Atmospheric Research 300 (2024) 107207

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Evaluation of the GRAMM/GRAL model for high-resolution wind fields in Heidelberg, Germany

Maximilian May [a,*], Simone Wald [a], Ivo Suter [c,d], Dominik Brunner [c], Sanam N. Vardag [a,b,*]

[a] Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany
[b] Heidelberg Center for the Environment, Heidelberg University, Heidelberg, Germany
[c] Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland
[d] ZHAW, School of Engineering, Meteorology, Environment and Aviation, Zürich, Switzerland

ARTICLE INFO

Keywords:
High-resolution simulation
Urban meteorology
Urban atmospheric transport

ABSTRACT

Independent verification of mitigation efforts for climate and air quality action in cities relies on inferring emissions from atmospheric concentration measurements. As emissions are dispersed in the atmosphere before they reach an instrument, the quantitative estimation of emissions requires an understanding of the atmospheric transport and associated uncertainties. In this study, we analyse the catalogue of steady-state flow fields generated by the Graz Mesoscale Model (GRAMM) coupled to the Graz Lagrangian Model (GRAL) for an entire year in Heidelberg, Germany. We use a loss function for the wind field selection, which assigns a best-matching catalogue entry to any given hour by exploiting observation data. We introduce a new loss function which finds an optimal balance between differences in wind speed and wind direction. We evaluate the performance of the model based on 15 meteorological measurement sites, of which 14 are in the inner high-resolution and building-resolving GRAL domain (12.5 km × 12.5 km, 10 m resolution). Performance metrics include mean bias (MB) and root mean square errors (RMSEs) of simulated and observed wind speed and wind direction for all individual stations. On average, we find a mean underestimation of wind speed of 0.14 ms^{-1} corresponding to about 7 % of the mean wind speed and a mean RMSE of 1.03 ms^{-1}. For wind direction, a mean overall bias smaller than 1° is achieved, but individual stations show larger biases (mean absolute bias: 37°), especially at stations where wind speeds are low on average. Evaluation benchmarks for mean biases of wind direction and wind speed of mesoscale models provided by the European Environmental Agency (EEA) are met at 11 and 14 out of 15 stations at low measurement heights, respectively. Recently suggested extended benchmarks for complex terrain are met at almost all stations. Additionally, for the first time, we analyse the model's ability to simulate the vertical wind profile and we analyse the benefit of implementing a wind profile measurement into the process. We find that the model does not fully capture the vertical profile in our setting. We further study the required measurement network size and find that a high number (> 6) of meteorological stations improves the selection of flow fields over the entire GRAL domain substantially. The conducted comprehensive analysis of the wind fields in the GRAL domain are the basis for detailed quantitative analysis of greenhouse gas and air pollutant emissions using the GRAMM/GRAL modelling framework.

1. Introduction

Cities and urban areas account for about 54 % of global population (UN Habitat, 2022) and contribute largely to greenhouse gas and air pollutant emissions. In future, this share is expected to further increase due to the continued global urbanization (UN Habitat, 2022). This imposes an immense responsibility on cities today and in the future, but offers many possibilities for focused environmental measures. Many cities already have ambitious goals to reduce emissions of greenhouse gases and air pollutants and to monitor the mitigation progress. The city of Heidelberg, for example, aims to reduce CO_2 emissions by focusing on 30 specific climate mitigation measures (City of Heidelberg, 2019). These cities actively contribute to climate action (Sustainable Development Goal 13), as well as to creating space for good health and well-being (Sustainable Development Goal 3) for their inhabitants (Li et al., 2018). To fully exploit the mitigation potential of urban areas, to

* Corresponding authors at: Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany.
E-mail addresses: maximilian.may@uni-heidelberg.de (M. May), sanam.vardag@uni-heidelberg.de (S.N. Vardag).

https://doi.org/10.1016/j.atmosres.2023.107207
Received 18 July 2023; Received in revised form 28 November 2023; Accepted 22 December 2023
Available online 30 December 2023

motivate, design and independently verify mitigation options, knowledge of the origin of the emissions and the atmospheric transport of these emissions is required. This knowledge may boost climate as well as air pollution action and enable effective planning and monitoring of mitigation action on a local level (Gurney and Shepson, 2021; Jungmann et al., 2022).

Concentration measurements in the urban area can provide an independent approach to estimating emissions at high resolution (Lauvaux et al., 2016) and, therefore, they can support city stakeholders in mitigation efforts (Mueller et al., 2021; Lauvaux et al., 2020). A measurement-based approach is vital if bottom-up emission inventories are not available, of poor quality or associated with large uncertainties, which is often the case for inventories at the sub-country and city level (Oda et al., 2019; Super et al., 2020). However, to infer the CO_2 emission strength from measured concentrations, one needs to separately account for the effects of atmospheric transport and emission strength on measured CO_2. Therefore, quantifying the quality of the atmospheric transport simulation is crucial to estimate emission strength. Assessing the atmospheric transport within the considered city, therefore enables verification of emission reductions and improvement of emission estimates and with that, contributes to a better understanding of the sources, sinks and mitigation options on local scale.

An appropriate model needs to be able to account for sub-urban atmospheric transport considering topography, land use as well as building structures. Furthermore, to be politically relevant, the model needs to be able to run for time periods of months and years such that different meteorological conditions and emission profiles can be considered. Combining city-wide simulations at building-resolving scale with long time periods is computationally not yet feasible with detailed computational fluid dynamics (CFD) models such as Large-Eddy-Simulation (LES) or Direct Numerical Simulation (DNS) models. A less accurate but computationally more efficient alternative is to use a Reynolds Averaged Navier Stokes (RANS) approach. An example of such a model is the GRAMM/GRAL system, which is capable of calculating high-resolution (10 m) wind fields over long time periods by using a so-called “catalogue approach” (Berchet et al., 2017a). These properties make the model well suited to support mitigation efforts in cities, if it simulates the atmospheric transport with high accuracy. Evaluating this requires an in-depth analysis of the meteorology of the atmospheric transport model before analysing the simulated concentration and finally estimating the emissions. So far, a long-term and city-wide evaluation of the GRAMM/GRAL model has only been performed for the cities of Zurich and Lausanne (Berchet et al., 2017a, 2017b), but their focus was on air pollution rather than on a comprehensive evaluation of high-resolution meteorology as performed here. Other analyses focused mainly on the evaluation of the GRAMM or GRAMM-SCI model (a new branch of the GRAMM model), the comparison of GRAMM-SCI with the Weather Research and Forecasting (WRF) model (Oettl, 2021; Oettl and Veratti, 2021) or on evaluating air pollution simulations (Almbauer et al., 2000; Oettl, 2014, 2015b; Romanov et al., 2020). To our knowledge an in-depth evaluation of GRAL wind fields over a large urban domain and time period has not yet been performed.

In this study, we analyse the wind fields generated by the GRAMM/GRAL model for a one-year period in Heidelberg. We analyse a period from 1st June 2021 to 1st June 2022 and apply the catalogue approach (see Sect. 2.2). We build on the results by Berchet et al. (2017a), but use a much denser network of weather stations in the GRAL domain. We evaluate improvements due to the availability of urban meteorology sites and evaluate the inner-city flow-field performance of GRAL. We suggest a new matching algorithm for the urban area of Heidelberg with an optimized loss function (see Sect. 2.2) and analyse how well the atmospheric transport within an urban area can be simulated with respect to mean bias (MB), mean absolute bias (MAB) and root mean square errors (RMSEs) of simulated and observed wind speed and wind direction (see Sect. 4). We further examine the vertical wind profile and analyse how many meteorological observations are necessary for the matching (see Sect. 4.4 and Sect. 4.5). Finally, we conclude and discuss possible next steps for model improvement.

2. Model description and catalogue-approach

2.1. GRAMM and GRAL model

GRAMM/GRAL is composed of the mesoscale model GRAMM and the coupled computational fluid dynamics model GRAL. For our study, we evaluate results obtained by employing GRAMM and GRAL versions 19.01 (Oettl, 2019a, 2019b). The non-hydrostatic GRAMM calculates mesoscale meteorological wind fields by solving the Reynolds-Averaged conservation equations of momentum, mass, potential temperature and humidity. The Reynolds stress tensor is computed using an eddy viscosity model. Turbulent viscosity is estimated using an algebraic turbulence model, which is described in Oettl (2019b). GRAMM considers a given topography and spatially resolved land-use classes, which the model translates into values of albedo, emissivity, soil moisture, surface roughness, heat conductivity and thermal diffusivity to simulate surface fluxes of heat, momentum, humidity and radiation. Steady-state wind fields can then be calculated in the model domain considering the influence of topography and land-use types.

GRAMM can be initialized by defining meteorological situations in terms of wind speed, wind direction and Pasquill-Gifford stability classes spanning from A: extremely unstable to G: extremely stable (Oettl, 2019b). These meteorological conditions constrain and initialize the wind speed, wind direction, vertical wind profile, temperature and pressure gradients of the GRAMM wind fields. Simulation of vertical transport uses a power law to compute initial wind profiles. Further details can be found in Oettl (2015b, 2019b).

We use 1008 meteorological situations as drivers for GRAMM. These situations are binned into 10° sectors for wind direction, separated into seven wind speed categories at 10 m above ground (0.25 ms^{-1}, 0.75 ms^{-1}, 1.5 ms^{-1}, 2.5 ms^{-1}, 4 ms^{-1}, 6 ms^{-1} and 7 ms^{-1}) and into seven Pasquill-Gifford classes following Berchet et al. (2017b). Physically unrealistic combinations such as stable or very unstable situations at high wind speeds are not simulated. For this study, we run GRAMM on a 100 m horizontal resolution to generate mesoscale meteorological fields. Thus, for every driving meteorological situation, a GRAMM wind field with 100 m resolution is computed.

For every one of the 1008 meteorological situations, the respective GRAMM wind field is used as input for the computation of GRAL wind fields. Within the GRAL domain, local effects of buildings and street canyons are considered additionally. GRAL is run at 10 m horizontal resolution within a model domain of about 12.5 km × 12.5 km (see Fig. 1). Vertically, the resolution is 2 m and the domain expands up to 800 m above the lowest elevation in the domain.

GRAL is nested within GRAMM, which means GRAMM wind fields are interpolated onto the GRAL coordinate grid, which means that three-dimensional wind fields of GRAMM are interpolated on the finer Cartesian grid and are used as inflow boundary for GRAL (see Fig. 1 for our GRAMM and GRAL domains in Heidelberg). The high resolution allows resolving street canyons and buildings. It is required to simulate the inner-city wind field considering the urban structures. The flow around buildings is explicitly simulated solving the RANS equations prognostically. Molecular viscosity, Coriolis and buoyancy forces are neglected and the recommended algebraic mixing-length model is used as eddy viscosity turbulence model (Oettl, 2019a). Passive tracers are released within the GRAL model to simulate their dispersion in a Lagrangian approach. We make use of a library of python scripts, which has been developed to ease the generation of required input files (Berchet et al., 2017b).

