CAMS Service Evolution

D2.4 Report on NMVOC emission inversion

Due date of deliverable	30.11.2025
Submission date	01.12.2025
File Name	Report on NMVOC emission inversion
Work Package /Task	CAMEO-D2.4-V1.0
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Flora Kluge
Revision number	1.0
Status	Final
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 the implementation of a satellite-retrieval based inversion system for surface fluxes of biogenic volatile organic compounds in the IFS global model. The scheme is based on formaldehyde (HCHO) satellite observations from multiple satellite instruments. As part of the work, an assimilation capacity for formaldehyde was developed for use in IFS-COMPO, which uses the 4DVAR data assimilation technique. The extension for HCHO assimilation applies the tangent linear and adjoint of the previously implemented simplified formaldehyde-isoprene chemistry scheme. The purpose of the adjoint simplified chemistry scheme is to enable a modification of the isoprene fields based on the assimilation of HCHO observations. The assimilation of polar-orbiting and geostationary HCHO satellite retrievals from TROPOMI, TEMPO, and GEMS is carried out and validated using ground-based and spaceborne formaldehyde observations. Globally, the assimilation has a neutral to slightly positive impact on IFS-COMPO HCHO simulations and negligible impact on other atmospheric compounds. Significant differences in IFS-COMPO HCHO fields due to the HCHO assimilation were not found for any of the evaluated time periods, air mass compositions or global regions. Based on the tangent linear and adjoint of the simplified HCHO chemistry scheme, also an inversion system for biogenic volatile organic compounds emissions was technically developed and applied for first research application with IFS-COMPO. First HCHO-satellite observation-based optimisations of the biogenic precursor emissions were successfully performed. In its first implemented version, the information content of the HCHO satellite retrievals at the surface is limited, such that the inversion system produces only negligible differences between bottom-up and top-down isoprene emissions. While a clear sensitivity of the implemented HCHO chemistry to isoprene emissions could be shown, this does not translate into the assimilation and the inversion system in their current implementation. Further analysis and optimisation of the implemented chemistry, the HCHO tangent linear and adjoint, the a priori error assumptions and the emissions is planned.

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 towards an emission inversion system for non-methane biogenic volatile organic compounds in the IFS. To this scope, an assimilation system for HCHO was developed and thoroughly tested that makes use of the tangent linear and adjoint of a simplified chemistry scheme. Further essential steps include the development of an inversion system for IFS-COMPO that simultaneously optimises the HCHO initial conditions and the underlying biogenic precursor emissions, and the assessment of its impact on the biogenic emissions as well as on related atmospheric tracers.

2.2.2 Work performed in this deliverable

In this deliverable the work as planned in the Description of Action (DoA, WP2.2 T2.3.3 and T2.3.4) was performed:

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.

Task 2.3.4: NMVOC emission inversion with IFS and assessment against other datasets. Simultaneously optimise the HCHO initial conditions and the mainly biogenic precursor emission. Define background error description for precursor emissions and enable the flux-inversion capabilities. Assess impact of flux inversion on HCHO, CO and O3 with independent data. Evaluate optimised biogenic fluxes against independent systems or bottom-up approaches. Use uncertainty information for biogenic emissions developed in WP5.

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 HCHO simplified chemistry

The following section describes the development and the evaluation of a formaldehyde 4D-Var assimilation system for the IFS global model based on tangent linear and adjoint of a simplified version of atmospheric formaldehyde chemistry. A simplified representation of the chemistry is required because the tangent linear and adjoint representation of the complete chemistry is expected to be computationally too expensive, and it is not available. The work in this section builds upon previous efforts from this work package that are described in more detail in CAMEO deliverable D2.3. Previously, a simplified formaldehyde chemistry scheme was developed and optimised for specific atmospheric conditions (biogenic air masses, see D2.3). Such biogenically-dominated air is typical for the major source regions of formaldehyde (tropical rain forests) and as such, its chemical conditions served as the best first-guess approach for operational assimilation of HCHO in IFS-COMPO. It is, however, not an ideal representation of formaldehyde atmospheric chemistry in air masses of non-biogenic composition, e.g. air masses affected by anthropogenic emissions. Section 3.1 reports on a slightly more complex simple HCHO chemistry that aims to expand the simple chemistry from remote to global atmospheric conditions, in particular to air masses under anthropogenic influence.

3.1 Optimisation of the simplified chemistry for high-NOX

The assimilation of HCHO within the global atmospheric composition model IFS-COMPO requires an atmospheric formaldehyde chemistry of significantly reduced complexity as compared to the standard chemistry in the IFS-COMPO TM5 and CB05 to allow for the development of the tangent linear and adjoint of the chemistry scheme. Such a reduction of the number of involved atmospheric tracers and chemical feedback mechanisms requires the application of artificial reactions and the optimisation of reaction parameters. The strong dependence of those parameters on the background atmospheric conditions necessarily restricts the optimal-fit of the scheme to specific atmospheric regimes. In the case of atmospheric formaldehyde, the main external dependencies of the reaction parameters are high and low NOx concentrations in conjunction with high and low HOx conditions. For formaldehyde source regions (tropical rain forests, high vegetation areas), usually low NOx/high HOx conditions apply. Consequently, as a first step towards HCHO assimilation, a simplified chemistry scheme was developed that was optimised for clean, biogenically dominated air masses (low NOx-high HOx, see D2.3):

Table 1: Formaldehyde simplified chemistry 1 (SC1) optimised for low NOx.

R1.1 ISOP + OH $\rightarrow$ a HCHO + IOX,	a=0.83, rate: 10^{-10} s^{-1}
R1.2 IOX + OH $\rightarrow$ b HCHO,	b=1.17, rate: 1.05 x 10^{-11} s^{-1}
R1.3 CH4 + OH $\rightarrow$ c HCHO,	c=1, rate as in standard model chemistry
L1 HCHO + OH $\rightarrow$ CO + 2 HO_2,	rate as in standard model chemistry
L2 HCHO + hv $\rightarrow$ products	rate as in standard model chemistry

However, global HCHO assimilation requires the expansion of this first chemistry scheme towards all-atmosphere conditions. To represent the globally varying, but significant impact of anthropogenic emissions, an optimisation of the involved reactions and parameters was performed under the assumption of increasing NOx concentrations towards high NOx/low HOx conditions.

Isoprene, being the main atmospheric formaldehyde precursor besides methane, has the main impact on its secondary production. Isoprene chemistry itself is strongly NOx dependent. As a result, the expansion of the previously developed simplified chemistry scheme (table 1) to other NOx regimes requires additional reactions and parameters. At the same time, the scheme's complexity continues to be as reduced as possible to still match its overall purpose, i.e. the application in operational data assimilation. Among all involved chemical steps, the reactions of the peroxy radical from from ISOP + OH ($ISOPO_2$) were found to be the most significant NOx dependent formaldehyde precursor reactions. Under low NOx conditions, $ISOPO_2$ mainly reacts with HO_2. This was accounted for in the low NOx simplified chemistry scheme by reaction R1.1. Under high NOx, $ISOPO_2$ increasingly reacts with NO instead of HO_2. This reaction has a different final formaldehyde yield than the $ISOPO_2$+HO_2 pathway due to 1) direct production of formaldehyde and 2) secondary production through intermediate isoprene oxidation products. As done in the first simplified scheme, this secondary formaldehyde production is again represented by a lumped tracer IOX. The NO_X-dependent two main reaction pathways of $ISOPO_2$ to formaldehyde need to be separately accounted for. The optimisation of the first chemistry scheme for low and high NO_X conditions therefore requires the inclusion of two separate $ISOPO_2$ reaction pathways to the first scheme (Table 2, reactions R2.2 and 2.3):

Table 2: Formaldehyde simplified chemistry 2 (SC2) optimised for both low and high NOX.

R2.1 ISOP + OH	$\rightarrow$ a $ISOPO_2$,	a=1, rate: 10^{-10} s^{-1}
R2.2 $ISOPO_2$ + HO_2	$\rightarrow$ b IOX,	b=1.89, rate: 1.65 x 10^{-11} s^{-1}
R2.3 $ISOPO_2$ + NO	$\rightarrow$ c HCHO + d IOX,	c=1.42, d=0.67, rate: 8.7 x 10^{-12} s^{-1}
R2.4 IOX + OH	$\rightarrow$ e HCHO,	e=1, rate: 5.57 x 10^{-12} s^{-1}
R2.5 CH4 + OH	$\rightarrow$ f HCHO,	f=1, rate as in standard model chemistry
L1 HCHO + OH	$\rightarrow$ CO + 2 HO_2,	rate as in standard model chemistry
L2 HCHO + hv	$\rightarrow$ products	rate as in standard model chemistry

Reaction rates R2.1 - R2.3 of chemistry scheme 2 correspond to the standard model chemistry (without temperature dependency). As for formaldehyde chemistry scheme 1, the reaction parameters and rates of reaction R2.4 were derived in box-model simulations and tuned with respect to IFS(CB05) chemistry. The simulations were performed by BIRA with the scope of minimising the least-square difference between cumulative HCHO from the proposed scheme and HCHO derived in pre-computed IFS-COMPO box model runs. The optimisation follows the derivation of the first simple chemistry scheme and is outlined in detail in deliverable D2.3, section 4.1. In the box model simulations, initial concentrations of 1ppb isoprene, 120 ppb CO, 28 ppb O_3 and varying NO2 of 0.1 ppb, 0.3 ppb, 1.0 ppb, and 3.0 ppb were assumed. Total NO_X as well as O_2, N_2, H_2O, and N_2O were kept constant during each simulation. For the first guess of the parameters, the coefficients a, b, c, and d were set to 1, and the rate of R2.4 was set to 10^{-11} s^{-1}. Additionally, different first-guesses were used to ensure convergence to a global minimum in the parameter space. The final reaction parameters are listed in Table 2.

3.2 Comparison of the two simplified schemes

The following section analyses the performance of the two simplified schemes in IFS-COMPO in more detail. The analysis is carried out based on two forecast experiments that were run for one example month (July 2023) using climatological biogenic emissions from CAMS-GLOB-

BIOv3.1 and each using one of the simplified HCHO chemistry schemes (table 3). The source code for each chemistry scheme can be found in Appendix A.

Over marine surfaces, the changes in isoprene chemistry evidently have no effect due to the absence of surface emissions and the two chemistry schemes agree within 0.3% (fig. 3.1, panel c). To focus on the actual impact of the modified isoprene oxidation, in the further analysis a land mask is applied on the data. Over terrestrial surfaces, the SC2 simplified chemistry reduces formaldehyde VCDs by -2.4% compared to simplified chemistry SC1. Since the SC1 already underestimates standard IFS-COMPO HCHO, this further degrades the comparison of simplified and full chemistry HCHO by 4% with SC1 resulting in a global median underestimation of standard IFS-COMPO HCHO of -15.5% and SC2 of -19.5% (fig. 3.2, panels a and b and fig. 3.3). These differences in HCHO from the two chemical schemes mainly result from different direct and indirect HCHO yields from isoprene oxidation in SC1 and SC2 both in low and high NOx air. This is discussed in more detail below.

Table 3: Forecast experiments for the evaluation of the two simplified HCHO chemistry schemes.

Name	experiment ID	branch	HCHO chemistry
SC1	b2un	cxfk_CY49R1_simple_chem_1	low NO_X simple scheme
SC2	b2uq	cxfk_CY49R1_simple_chem_2	variable NO_X simple scheme

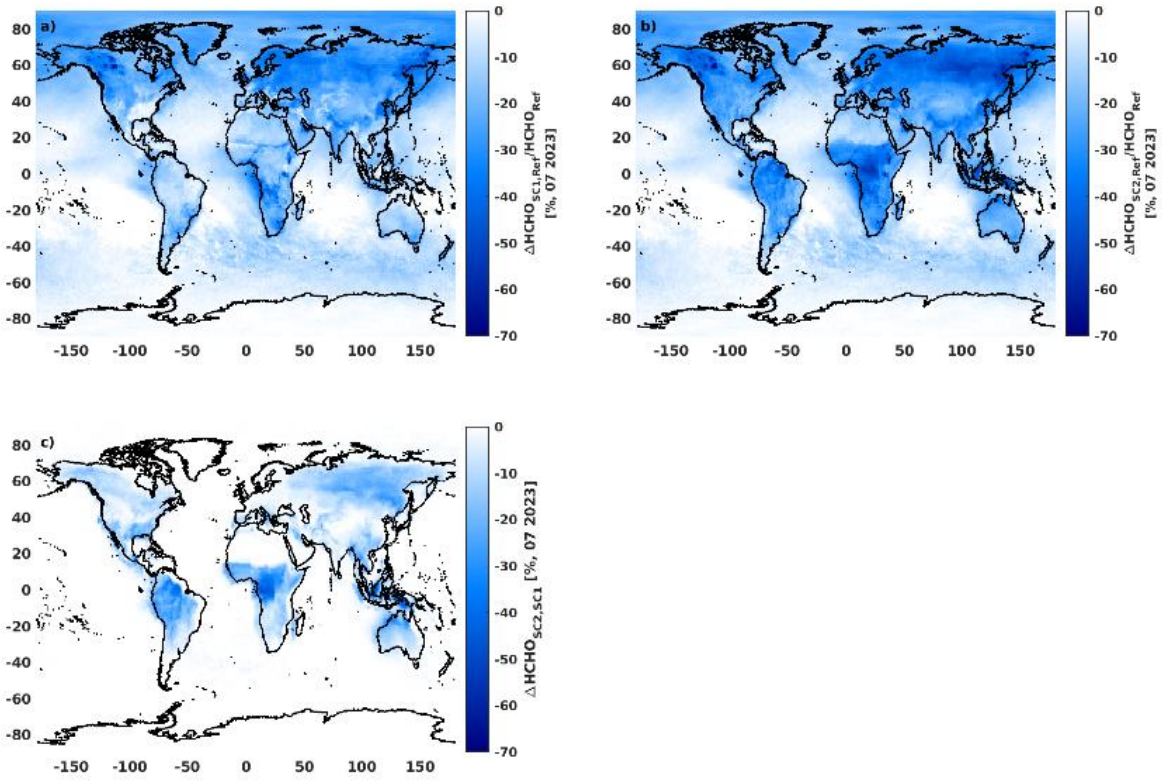


Fig. 3.1: Difference of the two simplified chemistry schemes and standard IFS-COMPO formaldehyde for July 2023. Panel a) and b) show the relative difference between HCHO from the two schemes and standard IFS-COMPO chemistry as ($HCHO_{SC1/2}$ – $HCHO_{ref}$)/$HCHO_{ref}$. Panel c) plots the relative difference of the two simplified formaldehyde schemes as ($HCHO_{SC2}$ - $HCHO_{SC1}$)/$HCHO_{SC1}$.

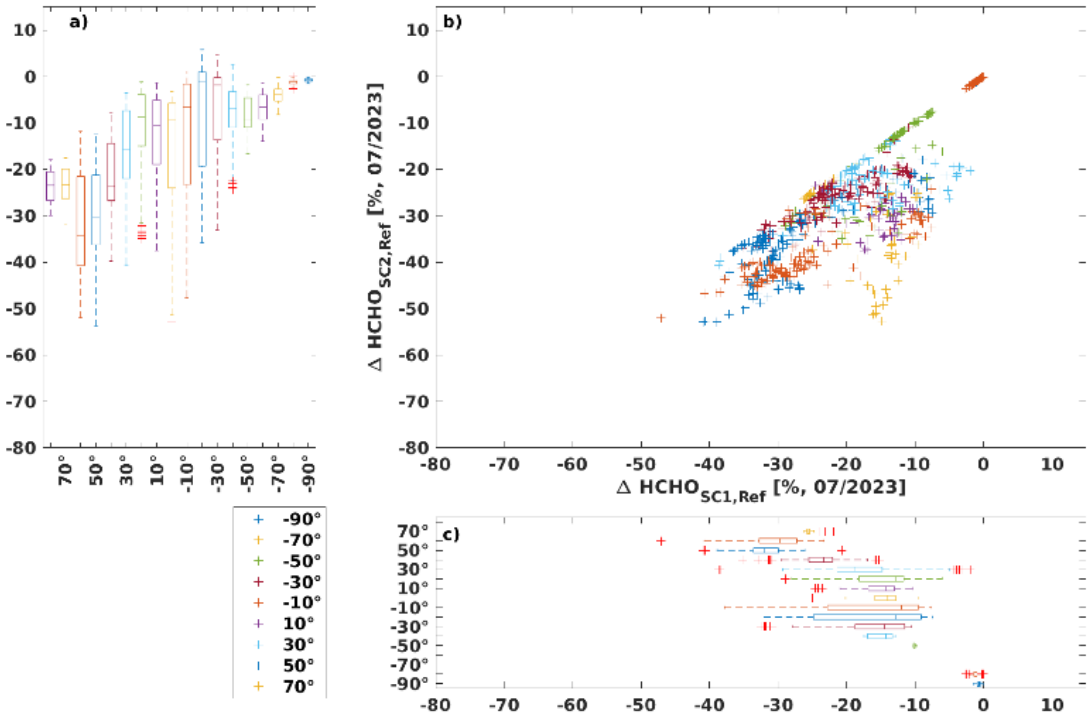


Fig. 3.2: Relative deviation of the simplified HCHO chemistry schemes to standard IFS-COMPO formaldehyde for simplified chemistry 1 ($(HCHO_{SC1}-HCHO_{ref})/HCHO_{ref}$, x-axis) and simplified chemistry 2 ($(HCHO_{SC2}-HCHO_{ref})/HCHO_{ref}$, y-axis) as a function of latitude (color-code) shown in panel b. The boxplots in panels a) and c) indicate the median relative difference for each degree latitude averaged over all longitudes (oceans masked out) with the boxes showing the lower 25 and higher 75 percentiles, dotted lines the minima and maxima, and red symbols statistical outliers. The line in each box marks the median. The data are averaged on a 10° latitudinal grid and filtered for land-only due to the marginal difference between the schemes over regions with no sources (i.e. marine areas, see fig. 3.1). The ocean mask leaves some data over large lakes, causing the entries with approximately 1:1 agreement at -10°N and -50°N. Evidently, the relative difference of SC1 to standard IFS-COMPO HCHO is smaller than of SC2.

Over remote isoprene source regions, the impact of the modified isoprene oxidation in the simple chemistries is globally the largest. This causes the largest deviations in HCHO to occur in the tropics with up to -54% difference between the schemes (fig. 3.1, panel c). Over the Amazonian rain forest (-15°N to 5°N, -55°E to -75°E), SC1 has a median VCD of $1.9\pm0.6\text{x}10^{16}$ molec cm^{-2} and SC2 is on average 18% smaller with a median VCD of $1.5\pm0.4\text{x}10^{16}$ molec cm^{-2} ($2.1\pm0.6\text{x}10^{16}$ molec cm^{-2} standard IFS-COMPO). In this region, both schemes underestimate the full chemistry HCHO by -10% (SC1) and -32% (SC2), respectively (fig. 3.2, panels a and c). The larger HCHO from SC1 over remote regions (low NOx limit) results from the larger direct and indirect HCHO yield from isoprene oxidation in SC1 compared to the only indirect and smaller HCHO yield from SC2 (2.0 vs 1.89). Smaller differences between the schemes are found outside of the tropics and for higher latitudes. Due to the optimisation of SC2 for a broader range of NOx conditions, the largest improvement for SC2 is expected over highly populated and polluted regions. At the high NOx limit (where ISOPO2 reacts mainly with NO), the HCHO yield from isoprene oxidation is 2.09 (partly indirect), which is higher than in SC1 (yield of 2). For mixed conditions with lower NOx, these yields are slightly lower. Over such mixed polluted regions like continental Europe or the US American and Chinese

coastlines, the two chemistries perform either similarly, or the low NOx scheme shows slightly better agreement to standard IFS-COMPO HCHO. In fact, over continental Europe (30°N to 70°N, -15°E to 40°E), HCHO from the two schemes agrees within 2%, despite some larger divergences over the Mediterranean coastlines (fig. 3.3).

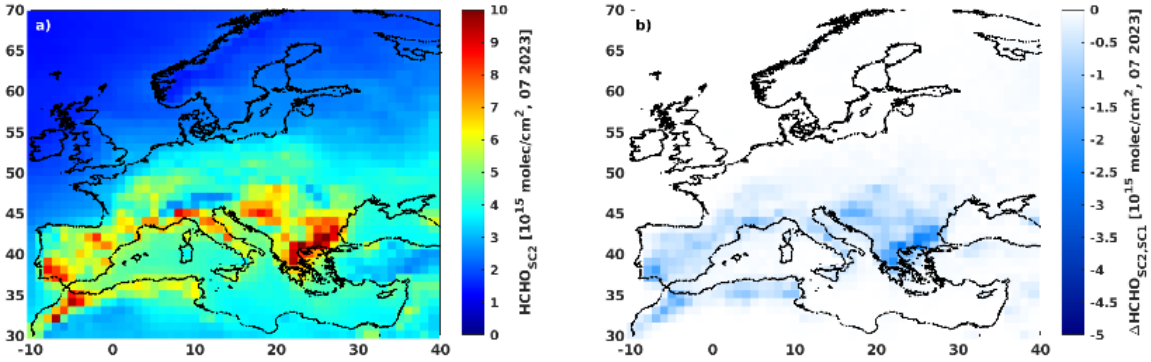


Fig. 3.3: Formaldehyde VCD from the simple chemistry SC2 over Europe for July 2023 (panel a) and the absolute difference HCHO$_{SC2}$ - HCHO$_{SC1}$ to formaldehyde from the simple chemistry SC1 (panel b).

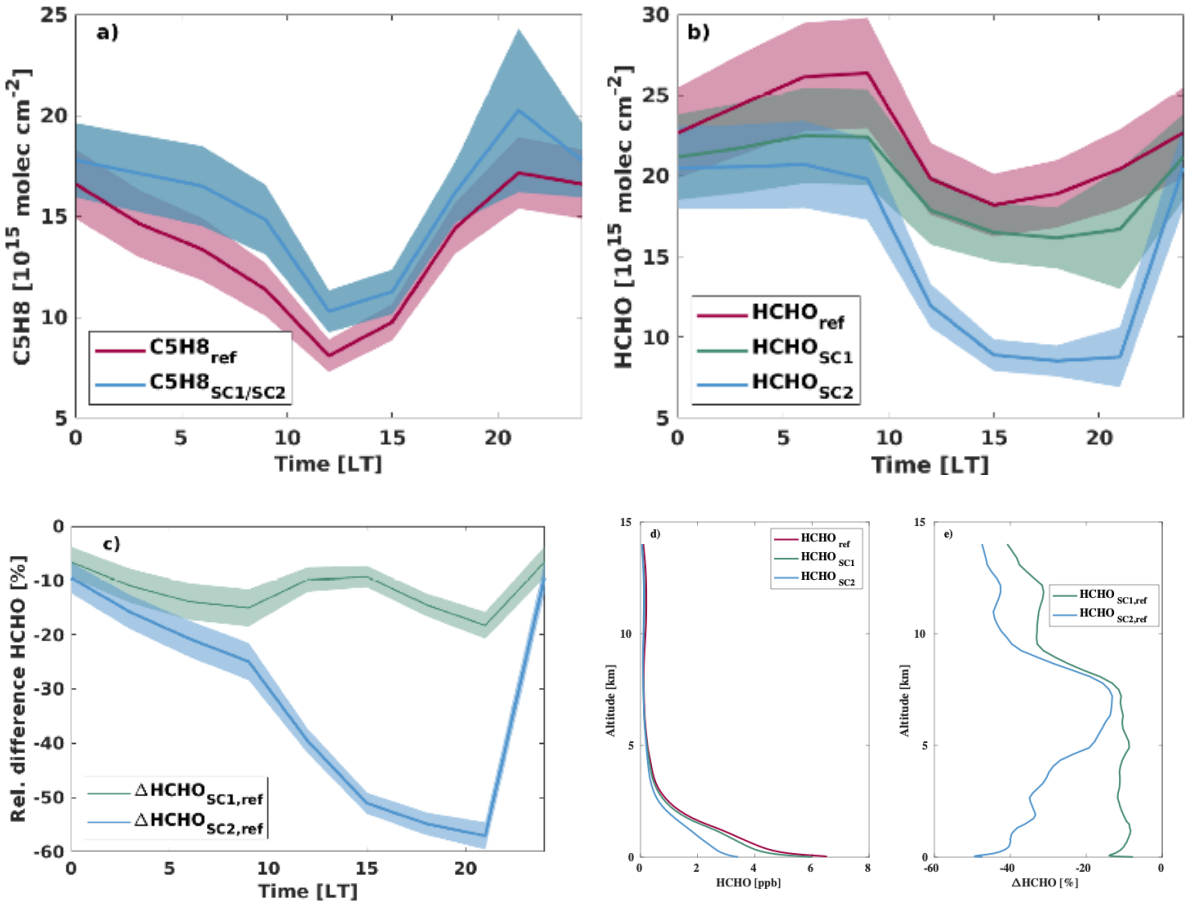


Fig. 3.4: Monthly averaged diurnal cycle of isoprene (panel a) and HCHO VCDs (panel b) above the Amazonian rain forest (-15°N to 5°N, -55°E to -75°E) *when using standard IFS-COMPO chemistry (red), simplified chemistry SC1 (green), and simplified chemistry SC2 (blue). Panel c) shows the relative deviation of simplified HCHO from reference IFS-COMPO*

as a function of daytime. Time is given as local time (UTC=LT+3h). Clearly visible is the increased deviation of $HCHO_{SC2}$ from the reference simulation during daytime. Panels d) and e) show the corresponding vertical profiles for the same region and time period, with panel e) plotting the relative deviation of each profile as $(HCHO_{SC1/2}-HCHO_{Ref})/HCHO_{Ref}$.

