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

Lukas Nicolai Pilz

born in Berlin

Oral examination: 14.01.2026

Monitoring urban greenhouse gas emissions using high-resolution models

by

LUKAS NICOLAI PILZ

Referees: Prof. Dr. André Butz
Prof. Dr. Werner Aeschbach

Abstract

Urban regions emit large amounts of CO_2, are drivers of national inventory uncertainty, and have a direct impact on the day-to-day life of their inhabitants. Therefore, an independent assessment of urban emissions supports climate action and governance. This work presents a complete pipeline for improving urban emissions estimates. It includes setting up, optimizing, and evaluating atmospheric transport as well as CO_2 and CO simulations and estimating emissions based on aircraft measurements.

We optimize the physics configuration of the WRF atmospheric transport model and show that its performance exceeds previous studies. We use this optimized configuration to generate MACRO-2018, a 1 km resolved dataset of meteorology as well as CO_2 and CO concentrations in German metropolitan areas for 2018. This dataset provides two different physics configurations and biogenic models. We evaluate its performance against high-precision CO_2 and CO measurements and find mean absolute biases of 5.2 ppm to 6.4 ppm for CO_2, exceeding the quality of comparable datasets. For CO, we find mean absolute biases of 25.4 ppb to 28.6 ppb. Finally, based on this dataset we set up a Bayesian inversion framework which assimilates aircraft measurements to optimize the emissions of Berlin and the background CO_2 concentration simultaneously. Using this approach, we reduce the emissions estimate uncertainty from a simpler and widely-used mass balance approach by a factor of five.

Zusammenfassung

Städtische Regionen emittieren große Mengen an CO_2, sind verantwortlich für erhebliche Unsicherheiten in nationalen Inventaren und haben einen direkten Einfluss auf das tägliche Leben ihrer Einwohner. Daher unterstützt die Verbesserung der Schätzungen von städtischen Emissionen globalen Klimaschutz und lokale Governance. Diese Arbeit präsentiert eine komplette Pipeline zur Verbesserung von urbanen Emissionsschätzungen. Sie umfasst die Einrichtung, Optimierung und Evaluation von atmosphärischen Transport-, CO_2- und CO-Simulationen und die Emissionsschätzung basierend auf Flugzeugmessungen.

Wir optimieren die physikalische Konfiguration des atmosphärischen Transportmodells WRF und zeigen, dass die Leistung frühere Studien übertrifft. Basierend auf dieser optimierten Konfiguration generieren wir MACRO-2018, einen Datensatz mit einer Auflösung von 1 km bestehend aus Meteorologie sowie CO_2- und CO-Konzentrationsfeldern in deutschen Ballungsräumen für das Jahr 2018. Dieser Datensatz beinhaltet zwei verschiedene physikalische Konfigurationen und zwei biogene Modelle. Wir bewerten die Leistung dieses Datensatzes anhand von hochpräzisen CO_2- und CO-Messungen und finden einen mittleren Betragsfehler von 5.2 ppm bis 6.4 ppm, was die Leistung vergleichbarer Datensätze übertrifft. Für CO finden wir einen mittleren Betragsfehler von 25.4 ppb bis 28.6 ppb. Schlussendlich setzen wir ein bayessches Inversionsframework basierend auf diesen Daten auf, welches Flugzeugmessungen assimiliert um gleichzeitig die Emissionen von Berlin und die Hintergrund-CO_2-Konzentration zu optimieren. Mit diesem Ansatz reduzieren wir die Unsicherheit der Emissionsschätzung gegenüber einem einfacheren, aber weit verbreiteten Massenbilanzansatz um den Faktor fünf.

Disclaimer

Parts of the content of this thesis have been published and publicly released as detailed below:

Pilz, L., C. Lüken-Winkels, M. Gałkowski, D. Ho, C. Gerbig, F. Chen, and S. N. Vardag (2026). "Evaluation of high-resolution WRF simulation in urban areas — Effect of different physics schemes on simulation performance in the Rhine-Main-Neckar area". *Atmospheric Research* 328, p. 108435. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2025.108435.

Author Contributions: ***Lukas Pilz:*** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. *Christopher Lüken-Winkels:* Writing – review & editing, Investigation. *Michał Gałkowski:* Writing – review & editing, Software, Methodology, Data curation. *David Ho:* Writing – review & editing, Software. *Christoph Gerbig:* Writing – review & editing, Resources, Data curation. *Fei Chen:* Supervision, Investigation. *Sanam N. Vardag:* Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization

Pilz, L., H. Denier van der Gon, T. Glauch, J. Marshall, and S. N. Vardag (2025a). *MACRO-2018 - High-Resolution Simulation of CO and CO2 concentrations over German Metropolitan Areas for 2018 using WRF-Chem*. Ed. by **L. Pilz**. World Data Center for Climate (WDCC) at DKRZ. DOI: 10.26050/WDCC/MACRO-2018.

Author Contributions: ***Lukas Pilz:*** Conceptualization, Methodology, Software, Simulation, Validation, Investigation, Formal analysis, Data curation. *Hugo Denier van der Gon:* Provision of emissions data. *Theo Glauch:* Provision of emissions data. *Julia Marshall:* Provision of emissions data. *Sanam N. Vardag:* Supervision, Resources, Project administration, Funding acquisition

Contents

1 Introduction

Climate change is one of the most serious and urgent threats to the stability of both natural and human systems alike (Ripple et al., 2025). Its observable and ever-growing impacts span nearly all components of the Earth's climate system: the atmosphere, ocean, cryosphere, and biosphere. Global average land-surface temperatures have increased by 1.5 °C to 1.6 °C over pre-industrial times (1850 to 1900; Kennedy, 2024). Losses and damages related to climate change are manifesting and will continue to intensify (IPCC, 2022a). These changes are already apparent in climate-dependent sectors of human society like agriculture or fishery (Blasiak et al., 2017; Jägermeyr et al., 2021). Long-term, adverse impacts of climate change on human civilization outweigh beneficial ones (Carleton et al., 2022; IPCC, 2023). Impacts are unevenly distributed, manifesting for example as intensified flooding in water-rich regions and increased drought frequency in arid zones, which threatens livelihoods and infrastructure, particularly in vulnerable areas (IPCC, 2022b). Projections suggest that without substantial mitigation, large populations may be forced out of their climate niche causing large-scale displacement and making conflicts more likely (Xu et al., 2020; Lenton et al., 2023).

Anthropogenic emissions of greenhouse gases (GHGs) are the main cause of climate change (IPCC, 2023). Their impact is measured in global warming potential (GWP) which quantifies the additional radiative forcing caused by these gases over a fixed amount of years. The most commonly used metric is the 100-year GWP (GWP_{100}). Human GHG emissions consist of carbon dioxide (CO_2, 75 % of GWP_{100}), methane (CH_4, 18 % of GWP_{100}), nitrous oxide (N_2O, 4 % of GWP_{100}), and fluorinated gases (2 % of GWP_{100}; Fig. 2.5 in IPCC, 2022b). The overall largest share of this global warming potential is caused by carbon dioxide emissions from energy generation through the combustion of fossil fuels.

As first posited by Svante Arrhenius in 1896 for CO_2, increasing concentrations of GHGs lead to a rise in Earth's average surface temperature. The underlying mechanism is GHGs absorbing and isotropically re-emitting the outgoing infrared radiation. An increase in GHG concentrations raises the effective atmospheric emission height of radiation into space, reducing its cooling efficiency due to the lower temperatures at higher altitudes. The most important part of the spectrum where this happens is the infrared region at the border of the atmospheric window, which is responsible for a large part of the energy loss of the Earth

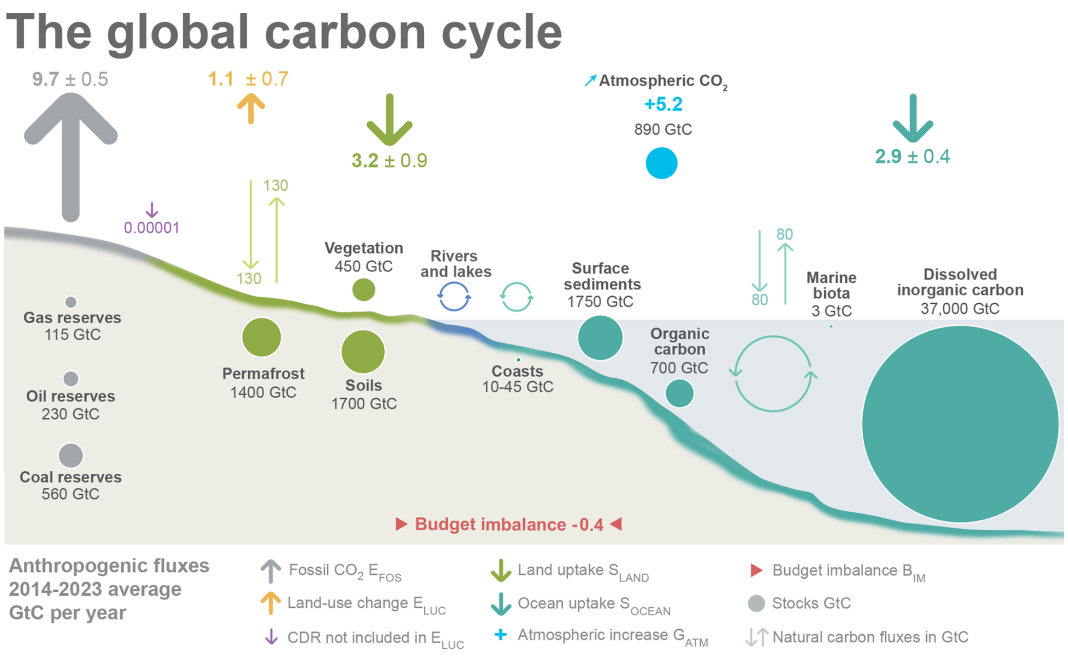


Figure 1.1: Representation of the perturbation of the global carbon cycle by anthropogenic activity. Fluxes (excl. CDR, which is reported only for the decade 2014 to 2023. Arrow size is not indicative of flux strength. Uncertainty of atmospheric CO_2 growth rate is ±0.02 Gt C a^{-1}. Figure reproduced from Friedlingstein et al. (2025) under CC BY 4.0 license.

system. At the same time, the increasing concentrations fill up non-saturated absorption lines which causes more downwelling radiation from the atmosphere onto the surface. Both of these effects contribute to effectively trapping a part of the long-wave radiation and lead to an increase of temperature of the surface-troposphere system. This increase in temperature eventually balances out the energy budget between the Earth's outgoing radiation from the top of the atmosphere with the incoming radiation from the sun. In the following, we focus on CO_2 since its emissions make up the bulk of the 100-year Global Warming Potential.

Anthropogenic CO_2 emissions contribute to the global carbon cycle, of which a yearly review is published as the Global Carbon Budget. Here, we present the results for the decade 2014 to 2023 as published in Friedlingstein et al. (2025, cf. Fig. 1.1). This cycle is dominated by the natural land and marine fluxes with 130 Gt C a^{-1} and 80 Gt C a^{-1} respectively. These fluxes circulate carbon between vast reservoirs containing thousands of gigatons on land and ten thousands of gigatons in the ocean. However, the anthropogenic carbon emissions disturb this balance of emission and uptake. They consist of fossil CO_2 emissions and land use, land use change, and forestry (LULUCF) emissions, accounting for 9.7 ± 0.5 Gt C a^{-1} and 1.1 ± 0.7 Gt C a^{-1} respectively. These get partially absorbed by land and ocean sinks of 3.2 ± 0.9 Gt C a^{-1} and 2.9 ± 0.4 Gt C a^{-1} leaving 5.20 ± 0.02 Gt C a^{-1} to accumulate in the

atmosphere. This constitutes a perturbation of the natural carbon cycle, which is approximately in equilibrium without anthropogenic interference. However small in comparison, this perturbation is responsible for the increase in atmospheric CO_2 leading to climate change. This puts the focus for combatting the root cause of climate change in order to avert negative consequences for human civilization squarely on reducing GHG emissions.

In order to respond to the growing threat of climate change, in 1992 the United Nations (UN) created the UN Framework Convention on Climate Change (UNFCCC) to which all 198 UN members are party. Its objective is to stabilize GHG concentrations in the atmosphere at levels deemed not dangerous, allowing ecosystems to adapt and reducing the impact on human populations. UNFCCC members meet yearly in the so-called Conference of Parties (COP), which first took place 1995 in Berlin. While international progress on this topic was slow, the 21st COP in 2015 marked a turning point, with the Paris Climate Accords being finalized and subsequently signed in 2016. Hailed as a breakthrough in combatting climate change, in the Accords the Parties agree to "[hold] the increase in global average temperature to well below 2 °C above pre-industrial levels [...]" (UNFCCC, 2015).

The mechanism of the agreement stipulates that countries determine autonomously which contributions they make towards the agreed-upon goal. These Nationally Determined Contributions (NDCs) are to be submitted every five years and are supposed to become more ambitious over time. However, these are not legally binding and there are no sanctions in case of shortfalls, which reduces the probability of compliance and limits the treaty's effectiveness. In order to incentivize action, a second pillar of the Accords is the five-yearly Global Stocktake. The Global Stocktake is supposed to monitor and evaluate the Parties' progress towards their NDCs and the compatibility of their NDC with the goal of the Paris Agreement. By way of "name and encourage" as coined by János Pásztor, the UN's former assistant secretary-general, this is supposed to increase countries' ambitions. The data the Global Stocktake is based on is provided by the Enhanced Transparency Framework (ETF), which is a legally binding provision of the Agreement. Thereby all Parties are required to submit Biennial Tranparency Reports (BTRs). These include the National Inventory Reports (NIRs), which include assessments of progress towards the NDCs as well as information about the Parties' progress towards other parts of the Paris Climate Accords.

Each NIR is a comprehensive budget of all GHG sources and sinks in the respective country. As laid out in decision 24/CP.19 of COP19 (Warsaw, 2013), NIRs are created following the 2006 Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories. These provide two approaches to calculate uncertainties: error propagation and Monte Carlo simulation, with the latter one being favored. These uncertainties for Germany amount to around 10 % at the national level, but are much larger at sub-national scales. In order to reduce these and pursuant to the decision 24/CP.19,

Annex I countries like Germany are required to submit a Quality Assurance/Quality Control Plan, which according to the guidelines may include inverse modeling approaches. Here, inverse modeling is referring to data assimilation from atmospheric measurements in order to estimate emissions and their uncertainties under the constraint of prior information. In order to streamline these efforts internationally, in 2015 the 17th World Meteorological Congress of the World Meteorological Organization (WMO) created the Integrated Global Greenhouse Gas Information System (IG3IS).

The German contribution to the IG3IS is the Integrated Greenhouse Gas Monitoring System (ITMS). It is a joint research project funded by the German Ministry of Research, Technology, and Space (BMFTR) between academic and non-academic partners across Germany, which is supposed to develop GHG assimilation capabilities at the German Weather Service (DWD). These assimilations are supposed improve the estimates of emissions and uncertainties within the NIR.

When accounting for anthropogenic emissions, their uneven distribution results in both, challenges and opportunities. While LULUCF emissions occur mostly in rural areas due to agriculture and deforestation, fossil fuel CO_2 emissions are more clustered. Crippa et al., 2021 report that in 2015, European CO_2 emissions of urban and sub-urban areas constitute about 60 % of all CO_2 emissions. However, the shares vary strongly between world regions. While in the Middle East, Russia, and Asia close to 75 % of all CO_2 emissions occur in urban and sub-urban areas, in North America, Africa, and Oceania the share is closer to 50 %. This notwithstanding, the emissions density of urban and sub-urban areas remains disproportionate, independently of the world region. In fact, while in 2015 only 1 % of the Earth's surface was covered by urban and sub-urban areas, they accounted for 50 % of global emissions (Ribeiro et al., 2019; Crippa et al., 2021). This is slated to increase as populations all over the globe continue to urbanize (Jiang and O'Neill, 2017).

According to the German NIR for 2023, Germany emitted 0.75 Gt CO_2eq with a total uncertainty of 2.8 % (Günther et al., 2025). This uncertainty increases to 12.1 % to 13.8 % according to the Monte Carlo simulation and error propagation approaches when adding up sectorial uncertainties. Of these emissions, CO_2 was responsible for 0.68 Gt CO_2eq with a combined sectorial uncertainty of 8.8 % to 9.7 %. According to approach two, the largest share of uncertainty for CO_2 originates in sectors which are associated with urban regions such as Public Electricity and Heat Production, Residential, and Road Transport (30.6 %; Günther et al., 2025). In conjunction with the future increase in urban populations, this makes urban areas an effective target for emissions monitoring and verification efforts (Jiang and O'Neill, 2017).

Moreover, investigating emissions from urban areas yields additional co-benefits. Indeed, any knowledge about the sub-national distribution of GHG emissions also helps to inform

the quantification of other GHG fluxes. This is because low uncertainties in one kind of emissions sector in methods like Bayesian inversions help constrain uncertainties in other sectors. And lastly, the focus on urban areas is a "think global, act local" approach to climate politics. Cities as stakeholders are important actors, as their actions directly shape the day-to-day life of large populations. Many cities have already pledged themselves to climate targets individually and as parts of organizations like C40 or The Global Covenant of Mayors for Climate & Energy (C40 Cities, 2025; Global Covenant of Mayors on Climate & Energy, 2025). Capable monitoring, reporting, and verification systems (MRVs) driving comprehensive accounting methodologies support local mitigation efforts, enhance trust in governance, and constitute valuable contributions to climate action (Jungmann et al., 2022; Ulpiani et al., 2025).

There are multiple preconditions to robust, independent estimates of urban CO_2 emissions. These include accurate transport modeling, reliable information on model uncertainty, high-quality measurements, and a well-constructed inversion which considers sources of error and biases. In this work, we present a full pipeline for optimizing urban emissions estimates. To that end, we investigate the sensitivity of the meteorological performance of the Weather Research and Forecasting Model (WRF) to different physics configurations in the Rhine-Neckar area in Chapter 2 (published in Pilz et al., 2026). In Chapter 3, we use these sensitivity studies to create the Metropolitan Area CO_x RecOrd of Germany 2018 (MACRO-2018), a high-resolution, high-fidelity, and long-term dataset of meteorological information as well as CO_2 and CO concentrations over Europe, focussing on German metropolitan areas. We evaluate its quality against high-precision measurements over the full year of 2018. Finally, in Chapter 4 we constrain the Berlin city emissions from data collected by aircraft measurements within the $[UC]^2$ campaign using a Bayesian inversion approach including a background estimation based on MACRO-2018 concentration and meteorological data. Each of the three chapters contains a specific introduction into the methods used therein at their start. In Chapter 5 we discuss the work done in this thesis and present possible future avenues of research.

2 Atmospheric Transport Modeling

Atmospheric transport modeling is a crucial part of GHG flux estimation using inverse methods. Schuh et al. (2019) and Munassar et al. (2023) show that for both, regional and global surface flux estimates, transport model error is the largest source of uncertainty. Urban regions contribute substantially to CO_2 emissions and are therefore a key target area for emissions monitoring, reporting, and verification systems (MRVs) using inverse modeling. This requires well-characterized atmospheric transport and urban areas are particularly challenging to model due to their heterogeneity and micro-climates. Determining the effect of different parametrization schemes on the (urban) atmospheric transport and its quality is an important step in improving emissions estimation. We, therefore, conduct a sensitivity analysis to determine the optimal physics configuration of the Weather Research and Forecasting Model (WRF) model in a German metropolitan area and present the results in this chapter. We start with giving some physical background and introducing the atmospheric transport model in Section 2.1. Continuing with Section 2.2, we discuss the setup of our simulations. Section 2.3 presents the measured data we use for evaluation and Section 2.4 the evaluation itself. Finally, Section 2.5 discusses the performance and possible causes. Please be aware that large parts of the content of the following chapter have been adapted from Pilz et al. (2026).

2.1 Background and Methods

In the first part of this section, we cover the physical basis of urban modeling. We start by presenting boundary layer physics in general, the structure of the boundary layer, and methods to analyze it. Next, we describe the impact urban regions have on the boundary layer and finally its diurnal dynamics. These descriptions are based on Tulapurkara (2005), Oke (2017), Roedel and Wagner (2017), and Stith et al. (2018).

In the second part, we present the Weather Research and Forecasting Model (WRF), which we use to simulate the urban meteorology. We first give a general overview over the model and then describe some of its parametrizations.

2.1.1 Boundary Layer Physics

Our studies focus on the meteorologically active part of Earth's atmosphere, the troposphere. It is the lowest of the five atmospheric layers, extending from the ground to the tropopause at a height between 6 km and 20 km, depending on latitude. In the mid-latitudes which we are interested in, the troposphere is approximately 9 km high. Dynamically, it is split into two layers: the so-called 'free atmosphere' or 'geostrophic layer', and the 'atmospheric boundary layer' or 'planetary boundary layer (PBL)'.

The free atmosphere is the topmost layer within the troposphere, which extends from the top of the PBL to the tropopause. Its other name, 'geostrophic layer', stems from the approximation of geostrophic balance between the forces acting on the winds therein. This is a theoretical state in which the Coriolis and the pressure gradient force balance exactly and there are no other forces like friction at action. Although this state is an ideal approximation and thus nonexistent in nature, wind flows in the free atmosphere outside of the tropics are close to in geostrophic balance most of the time. These near-geostrophic flows in the free atmosphere are responsible for the movement of large-scale weather systems.

The PBL is defined as that layer of the atmosphere which is affected by the Earth's surface due to friction and surface effects. It is the lowest part of the atmosphere and its height ranges from the surface to between 100 m and 3000 m depending on the meteorological situation. Due to the influence of surface effects, turbulence is a central phenomenon in the PBL. The fundamental description of boundary layer physics originates in fluid theory.

Before the development of boundary layer theory, classical hydrodynamics could not explain why solid objects experience drag when moving through fluids (Tulapurkara, 2005). This apparent contradiction between theory and observation became known as d'Alembert's paradox. Ludwig Prandtl resolved this paradox in his 1904 paper "Über Flüssigkeitsbewegung bei sehr kleiner Reibung", which was published in the proceedings of the Third International Congress of Mathematics held in Heidelberg the following year. Therein, Prandtl posited that any motion of fluids at the interface to objects must be constrained, at a microscopic scale, by the no-slip condition $v_{\mathrm{rel}} = 0$ of the fluid at the object's surface. This condition creates a 'boundary layer' close to the object's surface which mediates between the surface interaction and the standard hydrodynamic conditions above. This intermediate layer is defined by large shear stresses which eventually decrease with distance from the interface. In the 1920s, boundary layer theory was extended to turbulent flow and free shear flows like wakes and jets. Today, it provides the fundamental framework for understanding a wide range of fluid phenomena, including airflow over aircraft wings, fluid flow through pipes, and, importantly for us, the interaction of the atmosphere with the Earth's surface.

We will first provide an idealized view of the boundary layer stratification, which assumes a flat surface of little roughness after Roedel and Wagner (2017). The lowest part of the

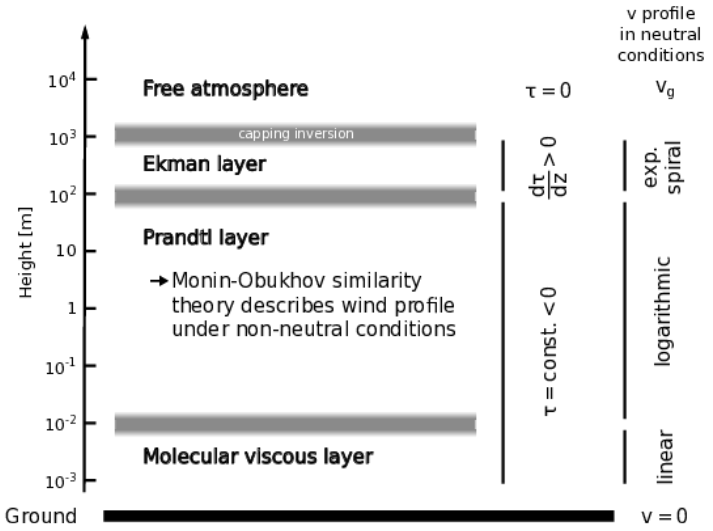


Figure 2.1: Structure of the planetary boundary layer. Boundary layer is split into molecular viscous layer, Prandtl layer and Ekman layer. Molecular viscous layer starts with no-slip condition at surface and linear velocity profile. Both molecular viscous and Prandtl layer have constant turbulent shear stress. Under neutral conditions, the Prandtl layer exhibits logarithmic wind profile. In the Ekman layer, turbulent shear stress decreases until the free atmosphere where it is zero. The wind vector turns and exponentially approaches the geostrophic winds. Figure after Roedel and Wagner (2017).

boundary layer is the so-called 'molecular viscous layer', in which energy dissipation due to molecular friction dominates. This layer is only millimeters thick and creates the connection to the lower boundary condition of the wind velocity, the no-slip condition $v_{\mathrm{rel}} = 0$. Starting from $v = 0$ at the surface, the molecular viscous layer exhibits a linear velocity profile. Above this zone are two layers, which are distinguished by their respective profiles of turbulent shear stress τ. These are the 'Prandtl layer' $\tau = \text{const.} < 0$ and the 'Ekman layer' with $\frac{\mathrm{d}\tau}{\mathrm{d}z} > 0$ until $\tau = 0$ in the free atmosphere. In the Prandtl layer, which is up to 100 m tall, the turbulent eddies grow until they reach the Ekman layer. There, above 100 m, the Coriolis forces increase until the eddies are suppressed. A characteristic phenomenon of the Ekman layer is the so-called Ekman spiral. This describes a change in wind direction from the geostrophic wind towards the surface caused by the frictional forces closer to the surface decelerating and deflecting the winds. This reduction is approximately exponential in nature. An overview over the layers can be found in Fig. 2.1.

The different parts of the boundary layer are closely connected to one another and turbulent exchanges between them are not impeded. During daytime, the dynamics within the boundary layer are dominated by surface heating creating vertical eddies ('thermals'; Oke, 2017). These rise through the layers of the PBL and hit the underside of the free atmosphere. Some of

these eddies overshoot into the free atmosphere and thus entrain warmer, drier air from above into the PBL, which produces a so-called 'capping inversion'. This inversion exhibits a sharp increase in (potential) temperature creating a stable layer at the top of the PBL, significantly reducing air exchange with the free atmosphere.

This stable layer makes the boundary layer interesting to study for us as it accumulates surface influences such as GHGs and air pollutants. Of main interest here is the mixing layer height, as it gives an indication of the volume in which these influences distribute. This is a crucial parameter, as a wrong mixing layer height severely impacts the estimation of pollutant concentrations. Applications interested in this are e.g. air quality monitoring or emissions monitoring of different kinds. We will focus on the retrieval of the mixing layer height from measurement data next for which we introduce Monin-Obukhov similarity theory.

Monin-Obukhov Similiarity Theory

In 1954, surface layer physics had made considerable progress from the days of Ludwig Prandtl and a lot of observational data was available. This progress, however, was mostly constrained to neutrally-stratified atmospheric conditions. For these, the logarithmic wind profile had been found to be a model with sufficient representation. However, while there was a lot of observational data on stable and unstable conditions, there was still no underlying theory.

The theory describing turbulent exchange which is still used to this day is Monin-Obukhov similarity theory (MOST; Monin and Obukhov, 1954) . It empirically describes non-dimensionalized mean flow using scale parameters. These scale parameters are the Obukhov length used to normalize height and the surface friction velocity (Monin and Obukhov, 1954). This Obukhov length characterizes the relative contribution of buoyancy and shear production to turbulent kinetic energy (TKE):

$$\mathscr{L} = \frac{u_\star^3}{\kappa \frac{g}{T} \frac{Q}{\rho c_p}} \tag{2.1}$$

with $u_\star$ being the surface friction velocity, $\kappa = 0.4$ the von Kármán constant, g the gravitational acceleration, T the temperature, Q the turbulent heat flux, c_p the heat capacity, and ρ the density.

From a dimensionality argument (and the Buckingham-π theorem; Buckingham, 1914), Monin and Obukhov (1954) derive the formulation

$$\frac{\mathrm{d}u}{\mathrm{d}z} = \frac{u_\star}{\kappa z} \phi\left(\frac{z}{\mathscr{L}}\right). \tag{2.2}$$

Here, ϕ is a 'universal function', which has to be empirically determined. Defining the height $z = z_0$ where $u(z_0) = 0$ as the roughness length and integrating from there to z yields:

$$u(z) = \frac{u_\star}{\kappa}\left(\ln\left(\frac{z}{z_0}\right) - \Psi\left(\frac{z}{\mathscr{L}}\right)\right). \tag{2.3}$$

This can be used to describe wind profiles for different conditions of atmospheric stability.

Importantly, MOST is only applicable to the Prandtl layer as it requires the friction velocity and Obukhov length to be independent of height. This limited applicability is because the altitude independence is guaranteed through constant heat and momentum fluxes, by which the Prandtl layer is defined. Further assumptions the theory makes are static flow and horizontal surface homogeneity.

Retrieval of Mixing Layer Height

Until the late 19$^{\text{th}}$ century, measurements of the PBL were restricted to using ground-based instrumentation (Stith et al., 2018). Abbott Lawrence Rotch changed this in 1894 when he performed the first upper-air observations with an instrument tied to a kite. Tethered balloons followed the kites and in the 1920s aircraft were used to perform direct measurements of the boundary layer. However, all of these measurement techniques were restricted to fair weather conditions and could only provide data after the instruments, which recorded it onto paper strips, were retrieved. The labor intensity associated with these kinds of measurements prompted the development of the radiosonde, an untethered balloon carrying instruments which transmit their data using radio waves, in 1924. These were further developed over the coming decade and since 1940 almost all upper-air soundings are performed by radiosondes. They have become the main source of information on the dynamics of the PBL.

There is a multitude of methods for retrieving the mixing layer height from vertical column data like that of radiosondes (Seibert, 2000). However, not all of them are useful for our case and would need data like turbulence flux profiles or cloud base height, which are not readily available. The most ubiquitous methods for mixing layer height retrieval are the potential temperature gradient method, the relative humidity gradient method, the parcel method, and the Bulk-Richardson method (Seidel et al., 2010; Li et al., 2021). Of those four, the Bulk-Richardson method has become one of the most widely used due to its robustness under varying atmospheric conditions (Seidel et al., 2012).

The Richardson number is a dimensionless variable quantifying the ratio of the buoyancy and shear flow term in fluid dynamics, which is a measure of turbulence generation vs. suppression. As such, it is akin to the Reynolds, Prandtl, or Rayleigh numbers used for

quantifying different regimes and phenomena. The fundamental number in meteorology is the Flux-Richardson number which is defined as

$$\mathrm{Ri}_f = -\frac{\varepsilon_A}{\varepsilon_S} \tag{2.4}$$

$$= -\frac{\kappa g Q}{c_p \rho \theta {u_\star}^3} \frac{z}{\phi\left(\frac{z}{\mathscr{L}}\right)} \tag{2.5}$$

after Roedel and Wagner (2017). Here, ε_A is the buoyancy energy production rate, ε_S is the shear energy dissipation rate, and θ is the virtual potential temperature. This number can then be simplified to the Gradient-Richardson number using K-Theory, which is a gradient ansatz for the heat and momentum flux. Inserting the turbulent shear stress and the heat flux

$$\tau = -\rho {u_\star}^2 = -\rho K \frac{\mathrm{d}u}{\mathrm{d}z} \tag{2.6}$$

$$Q = -c_p \rho K_Q \frac{\mathrm{d}\theta}{\mathrm{d}z} \tag{2.7}$$

and, additionally, Eq. (2.2) into Eq. (2.5) yields:

$$\mathrm{Ri} = \mathrm{Ri}_f \cdot \frac{K}{K_Q} \tag{2.8}$$

$$= \frac{g}{\theta} \frac{\mathrm{d}\theta/\mathrm{d}z}{(\mathrm{d}u/\mathrm{d}z)^2} \tag{2.9}$$

This number, in turn, gets approximated using finite difference methods to the Bulk-Richardson number, which after Seidel et al. (2012) and Vogelezang and Holtslag (1996) is:

$$\mathrm{Ri}_b(z) = \frac{g}{\theta_s} \frac{(\theta(z) - \theta_s) \cdot (z - z_s)}{(u(z) - u_s)^2 + (v(z) - v_s)^2 + (b{u_\star}^2)}. \tag{2.10}$$

with θ_s as virtual potential temperature at surface, z as height above ground level, $u(z), v(z)$ as horizontal wind velocity components, u_s and v_s as the surface wind velocities, b being a constant, and $u_\star$ as the surface friction velocity. Following Seidel et al. (2012), we set u_s, v_s, and $bu_\star$ to zero, which means our final Bulk-Richardson number reads:

$$\mathrm{Ri}_b(z) = \frac{g}{\theta_s} \frac{(\theta(z) - \theta_s) \cdot (z - 2\,\mathrm{m})}{u(z)^2 + v(z)^2}. \tag{2.11}$$

Seidel et al. (2012) then determine a critical Bulk-Richardson number of 0.25 to be the optimal threshold for defining the PBL mixing height. In the following, we use this formula and the associated threshold.

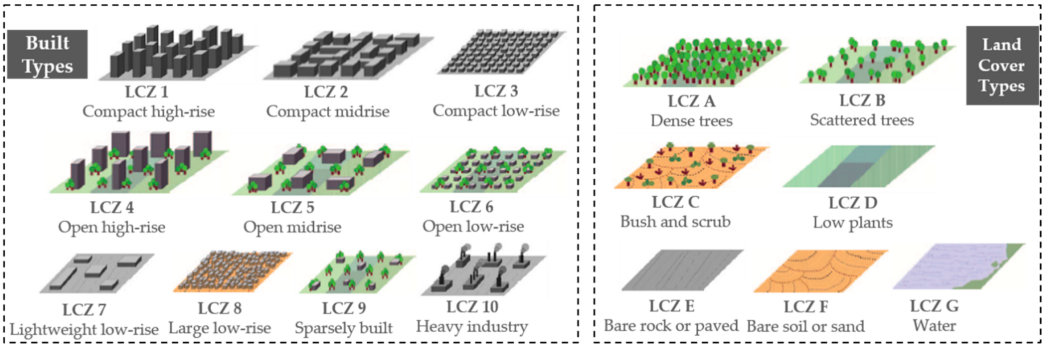


Figure 2.2: Overview of the different types of Local Climate Zones (LCZ). Built types vary depending on building density and height. Figure from Zhao et al. (2019), reproduced under CC BY 4.0 license.

Urban Meteorology

We now turn to the influence of urban areas on the stratification of the PBL. The mechanisms through which this influence is exerted are mass, momentum, and heat exchange with the surface. Urban areas complicate the simple picture laid out above. The resulting type of boundary layer is called 'urban boundary layer (UBL)'.

Urban environments differ significantly from the flat, grassy surface over which we previously defined the four layers of the PBL (cf. Fig. 2.1). They are not a homogenous category of surface with average characteristics which can be applied everywhere. Indeed, Stewart and Oke (2012) develop a classification distinguishing built areas into ten different LCZ depending on building height and density (cf. Fig. 2.2). The largest impact urban surfaces have on the boundary layer is through surface materials, buildings, and heat emissions.

Surfaces in urban regions are more often made of asphalt, concrete, and stone than in rural areas. These replace and seal natural surfaces like sand or earth, grass, and larger vegetation. The largest impacts thereof are the lack of moisture regulation (neither absorption nor retention), the lower albedos, and the comparatively high heat capacities. This leads to a phenomenon called the 'urban heat island' effect, significantly raising temperatures in urban areas compared to rural ones (Oke, 2017). In summer conditions, this can increase the cooling demand and even negatively impact the health of urban populations (Heaviside et al., 2017).

Buildings have many different impacts on the boundary layer. In first order, they simply increase surface roughness. However, buildings are not randomly placed in cities, indeed they usually follow streets. If placed sufficiently close to one another, this can form urban canyons. These canyons channel wind in street direction and block it perpendicularly to it.

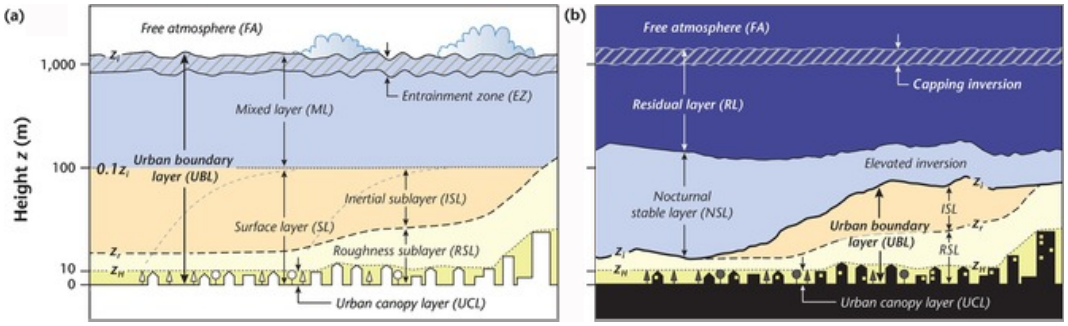


Figure 2.3: Structure of the UBL by day (Panel a) and by night (Panel b). Note that the height scale is logarithmic except for near the surface. Figure from Oke (2017), reproduced with permission of Cambridge University Press through PLSclear.

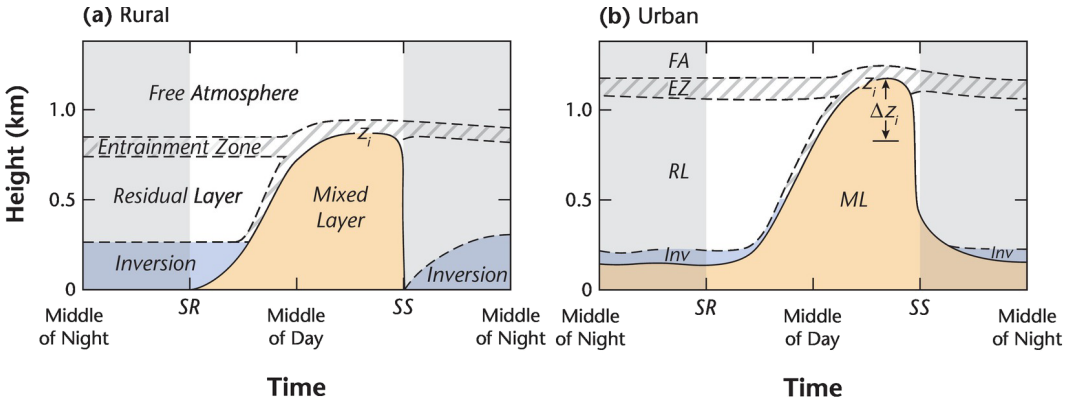
Due to their height, they also lead to a phenomenon called 'radiation trapping', which we discuss further in Section 2.1.2.

Direct anthropogenic heat emissions in urban areas are caused by the energy use of urban populations. These include heating and cooling demand for living and commercial spaces, industrial heat emissions, and heat generated by internal combustion engines.

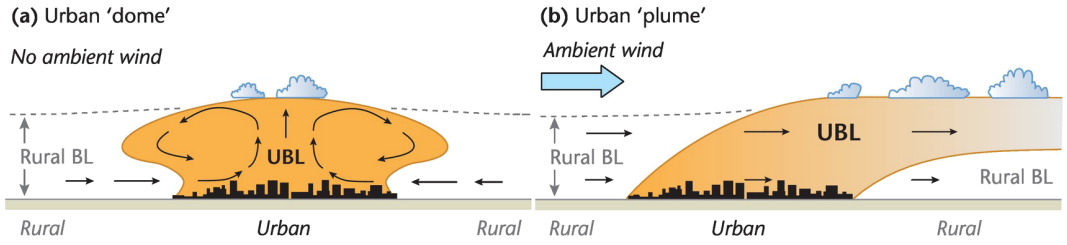
The influence of urban areas leads to a slightly more complicated structure of the UBL compared to the aforementioned conceptualization of the PBL. Due to the built up nature of the urban canopy, the surface roughness is much larger than on a flat plane of grass. Length scales can reach tens of meters depending on building height. This large roughness splits the Prandtl layer into a lower part, which is immediately influenced by urban orography like buildings and street canyons, and an upper part where MOST applies. These are called roughness sublayer (RSL) and inertial sublayer (ISL) respectively (cf. Fig. 2.3).

Since much of the dynamics of the PBL and UBL are driven by thermal phenomena, the largest influence on boundary layer mixing height comes from solar radiation. Indeed, Fig. 2.4i shows how closely the boundary layer diurnal cycle is tied to the change in solar radiation. When the sun rises and the turbulent sensible heat flux to the surface Q_H increases sharply, the buildup phase of the mixing layer starts. This causes the surface to heat up, which in turn causes heat flux $Q_\uparrow$ from the surface to the atmosphere. Depending on surface type, this heat flux splits into sensible and latent heat fluxes Q_S and Q_E. This energy then feeds the thermal eddies building up the mixing layer height until an equilibrium is reached. When the sun sets, Q_H changes sign which causes the boundary layer mixing height to collapse.

As shown in Fig. 2.4i, the boundary layer exhibits broadly similar behavior in urban and rural regions, yet several structural differences remain. These differences are driven by the larger roughness length of the urban canopy, more direct anthropogenic heat emissions by urban populations, as well as the indirect anthropogenic heat emissions due to heat storage



(i) Diurnal cycle of boundary layer



(ii) Boundary layer under wind conditions

Figure 2.4: Diurnal cycle of rural (Panel ia) and urban (Panel ib) boundary layer development in fair-weather conditions. The UBL shows a higher mixing layer height during stable phase and retains a residual during night. Subfigure (ii) shows an urban region exhibiting a 'dome' shape without ambient wind and a 'plume' under moderate winds. Figures from Oke (2017), reproduced with permission of Cambridge University Press through PLSclear.

in materials. Together, these effects not only cause the UBL mixing layer to extend to higher altitudes during the day, it also remains partially sustained throughout the night.

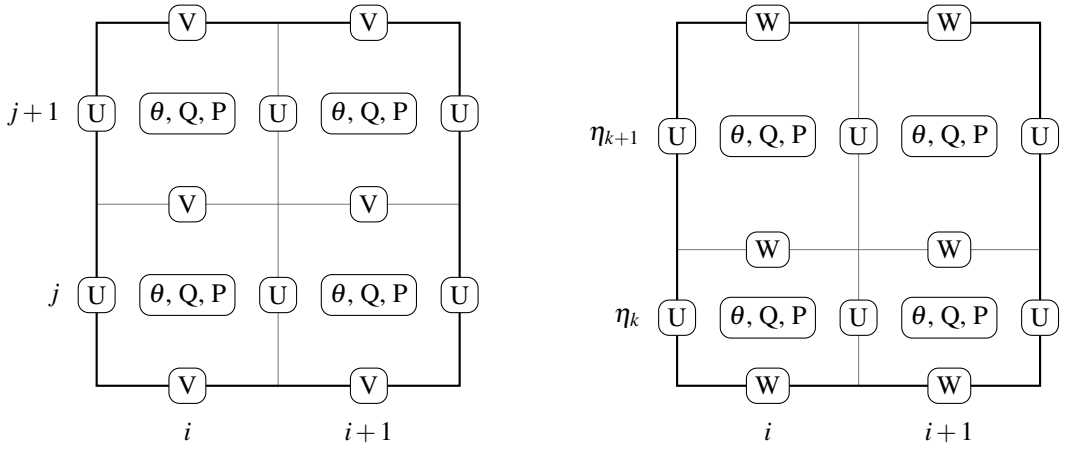
Since this often gets conflated, at this point it is important to mention the difference between the PBL or UBL height and the mixing layer height. The mixing layer height is the height of the layer driven by vertical thermal convection, wherein surface-induced turbulence leads to strong mixing. As it accumulates surface influences, this layer is of particular interest to studies of air quality or emissions. In contrast, while the mixing layer is part of the PBL and UBL heights, the latter also include additional layers such as ground-level inversion layers and the residual layer, where turbulence decays during the night. During cloudy conditions, the PBL and UBL heights tend to be close to the mixing layer height. As the WRF output variable effectively representing the mixing layer height is called `PBLH`, we use the terms PBL height and mixing layer height interchangeably.

This urban boundary layer also interacts with the surrounding area and the prevailing meteorological conditions. During calm conditions, as illustrated in Fig. 2.4ii, vertical mixing creates a local low-pressure system at surface level that draws in air from the surrounding rural region. Simultaneously, a high-pressure system forms at the top of the UBL causing a horizontal expansion of the mixed air masses containing urban influences. This circulation pattern is commonly referred to as an 'urban dome'. When stronger ambient winds are present, the vertical mixing still leads to rising surface level air masses. However, these air masses do not remain spatially confined but are advected with the prevailing winds resulting in a phenomenon called an 'urban plume'.

In this section, we have presented the key effects which must be considered when analyzing urban areas. However, due to the resolution of most atmospheric models, these can usually not be computed explicitly. Therefore, models use parametrizations to include the effects of these sub-grid-scale processes. We explain this in the following section for our model setup using WRF.

2.1.2 Weather Research and Forecasting Model

WRF is an atmospheric modeling system designed for meso-scale research and numerical weather prediction. This means it is intended to numerically simulate the Earth's atmosphere and coupled systems at resolutions of some tens of kilometers. Its development was started in the late 1990's and was driven by the National Center for Atmospheric Research (NCAR), National Center for Environmental Prediction (NCEP), and Environmental Research Laboratory (ERL) (both parts of the National Oceanic and Atmospheric Administration (NOAA)), US Air Force, the Naval Research Laboratory (NRL), the University of Oklahoma, and the Federal Aviation Administration (FAA) (UCAR, 2025). In 2006, it was put into active use and is still used today as the basis for operational forecast models like High Resolution Rapid



(a) horizontal grid

(b) vertical grid

Figure 2.5: Visualization of horizontal and vertical coordinate staggering in WRF's Arakawa-C grid. The x, y, and z wind velocity components are staggered a half-cell in the these same directions. Variables on mass points are (amongst others) temperature, humidity, and pressure. Figure created based on Fig. 3.2 of Skamarock et al. (2019).

Refresh (HRRR) and North American Model (NAM) at NCEP. As such, it is the successor to the MM5 and Eta models being used previously. The model is distributed with two dynamics solvers called the Advanced Research WRF (ARW) and Nonhydrostatic Mesoscale Model (NMM). Since it provides more configuration options and is routinely used for research, we use the ARW in the following. Nowadays, WRF (available in version 4.7.1) is open-source and can be found on github.com/wrf-model. However, institutional support for the model is decreasing and more development and maintenance is performed by the community as its successor Model for Prediction Across Scales (MPAS) continues to be developed.

The ARW core solves the compressible, non-hydrostatic Euler equations (Laprise, 1992). These are cast in flux-form, designed to conserve dry air mass and scalar mass (Skamarock et al., 2019). They are a derivative of the Navier-Stokes equations for a fluid with zero viscosity and thermal conductivity. For numerical reasons, a hydrostatic vertical coordinate is used. In WRF version 4, this coordinate can be a hybrid-sigma coordinate, which interpolates between a terrain-following and a pure-pressure coordinate. In this coordinate, the equations are then integrated using a Runge-Kutta 3 scheme (Skamarock et al., 2019). This integration is performed on two time-scales, a high-frequency acoustic time step for integrating the Euler equations and a low-frequency time step simulating computationally more expensive and slower phenomena like physical parametrizations. These parametrizations are modularized and consist of microphysics, cumulus parametrization, surface, planetary boundary layer, and atmospheric radiation physics models.

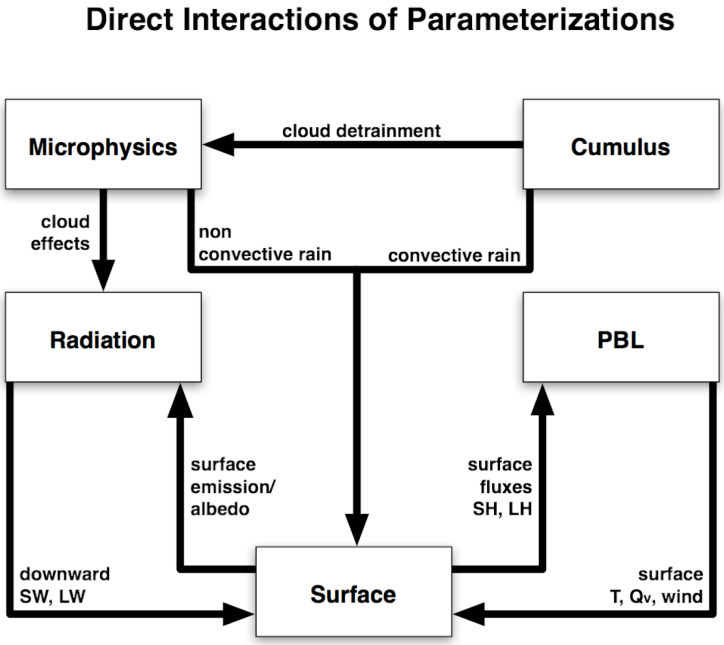


Figure 2.6: Overview over interaction of different physics parametrizations in WRF. Figure reproduced from Skamarock et al. (2019) under CC BY-NonCommercial 4.0 license.

In order to prevent checkerboard-pattern decoupling between cells, WRF employs an Arakawa-C grid staggering approach (Arakawa and Lamb, 1977). This means that vector quantities such as the wind components are defined at points that are staggered by one half-cell to the cell center points (cf. Fig. 2.5). These center points are also referred to as 'scalar points' or 'mass points'.

Some of the model categories above (microphysics, surface, etc.) have sub-models, which can be chosen and configured. For example, the surface treatment can be separated into the surface-layer model, the land surface model, and the urban parametrization amongst others like the ocean model. Of these, the surface layer model calculates friction velocities and exchange coefficients for the land surface model. The land surface model in turn uses these, information from other models, and data on the surface composition like land use to calculate surface heat and moisture fluxes. Since we are interested in urban areas in particular, the urban parametrization model is the topic of the next section. The descriptions in the following sections all follow Skamarock et al. (2019) closely.

Urban Parametrization

There are four physics schemes for urban parametrization implemented in WRF. Their description here is based on both Skamarock et al. (2019) and Joshi et al. (2025). These

physics schemes are the Bulk parametrization (Liu et al., 2006; Chen et al., 2011), the Single-Layer Urban Canopy Model (SLUCM; Kusaka et al., 2001; Kusaka and Kimura, 2004; Chen, 2006), the Building Environment Parametrization (BEP; Martilli et al., 2002), and the BEP-BEM (Building Energy Model; Salamanca and Martilli, 2009). While the Bulk parametrization merely conceptualizes urban areas as a surface with special properties like albedo, surface roughness, and thermal conductivity, the other schemes model these areas more in-depth. Indeed, both the SLUCM and BEP schemes not only add the possibility to simulate different types of urban surface (from WRF v4.3 on also LCZ categories), they also add processes which are simulated for each surface.

Individual urban categories can be assigned parameters, which change variables therein. Examples of this are road widths, roof height distributions (for BEP), non-vegetated fraction, or various heat capacities and thermal conductivities of roofs, roads, and walls. While a mix of roof types (like green roofs) can be simulated per grid cell by both SLUCM and BEP, both are limited to one type of road and wall. These get selected by the dominant urban category.

The SLUCM simulates the urban canopy as an idealized street canyon. This canyon is infinitely long and the buildings simulated in this model are of uniform height (Hang et al., 2024). Most importantly, this urban canopy is a single homogenous layer extending the whole first vertical simulation level (Hang et al., 2024). In contrast, the BEP resolves the urban canopy in more detail. It adds the possibility of choosing different street and building widths for different street direction angles, making surface roughness non-isotropic. It also separates the urban canyon into multiple layers vertically, which do not have to align with the vertical simulation levels.

Both schemes conceptualize the impact urban areas have on the meteorology primarily as driven by its orography. Indeed, urban street canyons are important features of the urban canopy and take a central role in its description. The impact street canyons have on the meteorology stems from their geometry modifying turbulent and radiation fluxes as well as surface roughness. An overview over the radiative impacts of the street canyon geometry is shown in Fig. 2.7. This figure displays the main features namely shading, energy storage, and radiation trapping. These can be simulated in principle by both, the SLUCM and BEP urban schemes.

Both these schemes output important diagnostic variables used to evaluate model performance, like 2 m temperature and 10 m wind velocities. These are not prognostic variables of the schemes, but rather get diagnosed. In the SLUCM scheme, the 2 m temperature is retrieved from a logarithmic interpolation between the surface- and the first model level temperature (and scaled by some correction factors; Joshi et al., 2025). The 10 m wind velocities get diagnosed similarly using a stability function of momentum from MOST for interpolation. In contrast, the BEP assigns the temperature and wind velocities of the lowest

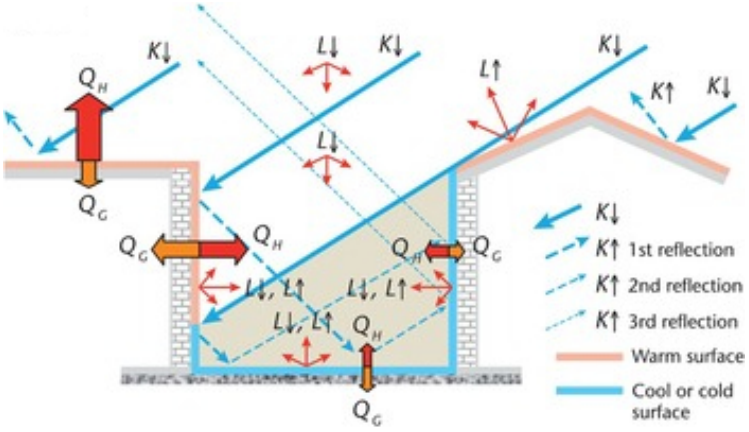


Figure 2.7: Schematic of the main energy fluxes in a simple urban canyon during the day. Important effects are shading, energy storage, and radiation trapping. The variables used here are Q_H (turbulent sensible heat flux), Q_G (substrate heat flux), $L_{\downarrow\uparrow}$ in- and outgoing long-wave radiation, $K_{\downarrow\uparrow}$ in- and outgoing short-wave radiation. Figure from Oke (2017), reproduced with permission of Cambridge University Press through PLSclear.

urban grid level to the diagnostic variables. This has to be kept in mind for any comparison of the two diagnostics.

The Building Energy Model (BEM) is an extension to BEP, which adds the simulation of windows, air conditioning, and direct anthropogenic heat emissions (Salamanca and Martilli, 2009). It can be configured in great detail to simulate different A/C units (coefficient of performance) and settings (comfort temperature, target humidity, etc.). Even parameters like average occupancy density or fraction of photovoltaic panels on roofs can be set. Due to the large uncertainty of these parameters and the high investment associated with researching suitable settings for our purpose, we decided to forgo using this scheme.

Planetary Boundary Layer Model

The vertical sub-grid-scale transport in WRF is treated by the PBL model. It gets the surface fluxes provided to it by the surface layer and land-surface models and uses them to compute the flux profiles in the vertical column. As output, it provides tendencies of temperature, moisture, and horizontal momentum in the column. The model is one-dimensional and works on the assumption of WRF being run at resolutions which do not explicitly resolve boundary layer eddies. These assumptions are not met at resolutions of a few hundred meters. For these scales an explicit TKE diffusion scheme is provided. In the following we will quickly go over the schemes we use in our simulations namely Mellor-Yamada-Janjic (MYJ), Yonsei University (YSU), and Bougeault-Lacarère (BouLac).

In general, there are two types of PBL schemes: prognostic and diagnostic. Prognostic schemes solve the energy balance equations explicitly to get an amount of TKE and use this to estimate the PBL height (mixing layer height). Diagnostic schemes compute estimates of flux profiles from surface fluxes and local stability parameters without solving additional prognostic equations and then diagnose the PBL height through explicit methods like Bulk-Richardson number presented in Section 2.1.1. The central difference between these schemes is how they calculate the eddy diffusivity. Prognostic schemes parametrize it directly over the evolving TKE and a mixing length, while diagnostic schemes calculate it from variables like PBL height, stability, and friction velocity.

