Data

Models

Inventories

PARIS

Process Attribution of Regional Emissions

GA 101081430, RIA

Synthetic data inversions complete and modified fluxes unblinded

Milestone M25			
Delivery due date Annex I	PM 24 | 31/12/2024		
Actual date of submission	31/10/2025		
Lead beneficiary: [acronym]	Work package: 6	Nature: Report	Dissemination level: PU
Responsible scientist	M. Rigby (UNIVBRIS)		
Contributors	A. van der Woude, D. Kivits, R. de Kok, I. Luijkx, W. Peters (WUR) E. Saboya, B. Murphy, P. Sutar (UNIVBRIS)		
Internal reviewers			
Version: 0.1			

Horizon Europe Cluster 5: Climate, energy and mobility
"This project has received funding from the European Union's Horizon Europe Research and Innovation programme under HORIZON-CL5-2022-D1-02 Grant Agreement No 101081430 - PARIS".

M25 – Synthetic data inversions complete and modified fluxes unblinded

Table of Content

PARIS Work Package 6: Separation of Biosphere and Fossil Fuel Carbon Dioxide Flux

Milestone 25: Synthetic data inversions complete and modified fluxes unblinded

October 2025

Auke van der Woude[1], Eric Saboya[2], Daan Kivits[1], Remco de Kok[1], Brendan Murphy[2], Prasad Sutar[2], Ingrid Luijkx[1], Wouter Peters[1], Matthew Rigby[2]

[1]Wageningen University (WU)

[2]School of Chemistry, University of Bristol (UNIVBRIS)

Pre-amble

This milestone report is a static rendering of an interactive Jupyter Notebook in which future analyses of this work package can be executed. There may be formatting errors compared to the original notebook.

Here, we describe the results of the first phase of the Horizon Europe PARIS project blind CO_2 verification challenge. The results should not be considered final. The primary aim of this milestone was to establish protocols and workflows for the final analysis, which will be reported on at the end of the PARIS project.

Executive Summary

Evaluating progress towards Net Zero requires continual monitoring and accurate quantification of national anthropogenic carbon dioxide (CO_2) emissions. Anthropogenic CO_2 emissions estimates are reported annually in national emissions inventories, with Annex I parties (those listed in Annex I of the United Nations Framework Convention on Climate Change) citing emissions uncertainties of 2-10% (Jones et al,. 2021; Andres et al., 2012). However, national CO_2 emissions uncertainties can potentially be larger at sub-national and sub-annual scales (Basu et al., 2016). Independent estimates of national anthropogenic CO_2 emissions are therefore needed for accurate evaluation of climate change mitigation measures.

Inverse modelling is a well-established approach that uses atmospheric trace gas measurements to infer emissions estimates through a "top-down" approach. This approach is more often used to derive emissions estimates for non-CO_2 gases across Europe (and other parts of the world). For CO_2, the influence of terrestrial biosphere fluxes make it particularly challenging to constrain terrestrial anthropogenic CO_2 emissions on national scales from atmospheric CO_2 mole fraction measurements alone. However, atmospheric CO_2 mole fraction measurements from a sufficiently dense measurement network could potentially be used to detect the relative net flux changes of CO_2.

Work package 6 of the PARIS project aims to develop methods for separating natural and anthropogenic CO_2 fluxes. Objective 6.2 of this work package is to "determine the detectability of change in CO_2 sources and sinks in Europe". The double-blind CO_2 verification challenge (Task 6.4) was designed to test our current atmospheric measurement network capability to meet this goal.

To achieve this work package objective, two different inverse models were used to identify different flux perturbations intentionally introduced to a reference set of CO_2 fluxes. Each flux perturbation represents a different scenario that tests the inverse model. A total of five flux perturbation scenarios were constructed, each of which represented a major change in emissions, commensurate with the size of change anticipated due to climate change mitigation measures (e.g., the halving of fossil fuel emissions from road transport in Germany), or with the magnitude of potential errors in national sources and sinks. Four scenarios represent area source perturbations. One scenario represents a point source perturbation.

Atmospheric mole fraction pseudo-observations were produced by each modelling group (University of Bristol – UNIVBRIS; Wageningen University – WU), using different atmospheric transport models to transport fluxes from each scenario; the Met Office NAME model for UNIVBRIS and IFS-STILT for WU. These sets of pseudo-observations were swapped between the two groups and used as the observations for the inversions performed by each modelling group. The pseudo-datasets were not labelled, so that each group performed the inversions "blind" to the perturbations that had been applied.

In general, we found it was extremely challenging to correctly identify each of the flux perturbations by inverse methods. The flux scenarios typically lead to mole fraction differences of ± 0.5 ppm at each station that are comparable in magnitude to atmospheric CO_2 mole fraction measurement uncertainties ($0.1 - 0.3$ ppm). We found that differences between the UNIVBRIS and WU sets of pseudo-observations lead to flux differences of around 50 Tg C in Germany, for example. This is larger than, or similar to, the (substantial) differences between the perturbed flux fields (around 5-50 Tg C). The discrepancy between the flux results from each inverse model could be due to differences between atmospheric transport models or mole fraction boundary conditions. Further investigation is required.

It is, however, possible to sometimes correctly identify the perturbation flux scenario by comparing pseudo-observations produced using the *same* atmospheric transport model and assuming uncertainties are within ± 0.5 ppm. It is evident that the mole fraction responses to these large flux perturbations are small. This could suggest the need for a denser network of measurement stations across Europe that may be more sensitive to large scale flux changes and/or the measurement of co-emitted tracer species. It is concerning that certain flux perturbation scenarios were challenging to identify - within uncertainties - when using current inverse frameworks.

The primary aims of this preliminary verification challenge were to establish protocols and data formats and carry out one complete cycle of producing pseudo-observations, carrying out inversions and comparing unblinded results. Based on lessons learned from this experiment, an improved protocol will be developed, a new set of emissions scenarios will be developed, and a wider group of modelling teams will be invited to join. Based on analysis of the next iteration of our experiment, we hope to better understand the limits of detectability of CO_2 flux changes using the current observing system.

Table of Contents

---

1. Background

Accurate monitoring of national greenhouse gas emissions is needed for ensuring Net Zero. Whilst Annex I parties to the United Nations Framework Convention on Climate Change (UNFCCC) are required to report their annual emissions estimates, they are not required to independently verify those estimates. Independent emissions evaluation is an important part of ensuring climate change mitigation measures are effective and national emissions targets are being met.

Inverse modelling is a well-established approach for deriving national emissions estimates from atmospheric trace gas measurements. This "top-down" approach complements the "bottom-up" (emissions factor-based and process model-based) approach used for reporting in national inventories. National scale top-down emissions estimates have been derived for anthropogenic fluxes of non-CO_2 gases in several previous studies (e.g. Saboya et al., 2024; Manning et al., 2023; Lunt et al., 2021). For CO_2, it is much more challenging to derive accurate top-down estimates from atmospheric CO_2 mole fraction measurements alone with current inverse modelling capabilities. Fluxes from the terrestrial biosphere are often nearly an order of magnitude larger than anthropogenic emissions (e.g. fossil fuel combustion). Without using atmospheric tracers of CO_2 (e.g. radiocarbon, atmospheric potential oxygen), it is known to be extremely challenging to derive accurate top-down anthropogenic emissions estimates on sub-annual timescales (Basu et al., 2016). Even current tracer-based methods are known to suffer from potential sources of bias. So far, inverse methods relying on atmospheric CO_2 mole fraction measurements have primarily been used to derive estimates of the net CO_2 flux (e.g., White et al., 2019).

There is an urgent demand for producing robust regional top-down CO_2 flux estimates (Friedlingstein et al., 2024). Top-down estimates can vary depending on the inverse method, the atmospheric transport model, selection of atmospheric measurements, and boundary condition mole fractions that are used. All of these factors contribute to uncertainties in the top-down emissions estimates. The impact of these sources of uncertainty can be assessed through observing system simulation experiments (OSSEs).

Objective 6.2 of PARIS Work Package 6: "separation of biosphere and fossil fuel carbon dioxide flux", seeks to "determine the detectability of changes in CO_2 sources and sinks in Europe". The double-blind verification challenge was designed to meet this objective by constructing five flux perturbation scenarios to test whether different inverse methods could correctly identify each scenario. In the first iteration of this verification challenge, we devised a set of five flux perturbations, each of which represented a major change in emissions, commensurate with the size of change anticipated due to climate change mitigation measures (e.g., the halving of fossil fuel emissions from road transport in Germany), or potential errors in reporting. Each group aimed to identify the correct flux perturbations that were applied to model mole fraction simulations shared by the other group, using their inverse model. Four of these scenarios introduced country-wide or domain-wide flux perturbations. And the remaining scenario introduced point-source perturbations across the entire domain. These OSSEs were performed by two different inverse models.

The RHIME (Regional Hierarchical Inverse Modelling Environment; Ganesan et al., 2014) and CarbonTracker ensemble Kalman filter (Van Der Laan-Luijkx et al., 2017) inverse models were used in this first CO_2 verification challenge. Each inverse model uses a different atmospheric transport model (ATM), boundary conditions method and inverse method.

Each modelling group produced hourly pseudo-observations for the atmospheric measurement sites in the ICOS network (and non-ICOS sites) for each flux perturbation scenario and the unperturbed reference (base) scenario for the year 2021. Pseudo-observations were swapped between participating modelling groups and used as the observation vector in each inverse model. As each modelling group uses a different ATM, differences in the background-subtracted pseudo-observations at each measurement site are a result of differences in the ATM's transport, underlying meteorology, and model resolution. This provides a proxy for atmospheric transport uncertainties from the real atmosphere.

Comparison of the posterior flux estimates with the base fluxes should theoretically reveal which signals are detectable from the set of perturbed flux scenarios. The outcomes of these OSSEs should provide information about the confidence level of posterior European CO_2 flux estimates for detecting CO_2 signals using the available network of atmospheric measurement stations.

The protocol for the double-blind verification challenge was published as part of PARIS Milestone 6.1. A summary of the protocol is stated below.

1.1 Summary of the verification challenge protocol

1. Each modelling group uses the same base flux field from the 2021 Carbon Tracker Europe High-Resolution (CTE-HR) product. The five perturbed flux scenarios are produced by intentionally introducing perturbations to the CTE-HR flux field.
2. Pseudo-observations are produced for ICOS and non-ICOS measurement stations across the UK and Europe for each set of flux fields for the entirety of 2021. The choice of sites should depend on whether modelling groups are confident that atmospheric transport models can accurately model the meteorology conditions present at each of these sites. For example, atmospheric mole fraction measurements for certain high-altitude sites are more challenging to model and usually omitted from some European greenhouse gas modelling studies.
3. Pseudo-observations produced by each group are swapped with each flux perturbation scenario ''blinded'' in advance.
4. Each modelling group performs inversions for the set of fluxes.
5. Modelling groups attempt to match each perturbed flux scenario with the inverse model output.
6. After both modelling groups have completed their inversions, the flux scenarios are revealed.

1.2 Outline of notebook

This Jupyter notebook is organised as follows. Section 2 gives an overview of the flux perturbation scenarios and pseudo-observations used in the double blind verification challenge. Section 3 summarises the RHIME and CarbonTracker inverse frameworks. The results from both modelling groups are presented and discussed in separate notebooks for the mole fraction analyses and the flux analyses. The outlook of the verification challenge is given below

---

2. Flux perturbation scenarios and pseudo-observations

2.1 Flux perturbation scenarios

The CarbonTracker Europe High-Resolution (CTE-HR) gridded flux product for the year 2021 is used as the basis of the flux perturbation scenarios for the verification challenge. The base or reference scenario consists of the unperturbed flux fields which span across Europe (15^{o}W to 35 ^{o}E; 33^{o}M to 72^{o}N) with an hourly time resolution and $0.1^o \times 0.2^o$ spatial resolution. The flux fields include estimates for the terrestrial and oceanic biosphere, fossil fuel combustion emissions and biomass burning emissions.

The five perturbation scenarios are as follows. The labels given to each scenario by the originating group in the blinded datasets are also provided.

1. **ATEN**: A 10% increase in total anthropogenic emissions within the CTE-HR domain. (Positive area source perturbation).

