# Proof-of-concept for multi-scale global IFS prototype

Nicolas Bousserez (ECMWF), Luca Cantarello (ECMWF)

coco2-project.eu

# D5.1 Proof-of-concept for multi-scale global IFS prototype

| | |
|---|---|
| **Dissemination Level**: | Public |
| **Author(s)**: | Nicolas Bousserez (ECMWF) and Luca Cantarello (ECMWF) |
| **Date**: | 26/12/2023 |
| **Version**: | 1.0 |
| **Contractual Delivery Date**: | 31/12/2023 |
| **Work Package/ Task**: | WP5/ T5.1 |
| **Document Owner**: | ECMWF |
| **Contributors**: | |
| **Status**: | Draft |
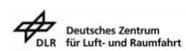
# CoCO2: Prototype system for a Copernicus CO₂ service

**Coordination and Support Action (CSA)
H2020-IBA-SPACE-CHE2-2019Copernicus evolution –
Research activities in support of a European operational
monitoring support capacity for fossil CO2 emissions**

**Project Coordinator:** Dr Richard Engelen (ECMWF)
**Project Start Date:** 01/01/2021
**Project Duration:** 36 months

**Published by the CoCO2 Consortium**

**Contact:**
ECMWF, Shinfield Park, Reading, RG2 9AX,
richard.engelen@ecmwf.int

# Table of Contents

# 1 Introduction

The European Commission is in the process of developing a $CO_2$ Monitoring and Verification Support (CO2MVS) capacity. This initiative will build a new component for its existing Copernicus Earth Observation programme. The core objective of the CO2MVS is to monitor and verify anthropogenic (human-caused) carbon dioxide and methane emissions leveraging all available remote-sensing and in situ observations as well as cutting-edge modelling, from the local to the global scale. The resulting improvement in the accuracy and reliability of emissions data will in turn help support policy makers in the design of efficient climate mitigation strategies in the context of the Paris agreement.

Demonstrator systems for this CO2MVS are being developed in the prototype system for a Copernicus $CO_2$ service (CoCO2) project. The multi-scale aspect of the approach, from local to global and daily to yearly emission estimates, is essential to provide both actionable information in near-real time as well as a longer-term trend analysis of the carbon budget to verify the impact of the Paris agreement implementation. The core of the global component of the inversion system will be based on the Integrated Forecast System (IFS) developed at the European Centre for Medium-range Weather Forecast (ECMWF). Adapting such a system, initially conceived for numerical weather forecasting, to the need of a global greenhouse gas (GHG) inversion system, presents several challenges. In particular, three important aspects to build a global multi-scale inversion prototype into the IFS are:

1) The extension of the current operational short-window (12-hour) assimilation to a longer window in order to accommodate the inversion of long-lived greenhouse gas emissions.
2) The integration of posterior emission products provided at various scales (from point source to global) from different systems into the global IFS system.
3) The seamless integration of the CO2MVS prototype into the current operational Copernicus Atmospheric Monitoring System (CAMS), ensuring both methodological robustness and computational efficiency for timely delivery of products.

In this report we describe the methodological approaches taken to address the challenges outlined above and provide a first proof-of-concept of an hybrid ensemble-variational inversion system based on the IFS. This system will be further developed and tested in the coming years at ECMWF, in particular in the framework of CORSO, in preparation for its operational implementation in 2026.

# 2 Methodological approach

## 2.1 Hybrid ensemble-variational algorithm

The current online 4D-Var IFS assimilation system uses a short 12-hour assimilation window, which, due to strong nonlinearities in the weather model, cannot be extended to longer periods without compromising computational efficiency and/or convergence properties. On the other hand, most offline global GHG inversion systems rely on long assimilation time-windows to take into account the long lifetime of GHG species and optimise observational constraints. Therefore, during the CoCO2 project, efforts have focused on developing a hybrid ensemble-variational method that extends the current short-window assimilation capabilities while maintaining computational efficiency. The latter will be critical to ensure timely delivery of the optimised emissions in an operational context. The approach consists in implementing a Kalman smoother algorithm wherein the initial state increment from the 4D-Var minimisation is propagated backward in time to update previous emissions using ensemble-based

covariances. In the first version of the inversion prototype, the emission-state covariance approximations are obtained using a modified version of the existing Ensemble of Data Assimilation (EDA) framework (see Fig. 1).
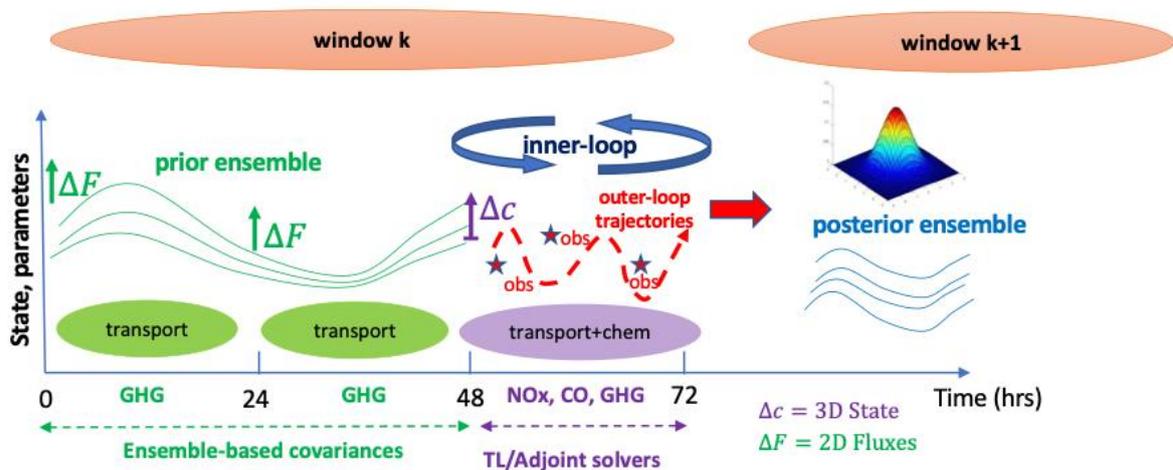