The YSU scheme (Hong et al., 2006) is a diagnostic, nonlocal-K scheme, which adds an explicit treatment of the entrainment layer to the older MRF PBL (Hong and Pan, 1996). Its parametrizations are based on study results using large-eddy models. It also implements a counter-gradient term, adding the treatment of the non-local effects of eddies in the transport of heat of momentum in the boundary layer. This extends the local formulation of K-Theory, which can be found in Eq. (2.6). Its effect is a cooling of the lower PBL and a warming of the upper PBL. It defines the PBL top using a critical Bulk-Richardson number criterion. The MYJ scheme (Janjic, 2002) implements the Mellor-Yamada Level 2.5 turbulence closure model (Mellor and Yamada, 1982). As such, it is a prognostic scheme using iteratively solved TKE production/dissipation differential equations. It is explicitly adjusted to improve flux transfer from the urban parametrization schemes. Finally, the Bougeault-Lacarère (BouLac) scheme (Bougeault and Lacarrère, 1989), too, is a prognostic TKE-based scheme that determines a turbulence mixing length based on buoyant parcel energies. It is also explicitly adapted to be compatible with the urban parametrizations.

Surface Layer Model

The surface layer model provides the friction velocities and exchange coefficients for momentum and moisture for the PBL model on the atmospheric side and the land-surface model on the ground side. It uses MOST to relate surface fluxes to mean gradients between the lowest model level and the surface. Using this stability theory, it calculates the 2 m temperature and 10 m wind diagnostics when not simulating over urban areas, where these get overwritten by urban parametrizations. We use two schemes, Revised MM5 Monin-Obukhov (MM5) and Monin-Obukhov (Janjic) (MO).

The MM5 scheme (Jiménez et al., 2012) is a newer version of the similarity theory used in the MM5 model. It updated the stability function and improved the consistency between Ri and $\frac{z}{\mathscr{L}}$. The MO scheme (Janjic, 2002) is based on the same formulations of stability functions as the MYJ boundary layer scheme. It adds an explicit viscous sub-layer, which is scaled according to roughness height.

Land Surface Model

The land-surface model (LSM) uses the exchange coefficients from the surface layer model, radiative fluxes from the radiation scheme, and precipitation from the microphysics and cumulus parametrization schemes to simulate various above and below-ground processes. It provides heat and moisture fluxes as output, which serve as lower boundary conditions for the PBL model. Our study looks at the Noah and the Noah-Multiparametrization Land Surface Model (Noah-MP) schemes.

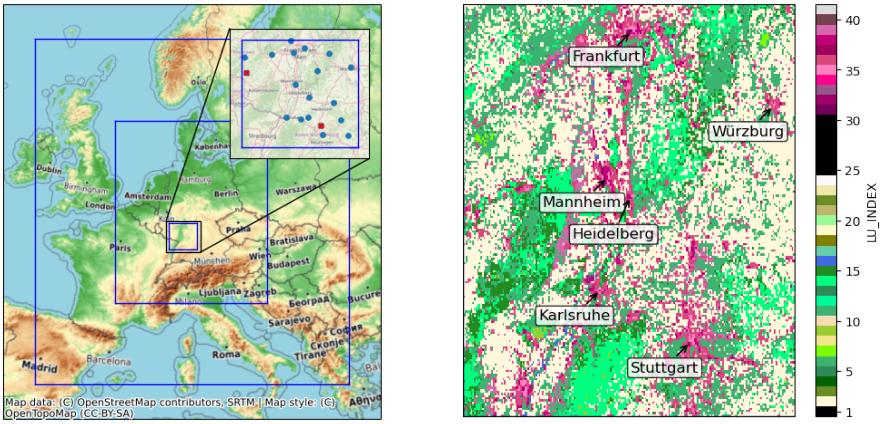
The Noah-MP scheme (Niu et al., 2011; Yang et al., 2011) is a newer, more complex version of the Noah scheme (Chen and Dudhia, 2001). Both have an internal model state and treat complex processes like runoff, evaporation, and water absorption of different soil types. They, furthermore, treat above and below-ground vegetation and snow cover. While the Noah scheme has four soil layers and a bulk snow layer, the Noah-MP scheme has three variable snow layers. It is also able to simulate dynamic vegetation, plant stomatal resistance, and even crop growth including irrigation.

2.2 Simulation Setup

In Section 2.1.2, we have presented a fraction of the available physics schemes one can choose from using ARW per physics parametrization. However, there is no clear answer as to which combination of schemes performs better, as that depends on simulation domain, meteorological situation, and so on. Therefore, we have to test different configurations in order to find a good match for our situation. In this section, we introduce the setup we used for the simulation ensemble which makes up this sensitivity study. It covers the study region used, the model configuration with regards to input and boundary conditions, as well as the ensemble of physics configurations itself. This section is adapted from Pilz et al. (2026).

2.2.1 Study Region

The Rhine-Main-Neckar area is one of Germany's largest metropolitan areas and as such responsible for 8.2 % of German CO_2 emissions according to TNO GHGco v4.1 (Super et al., 2020). There are two particularly large sources of emissions, the Großkraftwerk Mannheim (GKM), a hard-coal power plant and the BASF complex in Ludwigshafen, the worldwide largest contiguous chemical park. It is furthermore home to lots of meteorological and GHG observation equipment. The region is characterized as a so-called polycentric urban region (PUR; Meijers et al., 2017). These types of regions are defined as "clusters of historically and administratively distinct but proximate and well-connected cities" (Meijers et al., 2017).



(a) Domain location (b) Land use

Figure 2.8: The WRF domain setup is displayed in Panel (a) with the resolutions of the three domains being 25, 5, and 1 km. Within the smallest domain are the 19 DWD meteorological stations (blue dots) and two radiosonde stations (red crosses) we use to compare against the simulation. Panel (b) shows the dominant land use categories of the highest-resolution domain with the largest cities being tagged. The names of the US Geological Survey (USGS) land use categories are to be found in Table A.2. Map data copyrighted by OpenStreetMap contributors and available from openstreetmap.org. Figures taken from Pilz et al. (2026).

PURs are very common in Europe, housing 25 % of the population (Meijers et al., 2017) and thus a valuable target for modeling.

Our area of interest contains the large cities of Frankfurt (750,000 inh.), Stuttgart (630,000 inh.) as well as medium-sized cities like Mannheim and Karlsruhe (300,000 inh.) and small cities like Heidelberg (160,000 inh.) and Würzburg (127,000 inh.). In Fig. 2.8, we can see the proximity yet distinctness of the cities which is the defining criterion of a PUR. We also see that our domain consists of 16 % urban cells, which in turn are dominated by 67 % "open lowrise" cells, LCZ category 6 in Fig. 2.2. Our study region is favorable as it hosts many meteorological stations of the DWD, which allows us to thoroughly test the meteorological performance of the model.

2.2.2 Model Setup

The model we use for this study is WRF in version 4.3.1 with the ARW core. It provides the option to increase the simulation resolution in parts of the domain which one is interested in by adding higher-resolution nests. We start with the domain over Europe at a 25 km resolution, in which we embed a 5 km resolved domain centered on Germany. The innermost nest is centered on the Rhine-Main-Neckar region and has a resolution of 1 km. These three nests do not feed their results back to the lower-resolution domains (one-way nesting).

We use 42 vertical layers with 14 levels below 1.5 km like in previous studies, for example Lian et al. (2018). The Single-Layer Urban Canopy Model (SLUCM; Chen, 2006) urban parametrization does not allow any vertical levels below the highest building height. Because of this the lowest level for the SLUCM configurations is set to 90 m whereas it is at 15 m for the configurations using Building Environment Parametrization (BEP; Martilli et al., 2002; cf. Table A.1).

WRF provides an option for data assimilation, which can be used to keep the model dynamics closer to reality. Different methods of data assimilation are implemented like grid nudging, observation nudging, or spectral nudging. All of these options add an additional forcing term to the model equations, which keeps the model close to the given data. For our simulations, we use grid-nudging, which we restrict to the outer two domains above the PBL. This is done in order to allow WRF to develop finer-scale structures at higher resolutions while keeping it close to reality overall. Grid-nudging in this way is a common setting used for example by Lian et al. (2018) or Ho et al. (2024). In preparatory studies, we found the best temporal resolution of the nudging to be three hours, striking a balance between simulation performance and storage constraints. The analysis data used for grid-nudging is the ECMWF Reanalysis v5 (ERA5) data presented in the following Section 2.2.2 (Hersbach et al., 2020a).

However, there are still some residual long-term deviations in non-nudged variables. One example is long-term drifts in soil-water content due to drifts in the LSM water budget (Ho et al., 2024). We restrict the impact of these deviations by reinitializing the meteorology every seven days from ERA5 data. This means that we start a new simulation using initial condition data from ERA5. However, since the model needs some time to relax into its own dynamics from a foreign model state, we give it a spin-up period of 6 h to do so before we consider the model output as valid.

Input and Boundary Condition data

While it is possible to run global simulations with WRF, we use it as a limited area model. This means that we need lateral boundary condition data to inform the model about the meteorological conditions outside of its simulation domain. We also need to inform it about the initial meteorological state from which it is supposed to start.

For both of these applications, we use ERA5-reanalysis data provided by European Center for Medium-Range Weather Forecasts (ECMWF; Hersbach et al., 2020b). In particular, we use ERA5 hourly data on single levels (C3S, 2018). ERA5 is a reanalysis dataset which combines a state-of-the-art meteorological model used for forecasting with measurement data. In order to get the simulated data as close as possible to reality, it assimilates data streams like satellite, in-situ, and radiosonde measurements using 4D-Var. ERA5 uses a 12-

hour 4D-Var assimilation window (0 UTC to 12 UTC and 12 UTC to 24 UTC), during which model fields are adjusted to best fit the available observations. The assimilated variables include surface pressure, temperature, and humidity from weather stations as well as wind data from radar, aircraft, and radiosonde soundings. Many in-situ data which are assimilated in ERA5 are provided by the WMO Information System (WIS). The DWD ingests all of the data of its weather stations into the WIS. This means that the ERA5 data is not independent of the DWD data. In general, reanalysis like ERA5 have lower resolution than operational forecasts (31 km vs. 9 km) but provide a temporally consistent estimate of the atmospheric state by assimilating past observations.

The ERA5 data is provided for the time from 1940 to present. It is computed using the cycle CY41R2 ECMWF Integrated Forecast System (IFS) model in a hybrid sigma/pressure coordinate with 137 levels similarly to WRF. Internally, the ECMWF models use spherical harmonics grids. The dynamical core is computed on the truncated linear spectral grid TL639 (639 points, 31 km). We are running our simulations on the Levante supercomputer at the German Climate Computation Center (DKRZ). In order to save time and bandwidth, we decided to use the locally provided ERA5 data files instead of downloading them. These files are, however, provided on the native spherical harmonics grids which is why they have to first be converted to lat/lon using Climate Data Operators (CDO).

We replace the default USGS GMTED topographic data at 30″ resolution with Copernicus digital elevation model (DEM) data (Copernicus, 2022), which is available at resolutions of up to 90 m. Within the European Union (EU), we also replace the default USGS land use data at 30″ resolution using `wps_xr` (Pilz, 2025b). In a first step, it is replaced with Coordination of Information on the Environment (CORINE) data re-categorized to USGS at a resolution of 250 m (Breuer et al., 2021). This data is then augmented in the urban regions with information from the global LCZ dataset, using the `W2W` software (Demuzere et al., 2022). Due to a bug in WRF from version 4.3 on, which is fixed with commit `d96478d` (#2153, released in WRF v4.7.0), the USGS LCZ data was overwritten with default USGS urban category `LU_INDEX = 1` where `FRC_URB2D > 0.5`. However, as this only directly affects one of our stations (Offenbach-Wetterpark), this bug is neglected in the following analysis.

Ensemble of Physics Configurations

We simulate four contiguous one-month periods — April, July, September, and December of 2020. In comparison to previous studies, this is relatively long total simulation time. These months were selected using the phenological calendar of 2020 to represent the four meteorological seasons. All periods are run using 16 different physics configurations of WRF. Within these configurations, the land-surface model (LSM), surface layer model (SLM),

	Boundary Layer	Land Surface	Surface Layer	Urban
bep_bl_n_mm5	Bou-Lac	Noah	MM5	BEP
bep_bl_n_mo	Bou-Lac	Noah	Monin-Obukhov	BEP
bep_bl_nmp_mm5	Bou-Lac	Noah-MP	MM5	BEP
bep_bl_nmp_mo	Bou-Lac	Noah-MP	Monin-Obukhov	BEP
bep_myj_n_mo	MYJ	Noah	Monin-Obukhov	BEP
bep_myj_nmp_mo	MYJ	Noah-MP	Monin-Obukhov	BEP
bep_ysu_n_mm5	YSU	Noah	MM5	BEP
bep_ysu_nmp_mm5	YSU	Noah-MP	MM5	BEP
slucm_bl_n_mm5	Bou-Lac	Noah	MM5	SLUCM
slucm_bl_n_mo	Bou-Lac	Noah	Monin-Obukhov	SLUCM
slucm_bl_nmp_mm5	Bou-Lac	Noah-MP	MM5	SLUCM
slucm_bl_nmp_mo	Bou-Lac	Noah-MP	Monin-Obukhov	SLUCM
slucm_myj_n_mo	MYJ	Noah	Monin-Obukhov	SLUCM
slucm_myj_nmp_mo	MYJ	Noah-MP	Monin-Obukhov	SLUCM
slucm_ysu_n_mm5	YSU	Noah	MM5	SLUCM
slucm_ysu_nmp_mm5	YSU	Noah-MP	MM5	SLUCM

Table 2.1: WRF namelist settings describing the physics schemes used by the different configurations. Not all combinations are compatible, so this is a subset of the 24 theoretical combinations of settings. Table taken from Pilz et al. (2026).

urban canopy model (UCM), and planetary boundary layer (PBL) model implementations used by WRF are varied. Each of these models is available in multiple implementations (schemes) in WRF. The specific configurations used in this study are summarized in Table 2.1.

2.3 Measured Data

In this section, we present which measured data we use to evaluate the performance of our atmospheric transport model. The density of measurement stations is comparably high with 20 DWD stations being located in our simulation domain. This section is adapted from Pilz et al. (2026).

2.3.1 Planetary Boundary Layer Height

In order to ensure compatibility between PBL height definitions of measured and simulated data, we retrieve these using the Bulk-Richardson method as laid out in Section 2.1.1. To determine the virtual potential temperature, we use the `virtual_temperature` function of `metpy.calc` (May et al., 2022). The surface virtual potential temperature is calculated using the potential temperature diagnostic at two meters (`TH2`) and the water vapor mixing ratio (`QVAPOR`) in the lowest model level. Here, we make the assumption that the water vapor mixing ratio in the lowest model level is similar to the ratio at 2 m. This assumption carries some uncertainty. We found deviations of about 7.5 % for realistic conditions in our setup adding uncertainty to the retrieved PBL height. However, the necessary variable to calculate this exactly (`Q2`) is not available in our simulation outputs.

In order to calculate the PBL height from the Bulk-Richardson number profile, we interpolate it to the point where it first crosses the critical Bulk-Richardson number using `xgcm` (Abernathey et al., 2025). This is an empirical number, whose choice is studied in Vogelezang and Holtslag (1996) and Seidel et al. (2012). In accordance with their findings, we choose this number to be 0.25. However, sometimes between sensible PBL heights, we find spurious low values. This is caused by the Bulk-Richardson number at the first level being slightly higher than 0.25. These profiles first decrease from a little over 0.25 until they start to rise again, crossing 0.25 again at a higher, more realistic, point. In the instances where the crossing of the critical Bulk-Richardson number is between the first and second level, after communication with the data providers, we set the Bulk-Richardson number of the first model level to zero.

The observation time associated with each sounding, which is stated in the Integrated Global Radiosonde Archive, version 2 (IGRA) database is a time close to the completed ascent of the radiosonde, which is one hour and 15 min after the radiosonde release (06:00, 12:00, 18:00, and 00:00 UTC). In order to be closer to the time this radiosonde passed the PBL height, we change this time to be only 15 min after launch after discussion with the data providers. We also select only the measurement times when both stations (Stuttgart/Schnarrenberg and Idar/Oberstein) released radiosondes. This means the radiosonde measurement times we use in our evaluation are 05:00, 11:00 and 23:00 UTC.

2.3.2 Wind Data

We compare the 2 m temperature and 10 m wind diagnostics output of our model against the same variables measured at DWD stations within our domain. As commonly done, e.g., in Lin et al. (2021), we compare our model's instantaneous wind diagnostics output to the hourly averaged measured wind data. These wind data, however, are not useable in their raw form and have to be corrected first. Natively, they are reported on the hour as an average of the previous 60 min. Therefore, we assign their measurements to 30 min after the hour and interpolate them to the full hour to correspond to the simulation data.

An additional peculiarity of the raw wind data provided by DWD concerns the measurement height. While the 2 m air temperature is recorded exactly two meters above a flat, grassy surface, the 10 m wind data are not measured at a uniform height of 10 m above ground level. The 20 stations initially in our dataset measure the wind velocity and direction at 10, 12, 14.5, 15, and 29 m. These have to be corrected down to 10 m in order to be comparable to our simulated 10 m diagnostics.

We do this correction using an approximation of Eq. (2.3) for neutral atmospheric conditions as we do not have any independent information about the atmosphere's stability. The approximation we are using is (compare Liu et al., 2021):

$$u(z) = \frac{u_\star}{\kappa}\left(\ln\left(\frac{z}{z_0}\right) - \Psi\left(\frac{z}{\mathscr{L}}\right)\right) \quad \wedge \quad \Psi = 0$$
$$\Rightarrow u(10\,\mathrm{m}) = u(z)\frac{\ln(10\,\mathrm{m}/z_0)}{\ln(z/z_0)}$$

The wind profile Ψ is set to zero as we assume neutral conditions. Furthermore, we assume $z_0 = 0.03\,\mathrm{m}$, the surface roughness length of grass. This is the same value that ERA5 is using for the estimation of the Obukhov length at SYNOP stations (ECMWF, 2021).

When applying this correction, the measured wind velocities decrease by 3, 6, 6.5, and 15 % at the stations measuring at 12, 14.5, 15, and 29 m respectively. As the correction at the station measuring at 29 m (Weinbiet) is more than double that of the other stations, we decide to exclude this stations from our analysis. This leaves us with a final collection of 19 DWD stations which were active and measuring during our simulation time period.

These 19 stations are categorized into 'urban' and 'rural' sites by analyzing their surrounding land use. This is necessary as the effect of the urban influence is non-local through transport even though WRF calculates urban influence only in the column directly above the cell. Stations are defined as 'urban', when the majority of the surrounding tiles have an dominant land use index which is urban (`LU_INDEX` > 30). The others are labeled as 'rural'. General station information is to be found in Table A.3.

2.3.3 Statistics

For the performance evaluation, we apply standard statistical measures like mean absolute bias (MAB), root mean squared error (RMSE) and correlation. Regarding correlation, we use the Pearson-R coefficient for linear variables and the circular correlation as defined in Jammalamadaka et al. (2001) for circular variables like wind direction. The measures which we use are defined as follows:

$$\Delta d_i = \begin{cases} s_i - o_i, & \text{if } s_i - o_i \leq |180| \\ s_i - o_i - 360, & \text{if } s_i - o_i > 180 \\ s_i - o_i + 360, & \text{if } s_i - o_i < -180 \end{cases}$$

$$\text{MAB} = \frac{1}{n}\sum_{i=0}^{n} |s_i - o_i| \text{ or } \frac{1}{n}\sum_{i=0}^{n} |\Delta d_i| \qquad \text{rMAB} = \frac{\text{MAB}}{\bar{o}}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n} (s_i - o_i)^2} \text{ or } \sqrt{\frac{1}{n}\sum_{i=0}^{n} (\Delta d_i)^2} \qquad \text{rRMSE} = \frac{\text{RMSE}}{\bar{o}}$$

with s_i being simulated values, o_i observed values, $\bar{o}$ average of observations, n number of samples, and Δd_i wind direction difference.

2.3.4 Taylor Density Diagrams

A visualization commonly used to compare model performance across multiple stations or physics configurations are Taylor diagrams (Taylor, 2001). Taylor diagrams can compare multiple different time series of interest to a reference time series. They exploit the fact that the bias-corrected RMSE is related to the correlation and standard deviation via the law of cosines. Indeed, with the RMSE separated into bias and bias-corrected RMSE like:

$$E^2 = \overline{E}^2 + E'^2, \tag{2.12}$$

$$E'^2 = \sigma_f^2 + \sigma_r^2 - 2\sigma_f \sigma_r R \tag{2.13}$$

can be identified with the law of cosines:

$$c^2 = a^2 + b^2 - 2ab\cos\phi\,. \tag{2.14}$$

Hereby, E is the RMSE, $\overline{E}$ is the bias, E' is the bias-corrected RMSE, σ_f is standard deviation of the forecast time series, σ_r is the standard deviation of reference time series, and R the

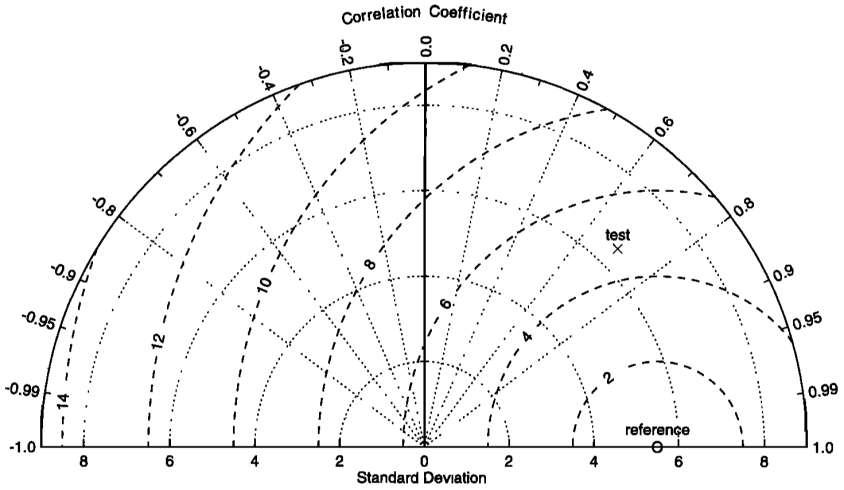


Figure 2.9: Example of a Taylor diagram. The radial distance from the origin is proportionate to the standard deviation, the azimuthal position describes the correlation. The relevant distance which is equivalent to the bias-corrected RMSE is the distance between the 'reference' point at a correlation of one and the 'test' point. Figure reproduced from Taylor (2001) with permission of Wiley through RightsLink.

Pearson-R correlation. Equation (2.13) derives from the definition of the bias-corrected RMSE from forecast f and reference r as

$$E'^2 = \frac{1}{n}\sum_{i=1}^{n}\left((f_i-\overline{f})-(r_i-\bar{r})\right)^2 \quad \wedge \quad f_i = \overline{f}+f_i' \tag{2.15}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(f_i'-r_i'\right)^2 \tag{2.16}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(f_i'^2+r_i'^2-2f_i'r_i'\right) \tag{2.17}$$

$$= \frac{1}{n}\sum_{i=1}^{n}f_i'^2+\frac{1}{n}\sum_{i=1}^{n}r_i'^2-\frac{2}{n}\sum_{i=1}^{n}f_i'r_i' \tag{2.18}$$

$$= \sigma_f^2+\sigma_r^2-2\,\mathrm{cov}(f,r) \tag{2.19}$$

$$= \sigma_f^2+\sigma_r^2-2\sigma_f\sigma_r R(f,r)\,. \tag{2.20}$$

An example of a Taylor diagram can be seen in Fig. 2.9. Importantly, the distance which is equivalent to the bias-corrected RMSE in the Taylor diagrams is not the distance to the origin. Instead, it is the distance to the reference point, which is located at a correlation of one.

We would usually mark a station or configuration using a marker at the point where their (normalized, if multiple stations) standard deviation and correlation are. However, since we

have 19 meteorological stations and 16 simulation configurations, it is not feasible to plot all 304 markers. In order to retain some visual clarity, we use averages but augment these with the kernel density estimates (KDEs) over the full data behind the averages. These KDEs are computed over all stations and configurations using one scheme of the most sensitive model. We find that this reduces the amount of data displayed in a Taylor diagram to a reasonable level.

To identify the aforementioned most sensitive model, for all schemes of this model we average the MABs of all configurations using this scheme. This yields one averaged MAB value for each individual scheme of a model. Now we can evaluate the range of these MAB values. The largest range of MAB values between the schemes of a model indicates the model most sensitive to change in scheme. The schemes of this model are then assigned different colors in this diagram. This so-called "Taylor density diagram" presents large amounts of information in a visually intuitive way. It reduces the visual clutter of large datasets while retaining information about the dataset's variability behind the averages.

2.4 Evaluation

We now present the model data and evaluate it against data measured by the DWD stations. First, we get an overview over the dataset and look at aggregated statistics. Then, we look at the individual meteorological variables. Finally, we investigate the robustness of our findings. This section is adapted from Pilz et al. (2026).

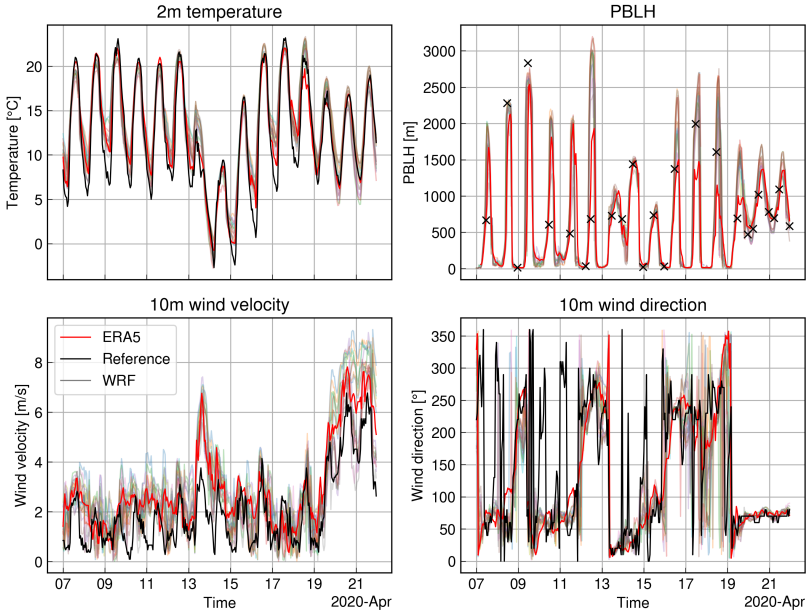
2.4.1 Time Series

In order to give the reader an impression of the data, we present an example two-week period in this section. This is done purely for illustrative purposes. The statistics for the full time series are to be found in Table A.4. We present simulations and measurements for the two stations in our domain where we have radiosonde soundings to measure PBL height and all meteorological variables. These are the stations Stuttgart/Schnarrenberg and Idar/Oberstein and for both stations we present the time period between Apr. 7th and Apr. 21st of 2020. The two stations differ in land use type, as Idar/Oberstein is a station classified as rural and Stuttgart/Schnarrenberg is an urban station.

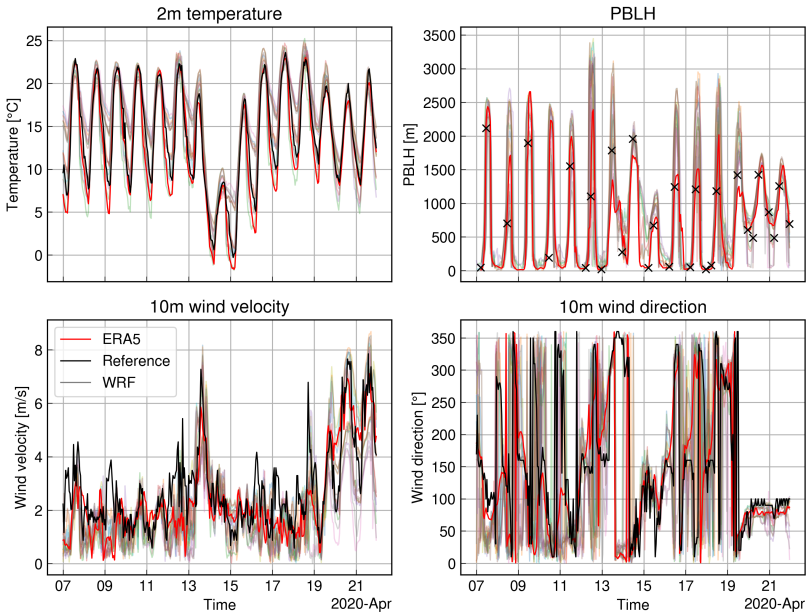
Figure 2.10 shows that within the given time period, the ensemble of simulations captures the diurnal and the synoptic variations of the 2 m temperature well. However the stations do differ in their performance with regards to this variable. While we generally see a good agreement between simulations and measurements during the day, we see more variation at night. There, the ensemble splits into two distinct regimes at Stuttgart/Schnarrenberg, which it does not do at Idar/Oberstein. These regimes are configurations using different urban parametrization schemes (SLUCM and BEP). The ERA5 record which assimilates these measurements matches them closely, however it does overestimate the nighttime temperatures slightly at Idar/Oberstein while underestimating them at Stuttgart/Schnarrenberg which is an indication of the limited resolution of ERA5.

When comparing the PBL height measurements, we immediately see the main challenge, which is the sparse temporal resolution of the measured data. Radiosondes measurements which we retrieve PBL heights from are only taken at 5:00, 11:00 and 23:00 UTC at both stations, while Idar/Oberstein launches another one at 17:00 UTC. While the data in this period matches well, we do see a small overestimation of PBL heights by the WRF simulation during this limited time period. Again, the ensemble shows a larger spread in the simulations at Stuttgart/Schnarrenberg than at Idar/Oberstein (here especially during transition phases) and a higher maximum PBL height at Stuttgart/Schnarrenberg. This fits to the more urban characteristic of Stuttgart/Schnarrenberg.

Looking at wind velocity, larger synoptic changes are simulated correctly, contrary to the short-term ones. At Idar/Oberstein, the ERA5 data and the simulation ensemble both deviate



(a) Idar/Oberstein



(b) Stuttgart/Schnarrenberg

Figure 2.10: Qualitative overview of simulations and measurements of meteorological parameters at rural station Idar/Oberstein and urban station Stuttgart/Schnarrenberg in time period of Apr. 7th to 21st 2020. ERA5 data are shown as red lines, measurements by the DWD meteorological station as black lines and black crosses are PBL heights retrieved from radiosonde launches. Colored lines are the WRF ensemble to give an impression of performance and spread. Please note that in order to discourage comparisons of individual scheme performance which would not be representative, legend for the colored lines is omitted. Figures taken from Pilz et al. (2026).

from measurements but are close to one another. This could be the result of local effects not being resolved in our simulations. The deviation is also more pronounced during night time.

Wind direction has a larger ensemble spread at Stuttgart/Schnarrenberg than at Idar/Oberstein despite similarly steady meteorological conditions. Both, ERA5 and our simulations exhibit similar performances here.

2.4.2 Statistics

Figure 2.11 gives an overview over all aggregated statistics for the whole dataset. For all meteorological variables (relative) MAB, (relative) RMSE and correlation are shown. We calculate the relative MAB and RMSE in order to make these variables comparable across seasons. We do this as, for example, the PBL height MAB is much smaller during winter. This, however, is not due to a more correct estimation, but it is rather driven by generally lower PBL heights. In this plot, we average over all available stations.

While there is seasonal variability in overall WRF simulation performance, the choice of optimal physics configuration is dependent on both variable of interest and season. This indicates that, at least in terms of relative performance, seasons do not have an outsized influence. While this variability across seasons and variables is present, some configurations like the one using the SLUCM, MYJ, Noah-MP, and MO schemes give overall better results.

In general, for 2 m temperature, PBL height and 10 m wind velocity we do find larger ensemble spread in winter than for other seasons. We also find the performance of the best configuration (in terms of MAB) to differ by 7 % for wind direction, 16 % for 2 m temperature and wind velocity and 26 % for PBL height from the ensemble average. These deviations increase two- and three-fold when comparing these relative deviations from ensemble mean in urban vs. rural areas for variables strongly influenced by urban surfaces (like 2 m temperature and PBL height).

One general characteristic shared by all the following Taylor density diagrams is the ensemble spread being more dominated by station characteristics (density contour lines) than by choice of physics scheme (markers). The software used for all Taylor diagrams in this thesis is based on Copin (2021).

2.4.3 2m Temperature

With regards to 2 m temperature, the model with the highest sensitivity (cf. Section 2.3.4) is the urban canopy model (UCM). This large difference between SLUCM and BEP configurations is clearly visible in Fig. 2.12. The diurnal cycle is well-captured by SLUCM in April, July, and September. However, in December there is a unique nighttime and afternoon

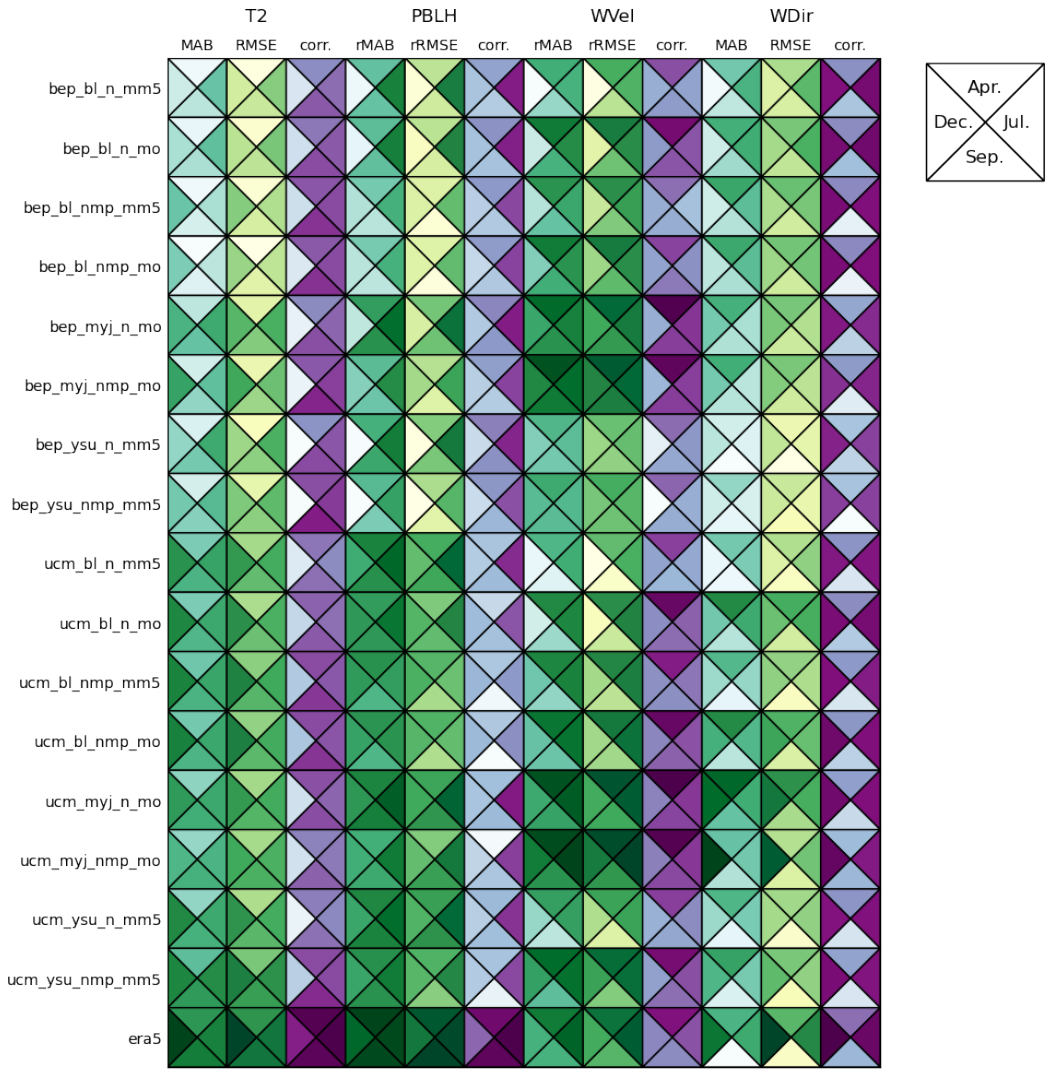


Figure 2.11: Relative MAB, relative RMSE, and Pearson-R correlation for 10 m wind velocity and PBL height and absolute MAB, RMSE, and Pearson-R correlation for 2 m temperature and 10 m wind direction across all WRF configurations and ERA5 data are shown in this Portrait plot (Gleckler et al., 2008). Each tile is split into the four simulation periods of April, July, September, and December of 2020. Each column has its own linearly normalized color map between highest and lowest value. Dark colors indicate better performance (i.e. lower MAB, RMSE and higher correlation). Figure taken from Pilz et al. (2026).

underestimation. In contrast to that, while the BEP configurations exhibit a full-day positive bias in December, the bias is restricted to nighttime in other months.

As the Taylor diagrams only display bias-corrected RMSEs, the performance difference is not visible in December (cf. Fig. A.1). However, the sensitivity to the UCM is clearly visible in the other panels of the Taylor density diagrams. Since the diurnal cycle amplitude is lower in winter, the correlation performance is worse than in the other months (cf. Figs. 2.11 and 2.12). In general, these differences between UCM are driven by urban stations, which is visible in Fig. A.2.

These deviations at urban stations could have different causes, either actual differences in the performance of the urban parametrization or differences in the vertical layer distribution of the SLUCM and BEP runs. The vertical layer distributions are different between the two parametrizations, as the implementation of the SLUCM scheme restricts the height of the first model level to be higher than the highest building height in the domain, as described in Section 2.2.2. We, thus, repeat a simulation of December 2020 using the YSU, Noah-MP, MM5 and BEP configuration with the vertical layer distribution of the SLUCM runs. We find that this change in vertical level configuration explains about half of the observed bias between SLUCM and BEP (cf. Fig. A.3). Therefore, this bias is attributable to a mixture of denser vertical layers in the BEP scheme and a difference in physical parametrization.

Finally, biased temperatures can be an indicator of a bias in other diagnostics like PBL height, which we analyze next. The best-performing ensemble member with respect to 2 m temperature is the SLUCM, YSU, Noah-MP, and MM5 configuration. In December, the best configuration uses the BouLac scheme for PBL and MO as SLM scheme instead.

2.4.4 Planetary Boundary Layer Height

For PBL height, the model with the largest sensitivity (see Section 2.3.4) is the UCM. This is clearly visible in the Taylor density diagrams of PBL height, Fig. 2.13. Similarly to the case of 2 m temperature, there is a large difference of performance at urban vs rural stations. This is clearly visible by the large spread in KDEs especially in September. Table A.4 shows that the average MAB of all WRF simulations is larger at the more urban Stuttgart/Schnarrenberg station than at Idar/Oberstein.

Figure A.4 clearly illustrates the large seasonal dependency of PBL height. We clearly see the much lower diurnal cycle amplitude of the December PBL heights compared to the other months. In the simulations, we also see earlier PBL height growth and later teardown for July than for April, September, and December. This is accompanied by higher daytime and lower nighttime PBL heights. These features are even more exaggerated in the BEP than in the SLUCM simulations. During daytime there is a higher PBL height estimation for BEP than for SLUCM. We also see a full-day bias of BEP PBL heights in December. The

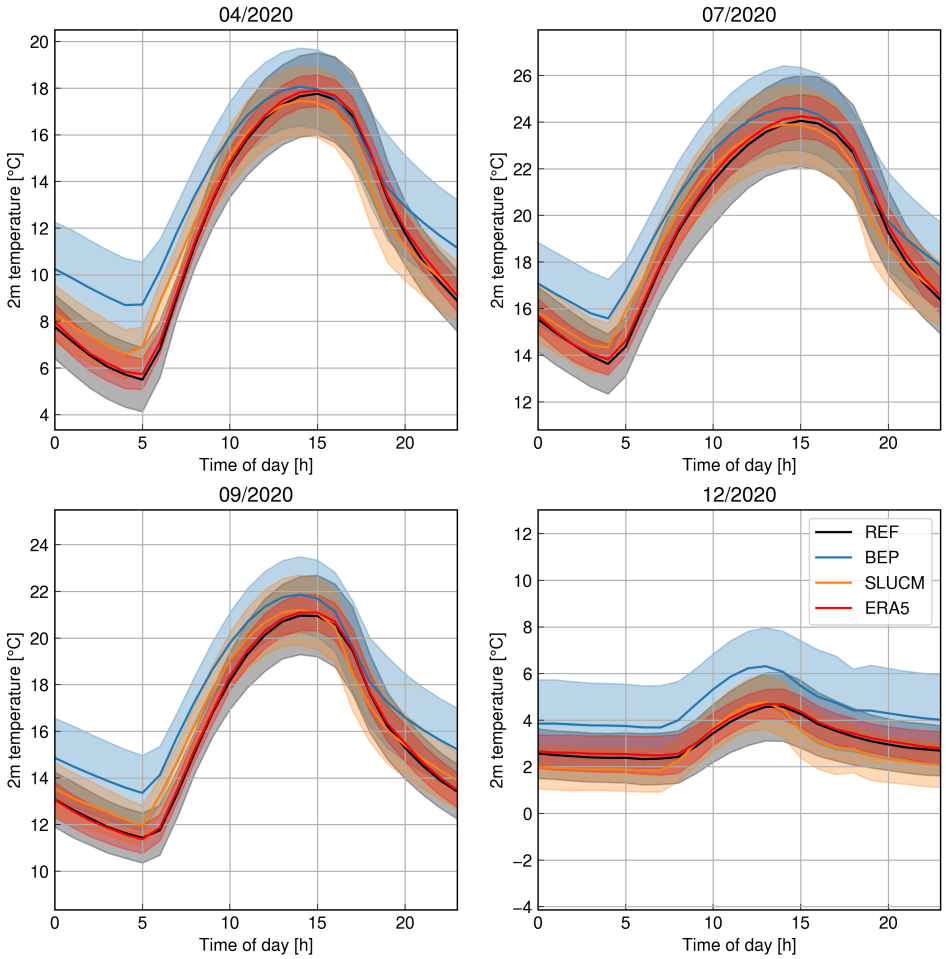


Figure 2.12: Diurnal cycle of 2 m temperature averaged over all stations. In black, the measurements from DWD stations, in red the ERA5 data, in orange the average over all simulations using SLUCM, and in blue the average over all simulations using BEP. Errors are one standard deviation of all diurnal cycles of each configuration and station. Figure taken from Pilz et al. (2026).

different PBL dynamics as well as the full-day bias in December are likely connected to the 2 m temperature misestimation. As we found in Pilz et al. (2026), "[this] is in line with the 2 m temperature bias between BEP and SLUCM, as warmer nighttime 2 m temperatures can lead to earlier and more effective PBL growth due to higher initial temperatures as well as deeper PBL mixing over the day".

For ERA5, the performance is similar between the two stations with smaller biases at Stuttgart/Schnarrenberg than at Idar/Oberstein except for July. The best-performing WRF ensemble member with respect to PBL height is the SLUCM, MYJ, Noah and MO configuration.

2.4.5 10m Wind Velocity

With regards to 10 m wind velocity, the model with the highest sensitivity (see Section 2.3.4) is the PBL model. A few observations spring to mind when looking at the Taylor density diagram in Fig. 2.14. Firstly, compared to for example 2 m temperature, the correlations of the 10 m wind velocities are generally smaller. We attribute this to the generally less pronounced diurnal cycle of wind velocities. Secondly, we find the MYJ scheme to outperform the BouLac and YSU schemes not only in terms of correlation. The MYJ scheme also has a smaller spread of KDE in relative standard deviation, indicating more consistent performance across stations.

Turning to the diurnal cycle in Fig. A.5, we find much smaller amplitudes in December compared to other months. However, we find generally larger uncertainties even in measurements which are caused by larger heterogeneity across stations compared to other variables. In general, there is an overestimation of 10 m wind velocities by the SLUCM scheme especially during the night. In contrast, for the BEP scheme daytime wind velocities are underestimated.

When looking at urban stations in particular, these aforementioned over- and underestimations by SLUCM and BEP become more pronounced (cf. Figs. A.6 and A.7). We hypothesize that this is the cause for the overestimation of 2 m temperatures by BEP. This is discussed further in Section 2.5. On average, though, the MAB of day and night time is the same (cf. Table A.6). Looking at Table A.4, we also see that the simulation performance is actually higher at urban than at rural stations for ERA5 and for the best WRF ensemble members. The best-performing WRF ensemble member with respect to 10 m wind velocity is the configuration using MYJ, Noah-MP, and MO without a clear preference between SLUCM and BEP.

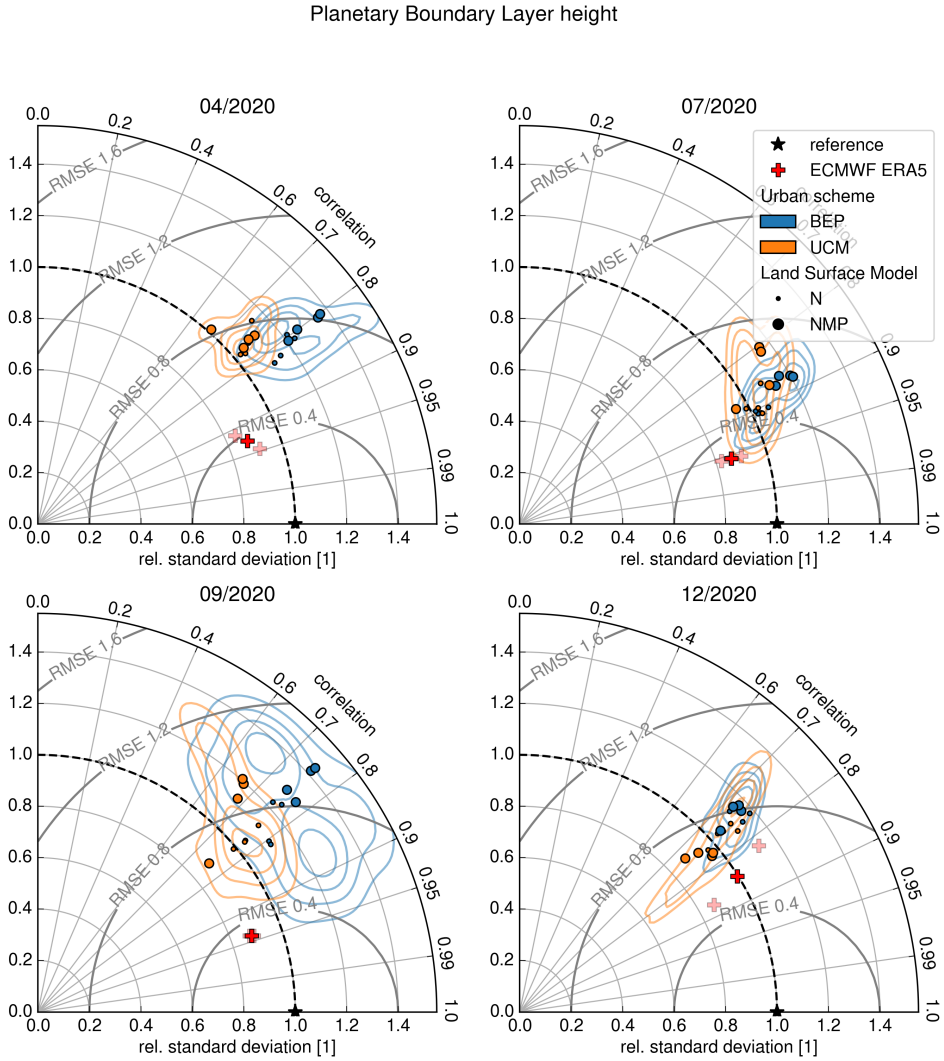


Figure 2.13: Taylor density diagram (see Section 2.3.4) of PBL height for all four simulated months of 2020. ERA5 data are plotted in red, configurations using SLUCM scheme in orange, and configurations using the BEP scheme in blue. Configurations are distinguished into using Noah or Noah-MP LSM by different markers. Behind each marker is the average over one simulation configuration averaged over all DWD stations. Spread of all configurations and stations using the respective UCM scheme is illustrated as KDEs except for ERA5, where there are only two values contributing to the average. Figure taken from Pilz et al. (2026).

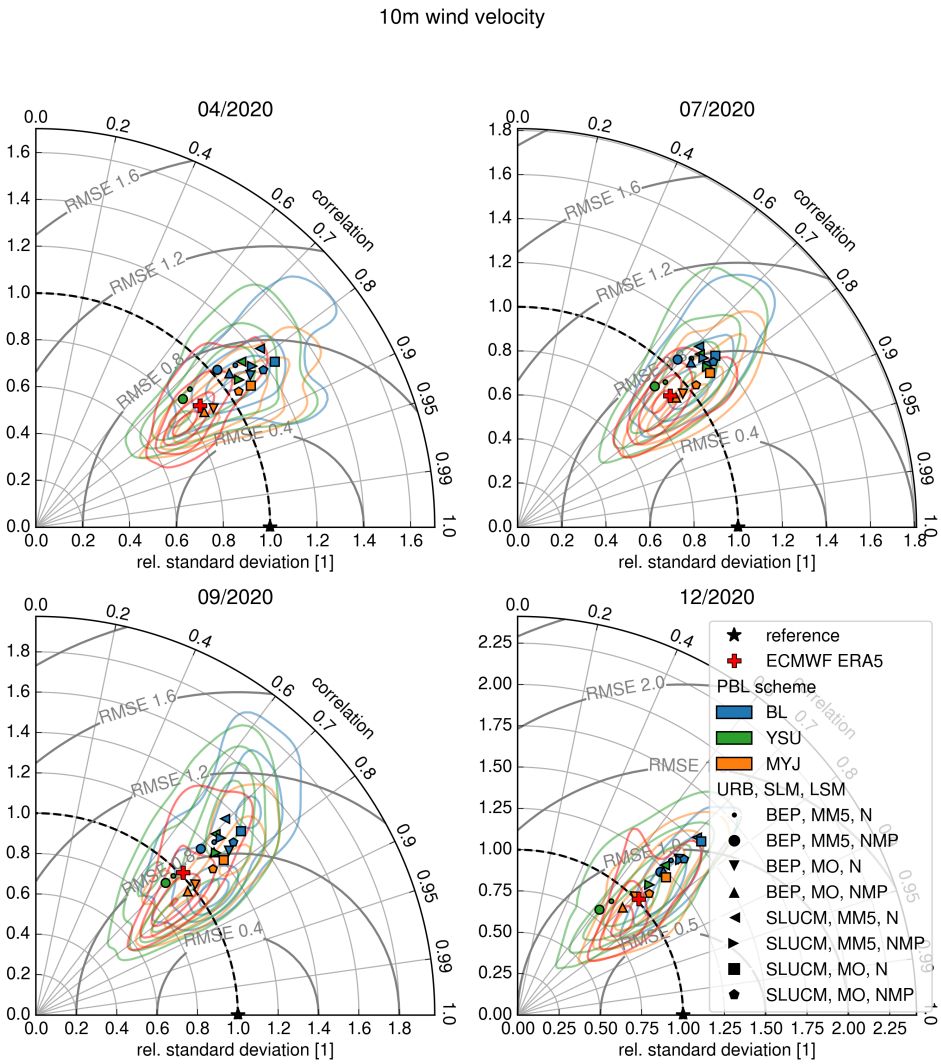


Figure 2.14: Taylor density diagram (see Section 2.3.4) of 10 m wind velocity for all four simulated months of 2020. ERA5 data are plotted in red, configurations using BouLac scheme in blue, configurations using the YSU scheme in green, and configurations using the MYJ scheme in orange. Configurations are distinguished by different markers. Behind each marker is the average over one simulation configuration averaged over all DWD stations. Spread of all configurations and stations using the respective PBL scheme is illustrated as KDEs. Figure taken from Pilz et al. (2026).

2.4.6 10m Wind Direction

We find the 10 m wind direction to be most sensitive (see Section 2.3.4) to choice of PBL scheme. Since wind direction is a circular quantity (see Jammalamadaka et al., 2001), the fundamental equations of the Taylor diagrams do not hold. There is also no inherent diurnal cycle in wind direction. However, we can still look at the ensemble spread of wind direction (as defined by standard deviation over all configurations averaged over all DWD stations). Figure A.8 clearly shows the wind direction ensemble spread decreases with increasing wind velocity. This can be explained by local influences which get estimated differently over all configurations outweighing synoptic drivers. In order to get more robust results, in the following we only look at situations with wind velocities above 1 m s^{-1}.

We see in Table A.4 that both WRF and ERA5 exhibit larger deviations over urban than over rural areas. This, too, can be explained by local influences like heterogeneity in materials or orography can influence wind deflection or heat fluxes. Table A.6 shows a better performance during daytime than nighttime, which is a corollary to wind velocities being higher during the day. We find the best-performing WRF ensemble member with respect to 10 m wind direction to be the SLUCM, MYJ, Noah, and MO configuration.

2.5 Discussion

This section is adapted from Pilz et al. (2026).

In this chapter, we have presented our analysis of high-resolution WRF simulations over a PUR in Europe. We have performed this analysis in order to optimize the setup of our WRF-Chem simulations of CO_2 and CO dispersion over German metropolitan areas. In this study, we have varied multiple different schemes for the land surface, surface layer, and planetary boundary layer models as well as urban parametrization. This large analysis spans all four seasons of the year 2020, investigating a whole month for each season. It allows us to draw conclusions on simulation performance with respect to different variables of interest and describe shortcomings in physical parametrizations.

Our ensemble consists of 16 different physics configurations of WRF and the simulation domain encompasses 19 meteorological stations of the DWD. This means that there are 304 combinations of configuration and station to evaluate. This would not be feasible without our introduction of Taylor density diagrams, augmenting traditional Taylor diagrams by a KDE. The KDE allows us to retain information about the spread of the data underlying the presented averages. We find the ensemble spread in our simulations to be dominated by station characteristics over physics configurations. This shows the importance of using as many meteorological stations as possible in these kinds of studies.

Overall, our WRF ensemble outperforms previous studies. The best-performing ensemble members (see ends of Section 2.4.3 to Section 2.4.6) have MAB values of 1.4 K to 1.7 K for 2 m temperature, 148 m to 249 m for PBL height, 0.8 m s^{-1} to 1.0 m s^{-1} for 10 m wind velocity, and 36.4° to 42.1° for wind direction. In urban areas, these become 1.2 K to 1.6 K, 168 m to 272 m, 0.8 m s^{-1} to 1.0 m s^{-1}, and 34.0° to 40.7°. The wind velocity MABs are lower than reported in Solbakken et al. (2021). Furthermore, all variables outperform the RMSE and correlation values reported in Ho et al. (2024, cf. Table A.5). We also find the performance of our WRF simulations to be much closer to ERA5 performance than the WRF simulations without observation nudging reported in Lian et al. (2018). Note, again, that the ERA5 performance is artificially inflated by it assimilating the data we compare against. For an actual comparison of WRF and ERA5 performance, independent measurement data would be necessary. However, our overall good performance compared to previous studies highlights the benefit of high-resolution simulations and high-quality input data.

In general, we find that WRF performance depends on both variable of interest and location. Regarding the variables' sensitivities, we find 2 m temperature and PBL height to be most sensitive to urban parametrization, while the 10 m wind variables are most sensitive to PBL schemes. This contrasts previous findings of Jänicke et al. (2017), who report larger sensitivity of 2 m temperature to PBL scheme selection than to urban parametrization.

Again, the WRF ensemble members' performance is dependent on variable of interest, which makes picking a best-performing configuration difficult. However, one configuration which performs well in general and especially when it comes to wind variables, which are of interest to us regarding CO_2 dispersion, is the SLUCM, MYJ, Noah-MP, and MO configuration.

Comparing the different configurations, we find a large difference in performance between BEP and SLUCM. This is not a new finding, as e.g. Ribeiro et al. (2021) have also reported similar performance differences. In Pilz et al. (2026), we have laid out three possible reasons as to the cause of this underestimation. This phenomenon could be caused by either or a combination of the following: The first possibility is a difference in diagnostic computation between the two schemes (cf. Section 2.1.2). Secondly, we found that the difference in vertical resolution explained about half of the bias. And finally, the default configuration of the urban parametrization within the BEP scheme could be at fault. This faulty configuration may overestimate surface friction in the urban scheme, which can be compounded over multiple vertical urban layers by the higher vertical resolution. The overestimation of surface friction in conjunction with radiation trapping in urban canyons could lead to higher 2 m temperatures and PBL height biases. This is in line with Ribeiro et al. (2021), who also find urban friction overestimation to be responsible for an underestimation of wind velocity.