2.2. Selection of wind situations

The GRAMM/GRAL model is capable of simulating hourly steady-

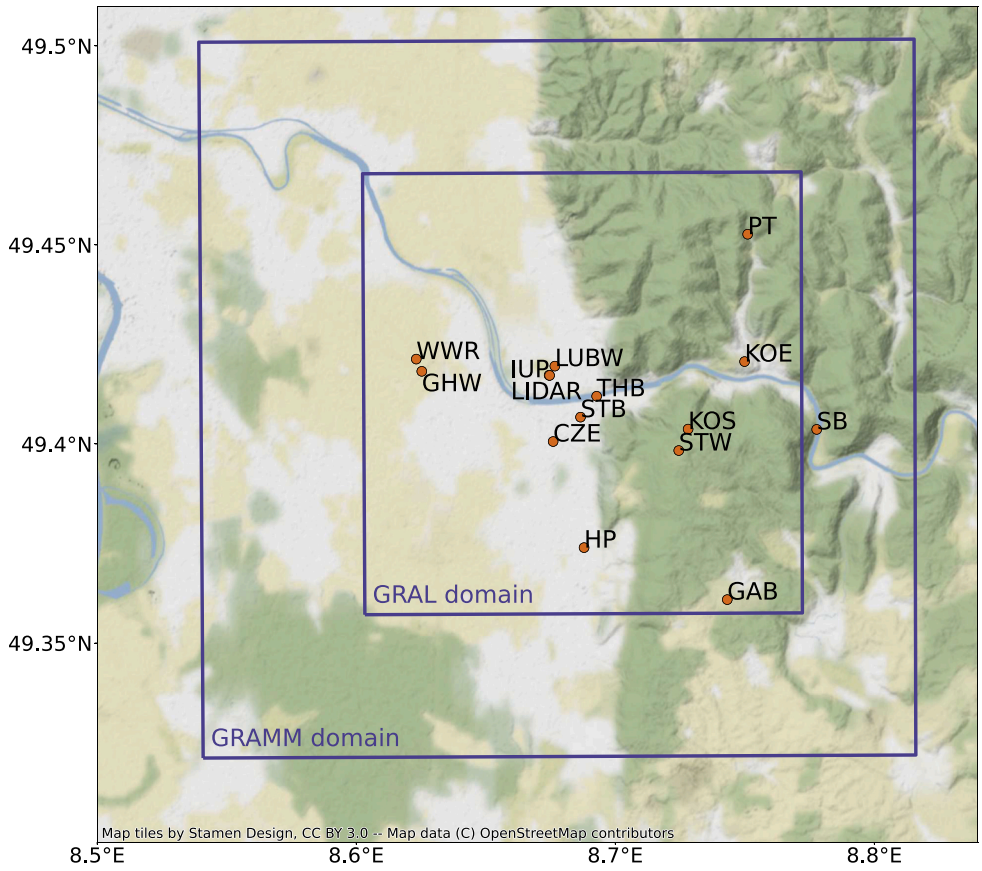

Fig. 1. GRAMM and GRAL domains used for Heidelberg and all meteorological stations used for the matching.

state wind fields. To simulate the dynamics of time-dependent meteorological fields, a catalogue approach is followed (Berchet et al., 2017a). This approach allows to simulate the atmospheric wind fields over long time scales with reasonable computational costs. For the catalogue approach, the model GRAMM is initialized with 1008 different meteorological situations differing in wind speed, wind direction and stability classes (see Sect. 2.1). A wide range of meteorological situations must be chosen to cover the full range of naturally occurring weather conditions in the domain of interest. The set of all resulting wind fields simulated by GRAMM, constitutes the "GRAMM catalogue".

The GRAL model uses the catalogue of GRAMM wind fields as input and calculates higher resolution wind fields taking the flow around buildings into account and thus, forming the "GRAL catalogue". Simultaneously, the particle dispersion can be calculated.

A time series of hourly wind fields (and concentrations) can be obtained by matching simulated to measured wind speed and wind direction at the locations of the measurement stations and selecting the situation from the catalogue which best matches the observed wind speeds and directions. This hourly matching procedure is able to save computational costs as the wind fields have to be computed only once and the matching is a computationally cheap minimisation procedure. Note that a matching frequency of less than hourly would be inconsistent with the physical assumptions of a RANS model, which can only represent temporally averaged but not instantaneous turbulent structures. This also implies that this hourly catalogue approach is only valid if the large scale wind fields do not change at time scales shorter than one hour.

Technically, for every hour h and every catalogue entry c, the matching algorithm calculates a loss function $\mathrm{L}_\lambda(h,c)$ based on differences between measured and simulated wind fields at the measurement sites for each catalogue entry. The entry which minimises the loss function $L_\lambda(h,c)$ is then selected as wind field for the respective hour. In principle, different loss functions are possible including L^1 and L^2 norm with and without weighting factors for high or low wind speeds. We have tested different formulations, including L^1 and L^2 norm as well as the one used by Berchet et al. (2017a) and found that we obtain best results in terms of RMSE and mean bias of the wind speed and wind direction if we use a compound loss function optimizing for wind direction and wind speed separately (see Eq. 1). Eq. 1 computes the loss L_λ for each measurement hour h and catalogue entry c based on the horizontal components of the wind vectors both measured (v_{hs}) and simulated ($\boldsymbol{u}_{cs}$) at the locations of all stations s:

$$L_\lambda(h,c) = \sum_s \left\{ \lambda \cdot \frac{d(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})}{\max\limits_{h,s,c} d(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})} + (1-\lambda) \cdot \frac{a(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})}{\max\limits_{h,s,c} a(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})} \right\}. \tag{1}$$

$a(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})$ denotes the angle between two vectors. $d(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs})$ is the difference of the magnitudes of the wind vectors multiplied with station-specific weights, which are defined as the inverse of the average wind speed thresholded to the median of all hourly measurements of wind speeds:

$$d(\boldsymbol{v}_{hs},\boldsymbol{u}_{cs}) = |\,\|\,\boldsymbol{v}_{hs}\,\| - \|\,\boldsymbol{u}_{cs}\,\|\,| \cdot \frac{1}{\langle v_s \rangle_h} \tag{2}$$

with

$$v_s = \begin{cases} v_{\mathrm{med}} & \text{if } \|\,\boldsymbol{v}_{hs}\,\| < v_{\mathrm{med}} \\ \|\,\boldsymbol{v}_{hs}\,\| & \text{else} \end{cases} \tag{3}$$

and v_{med} defined as the median of measured hourly horizontal wind speeds at all stations. $\|\cdot\|$ denotes the L^2 norm and $\langle v_s \rangle_h$ the averaged measured wind speed at the station over all hours. The station-specific weights are introduced in order to avoid that exposed stations with high wind speeds dominate the entire matching algorithm. By weighting

with the inverse average wind speed at the stations, the relative contribution of stations with large observed wind speeds is reduced. To prevent, in turn, overweighting the contribution of stations with low wind speeds, we clip all wind speeds to the median of all measurements before taking the station-average.

In eq. 1, the parameter $\lambda \in [0,1]$ determines how strongly the contributions from the magnitude distance and angular distance terms are weighted relative to each other. We first calculate the loss function for a set of possible λ values covering the full range from 0 and 1 and choose the best value λ^* ex post.

This is done by evaluating the best catalogue entries c^* as the ones minimising $L_\lambda(h,c)$ for each hour and used λ value:

$$c^* = c_{h\lambda} = \underset{c}{argmin}\, L_\lambda(h,c). \tag{4}$$

After this pre-selection of catalogue entries, λ^*, for which the mean L_2 norm between all $\boldsymbol{v}_{hs}$ and matched $\boldsymbol{u}_{c^*s}$ is smallest, is determined by minimising the average L^2 norm between the corresponding vectors:

$$\lambda^* = \underset{\lambda}{argmin} \left\langle L^2(\boldsymbol{v}_{hs}, \boldsymbol{u}_{c^*s}) \right\rangle. \tag{5}$$

Here, $\langle\cdot\rangle$ denotes the mean of the time average of each station.

The compound loss function aims at optimizing the two quantities which we are interested in and which we evaluate, namely wind speed and wind direction. Instead of simply applying the L^1 or L^2 norm, which optimizes only the RMSE of the wind speed vectors, or variations thereof (as done by Berchet et al. (2017a)), we only use the L^2 norm ex post to optimize the free parameter λ. As the contributing quantities have different units, we chose to define a unit-free loss function in which both terms are normalised such that their contributions lie in the range between 0 and 1 and are dimensionless. We achieve this by dividing the contributing terms by the maximum of all found contributions of the respective quantity defined across all hours, stations and catalogue entries. Note that a value of $\lambda = 0.5$ does not imply that speed and direction contributions are weighted equally because their physical contributions cannot be compared. For our setting, we find that values of $\lambda^* \approx 0.7$ are obtained from Eq. 5. Therefore, we use this λ^* for the entire matching period.

In order to prevent the selection of unphysical stability classes (e.g. unstable wind field during night), we have constrained the possible selected stability class dependent on large-scale wind speed of the catalogue and global radiation in the respective hour as given in Table A.3. This is the same pre-selection of catalogue entries as was already used by Berchet et al. (2017a). The large-scale wind speed refers to the model input and global radiation was measured at the Institut für Umweltphysik (IUP) (see Table 1). The pre-selection avoids choosing vastly different stability classes in successive hours. Note that this selection procedure hardly influences the overall performance as can be seen by the benchmarks listed in Table 2 (benchmarks when using pre-selected and full catalogue).

3. Heidelberg set-up and input data

3.1. Site description Heidelberg

Heidelberg is a medium-sized city counting about 160,000 inhabitants and is located in the densely populated Rhine-Neckar Metropolitan Region in southwestern Germany. The flat Upper Rhine Plain extends from Heidelberg to about 40 km further west. In the west part of Heidelberg, the Upper Rhine Plain ends by sharply defined mountains in the Odenwald (Königstuhl with 568 m a.s.l. and Heiligenberg with 445 m a.s.l.), through which the river Neckar has rutted a steep valley. The wind field in Heidelberg is influenced by mesoscale phenomena such as the deflection of air masses at mountains in the East of the city as well as by mountain-valley circulation patterns. On top of that, local features such as small valleys and urban canopy shape the wind field. This makes the simulation of meteorological variables especially challenging and calls for resolving topography at high resolution.

The GRAMM/GRAL model is especially designed to suit such complex topographies (Oettl, 2015a; Oettl et al., 2001, 2007). The GRAMM model domain is centered at the city and expands 20 km × 20 km (see Fig. 1) with a horizontal resolution of 100 m in 22 vertical layers with increasing thickness from bottom (10 m) to top (2470 m) above the lowest elevation within the domain. The domain comprises parts of the Upper Rhine plain as well as the Odenwald to consider the dominant topography features affecting the wind flows.

3.2. Model input data

GRAMM and GRAL utilize specific input data. For our Heidelberg setting, we briefly describe the used input data for topography, land use, building height, as well as meteorological data in the following.

The topography of the model domain is constructed from two data sets. For the GRAL domain of Heidelberg we use a 5 m-resolved digital terrain model from the *Landesamt für Geoinformation und Landentwicklung Baden-Württemberg* (LGL). In the GRAMM domain, we use the *ASTER Global Digital Elevation Model* by North American Space Agency (2019), which provides topographic data on 30 m resolution. Both data sets are joined such that they are interpolated to a regular 5 m grid of topography. Information on land use is taken from the European land cover data set CORINE version CLC2018 at a resolution of 100 m (Copernicus Programme, 2018). For the building data, we use level of detail 1 data (LoD1). Within a 4 km × 4 km box centered around Heidelberg, we have the LoD1 data from the LGL. Outside this box, we have extracted OpenStreetMap (OSM) (OpenStreetMap Contributors, 2004–2023) building shapes. The quality of OSM buildings has been

Table 1
Stations and characteristics of stations providing meteorological data for the selection of the simulated wind field.

Name	Station ID	Surface height [m.a.s.l.]	Meas. height above surface [m]	Remark
Grenzhöfer Weg	GHW	105	4	On bridge over railway tracks
Königstuhl	KOS	562	3	Mountain site in the forest
Peterstal	PT	350	3	Mountain slope site in the forest
Stadtbücherei	STB	124	10	Urban, roof with small platform
Wasserwerk Rauschen	WWR	106	10	Open field
Czernyring	CZE	140	10	Urban, roof
Landesstelle für Umwelt Ba-Wü	LUBW	112	10	Urban, street canyon
Hospital	HP	112	8	Suburban, street canyon
Sternwarte	STW	571	10	Mountain slope site in the forest
Schlierbach	SB	116	3	Valley site
Theodor-Heuss-Brücke	THB	114	8	Valley site on Neckar bridge
Köpfel	KOE	224	6	Open land
Gaiberg	GAB	321	10	Open land
Institut für Umweltphysik	IUP	143	7	Urban, roof, Global radiation measurements
LIDAR	LIDAR	143	10, 20, 42, 100, 150, 200	Urban, roof

Table 2
Mean observed (obs.) wind speed and benchmarks for all individual stations and their overall average. The values in brackets are the results when using the full catalogue, while the default values are the results when applying the pre-selection depending on the measured global radiation. The six different LIDAR heights were weighted each with a factor $\frac{1}{6}$ such that the LIDAR contributes to the average as much as any other station.

