Impact of optimized mixing heights on simulated regional atmospheric transport of CO_2

R. Kretschmer[1], C. Gerbig[1], U. Karstens[1], G. Biavati[1], A. Vermeulen[2], F. Vogel[3], S. Hammer[4], and K. U. Totsche[5]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany
[2]Energy research Centre of the Netherlands, Petten, the Netherlands
[3]Laboratoire des sciences du climat et l'environnement, Gif-sur-Yvette, France
[4]Heidelberg University, Heidelberg, Germany
[5]Friedrich Schiller University Jena, Jena, Germany

Correspondence to: R. Kretschmer (rkretsch@bgc-jena.mpg.de)

Abstract. The mixing height (MH) is a crucial parameter in commonly used transport models that proportionally affects air concentrations of trace gases with sources/sinks near the ground and on diurnal scales. Past synthetic data experiments indicated the possibility to improve tracer transport by minimizing errors of simulated MHs. In this paper we evaluate a method to constrain the Lagrangian particle dispersion model STILT (Stochastic Time-Inverted Lagrangian Transport) with MH diagnosed from radiosonde profiles using a bulk Richardson method. The same method was used to obtain hourly MHs for the period September/October 2009 from the Weather Research and Forecasting (WRF) model, which covers the European continent at 10 km horizontal resolution. Kriging with external drift (KED) was applied to estimate optimized MHs from observed and modelled MHs, which were used as input for STILT to assess the impact on CO_2 transport. Special care has been taken to account for uncertainty in MH retrieval in this estimation process. MHs and CO_2 concentrations were compared to vertical profiles from aircraft in situ data. We put an emphasis on testing the consistency of estimated MHs to observed vertical mixing of CO_2. Modelled CO_2 was also compared with continuous measurements made at Cabauw and Heidelberg stations. WRF MHs were significantly biased by ∼10–20 % during day and ∼40–60 % during night. Optimized MHs reduced this bias to ∼5 % with additional slight improvements in random errors. The KED MHs were generally more consistent with observed CO_2 mixing. The use of optimized MHs had in general a favourable impact on CO_2 transport, with bias reductions of 5–45 % (day) and 60–90 % (night). This indicates that a large part of the found CO_2 model–data mismatch was indeed due to MH errors. Other causes for CO_2 mismatch are discussed. Applicability of our method is discussed in the context of CO_2 inversions at regional scales.

1 Introduction

Atmospheric tracer transport models are a crucial tool to predict air quality and atmospheric composition. This information is needed for environmental authorities and political decision makers. In addition, such models are regarded as an important tool to verify budgets of greenhouse gases and most importantly CO_2 (Nisbet and Weiss, 2010). Within the top-down approach dispersion models are used to close the scale gap between global models and point observations by simulating regional greenhouse gas transport (Dolman et al., 2009; Gerbig et al., 2009). Such models are needed to extract source and sink information in CO_2 signals, often obtained by observations performed within the planetary boundary layer (PBL), that show large variability near the omni-present source/sink processes targeted by the inversion estimates (Lin et al., 2003; Gerbig et al., 2003a, b, 2006). This signal variability is not only a consequence of variations of the terrestrial fluxes, but also of vertical mixing by atmospheric turbulence, which is hoped to be resolved adequately by the transport model. Denning et al. (1995) demonstrated the impact of PBL parameterizations used within transport

models on the distribution of atmospheric CO_2 due to the co-variance of photosynthesis/respiration and the mixing height (MH), both being a function of incoming solar radiation, at seasonal and diurnal scales. The MH is usually defined as the height up to which tracers emitted from surface get well mixed within about an hour (Seibert et al., 1998). On these short timescales the MH proportionally affects tracer concentrations in the PBL. For instance, the footprint of a CO_2 measurement, i.e. the spatially integrated surface influence on the measured signal, drops to 30 % after one day; thus the footprint very close ($\sim$50–150 km) to the observation site is most important and there the footprint simply scales with 1/MH (Gerbig et al., 2003b, 2008). As a consequence the MH is one of the most important parameters in air pollution and greenhouse gas transport modelling at regional scales and at the same time considered to be one of the major sources of uncertainty in CO_2 transport modelling (Stephens and Keeling, 2000; Gerbig et al., 2009). For instance, previous model–model and model–data comparisons of mesoscale models found differences in simulated MH $\sim$25–30 % during daytime over land (Sarrat et al., 2007a, b; Gerbig et al., 2008; Hu et al., 2010, Kretschmer et al., 2012). Gerbig et al. (2008) showed that MH discrepancies of this size lead to uncertainties of 3 ppm in CO_2, which corresponds to about 30 % uncertainty in regional fluxes, simulated in summertime over a domain covering most of Europe. During stable conditions mixing is sporadic and weak such that a clear definition of a MH is difficult (Seibert et al., 2000). Nevertheless, wind shear caused by surface friction can very well lead to the development of a mixing layer, and thus a MH can be diagnosed (Stull, 1988; Vogelezang and Holtslag, 1996; Seibert et al., 2000). As a consequence, model errors in MH at night are at least a factor two larger and are substantially biased (Gerbig et al., 2008), which has been shown to cause biases in simulated CO_2 concentrations (Kretschmer et al., 2012), and which in turn leads to potentially serious systematic errors in the retrieved fluxes. For daytime data such biases of the transport model are usually neglected in inversions, while night-time data obtained within the PBL are not used, to avoid biases in the inferred surface fluxes (e.g. Broquet et al., 2011). Because night-time data also contain useful information to constrain respiration fluxes and other emission sources, like anthropogenic fossil fuel CO_2 fluxes, this can be regarded as a major deficit of current CO_2 inversions (Dolman et al., 2009).

Previous studies have shown in synthetic data experiments the possibility to improve the simulated CO_2 transport by considering observed MH (Kretschmer et al., 2012, 2013). Kretschmer et al. (2013) interpolated MHs from point observations in space–time to a domain covering most of Europe using KED (kriging with external drift), which uses simulated MHs as a covariate to add physical constraints to the interpolation. This geostatistical approach yields optimized MH fields at the resolution of the meteorological driver fields produced by the Weather Research and Forecasting (WRF) model. Output from WRF and KED MHs were then used to drive the Lagrangian particle dispersion model STILT to simulate turbulent transport of CO_2. In a synthetic data experiment Kretschmer et al. (2013) demonstrated the effectiveness of this method to largely reduce bias and random errors in simulated CO_2 time series caused by MHs errors. The experiment assumed that the true MH was known in a European network of about 60 MH observations two times a day, comparable to existing radiosonde observations. Note that radio sounding networks have relatively good data coverage and are often used as benchmark for novel approaches for MH detection (Seibert et al., 2000).

In this paper we follow the approach of Kretschmer et al. (2013), using MHs derived from radio soundings in the Integrated Global Radiosonde Archive (IGRA) (Durre and Yin, 2008). Tracer transport simulated using STILT and driven by WRF meteorology is compared to observations made during the IMECC (Infrastructure for Measurements of the European Carbon Cycle[1]) aircraft campaign and continuous measurements made at Cabauw (CBW) and Heidelberg (HEI). We selected these two sites as they are known for their complexity – here an accurate model approximation of the MH is most relevant. At both sites high-quality, continuous observations exist for several decades. CBW has the further advantage of providing co-located meteorological measurements up to 200 m (a.g.l.), allowing MH detection in stable boundary layers. The use of real observations introduces two further complications in comparison to a synthetic data experiment: (1) the true MH is not known exactly due to measurement errors, data limitations and methodological uncertainty (Seidel et al., 2012) and (2) a verification of the method by comparing simulated and observed CO_2 abundances is rendered difficult as the observed signal is a result of both surface fluxes and transport. The first is the very same quantity that we have limited knowledge about and that an atmospheric inversion tries to solve for (or optimize). The latter is affected by other uncertainties besides those in MH (Gerbig et al., 2009), e.g. deep convection, or horizontal advection.

We tried to cope with complication (1) by objectively estimating the uncertainty for each individual MH estimate based on an analysis of high-resolution radiosonde profiles, which are part of the UK Meteorological Office (UKMO) database. This MH uncertainty was propagated through the KED estimation and evaluated with observed meteorology and CO_2 measurements obtained during the IMECC campaign in September/October 2009, to assure the consistency of the estimated MH to effective CO_2 mixing in the atmosphere, something we henceforth call the "effective MH". In addition we performed a cross-validation of the KED MHs using the IGRA MHs and compare to independent UKMO radiosondes not part of IGRA.
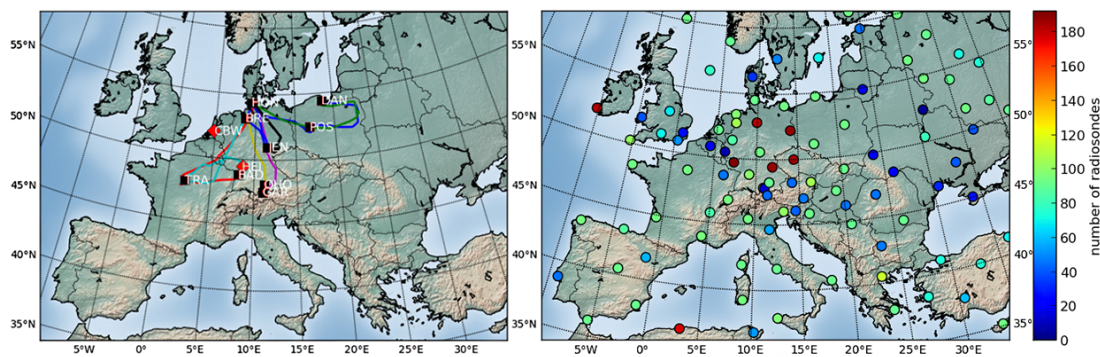
---

[1]Website: imecc.ipsl.jussieu.fr

Figure 1. Maps of the simulation domain. The left map shows flight tracks of the IMECC campaign. Coloured lines indicated individual flights. Black squares mark locations of profiles from start/landing of the aircraft or spiral flights. Heidelberg (HEI) and Cabauw (CBW) ground measurement stations are also shown (red diamonds). The right map shows the positions of radiosonde launch sites (coloured circles). Circle colours indicate the number of radio soundings available in the period 24 August 2009 to 10 October 2009. Station abbreviations are shown for the additional sites: Baden-Baden (BAD), Bremen (BRE), Garmisch-Partenkirchen (GAR), Gdańzk (DAN), Jena (JEN), Oberpfaffenhofen (OHO), Poznań (POS), Traînou (TRA).

The second complication is more difficult to tackle, because uncertainties in prior fluxes were shown to have substantial impact on simulated CO_2 concentrations (Peylin et al., 2011). To isolate the effect of transport errors on the CO_2 concentrations we prescribed the same CO_2 fluxes for all simulations – more specifically we compare results of two model setups with different PBL parameterizations, the Yonsei University Scheme (YSU, K-diffusion, Hong et al., 2006) and the Mellor–Yamada–Janjic scheme (MYJ, Turbulent Kinetic Energy, Janjic, 2002), prescribing the same vegetation and anthropogenic CO_2 fluxes. In addition, we utilize the independent auxiliary tracer CO, to assess the model performance in simulating trace gas transport. These two schemes are regarded as appropriate for the purpose of our study, because, firstly, they differ conceptually – the YSU scheme is based on K-diffusion and MYJ is solving for the budget of turbulent kinetic energy (TKE), and, secondly, it is known that MYJ produces weaker vertical mixing compared to YSU and other schemes (Hu et al., 2010); thus a significant divergence in simulated transport of CO_2 can be expected.

The objectives of our paper are to clarify the following questions: (1) What is the mismatch in CO_2 transport and can we improve CO_2 transport by reducing errors in MHs? (2) What is the model–data mismatch in MH of high resolution WRF simulations? (3) Can we adequately predict the MH from a limited set of data samples affected by measurement uncertainties for the whole European simulation domain? (4) Is the KED predicted MH consistent with the effective MH?

The content of the paper is structured in the following way. We start by introducing the IMECC campaign and continuous measurement site data. Some effort is devoted to explaining MH derivation from IGRA data and its uncertainty, as well as the KED approach. The method section concludes

with a summary of the WRF-STILT modelling system and flux inventories. In the first part of the results section we present comparisons of WRF MHs to IGRA data and evaluation of KED optimized MHs as a prerequisite for tracer transport with STILT. The second part shows the comparison of CO_2 mixing ratios from our four simulations (two PBL schemes, henceforth called STILT/MYJ, STILT/YSU and each with and without using optimized MHs from KED) with aircraft and ground-based in situ measurements. We then discuss our results with respect to flux and transport uncertainties, followed by a discussion of the potential of the method for regional CO_2 inversions and an outlook on further research.

2 Data and methods

2.1 Tracer observations and radio soundings

Figure 1 gives an overview on the simulation domain and location of available data sources, which are presented in the following.

2.1.1 IMECC campaign

The IMECC campaign was the first European aircraft campaign to calibrate six ground-based Fourier transform spectrometer (FTS) instruments that are used to retrieve column-averaged CO_2 for comparison to satellite measurements, e.g. from GOSAT. During the campaign from 28 September to 9 October 2009 eight flights were conducted. In 20 flight hours 12 000 km were flown with a special emphasis on spiral flights to retrieve vertical profiles usually covering an altitude from ∼0.3–13 km (Fig. 1, left). Measurements were also made during start and landing of the aircraft, providing some profiles starting from the surface, which allowed us to

evaluate tracer concentrations in stable boundary layers. CO_2 and CO was measured at 0.5 Hz and 1 Hz with a precision of 0.1 ppm and 2 ppb, respectively. From the measured meteorology and tracer profiles we selected five profiles for model–data comparison as further explained in Sect. 3.2.1. The detailed setup of the measurement equipment is described in Geibel (2011). Height above ground information was estimated from the aircraft altimeter and using output from a digital elevation model, i.e. the global 30 arc seconds topography map (GTOPO30, http://www1.gsi.go.jp/geowww/globalmap-gsi/gtopo30/gtopo30.html).

2.1.2 Ground observations: Cabauw and Heidelberg

The Cabauw 213 m tall tower with inlets for CO_2 and CO measurements at 20, 60, 120 and 200 m (a.g.l.), has been operated by the ECN (Energy Research Centre of the Netherlands) since 1992. CBW is located 25 km southwest of Utrecht, Netherlands (51.97° N, 4.93° E, −0.7 m a.s.l.) in an area of managed grassland. The area of 100 km around the tower contains a population of more than 7 million people. This local influence of significant contributions of sources and sinks makes the concentration footprint area of Cabauw one of the most intensive and complex source areas of greenhouse gases in the world, causing complex patterns in observed signals (Vermeulen et al., 2011). The sampling setup is described in Vermeulen et al. (2011). Meteorological observations of standard parameters like wind speed, temperature and humidity are made at altitudes 2, 10, 20, 40, 80, 140, 180, 200 m (Ulden and Wieringa, 1996). We use these meteorological measurements similarly to Vogelezang and Holtslag (1996) to obtain MH as is further explained in Sect. 2.2.

