CO2MVS RESEARCH ON SUPPLEMENTARY OBSERVATIONS

D2.6 Optimized B matrix parameters for IFS greenhouse gas emission inversions

Due date of deliverable	31 December 2024
Submission date	December 2024
File Name	CORSO-D2-6-V2
Work Package /Task	WP-2 / Task 2.3
Organisation Responsible of Deliverable	ECMWF
Author name(s)	A.J. Visser, N. Bousserez
Revision number	2.0
Status	Issued
Dissemination Level / location	Public

The CORSO project (grant agreement No 101082194) is funded by the European Union.

1 Executive Summary

The Integrated Forecasting System (IFS) has recently been extended to allow for emission estimation using inverse modelling of CO_2. Satellite observations of CO and NO_2, gases co-emitted with CO_2 in fossil fuel combustion, generally show more clear anthropogenic enhancements than CO_2. Their applicability to place additional constraints on anthropogenic CO_2 emissions in inversion systems is currently being investigated in the second work package of the CORSO project.

The potential impact of a satellite observation to correct underlying emissions depends, among other things, on the uncertainty associated with the emissions input datasets (the prior uncertainty). Additionally, (satellite) observations are generally sparse in space and time, so in order to maximize their impact in data assimilation systems, observational information must be spread to neighbouring grid cells and time steps in a statistically consistent manner. This is typically done using exponential decay functions with characteristic length and time scales, that describe the decay in uncertainty correlations in space and time. Similarly, propagating information from satellite observations co-emitted species to CO_2 emissions requires an estimate of the CO_2:CO:NO_x emission ratio uncertainties.

This report describes the derivation of prior uncertainties of anthropogenic CO_2 emissions and biogenic CO_2 uptake and release, and their spatial and temporal error correlation characteristics. For biogenic CO_2 exchange fluxes, these statistics have been derived based on a 20-year daily dataset of residuals between IFS and an independent product (FLUXCOM). The anthropogenic error statistics were derived from an ensemble of CO_2 emission fluxes, generated by perturbing input parameters in a Fossil Fuel Data Assimilation System (FFDAS, this dataset has been described in CORSO deliverable D1.4). The cross-species emission error correlation, reflecting emission ratio uncertainties, have been analysed in a separate CORSO report (Super and Hohenberger 2024).

We find that prior uncertainties in biogenic CO_2 uptake and release fluxes display spatial and temporal variability in line with their seasonal cycles and known difficulties of bottom-up models to capture temporal variability of these fluxes in semi-arid regions. Spatial error correlation length scales have been estimated for the first time at a high spatial resolution (0.25 degrees) based on IFS-FLUXCOM residuals, and the results are consistent with literature-reported values. Error correlation time scales of 23 (GPP) and 25 (REC) days were found, which also agrees well with previous analyses. Anthropogenic relative prior uncertainties based on the FFDAS results are ~0.5 on a daily basis. Spatial error correlation in the anthropogenic emission ensemble is highly variable in space and time and shows strong anisotropy, which must be incorporated with alternative methods to the isotropic spatial error correlation model currently used in the IFS. In contrast, prior errors are highly correlated in time, ranging from correlation coefficients larger than 0.9 in the first days up to 0.6-0.9 after several weeks, depending on the country.

This work paves the way for optimizing anthropogenic CO_2 emissions in the Integrated Forecasting System using a long-window emission inversion framework currently under development as part of the CO2MVS.

Table of Contents

2 Introduction

2.1 Background

To support EU countries in achieving greenhouse gas emission reduction targets, the EU and European Commission (EC) recognised the need for an objective way to monitor anthropogenic CO_2 emissions and their evolution over time. Such a monitoring capacity will deliver consistent and reliable information to support informed policy- and decision-making processes, both at national and European level. To maintain independence in this domain, it is seen as critical that the EU establishes an observation-based operational anthropogenic CO_2 emissions Monitoring and Verification Support (MVS) (CO2MVS) capacity as part of its Copernicus Earth Observation programme.

The CORSO research and innovation project will build on and complement the work of previous projects such as CHE (the CO_2 Human Emissions), and CoCO2 (Copernicus CO_2 service) projects, both led by ECMWF. These projects have already started the ramping-up of the CO2MVS prototype systems, so it can be implemented within the Copernicus Atmosphere Monitoring Service (CAMS) with the aim to be operational by 2026. The CORSO project will further support establishing the new CO2MVS addressing specific research & development questions.

One objective of CORSO is to investigate the use of co-emitted (CO, NO_x) to better separate fossil fuel emissions from the other sources of atmospheric CO_2. CORSO will deliver improved estimates of emission factors/ratios and their uncertainties as well as the capabilities at global and local scale to optimally use observations of co-emitted species to better estimate anthropogenic CO_2 emissions via inverse modelling.

This report focuses on the setup of the prior error covariance matrix of the global inverse modelling system of the CO2MVS, which is used to propagate observations in a statistically consistent manner space, time and across species within the data assimilation system.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverables

Deliverable D2.6 describes the method to derive optimized prior error covariance matrix (or B matrix) parameters for the Integrated Forecasting System (IFS) and the resulting datasets. This is an important component of ongoing developments to introduce greenhouse gas emission monitoring capacities in the IFS as part of the CO2MVS. More specifically, in this deliverable we:

(1) Introduce the prior error covariance matrix for the CO2MVS
(2) Introduce methodologies to calculate optimized spatial and temporal prior error correlation using flux ensemble datasets
(3) Discuss the applicability of conventional IFS tools to model spatial error correlation structures in surface fluxes of greenhouse gases.

2.2.2 Work performed in this deliverable

The work performed in this deliverable is part of CORSO Task 2.3, with the goal to model and test the prior error covariance matrix of the IFS. This includes the derivation of B matrix model parameters to capture spatial and temporal error correlation for biogenic and anthropogenic CO_2 fluxes, as well as the correlation between emissions of CO_2 and co-emitted species (NO_x and CO). Note that these cross-species correlations were the subject of a separate report produced in the context of CORSO task 1.4 and will not be considered here.

2.2.3 Deviations and counter measures

There were several modifications to the work description in Task 2.3:

- The CO2MVS IFS setup is still under development and not yet ready for testing the B matrix model parameters derived in CORSO T2.3, so this part could not yet be performed. Implementation and testing of the B matrix is an important part of the CO2MVS prototype system, which will be performed in the final year of the project as part of CORSO and a broader, ongoing effort at ECMWF.
- Although not explicitly stated in the proposal, B matrix model parameters are also derived for biogenic fluxes of CO_2 (gross primary productivity and ecosystem respiration).

3 The CO2MVS background error covariance matrix

The background error covariance matrix (or B matrix for short) is a term in the 4D-Var cost function, whose role is to spread information to unobserved neighbouring grid cells in a statistically consistent manner. This matrix is never formed explicitly but modelled for computational efficiency. In the IFS, a wavelet model of the B matrix is used, which allows for spatial correlation structures that vary in space (Fisher 2003):

$$x - x_b = E_b^{1/2} \sum_j^N \psi_j \times \left[C_j^{1/2}(\lambda, \phi) \chi_j \right]$$

where $x - x_b$ is the departure of model variables from the background, E_b is diagonal matrix of prior error standard deviations, Ψ_j is the wavelet function for the wavelet transform on the sphere at wavenumber j (reflecting spatial scale), $C_j^{1/2}$ is the square root of the vertical correlation matrix for latitude/longitude pair (λ, ϕ) and χ_j is the control variable.