Station ID	Mean obs. Wind speed [ms^{-1}]	RMSE vel. [ms^{-1}]	RMSE dir. [°]	Bias vel. [ms^{-1}]	Bias dir. [°]	Abs. bias dir.[°]	Corr. coeff.
GHW	1.76	1.13 (1.11)	58.99 (57.01)	0.03 (0.00)	−8.67 (−6.38)	43.61 (41.53)	0.42 (0.43)
KOS	1.18	0.68 (0.65)	49.52 (50.26)	−0.16 (−0.23)	−15.97 (−14.76)	31.68 (33.02)	0.45 (0.50)
PT	0.80	0.74 (0.74)	76.52 (57.00)	0.14 (0.017)	0.32 (5.04)	57.11 (38.65)	0.27 (0.17)
STB	2.20	0.84 (0.82)	38.68 (36.26)	−0.23 (−0.31)	−3.87 (−3.05)	25.77 (24.41)	0.63 (0.66)
WWR	2.17	1.02 (0.96)	48.48 (48.14)	−0.42 (−0.43)	−5.52 (−4.37)	31.81 (30.61)	0.78 (0.81)
CZE	2.50	1.01 (0.94)	36.49 (34.09)	0.25 (0.15)	5.85 (5.71)	23.29 (21.28)	0.81 (0.83)
LUBW	1.34	0.57 (0.52)	47.51 (40.55)	−0.11 (−0.15)	−7.52 (−6.27)	33.28 (28.02)	0.76 (0.78)
HP	1.10	0.60 (0.60)	66.32 (69.45)	0.018 (0.01)	35.98 (37.53)	55.03 (57.44)	0.69 (0.73)
STW	2.45	0.91 (0.91)	47.82 (45.38)	0.02 (−0.02)	−18.36 (−18.07)	36.44 (34.64)	0.58 (0.69)
SB	0.93	0.75 (0.72)	66.70 (58.26)	−0.04 (−0.13)	2.97 (0.90)	50.31 (42.79)	0.32 (0.31)
THB	2.45	1.71 (1.63)	50.58 (43.65)	−0.81 (−0.77)	−7.06 (−3.18)	37.72 (31.55)	0.26 (0.29)
KOE	2.03	1.17 (1.03)	63.00 (55.14)	0.42 (0.26)	18.12 (15.50)	45.78 (39.72)	0.66 (0.71)
GAB	2.58	1.08 (0.96)	55.01 (41.69)	−0.24 (−0.34)	2.77 (2.17)	38.56 (28.24)	0.65 (0.74)
IUP	2.70	1.15 (1.11)	41.68 (33.63)	−0.41 (−0.49)	−7.83 (−4.06)	27.80 (22.26)	0.77 (0.81)
LIDAR10	2.80	1.25 (1.19)	34.35 (33.74)	0.46 (0.36)	1.74 (1.72)	21.98 (21.55)	0.44 (0.46)
LIDAR20	3.33	1.21 (1.17)	33.21 (32.57)	0.21 (0.09)	4.13 (3.84)	20.73 (20.45)	0.47 (0.50)
LIDAR42	3.91	1.33 (1.29)	33.51 (33.68)	−0.09 (−0.22)	5.74 (4.67)	20.76 (20.93)	0.47 (0.50)
LIDAR100	4.94	1.91 (1.86)	32.56 (32.47)	−0.82 (−0.95)	4.59 (3.70)	20.08 (20.01)	0.40 (0.43)
LIDAR150	5.75	2.35 (2.33)	34.08 (34.17)	−1.26 (−1.40)	1.28 (−0.24)	20.57 (20.59)	0.35 (0.39)
LIDAR200	6.34	2.82 (2.79)	36.45 (36.29)	−1.60 (−1.75)	−4.74 (−5.49)	22.75 (22.77)	0.31 (0.34)
Avg. values	**2.05**	**1.03 (0.96)**	**52.10 (46.96)**	**−0.14 (−0.20)**	**−0.44 (0.42)**	**37.29 (33.01)**	**0.56 (0.59)**

evaluated by comparing the OSM and LGL products within the 4 km × 4 km box and no significant mismatches have been found so that a good representation of building areas is expected for the surroundings of Heidelberg. However, where the information on building heights in the OSM product is missing, a default value of 10 m has been used, which is based on an estimation of the LGL data. Other physical obstacles such as trees, cars, containers or construction sites are not taken into account due to lack of data availability.

3.3. Meteorological data

The meteorological data comprises data on wind direction and wind speed collected by the city of Heidelberg, the winter service Heidelberg, the University of Heidelberg and the Pädagogische Hochschule Heidelberg. These data streams have been collected and provided by the Digital-Agentur Heidelberg. In total, we use 14 meteorological stations and additional LIDAR measurements at one site. The LIDAR uses Doppler-shifted back-reflection to measure wind velocity and direction at six height layers up to 200 m. The LIDAR measurements at each height are weighted by one sixth in order not to give too much weight to this location. All meteorological measurements are averaged to hourly means such that they can be compared to the hourly simulations. The locations in the simulated domain are shown in Fig. 1 and they are listed in Table 1. 14 of the 15 stations lie within the GRAL domain. An overview of the characteristics of the stations can be taken from Table 1.

While the catalogue of wind fields is simulated independently of measurements, after the simulation, the wind fields are matched to the observations at the stations to select the best-fitting wind field. Afterwards, the wind measurements at the stations are used to evaluate the performance of the simulated wind fields (Sect. 4). Using meteorological measurements for selection and verification is only valid if a large number of measurement sites is available such that the effect of every individual site on the overall agreement is small. Our analysis in Heidelberg benefits from a large amount of measurement stations in the GRAL domain, leading to a small influence of individual measurement stations on the chosen wind field. We show this in Sect. 4.5 and in Fig. B.9.

4. Evaluation of wind fields

We first describe typical simulated flow patterns for day and night time conditions qualitatively (Sect. 4.1) and then compare simulated and hourly matched wind fields to meteorological observations in the city for the entire year in terms of wind direction and wind speed (Sect. 4.2–4.5). We discuss performance and mismatches.

4.1. Flow patterns in Heidelberg

Fig. 2 shows typical and frequently selected flow fields at 10 m height above the surface for the afternoon (12–5 p.m. UTC, left) and the night time (0–5 a.m. UTC, right). The two representative situations correspond to the most frequently selected combination of wind speed and direction during the respective time period. In the north, west and south of the domain, the city is surrounded by agricultural fields on flat terrain with hardly any buildings such that the wind field in these parts of the domain is homogeneous and representative of the mesoscale wind flow. In the east of the domain, the wind field is strongly influenced by the mountains. Wind speed above surface is larger over the mountains and the wind direction changes as the wind flows around the mountains and through the Neckar valley. In the central urban area of Heidelberg, the wind is deflected by buildings (see zoom-in in Fig. 2c and d). One can see that the inner-city wind field captures the flow through the street canyons during day and night.

During day, the representative flow field is driven by wind from the south-west ($v = 2.5$ ms^{-1}, Stability class 4, see Fig. 2a). For this wind situation, we see that the wind is deflected into the Neckar valley where it slows down. On the windward side of the mountains the wind nestles around the mountain contours. The wind field outside the city center in the north, south and west is uniform as it is hardly influenced by urban structures or the mountains. Fig. 2c zooms into the simulated wind flow in an area of 300 m × 1000 m for a typical daytime wind situation. One can see that the buildings deflect and slow down the wind in the street canyons. If wind direction and street angle are not in parallel, the wind is deflected and slowed down in the street canyon.

During night, the representative flow field corresponds to wind from

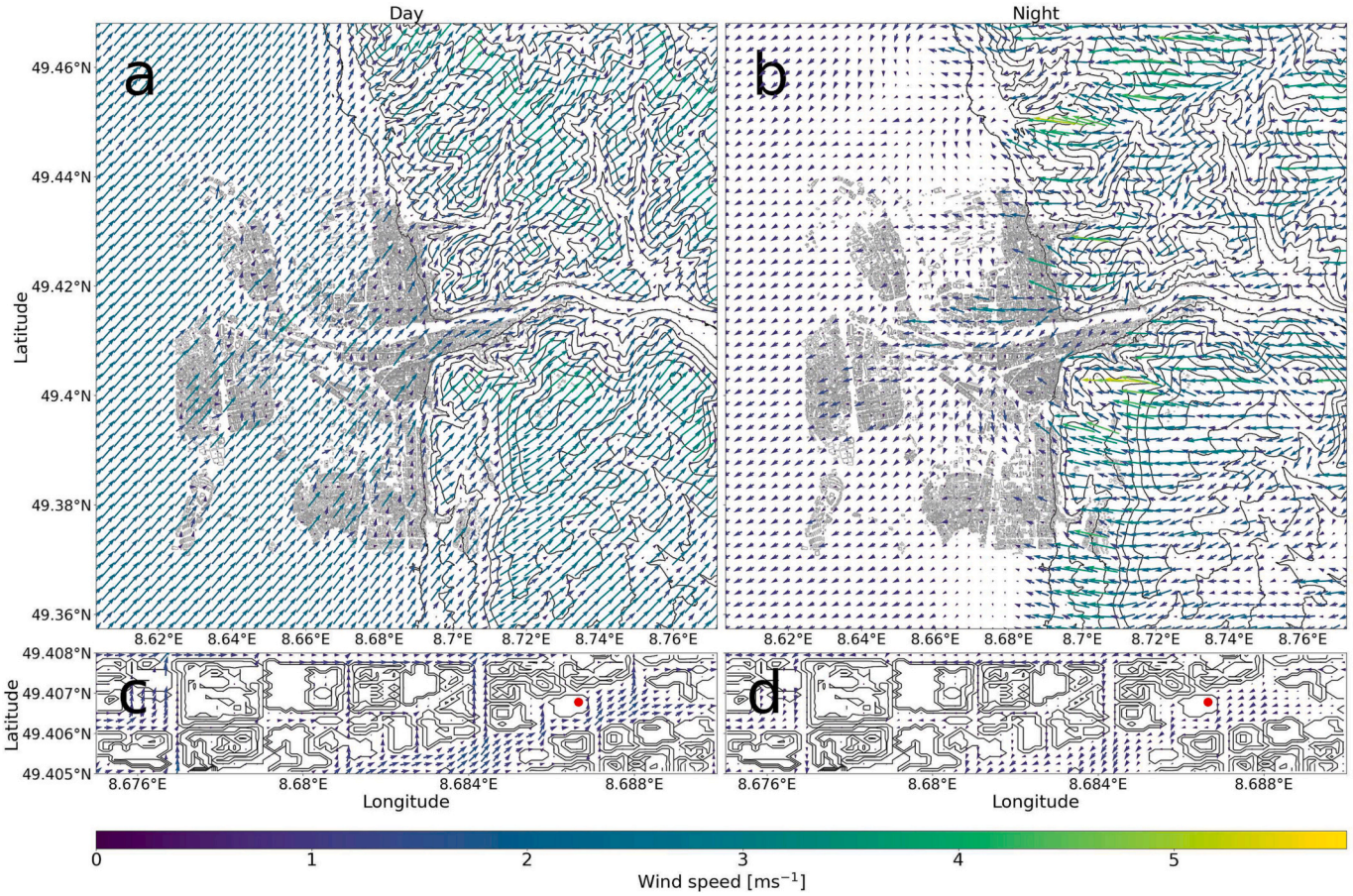

Fig. 2. Typical wind fields for day (a + c, 12–17 UTC) and night conditions (b + d, 0–5 UTC) at 10 m height above surface. During the day, wind often comes from the south-west (here: v=2.5 ms^{-1}, Stability class 4). During night, wind typically comes from the East (here: v=1.5 ms^{-1}, Stability class 6). Panel a and b show the entire GRAL domain and, for legibility, only every 24th wind vector in both dimensions. Panel c and d zoom into the flow field on an area of 300 m × 1000 m to show street canyon channelling by including only every second vector in both dimensions. The red dot denotes the station STB. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the east (v = 1.5 ms^{-1}, Stability class 6, see Fig. 2b). We observe the highest wind speeds on the mountain tops. In the Neckar valley winds are slower and are channelled along the river Neckar. In the west of the domain the wind speed is very low and as a consequence wind direction is more variable. Differences in wind vectors across the domain are smaller during the day than during the night period due to the higher stability during nighttime conditions. Zooming-in on the urban area at nighttime conditions (Fig. 2d) one again observes street canyon channelling. One can observe that the general wind field is flipped by about 180° for day and night. This is even more pronounced in the street canyons where the buildings direct the wind.

