TUM School of Computation, Information and Technology

Understanding urban GHG emissions: a high-resolution mesoscale GHG modeling infrastructure for cities based on WRF

Xinxu Zhao

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften

genehmigten Dissertation.

Vorsitz:		Prof. Dr.-Ing. habil. Dr. h.c. Alexander W. Koch
Prüfende	1.	Prof. Dr.-Ing. Jia Chen
der Dissertation	2.	Prof. Christoph Knote, Ph.D.

Die Dissertation wurde am 14.04.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 16.08.2022 angenommen.

Abstract

In order to meet ambitious carbon neutrality goals at national and city levels, several atmospheric networks have been established to monitor greenhouse gas (GHG) concentrations over cities, and atmospheric models are used to simulate the transport of tracer gases and help interpret these measurements. Here an urban modelling infrastructure based on the Weather Research and Forecasting model (WRF) was developed coupled with GHG modules to simulate atmospheric transport and predict column-averaged abundances of GHGs. This infrastructure includes the initialization of tagged tracers, model settings, model inputs, etc., which were all optimized. The software framework has been packaged into so-called containers for future re-productivity of the simulations. Two evaluation cases were studied to assess the performance of this WRF-based infrastructure in the vicinity of Berlin and Munich, respectively. Simulations were performed with high spatial resolutions of 1 km for Berlin and up to 400 m for Munich. By combining the differential column method (DCM) with footprints generated by a particle transport model (i.e., Stochastic Time-Inverted Lagrangian Transport Model (STILT)), column measurements from permanent sensor networks and short-term measurement campaigns can be better interpreted. Thus, a deeper understanding of the sources and sinks of GHGs and their transports over cities is enabled. Evidence of missing or underestimated sources in the emission inventory was found when this method was used in conjunction with knowledge of local emission sources, highlighting the potential to improve inventories.

Contents

1. Introduction

1.1. Background and motivation

Human activities have resulted in an increase of the global average temperature relative to pre-industrial levels of approximately 1.1 °C, a number which is rising with a rate of 0.2 °C per decade and expected to reach around 1.5 °C between 2030 and 2052 [1], [2]. Such a temperature rise are leading to unpredictable and severe extreme weather events (e.g., hurricanes, flash flood and droughts), which are significantly altering to the earth system and human lives. Obviously, the climate system is being changed, primarily as a result of human-induced greenhouse gas (GHG) emissions.

To combat the effects of climate change and reduce GHG emissions efficiently and effectively, the Paris Agreement was established in 2015, which includes 192 parties and provides a globally binding framework for carbon reduction and climate adaptation under financial and technical support. The key element of this agreement is to limit the long-term rise in the global average temperature to well below 2 °C and the increase of 1.5 °C above pre-industrial levels for reducing the risks and impacts relevant to climate change. To achieve the long-term temperature goal set by Paris Agreement, effective and efficient adaptations at national, regional and local levels are needed. It contributes to deep reductions in atmospheric emissions of GHGs in the coming decades [3].

The share of GHG emissions released from urban areas has continued to increase as a result of urbanization [4]–[7]. At present, 55 % of the global population resides in urban areas [8], a number that is projected to rise to 68 % by 2050 [9]. Meanwhile, urban areas cover less than 3 % of the land surface worldwide [10] but consume over 66 % of the world's energy [1], [11] and generate more than 70 % of anthropogenic GHG emissions [12]. Carbon dioxide (CO_2) emissions from energy use in cities are estimated to comprise over 75 % of the global energy-related CO_2, with a rise of 1.8 % per year projected un-

der business-as-usual scenarios between 2006 and 2030 [4]. Methane (CH_4) emissions from energy, waste, agriculture, and transportation in urban areas make up approximately 21 % of the global CH_4 emissions [12], [13]. As emission hotspots, urban areas therefore play a vital role in addressing the challenge of carbon mitigation. It is crucial to develop appropriate science-based methods for understanding and projecting the effects of GHG emissions on urban areas by estimating city carbon emissions, developing effective and coherent adaptation actions, formulating mitigation strategies, and monitoring their success.

As the continent with the highest population density, Europe plays a major role in future mitigation efforts. In recognition of this fact, the European Commission aims to make Europe climate neutral by 2050 [14]. Furthermore, member countries of the European Union (EU) also adopt individual strategies towards that goal. For example, the German government plans that the national GHG emissions will be reduced by more than 65 % compared to 1990 by 2030, achieving climate neutrality by 2045. Local-scale initiatives are also put in place. Munich, currently the third-largest city in Germany with over 1.5 million inhabitants, has set an even more ambitious goal, aiming to be climate neutral by 2030.

To confront the challenge of carbon mitigation in cities and reach the goals set by individual municipalities, a multitude of urban atmospheric measurement networks have been built worldwide to optimize urban emissions [15]. By using their measurements, GHG concentrations can be monitored in and around cities, and more accurate emission estimates can be derived by accompanying atmospheric transport modelling and statistical techniques [16], [17]. Based on these quantitative assessments, more reliable scientific guidance can be provided for policy-makers to plan local emission reductions effectively and monitor mitigation efforts.

1.2. State of the Art

Currently, two standard quantitative approaches are widely used for estimating GHG emission fluxes, namely the bottom-up approach and the top-down approach [18]–[20].

With the bottom-up approach, the total emission fluxes are estimated based on statistical activity data from individual sectors (i.e., a quantitative measure

of the activity that can emit GHGs, like power plants and traffic) and the corresponding emission factors [21]. This approach is widely used for generating global and national sector-by-sector emission inventories (e.g., the Emissions Database for Global Atmospheric Research (EDGAR)). The spatial and temporal distributions of emission estimates produced using this technique often hold large uncertainties. This is due to missing, simplified or incomplete knowledge and understanding of emission processes, or their considerable heterogeneity in space and time [22]–[24]. Moreover, these uncertainties are associated with the large variability in source-specific and country-specific emission factors and grow larger at sub-national scales when estimating the disaggregation of the national annual totals in space and time. The uncertainty on the national fossil-fuel CO_2 emission estimates can range from a few percent (e.g., 3 %–5 % for the US) to a maximum of over 50 % for countries with fewer resources for data collection and a poor statistics [25].

The top-down approach can not only provide estimated global fluxes but also be used to verify the consistency and assess the uncertainties of bottom-up emission inventories [22], [23], [26]. With this respect, emission fluxes are estimated using additional information provided by measurements of atmospheric concentrations. Prior emission estimates, usually estimated by the bottom-up approach, are transported in the atmosphere with a transport model, and the resultant concentrations are compared to the atmospheric composition measurements [27]. However, it is hard to quantify the statistical or even systematic errors attached to both atmospheric observations and prior knowledge about the distribution of emissions and sinks [28].

In any case, the emission estimates are then optimized using either a mass-balance approach [29] or other inverse techniques [30]. In terms of GHG emissions estimation for cities or an area of interest, the inversion method in the top-down approach has frequently been applied in modelling studies accompanied by urban measurement networks, e.g., in California [31], Paris [17], Boston [32], Berlin [24] and Indianapolis [30]. Inversion models still show considerable potential for improvement. The uncertainties here can be attributed to biases in background concentrations, the difficulty of modelling transport in complex urban environments, and limited knowledge on the characteristics and spatial distributions of emission sources, such as missing stationary point sources and inner-city traffic.

Column measurements are suggested to provide a promising route to improving the detection of CO_2 emitted from major source regions [33], possibly avoid-

ing extensive surface measurements near such regions. Such measurements, i.e., measurements of concentration averaged over a column of air, are performed to help to disentangle the effects of atmospheric mixing from the surface exchange [26] and decrease the biases associated with estimates of carbon sources and sinks in atmospheric inversions [34]. Compared to surface values, urban enhancements in columns are less sensitive to boundary-layer heights [26], [33], [35], and column observations have the potential to mitigate mixing height errors in an atmospheric inversion system [36]. Atmospheric GHG column measurements combined with inverse models are thus an attractive method for analyzing GHG emissions and can be used to analyze their spatial and temporal variability [18], [35], [37], [38].

The differential column method (DCM) was proposed [39] to focus the top-down approach on concentration differences caused by local and regional emission sources, in particular to quantify urban emissions. This method evaluates the differences between column measurements at different sites. Chen et al. (2016) [39] applied the DCM using the compact Fourier-transform spectrometers (FTSs) EM27/SUN (Bruker Optik, Germany) and demonstrated the capability of differential column measurements for determining urban and local emissions in combination with column models. Citywide GHG column measurement campaigns have been carried out, e.g., in Boston [40], Indianapolis [41], San Francisco [42], Berlin [43], Munich [44], and Hamburg [45]. However, only a few studies have combined differential column measurements with high-resolution models. Toja-Silva et al. (2017) [46] simulated the column data at upwind and downwind sites of a gas-fired power plant in Munich using the computational fluid dynamic (CFD) model and compared them with the column measurements. Viatte et al. (2017) [47] quantified CH_4 emissions from the largest dairies in the southern California region, using four EM27/SUNs in combination with the Weather Research and Forecasting model (WRF) in the large-eddy simulation mode. Vogel et al. (2019) [48] deployed five EM27/SUN spectrometers in the Paris metropolitan area and analyzed the data with the atmospheric transport model framework CHIMERE-CAMS. Jones et al. (2021) [30] carried out a field campaign in Indianapolis, Indiana, in which five EM27/SUNs were deployed throughout the city for five days.

To aid in reaching the goal of climate neutrality and track emissions in Munich, our group has established the first worldwide automatic urban sensor network (MUCCnet: Munich Urban Carbon Column Network, accessible via `http://atmosphere.ei.tum.de/`, [49], [50]) for continuous, long-term monitoring

of GHGs in and around Munich. These as well as other data need to be modelled so as to aid in their interpretation and better understand the processes driving the emission and uptake of GHGs around the city.

1.3. Purpose and structure of this work

The main purpose of this work is to develop a modelling infrastructure for urban areas to accompany measurement networks (e.g., MUCCnet) or used to interpret short-term measurement campaigns. This infrastructure is developed using WRF, which can reach to a high spatial resolution, e.g., of up to 400 m. It is the basis for a quantitative understanding of the processes responsible for the emission and consumption of CO_2, CH_4, and CO in cities of interest, i.e., Berlin and Munich in this work. The results provided by this infrastructure will help interpret the observations, to provide guidance for local emission reduction strategies and to further complement and improve general inventories in the future.

The discussion in this work starts with the fundamentals of the climate system, its structure and related research methods in Sect. 2. Furthermore, an overview of WRF is presented in Sect. 3, and the construction of this modelling infrastructure is described in Sect. 4. This section describes in particular the treatment of urban canopy in WRF, the initialization of meteorological and concentration fields, the modelled biogenic fluxes, the development, operation and compilation of WRF, and the containerization of this modelling infrastructure. Two evaluation cases are performed to assess the performance of this infrastructure by comparing the model output to measurements from local weather stations in Sect. 5 and the measurement campaigns presented in Sect. 6. In Sect. 7, the differential column method (DCM) is applied for both evaluation cases to further assess the model performance, in particular when eliminating the influence brought by global or regional background. Furthermore, this method is discussed to help track missing or underestimated emissions in and around a city of interest through combining footprints and local knowledge. Section 8 summarizes this work, draws conclusions, and discusses the prospects for further research.

Two case studies described in Sects. 5, 6, and 7 are published and most of the details can be found in Zhao et al., (2019, 2021, 2022) [51]–[53].

2. Theoretical background

Before providing a specific introduction to the modelling framework designed in this study, this section provides an overview of the basics of the climate system and its key components (Sect. 2.1), the atmosphere and its structure (Sect. 2.2), climate change and its causes (Sect. 2.3), and the methods used to monitor and address the impacts and challenges posed by climate change (Sect. 2.4). These fundamentals are the basics of understanding the state of the earth system, as well as its operation mechanisms, and further contribute to a in-depth study of atmospheric models described in the following sections.

2.1. Climate system

Literally, the term 'climate change' refers to a change in the climate system. To face and tackle the challenges and damage caused by climate change, it is significant to understand the components of the climate system, for example, how these components interact with each other inside the system, how the system is driven by external forces, and why the change becomes unpredictable. More importantly, with a progressively deeper understanding of the climate system, it is now possible for humans to mitigate the effects of climate change using scientific methods.

Before delving into the system, it is worth noting that the generic term 'climate' refers to long-term regional or even global weather conditions (e.g., global average temperature, seasonal rainfall patterns), while another term 'weather' denotes local atmospheric status over a short time-span, such as clouds, winds and snow. Climate system is a complicated system representing what is currently happening in reality, past and more generally, climate here can be understood as the state of the climate system as a whole.

The climate system is made up of five main components, namely, the atmosphere, the hydrosphere, the cryosphere, the biosphere, and the geosphere

(land surface) [54]. The following lists key features of each component:

The atmosphere is the key component in the climate system, containing gases, solid and liquid particles, and clouds. They interact with the incoming and outgoing solar radiation in a complex manner. Details are described in Sect. 2.2.

The hydrosphere refers to the area that comprises all liquid surface and subterranean water. The oceans, covering 72 % of the Earth's surface, can store and transport a certain amount of energy, as well as dissolve and store GHGs.

The cryosphere denotes the part of the Earth that consists of frozen water and contributes to the thermal inertia and the albedo of the climate system.

The biosphere is made up of the marine and terrestrial biospheres, and plays a key role in the carbon cycle by influencing the uptake and release of GHGs. Large amounts of carbon are transferred back and forth between the atmosphere and the biosphere. Marine and terrestrial vegetation can store a certain amount of carbon during photosynthesis and preserve it in the form of fossils.

The geosphere represents the solid area of the Earth's surface, through which energy absorbed from the Sun is absorbed into the soil and transferred back to the atmosphere as long-wave (infrared) radiation, further heating the atmosphere.

In brief, among these components, physical, chemical and biological interaction processes always take place. The main source of energy in the climate system is the Sun. Once the energy and the radiation released by the Sun reaches the planet (i.e., the incoming solar radiation), around 29.5 % of the energy gets reflected directly back to the outer space, 75 % of which is reflected directly by the cloud and scattered into space, with the rest being reflected by the ground. Another 47 % of the incoming solar radiation is absorbed by the ground and oceans. The last 23.5 % is absorbed by components in the atmosphere, e.g., water vapor, ozone, and clouds, some of which gets involved in the GHG effects and sticks longer before leaving the planet. To maintain a relatively stable average global temperature, the energy balance on Earth needs to be somehow balanced. To be specific, most of the absorbed energy by the land and oceans is then re-emitted and escapes to the space as outgoing long-wave radiation. The rest is exchanged with the atmosphere to drive natural

processes, including photosynthesis and evaporation. [7]

2.2. Atmosphere

As the key component in the climate system, the atmosphere plays a significant role in the general circulation, which is not an isolated system and always interacts with other components in the climate system (see Sect. 2.2). Owing to the complexity of the composition and structure in the atmosphere, the following lists their key characteristics:

A. Chemical compositions in the atmosphere

It is well known that the atmosphere consists of Dinitrogen (N_2, 78 %) and Oxygen (O_2, 21 %), Argon (Ar, 0.9 %), accounting for approximately 99.9 % of the total atmosphere by volume. The remaining volume is composed of Water Vapor (H_2O), Carbon Dioxide (CO_2), Neon (Ne), Methane (CH_4) and, etc., the majority of which are unreactive chemically.

Among these, the trace gases whose concentrations in the atmosphere can be altered by human activities and trap heat in the atmosphere, are collectively referred to as greenhouse gases (GHGs), including CO_2, CH_4, H_2O, Ozone (O_3), Nitrous Oxide (N_2O), Chlorofluorocarbons (CFCs), Hydro-fluorocarbons (HFCs) and, etc. GHGs play key roles in atmospheric energetics by absorbing outgoing radiations from the Earth and blocking them escaping to the space consequently.

B. Vertical construction of the atmosphere

The atmosphere consists of four distinct layers, defined by the vertical profile of temperature, named in order from the ground upwards: troposphere, stratosphere, mesosphere, and thermosphere. Here is a detailed description of their key features:

Troposphere: from the ground surface upward to 10 km; the closest layer to human life, including approximately 75 % of the atmospheric mass; around 99 % of water vapor confined in this layer; the majority of weather events occur and the general air circulation globally formed in this layer.

Stratosphere: from 6 to 20 km above the ground level to around 50 km; the increase of the temperature with the altitude due to the O_3 layer;

highly stratified and poorly mixed trace gases caused by the stable temperature configuration, whose vertical motion is largely negligible.

Mesosphere: from 50 to 85 km; the decrease of temperature with the altitude.

Thermosphere: from around 90 km to 500-600 km; the sharp climb of the temperature with altitude; the hottest layer due to the absorption of ultraviolet radiation from the Sun.

2.3. Climate change

Changes in the climate system are triggered in two ways, either driven by internal activities or/and forced by external factors. In the case of internal activities, changes in a single climate component and/or interactions between components can both lead to climate change. The external forcing includes changes in solar radiation, the slow drift of continents, volcanic routines, and etc.

With the influence of the internal interactions and the external forcing, the climate system has been changing throughout history and can usually be maintained in a status with an energy balance, the most intermediate effect of which is a slight rise in global average temperature. The last major global temperature increase on Earth occurred during the Ice Age cycle, rising by 4–5 °C, which took around 7,000 years. However, since the pre-industrial era of the 1970s, the rate of warming has been more than ten times larger than that at the end of the Ice Age cycle, which has been scientifically proven to be the result of human impact on the climate system [54]. This is also commonly referred to as 'global warming'. Global warming is induced by the rapid rise in anthropogenic emissions, leading to an observable increase in atmospheric concentrations of GHGs. In 2019, the atmospheric CO_2 concentration reached its highest level in the last two million years, alongside higher values of CH_4 and N_2O than at any time in the last 800,000 years [1].

Human-induced climate change has a variety of impacts on the state of many aspects of the climate system. It encompasses sea level rise and extreme climate events observed worldwide, such as more intensive and severe droughts, heatwaves, compound floods, and stronger hurricanes, which are significantly threatening human health, and imposing heavy social and economic costs.

These impacts are irreversible with time and will become more severe in the coming decades.

In recent decades, the entire global climate system has been warming rapidly, because of human interventions, which exacerbates the greenhouse effect – that is, the atmosphere is trapping more of the heat radiated from Earth into space. GHGs contribute to these greenhouse effects by blocking heat from escaping. A metric 'global warming potential' (GWP) is used to assess the effect of each gas on the warming of the atmosphere by quantifying their radiative efficiency in terms of CO_2 equivalent. In other words, how much mass of CO_2 can cause the same GHG effect as the target gas in the same time-span. By definition, CO_2 has a GWP of 1 regardless of the time period used (usually 100 years), regarded as the reference. The higher GWP is associated with larger infrared absorption and a long lifetime in the atmosphere.

Four main GHGs and their features are listed as follows:

Carbon dioxide (CO_2): As a primary GHG emitted by human activities, atmosphere CO_2 raises considerably since the pre-industrial era. Naturally, sinks and sources of CO_2 exist in the carbon cycle, while human activities alter this cycle by adding more CO_2 to the atmosphere and influencing the nature sinks from the atmospheric CO_2 (e.g., photosynthesis, CO_2 dissolution). The main sources of human-induced CO_2 emissions are fossil fuels for electricity generation, space heating, and transportation.

Since the turn of the century, Germany has been the largest CO_2 emitter in the EU, producing around 739 million tonnes of CO_2 in 2020, even more than the summed emissions of the following two large EU emitters – Italy and Poland. In Germany, the largest share of CO_2 emissions can be attributed to energy industries, accounting for around 30 % of the national GHG emissions in 2020. Through the implementation of relevant climate mitigation policies, almost 67 million tones of CO_2 (9.4 %) are reduced in 2020 compared to 2019, with the energy sector contributing the largest reduction by around 38 million tonnes of CO_2 (the data provided Umweltbundesamt).

Methane (CH_4): CH_4 is the second important GHG and can be emitted by both natural sources (e.g., wetland and termites) and human activities. Approximately 64 % of CH_4 emissions are caused by anthropogenic emissions, including landfills, waste management, livestock, biomass burning, and fossil fuel combustion. Even though the CH_4 emissions are overall

less than CO_2, it can absorb much more energy with a GWP of 28-36 over 100 years and of 84 over 20 years [55]. Due to the methane removal from the atmosphere, mostly by the chemical reaction with hydroxyl radicals, the lifetime of CH_4 is around 12 years.

CH_4 emissions accounted for 11 % of the total anthropogenic GHG emissions worldwide in 2017, 10 % in EU, and 6 % in Germany [1]. The emissions from the energy sector contribute around 20 % of human-induced CH_4 emissions in the EU (released in 2021 by Europe Commission [2]). The main sources of the energy sector include oil and fossil gas upstream exploration and production, gas transmissions, coal mines in operation, etc. The EU plans to reduce its CH_4 emission from the energy sector by 58 % by 2030 compared to 2020 (according to the 'Impact Assessment for the Climate Target Plan 2030').

Water Vapor (H_2O): As the most abundant nature GHG in this case, H_2O here represents the water in the gaseous form. The warming of the climate system alters the amount of water vapor in the atmosphere and indirectly contributes to the GHG effect. Specifically, the warmer the atmosphere, the more water evaporates from the land surface, which makes more heat being trapped on Earth, leading to more warming.

Nitrous oxide (N_2O): Compared to the other three GHGs, N_2O only takes a small share of the atmosphere but with a long lifetime of around 120 years, In addition, it has a GWP of approximately 300 times that of CO_2. Since the pre-industrial era, N_2O has raised by 16 %, one third of which is induced by human actions, including livestock manure, soil cultivation, biomass and fossil fuel combustion, and industries.

2.4. Research methods of the climate system

To monitor the current status of carbon emissions regionally or globally and face the challenges induced by climate change effectively, climate observations and associated modelling have been being developed rapidly in recent decades.

[1] Data here is reported by DVGW and can be accessible via https://www.dvgw.de/medien/dvgw/en/topics/DVGW-Brochure-Methane-emissions.pdf

[2] The press is assessed by https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_6684

It is essential to advance the understanding of the processes driving the climate system. Models and observations are interdependent and complementary; specifically, the evaluation of adaptation of models is tied to the accuracy of observations, while the model can help to interpret the observed values more comprehensively. The following content provides details of the observation tools and models applied in the climate system, respectively.

A Observations of the climate system:

To monitor essential meteorological variables and the amount of GHGs in the atmosphere, a number of observational networks have been established worldwide. These collected or retrieved measured values can be used to interpret the reality and provide the basis for other applications, e.g., environmental impact assessments and weather forecasting, and help to validate climate models. The instruments used in the climate system can generally be classified into two categories, namely in-situ measurements and remote sensing observations.

For in-situ observations, they can capture the signal of the target variables in the vicinity of the measurement point. Although they cannot be applied to large areas and capture vertical distributions, they are useful to calibrate and validate satellite measurements and algorithms, as well as to assimilate reanalysis models, particularly during the required time spans. A variety of the in-situ instruments are being developed and used to collect data, including airborne measurements, surface wind anemometers, weather balloons, etc.

In contrast, remote sensing instruments allow monitoring of a test object from a certain distance and can capture signals passively or actively. The passive one refers to measuring signals emitted or scattered by objects naturally, such as infrared devices, while the active one is to take the measurements by recording the properties of emitted acoustic or electromagnetic energy that is reflected from an object or the land surface and further back to the sensor, e.g., wind LiDARs.

Thanks to the development of space measurement techniques in recent decades, weather satellites have been launched since 1960, and the first satellite mission designed to measure CO_2 was on board in 1996. They measure the short- and long-wave radiation reflected by the Earth and the atmosphere, and the different wavelengths of the radiation captured by the satellites are further retrieved to reveal the composition of the at-

mosphere. Such satellite measurements hold a high spatial coverage and are particularly useful over areas without ground measurements stations available, while their retrieval uncertainties and systematic signal interference are difficult to attenuate.

B Models of the climate system:

Due to the considerable complexity of the climate system, feedback among all these climatic components, and external forcing, the complex issues associated with climate change cannot be fully explained, let alone further addressed, if only observed data is used to understand the climate system. Mathematical and physical models are therefore being developed to explore the complex climate system and better understand the general circulation of the atmosphere. Over time, different discrete components of the climate system have been studied and coupled to build this complex climate system, including the atmosphere, land surface, carbon cycle, atmospheric chemistry, etc. In other words, even though atmospheric science has a considerable history, dating back to the 1870s, its main research focus has been on forecasting weather conditions. Research related to human air pollution and GHGs began in the 19^{th} century, and these issues have become increasing recognized in recent decades (especially since the 1970s), with their growing impact on human and social life.

There are many different types of climate models, including global and regional circulation models (GCMs and RCMs) defined based on the domain size of interest, as well as complex and coupled climate models designed for specific climate studies, like fully couple GCMs, i.e., Earth System Models [56]. Similar to the definition of 'climate' mentioned in Sect. 2.1, the climate models focus on how the average weather conditions change or will change. On the contrary, weather models are developed to reproduce (or/and predict) day-by-day changes in weather over a specific area at a particular point in time or in the near future.

The model-based studies use mathematical representations to describe physical (such as cloud, radiative and boundary-layer processes), chemical, dynamical, gas (aerosol), transport, and surface processes in the atmosphere, which are numerically resolved in both space and time [57]. To explore the climate system by using models to interpret what has been observed, atmospheric models need to be aligned with physical reality

by adding more complex components, especially radiation schemes that are the main external forcing in the climate system. The entire Earth in the model is separated into individual three-dimensional grid cells and the modelled results in each of them are passed on to its neighbors to simulate the exchange of matters and energy over time.

The main input here is the incoming and outgoing solar radiation, and human-induced emissions (of gases and aerosols) from the ground surface [58]. Energy from incident radiation flows through the climate system, mainly in the atmosphere and hydrosphere, in the form of internal heat, potential energy, latent energy, and kinetic energy. Some of this can be stored and sequestered in the biosphere, geosphere, and cryosphere of the climate system, and eventually radiated back into space in the form of infrared radiation [59].

To understand, monitor, and ultimately predict human impacts on climate, it is necessary to study it systematically and comprehensively. Depending on the domain sizes of interest, atmospheric models can be categorized as global and regional models. Obviously, global models cover the entire planet and can produce a comprehensive picture of global meteorological phenomena, but usually at a coarse spatial resolution of a few tens of kilometers and require a great deal of computational power. On the contrary, regional models use the data from global reanalysis models to initialize their lateral boundary conditions, and produce much finer modelled meteorological fields for local studies, ranging from hundreds of meters to hundreds of kilometers. In addition, with the increased requirements of regional models, the more complex geographical background on the Earth's surface (water, ground, snow, and rocks) are taken into account by dividing the model planet into different latitude bands, rather than treating the planet as one big average.

In this work, my interest focuses on modelling the concentration of GHGs and further understanding the emissions in and around a city accompanied by the observed values provided by a measurement campaign or network. This modelling infrastructure is built based on a mesoscale model [3] over urban areas of interest, which is described in detail in the following sections.

[3] 'Mesoscale model' means its spatial resolutions range from 2 km to 2000 km.

3. Fundamentals of WRF

As mentioned in Sect. 2, the model-based methods can help understand how the climate system works by reproducing processes and feedback among its internal components. Theoretically, these models are built by numerically solving the mathematical equations of the climate system, which are defined according to physical, dynamical, chemical, and biological principles [60]. Furthermore, external observations and other coarse reanalysis models are used as inputs of the boundary conditions and external forcing. As a result, they can provide simulated fields well discrete in space and time. In this section, the basic theory of the model-based methods will be presented.

3.1. Theory

3.1.1. Euler and Lagrangian frames

When modeling state variables (e.g., wind, concentration, pressure) in the climate system using mathematical equations, two frames are applied to describe motions of air parcels in coordinates and directly influence the mathematical descriptions of models: Euler frame and Lagrangian frame [61].

Here it is worth clarifying the concept of 'air parcel' beforehand. This item stands for a group of molecules of a trace gas or an aerosol with well-defined characteristics, including mass, pressure, humidity, temperature, etc. Its movement can be traced by observed winds [62] and described in four dimensions: longitude, latitude, altitude, and time.

A. **Euler frame**: When using the Euler frame, the air particles in the model are traced as a continuum in a fixed three-dimensional domain (longitude, latitude, and altitude), whose movement follows the governing equations (i.e., conservation of mass, momentum, and energy) [63]. The objective of this frame is to study the distribution of state variables. This frame

is widely integrated into the dynamical core of several models, including Integrated Forecasting System (IFS), and WRF [64], [65].

B. **Lagrangian frame**: In contrast, the models based on Lagrangian frames are used to track the origins of air parcels to a fixed coordinate system by simulating their trajectories driven by known meteorological conditions. The modelled air parcel here is pertained to be a moving element, as a discrete phase with its own properties [63], [66]. And the governing equations applied in the models are described as a function of time and formulated based on its own pressure and velocity. This kind of Lagrangian models requires external wind data as inputs and is widely used to reproduce or predict the overall particle dispersion patterns.

Suppose an air parcel is modelled in the Lagrangian frame and driven from south to north by initialized winds. In this case, the time rate of change of a state variable S along a trajectory (i.e., the total derivative: dS/dt) can be defined as the sum of the time rate of change of S at the starting point in the south (i.e., the local derivative: $\partial S/\partial t$) and the time rate of change of S due to the transport that results from a south-north scalar velocity (v) and leads to a spatial gradient of S [57]:

$$\frac{dS}{dt} = \frac{\partial S}{\partial t}\frac{dt}{dt} + \frac{\partial S}{\partial y}\frac{dx}{dt} = \frac{\partial S}{\partial t} + v\frac{\partial S}{\partial y} \tag{3.1}$$

Here v represents the horizontal wind component towards north. When dS/dt is zero, it means that the state variable S does not change with time when travelling from A to B, while non-zero indicates that external processes (e.g., chemical, physical, or/and dynamical processes) impact on the air parcel.

Correspondingly, when the time rate of change of a state variable S is described in three dimensions of the Euler frame, it can be expressed as follows [57], [66],

$$\frac{dS}{dt} = \frac{\partial S}{\partial t} + u\frac{\partial S}{\partial x} + v\frac{\partial S}{\partial y} + w\frac{\partial S}{\partial z} \tag{3.2}$$

In this study, the modelling infrastructure is built using WRF based on the Euler frame. The modelled fields are reproduced following a set of continuity equations, thermodynamic energy equations, etc. The changes in the concentrations of gases are solved by the species continuity equation [1]. Moreover, the

[1] Species continuity equations express the atmospheric transport of trace gases using partial differential equations (i.e., time- and space-dependent processes) [57]

finer meteorological fields provided by WRF are converted and used as inputs to the Stochastic Time-Inverted Lagrangian Transport model (STILT). This can further help to localize the origins of unknown sources and be used in inversion models to investigate emission estimations over the area of interest.

3.1.2. Pressure structure

The reconstruction of the vertical pressure structure plays a key role in the reproduction of the atmospheric state in the model. In theory, an air parcel is normally balanced against three forces: a downward force from the air pressure [2] above it on the top of the air parcel F_{top}, an upward force from the air pressure below F_{bot}, and a downward force from its own weight G_{air}. If the air parcel is kept in a steady state (i.e., moving at a constant speed or rest), the net vertical force on this air parcel should be zero, and the mathematical expression can be written as follows:

$$F_{top} - F_{bot} + G_{air} = 0 \rightarrow P_{top} - P_{bot} = -\frac{G_{air}}{A} \rightarrow \int_{bot}^{top} P dl = -\rho g \Delta l \tag{3.3}$$

in which Δl is the height of the air parcel, A is the area of the air parcel, g stands for the acceleration of gravity, ρ represents the density of the air parcel, and P_{bot} and P_{top} mean the air pressure below and above the air parcel, respectively. Equation 3.3 can be further expressed as,

$$\frac{\partial P}{\partial l} = -\rho g \tag{3.4}$$

where $\partial P / \partial l$ is the change in pressure between the top and the bottom of the air parcel with respects to height. This equation is also named as 'hydrostatic [3] equation' and the pressure that solely depends on the weight of air above a certain altitude, is called the 'hydrostatic air pressure'. This equation stands for a balance between the vertical pressure gradient force and gravity without net vertical accelerations and gives a relationship between the change in pressure with its height and gravity. This relationship can be held well in the motions of large-scale atmosphere systems, i.e., with low resolutions. This hydrostatic

[2] Air pressure: the weight of air above a horizontal plant divided by the area of the plane.
[3] The word 'hydrostatic' here means fluids at rest.

balance is destroyed when the perturbation pressure is present as a driving force.

When building the pressure vertical structure in the model, two types of solvers are commonly used: the hydrostatic solver and the non-hydrostatic solver.

A. Hydrostatic solver: As mentioned above, the hydrostatic equation implies that the upward pressure gradient force is balanced by the downward gravitational force. In models with low resolutions, the hydrostatic solver can build the closest approximation to the realistic atmosphere and well resolve the motions. When applying this solver, the momentum equation along the vertical direction is not considered [4]. This makes the model unable to capture vertical motion features. That is, this solver is not suitable for regional or high-resolution models, while works well and efficiently with low resolutions. As implemented by ECWMF [67], the hydrostatic solver has been used extensively in their weather prediction models over the last 30 years and it is well suited to models at a horizontal resolution of over 10 km. However, such an assumption is not reasonable for models with higher resolutions, in which case the non-hydrostatic solver should be applied.

B. Nonhydrostatic solver: As expressed by its name, the non-hydrostatic solver retains the full vertical momentum equation for use in high-resolution models, particularly when the resolution is less than 3 km. Furthermore, model time steps [5] are constrained by the model's vertical and horizontal resolutions. In the explicit numerical models or schemes of the models, their time step are required to be smaller than the horizontal resolutions divided by the speeds of the fastest motion within their domains [57]. Thus, the time step defined in the non-hydrostatic model with a higher spatial resolution is quite small, even less than 1 second in WRF. This makes the model more computationally expensive than that in the hydrostatic model with a relatively coarse resolution.

[4] When the horizontal resolution is over 3 km (low resolutions), the vertical acceleration is much smaller than the gravitational accelerations. In this case, the vertical acceleration term is simplified and often removed from the momentum equation [57].

[5] Model time steps are the finite sizes of the time period of interest for advancing the state variables and resolving equations.

3.2. WRF model

The Weather Research and Forecasting model, well known as WRF, is a meso-scale numerical weather prediction system widely used for supporting related atmospheric research and operational forecasting. It is a 'community model' developed with the support of the National Center for Atmospheric Research (NCAR) and the National Oceanic and Atmospheric Administration (NOAA). All the source codes are now officially released in the GitHub repository: `https://github.com/wrf-model/WRF`, which can be easily cloned. The first version was released publicly in Dec. 2000 and it has been developed and updated (Version 4.3 in Dec. 2021) due to the improvements to typical new physics schemes and climate diagnostics. Furthermore, the model cannot be driven by any graphical user interface (GUI) and can only be invoked from UNIX terminals and commands, mostly in the Fortran language. WRF is not only used to study atmospheric phenomena under idealized conditions, but is also capable of reproducing or predicting real-time atmospheric processes. With the development of this community modelling system, WRF has been further developed to adapt to other extended applications, such as convection forecasting, high-resolution hurricane simulations, chemical reactions and data assimilation using observational data.

WRF is built based on the fully compressible non-hydrostatic Euler equations [64] that have been explained in Sect. 3.1.2. Two dynamical solvers are applied in WRF, Advanced Research WRF (WRF-ARW) and the non-hydrostatic meso-scale WRF model (WRF-NMM). It deals with how the mode does time integration and resolves the governance equations numerically, including advection, radiations, pressure-gradients, Coriolis, buoyancy, filters, etc.

As a numerical model, it starts with a specific atmospheric state S_{t_0} (i.e., model initial and boundary conditions) and advances in time to look for a solution for n time steps Δt. This is based on the governing equations of the model, and physical and dynamic schemes employed. Then the change of the atmospheric stage at a certain time ΔS_{t_n} is forced by a group of dynamical and chemical atmospheric processes F. The atmospheric state at time t_n (S_{t_n}) can be obtained as follows:

$$\begin{aligned} S_{t_n} &= S_{t_0} + \Delta S_{t_n} \\ \Delta S_{t_n} &= n \times \Delta t \times F \end{aligned} \tag{3.5}$$

The term F mentioned in Eq. 3.5 varies with time and is impossible to directly go to a certain stage in a time step. The stage in the future (i.e., S_{t_n}) is obtained

after a few time steps, namely time steps for integration in the model. The entire run can be understood as: how the data is represented and the governance equations are solved on each grid cell.

3.3. WRF operation

The running process of WRF can mainly be divided into two parts, namely WRF prepossessing (WPS) and the WRF main run, in which WRF initialization, restart run and etc., are included. This section contains a detailed descriptions of these two parts.

3.3.1. Preprocessing (WPS)

In the model workflow, WPS is applied to prepare suitable meteorological fields for the WRF main run, primary for real-data simulations. In WPS, the required time-varying meteorological fields (e.g., temperature, pressure, winds) and geographical data can be obtained from reanalysed datasets and static global or local database. In this step, these database are horizontally interpolated as the initial and boundary conditions for the WRF main run. Three main programs are implemented in WPS: *geogrid.exe*, *ungrib.exe* and *metgrid.exe* (see Fig. 3.1).

'geogrid.exe': The *geogrid* program is used to define the geographic information of domains and their characteristics, including map projections, dimensions and horizontal resolutions. The map projection defined here is used to flatten the real curved earth surface onto a rectangle in a plane and is chosen according to the latitudinal range in which the domain is located. In this study, the models are developed for three cities, namely Munich and Berlin, where the Lamber Conformal Conic projection (LCC) was applied to fit the mid-latitude region. Each grid cell of the domains is initialized by the external static or time-invariant geographical fields, including land use and land cover categories, vegetation fractions, topography, soil types, etc. Specifically, these static land fields are read and interpolated horizontally when this program is called. As the focus of this study is on urban areas, urban topographic maps need to be generated and written into the data format, which is the same as the common global static geographical field used in WRF (see Sect. 4.1). In addition,

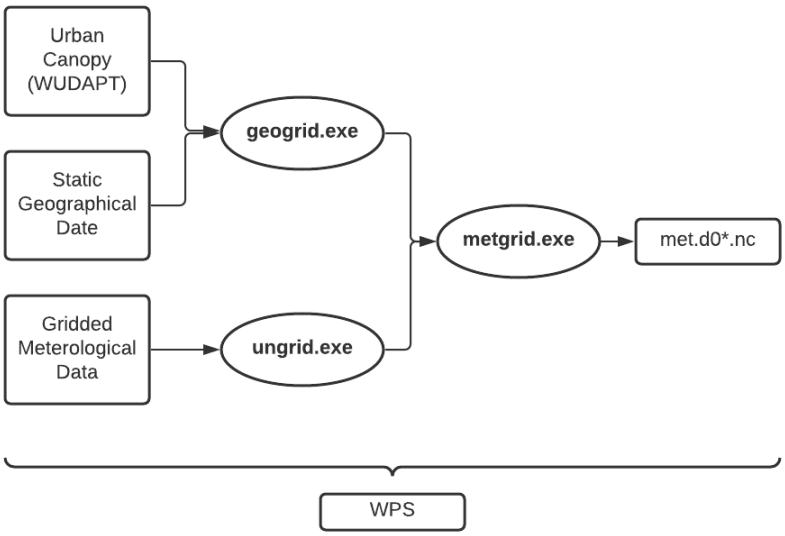


Figure 3.1.: The Workflow of WPS.

map scale factors and Coriolis parameters are calculated for each grid cell in this process.

'ungrib.exe': As its name 'ungrib' expresses, this program does 'undo' the format of *GRIB* held by the meteorological input and converts the dataset in this format to another format that can be more easily processed by the *metgrid* program. In this process, the gridded meteorological and land surface data is extracted and written in another intermediate format. Apart from it, some basic derivations are also included in this part, such as the calculation of relative humidity based on temperature, pressure and Water vapor mixing ratio. In addition, a Variable table called *'Vtable'* is required to tell the program which fields should be extracted from the *GRIB* input files and replaced when the meteorological data source is changed. In this study, four different *'Vtable'* files were used and toggled, when ECWMF reanalysis Version 5 (ERA5) is used for the initialization of the model (details in Sect. 4.2.1).

'metgrid.exe': When running this program, two groups of outputs from the *geogrid* and *ungrib* programs are combined and further processed on the model grid. The primary purpose of this program is to horizontally interpolate the meteorological data extracted by the *ungrib* program onto the model domain which is defined by using the *geogrid* program. An-

other purpose is to rotate the horizontal wind fields to let the u (westerly) and v (southerly) components with respect to the x and y directions in the grid cell. After processing this program, one output per domain per time period with the format of *NetCDF* (*.nc*) can be obtained. These final outputs contain the static fields from the *geogrid* program and the interpolated meteorological fields, are further passed off to the *real* program (Sect. 3.3.2).

When implementing horizontal interpolations of grid cells using the *metgrid* program, there are four ways used to define the horizontal grid structure of WRF (see Fig. 3.2) as follows:

(a) **The u staggered points**: The u-components of the horizontal wind field are interpolated to the u staggering, shown by the blue letters in Fig. 3.2;

(b) **The v staggered points**: The v-components of the horizontal wind field are interpolated to the v staggered points, depicted by the green letters;

(c) **The mass grid**: Other meteorological fields with scalar values are interpolated to the centers of the grid cells (also named as the mass grid), marked as the symbols θ with red;

(d) **The staggered grid**: The corners of the mass grids are defined as the staggered grids (see black dots). As can be seen in Fig. 3.2, one more point is shown in each direction beyond the dimensions of the mass grid. This kind of definition of grid cells along the vertical direction is of importance, since it is used to discrete the vertical coordinates following the pressure distribution (details in Sect. 3.1.2).

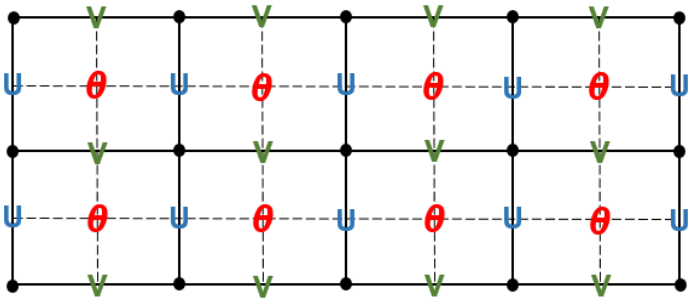


Figure 3.2.: Horizontal grid structure in the WRF domain.

3.3.2. Main run

'real.exe': As the pre-processor of the WRF main run, the *real* program takes the outputs from WPS and transforms them to files suitable for inputs to the WRF program. The main function of the *real* program is to handle the vertical interpolation. This program is able to read meteorological data with any of the traditional vertical coordinates, i.e., the data on terrain following the Isobaric, Sigma or Hybrid coordinates. The data from WPS is then processed in a strictly monotonic way on the pressure vertical coordinate, either increasing or decreasing. When running this program, a number of meteorological fields are initialized properly and a full set of physics in micro-physics, radiations, convection, boundary layer and surface conditions, are utilized within the model.

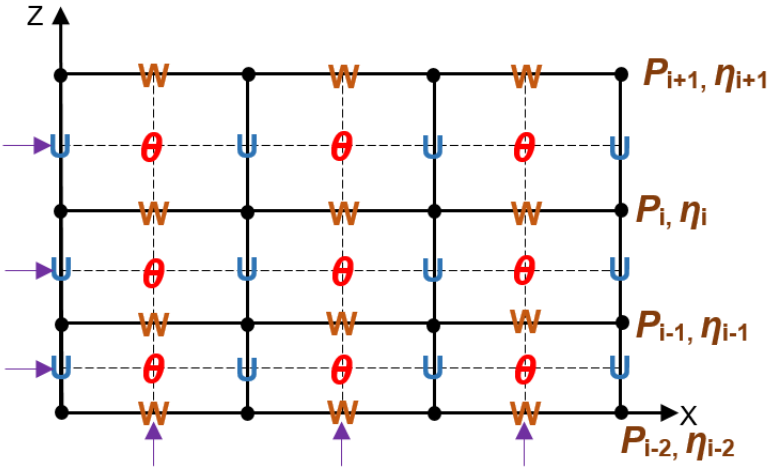


Figure 3.3.: Vertical grid structure in the WRF domain.

The outputs produced by this program are a number of files named *'wrfinput_d0*'* suffixed with the number of the domains, which include the initial state of each requested domain, and another file (called *'wrfbdy_d01'*), containing the lateral boundary conditions for the outermost domain. The initial condition file contains data for only a single time period (i.e., the start point of the run), while the lateral boundary file includes slices with the number of $n-1$ if the real program is processed n times.

In *'wrfbdy_d01'*, two groups of the variables are included. One is the initial conditions of one scalar variable at all time points along one horizontal direction with one of the following suffixes in its name: *'_BXS'*, *'_BXE'*,

'*_BYS*' and '*_BYE*'. Another one is the tendency on how one field is taken from the previous time point to the current one, for which additional letter *T* is included in its suffix of its name: '*_BTXS*', '*_BTXE*', '*_BTYS*' and '*_BTYE*' (see Fig. 3.4). In the name of these suffixes, *B* stands for boundary conditions, *X*/*Y* represents the longitude/latitude horizontal index and *S*/*E* means the start/end of the corresponding direction. The calculation on the tendency is mentioned in Sect. 4.2.3.

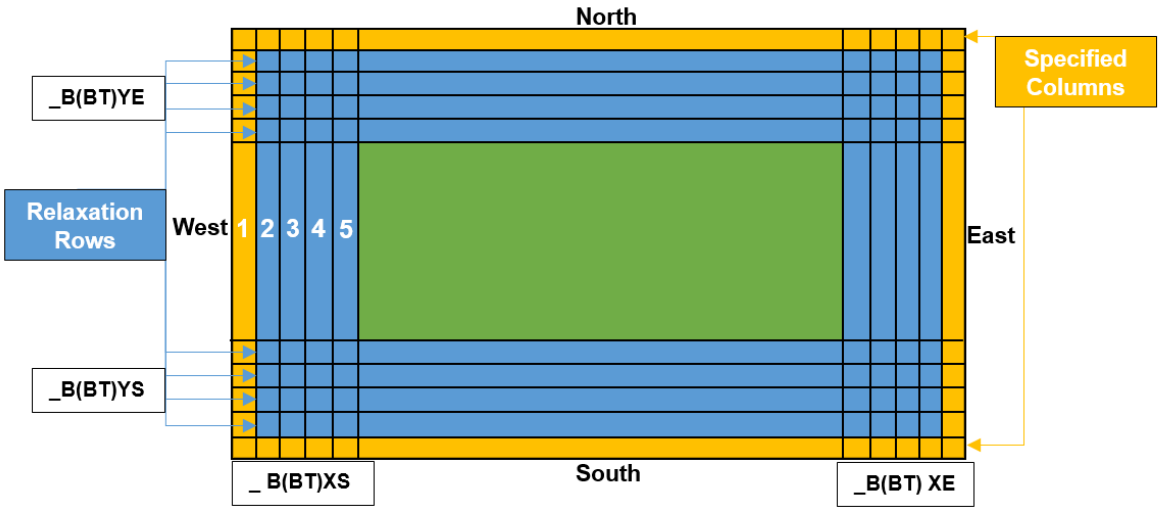


Figure 3.4.: Horizontal grid structure in the WRF domain.

'wrf.exe': The *wrf* program is applied for numerical time integration by taking the initial and boundary conditions provided by the *real* program. The output from the *wrf* program follows the hybrid vertical coordinate, which is a terrain following the vertical coordinate near the surface. In the setting of the model, the *eta* η levels are used to define the vertical coordinate. That is, the eta level where the eta value equals to zero stand for the top of the model, and the value of one represents the surface. The eta value at a specific layer η_i can be obtain as follows,

$$\begin{aligned} \eta_i &= \frac{P_i - P_{tp}}{\mu} \\ \mu &= P_{sfc} - P_{top} \end{aligned} \tag{3.6}$$

where P stands for the hydro-static pressure, and P_{sfc} and P_{tp} are the surface and top of the model domain, respectively. In this study, 50 hPa is always chosen as the pressure at the top layer of the model domain.

3.3.3. Data processing

The modelled fields gridded to mass points (i.e., θ in Figs. 3.2 & 3.3) normally consist of two separated components: a time-independent base state and a perturbation from the base state. The base state field is a function of the topography with several handful constants defined.

Geopotential height Geopotential is defined on full levels and can be computed using two variables from the WRF output files: the base state geopotential PHB [m^2/s^2] and the perturbation geopotential PH [m^2/s^2]. Then the geopotential height H_{gph} can be calculated using the acceleration of gravity g:

$$H_{gph} = \frac{PB + PHB}{g} \tag{3.7}$$

Pressure The hydrostatic pressure P in WRF with the unit of Pa is made up of two parts and can be obtained as follows,

$$P = P_p + P_b \tag{3.8}$$

Here P_p is the perturbation pressure and P_b represents the base state pressure, which can be read from the variables P [Pa] and PB [Pa] of the WRF output files.

Temperature The potential temperature is generated as a predicted variable in WRF, instead of temperature. The reason for it is that the density which cannot be measured directly, can be excluded from the governing equations. In the model, the temperature T_{tot} [K] is calculated using the potential temperature $theta$:

$$\begin{aligned} theta &= T + 300 \\ T_{tot} &= theta \times (\frac{P}{1000})^{\frac{2}{7}} \end{aligned} \tag{3.9}$$

where $theta$ is the equivalent potential temperature which can be taken from the variable *T* in the WRF output files, 1000 hPa is taken as a reference pressure.

4. WRF modelling infrastructure

One of the key purposes in this study is to generate a modelling infrastructure over an area of interest. This section includes all the significant components consisted of the modelling infrastructure, including geographical input, initial and boundary conditions for meteorological and concentration fields, etc. Further, details of the modelling structure and its operation are documented at the end of this section.

4.1. Urban canopy: WUDAPT to WRF

The geographical input of the modelling infrastructure is obtained from one of the widely used database provided by UCAR, i.e., the modified International Geosphere-Biosphere Programme global land cover classification based on the Moderate Resolution Imaging Spectroradiometer (*IGBP-Modified MODIS* [68]). This IGBP-Modified MODIS database has a resolution of 30 arc seconds (around 1 km) and can provide 20 land-cover categories that are mainly related to different types of vegetation (e.g., evergreen forest and mixed forest). However, the IGBP-Modified MODIS database, as well as other common geographical databases, only classify one land-use category defined for urban areas, as shown in Fig. 4.1(a).

To better capture the urban landscape features and improve the model performance over urban areas [69]–[71], extra morphological urban parameters for the innermost domain (urban areas of interest) are provided, which can be further called by the urban canopy multi-layer scheme available in WRF (i.e., *sf_urban_physics* = 2 in the *'namelist.input'* of WRF [69]). This is done by reclassifying the land-cover categories for urban areas, while keeping the other land-cover categories unmodified. The re-categorized land-cover types are derived from the European Local Climate Zones (LCZ) map [72], extracted for the high resolution domain.

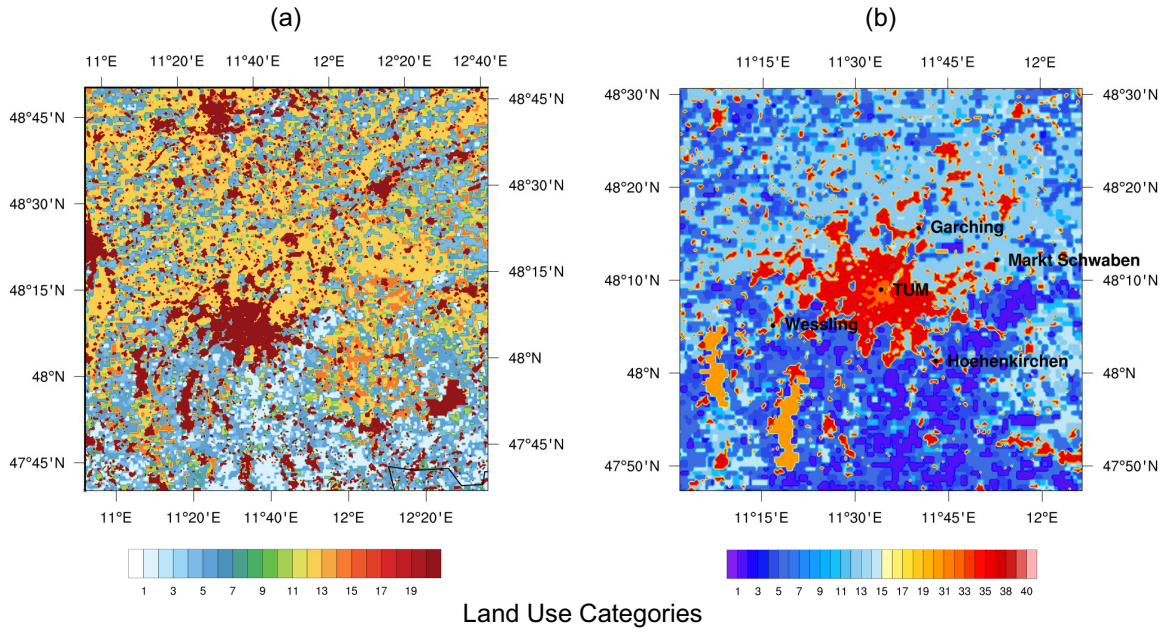


Figure 4.1.: The land use land cover index map (a) without adapting and (a) adapting with LCZ classes provided by the *'geo_em.d03.nc'* files in WRF.