*Figure 1: Schematic of the hybrid ensemble-variational inversion system. At each cycle k, the 4D-Var state increment for CH4 and CO2 ($\Delta c$) is propagated backward in time using the EDA ensemble to update the prior fluxes ($\Delta F$) in previous assimilation windows. The posterior EDA state ensemble is propagated forward in time to the beginning of the next assimilation window (k+1) and the algorithm is repeated.*

In the EDA system for GHG species developed at ECMWF, multiple 4D-Var minimisations are run in parallel in each 12-hour window to generate an ensemble of jointly optimised meteorological states, atmospheric concentrations, and surface emissions. For each species the estimate of the emissions is achieved by optimising a 2D scaling factor field which is applied to the surface fluxes. Perturbations are applied to multiple elements of the system to account for various sources of uncertainty and their propagation in time: observations, physical tendencies (a proxy for model error), sea surface temperatures, and prior emissions.

Uncorrelated multiplicative perturbations to the prior emissions sampled from a log-normal distribution are applied within each assimilation window. The median and standard deviation of the distribution used in the simulations reported in this deliverable have been set to 1 and 0.2, respectively.

A hybrid background covariance matrix is used in the EDA to account for the flow-dependent component of the model uncertainty provided by the ensemble, which is given a weight of 30% in our experiments.

Fig. 2 below illustrates the structure of the EDA system.
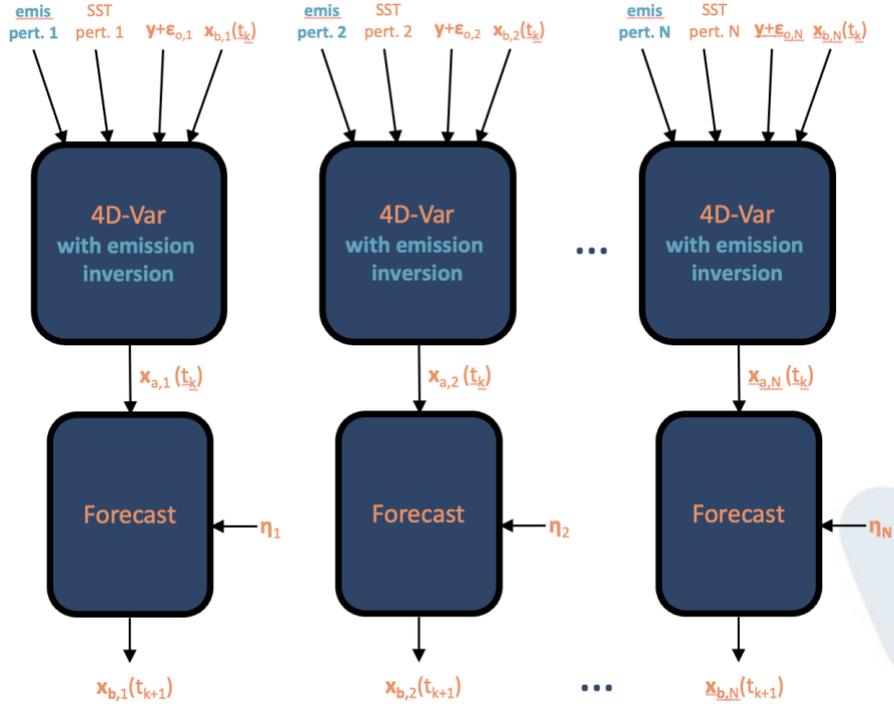


*Figure 2: Schematic representation of the EDA approach extended to the emission inversion system. Multiple 4D-Var minimisations are performed in parallel in which the observations, the sea surface temperature, the initial emissions, and the model physical tendencies are perturbed (the latter during the forecast integration). The surface fluxes optimisation is performed within each 4D-Var though the optimisation of a 2D scaling factor fields applied to the initial emissions.*

This adapted EDA system forms the basis of the hybrid ensemble-variational algorithm, in which the (prior) EDA ensemble is used to construct an extended 4D covariance matrix between the fluxes in previous window and the initial state of the current 4D-Var window. The algorithm consists of the following steps:

1. Retrieve the short-window 4D-Var state increment ($\delta C_a$):

$$\delta C_a = \mathrm{B} \mathrm{H}^{\mathrm{T}} (\mathrm{H} \mathrm{B} \mathrm{H}^{\mathrm{T}} + \mathrm{R})^{-1} (y - h(C_b)) \tag{1}$$

2. Normalize the increment using an approximation of the **B** matrix ($\widetilde{\mathbf{B}}$):

$$\delta C_{norm} = \widetilde{\mathrm{B}}^{-1} \mathrm{B} \mathrm{H}^{\mathrm{T}} (\mathrm{H} \mathrm{B} \mathrm{H}^{\mathrm{T}} + \mathrm{R})^{-1} (y - h(C_b)) \tag{2}$$

3. Propagate the normalized increment ($\delta C_{norm}$) using the EDA-based flux-state covariance:

$$\delta F_a = \frac{1}{N-1} (F_b - \overline{F_b})(C_b - \overline{C_b})^T \delta C_{norm}, \tag{3}$$

where **F**$_b$ represents the matrix of column vectors of the *n* ensemble members of prior fluxes in all previous windows and **C**$_b$ the matrix of column vectors of the *n* prior state concentrations at initial time $t_0$ of the current 4D-Var window. The operator **H** represents the full linear model (tangent-linear) including the observational operator, while **R** represents the observational error covariance matrix. In the current implementation the approximation $\widetilde{\mathbf{B}}$ and its inverse is obtained from the prior ensemble **C**$_b$. In an upcoming version $\widetilde{\mathrm{B}}^{-1}$ will consists of the inverse

of the wavelet operator, to ensure consistency with the $B$ model used in the 4D-Var minimization.

