Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Maksym Gachkivskyi

born in: Kyiv, Ukraine

Oral examination: 21.10.2025

Assessing Fossil CO_2 Emissions in Europe:
How Reliably Do ICOS $^{14}CO_2$ and CO Observations Constrain Atmospheric Inversions?

Referees: Prof. Dr. Werner Aeschbach

Prof. Dr. André Butz

Abstract

Accurate quantification of country-scale fossil fuel CO_2 ($\Delta ffCO_2$) emissions is essential for monitoring efforts to mitigate climate change. This thesis employs the regional isotope budget approach (RIBA) to calculate $\Delta ffCO_2$ observations based on flask, CO, and integral samples, and assesses their constraint on top-down fossil emission estimates within a Bayesian inversion framework. Flasks provide a limited number of precise hour-long observations, CO-based estimates are less precise but recorded every minute, and integral samples yield reliable two-week averages. The coverage of currently available $^{14}CO_2$ data from the Integrated Carbon Observation System (ICOS) infrastructure mainly includes Germany and nearby regions. When comparing to the emission inventories by the Global Carbon Project (GCP) and the Emissions Database for Global Atmospheric Research (EDGAR), corresponding winter emissions estimated from the atmospheric data differ by 0–4% from GCP but by 10–14% from EDGAR, reducing the discrepancy between the GCP and EDGAR inventories from 21% to about half. Incorporating $\Delta ffCO_2$ observations decreases the GCP inventory uncertainty by 30% in the Germany+ domain. Uncertainty analysis indicates that biases in $^{14}CO_2$ background are critical for the RIBA. A background bias of 2‰ results in estimated flux variations of 20%. Since integral-based samples have lower requirements for hourly-specific model transport, they are found to be the most suitable proxy for constraining trends and seasonal fossil fuel emissions. They effectively track trends and align with realistic expectations derived from bottom-up inventories. Flask-based estimates perform equally well if sampled more frequently than weekly, but demand more analytical resources. CO-based estimates are found to be less suitable due to their additional dependence on CO background estimates and unmodeled air chemistry. The two main recommendations for ICOS to monitor country-scale fossil emissions and their trends are: (a) diversifying European $^{14}CO_2$ background observations to reduce bias risk, and (b) prioritizing integral $^{14}CO_2$ sampling.

Zusammenfassung

Eine genaue Quantifizierung der fossilen CO_2 Emissionen ($\Delta ffCO_2$) auf Länderebene ist für die Überwachung der Bemühungen zur Eindämmung des Klimawandels entscheidend. Diese Arbeit verwendet den regionalen Isotopenhaushaltsansatz (RIBA), um $\Delta ffCO_2$ Beobachtungen auf der Grundlage von Flask-, CO- und Integralproben zu berechnen, und bewertet deren Auswirkungen auf Top-down-Schätzungen fossiler Emissionen innerhalb eines Bayes'schen Inversionssystems. Flasks liefern eine begrenzte Anzahl präziser eine Stunde langer Beobachtungen, CO-basierte Schätzungen sind weniger präzise, werden jedoch minütlich aufgezeichnet, und Integralproben liefern zuverlässige Zwei-Wochen-Durchschnittswerte. Die Abdeckung der $^{14}CO_2$ Daten aus der Infrastruktur des Integrated Carbon Observation System (ICOS) umfasst hauptsächlich Deutschland und benachbarte Regionen. Beim Vergleich zu den Emissionsinventaren des Global Carbon Project (GCP) und der Emissions Database for Global Atmospheric Research (EDGAR) weichen die entsprechenden Winteremissionen, die aus den $^{14}CO_2$ Daten abgeleitet wurden, um 0–4% vom GCP, aber um 10–14% von EDGAR ab, wodurch die Diskrepanz zwischen den GCP- und EDGAR-Inventaren von 21% auf etwa die Hälfte reduziert wird. Durch die Einbeziehung von $\Delta ffCO_2$ Beobachtungen wird die Unsicherheit des GCP-Inventars im Bereich Deutschland+ um 30% verringert. Die Unsicherheitsanalyse zeigt, dass Bias in den $^{14}CO_2$ Hintergrundwerten für RIBA von entscheidender Bedeutung sind. Ein Hintergrund-Bias von 2‰ führt zu abgeleiteten Flussvariationen von 20%. Da integralbasierte Proben geringere Anforderungen an den stundenspezifischen Modelltransport stellen, erweisen sie sich als der am besten geeignete Proxy für die Ermittlung von $\Delta ffCO_2$. Sie verfolgen Trends effektiv und stimmen mit realistischen Erwartungen überein, die aus Bottom-up-Inventaren abgeleitet wurden. Flask-basierte Schätzungen funktionieren bei einer häufigeren Probenahme als wöchentlich ebenso gut, erfordern jedoch mehr analytische Ressourcen. CO-basierte Schätzungen sind aufgrund ihrer zusätzlichen Abhängigkeit von CO-Hintergrundschätzungen und nicht modellierter Luftchemie weniger geeignet. Die beiden wichtigsten Empfehlungen für ICOS zur Überwachung der fossilen Emissionen und ihrer Trends auf Länderebene lauten: (a) Diversifizierung der europäischen $^{14}CO_2$ Hintergrundbeobachtungen zur Verringerung des Bias-Risikos und (b) Priorisierung der Integralprobenahme von $^{14}CO_2$.

List of Abbreviations

Abbreviation	Full Term
AMS	Accelerator Mass Spectrometry
AoC	area of constraint
CRL	Karl Otto Münnich Central Radiocarbon Laboratory in Heidelberg
EDGAR	Emissions Database for Global Atmospheric Research
FCL	Flask and Calibration Laboratory in Jena, Germany
$\Delta ffCO_2$	fossil fuel carbon dioxide
FWD	forward simulations
GCP	Global Carbon Project
GHG	greenhouse gas
ICOS	Integrated Carbon Observation System
MDM	model-data-mismatch
MICADAS	Mini Radiocarbon Dating System
NOAA	National Oceanic and Atmospheric Administration
RIBA	The regional isotope budget approach
RMSD	Root Mean Square Deviation
SNR	signal-to-noise ratio
STILT	Stochastic Time-Inverted Lagrangian Transport
WMO	World Meteorological Organization

Contents

Chapter 1

Introduction

1.1 Motivation

According to the Copernicus Climate Change Service and the World Meteorological Organization (WMO), 2024 marked the warmest year on record, with the global average temperature surpassing $1.5°C$ above pre-industrial levels for the first time since records began (Copernicus Climate Change Service & World Meteorological Organization, 2025). A broad and robust scientific consensus, underpinned by extensive evidence including Charles David Keeling's pioneering work at Mauna Loa Observatory in 1958 (Hofmann et al., 2009; Lan & Keeling, 2025), which documented the increasing atmospheric CO_2 concentration, firmly establishes human activities as the primary driver of this observed warming (Lynas et al., 2021).

In response to the urgent threat of climate change and to mitigate its profound effects on ecosystems and human society, the international community adopted the Paris Agreement in 2015 under the auspices of the United Nations Framework Convention on Climate Change (UNFCCC) (UNFCCC, 2015). This landmark agreement sets a central goal to limit the increase in the global average temperature to "well below $2°C$" above pre-industrial levels, while pursuing efforts to restrict the increase to an aspirational target of $1.5°C$. To ensure accountability and track progress towards these ambitious targets, member nations of the UNFCCC are mandated to submit yearly National Inventory Reports (e.g. Grubb, 2019). These reports are compiled using "bottom-up" methodologies and consumption statistics, adhering to standardized guidelines provided by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2006). The rigorous evaluation of these anthropogenic carbon activities is crucial for informing policy, setting emission reduction targets, and assessing the efficacy of mitigation strategies.

Despite these commitments, current projections indicate a significant shortfall in achieving the Paris Agreement's goals. The latest evaluation of Nationally Determined Contribu-

tions and proposed legislative changes, as outlined in the International Energy Agency's (IEA) Announced Pledges Scenario (IEA, 2024), projects the average global temperature increase by 2100 to be approximately 1.7°C above pre-industrial levels. More concerning, an analysis of policies already implemented or currently being developed, presented in the IEA's Stated Policies Scenario, forecasts an increase of 2.4°C by 2100. To contextualize these projections, the IPCC estimated in 2021 the remaining global carbon budget from the beginning of 2020 necessary to limit global warming to the Paris Agreement targets (Masson-Delmotte et al., 2021). According to these estimations, a carbon budget of 400 Gt CO_2 offers a 67% likelihood of limiting the global temperature increase to 1.5°C. However, with current global emissions in the order of 35 Gt CO_2 per year (Friedlingstein et al., 2024), this amount diminishes quickly.

This underscores the critical importance of reliable greenhouse gas (GHG) emissions reporting. However, the "bottom-up" approach currently employed in these national reports, which relies on multiplying activity data by emission factors, is inherently susceptible to systematic errors and constitutes a significant source of uncertainty (Super et al., 2020; Solazzo et al., 2021). Key challenges include inconsistent data quality, incomplete reporting, and the inherent difficulty in accurately separating anthropogenic emissions from the natural variability of carbon sinks (Andres et al., 2012; Friedlingstein et al., 2024). These limitations necessitate the development and application of independent verification methods for GHG emissions.

Independent verification of GHG emissions is crucial for assessing the effectiveness of mitigation strategies. Inverse atmospheric modeling offers a "top-down" approach for independently validating national emission inventories (e.g. Rödenbeck et al., 2003; Bergamaschi et al., 2015; Basu et al., 2020). These models leverage atmospheric concentration data to infer surface fluxes, using Bayesian statistics (Bayes, 1763) to update initial estimates from inventories with real-world observations. This method is crucial for verifying the efficacy of climate policies and identifying potential reporting discrepancies (Friedlingstein et al., 2024).

The success of inverse modeling is contingent on reliable, high-quality data, which is provided by infrastructures like the Integrated Carbon Observation System (ICOS, ICOS RI, 2020). The ICOS Atmosphere network ensures high-precision and intercompatible data across its stations through standardized measurement protocols, centralized calibration against international reference standards, and rigorous quality control. The resulting standardized datasets are essential for modeling applications. A key scientific tool for validation is radiocarbon (^{14}C), an ideal tracer for distinguishing fossil fuel CO_2 ($\Delta ffCO_2$) from biogenic sources (Levin et al., 2003; Levin & Rödenbeck, 2008; Graven, 2015; Turnbull et al., 2015; Li et al., 2025). These observations are particularly valuable on regional

and continental scales for monitoring the effectiveness of emission reduction efforts and disentangling anthropogenic from ecosystem-driven carbon fluxes. This can be achieved for example through a regional isotope budget approach (Levin et al., 2003; Maier et al., 2023).

1.2 Objectives and structure

This thesis investigates the information content of existing $^{14}CO_2$ observations conducted within the ICOS Atmosphere network, with a focus on assessing their potential to constrain fossil fuel CO_2 emissions at the country to continental scale. Rather than aiming for a full and policy-ready quantification of actual emissions—which would require multiple transport models and a broad ensemble of prior inventories—I deliberately use a single atmospheric transport model (STILT, Lin et al., 2003) and a limited set of inventories to isolate and compare the relative performance of different $\Delta ffCO_2$ proxies. These include ^{14}C-based flask and integral estimates as well as ^{14}C-calibrated CO-based proxies (Maier et al., 2023). Because my focus is on the comparative signal strength and consistency of these proxies across the existing ICOS network, these results are expected to be largely independent of the specific model or inventory used. In addition, to evaluate which of the proxies are most promising for use in inversion frameworks (CarboScope, Rödenbeck et al., 2003), I assess key uncertainty drivers—such as the influence of background $^{14}CO_2$ variability—on the detectability of fossil fuel signals. The ultimate goal of this work is to inform future decision making on ICOS network development, and sampling strategy, thereby strengthening the long-term capacity of ICOS to monitor fossil CO_2 emissions across Europe.

This thesis is structured as follows. Chapters 2 and 3 provide the theoretical and methodological framework for the regional isotope budget approach, detailing the specific tools and models used. Chapter 2 includes also a description of the ICOS measurement network and the different proxies investigated. Chapter 4 discusses the information content of the observational data series and the effects of measurement errors in a series of synthetic studies. In Chapters 5, 6, and 7, I present the results of regional inversions based on flask-, CO-, and integral-based $\Delta ffCO_2$ estimates. I use findings from previous chapters to evaluate proxy effectiveness in reproducing core features of emission fluxes, such as seasonality or trend, and investigate the effects of various uncertainties on the posterior estimates. Chapter 8 discusses the main problems of the regional isotope budget approach and compares the inversion results of the proxies for the overlapping time period. Finally, Chapter 9 evaluates the potential of the investigated proxies within the ICOS network and offers recommendations for improving future monitoring strategies. The appendix contains additional figures for corresponding evaluations.

Chapter 2

Carbon cycle and ^{14}C observations

2.1 Natural carbon cycle and anthropogenic disturbance

The carbon cycle (Fig. 2.1) is a biogeochemical process in-between the Earth's biosphere, geosphere, hydrosphere, and atmosphere. Cycling carbon between these reservoirs sets the foundation for life on Earth (Archer, 2011). This cycle comprises fast and slow exchange processes. Fast (biological) carbon cycles occur on the time scales from days to years, e.g. moving carbon between the atmosphere and biosphere. Slow (so-called geological or deep) carbon cycles, spanning millions of years, transport carbon through the Earth's crust, soil, ocean, and atmosphere (Archer, 2011). The ocean, representing the largest active carbon reservoir on the Earth's surface, plays a major role as a sink of atmospheric carbon. The terrestrial and oceanic biospheres interact with atmospheric carbon by removing CO_2 through photosynthesis and releasing it back during respiration, driving atmospheric CO_2 seasonality (Canadell et al., 2021). In general, the atmosphere acts in the short term as a carbon sink by accumulating carbon from natural and anthropogenic sources and distributing it to other reservoirs over time. In the atmosphere, carbon exists mainly as carbon dioxide (CO_2) and methane (CH_4), both potent greenhouse gases. Although CH_4 has a higher greenhouse effect per volume, CO_2 contributes more significantly to radiative forcing due to its higher concentration and longer atmospheric lifespan ($\approx$ 9-12 years and over 1000 years, respectively) (Ciais et al., 2013; Szopa et al., 2021).

In an undisturbed state, the preindustrial carbon cycle maintained a long-term equilibrium, with fluxes between reservoirs largely balancing out (Archer, 2011). However, human activities have significantly perturbed the natural carbon cycle during the last two centuries, notably through land use changes and the extraction and combustion of fossil fuels, i.e. coal, petroleum, and natural gas (Friedlingstein et al., 2024). Since the Industrial Revolution and particularly after World War II, massive amounts of geospheric carbon have been redistributed, increasing atmospheric carbon, mainly as CO_2 , by 52% compared to preindustrial

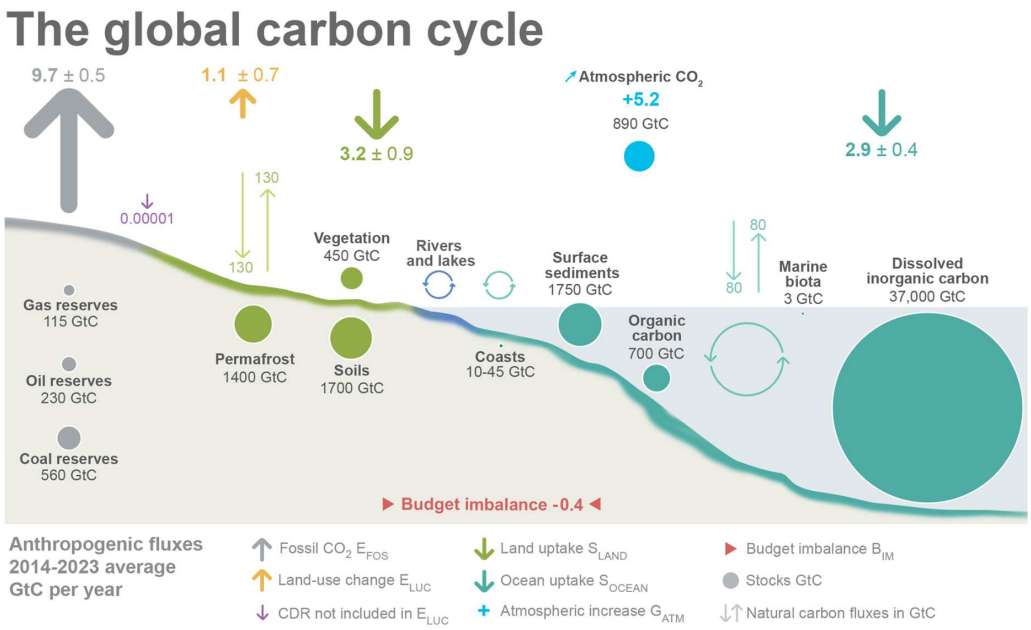

Figure 2.1. Schematic representation of the global carbon cycle disturbances, including estimated gross fluxes (Gt C yr^{-1}) and reservoir stocks (Gt C). Anthropogenic perturbations are shown atop the natural carbon cycle. All numbers are from Canadell et al., 2021, except for coastal carbon stocks, which are from Price and Warren, 2016. Figure adapted from Friedlingstein et al., 2024.

levels: 422 ppm in 2024 compared to approximately 278 ppm in 1750 (Friedlingstein et al., 2024). This has led to global warming and other effects such as a reduction in ocean pH, which fundamentally altered marine chemistry and endangered marine ecosystems (Doney et al., 2009).

However, even with the current unprecedented disturbance of the natural carbon cycle by human activities, the Earth system will eventually establish a new equilibrium, with atmospheric CO_2 levels approaching preindustrial concentrations over vast timescales (Archer et al., 2009). These long-term processes are predominantly governed by the ocean carbon pump, which would take thousands to millions of years to absorb most of the excess atmospheric carbon and store it as dissolved inorganic carbon, with a fraction eventually being incorporated into carbonate sediments (Archer et al., 2009). Unfortunately, the time scales involved are far too long to mitigate the immediate and pressing impacts of climate change on human societies, highlighting the need for immediate climate action (Lee et al., 2024).

In contrary to the natural long-term equilibrium processes described above, the atmospheric CO_2 concentrations could be influenced rather fast by changing anthropogenic forcing. However, it is first necessary to determine which part of the overall CO_2 signal was caused by anthropogenic emissions. The extraction of the fossil signal from the total CO_2 concentration proves to be challenging due to the significant influence of the biosphere on

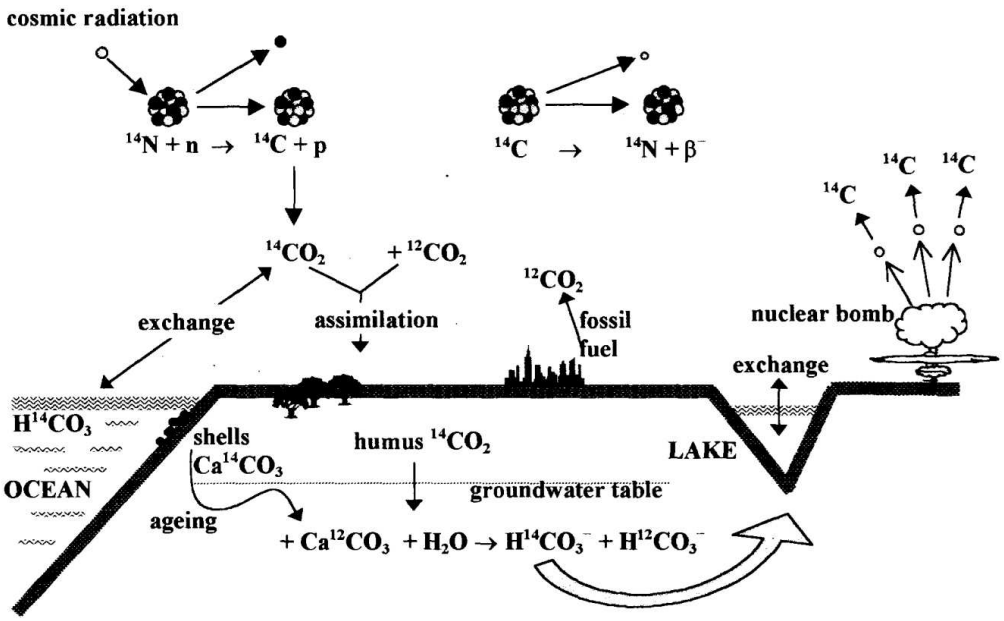

Figure 2.2. The natural origin and distribution of ^{14}C. Its atmospheric concentration has been altered by the addition of ^{14}C depleted CO_2 from fossil fuel combustion and the production of ^{14}C from nuclear fission and fusion reactions. Figure from Mook, 2000.

atmospheric CO_2 concentrations, with, for example, the photosynthetic removal signal being more than 10 times higher compared to the fossil signal in high summer (Maier et al., 2023). Radiocarbon ^{14}C , a radioactive carbon isotope, serves as a crucial tracer for distinguishing fossil fuel-induced carbon fluxes from ecosystem fluxes (Levin et al., 1989, 2003; Levin & Rödenbeck, 2008; Graven, 2015; Turnbull et al., 2015).

^{14}C is naturally produced in the transitional zone between the lower stratosphere and the upper troposphere (Libby, 1946; Mook, 2000). This occurs through a nuclear reaction where atmospheric nitrogen interacts with thermal neutrons, which are themselves generated by high-energy cosmic-ray protons that react with atmospheric molecules (Fig. 2.2). The newly formed ^{14}C rapidly oxidizes to ^{14}CO and $^{14}CO_2$, mixing with atmospheric CO_2 and participating in all CO_2 exchange processes. In an undisturbed natural carbon cycle, there is a balance between the production of ^{14}C from cosmic rays and its radioactive decay (half-life of $\sim$5700 years, Be et al., 2013).

Industrialization marked the first significant anthropogenic disturbance to the natural $^{14}C/^{12}C$ ratio. Fossil fuels, formed millions of years ago, are devoid of ^{14}C due to its short half-life compared to the age of the fuels. Consequently, fossil fuel combustion dilutes the natural $^{14}C/^{12}C$ ratio in rapidly exchanging carbon reservoirs, including the atmosphere, the upper oceans and the biosphere. This phenomenon is known as the Suess effect and was already reported in the 1950s (Suess, 1955, start of the series in Fig. 2.3).

A second major disturbance to the atmospheric ^{14}C levels was caused by extensive

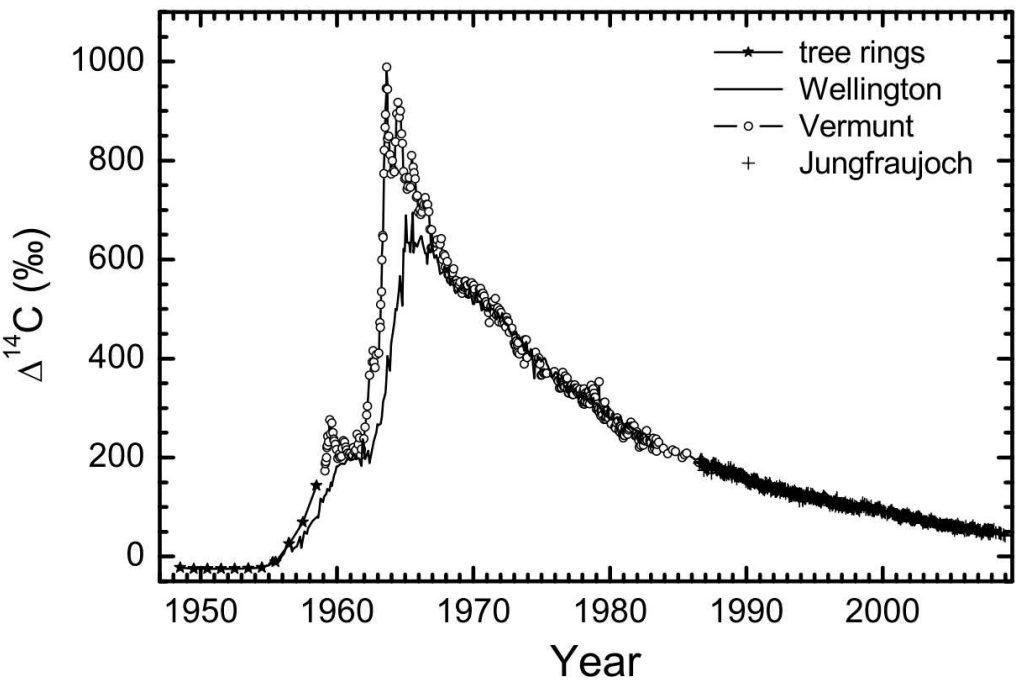

Figure 2.3. Influence of the nuclear weapons testing on the atmospheric ^{14}C concentration in the northern and the southern hemisphere. First signs of Suess effect can be seen as values below 0‰ in the early 1950s. Figure from Levin et al., 2010.

nuclear weapons testing from 1956 until the 1963 Nuclear Test Ban Treaty (Levin & Rödenbeck, 2008; Levin et al., 2010). Atmospheric tests doubled the $^{14}C/^{12}C$ ratio in the troposphere due to additional ^{14}C production from thermal neutron reactions (Fig. 2.3). Subsequent uptake of this bomb-produced $^{14}CO_2$ into fast-exchanging reservoirs masked the decline caused by the Suess effect for several decades (Levin & Rödenbeck, 2008; Levin et al., 2010). Although nuclear power plants still act as a source of anthropogenic ^{14}C (Knaack, 2025), more recently, fossil fuel CO_2 emissions have again become the main driving force that contribute to changes in atmospheric ^{14}C levels (Levin & Rödenbeck, 2008; Levin et al., 2010).

2.2 Isotopic composition of carbon, fractionation and $^{14}CO_2$ units

The chemical element carbon $_6C$ has three natural isotopes with different number of neutrons and hence different masses: ^{12}C, ^{13}C and ^{14}C.

The isotopic composition of a carbon reservoir can be described by the isotopic ratio of

a rare isotope to an abundant one:

$$^{i}R = \frac{\text{abundance of rare isotope}}{\text{abundance of abundant isotope}} \tag{2.1}$$

Instead of reporting absolute isotope ratios, it is common practice to express them using the δ notation, which quantifies the relative deviation of an isotope ratio in a sample from that of a defined standard. Given that such deviations are typically small, the δ values are reported in ‰:

$$\delta = \left(\frac{^{i}R_{Sample}}{^{i}R_{Standard}} - 1 \right) 1000‰ \tag{2.2}$$

The standard reference material for δ^{13}C measurements is Vienna Pee Dee Belemnite (VPDB) with the $^{13}C/^{12}C$ ratio of 1.12372% (Mook, 2000).

The definition of δ^{14}C was adapted to the nature of the measurement of ^{14}C. Due to the low abundance of the ^{14}C isotope, for a long time the only reliable method to determine its content was the low-level decay counting with the anti-coincidence configuration of the radiation detectors (Anderson et al., 1951; Libby, 1955; Kromer & Münnich, 1992). Hence, the ratios ^{i}R in Eq. 2.2 are replaced with specific activity A (per gram of carbon). The absolute internationally defined standard specific activity of ^{14}C A_{abs} is equal to 95% of the specific activity of the National Bureau of Standards (NBS) oxalic acid, corrected for decay loss since 1950 ($A_{abs} = 0.95 \cdot 0.238$ Bq/gC, Stuiver & Polach, 1977). The δ value of a sample $\delta^{14}C_S$ with its specific activity A_S is equal to:

$$\delta^{14}C_S = \left(\frac{A_S}{A_{abs}} - 1 \right) 1000‰ \tag{2.3}$$

The introduction of the Accelerator Mass Spectrometry (AMS) measurement technique allowed direct calculation of ^{14}R so that both the general δ notation (Eq. 2.2) and the specific activity (Eq. 2.3) definitions can be used if related to the isotopic ratio of the NBS oxalic acid reference material (Bonani et al., 1987; Hammer et al., 2017).

Differences in isotopic properties lead to fractionation effects, for example, during phase transitions, which must be taken into account.

Differences in isotopic masses lead to changes in the mass-dependent physical and chemical properties of carbon isotopologues (Mook, 2000), which in turn causes fractionation, i.e., separation of isotopologues, in physical and chemical processes. Mass-dependent fractionation processes are mainly differentiated into equilibrium and kinetic fractionation. For example, in a kinetic fractionation process, molecules containing the lighter isotope ^{12}C might diffuse through a medium or chemically react slightly faster than those containing ^{14}C due to the lower velocities of heavier molecules with the same thermal energy, leading to the separation of isotopologues. Equilibrium fractionation, on the other hand, is typically characterized by an enrichment of heavier isotopes in certain phases of matter due to higher binding energies and stronger chemical bonds (Mook, 2000).

Every exchange process between compounds or changes in the state of matter, for example evaporation or melting, results in fractionation effects (Mook, 2000). These are quantified by a fractionation factor α comparing isotopic ratios ^{i}R in compound of state A with that in B:

$$ {}^{i}\alpha_{A/B} = \frac{{}^{i}R_A}{{}^{i}R_B} \tag{2.4} $$

For carbon, the relative mass difference of ^{13}C to ^{12}C is twice as small as between ^{14}C and ^{12}C. Thus, the fractionation of ^{14}C to ^{12}C is routinely given as the square of the same effect between ^{13}C and ^{12}C (Eq. 2.5) (Stenström et al., 2011). Although there is some experimental evidence suggesting that this power factor is closer to 1.9 (Stuiver & Robinson, 1974; Fahrni et al., 2017), by convention a normalization factor of 2 is still used.

$$ {}^{14}\alpha_{A/B} = {}^{13}\alpha^2_{A/B} \tag{2.5} $$

To make different ^{14}C measurements independent of fractionation processes and therefore compatible with each other, a normalization factor based on the $\delta^{13}C$ value of each sample ($\delta^{13}C_S$) was introduced to account for fractionation. By convention, samples are normalized to a δ^{13}C value of $-25‰$ w.r.t. VPDB (Stuiver & Polach, 1977). The normalized activity A_{SN} of a ^{14}C sample is equal to:

$$ A_{SN} = A_S \left(\frac{\left(\frac{^{13}C}{^{12}C}\right)_{\delta^{13}C=-25}}{\left(\frac{^{13}C}{^{12}C}\right)_{Sample}} \right)^2 = A_S \frac{\left(1 - \frac{25}{1000}\right)^2}{\left(1 + \frac{\delta^{13}C_S}{1000}\right)^2} = A_S \left(\frac{0.975}{1 + \frac{\delta^{13}C_S}{1000}} \right)^2 \tag{2.6} $$

Combining this normalization with Eq. 2.3 results in the Δ notation following the definition of Stuiver and Polach, 1977, which is used in this study:

$$ \Delta^{14}C_S = \left(\frac{A_{SN}}{A_{abs}} - 1 \right) 1000‰ = \left(\frac{A_S}{A_{abs}} \left(\frac{0.975}{1 + \frac{\delta^{13}C_S}{1000}} \right)^2 - 1 \right) 1000‰ \tag{2.7} $$

2.3 ICOS $^{14}CO_2$ measurements

This section provides a brief introduction to the Integrated Carbon Observation System (ICOS) and its atmospheric network for $^{14}CO_2$ observations. It describes the collection and analysis of the two different $^{14}CO_2$ sample types — integrated and flask samples. Finally, an overview of the time periods covered by the data sets available for this work is given.

The Integrated Carbon Observation System (ICOS ERIC, 2025) is a European research infrastructure supported by 16 member states, established in 2008 and formalized as a European Research Infrastructure Consortium (ERIC, European Commission, Directorate-General

for Research and Innovation, 2025) in 2015. ICOS quantifies greenhouse gas concentrations in Europe and its surrounding regions. ICOS operates nearly 200 observation stations in the fast-cycling carbon reservoirs of the atmosphere, ecosystems, and oceans. The observations are conducted by the national ICOS networks and coordinated from the ICOS Head Office in Helsinki, as well as domain-specific Thematic centers and central laboratories. ICOS data and results are made openly available via a central ICOS data portal. More details on the structure, objectives, and operation of ICOS can be found in Heiskanen et al., 2022.

The scientific mission of ICOS is to deliver long-term, high-quality observations essential to understanding the European carbon cycle and anthropogenic greenhouse gas emissions, thus supporting climate change mitigation efforts. At atmospheric stations, ICOS emphasizes strict adherence to data quality and compatibility, ensuring that the measurements align with the standards established by the WMO for greenhouse gas observations (ICOS RI, 2020). To guarantee data consistency at its stations, ICOS employs standardized instrumentation, measurement protocols, and data processing procedures defined in the ICOS Atmosphere Station Specifications Handbook (ICOS RI, 2020). This high level of harmonization is essential to prevent systematic biases across the observation network, which could otherwise be misinterpreted by atmospheric inversion models as artificial greenhouse gas sources or sinks.

The design of the ICOS Atmospheric Network aims to achieve spatially homogeneous coverage throughout Europe, minimizing significant spatial gaps. However, as a nationally funded European research infrastructure, coverage is limited to the territories of member countries. Priority is given to continental tall tower stations to improve estimates of land-based greenhouse gas fluxes, which exhibit considerable spatio-temporal variability due to natural and anthropogenic fluxes. A smaller subset of coastal and mountain stations is added to capture the inflow of greenhouse gases to the European continent and serve as clean-air background reference sites. The locations of the ICOS stations are selected to minimize the direct influence of strong anthropogenic surface emissions. Key guidelines for station selection recommend a nominal spacing of approximately 300 km between stations, with a minimum distance of 50 km. Furthermore, it is recommended to avoid complex terrain and significant anthropogenic sources (e.g. urban areas) within 40 km – particularly if located upwind relative to prevailing winds – to ensure compatibility with current atmospheric transport models. A minimum tower height of 100 meters above ground level is required for the top observational level to increase the representativeness of the station. All stations undergo a two-step labeling process to ensure the adherence to ICOS specifications, data transfer protocols, and quality standards (ICOS RI, 2020).

ICOS atmosphere stations use commercially available instruments for continuous high-resolution measurements of, e.g., CO_2 , CH_4, CO, and main meteorological parameters.

These instruments undergo rigorous testing at the Atmosphere Thematic Center (ATC) in France before deployment (Kwok et al., 2015). Calibration gases for in situ measurements are prepared and calibrated at the Flask and Calibration Laboratory (FCL) in Jena, Germany, ensuring compatibility within the ICOS network and traceability to WMO calibration scales.

ICOS atmospheric stations are classified into two types: Class 1, equipped for comprehensive mandatory measurements that include integrated $^{14}CO_2$ and flask sampling, and Class 2, which conducts only a subset of Class 1 measurements without flask analysis and $^{14}CO_2$ observations. For our analysis, only Class 1 stations are relevant due to their capacity to collect ^{14}C samples in the form of flask samples and/or integrated samples. Fig. 2.4 shows the locations of the subset of ICOS Class 1 stations that have a sufficiently long $^{14}CO_2$ observation record to be used in the current study. The ICOS Class 1 sub-network covers northern France, the Benelux countries, Germany, and parts of Scandinavia. The station coordinates and heights of the intake lines used for the $^{14}CO_2$ sampling are summarized in Table 2.1.

2.3.1 ICOS $^{14}CO_2$ flask samples

Flask sampling in the ICOS network serves three main purposes: independent quality control for in situ observations of greenhouse gases, providing data on atmospheric components not continuously monitored, and collecting samples for $^{14}CO_2$ analysis to determine fossil fuel CO_2 ($\Delta ffCO_2$). 3 L glass flasks with 1.5 bar pressure are collected every third day, typically between 11:00 and 15:00 local time, from the highest intake of a station. Each sample represents a one hour integrated mean, collected during well-mixed atmospheric conditions (wind speeds > 2 ms^{-1}) to ensure a well-defined footprint. The ICOS-approved automatic flask sampler, developed by Max Planck Institute for Biogeochemistry in Jena, is used for this purpose. The FCL analyzes the collected flask samples for various atmospheric constituents (e.g. Jordan & Brand, 2003; Van der Laan et al., 2009), while precise $^{14}CO_2$ analysis is conducted at the Karl Otto Münnich Central Radiocarbon Laboratory (CRL) in Heidelberg. The ICOS flask sampling strategy is further detailed in Levin et al., 2020.

Analyzing trace amounts of $^{14}CO_2$ in 3 L flask air samples requires specialized preparation and measurement techniques. The initial step involves extracting CO_2 from the air mixture. The traditional cryogenic freezing method with liquid nitrogen, which was used for the samples analyzed in this study, is an effective technique, but labor intensive in its nature (Lux, 2018). Subsequent to extraction, the gaseous CO_2 must be reduced to solid graphite, a critical prerequisite for high-precision AMS measurements. This reduction typically occurs through a Bosch reaction ($CO_2 + 2H_2 \rightarrow C + 2H_2O$), catalyzed by materials such as iron or cobalt (Němec et al., 2010). This preparation enables the AMS analysis of ^{14}C from just a

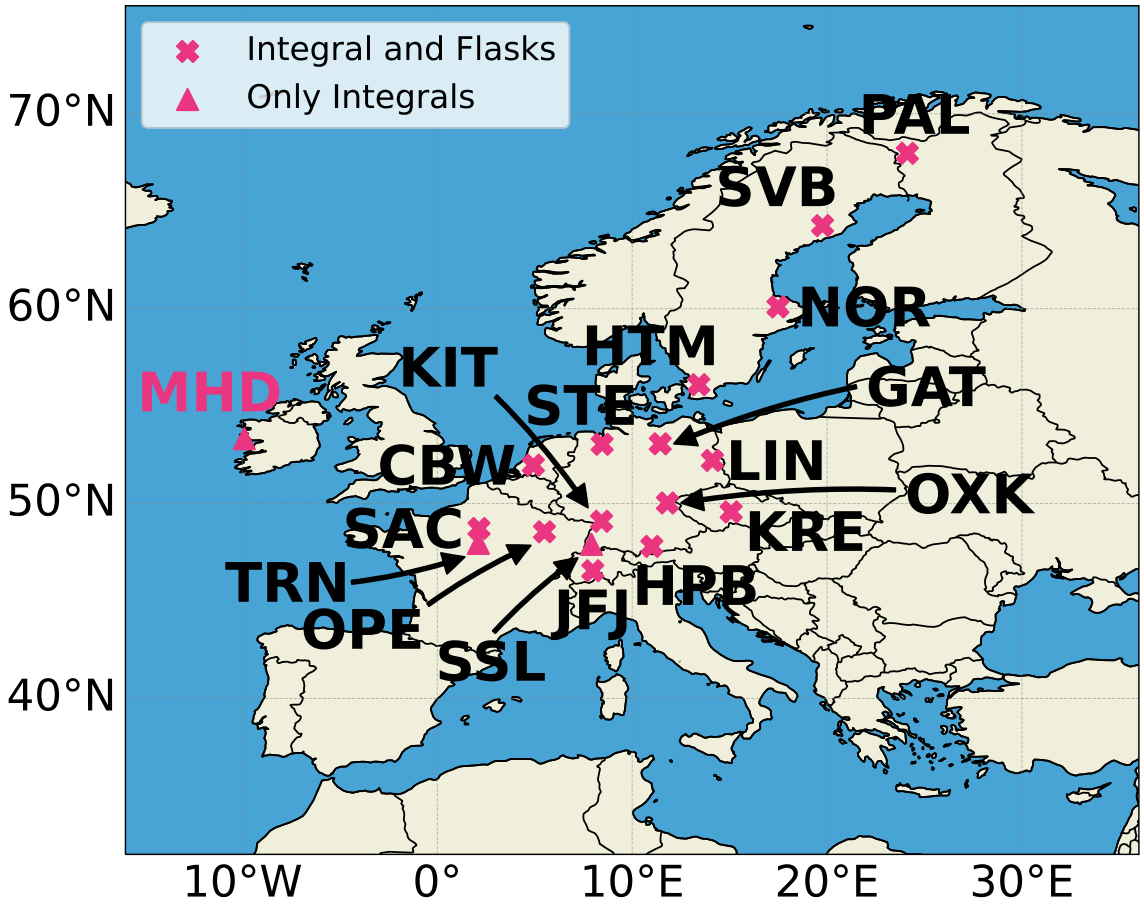

Figure 2.4. Map of ICOS Atmosphere Class 1 stations. Stations with both integrated and flask measurements are marked with a cross. Stations with only integrated samples are marked with a triangle. In situ CO observations are available at all Class 1 stations. The tag for the European marine background station Mace Head (MHD) is indicated in pink.

few liters of air, making it suitable for samples collected by the ICOS $^{14}CO_2$ flask sampling network.

All ^{14}C flask measurements in this study were performed using the Mini Radiocarbon Dating System (MICADAS) at the Curt-Engelhorn-Center for Archaeometry in Mannheim (Kromer et al., 2013). The MICADAS (Fig. 2.5, Synal et al., 2007; Wacker et al., 2010; Synal, 2022) is a low energy 200 kV AMS device designed for high precision analysis of ^{14}C, directly measuring $^{14}C/^{12}C$ ratios. Its architecture comprises a sample extraction unit, two mass spectrometers separated by a particle accelerator, and multiple detectors. Graphite samples are loaded onto magazines and automatically fed into the ion source, where a Cs^+ beam ionizes the graphite to C^- ions. These ions are then selected by a first mass spectrometer and injected into a tandem accelerator. Within the tandem accelerator, a stripper gas simultaneously converts negative ions to positive ions and fragments interfering molecules (e.g. $^{13}CH^-$), crucial for the unambiguous detection of ^{14}C. The accelerated ion beam then

Table 2.1. ICOS stations with CO_2 , $\Delta^{14}C$, and CO measurements, including corresponding heights used in this study.

Station, country code	ID	Lon (°)	Lat (°)	Station height (m a.m.s.l.)	Intake height (m a.g.l.)
Cabauw, NL	CBW	51.97 N	4.93 E	0	207
Gartow, DE	GAT	53.07 N	11.44 E	70	341
Hohenpeißenberg, DE	HPB	47.80 N	11.01 E	934	150
Hyltemossa, SE	HTM	56.10 N	13.42 E	115	150
Jungfraujoch, CH	JFJ	46.55 N	7.98 E	3572	14
Karlsruhe, DE	KIT	49.09 N	8.43 E	110	200
Křešín u Pacova, CZ	KRE	49.57 N	15.08 E	534	250
Lindenberg, DE	LIN	52.21 N	14.12 E	73	98
Mace Head, IE	MHD	53.33 N	-9.90 E	0	15
Norunda, SE	NOR	60.09 N	17.48 E	46	100
Observatoire pérenne de l'environnement, FR	OPE	48.55 N	5.50 E	390	120
Ochsenkopf, DE	OXK	50.03 N	11.81 E	1022	163
Pallas, FI	PAL	67.97 N	24.12 E	565	12
Saclay, FR	SAC	48.72 N	2.14 E	160	100
Schauinsland, DE	SSL	47.91 N	7.91 E	1205	12
Steinkimmen, DE	STE	53.04 N	8.46 E	29	252
Svartberget, SE	SVB	64.26 N	19.77 E	269	150
Trainou, FR	TRN	47.96 N	2.11 E	131	180

passes through a second mass spectrometer, separating the isotopes into distinct beams for current measurements of ^{12}C and ^{13}C. A final energy charge filter isolates ^{14}C ions, which are then counted by a single particle detector using a gas ionization chamber, ensuring high specificity and precision in the ^{14}C event detection. This setup allows for an accurate ^{14}C analysis from small amounts of carbon derived from air samples.

2.3.2 Two-week integrated $^{14}CO_2$ samples

In addition to flask sampling, Class 1 stations also collect two-week integrated samples from the same height as flask samples through chemical absorption of CO_2. This involves continuous pumping of ambient air through a rotating glass tube containing a NaOH base solution and filled with Raschig glass rings to increase the contact surface between the solution and

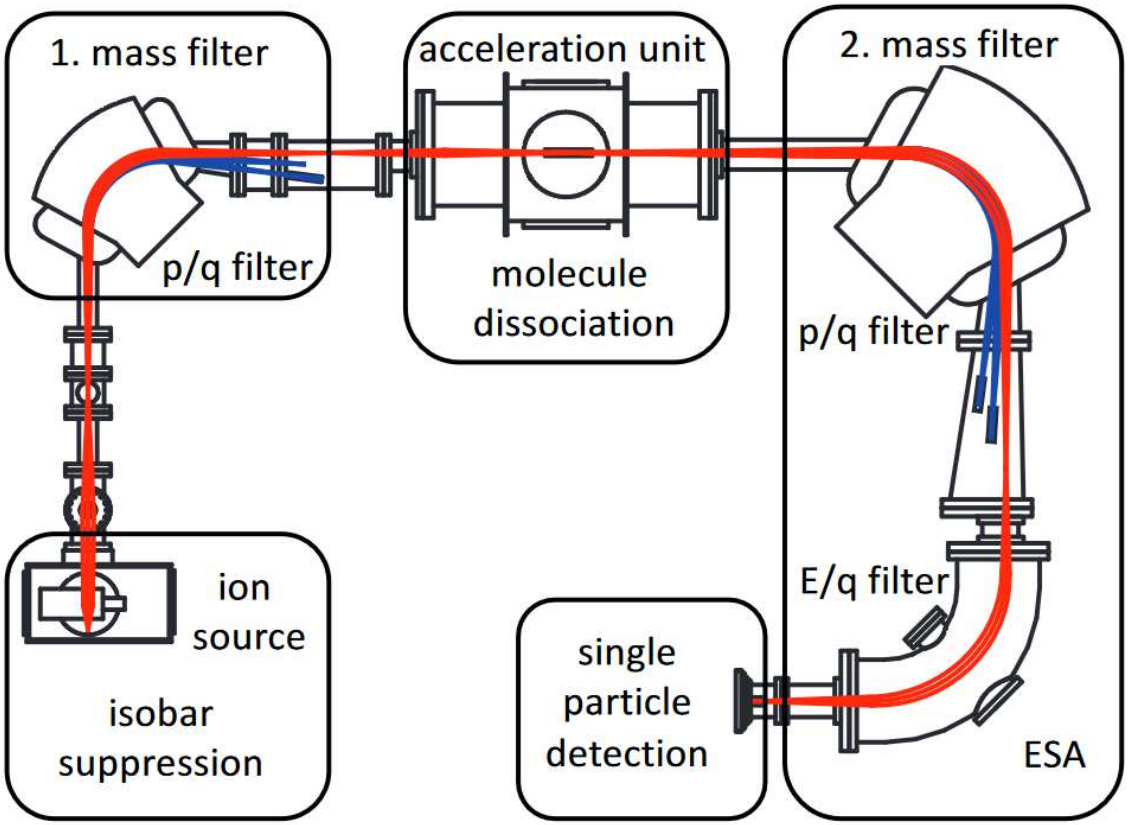

Figure 2.5. Schematic layout of the MICADAS AMS device. The red line illustrates the pathway of measured ^{14}C through the system. In blue, the pathway of other elements with atomic masses 12, 13, or 14 is shown. Reproduced from Synal, 2022.

pumped air, ensuring a high absorption rate of CO_2 (Levin et al., 1980). These integrated samples are then sent to CRL in solution form, where the absorbed CO_2 is extracted from NaOH by adding half-concentrated H_2SO_4 (40%) to the sodium carbonate solution.

Subsequently, these samples are measured using gas proportional counting in the Heidelberg low-level counting (LLC) laboratory, using an anticoincidence shield to minimize interference from cosmic rays (Libby, 1955; Schoch et al., 1980). Samples for LLC typically require larger sample amounts compared to flasks (around 25 m^3 of atmospheric air) and longer counting periods due to ^{14}C's long half-life of 5700 years (Be et al., 2013) and low natural abundance ($\approx 10^{-10}\%$, Schuur et al., 2016). The CRL's counting facilities are described in depth by Kromer and Münnich, 1992. LLC is a mature and powerful technique central to radiocarbon dating, and its results have shown good overall agreement with international AMS laboratories (Hammer et al., 2017).

2.3.3 ICOS $^{14}CO_2$ samples used in this work

Table 2.2 lists an overview of the $^{14}CO_2$ ICOS data series that were used in this thesis (Levin et al., 2024; Emmenegger et al., 2025b; Frumau & Hensen, 2025a; Hatakka, 2025a; Heliasz & Biermann, 2025a; Kubistin et al., 2025k, 2025d, 2025b, 2025c, 2025l, 2025a; Larmanou et al., 2025a; Lehner & Molder, 2025a; Marek et al., 2025a; Ramonet et al., 2025a, 2025c, 2025e; Schmidt et al., 2025). In chapters 5 to 7, three different $\Delta ffCO_2$ estimation approaches will be analyzed in view of their suitability for the inversion of fossil CO_2 fluxes within the ICOS network. The three approaches examined are introduced in detail later and are based on $^{14}CO_2$ information from flask samples or integral samples, and, in one of the estimates, additionally on the measurement of CO concentrations (Emmenegger et al., 2025a; Frumau & Hensen, 2025b; Hatakka, 2025b; Heliasz & Biermann, 2025b; Kubistin et al., 2025e, 2025f, 2025g, 2025h, 2025i, 2025j; Larmanou et al., 2025b; Lehner & Molder, 2025b; Marek et al., 2025b; Martin, 2025; Ramonet et al., 2025b, 2025d). The ICOS CO data series used in this work are therefore also listed in Table 2.2.

Table 2.2. Observation availability for $\Delta^{14}C$ and CO at the investigated stations. Dates in brackets note the $\Delta ffCO_2$ availability if components other than $\Delta^{14}C$ (e.g., nuclear correction, see Sect. 3.1) were not available for the period. Utilization in corresponding inversions is marked by: F - flask, C - CO-based, I - integral (suitability of the individual flask stations for CO-based inversion is discussed in Sect. 6.2.3). (*) KIT is used only in sensitivity studies due to potential ^{14}C contamination.

Site	$\Delta^{14}C$ integrals	$\Delta^{14}C$ flasks	CO in situ	Inversion set
CBW	09.05.11 (17.12.14) - 06.11.24 (03.01.24)	13.01.22 - 13.12.24	30.09.21 - 22.04.25	F C I
GAT	20.05.21 - 05.01.25 (09.01.24)	01.07.21 - 03.12.24	11.04.17 - 10.04.25	F C I
HPB	18.02.15 - 14.08.24 (20.12.23)	28.08.19 - 06.12.24	18.07.19 - 22.04.25	F C I
HTM	30.07.15 - 07.02.24 (09.01.24)	28.11.20 - 10.11.24	24.03.20 - 22.04.25	F C I
JFJ	23.07.86 (22.12.14) - 09.12.24 (25.12.23)	28.11.20 - 09.01.24	12.12.16 - 22.04.25	F X I
KIT*	01.02.18 - 27.09.21	26.06.19 - 01.12.24	26.06.19 - 22.04.25	X X X
KRE	29.03.17 - 11.12.24 (10.01.24)	21.12.23 - 15.11.24	12.04.17 - 22.04.25	F C I
LIN	15.03.18 - 17.12.24 (02.01.24)	07.09.20 - 18.11.24	11.08.20 - 22.04.25	F C I
MHD	09.10.00 - 17.10.22	-	-	X X I
NOR	26.05.15 - 04.02.25 (09.01.24)	02.06.21 - 20.10.24 (14.12.23)	01.04.17 - 22.04.25	F X I
OPE	25.03.11 (15.12.14) - 20.01.25 (08.01.24)	09.09.20 - 06.11.24	18.08.16 - 22.04.25	F C I
OXK	21.04.21 - 03.02.25 (06.06.23)	02.06.21 - 03.11.24 (20.12.23)	25.09.19 - 22.04.25	F X I
PAL	14.12.17 - 13.03.24 (03.01.24)	03.10.22 - 03.12.24 (03.12.23)	16.09.17 - 22.04.25	F X I
SAC	11.05.18 - 13.11.24 (22.11.23)	29.04.19 - 03.08.22	31.05.17 - 22.04.25	X X I
SSL	09.12.76 (22.12.14) - 10.06.24 (18.12.23)	-	-	X X I
STE	13.07.19 - 20.09.24 (01.01.24)	01.07.21 - 01.12.24	22.07.19 - 22.04.25	F C I
SVB	09.02.16 - 11.02.25 (21.12.23)	03.06.21 - 01.11.24 (01.12.23)	01.06.17 - 22.04.25	F X I
TRN	17.11.07 (09.12.14) - 28.05.24 (11.01.24)	-	-	X X I

Chapter 3

Methods

3.1 The regional isotope budget approach for ΔffCO$_2$ estimation

The regional isotope budget approach (RIBA) is a well-established technique for the estimation of recently added CO_2 from fossil fuel burning and cement production ($\Delta ffCO_2$) at regional or continental scales (Levin et al., 2003; Turnbull et al., 2006; Levin et al., 2011; Turnbull et al., 2015; Berhanu et al., 2017; Major et al., 2018; Zhou et al., 2020; Maier et al., 2023). This method exploits the fact that the combustion of ^{14}C-free fossil fuels causes a difference in $\Delta^{14}C$ measurements between a "clean" background and an observation site. This measurement-based top-down approach provides an independent estimate of recently added $\Delta ffCO_2$ concentrations, which, combined with atmospheric transport inversions (Sect. 3.2.2), can be used to study bottom-up fossil CO_2 emission inventories. The RIBA is discussed in great detail along with its fundamental assumptions and shortcomings in Maier et al., 2023, to which I refer the interested reader. In the following, I present the basic concept of RIBA.

A measured CO_2 signal (C_{meas}) at an observation site can be broken down as the sum of different CO_2 contributions (Levin et al., 2003; Turnbull et al., 2006; Maier et al., 2023):

$$C_{meas} = C_{bg} + C_{ff} + C_{resp} + C_{photo}(+C_{ocean} + C_{strato}) \tag{3.1}$$

The background contribution (C_{bg}) is the largest, representing the globally increasing and well-mixed hemispheric CO_2 concentration influenced by all global CO_2 fluxes (e.g., Mauna Loa CO_2 record, Lan & Keeling, 2025). Other contributions to the measured CO_2 signal originate from recent fossil fuel combustion and cement production (C_{ff} or $\Delta ffCO_2$), biosphere respiration (C_{resp}), and photosynthesis uptake (C_{photo}) within the observational site's catchment area. In addition, stations may be influenced by oceanic fluxes (C_{ocean}) and stratospheric air intrusions (C_{strato}).

As discussed in Sect. 2.2, the Δ-notation allows a direct comparison of different reservoirs by correcting for radioactive decay and accounting for mass-dependent isotopic fractionation via δ^{13}C normalization. Thus, the sum of products of individual CO$_2$ contributions and their characteristic isotopic Δ signatures is a conserved quantity (Tans et al., 1993). Consequently, a balance equation similar to Eq. 3.1 can be written for Δ^{14}C :

$$C_{\text{meas}} \cdot \Delta^{14}C_{\text{meas}} = C_{\text{bg}} \cdot \Delta^{14}C_{\text{bg}} + C_{\text{ff}} \cdot \Delta^{14}C_{\text{ff}} + C_{\text{resp}} \cdot \Delta^{14}C_{\text{resp}} + C_{\text{photo}} \cdot \Delta^{14}C_{\text{photo}} \\ (+C_{\text{ocean}} \cdot \Delta^{14}C_{\text{ocean}} + C_{\text{strato}} \cdot \Delta^{14}C_{\text{strato}}) + C_{\text{meas}} \cdot \Delta^{14}C_{\text{nuc}} \quad (3.2)$$

Here, the additional $\Delta^{14}C_{\text{nuc}}$ is the expected contribution to $\Delta^{14}C_{\text{meas}}$ from nuclear facilities, with negligible influence on overall CO$_2$ concentrations.

Nuclear installations emit ^{14}C to the atmosphere primarily in the forms of ^{14}CO$_2$ or ^{14}CH$_4$. Although ^{14}CH$_4$ does not affect the ΔffCO$_2$ calculations in the regional isotope budget approach, ^{14}CO$_2$ does so by increasing atmospheric ^{14}C and therefore potentially masking fossil signals. Thus, accounting for nuclear ^{14}C emissions is necessary, which will be discussed further in Sect. 3.1.3. Another source of ^{14}CO$_2$ which needs to be accounted for in RIBA is the release of ^{14}C through biospheric heterotrophic respiration (Caldeira et al., 1998; Randerson et al., 2002; Naegler & Levin, 2009b). Historical atmospheric nuclear weapon testing significantly increased Δ^{14}C in the atmosphere, nearly doubling its concentration (Naegler & Levin, 2006) in the 1960s. Since then, much of this excess has been assimilated by other carbon reservoirs, including the biosphere. ^{14}C-enriched CO$_2$ (in comparison to modern atmosphere) respired from decomposing organic material with long reservoir times (e.g., in soils) can also mask fossil signals. This requires a biospheric correction, which is discussed in Sect. 3.1.2.

Furthermore, I neglect the potential influence of oceanic fluxes ($C_{\text{ocean}} \cdot \Delta^{14}C_{\text{ocean}}$), since my primary focus is on measurements carried out on the European continent, where due to the location of the background station, all oceanic influences should already be incorporated into the background signal (Sect. 3.1.1). Due to the regional scale of my investigation, I assume that the contribution of stratospheric air enriched with ^{14}C does not change significantly in the European domain and is therefore already incorporated into $C_{\text{bg}} \cdot \Delta^{14}C_{\text{bg}}$. Small latitudinal changes in the concentration of ^{14}C in Europe (Lingenfelter, 1963) are neglected.

Fossil fuels, millions of years old, are devoid of ^{14}C, therefore, their Δ^{14}C signature is equal to -1000‰. This allows for a direct calculation of C_{ff} from Eq. 3.2. However, both biospheric components (photosynthesis and respiration) are often uncertain and are typically not known separately. Using Eq. 3.1, I can eliminate C_{photo} from Eq. 3.2. As suggested by Maier et al., 2023 for the ICOS stations, $\Delta^{14}C_{\text{photo}}$ is set to the on-site measured Δ^{14}C signature as it represents local biosphere that is photosynthesized. The Δ-notation accounts

for the fractionation between the reservoirs. Consequently, the C_{ff} for the ICOS sites can be written as:

$$C_{\text{ff}} = C_{\text{bg}} \cdot \frac{\Delta^{14}C_{\text{bg}} - \Delta^{14}C_{\text{meas}}}{\Delta^{14}C_{\text{meas}} + 1000‰} + C_{\text{meas}} \cdot \frac{\Delta^{14}C_{\text{nuc}}}{\Delta^{14}C_{\text{meas}} + 1000‰} + C_{\text{resp}} \cdot \frac{\Delta^{14}C_{\text{resp}} - \Delta^{14}C_{\text{meas}}}{\Delta^{14}C_{\text{meas}} + 1000‰} \tag{3.3}$$

C_{meas} in flasks is measured directly by FCL (see Sect. 2.3) and for two-week integral samples it is set to an average of co-measured in situ CO$_2$ during the integration period at the corresponding intake height.

In the following, I will describe the individual contributions to Eq. 3.3, including the estimation of the marine background (Sect. 3.1.1), the biospheric component (Sect. 3.1.2), and the correction for potential nuclear contamination (Sect. 3.1.3). Note that background estimation and nuclear corrections have been improved compared to the methods described by Maier et al., 2023.

3.1.1 Choice and construction of the ^{14}CO$_2$ background time series

The RIBA estimates excess concentrations of ΔffCO$_2$ at measurement stations relative to a "clean" background site. The choice of this background reference is crucial, as the resulting ΔffCO$_2$ estimates are directly proportional to the difference in Δ^{14}C (or ΔCO) between the observation and the reference site.

3.1.1.1 Mace Head (MHD) as background site for Europe

Previous analyses showed that the majority of STILT footprints for a subset of nine ICOS stations in 2018 exited the domain over the Atlantic Ocean, making Mace Head (MHD; 53.33° N, 9.90° W, 5 m a.s.l.) a suitable marine background station for the European domain (Maier et al., 2023). Therefore, MHD was used as the default background site in earlier studies (Maier et al., 2024b, 2024a). To ensure that MHD measurements reflect uncontaminated marine air, a wind sector filter (190°–300° and > 4 m/s) is applied to remove local or continental pollution.

As the RIBA uses only one background reference regardless of wind direction, this will introduce biases, especially for air masses entering Europe from nonwestern boundaries. To assess this, a sensitivity analysis was performed during this PhD project, simulating three virtual stations at high, mid, and low latitudes of the eastern STILT domain boundary. Using the global TM3 model driven by GCP emissions with European emissions in the STILT domain (16°W - 36°E and 32°N - 74°N, see Sect. 3.2.1) set to zero, I calculated the accumulated ΔffCO$_2$ concentrations for the virtual stations on the eastern boundary of the modeled

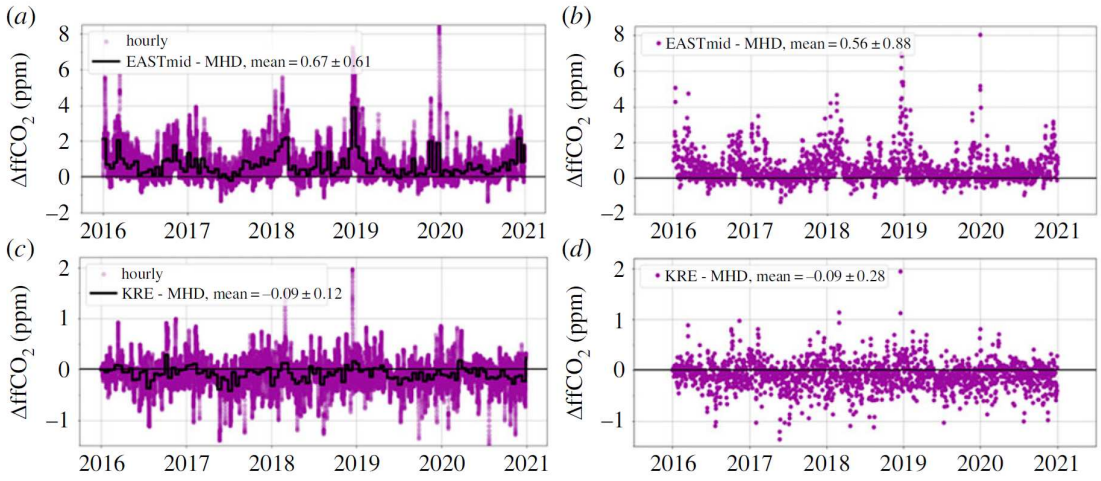

Figure 3.1. The offsets of fossil CO_2 concentration with respect to MHD due to the emissions from outside of European domain (see Fig. 2.4). Panels (a) and (b) show the data from a virtual station at mid-latitudes of the eastern boundary (EASTmid), for all hours with biweekly means (a) as well as for 13h UTC data only (b). Panels (c) and (d) depict the analogous data for the easternmost ICOS station Křešín (KRE), Czechia. Mean offsets and standard deviations in the legends are in ppm. Adapted from (Maier et al., 2023, Fig. 3)

European domain, and the residual concentrations of $\Delta ffCO_2$ were calculated by subtracting the MHD time series from these. Figure 3.1 shows the residual $\Delta ffCO_2$ concentrations at the virtual station at mid-latitudes (EASTmid, 55° N, 34° E, 150 m a.g.l.) and at the easternmost ICOS $\Delta^{14}C$ station Křešín (KRE, 49.57° N, 15.08° E, 250 m a.g.l.). These results have been developed as part of this PhD project and have already been incorporated into the publication by Maier et al., 2023.

The residual $\Delta ffCO_2$ signals demonstrate that a non-negligible offset of around 0.6 ppm $\Delta ffCO_2$ arises when MHD is used as the background for stations at the eastern boundary of the modeled European domain, particularly for easterly inflow. However, at KRE the mean offset has decreased to essentially zero (-0.09±0.28 ppm). Compared to a mean $\Delta ffCO_2$ signal of 2.4 ppm for integral-based measurements (Sect. 7.2.1), this offset poses a bias of about 4% and loses its relevance further the more westward the stations are located. Following Maier et al., 2024a, the absolute bias is neglected in the RIBA, but the MHD background uncertainty was increased by 0.28 ppm $\Delta ffCO_2$ (or 0.64‰ in $\Delta^{14}C$) to reflect this representativeness error. No significant latitudinal gradient was found along the three virtual stations at the eastern boundary.

3.1.1.2 Construction of a smoothed MHD-based $\Delta^{14}C$ background

Until 2022, the $\Delta^{14}C$ background time series was constructed from biweekly integrated samples at MHD. To fill temporal gaps, suppress short-term noise, and ensure continuity, data was smoothed using the ccgcrv algorithm (Press et al., 1988; Thoning et al., 1989). This method fits a long-term trend using a quadratic polynomial and captures seasonality via harmonic functions. A low-pass filter in Fourier space is applied to the residuals and added back to restore inter-annual variations.

Due to the biweekly sampling frequency and relatively high measurement uncertainty, short-term variations were not included. The resulting smoothed MHD-based background served as the main background estimate in earlier phases of this PhD.

However, in the late 2022, $\Delta^{14}C$ sampling at MHD was suspended due to safety issues with the tower to which the inlet system and the meteorological observations were attached. This period coincided with an expected increase in stratospheric $\Delta^{14}C$ input after the 2020 solar minimum (Clette & Lefèvre, 2015), with a typical delay of $\sim$2 years due to stratosphere–troposphere exchange (Holton et al., 1995). These developments motivated the construction of an updated and more robust background estimate.

3.1.1.3 Northern hemispheric background construction from multiple stations

To extend the $\Delta^{14}C$ background estimate and improve its representativeness, I derived a mean inter-annual trend from multiple Northern Hemisphere clean-air stations: Alert, CA (ALT, Levin & Hammer, 2022), Jungfraujoch, CH (JFJ), Mace Head, IE (MHD), Niwot Ridge, US (NWR, operated by National Oceanic and Atmospheric Administration, NOAA, Miller et al., 2025), Pallas, FI (PAL), and Svartberget, SE (SVB). Flask or integrated measurements were smoothed using ccgcrv, and offsets relative to MHD (2014–2022) were calculated and subtracted to align all $^{14}CO_2$ time series to MHD's latitudinal level.

The resulting trend series were averaged to produce a mean Northern Hemisphere $\Delta^{14}C$ trend, which reflects the large-scale stratospheric influence and mitigates the effect of data gaps at individual stations. A Gaussian filter with $\sigma = 50$ days was applied to remove edge artifacts from station-specific start and end dates. Linear extrapolation was used to extend the mean trend into the final quarter of 2024. The background uncertainty was increased from 2.1‰ to 2.4‰ in this period to account for the extrapolation. The seasonal cycle was adopted from the long-term mean seasonal pattern observed at MHD (2000–2022).

Figure 3.2 illustrates this new northern hemispheric background, as well as the key data for its construction. Panels (a)–(d) show the station-specific inter-annual trends in comparison to MHD; panel (e) compares the resulting background curve (green line, uncertainty in gray shading) to the previously used background of MHD only from Maier et al., 2023

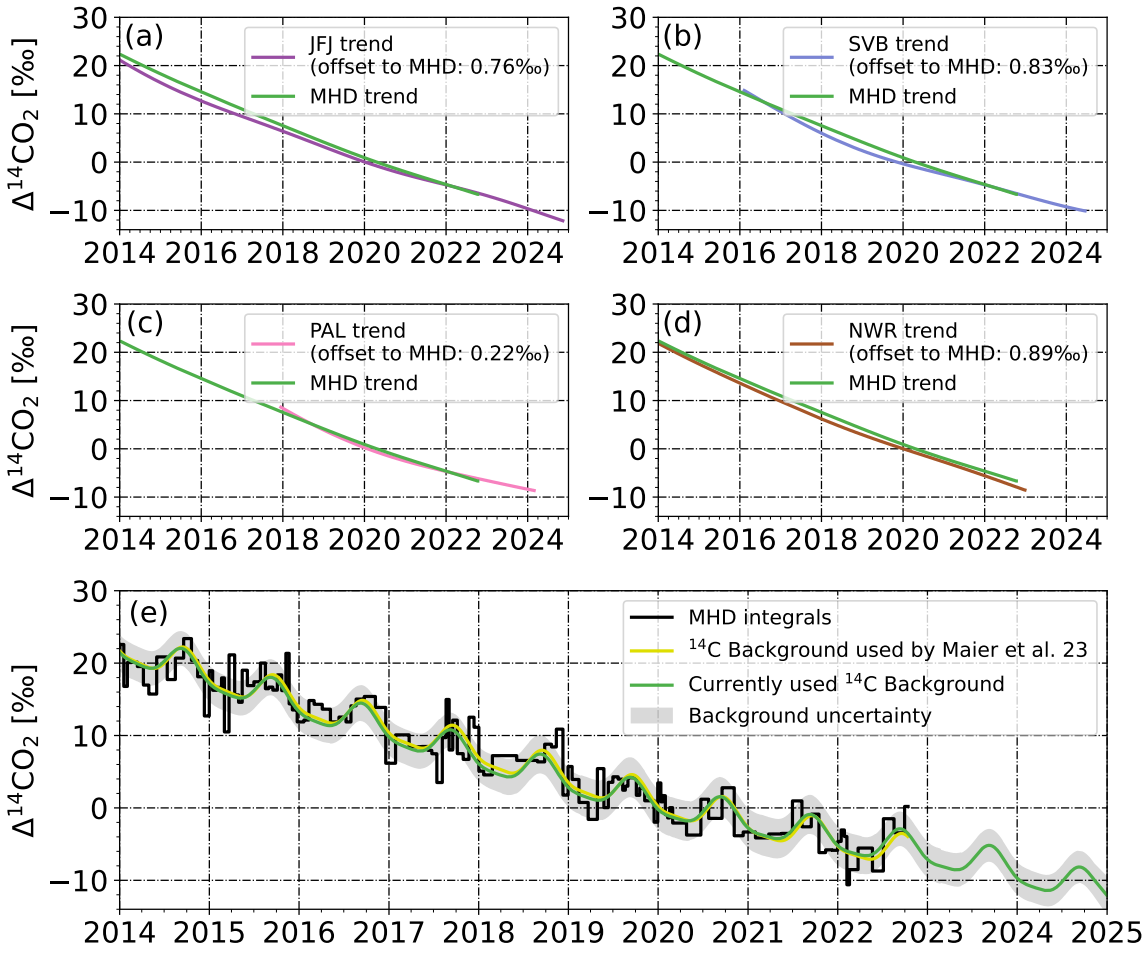

Figure 3.2. Comparison of different Northern Hemisphere $\Delta^{14}C$ background trends. Panels (a)-(d): Inter-annual $\Delta^{14}C$ trend of clean air sites extracted using ccgcrv smoothing routine used to calculate mean northern hemispheric trend. The offset to the mean MHD trend is noted in the legend. Panel (e): Comparison of the $\Delta^{14}C$ background curve used in this study (green line with gray shading representing uncertainty) against the background estimation from Maier et al., 2023 (yellow line) and individual MHD $\Delta^{14}C$ integral values (black line).

(yellow) and the individual $\Delta^{14}C$ samples from MHD (black line).

The influence of enhanced natural $\Delta^{14}C$ production becomes evident in 2022-2024, where less steep slopes can be observed at several background stations (JFJ, PAL, SVB). In addition, the late MHD samples show a tendency to exceed the background estimate. Although the assumption of fast meridional mixing justifies the use of a mean trend for the Northern Hemisphere (Warneck, 1999), the robustness of long-term station-specific latitudinal offsets for this particular period is uncertain. As can be seen in panels (a) - (d) of Figure 3.2, the latitudinal offsets are not constant in time. Furthermore, the fixed seasonal cycle of MHD makes the background estimate less sensitive to short-term changes. This is exemplified by the difference between the northern hemispheric ^{14}C background (green)

and the MHD-based background (yellow) in panel (e) in Figure 3.2. Thus, a potential bias in the $\Delta^{14}C$ background, particularly in the years 2023–2024 without the MHD observations, cannot be ruled out.

3.1.2 Biospheric correction

As stated previously, a significant correction in $\Delta ffCO_2$ calculations using the regional isotope budget approach RIBA involves biosphere respiration, which is enriched in $^{14}CO_2$ relative to today's atmospheric CO_2. The $^{14}CO_2$ enrichment in the plant's assimilated biomass results from higher atmospheric $^{14}CO_2$ levels in the past caused by nuclear bomb testing in 1960s (Fig. 2.3, Levin et al., 2022). Thus, the enrichment of heterotrophically respired CO_2 varies with the age of the respired biomass. This additional $^{14}CO_2$ masks the $\Delta ffCO_2$ signal.

The isotopic signature of respiration, $\Delta^{14}C_{\text{resp}}$, varies between autotrophic and heterotrophic processes. Earlier studies estimated heterotrophic $\Delta^{14}C$ signatures using mean terrestrial carbon residence times (Turnbull et al., 2006). Field measurements indicate $\Delta^{14}C_{\text{soil}}$ values of 48.2-56.7‰ in boreal forests in 2012 (Palonen et al., 2018) and a mean $\Delta^{14}C_{\text{resp}}$ of 32.0 ± 7.4‰ in tropical rainforests for year 2019 (Chanca, 2022), consistently tens of permil higher than contemporary atmospheric $\Delta^{14}C$ signatures (approximately 30‰ in 2012 and 2‰ in 2019 for the northern hemisphere, see Fig. 3.2). In this thesis, I used the method suggested by Maier et al., 2023, i.e., the approximated autotrophic $\Delta^{14}C_{\text{resp}}$ signature with background air values, while the $\Delta^{14}C$ signature of the heterotrophic respiration contributions were modeled assuming a mean age distribution of the biomass pools (e.g., Naegler and Levin, 2009a).

The Vegetation Photosynthesis and Respiration Model (VPRM, Mahadevan et al., 2008) coupled with STILT was used to simulate C_{resp} for ICOS sites. These simulations show higher C_{resp} signals in summer than in winter, typically ranging from 2 to 8 ppm at ICOS sites. The respiration correction is particularly crucial for ICOS sites with low $\Delta ffCO_2$ signals and strong biospheric influence. In winter, respiration and fossil CO_2 signals are comparable in magnitude ($C_{\text{resp}}/C_{\text{ff}} \approx 1$), leading to less than 5% $\Delta ffCO_2$ masking assuming realistic $^{14}C_{\text{resp}}$ enrichment of below 40‰ relative to ambient air. For summer, however, this effect can reach nearly 20% $\Delta ffCO_2$ masking due to the much higher biospheric signals ($C_{\text{resp}}/C_{\text{ff}} \approx 6$) assuming the same $^{14}C_{\text{resp}}$ enrichment (Maier et al., 2023). An uncertainty of 100% is assumed for the respiration component, resulting in an overall biospheric correction contribution to the expected uncertainty of $\Delta ffCO_2$ of less than 0.5 ppm for typical ICOS sites (Maier et al., 2023).

3.1.3 Nuclear correction

Estimating $\Delta ffCO_2$ from ^{14}C measurements requires accounting for ^{14}C emissions from nuclear facilities (Kuderer et al., 2018). The ICOS Carbon Portal provides a dedicated Jupyter notebook to calculate the so called nuclear contributions ($\Delta^{14}C_{nuc}$) at each ICOS Class 1 site resulting from all European nuclear facilities. This tool maps 3-hourly Stochastic Time-Inverted Lagrangian Transport (STILT, see Sect. 3.2.1) model footprints from each ICOS station with annual mean $^{14}CO_2$ emissions from the European Radioactive Discharges Database (Maier et al., 2023; RADD, 2025). This methodology, assuming constant annual discharge rates (Maier et al., 2023), introduced systematic biases because nuclear power plants (Boiling Water Reactors, BWRs; and Pressurized Water Reactors, PWRs) exhibit time-dependent emission profiles (Knaack, 2025). This often led to overestimation of PWR ($\sim$22%) and underestimation of BWR discharges ($\sim$20%), biasing $\Delta ffCO_2$ estimates near such facilities (Knaack, 2025).

Recent advancements by Knaack, 2025 addressed these biases using simplified reactor type-specific time-dependent $^{14}CO_2$ discharge profiles which are based on reactor-specific operating conditions. This approach utilizes annual $^{14}CO_2$ discharge data (Laemmel et al., 2025), hourly electricity generation, and refueling outages (Entso-E, 2025). For BWRs, the hourly discharge rate of $^{14}CO_2$ correlates with the energy production rate and was therefore calculated from the annual energy production data and the annual discharge rates. During periods of no energy production, a specific zero energy $^{14}CO_2$ production rate is applied. Samples taken at the start of refueling outages are flagged due to potentially large unknown ^{14}C releases during reactor depressurization. For PWRs, the approach uses annual ^{14}C discharge data and outage periods, assuming that 23% of the ^{14}C discharge is $^{14}CO_2$ (the rest being $^{14}CH_4$). The normal operational baseline discharge is 10-40% of the mean annual rate depending on the reactor type and is set to 37% (median over available reactor data), although newer data suggest that this could be an overestimation by $\sim$15%, leading to a positive bias in the PWR nuclear correction ($\sim$40%, not published). However, since the PWR component is not large ($\sim$20% of total nuclear influence), this error in total should be rather small. During refueling outages, discharge is assumed to double the annual mean, with a significant spike (50 times the annual average) at the onset of depressurization due to rapid ^{14}C release. For the ICOS flasks samples on average, this methodology has reduced the estimates of $\Delta ffCO_2$ by 0.28 ppm, correcting a slight overestimation of the previous average annual discharge rates. An uncertainty of 100% is applied to the modeled $\Delta^{14}C_{nuc}$ contributions to account for the inherent temporal and transport uncertainties. Due to data availability, this method corrects only for nuclear influence inside the European domain. I assume that the background station captures all external nuclear contributions, thus accounting for them during the $\Delta ffCO_2$ calculation.

Beyond methodological improvements for nuclear power plants, the analysis shows that the La Hague nuclear fuel reprocessing site is the dominant contributor of ^{14}C to almost all ICOS stations, accounting for approximately 40% of all nuclear corrections at the ICOS sites (Knaack, 2025). This highlights the critical need for a more rigorous investigation of its specific $^{14}CO_2$ discharge characteristics to further enhance the accuracy of $\Delta ffCO_2$ estimation. The finalized data flagging procedure in this study involves: 1) dropping values without nuclear correction, 2) flagging data with revision or maintenance flags indicating a high unknown contamination risk, 3) dropping flasks with high nuclear contributions ($>$ 2‰), and 4) discarding flasks with values significantly higher than the background ($>$ 2 times the ^{14}C measurement error plus background error), indicating potential unaccounted nuclear influence.

3.2 Transport model and Bayesian inversion

In situ measurements of atmospheric mole fractions are crucial for understanding the spatial distribution and magnitude of greenhouse gas sources and sinks. However, atmospheric transport significantly alters these surface influences, mixing them across vast regional and continental areas. Trace gas concentrations within the Planetary Boundary Layer (PBL) are particularly sensitive to surface emissions. The highly inhomogeneous patterns of sources and sinks in the near-field of a measurement site cause large variability in the measured concentration data. Given that the PBL ventilates in about four days (Cotton et al., 1995), this near-field surface area can be extensive, potentially covering hundreds to thousands of kilometers (regional to continental scales) with which the air in the PBL interacts before being observed. Therefore, accurate atmospheric transport models are essential to connect observed trace gas concentrations to surface emissions and sinks.