With regard to the LCZ typologies, they have been developed as a baseline for the classification of urban ground cover and structures, spanning from hundreds of meters to several kilometers horizontally. A set of numerical values are used to represent these typologies with a number of surface landscape properties, such as average building heights and impervious surface fraction. In total, there are 17 urban and vegetation types distinguished based on urban canopy features and vegetation landscapes. For LCZ, the numerical values from '1' to '10' belong to building types and the letters from 'A' to 'G' represent rural land cover types.

Generally speaking, there are two methods used to generate LCZ maps that are further converted into geographical inputs of WPS.

The first method is to obtain the LCZ map for urban areas of interest through clipping it from the EU LCZ map using SAGA GIS and the detailed implementation of this method is shown in Fig. 4.5. This EU map is provided by Demuzere et al (2021) [73] and generated by the protocol supported by the World Urban Database and Access Portal Tools project (WUDAPT, as available via `http://www.wudapt.org/`, last access: 03 Jan. 2022). For the generated LCZ maps, the LCZ classifications are used to distinguish landscape types, which include ten land-cover types defined for urban areas (e.g., com-

pact high-rise and compact low-rise), and another seven land-cover types for vegetation. These classifications are illustrated in Fig. 1 of Demuzere, M. et al (2019)[72].

Another method is to generate the LCZ maps by using the online LCZ generator. This tool [74] is available to the public from 2020 via `https://lcz-generator.rub.de`. In short, it is used to classify urban areas of interest according to the LCZ classifications, through capturing the morphological urban features using available multi-spectral images from satellites, like Landsat 8. The LCZ typologies are identified by a random forest classifier [75] based on pre-defined training areas (TAs).

In this work, the first method based on the EU LCZ map [73] is applied for Munich, owing to the unavailability of the online generator in 2019. After clipping the Munich area of interest (i.e., the innermost domain of WRF) from the EU LCZ map, the morphological information is then extracted and transformed into the format used in WPS using SAGA GIS [69]. The second method is applied to generate the LCZ map for the Hamburg campaign carried out by my colleagues in 2021.

A detailed description of how to generate the map and convert it to the geographical input of WPS is shown in Fig. 4.5. Three main steps are described as follows,

a. **Generating the LCZ map for urban areas of interest**: The first step is to prepare the LCZ map with a format of *GeoTiff* for a city of interest, either through clipping it directly from the EU LCZ map [73] or generating a new one using the online LCZ generator [74]. The resulting product in this step is a gridded map, including three bands: 'LCZ typology', 'LCZ Filter' and 'Class Probability' [74].

   As mentioned above, the LCZ map for Munich is generated by clipping the EU LCZ map [73]. This *GeoTiff* file can be read and clipped by SAGA GIS [76], as shown in Fig. 4.5(A). It is to be noted that the range of a clipped map should be close to the area covered by the innermost domain of WRF (i.e., the target region of interest). Due to the universality of urban landscape features in the LCZ typologies, the TAs of one city can be used to generate the urban landscape of another one, which is the fundamental of how the EU LCZ map was produced [72], [74], [77]. Nevertheless, the local landscape features cannot be captured well if no TA in and around the area of interest is chosen, which can result in mismatches between

the classified LCZ typologies and the reality.

For the Hamburg campaign in 2021, the key prerequisite for using this online tool is the preparation of a *.kmz* file using Google Earth Pro, which includes the pre-defined TAs for all the required LCZ typologies. When selecting TAs, four key properties are considered here: pervious vegetation or soil fractions, building fractions, impervious fractions and mean building heights. Other properties should be considered as well, such as sky view factor, aspect ratio, surface albedo and height of roughness elements, while excluded in this case due to the absence of relevant information.

Based on the personal experience and suggestions provided by Demuzere et al, 2020 [74], five notes for such selections are listed here:

a. For each LCZ topology, 5 to 15 TAs should be selected over the area of interest, since the selection of fewer TAs may not contain sufficient landscape features of one topology, whereas if more than 20 TAs for one topology are selected, the resulting map will not be greatly affected but computational resources are wasted;

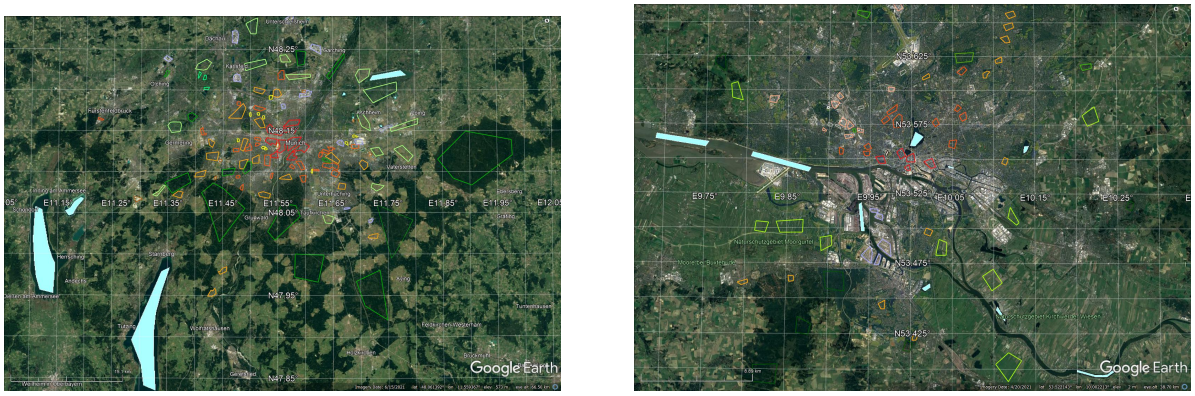
b. Even though only urban LCZ typologies are finally used in WPS, the rest 7 vegetation typologies should be carefully classified as they affect the delineation of the urban footprints;

c. Due to the limited pixel size of Landsat images (100-200 m on a side) that can be used in the LCZ classifications, the optimal size of one TA should be less than 1 km^2 and keep the narrowest point less than 200 m;

d. The polygons used to form a TA should be simple, avoiding complex shapes to maintain the homogeneity of the spectral information in the Landsat image and to maximize the number of pixels that can fill in the TA;

e. A buffer of 100 m or more should be maintained between selected TAs, which helps to avoid confusing information with mixed pixel signals to the classifier;

Figure 4.2 shows the selected TAs used to generate the LCZ maps for Munich (left) and Hamburg (right). Next, these LCZ maps can be either applied to the random forest classification tool in SAGA GIS [69] or up-



(a) Munich **(b)** Hamburg

Figure 4.2.: Training areas for (a) Munich and (b) Hamburg. Different colors represent the TAs for different urban typologies. Map provided by © Google Earth.

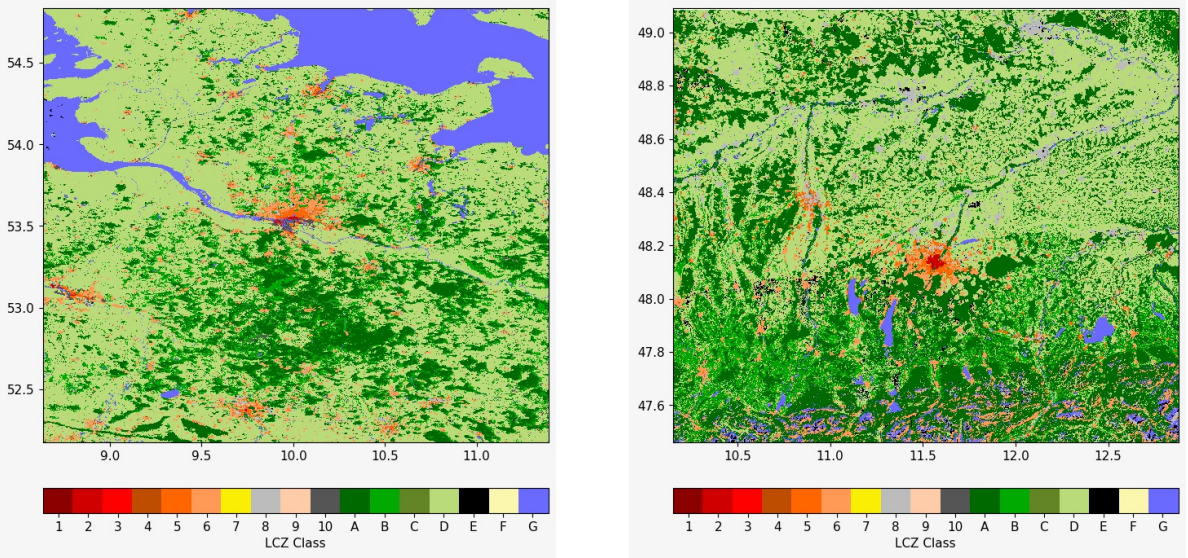
loaded to the online LCZ generator as an input [74]. In this work, the generated LCZ maps are presented in Fig. 4.3.

b. **Converting into geographical inputs of WPS**: Due to the restriction of the applicable WRF version coupled with GHG modules, this modelling infrastructure has to be built based on the version of WRF lower than Version 4. In this case, the urban canopy scheme related to the LCZ classifications is still not adapted to the source code of WRF and the tool WUDAPT to WRF (W2W) programmed with Fortran is applied. From the version of 4.3 onward, the LCZ typologies have been adapted to the source code of WRF and a well-packaged python tool can be used for such conversions.

   This conversion starts with the extraction the LCZ typologies from the LCZ map into a *text* file that can be read by the W2W Fortran tool. As shown in Fig. 4.5(B), the typologies are extracted from the *GeoTiff* file and stored to a *text* file by using the *Grid Values to Points* tool in SAGA GIS. The generated *text* file includes the geographical location of each grid cell centroids and its corresponding LCZ typology.

   Next is to obtain the geographical file by running the *rd_wrf_binary.exe* program. It is an input of the geogrid program in WPS. This program is used to read the *text* file and interpolates the LCZ typologies on the regular grids as required by WPS.

   Before compiling and running this program, the parameters in the Fortran program file *'rd_wr_binary.f90'* need to be adjusted according to the



(a) Munich **(b)** Hamburg

Figure 4.3.: The LRZ maps for (a) Munich and (b) Hamburg using the online LRZ Generator © 2020 Ruhr-University Bochum [74].

```
====================
name=LANDUSEF
        priority = 2
        dest_type = categorical
        z_dim_name = land_cat
        interp_option = default:nearest_neighbour
        abs_path = default:/gpfs/scratch/pr48ze/ga62kuy2/WRF-
urban/WPS/GEOG/landuse_Munich/
====================
```

Figure 4.4.: The required information in the *'GEOGRID.TBL'* of WPS for the urban topography in Munich.

basic information of the *text* file (e.g., number of lines) and the model configuration of WRF, including the projections and domain coordinates. As a result, the output here will be saved with a *binary* format that can be used in WPS [69]. This *binary* file is named as *'landuse_urban'* in default and then saved to the geographical input folder of WPS and renamed as *'00001–[number of X points].00001–[number of Y points]'*.

In this work, this *'landuse_urban'* file was saved in the *'GEOG'* folder of WPS under the name of *'Landuse_Munich'*, and the text file is further renamed as *'00001–00195.00001–00195'* for Munich.

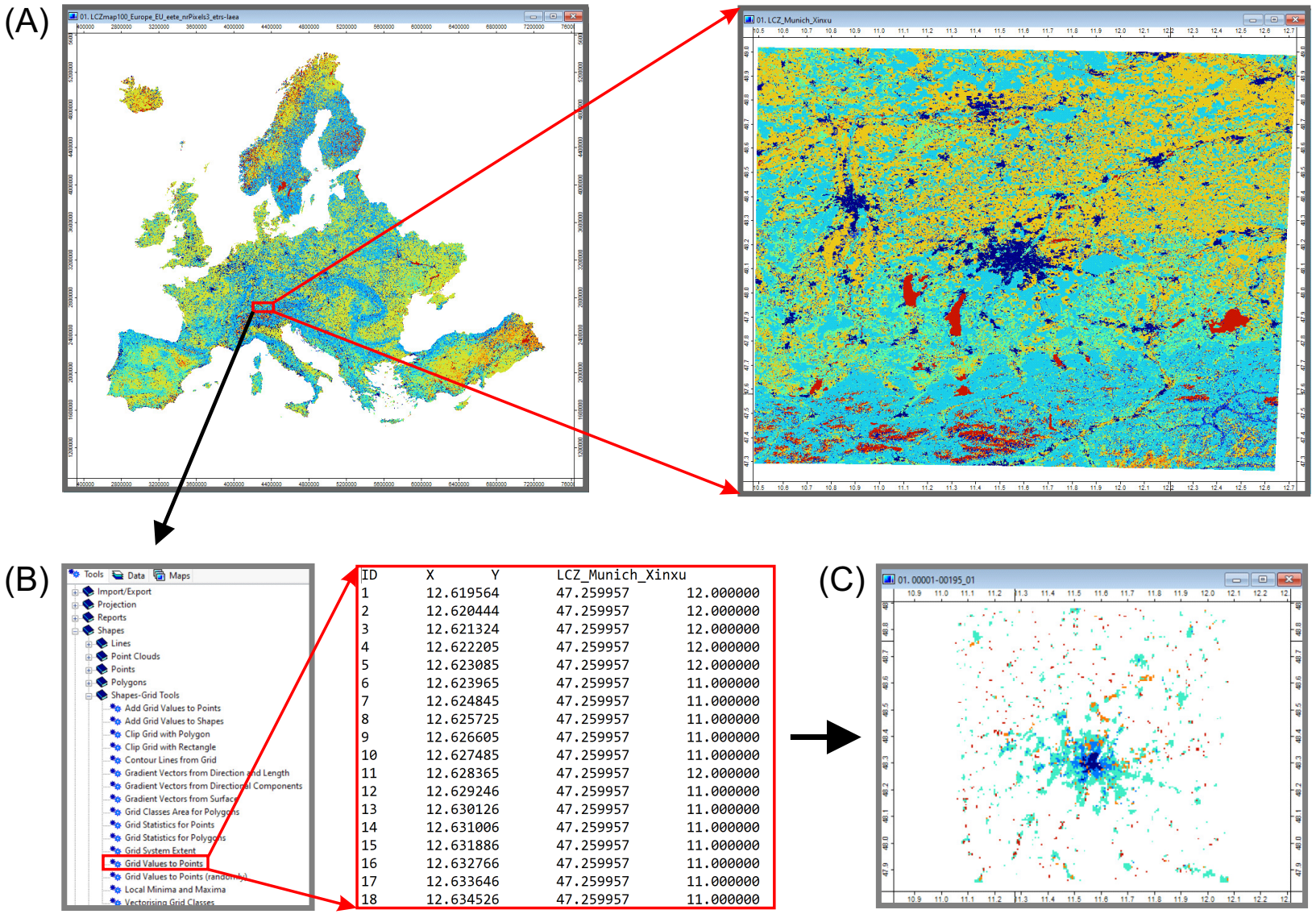


Figure 4.5.: The steps on generating the geographical input used for the input of the *geogrid* program in WPS.

Table 4.1.: List of geographical instructions in the *'index'* file (for Munich).

```
type = categorical
category_min = 31
category_max = 40
missing_value = 0.
projection = regular_ll
dx = 0.005549
dy = 0.005549
known_lon = 11.028
known_lat = 47.824
wordsize = 1
tile_x = 195
tile_y = 195
tile_z = 1
units = "category"
description = "10-category UCZ"
mminlu = "MODIFIED_IGBP_MODIS_NOAH"
```

Apart from this renamed *binary* file, an *index* file including the attributes to describe the land use instructions, should be created and saved in the same *'Landuse_Munich'* folder, shown in Table. 4.1. Moreover, before running the *geogrid* program, the *'GEOGRID.TBL'* in the geogrid folder of WPS should be modified and detailed information of this created geographical data is included, as shown in Fig. 4.4.

c. **The incorporation of LCZ classes to WRF Version 3.9.1.1**: Before WRF is updated to Version 4.3, the LCZ schemes are not incorporated to the source code of WRF and only 3 urban classes are defined with the numbers of '31', '32', '33' in the MODIS classifications: 'Low Density Residential', 'High Density Residential' and 'Industrial of Commercial', respectively.

   To insert the LCZ scheme in the source code of WRF, a new set of numerical values used to stand for the LCZ typologies, are defined by replacing the old typologies with the numerical values of 31-33 with 31-40. Table 4.2 lists 8 module scripts in the *'phys'* folder and 4 tables in the *'run'* folder that should be modified. In these files, the numerical values relevant to urban typologies should be redefined to merge the information and calculations related to the LCZ classifications. Moreover, the parameters

Table 4.2.: List of the edited module scripts in the *'phys'* and *'run'* folders.

1) module_physics_int.F
2) module_sf_clm.F
3) module_sf_noahdrv.F
4) module_sf_noahlsm.F
5) module_sf_noahmpdrv.F
6) module_sf_noahmplsm.F
7) module_sf_urban.F
8) module_surface_driver.F
9) LANDUSE.TBL
10) MPTABLE.TBL
11) URBTABLE.TBL
12) VEGTABLE.TBL

listed in these 4 tables of the *'run'* folder and related to urban topographies, should consider to be adapted to specific cases (cities), including roof levels (building heights) for each urban type, roof (or building) widths, road widths, standard deviation of roof heights, etc.

Regarding this incorporation of the LCZ classifications to WRF Version 3.9.1.1, I would like to express my sincere gratitude to Dr. Martilli Alberto, who provided me with the technical support on the edits of the WRF source code.

After adapting the module scripts using the LCZ typologies and adding the urban topography in the *'GEOG'* folder of WPS, WRF should be recompiled and the created programs can be used to generate new geographical files named as *'geo_em.d0∗'*. As seen in Fig. 4.1(b), the land use indexes in the new generated geographical files range from '1' to '40' and the LCZ typologies are represented by the numerical values from '31' to '40'.

Table 4.3 lists the coverage percentages of all the LCZ typologies in and around Munich (Fig. 4.3(a)). Over the area of interest around Munich, approximately 12 % of the entire domain is covered by building types, the largest of which are open low-rise buildings.

After implementing all the steps described above, the generated LCZ topography is converted to the geographical input file with the binary format which WPS can be read and used as an input. In addition, this *binary* input file can

Table 4.3.: List of LCZ types and the coverage percentages of each type for Munich.

LCZ type	Number	Category	Percent
LCZ 1	31	Compact high-rise	0
LCZ 2	32	Compact mid-rise	0
LCZ 3	33	Compact low-rise	0.35 %
LCZ 4	34	Open high-rise	0
LCZ 5	35	Open mid-rise	1.53 %
LCZ 6	36	Open low-rise	8.03 %
LCZ 7	37	Lightweight low-rise	0
LCZ 8	38	Large low-rise	0.99 %
LCZ 9	39	Sparsely built	0.94 %
LCZ 10	40	Heavy industry	0
LCZ E	41	Rock and paved	0

be mapped using SAGA GIS, as shown in Fig. 4.5(C).

4.2. Initialization

After preparing the geographical inputs of WRF by running the *geogrid* program (see Sect. 3.3.1), external reanalysis databases provide meteorological inputs of WRF that are taken and interpolated horizontally by calling the *ungrib* program in WPS. Section 4.2.1 includes details of these external reanalysis databases.

The near-surface fluxes are filled into grid cells by saving the values into the auxiliaries, named *'wrfchem'i_d0*'* and suffixed with date times (i.e., YYYYMMDDHH) and the number of domains. The sources of these fluxes and the generation of the auxiliaries *'wrfchemi_d0*'* are described in Sect. 4.2.2.

Furthermore, the initialization of tracers, including background fields and near-surface emission fluxes from different processes, should be implemented before running the *real* program. Specifically, the background fields are initialized by global reanalysis database and fed into the files that contain initial and boundary conditions, i.e., *'wrfinput_d0*'* and *'wrfbdy_d01'* (see Sect. 4.2.3).

Table 4.4.: List of meteorological fields used for initializing WRF.

	Institution	Spatial Resolution	Spatial Coverage	Temporal Resolution	Valid Period	Vertical level
ERA5 (pressure level)	ECWMF	~31 km (2.5 °)	Global	Hourly	1979 - now	37
ERA5 (model level)	ECWMF	~31 km (2.5 °)	Global	Hourly	1979 - now	137
ERA5-Interim	ECWMF	~79 km (5 °)	Global	6-hourly	1979 - Aug., 2019	60
GFS-Final (FNL)	NCEP	~28 km (2.5 °)	Global	6-hourly	1999 - now	26
GDAS-Final (FNL)	NCEP	~4 km (0.25 °)	Global	6-hourly	2015 - now	26
NAM	NCEP	12 km	North America	6-hourly	2012 - now	39
HRRR	NOAA	3 km	North America	Hourly	2014 - now	50
NNRP	NCEP/ NCAR	60 km	Global	6-hourly	1948 - now	28

4.2.1. Meteorological fields

In short, the time-varying meteorological fields are extracted from the global or large-scale reanalysis databases, and aggregated horizontally to the grid cells of the designed domains. It is implemented through calling the *ungrib* and *metgrid* programs of WPS (Sect. 3.2).

These databases should be provided or converted with the format of *GRIB*. Table 4.4 lists a group of meteorological databases that can be used in WPS. In this work, the initial WRF test in Berlin was built by taking the Global Forecast System (GFS) model to initialize the meteorological fields, while the European Center for Medium Range Weather Forecasts (ECWMF) Reanalysis Version 5 (ERA5) was used for the upgraded WRF test for Munich. The following shows a more detailed description of these two databases used:

GFS: The meteorological fields used in Berlin are obtained from GFS at a horizontal resolution of 0.5 °, with 64 vertical layers and a temporal resolution of 3 hours [1]. GFS uses hydrostatic equations for the prediction of atmo-

[1]The NOAA's National Center for Environmental Information is available via `https://www.ncdc.noaa.gov/`, last access: 03 Jan 2022).

spheric conditions, and its output includes large amounts of atmospheric and land–soil variables, wind fields, temperature, precipitation, and soil moisture, etc. GFS is one of the meteorological databases widely used by WRF, and no additional steps are required to fill the database into the model, whereas this is not the case with ERA5.

ERA5: The meteorological initial and lateral boundary conditions for Munich are obtained from ERA5, the fifth-generation reanalysis. ECWMF produces ERA5 by combining and assimilating a variety of observations from more than 20 satellites, radiosondes, aircraft measurements, etc. [78]. This database provides hourly meteorological driving data to WRF, extracted at approximately 31 km horizontal resolution and with 137 vertical levels from ground level to 0.01 hPa [2]. Since the primary application of ERA5 is not specific to WRF, additional processes are required when running the *ungrib* and *metgrid* programs and the detailed processing is shown in the workflow of Sect. 4.8.

4.2.2. Anthropogenic emission fluxes

In WRF, surface emission fluxes initialized by external emission inventories, are emitted from both point and area sources, and further spread to surrounding areas driven by associated meteorological fields (mainly winds). Here the emission inventories refer to mapping of annual estimated gridded anthropogenic emissions of GHGs and air pollutants globally or over a certain region. With the bottom-up approach (Sect. 1.2), the total emission fluxes are estimated based on statistical activity data from individual sectors (e.g., power plants and traffic) and their emission factors.

The fluxes extracted from the inventories are mapped to grid cells in each domain and saved in hourly chemical emission auxiliaries named *'wrfchemi_d0*'* suffixed with the number of domains and the hourly full time point. Table 4.5 shows a list of available emission inventories, including their major features and available species. In this work, EDGAR Version 4.1 and TNO_GHGco_v1.1 are used for the initialization of WRF for Berlin and Munich, respectively.

[2] Copernicus Climate Change Service can be assessed via: `https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.`, last access: 03 Jan 2022.

Table 4.5.: List of available emission inventories

	Institute	Temperal Resolution	Temperal Coverage	Spatial Resolution	Spatial Coverage	Species	Number of Sectors
EDGAR	JRC	Annual	1970 - now	0.1 ° × 0.1 °	Global	CH_4, CO_2, N_2O, F-gases, NMVOC, etc.	12
IER	Uni of Stuttgart Germany	Annual	2000; 2008	10 km × 10 km	Europe	CH_4, CO_2, CO, NOx, $PM_2 \cdot 5$	12
TNO-MACCIII	TNO	Annual	2015	∼7 km × 7 km	Europe	CH_4, CO_2, CO, NOx, SO_2, NMVOC, NH_3, etc.	14
CAMS-REG	TNO	Annual	2000 - 2017	∼6 km × 6 km	Europe	CH_4, CO, CO_2, NMVOC, PM, etc.	14
TNO_GHGco	TNO	Annual	2005 - 2015	∼1 km × 1 km	Europe	CH_4, CO_2, CO, NOx, NMVOC	14
EMEP/EEA	EMEP/EEA	Annual	1990 - now	0.1 ° × 0.1 °	Europe	NOx, PM, SOx. etc.	14
EPA	EPA	Annual	1990 - now	0.1 ° × 0.1 °	America	CO_2, CH_4, NOx	14
REAS	National Institute for Environmental Studies, Japan	Monthly	1950 - 1955; 1980 - 2015; 2020	0.25 ° × 0.25 °	Asia	CO_2, CO, NOx, SO_2, NMVOC, PM, NH_3, OC	5
MEIC	Tsinghua Uni China	Monthly	1990 - 2017	0.25 ° × 0.25 °	China	CO_2, NH_3, NMVOC, SO_2, NOx, CO, PM, etc.	5
MIX	MICS-Asia	Monthly	2008, 2010	0.25 ° × 0.25 °	Asia	SO_2, NOx, CO, NMVOC, NH_3, PM, BC, OC, CO_2	5
CAPSS	National Institute of Enbironmental work of Korea	Monthly	2008-2010	0.25 ° × 0.25 °	Korea	CO_2, SO_2, NOx, CO, NMVOC, PM_{10}	12
Vulcan	North American Carbon Program	Annual	2010-2015	1 km × 1 km	America Alaska	CO_2	10

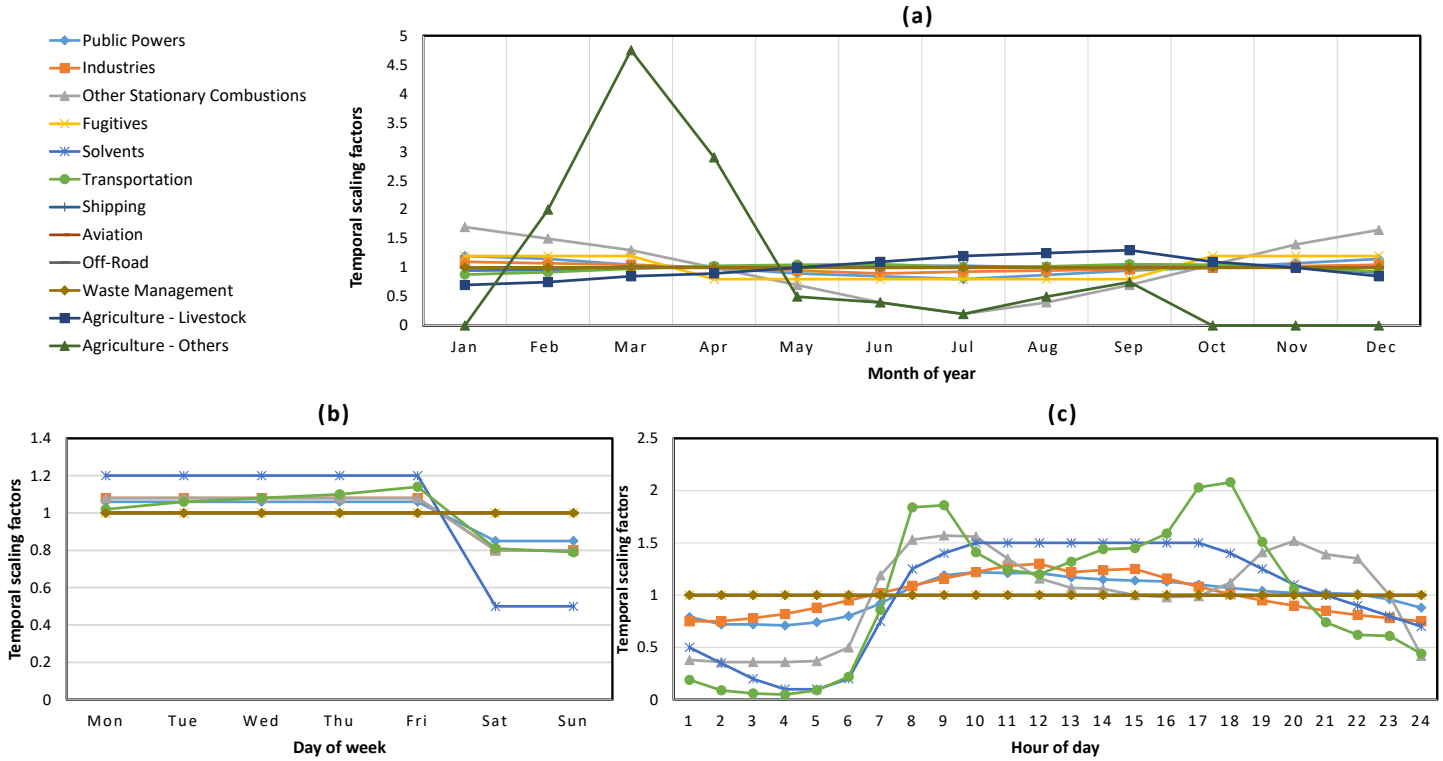


Figure 4.6.: Variation of temporal scaling factors used in TNO_GHGco, down-scaling (a) from yearly to monthly, (b) from weekly to daily and (c) from daily to hourly.

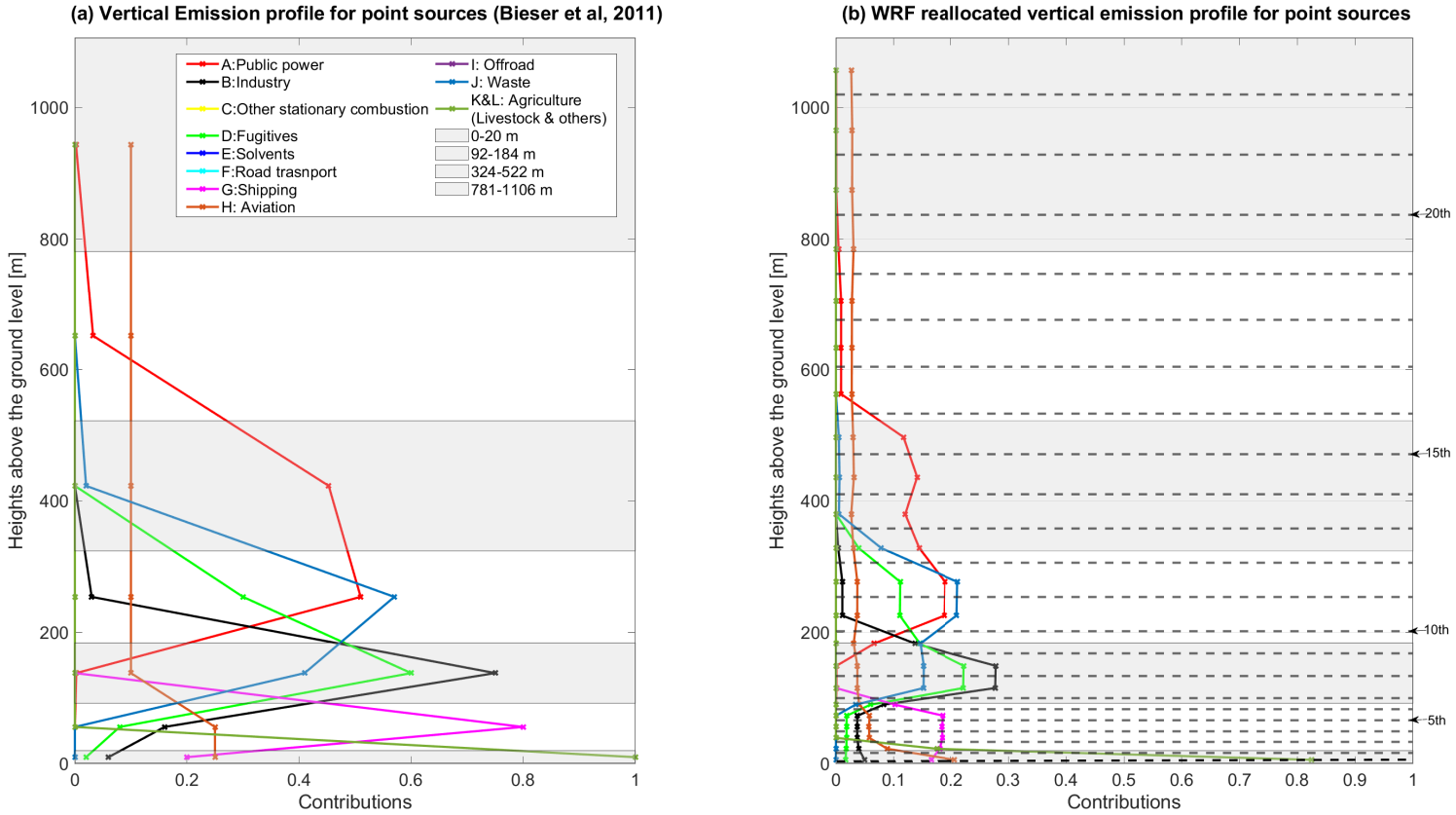


Figure 4.7.: Vertical emission profiles for point sources (a) are re-allocated to the WRF vertical layer (b). The alternating gray and white backgrounds stands for the seven vertical layers provided by Bieser et al. (2011) [79]. The dashed lines in (b) represent the first 22 vertical layers above the ground in the WRF setup.

EDGAR: For the initial test in Berlin, the external database 'Emission Database for Global Atmospheric research (EDGAR) Version 4.1' was used to provide initial and boundary conditions of the anthropogenic tracer, named as CO2_ANT in WRF. EDGAR Version 4.1 provides annually varying global anthropogenic GHG emissions and air pollutants at a spatial resolution of 0.1 ° [80], [81], whose source sectors include industrial processes, on-road and off-road sources in transport, large-scale biomass burning, and other anthropogenic sources [82].

Here the time factors for seasonal, weekly, daily and diurnal variations defined by the time profiles are applied [3], shown in Fig. 4.6. However, considerable uncertainties are to be expected in applying these time factors. This temporal variation set is derived based on western European data such that the representativity for other European countries and even other world regions may be quite poor.

The coarse emission fluxes used for the initialization of the anthropogenic tracer in WRF-Chem can cause problems when locating emission points within the high-resolution model grid and can weaken the impact from the real high-emission hotspots in the fine domain of the Berlin study (see Sect. 5.2.1).

The chemical sink for atmospheric CH_4 (e.g., photochemistry in the stratosphere) can be ignored in the model, owing to its relatively long lifespan (9.5 ± 1.3 year [83]), the small-scale domains and the limited simulation period (10 days) in this case study. The details on the model study for Berlin is described in Sect. 5.1.1.

TNO_GHGco_v1.1: For the modelling infrastructure applied for Munich, the first version of the TNO GHG and co-emitted species emission database (TNO_GHGco_v1.1 [84]) is used to initialize near-surface anthropogenic fluxes as tagged tracers. This database provides annual gridded anthropogenic emissions for 2015 at a horizontal resolution of 0.05 ° × 0.1 ° (latitude, longitude, approximately 6 km × 6 km), covering most of the European area and a part of North Africa.

The sources of emissions are classified into fourteen sectors following the GNFR categories ('A' to 'L'), such as industries, public powers and transportation. Sector-dependent vertical profiles are also provided, which are

[3] The temporal profiles are published via `https://edgar.jrc.ec.europa.eu/dataset_temp_profile` (last access: 03 Jan 2022)

applied to the point sources below approximately 1 km above the ground level.

Furthermore, a high-resolution version of TNO_GHGco_v1.1, was gridded at 1/120 ° × 1/60 ° (latitude, longitude, around 1 km × 1 km) is available for central Europe (i.e., all of Germany and parts of France, Poland and the Netherlands, etc.). This version of TNO_GHGco_v1.1 was prepared to support model studies at the local scale in the CO_2 Human Emissions project (CHE [85]). Since the outermost domain for Munich study (see Fig. 5.2 of Sect. 5) is not fully covered by the high-resolution version of TNO_GHGco_v1.1, the lower resolution (6 km) emissions are used for the outermost domain (D01), and high-resolution version for the other two domains (D02 and D03 in Fig. 5.2).

To prepare the input for WRF-Chem (i.e., the files with the name of *'wrfchemi_d0*'*), the required temporal disaggregation of the annual emissions was performed based on time-dependent scaling factors for monthly, weekly, and diurnal variations [84], [86], shown in Fig. 4.6. These default temporal profiles are provided per GNFR sector, including the variations of monthly scaling factors, weekly scaling factors and hourly scaling factors. It helps to capture the temporal emission characteristics of different sectors when downscaling the annual emission fluxes to hourly ones.

In addition, the near-surface emission fluxes from point sources per GNFR in TNO_GHGco_v1.1 are released from different heights above the ground, using the vertical profiles provided in Table 2 of Brunner et al. (2019)[87]. The layers in this vertical emission profiles (see Fig. 4.7(a)) are re-allocated to the model layers of of WRF, shown in Fig. 4.7(b).

For Munich, 11 emission sectors from GNFR are re-aggregated (see the categories in the legend of Fig. 4.7) to five major sectors applied in different tagged tracers. The details on the aggregations are shown in Table 4.6 for CO_2 and Table 4.7 for CH_4.

In this work, the emission fluxes provided by TNO_GHGco_v1.1 are fed into the geographical coordinates of WRF using Matlab, and further saved as *NetCDF* files. It is worth noting that the individual *.nc* files generated contain hourly emission maps from different sectors of one domain and the variable *Times* defined in this file should be consistent with the date time within the file name. When calling the *wrf* program, the model usually searches for the *'wrfchemi_d0*'* files for a certain domain at each full

Table 4.6.: Aggregations of emission categories for CO_2 from GNFR to WRF.

GNFR	Categories in WRF for CO_2				
	A. Power Plants	B. Industry	C. Other stationary combustion	D. Road Transport	E. Others
A: Public power	■				
B: Industry		■			
C: Other stationary combustion			■		
D: Fugitives					■
E: Solvents					■
F: Road transport				■	
G: Shipping					■
H: Aviation					■
I: Offroad					■
J: Waste					■
K&L: Agriculture (Livestock & others)					■

hour, and then reads emission fluxes from these files. The detailed implementation of defining and using these external emissions is described in Sect. 4.8.

4.2.3. Background concentration fields

In WRF, the initial and boundary conditions of background concentrations for GHGs are provided by external global databases, which can provide four dimensional concentration fields from global transport models. These fields are re-gridded to the geographical coordinates of WRF and applied to be the initial conditions of the corresponding tracers in the files with the name of *'wrfinput_d0*'* suffixed with the number of domains.

Moreover, the boundary conditions of each trace gas are calculated over the relaxation zones (see Fig. 3.4) and further saved in *'wrfbdy_d01'*. These calculations should be implemented after running the *real* program for the targeted period. A detailed description about the running of the entire modelling framework is mentioned in Sect. 4.8.

In this work, the estimated mixing ratios of CO_2 and CH_4 for both Berlin and Munich are provided by Integrated Forecasting System (IFS). IFS has been implemented by ECWMF to provide operational weather conditions and as such

Table 4.7.: Aggregations of emission categories for CH_4 from GNFR to WRF.

GNFR	Categories in WRF for CH_4					
	A. Power Plants	B. Industry	C. Agriculture	D. Waste Management	E. Fugitives and solvents	F. Others
A: Public power	■					
B: Industry		■				
C: Other stationary combustion						■
D: Fugitives					■	
E: Solvents					■	
F: Road transport						■
G: Shipping						■
H: Aviation						■
I: Offroad						■
J: Waste				■		
K&L: Agriculture (Livestock & others)			■			

extended to simulations of chemically interactive gases, GHGs and etc., and is a key part of the Copernicus Atmosphere Monitoring Service (CAMS) [88].

For Berlin, these fields were provided by IFS Cycle 40R1, with a spatial resolution of 0.8 ° on 137 vertical levels and with a temporal resolution of 6 hours. For Munich, the initial and boundary conditions of CO_2 and CH_4 were given by the IFS 45r1 [4]. The IFS Cycle 45r1 provides 6-hourly estimated mixing ratios of GHGs with a horizontal resolution of approximately 40 km (following the N256 reduced Gaussian grids from ECWMF) and a vertical resolution of 60 levels [88], [89].

4.3. Biogenic fluxes implemented in WRF-Chem

Biogenic fluxes of CO_2 have been found to largely influence the natural carbon cycle and urban carbon pools, estimations of which cannot be neglected in inverse models [90], [91]. Accurate estimations of net ecosystem exchange (NEE) and its two separated components, namely gross primary production (GPP) and respiration (RES), are therefore vital to provide insights of terrestrial biosphere and help optimize analysis of carbon fluxes over the area of interests.

[4]Atmosphere Monitoring Service is available via `https://confluence.ecmwf.int/pages/viewpage.action?pageId=116963341`.

Traditionally, two types of models are used to estimate biospheric fluxes: process based biochemical models and light-use-efficiency (LUE) models. Due to the development of space measurement techniques in recent years, a fresh concept used for the estimation of biospheric fluxes, solar-induced fluorescence (SIF), is being investigated and has proven to be an effective proxy of photosynthesis [27], [91]–[93]. SIF has been retrieved from satellite measurements and provided estimated GPP, such as OCO-2 SIF based GPP (GOSIF [94]), TanSat based GPP [95], and TROPOMI SIF based GPP [96].

In this work, biogenic CO_2 fluxes are implemented online utilizing the Vegetation Photosynthesis and Respiration Model (VPRM [90], [97]), a simple diagnostic LUE model. VPRM does not reproduce the physiological processes of vegetation, but rather calculates GPP using the input of meteorological variables and vegetation indices derived from remote sensing. RES is estimated using a simple linear model related to the air temperature and vegetation-specific parameters [90]. Then the hourly modelled NEE reflects the biospheric fluxes between the terrestrial biosphere and the atmosphere (i.e., RES − GPP).

In detail, the entire VPRM calculation is based on the satellite-derived indices, short wave radiation, and air temperature (i.e., the temperature at 2 m above the ground level (T2)) as provided by WRF [98]. The indices (i.e., Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI)) are taken from reflectance data measured by MODIS, specifically product MOD09A1 Version 6. MODIS has spatial resolutions of 0.5–1 km depending on the wavelength band and a temporal resolution of eight days [99]. The MODIS reflectance data are aggregated and transformed onto the LCC projection within the VPRM preprocessor, and the vegetation is classified following the Synergetic Land Cover Product (SYNMAP) data with a resolution of 1 km [100]. Then, the data, including the high-resolution vegetation indices, are available on the model domain. The equations and relevant descriptions regarding the estimation of RES and GPP can be found in Mahadevan et al. (2008) [90].

The VPRM subroutine in WRF-Chem is used to drive the calculation of NEE and included in the module file called *'module_ghg_fluxes.F'*. As claimed in this subroutine, 8 types of vegetation are classified using SYNMAP in VPRM, namely evergreen forest, deciduous forest, mixed forest, schrubland, savanna, cropland, grassland and others. Its map of the vegetation classifications over Munich is shown in the middle panel of Fig. 4.8. The air temperature with the unit of Celsius is used for the calculation of RES and obtained from the fields of T2 in WRF. All the characterizations in the calculation of RES and GPP are

collected based on the tall tower measurements around EU and updated by Dr. Michal Galkowski (MPI, Jena) in 2020.

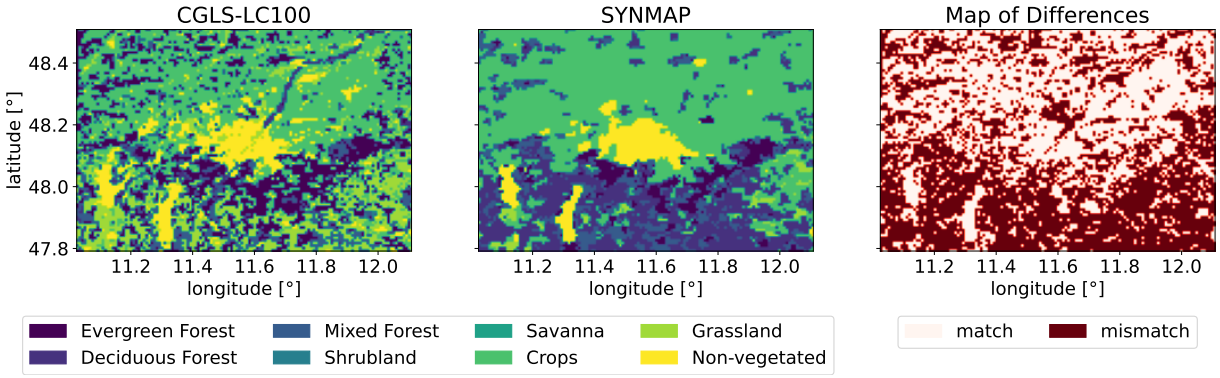


Figure 4.8.: The maps of the vegetation classifications with (left) and without (middle) refining using CGLS-LC100, and their difference (right).

Since the vegetation inside urban areas cannot be classified using the SYNMAP data, biogenic fluxes cannot be assigned with VPRM over cities [32]. Thus, the vegetation classification using the Dynamic Land Cover map of the Copernicus Global Land Service at a resolution of 100 m (CGLS-LC100) is extended and refined. This refined classification is used for our innermost domain to better capture the urban biogenic signals of CO_2. The CGLS-LC100 land cover map has been implemented to refine the vegetation classification over the innermost domain.

Figure 4.8 shows the refined map using CGLS-LC100 (left), the original map classified by SYNMAP (middle) and the difference between the two maps (right). As seen in the middle plot, there is no vegetation type classified in the middle area of the domain. After refining, more vegetation types can be recognized inside the city of Munich. Table 4.8 shows the details on how the Copernicus classes were reclassified to the eight vegetation fraction classes used in the VPRM preprocesser. As delving into the calculation of RES, it is inadequate to use a simple linear function with surface air temperature to calculation RES in VPRM. The dependence of RES should be not only on temperature, but also on EVI, water stress scaling factor, soil moisture and etc. [101], [102]. After such a modification for calculating RES, the updated VPRM is able to produce higher RES at nighttime during the growing season (June-September), which matches better compared to the RES values provided by the original VPRM.

Table 4.8.: Mapping the Copernicus classes to SYNMAP classes used in the VPRM prepossesser.

	SYNMAP VPRM Classes							
	Evergreen	Deciduous	Mixed Forest	Shrubland	Savanna	Cropland	Grassland	Others
Copernicus classes	Coniferous forest	Broad-leaved forest	Mixed forest	Moors and heathland	Beaches-dunes-sands	Rice fields	Natural grasslands	Continuous urban fabric
	-	-	-	Sclerophyllous vegetation	Bare rocks	Vineyards	Pastures	Discontinuous urban fabric
	-	-	-	Transitional woodland-shrub	Sparsely vegetated areas	Fruit trees and berry plantations	Green urban areas	Industrial or commercial units
	-	-	-	-	-	Olive groves	-	Road and rail networks and associated land
	-	-	-	-	-	Annual crops associated with permanent crops	-	Port areas
	-	-	-	-	-	Non-irrigated arable land	-	Airports
	-	-	-	-	-	Complex cultivation patterns	-	Construction site
	-	-	-	-	-	Land principally occupied by agriculture with significant areas of natural vegetation	-	Sport and leisure facilities
	-	-	-	-	-	Agro-forestry areas	-	Glaciers and perpetual snow
	-	-	-	-	-	-	-	Water courses
	-	-	-	-	-	-	-	Water bodies
	-	-	-	-	-	-	-	Coastal lagoons
	-	-	-	-	-	-	-	Sea and ocean
	-	-	-	-	-	-	-	NODATA
	-	-	-	-	-	-	-	Mineral extraction sites
	-	-	-	-	-	-	-	Dump sites
	-	-	-	-	-	-	-	Permanently irrigated land
	-	-	-	-	-	-	-	Inland marshes
	-	-	-	-	-	-	-	Peat bogs
	-	-	-	-	-	-	-	Salt marshes
	-	-	-	-	-	-	-	Salines
	-	-	-	-	-	-	-	Intertidal flats
	-	-	-	-	-	-	-	Estuaries
	-	-	-	-	-	-	-	Burnt areas

This underestimation of the original VPRM results in missing modelled peaks in concentrations on the early morning (see Sect. 6.4). To sum up, the calculation of RES in VPRM supposes to be improved. Since this adaptation is not easy to perform statistically, the current VPRM subroutine (in 2022) still holds the simple calculation and will be improved soon in WRF-Chem.

4.4. WRF-GHG development & operation

WRF is a numerical weather prediction system and can be used for both atmospheric research and operational forecasting on a mesoscale range from tens of meters to thousands of kilometers [103]. To produce high-resolution regional simulations of atmospheric CH_4 passive tracer transport, WRF was coupled with the Vegetation Photosynthesis and Respiration module (WRF-VPRM [104]). WRF-VPRM has been widely employed in several studies in which both the generally good agreement of the simulations with measurements and model biases were assessed in detail [18], [105]–[107]. Biogenic carbon fluxes given by VPRM tend to underestimate urban ecosystem carbon exchange, owing to the incomplete understanding of urban vegetation and to conditions related to urban heat islands and altered urban phenology [108].

WRF-VPRM was later extended to WRF-GHG [98], which can simulate the regional passive tracer transport for GHGs (CH_4, CO_2 and carbon monoxide (CO)). Relatively few studies using WRF-GHG have been published as of yet. Pillai et al. (2016) [109] utilized a Bayesian inversion approach based on WRF-GHG at a high spatial resolution of 10 km for Berlin to obtain anthropogenic CO_2 emissions and to quantify the uncertainties in retrieved anthropogenic emissions related to instruments (e.g., CarbonSat) and modelling errors. An observation system simulation experiment was studied in Pillai et al. (2016) [109] based on synthetic data rather than on real observations, as in this study.

It is worth noting here: the entire WRF modelling system is made up of four main components, namely WRF, WPS, WRF-Chem and WRF Data Assimilation (WRFDA). WPS and WRF have been described in Sect. 3.3. WRF-Chem is WRF coupled with Chemistry, including simulations of emission transport, mixing and chemical reactions of trace gases and aerosols, simultaneously driven by the modelled meteorological fields from WRF. Technically, it is an additional compressed package provided by NOAA and needs to be added into the *'WRF'* folder before the model compilation. Currently, it is released as part of the

WRF compacted package, especially in the official GitHub. Unlike the regular WRF, the operation of chemical part in WRF requires additional gridded input data as input, like *'wrfchemi_d0*'* in this work. These generated files need to meet with requirements of individual cases, and the source code of WRF-Chem and model configurations in *'namelist.input'* have to be modified accordingly. Regarding WRFDA, it is a state-of-the-art atmospheric data assimilation system. Here, observations (such as wind LiDAR and aircraft measurements) can be combined with global reanalysis weather databases, such as ERA5 and GFS, to provide more accurate simulated meteorological fields compared to WRF modelled results.

In this work, the modelling infrastructure is developed to provide a quantitative understanding of the processes responsible for the emission and consumption of CO_2 and CH_4. Throughout the working process, there are two phases, the initial test in Berlin based on WRF Version 3.2 and the built-up of the modelling infrastructure based on WRF Version 3.9.1.1, receptively.

For the initial test in Berlin, it was implemented using WRF Version 3.2. A detailed description of running WRF-GHG can be found in Beck et al. (2012) [98] and thus, only the initialization process for this case in particular is summarized here. One daily simulation with WRF-GHG is normally performed for a 30-hour time period, including a 6-hour spin-up for the meteorology from 18:00 to 24:00 UTC of the previous day and a 24-hour simulation of the tracer transport on the actual simulation day [98]. As for the boundary conditions, a small constant offset needs to be added into the WRF boundary files (i.e., *'wrfbdy_d01'*) for the biospheric CO_2 and the soil sink CH_4 tracers at the start of each run because these tracers can result in a net sink. When the concentrations become negative, the advected tracer fields will "disappear", as the WRF code does not allow tracers with negative values. An offset applied in the initialization process helps to avoid this problem and later is subtracted in the post-processing. As for the initial conditions, the meteorological conditions are initialized with external data sources each day to update the WRF meteorological fields properly. The tracers for the total and background CO_2 and CH_4 flux fields are initialized only once, at the first day of the simulation period, using CAMS as an external data source. Furthermore, the lateral boundary conditions of the outermost domain D01 are also initialized by global reanalysis concentration profiles.

