (austral winter) for the time of satellite overpass. The geographical distribution of the model bias is plotted in the figure at the left, a scatterplot of the model bias is plotted in the figure to

the right. The line model = retrieval is plotted in blue, the red-striped line is a linear fit to the data.
Formaldehyde VCD's during the Australian winter are significantly lower than during summer by ~0.5 x 10^{16} molecules cm^{-2}. The VCD measurements peak at the north-west coast with VCDs exceeding ~1.0 x 10^{16} molecules cm^{-2}. This is only partly predicted by IFS-COMPO and could be the result of fires. The fire season of this part of the Australian coast is between June and August [Australian Government Bureau of Meteorology] and therefore overlaps significantly with our researched period. IFS-COMPO has a smaller underestimation in the western Australian outback of ~-0.1 x 10^{16} molecules cm^{-2}.

Model performance comparison

Above findings from all three global regions are summarised in two normalised Taylor Diagrams[1] to directly compare the model performance over different regions and seasons (figs. 24 and 25)

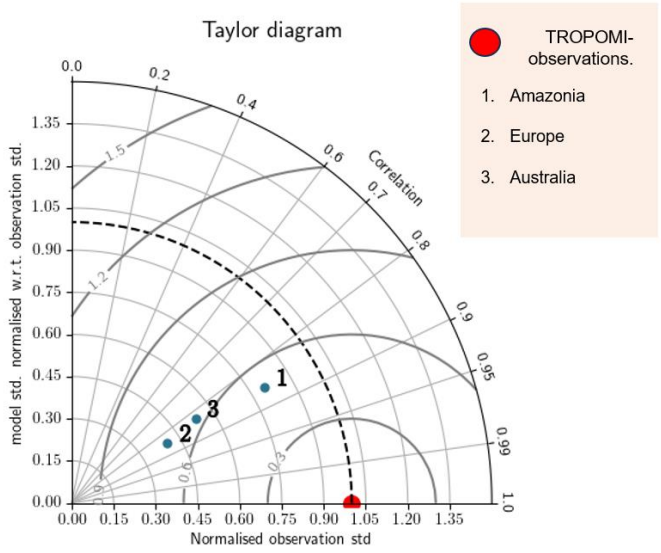


Figure 24. A normalised Taylor Diagram summarising the model performance with respect to TROPOMI observations for the months May/June/July of 2019.

[1] A normalized Taylor Diagram gives a neat graphical summary of the correlation between model values and observations, the Root mean square error (RMSE) and the model standard deviation normalized to the observation standard deviation. The red dot are the normalized TROPOMI-observations. The numbered dots are model values over different regions. The distance from the red dot to a numbered dot is the RMSE of the model values. The radial coordinates of a model dot correspond to the model standard deviation (radius) and *scaled* correlation (angle).

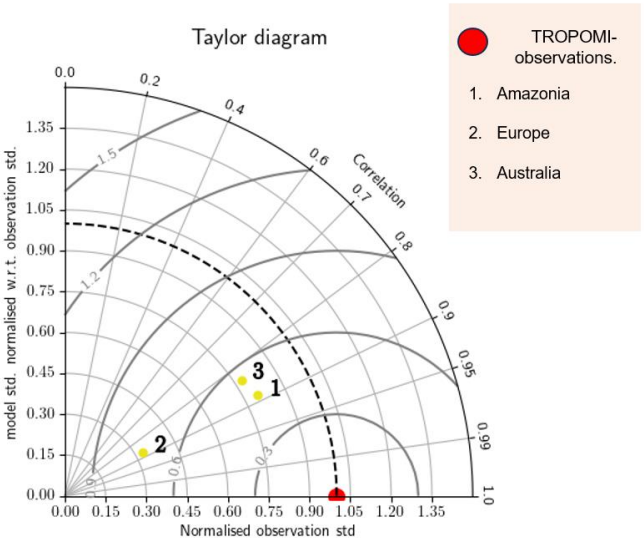


Figure 25. A normalised Taylor Diagram summarising the model performance with respect to TROPOMI observations for the months November/December/January of 2019.

IFS-COMPO mean VCDs have a lower standard deviation compared to observations for all regions and seasons. IFS-COMPO has trouble capturing the full range of observed formaldehyde VCDs, i.e. the simulations are too uniform. This is especially true for Europe with a normalised standard deviation of ~0.3 during both seasons.

Throughout both seasons, the Amazon rainforest and Europe maintain stable positions in the Taylor diagrams. Australia, however, improves its normalised standard deviation during November/December/January.

For all regions and seasons, the IFS-COMPO agrees well with the observations with a correlation between 0.8 and 0.9.

The normalised RMSE of IFS-COMPO is lowest over the Amazon Rainforest, ~0.5, for both seasons. IFS-COMPO over Europe has the highest normalised RMSE around ~0.7.

Conclusion

IFS-COMPO captures the formaldehyde VCDs observed by TROPOMI reasonably well in regions dominated by biogenic sources like the Amazon rainforest. Formaldehyde columns over the rainforest are overestimated in both seasons, especially in western Amazonia during the wet season, when VCDs are very low. IFS-COMPO underestimates the VCDs over Europe during summer but overestimates the low formaldehyde VCDs during winter. Australia's VCDs peak during the Australian summer, when IFS-COMPO underestimates the formaldehyde values along the more populated south-east coast and overestimates the outback and the north coast. VCDs during the Australian winter are roughly twice as low and are underestimated along the north-west coast. This could potentially be a consequence of the large uncertainties biomass burning inventories are affected off. To further assess remaining differences of observed and simulated HCHO, sensitivity studies (e.g. using alternative emission inventories) could be performed and more observational data could be employed (e.g. ground-based data sets like FTIR).
Together with the assessment against aircraft campaigns above evaluations suggest that the model quality to simulate HCHO columns may be sufficient in regions dominated by biogenic emissions to attempt the use of spaceborne HCHO observations for BVOC emissions inversion.

4 Develop and assess a simplified HCHO chemistry

4.1 A simplified chemistry mechanism applicable for biogenic hydrocarbons under tropical (low-NOx) conditions, using box-model calculations, tuned with respect to IFS(CB05) chemistry

By optimizing the HCHO precursor emissions in addition to the concentration itself in the data assimilation, we expect to improve simulated HCHO. The largest single source of HCHO over many continental areas are biogenic VOCs, especially over remote vegetated regions like the Amazon rainforest. Satellite observations of HCHO will be used to constrain the emissions of these BVOCs. This is because HCHO has a lifetime of the order of 4 hours, i.e. much shorter than the frequency of observations of low-Earth orbit (LEO) satellites with a global cover. Additionally, we assume that isoprene emissions are a good proxy for all biogenic VOC emissions. Indeed, they make up about half of the total biogenic VOC budget (Guenther et al. 2012). Other major biogenic compounds in terms of emissions include monoterpenes, methanol, acetone, acetaldehyde, and ethanol.

To adjust the isoprene emissions using assimilated formaldehyde measurements, an inversion mechanism needs to be developed that links the isoprene emissions to the formaldehyde columns at the time of observation. However, this requires a representation of the chemical degradation of isoprene, and subsequent transport of the intermediate species. A detailed representation of those processes is too complex and therefore too computationally expensive to use in a daily, global inversion mechanism with the IFS. It is therefore necessary to develop an ad-hoc simplified representation of the chemical processes that can be used in the inversion mechanism whilst not diverging too much from the full chemistry formaldehyde concentrations during the twelve-hour assimilation window.

The setup of this simplified chemistry for the degradation of isoprene consists of only two core reactions. Firstly, the oxidation of isoprene by OH, which results in the immediate formation of HCHO as well as of an intermediate oxidation product (IOX). The stoichiometric coefficient of HCHO in this reaction (α) is the primary production of HCHO. Secondly, the product IOX is oxidised by OH, at a slower rate, and this results in a delayed secondary production of HCHO, with stoichiometric coefficient ß, equal to the secondary HCHO formation yield from isoprene. The simplified chemistry scheme is completed with the addition of HCHO production from methane oxidation following the full chemistry, which contributes to the background HCHO column. Additionally, HCHO sinks are included by oxidation with OH and by photolysis. These last three reactions are the same as in the full chemistry scheme. The simplified chemistry scheme is shown in Fig. 26.

Simplified isoprene chemistry scheme

ISOP + OH → α HCHO + IOX	rate K_1 (Fixed to 10^{-10} s^{-1})
IOX + OH → β HCHO	rate K_2 (unknown)
CH_4 + OH → HCHO	from full chemistry
HCHO + OH → CO + 2 HO_2	from full chemistry
HCHO + *hv* → products	photolysis rate (calculated)

Figure 26. A simplified chemistry scheme for the biogenic oxidation chain to formaldehyde.

The simplified chemistry scheme displayed in Fig. 26 has four free parameters: the two stoichiometric coefficients for the primary and secondary production of HCHO (α and ß, respectively), and the two reaction rates (K_1 and K_2) for the oxidation of isoprene and for the

oxidation of IOX. The goal of this simplified scheme is to mimic the HCHO production from the full IFS chemistry. Consequently, we can optimise these four parameters by comparing box-model simulations of the full chemistry with the HCHO yield of the simplified scheme.

In this WP, we have set up box model runs by using the Kinetic PreProcessor (KPP) framework (version 3.1.1; Sandu et al. 2024). The chemistry in the box-model simulations is set up based on .kpp and .eqn files describing the IFS-COMPO chemistry. The initial concentrations are based on standard air composition with 1% H_2O content by volume, 150 ppb CO, and 20 ppb O_3. Additionally, we impose an initial concentration of 0.1 ppb of NO_2, thereby mimicking low-NOx conditions in the tropics. Because the NOx levels are impacted by the VOC chemistry, we keep the total NOx (= NO + NO_2) fixed to 0.1 ppb throughout the simulation. In real conditions, there would be various sources and sinks contributing to the NOx concentration, hence keeping the total NOx fixed will avoid unwanted feedback effects on the simulation results. Similarly, OH concentrations have a large impact on VOC chemistry. To limit the influence of OH on the box model simulations, we fix the OH concentration to a simple diurnal profile. Finally, an initial isoprene concentration of 1 ppb is assumed. Photodissociation rates are based on the TUV radiative transfer and photodissociation calculation package, in which the conditions are taken to be 30° N on 15th of July. The simulation starts at noon (12 pm local time) and is run for a total of 2.5 days with a time step of 6 minutes.

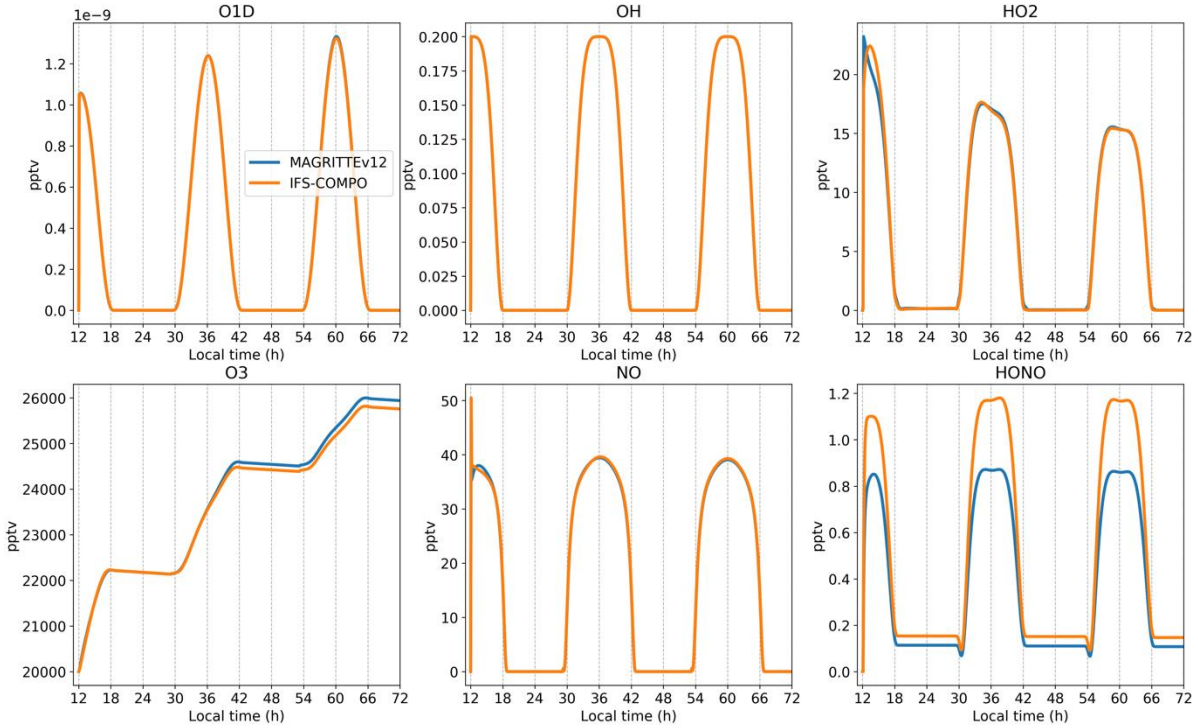


Figure 27: Evolution of concentrations of several inorganic compounds during box-model simulations using the chemistry of MAGRITTEv1.2 (blue) and IFS-COMPO (orange). From upper left to lower right, we display the concentrations of O(^{1}D) (first excited state of the oxygen atom), OH, HO_2, O_3, NO, and HONO (all in pptv).