The PBL's mixing behavior changes significantly between day and night (Fig. 3.3). During the daytime, convective turbulence drives the formation of a well-mixed layer. As sunset approaches, thermal convection and turbulence diminish, thus, a "residual layer" forms that retains the properties of the formerly mixed layer. Throughout the night, a stable boundary layer forms, growing upwards from the ground, transitioning smoothly into the residual layer (Stull, 2012). Current atmospheric transport models represent the daytime mixed layer relatively accurately (Trusilova et al., 2010), often leading to a preference for daytime over nighttime sampling for applications in inverse modeling. A critical implication for trace gas measurements is that daytime surface emissions mix into a much larger air volume than nighttime emissions. This means that nighttime measurements, despite their challenges, offer higher sensitivity to regional surface fluxes. Lagrangian particle dispersion models offer a potential solution by simulating turbulence and capturing subgrid-level transport

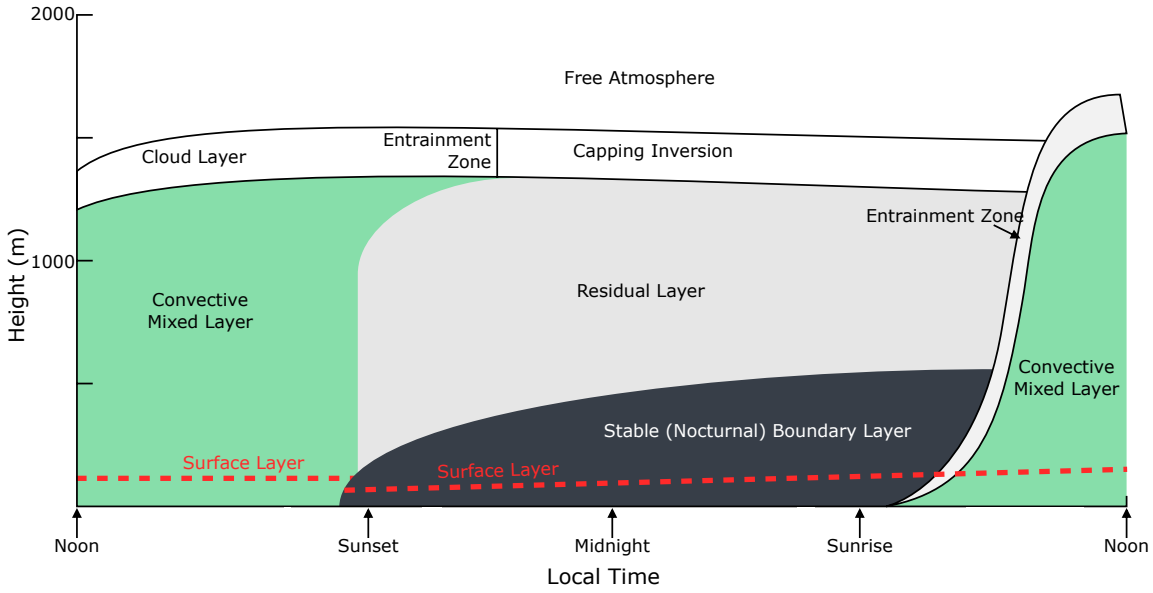

Figure 3.3. The Planetary Boundary Layer (PBL) in high-pressure regions typically consists of three main parts: a turbulent mixed layer during the day, a less turbulent residual layer that holds air from the former mixed layer, and a stable boundary layer that forms at night with sporadic turbulence. Adapted from (Stull, 2012).

processes.

3.2.1 STILT

The Stochastic Time-Inverted Lagrangian Transport model (STILT, Gerbig et al., 2003; Lin et al., 2003), a Lagrangian particle dispersion model (Zannetti, 1990), was the main atmospheric transport model used in this study. STILT determines the upstream influence region, or ”footprint,” of atmospheric measurement sites by tracking an ensemble of particles (e.g., 100 particles hourly) backward in time (Fig. 3.4) from receptors (observation sites). Convolving these footprints with surface fluxes (e.g., emission inventories) simulates atmospheric tracer concentrations at observation stations (forward or FWD simulations). The back-trajectories are calculated using meteorological fields and a stochastic representation of turbulent motions, potentially resolving subgrid processes as particle locations are calculated without grid cell restrictions. These back-trajectories are computed for 10 days or until particles exit the defined model domain (”stilt03” domain in CarboScope notation: 16°W - 36°E and 32°N - 74°N). Particles in the PBL are aggregated to footprints with a specified grid resolution (e.g., 0.25° x 0.25°). STILT is driven by meteorological fields from the fifth generation of the European Center for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5, Soci et al., 2024). This approach addresses challenges from ”near-field” variability and complex PBL dynamics, which often hinder traditional gridded

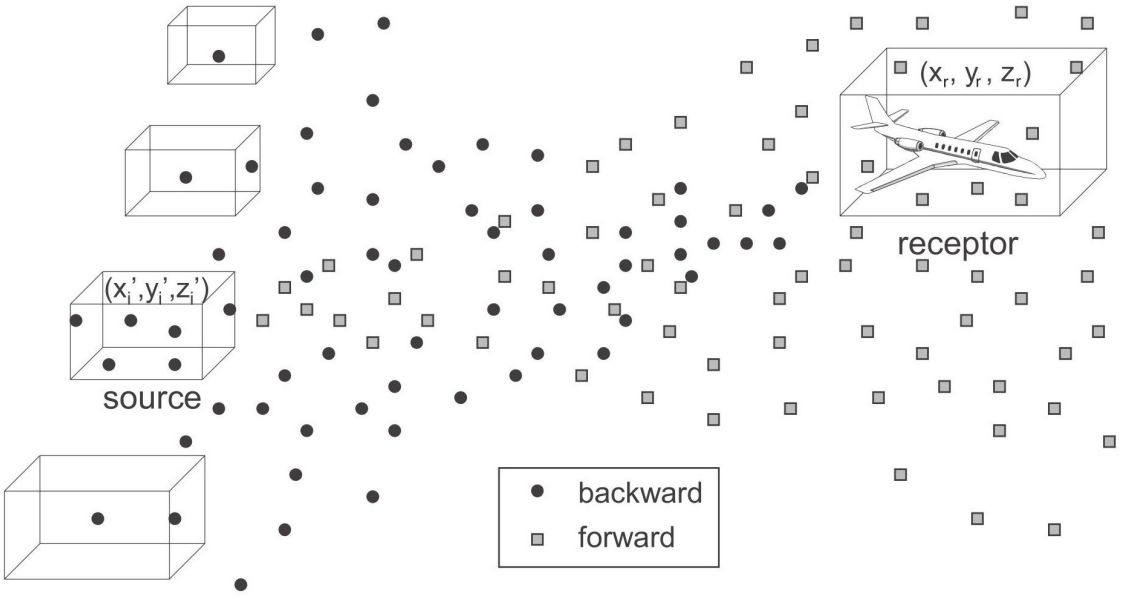

Figure 3.4. Comparing backward and forward time simulations for atmospheric transport. A single backward release of particle ensemble from a receptor (e.g. measurement site) identifies its potential spatio-temporal source region. In contrast, forward time simulations are computationally far more extensive as those require multiple particle releases from the entire domain to achieve the same. An empirical test of STILT model reversibility verified this by comparing particle counts in source/receptor boxes from backward and forward runs, accounting for air density differences. Adapted from (Lin et al., 2003).

models, by simulating turbulence and capturing subgrid-scale transport via a Markov chain process for turbulent velocity statistics (Lin et al., 2003). This receptor-oriented framework is particularly suitable for representing the ICOS ground-based network of tall towers.

A critical aspect of STILT simulations involves the management of boundary conditions. For particles leaving the European model domain, concentrations are set to zero, completely neglecting all fluxes outside the domain. This procedure on the model side mimics the calculations of $\Delta ffCO_2$ estimates using the RIBA, where the clean MHD marine background curve is subtracted from the data (Sect. 3.1). As discussed in Sect. 3.1.1, this assumption holds only for air masses originating from the western boundary, but Maier et al., 2023 showed that the representativeness bias for other boundaries is rather small.

Although STILT robustly derives footprints and simulates concentrations, transport model errors, including the accuracy of the underlying meteorological data, pose significant challenges for inverse modeling. As discussed in Lin and Gerbig, 2005, these errors can arise from various sources, including inadequate PBL dynamics representation and advection uncertainties. Past quantification approaches, such as comparing multiple atmospheric models or analyzing residuals between simulated and observed tracer time series, have limitations.

For example, collaborative model development can lead to similar parameterizations, and residual analysis often confuses transport errors with assumed tracer flux errors (Gerbig et al., 2008). Gerbig et al., 2008 further highlighted that vertical mixing uncertainty propagating into the mixing ratio uncertainties of CO_2 can be around 3 ppm, while advection errors during active vegetation periods can contribute up to 5 ppm, substantially exceeding typical measurement uncertainties. To mitigate these transport errors, Gerbig et al., 2008 outlined three main strategies: quantifying and propagating the errors via application of typical transport tracers (e.g. Radon-222, Gachkivskyi et al., 2025; Maier et al., 2025), improving the transport model, or employing less sensitive approaches. However, since the main focus of this study is to evaluate observation data content within the same inversion system, I concentrate on afternoon hours (for flasks and CO-based $\Delta ffCO_2$), which typically exhibit the lowest transport-related errors (Geels et al., 2007; Peng et al., 2023), or assume the transport processes aggregated over longer periods of time are representative of mean atmospheric conditions (integral- and CO-based $\Delta ffCO_2$, see Sect. 3.2.2.1). Additionally, I use different ratios of model-data mismatch uncertainty to prior uncertainty to investigate its effect on the inversion system (parameter μ, Sect. 3.2.2).

3.2.2 CarboScope inversion system

The forward model, a crucial component in inverse problems, translates GHG fluxes e.g. from emission inventories (in this study from GCP or EDGAR, Sect. 3.2.3) into atmospheric concentrations at a observation sites. In GHG inversions, this typically involves atmospheric transport models (e.g., STILT, Sect. 3.2.1) that simulate GHG transport, mixing, and observation at measurement sites, considering atmospheric dynamics like winds, convection, and turbulence. These forward simulations (FWD) are represented as:

$$\mathbf{c}_{\text{mod}} = \mathbf{A}\mathbf{f} + \mathbf{c}_{\mathbf{ini}} \tag{3.4}$$

where $\mathbf{A}$ is a transport matrix that relates surface fluxes ($\mathbf{f}$) to modeled concentrations ($\mathbf{c}_{\text{mod}}$), and $\mathbf{c}_{\mathbf{ini}}$ accounts for initial or background concentrations (Rödenbeck, 2005a).

However, both transport models and emission inventories are prone to errors (Lin & Gerbig, 2005; Gerbig et al., 2008; Super et al., 2020), which requires independent validation of simulations. This validation can be achieved by estimating surface fluxes from measured ambient GHG concentrations, posing a classic inverse problem: deducing cause from observed quantities. The fundamental relationship is expressed as:

$$\mathbf{c}_{\text{meas}} = \mathbf{c}_{\text{mod}} + \epsilon \tag{3.5}$$

where $\mathbf{c}_{\text{meas}}$ represents the measured concentrations and ϵ indicates errors (model-data-mismatch). The fluxes that I aim to retrieve are parameterized in the CarboScope system using a linear

flux model:

$$\mathbf{f} = \mathbf{f_{fix}} + \mathbf{Fp} \tag{3.6}$$

Here, $\mathbf{f_{fix}}$ is an a-priori (prior) flux estimate (typically in fossil fuel inversions from an emission inventory), the matrix $\mathbf{F}$ contains prior data on flux uncertainties and correlations (defining the a-priori covariance matrix $\mathbf{Q_{f,pri}} = \frac{1}{\mu}\mathbf{FF^T}$, where μ is a scaling factor), and $\mathbf{p}$ is a dimensionless vector parameter representing adjustments to this prior (Rödenbeck, 2005a). This parameterization defines the a-priori probability distribution of fluxes, assuming $\mathbf{p}$ has an a-priori mean of zero and unit variance. The total modeled concentration then becomes:

$$\mathbf{c}_{\text{mod}} = \mathbf{c_{mod,fix}} + \mathbf{AFp}, \text{with} \tag{3.7}$$

$$\mathbf{c_{mod,fix}} = \mathbf{Af_{fix}} + \mathbf{c_{ini}} \tag{3.8}$$

To find the optimal value of the parameter $\mathbf{p}$, a Bayesian inversion is used. Bayes' theorem (Bayes, 1763; Joyce, 2003), a fundamental principle of probability theory, describes the probability of an event given its probabilistic dependence on another event and the likelihood of that other event. It combines prior information with new observations to yield revised (posterior) information. Formally, for events A (e.g., flux estimation of emission inventory) and B (e.g., observations), Bayes' theorem states:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.9}$$

where $P(A|B)$ is the posterior probability of A given B, $P(B|A)$ is the probability of B given A (as given by the model), $P(A)$ is the prior probability of A before considering observations B (e.g. from prior uncertainty), and $P(B)$ is the likelihood of observing B (e.g. from measurement errors). This theorem is widely applied across various fields, including inverse problems, where it refines estimates of underlying physical states using observational data.

The following section is adapted from Rödenbeck, 2005a. In the context of inverse problems, as implemented in the CarboScope inversion system (Rödenbeck et al., 2003; Rödenbeck et al., 2023; Maier et al., 2025), Bayes' theorem is defined as:

$$P(\mathbf{p}|\mathbf{m}) = \frac{P(\mathbf{m}|\mathbf{p})P(\mathbf{p})}{P(\mathbf{m})} \tag{3.10}$$

where $\mathbf{m} = \mathbf{c_{obs}} - \mathbf{c}_{\text{mod}}$ represents the Model-Data-Mismatch (MDM) with $\mathbf{Q_m}$ as its covariance matrix. If the true value of $\mathbf{p}$ is known, MDM is attributed solely to random errors in measured and modeled concentrations, with a zero mean, leading to a (conditional) probability distribution of $\mathbf{m}$ proportional to $\exp(-\frac{1}{2}\mathbf{m^T Q_m^{-1} m})$. Furthermore, from the definition of $\mathbf{p}$, its (unconditional) probability distribution $P(\mathbf{p})$ is proportional to $\exp(-\frac{\mu}{2}\mathbf{p^T p})$. The

scaling factor μ (being 1 in the base setup) scales the impact of a-priori constraints (the ratio between a-priori and data constraints determining how strongly the solution is regularized by a-priori information). μ was introduced for ease of sensitivity tests, as the overall magnitudes of both $\mathbf{Q_m}$ and the a-priori flux covariance matrix $\mathbf{Q_{f,pri}}$ are often ill-determined in reality. With this, I define a cost function J:

$$J = -\ln\left(P(\mathbf{p}|\mathbf{m})\right) \tag{3.11}$$

$$= \frac{1}{2}\mathbf{m^T Q_m^{-1} m} + \frac{\mu}{2}\mathbf{p^T p} + C \tag{3.12}$$

where C is an additive constant that contains all $\mathbf{p}$-independent terms (e.g., $P(\mathbf{m})$ and normalization). As seen from Eq. 3.12, J is constructed from two essential components: a data constraint term $\mathbf{m^T Q_m^{-1} m}$, which penalizes large MDM and where the inverse of the covariance matrix Q_m^{-1} acts as a weighting factor (smaller uncertainties lead to larger penalties), and an a-priori flux constraint ($\frac{\mu}{2}\mathbf{p^T p}$), which regularizes the problem by penalizing large deviations of $\mathbf{p}$ from its a-priori expectation value (zero).

The goal of the inversion is to find the most probable value of $\mathbf{p}$ given the observed MDM. Hence, $\langle \mathbf{p_{post}} \rangle$ is found by maximizing $P(\mathbf{p}|\mathbf{m})$ with respect to $\mathbf{p}$ which is equivalent to minimizing the cost function J:

$$\left.\frac{\partial J}{\partial \mathbf{p^T}}\right|_{\mathbf{p}=\langle \mathbf{p_{post}} \rangle} = 0 \tag{3.13}$$

The cost function is numerically minimized using an iterative conjugate gradient algorithm (Rödenbeck, 2005a).

Formally, after inserting the previous equations into Eq. 3.12 and taking the derivative, the optimal parameter values ($\mathbf{p_{post}}$) are calculated:

$$\mathbf{p_{post}} = \mathcal{A}^{-1}\mathbf{b} \quad \text{where} \quad \mathcal{A} = \mathbf{F^T A^T Q_m^{-1} AF} + \mu\mathbf{1} \quad \text{and} \quad \mathbf{b} = \mathbf{F^T A^T Q_m^{-1}}(\mathbf{c_{obs}} - \mathbf{c_{mod,fix}}) \tag{3.14}$$

Consequently, the posterior fluxes ($\mathbf{f_{post}}$) in the original notation are (Rödenbeck, 2005a):

$$\mathbf{f_{post}} = \mathbf{f_{fix}} + \mathbf{F}(\mathbf{F^T A^T Q_m^{-1} AF} + \mu\mathbf{1})^{-1}\mathbf{F^T A^T Q_m^{-1}}(\mathbf{c_{obs}} - \mathbf{c_{mod,fix}}) \tag{3.15}$$

3.2.2.1 Averaging intervals and data-density weighting

Integrated samples represent the average atmospheric ΔffCO$_2$ signal over approximately two-week periods. To align these with the model output, the ΔffCO$_2$ value of each integration period was assigned to all individual hours during this period. On the model side, each integrated sample was then divided into ”averaging groups,” where hourly MDM values were replaced by weighted averages over the corresponding period for cost function calculations (Rödenbeck, 2005b, update 14.007). Crucially, the MDM vector length and the

original hourly MDM errors are preserved. This technique was also applied to CO-based ΔffCO$_2$ estimates. Maier et al., 2024b demonstrated that one-week averaging of hourly CO-based ΔffCO$_2$ data substantially reduces MDM spikes caused by faulty point source emission representation (e.g. due to transport model errors).

Additionally, when comparing inversions using different data types (e.g., flasks and integrated samples), it is crucial to prevent sampling method differences from disproportionately shifting data constraints in one of the inversions. This effect is evident in the contribution of MDM to the cost function J (J^*, Rödenbeck, 2005a):

$$J^* = \frac{1}{2} \sum_{i=1}^{N^*} \frac{(c_{obs,i} - c_{mod,i})^2}{\sigma_i^2} \tag{3.16}$$

where the sum is over all individual values in the considered period. If σ_i^2 are approximately equal, J^* increases proportionally to N^*, which on a weekly scale is typically 1 (or less) for flasks, 7×5 for CO-based data (afternoon hours only) or 7×24 for integrals (per week). Without adjustment, inversions with higher data density would experience disproportionately stronger data constraints simply due to larger data volume.

To mitigate these differences, I employed data density weighting. As described by Rödenbeck, 2005a, this technique ensures that all observations within a specified interval (typically one week, the time scale of synoptic weather patterns) contribute the same constraint as a single observation per week. This method artificially increases the MDM error based on the number of observations in the surrounding time interval:

$$\sigma_i^* = \sqrt{N^*} \cdot \sigma_i \tag{3.17}$$

This can also prevent sites with high-density continuous measurements from disproportionately influencing inversion results compared to, for instance, flask sites, if used within the same inversion. In this study, it leads to more comparable results of different proxies. Additionally, this method approximately accounts for temporal correlations between consecutive observations within a typical timescale of synoptic events (one week). More details on data density weighting are available in Rödenbeck, 2005a.

3.2.2.2 Used parameters and degrees of freedom

I used the standard CarboScope approach of exponentially decaying spatial correlation in longitudinal and latitudinal directions, which can be interpreted as spatial smoothing of the previously uncorrelated a-priori flux regions (pixels) (Rödenbeck, 2005a). In the inversions of this study, a relatively large spatial correlation length (compared to the extent of the investigated area, Sect. 4.2) of approximately 380 km in both directions was used, though this is comparable to previous setups for the European domain (Rödenbeck et al., 2023;

Maier et al., 2025). This choice aligns with the study's goal of resolving country-level emissions and temporal variability rather than small-scale spatial structures. The full spatial and temporal resolution of the resulting fluxes (model pixels and daily time steps) are not the primary quantities of interest of this study and are furthermore not well constrained by the available data. Therefore, the spatial aggregation primarily focuses on Germany and surrounding areas, as this region is best constrained by the measurement data currently available (Sect. 4.2).

Time correlations can be introduced by convolving the time series with a pulse response function. In CarboScope, this is implemented in the frequency domain (Fourier series), where the time correlations become a weighting pattern among the Fourier terms. As a temporal correlation filter, I used the one corresponding to a truncated triangular response function that linearly rises/falls away from the central peak and reaches zero at $\pm 1/\nu_{\text{low}}$ in the time domain. This function represents a low-pass filter applied to the Fourier decomposition of time series in the frequency domain, removing frequencies equal to or higher than ν_{low} (Rödenbeck, 2005a). Filters are referred to here in the format "FxT", with "F" and "T" denoting the "Filter" and "Truncated triangular" properties, whereas "x" defines the cut-off frequency: $\nu_{\text{low}} = x/\text{yr}$. In synthetic studies a range of temporal correlation lengths was explored, from biweekly (CarboScope notation would be "Filt24T") through five-yearly ("F0.2T") to a single temporal parameter for the entire period ("GLT") (Rödenbeck, 2005a). Synthetic studies demonstrated that a half-yearly correlation length ("F2T", Sect. 4.4) yields optimal inversion results and allows for investigation of the seasonal cycle in the ΔffCO_2 data.

In the CarboScope inversion system, transport model uncertainties are chosen depending on a observation site type. For the ICOS sites used in this study, a single station type (continental tall towers) was assumed, with an associated transport error of 1 ppm. While this classification is largely valid, it might slightly underestimate the error for certain high-altitude locations (e.g., JFJ or OXK) due to more complex transport processes. Furthermore, as described in Sect. 3.2.1, this assumed transport error may be significantly underestimated. However, it is challenging to disentangle the influence of emission inventory and transport errors, which are not well known, on the posterior fluxes. Consequently, instead of increasing the prescribed transport error, a range of scaling parameters μ (see Sect. 3.2.2), in the form of different ratios of MDM uncertainty to prior uncertainty, were investigated. For simplicity, the measurement error for all measurements is uniformly set to 1.5 ppm, a value consistent with typical uncertainties in ^{14}C-based ΔffCO_2 measurements. All STILT and CarboScope calculations carried out in this thesis were done using the fourth High Performance Computer System for Earth System Research (HLRE-4) "Levante" (DKRZ, 2025).

3.2.3 Apriori Information

According to the United Nations Framework Convention on Climate Change (Grubb, 2019), countries are required to report their greenhouse gas emissions to ensure transparency, accountability, and progress toward climate goals, which were further disaggregated into spatio-temporal data sets of the emission fluxes. Accurate and high spatio-temporal resolution of these fossil estimation products coupled with precise atmospheric transport models is crucial for reliable simulation of atmospheric CO_2 concentrations. This thesis utilizes two such datasets. The Global Carbon Project - Gridded Fossil Emissions Dataset (GCP-GridFED, or GCP, Jones et al., 2021) was used as the main emission inventory product in all sensitivity runs. In addition, the Emissions Database for Global Atmospheric Research (EDGAR v4.3.2 or EDGAR, Janssens-Maenhout et al., 2019) was used in Bayesian inversions with unmodified priors to provide a comparison basis with the GCP results (see Sect. 5.2.4, 6.3.4 and 7.2.4).

The GCP-GridFED (Jones et al., 2021) is designed to be consistent with the annual national emission estimates compiled by the Global Carbon Project for its annual Global Carbon Budget (Friedlingstein et al., 2024). The methodology involves scaling an existing gridded monthly emission distribution (specifically, EDGAR v4.3.2 for 2010) to match the national annual totals from GCP's National Annual Emissions (GCP-NAE) dataset, which is based on the United Nations Framework Convention on Climate Change (UNFCCC) national submissions, International Energy Agency (IEA) energy statistics, and global cement production data. For recent years (from 2019 onward), GCP incorporates dynamic monthly seasonality from datasets like Carbon Monitor and applies corrections for inter-annual climate variability or global events, such as the COVID-19 pandemic. The resulting data are available with $0.1^\circ \times 0.1^\circ$ spatial resolution and monthly temporal resolution, interpolated to daily fluxes.

EDGAR (Janssens-Maenhout et al., 2019) is a globally derived anthropogenic emission inventory. It incorporates updated annual energy statistics from the IEA and British Petroleum on fossil fuel consumption. Then these national-level emissions are spatially and temporally disaggregated based on fuel type, emission category, and country-specific emissions using the COFFEE method (Steinbach et al., 2011). EDGAR provides data at a $0.1^\circ \times 0.1^\circ$ spatial grid and an hourly temporal resolution.

Both GCP-GridFED and EDGAR represent state-of-the-art products with global coverage of fossil CO_2 emissions, guided by IPCC methodologies (IPCC, 2006). They can be used with the STILT transport model and the CarboScope inversion system when aggregated to a $0.25^\circ \times 0.25^\circ$ grid. The fundamental distinction between the two products lies in their construction philosophies. EDGAR is a self-contained bottom-up inventory with national and gridded emissions calculated directly from activity data and emission factors. In

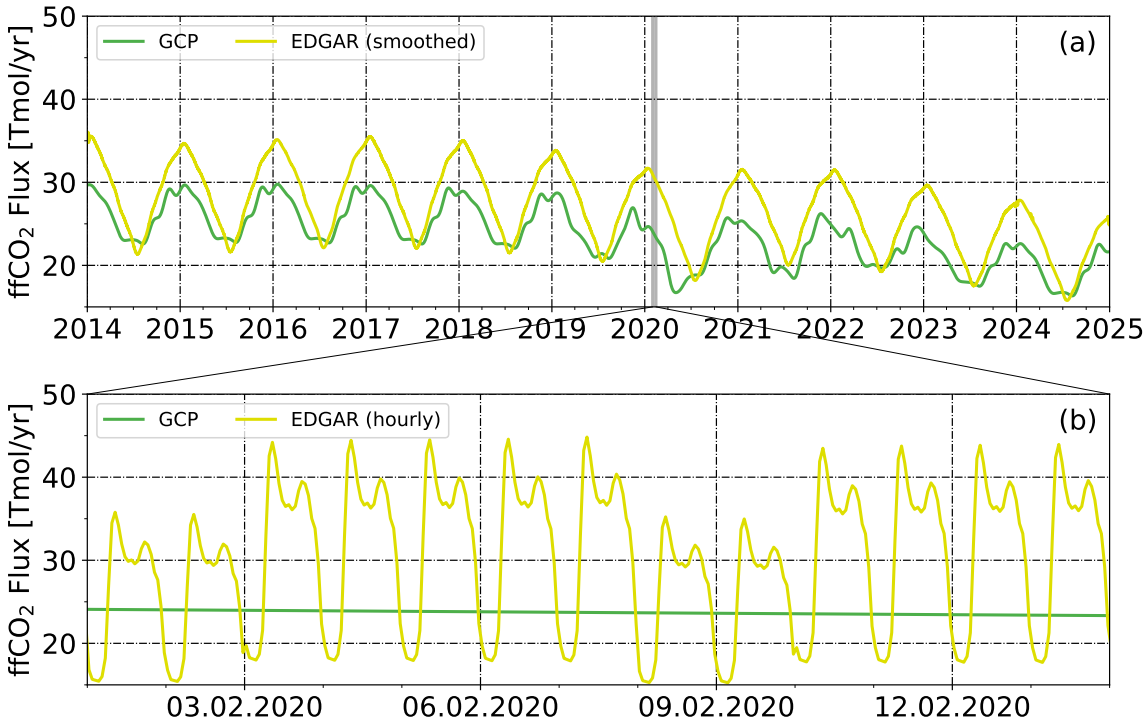

Figure 3.5. Fossil CO_2 fluxes for the Area of Constraint (AoC, 5°E - 16°E and 47°N - 56°N, see Sect. 4.2) from the GCP and EDGAR emission inventories. Panel (a) shows the emissions for the investigated time period with 14 days smoothed EDGAR emissions. Panel (b) shows zoom-in over 2 weeks with hourly EDGAR emissions as they were used in the inversion.

contrast, GCP-GridFED is a "constrained" gridded product, designed to align gridded fossil CO_2 emissions with the national annual totals of the Global Carbon Project. This leads to differences in national total emission fluxes and their spatio-temporal distribution (Fig. 3.5). For the investigated region (5°E - 16°E and 47°N - 56°N, see Sect. 4.2), EDGAR exhibits 11-19% higher annual fluxes in 2014-2024 compared to GCP annual means, with the largest difference during the COVID pandemic (2020 and 2021). Mean summer EDGAR emissions are typically closer to GCP estimates (EDGAR 3-15% higher) than winter fossil fluxes (EDGAR 15-23% higher).

Uncertainties in GHG budgets can hinder the understanding of modeling results. The gridded GCP product adopts error estimates from the Global Carbon Budget (Friedlingstein et al., 2024), specifying a 5% uncertainty (1σ) for the 42 Annex I countries reporting annually to the UNFCCC and a 10% uncertainty for other countries. EDGAR estimates the uncertainty in global total anthropogenic CO_2 emissions at $\pm 4.5\%$ (1σ), with lower uncertainties for EU15 countries at 2.5% (Janssens-Maenhout et al., 2019). However, as shown in Fig. 3.5, the annual emission difference for the Area of Constraint (AoC, Sect. 4.2) between these products exceeds their combined uncertainty estimates.

Furthermore, Super et al., 2020 highlighted that significant uncertainties arise when total national fossil emissions are disaggregated into regional resolutions, with the uncertainty for the entire AoC expected to be around 1%, but increasing to 40% for grid cell disaggregation. Although GCP uses 5-10% uncertainties for its gridded emissions, Jones et al., 2021 note that spatially averaged uncertainties from disaggregating national emissions to grid cells can range from 20% to 75% (1σ) at spatial resolutions from 1 km to 1°. This is attributed to incomplete proxy data coverage (e.g., unmapped point sources), poorly constrained nonlinearities (e.g., emission intensity differences between rural and urban areas of equal density), shortcomings in continuous proxy values (e.g., poorly constrained population density), or inappropriate spatial representativeness.

In conclusion, the uncertainties of emission inventories are not fully known, especially at the grid scale with sub-annual temporal resolution. These facts should be considered when evaluating posterior results and comparing them to emission inventories. Similar to the treatment of transport errors, the parameter μ (Sect. 3.2.2) is used to scale prior errors and investigate their effects on the posterior.

3.2.4 Evaluation Matrix

Beyond the Bayesian inversions performed with unmodified priors, which utilized my best estimate for the prior information and uncertainty, a set of inversion experiments called the Evaluation Matrix was performed utilizing specifically modified priors. The goal of the Evaluation Matrix design was to investigate the influence of the observations on certain features of the posterior fluxes compared to the GCP inventory. The Evaluation Matrix approach investigates this systematically by removing these pieces of information from the prior. The inversion algorithm then inferred the missing flux component solely from the provided observational data. Flux components describing main features of the emission field were analyzed: seasonality and inter-annual trends of fossil fluxes, as well as absolute and annual $\Delta ffCO_2$ emissions. For each component analysis, a specific modified prior was constructed based on the GCP inventory. The appropriate CarboScope inversion system filter and the characteristics of these modified priors are noted in Table 3.1. The modified priors used in this study are shown in Figure 3.6 and are described in the following. To give the inversion flexibility to compensate for the modifications to the priors, the prior uncertainty was increased to 40% for the experiments of the Evaluation Matrix inversions (see Sec. 4.5). The analysis method of the Evaluation Matrix for each of the four flux components was designed as follows.

The seasonality of the $\Delta ffCO_2$ flux can be described with two parameters, the amplitude and the phase of the seasonal cycle. The modified prior for this analysis called the

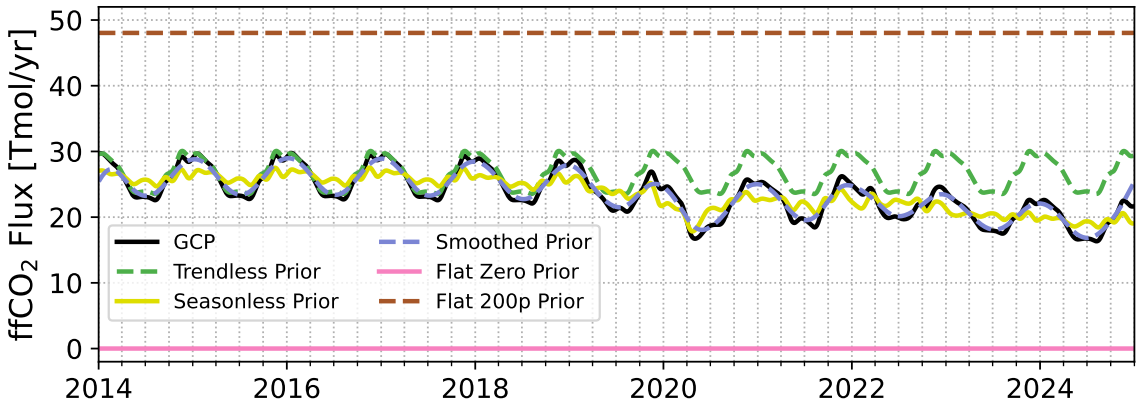

Figure 3.6. Modified priors used in this study, calculated based on the GCP inventory (Sect. 3.2.3) utilizing an appropriate time correlation CarboScope filter (Sect. 3.2.2.2). For descriptions and intended usage see Tab. 3.1.

Table 3.1. Description of the priors used in the sensitivity study and and their intended use. The name refers to the designation used throughout this thesis. All modified priors are based on the GCP inventory and are calculated by applying the corresponding Carbo Scope filter.

Prior name	Carbo Scope notation	Characteristic	Application
GCP	None	Global Carbon Project $\Delta ffCO_2$ inventory	Optimal Bayesian Inversion
Trendless Prior	`-Filt52TdZs`	Mean seasonal cycle with removed inter-annual trend	Analysis of the inter-annual trend
Seasonless Prior	`-Filt4Tma`	Inter-annual trend with short-term variability, no seasonal cycle	Analysis of the seasonality
Smoothed Prior	`Filt4Txx`	GCP with removed short-term variability	Analysis of the deviations from GCP
Flat Zero Prior	`LT` $\times 0$	Removed all time dependence, zero emissions, present spatial emission pattern	Analysis of the data driven inversion
Flat 200p Prior	`LT` $\times 2$	Two times mean $\Delta ffCO_2$ emissions over the time period, no temporal features	Analysis of the data driven inversion

Seasonless Prior was calculated by removing the seasonal cycle from the GCP prior and provides the inversion system with only the inter-annual trend and the short-term variability information (Fig. 3.6 yellow solid line). The posterior fluxes as well as the undisturbed

GCP flux were divided by the Seasonless Prior to evaluate and compare their seasonality. The relative amplitudes of the winter maximum and the summer minimum as well as their positions were compared.

The inter-annual trend in the measurement data was analyzed using the so-called Trendless Prior (green dashed line in Fig. 3.6), which only contains the mean seasonal cycle and the short-term variability of the GCP inventory with the inter-annual trend removed. Similarly as for the analysis of the seasonality, the resulting posteriors were divided by the Trendless Prior. The slopes of the linear regressions of the absolute and relative flux estimates were compared to the inter-annual trend of the GCP.

The absolute and annual mean ΔffCO$_2$ emissions were analyzed using two different modified priors, the Flat Zero Prior and the Flat 200p Prior (pink solid line and brown dashed line, respectively, in Fig. 3.6). For the Flat Zero Prior, I calculated the mean pixel-wise emissions of the GCP inventory over the whole investigated time period and then multiplied it by zero. However, it is important to note that the Flat Zero Prior (and later the Flat 200p Prior) constructed this way retained the pixel-wise time-averaged prior uncertainties of the original GCP inventory with higher uncertainty in high emission regions. The Flat 200p Prior was calculated in the same way as the Flat Zero Prior, but the emissions were multiplied by the factor of two instead of zero, amounting to 200% of the GCP mean.

The posteriors of these inversions were then aggregated to annual mean ΔffCO$_2$ emissions and compared to the GCP inventory yearly means. This last step in the postprocessing was done to mimic the national emission total before they are disaggregated in spatial and temporal domains, because the disaggregation causes additional errors in the emission inventories.

Additionally, I used Root Mean Square Deviation (RMSD) to compare the not aggregated posterior emissions of the Flat Zero and Flat 200p Priors with the smoothed variant of the GCP inventory (the Smooth Prior, blue dashed line in Fig. 3.6). This was done to facilitate a fair comparison, as the applied time-correlation parameter "F2T" (Sect. 3.2.2.2), which only allowed for the seasonal oscillations in the data-driven posterior adjustments to the prior (see Sect. 4.4), none of the short-term variability of the GCP inventory could be reproduced by the inversion. This allowed me to quantify the overall difference between the data-driven inversions and this inventory.

It should be emphasized that the inversions based on the real data do not have an objective truth to compare with, restricting the evaluation of the flux features to the potentially biased GCP inventory (Super et al., 2020). To determine the significance of the calculated data-driven posterior deviations from the GCP inventory, this Evaluation Matrix was also applied to the synthetic data inversions (Sect. 4.1) with known truth (Sect. 4.6). The Evaluation Matrix experiments with synthetic data show the magnitude of the expected deviations

based only on the measurement data errors without the transport error effects. Comparing the corresponding Evaluation Matrix results from real and synthetic measurement data allows us to assess how the deviations in the real-data posterior compare to those expected solely from observational uncertainties, under the idealized assumption of zero model uncertainty.

Chapter 4

Sensitivity studies

4.1 Synthetic Runs

Before analyzing $\Delta ffCO_2$ inversion results based on actual observational data, it is essential to investigate the validity range of key inversion parameters, such as the spatial and temporal domains in which robust conclusions can be drawn or meaningful variability can be detected. To establish this range, sensitivity analyses were carried out in the form of so-called synthetic inversion runs. In synthetic runs, the same parameters are used as in regular real-data inversions, but instead of working with the measurement data, synthetic $\Delta ffCO_2$ values are utilized. Synthetic $\Delta ffCO_2$ data are generated from forward (FWD) runs of the transport model (Sect. 3.2.1) employing $\Delta ffCO_2$ emission fluxes from bottom-up emission inventories (e.g. from GCP or EDGAR, Sect. 3.2.3), such that the true fluxes behind these data is known. With that, the posterior flux estimates of the synthetic inversions can be analyzed through comparison with their known truth. By setting the measurement uncertainties of the synthetic data to zero and using the same transport model in the FWD runs (Sect. 3.2.1) as in the inversion process, thereby eliminating transport model uncertainties, one can assess the performance of the inversion and its dependence on inversion parameters in an idealized setting. Additionally, the prior fluxes of the synthetic inversions can be modified (e.g., removed seasonal cycle or altered inter-annual trend) to test the inversion system's ability to reproduce the modified features from the available data. Furthermore, the impact of measurement uncertainties on the inversion results can be assessed by randomly perturbing data obtained from the FWD runs. This enables to quantify the variability in the posterior $\Delta ffCO_2$ fluxes that arises solely from observational errors. This variability represents one of the two key components required to evaluate any deviations between prior and posterior fluxes in inversions based on real data. The second source of variability stems from transport and representation errors in the model. However, since I use the same transport model for forward simulations and inversion, this component cannot be evaluated within

current framework. Nevertheless, conducting the sensitivity studies enables a well-founded determination of the most critical inversion parameters based on the characteristics of the observational network: its spatial and temporal coverage and its associated measurement uncertainties.

To simplify the interpretation of the sensitivity experiments, I define a truth recovery rate ρ of the synthetic inversion runs similar to Rödenbeck et al., 2023:

$$\rho\left(x,y,t\right) = \frac{f_{post}^{ffCO_2}\left(x,y,t\right) - f_{pri,manip}^{ffCO_2}\left(x,y,t\right)}{f_{true}^{ffCO_2}\left(x,y,t\right) - f_{pri,manip}^{ffCO_2}\left(x,y,t\right)} \cdot 100\% \tag{4.1}$$

where $f_i^{ffCO_2}\left(x,y,t\right)$ denotes the ΔffCO$_2$ fluxes for the posterior (index 'post'), the manipulated prior (i = pri,manip) and the known truth (i = true). The truth recovery rate ρ is 100% when the posterior flux completely reconstructs the known truth and 0% when the posterior flux remains the same as the manipulated prior flux.

4.2 Analysis of the Area of Constraint

One of the most important parameters for the interpretation of the atmospheric modeling results is the spatial Area of Constraint (AoC). In this study, I use the samples collected by the ICOS measurement network (Sect. 2.3) and are restricted by its spatial extent. Therefore, I must determine which area can be best examined within the given station distribution.

Similarly to previous studies using the RIBA (Levin & Rödenbeck, 2008; Maier et al., 2024b) the greater investigation domain of this study is the European domain with boundaries of 16°W - 36°E and 32°N - 74°N. However, within this domain, the AoC is limited to the surroundings of the Δ^{14}C sampling locations of the ICOS measurement network. The ICOS network provides decent coverage across western Central Europe and some of Northern Europe (Tab. 2.1) but the network does not provide information about Southern and Eastern Europe (Fig. 2.4). In addition, the temporal data coverage resulting from the measurement frequency and length of the time series plays an important role in the inversion results, which will be discussed in a later chapter (Sect. 4.3).

To investigate the spatial coverage of the ICOS network and define a suitable AoC, several synthetic studies were conducted for each of the ΔffCO$_2$ proxies and evaluated using the previously defined ρ metric (Eq. 4.1). Figure 4.1 shows the spatial distribution of ρ for the temporally averaged posterior ΔffCO$_2$ fluxes in 2022. The year 2022 was chosen because it contains a homogeneous observation coverage for all three proxies compared to the other years. For the AoC tests, synthetic observation data were derived from a forward run using the GCP inventory. The data coverage in these synthetic AoC studies corresponds to the real data coverage to conduct the AoC studies as close to reality as possible. These

synthetic observations have then been inverted using the Flat Zero Prior (Sect. 3.2.4) with prior uncertainties of either 40% or 4% as input. It is worth reiterating here that the varying uncertainties of the prior are interpreted by the inversion algorithm as weighting that determines the relative influence of the prior information compared to the observational data. The posterior fluxes may differ from the prior by several times the specified prior uncertainty. To best match the synthetic data availability with the real measurement data, the AoC runs for different proxies use different station sets. The CO-based $\Delta ffCO_2$ approach uses the smallest set- CBW, GAT, HPB, HTM, KRE, LIN, OPE, STE (Tab. 2.1). Stations JFJ, NOR, OXK, PAL, and SVB have been added to the flask-based inversion. The integral-based run uses both of these sets, plus TRN, SSL, SAC and MHD stations. As to the reasons why not all flask collecting stations are used in the final CO-based inversions, see Sect. 6.2.3.

Figure 4.1 shows that the AoC is centered in those regions where the footprint of the ICOS stations overlaps. The truth recovery rate ρ decreases with increasing distance from the center of the ICOS observation network. For prior uncertainty of 40%, the AoC sensitivity tests for the three $\Delta ffCO_2$ proxies reach 100% truth recovery in the center of the ICOS network. With lower prior uncertainty, ρ decreases to values of about 50-60% in the center of the ICOS network, but the shape of the AoC remains similar. The general patterns for the three $\Delta ffCO_2$ proxies are similar, which is expected since all runs use the same core station set. The AoC is more limited in the case of the CO-based synthetic inversion, since it uses the smallest station set. The differences caused by the different time coverage of the different $\Delta ffCO_2$ proxies (i.e., flasks provide $\Delta ffCO_2$ values for one hour, integrals provide 2-weekly 24h means, and CO-based $\Delta ffCO_2$ stands for weekly afternoon hours) do not seem to influence synthetic AoC sensitivity tests. However, this may be different for inversions with real data, as by definition the sensitivity tests are not impacted by transport errors, and all types of sampling will lead to correct values without any bias. The result figures for other years can be found in the Appendix (Sect. A.1).

The AoC test runs with the prior uncertainty set to 40% give an overly optimistic picture of the potential recovery area, as there is no inherent bias in these synthetic runs. With the transport and measurement errors in the system, only regions with a high density of observations can be constrained with the necessary confidence. Therefore, I have chosen the center of the ICOS network with the boundaries of 5°E - 16°E and 47°N - 56°N as the AoC for this study (black square in Fig. 4.1) and it roughly coincides with the borders of Germany and Benelux. Within the AoC, all sensitivity studies achieve a high degree of truth recovery rate even in tests with a lower prior uncertainty of 4%. No reliable constraints can be made for areas outside the AoC, and all future investigations will be restricted to this AoC. This region will be referred to either as AoC or Germany+ domain.

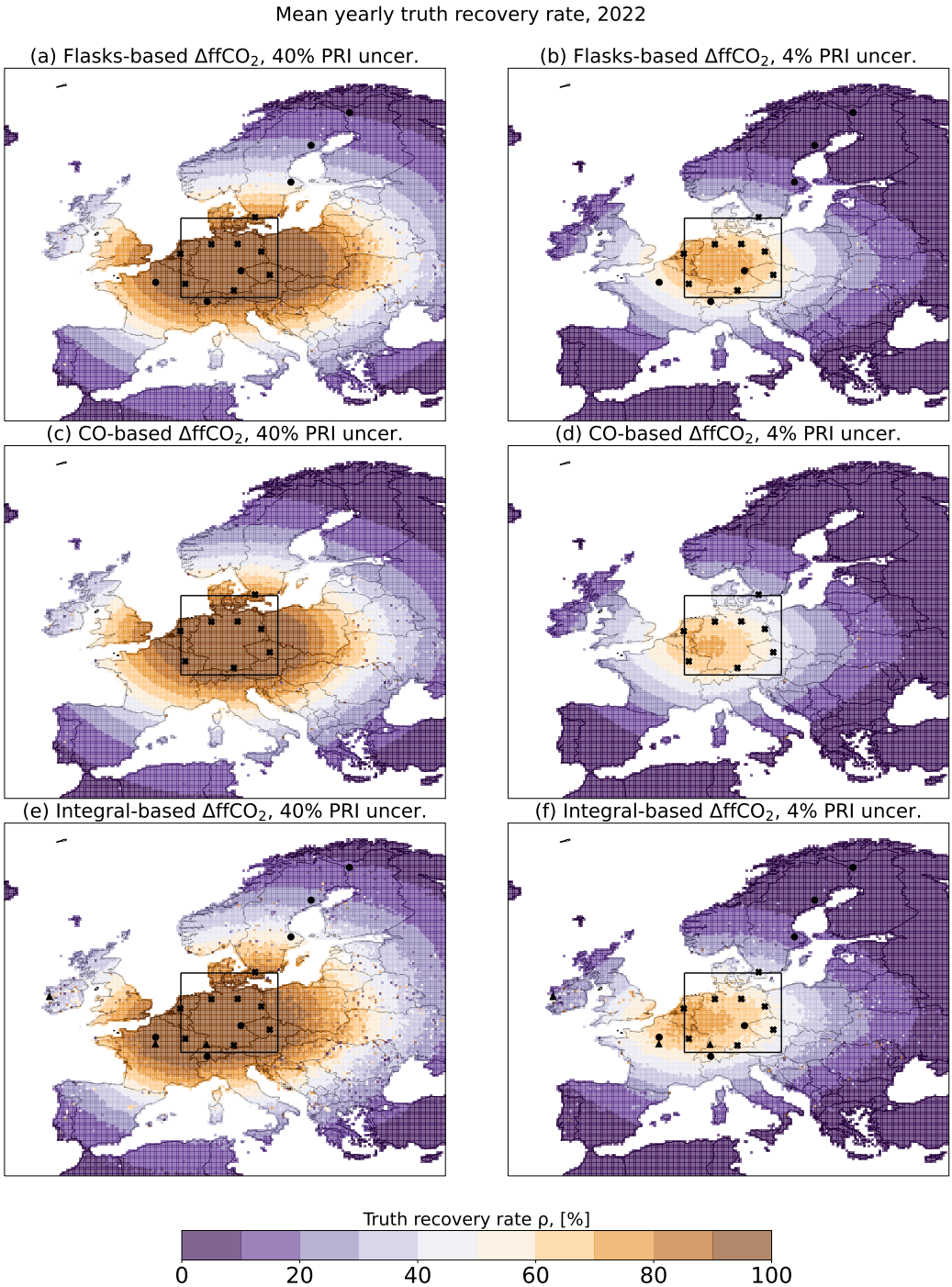

Figure 4.1. Temporally averaged truth recovery rate ρ maps of the AoC tests for the year 2022 for all $\Delta ffCO_2$ proxies: flask-based (panels (a) and (b)), CO-based (panels (c) and (d)), and integral-based (panels (e) and (f)). The Flat Zero Prior (Sec. 3.2.4) was used in all runs with different prescribed prior uncertainty (PRI uncer.): 40% (panels (a), (c) and (e)) and 4% (panels (b), (d) and (f)). The truth recovery rate ρ ranges from 0%, where the emissions were not changed in the posterior compared to the prior, to 100%, where the truth could be entirely recovered. Stations marked with X are used in all setups, circles - only in flask-based and integral-based runs, and triangles only in integral-based inversions. The region of maximum truth recovery (5°E - 16°E and 47°N - 56°N) is marked with a black square and covers mainly Germany and Benelux.

4.3 Analysis of the data density

The ICOS measurement network not only restricts the spatial domain of the reliable inversion results, but also restricts the temporal period of the investigation. For example, in 2019 there are only a few stations with flask samples (1 flasks per month on average in 3 stations in 2019, compared to 2 flasks monthly in 14 sites in 2022, Tab. 2.2), which is not enough to support the AoC defined in the previous chapter (Sect. 4.2). In addition, there may be time periods with different sampling frequencies, which would also lead to a varying data constraint. In this chapter, I investigate which time periods have the necessary data coverage to provide meaningful inversion results and which time periods cannot be analyzed because of insufficient data coverage.

First, the sampling periods for each of the three $\Delta ffCO_2$ proxies are discussed in more detail.

The monthly sampling frequency for flask samples is shown in Figure 4.2. In the years 2019 and 2020, only HPB, KIT, and SAC have flask measurements, with more stations starting flask sampling in 2021. Figure 4.2 clearly shows the highly inhomogeneous sampling pattern, with generally more samples available in winter than in summer. The year 2024 stands out with the highest sampling rates for the majority of the stations (approximately 1 flask every 3 days). These additional flasks were collected as part of the CORSO project (Levin et al., 2024). The JFJ, NOR, OXK, PAL, SAC and SVB stations do not have any $\Delta ffCO_2$ values in 2024 due to the missing nuclear correction (Sect. 3.1.3). As the STILT footprints normally used for the calculation of nuclear corrections were not yet available, the nuclear corrections for the stations included in the CORSO program were calculated using FLEXPART (Bakels et al., 2024).

The days per month with CO-based $\Delta ffCO_2$ values available at ICOS stations that were investigated in this study and passed the preliminary quality check (Sect. 6.2.3) cover nearly the entire investigation period starting at the end of 2019 (Fig. 4.3). The only significant exception is the delayed start of the data series from the CBW station, for which data are only available starting from the end of 2021.

Figure 4.4 shows that the collection of integrated $\Delta^{14}C$ samples started earlier than the collection of the flask samples. However, because of the lack of the nuclear correction (Sect. 3.1.3) needed for the calculation of $\Delta ffCO_2$, no $\Delta ffCO_2$ values are available for 2014 and 2024. Two major increases in the number of stations with available $\Delta ffCO_2$ values can be observed: one around 2018 when sampling started at KRE, PAL and SAC stations, and the other in 2021 when GAT, LIN and OXK stations were added to the network. Due to the sampling nature of the integrated $\Delta^{14}C$ measurements, the stations with integral-based $\Delta ffCO_2$ values probe all atmospheric situations equally during the sampling period, result-

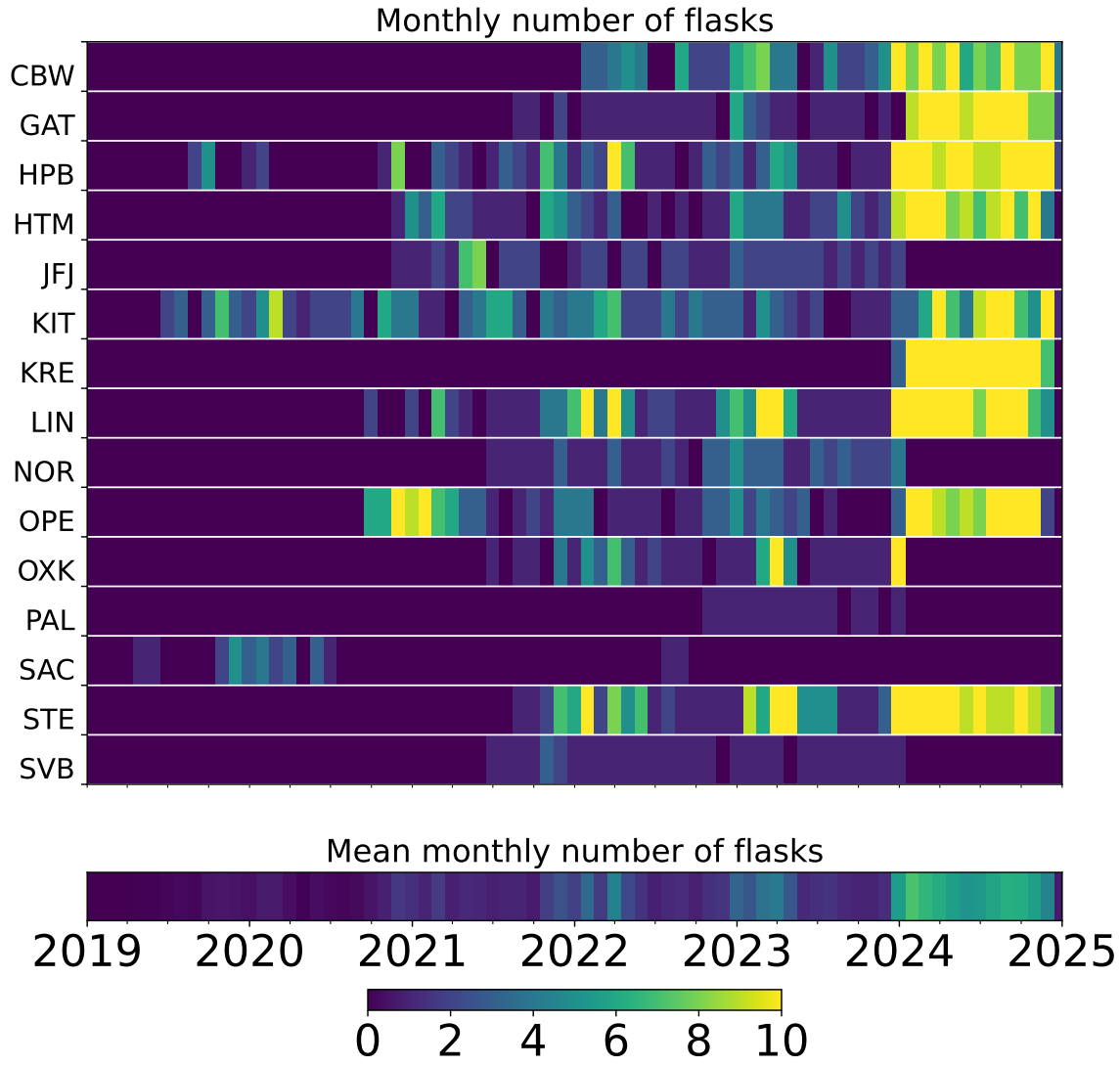

Figure 4.2. Monthly number of flask samples collected at the ICOS stations with available ΔffCO$_2$ values in the time period covered by this study. The higher number of observations in 2024 is attributed to the CORSO project (Levin et al., 2024). The bar at the bottom indicates the network-wide mean monthly availability of flask-based estimates per station.

ing in a homogeneous temporal coverage, but with only one value for the entire measurement period. Thus, stations with an integrated sample collection implemented have almost complete coverage until 2021. In the following time period, more stations were added to the network, but on average Δ^{14}C data are available only for every second sample due to a limit in the capacity for integrated sample analysis (these ”missing” samples will be measured as soon as the additional analysis capacity becomes available). This variation in the availability of the ΔffCO$_2$ data can lead to unexpected changes in the data constraint of the inversion system. This illustrates the need for further analysis of the data constraint using the synthetic inversions.

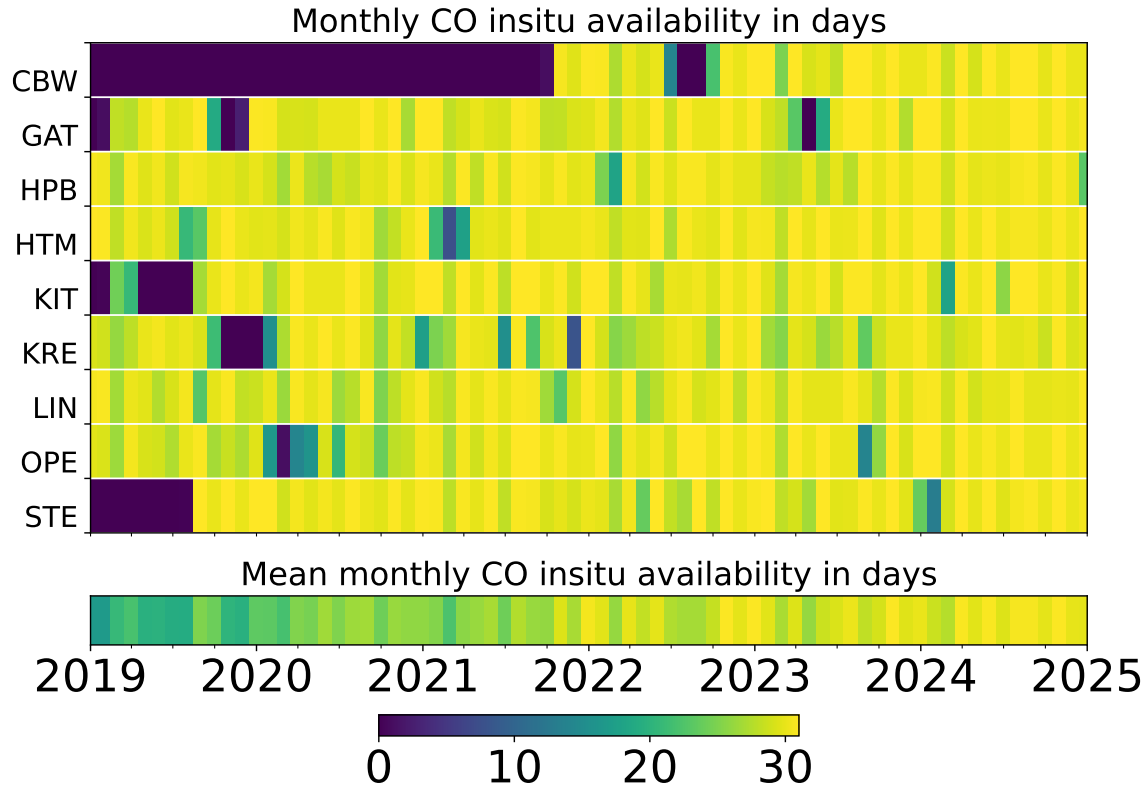

Figure 4.3. Monthly CO-based $\Delta ffCO_2$ availability in days at the ICOS stations in the time period covered by this study. The bar at the bottom indicates the network-wide mean monthly availability of CO-based estimates per station.

In the following, a series of synthetic runs were performed to determine a coherent time period with sufficient data coverage in the ICOS network for each $\Delta ffCO_2$ proxy to reliably constrain the AoC (Sect. 4.2). Here, the GCP inventory was used as known truth and the Flat Zero Prior (Tab. 3.1) as input. I can use the synthetic run test for this purpose, as the synthetic data was constructed to have the same temporal coverage as the real data for the corresponding proxy (Sect. 4.1). The standard station set was used for each proxy as defined in Table 2.2. The posterior fluxes were spatially aggregated to the chosen AoC (Sect. 4.2) and the ρ metric (Eq. 4.1) was applied. Figure 4.5 shows the time series of the truth recovery rate ρ in the upper part with the mean monthly data availability presented as a heat map in the lower part for the three $\Delta ffCO_2$ proxies in separate panels. For each proxy, the results were presented with different prior uncertainties, i.e. different weightings between prior and data constraint. The application of different prior uncertainties aims to discern various aspects of the influence of data coverage on ρ, however, they should contain the same information. For the case of 40% prior uncertainty, I expect a high level of ρ even with very limited data. A recovery rate lower than 100% will indicate a severe lack of data. The cases of 10% and 4% prior uncertainty would show more structure in ρ, allowing one to distinguish between

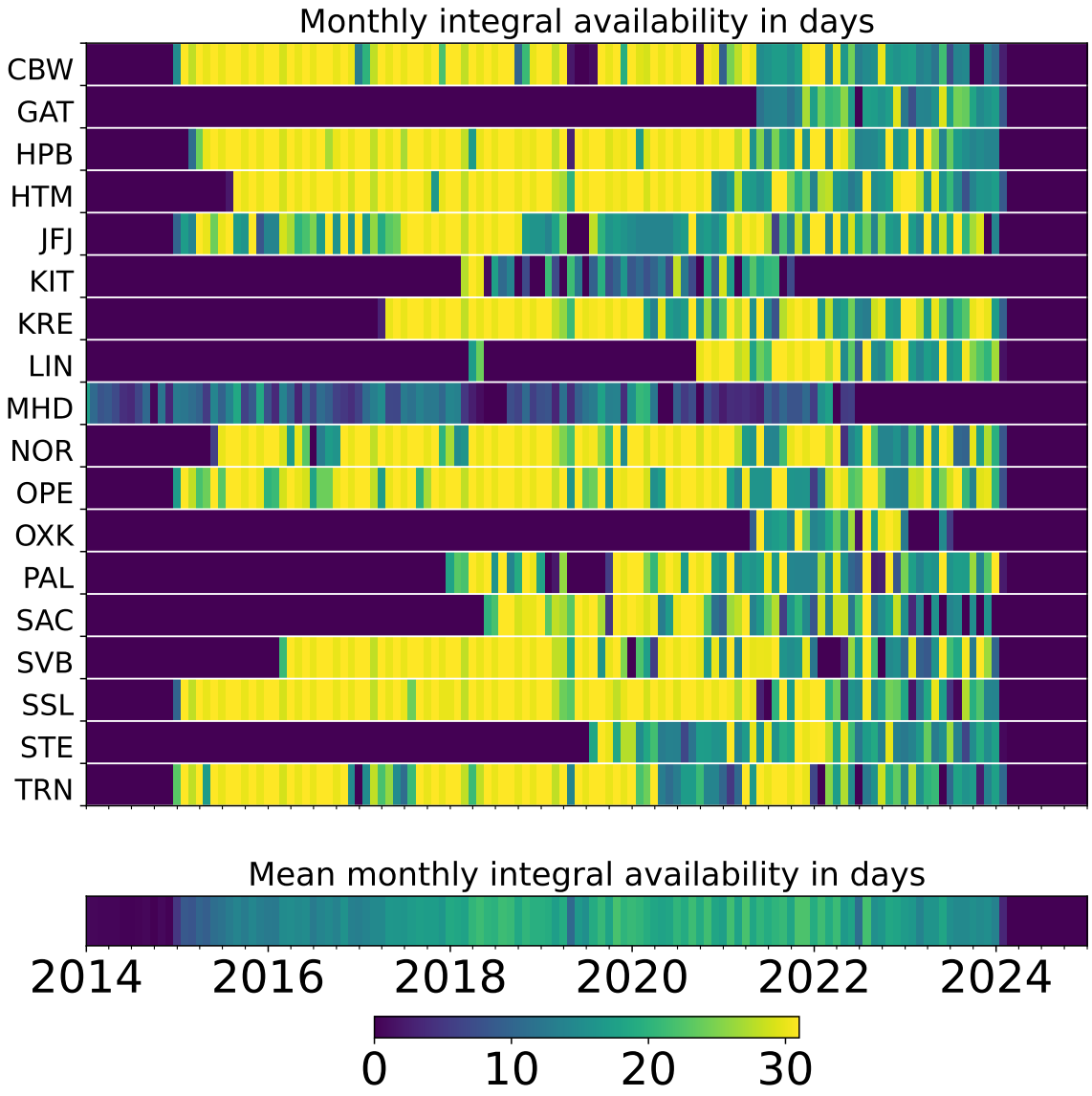

Figure 4.4. Monthly $\Delta ffCO_2$ availability based on the integral $\Delta^{14}C$ samples at the ICOS stations in the time period covered by this study. The absence of observations in 2014 and 2024 can be attributed to the missing nuclear corrections needed for the calculation of $\Delta ffCO_2$. The bar at the bottom indicates the network-wide mean monthly availability of integral-based estimates per station.

periods with different data constraints. I do not expect the corresponding posteriors to reach the recovery rate of 100%, as a deliberately wrongful prior with low uncertainty will exert a strong influence on the posterior. Regarding the prior uncertainty of only 1%, periods of higher ρ would point to the time periods of best data coverage for the given proxy.

As the Flat Zero Prior was used, the inversion model does not have prior information about seasonality. To allow the inversion to make changes in the temporal domain, a time correlation filter of F2T was applied to the inversion, which allows for the resolution of the

seasonal cycle but none of the shorter frequencies (Sect. 3.2.2.2). More on the appropriate time filters can be found in Section 4.4. This is the reason for the short-term variability in Figure 4.5, where the posterior with the highest prior uncertainty of 40% oscillates around the recovery rate of 100% with an amplitude of approximately 5%, slightly but consistently overestimating the truth due to the application of the Flat Zero Prior (see Sect. 4.6.3).

Figure 4.5, panel (a), depicts the temporal development of the truth recovery rate ρ for synthetic flask-based $\Delta ffCO_2$ data. The posterior of the inversion run with a 40% prior uncertainty achieves full truth recovery ($\sim$100%) for the entire period except for the year 2019. It shows that in the absence of any transport or measurement errors and with low restriction by the prior (i.e., prior uncertainty of 40%) the inversion system needs only one or two data points per month per station to reconstruct the truth. The effect of the number of samples on ρ is most clearly visible for the prior uncertainty of 4%, where the increase in the number of flask samples in winter 21/22 to 3-5 flasks per month and station leads to a significant increase in the recovery rate from an average of 25% up to 50% and above. The winters of 2022 and 2023 have slightly higher ρ values compared to the corresponding summers due to the respectively lower sampling rate in the summers. 2024, the year of the CORSO project (Levin et al., 2024) and thus the year of a significantly increased number of flask samples, shows another maximum in the truth recovery rate, only to fall again at the end of 2024, due to the decreasing number of flasks, as flask analysis has not yet progressed so far and the boundary effect as no data for 2025 is available yet. Based on these findings, I conclude that the higher sampling frequency improves the inversion results and that the inversion system has a sufficient level of flask-based data availability starting from November 2021.

The truth recovery rate of CO-based synthetic inversions (Fig. 4.5, panel (b)) for the prior uncertainty of 4% shows consistently high values throughout the entire analysis period due to the steady availability of CO data. All ρ values show an increase in the second half of 2021 when the CBW station started CO measurements. Comparing the absolute values of the recovery rate for posteriors with the prior uncertainty of 4% between the flask- and CO-based inversions, I can see that flask-based results have on average higher ρ values due to the larger number of measurement sites included in the inversion. Nevertheless, for CO-based $\Delta ffCO_2$ inversions, the time frame from January 2019 to the end of 2025 will be investigated in the further examinations, because the increase in 2021 is not as significant as for the flask-based proxy.

As with other proxies, synthetic inversion based on integral samples with prior uncertainty of 40% rapidly reaches full truth recovery as long as there are at least few data points available (Fig. 4.5, panel (c)). In this case, ρ is already maintained at about 100% in the second quarter of 2015 and stays at this level until the drop in 2024 caused by missing analysis results of samples. At lower prior uncertainties, one can see a structure in the recovery rate

line that correlates with the data coverage. From the second half of 2015 to the beginning of 2018 the ρ metric fluctuates around the average of 40% in the case of 4% prior uncertainty. In the second half of 2018, as more stations come online, the next plateau of truth recovery rate is reached at values of approximately 60% until they drop in the second half of 2023 due to the absence of data. In this case, it is not clear which time period has sufficient data coverage for the inversion system. The segment from July 2018 to July 2023 clearly has greater data availability and should be considered as the main analysis period. Nevertheless, the period from July 2015 to July 2018 should not be discarded, as it still may possess sufficient truth recovery and may provide useful information about the inventory; however, it should be treated with more care.

4.4 Optimization of temporal correlation length

An essential model parameter that influences the posterior fluxes is the level of temporal flexibility allowed in the inversion system - i.e., the degree of smoothing applied to the fluxes over time. In the synthetic runs, any short-term variations in the flux field can theoretically be resolved by the inversion if the data density is sufficient and if a high temporal flexibility of the posterior flux is permitted. However, in reality, measurement and transport errors within the modeling system set limits. In the CarboScope modeling framework, the temporal flexibility of the posterior is determined by the characteristic frequency of the filter used in the Fourier space to construct the temporal variability of the posterior (Sect. 3.2.2.2). With real data, high degrees of temporal freedom can, however, lead to overfitting of the observations, causing additional variability of the posterior fluxes, which the inversion generates by optimizing signals that are below measurement uncertainties. Conversely, too strict regularization in the time domain - for instance, allowing no changes to the temporal trend compared to the prior - will result in suboptimal utilization of the data's information content, skewing the Bayesian inversion towards the prior. For these reasons, this parameter is essential for producing a robust and informative posterior.

This sensitivity experiment aims to determine the optimal degree of temporal regularization of the posterior fluxes under varying levels of measurement uncertainty. To this end, synthetic pseudo-datasets were constructed by adding random, normally distributed errors to the otherwise error-free synthetic pseudo observations used in the previous section. Three standard deviations - 1.5 ppm, 3 ppm and 4.5 ppm - were selected to represent the uncertainty ranges of the $\Delta ffCO_2$ proxies considered in this study: $\Delta^{14}C$ based (Sect. 3.1), CO based (Sect. 6.3.1), and a high-end uncertainty scenario, respectively. For each error scenario, three independently perturbed datasets were generated, yielding a total of nine synthetic datasets.

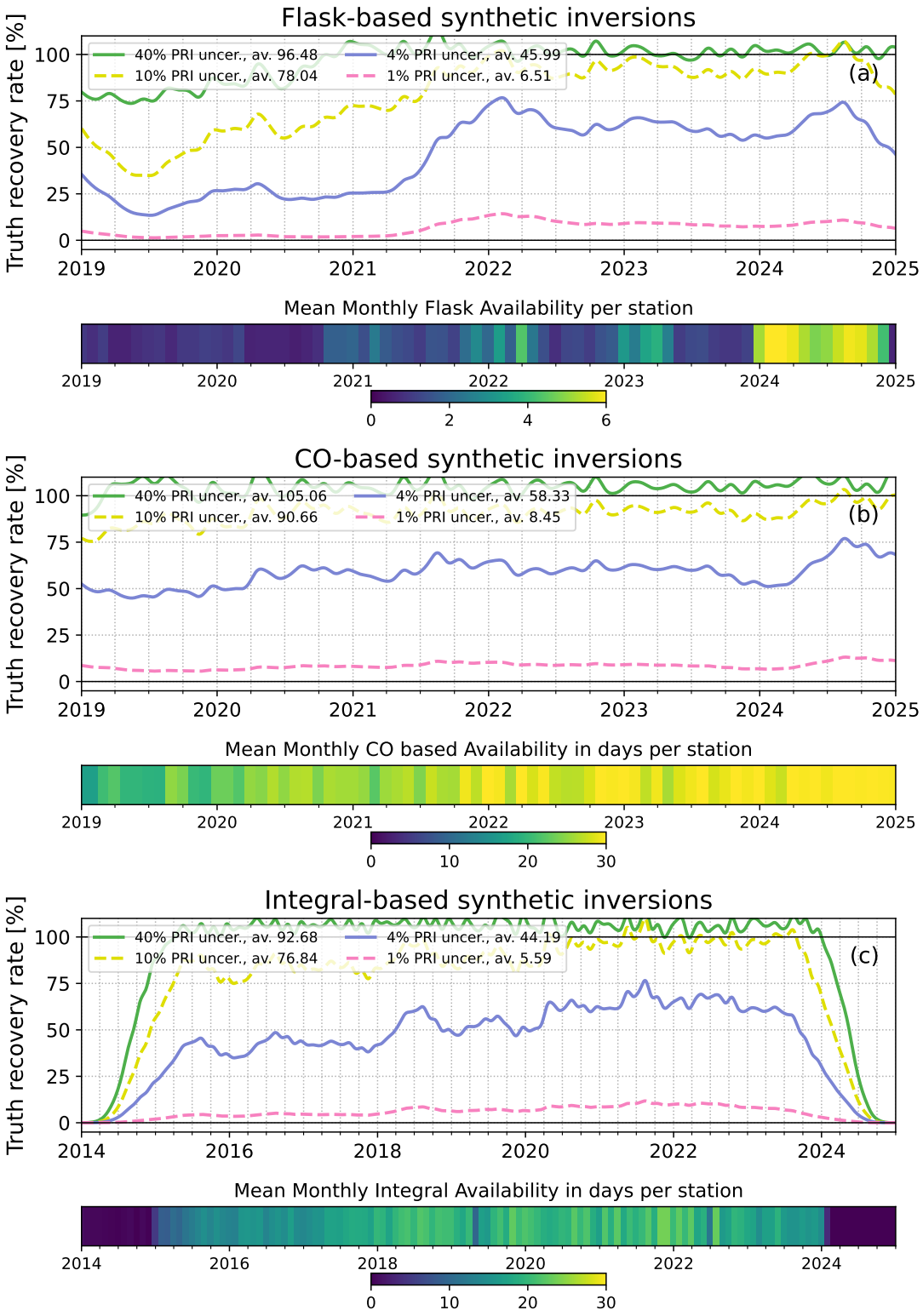

Figure 4.5. Truth recovery rate ρ in the AoC for the whole time period of the conducted study for three $\Delta ffCO_2$ proxies: flask-based $\Delta ffCO_2$ (panel (a)), CO-based $\Delta ffCO_2$ (panel (b)), and integral-based $\Delta ffCO_2$ (panel (c)). The runs with different prior uncertainties are represented by colored continuous and dashed lines in each plot. The mean recovery rate over the whole time period for each line is noted in the legend (av.). The color bars in the lower part of each panel denote mean monthly data availability for each proxy (for station-wise data availability see Figures 4.2 - 4.4).

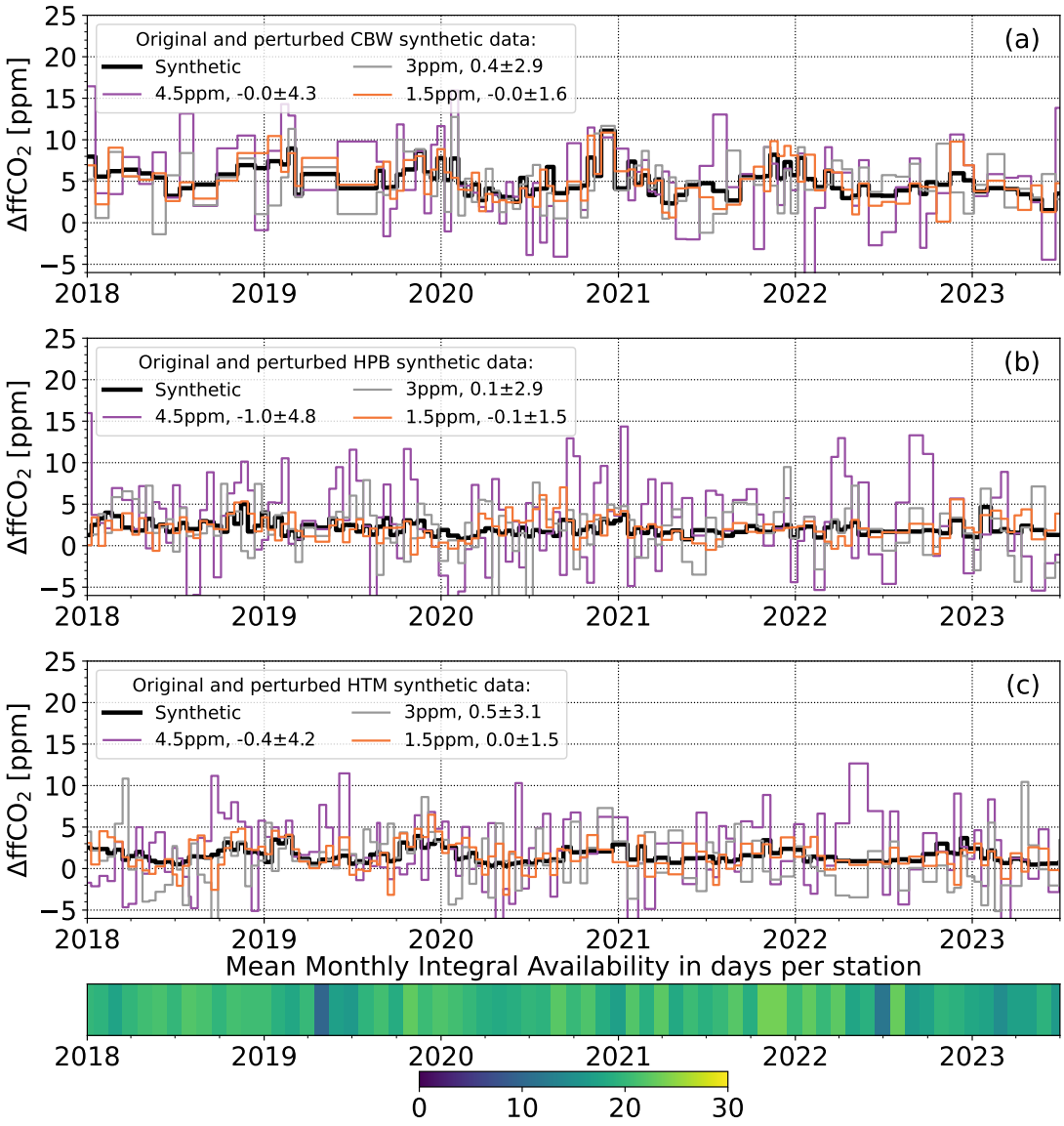

Figure 4.6. Original and perturbed synthetic data from three example sites: CBW (a), HPB (b), and HTM (c). Original synthetic values from a GCP inventory FWD run mimicking the integral sampling intervals (black line) are shown with perturbed synthetic data at 4.5 ppm (purple), 3 ppm (gray), and 1.5 ppm (orange). The legend provides the actual mean synthetic-perturbed mismatch and standard deviation in ppm for each series. The color bar at the bottom indicates the network-wide mean monthly availability of integral-based estimates per station.

Because the perturbations were generated randomly, each pseudo-dataset differs. Although an ideal approach would involve a large-scale Monte Carlo analysis with thousands of datasets, this was not feasible due to computational limitations. The analysis focused on comparing posterior results derived from the same pseudo-observations but with varying temporal regularization of the inversion system. Also due to computational limits, only the

integral-based synthetic proxy was analyzed, with the assumption that results for the other proxies would yield similar conclusions due to the inherent absence of the transport errors in these studies.

Figure 4.6 displays the original and perturbed synthetic data for the CBW, HPB and HTM stations, using one representative pseudo-dataset from each of the three perturbation scenarios. As expected for purely statistical noise, the mean bias relative to the original signal remains close to zero. The standard deviations of the deviations align well with the nominal values of the respective perturbations, confirming the intended error characteristics.

The perturbed synthetic data sets were used in an inversion using the Flat Zero Prior (Tab. 3.1 and a 40% prior uncertainty, ensuring a high weight on the data compared to the prior information. Ten temporal correlation lengths ranging from F24T (cutoff frequency of 24/yr, approx. biweekly correlation length) to F0.2T (cutoff frequency of 0.2/yr, about 5 year correlation length) and GLT/Global (one global degree of temporal freedom) were tested, resulting in a total of 30 inversions.

Figure 4.7 shows the resulting posterior $\Delta ffCO_2$ fluxes for the representative perturbed data sets depicted in Figure 4.6. In each panel, only five of the ten temporal correlation lengths tested are shown. As expected, the posteriors derived from datasets with lower noise levels reproduce the truth (GCP) more accurately (panel (c)), while those based on the high-perturbation scenarios exhibit substantial deviations from the truth (panel (a)).

A closer inspection of the F12T posteriors (green dashed lines in Figure 4.7) reveals considerable variability at the annual scale, leading to unrealistic emission patterns—even in the low-uncertainty case (panel (c)). This behavior clearly indicates overfitting to noise. Reducing the frequency content via lower harmonic filters leads to smoother posteriors, as each inferred value incorporates information from a broader temporal window. In the high-noise cases (panels (a) and (b)), the overall shape of the posterior remains similar to the truth but is consistently biased toward higher emissions—a direct result of the random distribution of positive deviations in these specific pseudo-observations. The GLT filter (brown dashed line), by contrast, yields a posterior equivalent to a long-term average, since each observation contributes uniformly across the entire analysis period.