The second surface observation site is located in the suburbs of Heidelberg (49.417° N, 8.675° E, 116 m a.s.l.), within the highly populated Upper Rhine valley in southwestern Germany. The inlet for CO_2 and CO in situ measurements is situated on the roof top of the Institut für Umweltphysik, University of Heidelberg ∼30 m (a.g.l.). Levin et al. (2011) have shown the strong link between PBL mixing and observed CO_2 variability at HEI. Concentration observations are performed using the Heidelberg Combi-GC as described by Hammer et al. (2008).

2.1.3 Radiosonde data

Radiosondes (RS) are usually released one hour before the synoptic hours; most launches took place prior to the synoptic hours 00:00 and 12:00 UTC. The balloon rises with a speed of ∼5 m s^{-1}, i.e. it takes one radiosonde about 10–15 min to sample the whole PBL. The data consist of vertical profiles of pressure, temperature, relative humidity, humidity mixing ratio, sonde position, wind speed and wind direction for altitudes up to 20–30 km. The reported height is accurate to within ±40 m (assessed at UKMO station Aberporth, see http://badc.nerc.ac.uk/). Data are usually reported

at standard pressure levels at 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10 hPa. In IGRA additional levels are included whenever significant deviations from linearity in the logarithm of pressure between two standard levels are observed. We use the derived data set version 2 of IGRA (Durre and Yin, 2008) maintained by the NOAA National Climatic Data Center available at http://www1.ncdc.noaa.gov/pub/data/igra/derived-v2/. This special version of IGRA is thoroughly quality controlled and besides the standard meteorological variables provides derived quantities useful for studies of vertical structure, including geopotential height, derived moisture variables, and calculated vertical gradients of several variables (Durre and Yin, 2008). On average, IGRA soundings had 16 data levels (typically 13–19 levels) below a height of 500 hPa. Here we consider only soundings from which a MH > 0 m using a bulk Richardson number method (see Sect. 2.2) could be detected and with non-zero surface wind-speed measurement (see Eq. 1). In the period 24 August to 9 October 2009 we used 6722 (3417 daytime and 3305 night-time) soundings.

In order to obtain estimates on typical uncertainties in MH retrieved from radio soundings we assessed instrument noise from signal standard deviations in UKMO high resolution radiosonde profiles as further explained in Sect. 2.2.1. These we obtained from the British Atmospheric Data Centre (BADC), the Natural Environment Research Council's (NERC) designated data centre for the atmospheric sciences (http://badc.nerc.ac.uk/). The UKMO soundings contain 2 Hz data from UK stations and also from Gibraltar, St. Helena and the Falklands. Two of the sounding stations were not included in the IGRA database, providing an opportunity for validating the KED spatial interpolation of MHs derived from the IGRA soundings (Sect. 3.1.2).

2.2 Diagnosing the mixing height

It is known that methods to diagnose the MH from profiles of meteorological variables detect different features in a given profile. An obvious example is the detection of the top of the residual layer instead of the mixing layer in stable conditions, leading to systematically different MH estimates (Seidel et al., 2010, 2012; Seibert et al., 2000). This necessitates the consistent use of one single method to avoid methodological differences in comparisons. Furthermore, for a method to be useful for our purposes it is important that it provides estimates consistent with the effective MH of trace gases as further discussed in Sect. 3.2.1. Bulk Richardson number methods (Ri methods) were suggested for air pollution studies because they better correspond with the effective MH than other methods (Seidel et al., 2010, 2012; Seibert et al., 2000). Ri methods are also considered suitable for convective and stable boundary layers and allow for automatic processing of large amounts of data (Seidel et al., 2012). Note that there are different implementations of Ri methods with associated parameter values, most importantly for the critical

Richardson number Ri_c (Vogelezang and Holtslag, 1996). Here, we estimated the MH using the Ri method suggested by Vogelezang and Holtslag (1996):

$$Ri_g(h) = \frac{(g/\theta_{vs})(\theta_{vh} - \theta_{vs})(h - z_s)}{(u_h - u_s)^2 + (v_h - v_s)^2}, \tag{1}$$

where Ri_g is the Richardson number evaluated at each height h above the surface height s (here 17 m) given profiles of virtual potential temperature (θ_v), wind components (u, v) and height above ground (z), as well as the gravitational acceleration ($g = 9.81\,\mathrm{m\,s^{-2}}$). A given profile of Ri_g was linearly interpolated to the MH where $Ri_g = Ri_c$, with $Ri_c = 0.25$, which is the common value recommended in the literature (Vogelezang and Holtslag, 1996; Seibert et al., 2000; Seidel et al., 2012). For the value of the surface level height, Vogelezang and Holtslag (1996) tested 20, 40, and 80 m and found little sensitivity. Here we have chosen the height at the centre of the first WRF vertical layer, which is ∼17 m. IGRA profiles were linearly interpolated on a logarithmic pressure scale to 17 m from the surface measurement and the first upper air level > 17 m.

2.2.1 Estimation of MH uncertainty

The uncertainty of MH diagnosed from Eq. (1) was approximated following the method introduced by Biavati et al. (2014). Here we briefly summarize the basic steps of applying the method to our MH retrievals. First, the variability in the 2 Hz UKMO RS signal of the individual profiles of pressure, temperature, relative humidity and wind components was estimated for each radio sounding. This was accomplished by applying a running standard deviation on each of the available RS profiles, after removing local trends by subtracting a running mean of three data points from each data point in a given profile. The typical noise of each variable was then assumed to be uniform for all UKMO and IGRA soundings. Second, this noise of the RS signal was propagated as error variance through all calculations needed to get the Ri_g profiles, including Eq. (1) using standard statistical error propagation. This results in a discrete profile of errors, i.e. one error estimation for each discrete sample point in a given Ri_g profile. These error profiles were computed for all considered Ri_g profiles and were then used to estimate the MH uncertainty as follows. All profiles of a given radio sounding including the height above ground z and the Ri_g profiles are conceived as sequences of real numbers (data points) with common indices. The profile of errors for a given Ri_g profile has then been used to estimate the uncertainty of localizing the MH σ_{MH} within that profile:

$$\sigma_{MH}^2 = \frac{1}{l_1 + l_2} \sum_{k=m-l_1}^{m+l_2} (z_m - z_k)^2. \tag{2}$$

Here, z_m is the value at index m of the height profile of the given radio sounding at which the MH was localized using the Richardson method as described in Sect. 2.2. The numbers $l_1, l_2 \in \mathbb{N}$ are found by considering the heights z_k which are elements of the set U of physically consistent data points in the profile surrounding z_m, i.e. they are within the confidence neighbourhood of z_m:

$$U_{z_m} = \{z_i : i \in \{m-1, m, m+1\}\} \cup \\ \{z_j : m - l_1 \le j \le m + l_2,\ \zeta(m, j) \le \gamma\}, \tag{3}$$

where ζ is the measure of confidence based on Welch's t test statistic:

$$\zeta(i, j) = \frac{|Ri_g(z_i) - Ri_g(z_j)|}{\sqrt{\sigma_i^2 + \sigma_j^2}}. \tag{4}$$

In the denominator we use the variances of the error profile which corresponds to the Ri_g profile of a given sounding as described above. From Monte Carlo simulations it was found that values of $0 < \gamma \le 3$ are physically consistent. For our purposes we set $\gamma = 2$. The advantage of this method is the ability to express the uncertainty on a per sounding basis, instead of deriving a statistic of general uncertainty over all profiles as was done e.g. by Seidel et al. (2012). To account for additional uncertainty caused by the low resolution of IGRA RS we assumed additional 50 m uncertainty, which is based on the analysis of Seidel et al. (2012). The estimated uncertainties are shown in Fig. 2. This combined MH uncertainty is further propagated through the KED estimation as explained in the following section. The usefulness of the derived MH uncertainties will be evaluated in Sect. 3.1.

2.2.2 Optimizing modelled MHs

Following the method proposed by Kretschmer et al. (2013) we need to predict fields of the MH covering the full European domain at 10 km spatial and hourly temporal resolution, which are then used as input fields for the STILT model (Sect. 2.3). The geostatistical approach kriging with external drift (KED) allows us to objectively take MH uncertainty (Eq. 2) and interpolation uncertainty into account. The purpose of the external drift is to guide the interpolation with data that are more easily obtained than the actual observations and which add some physical process information to the linear prediction system. Since kriging is a well-established geostatistical approach, we summarize the main steps and the reader is referred to the standard literature (e.g. Cressie, 1993; Wackernagel, 1995). The MH $Zi^*(u, t)$ at unobserved location in space u and at time t is predicted by

$$Zi^*(u, t) = \sum_{i=1}^{n} w_i(u, t) Zi(u_i, t_i), \tag{5}$$

where the value of $Zi^*(u, t)$ is obtained from a weighted combination of n observed MHs Zi diagnosed with Eq. (1). Kriging methods like KED solve for the weights w_i such
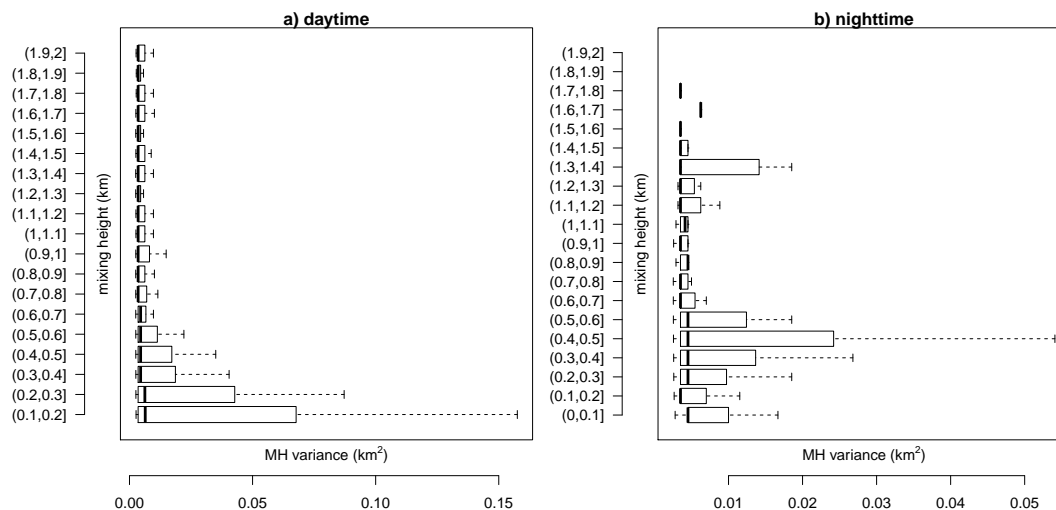
Figure 2. Uncertainty of the calculated MH with height (a.g.l.) shown in $100\,\mathrm{m}$ bins for daytime **(a)** and night-time **(b)**. The uncertainties were estimated using Eq. (2). All available IGRA observations were used for the calculations. Daytime values are shown starting from $100\,\mathrm{m}$ (a.g.l.). Boxes denote the central 50 % of the data, whiskers are at ± 1.5 times the inter-quantile range and outliers are not shown. Black line within boxes indicate the median value.

that the interpolation error is minimal (Best Linear Unbiased Estimator, BLUE). The map of MH Zi is conceptualized as random field composed of deterministic mean, referred to as trend, and spatially coloured random noise (auto-correlated). For this MH field second-order stationarity is assumed, i.e. the auto-correlated part depends on the separation distance only and is translation invariant throughout the estimation window, which is a smaller part of the domain. The auto-correlation of the residuals (signal trend) is usually modelled by one of several permissible variogram functions as explained below. The trend ($m^*(u, t)$) in KED is assumed to be a linear combination of external drift S such that it satisfies

$$m^*(u, t) = \alpha^* + \beta^* S(u, t)$$
$$E[\mathrm{Zi}(u, t)] = m^*(u, t). \tag{6}$$

The coefficients (α^*, β^*) are first solved for by ordinary least squares by evaluating the simulated MHs at observation points, and then in a second iteration predicted by KED together with the KED weights taking space–time auto-correlation of the MHs into account. Since the second iteration usually has only a minor impact on prediction skill (Hengl et al., 2007), we omit this step. Here S is obtained from MHs estimated from WRF simulated meteorology by using Eq. (1). As was shown in Kretschmer et al. (2013), KED estimates are better able to resemble a realistic fine-scale variability in MH field when using the covariate compared to ordinary kriging, which only uses a model of auto-correlation and the observed MH samples for interpolation. Kriging is mathematically equivalent to data assimilation techniques used in numerical weather prediction (Optimal Interpolation, 3DVAR; Kalnay, 2002; Wikle and Berliner,

2007), and S has a similar role as the background in these approaches. Therefore, one can interpret KED as a way to correct or optimize the background MHs in the sense of model–data fusion (Wackernagel, 1995). We prefer the term MH optimization here as it emphasizes the need for a high-quality background field, i.e. the variability in the predicted MHs can only be as realistic as provided by the meteorological model owing to the highly underdetermined problem. In this regard Hengl et al. (2007) highlight the importance of a high correlation of background and observations, also discussed in the context of MH prediction by Kretschmer et al. (2013).

The underlying auto-correlation of the MH field is approximated with a variogram model fitted to the sample variogram, which is a function of the distance in space (h_u) and time (h_t) between any pair of MH sample-trend residuals ($\mathrm{Res}_{\mathrm{Zi}}$) of the conditioning data binned in distance classes of size (N):

$$\hat{\gamma}(h_\mathrm{u}, h_\mathrm{t}) =$$
$$\frac{1}{2N(h_\mathrm{u}, h_\mathrm{t})} \sum [\mathrm{Res}_{\mathrm{Zi}}(u, t) - \mathrm{Res}_{\mathrm{Zi}}(u + h_\mathrm{u}, t + h_\mathrm{t})]^2. \tag{7}$$

Here, the assumption is that as auto-correlation decreases, the dissimilarities (semivariances) between residual pairs increase with separation distance and are bounded by a maximum value called the sill. Space and time variogram were calculated separately (by setting $h_\mathrm{t} = 0$ and $h_\mathrm{u} = 0$ for the space and time sample variograms, respectively), and then combined linearly with coefficients obtained from the sill values and a global sill to yield the product–sum variogram model, which allows for space–time interaction as described in detail in De Cesare et al. (2001). The global sill is found

by evaluating Eq. (7) beyond the distances in space and time where the respective sills were reached.