Figure 27 shows the results of the box model simulations for inorganic compound concentrations. We compare the IFS-COMPO full chemistry and the MAGRITTEv1.2 chemical network (Muller et al. 2019). The latter has a dedicated mechanism for the chemical degradation of isoprene, hence serves as a reference against the IFS-COMPO chemistry. For the inorganic compounds, a good agreement is found between the two chemical networks. For both runs, the OH concentrations are fixed to a diurnal profile with a maximum of 0.2 ppt at noon local time and fixed to zero at night.

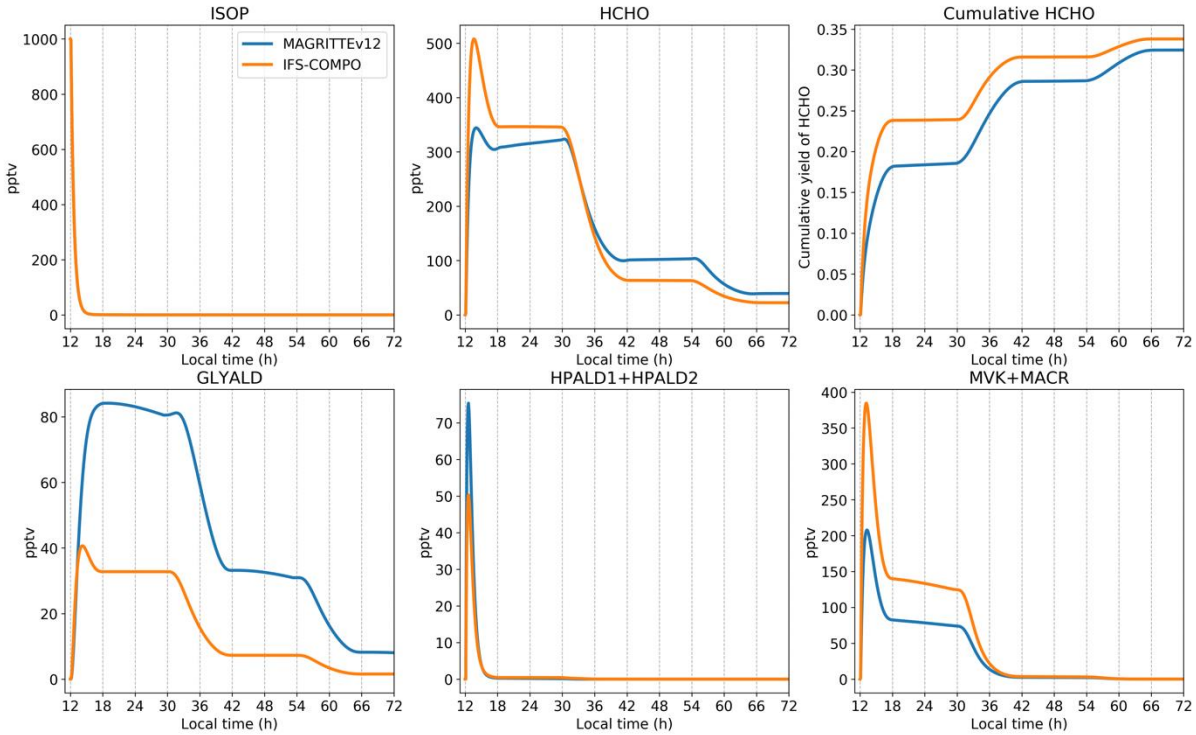


Figure 28: Evolution of concentrations for a list of organic compounds during box-model simulations using the chemistry of MAGRITTEv1.2 (blue) and IFS-COMPO (orange). From upper left to lower right, we display the concentrations of isoprene (C_5H_8), HCHO, cumulative HCHO production, glycolaldehyde (GLYALD), the sum of the two HPALDs, and the sum MVK+MACR.

The evolution of key organic compounds is shown in Fig. 28. As expected, isoprene is quickly oxidised, since its lifetime is typically lower than 1 hour during daytime in the simulation conditions. The HCHO production for the two chemistry models is relatively similar, although the IFS-COMPO mechanism leads to a higher primary production of HCHO. The total cumulative HCHO production after 60 hours is similar in the two models. In terms of intermediate species, we find higher concentrations of glycolaldehyde and HPALDs in the MAGRITTE mechanism. The sum of MVK and MACR (major isoprene oxidation products), which is included under the variable “ISPD” in the IFS-COMPO mechanism, shows higher concentrations in the IFS-COMPO simulation. This is attributed to the much faster 1,5 H-shift rates of isoprene hydroxyperoxy radicals in the IFS-COMPO mechanism, compared to MAGRITTE.

For the development of the simplified chemistry mechanism to be used in the HCHO data assimilation system, we substitute the organic chemistry of IFS-COMPO with the simplified scheme described in Fig. 26. Within the box model simulations, we exclude HCHO production from methane oxidation in the box model simulations aiming at deriving the parameters α, ß, K_1 and K_2. The first reaction in Fig. 22, ISOP+OH, is a fast reaction that is also present in the IFS-COMPO chemistry and produces the peroxy radicals ISOPBO2 and ISOPDO2. Its rate is well defined and is equal to about 10^{-10} s^{-1}. Consequently, for simplicity, we fix the value for K_1 = 10^{-10} s^{-1}. This leaves three parameters to be optimised.

The optimisation of the parameters is performed by minimising the squared difference between the cumulative HCHO distributions for the box models of the IFS-COMPO full chemistry and the simplified scheme. By taking the cumulative HCHO distribution as proxy,

the optimisation focuses on matching the HCHO production between the two chemical networks. We used the Nelder-Mead simplex algorithm (implemented using SciPy; Gao & Han, 2012) to find the minimum in the parameter space.

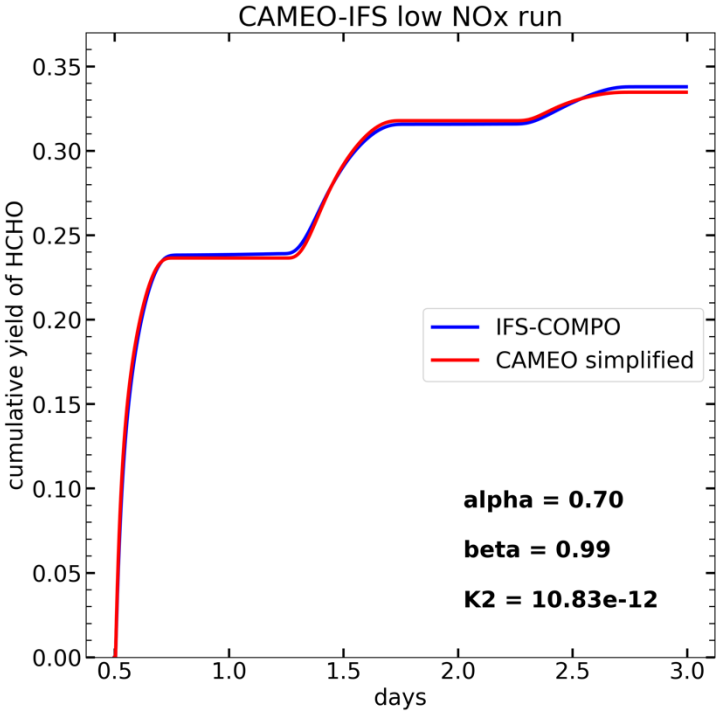


Figure 29: Cumulative yield of HCHO (per carbon atom) from isoprene oxidation from box-model simulations using the full chemistry (blue) and the simplified scheme (red). The optimised parameters are displayed in the bottom right corner. This optimisation was performed for low-NOx (100 ppt) conditions.

The result of the optimisation for low-NOx (100 ppt) conditions is shown on Fig. 29. The simplified chemistry succeeds very well in reproducing the temporal evolution of HCHO formation, despite its simplicity. In this framework, the primary HCHO production is mainly governed by the first reaction in Fig. 26, which is mainly determined by the coefficient α. Indeed, this reaction occurs very fast and all the isoprene, as can be seen in Fig. 27, is rapidly oxidised. The secondary production of HCHO from IOX is determined by ß and K_2, which contribute significantly to the total HCHO yield spread out over the full run.

However, in realistic conditions other VOCs contribute significantly to the HCHO abundances. In particular, and most importantly, monoterpenes ($C_{10}H_{16}$) are also emitted by terrestrial vegetation. Their contribution to HCHO formation is expected to be about 5 times smaller compared to isoprene to the HCHO concentrations. Indeed, following CAMS-GLOB-BIOv3.1 (Sindelarova et al. 2022), monoterpene emissions are about 5 times lower than isoprene emissions over the Amazon (in terms of mass). Neglecting these emissions would therefore lead to an underestimation of HCHO formation from biogenic VOC emissions. We tentatively account for this by including an initial monoterpene concentration in the box model simulations, equal to 10% of the initial isoprene concentration (i.e. 20% by mass). The parameter estimation is repeated while assuming that the total HCHO yield is due to isoprene only. The results of this optimisation are shown in Fig. 30.

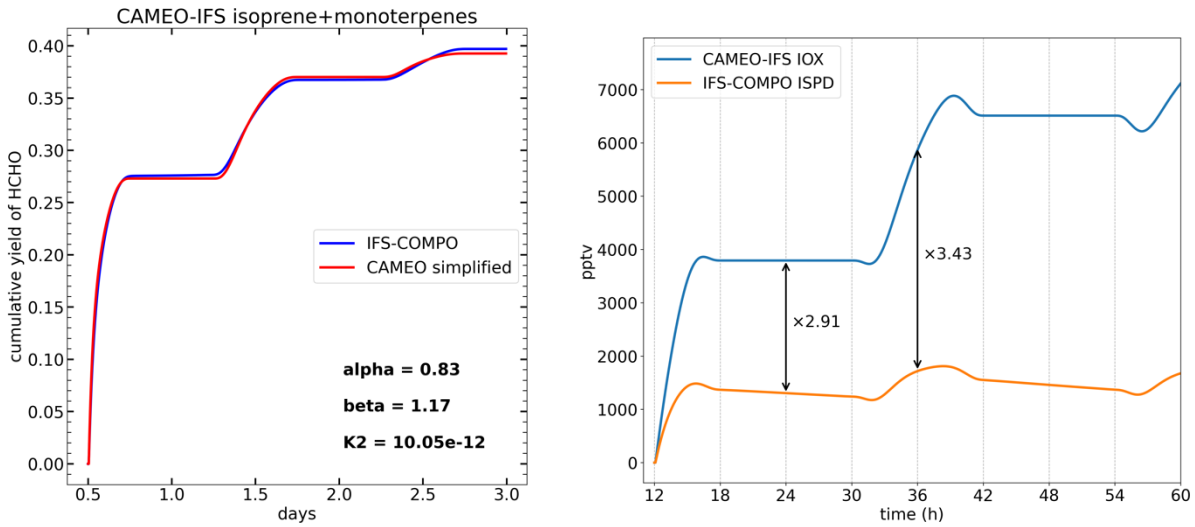


Figure 30: Left: Same as Fig. 29, but by including an initial monoterpene concentration of 100 ppt. The total HCHO yield is slightly increased, which is reflected in the higher values for the coefficients α and ß. Right: Evolution of the isoprene to IOX ratio in conditions with ongoing isoprene emissions.

The inclusion of monoterpenes in the box-model simulations increases the cumulative yield after 3 days to about 0.4, compared to 0.35 for isoprene only. This increase in the HCHO production is reflected by an increase in the coefficients of α and ß. With the co-occurrence of several BVOC sources in tropical regions, we expect the parameters in Fig. 30 to produce better results for the HCHO production of the full chemistry simulations.

Parameter	**Value determined in box model**
α	0.83
β	1.17
K_1 (Isop. oxidation rate)	1e-10 (fixed)
K_2 (IOX oxidation rate)	10.83e-12

Table 1. Parameters in the simplified chemistry scheme as determined by box model simulations for the oxidation of isoprene and monoterpenes in Fig. 30. These values are used for the data assimilation.

4.2 Implementation of the simplified chemistry in IFS and comparison of its performance against the full-chemistry configuration

Implementation in the IFS

The simplified chemistry scheme discussed in section 4.1 is implemented in IFS-COMPO as new subroutine *TM5_SIMPLE_HCHO_CHEM.* This subroutine is available from the ifs-source branch "*nkmw_CY48R1.M_simplified_chem_methane_source"* in the routine *"tm5_simple_hcho_chem.F90".* The subroutine is called from the subroutine *"chem_tm5.F90"* (line 845), directly after solving the tropospheric chemistry, using the KPP-solver code. The full "*tm5_simple_hcho_chem.F90"* source code as used for this report is included in Appendix A.

The simplified chemistry subroutine takes as input arrays the initial concentration of the chemical tracers from both the full and the simplified chemistry (molec cm^{-3}). Additionally, the reaction and photolysis rates of the full chemistry are taken as input. The subroutine outputs an array with the updated concentrations of the chemical tracers.

The script is called for each model time step and contains a single spatial loop that runs over all the model grid points. For each model grid point, the local reaction and photolysis rates are selected. Then, using the simplified chemical reactions, the concentrations of the involved tracers are updated, i.e. IOX*, ISOP* (isoprene) and HCHO*2 (formaldehyde).