- UNIVBRIS label: Zeus
- WU label: Hades

2. **DFIN**: Doubling forest carbon uptake in Finland. (Positive area source perturbation).
   - UNIVBRIS label: Apollo
   - WU label: Poseidon
3. **HFRA**: A 50% decrease in industry-related anthropogenic emissions in France. (Negative area source perturbation).
   - UNIVBRIS label: Ares
   - WU label: Athena
4. **HGER**: A 50% decrease in transport-related anthropogenic emissions in Germany. (Negative area source perturbation).
   - UNIVBRIS label: Hermes
   - WU label: Hephaestus
5. **PTEN**: Shutting down the top 10% emitting power plants within the CTE-HR domain. (Negative point source perturbation).
   - UNIVBRIS label: Demeter
   - WU label: Aphrodite

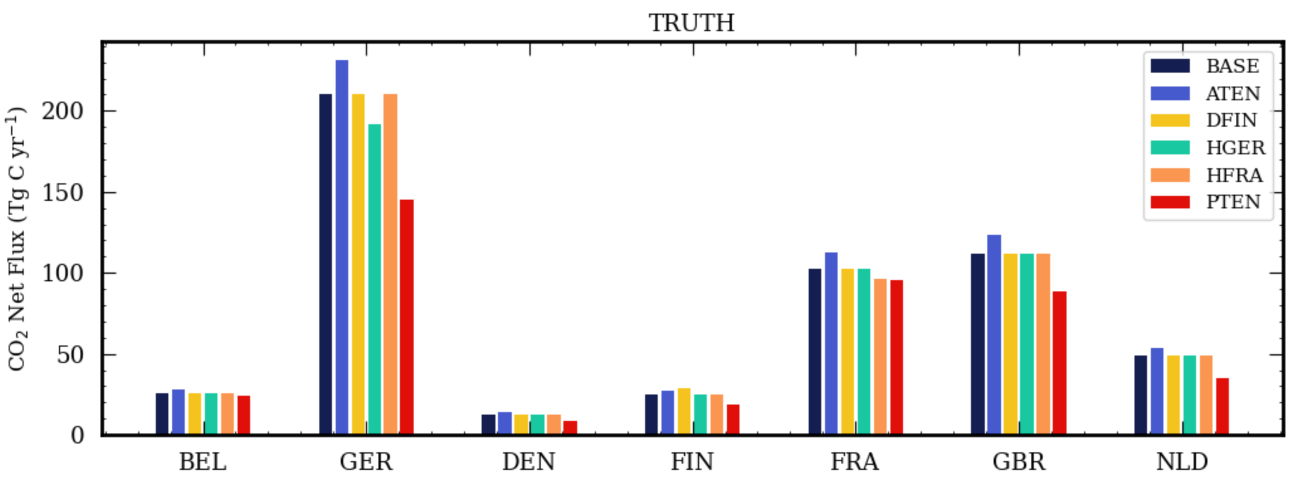


Figure 2.1: The ''true'' 2021 net CO_2 fluxes for key European countries for each of the PARIS verification challenge flux scenarios.

There are notable differences between each of the perturbed flux scenarios and the base fluxes for certain countries in the 'true', or expected, fluxes shown in Fig. 1. In Germany, for example, PTEN has a larger perturbation from BASE than the HGER scenario, which is halving German transport emissions across 2021. For other countries, HGER does not show a perturbation from the BASE scenario which is expected since this perturbation is only for Germany. In France, the PTEN and HFRA scenarios have a similar perturbation from BASE but are smaller in magnitude than the deviation arising from the ATEN scenario. Looking at the perturbations from these scenarios in other countries is key to discerning the correct scenario since HFRA is specific to France.

It is not too challenging to attribute the correct flux scenario when examining the set of ''true'' fluxes. Doing this attribution from fluxes inferred from atmospheric mole fraction measurements will be more challenging as it will depend on the sensitivity of the European atmospheric measurement network to CO_2 fluxes across Europe - particularly to the regions where the flux perturbations are most profound - and whether the corresponding atmospheric mole fraction perturbation is larger than the typical observational uncertainty.

2.2 Pseudo-observations

We can model the contributions to an atmospheric trace gas mole fraction by using a mass-balance approach. The atmospheric mass-balance for CO_2 can be written as:

$$C_a = C_{bg} + C_s - C_r.$$

Where atmospheric (pseudo-) observations (C_a) can be equated to the sum of background mole fractions (C_{bg}) and a component representing mole fraction changes that have not yet been mixed into the background atmosphere. The latter being the contributions of source (carbon-emitting; C_s) and sink (carbon-removal; C_r) processes.

Hourly pseudo-observations for the year 2021 were created for ICOS and non-ICOS measurement stations by combining the different flux fields with footprints from two different ATMs: NAME (Numerical Atmospheric dispersion Modelling Environment; Jones et al., 2007) v7.2 and IFS-STILT (Van der Woude et al., 2023). Pseudo-observations were created for the six sets of flux fields (five flux perturbation scenarios plus the base fluxes). The pseudo-observations for the different scenarios were then ''blinded'' - so that each group would not know which scenario they corresponded to - and then swapped between the two modelling groups and used as the observation vectors for the inversions.

We briefly describe the pseudo-observations produced by each modelling group and how they compare with the real observations at the different stations.

2.2.1 NAME-UKV simulations

The UK Met Office Lagrangian particle dispersion model, NAME, was used to quantify the relationship between surface emissions and atmospheric mole fractions measured at each station. "Footprints" of surface emission sensitivities were calculated from ensembles of particle back-trajectories (30,000 particles per footprint). Each grid cell of the footprint describes the influence of potential emissions from that grid cell on the measured mole fractions at the measurement site at a certain time (Manning et al., 2011).

Hourly footprints were calculated as described as in White et al. (2019) with a $\sim 25 \times 25$ km^2 ($0.352^o \times 0.234^o$) spatial resolution over a model domain spanning approximately 98^o W to 40^o E and 11^o N to 79^o N using a 30-day air-history. The first 24 hours of the air-history are not time-integrated to capture the diurnal variability of the terrestrial biosphere fluxes. The remaining 29 days of the air-history are time-integrated.

Meteorological fields from the Met Office Unified Model (UM) underlie the footprints with hourly, high-resolution (up to ~ 1.5 km ($0.0135^o \times 0.0135^o$) with 57 vertical levels up to ~ 12 km) UKV meteorological fields used for over the British Isles and 3-hourly UM ($0.1406^o \times 0.0938^o$ with 59 vertical levels up to ~ 30 km) global meteorological fields used for the rest of the modelling domain.

Footprints were combined with the gridded fluxes to simulate background-subtracted atmospheric CO_2 mole fractions at each measurement site. For the part of the NAME domain that does not encompass Europe, there are no fluxes. We used flux fields from other datasets for this part of the domain to ensure a like-for-like comparison. Anthropogenic fluxes from EDGAR v8.0, terrestrial biosphere fluxes from ORCHIDEE v2.0 and ocean fluxes from Jena Carboscope were used to fill the rest of the model domain. On average, this part of the domain contributed an extra 0.5 ppm to the pseudo-observations.

Background mole fraction estimates for each station were calculated by combining the NAME model particle end locations at each cardinal boundary at the edge of the NAME domain with global 3D CAMS (Copernicus Atmosphere Monitoring Service) v22r1 global inversion-optimized CO_2 mole fraction fields.

Pseudo-observations were produced for the following stations (defined in Appendix A): BSD, CMN, GAT, HEL, HPB, HTM, JFJ, KIT, KRE, LIN, LUT, MHD, NOR, OPE, OXK, PAL, RGL, SAC, STE, TAC, TOH, UTO.

2.2.2 IFS-STILT simulations

To analyse uncertainties related to atmospheric transport, we also use the Stochastic Time-Inverted Lagrangian Transport Model (STILT). We use a version similar to Botia et al. (2022) and Van der Woude et al. (2023), with a spatial resolution of $0.1^o \times 0.2^o$, which covers exactly the fluxes as shown above. STILT is driven by IFS forecast fields at $0.25^o \times 0.25^o$ resolution, with 137 layers in the vertical.

Similar to NAME, STILT is a Lagrangian transport model, meaning that footprints are created from back-trajectories. We create footprints for each station and observation hour, for which we release 250 particles at each receptor (i.e. observation hour and station). We follow these 250 particles backwards in time for 10 days, and record when and where they exchange with the surface. It is assumed that the receptor is sensitive to fluxes at that location and time. The footprints are convolved with hourly underlying flux fields to create enhancements in CO_2 due to local fluxes. We only use atmospheric observations for times at which atmospheric conditions are well-mixed (generally 12:00 to 16:00 local time for continental sites). The background atmospheric CO_2 mole fractions are based on the optimised atmospheric mole fractions from the CTE contribution to the 2023 Global Carbon Budget (Friedlingstein et al., 2023). For this, we traced each of the 250 particles to their ''end-point'' (i.e. where they left the domain, or where they exchanged with the surface). The average atmospheric mole fraction of CO_2 over the 250 particles is used as the background value.

3. Inverse methods

This section briefly describes the different inverse methods used by the University of Bristol (using RHIME) and by Wageningen University (using CarbonTracker) modelling groups.

3.1 RHIME

RHIME (Regional Hierarchical Inverse Modelling Environment; Ganesan et al., 2014) has been frequently used for trace gas inversions of various atmospheric species across the globe. In this inverse framework, Bayesian inference uses MCMC (Markov Chain Monte Carlo) to quantify a mean scaling and confidence intervals of a priori emissions for 100 spatial basis functions in the inversion domain. The span of the RHIME inversion domain is the same as the NAME domain. A scaling of boundary conditions on the four cardinal boundaries of the NAME domain is also derived in the inversion. I.e. the parameters vector comprises of 100 state variables for the fossil fuel basis functions, 100 variables for NEP, 100 variables for ocean fluxes and 100 variables for fire fluxes, plus 4 basis functions for the model domain boundaries.

As mole fraction uncertainties were not provided with the pseudo-observations, we use 0.1% of the simulated mole fraction value (~ 0.43 ppm). This is similar to the atmospheric mole fraction measurement uncertainties reported by ICOS. Boundary condition fields representing mole fractions along the edges of the NAME model domain are specified using CAMS (Copernicus Atmosphere Monitoring Service) v22r1 global inversion-optimized fluxes for each month. A scaling is calculated for each boundary in each 1-month period of inference to derive posterior boundary condition mole fractions. This ensures the CAMS data are not systematically underestimating or overestimating the baseline mole fractions at each of the measurement stations.

RHIME uses one hyperparameter that characterizes the probability density function (PDF) of the model error. The RHIME framework allows uncertainties in the scaling parameters to be included in the model. The a priori scaling for NEP is sampled from a normal PDF $\sim N(\mu = 1, \sigma = 1.0)$, fossil fuel combustion from a truncated normal PDF $\sim TN(\mu = 1, \sigma = 15.0)$, ocean and fire fluxes from a normal PDF $\sim N(\mu = 1, \sigma = 0.0001)$ which effectively 'fixes' the fluxes from these sectors. The boundary conditions sample from a truncated normal probability distribution function $\sim TN(\mu = 1, \sigma = 0.001)$ for each of the boundaries, and the model uncertainty hyperparameter from a uniform distribution bounded between 0.1 and 3.0. The model uncertainty is calculated as the sum in quadrature of the observational uncertainty and the model uncertainty - treated as the mean residual between pseudo-observations and forward-simulated mole fractions. A No-U-Turn (NUTS) sampler (Hoffman et al., 2014) for the a priori emissions and a slice sampler for estimating the model hyperparameter. The samplers used a total 8000 iterations (discarding the first 1000) with 3 chains running in parallel. A Gelman-Rubin diagnostic is used to check for parameter convergence in the chains.

3.2 CarbonTracker

This system is based on CarbonTracker Europe (CTE, van der Laan-Luijkx, 2017) and its Lagrangian version, CTE-Lagrange (He et al., 2018) and uses the ensemble Kalman Filter to solve for regional carbon fluxes. On global scales, for which CTE is designed, fossil fuel emissions have a small uncertainty, about 5%, and the main uncertainty in the global carbon budget is in the ocean and biogenic fluxes. On the contrary, on the regional scale of CTE-HR-inv (shown in Fig. 2), fossil fuel emissions are more expected to be more uncertain. This is because, whilst yearly, national fossil fuel use is quite well constrained, spatio-temporal upscaling of these fluxes adds uncertainty. Over the domain (Fig. 2), the ocean plays a minor role. Therefore, we have chosen to optimise both fossil fuel emissions and biogenic fluxes in CTEHR-inv. Although the system is in principle similar to CTE-Lagrange (He et al., 2018), some key differences are present. Besides a code overhaul, from the R programming language to Python, we use both different prior assumptions and a different state vector. These are explained in more detail below.

For the biosphere fluxes, we expect the errors in SiB4 to be land-use type specific. Additionally, SiB4 does not include forest age, and all pixels are assumed to be in steady-state (i.e. no net growth). To account for this, we assume errors in SiB4 to be forest age-related. Finally, different countries have different management regimes, resulting in different forest structures, which are not present in SiB4. We thus expect biosphere fluxes in different countries to behave differently. Based on these considerations, we optimise the net ecosystem exchange (NEE) per country per age-class. On each of these 503 parameters, we put an uncertainty of 2 umol/m^2/s. To limit the degrees of freedom, we apply a decay in temporal correlation with a half-time of 4 days. This synoptic time scale reflects the time scales in which weather patterns, and thus short-term variability in the biosphere, change. We have chosen to optimise an additive scaling factor on NEE, as prior NEE can be both positive and negative, depending on the season, and is very small in the shoulder seasons. A very small NEE signal requires a very large multiplicative scaling factor. These potential issues are resolved with the additive scaling factor.

Errors in total anthropogenic emissions stem mainly from errors in reporting and the use of the activity data. We therefore assume that these errors are unique for each of the 43 countries in the domain. Contrary to NEE, of which the sign can flip depending on the season, anthropogenic emissions are always positive. Therefore, we have chosen to use multiplicative scaling for the anthropogenic emissions. This choice also results in smaller uncertainties at times when total prior emissions are smaller. The errors in the reporting and activity data that inform the anthropogenic emissions are slowly varying parameters, such as energy efficiency of the industry. We expect these parameters to change only slowly over time. We therefore apply a large temporal correlation, with a half-time of two weeks. We put an uncertainty of 20% on the anthropogenic emissions. We take the background as the optimised 3D mole fractions from CTE2022. Although these mole fractions are optimally consistent with the atmospheric growth rate, they are not a perfect representation of the CO_2 distribution. We optimise the background additively over 8 regions, based on the wind-sector (NNE, NE, SE, SSE, SSW, SW, NW, NNW). We expect the errors in the background to vary over synoptic timescales and have thus chosen a temporal correlation with a half-time of 6 days and an uncertainty of 0.3 ppm.