## 2.2 Assimilation of multi-scale posterior emission products

The spatial resolution of the IFS global inversion is currently limited to about ~40 km, which is the resolution at which the tangent-linear model is run during the 4D-Var minimisation. Previous studies carried out as part of the CoCO2 project (see, e.g., D4.1) have highlighted the need for high-resolution modelling to accurately represent plumes originated from point source emissions as well as for regional limited area inversions. Before the IFS can reach spatial resolutions that meet those requirements (~5 km), it is therefore desirable to adopt an alternative solution to integrate higher resolution transport information into the global inversion system. The proposed approach relies on assimilating posterior emission products derived from high-resolution regional inversion systems and lightweight local plume inversion approaches as observations into the global IFS system. Such methodology requires two steps:

1. Estimate of the averaging kernel matrix and retrieval errors associated with the emission retrieval method.
2. Assimilation of the retrieved emissions into the global IFS system.

The following two sections describe the theoretical framework of the approach.

## 2.2.1 Approximation of averaging kernels

The approach chosen for 1. is to approximate the averaging kernel matrix using an ensemble of perturbed inversions. Formally, the emission retrieval can be written:

$$R(x, \varepsilon) = x_b + KH\,(x - x_b) + K\varepsilon, \tag{4}$$

where **K** is Kalman gain matrix, **H** the linear observational operator, **x** and **x**$_b$ the true and prior emissions, respectively, and $\varepsilon$ the observational errors.

It can be rewritten:

$$R\begin{pmatrix} x \\ \varepsilon \end{pmatrix} = \begin{pmatrix} x_b \\ 0 \end{pmatrix} + Q\left[\begin{pmatrix} x \\ \varepsilon \end{pmatrix} - \begin{pmatrix} x_b \\ 0 \end{pmatrix}\right], \tag{5}$$

with:

$$Q = \begin{bmatrix} KH & K \end{bmatrix} \tag{6}$$

With this notations, one can write the observational operator ($h$) and the covariance matrix of retrieval errors (**E**) as follows:

$$h(x) = x_b + KH\,(x - x_b) \tag{7}$$

$$E = \overline{(K\,\varepsilon)(K\,\varepsilon)^T},$$

where $^{-}$ represents the average operator. In order to compute **Q**, let us consider $n$ perturbed prior and observation (**y**):

$$x_i = x_b + \delta x_i \tag{8}$$

$$y_i = y + \varepsilon_i, i = 1, \dots, n$$

One has:

$$Q\begin{pmatrix} \delta x_i \\ \varepsilon_i \end{pmatrix} = R\begin{pmatrix} x_b + \delta x_i \\ \varepsilon_i \end{pmatrix} - \begin{pmatrix} x_b \\ 0 \end{pmatrix} \tag{9}$$

$$\Delta U = \begin{pmatrix} \delta x_i \\ \varepsilon_i \end{pmatrix}_{i=1,\dots,n} \tag{10}$$

$$\Delta Z = Q \Delta U \tag{11}$$

$$Q \approx \Delta Z (\Delta U)^+, \tag{12}$$

where $^+$ represents the pseudo-inverse operator. Therefore, the ensemble-based approximation of **Q** is a least-square estimation which uses a Monte-Carlo ensemble of pseudo-inversions.

Several types of inversion products need to be distinguished:

1. **Ensemble Kalman Filters**: the required ensemble information is a by-product of the system.
2. **Variational inversion systems**: the ensemble can be produced based on an EDA approach.
3. **Analytical inversion systems**: the matrices **Q** (i.e., **K** and **H)** can be provided directly.
4. **Non-Bayesian local plume inversions**: the averaging kernel matrix (**KH**) used is a Dirac function. The error matrix **E** is obtained using uncertainty estimation techniques tailored to the local inversion method.

Figure 3 illustrates the multi-scale approach adopted to integrate information from regional and local inversion products into the global IFS inversion system.
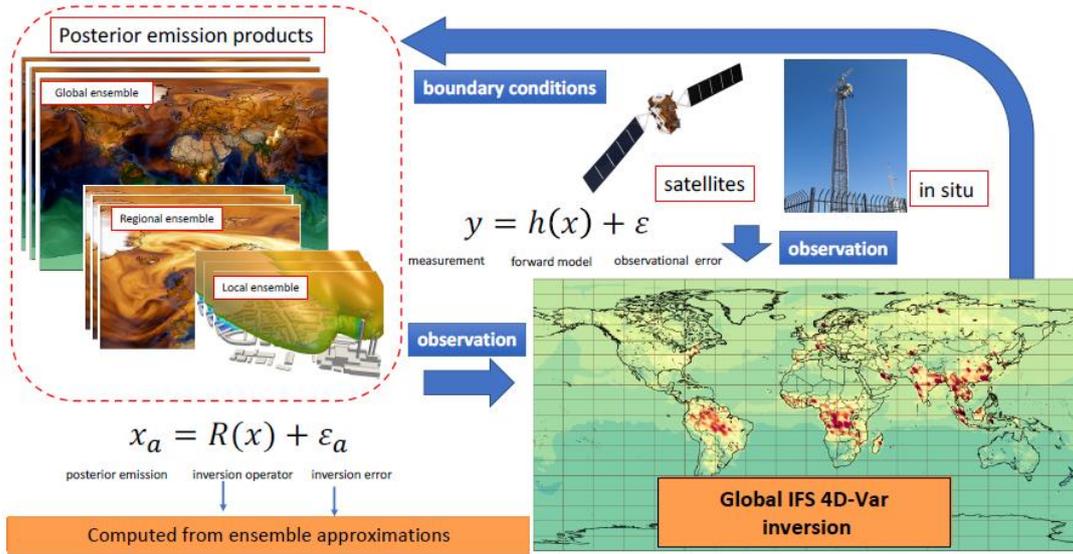
*Figure 3: Schematic of the multi-scale approach for the integrated global IFS inversion system. The IFS global inversion system provides boundary conditions to the regional and local inversion systems. The high-resolution posterior emissions obtained from those systems are in turn assimilated as observations into the global IFS model, enabling a two-way propagation of information.*