Vertical spatial error correlation is described in the wavenumber-dependent background error correlation matrices C_j. This matrix determines how a deviation from the background at one location may affect surrounding grid cells. Since C_j depends on the spatial scales (reflected by the wavenumber), horizontal correlation structures can be reconstructed from the N vertical correlation matrices. A diagnostic length scale may be calculated to describe the extent of spatial error correlation at each grid cell.

The conventional IFS 4D-Var application has a 12-hour assimilation window. This is insufficient for the CO2MVS: due to the long atmospheric lifetime of greenhouse gases, a longer assimilation window (on the order of several weeks to months) is needed to trace back observed greenhouse gas columns to their emissions. In order to back-propagate initial concentration increments in the active 12-hour assimilation window to emissions in previous windows, a correlation term associated with the atmospheric transport must be added to the equation above. This part of the B matrix is not treated in this deliverable. Additionally, a prior temporal error correlation term in the B matrix should be added to the equation above to spread information across emissions in each window.

Another modification with respect to the conventional B matrix formulation is the addition of a cross-species error correlation operator. This adds CO2 emission constraints from co-emitted air pollutants.

4 Methods

4.1 Input data

This section details the input data used to derive prior estimates of B matrix model parameters. Table 1 includes an overview of the data sources, the required degree of processing, as well as some metadata (sectoral distribution, spatial and temporal resolution). The following subsections contain a brief description of the data sources. The methodology to derive B matrix parameters based on this input data is introduced in Section 4.2.

Table 1: Overview of the data sources used to derive B matrix parameters.

B matrix component	Data source (provider, deliverable)	Processing required	Sectoral distribution	Spatial resolution	Temporal resolution
	Anthropogenic emissions (CO_2, CO, NO_x)				
Prior error standard deviation (CO_2)	CO_2: 4D emission ensemble FFDAS + temporal profiles (iLab/BSC, D1.4), CO, NO_x: global emission uncertainties (TNO)	CO_2: calculate ensemble variance CO, NO_x: none	Total anthropogenic flux (CO_2, CO, NO_x)	FFDAS grid (0.1°×0.1°)	CO_2: daily CO, NO_x: constant
Spatial (error correlation length scale)	4D emission ensemble: FFDAS + temporal profiles (iLab/BSC, D1.4)	Calculate spatial autocorrelation, derive diagnostic length scale	Total anthropogenic CO_2 flux	FFDAS grid (0.1°×0.1°), regridded to IFS inner-loop resolutions	Daily
Temporal (error correlation time scale)		Calculate temporal autocorrelation, derive length scale	Total anthropogenic CO_2 flux		
Cross-species error correlation	CO_2:CO:NO_x error correlations (TNO)	None	Total anthropogenic flux (CO_2, CO, NO_x)	FFDAS grid (0.1°×0.1°)	Constant
	Biogenic CO_2 fluxes				
Prior error variance	Ensemble of IFS – FLUXCOM residuals	Calculate ensemble variance	Gross Primary Productivity (GPP), ecosystem respiration (REC)	FLUXCOM v2 grid (0.25°×0.25°), regridded to IFS inner-loop resolutions	Monthly
Spatial (error correlation length scale)		Calculate spatial autocorrelation, derive diagnostic length scale			
Temporal (error correlation time scale)		Calculate temporal autocorrelation, derive length scale			

4.1.1 Biogenic CO_2 flux ensemble

The CO2MVS prototype offers the capability to separately constrain Gross Primary Productivity (IFS variable name: GPP) and ecosystem respiration (IFS variable name: REC), which together comprise CO_2 net ecosystem exchange (NEE). GPP and REC are simulated online in the IFS, using a photosynthesis parameterization based on the Farquhar model, with a plant functional type-dependent bias correction applied to REC (ECMWF, 2023).

To derive B matrix parameters for GPP and REC, we probe the IFS uncertainty by calculating the residual against biogenic flux estimates from the FLUXCOM v2 product (Jung et al. 2020). This gridded data product is based on upscaled eddy covariance observation using various various types of land surface remote sensing observations and machine learning techniques.

The IFS-FLUXCOM residuals are calculated on a daily basis for 2001-2020 at a spatial resolution of 0.25°×0.25°. We subsequently group these data by month to create ~600-member ensembles for each grid cell. An example spatial distribution of residuals is shown in Figure 1.

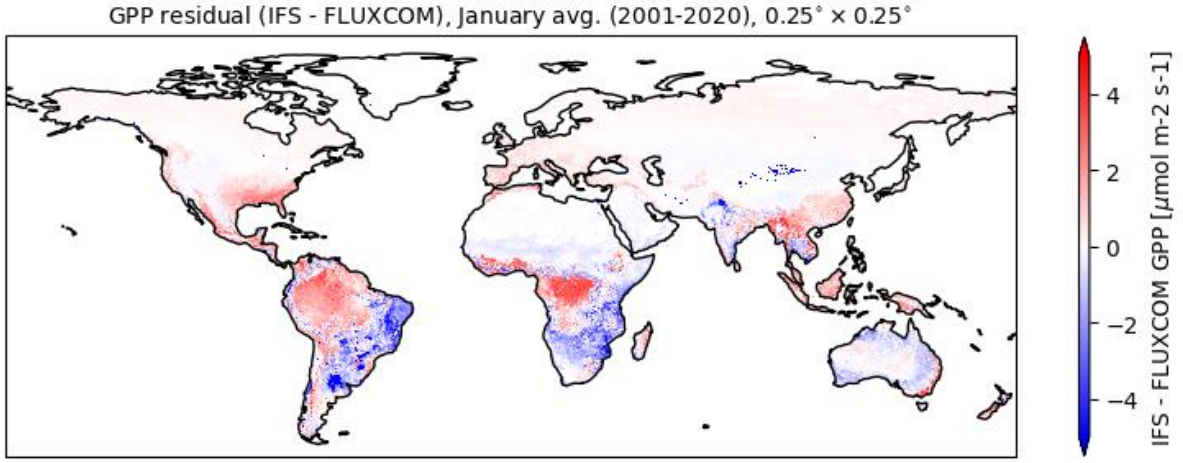

Figure 1: Spatial distribution of ensemble-average IFS-FLUXCOM residuals for GPP in January.

4.1.2 Anthropogenic CO_2 flux ensemble

An ensemble of anthropogenic CO_2 fluxes is taken from CORSO deliverable D1.4 (Kaminski and Voßbeck 2024). This dataset is generated using the Fossil Fuel Data Assimilation System (FFDAS, Rayner et al. 2010), which optimizes parameters in a process-based model of sectoral fossil fuel emissions against national emission totals based on observations of various spatial proxies of emissions, such as nighttime light intensity, population density, road networks and meteorological data. The ensemble is derived by propagating input parameter uncertainty to the total anthropogenic CO_2 flux in the FFDAS. Temporal emission variability is taken from CORSO M1 (Guevara 2023). This dataset has a spatial resolution of 0.1°×0.1° and a temporal resolution of 1 day. The ensemble mean is subtracted from each ensemble member to create an emission error ensemble.

4.2 Derivation of B matrix model parameters

The prior error statistics are calculated at three inner-loop resolutions (T_l95, T_l159, T_l255, corresponding to a resolution at the equator of 209 km, 125 km and 78 km, respectively) used for cost function minimization.

4.2.1 Prior error standard deviation

The prior uncertainty is characterized by the prior error ensemble standard deviation, is calculated for an n-sized ensemble by averaging the squared deviation of each ensemble member m from the mean:

$$\sigma(X) = \sqrt{\frac{1}{n-1}\sum_{m=1}^{n}(x_m - \bar{x})^2}$$

4.2.2 Error correlation length scale

The wavelet B matrix model used in the IFS allows for spatial variability in the modelled vertical and horizontal error correlation structures (Fisher 2003). These statistics are stored in so-called wavelet files, consisting of power spectra and vertical correlation matrices for several spectral bands. For this report, these files are generated based on an ensemble using an approach similar to Ensemble of Data Assimilations (EDA) applications applied operationally in IFS NWP analyses. The horizontal error correlation structure and error correlation length scale at the relevant inner-loop resolution can be diagnosed from the wavelet file components.