4. Results

Here, a short summary of the findings from the mole fraction analyses and the flux analyses is given (see Appendices).

We showed that both RHIME and CTE show similar enhancements in atmospheric CO_2 mole fractions due to European fluxes. This indicates that both systems are sensitive to European fluxes in a similar manner. However, we also find that the two systems differ in their background mole fractions. This is shown in the figure below.

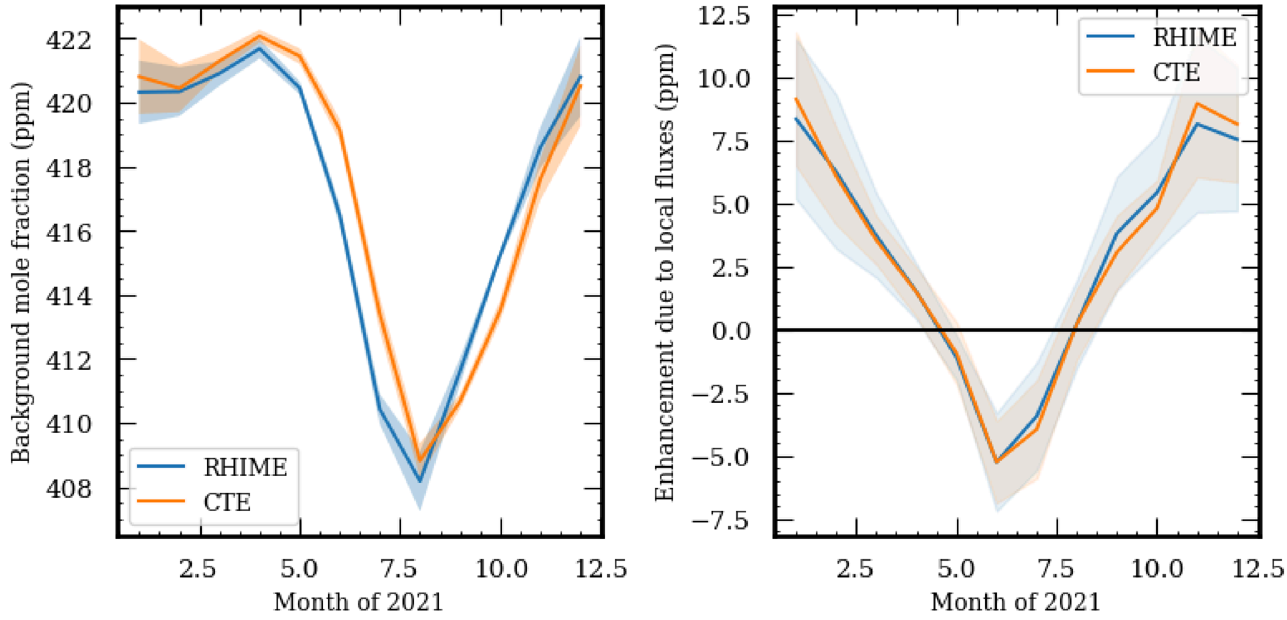


Figure 4.1: Left: Monthly mean background mole fractions simulated by the CTE and RHIME systems in 2021. Right: Monthly mean above or below-baseline enhancements due to regional fluxes in 2021.

Despite a similar sensitivity to atmospheric mole fractions over the domain, there are some differences in how the systems see the scenarios. For the two systems, the base case and the PTEN scenario (i.e. the scenario with the largest flux change) is shown below. A difference in the background is seen, but also on top of that a slight difference in enhancement.

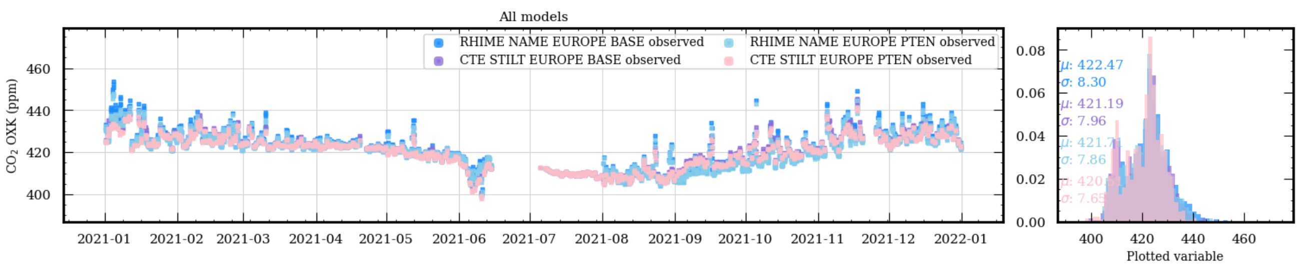


Figure 4.2: Mole fractions simulated by RHIME and CTE for the BASE and PTEN scenarios at the Ochsenkopf station in Germany (OXK). Right: histogram of the simulated mole fractions for all hours in 2021.

Not accounting for uncertainties and differences due to transport, and with known scenarios, we can discern the scenarios by looking at the difference in atmospheric mole fractions caused by only the fluxes, and attributing these differences to the scenarios manually. We see this in the following figure, in which the difference between the base case, and the scenarios is shown.

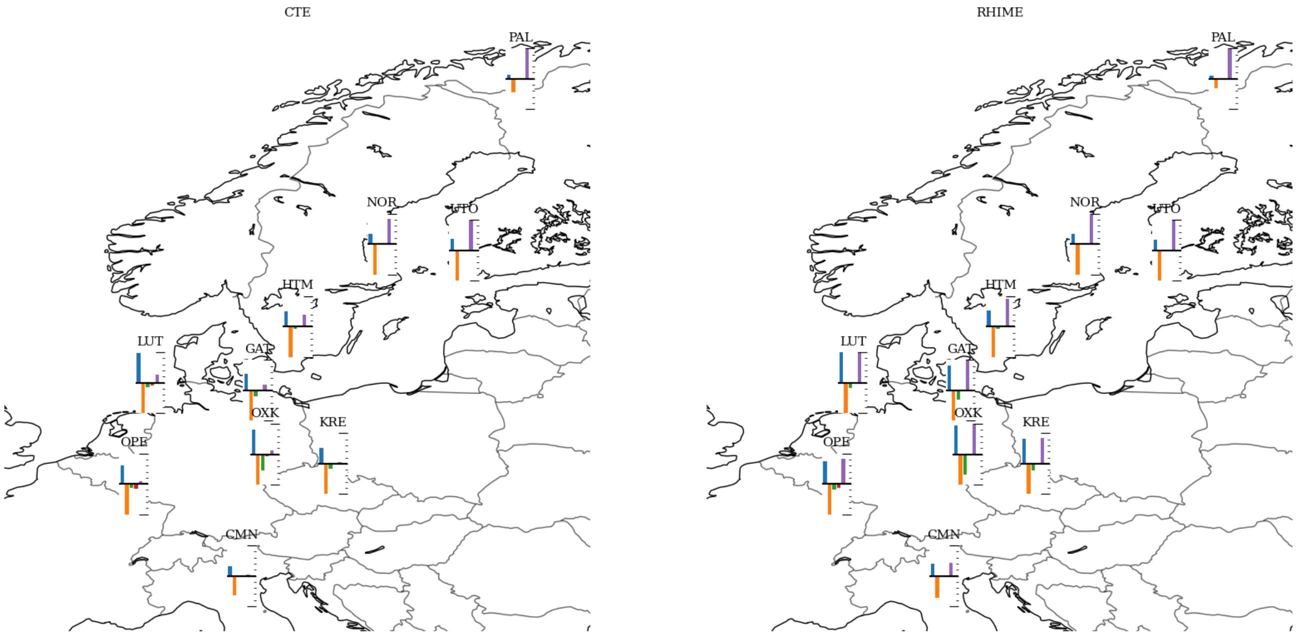


Figure 4.3: Mean mole fractions compared to the BASE scenario for CTE (left) and RHIME (right) for the five perturbation scenarios during winter (DJF). Deviations up to 0.5 ppm are shown. PTEN=blue, DFIN=purple, ATEN=green, HGER=orange, HFRA=red.

The figure shows the difference between the base scenario and the 5 perturbation scenarios in winter time. Knowing e.g. that PTEN is the largest increase in total flux, and that DFIN increases fluxes in the winter, we can attribute the two positive changes (blue and purple) to these scenarios. With purple being more distinct in the north of Europe, we (correctly) assume that purple is DFIN. For further explanation see the mole fraction results Appendix.

However, the differences in atmospheric mole fractions are small compared to the uncertainties and transport differences (the figure above shows differences up to 0.5 ppm, whereas we that background differences can be up to 3 ppm) that it is hard for the atmospheric inversions to attribute the differences to the correct scenario, given differences in background and uncertainties in atmospheric transport.

The differences in fluxes due to the different background and transport model are shown in the figure below. It shows the "true" fluxes, which both models use as prior, and the fluxes that both models found in the BASE scenario (i.e. no perturbations applied to the "true" fluxes). The difference between the black and both blue bars is indicative of the uncertainty in posterior fluxes due to transport only. The largest difference is for the UK, where RHIME derives fluxes approximately twice as large as truth, when using CTE pseudo-observations. This is potentially because Mace Head Ireland was not included in this iteration of the study, which may have influenced RHIME's ability to derive the UK baseline. For the other countries, differences with the true fluxes are on the order of tens of Tg/yr.

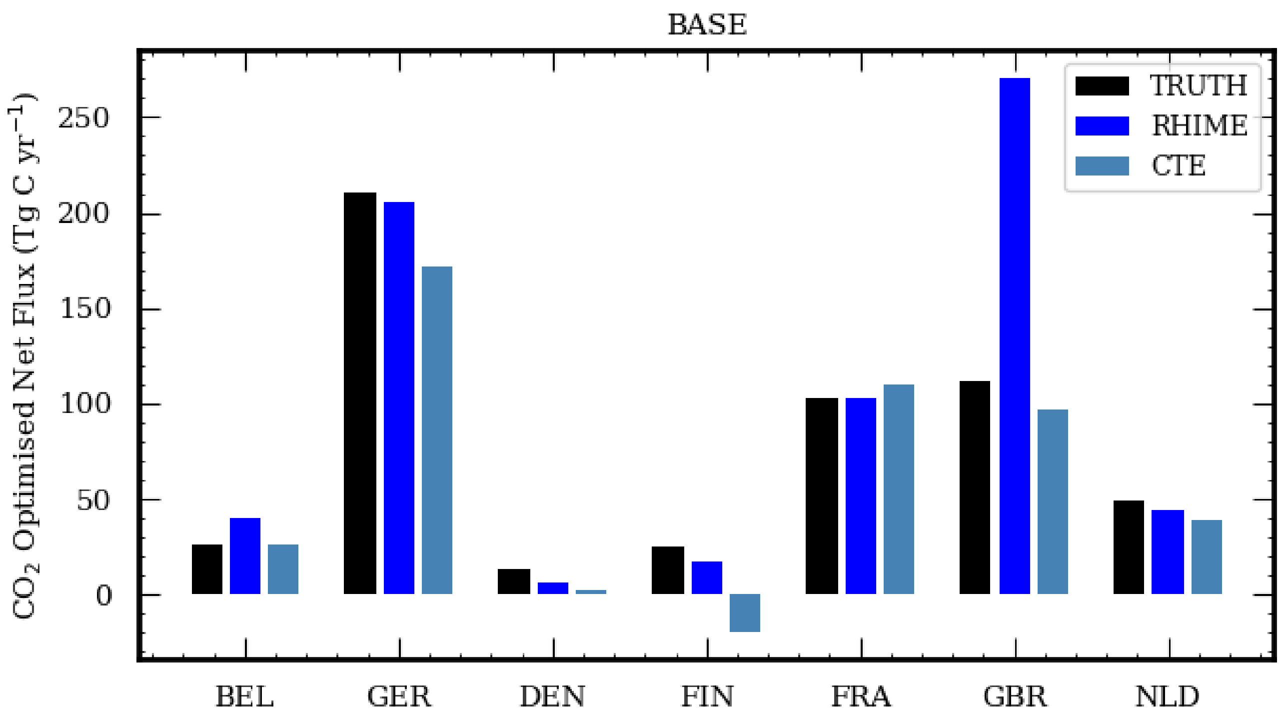


Figure 4.4: 2021 mean posterior fluxes for key European countries for the base scenario. No difference between the black and blue bars would indicate a perfect model

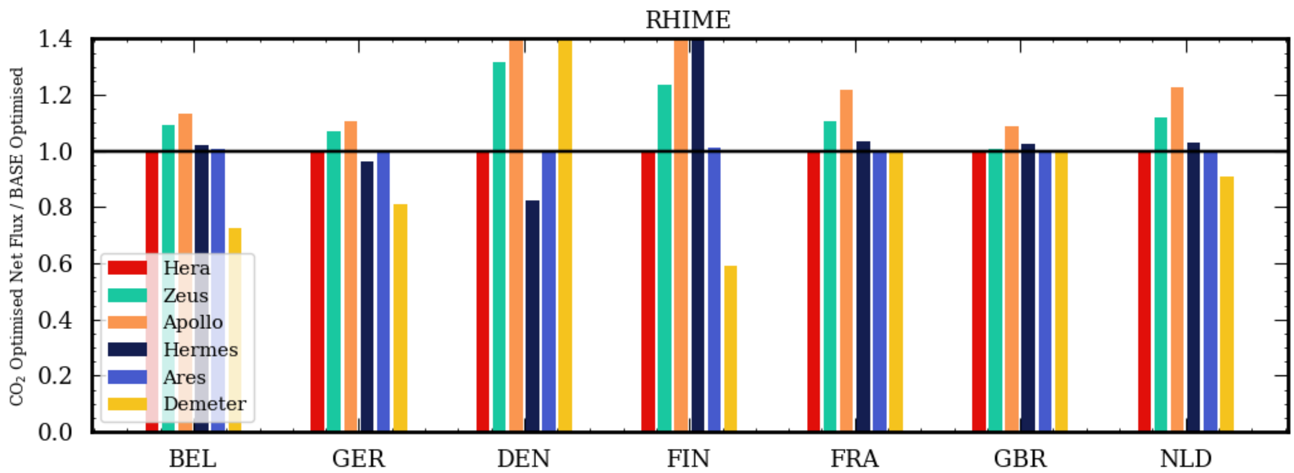


Figure 4.5: 2021 mean posterior fluxes for key European countries for each of the PARIS verification challenge flux scenarios using RHIME, relative to the optimised BASE scenario. Scenario names are given according to their unblinded definitions, which are provided in Section 2.1.