This extensive analysis enables us to choose the physics configuration for WRF such that atmospheric transport is as accurate as possible for a dedicated trace gas run, the topic of the next chapter.

3 MACRO-2018: Modeled COx Concentrations

The previous chapter presented the sensitivity studies optimizing the meteorological modeling and laying the groundwork for choosing a favorable model configuration. This chapter focuses on the presentation of the MACRO-2018 dataset and its genesis. This dataset includes meteorology as well as CO_2 and CO concentrations at up to 1 km resolution for German metropolitan areas for the full year of 2018. It includes two separate simulations with different physics configurations as well as two different biogenic models each. First the emissions input data for the simulation is described in Section 3.1. Then, we describe the data used for boundary conditions of the simulation in Section 3.2 and the configuration of the simulation itself in Section 3.3. Then, in Section 3.4, we compare the simulated CO_2 and CO concentration values to measurements. We first give an overview over the measurements and their preprocessing (Sections 3.4.1 and 3.4.2), then evaluate the two different versions of biogenic emissions (Section 3.4.3) and continue with more general comparisons against measurements (Section 3.4.4). Finally, in Section 3.6 we compare its performance to previous studies and in Section 3.7 we discuss the results.

3.1 Emissions Data

In order to simulate the dispersion of the CO_2 and CO concentration fields, information about their fluxes are required. Here, we focus on the ones for CO_2. In general, these are divided into two categories, anthropogenic emissions and biogenic fluxes. In the following section, we discuss the emissions inventories which we use to generate the MACRO-2018 dataset.

3.1.1 Biogenic Emissions

To calculate biogenic fluxes, we use the Vegetation Photosynthesis and Respiration Model (VPRM), which is conceptualizes net biospheric fluxes (net ecosystem exchange (NEE)) as the difference between gross primary production (GPP) and ecosystem respiration. GPP

refers to the total amount of carbon fixed by an ecosystem into biomass through photosynthesis. Photosynthesis, in turn, is the process by which plants convert light energy, water, and carbon dioxide into carbohydrates, releasing oxygen as a byproduct. It occurs when sufficient nutrients, water, sunlight, and CO_2 are available. Rates of photosynthesis can decline under environmental stress conditions like drought, nutrient deficiency, and extreme temperatures.

To meet their energy demands, plants break down stored carbohydrates through a process known as autotrophic respiration. Autotrophic respiration results in CO_2 emission, however these are typically offset by higher rates of photosynthetic CO_2 uptake during the day. The other component of respiration is the heterotrophic (soil) respiration which is caused by microorganisms.

The underlying model of VPRM was comprehensively described in Glauch et al. (2025) and we paraphrase this description here. It consists of the following two equations for GPP and ecosystem respiration

$$\begin{aligned} \mathrm{GPP} &= \varepsilon \cdot \frac{1}{1+\mathrm{PAR}/\mathrm{PAR}_0} \cdot \mathrm{PAR} \cdot \mathrm{EVI} \\ \mathrm{R_{eco}} &= \alpha \cdot \max\left(\mathrm{T}, \mathrm{T_{low}}\right) + \beta \end{aligned}$$

with photosynthetically active radiation (PAR), enhanced vegetation index (EVI), and ε being the light-use efficiency. In the respiration model, $\mathrm{T_{low}}$ is a minimum soil temperature for winter. Furthermore, α, β, PAR_0 and λ are empirical parameters which have to be fitted for each vegetation type. The light-use efficiency ε is calculated as

$$\begin{aligned} \varepsilon &= \lambda \cdot \mathrm{T_{scale}} \cdot \mathrm{W_{scale}} \cdot \mathrm{P_{scale}} \\ \mathrm{T_{scale}} &= \frac{(\mathrm{T}-\mathrm{T_{min}})(\mathrm{T}-\mathrm{T_{max}})}{(\mathrm{T}-\mathrm{T_{min}})(\mathrm{T}-\mathrm{T_{max}}) - (\mathrm{T}-\mathrm{T_{opt}})^2} \\ \mathrm{W_{scale}} &= \begin{cases} \frac{\mathrm{LSWI}-\mathrm{LSWI_{min}}}{\mathrm{LSWI_{max}}-\mathrm{LSWI_{min}}}, & \text{grassland} \\ \frac{1+\mathrm{LSWI}}{1+\mathrm{LSWI_{max}}}, & \text{all other classes} \end{cases} \end{aligned}$$

with λ being a correction parameter, $\mathrm{P_{scale}}$ the vegetation type dependent effect of leaf age, $\mathrm{T_{scale}}$ the temperature dependence of photosynthesis and $\mathrm{W_{scale}}$ the canopy water content. $\mathrm{T_{min}}$, $\mathrm{T_{max}}$ and $\mathrm{T_{opt}}$ represent the literature-derived values of minimal, maximal, and optimal temperatures for photosynthesis and $\mathrm{W_{scale}}$ is dependent on the leaf surface water index (LSWI).

While VPRM is a comparably simple model, it is able to simulate the difference of GPP uptake and respiration fluxes, net ecosystem exchange (NEE), using few inputs. Indeed, the only meteorological input parameters needed are temperature and PAR, as all other

meteorological influences are proxied by EVI and LSWI measurements from satellites. These measurements have to be filtered and interpolated in order to make up for gaps in the data. The procedure of using the model for a specific location consists of two steps. Before the model can be applied, its parameters have to be fit for all land use classes. Flux tower measurements (and their surrounding land use) are used in order to constrain the α, β, PAR_0 and λ parameters. When evaluating the model for a specific location, the land use data for the location in question have be determined. Then, the respective fitted parameters are selected. Finally, for the times of interest, the temperature and solar irradiance data are used to evaluate the model. For the MACRO-2018 dataset we used two versions of the same model in order to evaluate model uncertainties.

Version 1 The first version of flux data was generated by Prof. Dr. Julia Marshall (DLR) for the CoCO2 project and uses Moderate-resolution Imaging Spectroradiometer (MODIS) data measured using the Terra satellite and Synergetic Land Cover Product (SYNMAP) land use data. Here, the 8-day reflectance product from Terra (MOD09A1, version 6) is used for EVI and LSWI, which is provided on 500 m to 1000 m pixels. These pixels are then gap-filled in time using a locally weighted scatter plot smoothing (LOWESS) filter. As meteorological products, the 2 m temperature and downwelling shortwave radiation are extracted from analysis and short-term forecast fields from the ECMWF IFS model. The parameters of this product are presented in Gerbig and Koch (2024).

Version 2 This newer version of flux data was generated by Dr. Theo Glauch (DLR; Glauch et al., 2025). For EVI and LSWI, it uses the 1-day MODIS products from both, Terra and Aqua satellites (MOD09GA, MYD09GA). As land use data input the 100 m resolved Copernicus Dynamic Land Cover Collection 3 data (Buchhorn et al., 2020) are used. To fit the model parameters, a two-step protocol is applied with the respiration term (α, β) being fit only for the nighttime data. After these parameters, the GPP term (PAR_0, λ) is fit using fixed α and β over the daytime data. Furthermore, the T_{min} and α parameters in the respiration function (cf. Section 3.1.1) have been changed in a way which makes the new model more temperature-sensitive. The effect of using the high-resolution land use data is that especially in areas of sparse biogenic influence, like cities, this version of VPRM exhibits much more activity. The temperature products used here are the 2 m temperature and downwelling shortwave radiation products from ERA5-Land (C3S, 2019) at 9 km resolution.

3.1.2 Anthropogenic Emissions

The best available estimates for anthropogenic CO_2 emissions and their locations come from bottom-up inventories. These inventories use two kinds of reporting to know where emissions

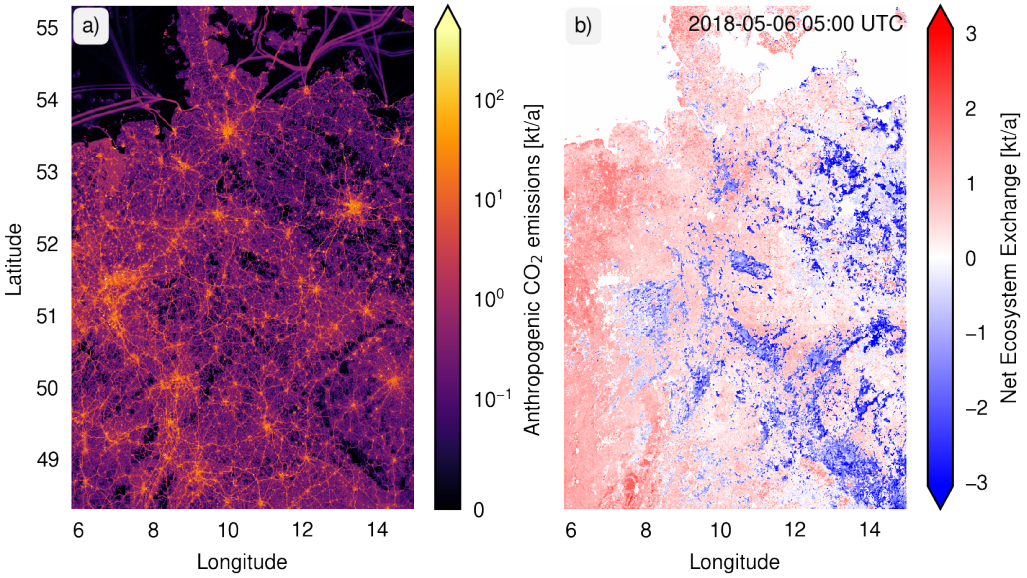


Figure 3.1: Overview of anthropogenic and biogenic emissions. The total CO_2 emissions of the Netherlands Organization for Applied Scientific Research (TNO) GHGco v4.1 inventory for the year 2018 on a logarithmic scale are displayed in Panel a). Panel b) displays the biogenic emissions from VPRM v1 of 6th of May, 2018 05:00 UTC scaled up to match units.

occur. First, they use country-level reporting (for example NIRs) required by supranational entities like the UNFCCC in the context of the Global Stocktake. Additionally, they use company-level reporting required by European law. The latter information is contained in databases like the German Pollutant Release and Transfer Register (PRTR) which is accessible via Thru.de and the European E-PRTR. They combine these data sources with proxies like nighttime lights for residential heating to distribute these emissions spatially.

We use the gridded TNO GHGco v4.1 inventory (Super et al., 2020). This inventory includes fossil- and biofuel emissions of CO_2, CH_4, CO, and NOx for the year of 2018. It is based on NIRs and reporting to European Monitoring and Evaluation Programme/Centre on Emission Inventories and Projections (EMEP/CEIP). For the year 2023, the NIR of 2025 reports a combined sectorial CO_2 emissions uncertainty of 7.6 % to 8.3 % (depending on method) for total German emissions (Günther et al., 2025). These emissions (grouped into Nomenclature for Reporting (NFR) sectors) are then gap-filled using the Greenhouse Gas – Air Pollution Interactions and Synergies (GAINS) and Emissions Database for Global Atmospheric Research (EDGAR) models. After being localized using NFR sector specific proxies and the temporal profiles being extracted, they are grouped into Gridded Nomenclature for Reporting (GNFR) sectors. These localized uncertainties on the scale of the simulation

domains will be considerably higher than the NIR uncertainties. In the end, all emissions are reported per GNFR sector and source type (point- or area source).

The TNO GHGco v4.1 inventory we use consists of a zoom and a non-zoom inventory. While the zoom inventory has a resolution of approx. $1 \times 1\,\text{km}$, the non-zoom inventory has a resolution of approx. $6 \times 6\,\text{km}$. And whereas the non-zoom part is available for all of Europe (30.025 °N, −29.95 °E to 71.975 °N, 59.950 °E), the zoom inventory only includes Germany, the Netherlands and Belgium completely. The zoom region's lower-left corner is at 47.004 °N, −1.992 °E (approximately at Nantes) and its upper right corner is at 55.996 °N, 18.992 °E (in the Baltic Sea at the longitude of Gdańsk). Figure 3.1 gives an overview over both the zoom inventory anthropogenic emissions and the biogenic emissions of VPRM v1.

3.1.3 Regridding

Regridding is the process of transferring data from one coordinate reference system (CRS) to another. Most inventory data including both our anthropogenic and biogenic emissions data are given on the WGS84 geographic CRS (latitude/longitude). However, our WRF simulations are defined on a projected CRS. We chose the Lambert Conformal Conic (LCC) projection, as it is widely used and can be easily adjusted to fit the desired simulation region. This means we have to transfer the emissions between these projections in order to build WRF input files.

The (non-rotated) LCC projection is an angle-preserving but not area-preserving (conformal) projection. It works by projecting points on the Earth's surface onto a cone's surface and then unrolling it. WRF uses a spherical datum for the WGS84 latitude/longitude coordinate system, so the cone is intersecting this sphere. In the non-rotated LCC projection, the cone's apex is above the north pole. Amongst the parameters of the projection are two "standard parallels", which are the latitudes at which the cone intersects the sphere (cf. Fig. 3.2). These two intersections lead to smaller projection errors over larger areas.

Because we want to regrid emissions fields, we care about the location and amount of emissions being conserved as much as possible between the projections. Tools like Earth System Modelling Framework (ESMF) and its derivatives `xESMF` and `ESMPy` exist and provide conservative interpolators. However, the documentation does not provide any information about whether they take the grid curvilinearity into account or approximate cells into squares. In order to improve the emissions localization by heeding the curvilinearity of the grid, we chose to implement our own regridding utility.

Our utility works similarly to the `emiproc` utility developed by Lionel et al. (2025). The common functionality is that each individual cell is modeled as a polygon with four corners. In our utility, each polygon can be assigned additional points on the edges which refine the

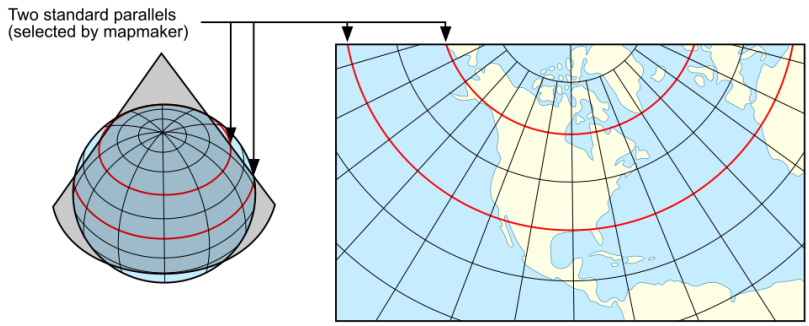


Figure 3.2: Visualization of the LCC projection after a USGS image. Two red "standard parallels" mark the intersection of the cone with the Earth's surface. Image from wikimedia (Mysid), part of public domain.

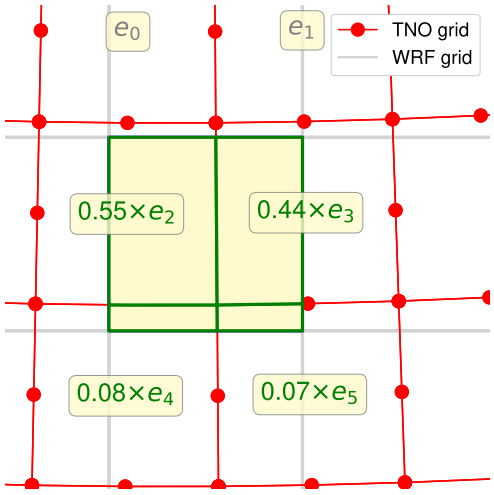


Figure 3.3: Visualization of the regridding from lat/lon grid to WRF model grid (LCC projection). It depicts how one of the WRF model grid cell emissions is computed using polygonal intersection from the lat/lon gridded emissions. The emissions in that cell are the sum of all four highlighted partial emissions.

curvilinear shape. These get then intersected with the polygons of the other projection using the `geopandas` library.

In preparatory studies, we found the curvilinearity of the grid to be too high and adjusted the LCC projection's standard latitudes in order to further reduce regridding errors. The `MAPFAC_M` variable in WRF, which describes the deviation of the projection from the sphere and is supposed to be around one, are for us between 0.999 and 1.023. As visible in Fig. 3.3, the lat/lon grid is barely curvilinear anymore in our projection.

Table 3.1: Errors due to regridding in all five focus regions including only Berlin. Compared are summed up TNO emissions with summed up wrfchemi files. Absolute and relative errors are given in kilotonnes and percent, respectively. The relative error is defined as (TNO − WRF)/TNO.

		Traffic	Area	Point	Biofuel	CO_2 Total	CO Total
Rhine-Neckar	kt	-161	-101	32	-64	-295	-2
	%	-0.97	-0.53	0.13	-0.72	-0.42	-0.8
Berlin	kt	-58	-35	50	-19	-62	-1
	%	-0.53	-0.25	0.13	-0.27	-0.09	-0.39
Berlin City	kt	-1	21	7	4	30	0
	%	-0.05	0.34	0.14	0.24	0.2	0.0
Rhine-Ruhr	kt	-151	-82	176	-58	-117	-1
	%	-0.86	-0.3	0.12	-0.47	-0.06	-0.22
Nuremberg	kt	-88	-44	1	-28	-158	-1
	%	-1.25	-0.68	0.12	-0.84	-0.88	-0.95
Munich	kt	-85	-71	6	-44	-195	-1
	%	-0.81	-0.86	0.13	-1.19	-0.72	-0.95

This utility was extensively tested using unit tests which assure a relative reallocation accuracy of total emissions in our test dataset of 0.01 ‰. However, upon reviewing the final regridded emissions, slightly higher inaccuracies were found (cf. Table 3.1). This demonstrates the need for not only unit- but integration tests. The emissions regridding was also qualitatively verified by reviewing all hourly emissions for all available emissions fields over the whole year. This was done in order to preclude obvious reallocation errors like shifts of the emissions over the year due to faulty grid definitions.

We evaluate the emissions regridding error for each of the focus domains. The error consists of two parts, which are caused by the regridding of area sources and the allocation of the regridded inventory to time and height. Since point sources are not subject to spatial regridding (other than bilinear interpolation), their errors are considerably smaller than the other sectors. Table 3.1 shows the regridding errors of anthropogenic emissions are below 1 % in all domains. However, the combined sectorial uncertainties of the German NIR which constitute an extreme lower bound for the localized emissions are considerably higher at 7.6 % to 8.3 % (Günther et al., 2025). Since our reallocation error is well below that, it is neglected in the following.

3.2 Initial and Boundary Conditions

Because our WRF simulations are not global, but only simulate a limited area, they need data providing initial and boundary conditions. This is required for both meteorological as well as CO_2 and CO concentration data. The ERA5 data we use for meteorology is presented in Section 2.2.2. Here we will only present the concentration data.

The CO_2 and CO concentration data used as initial and boundary conditions is the Copernicus Atmosphere Monitoring Service (CAMS) GHG short-term forecast data (experiment ID gqpe — IFS cycle CY43R1). These concentration fields assimilate satellite data from Greenhouse Gases Observing Satellite (GOSAT) measured by the Thermal and Near infrared Sensor for Carbon Observation (TANSO) instrument. The data used in this thesis was post-processed as described in Ho et al. (2024). The interpolation onto the 3D WRF grid was performed using `xESMF`. It was then integrated into our WRF coupled with Chemistry (WRF-Chem) model in the `CO2_BCK` and `CO_BCK` fields.

3.3 Simulation Configuration

The MACRO-2018 dataset contains two separate runs of WRF using the same input files but two different physics configurations which were identified with guidance from Chapter 2. Two simulation configurations create a mini-ensemble useful for a first-order estimation of physics-based model uncertainties. The configurations differ in the boundary layer schemes and surface layer schemes they use. One uses the YSU boundary layer scheme and MM5 surface layer scheme while the other uses MYJ and MO respectively. Both of them use the Noah-MP LSM. Further information on the differences between these physics schemes are presented in Chapter 2 and Pilz et al. (2026). However, they both use the BEP urban parametrization instead of the SLUCM parametrization as suggested by our studies presented in Chapter 2. This was chosen to allow for the higher vertical model resolution at low levels, which is not possible in SLUCM. This increased resolution is important in order to accommodate Lagrangian particle dispersion model (LPDM) use at urban building heights.

In general, the model setup is the same as presented in Chapter 2, however some adjustments were made in order to improve on it. Firstly, the land use bug mentioned in Section 2.2.2 was fixed, which improves the land use fidelity in urban areas. Secondly, the revised `URBPARM_LCZ.TBL` released in WRF v4.6.0 (issue `#1954`, PR `#1969`) was used in our simulations. Furthermore, the model was updated from v4.3.1 to v4.3.3 to include more bugfixes without major changes to its behavior. Lastly, its code was modified to accept the new input and concentration fields, which necessitated adjusting `Registry/registry.chem` and `chem/module_ghg_fluxes.F`.

One change compared to the prior chapter is that we couple the chemistry module to WRF (WRF-Chem; Grell et al., 2005). This module adds a lot of different schemes allowing for the online calculation of complex processes in atmospheric chemistry. We use it in tracer mode, meaning that all chemistry for both CO_2 and CO is turned off and they are only advected with the meteorology. While CO_2 barely has any relevant chemistry on the time scales we are evaluating, CO is a precursor in surface-level O_3 production and a product of CH_4 and NMVOC oxidation. However, due to the large computational demand explicitly modeling this chemistry would add and the large existing uncertainties in the CO inventory, we decided to forgo the chemical modeling. This is in line with previous studies like Callewaert et al. (2022) and Lama et al. (2022).

The model was run on the DKRZ supercomputer Levante using 156 processors per configuration. The run for the MYJ configuration took 3 days, 17 hours for spin-up and 47 days, 17 hours for the simulation. The YSU configuration took 3 days, 17 hours for spin-up and 43 days, 14 hours for the simulation. The integrity of the run was verified in multiple ways. Firstly, the CO_2 and CO field continuity was assured by scripts at every reinitialization. Secondly, a qualitative data integrity assurance was performed using a self-built dashboard using the Python package `panel` (see Fig. 3.4). Here, all domains were plotted in parallel. The variables were separated into two groups, 2D verification and 3D verification. The 2D verification group included `T2`, `U10`, and `V10` as well as the lowest levels of `CO_TOTAL`, `CO2_TOTAL`, `CO2_TOTAL_V2`, and `CO2_ANTHRO`. The fields of the 2D verification group were verified for every hour of the simulation. In order to prevent errors around the reinitializations, the 3D verification group was used. It consisted of `T`, `U`, `V`, as well as all the previously mentioned CO_2 and CO fields. Here, at least three vertical levels were verified for all hours of every Sunday and Monday (before and after the ERA5 reinitialization) in the simulation.

3.4 Evaluation

In order to evaluate the simulation output, we use high-precition in-situ concentration measurements. First, Section 3.4.1 describes how we pre-process and filter the measurement data. Then, Section 3.4.2 gives an overview over the processed measurement data. We start with evaluating the different VPRM versions in Section 3.4.3 and choose one with which to perform the evaluation. Then, in Section 3.4.4, we compare our simulation data to the measurements. And finally, we evaluate the realism of our dataset in Section 3.4.5. The length of our dataset of one year, the size of our study domain (Central Europe) and the two physics schemes as well as the two biogenic models in each present a unique opportunity for model evaluation supporting further developments.

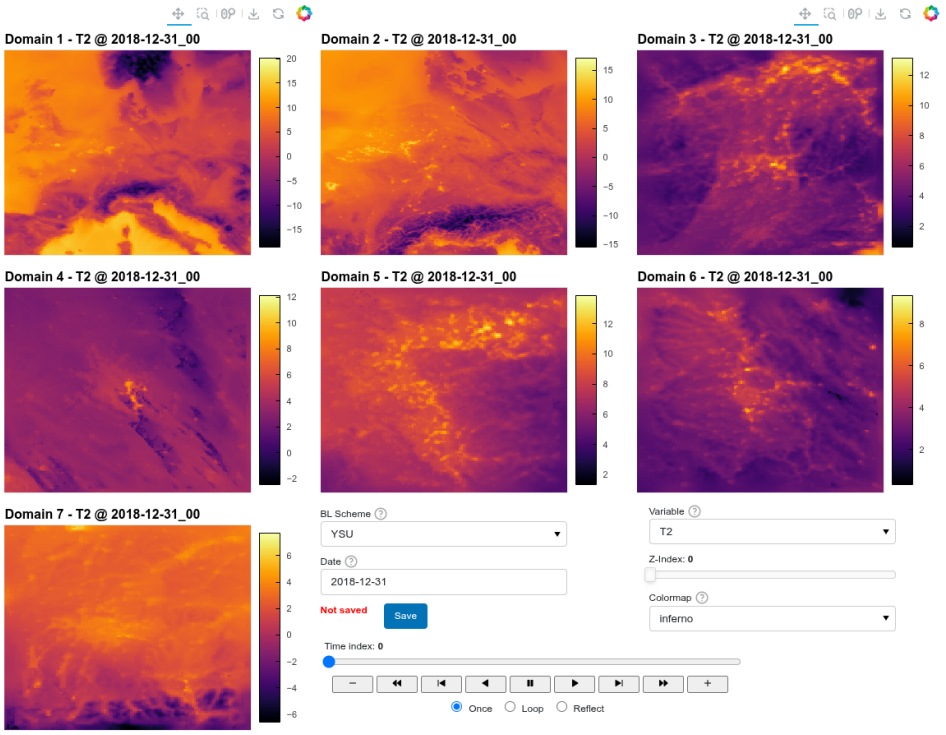


Figure 3.4: Screenshot of dashboard used for MACRO-2018 simulation verification. The panels in this dashboard display all seven domains at once. Menu enables selection of simulation output (BL Scheme) and date. Slider selects individual hours within the day and can start parallel playback of all domains. Selectors on the right determine content of panels. Variable of interest, simulation height level (Z-Index) and color map can be selected. Additional 'Save' button saves day as already verified which marks it in 'date' selector. Dashboard features include parallel preload of next day in order to reduce load times.

3.4.1 Measurement Data

First, we present the measurement data we compare our sampled simulations against. The data we use are a European version of the NOAA GLOBALVIEW Obspack data for CO_2 (ICOS RI et al., 2024) and the GLOBALVIEWplus-like Obspack product from NOAA (Schuldt et al., 2025) for CO. These Obspack datasets are a collection of hourly averaged concentration measurement data from Integrated Carbon Observation System (ICOS) and non-ICOS stations. They are measured by high-precision cavity ring-down spectrometer (CRDS) instruments with average measurement precisions of below 0.1 ppm for CO_2 and below 2 ppb for CO. As they are so low, these uncertainties will be disregarded in the following. The data is delivered with auxiliary variables used for quality assurance like the number of data points the hourly averages are computed from and a quality control flag. In order to improve representativeness of the data, we filter the number of measurements used for the average to be four or more. Also the quality control flag is filtered to `Flag 'O': data correct after manual quality control`. From our initial dataset of 40 stations (98 inlets), this excludes 8 stations (19 inlets) completely.

In order to compare over the whole simulation period, we filter the remaining inlets so that they have at least half of all possible measurements for each month. We also only keep inlets with a height of at least 10 m. This means that for CO_2 20 stations (52 inlets) remain while 14 stations (27 inlets) only have valid data for parts of the year. Two stations (`SAC` and `HFD`) appear in both the full-year and the part-year dataset because only some inlets do not provide full-year coverage. At the `BRM` station, one inlet does not provide any data at all and is thus excluded. For CO, 9 stations (24 inlets) remain while 6 inlets only provide partial data. These are the lowest inlets at `KRE` and `UTO`, the second lowest at `SAC` and inlets two to four at `LIN`. We continue our evaluation with the full-year dataset containing 20 stations (52 inlets) for CO_2 as well as 9 stations (24 inlets) for CO. An overview of the stations and their data coverage is to be found in Panel b) of Fig. 3.5.

3.4.2 Measurement Data Overview

First of all, we give an overview over the 54 inlets at 20 stations in our dataset. Of these 20 stations, five are surface stations with only one inlet while the other 15 are tower stations with two or more inlets. However, as we can see in Panel b) of Fig. 3.5, not all of the stations have valid data for all inlets over the whole year.

In order to extract comparable simulation data, we sample the MACRO-2018 dataset at the positions where the stations are measuring. The general problem when it comes to the question of where to interpolate vertically is that our WRF simulations systematically underestimate peaks in topography. This underestimation is caused by the simulation grid

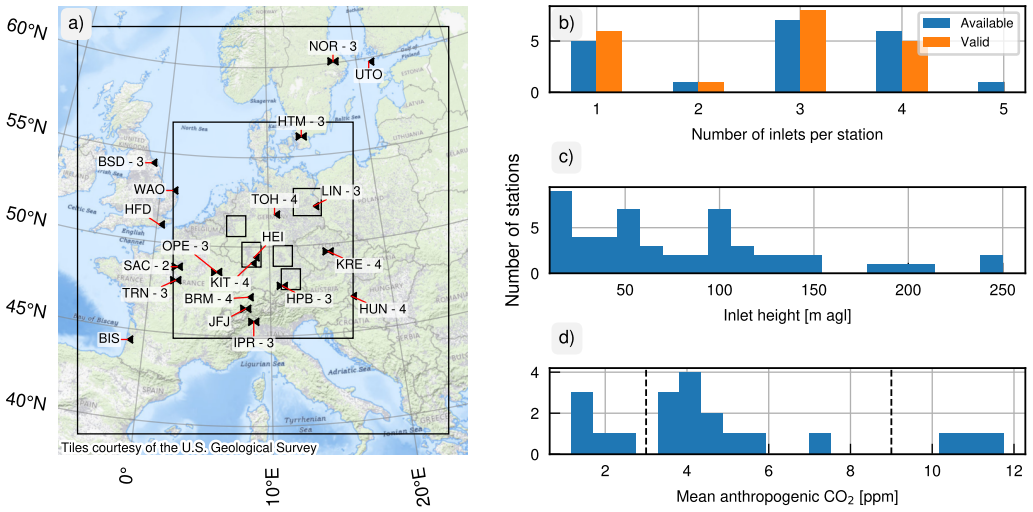


Figure 3.5: Overview over the measurement data uses in this study. Panel a) shows the locations of the Obspack stations and the number of available inlets for towers. Stations measuring CO_2 are marked with a marker pointing to the left. Additionally, CO is measured at all station which are marked with a marker pointed to the right. Panel b) compares the number of inlets per station before and after validation (cf. Section 3.4.1). Panel c) illustrates the mixed inlet heights between stations. Lastly, the dashed lines in Panel d) show the 3 and 9 ppm cutoff values used to categorize stations into 'remote', 'rural', and 'urban'. Map data copyrighted by OpenStreetMap contributors and available from openstreetmap.org.

resolution, onto which the input topography gets averaged. This means there are two ways of interpolating the simulation data to best fit the measurements. We can either interpolate it onto height above sea level or height above ground level. Using the height above sea level interpolation would be more accurate for remote stations on top of mountains. This is because they see more of the background and less surface-level emissions which would be overestimated if we were to sample above ground level. On the other hand, urban stations may be systematically underestimating concentrations since the sampling would be too high above the ground. Since metropolitan areas are our focus regions and we are interested in anthropogenic emissions, we choose the interpolation to height above ground level. We perform this interpolation of the simulation data onto inlet heights (above ground level) using the `xgcm` and `xwrf` utilities (Abernathey et al., 2025; Pilz et al., 2025b).

As there is no consistent station characterization given by the dataset provider, we use the MACRO-2018 data to generate one. We classify the stations into the categories: 'urban', 'rural', and 'remote' based on the signals they measure. It is important to note that these classifications are only defined relative to other stations in this dataset.

The main metric we use to classify stations this way is the average anthropogenic CO_2 concentration. For tower stations we use the maximum of this metric over all heights. We use averages instead of medians as we want a genuine integration of the signal including, amongst others, peaks from strong point sources. Panel d) in Fig. 3.5 demonstrates our stations splitting into three groups. The 'urban' stations have average anthropogenic signals over 9 ppm and are `KIT`, `HEI` and `SAC`. They are also the stations with the highest CO concentrations. While these signals do not need to be from urban areas but could also originate from industrial sources, we still call this category of stations 'urban'. The lowest signals are at the 'remote' stations with average anthropogenic signals under 3 ppm (`JFJ`, `HTM`, `BIS`, `UTO`, and `NOR`). These stations are also 5 of the 6 stations with the lowest CO concentrations. The rest of the stations fall in between. There are two stations, `TRN` and `IPR`, which are located close to urban areas (within max. 50 km) but do not see a lot of urban signal based on the MACRO-2018 dataset. These are classified as 'rural' in this case.

3.4.3 Evaluation of VPRM Versions

Now with a consistent dataset, we evaluate the performance of the two different VPRM versions in our data, which are presented in Section 3.1.1. This dataset provides a unique opportunity to investigate concentrations generated by the two different biogenic emissions models on a nearly continental scale and over a whole year, as they are usually only evaluated against flux tower data or over smaller simulation domains. The results of our evaluation directly support the further development of VPRM v2 by Dr. Theo Glauch.

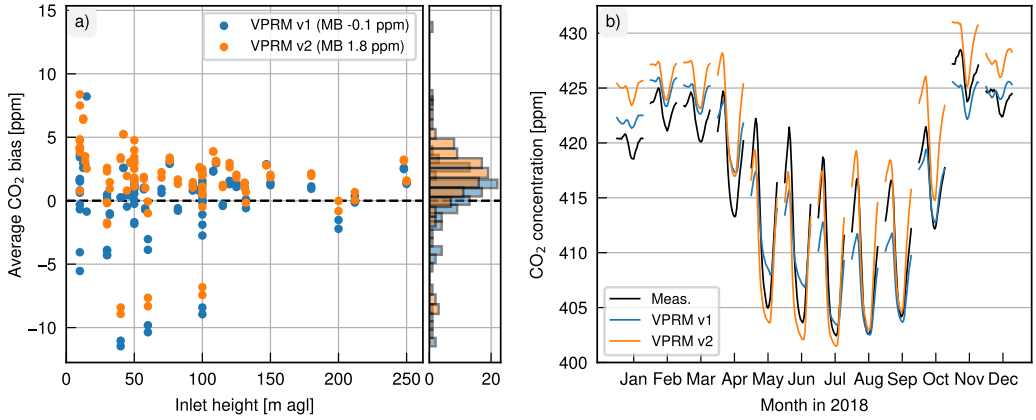


Figure 3.6: Comparison of VPRM version performance. Panel a) illustrates the average CO_2 bias over the inlet heights. Here, we clearly see the larger mean bias of VPRM v2. Panel b) displays the monthly diurnal cycle of total CO_2 concentrations using the different VPRM versions. Again, the positive bias of the VPRM v2 is visible especially in winter. However, it is good at capturing the diurnal cycle amplitude across the year and especially in summer, VPRM v2 has a higher fidelity than VPRM v1.

First we assess aggregate statistics before looking into the details of the performance. Over all 54 inlets (20 stations), we find a MAB for VPRM v1 of 5.2 ppm and for VPRM v2 of 6.4 ppm. This is mostly due to a higher mean bias in the VPRM v2 of 1.8 ppm vs −0.1 ppm for VPRM v1. We can see this in the overview in Panel a) of Fig. 3.6.

Panel b) of Fig. 3.6 displays the monthly diurnal cycle average over all stations. Here, we can see structural differences between the VPRM versions. Indeed, VPRM version 2 consistently overestimates winter and nighttime CO_2 (except for May, June, and July). However, this version captures the diurnal cycle amplitude more accurately. While it is consistently underestimated by VPRM v1 with a MAB of 6.0 ppm (mean bias (MB) −5.9 ppm), version 2 has a MAB of 4.1 ppm (MB 2.0 ppm; cf. Table 3.2). We only see minor differences between the two boundary layer schemes used. For differences between the VPRM versions, please refer to Section 3.1.1.

There are two possible sources of the performance difference of the simulated CO_2 concentrations between VPRM version 1 and 2. One possibility is the PBL height being underestimated by the model over the winter months, which overestimates the biogenic fluxes in general. However, the sensitivity studies performed in Chapter 2 and especially Fig. A.4 rather suggest a PBL height overestimation of the model. Additionally, our data includes two different PBL schemes and both have a similar mean bias (cf. Table 3.2). While not a perfect error estimate, this increases our confidence that the model PBL height estimation is not the problem.

Table 3.2: Overview of cumulative statistics of total CO_2 concentration using the different versions of VPRM. These statistics are averaged over all inlets.

	VPRM v1		VPRM v2	
	YSU	MYJ	YSU	MYJ
Mean Bias	−0.1	−0.2	1.8	1.8
MAB	5.1	5.2	6.2	6.3
RMSE	8.0	8.4	9.0	9.7
Diurnal Amplitude MB	−5.5	−5.4	1.8	2.1
Diurnal Amplitude MAB	5.7	5.4	3.8	3.8

The more likely explanation for these differences appears to point to the biogenic emissions models themselves. Changes to the respiration function, which increase the night- and winter time respiration, affect the flux estimates. These adjustments, however, are based on flux tower data, and since the representativeness of the flux towers across Europe varies, the VPRM fluxes in some regions are highly uncertain (Glauch et al., 2025). This leads us to conclude that the extrapolation errors from the flux towers where the parameters were fitted at to the whole of Europe is the primary source of the increased mean bias in VPRM v2 compared to v1. After communication with the provider of the VPRM v2 dataset, Dr. Theo Glauch, we continue our evaluation for the rest of this chapter with the VPRM version 1 data as it demonstrates higher average performance.

3.4.4 Comparison to Measurements

In this section, we compare the total simulated CO_2 concentrations (using the VPRM v1 biogenic emissions) to the Obspack measurements. The metrics we use are presented in Section 2.4.2.

First, we look at the overall data before going into more detail. In Fig. 3.7, Panels a) and b) show a 2D histogram of the simulated CO_2 and CO concentrations at all inlets against the measured concentrations. Across all inlets, for CO_2 concentrations we find a Pearson-R correlation of 0.76 between measurements and simulation while it is 0.61 for CO. In general, lower CO_2 and CO concentrations are simulated more accurately, with differences from the 1:1 line (grey) increasing for higher concentrations. There are both overestimations and underestimations of concentrations, with underestimations being more prevalent especially at higher concentrations. This can be attributed to high concentrations occurring more often at nighttime than at daytime and PBL height estimation at night being more difficult. They also occur in situations where the sensors measure large anthropogenic influences (especially for CO, as it has no biogenic component in our simulation). Here also emissions inventory

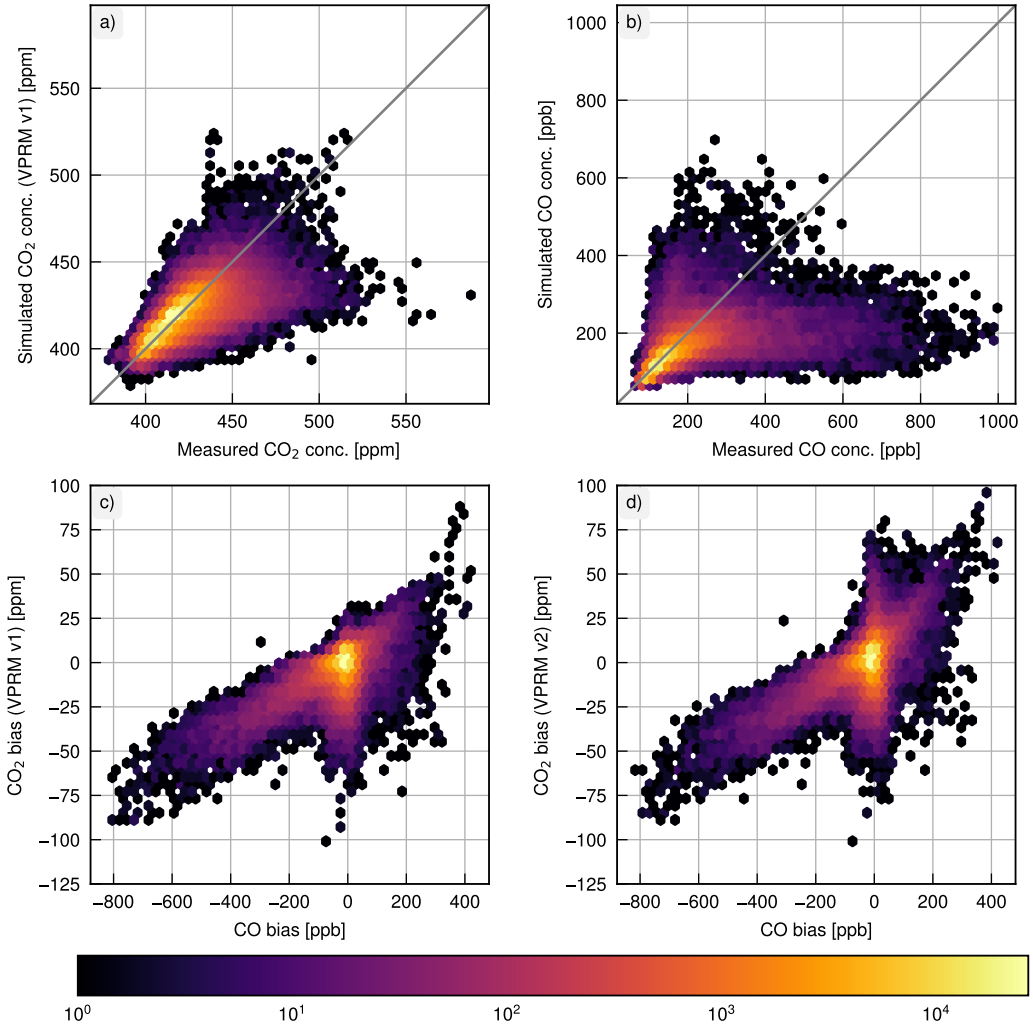


Figure 3.7: Histogram of simulated and measured CO_2 and CO incl. biases. Upper Panels (a, b) display 2D histogram of simulated CO_2 and CO concentrations compared to measured ones. Lower Panels (c, d) display CO_2 biases over CO biases for VPRM v1 and v2. Color scale for both is logarithmic. In the comparison of CO_2 and CO biases, two correlation regimes are apparent. A linear one and one of CO_2 biases without CO biases.

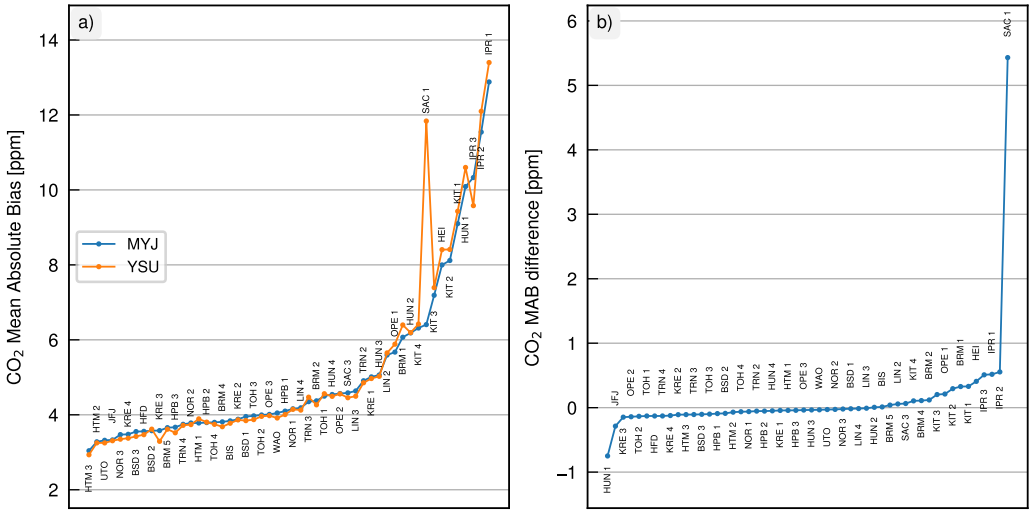


Figure 3.8: Visualization of MAB of all inlets in our dataset. Panel a) displays MYJ and YSU MABs with inlets sorted by MYJ MAB. The difference in MAB between the two versions of the dataset is displayed in Panel b). Clearly visible is the outlier inlet which is the lowest in SAC.

errors in both space and time contribute to concentration misestimations. This very high anthropogenic influence compared to CO_2 is why the CO concentration correlation is lower.

Looking at correlations between CO_2 and CO biases in Panels c) and d) of Fig. 3.7, we find two different correlation regimes. On one side there is a mostly linear relationship between the biases of the two gases, hinting at common underlying factors like emissions misestimation in the inventory or boundary layer simulation issues. On the other hand, there is a second regime of CO_2 biases without accompanying CO biases. This regime is more pronounced for VPRM v2 than v1. Since these CO_2 biases are not associated with CO biases, they are more likely caused by misestimations in biogenic CO_2 fluxes, especially of respiration, than by misestimations in PBL height. We can also see this in Fig. A.9, which shows the same histogram with measurements filtered for the stable PBL development phase (find definition below). Here, the second correlation regime is also much larger for VPRM v2 than v1.

The performance metrics for MACRO-2018 at the Obspack stations for CO_2 and CO are to be found in Tables A.8 and A.9. The inlet-wise performance can be found in Tables A.10 and A.11. First, we compare the two physics configurations we used for the simulations. They differ in choice of PBL and surface layer schemes. One used MYJ and MO, while the other used YSU and MM5. Overall, the differences in CO_2 concentration MABs between the two configurations are marginal, however one inlet exhibits a very large difference (cf. Fig. 3.8). This inlet is the lowest one (15 m agl) at the Saclay station (SAC). While there

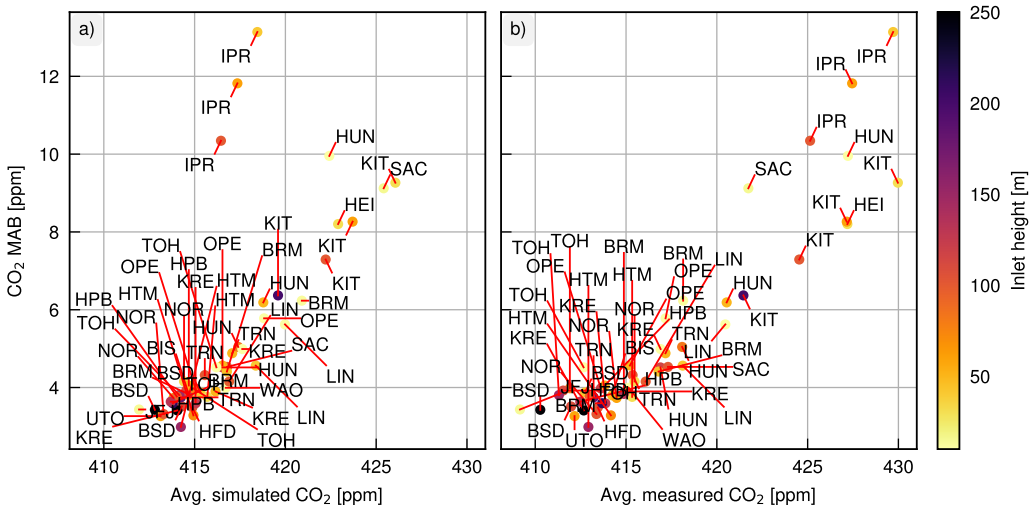


Figure 3.9: Correlation of CO_2 MAB with average CO_2 signal. Panel a) displays correlation of CO_2 MAB at each inlet with average simulated signal and Panel b) displays the same with average measured signal. The CO_2 MABs show correlations of 0.72 and 0.92 with simulated and measured signal. For the simulated signal, the IPR station deviates from the clear linear relationship with the other inlets. This points to systematic differences at that station in particular.

are differences in MAB values of −0.7 ppm to 0.6 ppm (average: −0.02 ppm) for all other inlets, the lowest inlet at Saclay exhibits a 5.4 ppm higher MAB for the YSU configuration than for the MYJ one. These differences likely result from a combination of factors from the measurement height not accurately representing the surrounding area to limitations of the simulation. In particular, the YSU simulation tends to underestimate the PBL height compared to MYJ, which in combination with the high emissions in Paris may amplify the discrepancy. This shows that users of the dataset have to investigate which configuration is suitable for their location and use case.

In general, the CO_2 performance at stations which are more remote is higher than at stations which are more urban. This is to be expected, as urban areas are heterogeneous in land use and emissions, complicating both meteorological and emissions modeling. Looking at correlations between MAB and other variables, the highest correlation is with the average signal strength itself. As visible in both panels of Fig. 3.9, the MAB of most stations correlates very tightly to the measured signal strength. This is either caused by misestimations of PBL height or problems with the biogenic respiration model. However, there is one stations which stands out.

This station is IPR (Ispra, Italy), where the simulated CO_2 underestimates the measurements by 10 ppm on average (cf. Table A.8). It is located next to Lago Maggiore, approximately 60 km from the next major city, Milan. There are several factors which can

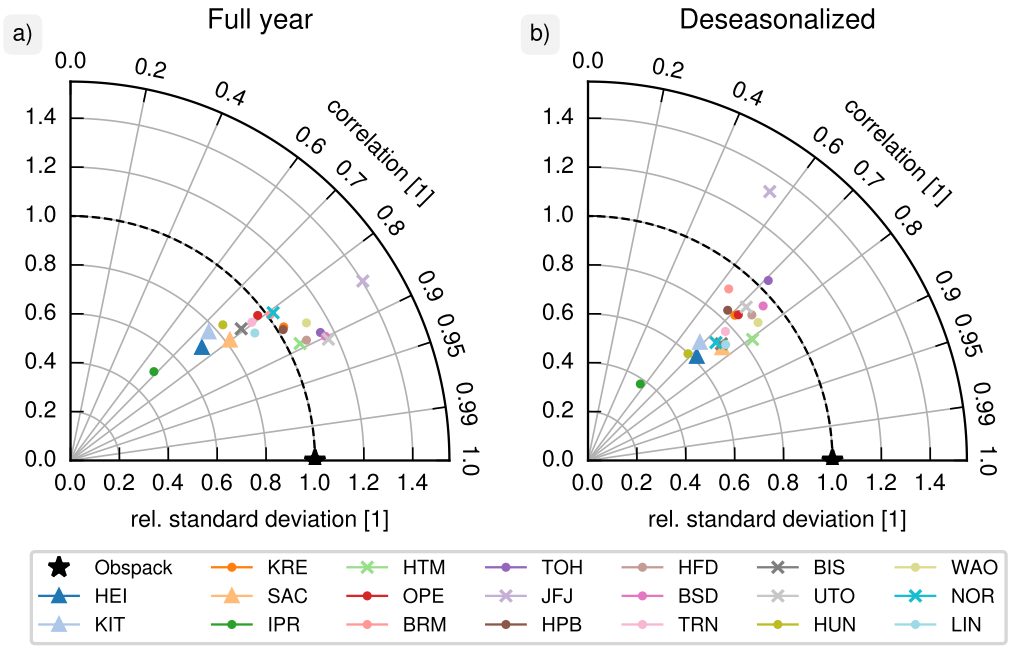


Figure 3.10: Taylor diagrams of VPRM v1 and MYJ configuration. The statistics in Panel a) are computed over the whole year whereas they are deseasonalized by computing monthly statistics and averaging them in Panel b). Triangle markers are for urban stations, dots for rural ones and 'x' markers are for remote stations.

explain the outstanding deviations at this station. The first is that the emissions inventories could be incorrect, e.g. by under-reporting of companies, and there are actually be more emissions in reality than in the TNO inventory. Indeed, there are two cement factories at 5.1 and 8.4 km distance whose emissions impact local measurements. Furthermore, as point sources, these can also be underestimated by virtue of the simulation resolution.

The resolution of our simulation is also able to affect our results in another way. As we resolve the simulation only at 5 km, the Alps, which the IPR station is nestled against, appear smoothed out. For example Monte Leone, the highest mountain of the Lepontine Alps, which are the closest mountain ridge to Ispra, stands at 3553 m, while in our simulation the elevation at the same location is only 2317 m. This can cause the accumulation of CO_2 south of the Alps to be severely underestimated, as air-mass (and pollution) trapping by Alpine topography is a well-known phenomenon (Diémoz et al., 2019). Users of the MACRO-2018 dataset have to be careful using the data around regions of high topographic complexity because of the limited resolution.

Looking more at the bias-corrected performance, we turn to the Taylor diagrams of our simulation in Fig. 3.10. Here, we have to deseasonalize the data because the base assumption

of the Taylor diagrams is normally distributed data. We remove the seasonal cycle by computing standard deviation and correlation per month and then averaging them. As expected, the correlation performance drops for the deseasonalized data, since the seasonal cycle is not artificially boosting it. While the three urban stations do cluster together, rural and remote stations do not show large differences between them. The previously discussed outlier, IPR, is visible and additionally the JFJ (Jungfraujoch) station shows large deviations.

The JFJ station is the highest measurement station of the dataset with an elevation above sea level of 3570 m. It is an outlier in this case as the simulation overestimates the standard deviation of the time series. The probable cause of this is the sampling of the MACRO-2018 data being performed above ground level. This works well for low-elevation stations, as it takes care of deviations from the topography due to spatial resolution, however at high elevations it breaks down due to the overwhelming smoothing effect of resolution on topography. Indeed, while the station's inlet in reality is at an elevation of 3583.9 m, we sample our simulation at 2753.8 m above sea level.

The CO performance metrics are listed in Table A.9. The MACRO-2018 datasets exhibits good correlations to measurements of above 0.80 except for the stations IPR, SAC, and JFJ. The performance problems at these three stations were already discussed above for CO_2 and the above arguments also apply for CO. With average CO concentrations over the whole year of 135 ppb to 160 ppb for the other stations, we find comparatively small mean biases of around 10 %. Most biases except for JFJ and SAC are negative, which could be caused by model transport, inventory issues, or the missing secondary production by photochemistry. Indeed, Fisher et al. (2017) and Huijnen et al. (2019) show that secondary production is a relevant source of CO concentration, which could explain some of our biases.

The MAB of all stations except IPR and SAC are 15 % or lower, which is a good agreement. However, the KRE station exhibits a rather high MAB of 23 ppb. This MAB is driven by the large mean bias at that station of −17 ppb. Apparently there is much more measured than simulated CO at this station, which is why the relative MAB is only 15 %. The largest source of CO in the vicinity of the station is the E50 highway, which could be too smoothed out in our simulations to have its influence on the measurements accurately modeled. Another possibility is that the CO emissions in the Czech Republic in general could be underestimated in the TNO GHGco v4.1 inventory.

Analysis across station types and times of day One characteristic of meteorological modeling, which is important to emissions inversions, is the reduction of transport error. In order to achieve this, inverters usually choose to only use afternoon measurements, as the PBL height is relatively stable, thus leading to comparatively low simulation errors. However, there is no consensus in the community as to what amounts to 'stable' conditions (McKain

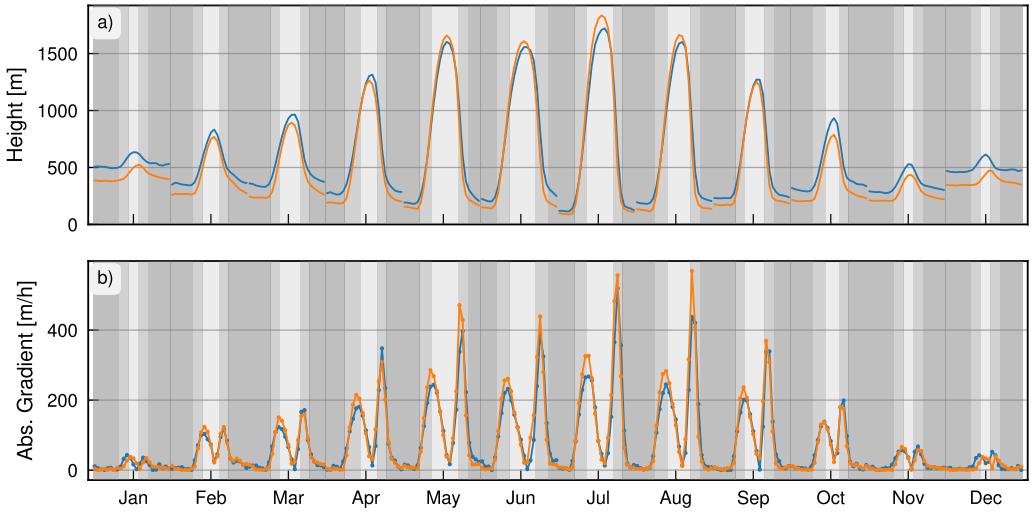


Figure 3.11: Diurnal cycle of PBL heights (a) and their absolute gradients (b) averaged over all stations in the UTC+1 timezone. Two very distinct peaks in the absolute gradient are visible in Panel b), one from PBL buildup in the morning and the other one from its teardown. Grey zones are in order of darkness, descending: Night, transition (buildup and teardown) and stable.

et al., 2015; Lauvaux et al., 2016; Miles et al., 2017; Kunik et al., 2019; Lian et al., 2024; Monteiro et al., 2024). It is not clear whether daylight savings time (DST) or local standard time (LST) is supposed to be used and all definitions used in the aforementioned papers are static over the whole year. So, we use the wealth of data of the MACRO-2018 dataset in order to define this more clearly.