The posterior fluxes of these runs were then compared to the truth fluxes via the root mean square deviation (RMSD). Only the period of the highest data constraint for the integral-based samples (01.07.2018 - 01.11.2023) was considered for the calculation of the RMSD's. It is assumed that an insufficient number of degrees of freedom will result in a high RMSD, as the posterior cannot adequately reflect the information content of the data. As temporal flexibility increases, the RMSD is expected to decrease, since the posterior can increasingly align with the underlying truth. However, beyond a certain point, the RMSD will rise again due to overfitting of the noise in the error-prone data. Identifying the minimum

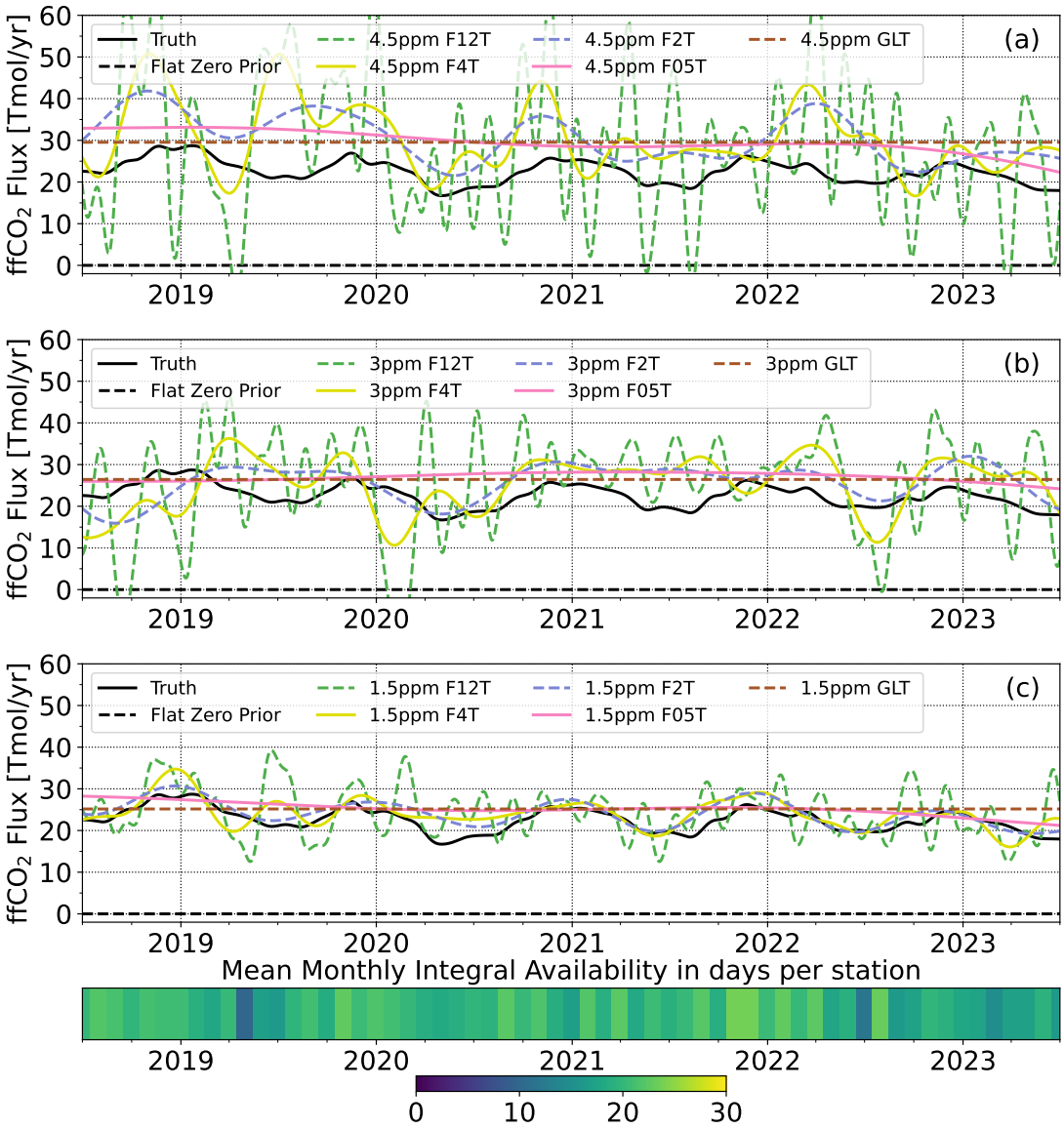

Figure 4.7. Analysis of different time correlation filters in dependence of the perturbation strength (Fig. 4.6). The black solid and dashed lines are respectively the truth (GCP) and the Flat Zero Prior that was used as input. The colored lines denote the posteriors that were generated using the same perturbed data set and prior uncertainty of 40%, but different time correlation filters, ranging from GLT (one global degree of temporal freedom) to F12T (cutoff frequency of 12/yr, Sect. 3.2.2.2).The color bar at the bottom indicates the network-wide mean monthly availability of integral-based estimates per station.

in the RMSD curve thus reveals the optimal time correlation length—corresponding to a characteristic frequency filtering in the Fourier domain.

The RMSD's between the true values and the posteriors with F24T filter (Fig. 4.8) reach the highest levels among all tested cases, as the variability is mainly caused by the added pseudo-variability of the synthetic data. As the degrees of freedom decrease, the RMSD

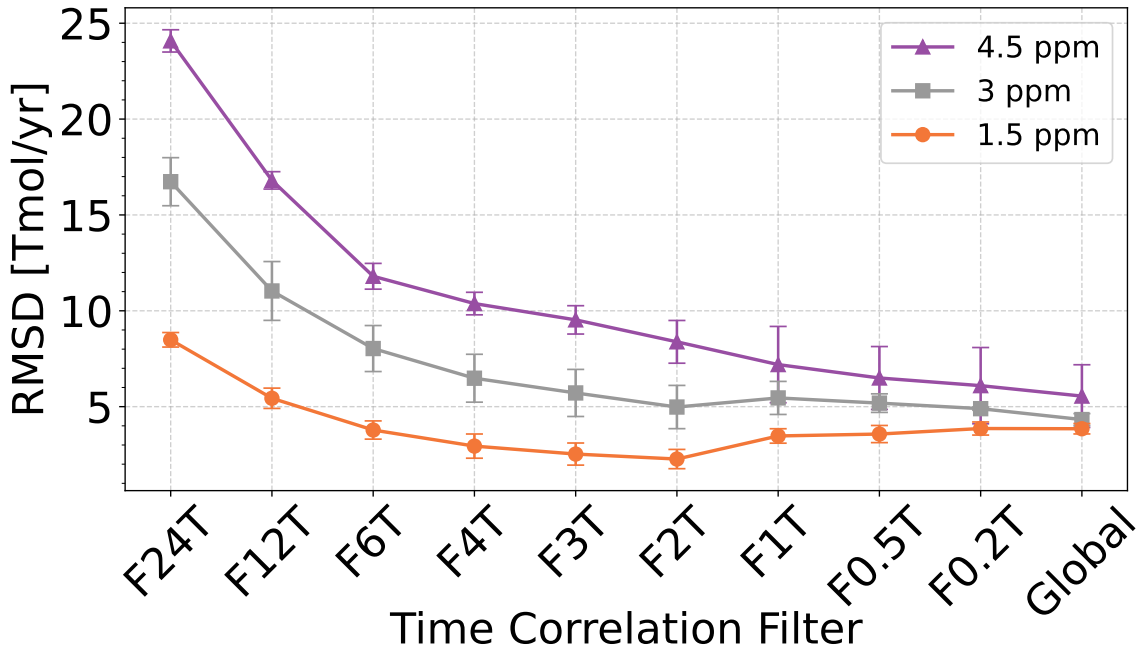

Figure 4.8. Root Mean Square Deviation (RMSD) between the truth and the respective posteriors over the time period with the highest data constraint (01.01.2018 - 01.07.2023). Each data point was calculated as the mean of three RMSD's, which were based on posteriors done with the same time correlation filter and data with the same nominal perturbation but different de facto biases due to the randomized nature of the pseudo-data creation. Errors are the standard deviation of the three corresponding RMSD's. The colored lines denote the three different random, normally distributed perturbations.

also decreases. For the case with 4.5 ppm random, normally distributed perturbation, the lowest RMSD is achieved with the Global filter with no flexibility at all. The RMSD for the 3 ppm perturbation reaches a local minimum at F2T, then rises, and finally decreases again, with the Global filter remaining the absolute minimum. Only the RMSD curve for the 1.5 ppm perturbation follows the expected pattern. It initially decreases due to reduced overfitting of the error-prone data, then reaches the absolute minimum at F2T, reflecting the seasonal oscillation of the true data and ignoring any short-term variability. Ultimately, it increases again because of the loss of contained information in the data by the posterior (missing seasonality).

High uncertainty in the data suggests the need to reduce the flexibility of the posterior. For observations with high error, longer time-period averaging is necessary to prevent overfitting, as the random error dominates the higher frequency information. Conversely, reducing the random errors allows for the extraction of short-term information. Based on this synthetic study, the optimal range of the temporal degrees of freedom for the 4.5 ppm case is on the scale of inter-annual variability (i.e., F0.5T or lower in Fig. 4.8), with no

possibility of resolving any seasonality. The set with 3 ppm nominal perturbation already allows for the investigation of seasonality. The 1.5 ppm case may be suitable for examining short-term variability up to seasonal resolution (F4T to F2T in Fig. 4.8).

However in reality the mismatch between model and actual atmospheric transport could lead to additional uncertainty. This will reduce the effective information content of the data and suggest further regularization of the inversion. Nevertheless, the seasonal signal may be strong enough to be elevated from the noise and is an important parameter of the climate system. Therefore, I will apply F2T time correlation filter in following inversions if not stated otherwise. In time-space the F2T filter corresponds to the temporal correlation of 6 months. This will allow for the investigation of the seasonal patterns and minimize the overfitting caused by the errors. Unfortunately, it will prohibit investigations of any variability on the shorter time scales.

4.5 Used prior uncertainty

Before analyzing inversions using actual atmospheric measurements as basis, one has to define the range of the tolerable prior uncertainties. Here I want to differentiate between two cases with different aims. First, the case where the information content of the measurements should be extracted to the full extent minimizing influence of the prior information. This is done with the aim to achieve a completely data driven inversion. Usually in such inversions a strongly modified prior is used to investigate a specific facet of the inversion system or data properties. Second, the case where beside the measurements, the weight of the prior information is used as attributed to it in literature, to achieve a posterior corresponding to optimal prior and data inputs.

In terms of synthetic inversion to achieve a completely data driven posterior, the inversion system should be able to reconstruct the known truth with very little or even no information from the prior. As it was shown in figure 4.5 that this can be already accomplished with the prior uncertainty of 40%, virtually independent of the variability in the data density. Theoretically, one could increase the uncertainty further to achieve even greater independence from the prior and it would work with the synthetic data, having no transport or measurement induced bias between the sampling sites. However in Bayesian terms this would be equal to decreasing the measurement errors, and hence forcing the inversion system to reduce the model-data-mismatch even further. Because of this, with the real measurements there is again a risk that the inversion system will start to overfit the error prone measurement data. Thus I conclude that 40% prior uncertainty provides enough flexibility to the inversion system for the cases where the posterior should be driven by the data.

If the goal of the specific analysis is to find the optimal posterior considering best es-

timated uncertainties of unmodified prior emissions and observational data, then I need to refer to the published values for prior uncertainties. As already discussed in the section 3.2.3 the nominal uncertainty of the total fossil CO_2 emissions for the Germany and surroundings is 1% as reported by Super et al., 2020, whereas the uncertainty for this area reported for the emission inventories used in this study is 5% for the GCP (Jones et al., 2022) and 2.5% for EDGAR (Janssens-Maenhout et al., 2019). However during the inversion procedure the total emission is disaggregated in the temporal and spatial domains with uncertainty for each grid cell rising to 20-75% (Super et al., 2020; Jones et al., 2022). In this study I do not analyze the emissions grid-wise, but rather investigate the AoC as a whole. But all model calculations are performed on gird cell level, what may lead to an increase in model-data-mismatch due to the discrepancies in the emissions between inventory and reality in the local catchment areas of the measurement sites. To account for this, I will use 4% prior uncertainty as conservative estimation in the inversions with unmodified priors (Sect. 5.2.4, 6.3.4 and 7.2.4), which is within reasonable error margins for emission inventories and is even lower than discrepancies in cross-inventory comparison (Sect. 3.2.3). For other inversions, to account for the disaggregation errors and to achieve data-driven posteriors, 40% prior uncertainty will be applied.

4.6 Application of the Evaluation Matrix on the perturbed synthetic data

As previously described in Section 3.2.4, the Evaluation Matrix is a set of experiments developed to investigate key features of the posterior fluxes estimated by the inversion framework. This approach will be applied to all real data inversions for the different $\Delta ffCO_2$ proxies (see Sections 5.2.2, 6.3.2 and 7.2.2). However, because real data inversions lack an objective emission truth, an evaluation measure is needed to determine the significance of the differences between the real data Evaluation Matrix posteriors and the GCP inventory used in the Evaluation Matrix. In this chapter, I therefore apply the Evaluation Matrix to posterior fluxes derived from perturbed synthetic integral data constructed from a known truth, i.e. the GCP inventory. To simulate the influence of measurement uncertainties of different magnitudes, the same perturbed synthetic data sets were used as in Section 4.4 (Fig. 4.6). However, to increase the statistical significance of the resulting posterior variability, the number of pseudo-data sets for the random normally distributed perturbation of 1.5 ppm that represents the expected uncertainty of the $\Delta^{14}C$-based $\Delta ffCO_2$ data was increased from n = 3 to n = 30. These 30 pseudo-data sets were used to derive the estimated variability of the Evaluation Matrix results but are not shown in the figures. The number of pseudo-data

sets for nominal perturbations of 3 ppm and 4.5 ppm was not increased to reduce the computational load on the inversion system. These results thus provide an indication of how the Evaluation Matrix parameters change with increasing random uncertainty of the measurements. Furthermore, assuming that for synthetic posteriors with no inherent transport model errors, perturbations have a greater effect than sampling routine differences, this synthetic study was conducted solely with integral data to further reduce the computational load. The results are presumed applicable to other $\Delta ffCO_2$ proxies during periods of high data constraint.

4.6.1 Seasonal cycle of $\Delta ffCO_2$ emission trend derived from synthetic inversions

Figure 4.9 shows the absolute posterior $\Delta ffCO_2$ fluxes resulting from the application of the Evaluation Matrix seasonal cycle investigation on perturbed synthetic data in the AoC (see Sect. 3.2.4). For this investigations, I used the Seasonless Prior (Fig. 3.6) with 40% prior uncertainty. Similarly to the analysis of the time-correlation filters, the posteriors calculated using the perturbed synthetic data with a nominal error of 4.5 ppm (panel (a)) deviate strongly from the truth. For all three experiments performed, both amplified amplitudes and shifted maxima and minima positions of the seasonal cycle can be observed. As the magnitude of the perturbation decreases to 3 ppm (panel (b)) and 1.5 ppm (panel (c)), the visible deviation of the seasonal cycle's phase and amplitude from the truth decreases as well, especially in the period from mid-2018 to mid-2023 corresponding to the period with the highest data density (Sect. 4.3).

To extract and visualize the amplitude and phase of the seasonal cycle, the posteriors and the truth were divided by the Seasonless Prior. The resulting relative seasonal cycles are depicted in the Figure 4.10. The position of the maxima and minima of the posteriors are marked with orange and purple circles. The corresponding extrema of the GCP truth are marked by diamonds. In some years, the shape of the seasonal cycle was distorted to such a degree that not all extrema corresponding to the truth could be identified in the posteriors. This is particularly evident in the posteriors with nominal perturbation of 4.5 ppm (panel (b)), but also present in some instances of the posteriors with nominal perturbation of 3 ppm (e.g. the absent 19/20 winter maximum for the "3 ppm ver 1" line in panel (c)). All corresponding extrema could be identified in the three depicted posteriors with a nominal perturbation of 1.5 ppm. However, seven of thirty calculated posteriors had missing at least one extremum (not shown).

The summer and winter extrema were analyzed separately. Due to lower fossil emissions in summer but the same nominal perturbations applied to the synthetic data, I expect summer

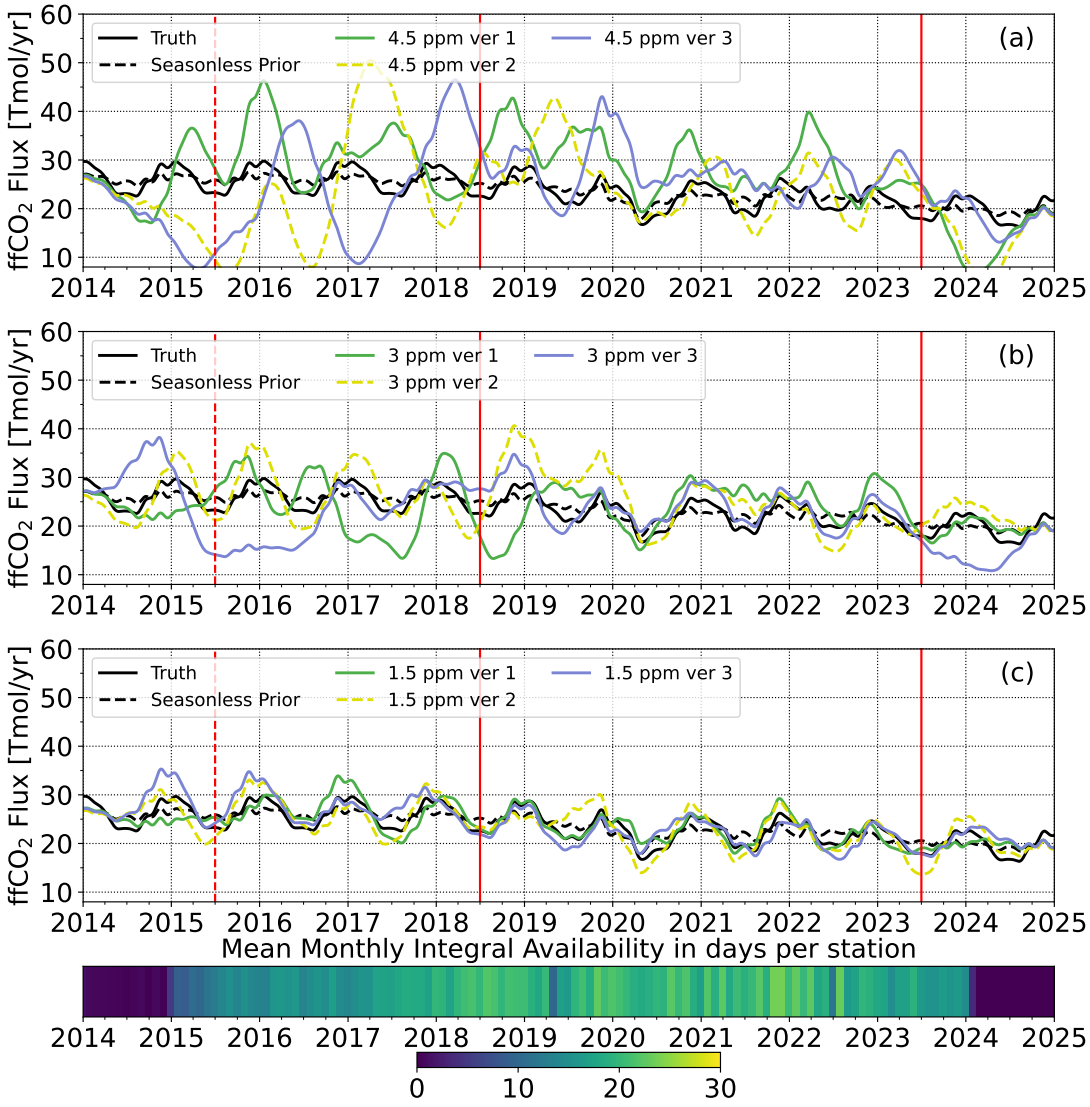

Figure 4.9. Analysis of the seasonality derived from perturbed synthetic data (Fig. 4.6). Panels (a-c) show absolute $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Seasonless Prior (black dashed line) that was used as input. The three colored lines represent posteriors with a 40% prior uncertainty, derived from synthetic datasets with the same nominal perturbation. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

fluxes to be more influenced by the perturbation due to the lower signal to noise ratio (SNR). Similar effect is observed in the real measurements, with comparable measurements errors during the whole year but lower $\Delta ffCO_2$ signals in summers.

Table 4.1 and 4.2 display the positions of winter maxima and summer minima for the

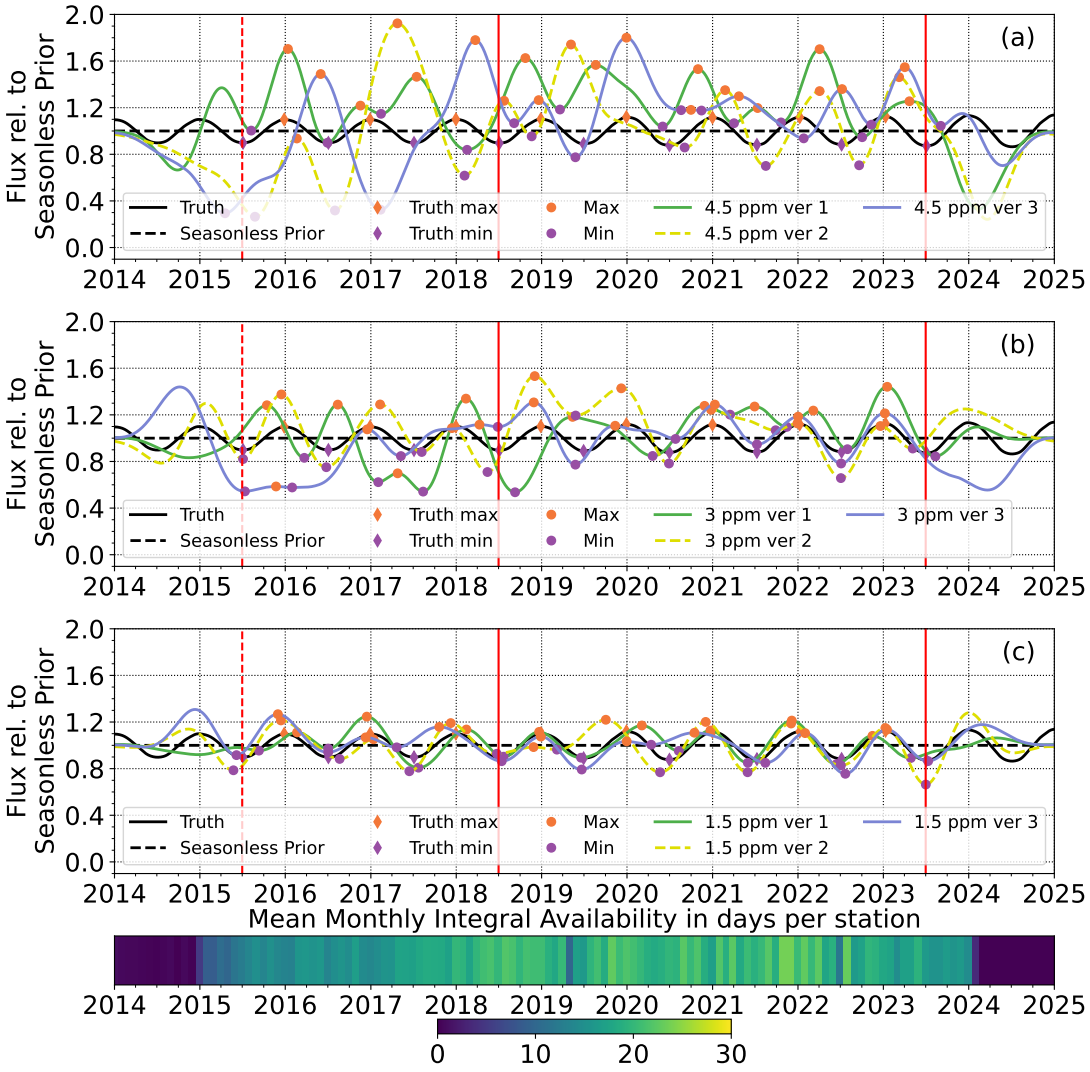

Figure 4.10. Analysis of the seasonality derived from perturbed synthetic data (Fig. 4.6) relative to the Seasonless Prior. Panels (a-c) show relative $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Seasonless Prior (black dashed line) that was used as input. The three colored lines represent posteriors with a 40% prior uncertainty, derived from synthetic datasets with the same nominal perturbation. Orange and purple circles mark the maxima and minima of the posteriors, with corresponding truth extrema marked by diamonds. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

truth and corresponding mean positions for the posteriors averaged over available inversions with the same nominal perturbations ($\overline{\text{Date}}$ columns). Additionally, the columns Δ and σ show the mean difference between means and the truth as well as standard deviation of the

differences. The first three rows correspond to the period with lower data constrain and were stated for the completeness. The measure for the significance of the deviation from the truth will be derived from the period between 01.07.2018 and 01.07.2023. The statistics of the last row was calculated by averaging over all maxima or minima across all extrema positions of the given random perturbation for this period. This was done to further increase the statistical significance of Δ and σ with the assumption that these deviations were not correlated but caused by the same effect.

For the perturbation scenario with a nominal standard deviation of 1.5 ppm, based on 30 independent realizations, the mean deviation in the timing of seasonal extrema relative to the true fluxes is already very close to zero. This confirms that random noise largely averages out when a sufficient number of realizations is used. Furthermore, the associated standard deviation of the extrema positions lies between 35 and 41 days, providing a robust first-order estimate of the uncertainty in seasonal timing introduced by observational noise at this error level.

In contrast, for the 3.0 ppm and 4.5 ppm perturbation scenarios, only three realizations were performed. Consequently, the resulting mean deviations and standard deviations are not statistically significant and should be interpreted purely as indicative figures. Nonetheless, the results qualitatively align with expectations: both the mean offset and the standard deviation of the extrema positions increase with higher levels of observational noise.

As already observed in the Fig. 4.10, higher nominal perturbation usually led to higher Δ and σ values. Even for the 1.5 ppm posteriors, the variance of the position around the truth is usually higher than one month. Furthermore, as expected, the variability of the summer minima (Table 4.2) were slightly higher then σ of the winter maxima for all three nominal perturbations. This supports the assumption that summer fluxes are more severely influenced by the random uncertainty of the data in comparison to the winter fluxes.

Table 4.3 and 4.4 contain the statistical evaluation of the posterior amplitudes for winter and summer relative to the Seasonless Prior. They have similar structures as Table 4.1 and 4.2. The relative mean amplitude of the runs with the same nominal perturbation normalized with the Seasonless prior are displayed in the Mean columns, with Δ and σ depicting the mean difference to the truth, i.e. the GCP inventory, and the standard deviation of the differences, respectively. The last row contains the statistics over the high data density period 2018-2023. Note that, the amplitude of the GCP seasonal cycle relative to the Seasonless Prior is increasing over time, from 0.90 in summer 2015 to 0.87 in summer 2023 and from 1.10 in winter 15/16 to 1.13 in winter 22/23. This is not caused by an increase in the absolute amplitude of the seasonal cycle of the inventory but by the overall decreasing fossil fluxes over the investigated time period (see Figure 3.6). As this effect is small, I ignored its influence on the Δ and σ values in this tests.

Table 4.1. Annual positions of flux winter maxima for the GCP (Date) and corresponding mean posterior positions ($\overline{\text{Date}}$) from synthetic data with same nominal perturbation. Δ and σ represent the mean difference and standard deviation of differences between corresponding maxima and the truth in days. The first three rows cover the secondary investigation period (from 2015-2018), while the next five rows cover the primary period (from 2018-2023). The last row provides overall statistics for all primary investigation period maxima sharing the same nominal perturbation. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Winter	GCP	1.5 ppm*			3 ppm			4.5 ppm		
	Date	$\overline{\text{Date}}$	Δ	σ	$\overline{\text{Date}}$	Δ	σ	$\overline{\text{Date}}$	Δ	σ
15/16	27.12	13.01	16	35	17.11	-40	26	12.03	75	59
16/17	29.12	28.12	-1	56	23.11	-36	76	05.02	38	79
17/18	30.12	07.01	8	33	11.02	42	48	09.02	40	156
18/19	28.12	25.12	-3	31	24.01	27	77	15.01	18	82
19/20	30.12	01.01	-3	48	24.11	-35	13	25.10	-66	66
20/21	01.01	07.01	6	42	21.12	-10	18	04.02	35	73
21/22	08.01	05.01	-3	25	22.01	14	31	05.05	117	45
22/23	11.01	05.01	-6	35	04.01	-6	12	01.04	81	18
Global mean over winters 18/19 - 22/23										
N.A.	N.A.	N.A.	-1	35	N.A.	0	47	N.A.	44	87

Similar to the analysis of the seasonal extrema positions, the amplitudes of the posterior fluxes show an expected increase in deviation from the true values with higher nominal measurement uncertainties. While this general trend is apparent, the limited number of realizations for the 3.0 ppm and 4.5 ppm scenarios does not allow for statistically robust conclusions. For example, the standard deviation (σ) of the amplitude deviations in summer for the 4.5 ppm scenario is lower than that of the 3.0 ppm case — and even lower than the winter counterpart of the 4.5 ppm case, which contradicts the expected pattern.

Another notable observation also for the 1.5 ppm perturbations is a tendency for amplitudes to be slightly overestimated in winter and underestimated in summer. However, these deviations remain small compared to their corresponding standard deviations (σ = 0.08 for winter and σ = 0.10 for summer) and are therefore not statistically significant. If these patterns are not the result of random noise, they may suggest an asymmetric influence of measurement errors on the inversion system. At this stage, however, my limited analysis

Table 4.2. Annual positions of flux summer minima for the GCP (Date) and corresponding mean posterior positions ($\overline{\text{Date}}$) from synthetic data with same nominal perturbation. Δ and σ represent the mean difference and standard deviation of differences between corresponding minima and the truth in days. The first three rows cover the secondary investigation period, while the next five rows cover the primary period. The last row provides overall statistics for all primary investigation period maxima sharing the same nominal perturbation. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Summer	GCP	1.5 ppm*			3 ppm			4.5 ppm		
	Date	$\overline{\text{Date}}$	Δ	σ	$\overline{\text{Date}}$	Δ	σ	$\overline{\text{Date}}$	Δ	σ
2015	04.07	25.07	21	52	08.07	4	4	07.07	3	57
2016	03.07	30.07	27	62	05.04	-89	60	15.07	12	16
2017	04.07	22.07	17	81	10.07	5	43	12.06	-23	169
2018	04.07	21.07	17	71	07.07	3	49	25.07	21	117
2019	29.06	05.07	6	56	26.05	-34	1	22.04	-68	33
2020	01.07	01.07	0	43	14.06	-16	41	29.07	29	41
2021	09.07	11.07	2	29	28.06	-12	80	25.08	47	122
2022	07.07	29.06	-8	24	13.07	7	13	11.10	97	22
2023	04.07	03.07	-1	45	24.06	-10	49	03.09	61	0
Global mean over summers 2018 - 2023										
N.A.	N.A.	N.A.	0	41	N.A.	-9	51	N.A.	34	96

does not provide sufficient evidence to support such a conclusion and further investigation would be required.

Summarizing the finding of this chapter, a phase deviation of about 35-40 days and an amplitude difference of 8%-10% relative to the Seasonless Prior can be estimated by the statistical errors of the thirty 1.5 ppm perturbed synthetic inversion runs. Summer fluxes are slightly more impacted by the random uncertainties. These characteristic values were used in the sections 5.2.2.1, 6.3.2.1 and 7.2.2.1 to compare the GCP extrema positions and amplitudes to the real data inversions.

Table 4.3. Amplitudes of flux winter maxima for the GCP (Value) and corresponding mean posterior positions (Mean) from synthetic data with same nominal perturbation relative to the Seasonless Prior. Δ and σ represent the mean difference and standard deviation of differences between corresponding maxima and the truth. The first three rows cover the secondary investigation period, while the next five rows cover the primary period. The last row provides overall statistics for all primary investigation period maxima sharing the same nominal perturbation. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Winter	GCP	1.5 ppm*			3 ppm			4.5 ppm		
	Value	Mean	Δ	σ	Mean	Δ	σ	Mean	Δ	σ
15/16	1.10	1.14	0.04	0.10	1.08	-0.02	0.35	1.38	0.28	0.32
16/17	1.10	1.13	0.03	0.11	1.22	0.12	0.10	1.57	0.47	0.35
17/18	1.10	1.15	0.05	0.07	1.18	0.08	0.11	1.50	0.40	0.21
18/19	1.10	1.13	0.03	0.08	1.34	0.24	0.15	1.54	0.44	0.20
19/20	1.12	1.10	-0.02	0.09	1.27	0.15	0.16	1.68	0.56	0.12
20/21	1.12	1.15	0.03	0.06	1.27	0.15	0.02	1.39	0.28	0.10
21/22	1.12	1.16	0.04	0.07	1.18	0.07	0.05	1.47	0.35	0.17
22/23	1.13	1.15	0.02	0.08	1.25	0.13	0.14	1.42	0.30	0.12
Global mean over winters 18/19 - 22/23										
N.A.	1.12	1.15	0.03	0.08	1.26	0.15	0.13	1.49	0.37	0.18

4.6.2 Inter-annual $\Delta ffCO_2$ emission trend derived from synthetic inversions

For the investigation of the inter-annual trends, the posterior fluxes were calculated using the Trendless Prior featuring a mean seasonal cycle (Tab. 3.1) and synthetic data with various magnitude of perturbation (1.5 - 4.5 ppm, Fig. 4.6). The prior uncertainty was again set to 40% to ensure that the posterior fluxes are mainly data-driven. However, for this sensitivity test the temporal correlation was increased to reduce the posteriors flexibility in the time domain. For this, I used two different frequency filters: F1T and F0.5T. F1T corresponds to one degree of freedom per year and F0.5T to one degree of freedom every two years (Sect. 3.2.2.2). These two time-correlation filters offer enough flexibility to replicate the inter-annual trend without significantly changing the mean seasonal cycle included in the Trendless Prior. The F1T filter offers more flexibility, allowing the inversion system to adjust shorter periods in the posteriors compared to F0.5T. The allowed flexibility of the

Table 4.4. Amplitudes of flux summer minima for the GCP (Value) and corresponding mean posterior positions (Mean) from synthetic data with same nominal perturbation relative to the Seasonless Prior. Δ and σ represent the mean difference and standard deviation of differences between corresponding minima and the truth. The first three rows cover the secondary investigation period, while the next five rows cover the primary period. The last row provides overall statistics for all primary investigation period maxima sharing the same nominal perturbation. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Summer	GCP	1.5 ppm*			3 ppm			4.5 ppm		
	Value	Mean	Δ	σ	Mean	Δ	σ	Mean	Δ	σ
2015	0.90	0.88	-0.02	0.11	0.68	-0.22	0.14	0.52	-0.38	0.34
2016	0.90	0.90	0.00	0.09	0.72	-0.18	0.11	0.61	-0.29	0.29
2017	0.90	0.84	-0.06	0.10	0.76	-0.14	0.15	0.70	-0.20	0.34
2018	0.90	0.89	-0.01	0.09	0.78	-0.12	0.24	0.95	0.06	0.09
2019	0.89	0.88	-0.01	0.09	0.98	0.09	0.21	0.98	0.09	0.21
2020	0.88	0.86	-0.02	0.12	0.87	0.00	0.09	1.03	0.15	0.13
2021	0.88	0.84	-0.04	0.10	1.07	0.19	0.10	0.90	0.02	0.15
2022	0.88	0.85	-0.03	0.09	0.78	-0.10	0.10	0.90	0.02	0.15
2023	0.87	0.88	0.01	0.08	0.88	0.01	0.03	1.04	0.17	0.00
Global mean over summers 2018 - 2023										
N.A.	0.88	0.86	-0.02	0.10	0.89	0.01	0.18	0.96	0.07	0.15

F1T filter might be too high for the investigations of the inter-annual trend but it can serve as indicator whether observed changes were caused by some shorter-term deviations or by a genuine trend. The posteriors for the 3 ppm and 4.5 ppm perturbations (three data series for each perturbation) are presented here to show the influence of the higher measurement errors, but are not evaluated quantitatively. The trend variations of the 1.5 ppm perturbation posteriors (30 in total) will serve as comparison measure for the real data inversions (Sect. 5.2.2.2, 6.3.2.2 and 7.2.2.2).

Figure 4.11 shows absolute posterior fluxes for the Germany+ domain with the F0.5T filter (F1T is not shown). Posteriors with the nominal 1.5 ppm perturbation successfully recovered the truth trend. As noted in the section 4.6.1, the 4.5 ppm inversions overestimate the truth fluxes but still show signs of the downward trend after 2020. These features were especially evident in the figure 4.12, which depicts the posterior fluxes relative to the Trendless Prior. The residual short-term variability of the truth (black line in Fig. 4.12), such as

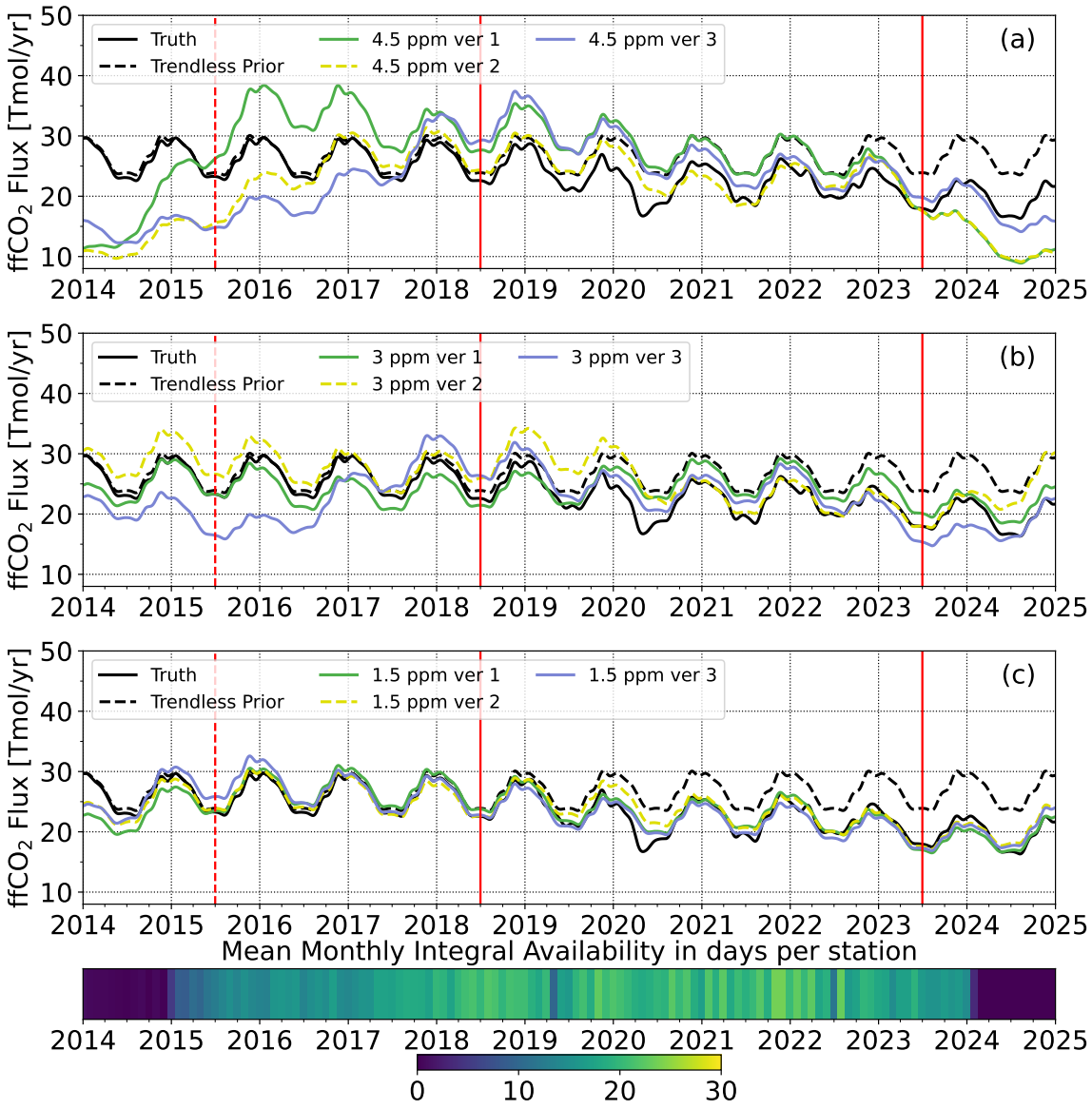

Figure 4.11. Analysis of the inter-annual trends derived from perturbed synthetic data (Fig. 4.6). Panels (a-c) show absolute $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Trendless Prior (black dashed line) that was used as input. The three colored lines represent posteriors with a 40% prior uncertainty and "F0.5T" time correlation filter (Sect. 3.2.2.2), derived from synthetic datasets with the same nominal perturbation. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

the sharp emission reduction during the 2020 COVID-19 lockdown, cannot be reconstructed with given reduced degrees of freedom in the temporal domain.

To provide a comparison for real-data inversions, I calculated the linear regression slopes

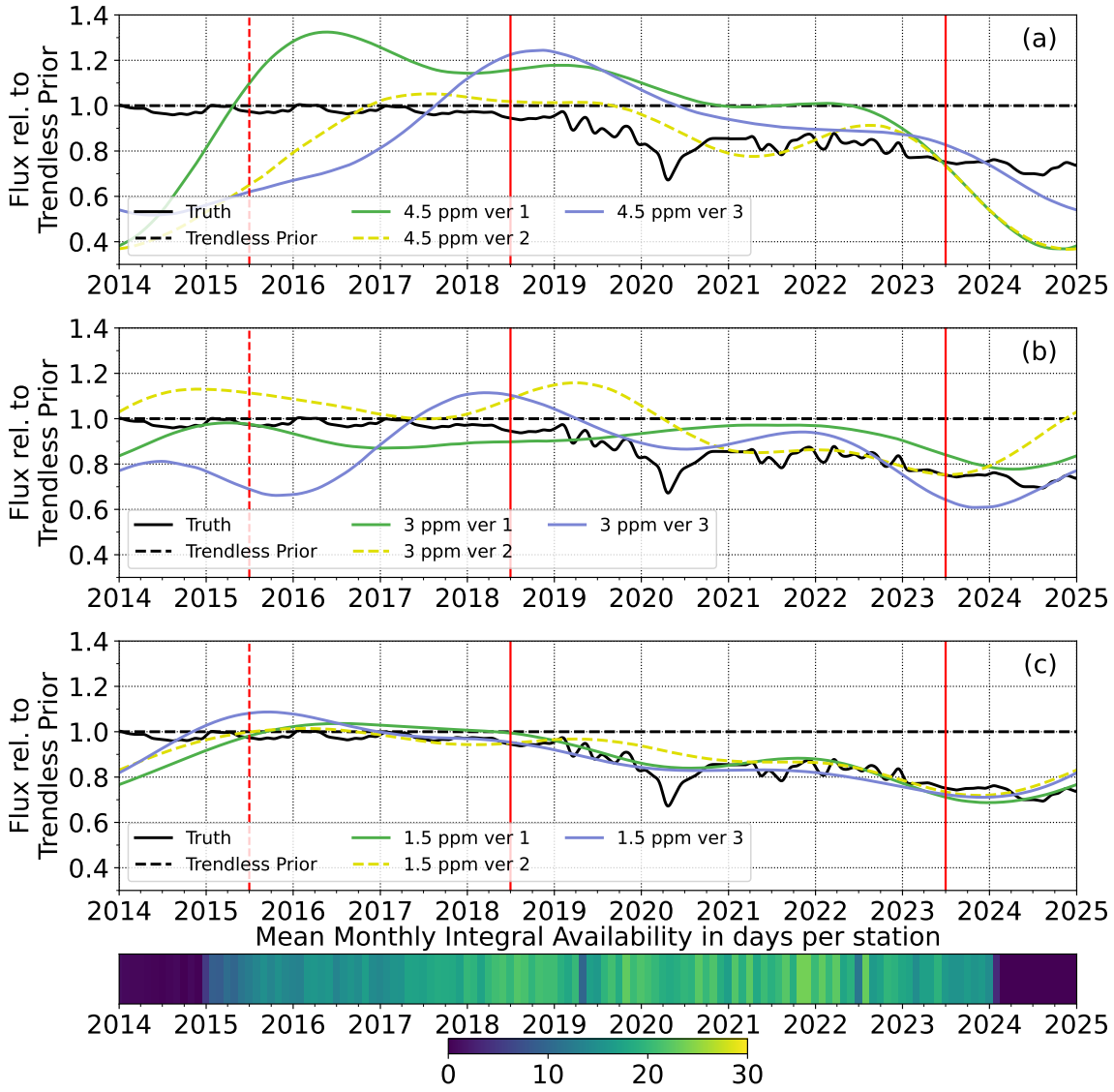

Figure 4.12. Analysis of the inter-annual trends derived from perturbed synthetic data (Fig. 4.6) relative to the Trendless Prior. Panels (a-c) show relative $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Trendless Prior (black dashed line) that was used as input. The three colored lines represent posteriors with a 40% prior uncertainty and "F0.5T" time correlation filter (Sect. 3.2.2.2), derived from synthetic datasets with the same nominal perturbation. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

for thirty posterior time series corresponding to the 1.5 ppm perturbation (Tab. 4.5). I focused on the standard deviation of these slopes as a measure of the system's sensitivity to observational errors. As this analysis used only synthetic data based on integral samples, the

period of the calculated regression was adapted to the real datasets of integral-based, CO-based, and flask-based $\Delta ffCO_2$. For the integral-based inversions the integral's main investigation period (01.07.2018–01.07.2023, 5 years) will be used for the linear regression. For the CO-based inversions, the usual period, over which the real data inversions are calculated (01.01.2019 - 01.01.2025, 6 years, Sect. 4.3), does not possess high enough data density in the integral-based data (colorbar in Fig. 4.11). Therefore, 01.07.2017–01.07.2023 (6 years) period was used instead. Similarly, for the flask-based inversions (with usual real data inversion range of 01.11.2021-01.01.2025) linear regressions over 01.07.2020 - 01.07.2023 (3 years) period were used here for the estimation of the variability of slopes. This period is 2 month shorter compared to the real flask period to minimize the effect of the COVID-19 rapid flux drop in the middle of 2020. As in these synthetic there are no transport errors, I expect that the perturbation will have the dominant effect on the posteriors, rendering the effects of the shifted time periods negligible.

The absolute slopes for the 5- and 6-year periods overestimate the GCP-based trend, which might be a result of the still not sufficient number of realizations in terms of Monte-Carlo simulations (Tab. 4.5, upper half). The 3-year period, however, corresponds well with the truth. The F1T slope is slightly smaller than the F0.5T slope, but this difference is not significant.

The standard deviations of the slopes are significantly different across the periods. The shortest 3-year period shows the largest variation (0.60 Tmol/yr for F0.5T and 0.73 Tmol/yr for F1T), 2–3 times larger than the variations for the 5- and 6-year periods (0.24–0.31 Tmol/yr). This demonstrates that short time series are more susceptible to observational errors when calculating posterior trends. The 5-year period's variability is also 25% higher than the 6-year period's. The emission reduction trends based on the fluxes relative to the Trendless Prior (Tab. 4.5, lower half) show the same patters.

4.6.3 Absolute and annual mean $\Delta ffCO_2$ emissions derived from synthetic inversions

This sensitivity test explores the capability of the inversion system to estimate the absolute and annually aggregated fossil CO_2 emissions for the AoC, using both the Flat Zero Prior and the Flat 200p Prior as input (not shown) with F2T time-correlation filter and 40% prior uncertainty. Synthetic integral-based data with 1.5-4.5 ppm perturbation magnitudes (Fig. 4.6) were used as input data for the inversions. As with the previous analysis, the quantitative measure for the deviations due to the observational errors (for comparison with the real data inversions, Sect. 5.2.2.3, 6.3.2.3 and 7.2.2.3) are derived from the thirty 1.5 ppm perturbation posteriors. Due to the inability of the posteriors to reproduce the short-term variability of

Table 4.5. Slopes of linear regressions for GCP and posterior fluxes in the Germany+ domain. Posterior estimates were derived from 30 synthetic datasets with a 1.5 ppm perturbation, using a Trendless Prior and F0.5T and F1T temporal correlation filters. Mean posterior slopes are shown with their standard deviations in parentheses. For GCP, the value in parentheses represents the slope error. The upper and lower halves of the table show trends for absolute fluxes (Fig. 4.11) and fluxes relative to the Trendless Prior (Fig. 4.12), respectively, for 3-, 5-, and 6-year periods.

Flux type	01.07.20-01.07.23	01.07.17-01.07.23	01.07.18-01.07.23
	Absolute trend [Tmol/yr]		
GCP	-0.54(0.08)	-0.87(0.03)	-0.74(0.04)
F05T (mean, #30)	-0.54(0.60)	-1.01(0.24)	-0.90(0.30)
F1T (mean, #30)	-0.49(0.73)	-0.98(0.24)	-0.88(0.31)
	Relative trend [%]		
GCP	-1.9(0.1)	-3.2(0.1)	-2.8(0.1)
F05T (mean, #30)	-2.0(2.2)	-3.8(0.9)	-3.4(1.1)
F1T (mean, #30)	-1.8(2.7)	-3.7(0.9)	-3.3(1.2)

the truth, the posterior absolute fluxes were compared to the Smooth Prior (Tab. 3.1).

Figure 4.13 and 4.14 depict the absolute and relative to the Smooth Prior fluxes in the Germany+ domain. Similar to the analysis in section 4.6.1, the posteriors for the higher nominal perturbations show increased amplitude and shifted positions of the extrema.

The average RMSD's between the posteriors and the truth in the main investigation period (Tab. 4.6) represent characteristic values for the deviations from the truth induced solely by the observational errors in these synthetic runs, and as expected they increase with higher perturbation magnitude. The relative RMSD of 11%±2% can be already caused by the perturbations of 1.5 ppm magnitude.

Panels (a) to (c) in Fig. 4.15 show the annual mean fluxes of the truth (in black) as well as the three example posterior fluxes in colored lines which I have obtained when using the Flat Zero Prior as input. As expected the annual mean posterior fluxes deviate stronger from the true annual mean flux with increasing random uncertainty in the pseudo-data. It can also be noted that the annual mean fluxes of the posteriors for all perturbation levels show a tendency to positive biases.

Tables 4.7 and 4.8 summarize the average relative deviations of the annual mean fluxes across all realizations of the perturbed pseudo-data, grouped by perturbation level and by

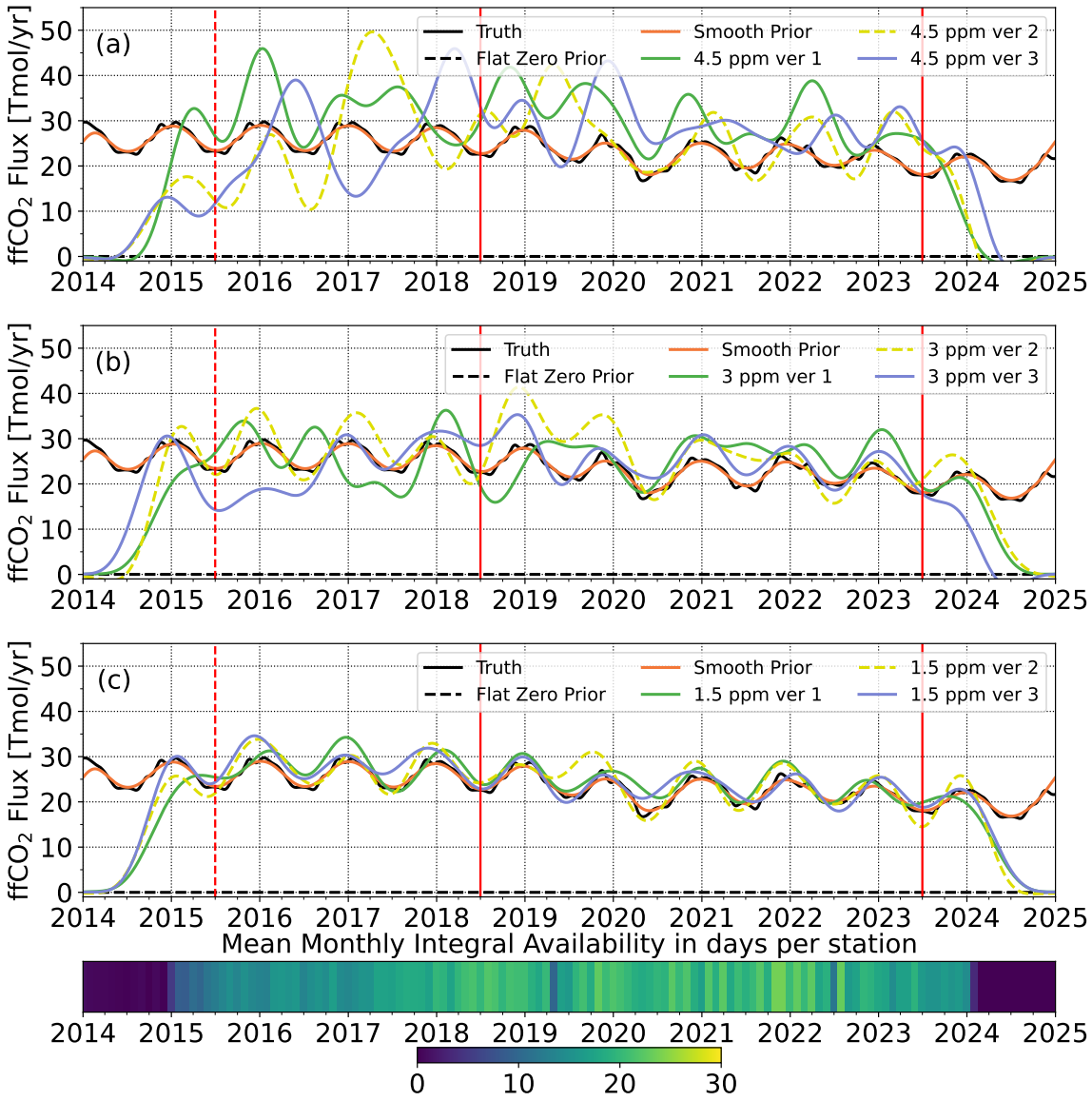

Figure 4.13. Analysis of the absolute emissions derived from perturbed synthetic data (Fig. 4.6). Panels (a-c) show absolute $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Flat Zero Prior (black dashed line) that was used as input and the Smooth Prior (orange solid line) that was used for comparison. The three colored lines represent posteriors with a 40% prior uncertainty, derived from synthetic datasets with the same nominal perturbation. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

the two applied Flat priors (Zero and 200p), each relative to the true values. Based on these results, the positive bias of the posterior fluxes estimated using the Flat Zero prior is evident.

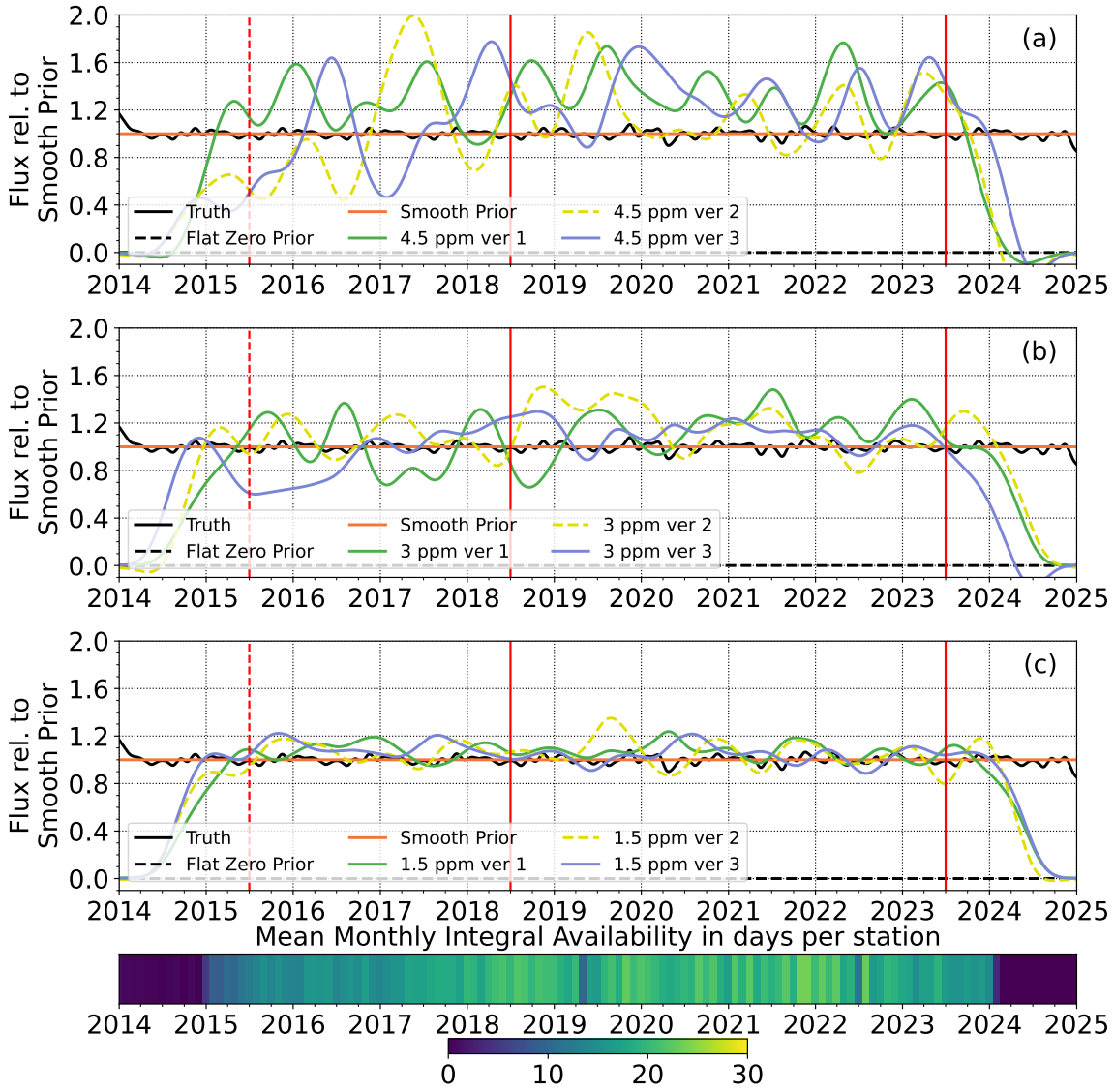

Figure 4.14. Analysis of the emissions derived from perturbed synthetic data (Fig. 4.6) relative to the Smooth Prior. Panels (a-c) show relative $\Delta ffCO_2$ fluxes in the Germany+ domain for decreasing perturbations: 4.5 ppm (a), 3 ppm (b), and 1.5 ppm (c). The truth (GCP, black solid line) is shown alongside the Flat Zero Prior (black dashed line) that was used as input and the Smooth Prior (orange solid line) that was used for comparison. The three colored lines represent posteriors with a 40% prior uncertainty, derived from synthetic datasets with the same nominal perturbation. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

A symmetrical negative bias is observed in the posterior fluxes derived with the Flat 200p prior. To further examine this bias behavior—referred to in the following as the "over- and undershoot" effect, additional inversions (not shown) were conducted using Flat 50p, 100p

Table 4.6. Means and standard deviations of the RMSD between the synthetic perturbed posterior fluxes and the truth relative to the Smooth Prior. Three nominal perturbations were used to simulate the measurement errors of various degrees. Two different priors were used as an input to the inversions: the Flat Zero Prior and the Flat 200p Prior (Tab. 3.1). Only the fluxes in the period from 01.01.2018 to 01.07.2023 were considered for the RMSD calculations. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Prior	1.5 ppm*		3 ppm		4.5 ppm	
	Mean	Std	Mean	Std	Mean	Std
Zero	0.11	0.02	0.19	0.04	0.32	0.04
200p	0.11	0.02	0.16	0.06	0.29	0.02

and Flat 150p priors, corresponding to 1/4, 1/2 and 3/4 of the Flat 200p prior fluxes. The results revealed a linear and symmetric response in the posterior fluxes: as the prior emissions increased, the posteriors increasingly underestimated the truth; conversely, lower priors led to overestimation. The posterior derived from the Flat 100p prior aligned most closely with the truth, within the expected range of variability caused by the measurement perturbations. Interestingly, similar behavior was later identified by other CarboScope modeling teams for different trace gases, following my inquiry on this matter. This suggests that the over- and undershoot phenomenon may be a fundamental property of this inversion system. However, it appears to manifest only in sensitivity studies with extreme and intentionally minimal prior constraints. In realistic inversions with well-informed priors, this effect does not appear to be relevant and is unlikely to introduce a significant bias in the final flux estimates.

Despite the systematic biases observed in the Flat Zero and 200p sensitivity tests, I maintain their use, as they are best suited to investigate the information content of the observations. To account for the over- and undershoot behavior observed in the sensitivity tests, both the mean deviation from the truth and its standard deviation were used to evaluate the real-data posteriors.

Table 4.7. Mean annual emissions of the synthetic perturbed integral-based posteriors relative to the truth (GCP) using the Flat Zero Prior. The means and the standard deviations were calculated from the sets of the posteriors with the same nominal perturbation. The last row provides the statistics over the years 2018-2023. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Year	1.5 ppm Zero*		3 ppm Zero		4.5 ppm Zero	
	Mean	Std	Mean	Std	Mean	Std
2018	1.08	0.05	1.11	0.11	1.24	0.16
2019	1.06	0.06	1.20	0.14	1.37	0.08
2020	1.06	0.07	1.12	0.01	1.25	0.17
2021	1.06	0.06	1.20	0.04	1.16	0.07
2022	1.05	0.05	1.06	0.09	1.22	0.10
2023	1.09	0.05	1.08	0.09	1.21	0.09
Global mean over 2018 - 2023						
Total	1.07	0.06	1.13	0.11	1.24	0.14

Table 4.8. Mean annual emissions of the synthetic perturbed integral-based posteriors relative to the truth (GCP) using the Flat 200p Prior. The means and the standard deviations were calculated from the sets of the posteriors with the same nominal perturbation. The last row provides the statistics over the years 2018-2023. (*) 1.5 ppm statistics based on thirty perturbed datasets instead of three and were used later for the quantitative analysis in the real-data inversions.

Year	1.5 ppm 200p*		3 ppm 200p		4.5 ppm 200p	
	Mean	Std	Mean	Std	Mean	Std
2018	0.95	0.05	0.98	0.11	1.11	0.16
2019	0.95	0.06	1.09	0.14	1.26	0.08
2020	0.94	0.07	1.01	0.01	1.13	0.17
2021	0.94	0.06	1.08	0.04	1.04	0.07
2022	0.94	0.05	0.95	0.09	1.12	0.10
2023	0.93	0.05	0.93	0.09	1.06	0.09
Global mean over 2018 - 2023						
Total	0.94	0.06	1.01	0.11	1.12	0.14

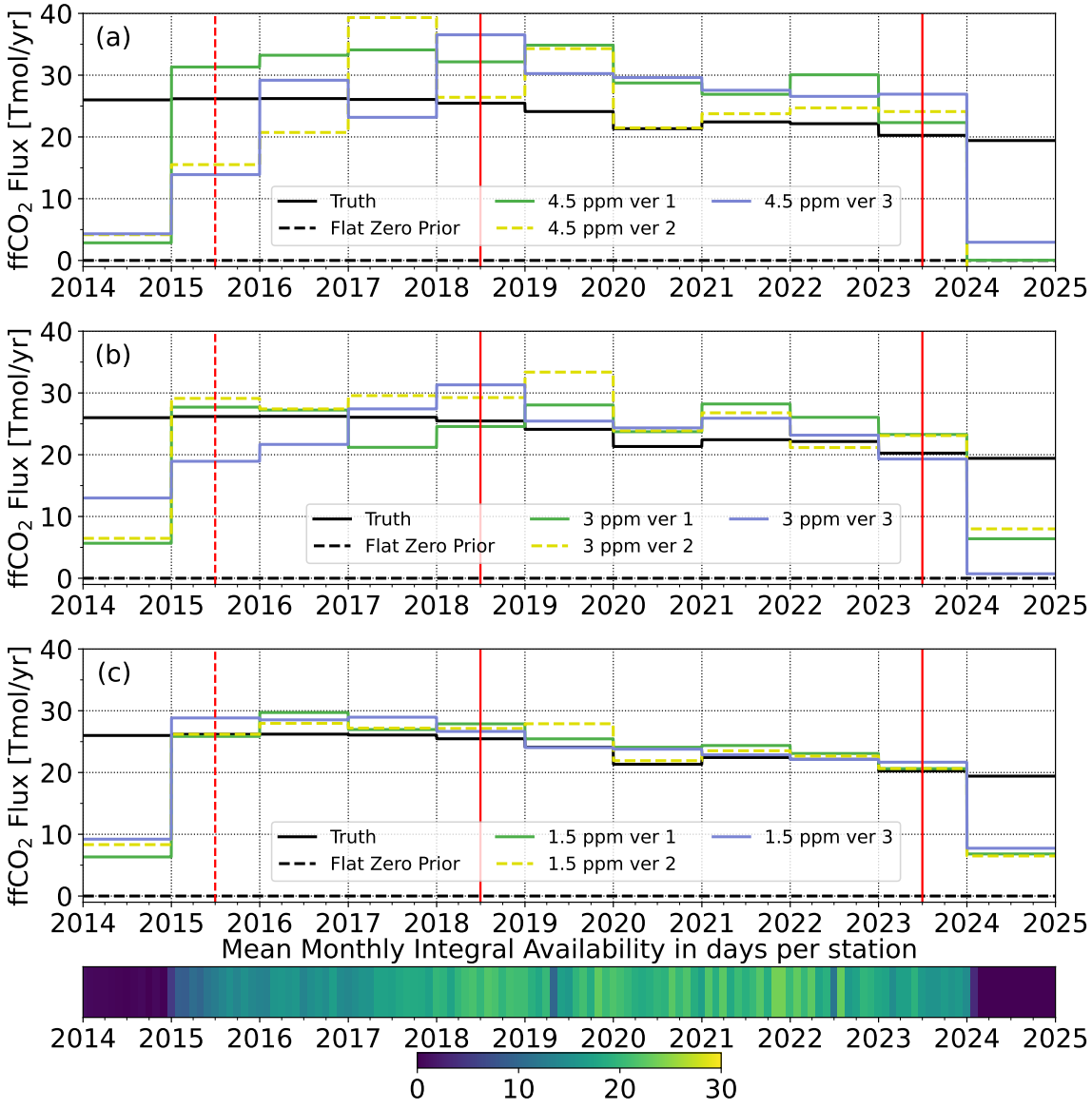

Figure 4.15. Annual mean $\Delta ffCO_2$ emissions derived from synthetic integral-based inversions. Panels (a)-(c) display yearly flux values in Germany+ domain, with decreasing perturbations: panel (a) - 4.5 ppm, panel (b) - 3 ppm, panel (c) - 1.5 ppm. Truth (black solid line) and the Flat Zero Prior (black dashed line) are shown alongside posterior results (colored lines) derived from the perturbed data. Prior uncertainty is 40%. Red lines indicate the main investigation period (01.07.18-01.07.23), with the dashed red line marking the secondary analysis period (from 01.07.15). The color bar at the bottom shows the network-wide mean monthly availability of integral-based estimates per station.

Chapter 5

Flask-based ΔffCO$_2$ inversions

5.1 Introduction and motivation

Flask samples offer several advantages for estimating ΔffCO$_2$ signals. First, the semi-automated sampling routines implemented in ICOS flask samplers allow for the evaluation of specific events or meteorological conditions. Furthermore, ICOS has developed advanced tools that use forecasted back-trajectories, which automatically trigger the sampling system (ICOS RI, 2020). This enables air masses originating from predefined areas of interest to be sampled, allowing for highly targeted data collection. Similar techniques can be used to avoid air masses that were potentially contaminated by nuclear facilities, ensuring that nuclear corrections remain small compared to continental or regional changes in the Δ^{14}C content of the sampled air mass. This approach was recently implemented during the CORSO project (Levin et al., 2024), which increased ^{14}CO$_2$ observations at 10 central western ICOS stations by a factor of five for the year 2024. As described in ICOS RI, 2020, most of the flasks were collected in the afternoon hours. The afternoon period is chosen to reduce the model-data-mismatch related to the transport model error, as atmospheric transport models generally exhibit their best performance during this time of day (Geels et al., 2007; Peng et al., 2023). Finally, the same air samples can be evaluated for other ΔffCO$_2$ proxies such as CO (Sect. 6), nitrogen oxides (NO$_x$) (Lopez et al., 2013; Jäschke, 2021; Rosendahl, 2022), and Atmospheric Potential Oxygen (APO) (Pickers et al., 2022; Rödenbeck et al., 2023), which allows for a cross-validation of other ΔffCO$_2$ estimates with ^{14}C-based ΔffCO$_2$ estimates.

Despite the advantages mentioned above, flask-based ΔffCO$_2$ series also have limitations. A primary drawback is the short integration time of one hour. If the model transport deviates from real atmospheric conditions during this sampling hour, it can inflate the error in the model-data-mismatch (MDM) and potentially skew the results of the inversion. Maier et al., 2024b recently demonstrated that, in the case of semi-urban inversions, misalignment

between the modeled and actual meteorology severely impairs the model's ability to accurately represent individual hours, effectively preventing the use of flask-based $^{14}CO_2$ observations. However, at ICOS stations, which are intentionally located at greater distances from local emission hotspots, meteorologically-driven model-data-mismatches are expected to be significantly smaller. In addition, if such an MDM is caused by statistical errors rather than systematic bias, a sufficiently high number of samples will remedy this disparity, as statistical MDM will average out to zero. The standard $^{14}CO_2$ flask analyzing frequency in (ICOS RI, 2020) is one hourly sample every two weeks, though in reality this is often less frequent, especially in summer as can be seen in flask coverage overview in Section 4.3. Such a sampling frequency is lower as the typical duration of the synoptic weather conditions, that are usually on the timescale of 2-5 days, or up to 10 days for some large scale events (Garreaud & Aceituno, 2007; Franzke et al., 2020). Thus, there is a high probability that the sampled flasks belong to systematically different atmospheric conditions, potentially sampling different catchment areas. As a consequence, this infrequent sampling leads to the posterior results of the inversion system relying on individual flasks, thereby limiting the resilience of the inversion system to errors. This underlines the importance of projects like CORSO (Levin et al., 2024), within which the sampling frequency was increased to about one flask every third day, increasing the robustness of the inversion results.

In this chapter, I first compare the observed flask-based $\Delta ffCO_2$ concentrations with the forward-modeled $\Delta ffCO_2$ signals (Sect. 5.2.1). This provides a first qualitative insight into the consistency between the observations and the modeled fossil CO_2 patterns. I then present the key characteristics of the posterior fossil fuel CO_2 fluxes over the Germany+ domain as derived from the flask-based $\Delta ffCO_2$ observations using the Evaluation Matrix (Sect. 5.2.2). Thereafter, the sensitivity of the posterior results to the $\Delta^{14}C$ background assumptions (Sect. 5.2.3.1) is investigated and I assess the influence of individual measurement sites (Sect. 5.2.3.2). This allows for the identification of the most reliable subsets of observations and time seasons for constraining fossil fuel emissions. Finally, I perform two full Bayesian inversions using the unmodified GCP and EDGAR prior fluxes, along with their best-estimate uncertainties, to derive optimized fossil fuel emissions for the constraint area (Sect. 5.2.4).

5.2 Results for flask-based $\Delta ffCO_2$ estimates

5.2.1 Forward runs and model-data-mismatch

First, I compare the $\Delta ffCO_2$ derived from the flask observations with the forward simulated $\Delta ffCO_2$ concentrations (FWD) calculated as described in Section 3.2. The GCP inventory

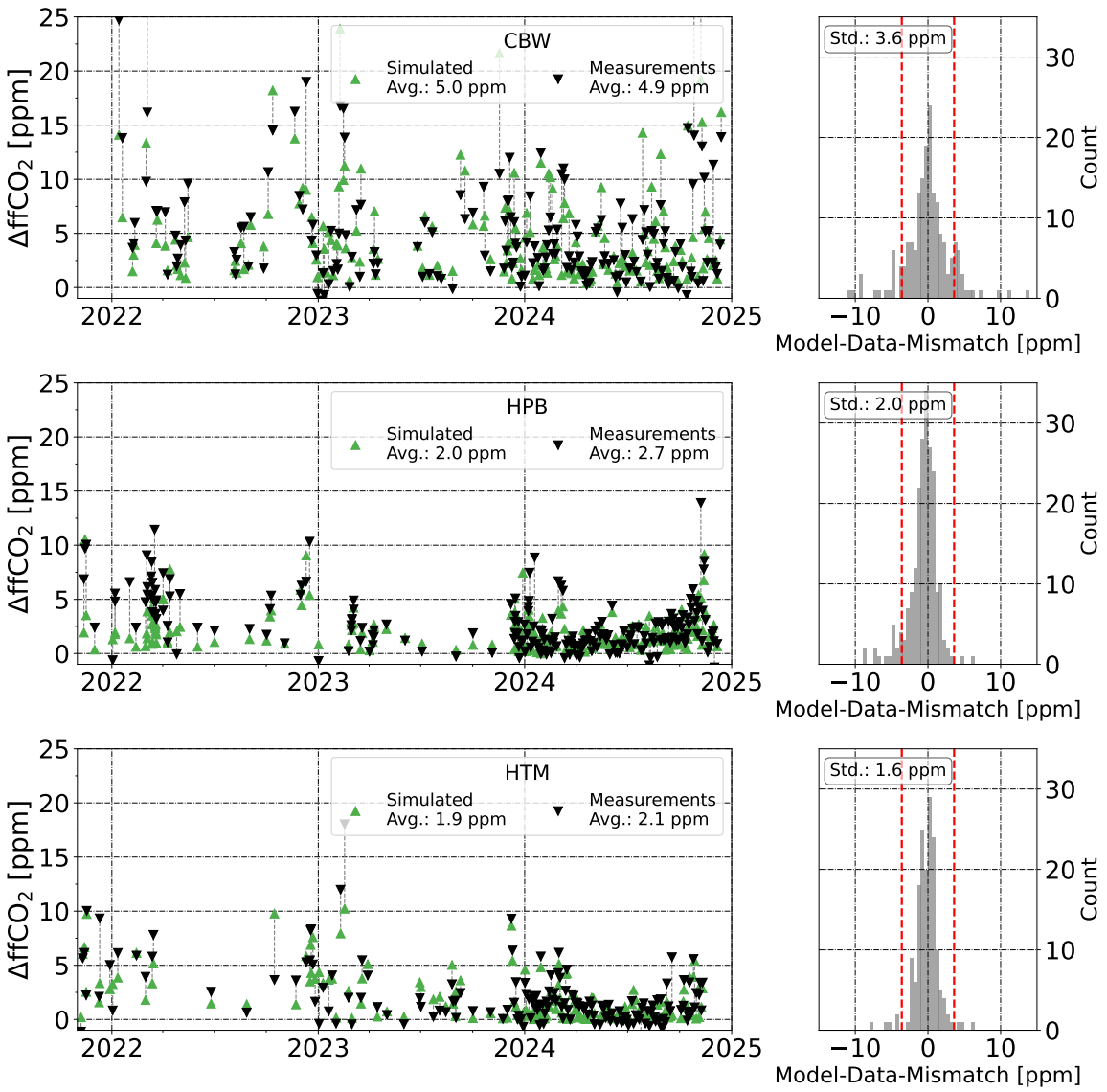

Figure 5.1. The flask-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ concentrations based on the GCP inventory (green triangles) for three example measurement sites, CBW, HPB and HTM. Observed and simulated $\Delta ffCO_2$ concentrations are connected by dashed lines. Average concentrations over the depicted time period are given in the legend. The right panels show the model-data-mismatch distribution and its variance is given in the legend. The red dashed lines show the 3 σ range of the mean uncertainty of the observational data. The shown time period has the highest data density (01.11.2021 - 01.01.2025, Sect. 4.3).

was used for fossil emissions in the FWD runs. The simulated concentrations are given for the exact sampling times and locations of the observations.

Figure 5.1 shows the measured and simulated data for three example observational sites. Corresponding figures for the other stations are in the appendix (Sect. A.2). These sites were chosen as they represent different types of catchment areas monitored by the

ICOS stations. Cabauw (CBW) site usually measures elevated levels of $\Delta ffCO_2$ in comparison to other stations in the network due to its proximity to Rotterdam, situated about 20 km to the west of CBW (see Fig. 2.4). The average observed $\Delta ffCO_2$ concentration in CBW is 4.9 ppm which is very well reproduced by the GCP-based forward modeled $\Delta ffCO_2$ concentrations averaging to 5.0 ppm. Although the observed and modeled mean values agree well, the MDM standard deviation of 3.6 ppm is three times larger than the average uncertainty of the $\Delta ffCO_2$ observations and corresponds to about 70% of the measured average $\Delta ffCO_2$ signal. The Hyltemossa (HTM) station in Sweden is located far away from major emission sources and shows lower $\Delta ffCO_2$ signals. Here too, the observed and FWD modeled $\Delta ffCO_2$ concentration agree well, with 2.1 ppm and 1.9 ppm respectively. Corresponding to the smaller overall signal, the standard deviation of the MDM also decreases, but remains proportionally comparable to CBW at around 75%. In contrast to the the former two stations, Hohenpeißenberg (HPB) site is situated at 934 m a.s.l., within a more complicate topography in the German Alpine foothills (see Tab. 2.1) causing additional challenges for the transport model. This is reflected in the larger deviation between the observed and the FWD modeled mean $\Delta ffCO_2$ concentration of 2.7 ppm and 2.0 ppm respectively. The standard deviation of the MDM remains also for this stations at 75% of the observed $\Delta ffCO_2$ signal.

The statistics of the observed and simulated $\Delta ffCO_2$ flask data in the ICOS network is shown in Table 5.1. For stations in central Europe, excluding HPB, the averages of simulated ($\bar{X}_{sim}$) and observed ($\bar{X}_{obs}$) $\Delta ffCO_2$ during the depicted time period from 2021 to 2024 agree quite well with each other compared to absolute signals. This relatively small FWD MDM already highlights two key findings. First, the GCP prior fossil CO_2 emissions are already in reasonable agreement with the observations, even before optimization through the inversion algorithm; secondly, the applied STILT transport model performs quite well on average. Larger deviations are found between the modeled and observed concentrations of HPB, which have previously been attributed to its location in complex terrain. The Scandinavian stations NOR, SVB, and PAL show decreasing concentrations of $\Delta ffCO_2$ with increasing latitude, leading to a higher MDM relative to the overall signal.

The variance ($\sigma_{sim-obs}$) of MDM is higher than the 1 σ uncertainty of the observations and correlates, for the non-background stations in central Europe, with the $\Delta ffCO_2$ signal strength. The proportion of data points with an MDM that exceeds three times the mean observational uncertainty ($|\bar{X}_{sim} - \bar{X}_{obs}| > 3\bar{e}$ in Tab. 5.1) also increases with the mean concentrations of $\Delta ffCO_2$. This suggests that the MDM arises not only from observational uncertainties but also from uncertainties in atmospheric transport and sensitivity to emissions. This again emphasizes the importance of the model's representativeness on an hourly timescale, which remains a key challenge for flask samples, as discussed in Section 5.1. The

Table 5.1. Statistics of the observed and simulated $\Delta ffCO_2$ concentrations of the flask samples in the ICOS network from 2021 to 2024. $\bar{X}_{obs}$ and $\bar{X}_{sim}$ depict mean observed and simulated signals for the corresponding site. $\bar{X}_{sim}$ - $\bar{X}_{obs}$ and $\sigma_{sim-obs}$ stand for the mean model-data-mismatch (MDM) and its standard deviation. $\bar{e}$ and SNR denote mean observation error and mean Signal-to-Noise Ratio for a single observation. $|\bar{X}_{sim} - \bar{X}_{obs}| > 3\bar{e}$ shows the approximate portion of the data with higher absolute MDM then triple of the $\bar{e}$. (*) KIT was not used in regular inversions, only in sensitivity studies.