As an example of the outcome of the flux inversions, the above figure shows the mean fluxes for key European countries for each of the PARIS verification challenge flux scenarios using RHIME, relative to the optimised BASE scenario. The scenario labelled "Demeter" has reduced fluxes compared to the base most countries. Therefore, this was identified as the PTEN scenario. However, other scenarios are more difficult to discern, with less clear patterns between countries. This finding highlights the challenges in distinguishing between scenarios based on the available data. Results for CTE and more in depth analysis is presented in the Appendix.

5. Outlook

5.1 Magnitude of perturbations

We found that the magnitude of the fluxes estimated by the systems differs significantly from the truth in some countries. We attribute this to differences in transport used, but also to the challenge of discerning anthropogenic fluxes using only CO_2 mole fraction data.

We find strong negative correlation in our solutions between the optimised fossil fuel and biosphere fluxes in our inversions. This is seen in Figure 5.1, in which the fossil and biogenic fluxes from Germany for all 150 realisations (members) from the CTE inversion system are shown. This shows that realisations with high fossil fuel emissions also have low biogenic emissions (=higher uptake), and vice versa.

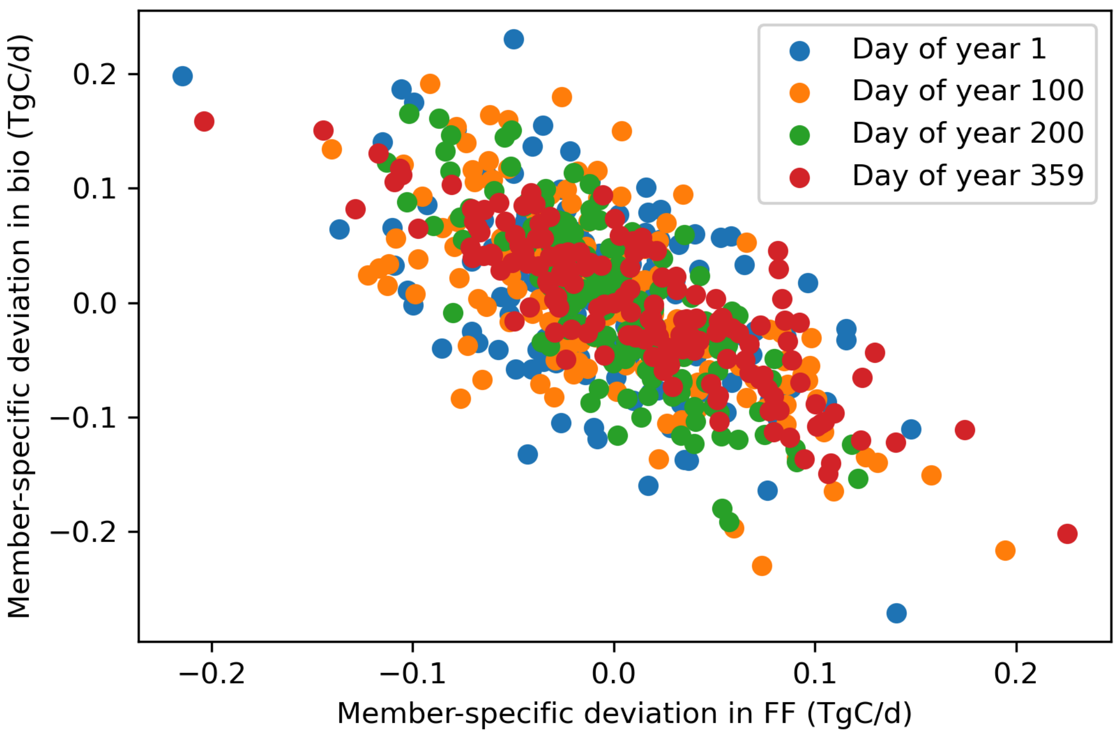


Figure 5.1: Deviations in fossil fuel flux (x-axis) and biogenic flux (y-axis) for each of the 150 realisations of a CTE simulation

This finding demonstrates the challenge of attempting to disentangle fossil and biosphere fluxes with mole fraction data alone. Additional tracers, such as radiocarbon or oxygen, might help to improve the separation of these fluxes. As the current project also focuses on the additional value of oxygen, we aim to use oxygen to distinguish between biogenic and fossil fuel fluxes in a next installment of the verification games.

5.2 Discernability of the scenarios

The verification challenge aimed to test whether two inverse methods could detect changes in CO_2 sources and sinks in Europe based on pseudo-observations generated by another transport model. Most scenarios, despite representing relatively large flux perturbations, were challenging to attribute to a particular scenario, given the uncertainties introduced due to model transport. This shows the challenges with using atmospheric CO_2 for monitoring and verification. Improvements may be possible using novel tracers or improvements in inversion configurations. These factors will be tested in the next iteration of the challenge.

5.3 Limitations of the current version of the verification challenge

Whilst performing the first verification challenge, we encountered problems or discovered limitations in the protocol that should be taken into consideration when next developing the verification challenge. These are:

- In the DFIN scenario, we wanted to increase uptake in Finland. For this, we multiplied NEE by 2. However, as CTE-HR is neutral (slightly positive) in Finland in 2021, this resulted in an extra source. Additionally, this increased TER at night, resulting in slightly larger mole fractions (due to [the opposite of] the diurnal rectifier effect). For the next round, we agreed to multiply GPP by 1.2 (20% increase).
- During the year in which we did these simulations, a bug was introduced in the data used to create the fluxes in CTE-HR. This resulted in an update to CTE-HR. The fluxes for PTEN were made after this update. Therefore, these fluxes are different in Ukraine and Albania (as seen in Figure 3). For a next round, we will unify these fluxes. However, we found that these countries have only a minor influence on the atmospheric mole fractions at the ICOS stations, as most stations are more than 1500km from these countries, and the dominant wind direction in Europe is from the west.
- The colocated flux processes mean it is challenging to discern between anthropogenic and NEP in the optimisation using mole fraction data alone. Inversions generally found strong correlations between fossil fuel and biogenic fluxes.
- Differences in the modelling capacities of the different groups meant certain measurement sites were not simulated. To avoid confusion and ensure that the inverse model setups are close, the sites that should be included should be agreed upon in advance.

References

Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., & Treanton, K. (2012). A synthesis of carbon dioxide emissions from fossil-fuel combustion. Biogeosciences, 9(5), 1845–1871. https://doi.org/10.5194/bg-9-1845-2012

Basu, S., Miller, J. B., & Lehman, S. (2016). Separation of biospheric and fossil fuel fluxes of CO2 by atmospheric inversion of CO2 and 14CO2 measurements: Observation System Simulations. Atmospheric Chemistry and Physics, 16(9). https://doi.org/10.5194/acp-16-5665-2016

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., … Zheng, B. (2023). Global Carbon Budget 2023. Earth System Science Data, 15(12). https://doi.org/10.5194/essd-15-5301-2023

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K. R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., & Weiss, R. F. (2014). Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. Atmospheric Chemistry and Physics, 14(8). https://doi.org/10.5194/acp-14-3855-2014

He, W., Velde, I. R. Van Der, Andrews, A. E., Sweeney, C., Miller, J., Tans, P., Laan-Luijkx, I. T. Van Der, Nehrkorn, T., Mountain, M., Ju, W., Peters, W., Chen, H., Van Der Velde, I. R., Andrews, A. E., Sweeney, C., Miller, J., Tans, P., Van Der Laan-Luijkx, I. T., Nehrkorn, T., … Chen, H. (2018). CTDAS-Lagrange v1.0: A high-resolution data assimilation system for regional carbon dioxide observations. Geoscientific Model Development, 11(8), 3515–3536. https://doi.org/10.5194/gmd-11-3515-2018

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., & Pätz, H. W. (2010). The global chemistry transport model TM5: Description and evaluation of the tropospheric chemistry version 3.0. Geoscientific Model Development, 3(2). https://doi.org/10.5194/gmd-3-445-2010

Hoffman, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15.

Jones, A., Thomson, D., Hort, M., & Devenish, B. (2007). The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. In Air Pollution Modeling and Its Application XVII (pp. 580–589). Springer US. https://doi.org/10.1007/978-0-387-68854-1_62

Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F. F., Quéré, C. Le, Le Quéré, C., Dou, X., Liu, Z., Pickers, P., Ciais, P., Patra, P. K., Chevallier, F. F., & Le Quéré, C. (2021). Gridded fossil CO2 emissions and related O2 combustion consistent with national inventories 1959–2018. Scientific Data, 8(1), 1–23. https://doi.org/10.1038/s41597-020-00779-6

Lunt, M. F., Manning, A. J., Allen, G., Arnold, T., Bauguitte, S. J.-B., Boesch, H., Ganesan, A. L., Grant, A., Helfter, C., Nemitz, E., O'Doherty, S. J., Palmer, P. I., Pitt, J. R., Rennick, C., Say, D., Stanley, K. M., Stavert, A. R., Young, D., & Rigby, M. (2021). Atmospheric observations consistent with reported decline in the UK's methane emissions (2013–2020). Atmospheric Chemistry and Physics, 21(21), 16257–16276. https://doi.org/10.5194/acp-21-16257-2021

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., & Derwent, R. G. (2011). Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach. Journal of Geophysical Research Atmospheres, 116(2), 1–19. https://doi.org/10.1029/2010JD014763

Manning, A. J., Redington, A. L., Say, D., O'Doherty, S., Young, D., Simmonds, P. G., Vollmer, M. K., Mühle, J., Arduini, J., Spain, G., Wisher, A., Maione, M., Schuck, T. J., Stanley, K., Reimann, S., Engel, A., Krummel, P. B., Fraser, P. J., Harth, C. M., ... Arnold, T. (2021). Evidence of a recent decline in UK emissions of hydrofluorocarbons determined by the InTEM inverse model and atmospheric measurements. Atmospheric Chemistry and Physics, 21(16), 12739–12755. https://doi.org/10.5194/acp-21-12739-2021

Saboya, E., Manning, A. J., Levy, P., Stanley, K. M., Pitt, J., Young, D., Say, D., Grant, A., Arnold, T., Rennick, C., Tomlinson, S. J., Carnell, E. J., Artioli, Y., Stavart, A., Spain, T. G., O'Doherty, S., Rigby, M., & Ganesan, A. L. (2024). Combining Top-Down and Bottom-Up Approaches to Evaluate Recent Trends and Seasonal Patterns in UK N 2 O Emissions. Journal of Geophysical Research: Atmospheres, 129(14). https://doi.org/10.1029/2024JD040785

van der Laan-Luijkx, I. T., van der Velde, I. R., van der Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Fang Zhang, H., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., & Peters, W. (2017). The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: Implementation and global carbon balance 2001-2015. Geoscientific Model Development, 10(7). https://doi.org/10.5194/gmd-10-2785-2017

van der Woude, A. M., Peters, W., Joetzjer, E., Lafont, S., Koren, G., Ciais, P., Ramonet, M., Xu, Y., Bastos, A., Botía, S., Sitch, S., de Kok, R., Kneuer, T., Kubistin, D., Jacotot, A., Loubet, B., Herig-Coimbra, P. H., Loustau, D., & Luijkx, I. T. (2023). Temperature extremes of 2022 reduced carbon uptake by forests in Europe. Nature Communications, 14(1), 6218. https://doi.org/10.1038/s41467-023-41851-0

Appendix A: Station ID definitions

- BIR: Birkenes, Norway
- BSD: Bilsdale, UK
- CBW: Cabauw, Netherlands
- CMN: Monte Cimone, Italy
- GAT: Gartow, Germany
- HEL: Helgoland, Germany
- HFD: Heathfield, UK
- HPB: Hohenpeissenberg, Germany
- HTM: Hyltemossa, Sweden
- HUN: Hegyhatsal, Hungary
- KIT: Karlsruhe, Germany
- KRE: Kresin u Pacova, Czech Republic
- LIN: Lindenberg, Germany
- LUT: Lutjewad, Netherlands
- NOR: Norunda, Sweden
- OPE: Observatoire perenne de l'environment, France
- OXK: Ochsenkopf, Germany
- PAL: Pallas, Finland
- RGL: Ridge Hill, UK
- SAC: Saclay, France
- SMR: Hyytiala, Finland

- SSL: Schauinsland, Germany
- STE: Steinkimmen, Germany
- TAC: Tacolneston, UK
- TOH: Torfhaus, Germany
- TRN: Trainou, France
- UTO: Uto - Baltic Sea, Finland
- WAO: Weybourne, UK
- WES: Westerland, Germany

PARIS Work Package 6: Separation of Biosphere and Fossil Fuel Carbon Dioxide Flux

Milestone 25: Synthetic data inversions complete and modified fluxes unblinded

Appendix B: Overview of mole fraction results

October 2025

Auke van der Woude[1], Eric Saboya[2], Daan Kivitss[1], Remco de Kok[1], Brendan Murphy[2], Prasad Sutar[2], Ingrid Luijkx[1], Wouter Peters[1], Matthew Rigby[2]

[1]Wageningen University (WU)

[2]University of Bristol (UNIVBRIS)

---

1. Atmospheric Mole Fraction Results from the Verfication Challenge

The flux results presented in the `results_flux` notebook are key to identifying the different flux scenarios and comparing the net flux totals inferred by each inverse model. The a priori atmospheric CO2 mole fractions at the different sites, however, can tell us about differences between the atmospheric transport models and modelled baseline mole fractions at each site. The optimised, or a posteriori, atmospheric CO2 mole fractions at each site demonstrate how each inverse model optimises the mole fractions at each site.