4.2. Wind direction

We now analyse the overall ability to simulate the domain-wide wind field in GRAMM/GRAL by comparing the measured and simulated wind roses over the course of one year (see Fig. 3).

When we first look at the observed wind roses across the domain, one can see that the wind roses of the individual stations differ from each other (see Fig. 3a). This again shows that the wind field is strongly influenced by the mountains in the east, by the steep valley and further by the urban structures in the center of the domain. At many stations one can observe that there are two dominating wind directions, which correspond to the two dominating wind regimes in Heidelberg during day and night presented in the previous section (see Sect. 4.1). Mountaintop stations showing this bi-directional behavior are KOS, KOE, GAB and to a smaller extent STW, which is very close to KOS but measures higher above the ground (see Fig. 1 for the precise locations of all stations; for reasons of legibility, locations of windroses are adjusted in Fig. 3). The dominant wind directions of the mountaintop stations are similar to the LIDAR measurement at 231 m above ground, which is located in the city, but due to the high altitude of the measurement is mostly determined by the large scale wind fields. Stations in the urban canopy such as LUBW, STB, CZE and IUP also show the bi-directional behavior. However, due to street chanelling, the wind roses follow the direction of streets at urban stations. PT is located on the slope of a mountain and the dominant wind directions follow the direction of the local slope. SB and THB are at the side or above the river Neckar, respectively (THB is on a bridge) and exhibit predominant wind directions in line with the Neckar valley.

When comparing measured (Fig. 3a) and simulated (Fig. 3b) wind fields, one can see a generally good agreement between simulations and measurements. Most valley channelling, as well as street canyon channelling is captured by the model. Simulations show similar predominant wind directions as the observations. As stations are distributed over the entire GRAL domain, the general good agreement suggests that the model is able of simulating the wind field over the entire domain well. This is especially true for stations located centrally in the urban areas.

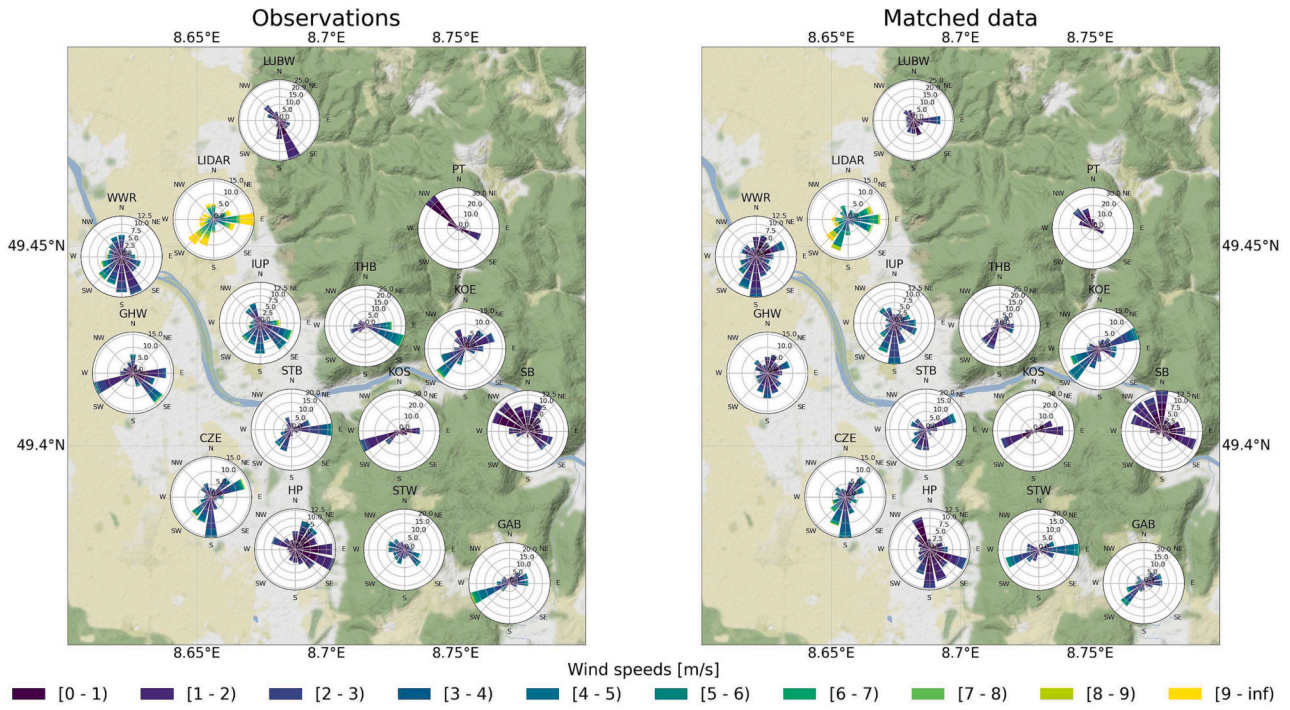

Fig. 3. Windroses of hourly measured data (a) and matched simulation results (b) for the evaluated time period of 1 year (01 June 2021 to 01 June 2022). GRAL wind fields were used for all stations located in the GRAL domain. For the LIDAR station, we show the data measured and simulated 200 m above the instrument. For the only station outside the GRAL domain (SB), the GRAMM wind fields were used. Note that when measurement stations were not working, we flagged and excluded the simulation results during the affected time period.

These stations are typically mounted on roofs and in street canyons. We find best matches in terms of wind directions for the stations STB, CZE, IUP and LIDAR with respect to RMSE and absolute bias (cf. Table 2). These stations are all located on top of buildings in built area demonstrating that GRAL is able to capture the wind deflection in built areas and street canyons.

Notable differences in wind direction patterns are found at the southern stations HP and STW. HP is located at a street junction and at the foot of the mountain. STW is geographically close to HP and is located at a mountain slope. While the observations at these stations show winds from all wind directions throughout the year, the simulation suggests two dominant, but two different wind directions for both stations. The station STW is surrounded by trees, which are not considered in the GRAL model and which could lead to reproducing the mesoscalic patterns rather than a deflection of wind direction due to trees in the simulation.

In general, we find that deviations in wind direction are higher for stations with low mean wind speeds, which is statistically expected (Papadopoulos et al., 1992). We observe a decrease of MAB with mean wind speed (see Fig. 4a). While wind direction biases can be large for low wind speeds (<1 ms^{-1}), they strongly decrease for higher wind speeds (>6 ms^{-1}). Station dependent differences occur, but all stations follow the general pattern of decreasing spread in bias with increasing wind speed.

To quantitatively evaluate mesoscale models, the European Environmental Agency (EEA) suggests benchmarks of mean bias of $\pm 10^\circ$ and mean absolute error of 30° for the wind direction (European Environment Agency, 2011). However, as pointed out by Oettl and Veratti (2021); Veratti et al. (2021), this benchmark is out of reach in areas with frequent low-wind speed conditions and in complex topography. Oettl and Veratti (2021), therefore, suggest a wind speed dependent wind direction benchmark for MAB of $\frac{46}{max(u,0.5)} + 25$, which is 117° for mean

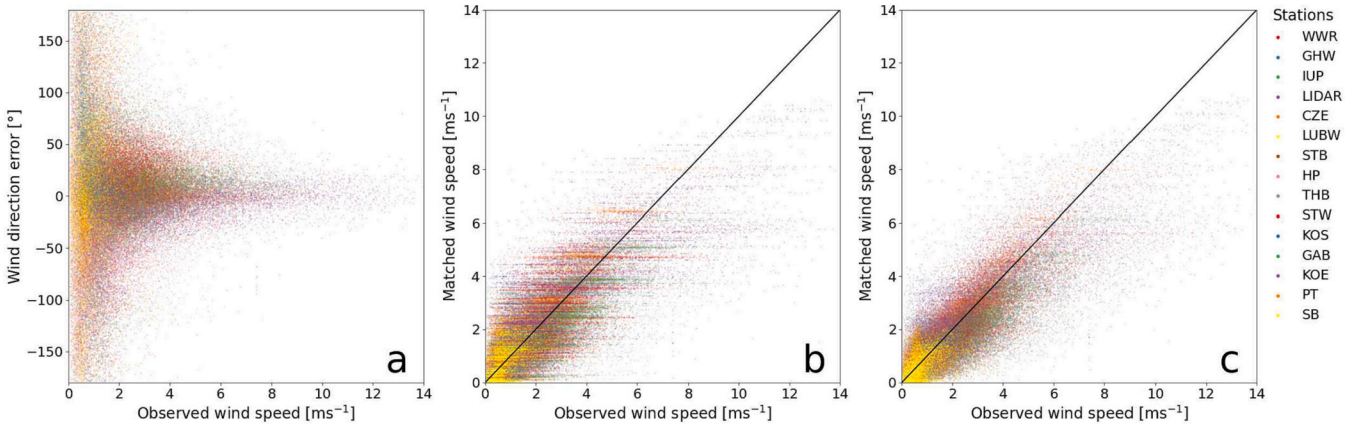

Fig. 4. a) Error in wind direction vs. observed wind speed, b) Matched vs. observed wind speeds with 1:1 line for all observation positions and times, c) same as b), but using an ensemble mean for the simulated wind speed ($n = 5$), see text. Stations are colour-coded as indicated.

wind speeds u below 0.5ms^{-1} and decreases to e.g. 32° for wind speeds of 7ms^{-1}.

The mean bias over all stations within the entire year analysed is −0.44°. But, for the individual sites, this bias is between −18° and 36° (see Table 2). From the 15 stations in total, 11 individually fulfill the high EEA criteria for mean bias but only 4 fulfill the criteria for mean absolute error. However, all sites fulfill the benchmark criteria for complex terrain suggested by Oettl and Veratti (2021). In terms of these benchmarks, we find that our results are comparable to similar studies in the Po valley or in Alpine regions (Oettl, 2021; Oettl and Veratti, 2021), which however focused on GRAMM wind fields rather than on the more heterogeneous GRAL windfields in urban areas.

4.3. Wind speed

Fig. 5 shows an exemplary two-week period of the observed and simulated wind speeds at 14 meteorological stations and at six LIDAR heights. Fig. 6 depicts the mean diurnal cycle for the entire year. Generally, the mean absolute wind speed increases with height above the surface. The highest wind speeds are obtained from the LIDAR measurements 200 m above the instrument. The lowest wind speeds are found at stations which are close to the surface and surrounded by large buildings (e.g. HP and LUBW) or trees (PT, STW, KOS). This behavior is expected (Oke et al., 2017). In most cases, the synoptic variability is well captured by the simulation. The onset of higher and lower wind speeds is reproduced by the model and the characteristics of the different stations are captured correctly suggesting that the model is able to simulate the urban wind speeds correctly.