Site	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]	$\bar{e}$ [ppm]	$\lvert\bar{X}_{sim} - \bar{X}_{obs}\rvert > 3\bar{e}$ [%]	SNR [rel.]
CBW	4.9	5.0	0.1	3.6	1.2	20%	4.1
GAT	2.1	2.1	0.0	1.8	1.1	10%	1.9
HPB	2.7	2.0	-0.7	2.0	1.2	10%	2.2
HTM	2.1	1.9	-0.2	1.6	1.2	0%	1.7
JFJ	0.8	0.7	-0.1	1.7	1.3	0%	0.6
KIT*	4.6	5.0	0.4	4.1	1.3	20%	3.6
KRE	1.5	1.5	0.1	1.4	1.2	0%	1.2
LIN	4.0	3.7	-0.3	3.1	1.2	20%	3.3
NOR	1.1	1.1	0.0	1.6	1.3	0%	0.8
OPE	2.6	2.2	-0.4	2.5	1.2	10%	2.1
OXK	3.6	3.3	-0.3	2.2	1.3	10%	2.8
PAL	-0.0	0.2	0.2	0.9	1.2	0%	-0.0
STE	3.9	3.3	-0.6	2.3	1.2	1%	3.3
SVB	0.7	0.2	-0.5	1.4	1.3	0%	0.6

mean signal-to-noise ratio (SNR) is generally low, with few remote stations having an SNR lower than 1. This underscores the sensitivity of the $\Delta^{14}C$ data to the measurement errors investigated in Section 4.6.

All stations shown in Table 5.1 were used in the regular inversion except the KIT site. This site was excluded due to the high probability of $\Delta^{14}C$ contamination from research activities on the campus of the Karlsruhe Institute of Technology. Such a $\Delta^{14}C$ contamination would mask some of the $\Delta ffCO_2$ signal, potentially leading to a higher MDM and hence to lower fossil fluxes in the posteriors. Efforts were made to sample flasks only during wind conditions that would avoid the influence of the campus sector, but due to its immediate proximity, this was not possible. The effect on the posteriors when including the KIT site in the data set was investigated in Section 5.2.3.2.

5.2.2 Application of the Evaluation Matrix to the flask-based inversions

The flask-based $\Delta ffCO_2$ inversions were analyzed according to the set of experiments described in Section 3.2.4 to assess the information content of the observation data. The differences between the GCP emission inventory assumed to be the truth in the Evaluation Matrix and the resulting posterior fluxes were classified using the results of the corresponding integral-based sensitivity experiments applied to the perturbed pseudo-data sets in Section 4.6. As noted there, I assume that sampling routine differences can be neglected for the synthetic inversions with no inherent transport model errors, and the results from those sections are applicable for flask-based inversions as well.

5.2.2.1 Seasonal cycle of $\Delta ffCO_2$ emission trend derived from flask-based inversions

The posterior fluxes derived from flask-based $\Delta ffCO_2$ observations for the Germany+ domain, estimated using the Seasonless Prior and a prior uncertainty of 40%, are shown in Figure 5.2. Panel (a) displays the absolute posterior fluxes in comparison to the GCP prior. Notably, the posterior fluxes before November 2021 exhibit unrealistic seasonal structure and absolute values due to the limited availability of flask data, which leads to overfitting of isolated model–data-mismatches. The summer minimum of 2021 is missing in the posterior. Due to the six-month time correlation applied in this study (F2T, see Sect. 4.4), this anomaly influenced the emission maximum in 21/22 winter, preventing a robust analytical determination of its phase and amplitude. Nonetheless, visual inspection indicates that the posterior winter maxima align closely with the GCP inventory in both amplitude and phase.

Panel (b) of Figure 5.2 shows the fluxes normalized to the Seasonless Prior, used to quantitatively describe the seasonality in Table 5.2. The amplitude of the posterior winter maxima matches that of the GCP inventory within 10%. This is slightly above the expected range based on the 1.5 ppm synthetic error scenario (8%; Tab. 4.3) derived for the 30 random perturbation of the true GCP flux. The phase shifts for all winter maxima remain within the uncertainty range derived from the synthetic studies (35 days; Tab. 4.1).

The positions of summer minima also occur within the expected temporal uncertainty range (41 days; Tab. 4.2). While the posterior minimum for summer 2022 agrees well in amplitude with the GCP value, the subsequent years — 2023 and 2024 — show progressively deeper summer minima, with the 2024 value approximately 30% below the GCP. These deviations are unlikely to be driven solely by data sparsity: while flask availability was indeed lower in 2023, the summer of 2024 was well-sampled across the network (see bottom color bar in Fig. 5.2) due to intensified sampling under the CORSO project (Levin et al., 2024).

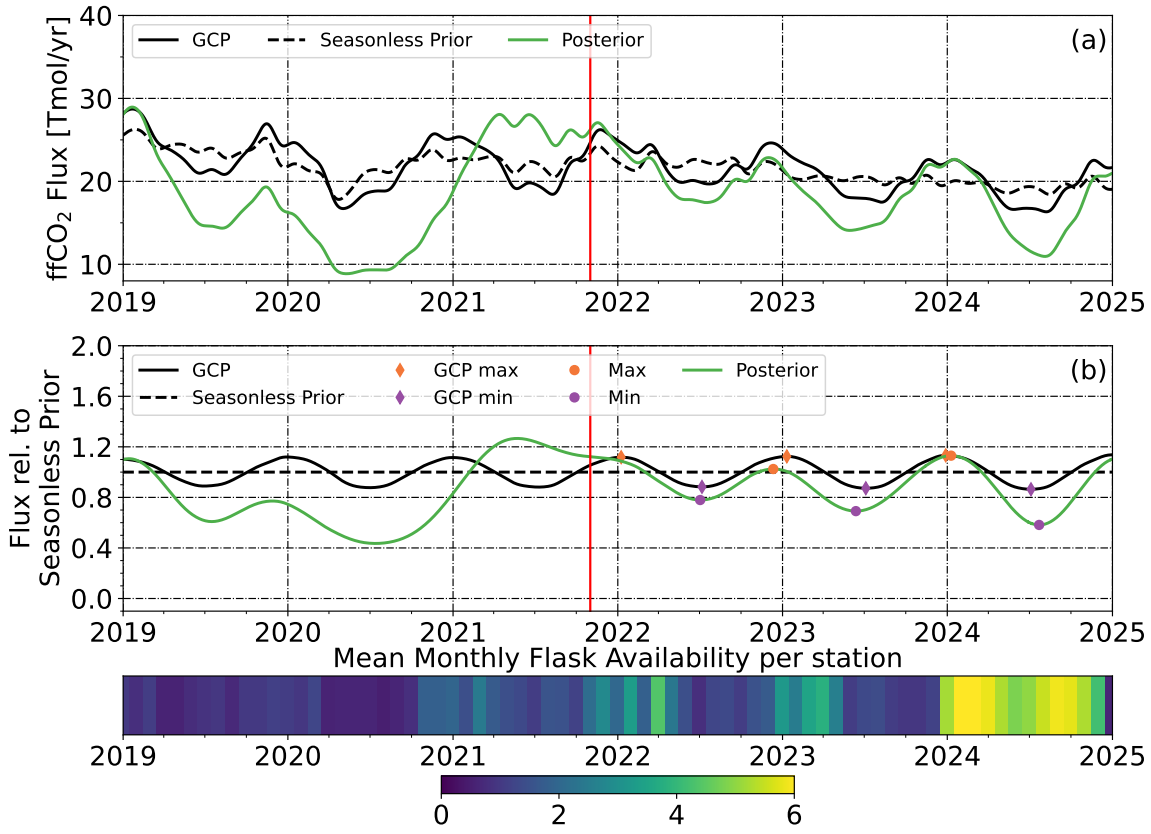

Figure 5.2. Results of the seasonal cycle investigation of the $\Delta ffCO_2$ flux estimation derived from flask-based inversions in the Germany+ domain. Panel (a) shows the absolute and panel (b) the relative to the Seasonless Prior $\Delta ffCO_2$ fluxes. The GCP inventory (black solid line) and the Seasonless Prior used as input (black dashed line) are shown alongside posterior results (green line). Orange and purple circles mark the maxima and minima of the posterior, with corresponding GCP extrema marked by diamonds. Red line indicates the main investigation time frame starting on 01.11.2021. The lower section of the figure depicts the corresponding timeline of mean monthly flask sample number per station as a heat map.

A possible explanation lies in the definition of the $\Delta^{14}C$ background used in the calculation of the $\Delta ffCO_2$ observations. As described in Section 3.1.1, the current $\Delta^{14}C$ background is based on a combination of trends from multiple clean air sites in the Northern Hemisphere and the seasonality derived from previous years of the MHD $\Delta^{14}C$ data. The discontinuation of the MHD time series may thus have introduced biases, especially in summer, when the signal-to-noise ratio is lower. This issue is discussed in more detail in Section 5.2.3.1, where the uncertainty of the $\Delta^{14}C$ background and its impact on posterior fluxes is systematically assessed.

Table 5.2. Positions (”Date” columns) and relative to the Seasonless Prior amplitudes (”Ampl.” columns) of the seasonal extrema extracted from the GCP inventory and the normalized posterior using flask-based $\Delta ffCO_2$ data as well as the difference between them.

Season	GCP		Flask-based $\Delta ffCO_2$		Posterior - GCP	
	Date	Ampl. [rel.]	Date	Ampl. [rel.]	Phase shift [days]	Ampl. diff. [rel.]
Winter 21/22	08.01	1.12	N.A.	N.A.	N.A.	N.A.
Winter 22/23	11.01	1.12	11.12	1.02	-31	-0.10
Winter 23/24	28.12	1.13	09.01	1.13	12	0.00
Summer 2022	07.07	0.88	03.07	0.78	-4	-0.10
Summer 2023	04.07	0.87	12.06	0.69	-22	-0.18
Summer 2024	04.07	0.86	22.07	0.58	18	-0.28

5.2.2.2 Inter-annual $\Delta ffCO_2$ emission trend derived from flask-based inversions

The inter-annual $\Delta ffCO_2$ emission trend of the flask-based inversion posteriors was examined using the Trendless Prior (Tab. 3.1). As stated before (Sect. 3.2.4 and 4.6), the Trendless Prior has a mean seasonal cycle but no inter-annual trend. The prior uncertainty in this sensitivity test was set to 40%. The absolute posterior fluxes (panel (a)) and the fluxes relative to the Trendless Prior (panel (b)) for the Germany+ domain are depicted in Figure 5.3. As described in Section 3.2.4, the temporal correlation was increased to one year (F1T) and two years (F0.5T) to limit the flexibility of the posteriors to inter-annual trends.

The $\Delta ffCO_2$ emission trends of the posteriors resemble the GCP inventory trend much closer then the absent trend of the constructed Trendless Prior. This confirms that the downward trend of the GCP inventory is also contained in the flask-based observations. As depicted in Figure 5.3, the relative posterior trends suggest even stronger emission reductions compared to the inventory. However, as follows from the analysis of the differences between posterior and GCP seasonal cycles (Sect. 5.2.2.1), only the summer half-years show significantly lower emission values compared to the GCP summer fluxes. This is visible in the run of the posterior curve based on the F1T filter shifting the 23/24 winter flux closer to the GCP inventory and summer 2023 and 2024 fluxes even further away from it.

A linear regression analysis of emission estimates starting from 01.11.2021 revealed a consistent negative trend. For absolute fluxes, the GCP model estimated a decrease of –1.6±0.1 Tmol/yr, whereas the F1T and F0.5T posteriors showed a stronger decline at –2.1±0.1 Tmol/yr and –2.2±0.1 Tmol/yr, respectively. This pattern is consistent when looking at relative trends. The GCP showed a –5.1%/yr decrease compared to –6.9%/yr

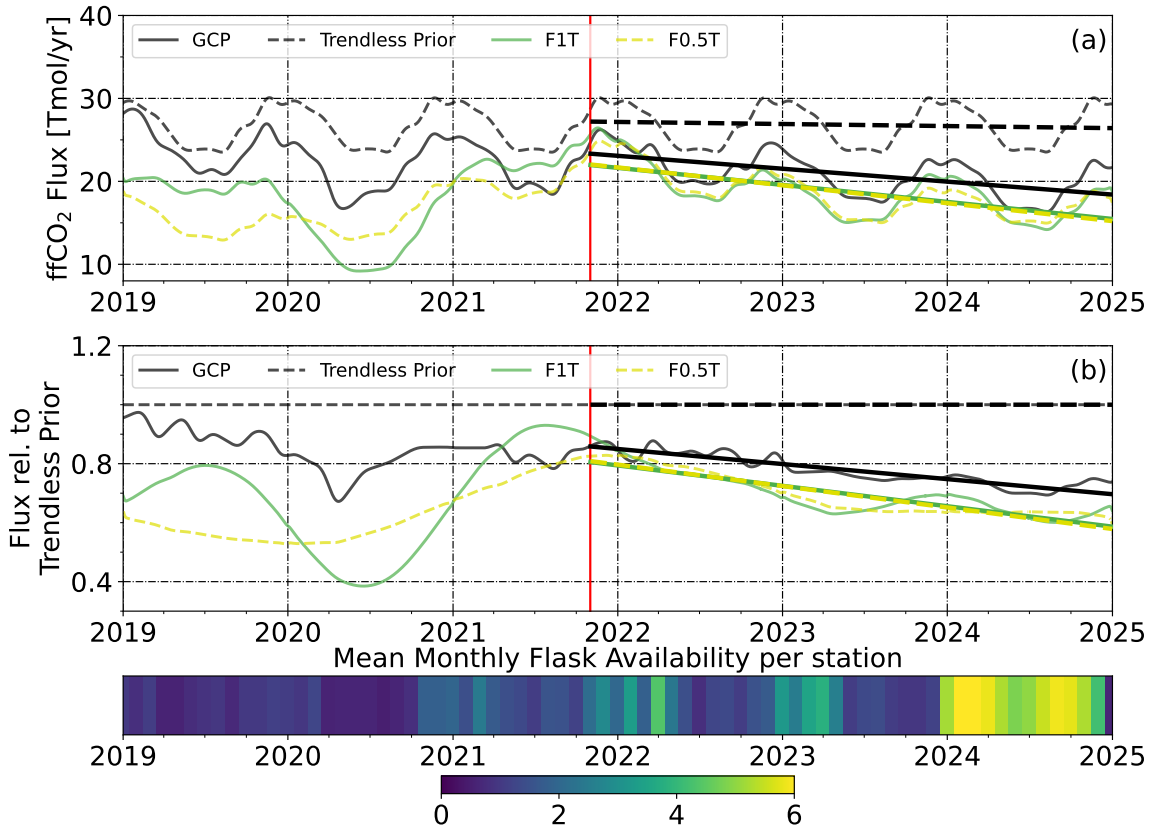

Figure 5.3. Results of the inter-annual trend analysis for the flask-based inversions. Panel (a) shows the absolute and panel (b) the relative to the Trendless Prior $\Delta ffCO_2$ fluxes in the Germany+ domain. The GCP inventory (black solid line) and the Trendless Prior (used as input, black dashed line) are shown alongside posterior results for F1T (green solid line) and F0.5T (yellow dashed line) filters. The straight lines show linear regressions of the correspondingly colored flux curves. Red line indicates the main investigation time frame (from 01.11.21). The lower section of the figure depicts the corresponding timeline of mean monthly flask sample number per station as a heat map.

(F1T) and –7.2%/yr (F0.5T). It's important to note that these deviations from the GCP's trend estimate are within the expected range of variation caused by measurement errors over three-year period, which are estimated to be approximately ∼0.6 Tmol/yr for F0.5T and ∼7 Tmol/yr for F1T (Sect. 4.6.2).

These findings demonstrate that flask-based inversions could be a viable method for the estimation of inter-annual trends in the future, provided that the previously discussed potential bias in the ^{14}C background estimation (see Sect. 5.2.2.1) is corrected for a longer time series.

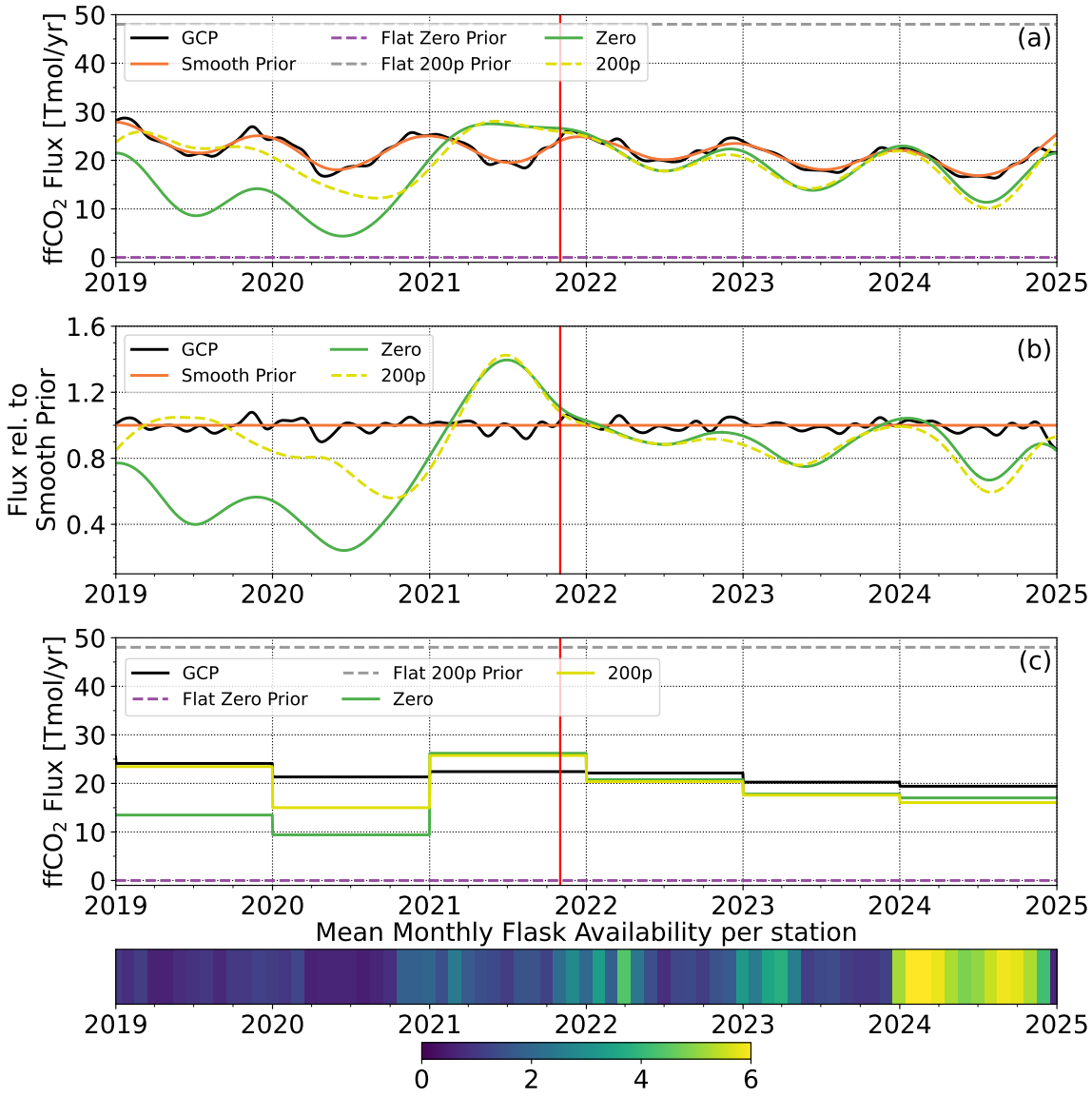

Figure 5.4. Results of the analysis of absolute and annually aggregated emission fluxes for the flask-based inversions in the Germany+ domain. Panel (a) shows the absolute and panel (b) the relative to the Smooth Prior ΔffCO$_2$ fluxes. Panel (c) shows yearly aggregated emissions from the panel (a). The Flat Zero and the Flat 200p Priors (used as inputs, respectively, purple and grey dashed lines) as well as the Smooth Prior (orange solid line) are shown alongside posterior results (green solid and yellow dashed lines). Red line indicates the start of main investigation time frame (from 01.11.21). The color bar indicates the mean monthly availability of integral-based estimates per station.

5.2.2.3 Absolute and annual mean ΔffCO$_2$ emissions derived from flask-based inversions

This sensitivity test for absolute and annually aggregated mean fluxes for the Germany+ domain for the flask-based ΔffCO$_2$ data set uses the Flat Zero and Flat 200p Priors (Tab.

Table 5.3. Annual mean $\Delta ffCO_2$ emissions of flask-based posteriors relative to the GCP inventory for the full years in the main investigation time frame starting on 01.11.2021.

Prior	2022	2023	2024
Flat Zero	0.94	0.88	0.88
Flat 200p	0.92	0.87	0.83

3.1), each with a prior uncertainty of 40%. The inversion system is thus provided with minimal prior information to ensure data-driven posteriors. Figure 5.4 depicts the resulting posterior $\Delta ffCO_2$ fluxes compared to the GCP fluxes.

As already discussed in the analysis of the seasonality (Sect. 5.2.2.1), the summer minima are significantly more pronounced in the posteriors compared to the GCP, suggesting a much steeper mean downward trend (Fig. 5.4, panel (a)). However, the magnitude of the winter emissions generally corresponds to the GCP fluxes.

The RMSD between the posteriors and the Smooth Prior (to avoid the influence of the short-term variability in the GCP, see Sect. 3.2.4) relative to the latter is equal to 0.10 for the Zero posterior and 0.12 for the 200p posterior for the main investigation period. These values are on the upper edge of the GCP deviation tolerance of approximately 0.10 that could be caused by measurement errors in the order of 1.5 ppm (Sect. 4.6.3). Nevertheless, there are hints to a seasonal dependence of the observed mismatch, that could be attributed to a potential bias in the ^{14}C background in 2022 and 2023 (Sect. 5.2.3.1).

Table 5.3 shows the numeric values of the mean annual posterior fluxes relative to the GCP inventory. For 2022, where the $\Delta^{14}C$ background should still be closely aligned with the MHD background, the flask-based $\Delta ffCO_2$ observations reproduce the annual mean emissions from the GCP inventory to 94% or 92%, respectively. As expected from the findings in the previous subsections, the increasing underestimation of summer emission minima consequently leads to an increasing underestimation of annual mean flux values.

Within the main investigation period, the Flat Zero and the Flat 200p priors show the highest discrepancy in the year 2024, the year with the highest data density (Fig. 5.4). As discussed in Section 4.6.3, such an enhanced sensitivity to different model inputs is probably caused by some systematic but yet unknown feature of the inversion system that is independent of the origin of the data, as it was observed in the posteriors based both on synthetic and real data. Based on the correlation observed here, though rather speculative, I might suggest that the high data density in 2024 may interact with the model in a way that it becomes more sensitive to this systematic effect compared to the other investigated periods.

5.2.3 Uncertainty investigation for the flask-based $\Delta ffCO_2$ inversions

5.2.3.1 Sensitivity of the posterior fluxes to the European $\Delta^{14}C$ background estimate

The European $\Delta^{14}C$ background plays a pivotal role in calculating the $\Delta ffCO_2$ values (Sect. 3.1.1). In this section I investigate how sensitive the inversion results are to the biases in the $\Delta^{14}C$ background estimation. There are two main reasons why the actual atmospheric $\Delta^{14}C$ background may deviate from the background estimated in Sect. 3.1.1, especially during the time period covered by the flask observations.

First, the year 2020 marked the most recent solar minimum (Clette & Lefèvre, 2015). During such periods of low solar activity, the reduced solar magnetic shielding allows a greater flux of galactic cosmic rays to penetrate the atmosphere of the Earth. This leads to an increased production of $\Delta^{14}C$ in the stratosphere. Given the typical stratosphere–troposphere exchange timescale of approximately two years (Holton et al., 1995), an effect in atmospheric $\Delta^{14}C$ concentrations and seasonality is expected to manifest around 2022.

Second, $\Delta^{14}C$ measurements at the Mace Head (MHD) station on the west coast of Ireland – the single most important station to quantify the European $\Delta^{14}C$ background estimate – were discontinued in spring 2022 due to safety-related issues at the sampling tower.

Unlike measurement errors, which affect specific samples, or instrument-induced systematic errors, which may affect individual sites, biases in the background influence the entire measurement network simultaneously. To analyze the influence of the background biases, the $\Delta^{14}C$ background curve was shifted by $\pm$2‰. This range approximately corresponds to the span of the the $\Delta^{14}C$ background uncertainty from the ccgcrv smoothing routine (Sect. 3.1.1). The flask-based $\Delta ffCO_2$ inversion using the Flat Zero prior with a prior uncertainty of 40% was repeated using two biased $\Delta^{14}C$ backgrounds. The other inversion parameters were not changed.

The positive offset of the $\Delta^{14}C$ background resulted in higher $\Delta ffCO_2$ values, leading to a higher flux estimation for the AoC (Fig. 5.5). The negative shift in the $\Delta^{14}C$ background had the opposite effect. On average, these changes to the background led to a 20% increase or decrease in the flux estimation during the investigation period relative to the mean flux. These changes are significant and much higher than the inventory uncertainty ($\sim$1-5%, Sect. 3.2.3), and greater than the deviations that could be induced by the typical measurement errors (e.g. Sect. 4.6.3). This indicates that the inversion system is highly sensitive to biases in the background and that the accurate background $\Delta^{14}C$ data are absolutely crucial for robust flux estimations based on the regional isotope budget approach.

The inversion system is symmetrically affected by the $\pm$2‰ biases, as can be seen from the good agreement of the +2‰ posterior compared to the -2‰ posterior, which was reflected along the x-axis after subtraction of the Standard posterior (Fig. 5.5, panel (b)). Dis-

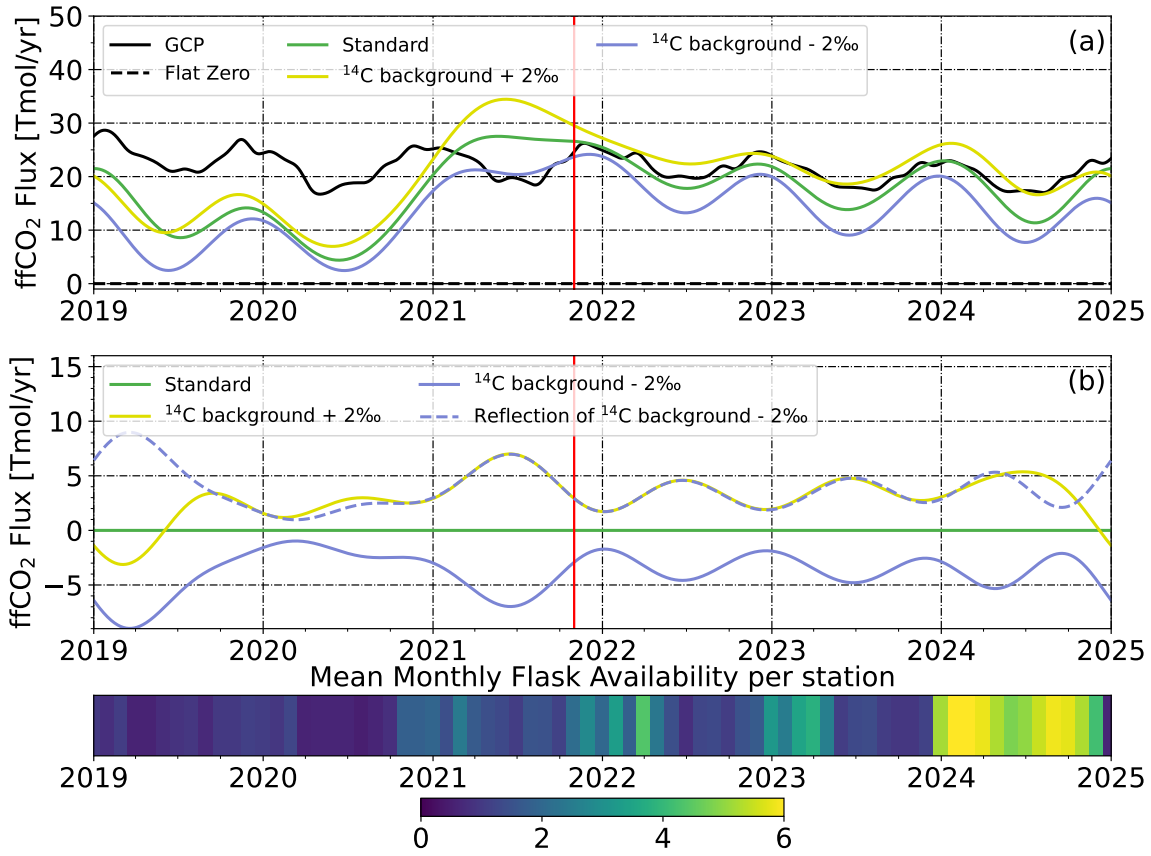

Figure 5.5. Results of the sensitivity analysis of the flask-based inversions to the $\Delta^{14}C$ background estimate. The GCP inventory (black solid line) and the Flat Zero Prior (used as input, black dashed line) are shown alongside posterior results based on the standard $\Delta^{14}C$ background (green line), the background with a $\Delta^{14}C$ offset of 2‰ (yellow line) and -2‰ (blue line). Red line indicates the main investigation time frame starting on 01.11.2021. Panel (a) shows the absolute flux values. Panel (b) depicts the difference between the Standard posterior and the flux estimations based on the shifted background. The blue dashed line shows the reflection of the blue line along the x-axis.

crepancies arise only at the boundaries of the investigated period and during the period with lower data density.

The key observation for flask-based posterior fluxes is that a constant bias in the $\Delta^{14}C$ background leads to seasonally varying deviations in flux space. A $\pm$2‰ shift in the $\Delta^{14}C$ background results in symmetric flux changes of 10-15% in winters and 25-30% in summers, relative to mean flux. This increased summer sensitivity arises from the lower absolute $\Delta ffCO_2$ signals during these months, making a given change in $\Delta^{14}C$ proportionally more impactful, even when the absolute ppm change is similar. This finding is consistent with the previously observed pattern that flask-based posterior fluxes match the GCP inventory more closely in winter than in summer. At the same time, the posterior with +2‰ back-

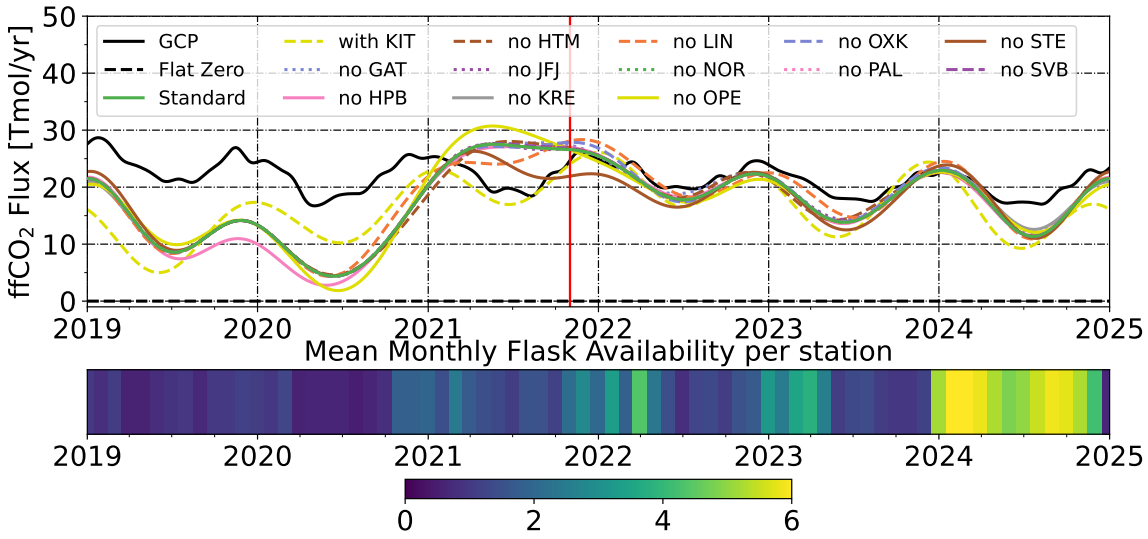

Figure 5.6. Results of the sensitivity analysis of flask-based ΔffCO$_2$ inversions on individual measurement stations. The panel shows posterior ΔffCO$_2$ fluxes for the Germany+ domain based on flask-derived estimates using a Flat Zero prior with 40% prior uncertainty. The standard posterior (green solid line) is compared to leave-one-out setups where individual ICOS stations were excluded (other lines), as well as a configuration including the KIT site (yellow dashed line). The GCP inventory (black solid line) and the Flat Zero prior (black dashed line) are shown for reference. The color bar indicates the mean monthly availability of flask-based ΔffCO$_2$ data per station.

ground bias closely tracks the GCP emissions for the Germany+ domain in the last years of the investigated period (considering the ”overshoot” effect from the Flat Zero Prior, Sect. 4.6.3). This can potentially suggest that there might be a systematic bias in the European ^{14}C background estimate constructed in Section 3.1.1 after measurements at the MHD site were discontinued.

5.2.3.2 Influence of individual stations on the flask-based posterior

To further investigate the inversion results, I evaluated the influence of individual ICOS measurement stations on the posteriors. Significant changes in posteriors upon station removal or addition may indicate an unstable inversion system. Such reliance on a single station suggests suboptimal station density or systematic differences (e.g., erroneous atmospheric transport, biased measurements, representation errors).

Figure 5.6 displays posteriors calculated using the Flat Zero Prior (40% prior uncertainty) and the F2T time correlation filter and various sets of measurement stations. The Standard posterior (green solid line) utilizes the standard flask-based inversion station set (Tab. 2.1). Other inversions excluded one of the stations from the core set or included the

KIT station.

Generally, posteriors with different station sets closely resemble the standard case, indicating that the measurement data from different stations do not contradict each other and provide some overlap in the data coverage both spatially and temporally. The current station set establishes a stable inversion system with reliable posteriors. However, minor deviations from the Standard posterior are observed.

Including the KIT station resulted in lower summer flux values and an earlier summer minima in 2023 and 2024. Due to the higher number of flask samples and generally higher $\Delta ffCO_2$ signals, KIT provided additional constraint on the seasonal cycle during periods of low data coverage (before 2022), leading to high posterior dependence on this station. As mentioned previously, $\Delta^{14}C$ contamination of unknown magnitude in immediate proximity to the KIT station is suspected due to the incineration of slightly radioactive waste at the KIT campus which might sometimes contain ^{14}C . The resulting posterior with KIT included supports this assumption as the $\Delta^{14}C$ source would mask a portion of the $\Delta ffCO_2$ signal leading to the lower estimated fluxes (as visible e.g. in Fig. 5.6 in summer 2023 and 2024) and validate the exclusion of the KIT site from the core station set.

At the start of the main investigation period, i.e. from Nov. 2021 to Jun. 2022, STE significantly influences posterior calculations, leading to lower emissions upon its removal. Located centrally in the Germany+ domain, STE and LIN stations constrain a vital area and provide a high sample number during these periods (Fig. 4.2). These features could explain the high reliance of the posteriors on these sites. However, considering that these deviations are minor compared to expected deviations from 1.5 ppm measurement errors (e.g. Sect. 4.6.1), these stations were retained to maintain data constraint. This evaluation highlights the need for more stations with high sample numbers.

5.2.4 Flask-based Bayesian inversions with unmodified GCP and EDGAR priors

This section presents the results from two Bayesian inversions using flask-based $\Delta ffCO_2$ estimates and state of the art priors and accounting for the findings in the previous sections. The GCP inventory with daily resolution and the hourly EDGAR emission inventory served as prior information (Sect. 3.2.3). A 4% prior uncertainty (relative to the GCP fluxes) was assigned to both unmodified inventories, consistent with reported uncertainties for larger domains (Janssens-Maenhout et al., 2019; Jones et al., 2022). Given the previous observations of noticeably low summer posterior fluxes (e.g., Sect. 5.2.2.1) and their susceptibility to biases in the $\Delta^{14}C$ background (Sect. 5.2.3.1), this analysis focused on winter months (November - February). The 2024/25 winter was excluded to avoid boundary effects.

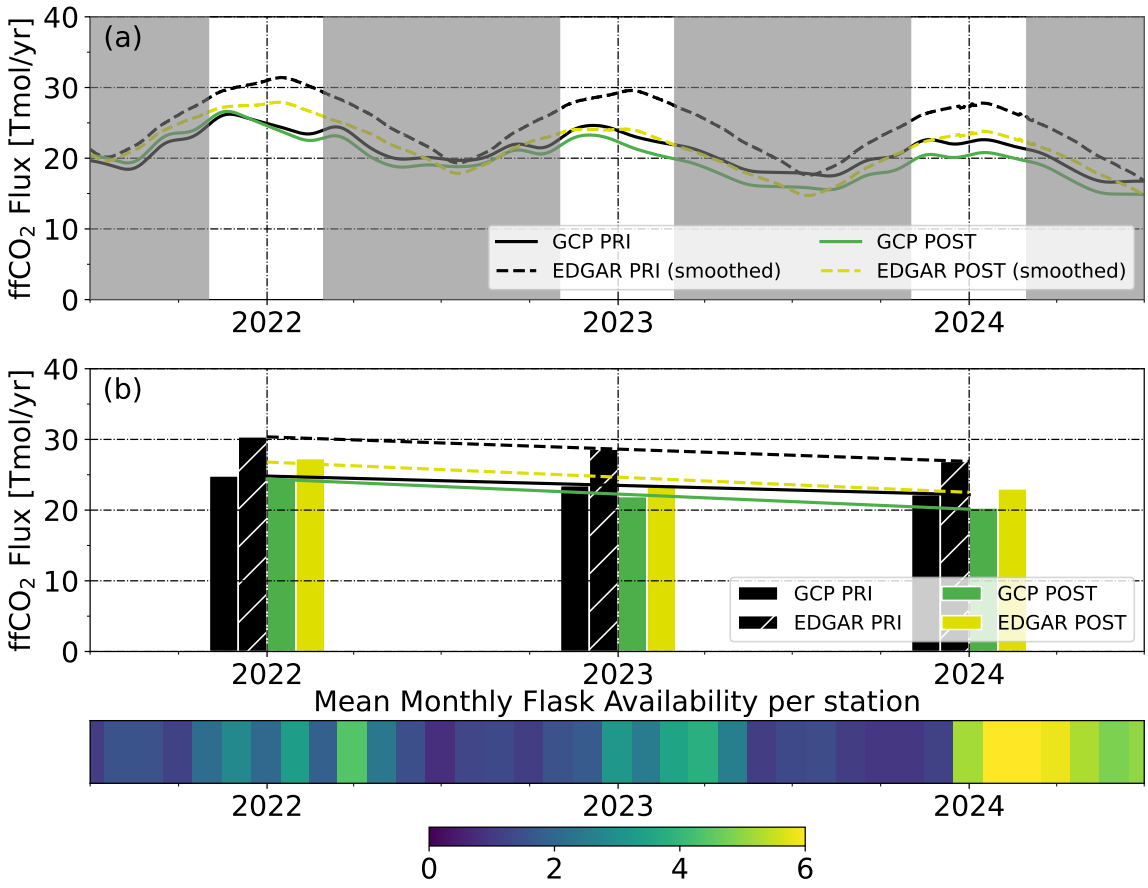

Figure 5.7. Wintertime fossil fuel emissions in the Germany+ domain from Bayesian inversions using flask-based $\Delta ffCO_2$ estimates and two different prior inventories: GCP (solid lines/bars) and EDGAR (dashed lines/hatched bars). Panel (a) shows smoothed daily (GCP) and hourly (EDGAR) prior and posterior fluxes; non-winter periods are shaded in gray. Winter periods are defined as November to February. Panel (b) displays the corresponding mean fluxes per winter season. The trends estimated from the mean winter emissions are shown as solid and dashed lines in the respective colors. The color bar indicates the mean monthly availability of flask-based estimates per station.

Figure 5.7 illustrates the prior fluxes as well as the inversion results for the period with sufficient data density (Sect. 4.3) for the Germany+ domain. The grayed out periods in panel (a) indicate the summer periods which are not included in the evaluation of this experiment. Table 5.4 outlines the numerical values of these inversions. The EDGAR inventory shows considerably larger fossil fluxes in winter compared to the GCP, while both agree in summer. Both posteriors indicate lower emissions than their corresponding priors. The GCP posterior shows emissions for Germany+ domain 0.2-1.9 Tmol/yr (1-9%) lower than the GCP inventory. The EDGAR posterior flux reductions are greater, ranging from 3.1 to 4.9 Tmol/yr (10-17%) relative to EDGAR. The difference between the two posterior flux es-

Table 5.4. Left side: Winter mean $\Delta ffCO_2$ emissions (in Tmol/yr) and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective flask-based Bayesian inversion posteriors. Right side: Relative deviations of winter mean $\Delta ffCO_2$ emissions from the GCP for the corresponding winter seasons. Winter periods refer to November through February of the indicated winter season.

Flux type	Absolute fluxes, [Tmol/yr]			Deviation from GCP, [%]		
	2021/22	2022/23	2023/24	2021/22	2022/23	2023/24
GCP Prior	24.8(0.9)	23.5(0.8)	22.2(0.8)	0(4)	0(4)	0(4)
GCP Posterior	24.6(0.4)	21.9(0.5)	20.3(0.5)	-1(2)	-7(2)	-9(2)
EDGAR Prior	30.4(1.0)	28.6(1.0)	26.9(0.9)	23(4)	22(4)	21(4)
EDGAR Posterior	27.3(0.5)	23.7(0.5)	23.0(0.6)	10(2)	1(2)	4(3)

timates is notably smaller than between corresponding prior fluxes. For the investigated winters, the GCP-EDGAR inventories difference ranges from 21-23% (relative to GCP), while the posterior difference is 8-13%, with EDGAR prior and posterior showing higher emissions than their GCP counterparts. Even with realistic prior uncertainties, the inversion significantly reduces posterior fluxes beyond 4% uncertainty, particularly for the EDGAR inventory, suggesting strong data constraints and data-driven results. It remains uncertain to what extent the lower posterior emissions compared to the two priors are still affected by a potentially biased ^{14}C background. Although this bias has a much lesser extent in the winters than during the summers, 1.5 Tmol/yr (2022/23) and 1.9 Tmol/yr (2023/24) emission differences between GCP prior and posterior could be explained by a ^{14}C background bias of 1-2‰ (see Sect. 5.2.3.1). Nevertheless, one can conclude that the optimized fluxes correspond more closely with GCP than with EDGAR, suggesting that EDGAR probably overestimates winter emissions.

The posterior uncertainty was reduced by 0.3 to 0.5 Tmol/yr, or 30-50% relative to the prior uncertainty (Tab. 5.4). Only for the winter of 2022/23 the mean posterior emissions were in the 3σ confidence range of each other, suggesting that the initial discrepancy between the priors was not completely eliminated by the inversion.

Emission reduction trends in priors and posteriors were also compared as slopes of linear fits for mean winter emissions (Fig. 5.7, panel (b)). Over three years, GCP suggests a yearly 1.32(0.03) Tmol/yr flux reduction in the Germany+ domain, while EDGAR suggests a 1.73(0.01) Tmol/yr annual reduction. The uncertainty of the trend in both inventories is very low ranging between 0.5 and 2%, due to a nearly constant emission reduction rate reported by the inventories. The posteriors suggest similar but slightly higher emission reduc-

tion trends: 2.1(0.3) Tmol/yr for GCP posterior and 2.1(0.9) Tmol/yr for EDGAR posterior. The uncertainties of the posterior emission trends are around 15% and 40%, respectively.

Three years might be too short period for a reliable validation of mean winter emission trends. Variations in the posterior flux estimates, e.g., due to uncertainties in the underlying $\Delta ffCO_2$ data, further complicate this. The higher trend of emission reduction observed in the posteriors, compared to the priors, could also be attributed to a potential bias in ^{14}C background estimation, even for winter fluxes. Overall, data-driven posteriors suggest lower emissions in the Germany+ domain, with absolute emission values closer to the GCP inventory.

Chapter 6

^{14}C calibrated CO-based ΔffCO$_2$ inversions

6.1 Introduction and motivation

As discussed in Section 2.1, ^{14}C is a potent proxy for the identification of fossil CO_2 . However, it has limitations. ΔffCO$_2$ estimates from ^{14}C measurements suffer from low temporal resolution (typically one data point per week or month) and sparse spatial coverage due to the resource-intensive nature of Δ^{14}C sample collection and analysis. Continuous ^{14}C observation techniques are in the early stages, with current working prototypes exhibiting approximately six times lower accuracy than conventional methods (Galli et al., 2013; McCartt et al., 2015; Fleisher et al., 2017; Delli Santi et al., 2022; Jiang & McCartt, 2024), bringing them slowly toward a useful option for ground-based measurement networks, but up to now no field deployment has been carried out.

Additional proxies that are co-emitted during fossil fuel combustion, the so-called co-emitted species, can provide additional or denser temporal ΔffCO$_2$ information, bridging the gaps between individual ^{14}C samples or serving as Δ^{14}C-calibrated ΔffCO$_2$ proxies. These species are often more readily available from continuous, in situ, or remote sensing measurements. Prominent examples include direct measurements of NO_x and CO (e.g., Palmer et al., 2006; Levin & Karstens, 2007; Lopez et al., 2013; Konovalov et al., 2016; Goldberg et al., 2019; Jäschke, 2021; Rosendahl, 2022; Maier et al., 2024a; Scarpelli et al., 2024) or combined O_2 and CO_2 proxies such as Atmospheric Potential Oxygen (APO) (Pickers et al., 2022; Rödenbeck et al., 2023), which can be used to deduce ΔffCO$_2$ emissions from concurrent oxygen consumption.

During the past two decades, efforts to continuously measure fossil CO_2 using CO as a co-emitted proxy have mainly focused on urban environments and are well-established (e.g., Gamnitzer et al., 2006; Levin & Karstens, 2007; Vogel et al., 2010, 2018). However,

ICOS atmospheric Class 1 stations (ICOS RI, 2020) are located to measure predominantly regional background air masses, resulting in significantly lower signals compared to urban environments. (Maier et al., 2024a) demonstrated for one station (OPE) that, given a sufficient data density, this approach is also applicable to an ICOS station. Furthermore, in his study, the application of CO-based $\Delta ffCO_2$ data significantly improved inversion results for the estimation of semi-urban Rhine valley emissions compared to flask-based inversion (Maier et al., 2023).

This provides motivation to investigate whether CO-based $\Delta ffCO_2$ estimates can also be applied to the entire ICOS network and used to constrain the inversions. However, as outlined in previous studies and most explicitly in Maier et al., 2024a, CO is not an ideal co-emitted $\Delta ffCO_2$ proxy. The ideal co-emitted proxy has a constant emission ratio to $\Delta ffCO_2$ for all combustion processes, no non-fossil sources, and a similar atmospheric residence time to CO_2 . None of these properties are true for CO. However, in the urban and semi-urban use case of the CO proxy, the shorter CO lifetime and the non-fossil sources lead to a non-dominant additional source of uncertainty, and the varying $\Delta CO/\Delta ffCO_2$ emission ratios average out over longer periods of time. If the atmospheric ratio is derived from representatively collected flask samples at a given station, it captures the station-specific effective $CO/\Delta ffCO_2$ ratio, with its associated uncertainty reflecting the observed variability.

The following chapter investigates the benefit of extending the application of ^{14}C - calibrated CO-based $\Delta ffCO_2$ estimates to the ICOS network (Sect. 2.3). I determine the CO-based $\Delta ffCO_2$ data for all suitable ICOS sites and estimate the fossil CO_2 emissions in the Germany+ domain (Sect. 6.2). The ”Evaluation Matrix” experiment set (Sect. 6.3.2) is thereby kept identical to those used for flask-based (Sect. 5) and integral-based (Sect. 7) inversions to allow for a comparison of all three $\Delta ffCO_2$ estimates in Sect. 8. Additionally, I investigate the sensitivity of the CO-based inversions to variation of several core parameters (Sect. 6.3.3). The final step involves two Bayesian inversions, performed with the unmodified GCP and EDGAR prior fluxes and their respective realistic prior uncertainties (Sect. 6.3.4).

6.2 CO-based $\Delta ffCO_2$ estimations

I calculated ^{14}C-calibrated CO-based $\Delta ffCO_2$ data using methods from (Maier et al., 2024a). Flasks with determined $\Delta ffCO_2^{flask}$ values were analyzed for excess CO (ΔCO^{flask}) against the marine background (MHD, see Sect. 6.2.1). I then calculated a station-specific average atmospheric ΔCO^{flask} / $\Delta ffCO_2^{flask}$ ratios ($\langle R_{flasks} \rangle$) from all afternoon samples, neglecting any spatiotemporal variability in the ratios (except for several stations where wind dependent ratios need to be applied, see Sect. 6.2.2.2). These ratio are applied to hourly in

situ excess CO concentrations (against the same background) to estimate hourly CO-based $\Delta\mathrm{ffCO}_2^{hourly}$:

$$\Delta\mathrm{ffCO}_2^{hourly} = \frac{\Delta\mathrm{CO}^{hourly}}{\langle \mathrm{R}_{flasks} \rangle} \tag{6.1}$$

6.2.1 CO background

In addition to the need to determine representative atmospheric $\Delta\mathrm{CO}/\Delta\mathrm{ffCO}_2$ ratios, the CO-based ΔffCO$_2$ approach requires a representative background value of CO against which the local CO enhancement can be determined. As seen for the effects of the ^{14}C background, biases in the estimation of the European background have significant effects on the ΔffCO$_2$ estimates.

For the determination of the effective atmospheric $\Delta\mathrm{CO}/\Delta\mathrm{ffCO}_2$ ratio it is crucial that the CO and ^{14}C excesses are determined with respect to the same representative European background. Thus, similar to the ^{14}C background (Sect. 3.1.1), the marine sector of the MHD station was chosen to estimate the CO background concentration of clean air masses prior to their arrival in Europe. However, in summer the atmospheric lifetime of CO (approximately two months (Khalil et al., 1999)) is considerably shorter than that of ^{14}C or CO$_2$ and is comparable to hemispheric mixing times (Schlesinger & Bernhardt, 2020). Therefore, unlike ^{14}C , the CO background was derived solely from MHD CO data, rather than using a hemispheric trend. Unlike (Maier et al., 2024a), who based their CO background estimate on the weekly MHD CO flasks from the NOAA measurement network (Petron et al., 2022), I used in situ CO measurements collected by the University of Galway (Martin, 2025) with applied marine sector wind selection criteria. I chose the selection criteria to match those for MHD ^{14}C integral sampling (wind speed >4 m/s, sector 190°-300°).

From this selected data set, daily afternoon averages (12h-16h) were calculated and smoothed using the same ccgcvr fitting routine that was used for the Δ^{14}C data (Press et al., 1988; Thoning et al., 1989). I choose the same fitting parameters for the CO as for the Δ^{14}C background, except for short-term variability filtering. Due to its short lifetime and higher variability, sub-annual CO variations need to be retained.

To determine the optimal short-term frequency cut-off for smoothing the marine MHD afternoon CO record, a variational approach was employed. The idea was to vary the smoothing window between 10 and 200 days and evaluate for each setting how well the resulting MHD CO background correlates with the lower envelope of CO observations from all available ICOS stations. I estimated the lower envelope of the ICOS stations as the 10th percentile of in situ CO data in a moving 10 day window. Panels (a) and (b) in Figure 6.1 depict the CO in situ records for two ICOS stations Lindenberg (LIN, panel (a)) and Pal-

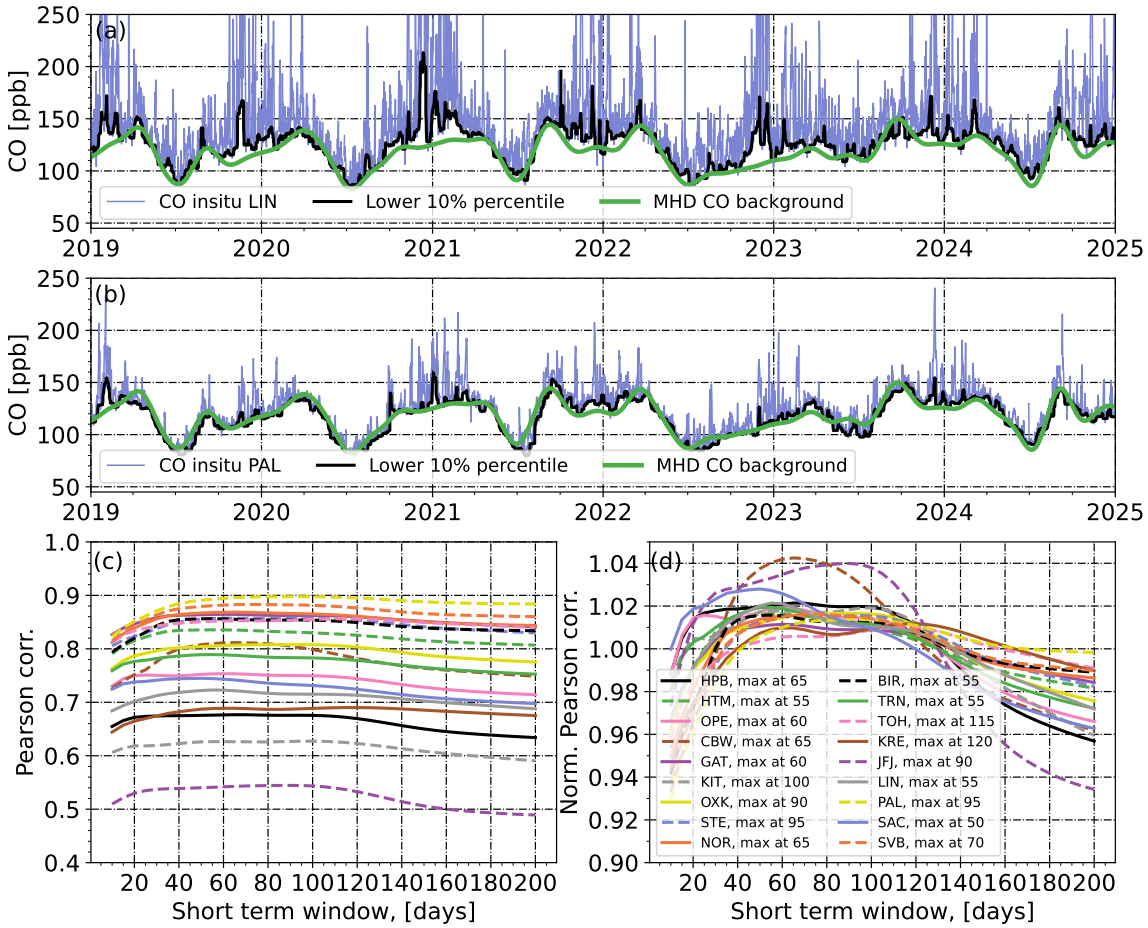

Figure 6.1. Results of the optimal short-term frequency cut-off determination for the MHD CO background. Panels (a) and (b) display in situ CO data (blue line) from LIN and PAL stations, respectively, alongside with their lower 10th percentile of in situ CO data over a 10-day moving window (black line) and the MHD CO background curve with a 75-day short-term variability cut-off (green line). Panels (c) and (d) share a common legend. Panel (c) shows the absolute Pearson correlation, while Panel (d) shows the correlation normalized to its mean, between the ICOS station's lower 10th percentile and the MHD CO background curve for various short-term variability cut-offs. The cut-off yielding maximal correlation is noted in the legend.

las (PAL, panel (b)) together with their lower envelope estimates given by the black lines in both panels. The optimal cut-off frequency is identified as the one yielding the highest absolute Pearson correlation across all stations in the network (see Fig. 6.1, panel (c)). Depending on the station's general CO signal levels, the correlations range between 0.5 and 0.9, with higher values typically observed at more remote or less polluted sites. The normalized Pearson correlations (Fig. 6.1, panel (d)) indicate a plateau in the correlation curves for most stations, ranging from approximately 40 to 110 days. The average cutoff over all stations of 75 days was selected for the final CO background curve and is depicted in green

in panels (a) and (b) in Figure 6.1.

Before 2024 MHD was not a classified ICOS station and therefore did not use calibration cylinders from the ICOS FCL. As a result, the MHD CO measurements had an offset compared to the international WMO CO X2014A scale. I estimated this offset to be constant at 15.16 ppb using MHD CO flask measurements from the NOAA network (Schuldt et al., 2025), which are also reported on the WMO CO X2014A scale (NOAA Global Monitoring Laboratory, 2024). All MHD in situ CO data have been corrected for this offset prior to the construction of the MHD CO background curve (solid green lines in Fig. 6.1, panels (a) and (b)).

The residual standard deviation of the in situ CO concentrations around this constructed MHD CO background with a cut-off frequency of 75 days is 10.3 ppb. Similar to the ^{14}C background (Sect. 3.1.1), additional uncertainty of the background estimate comes from unsuitability of the MHD estimate to serve as background for air masses entering the European domain through all domain boundaries other then western. The $\Delta ffCO_2$ representativeness uncertainty was estimated as 0.28 ppm by (Maier et al., 2024a). For CO, they recalculated it using the CO/$\Delta ffCO_2$ emission ratio of roughly 18 ppb/ppm in 2020 for the Eastern boundary taken from the TNO emission inventory, resulting in an additional contribution to the CO background uncertainty of 0.28 ppm · 18 ppb/ppm = 5.04 ppb. Together with the residual CO standard deviation, this results in a final error estimate of the CO background of 11.5 ppb.

6.2.2 ΔCO to $\Delta ffCO_2$ ratios

6.2.2.1 ^{14}C calibrated $\Delta CO/\Delta ffCO_2$ ratios

The ΔCO / $\Delta ffCO_2$ ratios were calculated using a modified version of the method used by (Maier et al., 2024a).

ΔCO and $\Delta ffCO_2$ were determined as excess concentrations above their corresponding background (Sect. 3.1.1 and 6.2.1) from all suitable flask samples. $\Delta ffCO_2$ data from flasks was calculated according to the procedure described in Sect. 3.1. Only flasks collected during afternoon hours (11:00-18:00) were used for the $\Delta CO/\Delta ffCO_2$ ratio determination, with the exception of JFJ station, for which only nighttime samples (22:00-4:00) were considered since it is a high-elevation mountain site. For HTM, STE, and LIN stations, the respective flasks were separated into two sets based on prevailing wind sectors during sampling, and $\Delta CO/\Delta ffCO_2$ ratios were calculated independently for each of the wind sectors (Sect. 6.2.2.2), accounting for heterogeneous distributions of CO and $\Delta ffCO_2$ sources.

Maier et al., 2024a recommend using the slope of the regression line between ΔCO and $\Delta ffCO_2$ for unbiased $\langle R_{flasks} \rangle$ estimates (black dashed lines in Fig. 6.2 or in Sect. A.3), in

contrast to methods considering error-weighted means or medians. The regression line for each site was calculated using the weighted total least squares algorithm (Krystek & Anton, 2007; Wurm, 2021), which accounts for uncertainties in both ΔCO and ΔffCO$_2$. The regression intercept was set to zero, assuming the MHD backgrounds for CO and ^{14}CO$_2$ are representative for ICOS sites under well-mixed, clean air conditions.

In addition to the Maier et al., 2024a method, outliers were removed from the flask pool during the calculation of $\langle \mathrm{R}_{flasks} \rangle$ using the Interquartile Range (IQR) method, also known as Tukey's Fences (Tukey et al., 1977). First, $\langle \mathrm{R}_{flasks} \rangle$ was calculated as described above. This ratio was then used to determine the CO-based ΔffCO$_2$. The 25th (Q_1) and 75th (Q_3) percentiles of the differences between ^{14}C-based and CO-based ΔffCO$_2$ estimates ($\Delta\Delta\mathrm{ffCO_2}$) were calculated for each station's entire flask pool. The IQR ($Q_3 - Q_1$) was then determined. Outliers were defined as observations outside the range

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$$

with k being some non-negative constant, usually 1.5 or 3, depending on the strictness of the desired flagging. Here, I used $k = 3$, which roughly corresponds to the 4.5σ criteria for Gaussian distributions, to ensure that only a small fraction of flasks were removed.

If outliers were detected, the sample furthest from the linear regression line was identified, considering applicable uncertainties. For each flask, the distance d to the regression line, taking into account individual uncertainties, was calculated as:

$$d_i = \frac{\sqrt{\sigma^2_{\Delta\Delta\mathrm{ffCO_2}} + \mathrm{RMSD}^2_{\Delta\Delta\mathrm{ffCO_2}}}}{|\Delta\Delta\mathrm{ffCO_2}|_i} \tag{6.2}$$

where $\sigma_{\Delta\Delta\mathrm{ffCO_2}}$ is the combined error of $\Delta\Delta\mathrm{ffCO_2}$ uncertainty (based on measurement errors), and $\mathrm{RMSD}_{\Delta\Delta\mathrm{ffCO_2}}$ is the root mean square deviation of $\Delta\Delta\mathrm{ffCO_2}$, representing variability of the data around the regression line between ^{14}C- and CO-based ΔffCO$_2$ estimates (e.g. panels (b) and (d) in Fig. 6.2. A low d value indicates a large difference between two ΔffCO$_2$ estimates relative to their errors. The flask with the lowest d value was removed, and $\langle \mathrm{R}_{flasks} \rangle$ recalculated. The whole procedure was iteratively repeated until no outliers were found by the IQR method. This process flagged only 3% of all flasks in the network (e.g. red crosses in Fig. 6.2).

The final ratios and statistics for each station are presented in Table 6.1. The RMSD between ^{14}C-based and CO-based ΔffCO$_2$ in flasks (Fig. 6.2, panels (b) and (d)) quantifies the uncertainty of the resulting CO-based ΔffCO$_2$ time series. This RMSD reflects the vertical scattering of the ^{14}C-based and CO-based ΔffCO$_2$ data around their regression line, which has by construction a slope of 1 (as the mean $\langle \mathrm{R}_{flasks} \rangle$ was used for CO-based ΔffCO$_2$ calculations). The resulting uncertainty represents the uncertainty of the mean

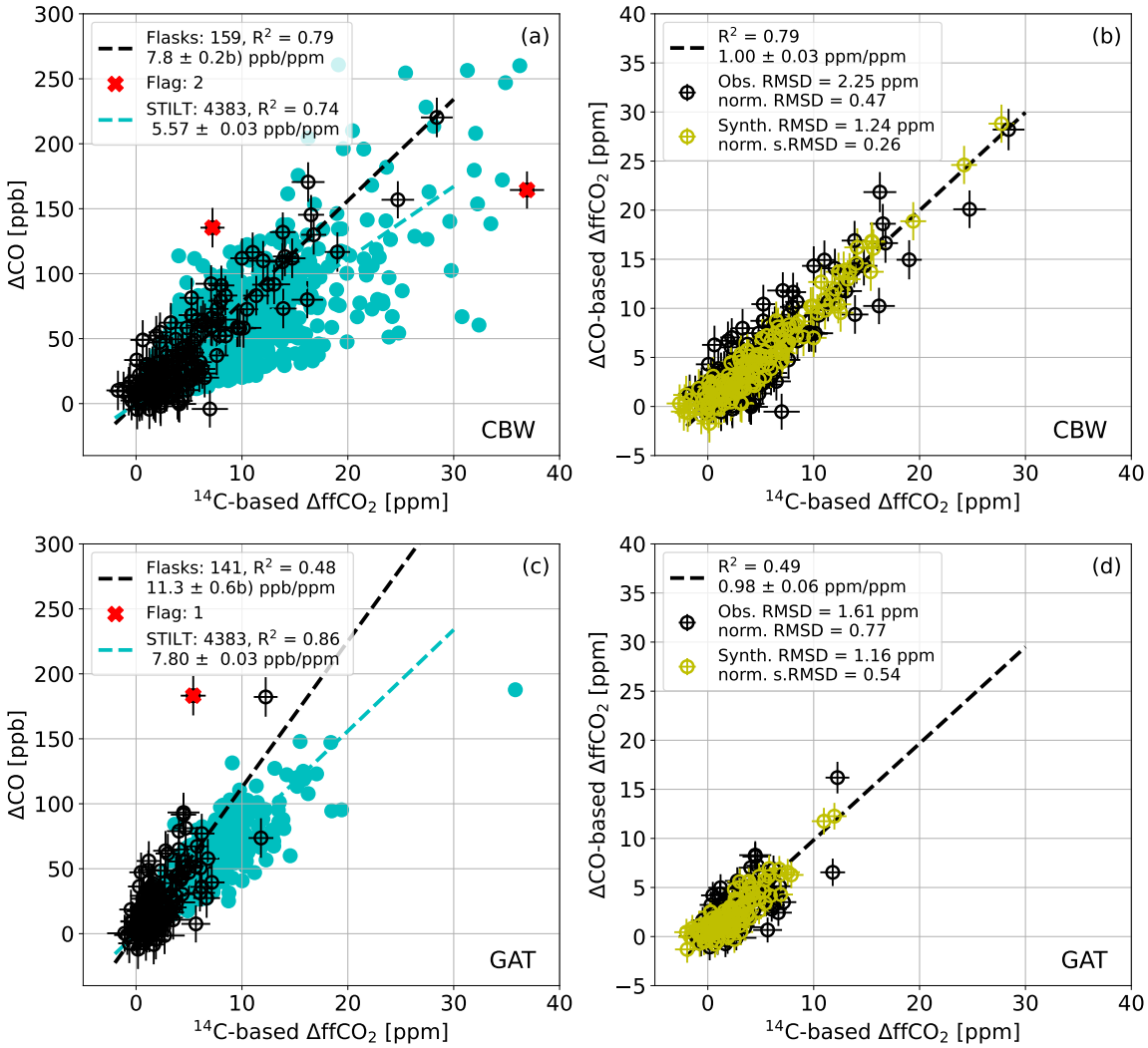

Figure 6.2. Calculation of ^{14}C-calibrated CO-based $\Delta ffCO_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (calculated slopes are denoted in legend). Panels (a) and (c) display ΔCO and $\Delta ffCO_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for CBW and GAT stations, respectively. Flagged flasks (red crosses) were excluded from the regression according to IQR outlier identification. In panels (b) and (d), ΔCO^{flask} was converted into $\Delta ffCO_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

$\langle R_{flasks} \rangle$, caused by combined variability in the ΔCO and $\Delta ffCO_2$ relationship within the station's catchment area and measurement errors.

Table 6.1. ΔCO and ΔffCO$_2$ ratios and their uncertainties derived from flask observations ($\langle R_{flasks} \rangle$) and STILT-TNO model estimates ($\langle R_{model} \rangle$, see Sect. 6.2.4) as well as coefficient of determination of the linear regression (R^2) for the flask-based ratios. Number of flask that were used for the calculation of the corresponding ratios as well as number of flagged flasks is listed. RMSD between ^{14}C-based and CO-based ΔffCO$_2$ in flasks is used as the uncertainty of the resulting CO-based ΔffCO$_2$ time series. $\langle R_{model} \rangle$ were derived from TNO inventory (Dellaert et al., 2019) and weighted with STILT footprints according to the procedure described in Maier et al., 2024a (see Sect. 6.2.4). (*) stations were excluded from inversions (see Sect. 6.2.3). (†) stations were used only in sensitivity studies (see Sect. 6.3.3).

Station	Num. flasks (flag)	$\langle R_{flasks} \rangle$ [ppb/ppm]	R^2	RMSD [ppm]	$\langle R_{model} \rangle$ [ppb/ppm]
CBW	159 (2)	7.81 (0.22)	0.79	2.25	5.57 (0.03)
GAT	141 (1)	11.25 (0.61)	0.48	1.61	7.80 (0.03)
HPB	251 (0)	10.67 (0.41)	0.38	2.02	11.16 (0.07)
HTM (all)†	191 (0)	14.30 (0.53)	0.51	1.76	9.43 (0.04)
HTM (wind sec. 1)	74 (0)	11.82 (0.57)	0.64	1.81	9.43 (0.04)
HTM (wind sec. 2)	49 (0)	20.11 (1.40)	0.52	1.57	9.43 (0.04)
JFJ*	63 (4)	5.71 (2.71)	-0.22	3.69	9.07 (0.04)
KIT†	228 (3)	8.04 (0.21)	0.65	2.52	7.10 (0.04)
KRE	114 (0)	18.85 (1.32)	0.32	1.17	9.80 (0.05)
LIN (all)†	248 (1)	14.82 (0.31)	0.49	2.67	6.05 (0.05)
LIN (wind sec. 1)	103 (4)	9.81 (0.35)	0.71	1.98	6.05 (0.05)
LIN (wind sec. 2)	69 (0)	20.91 (0.64)	0.45	2.54	6.05 (0.05)
NOR*	58 (2)	28.07(4.55)	-2.28	1.53	10.33 (0.04)
OPE	203 (2)	10.80 (0.37)	0.70	1.75	9.12 (0.05)
OXK*	73 (0)	11.09 (0.63)	0.11	2.53	8.93 (0.04)
PAL*	13 (0)	31.62 (20.88)	-1.95	0.87	10.59 (0.05)
SAC*	26 (0)	7.83 (0.46)	0.73	3.45	4.65 (0.02)
STE (all)†	236 (4)	10.18 (0.28)	0.43	2.40	7.28 (0.03)
STE (wind sec. 1)	36 (0)	20.30 (1.03)	0.22	2.46	7.28 (0.03)
STE (wind sec. 2)	125 (3)	8.63 (0.30)	0.54	2.34	7.28 (0.03)
SVB*	32 (0)	13.97 (3.26)	-0.55	1.67	10.05 (0.04)

6.2.2.2 Wind sector dependent ratios

A heterogeneous distribution of emission sources (e.g. transport, industry) in the station's catchment area can cause systematic patterns in the ΔCO and $\Delta ffCO_2$ ratios (e.g., Jäschke, 2021; Rosendahl, 2022), potentially biasing the mean $\langle R_{flasks} \rangle$ estimate and the resulting CO-based $\Delta ffCO_2$ data. To address spatial emission heterogeneities in the catchment areas of the ICOS sites, I developed a wind sector-based correction for the $\Delta CO - \Delta ffCO_2$ relationship.

First, I examined whether the ratio varied systematically with wind direction and speed using concurrent meteorological data (Fig. 6.3, panels (a) and (b)) (Heliasz & Biermann, 2024; Kubistin et al., 2024b, 2024a). This revealed wind-direction-dependent differences in the ratio for LIN (Fig. 6.3), HTM, and STE (Sect. A.4) stations. Next, I analyzed the wind direction distributions at each site to identify characteristic flow sectors associated with distinct $\Delta CO / \Delta ffCO_2$ ratios (Fig. 6.3, panel (c)). The preliminary sector boundaries were defined and the flask samples with valid meteorological data were assigned accordingly.

For each sector, I calculated the mean ratio ($\langle R_{flasks} \rangle$) and R^2 of a linear fit. Sector boundaries were iteratively optimized to maximize R^2 (Fig. 6.3, panels (d) and (e)). Final sector definitions were:

- STE: 30°–80°, remainder;
- HTM: 120°–280°, remainder;
- LIN: 135°–315°, remainder.

This method significantly improved the correlation at all three sites and stabilized the $\Delta CO / \Delta ffCO_2$ estimates. It enabled inclusion of stations like LIN, previously excluded due to exceeding the allowable $\langle R_{flasks} \rangle$ variability (see Sect. 6.2.3). The impact on inversion performance is discussed in Sect. 6.3.3.3.

6.2.3 Station selection criteria

The core assumption for calculating ^{14}C-calibrated CO-based $\Delta ffCO_2$ data from in situ CO observations is the existence of a characteristic $\langle R \rangle$ at each site (or two with wind selection; Sect. 6.2.2.2) that can be reliably estimated from ^{14}C-flask measurements. This method can be applied to all sites meeting these criteria. However, if $\langle R \rangle$ cannot be reliably estimated, e.g., due to low $\Delta ffCO_2$ signals or high spatio-temporal variability, the method is not applicable. In the following section, I define a quantitative criteria for station suitability for this approach.

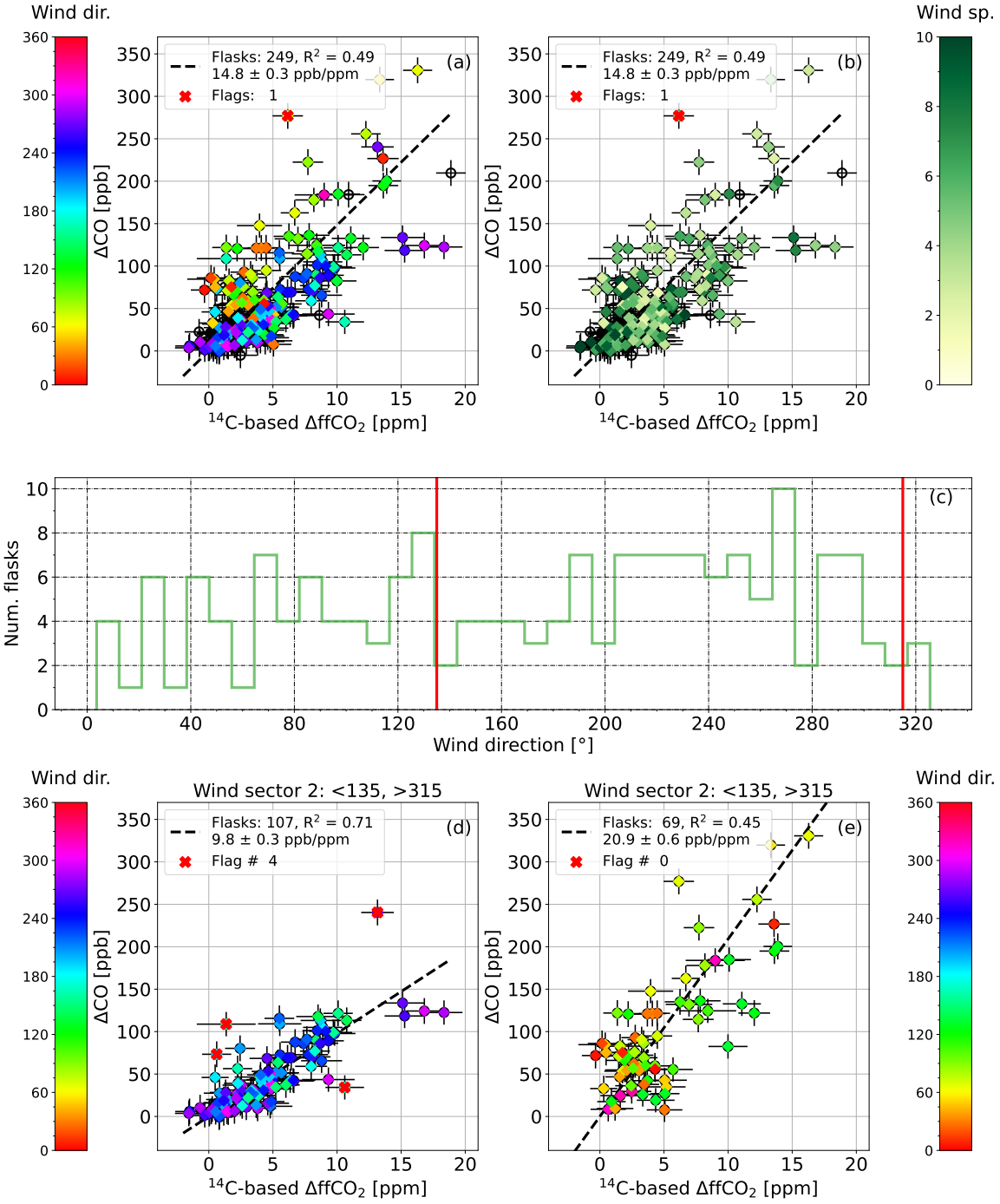

Figure 6.3. Selection of appropriate wind sectors for calculating wind-dependent ΔCO to ΔffCO$_2$ ratios at LIN station. Panels (a) and (b) display the complete flask dataset, while panels (d) and (e) show the dataset split by mean wind direction during sampling (excluding flasks with missing meteorological data), along with linear regressions (black dashed line, slope in legend) for corresponding flask pools. Color gradients in panels (a), (d), and (e) indicate mean wind direction, and in panel (b), mean wind speed during flask sampling. Panel (c) illustrates the distribution of wind conditions during flask sampling. Red lines denote the boundaries of the wind sectors used to divide the flask dataset into two groups.

First, I quantify the ratio variability caused by measurement errors using a Monte-Carlo experiment, similar to Maier et al., 2024a. I calculated "ideal" ΔCO values from the average $\langle \mathrm{R} \rangle$ and ^{14}C -based ΔffCO$_2$ data. Perturbations were then applied to these "ideal" data by drawing two random number sets from Gaussian distributions with zero mean and σ equal to typical measurement errors of ΔCO and ^{14}C-based ΔffCO$_2$. The perturbed ratio was recalculated and applied to the ΔCO data. Finally, I calculated the Root Mean Square Deviation (RMSD) between this error-prone CO-based ΔffCO$_2$ and the initial ^{14}C -based ΔffCO$_2$ data. This procedure was repeated 10,000 times to obtain the mean RMSD over all runs. This synthetic RMSD (RMSD^{synth}, Tab. 6.2, Fig. 6.2, right panels, yellow dots) represents variability solely from measurement errors, uninfluenced by possible ratio variability due to catchment area inhomogeneities (as only one ratio was used for all synthetic data sets).

To assess whether ΔffCO$_2$ signals are sufficient for reliable $\langle \mathrm{R} \rangle$ estimation, I define a threshold based on station-specific RMSD^{synth} normalized by the mean $\Delta\mathrm{ffCO}_2^{flask}$ signal (RMSD$^{synth}_{norm}$). If RMSD$^{synth}_{norm}$ exceeds the value of 1 (RMSD$^{synth}_{norm} > 1$), error-induced variation surpasses the mean signal, indicating that $\langle \mathrm{R}_{flasks} \rangle$ will be highly sensitive to measurement errors and therefore unreliable. I use RMSD^{synth} (estimated from 10,000 datasets) instead of RMSD calculated from measurement data (RMSD^{meas}) because RMSD^{meas} is a single realization and can vary in small flask sets due to the statistical nature of measurement errors. The stations excluded by this criterion are SVB, NOR, PAL and JFJ (Tab. 6.2, marked with (*)), i.e., all remote stations with low fossil fuel contributions.

The second criterion addresses source variability within the station's catchment area. Since synthetic data alone cannot estimate this variation, I compare RMSD^{meas} to its synthetic counterpart, since RMSD^{meas} incorporates both measurement error and $\langle \mathrm{R} \rangle$ variability. If RMSD^{meas} is more than twice larger than RMSD^{synth} (RMSD$^{meas} > 2 \cdot$RMSD^{synth}), $\langle \mathrm{R}_{flasks} \rangle$ variability in the catchment area dominates the observed variations in the data, making the one-ratio approach unsuitable. The JFJ, LIN, PAL and SAC stations were excluded by this criterion (Tab. 6.2, marked with (†)). Notably, applying wind-dependent ratios to LIN data reinstated it, since both sectors then fell within criteria boundaries (STE and HTM both passed, but wind-dependent ratios lead to improvements in the correlation between ΔCO and ΔffCO$_2$).

OXK station (Tab. 6.2, marked with ($^{+}$)) is a special case; it was excluded despite passing both criteria. ΔCO and ΔffCO$_2$ from OXK exhibited the lowest correlation in the measurement network ($\mathrm{R}^2 = 0.11$, Tab. 6.1 and Fig. A.19, panels (a) and (b)). I speculate, that the defined criteria failed due to several flasks with high ΔffCO$_2$ signals appearing anomalous for this site. Additionally, OXK is a mountain station (1022 m a.m.s.l.) prone to large catchment areas and complex atmospheric mixing. Consequently, OXK was excluded from the

Table 6.2. Absolute (RMSD^{meas}) and normalized by mean $\Delta\mathrm{ffCO}_2^{flask}$ signal (RMSD$^{meas}_{norm}$) RMSDs between ^{14}C -based and CO-based ΔffCO$_2$ in flasks, as well as synthetic RMSD (RMSD^{synth} and RMSD$^{synth}_{norm}$). RMSD^{synth} represent variability induced only by measurement errors, unlike RMSD^{meas} and normalized RMSD$^{meas}_{norm}$, which are also influenced by $\langle\mathrm{R}\rangle$ variability. (*) stations were excluded from inversions due to the low signals (RMSD$^{synth}_{norm}$ < 1). (†) stations were excluded from inversions due to the high $\langle\mathrm{R}\rangle$ variability in the catchment area (RMSD$^{meas}_{norm}$ > 2·RMSD$^{synth}_{norm}$). ($^+$) OXK was excluded due to the low correlation between ΔCO and ΔffCO$_2$ data ($\mathrm{R}^2 = 0.11$, Tab. 6.1) and KIT was excluded due to the possible ^{14}C contamination.