As described in Section 2.1.1, the mixing layer has two transition phases. It grows in the morning and then collapses in the evening. We identify these two phases by the gradient in PBL height. Here, we use the native WRF PBL height diagnostics to do this in contrast to Chapter 2. First, we select all stations which are in timezone UTC+1 and calculate the absolute average gradient in PBL height across all months. The result is visible in Panel b) of Fig. 3.11. We now only use the data from the MYJ simulation, as the only difference between MYJ and YSU is the shift of YSU peaks one hour later in January and one hour earlier in August. We then define the stable phase of the PBL development to be from one hour after the morning gradient peak to one hour before the evening gradient peak. Secondly, we identify the nighttime to be the last (in the mornings) and first (evening) mixing layer height gradient to be below 25 % of the morning peak. The transition phases are then the buildup phase between end of nighttime and the stable phase as well as the teardown phase between the end of the stable phase and the beginning of nighttime.

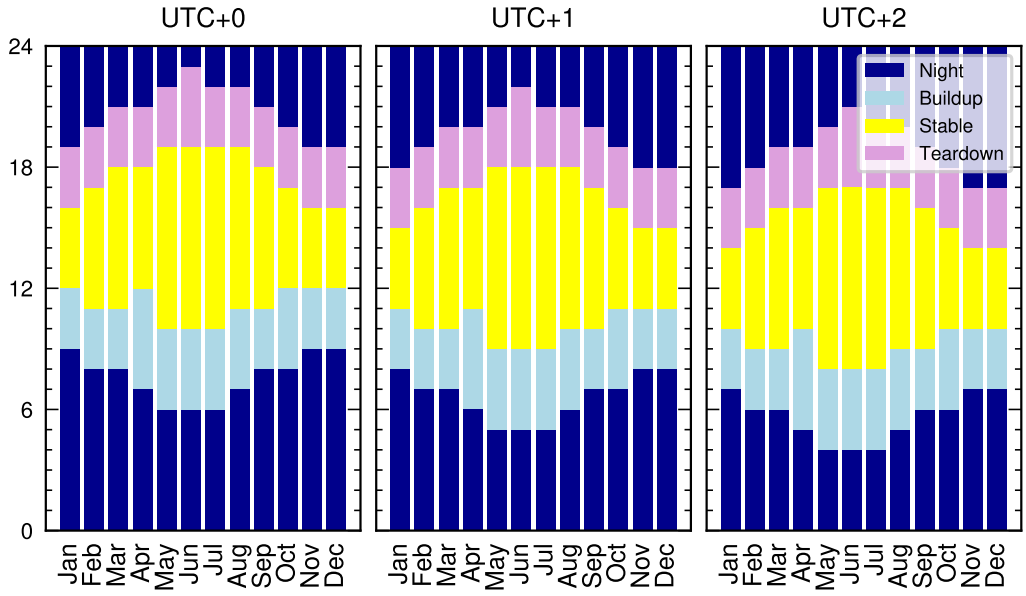


Figure 3.12: Exact definition of PBL height development phases for stations in timezones UTC+0, UTC+1, and UTC+2 by month. The phase definitions were created using the UTC+1 stations and then transferred to the UTC+0 and UTC+2 stations.

We then assign these to all stations shifted according to station timezone. The definition of the phases across the three timezones (UTC+0, +1 and +2) is provided in Fig. 3.12. Please note that we moved the BIS station from timezone UTC+1 to UTC+0 because it is geographically further west than other UTC+0 stations even though France uses UTC+1. As expected, Fig. 3.12 shows that stable conditions increasing in length over the year with May, June, and July having the longest stable phases. Additionally, except for the month of April, the starts and lengths of all phases behave monotonically. This, in conjunction with the general agreement of our definition with the previously cited studies, shows that we have developed a robust metric.

We now use this metric to analyze the simulation performance by station type and phase of PBL development. Results of this are to be found seen in Fig. 3.13 and Table 3.3. In general, we find that as expected urban areas are more difficult to simulate than rural areas and nighttime more than stable conditions. For stable conditions, in the MYJ simulation we find MAB values for CO_2 of 5.6, 4.3, and 3.2 ppm for urban, rural, and remote stations. At night, these increase to 7.8, 5.4, and 3.7 ppm. Pearson-R correlations in stable conditions are 0.83, 0.88, and 0.90 while at night they drop to 0.70, 0.78, and 0.83. The results are slightly different for CO, where the only available urban station is SAC and the IPR station errors are dominating the rural stations (cf. Table 3.4)

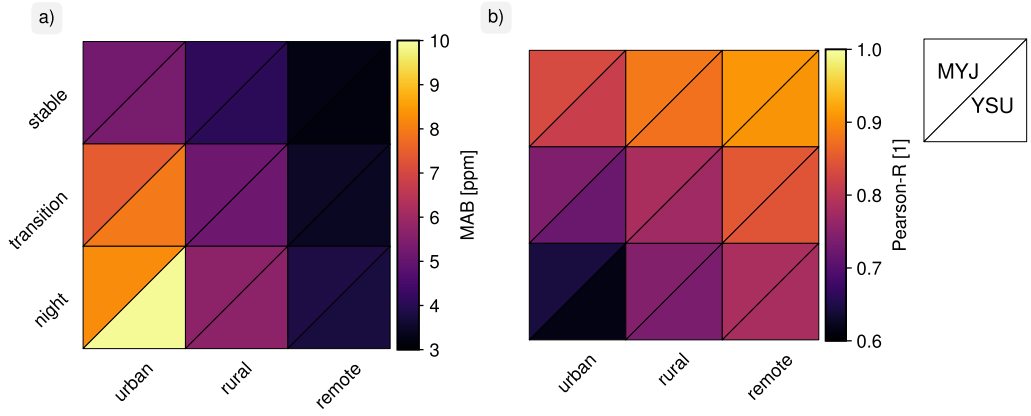


Figure 3.13: Overview of MAB (Panel a) and Pearson-R correlation (Panel b) of the simulated CO_2 concentrations split into station types and PBL phases. The performance dependency on PBL phase and station type is clearly visible. This figure was generated using `portraitpy` (Pilz, 2025a).

Table 3.3: Comparison of CO_2 concentration simulation performance across stations and phases of PBL development. All entries are averaged over all available inlets.

		urban		rural		remote		all	
		MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
stable	MB [ppm]	0.4	-0.4	1.6	1.4	1.6	1.6	1.4	1.2
	MAB [ppm]	5.2	5.3	4.1	4.1	3.2	3.2	4.1	4.1
	RMSE [ppm]	7.6	7.9	5.4	5.4	4.1	4.0	5.5	5.5
	Pearson-R [1]	0.83	0.81	0.88	0.87	0.91	0.91	0.88	0.87
transition	MB [ppm]	-2.7	-2.7	-0.2	-0.5	0.6	0.4	-0.4	-0.6
	MAB [ppm]	7.4	7.9	5.1	5.2	3.5	3.4	5.2	5.2
	RMSE [ppm]	11.2	11.9	7.2	7.2	4.7	4.7	7.3	7.4
	Pearson-R [1]	0.74	0.72	0.78	0.77	0.85	0.84	0.79	0.78
night	MB [ppm]	-3.8	-1.7	-0.6	-0.8	-0.0	-0.2	-0.9	-0.8
	MAB [ppm]	8.2	9.9	5.7	5.7	3.8	3.8	5.7	5.9
	RMSE [ppm]	12.2	14.3	7.8	7.8	5.4	5.3	7.9	8.2
	Pearson-R [1]	0.64	0.61	0.74	0.73	0.78	0.78	0.74	0.73
all	MB [ppm]	-2.2	-1.4	0.1	-0.1	0.6	0.5	-0.1	-0.2
	MAB [ppm]	7.1	8.1	5.1	5.1	3.6	3.5	5.1	5.2
	RMSE [ppm]	10.7	12.2	7.1	7.1	4.9	4.8	7.2	7.4
	Pearson-R [1]	0.75	0.73	0.82	0.81	0.85	0.85	0.81	0.81

Table 3.4: Comparison of CO concentration simulation performance across stations and phases of PBL development. All entries are averaged over all available inlets.

		urban		rural		remote		all	
		MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
stable	MB [ppb]	-8.7	-14.0	-16.0	-19.1	-9.1	-10.1	-13.4	-16.0
	MAB [ppb]	16.8	19.0	22.6	24.1	13.2	13.5	19.4	20.6
	RMSE [ppb]	26.8	30.0	37.5	40.0	19.3	19.6	31.3	33.2
	Pearson-R [1]	0.87	0.87	0.79	0.77	0.88	0.89	0.82	0.81
transition	MB [ppb]	1.1	15.7	-18.1	-21.0	-8.9	-9.9	-13.8	-14.7
	MAB [ppb]	24.9	38.2	26.7	28.3	13.5	13.8	22.7	24.9
	RMSE [ppb]	38.8	57.8	43.6	46.1	20.1	20.6	36.3	39.6
	Pearson-R [1]	0.75	0.68	0.77	0.75	0.88	0.88	0.80	0.78
night	MB [ppb]	5.7	56.7	-27.0	-29.6	-9.1	-9.9	-19.1	-16.7
	MAB [ppb]	34.4	79.7	36.6	38.2	14.4	14.8	30.0	34.8
	RMSE [ppb]	52.4	105.4	58.2	60.8	22.4	23.1	47.3	53.5
	Pearson-R [1]	0.64	0.66	0.75	0.72	0.85	0.84	0.77	0.75
all	MB [ppb]	0.3	27.0	-22.0	-24.8	-9.1	-10.0	-16.4	-16.2
	MAB [ppb]	27.2	52.9	30.5	32.1	13.9	14.2	25.4	28.6
	RMSE [ppb]	43.7	81.4	50.5	53.0	21.1	21.6	41.3	46.2
	Pearson-R [1]	0.71	0.67	0.76	0.74	0.86	0.86	0.79	0.77

Monteiro et al. (2024) tried to find some better metrics to decide which observations to include in atmospheric inversions for emissions estimation. They used data from the non-growing season and identified a wind velocity criterion of $\geq 5\,\mathrm{m\,s^{-1}}$. We now filter our data by situations, where the simulated wind velocity exceeds $5\,\mathrm{m\,s^{-1}}$ and see the results for the same analysis in Tables A.12 and A.13. We can see that this criterion significantly reduced the MAB for all phases of PBL development. Indeed, the MAB values now mostly depend on station type and not PBL development phase. This is because at winds of $5\,\mathrm{m\,s^{-1}}$, the atmosphere is well-mixed without a clearly defined mixing height. However, we also see that the correlations at nighttime are still lower than correlations at stable conditions or in transitions. This indicates that the nighttime CO_2 concentrations even for this subset of measurements are still not as well-represented as CO_2 concentrations in other times. Using our data we can more clearly separate these effects and confirm this well-known result.

Vertical Gradient

Now we can have a closer look at the individual inlets. Figure 3.14 gives an overview of the simulation performance at each individual inlet over inlet height above ground level. We compute the vertical gradient as the concentration difference between the two consecutive available heights divided by the distance between these inlets in meters. These concentration

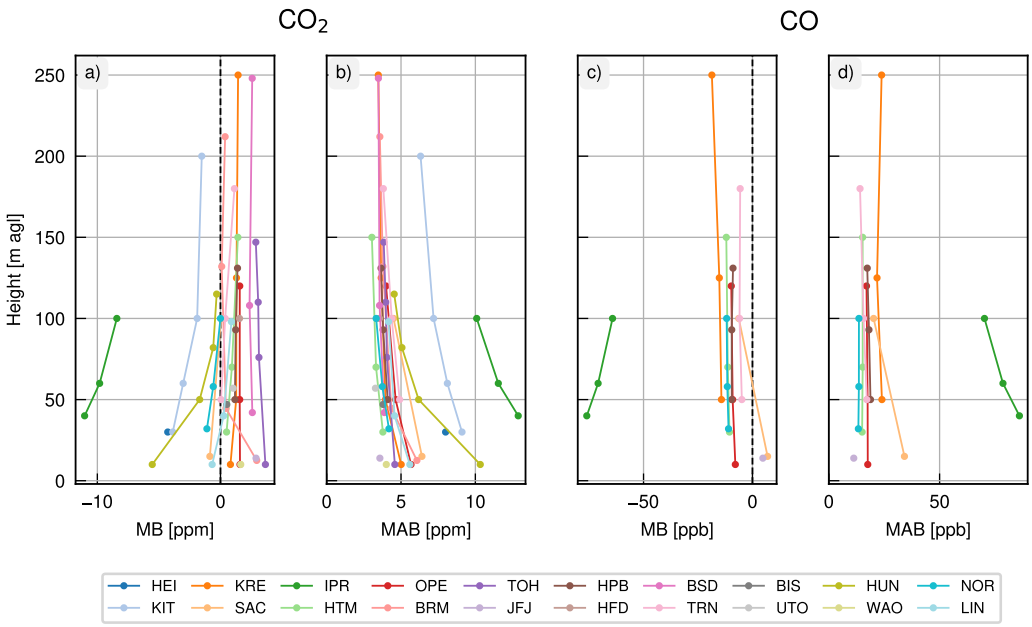


Figure 3.14: Mean biases (MBs) and mean absolute biases (MABs) of the MYJ simulation for both CO_2 and CO over all available inlet heights. This is a visualization of Table A.10. The increasing performance with increasing inlet height is clearly visible, as are the outliers, `IPR` and the lowest `HUN` inlet.

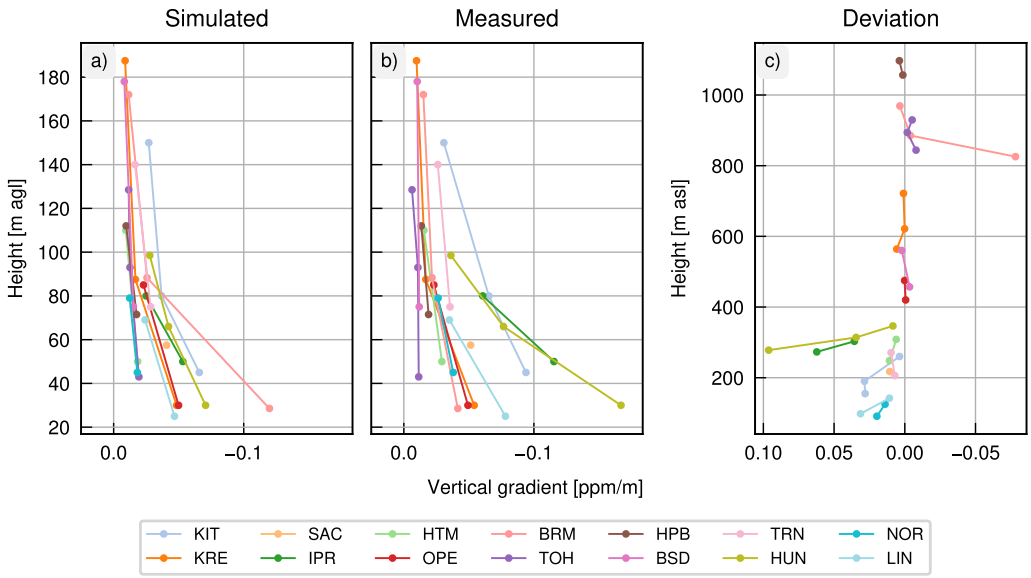
gradients are then assigned to the mid-point between these two inlets. The resulting gradients are to be found in Fig. 3.15.

As we can see in Fig. 3.15, vertical gradients are higher at low levels and then decrease with height. Looking at the deviations of CO_2 in Fig. 3.15, some station outliers, namely the stations IPR, BRM and HUN, are visible. As stated above, the IPR station (Ispra) is in a topographically and emissions-wise challenging spot at our simulation resolutions. Similarly, the BRM station (Beromünster) is located at the northern border of the Alps in Switzerland, 20 km to the North-West of Lucerne and 30 km to the South-West of Zurich. While it is also situated in close proximity to complex topography, the reason for the CO_2 concentration overestimation at the lowest inlet causing this gradient mismatch is probably the proximity to the A-2, the Gotthard-Highway. This highway is one of the most used ones in Switzerland and the 5 km resolution of our simulation causes the regridding to push some of the emissions into surrounding grid cells. Lastly, the HUN station (Hegyhátsál) is located in the Hungarian countryside. Here, the simulation performance could be affected by the VPRM misestimating biogenic fluxes due to the flux tower coverage over Eastern Europe in general and Hungary in particular being sparse. The only notable source of CO_2 in its vicinity is the trans-European E65 highway which it is located only 500 m away. Another reason for the gradient underestimation in our simulations, thus, may be the smoothing out of the E65 highway's emissions to the whole 15 km grid cell.

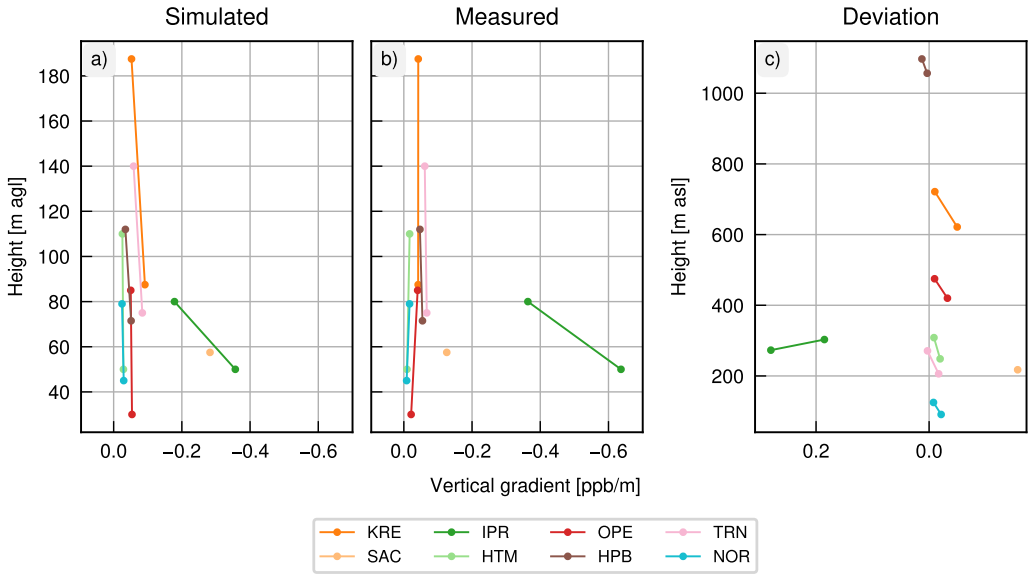
The general trend of vertical gradient deviations, however, hints at a systematic underestimation of the vertical gradient at lower elevations. This correlates with proximity to emissions in some cases, like KIT and SAC. However, also stations like NOR and HTM which we classified as remote see this underestimation of gradient. Looking at Fig. 3.14, we do not see a clear trend in biases between the stations seeing this underestimation. While stations like KIT, SAC, and NOR have negative biases, other stations like HTM or TRN see slight positive biases. This leads to the conclusion that this underestimation may be caused by a misestimation of the vertical diffusion coefficients in WRF.

3.4.5 Evaluation of Realism

The dataset produced is unique in terms of time period and area covered and is of high interest for many synthetic studies with various applications, which is why we want to evaluate its realism. We do this by looking at two main aspects. First, we analyze the features of our simulated data and check whether they are similar to ones of the measured data. In the second part, we check whether the simulated data is within known year-to-year variability. In this section, we do not directly compare with measurements as in the previous section.



(a) Vertical gradients of CO_2



(b) Vertical gradients of CO

Figure 3.15: Simulated and measured vertical gradients and their deviation for the MYJ simulation of both, CO_2 and CO. Simulation and measurement are plotted over inlet height above ground while deviations are plotted over inlet height above sea level. For CO_2, this clearly shows the `BRM` station outlier as well as `IPR` and `HUN`. A general underestimation of vertical gradient is seen for lower station heights. Vertical gradients are captured well for CO except for `IPR` and `SAC`.

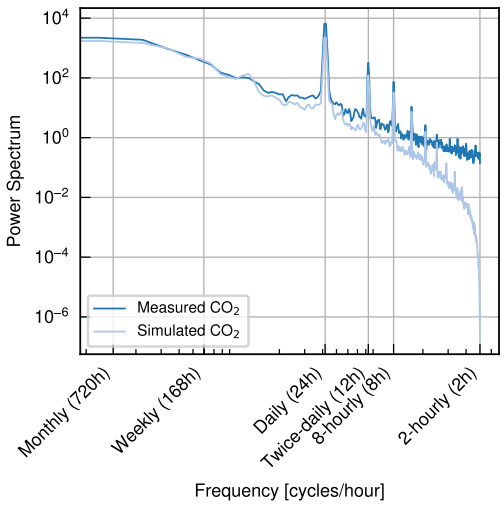


Figure 3.16: Power spectrum of average CO_2 concentration time series. Clearly visible are two main features, the seemingly regular peaks and the increasing deviation of the tail.

CO_2 Concentration Time Series

Our aim in assessing the realism of the CO_2 concentration time series is to take a step back from looking at whether the exact plume which passed over our measurement stations also passes over it at the exactly correct time in the simulations; rather we want to know whether, for example, these kinds of plumes come at similar intervals in our simulations as in measurements. This is why we use a power spectrum analysis for looking at the CO_2 concentration time series.

Due to the large amount of atmospheric noise on our data, we estimate the power spectral density of the averaged CO_2 time series using Welch's method (Welch, 1967). Figure 3.16, shows the spectrum which is cut off at 2 h, consistent with the Nyquist frequency of an hourly dataset. Additionally, there are some regular peaks clearly visible in the spectrum and we observe a deviation from the measured spectrum which is increasing at higher frequencies.

These peaks, which are present in the simulated and measured data, occur at 23.68, 12, 8.03, 6, 4.8, 4, 3.4, and 3 h. While some deviations from integer values like 8.03 and 23.68 h are probably due to the fast Fourier transform (FFT) grid not including an integer 24 h frequency, there are some that are not dividers of the base frequencies 24 h (diurnal cycle) like 4.8 and 3.4 h. We now reconstruct the signal of these peaks by doing an inverse FFT. However, since Welch's method does not have an inverse, we use these frequencies from Welch's method to find the closest ones in a standard FFT (where we find an additional peak at 2.7 h), extract the values of the peaks (including a buffer) and do an inverse FFT. Because of the noise on the measurements, it is not possible to extract all simulated peaks from their

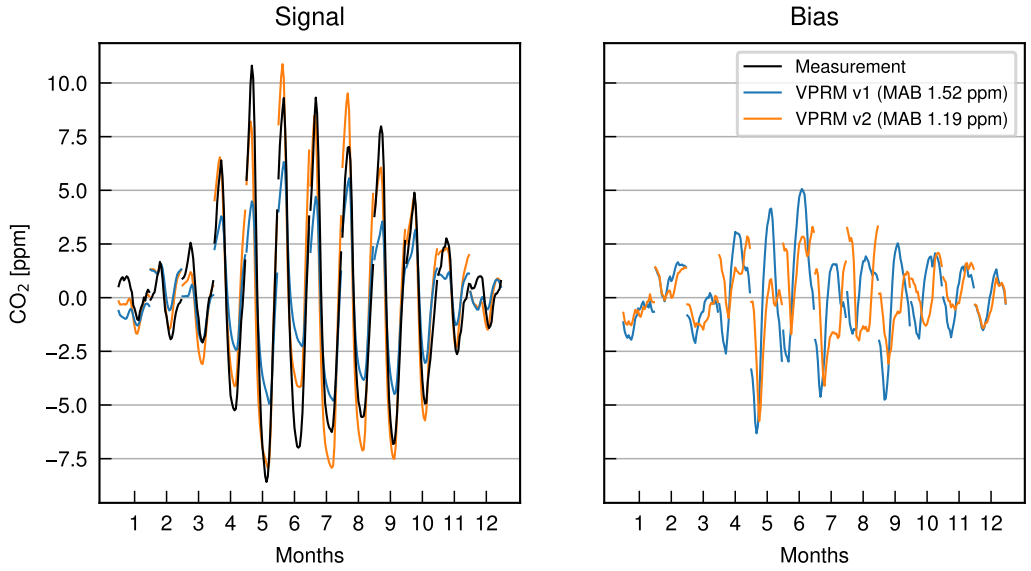


Figure 3.17: Visualization of the reconstructed signals from both VPRM models (MYJ configuration) and the trend-corrected diurnal cycle from measurements. Here, as already highlighted in Section 3.4.3, the VPRM v2 performs overall slightly better than VPRM v1, more accurately matching the diurnal cycle amplitude especially during summer.

spectrum. We, thus, compute the diurnal cycle of the time series corrected by a weekly rolling average trend. Figure 3.17 shows the diurnal cycle of the signals reconstructed from the sampled frequencies and the one computed from the measurements. We can clearly see the diurnal cycle amplitude increasing over the year and also its shift while some residual from the non-optimal biogenic model remains. This shows that our simulations realistically model the diurnal cycle but also hints at further development of the biogenic model being a fruitful avenue of progress.

The next feature we are going to analyze is the deviation in the tails between the simulated and measured spectra. This deviation is reminiscent of the spectral analysis of the kinetic energy in Skamarock (2004). In order to test the theory that the deviation-point is dependent on the resolution, we sample all stations within our 1 km domain also in the 5 km one. This 5 km domain has not only a coarser resolution, it also has a larger timestep than the 1 km one. Figure 3.18 shows a comparison between the spectra of the two simulation resolutions. While we do see the 1 km simulation performing better at timescales between 24 and 2.5 h, the 5 km simulation does not produce the same deviation at high frequencies as the 1 km one. As the 1 km resolution is part of the so-called gray zone of turbulence, the lower performance at higher frequencies may be due to a partial resolution of convection (Han and Hong, 2018; Honnert et al., 2020; Shi and Wang, 2022). This could produce high-frequency structures omitted by the 5 km resolution simulation.

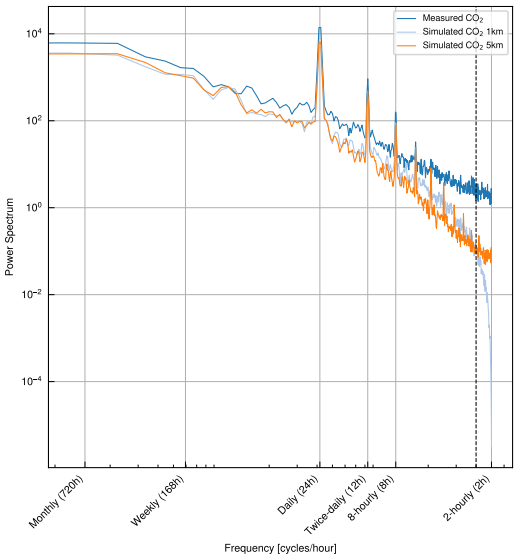


Figure 3.18: Comparison of power spectra of 1 and 5 km simulations and measurements. The 1 km simulation shows higher performance between 24 and 2.5 h. Its deviation below 2.5 h could be caused by partial resolution of convection.

CO_2 Seasonal Cycle

Next, we analyze the seasonal cycle of our simulated CO_2 concentration timeseries. Following Thoning et al. (1989), we calculate daily averages in order to filter out diurnal variations. However, since our stations are not background stations like Mauna Loa, we forgo further filtering of the data. We then use the following function to extract the seasonal cycle amplitude:

$$CO_2(t) = a \cdot \cos(b \cdot t + c) + d \cdot t + e. \tag{3.1}$$

It models the seasonal cycle as a cosine and adds a linear CO_2 increase over the whole year. We call the parameter a the seasonal cycle amplitude (even though it only is half the amplitude).

The daily averages of the simulated and measured CO_2 concentrations can be found in Figs. A.10 to A.15. Here, the daily averages of the simulated CO_2 concentrations are displayed in solid blue and the daily averages of the measured CO_2 concentrations in solid black. The dashed lines in the same colors are the fits of the respective concentrations with the function of Eq. (3.1). The amplitudes of the simulated timeseries are then subtracted from the ones of the measured timeseries. This deviation of the amplitude parameter is displayed in Fig. 3.19.

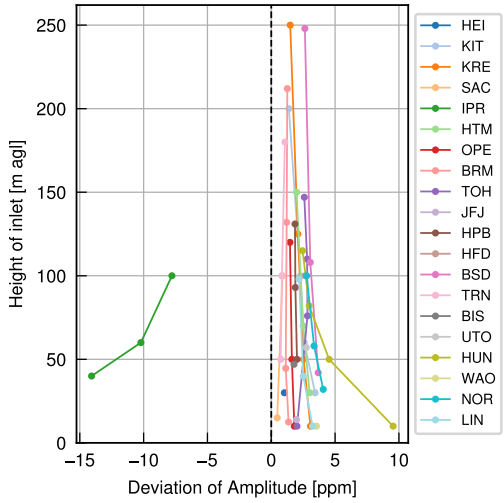


Figure 3.19: Deviation of simulated seasonal cycle amplitude fit parameter from measured one. Calculated as $a_{sim} - a_{meas}$. The stations IPR and HUN are clear outliers. The general overestimation of the seasonal cycle amplitude could be caused by the heat wave in 2018 (Ramonet et al., 2020; Zscheischler and Fischer, 2020).

Except for the station IPR, where our simulations underestimate the seasonal cycle amplitude, there is a general tendency toward overestimation. The seasonal cycle amplitude of 10.7 ppm averaged across all inlets is underestimated by our simulations by 1.6 ppm. A likely source of this bias in amplitude is the heat wave of 2018 (Zscheischler and Fischer, 2020), which Ramonet et al. (2020) have shown reduced the seasonal cycle of CO_2 over Central Europe by 1.4 ± 0.5 ppm. This suggests that the overestimation of CO_2 concentration seasonal cycle in our model may stem from the VPRM model and especially its respiration function not performing well in extreme situations.

In order to gauge how realistic our simulations of 2018 are, we do this fit procedure for all years with at least half data each month for each available inlet of the VPRM v1 dataset using the MYJ configuration. We then assume a Gaussian distribution for each of these parameters and check how well the 2018 parameters of our simulations fit to this distribution. To that end, we do a simple two-sided p-value calculation and select 5 % as significance criterion:

$$z_x = (par_{x,sim} - \mu(par_{x,meas,\neg 2018}))/\sigma(par_{x,meas,\neg 2018})$$
$$p_x = 2 \cdot (1 - \mathrm{CDF}(z_x))$$

with $par_{x,sim}$ the parameter fitted to simulated data to compare, $par_{x,meas,\neg 2018}$ the same parameters fitted to measured data for all years but 2018 with μ and σ its mean and standard

deviation. Furthermore, CDF is the cumulative distribution function of the normed normal distribution.

We find that the p-value meets the significance criterion at 20 of 52 inlets ($p >= 0.05$, "realistic"), while it falls short at 32 of 52 inlets ($p < 0.05$, "unrealistic"). Stratifying this by station type, we find the simulation to be "realistic" at all 7 urban inlets, while this is only the case at 12 of 36 rural and 1 of 9 remote inlets. We can explain this by the larger biospheric signal seen at the rural and remote inlets compared to the urban ones (cf. Section 3.4.3). One further confounding factor is the year 2018 being a particularly extreme climatic year in central Europe with a large heat wave (Zscheischler and Fischer, 2020). In order to assess how extreme this year was, we calculate the p-value of the 2018 measurements compared to all other years. Here, we find that even for the measured data seven rural inlets fall short of the significance criterion and are deemed "unrealistic". This indicates that this year was a particularly difficult year to model for a relatively simple biospheric model like VPRM.

3.5 Data Access

As of writing of this thesis, the MACRO-2018 dataset has been published on the Wold Data Center for Climate (WDCC) under Pilz et al. (2025a). There, a post-processed version is available as a part of WDCC under 10.26050/WDCC/MACRO-2018. The raw data are also available on WDCC under https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1170_dsg0001.

A second post-processed version is available in the DKRZ-provided S3 bucket under the Endpoint URL `https://s3.eu-dkrz-1.dkrz.cloud/` and path `s3://bb1170/public/MACRO-2018/`. It can be easily accessed using the `python` packages `xarray` and `zarr` as seen in Fig. 3.20.

3.6 Comparison to Literature

Next, we compare the performance of the MACRO-2018 dataset to studies published between 2012 and 2025. Since, however, few previous studies simulated CO_2 and CO concentrations with a comparable setup for a comparable length of time, we have to adjust some metrics or time periods we calculate our statistics over. We also try to find the closest comparison in station types to the stations the other studies evaluated their data at. For better readability, and because of the not well-understood deviation of YSU from measurements at the lowest `SAC` inlet, we only compare the MYJ configuration of MACRO-2018.

Ganshin et al. (2012) use a setup coupling the LPDM FLEXPART with the Eulerian model National Institute for Environmental Studies Japan Transport Model (NIES-TM). They compare at four stations over different time periods all over one year with three different model setups. We compare our results only with the best-performing, 1 km flux resolution

```python
import xarray as xr
blscheme = "MYJ" # or YSU
domain = 4 # available domains are 1-Europe, 2-Germany, 3-Rhine-Neckar, 4-Berlin, 5-Rhine-Ruhr, 6-Nuremberg, 7-Munich
s3_path = f"s3://bb1170/public/MACRO-2018/{blscheme}/wrfout_d{domain:02d}.zarr"

s3_options = dict(
        anon=True,
        client_kwargs=dict(
            endpoint_url="https://s3.eu-dkrz-1.dkrz.cloud/"
        )
    )
cache_options = dict(
        cache_storage="/tmp/zarr_cache"
    )

# Open the datatree directly
ds = xr.open_datatree(
        f"simplecache::{s3_path}",
        engine="zarr",
        consolidated=True,
        storage_options=dict(
            s3=s3_options,
            simplecache=cache_options,
        )
    )
```

Figure 3.20: Access to MACRO-2018 dataset on S3 at DKRZ using `python`. Necessary dependencies for this are `xarray`, `zarr`, and `s3fs`.

coupled setup. Comparing Fig. 2 from Ganshin et al. (2012) (which is created from only daytime data) to our Fig. 3.10 (which includes all PBL development phases), the standard deviations clearly match better in our data. Using only daytime measurements, Ganshin et al. (2012) find correlations of around 0.69 to 0.74 while the MACRO-2018 dataset exhibits correlations over all investigated Obspack inlets during stable PBL phases of 0.81 to 0.94.

Nehrkorn et al. (2013) simulate CO_2 using WRF-Chem over Salt-Lake City, Utah at 1.3 km resolution for October 2006. They find mean biases of −16.07 ppm to −13.3 ppm and RMSEs of 27.3 ppm to 34.1 ppm. Trying to make the comparison as similar as possible, we select the urban stations for October 2018. Here, the MACRO-2018 dataset has biases of −4.5 ppm to 3.2 ppm (average: −1.1 ppm) and an RMSEs of 4.1 ppm to 12.7 ppm (average: 9.6 ppm).

Bréon et al. (2015) use the Chimere transport model (Menut et al., 2013) to simulate CO_2 concentrations at 2 km resolution. Their study region is the greater Paris area and includes the 180 m inlet of the `TRN` station, which is also available in the Obspack dataset. They report a bias of −2.8 ppm and a correlation of 0.81 for the period of November 27th to December 27th 2010. Comparatively, for the same period in 2018, we find a bias of 1.8 ppm and a correlation of 0.80 at the same inlet.

Boon et al. (2016) also use the Chimere model to simulate CO_2 concentrations in the vicinity of London at 2 km resolution. Unfortunately, they do not compare at Obspack stations. However, three of their stations are urban and one is rural. They calculate the aggregated statistics for the stable PBL development phase from July to September 2012. We select the same PBL development phase and months to get the best possible comparison. For the urban stations, while Boon et al. (2016) find mean biases and RMSEs of −9.1 ppm to −5.5 ppm and 9.0 ppm to 11.7 ppm, we find biases of −2.2 ppm to 1.4 ppm (average: −0.5 ppm) and

RMSE values of 3.0 ppm to 6.5 ppm (average: 5.6 ppm). For their rural station, they find a mean bias of −5.3 ppm and a RMSE of 8.4 ppm. The MACRO-2018 dataset has mean biases of −1.0 ppm to 1.0 ppm (average: 0.0 ppm) and RMSEs of 2.4 ppm to 6.2 ppm (average: 3.7 ppm).

Feng et al. (2016) use WRF-Chem in four different configurations at resolutions of 1.3 km to 4 km. They compare the forward simulated CO_2 concentrations to the ones measured during the CalNex-LA campaign from mid-May to mid-June 2010. Both of the sites they compare their simulations against can be classified as urban, as they are in Los Angeles and very close to the city respectively. They compare the stable PBL phase measurements to their simulations and, for hourly concentrations, get biases of −1.38 ppm to 8.91 ppm (average: 4.5 ppm) and RMSEs of 9.13 ppm to 19.64 ppm (average: 13.7 ppm). Looking at stable PBL phase measurements between 15th of May and 15th of June, the MACRO-2018 dataset exhibits biases of −2.5 ppm to 0.1 ppm (average: −1.7 ppm) and RMSEs of 7.8 ppm to 8.9 ppm (average: 8.6 ppm).

Another study investigating CO_2 dispersion in Europe is Lian et al. (2021). They use WRF-Chem v3.9.1 centered over Paris for the time period December 2015 to November 2016. Lian et al. (2021) simulate CO_2 using the Institute of Energy Economics and the Rational Use of Energy, University of Stuttgart (IER) and AirParif inventories (at 5×5 km a 1×1 km for the Île de France region) at 1 km resolution. Their physics configuration also uses the BEP urban parametrization and the MYJ PBL scheme. The inlets they evaluate their CO_2 concentration at include two `SAC` inlets that are also in our Obspack data. While they find RMSE values of 10.23 and 7.64 ppm for the 15 m and 100 m Saclay inlets, the MACRO-2018 dataset has RMSEs of 9.76 and 6.42 ppm. As for Pearson correlation coefficients, our dataset has coefficients of 0.76 and 0.84 versus coefficients of 0.72 and 0.74 in Lian et al. (2021).

The most recent paper on CO_2 concentration simulation is Bisht et al. (2025). They simulate CO_2 concentrations over Japan at up to 1 km resolution using WRF-GHG v4.2.1, the EDGAR inventory for anthropogenic, and VPRM for biogenic emissions. These simulations are compared with one continuous monitoring site for February 2018 and February 2020 and two additional ones for May 2018. However, for these comparisons, the only aggregate statistical metric they report is the Pearson correlation coefficient, which is a shortfall in reporting quality even compared to papers over ten years old. Nevertheless, for February 2018, they find a correlation of 0.38 and one of 0.66 for the 1 and 27 km domains while the correlations for the February 2020 simulations are 0.66 and 0.74. For May 2018, the correlations at the Kisai, Mt. Dodaira and Yoyogi stations are 0.38 to 0.47, 0.22 to 0.29, and 0.20 to 0.25 for different emissions inventories. Since the Kisai and Yoyogi stations are located within the greater Tokyo area, we compare them with our urban stations. The Mt. Dodaira station is located on a mountain in the vicinity of Tokyo and thus be compared to our rural stations.

For February 2018, our simulations exhibit correlations of 0.63 to 0.84 (average: 0.71) for urban stations. For May 2018, the correlations for rural stations like their Mt. Dodaira station are 0.44 to 0.82 (average: 0.62) and for urban ones 0.49 to 0.71 (average: 0.62).

Finally, the study which is the closest to our setup is Wang et al. (2025) (preprint at time of writing). They simulate CO_2 dispersion over Belgium and surrounding regions at 3 km resolution between 1st of June and 31st of August 2018 and compare their simulations to ICOS stations which are part of Obspack. Their physics setup is based on Kuik et al. (2016) with BEP as urban parametrization and YSU as PBL scheme. We compare the MACRO-2018 dataset to their results at the inlets which we both evaluated in Table 3.5. They simulate multiple emissions inventories, CAMS, EDGAR, and TNO, however without any vertical distribution of emissions. Our simulation performance is on par with or exceeds all of these inventories with regards to RMSE and correlation. They also present simulations assuming all emissions are from point sources and distributed them vertically based on plume rise models of Brunner et al. (2019) (CAMS_P, EDGAR_P). Compared to these unrealistic inventories, MACRO-2018 performance is exceeded at the lower inlets of `KIT`, an urban station. However, our performance is higher at the other inlets. Overall, as evidenced by the higher performance at the more rural stations, one major contributor to our simulation performance is the comparatively good biogenic emissions model. However, as also pointed out by Wang et al. (2025), the extreme weather conditions in 2018 could show limitations in performance of the relatively simple respiration model in VPRM. This is also suggested by the general overestimation of the seasonal cycle in our simulations which fits findings by Ramonet et al. (2020), as presented in Section 3.4.5. This, in conjunction with the higher summer daytime performance of VPRM v2, suggests that there is large potential for developments in the biogenic model. Wang et al. (2025) attribute the large performance differences between the inventories at `KIT` to an oil refinery at a distance of 6.5 km. If this is correct, the comparatively lower performance of the MACRO-2018 dataset to the artificial inventories could be caused by the emissions time factors not being accurate for this particular point source.

Table 3.5: Comparison of RMSE and correlation values for simulation of CO_2 concentrations during June to August 2018, following the analysis by Wang et al. (2025). The best performing simulation for each inlet height is highlighted in grey.

	RMSE						Pearson-R					
	CAMS_P	CAMS_S	EDGAR_P	EDGAR_S	TNO	MACRO-2018	CAMS_P	CAMS_S	EDGAR_P	EDGAR_S	TNO	MACRO-2018
KIT_30m	15.3	38.7	15.2	57.4	26.6	17.7	0.79	0.57	0.78	0.53	0.65	0.77
KIT_60m	12.0	22.0	11.8	30.6	17.6	13.7	0.78	0.65	0.78	0.61	0.71	0.74
KIT_100m	10.6	15.0	10.3	19.0	13.3	11.1	0.74	0.63	0.75	0.59	0.70	0.71
KIT_200m	9.9	10.8	9.9	11.3	10.6	9.5	0.63	0.57	0.63	0.56	0.59	0.62
TRN_50m	10.0	10.0	10.1	10.1	10.1	9.3	0.62	0.62	0.63	0.63	0.62	0.63
TRN_100m	8.3	8.3	8.3	8.3	8.3	7.7	0.56	0.56	0.56	0.57	0.56	0.55
TRN_180m	7.0	7.0	7.0	7.0	7.0	6.2	0.44	0.45	0.45	0.45	0.45	0.49
SAC_60m	7.7	7.6	7.8	7.7	7.8	7.2	0.69	0.70	0.69	0.70	0.71	0.74
SAC_100m	7.5	7.5	7.5	7.6	7.5	6.6	0.62	0.63	0.63	0.63	0.64	0.71
OPE_50m	9.5	9.5	9.6	9.6	9.6	7.4	0.46	0.47	0.46	0.47	0.47	0.57
OPE_120m	6.9	6.8	6.9	6.8	6.8	5.5	0.46	0.47	0.46	0.47	0.47	0.56

3.7 Discussion

In this chapter, we have presented MACRO-2018, a unique dataset containing meteorology as well as CO_2 and CO concentrations at 1 km resolution over German metropolitan areas, which provides an extensive simulation period of one year and a large spatial extent across Central Europe. It, furthermore, includes two separate physics configurations and two different biogenic models which enables atmospheric and biogenic model performance investigations. We have evaluated its performance against high-precision CO_2 and CO concentration measurements from Obspack datasets (ICOS RI et al., 2024; Schuldt et al., 2025).

Both the MYJ and YSU physics configurations demonstrate similar performances with the largest differences at `SAC` station. Overall, the dataset performance in CO_2 MAB (using VPRM v1) is 5.1 ppm for MYJ and 5.2 ppm for YSU. For CO, the MABs are 25.4 and 28.6 ppb for MYJ and YSU. The biogenic emissions we use are modeled using VPRM and are available in two versions which use different satellite data products and parameters (Marshall, 2022; Glauch et al., 2025). Due to the large simulation domain and long time period, MACRO-2018 presents a unique opportunity to validate the developments in VPRM v2. Here, we show a lower performance during night time which hints at too large respiratory fluxes, however a high performance in the summer months at daytime. This comparison of forward modeled concentrations over such a long timeframe at nearly continental scale is unique and our results are actively supporting the further development of VPRM.

In order to make our evaluation more robust, we have categorized the stations of the Obspack dataset into 'urban', 'rural', and 'remote' based on their exposure to anthropogenic CO_2 concentrations. We show that performance at urban stations is lower than at rural ones due to the more difficult meterological situation and the higher heterogeneity in emissions. As such, while remote stations usually exhibit MABs for CO_2 of 3.2 ppm to 3.8 ppm, urban stations' CO_2 MABs range from 5.2 ppm to 8.2 ppm for the MYJ configuration.

To enable robust performance evaluation, we have developed a coherent meteorological definition of diurnal PBL stability phases based on PBL height gradients. This definition has the potential to improve measurement data filtering, increasing the amount of data available for applications like emissions estimations. We show that when applying this definition, the performance of the MACRO-2018 dataset for CO_2 behaves as expected, depending on both PBL stability phase (best at stable conditions, worst at night) and station type (best at remote and worst at urban stations). For CO, however, the dependency on PBL stability phase and station type is less pronounced due to lower amount of measurements and CO in general having a less clear diurnal cycle in emissions.

We have also evaluated the realism of the dataset. To that end, we have investigated the Fourier transform of the CO_2 concentration time series and the seasonal cycle. We show that our simulated concentrations exhibit the most important features of the Fourier transform of the measured ones. Here, the 1 km resolution matches the measurements more closely for sub-daily to 2.5-hourly frequencies than the 5 km resolution. Furthermore, we have found that the dataset reproduces a realistic seasonal cycle at the urban stations compared to previous years.

These results have to be seen in the context of the very challenging meteorological situation of the year 2018. Especially the biospheric development was significantly impacted by the large heat wave across the whole year (Zscheischler and Fischer, 2020). This is a difficult situation for the VPRM models as they were not developed for extreme conditions. Improvements in construction of the model can help here. Promising avenues are a more realistic, non-linear, respiration function (Niu et al., 2024) and spatially disaggregated parameters for land use classes. The latter ones are important as, at the moment, land use classes like 'grassland' or 'cropland' are treated the same over all of Europe, even though they might contain widely varying species of plants depending on their geographical location, exhibiting different characteristics. The data generated in MACRO-2018 are actively supporting these developments.

Finally, we compare our simulated CO_2 and CO concentrations against previous studies. Here, we find the performance of our dataset to match or exceed all presented results. This shows that it is a valuable asset for the scientific community and is ready to be used for different applications. Forms in which it is already being used are the comparison with high-precision measurements by ICOS station staff, preparatory studies for satellite mission planning, and investigations of urban emissions monitoring systems. These investigations include observing system simulation experiments (OSSEs) and the analysis of long-path dual-comb spectrometer measurements over urban areas.

In conclusion, the MACRO-2018 dataset created in this thesis is a high-resolution, high-fidelity, and realistic dataset whose performance is similar to or better than all comparable previous forward modeled CO_2 concentration results we were able to find. It delivers two full simulation runs of different physics configurations including two different biogenic models each, creating a mini-ensemble. This ensemble not only supports the development of the biogenic models but also enables a first-order estimate of model and transport uncertainties. While there are some discrepancies to measurements, these behave mostly as expected. Possible sources of these deviations include atmospheric transport (`IPR`, `JFJ`), errors in emissions inventories (VPRM v1 vs. v2) and even completely neglected sources of CO_2 emissions like human respiration, which can contribute about 7 % to 18 % in urban areas (Stagakis et al.,

2023; Dröge et al., 2024). These shortcomings in the inventories can be resolved to some degree by conducting a data assimilation, which is the topic of the next chapter.

4 Emissions Estimation

The previous chapter presented the MACRO-2018 dataset and its genesis. This chapter focuses on the optimization of a CO_2 emissions estimate for Berlin. We base this analysis on aircraft measurements which were performed in the context of the Urban Climates Under Change ([UC]^2) campaign around Berlin in 2018, published in Scherer et al. (2019) and Klausner et al. (2020). Since aircraft can cover large areas in a short amount of time, this data provides valuable insights into urban boundary layer meteorology and trace gas dispersion. Given its high cost of generation, it is crucial to analyze it as thoroughly as possible. We do this by integrating it into a Bayesian inversion framework combining traditional emissions estimation with a background estimation. The Bayesian inversion is driven by the meteorology of MACRO-2018 and by its forward modeled concentration data which is used as a prior background estimate. We compare the emissions estimate from this Bayesian inversion to a simpler mass balance approach. In Section 4.1, we present the methodology of the two different emissions estimation approaches under comparison, mass balance and Bayesian inversion. Continuing, in Section 4.2, we present the measurement data and resulting mass balance emissions estimate. We then introduce our Bayesian inversion framework in Section 4.3. Next, in Section 4.4 we compare simulated to measured data, explain the setup of our Bayesian inversion and present its results. In Section 4.5, we end with discussing outcomes and differences between the estimation approaches.

4.1 Background and Methods

This section will present both, the mass balance approach to emissions estimation as well as the Bayesian inversion approach. We start with discussing the theory and assumptions behind mass balance following Klausner et al. (2020). Then we dive into the theoretical basis of the Bayesian inversion based on Rodgers (2000). Finally, we will discuss FLEXPART, the Lagrangian particle dispersion model driving our Bayesian inversions.

4.1.1 Mass Balance

The mass balance approach to emissions estimation, fundamentally, is derived from conservation of mass. It uses inflow and outflow measurements over an area in order to determine

a change in concentration of the same air mass. This change is then attributed to the flux within the area between the measurements. In urban areas, the influence of biospheric fluxes is expected to be small in comparison to anthropogenic emissions (in our case approx. 10 %). Thus, while we will use the term 'emissions' in the following, we mean CO_2 fluxes which also include a small biospheric component.

The mass balance method as used in Klausner et al. (2020) tries to capture the concentrations within the urban plume. The concept and physics of the urban plume were presented in Section 2.1.1. The method estimates emissions by measuring the flow of trace gas enhancements through a plane normal to the prevailing winds. This plane is vertically bounded by the UBL height, since its entrainment layer at the top restricts vertical GHG transport. Therefore, only measurements during phases where a well-defined mixing layer spans the whole UBL up to the entrainment layer can be used. Aircraft-based methods determine these enhancements by flying upwind and downwind transects from the target and comparing the respective concentrations.

The flux (in mass per time) through said normal plane is computed as:

$$f = \int_0^{\mathrm{PBL}} \mathrm{d}z \int_a^b \mathrm{d}x (c(x,z) - c_{\mathrm{bck}}(x,z)) \frac{p(x,z)}{T(x,z) \cdot R} M \cdot u(x,z)$$

with (a,b) being the horizontal extent of the plume, $c_{(\mathrm{bck})}$ the (background) concentrations, p the pressure, T the temperature, R the ideal gas constant, M the molar mass, and u the component of the horizontal wind velocity perpendicular to the normal plane.

This approach assumes a static meteorological system with a well-defined mixing layer extending the UBL for several hours before and after the flight. This significantly reduces the number of useable meteorological situations and introduces large uncertainties in the estimated emissions. The largest contributions to this uncertainty are the variation in background, wind velocity and direction, and PBL height.

4.1.2 Bayesian Inversion

Bayesian inversions are a fundamental tool in atmospheric science. They enable the estimation of parameters of a linear system from observational data under the consideration of prior knowledge and uncertainties.

The system under investigation looks like this:

$$y = F(x) + \varepsilon \tag{4.1}$$

with y being the measurements of the system F at state x with measurement error ε. Here, F is also called the forward model. Our objective is to estimate the system state x using

information we gain from measurements y. This requires inverting Eq. (4.1), which we do using Bayes' theorem. The following description of how Bayesian inversions work is based on Rodgers (2000).

Bayes' theorem, which is the basis for this approach, reads:

$$P(x|y) = \frac{P(y|x)\,P(x)}{P(y)}\,. \tag{4.2}$$

What we are interested in is the posterior probability density function (PDF) $P(x|y)$, the probability of system state x given measurements y. In order to find this, we update our prior knowledge expressed in the prior PDF $P(x)$ with the information gained from measurements. The PDF $P(y|x)$ describes the probability to obtain observations y given system state x. Finally, $P(y)$ is the PDF of the measurements y and is usually disregarded as it is just a normalizing factor.

More explicitly, assuming a linear forward model, we obtain:

$$y = F(x) + \varepsilon = \mathbf{K}x + \varepsilon\,. \tag{4.3}$$

Assuming Gaussian PDFs and using the forward model to obtain the most likely observation given a certain state, we get

$$-2\ln P(y|x) = (y - \mathbf{K}x)^T \mathbf{S}_e{}^{-1} (y - \mathbf{K}x) + c_1 \tag{4.4}$$

with the covariance matrix of the model-measurement part $\mathbf{S}_e$ (also called model-data mismatch covariance matrix). We construct $P(x)$ as a Gaussian distribution centered around our prior assumption x_a

$$-2\ln P(x) = (x - x_a)^T \mathbf{S}_a{}^{-1} (x - x_a) + c_2 \tag{4.5}$$

with x_a the a priori value of the system state and $\mathbf{S}_a$ the prior covariance matrix.

We now substitute Eqs. (4.4) and (4.5) in Eq. (4.2) to find

$$J(x,y) := -2\ln P(x|y) = (y - \mathbf{K}x)^T \mathbf{S}_e{}^{-1} (y - \mathbf{K}x) + (x - x_a)^T \mathbf{S}_a{}^{-1} (x - x_a) + c_3\,. \tag{4.6}$$

This equation J is central to the implementation of this approach, as it is the cost function of our problem, meaning that minimizing J maximizes $P(x|y)$. Varying states x while keeping measurements y constant until J is minimal now yields exactly what we were looking for, which is the most probable state x given measurements y.

This most probable state given the measurements is called the posterior and is denoted by $\hat{x}$. However, the formal solution to the Bayesian problem is the complete PDF $P(x|y)$ of

which $\hat{x}$ is only the expected value. As such, it is associated with the so-called posterior covariance $\hat{\mathbf{S}}$ via

$$-2\ln P(x|y) = (x-\hat{x})^T \hat{\mathbf{S}}^{-1}(x-\hat{x})\,. \tag{4.7}$$

Equating linear and quadratic terms of Eqs. (4.6) and (4.7) and following the derivation in Rodgers (2000) yields us the following form for the posterior and its covariance:

$$\begin{aligned}\hat{x} &= x_0 + \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_e)^{-1}(y-\mathbf{K}x_0)\\ \hat{\mathbf{S}} &= \mathbf{K}^T{\mathbf{S}_e}^{-1}\mathbf{K} + {\mathbf{S}_a}^{-1}\,.\end{aligned}$$

Since the covariance matrices are quite central to Bayesian inversions, we discuss them in the following section.

Covariance Matrix

In general, the covariance is a measure of joint variability of two random variables. For the variables X and Y, is defined as

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

with $E[.]$ being the expected value operator. It is closely related to the Pearson correlation coefficient, which is essentially a normalized version of the covariance

$$\begin{aligned}Corr(X,Y) &= \frac{Cov(X,Y)}{Std(X)Std(Y)}\\ \Rightarrow\ Cov(X,Y) &= Corr(X,Y)\cdot Std(X)\cdot Std(Y)\\ &= \rho_{X,Y}\cdot\sigma_X\cdot\sigma_Y\,.\end{aligned}$$

In our case, where the random variables are vector elements, the covariance is a matrix which is defined as

$$S_{ij} = \sigma_i\cdot\rho_{i,j}\cdot\sigma_j$$

or

$$\mathbf{S} = \begin{pmatrix}\sigma_0^2 & \sigma_0\sigma_1 & \cdots\\ \sigma_0\sigma_1 & \sigma_1^2 & \\ \vdots & & \ddots\end{pmatrix} \odot \begin{pmatrix}1 & \rho_{0,1} & \cdots\\ \rho_{0,1} & 1 & \\ \vdots & & \ddots\end{pmatrix}$$

with $\odot$ being the element-wise multiplication operator.