Quantitatively, we find that stations with higher wind speeds show larger RMSEs and larger mean biases, but better correlations as is statistically expected (see Table 2). On average, the simulated wind speeds are slightly lower than the observed wind speeds (mean bias: −0.14ms^{-1}). This is only about 7 % of the mean wind speed of all stations and, therefore, less than half as large as found in previous studies (Berchet et al., 2017a). When plotting matched over observed wind speeds, we find that the data points scatter around the 1:1 line. (see Fig. 4b). However, the scatter is rather large and station-dependent differences exist as some stations are over- and some are underestimated.

Generally, we observe a small overestimation of lower wind speeds and an underestimation of high wind speeds. Especially, for situations with wind speeds larger than 8 ms^{-1}, we strongly underestimate wind speeds in the simulation. The reason is the choice of wind speed categories, where the highest prescribed synoptic wind speed is 7 ms^{-1}. Even though these situations do not occur very often, it is advised to extend the catalogue entries to higher synoptic wind speeds. We further observe a horizontal striping pattern, which indicates that the same wind speed category has to cover a range of different observed wind speeds for a location like Heidelberg. This is an obvious consequence from the limited number of catalogue entries for a large amount of hours of the year. To circumvent this pattern one can increase the number of catalogue entries and with that reduce the spacing between synoptic catalogue entries. As this requires the simulation of new GRAMM and GRAL runs, we have introduced an alternative ensemble approach, in which we select the five wind situations with the lowest loss according to Eq. 1. We average wind speed and wind direction of these five wind fields at all stations. By averaging over the five best wind fields, we expand the parameter room of possible wind speeds and directions at the stations and with that the horizontal striping pattern decreases (see Fig. 4c). When using this ensemble match, the simulated wind speed and direction change only marginally. Quantitatively, the overall benchmarks change from 47° and 0.96 ms^{-1} (RMSE of wind direction and wind speed) and 33° and 0.20 ms^{-1} (MAB for wind direction and MB for wind speed) to 47° and 0.92, and 33° and 0.26, respectively. To illustrate these marginal changes, we include this ensemble match in Fig.s 5–7 in addition to the matching result that we evaluate primarily in this study.

Fig. 6 shows the diurnal cycle in wind speed with generally larger values during the day for most stations. For the catalogue approach to be applied, we assume an hourly steady state, which actually cannot always be fulfilled e.g. during sunset and sunrise. However, we do not identify a time of the day when the model is not able to simulate the wind field. Stations measuring at high altitudes (LIDAR at 150 m and 200 m, as well as KOS on the mountain Königstuhl) show hardly any diurnal cycle in the observations.

The wind speed pattern is well captured in terms of magnitude, timing and diurnal cycle (see Figs. 5 and 6) and gives confidence in the capability of the model to simulate urban flow realistically over large domains and throughout the day.

We find that at some stations such as WWR, IUP or LIDAR we slightly underestimate the wind speed throughout the day. However, only few stations show larger discrepancies – namely the LIDAR, THB and PT. The simulation of the highest LIDAR measurements shows a diurnal cycle, which is not visible in the observations at that altitude. This hints at a failure of the current simulation to correctly capture the vertical profile. We will investigate this further in section 4.4. Furthermore, the simulated wind speed at the station THB over the river Neckar shows an opposite diurnal cycle, with slightly lower values during the night.

The THB site is located on a Neckar bridge within the Neckar valley. We see a strong underestimation of simulated wind speeds of 0.81 ms^{-1}, which corresponds to about 33% of the measured wind speed. Evaluating the diurnal cycle of wind speed (Fig. 6), the model is not able to capture the different behavior in the diurnal cycle of this station. We find that the highest differences between observations and simulations occur at night when the atmosphere encounters stable conditions and when the wind comes from the east, which is from the Neckar valley (see Fig. 2b). Possible reasons for this mismatch might be related to either not capturing the valley channelling in a complex topography, or due to the effect of the river Neckar itself. We have analysed the effect of default thermal conductivity of land-use type “river” in our model by setting it to the literature value for water, and found that the wind velocity changes substantially (up to 1 ms^{-1}) locally over the river especially for stable conditions - however the local effect at the station THB is very small (< 0.1 ms^{-1}) for the selected wind situations. Further, the station SB is also located in the valley very close to the river Neckar and captures the nocturnal hours correctly. While the thermal conductivity might not fully explain the mismatch at THB, our results suggest that it might have a large influence locally. Further analysis of simulated wind fields over water bodies may answer to which degree the thermal conductivity can influence the flow field. Finally, as only Level of detail 1 data are used as building shapes, the bridge is treated as solid building without air flow under the bridge, this might influence the wind speed. However, we have varied the height of the bridge between 1 m and 9 m and found that the difference in wind speed at the station ($v_{\mathrm{THB,1m}} - v_{\mathrm{THB,9m}}$) is positive and never exceeds 0.2 ms^{-1}. Thus, it cannot explain the mismatch at station THB.

The PT station is a forest station located on the mountain slope. As PT exhibits very small wind speeds, their RMSEs of wind speed are not large – however the behavior of wind gusts cannot be reproduced in the model correctly and the mean absolute bias in wind direction is large (57.11°). The bad performance of the model at this station may be due to the fact that trees are not taken into account as shapes, such that the flow around and above the trees is not represented well. However, also a deficiency of the model to simulate mountain-valley circulation cannot be ruled out. To identify possible reasons for better or worse model performance

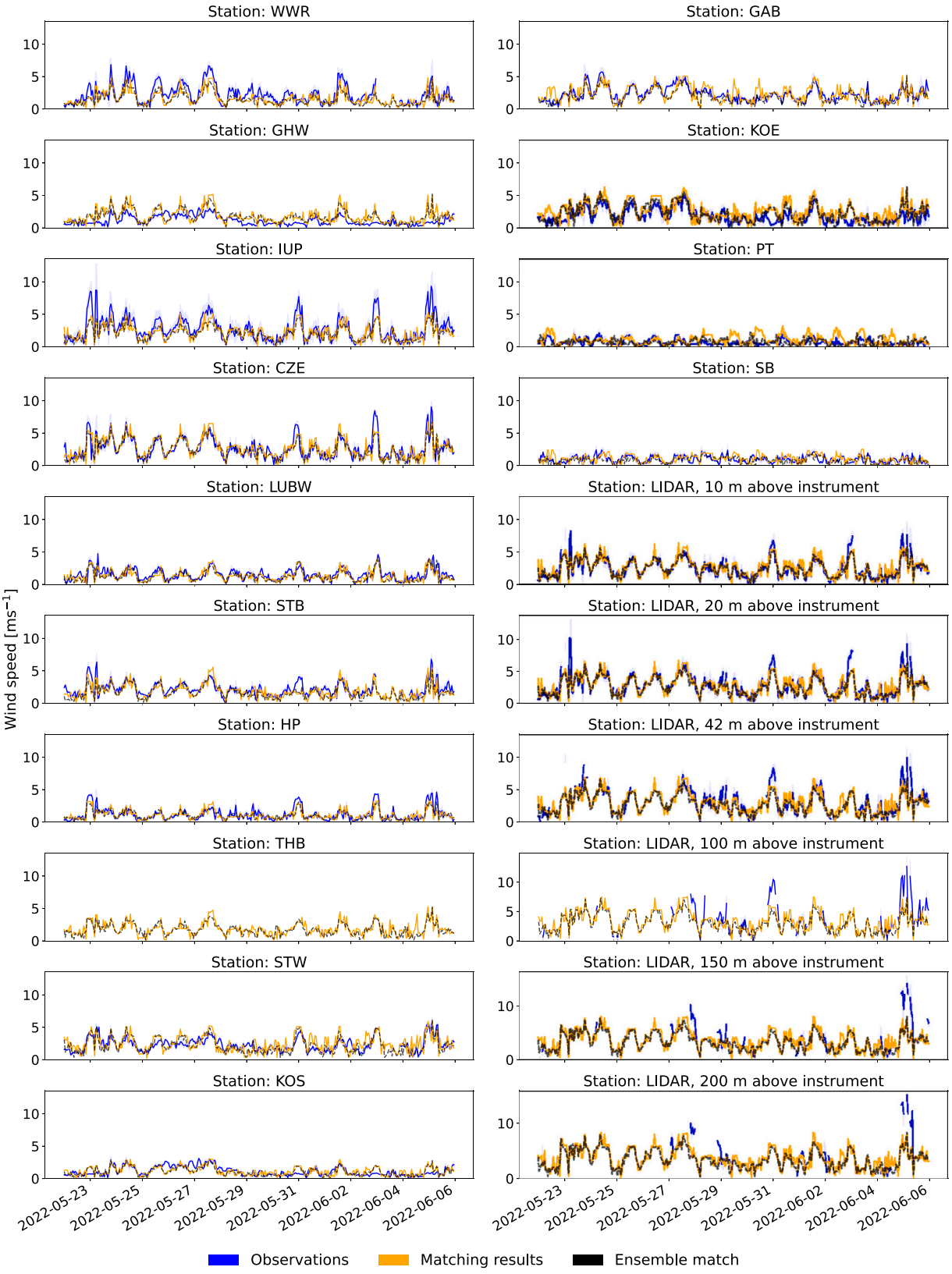

Fig. 5. Exemplary two-week period from May 22nd 2022 to June 5th 2022. Note that all measuring heights of the LIDAR stations are depicted here, but only an averaged LIDAR measurement is used for the matching. The period was chosen as many measurement stations were running during these two weeks.

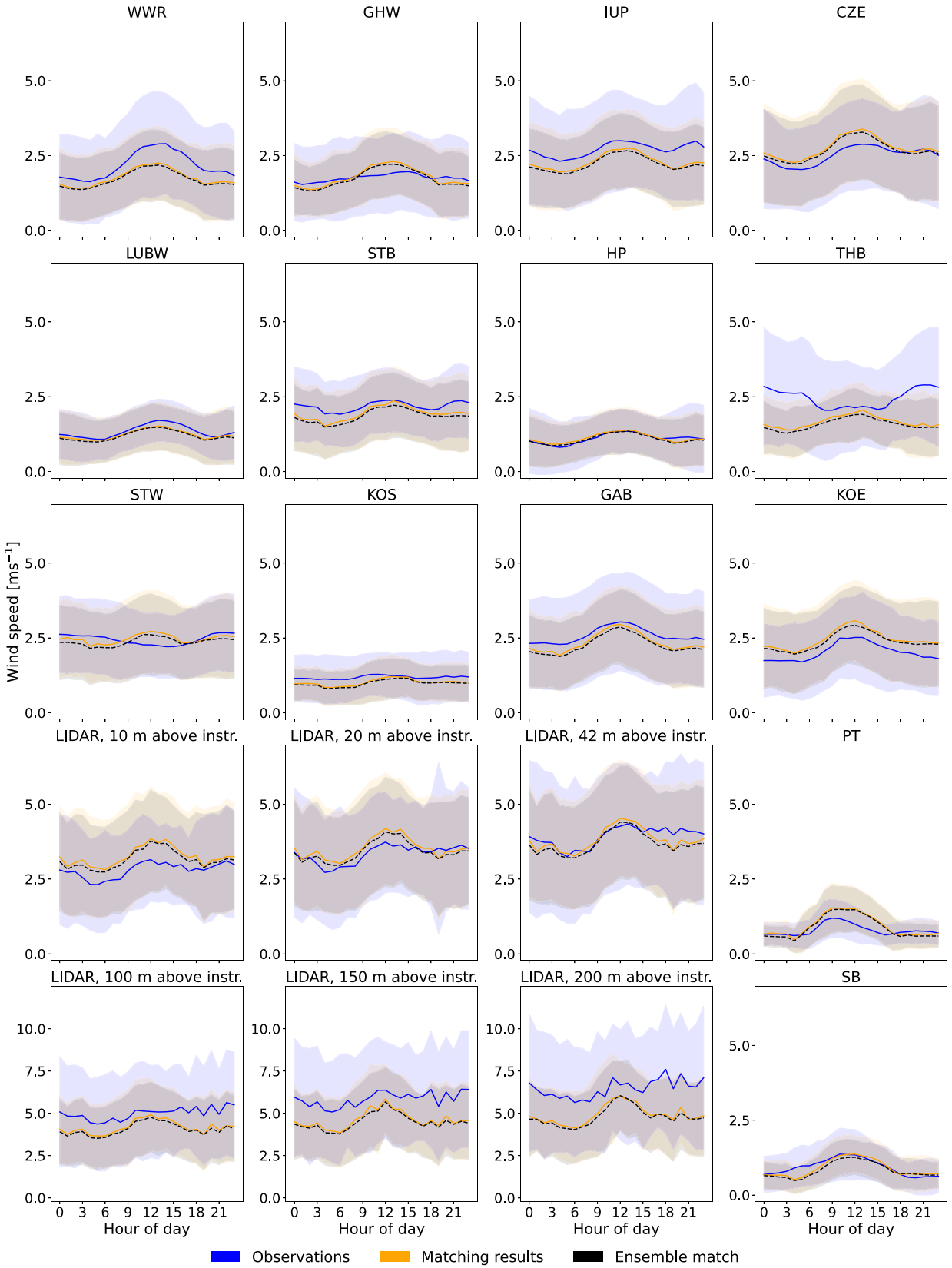

Fig. 6. Diurnal cycle of the annual wind speeds at all stations. Only times for which observation data is available are taken into account. Shaded areas represent the 1-sigma ensemble uncertainty.

at different stations, we grouped the stations in different categories. We analysed how well GRAMM/GRAL captures slope winds by comparing the statistics of stations located at the foot of or on a mountain (PT, KOE, STW, KOS, SB and THB) to all other stations. We find that RMSEs and correlation between simulated and measured wind speed are slightly worse, suggesting that the simulation of mountain-valley circulation is challenging even with GRAMM/GRAL, which was developed especially for complex topographies. However, differences are statistically not significant.