Due to the strong diurnal cycle in MH (a result from the solar insulation driven turbulent mixing) in combination with limitations in temporal resolution of the conditioning data (the MH observations), special care has to be taken to derive the residuals needed in Eq. (7). Similar to Kretschmer et al. (2013) we calculated spatial sample variograms for day and night observations separately, i.e. 12:00 UTC and 00:00 UTC. MHs derived from WRF simulations were sampled at the grid box closest to a given IGRA profile. This was repeated for each of the WRF PBL schemes used (YSU and MYJ). We estimated the deterministic component of the observations in accordance to Eq. (6) by fitting a weighted linear regression model to the observed MH as a function of the WRF MHs, taking the reciprocal of the estimated MH uncertainty from Eq. (2) as weights. The resulting regression residuals were taken to evaluate Eq. (7). Variogram models were fitted to each sample variogram shown in Fig. 3. The day variogram model was used to predict hours 9 to 16, which was found suitable for this domain by Kretschmer et al. (2013). Since the 12 h resolution of the IGRA data is too coarse to constrain the variogram model sufficiently, we make use of the hourly MHs from the WRF simulations, assuming that the resulting semivariance closely resembles the temporal auto-correlation properties of the observed signal adequately. This assumption is reasonable since we have chosen PBL schemes that were shown to realistically simulate PBL dynamics (Hu et al., 2010). We sampled WRF MHs again separately for day and night. Because we have used WRF MHs as the MH data points, i.e. synthetic data, instead of real IGRA data we cannot use the same WRF MHs as trend to calculate the residual terms (Res_{Z_i}) in Eq. (7). Instead we follow the usual procedure to model the diurnal pattern in the WRF MHs as deterministic trend using an oscillating sinusoidal function. The residuals between this model and the WRF MH were then used to compute Eq. (7) for the time domain. MH uncertainty obtained from Eq. (2) is considered in the MH prediction by adding this uncertainty to the diagonal elements of the covariance matrix used in the KED system of linear equations as suggested by Wackernagel (1995). This covariance matrix contains covariances between any pair of MH observations that were computed by subtracting the semivariances from the sill values obtained from the variogram model fit.

The described computations were executed using a modified version of the Edinburgh Space Time statistics (Spadavecchia, 2009) and the geoR package for the R programming language (Diggle and Ribeiro, 2007; Ribeiro and Diggle, 2001).

2.3 Transport modelling

Our transport modelling system consists of the Eulerian WRF model, which provides hourly meteorological driver fields on a 10 km grid for the STILT model. STILT is a receptor oriented Lagrangian particle dispersion model introduced by Lin et al. (2003). Since we basically use the same modelling system as in Kretschmer et al. (2013), we give only a brief summary here. The STILT model calculates for a given grid box i, j of the domain, the rate of change in tracer concentration at the receptor r, e.g. a tall tower observation site, over time step m from a footprint function f and the surface fluxes F (Gerbig et al., 2003b; Lin et al., 2003):

$$\Delta C_{m,i,j}(x_r, t_r) = f(x_r, t_r | x_i, y_i, t_m) F(x_i, y_i, t_m). \quad (8)$$

The footprint element $f(...)$ relates fluxes at a specific location and time to changes in the mixing ratio along the particle trajectory. The flux function $F(...)$ represents any combination of offline flux input and online calculated fluxes, which are described in Sect. 2.4. The footprint is calculated by releasing ensembles of 100 particles at the receptor and, for each particle, advection with the mean wind provided by WRF is computed. As particles move further away from the receptor the grid is aggregated to a coarser resolution to account for effect of undersampling caused by the relatively small ensemble size. Turbulent diffusion is implemented as a stochastic process which is added to the mean particle trajectory. Mass fluxes related to moist convection (updraft, downdraft, and entrainment fluxes) are provided by WRF through the Grell–Dévényi scheme (Grell and Dévényi, 2002), and are applied in STILT in a stochastic way (Nehrkorn et al., 2010). The WRF setup is summarized in Table 1.

STILT determines the mixing height offline from the meteorological driver fields using a Ri method, or MHs are prescribed externally (Lin et al., 2003). For the control simulations we determine the MH from WRF output using the Ri method (Eq. 1). To reduce the impact from MH errors on tracer transport we use the KED optimized MHs (see Sect. 2.2.2) as external input for STILT.

The results of Kretschmer et al. (2013) indicate that the dominant effect of MH errors on the transport simulation is the turbulent diffusion of tracer particles up to a wrong altitude, suggesting that potential physical inconsistencies and side effects affect the tracer concentrations in the mixing layer to a minor extent. Such physical inconsistencies involve other meteorological input variables used for the turbulence calculations. The profiles of vertical velocity variance σ_w, which determines the amount of random deviation from the mean vertical wind for a given particle, and the Lagrangian timescale T_L, which describes the decorrelation in the particle movement (Lin et al., 2003), depend not only on the MH, but also on roughness length, Monin–Obukhov length, convective velocity scale, and frictional velocity, following Hanna (1982). However, only the mixing height determines the altitude at which strong turbulent mixing changes from high values within the mixing layer to lower values for the free troposphere. Here our assumption is that the potential impact of the other meteorological input variables on resulting tracer profiles is small. To support this assumption, we
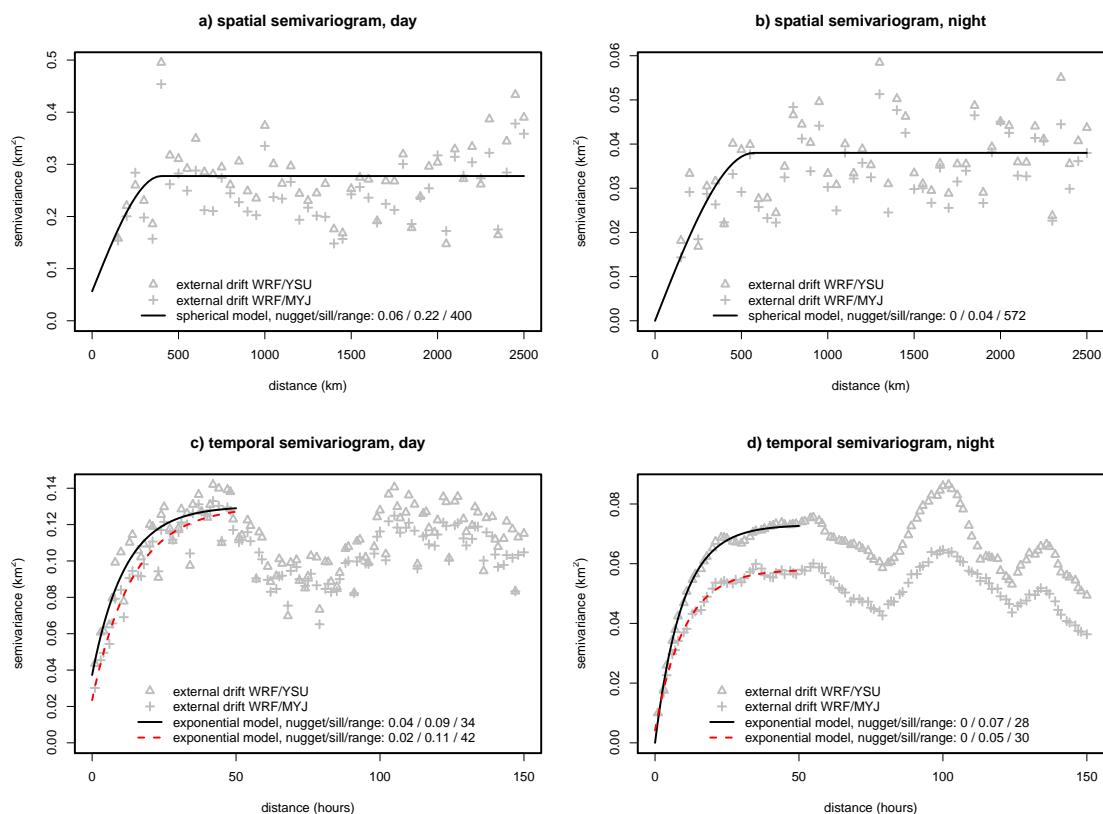
Figure 3. Spatial **(a, b)** and temporal **(c, d)** sample variograms as calculated with Eq. (7) for day and night separately. Spatial sample variograms were calculated from model/observation linear regression residuals for each WRF setup using MYJ (grey pluses) and YSU (grey triangles) PBL scheme. The time variograms were calculated from WRF MHs subtracting a diurnal trend **(c, d)**, notable differences between PBL setups were found; thus we fitted separate variogram models for each PBL scheme **(c, d)**. KED was set up to use 25 neighbours in space and 3 in time, i.e. 75 observations were used to predict each of the 141.12×10^6 grid cells of one WRF simulation. The maximal distance between any pair of observations were typically below 2500 km in space and 48 h in time because of the 12 h temporal resolution of the conditioning data. The variogram models were fitted to cover these ranges. A temporal trend is observed at synoptic scales beyond 50 h, which does not affect the fitted variogram models **(c, d)**.

Table 1. Setup of WRF options.

| Option	Setting
Model code version	3.0.1.1
Time step integration	1 min, third-order Runge–Kutta, output interval 1 h
Grid definition	280×400 (north–south × west–east), 10 km spacing, Arakawa C
Vertical coordinates	41 levels (20 below 2 km), terrain following, eta coordinates, pressure top 50 hPa
Basic equations	Non-hydrostatic, compressible
Microphysics	WRF single moment class 5
Atmospheric radiation	Rapid Radiative Transfer Model (RRTM, long wave), Mesoscale Model 5 (MM5, Dudhia, short wave)
Cumulus parameterization	Grell–Dévéni
Land-surface model (LSM)	Noah LSM, 4 soil layers
PBL scheme	Yonsei University (YSU setting), Mellor–Yamada–Janjic (MYJ setting)
Surface layer scheme	Monin–Obukhov similarity (YSU setting), Monin–Obukhov (Janjic Eta, MYJ setting)
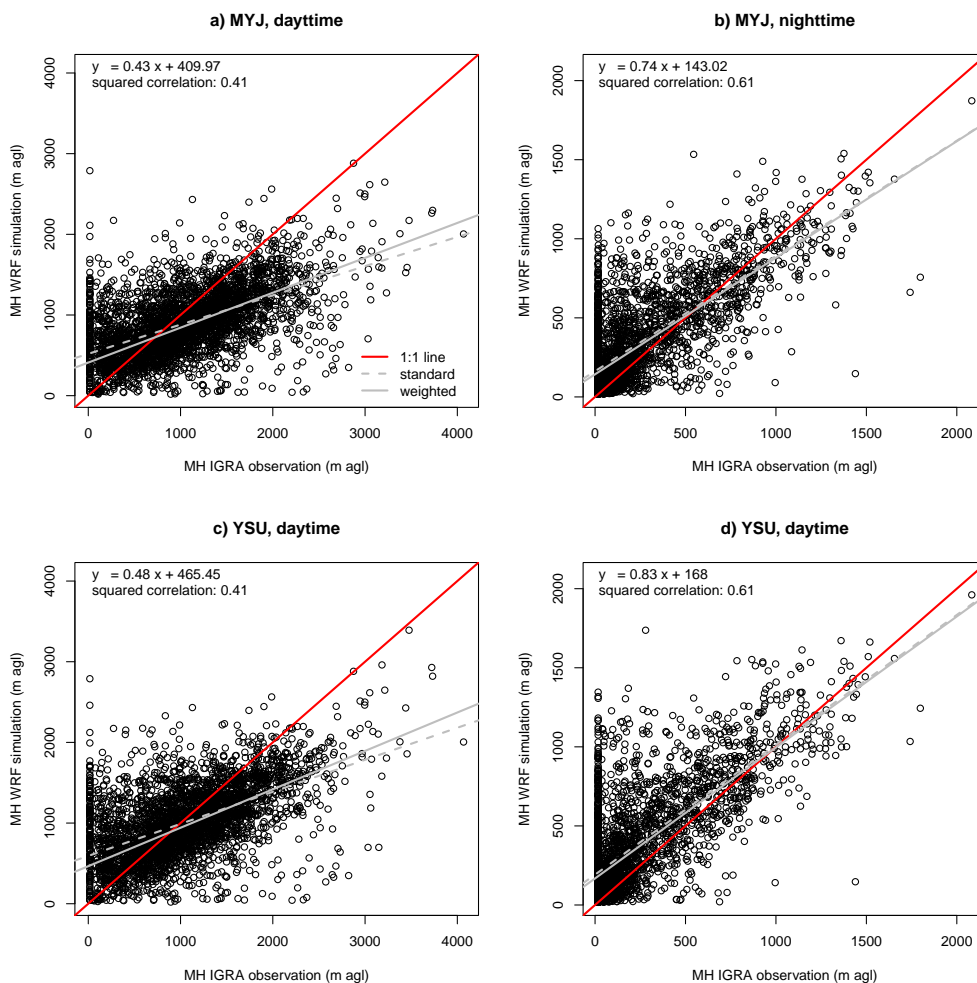
Figure 4. Comparison of WRF- and IGRA-derived MHs using the *Ri* method. We used about 6700 samples almost evenly distributed among day and night (00:00 and 12:00 UTC). The left column **(a, c)** shows 12:00 UTC and the right column **(b, d)** 00:00 UTC data. The function of a weighted linear regression model together with the resulting r^2 is shown in the upper left corner of each plot (dashed grey line). The weights were set to the reciprocal of the MH estimation uncertainty. The regression function neglecting this uncertainty is indicated as a solid grey line. The squared correlation coefficients for the latter are shown in Table 2.

would need to assimilate these additional variables, but in comparison to MHs these are harder to measure. Another side effect is the de- or entrainment of air particles to or from the layer above the mixing layer when the MH is changed, in combination with wind shear typically present in that region, which could lead to an alteration of the particles' trajectories within the mixing layer, resulting in modified surface influence. Lowering the MH leads to a more local flux influence since horizontal wind speed decreases with decreasing distance to the surface; thus we can expect a minor alteration of mixing ratios. More sensitivity has to be expected in the case of strong wind shear near the MH together with an increase in the MH. This is likely to happen during night-time when low-level jets can develop near the MH. For CO_2 we can expect minor negative impact on mixing ratios, because the dominating respiration flux is spatially rather homogeneous.

During daytime the NEE is less homogeneous, but the mixing layer is generally deeper and thus the impact of a slight increase of the MH is expected to cause only small alterations of the mean horizontal trajectories. These assumptions are supported by the results of Kretschmer et al. (2013).

In the present study transport was simulated hourly for Cabauw and Heidelberg in the period of 1 September to 9 October 2009, and for receptors located along the IMECC flight track (Fig. 1). The STILT domain was set up to cover most of Europe on a Cartesian grid at $1/12° × 1/8°$ ($\sim$10 km × 10 km) as in Gerbig et al. (2008) with 41 vertical levels similar to our WRF setup.