The diagnosed spatial error correlation length scale is modelled using a second-order autoregressive (SOAR) function (Pannekoucke et al. 2008):

$$\rho(\delta x, L) = \left(1 + \frac{\delta x}{L}\right) e^{-\delta x/L}$$

Where ρ is the spatial error correlation, δx is the lag distance and L is the correlation length scale.

4.2.3 Error correlation time scale

The CO2MVS B matrix propagates information in time by modelling the temporal autocorrelation of the background error. This temporal autocorrelation is derived as follows (Kountouris et al. 2015):

$$\rho(t, \delta t) = \frac{\sum_{t=1}^{T}(x_{l,t} - \bar{x_l})(x_{l,t+\delta t} - \bar{x_l})}{\sum_{t=1}^{T}\left(x_{l,t} - \bar{x_l}\right)^2}$$

where $x_{l,t}$ is the model-observation difference at grid cell l and time t, and δt is lag time. If the temporal autocorrelation decreases exponentially with lag time, this can be captured by fitting an exponentially decaying model with time scale τ:

$$\rho(\delta t, \tau) = (1 - \alpha)e^{-\delta t/\tau}$$

Note that we calculate anthropogenic temporal error correlation by correlating over the ensemble instead of the time dimension.

4.2.4 Cross-species error correlation

The methodology to derive cross-species emission error correlation between CO_2 and the co-emitted species CO and NO_x is described elsewhere (Super and Hohenberger 2024).

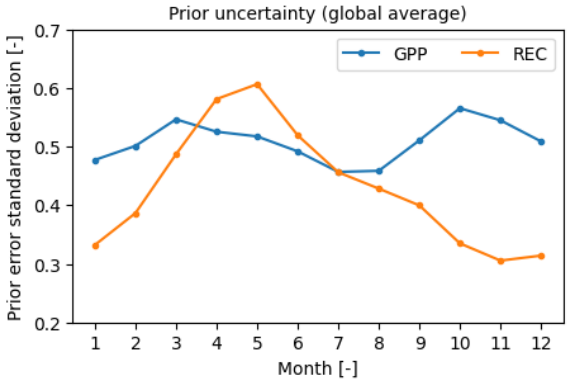

Figure 2: Seasonal cycle of the global average prior uncertainty derived from the GPP/FLUXCOM prior error ensemble interpolated to a T_l255 grid.

5 Results

5.1 Derivation of B matrix parameters for biogenic CO_2 fluxes

5.1.1 Prior error standard deviation

The prior uncertainties, currently described as a globally constant value in IFS inversions, display spatio-temporal variability following geographical and climatological patterns. This is summarized in Figures 2-4, which respectively display the globally average ensemble standard deviation per month (Fig. 2), and spatial distribution of the prior uncertainty in different months for gross primary productivity (Fig. 3) and ecosystem respiration (Fig. 4). The

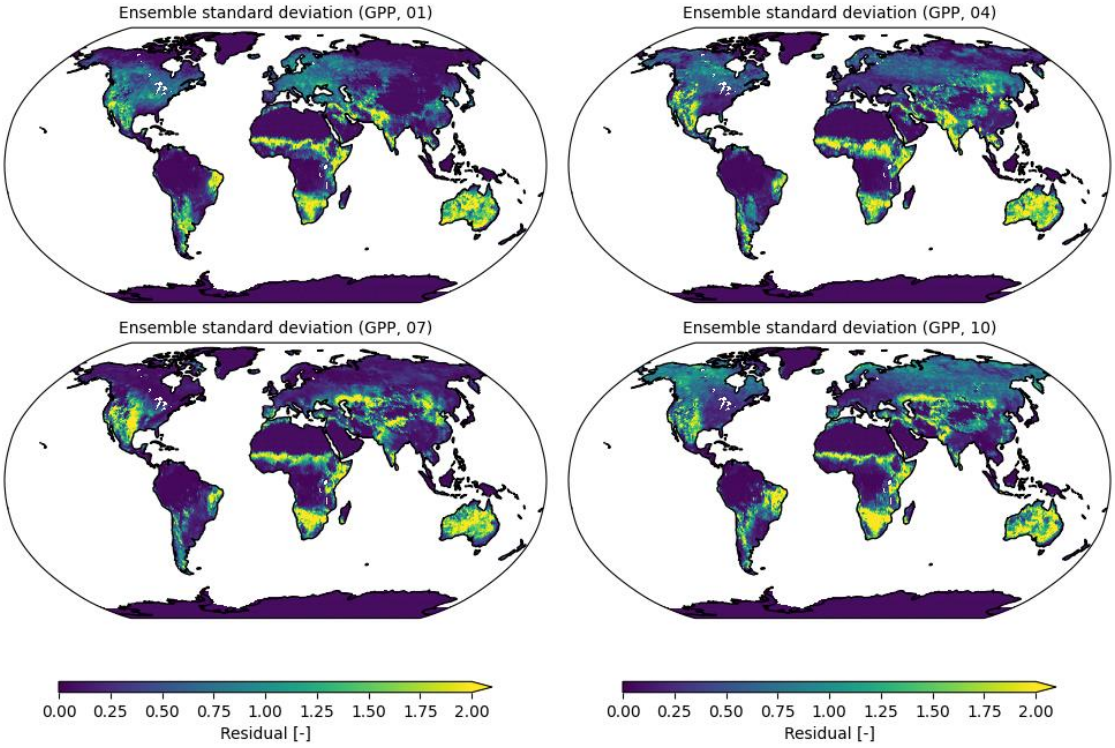

Figure 3: Spatial distribution of the prior error standard deviation for GPP in January, April, July and October (spatial resolution corresponds to a T_l255 grid with a resolution of 78 km × 78 km). Note: prior flux uncertainties are derived for every model grid cell, including regions with (near-)zero ecosystem-atmosphere CO_2 fluxes such as oceans (here filtered out) or arid regions such as the Sahara.