The decreased $HCHO_{SC2}$ relative to $HCHO_{SC1}$ and $HCHO_{ref}$ are investigated by examining the diurnal patterns of the different chemistry schemes. Input isoprene is the same for both simple chemistries (fig. 3.4, panel a). However, the modified isoprene oxidation and reaction parameters in SC2 appear to cause a significantly stronger decay in HCHO concentrations during daytime (fig. 3.4, panel b). This leads to an increased bias to $HCHO_{ref}$ up to -57% at 21LT (fig. 3.4, panel c). Contrarily, $HCHO_{SC1}$ and $HCHO_{ref}$ have much less absolute diurnal variation and an approximately constant daily bias, that ranges between -7% (00 LT) and -18% at 21 LT. This strong decay of $HCHO_{SC2}$ causes the globally more pronounced HCHO underestimation of SC2. Since the differences of the chemistry schemes are driven by isoprene oxidation, the formaldehyde concentrations differ the most close to the surface (where most atmospheric isoprene is located) and show better agreement in the free troposphere (fig. 3.4, panels d and e). Interestingly, over the tropics, the difference of the vertical profiles increases again in the upper troposphere. Potentially, this is caused by oxidation of upwards transported isoprene, which has been reported to cause elevated concentrations of oxidised bVOCs in the tropical upper troposphere (e.g. Tripathi et al., 2025).

The analysis does not show a significantly improved agreement of simplified and standard IFS-COMPO formaldehyde for any region or air mass composition when using the extended scheme SC2. In particular the strongly diverging diurnal behaviour of should be further investigated and potentially finetuned. Thie large underestimation of HCHO in SC2 is unexpected, because the total yield of HCHO from ISOP+OH in the two mechanisms is not very different (2 HCHO molecules per isoprene molecule in SC1 and 1.89 to 2.09 HCHO molecules per isoprene molecule in SC2, depending on the NOx level. However, HCHO production is partly direct in SC1, whereas it is entirely indirect in SC2 at the low-NOx limit. Therefore, the most likely cause for the HCHO underestimation with SC2 is the treatment of the isoprene oxidation intermediates. This compound is comparably long-lived (lifetime of days) and its concentration is dependent on the initial conditions used in the assimilation.

At this point, there is no clear indication that an integration of the more complex chemistry scheme into IFS-COMPO would yield improved HCHO assimilation results, such that the current assimilation scheme of HCHO in IFS-COMPO is based on chemistry scheme SC1.

4 Formaldehyde assimilation

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 was developed that makes use of the tangent linear and adjoint of a simplified HCHO chemistry scheme and inversion of non-methane VOCs in the IFS. Previous steps included the build-up of the necessary software infrastructure to enable HCHO assimilation in the CAMS data assimilation system and the testing of its assimilation without applying a tangent linear and adjoint of HCHO or NMVOC inversion (deliverable D2.3). Building upon this work, tangent linear and adjoint codes of the simplified scheme were derived and included into the IFS-COMPO assimilation system in the atmospheric composition section of the model (ifs-source/arpifs/chem/). The detailed tangent linear and adjoint code is included in Appendix B and C.

Sections 4.1 and 4.2 analyse the impact of the assimilation of HCHO satellite retrievals on IFS-COMPO HCHO as well as on its atmospheric precursors and other key atmospheric tracers. The evaluation makes use of different satellite retrievals and ground-based

measurement stations. The individual impact of the assimilated satellite observations is analysed by separately discussing the assimilation of polar-orbiting retrievals (TROPOMI S5P) and the assimilation of geostationary retrievals (TEMPO, GEMS). Section 4.3 focuses on the technical evaluation of the HCHO assimilation system. Key input parameters, such as the assimilation window and averaging kernels, are analysed in detail.

All IFS-COMPO simulations are performed using (1) a reference atmospheric composition simulation based on CAMS cycle 49R1 (assimilation of O_3, CO, NO_2, $SO_{2,volc}$ and AOD, experiment ID *ipie*) and (2) the same configuration with additionally HCHO assimilation (experiment ID *ipmj*). Both simulations use climatological biogenic and anthropogenic emissions from CAMS-GLOB-BIOv3.1 and CAMS-GLOB-ANTv6.1, respectively. The ECMWF super-obbing software is applied and negative satellite retrievals are included in the calculation of the super-observations to avoid a false-positive observation offset. For TROPOMI, the assimilation takes into account the averaging kernels of the retrievals. For TEMPO and GEMS, at the time of writing, HCHO averaging kernels were not provided by the data producers. No bias correction is applied to the NRT HCHO retrievals in the IFS assimilation tests, in contrast to a recent study carried out by Oomen et al., (2024) who applied a bias correction to TROPOMI HCHO.

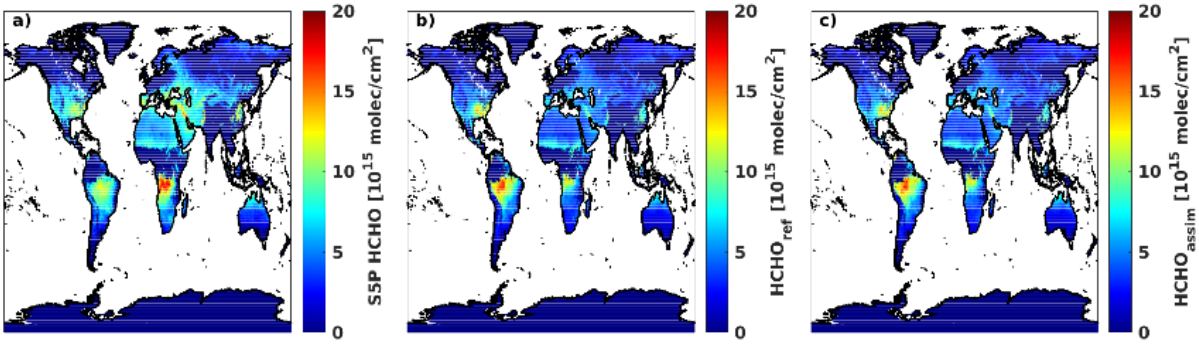


Fig. 4.1: Global formaldehyde VCDs for July 2025 with TROPOMI retrievals (panel a), IFS-COMPO reference simulations (panel b) and IFS-COMPO HCHO assimilation (panel c).

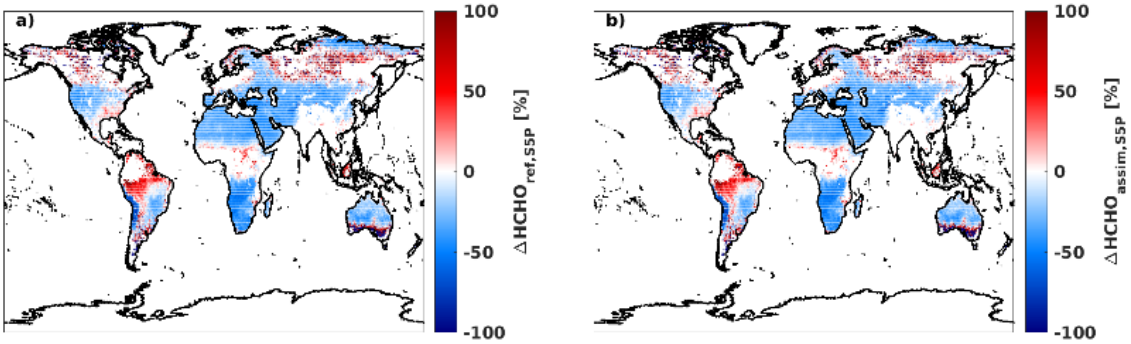


Fig. 4.2: Difference ($HCHO_{IFScompo}$–$HCHO_{S5P}$)/$HCHO_{S5P}$ between TROPOMI and IFS-COMPO formaldehyde VCDs for July 2025 for the reference simulations (panel a) and HCHO assimilation (panel b).

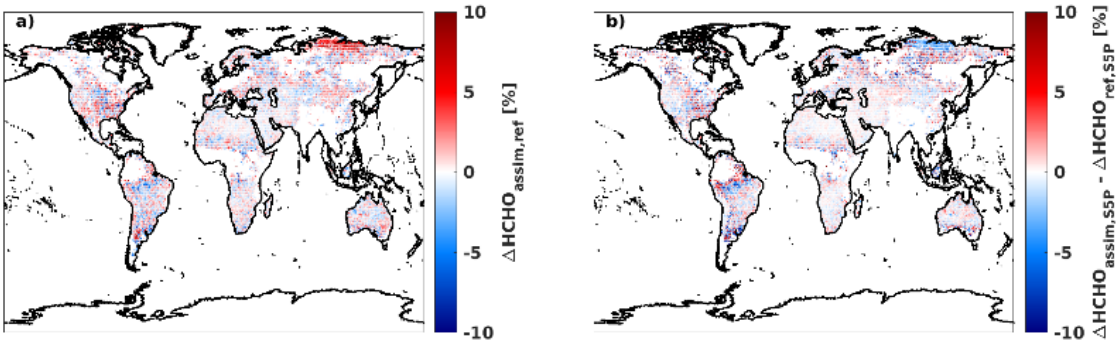


Fig. 4.3: Relative difference ($HCHO_{assim}$–$HCHO_{ref}$)/$HCHO_{ref}$ of IFS-COMPO HCHO due to the assimilation (panel a) and impact of the assimilation on the comparison to TROPOMI retrievals (panel b). For the latter, the absolute difference abs($HCHO_{IFScompo}$–$HCHO_{S5P}$) of each IFS-COMPO configuration is subtracted from each other, such that blue colour indicates an improvement and red indicates a degradation due to the assimilation.

4.1 Impact on IFS-COMPO HCHO

4.1.1 Assimilation of TROPOMI HCHO retrievals

Between May and October 2025, IFS-COMPO formaldehyde has a global monthly median bias to TROPOMI HCHO ranging between +27% and +33% (e.g. positive model bias of 31.18% for July 2025). This divergence is slightly reduced by 0.03% up to 0.5% when assimilating TROPOMI formaldehyde (e.g. a bias reduction from 31.18% to 31.03% in July 2025, figs. 4.1 and 4.2). There is no clear latitudinal pattern of the assimilation increments (fig. 4.3). Globally, the difference between TROPOMI retrievals and IFS-COMPO HCHO is most pronounced over biogenic source regions, where positive deviations of up to 60% are found for July 2025 (fig. 4.4, panel a). The relative change in IFS-COMPO HCHO due to the assimilation of TROPOMI retrievals does not significantly improve this bias. In fact, the latitudinal profile shapes of analysis increments and model-observation bias show no significant correlation. The assimilation improves the analysis over some regions with pronounced bias to the TROPOMI observations (e.g. North Russia around 70°N), but has much less impact in other regions with similar or even larger biases (fig. 4.4). This is particularly true for tropical latitudes between -10°N and 10°N, where the globally largest bias between IFS-COMPO and TROPOMI is located. The analysis of selected global regions yields similar findings, e.g. over continental Europe (fig. 4.5). Here, TROPOMI observes significantly higher HCHO in comparison with IFS-COMPO during July and August 2025, causing a largely increased bias to IFS-COMPO (fig. 4.5, panel a), which is mostly unaffected by the assimilation (fig. 4.5, panel b). The analysis increments show a clear altitudinal dependency with the largest assimilation impact found at the surface, where HCHO concentrations are the highest, the satellite instrument sensitivity the lowest (Appendix E, fig. E6) and the standard deviation of the background error is the largest (fig. 4.5, panels c, f, g). During July 2025, when IFS-COMPO underestimates TROPOMI observations, the assimilation increases surface HCHO by up to 4% in correspondence to the model-observation bias (fig. 4.5, panel c). At the same time, free troposphere HCHO is slightly reduced by -1% and the change in surface HCHO is not large enough to translate into significantly improved HCHO VCDs. Over North America (10°N to 70°N and -160°E to 60°E), the opposite pattern is observed with surface HCHO decreasing by up to 4% due to the assimilation (fig. 4.5, panels d-f). For all global

regions and during the entire analysed period, IFS-COMPO correlates well with the observations and the impact of HCHO assimilation is either negligible or follows the model-observation bias.

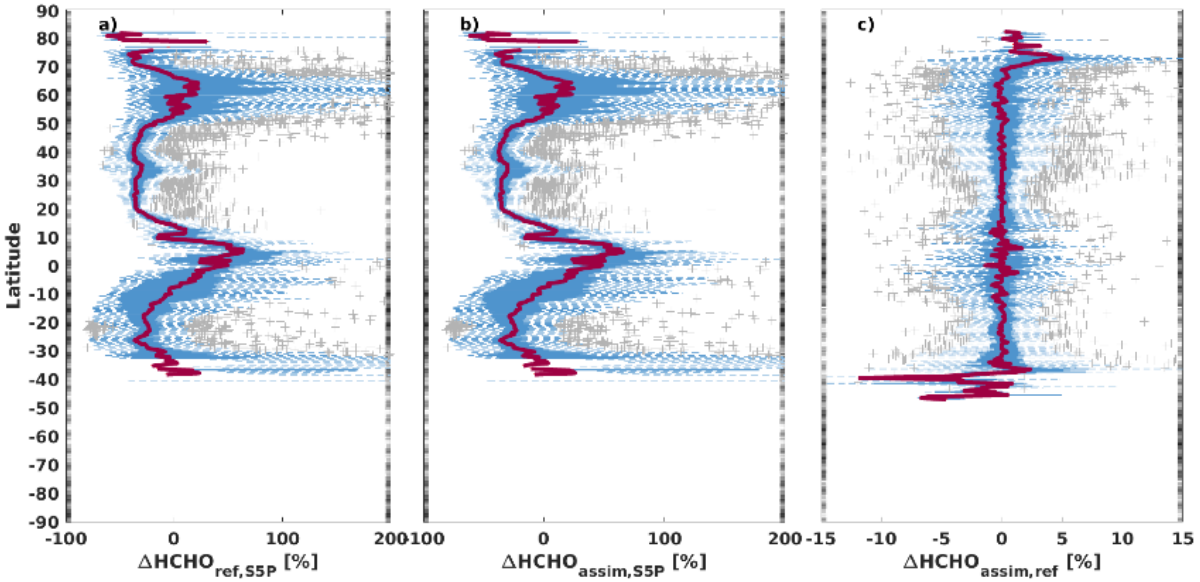


Fig. 4.4: Relative difference ($HCHO_{IFScompo}$-$HCHO_{S5P}$)/$HCHO_{S5P}$ of TROPOMI HCHO and IFS-COMPO HCHO as a function of latitude for July 2025 with reference analysis in panel a), HCHO assimilation in panel b) and the difference ($HCHO_{assim}$-$HCHO_{ref}$)/$HCHO_{ref}$ between reference analysis and HCHO assimilation analysis in panel c). Globally, the IFS-COMPO bias follows isoprene concentration patterns with the largest divergence from the observations in the tropics, where bVOC emissions are the largest.

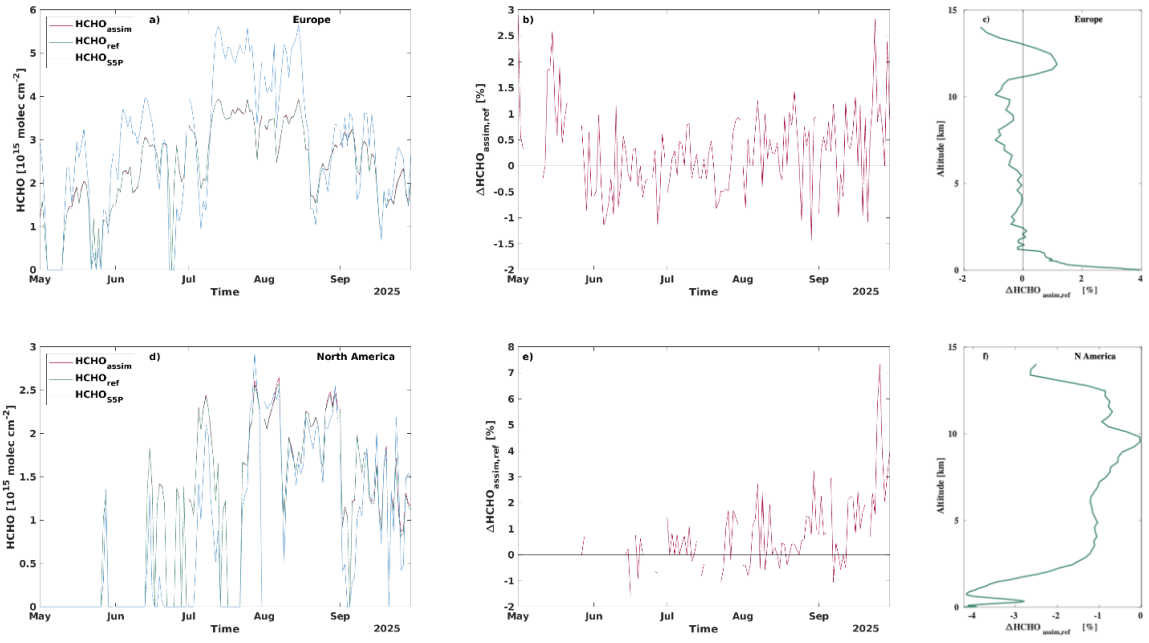


Fig. 4.5: Comparison of HCHO VCDs above continental Europe between 43°N to 53°N and 0°E to 30°E (panels a to c) and North America between 10°N to 70°N and -160°E to 60°E

(panels d-f) from May to October 2025 with observations from TROPOMI S5P (blue), reference simulations without assimilation (green), and assimilated HCHO (red) as well as the relative difference between the IFS-COMPO simulations for the same regions and time period (panels b, e). Panels c and f show the relative change in the vertical profile due to the assimilation (monthly average for July 2025). In accordance to the observation bias of the reference simulations, the assimilation causes an increase in surface HCHO over continental Europe and a decrease of surface HCHO above North America. The simulation above North America also includes TEMPO satellite retrievals. Panel g) shows the vertical profile of the standard deviation of the background error for HCHO as used for the assimilation and the inversion.

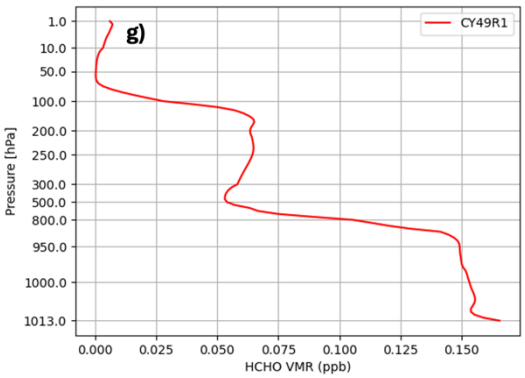


Fig. 4.5 continued.

4.1.2 Assimilation of TEMPO and GEMS HCHO retrievals

Operational geostationary satellite observations of formaldehyde from GEMS and TEMPO have recently become available and have been included into the assimilation for first evaluations. Both instruments provide daily formaldehyde observations at 1h time resolution during daytime. GEMS level 2 data are provided by the Korean National Institute for Environmental Research NIEC (https://nesc.nier.go.kr/en/html/index.do). TEMPO level 2 HCHO is made available by NASA (https://tempo.si.edu/index.html).

Previous evaluations of TROPOMI and TEMPO formaldehyde retrievals reported good agreement of both instruments, with TROPOMI HCHO on average being slightly higher. Comparisons of TEMPO retrievals and ground-based observations from the PANDORA network equally showed overall good agreement (Henderson et al., 2024, *https://tempo.si.edu/presentations/August2024/Oral/D2_1000_Henderson_Validation_final.pdf*). For GEMS HCHO, first evaluations report a negative bias compared to ground-based observations in the order of 30% and higher (Lee et al., 2024). Better agreement was found to TROPOMI retrievals (10-16% positive bias) with some latitudinal dependency and better correlations at the higher latitudes (Northeast Asia).

This section reports on a first evaluation of the assimilation of geostationary formaldehyde retrievals from GEMS and TEMPO in IFS-COMPO based on April and May 2025.

TEMPO

TEMPO formaldehyde over North America shows enhanced VCDs towards the eastern coastline as well as over Mexico and the Caribbean during April 2025. Formaldehyde decreases towards higher latitudes and over the oceans (Fig. 4.6, panel a). This pattern is well captured by IFS-COMPO (Fig. 4.6, panels b and c). However, the comparison of TEMPO and reference IFS-COMPO formaldehyde (no assimilation) generally shows an

underestimation of background formaldehyde by IFS-COMPO both over the continent and adjacent marine regions (Fig. 4.7, panel a). At the same time, over some source regions, an overestimation of formaldehyde is found, in particular over the Yucatan Peninsula. This bias largely persists also when assimilating TEMPO formaldehyde in IFS-COMPO (Fig. 4.7, panel b). However, the assimilation reduces the observed bias both by increasing formaldehyde outside of source regions, where the observations are underestimated in the standard IFS-COMPO configuration, and by decreasing overestimated VCDs, in particular over southern Mexico. As a result, the assimilation decreases the model bias and improves IFS-COMPO formaldehyde over the entire region (Fig. 4.8).

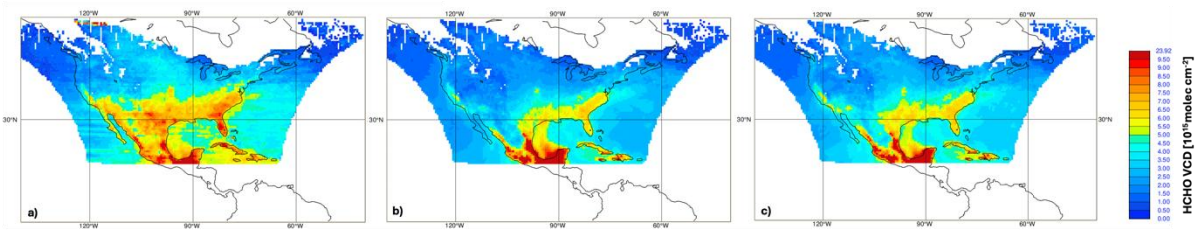


Fig. 4.6: Formaldehyde VCDs observed from TEMPO (panel a) and simulated by IFS-COMPO without assimilation (panel b) and when assimilating TEMPO retrievals (panel c) for April 2025.

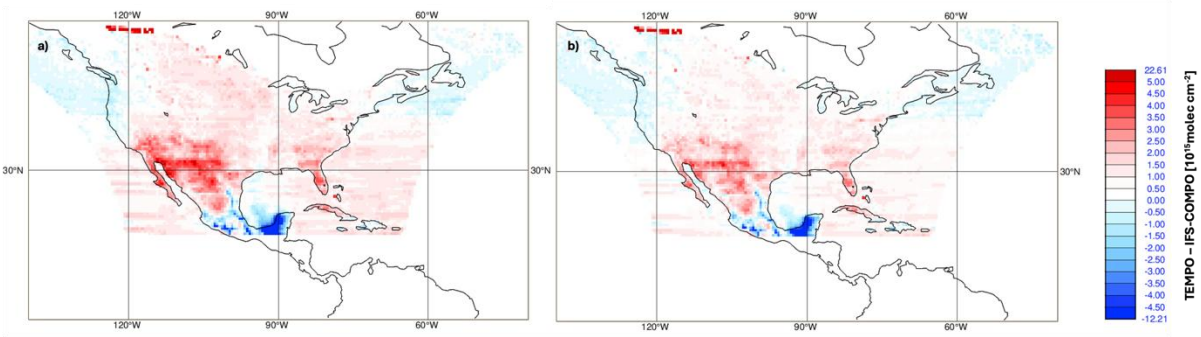


Fig. 4.7: Difference of TEMPO HCHO observations and IFS-COMPO HCHO without HCHO assimilation (panel a) and when assimilating TEMPO observations (panel b) for April 2025. Red indicates larger observed than simulated HCHO, blue shows smaller observed HCHO.

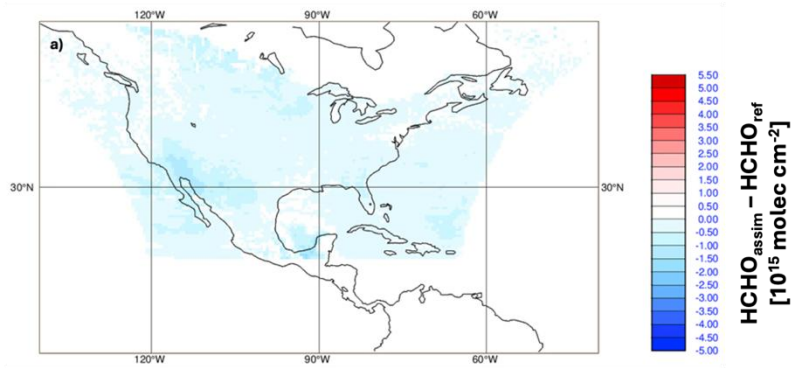


Fig. 4.8: Impact of the assimilation of TEMPO retrievals with red indicating a larger bias of IFS-COMPO HCHO when assimilating the retrievals and blue showing a reduced bias due to the assimilation.