These uncertainties are important because they determine how strongly the posterior is driven by measurements and prior information. While they are central to Bayesian inversions, they can be difficult to obtain directly from a priori information. In practice, while parts of the model-data mismatch covariance can be derived from measurement statistics, estimating the prior covariance from data is rarely possible. Therefore, the prior covariance is often defined by operator's choice or selected to balance inversion stability and physical interpretability of the results.

Evaluation

One of the important points to analyze regarding the performance of our inversion is the construction of our state space. There are two important concepts which help with its evaluation, the gain matrix and the averaging kernel matrix. These concepts are sometimes known under different names, as the gain matrix is sometimes also called the 'generalized inverse of **K**' and the averaging kernel matrix has names ranging from 'model resolution matrix' to 'resolving kernel' (Rodgers, 2000).

The derivation of these concepts is performed in Rodgers (2000) with their respective definitions being

$$\mathbf{G} = \left(\mathbf{K}^T \mathbf{S}_e^{-1} \mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1} \mathbf{K}^T \mathbf{S}_e^{-1}$$
$$\mathbf{A} = \mathbf{G}\mathbf{K}.$$

These concepts are interesting to us as they allow us to formulate versions of the posterior state vector like

$$\hat{x} = x_0 + \mathbf{G}(y - \mathbf{K}x_0) \tag{4.8}$$
$$= (\mathbb{1} - \mathbf{A})x_0 + \mathbf{G}y. \tag{4.9}$$

This shows that the gain matrix describes the influence of the measurements on our posterior. In order to gain a better understanding for the averaging kernel, we can substitute $y = \mathbf{K}x_{\text{true}} + \varepsilon$ in Eq. (4.9):

$$\hat{x} = \mathbf{A}x_{\text{true}} + (\mathbb{1} - \mathbf{A})x_0 + \mathbf{G}\varepsilon. \tag{4.10}$$

The averaging kernel matrix, thus, describes the posterior's distance between the prior and the true state. The gain matrix has a more descriptive function and there is no obvious form to optimize for. However, the influence the measurements have on the posterior should be realistic. In contrast, there is a clear ideal averaging kernel matrix, which would be an identity

matrix. This would mean that the estimated state is as close to the true state as is possible considering the measurement error and forward model construction. Large off-diagonal elements, however, indicate interdependencies between the state vector elements and hint at possible improvements in state vector construction. If these off-diagonal elements exceed the bounds of negative and positive one, the simulation is considered not to be physically sensible.

Another metric being used to assess inversion performance is the error reduction. It quantifies how the posterior uncertainties compare to the prior ones and is defined for each state vector element as:

$$\text{Err. Red.}_i = (\sqrt{\mathbf{S}_{a,ii}} - \sqrt{\mathbf{S}_{e,ii}})/\sqrt{\mathbf{S}_{a,ii}}\,.$$

An issue with this metric is that it can be arbitrarily inflated by choosing larger prior uncertainties. However, if the uncertainties are chosen in a well-founded and considerate way, the metric illustrates how much more informed the state vector estimate has become due to the measurements being assimilated.

Until now, we have derived the general Bayesian inversion of a linear model. Applying this to a traditional investigation of GHG emissions, excluding background corrections, the individual terms would be:

x – emissions

y – measured enhancements

$\mathbf{K}$ – matrix containing footprints (sensitivity of emissions x to measurements y)

$\mathbf{S}_a$ – prior covariance matrix (covariance of emissions)

$\mathbf{S}_e$ – model-data mismatch covariance matrix (combined covariance of modeled concentrations and measurements).

However, in Section 4.3 we extend this for our specific use case.

4.1.3 FLEXPART

We now present the Lagrangian particle dispersion model (LPDM) we use to linearize the forward model in our Bayesian inversion framework. Our LPDM is called FLEXPART-WRF (Brioude et al., 2013). It is an adaptation of the Flexible Particle dispersion model (FLEXPART; Pisso et al., 2019) which is able to ingest WRF meteorology. This is of great advantage, as our WRF meteorology is of much higher resolution than the ERA5

meteorology, which would be used instead. LPDMs in general have been extensively used to study transport processes in the atmosphere. These studies range from investigating radionuclide fallout in the context of emergency response systems (Stohl et al., 2012; Arnold et al., 2013) over short- and long-range pollutant transport (Bahreini et al., 2009; Cooper et al., 2010) to determining flux footprints of measurement stations (Flesch et al., 1995; all as cited in Brioude et al., 2013).

As the name suggests, LPDMs are Lagrangian models, meaning they stochastically simulate the advection of individual tracer particles through an Eulerian meteorology instead of computing Eulerian concentration fields. However, they are constructed to be consistent with Eulerian transport probability density functions. This focus on individual particles makes them very well suited for investigating individual point events. These events could be emissions from sources like power stacks or measurements from receptors like in-situ concentration sensors or aircraft.

The capability to investigate not only sources but also receptors stems from the fact that LPDM in general as well as FLEXPART in particular can integrate the advection of these particles not only forwards but also backwards in time. This means we can spawn pseudo-particles associated with certain substance masses at the measurement times and locations (receptors), integrate their transport backwards with the provided meteorological fields and study their origin. In order to better sample the particles' distribution, it is also possible to increase the number of pseudo-particles.

The sensitivities FLEXPART returns are volumetric data on the whole emissions grid. This data is a function of the residence time of the particles in the specific cells which uses a unit akin to ppm/(mol/km^2/h), so concentration per emission. Multiplying these so-called footprints with the emissions fields then yields the enhancement in ppm caused by the emissions at the specified receptor and to the specified time we spawned our particles at.

However, in order to get a footprint for the whole measurement, one would have to theoretically integrate the meteorology backwards ad infinitum. This means there is always a residual of the concentration contribution from before the end point of the backwards integration, which has to be treated. We call this contribution the 'background'. We present the concrete implementation of the Bayesian inversion framework and how we leverage the forward modeled concentrations from the MACRO-2018 dataset in order to constrain this background in Section 4.3.

4.2 Case Study: Aircraft Campaign

The following section presents the aircraft campaign measurements and emissions estimate which was published in Klausner et al. (2020).

4.2.1 Aircraft Measurements

The [UC]^2 campaign was a large measurement campaign focussing on urban meteorology in Berlin, Hamburg, and Stuttgart which included the aircraft measurements we use in the following (Scherer et al., 2019). It was conducted from the 16th to 26th of July 2018. The aircraft measurements used a Cessna 208B Grand Caravan from The German Aerospace Center (DLR), which was fitted with a meteorological sensor package (METPOD) and multiple trace gas analyzers. The relevant analyzer for our purposes is the Picarro (G1301-m) CRDS, which measures CO_2, CH_4, and water vapor. Information on the calibration procedure is provided in Klausner et al. (2020).

Within the time of the campaign, from 16th to 26th of July 2018, five flights took place (cf. Fig. 4.1). The purpose of these flights was to fly upwind transects measuring background air and then fly downwind of the city in order to transect the urban plume. The flights were carried out on the 18th, 20th, 23rd, 24th, and 25th of July. The three flights on the 18th, 23rd, and 25th of July are excluded because either the urban plume was missed or the PBL was not clearly separated from the free troposphere. Furthermore, the flight on the 23rd of July does not yield any emissions estimate because of low wind velocities. This leaves the flight of the 20th of July, which serves as a proof of concept for establishing the Bayesian inversion and its performance compared to the mass balance approach. We investigate this flight in the following.

4.2.2 Results

Figure 4.2 displays the CO_2 measurements during the transects, which were flown at 489, 969, and 1601 m altitude and the vertical profile measured over Tempelhofer Feld. The PBL height needed for the flux estimation was identified by Klausner et al. (2020) from the meteorological variables of the vertical profile shown in Panel b) to be 2737 ± 2 m. It is assumed that the PBL height does not vary between the time the vertical profile was recorded (12:10 to 12:20 UTC) and the downwind transects (12:30 to 13:45 UTC).

In order to get an estimate of the background uncertainty, Klausner et al. (2020) use two methods of background estimation. First, the background is linearly estimated (dashed lines in Fig. 4.2, Panel a). Here, the plume is identified as "when the measured mixing ratios exceeded the inward running mean (interval of ±30 s) plus one standard deviation [...] for at least five consecutive measurement points" (Klausner et al., 2020). The background is then calculated using the average of 60 measurement points outside the plume. The second method uses the Hybrid Single-Particle Lagrangian Integrated Trajectory Model (HYSPLIT) in order to project the upwind trajectory onto the downwind wall. These projected measurements are also depicted in Panel a) of Fig. 4.2 as the solid faint purple line.

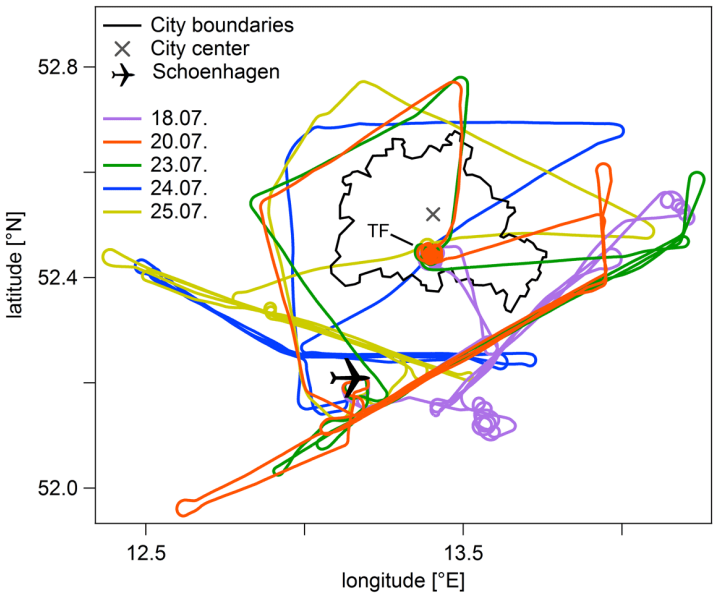


Figure 4.1: Flight tracks of all DLR Cessna research flights within the $[UC]^2$ campaign. Respective upwind and downwind transects as well as the vertical spiral flown over Tempelhofer Feld are visible for all flights. The flight used for this work is the one performed on July 20th. Figure taken from Klausner et al. (2020) under CC BY 4.0 license.

These different approaches yield emissions of 2.12 tCO_2/s for the first and 1.39 tCO_2/s for the second approach. The additional sources of uncertainty are wind direction (±2°), wind velocity (±0.3 $m\,s^{-1}$), and the PBL uncertainty, which is assumed to be between 5 % and 10 %. The flux uncertainties resulting from the combination of these different uncertainties are 52, 15, and 9 %. Klausner et al. (2020) then use the flux estimate of the HYSPLIT background method and include the linear background estimations only in the uncertainty estimation, which yields a final CO_2 flux estimate of $1.39 \pm 0.76\,tCO_2/s$, which is equivalent to $44 \pm 24\,Mt\,a^{-1}$.

This flux estimate is approximately double the Berlin city emissions derived from the CAMS-REG and the EDGAR inventories but agrees within the uncertainty. Sampling the inventories by weighting them with the footprints of the measurements, the estimates get closer. This indicates a two-fold limitation of the mass balance method. Firstly, the area this method is sensitive to cannot be exactly constrained to just the administrative boundaries of a city. Secondly, the uncertainty of the emissions estimate is quite large. Both of these limitations are inherent to the method and cannot be easily resolved. This is why we use a different approach in the following — Bayesian inversion.

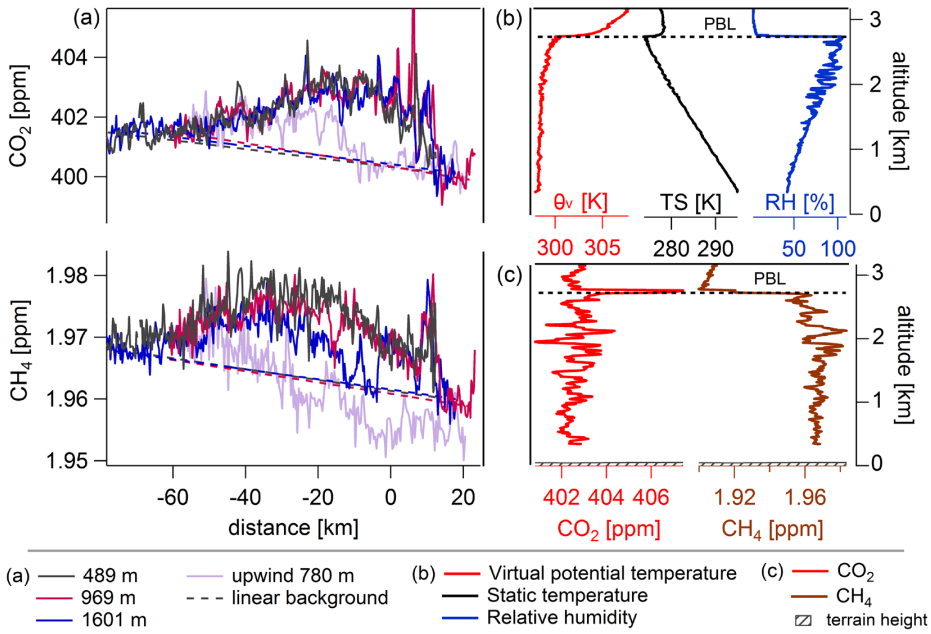


Figure 4.2: GHG mixing ratios and meteorological variables measured during the vertical profile over Tempelhofer Feld of the July 20th flight. Upwind and downwind mixing ratios of CO_2 and CH_4 in Panel a). The x-axis is distance of measurement to point 'A' at 52.41 °N and 13.95 °E, the curve in the downwind flight track. Upwind transect is projected onto downwind wall using HYSPLIT simulations. Dashed lines are backgrounds estimated as described in Klausner et al. (2020). Panels b) and c) depict vertical profiles of meteorological variables and trace gases. Figure taken from Klausner et al. (2020) under CC BY 4.0 license.

4.3 Bayesian Inversion Framework

We now extend the Bayesian inversion as presented in Section 4.1.2 for our problem of estimating GHG emissions of Berlin. In our case, y is a vector of the in-situ aircraft measurements and x is a vector representing the system state we want to estimate. The forward model F describes the relationship between our system state and the measured concentrations. Normally, these measurements are defined as enhancements against some background which is subtracted from the measurements in advance. However, we integrate this background into our forward model and system state.

The forward model F has to linearize the relationship between emissions, background, and measurements driven by meteorological transport. This is usually done by using an LPDM to calculate 'footprints' of the measurements (cf. Section 4.1.2). These footprints describe the sensitivity of the concentration measurements to the emissions across our spatial domain in a linear fashion. We aggregate these emissions from our simulation resolution into larger cells, striking a balance between the fidelity of our system and the information content afforded by the measurements. This is the first part of our forward model and yields enhancements which are caused by the emissions.

The second part of our forward model is an added background concentration value for every measurement. Since the MACRO-2018 dataset includes forward modeled concentrations, we can use the CAMS background as a first estimate of this inversion-specific background. The reason this remains a first estimate is that this CAMS background is only transported from outside the Europe domain. The background concentrations we want to estimate in the inversion, however, also include a part caused by fluxes between the edge of the footprints and the boundary of the Europe domain. These enhancements have to be added by the inversion.

Our forward model is, thus, a matrix of shape $m \times (t \cdot c + m)$ with m being the number of measurements, t the number of time steps, and c the number of emissions cells. This matrix projects between the state space of size $t \cdot c + m$ with the additional m elements being the background concentrations and the measurement space of size m. Thus, our version of Eq. (4.3) looks like:

$$\begin{pmatrix} m_0 \\ m_1 \\ \vdots \end{pmatrix} = \begin{pmatrix} f_{m0,c0} & f_{m0,c1} & \cdots & 1 & 0 & \cdots \\ f_{m1,c0} & f_{m1,c1} & & 0 & 1 & \\ \vdots & & \ddots & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} e_{c0} \\ e_{c1} \\ \vdots \\ b_{m0} \\ b_{m1} \\ \vdots \end{pmatrix} + \varepsilon$$

with ε being the measurement uncertainties and $f_{mi,cj}$ the footprint fractions of state vector cell c_j contributing to measurement m_i. The state vector elements for our inversion are the emissions e_{cj} of cell c_j and background concentrations of measurements m_i. A similar approach was also used by Lopez-Coto et al. (2020).

4.4 Optimization of Emissions

Now, we use the MACRO-2018 dataset to optimize a bottom-up emissions estimate for Berlin using measurements from the $[UC]^2$ aircraft data. As a first step, we evaluate the simulation fidelity of concentrations and meteorological variables to the measured data in order to gauge the suitability of our dataset for this task.

4.4.1 Simulation Evaluation

We start our evaluation by sampling the MACRO-2018 dataset along the flight track in time and space and comparing the extracted data to the measurements. Figure 4.3 depicts the CO_2 concentrations sampled from the MACRO-2018 dataset compared to the aircraft measurements. Panels a) and b) display the concentrations using the VPRM v1 and v2 biogenic fluxes respectively. As described in Section 3.4.3, the VPRM v2 emissions capture the summer daytime CO_2 uptake better. This is why they have a smaller mean bias (0.5 and 0.7 ppm) than the VPRM v1 data (1.7 and 1.9 ppm). However, after subtracting this bias, both VPRM versions exhibit a similar fidelity across the transects, as evidenced by the similar cMAB (corrected MAB) values (0.5 ppm for v1 and 0.6 ppm for v2). When comparing the two physics configurations we find the MYJ configuration to be exhibiting a slightly lower mean bias by 0.2 ppm indicating a model uncertainty of at least that magnitude. The largest deviation of the simulated concentrations from measurements occurs during the vertical profile, which was flown over the Tempelhofer Feld.

The flight path of the aircraft is displayed in Panel a) of Fig. 4.4. Here, we can clearly make out the corkscrew pattern flown for the vertical profile as well as the upwind and three downwind transects. The other panels in this figure display the comparison of the aircraft measurements to simulated variables extracted from a column of the simulation at 12:15 above the center of the spiral. Some residuals of the spiral flight pattern are visible in the wind velocity and CO_2 concentration fields. However, all in all (neglecting fine-scale structures) there is good agreement between the simulations and measurements up to 2 km height.

Above 2 km, deviations increase (cf. Table 4.1). These deviations stem from the measurements showing a strong capping inversion at the top of the boundary layer (2737 ± 2 m as estimated from vertical profile by Klausner et al., 2020). This capping inversion is typical

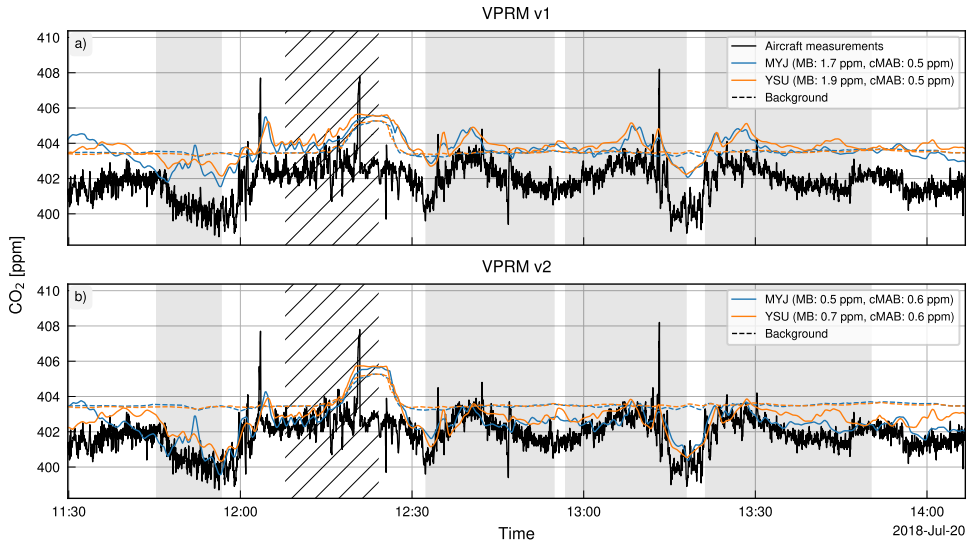


Figure 4.3: Comparison of aircraft measured CO_2 concentrations to concentrations extracted from MACRO-2018. Panel a) displays VPRM v1 and Panel b) VPRM v2 biospheric emissions. Black line represents aircraft measurements, the blue and orange lines the MYJ and YSU configurations. Hatched background marks vertical profile flown over Tempelhofer Feld, solid backgrounds mark transects. Transects are first upwind (780 m height), then three downwind transects (489, 969, and 1601 m height). The cMAB statistic is the bias-corrected MAB. Aggregated statistics are calculated across the transects only.

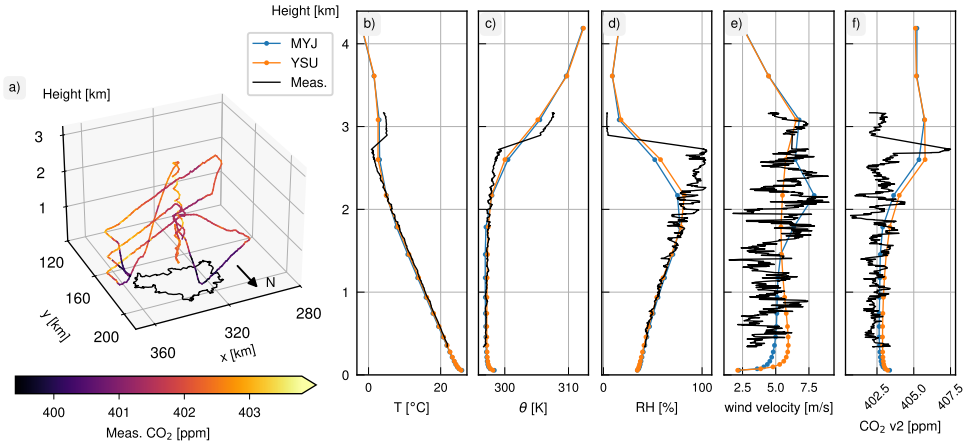


Figure 4.4: Display of aircraft track and comparison of vertical profile flown over Tempelhofer Feld to MACRO-2018 column. Panel a) displays the aircraft flight track and the CO_2 concentration measurements. Clearly visible are the upwind transect NW of Berlin, the corkscrew vertical profile and the three downwind transects SE of Berlin. Panels b) to f) show air temperature, potential temperature, relative humidity, wind velocity, and CO_2 concentrations compared to aircraft measurements. While the aircraft measurements exhibit a very clear inversion at the top of the boundary layer, the vertical resolution of our simulation is too coarse to resolve this.

Table 4.1: Mean absolute bias of variables in MACRO-2018 compared to flight measurements for the vertical profile on July 20th 2018, separated into values below and above 2 km altitude. All variables show better matches at height except for wind velocity, which is more strongly influenced by the surface.

	below 2km		above 2km	
	MYJ	YSU	MYJ	YSU
T [°C]	0.5	0.3	1.2	1.1
θ [K]	0.2	0.1	1.3	1.2
RH [%]	3.4	3.4	23.0	20.8
wind velocity [m/s]	1.3	1.4	1.1	1.1
CO_2 v2 [ppm]	0.6	0.7	1.9	2.1

for urban boundary layers and separates the mixing layer below from the free atmosphere above (cf. Section 2.1.1). Our simulations, however, exhibit a lower PBL height (MYJ: 1900 m, YSU: 2250 m; from `PBLH` diagnostics) and a less strong inversion (as visible in Fig. 4.4, Panel c). A partial explanation is the much lower vertical resolution at these heights contributing to more smoothed-out variables. However, the main features of the variables are well-captured. The temperature, potential temperature, and relative humidity profiles are similar to measurements except for the lower simulated capping inversion. The CO_2 concentrations in the boundary layer are close to the measurements as well, except for the peak caused by the capping inversion at the top of the PBL, which is smeared out in the simulation. Above the PBL height, where air masses belong to the larger background, however, the CO_2 concentrations are overestimated. This indicates that the model background we use as background concentration prior might not be perfect and will need to be optimized in the inversion itself.

Comparing the two physics configurations, we find a similar performance of MYJ and YSU. The CO_2 concentrations within the boundary layer of the MYJ simulation are a bit lower compared to YSU. This could be caused by the lower boundary layer height estimation reducing the volume where CO_2 uptake takes place. Compared to CO_2, the deviations of wind velocity are larger. Here we find the clearest differences between the MYJ and YSU configurations. The lower boundary layer height notwithstanding, for wind velocities the MYJ configuration is more in agreement with measurements than YSU (cf. Table 4.1). Therefore, we will continue with the MYJ configuration.

In conclusion, we use the MYJ configuration and the VPRM v2 emissions for our Bayesian inversion. Following from the above analysis, we can conclude that the meteorology and emissions used in the MACRO-2018 dataset represent a solid basis for our inversion.

4.4.2 Inversion Setup

We now use the meteorological data from the MACRO-2018 dataset and the same prior emissions input files we used to calculate the concentrations in MACRO-2018 in order to run our Bayesian inversions. The measurement footprints will provide the link between emissions and measured concentrations (cf. Section 4.3). For the integration of the meteorological fields, we use FLEXPART-WRF as described above (cf. Section 4.1.3). The spawn points of the footprints are evenly spaced at 5 min intervals and integrated backwards for 24 h. Along the transects, this 5 min spacing corresponds approximately to 10 km distance. These footprints were kindly provided by Christopher Lüken-Winkels.

As we do not have enough information to invert every one of the simulation grid cells on 1×1 km resolution, they have to be aggregated into larger state vector elements. Since this aggregation is already a constraint we place on the system, we aim to make it as task-driven as possible. Our objective is the Berlin city emissions, so one state vector element includes the administrative boundary of Berlin. The other elements are border cells and we divide them according to the footprint influence. We separate the elements at the WRF domain boundaries in order to distinguish between near- and far-field contributions. Furthermore, we divide them into upwind and downwind elements as the sensitivity of the measurements varies starkly between them. This results in five emissions state vector elements which we use in our inversion. The exact shape and size along with the flight track and footprint release points is displayed in Fig. 4.5. We use the measurements taken during the one upwind and three downwing transects, which were flown at 780, 489, 969, and 1601 m altitude. We exclude the measurements at 11:45 UTC and 13:45 UTC due to data quality issues.

The second part of the forward model is the background (cf. Section 4.3). Here, we can leverage our Eulerian background concentrations which are part of the MACRO-2018 dataset. We sample the `CO2_BCK` field at the measurement times and locations used for the footprints and use this as our background prior. Without the forward modeled CO_2 concentrations, in a purely Lagrangian approach, this would be more difficult. It would require a background calculation based e.g. on footprints at downwind locations or background measurements, both of which may be associated with large uncertainties.

The inversion additionally requires the compilation of uncertainty information in the covariance matrices. Their design is important for the inversion result (cf. Section 4.1.2) as they constrain the state vector optimization. The two matrices we have to construct are the prior covariance matrix and the model-data mismatch covariance matrix (cf. Eq. (4.6)). We will focus on this next.

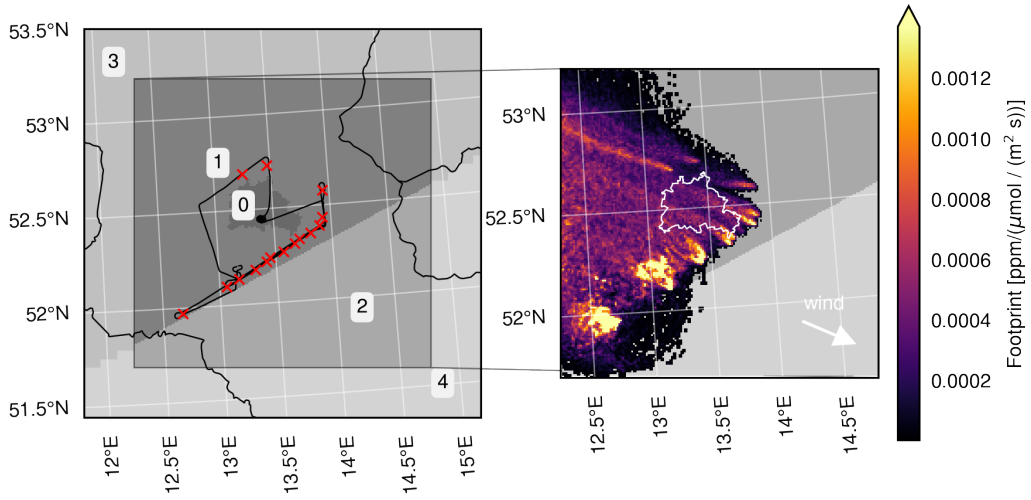


Figure 4.5: Setup of the emissions elements of the state vector. Panel a) displays the Berlin domain and parts of the Germany domain (note the lower resolution). These domains are split into five state vector elements. The state vector elements are Berlin city, Berlin upwind, and Berlin downwind as well as Germany upwind and Germany downwind (in order). Furthermore, the flight track as well as the footprint releases are displayed. The first upwind measurement (11:45 UTC) and one of the later downwind measurements (13:45 UTC) had to be excluded due to data quality issues. Panel b) illustrates the surface level footprints of the selected measurements and the Berlin administrative boundary.

Prior Covariance

The prior covariance matrix consists of both, the prior emission and the prior background covariances. Where possible we present a data-driven estimation of the associated uncertainties and correlations. We first discuss the prior emissions uncertainties before turning to the background in the second part.

Emissions uncertainty The emissions uncertainties arise from the underlying emissions inventories, in this case both TNO and VPRM. However, we do not have enough information about either inventory to construct our prior covariance purely from the inventory uncertainties. One effect, which should be similar for both emissions inventories, though, is the scaling of the uncertainty with the cell size. Since continental-scale biogenic fluxes and national-level anthropogenic emissions are much better constrained than their gridded variants at 1 km resolution, larger cells should have smaller uncertainties than smaller ones.

When inquiring about inventory uncertainties, the inventory developers, Dr. Ingrid Super and Hugo Denier van der Gon, provided the uncertainty for the spatial distribution and the national emissions of an inventory created in the context of the AVENGERS project. These data are provided on an approximately 7 km resolved lat/lon grid. We aggregate these to our state vector elements by using sum of squares. As uncertainties in temporal and height distribution of emission are not included in this estimate, this estimate cannot be used to determine the total uncertainties. We can, however, still use it to get the relative difference in uncertainty between state vector elements of different sizes. In effect, we will scale the uncertainties of the other state vector elements relative to element zero (Berlin city). Doing this, state vector elements one and two (Berlin domain) get scaled by two-thirds, while the latter two (Germany domain) get scaled by one tenth due to their larger areas. To get a final uncertainty for the inversion, we now only need to set a sensible value for the Berlin state vector element.

As we lack the data to directly calculate the uncertainty, we do an investigation of the numerical stability and the sensitivity of the posterior emissions estimate for Berlin to the prior uncertainty. Our goal is to have an emission estimate that is as measurement-informed as possible by having a weak prior constraint, while still being physically sound. This means we run inversions with various relative emissions uncertainties for the Berlin element (25, 50, 75, 100, and 125 %). Evaluating their averaging kernels, we see that at uncertainties higher than 75 %, there are off-diagonal elements which exceed the lower bound of negative one, which is considered physically sound. Therefore, we use 75 % as our emissions uncertainty for Berlin, which gets downscaled to the other state vector elements as shown above.

Background uncertainty The background of our inversion includes all concentration contributions that are outside of the footprint. To estimate the background uncertainty from outside the Europe domain we use the MACRO-2018 dataset's `CO2_BCK` field. For each measurement, we can sample the background concentrations (from outside the Europe domain) for 75 min around the measurement itself. We choose this time period as this is approximately the time needed by the aircraft for all three transects. We extract the background variability for each measurement (standard deviation). To additionally account for inner-European emissions that are not covered by the footprint, we increase this by 0.5 ppm.

Correlations of the far-field contribution to the measurements are expected to be large. We extract the correlations of the sampled time series and use it as a lower bound. Then we increase the correlation such that the local CO_2 signals do not get imprinted onto the posterior background, meaning that the correlation between the background and the measurements is not changed by the inversion.

Model-Data Mismatch

The model-data mismatch includes the covariance of the model concentrations and the measurements. Here, the uncertainties include the uncertainty of the measurement instrument and that of the model transport. We construct our model-data mismatch error as in Boschetti et al. (2018), who use:

$$\mathbf{S}_e = \mathbf{C}_s \mathbf{C}_t \varepsilon_{\text{tran}}^2 + \varepsilon_{\text{meas}}^2$$

with $\mathbf{C}_s$ being inter-species covariance (which we do not have), $\mathbf{C}_t$ being temporal correlation, $\varepsilon_{\text{tran}}$ being transport error and $\varepsilon_{\text{meas}}$ being measurement error. Since we explicitly sample non-homogeneous air masses (plume transects) and work on urban scales and not on continental ones like Boschetti et al. (2018), we do not assume any temporal correlations of our measurements. Since the MACRO-2018 dataset includes two simulations with different physics configurations, we use this mini-ensemble to estimate the transport error. As such, we use the mean absolute differences between the `CO2_TOTAL_V2` concentrations for the MYJ and YSU configurations as an estimate which are on average 0.49 ppm. This then is combined with the instrument measurement uncertainty of 0.15 ppm using sum of squares to a total uncertainty of 0.51 ppm.

4.4.3 Emissions Estimate

With our uncertainties contsructed, we can now turn to the emissions estimate. Here, we try to include the uncertainty associated with the choice of prior covariance into our estimate. In

Table 4.2: Details of inversion result. Change in emissions and error reduction are both relative to prior. All values of the averaging kernel diagonal are below one and the first two are close to one suggesting high performance.

	Prior emissions [Mt a^{-1}]	Posterior emissions [Mt a^{-1}]	Change in emissions [%]	Error reduction [%]	Averaging kernel [1]
Berlin city	9.1	16.5	82.0	36.6	0.60
Berlin upwind	−34.2	−23.1	32.5	57.7	0.82
Berlin downwind	4.1	4.1	−0.6	0.2	0.00
Germany upwind	−169.7	−163.4	3.7	2.5	0.05
Germany downwind	−1908.0	−1907.5	0.0	0.0	0.00

order to gauge the dependence of our inversion output to this choice, we perform a small sensitivity study where we vary the prior uncertainty between 50, 75, and 100 %. For the chosen prior covariance of 75 %, our Bayesian inversion yields an emissions estimate of $16.5 \pm 4.3\,\mathrm{Mt\,a^{-1}}$. The maximum change of the emissions estimates compared to using 50 % or 100 % is $2.55\,\mathrm{Mt\,a^{-1}}$. We add this amount using sum of squares to the posterior uncertainty, which yields the final emissions estimate of $16.5 \pm 5.0\,\mathrm{Mt\,a^{-1}}$.

Figure 4.6 compares this estimate to the mass balance estimate from Klausner et al. (2020), the average daily Berlin emissions from the self-reported inventory, TNO, and the CAMS and EDGAR inventories also used in Klausner et al. (2020). We can clearly see that the Bayesian inversion is close to the average daily emissions from inventories, indicating a realistic estimation. Additionally, the uncertainty has been reduced from $24\,\mathrm{Mt\,a^{-1}}$ for the mass balance approach to $5.0\,\mathrm{Mt\,a^{-1}}$ for our Bayesian inversion. This is a decrease by nearly a factor of five.

Figure 4.7 and Table 4.2 give an overview of the inversion result. In Fig. 4.7, we see an overview over the Bayesian inversion and some details on its performance. It shows that the error reduction is the largest for the Berlin city and the Berlin upwind cell (36.6 and 57.7 %) while the other cells' error reductions are two orders of magnitude smaller. These are also the cells whose emissions got adjusted the most (82.0 and 32.5 %). This is expected, as the Berlin city and the Berlin upwind cell are the ones the inversion is most sensitive to as seen by the largest footprints. Turning our attention to the averaging kernel (Table 4.2), we see that the elements on the diagonal are all close to and below one, which is good. However, we also see large off-diagonal elements between the Berlin city (zero), Berlin upwind (one), and Germany upwind (three) cells (cf. Fig. 4.7). These off-diagonal elements indicate that the emissions estimates of the Berlin city cell and the upwind cells are not independent of one another. This is caused by the measurement footprints being sensitive to all three cells simultaneously. In effect, there are just too few measurements and meteorological situations to perfectly separate the Berlin city cell from the surrounding cells.

The average posterior background concentrations get adjusted downwards from the prior ones of 403.5 ppm to 402.3 ± 0.3 ppm. To ensure that the measurements do not unduly influence the background, we check the posterior background–measurement correlation and find it to be the same as the prior background–measurement one. This indicates that the inversion is not over-fitting the background to the measured concentrations. The direction of this change in background concentration is expected since the emissions between the footprint boundary and the Europe domain boundary, which we expect the inversion to add, are mostly negative. Additionally, the value the posterior background concentrations are corrected to (402.3 ± 0.3 ppm) matches closely to the median of the observed CO_2 concentrations above the PBL with 402.6 ppm (cf. Fig. 4.4). We expect these concentrations above the background to constitute the far-field background of our measurements, since the near-field emissions influence is mostly limited to the mixing layer. This adjustment by our inversion is remarkable, as we only use measurements well within the PBL up to 1.6 km. This notwithstanding, the inferred posterior background is very close to the observed background at 3 km height. This validation with independent data speaks to the robustness of our method.

In order to test the importance of assimilating the background and the emissions at the same time, we conduct an experiment where we set the prior background uncertainty to zero. This means that the background does not get changed at all by the inversion and just adds an offset to the measurements. Under these conditions, the inversion returns a posterior emissions estimate of $10.6 \pm 4.0\,\mathrm{Mt\,a^{-1}}$. These emissions are 40 % below our estimate considering the background concentrations. Thus, assimilating the background concentrations is crucial as it avoids systematic errors implicit in the inversion construction.

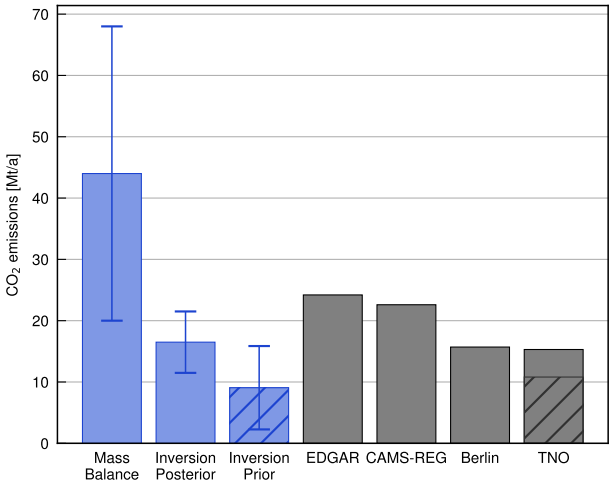


Figure 4.6: Comparison of emissions estimate from mass balance ($44 \pm 24\,\mathrm{Mt\,a^{-1}}$) and Bayesian inversion ($16.5 \pm 5.0\,\mathrm{Mt\,a^{-1}}$). Prior of the Bayesian inversion is displayed with $9.1 \pm 6.8\,\mathrm{Mt\,a^{-1}}$ (including $1.7\,\mathrm{Mt\,a^{-1}}$ biogenic uptake). Error reduction of 36.6 % for the Berlin state vector element is clearly visible. This is compared to EDGAR, CAMS-REG, the self-reported and the TNO average yearly emissions for the city of Berlin for 2018 (24.2, 22.6, 15.7, and $15.3\,\mathrm{Mt\,a^{-1}}$). Hatched part of the TNO inventory bar is the Berlin emissions after applying the time factors, which is the anthropogenic part of the prior emissions ($10.8\,\mathrm{Mt\,a^{-1}}$).

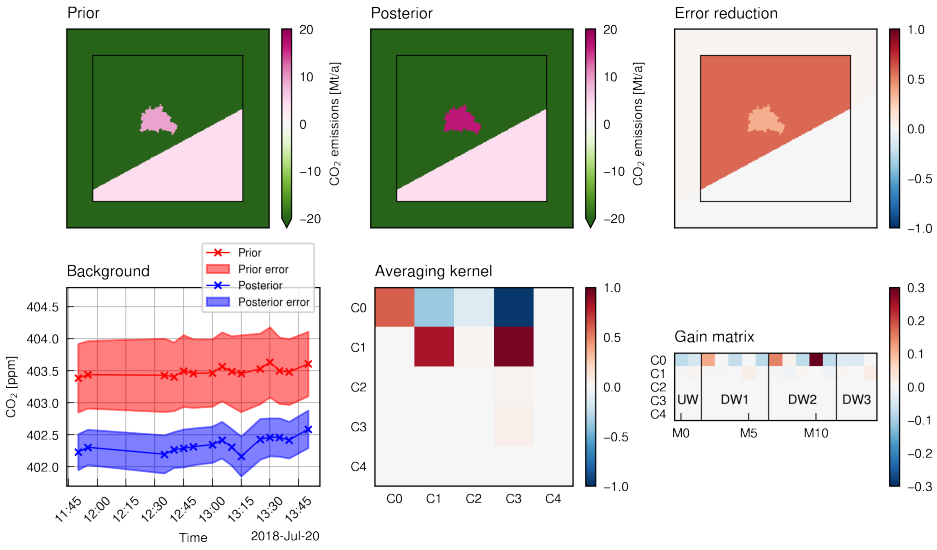


Figure 4.7: Averaging kernel emissions part shows positive and below one diagonal and less ideal off-diagonal elements. Gain matrix illustrates the large influence of the measurements (M0-M14) on the Berlin city emissions estimate and the smaller influence on the Berlin upwind cell. This is reflected in the relative emissions adjustment from prior to posterior. Transects are marked by 'UW' (upwind) and 'DW' (downwind) labels.

4.5 Discussion

In this chapter we have improved an emissions estimate for the city of Berlin based on aircraft measurements. The original estimate using a mass balance method was presented by Klausner et al. (2020) and resulted in emissions of $44 \pm 24\,\mathrm{Mt\,a^{-1}}$. As reported by Klausner et al. (2020), this estimate matches closer to the inventories when sampling the inventories using the footprints of the measurements and not the administrative boundary of Berlin.

We have now used the meteorology and concentration fields provided by the MACRO-2018 dataset and have performed a Bayesian inversion, yielding a posterior emissions estimate. This inversion is based on an LPDM transport linearization and explicitly includes the background concentrations of each measurement as state vector elements. We have used the concentration fields included in the MACRO-2018 dataset in order to get first-order estimates of the backgrounds. Furthermore, we have constructed the uncertainties based on data and, where necessary, expert judgement. We have performed a sensitivity study in order to reduce the impact of arbitrary choices on our result and have added the difference in posterior emissions on top of our uncertainty. Using this approach, we get an emissions estimate of $16.5 \pm 5.0\,\mathrm{Mt\,a^{-1}}$.

Our posterior emissions estimate increases the total prior emissions of $9.1 \pm 6.8\,\mathrm{Mt\,a^{-1}}$ by approx. 80 %. This speaks to there being either less biogenic uptake or more anthropogenic emissions than estimated in our inventories. However, the posterior emissions are already larger than just the anthropogenic part of the prior emissions (hatched part of TNO bar in Fig. 4.6; $10.8\,\mathrm{Mt\,a^{-1}}$). Since we can most likely neglect biogenic respiration as a significant source of CO_2, our findings point to the anthropogenic emissions being higher than reported in the TNO inventory. This can either be caused by a misestimation of total emissions of some sources in the inventory or by the time factors, which get applied to a whole emissions sector across all of Europe, being wrong in this particular area. Another possible explanation are missing emissions sources like human respiration.

The posterior emissions estimate of the Berlin state vector element shows an error reduction of 36.6 %. This indicates that the measurements we use in the inversion significantly reduce the emissions uncertainty. However, the largest error reduction in our state vector is in the Berlin upwind cell with 57.7 %. This is due to the measurements having a larger sensitivity to the Berlin upwind emissions than the actual city emissions. Said sensitivity was also found and discussed in Klausner et al. (2020). Klausner et al. (2020) compute footprints for the aircraft measurements and find that they are sensitive to large areas outside the administrative boundaries of Berlin. Our method, however, is able to leverage the footprints to restrict the emissions estimate to the Berlin administrative boundaries. We, furthermore, achieve

a posterior emissions uncertainty which is a factor of five smaller than the one of the mass balance method.

While our emissions estimate is obviously only valid for the 20th of July 2018, in order to stay comparable to Klausner et al. (2020) we also compare it to the average daily emissions reported in the inventories. Our posterior emissions estimate fits the annual 2018 emissions reported in the TNO GHGco v4.1 inventory (Super et al., 2020) and the self-reported Berlin 2018 emissions estimates (Amt für Statistik Berlin-Brandenburg, 2025) of 15.3 and 15.7 Mt a^{-1} within the uncertainty. It is lower than the emissions reported in the EDGAR 2015 and the CAMS-REG 2015 inventory presented in Klausner et al. (2020). Our emissions estimate being in between the inventory estimates indicates that it is realistic. However, it should be reiterated that the inventory estimates represent the average daily emissions of Berlin based on the total 2018 emissions, and not the estimates for this particular day. In order to be able to make judgments about the quality of the inventories' emissions estimates, more flights are needed.

We have shown that our Bayesian inversion framework, assimilating emissions and background concentrations at the same time, is robust. The background estimation is able to reconstruct a background concentration from above the PBL while only assimilating measurements from within the PBL. This is good evidence not only for our method working but also for the usefulness of forward modeled datasets like MACRO-2018, without which this would not have been possible. This approach of combining forward concentration modeling and emissions estimation using Bayesian inversions has the potential to help future measurement campaigns extract the maximum possible amount of information about emissions from their measurements.

5 Conclusion and Outlook

Emissions monitoring, verification, and reporting is a key component of global climate action and helps hold governments accountable to their climate promises. Germany's contribution to the IG3IS coordinating these efforts internationally is the ITMS. In the context of the ITMS, we focus on urban monitoring as emissions sectors associated with urban areas are responsible for 31 % of the CO_2 emissions uncertainty in the German NIR. Reliable, high-quality emissions quantification requires high-performance modeling which is validated against independent data and an inversion framework which is constructed to treat sources of error and biases.

This thesis presents the development of the MACRO-2018 dataset and its use for emissions estimation in urban areas. We present the setup of WRF and the corresponding sensitivity studies investigating its meteorological performance in the Rhine-Main-Neckar area, as published in Pilz et al. (2026). This investigation forms the basis of the WRF-Chem simulations which generate the high-quality, long-term MACRO-2018 dataset published as Pilz et al. (2025a). Finally, this dataset is used to drive and constrain a Bayesian inversion optimizing CO_2 emissions using aircraft measurements over the city of Berlin.

The setup of our atmospheric transport model is presented in Chapter 2. Therein, we describe the establishment of a new WRF simulation setup on the supercomputer Levante at DKRZ. We conduct a sensitivity analysis of different physics configurations in order to optimize the atmospheric transport simulation for our use case. To this end, we evaluate the meteorological performance of 16 physics configurations at 19 meteorological stations in the Rhein-Main-Neckar area of Germany. This area not only provides a lot of meteorological stations to compare our simulation against, it is also a polycentric urban region (PUR) and as such representative of other European metropolitan areas like the Rhein-Ruhr area. We identify well-performing configurations, discuss possible reasons, and compare to previous studies, demonstrating that our simulation performance exceeds theirs.

This analysis is the basis for generating Metropolitan Area CO_x RecOrd of Germany 2018 (MACRO-2018; Pilz et al., 2025a), which we present in Chapter 3. It provides two separate physics configurations and biogenic models enabling first-order estimates of model uncertainty. We present its data consisting of hourly output at up to 1 km resolution over the full year of 2018 for meteorology as well as CO_2 and CO concentrations. Available CO_2

concentration fields include biogenic contributions, background as well as traffic, point, and area emissions driven by the VPRM and TNO inventories. We evaluate the CO_2 and CO concentration performance against high-precision Obspack measurements. To that end, we develop a new, robust, method of disambiguating different PBL development phases. This is necessary as studies usually only use a fixed time window of, for example 13 UTC to 17 UTC. We could show that this does not match with the underlying physical reality, leading to either useful data being discarded or data during unstable conditions being used. Our new definition now gives a dynamical definition of which data is useful and can be adapted by the wider community improving measurement filtering.

We benchmark MACRO-2018 against previous simulations of CO_2 concentrations and demonstrate it performing on par with or better than all previous studies we were able to find. By providing validated long-term, high-resolution meteorology as well as CO_2 and CO concentration fields over Europe of the highest quality, MACRO-2018 offers a powerful foundation for various kinds of scientific investigation. It is already being actively employed for multiple different applications. These include satellite mission planning, investigating the impact of biogenic models at sub-continental scale, the optimization of MRV systems of urban CO_2, and the optimization of urban CO_2 emissions as done in this thesis.

This high-quality data provides a solid basis for the optimization of urban CO_2 emissions. In Chapter 4, we use it to estimate CO_2 emissions of Berlin using aircraft measurement data collected in the context of the $[UC]^2$ campaign. Since these aircraft data provide important information but are also expensive to generate, we build a Bayesian inversion framework to extract the maximum possible amount of information from these measurements. This framework allows us to optimize CO_2 emissions and CO_2 background values simultaneously. The meteorological data of MACRO-2018 is used to drive the LPDM FLEXPART, and its CO_2 concentration data is used for prior estimates of background concentrations. Using this approach, we reduce the emissions uncertainty over a previous emissions estimate based on the same measured data using a mass balance approach from $24\,\mathrm{Mt\,a^{-1}}$ to $5.0\,\mathrm{Mt\,a^{-1}}$. Furthermore, we find the optimized background concentration values to closely match measured concentrations outside of the PBL, unseen by the inversion. The value of our optimized posterior emissions being in between the average daily emissions reported by multiple different inventories indicates its realism, while the reproduction of unseen background concentrations demonstrates the robustness of our method. We find CO_2 emissions for Berlin of $16.5 \pm 5.0\,\mathrm{Mt\,a^{-1}}$, which is 8 % above daily average emissions from TNO and 32 % below EDGAR. The Bayesian inversion framework we have developed is ready-to-use for future measurement campaigns. This approach of combining forward concentration modeling and emissions estimation using Bayesian inversions has the potential to help future

campaigns extract the maximum possible amount of information about emissions from their measurements.

There are multiple avenues for future developments to build upon this work. In order to further improve the meteorological modeling, larger ensemble sizes using more physics configurations could be used. These configurations should include previously untested PBL schemes like, for example, MYNN (Nakanishi and Niino, 2009) or Shin-Hong (Shin and Hong, 2015). Additionally, comprehensive investigation into the observed underestimation of wind velocities when using the BEP urban parametrization should be carried out, as this is a long-standing issue. Furthermore, new meteorological models are emerging showing strong performances, which partially exceed WRF. Therefore, switching to a different underlying model like ICON-ART, whose driver ICON shows high fidelity (Magnusson et al., 2022; Loprieno et al., 2026) could be considered. However, especially the combination of urban parametrization and tracer dispersion in these models should be investigated.

A new version of the MACRO-2018 dataset should include a larger meteorological ensemble. One should investigate the possibility of adding CO chemistry to the simulation in order to improve CO concentration fidelity. Furthermore, we clearly show the importance of the biogenic models to simulation fidelity. While their optimization is already ongoing and MACRO-2018 is actively supporting this development, this is an avenue of research which should be prioritized. Regarding the anthropogenic emissions, a future version of MACRO-2018 could also include more emissions inventories like EDGAR or CAMS to have an ensemble of anthropogenic influences. One should furthermore consider including CO_2 fluxes from human respiration, as they represent a significant part of total CO_2 fluxes in urban areas (7 % to 18 % Stagakis et al., 2023; Dröge et al., 2024).

This work has paved the way to improve emissions estimation based on future aircraft campaigns. Increasing the amount of flights will lead to a higher temporal representativeness of the emissions estimate, which is important for emissions accounting on the annual level. It will also reduce the underconstraint of the spatial dimension evident in the averaging kernel. Our Bayesian inversion framework is able to be extended to support the disaggregation of anthropogenic and biogenic emissions using a multi-species approach including CO and isotopologues of CO_2. Finally, beyond aircraft measurements, different kinds of observations should be investigated as tools for long-term urban emissions MRV systems. Multiple new satellites include target modes, which enable the observation of areas of interest at very high spatial resolutions on the kilometer scale (Guanter et al., 2015; Taylor et al., 2020; Cogliati et al., 2021; Tanimoto et al., 2025). Using these for urban areas can increase the coverage and spatial disaggregation of emissions, opening the door to operationalized urban emissions accounting. Further important methods include in-situ CO_2 concentration measurement networks, which are already being deployed worldwide (Shusterman et al., 2016; Grange

et al., 2025; Kim et al., 2025; Aigner et al., 2025). These have the advantage of uninterrupted temporal coverage, high signal-to-noise ratios, and comparably simple characterization. However, since they are static, the structure of the network design has to be thoroughly studied.

In summary, this thesis establishes a robust framework for high-resolution, top-down urban CO_2 emissions estimation from model sensitivity study to inversion-based quantification. Through the development and validation of the MACRO-2018 dataset, we demonstrate that the careful comparison of simulated and measured concentrations is important for improvements in both models and emissions inventories. Through its application to Berlin, we show that rigorous model optimization allows quantifying model-data mismatch errors central to urban emissions estimation. Furthermore, we show that the use of a well-optimized atmospheric transport model and concentration dataset for emissions estimation is able to significantly reduce emissions uncertainties, contributing to independent emissions monitoring in the spirit of the Paris Agreement. Beyond its immediate results, the methods, datasets, and tools developed here form a flexible foundation for future work on regional atmospheric modeling and urban emissions monitoring. As cities become increasingly important to climate action, high-quality scientific evidence and analysis will play a central role in supporting credible emissions mitigation verification.

A Appendix

Sensitivity Studies

index	SLUCM levels	BEP levels	index	SLUCM levels	BEP levels
1	0 m	0 m	22	2869.2 m	4959.1 m
2	89.9 m	15.0 m	23	3167.0 m	5641.5 m
3	181.6 m	34.4 m	24	3522.9 m	6378.2 m
4	275.4 m	59.5 m	25	3965.9 m	7171.6 m
5	371.2 m	91.8 m	26	4549.8 m	8023.8 m
6	469.6 m	133.3 m	27	5324.6 m	8863.8 m
7	570.5 m	186.6 m	28	6345.8 m	9681.9 m
8	674.5 m	254.6 m	29	7475.0 m	10478.1 m
9	781.8 m	341.0 m	30	8559.3 m	11252.5 m
10	892.8 m	450.2 m	31	9612.4 m	12005.1 m
11	1008.0 m	586.9 m	32	10629.7 m	12735.8 m
12	1128.0 m	756.6 m	33	11611.3 m	13450.4 m
13	1253.4 m	964.6 m	34	12557.0 m	14165.1 m
14	1385.1 m	1216.2 m	35	13471.2 m	14879.7 m
15	1524.1 m	1515.6 m	36	14385.4 m	15594.3 m
16	1671.6 m	1865.4 m	37	15299.5 m	16308.9 m
17	1829.2 m	2266.1 m	38	16213.7 m	17023.5 m
18	1998.9 m	2715.6 m	39	17127.9 m	17738.1 m
19	2183.4 m	3209.3 m	40	18042.1 m	18452.7 m
20	2386.3 m	3745.8 m	41	18956.3 m	19167.4 m
21	2612.5 m	4328.1 m	42	19870.5 m	19882.0 m

Table A.1: Table of WRF level heights output by `real.exe`. Table taken from Pilz et al. (2026).