Similarly, we analysed if we find differences of the model's ability to simulate the wind flow over different land-use types (urban vs. agricultural vs. forest) or during different stability conditions (stable vs. neutral vs. unstable) and found only insignificant differences between these groups.

The EEA provides a benchmark for the model performance of wind speed in terms of mean bias and RMSE. The EEA benchmark for the mean biases is $\pm$ 0.5ms^{-1} and for the RMSE 2 ms^{-1}. As with wind direction, the EEA does not differentiate between simple and complex terrain. In contrast, Oettl (2021) states that the US-EPA distinguishes between simple terrain, which has the same benchmarks as the EEA, and complex terrain with benchmarks of 1.5 ms^{-1} for the mean bias and 2.5 ms^{-1} for the RMSE error.

The mean bias over all stations for the entire year is -0.14 ms^{-1}. For the individual sites, this bias can however be larger. It ranges from -1.6 to 0.46 ms^{-1}. Except THB, all sites measuring near the surface meet the high EEA-standards for the mean bias of wind speed and all of them meet the EEA-RMSE standard. However, all of them satisfy both the benchmarks for complex terrain for mean bias and for RMSE. At the LIDAR measurements at heights 100 to 200 m, the benchmarks are not met either. The EEA mean bias is missed at the three highest measurements and the EEA RMSE at the highest two measurement heights. Both benchmarks for complex terrain are missed only at the highest LIDAR measurement. Given the complexity of the terrain in Heidelberg and comparing with other studies in complex terrain (Oettl and Veratti, 2021; Veratti et al., 2021), we can confirm a good overall model performance in terms of mean bias and RMSE for wind speed in Heidelberg - especially near ground.

4.4. Vertical wind profile

So far, we have focused the analysis on horizontal wind field measured typically in a few meters height above surface or above buildings. However, in order to be able to draw conclusions on the model's ability to disperse particles within the domain, the stability and thus vertical profile of meteorological fields is important as they define the horizontal and vertical dispersion behavior. Previous studies encouraged an analysis of the vertical profile to determine if horizontal wind fields can constrain the weather situation accurately (Berchet et al., 2017a). To our knowledge, this is the first analysis of the vertical profile of GRAMM/GRAL. The initial vertical wind profile is described by a power law. The power law exponent is a function of the Monin-Obukhov length, which itself depends on surface roughness, stability class and constants which were empirically derived (Oettl, 2019b). Further details on how GRAMM/GRAL simulates the vertical wind profile can be found in Oettl (2015b, 2019b,a).

We analyse how well the model is able to represent the vertical profile and how the matching result is influenced by including measurements at multiple vertical levels. Fig. 7 shows the measured and simulated profiles for the LIDAR measurements at 10 m, 20 m, 42 m, 100 m, 150 m and 200 m above the instrument. The instrument (1 m height) was mounted on a building roof of 30 m height. In general, we observe a typical wind profile with wind speeds increasing with height. As expected, in the evening (18 h UTC), the observed wind profile shows more stable conditions than during the day (12 h UTC) when strong turbulent mixing reduces the wind speeds at higher elevations. For the lower heights of 10 m, 20 m and 42 m above the instrument (corresponding to 41 m, 51 m, 73 m above ground level) the observed and matched wind speeds agree very well on average. However, for the higher measurements (131 m, 181 m and 231 m above ground level), the observations are substantially higher than the simulated wind speeds. This behavior is consistent throughout the year (not shown here) and throughout the day, but largest in the evening when we expect more stable conditions (see Fig. 7). Thus, in our setting the simulations likely overestimate vertical mixing especially during evening hours, which in turn would lead to an underestimation of concentration at ground level. Berchet et al. (2017a) speculated their observed overestimation of near ground concentration was due to insufficient vertical mixing in unstable conditions, which is not what we observe in Heidelberg. However, in contrast to the evaluation of horizontal wind fields at ground level, this finding is only based on the vertical profile measurement at one position.

We have further analysed the effect of the matching routine itself on the simulated vertical profile (see Sect. Appendix C), but found that weighting the vertical wind measurements differently does not influence the vertical wind profile substantially. We expect the mismatch therefore to be an effect of the model and its current settings.

We analysed the effect of varying the surface roughness on the vertical profile, but found that it changes the matched profile only marginally and therefore does not explain the mismatch between simulation and observation. As the mismatch may influence the concentration at ground level, the detected mismatch should be further

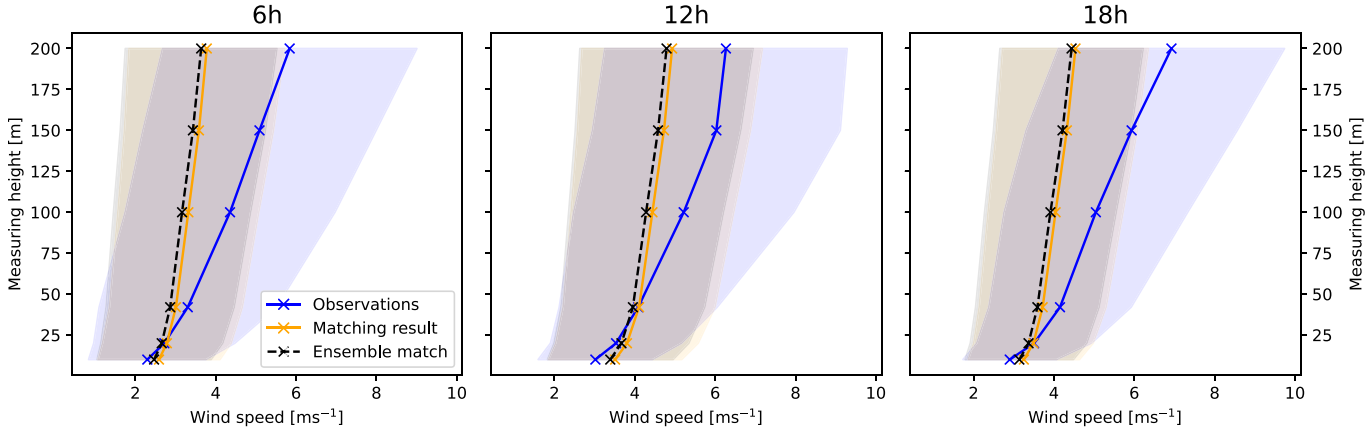

Fig. 7. Average wind speed profiles from observations and the matching results above the operation site of the LIDAR with standard deviations (1σ intervals) for different times of the day (2 h intervals around the indicated time) during the time period in which the LIDAR instrument was operating (January 24 to June 09, 2022).

investigated in the future. A possible reason for the mismatch is that the Monin-Obukhov theory was derived for flat homogenous terrain. In complex terrain comprising complex topographic and urban structures, the power law does not seem appropriate. It remains an open question if adjusting the parameters in the applied power law, which have been found empirically for flat homogeneous terrain, is sufficient to better represent the vertical profile in Heidelberg, or if instead the power law is not appropriate at all for our site. To further analyse this matter, measurements of vertical profiles at different positions in the domain are required.

4.5. Measurement network size

GRAMM/GRAL obtains its temporal dynamic by chaining the hourly steady-state wind fields. The hourly selection process of the wind fields is described in Sect. 2.2, and is based on minimising differences in wind speed and wind direction at all measurement stations. We expect that the flow field can be well reconstructed if there are enough measurement sites providing information about the flow field. However, as large measurement networks are costly, it is important to know how many stations are actually needed to obtain wind fields with sufficient accuracy. This analysis allows to efficiently balance additional benefit against the cost of installing and maintaining more measurement stations.

Therefore, we perform a leave-n-out experiment, in which we analyse the benchmarks RMSE and mean (absolute) bias of wind speed and wind direction at all stations. Fig. 8 illustrates the average performance in terms of RMSE and mean bias when employing n sites (black lines, left axes), as well as the enhancement in wind simulation performance achieved by adding a specific site to the n observation sites (colored lines, right axes). These benchmarks are derived from an exhaustive analysis of matching results for all possible leave-n-out experiments. In other words, for each set of observation sites (combinations for n sites), a new matching is performed, a new series of weather situations is generated, and the benchmarks for this sequence are calculated. We find that the overall RMSE decreases with increasing number of stations. This means that the performance of the selected wind field improves with increasing number of stations. However, one can see that RMSE approximately follows a power law, suggesting that changes in the benchmarks are large when there is only a small number of stations available, while the effect weakens as the number of stations increases. In particular, one can see that the RMSEs of wind speed and wind direction decrease with number of stations added as the matching algorithm minimises hourly absolute differences in wind speed and wind direction. The MB of wind speed converges to the average value when all stations are used (see Table 2). This value is not zero as the matching procedure does not optimize for total bias only (see Eq. 1). However, the MAB of wind speed is small compared to its RMSE.

Analysing the shape of the RMSE and M(A)B in Fig. 8, we find that for wind direction, we are not in saturation even for 15 stations. We expect that adding stations would further improve the result in terms of wind direction. For wind speed, the slope of the mean bias curve flattens after 6 stations. However, for RMSE, the wind speed curve does not completely flatten even for 15 stations, even though the decrease in magnitude of the slope is most pronounced for up to ~ 6 stations. This experiment highlights the importance of having sufficient meteorological stations within the high-resolution GRAL domain. We find that for Heidelberg the simulation of the GRAL domain with 12.5 km $\times$ 12.5 km, benefits from having 14 stations, roughly corresponding to 1 station every 3 km $\times$ 3 km. We further investigate the benefit of individual meteorological stations for the selection of wind fields.