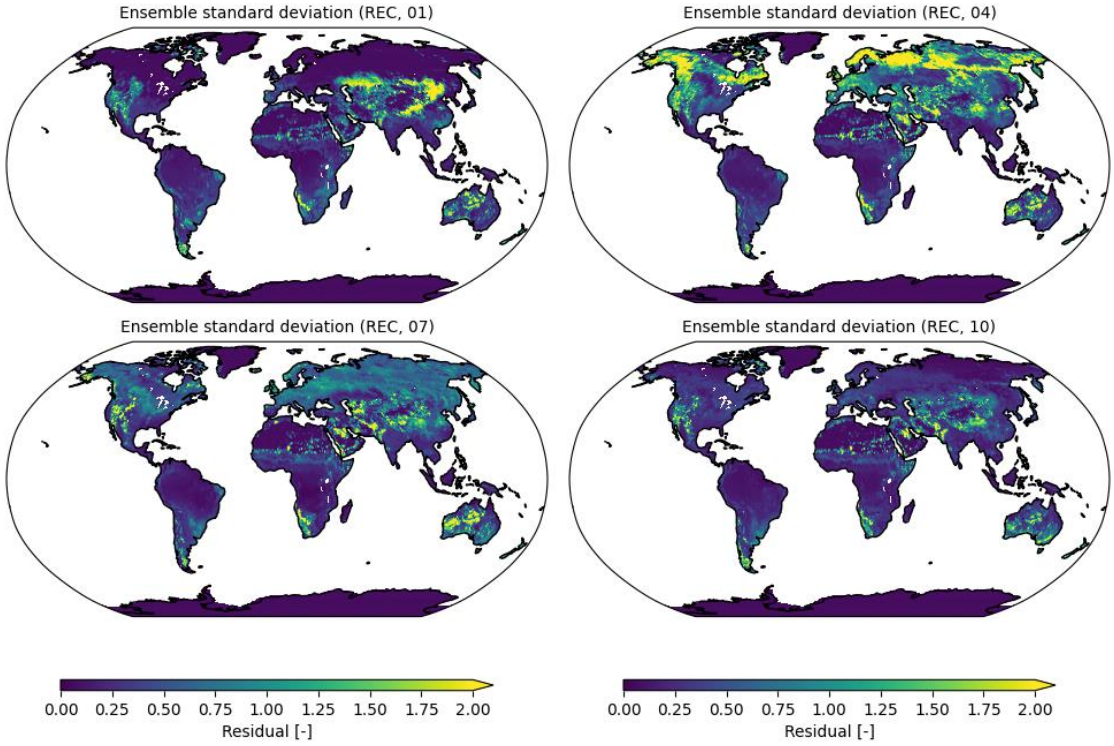

Figure 4: As Figure 3, but for ecosystem respiration.

global average prior error uncertainties for GPP and REC show different seasonal variability (Fig. 2). The seasonal variability in GPP prior uncertainties is relatively low throughout the year, which reflects the difficulties land surface models have to capture large inter-annual variability in biogenic CO_2 fluxes in semi-arid regions (e.g. the Sahel, southern Africa, Australia, Eastern South America and the western US), which has been reported elsewhere (e.g. Poulter et al. 2014, Humphrey et al. 2021). The two small peaks in the global average GPP prior error standard deviation in Figure 2 are caused by relatively large prior error standard deviations (up to 1) in the northern hemisphere at the start and end of the growing season (Fig. 3). In contrast to GPP, the prior uncertainty seasonal cycle for ecosystem respiration displays a distinct peak in April and May (Figure 2). This is driven by high prior uncertainties in northern Eurasia and North America in these months. In contrast to GPP, there are fewer areas with persistently high prior uncertainties throughout the year (Fig. 3), leading to lower globally average standard deviations outside the seasonal peak. This analysis was performed at the highest inner-loop resolution of the IFS cost function minimization; the spatial distribution of the prior uncertainty is similar when analysed at coarser resolutions, but the absolute values do show some differences, most notably for small-scale and high prior uncertainties.

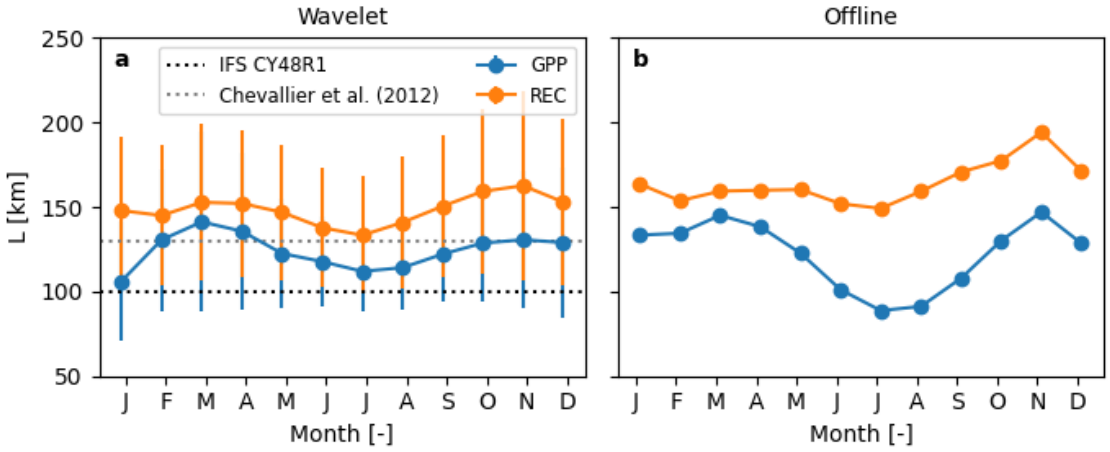

Figure 5: Seasonal variability of the prior error correlation length scale for GPP and REC, diagnosed from the horizontal error correlation in the T159 wavelet file (panel a) and from an offline analysis (panel b). Panel a additionally shows the current IFS error correlation length scale for GPP and REC, and a length scale for net ecosystem exchange derived from Chevallier et al. (2012).

5.1.2 Prior error correlation length scales

Like the prior uncertainty, the prior error correlation length scale (hereafter: length scale, diagnosed from IFS-FLUXCOM GPP and REC residuals, varies in time and space. The GPP length scale varies between 100 and 140 km, peaking in February-April with a smaller secondary peak at the end of the year. REC length scales follow a similar seasonal pattern but are on average ~25 km larger. This reflects the influence of spatial variability of solar radiation, one of the driving variables of simulated GPP, which adds spatial heterogeneity to the prior errors. On average, the length scales exceed the IFS length scale of 100 km used for GPP and REC. The wavelet-derived length scales agree well with the globally constant length scales used in other systems such as ORCHIDEE (130 km, Chevallier et al. 2012, highlighted in Figure 5a) and CarboScope-Regional (100 km, Kountouris et al. 2018), both derived based on spatial error correlation analysis using in-situ net ecosystem (NEE) observations. Additionally, length scales with similar magnitude and seasonality are derived from offline spatial correlation analysis using the IFS-FLUXCOM error ensemble at a spatial resolution of 0.25°×0.25°, shown in Figure 5b. This suggests the temporal variability derived here is robust and calls for including temporal variability in spatial error correlation patterns in the B matrix.

Besides temporal variability, the prior error correlation length scale varies considerably in space. To illustrate this, Figure 6 shows maps of GPP and REC length scales April and July. GPP and REC length scales peak in the Northern Hemisphere in April, reflecting spatially homogeneous error patterns associated to the onset of the growing season. For REC, spatial correlation of the prior error peaks in the Northern Hemisphere in spring, while Southern Hemispheric length scales peak in austral winter. In the eastern Amazon, GPP length scales are consistently higher than in surrounding regions, again indicating spatial homogeneity of the prior error. Previous studies did not consider spatial error correlation on a grid cell basis but rather focused on global averages or aggregated parameters per ecosystem type (Chevallier et al. 2012, Kountouris et al. 2018). This analysis suggests the potential for including more spatial detail in prior error covariance matrices for biogenic flux inversions, which is enabled by the wavelet B matrix formulation in the IFS.