Then, for the other days within the simulation period, these tracers for the total and background CO_2 and CH_4 fluxes are directly taken from the final

WRF output at 24:00 UTC of the previous day to make the entire simulation continuous. The CO_2 tracer for VPRM and the CH_4 tracer for soil uptake are also initialized with a constant offset to avoid the appearance of negative values caused, for example, by the vegetation respiration [98]. In terms of the other flux tracers, the tracer variables are initialized each day, using external data sources to provide the updated emission data for each tracer.

When turning to build the modelling infrastructure for Munich, it is based on WRF-Chem Version 3.9.1.1 with an updated GHG module [98]. Compared to the originally developed WRF-GHG, in which the GHG related modules had to be explicitly coupled to WRF Version. 3.2 (cf. Zhao et al. (2019) [86]), these modules have been added to the official WRF-Chem repository since WRF-Chem Version. 3.4. Specifically, chemical options 17 & 16 allows for GHG or CO_2 tracers to be included. It is worth noting that the GHG module does not take into account the atmospheric chemical reactions (i.e., it treats GHGs as passive tracers). That is, however, not expected to produce significant biases, due to the long lifetimes of GHGs compared to the relatively short residence time of tracers in the regional domain. To simulate the fluxes in the online mode, emission models are coupled with WRF and driven by internal model variables. Additionally, fluxes from external emission inventories (see Sect. 4.2.2) are included as input and these surface fluxes are transported as passive tracers in WRF [98].

4.5. Operation with HPC

WRF is a mesoscale model and can capture more features of atmospheric transport, especially near ground levels, compared to global weather models. Such models generally consume great amounts of computational power, depending on the spatial resolutions of models. For instant, the model at a 10 km resolution that covers the same area and maintains the same vertical layers, requires 27 times more computational power than the one with a resolution of 30 km. When the spatial resolution reaches 1 km, thousands of times more computational power is required relative to the one with a resolution of 10 km.

The modelling infrastructure produced in this work is compiled and run in the High Performance computing (HPC) provided by the Leibniz Supercomputing Center (Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ)). LRZ is the Information Technology provider for universities

in Munich and one of European Supercomputing Center. There are three major components in its HPC: SuperMUC-NG, Linux Cluster and Computer Cloud. The following includes main features for each component of HPCs at LRZ:

Linux Cluster: Linux cluster is the most common used platform for all the HPC users. Currently, three massive parallel clusters are available: CoolMUC-2 (cm2), CoolMUC-3 and Teramem. For each of these three groups of machines, they have their own module system and hardware setup. It is important to choose a more appropriate machine for individual cases by taking into account different aspects (e.g., available modules, versions of libraries and compilers, number of cores per node, node usage). In this work, most of the WRF tests for Munich took place on cm2, while the Berlin study was run in ivyMUC (retired in Sept. 2021). All runs related to Computational Fluid Dynamics (CFD) were operated in CoolMUC-3.

SuperMUC-NG: Running in SuperMUC-NG is only allowed if sufficient tests have been completed on a Linux Cluster or other machine. Computational resources are allocated, by requesting a specific partition, and further subtracted from the project budget.

Computer Cloud: Computer cloud can provide users with enough freedom to run services and test software, while using less computational resources. Thanks to the freedom in the cloud, the *cron* job (i.e., job schedulers similar to UNIX operating systems) and *Docker* runs can be operated by users as administrators.

Here are the descriptions of characteristics of the Linux Clusters used most often in this work:

File system: Three working directories on Linux Clusters are available: *'Home'*, *'Project'* and *'SRATCH'*. With the restriction of space in *'Home'* (100 GByte) and *'Project'* (1 TByte), most of tests and long-period runs have been operated in *'SRATCH'* (1,400 TByte). But *'SRATCH'* is not safe to save data, useful results, and even executable, since any file or folder can be unpredictably deleted in *'SRATCH'* to avoid overflow of the large scale storage.

Module system: Due to the restrictions of management licences, LRZ provides a large amount of structured software (including libraries, compilers, parallel run-time environments and open source tools) to its HPC users. All these programming are included in the module system, public and easily called to manage the environment.

In the HPC system at LRZ, a default set of module packages related to the setup of system environment are loaded. Once logging into nodes, all commands related to module packages are available automatically. It should be noted, however, that different machines offer different sets of default files and different versions of module packages. Regarding the application of CFD, it should be run using the IP address (*'lxlogin8.lrz.de'*) while for WRF, another platform (*'lxlogin2.lrz.de'*) is recommended.

Since the entire workflow of the WRF modelling infrastructure (Sect. 4.8) is completed and its operation requires more computational resources for long-term runs, it is much applicable if it can be compacted as a *Docker* container. However, owing to the strict restrictions of management licences at LRZ, *Docker* cannot be built and tested at LRZ. Thanks to the support from the Large-scale EXecution for Industry & Society project (LEXIS), I am able to get amount of computational resources from IT4Innovations that is the super-computing center from Czech, and test the *Docker* container efficiently.

4.6. Compilation with HPC

Three main steps are required when compiling the WRF in cm2 at LRZ:

1. **Set up Linux environment**: Apart from the modules loaded automatically, another important module required in this work is a bundle-package: *netcdf-hdf5-all/4.7_hdf5-1.10-intel19-impi*. It includes five libraries for the compilation of WRF: *NetCDF* Version 4.7.4, *HDF5* Version 1.10.7, *Intel* Version 9, *Parallel-NetCDF* Version 1.12.1, and *NetCDF-Fortran* Version 4.5.3.

   Apart from loading these necessary libraries, the paths to the *'LIBS'* and *'INCLUDE'* folders of *NetCDF* and *Jaspers* need to be claimed before the configuration and the compilation of WRF. Details on the compilation of WRF can be found in the User Guide of WRF [5].

2. **Adapt the GHG related modules**: Although a few commonly used emission variables and tracers are defined in source code of WRF-Chem by default, additional trace gases and emission tracers of interest should be updated and defined in the *'Registry.chem'* file of WRF to meet the requirements of individual cases.

[5]The WRF user guide used to guide this work is available through `https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v3/contents.html`.

In the source code of WRF, a file called *'Registry.chem'* is saved in the *'Registry'* folder, and used to define the attributes and properties of required trace gases and their emission inputs. This file contains the dimensions of these variables, how they should be treated, whether they will be saved in outputs, etc. All the updates in *'Registry.chem'* refer to modifications of the WRF source code via the *cpp* directive, which is equivalent to generating new source code during the compilation of WRF. For each variable proposed in *'Registry.chem'*, a number of elements should be listed, shown in Tables. 4.9-4.12. Tables 4.9 & 4.10 list the emission variables of inputs for tracer gases, while Tables 4.11 & 4.12 show the definition of tracer gases.

After proposing these new variables in *'Registry.chem'*, they should be included in specific packages for associating these newly defined variables with specific physical and chemical schemes. In this work, these newly defined tracers are associated with GHG related packages by declaring in the package lines of *'Registry.chem'* as follows,

```
# Anthropogenic CO2, CO and CH4 emissions:
package eco2    emiss_opt == 16 -   emis_ant: e_co2, e_ch4
package eghg    emiss_opt == 17 -   emis_ant: e_co2, e_co2_a, e_co2_b, e_co2_c,
    e_co2_f, e_co2_o, e_co2_bf, e_co2_ff, e_ch4, e_ch4_a, e_ch4_b, e_ch4_c,
    e_ch4_e, e_ch4_f, e_ch4_o

# GHG chem packages, passive tracers
package co2_tracer  chem_opt == 16  -   chem: co2_ant, co2_bio, co2_bck
package ghg_tracer  chem_opt == 17  -   chem: co2_ant, co2_bio, co2_oce, co2_bck,
    co2_bf, co2_ff, co2_a, co2_b, co2_c, co2_f, co2_o, ch4_ant, ch4_bio, ch4_bck,
    ch4_a, ch4_b, ch4_c, ch4_e, ch4_f, ch4_o, bio_ch4wet, bio_ch4soil, bio_ch4term
```

The newly defined tracers are used as inputs or considered in the model when setting chemical and emission mechanisms as '17' (i.e., *Chem_opt* and *emiss_opt* in *'namelist.input'*). As shown in the Fortran commands below, these variables and their associated packages are linked to the detailed calculations of the GHG-related solvers by inserting the corresponding solver routine (*'add_ghg_fluxes'*) encoded in the module file named *'module_ghg_fluxes.F'* in the *'Chem'* folder. In this subroutine, the *emis_ant* arrays refer to the emission fluxes of specific emission tracers, which are declared in *'Registry.chem'* as operating under a certain emission mechanisms.

```
! Solver routine of 'add_ghg_fluxes'
call wrf_debug(15,'add_ghg_fluxes')
```

```
! For both GHG options
DO j=jts,jte
  DO i=its,ite
!  3D anthropogenic fluxes
    DO k=kts,min(config_flags%kemit,kte)
       conv_rho=8.0461e-6/rho_phy(i,k,j)*dtstep/dz8w(i,k,j)
       !8.0461e-6 = molar_mass(air)/3600, [g/mol/s]
       chem(i,k,j,p_co2_ant)=chem(i,k,j,p_co2_ant)+conv_rho*emis_ant(i,k,j,p_e_co2)
       chem(i,k,j,p_ch4_ant)=chem(i,k,j,p_ch4_ant)+conv_rho*emis_ant(i,k,j,p_e_ch4)
! Additional tagged tracers for Munich WRF-GHG campaign
     chem(i,k,j,p_co2_bf)=chem(i,k,j,p_co2_bf)+conv_rho*emis_ant(i,k,j,p_e_co2_bf)
     chem(i,k,j,p_co2_ff)=chem(i,k,j,p_co2_ff)+conv_rho*emis_ant(i,k,j,p_e_co2_ff)
       chem(i,k,j,p_co2_a)=chem(i,k,j,p_co2_a)+conv_rho*emis_ant(i,k,j,p_e_co2_a)
       chem(i,k,j,p_co2_b)=chem(i,k,j,p_co2_b)+conv_rho*emis_ant(i,k,j,p_e_co2_b)
       chem(i,k,j,p_co2_c)=chem(i,k,j,p_co2_c)+conv_rho*emis_ant(i,k,j,p_e_co2_c)
       chem(i,k,j,p_co2_o)=chem(i,k,j,p_co2_o)+conv_rho*emis_ant(i,k,j,p_e_co2_o)
       chem(i,k,j,p_co2_f)=chem(i,k,j,p_co2_f)+conv_rho*emis_ant(i,k,j,p_e_co2_f)
       chem(i,k,j,p_ch4_a)=chem(i,k,j,p_ch4_a)+conv_rho*emis_ant(i,k,j,p_e_ch4_a)
       chem(i,k,j,p_ch4_b)=chem(i,k,j,p_ch4_b)+conv_rho*emis_ant(i,k,j,p_e_ch4_b)
       chem(i,k,j,p_ch4_c)=chem(i,k,j,p_ch4_c)+conv_rho*emis_ant(i,k,j,p_e_ch4_c)
        chem(i,k,j,p_ch4_o)=chem(i,k,j,p_ch4_o)+conv_rho*emis_ant(i,k,j,p_e_ch4_o)
       chem(i,k,j,p_ch4_e)=chem(i,k,j,p_ch4_e)+conv_rho*emis_ant(i,k,j,p_e_ch4_e)
       chem(i,k,j,p_ch4_f)=chem(i,k,j,p_ch4_f)+conv_rho*emis_ant(i,k,j,p_e_ch4_f)
! 2D biospheric fluxes:
    if (k == 1) then
         chem(i,1,j,p_co2_bio)=chem(i,1,j,p_co2_bio)+conv_rho*
    (eghg_bio(i,1,j,p_ebio_gee)+ eghg_bio(i,1,j,p_ebio_res))
         ! both uptake and release
     end if
    ENDDO
  ENDDO
ENDDO
! For the GHG_TRACER option only
IF(config_flags%chem_opt == GHG_TRACER) THEN
    DO j=jts,jte
      DO i=its,ite
!  3D anthropogenic fluxes
        DO k=kts,min(config_flags%kemit,kte)
           conv_rho = 8.0461e-6/rho_phy(i,k,j) * dtstep/dz8w(i,k,j)
! 2D biospheric fluxes:
        if (k == 1) then
           chem(i,1,j,p_ch4_bio)=chem(i,1,j,p_ch4_bio)+conv_rho*
    (eghg_bio(i,1,j,p_ebio_ch4wet)+eghg_bio(i,1,j,p_ebio_ch4soil)+
    eghg_bio(i,1,j,p_ebio_ch4term))
           chem(i,1,j,p_bio_ch4wet)=chem(i,1,j,p_bio_ch4wet)+conv_rho*
    eghg_bio(i,1,j,p_ebio_ch4wet)
              chem(i,1,j,p_bio_ch4soil)=chem(i,1,j,p_bio_ch4soil)+conv_rho*
    eghg_bio(i,1,j,p_ebio_ch4soil)
              chem(i,1,j,p_bio_ch4term)=chem(i,1,j,p_bio_ch4term)+conv_rho*
    eghg_bio(i,1,j,p_ebio_ch4term)
        end if
      ENDDO
    ENDDO
  ENDDO
END IF
```

Table 4.9.: List of emission variables for trace gases of CO_2 used in the modelling infrastructure for Munich.

DName[a]	Emission Variables							
	Meta Name[b]	Description	Unit	Type	Stag[c]	IO[d]	Dims[e]	Use
E_CO2	e_co2	Anthropogenic CO_2 fluxes	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
EBIO_GEE	ebio_gee	Biospheric VPRM CO_2 uptake	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	-[d]	ivjf	eghg_bio
EBIO_RES	ebio_res	Biospheric VPRM CO_2 release	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	-	ivjf	eghg_bio
E_CO2_BF	e_co2_bf	TNO CO_2 emissions from biomass fuel	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_FF	e_co2_ff	TNO CO_2 emissions from fossil fuel	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_A	e_co2_a	TNO CO_2 emissions from public power stations	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_B	e_co2_b	TNO CO_2 emissions from industry	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_C	e_co2_c	TNO CO_2 emissions from other stationary combustion	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_F	e_co2_f	TNO CO_2 emissions from road transport	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CO2_O	e_co2_o	TNO CO_2 emissions from other tracers	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant

[a] 'Dname' is the name shown in the read-in and output files;
[b] 'Meta Name' stands for the name of the variable inside of WRF;
[c] 'Stag' are the string used to indicate stagger dimensions of the variables;
[d] 'IO' shows whether and how the variable is available in initial input (i), history (h), restart (r) and nesting. The digit number from 1 to 9 denote which auxiliary input is related to and '5' here represent the *'wrfchemi_d0*'*
[e] 'Dims' represent the number and the order of the dimension;
[f] '-' means that the biogenic fluxes is not directly fed by external database, while generated online using the VPRM module in WRF-Chem;

Table 4.10.: List of emission variables for trace gases of CH_4 used in the modelling infrastructure for Munich.

DName	Emission Variables							
	Meta Name	Description	Unit	Type	Stag	IO	Dims	Use
E_CH4	e_ch4tst	Anthropogenic CH_4 fluxes	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
EBIO_CH4WET	ebio_ch4wet	Biogenic CH_4 wetland emissions	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	-	ivjf	eghg_bio
EBIO_CH4SOIL	ebio_ch4soil	CH_4 soil uptake fluxes	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	-	ivjf	eghg_bio
EBIO_CH4TERM	ebio_ch4term	CH_4 termite emissions	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	-	ivjf	eghg_bio
E_CH4_A	e_ch4_a	TNO CH_4 emissions from public power stations	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CH4_B	e_ch4_b	TNO CH_4 emissions from industry	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CH4_C	e_ch4_c	TNO CH_4 emissions from other stationary combustion	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CH4_E	e_ch4_e	TNO CH_4 emissions from waste management	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CH4_F	e_ch4_f	TNO CH_4 emissions from agriculture	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant
E_CH4_O	e_ch4_o	TNO CH_4 emissions from other tracers	$mol \cdot km^{-2} \cdot hr^{-1}$	real	Z	i5r	i+jf	emis_ant

Table 4.11.: List of Chemical scalars of tracers for trace gases of CO_2 used in the modelling infrastructure for Munich.

DName	Tracer Variables							
	Meta Name	Description	Unit	Type	Stag	IO	Dims	Use
CO2_ANT	co2_ant	Mixing ratio of anthropogenic CO_2	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}[a]	ikjftb	chem
CO2_BIO	co2_bio	Mixing ratio of VPRM CO_2	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_BCK	co2_bck	Mixing ratio of background CO_2	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_BF	co2_bf	Mixing ratio of CO_2 from biomass fuel	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_FF	co2_ff	Mixing ratio of CO_2 from fossil fuel	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_A	co2_a	Mixing ratio of CO_2 from public power station	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_B	co2_b	Mixing ratio of CO_2 from industry	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_C	co2_c	Mixing ratio of CO_2 from other stationary combustion	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_F	co2_f	Mixing ratio of CO_2 from road transport	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CO2_O	co2_o	Mixing ratio of CO_2 from others	ppmv	real	-	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem

[a] 'i0{12}rhusdf' in the IO stream donates that the variable is saved and used as the combination of the initial (i), history (h), restart (r) and another four nesting options, namely upscale (u), downscale (d), smooth (s) and force (f). Specifically, if this variable needs to be fed from the course domain to its inner one, it refers to a 'downscale' interpolation, while back from the inner domain to its outer one, it is an 'upscale' interpolation. Then if the lateral boundary conditions of this variable are required for each parent time step, 'f' means the lateral boundary forcing and 's' is for an option for each feedback used for smoothing the data inside the inner/nest domain.

Table 4.12.: List of Chemical scalars of tracers for trace gases of CH_4 used in the modelling infrastructure for Munich.

DName	Tracer Variables						
	Meta Name	Description	Unit	Type	IO	Dims	Use
CH4_ANT	ch4_ant	Mixing ratio of anthropogenic CH_4	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_BIO	ch4_bio	Mixing ratio of biogenic CH_4	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
BIO_CH4WET	bio_ch4wet	Mixing ratio of natural biogenic CH_4 from wetland	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
BIO_CH4SOIL	bio_ch4soil	Mixing ratio of natural biogenic CH_4 from soil	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
BIO_CH4TERM	bio_ch4term	Mixing ratio of natural biogenic CH_4 from termit	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_BCK	ch4_bck	Mixing ratio of background CH_4	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_A	ch4_a	Mixing ratio of CH_4 from public power station	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_B	ch4_b	Mixing ratio of CH_4 from industry	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_C	ch4_c	Mixing ratio of CH_4 from other stationary combustion	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_E	ch4_e	Mixing ratio of CH_4 from waste management	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_F	ch4_f	Mixing ratio of CH_4 from agriculture	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem
CH4_O	ch4_o	Mixing ratio of CH_4 from others	ppmv	real	i0{12}rhusdf = {bdy_interp:dt}	ikjftb	chem

These fluxes are initialized from the input files named *'wrfchemi_d0*'* suffixed with the number of domains and the date time. The generation of these chemical input files is described in Sect. 4.7. The emission variables are then converted from emission arrays to the increase in the mixing ratios at the bottom layer of the domain (i.e., chemical arrays) by using pre-defined conversion rates. Moreover, if there is any change of the source code in this step, the WRF should be reconfigured and recompiled.

3. **Build the executable of WRF and WPS**: After the necessary tracers and their associated calculations have been declared and defined in the WRF source code, WRF can be configured and compiled. Four executable are then generated in the *'main'* folder and linked to both *'test/em_real'* and *'run'* folders. The three executable used in WPS can be created on the basis of a complete WRF compilation, since these programs in WPS require *I/O Application Programming Interface* library from WRF, but the *'Registry.chem'* in WRF is not shared when compiling WPS. That means that all the edits in Step 2 are not affected the compilation of WPS.

4. **Visualize and analyze WRF modelled results**: The programming tools are used to plot and do data analysis of modelled results by using *R*, *Matlab* and *Python*. Moreover, there are a number of compressed tools and commands that can be used to visualize the WRF outputs, such as *Ncview* and *NCL* commands in the Linux based system. *Panoply* is a common tool for visualizing *NetCDF* files in Windows, especially for drawing geographical maps.

4.7. Modelling infrastructure for Munich

Since the establishment of the firstly fully-automatic urban GHG measurement network in Munich in 2019 [110], a high-resolution modelling infrastructure has been built which is the basis for a quantitative understanding of the processes responsible for the emission and consumption of CO_2, CH_4, and CO in Munich. The results of this infrastructure are expected to provide guidance for local emission reduction strategies.

The entire modelling infrastructure is designed based on the five programs built in WRF (Sect. 3.3) and made up of three key parts: WPS (Sect. 3.3.1), the WRF main run (Sect. 3.3.2) and the post-processing. As shown in the main workflow in Fig. 4.9, the main inputs to this modelling infrastructure

are obtained by downloading from online servers: the UCAR server for the static geographical database, the CDS server for the meteorological inputs and the DSS container from LRZ for the initialization of background fields and emission fluxes. After the internal operation of this modelling framework, the WRF output is further processed to animate modelled concentration fields of trace gases and to validate them by comparison with available measurements. Moreover, the simulated hourly meteorological fields are converted to the *ARL* data format [6] and used as the driver for particle transports in STILT (see Sect. 7.2.2).

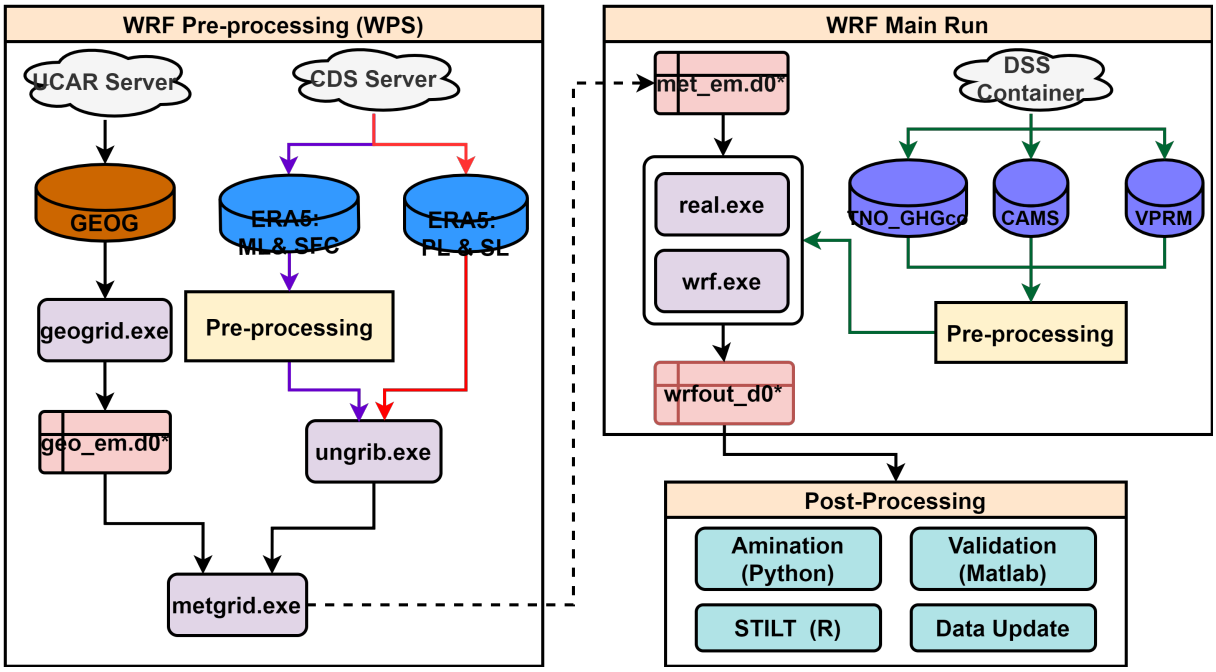


Figure 4.9.: Workflow of the modelling infrastructure in Munich for one day.

Descriptions of running WRF-GHG can be found in Beck et al. (2012) [98] and the detailed operation process has been described in Sect. 4.4. Thus, only the detailed initialization process for individual tracers and the details of this modelling infrastructure are documented in this section. In short, one daily simulation is normally implemented for a 30-hour time period, including a 6-hour spin-up run for the meteorology from 18:00 to 24:00 UTC of the previous day (Fig. 4.11(c)) and a 24-hour simulation of the tracer transport on the actual simulation day (Fig. 4.11(d)) [98].

[6]The detailed information of the *ALR* data format can be found via `https://www.ready.noaa.gov/hysplitusersguide/S141.htm`.

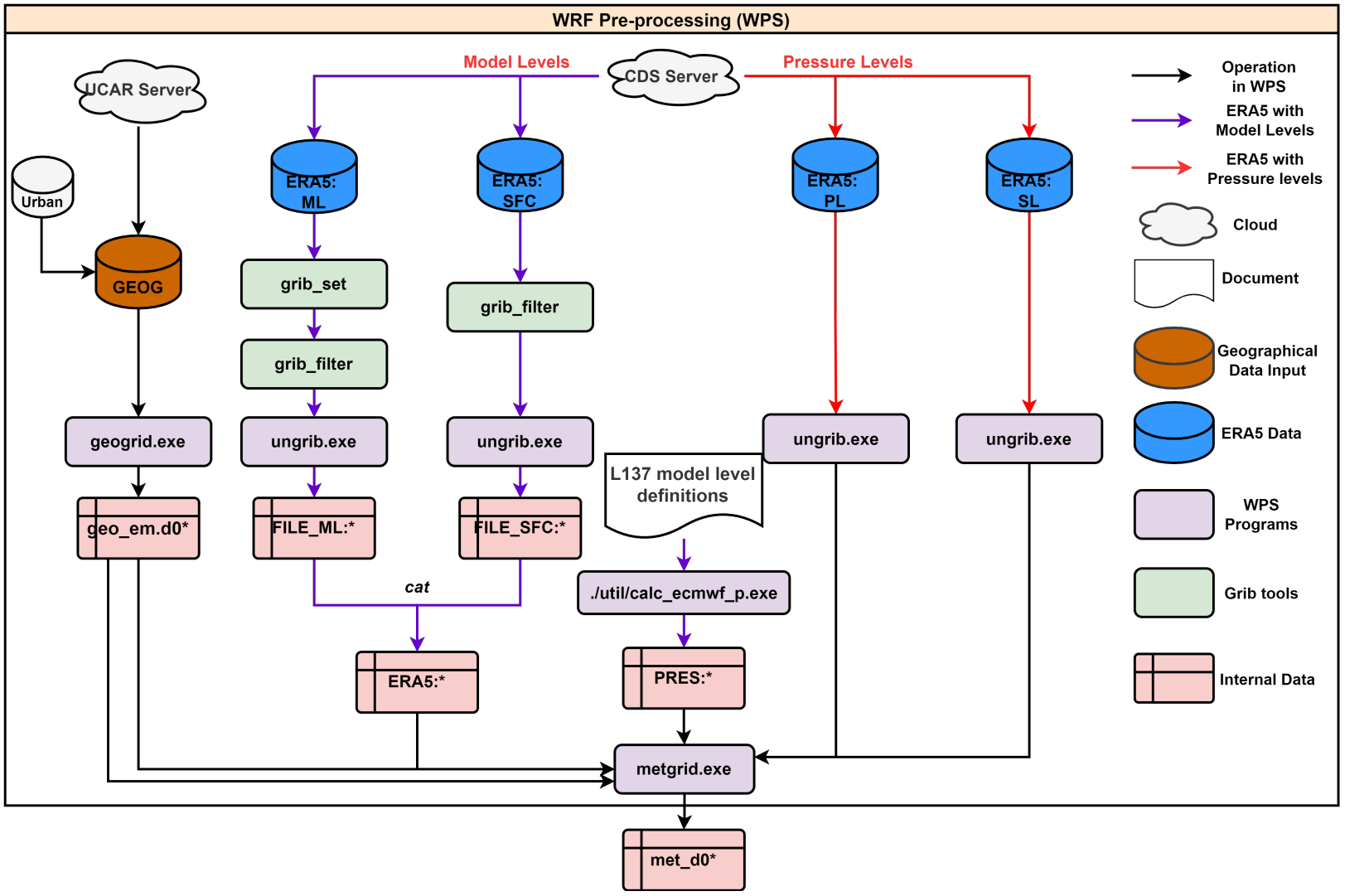


Figure 4.10.: Workflow of WPS in the modelling infrastructure in Munich for one day.

The entire infrastructure starts with initializing the meteorological conditions. External data sources (GFS for Berlin and ERA5 for the Munich modelling infrastructure) are used each day to update the meteorological fields properly. During the running of WPS, three steps are proceeded centered around three main programs described in Sect. 3.3. Different from the common run of WPS there are two internal steps required to be care about when running with ERA5:

a. **Add the LCZ urban typologies**: Details on the generation of LCZ maps and their impact on building WRF are descried in Sect. 4.1. When running the *geogrid* program, the *geog_data_res* setting needs to be adjusted in *'namelist.wps'*. That is, apart from the global static geographical database (i.e., *modis_landuse_20class_30s_with_lakes* in this infrastructure), the newly generated land use is claimed by adding its folder name in this option (i.e., *'land_munich'*).

b. **Initialize with ERA5**: As mentioned in Sect. 4.2.1, ERA5 is not specifically provided for initializing WRF. Two types of ERA5 databases are used in WRF: ERA5 model level data (137 vertical levels) and pressure level data (37 vertical levels). The difference of these two products is that the pressure level product is a near-real-time database that assimilates as many observations as possible in the upper air and near surface, whereas the model level product has a more granular vertical layer with a two month time lag for updates. Additionally, only authorized users have the right to download model level products after their data migration to MARS on Oct. 2021.

   In this work, most of the tests in Munich were initialized with the ERA5 model level product. To meet with the requirements of long-term WRF tests, both downloading options have been included in this modelling infrastructure and they can be flexibly selected when building and running the tests based on the *Docker* container (Sect. 4.8).

   In terms of the detailed operations of downloading ERA5 databases, the *Python* scripts are used to access the *GRIB* format data from the CDS server. The *ecCodes Grib* tools that are developed by ECWMF for decoding and encoding *Grib* files, are further invoked for further data processing, i.e., temporal split from monthly or daily to hourly and the format conversion from *GRIB2* to *GRIB1*. Here the hourly *GRIB1* files are eventually required.

Then run the *ungrib* program to read the meteorological fields and save them into the intermediate files. In this step, the *'Vtable'* files need to be changed with different databases. The purple arrows drawn in Fig. 4.10 indicate the steps used when running with the ERA5 model level. Since no pressure definition in the model level database is provided, an additional file called *'ecmwf_coeff'* is required to define the 137 model-level pressure information along its defined altitudes, as shown in Fig. 4.10.

At last, the hourly meteorological files are generated in this step, which contains the gridded meteorological and geographical fields along the vertical distribution provided by the initialized meteorological databases. Next, turn to the data preparation of the WRF main run, specifically for the initial and boundary conditions of individual tracers and inputs of the online calculation processes, such as VPRM (Sect. 4.3).

Regarding the boundary and initial conditions for the biospheric (i.e., CO2_BIO and CH4_BIO) and its associated tracers (e.g., CH4_WET, CH4_SOIL) as shown in Tables 4.11 & 4.12, constant offsets need to be added in the initialization process helps to avoid this problem and later is subtracted in the post-processing, 400 ppm for CO_2 and 1.9 ppm for CH_4, which has been mentioned in Sect. 4.4. This progress is implemented by *Matlab*, simultaneously with the initialization of initial and boundary conditions for other tracers.

The tracers of the total and background CO_2 and CH_4 flux fields are initialized only once, at the first day of the simulation period by aggregating external data source (like CAMS in this work) to the initial files (i.e., *'wrfinput_d0*'*). Furthermore, the lateral boundary conditions of the outer domain D01 in *'wrfbdy_d01'* are also taken from the same source. As introduced in Sect. 3.3.2, two types of variables for each tracer need to be initialized. One type is the initial condition for the relaxation zones shown in Fig. 3.4 and the other is their lateral boundary condition (i.e., the temporal tendencies of the variable between 6 hours). Then, for the other days within the simulation period, these tracers of the total and background fluxes are directly taken from the WRF output at 24:00 UTC of the previous day to make the entire simulation continuous. The biospheric CO_2 and CH_4 tracers are always simultaneously initialized with a constant offset [98]. In terms of the other flux tracers, the tracer variables are initialized each day, using external data sources to provide the updated emission data for each tracer. This is done through inserting emission fluxes using the *'wrfchemi_d0*'* files.

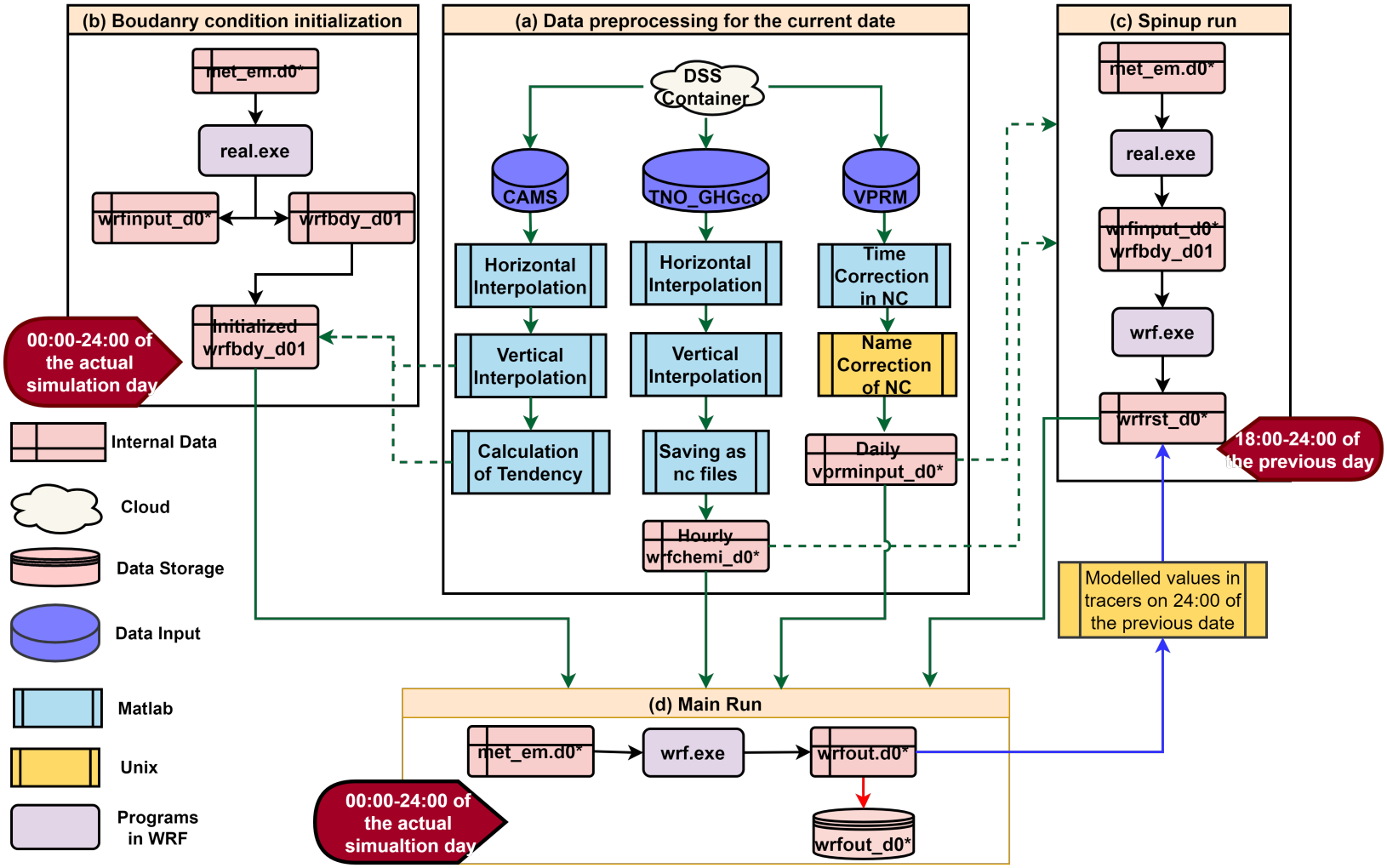


Figure 4.11.: Workflow of the WRF run in the modelling framework in Munich for one day (i.e., the actual simulation day in the workflow).

As shown in Fig. 4.11, all these data preparation steps described above are implemented by *Matlab*. In brief, the hourly *'wrfchemi_d0*'* files are generated by downscaling the annual emission inventory (TNO_GHGco) in space and time (Fig. 4.11(a)). The external global reanalysis flux database (CAMS) are interpolated in space horizontally and vertically and saved in *'wrfbdy_d01'* that include the entire simulation days (Fig. 4.11(b)). The daily input of VPRM online mode is a *NetCDF* file named *'vprminput_d0*'* suffixed with the number of domains and the date time (Fig. 4.11(a)). Here the date time in its name need to be consistent with the *Time* variable inside the *NetCDF* file and the start time of *'namelist.input'*. Otherwise, the model is not able to open these input files. At the end, the modelled results at 24:00 UTC of the previous day are read and copied to the spinup outputs (see yellow box in Fig. 4.11). These outputs are generated by 6-hour spinup run and called *'wrfrst_d0*'* suffixed with the number of domains and the start time of the actual simulation day (Fig. 4.11(c)). This process is completed by calling the *ncks* commands [7].

4.8. Docker-WRF container for Munich

As documented in Sect. 4.7, the entire workflow of this modelling infrastructure is complicated with a series of scripts to be run for internal data processing and amount of steps in the spinup and main WRF run. However, the WRF workflow is commonly run manually, without a sophisticated orchestration and data management solutions. Thus, one of the main purposes in this work is to create a semi-operational modelling infrastructure which can be easily run even by modellers who do not have a comprehensive understanding of the internal mechanisms, when all the necessary inputs are in place.

The simulations are automated by orchestration of the chain 'pre-processing (WPS) – WRF main runs – post-processing'. This infrastructure is built to reduce manual intervention and to solve the problems of running the next long-term study in Munich with better accessibility and efficiency.

In this work, *Docker* is used to build such a semi-operational infrastructure. As an open source platform, *Docker* enables developers to package their applications into containers. Each container can include executable/programs with operating system libraries and dependencies coded in any environment.

[7] The *ncks* commands are described in details under `https://linux.die.net/man/1/ncks`.

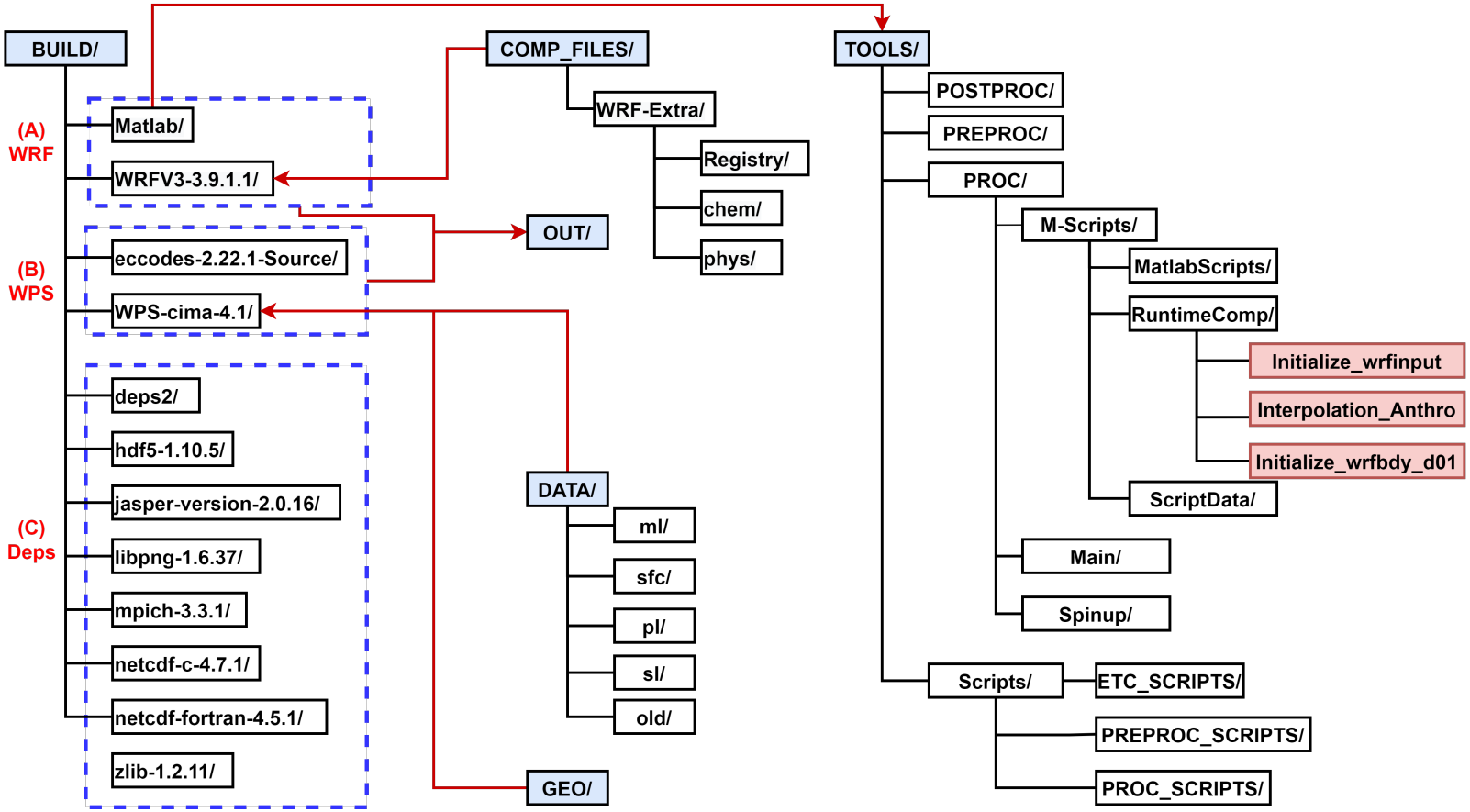


Figure 4.12.: File structure of the *Docker* container designed for the modelling infrastructure of Munich.

That is, the container can be isolated at the process levels and have its own file system. By using a series of reusable images, *Docker* can deploy, replicate, move and back up workloads in a streamlined way, which is more portable and flexible. Compared to other virtual machine, the *Docker* container offers a lightweight and speedy option, since it can isolate the execution environment by sharing the underlying operating system kernel, rather than building its own operating system.

This application is supported by the Large-scale EXecution for Industry & Society project (LEXIS), which has built a state-of-the-art engineering platform at the confluence of HPC, cloud and big data. Thanks to their technical support, I am able to build this container as an open call in their project and tests the container with enough computation resources on Linux Clusters from the IT4Innovation National Supercomputing Center in Ostrava, Czech.

The WRF modelling infrastructure in this work is supposed to be automated/orchestrated for long-term simulations by building the *Docker* container based on the layers provided by the WRF GFS computation over Italy. Section 4.7 describes in detail the whole workflow, while in this section, the built-up of the container will be documented from the perspective of the file structure (see Fig. 4.12). Three main steps are included in this container, respectively, pre-processing, HPC WRF run and post-processing with *Matlab*, *R* and *python*. overall, six main folders are distributed marked as blue in the *'root'*, namely *'BUILD'*, *'COMP_FILES'*, *'DATA'*, *'GEO'*, *'OUT'* and *'TOOLS'*.

The running of this container starts with the built-up of dependencies and the compilation of WRF. The version of WRF and WRF-Chem used in this infrastructure is the Version 3.9.1.1, rather than Version 4.1 that has been built in the existing layers of the Italy case. But the version of WPS is not affected. In this case, therefore, only WRF is re-constructed in the new layer with version-specific dependencies, which should be highlighted since it will directly affect on the compilation of WRF and the operation of the container. The detailed information is described in the thesis of Alp Berkman [111]. Then the source code of WRF and WRF-Chem require to be edited accordingly (see Sect. 4.2.3). Further, WRF are configured and compiled in the container. All these steps are run and required dependencies are built in the *'BUILD'* folder. The edited source code of WRF are saved in the *'COMP_FILES'*, where all the edits related to the GHG modules described in Sect. 4.5 are saved.

Next is the running of WPS. The detailed steps are shown in Fig. 4.10 and de-

scribed in Sects. 3.3.1 and 4.7. In the container, the *'DATA'* and *'GEO'* folders are used to save the downloaded ERA5 and geographical databases. The output generated here is the hourly meteorological fields saved in the *'OUT'* folder temporally. All these downloaded and intermediate files will be deleted after the generation of the WRF outputs.

Subsequently, the processing of biospheric, anthropogenic, background and total tracers is initiated. In this part, all the relevant scripts used are saved in the *'TOOLS'* folder, marked as red rectangles in Fig. 4.12. They refer to the three groups of *Matlab* scripts used to generate or process the initial and lateral boundary conditions for different tracers. The output here includes the hourly gridded near-ground fluxes saved in the *'wrfchemi_d0*'* files, as well as the daily *'vprminput_d0*'* files for the 18:00 UTC used in the spinup run and the 00:00 UTC in the main run. In addition, the concentration fields taken from CAMS are fed into the initial and lateral boundary conditions for the background and total traces in the *'wrfinput_d0*'* and *'wrfbdy_d01'*.

Further, the spinup and main runs are operated in order following the workflow shown in Fig. 4.11. The final output (i.e., '*wrfout_d0**') in the container is stored in the *'OUT'* folder. This folder can also be mounted in the local computer and can further used for data post-processing. Figure A.1 shows the workflow of the sequence of running scripts in the container, allowing a clear overview of the content described in this section.

5. Assessment of meteorological fields [51]–[53].

Overall, this work shows two cases to evaluate the performance of the WRF modelling infrastructure in and around urban environments. Here, the model performance is assessed from three aspects, namely the assessments of meteorological and concentration fields and discussions based on Differential Column Method (DCM). They are presented in detail in Sects. 5, 6, and 7, respectively.

The first case is an initial model test designed for the city of Berlin (Sect. 5.1.1), validated by comparing the modelled results to the observed values provided by a measurement campaign in Berlin in 2015 (Sect. 5.2.1). This case study presents the potential of a WRF modelling infrastructure to reproduce the column-average abundances of CO_2 and CH_4 in and around Berlin. For another case, a simulation of Munich (Sect. 5.1.2) is carried out for a period of one month based on an updated version of the modeling infrastructure described in Sect. 4.7.

This section focuses on the evaluation of modelled meteorological fields. Firstly, the model configurations (Sect. 5.1) and the details of the measurement campaigns for Berlin and Munich (Sect. 5.2) are described. The comparisons of wind fields and the behavior of other meteorological fields are then discussed in Sects. 5.3 & 5.4.

5.1. Model configurations

5.1.1. Berlin

For the Berlin study, WRF Version 3.2 coupled with GHG modules is used to quantify the uptake and emission of atmospheric GHGs around Berlin at a

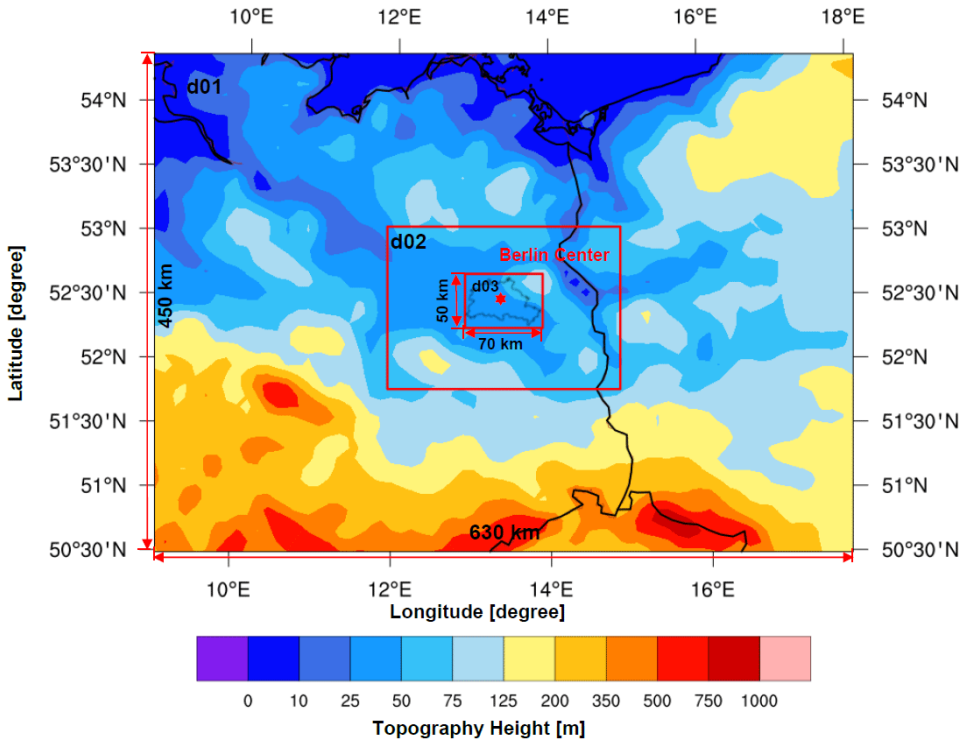


Figure 5.1.: The topography map for the three domains in our study. The domain D03 is centered over Berlin, at 13.383 °N, 52.517 °E, and is marked with a red star. The boundary of Berlin from GADM (available at `https://gadm.org/` (last access: 11 Dec. 2021); © GADM maps and data) is depicted in the innermost domain.

high resolution of 1 km. As shown in Fig. 5.1, three domains are set up here, whose dimensions are 70 × 50 horizontal grid points with a spacing of 9 km for the coarsest domain (D01), 3 km for the middle domain (D02) and 1 km for the innermost domain (D03). WRF uses a terrain-following hydrostatic pressure vertical coordinate [112].

In this Berlin case, 26 vertical levels are defined from the surface up to 50 hPa, 14 of which are in the lowest 2 km of the atmosphere. The innermost domain, D03, envelops all five measurement sites (described in Sect. 5.2.2) to assess the simulation by comparing with the measured data. Berlin lies in the North European Plain on flat land (crossed by northward-flowing watercourses), which avoids the vertical interpolation problems caused by topography differences (Fig. 5.1). The LCC projection is selected as a map projection. The simulated time span is from 18:00 UTC on 30 June to 00:00 UTC on 11 July in 2014.

The meteorological initial conditions and lateral boundary conditions were pro-

vided by GFS (Sect. 4.2.1).The initial and lateral boundary conditions for the modelled concentration fields were implemented using the CAMS (Sect. 4.2.3). The biogenic CO_2 emission was calculated online using VPRM (Sect. 4.3). The external dataset EDGAR V.4.1 was used as the input of human fluxes (Sect. 4.2.2).

5.1.2. Munich

WRF Version 3.9.1.1 is used to simulate the transport processes of atmospheric GHGs around Munich at a resolution up to 400 m. The model is configured in a three-domain nested configuration, with horizontal resolutions of 10 km for the outermost domain (D01), 2 km for the intermediate domain (D02) and 400 m for the innermost domain (D03), as illustrated in Fig. 5.2. The spatial grids are assigned using the LCC projection.

The simulations are carried out with model integration time steps of 30 seconds, 6 seconds 1.2 seconds for each domain, with model outputs saved at time intervals of 3 hours, 1 hour, and 15 minutes, respectively. 46 vertical levels are defined from the surface up to 50 hPa, 21 of which are in the lowest 1 km of the atmosphere.

To better capture the urban landscape features and improve the urban model performance ([69]–[71]), extra morphological urban parameters for the innermost domain (D03, area of Munich, Fig. 5.2) are provided, which enables the model to use the urban canopy multi-layer scheme in WRF [69].

The realistic meteorological driving data was taken from ERA5 [78] (Sect. 4.2.1). The initial and lateral boundary conditions in the simulated background concentration fields of CO_2 and CH_4 were taken from IFS Cycle 45r1 (Sect. 4.2.3). TNO_GHGco_v1.1 was used to initialize anthropogenic fluxes as tagged tracers (Sect. 4.2.2). Biogenic CO_2 fluxes were implemented online utilizing VPRM (Sect. 4.3).