The comparison of the above findings to the assimilation of TROPOMI HCHO retrievals over North America overall shows similar biases of both instruments to the IFS-COMPO simulations (fig. 4.9). Both instruments show generally consistent HCHO variability over the analysed region and their assimilation leads to similar increments over the same areas. Differences are mostly found for higher latitudes above 50°N, where TROPOMI VCDs are generally larger than TEMPO causing an even larger negative offset of IFS-COMPO HCHO (fig. 4.9, panel f).

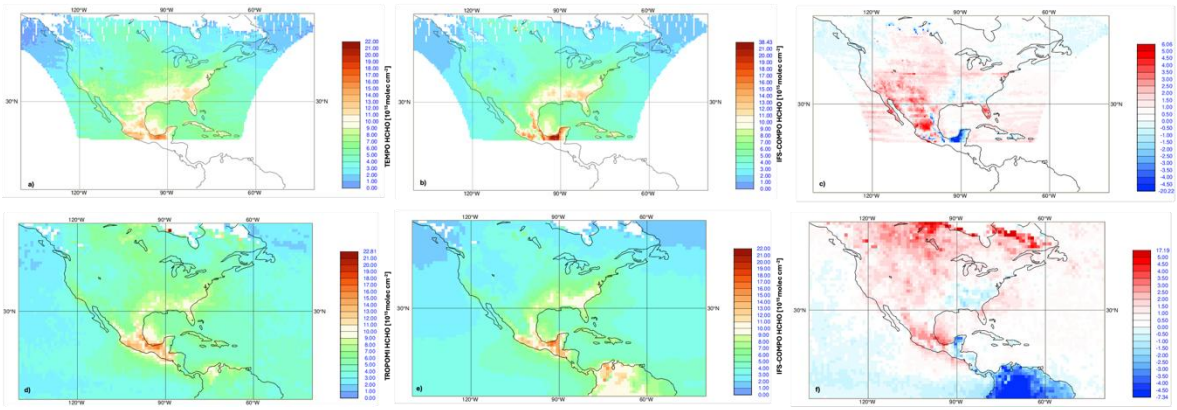


Fig. 4.9: Comparison of the assimilation of TEMPO (panels a-c) and TROPOMI HCHO retrievals (panels d-f) in IFS-COMPO for May 2025 with satellite retrievals (panels a ,d), IFS-COMPO simulations with HCHO assimilation (panels b, e) and the difference of retrievals and IFS-COMPO analysis (panels c, f).

GEMS

The monitoring of GEMS formaldehyde retrievals reveals some major discrepancies between GEMS and TROPOMI over the analysed area and time period (May 2025). In particular towards the borders of the field of view of GEMS, there are increasing differences in the retrievals of the satellite instruments (e.g. over India, fig. 4.10, panels a and c). Additionally, apparent retrieval artefacts of GEMS formaldehyde around the equator cause a regionally confined, positive offset and are currently not filtered out by IFS-COMPO quality control. These findings show that assimilation tests with GEMS HCHO data are currently not meaningful. Interestingly, the observation artefacts along the equator are not assimilated. These findings will be re-assessed after future retrieval updates.

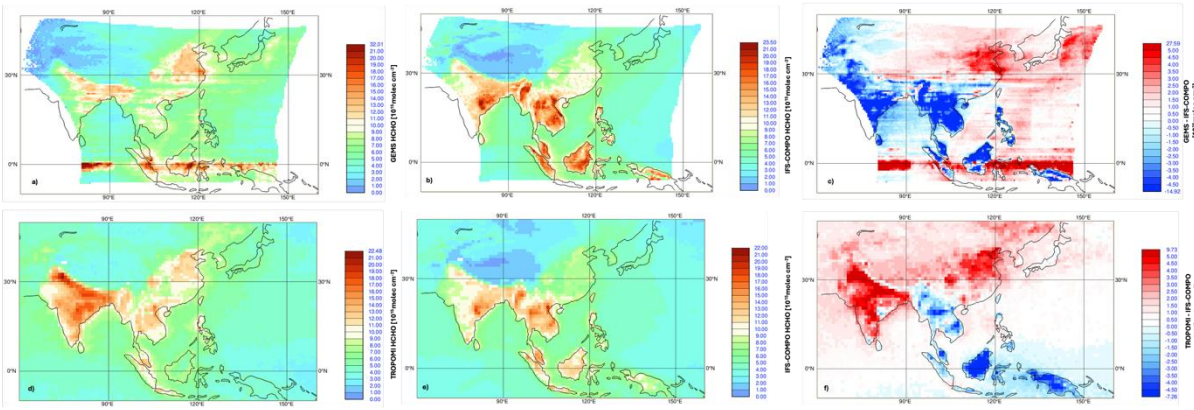


Fig. 4.10: *Comparison of the assimilation of GEMS (panels a-c) and TROPOMI HCHO retrievals (panels d-f) in IFS-COMPO for May 2025 with satellite retrievals (panels a, d), IFS-*

COMPO simulations with HCHO assimilation (panels b, e) and the difference of retrievals and IFS-COMPO analysis (panels c, f).

4.1.3 Evaluation based on independent observations

This section makes use of independent observations to analyse the impact of combined assimilation of all currently included formaldehyde satellite retrievals (polar-orbiting: TROPOMI S5P; geostationary: TEMPO and GEMS). IFS-COMPO HCHO (simulation details at the beginning of section 4) is evaluated relative to 10 ground-based NDACC FTIR HCHO measurement stations in different global regions. For the comparison, the IFS-COMPO simulations are gridded relative to the ground-based stations and their elevation above sea level and smoothed according to instrument sensitivity and averaging kernels of the observations. The statistical analysis is provided by Bavo Langerock (Royal Belgian Institute for Space Aeronomy (BIRA), CAMS-82-bis) with data support by the CAMS2-27 project. The impact of atmospheric seasonality on HCHO concentrations is taken into account by performing the evaluation for a total time span of 6 months (May – September 2025). Subject to data availability, the chosen NDACC FTIR stations aim to cover the largest possible range of different HCHO VCDs (fig. 4.11). The precise locations of the measurement sites are:

1. Remote, pristine regions: Ny-Alesund (1, Svalbard), Thule (2, Greenland), Sodankylä (3, Finland), Jungfraujoch (4, Swiss)
2. Remote with biogenic impact: Maido (5, La Reunion), Lauder (6, New Zealand)
3. Biogenic with anthropogenic impact: Wollongong (7, Australia), Toronto (8, Canada), Boulder (9, USA)
4. Anthropogenic regions: Xianghe (10, China)

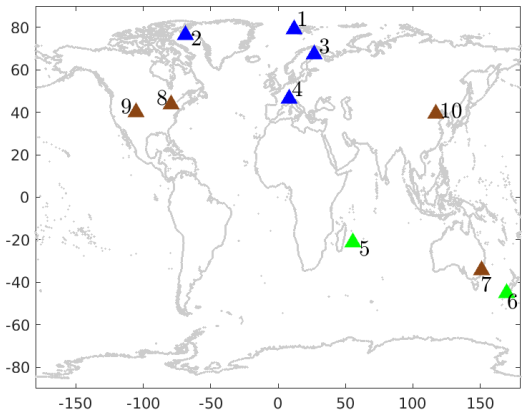


Fig. 4.11: Geographic locations of the evaluated NDACC FTIR stations with the characteristic atmospheric conditions indicated by the colour-coding (clean air masses in blue, biogenic in green, and anthropogenic in brown).

More details on the employed FTIR instruments, their precise locations and measurement methods can be found in the description of the Network for the Detection of Atmospheric Composition Change (NDACC, see https://ndacc.larc.nasa.gov/instruments/ftir-spectrometer).

It is noteworthy that the chosen time span covers spring to autumn (summer) for the northern hemispheric measurement stations, but autumn to spring (winter) for the observations in the southern hemisphere. As discussed later in detail, this seasonality reflects in the respective time series of HCHO VCDs. As one consequence, HCHO VCDs from the southern stations (southern hemisphere, local winter) are not significantly larger than the observations from the pristine stations (northern hemisphere, local summer). Due to local wintertime (and in case of La Reunion also the strong spatial emission confinement (small island) and high elevation of the measurement site of 2160m above sea level, i.e. mountainous conditions instead of tropical rain forest), neither the measurements on La Reunion nor in Lauder are optimal to evaluate the assimilation results in air masses affected by strong biogenic emissions. The analysis rather shows pristine atmospheric conditions over the two stations for the evaluated time period, with methane photochemistry inducing a slight seasonal cycle for the Maido site.

Over the US, the assimilation makes use of TROPOMI and TEMPO observations (FTIR stations Boulder, Toronto). Over China, retrievals from TROPOMI and GEMS are employed (FTIR station Xianghe). For the remaining sites, the assimilation relies on TROPOMI observations only.

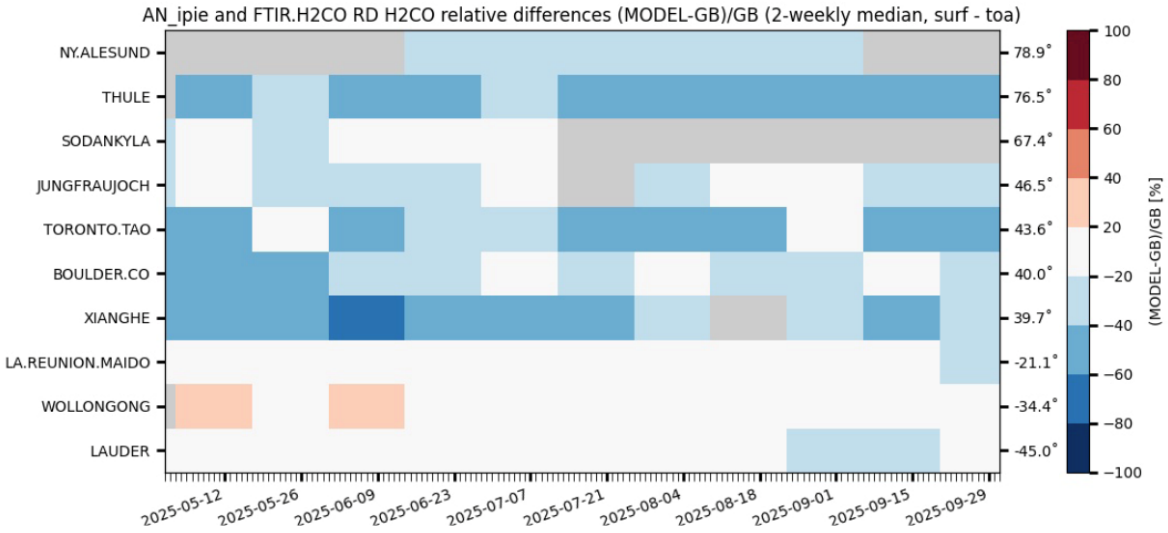


Fig. 4.12: Relative differences of HCHO VCDs from IFS-COMPO and FTIR observations for all analysed sites covering May to September 2025. IFS-COMPO simulations are performed using a standard atmospheric composition setup from CY49R1 (no HCHO assimilation). The sites are sorted by decreasing latitude, i.e. from North to South. Periods of missing observations are plotted in grey. Plot courtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

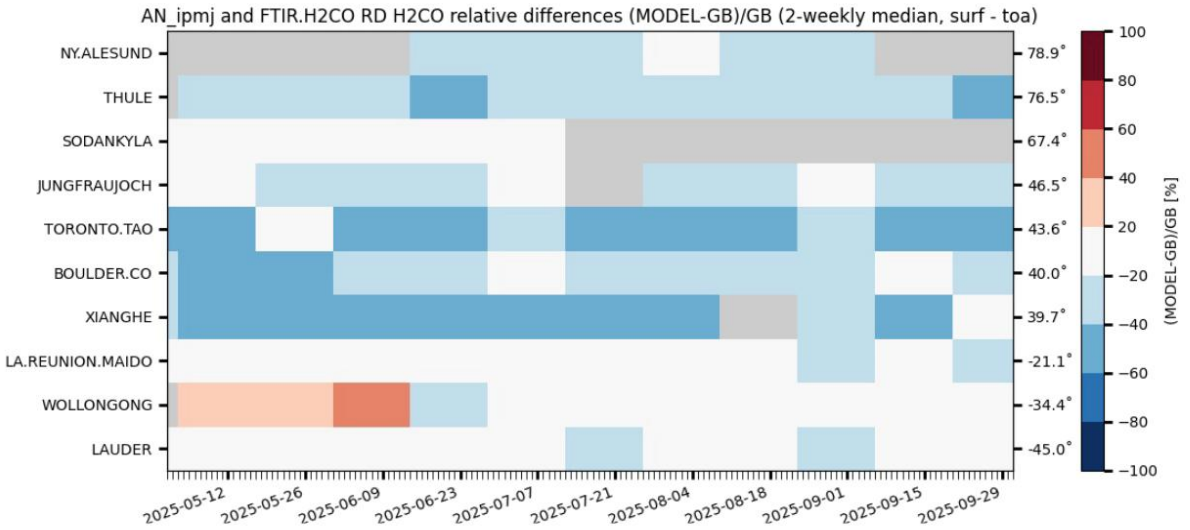


Fig. 4.13: Same as above but with active assimilation of HCHO retrievals. Except for two periods over Wollongong, the assimilation has neutral to positive impact on the comparison. Plot courtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

Globally and for both IFS-COMPO configurations, simulated HCHO shows a negative bias compared to the ground-based observations (figs. 4.12 - 4.15). Among all analysed sites, Wollongong possibly has the most complex air masses with coastal, background, biogenic, and biomass burning impact. This site appears to be the only location where HCHO is overestimated and not underestimated (fig. 4.14, panel b). Across all stations and for the entire analysed time span, the Spearman correlation coefficients of observations and simulations range from 0.58 (Toronto) to 0.89 (Thule). The normalized mean bias of IFS-COMPO seems larger in the northern hemisphere (figs. 4.12 and 4.13, upper 7 rows) than in the southern hemisphere (La Reunion, Wollongong, and Lauder). Presumably, this pattern is not reflecting a latitudinal dependency, but rather a consequence of the atmospheric composition over the different measurement stations. In the northern hemisphere, more stations may be affected by elevated NOx levels. In such regimes, the relative model bias appears to be larger than over pristine or biogenic HCHO source regions as they are more likely to be found over the three southern hemisphere stations. It is noteworthy, that the model-observations bias appears constant in time for the analysed period (figs. 4.12 and 4.13), which indicates that IFS-COMPO well captures the seasonal HCHO patterns.

The assimilation of HCHO retrievals does not significantly alter above main findings. Globally and for the entire period, the assimilation reduces the median normalized model bias by 1.1% from -24.4% to -23.3%. At the same time, the assimilation decreases the Spearman correlation coefficient by 0.02 from 0.77 to 0.75 (all sites average). As can be deduced from table 4, this decreased correlation is mainly a result of the evaluation of the Thule measurement site and -to a lesser extent- the other FTIR stations in clean pristine regions. For all these, the assimilation improves the relative model-bias, but at the same time decreases the Spearman coefficient. For seven out of the ten stations, the median normalized mean bias (MNMB) is decreased by the assimilation. For the remaining three (Xianghe, Toronto, Wollongong), the assimilation slightly increases the MNMB. All these stations are located within source regions and characterised by biogenic emissions with slightly elevated NOx levels.

Table 4: Statistics of the IFS-COMPO-FTIR evaluation for all analysed stations and averaged over May 1st to September 30th, 2025. IFS-COMPO simulations were gridded onto the ground grid of the measurement stations and smoothed according to instrument sensitivity and

averaging kernels. Statistical analysis courtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

Site and elevation above sea level (msl)	FTIR	IFS-COMPO, assimilation			IFS-COMPO, reference			Assimilation impact	
	median [10^{15}molec cm^2]	median [10^{15}molec cm^2]	correlation model, FTIR	NMB (Model-FTIR)/FTIR [%]	median [10^{15}molec cm^2]	correlation model, FTIR	NMB (Model-FTIR)/FTIR [%]	Delta correlation	Delta median NMB [%]
Ny-Alesund (15msl)	2.79	1.72	0.83	-35.13	1.64	0.85	-39.12	-0.02	+3.99
Thule (30msl)	2.13	1.35	0.70	-34.44	1.15	0.89	-44.45	-0.17	+10.01
Sodankylä (179msl)	2.80	2.38	0.77	-15.67	2.32	0.79	-18.68	-0.02	+3.01
Jungfraujoch (3580msl)	1.68	1.40	0.76	-20.58	1.38	0.76	-21.46	0	+0.88
Toronto (174msl)	12.50	6.89	0.61	-44.46	6.74	0.58	-44.01	+0.03	-0.45
Boulder (1634msl)	9.20	7.22	0.81	-25.55	7.04	0.79	-27.35	+0.82	+1.80
Xianghe (35msl)	15.15	8.75	0.74	-49.10	8.34	0.78	-48.69	-0.04	-0.41
La Reunion, Maido (2155msl)	1.70	1.53	0.73	-8.31	1.53	0.76	-8.95	-0.03	+0.64
Wollongong (30msl)	3.34	3.67	0.86	10.17	3.61	0.86	6.93	0	-3.24
Lauder (370msl)	1.48	1.36	0.85	-8.79	1.32	0.69	-10.88	+0.16	+2.09
median	2.79	2.05	0.75	-23.06	1.98	0.77	-24.41	-0.02	1.35

Summarized comparison statistics can be found in table 4. Timeseries of the simulations and observations for each measurement site are included in Appendix E. Overall, the evaluation shows a neutral to slightly positive impact of assimilating HCHO in IFS-COMPO. The relatively smallest impact is found for the more polluted sites, the largest improvements are observed over the clean, pristine sites. The relatively small impact of the assimilation in air with higher NOx levels may be a result of the HCHO chemistry scheme, that the tangent linear and adjoint are based on. As laid out in detail in section 3 and CAMEO deliverable D2.3, the currently employed scheme was optimised for high VOC, low NOx conditions. Isoprene was identified as predominant HCHO precursor and the isoprene oxidation pathway assumed to occur primarily via ISOPO2+HO2 reaction. The latter is not representative for anthropogenic conditions with high NOX concentrations. This may at least in part explain the relatively smaller impact of the assimilation in anthropogenic air masses. The currently employed chemistry scheme is better suited to improve IFS-COMPO HCHO in remote terrestrial air masses characterised by biogenic emissions.

It is noteworthy, that despite these chemistry-related limitations and the generally low impact, the assimilation does improve the comparison for three of the four more polluted sites (however at minor extent) and in fact shows relatively larger impact on IFS-COMPO HCHO over the clean, pristine NH sites, than over the two remote measurement stations in the southern hemisphere.

To better evaluate the assimilation in biogenic air masses, seasons and regions with spatially extended and high isoprene emissions would be ideal. The time span for the analysis (May-September 2025) was chosen with respect to satellite retrieval availability, which in the case

of TEMPO only started in April 2025. This month was used as spin-off for the analysis experiments and the data evaluation was started in May 2025. The evaluation of the assimilation will be continued over the coming months (end of 2025, beginning of 2026) and, where possible, expanded to measurement sites affected by higher biogenic emissions. Further, a more generalised chemistry scheme, that accounts for varying NOx concentrations, was recently developed (sect. 3). Subject to further finetuning of the new scheme (in particular with respect to its capability to simulate daytime HCHO concentrations), its impact on the assimilation, in particular in anthropogenic air masses, is planned to be tested in the coming months by developing a respective tangent linear and adjoint and integrating the adapted scheme into the assimilation framework.

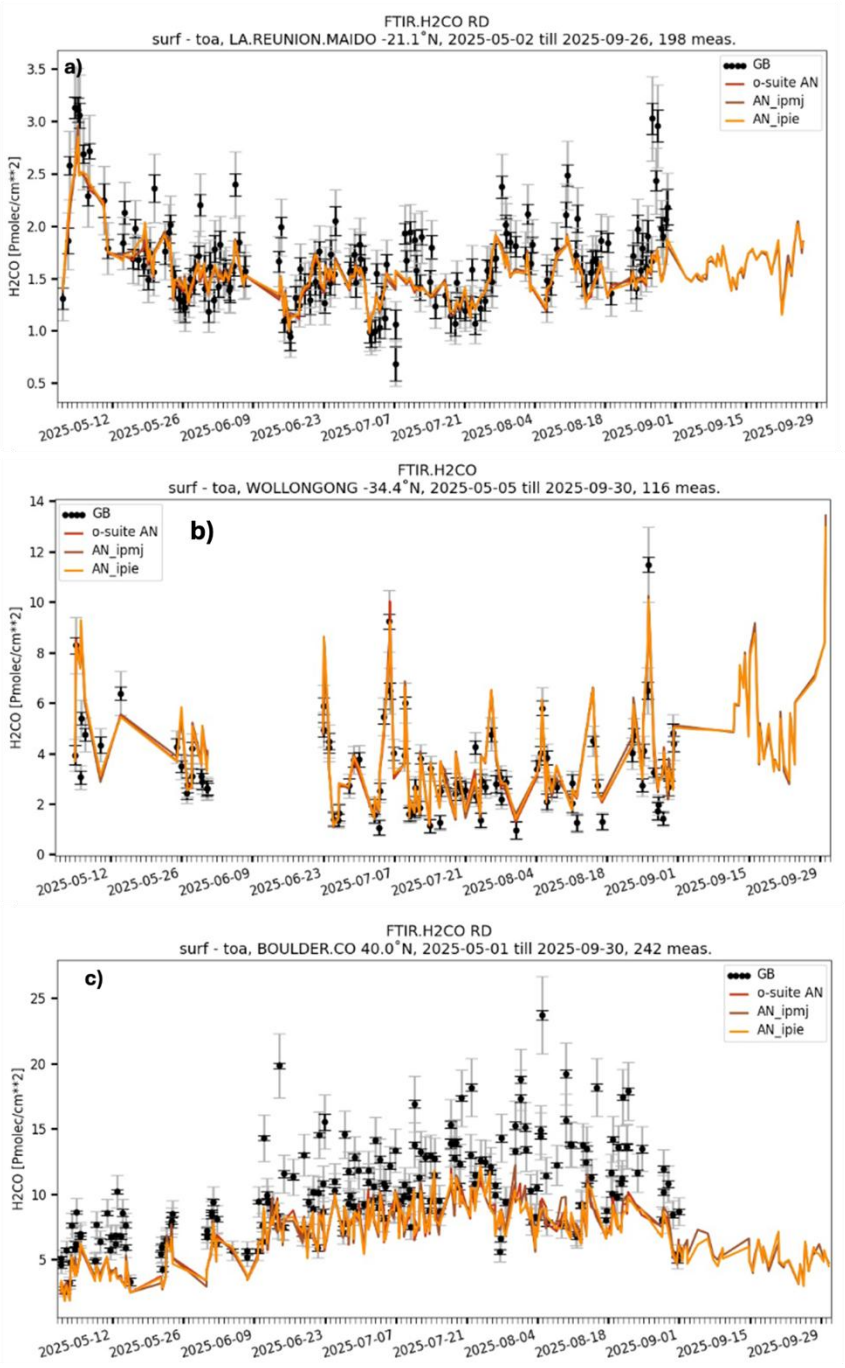


Fig. 4.14: Timeseries of HCHO VCDs over Maido, Wollongong, and Boulder with observations (black), IFS-COMPO reference run (yellow), and HCHO assimilation (brown). The data are

averaged in two-week intervals ranging from May 2025 to September 2025. Additionally plotted is the current CAMS o-suite, that mainly corresponds to the reference model setup.