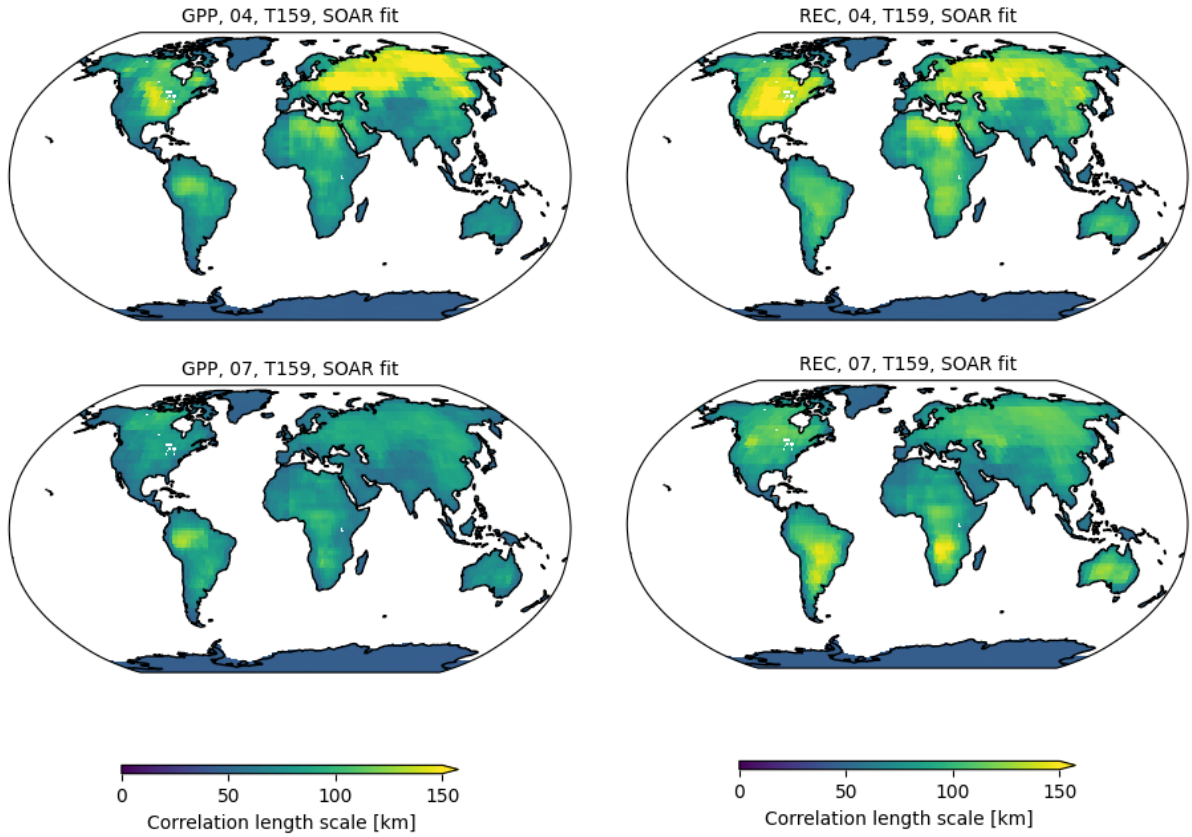

Figure 6: Spatial distribution of error correlation length scales diagnosed from IFS wavelet files on a Tl159 grid for GPP (left column) and REC (right columns) for two different months (top row: April; bottom row: July). Note: B matrix model parameters are derived for every model grid cell, including regions with (near-)zero ecosystem-atmosphere CO_2 fluxes such as oceans (here filtered out) or arid regions such as the Sahara.

5.1.3 Prior error correlation time scales

Biogenic flux errors are correlated in time on temporal scales of up to several weeks, and on annual time scales to a smaller extent. The global average temporal autocorrelation function for GPP in 2001 (Fig. 7a) shows positive error correlation until a lag time of ~70 days, subsequent negative autocorrelation of up to -0.25 until ~290 days, and then increasing again up to ~0.5 at a lag time of 365 days due to yearly recurring IFS-FLUXCOM residuals. Since the foreseen CO2MVS time window is several weeks to months, it is sufficient to include only the initial decrease in temporal error correlation. The initial decay of the temporal autocorrelation function can be reasonably described using an exponentially decaying curve (see Eqn. 5) with a time scale of 23 days. The described GPP temporal error correlation is similar for ecosystem respiration, with a similarly constant but slightly higher average time scale (~25 days). The seasonal variability in temporal autocorrelation was also observed in other studies (Chevallier et al. 2012, Kountouris et al. 2018), and the derived length scales fall within the range of 20-30 days reported in the literature (Broquet et al. 2013, Chevallier et al. 2018, Kountouris et al. 2015, 2018), lending credibility to the values derived in this analysis.

The large variability in the annual averaged time scales presented in Figure 7c-d points to large spatial variability in the error correlation time scales. The error correlation time scales were derived on a yearly basis for 2001-2020, and Figure Y shows the average time scales for GPP and REC over this period. Unlike for the spatial error correlation patterns, spatial variability in the temporal error correlation is less easily discernible. Notable features are the high inter-annual variability in the time scales in particular regions in the Southern Hemisphere,

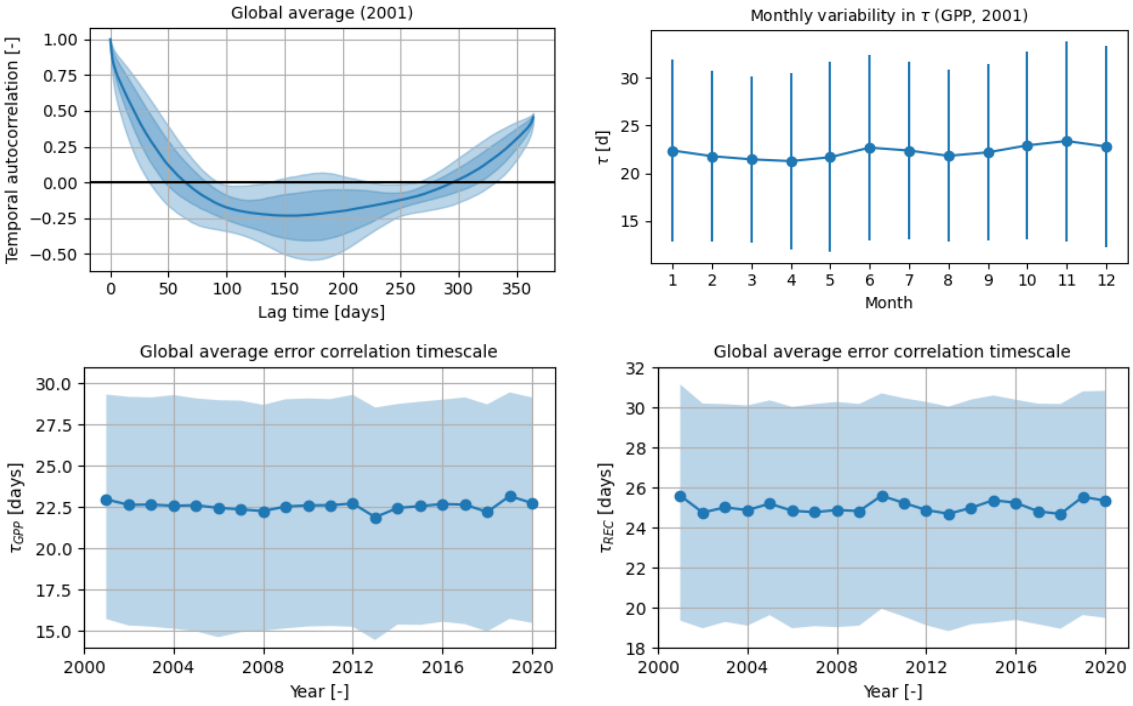

Figure 7: Temporal autocorrelation in biogenic fluxes and the resulting prior error correlation. Panel a shows the spatially averaged temporal autocorrelation for GPP in 2001 (dark and light shaded indicate the 50% and 80% spatial variability of the mean, respectively. Panel b displays the monthly variability in the error correlation time scale for GPP in 2001, while panels c and d show the yearly variability in the error correlation time scales for GPP and REC, respectively.

which must be corroborated further prior to application in the IFS. Given the uncertainty in spatial variability of the error correlation time scale, we suggest first experimenting with globally average correlation time scales of 23 and 25 days for GPP and REC, respectively.