This notebook uses the intercomparison tool, FLUXY, as the underlying library for comparing the different sets of mole fractions at certain sites.

1.1 Baseline mole fraction comparisons

Accurate baseline mole fractions at each site are key to ensuring the inverse models can correctly infer the true flux perturbations in the model domain. Mole fraction baselines that are too low might lead to fluxes being over-estimated by the inverse model, and mole fraction baselines that are too high might lead to under-estimated fluxes being inferred. Furthermore, large differences in the modelled baseline mole fractions at each site could lead to biases between the fluxes inferred from the inverse models. We therefore begin our comparison of the mole fractions by considering the different modelled baseline mole fractions at certain key measurement stations in the Europe model domain.

We first do this over the full domain, so for all stations. As fluxy is not made for this, we temporarily use other packages

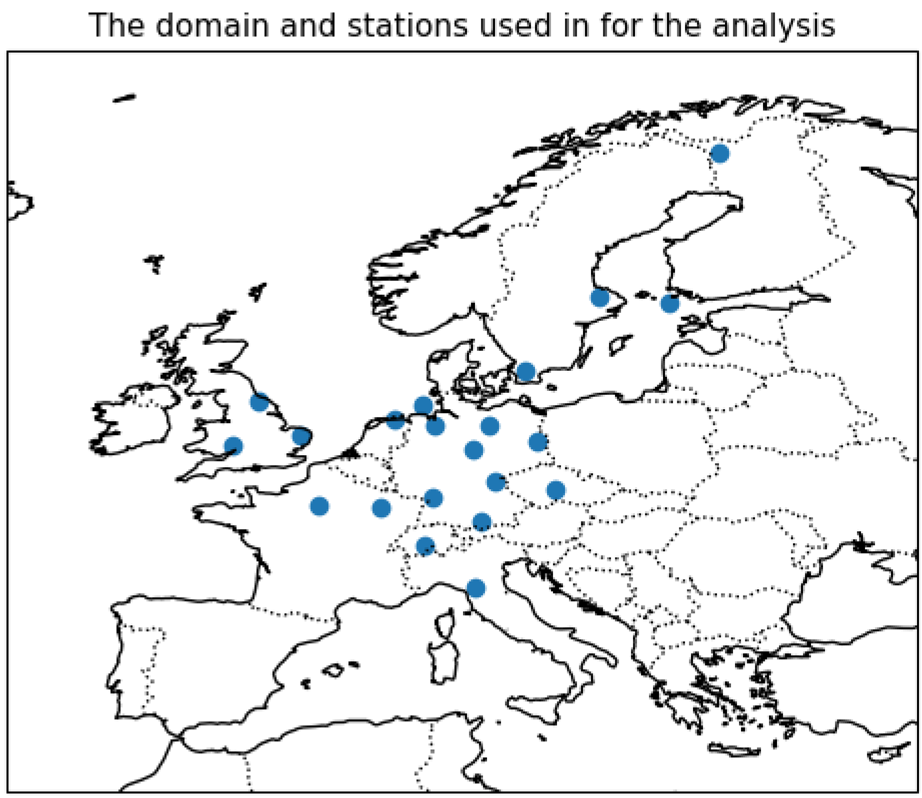


Figure B1: The stations (blue dots) used in the verification games

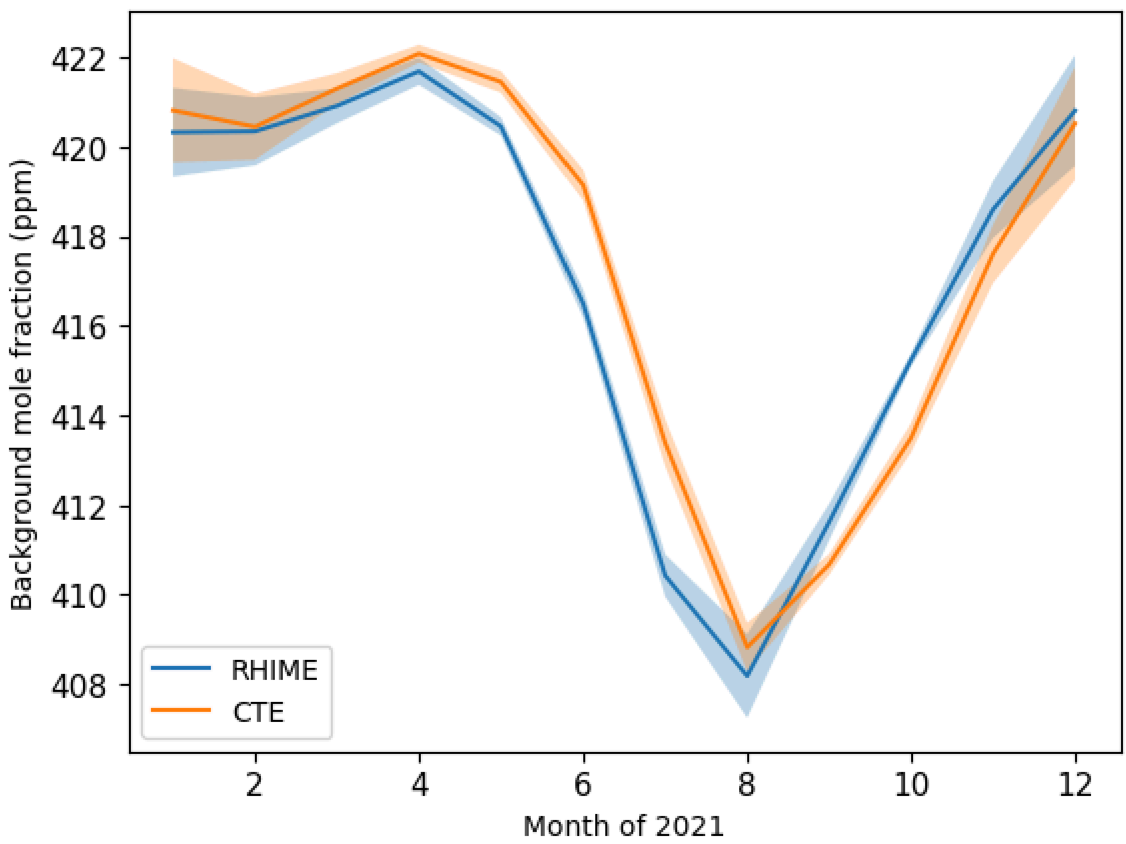


Figure B2: Monthly mean time series over all stations of the background mole fraction of CO2 for both systems used

In the above figure, the orange and blue line show the mean background mole fractions at European stations from the RHIME and CTE systems respectively. The inter-station spread is shown in the opaque band around the solid line for each of the systems. From this figure, we see that although the background mole fractions are similar, there is a difference of about 1 ppm throughout the year. This can be attributed to different domains in the transport model, but also different wind fields used for the transport. Having noted this, we next look at the European mean enhancement due to local fluxes

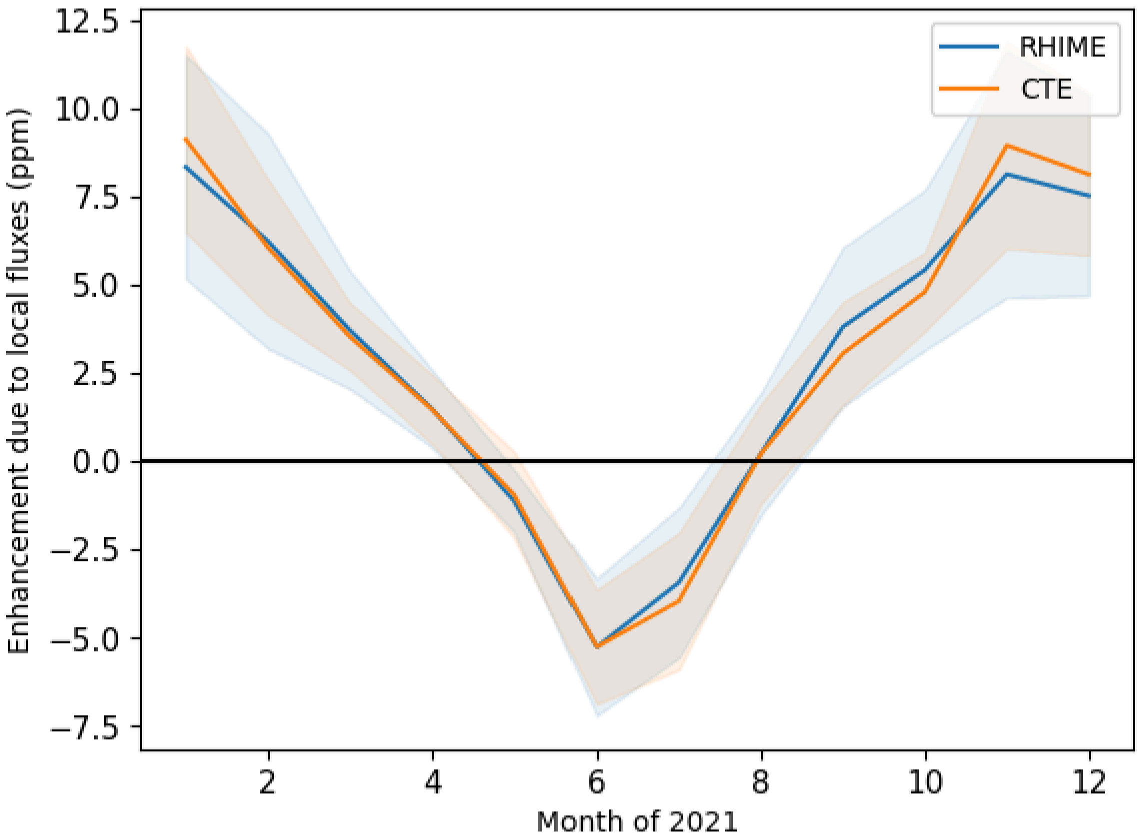


Figure B3: Monthly mean time series of the enhancements due to local (i.e. European) fluxes averaged over all stations in the domain for both systems

In the above figure, the mean mole fraction enhancement due to local fluxes (i.e. within the European domain) is shown. The blue line shows RHIME and the orange line shows CTE, with the opaque bands showing the inter-station spread. From this, we see that the local enhancement on the stations is very similar for each of the systems. This local enhancement is caused only by fluxes in Europe, and indicates that the two systems shown here are very similar in their sensitivity to European fluxes. Note that the positive enhancement in winter is due to the dominance of respiration and fossil fuels, and the negative enhancement is due to uptake of carbon by vegetation. Having looked at the Europe-mean sensitivity of the systems, we will look at individual stations to further examine the differences.