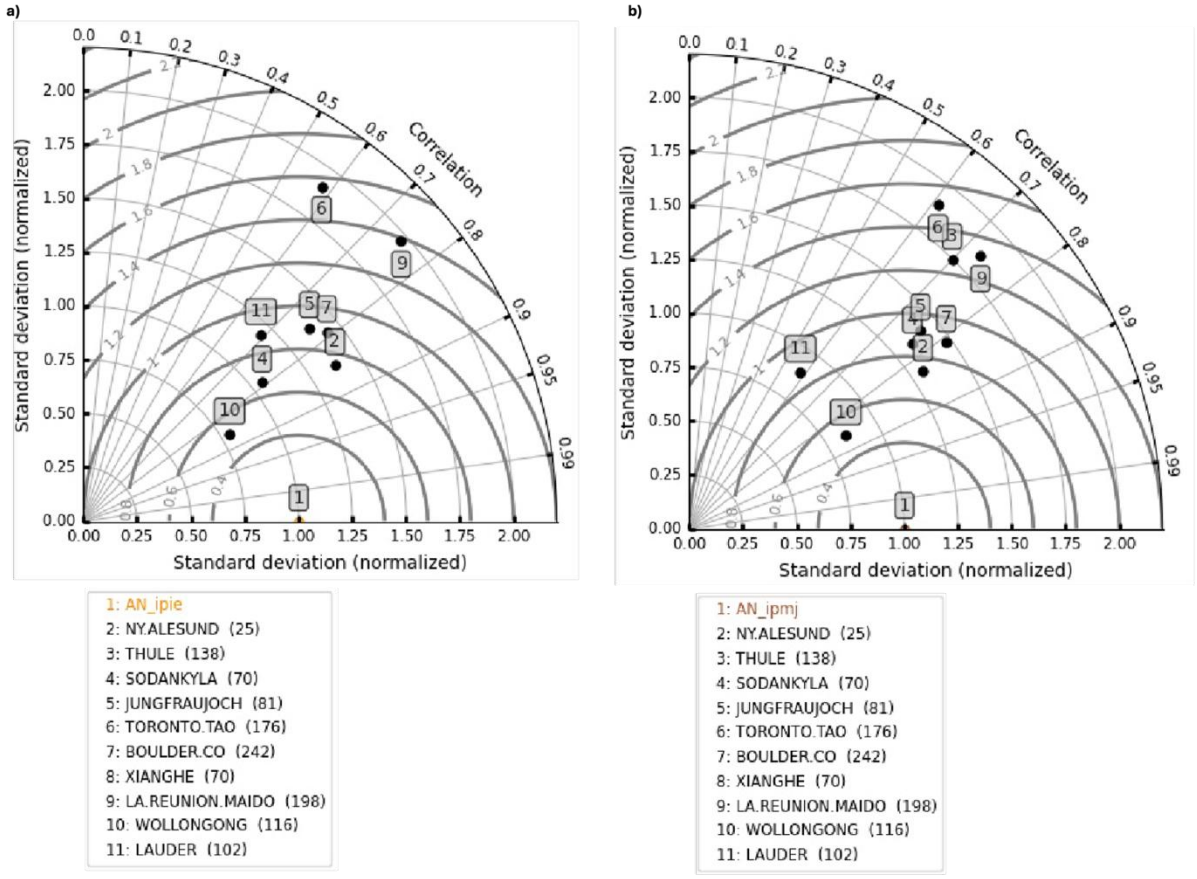


Fig. 4.15: Taylor diagrams for the FTIR-IFS-COMPO evaluation for the IFS-COMPO reference setup (panel a, experiment ipie), and HCHO assimilation (panel b, experiment ipmj). The data are averaged over the entire time span from May 1st to September 30th, 2025. IFS-COMPO simulations were re-gridded onto the ground grid of the measurement stations and smoothed according to instrument sensitivity and averaging kernels. Plot curtesy by Bavo Langerock (BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

4.2 Impact on key atmospheric tracers

This section analyses the impact of formaldehyde assimilation on tropospheric isoprene, carbon monoxide and ozone. In addition to the model-internal comparison of the simulations with and without HCHO assimilation, the results are also validated relative to observations. For ozone and CO, ground-based measurement stations from the NDACC FTIR network are used. The ozone evaluation additionally includes tropospheric ozone sonde measurements to enable an altitudinally resolved validation. For isoprene, satellite observations from the CrIS satellite instrument are used. These isoprene retrievals are available as multi-annual monthly means from the years 2013-2020 (Wells at al., 2020). The validation is based on the same IFS-COMPO experiments as above (see beginning of section 4 for the simulation details).

4.2.1 Isoprene

The assimilation of formaldehyde potentially impacts atmospheric isoprene concentrations (1) through the chemical link of the two species introduced by the adjoint and tangent linear simplified HCHO-isoprene chemistry scheme (sect. 3) and (2) indirectly e.g. via changed OH levels, that impact the abundance and lifetimes of related bVOC species and hence of isoprene. However, due to the overall small impact of the formaldehyde assimilation scheme in its current version, changes in atmospheric isoprene, if any, are expected to be minimal. In fact, the comparison of IFS-COMPO isoprene VCDs with and without HCHO assimilation shows only minor changes due to the assimilation. Terrestrial median isoprene VCDs are decreased by 0.6% from 4.80x10^{13}molec cm^{-2} to 4.78 x10^{13}molec cm^{-2} (six-month average from May to October 2025). The predominant impact is found in the tropics, where isoprene concentrations are the largest. Here, the assimilation causes a median reduction in isoprene in the order of 1.2±3.8% (Amazonia between -10°N and 5°N and -55°E and -75°E). However, even in these latitudes, the difference between reference isoprene and isoprene with HCHO assimilation is not large enough to systematically modify isoprene concentrations or to significantly affect the comparison to the CrIS isoprene observations (fig. 4.16). When compared to the observations, IFS-COMPO appears to overestimate isoprene over source regions (e.g. tropics) and to underestimate isoprene elsewhere (fig. 4.16, panel c). When interpreting the absolute deviation of simulations and retrievals, it should be taken into account that the multi-annual mean of the retrievals may cause deviations in the data sets that are unrelated to the assimilation. However, the source overestimation and background underestimation of IFS-COMPO isoprene was found for the entire period (May to October), such that the results rather point to a systematic than coincidental pattern. In fact, similar results were also obtained for IFS-COMPO formaldehyde (e.g. when comparing IFS-COMPO formaldehyde and TEMPO satellite retrievals over North America, sect. 4.1.2). In either case, HCHO assimilation does not appear to significantly reduce the bVOC overestimation in tropical latitudes by IFS-COMPO.

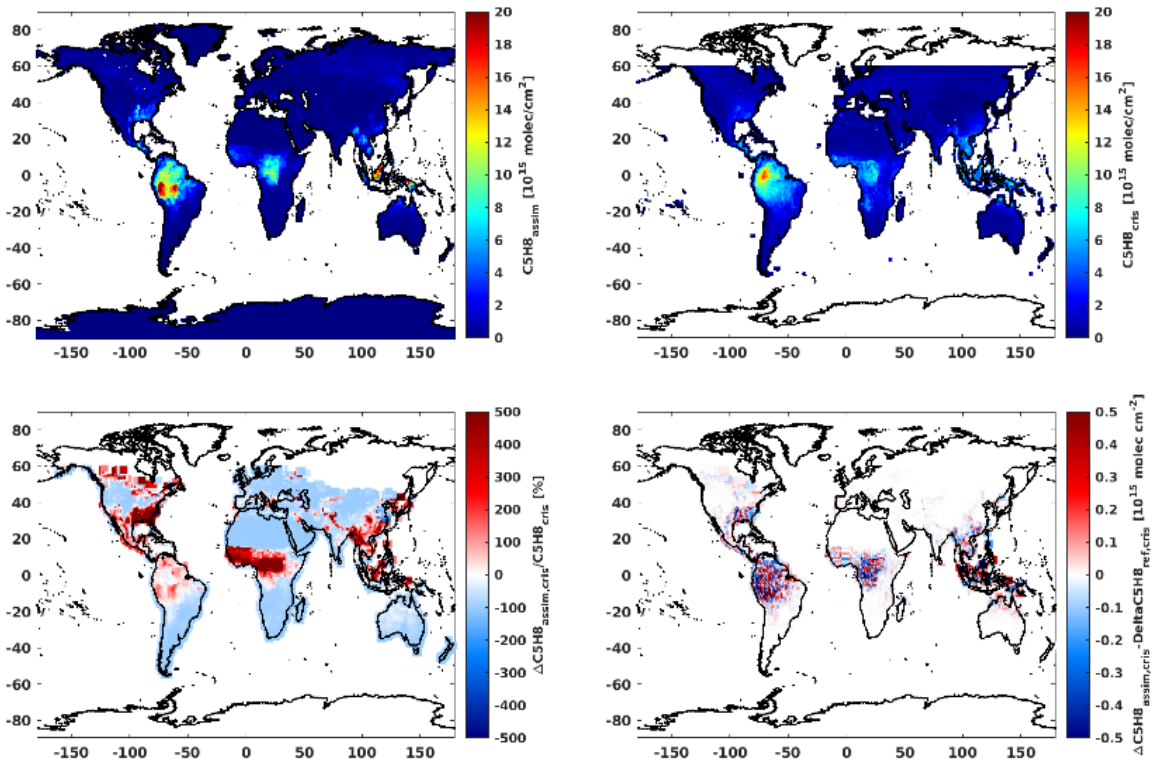


Fig. 4.16: Global isoprene VCDs when assimilation HCHO in IFS-COMPO (panel a) and retrieved from the CrIS instrument (panel b) as well as their relative difference (panel c). Panel

d) shows the impact of the HCHO assimilation on the comparison with the colour-code corresponding to the difference of the absolute deviation of each IFS-COMPO configuration from the observations. As such, red colour indicates a degradation and blue indicates an improved comparison due to the assimilation. All IFS-COMPO simulations are a monthly mean for July 2025. The CrIS retrievals are a multi-year average for July (2013-2020, Wells et al., 2020). Simulations above marine surfaces are masked out for better visibility.

4.2.2 Carbon monoxide

The oxidation of non-methane volatile organic compounds (NMVOCs) is a significant atmospheric source of carbon monoxide with estimated yields of 870Tg CO per year (Muller et al., 2005). For formaldehyde, the link to CO mainly originates from the oxidation of methane and major bVOCs, most notably isoprene. As a result, the assimilation of formaldehyde satellite retrievals in IFS-COMPO may alter CO both via its impact on OH levels and via modification of isoprene concentrations, which impacts the subsequent isoprene-to-CO oxidation chain. Since CO is one of the final products of bVOC oxidation, the small impact of HCHO assimilation on the individual bVOC species may potentially accumulate and therefore cause relatively larger changes in CO than found for e.g. isoprene. Since all bVOC-related CO production processes are mostly occurring within the lower atmosphere, where VOC concentrations are the largest, the present analysis focuses on tropospheric CO only and compares IFS-COMPO CO VCDs with and without HCHO assimilation to respective observations from NDACC FTIR stations.

The relative difference to the FTIR observations is less than 10% both when assimilating HCHO and without assimilation. CO (and NO_2, O_3, AOD, and $SO_{2,volc}$) is assimilated in both experiments using IASI and TROPOMI observations. The correlation coefficients to the observations range from 0.71 to 0.98. Relative to the measurements, IFS-COMPO CO with HCHO assimilation has a median positive offset of 4.2% (4.0% without HCHO assimilation) with an average tropospheric CO VCD of 1480×10^{15} molec cm^{-2} (1477×10^{15} molec cm^{-2} without HCHO assimilation). Collocated observations from the same time period are slightly lower with a median VCD of 1423×10^{15} molec cm^{-2}. Across all analysed sites, the assimilation of formaldehyde leads to minor changes in tropospheric CO of less than 0.5% by slightly increasing tropospheric VCDs. The correlation to the observations is either unaffected by the assimilation or slightly degraded (over Xianghe and Ny-Alesund). The normalized relative difference to the observations decreases for some sites (Ny-Alesund, Toronto, La Reunion, Lauder), but increases for the others (Thule, Bremen, Jungfraujoch, Boulder, Xianghe, Wollongong). Over Ny-Alesund, HCHO assimilation causes the largest improvement in tropospheric CO. Over this site, the relative difference to CO observations is reduced by 0.5% from 1.2% to 0.7% (however with a much larger standard deviation of 8.6%).

For all sites, the change in tropospheric CO is well within the standard deviation. For no region, significant improvements or degradations in tropospheric CO are observed. One example time series over Wollongong (Australia) is shown in fig. 4.17.

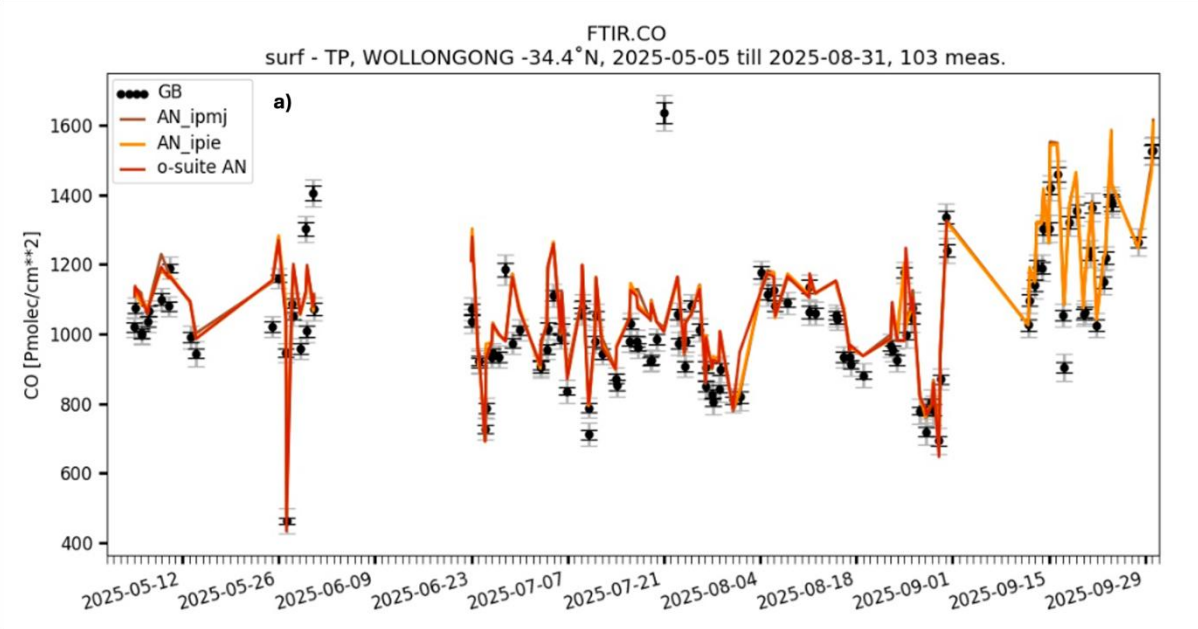


Fig.4.17: Time series of tropospheric CO over Wollongong (Australia) from May 2025 to September 2025 with observations in black and IFS-COMPO simulations in yellow (reference run, no HCHO assimilation) and red (HCHO assimilation). Additionally plotted is the operational IFS-COMPO analysis (CAMS cycle 49R1), which corresponds mostly to the setup of the reference experiment ipie. Plot courtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

4.2.3 Ozone

Assimilation of formaldehyde satellite retrievals may potentially impact ozone levels via VOC oxidation, modified OH and CO levels, and trace gas lifetimes. Again, the largest impact is expected to be found within the lowermost atmosphere, where VOC concentrations are the largest. As found for isoprene and CO, the evaluation shows no significant modification of tropospheric ozone when assimilating formaldehyde for any global region or altitude. The validation of ozone sonde profiles across all latitudes reveals no significant changes of ozone, in particular in the tropics where the largest impact is expected due to locally high VOC concentrations (fig. 4.18).

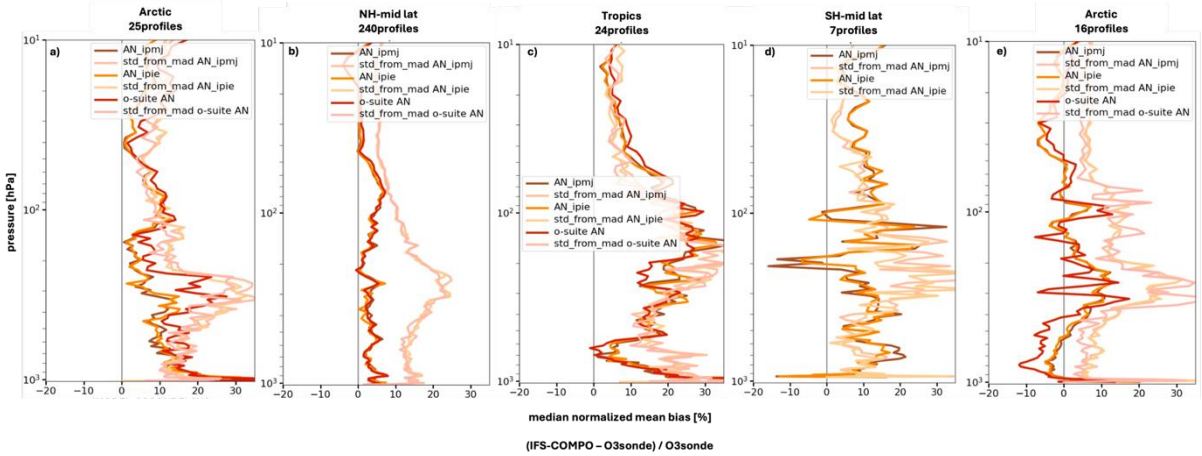


Fig. 4.18: Median bias of IFS-COMPO tropospheric ozone relative to global ozone sonde profiles for May 2025 to September 2025 showing simulations with HCHO assimilation

(experiment ipmj, brown), without HCHO assimilation (experiment ipie, yellow) and the operational IFS-COMPO analysis from CAMS cycle 49R1 (red). Plot curtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project.

4.3 Assimilation parameters and satellite instruments

This section evaluates technical details of the HCHO assimilation in IFS-COMPO. The analysis is again based on IFS-COMPO experiments ipmj (HCHO assimilation) and ipie (no HCHO assimilation) as described at the beginning of section 4.

4.3.1 Observation quality control

After application of standard IFS-COMPO observational filtering, data thinning, variational quality control and background-error checks, the assimilated TROPOMI HCHO observations are spatially homogeneously distributed (quality flag 0). There are no systematic longitudinal inhomogeneities that may potentially cause varying sensitivities of the assimilation system to different longitudinal/latitudinal regions (fig. 4.19). The slight decreasing number of observations towards the poles is a direct consequence of the decreasing instrument sensitivity at higher latitudes.

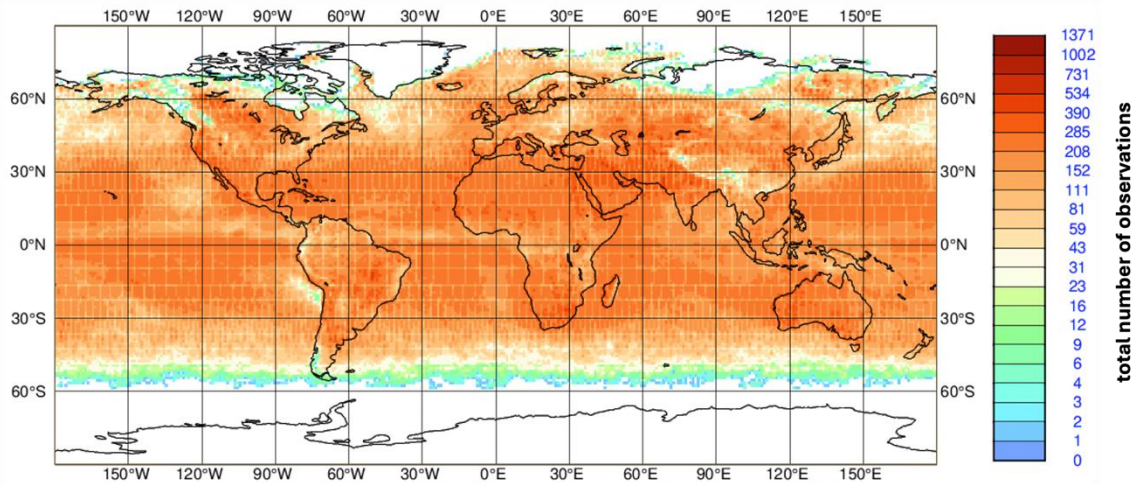


Fig.4.19: Global overview on the total number of assimilated TROPOMI HCHO observations for May 2025.

4.3.2 TROPOMI averaging kernels

The impact of TROPOMI S5P HCHO averaging kernels on the assimilation was tested by performing IFS-COMPO assimilation experiments that exclude or include the averaging kernels for one example month (April 2025). Based on the simulations, a clearly improved agreement of IFS-COMPO analysis formaldehyde and TROPOMI observations was found when taking the averaging kernels into account. Globally, the use of the averaging kernels reduces assimilated formaldehyde. The largest impact is found in tropical latitudes (fig. 4.20, panel c). Consequently, the assimilation improves in this region when using averaging kernels (Fig. 4.20, panel d). The same applies to background formaldehyde, e.g. over marine areas. For no region, a degradation of the assimilation was found when applying the averaging kernels.

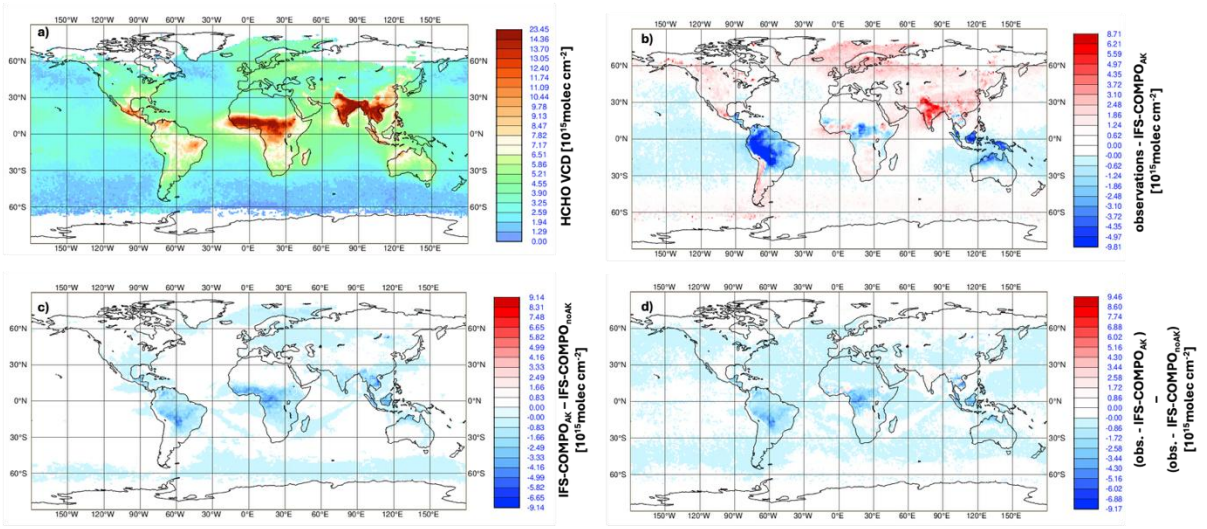


Fig. 4.20: Global overview on the impact of TROPOMI averaging kernels on the assimilation of formaldehyde in IFS-COMPO for April 2025. The assimilated TROPOMI HCHO observations are shown in panel a. Panel b) plots the absolute difference of the observations and simulated formaldehyde using the standard set up that includes the TROPOMI averaging kernels. Panel c) plots the difference of simulated formaldehyde when including and excluding averaging kernels and panel d) shows the resulting increased/decreased difference to the observations. For the latter, red colour indicates regions where the use of averaging kernels degrades the comparison and blue all areas where the assimilation with averaging kernels improves the agreement to the observations.

4.3.3 Assimilation window

Due to their differing observational setups, geostationary and polar-orbiting satellites provide observations at different spatial and temporal grids. Geostationary instruments stay at a fixed position relative to the earth and hence yield timely continuous coverage of a fixed global area. Satellites in low earth polar orbits provide global observations twice daily. Each location is observed at about the same local time every day (13:30LT for TROPOMI S5P). This introduces a spatial time-dependence of the observations, that may potentially impact their assimilation in IFS-COMPO. Generally, observations that were obtained towards the end of the assimilation window have larger impact on the analysis than those from the beginning of the 4DVar-window, as they are the most up-to-date observations (McNally, 2019). This implies, that the impact of each observation on the assimilation also depends on the start time of the assimilation window, with benefits potentially offset by growing model error during the assimilation window. Such time-dependency may have larger impact on the assimilation of short-lived tracers with high timely dynamics, as is the case for formaldehyde. In fact, the average atmospheric lifetime of formaldehyde of about three hours only is much shorter than the 12-hour 4DVar assimilation window. With growing distance of the time of the observation to the start of the assimilation, the capacity to accurately model the intermediate atmospheric chemistry and the resulting strongly pronounced diurnal cycle of formaldehyde may have increasing impact. If simultaneously to the concentrations also the emissions are optimised (and not kept constant like in the present analysis), the same findings are expected. No or little impact of observations would be expected near the beginning of the assimilation window, because variational changes in the emissions need some time to translate into changed concentrations. The representation of the atmospheric formaldehyde sources and sinks, i.e. the validity of the simplified formaldehyde chemistry scheme that is used for the assimilation, may equally time-dependently impact the assimilation. The following section therefore analyses the influence of the observation time on the assimilation of TROPOMI HCHO

retrievals. For better understanding, the IFS-COMPO assimilation window and the handling of time-dependence in ECMWF's analysis and forecasting system are briefly laid out below.

IFS-COMPO applies a twelve-hour assimilation window, during which the atmospheric state is mainly characterised by the background vector (a three-hour forecast from the previous analysis) and by the observations. Both of these are characterised by their individual error covariance matrices, which determine the weighting of the different variables to obtain the best-possible estimate of the atmospheric state. For IFS-COMPO, this is done by four-dimensional variational data assimilation (4DVar). 4DVar accounts for the time-dimension by deriving and minimizing a global cost function with the scope of optimising the initial conditions until obtaining the model trajectory that best fits both background and observations over the time window. This approach is an extension of three-dimensional variational data assimilation, which assumes that background and available observations all occurred at the time of the analysis. One obvious advantage of considering the time-dimension of the problem is the consideration of the time distribution of the observations. This is particularly impactful for dynamic atmospheric components with strong variability in space and time, e.g. tracers with a short atmospheric lifetime as is the case for formaldehyde.

In operational IFS-COMPO (CAMS cycle 49R1), two 12-hour assimilation windows are run from 3-15 and 15-3 UTC, respectively, to produce the 12/18 UTC and 00/06 UTC analysis and forecast fields. For each assimilation window, two minimizations are performed at reduced spectral resolution. This means, that the analysis at 00UTC uses observations from the time window 15-3UTC by extracting two sets of observations from the time slots 15-9UTC and 21-3UTC, respectively. For the background, the 3-hour forecast from the previous analysis is used, i.e. for the analysis at 00UTC, the model forecast from the 12UTC analysis is used as background (Appendix E, fig. E2). Equally, the 12UTC analysis uses the 3-hour forecast from the previous 00UTC analysis and observations are extracted for 3-9UTC and 9-15UTC time windows.