Station	RMSD^{meas}	RMSD$^{meas}_{norm}$	RMSD^{synth}	RMSD$^{synth}_{norm}$
CBW	2.25	0.47	1.24	0.26
GAT	1.61	0.77	1.16	0.54
HPB	2.02	0.84	1.26	0.49
HTM	1.76	0.74	1.21	0.54
HTM (wind sec. 1)	1.81	0.52	1.25	0.38
HTM (wind sec. 2)	1.57	0.64	1.25	0.54
JFJ*†	3.69	-18.22	1.44	1.50
KIT$^+$	2.52	0.55	1.36	0.30
KRE	1.17	0.84	1.22	0.81
LIN†	2.67	0.72	1.27	0.31
LIN (wind sec. 1)	1.98	0.42	1.32	0.28
LIN (wind sec. 2)	2.54	0.54	1.33	0.27
NOR*	1.53	1.54	1.27	1.35
OPE	1.75	0.70	1.28	0.53
OXK$^+$	2.53	0.69	1.32	0.36
PAL*†	0.87	2.40	1.15	-30.62
SAC†	3.45	0.71	1.42	0.30
STE	2.40	0.60	1.26	0.32
STE (wind sec. 1)	2.46	0.61	1.25	0.26
STE (wind sec. 2)	2.34	0.46	1.32	0.27
SVB*	1.67	1.64	1.34	1.33

standard CO-based inversion station set. KIT station was excluded as well following the argument of possible ^{14}C contamination made in Section 5.2.1. Its influence on the inversion system was investigated in Section 6.3.3.2.

6.2.4 ΔCO / $\Delta ffCO_2$ ratios derived from emission inventory

CO-based $\Delta ffCO_2$ offers high temporal resolution and the potential for increased spatial coverage, since in situ CO measurements are mandatory for ICOS Class 2 stations (ICOS RI, 2020). This would be beneficial if reliable effective atmospheric ΔCO / $\Delta ffCO_2$ ratios were available. Such ratios could be derived from high-resolution, sector-specific emission inventories or remote sensing, combined with footprint analyses, and estimates of non-fossil CO sources and CO air chemistry. Although high-resolution bottom-up inventories like EDGAR (Janssens-Maenhout et al., 2019) and CAMS (Kuenen et al., 2022) could provide this information, they often have significant uncertainties in the temporal and spatial disaggregation of national signals (Super et al., 2020). This leads to incorrect ratio estimates (Gamnitzer et al., 2006; Maier et al., 2024a), which highlights the need for observation-based inventory-independent estimates of ΔCO / $\Delta ffCO_2$ emission ratios.

Table 6.1 shows the modeled ratio ($\langle R_{model} \rangle$), calculated using the TNO emission inventory (Dellaert et al., 2019) and weighted by STILT footprints, as described in Maier et al., 2024a. The STILT-TNO ratios systematically underestimate the measurement-based flask ratios ($\langle R_{flasks} \rangle$), resulting in an average $\Delta ffCO_2$ overestimation of 22% when applied to continuous CO measurements, potentially due to not considered CO air chemistry and non-fossil CO sources in the inventories. Therefore, I conclude that only $\Delta^{14}C$-calibrated ratios (Sect. 6.2.2) currently provide $\Delta ffCO_2$ with acceptable uncertainties.

6.3 Results for CO-based $\Delta ffCO_2$ estimates

6.3.1 Forward runs and model-data-mismatch

Hourly ^{14}C-calibrated CO-based $\Delta ffCO_2$ series were calculated as described in Sect. 6.2. For the inversion system, only afternoon hours were used, as transport models perform best during this period (Geels et al., 2007; Peng et al., 2023). Following (Maier et al., 2024a), hourly CO-based $\Delta ffCO_2$ data were aggregated weekly to mitigate hourly mismatches between the transport model and actual meteorological conditions. The impact of these choices is explored in Sect. 6.3.3. Only stations which had passed the station selection criteria outlined in Sect. 6.2.3 were included in this analysis and the inversion setup. Measurement data were compared to simulations generated using the STILT transport model and the GCP emission inventory. These forward (FWD) runs (see Sect. 3.2) were calculated at exact sampling times and locations, and aggregated identically to the measurements.

Figure 6.4 illustrates the forward-simulated and observed $\Delta ffCO_2$ concentrations at three representative ICOS stations (CBW, HPB, and HTM), based on CO-derived estimates. The corresponding plots for all additional stations are provided in Appendix A.5.

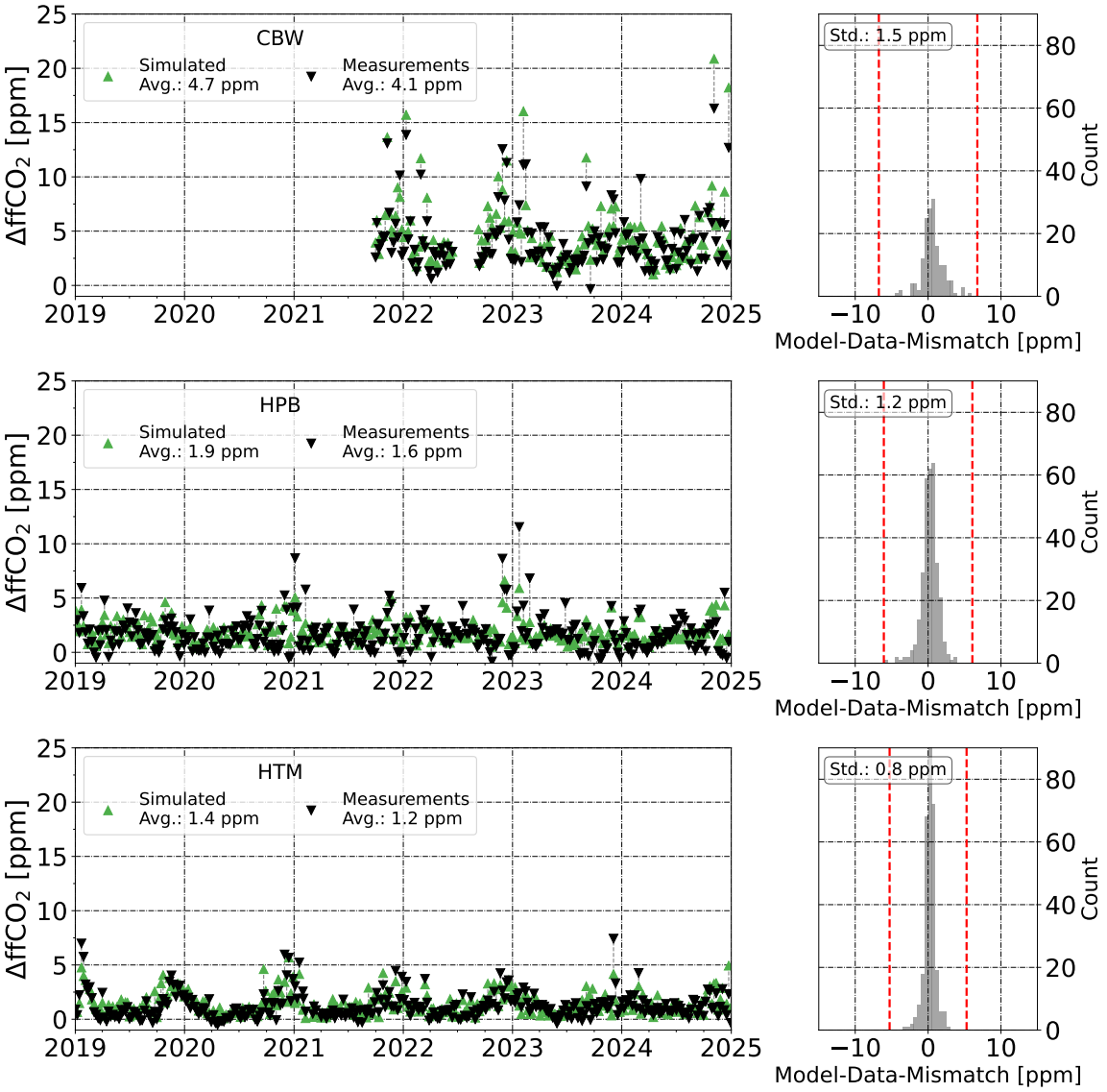

Figure 6.4. The CO-based ΔffCO$_2$ observations (black inverse triangles) and simulated ΔffCO$_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the RMSD-based measurement error.

Each left panel displays weekly-averaged ΔffCO$_2$ values, comparing the CO-based simulations (green triangles) with the observed concentrations (black triangles). The right-hand panels show the corresponding model–data mismatch (MDM) distributions, where the standard deviation is given in the top-left corner of each histogram.

Compared to the flask-based results (see Sect. 5.2.1), the variability of the MDM in the CO-based estimates is significantly reduced—approximately half as large—due to the use of weekly-averaged rather than instantaneous data, potentially indication better transport

Table 6.3. Statistics of the measured CO-based and simulated $\Delta ffCO_2$ signals in the ICOS network form 2019-2024. $\bar{X}_{obs}$ and $\bar{X}_{sim}$ depict mean measured and simulated signals for the corresponding site. $\bar{X}_{sim}$ - $\bar{X}_{obs}$ and $\sigma_{sim-obs}$ stand for the mean model-data-mismatch (MDM) and its standard deviation. $\bar{e}$ and SNR denote RMSD-based observation error and mean Signal-to-Noise Ratio. $|\bar{X}_{sim} - \bar{X}_{obs}| > 3\bar{e}$ shows the approximate portion of the data with higher absolute MDM then triple of the $\bar{e}$. (*) KIT was not used in regular inversions, only in sensitivity studies.

Site	$\bar{X}_{obs}$, [ppm]	$\bar{X}_{sim}$, [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$, [ppm]	$\sigma_{sim-obs}$, [ppm]	$\bar{e}$, [ppm]	$\lvert\bar{X}_{sim} - \bar{X}_{obs}\rvert > 3\bar{e}$, [%]	SNR, [rel.]
CBW	4.1	4.7	0.6	1.5	2.3	0%	1.8
GAT	1.9	2.3	0.4	1.2	1.6	0%	1.2
HPB	1.6	1.9	0.2	1.2	2.0	0%	0.8
HTM	1.2	1.4	0.2	0.8	1.8	0%	0.7
KIT*	4.9	5.4	0.6	2.5	2.5	0%	1.9
KRE	1.5	2.0	0.5	0.8	1.2	0%	1.3
LIN	2.5	3.3	0.9	1.8	2.7	0%	0.9
OPE	1.8	1.8	0.0	1.1	1.8	0%	1.1
STE	2.9	3.0	0.1	1.2	2.4	0%	1.2

model representativeness of mean weeky signals. Notably, the average observed $\Delta ffCO_2$ concentrations in the CO-based estimates are lower than in the flask-based results. This difference arises from the targeted flask sampling strategy used in the earlier, pre-CORSO years, which aimed to maximize the $\Delta ffCO_2$ signal (e.g., by sampling during optimal meteorological conditions or high-emission episodes). In contrast, the CO-based estimates rely on continuous high-frequency CO measurements and therefore reflect less biased, more representative atmospheric conditions.

Interestingly, the MDM in the CO-based approach shows a consistently positive bias across all stations, with simulated concentrations exceeding the observations. This contrasts with the flask-based MDM, where the deviations were more symmetrically distributed.

Due to high measurement uncertainty, the Signal-to-Noise Ratio (SNR) is generally low. Even for high-signal stations like CBW and KIT, SNR is below 2. Furthermore, stations including LIN, HPB, and HTM have an SNR below 1. The largest MDM is observed at LIN, suggesting that even with wind-dependent ratios, high ΔCO to $\Delta ffCO_2$ variability in the catchment area might not be fully compensated. This is supported by tests investigating the influence of individual stations, which also indicate a particularly high influence of LIN on the inversion system (Sect. 6.3.3.2).

6.3.2 Application of the Evaluation Matrix to the CO-based inversions

In the following sections, I evaluate core posterior features (e.g. seasonality and inter-annual trend) extracted from the data, as described in Sect. 3.2.4. This evaluation will classify these features based on the results from the synthetic application of this experiment set (Sect. 4.6). Because the synthetic inversions have no inherent transport model errors, I assume sampling routine differences are negligible. This allows us to apply the results from those sections to CO-based inversions.

6.3.2.1 Seasonal cycle of $\Delta ffCO_2$ emission trend derived from CO-based inversions

The seasonal cycle of posterior fluxes (Fig. 6.5) exhibits a peak-to-peak amplitude 2 to 3 times larger than the GCP inventory. This increase is primarily driven by significantly lower summer minima, with four summers showing a four-fold greater flux reduction relative to the Seasonless Prior compared to the GCP (Tab. 6.4). On the other hand, winter fluxes deviate less from the GCP. Relative maxima deviations of winters 20/21, 22/23, and 23/24 from the GCP are all below 10% relative to the Seasonless prior, showing no significant deviation based on the synthetic tests (Sect. 4.6.1), while winters 19/20 and 21/22 exhibit a significant underestimation compared to the GCP fluxes (17% and 11% respectively). All extrema, except for summer 2024 with phase shift of 54 days (Tab. 6.4), show no significant phase shift w.r.t. the GCP being lower than expected phase shift variation due to the magnitude of the measurement errors (35 days for winter and 31 days for summer, Sect. 4.6.1).

The year 2019, including winter 19/20, may be influenced by lower data density than the rest of the investigation period (colorbar or Fig. 4.5). Beyond this, no clear anomalies in the posterior can be attributed to data density changes. Extrema positions generally align with the GCP, except for summer 2024, which may be affected by particularly low fluxes in the period afterward.

In conclusion, I observe a clear seasonal influence on posterior results, indicating a strong underestimation of fossil emissions during summer. Winter fluxes are mostly within range of the GCP considering results of synthetic Evaluation matrix (Sect. 4.6.1), with some instances of underestimation but no significant phase shifts.

6.3.2.2 Inter-annual $\Delta ffCO_2$ emission trend derived from CO-based inversions

Fossil fuel emission trends are investigated using CO-based $\Delta ffCO_2$ data. As shown in Figure 6.6, posterior fluxes from the F1T and F0.5T inversions are compared to the GCP inventory and a Trendless prior. I quantify the changes in emissions during the main period by fitting linear trends to the posterior time series.

As seen from Fig. 6.6, the inter-annual posterior trends derived from CO-based $\Delta ffCO_2$

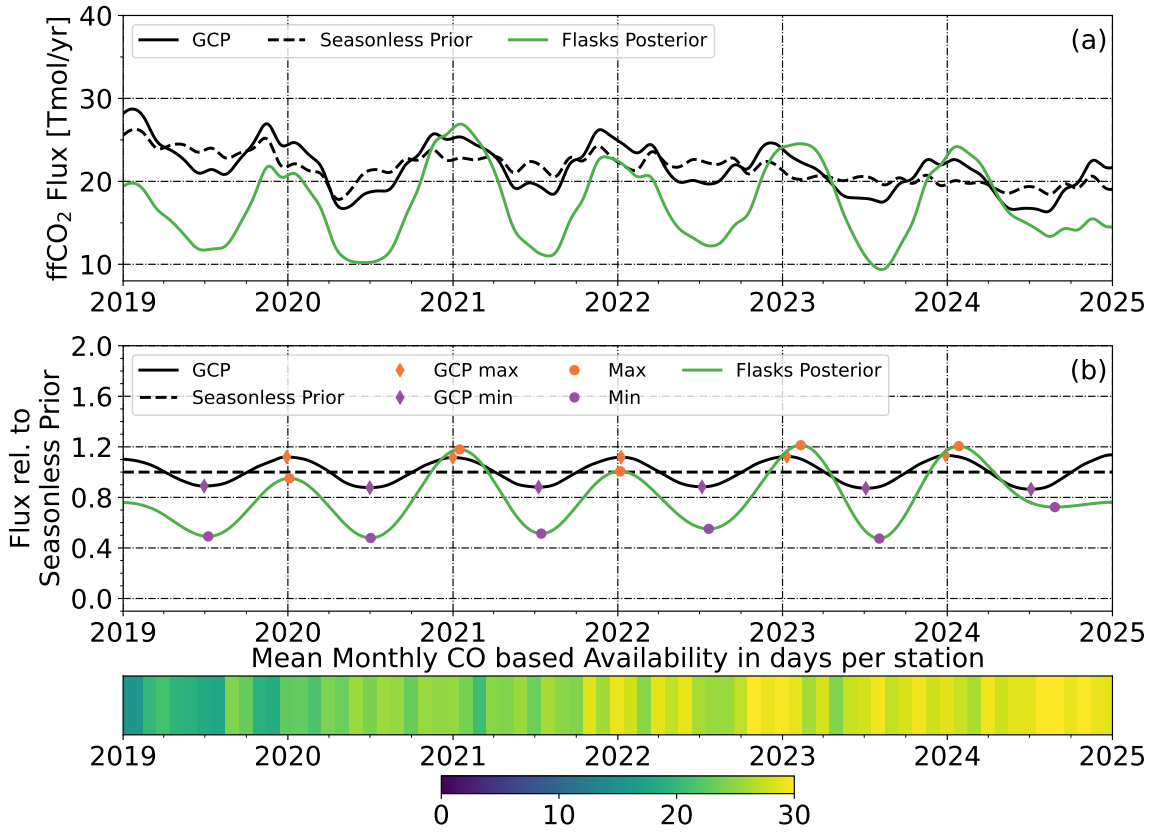

Figure 6.5. Results of the seasonal cycle evaluation of the CO-based posterior ΔffCO$_2$ fluxes. The absolute (panel (a)) and relative to the Seasonless Prior (panel (b)) fluxes in the Germany domain using CO-based ΔffCO$_2$ data show the GCP inventory (black solid line), the Seasonless Prior (used as input, black dashed line) and posterior results (green line). Orange and purple circles mark the maxima and minima of the posterior, with corresponding GCP extrema marked by diamonds. The color bar at the bottom indicates the network-wide mean monthly availability of CO-based estimates per station.

data with F0.5T and F1T are closer to the GCP inventory than the Trendless Prior throughout all the period. The differences between F1T and F0.5T posteriors are similar to those in flask-based inversions (Sect. 5.2.2.2), with F1T oscillating slightly around F0.5T.

However, their absolute values are significantly lower compared to the GCP due to underestimated summer fossil fluxes (see Sect. 6.3.2.1). From summer 2021 to winter 2025, the F0.5T posterior shows a downward trend with a consistent ∼13% offset from the GCP trend. Fluxes from the first year of the posterior, which coincides with the lowest data density (Fig. 6.6, colorbar), are 20-40% lower than the GCP estimate.

Linear regression slopes for the entire period are essentially zero for both posteriors (0.01±0.03 Tmol/yr for F1T and 0.11±0.03 for F0.5T, not shown),implying a significant discrepancy compared to the downward GCP trend of -0.72±0.01 Tmol/yr. This is larger than the synthetic estimate for trend variations induced by measurement uncertainty (∼0.4

Table 6.4. Positions (”Date” columns) and relative to the Seasonless Prior amplitudes (”Value” columns) of the seasonal extrema extracted from the GCP inventory and the posterior of the CO-based inversion as well as the difference between them. The difference between the positions is in days.

Season	GCP		Flask-based $\Delta ffCO_2$		Obs. - GCP	
	Date	Ampl., [rel.]	Date	Ampl., [rel.]	Phase shift, [days]	Ampl. diff., [rel.]
Winter 19/20	30.12	1.12	04.01	0.95	5	-0.17
Winter 20/21	01.01	1.12	15.01	1.18	14	0.06
Winter 21/22	08.01	1.12	06.01	1.01	-2	-0.11
Winter 22/23	11.01	1.12	10.02	1.21	30	0.09
Winter 23/24	28.12	1.13	26.01	1.21	29	0.08
Summer 2019	29.06	0.89	08.07	0.49	9	-0.40
Summer 2020	01.07	0.88	02.07	0.48	1	-0.40
Summer 2021	09.07	0.88	15.07	0.51	6	-0.37
Summer 2022	07.07	0.88	21.07	0.55	14	-0.33
Summer 2023	04.07	0.87	03.08	0.47	30	-0.40
Summer 2024	04.07	0.86	27.08	0.72	54	-0.14

Tmol/yr, Sect. 4.6.2).

If the 2019 fluxes are excluded, the posterior trend estimates (F1T: -0.62±0.04 Tmol/yr; F0.5T: -0.89±0.03 Tmol/yr, Fig. 6.6) align closely with the GCP trend of -0.56±0.04 Tmol/yr. Relative fluxes to the Trendless Prior show similar results: a -2.1% GCP emission reduction rate versus -2.3% (F1T) and -3.4% (F0.5T) for posteriors.

These findings highlight two key points. First, 2019 data density may be insufficient for reliable posterior results, contrary to the conclusions presented in Sect. 4.3, possibly due to effects of transport errors, which are not accounted for in the synthetic runs. Second, like with flasks, CO-based inversions can estimate mean emission trends during periods of high data density.

6.3.2.3 Absolute and annual mean $\Delta ffCO_2$ emissions derived from CO-based inversions

Figure 6.7 depicts the absolute and annually aggregated fossil fuel CO_2 fluxes derived from CO-based inversions for the Germany+ domain. To limit the prior information available to the inversion system and achieve data-driven posteriors, inversions were conducted using

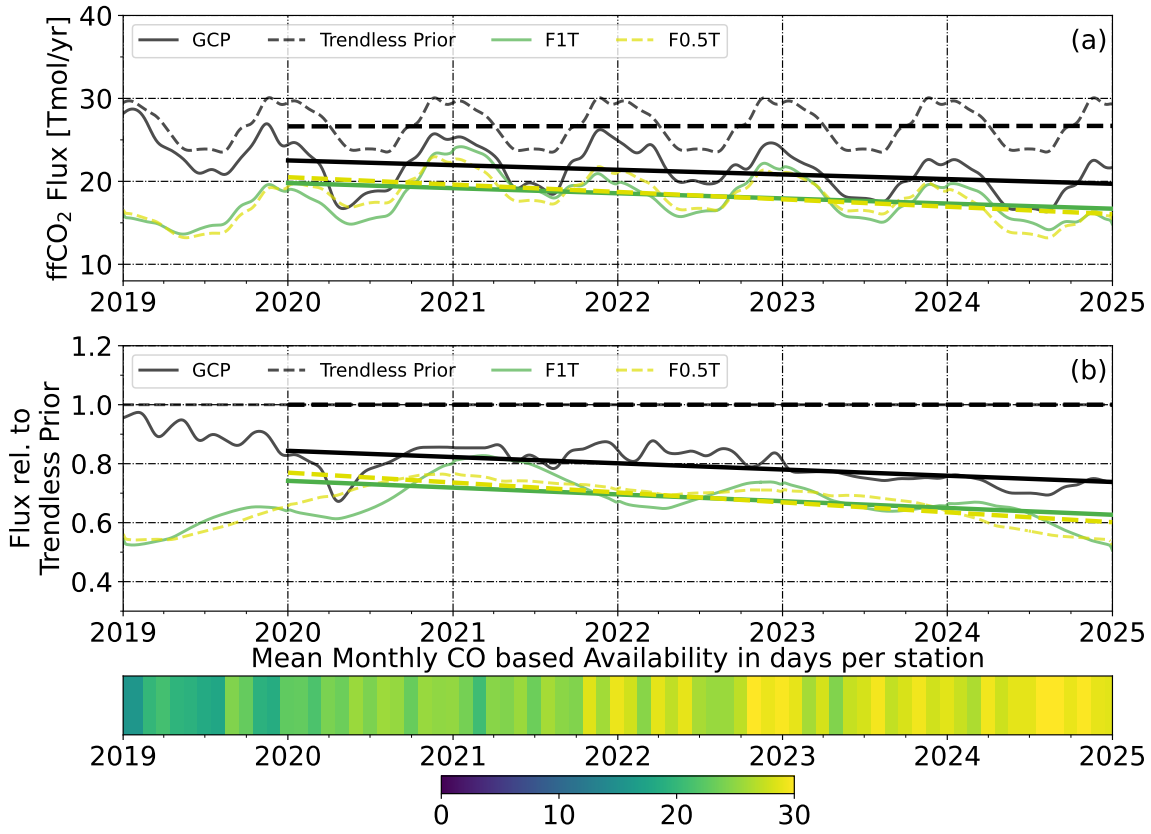

Figure 6.6. Results of the inter-annual trend analysis for the CO-based inversions. The absolute (panel (a)) and relative to the Trendless Prior (panel (b)) fluxes in the Germany+ domain. The GCP inventory (black solid line) and the Trendless Prior (used as input, black dashed line) are shown alongside posterior results for F1T (green solid line) and F0.5T (yellow dashed line) filters. The lines show linear regressions of the correspondingly colored flux curves, excluding 2019 with the lowest data density. The color bar at the bottom indicates the network-wide mean monthly availability of CO-based estimates per station.

the Flat Zero and the Flat 200p Prior with 40% prior uncertainty (Sect. 3.2.4).

The seasonal and trend features from previous sections are evident in the posteriors (Fig. 6.7, panel (a)), even with low information priors like the Flat 200p and Flat Zero. Consistent with the seasonality analysis (Sect. 6.3.2.1), summer fluxes are significantly underestimated, while winter fluxes generally align with the GCP inventory. Except for 2019 and late 2024, the general trend direction matches the GCP, but low summer fluxes cause a downward emission offset (Sect. 6.3.2.2). Given high uncertainties and the amplified seasonal cycle, prominent posterior features like reduced emissions during the COVID-19 lockdown are not discernible. RMSD values between the Smooth Prior and posteriors (0.19 for the Flat Zero, 0.26 for the Flat 200p) are significantly higher than the potential deviations caused by the measurement errors (0.11 ± 0.02, Sect. 4.6.3).

Similar to the trend (Fig. 6.6), annual posterior emissions significantly underestimate

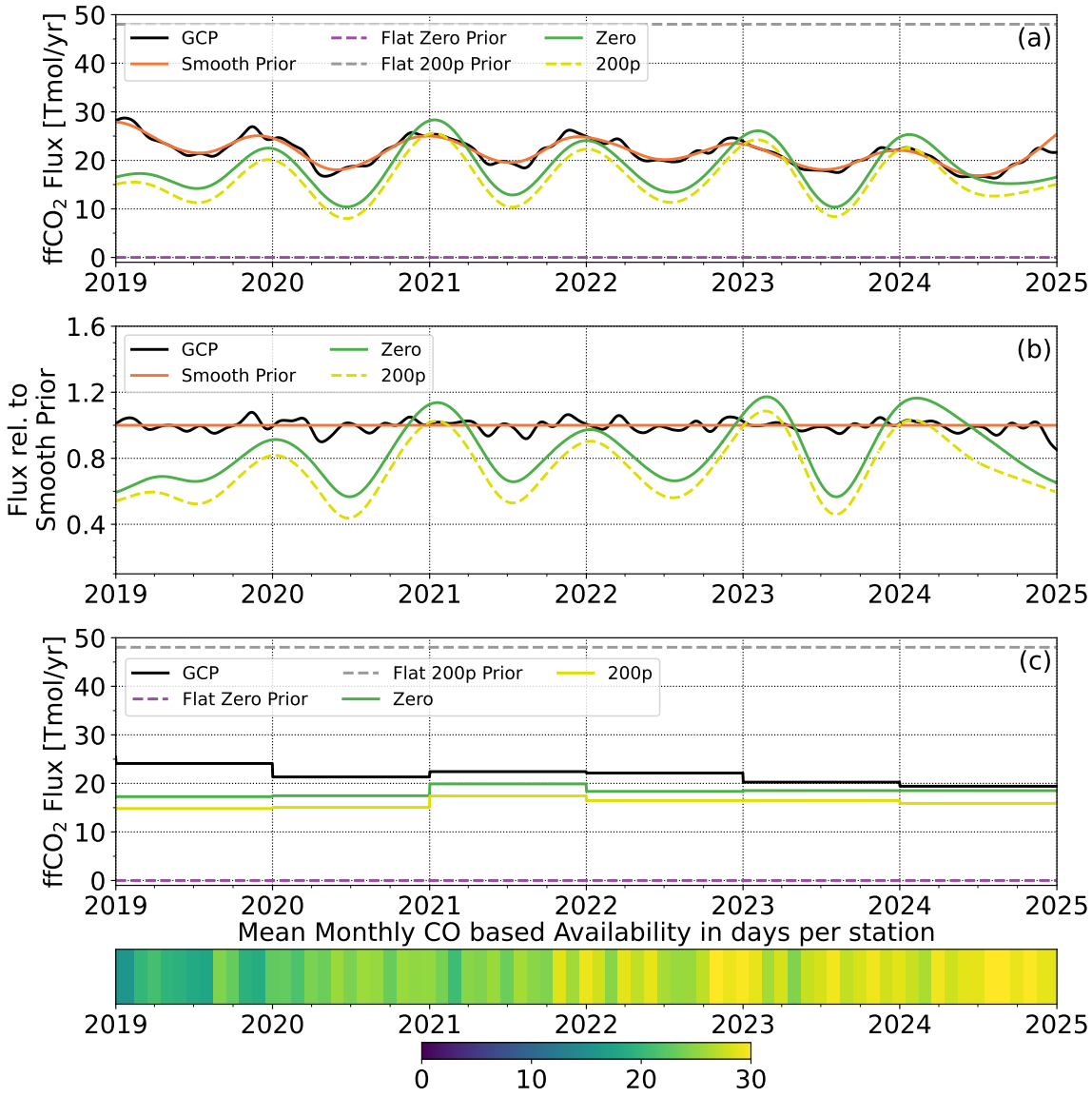

Figure 6.7. Results of the analysis of absolute and annually aggregated emission fluxes for the CO-based inversions in the Germany+ domain. Panel (a) shows the absolute and panel (b) the relative to the Smooth Prior $\Delta ffCO_2$ fluxes. Panel (c) shows yearly aggregated emissions from the panel (a). the Flat Zero Prior and the Flat 200p Prior (used as inputs, respectively, purple and grey dashed lines) as well as the Smooth Prior (orange solid line) are shown alongside posterior results (green solid and yellow dashed lines). The color bar indicates the mean monthly availability of CO-based estimates per station.

the GCP inventory (Fig. 6.7, panel (c)), falling outside the boundaries defined by the synthetic experiments (0.94±0.06 for the Flat 200p Prior and 1.07±0.06 for the Flat Zero Prior relative to yearly GCP fluxes). The underestimation from expected values ranges approximately from 10% to 30% (Tab. 6.5), with 2019 and 2020 showing the greatest underestimation. Based on the findings from the previous sections, which indicate a systematic

Table 6.5. Annual mean CO-based posterior $ffCO_2$ fluxes relative to the GCP inventory for the full years.

Prior	2019	2020	2021	2022	2023	2024
Flat Zero	0.72	0.82	0.89	0.83	0.91	0.95
Flat 200p	0.61	0.71	0.78	0.74	0.81	0.82

underestimation of summer fluxes in the posterior, it is clear that this bias also influences the annual mean emissions. Notably, the extent of underestimation is more pronounced in the earlier years of the study period, while the discrepancy decreases over time, with the smallest deviation observed in 2024. This characteristic is revisited and explained further in the section discussing the uncertainties related to the CO-based $\Delta ffCO_2$ proxy (Sect. 6.3.3).

The ”over- and undershooting” posterior patterns observed with the Flat Zero and Flat 200p Priors reoccurred here (Sect. 4.6.3). The difference between the two posteriors ranges from 9% to 13%, with 2024 exhibiting the largest discrepancy. This difference is considerably higher than that observed in flask-based inversions and closer to the integral-based synthetic inversions (avg. 13%, Sect. 4.6.3).

6.3.3 Uncertainty investigation

6.3.3.1 Effect of CO-based $\Delta ffCO_2$ averaging

As described in Sect. 3.2.2.1, the standard CO-based $\Delta ffCO_2$ inversion setup utilizes afternoon hours aggregated on a weekly basis. This procedure, suggested by Maier et al., 2024b, addresses the transport model's inability to accurately simulate hour-specific afternoon $\Delta ffCO_2$ concentrations. This deficit results from transport model errors in combination with spatially heterogeneous emissions in the stations' catchment areas—particularly from point sources. The presence or absence of such sources within a station's footprint can significantly influence $\Delta ffCO_2$ concentrations, and discrepancies between real and modeled atmospheric conditions can lead to substantial model–data mismatches (MDM). Aggregating the CO-based $\Delta ffCO_2$ signal to weekly averages mitigates the need for hour-level transport accuracy and instead requires only a reliable estimate of the average impact of these emission sources. Maier et al., 2024b investigated this effect in Heidelberg, a highly heterogeneous urban environment with a high density of nearby point sources. The following section analyzes this effect on stations within the ICOS network, which are less influenced by surrounding point sources due to their locations in the hinterland (ICOS RI, 2020).

Similar to patterns observed by Maier et al., 2024b (Figure C2), shorter aggregation intervals lead to lower emissions in the Germany+ domain. Fig. 6.8 shows posterior fossil fuel

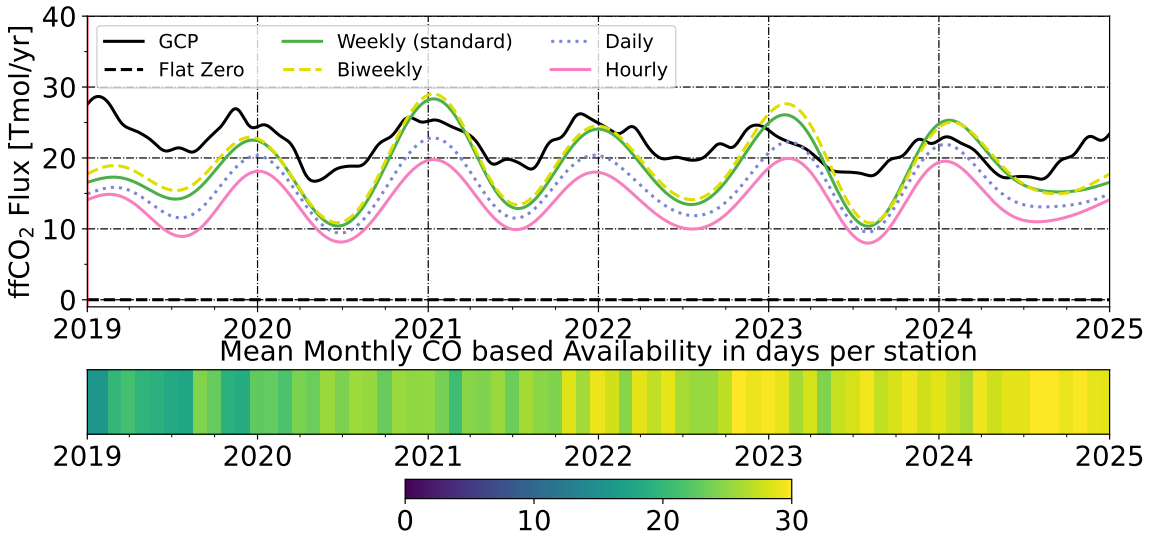

Figure 6.8. Posterior CO-based ΔffCO$_2$ emissions for the Germany+ domain using a Flat Zero Prior (black dashed line) with 40% prior uncertainty in dependence of averaging interval. The posterior of the standard setup with weekly averaging is shown as green solid line along with the GCP inventory as reference (black solid) and posteriors with varying averaging periods: biweekly (yellow dashed line), daily (blue dotted line) and hourly averaging (pink solid line). The color bar at the bottom indicates the network-wide mean monthly availability of CO-based estimates per station.

emissions for the Germany+ domain derived using CO-based ΔffCO$_2$ estimates with varying temporal aggregation intervals—ranging from hourly to biweekly. All inversions were performed using a Flat Zero Prior with 40% prior uncertainty. The posterior fluxes exhibit a strong dependency on the aggregation interval, particularly for shorter periods. Aggregation windows longer than one week have a negligible influence on the posterior estimates compared to the standard weekly setup. Notably, a seasonal pattern emerges, with winter fluxes being more strongly affected by aggregation than summer fluxes. The underlying reasons for this aggregation- and season-dependent emission reduction are further explored in the subsequent section.

Standard weekly aggregated ΔffCO$_2$ model forward results based on the GCP inventory are, on average, higher than CO-based ΔffCO$_2$ concentrations (see Tab. 6.3). This results in an MDM distribution with a positive mean shift and more positive outliers than negative ones. This effect leads to a systematic underestimation of the GCP inventory by CO-based posteriors, as observed previously (see Sect. 6.3.2). Comparing the difference between simulated (X_{sim}) and measurement data (X_{obs}) in weekly aggregated (Tab. 6.6, winter or summer) and hourly (Tab. 6.7) data, the mean MDM ($\bar{X}_{sim}$ - $\bar{X}_{obs}$) does not change significantly, as expected. However, the width of the MDM distribution ($\sigma_{sim-obs}$) is much higher (about

Table 6.6. Statistics of the weekly aggregated measured CO-based and simulated $\Delta ffCO_2$ signals in the ICOS network form 2019-2024. $\bar{X}_{obs}$ and $\bar{X}_{sim}$ depict mean measured and simulated signals for the corresponding site. $\bar{X}_{sim}$ - $\bar{X}_{obs}$ and $\sigma_{sim-obs}$ stand for the mean model-data-mismatch (MDM) and its standard deviation.

Site	Weekly aggregation					
	Winter (December, January, February)			Summer (June, July, August)		
	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]
CBW	5.1	0.9	1.8	2.8	0.0	0.9
GAT	3.0	0.3	1.8	1.3	0.2	0.6
HPB	1.9	0.1	1.4	1.9	-0.3	1.0
HTM	2.2	-0.2	0.9	0.6	0.3	0.4
KRE	2.4	0.3	1.0	1.0	0.5	0.5
LIN	4.4	0.6	2.4	1.4	0.5	0.7
OPE	3.2	-0.8	1.3	1.4	-0.2	0.8
STE	4.2	-0.3	1.5	1.9	0.0	0.8

a factor of 2.3 higher in both summer and winter) for hourly data than for weekly data. Since neither MDM distribution is centered at zero, a larger $\sigma_{sim-obs}$ increases the frequency and magnitude of positive deviations in the tails of the distribution. As described in Sect. 3.2.2, MDM contributes quadratically to the inversion system's cost function, meaning that large deviations have a disproportionate influence on the posterior calculation. Consequently, the broader and more skewed MDM distribution associated with hourly and daily aggregation leads to systematically lower posterior flux estimates compared to weekly aggregation.

The ratio $\sigma^{hourly}_{sim-obs}/\sigma^{weekly}_{sim-obs}$ for core summer (Tab. 6.6 and 6.7, right half) is approximately equal to that for core winter (Tab. 6.6 and 6.7, left half). However, the absolute change from $\sigma^{weekly}_{sim-obs}$ to $\sigma^{hourly}_{sim-obs}$ in winter is twice as high as in summer, implying a broader MDM distribution in winter than in summer, and thus a larger reduction in absolute fluxes. Nevertheless, the flux changes relative to the corresponding period do not exhibit such a strong seasonal dependence (approx. 5% difference in relative fluxes between summer and winter).

6.3.3.2 Influence of individual stations on the CO-based posterior

Figure 6.9 shows the results of a leave-one-out sensitivity test for the ICOS stations contributing to the CO-based $\Delta ffCO_2$ inversion in the Germany+ domain. Each colored line

Table 6.7. Statistics of the hourly measured CO-based and simulated $\Delta ffCO_2$ signals in the ICOS network form 2019-2024. $\bar{X}_{obs}$ and $\bar{X}_{sim}$ depict mean measured and simulated signals for the corresponding site. $\bar{X}_{sim}$ - $\bar{X}_{obs}$ and $\sigma_{sim-obs}$ stand for the mean model-data-mismatch (MDM) and its standard deviation.

Site	Weekly aggregation					
	Winter (December, January, February)			Summer (June, July, August)		
	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ - $\bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]
CBW	5.2	0.7	5.1	2.7	0.0	2.6
GAT	3.1	0.2	3.2	1.3	0.2	1.3
HPB	1.9	0.1	2.6	1.9	-0.3	1.7
HTM	2.3	-0.2	2.0	0.6	0.3	1.0
KRE	2.5	0.2	2.4	1.0	0.5	1.2
LIN	4.5	0.5	6.5	1.4	0.5	1.7
OPE	3.3	-0.8	2.8	1.4	-0.2	1.4
STE	4.4	-0.3	4.3	2.0	0.0	2.2

represents a posterior flux time series derived by excluding one station at a time from the inversion setup. The standard setup (green line) includes all stations, while the black solid and dashed lines show the GCP inventory and the Flat Zero Prior, respectively.

In general, the posterior fluxes are robust against the omission of individual stations, with only minor variations. However, two stations exhibit a more pronounced impact.

First, the exclusion of LIN (orange dashed line) results in posterior fluxes that show improved agreement with the GCP inventory, particularly at the beginning and end of the time series. Additionally, the seasonal cycle becomes more clearly defined. LIN is the only station for which a wind-sector-dependent CO/$\Delta ffCO_2$ ratio had to be applied, with sector-specific ratios differing by more than a factor of two (Tab. 6.2.2). Moreover, LIN displayed the strongest discrepancies in forward-modeled versus observed CO-based $\Delta ffCO_2$ signals. These findings indicate that LIN has a disproportionately strong influence on the inversion and that the applied correction does not fully account for local emission heterogeneity. Thus, LIN should be excluded from the CO-based $\Delta ffCO_2$ station network used in inversions.

Second, the removal of STE (yellow solid line), another site with a wind-sector-dependent CO/$\Delta ffCO_2$ ratio, leads to notable changes in the posterior during the early years of the time series. Although the effect is smaller than for LIN, it still suggests that STE also introduces some inconsistency due to emission inhomogeneities in its catchment area. The influence is, however, less systematic and fades toward the later years.

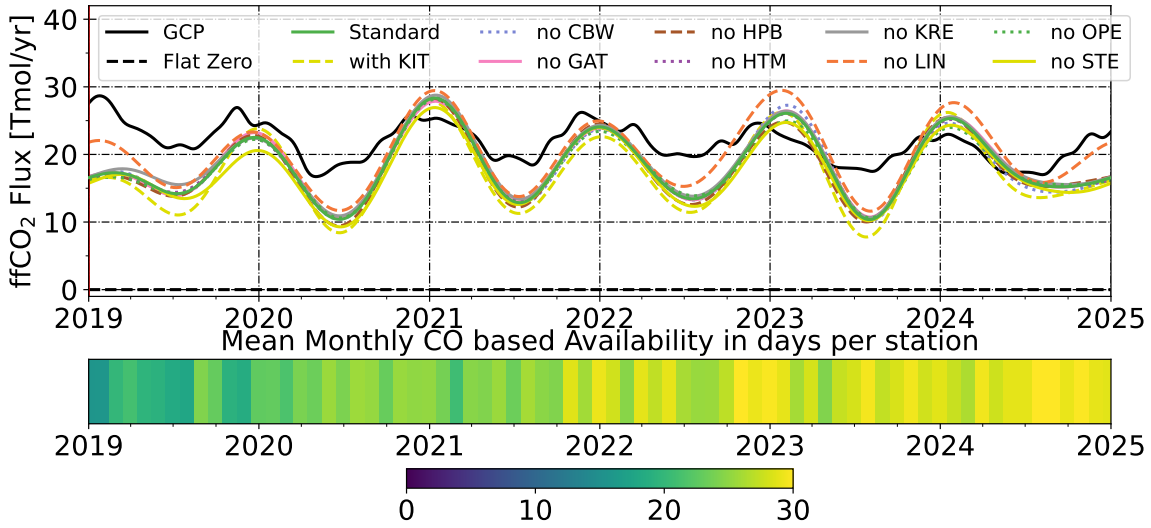

Figure 6.9. Posterior $\Delta ffCO_2$ fluxes for the Germany+ domain based on CO-derived estimates using a Flat Zero prior with 40% prior uncertainty. The standard posterior (green solid line) is compared to leave-one-out setups, where individual ICOS stations were excluded (other lines), as well as a configuration including the KIT site (yellow dashed line). The GCP inventory (black solid line) and the Flat Zero prior (black dashed line) are shown for reference. The color bar indicates the mean monthly availability of CO-based $\Delta ffCO_2$ data per station.

Together, these results underline the limitations of simplified sector-based corrections for sites affected by complex source environments and highlight the importance of carefully selecting stations for CO-based inversions.

6.3.3.3 Effect of different treatment of stations with wind-dependent ratios

As described in Sect. 6.2.2.2, wind-dependent $\langle R_{flasks} \rangle$ values were applied during CO-based $\Delta ffCO_2$ calculations for HTM, LIN, and STE sites to mitigate the impact of emission source inhomogeneities within the stations' catchment areas. Due to the novelty of this procedure and given LIN's high influence on the posterior (Sect. 6.3.3.2), I evaluated the impact of this and alternative methods on posterior emissions.

The Standard CO-based inversion setup (Fig. 6.10, green line) for the German+ domain employs wind-dependent ratios. Alternatively, in the "No Wind Selection" scenario (yellow dashed line), HTM, LIN, and STE were treated like other network stations, applying one mean $\langle R_{flasks} \rangle$ across all wind conditions. Additionally, I investigated the posterior effect of using only data from the wind sector with a lower $\langle R_{flasks} \rangle$ ("Only low R sector", blue dotted line) or uniformly applying this lower $\langle R_{flasks} \rangle$ to the entire data series ("Low R for all conditions", pink solid line). These last two investigations were motivated by the

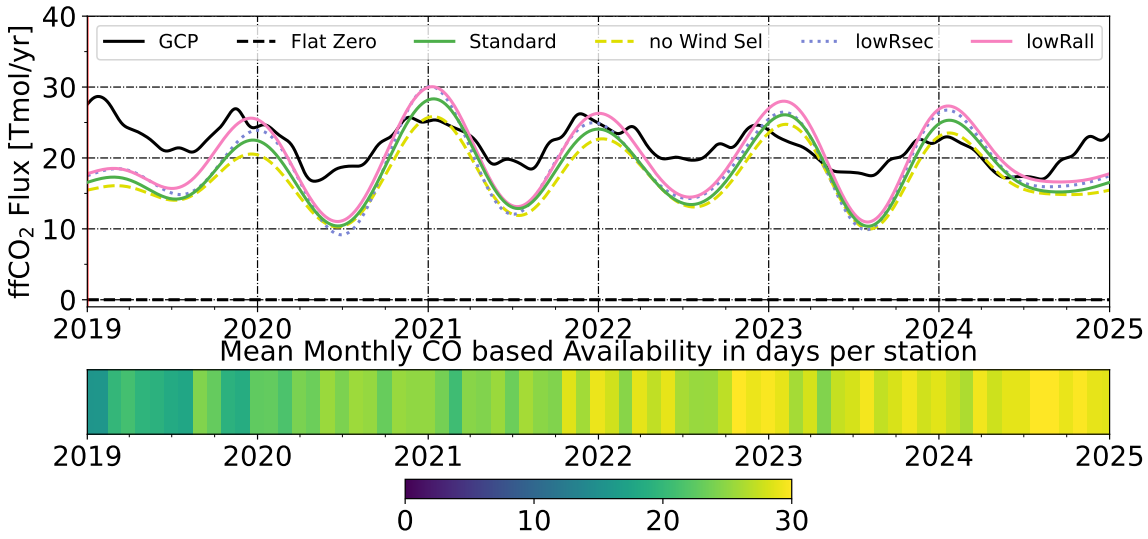

Figure 6.10. Posterior CO-based ΔffCO$_2$ emissions for the Germany+ domain. The standard setup (green solid line) uses the Flat Zero Prior (black dashed line) with 40% uncertainty and wind-dependent $\langle R_{flasks} \rangle$ for HTM, LIN, and STE. The figure also shows the GCP inventory (black solid line) as reference and posteriors from other wind-related scenarios: "No Wind Selection" (yellow dashed line), "Only low R sector" (blue dotted line), and "Low R for all conditions" (pink solid line). The color bar at the bottom indicates the network-wide mean monthly availability of CO-based estimates per station.

assumption that high $\langle R_{flasks} \rangle$ values (above 20, leading to lower CO-based ΔffCO$_2$ data) could be influenced by local signals not captured by the transport model or by non-fossil CO sources near the stations. The ratios used in these investigations are provided in Tab. 6.1.

The "No Wind Selection" posterior yields, on average, 5% lower emissions than the Standard setup. In this case, periods with lower $\langle R_{flasks} \rangle$ in the Standard setup receive a higher mean $\langle R_{flasks} \rangle$, resulting in lower ΔffCO$_2$ values, and vice versa for periods with initially higher $\langle R_{flasks} \rangle$. While CO-based ΔffCO$_2$ data for HTM and LIN are nearly evenly split between two wind sectors (HTM: 42% and 58%; LIN: 44% and 56%), mean HTM signals were only slightly affected while LIN signals became lower. This leads to a stronger underestimation of the GCP FWD run in the whole German+ domain, as higher ΔffCO$_2$ values were more significantly impacted by applying the mean ratio in the "No Wind Selection" setup. Additionally, 98% of STE data experienced a reduction in CO-based ΔffCO$_2$ values, as they were initially calculated using a lower $\langle R_{flasks} \rangle$ in the Standard setup. This largely explains the predominantly lower emissions in the corresponding posterior and an even greater discrepancy with the GCP inventory compared to the Standard setup.

On the other hand, applying the lower $\langle R_{flasks} \rangle$ to all wind conditions in the "Low R

for all conditions" scenario, as expected, leads to approximately 8% higher fluxes in the resulting posterior. However, this would only marginally improve the strong summer underestimation of the GCP inventory while increasing the winter discrepancy, potentially overestimating the influence of high CO yield fossil sources (e.g., heating).

Removing CO-based ΔffCO$_2$ data from the higher $\langle \mathrm{R}_{flasks} \rangle$ wind sector in the "Only low R sector" setup magnifies the seasonal peak-to-peak amplitude in the posterior. This might suggest additional seasonality in the transport model, non-fossil CO emissions, or other unaccounted effects.

Overall, all wind direction influenced posteriors generally lie within a $\pm$10% band around the Standard posterior. This effect is minor compared to effects from investigating other system properties (e.g., measurement errors, Sect. 4.6). The seasonal effect relative to the Standard posterior is small for "Low R for all conditions" and "No Wind selection" cases (Fig. 6.10, panel (b)). However, the "Only low R sector" setup further amplifies the seasonal amplitude compared to the GCP inventory. This analysis underscores the importance of inhomogeneities for CO-based ΔffCO$_2$ and confirms wind-selected ratios as a viable option to reduce their influence. Nevertheless, their effect cannot explain the strong underestimation of summer fluxes in the Standard setup.

6.3.3.4 Sensitivity of the posterior fluxes to the European Δ^{14}C and CO background estimates

To assess the sensitivity of the CO-based posterior flux estimates to a possible bias in the assumed European Δ^{14}C background, a perturbation experiment analogous to that performed for the flask-based analysis (Sect. 5.2.3.1) was conducted. Since the same Δ^{14}C background is used to derive both flask- and CO-based ΔffCO$_2$ estimates, its influence must be evaluated consistently across proxies.

In this experiment, the European Δ^{14}C background was systematically shifted by $\pm$2‰. Based on the perturbed backgrounds, new flask-based ΔffCO$_2$ estimates were calculated, which were then used to update the CO/ΔffCO$_2$ ratios. These revised ratios were subsequently used to produce continuous biased CO-based ΔffCO$_2$ time series. Finally, two sensitivity inversions were performed using these adjusted input datasets.

The results of this experiment are illustrated in Fig. 6.11. The upper panel shows the absolute posterior fluxes for the unperturbed and perturbed background cases, while the lower panel highlights the differences relative to the standard inversion. The seasonal asymmetry in the response is clearly visible.

As in the flask-based case, the posterior fluxes responded symmetrically to the $\pm$2‰ bias in the ^{14}C background, with a magnitude of change in the range of 3–5 Tmol/yr per $\pm$2‰ shift in the background. However, the seasonal pattern of sensitivity differs. For the

CO-based posterior fluxes, the impact of the background offset is largest during the winter half-year and smaller during summer. This contrasts with the flask-based posteriors, which showed the strongest sensitivity during summer and relatively muted response in winter.

For flask and integral samples (see Sect. 7.2.3.1), a ± 2‰ ^{14}C background bias directly affects the $\Delta ffCO_2$ values, leading to uniformly positive of negative offsets in absolute $\Delta ffCO_2$ estimates. The seasonality in the corresponding tests originates from higher relative changes of (smaller) summer signals compared to winter. For CO-based $\Delta ffCO_2$ data, the ^{14}C background bias changes $\Delta CO/\Delta ffCO_2$ ratios. In this case the relative change for a single station's CO-based $\Delta ffCO_2$ values is constant (as the same ratio is applied to all ΔCO values), but the absolute signal change is greater in winter due to higher ΔCO signals in this season.

Based on investigations of ^{14}C -based $\Delta ffCO_2$ proxies (e.g., Sect. 5.2.2.1 and 7.2.2.1), I suspect the $\Delta^{14}C$ background estimate is biased towards lower values following the discontinuation of the MHD data series. The bias appears to be approximately 1‰ in 2023, increasing to 2‰ in 2024. Unfortunately, most flasks used to calculate $\Delta CO/\Delta ffCO_2$ ratios were collected during the CORSO project (Levin et al., 2024) in 2024. This may have systematically biased all CO-based $\Delta ffCO_2$ estimates. This bias would also explain the consistently positive MDM observed for CO-based $\Delta ffCO_2$ across all stations (Sect. 6.3.1), highlighting the high susceptibility of this and other proxies to background biases.

In contrast to the flask-based $\Delta ffCO_2$ estimates, the CO-based approach requires not only an estimate of the $\Delta^{14}C$ background but also of the marine CO background. Therefore, a second sensitivity experiment was conducted to quantify the influence of potential biases in the marine CO background on the estimated posterior $\Delta ffCO_2$ fluxes. Figure 6.12 illustrates the posterior results obtained when shifting the marine CO background by ± 15 ppb and recalculating the full chain of CO-based $\Delta ffCO_2$ estimates accordingly (including calculations of new ratios).

Unlike the previously discussed background sensitivity experiments, the effect of the CO background bias is clearly asymmetric. A reduction of 15 ppb leads to only minor changes in the posterior fluxes during winter (0–2 Tmol/yr), while summer fluxes are increased by 5–7 Tmol/yr. In contrast, increasing the background by 15 ppb results in a strong summer reduction of up to 15 Tmol/yr, occasionally pushing posterior fluxes close to or even below zero. Effect on winter fluxes is minor (0–3 Tmol/yr).

This seasonal asymmetry can be explained by the fact that a fixed background offset represents a much larger relative change during summer, when the typical CO enhancements at the ICOS stations are lower (approximately 15 ppb), compared to winter (typically 35 ppb). Thus, the same absolute bias in the background leads to a much stronger relative distortion of the CO-based signal in summer.

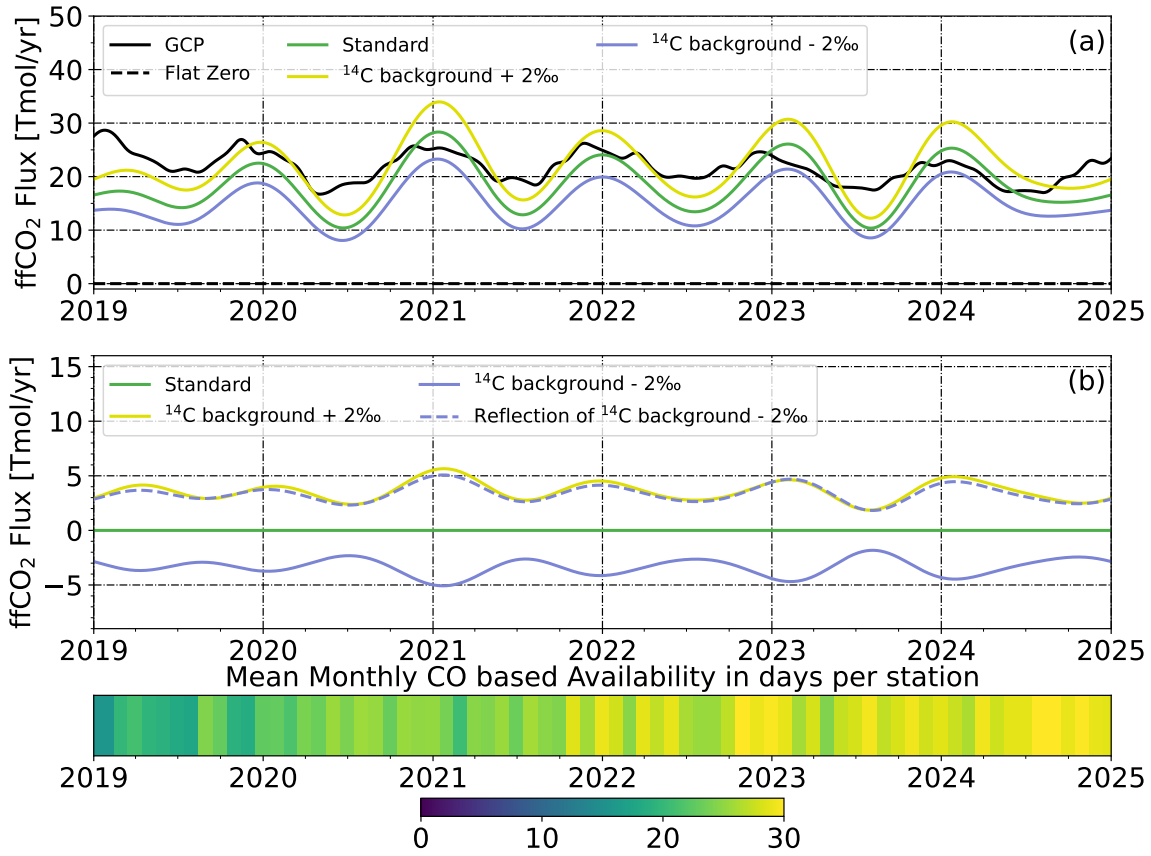

Figure 6.11. Posterior $\Delta ffCO_2$ emissions in the Germany+ domain based on CO-derived estimates using three different assumptions for the $\Delta^{14}C$ background: unmodified (standard, green), increased by 2‰ (yellow), and decreased by 2‰ (blue). Panel (a) shows the absolute fluxes, while panel (b) displays the deviations from the standard posterior. The dashed blue line in panel (b) represents the absolute value. The color bar indicates the mean monthly data availability in days per station.

This experiment highlights the high sensitivity of CO-based inversions to biases in the marine CO background. The strong seasonal cycle seen in the standard posterior fluxes may point to a summer overestimation of the background or suggest that a portion of the marine background signal does not reach the measurement stations—potentially due to atmospheric sinks or mixing processes—leading to an effective overestimation of the CO background during this time of year. Note, that the posterior winter estimates are less sensitive to biases in the CO background and can, accounting for this additional uncertainty sill be used to estimate winter fossil emissions in the Germany+ domain.

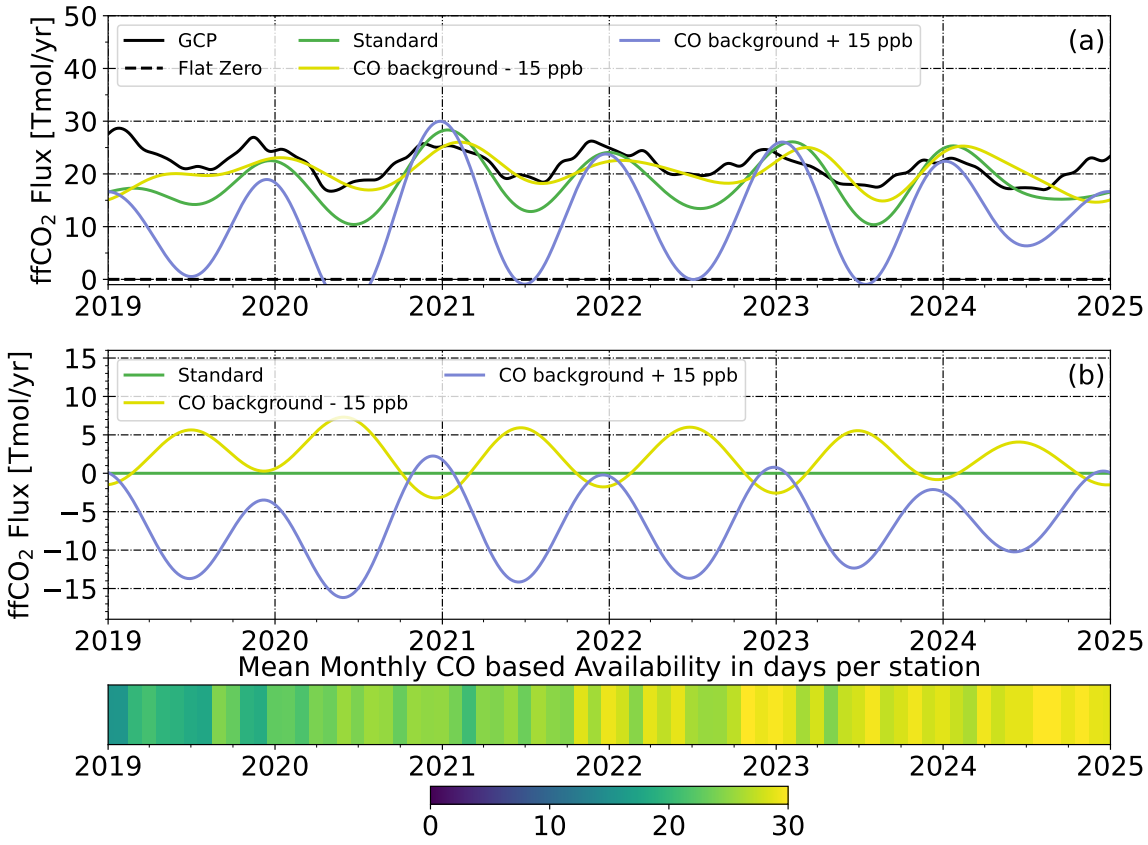

Figure 6.12. Posterior fossil fuel $\Delta ffCO_2$ emissions for the Germany+ domain based on CO-derived estimates using a Flat Zero prior and 40% prior uncertainty. Panel (a) shows posterior fluxes resulting from a standard CO background (green) and from backgrounds shifted by ±15 ppb (yellow and blue), along with the GCP inventory (black) as reference. Panel (b) presents the flux differences relative to the standard case. The color bar indicates the monthly average availability of CO-based estimates per station.

6.3.4 CO-based Bayesian inversions with unmodified GCP and EDGAR priors

Here, I present results from two Bayesian inversions using unmodified daily GCP and hourly EDGAR prior fluxes (Sect. 3.2.3), each with a 4% prior uncertainty (relative to the GCP emissions) consistent with reported uncertainties for larger domains (Janssens-Maenhout et al., 2019; Jones et al., 2022), and driven by CO-based $\Delta ffCO_2$ data. Similar to the flask-based inversions (Sect. 5.2.4), I focus on winter periods (November - February), excluding 2018/19 and 2024/25 to avoid boundary effects.

The winter prior fluxes as well as the inversion results in the Germany+ domain for the period with sufficient data density (Sect. 4.3) are depicted in Fig. 6.13 with the summer periods not being included in the evaluation (the grayed out periods in panel (a)). CO-based inversions exhibit patterns consistent with flask-based results (Sect. 5.2.4). Posteriors indi-

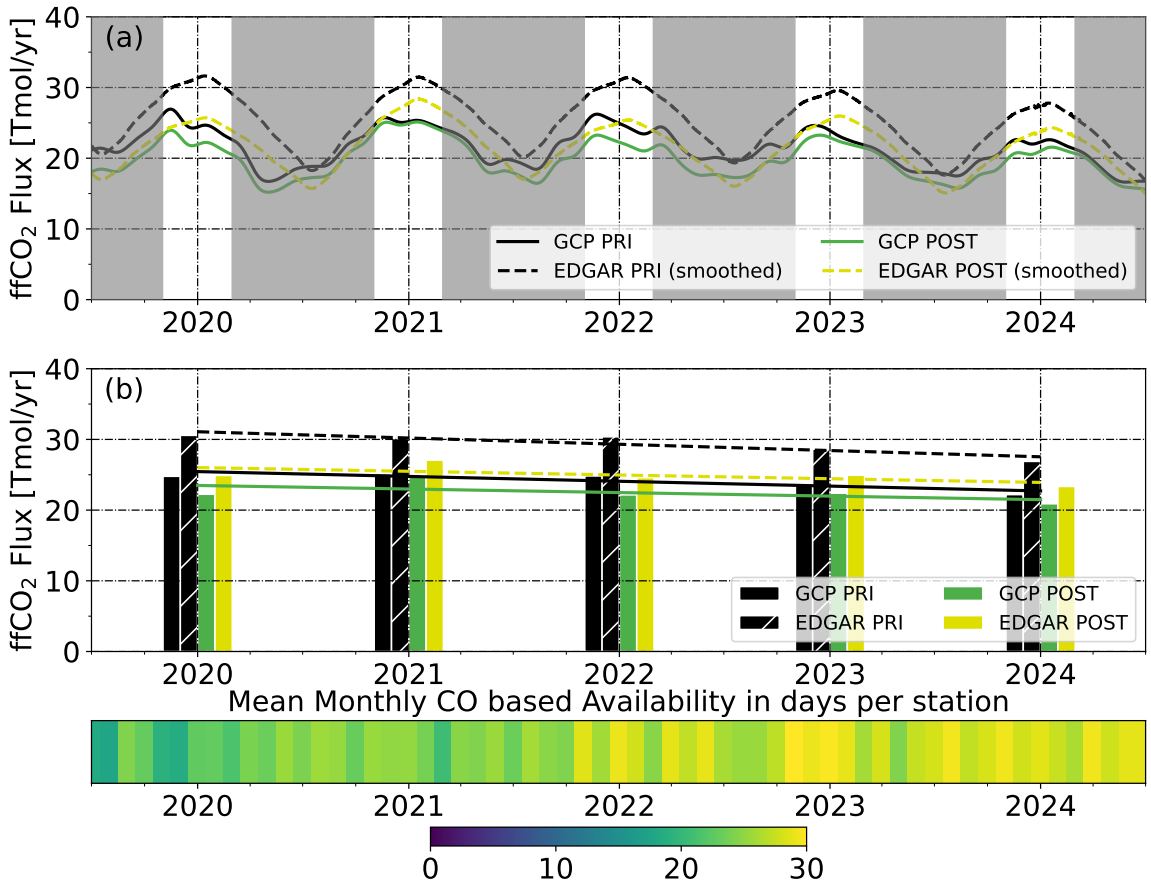

Figure 6.13. Wintertime fossil fuel emissions in the Germany+ domain from Bayesian inversions using CO-based $\Delta ffCO_2$ estimates and two different prior inventories: GCP (solid lines/bars) and EDGAR (dashed lines/hatched bars). Panel (a) shows smoothed daily (GCP) and hourly (EDGAR) prior and posterior fluxes; non-winter periods are shaded in gray. Winter periods are defined as November to February. Panel (b) displays the corresponding winter mean fluxes per winter season. The trends estimated from the mean winter emissions are shown as solid and dashed lines in the respective colors. The color bar indicates the mean monthly availability of CO-based estimates per station.

cate lower emissions than their corresponding priors (Tab. 6.8 and 6.9): posterior winter flux adjustments to the GCP range from 0.4-2.7 Tmol/yr (2-10%), while EDGAR adjustments are 3.1-5.8 Tmol/yr (12-23%) in the corresponding posterior. EDGAR winter emissions during the investigated period are 4.7-5.8 Tmol/yr (20-24%) higher than the GCP prior, over twice the difference between the posteriors (2.3-2.6 Tmol/yr, 9-11%), with EDGAR posterior values being consistently higher. Posterior uncertainty was reduced by 0.2 to 0.4 Tmol/yr, or 20-40% relative to the prior uncertainty. Mean winter posterior emissions of both priors consistently stayed within 3σ confidence range. For the EDGAR inventory, in particular, the inversion led to a substantial reduction in posterior fluxes, pushing them below 4% prior

Table 6.8. Winter mean $\Delta ffCO_2$ emissions (in Tmol/yr) and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective CO-based posterior inversions. Winter periods refer to November through February of the indicated winter season.

Flux type	Absolute fluxes, [Tmol/yr]				
	2019/20	2020/21	2021/22	2022/23	2023/24
GCP Prior	24.8(0.9)	25.1(0.9)	24.8(0.9)	23.5(0.8)	22.2(0.8)
GCP Posterior	22.2(0.7)	24.7(0.6)	22.2(0.6)	22.4(0.5)	20.9(0.6)
EDGAR Prior	30.6(1.0)	30.1(1.0)	30.4(1.0)	28.6(1.0)	26.9(0.9)
EDGAR Posterior	24.9(0.7)	27.1(0.7)	24.6(0.6)	24.9(0.6)	23.3(0.6)

Table 6.9. Deviations of winter mean $\Delta ffCO_2$ emissions from the GCP for the corresponding winter and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective CO-based posterior inversions. Winter periods refer to November through February of the indicated winter season.

Flux type	Deviation from GCP, [%]				
	2019/20	2020/21	2021/22	2022/23	2023/24
GCP Prior	0(4)	0(4)	0(4)	0(4)	0(4)
GCP Posterior	-10(3)	-2(2)	-10(2)	-5(2)	-6(3)
EDGAR Prior	23(4)	20(4)	23(4)	22(4)	21(4)
EDGAR Posterior	1(3)	8(3)	-1(2)	6(3)	5(3)

uncertainties. Consistent underestimation of the prior fluxes can be explained by potential ^{14}C background bias (see Sect. 8.1), which for the CO-based data would influence the whole data series through the $\Delta CO/\Delta ffCO_2$ ratio. However, the emission reductions in the EDGAR-based posteriors still would be significant, suggesting potential overestimation of emission fluxes by the EDGAR inventory in CO-based inversions.

Mean winter emission reduction trends over five years, calculated as linear regression slopes (Fig. 6.13, panel (b)), generally align between inventories and posteriors. GCP reports an average annual emission reduction of 0.7(0.2) Tmol/yr, while EDGAR suggests a slightly higher rate of 0.9(0.3) Tmol/yr. Both posteriors indicate slightly lower reduction rates of 0.5(0.4) Tmol/yr.

Even with a five-year period, posterior emission trend uncertainties remain high, preventing reliable validation of inventory-based mean winter emission trends. Notably, in-

ventory slope errors are significantly higher here (31% for GCP, 28% for EDGAR) than for the last three years (2% for GCP, 0.5% for EDGAR, Sect. 5.2.4) due to COVID-related flux anomalies (sharp decrease in 2020, subsequent increase in 2021, and further reductions thereafter). Posterior emission trend uncertainties are approximately 84% and 71%. While high posterior uncertainties might be partially attributed to COVID-induced flux variations, other factors such as ^{14}C and CO background biases (Sect. 6.3.3.4) and CO to $\Delta ffCO_2$ ratio inhomogeneities in station catchment areas (e.g., Sect. 6.3.3.3) also contribute. Nevertheless, data-driven posteriors align more closely with GCP than EDGAR, suggesting overall lower winter emissions in the AoC.

Chapter 7

Integral-based $\Delta ffCO_2$ inversions

7.1 Introduction and motivation

Integral sampling represents a third complementary approach for estimating fossil fuel CO_2 ($\Delta ffCO_2$) fluxes, alongside flask-based (Sect. 5) and CO-based (Sect. 6) methods. Unlike flask sampling, which targets specific times or meteorological conditions, integral samplers continuously collect air for longer periods, usually two weeks. This results in highly representative samples that average over full synoptic cycles and diverse atmospheric transport conditions.

A key strength of this approach is its reduced sensitivity to short-term meteorological misrepresentation in transport models, since the time-averaged nature of the samples inherently smooths out high-frequency variability. This leads to greater robustness in the posterior estimates, particularly when the number of samples is limited. Moreover, integral samples are not subject to the selection bias introduced by targeted flask sampling, making them well suited to derive $\Delta ffCO_2$ fluxes on climatological scales.

However, integral sampling also introduces specific challenges. Continuous collection over long intervals dampens high-frequency signals, thereby lowering the signal-to-noise ratio (SNR) of fossil fuel enhancements. As a result, small emission events may become undetectable. Additionally, because sampling is non-selective, the potential for contamination from nuclear power plant emissions is higher compared to targeted flask sampling, where meteorological filtering can be applied. This increases the nuclear-caused uncertainty of the $\Delta ffCO_2$ estimates.

A distinct feature of integral sampling is that it captures the entire diurnal cycle of atmospheric CO_2 , including conditions at night and early in the morning that are typically excluded from atmospheric inversion systems. This improves the representativeness of the sampled air masses but raises the question of how well transport models, whose evaluation is often limited to afternoon observations when mixing is strongest, can accurately simulate the

$\Delta ffCO_2$ signal in these time-averaged samples. Since inversion frameworks generally assume that the transport error is smallest during well-mixed afternoon periods, using 24-hour averaged observations might introduce additional model-data mismatch (MDM), potentially limiting the interpretability of the results.

In this chapter, I first present a comparison between the observed integral-based $\Delta ffCO_2$ signals and those simulated using forward atmospheric modeling (Sect. 7.2.1). This analysis serves as a consistency check between the measured and expected signals. I then apply the Evaluation Matrix inversion framework, identical to the one used in previous chapters for flask-based (Sect. 5.2.2) and CO-based (Sect. 6.3.2) data, to derive posterior $\Delta ffCO_2$ fluxes over the Germany+ domain (Sect. 7.2.2). The resulting posterior is compared to both flask-based and CO-based posteriors to assess the strengths and limitations of the integral data set in constraining fossil fuel emissions. Particular attention is paid to seasonal performance, spatial representativeness, and the sensitivity of posterior fluxes to key parameters such as the assumed background concentration (Sect. 7.2.3). Finally, I perform two Bayesian inversions with unmodified GCP and EDGAR prior fluxes with realistic prior uncertainties (Sect. 7.2.4).

7.2 Results for integral-based $\Delta ffCO_2$ estimates

7.2.1 Forward runs and model-data-mismatch

Two-weekly integral-based $\Delta ffCO_2$ estimates for CBW, HPB, and HTM sites (additional stations in Sect. A.6) are presented in Figure 7.1. These are shown alongside the forward modeled estimates, calculated using the GCP inventory and the STILT transport model (Sect. 3.2.1), and aggregated according to integrated sample times.

Most stations exhibit relatively low MDM between -0.3 ppm and 0.5 ppm, presented in the Table 7.1 as $\bar{X}_{sim} - \bar{X}_{obs}$, indicating good correspondence between observation-based and modeled estimates. However, CBW, LIN, and SAC stations show significantly higher positive MDM of 1.3 ppm, 1.3 ppm and 3.3 ppm, respectively. Although these stations are located near major fossil emission areas (CBW – Rotterdam, LIN – Berlin, SAC – Paris), their measurements suggest much lower integrated $\Delta ffCO_2$ concentrations than the corresponding simulations. For stations with elevated simulated signals, such as CBW, KIT, and SAC (consistent with flask-based, Sect. 5.2.1, and CO-based, Sect. 6.3.1, $\Delta ffCO_2$ observations), the variances of MDM ($\sigma_{sim-obs}$) are much higher than the mean measurement errors ($\bar{e}$): 1.8 ppm, 2.7 ppm and 2.6 ppm opposed to 1.4 ppm, respectively. This suggests a higher sensitivity to transport errors due to their proximity to point sources. In combination with potentially large heterogeneity of the $\Delta ffCO_2$ sources in the catch-

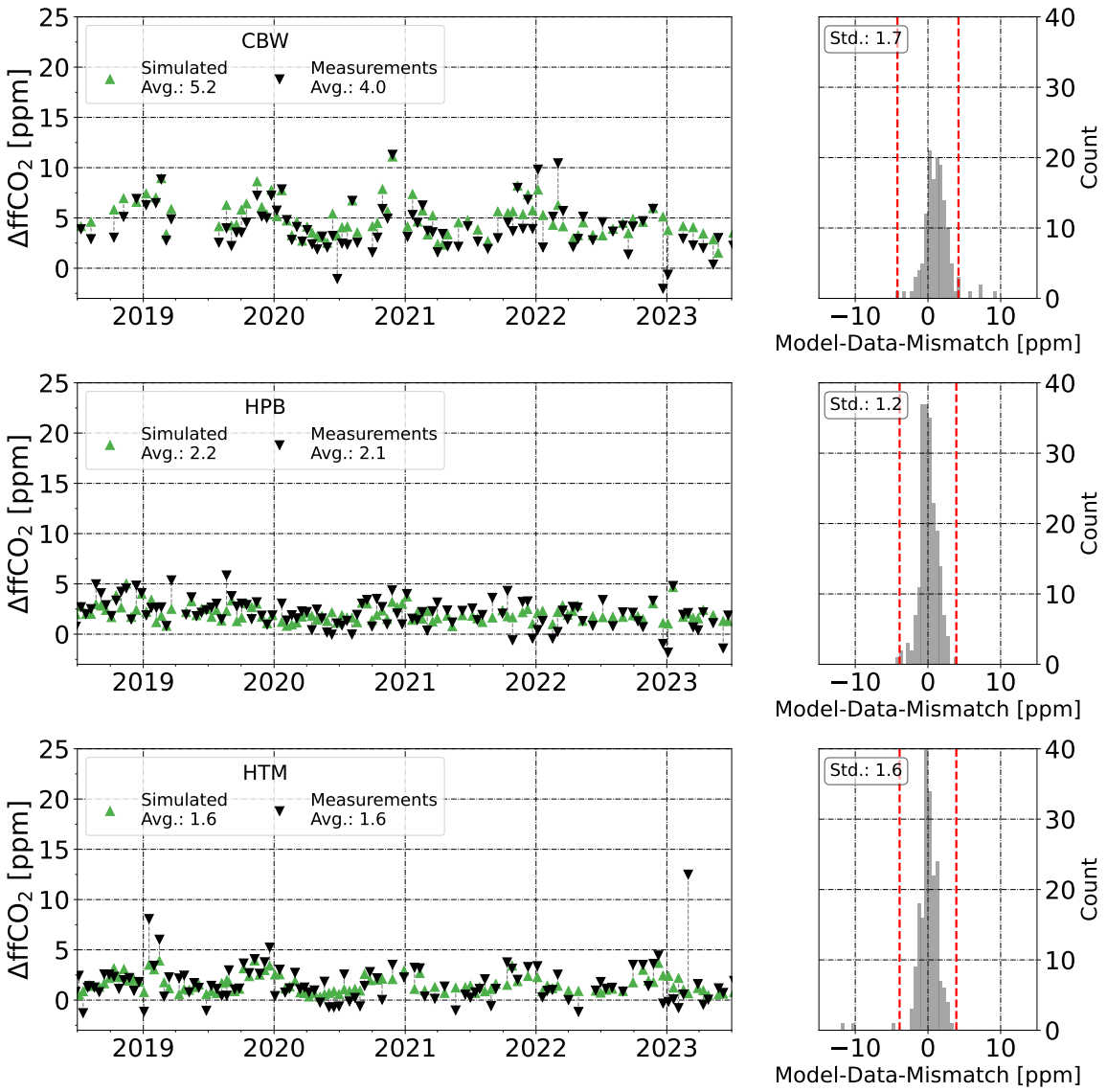

Figure 7.1. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed lines. The average signal over the depicted time period is noted in the legend. The panels on the right demonstrate MDM distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sect. 4.3).

ment area for LIN (previously noted in flask-based, Sect. 5.2.3.2, and CO-based inversions, Sect. 6.3.3.2), this could explain the high MDM at these sites. For other stations, mean measurement errors $\bar{e}$ (1.2-1.4 ppm) generally exceed $\sigma_{sim-obs}$ (0.9-1.4 ppm), implying that measurement errors alone could account for the observed discrepancy between model and observations (except for HTM with an MDM variance of 1.6 ppm, likely due to a few high $\Delta ffCO_2$ outliers). The SNR for most ICOS stations ranges from 1 to 2, but can reach 3.9

Table 7.1. Statistics of the integral-based observed and simulated $\Delta ffCO_2$ concentrations in the ICOS network from 2018-2023. $\bar{X}_{obs}$ and $\bar{X}_{sim}$ denote mean measured and simulated concentrations for the corresponding site. $\bar{X}_{sim} - \bar{X}_{obs}$ and $\sigma_{sim-obs}$ denote the mean MDM and its standard deviation. $\bar{e}$ and SNR denote RMSD-based measurement error and mean Signal-to-Noise Ratio. $|\bar{X}_{sim} - \bar{X}_{obs}| > 3\bar{e}$ shows the approximate portion of the data with higher absolute MDM then the triple of $\bar{e}$. (*) KIT was not used in regular inversions, only in sensitivity studies.