## 2.2.2 Assimilation of retrieved emissions into the global IFS

The equation to assimilate an external regional/local inversion product as observation into the IFS is as follows:

$$z_a = z_b + \frac{1}{N-1} Z_b (AZ_b)^T \left( \frac{1}{N-1} AZ_b(AZ_b)^T + E \right)^{-1} (x_{ret} - Az_b), \qquad (13)$$

where $z_b$ represents the prior fluxes of the global IFS model (mean of the ensemble distribution), $Z_b$ is the matrix of column vectors associated with each ensemble member departure from the mean, $x_{ret}$ is the vector of retrieved regional/local emissions, $A$=$KH$ is the averaging kernel matrix of the regional/local emission retrieval method computed following the approach described in 2.2.1, and $E$ is the matrix of retrieval errors. In this equation, the ensemble $Z_b$ is obtained from the posterior 4D IFS ensemble generated by the global hybrid ensemble-variational inversion system described in Sec. 2.1. Therefore, the assimilation of external posterior emission products into the global inversion system can be performed at any time during or after the long-window assimilation procedure, making this approach flexible for fast and continuous uptake of new high-resolution posterior emission estimates.

# 3   Practical integration into existing IFS framework

## 3.1 Long-window inversion tasks

This section describes the modifications of the IFS ecflow workflow that were required to implement the long-window inversion algorithm.

A new family (long_window) has been added to the 'main' family and is executed after the 'an' family is complete. The long_window family contains two tasks:

- prior_ens_data: this task collects and stores locally the model fields needed for the long-window inversion, namely: (1) the prior (type=fc, file conc_prior) and posterior

(type=4v, file conc_post) 3D concentrations from each ensemble member valid at the initial time of the current 4D-Var assimilation window; (2) the optimised scaling factors (type=4v, file sf_prior) archived in each of the previous windows; (3) the scaling factor optimised in the 4D-Var minimisations of the current cycle (type=4v, file sf_prior_last). A python script named 'sf_rescale.py' adds the initial emission perturbations to the sf_prior_last scaling factors and save them to the FDB.

- emis_an: this task performs the long-window inversion using conc_prior, conc_post and sf_prior as inputs and following the method outlined in Sec. 2.1. The output of the long window is named sf_post and has the same structure (dates, ensemble members and levels) as sf_prior. The fields in sf_prior_last are appended to sf_post at the end of the long-window inversion.

Figure 4 below shows the long-window inversion tasks on ecflow_ui.
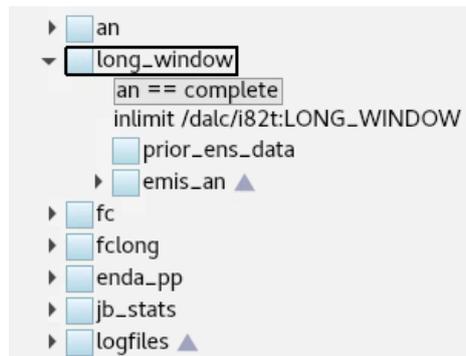


*Figure 4: Ecflow chart showing the new long-window inversion tasks inserted in the standard 4D-Var algorithm.*

Furthermore, two tasks have been modified in the lag family inside the 'archive' and 'archive_fields' families to (re-)archive the scaling factors optimised by the long-window inversion:

- ecvanml: in this task, the scaling factors in sf_post are re-archived, hence overwriting the previous scaling factors for each of the past cycle.
- longfc/sfc: in this task, the scaling factors of all previous windows archived by ecvanml are retrieved together with the prior (unperturbed) emissions and subsequently multiplied together to compute and archive the posterior emissions.

## 3.2 Offline posterior emissions assimilation

The assimilation of external posterior emission products into the global IFS system is performed offline and uses the ensemble generated from the hybrid long-window ensemble-variational algorithm, as described in Sections 2.1 and 2.2.2. The implementation of the method is done in python and consists of the following steps:

1. For each new external emission product to be assimilated, define a period centred around the date of that product. The range of that period should define the temporal localisation of the impact of the assimilation using some objective criteria.
2. Retrieve the IFS posterior emission ensemble from MARS for the period selected in 1.
3. Assimilate the external posterior emission product into the global IFS ensemble using the EnKS method formulated in Sec. 2.2.2.
4. Archive the updated IFS posterior ensemble on MARS.

From an operational point of view, the assimilation of external posterior emission products can be envisaged in a reanalysis or in a near real-time mode. In a reanalysis mode, an ensemble

of local and regional posterior emissions are assimilated all at once after a long-window (e.g., yearly) IFS global inversion has been performed. In a near-real time mode, the assimilation of external posterior emission products is carried out continuously whenever those products become available.

# 4  Preliminary demonstration

In the two sections below we provide preliminary examples of application of the long-window inversion and posterior emissions assimilation approaches described in Sec. 2.1 and 2.2, respectively.