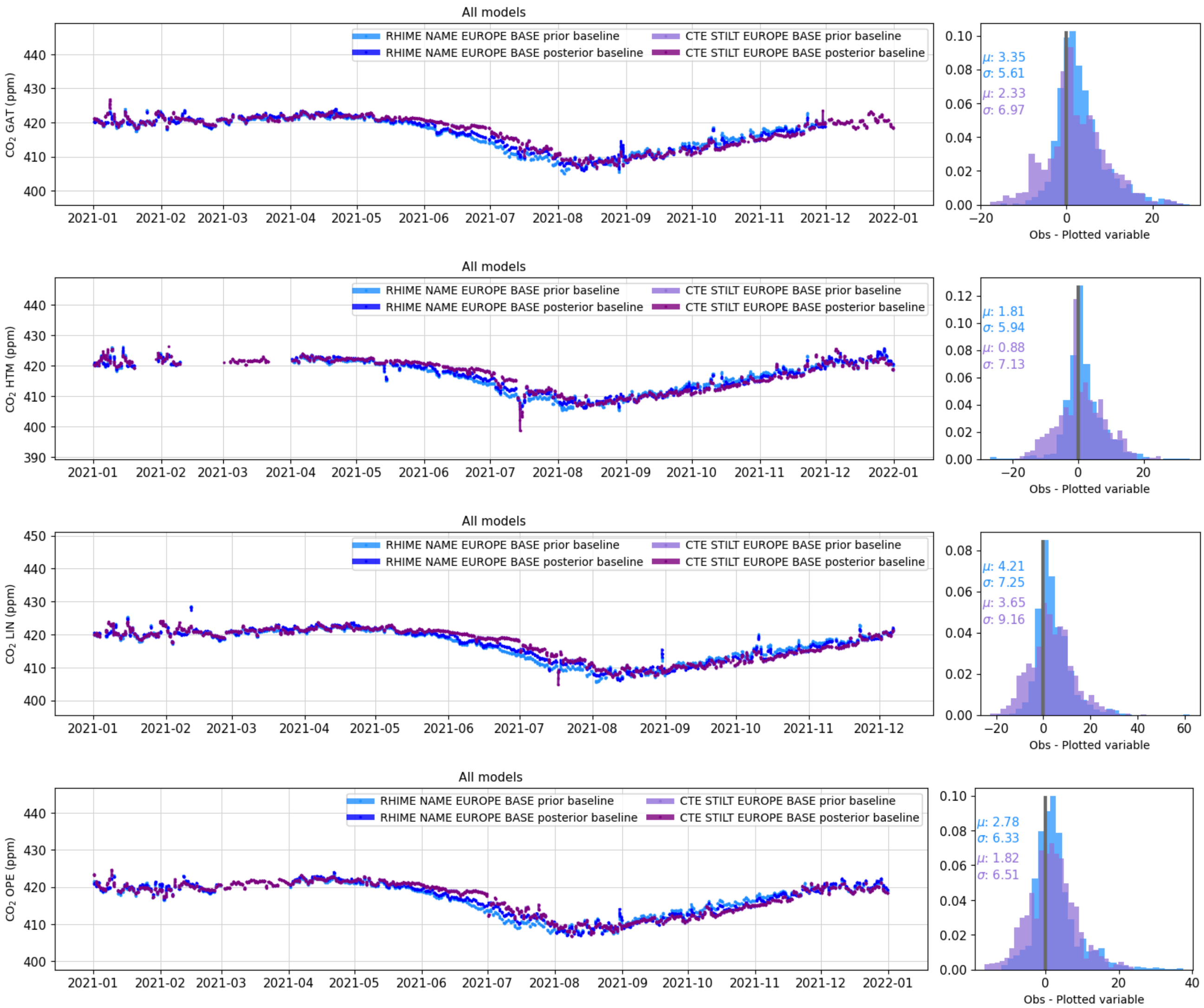


Figure B4: The 2021 a priori and optimised baseline mole fractions from RHIME and CTE at key measurement stations in the model domain.

1.2 Pseudo-observation comparison

Having established differences between the differnt atmospheric baseline mole fractions at the measurement stations of interest, we can now compare the impact of the two largest perturbation scenarios (ATEN -a 10% increase in anthropogenic emissions across the domain- and PTEN -shutting down the highest 10% emitting power plants in the model domain) on the simulated mole fractions. The ''observed'' atmospheric mole fraction at each site are those produced by the atmospheric transport model used by the other inverse model. In other words, the pseudo-observations used in RHIME are the fluxes transported by STILT along with the CTE baseline mole fractions. And the pseudo-observations used in CTE are the fluxes transported by NAME along with the RHIME baseline mole fractions (plus a bias mole fraction term accounting for the contributions from the model domain outside of Europe).

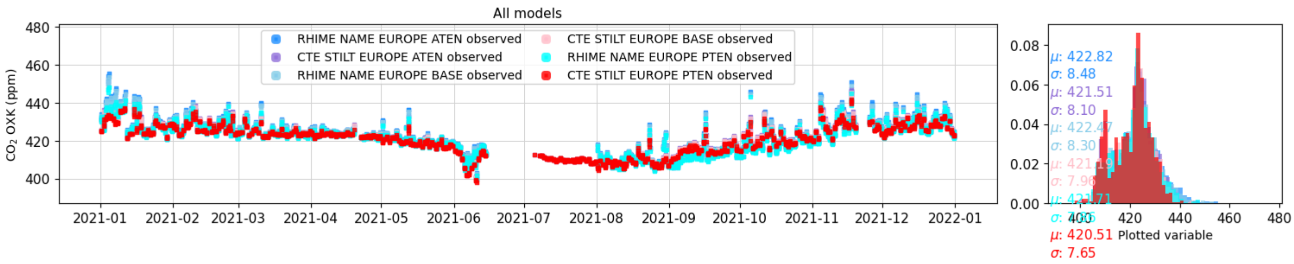


Figure B5: The 2021 pseudo-observations for the ATEN, BASE and PTEN scenarios at key measurement stations.

1.3 Verification games

Having noted the difference at site and European level, we can play some manual verification games. Here, we plot the change in atmospheric mole factions due to the change in fluxes in the scenario for a few stations. We plot only wintertime fluxes, as in winter fossil fuel signals are higher. This thus makes it a bit easier for us to detect the signals. The goal is to link the change in mole fraction to each scenario. Note that here, transport differences between the transport models are ignored. Also note that the y-axes are cut off at +/- 0.5 ppm for visibility. This 0.5 ppm is smaller than the uncertainty generally put on atmospheric CO2 observations.

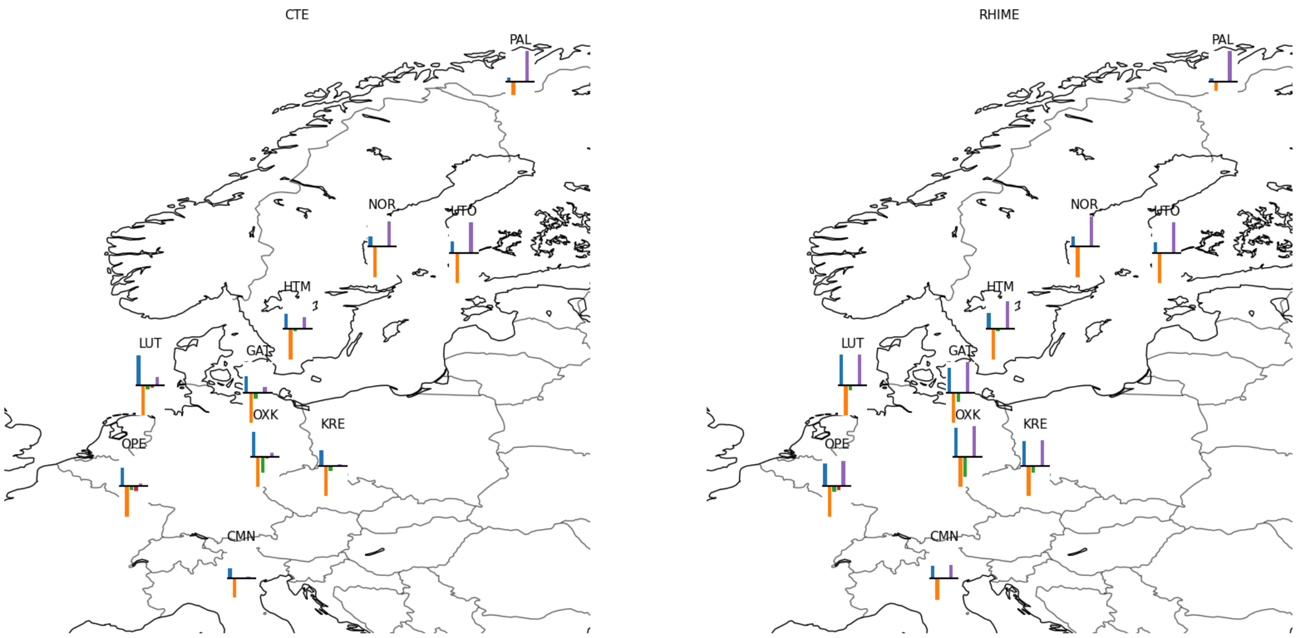


Figure B6: Blinded wintertime-mean atmospheric mole fraction enhancements compared to the base for the 5 scenarios for CTE (left) and RHIME (right). The bar plots are cut off at +/- 0.5 ppm

We see large positive differences in the blue scenario, and large negative differences in the orange scenario. Green and red have small differences, and purple has large positive differences in the north of Europe. Thus, the blue scenario must be one in which fluxes are increased by a large fraction, or uptake is reduced. Conversely, in the orange scenario, biosphereric uptake throughout Europe has to be increased, or fossil fuel emissions reduced. As we see the influence of the purple scenario decrease towards the south of Europe, we know the flux perturbation has to be in the north.

Knowing the scenarios and their relative influence on the total flux, we can thus see that blue must be ATEN, orange PTEN and purple DFIN. Note here that in the DFIN scenario, the hourly NEE over Finland was multiplied by 2. This increases all biogenic fluxes. As the biosphere in wintertime is a source of CO2, this source is also multiplied by two, increasing atmospheric mole fractions.

In both transport models, the green scenario mainly shows in Germany. As we know one of the two remaining scenarios has to be HGER, we have to presume that green is HGER, and thus red (only very slightly visible at OPE) is HFRA. The figure below shows us that this is, indeed, correct. thus, with a limited amount of known possible scenarios, we can identify the scenario by the response in the atmosphere.

Note that the differences between the models are generally small, with the only discernible differences in the purple bars. Finally, note that the differences are small compared to the transport and observation uncertainty, which are generally put at ~2 ppm, and thus 4 times larger than the y-axis cutoff shown here.

Figure B7: The legend (i.e. result) from the manual verification games in Figure 1.6.

In the flux results notebook we analyse the fluxes that the inverse systems find, given the observations used for the previous plots, but without knowing the possible scenarios

PARIS Work Package 6: Separation of Biosphere and Fossil Fuel Carbon Dioxide Flux

Milestone 25: Synthetic data inversions complete and modified fluxes unblinded

Appendix C: Overview of flux inversion results

October 2025

Eric Saboya[1], **Auke M. van der Woude**[2], **Daan Kivits**[2], **Remco de Kok**[2], **Brendan Murphy**[1], **Prasad Sutar**[1], **Ingrid Luijkx**[2], **Wouter Peters**[2], **Matthew Rigby**[1]

[1]University of Bristol (UNIVBRIS)

[2]Wageningen University (WU)

---

1. Flux Results from the Verfication Challenge

The set of optimised CO2 fluxes inferred by each inverse model allow us to evaluate the capabilities of the two inverse models (and the 2021 UK and European measurement network). We can determine the capability of an inverse model by assessing whether we can attribute the set of optimised fluxes to the correct flux scenario, whether the same conclusions about the flux scenarios can be drawn by each inverse model, and whether the set of optimised CO2 fluxes from each inverse model are statistically significantly different.

The optimised fluxes for certain European countries are key when it comes to attributing the different flux scenarios. The flux perturbation scenarios 'HFRA' (halving industry-related anthropogenic emissions in France), 'HGER' (halving transport emissions in Germany) and 'DFIN' (doubling forest carbon uptake in Finland) are country specific. Whilst the flux perturbation scenarios 'ATEN' (a 10% increase in anthropogenic emissions in Europe) and 'PTEN' (shutting down the top 10% emitting power plants in Europe) are domain wide, the pertubation from these scenarios will be more pronounced in certain countries.

This notebook presents the flux results from the blind verification challenge for RHIME and CTE. As the results are very similar for both systems, we focus our analysis of the blind verificatino games on RHIME, but show the results for CTE as well

We first analyse if the systems find similar total fluxes for the different countries by plotting the 'base' scenario (i.e. no flux perturbation). We expect both systems to show fluxes that are similar to the truth

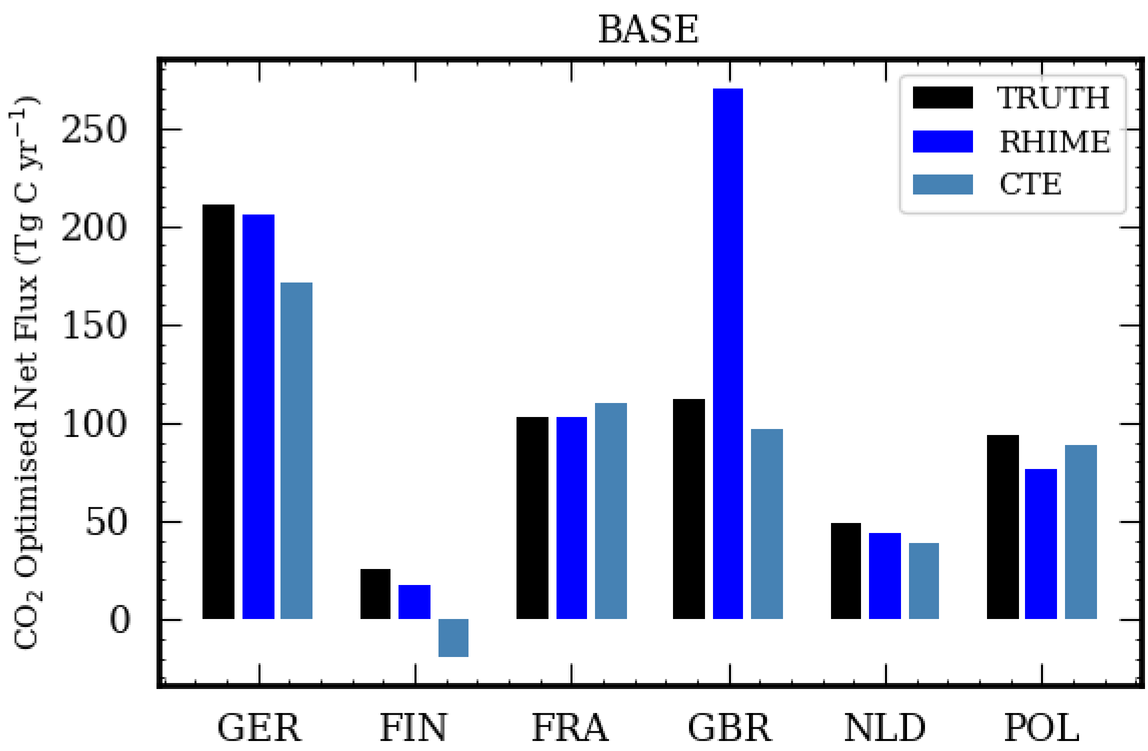


Figure C1.1: The 2021-averaged optimised fluxes from RHIME and CTE blinded for the six flux scenarios (BASE plus five perturbed scenarios). From the figure, we see that the optimised net flux (fossil fuel + biosphere) for most countries is similar to the truth. Some key differences should be noted however. For Finland, CTE finds a net sink and for Great Brittain, Rhime finds a too large source of CO2. Note that these fluxes are estimated using simulated mole fractions from the other system, so these differences can be seen as a metric for the cumulative sum of all uncertainties and choices in the models. Having seen the differences in the mole fractions in this notebook, we attribute the differences mainly to differences in the background used by the systems. Next, we use RHIME to analyse the blind verification games.