This section analyses the impact of the observation time by shifting the assimilation windows from their starting times at 3 and 15UTC, respectively, by six hours to 9UTC and 21UTC (as used for ECMWF NWP). As a result, the 12UTC analysis now uses observations from 9-21UTC instead of 3-15UTC. This causes minor differences in the distribution of assimilated observations. With the shifted time window, more observations are included over western longitudes and less observations over the east (fig. 4.21). As a result, the difference of simulated formaldehyde and observations is slightly reduced over South/North America and slightly increased over Asia and Australia (fig. 4.22, panel b). On a global scale, the shift of the assimilation window has negligible to minor negative impact on the assimilation results. A significant dependency of HCHO assimilation on the assimilation window was not found.

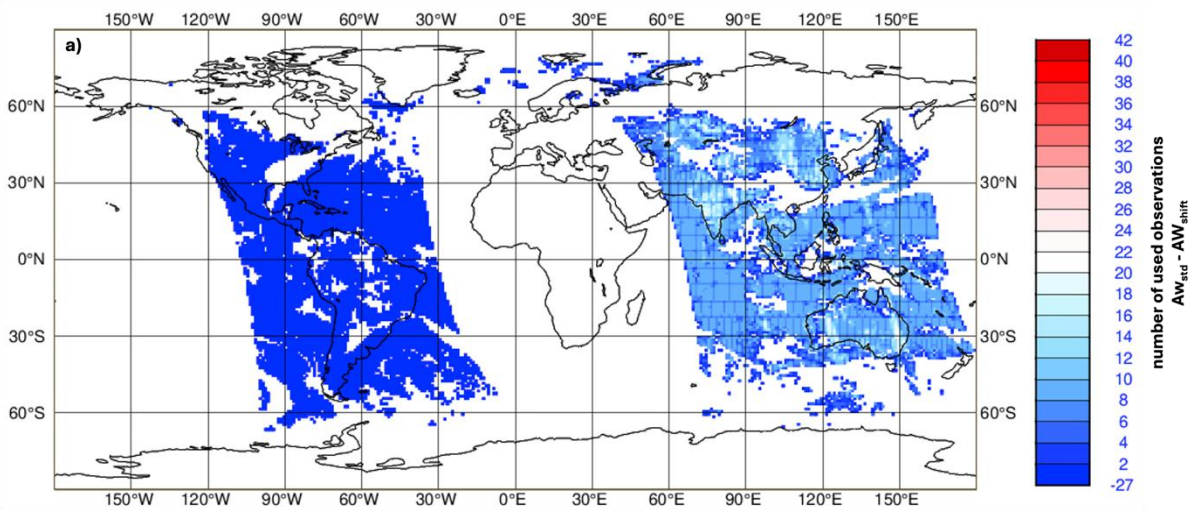


Fig.4.21: Difference of the total number of used TROPOMI satellite retrievals when shifting the assimilation window from its start at 3UTC and 15 UTC (AW_{std}) to 9UTC and 21UTC (AW_{shift}) for the time period 05/04/2025, 3UTC to 06/04/2025, 9UTC. The shift causes an increased number of included observations over the western hemisphere and less over eastern longitudes.

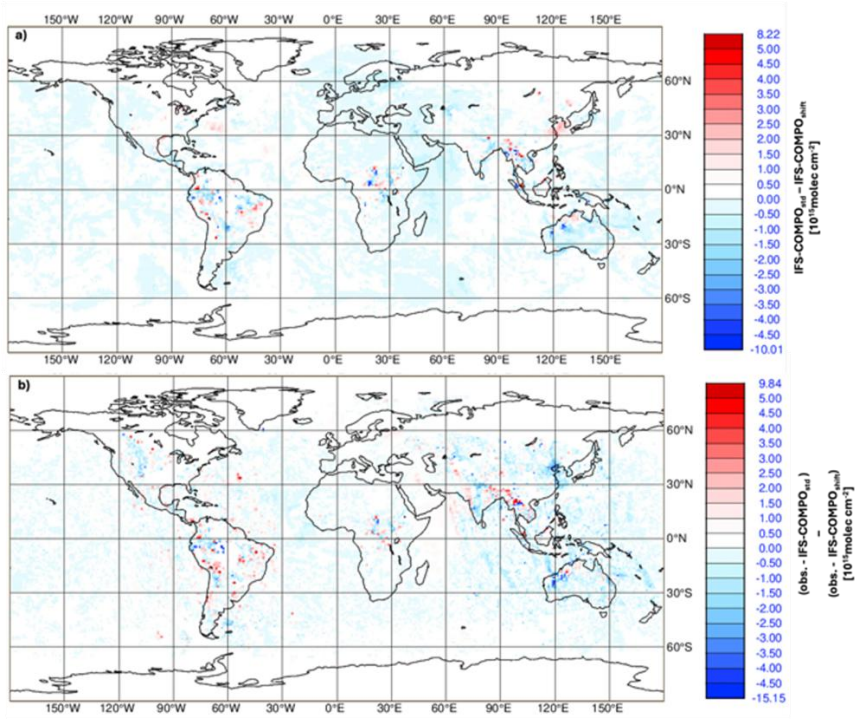


Fig.4.22: Impact of the assimilation window on the assimilation of HCHO in IFS-COMPO (panel a) with red indicating regions where the shift decreases HCHO and blue showing regions where the shift increases HCHO. Panel b) shows the absolute observation bias of IFS-COMPO formaldehyde when applying the standard assimilation window with start at 3 UTC and 15 UTC and when using a shifted window starting to 9 UTC and 21UTC. Red indicates regions where the shift improves the comparison to the satellite observations. Blue shows regions where the difference to the observations is larger when shifting the time window.

5 Emission inversion of biogenic VOCs in IFS-COMPO

5.1 Implementation of an HCHO satellite retrieval-based bVOC inversion system

The assimilation of TROPOMI HCHO retrievals was integrated into the IFS-COMPO flux inversion system. The system was modified to enable biogenic flux inversions and successfully tested. Currently, the biogenic emission inversion is based on CAMS cycle 49R1. The use of the recently developed online biogenic emission module in the inversion to complement the inventory-based emissions requires additional modifications. This is planned in a next step, when a first test version based on CY50R1 will become available. This will also enable the integration of additional satellite instruments, in particular HCHO from the geostationary satellites. At this point, the biogenic emission inversion relies on TROPOMI HCHO observations and climatological emissions from CAMS-GLOB-BIOv3.1. The isoprene flux inversion branch is *cxfk_CY49R1.M_isop_emis_inv_v0* (scripts, source, suites, and default). It contains the latest version of the emission inversion in IFS-COMPO and the formaldehyde assimilation schemes based on the tangent linear and adjoint from the simplified HCHO chemistry SC1 (low NOx). The inversion requires a configuration file, that specifies the control variable (species and sectors to be optimised) as well as the associated a-priori error assumptions (standard deviation, spatial correlation length scales). An example configuration file for isoprene inversion can be found in appendix D. The inversion products are posterior scaling factors by which the a-priori isoprene emissions need to be multiplied. They are outputted for each assimilation window (0UTC and 12UTC) and stored as grib fields, for which a corresponding number has to be specified in the configuration file. In the basic biogenic emission inversion setup, 50% uncertainty is assumed for the a-priori errors. Higher uncertainties are planned to be tested in the future as part of the background error optimisation. For the observation errors, the recommendations from the data providers are followed. The impact of the isoprene emission inversion based on TROPOMI S5P HCHO retrievals will be discussed in detail in the following section.

5.2 Impact of the HCHO-based emission inversion on a priori emissions

This section evaluates the impact of the HCHO satellite retrieval-based emission inversion system on the biogenic emissions. The tests were carried out for July 2023 using model cycle CY49R1 and branch *cxfk_CY49R1.M_isop_emis_inv_v0* (experiment ID *it5z*). Fig. 5.1 shows the resulting scaling factors from the inversion for each analysis cycle separately. Clearly visible is the diurnal cycle of the satellite retrievals, which causes the 0UTC inversion to mainly optimise emissions in easterly longitudes and the 12UTC cycle to mainly optimise the western hemisphere. For the optimisation, the a-priori emissions are meant to be multiplied by the scaling factors, such that a scaling factor of 1 indicates no change to the reference emissions. Larger scaling factors than 1 increase the reference emissions, smaller scaling factors cause an emission reduction.

The optimisation is expected to minimize the difference of the satellite retrievals and the simulation (fig. 5.2), given their respective uncertainties. Correspondingly, isoprene emissions are expected to be reduced by the inversion in regions where IFS-COMPO overestimates formaldehyde when comparing to TROPOMI HCHO and, equally, isoprene should be increased over regions where IFS-COMPO underestimates formaldehyde. This is applied correctly by the inversion system, as shown below exemplarily for the 0UTC and 12UTC analysis windows on 04/07/2023. For the 0UTC cycle, the comparison of simulated and observed formaldehyde shows an overestimation of HCHO by IFS-COMPO over the northern South American coast (around 10°N). The corresponding scaling factors reduce the a priori emissions. Equally, there is an underestimation of HCHO by IFS-COMPO over China in the order of $3x10^{15}$ molec cm^{-2}. The corresponding scaling factors increase isoprene emissions over the region.

Generally, the inversion shows expected impact on isoprene emissions by following the observational biases. However, the scaling factors are extremely small and in the order of 10^{-6} which results in negligible changes of the a-priori isoprene emissions in the same order of magnitude. The resulting changes of its atmospheric concentrations are equally negligible. At this stage, the presented analysis should be understood as a proof of concept. An inversion system for biogenic emission inversion was correctly implemented into IFS-COMPO. However, the inversion results clearly show that there is need for further optimisation of the system and more detailed understanding of its optimal integration into IFS-COMPO. Reasons for the small impact of the satellite observations on the underlying emissions potentially lie in not optimal background, prior and observation uncertainty estimates, all of which directly impact the weighting of the involved variables for the optimisation. Equally, a systematic deficiency may be caused by the indirect link of formaldehyde and isoprene atmospheric chemistry, which could potentially not be strong enough in IFS-COMPO to significantly modify isoprene emissions based on formaldehyde assimilation. Further, the applied simplified HCHO chemistry may not be sufficiently accurate to approximate a realistic isoprene formaldehyde chemical relationship. An indication for potential deficiencies of the simplified HCHO chemistry scheme are the results of the initial condition optimisation, which has shown to equally only cause minor changes in atmospheric formaldehyde (section 4).

To systematically analyse the questions raised above, after the implementation of the inversion system into IFS-COMPO, a sensitivity study was performed that analyses the isoprene-formaldehyde link of (a) standard IFS-COMPO and (b) the simplified IFS-COMPO HCHO chemistry scheme. The results of this analysis are evaluated in detail in the next section.

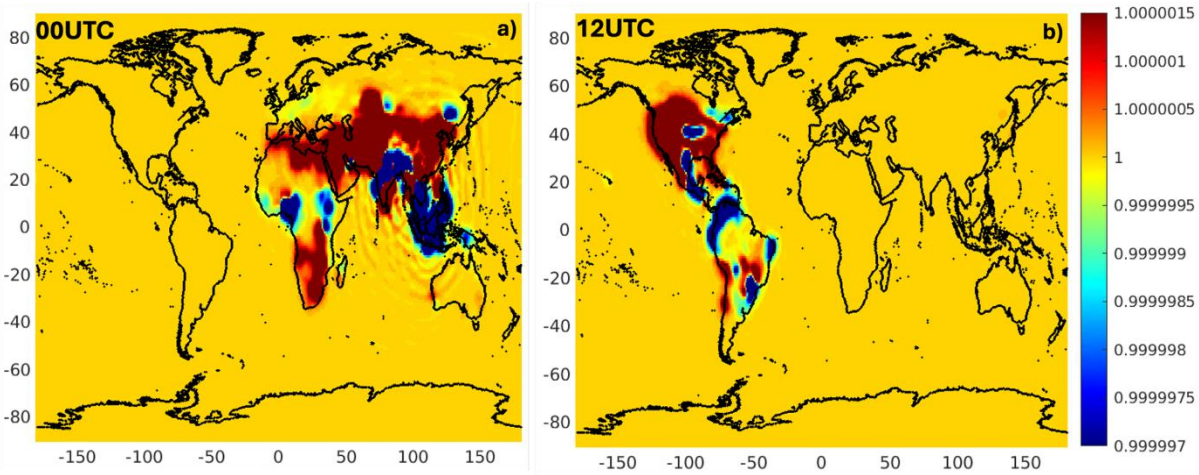


Fig. 5.1: Scaling factors for isoprene emissions from the 0UTC (panel a) and 12UTC (panel b) assimilation windows for 04/07/2023 with a scaling factor of 1 implying no change to the reference emissions.

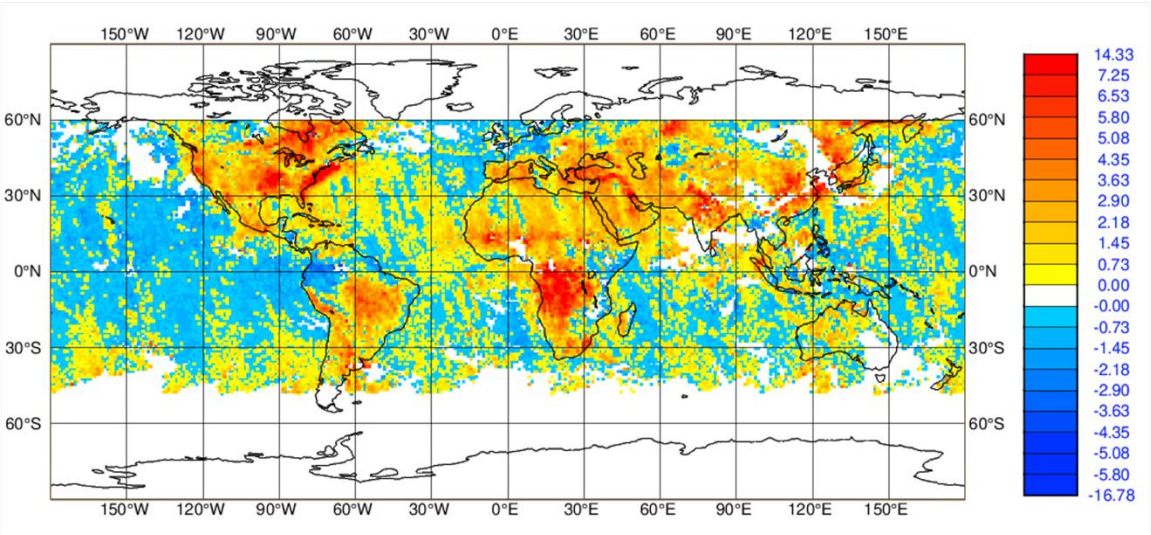


Fig. 5.2: Comparison of TROPOMI HCHO retrievals and assimilated IFS-COMPO HCHO on 04/07/2023 with red colour indicating larger observed formaldehyde and blue colour indicating less observed formaldehyde than simulated by IFS-COMPO.

5.3 Sensitivity study of isoprene – formaldehyde chemistry in IFS-COMPO

The small impact of the bVOC inversion on the posterior emissions may be caused by various factors ranging from the inversion system itself (i.e. its integration into the IFS) to not-optimal uncertainty assumptions (e.g. with respect to the background and a-priori errors). One other potential reason may be the strength of the link of formaldehyde and isoprene in the IFS. The bVOC inversion performs an indirect flux optimisation, where formaldehyde satellite observations are used to optimise emissions of a different precursor species. The strength of their chemical link therefore directly affects the impact of the inversion. While over most continental regions, atmospheric formaldehyde in fact originates from isoprene emissions, there are secondary, isoprene-unrelated sources, and furthermore the chemical conversion from isoprene emissions to formaldehyde is not spontaneous, but delayed in time and hence also space. Finally, it is also possible that the implemented HCHO chemistry scheme over-simplifies bVOC chemistry to an extent that the formaldehyde - isoprene sensitivity in the simplified chemistry is insufficient for its purpose. An indication for the latter may be that not only the inversion system, but also the HCHO assimilation itself only shows minor impact on the discussed species. To improve the understanding of such method-related factors, the following section systematically investigates the link of isoprene and formaldehyde both in the IFS standard configuration and when using the simplified HCHO chemistry scheme. If this sensitivity is found to be insufficient or largely deviating between the two chemical schemes, further optimisation of the simplified atmospheric chemistry of formaldehyde and isoprene may be necessary.

The sensitivity of IFS-COMPO formaldehyde concentrations to isoprene emissions and the impact of atmospheric chemistry on this link is tested by scaling a-priori isoprene emissions by differing factors and analysing the resulting variance of full-IFS and simplified-IFS formaldehyde. The evaluation is performed using a standard isoprene emission dataset from CAMS-GLOB-BIOv3.1. For the tests, the emissions from this dataset were left unchanged (reference), decreased by 50% (SF05), and increased by 100% (SF2, fig. 5.3). These scaled emission maps are then used for the biogenic VOC emissions in three forecast experiments that where run for a total duration of one month (July 2023, table 5). Each experiment simultaneously simulates HCHO according to (a) the standard IFS-COMPO chemistry and (b) the simplified HCHO chemistry scheme SC1 (sect. 3).

Table 5: Forecast experiments for the sensitivity tests of the simplified HCHO chemistry.

name	experiment ID	branch	isoprene emissions
ref	b2un	cxfk_CY49R1_simple_chem_1	standard
SF05	b2up	cxfk_CY49R1_simple_chem_1	decreased by factor 0.5
SF2	b2uo	cxfk_CY49R1_simple_chem_1	increased by factor 2

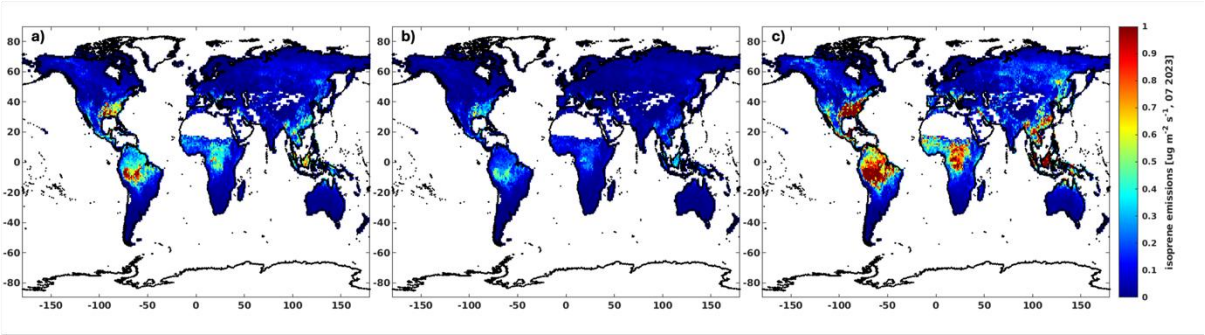


Fig. 5.3: CAMS-GLOB-BIOv3.1 isoprene emissions for July 2023 as used for the sensitivity tests with reference emissions (panel a), reduced emissions by 50% (panel b), and increased emissions by 100% (panel c).

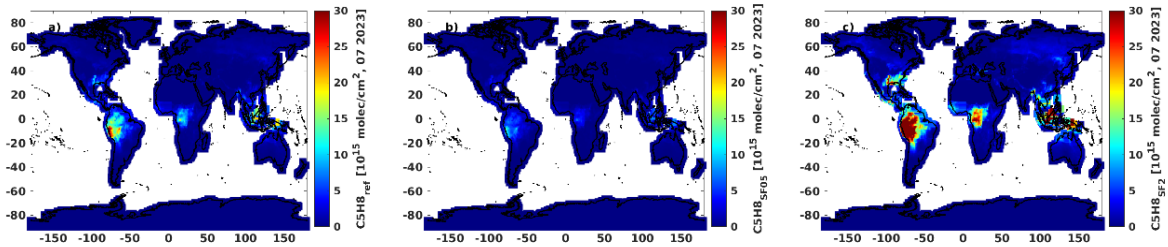


Fig. 5.4: Vertical column densities of IFS-COMPO isoprene VCDs using reference emissions (panel a), 50% reduced emissions (panel b), and 100% increased emissions (panel c) for July 2023. In correspondence to the global isoprene emission hotspots (fig. 5.3), the main impact on the VCDs is found in tropical latitudes.

The following section firstly discusses the analysis on a global scale, followed by a more detailed discussion of a selected region (Amazonian rain forest), where isoprene emissions are the pre-dominant biogenic VOC source both in the simplified and in the full chemistry IFS-COMPO configuration. In this region, the sensitivity of formaldehyde to the change in isoprene emissions is expected to be more pronounced than in air masses affected by small (e.g. oceans) or non-biogenic emissions. This is particularly true for the standard IFS-COMPO configuration, where non-isoprene sources potentially have stronger impact on formaldehyde concentrations due to the more complex HCHO source chemistry.

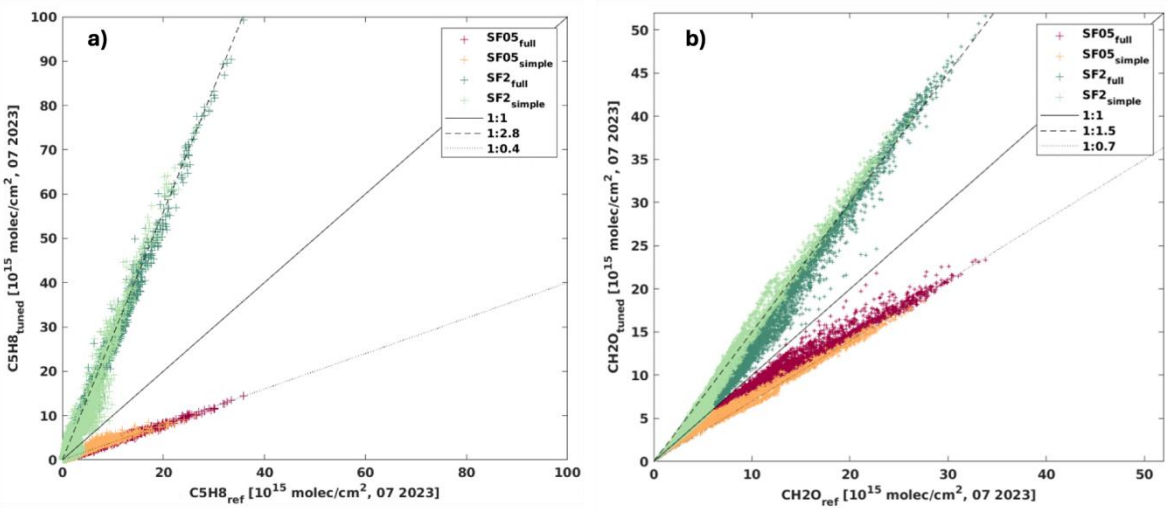


Fig. 5.5: Absolute change in global isoprene and formaldehyde VCDs due to modified isoprene emissions for July 2023 with different colours indicating standard IFS-COMPO chemistry (dark red, dark green) and simplified formaldehyde chemistry (bright red, bright green). Simulations applying standard CAMS-GLOB-BIOv3.1 isoprene emissions ('ref') are plotted on the x-axis, simulations using tuned isoprene emissions are shown on the y-axis ('tuned'). Data using 50% reduced isoprene emissions is marked with SF05, results based on 100% increased isoprene are plotted as SF2. Evidently, the change in emissions directly reflects in the atmospheric VCDs of both tracers by causing increased/decreased VCDs.

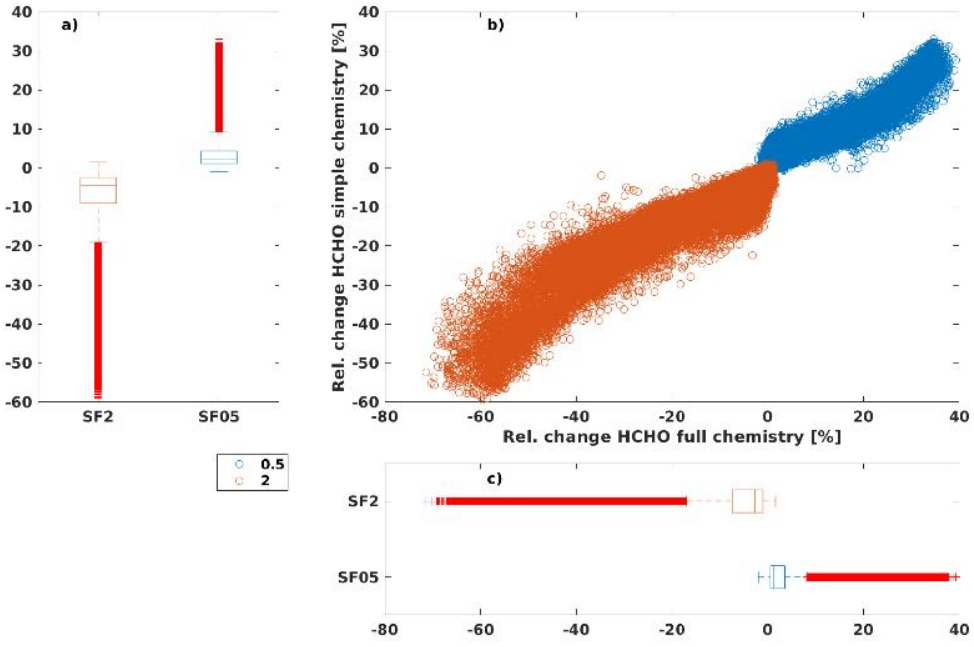


Fig. 5.6. Global sensitivity of the two HCHO chemistry configurations (simple and full IFS-COMPO chemistry) to the changes in isoprene emissions with SF05 indicating a 50% emission reduction (blue) and SF2 an 100% increase in isoprene emissions (orange) for July 2023. Relative changes are calculated for each chemistry scheme according to ($HCHO_{tuned}$-$HCHO_{ref}$)/$HCHO_{ref}$ with 'tuned' indicating simulations with changed isoprene emissions and 'ref' simulations using the reference emissions. For both HCHO chemistry configurations, the median change in global HCHO is smaller than 5% and the difference between the schemes is minimal.