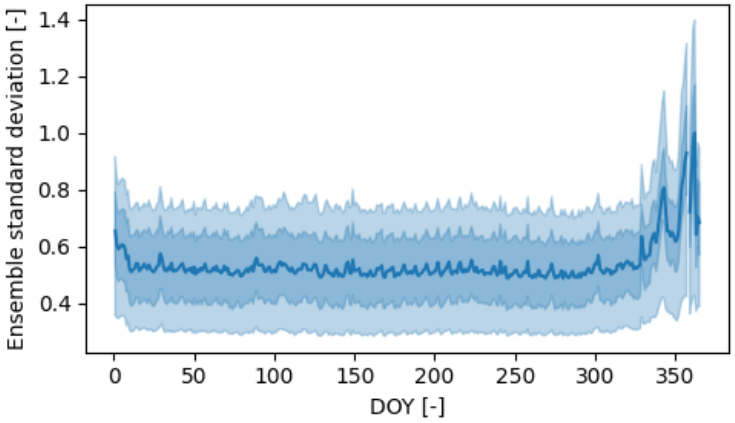

Figure 8: Prior anthropogenic CO_2 emission uncertainty derived from the FFDAS emission ensemble (diagnosed as the ensemble standard deviation).

5.2 Derivation of B matrix parameters for anthropogenic fluxes of CO_2 and co-emitted species (CO, NO_x)

5.2.1 Prior error variances

The one-year FFDAS anthropogenic CO_2 emission dataset enables calculating spatially explicit prior uncertainties on a daily basis as the ensemble standard deviation. This prior uncertainty displays little variability throughout the year, except for a notable peak in winter (Fig. 8). When looking at the spatial variability of the prior uncertainty, displayed in Figure 9, the prior uncertainties are larger in most regions in the world, most notably in South and East Asia and South America, with differences up to 0.3, and to a smaller extent in Europe and North America as well. In other countries (e.g., Angola, Democratic Republic of the Congo, Nigeria), the prior uncertainty does not vary between those dates. The TNO cross-species error correlation dataset also contains prior uncertainty estimates for anthropogenic CO_2 emissions, based on IPCC uncertainty estimates, which are substantially smaller (~0.2) than the FFDAS-derived uncertainties. The most appropriate uncertainty estimates must be evaluated using real-observation inversion test experiments in the IFS, which are planned for early 2025.

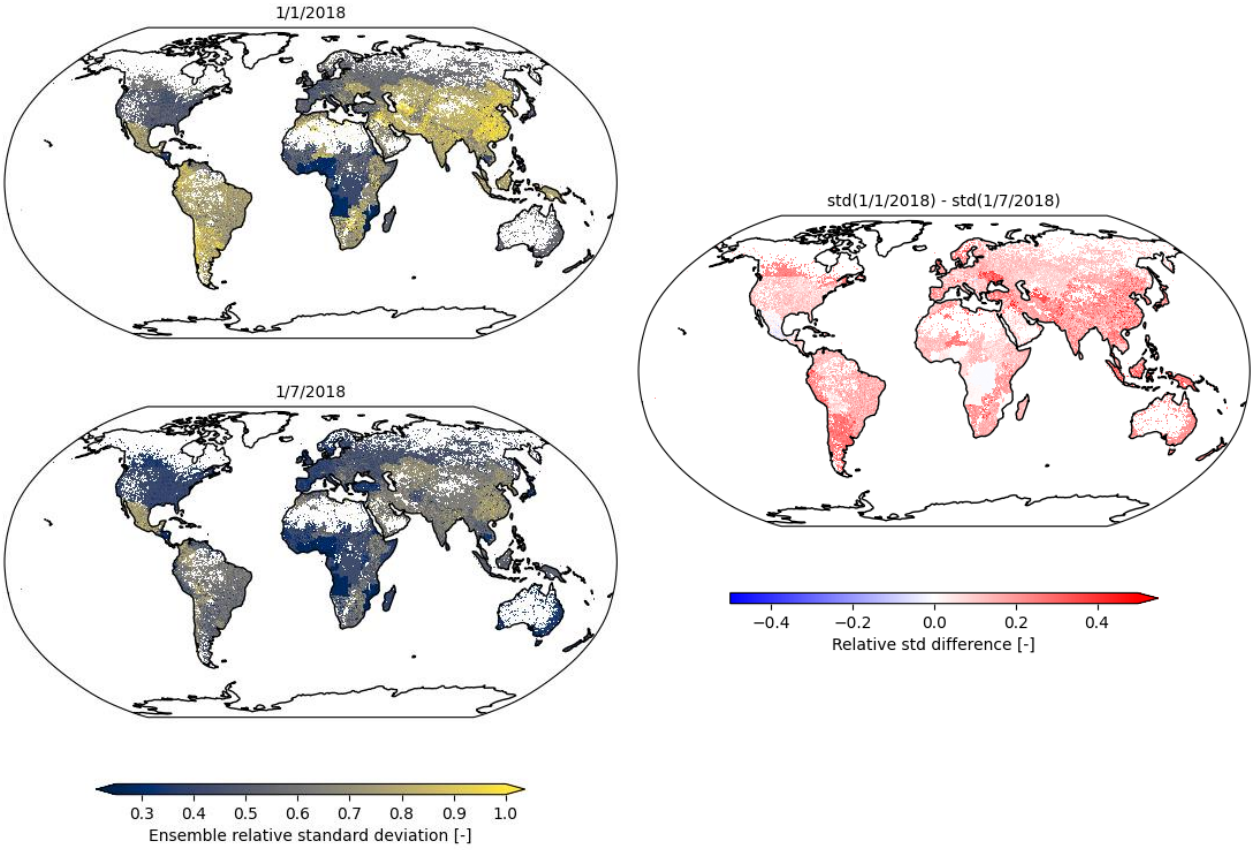

Figure 9: Spatial distribution of prior uncertainties for two days (upper left: January 1st, 2018; lower left: July 1st, 2018), and the absolute difference in prior uncertainty between those days (right).

5.2.2 Prior error correlation length scales

We first analyze spatial error correlation in the FFDAS ensemble at its native resolution of 0.1°×0.1° for some high-emitting cities (see Fig. 10). In several cities (e.g. Hong Kong, Shanghai, Singapore and Tokyo), the four-direction average spatial error correlation in drops quickly to near-zero values, while the spatial error is correlated over larger distances in some other cities (e.g. Los Angeles, Chicago and Tehran). In many cities, there are substantial differences between the spatial error correlation derived in different directions. This points at anisotropy in the spatial error correlation structures, which cannot be represented in the isotropic wavelet B model. To illustrate this further, we display spatial error correlation length scales diagnosed from wavelet files at two different resolutions in Figure 11. The length scales at the Tl255 resolution are markedly smaller than at Tl159, meaning that the diagnosed length scales decrease with increasing resolution. The resolution-dependence of the length scale is a sign that the spatial error correlation patterns are largely averaged out in the interpolation to the IFS inner loop resolutions. This also the case biogenic length scales in regions with small-scale spatial error correlation structures. However, length scales for larger-scale prior error correlation structures converge to similar values for biogenic fluxes when diagnosed at different spatial resolutions. We will revisit the issue of spatial error correlation in anthropogenic emissions in the concluding section of this report.