2.4 Boundary conditions and input fluxes

For the transport simulation of any tracer, lateral boundary conditions and surface fluxes need to be prescribed. These

Table 2. Results of the comparison of WRF and IGRA RS-derived MHs based on Eq. (1) (upper panel) and cross-validation results (lower panel) for day (12:00 UTC) and night (00:00 UTC). Results are shown for both PBL parameterizations: MYJ and YSU. The values in the tables denote the results without and with considering the MH uncertainty shown in the format: without/with MH uncertainty considered. The MH uncertainty was derived as described in Sect. 2.2.1. Bias and random error were normalized by the mean observed IGRA derived MH (columns % Bias, % RMSE). When taking MH uncertainty into account the normalization was done with a weighted mean. Column % SD refers to the number of observations within one standard deviation of the KED estimate and % CI to the observations within the 95.42 % confidence interval.

Comparison of WRF MH to IGRA derived MH								
Simulation	Time	Bias	% Bias	RMSE	% RMSE	r^2	% in SD	% in CI
MYJ	day	−144/−219	−14/−20	543/519	52/42	0.32/0.41	NA	NA
YSU	day	−36/−116	3/10	535/487	51/42	0.30/0.41	NA	NA
MYJ	night	108/89	54/40	235/222	116/92	0.56/0.61	NA	NA
YSU	night	156/135	77/61	277/259	137/116	0.54/0.61	NA	NA
Cross-validation results of KED derived MH								
Simulation	Time	Bias	% Bias	RMSE	% RMSE	r^2	% in SD	% in CI
MYJ	day	5/35	0/3	521/458	50/40	0.31/0.42	69/69	92/91
YSU	day	7/42	1/4	518/447	50/39	0.32/0.45	73/74	95/94
MYJ	night	4/13	2/5	195/196	96/85	0.57/0.62	78/78	93/93
YSU	night	5/15	2/6	192/195	94/84	0.59/0.63	81/80	95/95

fields were re-projected and aggregated to the STILT grid taking mass conservation into account. In the following we give a summary of the input data we used for individual tracers.

2.4.1 CO$_2$

The CO$_2$ boundary conditions were taken from 6-hourly analysed fields of 2009 from the Jena Inversion[2] version 3.3 on a 4° × 5° grid with 19 vertical levels (Rödenbeck, 2005).

To prescribe combustion fluxes we make use of the 2005 Emission Database for Global Atmospheric Research (EDGAR) on a 0.1° × 0.1° grid to consider anthropogenic flux contributions (source: EC-JRC/PBL; EDGAR version 4.1. http://edgar.jrc.ec.europa.eu/, 2010). Similar to Steinbach (2010) we extrapolated country total emissions to the year 2009. The extrapolation is based on BP statistics obtained from http://www.bp.com/statisticalreview. These emissions were then spatially distributed down to the grid level based on the 2005 data set. Time factors were obtained from the EDGAR database and then applied to yearly fluxes to resolve the daily cycle. The time factors are based on the step-function time profiles published on the EDGAR website (http://themasites.pbl.nl/tridion/en/themasites/edgar). These were modified before they were applied to yearly fluxes in order to resolve the daily cycle. The modification of the temporal factors involves a better global representation and a smoothing of the monthly transitions (see Steinbach, 2010 for further details).

Contributions from oceanic fluxes were accounted for by including the Takahashi et al. (2009) climatological inventory for the reference year 2000 and revised in October 2009 provided monthly with a spatial resolution of 4° × 5°.

Vegetation fluxes of the net ecosystem exchange (NEE) were calculated within STILT based on the Vegetation Photosynthesis and Respiration Model (VPRM; Mahadevan et al., 2008). VPRM is a diagnostic model that uses as input shortwave radiation and 2 m temperature, both calculated within WRF (variables SWDOWN and T2), and two vegetation indices. These indices, the enhanced vegetation index (EVI) and the land surface water index (LSWI) are obtained from 500 m, 8-daily MODIS (Moderate Resolution Imaging Spectroradiometer) satellite surface reflectance data (http://modis.gsfc.nasa.gov). VRPM indices are scaled with parameters optimized against eddy covariance flux measurements for Europe (Pillai et al., 2011) to derive respiration and Gross Ecosystem Exchange (GEE) fluxes separately (Mahadevan et al., 2008). There is one parameter set for each of the eight vegetation classes used. Fractional vegetation coverage for each model grid cell was derived from SYNMAP (Jung et al., 2006) with a horizontal resolution of ∼1 km^2. VPRM fluxes were calculated once from SWDOWN and T2 produced by the WRF-YSU simulation and then used for all other STILT simulations, which effectively resembles an offline flux model similar to the other offline fluxes (EDGAR, ocean fluxes). This approach ensures that all transport simulations use consistent biospheric fluxes and thus facilitates the interpretation of the results, avoiding the impact from e.g. changes in temperature and cloud cover due to MH alterations on the fluxes. In contrast to CO, emissions due to

---

[2]Available at www.bgc-jena.mpg.de/~christian.roedenbeck/download-CO2-3D/

fire were not considered explicitly, because they are known to have only negligible impact on CO_2 concentrations.

2.4.2 CO

CO is used as auxiliary tracer to isolate the impact from combustion fluxes on CO_2 signals. Initial and boundary conditions for the CO transport were obtained from re-analysed 4-D fields provided as part of the Monitoring Atmospheric Composition and Climate (MACC) project (source: http://data-portal.ecmwf.int/data/d/macc_reanalysis/). The MACC re-analysis is provided on a 6-hourly $1.125° \times 1.125°$ grid with 60 vertical levels. Similar to the CO_2 tracer we obtain anthropogenic emission from EDGAR (Sect. 2.4.1). The impact of fire emissions are considered by including flux fields from the Global Fire Emissions Database (GFED; source: http://www.globalfiredata.org) in version 3.1. We used the 3-hourly fields on a $0.5° \times 0.5°$ grid (Mu et al., 2011). The major sink for CO is atmospheric destruction by hydroxyl radical OH which is computed within STILT. Similar to Gerbig et al. (2003b) we estimate the OH on a given particle location based on a climatological OH field. The soil uptake of CO is an order of magnitude smaller than the OH reaction and is therefore neglected.

2.5 Statistical measures

To summarize the performance of the transport simulation and MH optimization we report some commonly used statistical measures. Bias b is computed as the mean difference between pairs of estimated and observed quantity, such that the bias is negative when the estimations are on average smaller than the observation and positive if the estimates were greater, respectively. The random error denotes the standard deviation s of these differences. In addition, we report the root mean squared error as a function of bias and random error: $RMSE = (b^2 + s^2)^{1/2}$. To test the statistical significance of the bias we performed one-sample, two-tailed t tests. With the null hypothesis that the sample was unbiased ($b = 0$) with a significance level of 0.05. Weighted statistics were calculated setting weights to the reciprocal of the estimated MH uncertainty variance (see Sect. 2.2.2).

3 Results

3.1 Evaluation of mixing heights

3.1.1 Comparison of WRF MHs to IGRA RS

Figure 4 shows the comparison of MH estimated from IGRA profiles for day and night-time and both WRF simulations. Daytime WRF MHs are in general lower than observed. Most night-time MHs are below 500 m although there are quite a number of MH above that threshold, which is reasonably captured by both WRF simulations. However, all of the

Table 3. Results of STILT–data CO_2 comparison at Cabauw (upper half) and Heidelberg (lower half), considering STILT/MYJ and STILT/YSU simulations (rows: MYJ, YSU). Also shown are the statistics of the STILT simulations using KED-optimized MHs (rows: MYJ KED, YSU KED). Bias and RMSE were normalized with the observed CO_2 subtracting STILT simulated background concentrations (columns % Bias and % RMSE). Day denotes hours 10:00–12:00 UTC and night 20:00–04:00 UTC.

CO_2 Cabauw, 20 m						
Simulation	Time	Bias	% Bias	RMSE	% RMSE	r^2
MYJ	day	1.45	9	4.95	31	0.64
MYJ KED	day	0.84	5	4.60	29	0.69
YSU	day	1.17	7	4.81	30	0.66
YSU KED	day	0.83	5	4.58	28	0.69
MYJ	night	−8.96	−56	16.34	101	0.53
MYJ KED	night	−2.95	−18	17.45	108	0.59
YSU	night	−9.96	−62	17.77	110	0.42
YSU KED	night	−1.02	−6	22.4	139	0.42

CO_2 Heidelberg, 30 m						
Simulation	Time	Bias	% Bias	RMSE	% RMSE	r^2
MYJ	day	−1.89	−10	4.38	23	0.75
MYJ KED	day	−1.87	−10	4.57	24	0.70
YSU	day	−1.68	−9	5.08	27	0.6
YSU KED	day	−1.31	−7	4.57	24	0.67
MYJ	night	−6.91	−36	11.84	62	0.58
MYJ KED	night	−2.54	−13	11.67	61	0.51
YSU	night	−7.04	−37	12.32	65	0.53
YSU KED	night	−1.53	−8	11.12	58	0.53

plots exhibit large scatter, especially during daytime. Correspondingly, the explained variability is rather low at day with ∼30 % for both PBL schemes compared to night-time with over 50 %. Unlike the simulations, IGRA-diagnosed MH seems to detect surface-based inversion layers fairly often, which leads to MH detection at the first upper air level (17 m). Taking MH uncertainty (Eq. 2) as weights for the linear regression into account seems to downweight some of these rather low IGRA MHs, which is likely caused by the poor vertical resolution of IGRA RS that affects especially the night-time (Seidel et al., 2012). The weighting results in improved correlation coefficients by ∼10 %.

Table 2 summarizes some statistics computed from the comparison taking the MH uncertainty into account. In general all simulations exhibit significant bias and substantial random error especially at night. As could be expected YSU produces the best correspondence to IGRA MHs at day, with bias ∼10 %, albeit random errors of 40 % occur. MYJ exhibits greater daytime bias of ∼20 %, which is expected to bias the simulation of vertical tracer diffusion notably. At night YSU has a large bias of 60 % while MYJ exhibits a bias of only 40 %. Random errors for both PBL schemes approach 100 % at night. These numbers confirm the large model uncertainty in MH during night-time; thus we expect a corresponding transport model error, which will lead to

too much or too little accumulation of tracer mass in the SBL, respectively. Thus, the use of optimized MH should have most potential for improvement in stable/wind-shear-driven conditions. The observed mismatch in MHs are comparable to the findings of Gerbig et al. (2008). They compared radiosonde based MHs to ECMWF-analysed meteorology ($\sim 35\,km^2$ resolution).

3.1.2 Evaluation of MH optimization

We assessed the skill of the KED optimization by cross-validation, such that each of the 6722 IGRA samples was temporarily excluded from the data set and then estimated with the remaining data. Table 2 shows the results of the cross-validation, which was executed with and without taking MH uncertainty into account to test the validity of the assumptions stated in Sect. 2.2.1. A small but significant bias on the order of a few percent remains for some simulations, which decreases when taking MH uncertainty into account. Compared to the random errors of the unoptimized WRF MH (see Table 2, upper panel) the random errors of the optimized WRF MH are slightly smaller during the day (see Table 2 lower panel), but become notably smaller when considering MH uncertainty, too. The correlation values stay at the level before the optimization with rather low values for the case neglecting MH uncertainty. The KED errors were reasonably estimated – from a normal distribution we would expect 68 % of the observed MH to lie within one standard deviation and 95 % within the confidence interval of the KED estimate. This result suggests the possibility of propagating these uncertainties through a CO_2 inversion as discussed in Sect. 4.3. Figure 5 shows an example highlighting the differences in innovation when MH uncertainty is taken into account, which also leads to alterations in the spatial distribution of KED errors.

We compared estimated MHs to 73 UKMO RS profiles (32 day, 41 night) from the stations Castor Bay and Albemarle (shown in Fig. 1). Castor Bay was typically 380 km and Albemarle 230 km away from the next IGRA station. The results were comparable to the cross-validation (not shown). In general, daytime MHs are less affected by the MH optimization while large reduction in bias and random errors occurred at night, together with substantial increases in correlation. In contrast to the cross-validation, we observed KED variances that are usually too conservative, with > 80 % in one standard deviation of KED error, which might be due to the small sample sizes.

3.2 Comparison to IMECC campaign data

3.2.1 Effective CO_2 mixing height

In order to test whether the KED MHs are consistent with the actual vertical mixing of CO_2, we estimated an effective MH from IMECC vertical profiles. We define MHs to
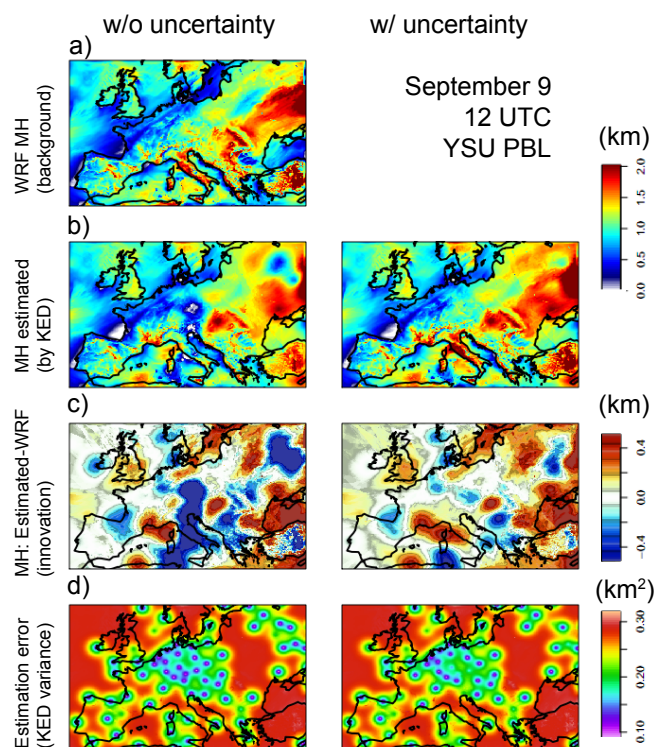

Figure 5. Example of the KED estimation. Shown are the MHs produced by the WRF YSU simulation (**a**) for each $10 \times 10\,km^2$ pixel of the simulation domain on the 9 September 2009, 12:00 UTC used as external drift (background field) in the KED estimation. The maps in (**b, c, d**) (left column) show estimation results neglecting uncertainty of observed MH in the KED estimation and the right column the results when accounting for this uncertainty. The KED optimized MHs are shown in row (**b**). The innovation for each pixel is computed as the difference of optimized MH and background field (row **c**). Also shown in row (**d**) is the resulting KED error variance. Less error variance is observed near the locations of radiosondes.

be consistent if they are within one standard deviation of the KED error. However, there is no well-established method known to us to objectively diagnose effective MHs from CO_2 profiles. Here we followed a visual approach. First, we selected profiles which have a good vertical CO_2 data coverage in the lower parts of the PBL. We omitted profiles where data in the PBL are sparse or missing, such that in the remaining profiles gradients from vertical mixing are visible. The analysis of profiles obtained during stable conditions, i.e. mainly night-time and early morning, was hampered by poor data coverage and by the fact that the relationship between the CO_2 profile and the Richardson number is not well understood. During these stable conditions we often observed the absence of mixing, caused by strong temperature inversions and low friction velocities. The definition of a MH is difficult in such situations (Seibert et al., 2000) and CO_2 concentrations show large gradients in the lowest $\sim 500\,m$. Therefore we decided to analyse profiles that have been taken likely in well-mixed conditions. The profiles that were used for the

Table 4. Comparison of CO_2 and MH model to model differences shown for Cabauw. The differences were calculated by subtracting the simulations using YSU PBL scheme from the ones using MYJ PBL scheme for each experiment, i.e. with and without MH optimization respectively. Day denotes hours 10:00–12:00 UTC and night 20:00–04:00 UTC.