5.3.1 Global sensitivity

Evidently, the change in isoprene emissions directly reflects in the global isoprene abundance by largely increasing/decreasing isoprene concentrations and vertical column densities (fig. 5.4). This change in isoprene concentrations translates into modified formaldehyde VCDs of both simplified and full IFS-COMPO chemistries, however at a more moderate extent. This is discussed in more detail below.

On a global scale, a decrease in isoprene emissions by 50% causes median decreases in global isoprene VCDs of 56% (fig. 5.5, panel a), which in turn cause a median decrease in formaldehyde of 1% (simplified chemistry) and 2% (full chemistry, fig. 5.6). Equally, increasing isoprene emissions by 100% causes 140% larger median isoprene VCDs, but only 2% (simplified) and 4% (full chemistry) change in formaldehyde. The evaluation shows that on a global scale, extreme changes in isoprene emissions are required to significantly alter formaldehyde concentrations in IFS-COMPO independent on the applied HCHO chemistry. This finding is to be expected, since HCHO is significantly dependent on isoprene only in regions, where isoprene is the main formaldehyde precursor, i.e. regions with high biogenic emissions. In the background and marine atmosphere, methane is the main formaldehyde precursor and the chemical HCHO production has a correspondingly weaker link to isoprene emissions.

In fact, there are no isoprene emissions above marine surfaces. Therefore, the changes in isoprene emissions mainly alter terrestrial bVOC and the (mostly) unaffected marine HCHO decreases the globally averaged sensitivity. This dependency of HCHO sensitivity on air mass composition and origin also reflects in a pronounced latitudinal sensitivity pattern, with higher sensitivity over isoprene source regions and smaller sensitivity elsewhere, e.g. marine air masses (fig. 5.7). Between 40°N to 70°N, where large biogenic emissions occur during the analysed period (July), the relative impact of increased isoprene emissions on formaldehyde concentrations is larger than in other latitudes for both chemistry schemes and emission perturbations. In the tropics, standard IFS-COMPO HCHO clearly shows larger median isoprene-sensitivity then simplified HCHO (fig. 5.7, panel b). The same pattern is found for the 50% reduction in isoprene emissions (Appendix E, fig. E5). At the same time, the simplified chemistry is either comparably or even more sensitive to changes in isoprene concentrations (instead of emissions) for all latitudes. This becomes evident when comparing fig. 5.7 left versus right panels, where the sensitivity to changes in isoprene emission is shown on the left, and the sensitivity to changes in isoprene VCDs is plotted on the right. Clearly, the reduced sensitivity with respect to isoprene emissions that is found for the tropics does not translate into a generally reduced isoprene sensitivity. As such, it is probably not resulting from the simplified isoprene-formaldehyde chemistry scheme, but is caused by other factors, which are discussed in more detail below. The above findings are of particular relevance, since tropical latitudes are a region where high biogenic emissions occur and the inversion is expected to have a larger impact on the emission. The following section therefore analyses the link of isoprene and formaldehyde above the Amazonian rain forest in more detail.

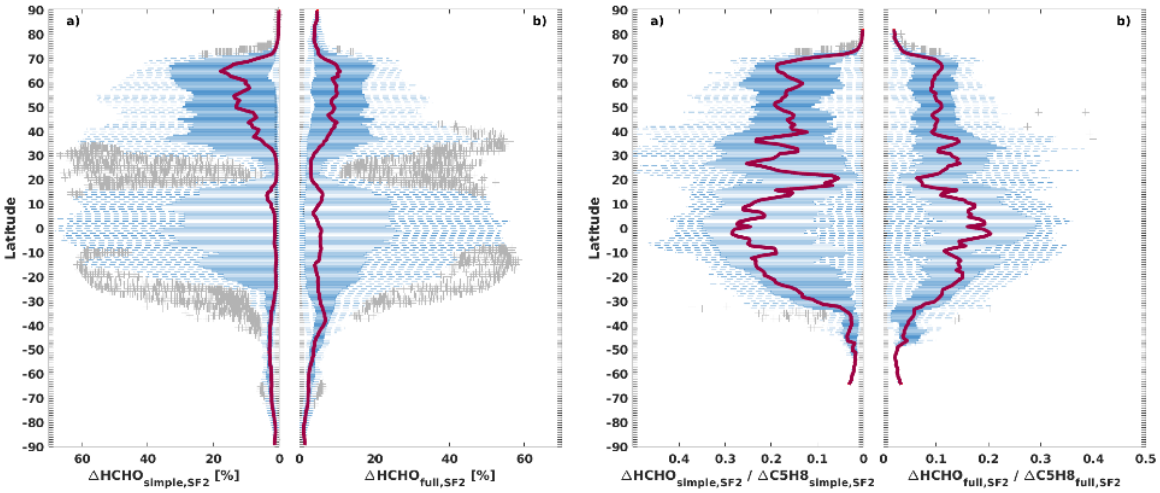


Fig. 5.7: Left: Latitudinal dependency of the relative change in formaldehyde due to an increase in isoprene emissions of 100% for simplified (panel a) and standard IFS-COMPO configuration (panel b) for July 2023. The red line marks the median relative change for a given latitude. Blue horizontal boxes indicate the lower and higher 25/75 percentile, dashed lines the minima and maxima and grey colour data outliers. The relative change is calculated for each chemistry scheme according to ($HCHO_{SF2}$-$HCHO_{ref}$)/$HCHO_{ref}$ with SF2 indicating simulations with changed isoprene emissions and 'ref' simulations using the reference emissions. Right: Relative change in formaldehyde as a function of the relative change in precursor isoprene concentrations for simplified (panel a) and standard IFS-COMPO configuration (panel b) for July 2023.

5.3.2 Sensitivity above the Amazonian rain forest

Above the Amazonian rain forest, a 100% increase of isoprene emissions translates into 166% increased isoprene VCDs (both chemistries) and 52%/58% increased HCHO VCDs (full/simple chemistry scheme). Equally, a reduction of isoprene emissions by 50% causes 60% reduced isoprene VCDs and 27%/33% decreased HCHO (full/simple chemistry, monthly median for July 2023). As a rule of thumb, above bVOC source regions, changes in isoprene emissions appear to translate into IFS-COMPO formaldehyde VCDs via a factor of 0.5. For the analysed time period, the resulting isoprene VCDs are larger in the simplified chemistry than standard chemistry for all times of the day, whereas HCHO from the simplified chemistry has a constant negative bias with respect to standard IFS-COMPO HCHO (fig. 5.8).

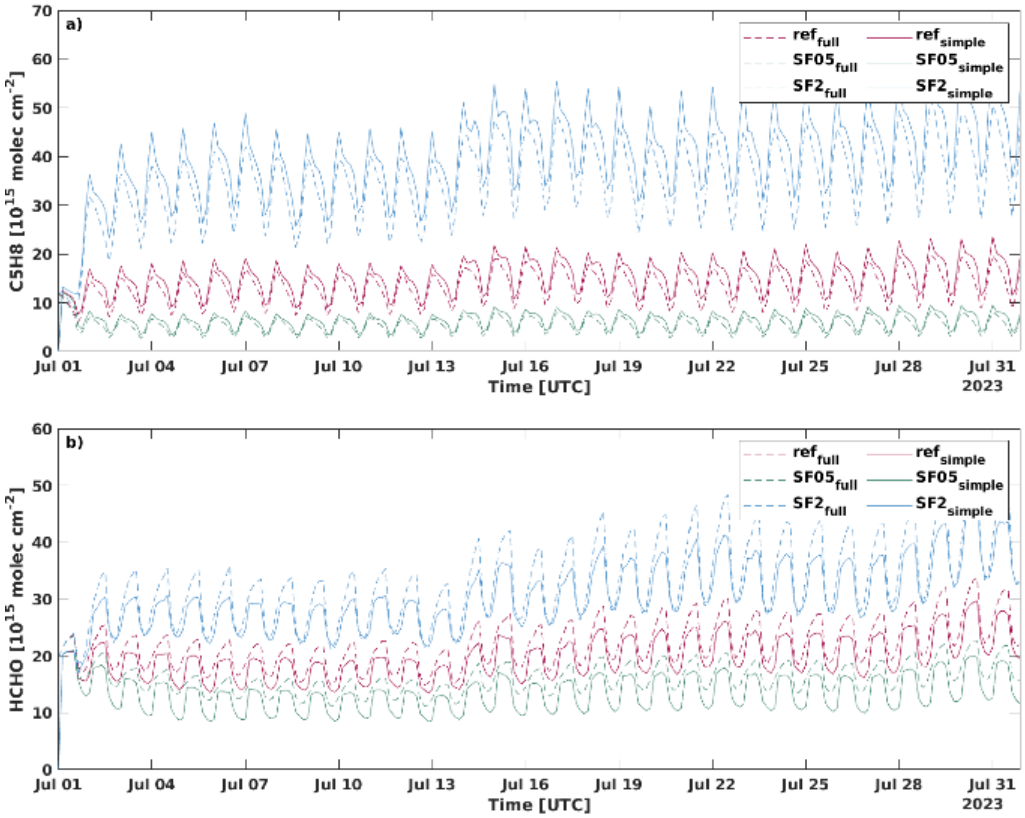


Fig. *5.8: Timeseries of isoprene (panel a) and formaldehyde (panel b) VCDs above the Amazonian rain forest (-10°N to 5°N and -75°E to -55°E) for unchanged isoprene emissions (red), reduced isoprene emissions (green) and increased emissions (blue) during July 2023. Local time corresponds to UTC-3h. Plotted are both standard IFS-COMPO chemistry (dotted lines) and isoprene from the simplified HCHO chemistry (solid lines). The simulations start from the same initial conditions (time=0) and then subsequently diverge due to the different emission scenarios (time step 3h onwards).*

For both chemistry configurations and all isoprene emission scenarios, formaldehyde reaches peak concentrations during local morning hours (fig. 5.9, panel b), 9h after peak isoprene concentrations (fig. 5.9, panel a). A minor negative bias of absolute HCHO of the simplified chemistry configuration persists for all times of the day. This difference between the chemistry schemes has been discussed in detail in deliverable D2.3 and section 3. Full IFS-COMPO formaldehyde and simplified chemistry formaldehyde both show a clear sensitivity to the change in atmospheric isoprene concentrations for all times of the day. However, based on the diurnal cycle, it becomes evident that the sensitivity is to some extent time dependent, with deviations between the chemistry schemes between 9LT and 21LT (fig. 5.9, panel c). During this half of the day, the isoprene-sensitivity of simplified HCHO increases by up to 13%. Contrarily, the sensitivity of standard IFS-COMPO HCHO decreases by 11%. Evidently, during daytime and above bVOC source regions, the link of isoprene and formaldehyde in the simplified chemistry exceeds that of full chemistry IFS-COMPO. This causes a somewhat opposed diurnal cycle of the isoprene-formaldehyde sensitivity of both schemes, with a maximum sensitivity of full IFS-COMPO at 9LT and of the simplified chemistry at 21LT.

The largest relative change in isoprene VCDs occurs between 12LT and 15LT (figs. 5.10 and 5.11, panel c). This isoprene pattern is a direct consequence of the superimposed diurnal bVOC emission cycle in IFS-COMPO. It accounts for the dependency of biogenic emissions

on temperature and photosynthetically active radiation, i.e. the amount of available sunlight. As a result, isoprene emissions in IFS-COMPO peak at local midday. This appears to quickly lead to relatively larger changes in isoprene VCDs at 15LT as compared to the rest of the day. 6h later, at 21LT, this maximum isoprene variability translates into maximum HCHO variability in the simplified chemistry (fig. 5.9, panel c and fig. 5.10, panel a). For standard IFS-COMPO, this link of isoprene concentration variability and formaldehyde variability seems less direct. Peak formaldehyde sensitivity occurs at 9LT, i.e. 18h after the maximum isoprene concentrations changes (fig. 5.11, panel a). This timely delay in the formaldehyde reaction to isoprene emission changes in standard IFS-COMPO may cause the 6% reduced median sensitivity as compared to the simple scheme. It also reflects in the global picture (fig. 5.7, right), which shows generally smaller sensitivity of standard IFS-COMPO to isoprene concentration changes.

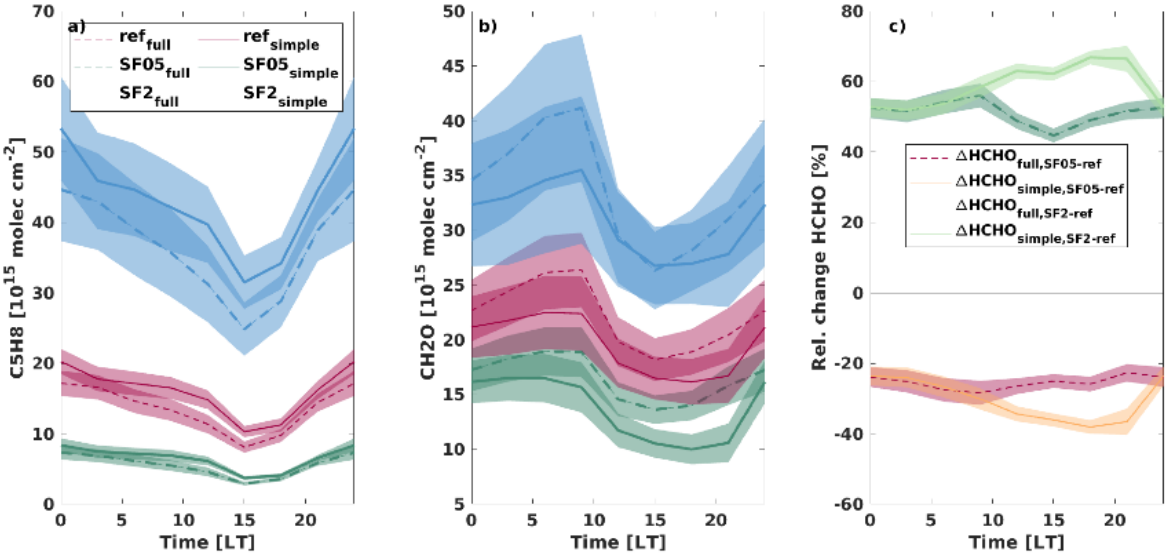


Fig. 5.9: Impact of 100% increased (SF2) and 50% reduced (SF05) versus standard (ref) isoprene emissions on the monthly averaged diurnal cycles of isoprene (panel a) and formaldehyde (panel b) based on full IFS-COMPO (full) and simple HCHO chemistry (simple) over the Amazonian rain forest for July 2023. Panel c) shows the median relative change in HCHO as a function of time due to the changed emissions according to $HCHO_{full/simple,SF05/2\text{-}ref}=(HCHO_{full/simple,SF05/2}-HCHO_{full/simple,ref})/HCHO_{full/simple,ref}$ *for full and simple chemistry and emission scenarios SF05 and SF2, respectively.*

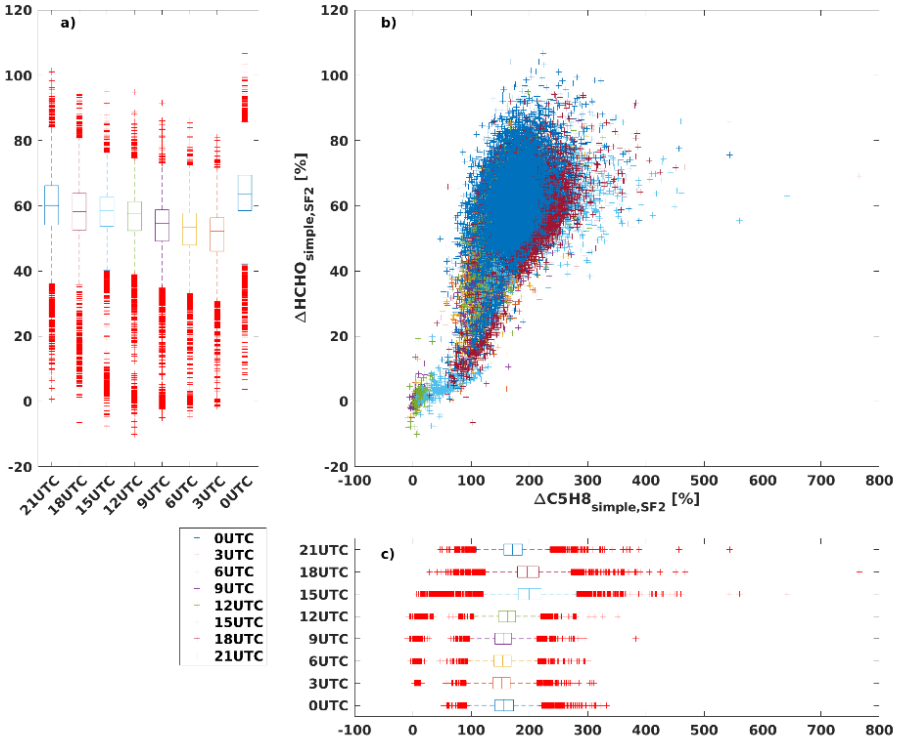


Fig. 5.10: Sensitivity of simplified chemistry HCHO to changes in isoprene VCDs above the Amazonian rain forest as a function of the daytime. Local time corresponds to UTC-3h.

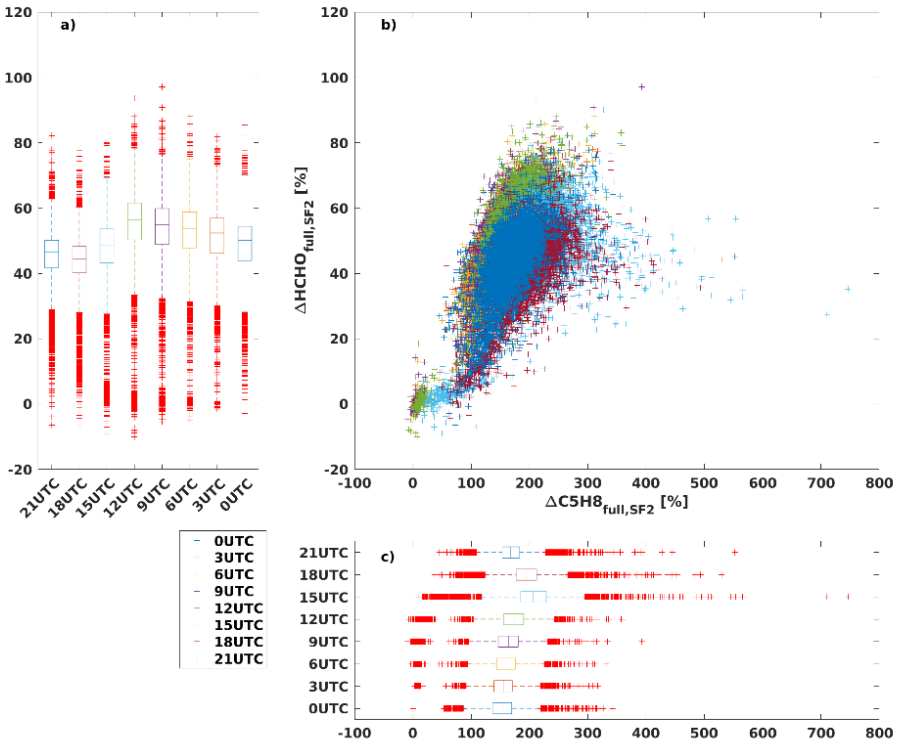


Fig. 5.11: As above but using the standard IFS-COMPO chemistry configuration.

5.3.3 Summary

The above discussion shows that despite some diurnal variations, the sensitivity of the chemistry schemes to isoprene emissions is either comparable (nighttime) or up to 16% larger in the simplified chemistry (daytime) above biogenic source regions. Averaged over the entire day, the median difference of 6% between the HCHO chemistry schemes is within the standard deviation (6.1%). Clearly, the reduced sensitivity of simplified HCHO in tropical latitudes (fig. 5.7) neither originates from a diurnal pattern, nor from systematic insufficiencies of the simplified chemistry to link isoprene and formaldehyde in biogenic air masses. Consequently, it has to result from differing performances of the chemistry schemes outside of biogenically dominated air, most probably either of marine or anthropogenic composition. The latter would be expected to also effect the comparison in the northern hemisphere, where total anthropogenic emissions are larger than in the southern hemisphere. However, this does not reflect in the data. Contrarily, the longitudinally averaged northern hemisphere sensitivity of simplified HCHO to isoprene emission changes exceeds that of standard IFS-COMPO HCHO. Therefore, since it cannot be traced back to biogenic or anthropogenic air, the reason for the reduced sensitivity of the simplified HCHO in the tropics is most probably related to the chemistry performance in background and marine air masses. These make up a considerable fraction of the air masses found in tropical latitudes.

In fact, when filtering the data for terrestrial-only and marine-only, the isoprene-sensitivity of the chemistry schemes shows significant deviations, which potentially explain the above findings. In continental air, the simplified chemistry shows larger isoprene sensitivity than standard IFS-COMPO for all global latitudes, including the tropics (fig. 5.12). However, the simplified scheme shows negligible sensitivity to isoprene emission changes in background and marine air (fig. 5.13). Standard IFS-COMPO, on the other side, also reacts to isoprene emission changes in background and marine air masses. At the same time, HCHO from the simple chemistry has higher sensitivity regarding isoprene concentration (instead of emission) changes for all air masses and latitudes (fig. 5.7, right). This necessarily implies, that the simple scheme is highly sensitive towards bVOC chemistry close to emission sources (potentially related to NOx chemistry), but notably loses sensitivity with growing distance from the emission locations.

Generally, the analysis has shown a superior sensitivity of the simplified chemistry HCHO to isoprene concentration changes for all global regions and air mass compositions and a comparable to superior sensitivity to isoprene emission changes in terrestrial air. This includes in particular biogenic emission hotspots such as tropical rain forests. At the same time, the scheme is not suited to represent timely delayed formaldehyde production related to isoprene oxidation, which causes missing sensitivity in remote and marine air. While further tuning of the chemistry scheme may be needed to improve this aspect, the analysis does not suggest a generally insufficient isoprene-formaldehyde link in the HCHO tangent linear and adjoint chemistry.

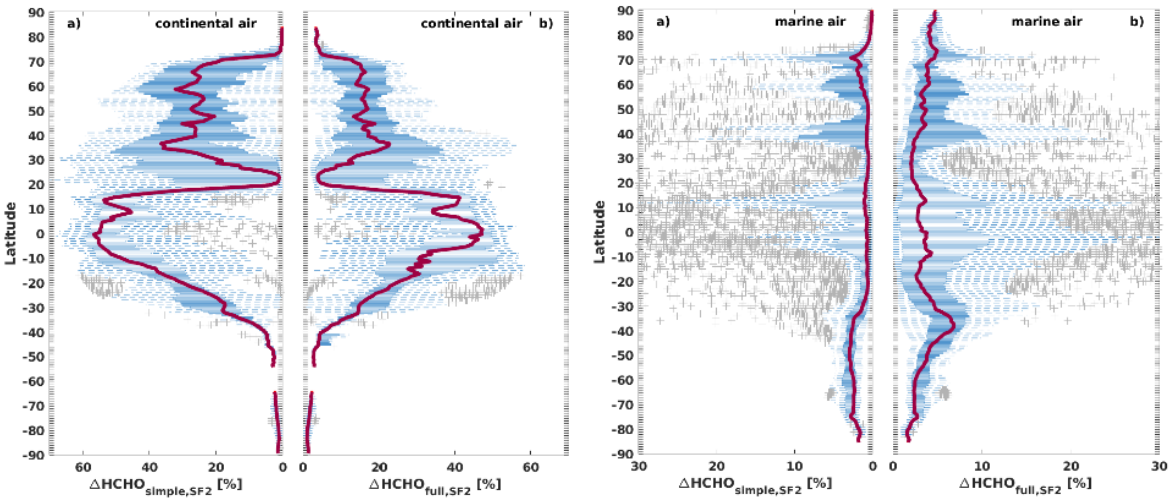


Fig. 5.12: Relative change in continental (left) and marine (right) HCHO due to an increase in isoprene emissions by 100% for simplified (panel a) and standard IFS-COMPO configuration (panel b) for July 2023. 25% water coverage is chosen as threshold between the regimes.