The estimated fluxes for each (blinded) scenario for both systems are shown below.

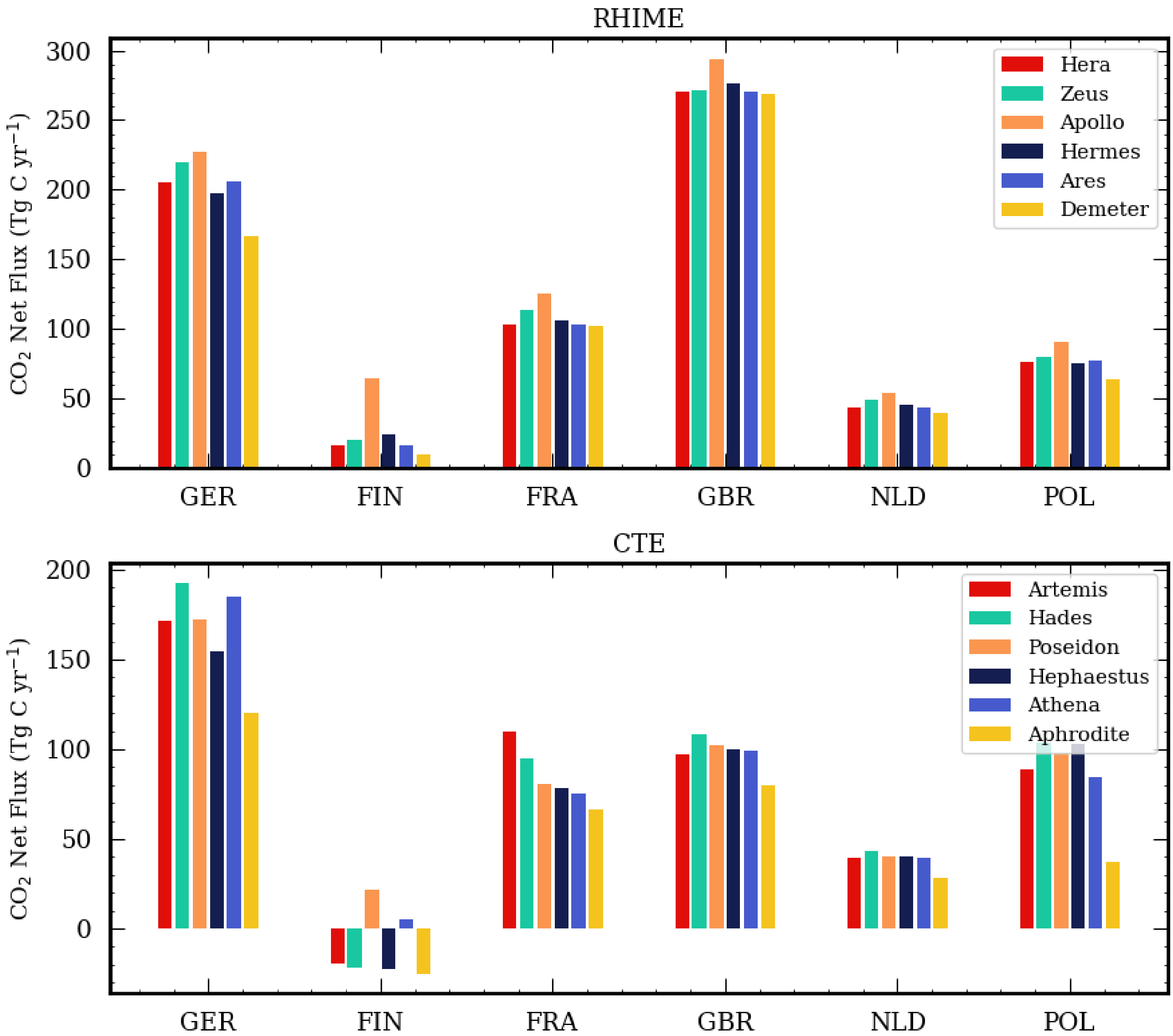


Figure C1.2: The 2021-averaged optimised fluxes from RHIME and CTE blinded for the six flux scenarios (BASE plus five perturbed scenarios).

To better compare different countries, we divide the fluxes by the base (i.e. no flux perturbation), which is shown in the next figure.

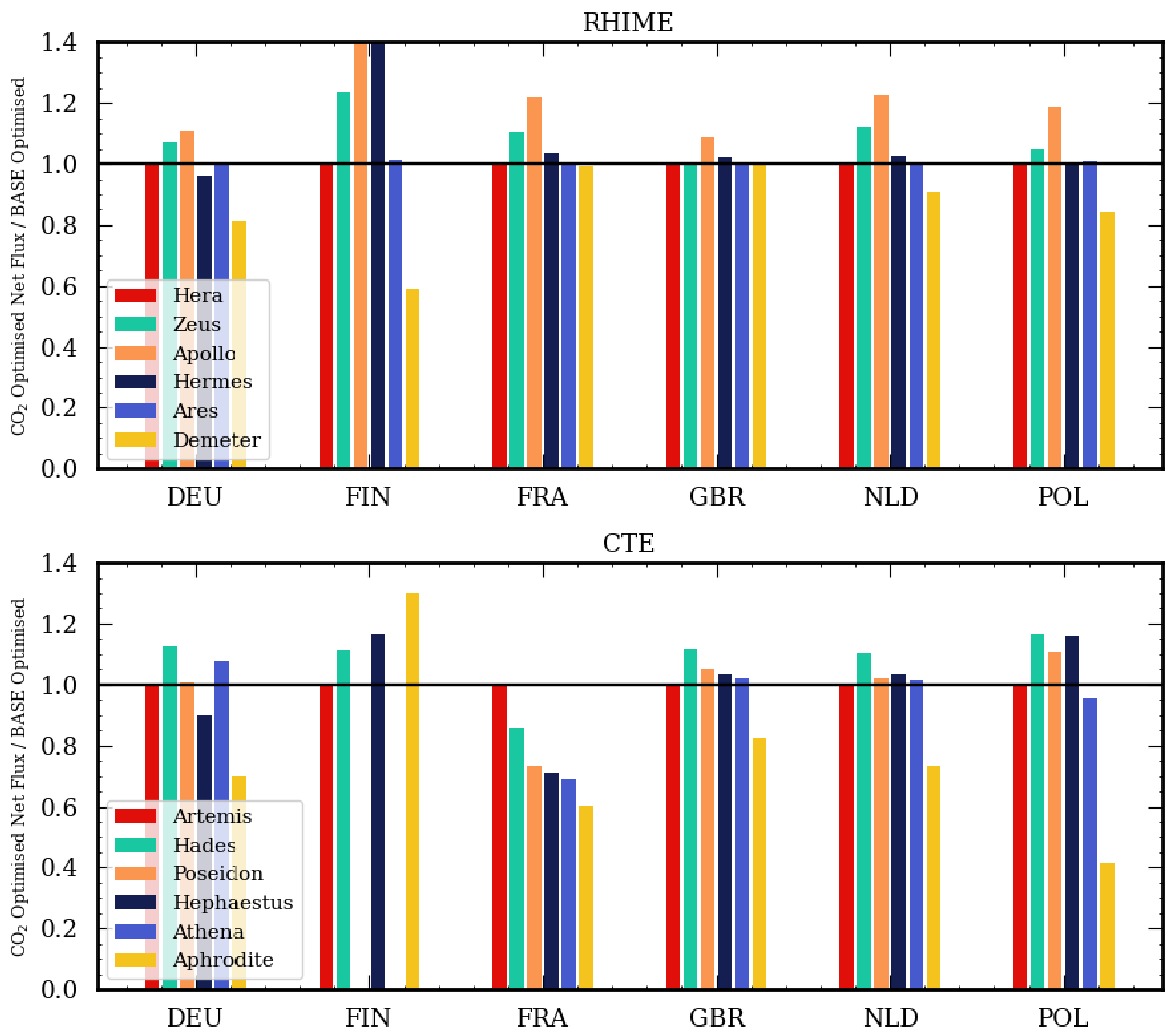


Figure C1.3: The 2021-averaged optimised fluxes divided by the BASE optimised flux from RHIME and from CTE. Flux scenarios are blinded for the six flux scenarios (BASE plus five perturbed scenarios).

Using Figures C1.2 and C1.3 we can attribute the blinded flux scenarios. We can deduce Demeter is clearly PTEN since this scenario has a perturbation with an optimised flux-to-optimised base ratio less than 1 across all countries. Hermes is the only other scenario to have an optimised flux-to-optimised base ratio less than 1 in Germany. However, Hermes does have ratios slightly higher than 1 for Belgium, France, Great Britain and the Netherlands but not high enough to conclusively be the ATEN scenario. Therefore Hermes could be HGER.

Apollo and Zeus have optimised flux-to-optimised base ratios higher than 1 across these selected countries. One of these must be ATEN and the other DFIN since these are the only positive perturbation scenarios. From Fig. C1.2. we see that Apollo has the largest perturbation for Finland so we could tenatively match Hera to DFIN, which then would make Zeus the ATEN scenario. However, this decision is based on the results from one country. Ignoring Finland would likely suggest that Apollo is ATEN and DFIN would be Zeus.

Ares and Hera must be BASE or HFRA. Looking at France we see that both Ares and Hera are exactly the same. This suggests the HFRA perturbation could not be detected by RHIME using the current measurement network. We see that for the other countries Ares and Hera are also exactly the same, except in Finland and Belgium which have an almost inperceptible difference where Ares is slighly higher than Hera. However, this difference is so small we cannot reasonably attribute these blinded scenarios.

In summary,

1. Apollo --> _DFIN*_
2. Ares --> could not be reasonably determined
3. Hera --> could not be reasonably determined
4. Hermes --> _HGER_
5. Demeter --> PTEN
6. Zeus --> _ATEN*_

Where italics indicate a tentative result, and an asterisk suggests these could be swapped around