In the test performed for this report, the ISOP* and HCHO* concentrations are re-set every day using the isoprene and formaldehyde outputs of the full chemistry. Since IOX* does not correspond to a single, full-chemistry tracer, but instead to an abstract collection of all the oxidation products of isoprene, a choice has to be made how to update its concentration. Within the framework of developing an optimal simplified chemistry scheme, different trace gases were analysed for their use as proxy tracers for all the combined intermediate isoprene oxidation products (IOX). Test experiments were performed using ethene, acetylene, ISPD, and glyoxal. ISPD was found to be preferable due to its close correlation with isoprene oxidation and its larger concentrations as e.g. found for glyoxal. Still, ISPD underestimates the total sum of all isoprene oxidation products. Box model evaluations of the evolution of the ISPD/IOX ratio in conditions with ongoing isoprene emissions found ratios in the range 2.9 - 3.4 after 12 to 24 hours (fig. 30, right panel). For these experiments, the isoprene concentration was fixed to a diurnal profile rather than an initial concentration only. Following these findings, the presented simple chemistry scheme uses ISPD multiplied by a factor of three to optimally represent IOX (ISPD*3).

Evaluation of the Simplified Chemistry

The goal of the simplified chemistry is to match formaldehyde from the full chemistry as close as possible within the CAMS assimilation window of 12 hours. In this section, we discuss the evaluation results where we evaluate the difference between using the simplified chemistry and the full chemistry.

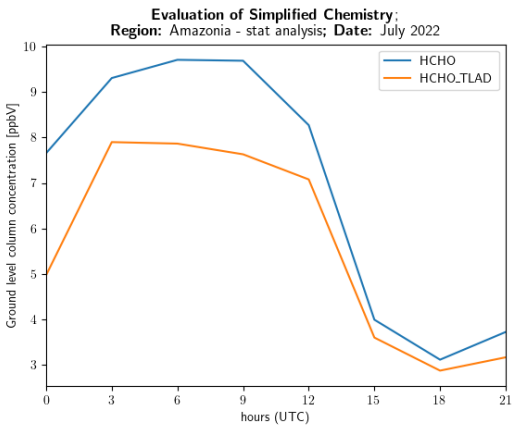
2 To distinguish similar chemical species concentrations in the simplified chemistry scheme from their concentrations calculated using the full chemistry, an asterisk is added in the text. In the code the subscript '_TLAD' is used in anticipation of the variables use in the Trans-Linear ADjoint code.

The first focus of the research project is Amazonia, where biogenic sources of formaldehyde dominate, creating an ideal test region for the project goal of using formaldehyde observations to constrain biogenic VOC emissions (fig. 31).

Figure 31. A map of the target area - a large part of the remote Amazon Rainforest with easily defined boundaries for data analysis. The longitude limits are [-74.2,-52.0], the latitude limits are [-9.1,2.6].

Diurnal cycle

In Figure 32 the monthly mean formaldehyde diurnal cycle based on full and simplified chemistry is plotted for July 2022. Midday in the project region is around 16 UTC. Clearly, simple and full chemistry formaldehyde have similar diurnal cycles, with a strongly defined peak in early morning hours and pronounced loss following local midday. The simplified chemistry seems to capture the diurnal variation well both at surface level fig. 32, left) and in the vertical column density (fig. 32, right). The same was found for December (fig. 33). Generally, the pronounced early morning peak appears slightly counter intuitive, as the production of HCHO through oxidation of isoprene is mainly taking place during daytime. Also, isoprene emissions follow a diurnal cycle with their maximum during daytime. However, the lifetime of HCHO is also shortest during daytime, implying quick loss of HCHO after its production. Nevertheless, there is indication that the diurnal cycle is too strong.



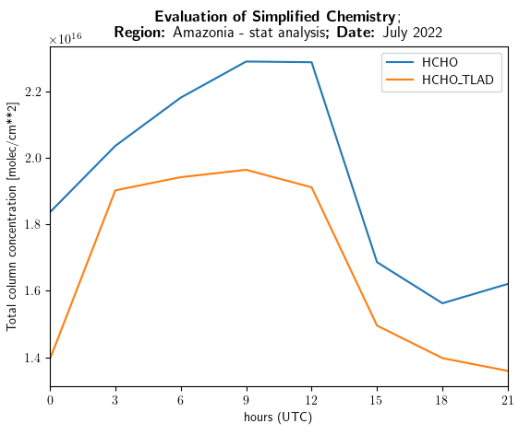


Figure 32. The diurnal cycle of formaldehyde over the Amazon from the global IFS-COMPO model averaged over July 2022. The surface level mixing ratio is shown at the left, the total column is shown at the right. The blue line is the formaldehyde concentration of the full chemistry. The orange line is the formaldehyde concentration from the simplified chemistry.

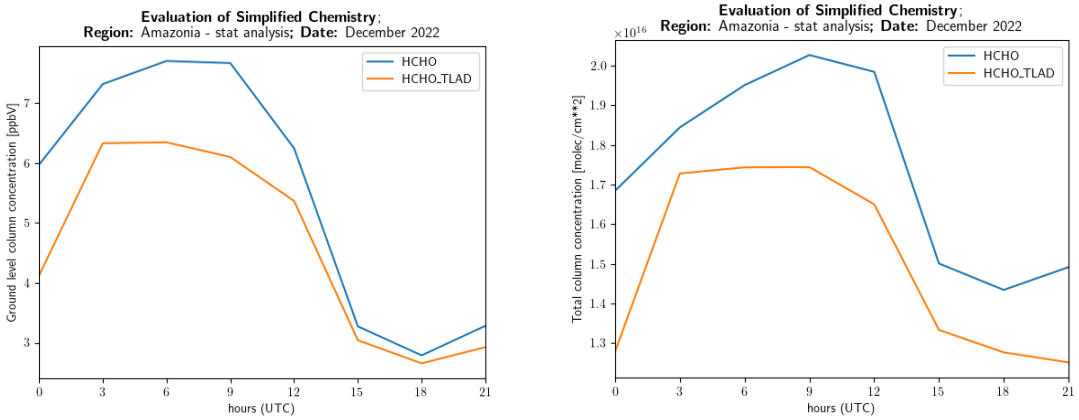


Figure 33. The diurnal cycle of formaldehyde over the Amazon from the global CAMS-model averaged over December 2022. The ground level concentration is shown at the left, the total column is shown at the right. The blue line is the formaldehyde concentration of the full chemistry. The orange line is the formaldehyde concentration as a result of the simplified chemistry.

Depicted in Figure 34 are in addition to the two formaldehyde mixing ratios, the isoprene mixing ratio of the full and the simplified chemistry, as well as the proxy tracer IOX of the delayed oxidation channel from isoprene to formaldehyde. Plotted are the average mixing ratios over Amazonia for the month July 2022. The isoprene mixing ratio of the simplified chemistry stays very close to that of the full chemistry. This is to be expected as the isoprene-concentration of the simplified chemistry is set to the value of the full chemistry every day in the model. Secondly, both the isoprene and IOX mixing ratios peak around ~6 hours earlier than formaldehyde. This is the time it takes for the isoprene to oxidise into formaldehyde.

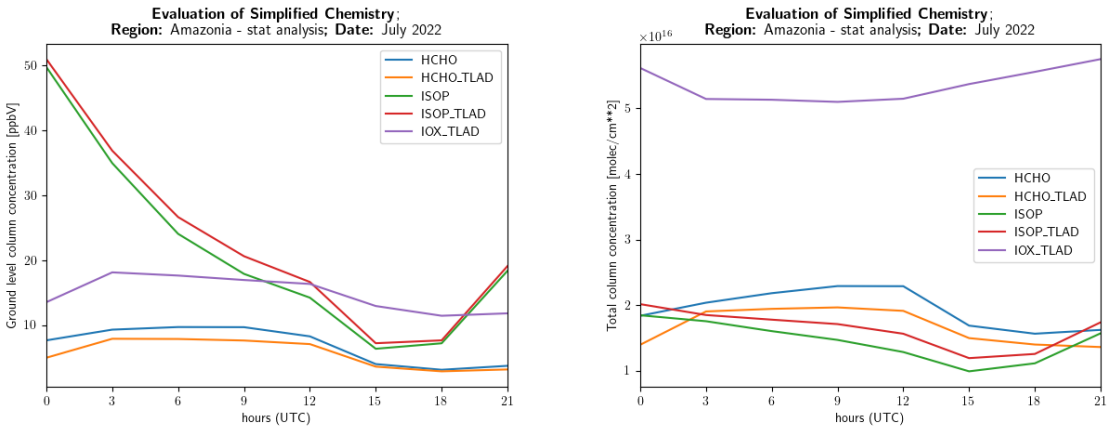


Figure 34. The diurnal cycle of formaldehyde resulting from the full chemistry (blue), formaldehyde resulting from the simplified chemistry (orange), isoprene in the full chemistry (green), isoprene from the simplified chemistry (red) and the proxy tracer of the delayed

oxidation channel in the simplified chemistry scheme IOX (purple). All concentrations are averaged over the project region (Amazonia) and the month of July 2022. The left plot shows the mixing ratios at the surface level. The right plot shows the VCD. For the calculation of the total columns, IOX is assumed to have the molecular mass of ISPD.

Depicted in Figure 35 are again the mixing ratios of isoprene, IOX and formaldehyde averaged over a month in the wet season (December 2022). The graphs are almost identical of shape compared to the concentration during the dry season but have a lower mixing ratio.

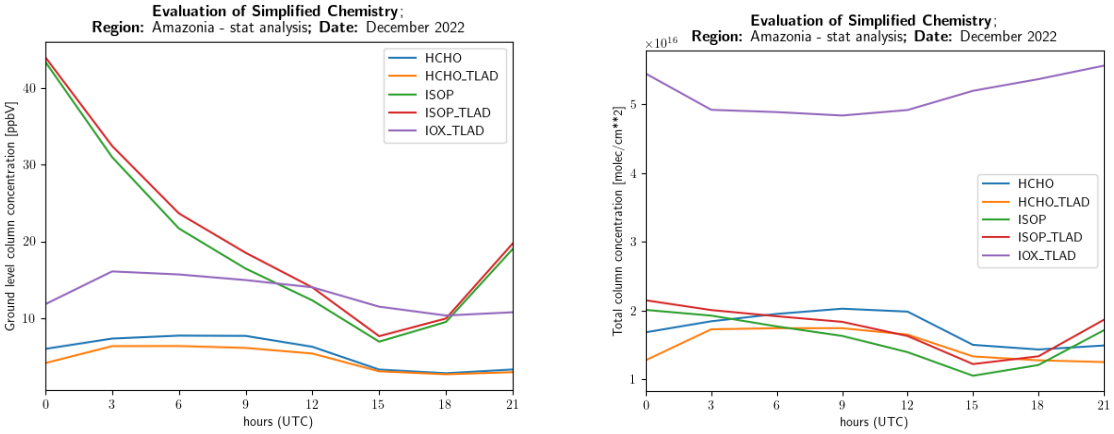


Figure 35. The diurnal cycle of formaldehyde resulting from the full chemistry (blue), formaldehyde resulting from the simplified chemistry (orange), isoprene in the full chemistry (green), isoprene from the simplified chemistry (red) and the proxy tracer of the delayed oxidation channel in the simplified chemistry scheme IOX (purple). All concentrations are averaged over the project region (Amazonia) and the month of December 2022 – the wet season. The left plot shows the concentrations at the ground level. The right plot shows the total column concentrations. The left plot shows the mixing ratios at the surface level. For the calculation of the total column concentration, IOX is assumed to have the molecular mass of Glyoxal.

Geographic distribution

Finally, we will assess if the error due to using a simplified chemistry has any significant geographic distribution. See Figure 36 for the difference between formaldehyde using the full chemistry and the simplified chemistry at the time closest to the TROPOMI overpass around 17.30 UTC. We therefore use the model output of 18.00 UTC.

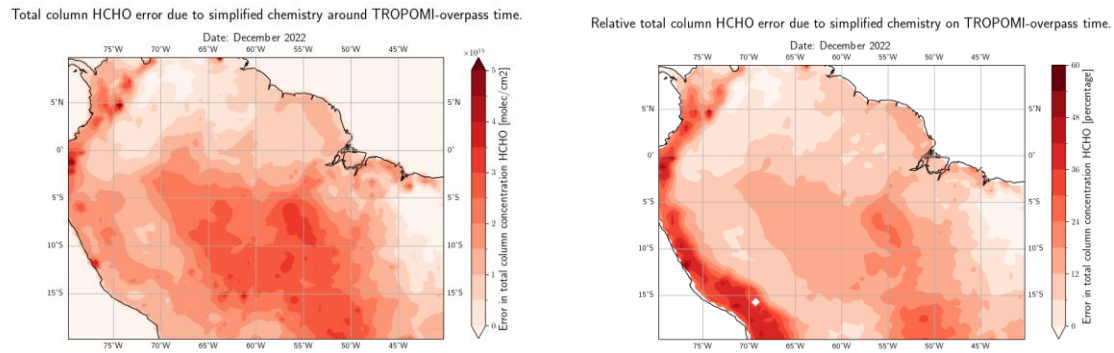


Figure 36. The geographical distribution of the error in total column formaldehyde at 1800 UTC for December 2022 resulting from using the simplified chemistry scheme. Depicted is HCHO due to full chemistry minus HCHO due to simplified chemistry. (Left) The absolute error in molec/cm2. (Right) The relative error with respect to the full chemistry HCHO.