MH Optimization	Time	CO_2 Bias (ppm)	CO_2 RMSE (ppm)	CO_2 r^2	MH Bias (m)	MH RMSE (m)	MH r^2
off	day	0.27	2.42	0.86	−116.56	189.96	0.74
on	day	−0.06	2.18	0.86	5.45	122.09	0.84
off	night	2.16	17.77	0.5	−90.79	178.94	0.85
on	night	−0.92	26.29	0.46	−14.04	105.14	0.88

analysis are shown in Fig. 6. During well-mixed conditions the profiles of CO_2 concentrations are constant with height within the mixed layer, and exhibit a gradual change to free tropospheric values above. Thus, we plotted the CO_2 gradients as a function of altitude and then selected the height above the surface layer where the gradient from ML to free troposphere was observed as the effective MH. This visual approach requires that turbulence in the entrainment zone causes well distinguishable gradients in CO_2 that can be detected. Thus a larger uncertainty is expected for cases when mixing layer and free troposphere (or residual layer) have nearly the same concentration. As indicated in Fig. 6, optimized MHs are in most cases consistent with the effective MH. This result holds for both PBL schemes. WRF generally reproduces the observed potential temperature profiles well and accordingly the optimized MH has a minor impact at day, except one instance at Hohn, 2 October 2009, 12:17 UTC when the WRF MH is 200–300 m too high. This error is significantly reduced in the optimized MHs. Another instance at Hohn, 9 October 2009, 11:34 UTC shows a well-mixed layer of CO_2 up to only ∼600 m, while the simulations show much larger mixing heights. Here MYJ had a better agreement with the observed CO_2 profile, but the MH optimization caused an increase of the MH. Such large deviations are an indication of low correlation between WRF background and conditioning data from IGRA RS. Here the effective MH is not within one standard deviation of the KED error, but well within the confidence interval.

3.2.2 Simulated CO_2 profiles

Here we focus on the STILT simulated CO_2 profiles during the IMECC campaign (Fig. 6). During well-mixed conditions both STILT simulations are able to reproduce the observed CO_2 profile reasonably, but tend to be ∼1 ppm too high in the mixing layer. Because differences in initial and optimized MHs were small, the impact on CO_2 simulations was limited, but usually the mean CO_2 concentration in the mixing layer is corrected towards the observations. A notable example is the profile at Hohn, 2 October 2009, where the CO_2 concentration was adjusted by about 0.5 ppm, but a 1 ppm model–data mismatch in CO_2 remains. The profile near Gdańsk, 30 September 2009, 10:37 UTC shows a large model mismatch in CO_2 concentrations throughout

the mixed and residual layer. The effect of optimization in the residual layer was in general negligible.

3.3 Comparison to ground measurements

While the IMECC profiles give a good vertical coverage, they provide only snapshots. Thus the continuous observations from Cabauw and Heidelberg are useful to better understand the impact of MHs on CO_2 time series as shown in Fig. 7. The average diurnal cycle of KED estimated MH are shown in Fig. 8. Table 3 gives an overview of the statistics.

3.3.1 Cabauw

Figure 7a and b show the complete CBW time series at the 20 m level, which is mostly affected by the MH during day and night. The CO_2 signal at CBW shows a pronounced amplitude of 30 ppm on average, caused by the accumulation of respiration and nearby anthropogenic sources during night. The diurnal and synoptic variability in the time series is well represented in all simulations with r^2 typically between 0.5 and 0.7. These values generally improve by ∼3–6 % when using optimized MHs (see Table 3). Large negative biases were produced at night of −9 to −10 ppm, when the optimization of MHs is most effective reducing biases from −56 % to −18 % (STILT/MYJ) and from −62 % to −6 % (STILT/YSU; see Table 3), which was expected from the MH cross-validation results. There are positive biases at day of 1.17 ppm (STILT/YSU) and 1.45 ppm (STILT/MYJ) about 9 and 7 % of the regional signal, i.e. here the bias reduction due to MH optimization is at 29–44 %. The overall size of the errors is comparable to previous simulation results (e.g. Broquet et al., 2011). Note that we used prior fluxes for the regional simulations, which were not optimized against observations of CO_2 through inversions. The random errors increase during events when using optimized MH fields, e.g. 7 to 9 September or 25 to 28 September (Fig. 7a and b). The events are characterized by especially large random errors and model–model differences. The error is most obvious in the STILT/YSU simulation, which is further discussed in Sect. 4. At the highest inlet (200 m) the signal amplitude is on average a factor 6 smaller (not shown), and model bias and random errors were below 5 ppm respectively with $r^2 \sim 0.5$. At this altitude the overall bias of the STILT/MYJ reduced
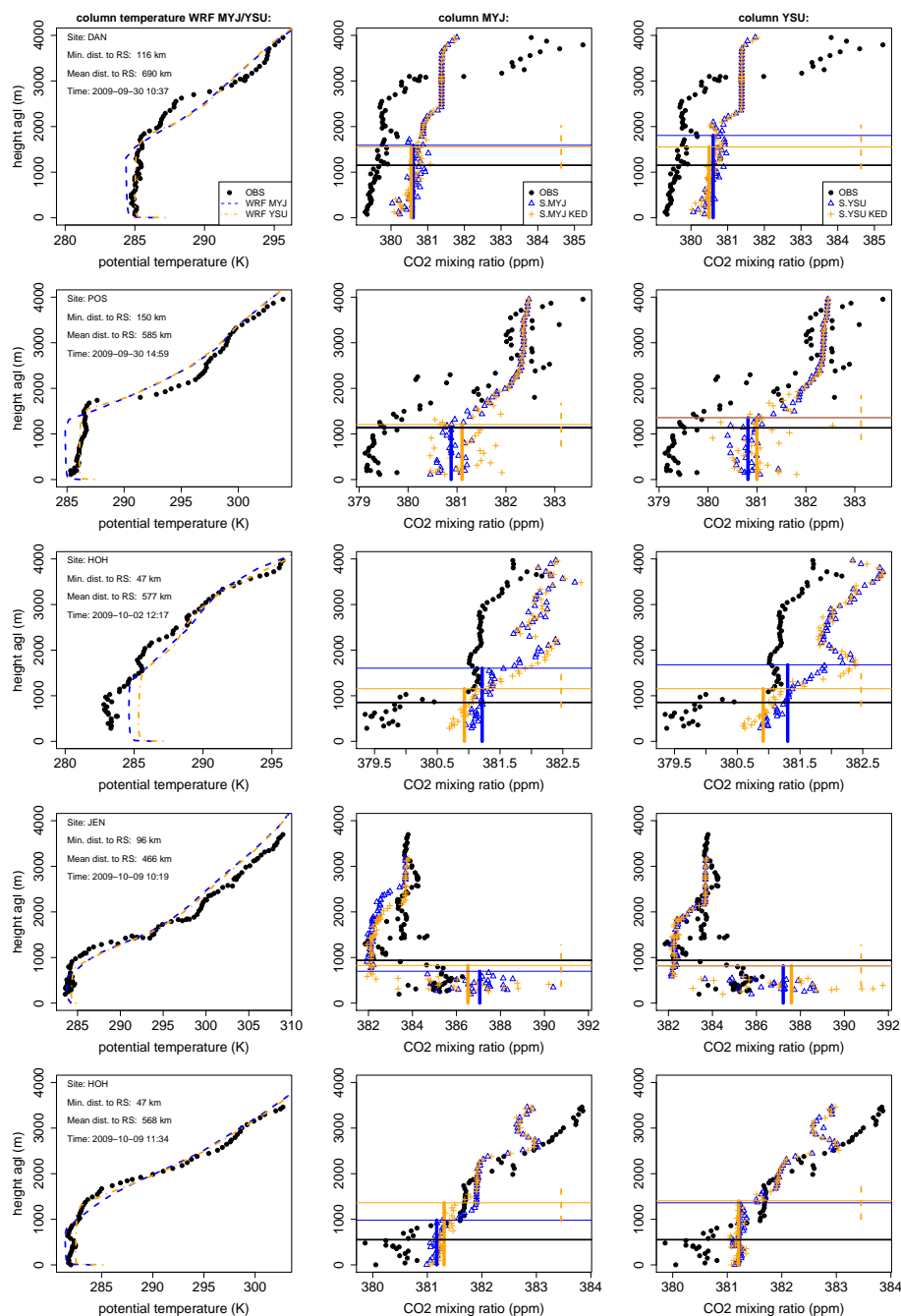
Figure 6. Measured and simulated IMECC profiles measured in daytime with enhanced mixing. The left column shows observed (black dots) and WRF simulated potential temperature. Shown are both WRF simulations using the YSU PBL (orange dash-dotted lines) and the MYJ PBL (blue dashed lines) parameterization schemes. The statistics in the temperature plots indicate the location and time of the profile. Site abbreviations correspond to the red squares on the map in Fig. 1. Also shown are the minimum and average distances to the next IGRA radiosonde used for optimization. The comparisons of STILT simulated and observed CO_2 profiles are shown in the middle and right columns. A new STILT receptor was defined every 10 km horizontal or every 100 m altitude change of the airplane. Horizontal lines indicate MHs. Observed MHs were derived by analysing gradients in the CO_2 profile (see text for details, black lines). STILT MHs are averages from all receptors of a given profile. STILT CO_2 profiles are shown with (orange crosses) and without optimized MHs (blue triangles). One standard deviation of the KED error is shown as an orange dotted bar in the right of each plot. Thick vertical lines from the surface (0 m a.g.l.) to each MH correspond to mixing layer averaged CO_2 (blue and orange lines).
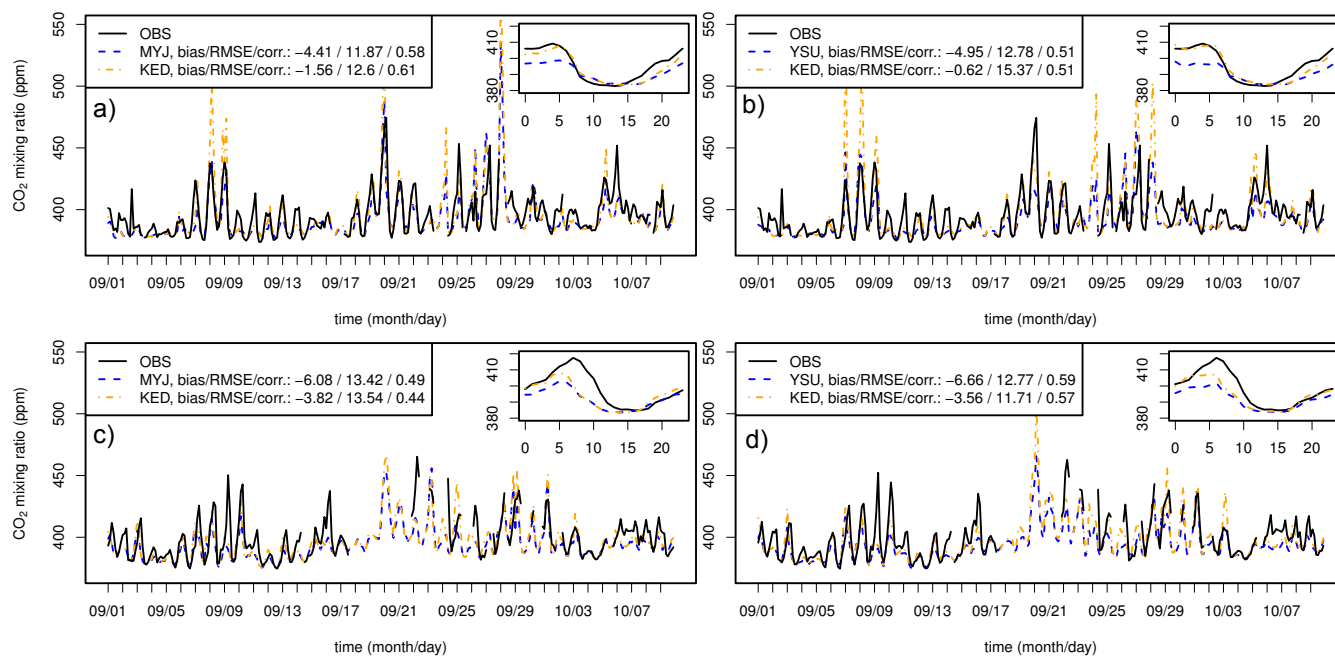
Figure 7. CO_2 time series observed at Cabauw 20 m a.g.l. (**a, b**) and Heidelberg 30 m a.g.l. (**c, d**, black lines). STILT simulations are shown with (orange dash dotted lines) and without optimized MHs (blue dashed lines). Inlets in the upper right corner of each plot show the averaged diurnal cycle. In the upper left corner of each plot are summary statistics calculated for the full times series of the CO_2 mismatch: STILT observations together with STILT/observation squared correlation coefficient (r^2). Data shown are 3 h averages. Dates on the *x* axis refer to the analysis year 2009.

by 1 ppm due to optimized MHs, while STILT/YSU was not affected. Night-time improvement,.i.e. reduction in bias and random error, is less notable than at the 20 m inlet, because the receptor is located below the MH. In general, CO_2 biases between PBL schemes were in line with our MH comparison (see Sect. 3.1.1 and Table 2), i.e. better results of STILT/YSU at day and worse at night than STILT/MYJ (Table 3). The general reduction in model–data mismatch indicates the effectiveness of the MH optimization. Remaining CO_2 mismatches are of comparable size for both PBL schemes; this is to be expected when estimated MHs converge (Fig. 8).

3.3.2 Heidelberg

Similar to CBW, the HEI CO_2 time series shows a pronounced daily amplitude at the ca. 30 m inlet height of about 30 ppm on average (Fig. 7c and d). Correlation between CO_2 observations and all simulations were generally good $r^2 > 0.5$. All simulations produced too low concentration at day and especially night with biases of ~6 ppm. The phasing of the mean diurnal cycle exhibits notable mismatch in all simulations, the MH optimization cannot resolve this issue, probably due to the low temporal resolution of the conditioning data (Sect. 2.2). Nevertheless, the usage of optimized MHs reduces the overall bias by 40–50 % (Table 3 and Fig. 7c and d). We observed again a most notable reduction of night-time bias, especially in the early morning. The

effect of MH optimization on correlations and random errors was mixed (Table 3). Similar to CBW slight deterioration is observed at the event with increased errors and model–model differences centred around 10 October 2009 (Fig. 7c and d), albeit less obvious.