Focussing on RMSE of wind speed and wind direction, we find that the stations which contribute most to an improvement, i.e. that including them reduces the RMSE most (right axes and colour code in Fig. 8), are the stations CZE and IUP followed by LUBW, WWR and STB. These stations are mainly located centrally in the GRAL domain. This result suggests that central meteorological stations best capture the main wind patterns in the entire domain. Including the station HP actually has

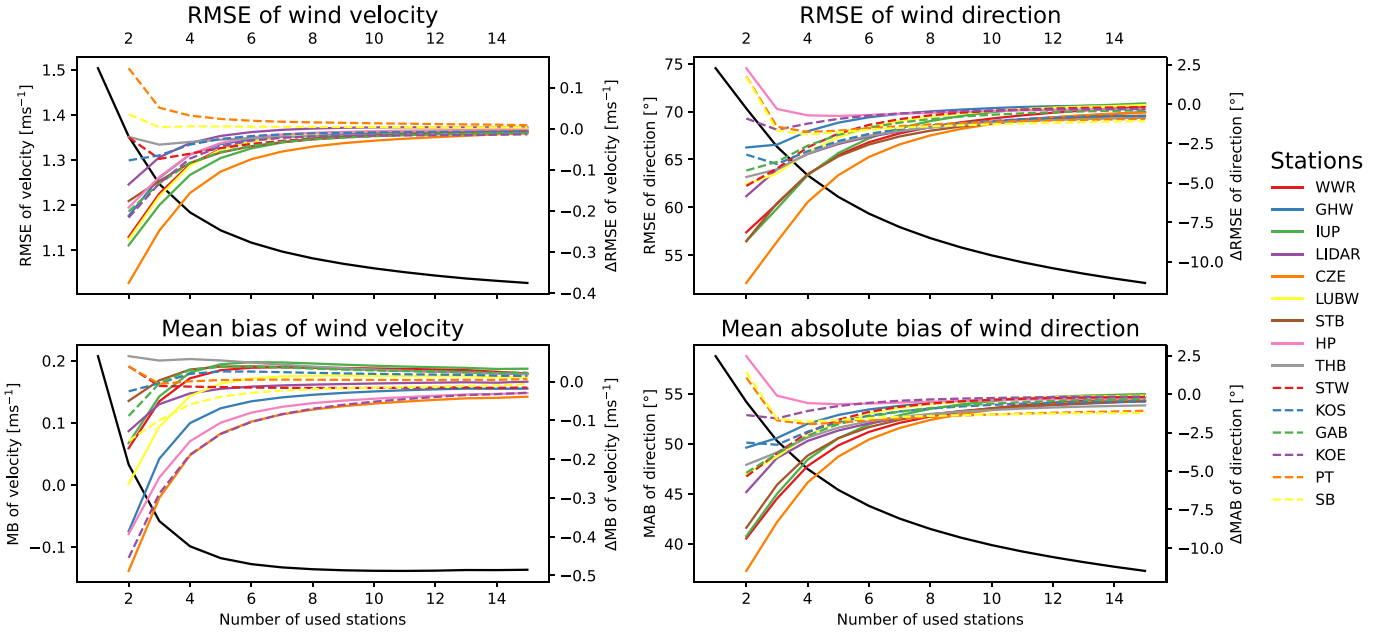

Fig. 8. RMSE and M(A)B of wind speed and direction (black lines and left y axes). Colored lines show the average effect of adding the respective station as n^{th} station to the set of stations used for the matching algorithm (right y axes).

a detrimental effect on the quality of the matching result and the inclusion of the valley station SB, the only station outside the GRAL domain, has no effect on the RMSE of wind direction and wind speed. This means that the model is not able to capture the signals measured at the station due to model or instrumental uncertainties. For a large number of stations, as the case in Heidelberg, the effect of individual stations is so low that we do not necessarily need to discard individual measurement sites from the matching procedure.

The large number of stations is also the reason why it is valid to use the same data for selection of the wind situations as well as for evaluation of the model, as exemplary shown in a leave-one-out cross-validation in Fig. B.9.

Overall, it is evident that for a small number of stations, the choice of stations is of large importance, but for a large number of stations the effect of adding a station is marginal. This is statistically expected, but highlights the importance of selecting measurement sites carefully, if only a small number of stations is available.

5. Conclusion and Outlook

In this study we analysed the performance of the GRAMM/GRAL model to simulate high-resolution meteorological conditions in terms of wind speed and wind direction over a large domain over Heidelberg for a period of one year. We have used the catalogue-approach to simulate a dynamic wind field by hourly selecting the best fitting pre-computed steady state wind field. For this selection, we have introduced and motivated a new compound loss function. We have demonstrated the model's ability to simulate the meteorological conditions in a complex topographic setting with site-specific mountain-wind systems, synoptic flow patterns and urban structures. The in-depth analysis of GRAL wind fields by comparison with 14 sites within the GRAL domain is novel. We found that the model is able to capture local wind patterns such as valley channelling and street canyon channelling very well with small mean biases (-0.14 ms^{-1}) and RMSE (1 ms^{-1}). Thus, we have shown that the matching procedure can be successfully applied to urban meteorological wind stations and does not require high quality or undisturbed wind station data as provided by e.g. the German Weather Service. This facilitates setting up GRAMM/GRAL in other urban areas if local sensor data is available. While the model is able to meet the high EEA benchmarks for mesoscale models for some measurement stations, it fails for other stations, which is a consequence of the complex terrain in Heidelberg. Other proposed benchmarks, which take into account the complexity of the topography, are met at almost all stations. In comparison to studies in other complex terrain (Oettl and Veratti, 2021; Veratti et al., 2021), the model is able to simulate the horizontal wind fields in Heidelberg very well. Increasing the number of synoptic forcing entries may further increase the model's ability to simulate the wind fields correctly. We especially expect to capture higher wind speeds only if the synoptic forcing is extended to higher wind speeds of up to 10 ms^{-1}.

We find that the model is not able to simulate the wind flow at the stations PT, THB and STW well. This might either be due to a deficiency of the model to simulate slope winds. It may additionally be a consequence of trees currently not being considered and/or of spurious behavior of the wind fields over water bodies.

In future, we suggest to take trees into account explicitly in the model, for which the model is already prepared. Also, the wind field over the Neckar at the THB station is not well captured. This mismatch calls for further investigating whether this is a common feature over rivers and other water bodies or this is a site-specific artefact.

For the first time, we have further analysed the vertical wind profile in a real urban setting by comparing the model results with wind LIDAR measurements. As the vertical wind profile changes with atmospheric stability, which in turn influences the dispersion of concentration (Oettl, 2019a), the correct estimation of the vertical profile is important in order to correctly simulate the inner-city concentration enhancements. Even though both, observations and model, show the expected increase of wind speed with height, we have found that the model is not able to capture the vertical wind profile fully. The simulations tend to show lower wind speeds at higher altitudes suggesting a larger vertical mixing than the observations. The discrepancy remained high also when weighting the LIDAR measurements stronger in the matching algorithm. This suggests that the mismatch in vertical profile is not an artefact caused by the matching, but rather by the model or model initialisation itself. Additional tests with adapted input height for synoptic forcing and wind profile comparisons at different location will be required to understand the discrepancies of the measured and simulated profiles and between simulated and measured concentration of air pollution or greenhouse gases. As the reason for the mismatch remains unclear and the possible implications for model users are important, we encourage further investigation of the systematics behind this mismatch.

We have shown the importance of utilizing a sufficient number of stations in an urban environment to select a flow field which is representative of the entire model domain. We have found that while the wind direction continues to benefit with increasing the number of stations to 15, the additional benefit for the wind speed flattens at about 6 stations. In principle, stations which are located centrally in the domain are most valuable for selecting the wind field.

Due to the reasonable computational costs, GRAMM/GRAL offers an ideal tool to plan and verify local mitigation on hourly time scales and in a large city-wide complex domain if it is able to capture the atmospheric transport processes correctly. The presented evaluation of the meteorological fields of the GRAMM/GRAL model is the first evaluation with a focus on the high resolution wind field. We have verified the ability of GRAMM/GRAL to capture city-wide horizontal wind fields and highlighted mismatches in the vertical profile, which may influence the simulation of air pollutants or greenhouse gases. This analysis lays the foundation for simulating the dispersion of emitted substances on high resolution.

Author contributions

M.M. performed the GRAL simulations, developed the analysis software and conducted the formal analysis, investigation and data curation. S.W. performed the GRAMM simulations and wrote parts of the software code. I.S. and D.B. supported the simulation set-up and interpretation of data and shared software code. S.N.V. conceptualised the work, set up the simulations and supervised the work. M.M. and S.N.V. wrote the original draft. All authors edited and reviewed the original draft.

CRediT authorship contribution statement

Maximilian May: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Simone Wald:** Methodology, Software, Writing – review & editing. **Ivo Suter:** Software, Writing – review & editing. **Dominik Brunner:** Software, Writing – review & editing. **Sanam N. Vardag:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Meteo data will be provided upon request. Topography, and building data cannot be made available by the authors. Code can be found on Github.

Acknowledgements

We thank Prof. Dr. André Butz for valuable discussions on the model results. For the provision of weather data, we acknowledge Heidelberg University of Education, Dep. of Geographie - Research Group for Earth Observation - rgeo (Head: Prof. Dr. Alexander Siegmund, Operator: Nils Hevendehl), Digital-Agentur Heidelberg (Esther Euteneuer) and City of Heidelberg, Office for Environmental Protection, Trade Inspection and Energy (Kai Schaupp, Dr. Raino Winkler). This research was partly funded by the German Research Foundation (DFG) within the Excellence Strategy, ExU 5.2 as granted by the Heidelberg Center for the Environment. Our project received funding from the Ministry of Science, Research, and Arts Baden-Wuerttemberg. For the publication fee we acknowledge financial support by DFG within the funding programme "Open Access Publikationskosten" as well as by Heidelberg University.

Appendix A. Allowed stability classes by global radiation

To pre-select the allowed catalogue entries by stability class, we apply the same scheme as Berchet et al. (2017a). The allowed stability classes ranging from A, "extremely unstable", to G, "extremely stable". Stability class D describes "neutral" conditions. The allowed stability classes by wind speed and global radiation are listed in Table A.3.

Table A.3
Allowed stability classes by wind speed of the large scale forcing as given as input to GRAMM and by global radiation as measured at the IUP station. This matrix is used to apply the pre-selection in the matching routine. It also shows that not all combinations of wind speed and stability need to be simulated, thus those combinations are not part of the catalogue.

Wind speed [ms^{-1}]	Global radiation [Wm^{-2}]				
	> 925	925–675	675–175	175–20	<20
0–2	A	A	B	D	F/G
2–3	A	B	C	D	E/F
3–5	B	B	C	D	D
5–6	C	C	D	D	D
>6	C	D	D	D	D