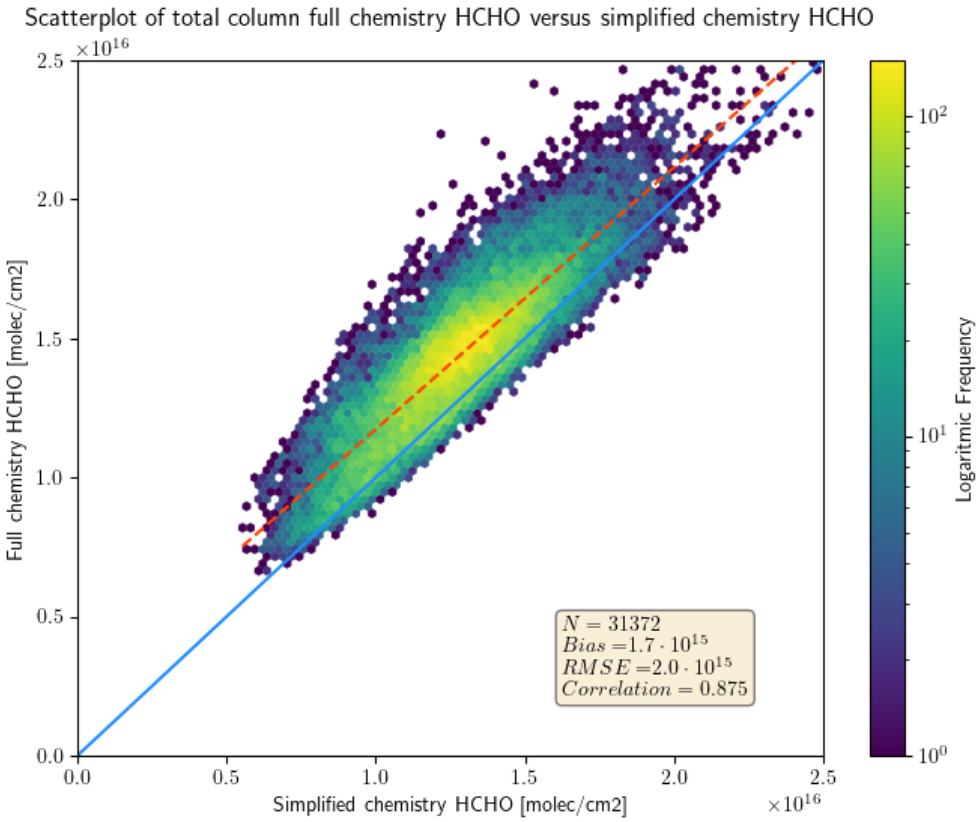


Figure 37. A scatterplot of the HCHO VCDs over Amazonia at 18.00 UTC for December 2022. Clearly visible is the small positive offset of full chemistry HCHO (red dotted line).

The deviation of simple and full chemistry HCHO in the project area ranges from 0.5 to 2.5 molec cm^{-2} (fig. 36). The absolute discrepancy peaks around the heavily populated city Bogotá in the north. On average, the discrepancy gets larger to the south of the project area where the rainforest slowly transforms into agricultural land. The relative discrepancy is highest over the Andes mountains where HCHO is generally low due to decreased surface sources.

Conclusion

A direct comparison of full and simple chemistry HCHO VCDs for December 2022 confirms above findings. The two chemistry schemes show a good correlation of 0.875 and an average negative offset of the simplified scheme of 1.7×10^{15} molec cm^{-2} relative to the full chemistry configuration (fig. 37). The comparison has proven that the simple scheme captures well both the geographic as well as the diurnal patterns of full chemistry HCHO. Absolute formaldehyde (both concentrations and VCDs), however, are slightly underestimated. This is most probably a consequence of missing photochemical sources either within the isoprene oxidation chain or independent of isoprene. Also, we note that the simple scheme does not (and is not meant to) capture any anthropogenic formaldehyde enhancement at this stage of the project.

5 Assimilation of HCHO data in 4DVar context and derivation of TL/AD of a simplified HCHO chemistry

IFS-COMPO uses incremental four-dimensional variational (4D-Var) data assimilation with a 12-hour assimilation window. Within the CAMEO project, an assimilation system for HCHO data is being developed that makes use of tangent linear and adjoint of a simplified HCHO chemistry scheme and inversion of non-methane VOCs in the IFS. In a first step, the impact of HCHO assimilation on other trace gases is analysed by only assimilating TROPOMI HCHO observations without applying the tangent linear and adjoint of HCHO or NMVOC inversion. Secondly, the impact of the tangent linear and adjoint code on the HCHO assimilation will be tested. As a final step, the impact of the assimilation of HCHO will be analysed by making use of the tangent linear and adjoint and the newly developed IFS NMVOC inversion. If successful, this final assimilation system of HCHO is meant to be implemented in the operational CAMS system.

This deliverable reports on the impact of the assimilation of HCHO TROPOMI observations without use of tangent linear and adjoint or NMVOC inversion. The further steps will be tested in the remainder of the project.

5.1 Assessment of the impact of HCHO assimilation in IFS-COMPO

Due to the short atmospheric lifetime of HCHO against photolysis and oxidation of ~2-3 hours only, the assimilation of HCHO (without tangent linear and NMVOC inversion) is assumed to have no major impact on the atmospheric chemistry. However, some impact on correlated trace gases is expected. This is tested below for the major atmospheric precursor of HCHO, isoprene, as well as for ozone due to its close chemical correlation with atmospheric VOCs. The following analysis validates monthly averaged VCDs of atmospheric HCHO, isoprene, and ozone a) when assimilating TROPOMI HCHO observations and b) without HCHO assimilation, but instead employing the standard IFS-COMPO chemistry. For this purpose, four test experiments have been set up based on CAMS cycle 48R1 (table 2). The results are discussed below. Currently, the assimilation of HCHO in CAMS makes use of TROPOMI HCHO without application of the bias correction that is described in section 3.2.

Experiment ID	branch	Time period	comment
icir	std_CY48R1.M_HCHOneg	July 2023	Reference run without HCHO assimilation
ieg6	std_CY48R1.M_HCHOneg	December 2023	Reference run without HCHO assimilation
ichf	std_CY48R1.M_HCHOneg	July 2023	HCHO assimilation
iegt	std_CY48R1.M_HCHOneg	December 2023	HCHO assimilation

Table 2: Test experiments with assimilation of TROPOMI HCHO observations and reference experiments without HCHO assimilation using the standard IFS-COMPO chemistry of CAMS cycle 48R1.

The assimilation of TROPOMI HCHO observations causes a small increase in global simulated HCHO in the order of 0.1±1.1% (December) and 1.3±3.4% (July), respectively (fig. 38). While no clear geographic pattern can be observed for December, the assimilation of

HCHO in July decreases northern hemisphere HCHO by on average 0.2±2.1%, while simultaneously increasing southern hemispheric formaldehyde by 3.7±3.5%. However, this impact appears most pronounced towards the poles and over marine regions, where no significant surface sources of HCHO exist and only a small atmospheric background of HCHO from methane oxidation is expected. In terrestrial regions and over formaldehyde source areas (e.g. tropical rainforests), the impact of activating HCHO assimilation is much smaller, also in the southern hemisphere. As an example, we observe an average HCHO increase of 0.5±2.0% and 1.4±1.7% for December and July, respectively over the Amazon rainforest. Notable is the significant reduction in simulated HCHO over western Canada in July 2023 (fig. 32, panels d to f, black arrows). Here, extended wildfires occurred throughout the entire month (fig. 39), causing significant enhancements in IFS-COMPO HCHO up to 21x10^{15} molec cm^{-2} (no assimilation) or 13x10^{15} molec cm^{-2} (HCHO assimilation). Assimilating TROPOMI HCHO reduces HCHO by 15%, both in vicinity of the fires as well as when the plumes are transported downwind.

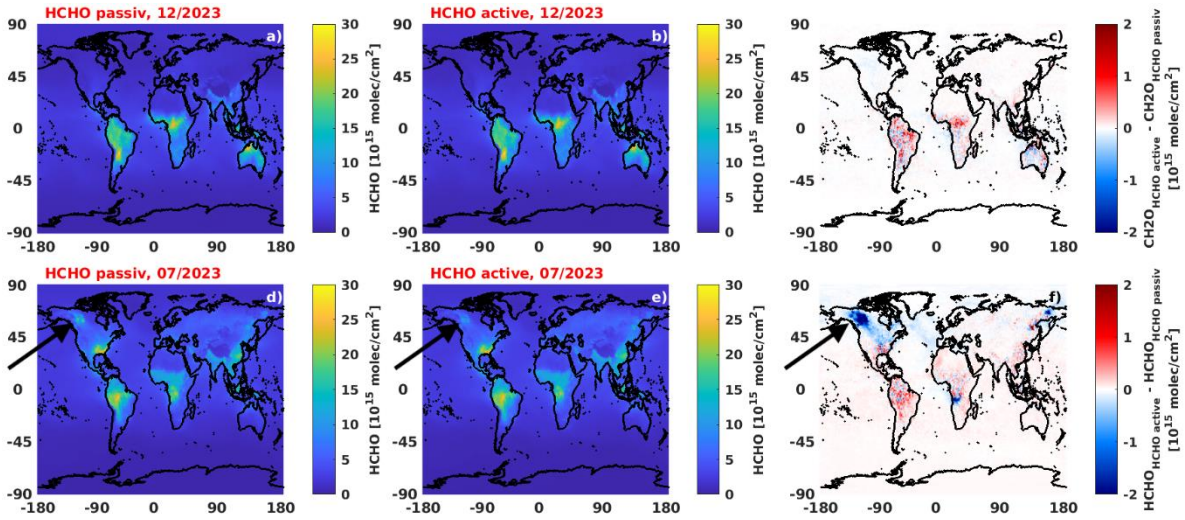


Fig. 38: Monthly averaged HCHO VCDs for December 2023 (panels a to c) and July 2023 (panels d to f) based on CAMS 48R1 without (panels a, d) and with HCHO assimilation (panels b, e as well as the difference of both (panels c, f). In the difference plots, red colour indicates regions where the assimilation of HCHO increased simulated HCHO and blue colour indicates regions where the assimilation decreased simulated HCHO. Black arrows show areas of extended wildfires in northwest Canada in July 2023.

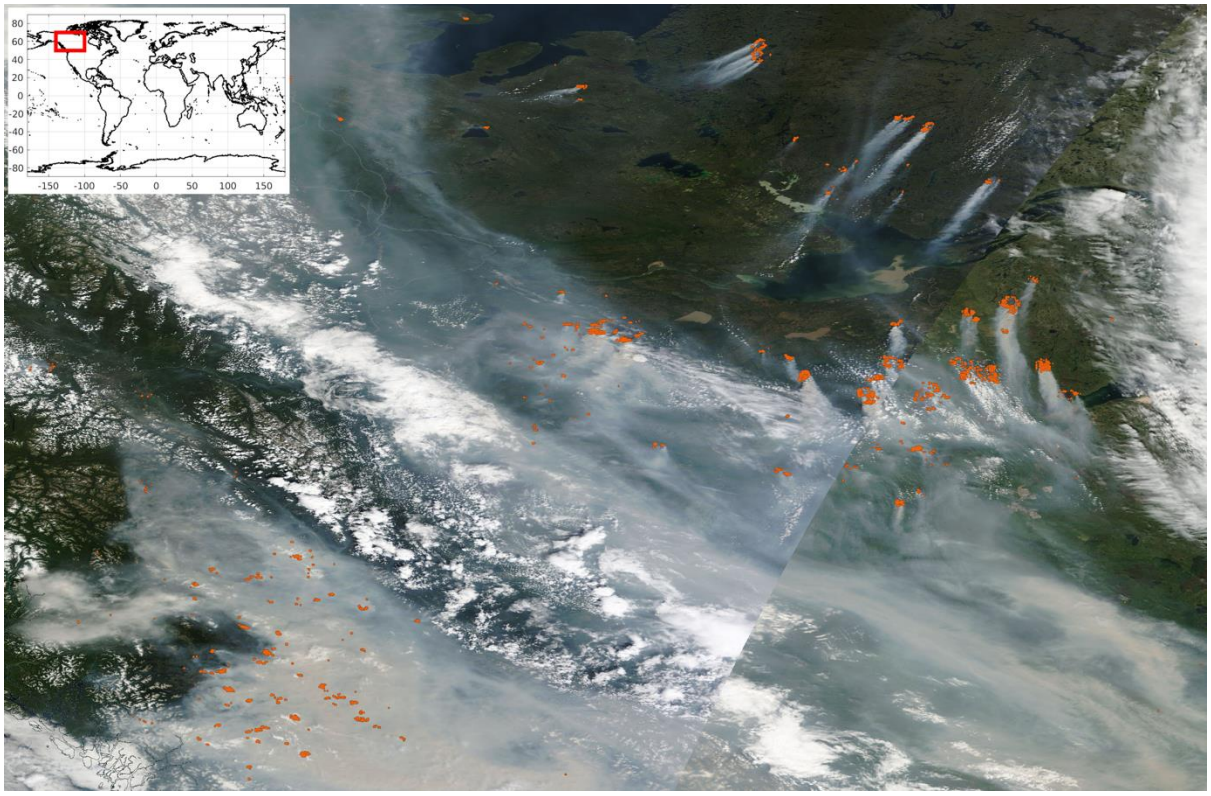


Fig. 39: Biomass burning events (red dots) and smoke plumes as detected by the MCD14 MODIS instrument on the Terra and Aqua satellites over northwest Canada between 50 and 70°North and 110 and 130°West for 13 July 2023. MODIS satellite image taken from NASA WORLDVIEW; see https://worldview.earthdata.nasa.gov/.

The assimilation of HCHO in absence of tangent linear and adjoint formulation, may still indirectly impact atmospheric isoprene, the main HCHO precursor, through changes in ozone and OH that impact the isoprene lifetime (fig. 40). However, neither a clear seasonal pattern (fig. 40, top versus bottom row) nor a significant geographic dependence can be observed -in particular not in the main bVOC source regions, i.e. the tropics. There, simulated isoprene decreases on average by 3.8±7.3% (December) and 1.5±8.0% (July) when assimilating HCHO. This pattern is most pronounced over the Amazon rainforest, where CAMS cycle 48R1 is potentially overestimating biogenic VOCs (see section 3.4).