6 Conclusion

In this deliverable, we have reported on the implementation of a satellite-retrieval based inversion system for biogenic volatile organic compound emissions in the IFS global atmospheric composition model. The system relies on a tangent linear and adjoint based formaldehyde assimilation system. Technical details of the assimilation, the included satellite instruments, and the assimilation impact on atmospheric composition in IFS-COMPO were discussed in detail. Globally, the assimilation of HCHO satellite retrievals showed either negligible or minor positive impact on IFS-COMPO HCHO and generally negligible impact on atmospheric precursors (isoprene) and related tracers (CO, ozone). The largest impact on HCHO was found close to the surface and for regions with combined geostationary and polar-orbiting satellite observations. Based on the tangent linear and adjoint of a simplified HCHO chemistry scheme, an HCHO based inversion system for biogenic VOCs was successfully developed and tested in CAMS model cycle 49R1. Globally, the inversion system showed negligible impact on the bottom-up emission estimates. Evidently, this is a consequence of the low impact of the satellite retrievals on the HCHO fields in the assimilation, which translates into equally low impact of the observations on the surface emissions.

Potential reasons were suspected in the strength of the chemical link of biogenic emissions to formaldehyde, which may be insufficiently represented by the simplified HCHO chemistry scheme. Therefore, a sensitivity study was performed that aimed to systematically investigate the sensitivity of formaldehyde to bVOC emissions both in standard IFS-COMPO and in the assimilation system. The sensitivity study clearly indicates that simplified and full IFS-COMPO HCHO chemistry show comparable sensitivity to isoprene emission changes above isoprene source regions and a globally higher sensitivity of the simplified chemistry to isoprene concentration changes. Changes in isoprene emissions were found to quickly translate into modified formaldehyde VCDs via a factor of 0.5 (100% change of isoprene emissions causes 50% modified formaldehyde VCDs). A systematic insensitivity of the currently implemented formaldehyde tangent linear and adjoint to isoprene emissions above isoprene source regions could be ruled out.

At the same time, the isoprene-HCHO sensitivity is significantly smaller in non-biogenic air, where large changes in isoprene emissions only have negligible effect on HCHO concentrations in the simplified chemistry. As a result, large changes in isoprene emissions

are required to significantly alter formaldehyde concentrations in IFS-COMPO on a global scale (independently on the applied HCHO chemistry). This higher sensitivity of the simplified chemistry in biogenic source regions and the lower sensitivity in background air does currently neither reflect in the inversion nor in the assimilation results. The scaling factors for the posterior emissions were negligible for all global regions and latitudes independent of air mass composition.

While it could be shown that an insufficient sensitivity of the applied HCHO chemistry scheme cannot fully explain the inversion results, it may still contribute to the findings. The presented sensitivity study clearly indicates that further tuning of the simplified HCHO chemistry and of its implementation into the IFS may be needed to improve the assimilation and inversion results and corresponding tests are planned to be carried out in the coming months. This includes testing of the variable NOx optimised simplified HCHO chemistry as well as additional sensitivity studies to investigate how the reported isoprene emission sensitivity of the simple chemistry in forecast simulations translates into the assimilation. Additionally, the HCHO data assimilation system will be finetuned and optimised aiming to increase the impact of the satellite retrievals on the surface fields and to optimise the error matrices by analysing the associated uncertainties of background, observations, and prior emissions.

7 Acknowledgements

Figures and statistical analysis in section 4.1.3 courtesy of Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project. The data used in this publication were obtained from the Network for the Detection of Atmospheric Composition Change (NDACC) and are available through the NDACC website www.ndacc.org.

8 References

Lee, Gitaek T., et al. "First evaluation of the GEMS formaldehyde product against TROPOMI and ground-based column measurements during the in-orbit test period." *Atmospheric Chemistry and Physics* 24.8 (2024): 4733-4749.

McNally, Anthony P. "On the sensitivity of a 4D-Var analysis system to satellite observations located at different times within the assimilation window." *Quarterly Journal of the Royal Meteorological Society* 145.723 (2019): 2806-2816.

Oomen, Glenn-Michael, et al. "Weekly derived top-down volatile-organic-compound fluxes over Europe from TROPOMI HCHO data from 2018 to 2021." *Atmospheric Chemistry and Physics* 24.1 (2024): 449-474

Tripathi, Nidhi, et al. "Impacts of convection, chemistry, and forest clearing on biogenic volatile organic compounds over the Amazon." *Nature Communications* 16.1 (2025): 4692.

Vigouroux, Corinne, et al. "TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations." *Atmospheric Measurement Techniques* 13.7 (2020): 3751-3767.

Wells, Kelley C., et al. "Satellite isoprene retrievals constrain emissions and atmospheric oxidation." *Nature* 585.7824 (2020): 225-233.

9 List of Abbreviations

CAMS Copernicus Atmosphere Monitoring Service

GEMS Geostationary Environment Monitoring Spectrometer

HCHO/CH2O Formaldehyde

IFS Integrated Forecasting System

IFS-COMPO CAMS atmospheric composition forecasting system

SC1 HCHO simplified chemistry scheme 1 (high VOC, low NOx)

SC2 HCHO simplified chemistry scheme 2 (variable VOC and NOx level)

S5P Sentinel 5-precursor

TEMPO Tropospheric emissions: Monitoring of Pollution

TL&AD Tangent linear and adjoint

TROPOMI TROPOspheric Monitoring Instrument

VCD Vertical Column density

VMR Volume Mixing Ratio

10 Appendix A. Simplified HCHO Chemistry Schemes

The FORTRAN source code that implements the simplified chemistries in the IFS are integrated at “ifs-source/arpifs/chem/tm5_simple_hcho_chem.F90”. The source codes are also available from the ifs-source branches *cxfk_CY49R1_HCHO_simple_chem_1* (low NOx) and *cxfk_CY49R1_HCHO_simple_chem_2* (variable NOx).

```
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

  IMPLICIT NONE

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

  TYPE(TYPE_GFLD)   ,INTENT(INOUT):: YGFL
  INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON
  REAL(KIND=JPRB),INTENT(IN)    :: PDT
  REAL(KIND=JPRB),INTENT(IN)    :: PRR(KLON,NREAC)
  REAL(KIND=JPRB),INTENT(IN)    :: PRJ(KLON,NPHOTO)
  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
 89  DO JL=KIDIA,KFDIA
 90
 91    ! R1.1: ISOP+OH reaction rate.
 92    ZRR_1=PRR(JL,KC76)* MAX(PY(JL,IOH),1.E-30_JPRB)
 93
 94    ! R1.2: IOX+OH reaction rate.
 95    ZRR_2= 10.05E-12_JPRB * MAX(PY(JL,IOH),1.E-30_JPRB)
 96
 97    ! L1/2: HCHO loss rates
 98    ZRJA=PRJ(JL,JACH2O)
 99    ZRJB=PRJ(JL,JBCH2O)
100    ZRR_3=PRR(JL,KFRMOH)* MAX(PY(JL,IOH),1.E-30_JPRB)
101
102    ! Methane loss rate:
103    ZRR_4 = PRR(JL,KCH4OH) * PY(JL,ICH4) * MAX(PY(JL,IOH),1.E-30_JPRB)
104
105    ! CHEMISTRYs
106    ! ----------
107
108    ! XO2 production:
109    ZPOX= PY(JL, IISOP_TLAD)*ZRR_1
110    ZLOX= ZRR_2
111
112    ! HCHO production and loss:
113    ZPHCHO= ZALPHA * PY(JL, IISOP_TLAD)*ZRR_1 &
114        & + ZBETA  * PY(JL, IIOX_TLAD)*ZRR_2 &
115        & + ZRR_4
116    ZLHCHO= ZRJA + ZRJB + ZRR_3
117    PY(JL, IHCHO_TLAD)=(PY(JL, IHCHO_TLAD) + ZPHCHO*PDT)/(1._JPRB+ZLHCHO*PDT)
118
119    ! Isoprene loss:
120    PY(JL, IISOP_TLAD)=1._JPRB/(1._JPRB+ZRR_1*PDT) * PY(JL, IISOP_TLAD)
121
122    ! IOX loss:
123    PY(JL, IIOX_TLAD)= (PY(JL, IIOX_TLAD)+ZPOX*PDT)/(1._JPRB+ZLOX*PDT)
124
125  ENDDO !JL
126
127  IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',1,ZHOOK_HANDLE)
128  END SUBROUTINE TM5 SIMPLE HCHO CHEM
```

Fig. A.1: Simplified HCHO chemistry scheme for low NOx conditions (SC1).

```
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.
  !     --------------
  !     01-2025 - Expanded simple chemistry to full NOx-regime.
  !        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, IISOPO2_TLAD, IIOX_TLAD, IHCHO_TLAD, IOH, ICH4, IHO2, INO
  !* reaction rates
  USE TM5_CHEM_MODULE, ONLY : KC76, KOHISPD, KFRMOH, KCH4OH, NREAC
  USE TM5_PHOTOLYSIS , ONLY : NPHOTO,&
    !* photolysis rates
     & JACH2O,JBCH2O

  IMPLICIT NONE

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

  TYPE(TYPE_GFLD)   ,INTENT(INOUT):: YGFL
  INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON
  REAL(KIND=JPRB),INTENT(IN)    :: PDT
  REAL(KIND=JPRB),INTENT(IN)    :: PRR(KLON,NREAC)
  REAL(KIND=JPRB),INTENT(IN)    :: PRJ(KLON,NPHOTO)
```

```
 70  REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,YGFL%NCHEM+4)  ! final concentrations
 71
 72  REAL(KIND=JPHOOK)    :: ZHOOK_HANDLE
 73
 74  REAL (KIND=JPRB)              :: ZRJA,ZRJB, ZRR_1,ZRR_2,ZRR_3, ZRR_4, ZRR_5, ZRR_6
 75  REAL (KIND=JPRB)              :: ZPOX, ZLOX, ZPHCHO, ZLHCHO, ZPISOPO2, ZLISOPO2
 76
 77  REAL (KIND=JPRB),PARAMETER    :: ZALPHA = 0 ! HCHO yield from ISOPO2 + HO2
 78  REAL (KIND=JPRB),PARAMETER    :: ZBETA  = 1.89 ! IOX yield from ISOPO2 + HO2
 79  REAL (KIND=JPRB),PARAMETER    :: ZGAMMA  = 1.42 ! HCHO yield from ISOPO2 + NO
 80  REAL (KIND=JPRB),PARAMETER    :: ZDELTA  = 0.67 ! IOX yield from ISOPO2 + NO
 81  REAL (KIND=JPRB),PARAMETER    :: ZEPSILON  = 1.0 ! HCHO yield from OX + OH !R2.4
 82
 83
 84
 85  INTEGER(KIND=JPIM) :: JL
 86
 87  IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',0,ZHOOK_HANDLE)
 88
 89
 90  DO JL=KIDIA,KFDIA
 91
 92    ! Reaction rates
 93    ! ----------
 94    ZRR_3=PRR(JL,KFRMOH)* MAX(PY(JL,IOH),1.E-30_JPRB) !L1: HCHO+OH loss
 95    ZRJA=PRJ(JL,JACH2O) !L2: HCHO photolysis
 96    ZRJB=PRJ(JL,JBCH2O) !L2: HCHO photolysis
 97    ZRR_1=PRR(JL,KC76)* MAX(PY(JL,IOH),1.E-30_JPRB) ! R2.1: ISOP+OH reaction rate
 98    ZRR_5= 1.65E-11_JPRB * MAX(PY(JL,IHO2),1.E-30_JPRB) ! R2.2: ISOPO2 + HO2 reaction rate
 99    ZRR_6= 8.70E-12_JPRB * MAX(PY(JL,INO),1.E-30_JPRB) ! R2.3: ISOPO2 + NO reaction rate
100    ZRR_2= 5.57E-12_JPRB * MAX(PY(JL,IOH),1.E-30_JPRB)  ! R2.4: IOX+OH reaction rate
101    ZRR_4 = PRR(JL,KCH4OH) * PY(JL,ICH4) * MAX(PY(JL,IOH),1.E-30_JPRB) ! R2.5: CH4+OH reaction rate
102
103    ! CHEMISTRY
104    ! ----------
105
106    ! IISOPO2 production and loss coefficients:
107    ZPISOPO2= PY(JL, IISOP_TLAD)*ZRR_1 !R2.1
108    ZLISOPO2= ZRR_5 + ZRR_6 !R2.2/2.3
109
110    ! IOX production and loss coefficients:
111    ZPOX= ZBETA * PY(JL, IISOPO2_TLAD)*ZRR_5 + ZDELTA * PY(JL, IISOPO2_TLAD)*ZRR_6
112    ZLOX= ZRR_2
113
114    ! HCHO production and loss coefficients:
115    ZPHCHO= ZALPHA * PY(JL, IISOPO2_TLAD)*ZRR_5 &
116        & + ZGAMMA * PY(JL, IISOPO2_TLAD)*ZRR_6 &
117        & + ZEPSILON * PY(JL, IIOX_TLAD)*ZRR_2 &
118        & + ZRR_4
119    ZLHCHO= ZRJA + ZRJB + ZRR_3
120
121    ! HCHO production and loss:
122    PY(JL, IHCHO_TLAD)=(PY(JL, IHCHO_TLAD) + ZPHCHO*PDT)/(1._JPRB+ZLHCHO*PDT)
123
124    ! IOX production and loss:
125    PY(JL, IIOX_TLAD)= (PY(JL, IIOX_TLAD)+ZPOX*PDT)/(1._JPRB+ZLOX*PDT)
126
127    ! ISOPO2 production and loss:
128    PY(JL, IISOPO2_TLAD)= (PY(JL, IISOPO2_TLAD)+ZPISOPO2*PDT)/(1._JPRB+ZLISOPO2*PDT)
129
130    ! Isoprene production and loss:
131    PY(JL, IISOP_TLAD)=1._JPRB/(1._JPRB+ZRR_1*PDT) * PY(JL, IISOP_TLAD)
132
133  ENDDO !JL
134
135  IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_HCHO_CHEM',1,ZHOOK_HANDLE)
136  END SUBROUTINE TM5 SIMPLE HCHO CHEM
```

Fig. A.2: Simplified HCHO chemistry scheme for variable NOx conditions (SC2).

11 Appendix B. Tangent linear code

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

```
SUBROUTINE TM5_SIMPLE_CH2O_CHEM_TL(JK,KLEV,KIDIA,KFDIA,KLON,RDT,RRR,RRJ,PY5,PY)

!**   DESCRIPTION
!     ----------
!
!   Part of TM5 routines for IFS chemistry:
!-----------------------------------------------------------------------
!   Eulerian backward Iteration
!   Chemistry solver for simplified CH2O test chemistry
!-----------------------------------------------------------------------
!
!
!
!**   INTERFACE.
!     ----------
!          *TM5_SIMPLE_CH2O_CHEM_TL* IS CALLED FROM *CHEM_TM5_TL*.

! INPUTS:
! -------
! KIDIA :  Start of Array
! KFDIA :  End  of Array
! KLON  :  Length of Arrays
! RDT   :  Time step in seconds
! RRR  (KLON,NREAC)      : reaction rates
! RRJ  (KLON,NPHOTO)     : photolysis rates
!
!
! OUTPUTS:
! -------
! PY (KLON,NCHEM_TLAD)        : final   volume ratios OF TRACERS          (mol/mol)
!
! LOCAL:
! -------
!
!
!     AUTHOR.
!     -------
!        Flora Kluge        *ECMWF*
!        VINCENT HUIJNEN    *KNMI*
!
```

```
!     MODIFICATIONS.
!     --------------
!        ORIGINAL : 2024-11-01



USE PARKIND1 , ONLY : JPIM     ,JPRB
USE YOMHOOK   ,ONLY : LHOOK, DR_HOOK, JPHOOK

USE ALGORITHM_STATE_MOD, ONLY : GET_NITER4D
USE YOMLUN   , ONLY : NULOUT
USE YOMCT3, ONLY : NSTEP

!* reaction rates and tracer indices
USE YOMTM5TLAD , ONLY :  NCHEM_TLAD, IC5H8, IISPD, ICH2O, IOH, ICH4, &
     &  KC76, KOHIOX, KFRMOH, KCH4OH, JACH2O, JBCH2O, &
     &  NREAC, NPHOTO

IMPLICIT NONE

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

! non local
INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON,JK, KLEV

!non local constants
REAL(KIND=JPRB),INTENT(IN)    :: RDT
REAL(KIND=JPRB),INTENT(IN)    :: RRR(KLON,NREAC)
REAL(KIND=JPRB),INTENT(IN)    :: RRJ(KLON,NPHOTO)
!non local variables
REAL(KIND=JPRB),INTENT(INOUT) :: PY5(KLON,NCHEM_TLAD)      ! tracer concentrations (trajectory)
REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,NCHEM_TLAD)       ! tracer concentrations (increment)

! * LOCAL
REAL(KIND=JPHOOK)    :: ZHOOK_HANDLE
INTEGER(KIND=JPIM)    :: ITER
!local constants
REAL (KIND=JPRB)              :: RRJA,RRJB, RRR_1,RRR_2,RRR_3,RRR_4,RLCH2O
REAL (KIND=JPRB),PARAMETER    :: RALPHA = 0.83_JPRB
REAL (KIND=JPRB),PARAMETER    :: RBETA  = 1.17_JPRB
!local variables
REAL (KIND=JPRB)              :: ZPOX , ZPCH2O
REAL (KIND=JPRB)              :: ZPOX5, ZPCH2O5

! * counters
INTEGER(KIND=JPIM) :: JL
```

```
 89
 90 !---------------------------------------------------------------------
 91
 92 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_CH2O_CHEM_TL',0,ZHOOK_HANDLE)
 93
 94 DO JL=KIDIA,KFDIA
 95
 96
 97 ! TANGENT-LINEAR CODE
 98 !=======================
 99
100 !constants
101    RRR_1 = RRR(JL,KC76)    * MAX(PY5(JL,IOH),1.E-30_JPRB)
102    RRR_2 = 10.83E-12_JPRB * MAX(PY5(JL,IOH),1.E-30_JPRB)
103    RRR_3 = RRR(JL,KFRMOH) * MAX(PY5(JL,IOH),1.E-30_JPRB)
104    RRR_4 = RRR(JL,KCH4OH) * MAX(PY5(JL,IOH),1.E-30_JPRB) * PY5(JL,ICH4)
105    RRJA  = RRJ(JL,JACH2O)
106    RRJB  = RRJ(JL,JBCH2O)
107    RLCH2O= RRJA + RRJB + RRR_3
108
109 !variables
110    PY5(JL,IC5H8)= 1._JPRB/(1._JPRB+RRR_1*RDT) * PY5(JL,IC5H8)
111    PY(JL,IC5H8) = 1._JPRB/(1._JPRB+RRR_1*RDT) * PY(JL,IC5H8)
112
113    ZPOX5 = PY5(JL,IC5H8) * RRR_1
114    ZPOX  = PY(JL,IC5H8) * RRR_1
115
116    PY5(JL,IISPD) = (PY5(JL,IISPD)+ZPOX5*RDT)/(1._JPRB+RRR_2*RDT)
117    PY(JL,IISPD)  = (PY(JL,IISPD) +ZPOX *RDT)/(1._JPRB+RRR_2*RDT)
118
119    ZPCH2O5 = RALPHA * PY5(JL,IC5H8) * RRR_1 &
120          & + RBETA  * PY5(JL,IISPD) * RRR_2 &
121          & + RRR_4
122    ZPCH2O =  RALPHA * PY(JL,IC5H8) * RRR_1 &
123          & + RBETA  * PY(JL,IISPD) * RRR_2
124
125    PY5(JL,ICH2O)=(PY5(JL,ICH2O)+ZPCH2O5*RDT)/(1._JPRB + RDT * RLCH2O)
126    PY(JL,ICH2O) =(PY(JL,ICH2O) +ZPCH2O *RDT)/(1._JPRB + RDT * RLCH2O)
127
128
129 ENDDO
130
131 IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_CH2O_CHEM_TL',1,ZHOOK_HANDLE)
132 END SUBROUTINE TM5_SIMPLE_CH2O_CHEM_TL
```

12 Appendix C. Adjoint code

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

```
SUBROUTINE TM5_SIMPLE_CH2O_CHEM_AD(JK,KLEV,KIDIA,KFDIA,KLON,RDT,RRR,RRJ,PY5,PY)

!**   DESCRIPTION
!     ----------
!
!   Part of TM5 routines for IFS chemistry:
!-----------------------------------------------------------------------
!   Eulerian backward Iteration
!   Chemistry solver for simplified CH2O test chemistry
!-----------------------------------------------------------------------
!
!
!
!**   INTERFACE.
!     ----------
!          *TM5_SIMPLE_CH2O_CHEM_AD* IS CALLED FROM *CHEM_TM5_AD*.

! INPUTS:
! -------
! KIDIA :  Start of Array
! KFDIA :  End  of Array
! KLON  :  Length of Arrays
! RDT   :  Time step in seconds
! RRR  (KLON,NREAC)      : reaction rates
! RRJ  (KLON,NPHOTO)     : photolysis rates
!
!
! OUTPUTS:
! -------
! PY (KLON,NCHEM_TLAD)        : final   volume ratios OF TRACERS           (mol/mol)
!
! LOCAL:
! -------
!
!
!
!     AUTHOR.
!     -------
!        Flora Kluge        *ECMWF*
!        VINCENT HUIJNEN    *KNMI*
!        TM5-community
!
!     MODIFICATIONS.
!     --------------
!        ORIGINAL : 2024-11-01


USE PARKIND1 , ONLY : JPIM     ,JPRB
USE YOMHOOK   ,ONLY : LHOOK, DR_HOOK, JPHOOK

USE ALGORITHM_STATE_MOD, ONLY : GET_NITER4D
USE YOMLUN    , ONLY : NULOUT
USE YOMCT3, ONLY : NSTEP

!* reaction rates and tracer indices
USE YOMTM5TLAD , ONLY :  NCHEM_TLAD, IC5H8, IISPD, ICH2O, IOH, ICH4, &
     & KC76, KOHIOX, KFRMOH, KCH4OH, JACH2O, JBCH2O, &
     & NREAC, NPHOTO


IMPLICIT NONE

!-----------------------------------------------------------------------
!*       0.1  ARGUMENTS
!             ---------
! non local
INTEGER(KIND=JPIM),INTENT(IN) :: KIDIA , KFDIA , KLON, JK, KLEV

!non local constants
REAL(KIND=JPRB),INTENT(IN)    :: RDT
REAL(KIND=JPRB),INTENT(IN)    :: RRR(KLON,NREAC)
REAL(KIND=JPRB),INTENT(IN)    :: RRJ(KLON,NPHOTO)
!non local variables
REAL(KIND=JPRB),INTENT(INOUT) :: PY5(KLON,NCHEM_TLAD)
REAL(KIND=JPRB),INTENT(INOUT) :: PY(KLON,NCHEM_TLAD)

! * LOCAL
REAL(KIND=JPHOOK)    :: ZHOOK HANDLE
```

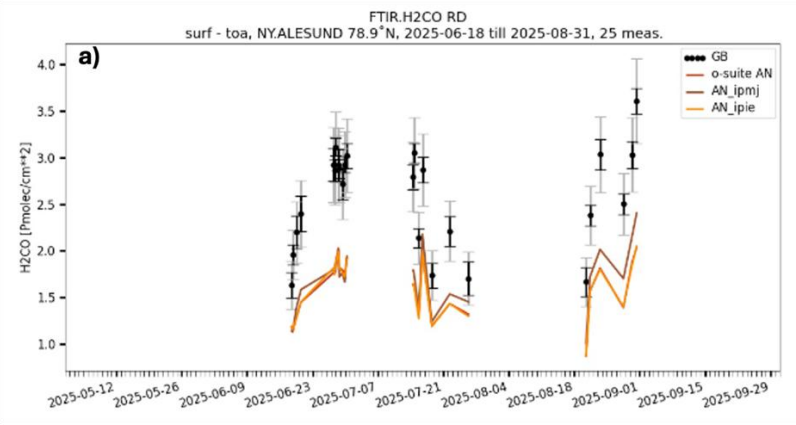
```
INTEGER(KIND=JPIM)    :: ITER
!local constants
REAL (KIND=JPRB)              :: RRJA,RRJB, RRR_1,RRR_2,RRR_3, RRR_4,RLCH2O
REAL (KIND=JPRB),PARAMETER    :: RALPHA = 0.83_JPRB
REAL (KIND=JPRB),PARAMETER    :: RBETA  = 1.17_JPRB
!local variables
REAL (KIND=JPRB)              :: ZPOX , ZPCH2O
REAL (KIND=JPRB)              :: ZPOX5, ZPCH2O5

! * counters
INTEGER(KIND=JPIM) :: JL

!-----------------------------------------------------------------------
!-----------------------------------------------------------------------
!-----------------------------------------------------------------------

IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_CH2O_CHEM_AD',0,ZHOOK_HANDLE)

DO JL=KIDIA,KFDIA

   !constants
   !-----------------------------------------------------------------------
   RRR_1 = RRR(JL,KC76)   * MAX(PY5(JL,IOH),1.E-30_JPRB)
   RRR_2 = 10.83E-12_JPRB * MAX(PY5(JL,IOH),1.E-30_JPRB)
   RRR_3 = RRR(JL,KFRMOH) * MAX(PY5(JL,IOH),1.E-30_JPRB)
   RRR_4 = RRR(JL,KCH4OH) * MAX(PY5(JL,IOH),1.E-30_JPRB) * PY5(JL,ICH4)
   RRJA  = RRJ(JL,JACH2O)
   RRJB  = RRJ(JL,JBCH2O)
   RLCH2O= RRJA + RRJB + RRR_3

   !forward trajectories
   !-----------------------------------------------------------------------
   PY5(JL,IC5H8) = 1._JPRB/(1._JPRB+RRR_1*RDT) * PY5(JL,IC5H8)

   ZPOX5 = PY5(JL,IC5H8) * RRR_1

   PY5(JL,IISPD) = (PY5(JL,IISPD)+ZPOX5*RDT)/(1._JPRB+RRR_2*RDT)

   ZPCH2O5 = RALPHA * PY5(JL,IC5H8) *RRR_1 &
         & + RBETA  * PY5(JL,IISPD) *RRR_2 &
         & + RRR_4

   PY5(JL,ICH2O)=(PY5(JL,ICH2O)+ZPCH2O5*RDT)/(1._JPRB + RDT * RLCH2O)

   ! adjoint
   !-----------------------------------------------------------------------
   ZPOX            = 0.0_JPRB

   PY(JL,ICH2O) =  PY(JL,ICH2O) / (1._JPRB + RDT * RLCH2O)
   ZPCH2O = 0.0_JPRB
   ZPCH2O = ZPCH2O + PY(JL,ICH2O) * RDT / (1._JPRB + RDT * RLCH2O)

   PY(JL,IC5H8) = PY(JL,IC5H8) + RALPHA * RRR_1 * ZPCH2O
   PY(JL,IISPD) = PY(JL,IISPD) + RBETA  * RRR_2 * ZPCH2O
   ZPCH2O       = 0.0_JPRB

   PY(JL,IISPD)= PY(JL,IISPD) * (1._JPRB/(1._JPRB + RDT * RRR_2))
   ZPOX        = ZPOX + RDT / (1._JPRB + RDT * RRR_2) * PY(JL,IISPD)

   PY(JL,IC5H8)=PY(JL,IC5H8) + RRR_1 * ZPOX
   ZPOX = 0.0_JPRB

   PY(JL,IC5H8)=PY(JL,IC5H8) * (1._JPRB/(1._JPRB+RDT*RRR_1))

ENDDO

IF (LHOOK) CALL DR_HOOK('TM5_SIMPLE_CH2O_CHEM_AD',1,ZHOOK_HANDLE)
END SUBROUTINE TM5_SIMPLE_CH2O_CHEM_AD
```