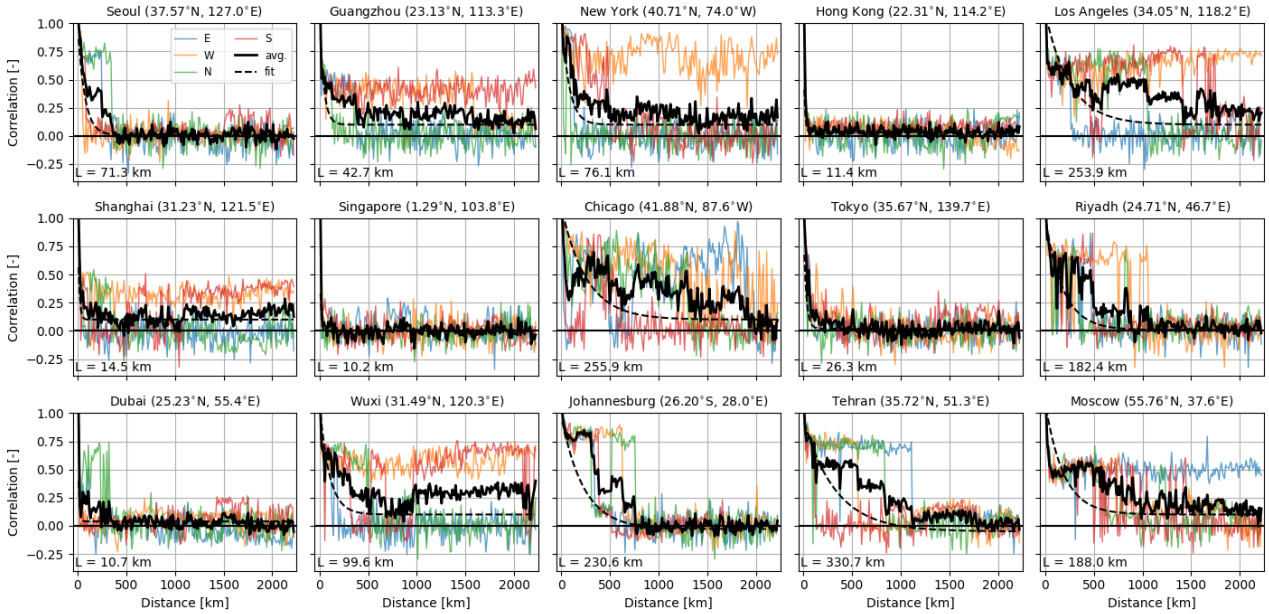

Figure 10: Distance correlograms for the 15 urban regions with the highest CO_2 emission footprint in the world based on Moran et al. (2018). Coloured lines show the spatial error correlation in four directions. The solid black line displays the four-direction average, and the dashed line shows the best exponential fit through the four-direction average spatial error correlation, with the error correlation length scale displayed at the bottom left of each panel.

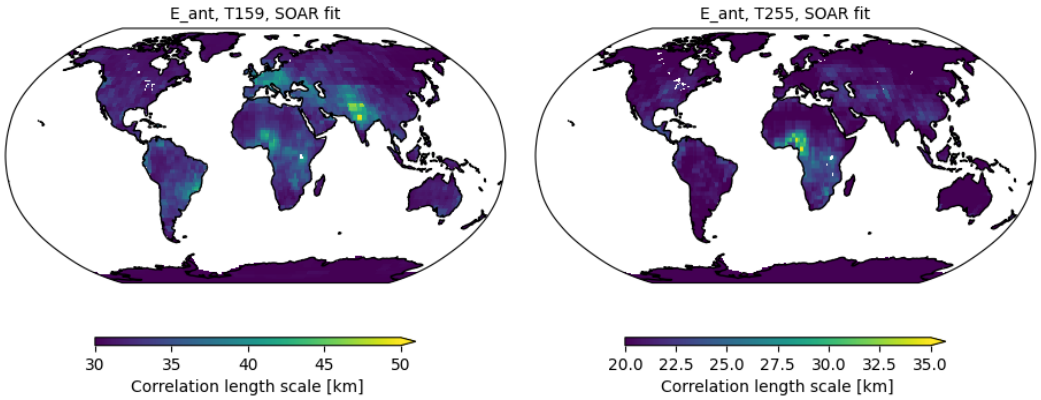

Figure 11: Spatial error correlation length scales for the anthropogenic CO_2 flux derived from the FFDAS yearly average emission ensemble at two different IFS inner-loop resolutions (left: T_l159; right: T_l255). Note the different colorbar ranges between the panels.

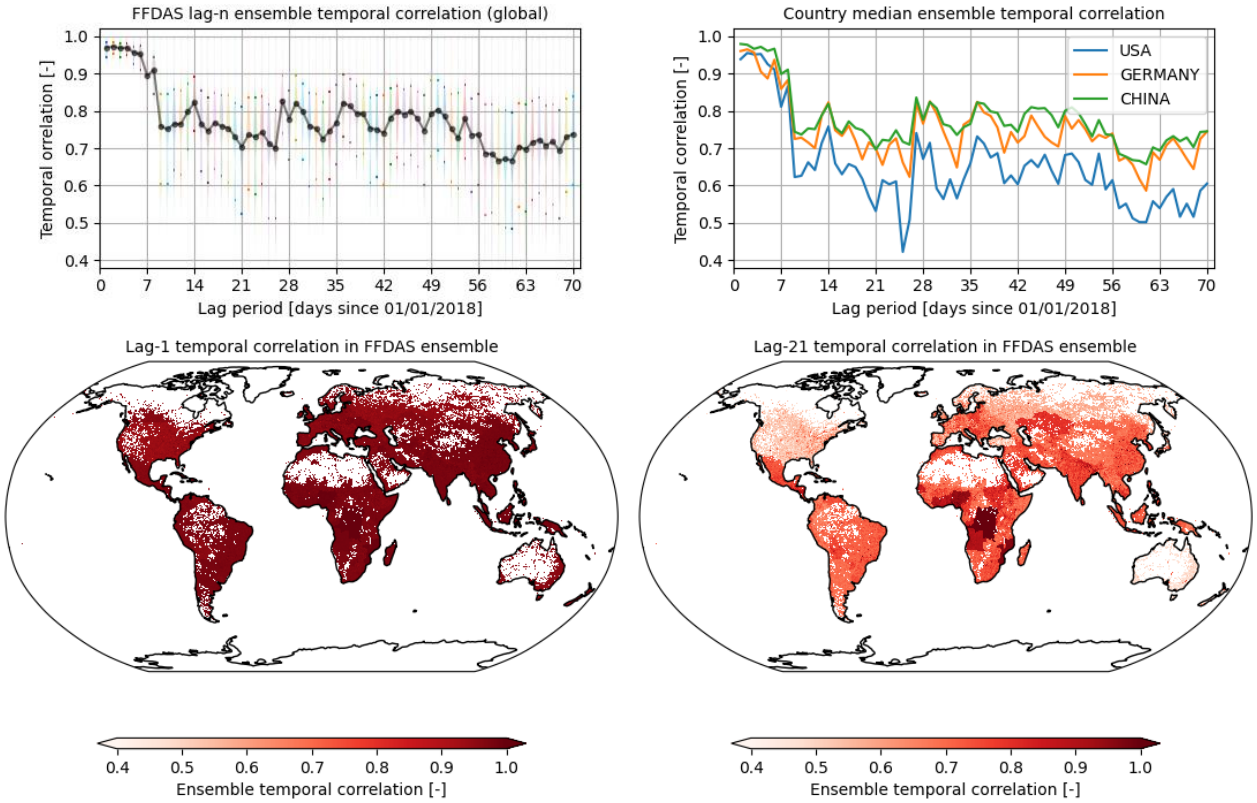

Figure 12: Temporal error correlation in the FFDAS anthropogenic CO_2 emission ensemble. Upper left panel: violin plots of the global ensemble temporal correlation (large black points and line show the global median, small points show the 80% quantiles). Upper right panel: median ensemble temporal correlation for three individual countries. Bottom panels: maps of the temporal autocorrelation for lag periods of 1 (left) and 21 (right) days.