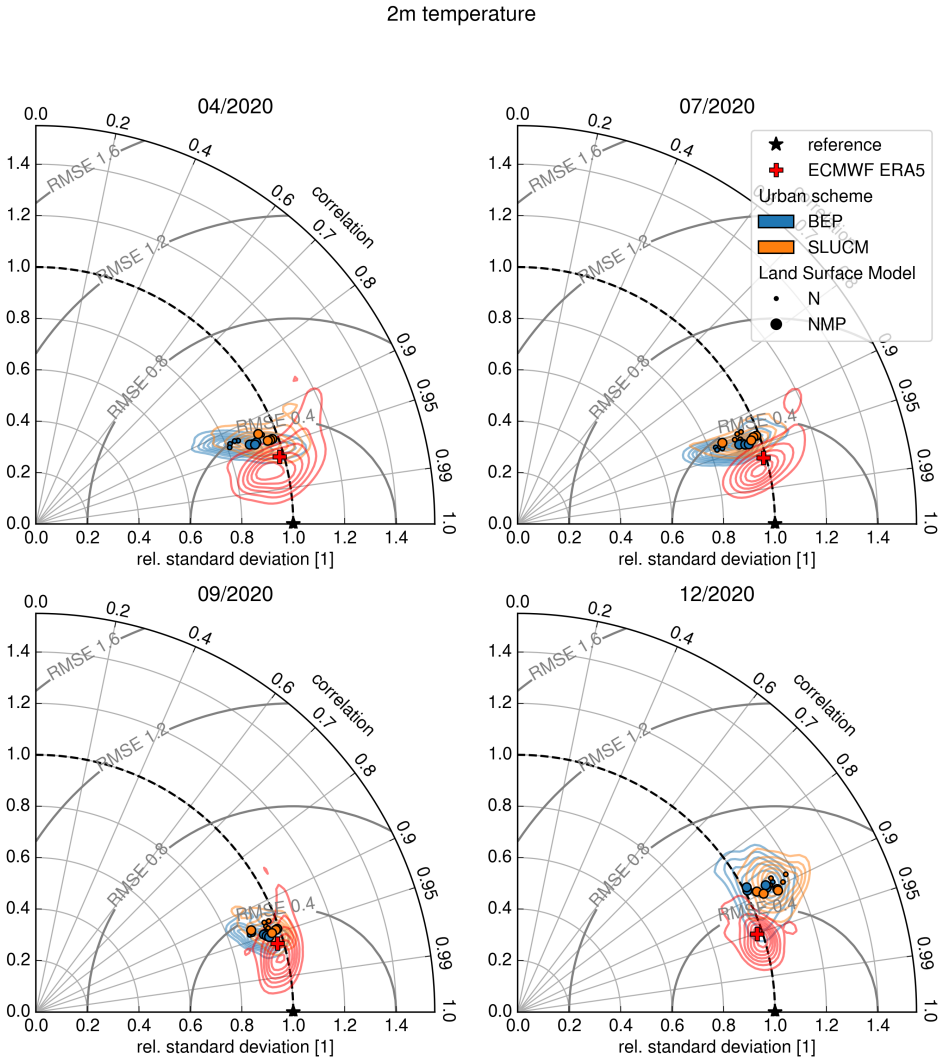


Figure A.1: Taylor density diagram (see Section 2.3.4) of 2 m temperature for all four simulated months of 2020. ERA5 data are show in red, configurations using SLUCM scheme in orange and configurations using the BEP scheme in blue. Configurations are distinguished into using Noah or Noah-MP LSM by different markers. Behind each marker is the average over one simulation configuration averaged over all DWD stations. Spread of all configurations and stations using the respective UCM scheme is shown as KDEs. There, some outliers namely the Stötten and Kleiner Feldberg/Taunus stations which are both at elevation are visible. Figure taken from Pilz et al. (2026).

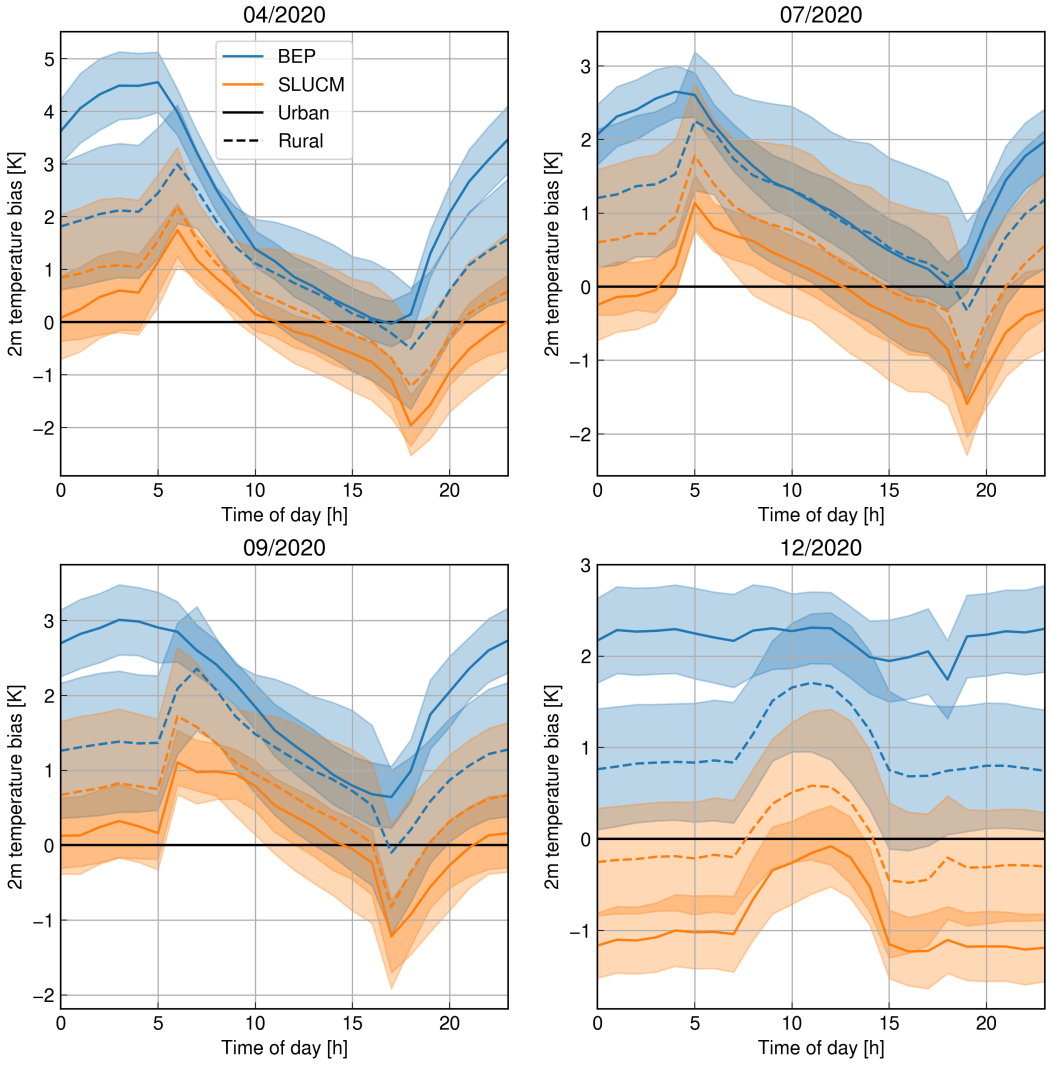


Figure A.2: Comparison of 2 m temperature bias for SLUCM and BEP runs stratified by urban and rural stations. Solid lines are average of urban stations, dashed lines are average of rural stations. Clearly, the BEP scheme shows a much larger difference of performance at urban vs at rural stations. Figure taken from Pilz et al. (2026).

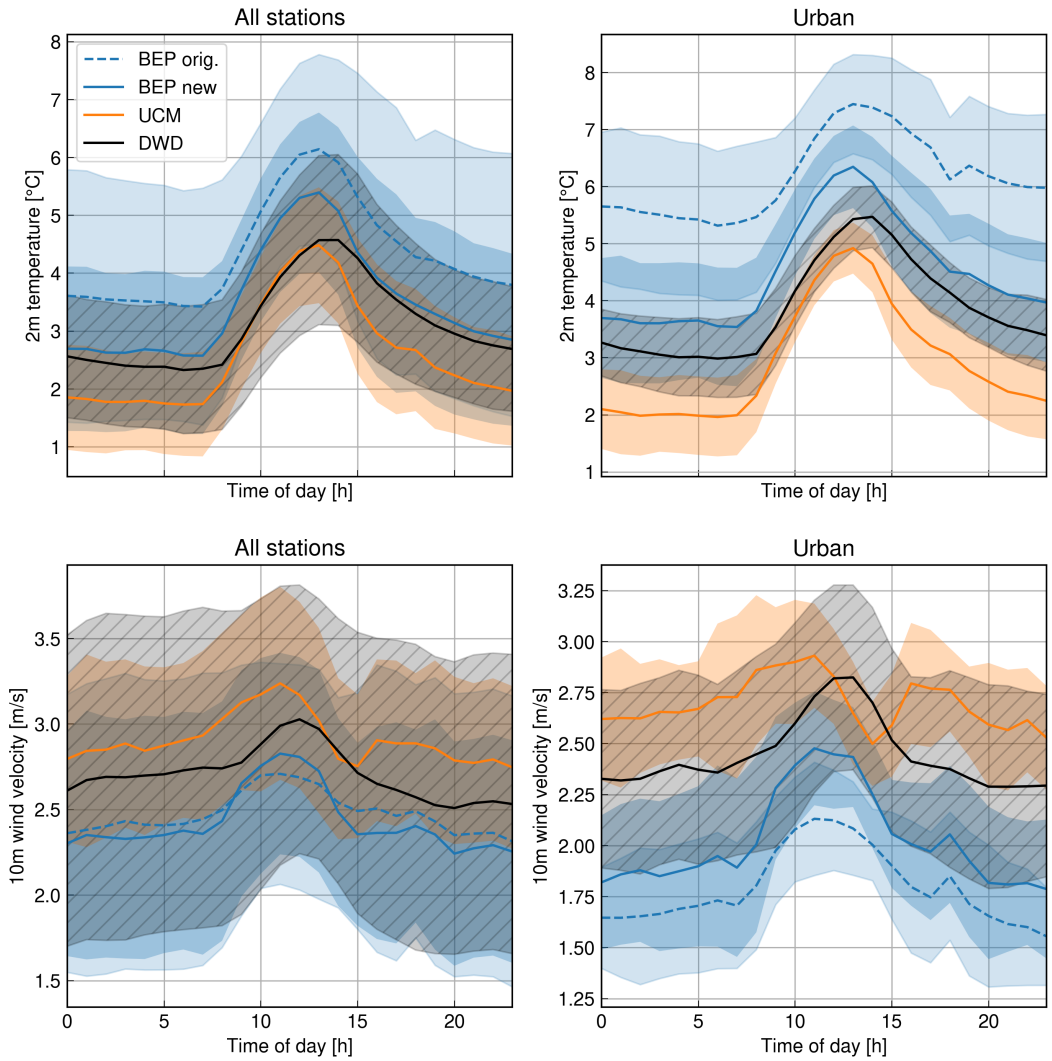


Figure A.3: Influence of vertical configuration on BEP 2 m-temperature and wind velocity. Vertical configuration was changed from the original BEP setup of the lowest layer at 15 m (dashed blue line) to the lowest layer at 90 m (solid blue line). Mean absolute biases improve for 2 m-temperature (Total: 1.4 K → 0.6 K, Urban: 2.3 K → 0.7 K) and remain approximately similar for wind velocity (Total: 0.7 m s^{-1} → 0.7 m s^{-1}, Urban: 0.7 m s^{-1} → 0.6 m s^{-1}). Figure taken from Pilz et al. (2026).

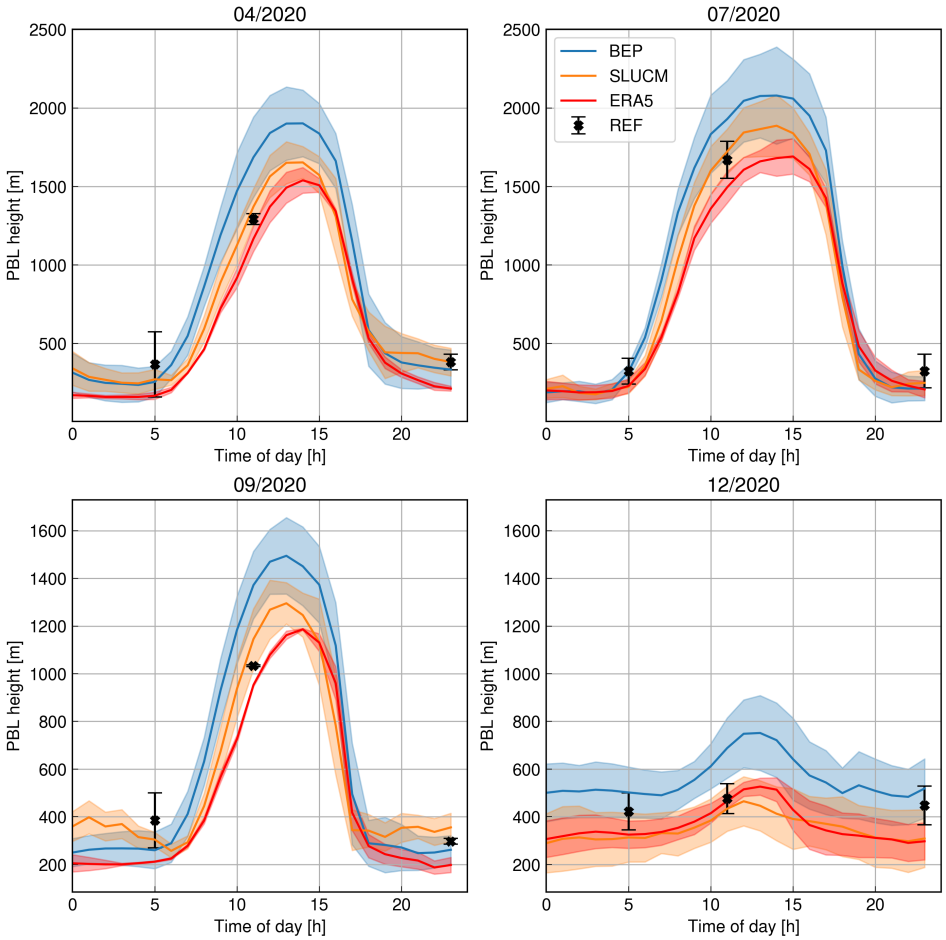


Figure A.4: Diurnal cycle of PBL height for SLUCM and BEP configurations averaged over both stations. Orange line is the SLUCM configuration, blue line is the BEP configuration, and red line is ERA5 data. Black markers are the PBL height measurements retrieved from radiosonde launches. Error bars and shading are one standard deviation of all diurnal cycles. Figure taken from Pilz et al. (2026).

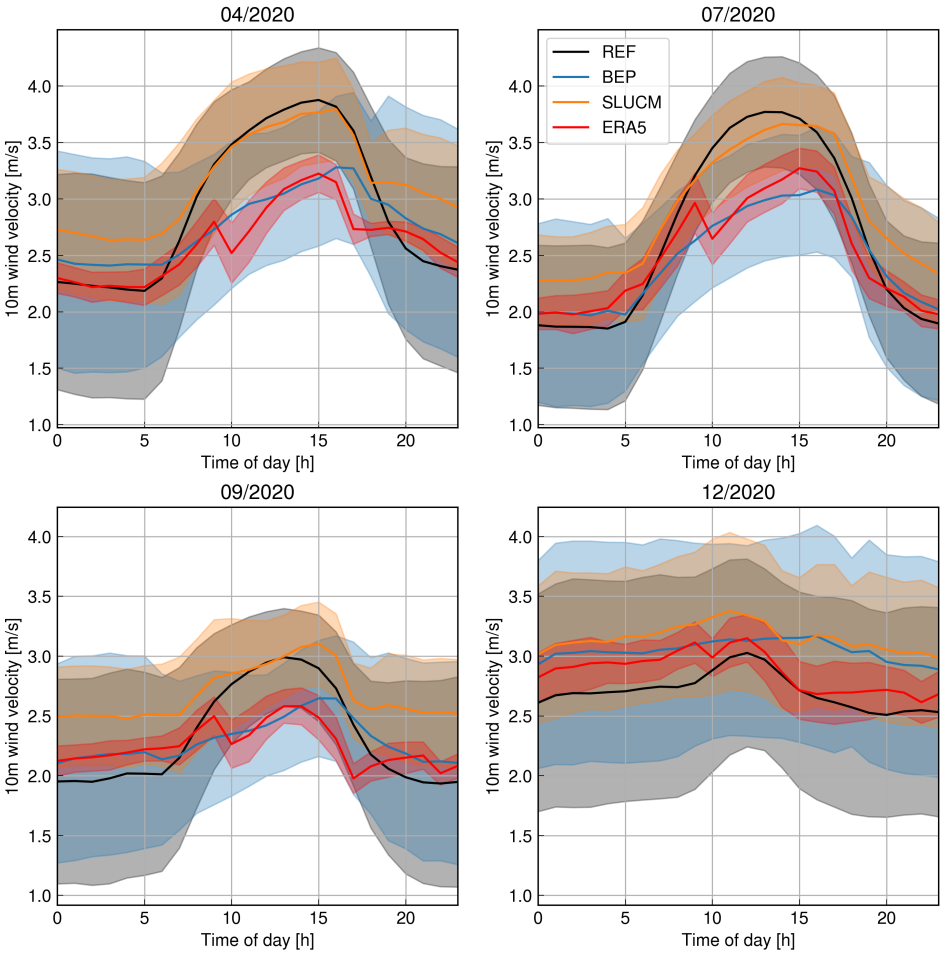


Figure A.5: Diurnal cycle of 10 m wind velocity for SLUCM and BEP configurations averaged over all stations. Orange line is the SLUCM configuration, blue line is the BEP configuration, red line is ERA5 data, and black line is the DWD station measurements. Shading is one standard deviation of all diurnal cycles. Figure taken from Pilz et al. (2026).

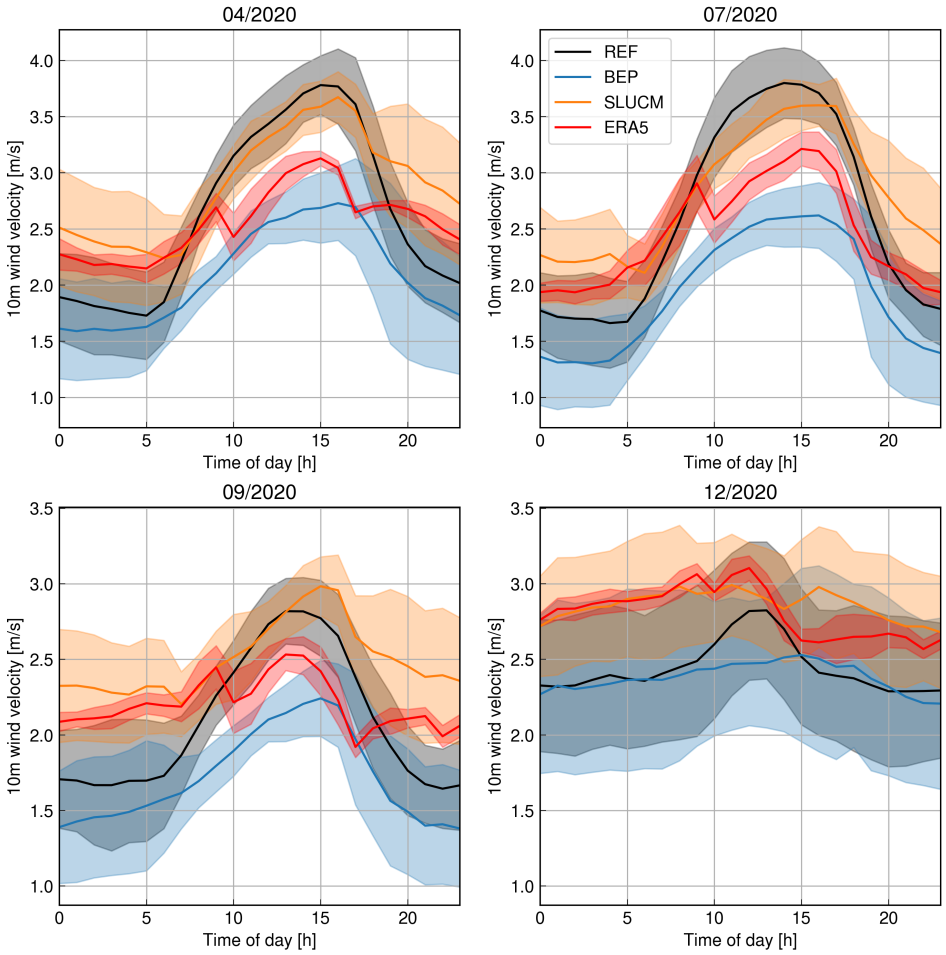


Figure A.6: Diurnal cycle of 10 m wind velocity for SLUCM and BEP configurations averaged over urban stations. Orange line is the SLUCM configuration, blue line is the BEP configuration, red line is ERA5 data, and black line is the DWD station measurements. Shading is one standard deviation of all diurnal cycles. Figure taken from Pilz et al. (2026).

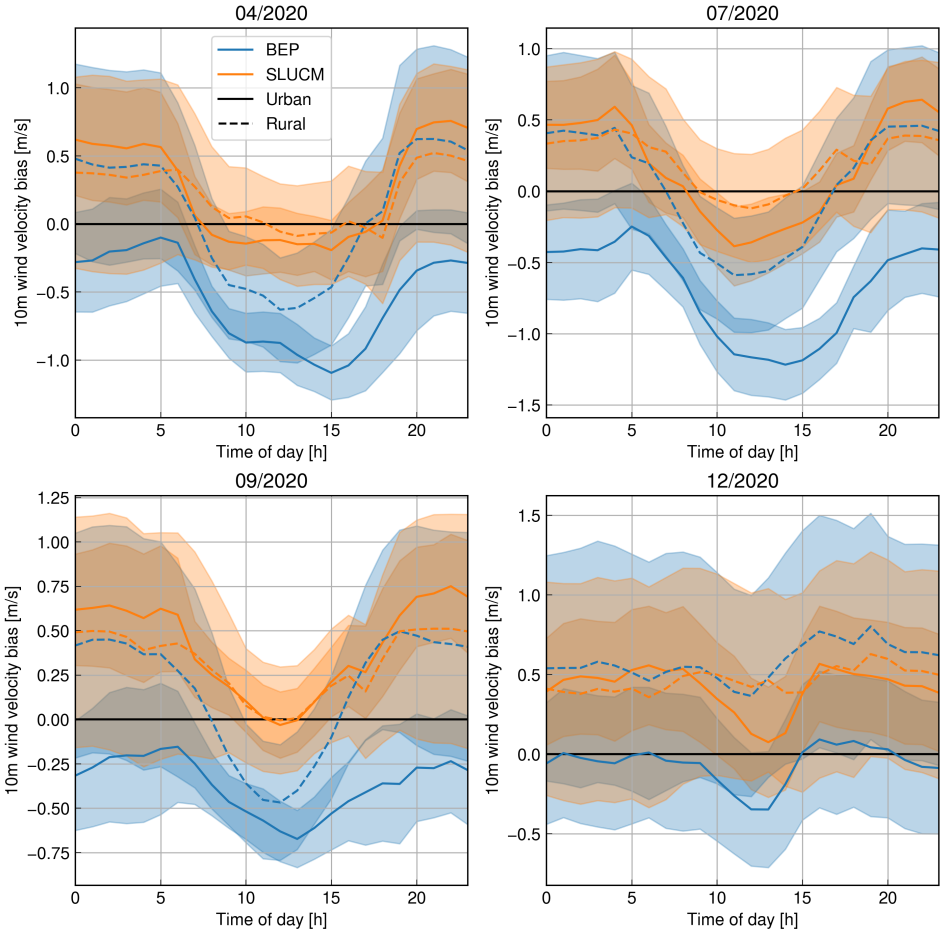


Figure A.7: Diurnal cycle of 10 m wind velocity bias for SLUCM and BEP configurations averaged both urban and rural stations. Orange line is the SLUCM configuration, blue line is the BEP configuration, red line is ERA5 data, and black line is the DWD station measurements. Solid lines are urban and dashed lines rural stations. Shading is one standard deviation of all diurnal cycles. Figure taken from Pilz et al. (2026).

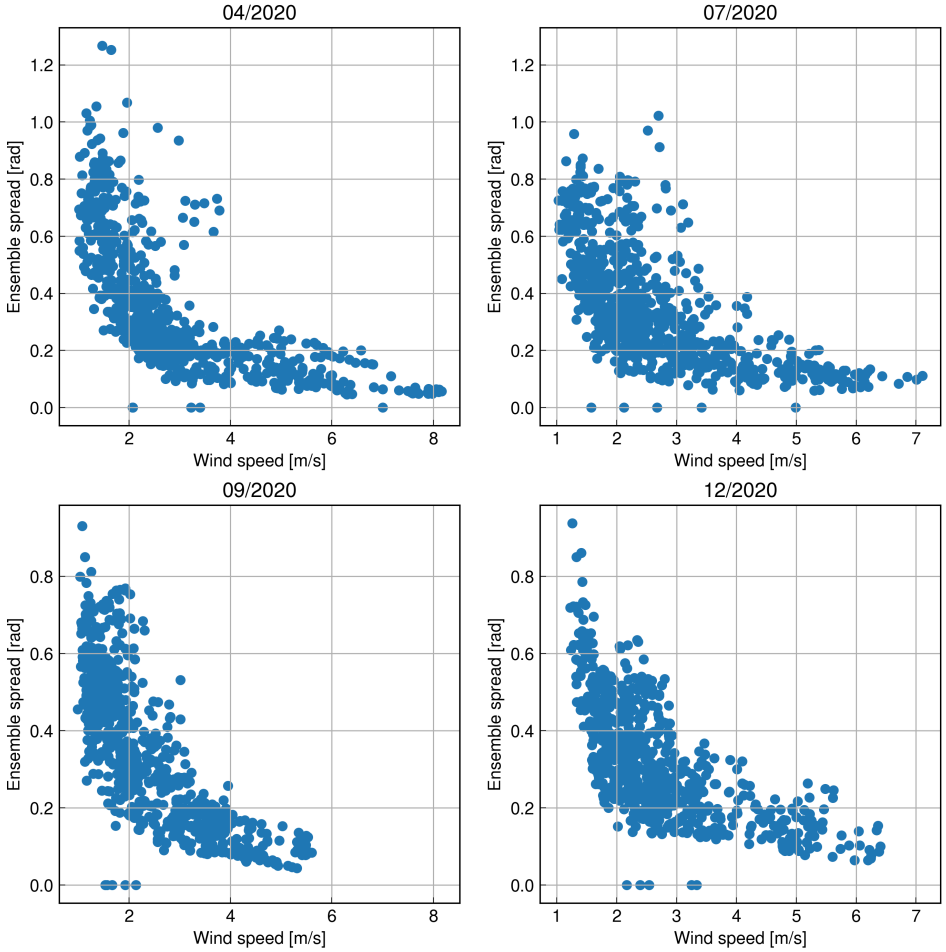


Figure A.8: Ensemble spread of 10 m wind direction against 10 m wind velocity. Ensemble spread is defined by circular standard deviation (see Jammalamadaka et al., 2001) across all configurations averaged over all DWD stations. Figure taken from Pilz et al. (2026).

Category	Description	Category	Description
1	Urban and Built-up Land	18	Wooden Wetland
2	Dryland Cropland and Pasture	19	Barren or Sparsely Vegetated
3	Irrigated Cropland and Pasture	20	Herbaceous Tundra
4	Mixed Dryland/Irrigated Cropland and Pasture	21	Wooded Tundra
		22	Mixed Tundra
5	Cropland/Grassland Mosaic	23	Bare Ground Tundra
6	Cropland/Woodland Mosaic	24	Snow or Ice
7	Grassland	31	LCZ 1: Compact high-rise
8	Shrubland	32	LCZ 2: Compact midrise
9	Mixed Shrubland/Grassland	33	LCZ 3: Compact low-rise
10	Savanna	34	LCZ 4: Open high-rise
11	Deciduous Broadleaf Forest	35	LCZ 5: Open midrise
12	Deciduous Needleleaf Forest	36	LCZ 6: Open low-rise
13	Evergreen Broadleaf	37	LCZ 7: Lightweight low-rise
14	Evergreen Needleleaf	38	LCZ 8: Large low-rise
15	Mixed Forest	39	LCZ 9: Sparsely built
16	Water Bodies	40	LCZ 10: Heavy industry
17	Herbaceous Wetland	41	LCZ E (LCZ 11): Rock and paved

Table A.2: Names of USGS land use categories in WRF. Table taken from Pilz et al. (2026).

Station ID	Name	State	Latitude	Longitude	Height [m ASL]	Wind Meas. Height [m AGL]	Urban/ Rural
3925	Pforzheim-Ispringen	BW	48.9329	8.6973	332.0	12.0	R
5871	Hahn	RP	49.9463	7.2645	497.0	10.0	R
2385	Idar-Oberstein	RP	49.6927	7.3263	376.0	14.5	R
5906	Mannheim	BW	49.5063	8.5584	98.0	10.0	U
4928	Stuttgart (Schnarrenberg)	BW	48.8281	9.2	314.0	12.0	U
2485	Kaisersbach-Cronhütte	BW	48.917	9.6871	489.0	12.0	R
3362	Mühlacker	BW	48.9721	8.8735	243.0	10.0	R
3761	Öhringen	BW	49.207	9.5176	276.0	15.0	U
4931	Stuttgart-Echterdingen	BW	48.6883	9.2235	371.0	10.0	U
5705	Würzburg	BY	49.7704	9.9576	268.0	12.0	U
1580	Geisenheim	HE	49.9859	7.9548	111.0	12.0	U
13674	Waibstadt	BW	49.2943	8.9053	237.0	10.0	R
2601	Kleiner Feldberg/Taunus	HE	50.2218	8.4469	822.0	12.0	R
1420	Frankfurt/Main	HE	50.0259	8.5213	100.0	10.0	R
4177	Rheinstetten	BW	48.9726	8.3301	116.0	10.0	R
3287	Michelstadt-Vielbrunn	HE	49.7177	9.0997	453.0	12.0	R
4887	Stötten	BW	48.6656	9.8648	734.0	12.0	R
7341	Offenbach-Wetterpark	HE	50.09	8.7862	119.0	15.0	U
7412	Neuhütten/Spessart	BY	50.0083	9.4238	340.0	10.0	R

Table A.3: Overview over the DWD measurement stations we compare the simulated 2 m temperature and 10 m wind velocity and wind direction data against. Urban stations are marked with “U” and rural stations with “R”. Details on the classification can be found in Section 2.3.2. State abbreviations are BW=Baden-Württemberg, BY=Bavaria, HE=Hessia, RP=Rhineland-Palatinate. Table taken from Pilz et al. (2026).

			avg. SLUCM	avg. BEP	best WRF simulation MAB	name	avg. ERA5
T2 [°C]	Apr.	Total	1.8	2.2	1.7	slucm_ysu_nmp_mm5	1.4
		Urban	1.7	2.6	1.6	slucm_ysu_nmp_mm5	1.2
		Rural	1.9	2.0	1.8	slucm_ysu_nmp_mm5	1.6
	Jul.	Total	1.5	1.7	1.4	slucm_ysu_nmp_mm5	1.3
		Urban	1.4	1.8	1.2	slucm_ysu_nmp_mm5	1.1
		Rural	1.6	1.7	1.5	bep_ysu_n_mm5	1.5
	Sep.	Total	1.6	1.9	1.5	slucm_ysu_nmp_mm5	1.3
		Urban	1.4	2.2	1.3	slucm_ysu_nmp_mm5	1.1
		Rural	1.7	1.8	1.6	bep_ysu_n_mm5	1.4
	Dec.	Total	1.4	1.8	1.3	slucm_bl_nmp_mo	1.0
		Urban	1.5	2.3	1.4	slucm_bl_nmp_mo	0.9
		Rural	1.3	1.5	1.3	slucm_bl_nmp_mo	1.1
PBLH [m]	Apr.	Total	271	346	249	slucm_bl_n_mm5	158
		Urban	293	452	271	slucm_bl_nmp_mm5	134
		Rural	249	240	208	bep_ysu_n_mm5	182
	Jul.	Total	268	308	220	slucm_myj_n_mo	191
		Urban	267	345	219	slucm_myj_n_mo	208
		Rural	270	270	226	slucm_myj_n_mo	174
	Sep.	Total	240	310	203	slucm_myj_n_mo	137
		Urban	247	367	206	slucm_myj_n_mo	137
		Rural	234	253	189	bep_ysu_n_mm5	138
	Dec.	Total	158	263	149	slucm_myj_n_mo	108
		Urban	167	362	139	slucm_bl_n_mo	102
		Rural	149	163	130	slucm_myj_n_mo	115
WVel [m/s]	Apr.	Total	1.0	1.0	0.9	slucm_myj_nmp_mo	1.0
		Urban	1.1	1.0	0.9	bep_bl_n_mo	0.9
		Rural	1.0	1.1	0.9	slucm_myj_nmp_mo	1.1
	Jul.	Total	0.9	1.0	0.8	slucm_myj_nmp_mo	0.9
		Urban	0.9	1.0	0.8	slucm_myj_nmp_mo	0.8
		Rural	0.9	1.0	0.8	slucm_myj_nmp_mo	0.9
	Sep.	Total	1.0	0.9	0.8	bep_myj_nmp_mo	0.9
		Urban	1.0	0.8	0.7	bep_myj_nmp_mo	0.8
		Rural	1.0	1.0	0.8	bep_myj_nmp_mo	1.0
	Dec.	Total	1.1	1.2	0.9	slucm_myj_nmp_mo	1.1
		Urban	1.1	0.9	0.8	bep_myj_nmp_mo	0.9
		Rural	1.2	1.3	1.0	slucm_myj_nmp_mo	1.2
WDir [°]	Apr.	Total	32	33	30	slucm_myj_n_mo	32
		Urban	39	39	36	slucm_myj_n_mo	38
		Rural	28	29	26	slucm_myj_n_mo	29
	Jul.	Total	33	33	32	slucm_bl_n_mo	32
		Urban	37	38	36	bep_bl_nmp_mm5	35
		Rural	30	31	29	bep_bl_n_mo	30
	Sep.	Total	35	35	33	slucm_myj_n_mo	36
		Urban	41	41	38	slucm_myj_n_mo	42
		Rural	31	31	30	slucm_myj_n_mo	33
	Dec.	Total	32	34	29	slucm_myj_nmp_mo	30
		Urban	37	39	35	slucm_myj_nmp_mo	33
		Rural	30	32	26	slucm_myj_nmp_mo	29

Table A.4: Comparison of Mean Absolute Bias between WRF simulations and ERA5 for 2 m-temperature, PBL height, wind velocity and wind direction rounded to instrument precision. Mean Absolute Bias values are given for each month for either averaged over all, only urban or only non-urban stations (Total, Urban, Rural). They are given for the average over all simulations using SLUCM and using BEP urban parametrization schemes. Also, the best WRF simulation is given with its name and MAB value. For comparison, ERA5 performance is given and rows where WRF outperforms ERA5 on full precision are highlighted in grey. Table taken from Pilz et al. (2026).

			avg. SLUCM	avg. BEP	best WRF simulation RMSE	name	avg. ERA5
T2 [°C]	Apr.	Total	2.3	2.9	2.2	slucm_ysu_nmp_mm5	1.8
		Urban	2.2	3.4	2.0	slucm_ysu_nmp_mm5	1.6
		Rural	2.4	2.6	2.3	slucm_ysu_nmp_mm5	2.0
	Jul.	Total	2.0	2.2	1.8	slucm_ysu_nmp_mm5	1.6
		Urban	1.8	2.2	1.6	slucm_ysu_nmp_mm5	1.4
		Rural	2.1	2.2	1.9	bep_ysu_n_mm5	1.8
	Sep.	Total	2.0	2.4	1.9	slucm_ysu_nmp_mm5	1.6
		Urban	1.8	2.7	1.6	slucm_ysu_nmp_mm5	1.4
		Rural	2.2	2.3	2.0	bep_ysu_n_mm5	1.8
	Dec.	Total	1.8	2.3	1.7	slucm_ysu_nmp_mm5	1.3
		Urban	1.9	2.8	1.8	slucm_ysu_nmp_mm5	1.1
		Rural	1.7	2.0	1.7	slucm_bl_nmp_mo	1.4
PBLH [m]	Apr.	Total	271	346	249	slucm_bl_n_mm5	158
		Urban	293	452	271	slucm_bl_nmp_mm5	134
		Rural	249	240	208	bep_ysu_n_mm5	182
	Jul.	Total	268	308	221	slucm_myj_n_mo	191
		Urban	267	345	215	slucm_myj_n_mo	208
		Rural	270	270	226	slucm_myj_n_mo	174
	Sep.	Total	240	310	203	slucm_myj_n_mo	137
		Urban	247	367	206	slucm_myj_n_mo	137
		Rural	234	253	189	bep_ysu_n_mm5	138
	Dec.	Total	158	263	149	slucm_myj_n_mo	109
		Urban	167	362	139	slucm_bl_n_mo	102
		Rural	149	163	130	slucm_myj_n_mo	115
WVel [m/s]	Apr.	Total	1.3	1.3	1.1	slucm_myj_nmp_mo	1.3
		Urban	1.4	1.3	1.1	bep_bl_n_mo	1.1
		Rural	1.3	1.4	1.1	slucm_myj_nmp_mo	1.4
	Jul.	Total	1.2	1.3	1.0	slucm_myj_nmp_mo	1.1
		Urban	1.2	1.3	1.1	slucm_myj_nmp_mo	1.0
		Rural	1.2	1.3	1.0	slucm_myj_nmp_mo	1.1
	Sep.	Total	1.3	1.2	1.0	bep_myj_nmp_mo	1.2
		Urban	1.3	1.0	1.0	bep_myj_nmp_mo	1.0
		Rural	1.3	1.3	1.0	bep_myj_nmp_mo	1.3
	Dec.	Total	1.5	1.5	1.2	slucm_myj_nmp_mo	1.3
		Urban	1.4	1.2	1.1	bep_myj_nmp_mo	1.1
		Rural	1.5	1.7	1.2	slucm_myj_nmp_mo	1.4
WDir [°]	Apr.	Total	47	48	45	slucm_myj_n_mo	46
		Urban	55	55	52	slucm_myj_n_mo	53
		Rural	42	43	41	slucm_myj_n_mo	42
	Jul.	Total	47	48	46	slucm_myj_n_mo	44
		Urban	52	53	50	bep_bl_nmp_mm5	49
		Rural	44	45	43	slucm_myj_n_mo	41
	Sep.	Total	50	50	48	slucm_myj_n_mo	51
		Urban	58	56	54	bep_myj_nmp_mo	57
		Rural	46	46	45	slucm_myj_n_mo	48
	Dec.	Total	46	48	43	slucm_myj_nmp_mo	42
		Urban	51	53	49	slucm_myj_nmp_mo	44
		Rural	43	46	39	slucm_myj_nmp_mo	40

Table A.5: Comparison of Root Mean Squared Error between WRF simulations and ERA5 for 2 m-temperature, PBL height, wind velocity and wind direction rounded to instrument precision. Root Mean Squared Error values are given for each month for either averaged over all, only urban or only non-urban stations (Total, Urban, Rural). They are given for the average over all simulations using SLUCM and using BEP urban parametrization schemes. Also, the best WRF simulation is given with its name and MAB value. For comparison, ERA5 performance is given and rows where WRF outperforms ERA5 on full precision are highlighted in grey. Table taken from Pilz et al. (2026).

			avg. SLUCM	avg. BEP	best WRF simulation MAB	name	avg. ERA5
T2 [°C]	Apr.	Day	1.5	1.7	1.4	slucm_ysu_nmp_mm5	1.2
		Night	2.2	3.0	2.1	slucm_ysu_nmp_mm5	1.8
	Jul.	Day	1.4	1.5	1.2	slucm_ysu_nmp_mm5	1.2
		Night	1.8	2.1	1.7	slucm_ysu_nmp_mm5	1.5
	Sep.	Day	1.4	1.7	1.3	slucm_ysu_nmp_mm5	1.2
		Night	1.8	2.3	1.7	slucm_ysu_nmp_mm5	1.5
	Dec.	Day	1.3	1.9	1.3	slucm_bl_nmp_mo	1.1
		Night	1.4	1.7	1.4	slucm_bl_nmp_mo	1.0
PBLH [m]	Apr.	Day	310	471	270	slucm_ysu_n_mm5	212
		Night	154	144	100	bep_bl_nmp_mo	64
	Jul.	Day	255	310	207	slucm_myj_n_mo	200
		Night	235	213	167	slucm_myj_n_mo	114
	Sep.	Day	278	385	234	slucm_myj_n_mo	171
		Night	133	163	108	slucm_ysu_nmp_mm5	79
	Dec.	Day	157	293	144	slucm_myj_n_mo	115
		Night	135	222	126	slucm_bl_nmp_mm5	89
WVel [m/s]	Apr.	Day	0.9	1.1	0.9	slucm_ysu_nmp_mm5	1.0
		Night	1.1	1.0	0.8	bep_myj_nmp_mo	1.0
	Jul.	Day	0.9	1.0	0.8	slucm_myj_nmp_mo	0.9
		Night	1.0	0.9	0.8	bep_myj_nmp_mo	0.9
	Sep.	Day	0.9	0.9	0.8	bep_myj_nmp_mo	0.9
		Night	1.1	0.9	0.8	bep_myj_nmp_mo	0.9
	Dec.	Day	1.1	1.1	0.9	slucm_myj_nmp_mo	1.1
		Night	1.1	1.2	0.9	slucm_myj_nmp_mo	1.1
WDir [°]	Apr.	Day	30	31	29	slucm_bl_nmp_mo	31
		Night	39	39	34	slucm_myj_n_mo	39
	Jul.	Day	31	32	30	slucm_ysu_n_mm5	30
		Night	41	40	37	slucm_myj_n_mo	39
	Sep.	Day	34	35	34	slucm_myj_n_mo	37
		Night	39	38	34	slucm_myj_n_mo	39
	Dec.	Day	32	33	29	slucm_myj_nmp_mo	30
		Night	34	36	30	slucm_myj_nmp_mo	32

Table A.6: Comparison of Mean Absolute Bias between WRF simulations and ERA5 for 2 m-temperature, PBL height, wind velocity and wind direction rounded to instrument precision. Mean Absolute Bias values are averaged over all stations and given for each month averaged over either daytime or non-daytime hours. These were selected based on the sunrise and sunset of the first of the respective month. They are given for the average over and all simulations using SLUCM and using BEP urban parametrization schemes. Also, the best WRF simulation is given with its name and MAB value. For comparison, ERA5 performance is given and rows where WRF outperforms ERA5 on full precision are highlighted in grey. Table taken from Pilz et al. (2026).

MACRO-2018

Known Issues

The main known issue with the MACRO-2018 dataset are the slight deviations in total emissions due to regridding (cf. Section 3.1.3). If totally perfect conservation of the field is required, these can then be scaled in post-processing using the factors given in Table 3.1. However, when doing so the pre-aggregated fields (`CO2_ANTHRO`, `CO2_TOTAL`, and `CO2_TOTAL_V2`) in the dataset should not be used.

Careful selection of the appropriate boundary layer scheme for the location of interest is an important consideration for users. Our studies find the YSU scheme to underperform the MYJ scheme at the 15 m inlet of `SAC` station (Saclay in the vicinity of Paris; cf. Section 3.4.4).

Users should be careful around topographically complex areas like the Alps. Here, the relatively low resolution in the Germany and Europe domains (5 and 15 km) can contribute to significant deviations like at `IPR` station (Ispra, Italy; cf. Section 3.4.4).

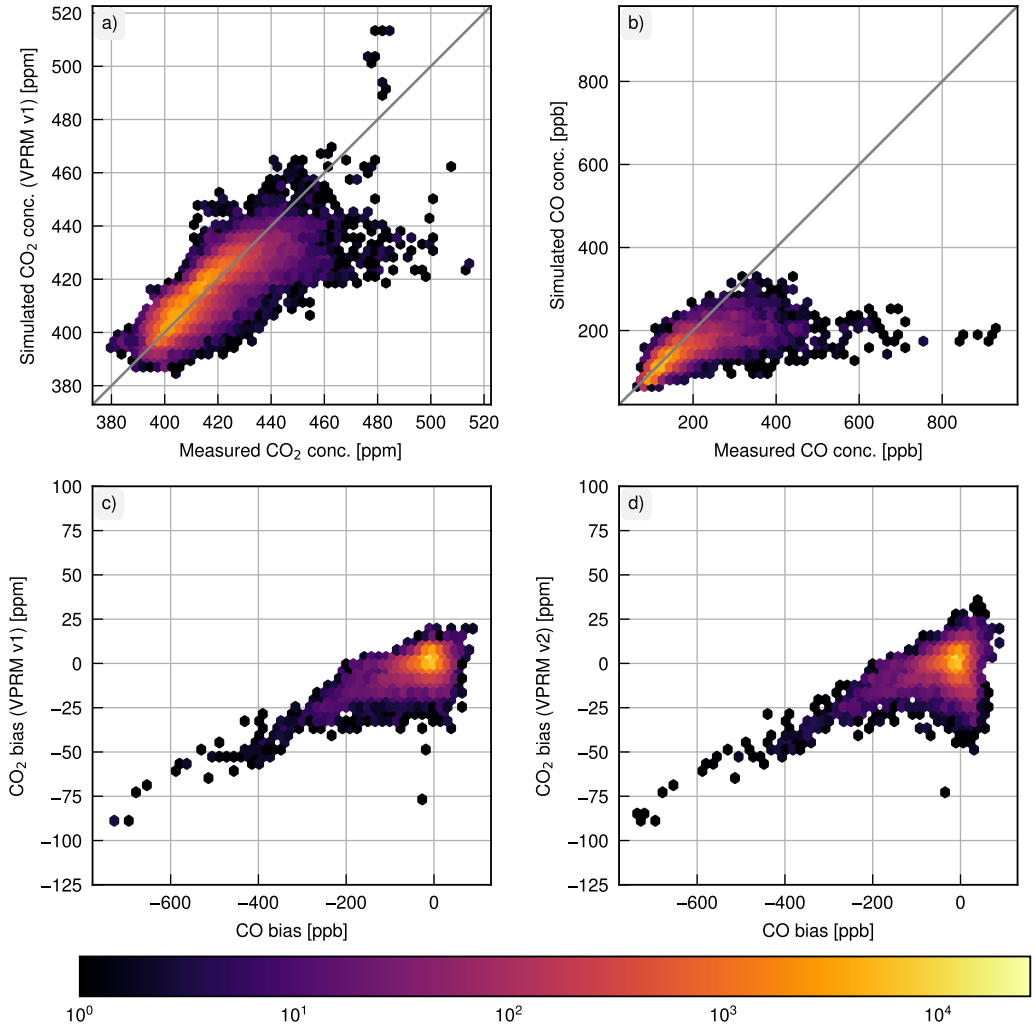


Figure A.9: 2D Histogram of simulated CO_2 and CO concentrations vs measurements only for stable PBL development phase. Lower panels show CO_2 vs. CO biases for both VPRM versions.