## 4.1  Long-window inversion

A first version of the long-window inversion prototype, whose algorithm is described in Section 2.1, has been implemented in python and tested for a short time period. Due to the I/O intensive tasks of retrieving, reading and archiving large ensemble of files from the EDA outputs, as well as the computational cost of applying covariance-vector products for numerous assimilation time-windows and high-dimensional vectors, the current version of the code is limited to a week-long assimilation period. Ongoing efforts focus on optimising the algorithm implementation by mitigating these costs through parallelisation of independent covariance computations and splitting of large emission output files into multiple ones.

A preliminary demonstration of the technical feasibility of the long-window inversion implementation has been carried out for methane emissions. Figure 5 shows the state increments, dCa (see Eq. 1), and corresponding emission scaling factor increments, dFa (see Eq. 3), for several past windows for a 72-hour inversion period. With the current version of the code and for a 10-members EDA ensemble, the computational wall-time is approximately 2 days. With the upcoming IFS 49R1 cycle improvements of the EDA convergence efficiency we expect this wall-time to be reduced by about 40%, making the long-window inversion implementation more suitable for operational purposes. As described in Section 4, future algorithmic developments will help further improve the computational efficiency of the ensemble generation by using a cheaper posterior square-root sampling approach in place of the EDA (see, e.g., Bousserez et al. (2020); Bousserez and Henze (2018)).

20190630 21:00 UTC



20190701 09:00 UTC



20190702 21:00 UTC
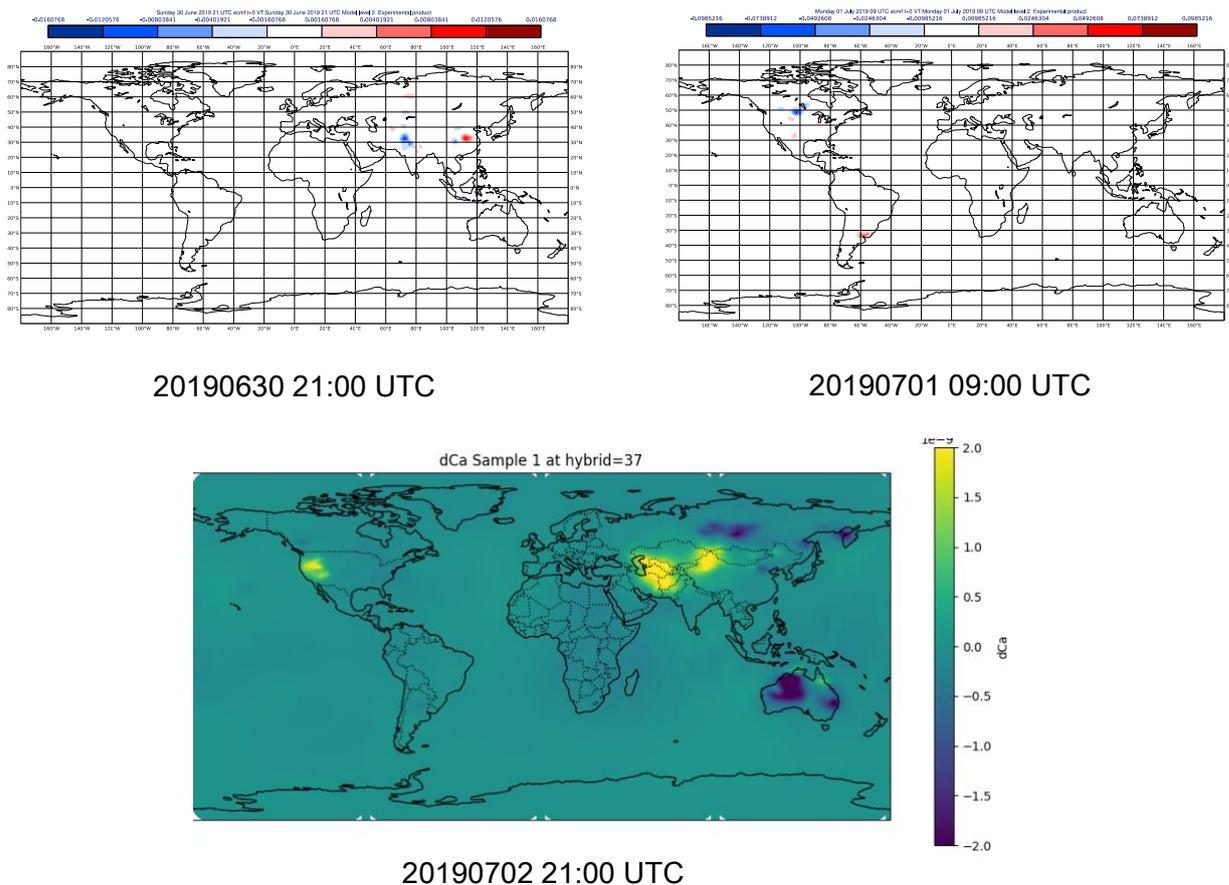
*Figure 5: Bottom: $CH_4$ state increment at the model surface level (dCa); Top: associated emission scaling factor increments (dFa) for different dates and times for a 72-hour long-window inversion.*

## 4.2 Emission hotspots assimilation

Preliminary experiments have been conducted to assimilate point source posterior emission estimates into the global IFS inversion system. These first tests involve the assimilation of emissions obtained from non-Bayesian local plume inversion techniques. Following the approach described in Section 2.2.1, we consider a Dirac function for the averaging kernel matrix of the emission retrieval, which is equivalent to assuming a direct observation of the emissions with some associated retrieval errors. The generalisation of the method to regional inversion products with spatially-resolved retrieval sensitivity matrices is under implementation.