5.2.3 Prior error correlation time scales

In contrast to spatial errors, correlation over the time dimension is more pronounced. The global average temporal error correlation, displayed in Figure 12 (upper left panel), is larger than 0.9 for the first seven days, then drops sharply to ~0.75, and it fluctuates around this value for up the 10 weeks analysed for this report. The error correlation decreases very slowly in some countries (e.g., Democratic Republic of Congo, Nigeria, see lower right panel in Fig. 12) due to flat temporal emission profiles for the dominant emission sources in those countries (e.g. wood-fired cooking) while temporally varying emission factors (e.g., transport) are not well described for these countries. The temporal error correlations for those countries are therefore likely overestimated in this dataset. The time correlogram for several high-emission countries (USA, Germany, China) shown in Figure 12 (upper right panel) follows a similar temporal pattern as described above, although the correlation drops to values slightly below the global average (~0.6) in the USA. This preliminary analysis demonstrates a significant temporal spread of observational constraints on the emissions, reducing the effective temporal resolution of the daily posterior emissions.

6 Conclusion and perspectives

This report discusses the derivation of optimized prior error covariance matrix parameters for the CO2MVS, in order to propagate information in space, time and across species within the IFS long-window inversion system in a statistically consistent manner. We have analysed spatial and temporal variability in prior uncertainties and error correlation length and time scales for biogenic fluxes (gross primary productivity and ecosystem respiration) and anthropogenic CO_2 emissions. We found that:

- Prior uncertainties in biogenic fluxes show spatial and temporal variability that are expected based on climatological patterns and known weaknesses of biospheric models (e.g. biogenic fluxes in semi-arid regions).
- Error correlation length scales for GPP and REC broadly agree with literature-reported values. For the first time, we report on prior error correlation length scales for biogenic fluxes at high spatial resolution.
- Biogenic fluxes are on average correlated over time scales of 23 (GPP) and 25 (REC) days, as well as over time scales of ~1 year. This is also in line with literature-reported values.
- Anthropogenic prior uncertainties derived from the FFDAS ensemble are derived at a high spatial resolution and displays little temporal variability but large spatial variability.
- Spatial error correlation for anthropogenic fluxes is highly anisotropic, which means the isotropic spatial error correlation model of the IFS B matrix is not appropriate.
- Prior errors of anthropogenic CO_2 emissions are moderately to strongly correlated in the time dimension, depending on the country.

The wavelet approach works well if (1) the error correlation structures are larger than the spatial resolution of the corresponding inner loop resolution of the cost function minimization, and (2) if the spatial error correlation structures are (largely) isotropic. The first condition is largely satisfied for biogenic CO_2 fluxes (GPP, REC), given that the error correlation length scales at different resolutions converge for larger correlation structures. In contrast, spatial error correlations from the FFDAS prior emission are mostly lower than the inner-loop resolutions, and additionally display strong anisotropy. This makes the wavelet approach unsuitable to derive spatial error correlations for anthropogenic fluxes.

Two alternatives can be explored to model spatial error correlations. Firstly, a hybrid B matrix formulation, which is a linear combination of a wavelet-B and an ensemble-based B, can be explored. This has the important advantage of capturing the full spatial variability in the error ensemble. Secondly, an updated inversion system in the IFS (CY50R1) enables optimization of individual source sectors. This includes the sectoral differentiation in B matrix model parameters. In this setup, point and line sources can be assigned low spatial error correlation length scales, while other sources characterised by larger spatial scale and more isotropic error structures can be treated using the wavelet model.

7 Data availability

The spatial and temporal error correlation scales are derived for internal use in the IFS and can be made available to CORSO partners on request. The FFDAS ensemble is available at the CORSO ftp server.

8 References

Broquet, G., Chevallier, F., Bréon, F.-M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P. (2013). Regional inversion of CO_2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmos. Chem. Phys., 13, 9039–9056, https://doi.org/10.5194/acp-13-9039-2013.

Chevallier, F., Wang, T., Ciais, P., Maignan, F., Bocquet, M., Altaf Arain, M., Cescatti, A., Chen, J., Dolman, A.J., Law, B. E., Margolis, H.A., Montagnani, L., Moors, E.J.. (2012). What eddy-covariance measurements tell us about prior land flux errors in CO_2-flux inversion schemes. *Global Biogeochemical Cycles, 26* (1), https://doi.org/10.1029/2010GB003974.

ECMWF (2023). IFS Documentation CY48R1 – Part VIII: Atmospheric Composition, https://doi.org/10.21957/749dc09059.

Fisher, M. (2003). Background error covariance modelling. Proceedings of seminar on recent developments in data assimilation for atmosphere and ocean, https://www.ecmwf.int/sites/default/files/elibrary/2003/74483-background-error-covariance-modelling_0.pdf.

Guevara, M., Val, A. (2024). Fluctuations of emission ratios in urban plumes. CORSO milestone report.

Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S.I., Frankenberg, C. (2021). Soil moisture-atmosphere feedback dominates land carbon uptake variability. Nature, 592, p. 65-69, https://doi.org/10.1038/s41586-021-03325-5.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020.

Kaminski, T., Voßbeck, M. (2024). First version of uncertainty ranges in flux space derived by an ensemble of emission model runs. CORSO deliverable report.

Kountouris, P., Gerbig, C., Totsche, K.-U., Dolman, A. J., Meesters, A. G. C. A., Broquet, G., Maignan, F., Gioli, B., Montagnani, L., and Helfter, C. (2015). An objective prior error quantification for regional atmospheric inverse applications, Biogeosciences, 12, 7403–7421, https://doi.org/10.5194/bg-12-7403-2015.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F., and Heimann, M. (2018). Atmospheric CO_2 inversions on the mesoscale using data-driven prior uncertainties: quantification of the European terrestrial CO_2 fluxes, Atmos. Chem. Phys., 18, 3047–3064, https://doi.org/10.5194/acp-18-3047-2018.

Moran, D., Kanemoto, K., Jiborn, M., Wood, R., Többen, J., Seto, K.C. (2018). Carbon footprints of 13000 cities. *Environmental Research Letters, 13* (6), 064041, https://doi.org/10.1088/1748-9326/aac72a

Pannekoucke, O., Berre, L., Desroziers, G. (2008). Background-error correlation length-scale estimates and their sampling statistics. *Quarterly Journal of the Royal Meteorological Society, 134* (631), p. 497-508, https://doi.org/10.1002/qj.288.

Poulter, B., Frank, D., Ciais, P., Myneni, R.B., Andela, Nl, Bi, J., Broquet, G., Canadell, J.G., Chevallier, F., Liu, Y.Y., Running, S.W., Sitch, S., Van der Werf, G.R. (2014). Contribution of

semi-arid ecosystems to interannual variability of the global carbon cycle. Nature, 509, p. 600-603, https://doi.org/10.1038/nature13376.

Super, I., Hohenberger, T. (2024). Global CO_2, CO and NO_x emission uncertainties and CO_2:CO:NO_x error correlations. CORSO report.

Document History

Version	Author(s)	Date	Changes
1.0	Auke Visser, Nicolas Bousserez	6 December 2024	
2.0	Auke Visser, Nicolas Bousserez	19 December 2024	Addressed reviewer comments and minor textual improvements

Internal Review History

Internal Reviewers	Date	Comments
ETHZ and CEA(LSCE)	Dec 2024	Comments addressed