Changes in total atmospheric ozone due to the assimilation of TROPOMI HCHO are only minor and stay well below 1% (fig. 41). For the tropospheric fraction of ozone, the impact of assimilating HCHO is expectedly slightly larger with median changes ranging from -1.1±1.3% to 0.5±0.5% depending on season and latitude (fig. 42).

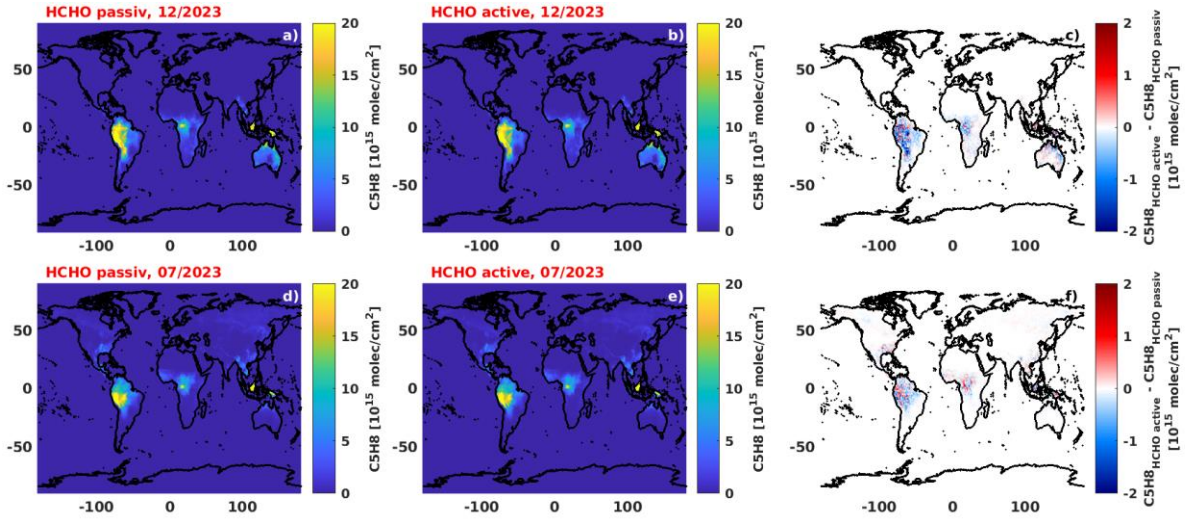


Figure 40: Monthly averaged isoprene VCDs for December 2023 (top) and July 2023 (bottom) based on CAMS 48R1 without (left) and with HCHO assimilation (middle) as well as the difference of both (right).

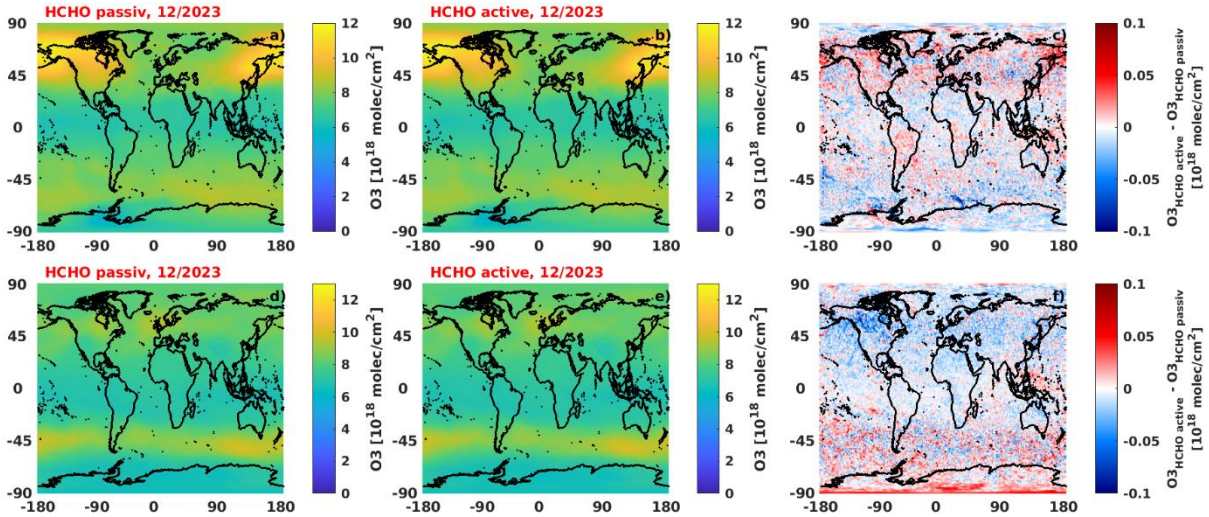


Figure 41: Monthly averaged ozone VCDs (troposphere and stratosphere) in Dobson units for December 2023 (top) and July 2023 (bottom) based on CAMS 48R1 without (left) and with HCHO assimilation (middle) as well as the absolute difference of both (right).

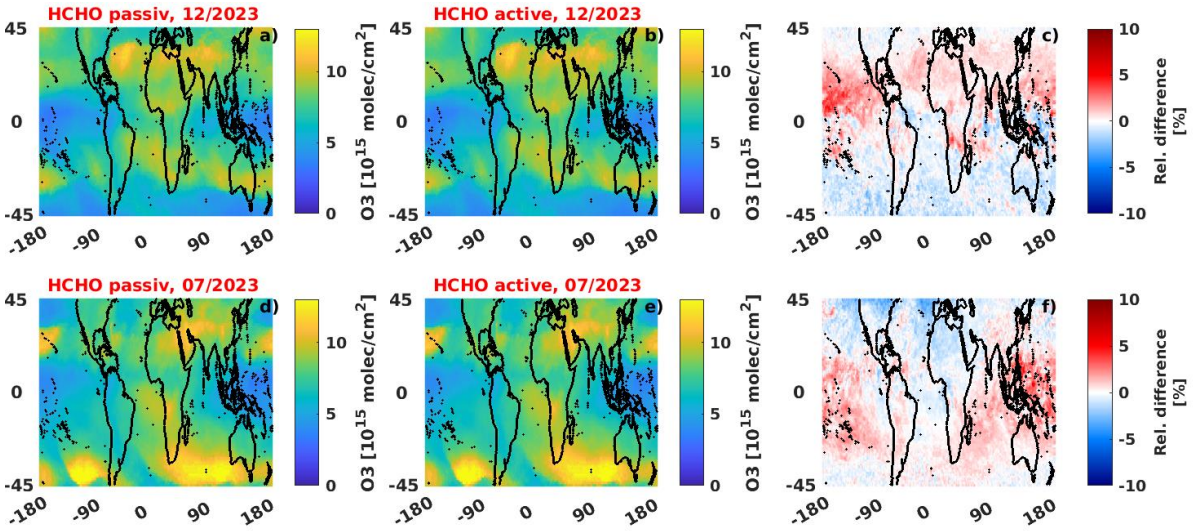


Figure 42: Monthly averaged ozone VCDs (troposphere only) in molec cm^{-2} for December 2023 (top) and July 2023 (bottom) between –45° and 45°North based on CAMS 48R1 without (left) and with HCHO assimilation (middle) as well as the relative difference of both (right). Difference in % expressed relative to the standard IFS-COMPO configuration without HCHO assimilation with red indicating larger O3 when assimilating HCHO and blue indicating smaller O3 due to the assimilation.

5.2 Comparison of assimilated HCHO against direct satellite retrievals and interpretation of the differences

The following section analyses the impact of HCHO assimilation using collocated satellite retrievals of HCHO from the TROPOMI instrument. The comparison is performed based on monthly averaged HCHO vertical column densities for the months July and December 2023 with particular focus on a series of larger wildfires in summer 2023. Since the assimilation of TROPOMI HCHO in the CAMS system does currently not make use of the bias correction according to section 3.2, it is also not applied to TROPOMI HCHO in the presented comparison.

For December 2023, satellite HCHO retrievals are mostly limited to latitudes south of 50°North (fig. 43, panels a to c). As laid out in detail in chapters 3 and 4, simulated HCHO appears to generally underestimate HCHO over marine surfaces and outside of prominent source regions (fig. 43, panels c, f). However, over tropical rainforests and in particular north of the Amazon river basin, simulated formaldehyde significantly exceeds the satellite observations. These main findings are unaffected by the assimilation of HCHO in IFS-COMPO (compare fig. 43, panels c, f). For December, the assimilation mostly affects HCHO over southern hemispheric terrestrial regions without clear tendency towards larger or smaller model to satellite biases (fig. 44, panel a).

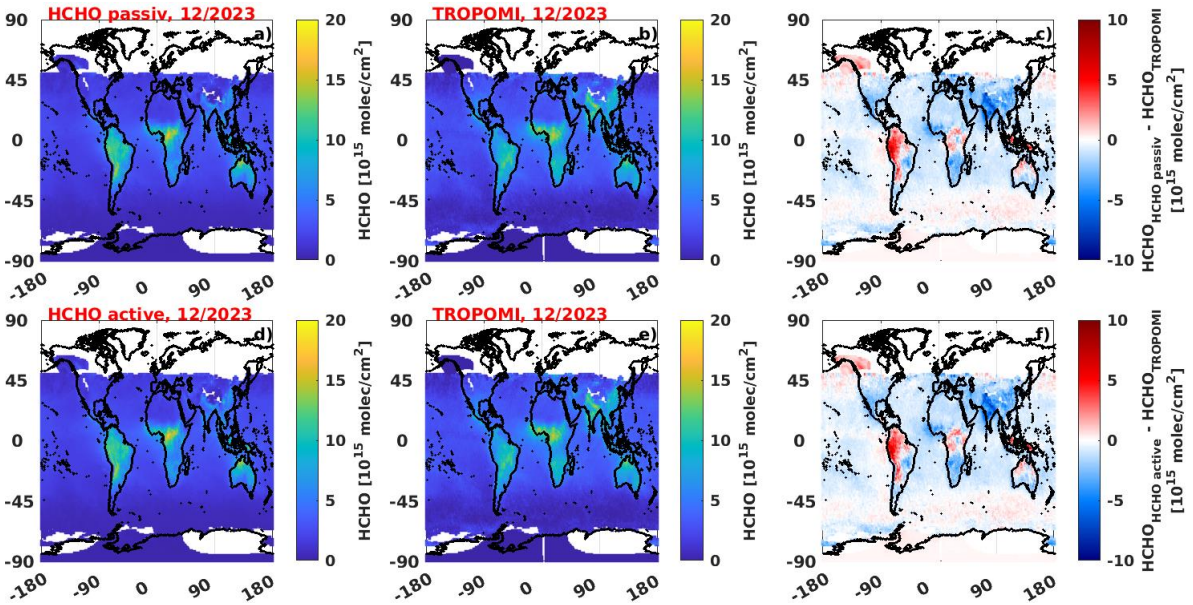


Figure 43: Comparison of monthly averaged formaldehyde VCDs from IFS-COMPO (CAMS 48R1) and the TROPOMI instrument without assimilation of HCHO (upper row) and when assimilating HCHO (lower row) for December 2023. For the difference plots (panels c, f), red colour indicates larger simulated HCHO than retrieved from the satellite and blue indicates larger satellite retrieved HCHO as compared to the simulations. The absolute difference of panels c) and f) is plotted in fig. 39, panel a).

For July 2023, the comparison of satellite retrieved HCHO and IFS-COMPO simulations with HCHO assimilation reflects the main findings from December. With better geographic coverage of the northern latitudes as compared to December, a slight overestimation of simulated HCHO over continental Europe and surrounding marine regions becomes evident (fig. 46). As discussed for December 2023, on a global scale the assimilation of HCHO in IFS-COMPO neither causes a significantly larger or smaller bias to the satellite observations (fig. 44, panel b). Minor improvements by 7% are observable over marine surfaces in moderate latitudes (-20° to 20°North) and over terrestrial regions without major surface BVOC emissions like North Africa or Australia (southern hemispheric winter). Towards the poles, where the sensitivity of the satellite observations decreases, the model-satellite difference shows a small increase due to the assimilation. Over BVOC source regions, e.g. northern South America, the agreement of simulated and observed formaldehyde slightly decreases by 1.4% due to the assimilation (10.5°South and 11.5°North and 58° to 78°West; fig. 45). Southeast of this region, the bias of retrieved and modelled HCHO is lowered by 2% when assimilating HCHO (28° and 4.5°South and 41° to 58°West; fig. 45). Notable is also a slightly increased positive model bias due to the assimilation of HCHO over east Canada. There, TROPOMI observes enhanced HCHO VCDs mostly as a consequence of extended wildfire occurring during July 2023 (fig. 47). More analysis is needed to better understand the discrepancies in and outside of source regions as well as during emission events, in particular such cases where the assimilation increases IFS-COMPO and TROPOMI differences. These topics will be investigated further within the context of including tangent linear and adjoint into the assimilation of HCHO in IFS-COMPO.