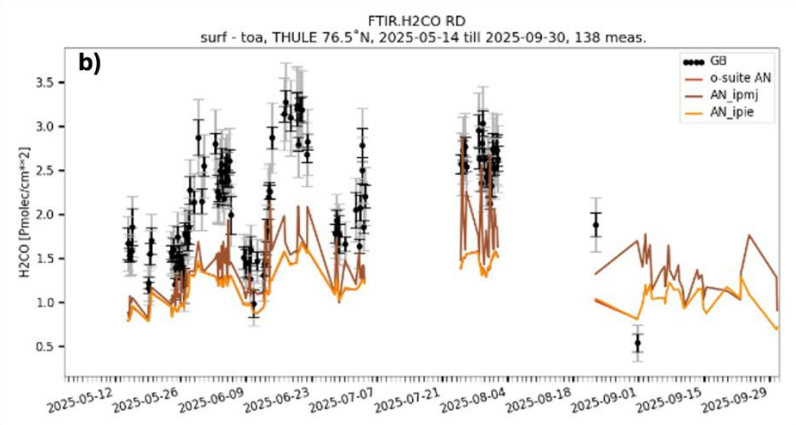
13 Appendix D. Configuration file for isoprene emission inversion

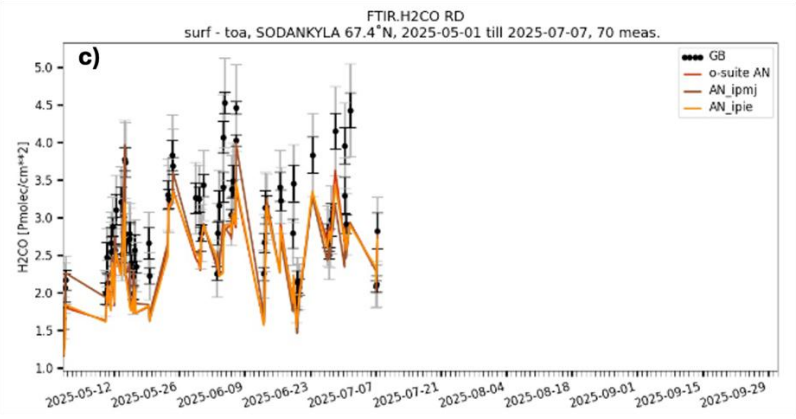
NAME COMMENT SPECIES GRIBCODE SECTOR_1 SECTOR_2 SECTOR_3 SECTOR_4 SECTOR_5 SECTOR_6 SECTOR_7 SECTOR_8 SECTOR_9 SECTOR_10 SECTOR_11 SECTOR_12 SECTOR_13 SECTOR_14 SECTOR_15 SECTOR_16 SECTOR_17 SECTOR_18 SECTOR_19 SECTOR_20 STD_ERR S_COR_LENGTH_SCALE_m T_COR_LENGTH_SCALE_sec LLOGNORM

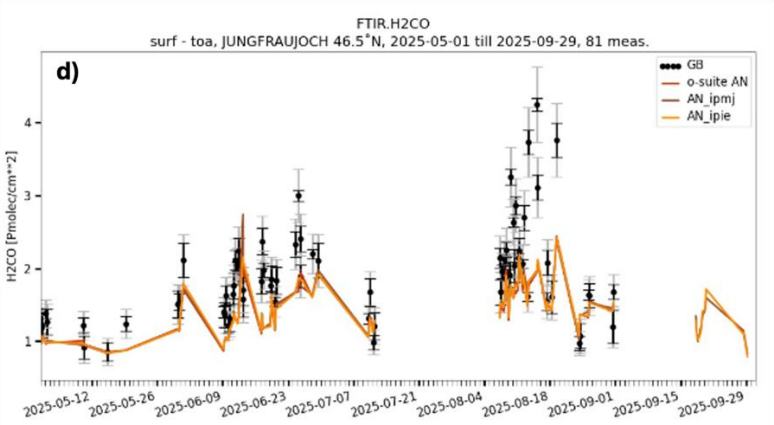
isoprene ISOP_emis C5H8 95 awb ene fef res ref com shp slv swd tnr tro none none none none none none none none none 0.5 100000. 86400. True

14 Appendix E. Supplementary figures and analysis









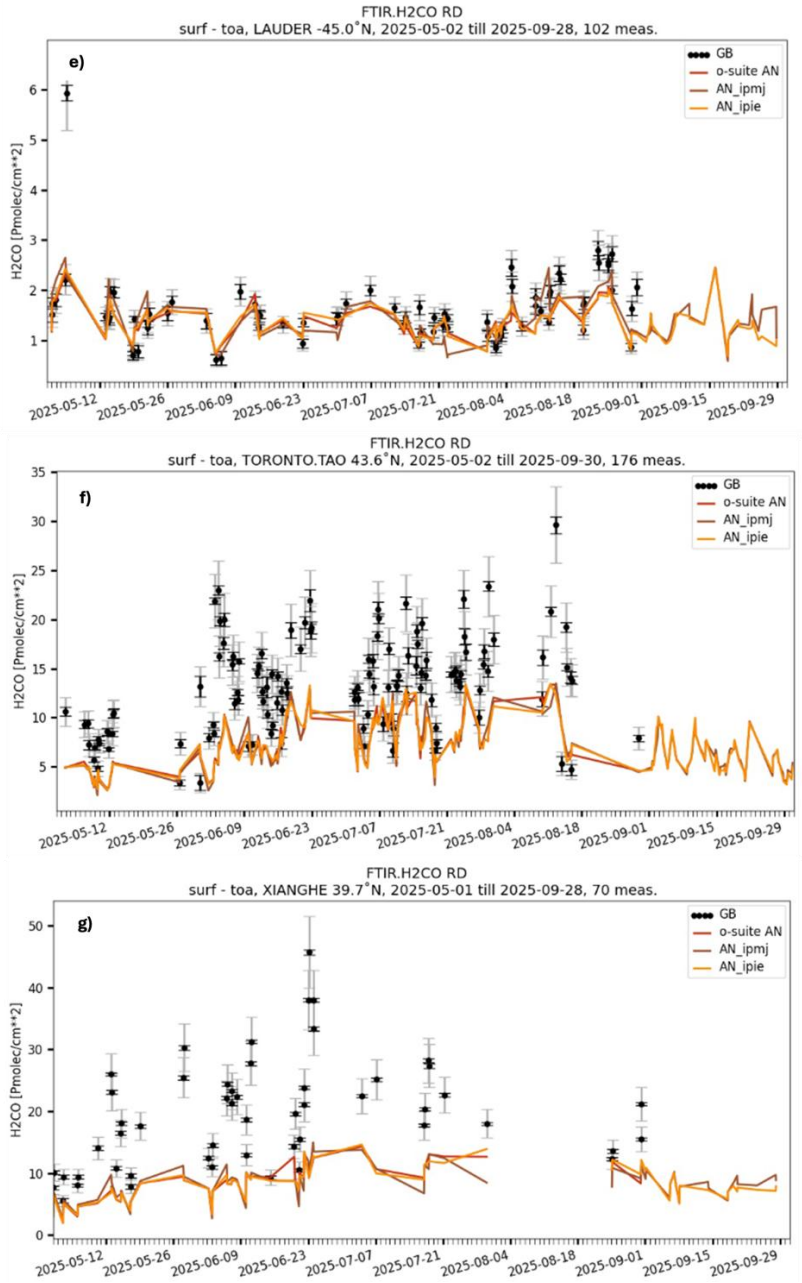


Fig. E1: Timeseries of HCHO VCDs over Ny-Alesund (panel a), Thule (panel b), Sodankyla (panel c), Jungfraujoch (panel d), Lauder (panel e), Toronto (panel f), and Xianghe (panel f) with observations (black), IFS-COMPO reference run (yellow), and HCHO assimilation (brown). The data are averaged in two-week intervals ranging from May 2025 to September 2025. Additionally plotted is the current CAMS o-suite, that mainly corresponds to the reference model setup. Plots curtesy by Bavo Langerock (NDACC,BIRA, CAMS-82-bis) with data support by the CAMS2-27 project

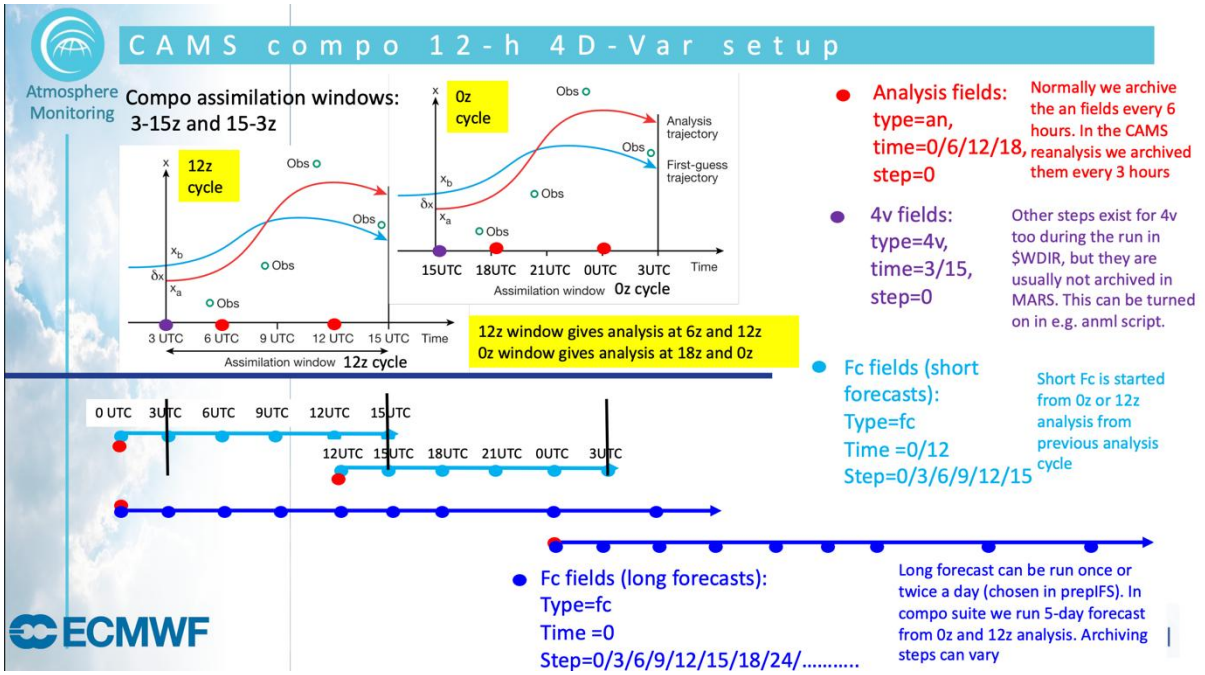


Fig. E2: Summarized overview on data assimilation cycling in CAMS COMPO 12-h 4D-Var configuration. Figure provided by Antje Inness.

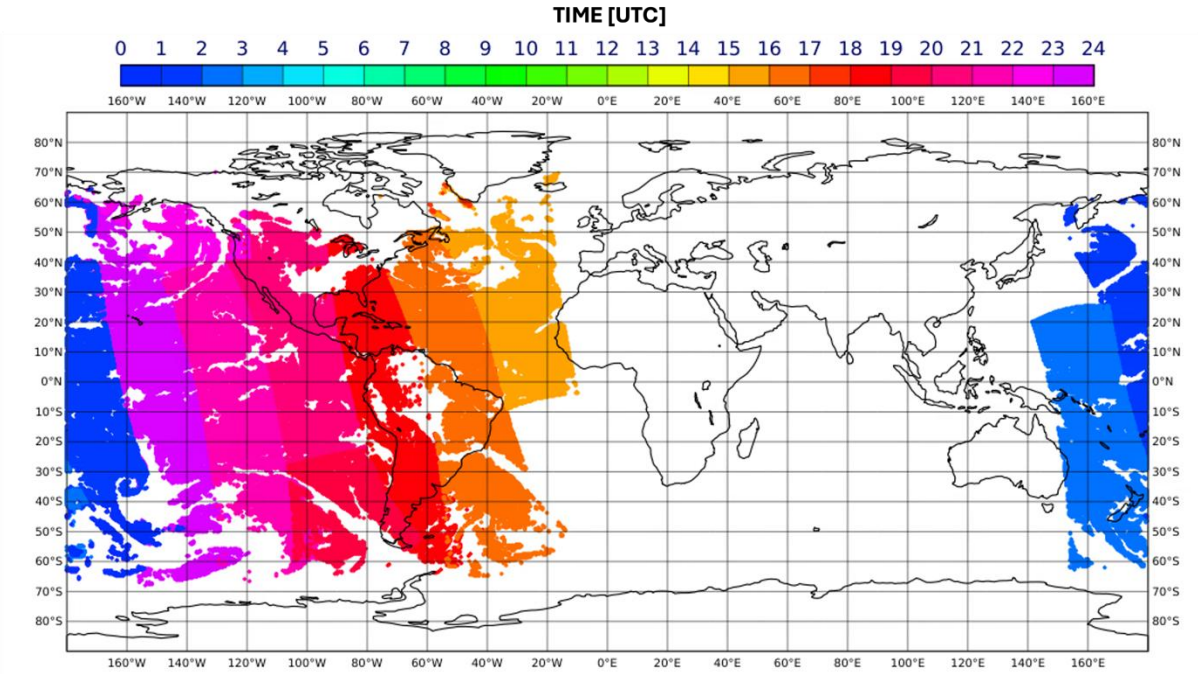


Fig. E3.: TROPOMI S5P orbits between 15UTC and 3UTC on 06/04/2025, that are included in the 00UTC IFS-COMPO analysis cycle color-coded as a function of time.

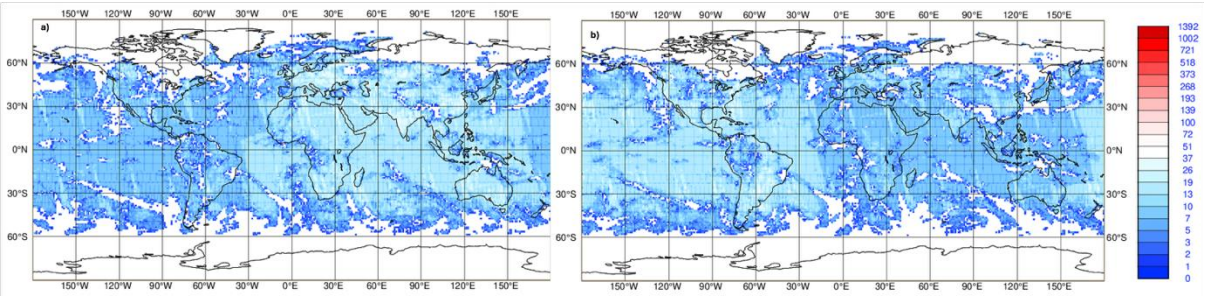


Fig. E4.: Global distribution of TROPOMI HCHO observations used in the 12UTC analysis (01/05/2025 3UTC to 01/05/2025 15UTC, panel a) and 00UTC analysis (30/04/2025, 15UTC to 01/05/2025, 3UTC).

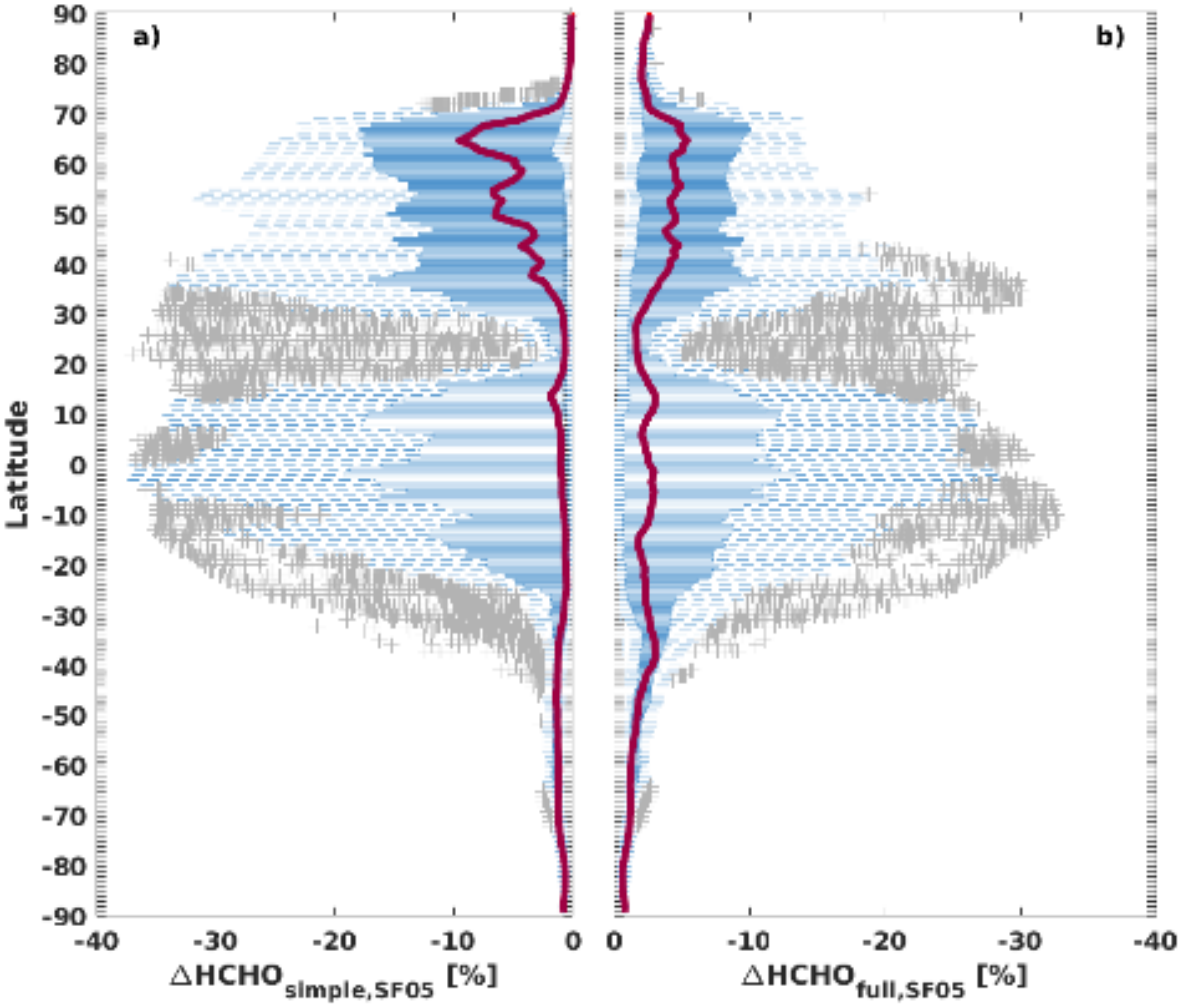


Fig. E5: Latitudinal dependency of the relative change in formaldehyde due to a decrease in isoprene emissions of 50% for simplified (panel a) and standard IFS-COMPO configuration (panel b) for July 2023.

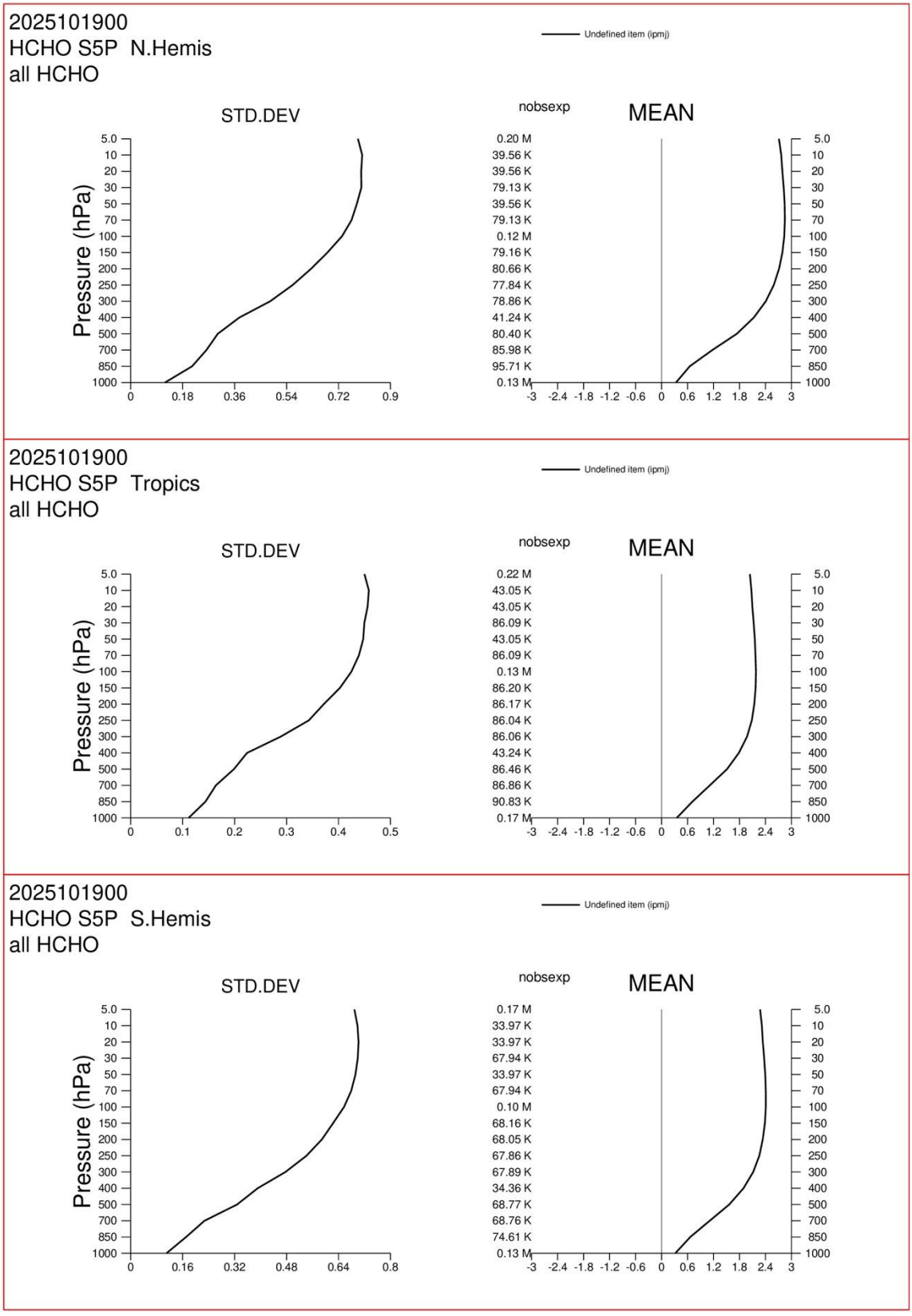


Fig. E6: Exemplary averaging kernels for the TROPOMI HCHO retrievals in different global regions on 19/10/2025.

Document History

Version	Author(s)	Date	Changes
0.1	Flora Kluge	15.11.2025	Initial version
1.0	Flora Kluge	26.11.2025	Version after internal review

Internal Review History

Internal Reviewers	Date	Comments
Jean-Francois Muller (BIRA-IASB), Yves-Marie Saint Drenan (Armines)	Nov 2025	