Appendix B. Leave-one-out cross-validation

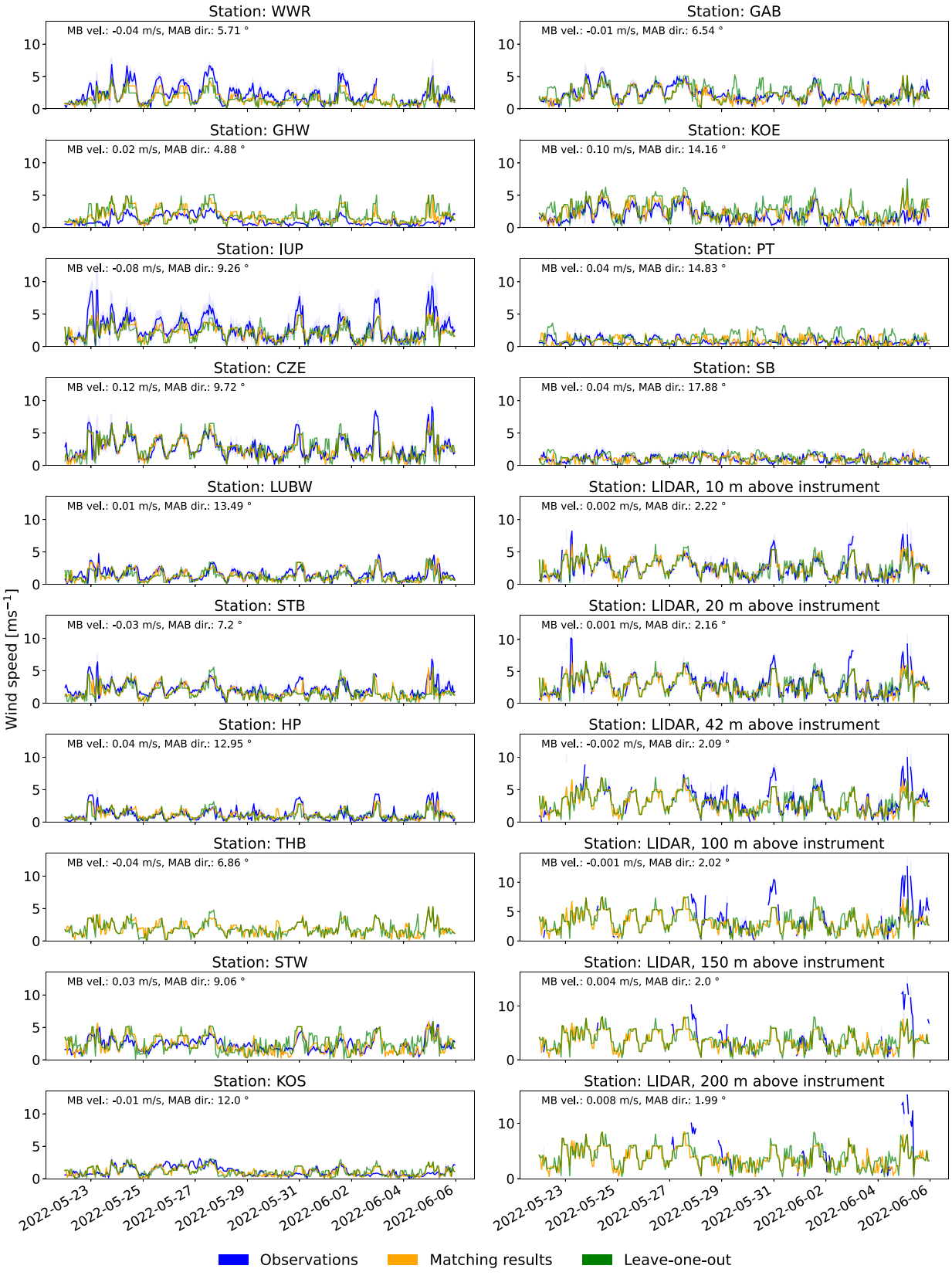

Fig. B.9. Results of the matching procedure used for the results discussed in this paper in comparison with the leave-one-out results. We display the same exemplary two-week period as in Fig. 5. For all stations, the MBs of wind speed and MABs of wind direction between the matching with the full set and the leave-one-out result. The similarity of the orange and the green lines is a visual indicator that the inclusion of the 15th station typically does not substantially change the result. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We performed a leave-one-out experiment, in which, for each station, we analysed the performance when the same station is not included in the set

of sites determining the matching result. Here, we show the two-week period also used in the main part of the paper, but comparing the results of the leave-one-out matching versus the full matching. The M(*A*)Bs at the individual stations between the two results are included in the panels of Fig. B.9, which shows the results for the same two-week period used in Fig. 5. One can see that differences are marginal and both matching results show the same features. For the entire year, the mean difference in wind speed between both matchings is 0.01 ms^{-1} and in wind direction −0.17°. The mean absolute difference in wind direction is 9.8° (LIDAR heights are always weighted as one sixth, as usual). The fact that the differences are so small confirms that it is justified to use the same stations for matching and validation.

Appendix C. Effect of matching on vertical profile

One possible explanation for the mismatch between observed and simulated vertical profile is that the wind measurements on ground are not sufficient for constraining the weather situation including the correct stability class. Therefore, we have analysed the effect of using more or less vertical profile information for the matching algorithm. We have run a matching without any LIDAR measurements included (scenario A), with six LIDAR measurements included (scenario B) and with six LIDAR measurements included, but each weighted one sixth (scenario C, as used in this analysis) and evaluate their results.

Between scenarios A and C, the chosen synoptic wind speed after matching differs in 9.4 % of the cases and the synoptic wind direction in 20.2% of the cases. Between scenarios B and C, these values are 8.5% and 39.8%, respectively. However, if only a slightly different wind situation is chosen, the effect on the overall matching result may be small. Therefore, we additionally calculated the benchmarks RMSE and M(*A*)B. The RMSE and mean bias of wind direction and speed differ by 0.01 ms^{-1} (RMSE wind speed) and − 0.33° (RMSE wind direction), 0.01 ms^{-1} (MB wind speed) and − 0.20° (MAB wind direction) in scenario A compared to C. The difference of scenario B compared to C is 0.01 ms^{-1} (RMSE wind speed) and 1.78°(RMSE wind direction), 0.01 ms^{-1} (MB wind speed) and 1.31° (MAB wind direction). Scenario C (i.e. LIDAR heights each weighted one sixth as used in this analysis) shows the best results in terms of RMSE of wind speed at all stations and second-best (after scenario A) for RMSE of wind direction. However, differences are smaller than 5% for all three scenarios suggesting that it is not due to the matching whether the vertical wind profile can be achieved, but rather an effect of the model and its current settings. Finally, we have tested if we can match the vertical profile, if we only use the six LIDAR measurements heights for the matching, but no other (ground) stations. We find that we still cannot capture the observed wind profile, which means that none of the simulated profiles captures the measured wind profile correctly. Therefore, the mismatch between observation and simulation is not due to the matching, but rather due to a deficiency of the model in its current set-up to offer a realistic parameter space of the vertical profile.

References

Almbauer, R., Piringer, M., Baumann, K., Oettl, D., Sturm, P., 2000. Analysis of the daily variations of wintertime air pollution concentrations in the city of Graz, Austria. Environ. Monit. Assess. 65, 79–87.

Berchet, A., Zink, K., Muller, C., Oettl, D., Brunner, J., Emmenegger, L., Brunner, D., 2017a. A cost-effective method for simulating city-wide air flow and pollutant dispersion at building resolving scale. Atmos. Environ. 158, 181–196.

Berchet, A., Zink, K., Oettl, D., Brunner, J., Emmenegger, L., Brunner, D., 2017b. Evaluation of high-resolution GRAMM–GRAL (v15. 12/v14. 8) NO x simulations over the city of Zürich, Switzerland. Geosci. Model Dev. 10 (9), 3441–3459.

City of Heidelberg, 2019. Der 30-Punkte-Aktionsplan für mehr Klimaschutz. https://www.heidelberg.de/hd/HD/service/22_11_2019+der+30-punkte-aktionsplan+fuer+mehr+klimaschutz.html, accessed: May 5, 2023.

Copernicus Programme, 2018. Corine Land Cover 2018. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018, accessed: May 5, 2023.

European Environment Agency, 2011. The Application of Models under the European Union's Air Quality Directive: A Technical Reference Guide. Publications Office of the European Union.

Gurney, K., Shepson, P., 2021. The power and promise of improved climate data infrastructure. In: Proceedings of the National Academy of Sciences, p. 118.

Habitat, U.N., 2022. World Cities Report 2022: Envisaging the Future of Cities.

Jungmann, M., Vardag, S., Kutzner, F., Keppler, F., Schmidt, M., Aeschbach, N., Gerhard, U., Zipf, A., Lautenbach, S., Siegmund, A., et al., 2022. Zooming-in for climate action—hyperlocal greenhouse gas data for mitigation action? Clim. Act. 1 (1), 8.

Lauvaux, T., Miles, N.L., Deng, A., Richardson, S.J., Cambaliza, M.O., Davis, K.J., Gaudet, B., Gurney, K., Huang, J., O'Keefe, D., et al., 2016. High-resolution atmospheric inversion of urban CO2 emissions during the dormant season of the Indianapolis flux experiment (INFLUX). J. Geophys. Res. Atmos. 121 (10), 5213–5236.

Lauvaux, T., Gurney, K., Miles, N., Davis, K.J., Richardson, S.J., Deng, A., Nathan, B.J., Oda, T., Wang, J.A., Hutyra, L., et al., 2020. Policy-relevant assessment of urban CO2 emissions. Environ. Sci. Technol. 54 (16), 10237–10245.

Li, M., Zhang, D., Li, C.-T., Mulvaney, K.M., Selin, N.E., Karplus, V.J., 2018. Air quality co-benefits of carbon pricing in China. Nat. Clim. Chang. 8 (5), 398–403.

Mueller, K., Lauvaux, T., Gurney, K., Roest, G., Ghosh, S., Gourdji, S., Karion, A., DeCola, P., Whetstone, J., 2021. An emerging GHG estimation approach can help cities achieve their climate and sustainability goals. Environ. Res. Lett. 16, 084003.

North American Space Agency, 2019. New version of the ASTER GDEM. https://www.earthdata.nasa.gov/news/new-aster-gdem, accessed: May 5, 2023.

Oda, T., Bun, R., Kinakh, V., Topylko, P., Halushchak, M., Marland, G., Lauvaux, T., Jonas, M., Maksyutov, S., Nahorski, Z., et al., 2019. Errors and uncertainties in a gridded carbon dioxide emissions inventory. Mitig. Adapt. Strateg. Glob. Chang. 24, 1007–1050.

Oettl, D., 2014. High resolution maps of nitrogen dioxide for the Province of Styria, Austria. Int. J. Environ. Pollut. 15 54 (2–4), 137–146.

Oettl, D., 2015a. A multiscale modelling methodology applicable for regulatory purposes taking into account effects of complex terrain and buildings on pollutant dispersion: a case study for an inner Alpine basin. Environ. Sci. Pollut. Res. 22 (22), 17860–17875.

Oettl, D., 2015b. Evaluation of the revised Lagrangian particle model GRAL against wind-tunnel and field observations in the presence of obstacles. Bound.-Layer Meteorol. 155 (2), 271–287.

Oettl, D., 2019a. Documentation of the Lagrangian Particle Model GRAL (Graz Lagrangian Model) Vs. 19.1.

Oettl, D., 2019b. Documentation of the prognostic mesoscale model GRAMM (Graz Mesoscale Model) Version 19.1.

Oettl, D., 2021. Development of the Mesoscale Model GRAMM-SCI: Evaluation of simulated Highly-Resolved Flow Fields in an Alpine and Pre-Alpine Region. Atmosphere 12 (3), 298.

Oettl, D., Veratti, G., 2021. A comparative study of mesoscale flow-field modelling in an Eastern Alpine region using WRF and GRAMM-SCI. Atmos. Res. 249, 105288.

Oettl, D., Almbauer, R., Sturm, P.-J., Piringer, M., Baumann, K., 2001. Analysing the nocturnal wind field in the city of Graz. Atmos. Environ. 35 (2), 379–387.

Oettl, D., Sturm, P., Anfossi, D., Castelli, S., Lercher, P., Tinarelli, G., Pittini, T., 2007. Lagrangian particle model simulation to assess air quality along the Brenner transit corridor through the Alps. Develop. Environ. Sci. 6, 689–697.

Oke, T., Mills, G., Christen, A., Voogt, J., 2017. Urban climates. Cambridge University Press.

OpenStreetMap Contributors, 2004–2023. OpenStreetMap. https://www.openstreetmap.org/, accessed: May 5, 2023.

Papadopoulos, K., Helmis, C., Amanatidis, G., 1992. An analysis of wind direction and horizontal wind component fluctuations over complex terrain. J. Appl. Meteorol. Climatol. 31 (9), 1033–1040.

Romanov, A., Gusev, B., Leonenko, E., Tamarovskaya, A., Vasiliev, A., Zaytcev, N., Philippov, I., 2020. Graz Lagrangian Model (GRAL) for pollutants tracking and estimating sources partial contributions to atmospheric pollution in highly urbanized areas. Atmosphere 11 (12), 1375.

Super, I., Dellaert, S., Visschedijk, A., Denier van der Gon, H., 2020. Uncertainty analysis of a European high-resolution emission inventory of CO2 and CO to support inverse modelling and network design. Atmos. Chem. Phys. 20 (3), 1795–1816.

Veratti, G., Bigi, A., Lupascu, A., Butler, T., Ghermandi, G., 2021. Urban population exposure forecast system to predict no2 impact by a building-resolving multi-scale model approach. Atmos. Environ. 261, 118566.