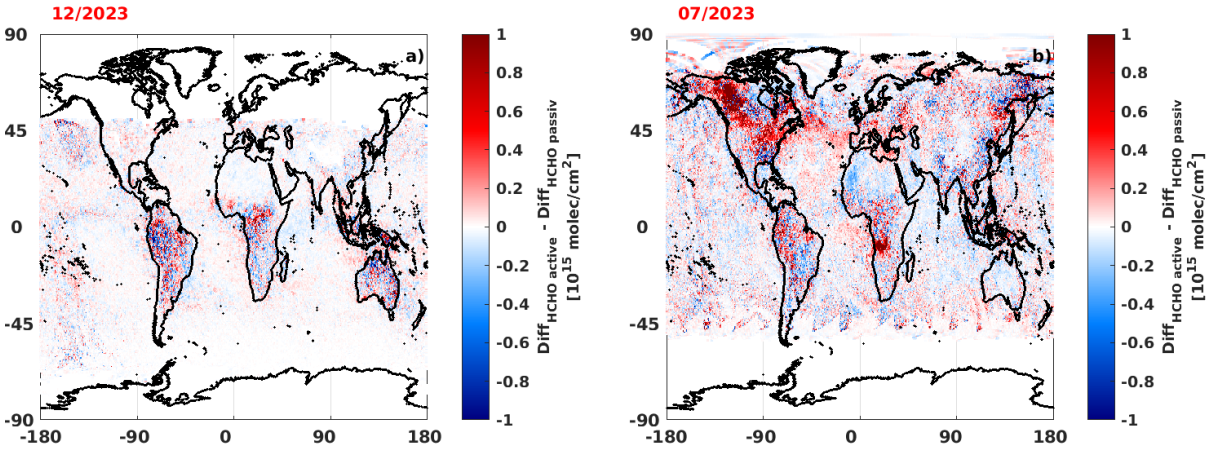


Figure 44: Absolute difference of simulated and retrieved HCHO with and without HCHO assimilation for December 2023 (panel a) and July 2023 (panel b). Data calculated according to abs($HCHO_{ifs,active}$ – $HCHO_{tropomi}$) - abs($HCHO_{ifs,passiv}$ – $HCHO_{tropomi}$) with $HCHO_{ifs,passiv}$ specifying IFS-COMPO HCHO simulations without HCHO assimilation and $HCHO_{ifs,active}$ with active HCHO assimilation. As such, red colour indicates an increased model to satellite divergence when assimilating formaldehyde, blue colour indicates a smaller model to satellite difference due to the assimilation of HCHO.

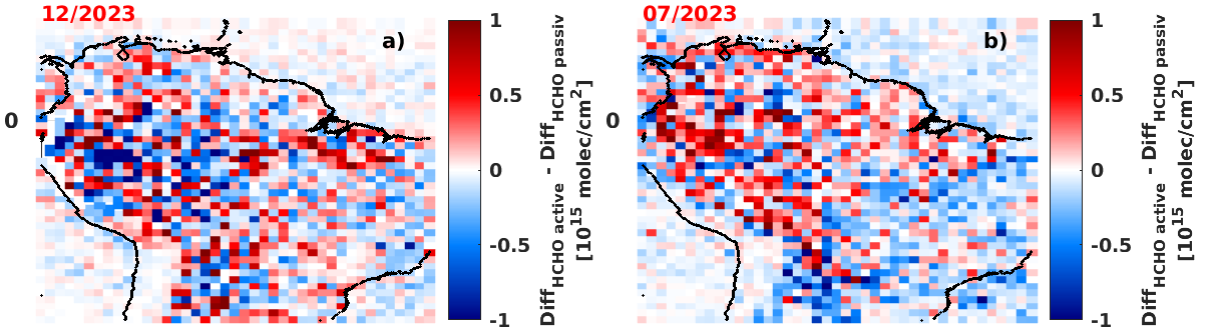


Figure 45: Zoom onto the South American rainforest region. Figure legend according to fig. 44 with blue indicating improved satellite and model agreement due to the assimilation of HCHO and red indicating a reduced agreement.

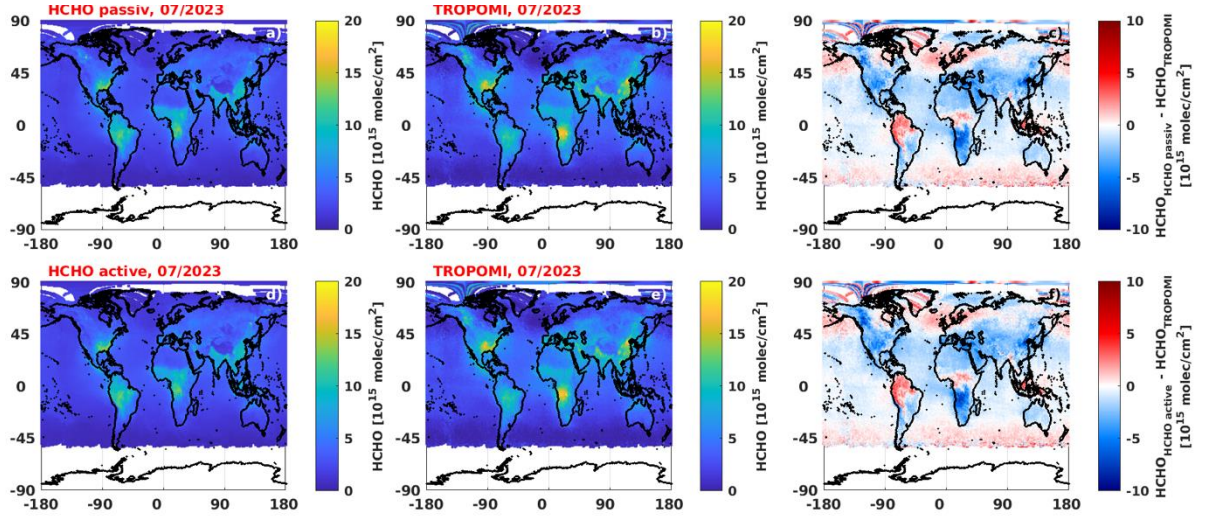


Figure 46: Comparison of monthly averaged formaldehyde VCDs from IFS-COMPO (CAMS 48R1) and the TROPOMI instrument without assimilation of HCHO (upper row) and when assimilating HCHO (lower row) for July 2023. The absolute difference of panels c) and f) is plotted in fig 39, panel b).

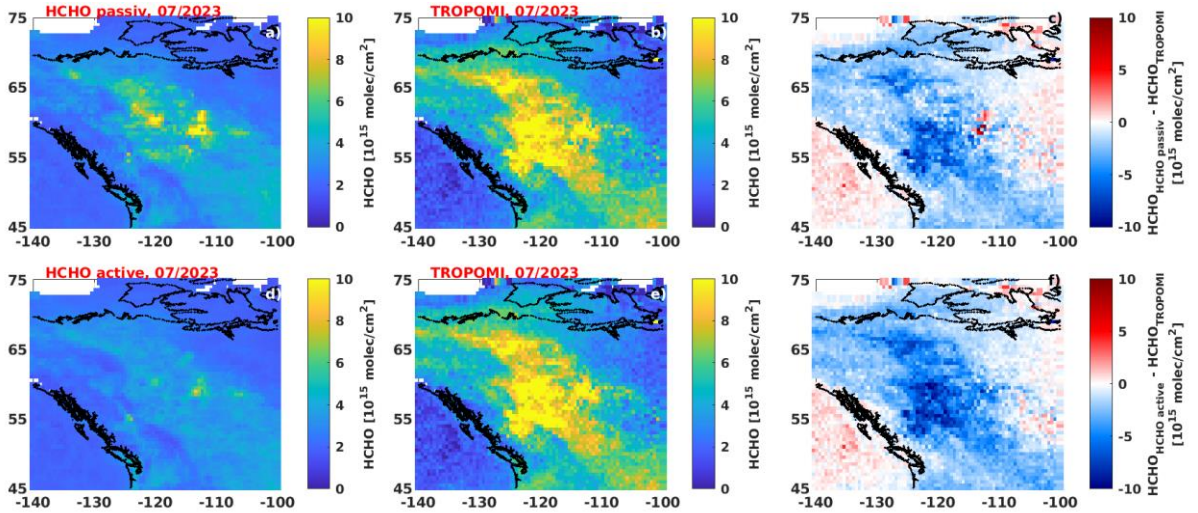


Figure. 47: Comparison of simulated (panels a, d) and satellite retrieved (panels b, e; same data) formaldehyde VCDs over western Canada in July 2023 without HCHO assimilation (upper row) and with HCHO assimilation (lower row).

5.3 Tangent linear and adjoint code of the linearized chemistry and corresponding HCHO assimilation tests

Working towards an emission inversion system of non-methane volatile organic compounds (NMVOCs) within the IFS, a tangent linear and adjoint code have been derived for the linearized HCHO chemistry scheme (see chapter 4). The tangent linear model will be used to calculate the forward model trajectory at a given time (and hence its cost function) and the adjoint model will derive the corresponding gradient of the cost function of the trajectory.

The tangent linear M' of model M with $x(t_{i+1}) = M[x(t_i)]$ has been computed such that $\partial x(t_{i+1}) = M'[x(t_i)] \cdot \partial x(t_i)$. The implementation of the tangent linear of the simplified HCHO chemistry scheme into the IFS is shown in detail in Appendix B. The adjoint M* of the tangent linear model M' of the simplified HCHO chemistry follows $\langle M'x, y\rangle = \langle x, M^*y\rangle$. The implementation into the IFS is shown in detail in Appendix C.

The tangent linear and adjoint codes have been implemented into the IFS system. They are currently tested and compiled using ifs-test. Once fully implemented and thoroughly tested, in a next step the HCHO initial conditions will be optimised simultaneously with the main biogenic precursor emissions (D2.4).

6 Conclusion

In this report we present first steps towards the implementation of an inversion system for biogenic volatile organic compound emissions in IFS-COMPO. Observational formaldehyde data from the TROPOMI instrument has successfully been prepared for use in IFS-COMPO and will be routinely monitored in the operational CAMS system from IFS cycle 49R1 onwards using observational filters and uncertainty parameters provided by the BIRA institute. The TROPOMI HCHO satellite observations have further been used for an extensive assessment of IFS-COMPO formaldehyde VCD simulations in CAMS cycle 48R1, and potential model biases have been analysed. Additionally, airborne formaldehyde measurements have been used to derive and analyse vertical formaldehyde profiles to enable an altitude resolved evaluation of IFS-COMPO formaldehyde mixing ratios. The comparison showed a general underestimation of measured formaldehyde in mixed polluted continental air, (e.g. Central Europe) and also over marine regions (e.g. Tropical Atlantic) throughout the entire troposphere, but most pronounced within the planetary boundary layer. The underestimation was less pronounced for background air (e.g. Patagonia) and air masses mainly affected by aged pollution plumes (e.g. East China Sea). Notable is the significant deviation from this global picture over the tropical rainforest north of the Amazon river delta, where we observe a significant overestimation of formaldehyde by IFS-COMPO. These general findings are confirmed by both, satellite and airborne formaldehyde observations. The overestimation of biogenic VOCs over the South American rainforest was reduced in CAMS cycle 49R1 compared to cycle 48R1, e.g. by updated formaldehyde photolysis rates.

Furthermore, a linearised and simplified chemistry scheme for HCHO has been developed. Employed sources, sinks, and reaction parameters have been optimised for low NOx-high HOX conditions dominated by biogenic emissions as they are commonly found over tropical rainforests. The scheme has been optimised to best match formaldehyde simulations based on the full IFS-COMPO using the modified CB05 tropospheric chemistry configuration while simultaneously reducing the complexity of the scheme to a minimum. Extensive testing showed reasonable agreement between full and simplified chemistry both in absolute formaldehyde as well as its diurnal cycle when considering one background formaldehyde source (methane) and one biogenic emission precursor trace gas (isoprene). Isoprene is treated both as a direct source of formaldehyde (by oxidation) and as secondary source (by oxidation of isoprene oxidation products).

Finally, first tests of the assimilation of HCHO observations in 4DVar context without tangent linear and adjoint and BVOC emission inversion have been performed. According to the short atmospheric lifetime of formaldehyde of few hours only, the impact of this assimilation on precursors and related trace gases (e.g. isoprene and tropospheric ozone) is only minor. Simultaneously, a tangent linear and adjoint code have been developed based on the proposed simplified HCHO chemistry scheme. As a next step, corresponding HCHO assimilation tests will be carried out with the objective to test the use of the tangent linear and adjoint of the simplified chemistry in the CAMS system.

7 References

Australian Government Bureau of Meteorology: Fire weather seasons. Accessed: 25/07/2024. http://www.bom.gov.au/weather-services/fire-weather-centre/fire-weather-season/.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmos. Meas. Tech., 11, 2395–2426, https://doi.org/10.5194/amt-11-2395-2018, 2018.

De Smedt, Isabelle, et al. "Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements." *Atmospheric Chemistry and Physics* 21.16 (2021): 12561-12593.

Douros, J., Eskes, H., van Geffen, J., Boersma, K. F., Compernolle, S., Pinardi, G., Blechschmidt, A.-M., Peuch, V.-H., Colette, A., and Veefkind, P.: Comparing Sentinel-5P TROPOMI NO2 column observations with the CAMS regional air quality ensemble, Geosci. Model Dev., 16, 509–534, https://doi.org/10.5194/gmd-16-509-2023, 2023.

Eskes, Henk, et al. "Evaluation of the Copernicus Atmosphere Monitoring Service Cy48R1 upgrade of June 2023." *EGUsphere* 2024 (2024): 1-57.

Fix, Andreas, et al. "CoMet: An airborne mission to simultaneously measure CO2 and CH4 using lidar, passive remote sensing, and in-situ techniques." *Epj web of conferences*. Vol. 176. EDP Sciences, 2018.

Andrés Hernández, M. Dolores, et al. "Overview: On the transport and transformation of pollutants in the outflow of major population centres–observational data from the EMeRGe European intensive operational period in summer 2017." *Atmospheric Chemistry and Physics* 22.9 (2022): 5877-5924.