CH_4 fluxes from wetlands were estimated using the Kaplan model [113], which is run online within WRF-Chem. This model calculates CH_4 emissions from anaerobic microbial production in wetlands as a fraction of heterotrophic respiration [114]. The flux estimates depend on the modelled soil moisture, soil temperature and the carbon pool from the Lund-Potsdam-Jena model, which is used for classifying the wetland fractions in the domain [115].

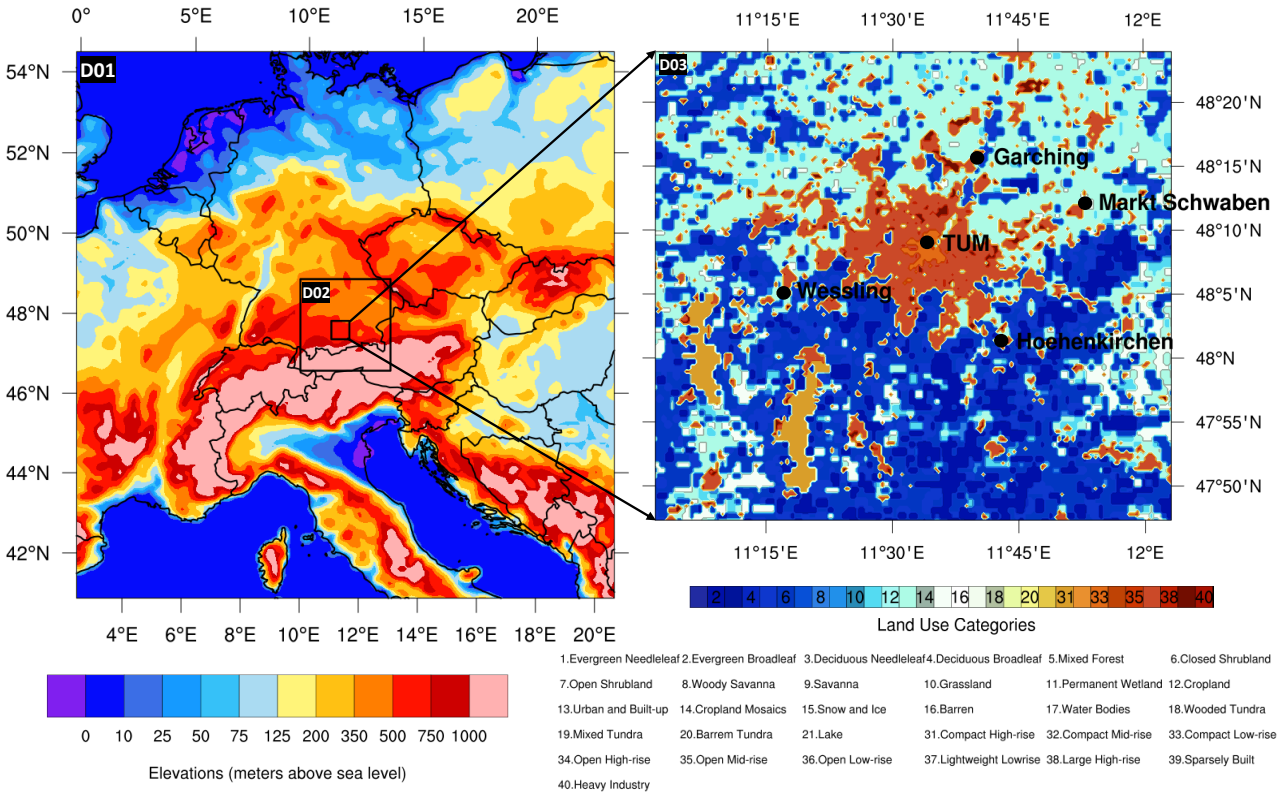


Figure 5.2.: Topography map for the entire domain area (left panel). The right panel shows the land use classification in D03, including the 21 IGBP-modified MODIS land-cover types (from 1 to 21, as illustrated in the color bar and labels), and 10 LCZ land-cover categories defined for the urban areas of D03 (from 30 to 40, as illustrated in the color bar and labels). The five measurement sites in the MUCCnet campaign are marked as black dots on the right panel.

5.2. Measurement campaign/network

The key instruments used to assess the model performance associated with the concentration fields are EM27/SUN Fourier Transform Spectrometers (FTS; Bruker Optics). By using the sun as a light source, an FTS EM27/SUN can measure the column-integrated amount of a tracer through the atmospheric column with excellent precision, yielding the column-averaged dry-air mole fractions (DMFs) of the target gases [116]. The measured DMFs of CO_2 and CH_4 are denoted by XCO_2 and XCH_4. Detailed information on the EM27/SUN instrument can be found, e.g., in Gisi et al. (2012) [117], Hedelius et al. (2016) [116], Chen et al. (2016) [39], Hase et al. (2016) [118] and Frey et al. (2019) [119].

5.2.1. Berlin measurement campaign

The measurement campaign in Berlin used to compare with the WRF modelled values, was performed from 23 June to 11 July, 2014 using five spectrometers [43]. It helps both test the precision of the modelling infrastructure (Sect. 4) and verify DCM as the initial attempt of the model analytic method (Sect. 7).

In this Berlin measurement campaign [43], these five portable EM27/SUN were used for atmospheric measurements based on solar absorption spectroscopy. Five sampling stations around Berlin were set up, four of which (Mahlsdorf, Heiligensee, Lindenberg and Lichtenrade) were roughly situated along a circle with a radius of 12 km around the center of Berlin. Another sampling site was closer to the city center and located inside the Berlin motorway ring at Charlottenburg (Fig. 5.3). Detailed information on this measurement campaign is given in Hase et al. (2015) [43].

5.2.2. Munich measurement network

The WRF modelling infrastructure for Munich is designed to study GHG concentrations in connection with the Munich Urban Carbon Column network (MUCCnet [110]). Five compact Fourier-transform infrared Spectrometers (FTIR; EM27/SUN) have been deployed in MUCCnet, four of which were located around Munich at a radius of 20 km for the 2018 period analysed in this study (this was later changed to 10 km in 2019). The fifth instrument has been set up

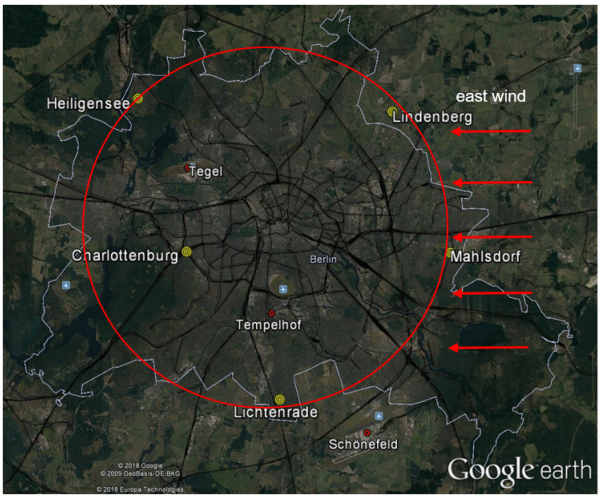


Figure 5.3.: Detailed locations of the five sampling sites. The five yellow circles stand for the five sampling sites, four of which (Mahlsdorf, Heiligensee, Lindenberg and Lichtenrade) were roughly situated along a circle with a radius of 12 km around the center of Berlin, marked as the black circle. The innermost domain of the WRF-GHG model contains all five measurement sites. The three wind measurement sites are marked by red circles. Map provided by © Google Earth, © GeoBasis DE/BKG and © Europa Technologies.

close to the center of Munich, at the TUM campus (see Fig. 5.2). All five total column measurement sites from MUCCnet are located inside D03.

In MUCCnet, the recorded interferograms are automatically transformed to spectra, converted to abundances of CO_2 and CH_4 in a column between the instrument and the end of the atmosphere in direction towards the sun, and further uploaded to the official website of MUCCnet. All five instruments are automatically operated and controlled using our universal enclosure systems and two software programs [50].

For the Munich case, simulations are compared to measurements collected during a campaign that was carried out from 1 to 30 August, 2018. Table 5.1 shows relevant parameters for assessing the measurement performance during that period, including the number of observations per day for each site and the ground-level wind information for each day, i.e., the daily mean of WS30 & the approximate change in WD30 during the day, provided by the LMU station. The criteria for this assessment in the table are based on Table 1 of a study for the Berlin region [43] and Table 2 of a study for Paris [48].

Table 5.1.: Summary of measurement performance from 1 to 30, August in 2018 at five measurement sites, including the number of measurement points for each site, overall data coverage for each measurement date (with the classifications from poor to excellent: +, ++, +++, ++++) based on the available observations, averaged wind speeds during the day time and wind directions at the ground level obtained from the LMU stations [43], [48].

Date	Quality	Number of Observations					Wind Speed	Wind Direction
		Garching (North)	TUM (Center)	Höhenkirchen (South)	Markt Schwaben (East)	Weßling (West)		
20180801 (Wed)	+	15	127	0	0	0	2.56	W-N-E
20180802 (Thu)	+	0	131	0	0	35	2.92	W-N-E
20180803 (Fri)	+	28	137	82	0	50	2.70	W-N-E
20180804 (Sat)	++++	127	131	125	125	60	2.11	N-E
20180805 (Sun)	++++	110	131	82	99	68	2.83	NW-NE
20180806 (Mon)	+++	114	126	124	126	88	3.19	NE
20180807 (Tue)	+	0	47	0	53	23	2.10	W-N-SE
20180808 (Wed)	++	89	92	76	106	37	3.30	W-N
20180809 (Thu)	++++	112	97	106	106	130	2.49	SE-N
20180810 (Fri)	+	0	0	0	0	0	2.19	N-W-S
20180811 (Sat)	+++	100	93	91	114	99	3.29	W
20180812 (Sun)	+	130	0	135	135	136	2.68	E-N
20180813 (Mon)	+	73	81	72	93	68	3.10	W-N-W
20180814 (Tue)	+	0	0	0	0	0	3.16	NW
20180815 (Wed)	+	57	94	0	89	46	2.26	SW-N-NE
20180816 (Thu)	++++	110	133	105	133	133	2.97	E-NE
20180817 (Fri)	+++	102	130	131	132	97	1.87	W-N-E
20180818 (Sat)	+++	75	69	100	98	71	2.56	NW
20180819 (Sun)	++++	127	130	129	132	127	2.03	W-N
20180820 (Mon)	++++	126	127	126	131	112	2.25	W-N
20180821 (Tue)	+++	109	131	115	130	109	2.44	NW-N-E
20180822 (Wed)	++++	129	130	128	130	129	2.26	N-E
20180823 (Thu)	++	60	102	75	72	72	3.14	W-N-S
20180824 (Fri)	+	0	0	0	0	0	3.79	W
20180825 (Sat)	+	0	0	0	0	0	3.16	W
20180826 (Sun)	+	23	83	28	0	0	2.28	W-N-E
20180827 (Mon)	+++	81	75	98	101	61	2.95	W
20180828 (Tue)	+++	66	72	104	99	75	2.18	NE
20180829 (Wed)	+++	79	86	55	86	94	1.71	SW-N-E
20180830 (Thu)	+	0	0	0	0	0	2.52	W

Specifically, all the measurement sites should work, otherwise the overall quality is set to be '+'. The overall quality levels are then determined based on the numbers of the available observations during one day: daily temporal coverage lower than 25% '+', from 25% to 50% '++', from 50% to 75% '+++', and from 75% to 100% '++++'. The numbers of the available observations at more than three sites during one day should be higher than the criteria. After assessing the measurement performance of this campaign, fifteen days are selected (with a quality level better than '++'; cf. Table 5.1) to make the model-measurement comparison. Details of the campaign and side-by-side calibrations are discussed in Sect. 4.1 and 5 of Detrich et al. (2020) [110]. There are in total 15 days with good measurement conditions that can be used for the model-measurement comparison: 4-6, 9, 11, 16-22 and 27-29 August, 2018.

Over the entire campaign period in 2018, the mean of the measured XCO_2 for all five sites is 404.4 ppm with a standard deviation of 1.2 ppm, ranging from 400.8 ppm to 408.1 ppm. For XCH_4, the measurements range from 1840.5 ppb to 1896.0 ppb, with a mean of 1865.5 ppb and a standard deviation of 9.1 ppb. Since the operation of the instruments is strongly influenced by weather conditions, such as sunlight [48], the spatial and temporal measurement coverage for some days (e.g., 1-3 August) is limited (see Table 5.1).

5.3. Comparisons for Berlin

Winds have a strong impact on the vertical mixing of GHGs and a direct influence on their atmospheric transport patterns. Hence, firstly the wind speeds and wind directions obtained from WRF are compared to the measurements, such that deviations between the simulated and measured wind fields are assessed in this section.

The wind measurements are not exactly co-located with the spectrometers mentioned in Sect. 5.2.1, but rather are located at three sampling sites (Tegel, Schönefeld and Tempelhof, respectively) and measure at a height of 10 m above the ground. The simulated wind speed at 10 m (ws_{10m}) and wind direction at 10 m (wd_{10m}) are calculated following the equations,

$$\begin{aligned} ws_{\mathrm{10m}} &= \sqrt{u_{\mathrm{10m}}^2 + v_{\mathrm{10m}}^2} \\ wd_{\mathrm{10m}} &= \arctan \frac{v_{\mathrm{10m}}}{u_{\mathrm{10m}}} \end{aligned} \tag{5.1}$$

where $u_{10\mathrm{m}}$ and $v_{10\mathrm{m}}$ are the components of the horizontal wind, towards the east and north respectively, which can be obtained from the WRF output files.

Figure 5.4 shows the comparisons of wind speeds (Fig. 5.4(a)) and wind directions (Fig. 5.4(b)) between simulations and observations at 10 m from 1 to 10 July and the model–measurement differences. EM27/SUN only operates in the daytime when there is sufficient sunlight [48], [117], [120], and the instrumental working periods are marked by gray shaded boxes in Fig. 5.4.

The measured (dashed lines) and simulated (solid) wind speeds (Fig. 5.4(a)) at 10 m show similar trends and demonstrate relatively good agreement over the 10-day time series, with a root-mean-square error (RMSE) of 0.93 m/s. Large uncertainties in wind speeds are found to appear always with the lower wind speeds, mostly at night. In terms of wind directions at 10 m, the simulated wind directions show similar but slightly underestimated fluctuations (Fig. 5.4(b)), which result in an RMSE of 60.83 °. Larger uncertainties in wind directions always exist during the low wind speed periods (Fig. 5.4(a,b)). During the instrumental working period (within the daytime), the simulations fit better with the measurements with relatively lower RMSEs of 0.69 m/s for wind speeds and 41.48 ° for wind directions.

It has been found that the measured wind fields (both wind speeds and wind directions) have more fluctuations compared to the simulations. This could be caused by really fast wind changes which the model, simulating a somewhat idealized environment, is not able to capture. To be specific, local turbulence given by urban canopy, buildings, etc., is not represented well in the model.

Furthermore, a brief analysis does here regarding the vertical distribution of wind fields. As seen from Fig. 5.5, above approximately 300 hPa, the lower the pressure is, the larger the wind speed is. While the wind speed decrease sharply with the increase of the height (below 300 hPa).

Overall, wind directions shift from northeast to northwest from morning to evening. As depicted in the trend of the vertical distribution of wind directions in Fig. 5.5, the surface wind shows a prevailing pattern towards the east with the increase of the altitude in the troposphere. Then above the troposphere (see Sect. 2.2), the prevailing pattern turns gradually to be in the opposite direction (the east).

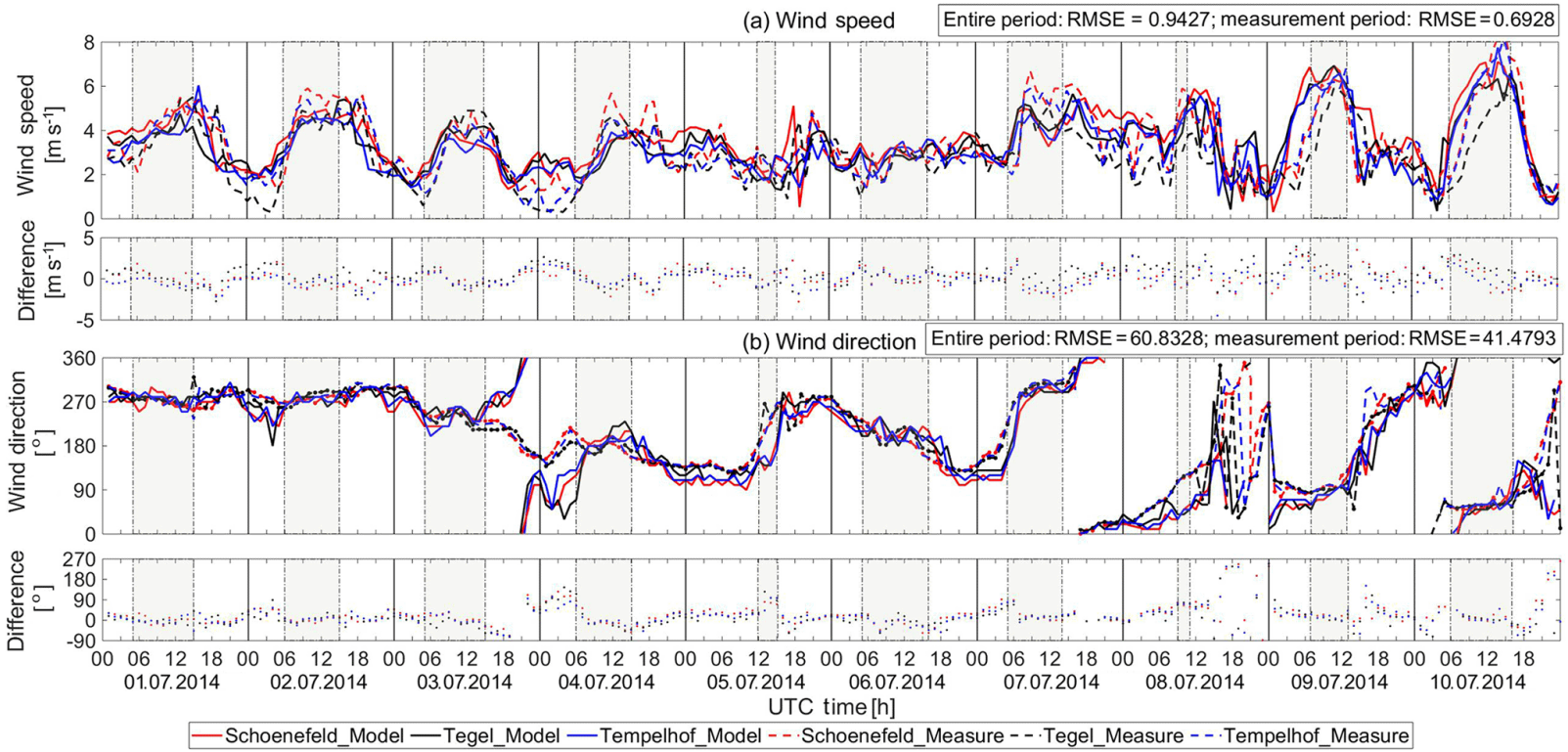


Figure 5.4.: Variation and differences between simulated and measured wind fields for (a) wind speeds and (b) wind directions from 1 to 5 July 2014 at the three measurement sites, Schönefeld (red lines), Tegel (black) and Tempelhof (blue) in Berlin. The solid lines represent the simulated wind fields provided by WRF and the dashed lines depict the measured wind fields. The differences in (a,b) are simulations minus measurements. FTS measurement time periods on each date are marked by gray shaded areas.

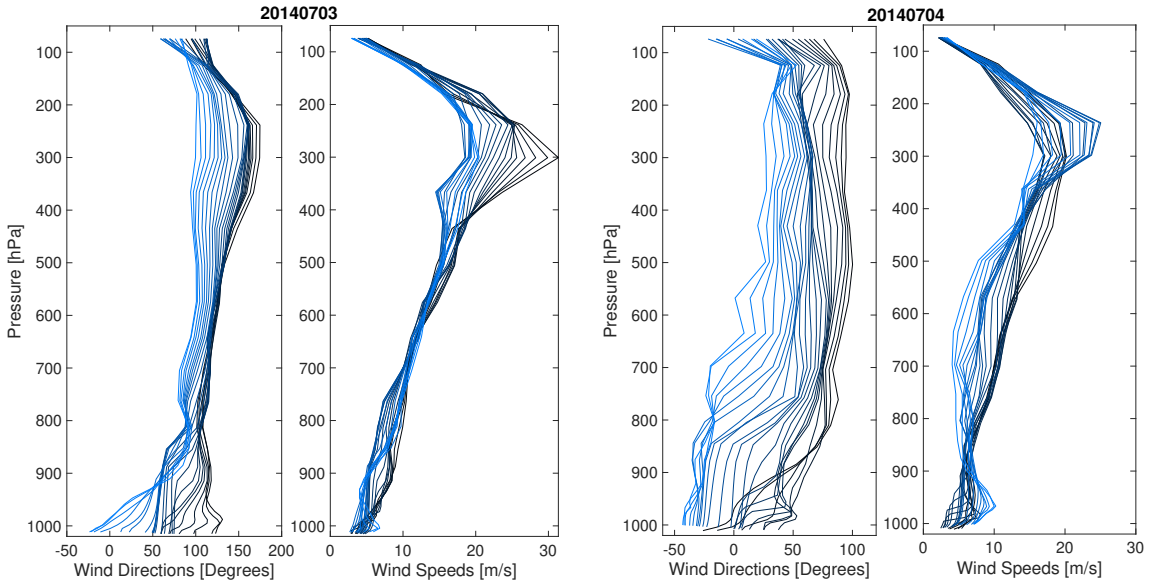


Figure 5.5.: The vertical distribution of wind fields (wind speeds and wind directions) on 3rd July (a,b) and 4th July (c,d) in Tegel. The colors from black to blue represent the time from morning to evening.

Apart from modelled wind profile, a brief analysis on PBL is presented here, as shown in Fig. 5.6. The vertical layers in WRF are defined following the pressure definition and the top-layer pressure (50 hPa) is already beyond the tropopause. The PBL normally varies from a few hundred meters (morning and night) to over one thousand meters (daytime, pick at noon), which is situated from the second vertical layer (morning and night) to the 13th layer (noon) of the model domain, respectively (see the left column of Fig. 5.6 of this response).

Figure 5.6 shows the wind fields within PBL on 3rd July. The PBL is situated either in the second or third layer (morning and night) or in the 13th layer (noon) in our domain. The wind speeds and wind directions closer to the PBL are generally higher than the surface wind fields. The WRF outputs do provide these simulated PBL values but there is a lack of observations to assess these simulated values. Avolio et al. (2017) [121] did some sensitivity analysis of the PBL in the WRF model with a case study of southern Italy, and concluded that the simulated PBL heights are mostly overestimated.

Due to the limitations of the available variables provided by this measurement campaign, no further model-measurement comparisons are made for other meteorological fields. For Munich, more variables are to be considered to assess

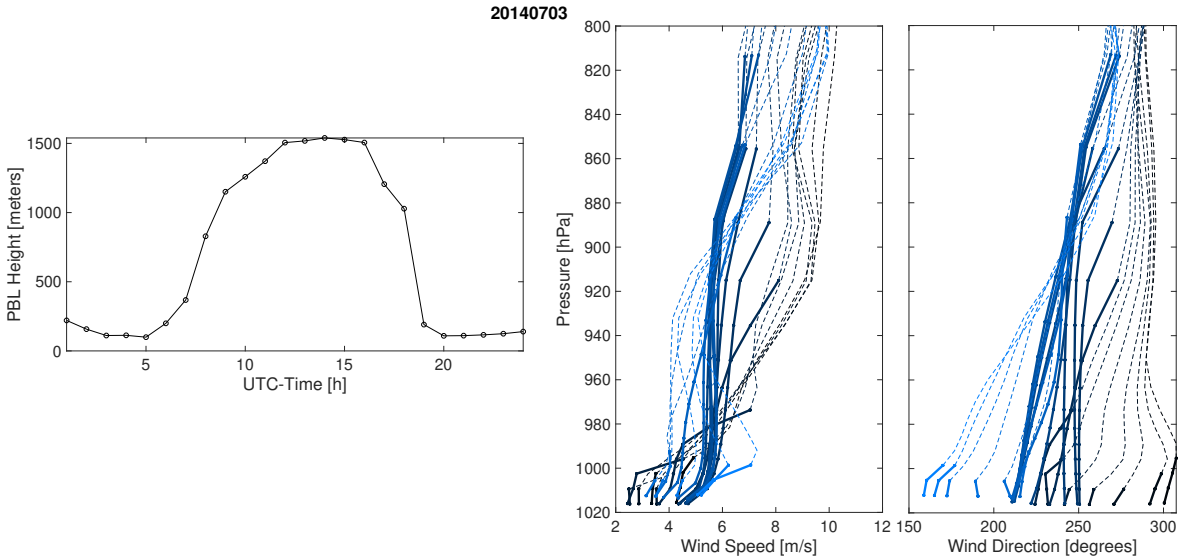


Figure 5.6.: Variations of the simulated PBL height (left side), and the wind speeds (middle) and wind directions (right) within PBL on 3 July at Tegel. The colors from black to blue represent the time from morning to evening. The bold solid lines represent the values within PBL

the model performance.

5.4. Comparisons for Munich

For the Munich case, some key meteorological parameters are compared to measured values provided by two local weather stations, to evaluate the modelled meteorological variables used for transporting the fluxes.

The first station is located at the Meteorological Institute of the Ludwig Maximilian University of Munich (LMU; latitude: 48.15 °, longitude: 11.57 °, altitude: 561 m), close to the center of Munich. This station can provide time series of meteorological variables second by second. The model here is compared to five meteorological variables measured at LMU: the temperatures at heights of 2 m and 30 m above the ground (T2 & T30), the precipitation, and the wind speed and direction at 30 m above the ground (WS30 & WD30).

The other station considered is run by the German Meteorological Service (Deutscher Wetterdienst in German; DWD). This automatic weather station is located at the Munich airport with the station ID of 01262 (latitude: 48.35 °,

longitude: 11.81 °, altitude: 446 m). The following variables measured at the airport is used for comparisons with the model: T2, precipitation, relative humidity, air pressure and the wind speed and direction at 10 m above the ground (WS10 & WD10 respectively).

5.4.1. Wind fields

As one of the key drivers for the transport of trace gases in the model, the simulated wind field directly impacts the transport patterns of the tracers. Thus, it is particularly important to assess the model performance with regards to the wind field.

Here, the measured WS30 & WD30 at the LMU station, and WS10 & WD10 at the DWD station are employed. The DWD station measures the winds every second, while the wind data given by DWD are recorded as 10-minute means. A cut-off wind speed threshold (0.5 m/s in the case) is applied to the values shown in Fig. 5.7, owing to large uncertainties in wind directions during low wind speed periods [86].

A comparison between the modelled and measured winds at the LMU station is shown in Fig. 5.7. Prevailing wind directions both in the simulations and measurements are either easterly or westerly during the daytime, while the prevailing winds at night are generally from the southwest. The measurements (panels (a.1,a.3,a.5)) show larger scatter in the wind direction over August compared to the simulations. Along the time series, the simulated (Fig. 5.7(b): red crosses) and measured (blue dots) wind speeds vary in a quite similar manner. But in general, the model overestimates wind speeds with a RMSE of 2.0 m/s and a mean absolute error (MAE) of 1.2 m/s.

Regarding a comparison of the wind directions between the model and measurements (see Fig. 5.7(c)), the model mostly follows the measured fluctuations of the wind directions (RMSE = 64.1 ° & MAE = 1.8 °) over time, but with the model performance in some periods reduced, e.g., between 24 and 26 August, when the variability of the wind direction is remarkably lower in the model.

Figure 5.8 shows a comparison between the modelled and the measured values obtained from the DWD station. As depicted in the wind roses of Fig. 5.8(a), the prevailing measured surface wind blows from the southwest, while the modelled directions show more variations. Mismatches of wind directions mainly

appear during daytime, but the overall trend can be reproduced by the model (red crosses in Fig. 5.8(c); RMSE = 72 ° and MAE = 17.83 °). In terms of wind speeds (see Fig. 5.8(b)), the model matches well with the measurements with a RMSE of 1.58 m/s and a MAE of 0.15 m/s. At nighttime, the measurement exhibits higher wind speeds under the same wind directions, compared to the model (see Fig. 5.8(a.5,a.6)).

5.4.2. Other meteorological fields

Apart from the wind fields mentioned above, other meteorological parameters are also expected to impact the behavior of the tracers in this model. Thus, similar comparisons regarding the other meteorological variables have also been performed and are presented in this section. Theses comparisons indicate that our model has the capability to provide reasonable simulated meteorological fields for driving the transport of trace gases.

The temperature at a height of two meters above the ground (T2) plays a key role in calculating biogenic fluxes in VPRM [90]. Precipitation is also an important point in assessing the model behavior, since the functionality of the instrument used for measuring the column concentrations is influenced by rainfall, as described in detail in Sect. 5.2.

Figure 5.9 shows comparisons of T2 and the precipitation between models and measurements. Over the entire simulation period, the modelled T2 (black solid lines in Fig. 5.9(a,b)) shows trends similar to the measurements (red dashed lines) from both weather stations. As illustrated in Fig. 5.9(b), a better agreement appears to exist between the modelled and those values from the LMU station, with a RMSE of 2.1 K and a MAE of -0.23 K, while the RMSE between the simulated values and those measured at DWD is 2.3 K and the simulated values are generally higher compared to the measurements, with a MAE of 1.15 K. During the daytime, the variations of the modelled and measured T2 generally match well, apart from the appearance of some peaks at noon measured by the LMU weather station (Fig. 5.9(b)). Major discrepancies of T2 between models and measurements occur during the nighttime, early morning and at specific days, e.g., 10 and 24 August.

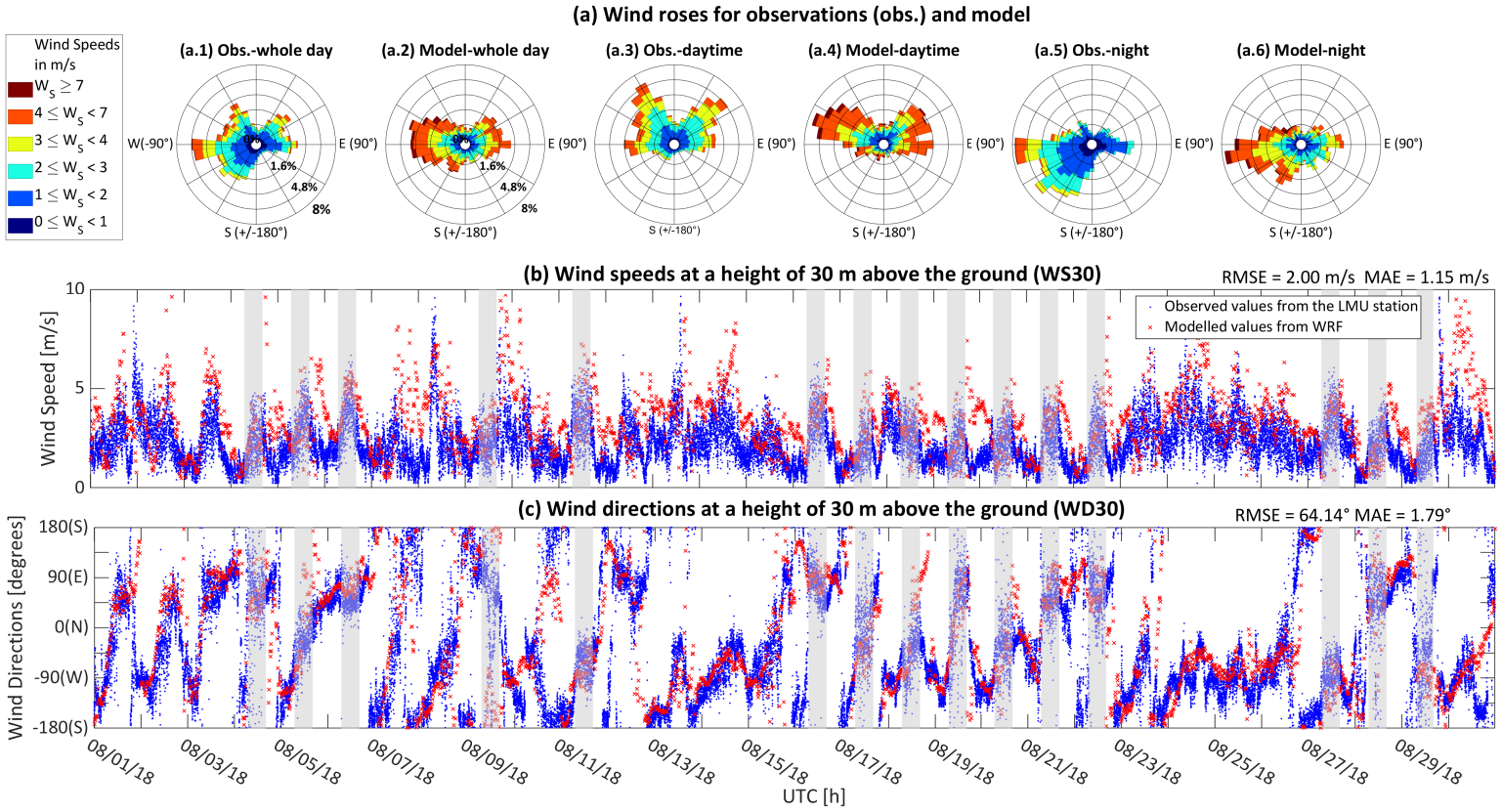


Figure 5.7.: Wind roses (a) and time series of simulated and observed wind speeds (b) and wind directions (c) at a height of 30 m above the ground at LMU. Wind roses for the measurements over August 2018 are plotted in Panels (a.1): 24 hours, (a.3): daytime from 6:00 UTC to 17:00 UTC only and (a.5): nighttime only, while Panels (a.2,a.4,a.6) represent the modelled values. Each wind rose indicates WS30, WD30 and the frequency (% scale) of wind coming from a particular direction during the targeted period. The blue dots in Panels (b,c) represent the measured values from the LMU station and the red crosses represent the simulation. The grey shaded areas mark the measurement periods used for the model-measurement comparison of column concentrations in Sect. 6.

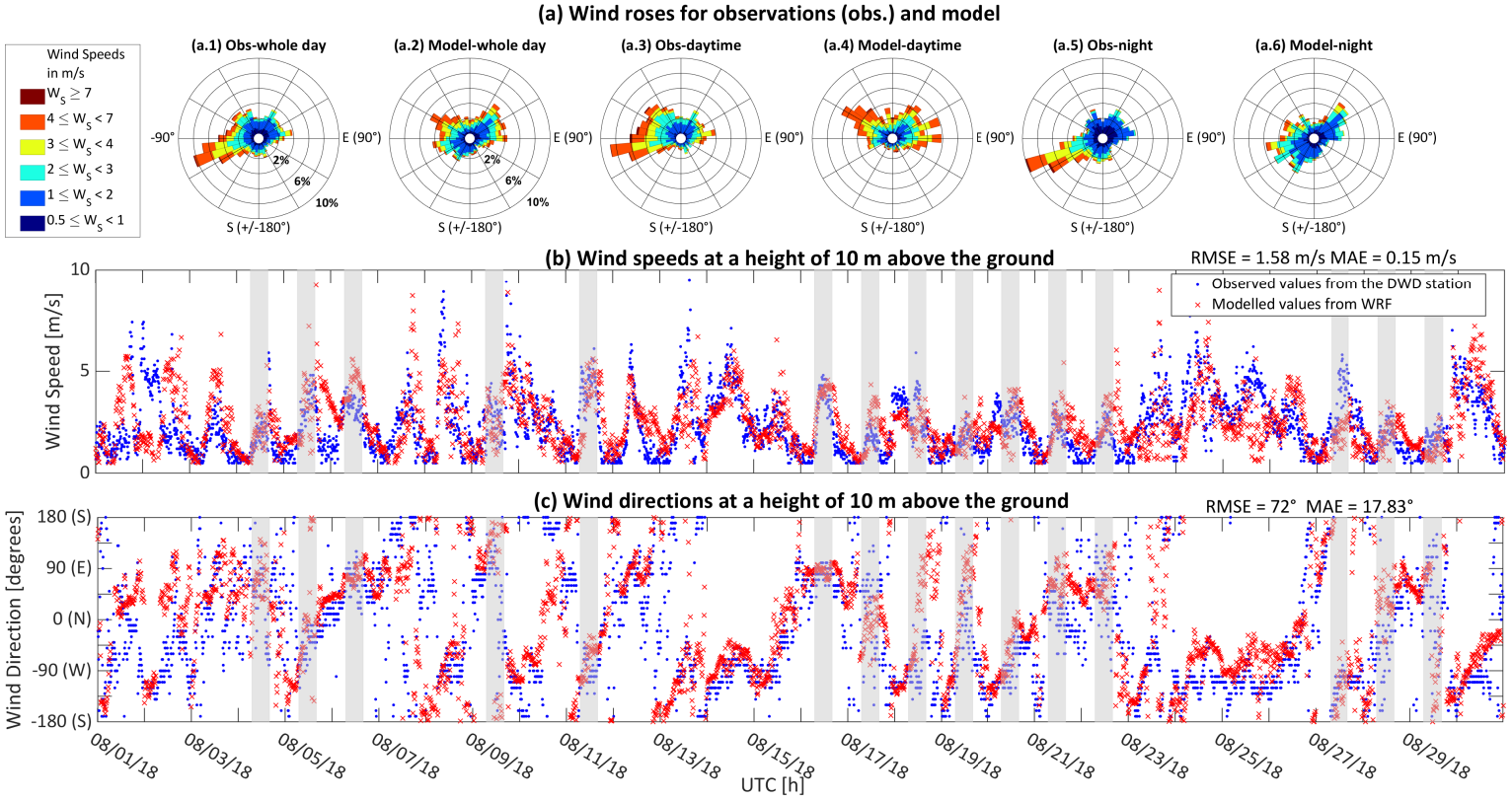


Figure 5.8.: Wind roses (a) and time series of simulated and measured wind speeds (b) and wind directions (c) at a height of 10 m above the ground. All measurements shown here have been recorded by the DWD station. Wind roses for the measurements over August 2018 are plotted in Panels (a.1): 24 hours, (a.3): daytime from 6:00 UTC to 17:00 UTC only and (a.5): nighttime only, while Panels (a.2,a.4,a.6) represent the models. Each wind rose indicates WS10, WD10 and the frequency of wind coming from a particular direction during the targeted period. The blue dots in Panels (b,c) represent the measured values from the DWD station and the red crosses represent the simulation. The grey shaded areas mark the measurement periods used for the model-measurement comparison of column concentrations.

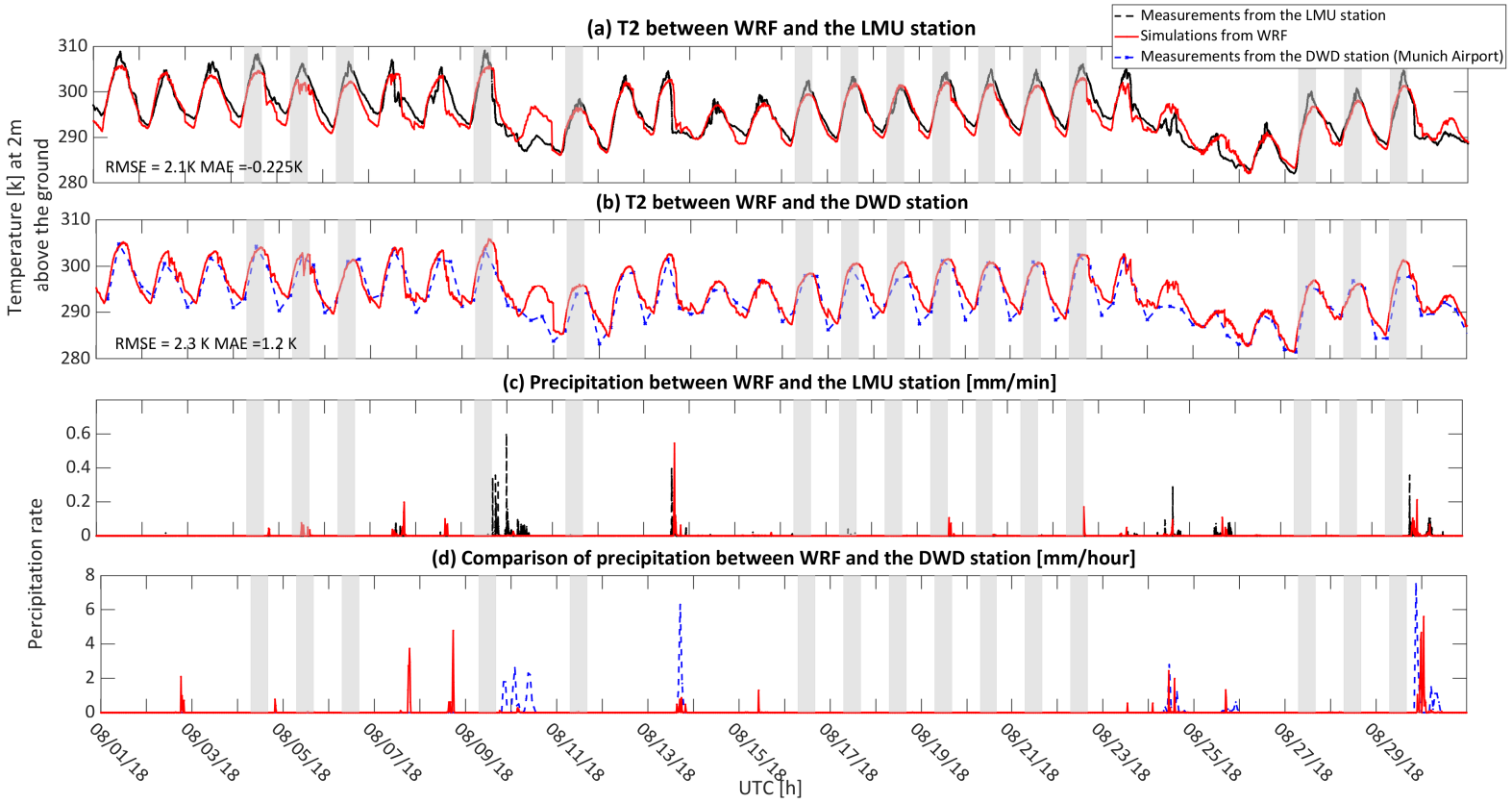


Figure 5.9.: Comparison of temperatures at a height of 2 meters above the ground (T2; upper two panels) and precipitation values (lower two panels) between the model and two stations from 1 to 30 August, 2018: Panels (a,c) display data from DWD and Panels (b,d) show values from the LMU station. In all subplots, the simulations are plotted using the red solid lines while the measurements are represented by the blue dashed lines. The grey shaded areas mark the measurement periods used for comparing to the simulations in Sect. 6.4

One of the potential reasons for the discrepancies between models and measurements on these two specific days is that the mesoscale model can not reproduce the precipitation on these two days well, especially for 10 August. The missing or weak precipitation potentially led to unrealistic temperatures. As evident in Fig. 5.9(c,d), most of the rainfalls in August, 2018 can be replicated by the model, but the magnitude of the precipitation is not always well estimated. The disagreement in precipitation between the model and measurements can be due to several potential error sources. For instance, the spatial resolution of the mesoscale model may be not sufficient to fully reproduce complex flow features within a micro-scale or to resolve orographic uplift, but these effects can have an impact on the measurements [122].

To sum up, for both Berlin and Munich cases, the WRF modelled meteorological fields used for driving trace gases, are comparable and reasonable well. The model is always able to reproduce the prevailing winds well, while slightly biases in wind speeds compared to the measurements. Most of the measured variations in wind directions along the time series can be captured by the model.

6. Assessment of concentration fields [51]–[53]

As described in Sect. 5.2, the precision of simulated concentration fields is assessed through comparing them to observations carried out using the high-resolution portable EM27/SUNs. Therefore, such model-measurement comparisons regarding concentration fields are discussed in this section, including the calculation of column concentrations, the model-measurement comparisons of concentration fields, and tracer analysis for both cases.

6.1. Model systematic errors

In the passive tracer transport simulation, the total concentration of each GHG is represented as a separate tracer, giving redundant information (with respect to the sum of all tracers for each GHG) and allowing for consistency checks. A variety of flux models and emission inventories implemented in the modules of WRF-Chem are used for the estimation of GHG fluxes. The flux values from external emission inventories are gridded and absorbed into the model.

In the transport process, the relationship among the changes in concentrations from different emission tracers, the total and background concentrations (Eq. 6.1) should then be satisfied, ideally with ΔCO_2 and ΔCH_4 computational errors during the simulation process being zero. Nonzero values of ΔCO_2 and ΔCH_4 reflect the limited precision of the tracer transport calculation in the GHG related modules of WRF-Chem.

$$\begin{aligned} CO_{2,\text{total}} &= CO_{2,\text{bgd}} + CO_{2,\text{anthro}} + CO_{2,\text{VPRM}} + \Delta CO_2 \\ CH_{4,\text{total}} &= CH_{4,\text{bgd}} + CH_{4,\text{anthro}} + CH_{4,\text{soil}} + \Delta CH_4 \end{aligned} \tag{6.1}$$

Taking the initial test for Berlin as a sample, Figure 6.1 thus shows the mean values (solid lines) and the 95 % confidence intervals of ΔCO_2 and ΔCH_4. As

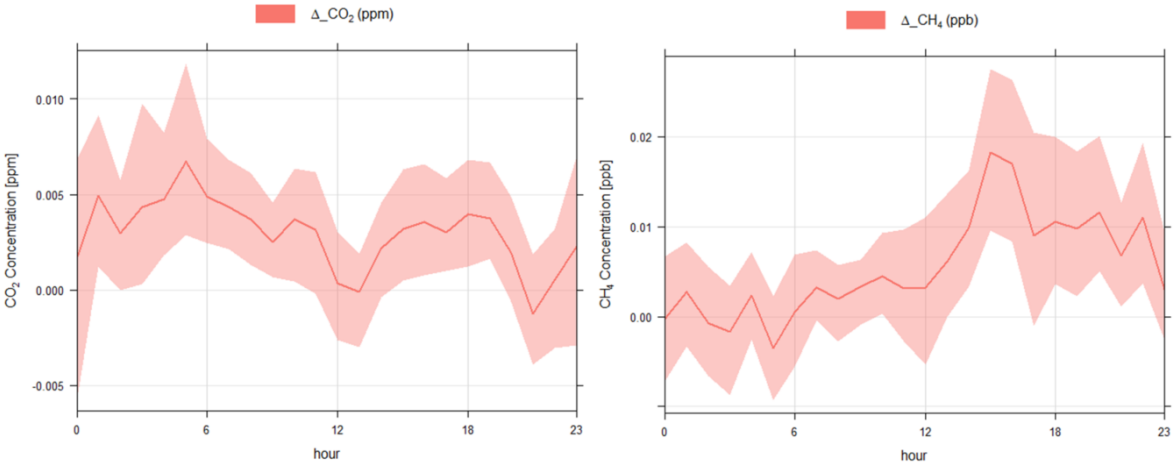


Figure 6.1.: The mean values (solid lines) and the 95 % confidence intervals of the computational error (a) ΔCO_2 and (b) ΔCH_4. ΔCO_2 and ΔCH_4 are calculated using Eq. 6.1.

depicted in Fig. 6.1, ΔCO_2 ranges from -0.005 to 0.01 ppm, while ΔCH_4 is in the range of -0.01 to 0.02 ppb. Divided by typical absolute values of the concentrations from different flux processes for XCO_2 (around 1 ppm) and XCH_4 (around $2-3$ ppb) depicted in Fig. 6.6, the relative computational error is found to be approx. 0.1 % for both CO_2 and CH_4. These tiny computational errors can be caused by the slight non-linearity of the advection scheme used in WRF, which makes the sum of the concentrations in CO_2 and CH_4 from all individual flux tracers not exactly equal to the concentration from the sum tracer, representing the total sum of all fluxes related to different processes.

6.2. Pressure-weighed column-averaged concentrations

As described in Sect. 5.2, EM27/SUN can measure the column-integrated amount of a tracer through the atmospheric column with excellent precision, yielding the column-averaged DMFs of the target gases [39], [116]. The measured DMFs of CO_2 and CH_4 are denoted by XCO_2 and XCH_4.

The characteristics of the instruments needs to be considered and used in a more accurate way, when comparing the modeled values to the measurements

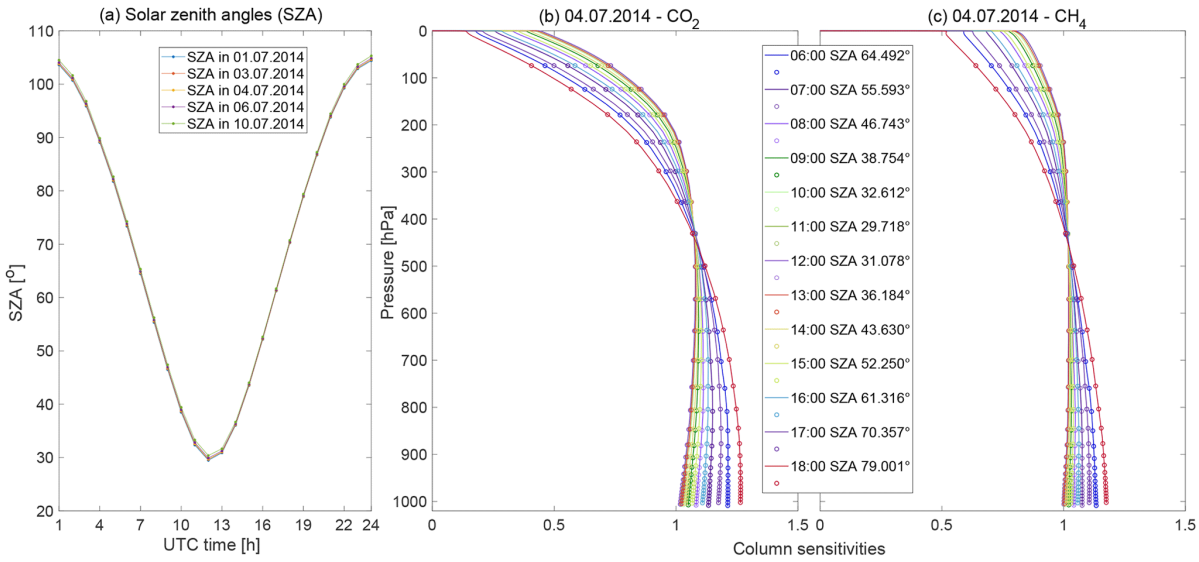


Figure 6.2.: Daily variations in SZA for five simulation dates (1, 3, 4, 6 and 10 July), and the vertical distributions of column sensitivities for (b) CO_2 and (c) CH_4 on 4 July. In (b,c), the solid lines represent the derived column sensitivities for EM27/SUN under different SZAs, and the circles stand for the values on model pressure levels.

here. The EM27/SUN records the spectra along a slant column from the Sun to the ground, instead of a vertical column perpendicular to the ground. Simulated concentration fields of CO_2 and CH_4 used for the model-measurement comparison in this study must therefore be aggregated along the slant columns from the ground to the Sun.

Furthermore, when comparing remote sensing observations to model data (or also databases from different remote sensing instruments to one another), limitations of the instruments in reconstructing the actual atmospheric state need to be taken into account. In general, this requires the *a-priori* profile that is used for the retrieval and the averaging kernel (AK) matrix, which specifies the loss of vertical resolution (fine vertical details of the actual trace gas profile cannot be resolved) and limited sensitivity (e.g., Rodgers and Connor, 2003 [123]). In the case of EM27/SUN, the spectrometers used in the network offer only a low spectral resolution of 0.5 cm^{-1}. Therefore, performing a simple least-squares fit by scaling retrieval of the a priori profile is generally appropriate. In this case, there is no need to specify a full averaging kernel matrix; instead, the specification of a total column sensitivity is sufficient. The total column sensitivity is a vector (being a function of altitude), which specifies

to which degree an excess partial column superimposed on the actual profile at a certain input altitude is reflected in the retrieved total column amount. This sensitivity vector is a function of a solar zenith angle (SZA; and ground pressure), mainly due to the fact that the observed signal levels in different channels building the spectral scene used for the retrieval are shaped by a mixture of weaker and stronger absorption. (If all spectral lines in the spectral scene are optically thin and too narrow to be resolved by the spectral measurement, the sensitivity would approach unity throughout.) In addition, constant a priori profile shapes are always used in the retrievals of EM27/SUN measurements [43].