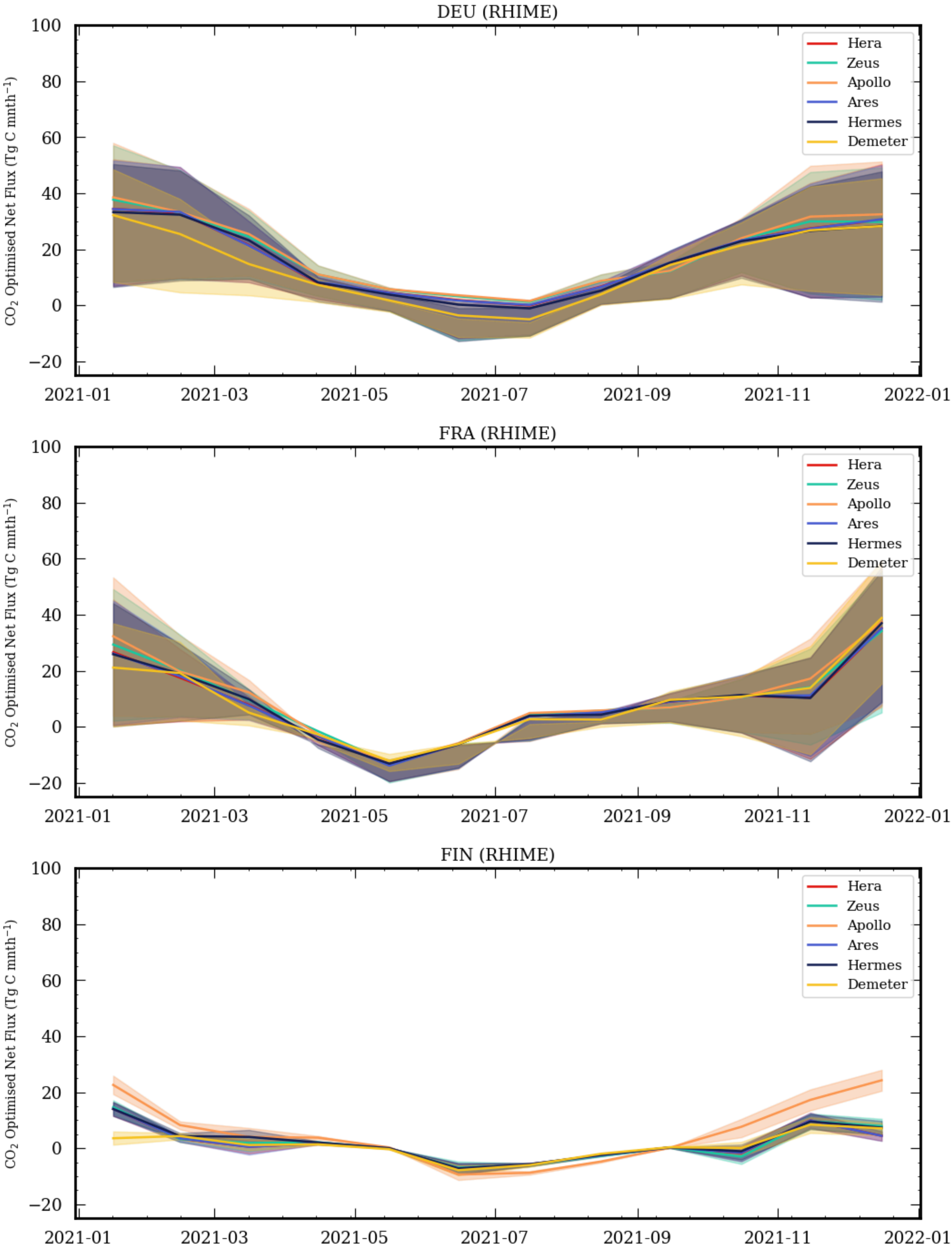
---

2. Statistical Significance

When attributing the optimised fluxes to the different scenarios we did not consider the uncertainties of the optimised fluxes. Figure 3.4 shows the 2021 monthly timeseries of the optimised fluxes with their 68% confidence intervals for each flux scenario for Germany, France and Finland - the countries with their own specific perturbations. The optimised fluxes from RHIME have overlapping uncertainties for most scenarios in each of these countries. We see it is particularly challenging to distinguish the different scenarios during the summer months where there is considerable overlap between the optimised flux values. This is not surprising since the flux perturbations were anthropogenic (apart from DFIN) and the terrestrial biosphere fluxes during the summer are much larger in magnitude than the anthropogenic fluxes. We observe that there are differences between the scenarios in the winter months, but not statistically different (i.e. overlapping uncertainty ranges).

In Finland, Apollo and Demeter are significantly different from the other flux perturbation scenarios in January - and Apollo in December - which could indicate that these particular results are significant. But interpreting them in the context of the other flux scenarios is still challenging.

Accounting for uncertainties in the optimised fluxes and considering the optimised fluxes at a monthly resolution shows that we cannot discern between the scenarios or draw firm conclusions about the results. Considering the results at an annual timescale could potentially help differentiate between the scenarios but then begs the question of what we would want to use regional inversions of CO_2 mole fractions in Europe for, since the annual bottom-up inventories used by different countries have much smaller uncertainties than derived here.




```
Only plotting one system is supported
```

Figure C2.1: The 2021 monthly optimised fluxes from RHIME and from CTE for Germany (DEU), France (FRA) and Finland (FIN). Flux scenarios are blinded for the six flux scenarios (BASE plus five perturbed scenarios).

3. The Reveal

Here, we reveal the flux perturbation scenarios to see whether what was inferred above is true.

RHIME

The unblinded results for RHIME are shown in Fig. 3.5

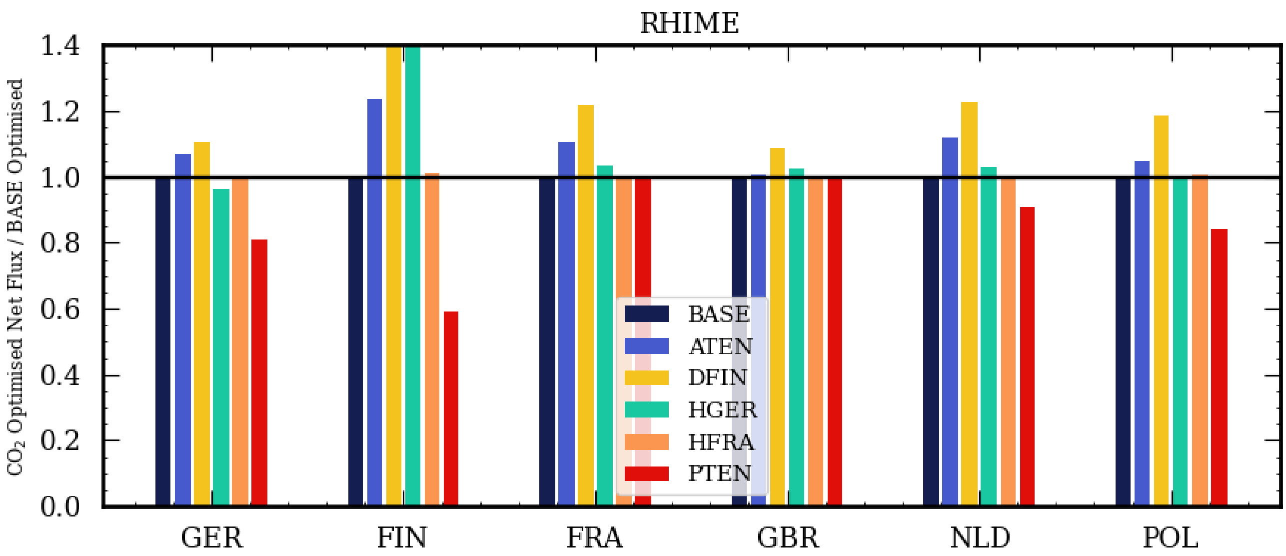


Figure C3.1: The unblinded 2021 optimised fluxes-to-optimised base flux from RHIME.

We deduced from the blinded scenarios that:

1. Apollo --> *DFIN**
2. Ares --> could not be reasonably determined
3. Hera --> could not be reasonably determined
4. Hermes --> *HGER*
5. Demeter --> PTEN
6. Zeus --> *ATEN**

From Fig. 3.4 we see that

1. Apollo = DFIN [correct]
2. Ares = HFRA [incorrect]
3. Hera = BASE [incorrect]
4. Hermes = HGER [correct]
5. Demeter = PTEN [correct]
6. Zeus = ATEN [correct]

We were (tentatively) able to correctly identify four of six the scenarios with RHIME when disregarding uncertainties. We could not reasonably identify which scenario was HFRA and BASE. We see from the plot of the true fluxes (Fig. 2.1) that the HFRA perturbation is small for France so given the sensitivity of the atmospheric measurements to France and the size of this perturbation it is maybe not surprising this scenario could not be identified. Indeed we also see the PTEN perturbation in France (which is also small) is also indistuinguishable from the BASE scenario.

The large scale perturbation scenarios - ATEN and PTEN - were correctly identified in RHIME. These perturbation scenarios were domain-wide rather than country specific. This again tells us that given the magnitude of these perturbations, they need to be across a sufficient area to be *detected* in the inversion. We also had prior knowledge of what the flux perturbations were and could logically deduce some of them.

---

4. Can Optimised Fluxes be Correctly Quantified?

Whilst we've been able to identify and detect some of the perturbations, we should check whether we've correctly quantified the perturbation. We shall check this for ATEN and PTEN in Figures C4.1 and C4.2, respectively.

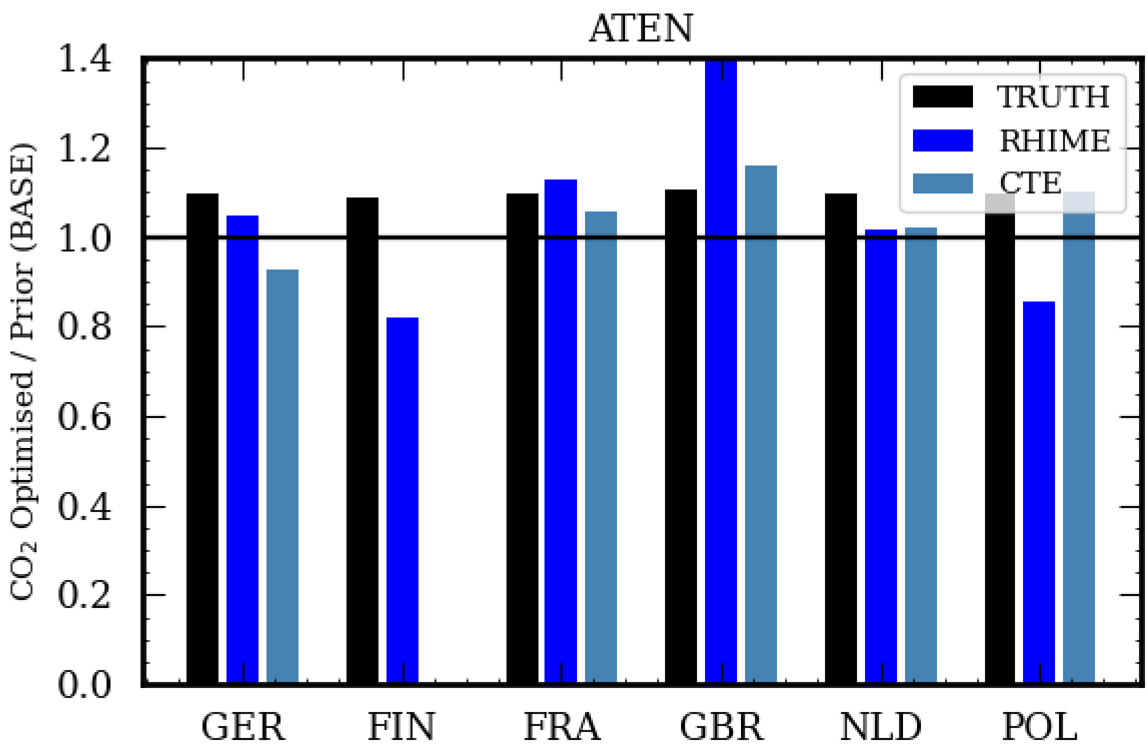


Figure C4.1: ATEN optimised fluxes divided by their prior. We would expect the inverse models to have values matching the truth (black) for each country.

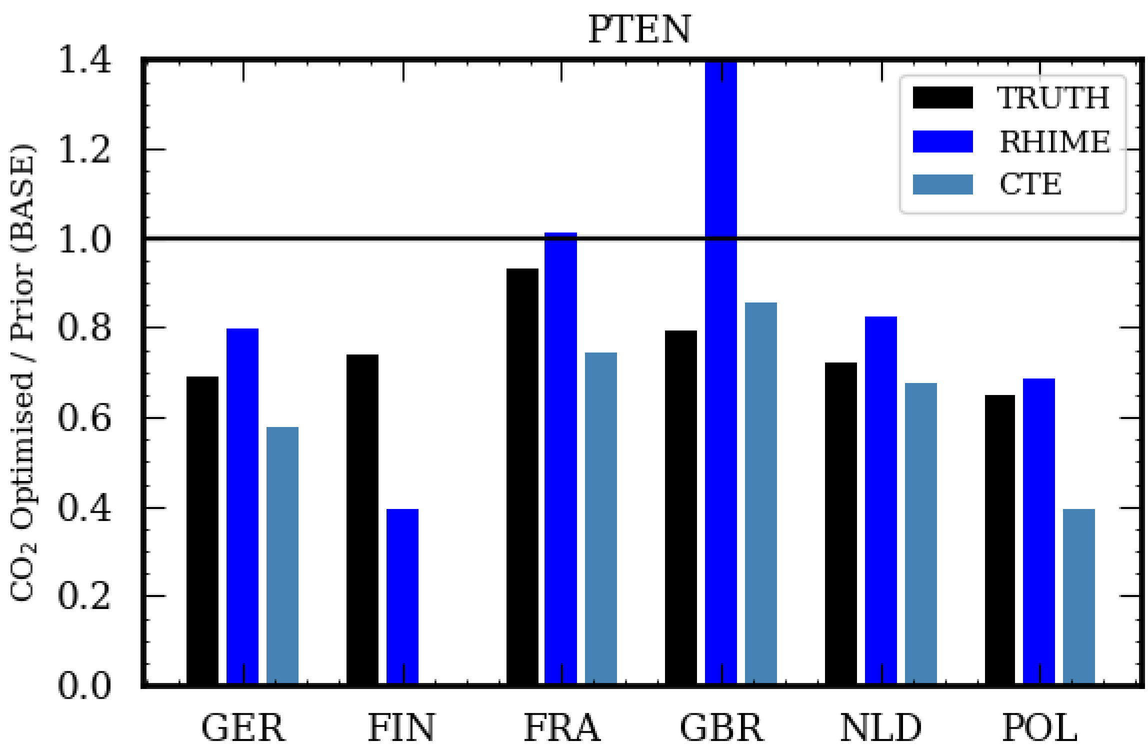


Figure C4.2: PTEN optimised fluxes divided by their prior. We would expect the inverse models to have values matching the truth (black) for each country.

From these figures, we can see some significant differences between the optimized fluxes and the 'true' fluxes from the scenario. We discuss this in the outlook