Figure 6 shows an example of IFS global posterior scaling factor increments obtained from the assimilation of a pseudo posterior emission from the Bełchatów power-plant in Poland (Longitude: 19.33; Latitude 51.26) on 2019/07/01, 09:00UTC. The retrieved point source posterior emission (94 kt/day) and associated retrieval error (10 kt/day) were taken from an estimate on 2020/06/27 based on OCO-3 observations from Nassar et al. (2022). The date (i.e., month and day) was chosen so as to be as representative as possible to the period ran in our preliminary global inversion tests. The spatial smoothing of the increment seen in Fig. 6 is produced by the spatial correlations in the global IFS 4D ensemble produced by the long-window inversion. The EnKS-like assimilation of the estimated hotspot emissions (see Sec. 2.2.2) takes only a few second once the posterior IFS ensemble has been retrieved from MARS. This makes the approach very efficient in the context of continuous assimilation of regional and local emission products for an operational system.

Interpolated dSa Distribution on 2D Global Map
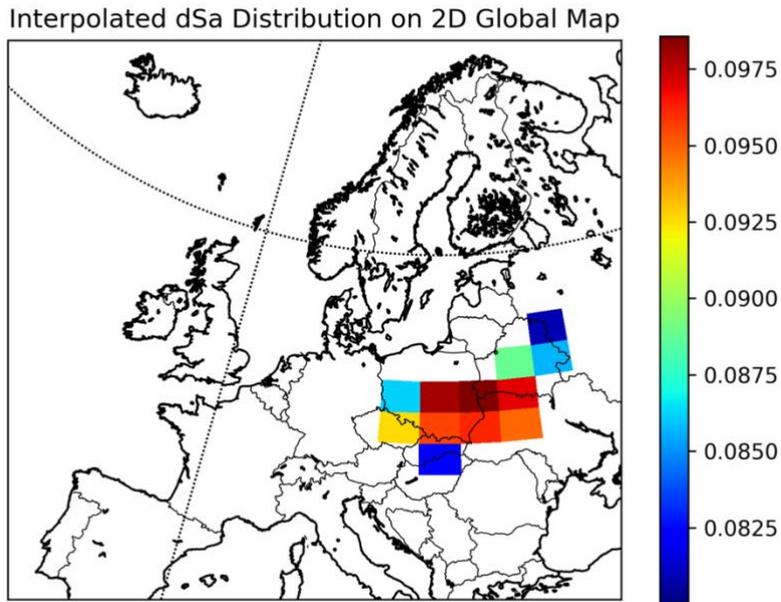


*Figure 6: Scaling factor increments obtained from assimilating a pseudo point source estimate for the Bełchatów power-plant (Poland). The spatial smoothing of the increment is produced by the spatial correlations in the global IFS 4D ensemble produced by the long-window inversion algorithm.*

# 5  Validation of prototype inversion system

An important element of the operational IFS global inversion prototype is the Evaluation and Quality Control (EQC) system. In this section we provide a few examples based on the EQC proposed in D6.7. We illustrate the approach using results from short-window NOx and $CH_4$ inversion experiments, as they were conducted earlier in the CoCO2 project than the $CO_2$ inversions and have thus undergone more extensive testing. Three methods of evaluation are presented here: comparison between inversion products (Section 6 of D6.7), comparison to independent atmospheric observations (Section 3 of D6.7) and performance of forecast initialised with posterior emissions (Section 4 of D6.7)

Figure 7 shows an evaluation based on the comparison between the forecast performance of a state-only and a joint state-emission optimisation experiment for NOx. In those experiments, the TROPOMI $NO_2$ column retrievals (PAL product) are assimilated on day *j* and the TROPOMI observations on day *j+1* are used to evaluate the predicted $NO_2$ columns. The metric considered is the Root Mean Squared Error (RMSE) of the forecasted $NO_2$ columns. The results show improvements when the emissions are jointly optimised with the initial state, with a reduction in RMSE greater than 10% and 40% over some areas for Europe and East Asia, respectively.
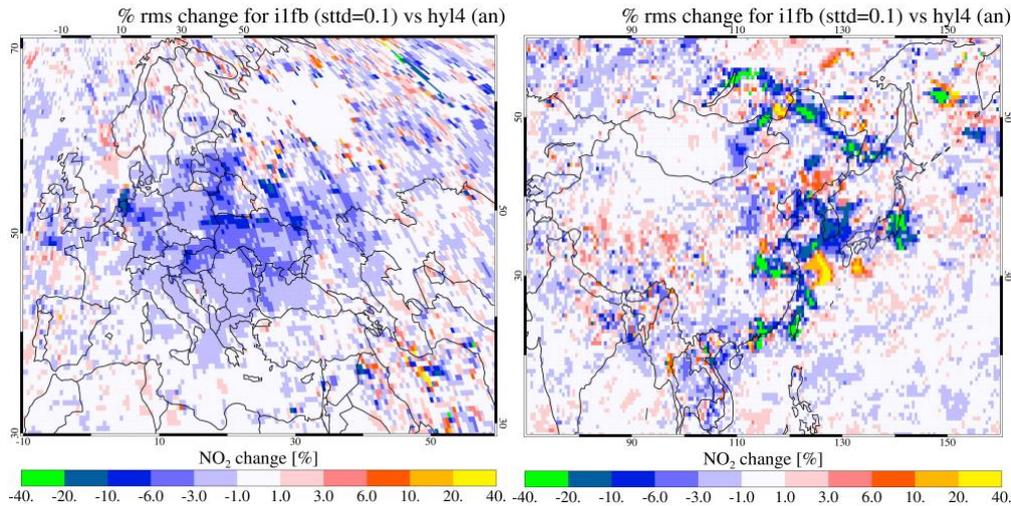
*Figure 7: Maps of relative change in Root Mean Square Error (RMSE) against TROPOMI NO2 observations (PAL product) between a state-only optimisation and a joint state-emission inversion. The RMSE are evaluated based on 24-hour IFS NO₂ tropospheric column forecasts using non-assimilated TROPOMI NO2 data. Green to blue colours indicate reduction in RMSE, while red to yellow colours indicate increase in RMSE.*