Taking the sample date 4 July (the best-quality day in terms of measurements) in the Berlin case as an example, the pressure-dependent column sensitivities for CO_2 (Fig. 6.2(b)) and CH_4 (Fig. 6.2(c)) are derived from measurements performed in Lindenberg [43]. As can seen from Fig. 6.2, the shape and values of the column sensitivities from Lindenberg closely resemble the results of Hedelius et al. (2016) [116] in Pasadena. Given the limited simulation period in July (10 days) of the measurement campaign in Berlin, the daily variations of SZAs for these ten days are mostly overlapping (Fig. 6.2(a)). That is, the SZAs are almost identical for each day in the study (at each hour), rendering the shape of column sensitivities (at a specific hour of the day) practically independent of the measurement date. The column sensitivities for 4 July (Fig. 6.2(b,c)) are taken as a basis for the smoothing process (see Eq. 6.2). Then the column sensitivities following the model vertical pressure axis can be derived through interpolation on the basic of the reference column sensitivities.

The smoothed concentration for a target gas G at a specific location and the l^{th} layer of the slant column from the site to the Sun ($G^{\mathrm{s}}_{\mathrm{sla},l}$) is calculated by following Eq. 6.2 (cf. [48], [86]),

$$G^{\mathrm{s}}_{\mathrm{sla},l}(x,y,t) = \mathrm{AK}_{G,l}(t) \times G_{\mathrm{sla},l}(x,y,t) + (1 - \mathrm{AK}_{G,l}(t)) \times G_{\mathrm{pri},l}) \tag{6.2}$$

where $G_{\mathrm{sla},l}$ are modelled concentrations for trace gas G at the l^{th} vertical layer following the slant column along the line of the Sun, $\mathrm{AK}_{G,l}$ is the fitted AK of the gas G at the l^{th} vertical layer, and time t, $G_{\mathrm{pri},l}$ stands for the mixing ratio of the *a-priori* profile at the l^{th} vertical layer. The *a-priori* profiles from the Whole Atmosphere Community Climate Model (WACCM) Version 6 in 2015 for Berlin and 2018 for Munich have been used in this work.

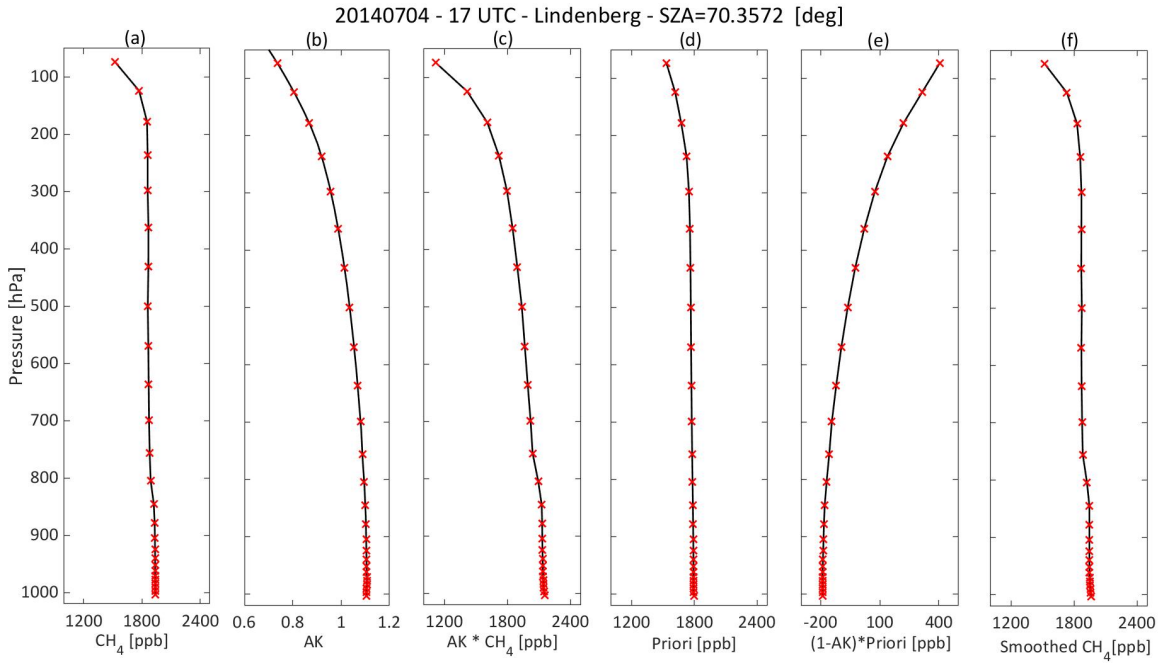


Figure 6.3.: The calculations during the smoothing process: (a) the modelled vertical concentration profile at 17:00 UTC on 4 July; (b) the column sensitivities (AK) derived from EM27/SUN and the red crosses stand for the interpolated AKs associated with the model pressure axis; (c) the vertical concentration profiles weighted by the interpolated AKs shown in (b); (d) the vertical profile of concentrations derived from the prior model (WACCM); (e) the vertical profile of the weighted (weight factors: 1-AK) prior concentrations; (f) the smoothed vertical concentration profile obtained by the sum of (c) and (e).

As seen in Fig. 6.3(b,c), the interpolated column sensitivities (the crosses) fit well with the distribution of column sensitivities with different SZAs for both CH_4 and CO_2. Since the best measurement date during this Berlin campaign (4 July) provides enough values to capture the vertical distribution features of column sensitivities accurately, the column sensitivities in this work can be directly interpolated, which is different from the Paris study [48] that calculated the column sensitivities based on formulas. Based on the smoothing process indicated in Eq. 6.2 by using the column sensitivities and the *a-priori* concentration fields, one example is depicted in Fig. 6.3 to provide a clear understanding.

Further, the modelled vertical concentration profiles are converted to pressure-weighted column-averaged concentrations. That is to say, for the trace gas G,

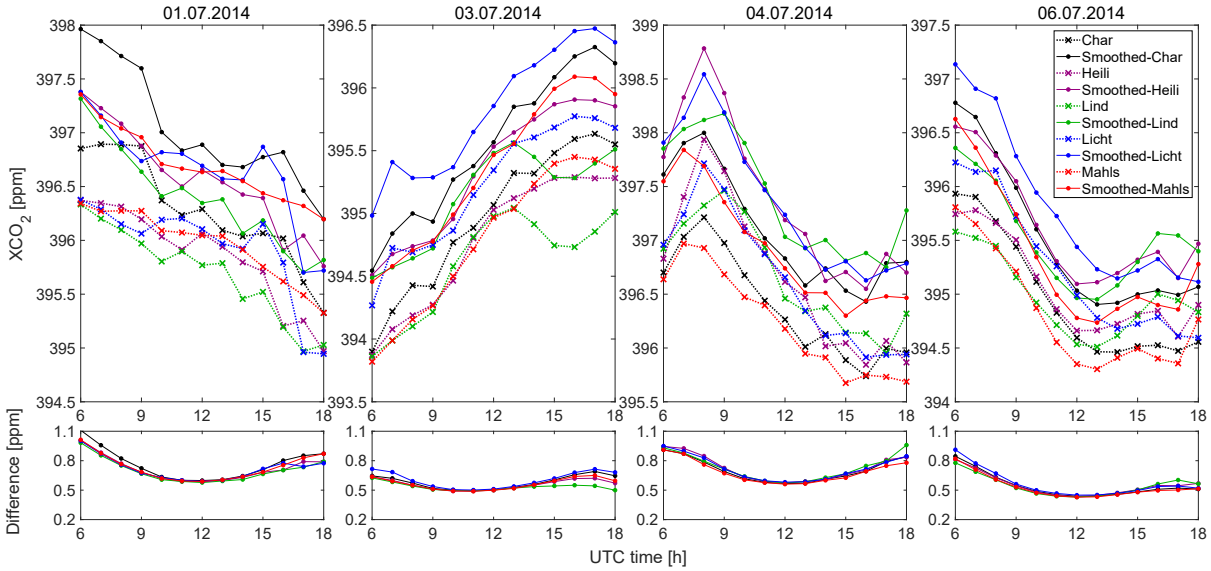


Figure 6.4.: Comparison of modelled XCO_2 with and without smoothing (using the column sensitives for EM27/SUN) for the first four simulated dates. The five colors stand for the concentrations from five sample sites. Dotted lines with the crosses represent the XCO_2 without smoothing, while solid lines with the circles stand for the smoothed values.

the simulated column-concentration at a specific location and time ($XG(x, y, t)$) can be calculated as follows:

$$XG(x, y, t) = \sum_{l=1}^{L_{unsta}} [w_l(x, y, t) \times G_l(x, y, t)] \tag{6.3}$$

where G_l stands for the simulated mole fraction at the location (x, y) and time t in the l^{th} unstaggered level of WRF (i.e., the middle of the adjacent model levels [124]). L_{unsta} is the total number of the unstaggered levels (i.e., 26 in Berlin and 45 in Munich), and w_l means the weight of the l^{th} vertical layer which can be obtained as:

$$w_l(x, y, t) = \frac{\Delta P_l(x, y, t)}{P_{sfc}(x, y, t) - P_{tp}} \tag{6.4}$$

where P_{tp} is the hydrostatic pressure at the top of the model (i.e., 50 hPa) and P_{sfc} is the surface pressure. ΔP_l denotes the pressure difference between the top and the bottom of the l^{th} vertical layer.

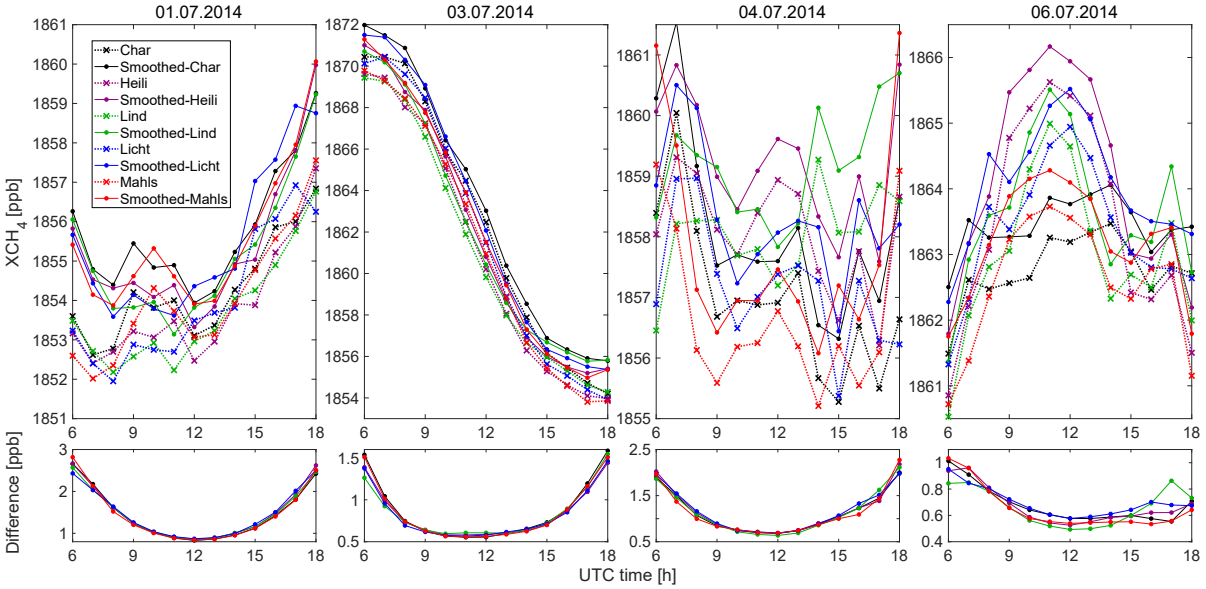


Figure 6.5.: Comparison of modelled XCH_4 with and without smoothing (using the column sensitives for EM27/SUN) for the first four simulated dates. The five colors stand for the concentrations from five sample sites. Dotted lines with the crosses represent the XCH_4 without smoothing, while solid lines with the circles stand for the smoothed values.

Figures 6.4 and 6.5 show the comparisons of the simulated XCO_2 and XCH_4 with and without smoothing for four sample dates of the Berlin measurement campaign. The simulated concentrations are only slightly enlarged after smoothing, at approximately 1–2 ppm for XCO_2 and 2 ppb for XCH_4, while the variations are mostly unchanged. Compared to the period with lower SZAs (at noon), the smoothed values in the morning and afternoon with higher SZAs hold relatively larger enlargements.

6.3. Comparisons for Berlin

For the Berlin study, in order to ensure measurement quality and enough sample points for further concentration comparisons, five measurement dates (1, 3, 4, 6 and 10 July) are selected with relatively good measurement qualities (from fair, "++", to very good, "++++") based on Hase et al. (2015) [43].

Figure 6.6(a) shows the measured and smoothed modeled variations in XCO_2 and XCH_4 for these 5 days. Compared to the measurements, the smoothed

simulated pressure-weighted column-averaged concentrations for CO_2 (XCO_2) show quite similar trends but with approximately 1–2 ppm bias, indicated by an RMSE of 1.25 ppm. The simulated XCO_2 values are overestimated for 1, 3 and 4 July, while on 6 and 10 July, the model is underestimated, which could be the result of uncertainties from the coarse anthropogenic surface emission fluxes, background concentrations from CAMS [125] and the ignorance of the influence from the line of sight of the Sun that is not included when studying the case of Berlin.

Figure 6.6(b) shows the comparison of the pressure-weighted column-averaged concentrations for CH_4 (XCH_4) between observations and smoothed simulations on the five selected dates (1, 3, 4, 6 and 10 July). It has been found that there is an approximate offset of 50–60 ppb between observations and models (RMSE is 58.11 ppb). The simulated XCH_4 is around 1860 ppb while the measured value is around 1810 ppb, which is comparable to the values (1790–1810 ppb) observed at two Total Carbon Column Observing Network (TCCON) measurement sites in June and July 2014 in Bremen in Germany [126] and Bialystok in Poland [127]. This bias of the simulated XCH_4 seems to be constant (around 2.7 %) each day. Thus, an offset applied to all sites for each simulation date is introduced here to compare the model and the measured data, effectively removing the bias, which can be attributed to a too high background XCH_4. The daily offset is assumed to be the difference between the smoothed simulated and measured daily mean XCH_4.

After applying the daily offset, the measured XCH_4 shows a somewhat better agreement and a similar trend but with larger variability compared to the simulation (RMSE is 3.17 ppb). The smaller variations from the simulation results can, for example, be caused by the error from the spatio-temporal treatment of emission maps, underestimated emissions from anthropogenic activities, the coarse wind data and/or the smoothing of actual extreme values in the simulation.

A major offset in modeled CH_4 concentration fields could potentially be attributed to the errors in the troposphere height and a general offset from CAMS. An illustration of the vertical distribution for CH_4 is provided in Fig. 6.7. In the CH_4 vertical concentration profile, it is found that the typical sharp decrease occurs at the tropopause height. Tukiainen et al. (2016) [128] also find the similar sharp decrease when using the AirCore to retrieve atmospheric CH_4 profiles in Finland. During the simulation, the background concentration values of CAMS are directly fitted to the WRF pressure axis without considering

the actual tropopause height; thus this could cause some error.

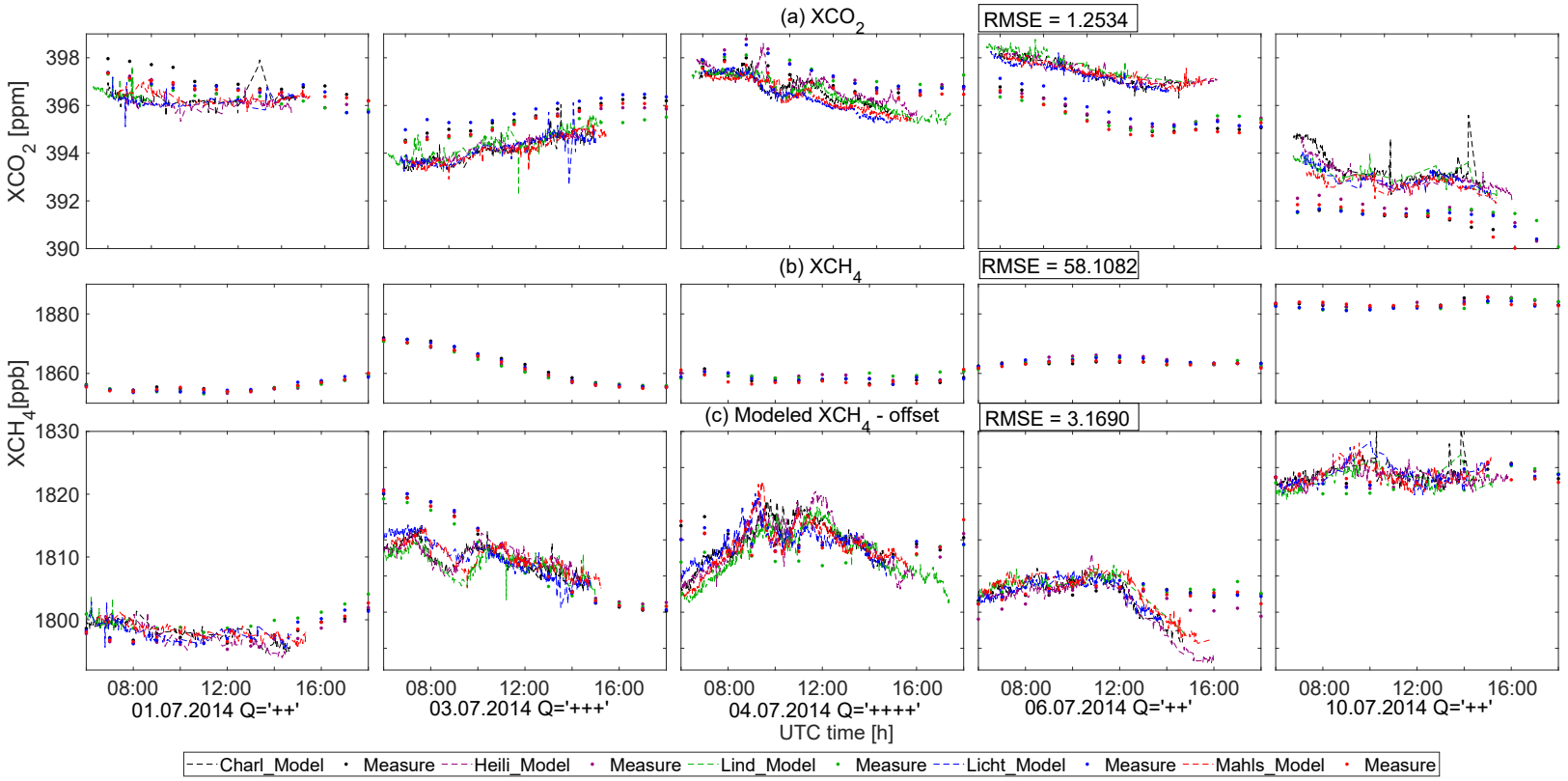


Figure 6.6.: Variations of the measured and smoothed simulated (a) XCO_2 and (b) XCH_4, on 1, 3, 4, 6 and 10 July 2014, for five sampling sites in Berlin: Charlottenburg (Charl: black markers), Heiligensee (Heili: purple markers), Lichtenrade (Licht: green markers), Lindenberg (Lind: blue markers) and Mahlsdorf (Mahls: red markers). The solid circles in (a & b) stand for the simulated values, and the dashed lines represent the measured concentrations. The solid circles represent the simulated XCH_4 after the subtraction of the daily offset in (c).

In contrast, the CO_2 vertical distribution shows decrease that is quite flat with the increase in pressure, and there is no need to consider the tropopause height during the grid treatment in the vertical layer. In terms of CAMS, the reports from MACC stated that CAMS has a bias and RMSE (approximately 50 ppb) in each part of the world, compared to the Integrated Carbon Observation System (ICOS) observed values in 2017 [129]. Galkowski et al. (2019) [130] also mentioned one CH_4 offset (approximately 30 ppb within troposphere) when initializing the concentration fields using CAMS. Apart from these two major potential reasons for this bias, the influence from the inaccurate simulated PBLs and the shape of the constant a priori profile used for the retrievals could both potentially contribute to the discrepancies for the concentration fields. Due to the lack of fine measured vertical concentration profiles, it is not easy to quantify these errors and attribute these potential reasons to this 2.7 % error quantitatively.

Thus, a DCM-based analysis is presented in Sect. 7.1, aiming at eliminating the bias from these relatively high initialization values for CH_4 and making it easier to assess the modelled results with respect to the measurements.

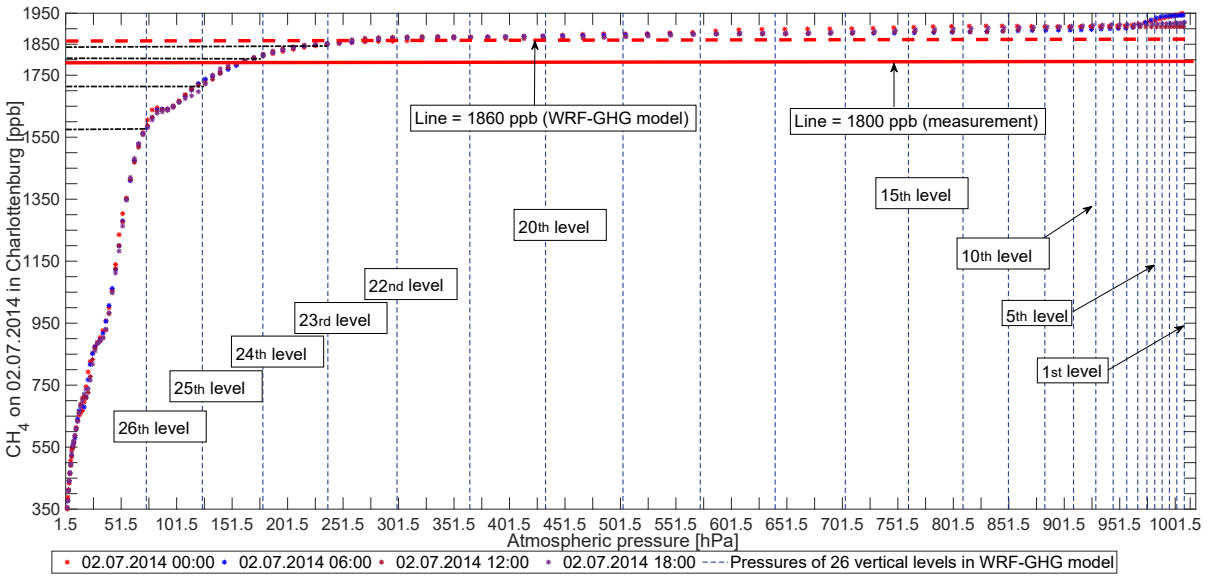


Figure 6.7.: The vertical distribution of CH_4 on 2nd July in Charlottenburg. The asterisks represent the XCH_4 field from CAMS. The vertical dashed lines show the values of atmospheric pressure corresponding to the 26 vertical levels in the model. Y–axis levels of 1800 and 1860 ppb, corresponding to the total column measurement and the modelled value, respectively, have been marked by red horizontal (solid and dashed) lines.

6.4. Comparisons for Munich

This section continues with the model-measurement comparison of slant column concentrations for the Munich measurement campaign (Sect. 5.2.2). It is worth noting in advance that the measured samples are filtered during the autonomous retrieval process in MUCCnet [110]. Specifically, to reduce uncertainties caused by high air masses, measurements are discarded when they are observed at SZA larger than 75 ° [131]. Thus, the measurement period each day ranges from 6:00 UTC to 17:00 UTC and lasts for approximately 11 hours in summer.

6.4.1. Daily mean concentrations

The daily mean measured and modelled values of $XCO^{S}_{2,sla}$ and $XCH^{S}_{4,sla}$ for the 15 studied days and their scatter plots are shown in Fig. 6.8. When producing time-averaged modelled values, the limited measurement period is considered, as mentioned above. For CO_2, the simulated smoothed column concentrations ($XCO^{S}_{2,sla}$, dotted lines in Fig. 6.8(a)) are overall overestimated compared to the measurements, with a mean bias (MB) of 3.7 ppm ± 0.9 ppm, the latter value giving the standard deviation of the MB over all measurement days.

This bias for $XCO^{S}_{2,sla}$ is mainly attributed to the initial and boundary conditions of the concentration fields in the model as provided by CAMS, which has also been seen in other studies. Galkowski et al.(2021)[132] found a similar bias between the CAMS product and airborne measurements in the free troposphere over Europe, with a MB of 3.7 ± 1.5 ppm. Moreover, Tu et al. (2020)[131] also reported a bias when comparing CAMS to their column measurements from the Collaborative Carbon Column Observing Network (COCCON) site at Kiruna, Sweden, with a MB of 3.7 ± 1.8 ppm. As can be seen in the scatter plot of $XCO^{S}_{2,sla}$ (Fig. 6.8(b)), the measurements generally exhibit more scatter (seen in the magnitude of the error bars in the x-direction) compared to the simulation (the error bars in the y-direction), and the slope of the linear regression is only 0.45. The smaller standard deviations in the model represent weaker fluctuations over the daily mean. This will be discussed further in Sect. 6.4.2 when looking into the model-measurement comparison at higher temporal resolutions.

Figure 6.8(c,d) shows that the daily mean modelled values of $XCH^{S}_{4,sla}$ (solid lines in Fig. 6.8(c)) agree well with the daily mean measurements.

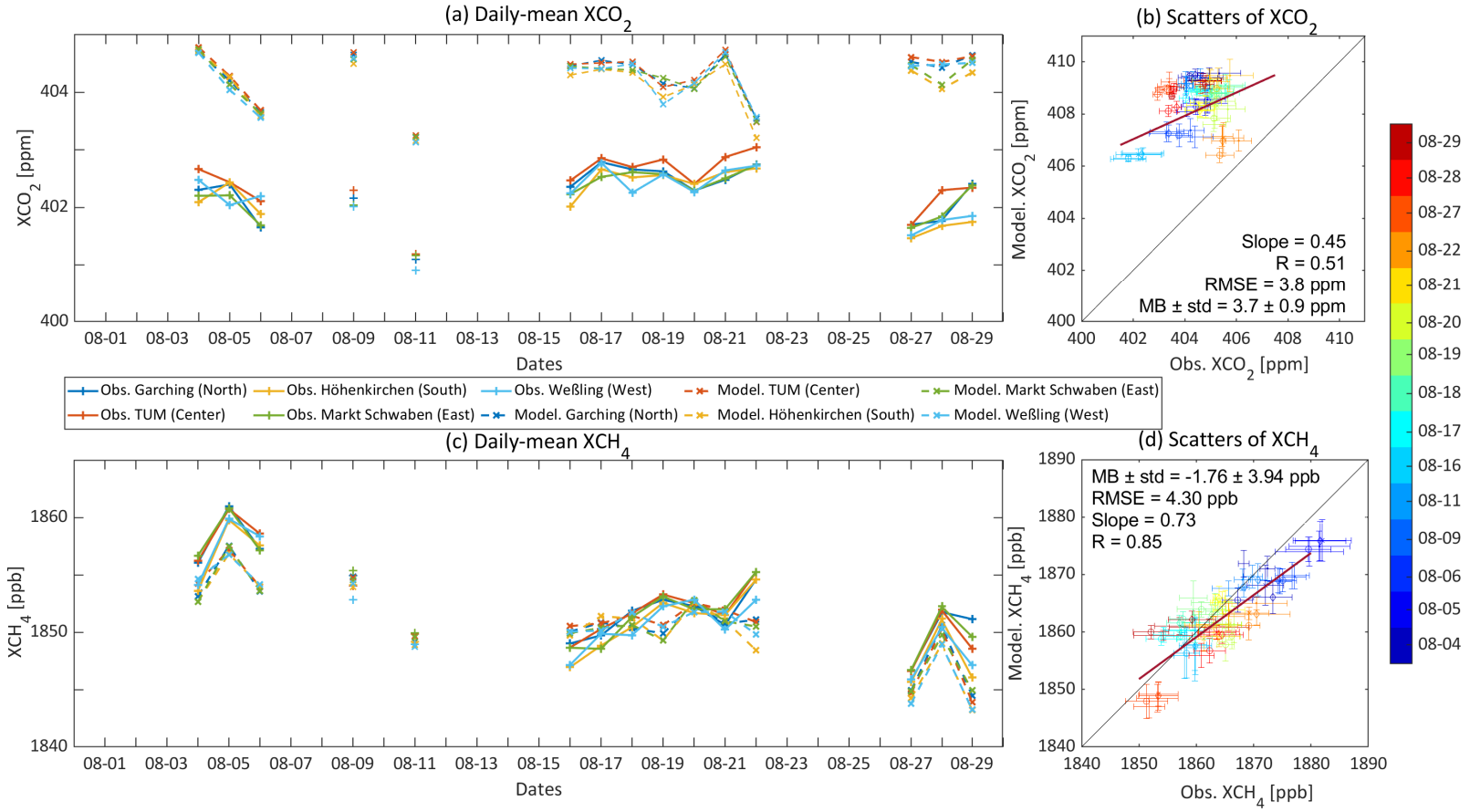


Figure 6.8.: Time series and scatter plots for XCO_2 (a & b) and XCH_4 (c & d). In panels (a) & (c), the dashed lines represent the daily mean modelled $XCO^S_{2,sla}$ and $XCH^S_{4,sla}$, while the solid lines denote the measurements. Colors in panels (a & c) mark the different measurement sites. In scatter plots (panels (b & d)), colors represent the values for different measurement days, as marked in the color bar. The error bar is the standard deviation of the measured and simulated values at each site.

The model is able to capture most of the variations in the daily mean values, while in general the observed values are slightly higher, with a linear regression slope of 0.73 and a negative MB (-1.8 $\pm$ 4.0 ppb). This small bias could be caused by the initial and lateral boundary conditions from CAMS, or due to unknown or underestimated emissions. Comparing CH_4 in the CAMS product with in-situ observations in the troposphere, Galkowski et al. (2021) [132] also reported a negligible MB, but a relatively large standard deviation (0 $\pm$ 14 ppb) in their setup.

6.4.2. Intra-day concentrations

In order to obtain a more detailed view on how the model behaves at higher temporal resolutions, the daily curves of $XCO^{S}_{2,sla}$ and $XCH^{S}_{4,sla}$ at five sites for seven consecutive days from 16 to 22 August are plotted in Figs. 6.9 & 6.10 against the corresponding modeled values with the hourly temporal resolution.

As mentioned in Sect. 6.2, the modelled and observed slant column concentrations used for the model-measurement comparisons are smoothed using the SZA-dependent AK based on Eq. 6.2. Due to the restriction of SZAs and the corresponding availability of measured values provided by the MUCCnet (cf. Sect. 5.2.2), these model-measurement comparisons of the total column concentrations for GHGs can only be made during the daytime, approx. from 6:00 UTC to 17:00 UTC (cf. Figs. 6.9 & 6.10). However, Figure 6.10 also illustrates the contributions to the total column concentrations of CO_2 from different tracers in the model at nighttime, using pressure-weighted column concentrations (Eq. 6.1) as a proxy. This helps understand the model and the measurements, e.g., the contribution of nighttime vegetation respiration to the changes in total column concentrations of CO_2. Apart from the modelled total column concentrations smoothed by the SZA-dependent AK, the contributions from other tracers related to biogenic and human activities shown in Figs. 6.9 & 6.10 along the full-time series are calculated following Eq. 6.1 without being smoothed by the AK.

Figure 6.9 shows the modelled CO_2 column concentrations from all tracers. In general, there is little difference in the column background concentrations among the five sites (black lines). Variations of the modelled total CO_2 corrected by the MB (solid red lines) are mostly dominated by biogenic activity (dashed blue lines), with only a minor influence predicted from anthropogenic emissions (dashed yellow lines).

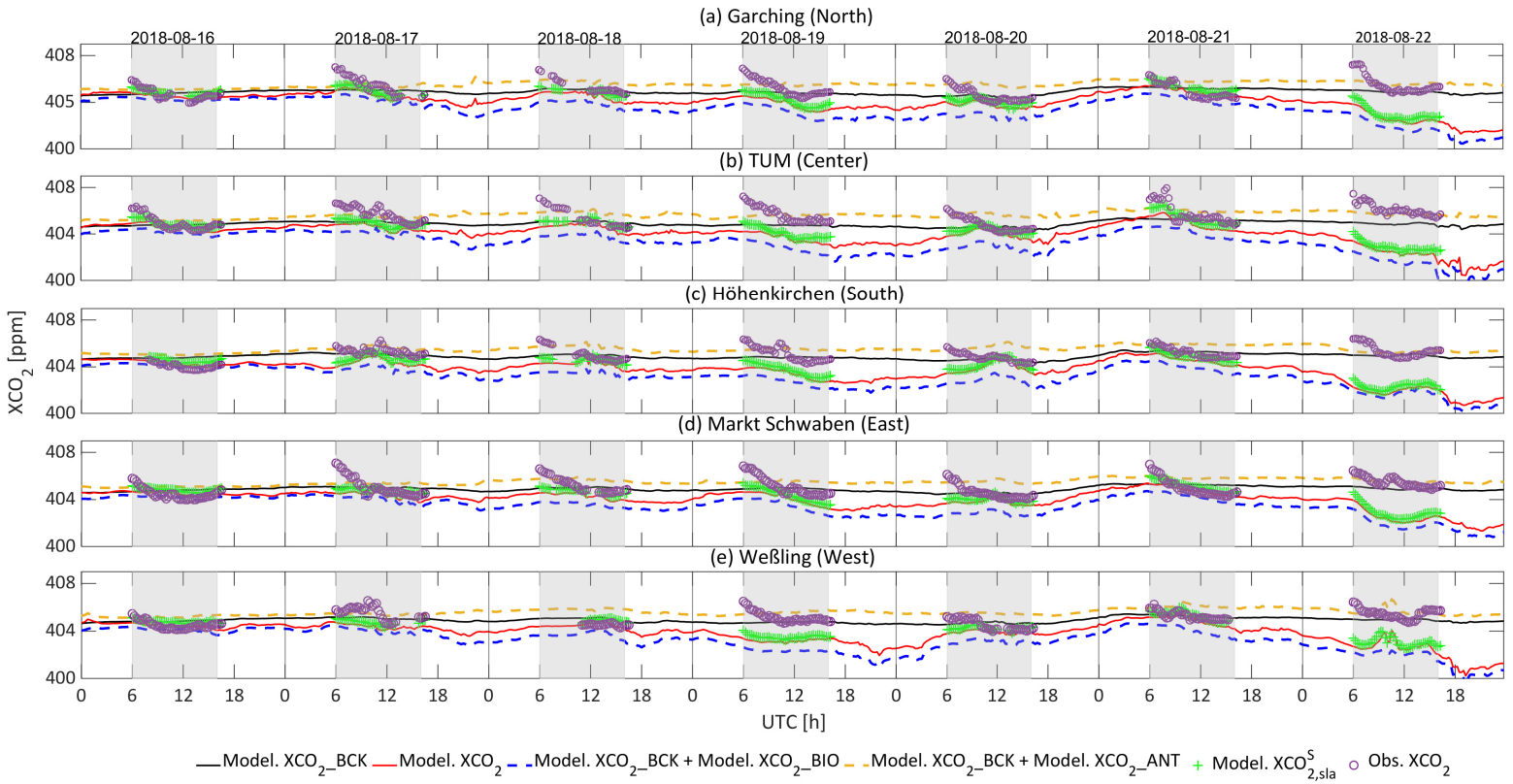


Figure 6.9.: Modelled $XCO^{S}_{2,sla}$, attribution of variations to different tracers, and the measurements at five sites of MUCCnet, from 16 to 22 August, 2018. The purple circles represent the column measurements from MUCCnet and the green '+' stands for the modelled $XCO^{S}_{2,sla}$ after subtracting the MB. The other lines in the plot give pressure-weighted modelled column concentrations along the full time series. This is calculated following Eq. 6.1 (i.e., without smoothing using the SZA-dependent AK) and all corrected by MB. The black curve represents the modelled background (XCO_2_BCK), and the red shows the modelled XCO_2. The dashed yellow and blue curves highlight the concentration changes caused by human activities (XCO_2_ANT) and biogenic activities (XCO_2_BIO). The grey shaded areas mark the measurement periods used for comparing observations to model results.

After smoothing, the modelled, bias-corrected $XCO^{S}_{2,sla}$ (green '+') is slightly higher than the corresponding XCO_2 modelled values (solid red lines), especially during the morning and afternoon, with a RMSE of 0.37 ppm and a MB with its standard deviation of 0.34 ± 0.13 ppm. This is caused by the steeper shape of the vertical profile of the AK under smaller SZAs. During the daytime from around 6:00 UTC to 17:00 UTC, there is no obvious difference between the modelled values with and without smoothing.

The modelled $XCO^{S}_{2,sla}$ (green '+') reproduces the variability in the measurements (purple 'o') reasonably well, with a RMSE of 1.33 ppm and a MB $\pm$ std of -0.79 ± 0.14 ppm, as it turns out. However, the measurements often show a steep decrease in concentration during the morning, while the model only shows slight declines. This difference could be mainly due to the underestimation of the modelled RES from VPRM. Specifically, during the growing seasons (June-September), VPRM [90] has been found to underestimate RES, especially at nighttime and overestimate GPP during the daytime [101], [133]. This causes an overestimation of the magnitude of NEE (i.e., the difference between RES and GPP). Such an overestimation of NEE could explain the difference between the modelled column concentrations and what is observed by MUCCnet. In this case, the observations suggested higher RES fluxes at nighttime than what was simulated. This led to much lower modelled column concentrations in the early morning, which was also seen in Hu et al. (2021) [101].

Compared to the modelled values for the other six dates, a slight rise was seen (around 1.5 ppm on average) in the morning of 21 August. This was induced by a combined effect of elevated background and biogenic tracer mole fractions. Closer analysis of this case (i.e., time series of XCO_2_BCK) has shown that background enhancement entered from the northwest of the outermost boundary in the morning of 19 August, and was transported into the innermost domain by the late afternoon of 20 August. This enhanced background signal contributes around two thirds of the modelled 1.5–ppm rise. The rest is the result of air masses with strong biospheric CO_2 signals coming from the northwest of the outermost domain (D01).

On 22 August, compared to the simulations for the other days, the modelled $XCO^{S}_{2,sla}$ shows a larger deficit with respect to the measurements in Fig. 6.9. The model can capture the variation during the day quite well, but produces too low XCO_2 values when bias corrected through subtraction of the MB.

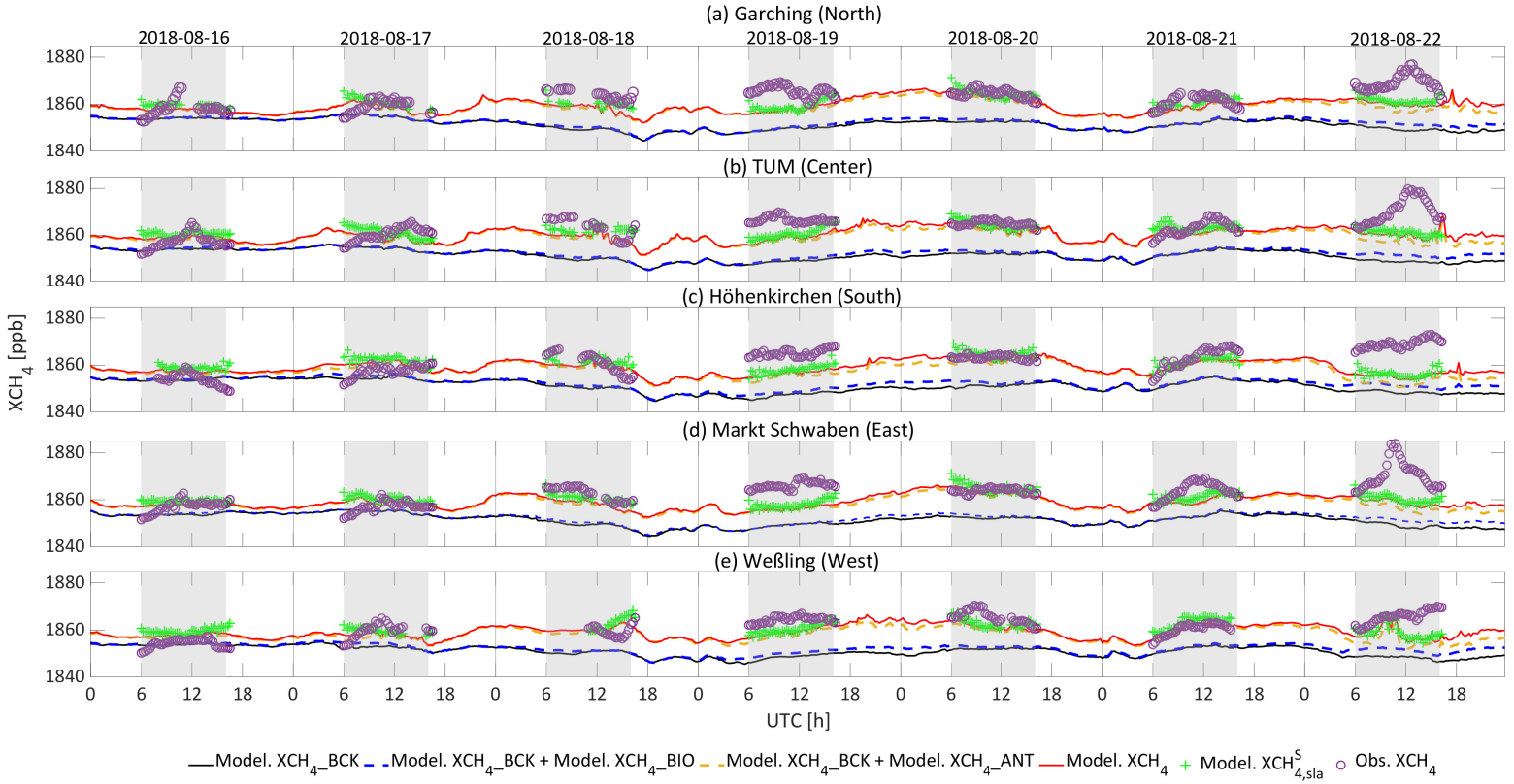


Figure 6.10.: Modelled $XCH^{S}_{4,sla}$, attribution of variations to different tracers and the measurements at five sites of the MUCCnet from 16 to 22 August, 2018. The purple 'o' represents the column measurements from MUCCnet and the green '+' stands for the modelled $XCH^{S}_{4,sla}$. The other lines in the plot show the modelled column concentrations calculated following Eq. 6.1, i.e., without smoothing using the SZA-dependent AK: the solid black for the modelled backgrounds (XCH_4_BCK) and the dashed red for the modelled XCH_4. The dashed yellow and blue lines highlight the concentration changes caused by human activities (XCH_4_ANT) and induced by biogenic activities (XCH_4_BIO), including the background. The grey shaded areas mark the measurement periods used for comparing to the simulations.

This more extreme mismatch is probably due to the advection of air masses heavily impacted by biogenic activities (and thus with less CO_2) in the model, coming into the domain from e.g., Italy, Slovenia and Croatia.

As can be seen in the map of the modelled NEE from VPRM for the outermost domain (D01) in Fig. 6.11, much stronger biogenic fluxes are found in the south of the outermost domain, compared to the other areas. More CO_2 was taken up and the affected air masses (i.e., with comparatively less CO_2) are able to reach Munich when the wind is strong enough to drive them past the Alps. A constant MB over all days, as was applied here, may not always be realistic, as can be seen here. An evaluation of the signals within MUCCnet using the DCM is still possible, however, as will be discussed in Sect. 7.2.4.

In terms of CH_4, the same visual analysis (cf. Fig. 6.10) is conducted. The variations of the modelled total XCH_4 (solid red) are dominated by anthropogenic activities (XCH_4_ANT, dashed yellow line) and these two variations mostly overlap, since no significant signal induced by wetland emissions (XCH_4_BIO, dashed blue) is predicted by the model. The time series of the modelled $XCH^{S}_{4,sla}$ (green '+') shows general agreement with the measurements (purple 'o'). The measurements seem to capture stronger emission signals (e.g., on 22 August), perhaps due to gaps in the knowledge of the spatio-temporal distribution of CH_4 emissions. The modelled values show little diurnal variability at all sites compared to the measurements (RMSE: 6.7 ppb and MB $\pm$ std: -3.3 $\pm$ 5.9 ppb). Regarding the strong enhancements observed by the instruments during the daytime, especially on 22 August, these might be the result of sources which are missing from inventories, or are underestimated in their magnitude. In Sect. 7.2, such unknown or underestimated emission sources over the domain using the DCM attempt to be detected.

6.5. Emission tracer analysis

This section shows contributions of different emission sources and sinks induced by biogenic and human activities for Berlin and a detailed analysis of different anthropogenic tracers for Munich. The analysis for Berlin focuses on understanding the major biogenic emissions from vegetation photosynthesis and respiration and human induced emissions in Sect. 6.5.1. When building the model for Munich, individual human emission sources are added as tracers (see Sect. 4.6). Thus, their contributions to the total signal can be discussed

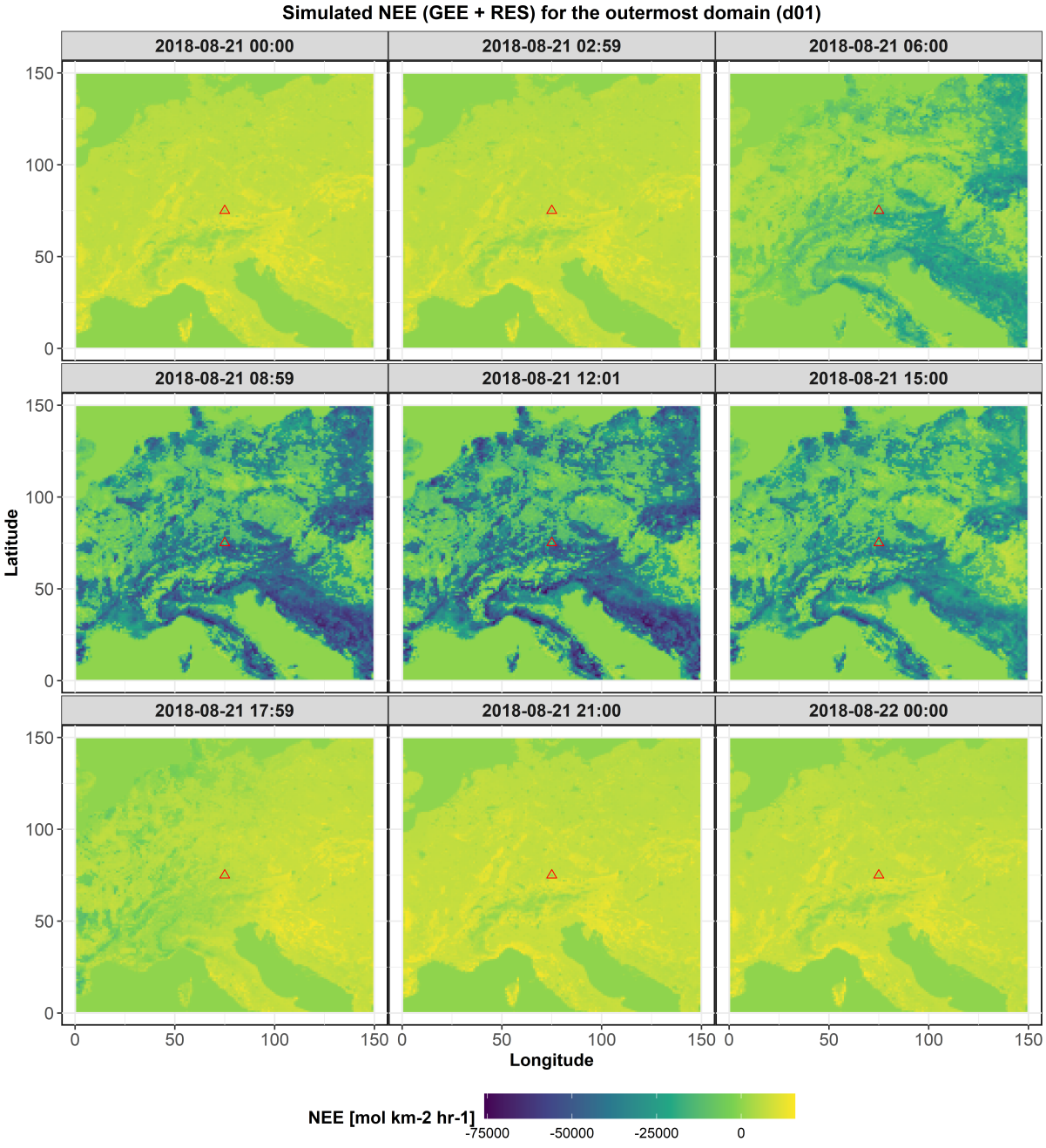


Figure 6.11.: The map of modelled NEE on 21 August for the outermost domain (d01).

in this work and shown in Sect. 6.5.2.

6.5.1. Berlin: biogenic and human activities

As described in Sect. 5.2, the various flux models implemented in WRF-Chem are advected as separate tracers, making it possible to distinguish the signals in concentration space for different source and sink categories for CO_2 and CH_4 [98]. Berlin is located in an area of low-lying, marshy woodlands with a mainly flat topography [134]. There is no wetland in Berlin according to the MODIS Land Cover Map [68]. The land covered by forests, green and open spaces (e.g., farmlands, parks and allotment gardens) accounts for 35 % of the total area in Berlin [135]. Additionally, 11 power plants are currently being operated in Berlin, 8 of which have a capacity of over 100 MW [136].

In accordance with the geographical characteristics of the district and potential emission sources in Berlin, this study focuses on understanding the major emissions caused by vegetation photosynthesis and respiration ($XCO_{2,VPRM}$) as well as anthropogenic activities ($XCO_{2,anthro}$) for CO_2 and by soil uptake ($XCH_{4,soil}$) as well as human activities ($XCH_{4,anthro}$) for CH_4.

As an instructive example of an analysis involving these tracers, Figure. 6.12 depicts the diurnal cycle of contributions from the different tracers mentioned above in Charlottenburg. The mean values, averaged over 9 days (from 2 to 10 July), as well as a 95 % confidential interval calculated in the averaging process are shown in Fig. 6.12. Figure 6.12(a) clearly shows a decline during the day and a rise at night in the XCO_2 enhancement over the background (blue: $XCO_{2,total} - XCO_{2,bgd}$), with a maximum decrease over the course of the day of around 2 ppm. The XCO_2 enhancement over the background reaches its daily peak during morning rush hour (07:00 UTC). The morning peak corresponds to XCO_2 changes from human activities, depicted by the black line from 04:00 to 07:00 UTC (marked by a red square in Fig. 6.11(a)). Before the evening rush hour (16:00 UTC), XCO_2 over the background then decreases, owning to biogenic uptake. Beginning in the evening, values increase again. The fluctuation in the evening (17:00–19:00 UTC) is dominated by XCO_2 enhancements from human activities, while the substantial rise from 19:00 UTC onward is generated by the VPRM tracer, specifically the accumulation of the vegetation respiration in the evening. XCO_2 is weaker compared to the strong biogenic uptake. To further highlight the role of anthropogenic activities in XCO_2 changes within the urban area, DCM is applied in Sect. 7.1.