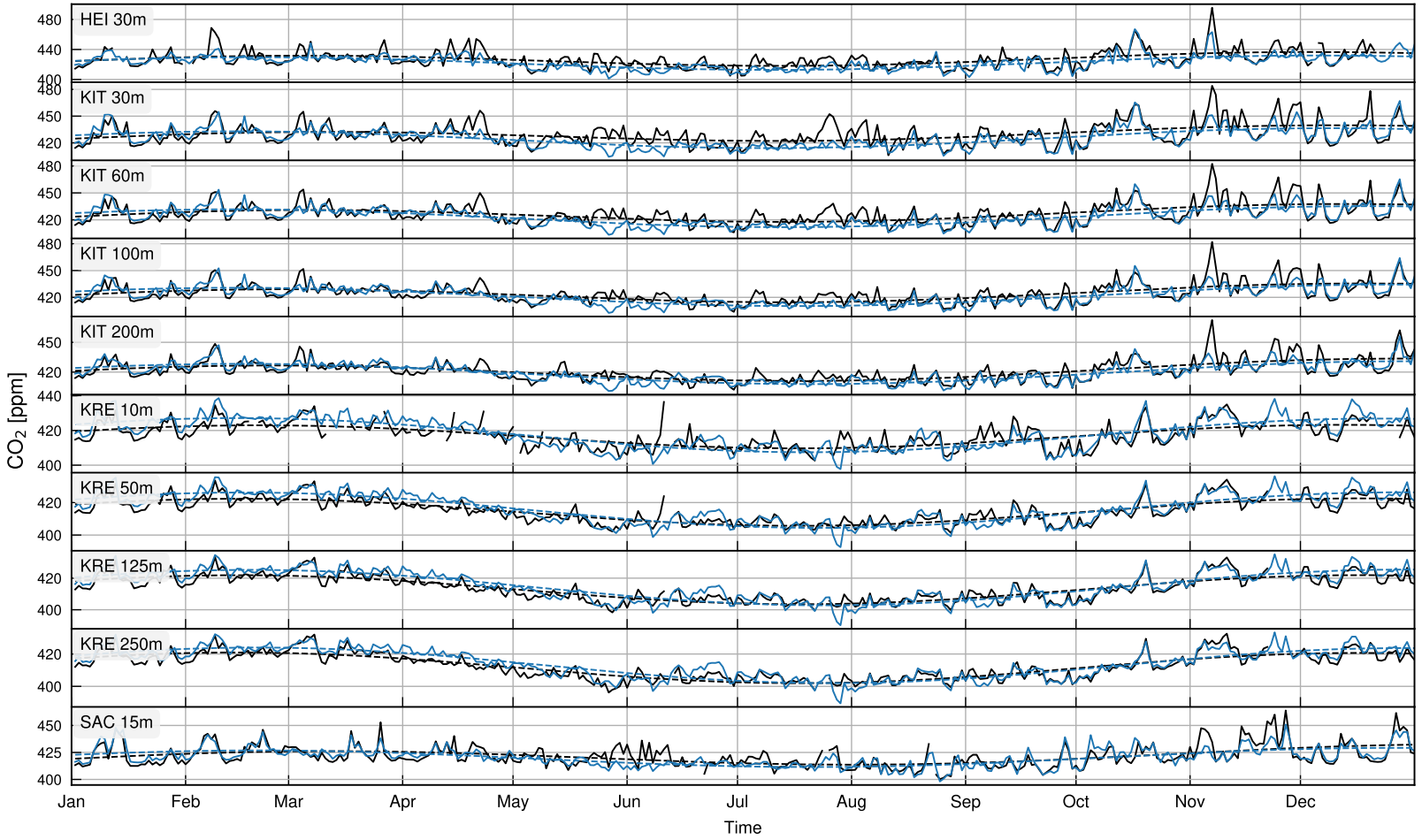


Figure A.10: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

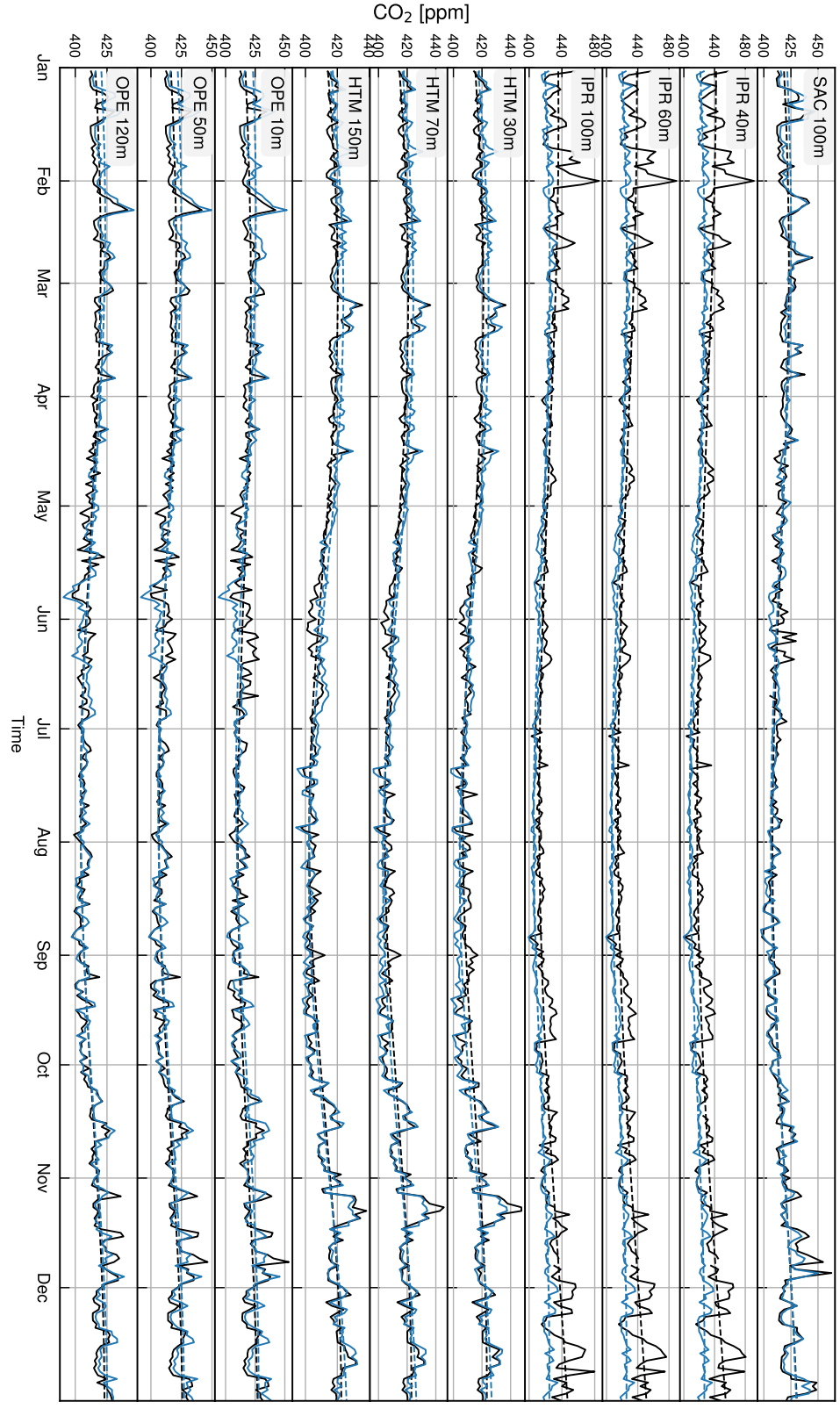


Figure A.11: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

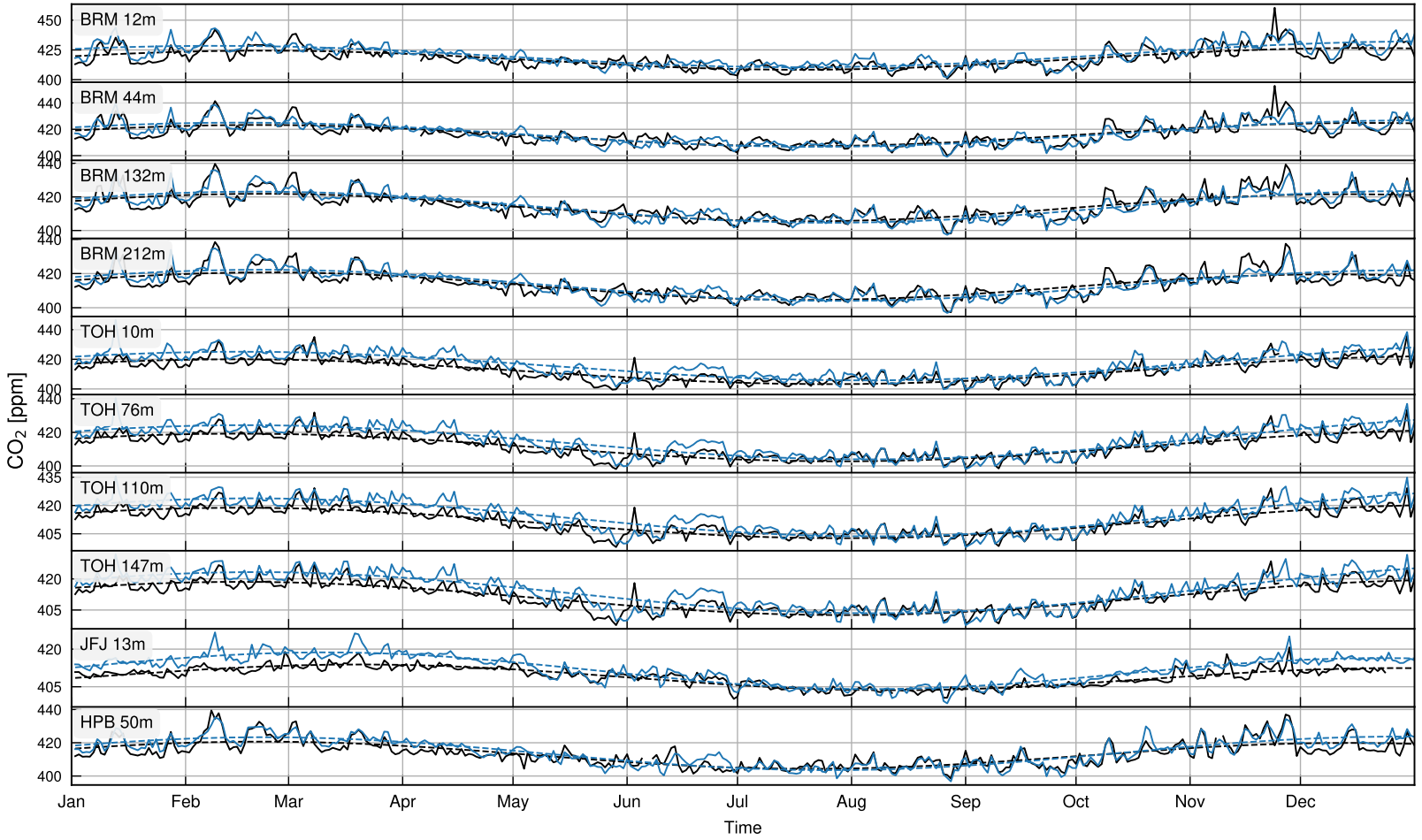


Figure A.12: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

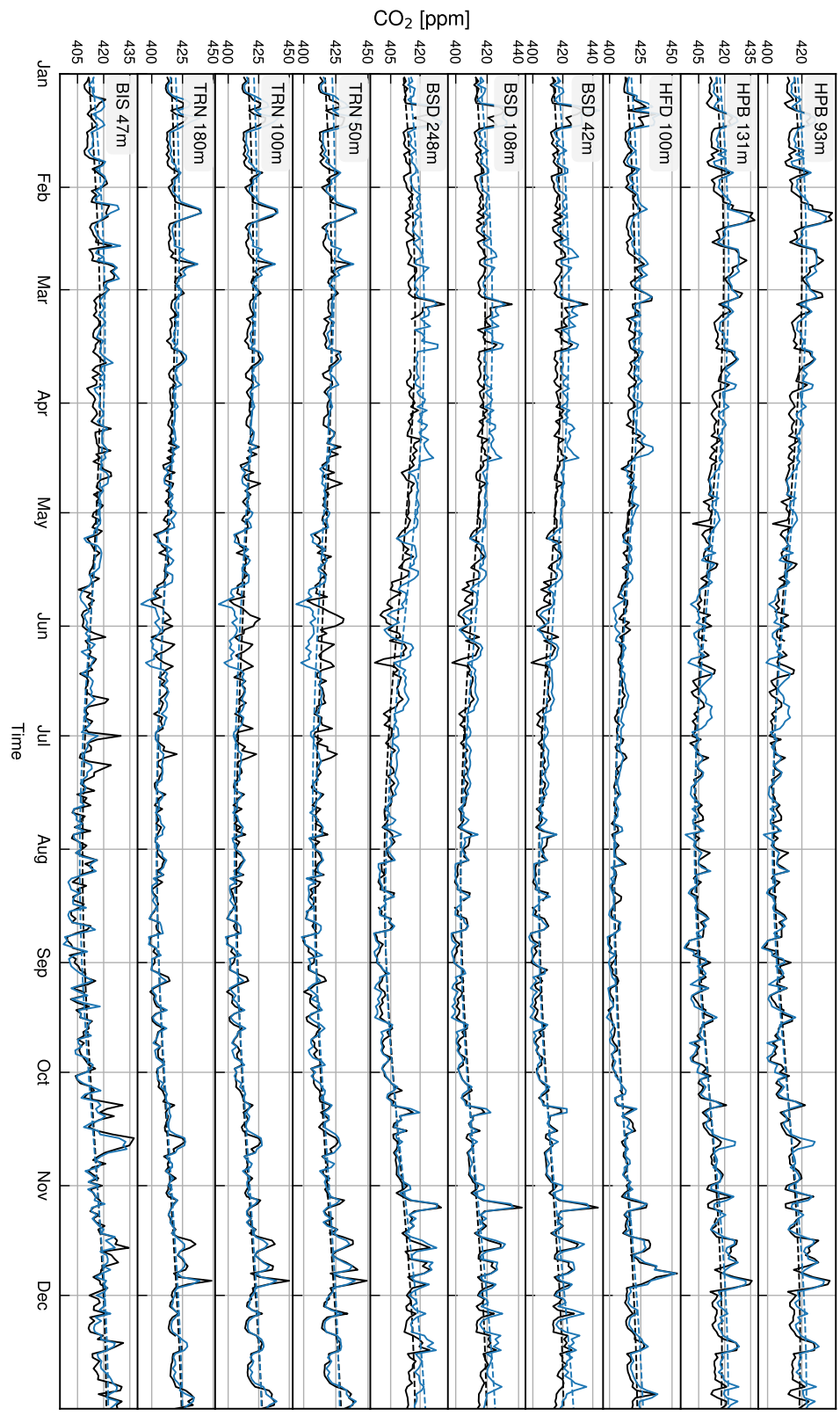


Figure A.13: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

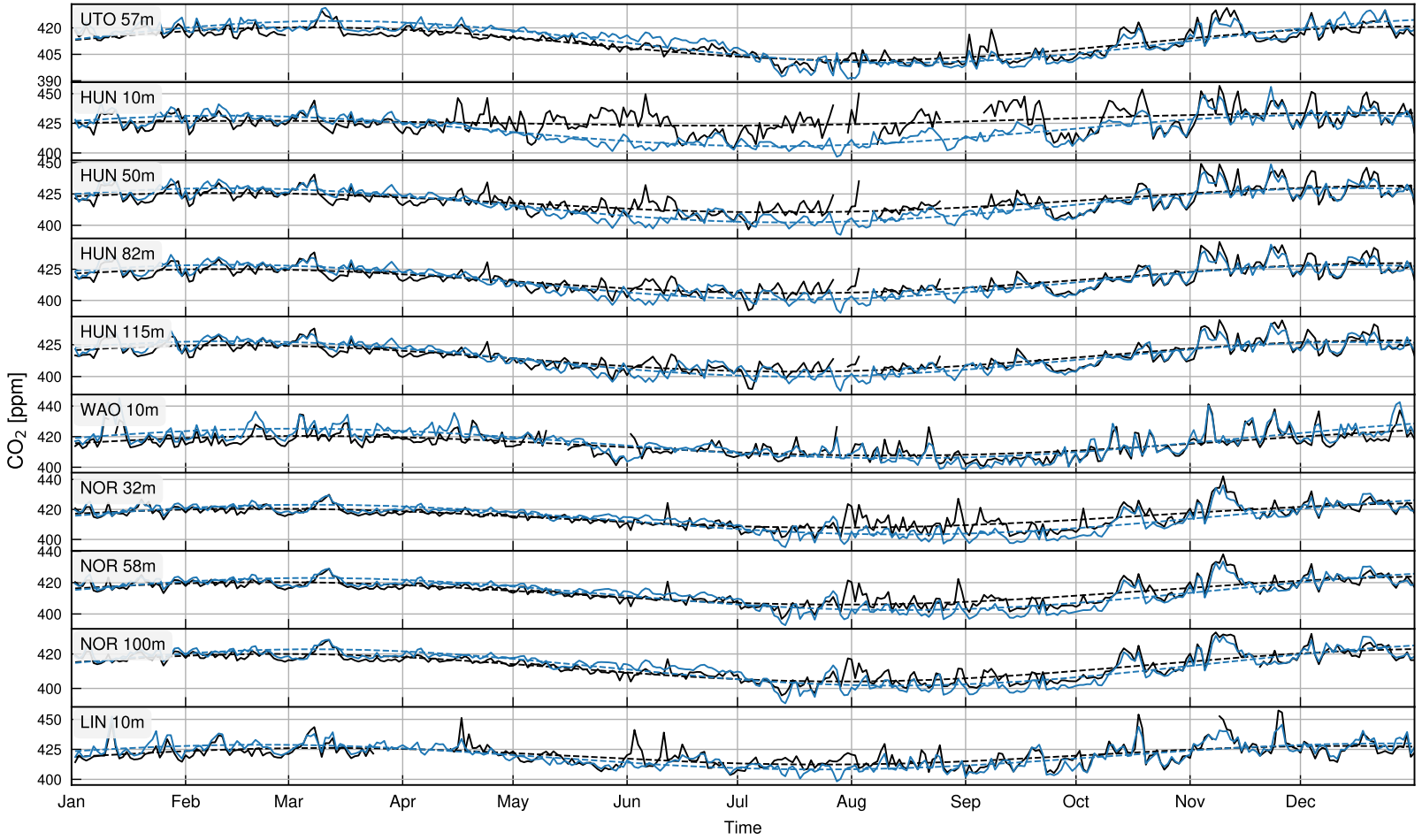


Figure A.14: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

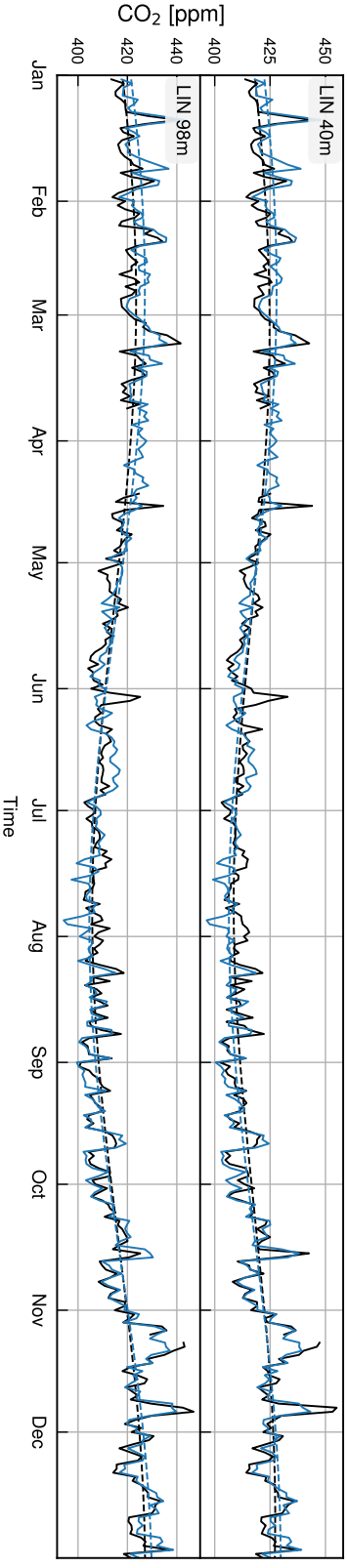


Figure A.15: Timeseries of simulated and measured daily average CO_2 concentrations (solid blue and black lines) and seasonal cycle fits (dashed blue and black lines).

Table A.7: Overview of all Obspack stations used in the evaluation of the MACRO-2018 dataset.

Station ID	Name	Category	Latitude	Longitude	Elevation [m asl]	Inlet heights [m agl]	Simulation res. [km]	Domain	Gases
LIN	Lindenberg	Rural	52.166302	14.122600	73.0	10/40/98	1	Berlin	CO_2
HFD	Heathfield	Rural	50.976700	0.230600	160.0	100	15	Europe	CO_2
BSD	Bilsdale	Rural	54.358601	-1.150400	382.0	42/108/248	15	Europe	CO_2
TRN	Trainou	Rural	47.964699	2.112500	131.0	50/100/180	15	Europe	CO_2, CO
BIS	Biscarrosse	Remote	44.378101	-1.231100	73.0	47	15	Europe	CO_2
UTO	Uto	Remote	59.783901	21.367201	8.0	57	15	Europe	CO_2
HUN	Hegyhátsál háttérszennye-zettség-mérő állomás	Rural	46.955799	16.652201	248.0	10/50/82/115	15	Europe	CO_2
WAO	Weybourne	Rural	52.950001	1.121000	17.0	10	15	Europe	CO_2
NOR	Norunda	Remote	60.086399	17.479401	46.0	32/58/100	15	Europe	CO_2, CO
JFJ	Jungfraujoch	Remote	46.549999	7.987000	3570.0	13	5	Germany	CO_2, CO
TOH	Torfhaus	Rural	51.808800	10.535000	801.0	10/76/110/147	5	Germany	CO_2
BRM	Beromunster	Rural	47.189602	8.175500	797.0	12/44/132/212	5	Germany	CO_2
SAC	Saclay	Urban	48.722698	2.142000	160.0	15/100	5	Germany	CO_2, CO
HTM	Hyltemossa	Remote	56.097599	13.418900	198.4	30/70/150	5	Germany	CO_2, CO
IPR	Ispra	Rural	45.806999	8.630000	223.0	40/60/100	5	Germany	CO_2, CO
KRE	Kresin u Pacova	Rural	49.583000	15.083000	534.0	10/50/125/250	5	Germany	CO_2, CO
OPE	Observatoire perenne de l'environnement	Rural	48.561901	5.503600	390.0	10/50/120	5	Germany	CO_2, CO
HPB	Hohenpeissenberg	Rural	47.801102	11.024500	985.0	50/93/131	1	Munich	CO_2, CO
HEI	Heidelberg	Urban	49.417198	8.675000	113.0	30	1	Rhine-Neckar	CO_2
KIT	Karlsruhe	Urban	49.091499	8.424900	110.0	30/60/100/200	1	Rhine-Neckar	CO_2

Table A.8: Statistics of CO_2 concentrations averaged over all inlets for each station and boundary layer scheme. Stations are sorted by average MAB. Average at bottom is computed over per-station values, which are averaged over the respective inlets. As such it differs from Table A.10.

	Inlets	Category	MB [ppm]		MAB [ppm]		RMSE [ppm]		Pearson-R	
			MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
UTO	1.0	Remote	1.0	0.9	3.3	3.3	4.2	4.2	0.90	0.90
HTM	3.0	Remote	0.9	0.7	3.3	3.2	4.1	4.1	0.90	0.90
JFJ	1.0	Remote	2.9	2.7	3.6	3.3	4.2	3.9	0.85	0.87
HFD	1.0	Rural	1.5	1.3	3.6	3.4	4.7	4.5	0.89	0.89
BSD	3.0	Rural	2.5	2.5	3.5	3.5	4.4	4.3	0.91	0.91
NOR	3.0	Remote	−0.6	−0.6	3.7	3.6	5.4	5.3	0.82	0.81
BIS	1.0	Remote	0.5	0.4	3.8	3.8	5.3	5.3	0.79	0.79
HPB	3.0	Rural	1.3	0.8	3.8	3.7	4.8	4.8	0.86	0.84
KRE	4.0	Rural	1.3	1.1	3.7	3.6	4.7	4.6	0.87	0.87
WAO	1.0	Rural	1.6	1.5	4.0	4.0	5.2	5.1	0.86	0.86
TOH	4.0	Rural	3.2	3.0	4.0	3.9	5.0	4.8	0.90	0.90
TRN	3.0	Rural	0.5	0.2	4.1	4.0	6.0	5.9	0.80	0.81
BRM	4.0	Rural	1.0	0.3	4.0	4.1	5.2	5.4	0.83	0.81
OPE	3.0	Rural	1.6	1.4	4.5	4.5	6.2	6.2	0.81	0.79
LIN	3.0	Rural	0.2	0.2	4.5	4.5	6.3	6.3	0.84	0.84
HUN	4.0	Rural	−2.0	−1.8	6.1	5.9	9.0	8.6	0.74	0.76
SAC	2.0	Urban	−0.4	3.5	5.2	7.1	7.5	10.1	0.80	0.75
KIT	4.0	Urban	−2.6	−3.2	7.2	7.4	10.6	10.9	0.75	0.74
HEI	1.0	Urban	−4.3	−4.3	8.0	8.4	12.2	12.9	0.76	0.72
IPR	3.0	Rural	−9.7	−10.2	11.4	11.9	16.2	16.9	0.69	0.65
Average			0.0	0.0	4.7	4.8	7.2	7.5	0.83	0.82

Table A.9: Statistics of CO concentrations averaged over all inlets for each station and boundary layer scheme. Stations are sorted by average MAB. Average at bottom is computed over per-station values, which are averaged over the respective inlets. As such it differs from Table A.11.

	Inlets	Category	MB [ppb]		MAB [ppb]		RMSE [ppb]		Pearson-R	
			MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
JFJ	1.0	Remote	4.8	3.4	11.3	10.6	15.8	15.0	0.72	0.73
NOR	3.0	Remote	−11.5	−11.8	13.4	13.7	18.6	19.1	0.90	0.89
HTM	3.0	Remote	−11.3	−12.6	15.1	15.6	25.0	26.0	0.88	0.88
TRN	3.0	Rural	−5.4	−7.5	15.0	15.1	22.1	22.8	0.85	0.85
OPE	3.0	Rural	−8.9	−10.9	17.0	17.6	25.9	27.0	0.82	0.81
HPB	3.0	Rural	−9.2	−12.0	17.8	19.1	27.4	29.6	0.80	0.78
KRE	3.0	Rural	−17.3	−19.5	23.6	24.9	38.7	40.6	0.80	0.79
SAC	2.0	Urban	0.3	26.6	25.6	44.2	39.9	64.3	0.72	0.63
IPR	3.0	Rural	−70.7	−75.6	78.2	82.5	135.9	142.8	0.58	0.50
Average			−14.4	−13.3	24.1	27.1	52.4	57.4	0.79	0.76

Table A.10: Statistics of CO_2 concentrations for each inlet and boundary layer scheme. Inlets are sorted by average MAB.

Station	Inlet height [m agl]	MB [ppm] MYJ	MB [ppm] YSU	MAB [ppm] MYJ	MAB [ppm] YSU	RMSE [ppm] MYJ	RMSE [ppm] YSU	Pearson-R MYJ	Pearson-R YSU
HTM	150	1.4	1.2	3.0	2.9	3.8	3.7	0.92	0.92
UTO	57	1.0	0.9	3.3	3.3	4.2	4.2	0.90	0.90
HTM	70	0.9	0.6	3.3	3.2	4.3	4.3	0.90	0.89
NOR	100	0.0	−0.2	3.3	3.3	4.6	4.7	0.86	0.85
KRE	250	1.4	1.3	3.5	3.3	4.3	4.2	0.90	0.90
BSD	248	2.6	2.5	3.5	3.4	4.4	4.3	0.90	0.90
JFJ	13	2.9	2.7	3.6	3.3	4.2	3.9	0.85	0.87
HFD	100	1.5	1.3	3.6	3.4	4.7	4.5	0.89	0.89
BSD	108	2.4	2.3	3.6	3.5	4.5	4.4	0.90	0.90
KRE	125	1.3	1.1	3.7	3.5	4.6	4.4	0.89	0.88
BRM	212	0.4	−0.1	3.6	3.6	4.9	5.0	0.83	0.81
HPB	131	1.4	1.0	3.7	3.6	4.6	4.6	0.86	0.85
NOR	58	−0.6	−0.7	3.7	3.7	5.6	5.6	0.81	0.80
TRN	180	1.1	0.9	3.8	3.7	5.3	5.2	0.84	0.84
HTM	30	0.5	0.2	3.8	3.7	5.1	5.2	0.86	0.85
HPB	93	1.2	0.8	3.8	3.7	4.8	4.8	0.85	0.84
BIS	47	0.5	0.4	3.8	3.8	5.3	5.3	0.79	0.79
TOH	147	2.9	2.8	3.8	3.8	4.8	4.7	0.90	0.89
BRM	132	0.1	−0.6	3.8	3.9	5.0	5.2	0.83	0.81
BSD	42	2.6	2.6	3.9	3.9	5.0	4.9	0.89	0.89
KRE	50	1.3	0.9	4.0	3.8	5.1	5.0	0.86	0.85
TOH	110	3.1	2.9	4.0	3.9	4.9	4.8	0.90	0.89
OPE	120	1.6	1.3	4.0	4.0	5.4	5.3	0.85	0.84
TOH	76	3.1	2.9	4.0	3.9	5.1	4.9	0.90	0.89
WAO	10	1.6	1.5	4.0	4.0	5.2	5.1	0.86	0.86
HPB	50	1.2	0.5	4.1	4.0	5.2	5.2	0.84	0.82
LIN	98	0.9	0.8	4.2	4.1	5.7	5.7	0.85	0.85
NOR	32	−1.1	−1.0	4.2	4.1	6.6	6.5	0.76	0.76
TRN	100	0.4	0.0	4.4	4.3	6.5	6.4	0.79	0.79
BRM	44	0.4	−0.5	4.4	4.5	5.8	6.0	0.81	0.79
HUN	115	−0.3	−0.4	4.5	4.5	6.4	6.3	0.84	0.84
TOH	10	3.6	3.4	4.6	4.5	5.8	5.6	0.87	0.87
SAC	100	0.0	−1.1	4.5	4.6	6.4	6.7	0.84	0.83
LIN	40	0.3	0.2	4.6	4.6	6.4	6.4	0.83	0.83
OPE	50	1.6	1.1	4.6	4.5	6.5	6.4	0.80	0.79
TRN	50	0.0	−0.2	4.9	4.9	7.4	7.4	0.76	0.76
KRE	10	0.8	0.6	5.0	5.0	7.5	7.4	0.75	0.75
HUN	82	−0.6	−0.8	5.1	5.0	7.3	7.3	0.81	0.81
LIN	10	−0.7	−0.3	5.6	5.7	8.7	8.6	0.78	0.79
OPE	10	1.6	1.7	5.7	5.9	8.4	8.8	0.72	0.70
HUN	50	−1.7	−1.8	6.2	6.2	9.5	9.4	0.74	0.74
BRM	12	2.9	2.6	6.1	6.4	8.1	8.6	0.76	0.71
KIT	200	−1.5	−2.2	6.3	6.4	9.3	9.6	0.74	0.73
	100	−1.9	−2.7	7.2	7.4	10.6	11.0	0.73	0.72
HEI	30	−4.3	−4.3	8.0	8.4	12.2	12.9	0.76	0.72
KIT	60	−3.0	−3.9	8.1	8.4	12.3	12.7	0.72	0.71
SAC	15	−0.9	8.2	6.4	11.8	9.8	18.2	0.76	0.67
KIT	30	−3.9	−3.9	9.1	9.4	14.2	14.5	0.73	0.72
HUN	10	−5.5	−4.1	10.3	9.6	17.6	15.8	0.60	0.68
IPR	100	−8.4	−8.9	10.1	10.6	14.5	15.2	0.70	0.66
	60	−9.8	−10.4	11.5	12.1	16.4	17.2	0.69	0.65
	40	−11.0	−11.5	12.9	13.4	18.4	19.1	0.67	0.63
Average		−0.1	−0.2	5.1	5.2	8.0	8.4	0.81	0.81

Table A.11: Statistics of CO concentrations for each inlet and boundary layer scheme. Inlets are sorted by average MAB.

Station	Inlet height [m agl]	MB [ppb] MYJ	MB [ppb] YSU	MAB [ppb] MYJ	MAB [ppb] YSU	RMSE [ppb] MYJ	RMSE [ppb] YSU	Pearson-R MYJ	Pearson-R YSU
JFJ	13	4.8	3.4	11.3	10.6	15.8	15.0	0.72	0.73
NOR	32	−11.0	−10.9	13.3	13.5	18.5	18.8	0.89	0.89
	58	−11.5	−12.0	13.6	13.9	18.8	19.3	0.89	0.89
	100	−11.8	−12.5	13.6	14.1	18.8	19.5	0.90	0.89
TRN	180	−5.6	−7.6	14.1	14.5	21.4	22.3	0.85	0.84
HTM	30	−10.5	−11.8	15.0	15.5	25.2	26.1	0.88	0.88
	150	−12.0	−13.3	15.3	15.8	25.1	26.1	0.89	0.89
	70	−11.3	−12.8	15.3	15.9	25.4	26.4	0.88	0.88
TRN	100	−5.9	−8.4	15.7	15.9	23.5	24.4	0.84	0.84
	50	−4.9	−6.9	17.1	17.0	25.3	25.9	0.83	0.83
OPE	120	−9.7	−11.7	17.0	17.7	25.7	26.8	0.81	0.80
	50	−9.1	−11.4	17.5	18.2	26.9	27.8	0.81	0.81
HPB	131	−8.9	−11.4	17.3	18.6	26.4	28.5	0.80	0.78
OPE	10	−7.9	−9.5	17.6	18.5	27.6	28.8	0.80	0.79
HPB	93	−9.5	−12.3	18.1	19.4	28.1	30.2	0.80	0.78
	50	−9.5	−12.9	18.8	20.2	29.3	31.6	0.78	0.77
SAC	100	−6.3	−13.9	20.2	22.4	31.4	35.0	0.81	0.79
KRE	125	−15.2	−17.4	21.8	23.1	34.1	36.1	0.81	0.80
	250	−18.6	−20.2	23.8	24.8	38.6	39.9	0.80	0.79
	50	−14.2	−17.2	24.0	25.4	39.1	41.5	0.79	0.78
SAC	15	6.9	67.9	34.1	83.4	56.0	127.8	0.61	0.54
IPR	100	−64.2	−69.3	70.3	74.6	122.7	129.4	0.56	0.46
	60	−70.9	−76.2	78.6	83.2	137.3	144.5	0.57	0.48
	40	−76.0	−80.4	86.0	90.1	151.3	157.8	0.58	0.51
Average		−16.5	−16.1	25.5	28.8	56.0	63.5	0.79	0.77

Table A.12: Comparison of CO_2 concentration simulation performance across stations and phases of PBL development. This includes only situations with winds over 5 m s^{-1}. All entries are averaged over all available inlets.

		urban		rural		remote		all	
		MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
stable	MB [ppm]	2.1	1.5	2.4	2.2	1.9	1.9	2.3	2.0
	MAB [ppm]	4.3	4.4	3.8	3.8	3.1	3.0	3.8	3.7
	RMSE [ppm]	5.8	6.1	4.8	4.8	3.9	3.7	4.8	4.8
	Pearson-R [1]	0.86	0.83	0.89	0.88	0.91	0.91	0.89	0.88
transition	MB [ppm]	0.6	-0.1	1.1	0.7	1.3	1.0	1.1	0.7
	MAB [ppm]	4.5	4.7	3.9	4.1	3.1	3.0	3.9	4.0
	RMSE [ppm]	6.5	7.2	5.1	5.4	3.8	3.8	5.1	5.4
	Pearson-R [1]	0.80	0.78	0.84	0.82	0.89	0.88	0.84	0.83
night	MB [ppm]	-0.2	-1.3	0.2	-0.2	0.5	0.2	0.2	-0.3
	MAB [ppm]	4.6	5.7	4.3	4.5	3.2	3.2	4.1	4.4
	RMSE [ppm]	6.6	8.6	5.5	5.8	4.1	4.2	5.4	5.9
	Pearson-R [1]	0.72	0.64	0.79	0.77	0.86	0.85	0.79	0.77
all	MB [ppm]	0.6	-0.4	0.8	0.4	1.1	0.9	0.8	0.4
	MAB [ppm]	4.5	5.2	4.2	4.4	3.2	3.1	4.0	4.3
	RMSE [ppm]	6.4	7.8	5.4	5.7	4.0	4.0	5.3	5.7
	Pearson-R [1]	0.80	0.73	0.84	0.82	0.89	0.88	0.84	0.82

Table A.13: Comparison of CO concentration simulation performance across stations and phases of PBL development. This includes only situations with winds over 5 $m\,s^{-1}$. All entries are averaged over all available inlets.

		urban		rural		remote		all	
		MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU
stable	MB [ppb]	-9.5	-11.1	-14.7	-18.5	-9.4	-10.4	-12.7	-15.5
	MAB [ppb]	13.5	14.5	18.7	21.4	12.4	12.8	16.4	18.4
	RMSE [ppb]	21.4	22.8	30.9	35.8	18.2	18.8	26.4	29.8
	Pearson-R [1]	0.90	0.88	0.77	0.74	0.90	0.90	0.82	0.80
transition	MB [ppb]	-9.2	-8.0	-17.5	-21.9	-10.0	-10.9	-14.6	-17.5
	MAB [ppb]	15.5	16.3	22.7	26.5	13.4	13.9	19.4	22.0
	RMSE [ppb]	24.7	25.3	36.7	42.7	20.4	21.2	30.9	35.0
	Pearson-R [1]	0.88	0.85	0.75	0.76	0.88	0.88	0.80	0.80
night	MB [ppb]	-8.9	-10.5	-22.7	-29.2	-10.9	-11.5	-18.1	-22.5
	MAB [ppb]	15.5	18.6	28.8	34.7	14.5	14.7	23.5	27.6
	RMSE [ppb]	24.3	28.6	46.1	55.2	23.1	23.6	37.5	43.8
	Pearson-R [1]	0.87	0.87	0.73	0.70	0.87	0.87	0.78	0.76
all	MB [ppb]	-9.2	-10.9	-20.9	-27.4	-10.2	-11.0	-16.8	-21.2
	MAB [ppb]	15.1	16.5	26.3	32.1	13.6	14.0	21.7	25.5
	RMSE [ppb]	23.8	25.6	43.1	52.3	21.1	21.8	35.1	41.2
	Pearson-R [1]	0.88	0.87	0.74	0.72	0.88	0.88	0.79	0.78

Emissions Estimation

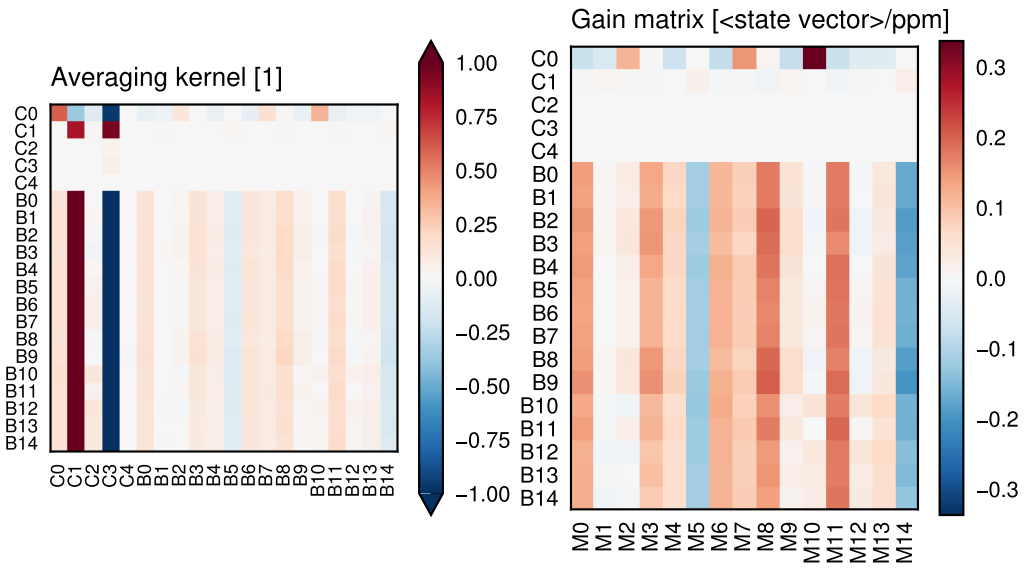


Figure A.16: Averaging kernel and gain matrix of full state vector. Averaging kernel shows large elements in the lower left off-diagonal block, which shows the influence of the emissions state vector elements on the background estimation. These elements are mostly in the upwind cells. On the other side of the diagonal, elements are smaller and the influence of the background elements on the emissions estimation is visible. The gain matrix shows the influence of the measurements on the state vector elements. Due to the large correlations between background elements, the influences are quite similar across them. The largest influences of the measurements on the emissions state vector elements are for the Berlin city cell.

Acronyms

$[UC]^2$	Urban Climates Under Change
ARW	Advanced Research WRF
ATC	Atmospheric Thematic Center
AVENGERS	Attributing and Verifying European and National Greenhouse Gas and Aerosol Emissions and Reconciliation with Statistical Bottom-up Estimates
BASF	Badische Anilin- und Sodafabrik
BEM	Building Energy Model
BEP	Building Environment Parametrization
BMFTR	German Ministry of Research, Technology, and Space
BouLac	Bougeault-Lacarère
BTR	Biennial Tranparency Report
CAMS	Copernicus Atmosphere Monitoring Service
CDO	Climate Data Operators
CDR	carbon dioxide removal
CoCO2	Prototype system for a Copernicus CO_2 service
COP	Conference of Parties
CORINE	Coordination of Information on the Environment
CRDS	cavity ring-down spectrometer
CRS	coordinate reference system
DEM	digital elevation model
DKRZ	German Climate Computation Center
DLR	The German Aerospace Center
DST	daylight savings time
DWD	German Weather Service
E-PRTR	European Pollutant Release and Transfer Register
ECMWF	European Center for Medium-Range Weather Forecasts
EDGAR	Emissions Database for Global Atmospheric Research
EMEP/CEIP	European Monitoring and Evaluation Programme/Centre on Emission Inventories and Projections

ERA5	ECMWF Reanalysis v5
ERL	Environmental Research Laboratory
ESMF	Earth System Modelling Framework
ETF	Enhanced Transparency Framework
EU	European Union
EVI	enhanced vegetation index
FAA	Federal Aviation Administration
FFT	fast Fourier transform
FLEXPART	Flexible Particle dispersion model
GAINS	Greenhouse Gas – Air Pollution Interactions and Synergies
GHG	greenhouse gas
GKM	Großkraftwerk Mannheim
GMTED	Global Multi-resolution Terrain Elevation Data
GNFR	Gridded Nomenclature for Reporting
GOSAT	Greenhouse Gases Observing Satellite
GPP	gross primary production
GWP	global warming potential
GWP_{100}	100-year GWP
HRRR	High Resolution Rapid Refresh
HYSPLIT	Hybrid Single-Particle Lagrangian Integrated Trajectory Model
ICON	Icosahedral Non-hydrostatic Model
ICOS	Integrated Carbon Observation System
IER	Institute of Energy Economics and the Rational Use of Energy, University of Stuttgart
IFS	ECMWF Integrated Forecast System
IG3IS	Integrated Global Greenhouse Gas Information System
IGRA	Integrated Global Radiosonde Archive, version 2
IPCC	Intergovernmental Panel on Climate Change
ISL	inertial sublayer
ISS	International Space Station
ITMS	Integrated Greenhouse Gas Monitoring System
KDE	kernel density estimate
LCC	Lambert Conformal Conic
LCZ	Local Climate Zones
LEO	low Earth orbit
LLM	large language model
LOWESS	locally weighted scatter plot smoothing

LPDM	Lagrangian particle dispersion model
LSM	land-surface model
LST	local standard time
LSWI	leaf surface water index
LULUCF	land use, land use change, and forestry
MAB	mean absolute bias
MACRO-2018	Metropolitan Area CO_x RecOrd of Germany 2018
MB	mean bias
MLH	mixing layer height
MLH	mixing layer height
MM5	Revised MM5 Monin-Obukhov
MO	Monin-Obukhov (Janjic)
MODIS	Moderate-resolution Imaging Spectroradiometer
MOST	Monin-Obukhov similarity theory
MPAS	Model for Prediction Across Scales
MRV	monitoring, reporting, and verification system
MYJ	Mellor-Yamada-Janjic
MYNN	Mellor-Yamada-Nakanishi-Niino
NAM	North American Model
NCAR	National Center for Atmospheric Research
NCEI	National Centers for Environmental Information
NCEP	National Center for Environmental Prediction
NDC	Nationally Determined Contribution
NEE	net ecosystem exchange
NFR	Nomenclature for Reporting
NIES-TM	National Institute for Environmental Studies Japan Transport Model
NIR	National Inventory Report
NMM	Nonhydrostatic Mesoscale Model
NOAA	National Oceanic and Atmospheric Administration
Noah-MP	Noah-Multiparametrization Land Surface Model
NRL	Naval Research Laboratory
OSM	OpenStreetMap
OSSE	observing system simulation experiment
PAR	photosynthetically active radiation
PBL	planetary boundary layer
PDF	probability density function

PR	pull request
PRTR	Pollutant Release and Transfer Register
PUR	polycentric urban region
RMSE	root mean squared error
RSL	roughness sublayer
SLM	surface layer model
SLUCM	Single-Layer Urban Canopy Model
STILT	Stochastic Time-Inverted Lagrangian Transport model
SYNMAP	Synergetic Land Cover Product
TANSO	Thermal and Near infrared Sensor for Carbon Observation
TKE	turbulent kinetic energy
TNO	Netherlands Organization for Applied Scientific Research
UBL	urban boundary layer
UCM	urban canopy model
UN	United Nations
UNFCCC	UN Framework Convention on Climate Change
USGS	US Geological Survey
VPRM	Vegetation Photosynthesis and Respiration Model
WDCC	Wold Data Center for Climate
WIS	WMO Information System
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting Model
WRF-Chem	WRF coupled with Chemistry
YSU	Yonsei University

List of Figures

List of Tables

Bibliography

Abernathey, R. P., J. J. M. Busecke, T. A. Smith, J. D. Deauna, A. Banihirwe, T. Nicholas, F. Fernandes, B. James, R. Dussin, D. A. Cherian, R. Caneill, A. Sinha, L. Uieda, W. Rath, D. Balwada, N. C. Constantinou, A. Ponte, Y. Zhou, T. Uchida, and J. Thielen (2025). `xgcm`. Version 0.8.1. DOI: 10.5281/zenodo.7348619.

Aigner, P., J. Chen, F. Böhm, M. Chariot, L. Emmenegger, L. Frölich, S. Grange, D. Kühbacher, K. Kürzinger, O. Laurent, M. Makowski, P. Rubli, A. Schmitt, and A. Wenzel (2025). "ACROPOLIS: Munich Urban CO2 Sensor Network". DOI: 10.5194/egusphere-2025-4157.

Amt für Statistik Berlin-Brandenburg (2025). *Energie- und CO_2-Bilanz*. URL: https://web.archive.org/web/20250927205413/https://download.statistik-berlin-brandenburg.de/7158c16f221e4d65/aa951a320938/energie-zeitreihe-2022-energiebilanz.xlsx.

Arakawa, A. and V. R. Lamb (1977). "Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model". In: *General Circulation Models of the Atmosphere*. Elsevier, pp. 173–265. DOI: 10.1016/b978-0-12-460817-7.50009-4.

Arnold, D., P. Seibert, H. Nagai, G. Wotawa, P. Skomorowski, K. Baumann-Stanzer, E. Polreich, M. Langer, A. Jones, M. Hort, S. Andronopoulos, J. G. Bartzis, E. Davakis, P. Kaufmann, and A. Vargas (2013). "Lagrangian Models for Nuclear Studies: Examples and Applications". In: *Lagrangian Modeling of the Atmosphere*. American Geophysical Union, pp. 329–348. DOI: 10.1029/2012gm001294.

Bahreini, R., B. Ervens, A. M. Middlebrook, C. Warneke, J. A. de Gouw, P. F. DeCarlo, J. L. Jimenez, C. A. Brock, J. A. Neuman, T. B. Ryerson, H. Stark, E. Atlas, J. Brioude, A. Fried, J. S. Holloway, J. Peischl, D. Richter, J. Walega, P. Weibring, A. G. Wollny, and F. C. Fehsenfeld (2009). "Organic aerosol formation in urban and industrial plumes near Houston and Dallas, Texas". *Journal of Geophysical Research: Atmospheres* 114:D7. ISSN: 0148-0227. DOI: 10.1029/2008jd011493.

Bisht, J. S. H., P. K. Patra, M. Takigawa, Y. Kanaya, M. Yamaguchi, T. Machida, and H. Tanimoto (2025). "High-Resolution Simulation of CO2 Using WRF-GHG Over the Kanto Region in Japan". *Journal of Geophysical Research: Atmospheres* 130:16. ISSN: 2169-8996. DOI: 10.1029/2025jd043589.

Blasiak, R., J. Spijkers, K. Tokunaga, J. Pittman, N. Yagi, and H. Österblom (2017). "Climate change and marine fisheries: Least developed countries top global index of vulnerability". *PLOS ONE* 12:6. Ed. by B. R. MacKenzie, e0179632. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0179632.

Boon, A., G. Broquet, D. J. Clifford, F. Chevallier, D. M. Butterfield, I. Pison, M. Ramonet, J.-D. Paris, and P. Ciais (2016). "Analysis of the potential of near-ground measurements of CO_2 and CH_4 in London, UK, for the monitoring of city-scale emissions using an atmospheric transport model". *Atmospheric Chemistry and Physics* 16:11, pp. 6735–6756. DOI: 10.5194/acp-16-6735-2016.

Boschetti, F., V. Thouret, G. J. Maenhout, K. U. Totsche, J. Marshall, and C. Gerbig (2018). "Multi-species inversion and IAGOS airborne data for a better constraint of continental-scale fluxes". *Atmospheric Chemistry and Physics* 18:13, pp. 9225–9241. ISSN: 1680-7324. DOI: 10.5194/acp-18-9225-2018.

Bougeault, P. and P. Lacarrère (1989). "Parameterization of Orography-Induced Turbulence in a Mesobeta–Scale Model". *Monthly Weather Review* 117:8, pp. 1872–1890. ISSN: 1520-0493. DOI: 10.1175/1520-0493(1989)117<1872:pooiti>2.0.co;2.

Bréon, F. M., G. Broquet, V. Puygrenier, F. Chevallier, I. Xueref-Remy, M. Ramonet, E. Dieudonné, M. Lopez, M. Schmidt, O. Perrussel, and P. Ciais (2015). "An attempt at estimating Paris area CO_2 emissions from atmospheric concentration measurements". *Atmospheric Chemistry and Physics* 15:4, pp. 1707–1724. DOI: 10.5194/acp-15-1707-2015.

Breuer, H., Z. Zempléni, and Á. Varga (2021). "Application of the complete CORINE land covers for modelling in WRF model". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu21-2940.

Brioude, J., D. Arnold, A. Stohl, M. Cassiani, D. Morton, P. Seibert, W. Angevine, S. Evan, A. Dingwell, J. D. Fast, R. C. Easter, I. Pisso, J. Burkhart, and G. Wotawa (2013). "The Lagrangian particle dispersion model FLEXPART-WRF version 3.1". *Geoscientific Model Development* 6:6, pp. 1889–1904. DOI: 10.5194/gmd-6-1889-2013.

Brunner, D., G. Kuhlmann, J. Marshall, V. Clément, O. Fuhrer, G. Broquet, A. Löscher, and Y. Meijer (2019). "Accounting for the vertical distribution of emissions in atmospheric CO2 simulations". *Atmospheric Chemistry and Physics* 19:7, pp. 4541–4559. ISSN: 1680-7324. DOI: 10.5194/acp-19-4541-2019.

Buchhorn, M., B. Smets, L. Bertels, B. D. Roo, M. Lesiv, N.-E. Tsendbazar, L. Li, and A. Tarko (2020). *Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Product User Manual*. en. Technical report. DOI: 10.5281/ZENODO.3938963.

Buckingham, E. (1914). "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations". *Physical Review* 4:4, pp. 345–376. ISSN: 0031-899X. DOI: 10.1103/physrev.4.345.

C3S (2018). *ERA5 hourly data on single levels from 1940 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/CDS.ADBB2D47.

– (2019). *ERA5-Land hourly data from 1950 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/CDS.E2161BAC.

C40 Cities (2025). *Cities Race to Zero*. URL: https://web.archive.org/web/20250829014228/https://www.c40.org/what-we-do/building-a-movement/cities-race-to-zero/.

Callewaert, S., J. Brioude, B. Langerock, V. Duflot, D. Fonteyn, J.-F. Müller, J.-M. Metzger, C. Hermans, N. Kumps, M. Ramonet, M. Lopez, E. Mahieu, and M. De Mazière (2022). "Analysis of CO2, CH4, and CO surface and column concentrations observed at Réunion Island by assessing WRF-Chem simulations". *Atmospheric Chemistry and Physics* 22:11, pp. 7763–7792. ISSN: 1680-7324. DOI: 10.5194/acp-22-7763-2022.

Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, J. Rising, A. Rode, H. K. Seo, A. Viaene, J. Yuan, and A. T. Zhang (2022). "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits". *The Quarterly Journal of Economics* 137:4, pp. 2037–2105. ISSN: 1531-4650. DOI: 10.1093/qje/qjac020.

Chen, F. (2006). "Current status of urban modeling in the community Weather Research and Forecast (WRF) model". In: *Sixth Symposium on the Urban Environment*.

Chen, F. and J. Dudhia (2001). "Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity". *Monthly Weather Review* 129:4, pp. 569–585. ISSN: 1520-0493. DOI: 10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2.

Chen, F., S. Miao, M. Tewari, J.-W. Bao, and H. Kusaka (2011). "A numerical study of interactions between surface forcing and sea breeze circulations and their effects on stagnation in the greater Houston area". *Journal of Geophysical Research: Atmospheres* 116, D12. ISSN: 2156-2202. DOI: 10.1029/2010JD015533.

Cogliati, S., F. Sarti, L. Chiarantini, M. Cosi, R. Lorusso, E. Lopinto, F. Miglietta, L. Genesio, L. Guanter, A. Damm, S. Pérez-López, D. Scheffler, G. Tagliabue, C. Panigada, U. Rascher, T. Dowling, C. Giardino, and R. Colombo (2021). "The PRISMA imaging spectroscopy mission: overview and first performance analysis". *Remote Sensing of Environment* 262, p. 112499. ISSN: 0034-4257. DOI: 10.1016/j.rse.2021.112499.

Cooper, O. R., D. D. Parrish, A. Stohl, M. Trainer, P. Nédélec, V. Thouret, J. P. Cammas, S. J. Oltmans, B. J. Johnson, D. Tarasick, T. Leblanc, I. S. McDermid, D. Jaffe, R. Gao, J. Stith,

T. Ryerson, K. Aikin, T. Campos, A. Weinheimer, and M. A. Avery (2010). "Increasing springtime ozone mixing ratios in the free troposphere over western North America". *Nature* 463:7279, pp. 344–348. ISSN: 1476-4687. DOI: 10.1038/nature08708.

Copernicus (2022). *Copernicus DEM*. DOI: 10.5270/esa-c5d3d65.

Copin, Y. (2021). `Taylor diagram for python/matplotlib`. Version 2018-12-06. DOI: 10.5281/zenodo.5548061.

Crippa, M., D. Guizzardi, E. Pisoni, E. Solazzo, A. Guion, M. Muntean, A. Florczyk, M. Schiavina, M. Melchiorri, and A. F. Hutfilter (2021). "Global anthropogenic emissions in urban areas: patterns, trends, and challenges". *Environmental Research Letters* 16:7, p. 074033. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ac00e2.

Demuzere, M., D. Argüeso, A. Zonato, and J. Kittner (2022). "W2W: A Python package that injects WUDAPT's Local Climate Zone information in WRF". *Journal of Open Source Software* 7:76, p. 4432. ISSN: 2475-9066. DOI: 10.21105/joss.04432.

Diémoz, H., F. Barnaba, T. Magri, G. Pession, D. Dionisi, S. Pittavino, I. K. F. Tombolato, M. Campanelli, L. S. Della Ceca, M. Hervo, L. Di Liberto, L. Ferrero, and G. P. Gobbi (2019). "Transport of Po Valley aerosol pollution to the northwestern Alps – Part 1: Phenomenology". *Atmospheric Chemistry and Physics* 19:5, pp. 3065–3095. ISSN: 1680-7324. DOI: 10.5194/acp-19-3065-2019.

Dröge, R., H. D. van der Gon, O. Perrussel, L. David, P. Aigner, D. Kuehbacher, J. Chen, J. Hinderer, D. Brunner, and L. Constantin (2024). *Final version of high-resolution city emission inventory for GHGs and co-emitted species for 2018, 2020 and 2022*. Deliverable. ICOS Cities. URL: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e50eb25a79&appId=PPGMS (visited on 2025-11-03).

Dury, A., A. Zonato, B. Kirk, N. Sobhani, C. He, T.-S. Lin, C. Li, J. Lewandoswki, J. Ruppert, J. Olson, A. Ukhov, L. FIta, L. Pilz, A. Martilli, M. Landreau, M. Chen, R. Conrick, R. Gilliam, J. Pleim, S. Osipov, S.-Y. Hong, T. Spero, and T. Mansell (2025). *WRF Version v4.7.0*. NCAR. URL: https://github.com/wrf-model/WRF/releases/tag/v4.7.0.

ECMWF (2021). "IFS Documentation CY47R3 - Part IV Physical processes". eng. In: ECMWF. DOI: 10.21957/eyrpir4vj.

Feng, S., S. Feng, S. Feng, T. Lauvaux, T. Lauvaux, S. Newman, P. Rao, R. Ahmadov, R. Ahmadov, A. Deng, L. I. Díaz-Isaac, R. M. Duren, M. L. Fischer, C. Gerbig, K. R. Gurney, J. Huang, S. Jeong, Z. Li, C. E. Miller, D. O'Keeffe, R. Patarasuk, S. P. Sander, Y. Song, K. W. Wong, K. W. Wong, and Y. L. Yung (2016). "Los Angeles megacity: a high-resolution land-atmosphere modelling system for urban CO_2 emissions". *Atmospheric Chemistry and Physics* 16, pp. 9019–9045. ISSN: 1680-7324. DOI: 10.5194/acp-16-9019-2016.

Fisher, J. A., L. T. Murray, D. B. A. Jones, and N. M. Deutscher (2017). "Improved method for linear carbon monoxide simulation and source attribution in atmospheric chemistry models illustrated using GEOS-Chem v9". *Geoscientific Model Development* 10:11, pp. 4129–4144. ISSN: 1991-9603. DOI: 10.5194/gmd-10-4129-2017.

Flesch, T. K., J. D. Wilson, and E. Yee (1995). "Backward-Time Lagrangian Stochastic Dispersion Models and Their Application to Estimate Gaseous Emissions". *Journal of Applied Meteorology and Climatology* 34:6, pp. 1320–1332. DOI: 10.1175/1520-0450(1995)034<1320:BTLSDM>2.0.CO;2.

Friedlingstein, P., M. O'Sullivan, M. W. Jones, R. M. Andrew, J. Hauck, P. Landschützer, C. Le Quéré, H. Li, I. T. Luijkx, A. Olsen, G. P. Peters, W. Peters, J. Pongratz, C. Schwingshackl, S. Sitch, J. G. Canadell, P. Ciais, R. B. Jackson, S. R. Alin, A. Arneth, V. Arora, N. R. Bates, M. Becker, N. Bellouin, C. F. Berghoff, H. C. Bittig, L. Bopp, P. Cadule, K. Campbell, M. A. Chamberlain, N. Chandra, F. Chevallier, L. P. Chini, T. Colligan, J. Decayeux, L. M. Djeutchouang, X. Dou, C. Duran Rojas, K. Enyo, W. Evans, A. R. Fay, R. A. Feely, D. J. Ford, A. Foster, T. Gasser, M. Gehlen, T. Gkritzalis, G. Grassi, L. Gregor, N. Gruber, Ö. Gürses, I. Harris, M. Hefner, J. Heinke, G. C. Hurtt, Y. Iida, T. Ilyina, A. R. Jacobson, A. K. Jain, T. Jarníková, A. Jersild, F. Jiang, Z. Jin, E. Kato, R. F. Keeling, K. Klein Goldewijk, J. Knauer, J. I. Korsbakken, X. Lan, S. K. Lauvset, N. Lefèvre, Z. Liu, J. Liu, L. Ma, S. Maksyutov, G. Marland, N. Mayot, P. C. McGuire, N. Metzl, N. M. Monacci, E. J. Morgan, S.-I. Nakaoka, C. Neill, Y. Niwa, T. Nützel, L. Olivier, T. Ono, P. I. Palmer, D. Pierrot, Z. Qin, L. Resplandy, A. Roobaert, T. M. Rosan, C. Rödenbeck, J. Schwinger, T. L. Smallman, S. M. Smith, R. Sospedra-Alfonso, T. Steinhoff, Q. Sun, A. J. Sutton, R. Séférian, S. Takao, H. Tatebe, H. Tian, B. Tilbrook, O. Torres, E. Tourigny, H. Tsujino, F. Tubiello, G. van der Werf, R. Wanninkhof, X. Wang, D. Yang, X. Yang, Z. Yu, W. Yuan, X. Yue, S. Zaehle, N. Zeng, and J. Zeng (2025). "Global Carbon Budget 2024". *Earth System Science Data* 17:3, pp. 965–1039. ISSN: 1866-3516. DOI: 10.5194/essd-17-965-2025.

Ganshin, A., T. Oda, M. Saito, S. Maksyutov, V. Valsala, R. J. Andres, R. E. Fisher, D. Lowry, A. Lukyanov, H. Matsueda, E. G. Nisbet, M. Rigby, Y. Sawa, R. Toumi, K. Tsuboi, A. Varlagin, and R. Zhuravlev (2012). "A global coupled Eulerian-Lagrangian model and 1 × 1 km CO2 surface flux dataset for high-resolution atmospheric CO2 transport simulations". *Geoscientific Model Development* 5:1, pp. 231–243. ISSN: 1991-9603. DOI: 10.5194/gmd-5-231-2012.

Gerbig, C. and F.-T. Koch (2024). *Biosphere-atmosphere exchange fluxes for CO2 from the Vegetation Photosynthesis and Respiration Model VPRM for 2022-2023*. ICOS-ERIC - Carbon Portal. DOI: 10.18160/R5HS-YKW0.

Glauch, T., J. Marshall, C. Gerbig, S. Botía, M. Gałkowski, S. N. Vardag, and A. Butz (2025). "pyVPRM: A next-generation Vegetation Photosynthesis and Respiration Model for the post-MODIS era". *EGUsphere*. DOI: 10.5194/egusphere-2024-3692.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008). "Performance metrics for climate models". en. *Journal of Geophysical Research: Atmospheres* 113:D6. ISSN: 2156-2202. DOI: 10.1029/2007JD008972.

Global Covenant of Mayors on Climate & Energy (2025). *Global Covenant of Mayors*. URL: https://web.archive.org/web/20251011142803/https://www.globalcovenantofmayors.org/.

Grange, S. K., P. Rubli, A. Fischer, D. Brunner, C. Hueglin, and L. Emmenegger (2025). "The ZiCOS-M CO2 sensor network: measurement performance and CO2 variability across Zurich". *Atmospheric Chemistry and Physics* 25:5, pp. 2781–2806. ISSN: 1680-7324. DOI: 10.5194/acp-25-2781-2025.

Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder (2005). "Fully coupled "online" chemistry within the WRF model". *Atmospheric Environment* 39:37, pp. 6957–6975. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2005.04.027.

Guanter, L., H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek, C. Straif, S. Fischer, S. Schrader, T. Storch, U. Heiden, A. Mueller, M. Bachmann, H. Mühle, R. Müller, M. Habermeyer, A. Ohndorf, J. Hill, H. Buddenbaum, P. Hostert, S. Van der Linden, P. Leitão, A. Rabe, R. Doerffer, H. Krasemann, H. Xi, W. Mauser, T. Hank, M. Locherer, M. Rast, K. Staenz, and B. Sang (2015). "The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation". *Remote Sensing* 7:7, pp. 8830–8857. ISSN: 2072-4292. DOI: 10.3390/rs70708830.