Site	$\bar{X}_{obs}$ [ppm]	$\bar{X}_{sim}$ [ppm]	$\bar{X}_{sim} - \bar{X}_{obs}$ [ppm]	$\sigma_{sim-obs}$ [ppm]	$\bar{e}$ [ppm]	$\lvert\bar{X}_{sim} - \bar{X}_{obs}\rvert > 3\bar{e}$ [%]	SNR [rel.]
CBW	4.0	5.3	1.3	1.8	1.4	10%	2.9
GAT	2.1	2.2	0.1	1.1	1.3	0%	1.6
HPB	2.1	2.2	0.1	1.2	1.3	0%	1.6
HTM	1.7	1.6	-0.1	1.6	1.3	0%	1.3
JFJ	0.6	1.0	0.4	1.1	1.3	0%	0.4
KIT*	5.5	5.6	0.1	2.7	1.4	10%	3.9
KRE	2.4	2.3	-0.1	1.0	1.3	0%	1.8
LIN	3.4	4.7	1.3	1.2	1.3	0%	2.6
MHD	-0.1	0.1	0.2	0.9	1.2	0%	0.0
NOR	0.4	0.9	0.5	1.0	1.3	0%	0.3
OPE	1.9	2.4	0.5	1.2	1.3	0%	1.5
OXK	2.1	2.5	0.4	1.1	1.3	0%	1.6
PAL	0.6	0.3	-0.3	1.2	1.3	0%	0.5
SAC	2.2	5.5	3.3	2.6	1.4	40%	1.6
SSL	1.7	1.8	0.1	1.2	1.3	0%	1.3
STE	3.2	3.1	-0.1	1.2	1.3	0%	2.4
SVB	0.5	0.4	-0.1	1.0	1.3	0%	0.4
TRN	1.6	2.0	0.4	1.4	1.4	0%	1.1

for KIT or drop below 1 for clean air or background sites (JFJ, MHD, NOR, PAL, SVB). A comparison of the observed average $\Delta ffCO_2$ concentrations for integral-based and CO-based measurements (Tab. 6.3) reveals a relatively good match for the corresponding sites, with the exception of LIN.

As with other proxies, KIT station was excluded from the standard inversion station setup due to potential on-site $\Delta^{14}C$ contamination, but its influence, as well as that of other stations, on the posterior will be investigated in Section 7.2.3.2. For the final inversion with unmodified priors (Sect. 7.2.4), SAC and LIN sites were additionally removed from the core

station set to avoid any potentially biased sources. The SAC station was excluded because its high MDM and $\sigma_{sim-obs}$ values suggest potential ^{14}C contamination from the Saclay Nuclear research facility in its catchment area. The LIN station was removed as it is likely influenced by large, brown coal power stations, "Jänschwalde" (approx. 80 km southeast) and "Schwarze Pumpe" (approx. 100 km southeast), which are likely not properly resolved by the transport model (see Sect. 7.2.3.2). These stations were, however, included in the "Evaluation Matrix" inversions.

7.2.2 Application of the Evaluation Matrix to the integral-based inversions

Following the methodology applied to flask-based (Sect. 5.2.2) and CO-based (Sect. 6.3.2) inversions, integral-based posteriors were analyzed according to the Evaluation Matrix as described in Sect. 3.2.4. This analysis aims to assess features such as seasonality and inter-annual trends of fossil CO_2 emissions embedded in the measurement data. Deviations from the GCP emission inventory were classified for significance using results from the synthetic experiments (Sect. 4.6).

7.2.2.1 Seasonal cycle of $\Delta ffCO_2$ emission trend derived from integral-based inversions

The seasonal cycle of the integral-based posterior depicted in Figure 7.2 (green line) largely corresponds to the GCP inventory (black line). Significant deviations in the winter emission maxima are observed only for the 16/17, 19/20, and 22/23 winters, which is reflected in the calculated relative amplitude difference of over 15% denoted in Table 7.2. For summers within the main investigation period (01.07.2018 – 01.07.2023), only 2020 and 2023 significantly deviate from the GCP inventory (relative amplitude difference over 20%). However, all summer minima in the secondary investigation period underestimate GCP emissions by 25% or more. This potentially suggests that, due to the lower SNR in summer, more data is required to sufficiently constrain the inversion system. The positions of the extrema are generally better captured by the measurement data, with only two winter (2018/19 and 2022/23) and two summer (2015 and 2023) periods showing significantly high deviations from the GCP extrema positions of 50 days or more, as expected from a 1.5 ppm observational uncertainty derived in Section 4.6.1 which resulted in a 1σ standard deviation for the position of the extrema of 35-40 days and a 1σ variation in the value for the relative amplitude of 0.08 to 0.10.

Observed deviations from the GCP inventory could not be attributed to any single measurement site (see Sect. 7.2.3.2). Deviations, such as the high amplitude in winter 16/17

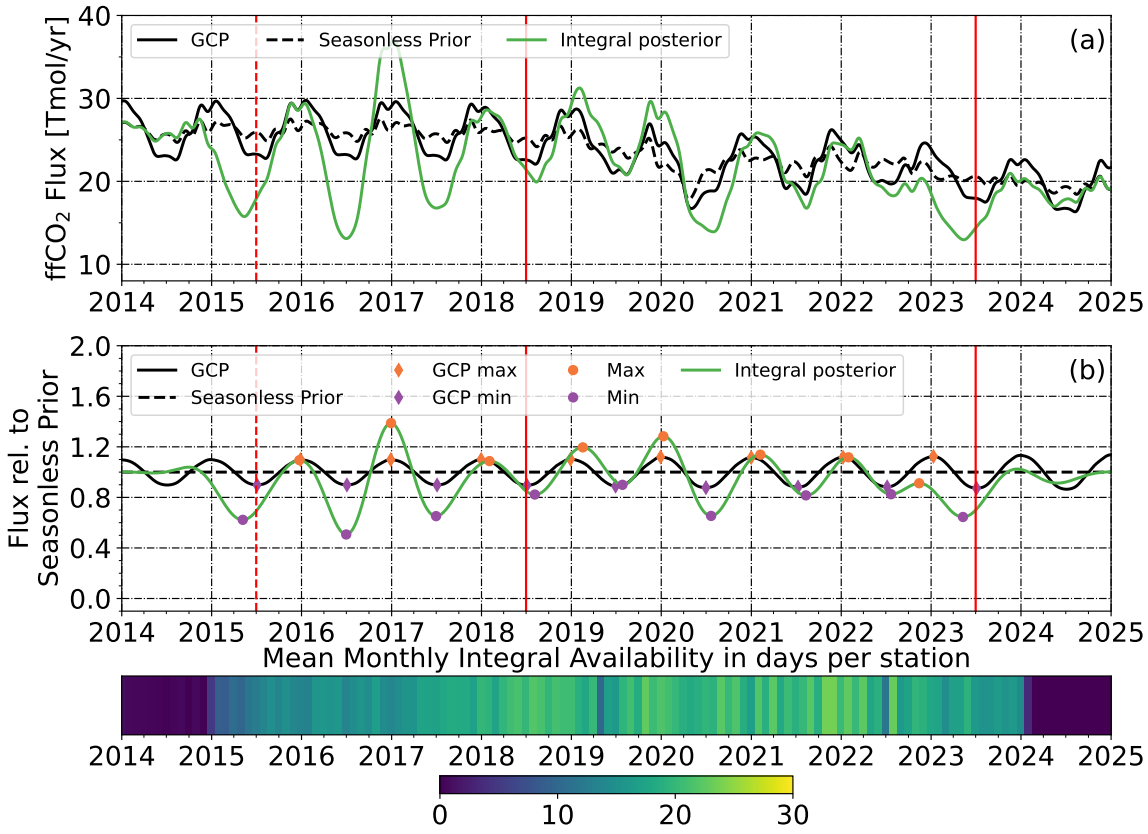

Figure 7.2. Results of the seasonal cycle evaluation of the integral-based posterior $\Delta ffCO_2$ fluxes. The absolute (panel (a)) and relative to the Seasonless Prior (panel (b)) fluxes in the Germany+ domain show the GCP inventory (black solid line), the Seasonless Prior (used as input, black dashed line) and the posterior results (green line). Orange and purple circles mark the maxima and minima of the posterior, with corresponding GCP extrema marked by diamonds. Red lines indicate the main investigation time frame, with a dashed red line for the secondary analysis period. The color bar at the bottom indicates the network-wide mean monthly availability of integral-based estimates per station.

and phase shifts in winter 18/19 (coinciding with a local data density minimum, see color bar in Fig. 7.2) and in summer 2015, could potentially be explained by insufficient data density. Other deviations may suggest systematic differences between measurement data and the inventory. For instance, absolute posterior fluxes in winter 19/20 showed values similar to the GCP inventory of the previous winter 18/19, whereas the GCP depicted a very skewed winter emission shape in 19/20 due to a much earlier emission decrease caused by the Covid-19 lockdown. Furthermore, the integral-based posterior suggests overall lower emissions in the summer with a later minimum compared to the GCP. These unique circumstances may lead to increased errors in the disaggregation of national totals (assimilated into emission inventories), as human and economic behavioral patterns significantly differed from business-as-usual scenarios.

Table 7.2. Positions (Date) and relative amplitudes (Ampl., relative to the Seasonless Prior) of the seasonal extrema extracted from the GCP inventory and the posterior of the integral-based inversion, as well as the difference between them.

Season	GCP		Integral-based $\Delta ffCO_2$		Posterior – GCP	
	Date	Ampl. [rel.]	Date	Ampl. [rel.]	Phase shift [days]	Ampl. diff. [rel.]
Winter 15/16	27.12	1.10	23.12	1.09	-4	-0.01
Winter 16/17	29.12	1.10	29.12	1.39	0	0.29
Winter 17/18	30.12	1.10	02.02	1.09	34	-0.01
Winter 18/19	28.12	1.10	16.02	1.20	50	0.10
Winter 19/20	30.12	1.12	09.01	1.28	10	0.16
Winter 20/21	01.01	1.12	06.02	1.14	36	0.02
Winter 21/22	08.01	1.12	30.01	1.12	22	0.00
Winter 22/23	11.01	1.12	14.11	0.91	-58	-0.21
Summer 2015	04.07	0.90	07.05	0.62	-58	-0.28
Summer 2016	03.07	0.90	30.06	0.51	-3	-0.39
Summer 2017	04.07	0.90	01.07	0.65	-3	-0.25
Summer 2018	04.07	0.90	05.08	0.82	32	-0.08
Summer 2019	29.06	0.89	27.07	0.90	28	0.01
Summer 2020	01.07	0.88	21.07	0.65	20	-0.23
Summer 2021	09.07	0.88	11.08	0.82	33	-0.06
Summer 2022	07.07	0.88	23.07	0.83	16	-0.05
Summer 2023	04.07	0.87	11.05	0.65	-54	-0.22

Another prominent deviation from the GCP inventory was observed in winter 22/23 and summer 2023. A similar, albeit less pronounced, effect was previously observed in the flask-based inversions (Sect. 5.2.2.1). There, the winter 22/23 posterior flux maximum was also shifted to earlier dates with a lower overall amplitude, and the summer 2023 minimum was considerably lower compared to the GCP. As demonstrated in the previous chapters, the extrapolation of the marine $\Delta^{14}C$ background — and potential biases therein — has had a significant impact on posterior flux estimates. In Sect. 7.2.3.1, this sensitivity will be investigated for the integral samples as well.

Overall, taking into account the exceptional circumstances of the COVID-19 pandemic in 2020 and a potential bias in the extrapolated $\Delta^{14}C$ background in 2023, I conclude that

the position and amplitude of the seasonal cycle in the integral-based $\Delta ffCO_2$ posterior do not significantly deviate from the GCP inventory.

7.2.2.2 Inter-annual $\Delta ffCO_2$ emission trend derived from integral-based inversions

The integral-based $\Delta ffCO_2$ data provide a valuable foundation for investigating long-term trends in fossil fuel emissions. Figure 7.3 shows posterior fluxes from the F1T and F0.5T inversions compared to the GCP inventory and the Trendless prior, analogous to the inter-annual trend investigation of synthetic data (Sect. 4.6.2). In this section, linear trends are fitted to the posterior time series to quantify emission changes over the most recent period (01.07.2018–01.07.2023), with particular attention to the influence of 2023.

For the full period from 2018 to 2023, the trend in absolute fluxes (Fig. 7.3, panel (a)) is negative for all time series but differs in magnitude: GCP shows a moderate decrease of –0.74±0.04 Tmol/yr, whereas the F1T and F0.5T posteriors indicate significantly stronger declines of –1.97±0.04 Tmol/yr and –1.88±0.03 Tmol/yr, respectively. The relative trends (Fig. 7.3, panel (b)) compared to the Trendless Prior yield consistent results: –2.8%/yr for GCP versus –7.4%/yr (F1T) and –7.1%/yr (F0.5T). Compared to the synthetic studies (Sect. 4.6.2), such difference in the trends is higher than the difference expected just from the influence of the measurement errors (∼0.3 Tmol/yr). These differences underline a pronounced sensitivity of the inversions to the data at the end of the time series.

To assess this further, the same trend analysis is repeated excluding the final year 2023 (not shown). For the reduced period (01.07.2018–01.07.2022), the posterior trends in absolute fluxes shift to –1.60±0.06 Tmol/yr (F1T) and –1.73±0.05 Tmol/yr (F0.5T), aligning more closely with the GCP trend of –0.78±0.06 Tmol/yr. The relative slopes also become more consistent: –2.9%/yr (GCP), –6.0%/yr (F1T), and –6.5%/yr (F0.5T). The difference is, however, still higher than the expected variation based solely on the measurement errors (∼0.3 Tmol/yr, Sect. 4.6.2).

Visual inspection of panels (a) and (b) in Figure 7.3, on the other hand, would generally support the claim that the integral-base posteriors are suitable for trend analysis: the F0.5T posterior closely tracks the GCP trend until well into 2022. The sharp downward deviation starting in 2023 dominates the full-period slope and appears inconsistent with prior behavior. Given the two-year temporal correlation length applied in the F0.5T setup, it is likely that potential anomalies — such as a biased $\Delta^{14}C$ background in 2023 — also affect the posterior estimates in late 2022. Here is also evident that the difference in the trend slopes (with excluded last year) comes solely from the high 2019 summer emission estimates in the posteriors, which can be influenced by the local minimum in the data density (colorbar in Fig. 7.3).

Taken together, these findings highlight the strong influence of the year 2023 on inferred

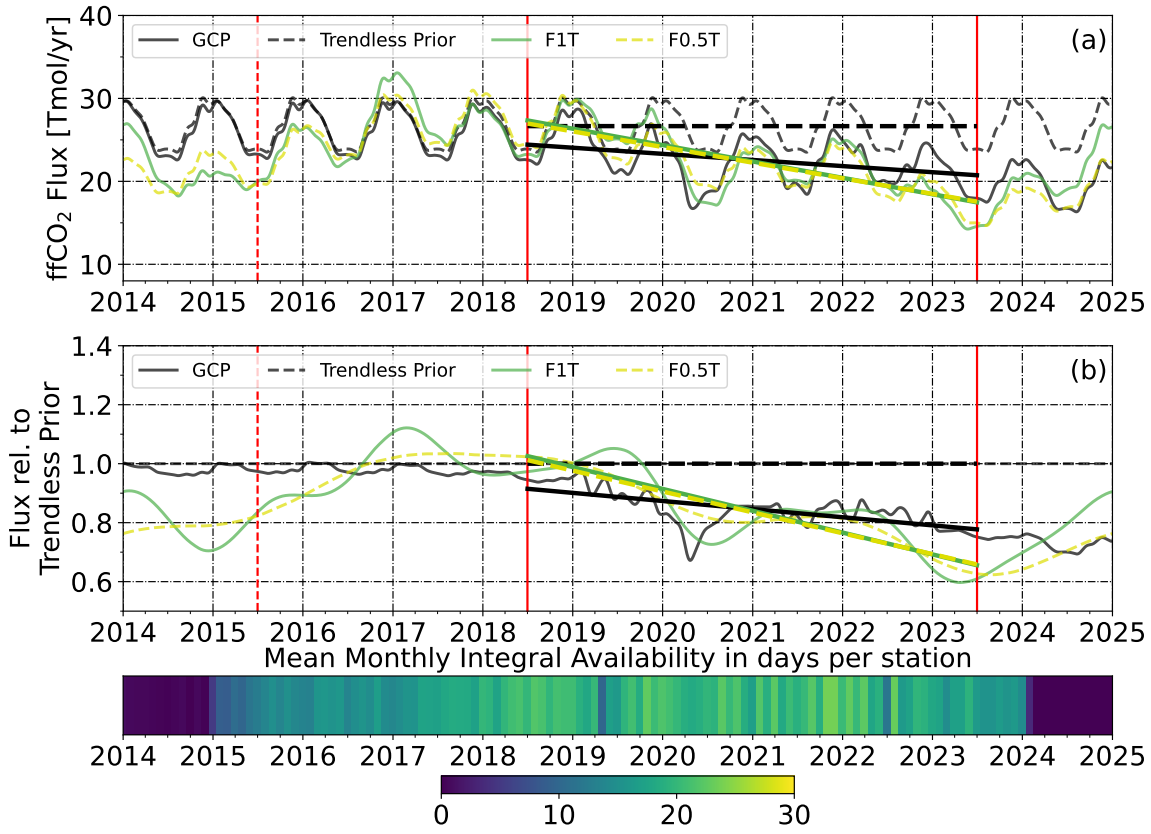

Figure 7.3. Results of the inter-annual trend analysis for the integral-based inversions. The absolute (panel (a)) and relative to the Trendless Prior (panel (b)) fluxes in the Germany+ domain show the GCP inventory (black solid line), the Trendless Prior (used as input, black dashed line) and posterior results for F1T (green solid line) and F0.5T (yellow dashed line) filters. The straight lines show linear regressions of the correspondingly colored flux curves. Red lines indicate the main investigation time frame, with a dashed red line for the secondary analysis period. The color bar at the bottom indicates the network-wide mean monthly availability of integral-based estimates per station.

multi-annual trends and emphasize the need for caution when interpreting trend magnitudes. The consistency between posterior and GCP trends prior to 2023, especially in the F0.5T inversion, demonstrates the potential of integral samples to resolve long-term fossil fuel emission changes — provided sufficient data quality and stable background assumptions.

7.2.2.3 Absolute and annual mean $\Delta ffCO_2$ emissions derived from integral-based inversions

The absolute and annually aggregated fossil fuel CO_2 fluxes derived from inversions based on integral $\Delta ffCO_2$ data are summarized in Figure 7.4. As part of the sensitivity analysis, inversions were conducted using a Flat Zero and a Flat 200p Prior, each with 40% prior

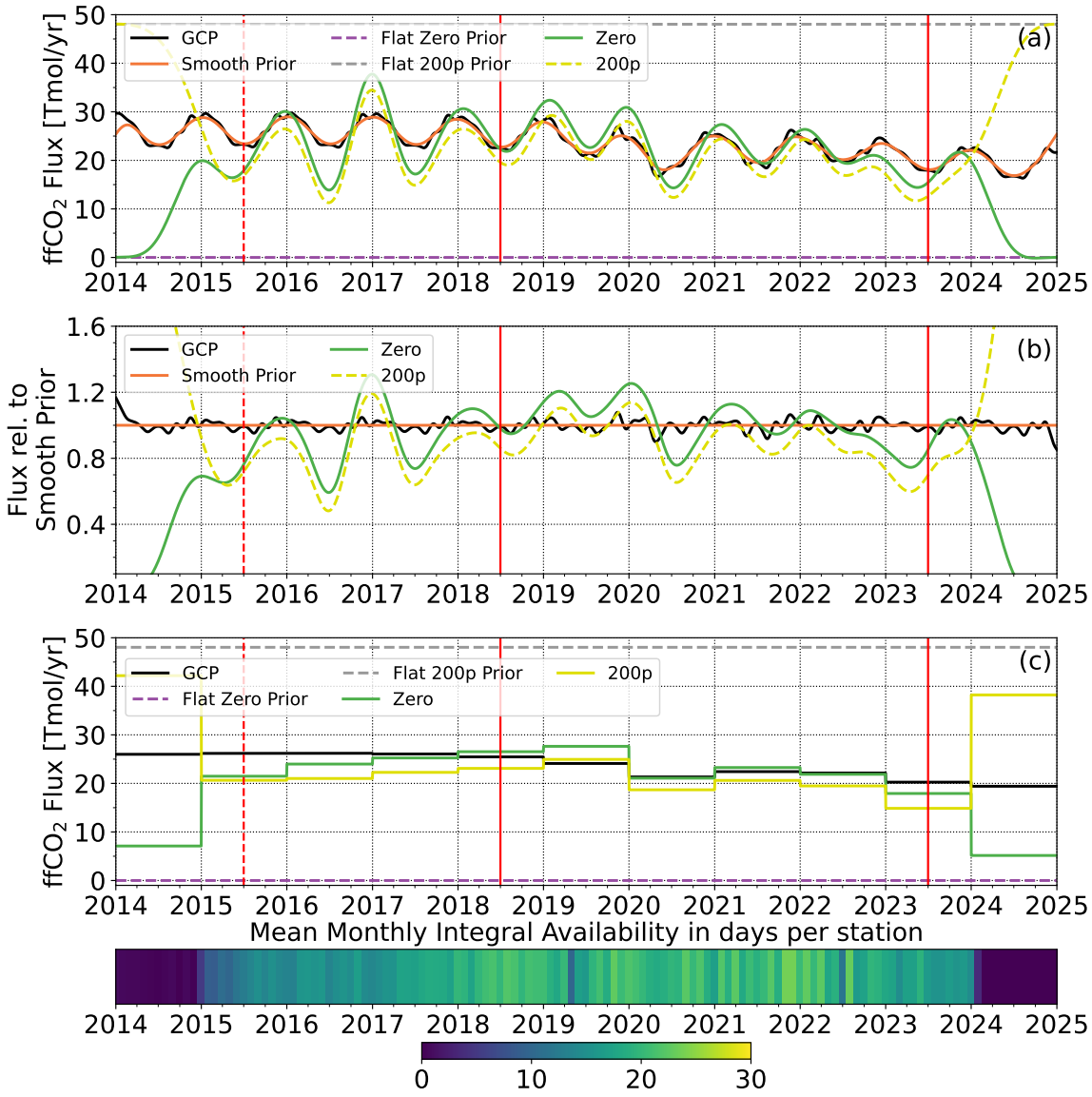

Figure 7.4. Results of the analysis of absolute and annually aggregated emission fluxes for the integral-based inversions in the Germany+ domain. Panel (a) shows the absolute and panel (b) the relative to the Smooth Prior $\Delta ffCO_2$ fluxes. Panel (c) shows yearly aggregated emissions from the panel (a). The Flat Zero Prior and the Flat 200p Prior (used as inputs, purple and gray dashed lines, respectively) as well as the Smooth Prior (orange solid line) are shown alongside posterior results (green solid and yellow dashed lines). Red lines indicate the main investigation time frame (01.07.12-01.07.23), with a dashed red line for the secondary analysis period (from 01.07.15). The color bar indicates the mean monthly availability of integral-based estimates per station.

uncertainty. The systematic behavior observed in the synthetic flat-prior experiments, the so-called over- and undershoot effect (Sect. 4.6.3), emerges again: posterior emissions tend to overestimate the true values when starting from a low-emission prior (Flat Zero), and

Table 7.3. Annual mean aggregated posterior integral-based $\Delta ffCO_2$ fluxes relative to the GCP inventory. Here I consider 2018-2022 to be the core evaluation period. 2015 and 2023 are not investigated as the sufficiently high data density was not maintained for the whole year.

Prior	2016	2017	2018	2019	2020	2021	2022
Flat Zero	0.92	0.97	1.04	1.15	0.99	1.04	0.99
Flat 200p	0.80	0.85	0.91	1.03	0.88	0.92	0.88

underestimate them when starting from a high-emission prior (Flat 200p) (Fig. 7.4, panel (a)).

To account for this effect, the real-data posteriors were interpreted in the context of the synthetic benchmark results (not shown). After considering the mean bias observed in the synthetic experiments ($\sim$7% overestimation for the Flat Zero Prior and $\sim$6% underestimation for the Flat 200p Prior, Sect. 4.6.3), it becomes clear that even with very limited prior information in form of the Flat Zero and Flat 200p Priors, the integral-based posterior estimates align remarkably well with the GCP inventory for the years with good data coverage and available $\Delta^{14}C$ background measurements at the MHD station (notably 2018–2022). Generally, the normalized RMSD for the main investigation period between the posteriors and the Smooth Prior (0.11 for Flat Zero and 0.13 for Flat 200p) can be entirely explained by the measurement errors (0.11 ± 0.02 in synthetic experiments, Sect. 4.6.3).

The annual fossil fuel emission estimates generally deviate from the GCP values by only a few percent (within $\pm$9%) as shown in Table 7.3 and Figure 7.4, panel (c), while a deviation of approximately 6% can already be expected due to the observational uncertainty of 1.5 ppm. Larger deviations occur in the years 2016–2017 and 2019. The underestimations in 2016–2017 coincide with periods of limited data availability. The discrepancy in 2019 is mainly attributed to a shifted seasonal cycle, particularly the timing of the 2018/19 winter maximum. In 2020, the year of the COVID-19 lockdown, posterior estimates suggest a more pronounced drop in fossil emissions than reported in the GCP inventory. Again, the year 2023 shows particularly low $\Delta ffCO_2$ emissions, which will be further discussed in the ^{14}C background sensitivity studies (Sect. 7.2.3.1).

In summary, once the systematic prior dependence identified in the synthetic tests is considered, the integral-based inversions prove to be highly capable of recovering general and annual fossil emission trends during periods with sufficient observational constraints. This underscores the utility of integral sampling for assessing absolute fossil CO_2 emissions in the Germany+ domain.

7.2.3 Uncertainty investigation

7.2.3.1 Sensitivity of the posterior fluxes to the European $\Delta^{14}C$ background estimate

As discussed in previous chapters, the European $\Delta^{14}C$ background is a key component in the determination of $\Delta ffCO_2$ values (see Sect. 3.1.1). Its extrapolation beyond spring 2022, when regular measurements at the Mace Head (MHD) background station were discontinued, introduces a potential source of systematic bias, particularly from 2023 onward. In both the flask-based and CO-based $\Delta ffCO_2$ inversions, the possibility of such a bias was considered as a likely explanation for deviations from inventory trends observed in recent years. In this section, I systematically assess the sensitivity of the integral-based $\Delta ffCO_2$ inversions to such background uncertainties.

The sensitivity experiment follows the same methodology as described in Section 5.2.3.1 for flask-based and Section 6.3.3.4 for CO-based inversions, in which the European $\Delta^{14}C$ background curve was shifted by $\pm 2‰$, reflecting the approximate uncertainty range in the background estimate. Inversions were repeated using the Flat Zero Prior and a prior uncertainty of 40%, with all other parameters kept constant. Figure 7.5 presents the resulting posterior fossil fuel CO_2 fluxes.

Similarly to the results of the flask-based experiment, a positive background offset ($+2‰$) leads to elevated $\Delta ffCO_2$ values and thus higher inferred fluxes. Consequently, the $-2‰$ background shift reduces the estimated fluxes. However, the amplitude of this effect is even more pronounced in the case of integral data. Panel (b) in Figure 7.5 reveals seasonal deviations of up to 10 Tmol/yr in summer, which corresponds to flux differences of more than 40% relative to the seasonal mean. During winter, the flux changes induced by the background shifts remain smaller but are still substantial, ranging from approximately 5 to 10 Tmol/yr.

The heightened sensitivity of the integral-based inversions can be attributed to the averaging nature of the sampling. As the ^{14}C signal is integrated over two-week periods, short-term high-concentration episodes are smoothed out, effectively reducing the signal-to-noise ratio of the $\Delta ffCO_2$ estimates. As a result, a given absolute bias in $\Delta^{14}C$ translates more directly into flux space.

The posteriors respond nearly symmetrically to the $\pm 2‰$ background shift, with only minor deviations occurring near the boundaries of the time series and during periods of limited data availability. In these cases, the posterior solutions tend to drift towards the Flat Zero Prior due to the reduced observational constraint. However, during the core investigation period (2018-2022), the differences between the $+2‰$ and the mirrored $-2‰$ posteriors remain small, indicating an approximately linear and symmetric system response to variations in the $\Delta^{14}C$ background.

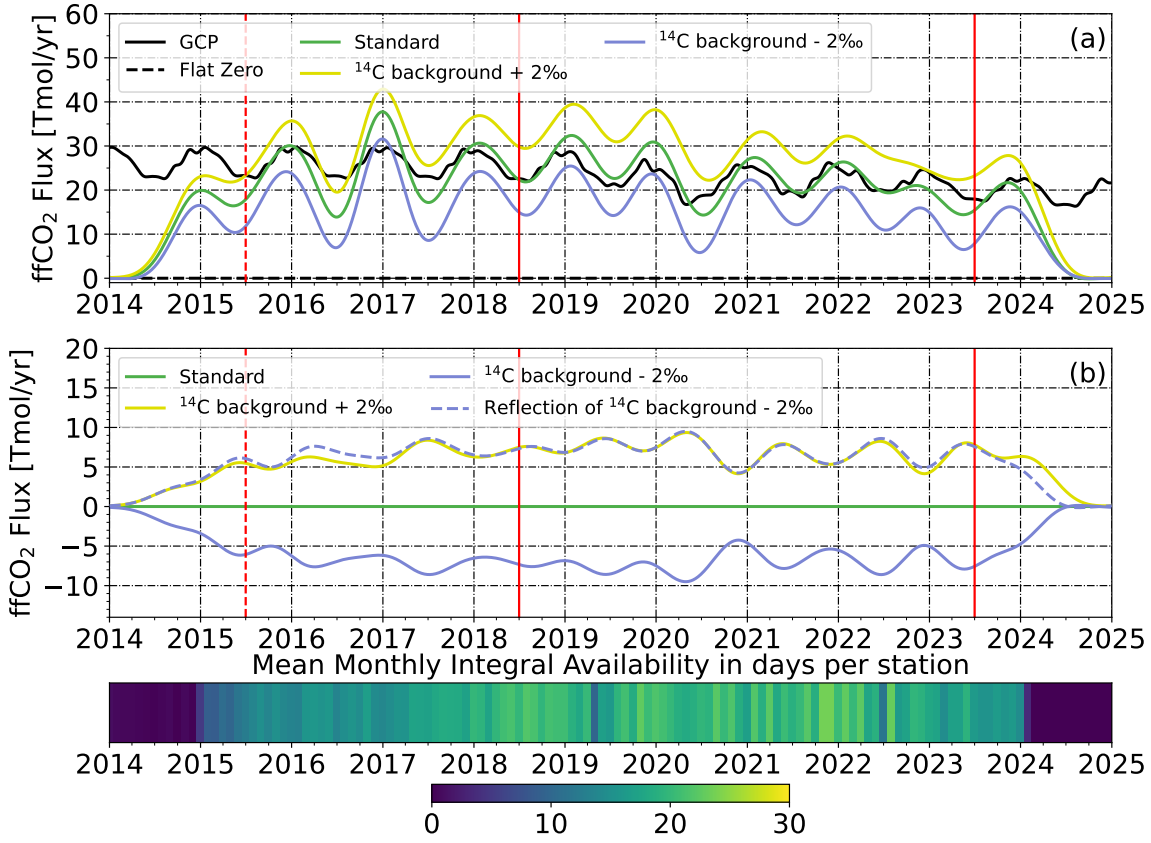

Figure 7.5. Sensitivity of the posterior fossil fuel CO_2 fluxes derived from integral-based observations to a $\pm$2‰ shift in the European $\Delta^{14}C$ background. Panel (a) shows absolute fluxes for the GCP inventory (black), the Flat Zero prior (black dashed), and the three posterior scenarios: standard background (green), $-$2‰ (yellow), and $-$2‰ (blue). Panel (b) displays the flux differences between the standard and perturbed posteriors. The dashed blue line shows the reflected difference of the $-$2‰ case for comparison. Vertical red lines indicate the boundaries of the main investigation period. The bottom color bar illustrates the monthly integral data availability.

These results reinforce the notion of the critical role of the $\Delta^{14}C$ background in RIBA. While the magnitude of the impact is already considerable for flask samples (Sect. 5.2.3.1), it is even more pronounced in the case of integral sampling. This underscores the urgent need for robust and continuous $\Delta^{14}C$ background measurements at key sites such as MHD to ensure the reliability of fossil fuel CO_2 flux estimates — especially when using data types that rely on long sampling periods.

7.2.3.2 Influence of individual stations on the integral-based posterior

The influence of individual stations on the integral-based posterior emissions was assessed by removing one station's data from the inversion (or adding it for KIT). Most stations do not

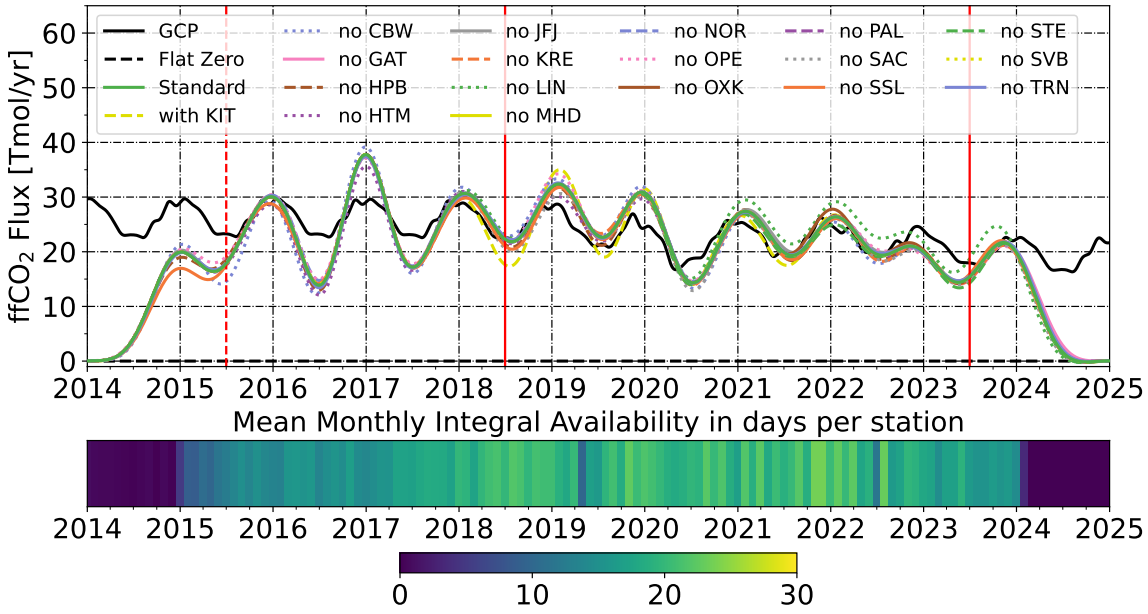

Figure 7.6. Posterior ΔffCO$_2$ fluxes for the Germany+ domain based on integral-based ΔffCO$_2$ estimates using a Flat Zero prior with 40% prior uncertainty. The standard posterior (green solid line) is compared to leave-one-out setups, where individual ICOS stations were excluded (other lines), as well as a configuration including the KIT site (yellow dashed line). The GCP inventory (black solid line) and the Flat Zero prior (black dashed line) are shown for reference. The color bar indicates the mean monthly availability of integral-based ΔffCO$_2$ data per station.

exhibit an excessive influence on the posterior results as depicted in Figure 7.6), suggesting robust data coverage in the Germany+ domain. In the following, I analyze how individual stations affect the previously observed discrepancies between posteriors and the GCP inventory (Sect. 7.2.2.3) and focus on KIT and LIN, which show significant influence compared to the rest of the network and did already show discrepancies for the other ΔffCO$_2$ proxies.

The emission maximum in winter 2016/17 is notably influenced by the CBW site (which suggests 5% lower emissions) and the HTM station (leading to 5% higher flux estimates for the domain). However, all stations generally agree on elevated emissions during this period. If the posterior is indeed biased, the underlying reasons for that have likely affected the entire measurement network, e.g., an incorrect background for this period or inaccurate atmospheric transport. The same applies to summer flux estimates in the secondary investigation period.

The LIN site significantly influences the posterior results in the final year, causing an average 10% decrease in ffCO$_2$ flux estimates. However, this influence supersedes the perceived discrepancy with the GCP inventory. Although excluding LIN data increases posterior fluxes, bringing them closer to GCP after mid-2022, the posterior shape remains dis-

torted, and the mismatch with GCP increases before 2022. Therefore, LIN data are not solely responsible for the discrepancy observed in the last year.

LIN's significant influence on posteriors was also observed in flask-based (Sect. 5.2.3.2) and CO-based (Sect. 6.3.3.2) inversions, where its exclusion led to higher posterior emissions. This is not surprising given its location east of Berlin, a major emitter, making it geographically important for constraining the Germany domain, as main air movement is typically from the west. Although LIN's mean MDM is high (third highest in the network, Tab. 7.1), its MDM variation for integrals ($\sigma_{sim-obs}$) is comparable to other ICOS stations.

Forward model runs match flask-based $\Delta ffCO_2$ observations at LIN (Sect. 5.2.1), but not CO-based (Sect. 6.3.1) or integral-based $\Delta ffCO_2$ (Sect. 7.2.1) data. However, flask data show high MDM variability ($\sigma_{sim-obs}$), suggesting potential misrepresentation of atmospheric conditions by the transport model. LIN is also located northwest of the "Jänschwalde" (approx. 80 km southeast) and "Schwarze Pumpe" (approx. 100 km southeast) power plants, Germany's largest brown coal power plants, which may contribute to high $\Delta ffCO_2$ variability. Both integrated fossil proxies show lower $ffCO_2$ concentrations than the forward model. Maier et al., 2021 demonstrated that incorrect emission levels of point sources can bias forward model $\Delta ffCO_2$ estimates, especially at night (relevant for integral samples), if the source is closer than 50 km to the station. Given their exceptional emission magnitudes, the aforementioned power plants might bias forward model predictions even at slightly longer distances. For CO-based $\Delta ffCO_2$ estimates, these plants can significantly contribute to the observed variability of the $\Delta CO / \Delta ffCO_2$ ratios. The presence of such high emitters within the LIN footprint could explain its significant influence on the Germany+ domain and raises questions about the ability of current transport models to reliably model point sources.

Even if its geographic location is excellent for constraining the Germany+ domain, in light of the high MDM and its variability (Tab. 7.1), potentially pointing to lacking transport model representativeness, the LIN station was removed from the final inversions with unmodified priors. Such a high dependence of the posterior on an individual station ($\approx$10%) makes the posterior highly susceptible to potential biases in that station's data.

Another prominent outlier from the station set is KIT. If the KIT site is added to the station set, the posterior experiences an increased seasonal amplitude: higher emissions in winter and lower in summer in the main investigation period. Although the mean MDM between simulations and measurements is low (0.1 ppm, Tab. 7.1), this site shows the highest variation of MDM throughout the ICOS network (2.7 ppm) due to its proximity to point sources and possible $\Delta^{14}C$ contamination. As explained previously, posterior fluxes are highly susceptible to outliers in the MDM. Given that KIT is not essential to constrain the inversion and introduces a potential bias by increasing the seasonal amplitude, its removal

from the core station set is hereby justified.

Similarly, the SAC station exhibits the highest absolute MDM (3.3 ppm) and the second-highest MDM standard deviation (2.6 ppm) after KIT (Tab. 7.1). This is likely due to complex emission patterns near Paris and potential on-site ^{14}C contamination from Saclay Nuclear Research Center, the campus of which harbors the measurement site. Such contamination would mask the $\Delta ffCO_2$ signal, potentially explaining high MDM. Furthermore, SAC has only little influence on the fluxes in the Germany+ domain and is not required to constrain it (Fig. 7.6) as it is located outside of the AoC (Fig. 2.4, Fig. 4.1). These issues advocate for the exclusion of SAC data from future inversions, including the inversions with unmodified priors, to avoid potential ^{14}C contamination induced bias.

7.2.4 Integral-based Bayesian inversions with unmodified GCP and EDGAR priors

This section presents results from two inversion runs based on integral-based $\Delta ffCO_2$ data, using the unaltered daily GCP and hourly EDGAR emission inventories as priors (Sect. 3.2.3). A prior uncertainty of 4% (relative to the GCP fluxes) was applied, in line with literature values for similarly aggregated domains (Janssens-Maenhout et al., 2019; Jones et al., 2022). Although the systematic anomalous behavior of the summer flux estimates was not observed in the integral-based inversion so far (Sec. 7.2.2.1), to be consistent with other investigations of this type (flask-based in Sect. 5.2.4 and CO-based in Sect. 6.3.4), I focused on winter months (November-February) from 2018-2023, representing the period with highest data constraint (Sect. 4.3). As discussed in Section 7.2.3.2, the data from SAC and LIN were excluded from these inversions.

Figure 7.7 depicts the prior $ffCO_2$ fluxes and inversion results in the Germany+ domain. Consistent with flask-based (Sect. 5.2.4) and CO-based (Sect. 6.3.4) inversions, both posteriors indicate lower mean winter fluxes than the corresponding inventories. Both absolute posterior values are closer to GCP estimates as shown in Table 7.4 and Table 7.5, suggesting a potential EDGAR winter flux overestimation. The GCP prior-posterior difference in mean winter emissions ranges from 0.5 to 2.9 Tmol/yr, or 2–12%, while for EDGAR, this difference is significantly higher: 3.1–5.8 Tmol/yr or 12–25%. Posterior mean winter flux estimates exhibit smaller differences (7–13%) than corresponding priors (16–23%), indicating strong data constraints and data-driven results for the integral-based inversions, particularly for the EDGAR posterior, where the adjustments are significantly higher than 4% assumed prior uncertainty. Uncertainty reduction in the posteriors compared to the priors ranges between 0.2 and 0.3 Tmol/yr corresponding to 20–30% relative uncertainty reduction. The mean winter posterior emissions are within 2–3σ confidence range of each other

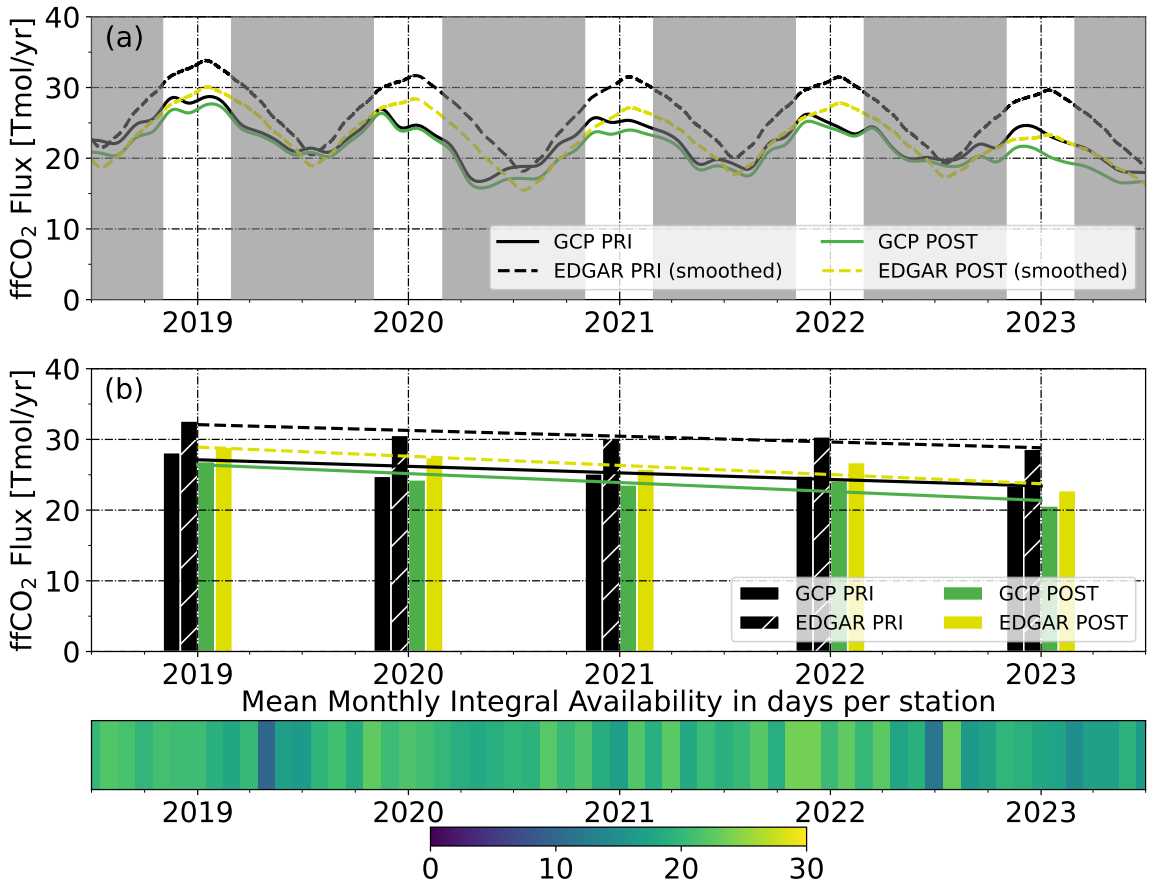

Figure 7.7. Winter fossil fuel CO_2 emissions in the Germany+ domain from Bayesian inversions using integral-based $\Delta ffCO_2$ estimates and two different prior inventories: GCP (solid lines/bars) and EDGAR (dashed lines/hatched bars). Panel (a) shows smoothed daily (GCP) and hourly (EDGAR) prior and posterior fluxes; non-winter periods are shaded in gray. Winter periods are defined as November through February. Panel (b) displays the corresponding winter mean fluxes per winter season. The trends estimated from the mean winter emissions are shown as solid and dashed lines in the respective colors. The color bar indicates the mean monthly availability of integral-based estimates per station.

for corresponding years.

As noted previously (e.g., Sect. 7.2.2.3), 2022/23 winter emission estimates for both posteriors show larger deviations from prior fluxes. Mean relative prior-posterior differences for winters 2018/19–2021/22 are 2–6% for GCP (Tab. 7.5) and 12–17% for EDGAR, whereas winter 2022/23 shows 12% and 25% deviations, respectively. For the GCP case, the "missing" ∼10% or 2.4 Tmol/yr emissions correspond to a ∼1‰ bias in the ^{14}C background (Sect. 7.2.3.1). This hypothesis is supported by the fact that the deviations in the flask-based inversions could be also explained by ^{14}C background bias of similar magnitude (Sect. 5.2.4).

Table 7.4. Winter mean $ffCO_2$ emissions (in Tmol/yr) and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective posterior inversions using integral-based $\Delta ffCO_2$ estimates. Winter periods refer to November through February of the indicated winter season.

Flux type	Absolute fluxes [Tmol/yr]				
	2018/19	2019/20	2020/21	2021/22	2022/23
GCP Prior	28.1(1.0)	24.8(0.9)	25.1(0.9)	24.8(0.9)	23.5(0.8)
GCP Posterior	26.9(0.7)	24.3(0.7)	23.6(0.6)	24.2(0.6)	20.6(0.6)
EDGAR Prior	32.6(1.1)	30.6(1.0)	30.1(1.0)	30.4(1.0)	28.6(1.0)
EDGAR Posterior	28.9(0.8)	27.5(0.7)	25.8(0.7)	26.7(0.7)	22.8(0.7)

Table 7.5. Deviations of winter mean $ffCO_2$ emissions from the GCP emission inventory for the corresponding winter and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective posterior inversions using integral-based $\Delta ffCO_2$ estimates. Winter periods refer to November through February of the indicated winter season.

Flux type	Deviation from GCP [%]				
	2018/19	2019/20	2020/21	2021/22	2022/23
GCP Prior	0(4)	0(4)	0(4)	0(4)	0(4)
GCP Posterior	–4(2)	–2(2)	–6(2)	–2(2)	–12(2)
EDGAR Prior	16(4)	23(4)	20(4)	23(4)	22(4)
EDGAR Posterior	3(2)	11(3)	3(2)	8(2)	–3(3)

Emission reduction trends were calculated using linear fits of mean winter emissions (Fig. 5.7, panel (b)). Over five winters, estimated annual emission reductions in the AoC are 0.9(0.3) Tmol/yr for GCP and 0.8(0.2) Tmol/yr for EDGAR (35% and 25% uncertainty, respectively). Both posteriors suggest a higher reduction rate of 1.3(0.4) Tmol/yr (30% trend uncertainty). This higher estimate may be attributed to notably lower 2022/23 winter emissions. Excluding this winter from the linear fits changes emission trend estimates to 1.0(0.6) Tmol/yr and 0.7(0.4) Tmol/yr for GCP and EDGAR priors, and 0.9(0.5) Tmol/yr and 0.8(0.4) Tmol/yr for respective posteriors (not shown). These revised estimates do not differ significantly within respective error boundaries. High uncertainties in prior emission reduction rates, similar to the CO case (Sect. 6.3.4), potentially relate to COVID-induced flux variations in 2020–2021. These findings support using integral-based inversions for

validating mean winter emission trends, once the issues with ^{14}C background are fixed.

Chapter 8

Discussion

This chapter discusses a key limitation of the Regional Isotope Budget Approach (RIBA): its dependence on a reliable ^{14}C background estimate. The effects of a potential 2023–2024 background bias on the posterior results for different $\Delta ffCO_2$ proxies are summarized in Section 8.1. Section 8.2 compares the inversions results using unmodified GCP and EDGAR priors for flask, integral, and CO-based $\Delta ffCO_2$ proxies for the overlapping time frame, assuming appropriate flux adjustments to counteract the effects of the ^{14}C background bias.

8.1 High ^{14}C background bias dependence of the RIBA

As already mentioned before (see Sect. 5.2.3.1, 6.3.3.4 and 7.2.3.1), a biased ^{14}C background can substantially affect $\Delta ffCO_2$ estimates derived via the RIBA and thus influence the inversion results. Unfortunately, no ^{14}C samples have been collected at MHD since summer 2022, rendering the central background station for the European continent unavailable for the years with an exceptionally high number of flask samples (Sect. 4.3). Moreover, this time span coincides with the interval in which I expect the most pronounced changes in natural ^{14}C production. The 2020 solar minimum (Clette & Lefèvre, 2015) led to increased stratospheric $\Delta^{14}C$ production, with expected tropospheric effects around 2022 due to typical stratosphere-troposphere exchange times (Holton et al., 1995). The change in the natural ^{14}C production is evident by the slight trend adjustments in 2023-2024 seen at the stations JFJ, PAL and SVB as well as in the last MHD integral samples that are systematically above the used ^{14}C background estimate (Fig. 3.2). These circumstances motivated the construction of a ^{14}C background record based on the trends observed at other available Northern Hemisphere remote stations with no major fossil CO_2 sources nearby located at similar latitudes.

Due to fast meridional mixing that acts on timescales of weeks to months (Warneck, 1999) this mean northern-hemispheric trend should consistently reflect the effect of en-

hanced stratospheric ^{14}C production. However, it cannot be said with certainty whether the mean ^{14}C latitude-dependent differences derived from long-term station comparisons remain valid for this particular period. In addition, the mean seasonality of the used ^{14}C background was derived from the long-term MHD seasonal cycle, making the background estimate rigid and unable to reflect potential short-term variations. In summary, it can be stated that there is in fact a potential for bias in the constructed ^{14}C background, particularly in the years 2023 and 2024.

The analyses of the ^{14}C background bias sensitivity (Sect. 5.2.3.1, 6.3.3.4, 7.2.3.1) showed a substantial sensitivity of the posterior fossil flux estimates for all three $\Delta ffCO_2$ proxies. In these sensitivity analyses, I investigated the impact of a ± 2‰ bias in $\Delta^{14}C$ background on the posteriors of each proxy. For the flask-based inversions, the ± 2‰ bias leads to an $\Delta ffCO_2$ emission change of $\sim \pm 2$ Tmol/yr in winter and $\sim \pm 4.5$ Tmol/yr in summer (approx. 8% and 20% of the corresponding GCP emissions). For the integral-based inversions, the ± 2‰ bias in $\Delta^{14}C$ background leads to even larger emission changes of $\sim \pm 6$ Tmol/yr in winter and $\sim \pm 8$ Tmol/yr in summer (approx. 22% and 40% relative to the GCP fluxes). The differing impact of a constant $\Delta^{14}C$ bias on the two proxies, as well as on the two seasons, can be explained by the absolute magnitude of the $\Delta ffCO_2$ enhancement signal. The smaller the $\Delta ffCO_2$ concentrations — for instance, in summer or in integrated samples compared to flask samples — the greater the relative influence of the background bias.

CO-based inversions are indirectly influenced by the ^{14}C background through its role in calculating the $\Delta CO/\Delta ffCO_2$ ratio, which affects the entire CO-based dataset. A ± 2‰ bias in the ^{14}C background systematically shifts the derived CO-based $\Delta ffCO_2$ concentrations, with a more pronounced effect in winter. Although the relative impact per sample remains consistent, the absolute change in winter $\Delta ffCO_2$ concentration is greater due to the larger ΔCO signals. This leads to a higher influence on the posterior estimates in winter of about ± 5 Tmol/yr and only ± 3 Tmol/yr in summer. Additionally, CO-based data are also influenced by CO background biases, directly affecting $\Delta ffCO_2$ estimates and leading to posterior $\Delta ffCO_2$ emission changes of 1-2 Tmol/yr in winter and up to 15 Tmol/yr in summer for ± 15 ppb CO background biases ($\sim$5-8% and $\sim$60% of the mean annual fluxes in the Germany+ domain). As discussed in Section 6.3.3.4, most ^{14}C flasks for ratio calculation were collected in 2024, and are therefore potentially susceptible to a bias in the background.

All three inversion results show indications of a ^{14}C background bias in 2023 and 2024 (see Sect. 5.2.2.1, 6.3.2.1 and 7.2.2.1), coinciding with the data gap at MHD. This suggests that efforts to remedy this by using global clean air ^{14}C sites for the northern hemispheric trend estimation were not entirely successful. Especially the observed deviations during summer in flask-based and integral-based inversions suggest a potential bias of 1‰ in 2023,

increasing to approximately 2‰ in 2024 (e.g., Sect. 5.2.4 and 7.2.4). These findings underscore MHD's critical role as a clean European marine background site and emphasize the need for its continued ^{14}C data series. The ^{14}C background determination for the years 2023 and 2024 is expected to become more robust once the ^{14}C dataset at MHD is resumed with the 2025 measurements and additional data from other Northern Hemisphere background stations are available.

8.2 Fossil fuel CO_2 proxies inter-comparison

Sensitivity studies on all three $\Delta ffCO_2$ proxies revealed that the phase of the seasonal cycle could be reliably constrained within the expected range of variability caused by proxy-specific observational uncertainties. However, the amplitude of the seasonal cycle has been significantly overestimated in recent years for flask-based and CO-based inversions (Sect. 5.2.2.1 and 6.3.2.1). As discussed in the previous section (Sect. 8.1), this overestimation is most likely due to a growing bias in the ^{14}C background estimate.

Consequently, the results of the sensitivity tests for annual mean emissions derived from flask-based and CO-based $\Delta ffCO_2$ data cannot be interpreted meaningfully. In contrast, the longer data record of the integral samples allows for the assessment of the seasonality as well as the annual mean fossil emissions during the period before the suspected background bias. Between 2016 and 2022, the integral-based data enabled the estimation of annual mean fossil CO_2 emissions in the Germany+ domain to within 10% of the corresponding GCP inventory values — even with minimal prior information (Sect. 7.2.2). This result strongly demonstrates the extent to which $^{14}CO_2$ based atmospheric observations alone can constrain fossil emissions at the regional scale, if the ^{14}C background data is reliable.

When comparing the results of the Bayesian inversions of the different fossil fuel CO_2 proxies using unmodified priors, I focused on the winter periods to reduce the influence of a potentially biased ^{14}C background, although the previous section has shown that even during winter a background bias cannot be entirely ignored.

To address this, the proxy-specific results presented here are also interpreted under a biased scenario, where the ^{14}C background is assumed to be negatively biased by 1‰±0.5‰ in 2023 and by 2‰±1‰ in 2024 affecting the flask-based and integral-based posterior emissions. The upper half of Table 8.1 summarizes the mean GCP prior and posterior winter $\Delta ffCO_2$ fluxes for the Germany+ domain across the different proxies. The upper part shows the unmodified posterior results, while the lower part of the upper half (marked with *) presents the background bias adjusted results for the last two winters, assuming the 1‰ and 2‰ biases, respectively. Note that for CO-based estimates, the bias correction in 2024 affects all previous years, since the sheer number of samples from 2024 dominates the CO

Table 8.1. Winter mean $\Delta ffCO_2$ emissions (in Tmol/yr) and their 1σ uncertainty (in parentheses) for the Germany+ domain, based on different prior inventories (GCP and EDGAR) and the respective posterior emissions based on three $\Delta ffCO_2$ proxies: ^{14}C-based flasks (Fl.) and integrals (Int.) and ^{14}C-calibrated CO-based $\Delta ffCO_2$. (*) Emissions are adjusted for the bias in $\Delta^{14}C$ background (1‰±0.5‰ in 2023 and 2‰±1‰ in 2024 for Fl. and Int., 1‰±0.5‰ in all years for CO). "NA" denotes winters outside of the main investigation period of the corresponding proxy. Winter periods refer to November through February of the indicated winter season.

Flux type	Proxy	Absolute fluxes, [Tmol/yr]					
		2018/19	2019/20	2020/21	2021/22	2022/23	2023/24
GCP Prior	-	28.1(1.0)	24.8(0.9)	25.1(0.9)	24.8(0.9)	23.5(0.8)	22.2(0.8)
GCP Posterior	Fl.	NA	NA	NA	24.6(0.4)	21.9(0.5)	20.3(0.5)
	CO	NA	22.2(0.7)	24.7(0.6)	22.2(0.6)	22.4(0.5)	20.9(0.6)
	Int.	26.9(0.7)	24.3(0.7)	23.6(0.6)	24.2(0.6)	20.6(0.6)	NA
	Fl.*	NA	NA	NA	24.6(0.4)	22.9(1.0)	22.3(1.5)
	CO*	NA	24.7(2.0)	27.2(1.9)	24.7(1.9)	24.9(1.9)	23.4(1.8)
	Int.*	26.9(0.7)	24.3(0.6)	23.6(0.6)	24.2(0.6)	23.6(2.1)	NA
EDGAR Prior	-	32.6(1.1)	30.6(1.0)	30.1(1.0)	30.4(1.0)	28.6(1.0)	26.9(0.9)
EDGAR Posterior	Fl.	NA	NA	NA	27.3(0.5)	23.7(0.5)	23.0(0.6)
	CO	NA	24.9(0.7)	27.1(0.7)	24.6(0.6)	24.9(0.6)	23.3(0.6)
	Int.	28.9(0.8)	27.5(0.7)	25.8(0.7)	26.7(0.7)	22.8(0.7)	NA
	Fl.*	NA	NA	NA	27.3(0.5)	24.7(1.0)	25.0(1.6)
	CO*	NA	27.4 (2.0)	29.6(2.0)	27.1(1.9)	27.4(1.9)	25.8(1.9)
	Int.*	28.9(0.8)	27.5(0.7)	25.8(0.7)	26.7(0.7)	25.8(2.2)	NA

to $\Delta ffCO_2$ correlation. For CO-based posteriors, I assume a mean ^{14}C background bias of 1‰±0.5‰ and apply positive 2.5±1.3 Tmol/yr correction to all years. The lower half of Table 8.1 shows the same results, but for the EDGAR prior and posteriors.

Table 8.2 displays the original and bias-adjusted relative deviations with respect to the winter GCP values. Comparing the winters of 2021/22 and 2022/23, during which posterior estimates could be derived from all three proxies, I find that in winter 2021/22 – the only period for which proxies are available and no background bias correction was required (ex-

Table 8.2. Deviations of winter mean $\Delta ffCO_2$ emissions from the GCP for the corresponding winter and their 1σ uncertainty (in parentheses) for the Germany+ domain. The GCP posteriors are based on three $\Delta ffCO_2$ proxies: ^{14}C-based flasks (Fl.) and integrals (Int.) and ^{14}C-calibrated CO-based $\Delta ffCO_2$. (*) Emissions are adjusted for the bias in $\Delta^{14}C$ background (1‰±0.5‰ in 2023 and 2‰±1‰ in 2024 for Fl. and Int., 1‰±0.5‰ in all years for CO). "NA" denotes winters outside of the main investigation period of the corresponding proxy. Winter periods refer to November through February of the indicated winter season.

Flux type	Proxy	Deviation from GCP, [%]					
		2018/19	2019/20	2020/21	2021/22	2022/23	2023/24
GCP Prior	-	0(4)	0(4)	0(4)	0(4)	0(4)	0(4)
EDGAR Prior	-	16(4)	23(4)	20(4)	23(4)	22(4)	21(4)
GCP Posterior	Fl.	NA	NA	NA	-1(2)	-7(2)	-9(2)
	CO	NA	-10(3)	-2(2)	-10(2)	-5(2)	-6(3)
	Int.	-4(2)	-2(2)	-6(2)	-2(2)	-12(2)	NA
	Fl.*	NA	NA	NA	-1(2)	-3(4)	-4(7)
	CO*	NA	-0(8)	8(8)	-0(8)	6(8)	5(8)
	Int.*	-4(2)	-2(2)	-6(2)	-2(2)	0(9)	NA

cept for CO-based) – the posterior $\Delta ffCO_2$ emissions deviate from the GCP inventory by between 0% and –2%. All three proxies agree within their respective uncertainty ranges.

In contrast, winter 2022/23 is subject to a direct bias correction due to the assumed downward shift in the ^{14}C background. Here, the bias-adjusted deviations from the GCP inventory reach 6%. A quantitative interpretation of these differences is not attempted as they depend directly on the assumed bias magnitude, which is only qualitatively estimated. However, under these assumptions, all three proxies produce posterior winter emission estimates that are consistent with the GCP bottom-up inventory at the ±6% level and are also mutually consistent within that range.

Considering the winter posterior fluxes for the same periods derived using EDGAR emissions as prior in Table 8.3, I find that for the same period the background bias-adjusted posteriors of all proxies reduce the EDGAR emissions between -4 and -14%.

Looking at the full time series of the bias-corrected proxy-specific posteriors, flask-based inversions consistently show excellent agreement with the GCP product, with devi-

ations $\leq$ –4%. In contrast, EDGAR-based posteriors are systematically reduced by –7% to –14% relative to the prior. CO-based posterior fluxes deviate by 0% to +8% from the GCP prior and by –2% to –11% from EDGAR, although all CO-based priors were assigned a broader uncertainty of $\pm$8%. The integral-based posteriors also show strong agreement with the GCP estimates, with deviations generally ranging from 0% to –4%, except in winter 2020/21. During this winter, a COVID-19 lockdown from mid-December to April likely contributed to a reduction of –6% compared to the GCP prior. However, compared to EDGAR, the integral-based posteriors consistently show reductions between –10% and –14% throughout the entire record.

The substantially stronger posterior adjustments relative to the EDGAR prior indicate that, under the chosen inversion settings and with prior uncertainties of 4%, the observational data exert sufficient influence to induce notable changes in the prior fluxes. The posterior winter fluxes based on the GCP and EDGAR priors differ only by 10%, demonstrating that the integration of atmospheric observations effectively reduced the average 21% prior discrepancy between the two inventories by half. This demonstrates that within the inversion framework used in this study, the observational data carry sufficient weight to induce substantial adjustments to the prior fluxes. The small reduction in GCP winter emissions for the Germany+ region in the GCP posteriors (0–6%) strongly suggests that GCP provides an accurate representation of the actual winter fossil emissions for this area. Integral-based GCP posteriors before the suspected ^{14}C background bias also support this, improving the winter emission estimate within GCP's reported 5% uncertainty (Jones et al., 2021), with one potential COVID-related discrepancy (–6% in winter 2020/21). Bias-adjusted flask-based winter estimates further confirm this finding. The assimilation of $^{14}CO_2$ -based $\Delta ffCO_2$ produces better quantified fossil winter emissions of the Germany+ domain, with a mean GCP posterior uncertainty over flasks and integrals of $\sim$0.6 Tmol/yr (Tab. 8.1), representing an uncertainly reduction of $\sim$30% from the prior estimate.

This finding is further supported by the comparison of prior and posterior winter trends, calculated using the error-weighted linear regressions over the winter emissions (not shown). For the winters covered by the integral samples (2018/19 to 2022/23), the GCP prior shows a trend of –0.9$\pm$3 Tmol/yr, while EDGAR reports –0.8$\pm$0.2 Tmol/yr. The integral-based bias-adjusted posterior fluxes are able to confirm these trends, yielding –0.8$\pm$0.4 Tmol/yr when using GCP priors and –0.7$\pm$0.3 Tmol/yr for EDGAR priors. In the most recent winters (2021/22 to 2023/24), the prior trends in the inventories show a significantly stronger decline with reduced inter-annual variability: GCP reports –1.32$\pm$0.03 Tmol/yr and EDGAR –1.74$\pm$0.01 Tmol/yr. The flask-based bias-corrected posteriors confirm this sharper reduction in fossil emissions, with trends of –1.3$\pm$0.3 Tmol/yr when using GCP priors and –1.1$\pm$0.9 Tmol/yr with EDGAR priors. Although the posterior trends carry substantially

Table 8.3. Deviations of winter mean $\Delta ffCO_2$ emissions from the EDGAR for the corresponding winter and their 1σ uncertainty (in parentheses) for the Germany+ domain. The EDGAR posteriors are based on three $\Delta ffCO_2$ proxies: ^{14}C-based flasks (Fl.) and integrals (Int.) and ^{14}C-calibrated CO-based $\Delta ffCO_2$. (*) Emissions adjusted for the bias in $\Delta^{14}C$ background (1‰±0.5‰ in 2023 and 2‰±1‰ in 2024 for Fl. and Int., 1‰±0.5‰ in all years for CO). "NA" denotes winters outside of the main investigation period of the corresponding proxy. Winter periods refer to November through February of the indicated winter season.

Flux type	Proxy	Deviation from EDGAR, [%]					
		2018/19	2019/20	2020/21	2021/22	2022/23	2023/24
EDGAR Prior	-	0(3)	0(3)	0(3)	0(3)	0(3)	0(3)
GCP Prior	-	-14(3)	-19(3)	-17(3)	-18(3)	-18(3)	-17(3)
EDGAR Posterior	Fl.	NA	NA	NA	-10(2)	-17(2)	-14(2)
	CO	NA	-19(2)	-10(2)	-19(2)	-13(2)	-13(2)
	Int.	-11(2)	-10(2)	-14(2)	-12(2)	-20(2)	NA
	Fl.*	NA	NA	NA	-10(2)	-14(3)	-7(6)
	CO*	NA	-10(7)	-2(7)	-11(6)	-4(7)	-4(7)
	Int.*	-11(2)	-10(2)	-14(2)	-12(2)	-10(7)	NA

larger uncertainties, they align more closely with the trend reported by GCP. The bias-corrected CO-based posteriors suggest the lowest decreasing trend of the three proxies for the 2019/20 to 2023/24 period (–0.5±0.4 Tmol/yr for both priors). Within the uncertainty ranges, they do not deviate significantly from the prior trends: -0.7±0.2 Tmol/yr for GCP and -0.9±0.2 Tmol/yr for EDGAR.

Chapter 9

Conclusion

In this concluding chapter, I summarize the main findings of this thesis regarding the suitability of different $\Delta ffCO_2$ proxies and inversion configurations to estimate country-scale fossil fuel CO_2 emissions. The results highlight the critical role of $^{14}CO_2$ background accuracy, the structural limitations of the Regional Isotope Budget Approach (RIBA), and the relative performance of flask-based, CO-based, and integral-based observations. Based on these insights, I provide recommendations for improving future ICOS sampling strategies and inversion system design.

9.1 Background diversification

This study underscores the substantial influence that potential biases in the European ^{14}C background can exert on posterior $\Delta ffCO_2$ flux estimates. Therefore, accurate knowledge of the ^{14}C activity in air masses prior to their arrival on the European continent is essential for the integrity of the RIBA. The current heavy reliance on a single station, Mace Head (MHD), is scientifically inadequate, particularly in light of the disproportionate investment in monitoring continental as opposed to background ^{14}C concentrations. To improve the robustness and credibility of future inversions, the ICOS atmospheric network is strongly advised to diversify its ^{14}C background observation sites. This recommendation remains pressing even if ^{14}C measurements at MHD resume in 2025.

Developing secondary background stations for the European domain should be considered a high priority. Given that most European air masses originate from the western boundary, existing Irish ICOS coastal Class 2 stations, such as Valentia Island (51.93°N, –10.24°E) or Malin Head (55.37°N, –7.34°E), may be suitable secondary sites for clean marine air. Although there is currently no ICOS coastal station on the western French coast, the Roc'h Trédudon station operated by Laboratory for Climate and Environmental Sciences (LSCE, 48.41°N, –3.89°E) on the Brittany peninsula represents a valuable non-ICOS

alternative.

9.2 Alternatives to the RIBA

The RIBA, first introduced by Levin et al., 2003 and further developed by Levin and Rödenbeck, 2008, has proven effective in urban environments such as Heidelberg, DE, where high $\Delta ffCO_2$ signals minimize the relative impact of background-related uncertainties. However, these early studies have already revealed increased uncertainty at more remote sites such as Schauinsland, DE, where the $\Delta ffCO_2$ signals are weaker. As most ICOS stations are located similarly in rural or remote areas, they are typically characterized by low $\Delta ffCO_2$ concentrations. This renders the entire ICOS network highly susceptible to biases in the ^{14}C background estimate, as discussed in Section 8.1, and exposes a fundamental limitation of the RIBA when applied to observations with low fossil signals.

In essence, the RIBA acts as a simplified transport model, assuming a common background for all observations. In this study, the $\Delta ffCO_2$ values derived via the RIBA were further processed using the STILT model, which explicitly accounts for time- and location-specific meteorology. When actual atmospheric transport conditions align with the implicit assumption of RIBA of predominantly westerly wind flow, both models yield consistent results. As shown by Maier et al., 2024a, this alignment holds for approximately two-thirds of the time. For the remaining one-third, deviations from assumed meteorology introduce systematic biases into the RIBA-based $\Delta ffCO_2$ estimates. Nevertheless, the spatial gradients across the network, i.e., the relative differences between stations, remain informative and usable.

This key limitation can be overcome by implementing coupled atmospheric ^{14}C–CO_2 models (e.g. Basu et al., 2020; Gómez-Ortiz et al., 2025), which directly simulate both tracers and no longer rely on pre-calculated $\Delta ffCO_2$ values from the RIBA. Such models can incorporate time-resolved, air-mass-specific background concentrations, and thus eliminate the need for a single, static background time series. Within the CORSO project, several European research groups are working towards this goal by integrating the entire atmospheric ^{14}C cycle into existing CO_2 transport models and inverse modeling frameworks.

Although these next-generation models offer a path beyond the limitations of the RIBA, they still require robust and accurate data on the temporal evolution of the European ^{14}C background. As with global models, European-scale applications must account for additional ^{14}C-specific influences — such as emissions from nuclear installations outside Europe, contributions from the stratosphere and air–sea exchange of radiocarbon. Unlike in the RIBA, these effects are no longer implicitly accounted for but must be explicitly modeled. Reliable and regionally resolved ^{14}C observations remain therefore an essential foundation

for any robust fossil fuel CO_2 estimation framework.

9.3 Suitability of $\Delta ffCO_2$ proxies for inverse modeling of country-scale fossil fuel CO_2 emissions

This thesis evaluated the suitability of three $\Delta ffCO_2$ proxies — ^{14}C-flask-based, CO-based, and ^{14}C-integral-based estimates — for use in inverse modeling of country-scale fossil fuel CO_2 emissions. The aim was to inform and improve the ICOS $^{14}CO_2$ sampling strategy by identifying which observation types are most compatible with current inversion frameworks.

To evaluate the proxies, three key criteria were defined:

1. Model representativeness

   Representativeness errors occur when the transport model fails to reproduce the actual atmospheric state at the time and location of sampling. These errors are represented by the standard deviation of the model-data-mismatch ($\sigma_{sim-obs}$). High $\sigma_{sim-obs}$ inflates the inversion cost function and results in biased flux estimates, particularly when the $\sigma_{sim-obs}$ distribution is not symmetric.

2. Signal-to-noise ratio (SNR)

   Higher $\Delta ffCO_2$ concentrations enhance the robustness of the inversion. The sampling method and strategy influence the observed $\Delta ffCO_2$ concentrations and hence the SNR.

3. Sensitivity to background biases

   Biases in the ^{14}C background strongly affect the RIBA. Although this dependency can be mitigated by employing coupled $^{14}CO_2$–CO_2 models, lower $\Delta ffCO_2$ signals inherently increase sensitivity to background errors.

Flask-based $\Delta ffCO_2$ estimates. Flask samples yield the highest $\Delta ffCO_2$ signals as a result of the possibility of targeted event-based sampling. This results in favorable SNR, especially under the standard ICOS flask sampling protocol, which targets afternoon conditions with potentially high $\Delta ffCO_2$. However, this advantage comes at the expense of model representativeness: the $\sigma_{sim-obs}$ values of the flask samples are typically 40–100% higher than those of the integral samples at the same sites, except at locations with well-mixed atmospheric conditions such as HTM and PAL. The low flask sampling frequency increases the susceptibility of individual flasks to transport model mismatches, particularly in heterogeneous environments.

While flasks provide flexibility and allow targeted sampling, they require a high sampling frequency to ensure robustness: ideally, several flasks per synoptic event (2–5 days). The CORSO campaign (Levin et al., 2024) demonstrated that this is feasible, while biweekly flask sampling remains inadequate given current modeling limitations. Flask-based inversions were shown to be the least sensitive to ^{14}C background biases due to their higher absolute $\Delta ffCO_2$ concentrations per sample. The bias-adjusted winter posterior fluxes showed good agreement with the GCP emission estimates, while consistently deviating from the EDGAR inventory by –7% to –14%.

CO-based $\Delta ffCO_2$ estimates. CO-based $\Delta ffCO_2$ estimates, derived from continuous CO measurements calibrated against ^{14}C samples, offer high temporal resolution and dense data coverage compared to the other proxies. However, this study showed that the inversion system does not benefit from such a fine temporal resolution, due to the limited ability of the transport model to simulate atmospheric conditions accurately at hourly scales. Therefore, weekly averaging of the CO-based $\Delta ffCO_2$ estimates was required to reduce the dependence on hour-specific transport accuracy and thus limit representativeness errors. In fact, $\sigma_{sim-obs}$ for the CO-based approach was comparable to integral data at most sites (within 20%), with the exception of LIN, where the $\Delta CO/\Delta ffCO_2$ ratios varied strongly within the footprint.

The CO-based proxy suffers from substantial conceptual limitations. Its reliability depends on both the CO and the ^{14}C background estimates, which introduces a dual sensitivity to background biases. In addition, the unmodeled, seasonally varying CO sinks, non-fossil CO sources, and poorly constrained emission ratios limit the robustness of CO-based $\Delta ffCO_2$ estimates. These problems are particularly relevant at rural ICOS sites, where signal strengths are low. Although CO-based methods have shown promise in urban-scale applications (e.g. Levin & Karstens, 2007; Lopez et al., 2013; Maier et al., 2024b), this study concludes that they are not suitable for continental-scale inversions without a fully implemented CO chemistry module in a coupled $^{14}CO_2$–CO_2 model.

Integral-based $\Delta ffCO_2$ estimates. Integral samples, which represent an average over two weeks of continuous sampling, do not require accurate hour-by-hour transport modeling. A fair representation of mean atmospheric transport during the integration period is sufficient for the model to replicate observed concentrations. $\sigma_{sim-obs}$ values are significantly lower than those of the flask samples and comparable to those of weekly averaged CO-based runs. Although $\Delta ffCO_2$ concentrations in integral samples are 20–40% lower than in flask samples, reduced variability and consistent coverage lead to stable SNR, particularly under standardized biweekly sampling. While integral-based inversions were most sensitive to ^{14}C background biases between all proxies, this drawback can be mitigated through the use

of coupled $^{14}CO_2$ –CO_2 models.

In the well-constrained and non-biased 2018–2022 period (presumably, due to the availability of MHD data), integral-based inversions successfully reproduced key characteristics of the GCP inventory, including annual mean emissions, seasonal cycles and multi-year trends, within the limits defined by synthetic studies. Even in Flat Prior sensitivity tests (Sect. 7.2.2.3), inversion results matched GCP values to within a few percent. The bias-corrected winter fluxes derived from unmodified prior inversions using integral samples showed strong agreement with GCP estimates, with deviations of only 0% to –4%, while systematically indicating lower emissions compared to EDGAR, with reductions between –10% and –14%.

Conclusion. Each of the evaluated proxies has distinct strengths and limitations. CO-based $\Delta ffCO_2$ estimates offer high temporal coverage, but are hampered by dual background dependencies, unresolved atmospheric chemistry, and poorly constrained tracer relationships. In light of these limitations, CO-based estimates currently represent the least reliable method to quantify fossil fuel CO_2 concentrations within the ICOS network.

Flask-based estimates provide strong signals and sampling flexibility. As demonstrated in this study, they represent a viable and robust option for estimating country-scale fossil fuel emissions — provided that the sampling frequency is sufficiently high (at least 2–3 samples per week).

Integral-based $\Delta ffCO_2$ estimates offer the most favorable balance of strengths and weaknesses. While their signal strength is generally lower, making them somewhat more susceptible to background biases, they exhibit superior model representativeness and yield reproducible inversion results across key evaluation metrics.

For future ICOS sampling strategies, particularly when aiming to constrain national- or continental-scale fossil fuel emissions, it is therefore my recommendation to:

(a) diversify the European ^{14}C background estimation to minimize systematic bias, and

(b) prioritize integral sampling as the default $^{14}CO_2$ observation method.

This recommendation is supported by the finding that flask and integral samples produce comparable inversion results, while integral sampling requires four to six times fewer analyses.

9.4 Enhancing the ICOS ^{14}C sampling spatial data coverage

Figure 4.5 illustrates the data constraint within the Germany+ domain. The increased constraint for integral-based inversions in 2018 resulted from the addition of new stations to the ^{14}C measurement network, as detailed in Section 4.3. Notably, despite only evaluating approximately every second integral sample from late 2021 to end of 2023, no significant decrease in data constraint was observed. This suggests that expanding the ^{14}C network with new stations is more beneficial to the characterization of the European $\Delta ffCO_2$ fluxes than increasing the evaluation frequency at existing sites.

Currently, the reliable data constraint is limited to the Germany+ domain (Sect. 4.2). If additional ^{14}C measurement funding becomes available, extending ^{14}C integral sampling to ICOS Class 2 sites could expand the current sampling station set from 18 to 46. This expansion would presumably broaden the constrained area to encompass Austria, Switzerland, Denmark, Eastern France, Central Spain, most of Great Britain, and Italy. Such an extension would leverage the existing ICOS infrastructure and require relatively simple and inexpensive instrumentation, such as micro-alkaline CO_2 collectors. These meaningful additions to European spatial data coverage would advance ICOS toward its goal of providing robust data to constrain long-term European greenhouse gas emissions.