Gao, F., Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput Optim Appl* **51**, 259–277 (2012). https://doi.org/10.1007/s10589-010-9329-3

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Hüneke, Tilman, et al. "The novel HALO mini-DOAS instrument: inferring trace gas concentrations from airborne UV/visible limb spectroscopy under all skies using the scaling method." *Atmospheric Measurement Techniques* 10.11 (2017): 4209-4234.

Kluge, Flora, et al. "Profiling of formaldehyde, glyoxal, methylglyoxal, and CO over the Amazon: normalized excess mixing ratios and related emission factors in biomass burning plumes." *Atmospheric Chemistry and Physics* 20.20 (2020): 12363-12389.

Lamarque, J-F., et al. "CAM-chem: Description and evaluation of interactive atmospheric chemistry in the Community Earth System Model." *Geoscientific Model Development* 5.2 (2012): 369-411.

Müller, J.-F., Stavrakou, T., and Peeters, J.: Chemistry and deposition in the Model of Atmospheric composition at Global and Regional scales using Inversion Techniques for Trace gas Emissions (MAGRITTE v1.1) – Part 1: Chemical mechanism, Geosci. Model Dev., 12, 2307–2356, https://doi.org/10.5194/gmd-12-2307-2019, 2019.

Müller, J.-F., Stavrakou, T., Oomen, G.-M., Opacka, B., De Smedt, I., Guenther, A., Vigouroux, C., Langerock, B., Aquino, C. A. B., Grutter, M., Hannigan, J., Hase, F., Kivi, R., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Röhling, A., Stremme, W., Strong, K., Sussmann, R., Té, Y., and Fried, A.: Bias correction of OMI HCHO columns based on FTIR and aircraft measurements and

impact on top-down emission estimates, Atmos. Chem. Phys., 24, 2207–2237, https://doi.org/10.5194/acp-24-2207-2024, 2024.

Myriokefalitakis, Stelios, et al. "Description and evaluation of a detailed gas-phase chemistry scheme in the TM5-MP global chemistry transport model (r112)." *Geoscientific Model Development Discussions* 2020 (2020): 1-64.

Rapp, Markus, et al. "SOUTHTRAC-GW: Gravity Waves Meet Sudden Stratospheric Warming." *Bulletin of the American Meteorological Society* 102.12 (2021): 1158-1166.

Rijsdijk, P., Eskes, H., Dingemans, A., Boersma, F., Sekiya, T., Miyazaki, K., and Houweling, S.: Quantifying uncertainties of satellite NO2 superobservations for data assimilation and model evaluation, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-632, 2024

Adrian Sandu, Rolf Sander, Michael S. Long, Robert M. Yantosca, Haipeng Lin, Lu Shen, & Daniel J. Jacob. (2024). KineticPreProcessor/KPP: The Kinetic PreProcessor (KPP) 3.1.1 (3.1.1). Zenodo. https://doi.org/10.5281/zenodo.11094840

Sindelarova, K., Markova, J., Simpson, D., Huszar, P., Karlicky, J., Darras, S., and Granier, C.: High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling, Earth Syst. Sci. Data, 14, 251–270, https://doi.org/10.5194/essd-14-251-2022, 2022.

Stavrakou, T., J. Peeters, and J-F. Müller. "Improved global modelling of HO x recycling in isoprene oxidation: evaluation against the GABRIEL and INTEX-A aircraft campaign measurements." *Atmospheric Chemistry and Physics* 10.20 (2010): 9863-9878.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, et al.: TROPOMI Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-13-3751-2020, 2020

Weller, Rolf, et al. "Meridional distribution of hydroperoxides and formaldehyde in the marine boundary layer of the Atlantic (48° N-35° S) measured during the Albatross campaign." *Journal of Geophysical Research: Atmospheres* 105.D11 (2000): 14401-14412.

Wendisch, Manfred, et al. "ACRIDICON–CHUVA campaign: Studying tropical deep convective clouds and precipitation over Amazonia using the new German research aircraft HALO." *Bulletin of the American Meteorological Society* 97.10 (2016): 1885-1908.

Williams, J. E., Huijnen, V., Bouarar, I., Meziane, M., Schreurs, T., Pelletier, S., Marécal, V., Josse, B., and Flemming, J.: Regional evaluation of the performance of the global CAMS chemical modeling system over the United States (IFS cycle 47r1), Geosci. Model Dev., 15, 4657–4687, https://doi.org/10.5194/gmd-15-4657-2022, 2022.

Zhu, L., González Abad, G., Nowlan, C. R., Chan Miller, C., Chance, K., Apel, E. C., DiGangi, J. P., Fried, A., Hanisco, T. F., Hornbrook, R. S., Hu, L., Kaiser, J., Keutsch, F. N., Permar, W., St. Clair, J. M., and Wolfe, G. M.: Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns, Atmos. Chem. Phys., 20, 12329–12345, https://doi.org/10.5194/acp-20-12329-2020, 2020.

8 List of Abbreviations

CAMS Copernicus Atmosphere Monitoring Service

FTIR Fourier-transform infrared spectroscopy

IFS Integrated Forecasting System

IFS-COMPO CAMS atmospheric composition forecasting system

S5P Sentinel 5-precursor

TL&AD Tangent linear and adjoint

TROPOMI TROPOspheric Monitoring Instrument

VCD Vertical Column density

VMR Volume Mixing Ratio

9 Appendix A. Simplified Chemistry Scheme source code.

The FORTRAN source code that implements the simplified chemistry in the IFS at “ifs-source/arpifs/chem/tm5_simple_hcho_chem.F90”. The source code is also available from the ifs-source branch nkmw_CY48R1.M_simplified_chem_methane_source, in routine *”tm5_simple_hcho_chem.F90.*

```
SUBROUTINE TM5_SIMPLE_HCHO_CHEM(YGFL,KIDIA,KFDIA,KLON,PDT,PRR,PRJ,PY)

  !**   DESCRIPTION
  !     ----------
  !
  !   Part of TM5 routines for IFS chemistry:
  !-----------------------------------------------------------------------------
  !   Eulerian backward Iteration
  !   Chemistry solver for simplified HCHO test chemistry
  !-----------------------------------------------------------------------------
  !
  !
  !
  !**   INTERFACE.
  !     ----------
  !          *TM5_SIMPLE_HCHO_CHEM* IS CALLED FROM *CHEM_tm5*.

  ! INPUTS:
  ! -------
  ! KIDIA :  Start of Array
  ! KFDIA :  End  of Array
  ! KLON  :  Length of Arrays
  ! PDT   :  Time step in seconds
  ! PRR  (KLON,NREAC)      : reaction rates
  ! PRJ  (KLON,NPHOTO)     : photolysis rates
  !
  !
  ! OUTPUTS:
  ! -------
  ! PY (KLON,NCHEM+3)        : final volume ratios OF TRACERS           (molec/cm3)
  !
  ! LOCAL:
  ! -------
  !
  !     AUTHOR.
  !     -------
  !        VINCENT HUIJNEN    *KNMI*
  !        FLORA KLUGE        *ECMWF*
  !        MIRO VAN DER WORP  *KNMI*
  !        TM5-community
  !
  !     MODIFICATIONS.
  !     --------------
  !        ORIGINAL : 2009-09-08


  USE PARKIND1 , ONLY : JPIM, JPRB
  USE YOMLUN , ONLY : NULERR
  USE YOMHOOK   ,ONLY : LHOOK, DR_HOOK, JPHOOK
  USE YOM_YGFL , ONLY : TYPE_GFLD
  USE TM5_TRACERS, ONLY :  IISOP_TLAD, IIOX_TLAD, IHCHO_TLAD, IOH, ICH4
  !* reaction rates
  USE TM5_CHEM_MODULE, ONLY : KC76, KOHISPD, KFRMOH, KCH4OH, NREAC
  USE TM5_PHOTOLYSIS , ONLY : NPHOTO,&
    !* photolysis rates
     & JACH2O,JBCH2O

```

```
58  IMPLICIT NONE
59
60  !----------------------------------------------------------------------
61  !*       0.1  ARGUMENTS
62  !             ---------
63
64  TYPE(TYPE_GFLD)   ,INTENT(INOUT):: YGFL
65  INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON
66  REAL(KIND=JPRB),INTENT(IN)    :: PDT
67  REAL(KIND=JPRB),INTENT(IN)    :: PRR(KLON,NREAC)
68  REAL(KIND=JPRB),INTENT(IN)    :: PRJ(KLON,NPHOTO)
69  REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,YGFL%NCHEM+3)  ! final concentrations
70
71  REAL(KIND=JPHOOK)    :: ZHOOK_HANDLE
72
73  REAL (KIND=JPRB)              :: ZRJA,ZRJB, ZRR_1,ZRR_2,ZRR_3, ZRR_4
74  REAL (KIND=JPRB)              :: ZPOX, ZLOX, ZPHCHO, ZLHCHO
75
76  REAL (KIND=JPRB),PARAMETER    :: ZALPHA = 0.83 ! HCHO yield from ISOP + OH
77  REAL (KIND=JPRB),PARAMETER    :: ZBETA  = 1.17 ! HCHO yield from OX + OH
78
79  ! * counters
80  INTEGER(KIND=JPIM) :: JL
81
82  !----------------------------------------------------------------------
83  !----------------------------------------------------------------------
84  !----------------------------------------------------------------------
85
86  IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',0,ZHOOK_HANDLE)
87
88  ! Production and loss chemistry for HCHO and its isoprene precursors
89  DO JL=KIDIA,KFDIA ! spatial loop.
90
91    ! DETERMINE REACTION RATES
92    ! ------------------------
93
94    ! Get ISOP+OH reaction rate.
95    ZRR_1=PRR(JL,KC76)* MAX(PY(JL,IOH),1.E-30_JPRB)
96
97    ! Get OX+OH reaction rate. Determined in box-model calculations by BIRA.
98    ZRR_2= 10.05E-12_JPRB * MAX(PY(JL,IOH),1.E-30_JPRB)
99
100   ! Get two HCHO photolysis rates and HCHO+OH reaction rate
101   ZRJA=PRJ(JL,JACH2O)
102   ZRJB=PRJ(JL,JBCH2O)
103   ZRR_3=PRR(JL,KFRMOH)* MAX(PY(JL,IOH),1.E-30_JPRB)
104
105   ! Methane loss rate:
106   ZRR_4 = PRR(JL,KCH4OH) * PY(JL,ICH4) * MAX(PY(JL,IOH),1.E-30_JPRB)
107
108   ! CHEMISTRYs
109   ! ----------
110
```

```
111     ! XO2 production:
112     ZPOX= PY(JL, IISOP_TLAD)*ZRR_1
113     ZLOX= ZRR_2
114
115     ! HCHO production and loss:
116     ZPHCHO= ZALPHA * PY(JL, IISOP_TLAD)*ZRR_1 &
117         & + ZBETA  * PY(JL, IIOX_TLAD)*ZRR_2 &
118         & + ZRR_4
119     ZLHCHO= ZRJA + ZRJB + ZRR_3
120     PY(JL, IHCHO_TLAD)=(PY(JL, IHCHO_TLAD) + ZPHCHO*PDT)/(1._JPRB+ZLHCHO*PDT)
121
122     ! Isoprene loss:
123     PY(JL, IISOP_TLAD)=1._JPRB/(1._JPRB+ZRR_1*PDT) * PY(JL, IISOP_TLAD)
124
125     ! IOX loss:
126     PY(JL, IIOX_TLAD)= (PY(JL, IIOX_TLAD)+ZPOX*PDT)/(1._JPRB+ZLOX*PDT)
127
128   ENDDO !JL
129
130   IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',1,ZHOOK_HANDLE)
131   END SUBROUTINE TM5_SIMPLE_HCHO_CHEM
```