4 Discussion

The KED estimation was most effective to remove biases from WRF MHs. In addition, KED MHs were found to be mostly consistent with actual CO_2 mixing. This is also supported by the generally favourable effect of optimized MHs on CO_2 transport, as averaged diurnal cycles were notably better captured. However, substantial deviations from observed CO_2 signals remained. Model–data mismatch in CO_2 concentrations can be caused by other factors that need to be considered, namely (1) errors in CO_2 fluxes and (2) remaining transport-related model errors, which includes horizontal advection and imperfect MH optimization. Factor (1) is likely since we used non-optimized (a priori) fluxes; thus a perfect match to observed CO_2 cannot be expected. This could to some extent even lead to a shift in phasing such as observed at HEI. Factor (2) is also possible, as we have shown that significant bias and errors remain in MHs (Table 2). In addition, transport model errors could be due to physical inconsistencies of the presented method for MH
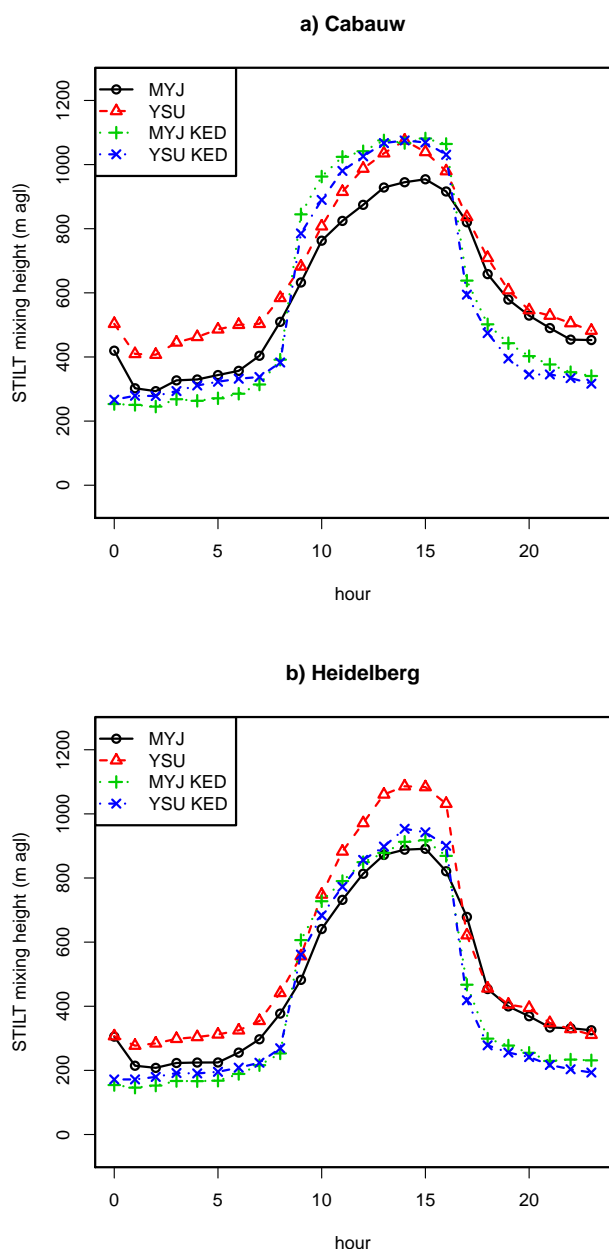
Figure 8. Comparison of averaged diurnal cycles of WRF derived (black dots, red triangles) and optimized MHs (green pluses, blue crosses) at Cabauw **(a)** and Heidelberg **(b)**. The KED optimization was done with the MYJ and YSU PBL scheme and using the MH observations. Average difference of MHs between the PBL schemes is minimized by the optimization as both converge to the conditioning data. Note that there is no MH observation co-located with CBW or HEI.

optimization (see Sect. 2.3). An improvement would be an assimilation of the MH directly in the meteorological model, which could be achieved by more classical assimilation techniques (e.g. 4D-VAR) and the introduction of a new observational operator that relates prognostic variables (e.g.

temperature) with MH observations, for instance one could use a Ri method such as presented in Eq. (1).

While Kretschmer et al. (2013) showed that an increased number of MH data to condition the KED optimization will eventually lead to further improvement, the observed deterioration of night-time RMSE especially at Cabauw can be caused by any of the mentioned factors. The following discussion focuses on CBW, because prominent RMSE features were observed there and additional meteorological observations were available, which were needed for the purpose of analysis.

4.1 Uncertainty in fluxes

Errors in CO_2 fluxes are an obvious reason for CO_2 model–data mismatch. Thus we need to evaluate the ability of our modelling system to adequately reproduce biospheric and anthropogenic flux components at CBW.

4.1.1 Uncertainty of VPRM parameters

Ahmadov et al. (2007) have shown the general ability of VPRM to calculate realistic fluxes in the European domain. However, we evaluated radiation and surface temperature, the two main WRF parameters that drive VPRM, to hourly measurements at CBW. Radiation was found to deviate by 50 % for some days, but was unbiased. Surface temperature was simulated well, but revealed a $1°$ K bias in the evening with higher temperatures in the YSU PBL, likely caused by a too deep mixing layer that does not cool down as fast as a shallow mixing layer. Considering characteristic VPRM parameters relevant for the site $\alpha = 0.088\,\mu\text{mole}\,\text{m}^{-2}\,\text{s}^{-1}\text{K}^{-1}$ and $\beta = 0.58\,\mu\text{mole}\,\text{m}^{-2}\,\text{s}^{-1}$, a change in temperature from 13 to 14 °C increases the respiration only slightly from 1.72 to 1.81 $\mu\text{mole}\,\text{m}^{-2}\,\text{s}^{-1}$. This effect seems negligible compared to the large overestimation of nocturnal CO_2 build-up. Other sources of uncertainty are the scaling of VPRM parameters from point to region and prior uncertainty, e.g. from VPRM in comparison to eddy flux residuals and the uncertainty of fitting VPRM parameters to eddy covariance data. The significant temporal variation of the peat-on-clay and clay-on-peat emissions depending on the soil moisture content are not captured by the alpha and beta parameters, but these emissions could have a notable impact on CO_2 emissions at CBW (Vermeulen et al., 2011). A detailed analysis of these factors is not trivial and is beyond the scope of the present study. For a discussion the reader is referred to the work of Mahadevan et al. (2008) and Ahmadov et al. (2007).

4.1.2 Uncertainty of regional signals

Because the general correspondence with the observations is high, we gain confidence that we can use the model to disentangle the CO_2 signal into its individual components. Figure 9 shows simulated GEE, respiration and fossil fuel
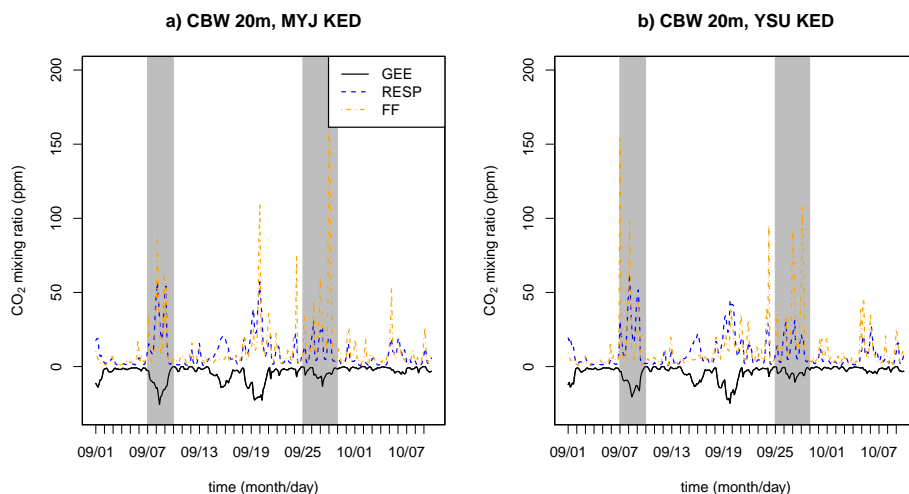
Figure 9. Simulated CO_2 tracers for the vegetation signal Gross Ecosystem Exchange (GEE, black line), respiration (RESP, blue dashed line) and the fossil fuel tracer (FF, orange dash dotted line) at Cabauw 20 m (a.g.l.) from the MYJ KED STILT simulation **(a)** and the YSU KED simulation **(b)**. The grey areas indicate the occurrence of two events characterized by comparably large model errors: period 1 (7 to 9 September 2009) and period 2 (25 to 28 September 2009). Airflow in both periods is qualitatively different, resulting in a strong vegetation signal in period 1 while in period 2 the FF signal dominates. In these periods quantitative model–model differences are notable although the same surface fluxes were used.

signals at CBW. The differences between these components point to strong sources of CO_2 in that part of the domain, as expected. Two events with specifically large amplitudes in the fossil fuel signal (henceforth referred to as period 1 and 2, marked grey in Fig. 9), correspond to notable deterioration in the KED simulations, which might be caused by overestimated fossil fuel fluxes. To test the relative impact of the regional fossil fuel signal ($CO_{2,ff}$) on STILT with optimized MH during these periods, we isolated the signal from regional biosphere–atmosphere fluxes ($\Delta CO_{2,veg}$) from measured CO_2 concentration ($CO_{2,meas}$) as in Gerbig et al. (2003b) using observed and STILT simulated CO signals:

$$\Delta CO_{2,veg} = CO_{2,meas} - \Delta CO_{2,ff} - CO_{2,bg}. \tag{9}$$

Here $CO_{2,bg}$ is the contribution from the background concentration advected to the receptor, which is simulated by STILT. The regional fossil fuel signal is then approximated by

$$\Delta CO_{2,ff} = (CO_{meas} - CO_{bg}) \frac{\Delta CO_{2,ff,mod}}{\Delta CO_{ff,mod}}. \tag{10}$$

The ratio on the right-hand side of Eq. (10) relates regional CO_2 and CO signals as simulated by STILT. Similarly, the advected background signal CO_{bg} was again obtained from STILT, which uses MACC reanalysis and accounts for chemical production/loss during transport from the lateral boundary to the receptor. CO_{meas} is the measured CO signal at CBW.

Figure 10 shows the resulting estimates for the signal from regional NEE fluxes using each of the four simulations of our experiment. The KED simulations exhibit slightly reduced

random errors and absolute bias changes relative to the simulations with unoptimized MHs from > 1 ppm to < 0.5 ppm. Period 1 shows less deterioration; thus fossil fuel emission might indeed be overestimated. However, especially the YSU simulation remains deteriorated. The MYJ–YSU differences during these periods and the general low amplitude of the models in the second period (Fig. 10) point to transport-related uncertainties, which is discussed below.

We observed a notable model mismatch of the CO_2 diurnal cycle at HEI (see Sect. 3.3.2). This effect could be the result of a wrong diurnal cycle in the used EDGAR fossil fuel emission inventory. Here we used the observed CO_2 vegetation signal that was estimated with Eq. (9) to compare the phase of the mean diurnal cycle to the simulations (not shown). We found a similar shift in the phasing of the diurnal cycle of the vegetation signal. This indicates that a potential error in the temporal variability of the EDGAR emissions cannot fully explain the mismatch in the diurnal cycle. We rather need to consider transport-related causes for these mismatches, which is discussed in Sect. 4.2.

4.1.3 Uncertainty in CO emissions

We also compared simulated CO directly to observations for indications of overestimated fossil fuel emissions shown in Fig. 11. Even before optimizing MHs, CO was close to observed levels and even higher in period 2, although nighttime MHs were likely too high in both PBL model setups. This is another indication that fossil fuel emissions might be overestimated. Especially in the Netherlands substantial uncertainties in fossil fuel inventories were shown by Peylin
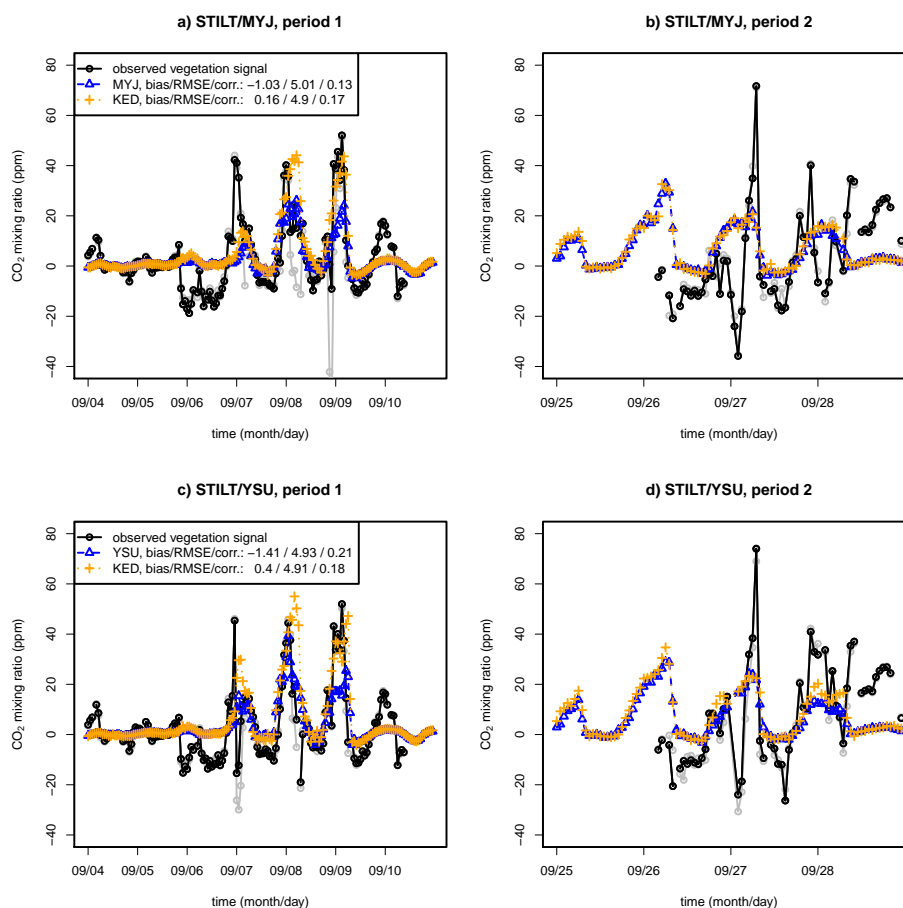
Figure 10. Approximated regional vegetation signal with (grey lines) and without (black lines) optimized MHs during period 1 and 2 of 2009 using the STILT/MYJ model **(a, b)** and the STILT/YSU model **(c, d)** at Cabauw 20 m (a.g.l.). The observed vegetation signal was derived from modelled and observed CO concentrations by Eqs. (9) and (10). Compared are also the modelled vegetation signals with (orange pluses) and without (blue triangles) using optimized MHs. For both of these variants we calculated the observed vegetation signals shown for comparison with black and grey circles respectively. In the upper left corner of **(a)** and **(c)** are summary statistics calculated for the full times series of the signal mismatch: STILT "observations" together with STILT/"observation" squared correlation coefficient (r^2). Statistics are shown for the approximations without MH optimization.

et al. (2011). However, error in fossil fuel signal can also be caused by mismatches in transport, which could lead to the advection of CO_2 from relatively strong fossil fuel emission sources. Such potential uncertainty in transport is discussed in the following section.