Bibliography

Anderson, E., Arnold, J., & Libby, W. (1951). Measurement of low level radiocarbon. *Review of Scientific Instruments*, *22*(4), 225–230.

Andres, R. J., Boden, T. A., Breon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., et al. (2012). A synthesis of carbon dioxide emissions from fossil-fuel combustion. *Biogeosciences*, *9*(5), 1845–1871.

Archer, D. (2011). *The global carbon cycle*. Princeton University Press.

Archer, D., Eby, M., Brovkin, V., Ridgwell, A., Cao, L., Mikolajewicz, U., Caldeira, K., Matsumoto, K., Munhoven, G., Montenegro, A., et al. (2009). Atmospheric lifetime of fossil fuel carbon dioxide. *Annual review of earth and planetary sciences*, *37*(1), 117–134.

Bakels, L., Tatsii, D., Tipka, A., Thompson, R., Dütsch, M., Blaschek, M., Seibert, P., Baier, K., Bucci, S., Cassiani, M., et al. (2024). Flexpart version 11: Improved accuracy, efficiency, and flexibility. *Geoscientific Model Development*, *17*(21), 7595–7627.

Basu, S., Lehman, S. J., Miller, J. B., Andrews, A. E., Sweeney, C., Gurney, K. R., Xu, X., Southon, J., & Tans, P. P. (2020). Estimating us fossil fuel co2 emissions from measurements of 14c in atmospheric co2. *Proceedings of the National Academy of Sciences*, *117*(24), 13300–13307.

Bayes, T. (1763). Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, (53), 370–418.

Be, M.-M., Chiste, V., Dulieu, C., Mougeot, X., Chechev, V., Kondev, F., Nichols, A., Huang, X., & Wang, B. (2013). Table of radionuclides (vol. 7-a= 14 to 245).

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A., et al. (2015). Top-down estimates of european ch 4 and n 2 o emissions based on four different inverse models. *Atmospheric Chemistry and Physics*, *15*(2), 715–736.

Berhanu, T. A., Szidat, S., Brunner, D., Satar, E., Schanda, R., Nyfeler, P., Battaglia, M., Steinbacher, M., Hammer, S., & Leuenberger, M. (2017). Estimation of the fossil fuel component in atmospheric co 2 based on radiocarbon measurements at the

beromünster tall tower, switzerland. *Atmospheric Chemistry and Physics*, *17*(17), 10753–10766.

Bonani, G., Beer, J., Hofmann, H., Synal, H.-A., Suter, M., Wölfli, W., Pfleiderer, C., Kromer, B., Junghans, C., & Münnich, K. O. (1987). Fractionation, precision and accuracy in 14c and 13c measurements. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *29*(1-2), 87–90.

Caldeira, K., Rau, G. H., & Duffy, P. B. (1998). Predicted net efflux of radiocarbon from the ocean and increase in atmospheric radiocarbon content. *Geophysical Research Letters*, *25*(20), 3811–3814.

Canadell, J. G., Monteiro, P. M., Costa, M. H., Da Cunha, L. C., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., et al. (2021). Global carbon and other biogeochemical cycles and feedbacks. *IPCC AR6 WGI, final government distribution*, chapter–5.

Chanca, I. (2022). *Theoretical and experimental approaches using 14c for estimating system diagnostic times in the central amazon rainforest* [Doctoral dissertation, Fluminense Federal University Niteroi].

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Costa, M., Cotrim da Cunha, L., Cox, P., Eliseev, A., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Monteiro, P., Patra, P., Piao, S., Rogelj, J., … Zickfeld, K. (2013). Carbon and other biogeochemical cycles. In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. Midgley (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 465–570). Cambridge University Press. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter06_FINAL.pdf

Clette, F., & Lefèvre, L. (2015, July). Silso sunspot number v2.0 [Published by WDC SILSO - Royal Observatory of Belgium (ROB)].

Copernicus Climate Change Service & World Meteorological Organization. (2025). European State of the Climate 2024. https://doi.org/10.24381/14j9-s541

Cotton, W. R., Alexander, G. D., Hertenstein, R., Walko, R. L., McAnelly, R. L., & Nicholls, M. (1995). Cloud venting—a review and some new global annual estimates. *Earth-Science Reviews*, *39*(3-4), 169–206.

Dellaert, S., Super, I., Visschedijk, A., & Denier van der Gon, H. (2019). High resolution scenarios of co2 and co emissions.

Delli Santi, M. G., Insero, G., Bartalini, S., Cancio, P., Carcione, F., Galli, I., Giusfredi, G., Mazzotti, D., Bulgheroni, A., Martinez Ferri, A. I., et al. (2022). Precise radio-

carbon determination in radioactive waste by a laser-based spectroscopic technique. *Proceedings of the National Academy of Sciences*, *119*(28), e2122122119.

DKRZ. (2025). Levante HPC System Documentation [Accessed: 2025-07-19]. *Deutsches Klimarechenzentrum*.

Doney, S. C., Fabry, V. J., Feely, R. A., & Kleypas, J. A. (2009). Ocean acidification: The other co2 problem. *Annual review of marine science*, *1*(1), 169–192.

Emmenegger, L., Harris, E., Leuenberger, M., & Steinbacher, M. (2025a). Icos atc co release from jungfraujoch (13.9 m), 2016-12-12–2025-03-31. https://hdl.handle.net/11676/1hGW506ud_jGSBdvl6Ibb-KK

Emmenegger, L., Leuenberger, M., & Steinbacher, M. (2025b). Icos atc 14c release analysed by icos crl from jungfraujoch (6.0 m), 2015-09-21–2024-11-25. https://hdl.handle.net/11676/AtmTz0jbjTlwKST7jLLrHSHh

Entso-E. (2025). Entso-E Transparency Platform.

European Commission, Directorate-General for Research and Innovation. (2025). European Research Infrastructure Consortium (ERIC) [Last accessed: July 7, 2025].

Fahrni, S. M., Southon, J. R., Santos, G. M., Palstra, S. W., Meijer, H. A., & Xu, X. (2017). Reassessment of the 13c/12c and 14c/12c isotopic fractionation ratio and its impact on high-precision radiocarbon dating. *Geochimica et Cosmochimica Acta*, *213*, 330–345.

Fleisher, A. J., Long, D. A., Liu, Q., Gameson, L., & Hodges, J. T. (2017). Optical measurement of radiocarbon below unity fraction modern by linear absorption spectroscopy. *The journal of physical chemistry letters*, *8*(18), 4550–4556.

Franzke, C. L., Barbosa, S., Blender, R., Fredriksen, H.-B., Laepple, T., Lambert, F., Nilsen, T., Rypdal, K., Rypdal, M., Scotto, M. G., et al. (2020). The structure of climate variability across scales. *Reviews of Geophysics*, *58*(2), e2019RG000657.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quere, C., Li, H., Luijkx, I. T., Olsen, A., et al. (2024). Global carbon budget 2024. *Earth System Science Data Discussions*, *2024*, 1–133.

Frumau, A., & Hensen, A. (2025a). Icos atc 14c release analysed by icos crl from cabauw (207.0 m), 2015-09-23–2024-10-23. https://hdl.handle.net/11676/v51We0vxPbXN6BWz2jKL9Tu

Frumau, A., & Hensen, A. (2025b). Icos atc co release from cabauw (207.0 m), 2021-09-30–2025-03-31. https://hdl.handle.net/11676/Cmow_6bfKVPIweNh4dFkh37F

Gachkivskyi, M., Karstens, U., Fischer, B., Kubistin, D., Müller-Williams, J., Lindauer, M., & Levin, I. (2025). Radon-222 monitoring at german icos atmosphere stations. *Earth System Science Data Discussions*, *2025*, 1–33.

Galli, I., Bartalini, S., Cancio, P., De Natale, P., Mazzotti, D., Giusfredi, G., Fedi, M., & Mando, P. (2013). Optical detection of radiocarbon dioxide: First results and ams intercomparison. *Radiocarbon*, *55*(2), 213–223.

Gamnitzer, U., Karstens, U., Kromer, B., Neubert, R. E., Meijer, H. A., Schroeder, H., & Levin, I. (2006). Carbon monoxide: A quantitative tracer for fossil fuel co2? *Journal of Geophysical Research: Atmospheres*, *111*(D22).

Garreaud, R. D., & Aceituno, P. (2007). Atmospheric circulation and climatic variability. *The physical geography of South America*, *45*, 59.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A., Dargaville, R., Aalto, T., Brandt, J., Christensen, J., et al. (2007). Comparing atmospheric transport models for future regional inversions over europe–part 1: Mapping the atmospheric co 2 signals. *Atmospheric Chemistry and Physics*, *7*(13), 3461–3479.

Gerbig, C., Körner, S., & Lin, J. (2008). Vertical mixing in atmospheric tracer transport models: Error characterization and propagation. *Atmospheric Chemistry and Physics*, *8*(3), 591–602.

Gerbig, C., Lin, J., Wofsy, S., Daube, B., Andrews, A., Stephens, B., Bakwin, P., & Grainger, C. (2003). Toward constraining regional-scale fluxes of co2 with atmospheric observations over a continent: 2. analysis of cobra data using a receptor-oriented framework. *Journal of Geophysical Research: Atmospheres*, *108*(D24).

Goldberg, D. L., Lu, Z., Oda, T., Lamsal, L. N., Liu, F., Griffin, D., McLinden, C. A., Krotkov, N. A., Duncan, B. N., & Streets, D. G. (2019). Exploiting omi no2 satellite observations to infer fossil-fuel co2 emissions from us megacities. *Science of The Total Environment*, *695*, 133805.

Gómez-Ortiz, C., Monteil, G., Basu, S., & Scholze, M. (2025). A co2–Δ 14 co2 inversion setup for estimating european fossil co2 emissions. *Atmospheric Chemistry and Physics*, *25*(1), 397–424.

Graven, H. D. (2015). Impact of fossil fuel emissions on atmospheric radiocarbon and various applications of radiocarbon over this century. *Proceedings of the National Academy of Sciences*, *112*(31), 9542–9545.

Grubb, M. (2019). United nations framework convention on climate change. In *The'earth summit'agreements: A guide and assessment* (pp. 61–74). Routledge.

Hammer, S., Friedrich, R., Kromer, B., Cherkinsky, A., Lehman, S. J., Meijer, H. A., Nakamura, T., Palonen, V., Reimer, R. W., Smith, A. M., et al. (2017). Compatibility of atmospheric 14co2 measurements: Comparing the heidelberg low-level counting facility to international accelerator mass spectrometry (ams) laboratories. *Radiocarbon*, *59*(3), 875–883.

Hatakka, J. (2025a). Icos atc 14c release analysed by icos crl from pallas (12.0 m), 2017-12-14–2024-09-25. https://hdl.handle.net/11676/JsgTpwkluSBYj__HUkGuYe6P

Hatakka, J. (2025b). Icos atc co release from pallas (12.0 m), 2017-09-16–2025-03-31. https://hdl.handle.net/11676/iJrpuTQy_EY6Nlw-T1vAb_i_

Heiskanen, J., Brümmer, C., Buchmann, N., Calfapietra, C., Chen, H., Gielen, B., Gkritzalis, T., Hammer, S., Hartman, S., Herbst, M., et al. (2022). The integrated carbon observation system in europe. *Bulletin of the American Meteorological Society*, *103*(3), E855–E872.

Heliasz, M., & Biermann, T. (2024). Icos atc meteo release from hyltemossa (150.0 m), 2017-09-26–2024-03-31. https://hdl.handle.net/11676/jiS8ewz0BBeSWrl09A7kWUo0

Heliasz, M., & Biermann, T. (2025a). Icos atc 14c release analysed by icos crl from hyltemossa (150.0 m), 2015-09-23–2024-09-06. https://hdl.handle.net/11676/K26g69zog4nycvv2BOkb

Heliasz, M., & Biermann, T. (2025b). Icos atc co release from hyltemossa (150.0 m), 2017-04-17–2025-03-31. https://hdl.handle.net/11676/opUhmOel7cewa0ko-kI407A2

Hofmann, D. J., Butler, J. H., & Tans, P. P. (2009). A new look at atmospheric carbon dioxide. *Atmospheric Environment*, *43*(12), 2084–2086.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., & Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of geophysics*, *33*(4), 403–439.

ICOS ERIC. (2025). Integrated Carbon Observation System: ICOS [Last accessed: July 7, 2025].

ICOS RI. (2020). Icos atmosphere station specifications v2.0 (editor: O. laurent). https://doi.org/10.18160/GK28-2188

IEA. (2024). *World Energy Outlook 2024* (Licence: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A)). https://www.iea.org/reports/world-energy-outlook-2024

IPCC. (2006). Guidelines for national greenhouse gas inventories. *Prepared by the National Greenhouse Gas Inventories Programme. Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K, editors. Published: IGES, Japan*.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G., Peters, J. A., et al. (2019). Edgar v4. 3.2 global atlas of the three major greenhouse gas emissions for the period 1970–2012. *Earth System Science Data*, *11*(3), 959–1002.

Jäschke, C. (2021). *Potentials and limitations of proxy to fossil fuel co2 ratios - a case study at the icos station near karlsruhe* [Master's thesis]. Heidelberg University.

Jiang, J., & McCartt, A. D. (2024). Mid-infrared trace detection with parts-per-quadrillion quantitation accuracy: Expanding frontiers of radiocarbon sensing. *Proceedings of the National Academy of Sciences*, *121*(15), e2314441121.

Jones, M., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F., & Le Quere, C. (2021). Gridded fossil co2 emissions and related o2 combustion consistent with national inventories 1959-2020.

Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Dou, X., Liu, Z., Pickers, P., Ciais, P., Patra, P. K., Chevallier, F., & Le Quéré, C. (2022, September). *Gridded fossil co2 emissions and related o2 combustion consistent with national inventories* (Version GCP-GridFEDv2022.2). Zenodo. https://doi.org/10.5281/zenodo.7016360

Jordan, A., & Brand, W. A. (2003). Technical report: Mpi-bgc, germany. *Report of the 11th WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracer Measurement Techniques (Tokyo, Japan, 25.-28. September 2001)*.

Joyce, J. (2003). Bayes' theorem.

Khalil, M., Pinto, J., & Shearer, M. (1999). Atmospheric carbon monoxide. *Chemosphere-Global change science*, *1*(1-3), ix–xi.

Knaack, T. T. (2025). *Improving capabilities of 14co2-based fossil fuel co2 estimates in europe* [Master's Thesis]. Heidelberg University.

Konovalov, I. B., Berezin, E. V., Ciais, P., Broquet, G., Zhuravlev, R. V., & Janssens-Maenhout, G. (2016). Estimation of fossil-fuel co 2 emissions using satellite measurements of" proxy" species. *Atmospheric Chemistry and Physics*, *16*(21), 13509–13540.

Kromer, B., Lindauer, S., Synal, H.-A., & Wacker, L. (2013). Mams–a new ams facility at the curt-engelhorn-centre for achaeometry, mannheim, germany. *Nuclear instruments and methods in physics research Section B: beam interactions with materials and atoms*, *294*, 11–13.

Kromer, B., & Münnich, K. O. (1992). Co2 gas proportional counting in radiocarbon dating—review and perspective. *Radiocarbon after four decades: An interdisciplinary perspective*, 184–197.

Krystek, M., & Anton, M. (2007). A weighted total least-squares algorithm for fitting a straight line. *Measurement Science and Technology*, *18*(11), 3438.

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., & Müller-Williams, J. (2024a). Icos atc meteo release from steinkimmen (252.0 m), 2019-08-27–2024-03-31. https://hdl.handle.net/11676/BTq7skY564wAfFr1gRMhqgNm

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., & Müller-Williams, J. (2025a). Icos atc 14c release analysed by icos crl from steinkimmen (252.0 m), 2019-07-27–2024-09-07. https://hdl.handle.net/11676/ejFtdDsVa7KQv15ZBu6dtWFx

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Schumacher, M. (2024b). Icos atc meteo release from lindenberg (98.0 m), 2017-02-21–2024-03-31. https://hdl.handle.net/11676/ejumpR8yaFWEp6XWrOUW88Hj

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Schumacher, M. (2025b). Icos atc 14c release analysed by icos crl from karlsruhe (200.0 m), 2018-03-28–2021-09-20. https://hdl.handle.net/11676/G7YOZGhmICr9kOfH0TYAmU

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Schumacher, M. (2025c). Icos atc 14c release analysed by icos crl from lindenberg (98.0 m), 2018-03-15–2024-12-03. https://hdl.handle.net/11676/a1BkYXp-CBFUMmNjV6XWKFoC

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., Schumacher, M., & Zwerschke, E. (2025d). Icos atc 14c release analysed by icos crl from hohenpeissenberg (131.0 m), 2015-09-24–2025-02-12. https://hdl.handle.net/11676/-7c3LnsTumL0m3g1Vr0X8fA2

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., Schumacher, M., & Zwerschke, E. (2025e). Icos atc co release from gartow (341.0 m), 2017-04-11–2025-03-31. https://hdl.handle.net/11676/lXTd77IBEj-kxQO8f4VG0Ot5

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., Schumacher, M., & Zwerschke, E. (2025f). Icos atc co release from hohenpeissenberg (131.0 m), 2017-02-14–2025-03-31. https://hdl.handle.net/11676/mrIys0VsZ9ahgf1_L4TzR3uL

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Zwerschke, E. (2025g). Icos atc co release from karlsruhe (200.0 m), 2019-01-31–2025-03-31. https://hdl.handle.net/11676/0PbsKIzLLSPTpGgR1igBzaay

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Zwerschke, E. (2025h). Icos atc co release from lindenberg (98.0 m), 2018-08-24–2025-03-31. https://hdl.handle.net/11676/9mohjNIQ_IHKO2oY5KdI1UhM

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Zwerschke, E. (2025i). Icos atc co release from ochsenkopf (163.0 m), 2019-09-25–2025-03-31. https://hdl.handle.net/11676/EB71bQD6HKNO0aKoerXBBMXc

Kubistin, D., Plaß-Dülmer, C., Arnold, S., Kneuer, T., Lindauer, M., Müller-Williams, J., & Zwerschke, E. (2025j). Icos atc co release from steinkimmen (252.0 m), 2019-07-22–2025-03-31. https://hdl.handle.net/11676/eCX_5R9pLYLINP1jyKp4fYIC

Kubistin, D., Plaß-Dülmer, C., Kneuer, T., Lindauer, M., & Müller-Williams, J. (2025k). Icos atc 14c release analysed by icos crl from gartow (341.0 m), 2021-05-20–2024-12-22. https://hdl.handle.net/11676/egzs14iC4LxiOgjdiJsDGzZt

Kubistin, D., Plaß-Dülmer, C., Kneuer, T., Lindauer, M., Müller-Williams, J., & Zwerschke, E. (2025l). Icos atc 14c release analysed by icos crl from ochsenkopf (163.0 m), 2021-04-21–2025-01-20. https://hdl.handle.net/11676/8nCqQNyIib6tMf8ykoOuXcnq

Kuderer, M., Hammer, S., & Levin, I. (2018). The influence of 14 co 2 releases from regional nuclear facilities at the heidelberg 14 co 2 sampling site (1986–2014). *Atmospheric Chemistry and Physics*, *18*(11), 7951–7959.

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., & Denier van der Gon, H. (2022). Cams-reg-v4: A state-of-the-art high-resolution european emission inventory for air quality modelling. *Earth System Science Data*, *14*(2), 491–515.

Kwok, C. Y., Laurent, O., Guemri, A., Philippon, C., Wastine, B., Rella, C., Vuillemin, C., Truong, F., Delmotte, M., Kazan, V., et al. (2015). Comprehensive laboratory and field testing of cavity ring-down spectroscopy analyzers measuring h 2 o, co 2, ch 4 and co. *Atmospheric Measurement Techniques*, *8*(9), 3867–3892.

Laemmel, T., Knaack, T., Heckel, A., Wenger, A., & Szidat, S. (2025, March). *Compilation of nuclear industry's gaseous radiocarbon emissions from 1950 to 2023*. Zenodo. https://doi.org/10.5281/zenodo.15034663

Lan, X., & Keeling, R. (2025). Atmospheric carbon dioxide data from noaa/gml and scripps institution of oceanography.

Larmanou, E., Marklund, P., Ottosson-Löfvenius, M., & Smith, P. (2025a). Icos atc 14c release analysed by icos crl from svartberget (150.0 m), 2016-02-09–2025-01-28. https://hdl.handle.net/11676/BbQgw7Ik1bGRI60OGbIblMGD

Larmanou, E., Marklund, P., Ottosson-Löfvenius, M., & Smith, P. (2025b). Icos atc co release from svartberget (150.0 m), 2017-06-01–2025-03-31. https://hdl.handle.net/11676/sWfTdyd6r_s6FVQPcaqwR3mm

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., Trisos, C., Romero, J., Aldunce, P., & Ruane, A. C. (2024). Climate change 2023 synthesis report summary for policymakers. *CLIMATE CHANGE 2023 Synthesis Report: Summary for Policymakers*.

Lehner, I., & Molder, M. (2025a). Icos atc 14c release analysed by icos crl from norunda (100.0 m), 2015-09-29–2025-01-21. https://hdl.handle.net/11676/FWluei90eXz8Ak7eiI0kuwGm

Lehner, I., & Molder, M. (2025b). Icos atc co release from norunda (100.0 m), 2017-04-01–2025-03-31. https://hdl.handle.net/11676/B8ztQh-xTs8cYnPCralUTA2t

Levin, I., Preunkert, S., Graven, H., Lewis, C., Miller, J., Turnbull, J., Xu, X., & Hammer, S. (2024). Database of existing d14co2 measurements [Funded by the European Union to enhance CO_2 emissions monitoring within the Copernicus program.]. Retrieved May 12, 2025, from https://www.corso-project.eu/objectives

Levin, I., & Hammer, S. (2022). Atmospheric station results archive from alert (20.0 m), 1987-11-16–2020-04-06. https://hdl.handle.net/11676/Dc6VHvDfc_N8y53V-04ZNGTs

Levin, I., Hammer, S., Eichelmann, E., & Vogel, F. R. (2011). Verification of greenhouse gas emission reductions: The prospect of atmospheric monitoring in polluted areas. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 1906–1924.

Levin, I., Hammer, S., Kromer, B., Preunkert, S., Weller, R., & Worthy, D. E. (2022). Radiocarbon in global tropospheric carbon dioxide. *Radiocarbon*, *64*(4), 781–791.

Levin, I., & Karstens, U. (2007). Inferring high-resolution fossil fuel co2 records at continental sites from combined 14co2 and co observations. *Tellus B: Chemical and Physical Meteorology*, *59*(2), 245–250.

Levin, I., Karstens, U., Eritt, M., Maier, F., Arnold, S., Rzesanke, D., Hammer, S., Ramonet, M., Vítková, G., Conil, S., et al. (2020). A dedicated flask sampling strategy developed for integrated carbon observation system (icos) stations based on co 2 and co measurements and stochastic time-inverted lagrangian transport (stilt) footprint modelling. *Atmospheric Chemistry and Physics*, *20*(18), 11161–11180.

Levin, I., Kromer, B., Schmidt, M., & Sartorius, H. (2003). A novel approach for independent budgeting of fossil fuel co2 over europe by 14co2 observations. *Geophysical Research Letters*, *30*(23).

Levin, I., Münnich, K. O., & Weiss, W. (1980). The Effect of Anthropogenic CO2 and 14C Sources on the Distribution of 14C in the Atmosphere. *Radiocarbon*, *22*(2), 379–391. https://doi.org/10.1017/S003382220000967X

Levin, I., Naegler, T., Kromer, B., Diehl, M., Francey, R., Gomez-Pelaez, A., Steele, P., Wagenbach, D., Weller, R., & Worthy, D. (2010). Observations and modelling of the global distribution and long-term trend of atmospheric 14co2. *Tellus B: Chemical and Physical Meteorology*, *62*(1), 26–46.

Levin, I., & Rödenbeck, C. (2008). Can the envisaged reductions of fossil fuel co 2 emissions be detected by atmospheric observations? *Naturwissenschaften*, *95*, 203–208.

Levin, I., Schuchard, J., Kromer, B., & Münnich, K. (1989). The continental european suess effect. *Radiocarbon*, *31*(3), 431–440.

Li, P., Lin, B., Cheng, Z., Li, J., Li, J., Chen, D., Zhang, T., Lin, R., Zhu, S., Liu, J., et al. (2025). Drivers and implications of declining fossil fuel co 2 in chinese cities revealed by radiocarbon measurements. *EGUsphere*, *2025*, 1–44.

Libby, W. (1955). *Radiocarbon dating*. University of Chicago Press.

Libby, W. F. (1946). Atmospheric helium three and radiocarbon from cosmic radiation. *Physical Review*, *69*(11-12), 671.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., & Grainger, C. (2003). A near-field tool for simulating the upstream influence of atmospheric observations: The stochastic time-inverted lagrangian transport (stilt) model. *Journal of Geophysical Research: Atmospheres*, *108*(D16).

Lin, J., & Gerbig, C. (2005). Accounting for the effect of transport errors on tracer inversions. *Geophysical Research Letters*, *32*(1).

Lingenfelter, R. E. (1963). Production of carbon 14 by cosmic-ray neutrons. *Reviews of Geophysics*, *1*(1), 35–55.

Lopez, M., Schmidt, M., Delmotte, M., Colomb, A., Gros, V., Janssen, C., Lehman, S. J., Mondelain, D., Perrussel, O., Ramonet, M., et al. (2013). Co, no x and 13 co 2 as tracers for fossil fuel co 2: Results from a pilot study in paris during winter 2010. *Atmospheric Chemistry and Physics*, *13*(15), 7343–7358.

Lux, J. T. (2018). *A new target preparation facility for high precision ams measurements and strategies for efficient 14co2 sampling* [Doctoral dissertation, University of heidelberg].

Lynas, M., Houlton, B. Z., & Perry, S. (2021). Greater than 99% consensus on human caused climate change in the peer-reviewed scientific literature. *Environmental Research Letters*, *16*(11), 114005.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., & Gottlieb, E. W. (2008). A satellite-based biosphere parameterization for net ecosystem co2 exchange: Vegetation photosynthesis and respiration model (vprm). *Global Biogeochemical Cycles*, *22*(2).

Maier, F., Falge, E., Gachkivskyi, M., Henne, S., Karstens, U., Kikaj, D., Levin, I., Manning, A., Rödenbeck, C., & Gerbig, C. (2025). How reliable are process-based 222 radon emission maps? results from an atmospheric 222 radon inversion in europe. *EGUsphere*, *2025*, 1–37.

Maier, F., Gerbig, C., Levin, I., Super, I., Marshall, J., & Hammer, S. (2021). Effects of point source emission heights in wrf–stilt: A step towards exploiting nocturnal observations in models. *Geoscientific Model Development Discussions*, *2021*, 1–25.

Maier, F., Levin, I., Conil, S., Gachkivskyi, M., Denier van der Gon, H., & Hammer, S. (2024a). Uncertainty in continuous Δ co-based Δ ffco 2 estimates derived from 14 c flask and bottom-up Δ co/Δ ffco 2 ratios. *Atmospheric Chemistry and Physics*, *24*(14), 8205–8223.

Maier, F., Levin, I., Gachkivskyi, M., Rödenbeck, C., & Hammer, S. (2023). Estimating regional fossil fuel co2 concentrations from 14co2 observations: Challenges and uncertainties. *Philosophical Transactions of the Royal Society A*, *381*(2261), 20220203.

Maier, F., Rödenbeck, C., Levin, I., Gerbig, C., Gachkivskyi, M., & Hammer, S. (2024b). Potential of 14 c-based vs. Δ co-based Δ ffco 2 observations to estimate urban fossil fuel co 2 (ffco 2) emissions. *Atmospheric Chemistry and Physics*, *24*(14), 8183–8203.

Major, I., Haszpra, L., Rinyu, L., Futó, I., Bihari, Á., Hammer, S., Jull, A. T., & Molnár, M. (2018). Temporal variation of atmospheric fossil and modern co2 excess at a central european rural tower station between 2008 and 2014. *Radiocarbon*, *60*(5), 1285–1299.

Marek, M. V., Vítková, G., & Komínková, K. (2025a). Icos atc 14c release analysed by icos crl from křešín u pacova (250.0 m), 2017-03-29–2024-11-27. https://hdl.handle.net/11676/_QjnCauVQV4MwgpI3H8jTyEY

Marek, M. V., Vítková, G., & Komínková, K. (2025b). Icos atc co release from křešín u pacova (250.0 m), 2017-04-12–2025-03-31. https://hdl.handle.net/11676/dVo5wZ07kZ1v6n3cfbUUs

Martin, D. (2025). *Unpublished picarro co data (instrument ids 955 and 695), received from damien martin upon request* [damien.martin@universityofgalway.ie].

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Pean, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., et al. (2021). Climate change 2021: The physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, *2*(1), 2391.

McCartt, A., Ognibene, T., Bench, G., & Turteltaub, K. (2015). Measurements of carbon-14 with cavity ring-down spectroscopy. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *361*, 277–280.

Miller, J. B., Dyonisius, M., Wolak, C., Cappa, P., Morgan, S., Andrews, A., Lan, X., Moglia, E., & Lehman, S. J. (2025). NOAA GML global greenhouse gas reference network measurements of the radiocarbon composition of atmospheric carbon dioxide ($\Delta^{14}co_2$) at manual (non-pfp) surface flask sampling sites, 2003-2021 [Data set]. https://doi.org/10.15138/9R46-MR76

Mook, W. G. (2000). Environmental isotopes in the hydrological cycle. *Principles and Applications, Volume I*.

Naegler, T., & Levin, I. (2006). Closing the global radiocarbon budget 1945–2005. *Journal of Geophysical Research: Atmospheres*, *111*(D12).

Naegler, T., & Levin, I. (2009a). Biosphere-atmosphere gross carbon exchange flux and the δ13co2 and Δ14co2 disequilibria constrained by the biospheric excess radiocarbon inventory. *Journal of Geophysical Research: Atmospheres*, *114*(D17).

Naegler, T., & Levin, I. (2009b). Observation-based global biospheric excess radiocarbon inventory 1963–2005. *Journal of Geophysical Research: Atmospheres*, *114*(D17).

Němec, M., Wacker, L., & Gäggeler, H. (2010). Optimization of the graphitization process at age-1. *Radiocarbon*, *52*(3), 1380–1393.

NOAA Global Monitoring Laboratory. (2024). Carbon Monoxide (CO) WMO Scale [Accessed on August 2, 2025].

Palmer, P. I., Suntharalingam, P., Jones, D. B., Jacob, D. J., Streets, D. G., Fu, Q., Vay, S. A., & Sachse, G. W. (2006). Using co2: Co correlations to improve inverse analyses of carbon fluxes. *Journal of Geophysical Research: Atmospheres*, *111*(D12).

Palonen, V., Pumpanen, J., Kulmala, L., Levin, I., Heinonsalo, J., & Vesala, T. (2018). Seasonal and diurnal variations in atmospheric and soil air 14co2 in a boreal scots pine forest. *Radiocarbon*, *60*(1), 283–297.

Peng, Y., Hu, C., Ai, X., Li, Y., Gao, L., Liu, H., Zhang, J., & Xiao, W. (2023). Improvements of simulating urban atmospheric co2 concentration by coupling with emission height and dynamic boundary layer variations in wrf-stilt model. *Atmosphere*, *14*(2), 223.

Petron, G., Crotwell, A. M., Crotwell, M. J., Dlugokencky, E., Madronich, M., Moglia, E., Neff, D., Thoning, K., Wolter, S., & Mund, J. W. (2022). Atmospheric carbon monoxide dry air mole fractions from the noaa gml carbon cycle cooperative global air sampling network, 1988-2021. https://doi.org/10.15138/33bv-s284

Pickers, P. A., Manning, A. C., Le Quere, C., Forster, G. L., Luijkx, I. T., Gerbig, C., Fleming, L. S., & Sturges, W. T. (2022). Novel quantification of regional fossil fuel co2 reductions during covid-19 lockdowns using atmospheric oxygen measurements. *Science advances*, *8*(16), eabl9250.

Press, W. H., Vetterling, W. T., Teukolsky, S. A., & Flannery, B. P. (1988). *Numerical recipes*. Cambridge University Press, London, England.

Price, J., & Warren, R. (2016). Literature review of the potential of "blue carbon" activities to reduce emissions.

RADD. (2025). European Commission RAdioactive Discharges Database: Standard Reports [Accessed: July 10, 2025].

Ramonet, M., Conil, S., Delmotte, M., Laurent, O., & Lopez, M. (2025a). Icos atc 14c release analysed by icos crl from observatoire pérenne de l'environnement (120.0 m), 2015-09-28–2025-01-06. https://hdl.handle.net/11676/SPsTcsvnB4c7VfEg2TnpVfVe

Ramonet, M., Conil, S., Delmotte, M., Laurent, O., & Lopez, M. (2025b). Icos atc co release from observatoire pérenne de l'environnement (120.0 m), 2016-08-18–2025-03-31. https://hdl.handle.net/11676/0XgPlP1-1DTSLEgWXW5MQXwv

Ramonet, M., Delmotte, M., & Lopez, M. (2025c). Icos atc 14c release analysed by icos crl from saclay (100.0 m), 2018-05-11–2024-10-29. https://hdl.handle.net/11676/hCausDA2KFK0EHwtTUlbPIQY

Ramonet, M., Delmotte, M., & Lopez, M. (2025d). Icos atc co release from saclay (100.0 m), 2017-05-31–2025-03-31. https://hdl.handle.net/11676/Prknecf5rYiJ2D9-enWh9HGm

Ramonet, M., Lopez, M., & Delmotte, M. (2025e). Icos atc 14c release analysed by icos crl from trainou (180.0 m), 2015-09-29–2024-10-03. https://hdl.handle.net/11676/-3FdVO02SaoiJBZrbgFnXanF

Randerson, J., Enting, I., Schuur, E., Caldeira, K., & Fung, I. (2002). Seasonal and latitudinal variability of troposphere Δ14co2: Post bomb contributions from fossil fuels, oceans, the stratosphere, and the terrestrial biosphere. *Global Biogeochemical Cycles*, *16*(4), 59–1.

Rödenbeck, C., Houweling, S., Gloor, M., & Heimann, M. (2003). Co 2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, *3*(6), 1919–1964.

Rödenbeck, C. (2005a). Estimating co2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport.

Rödenbeck, C. (2005b). Jena Inversion System - HowTo [Max Planck Institute for Biogeochemistry, Jena. Accessed: 2025-07-13].

Rödenbeck, C., Adcock, K. E., Eritt, M., Gachkivskyi, M., Gerbig, C., Hammer, S., Jordan, A., Keeling, R. F., Levin, I., Maier, F., et al. (2023). The suitability of atmospheric oxygen measurements to constrain western european fossil-fuel co 2 emissions and their trends. *Atmospheric Chemistry and Physics*, *23*(24), 15767–15782.

Rosendahl, C. L. (2022). *Proxy to fossil-fuel co2 emission ratios: In-situ versus inventory data* [Doctoral dissertation, Heidelberg University].

Scarpelli, T. R., Palmer, P. I., Lunt, M., Super, I., & Droste, A. (2024). Verifying national inventory-based combustion emissions of co 2 across the uk and mainland europe using satellite observations of atmospheric co and co 2. *Atmospheric Chemistry and Physics*, *24*(18), 10773–10791.

Schlesinger, W. H., & Bernhardt, E. S. (2020). The atmosphere. *Biogeochemistry*, 51.

Schmidt, M., Hoheisel, A., & Meinhardt, F. (2025). Icos atc 14c release analysed by icos crl from schauinsland (12.0 m), 2015-09-28–2024-05-27. https://hdl.handle.net/11676/e06vzI55aZdsbW-x1D23IPP3

Schoch, H., Bruns, M., Münnich, K. O., & Münnich, M. (1980). A multi-counter system for high precision carbon-14 measurements. *Radiocarbon*, *22*(2), 442–447.

Schuldt, K. N., Aalto, T., Aaltonen, H., Andrews, A., Apadula, F., Arnold, S., Baier, B., Bergamaschi, P., Biermann, T., Biraud, S. C., Blot, R., Brand, W. A., Bäni, L., Calzolari, F., Catoire, V., Chen, H., Chen, G., Colomb, A., Commane, R., … di Sarra,

A. G. (2025). Multi-laboratory compilation of atmospheric carbon monoxide data for the period 1984-2023 [obspack_co_1_GLOBALVIEWplus_v5.0_2025-01-22].

Schuur, E., Carbone, M., Hicks Pries, C., Hopkins, F., & Natali, S. (2016). Radiocarbon in terrestrial systems. *Radiocarbon and climate change: Mechanisms, applications and laboratory techniques*, 167–220.

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., et al. (2024). The era5 global reanalysis from 1940 to 2022. *Quarterly Journal of the Royal Meteorological Society*, *150*(764), 4014–4048.

Solazzo, E., Crippa, M., Guizzardi, D., Muntean, M., Choulga, M., & Janssens-Maenhout, G. (2021). Uncertainties in the emissions database for global atmospheric research (edgar) emission inventory of greenhouse gases. *Atmospheric Chemistry and Physics*, *21*(7), 5655–5683.

Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., & Mukai, H. (2011). The co 2 release and oxygen uptake from fossil fuel emission estimate (coffee) dataset: Effects from varying oxidative ratios. *Atmospheric Chemistry and Physics*, *11*(14), 6855–6870.

Stenström, K. E., Skog, G., Georgiadou, E., Genberg, J., & Johansson, A. (2011). A guide to radiocarbon units and calculations. *Lund University, Department of Physics internal report*, 1–17.

Stuiver, M., & Polach, H. A. (1977). Discussion reporting of 14c data. *Radiocarbon*, *19*(3), 355–363.

Stuiver, M., & Robinson, S. W. (1974). University of washington geosecs north atlantic carbon-14 results. *Earth and Planetary Science Letters*, *23*(1), 87–90.

Stull, R. B. (2012). *An introduction to boundary layer meteorology* (Vol. 13). Springer Science & Business Media.

Suess, H. E. (1955). Radiocarbon concentration in modern wood. *Science*, *122*(3166), 415–417.

Super, I., Dellaert, S. N., Visschedijk, A. J., & Denier van der Gon, H. A. (2020). Uncertainty analysis of a european high-resolution emission inventory of co 2 and co to support inverse modelling and network design. *Atmospheric Chemistry and Physics*, *20*(3), 1795–1816.

Synal, H.-A. (2022). Accelerator mass spectrometry: Ultra-sensitive detection technique of long-lived radionuclides. *Chimia*, *76*(1-2), 45–51.

Synal, H.-A., Stocker, M., & Suter, M. (2007). Micadas: A new compact radiocarbon ams system. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *259*(1), 7–13.

Szopa, S., Naik, V., Adhikary, B., Artaxo, P., Berntsen, T., Collins, W., Fuzzi, S., Gallardo, L., Kiendler-Scharr, A., Klimont, Z., Liao, H., Unger, N., & Zanis, P. (2021). Short-lived climate forcers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (pp. 817–922). Cambridge University Press. https://doi.org/10.1017/9781009157896.008

Tans, P. P., Berry, J. A., & Keeling, R. F. (1993). Oceanic 13c/12c observations: A new window on ocean co2 uptake. *Global biogeochemical cycles*, *7*(2), 353–368.

Thoning, K. W., Tans, P. P., & Komhyr, W. D. (1989). Atmospheric carbon dioxide at mauna loa observatory: 2. analysis of the noaa gmcc data, 1974–1985. *Journal of Geophysical Research: Atmospheres*, *94*(D6), 8549–8565.

Trusilova, K., Rödenbeck, C., Gerbig, C., & Heimann, M. (2010). A new coupled system for global-to-regional downscaling of co 2 concentration estimation. *Atmospheric Chemistry and Physics*, *10*(7), 3205–3213.

Tukey, J. W., et al. (1977). *Exploratory data analysis* (Vol. 2). Springer.

Turnbull, J. C., Miller, J., Lehman, S., Tans, P., Sparks, R., & Southon, J. (2006). Comparison of 14co2, co, and sf6 as tracers for recently added fossil fuel co2 in the atmosphere and implications for biological co2 exchange. *Geophysical research letters*, *33*(1).

Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., Davis, K. J., Lauvaux, T., Miles, N. L., Richardson, S. J., et al. (2015). Toward quantification and source sector identification of fossil fuel co2 emissions from an urban area: Results from the influx experiment. *Journal of Geophysical Research: Atmospheres*, *120*(1), 292–312.

UNFCCC. (2015). The Paris Agreement [Accessed: 2025-07-19]. *United Nations Framework Convention on Climate Change*.

Van der Laan, S., Neubert, R., & Meijer, H. (2009). A single gas chromatograph for accurate atmospheric mixing ratio measurements of co 2, ch 4, n 2 o, sf 6 and co. *Atmospheric Measurement Techniques*, *2*(2), 549–559.

Vogel, F., Hamme, S., Steinhof, A., Kromer, B., & Levin, I. (2010). Implication of weekly and diurnal 14c calibration on hourly estimates of co-based fossil fuel co2 ata moderately polluted site in southwestern germany. *Tellus B: Chemical and Physical Meteorology*, *62*(5), 512–520.

Vogel, F., Hammer, S., Levin, I., Maier, F., Karstens, U., Meijer, H., Chen, H., Nguyen, L., Ramonet, M., Peters, W., et al. (2018). Constraining emissions of co2 and co-emitted species in european cities. *EGU General Assembly Conference Abstracts*, 13350.

Wacker, L., Bonani, G., Friedrich, M., Hajdas, I., Kromer, B., Němec, M., Ruff, M., Suter, M., Synal, H.-A., & Vockenhuber, C. (2010). Micadas: Routine and high-precision radiocarbon dating. *Radiocarbon*, *52*(2), 252–262.

Warneck, P. (1999). *Chemistry of the natural atmosphere* (Vol. 71). Elsevier.

Wurm, M. (2021). A universal and fast method to solve linear systems with correlated coefficients using weighted total least squares. *Measurement Science and Technology*, *33*(1), 015017.

Zannetti, P. (1990). Lagrangian dispersion models. In *Air pollution modeling: Theories, computational methods and available software* (pp. 185–222). Springer.

Zhou, W., Niu, Z., Wu, S., Xiong, X., Hou, Y., Wang, P., Feng, T., Cheng, P., Du, H., Lu, X., et al. (2020). Fossil fuel co2 traced by radiocarbon in fifteen chinese cities. *Science of the Total Environment*, *729*, 138639.

Appendix A

Additional figures

A.1 Analysis of region of interest (ROI), 2014-2024

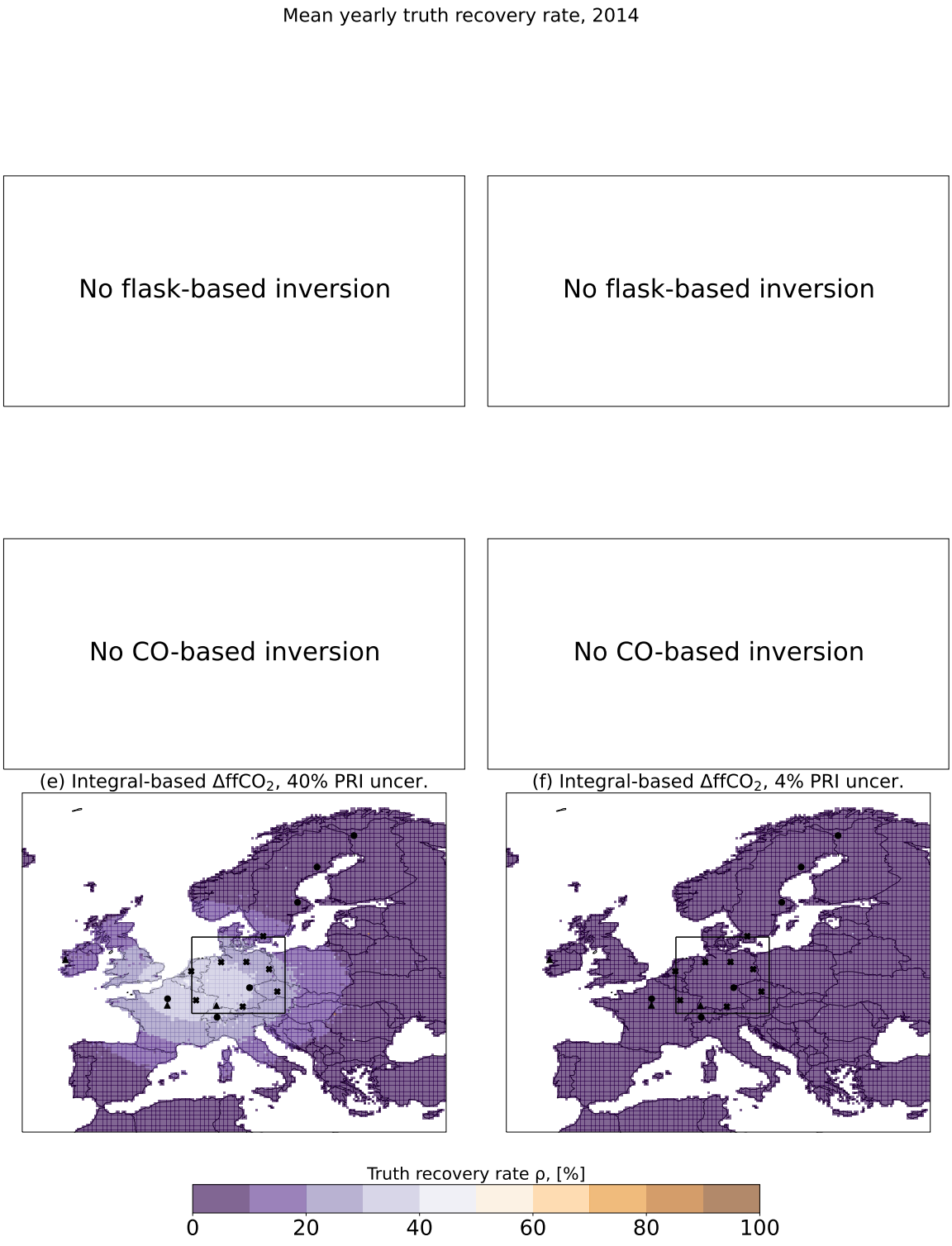

Figure A.1. Mean spacial constrain of the inversion in 2014 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

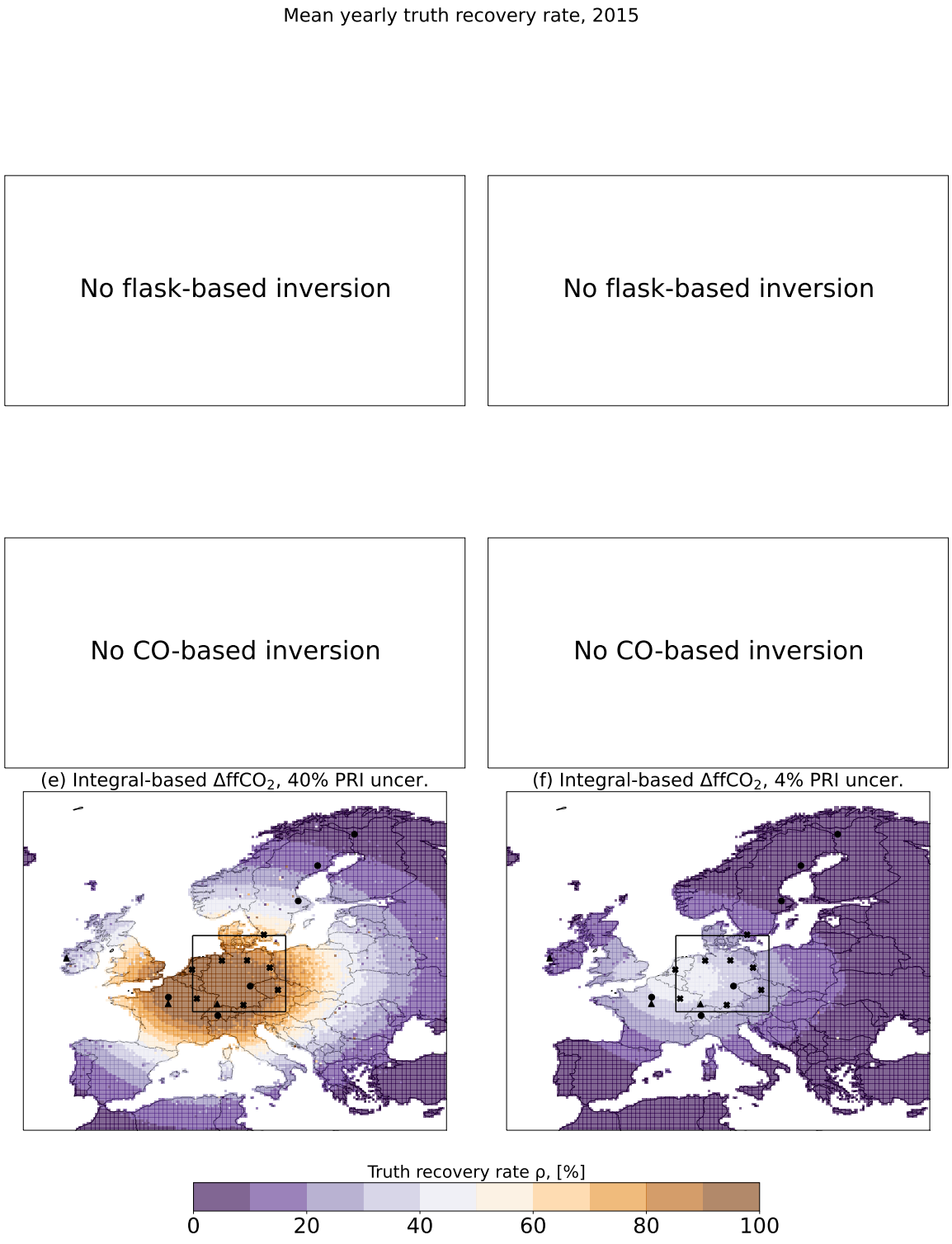

Figure A.2. Mean spacial constrain of the inversion in 2015 for three ΔffCO$_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

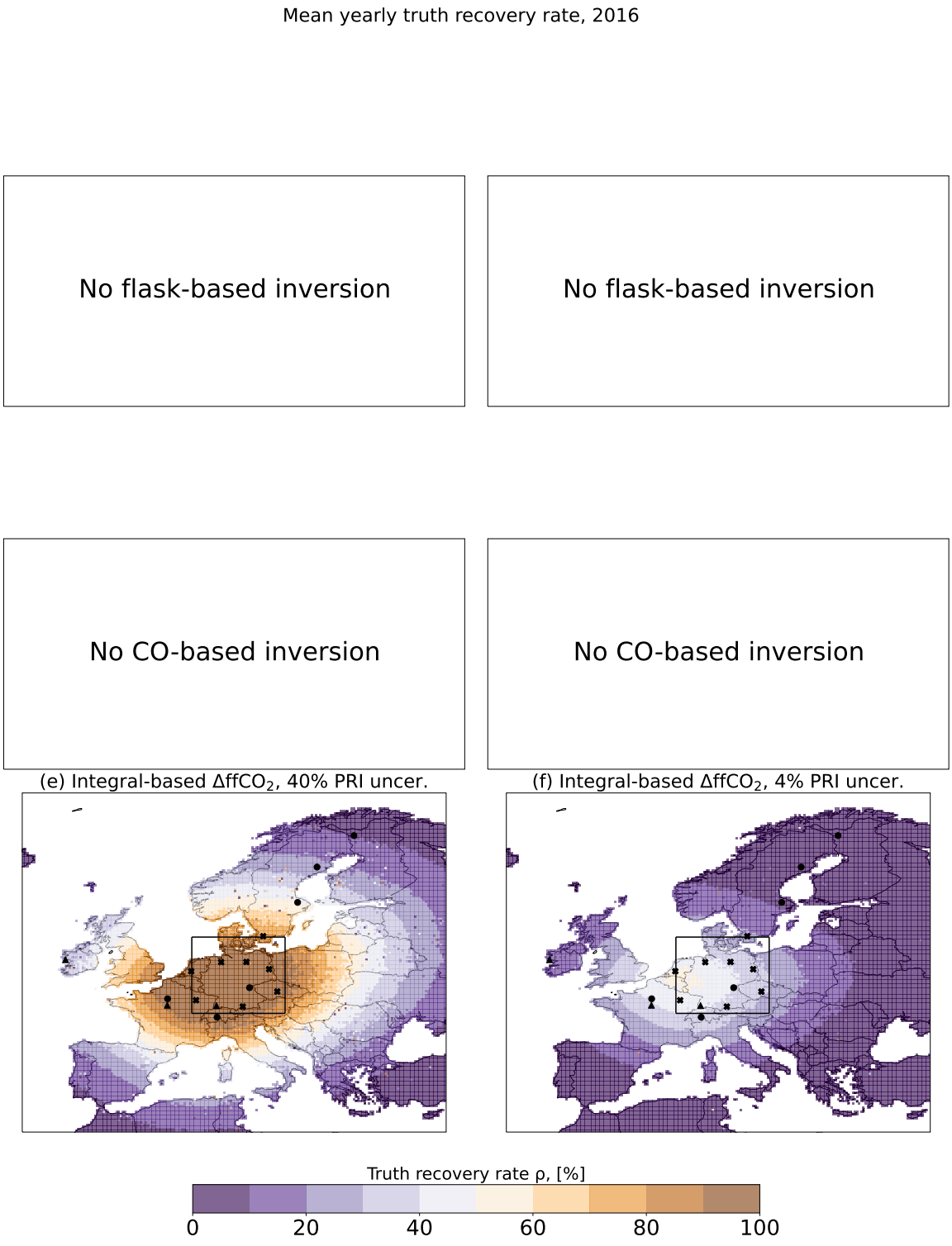

Figure A.3. Mean spacial constrain of the inversion in 2016 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

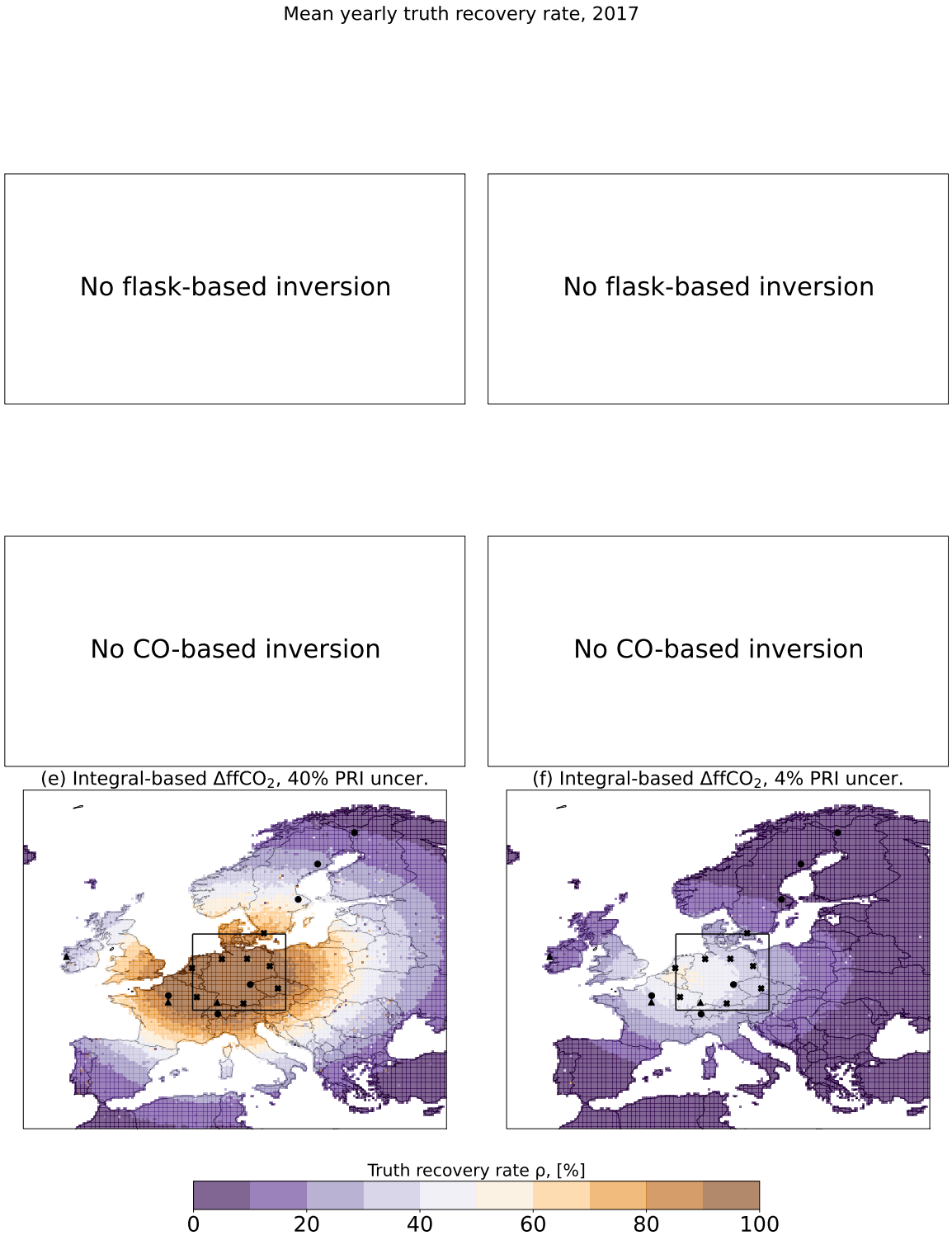

Figure A.4. Mean spacial constrain of the inversion in 2017 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

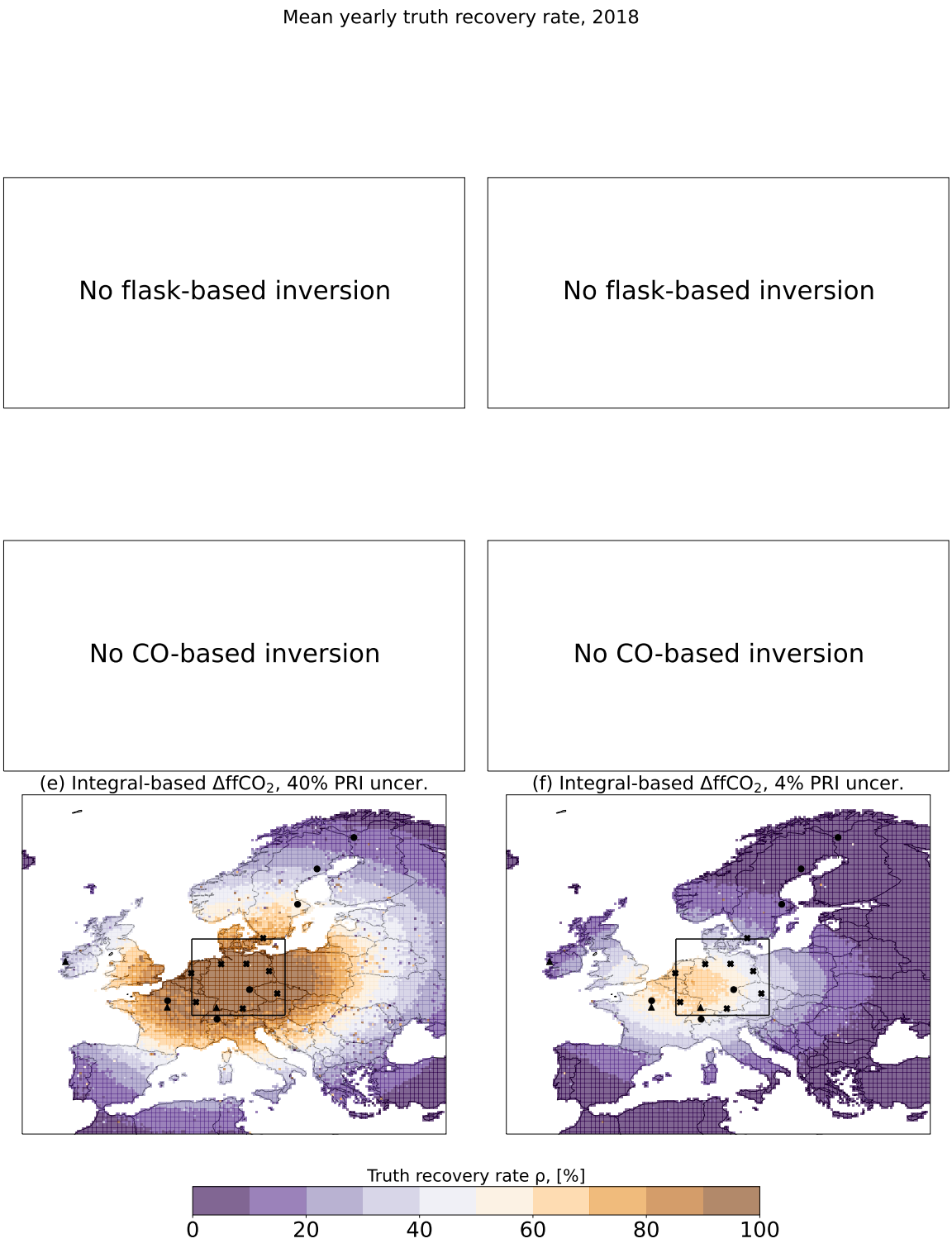

Figure A.5. Mean spacial constrain of the inversion in 2018 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

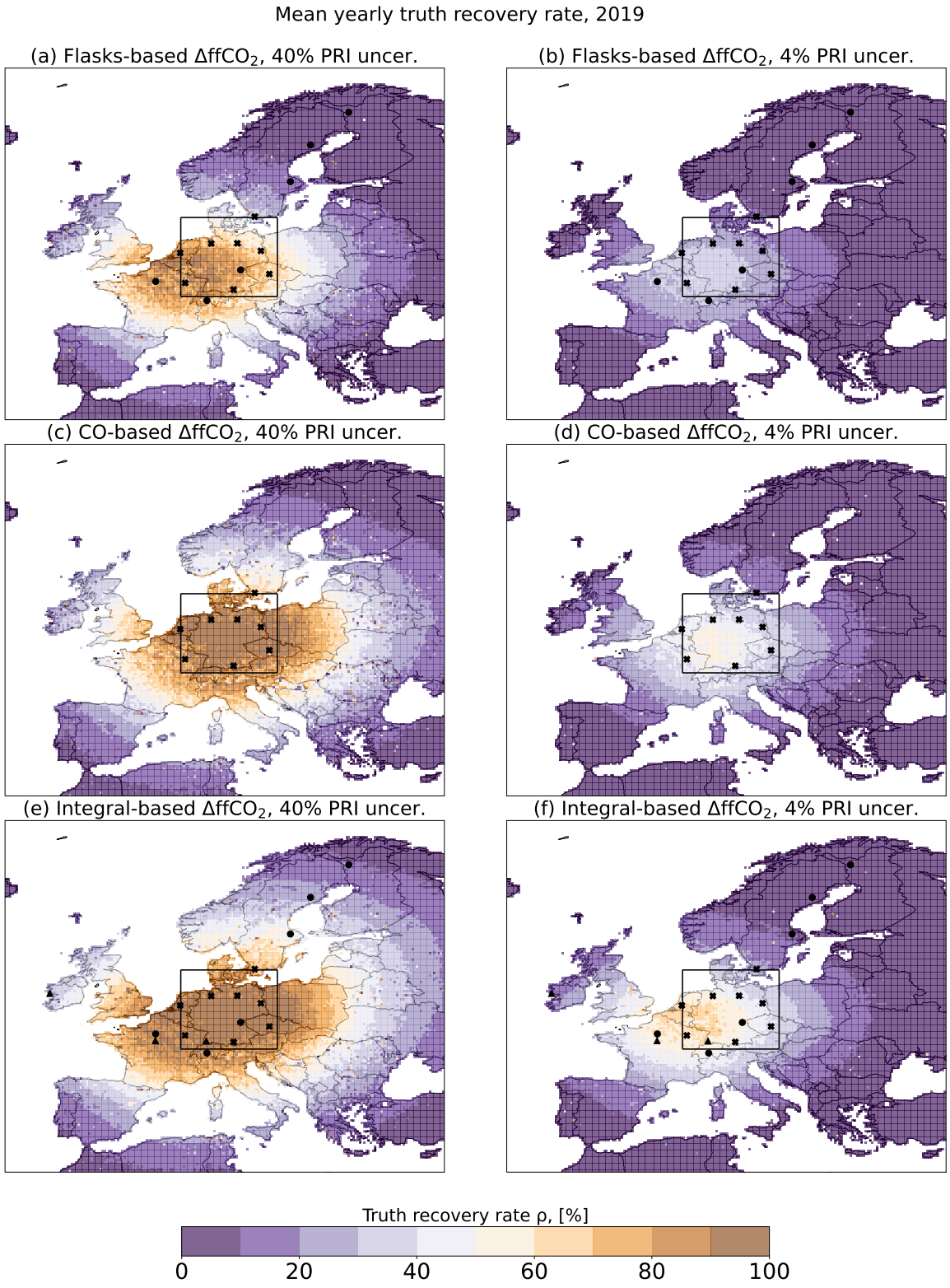

Figure A.6. Mean spacial constrain of the inversion in 2019 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

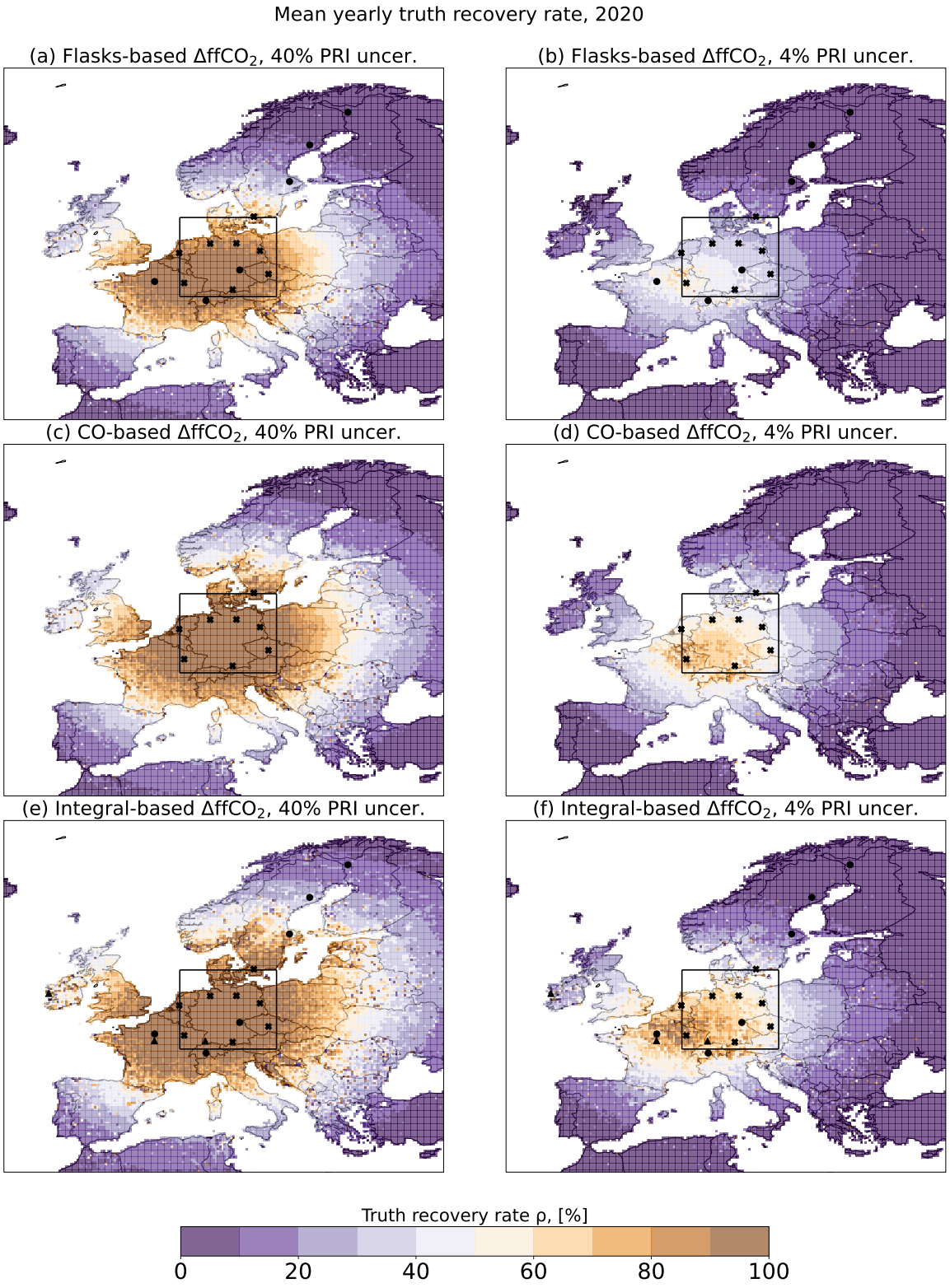

Figure A.7. Mean spacial constrain of the inversion in 2020 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

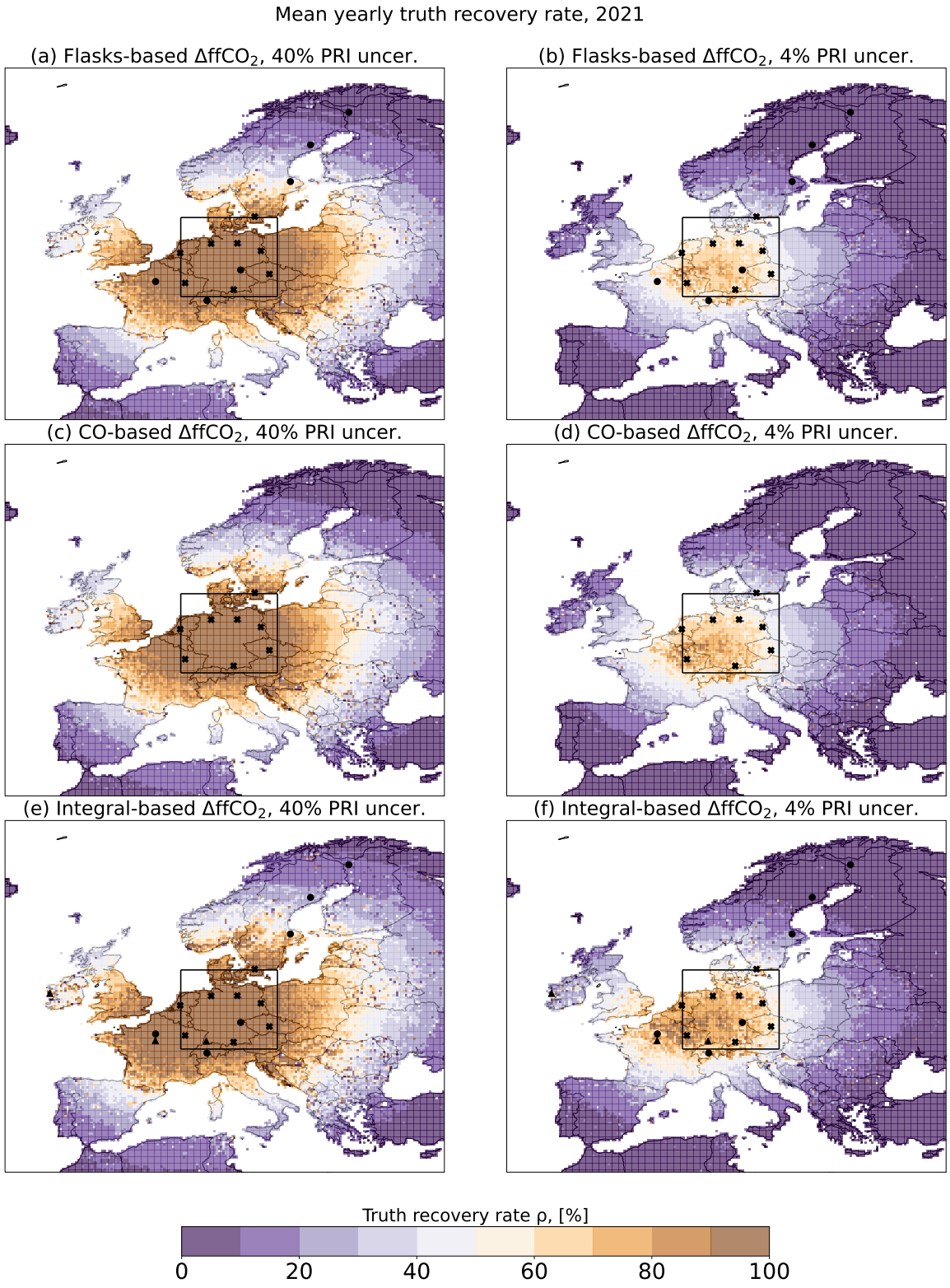

Figure A.8. Mean spacial constrain of the inversion in 2021 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

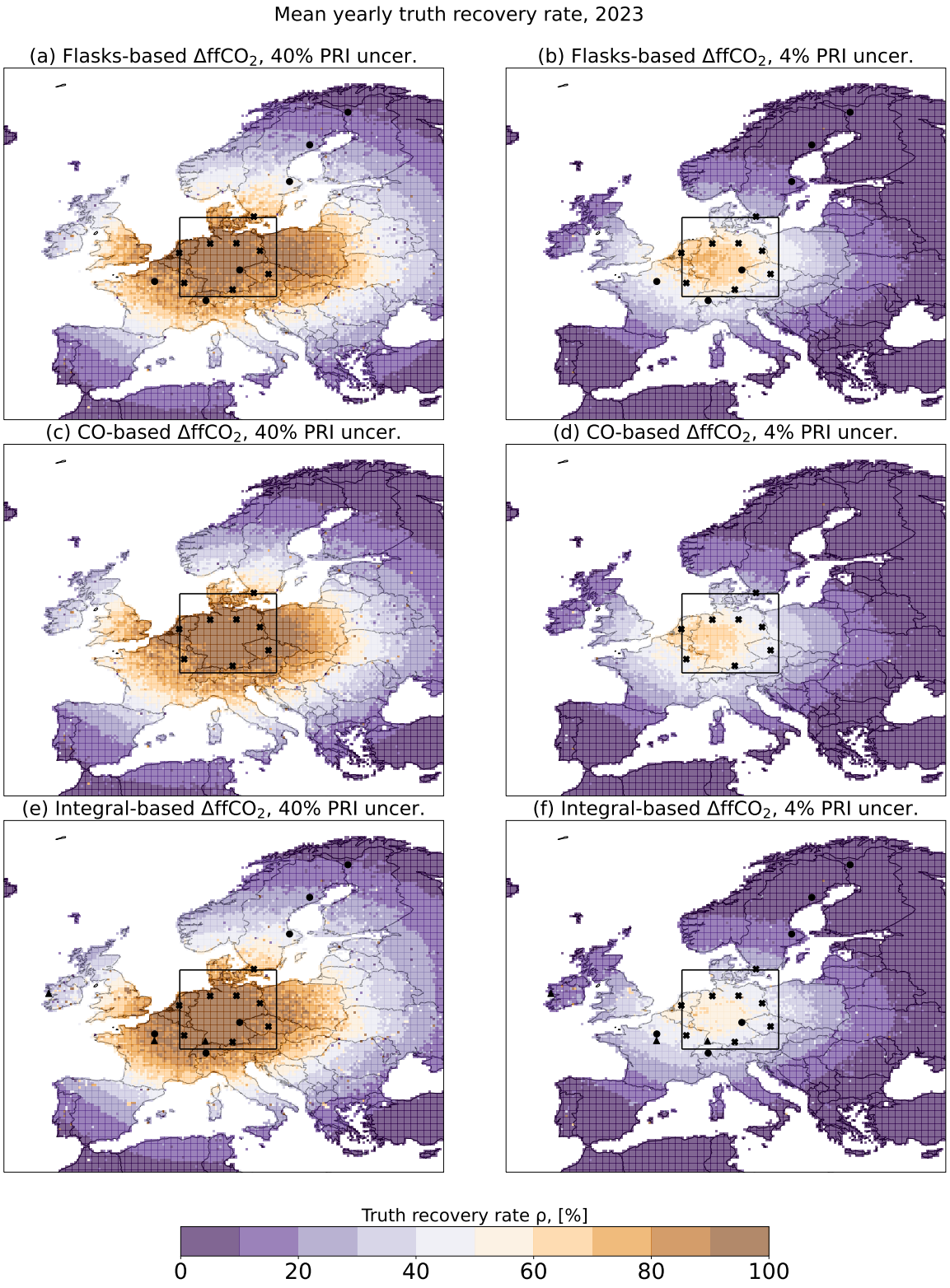

Figure A.9. Mean spacial constrain of the inversion in 2023 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

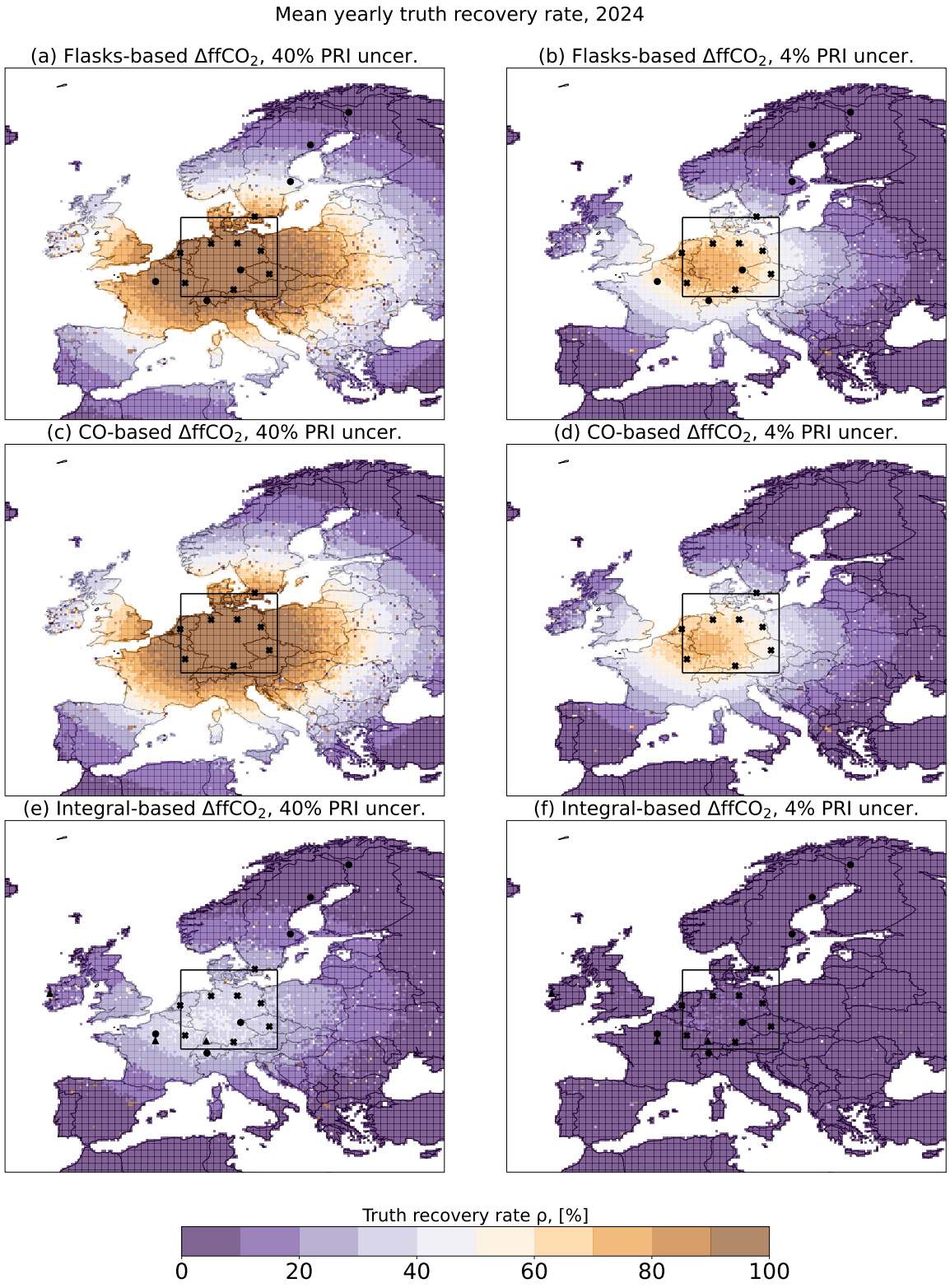

Figure A.10. Mean spacial constrain of the inversion in 2024 for three $\Delta ffCO_2$ proxies investigated in this study: flask-based (subplots a and b), CO-based (subplots c and d), integral-based (subplots e and f). Flat Zero Prior (3.2.4) was used in all runs with different prescribed prior uncertainty (subplots a, c, e - 40%, subplots b, d, f - 4%). The truth recovery rate ρ ranges from 0%, meaning emissions were not change in posterior compared to prior, to 100%, where the truth could be completely recovered. Stations marked with crosses are used in all setups, circles - only in flask- and integral-based runs, and triangles only in integral-based inversions. The region of maximum recovery is marked with black square.