10 Appendix B. Tangent linear code

Fortran code of the tangent linear model of the linearized HCHO chemistry (see fig. 22). The code can be found in ifs-source/arpifs/chem/tm5_simple_hcho_chem_tl.F90.

```
SUBROUTINE TM5_SIMPLE_HCHO_CHEM_TL(KIDIA,KFDIA,KLON,KINDICES_TLAD,PDT,PRR5,PRJ5,PY5,PY)

!**   DESCRIPTION
!     ----------
!
!   Part of TM5 routines for IFS chemistry:
!-----------------------------------------------------------------------------
!   Eulerian backward Iteration
!   Chemistry solver for simplified HCHO test chemistry
!-----------------------------------------------------------------------------
!
!
!
!**   INTERFACE.
!     ----------
!          *TM5_SIMPLE_HCHO_CHEM_TL* IS CALLED FROM *CHEM_tm5*.

! INPUTS:
! -------
! KIDIA :  Start of Array
! KFDIA :  End  of Array
! KLON  :  Length of Arrays
! PDT   :  Time step in seconds
! PRR  (KLON,NREAC)      : reaction rates
! PRJ  (KLON,NPHOTO)     : photolysis rates
!
!
! OUTPUTS:
! -------
! PY (KLON,NCHEM+3)        : final   volume ratios OF TRACERS          (mol/mol)
!
! LOCAL:
! -------
!
!
!     AUTHOR.
!     -------
!        VINCENT HUIJNEN    *KNMI*
!        TM5-community
!
```

```
41 !     MODIFICATIONS.
42 !     --------------
43 !        ORIGINAL : 2009-09-08
44
45
46
47 USE PARKIND1 , ONLY : JPIM     ,JPRB
48 USE YOMHOOK   ,ONLY : LHOOK, DR_HOOK, JPHOOK
49 !* reaction rates
50 USE YOMTM5TLAD
51 !USE CHEM_TM5
52
53 USE TM5_TRACERS, ONLY :  IISOP_TLAD, IIOX_TLAD, IHCHO_TLAD, IOH_TLAD, ICH4_TLAD
54 !USE TM5_CHEM_INI_TLAD
55
56 IMPLICIT NONE
57
58 !----------------------------------------------------------------------
59 !*       0.1  ARGUMENTS
60 !             ---------
61
62 INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON
63 INTEGER(KIND=JPIM),INTENT(IN) :: KINDICES_TLAD(NCHEM_TLAD)
64 REAL(KIND=JPRB),INTENT(IN)    :: PDT
65 REAL(KIND=JPRB),INTENT(IN)    :: PRR5(KLON,NREAC)
66 REAL(KIND=JPRB),INTENT(IN)    :: PRJ5(KLON,NPHOTO)
67 REAL(KIND=JPRB),INTENT(INOUT)    :: PY5(KLON,NCHEM_TLAD)
68 REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,NCHEM_TLAD)
69
70 ! * LOCAL
71 REAL(KIND=JPHOOK)    :: ZHOOK_HANDLE
72
73 INTEGER(KIND=JPIM)    :: ITER
74
75 REAL (KIND=JPRB)             :: ZRJA,ZRJB, ZRR_1,ZRR_2,ZRR_3,ZRR_4
76 REAL (KIND=JPRB)             :: ZPOX, ZLOX, ZPHCHO, ZLHCHO
77 REAL (KIND=JPRB)             :: ZRJA5,ZRJB5, ZRR_15,ZRR_25,ZRR_35,ZRR_45
78 REAL (KIND=JPRB)             :: ZPOX5, ZLOX5, ZPHCHO5, ZLHCHO5
79
80 REAL (KIND=JPRB),PARAMETER   :: ZALPHA = 0.83 ! HCHO yield from ISOP + OH
81 REAL (KIND=JPRB),PARAMETER   :: ZBETA  = 1.17 ! HCHO yield from OX + OH
```

```
82
83 ! * counters
84 INTEGER(KIND=JPIM) :: JL
85
86 !-----------------------------------------------------------------------
87 !-----------------------------------------------------------------------
88 !-----------------------------------------------------------------------
89
90 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',0,ZHOOK_HANDLE)
91
92
93
94 ! Production and loss chemistry for HCHO and its isoprene precursors
95 DO JL=KIDIA,KFDIA
96
97
98 ! TANGENT-LINEAR CODE
99 !========================
100
101    ZRR_15 = PRR5(JL,KC76)   * PY5(JL,IOH_TLAD)
102    ZRR_25 = PRR5(JL,KOHIOX) * PY5(JL,IOH_TLAD) ! IOX+OH RR from BIRA
103    ZRR_35 = PRR5(JL,KFRMOH) * PY5(JL,IOH_TLAD)
104    ZRR_45 = PRR5(JL,KCH4OH) * PY5(JL,ICH4_TLAD) * PY5(JL,IOH_TLAD)
105
106    ! photolysis hcho
107    ZRJA5 = PRJ5(JL,JACH2O)
108    ZRJB5 = PRJ5(JL,JBCH2O)
109
110    !isoprene
111    PY5(JL,IISOP_TLAD) = 1._JPRB/(1._JPRB+ZRR_15*PDT) * PY5(JL,IISOP_TLAD)
112    PY (JL,IISOP_TLAD) = 1._JPRB/(1._JPRB+ZRR_15*PDT) * PY(JL,IISOP_TLAD)
113
114    !iox
115    ZLOX5 = ZRR_25
116
117    ZPOX5 = PY5(JL,IISOP_TLAD) * ZRR_15
118    ZPOX  = PY (JL,IISOP_TLAD) * ZRR_15
119
120    PY5(JL,IIOX_TLAD) = (PY5(JL,IIOX_TLAD)+ZPOX5*PDT)/(1._JPRB+ZLOX5*PDT)
121    PY(JL,IIOX_TLAD)  = PY(JL,IIOX_TLAD) * (1._JPRB/(1._JPRB+PDT*ZLOX5)) + &
122                      & (PDT / (1._JPRB+PDT*ZLOX5)) * ZPOX
123
124    !hcho production
125    ZPHCHO5 = ZALPHA * PY5(JL,IISOP_TLAD) *ZRR_15 &
126          & + ZBETA  * PY5(JL,IIOX_TLAD)  *ZRR_25 &
127          & + ZRR_45
128
129    ZPHCHO = ZALPHA * ZRR_15 * PY(JL,IISOP_TLAD) &
130         & + ZBETA  * ZRR_25 * PY(JL,IIOX_TLAD)
131
132    !hcho loss
133    ZLHCHO5= ZRJA5 + ZRJB5 + ZRR_35
134
135    !equilibrium hcho
136    PY5(JL,IHCHO_TLAD)=(PY5(JL,IHCHO_TLAD)+ZPHCHO5*PDT)/(1._JPRB+ZLHCHO5*PDT)
137
138    PY(JL,IHCHO_TLAD) = PY(JL,IHCHO_TLAD) / (1._JPRB + PDT * ZLHCHO5) + PDT * ZPHCHO / (1._JPRB + PDT * ZLHCHO5)
139
140
141
142 ENDDO !JL
143
144
145 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM_TL',1,ZHOOK_HANDLE)
146 END SUBROUTINE TM5_SIMPLE_HCHO_CHEM_TL
```

11 Appendix C. Adjoint code

Fortran code of the adjoint model of the linearized HCHO chemistry (see fig. 22). The code can be found in ifs-source/arpifs/chem/tm5_simple_hcho_chem_ad.F90

```
SUBROUTINE TM5_SIMPLE_HCHO_CHEM_AD(KIDIA,KFDIA,KLON,KINDICES_TLAD,PDT,PRR5,PRJ5,PY5,PY)

!**   DESCRIPTION
!     ----------
!
!   Part of TM5 routines for IFS chemistry:
!-----------------------------------------------------------------------
!   Eulerian backward Iteration
!   Chemistry solver for simplified HCHO test chemistry
!-----------------------------------------------------------------------
!
!
!
!**   INTERFACE.
!     ----------
!          *TM5_SIMPLE_HCHO_CHEM_AD* IS CALLED FROM *CHEM_tm5*.

! INPUTS:
! -------
! KIDIA :  Start of Array
! KFDIA :  End  of Array
! KLON  :  Length of Arrays
! PDT   :  Time step in seconds
! PRR  (KLON,NREAC)      : reaction rates
! PRJ  (KLON,NPHOTO)     : photolysis rates
!
!
! OUTPUTS:
! -------
! PY (KLON,NCHEM+3)        : final   volume ratios OF TRACERS           (mol/mol)
!
! LOCAL:
! -------
!
!
!     AUTHOR.
!     -------
!        VINCENT HUIJNEN    *KNMI*
!        TM5-community
!
!     MODIFICATIONS.
!     --------------
!        ORIGINAL : 2009-09-08



USE PARKIND1 , ONLY : JPIM     ,JPRB
USE YOMHOOK   ,ONLY : LHOOK, DR_HOOK, JPHOOK
!* reaction rates
USE YOMTM5TLAD
!USE CHEM_TM5

USE TM5_TRACERS, ONLY :  IISOP_TLAD, IIOX_TLAD, IHCHO_TLAD, IOH_TLAD, ICH4_TLAD

!USE TM5_CHEM_INI_TLAD

IMPLICIT NONE

!-----------------------------------------------------------------------
!*       0.1  ARGUMENTS
!             ---------
```

```
62
63 INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON
64 INTEGER(KIND=JPIM),INTENT(IN) :: KINDICES_TLAD(NCHEM_TLAD)
65 REAL(KIND=JPRB),INTENT(IN)    :: PDT
66 REAL(KIND=JPRB),INTENT(IN)    :: PRR5(KLON,NREAC)
67 REAL(KIND=JPRB),INTENT(IN)    :: PRJ5(KLON,NPHOTO)
68 REAL(KIND=JPRB),INTENT(INOUT) :: PY5(KLON,NCHEM_TLAD)   ! tracer concentrations  (trajectory)
69 REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,NCHEM_TLAD)    ! tracer concentrations  (adjoint variable)
70
71 ! * LOCAL
72 REAL(KIND=JPHOOK)    :: ZHOOK_HANDLE
73
74 INTEGER(KIND=JPIM)    :: ITER
75 !INTEGER(KIND=JPIM)    :: IISOP_TLAD, IIOX_TLAD, IHCHO_TLAD, IFLOH, IFLCH4
76
77 REAL (KIND=JPRB)             :: ZRJA,ZRJB, ZRR_1,ZRR_2,ZRR_3, ZRR_4
78 REAL (KIND=JPRB)             :: ZPOX, ZLOX, ZPHCHO, ZLHCHO
79 REAL (KIND=JPRB)             :: ZRJA5,ZRJB5, ZRR_15,ZRR_25,ZRR_35, ZRR_45
80 REAL (KIND=JPRB)             :: ZPOX5, ZLOX5, ZPHCHO5, ZLHCHO5
81
82 REAL (KIND=JPRB),PARAMETER   :: ZALPHA = 0.83 ! HCHO yield from ISOP + OH
83 REAL (KIND=JPRB),PARAMETER   :: ZBETA  = 1.17 ! HCHO yield from OX + OH
84
85 ! * counters
86 INTEGER(KIND=JPIM) :: JL
87
88 !-----------------------------------------------------------------------
89 !-----------------------------------------------------------------------
90 !-----------------------------------------------------------------------
91
92 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',0,ZHOOK_HANDLE)
93
94
95
96 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! ADJOINT CODE !!!!!!!!!!!!!!!!!!!!!!!!!!!!
97 !=============================================================================================
98
99 DO JL=KIDIA,KFDIA
100
101    !forward trajectories
102    !=============================================================================================
```

```
103    ZRR_15=PRR5(JL,KC76)    * PY5(JL,IOH_TLAD)
104    ZRR_25 = PRR5(JL,KOHIOX) * PY5(JL,IOH_TLAD) ! IOX+OH RR from BIRA
105    ZRR_35=PRR5(JL,KFRMOH) * PY5(JL,IOH_TLAD)
106    ZRR_45=PRR5(JL,KCH4OH) * PY5(JL,ICH4_TLAD) * PY5(JL,IOH_TLAD)
107
108    ZRJA5=PRJ5(JL,JACH2O)
109    ZRJB5=PRJ5(JL,JBCH2O)
110    PY5(JL,IISOP_TLAD)=1._JPRB/(1._JPRB+ZRR_15*PDT) * PY5(JL,IISOP_TLAD)
111    ZPOX5= PY5(JL,IISOP_TLAD) * ZRR_15
112    ZLOX5= ZRR_25
113    PY5(JL,IIOX_TLAD)=(PY5(JL,IIOX_TLAD)+ZPOX5*PDT)/(1._JPRB+ZLOX5*PDT)
114    ZLHCHO5= ZRJA5 + ZRJB5 + ZRR_35
115    ZPHCHO5= ZALPHA * PY5(JL,IISOP_TLAD) *ZRR_15 &
116         & + ZBETA  * PY5(JL,IIOX_TLAD)   *ZRR_25 &
117         & + ZRR_45
118    PY5(JL,IHCHO_TLAD)=(PY5(JL,IHCHO_TLAD)+ZPHCHO5*PDT)/(1._JPRB+ZLHCHO5*PDT)
119
120    ! adjoint code
121    ! ============================================================================
122    ! initialise adjoint variables
123    ZPOX =  0.0_JPRB
124    ZPHCHO = 0.0_JPRB
125
126    !adjoint code
127    ZPHCHO = ZPHCHO + PY(JL,IHCHO_TLAD) * PDT / (1._JPRB + PDT * ZLHCHO5)
128    PY(JL,IHCHO_TLAD) =  PY(JL,IHCHO_TLAD) / (1._JPRB + PDT * ZLHCHO5)
129    PY(JL,IIOX_TLAD)  = PY(JL,IIOX_TLAD)  + ZBETA  * ZRR_25 * ZPHCHO
130    PY(JL,IISOP_TLAD) = PY(JL,IISOP_TLAD) + ZALPHA * ZRR_15 * ZPHCHO
131    ZPHCHO = 0.0_JPRB
132    ZPOX = ZPOX + PDT / (1._JPRB + PDT * ZLOX5) * PY(JL,IIOX_TLAD)
133    PY(JL,IIOX_TLAD)=  PY(JL,IIOX_TLAD) * (1._JPRB/(1._JPRB+PDT*ZLOX5))
134    PY(JL,IISOP_TLAD)=PY5(JL,IISOP_TLAD) + ZRR_15 * ZPOX
135    ZPOX = 0.0_JPRB
136    PY(JL,IISOP_TLAD)=PY(JL,IISOP_TLAD) * (1._JPRB/(1._JPRB+PDT*ZRR_15))
137
138 ENDDO
139
140
141 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM_AD',1,ZHOOK_HANDLE)
142 END SUBROUTINE TM5_SIMPLE_HCHO_CHEM_AD
```

Document History

Version	Author(s)	Date	Changes
1.0	Flora Kluge, Vincent Huijnen, Antje Inness, Glenn-Michael Oomen, Jean-François Müller, Trissevgeni Stavrakou, Miró van der Worp	09.09.2024	Initial version
1.1	Flora Kluge,	September 2024	Updated after internal review
1.2	Rhona Phipps	September 2024	Resolving formatting issues

Internal Review History

Internal Reviewers	Date	Comments
Ana Carvalho, Lennart Robertson (SMHI) and Dick Heslinga (TNO)	Sept 2024	Internal review comments