Günther, D., S. Adam, L. Ahrem, M. Bernicke, A. Bertram, T. Bolland, C. Bottcher, K. Brand, O. Can, M. Dreher, D. Drohsin, K. Dunger, C. Elsner, T. Fiedler, R. Fuß, A. Gensior, G. Gohlisch, S. Gores, U. Gromke, C. Grosse, J. Grubel, E. Gruneberg, D. Gunther, K. Hausmann, T. Hermann, P. Holting, P. Icha, K. Juhrich, K. Kessler, R. Kludt, K. Koppe, A. Koska, M. Kotzulla, B. Laggner, A. Laggner, J. Langenfeld, C. Lehmann, B. Lenzen, S. Leuthold, C. Liesegang, K. Martens, C. Mielke, J. Niebuhr, K. Pannier, S. Plickert, C. Proske, K. Oehmichen, J. Reichel, A. Reichart, T. Riedel, C. Rosemann, S. Ruter, S. Schiller, W. Stumer, C. Vos, N. Wellbrock, F. Wetzel, J. Zheng, and D. Ziche (2025). *Submission under the United Nations Framework Convention 2025. National Inventory Report for the German Greenhouse Gas Inventory 1990 – 2023*. Technical report. German Environment Agency. URL: https://web.archive.org/web/20251126103000/https://www.umweltbundesamt.de/sites/default/files/medien/11850/publikationen/40_2025_cc.pdf (visited on 2025-11-26).

Han, J.-Y. and S.-Y. Hong (2018). "Precipitation Forecast Experiments Using the Weather Research and Forecasting (WRF) Model at Gray-Zone Resolutions". *Weather and Forecasting* 33:6, pp. 1605–1616. ISSN: 1520-0434. DOI: 10.1175/waf-d-18-0026.1.

Hang, J., L. Zeng, X. Li, and D. Wang (2024). "Evaluation of a single-layer urban energy balance model using measured energy fluxes by scaled outdoor experiments in humid subtropical climate". *Building and Environment* 254, p. 111364. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2024.111364.

Heaviside, C., H. Macintyre, and S. Vardoulakis (2017). "The Urban Heat Island: Implications for Health in a Changing Environment". *Current Environmental Health Reports* 4:3, pp. 296–305. ISSN: 2196-5412. DOI: 10.1007/s40572-017-0150-3.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut (2020a). "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* 146:730, pp. 1999–2049. ISSN: 1477-870X. DOI: 10.1002/qj.3803.

– (2020b). "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* 146:730, pp. 1999–2049. ISSN: 1477-870X. DOI: 10.1002/qj.3803.

Ho, D., M. Gałkowski, F. Reum, S. Botía, J. Marshall, K. U. Totsche, and C. Gerbig (2024). "Recommended coupling to global meteorological fields for long-term tracer simulations with WRF-GHG". *Geoscientific Model Development* 17:20, pp. 7401–7422. ISSN: 1991-9603. DOI: 10.5194/gmd-17-7401-2024.

Hong, S.-Y., Y. Noh, and J. Dudhia (2006). "A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes". *Monthly Weather Review* 134:9, pp. 2318–2341. ISSN: 0027-0644. DOI: 10.1175/mwr3199.1.

Hong, S.-Y. and H.-L. Pan (1996). "Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model". *Monthly Weather Review* 124:10, pp. 2322–2339. ISSN: 1520-0493. DOI: 10.1175/1520-0493(1996)124<2322:nblvdi>2.0.co;2.

Honnert, R., G. A. Efstathiou, R. J. Beare, J. Ito, A. Lock, R. Neggers, R. S. Plant, H. H. Shin, L. Tomassini, and B. Zhou (2020). "The Atmospheric Boundary Layer and the "Gray Zone" of Turbulence: A Critical Review". *Journal of Geophysical Research: Atmospheres* 125:13, e2019JD030317. ISSN: 2169-8996. DOI: 10.1029/2019JD030317.

Howes, C., R. Easter, R. Arthur, J. Mirocha, M. Göbel, S. Walters, R. Cabell, S. Roh, H.-J. Songa, A. Valmassoi, T. Mansell, M. Honnorat, L. Pilz, T. W. Juliano, J.-H. Kim, P. A.

Jimenez, J. Lee, T. Brummet, and T. Raupach (2021). *WRF Version v4.2.2*. NCAR. URL: https://github.com/wrf-model/WRF/releases/tag/v4.2.2.

Huijnen, V., A. Pozzer, J. Arteta, G. Brasseur, I. Bouarar, S. Chabrillat, Y. Christophe, T. Doumbia, J. Flemming, J. Guth, B. Josse, V. A. Karydis, V. Marécal, and S. Pelletier (2019). "Quantifying uncertainties due to chemistry modelling – evaluation of tropospheric composition simulations in the CAMS model (cycle 43R1)". *Geoscientific Model Development* 12:4, pp. 1725–1752. ISSN: 1991-9603. DOI: 10.5194/gmd-12-1725-2019.

ICOS RI, P. Bergamaschi, A. Colomb, M. De Mazière, L. Emmenegger, D. Kubistin, I. Lehner, K. Lehtinen, M. Leuenberger, C. Lund Myhre, M. V. Marek, S. O'Doherty, S. M. Platt, C. Plaß-Dülmer, M. Ramonet, C. Rennick, M. Schmidt, S. Zähle, F. Apadula, S. Arnold, P.-E. Blanc, D. Brunner, H. Chen, Ł. Chmura, L. Chmura, S. Conil, P. Cristofanelli, T. Di Iorio enea, G. Forster, A. Frumau, C. Gerbig, F. Gheusi, S. Hammer, L. Haszpra, J. Hatakka, M. Heliasz, A. Hensen, A. Hoheisel, T. Kneuer, E. Larmanou, T. Laurila, A. Leskinen, I. Levin, J. Levula, M. Lindauer, M. Lopez, C. Lunder, I. Mammarella, G. Manca, A. Manning, D. Martin, F. Meinhardt, M. Molnár, M. Mölder, J. Müller-Williams, S. M. Noe, J. Nęcki, M. Ottosson-Löfvenius, C. Philippon, S. Piacentino, J. Pitt, P. Rivas-Soriano, B. Scheeren, M. Schumacher, M. K. Sha, P. Smith, G. Spain, M. Steinbacher, L. L. Sørensen, A. Vermeulen, G. Vítková, I. Xueref-Remy, S. Zaehle, A. di Sarra, F. Conen, S. Henne, V. Kazan, Y.-A. Roulet, T. Biermann, M. Delmotte, D. Heltai, O. Hermansen, K. Komínková, O. Laurent, P. Marklund, J.-A. Morguí, J.-M. Pichon, D. Sferlazzo, K. Stanley, P. Trisolino, G. Zazzeri, ICOS Carbon Portal, ICOS Atmosphere Thematic Centre, ICOS Flask And Calibration Laboratory, ICOS Flask And Calibration Laboratory, and ICOS Central Radiocarbon Laboratory (2024). *European Obspack compilation of atmospheric carbon dioxide data from ICOS and non-ICOS European stations for the period 1972-2024; obspack_co2_466_GVeu_v10.0_20240729*. ICOS RI. ICOS ERIC - Carbon Portal. DOI: 10.18160/X450-GTAY.

IPCC (2022a). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by H. O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. Cambridge University Press, Cambridge, UK and New York, NY, USA. DOI: 10.1017/9781009325844.

– (2022b). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by P. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa,

S. Luz, and J. Malley. Cambridge University Press, Cambridge, UK and New York, NY, USA. DOI: 10.1017/9781009157926.

– (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* Ed. by P. Arias, M. Bustamante, I. Elgizouli, G. Flato, M. Howden, C. Méndez-Vallejo, J. J. Pereira, R. Pichs-Madruga, S. K. Rose, Y. Saheb, R. Sánchez Rodríguez, D. Ürge-Vorsatz, C. Xiao, N. Yassaa, J. Romero, J. Kim, E. F. Haites, Y. Jung, R. Stavins, A. Birt, M. Ha, D. J. A. Orendain, L. Ignon, S. Park, Y. Park, A. Reisinger, D. Cammaramo, A. Fischlin, J. S. Fuglestvedt, G. Hansen, C. Ludden, V. Masson-Delmotte, J. R. Matthews, K. Mintenbeck, A. Pirani, E. Poloczanska, N. Leprince-Ringuet, and C. Péan. DOI: 10.59327/ipcc/ar6-9789291691647.

Jägermeyr, J., C. Müller, A. C. Ruane, J. Elliott, J. Balkovic, O. Castillo, B. Faye, I. Foster, C. Folberth, J. A. Franke, K. Fuchs, J. R. Guarin, J. Heinke, G. Hoogenboom, T. Iizumi, A. K. Jain, D. Kelly, N. Khabarov, S. Lange, T.-S. Lin, W. Liu, O. Mialyk, S. Minoli, E. J. Moyer, M. Okada, M. Phillips, C. Porter, S. S. Rabin, C. Scheer, J. M. Schneider, J. F. Schyns, R. Skalsky, A. Smerald, T. Stella, H. Stephens, H. Webber, F. Zabel, and C. Rosenzweig (2021). "Climate impacts on global agriculture emerge earlier in new generation of climate and crop models". *Nature Food* 2:11, pp. 873–885. ISSN: 2662-1355. DOI: 10.1038/s43016-021-00400-y.

Jammalamadaka, S. R., A. Sengupta, and A. Sengupta (2001). *Topics in Circular Statistics*. en. Google-Books-ID: sKqWMGqQXQkC. World Scientific. ISBN: 9789812779267.

Jänicke, B., F. Meier, D. Fenner, U. Fehrenbach, A. Holtmann, and D. Scherer (2017). "Urban–rural differences in near-surface air temperature as resolved by the Central Europe Refined analysis (CER): sensitivity to planetary boundary layer schemes and urban canopy models". *International Journal of Climatology* 37:4, pp. 2063–2079. ISSN: 1097-0088. DOI: 10.1002/joc.4835.

Janjic, Z. (2002). "Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso Model". *NCEP Office Note* 436.

Jiang, L. and B. C. O'Neill (2017). "Global urbanization projections for the Shared Socioeconomic Pathways". *Global Environmental Change* 42, pp. 193–199. ISSN: 0959-3780. DOI: 10.1016/j.gloenvcha.2015.03.008.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante (2012). "A Revised Scheme for the WRF Surface Layer Formulation". *Monthly Weather Review* 140:3, pp. 898–918. ISSN: 1520-0493. DOI: 10.1175/mwr-d-11-00056.1.

Joshi, P., T.-S. Lin, C. He, and K. Lamer (2025). "Urban weather modeling using WRF: linking physical assumptions, code implementation, and observational needs". *Geoscientific Model Development* 18:20, pp. 7869–7890. ISSN: 1991-9603. DOI: 10.5194/gmd-18-7869-2025.

Jungmann, M., S. N. Vardag, F. Kutzner, F. Keppler, M. Schmidt, N. Aeschbach, U. Gerhard, A. Zipf, S. Lautenbach, A. Siegmund, T. Goeschl, and A. Butz (2022). "Zooming-in for climate action—hyperlocal greenhouse gas data for mitigation action?" *Climate Action* 1:1. ISSN: 2731-3263. DOI: 10.1007/s44168-022-00007-4.

Kennedy, J. (2024). *State of the Global Climate 2024*. Technical report. World Meteorological Organization. URL: https://web.archive.org/web/20250812145206/https://wmo.int/sites/default/files/2025-03/WMO-1368-2024_en.pdf (visited on 2025-08-12).

Kim, J., W. M. Berelson, N. E. Rollins, N. G. Asimow, C. Newman, R. C. Cohen, J. B. Miller, B. C. McDonald, J. Peischl, and S. J. Lehman (2025). "Observing Anthropogenic and Biogenic CO2 Emissions in Los Angeles Using a Dense Sensor Network". *Environmental Science & Technology* 59:7, pp. 3508–3517. ISSN: 1520-5851. DOI: 10.1021/acs.est.4c11392.

Klausner, T., M. Mertens, H. Huntrieser, M. Galkowski, G. Kuhlmann, R. Baumann, A. Fiehn, P. Jöckel, M. Pühl, and A. Roiger (2020). "Urban greenhouse gas emissions from the Berlin area: A case study using airborne CO2 and CH4 in situ observations in summer 2018". *Elem Sci Anth* 8. Ed. by D. Helmig and L. Bruhwiler. ISSN: 2325-1026. DOI: 10.1525/elementa.411.

Knapp, M., B. Hemmer, R. Kleinschek, M. Sindram, T. Schmitt, L. Pilz, B. Burger, and A. Butz (2022). "Towards Carbon Dioxide emission estimation with a stationary hyperspectral camera". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu22-3924.

Kuik, F., A. Lauer, G. Churkina, H. A. C. Denier van der Gon, D. Fenner, K. A. Mar, and T. M. Butler (2016). "Air quality modelling in the Berlin–Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data". *Geoscientific Model Development* 9:12, pp. 4339–4363. ISSN: 1991-959X. DOI: 10.5194/gmd-9-4339-2016.

Kunik, L., D. V. Mallia, K. R. Gurney, D. L. Mendoza, T. Oda, and J. C. Lin (2019). "Bayesian inverse estimation of urban CO2 emissions: Results from a synthetic data simulation over Salt Lake City, UT". *Elementa: Science of the Anthropocene* 7. Ed. by D. Helmig and L. Bruhwiler. DOI: 10.1525/elementa.375.

Kusaka, H. and F. Kimura (2004). "Coupling a Single-Layer Urban Canopy Model with a Simple Atmospheric Model: Impact on Urban Heat Island Simulation for an Idealized Case". *Journal of the Meteorological Society of Japan. Ser. II* 82:1, pp. 67–80. ISSN: 2186-9057. DOI: 10.2151/jmsj.82.67.

Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura (2001). "A Simple Single-Layer Urban Canopy Model For Atmospheric Models: Comparison With Multi-Layer And Slab Models". *Boundary-Layer Meteorology* 101:3, pp. 329–358. ISSN: 1573-1472. DOI: 10.1023/A:1019207923078.

Lama, S., S. Houweling, K. F. Boersma, I. Aben, H. A. C. Denier Van Der Gon, and M. C. Krol (2022). "Estimation of OH in urban plume using TROPOMI inferred NO2/CO". DOI: 10.1002/essoar.10510109.1.

Laprise, R. (1992). "The Euler Equations of Motion with Hydrostatic Pressure as an Independent Variable". *Monthly Weather Review* 120:1, pp. 197–207. ISSN: 1520-0493. DOI: 10.1175/1520-0493(1992)120<0197:teeomw>2.0.co;2.

Lauvaux, T., N. L. Miles, A. Deng, S. J. Richardson, M. O. Cambaliza, K. J. Davis, B. Gaudet, K. R. Gurney, J. Huang, D. O'Keefe, Y. Song, A. Karion, T. Oda, R. Patarasuk, I. Razlivanov, D. Sarmiento, P. Shepson, C. Sweeney, J. Turnbull, and K. Wu (2016). "High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX)". *Journal of Geophysical Research: Atmospheres* 121:10, pp. 5213–5236. ISSN: 2169-8996. DOI: 10.1002/2015jd024473.

Lenton, T. M., C. Xu, J. F. Abrams, A. Ghadiali, S. Loriani, B. Sakschewski, C. Zimm, K. L. Ebi, R. R. Dunn, J.-C. Svenning, and M. Scheffer (2023). "Quantifying the human cost of global warming". *Nature Sustainability* 6:10, pp. 1237–1247. ISSN: 2398-9629. DOI: 10.1038/s41893-023-01132-6.

Li, H., B. Liu, X. Ma, S. Jin, Y. Ma, Y. Zhao, and W. Gong (2021). "Evaluation of retrieval methods for planetary boundary layer height based on radiosonde data". *Atmospheric Measurement Techniques* 14:9, pp. 5977–5986. ISSN: 1867-8548. DOI: 10.5194/amt-14-5977-2021.

Lian, J., F.-M. Bréon, G. Broquet, T. Lauvaux, B. Zheng, M. Ramonet, I. Xueref-Remy, S. Kotthaus, M. Haeffelin, and P. Ciais (2021). "Sensitivity to the sources of uncertainties in the modeling of atmospheric CO_2 concentration within and in the vicinity of Paris". English. *Atmospheric Chemistry and Physics* 21:13, pp. 10707–10726. ISSN: 1680-7316. DOI: 10.5194/acp-21-10707-2021.

Lian, J., O. Laurent, M. Chariot, L. Lienhardt, M. Ramonet, H. Utard, T. Lauvaux, F.-M. Bréon, G. Broquet, K. Cucchi, L. Millair, and P. Ciais (2024). "Development and deployment of a mid-cost CO2 sensor monitoring network to support atmospheric inverse modeling for quantifying urban CO2 emissions in Paris". *Atmospheric Measurement Techniques* 17:19, pp. 5821–5839. ISSN: 1867-8548. DOI: 10.5194/amt-17-5821-2024.

Lian, J., L. Wu, F.-M. Bréon, G. Broquet, R. Vautard, T. S. Zaccheo, J. Dobler, and P. Ciais (2018). "Evaluation of the WRF-UCM mesoscale model and ECMWF global operational forecasts over the Paris region in the prospect of tracer atmospheric transport

modeling". *Elementa: Science of the Anthropocene* 6. Ed. by D. Helmig and G. Pfister. DOI: 10.1525/elementa.319.

Lin, D., B. Khan, M. Katurji, L. Bird, R. Faria, and L. E. Revell (2021). "WRF4PALM v1.0: a mesoscale dynamical driver for the microscale PALM model system 6.0". *Geoscientific Model Development* 14:5, pp. 2503–2524. DOI: 10.5194/gmd-14-2503-2021.

Lionel, C., B. Dominik, T. Joel, K. Corina, S. Michael, and K. Erik (2025). "Emiproc: A Python package for emission inventory processing". *Journal of Open Source Software* 10:105, p. 7509. ISSN: 2475-9066. DOI: 10.21105/joss.07509.

Liu, L., S. N. Gadde, and R. J. Stevens (2021). "Universal wind profile for conventionally neutral atmospheric boundary layers". *Physical review letters* 126:10, p. 104502.

Liu, Y., F. Chen, T. Warner, and J. Basara (2006). "Verification of a Mesoscale Data-Assimilation and Forecasting System for the Oklahoma City Area during the Joint Urban 2003 Field Project". *Journal of Applied Meteorology and Climatology* 45:7, pp. 912–929. ISSN: 1558-8424. DOI: 10.1175/jam2383.1.

Lopez-Coto, I., X. Ren, O. E. Salmon, A. Karion, P. B. Shepson, R. R. Dickerson, A. Stein, K. Prasad, and J. R. Whetstone (2020). "Wintertime CO2, CH4, and CO Emissions Estimation for the Washington, DC–Baltimore Metropolitan Area Using an Inverse Modeling Technique". *Environmental Science & Technology* 54:5, pp. 2606–2614. ISSN: 1520-5851. DOI: 10.1021/acs.est.9b06619.

Loprieno, L. L., M. Raffa, A. Campanale, M. Adinolfi, and P. Mercogliano (2026). "Comparative study of the COSMO and ICON models performances over Italy: Key insights from the impactful year 2023". *Atmospheric Research* 327, p. 108339. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2025.108339.

Lüken-Winkels, C., L. Pilz, M. Cassiani, I. Pisso, and S. N. Vardag (2024). "High-resolution inversion of Berlin city emissions – A synthetic study using FLEXPART-WRF for network optimization within ITMS". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-9996.

Lüken-Winkels, C., L. Pilz, S. Cello, and S. N. Vardag (2025). "Designing CO2 sensor networks for German cities: Insights from synthetic studies in Berlin and Munich ". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu25-10382.

Magnusson, L., D. Ackerley, Y. Bouteloup, J.-H. Chen, J. Doyle, P. Earnshaw, Y. C. Kwon, M. Köhler, S. T. K. Lang, Y.-J. Lim, M. Matsueda, T. Matsunobu, R. McTaggart-Cowan, A. Reinecke, M. Yamaguchi, and L. Zhou (2022). "Skill of Medium-Range Forecast Models Using the Same Initial Conditions". *Bulletin of the American Meteorological Society* 103:9, E2050–E2068. ISSN: 1520-0477. DOI: 10.1175/bams-d-21-0234.1.

Marshall, J. (2022). "VPRM fluxes for CoCO2 WP2". Personal Communication.

Martilli, A., A. Clappier, and M. W. Rotach (2002). "An Urban Surface Exchange Parameterisation for Mesoscale Models". *Boundary-Layer Meteorology* 104:2, pp. 261–304. ISSN: 1573-1472. DOI: 10.1023/A:1016099921195.

May, R. M., K. H. Goebbert, J. E. Thielen, J. R. Leeman, M. D. Camron, Z. Bruick, E. C. Bruning, R. P. Manser, S. C. Arms, and P. T. Marsh (2022). "MetPy: A Meteorological Python Library for Data Analysis and Visualization". *Bulletin of the American Meteorological Society* 103:10, E2273–E2284. DOI: 10.1175/BAMS-D-21-0125.1.

McKain, K., A. Down, S. M. Raciti, J. Budney, L. R. Hutyra, C. Floerchinger, S. C. Herndon, T. Nehrkorn, M. S. Zahniser, R. B. Jackson, N. Phillips, and S. C. Wofsy (2015). "Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts". *Proceedings of the National Academy of Sciences* 112:7, pp. 1941–1946. ISSN: 1091-6490. DOI: 10.1073/pnas.1416261112.

Meijers, E., M. Hoogerbrugge, and R. Cardoso (2017). "Beyond Polycentricity: Does Stronger Integration Between Cities in Polycentric Urban Regions Improve Performance?" *Tijdschrift voor Economische en Sociale Geografie* 109:1, pp. 1–21. ISSN: 1467-9663. DOI: 10.1111/tesg.12292.

Mellor, G. L. and T. Yamada (1982). "Development of a turbulence closure model for geophysical fluid problems". *Reviews of Geophysics* 20:4, pp. 851–875. ISSN: 1944-9208. DOI: 10.1029/rg020i004p00851.

Menut, L., B. Bessagnet, D. Khvorostyanov, M. Beekmann, N. Blond, A. Colette, I. Coll, G. Curci, G. Foret, A. Hodzic, S. Mailler, F. Meleux, J.-L. Monge, I. Pison, G. Siour, S. Turquety, M. Valari, R. Vautard, and M. G. Vivanco (2013). "CHIMERE 2013: a model for regional atmospheric composition modelling". *Geoscientific Model Development* 6:4, pp. 981–1028. ISSN: 1991-9603. DOI: 10.5194/gmd-6-981-2013.

Miles, N. L., S. J. Richardson, T. Lauvaux, K. J. Davis, N. V. Balashov, A. Deng, J. C. Turnbull, C. Sweeney, K. R. Gurney, R. Patarasuk, I. Razlivanov, M. O. L. Cambaliza, and P. B. Shepson (2017). "Quantification of urban atmospheric boundary layer greenhouse gas dry mole fraction enhancements in the dormant season: Results from the Indianapolis Flux Experiment (INFLUX)". *Elementa: Science of the Anthropocene* 5. Ed. by D. Helmig and P. Palmer. ISSN: 2325-1026. DOI: 10.1525/elementa.127.

Monin, A. S. and A. M. Obukhov (1954). "Basic laws of turbulent mixing in the surface layer of the atmosphere". Trans. by K. McNaughton. *Tr. Akad. Nauk SSSR Geophiz. Inst.* 24:151, pp. 163–187.

Monteiro, V. C., J. C. Turnbull, N. L. Miles, K. J. Davis, Z. Barkley, and A. Deng (2024). "Assimilating Morning, Evening, and Nighttime Greenhouse Gas Observations in Atmospheric Inversions". *Journal of Geophysical Research: Atmospheres* 129:17. ISSN: 2169-8996. DOI: 10.1029/2024jd040998.

Munassar, S., G. Monteil, M. Scholze, U. Karstens, C. Rödenbeck, F.-T. Koch, K. U. Totsche, and C. Gerbig (2023). "Why do inverse models disagree? A case study with two European CO2 inversions". *Atmospheric Chemistry and Physics* 23:4, pp. 2813–2828. ISSN: 1680-7324. DOI: 10.5194/acp-23-2813-2023.

Nakanishi, M. and H. Niino (2009). "Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer". *Journal of the Meteorological Society of Japan. Ser. II* 87:5, pp. 895–912. ISSN: 2186-9057. DOI: 10.2151/jmsj.87.895.

Nayna Schwerdtle, P., E. Cavan, L. Pilz, S. D. Oggioni, A. Crosta, V. Kaleyeva, P. H. Karim, F. Szarvas, T. Naryniecki, and M. Jungmann (2023). "Interlinkages between Climate Change Impacts, Public Attitudes, and Climate Action—Exploring Trends before and after the Paris Agreement in the EU". *Sustainability* 15:9, p. 7542. ISSN: 2071-1050. DOI: 10.3390/su15097542.

Nehrkorn, T., J. Henderson, M. Leidner, M. Mountain, J. Eluszkiewicz, K. McKain, and S. Wofsy (2013). "WRF Simulations of the Urban Circulation in the Salt Lake City Area for CO_2 Modeling". *Journal of Applied Meteorology and Climatology* 52:2, pp. 323–340. DOI: 10.1175/jamc-d-12-061.1.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia (2011). "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements". *Journal of Geophysical Research* 116:D12. ISSN: 0148-0227. DOI: 10.1029/2010jd015139.

Niu, S., W. Chen, L. L. Liáng, C. A. Sierra, J. Xia, S. Wang, M. Heskel, K. F. Patel, B. Bond-Lamberty, J. Wang, G. Yvon-Durocher, M. U. F. Kirschbaum, O. K. Atkin, Y. Huang, G. Yu, and Y. Luo (2024). "Temperature responses of ecosystem respiration". *Nature Reviews Earth & Environment* 5:8, pp. 559–571. ISSN: 2662-138X. DOI: 10.1038/s43017-024-00569-3.

Oke, T. R. (2017). *Urban climates*. Ed. by G. Mills, A. Christen, and J. A. Voogt. Includes bibliographical references and index. Cambridge University Press, Cambridge. 1 p. ISBN: 9781139016476.

Pilz, L., H. Denier van der Gon, T. Glauch, J. Marshall, and S. N. Vardag (2025a). *MACRO-2018 - High-Resolution Simulation of CO and CO2 concentrations over German Metropolitan Areas for 2018 using WRF-Chem*. Ed. by L. Pilz. World Data Center for Climate (WDCC) at DKRZ. DOI: 10.26050/WDCC/MACRO-2018.

Pilz, L. (2025a). `portraitpy`. URL: https://github.com/ATMO-IUP-UHEI/portraitpy (visited on 2025-11-03).

– (2025b). `wps_xr`. URL: https://github.com/lpilz/wps_xr (visited on 2025-11-03).

Pilz, L., A. Banihirwe, J. Thielen, K. Paul, D. Cherian, K. FitzGerald, W. Rudisill, C. Lüken-Winkels, and J. Kent (2025b). xWRF. Version 0.0.5. URL: https://github.com/xarray-contrib/xwrf (visited on 2025-11-03).

Pilz, L., M. Galkowski, J. Fallmann, F. Chen, A. Butz, and S. N. Vardag (2023). "Optimizing High-Resolution Simulations with the Weather Research and Forecasting (WRF) Model for the German Rhine-Neckar Metropolitan Region". In: Wiley. DOI: 10.22541/essoar.167591054.45628540/v1.

Pilz, L., C. Lüken-Winkels, M. Gałkowski, D. Ho, F. Chen, and S. N. Vardag (2024). "High-resolution meteorological CO2 enhancements of German metropolitan areas using WRF". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-8915.

Pilz, L., C. Lüken-Winkels, M. Gałkowski, D. Ho, C. Gerbig, F. Chen, and S. N. Vardag (2026). "Evaluation of high-resolution WRF simulation in urban areas — Effect of different physics schemes on simulation performance in the Rhine-Main-Neckar area". *Atmospheric Research* 328, p. 108435. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2025.108435.

Pilz, L., S. Vardag, R. Kleinschek, S. Hammer, and A. Butz (2021). "Towards an integrated study of urban CO2 emissions". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu21-2724.

Pilz, L. N., S. N. Vardag, J. Fallmann, and A. Butz (2022). "Hochaufgelöste Simulation von meteorologischen Feldern in urbanen Räumen mit dem Weather Research and Forecasting (WRF) Modell". In: Copernicus GmbH. DOI: 10.5194/dach2022-69.

Pisso, I., E. Sollum, H. Grythe, N. I. Kristiansen, M. Cassiani, S. Eckhardt, D. Arnold, D. Morton, R. L. Thompson, C. D. Groot Zwaaftink, N. Evangeliou, H. Sodemann, L. Haimberger, S. Henne, D. Brunner, J. F. Burkhart, A. Fouilloux, J. Brioude, A. Philipp, P. Seibert, and A. Stohl (2019). "The Lagrangian particle dispersion model FLEXPART version 10.4". *Geoscientific Model Development* 12:12, pp. 4955–4997. DOI: 10.5194/gmd-12-4955-2019.

Ramonet, M., P. Ciais, F. Apadula, J. Bartyzel, A. Bastos, P. Bergamaschi, P. E. Blanc, D. Brunner, L. Caracciolo di Torchiarolo, F. Calzolari, H. Chen, L. Chmura, A. Colomb, S. Conil, P. Cristofanelli, E. Cuevas, R. Curcoll, M. Delmotte, A. di Sarra, L. Emmenegger, G. Forster, A. Frumau, C. Gerbig, F. Gheusi, S. Hammer, L. Haszpra, J. Hatakka, L. Hazan, M. Heliasz, S. Henne, A. Hensen, O. Hermansen, P. Keronen, R. Kivi, K. Komínková, D. Kubistin, O. Laurent, T. Laurila, J. V. Lavric, I. Lehner, K. E. J. Lehtinen, A. Leskinen, M. Leuenberger, I. Levin, M. Lindauer, M. Lopez, C. L. Myhre, I. Mammarella, G. Manca, A. Manning, M. V. Marek, P. Marklund, D. Martin, F. Meinhardt, N. Mihalopoulos, M. Mölder, J. A. Morgui, J. Necki, S. O'Doherty, C. O'Dowd, M. Ottosson, C. Philippon, S. Piacentino, J. M. Pichon, C. Plass-Duelmer, A. Resovsky, L. Rivier, X. Rodó, M. K. Sha, H. A. Scheeren, D. Sferlazzo, T. G. Spain, K. M. Stanley, M. Steinbacher, P. Trisolino,

A. Vermeulen, G. Vítková, D. Weyrauch, I. Xueref-Remy, K. Yala, and C. Yver Kwok (2020). "The fingerprint of the summer 2018 drought in Europe on ground-based atmospheric CO2measurements". *Philosophical Transactions of the Royal Society B: Biological Sciences* 375:1810, p. 20190513. ISSN: 1471-2970. DOI: 10.1098/rstb.2019.0513.

Ribeiro, H. V., D. Rybski, and J. P. Kropp (2019). "Effects of changing population or density on urban carbon dioxide emissions". en. *Nature Communications* 10:1, p. 3204. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11184-y.

Ribeiro, I., A. Martilli, M. Falls, A. Zonato, and G. Villalba (2021). "Highly resolved WRF-BEP/BEM simulations over Barcelona urban area with LCZ". *Atmospheric Research* 248, p. 105220. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2020.105220.

Ripple, W. J., C. Wolf, M. E. Mann, J. Rockström, J. W. Gregg, C. Xu, N. Wunderling, S. E. Perkins-Kirkpatrick, R. Schaeffer, W. J. Broadgate, T. M. Newsome, E. Shuckburgh, and P. H. Gleick (2025). "The 2025 state of the climate report: a planet on the brink". *BioScience*. ISSN: 1525-3244. DOI: 10.1093/biosci/biaf149.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. WORLD SCIENTIFIC. DOI: 10.1142/3171.

Roedel, W. and T. Wagner (2017). *Physik unserer Umwelt: Die Atmosphäre*. Springer Berlin Heidelberg. ISBN: 9783662542583. DOI: 10.1007/978-3-662-54258-3.

Salamanca, F. and A. Martilli (2009). "A new Building Energy Model coupled with an Urban Canopy Parameterization for urban climate simulations—part II. Validation with one dimension off-line simulations". *Theoretical and Applied Climatology* 99:3–4, pp. 345–356. ISSN: 1434-4483. DOI: 10.1007/s00704-009-0143-8.

Scherer, D., F. Ament, S. Emeis, U. Fehrenbach, B. Leitl, K. Scherber, C. Schneider, and U. Vogt (2019). "Three-Dimensional Observation of Atmospheric Processes in Cities". *Meteorologische Zeitschrift*, pp. 121–138. ISSN: , DOI: 10.1127/metz/2019/0911.

Schmitt, T., J. Kuhn, L. Pilz, R. Maiwald, M. May, R. Kleinschek, P. Edinger, S. Schmitt, F. Hase, D. W. T. Griffith, and A. Butz (2023). "Long open-path measurements of CO2 and CH4 with an 125HR FTS in an urban environment." In: Copernicus GmbH. DOI: 10.5194/egusphere-egu23-4403.

Schmitt, T. D., L. Pilz, R. Maiwald, M. May, B. A. Löw, R. Kleinschek, J. B. Wietzel, J. Kuhn, S. Schmitt, M. Schmidt, S. N. Vardag, F. Hase, D. W. T. Griffith, and A. Butz (2024). "A combined dataset of path-averaged and in-situ measurements of greenhouse gases to inform on the sensitivities to localized source patterns and transport effects in the urban atmosphere." In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-11113.

Schuh, A. E., A. R. Jacobson, S. Basu, B. Weir, D. Baker, K. Bowman, F. Chevallier, S. Crowell, K. J. Davis, F. Deng, S. Denning, L. Feng, D. Jones, J. Liu, and P. I. Palmer (2019). "Quantifying the Impact of Atmospheric Transport Uncertainty on CO2 Surface

Flux Estimates". *Global Biogeochemical Cycles* 33:4, pp. 484–500. ISSN: 1944-9224. DOI: 10.1029/2018gb006086.

Schuldt, K. N., T. Aalto, Arlyn Andrews, F. Apadula, S. Arnold, B. Baier, L. Bäni, P. Bergamaschi, T. Biermann, S. C. Biraud, R. Blot, W. A. Brand, Francescopiero Calzolari, G. Chen, Huilin Chen, A. Colomb, R. Commane, L. Condori, F. Conen, S. Conil, C. Couret, P. Cristofanelli, E. Cuevas, B. Daube, K. J. Davis, M. Delmotte, R. Dickerson, J. P. DiGangi, G. Diskin, M. Elsasser, L. Emmenegger, M. L. Fischer, G. Forster, M. Fuente-Lastra, T. Gehrlein, C. Gerbig, J. Hatakka, M. Heimann, M. Heliasz, D. Heltai, C. Hermans, O. Hermansen, A. Hoheisel, J. Holst, T. Di Iorio, D. A. Jaffe, A. Jordan, A. Karion, V. Kazan, P. Keronen, T. Kneuer, P. Kolari, K. Kominkova, E. Kort, E. Kozlova, P. B. Krummel, D. Kubistin, N. Kumps, R. L. Langenfelds, A. Lanza, O. Laurent, T. Laurila, J. Lavric, J. Lee, I. Lehner, K. Lehtinen, R. Leppert, A. Leskinen, M. Leuenberger, J. Levula, M. Lindauer, A. Lindroth, Mikaell Ottosson Löfvenius, Z. M. Loh, M. Lopez, C. R. Lunder, I. Mammarella, G. Manca, M. V. Marek, P. Marklund, M. Y. Martin, M. De Mazière, K. McKain, F. Meinhardt, Jean-Marc Metzger, N. L. Miles, C. E. Miller, J. B. Miller, M. Mölder, H. Moossen, J. Müller-Williams, C. L. Myhre, P. Nédélec, F. Obersteiner, O. Peltola, G. Petron, S. Piacentino, J. M. Pichon, P. Pickers, C. Plass-Dülmer, S. M. Platt, M. Ramonet, Xinrong Ren, S. J. Richardson, Louis-Jeremy Rigouleau, P. P. Rivas, M. Rothe, Yves-Alain Roulet, A. G. Di Sarra, D. Scharfe, B. Scheeren, M. Schmidt, M. Schumacher, T. Seifert, M. K. Sha, P. Shepson, C. D. Sloop, P. D. Smith, D. Steger, M. Steinbacher, B. Stephens, C. Sweeney, R. Taipale, H. Timas, P. Trisolino, J. Turnbull, B. Viner, G. Vitkova, A. Watson, S. De Wekker, D. Weyrauch, S. C. Wofsy, J. Worsey, C. Yver-Kwok, S. Zaehle, and A. Zahn (2025). *Multi-laboratory compilation of atmospheric carbon monoxide data for the period 1989-2023; obspack_co_1_GLOBALVIEWplus_v5.0_2025-01-22*. DOI: 10.25925/20241201.

Seibert, P. (2000). "Review and intercomparison of operational methods for the determination of the mixing height". *Atmospheric Environment* 34:7, pp. 1001–1027. ISSN: 1352-2310. DOI: 10.1016/s1352-2310(99)00349-0.

Seidel, D. J., C. O. Ao, and K. Li (2010). "Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis". *Journal of Geophysical Research: Atmospheres* 115:D16. ISSN: 0148-0227. DOI: 10.1029/2009jd013680.

Seidel, D. J., Y. Zhang, A. Beljaars, J.-C. Golaz, A. R. Jacobson, and B. Medeiros (2012). "Climatology of the planetary boundary layer over the continental United States and Europe". *Journal of Geophysical Research: Atmospheres* 117:D17. ISSN: 0148-0227. DOI: 10.1029/2012jd018143.

Shi, X. and Y. Wang (2022). "Impacts of Cumulus Convection and Turbulence Parameterizations on the Convection-Permitting Simulation of Typhoon Precipitation". *Monthly Weather Review* 150:11, pp. 2977–2997. ISSN: 1520-0493. DOI: 10.1175/mwr-d-22-0057.1.

Shin, H. H. and S.-Y. Hong (2015). "Representation of the Subgrid-Scale Turbulent Transport in Convective Boundary Layers at Gray-Zone Resolutions". *Monthly Weather Review* 143:1, pp. 250–271. ISSN: 1520-0493. DOI: 10.1175/mwr-d-14-00116.1.

Shusterman, A. A., V. E. Teige, A. J. Turner, C. Newman, J. Kim, and R. C. Cohen (2016). "The BErkeley Atmospheric CO$<*sub*$ >2$</*sub*$ > Observation Network: initial evaluation". *Atmospheric Chemistry and Physics* 16:21, pp. 13449–13463. ISSN: 1680-7324. DOI: 10.5194/acp-16-13449-2016.

Skamarock, W. C. (2004). "Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra". *Monthly Weather Review* 132:12, pp. 3019–3032. ISSN: 1520-0493. DOI: 10.1175/mwr2830.1.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang (2019). "A Description of the Advanced Research WRF Model Version 4". en. DOI: 10.5065/1DFH-6P97.

Solbakken, K., Y. Birkelund, and E. M. Samuelsen (2021). "Evaluation of surface wind using WRF in complex terrain: Atmospheric input data and grid spacing". *Environmental Modelling & Software* 145, p. 105182. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2021.105182.

Stagakis, S., C. Feigenwinter, R. Vogt, D. Brunner, and M. Kalberer (2023). "A high-resolution monitoring approach of urban CO2 fluxes. Part 2 – surface flux optimisation using eddy covariance observations". *Science of The Total Environment* 903, p. 166035. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.166035.

Stewart, I. D. and T. R. Oke (2012). "Local Climate Zones for Urban Temperature Studies". *Bulletin of the American Meteorological Society* 93:12, pp. 1879–1900. DOI: 10.1175/BAMS-D-11-00019.1.

Stith, J. L., D. Baumgardner, J. Haggerty, R. M. Hardesty, W.-C. Lee, D. Lenschow, P. Pilewskie, P. L. Smith, M. Steiner, and H. Vömel (2018). "100 Years of Progress in Atmospheric Observing Systems". *Meteorological Monographs* 59, pp. 2.1–2.55. DOI: 10.1175/AMSMONOGRAPHS-D-18-0006.1.

Stohl, A., P. Seibert, G. Wotawa, D. Arnold, J. F. Burkhart, S. Eckhardt, C. Tapia, A. Vargas, and T. J. Yasunari (2012). "Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition". *Atmospheric Chemistry and Physics* 12:5, pp. 2313–2343. ISSN: 1680-7324. DOI: 10.5194/acp-12-2313-2012.

Super, I., S. N. C. Dellaert, A. J. H. Visschedijk, and H. A. C. D. van der Gon (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, pp. 1795–1816. DOI: 10.5194/acp-20-1795-2020.

Tanimoto, H., T. Matsunaga, Y. Someya, T. Fujinawa, H. Ohyama, I. Morino, H. Yashiro, T. Sugita, S. Inomata, A. Müller, T. Saeki, Y. Yoshida, Y. Niwa, M. Saito, H. Noda, Y. Yamashita, K. Ikeda, N. Saigusa, T. Machida, M. M. Frey, H. Lim, P. Srivastava, Y. Jin, A. Shimizu, T. Nishizawa, Y. Kanaya, T. Sekiya, P. Patra, M. Takigawa, J. Bisht, Y. Kasai, and T. O. Sato (2025). "The greenhouse gas observation mission with Global Observing SATellite for Greenhouse gases and Water cycle (GOSAT-GW): objectives, conceptual framework and scientific contributions". *Progress in Earth and Planetary Science* 12:1. ISSN: 2197-4284. DOI: 10.1186/s40645-025-00684-9.

Taylor, K. E. (2001). "Summarizing multiple aspects of model performance in a single diagram". *Journal of Geophysical Research: Atmospheres* 106:D7, pp. 7183–7192. ISSN: 0148-0227. DOI: 10.1029/2000jd900719.

Taylor, T. E., A. Eldering, A. Merrelli, M. Kiel, P. Somkuti, C. Cheng, R. Rosenberg, B. Fisher, D. Crisp, R. Basilio, M. Bennett, D. Cervantes, A. Chang, L. Dang, C. Frankenberg, V. R. Haemmerle, G. R. Keller, T. Kurosu, J. L. Laughner, R. Lee, Y. Marchetti, R. R. Nelson, C. W. O'Dell, G. Osterman, R. Pavlick, C. Roehl, R. Schneider, G. Spiers, C. To, C. Wells, P. O. Wennberg, A. Yelamanchili, and S. Yu (2020). "OCO-3 early mission operations and initial (vEarly) XCO2 and SIF retrievals". *Remote Sensing of Environment* 251, p. 112032. ISSN: 0034-4257. DOI: 10.1016/j.rse.2020.112032.

Thoning, K. W., P. P. Tans, and W. D. Komhyr (1989). "Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985". *Journal of Geophysical Research: Atmospheres* 94:D6, pp. 8549–8565. ISSN: 0148-0227. DOI: 10.1029/jd094id06p08549.

Tulapurkara, E. G. (2005). "Hundred years of the boundary layer — Some aspects". *Sadhana* 30:4, pp. 499–512. ISSN: 0973-7677. DOI: 10.1007/BF02703275.

UCAR (2025). *Weather Research & Forecasting Model (WRF)*. URL: https://web.archive.org/web/20250929034254/https://www.mmm.ucar.edu/models/wrf (visited on 2025-09-29).

Ulpiani, G., N. Vetters, C. Thiel, N. Della Valle, Ş. Kılkış, V. Palermo, and G. Melica (2025). "Key recommendations for cities committed to climate neutrality". *npj Urban Sustainability* 5:1. ISSN: 2661-8001. DOI: 10.1038/s42949-025-00268-y.

United Nations Framework Convention on Climate Change (UNFCCC) (2015). "Paris Agreement". In: *UNFCCC, COP Report No. 21, Addenum, at 21, U.N. Doc. FCC-C/CP/2015/10/Add.1 (Jan. 29, 2016).*

Vogelezang, D. H. P. and A. A. M. Holtslag (1996). "Evaluation and model impacts of alternative boundary-layer height formulations". *Boundary-Layer Meteorology* 81:3, pp. 245–269. ISSN: 1573-1472. DOI: 10.1007/BF02430331.

Wang, J., S. Callewaert, M. Zhou, F. Desmet, S. Conil, M. Ramonet, P. Wang, and M. De Mazière (2025). "WRF-Chem simulations of CO2 over Western Europe assessed by ground-based measurements". DOI: 10.5194/egusphere-2025-4537.

Welch, P. (1967). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". *IEEE Transactions on Audio and Electroacoustics* 15:2, pp. 70–73. ISSN: 0018-9278. DOI: 10.1109/tau.1967.1161901.

Xu, C., T. A. Kohler, T. M. Lenton, J.-C. Svenning, and M. Scheffer (2020). "Future of the human climate niche". *Proceedings of the National Academy of Sciences* 117:21, pp. 11350–11355. ISSN: 1091-6490. DOI: 10.1073/pnas.1910114117.

Yang, Z.-L., G.-Y. Niu, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, L. Longuevergne, K. Manning, D. Niyogi, M. Tewari, and Y. Xia (2011). "The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins". *Journal of Geophysical Research: Atmospheres* 116, D12. ISSN: 2156-2202. DOI: 10.1029/2010JD015140.

Zhao, N., A. Ma, Y. Zhong, J. Zhao, and L. Cao (2019). "Self-Training Classification Framework with Spatial-Contextual Information for Local Climate Zones". *Remote Sensing* 11:23, p. 2828. ISSN: 2072-4292. DOI: 10.3390/rs11232828.

Zscheischler, J. and E. M. Fischer (2020). "The record-breaking compound hot and dry 2018 growing season in Germany". *Weather and Climate Extremes* 29, p. 100270. ISSN: 2212-0947. DOI: 10.1016/j.wace.2020.100270.

Scientific Contributions

Papers

Pilz, L., C. Lüken-Winkels, M. Gałkowski, D. Ho, C. Gerbig, F. Chen, and S. N. Vardag (2026). "Evaluation of high-resolution WRF simulation in urban areas — Effect of different physics schemes on simulation performance in the Rhine-Main-Neckar area". *Atmospheric Research* 328, p. 108435. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2025.108435.

Nayna Schwerdtle, P., E. Cavan, **L. Pilz**, S. D. Oggioni, A. Crosta, V. Kaleyeva, P. H. Karim, F. Szarvas, T. Naryniecki, and M. Jungmann (2023). "Interlinkages between Climate Change Impacts, Public Attitudes, and Climate Action—Exploring Trends before and after the Paris Agreement in the EU". *Sustainability* 15:9, p. 7542. ISSN: 2071-1050. DOI: 10.3390/su15097542.

Datasets

Pilz, L., H. Denier van der Gon, T. Glauch, J. Marshall, and S. N. Vardag (2025a). *MACRO-2018 - High-Resolution Simulation of CO and CO2 concentrations over German Metropolitan Areas for 2018 using WRF-Chem*. Ed. by **L. Pilz**. World Data Center for Climate (WDCC) at DKRZ. DOI: 10.26050/WDCC/MACRO-2018.

Open Source Software

Pilz, L., A. Banihirwe, J. Thielen, K. Paul, D. Cherian, K. FitzGerald, W. Rudisill, C. Lüken-Winkels, and J. Kent (2025b). `xWRF`. Version 0.0.5. URL: https://github.com/xarray-contrib/xwrf (visited on 2025-11-03).

Pilz, L. (2025b). `wps_xr`. URL: https://github.com/lpilz/wps_xr (visited on 2025-11-03).

Pilz, L. (2025a). `portraitpy`. URL: https://github.com/ATMO-IUP-UHEI/portraitpy (visited on 2025-11-03).

Bugfixes

Dury, A., A. Zonato, B. Kirk, N. Sobhani, C. He, T.-S. Lin, C. Li, J. Lewandoswki, J. Ruppert, J. Olson, A. Ukhov, L. FIta, **L. Pilz**, A. Martilli, M. Landreau, M. Chen, R. Conrick, R. Gilliam, J. Pleim, S. Osipov, S.-Y. Hong, T. Spero, and T. Mansell (2025). *WRF Version v4.7.0*. NCAR. URL: https://github.com/wrf-model/WRF/releases/tag/v4.7.0.

Howes, C., R. Easter, R. Arthur, J. Mirocha, M. Göbel, S. Walters, R. Cabell, S. Roh, H.-J. Songa, A. Valmassoi, T. Mansell, M. Honnorat, **L. Pilz**, T. W. Juliano, J.-H. Kim, P. A. Jimenez, J. Lee, T. Brummet, and T. Raupach (2021). *WRF Version v4.2.2*. NCAR. URL: https://github.com/wrf-model/WRF/releases/tag/v4.2.2.

Conferences

Lüken-Winkels, C., **L. Pilz**, S. Cello, and S. N. Vardag (2025). "Designing CO2 sensor networks for German cities: Insights from synthetic studies in Berlin and Munich ". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu25-10382.

Pilz, L., C. Lüken-Winkels, M. Gałkowski, D. Ho, F. Chen, and S. N. Vardag (2024). "High-resolution meteorological CO2 enhancements of German metropolitan areas using WRF". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-8915.

Lüken-Winkels, C., **L. Pilz**, M. Cassiani, I. Pisso, and S. N. Vardag (2024). "High-resolution inversion of Berlin city emissions – A synthetic study using FLEXPART-WRF for network optimization within ITMS". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-9996.

Schmitt, T. D., **L. Pilz**, R. Maiwald, M. May, B. A. Löw, R. Kleinschek, J. B. Wietzel, J. Kuhn, S. Schmitt, M. Schmidt, S. N. Vardag, F. Hase, D. W. T. Griffith, and A. Butz (2024). "A combined dataset of path-averaged and in-situ measurements of greenhouse gases to inform on the sensitivities to localized source patterns and transport effects in the urban atmosphere." In: Copernicus GmbH. DOI: 10.5194/egusphere-egu24-11113.

Pilz, L., M. Galkowski, J. Fallmann, F. Chen, A. Butz, and S. N. Vardag (2023). "Optimizing High-Resolution Simulations with the Weather Research and Forecasting (WRF) Model for the German Rhine-Neckar Metropolitan Region". In: Wiley. DOI: 10.22541/essoar.167591054.45628540/v1.

Schmitt, T., J. Kuhn, **L. Pilz**, R. Maiwald, M. May, R. Kleinschek, P. Edinger, S. Schmitt, F. Hase, D. W. T. Griffith, and A. Butz (2023). "Long open-path measurements of CO2 and CH4 with an 125HR FTS in an urban environment." In: Copernicus GmbH. DOI: 10.5194/egusphere-egu23-4403.

Knapp, M., B. Hemmer, R. Kleinschek, M. Sindram, T. Schmitt, **L. Pilz**, B. Burger, and A. Butz (2022). "Towards Carbon Dioxide emission estimation with a stationary hyperspectral camera". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu22-3924.

Pilz, L. N., S. N. Vardag, J. Fallmann, and A. Butz (2022). "Hochaufgelöste Simulation von meteorologischen Feldern in urbanen Räumen mit dem Weather Research and Forecasting (WRF) Modell". In: Copernicus GmbH. DOI: 10.5194/dach2022-69.

Pilz, L., S. Vardag, R. Kleinschek, S. Hammer, and A. Butz (2021). "Towards an integrated study of urban CO2 emissions". In: Copernicus GmbH. DOI: 10.5194/egusphere-egu21-2724.

Acknowledgements

This thesis would not have been possible without the overwhelming support of many people. I count myself very fortunate to be part of a community which encourages learning, supports initiative, and rewards curiosity. My gratitude goes out to:

Dr. Sanam Vardag, for her guidance, support, and dutiful supervision of this thesis; for sharing her experience, knowledge, and skills which I profited from immensely; for always pushing forwards and affording space for personal development at the same time, and for her dedication to climate action and interdisciplinarity even in the face of adversity.

Prof. Dr. André Butz for his mentoring and advice; for building a constructive scientific atmosphere and a welcoming community; for his scientific rigour and skepticism, and for being a great representative of Environmental Physics as a discipline.

Christopher Lüken-Winkels for sharing the journey through ITMS with me; for being a great sounding board for my good and not so good ideas; for his optimism during difficult times and his shared joy in easier ones; for his mentorship in Mario Kart and Bayesian inversions, and for being an awesome office mate.

My other office mates Dr. Tobias Schmitt, Dr. Benedikt Löw, Dr. Marvin Knapp, Moritz Sindram, and Leonie Scheidweiler for providing a curious and supporting atmosphere; for sharing their knowledge, and for entertaining some of my wilder ideas.

Dr. Tobias Schmitt for great discussions over tea; for sharing his knowledge and experience, and for always striving to improve.

Dr. Benedikt Löw for his all-round support; for getting me into making pasta, and for sharing his advice and encouragement with the final stretch of the thesis.

Thomas Plewa for great discussions during long walks and for being a fountain of knowledge about a lot of things including theoretical physics.

All the rest of the ATMO and IUP PhDs for a great atmosphere; for sharing the PhD-journey in good and not so good, and for the nice evenings.

The whole ATMO group including Ralph Kleinschek and Angelika Gassama for the up-lifting atmosphere; for constructive conversations, and for entertaining my Doppelkopf-addiction.

Dr. Benedikt Löw, Christopher Lüken-Winkels, Robert Maiwald, Valeria Politino, Karolin Voss for proofreading my thesis and the constructive input.

Hugo Denier van der Gon, Dr. Ingrid Super, Dr. Theo Glauch, and Prof. Dr. Julia Marshall for providing me with their respective emissions inventories and the support in using them.

Michał Gałkowski and David Ho for sharing their knowledge and experience with WRF; for great times during conferences and workshops, and for their invite to Jena, which I will definitely take them up on.

Everyone I met at ITMS and who supported me and my research or just lent an open ear. I will miss the great social atmosphere and scientific exchanges around the General Assemblies.

Dr. Joachim Fallmann for sharing his experience with WRF and for establishing the connection to Prof. Dr. Fei Chen at NCAR.

Prof. Dr. Fei Chen for hosting me at NCAR and his support with my research.

Dr. Cenlin He and Dr. Jimy Dudhia for advice and help with WRF during my stay.

All the friends I made during my time in Boulder, CO including but not limited to our foreign researcher crew Dr. Jakob Schlör, Dr. Anastasiia Chyhareva, and Dr. Alexander Lojko and my climbing partners, especially Tim Holt and Dr. Ling Wang.

Dr. Jöran Landschoff for being a great friend and a political partner in crime; for commiseration in trying times, and for always being engaged.

All my friends including but not limited to Dr. Alice Hesse, Christin Pross, Clara Löw, Lukas Wolff and especially my Heidelberg climbing crew Jan Kleinmann and Maike Strauch who all supported me emotionally during difficult times and were a much needed source of stability.

Valeria Politino for her all-round support during the final phase of my PhD; for her inexhaustible energy, and for being a source of great love and affection.

My closer and extended family for their love and support along the way.

This thesis also would not have been possible without institutional support. Here, my thanks go out to:

The DKRZ for the computing resources granted by the WLA under project IDs bb1170 and bm1400 and the support around computing, data storage, and publication, without which this project would not have been possible at all.

The German-American Fulbright Commission for granting me a visiting researcher scholarship in the Doktorand:innenprogramm, which enabled my research stay at NCAR in Boulder, CO.

The Heidelberg Center for the Environment, especially Dr. Max Jungmann and Dr. Tatjana Peskan-Berghöfer for providing financial support and great interdisciplinary exchange through the HCE College, and for organizing the 4EU+ project "Collegio Futuro" which taught me hands-on international and interdisciplinary research skills.

The Heidelberg Graduate School for Physics, especially Gesine Heinzelmann for providing financial support for the last months of the thesis and for helping me navigate the bureaucracy throughout the PhD.

This work is part of the joint project ITMS, funded by the German Ministry of Research, Technology, and Space (BMFTR) under reference number 01LK2102D. I thank the ECMWF for the great ERA5 and CAMS data and want to disclose that my results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

Finally, I want to apologize to everyone whom I forgot to mention and sincerely thank every person who contributed in any way to my PhD experience.

AI Disclosure

During the preparation of this work, large language models (LLMs) were used in a limited manner. Services such as Github Copilot and Claude AI were used to support writing code for the investigation of data. The original draft of the manuscript for this thesis was fully written by the author without any AI assistance. Tools like ChatGPT were used sparingly during the editing stage to suggest alternative phrasing with the aim of improving clarity and readability. These tools were only used on a small number of individual paragraphs and sentences.

All code used for generating tables and figures was carefully reviewed, tested, and validated to ensure correctness. All alternative formulations generated using LLMs were critically reviewed and adjusted. No AI tools were used to generate, modify, or adjust original scientific content presented in this thesis. After using these tools and reviewing and adjusting their output, the author takes full responsibility for the accuracy and originality of the content of this thesis.