Figure 8 shows the in situ evaluation of the average daily time series of posterior surface NO₂ concentrations for many air quality sites over China, for the period 31/03/2020-30/05/2020, for a state-only and a joint state-emission optimisation. Both time series of RMSE and biases are used in the comparisons. The results show small to no improvements of the simulated surface NO₂ concentrations when emissions are optimised for those AQ sites. Such differences between the previous satellite-based (i.e., TROPOMI NO₂) and this in situ evaluations could be due to representativeness error owing to the low spatial resolution of the IFS inversion, since the emission optimisation in the CAMS air quality configuration is performed at about 80 km horizontal resolution.
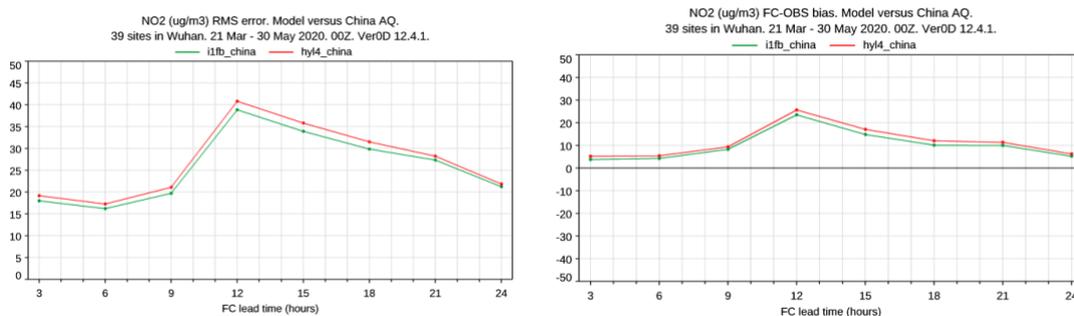


*Figure 8: Evaluation of posterior surface NO₂ concentrations obtained from a state-only optimisation (red line) and a joint state-emission optimisation (green). The figure shows average daily time series of Root Mean Square Error (RMSE) (left) and biases (right) for the period 31/03/2020-30/05/2020 based on observations from air quality stations over China.*

Figure 9 shows a comparison between posterior CH₄ emissions obtained from the global IFS inversion prototype (McNorton et al., 2022) and another inversion study (Zhang et al., 2020) over the Permian Basin (United States). While Zhang et al. (2020) report an estimate of 240 ± 40 Gg/month, the IFS optimisation yields an estimate of 190± 39 Gg/month, starting from a prior of 164 ± 3 Gg/month. Part of the differences between the IFS posterior estimate and Zhang et al. (2020) has been attributed to the significantly larger prior error used in the latter (McNorton et al., 2022).
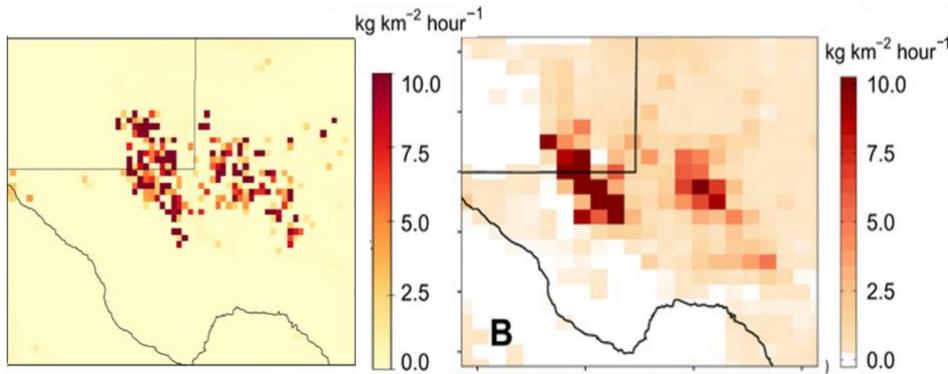
*Figure 9: Comparison between posterior emission estimates from the IFS inversion (left) and Zhang et al. (2020) (right) over the Permian Basin (United States) (McNorton et al., 2022).*

# 6  Next steps

The hybrid ensemble-variational data assimilation method outlined in Section 2.1 can be generalised to encompass parameters optimisation for process-based models, such as anthropogenic emission and biogenic flux models. In that context, the state increment $dC_a$ in Eq. (1) is replaced by an emission of flux increments and the flux increment $dF_a$ in Eq. (3) by a parameter increment, while the covariance matrix **B** represents the error covariance between the parameters and the emissions (or flux). The latter can be computed either using an ensemble (i.e., Monte-Carlo simulations) or an adjoint if available.

Additionally, the EnKS-like approach to assimilate regional and local posterior emission products described in Section 2.2.2 can also be used to assimilate complementary atmospheric GHG observations, including satellite observations that are not currently part of the CAMS observation suite and/or in situ measurements (e.g., ICOS sites data).