4.2 Uncertainty in transport

There are striking differences between the two PBL schemes. Because both STILT/MYJ and STILT/YSU use the same input fluxes, any CO_2 deviations between the two sets of modelling results have to be transport related. Thus, after optimizing the MH, we expect model differences in CO_2 to become smaller. Table 4 shows a comparison of actual model–model differences at CBW. While differences in MHs are well reduced when using MH optimization, an increase in model divergence of simulated CO_2 can be observed especially at

night-time which is contrary to our expectations. To understand the reasons for this behaviour we studied cases where model–model differences are most obvious. We find most striking differences during the events in period 1 and 2 (marked in Fig. 9). Figure 12 shows a comparison of STILT and observed wind speeds at CBW. Wind speed is about a factor 3 higher in period 1 and this is well captured by all simulations. Because of the higher transport speeds, the footprints capture a larger source area in the last 12 h before arrival at CBW. Good correspondence was found between modelled and measured local wind direction (not shown). Since CO_2 surface influence is more local in period 2, we expect a greater impact of the MH on the CO_2 signal. Similar to Vogelezang and Holtslag (1996) we estimated MHs at CBW from observed meteorology (Eq. 1). The data coverage is almost complete in the full period of interest, but only MH below 200 m can be detected. During period 1 CBW MHs were
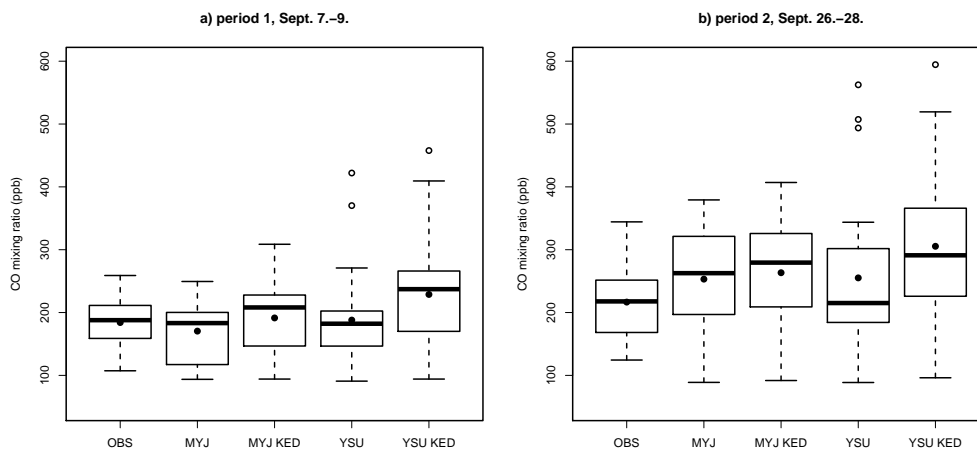
Figure 11. Comparison of observed (OBS) and simulated night-time CO concentrations using the STILT/MYJ simulation (MYJ), the STILT/MYJ simulation with optimized MHs (MYJ KED) and similarly the STILT/YSU simulations (YSU, YSU KED) during period 1 (**a**) and 2 (**b**). Boxes denote the central 50 % of the data, whiskers are at ±1.5 times the inter-quantile range and outliers are shown as open circles. Black line within boxes indicate the median value. Filled points show the mean values.
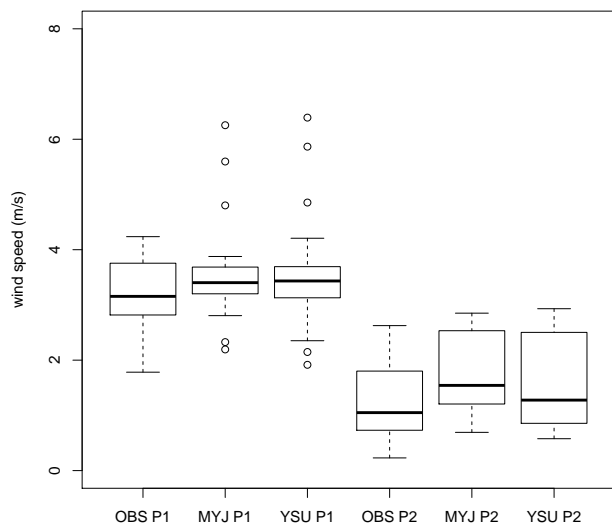

Figure 12. Comparison of observed (OBS) and simulated 20 m wind speed by STILT/MYJ (MYJ) and STILT/YSU (YSU) both using KED optimized MHs at Cabauw. Results are shown for period (P1) and period 2 (P2) shown as grey stripes in Fig. 9. Boxes denote the central 50 % of the data, whiskers are at ±1.5 times the inter-quantile range and outliers are shown as open circles. Black line within boxes indicate the median value.

on average 100 m, in contrast, KED estimated average MHs of 50 m. Contrarily, during period 2 KED and CBW observations agree on average MHs of 50 m, which could explain the little effect of the MH optimization on CO_2. On the one hand a local surface influence and correctly simulated MHs suggest overestimated CO_2 emissions. On the other hand the CBW site is surrounded by strong point sources, e.g. Amsterdam, Utrecht, Rotterdam; thus even minor mismatch in horizontal advection might have a large impact on CO_2 concentrations.

Obviously, KED MHs are too low in period 1. To test the impact of the local MH on CO_2, we used the CBW MHs as additional conditioning data for the KED prediction. The resulting CO_2 time series for periods 1 and 2 are shown in Fig. 13. The problem of deterioration is largely reduced in period 1, while period 2 is nearly unaffected, which could be expected from the MH comparison. However, a notable overestimation of CO_2 remains during two nights within period 1.

Overestimation was found to be stronger in STILT/YSU, and to find causes for this effect we studied cases where these model differences were most obvious. At 7 September, 00:00 UTC in period 1, STILT/YSU shows a large peak in fossil fuel CO_2, which is not seen as strong in MYJ (Fig. 9). Figure 14a shows southwesterly footprints in the 12 h before arrival at CBW. The STILT/MYJ footprint covers about twice the area, i.e. particles travel faster. Thus, STILT/YSU influence is more local and at the same time MHs are at 50 m while MYJ KED MHs are quickly above 100 m. In contrast, 28 September is an instance during period 2 where STILT/MYJ shows a stronger peak in CO_2 than STILT/YSU (Fig. 14b). Here the MHs are at about 50 m in the last 3 h before particles arrive at CBW in both simulations, which was found to be in line with CBW MHs. The horizontal advection was notably different in both simulations, which caused STILT/MYJ to capture emissions from the Rotterdam area when surface influence was still above 0.5 ppm µmole^{-1} m^{-2} s^{-1}, while STILT/YSU just missed these emissions. From these examples it is clear that uncertainty in simulated horizontal winds needs to be considered, for example using a method as introduced by Lin and Gerbig (2005).
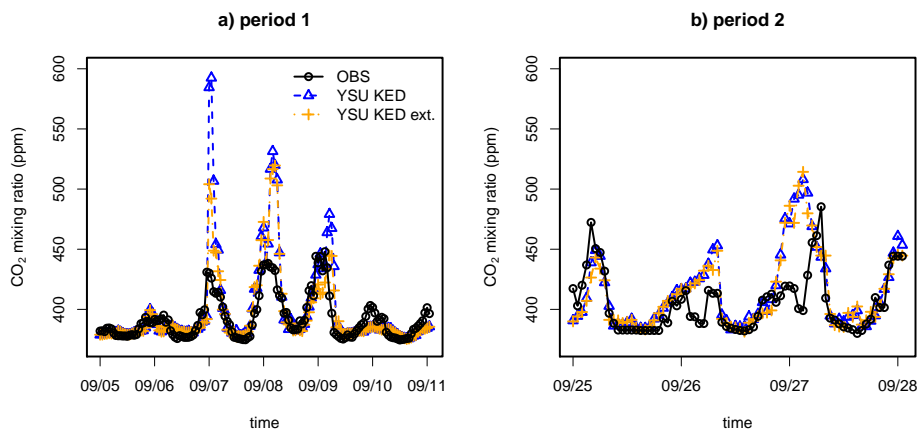
Figure 13. Observed and simulated CO_2 time series for period 1 (**a**) and 2 (**b**) of 2009 at Cabauw 20 m (a.g.l.). Shown here is the STILT/YSU simulations for which deterioration in simulated CO_2 was most notable when using optimized MHs (blue triangles). Using additional night-time MHs observed at Cabauw to condition the KED optimization reduces the overcompensation in period 1 notably (**a**, orange pluses). No such effect is observed in period 2 (**b**).

4.3 Potential of MH optimization for regional CO_2 inversions

The key advantage of geostatistical methods like KED lies in the provided estimation errors, which are a combination of the uncertainty related to spatio-temporal interpolation and uncertainty of estimating the MH from RS data. These errors can be propagated to CO_2 fluxes estimated from the transport inversion as suggested by Gerbig et al. (2008), providing improved and more reliable inversion results. In this approach the MH error variance from KED is added as an additional term in the stochastic calculation of each particle trajectory. Two STILT runs, one with and one without the additional stochastic process, yield ensembles of particle trajectories that provide two distributions of CO_2 mixing ratios with differing variances, and the differences in these variances provide an estimate of MH-induced CO_2 uncertainty. The additional CO_2 uncertainty variance is then added to the diagonal elements of the error covariance matrix. However, this method requires that the transport model has an unbiased representation of vertical mixing. Our results indicate that the systematic error in simulated CO_2 due to errors in MHs were largely reduced by the MH optimization for night time observations, and at least slightly reduced for daytime observations. Thus it can be expected that the proposed method has potential to yield more reliable results when applied for inverse estimation of surface–atmosphere exchange fluxes. However, substantial problems are associated with the spatial and temporal undersampling of the RS data (Kretschmer et al., 2013). The fact that including CBW MH measured at the tower yielded better results in simulated night-time CO_2 clearly indicates the potential for using MH data based on observations made in close proximity to the CO_2 measurement site. In contrast, the HEI comparisons showed large biases caused by a shift in the diurnal phasing, which can only be

resolved by adding MH observations in the hours between 00:00 and 12:00 UTC, e.g. from continuous retrievals of the MH.

4.4 Outlook

4.4.1 Additional MH data

The IGRA data set is limited especially in the temporal resolution. The KED estimation error is a tool that can help to guide the installation of future instrumentation. A particularly promising data set can be obtained from lidars or ceilometers, that are able to continuously observe aerosol backscatter signals, from which MH can be retrieved at relatively low cost (Eresmaa et al., 2006). Networks of lidars and ceilometers are already operated throughout Europe (Haeffelin et al., 2012). Within the Integrated Carbon Observation System (ICOS[3]) project a network of atmospheric measurement stations is built for which MH measurements are mandatory for level 1 continental stations. In addition, Jordan et al. (2010) demonstrated the retrieval of the PBL heights from satellite-based lidar data. A challenging task for the future will be to assure the consistency of effective tracer mixing and these different MH estimation methods. Future research should aim at the quantification of the uncertainty from different kinds of MH observations. In the present study we have shown the substantial impact of such uncertainties on KED estimates.

4.4.2 Improvements of MH estimation

The chosen KED approach is quite simplistic in terms of the underlying product–sum covariance model which assumes space–time stationarity. This assumption is likely to
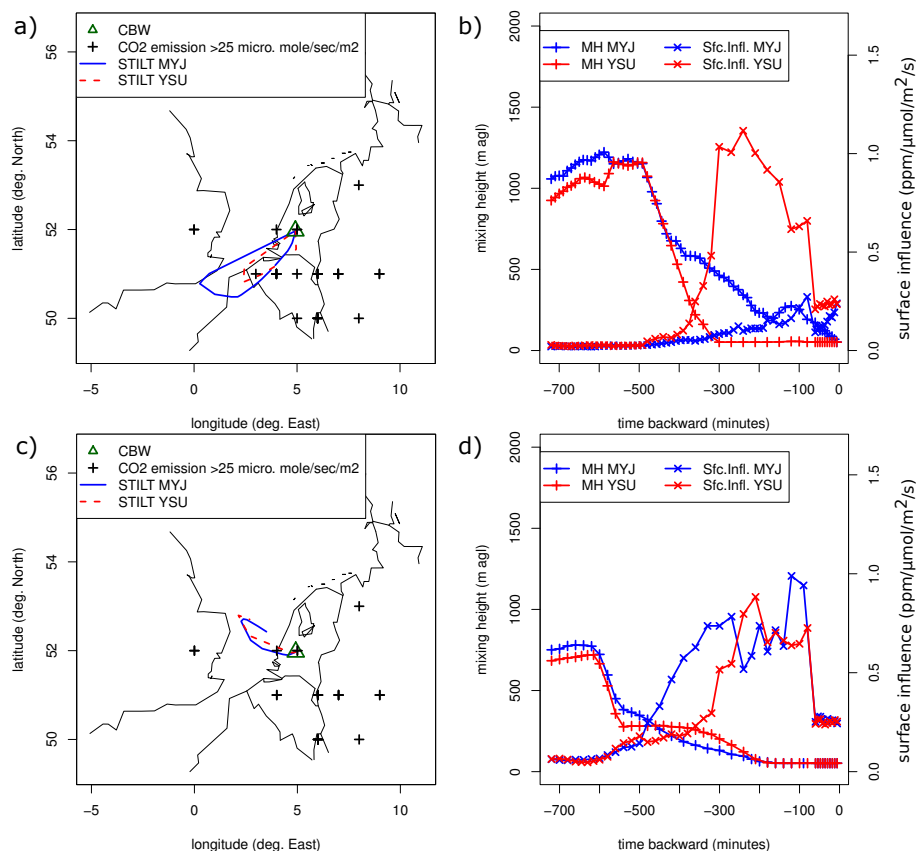
---

[3]www.icos-infrastructure.eu

Figure 14. Comparison of STILT simulated surface influence using KED-optimized MHs. Shown are maps around the Cabauw tall tower (green triangle) in **(a)** and **(c)** together with polygons indicating the area covered by particles with non-zero surface influence in the last 12 h before arrival at Cabauw for the STILT/MYJ (blue lines) and STILT/YSU (red dashed lines) simulations. These areas were approximated by convex hulls around horizontal particle distributions. Also shown are strong point sources in the EDGAR emissions (black pluses) on 7 September 2009, 00:00 UTC **(a, b)** and 28 September 2009, 02:00 UTC **(c, d)** when deviations in fossil fuel signals between the simulations were found most distinct (Fig. 9). Surface influence is also shown for the simulations STILT/MYJ (blue crosses) and STILT/YSU (red crosses) as a function of time before arrival at Cabauw **(b, d)**. The optimized MHs using WRF/MYJ (blue pluses) and WRF/YSU (red pluses) as external drift in KED are plotted for comparison. Although the same conditioning data were used, the estimated night-time MHs differ notably in **(b)**, leading to corresponding differences in surface influences.

be violated by processes like PBL mixing as the temporal partial ranges of the variogram model are likely changing during day/night transition times. Non-stationary covariance models exist, but their application to the special problem of PBL mixing is non-trivial and an improvement in predictions skill is not guaranteed (Paciorek and Schervish, 2006). In addition, such methods require the analysis of temporal correlation length scales from continuous MH observations, e.g. from ceilometers. Nevertheless, we believe such an analysis would be worth future research.