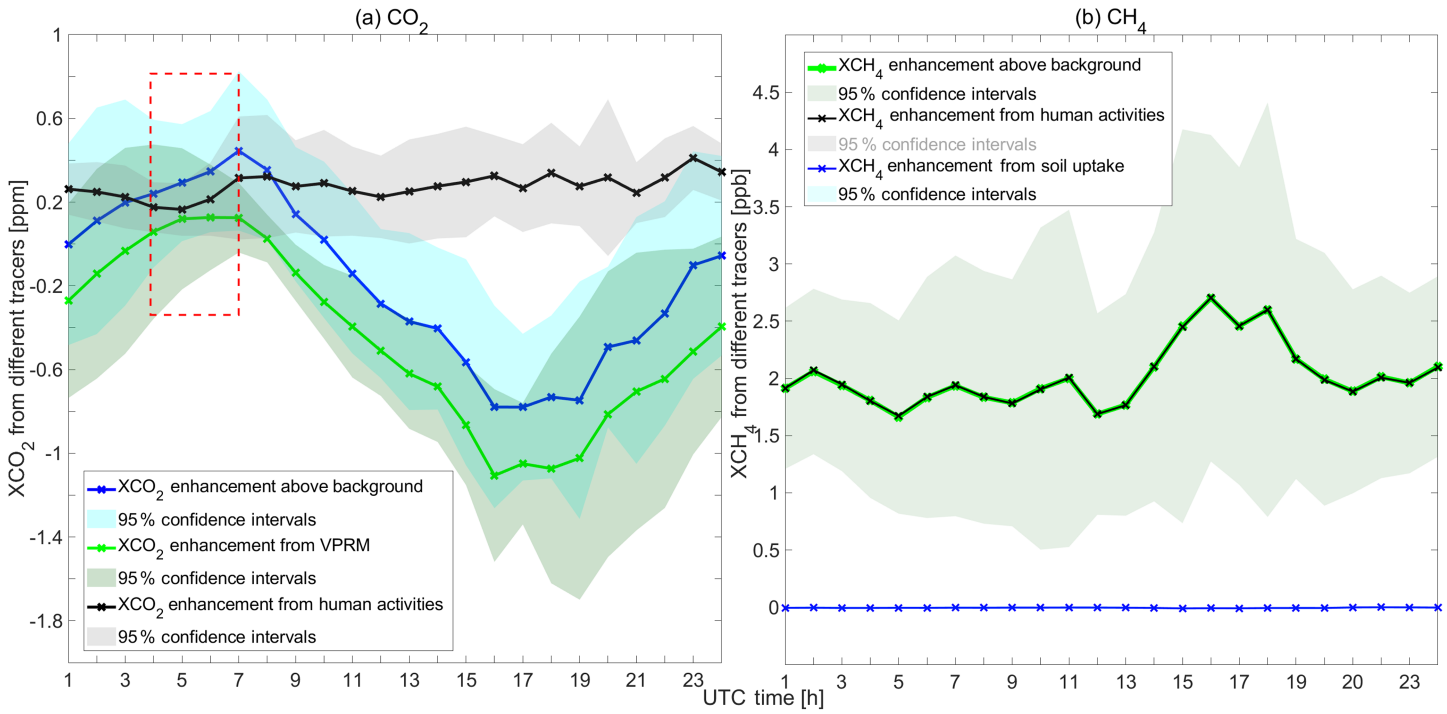


Figure 6.12.: The diurnal variations of the simulated changes in concentrations caused by different emission tracers in Charlottenburg in Berlin from 2014, averaged over a period of nine days (from 2 to 10 July 2014). The colored lines represent the concentration changes and the mean enhancement over background. (a): the mean hourly $XCO_{2,VPRM}$ (green line) and $XCO_{2,anthro}$ (black); (b): the mean hourly $XCH_{4,anthro}$ (black) and $XCH_{4,soil}$ (blue). The red box in (a) marks the morning peak of the XCO_2 enhancement over the background, as described in Sect. 6.3

More specifically, the downwind-minus-upwind column differences of CO_2 (ΔXCO_2) will be used to describe the XCO_2 enhancement over an upwind site, as the difference between the downwind and upwind sites can be attributed to urban emissions.

Turning to XCH_4, the variations in the mean hourly contributions from the anthropogenic (black line: $XCH_{4,anthro}$) and soil uptake tracer (blue line: $XCH_{4,soil}$) in Charlottenburg are shown in Fig. 6.12(b). The contributions by anthropogenic activities fluctuate slightly around 2 ppb in the morning and at noon; then a peak occurs at the start of the evening rush hour (16:00 UTC). After 18:00 UTC, values clearly decrease, reaching approximately 2 ppb. From 21:00 UTC, XCH_4 stabilizes, exhibiting only moderate fluctuations. The XCH_4 enhancement above the background (green: $XCH_{4,total} - XCH_{4,bgd}$) depends largely on the XCH_4 contributions by human activities. The changes in concentrations caused by the soil uptake tracer (blue), whose values fluctuate between 0.001 and 0.01 ppb, have almost no influence on the variation in the XCH_4 enhancement over the background in the urban area.

6.5.2. Munich: individual human activities

Beyond the major contributors to the concentration enhancements above the background as discussed above, the contributions from individual anthropogenic emission processes to understand how these processes impact concentrations quantitatively are also analyzed. To be specific, the GNFR emission categories from TNO_GHGco is used for separately advected tagged tracers. For CO_2 the categories are: "A. Power Plants", "B. Industry", "C. Other Stationary Combustion", "D. Road Transport" and "E. Other". For CH_4, the emission processes are: "A. Power Plants", "B. Industry", "C. Agriculture", "D. Waste Management", "E. Fugitives and solvents", and "F. Other".

The changes in concentrations induced by different human activities along the full time series are plotted in Fig. 6.12. For CO_2, the emissions from road transport (red) contribute the largest portion (around 37.4 %) of the total concentration enhancements caused by anthropogenic activities in August, 2018. This is consistent with the finding that over 30 % of the total GHG emissions are induced by on-road transportation for around one-third of 167 cities worldwide [137]. This finding may also be due to the seasonal features of emissions in Munich, as there is no heating and less electricity generation in summer.

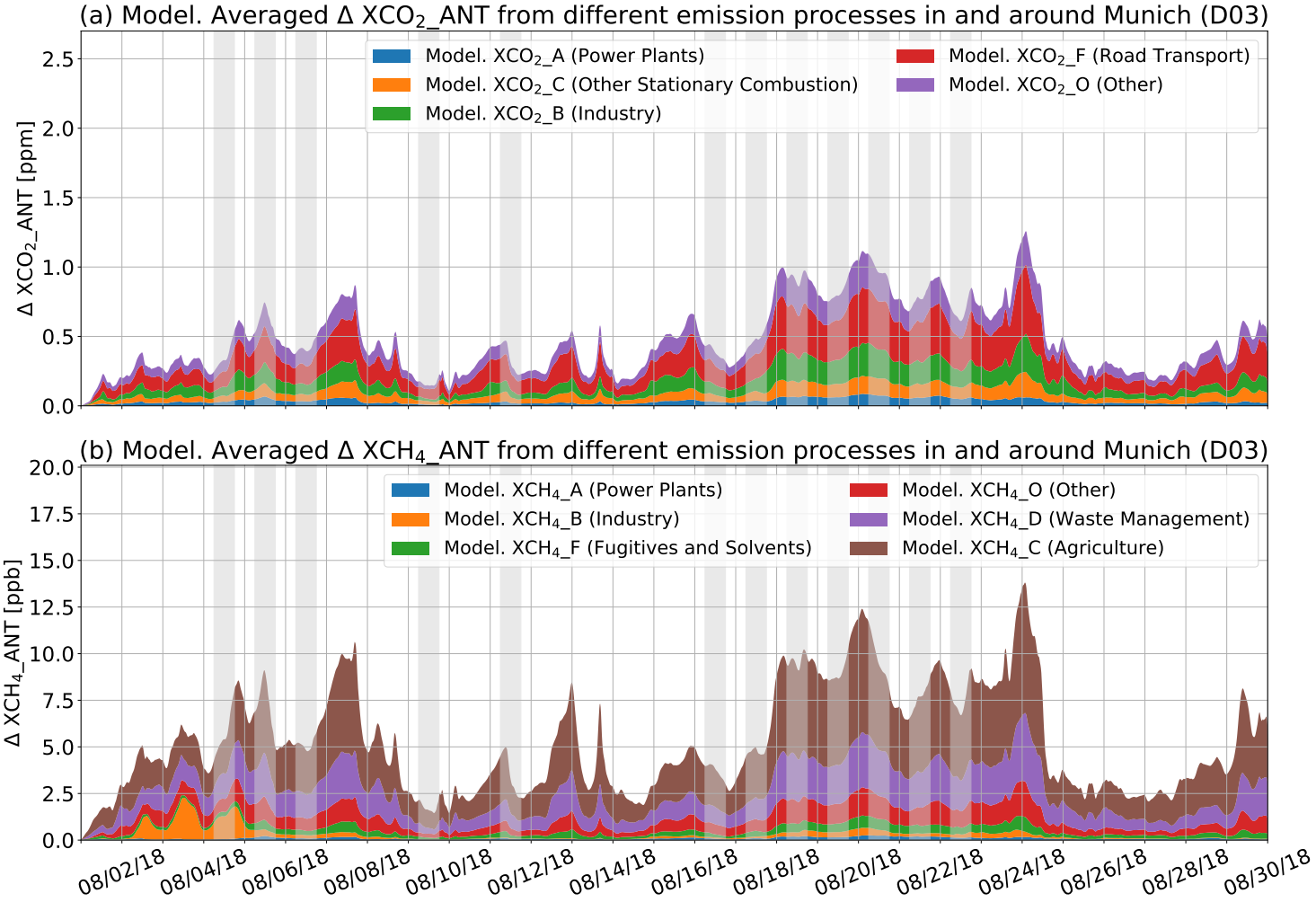


Figure 6.13.: Average modelled concentration contributions by individual anthropogenic emission processes of (a) CO_2 (XCO_2_ANT) and (b) CH_4 (XCH_4_ANT) from 1 to 30 August, 2018 in and around Munich (D03). The grey shaded areas mark the measurement periods used for comparing to the simulations.

For other contributors, power plants account for around 7 %, other stationary combustion for 13.1 %, and both industry and other contributors for approx. 21 %, respectively.

Figure 6.12(b) shows that for CH_4 the emissions are dominated by agriculture (brown) and waste management (purple), which are estimated to contribute around 50 % and 25 % of the total human-related concentration enhancements

To summarize, the simulated pressure vertical profile (along the line of the Sun) and the AK from the solar-viewing spectrometer (EM27/SUN) need to be used to obtain the smoothed pressure-weighted average concentration for further comparisons. For both cases, the simulated XCO_2 concentrations actually reproduce the observations well. But for Berlin, an approximately 1–2 ppm bias is found, which can be attributed to the coarse emission inventory, background concentrations from CAMS and the ignorance of the line of the sun sight for the simulation. For Munich, a bias (around 3.7 ppm) for the column CO_2 along the slant column exists in the model. This is caused by the initial and boundary conditions provided by CAMS. In addition, the measured column concentrations along the time series show larger fluctuations, which were attributed to stronger concentration enhancements induced by emission sources that are underestimated or not included in the initial emission inventory. In the comparison of column concentrations over the course of the day, the underestimated morning enhancements of CO_2 in the model may be caused by the underestimation of the modelled RES from VPRM. Regarding CH_4, some deviations can clearly be noted in the simulated XCH_4 compared to the measured XCH_4 for Berlin, mostly caused by the relatively high background concentration fields and the errors at the tropopause height, while this is not the case for Munich due to its improvement in recent years. Turning to the model-measurement comparison for Munich of the column CH_4 along the slant column, the model shows general agreement with the measurements, though the instruments see stronger emission signals that may be due to gaps in the prior knowledge of CH_4 emissions.

7. DCM-based model assessments [51]–[53]

In order to analyse the modelled results in more depth and learn about unknown emission sources, the model-measurement comparison using the differential column method (DCM [39]) is presented in this section. DCM can be employed to detect and estimate local emission sources within an area, based on calculated concentration differences between downwind and upwind sites [39]. The difference (ΔXG) of a specific gas G in column-averaged DMFs across the downwind and upwind sites is defined as,

$$\Delta \mathrm{X}G = \mathrm{X}G_{\mathrm{downwind}} - \mathrm{X}G_{\mathrm{upwind}} \tag{7.1}$$

where XG_{downwind} and XG_{upwind} represent the column-average DMFs at the downwind and upwind sites.

In this study, DCM allows for a cleaner interpretation of the behaviour of measurements and models, and is applied to them in the spirit of a post-processing analysis. This approach is not only useful for reducing the effect of the bias from the initial and boundary conditions of the modelled background concentration fields provided by CAMS (Sect. 4.2.3) when assessing the role of anthropogenic activities in XCO_2 changes more appropriately, but also for localizing the potential unknown or underestimated emissions sources around the measurement network.

A necessary prerequisite for DCM is distinguishing the upwind and downwind sites among all five sampling sites. Wind direction thus plays a pivotal role in the calculation of the downwind-minus-upwind column differences. With the restricted knowledge in the particle transport when doing the initial test for Berlin, a simple approach used for selecting upwind and downwind sites is used here based on the daily averaged winds from three wind sampling sites around Berlin (see Sect. 7.1), as marked by red circles in Fig. 5.3. Turning to the Munich case, the application of DCM compared to the Berlin study has

been improved through making use of a transport model at this point, together with wind information to select days for this method is applicable, presented in detail in Sect. 7.2.

7.1. Assessments for Berlin

7.1.1. With wind information

In the Berlin case, the hourly simulated vertically averaged wind directions are assumed as a standard to classify the sites into downwind and upwind sites. The tracer transport calculations in the first few hours are not stable in WRF-Chem. Thus, 3, 4, 6 and 10 July as the targeted dates are selected.

Date	Wind direction (degree)	Upwind sites	Downwind sites
3 July	272.55 ± 20.19	Charlottenburg–Heiligensee	Lindenberg–Mahlsdorf
4 July	206.93 ± 24.23	Lichtenrade	Heiligensee–Lindenberg
6 July	214.51 ± 26.38	Lichtenrade	Heiligensee–Lindenberg
10 July	38.03 ± 25.33	Mahlsdorf–Lindenberg	Heiligensee–Charlottenburg

Figure 7.1.: The selections of upwind and downwind sites for four dates.

The table shown in Fig. 7.1 lists the daily averaged wind directions with standard derivations and the details on the downwind and upwind sites for these four target dates. West wind is the prevailing wind direction on 3 July. That is to say, Mahlsdorf and Lindenberg are downwind sites, and the upwind sites corresponding to these are Charlottenburg and Heiligensee, described in Eq. 7.2. The wind on 10 July is northeasterly, and the combination of downwind and upwind sites are selected to be opposite of the ones on 3 July, see Eq. 7.4. The prevailing winds on 4 and 6 July are easterly. The upwind site is Lichtenrade, and the corresponding downwind sites are Heiligensee and Lindenberg, see Eq. 7.3. Based on the selection of downwind and upwind sites shown in Table 7.1 and Eq. 7.1, differential column concentrations (ΔXCH_4) are, therefore, calculated as

$$\begin{aligned} &\text{Western Wind (3 July)}: \\ \Delta \mathrm{XCH_4} = &(\mathrm{XCH_4^{Mahlsdorf}} - \mathrm{XCH_4^{Lindenberg}})/2 \\ &-(\mathrm{XCH_4^{Charlottenburg}} - \mathrm{XCH_4^{Heiligensee}})/2 \end{aligned} \tag{7.2}$$

$$\text{Northern Wind (4 and 6 July):}$$
$$\Delta\mathrm{XCH_4} = (\mathrm{XCH_4^{Heiligensee}} + \mathrm{XCH_4^{Lindenberg}})/2 - \mathrm{XCH_4^{Lichtenrade}} \tag{7.3}$$

$$\text{Northeastern Wind (10 July) :}$$
$$\Delta\mathrm{XCH_4} = (\mathrm{XCH_4^{Charlottenburg}} + \mathrm{XCH_4^{Heiligensee}})/2 - (\mathrm{XCH_4^{Mahlsdorf}} + \mathrm{XCH_4^{Lindenberg}})/2 \tag{7.4}$$

Figure 7.2 depicts the variations in the wind fields (wind speeds and wind directions) and $\Delta\mathrm{XCH_4}$ (corresponding to Eqs. 7.2, 7.3 and 7.4) on 3, 4, 6 and 10 July. As depicted in Fig. 7.2(a–d), the hourly vertically averaged simulated wind speeds and directions at downwind and upwind sites are homogeneous. Thus, it is reasonable to use the daily mean wind directions as the standard for the selection of downwind and upwind sites. The general trends in the simulated $\Delta\mathrm{XCH_4}$ values, shown in Fig. 7.2(e–h), seem to be roughly reproduced by the observations but slightly overestimated, with an RMSE of 1.3895 ppb. Yet DCM as presented here has the potential to highlight the role of anthropogenic activities, which is demonstrated, applying it to CO_2 tracers in the simulation. Thus, the analysis on anthropogenic and biogenic tracers for CO_2 will be especially prominent here.

As described above, 3, 4, 6 and 10 July are taken as examples (Fig. 7.3(a–d)). The variations in $\Delta\mathrm{XCO_2}$ (corresponding to Eqs. 7.2, 7.3 and 7.4) on 3, 4, 6 and 10 July are shown. In contrast to the variations in $\mathrm{XCO_2}$ values (Sect. 6.5.1; Fig. 6.12(a)), the simulated $\Delta\mathrm{XCO_2}$ (Fig. 7.3(a–d), blue lines) is not so much influenced by the $\mathrm{XCO_2}$ changes from the VPRM tracer (Fig. 7.3(a–d), green) but more closely follows the $\mathrm{XCO_2}$ changes from anthropogenic activities (red). With DCM, the role of human activities in $\mathrm{XCO_2}$ changes is highlighted, and the strong effect from the biogenic component is canceled out. The $\Delta\mathrm{XCO_2}$ measurements (Fig. 7.3(a–d), black) show similar trends as the simulation with an RMSE of 0.2973 ppm. To further understand the differences of $\Delta\mathrm{XCO_2}$ and $\Delta\mathrm{XCH_4}$ between measurements and simulations (see Fig. 7.2(e–h) and Fig. 7.3(a–d)), the comparison of hourly mean $\Delta\mathrm{XCO_2}$ and $\Delta\mathrm{XCH_4}$ values for these four targeted dates is illustrated in the right column of Fig. 7.3. Due to the restriction of measured wind information, the differences of simulated and measured wind directions at 10 m (Fig. 5.4(b)) with respect to the hourly mean $\Delta\mathrm{XCO_2}$ and $\Delta\mathrm{XCH_4}$ is illustrated here. It is found that the real hourly mean $\Delta\mathrm{XCO_2}$ and $\Delta\mathrm{XCH_4}$ values are generally higher than the simulated values. Extreme points are colored by red and blue in the right column of Fig. 7.3(e–f),

standing for large differences between measured and simulated wind directions at 10 m.

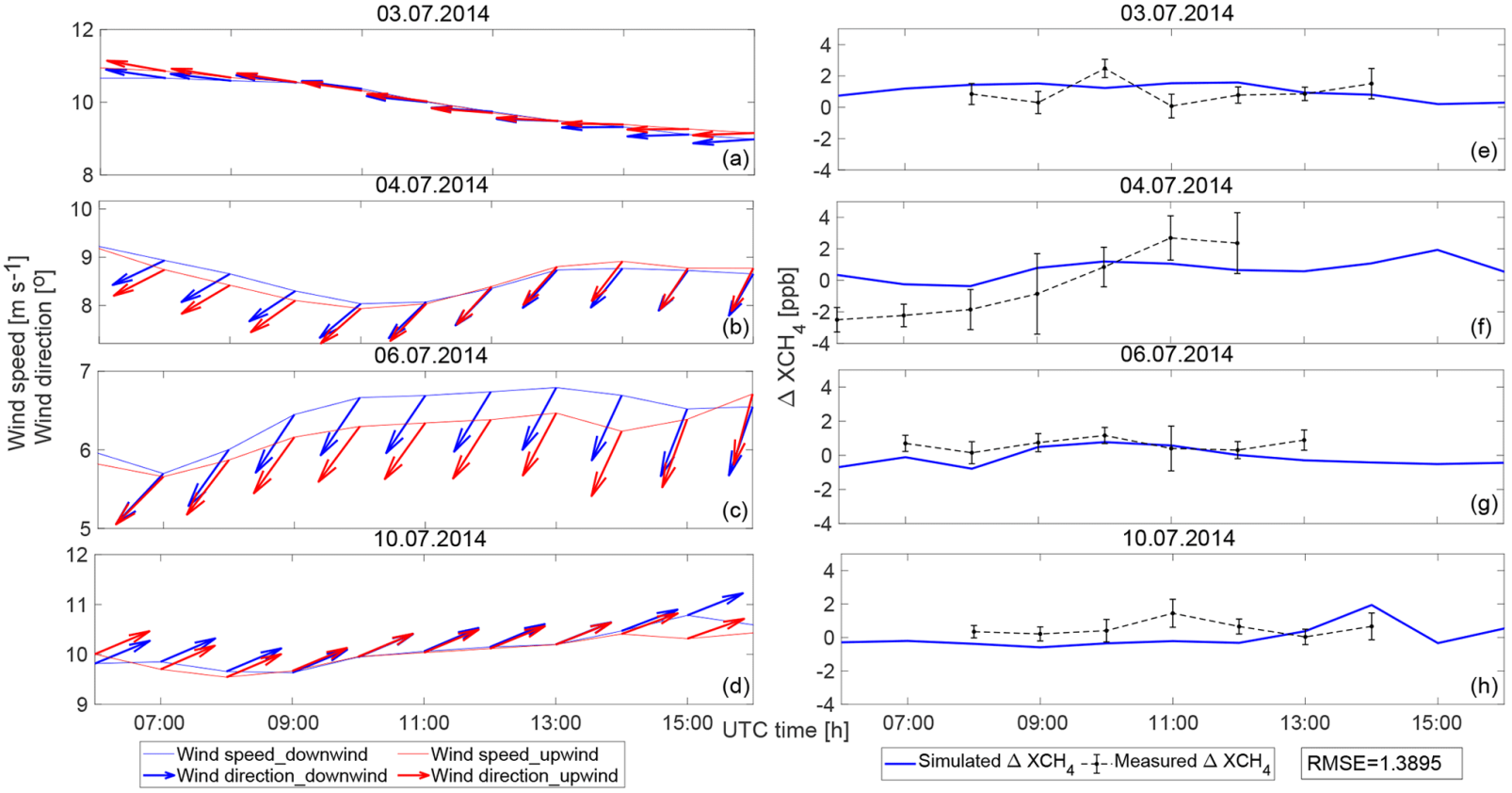


Figure 7.2.: Modeled wind fields for downwind (blue lines) and upwind (red lines) sites (a–d), and downwind-minus-upwind differential evaluation for measured (blue) and simulated (black lines) XCH_4 (e–h) on 3, 4, 6 and 10 July 2014. Based on the selection of downwind and upwind sites listed in Fig. 7.1, ΔXCH_4 is calculated using Eqs. 7.2, 7.3 and 7.4, depicted by blue lines for measurements and black lines for simulations. The black error bars in (e–h) are the standard derivations of the minute values of the hourly mean.

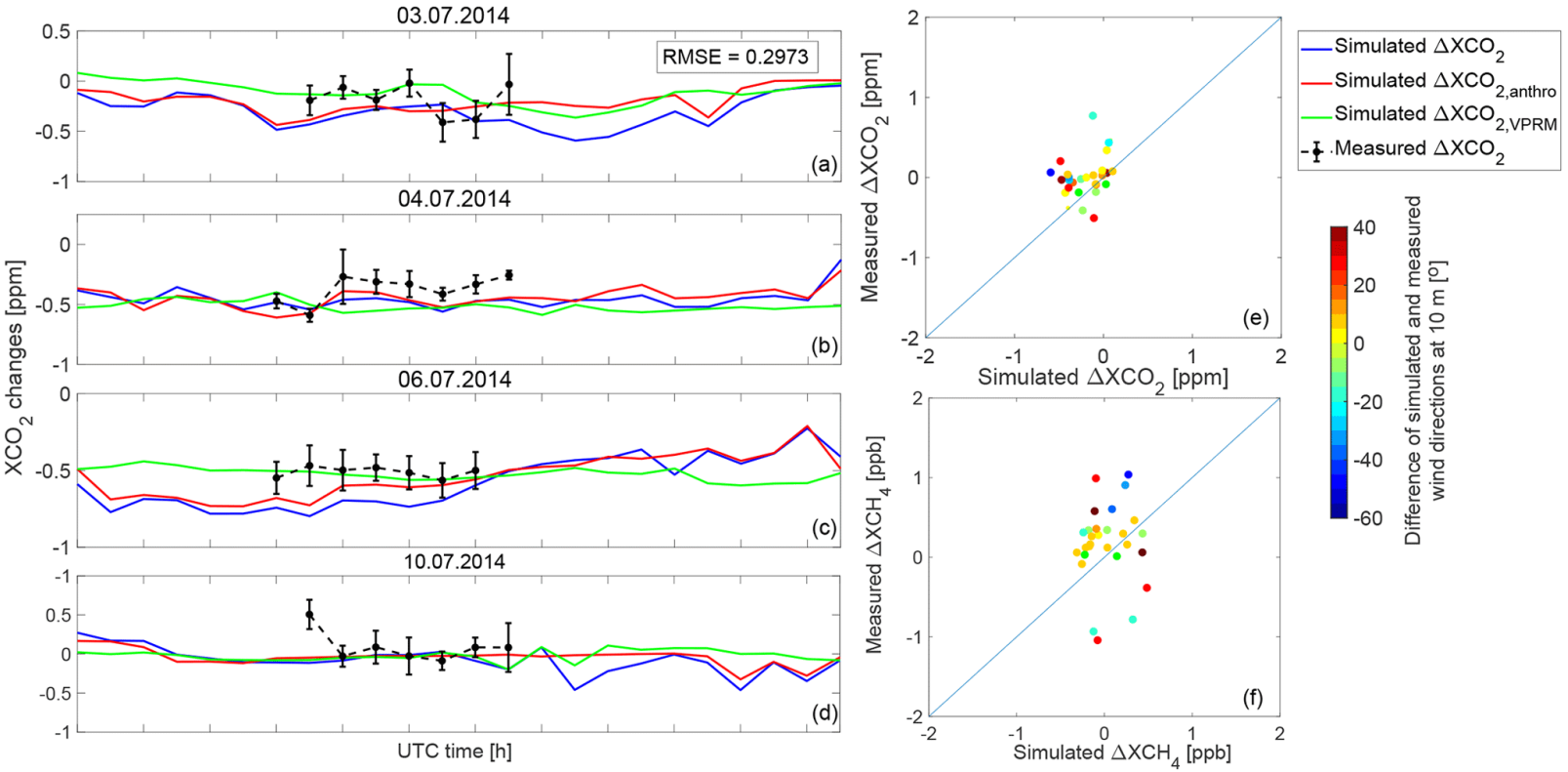


Figure 7.3.: Measured (black lines) and simulated (blue lines) ΔXCO_2 on 3, 4, 6 and 10 July 2014, and comparison of hourly mean ΔXCO_2 and ΔXCH_4 for these 4 days. The ΔXCO_2, calculated using Eqs. 7.2, 7.3 and 7.4, are depicted by blue lines in (a–d). The red and green lines show the variation in the differences between downwind and upwind sites in XCO_2 changes from anthropogenic and biogenic activities, respectively. The points in (e–f) are coded by the difference of the simulated and measured wind directions at 10 m. The black error bars in (a–d) are the standard derivations of the minute values of the hourly mean.

It can be seen that a large difference of wind directions is a necessary but insufficient condition for the bias of ΔXCO_2 and ΔXCH_4 between measurements and simulations. In future studies, this is suggested as something to be verified further. Therefore, DCM, as applied in this plot, reduces the model bias caused by the simulation initialization but introduces unpleasant effects which may be attributed to errors in the assumed or simulated wind directions.

7.1.2. Without wind information

As mentioned above, the wind direction impacts the distinction between downwind and upwind sites for DCM. Devising meaningful and accurate recipes for determining the wind directions is not easy, sometimes resulting in mixed-quality results (of Sect. 7.1.1). The simulated output provides the hourly wind and concentration fields. The instruments measure the concentration value every minute [43].

As simply assumed, the wind direction to be a constant value within 1 h (the hourly vertically averaged values) in this calculation also when it comes to selecting upwind and downwind sites. This may create inaccuracies in the calculation of the measured $\Delta\mathrm{XCH_4}$. As tested in this section, the upwind values in DCM are replaced by an all-site mean to provide a potential solution for the elimination of such problems while still applying the DCM. The mean of the column-averaged DMFs over all sampling sites ($\overline{\mathrm{XG}}_{\text{specific site}}$) is assumed to be the background concentration within the entire urban region, replacing the XCH_4 at the upwind site. The differences between the specific site and the mean of all the sites for each gas G ($\Delta\overline{\mathrm{XG}}_{\text{specific site}}$) is then evaluated, i.e.,

$$\Delta\overline{\mathrm{X}G}_{\text{specific site}} = \mathrm{X}G_{\text{specific site}} - \overline{\mathrm{X}G}_{\text{all sites}} \tag{7.5}$$

where $\mathrm{X}G_{\text{specific site}}$ is the column-averaged DMF at the respective sampling site.

Now this form of DCM is tested for the same four targeted dates (3, 4, 6 and 10 July). The distance between any two sampling sites is around 25 km. The general trends of the simulated (Fig. 7.4, blue lines) and measured (Fig. 7.4, black lines) $\Delta\overline{\mathrm{XCH_4}}$ values appear to be more similar with an RMSE of 0.6698 ppb compared to the comparison of $\Delta\mathrm{XCH_4}$ in Fig. 7.4(e–h) (RMSE of 1.3895 ppb).

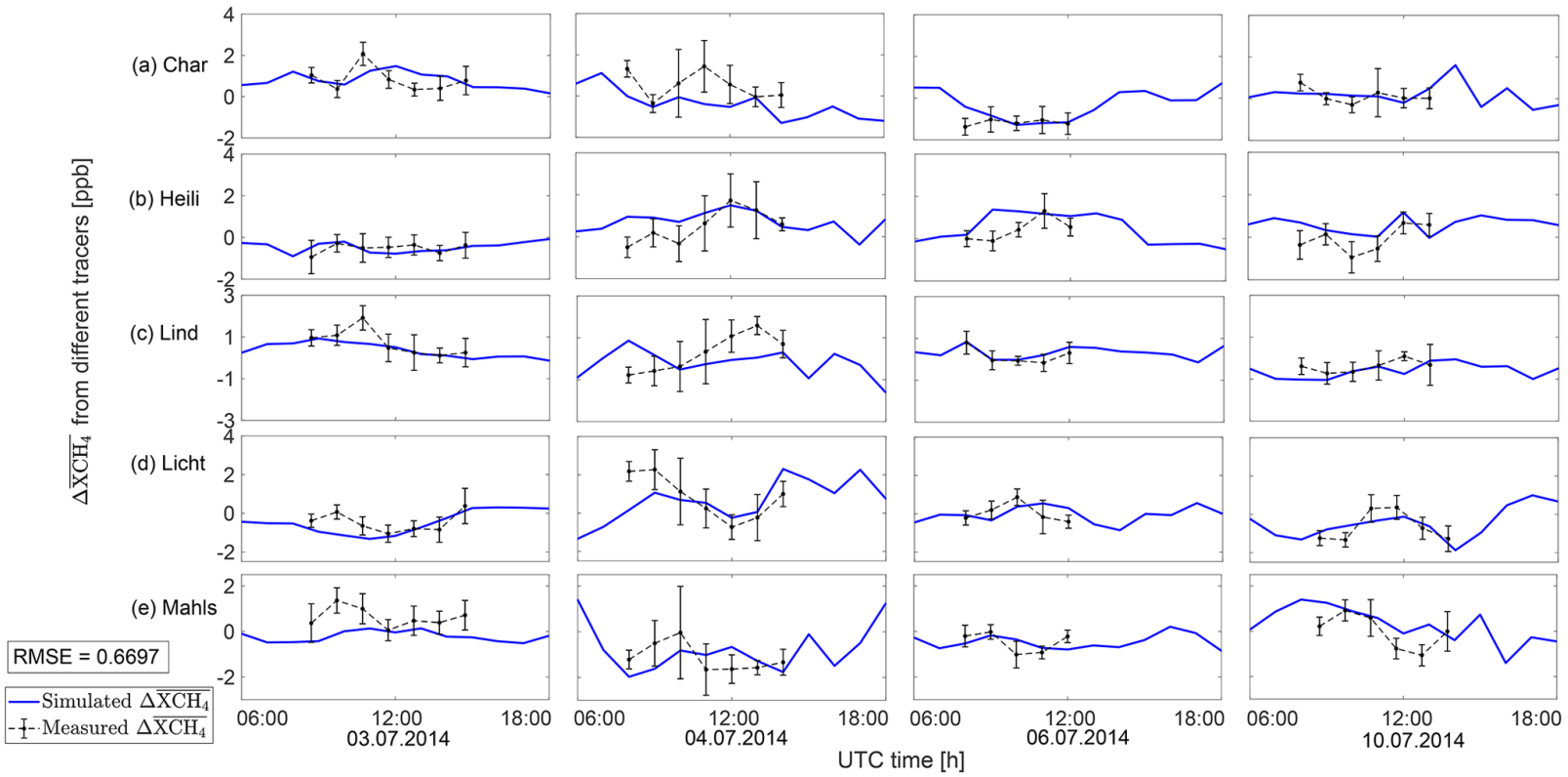


Figure 7.4.: Modeled (blue lines) and observed (black lines) site XCH_4 vs. site-mean XCH_4 data for five sampling sites: Charlottenburg (a: Char), Heiligensee (b: Heili), Lindenberg (c: Lind), Lichtenrade (d: Licht) and Mahlsdorf (e: Mahls). The black error bars in each subplot are the standard derivations of the minute values of the hourly mean.

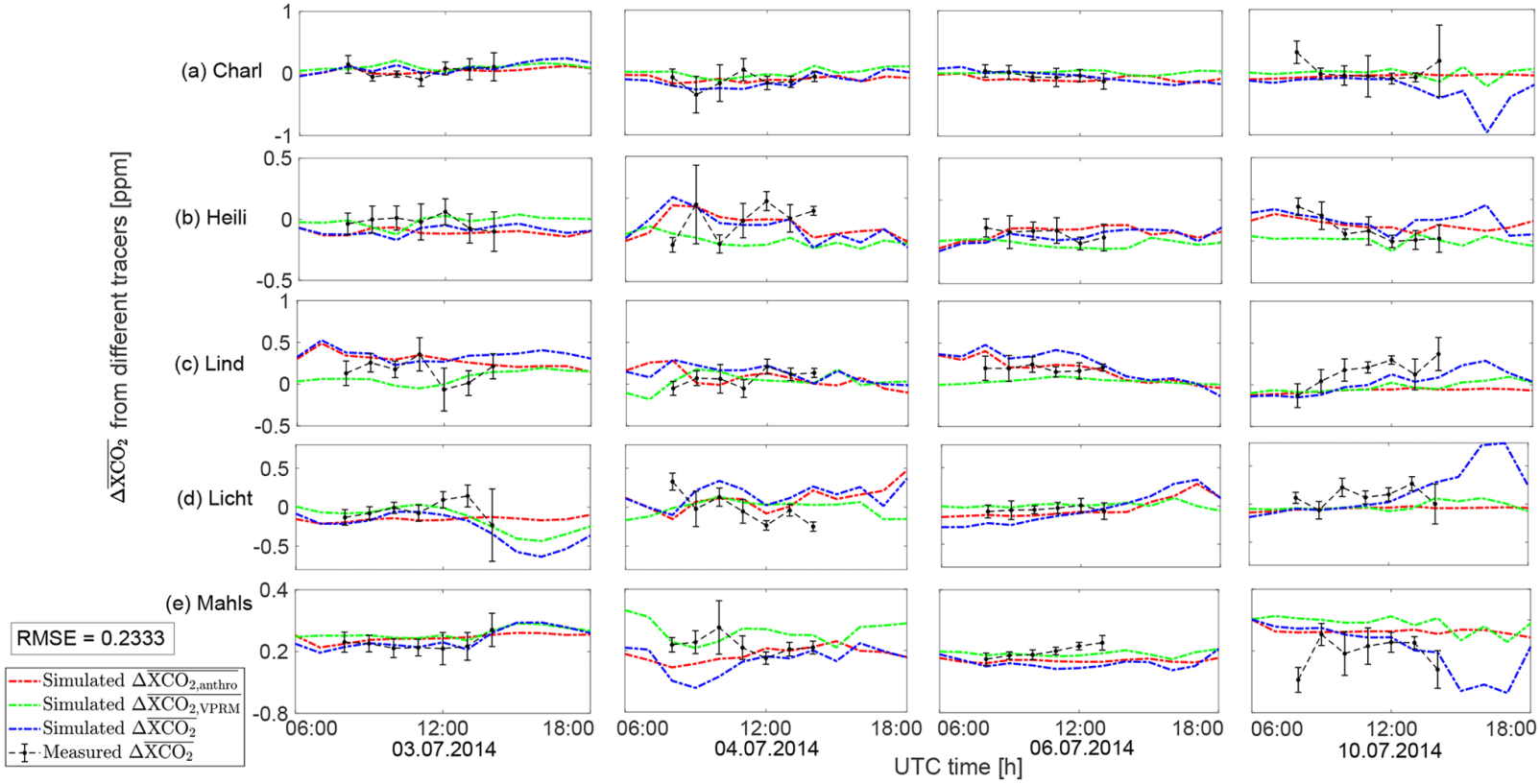


Figure 7.5.: $\Delta\overline{XCO_2}$ (blue lines for simulations and black for measurements) for five sampling sites (i.e., the difference between XCO_2 at the site and the mean XCO_2 of five sampling sites): Charlottenburg (a: Char), Heiligensee (b: Heili), Lindenberg (c: Lind), Lichtenrade (d: Licht) and Mahlsdorf (e: Mahls). Furthermore, the differences in the simulated $\Delta\overline{XCO_2}$ changes from biogenic (green lines) and anthropogenic (red lines) activities are shown. The black error bars in each subplot are the standard derivations of the minute values of the hourly mean.

The model–measurement bias can be caused by underestimated emissions from anthropogenic activities, the smoothing of actual extreme values in the simulation, and the ignorance of the line of the sun sight for the simulation. The variations in the XCH_4 at the five different sampling sites on the same day are similar (Fig. 7.4), but the measurements show more extreme values (e.g., 4 July) compared to the simulations. A further analysis in a future study is suggested to provide deeper insight into site-specific transport characteristics, e.g., understanding particle transport patterns by using a particle transport model described in Sect. 7.2.

As a final point in this analysis, simulated $\Delta\overline{\mathrm{XCO_2}}$ values for these four target dates (Fig. 7.5) is discussed. The $\Delta\overline{\mathrm{XCO_2}}$ values (blue line) on 3, 4, 6 and 10 July in five sampling sites are mainly dominated by the XCO_2 changes caused by the anthropogenic tracer (red) instead of the VPRM tracer (green). Compared to Fig. 7.3(a–d), the red line and blue line in Fig. 7.5 show a stronger similarity in their trends. With this form of DCM (compared to the original form Eq. 7.1), anthropogenic activities can be clearly shown to influence XCO_2 within urban areas. Meanwhile, the $\Delta\overline{\mathrm{XCO_2}}$ measurements (black lines) fit better with the simulation with an RMSE of 0.2333 ppm compared to the comparisons of $\Delta\mathrm{XCO_2}$ depicted in Fig. 7.3(a–d) (RMSE of 0.2973 ppm).

7.2. Assessments for Munich

As discussed in Sect. 7.1 and presented in Zhao et al. (2019) [86], DCM was shown to be a useful post-processing approach in model analysis and model-measurement comparisons, due to its ability to cancel out biases related to initial and boundary conditions. The aim of the DCM-based model evaluation for Munich is to compare the measurements to the model using DCM for selected days from 16 to 22 August, 2018, before attempting to track CH_4 emission sources based on the analysis with the help of the Stochastic Time-Inverted Lagrangian Transport model (STILT [138]; details in Sect. 7.2.2).

To begin with, the up- and downwind sites are selected based on the wind data presented in Fig. 5.7 (Sect. 5.4). The application of DCM has been improved significantly here compared to the case done for Berlin (Sect. 7.1): The transport model is made use of STILT (Sect. 7.2.2) at this point, together with wind information, to select dates for which DCM is applicable (Sect. 7.2.3). After comparing the modelled concentration gradients to the observations (Sect. 7.2.4),

the potential location(s) of unknown or underestimated CH_4 sources is further tracked by using STILT footprint contours, marking the air masses which arrive at the up- and downwind sites (Sect. 7.2.5).

7.2.1. Selection of up- and downwind sites

This method is started by dividing the measurement sites into down- and upwind sites day by day, based on measured wind directions.

As seen in Fig. 6.8, the measured concentration values at the center site of MUCCnet (TUM) are found to always be higher than the values observed at the other sites, owing to the dense distribution of emission sources close to the city center. To better understand the concentration gradients between the down- and upwind sites, here the center site chooses to be excluded.

The four remaining sites are grouped according to the wind directions observed at LMU. As shown in the wind rose of the measurements during the daytime of the simulation period (see Fig. 5.7(a.3)), the prevailing daytime wind directions (WD30) are either northeasterly or northwesterly. Table 7.1 shows the list of the down- and upwind sites for different prevailing wind conditions. The study period from 16 to 22 August can be divided into two groups of days: with northeasterly or easterly winds (during four days), and with northwesterly or westerly winds (during three days).

Concentration gradients between the down- and upwind sites are then calculated based on Table 7.1. Doing this, it is assumed that the surface prevailing winds measured in the city center are representative of the regional wind conditions over the model domain during the day, and that they are sufficiently stationary for application of DCM. When rapid, regional-scale horizontal and vertical wind shifts occur, as during a summer cold front passage around the Alps, these assumptions might fail.

Therefore, STILT [138] is set up (Sect. 7.2.2) and used (Sect. 7.2.3), as well as wind profiles from Radiosondes to assess transport patterns in particular for CH_4 concentration peaks on each day of the study period (Sect. 7.2.3). For further discussion of CO_2 and CH_4 measurements vs. simulations, and most importantly for the attempt to locate unknown CH_4 emissions (Sect. 7.2.5), the days where the assumptions for DCM are likely to be met, have thus been able to be selected. In particular, the method presented here ensures the validity of DCM for the interpretation of the CH_4 concentration peaks.

Table 7.1.: Table of up- and downwind sites depending on wind direction.

Wind Direction	Upwind Sites	Downwind Sites
Northeasterly/Easterly (NE/N)	Markt Schwaben (East) Garching (North)	Weßling (West)
Northwesterly/Westerly (NW/W)	Weßling (West)	Garching (North) Markt Schwaben (East) Höhenkirchen (South)

7.2.2. STILT model setup

Before applying DCM to any of the days from 16 to 22 August, its applicability for each day will be assessed by tracking the origin of air masses at different measurement sites with the transport model STILT (cf. Fasoli et al. (2018) [138]) and assessing the modelled vertical wind profiles used to generate footprints (see Sect. 7.2.3). Here STILT and its setup are briefly described.

In this work, STILT with *R* code base (Version 2, as available via `https://uataq.github.io/stilt/index.html#/`, last access: 11, Jan 2022) was implemented using around 168 core hours provided by the high performance computer center LRZ (see Sect. 4.5). To assure transport consistency with previously presented results, STILT was driven by the WRF meteorological fields generated for the second domain at a horizontal resolution of 2 km. In order to trace back the origin of air masses at a given spatio-temporal receptor point (corresponding to the time at which an instrument performs a given measurement), STILT uses ensembles of tracer particles which are propagated backwards in time. Specifically, the model provides the sensitivity of the analyzed slant columns to surface fluxes by counting the residence time of released particles when they traverse in the lower PBL before reaching the measurement location. These surface sensitivities are further aggregated over the STILT simulation time to produce footprint maps.

In the configuration of STILT, 500 particles have been released at 13 altitudes along the slant column namely at 20, 180, 350, 520, 700, 880, 1060, 1250, 1440, 1620, 1920, 2020 and 2220 m above the ground level for each simulation. STILT then yields so-called footprint contours (i.e., contours enclosing a certain percentage, e.g., 90 %, of the accumulated surface sensitivity) for each altitude layer. In order to yield an estimated effective footprint independent of emission height (depending on source and local conditions), these altitude-dependent footprint data have been aggregated for the different layers using the pressure differences between layers as weights [30].

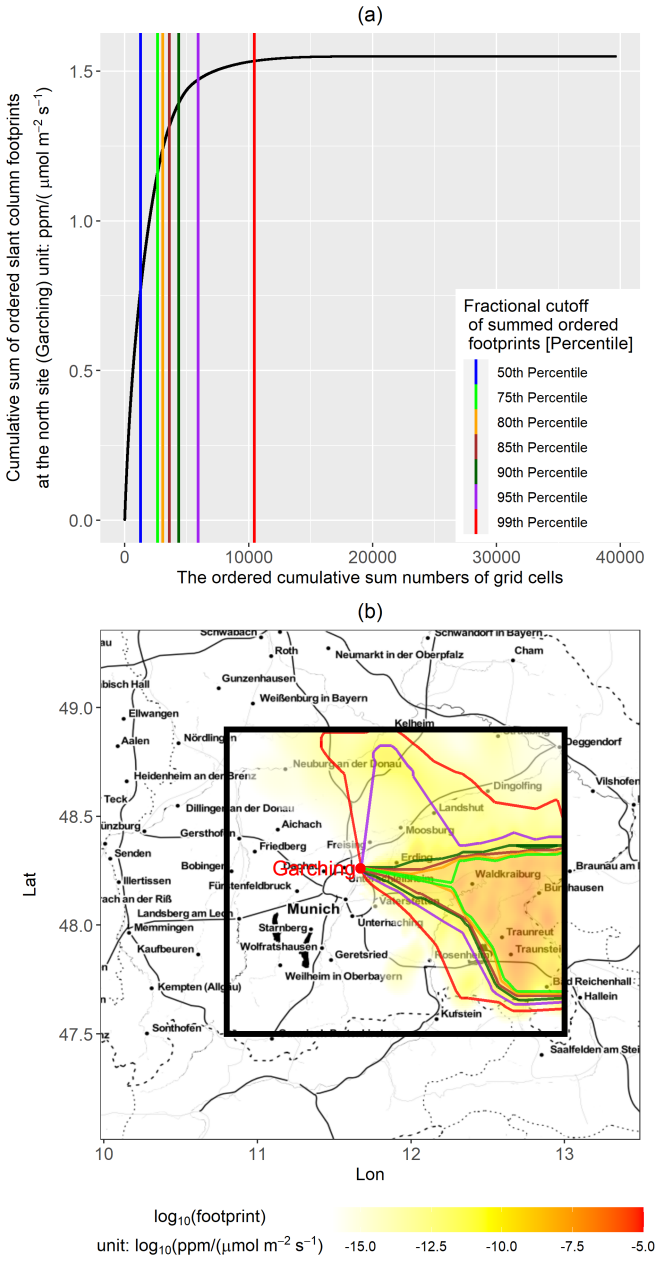


Figure 7.6.: (a) The cumulative sum of the sorted slant column footprints for one sample site and time: Garching at 11 am on 16 August, 2018 and running backward towards 18 hours. Six percentile thresholds were used for selecting points with footprints at or higher than threshold values; (b) The slant column footprint and its percentile contours: red for the 99th percentile, purple for the 95th percentile, brown for the 90th percentile, orange for the 85th percentile and green for the 80th percentile.

I have built an automated running framework using STILT to generate slant column footprints for a number of receptors over a certain period. Thanks to the support of my Bachelor student Moritz Makowski [139], the generated footprints in Munich can be found in our team webpage via `https://atmosphere.ei.tum.de/`. This framework has been applied for several cases, including Nile Delta, Egypt [27], San Francisco Bay, USA [42], [140], Hamburg, Germany [141] and Munich, Germany [142].

To define the percentile contours of footprints used in Munich, the following steps was performed:

(i) the values of the slant column footprint map were ordered and the cumulative sum of these sorted values was calculated (see Fig. 7.6.(a));

(ii) based on the plot of the cumulative sum of the sorted values, a threshold under a specific percentile was defined to exclude points with lower sensitivity, e.g., the 90th pecentile (brown line in Fig. 7.6.(a));

(iii) the footprint contours of the different percentiles were defined.

The generation of footprint percentile contours in this work is based on the approach used in Dayalu et al. (2020) [143]. Table 7.2 lists the receptor time used to generate the footprint contours in Figs. 7.6 & 7.8. As mentioned in Sect. 6.4.2, the receptor time (i.e., the time when the air parcels are released from the receptor) is defined based on the observed peak time during the day, and when this observed peak time cannot be read from Figs. 6.9 & 6.10, it is defined by calculating the transport time from a known upwind site to a relatively downwind site. However, when Höhenkirchen (south) site should be regarded as an upwind site under the Northeasterly/Easterly wind and the observed peak time is not readable from the figures, its receptor is not able to be estimated properly (see '-' in Table 7.2).

Figure 7.6 shows the footprint contours at the up- and downwind sites from 17 to 22 August, 2018. In this work, the footprint contours are generated to show the main area, across which air parcels released from the receptor travel, and can also be used to reflect the prevailing wind over the domain. The footprint contour generated by a steady prevailing wind always starts at the receptor and spreads gradually in the direction, from which the air parcels come. However, the unsteady wind results in the disordered travel of the air parcels, making it hard to see the location of the receptor and to specify the prevailing wind using its footprint contour. That is, with DCM, the air pack should theoretically pass

Table 7.2.: List of receptor time for five sites from 18 to 22 August, 2018.

Date	Receptor time (UTC)				
	Garching (north)	Markt Schwaben (East)	TUM (Center)	Weßling (West)	Höhenkirchen (South)
20180816	11:00	11:00	12:00	14:00	11:30
20180817	11:00	11:00	13:00	11:00	15:00
20180818	11:00	11:00	12:00	14:00	11:00
20180819	11:00	12:00	11:00	15:00	11:30
20180820	13:00	13:00	12:00	9:00	13:00
20180821	10:30	11:00	12:00	14:30	-
20180822	13:00	11:00	12:00	15:30	-

by the upwind site and further arrive at the downwind site with a relatively stable wind, after travelling through the urban area in which most of emissions are located. In such cases, the concentration enhancement between the down- and upwind sites can be led by the sources inside the city. Therefore, the applicable dates need to be selected within the measurement period to satisfy the assumption of this approach.

7.2.3. Date selection

Leveraging the footprint contours from STILT, the differences in the origin and path of air masses arriving at up- and downwind sites in defined ranges of time can be determined. Understanding these differences is a key prerequisite for determining the location of potential unknown GHG sources based on noteworthy signals in the downwind-upwind concentration gradients. Whenever air masses reaching up- and downwind sites have very different areas of influence, the upwind site cannot be used as a relative background site when calculating concentration gradients. When the footprints fully overlap, however, such that air passes over the upwind to the downwind site, and additional GHG contributions are from in between, DCM can be used.

But even when this condition is not strictly met, if differences in the footprint areas are small, the small non-overlapping parts are potential locations for unknown GHG emitters and sinks to be pinned down, rendering DCM effective. Signals coming from the overlapping area of the footprints, in contrast, will be visible at multiple measurement sites with a characteristic time delay. Clearly, for all this to hold, it requires to be checked whether the footprints are realistic. The main prerequisite for this is the accuracy of the WRF wind fields,

driving STILT, at different heights. This criterion is checked at the end of this section.

Assessment of air mass origin (footprints) at different sites on different days

The approach in Sect. 7.2.2 can be applied to understand up-/downwind differences and obtain information about GHG sources and transport in the target area. Here, the following strategy has been adapted: for each measurement site, the footprint contours are computed with the receptor time in STILT set to the time of the daily XCH_4 peak value [1]. Then, only days where the overlap of these footprints is large, are accepted. The peak times of the stations are usually different by only a few hours. The strategy results in three outcomes:

(i) days with unstable wind conditions in time or with large variations of wind directions from one site to another are excluded (unless the two effects compensate quite exactly, which is unlikely);

(ii) the large overlap and small differences in footprint contours allow for a clear localisation of potential origins of differences, as discussed above;

(iii) whether the peak is of the same origin at all sites, can be understood in this case;

CH_4 signals choose to be analysed at the end of this case (Sect. 7.2.5), as this promises a realistic possibility of tracking human emission sources (cf. Sect. 6.4.2) in this exploratory work. In contrast, the current state of the art makes it more difficult to trace anthropogenic effects in CO_2 signals, where biogenic activity plays a much larger role. The strategy described above is applied to all measurement days, and finally selected 16, 20 and 21 August 2018 as days suitable for further analysis (Sects. 7.2.4 & 7.2.5). Figure 7.7 and the Appendix (in particular Fig. A.2) lay out the reasons for our decision and show all the footprint contours. In the following paragraphs, the examples from Fig. 7.7 (16 and 22 August) further as typical days deemed appropriate (16 August) or inappropriate (22 August) for further analysis using DCM are discussed. Besides the footprints, Figure 7.7 also shows the peak times used as receptor times for STILT.

[1] In a few cases where the peak times have been hardly detectable, they have been inferred using the peak times at nearby sites and considering time delays derived from the daily surface wind speeds from LMU.