A.2 Additional flask FWD runs

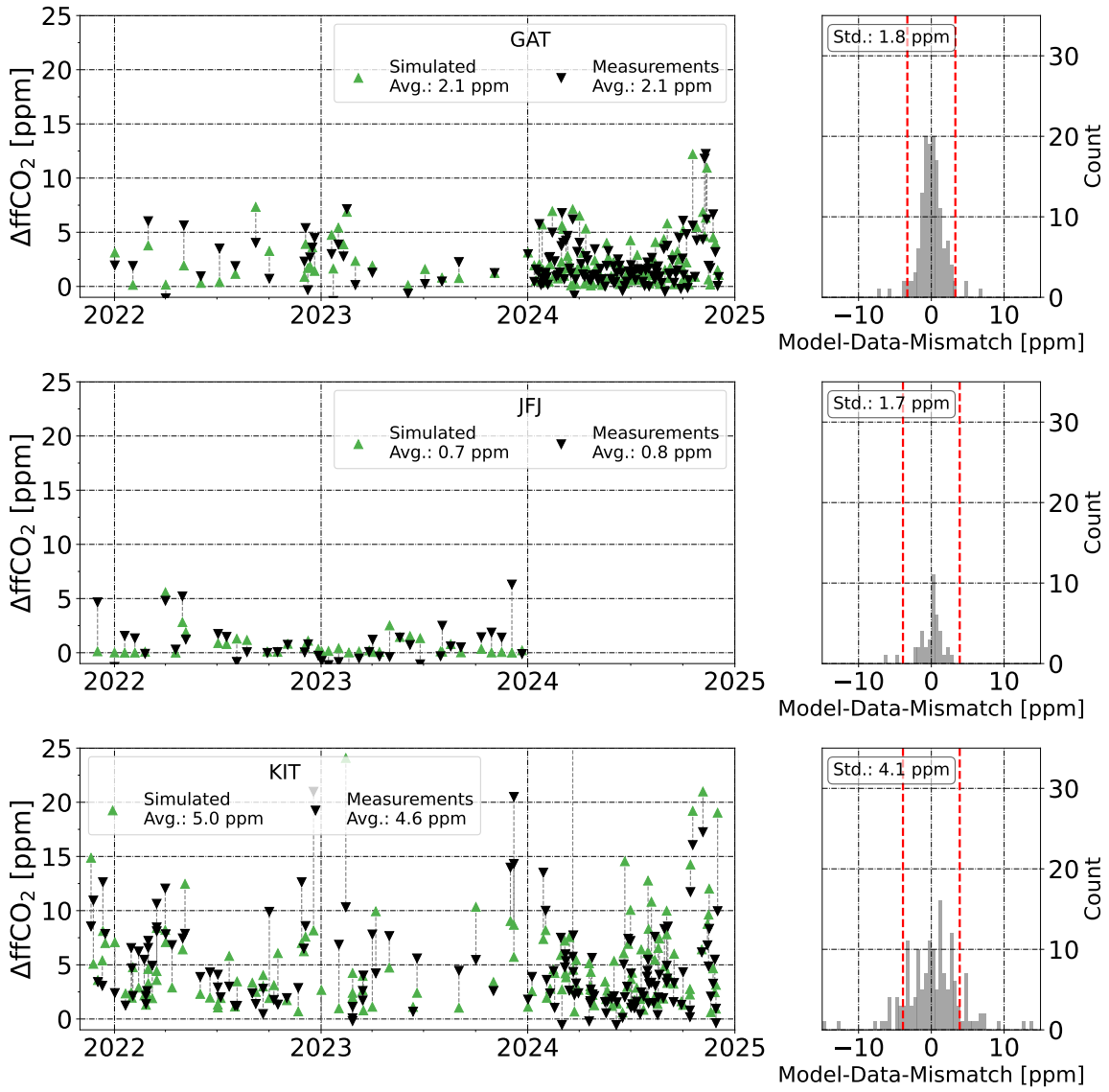

Figure A.11. The flask-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three measurement sites. Corresponding pairs are connected by dashed line. The panels on the right demonstrate model-data-mismatch distribution. The chosen time period has the highest data density (01.11.2021 - 01.01.2025, Sec. 4.3). The signal averages over this period are depicted in the legend.

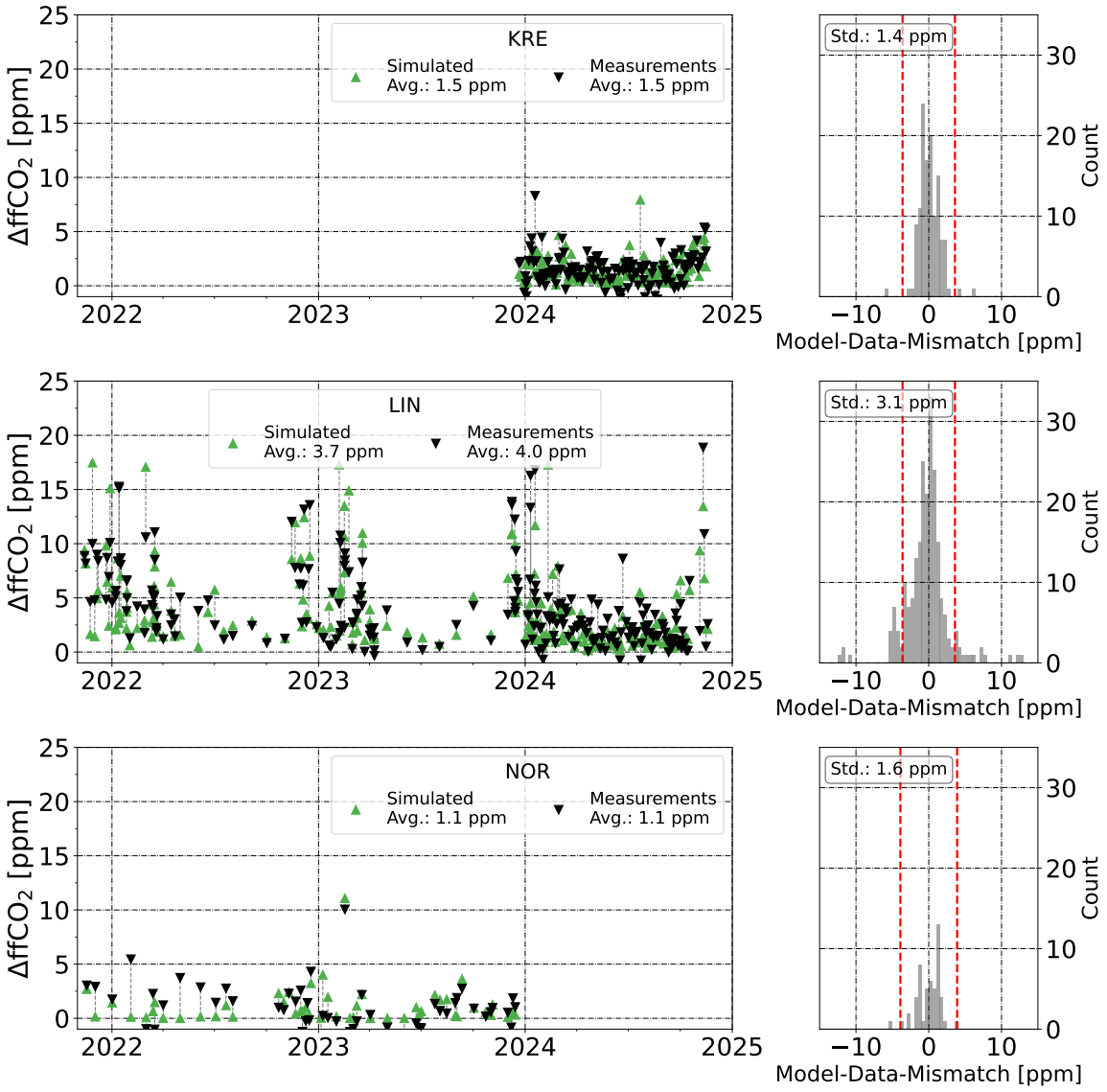

Figure A.12. The flask-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three measurement sites. Corresponding pairs are connected by dashed line. The panels on the right demonstrate model-data-mismatch distribution. The chosen time period has the highest data density (01.11.2021 - 01.01.2025, Sec. 4.3). The signal averages over this period are depicted in the legend.

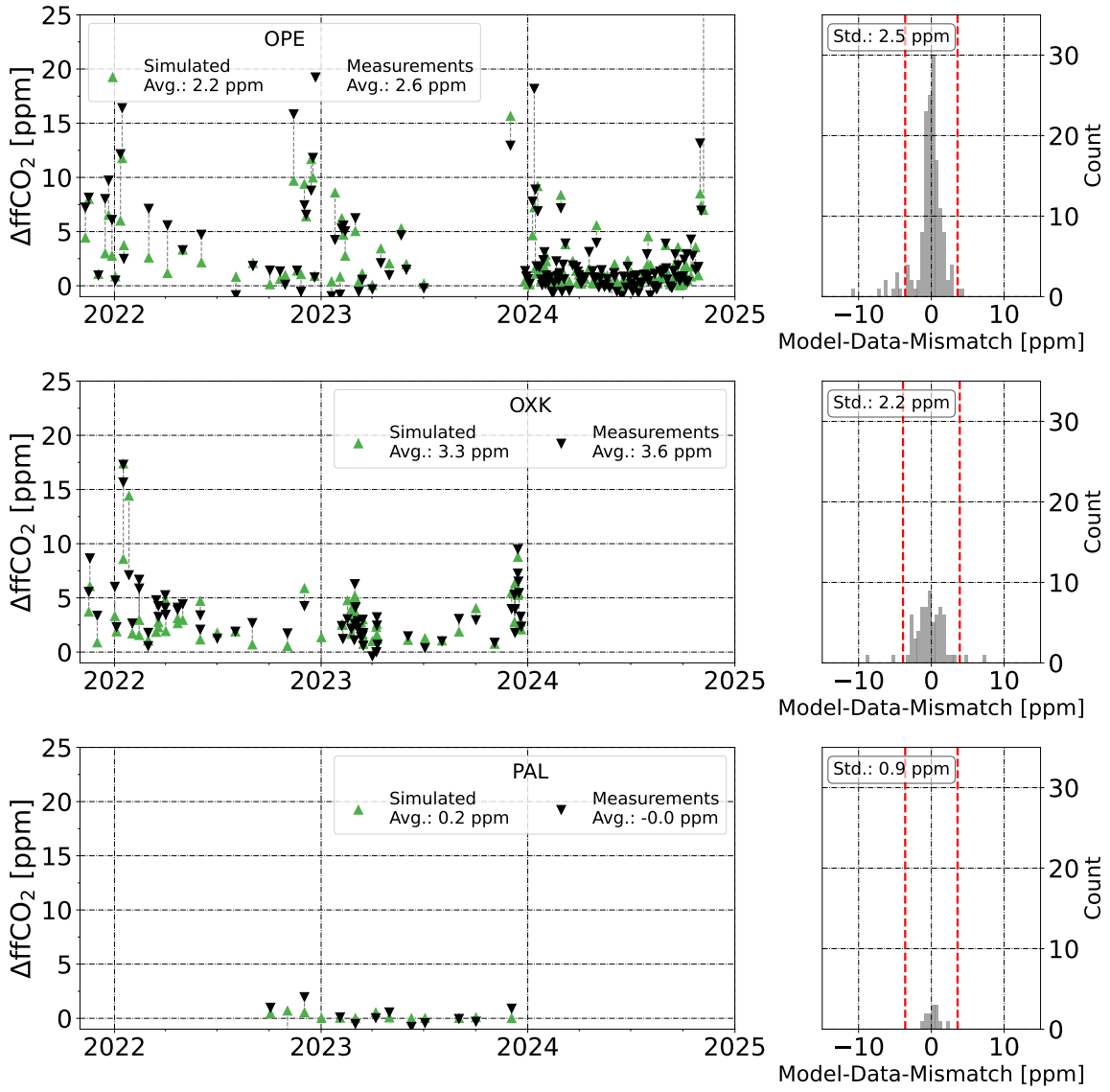

Figure A.13. The flask-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three measurement sites. Corresponding pairs are connected by dashed line. The panels on the right demonstrate model-data-mismatch distribution. The chosen time period has the highest data density (01.11.2021 - 01.01.2025, Sec. 4.3). The signal averages over this period are depicted in the legend.

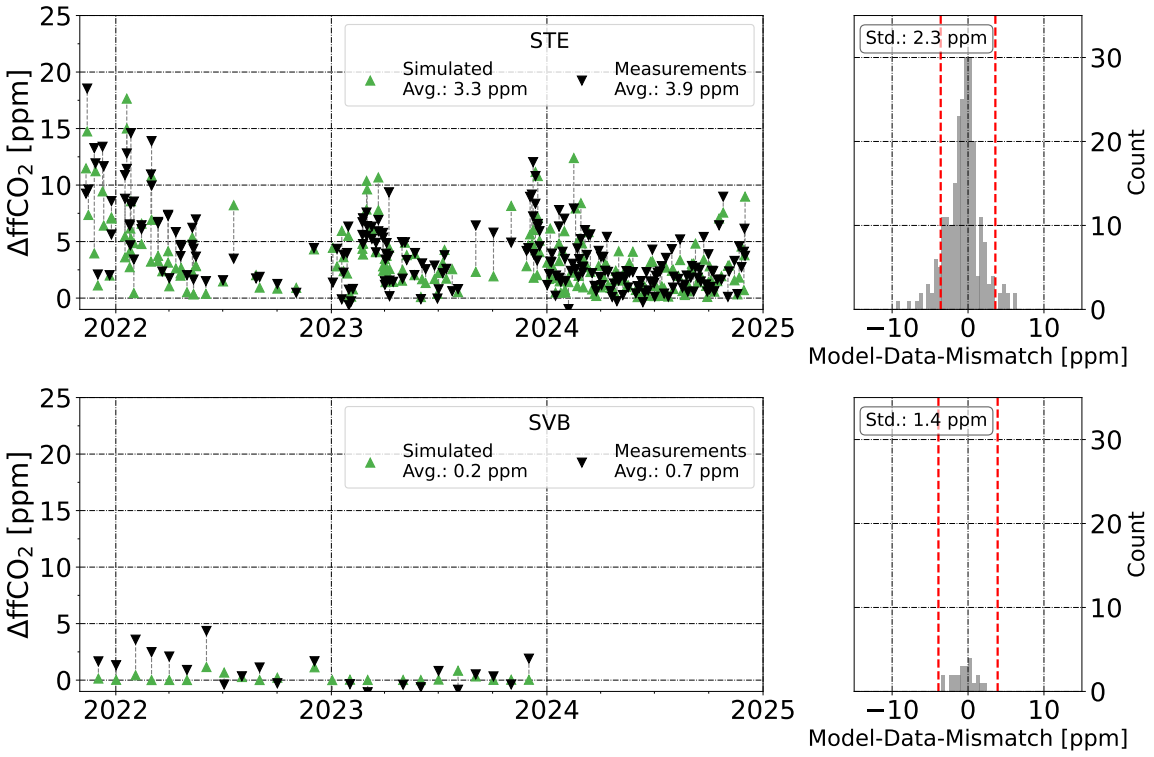

Figure A.14. The flask-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three measurement sites. Corresponding pairs are connected by dashed line. The panels on the right demonstrate model-data-mismatch distribution. The chosen time period has the highest data density (01.11.2021 - 01.01.2025, Sec. 4.3). The signal averages over this period are depicted in the legend.

A.3 Additional ^{14}C calibrated $\Delta CO/\Delta ffCO_2$ ratios

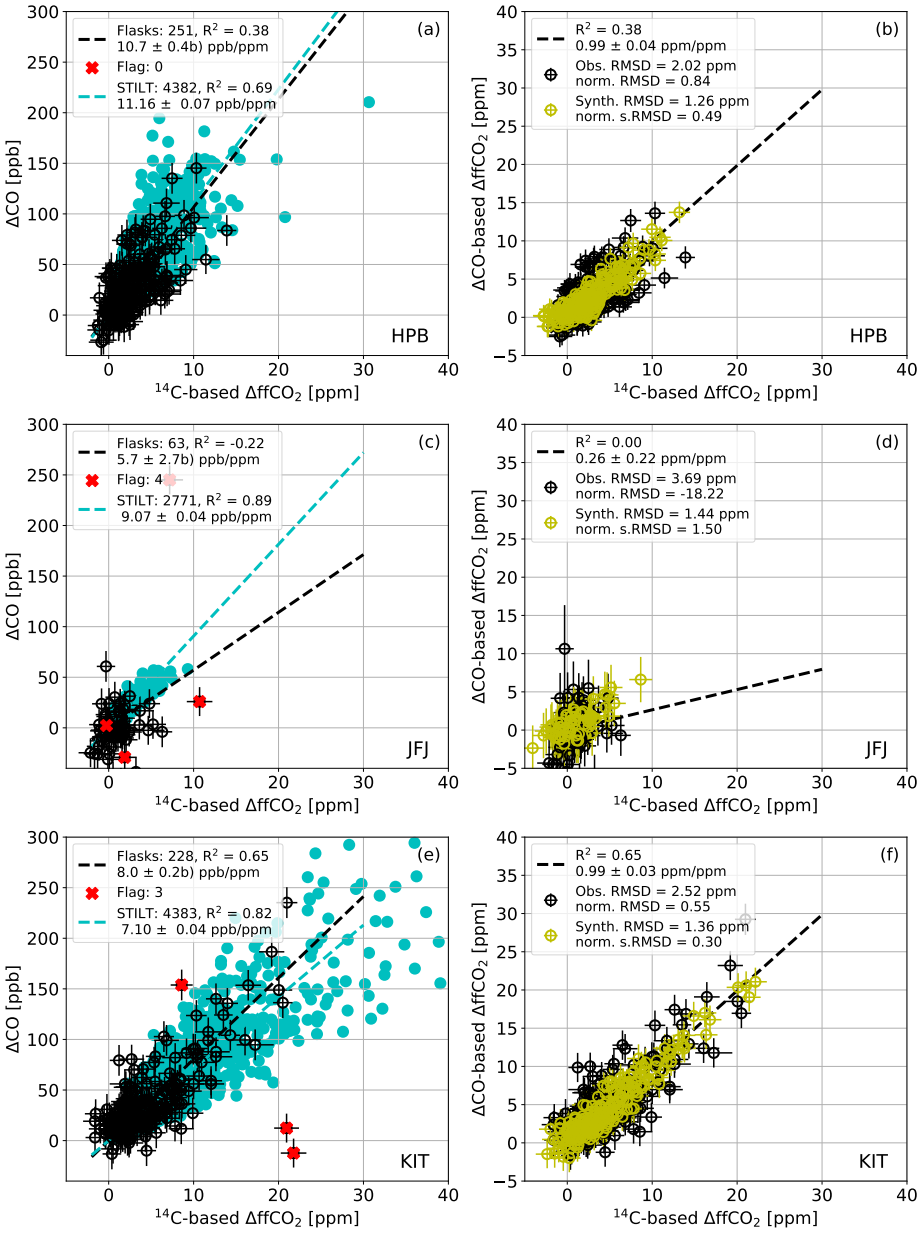

Figure A.15. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

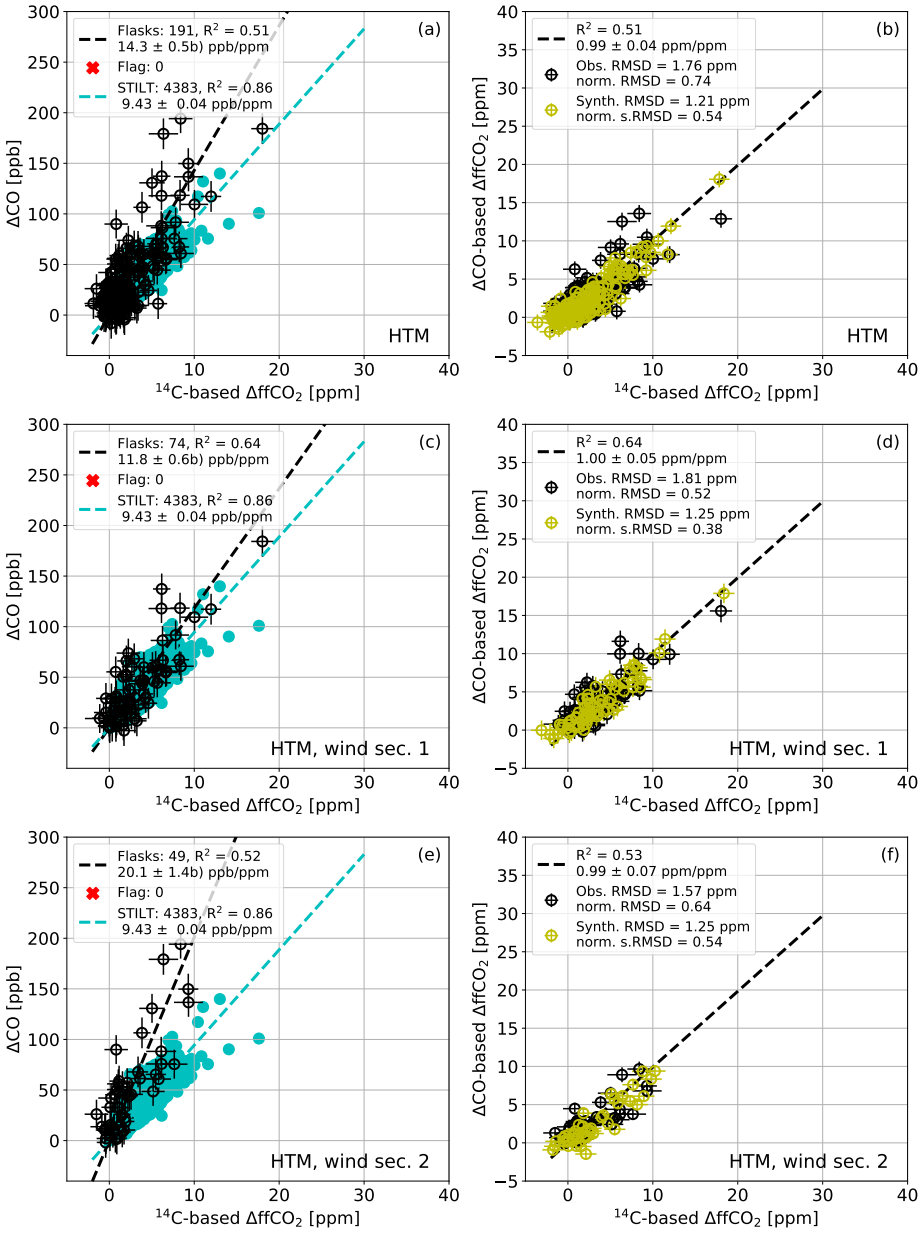

Figure A.16. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

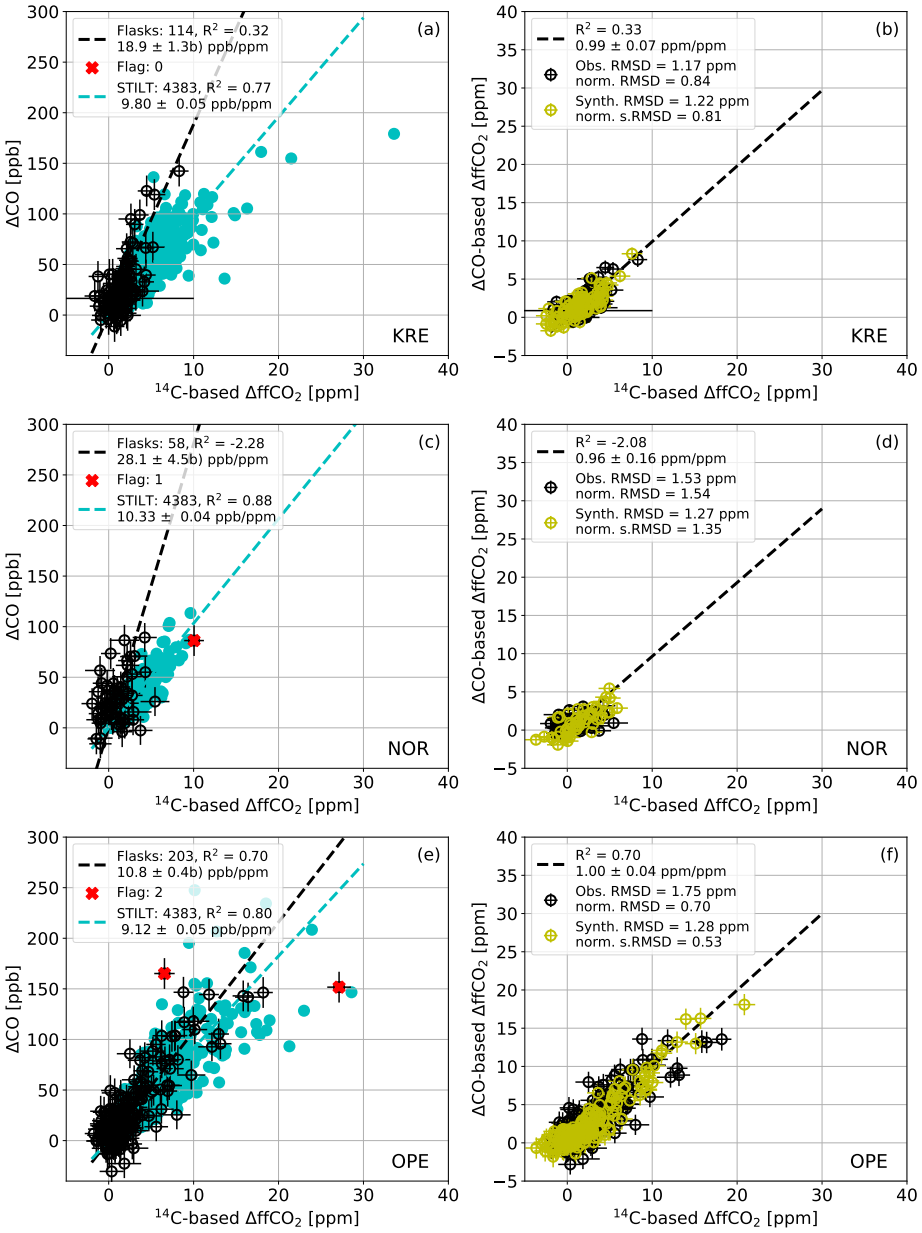

Figure A.17. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

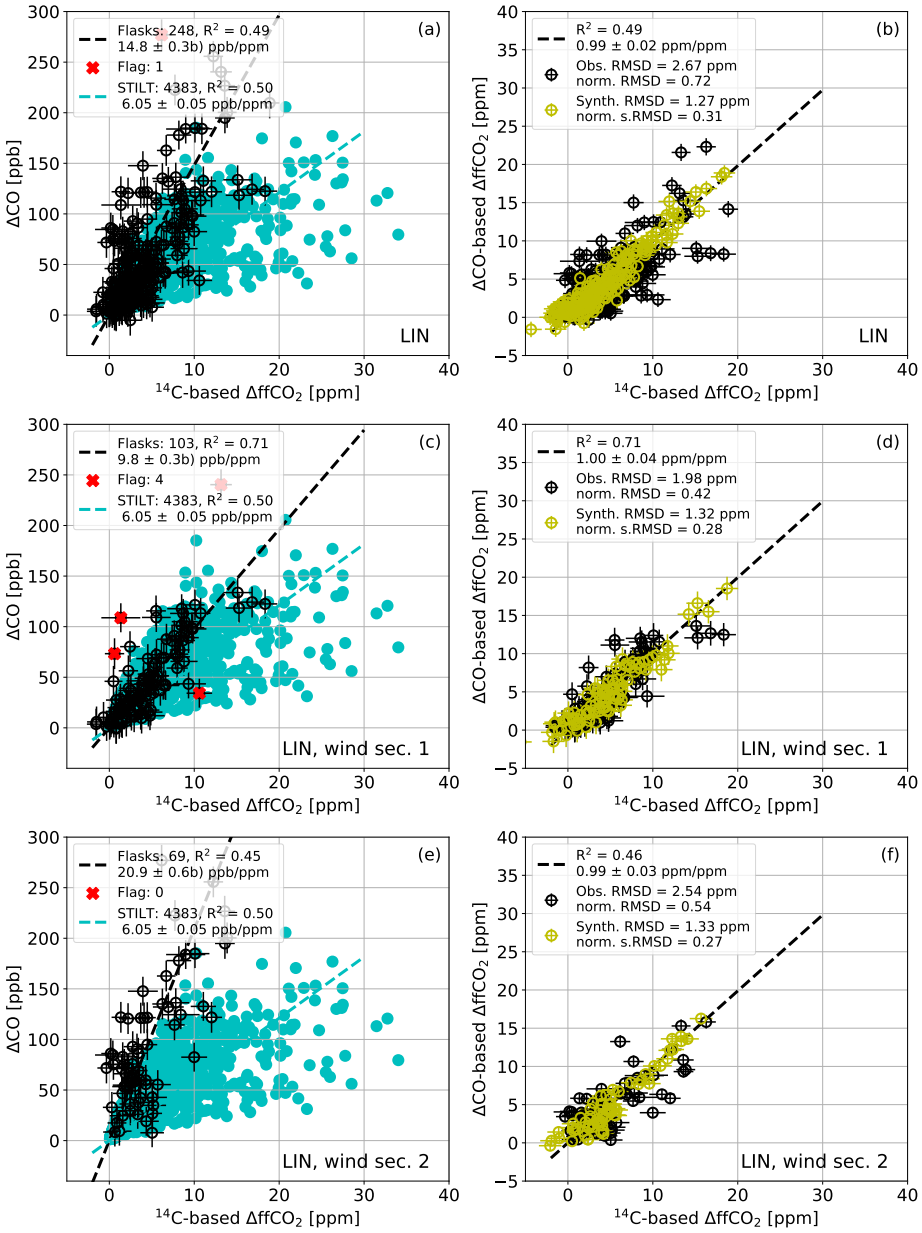

Figure A.18. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

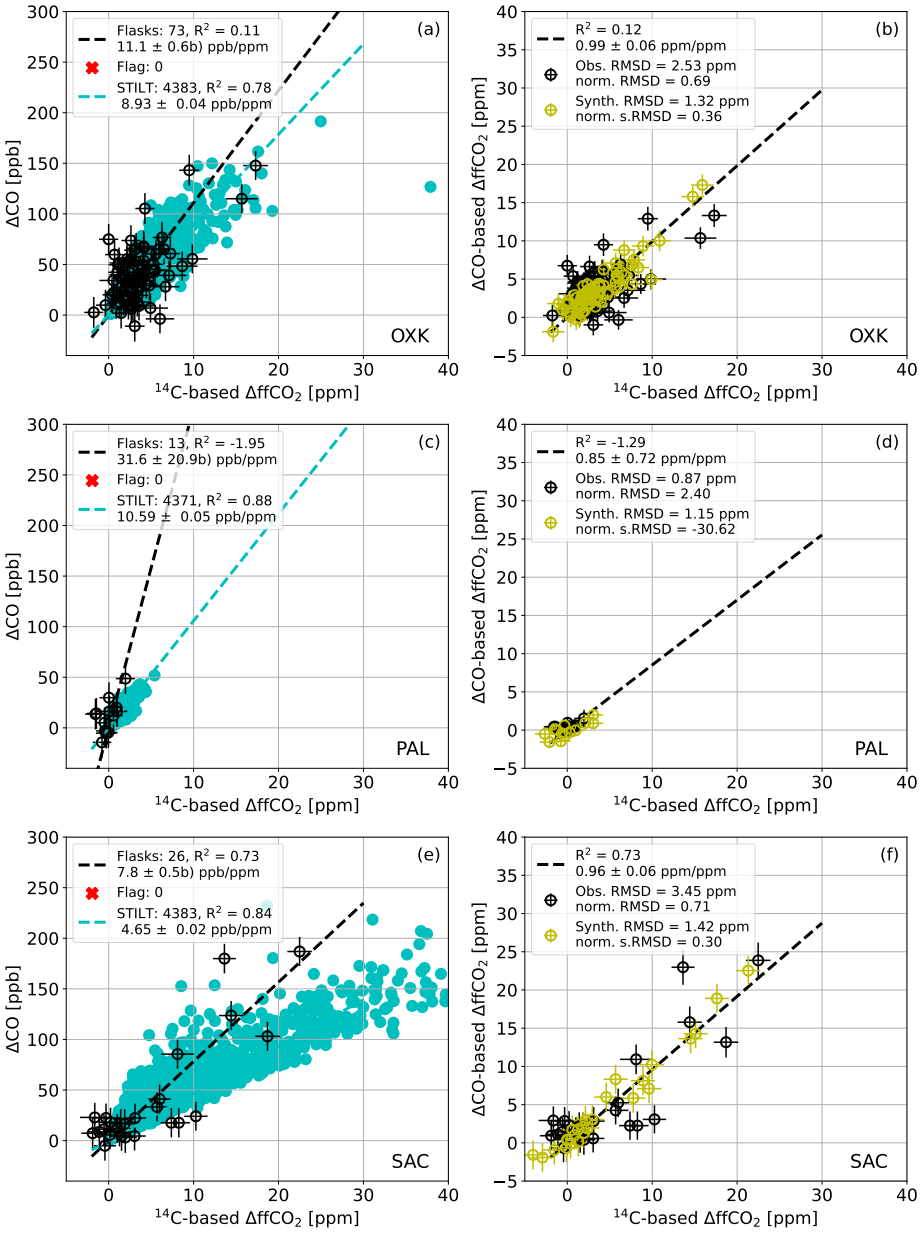

Figure A.19. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

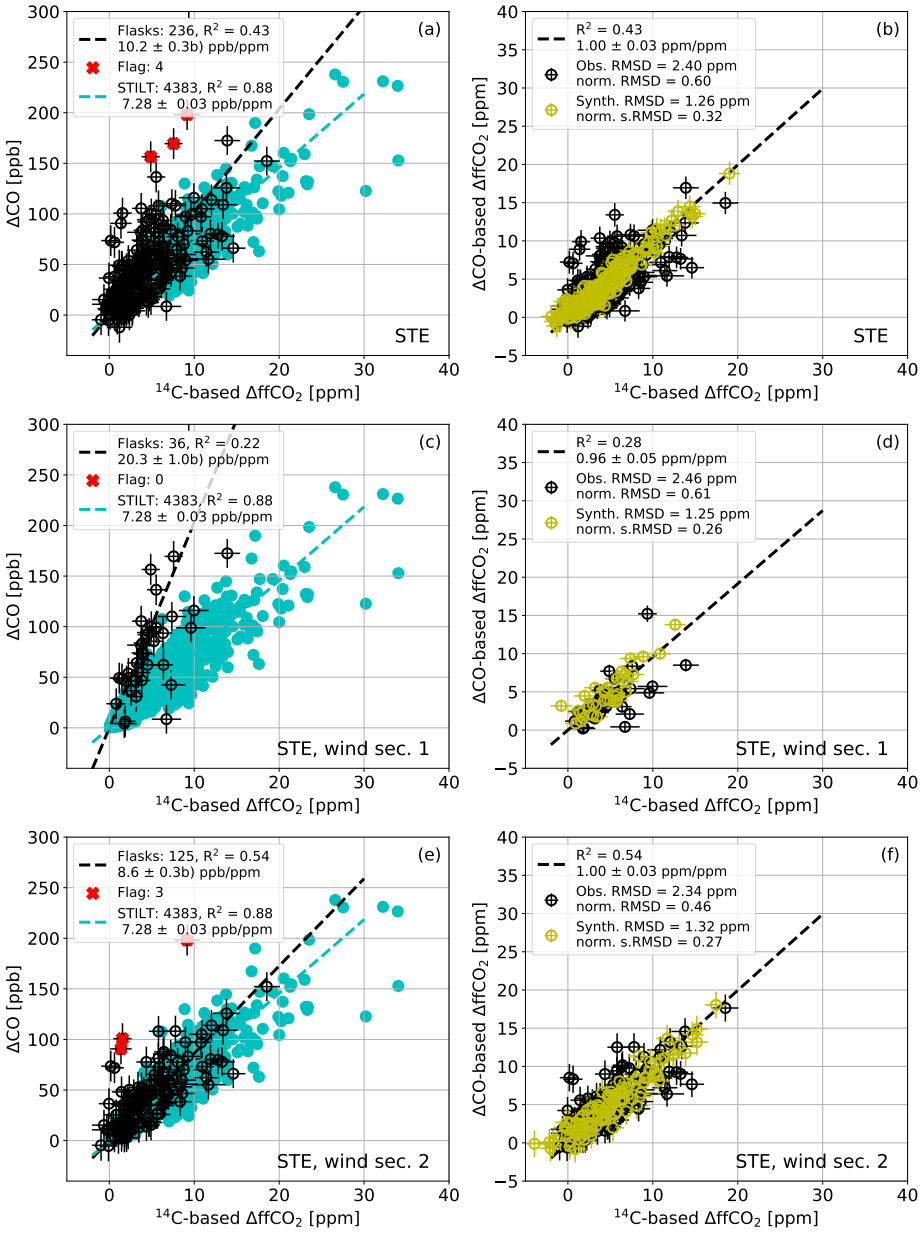

Figure A.20. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panels (a), (c) and (e) display ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panels (b), (d) and (f), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

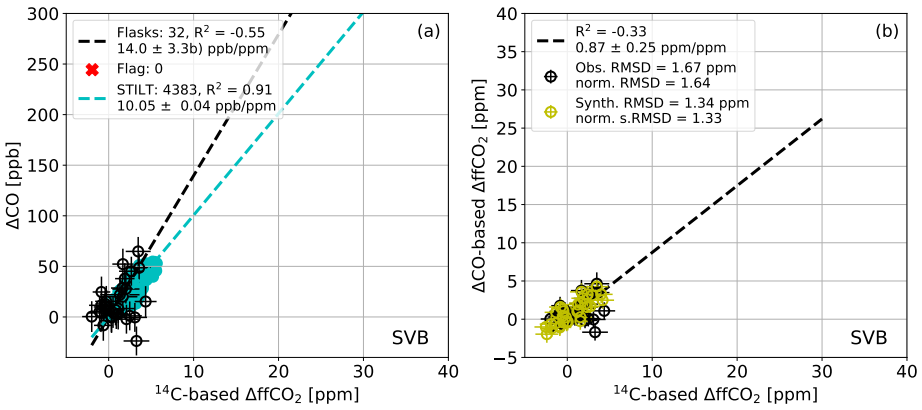

Figure A.21. Calculation of ^{14}C-calibrated CO-based ΔffCO$_2$. In all panels, dashed lines represent linear regressions of the data in corresponding colors (slopes in legend). Panel (a) displays ΔCO and ΔffCO$_2$ values from flasks (black circles with error bars) and based on STILT-TNO modeling (turquoise dots, see Sect. 6.2.4) for station denoted on bottom right. Flagged flasks (red crosses) were excluded from the regression. In panel (b), ΔCO^{flask} was recalculated into ΔffCO$_2$ (Eq. 6.1) using $\langle R_{flasks} \rangle$. RMSD of the corresponding data represents variability caused either by measurement errors alone (synthetic data, yellow circles with error bars, see Sect. 6.2.3) or by both measurement errors and ratio variability in the station's catchment area (real data, Obs. RMSD).

A.4 Wind dependent ΔCO to ΔffCO$_2$ ratios

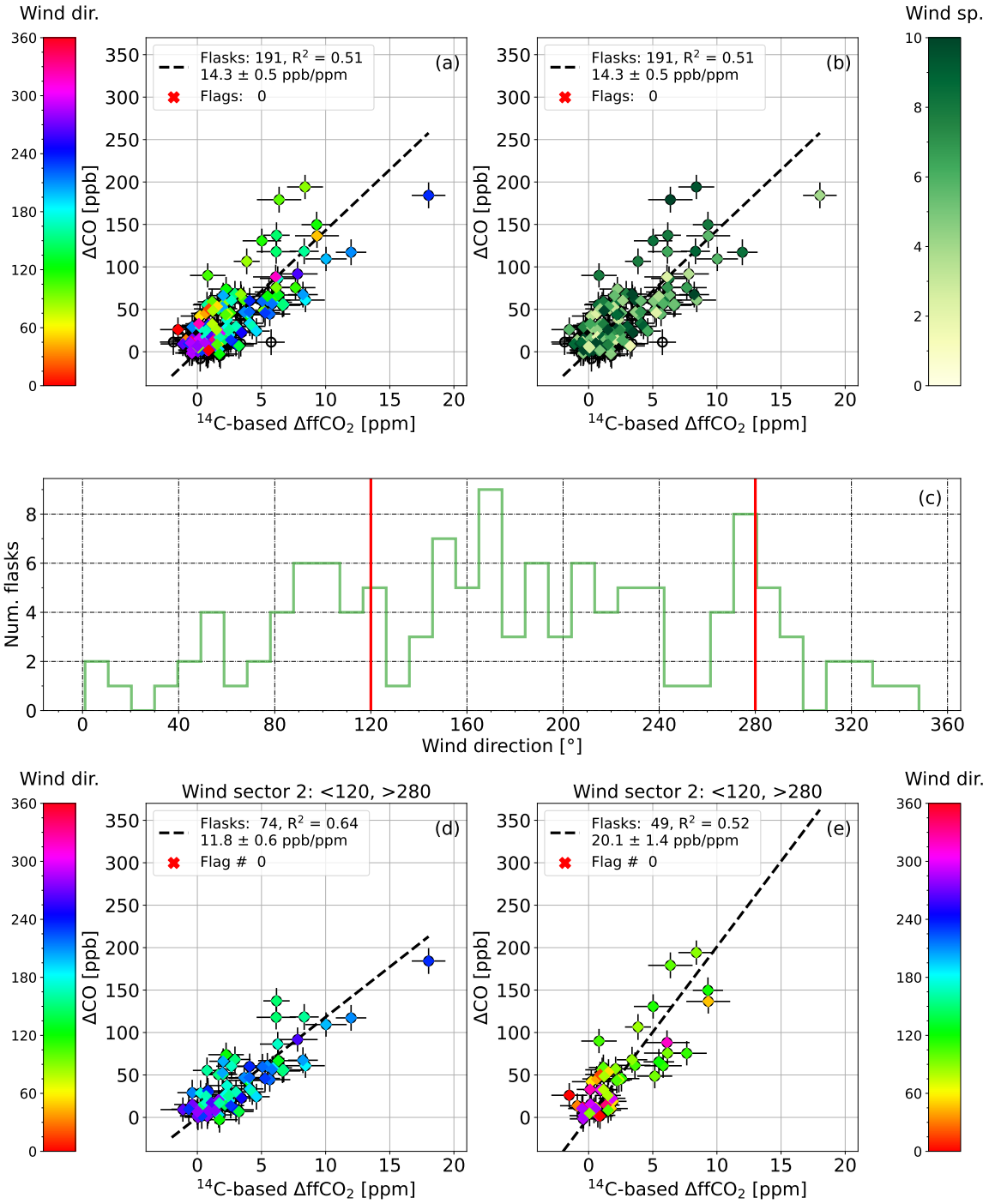

Figure A.22. Selection of appropriate wind sectors for calculating wind-dependent ΔCO to ΔffCO_2 ratios at HTM station. Panels (a) and (b) display the complete flask dataset, while panels (d) and (e) show the dataset split by mean wind direction during sampling (excluding flasks with missing meteorological data), along with linear regressions (black dashed line, slope in legend) for corresponding flask pools. Color gradients in panels (a), (d), and (e) indicate mean wind direction, and in panel (b), mean wind speed during flask sampling. Panel (c) illustrates the distribution of wind conditions during flask sampling. Red lines denote the boundaries of the wind sectors used to divide the flask dataset into two groups.

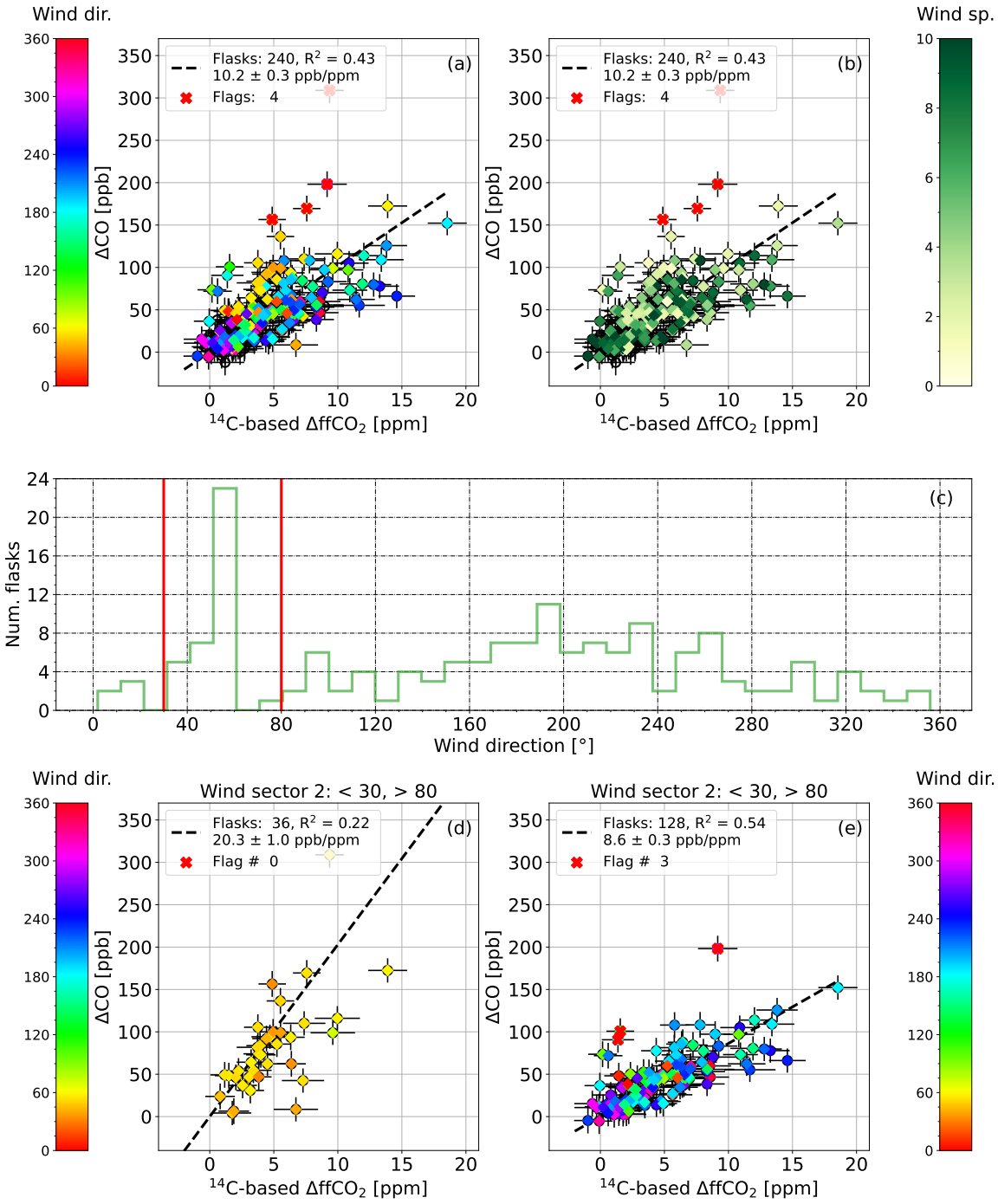

Figure A.23. Selection of appropriate wind sectors for calculating wind-dependent ΔCO to ΔffCO$_2$ ratios at STE station. Panels (a) and (b) display the complete flask dataset, while panels (d) and (e) show the dataset split by mean wind direction during sampling (excluding flasks with missing meteorological data), along with linear regressions (black dashed line, slope in legend) for corresponding flask pools. Color gradients in panels (a), (d), and (e) indicate mean wind direction, and in panel (b), mean wind speed during flask sampling. Panel (c) illustrates the distribution of wind conditions during flask sampling. Red lines denote the boundaries of the wind sectors used to divide the flask dataset into two groups.

A.5 Additional CO-based FWD runs

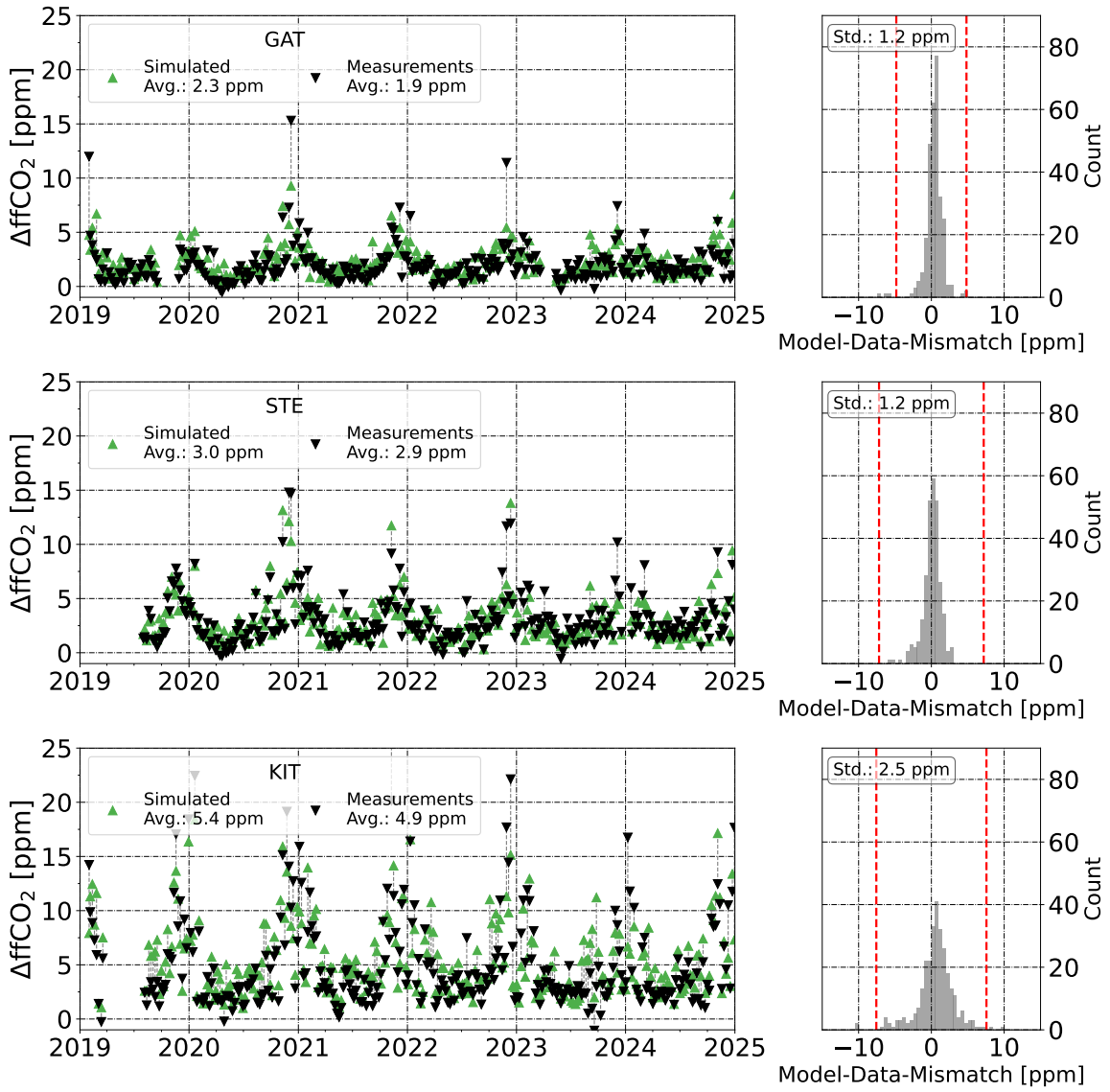

Figure A.24. The CO-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the RMSD-based measurement error.

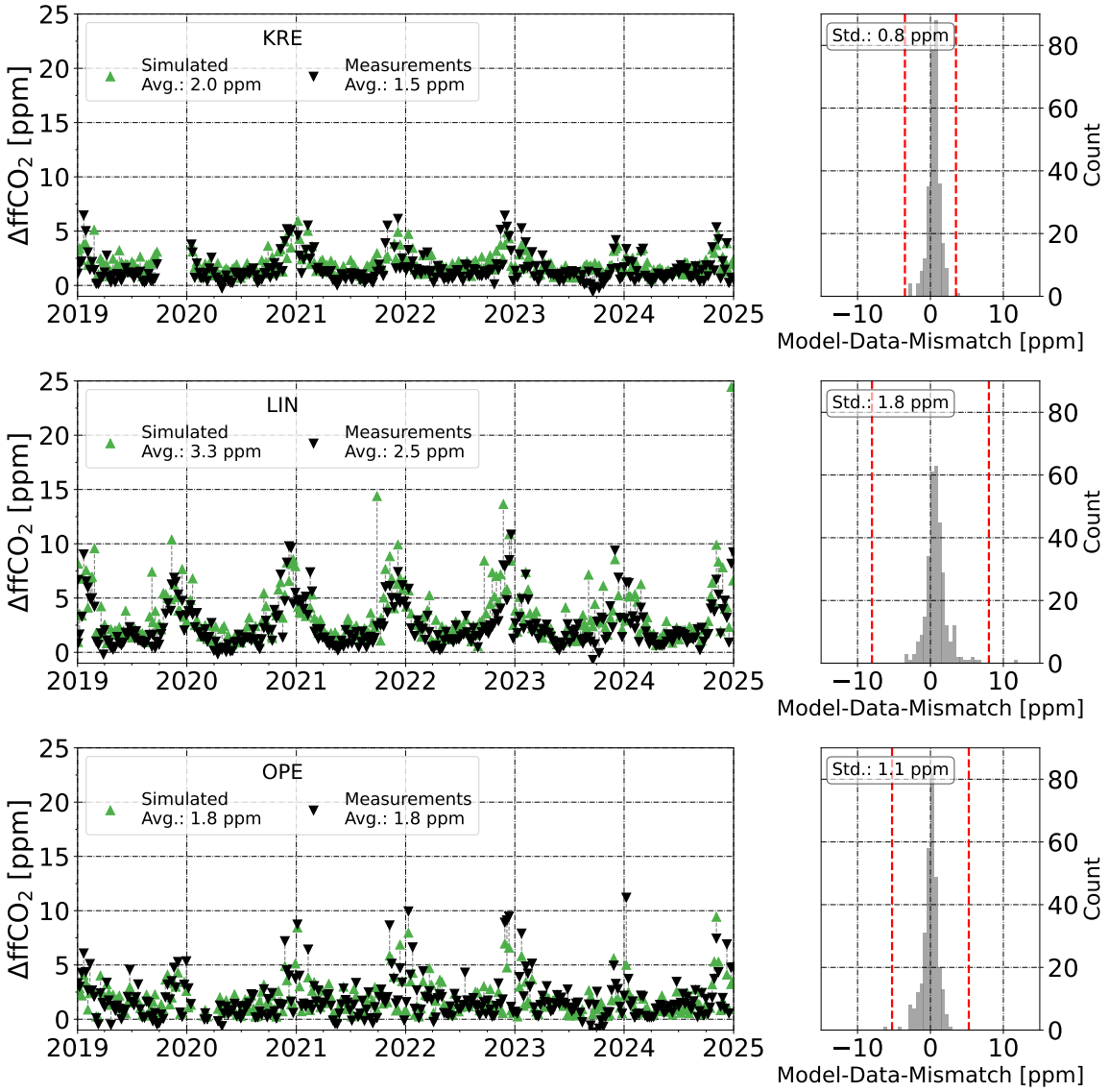

Figure A.25. The CO-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the RMSD-based measurement error.

A.6 Additional integral-based FWD runs

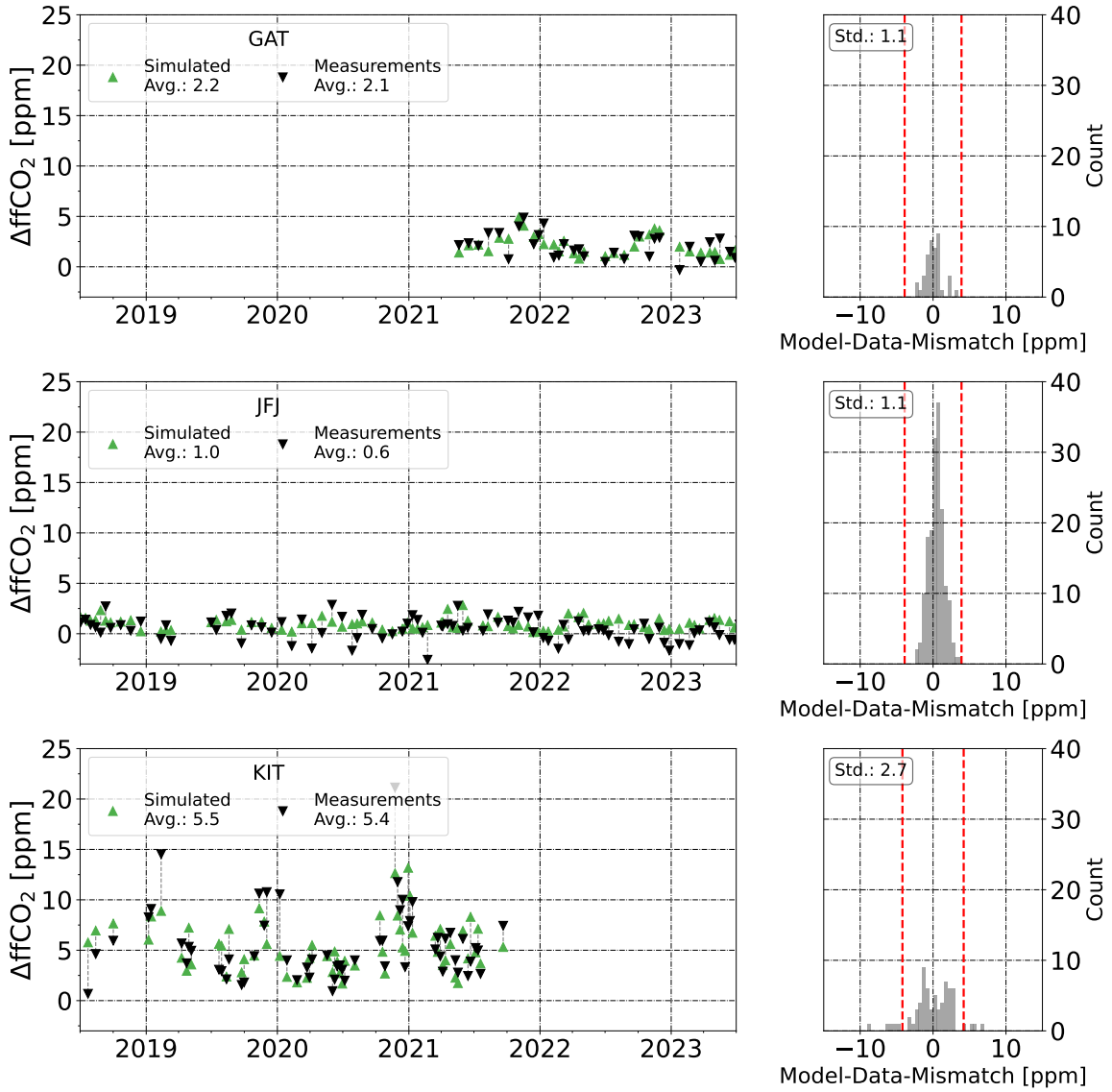

Figure A.26. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sec. 4.3).

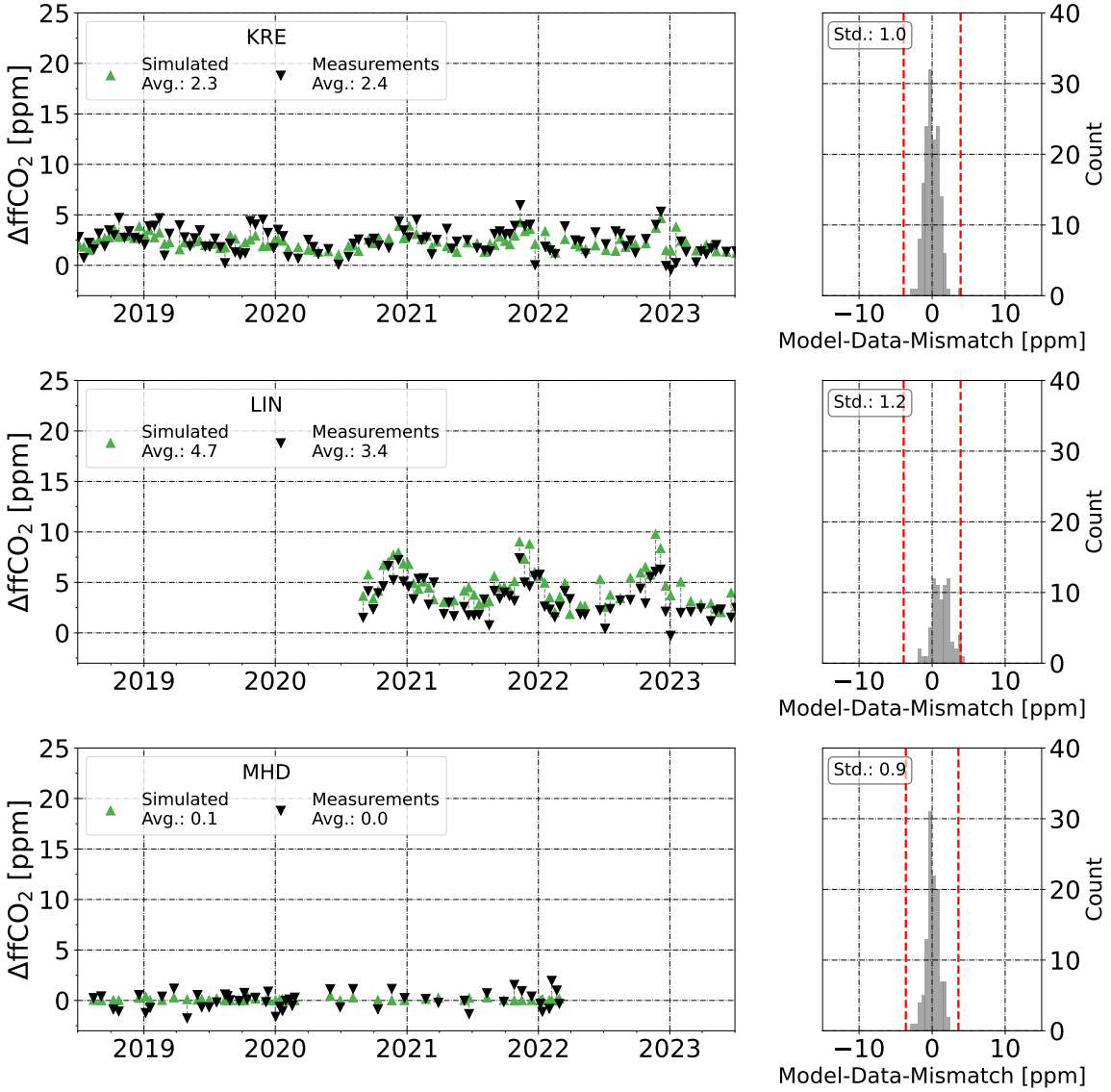

Figure A.27. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sec. 4.3).

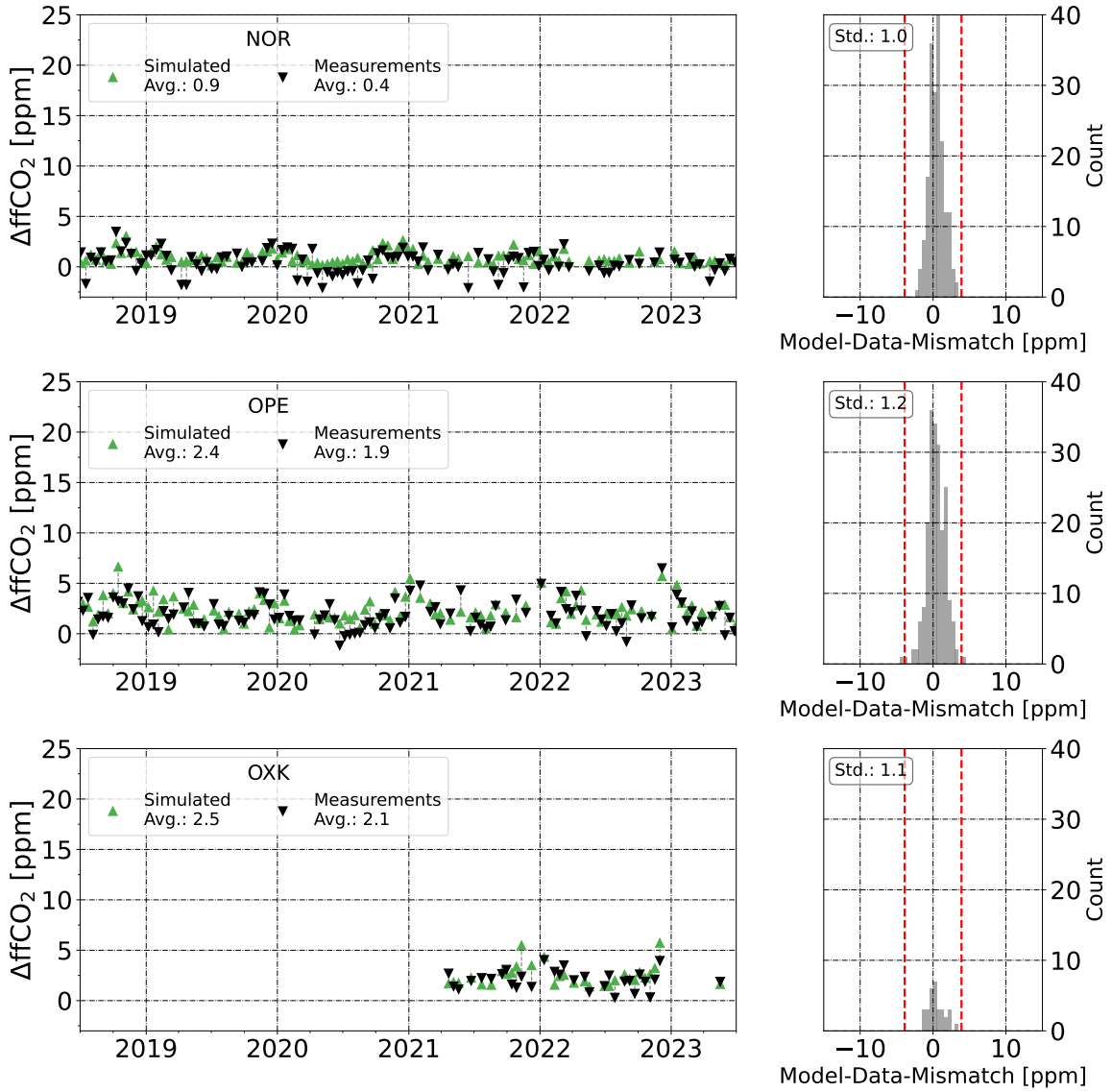

Figure A.28. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sec. 4.3).

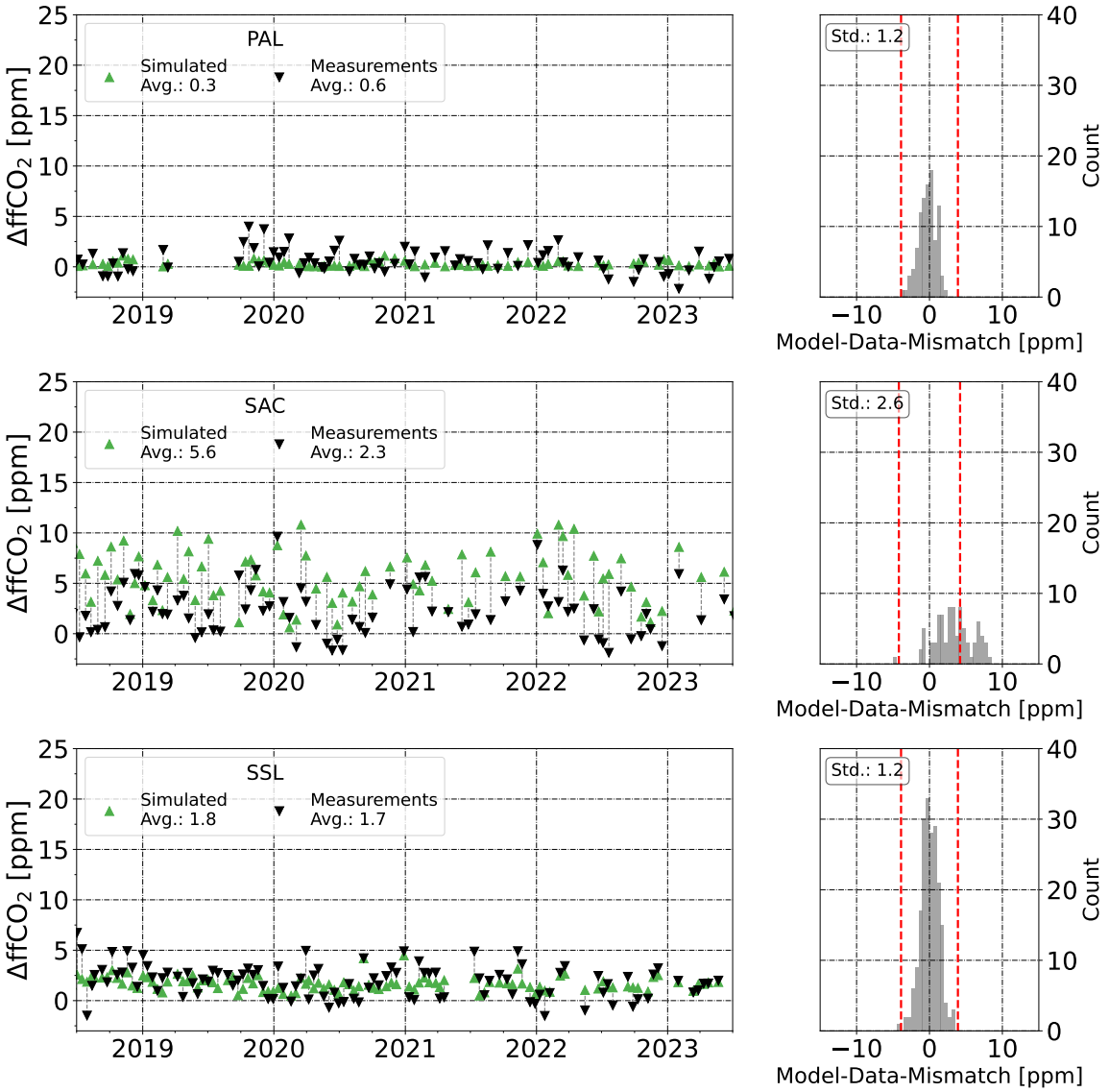

Figure A.29. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sec. 4.3).

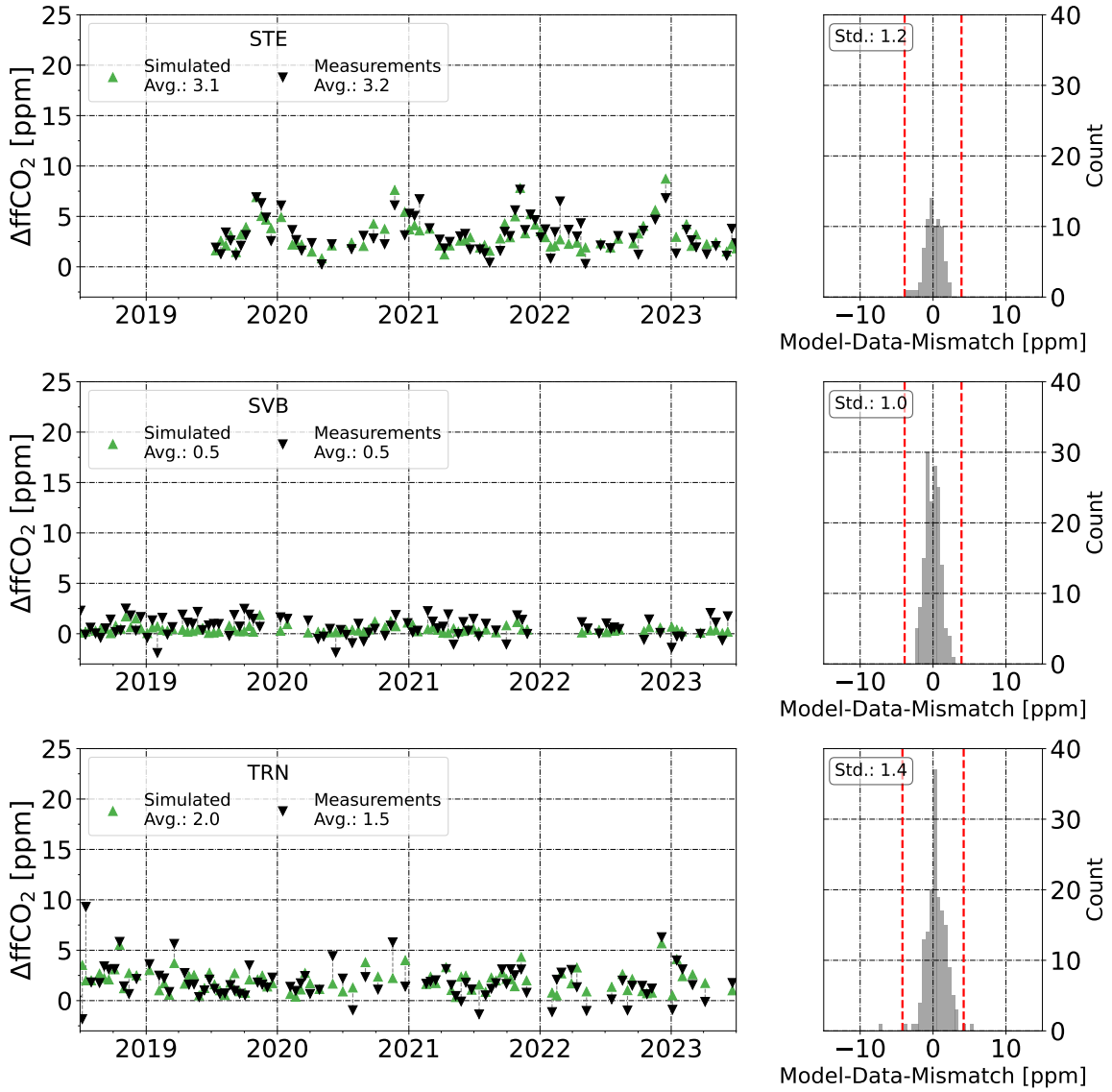

Figure A.30. The integral-based $\Delta ffCO_2$ observations (black inverse triangles) and simulated $\Delta ffCO_2$ values based on GCP inventory (green triangles) for three example measurement sites. Corresponding pairs are connected by dashed line. The average signal over the depicted time period are in the legend. The panels on the right demonstrate model-data-mismatch distribution including it variance with red dashed lines showing the triple of the mean measurement error. The chosen time period has the highest data density (01.07.2018 - 01.07.2023, Sec. 4.3).

Acknowledgments

Even though only my name appears at the front, work like this would never be possible without many helpful hands. I would like to take this opportunity to thank a few people who have made a particularly significant contribution to make this thesis happen.

Sam Hammer's contribution to my work was probably the most significant, especially following Ingeborg's untimely departure. You took the role of scientific advisor very earnestly, becoming my guiding star and role model in the scientific world. There were many phases during these years when I actually lost hope of getting meaningful results. But you always encouraged me to persevere and to look at problems from different angles, and you never let me despair. Thanks to your guidance, patience and all our discussions, my thesis changed a lot for the better during the last year. Your support during the final phase of my thesis saved me from many stressful nights and helped me produce the best thesis possible. Thank you.

None of this would have been possible without Ingeborg Levin. You welcomed me to the group and entrusted me with your passion project. You were always 'hart aber fair', never easy to please, but I think you would like this final result, which grew out of your tenacity with your students and your measurements. This is for you.

Christian Rödenbeck, you introduced me to the fascinating and complicated world of inverse modelling. There were many moments when the results made no sense to me, but you could calmly explain why they did. You always managed to find the positives in my results, even when others couldn't. And, of course, you always protected my code from the hungry penguins!

Thank you to the best research group. Wir sind klein aber fein! Susi, it was only thanks to your eagle eyes and reliable data management that I had such high-quality input for my models. Timo, thanks to your 'fancygste' work, I now have much more trust in the $\Delta ffCO_2$ data, even though it cost you a place on all the FBI's suspect lists. Ann-Kristin, seeing how you tackle your PhD challenges and overcome them (oh, that Paris Flask sampler...) gave me the push I needed to never give up. Fabian, even though you're no longer officially part of the group, I always felt I could ask you for advice and receive the most comprehensive answer imaginable. Hannes, the tea and the advice were very useful; but not as useful as the $\Delta ffCO_2$ tool, or my privilege of asking for any weird version of the data and actually receiving it.

Julian, you're the person who really understands me (except for my Senf preferences) – whether it's deep asteroid mines, Valhalla or kimchi, the chemistry just works! I had the best time at the institute after I moved into your office. Thanks.

Huge thanks to all the CRL laboratory technicians. You have poured all your hard work into the data, allowing people like me to use it without having to think about measurements. It's a first-class service.

I would like to thank my two reviewers, Prof. Dr. Werner Aeschbach and Prof. Dr. André Butz for undertaking this challenging yet essential task. I would also like to thank the other members of my committee, Prof. Dr. Ulrich Schwarz and Dr. Kai Schweda, for supervising my PhD defense.

From the beginning to the end, all of my wonderful, big, extended family supported me. From my wife's family to my brothers and sisters, and my grandparents — even if not all of you are with us anymore. Mama and Papa, it was thanks to you that I was able to go to Germany and receive such education at one of the world's best universities. You always wanted me to have a better and easier life than you did. You believed in me and trusted me. After all these years, I can say with confidence that it was all worth it. You made me who I am today.

And, of course, I am endlessly thankful to the love of my life, my favorite cat and my wonderful wife, Dasha. This time was hard for many different reasons: from geopolitical to deeply personal. You were always there. You supported me every step of the way. In my darkest moments and my brightest, you were there. And it goes both ways. Always. Until the end.

Declaration on the use of generative AI

This thesis was written with the assistance of several artificial intelligence (AI) tools. My use of these tools was limited to a supportive role, with all scientific content, data analysis, and conclusions remaining my original work.

The following AI tools were used:

- **Gemini:** Used for a broad range of tasks throughout the entire thesis, including:
  - Text Optimization and Proofreading: To refine text for clarity, conciseness, and scientific style, and to correct grammatical errors.
  - Coding Assistance: To generate snippets of Python code, which were then adapted and integrated into my own scripts for data analysis and image generation.
  - Formatting: To assist in formatting of Latex tables and other structured text.
- **DeepL:** Used to translate the abstract of this thesis into German.
- **Overleaf Writeful:** Used to receive stylistic suggestions and improve the flow of sentences during the writing process.

All text was reviewed, edited, and approved by me to ensure its accuracy, originality, and alignment with the research. Factual information, research findings, and all academic citations were sourced and verified by me from peer-reviewed literature. The use of these tools was limited to a role similar to that of a stylistic and grammatical editor, under my direct supervision.