Figure 7 represents a schematic of the envisioned operational global IFS inversion system, including all the components discussed thus far. At each iteration of the hybrid ensemble-variational algorithm the satellite data are assimilated into the 4D-Var IFS system and this information propagated backward in time using the ensemble-based prior covariance to update fluxes and emission in previous windows. Once the global posterior fluxes ensemble has been updated, additional satellite or in situ observations can be assimilated into the system using the EnKS-like approach. The latter produces another posterior ensemble, and the algorithm is iterated at the next 4D-Var assimilation window. At any stage in that process, the GHG emission or flux increments produced can be converted into corresponding increments of parameters, provided a model relating the parameters to the fluxes is available.
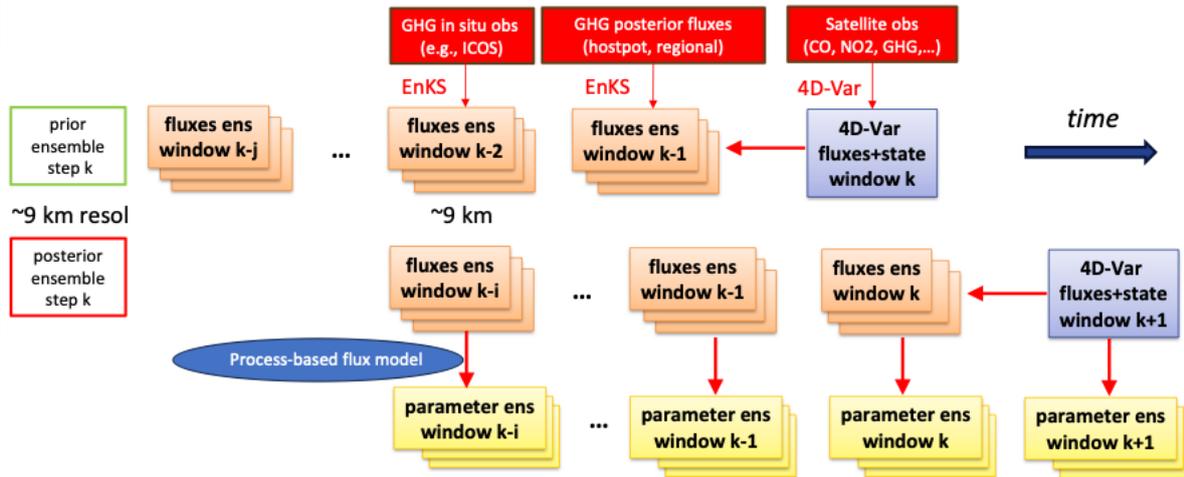
*Figure 7: Schematic representation of the extended ensemble-variational inversion workflow. At each iteration of the hybrid ensemble-variational algorithm the satellite data are assimilated into the 4D-Var IFS system to produce flux and 3D state increments. The 3D state increments are propagated backward in time using the ensemble-based prior covariance to update fluxes and emission in previous windows. The posterior fluxes ensemble thus obtained can be then used to assimilate additional in situ observations into the system using an EnKS-like approach. Additionally, parameters of processed-based models can be optimised using a similar method.*

The next steps in the development towards an operational multi-scale global inversion system will focus on three main aspects:

1. Parallelisation of the algorithm and optimisation of the ensemble generation, currently based on the EDA, using Hessian-based posterior square-root sampling approaches. The latter will enable to replace the ensemble of *n* costly 4D-Var minimisations by an ensemble of *n* perturbed forward trajectories. This will enable timely delivery of the posterior emission ensemble for operational production.
2. Implementation of efficient 4D localisation approaches to filter the sampling noise associated with the small ensemble size used. Such methods have been recently developed to be used in operational ensemble data assimilation systems (see, e.g., Ménétrier et al. (2015)).
3. Development of an extended prior error covariance (**B**) matrix to account for spatially-varying horizontal error correlation length scales, time and cross-species error correlations for the flux control vector. This activity forms a significant part of the ongoing Horizon Europe CORSO project. It will enable, in particular, to transfer information from co-emitters observations (i.e., CO, $NO_2$) to $CO_2$ emissions, providing more constraints on the source attribution problem.

The combination of a computationally efficient long-window ensemble-variational method, comprehensive prior information on spatial, temporal and co-emitters covariances, as well as cutting-edge noise filtering techniques, will provide the required tools to both meet the constraints of an operational system and deliver robust GHG flux quantifications. This new Copernicus product will help better monitor and understand trends in the carbon cycle to support policymakers in the implementation of the objectives of the Paris agreement.

# 7  Acknowledgments

# 8 References

Bousserez, N., 2019. Towards a Prototype Global CO2 Emissions Monitoring System for Copernicus. *arXiv preprint arXiv:1910.11727*

Bousserez N, Guerrette JJ, Henze DK., 2020. Enhanced parallelization of the incremental 4D-Var data assimilation algorithm using the Randomized Incremental Optimal Technique. *QJR Meteorol Soc*. 2020; 146: 1351–1371. https://doi.org/10.1002/qj.3740

Bousserez, N. and Henze, D.K., 2018. Optimal and scalable methods to approximate the solutions of large-scale Bayesian problems: theory and application to atmospheric inversion and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(711), pp.365-390.

McNorton, J., Bousserez, N., Agustí-Panareda, A., Balsamo, G., Cantarello, L., Engelen, R., Huijnen, V., Inness, A., Kipling, Z., Parrington, M., and Ribas, R., 2022. Quantification of methane emissions from hotspots and during COVID-19 using a global atmospheric inversion, *Atmos. Chem. Phys., 22, 5961–5981*, https://doi.org/10.5194/acp-22-5961-2022

Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre, 2015. Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization. *Mon. Wea. Rev.*, **143**, 1622–1643, https://doi.org/10.1175/MWR-D-14-00157.1.

Nassar, R, Moeini O, Mastrogiacomo J-P, O'Dell CW, Nelson RR, Kiel M, Chatterjee A, Eldering A and Crisp D., 2022. Tracking CO2 emission reductions from space: A case study at Europe's largest fossil fuel power plant. *Front. Remote Sens. 3:1028240*, doi:10.3389/frsen.2022.1028240

## Document History

| Version | Author(s) | Date | Changes |
|---------|-----------|------|---------|
| 1.0 | All authors | 27/12/2023 | Initial version |

## Internal Review History

| Internal Reviewers | Date | Comments |
|--------------------|------|----------|
| N/A | | |