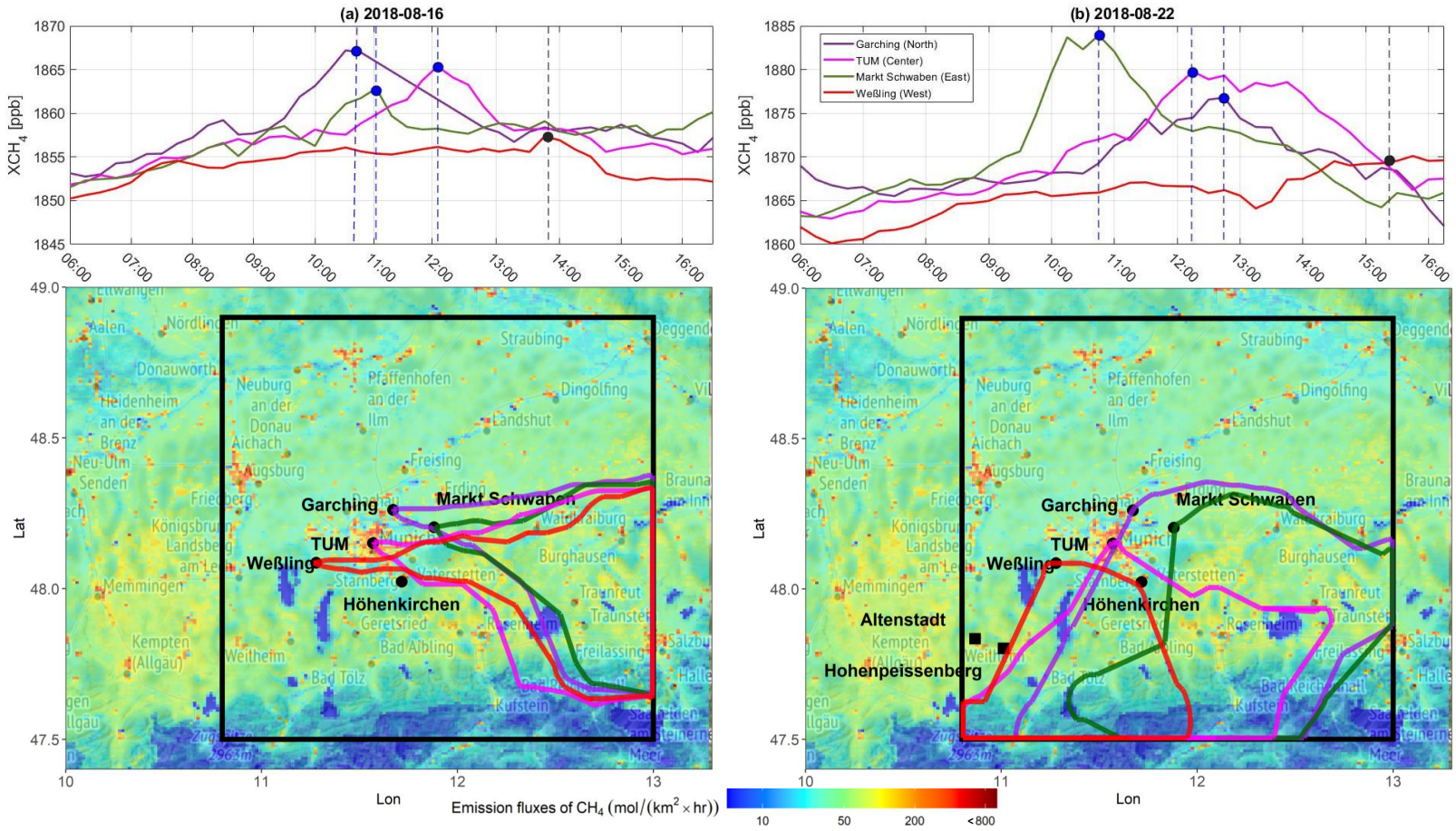


Figure 7.7.: Observed XCH_4 over time (upper panel) and the 90th percentile contours of column footprints (lower panel) on (a/left) 16 and (b/right) 22 August at up- and downwind sites with different colors: red for Weßling (West), green for Markt Schwaben (East), purple for Garching (North) and pink for TUM (Center). The peaks in the observations are marked by dots (blue: clear peaks, black: inferred peaks, see main content/footnote). Two black squares in (b) mark the measurement sites of IGRA (cf. Sect. 7.2.3). The background maps use tiles from Stamen Design (`https://maps.stamen.com/`, under CC BY 3.0, with data by OpenStreetMap, under ODbL, from 2021 Dec. 21). The map colouring reflects the emissions from the initial inventory (i.e. TNO_GHGco).

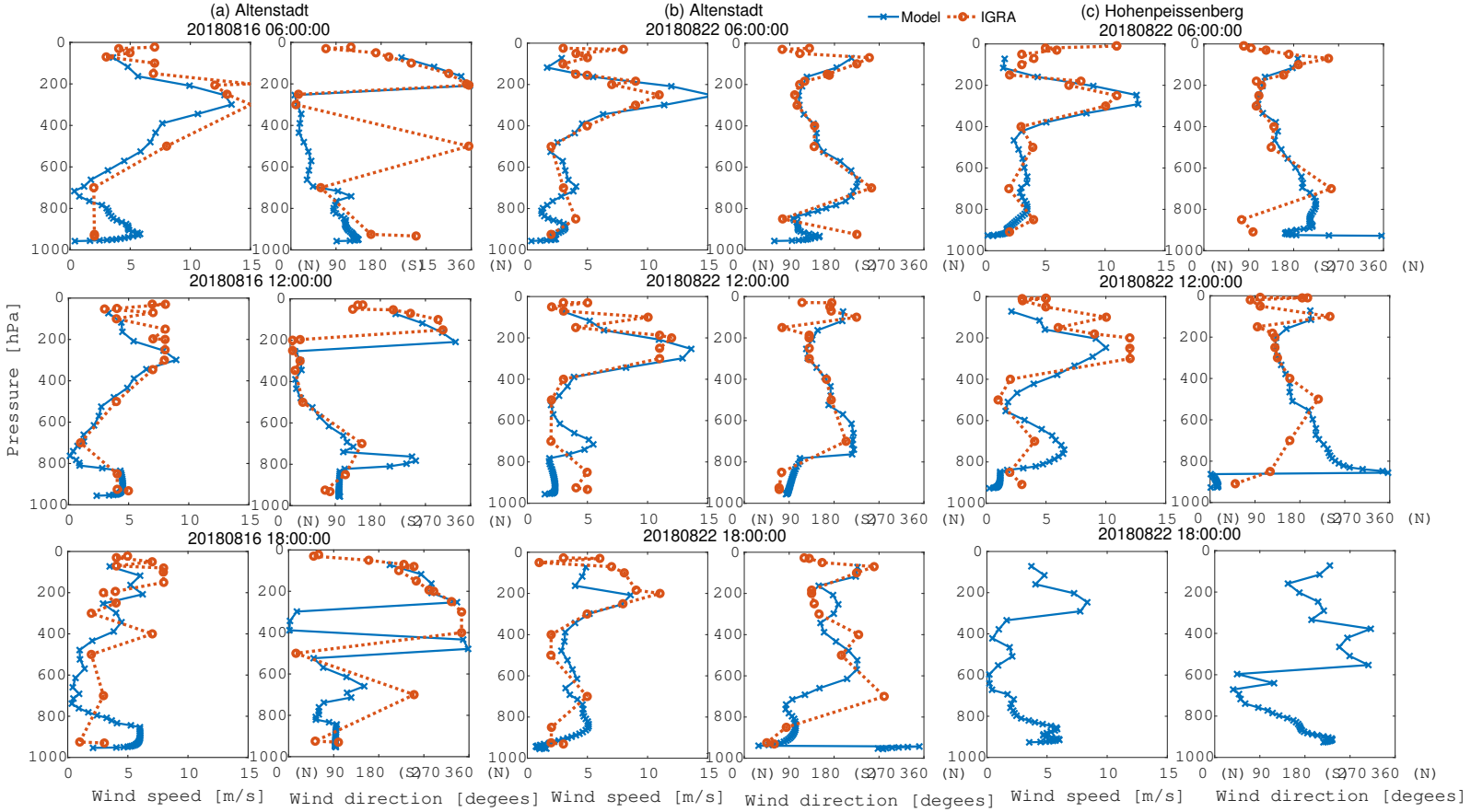


Figure 7.8.: Vertical wind comparison between the model (orange dashed line) and the measurements (blue solid line) from two IGRA stations, i.e., (a,b) Altenstadt on 16 and 22 August, and (c) Hohenpeißenberg on 22 August.

On August 16, with easterly wind prevailing, the instruments deployed upwind (Garching/Markt Schwaben) captured peak CH_4 signals in the first half of the day. A similar signal was then seen at TUM about an hour later, while the sensor at the downwind side (Weßling) did not detect a major peak. Using the knowledge of station locations and the observed peak times (blue dots in Fig. 7.7(a)), a peak at Weßling (Fig. 7.7(a), black dot) is predicted as follows: upwind and central sites captured the peaks at 11:00 and 12:00 UTC, respectively. These air masses would then possibly have reached the downwind site after another two hours (estimated from wind speeds as given in Table 5.1), corresponding to the distance between TUM and Weßling, i.e., at 14:00 UTC (Fig. 7.7(a), black dot). Note that the receptor times for all up- and downwind sites and dates, determined from observed peaks or estimates in this manner, are listed in Table 7.2. All in all, an interesting difference between the peak behaviour at the sites and a footprint situation warranting a useful analysis can be seen.

In the morning of 22 August (see Fig. 6.10), a similar rising signal to that on 16 August was detected at the upwind site (Markt schwaben) until 11:00 UTC, and was then detected at the North (Garching) and center (TUM) sites. On the afternoon, an increase in concentrations could be found at the downwind site (Weßling) until late afternoon (around 15:00 UTC). Thus, the particles is released in STILT at 11:00 UTC in Markt schwaben, 13:00 UTC at TUM and Garching, 15:00 UTC in Weßling. As can be seen in the footprint contours on 22 August in Fig. 7.7(b), the air masses passing by the downwind site (Weßling, red line) did not come from the same upstream as the upwind sites. The footprints (with receptor times corresponding to the blue dots in Fig. 7.7(b), top panel) are widespread, and any 'tracing experiment' would be poorly controlled. That is, in the model, Weßling (west) should not be considered as the relatively downwind site when taking Garching (north) and Markt Schwaben (east) as the upwind sites, despite the fact that analysis of the measured signals alone would suggest otherwise. The case does not conform to the definition of the up- and downwind sites in DCM and this date does not lend itself to the application of this approach to assess the model behavior.

In order to assess whether the WRF meteorological fields lead STILT to produce realistic footprints or not, the modelled height-dependent wind fields are further evaluated using observations. This assesses their accuracy in the vertical dimension more deeply than the wind-field comparison shown in Sect. 5.4.

Assessment of vertical wind profiles for calculating realistic footprints

Modelled and measured wind speed and direction profiles are shown in Fig. 7.8 for 16 and 22 August. The comparison uses measurements from two sites (Altenstadt and Hohenpeißenberg, marked as black squares in Fig. 7.7(b)), provided by IGRA. This archive collects radiosonde and pilot balloon observations along significant vertical levels historically and in near-real-time from around 800 distributed stations worldwide [144]. On both days at Altenstadt (Fig. 7.8(a,b)), the model reproduces the observed winds well, especially at noon, while outliers in wind direction always exist at the lowest level in the morning and some mismatches appear at 18:00 UTC. Regarding the wind profiles measured at Hohenpeißenberg, the modelled wind direction at higher altitudes agree quite well with the observations, with mismatches close to the ground level. It is worth noting that our domain is close to the Alps, with complex topography making meteorological modelling more challenging. The results need to be interpreted in this content.

7.2.4. Model-measurement comparison

After checking the prerequisites in the previous sections, the analysis of the differential concentrations (gradients) will focus on the selected dates of 16, 20 and 21 August.

Figures 7.9 and 7.10 show $\Delta XCO^{S}_{2,sla}$ and $\Delta\Delta XCH^{S}_{4,sla}$, respectively, for the days selected based on the prevailing wind directions. The modelled concentration gradients of $\Delta XCO^{S}_{2,sla}$ between the down- and upwind sites (blue soild in Fig. 7.9) are driven by both biogenic activities (light green dashed) and human activities (grey dashed). The biogenic part can be attributed to the special spatial distribution of biogenic sinks in Munich and its surroundings (see Fig. 7.11): The southeastern and southwestern parts around Munich are more biologically active and have greater carbon sinks, compared to the other areas. This is an interesting difference to Berlin, where the Berlin study [86] showed no such signal, corresponding to a relatively even distribution of biogenic fluxes over Berlin. In terms of $\Delta XCH^{S}_{4,sla}$, the variation of the modelled values over the day (blue solid in Fig. 7.10) is driven by human activities (grey dashed). In the model-measurement comparison of $\Delta XCO^{S}_{2,sla}$, the model was able to reproduce the general variations when comparing to the measurements (red solid with '+' in Fig. 7.9).

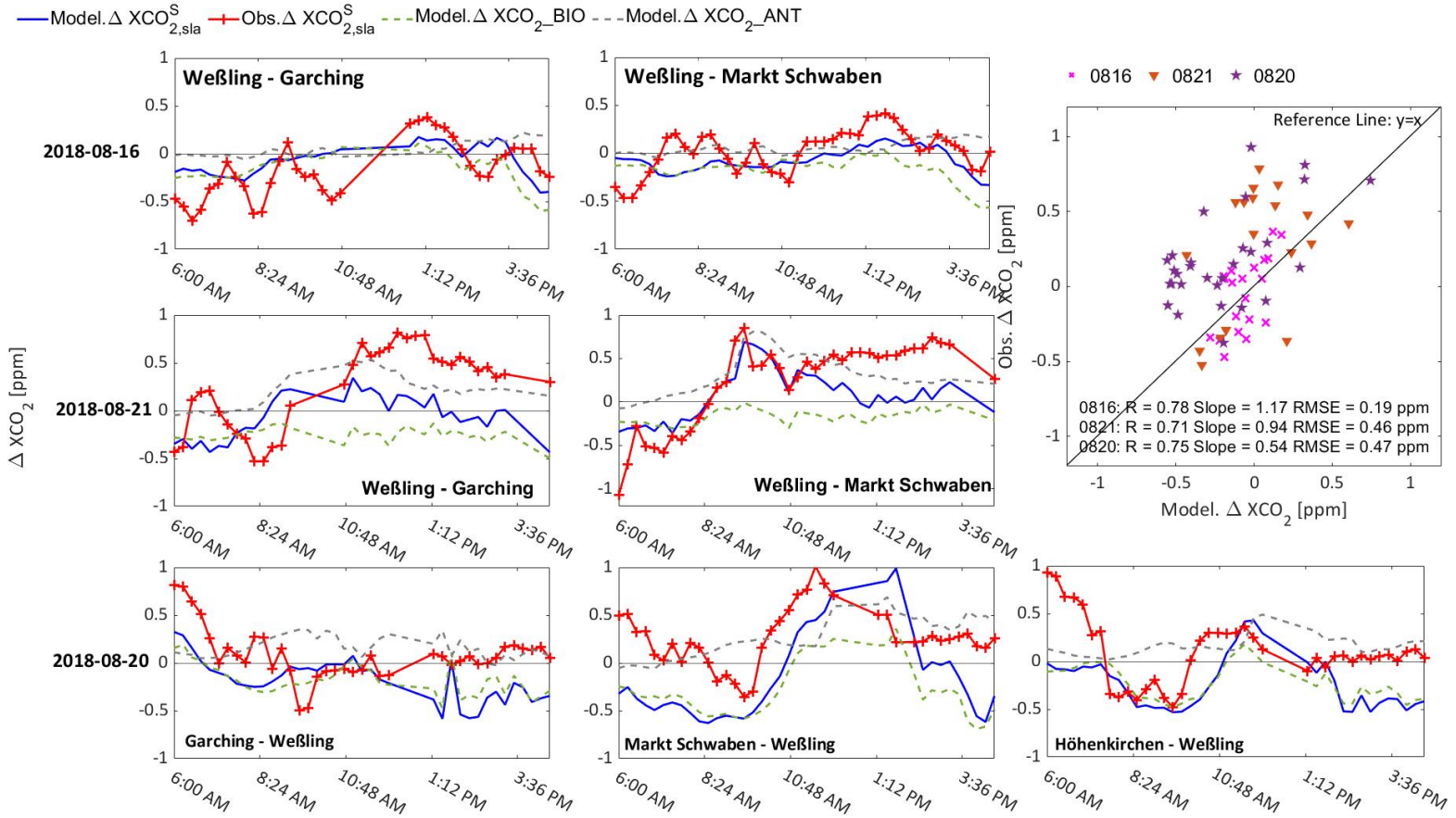


Figure 7.9.: Time series of $\Delta XCO^{S}_{2,sla}$ for three targeted days and their scatter plot: 16 and 21 August with NE/E winds in the upper two rows, and 20 August with W/NW winds in the bottom row. The column concentration differences between the down- and the upwind sites are plotted as red solid lines for measurements, blue solid for the modelled full signal values, light green dashed for the simulated biogenic signal and grey dashed for the modelled contributions related to anthropogenic activities.

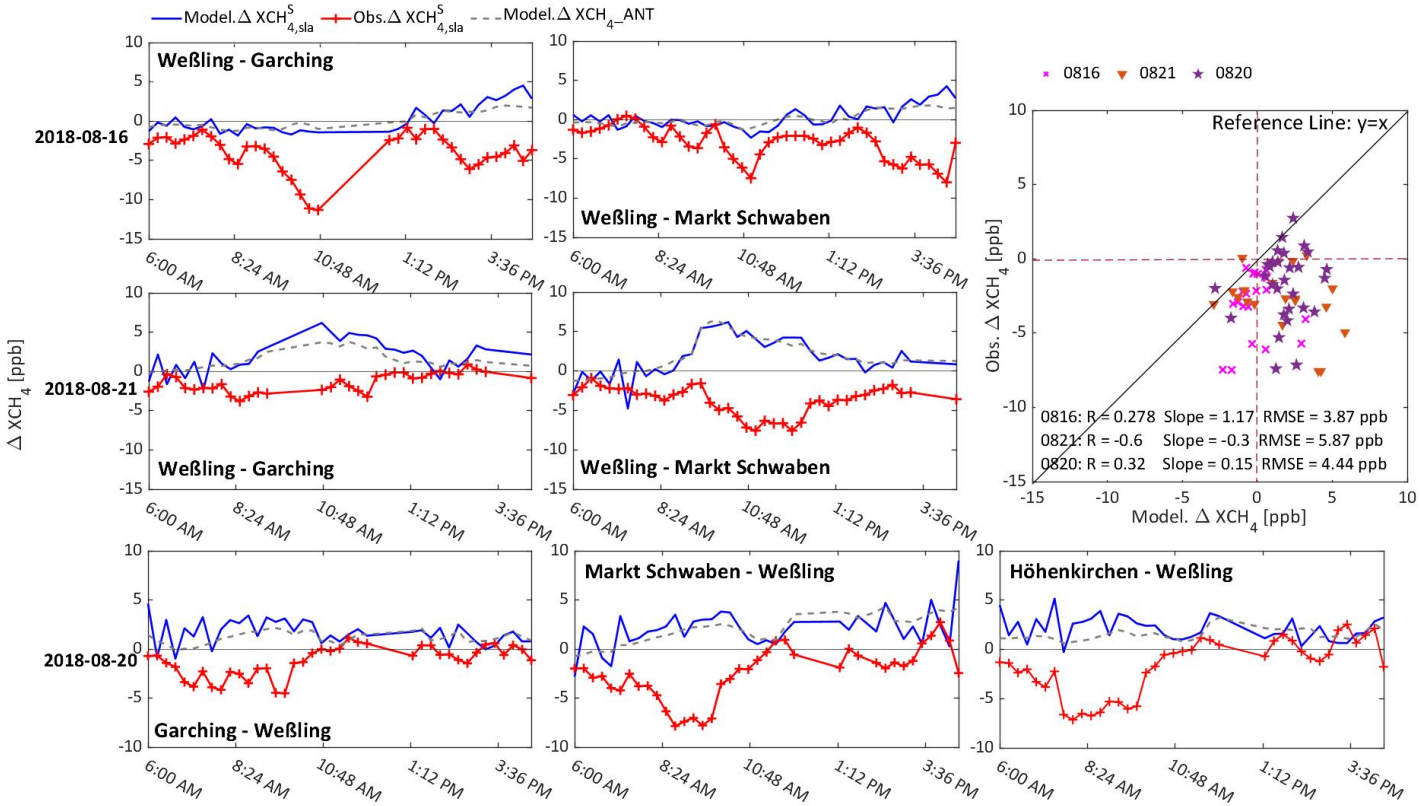


Figure 7.10.: Time series of $\Delta XCH^{S}_{4,sla}$ for three targeted days and their scatter plot: 16 and 21 August with NE/E winds in the upper two rows, and 20 August with W/NW winds in the bottom row. The column concentration differences between the down- and the upwind sites are plotted as red solid lines for measurements, blue solid for the modelled full signal values and grey dashed for the modelled contributions related to anthropogenic activities.

However, mostly underestimated the concentration gradients in its order of magnitude before around 9:00 UTC and after14:00 UTC, with a Pearson correlation coefficient of 0.74 and RMSE of 0.37 ppm. This could be caused by underestimated concentration gradients from biogenic fluxes in and around the city (see light green dashed). Especially, underestimating RES during nighttime (details in Sect. 6.4.2) could result in the underestimation of concentration gradients of CO_2 in the early morning.

In terms of $\Delta XCH^{S}_{4,sla}$, shown in Fig. 7.10, the modelled values generally show quite slight variations and are mostly positive during the day, while the measured concentration gradients between the down- and upwind sites are mainly negative. With NE/E winds, the instruments always measure strong signals at the upwind sites (i.e., Garching (north) and Markt Schwaben (east)) compared to the downwind sites, which cannot be reproduced by the model. As a large methane sink over the city is not expected, the most likely cause for this phenomenon is missing or underestimated emission sources located upstream of the upwind sites, located somewhere to the northeast or east of the Garching and Markt Schwaben stations.

Likewise, in the case with W/NW winds, the negative measured concentration gradients between the three down- and one upwind sites are found with -1.89 ppb in daily means and the model fails to reproduce these signals. Again, the measured column concentrations at the upwind site (i.e., Weßling) are generally higher than at the downwind sites. Especially in the morning of 20 August, a clear strong increase was captured at the upwind side (see Fig. 7.10). However, none of these features could be replicated by the model. The presence of an unknown or underestimated source of emissions located upstream areas Weßling is postulated as the most likely explanation.

7.2.5. Unknown emission sources localization

To further localize the underestimated or unknown emissions at the upstream areas of both upwind sides (i.e., the western area of Weßling and the eastern area of Markt Schwaben), the footprint contours are used to interpret the transport of air masses. Figure 7.12 maps the footprint contours of the up- and downwind sites with two different wind conditions (a) for 16 and (b) 20 August and their receptor times are listed in Table 7.2.

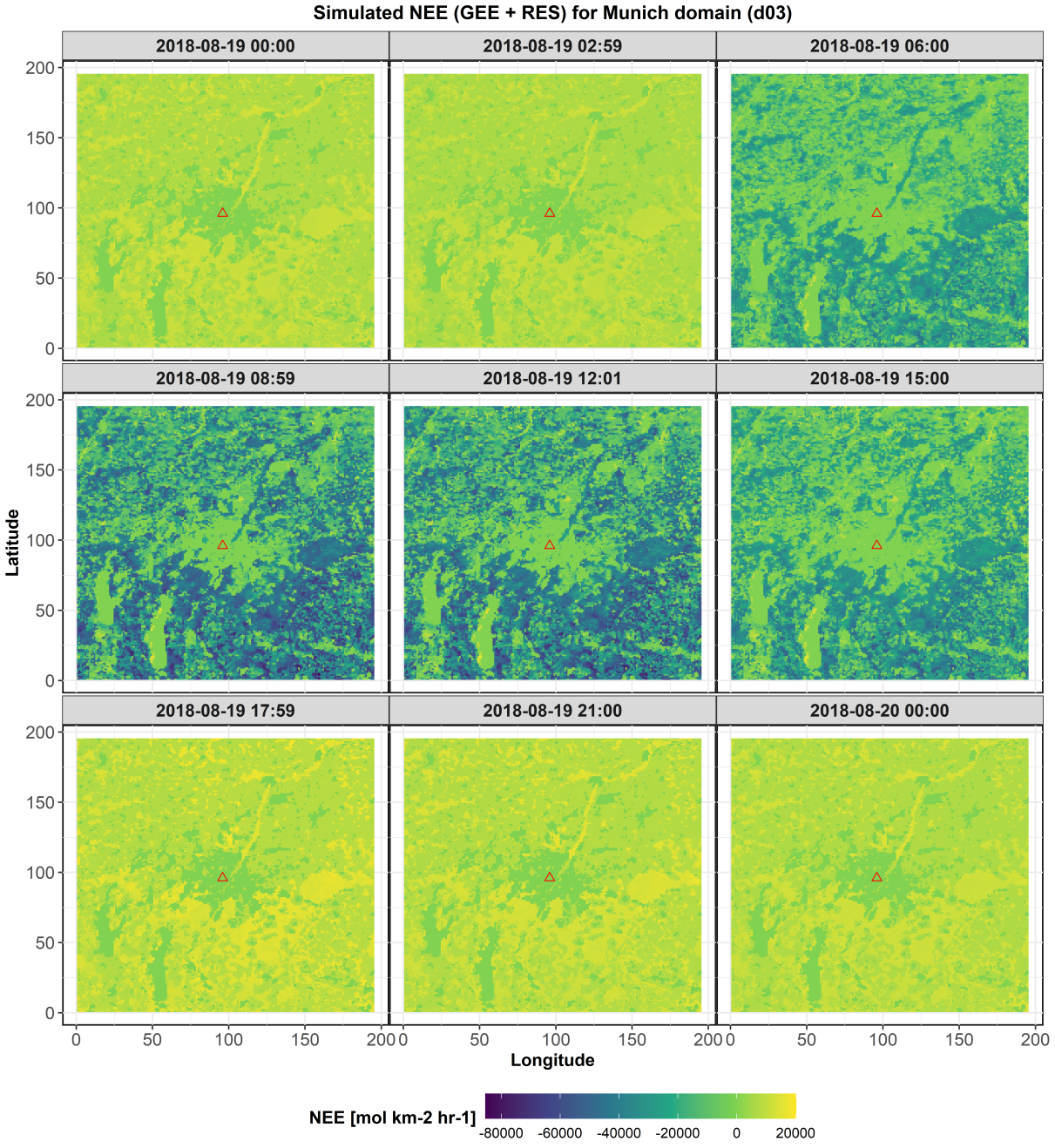


Figure 7.11.: The map of NEE on 19 August for the innermost domain (d03), modelled with VPRM using the new land use categories mentioned in Sect. 4.3.

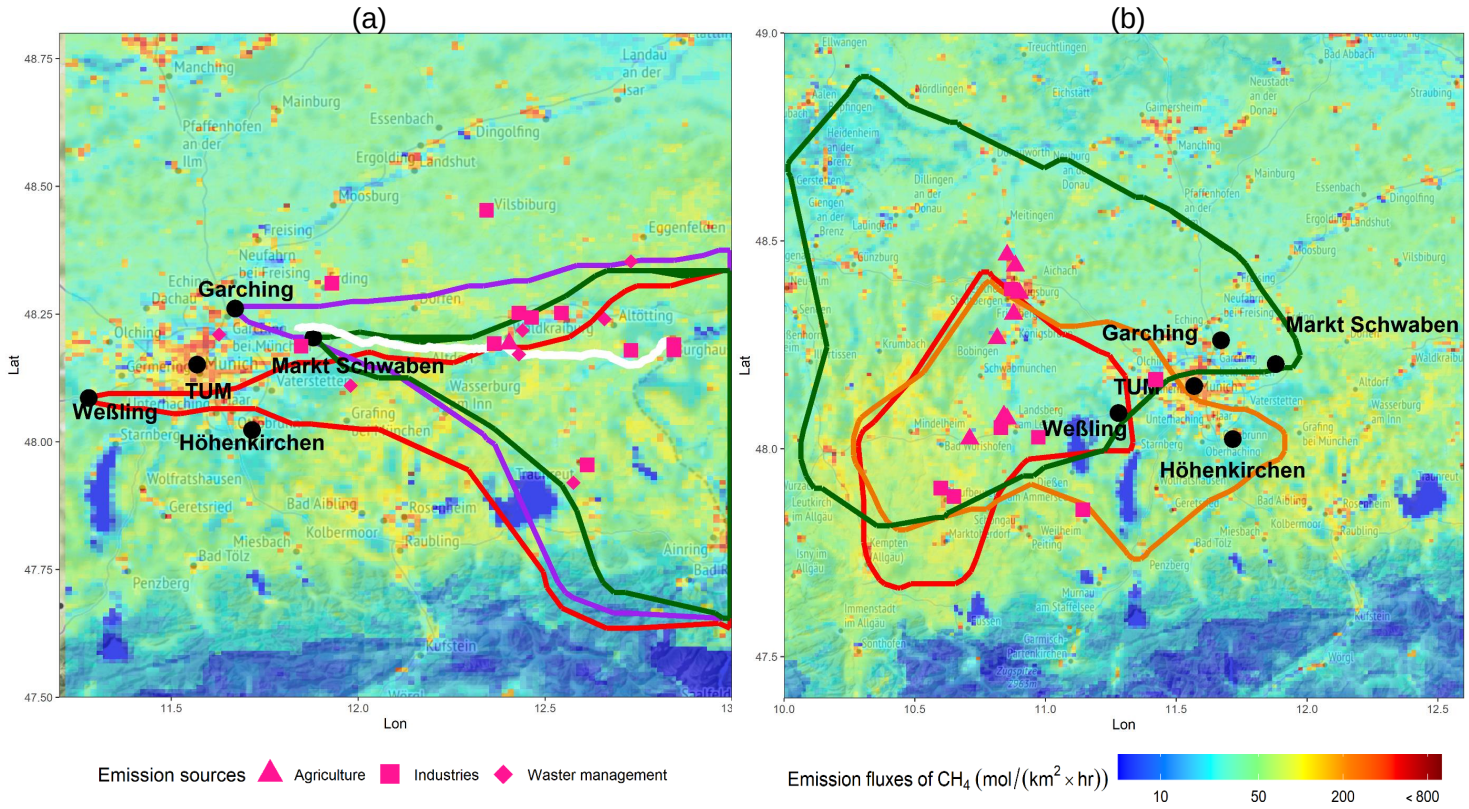


Figure 7.12.: Distribution of known emission sources located in the (a) eastern and (b) western area of Munich and the footprint contours for down- and upwind sites on (a) 16 and (b) 20 August: green for Markt Schwaben (East), red for Weßling (West), purple for Garching (North) and yellow for Höhenkirchen (South). The white solid line in (a) shows the location of the Burghausen-Finsing-Amerdingen high pressure natural gas pipeline. The background maps use tiles from Stamen Design (`https://maps.stamen.com/`, under CC BY 3.0, with data by OpenStreetMap, under ODbL, from 21 Dec, 2021). The map colouring reflects the emissions from the initial inventory (i.e., TNO-MACCco).

As mentioned in Sect. 6.5.2, the major contributors of CH_4 related to human activities are waste management, agriculture and industry. A number of sources from these sectors in and around the areas covered by the footprint contours are identified based on the European industrial emissions portal (available as `https://industry.eea.europa.eu/`, last access: 14 Dec. 2021), the initial emission inventory and local knowledge, and marked in Fig. 7.12. The waste management here refers to landfill and wastewater treatments, the emissions induced by agriculture come from livestock and the industry emissions are from heat and gas production, manufacturing etc. Another potential source of the observed signals that was not represented in the modelling framework was a high-pressure natural gas pipeline that passes through the eastern section of our domain. This pipeline was reportedly under construction and tested in August, 2018 (constructed by Bayernets GmbH).

Clearly, the exact cause of the observed peaks not reproduced in the model cannot not be determined yet. However, with a longer observation record and refinements to the modelling approach, it is possible to track down strong emitters of GHGs, which could inform mitigation strategies. Here, the long-term measurements from MUCCnet, which cover a wide range of weather conditions and complete season cycles, will help to complement and improve general inventories.

To learn more about emissions that may be missing from or underestimated in the initial emission inventory used in the model, the particle transport model STILT driven by the WRF meteorological fields was used to obtain the transport patterns backward in time. Through comparing the modelled $\Delta\mathrm{XCH_4}$ to the measurements, the area of unknown or underestimated sources of emissions can be suggested on the basis of the 90th percentile contours of flux footprints. This approach can be useful for optimizing the local emission inventory, especially when applied to a year-round time series.

8. Conclusion and outlook

In this work, an urban modelling infrastructure based on WRF has been developed to accompany column measurement networks or campaigns. It has been based on modern scientific computing solutions for reproducibility, i.e., containerisation with *Docker* for complex task orchestration and data management solutions. To complement the first worldwide permanent column measurement network designed in Munich (MUCCnet), this modelling infrastructure is currently applied in Munich, while it can be easily adapted to other urban areas by updating model configurations in WRF and providing suitable initial and boundary condition for tracers over the targeted periods.

Two case studies have been carried out in this work to assess the performance of this WRF-based infrastructure in the vicinity of urban areas, of Berlin and Munich in particular. The Berlin study has addressed a fundamental study [51], in that it has proven that WRF coupled with GHG modules is a suitable tool for precise GHG transport in urban areas. Building upon this, the usefulness of modelled concentration estimations has been demonstrated for the case of Munich in conjunction with MUCCnet [145]. Compared to the Berlin case [51], the modelling infrastructure used in Munich has been improved in terms of the introduction of tagged tracers and improved model inputs. Based on the updated version of WRF and WRF-Chem, urban canopy information has been taken into account and more emission tracers (e.g., biogenic emissions from wetland for XCH_4, traffic emission and strong point source emissions in urban areas) have been separated. The major contributions to the modelled concentrations over both target areas have been identified and discussed in this work.

Moreover, this work has shown that a model accompanied by measurements can highlight capabilities of DCM for analyzing urban CO_2 and CH_4 emissions and identify the main driver processes of such emissions. For both cases, this method emerges as an effective method of comparing models to observations independently of biases caused, e.g., by initial conditions. It allows the us-

age of the high-resolution WRF model to roughly detect and understand major sources of GHG emissions in and around urban areas. This work concludes with a refined application of DCM, aided by air-mass transport tracing with the STILT model. In an exploratory application to CH_4, this enabled us to identify unexpected signals in the measurements and some potential sources as well.

In summary, the modelling infrastructure presented in this work has shed new light on the interpretation of column measurements from observation networks or short-term campaigns. It shall foster a deeper understanding of GHG sources and sinks, and their transports over cities. To this end, its applications can be extended, e.g., to an interpretation of year-round MUCCnet column measurements in the future. This will definitely provide a deeper understanding on urban scale emissions.

In terms of the modelling infrastructure itself, the WRF outputs can be more detailed validated and compared with Celiometer and LiDAR and WRF data assimilation methods can be used to assimilate wind LIDAR data [146] and aircraft data from In-service Aircraft for a Global Observing System (IAGOS [1]) into the model. This will help provide more accurate wind information for tracer transports in WRF-Chem and the simulation of air particle trajectories in STILT. With regards to the underestimation of biogenic fluxes modelled in VPRM (see Sect. 6.4.2), some studies [101], [102] have demonstrated the benefits of more complicated calculations in the estimation of RES recently. These studies have, in particular, proven to improve the nighttime behavior of VPRM. It is therefore necessary to incorporate these development of VPRM in WRF-Chem when updating this infrastructure in the future.

As the main principal pilot city in the ICOS project, supported by the EU (Horizontal 2020 project ICOS Cities/PAUL, Grant ID:101037319), Munich has been selected for the study of GHGs on urban scales, which will contribute to addressing the European Green Deal challenge. Ongoing work includes applying this modelling infrastructure to the COVID-19 pandemic period of 2019 to 2021 to further assess the model by comparing it to the measurements from MUCCnet and IAGOS . It will be helpful to understand how the COVID pandemic impacted urban GHG emissions.

Another further promising direction for future studies may be the application

[1] IAGOS database is available at `https://www.iagos.org/iagos-core-instruments/package2d/` and supported by Dr. Christoph Gerbig

of DCM and model-based analysis to satellite measurements to assess gradients across column concentrations with a dense spatial sampling. The Orbiting Carbon Observatory-2/3 (OCO-2/3), operated by the National Aeronautics and Space Administration (NASA), has chosen Munich as a target area, which means that the satellite instruments will view the region continuously when they pass over the city. This infrastructure shall help interpret the co-located OCO-2 and OCO-3 satellite measurements and effectively track GHG emissions, contributing to the improvement of inventories.

In a more basic methodical approach, this WRF-based modelling infrastructure may also be combined with microscale atmospheric transport models to simulate crucial details of emission sources and transport patterns precisely, with the aim of tracing urban GHG emissions. All these promising directions will help our aim of gaining detailed knowledge on urban GHG emissions. This will hopefully be a significant contribution to the mitigation of the causes of climate change, carbon emissions and the problem caused by them.

A. Appendix

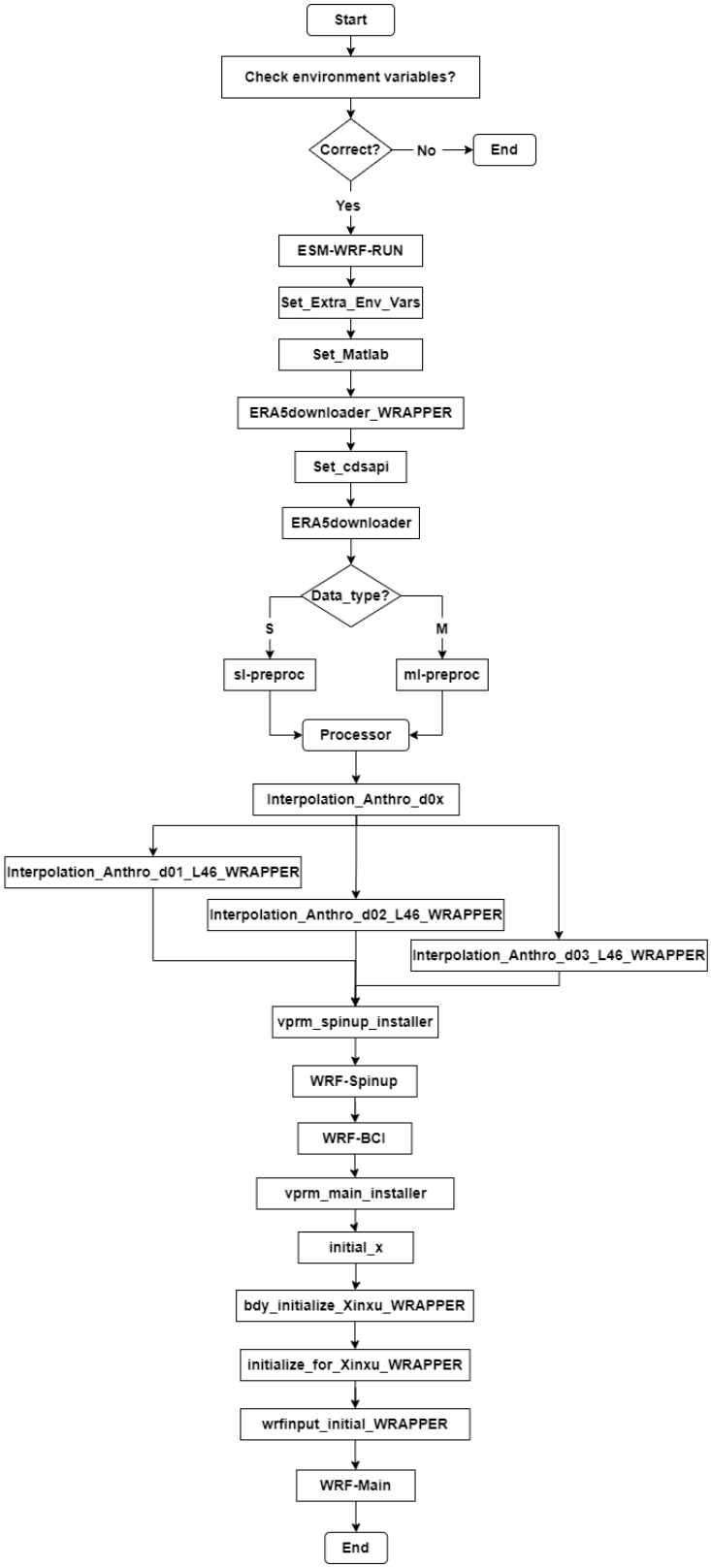


Figure A.1.: Workflow of the ordered scripts running in the *Docker* container

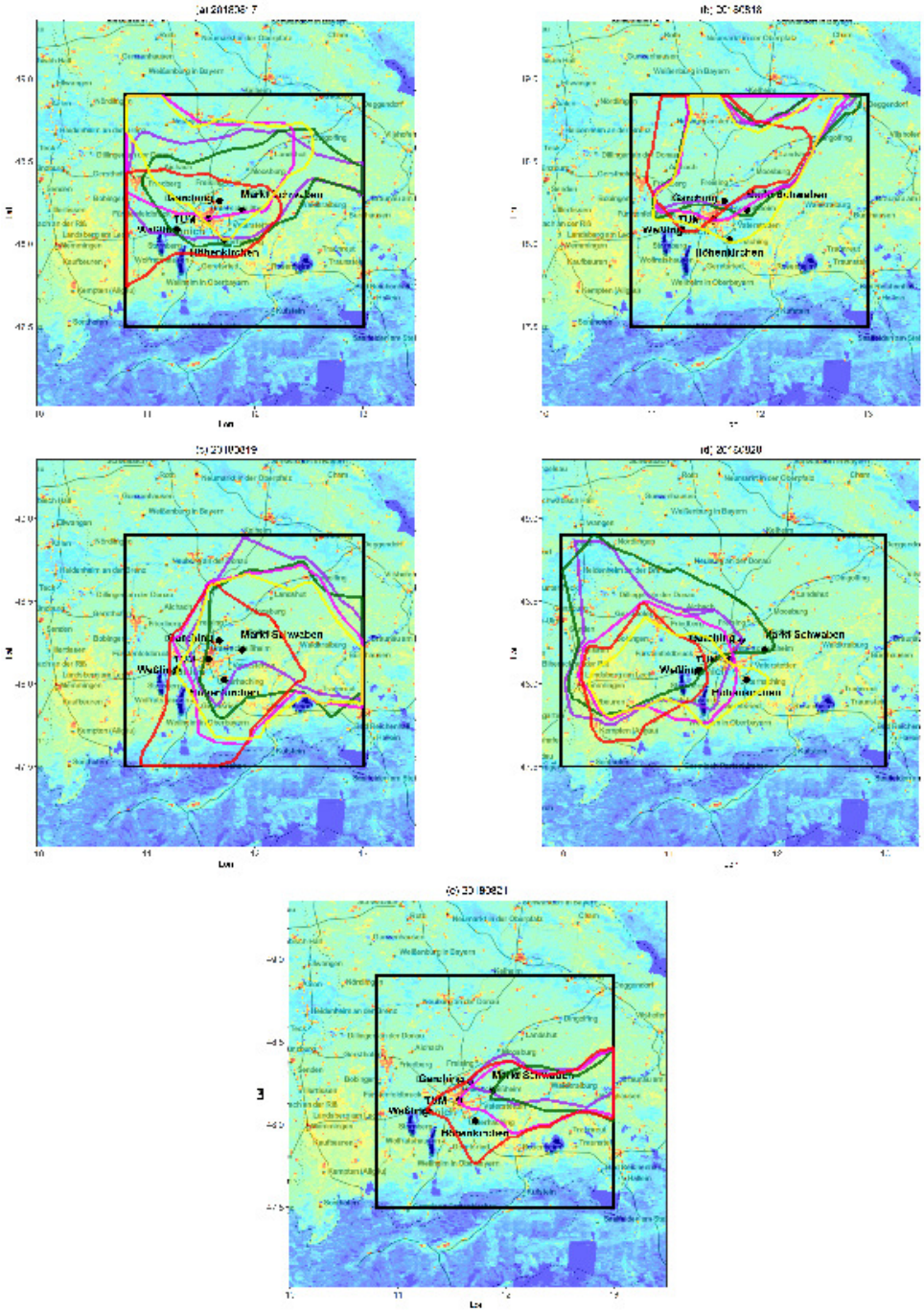


Figure A.2.: Footprint contours from 17 to 21 August at up- and downwind sites with different colors: red for Weßling (West), green for Markt Schwaben (East), purple for Garching (North) and pink for TUM (Center), yellow for Höhenkirchen (South).

List of Symbols

Chemical Symbols

Ar Argon

CFCs Chlorofluorocarbons

CH_4 Methane

CO_2 Carbon Dioxide

H_2O Water (Water Vapor)

HFCs Hydro-fluorocarbons

N_2O Nitrous Oxide

N_2 Dinitrogen

Ne Neon

O_2 Oxygen

O_3 Ozone

Other symbols

S A state variable

V Constant volume

Physics constants

XCH_4 The DMFs of CH_4

XCO_2 The DMFs of CO_2

ΔP_l The pressure difference between the top and the bottom of the l^{th} model layer

$\Delta G^{s}_{sla,l}$ The smoothed concentration for a target gas G at a specific location and the l^{th} layer of the slant column from the site to the Sun

ΔS_{t_n} The change of the atmospheric stage at a certain time

Δt Time step

η Vertical levels in WRF

$\mathrm{AK}_{G,l}$ The fitted AK of the gas G at the l^{th} layer and time t

$\mathrm{L}_{\mathrm{unsta}}$ The total number of the unstaggered levels

w_l The weight of the l^{th} vertical layer

XG The simulated column-concentration

dP The pressure difference between the top and the bottom of the air parcel

F Force from a group of dynamical and chemical atmospheric processes in WRF

F_{bot} An upward force from the air pressure below the air parcel

F_{top} A downward force from the air pressure above the air parcel

G A target Gas

g The acceleration of gravity

G_l The simulated mole fraction at the location (x, y) and time t in the l^{th} unstaggered level of WRF

$G_{\mathrm{pri},l}$ The mixing ratio of the a-priori profile at the l^{th} layer

$G_{\mathrm{sla},l}$ The modelled concentrations for trace gas G at the l^{th} layer following the slant column along the line of the Sun

G_{air} The weight of the air parcel

H_{gph} The geopotential height

l The height of the air parcel

P Hydro-static pressure

P_b The base state pressure

P_p The perturbation pressure

P_{bot} The air pressure at the bottom of the air parcel

P_{sfc} The surface pressure of the model domain

P_{top}	The air pressure at the top of the air parcel
P_{tp}	The pressure at the top of the model domain, i.e. 50 hPa
PH	The perturbation geopotential
PHB	The base state geopotential
rho	The density of the air parcel
S_t	A specific atmospheric state in WRF
T_{tot}	Temperature
$theta$	The equivalent potential temperature
u	The component of the horizontal wind towards east
v	The component of the horizontal wind towards north
w	The component of the vertical wind
wd	Wind Direction
ws	Wind Speed

List of Abbreviations

AK	Averaging Kernel
CAMS	Copernicus Atmosphere Monitoring Service
CAPSS	Clean Air Policy Support System
CDS	Climate Data Store from ECMWF
CFD	Computational Fluid Dynamics
CGLS-LC100	Dynamic Land Cover map of the Copernicus Global Land Service at a resolution of 100 m
CHE	CO_2 Human Emissions Projects
COCCON	Collaborative Carbon Column Observing Network
DCM	Differential Column Method
DMF	Dry-air Mole Fraction
DSS	Data Storage System
DWD	Deutscher Wetterdienst
ECWMF	European Centre for Medium-Range Weather Forecasts
EDGAR	Emission Database for Global Atmospheric Research
EEA	European Environment Agency
EMEP	European Monitoring and Evaluation Programme
EPA	United States Environmental Protection Agency
ERA5	ECMWF reanalysis V5
EU	European Union
EVI	Enhanced Vegetation Index
FTIR	Fourier transform infrared spectroscopy
FTS	Fourier Transform Spectrometer
GADM	Database of Global Administrative Areas
GCM	Global Circulation Model
GDAS	Global Data Assimilation System
GFS	Global Forecast System
GHG	Greenhouse Gases
GNFR	Gridding Nomenclature For Reporting

GPP	Gross Primary Production
GUI	Graphical User Interface
GWP	Global Warming Potential
HPC	High Performance Computer
HRRR	High-Resolution Rapid Refresh
IAGOS	In-service Aircraft for a Global Observing System
ICOS	Integrated Carbon Observation System
IER	Institute of Energy Economics and Rational Energy Use
IFS	Integrated Forecasting System
IGBP	International Geosphere-Biosphere Programme global land cover classification
IGRA	Integrated Global Radiosonde Archive
JRC	Joint Research Centre
LCC	Lamber Conformal Conic projection
LCZ	Local Climate Zones
LEXIS	Large-scale EXecution for Industry
Society project	
LMU	Ludwig Maximilian University of Munich
LRZ	Leibniz Supercomputing Centre
LSWI	Land Surface Water Index
LUE	Light-Use-Efficiency
MACC	Monitoring Atmospheric Composition and Climate
MAE	Mean Absolute Error
MB	Mean Bias
MEIC	Multi-resolution Emission Inventory for China
MICS	Model Inter-Comparison Study
MIX	Mosaic Asian Anthropogenic Emission Inventory
MODIS	Moderate Resolution Imaging Spectroradiometer
MPI	Max Planck Institute
MUCCnet	Munich Urban Carbon Column network
NAM	North American Mesoscale Forecast System
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NEE	Net Ecosystem Exchange
NNRP	National Nuclear Regulatory Portal
NOAA	National Oceanic and Atmospheric Administration
PBL	Planetary Boundary Layer

RCM	Regional Circulation Model
REAS	Regional Emission inventory in Asia
RES	Ecosystem Respiration
RMSE	Root Mean Square Error
SAGA GIS	System for Automated Geoscientific Analyses
SIF	Solar-Induced Fluorescence
STILT	Stochastic Time-Inverted Lagrangian Transport model
SYNMAP	Synergetic Land Cover Product
SZA	Solar Zenith Angle
T2	Temperature at a height of two meters above the ground
TA	Training Area
TCCON	Total Carbon Column Observing Network
TNO	Netherlands Organisation for applied scientific research
TNO_GHGco	TNO GHG and co-emitted species emission database
TUM	Technology University of Munich
UCAR	University Corporation for Atmospheric Research
VPRM	Vegetation Photosynthesis and Respiration Model
W2W	WUDAPT to WRF
WACCM	Whole Atmosphere Community Climate Model
WD	Wind Direction
WPS	WRF prepossessing
WRF	Weather Research and Forecasting
WRF-ARW	Advanced Research WRF
WRF-NMM	Non-hydrostatic Meso-scale WRF
WRFDA	WRF Data Assimilation
WS	Wind Speed
WUDAPT	World Urban Database and Access Portal Tools project

Bibliography

[1] V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, *et al.*, "IPCC, 2018: Summary for policymakers. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty", *World Meteorological Organization, Geneva, Switzerland*, vol. 10, 2018.

[2] M. Allen, M. Babiker, Y. Chen, H. de Coninck, S. Connors, R. van Diemen, O. Dube, *et al.*, "Summary for policymakers Global Warming of 1.5°C: an IPCC Special Report on the Impacts of Global Warming of 1.5°C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emissions Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change", *World Meteorological Organization*, pp. 1–24, 2018.

[3] V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, *et al.*, "Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change", *Cambridge University Press*, pp. 1–24, 2021.

[4] International Energy Agency, "Cities, Towns & Renewable Energy", Tech. Rep., 2009.

[5] C. Kennedy, J. Steinberger, B. Gasson, Y. Hansen, T. Hillman, M. Havranek, D. Pataki, *et al.*, "Greenhouse gas emissions from global cities", vol. 43, no. 19, pp. 7297–7302, 2009. DOI: 10.1021/es900213p.

[6] L. Parshall, K. Gurney, S. Hammer, D. Mendoza, Y. Zhou, and S. Geethakumar, "Modeling energy consumption and CO_2 emissions at the urban scale: Methodological challenges and insights from the United States",

Energy Policy, vol. 38, no. 9, pp. 4765–4782, 2010. DOI: 10.1016/j.enpol.2009.07.006.

[7] International Panel on Climate Change, *IPCC fifth assessment report: Climate change 2014 Synthesis Report. Summary for Policymakers*, 2014.

[8] United Nations Department of Economic and Social Affairs, *World Urbanization Prospects: The 2014 Revision*, 2014.

[9] United Nations Department of Economic and Social Affairs, *World Urbanization Prospects: The 2018 Revision*, 2018.

[10] L. Wu, G. Broquet, P. Ciais, V. Bellassen, F. Vogel, F. Chevallier, I. Xueref-Remy, *et al.*, "What would dense atmospheric observation networks bring to the quantification of city CO_2 emissions?", *Atmospheric Chemistry and Physics*, vol. 16, no. 12, pp. 7743–7771, 2016. DOI: 10.5194/acp-16-7743-2016.

[11] M. Fragkias, J. Lobo, D. Strumsky, and K. Seto, "Does size matter? Scaling of CO_2 emissions and US urban areas", *PLOS One*, vol. 8, no. 6, e64727, 2013. DOI: 10.1371/journal.pone.0064727.

[12] F. Hopkins, J. Ehleringer, S. Bush, R. Duren, C. Miller, C. Lai, Y. Hsu, *et al.*, "Mitigation of methane emissions in cities: How new measurements and partnerships can contribute to emissions reduction strategies", *Earth's Future*, vol. 4, no. 9, pp. 408–425, 2016. DOI: 10.1002/2016EF000381.

[13] P. Marcotullio, A. Sarzynski, J. Albrecht, N. Schulz, and J. Garcia, "The geography of global urban greenhouse gas emissions: An exploratory analysis", *Climatic Change*, vol. 121, no. 4, pp. 621–634, 2013. DOI: 10.1007/s10584-013-0977-z.

[14] EU Commission, "A Clean Planet for all-A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy", Technical report, European Commission, Brussels. Federal Ministry of Interior Building and Community (IBC)-Sustainable Building Assessment System (BNB), Tech. Rep., 2018.

[15] P. DeCola, O. Tarasova, D. Brunner, S. Maksyutov, A. Manning, F. Vogel, K. Gurney, *et al.*, "An Integrated Global Greenhouse Gas Information System (IG3IS) Science Implementation Plan Approved by EC-70", *World Meteorological Organization (WMO)*, no. No. 1218, 2018.

[16] T. Lauvaux, N. Miles, A. Deng, S. Richardson, M. Cambaliza, K. Davis, B. Gaudet, *et al.*, "High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX)", *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 10, pp. 5213–5236, 2016. DOI: 10.1002/2015JD024473.

[17] J. Staufer, G. Broquet, F. Bréon, V. Puygrenier, F. Chevallier, I. Xueref-Rémy, E. Dieudonné, *et al.*, "The first 1-year-long estimate of the Paris region fossil fuel CO_2 emissions based on atmospheric inversion", *Atmospheric Chemistry and Physics*, vol. 16, no. 22, pp. 14 703–14 726, 2016. DOI: 10.5194/acp-16-14703-2016.

[18] D. Pillai, C. Gerbig, R. Ahmadov, C. Rödenbeck, R. Kretschmer, T. Koch, R. Thompson, *et al.*, "High-resolution simulations of atmospheric CO_2 over complex terrain–representing the Ochsenkopf mountain tall tower", *Atmospheric Chemistry and Physics*, vol. 11, no. 15, pp. 7445–7464, 2011. DOI: 10.5194/acp-11-7445-2011.

[19] D. Caulton, P. Shepson, R. Santoro, J. Sparks, R. Howarth, A. Ingraffea, M. Cambaliza, *et al.*, "Toward a better understanding and quantification of methane emissions from shale gas development", *Proceedings of the National Academy of Sciences*, p. 201 316 546, 2014. DOI: 10.1073/pnas.1316546111.

[20] X. Newman S.and Xu, K. Gurney, Y. Hsu, K. Li, X. Jiang, R. Keeling, S. Feng, *et al.*, "Toward consistency between trends in bottom-up CO_2 emissions and top-down atmospheric measurements in the Los Angeles megacity", *Atmospheric Chemistry and Physics*, vol. 16, no. 6, pp. 3843–3863, 2016. DOI: 10.5194/acp-16-3843-2016.

[21] H. Wang, L. Fu, X. Lin, Y. Zhou, and J. Chen, "A bottom-up methodology to estimate vehicle emissions for the Beijing urban area", *Science of the total environment*, vol. 407, no. 6, pp. 1947–1953, 2009. DOI: 10.1016/j.scitotenv.2008.11.008.

[22] S. Montzka, E. Dlugokencky, and J. Butler, "Non-CO_2 greenhouse gases and climate change", *Nature*, vol. 476, pp. 43–50, 2011. DOI: 10.1038/nature10322.

[23] P. Bergamaschi, M. Corazza, U. Karstens, M. Athanassiadou, R. Thompson, I. Pison, A. Manning, *et al.*, "Top-down estimates of European CH_4 and N_2O emissions based on four different inverse models", *At-*

mospheric Chemistry and Physics, vol. 15, no. 2, pp. 715–736, 2015. DOI: 10.5194/acp-15-715-2015.

[24] T. Klausner, M. Mertens, H. Huntrieser, M. Galkowski, G. Kuhlmann, R. Baumann, A. Fiehn, *et al.*, "Urban greenhouse gas emissions from the Berlin area: A case study using airborne CO_2 and CH_4 in situ observations in summer 2018", *Elementa: Science of the Anthropocene*, vol. 8, 2020. DOI: 10.1525/elementa.411.

[25] R. Andres, T. Boden, F. Bréon, P. Ciais, S. Davis, D. Erickson, J. Gregg, *et al.*, "A synthesis of carbon dioxide emissions from fossil-fuel combustion", *Biogeosciences*, vol. 9, no. 5, pp. 1845–1871, 2012. DOI: 10.5194/bg-9-1845-2012.

[26] D. Wunch, G. Toon, J. Blavier, R. Washenfelder, J. Notholt, B. Connor, D. Griffith, *et al.*, "The total carbon column observing network", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1943, pp. 2087–2112, 2011. DOI: 10.1098/rsta.2010.0240.

[27] A. Shekhar, J. Chen, J. Paetzold, F. Dietrich, X. Zhao, S. Bhattacharjee, V. Ruisinger, *et al.*, "Anthropogenic CO_2 emissions assessment of Nile Delta using XCO_2 and SIF data from OCO-2 satellite", *Environmental Research Letters*, vol. 15, no. 9, p. 095 010, 2020. DOI: 10.1088/1748-9326/ab9cfe.

[28] C. Cressot, F. Chevallier, P. Bousquet, C. Crevoisier, E. Dlugokencky, A. Fortems-Cheiney, C. Frankenberg, *et al.*, "On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements", *Atmospheric Chemistry and Physics*, vol. 14, no. 2, pp. 577–592, 2014. DOI: 10.5194/acp-14-577-2014.

[29] A. Heimburger, R. Harvey, P. Shepson, B. Stirm, C. Gore, J. Turnbull, M. Cambaliza, *et al.*, "Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging", *Elementa: Science of the Anthropocene*, vol. 5, 2017. DOI: 10.1525/elementa.134.

[30] T. Jones, J. Franklin, J. Chen, F. Dietrich, K. D. Hajny, J. C. Paetzold, A. Wenzel, *et al.*, "Assessing urban methane emissions using column-observing portable fourier transform infrared (ftir) spectrometers and a novel bayesian inversion framework", *Atmospheric Chemistry and Physics*,

vol. 21, no. 17, pp. 13 131–13 147, 2021. DOI: 10.5194/acp-21-13131-2021.

[31] A. Turner, A. Shusterman, B. McDonald, V. Teige, R. Harley, and R. Cohen, "Network design for quantifying urban CO_2 emissions: assessing trade-offs between precision and network density", *Atmospheric Chemistry and Physics*, vol. 16, no. 21, pp. 13 465–13 475, 2016. DOI: 10.5194/acp-16-13465-2016.

[32] M. Sargent, Y. Barrera, T. Nehrkorn, L. Hutyra, C. Gately, T. Jones, K. McKain, *et al.*, "Anthropogenic and biogenic CO_2 fluxes in the Boston urban region", *Proceedings of the National Academy of Sciences*, vol. 115, no. 29, pp. 7491–7496, 2018. DOI: 10.1073/pnas.1803715115.

[33] K. McKain, S. Wofsy, T. Nehrkorn, J. Eluszkiewicz, J. Ehleringer, and B. Stephens, "Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region", *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8423–8428, 2012. DOI: 10.1073/pnas.1116645109.

[34] S. Olsen and J. Randerson, "Differences between surface and column atmospheric CO_2 and implications for carbon cycle research", *Journal of Geophysical Research: Atmospheres*, vol. 109, no. D2, 2004. DOI: 10.1029/2003JD003968.

[35] R. Kivi and P. Heikkinen, "Fourier transform spectrometer measurements of column CO_2 at Sodankylä, Finland", *Geoscientific Instrumentation, Methods and Data Systems*, vol. 5, no. 2, pp. 271–279, 2016. DOI: 10.5194/gi-5-271-2016.

[36] C. Gerbig, S. Körner, and J. Lin, "Vertical mixing in atmospheric tracer transport models: error characterization and propagation", *Atmospheric Chemistry and Physics*, vol. 8, no. 5, pp. 591–602, 2008. DOI: 10.5194/acp-8-591-2008.

[37] H. Ohyama, I. Morino, T. Nagahama, T. Machida, H. Suto, H. Oguma, Y. Sawa, *et al.*, "Column-averaged volume mixing ratio of CO_2 measured with ground-based Fourier transform spectrometer at Tsukuba", *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D18, 2009. DOI: 10.1029/2008JD011465.

[38] A. Ostler, R. Sussmann, P. Patra, S. Houweling, B. De, G. Stiller, F. Haenel, *et al.*, "Evaluation of column-averaged methane in models and TCCON with a focus on the stratosphere", *Atmospheric Measurement Techniques*, vol. 9, pp. 4843–4859, 2016. DOI: 10.5194/amt-9-4843-2016.

[39] J. Chen, C. Viatte, J. Hedelius, T. Jones, J. Franklin, H. Parker, E. Gottlieb, *et al.*, "Differential column measurements using compact solar-tracking spectrometers", *Atmospheric Chemistry and Physics*, vol. 16, no. 13, pp. 8479–8498, 2016. DOI: 10.5194/acp-16-8479-2016.

[40] J. Chen, J. Samra, E. Gottlieb, J. Budney, S. Wofsy, K. McKain, F. Hase, *et al.*, "Novel instrumentation for column measurements and regional eulerian modeling for network design in Boston", in *AGU Fall Meeting Abstracts*, 2013.

[41] J. Franklin, T. Jones, J. Chen, H. Parker, J. Hedelius, P. Wennberg, M. Dubey, *et al.*, "A three-dimensional observation network for determining urban emissions of CO_2 and CH_4", 2017 North American Carbon Program, North Bethesda, MD, USA, 2017.

[42] F. Klappenbach, J. Chen, A. Wenzel, A. Forstmaier, F. Dietrich, X. Zhao, T. Jones, *et al.*, *Methane emission estimate using ground based remote sensing in complex terrain*, ser. EGU General Assembly Conference Abstracts. 2021, EGU21–15 406. DOI: 10.5194/egusphere-egu21-15406.

[43] F. Hase, M. Frey, T. Blumenstock, J. Groß, M. Kiel, R. Kohlhepp, G. Mengistu Tsidu, *et al.*, "Application of portable FTIR spectrometers for detecting greenhouse gas emissions of the major city Berlin", *Atmospheric Measurement Techniques*, vol. 8, no. 7, pp. 3059–3068, 2015. DOI: 10.5194/amt-8-3059-2015.

[44] J. Chen, F. Dietrich, J. Franklin, T. Jones, B. André, A. Luther, R. Kleinschek, *et al.*, "Mesoscale column network for assessing GHG and NOx emissions in Munich", vol. 20, EGU General Assembly Conference Abstracts, 2018. DOI: 10.5194/EGU2018-10192-2.

[45] A. Forstmaier, J. Chen, F. Dietrich, J. Bettinelli, H. Maazallahi, C. Schneider, D. Winkler, *et al.*, "Quantification of methane emissions in hamburg using a network of ftir spectrometers and an inverse modeling approach [in review]", *Atmospheric Chemistry and Physics*, pp. 1–33, 2022. DOI: 10.5194/acp-2022-710.

[46] F. Toja-Silva, J. Chen, S. Hachinger, and F. Hase, "CFD simulation of CO_2 dispersion from urban thermal power plant: Analysis of turbulent Schmidt number and comparison with Gaussian plume model and measurements", *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 169, pp. 177–193, 2017. DOI: 10.1016/j.jweia.2017.07.015.

[47] C. Viatte, T. Lauvaux, J. Hedelius, H. Parker, J. Chen, T. Jones, J. Franklin, *et al.*, "Methane emissions from dairies in the Los Angeles Basin", *Atmospheric Chemistry and Physics*, vol. 17, no. 12, pp. 7509–7528, 2017. DOI: 10.5194/acp-17-7509-2017.

[48] F. Vogel, M. Frey, J. Staufer, F. Hase, G. Broquet, I. Xueref-Remy, F. Chevallier, *et al.*, "XCO_2 in an emission hot-spot region: the COCCON Paris campaign 2015", *Atmospheric Chemistry and Physics*, vol. 19, no. 5, pp. 3271–3285, 2019. DOI: 10.5194/acp-2018-595.

[49] L. Heinle and J. Chen, "Automated enclosure and protection system for compact solar-tracking spectrometers", *Atmospheric Measurement Techniques*, vol. 11, no. 4, pp. 2173–2185, 2018. DOI: 10.5194/amt-11-2173-2018.

[50] F. Dietrich, B. Chen J.and Voggenreiter, P. Aigner, N. Nachtigall, and B. Reger, "MUCCnet: Munich urban carbon column network", *Atmospheric Measurement Techniques*, vol. 14, no. 2, pp. 1111–1126, 2021. DOI: 10.5194/amt-14-1111-2021.

[51] X. Zhao, J. Marshall, S. Hachinger, C. Gerbig, M. Frey, F. Hase, and J. Chen, "Analysis of total column CO_2 and CH_4 measurements in Berlin with WRF-GHG", *Atmospheric Chemistry and Physics*, vol. 19, no. 17, pp. 11 279–11 302, 2019. DOI: 10.5194/acp-19-11279-2019.

[52] X. Zhao, J. Chen, J. Marshall, M. Galkowski, C. Gerbig, S. Hachinger, J. Gensheimer, *et al.*, *Quantifying the impact of urban greenhouse gas emissions for Munich during the COVID-19 pandemic using WRF V3.9.1.1, series = EGU General Assembly*. 2021. DOI: 10.5194/egusphere-egu21-13431.

[53] X. Zhao, J. Chen, J. Marshall, G. Christoph, M. Galkowski, H. Stephan, F. Dietrich, *et al.*, "Understanding greenhouse gas (GHG) column concentrations in Munich using WRF [in review]", *Atmospheric Chemistry and Physics*, 2022.

[54] A. Baede, E. Ahlonsou, Y. Ding, and D. Schimel, "The Climate System: an Overview", Climate change 2001: the scientific basis, Tech. Rep., 2001, pp. 38–47.

[55] T. Skytt, S. N., and B. Jonsson, "Global warming potential and absolute global temperature change potential from carbon dioxide and methane fluxes as indicators of regional sustainability–A case study of Jämtland, Sweden", *Ecological Indicators*, vol. 110, p. 105 831, 2020. DOI: 10.1016/j.ecolind.2019.105831.

[56] M. Kawamiya, T. Hajima, K. Tachiiri, S. Watanabe, and T. Yokohata, "Two decades of Earth system modeling with an emphasis on Model for Interdisciplinary Research on Climate (MIROC)", *Progress in Earth and Planetary Science*, vol. 7, no. 1, pp. 1–13, 2020. DOI: 10.1186/s40645-020-00369-5.

[57] M. Jacobson, "Introduction", in *Fundamentals of Atmospheric Modeling*, 2nd ed. Cambridge University Press, 2005, pp. 1–11. DOI: 10.1017/CBO9781139165389.002.

[58] Intergovernmental Panel On Climate Change, "Climate change 2007: the physical science basis", *Agenda*, vol. 6, no. 07, p. 333, 2007.

[59] K. Trenberth and D. Stepaniak, "The flow of energy through the Earth's climate system", *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 130, no. 603, pp. 2677–2701, 2004. DOI: 10.1256/qj.04.83.

[60] H. Goosse, P. Barriat, M. Loutre, and V. Zunz, *Introduction to climate dynamics and climate modeling*. Centre de recherche sur la Terre et le climat Georges Lemaitre-UCLouvain, 2010.

[61] J. Lin, D. Brunner, C. Gerbig, A. Stohl, A. Luhar, and P. Webley, *Lagrangian modeling of the atmosphere*. John Wiley & Sons, 2013.

[62] M. Dhadly and M. Conde, "Trajectories of thermospheric air parcels flowing over alaska, reconstructed from ground-based wind measurements", *Journal of Geophysical Research: Space Physics*, vol. 122, no. 6, pp. 6635–6651, 2017. DOI: 10.1002/2017JA024095.

[63] G. P. Brasseur and D. J. Jacob, *Modeling of atmospheric chemistry*. Cambridge University Press, 2017.

[64] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, and J. Powers, "A description of the advanced research WRF Version 2", National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology, Tech. Rep., 2005.

[65] J. Kukkonen, T. Olsson, D. M. Schultz, A. Baklanov, T. Klein, A. Miranda, A. Monteiro, *et al.*, "A review of operational, regional-scale, chemical weather forecasting models in europe", *Atmospheric Chemistry and Physics*, vol. 12, no. 1, pp. 1–87, 2012. DOI: 10.5194/acp-12-1-2012.

[66] J. R. Holton, *An introduction to dynamic meteorology (Fourth Edition)*. Elsevier Academic Press, 2004, vol. 88.

[67] N. Wedi and S. Maladel, "Non-hydrostatic modelling at ECMWF", ECMWF Newsletter No. 125, Tech. Rep., 2010, pp. 17–21.

[68] M. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets", *Remote sensing of Environment*, vol. 114, no. 1, pp. 168–182, 2010. DOI: 10.1016/j.rse.2009.08.016.

[69] O. Brousse, A. Martilli, M. Foley, G. Mills, and B. Bechtel, "WUDAPT, an efficient land use producing data tool for mesoscale models? Integration of urban LCZ in WRF over Madrid", *Urban Climate*, vol. 17, pp. 116–134, 2016. DOI: 10.1016/j.uclim.2016.04.001.

[70] J. Ching, G. Mills, B. Bechtel, L. See, J. Feddema, X. Wang, C. Ren, *et al.*, "WUDAPT: An urban weather, climate, and environmental modeling infrastructure for the anthropocene", *Bulletin of the American Meteorological Society*, vol. 99, no. 9, pp. 1907–1924, 2018. DOI: 10.1175/BAMS-D-16-0236.1.

[71] M. Mughal, "Modelling the Urban Heat Island in Singapore-state of the art WRF model technical details", *Technical Report Cooling Singapore*, vol. 500, 2020. DOI: 10.3929/ethz-b-000412358.

[72] M. Demuzere, B. Bechtel, A. Middel, and G. Mills, "Mapping Europe into local climate zones", *PloS one*, vol. 14, no. 4, e0214474, 2019. DOI: 10.1371/journal.pone.0214474.

[73] M. Demuzere, B. Bechtel, A. Middel, and G. Mills, "European LCZ map", 2021. DOI: 10.6084/m9.figshare.13322450.v2.

[74] M. Demuzere, J. Kittner, and B. Bechtel, "LCZ Generator: a web application to create Local Climate Zone maps", *Frontiers in Environmental Science*, 2021. DOI: 10.3389/fenvs.2021.637455.

[75] L. See, C. Perger, M. Duerauer, S. Fritz, B. Bechtel, J. Ching, P. Alexander, *et al.*, "Developing a community-based worldwide urban morphology and materials database (WUDAPT) using remote sensing and crowd sourcing for improved urban climate modelling", in *2015 Joint Urban Remote Sensing Event (JURSE)*, IEEE, 2015, pp. 1–4. DOI: 10.1109/JURSE.2015.7120501.

[76] O. Conrad, B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, *et al.*, "System for automated geoscientific analyses (SAGA) v.2.1.4", *Geoscientific Model Development*, vol. 8, no. 7, pp. 1991–2007, 2015. DOI: 10.5194/gmd-8-1991-2015.

[77] M. Demuzere, S. Hankey, G. Mills, W. Zhang, T. Lu, and B. Bechtel, "Combining expert and crowd-sourced training data to map urban form and functions for the continental US", *Scientific data*, vol. 7, no. 1, pp. 1–13, 2020. DOI: 10.1038/s41597-020-00605-z.

[78] H. Hersbach, B. Bill, B. Paul, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, *et al.*, "The ERA5 global reanalysis", *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020. DOI: 10.1002/qj.3803.

[79] J. Bieser, A. Aulinger, V. Matthias, M. Quante, and H. Van Der Gon, "Vertical emission profiles for Europe based on plume rise calculations", *Environmental Pollution*, vol. 159, no. 10, pp. 2935–2946, 2011. DOI: 10.1016/j.envpol.2011.04.030.

[80] M. Muntean, G. Janssens-Maenhout, S. Song, N. Selin, J. Olivier, D. Guizzardi, R. Maas, *et al.*, "Trend analysis from 1970 to 2008 and model evaluation of EDGAR V.4 global gridded anthropogenic mercury emissions", *Science of the Total Environment*, vol. 494, pp. 337–350, 2014. DOI: 10.1016/j.scitotenv.2014.06.014.

[81] G. Janssens-Maenhout, M. Crippa, D. Guizzardi, F. Dentener, M. Muntean, G. Pouliot, T. Keating, *et al.*, "HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution", *Atmospheric Chemistry and Physics*, vol. 15, no. 19, pp. 11 411–11 432, 2015. DOI: 10.5194/acp-15-11411-2015.

[82] E. Saikawa, H. Kim, M. Zhong, A. Avramov, Y. Zhao, G. Janssens-Maenhout, J. Kurokawa, *et al.*, "Comparison of emissions inventories of anthropogenic air pollutants and greenhouse gases in China", *Atmospheric Chemistry and Physics*, vol. 17, no. 10, pp. 6393–6421, 2017. DOI: 10.5194/acp-17-6393-2017.

[83] C. Holmes, "Methane feedback on atmospheric chemistry: Methods, models, and mechanisms", *Journal of Advances in Modeling Earth Systems*, vol. 10, no. 4, pp. 1087–1099, 2018. DOI: 10.1002/2017MS001196.

[84] I. Super, S. Dellaert, A. Visschedijk, and H. Denier van der Gon, "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*, vol. 20, no. 3, pp. 1795–1816, 2020. DOI: 10.5194/acp-20-1795-2020.

[85] H. van der Gon, J. Kuenen, E. Boleti, M. Muntean, M. G., J. Marshall, and J. Haussaire, "Emissions and natural fluxes Dataset", Tech. Rep., 2019.

[86] X. Zhao, J. Marshall, S. Hachinger, C. Gerbig, M. Frey, F. Hase, and J. Chen, "Analysis of total column CO_2 and CH_4 measurements in Berlin with WRF-GHG", *Atmospheric Chemistry and Physics*, vol. 19, no. 17, pp. 11 279–11 302, 2019. DOI: 10.5194/acp-19-11279-2019.

[87] D. Brunner, G. Kuhlmann, J. Marshall, V. Clément, O. Fuhrer, G. Broquet, A. Löscher, *et al.*, "Accounting for the vertical distribution of emissions in atmospheric CO_2 simulations", *Atmospheric Chemistry and Physics*, vol. 19, no. 7, pp. 4541–4559, 2019. DOI: 10.5194/acp-19-4541-2019.

[88] S. Rémy, Z. Kipling, J. Flemming, O. Boucher, P. Nabat, M. Michou, A. Bozzo, *et al.*, "Description and evaluation of the troposphere aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1)", *Geoscientific Model Development*, vol. 12, no. 11, pp. 4627–4659, 2019. DOI: 10.5194/gmd-12-4627-2019.

[89] P. Browne, P. de Rosnay, H. Zuo, A. Bennett, and A. Dawson, "Weakly coupled ocean–atmosphere data assimilation in the ECMWF NWP system", *Remote Sensing*, vol. 11, no. 3, p. 234, 2019. DOI: 10.3390/rs11030234.

[90] P. Mahadevan, S. Wofsy, D. Matross, X. Xiao, A. Dunn, J. Lin, C. Gerbig, *et al.*, "A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM)", *Global Biogeochemical Cycles*, vol. 22, no. 2, 2008. DOI: 10.1029/2006GB002735.

[91] D. Wu, J. Lin, H. Duarte, V. Yadav, N. Parazoo, T. Oda, and E. Kort, "A model for urban biogenic CO_2 fluxes: Solar-Induced Fluorescence for Modeling Urban biogenic Fluxes (SMUrF v1)", *Geoscientific Model Development*, vol. 14, no. 6, pp. 3633–3661, 2021. DOI: 10.5194/gmd-14-3633-2021.

[92] J. Gensheimer, A. J. Turner, P. Köhler, C. Frankenberg, and J. Chen, "A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet)", *Biogeosciences*, vol. 19, no. 6, pp. 1777–1793, 2022. DOI: 10.5194/bg-19-1777-2022.

[93] A. Shekhar, J. Chen, S. Bhattacharjee, A. Buras, A. O. Castro, C. S. Zang, and A. Rammig, "Capturing the impact of the 2018 European drought and heat across different vegetation types using OCO-2 solar-induced fluorescence", *Remote Sensing*, vol. 12, no. 19, p. 3249, 2020. DOI: 10.3390/rs12193249.

[94] X. Li, J. Xiao, and B. He, "Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests", *Remote Sensing of Environment*, vol. 204, pp. 659–671, 2018. DOI: 10.1016/j.rse.2017.09.034.

[95] S. Du, L. Liu, X. Liu, X. Zhang, X. Zhang, Y. Bi, and L. Zhang, "Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite", *Science Bulletin*, vol. 63, no. 22, pp. 1502–1512, 2018. DOI: 10.1016/j.scib.2018.10.003.

[96] P. Köehler, C. Frankenberg, T. S. Magney, L. Guanter, J. Joiner, and J. Landgraf, "Global retrievals of solar induced chlorophyll fluorescence with TROPOMI: first results and inter-sensor comparison to OCO-2", *Geophysical Research Letters*, vol. 45, no. 19, pp. 10 456–10 463, 2018. DOI: 10.1029/2018GL079031.

[97] J. Chen, C. Gerbig, J. Marshall, and K. Totsche, "Short-term forecasting of regional biospheric CO_2 fluxes in Europe using a light-use-efficiency model (VPRM, MPI-BGC version 1.2)", *Geoscientific Model De-*

velopment, vol. 13, no. 9, pp. 4091–4106, 2020. DOI: 10.5194/gmd-13-4091-2020.

[98] V. Beck, T. Koch, R. Kretschmer, J. Marshall, R. Ahmadov, C. Gerbig, D. Pillai, *et al.* "The WRF Greenhouse Gas Model (WRF-GHG)". (2012), [Online]. Available: https://www.bgc-jena.mpg.de/bgc-systems/pmwiki2/uploads/Download/Wrf-ghg/WRF-GHG_Techn_Report.pdf.

[99] E. Vermote, "MOD09A1 MODIS Surface Reflectance 8-Day L3 Global 500m SIN Grid V006", *https://modis.gsfc.nasa.gov/data/dataprod/mod09.php*, 2015. DOI: 10.5067/MODIS/MOD09A1.006.

[100] M. Jung, K. Henkel, M. Herold, and G. Churkina, "Exploiting synergies of global land cover products for carbon cycle modeling", *Remote Sensing of Environment*, vol. 101, no. 4, pp. 534–553, 2006. DOI: 10.1016/j.rse.2006.01.020.

[101] X. Hu, S. Gourdji, K. Davis, Q. Wang, Y. Zhang, M. Xue, S. Feng, *et al.*, "Implementation of improved parameterization of terrestrial flux in WRF-VPRM improves the simulation of nighttime CO_2 peaks and a daytime CO_2 band ahead of a cold front", *Journal of Geophysical Research: Atmospheres*, vol. 126, no. 10, 2021. DOI: 10.1029/2020JD034362.

[102] S. Gourdji, A. Karion, I. Lopez-Coto, S. Ghosh, K. Mueller, Y. Zhou, C. Williams, *et al.*, "A modified Vegetation Photosynthesis and Respiration Model (VPRM) for the eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models", *Journal of Geophysical Research: Biogeosciences*, 2021. DOI: 10.1029/2021JG006290.

[103] F. Chen, H. Kusaka, R. Bornstein, J. Ching, C. Grimmond, S. Grossman-Clarke, T. Loridan, *et al.*, "The integrated WRF-urban modelling system: development, evaluation, and applications to urban environmental problems", *International Journal of Climatology*, vol. 31, no. 2, pp. 273–288, 2011. DOI: 10.1002/joc.2158.

[104] R. Ahmadov, C. Gerbig, R. Kretschmer, S. Koerner, B. Neininger, A. Dolman, and C. Sarrat, "Mesoscale covariance of transport and CO_2 fluxes: Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model", *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D22, 2007. DOI: 10.1029/2007JD008552.

[105] R. Ahmadov, C. Gerbig, R. Kretschmer, S. Körner, C. Rödenbeck, P. Bousquet, and M. Ramonet, "Comparing high resolution WRF-VPRM simulations and two global CO_2 transport models with coastal tower measurements of CO_2", *Biogeosciences*, vol. 6, no. 5, pp. 807–817, 2009. DOI: 10.5194/bg-6-807-2009.

[106] D. Pillai, C. Gerbig, R. Kretschmer, V. Beck, U. Karstens, B. Neininger, and M. Heimann, "Comparing Lagrangian and Eulerian models for CO_2 transport–a step towards Bayesian inverse modeling using WRF/STILT-VPRM", *Atmospheric Chemistry and Physics*, vol. 12, no. 19, pp. 8979–8991, 2012. DOI: 10.5194/acp-12-8979-2012.

[107] R. Kretschmer, C. Gerbig, U. Karstens, and F. Koch, "Error characterization of CO_2 vertical mixing in the atmospheric transport model WRF-VPRM", *Atmospheric Chemistry and Physics*, vol. 12, no. 5, pp. 2441–2458, 2012. DOI: 10.5194/acp-12-2441-2012.

[108] B. Hardiman, J. Wang, L. Hutyra, C. Gately, J. Getson, and M. Friedl, "Accounting for urban biogenic fluxes in regional carbon budgets", *Science of the Total Environment*, vol. 592, pp. 366–372, 2017. DOI: 10.1016/j.scitotenv.2017.03.028.

[109] D. Pillai, M. Buchwitz, C. Gerbig, T. Koch, M. Reuter, H. Bovensmann, J. Marshall, *et al.*, "Tracking city CO_2 emissions from space using a high-resolution inverse modelling approach: a case study for Berlin, Germany", *Atmospheric Chemistry and Physics*, vol. 16, no. 15, pp. 9591–9610, 2016. DOI: 10.5194/acp-16-9591-2016.

[110] F. Dietrich, J. Chen, B. Voggenreiter, P. Aigner, N. Nachtigall, and B. Reger, "MUCCnet: Munich Urban Carbon Column network", *Atmospheric Measurement Techniques*, vol. 14, pp. 1111–1126, 2021. DOI: 10.5194/amt-14-1111-2021.

[111] A. Berkmann, "Applying WRF modelling framework using Docker", Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

[112] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, and J. Powers, "Description of the Advanced Research WRF Version 3", National Center for Atmospheric Research, Tech. Rep., 2008.

[113] J. Kaplan, G. Folberth, and D. Hauglustaine, "Role of methane and biogenic volatile organic compound sources in late glacial and Holocene fluctuations of atmospheric methane concentrations", *Global Biogeochemical Cycles*, vol. 20, no. 2, 2006. DOI: 10.1029/2005GB002590.

[114] V. Beck, C. Gerbig, T. Koch, M. Bela, K. Longo, S. Freitas, J. Kaplan, *et al.*, "WRF-Chem simulations in the Amazon region during wet and dry season transitions: evaluation of methane models and wetland inundation maps", *Atmospheric Chemistry and Physics*, vol. 13, no. 16, pp. 7961–7982, 2013. DOI: 10.5194/acp-13-7961-2013.

[115] V. Beck, T. Koch, R. Kretschmer, R. Marshall J.and Ahmadov, C. Gerbig, D. Pillai, and M. Heimann, "The WRF Greenhouse Gas Model (WRF-GHG). Technical Report No. 25, Max Planck Institute for Biogeochemistry, Jena, Germany.", 2012.

[116] J. Hedelius, C. Viatte, D. Wunch, C. Roehl, G. Toon, J. Chen, T. Jones, *et al.*, "Assessment of errors and biases in retrievals of XCO_2, XCH_4, XCO, and XN_2O from a 0.5 cm–1 resolution solar-viewing spectrometer", *Atmospheric Measurement Techniques*, vol. 9, no. 8, pp. 3527–3546, 2016. DOI: 10.5194/amt-9-3527-2016.

[117] M. Gisi, F. Hase, S. Dohe, T. Blumenstock, A. Simon, and A. Keens, "XCO_2-measurements with a tabletop FTS using solar absorption spectroscopy", *Atmospheric Measurement Techniques*, vol. 5, no. 11, pp. 2969–2980, 2012. DOI: 10.5194/amt-5-2969-2012.

[118] F. Hase, M. Frey, M. Kiel, T. Blumenstock, R. Harig, A. Keens, and J. Orphal, "Addition of a channel for XCO observations to a portable FTIR spectrometer for greenhouse gas measurements", *Atmospheric Measurement Techniques*, vol. 9, no. 5, pp. 2303–2313, 2016. DOI: 10.5194/amt-9-2303-2016.

[119] M. Frey, M. Sha, F. Hase, M. Kiel, T. Blumenstock, R. Harig, G. Surawicz, *et al.*, "Building the COllaborative Carbon Column Observing Network (COCCON): long-term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer", *Atmospheric Measurement Techniques*, vol. 12, no. 3, pp. 1513–1530, 2019. DOI: 10.5194/amt-12-1513-2019.

[120] M. Frey, F. Hase, T. Blumenstock, J. Groß, M. Kiel, G. Mengistu Tsidu, K. Schäfer, *et al.*, "Calibration and instrumental line shape characterization of a set of portable FTIR spectrometers for detecting greenhouse

gas emissions", *Atmospheric Measurement Techniques*, vol. 8, no. 7, pp. 3047–3057, 2015. DOI: 10.5194/amt-8-3047-2015.

[121] E. Avolio, S. Federico, M. Miglietta, T. Feudo, C. Calidonna, and A. Sempreviva, "Sensitivity analysis of WRF model PBL schemes in simulating boundary-layer variables in southern Italy: An experimental campaign", *Atmospheric Research*, vol. 192, DOI: 10.1016/j.atmosres.2017.04.003.

[122] E. Collier, T. Mölg, F. Maussion, D. Scherer, C. Mayer, and A. Bush, "High-resolution interactive modelling of the mountain glacier - atmosphere interface: an application over the Karakoram", *The Cryosphere*, vol. 7, no. 3, pp. 779–795, 2013. DOI: 10.5194/tc-7-779-2013.

[123] D. Rodgers and J. Connor, "Inter-comparison of remote sounding instruments", *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D3, 2003. DOI: 10.1029/2002JD002299.

[124] L. Fita, J. Polcher, T. Giannaros, T. Lorenz, J. Milovac, G. Sofiadis, E. Katragkou, *et al.*, "CORDEX-WRF v1.3: development of a module for the Weather Research and Forecasting (WRF) model to support the CORDEX community", *Geoscientific Model Development*, vol. 12, no. 3, pp. 1029–1066, 2019. DOI: 10.5194/gmd-12-1029-2019.

[125] H. Sembhi, H. Boesch, A. Agusti-Panareda, and S. Massart. "Final Report: MACC-III (Monitoring Atmospheric Composition and Climate 3)". (2016), [Online]. Available: https://atmosphere.copernicus.eu/sites/default/files/repository/MACCIII_FinalReport.pdf.

[126] J. Notholt, C. Petri, T. Warneke, M. Deutscher, M. Buschmann, C. Weinzierl, R. Macatangay, *et al.* "TCCON data from Bremen (DE)". (2014), [Online]. Available: https://doi.org/10.14291/tccon.ggg2014.bremen01.R0/1149275.

[127] M. Deutscher, J. Notholt, J. Messerschmidt, C. Weinzierl, T. Warneke, C. Petri, P. Grupe, *et al.* "TCCON data from Bremen (DE)". (2014), [Online]. Available: https://doi.org/10.14291/tccon.ggg2014.bialystok01.R1/1183984.

[128] S. Tukiainen, J. Railo, M. Laine, J. Hakkarainen, R. Kivi, P. Heikkinen, H. Chen, *et al.*, "Retrieval of atmospheric CH_4 profiles from Fourier transform infrared data using dimension reduction and MCMC", *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 17, pp. 10–312, 2016. DOI: 10.1002/2015JD024657.

[129] S. Basart, A. Benedictow, Y. Bennouna, A. Blechschmidt, S. Chabrillat, E. Cuevas, A. El-Yazidi, *et al.*, "Validation report of the CAMS near-real-time global atmospheric composition service: Period September-November 2019", *Copernicus Atmosphere Monitoring Service*, 2020.

[130] M. Galkowski, C. Gerbig, J. Marshall, F. Koch, J. Chen, S. Baum, A. Jordan, *et al.*, "Airborne in-situ measurements of CO_2 and CH_4 and their interpretation using WRF-GHG: results from the HALO CoMet 1.0 campaign.", in *EGU General Assembly Conference Abstracts*, vol. 21, Vienna, Austria, 8–13 April 2019, 2019.

[131] Q. Tu, F. Hase, T. Blumenstock, R. Kivi, P. Heikkinen, M. Sha, U. Raffalski, *et al.*, "Atmospheric CO_2 and CH_4 abundances on regional scales in boreal areas using Copernicus Atmosphere Monitoring Service (CAMS) analysis, Collaborative Carbon Column Observing Network (COCCON) spectrometers, and Sentinel-5 Precursor satellite observations", *Atmospheric Measurement Techniques*, vol. 13, pp. 4751–4771, 2020. DOI: amt-13-4751-2020.

[132] M. Gałkowski, A. Jordan, M. Rothe, J. Marshall, F. Koch, J. Chen, A. Agusti-Panareda, *et al.*, "In situ observations of greenhouse gases over Europe during the CoMet 1.0 campaign aboard the HALO aircraft", *Atmospheric Measurement Techniques*, vol. 14, no. 2, pp. 1525–1544, 2021. DOI: 10.5194/amt-14-1525-2021.

[133] S. Gourdji, A. Karion, I. Lopez-Coto, S. Ghosh, K. Mueller, Y. Zhou, C. Williams, *et al.*, "A modified Vegetation Photosynthesis and Respiration Model (VPRM) for the eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models", *Earth and Space Science Open Archive*, p. 50, 2021. DOI: 10.1002/essoar.10506768.1.

[134] A. Kindler, H. Klimeczek, and U. Franck, "Socio-spatial distribution of airborne outdoor exposures–An indicator for environmental quality, quality of life, and environmental justice: The case study of Berlin", in *Urban Transformations*, Springer, 2018, pp. 257–279. DOI: 10.1007/978-3-319-59324-1_14.

[135] SenStadtH, *Berlin environmental atlas: 06.01 actual use of built-up areas/ 06.02 inventory of green and open spaces/ 06.01.1 actual use / 06.02.1 actual use and vegetation cover (edition 2016)*, http://www.

stadtentwicklung.berlin.de/umwelt/umweltatlas/e_text/eke601.pdf, 2016.

[136] Fraunhofer-Gesellschaft, *Map of Power Plants | Energy Chart, data source: European Energy Exchange (EEX)*, https://www.energy-charts.de/osm.htm, Assessed August 14th, 2018, 2018.

[137] T. Wei, J. Wu, and S. Chen, "Keeping track of greenhouse gas emission reduction progress and targets in 167 cities worldwide", *Frontiers in Sustainable Cities*, p. 64, 2021. DOI: 10.3389/frsc.2021.696381.

[138] B. Fasoli, J. Lin, D. Bowling, L. Mitchell, and D. Mendoza, "Simulating atmospheric tracer concentrations for spatially distributed receptors: updates to the Stochastic Time-Inverted Lagrangian Transport model's R interface (STILT-R version 2)", *Geoscientific Model Development*, vol. 11, no. 7, pp. 2813–2824, 2018. DOI: 10.5194/gmd-11-2813-2018.

[139] M. Makowski, "Automating the Stochastic Time-Inverted Lagrangian Transport model, STILT", Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[140] F. Klappenbach, J. Chen, A. Wenzel, A. Forstmaier, X. Zhao, T. Jones, J. Franklin, *et al.*, *Novel methane emission estimation method for ground based remote sensing networks*, ser. EGU General Assembly Conference Abstracts. 2022. DOI: 10.5194/egusphere-egu22-10604.

[141] J. Chen, F. Dietrich, A. Forstmaier, J. Bettinelli, H. Maazallahi, C. Schneider, T. Röck-mann, *et al.*, *Multi-scale measurements combined with inverse modeling for assessing methane emissions of Hamburg*, ser. EGU General Assembly Conference Abstracts. 2022. DOI: 10.5194/egusphere-egu22-11548.

[142] F. Dietrich, J. Chen, B. Voggenreiter, and X. Zhao, *Greenhouse gas emission estimate using a fully-automated permanent sensor network in Munich*, ser. EGU General Assembly Conference Abstracts. 2020, p. 19 237.

[143] A. Dayalu, J. Munger, Y. Wang, S. Wofsy, Y. Zhao, T. Nehrkorn, C. Nielsen, *et al.*, "Evaluating China's anthropogenic CO_2 emissions inventories: a northern China case study using continuous surface observations from 2005 to 2009", *Atmospheric Chemistry and Physics*, vol. 20, no. 6, pp. 3569–3588, 2020. DOI: 10.5194/acp-20-3569-2020.

[144] I. Durre, X. Yin, R. Vose, S. Applequist, and J. Arnfield, "Enhancing the data coverage in the Integrated Global Radiosonde Archive", *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 9, pp. 1753–1770, 2018. DOI: 10.1175/JTECH-D-17-0223.1.

[145] X. Zhao, J. Chen, J. Marshall, M. Galkowski, C. Gerbig, S. Hachinger, F. Dietrich, *et al.*, "A semi-operational near-real-time Modelling Infrastructure for assessing GHG emissions in Munich using WRF-GHG", in *EGU General Assembly*, [online], 2020. DOI: 10.5194/egusphere-egu2020-13164.

[146] J. Kostinek, A. Roiger, M. Eckl, A. Fiehn, A. Luther, N. Wildmann, T. Klausner, *et al.*, "Estimating Upper Silesian coal mine methane emissions from airborne in situ observations and dispersion modeling", *Atmospheric Chemistry and Physics*, vol. 21, no. 11, pp. 8791–8807, 2021. DOI: 10.5194/acp-21-8791-2021.

Publications and supervised thesis

Own Publications

[P-1] S. Bhattacharjee, J. Chen, L. Jindun, and X. Zhao, *Kriging-based mapping of space-borne CO_2 measurements by combining emission inventory and atmospheric transport modeling*, ser. EGU General Assembly Conference Abstracts. 2020. DOI: 10.5194/egusphere-egu2020-10076.

[P-2] J. Chen, F. Dietrich, A. Forstmaier, J. Bettinelli, H. Maazallahi, C. Schneider, T. Röck-mann, *et al.*, *Multi-scale measurements combined with inverse modeling for assessing methane emissions of Hamburg*, ser. EGU General Assembly Conference Abstracts. 2022. DOI: 10.5194/egusphere-egu22-11548.

[P-3] J. Chen, F. Dietrich, A. Wenzel, N. Nachtigall, A. Forstmaier, F. Klappenbach, X. Zhao, *et al.*, *Permanent Urban Column Network for Carbon Emission Monitoring in Munich*. 2020, vol. 2020, B108–0024. DOI: 2020AGUFMB108.0024C.

[P-4] F. Dietrich, J. Chen, B. Voggenreiter, and X. Zhao, *Greenhouse gas emission estimate using a fully-automated permanent sensor network in Munich*, ser. EGU General Assembly Conference Abstracts. 2020, p. 19 237.

[P-5] F. Dietrich, J. Chen, A. Wenzel, A. Forstmaier, F. Klappenbach, X. Zhao, N. Nachtigall, *et al.*, *Urban methane emission estimate using measurements obtained by MUCCnet (Munich Urban Carbon Column network)*, ser. EGU General Assembly Conference Abstracts. 2021. DOI: 10.5194/egusphere-egu21-12210.

[P-6] F. Dietrich, J. Chen, A. Wenzel, M. Rißmann, A. Forstmaier, F. Klappenbach, X. Zhao, *et al.*, *Urban emission estimates and validation of satellite-measured urban GHG concentration gradients using MUCCnet data (Munich Urban Carbon Column network)*. 2021.

[P-7] A. Forstmaier, J. Chen, F. Dietrich, J. Bettinelli, H. Maazallahi, C. Schneider, D. Winkler, *et al.*, "Quantification of methane emissions in hamburg using a network of ftir spectrometers and an inverse modeling approach [in review]", *Atmospheric Chemistry and Physics*, pp. 1–33, 2022. DOI: 10.5194/acp-2022-710.

[P-8] F. Klappenbach, J. Chen, A. Wenzel, A. Forstmaier, X. Zhao, T. Jones, J. Franklin, *et al.*, *Novel methane emission estimation method for ground based remote sensing networks*, ser. EGU General Assembly Conference Abstracts. 2022. DOI: 10.5194/egusphere-egu22-10604.

[P-9] F. Klappenbach, J. Chen, A. Wenzel, A. Forstmaier, F. Dietrich, X. Zhao, T. Jones, *et al.*, *Methane emission estimate using ground based remote sensing in complex terrain*, ser. EGU General Assembly Conference Abstracts. 2021, EGU21–15 406. DOI: 10.5194/egusphere-egu21-15406.

[P-10] L. Lan, J. Chen, X. Zhao, Y. Bai, H. Ghasemifard, and S. Hachinger, *Continuous urban in-situ CO_2 measurement at Munich using TDLAS-WMS method and VCSEL laser.* Ser. EGU General Assembly Conference Abstracts. 2019, vol. 21. DOI: EGU2019-1025.

[P-11] L. Lan, J. Chen, X. Zhao, and H. Ghasemifard, "VCSEL-based atmospheric trace gas sensor using first harmonic detection", *IEEE Sensors Journal*, vol. 19, no. 13, pp. 4923–4931, 2019. DOI: 10.1109/JSEN.2019.2901793.

[P-12] L. Lan, H. Ghasemifard, Y. Yuan, S. Hachinger, X. Zhao, S. Bhattacharjee, X. Bi, *et al.*, "Assessment of urban CO_2 measurement and source attribution in Munich based on TDLAS-WMS and trajectory analysis", *Atmosphere*, vol. 11, no. 1, p. 58, 2020. DOI: 10.3390/atmos11010058.

[P-13] M. Rißmann, J. Chen, G. Osterman, X. Zhao, F. Dietrich, M. Makowski, and M. Kiel, "Validation of OCO-2 target observations using MUCCnet - Is it possible to capture urban XCO_2 gradients from space? ", *Atmospheric Measurement Techniques*, 2022. DOI: 10.5194/amt-15-6605-2022.

[P-14] A. Shekhar, J. Chen, J. C. Paetzold, F. Dietrich, X. Zhao, S. Bhattacharjee, V. Ruisinger, *et al.*, "Anthropogenic CO_2 emissions assessment of Nile Delta using XCO_2 and SIF data from OCO-2 satellite", *Environmental Research Letters*, vol. 15, no. 9, p. 095 010, 2020. DOI: 10.1088/1748-9326/ab9cfe.

[P-15] X. Zhao, J. Chen, J. Marshall, G. Christoph, M. Galkowski, H. Stephan, F. Dietrich, *et al.*, "Understanding greenhouse gas (GHG) column concentrations in Munich using WRF [in review]", *Atmospheric Chemistry and Physics*, 2022.

[P-16] X. Zhao, J. Chen, J. Marshall, M. Galkowski, C. Gerbig, S. Hachinger, F. Dietrich, *et al.*, *A semi-operational near-real-time modelling infrastructure for assessing GHG emissions in Munich using WRF-GHG*, ser. EGU General Assembly Conference Abstracts. 2020. DOI: 10.5194/egusphere-egu2020-13164.

[P-17] X. Zhao, J. Chen, J. Marshall, M. Galkowski, C. Gerbig, S. Hachinger, J. Gensheimer, *et al.*, *Quantifying the impact of urban greenhouse gas emissions for Munich during the COVID-19 pandemic using WRF V3.9.1.1*, ser. EGU General Assembly Conference Abstracts. 2021. DOI: 10.5194/egusphere-egu21-13431.

[P-18] X. Zhao, J. Marshall, S. Hachinger, C. Gerbig, and J. Chen, *Analysis for Total Column CO_2 and CH_4 combining WRF-GHG Model with Differential Column Methodology (DCM).* Ser. European WRF-Chem User Workshop 2019. 2019.

[P-19] X. Zhao, J. Marshall, S. Hachinger, C. Gerbig, and J. Chen, *Analysis for Total Column CO_2 and CH_4 in Berlin using WRF-GHG combined with Differential Column Methodology (DCM).* Ser. EGU General Assembly Conference Abstracts. 2019, vol. 21. DOI: EGU2019-17611-1.

[P-20] X. Zhao, J. Marshall, S. Hachinger, C. Gerbig, M. Frey, F. Hase, and J. Chen, "Analysis of total column CO_2 and CH_4 measurements in Berlin with WRF-GHG", *Atmospheric Chemistry and Physics*, vol. 19, no. 17, pp. 11 279–11 302, 2019. DOI: 10.5194/acp-19-11279-2019.

Supervised and co-supervised theses

[T-1] A. Berkmann, "Applying WRF Modelling Framework using Docker", Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

[T-2] F. Dobler, "Validation of CO_2 WRF-Chem simulations with OCO-2 data for Munich", Study Project, TUM Engineering Faculty Building Geo Environment, Technical University of Munich, Germany, 2021.

[T-3] X. Guo, “Quantifying Carbon Dioxide Emission for Munich using a Bayesian Inversion Model”, Master thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[T-4] Y. Huang, “SIF-driven Net Ecosystem Exchange Estimate”, Master thesis & Forschungspraxis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

[T-5] E. Kim, “Building a CO_2 Inversion Modelling Framework for Munich”, Master thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

[T-6] E. Lang, “Utilization of the STILT model to visualize the Footprint and Trajectories of Trace Gases”, Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[T-7] S. Lyu, “Analysis of Methane Emissions in Munich based on STILT Model”, Master thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2019.

[T-8] C. Ma, “Development of the automatic WRF-GHG Modeling Framework in Munich”, Master thesis & Forschungspraxis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[T-9] M. Makowski, “Automating the Stochastic Time-Inverted Lagrangian Transport Model, STILT”, Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[T-10] A. Uzun, “Wind Error Analysis”, Forschungspraxis & Master thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

[T-11] C. Wang, “Meso-scale Atmospheric Modelling”, Hauptseminar, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2020.

[T-12] D. Wu, “Inverse Modeling”, Hauptseminar, Technical University of Munich, Germany, 2021.

[T-13] Q. Xu, “A brief Introduction of WRF-GHG”, Hauptseminar, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2020.

[T-14] H. Zhang, “Atmospheric Greenhouse Gas Simulation in Munich using WRF Model”, Forschungspraxis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2018.

[T-15] X. Zhang, “Generating Local Climate Zone Maps of the city Munch”, Bachelor thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2021.

[T-16] W. Zhao, “Atmospheric Greenhouse Gas Simulation in Munich using WRF-GHG Model during the COVID-19 pandemic”, Master thesis, Professorship for Environmental Sensing and Modeling, Technical University of Munich, Germany, 2022.

Acknowledgments

First and foremost, I would like to sincerely thank my supervisor Prof. Dr. Jia Chen, for her guidance in my academic and life over the past few years. Without her help, advice, expertise and encouragement, this work and related researches would not have happened. And special thanks to my mentor, Dr. Stephan Hachinger, who has always provided me with unconditional support and guidance to me and is the best mentor I have ever known.

I own my deepest gratitude to my collaborators from Max Plank Institute, Dr. Christoph Gerbig, Dr. Julia Marschall, Dr. Michal Galkowski, etc. Without their continuous optimism, enthusiasm, encouragement and support for all WRF related studies, this work would have been difficult to complete.

Many thanks and appreciations also go to my colleagues from ESM who have willingly helped me out with their ability, especially Florian Dietrich, Ankit Shehar, Johannes Gensheimer, Dr. Lijuan Lan, Dr. Homa Ghasemifard, Adrian Wenzel, Xiao Bi, etc.

For my parents Erling Zhang and Qingmin Zhao, thanks for all your support and understanding over the past four years, without which I would have stopped this work some time ago. Thank you to all my friends and other family members who have helped me survive all the stress and not letting me give up, especially my dear cousins Dr. Qi Cai and Yi Song. Special thanks to my closest friends, Sisi Liu, Jing Wang, Yihao Cao, Kun Wang, Chengyu Peng, Dr. Shengjia Wang, Dr. XingChen Dong, YuChen Wang and Lin Chen for their encouragement and everlasting unconditional support.

In addition, special wishes to thank Allen (Lun) Deng and TheShy (Seung-lok Kang), who have given me their strong moral support to get me through the tough time and inspired me to keep going over the last four years.

And finally, to Doubao, the loyalest friend who has always supported me and stayed with me during the toughest period of the COVID pandemic in the past two years.