We found indications of overestimation in the KED variances, which could be caused by biases in the semivariance model due to MH uncertainty. Methods exist that allow one to account for such biases in the estimation of the covariance model (e.g. Christensen, 2011). However, MH uncertainty itself is a function of the PBL mixing process, e.g. day/night difference of error magnitude. In such a situation the bias

correction of the variogram model may be flawed, which similarly can cause biased estimates. The use of variance-stabilizing transformations are suggested to circumvent such complications (Christensen, 2011). Additional research in this direction could lead to improved estimation of uncertainty, which in turn may have a positive impact on inversion results.

5 Conclusions

We evaluated a method to interpolate MH estimated from data of the IGRA database using the KED approach. The impact of the resulting optimized MHs on regional tracer transport was assessed by comparing two STILT simulations with different kinds of WRF PBL parameterizations (MYJ and YSU) to observations.

Referring to the questions posed in the introduction we summarize and conclude:

1. Significant biases in CO_2 at CBW, where the diurnal evolution of CO_2 was captured well, were approximately 7–9 % (day) and 60 % (night) of the regional signal. The use of optimized MH data yielded bias reductions of 29–44 % (day) and 68–90 % (night), respectively. While bias reduction at HEI was similarly effective at night, daytime biases remained due to errors in phasing of the CO_2 signal. The effect on random errors and r^2 was mixed. At CBW, RMSE was even increased by using optimized fields. The reasons for this effect are likely related to the sensitivity on the external drift and errors in horizontal advection. In addition, we found indications of overestimated fossil fuel emissions.

2. Simulated MHs were significantly biased with values of -20 % (MYJ) to 10 % (YSU) at day and 40–60 % at night, respectively. RMSE was of the order of 90–116 % with larger values for the YSU scheme.

   Because of these results, we conclude that mismatches in MHs lead to significant bias and random error in tracer concentrations. Therefore, biased flux estimates due to mismatches in simulated MHs have to be expected, when using comparable high-resolution transport models in regional inversions.

3. We conclude that the IGRA data set generally provided enough conditioning data to estimate nearly unbiased MH fields. Biases were reduced from up to 77 % to below 10 % in the cross-validation. RMSE was reduced by up to 30 %, r^2 was unaffected at ∼40–60 %. However, comparison at CBW showed a sensitivity of the estimation to the WRF model in terms of differences in advection and the MHs used as external drift. The latter effect could be mitigated by using additional night-time MHs observed at CBW. Additionally, we found that models could not reproduce the phasing of the CO_2 diurnal cycle at HEI, which is likely the result of improper modelling of the timing of MH development. Due to temporal gaps in the conditioning observation data, the MH optimization was unable to resolve this problem.

4. KED MH estimates were in general consistent with effective CO_2 mixing as derived from observed CO_2 profiles of the IMECC aircraft. The CO_2 MH was usually within one standard deviation of the KED error. This finding indicates that the KED variance is an adequate measure to quantify MHs uncertainty.

The effect of MH optimization were generally favourable and the method provides error estimates that can be propagated through a CO_2 inversion to obtain reliable posterior fluxes (Gerbig et al., 2008). Thus we recommend the consideration of the proposed MH optimization for future regional inversions. In particular, the effective removal of night-time bias could potentially allow for the use of night-time greenhouse gas observations. These are currently neglected by inversions, due to large transport model errors. However, these observations could provide important information to constrain flux estimates, especially to gain further insight spatiotemporal variability of anthropogenic fossil fuel CO_2 emissions on regional scales.

References

Ahmadov, R., Gerbig, C., Kretschmer, R., Körner, S., Neininger, B., Dolman, A. J., and Sarrat, C.: Mesoscale covariance of transport and CO_2 fluxes: evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model, J. Geophys. Res.-Atmos., 112, D22107, doi:10.1029/2007JD008552, 2007.

Biavati, G., Feist, D. G., Gerbig, C., and Kretschmer, R.: Error estimation for localized signal properties: application to atmospheric mixing height retrievals, Atmos. Meas. Tech. Discuss., in preparation, 2014.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO_2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, J. Geophys. Res., 116, D23303, doi:10.1029/2011JD016202, 2011.

Christensen, W. F.: Filtered kriging for spatial data with heterogeneous measurement error variances, Biometrics, 67, 947–957, doi:10.1111/j.1541-0420.2011.01563.x, 2011.

Cressie, N. A. C.: Statistics for spatial data, J. Wiley, New York, 1993.

De Cesare, L., Myers, D. E., and Posa, D.: Product-sum covariance for space-time modeling: an environmental application, Environmetrics, 12, 11–23, doi:10.1002/1099-095X(200102)12:1<11::AID-ENV426>3.0.CO;2-P, 2001.

Denning, A., Fung, I., and Randall, D.: Latitudinal gradient of atmospheric CO_2 due to seasonal exhchange with land biota, Nature, 376, 240–243, 1995.

Diggle, P. J. and Ribeiro Jr., P. J.: Model Based Geostatistics, Springer, New York, 2007.

Dolman, A. J., Gerbig, C., Noilhan, J., Sarrat, C., and Miglietta, F.: Detecting regional variability in sources and sinks of carbon dioxide: a synthesis, Biogeosciences, 6, 1015–1026, doi:10.5194/bg-6-1015-2009, 2009.

Durre, I. and Yin, X.: Enhanced radiosonde data for studies of vertical structure, B. Am. Meteorol. Soc., 89, 1257–1262, doi:10.1175/2008BAMS2603.1, 2008.

Eresmaa, N., Karppinen, A., Joffre, S. M., Räsänen, J., and Talvitie, H.: Mixing height determination by ceilometer, Atmos. Chem. Phys., 6, 1485–1493, doi:10.5194/acp-6-1485-2006, 2006.

Geibel, M.: Measurement of climate-relevant trace gases via infrared spectroscopy, Ph.D. thesis, Friedrich-Schiller-Universität, Jena, 2011.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res.-Atmos., 108, 4756, doi:10.1029/2002JD003018, 2003a.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res.-Atmos., 108, 4757, doi:10.1029/2003JD003770, 2003b.

Gerbig, C., Lin, J. C., Munger, J. W., and Wofsy, S. C.: What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, Atmos. Chem. Phys., 6, 539–554, doi:10.5194/acp-6-539-2006, 2006.

Gerbig, C., Körner, S., and Lin, J. C.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, Atmos. Chem. Phys., 8, 591–602, doi:10.5194/acp-8-591-2008, 2008.

Gerbig, C., Dolman, A. J., and Heimann, M.: On observational and modelling strategies targeted at regional carbon exchange over continents, Biogeosciences, 6, 1949–1959, doi:10.5194/bg-6-1949-2009, 2009.

Grell, G. A. and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, 38-1–38-4, doi:10.1029/2002GL015311, 2002.

Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G., Lolli, S., O'Dowd, C., Sauvage, L., Xueref-Rémy, I., Wastine, B., and Feist, D.: Evaluation of mixing-height retrievals from automatic profiling lidars and ceilometers in view of future integrated networks in Europe, Bound.-Lay. Meteorol., 143, 49–75, 2012.

Hammer, S., Glatzel-Mattheier, H., Müller, L., Sabasch, M., Schmidt, M., Schmitt, S., Schönherr, C., Vogel, F., Worthy, D. E., and Levin, I.: A gas chromatographic system for high-precision quasi-continuous atmospheric measurements of CO_2, CH_4, N_2O, SF_6, CO and H_2, available at: http://www.iup.uni-heidelberg.de:8080/institut/studium/lehre/physik4/forschung/Forschungsdatenbank/groups/kk/en/GC_Hammer_25_SEP_2008.pdf (last access: 17 February 2014), 2008.

Hanna, S.: Atmospheric Turbulence and Air Pollution Modelling, Springer, 1st Edn., doi:10.1007/978-94-010-9112-1_7, 1982.

Hengl, T., Heuvelink, G. B., and Rossiter, D. G.: About regression-kriging: from equations to case studies, Comput. Geosci., 33, 1301–1315, doi:10.1016/j.cageo.2007.05.001, 2007.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, 2006.

Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of three planetary boundary layer schemes in the WRF Model, J. Appl. Meteorol. Clim., 49, 1831–1844, doi:10.1175/2010JAMC2432.1, 2010.

Janjic, Z. I.: Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso model, Office Note 437, NCEP Office, 61 pp., 2002.

Jordan, N. S., Hoff, R. M., and Bacmeister, J. T.: Validation of Goddard Earth Observing System-version 5 MERRA planetary boundary layer heights using CALIPSO, J. Geophys. Res.-Atmos., 115, D24218, doi:10.1029/2009JD013777, 2010.

Jung, M., Henkel, K., Herold, M., and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sens. Environ., 101, 534–553, 2006.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, UK, 2002.

Kretschmer, R., Gerbig, C., Karstens, U., and Koch, F.-T.: Error characterization of CO_2 vertical mixing in the atmospheric transport model WRF-VPRM, Atmos. Chem. Phys., 12, 2441–2458, doi:10.5194/acp-12-2441-2012, 2012.

Kretschmer, R., Koch, F., Feist, D., Biavati, G., Karstens, U., and Gerbig, C.: Toward Assimilation of Observation-Derived Mixing Heights to Improve Atmospheric Tracer Transport Models, in Lagrangian Modeling of the Atmosphere, edited by: Lin, J., Brunner, D., Gerbig, C., Stohl, A., Luhar, A., and Webley, P., American Geophysical Union, Washington, DC, doi:10.1029/2012GM001255, 2013.

Levin, I., Hammer, S., Eichelmann, E., and Vogel, F. R.: Verification of greenhouse gas emission reductions: the prospect of atmospheric monitoring in polluted areas, Philos. T. Roy. Soc. A, 369, 1906–1924, 2011.

Lin, J. C. and Gerbig, C.: Accounting for the effect of transport errors on tracer inversions, Geophys. Res. Lett., 32, L01802, doi:10.1029/2004GL021127, 2005.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: the Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res.-Atmos., 108, 4493, doi:10.1029/2002JD003161, 2003.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochem. Cy., 22, GB2005, doi:10.1029/2006GB002735, 2008.

Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3 hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, J. Geophys. Res., 116, D24303, doi:10.1029/2011JD016245, 2011.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and

forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model, Meteorol. Atmos. Phys., 107, 51–64, 2010.

Nisbet, E. and Weiss, R.: Top-down versus bottom-up, Science, 328, 1241–1243, doi:10.1126/science.1189936, 2010.

Paciorek, C. J. and Schervish, M. J.: Spatial modelling using a new class of nonstationary covariance functions, Environmetrics, 17, 483–506, doi:10.1002/env.785, 2006.

Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P., and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: model intercomparison, Atmos. Chem. Phys., 11, 6607–6622, doi:10.5194/acp-11-6607-2011, 2011.

Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., and Lavrié, J. V.: High-resolution simulations of atmospheric CO_2 over complex terrain – representing the Ochsenkopf mountain tall tower, Atmos. Chem. Phys., 11, 7445–7464, doi:10.5194/acp-11-7445-2011, 2011.

Ribeiro Jr., P. J. and Diggle, P. J.: geoR: a package for geostatistical analysis, R news, 1, 14–18, 2001.

Rödenbeck, C.: Estimating CO_2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport. Technical Report 6, Tech. rep., Max-Planck Institut für Biogeochemie, available at: http://www.bgcjena.mpg.de/uploads/Publications/TechnicalReports/tech_report6.pdf, (last access: 17 February 2014), 2005.

Sarrat, C., Noilhan, J., Dolman, A. J., Gerbig, C., Ahmadov, R., Tolk, L. F., Meesters, A. G. C. A., Hutjes, R. W. A., Ter Maat, H. W., Pérez-Landa, G., and Donier, S.: Atmospheric CO_2 modeling at the regional scale: an intercomparison of 5 meso-scale atmospheric models, Biogeosciences, 4, 1115–1126, doi:10.5194/bg-4-1115-2007, 2007a.

Sarrat, C., Noilhan, J., Lacarrère, P., Donier, S., Lac, C., Calvet, J. C., Dolman, A. J., Gerbig, C., Neininger, B., Ciais, P., Paris, J. D., Boumard, F., Ramonet, M., and Butet, A.: Atmospheric CO_2 modeling at the regional scale: application to the CarboEurope regional experiment, J. Geophys. Res.-Atmos., 112, D12105, doi:10.1029/2006JD008107, 2007b.

Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, Ph.: Mixing height determination for dispersion modelling, Report of Working Group 2, in: Harmonization in the Preprocessing of meteorological data for atmospheric dispersion models, COST Action 710, CEC Publication EUR 18195, 145–265, 1998.

Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmos. Environ., 34, 1001–1027, 2000.

Seidel, D. J., Ao, C. O., and Li, K.: Estimating climatological planetary boundary layer heights from radiosonde observations: comparison of methods and uncertainty analysis, J. Geophys. Res.-Atmos., 115, D16113, doi:10.1029/2009JD013680, 2010.

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J.-C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental United States and Europe, J. Geophys. Res., 117, D17106, doi:10.1029/2012JD018143, 2012.

Spadavecchia, L. P.: Estimation of landscape carbon budgets: combining geostatistical and data assimilation approaches, Ph.D. thesis, University of Edinburgh, 2009.

Steinbach, J.: Enhancing the usability of atmospheric oxygen measurements through emission source characterization and airborne measurements, Ph.D. thesis, Friedrich-Schiller-Universität, Jena, 2010.

Stephens, B. B. and Keeling, R. F.: The influence of Antarctic sea ice on glacial-interglacial CO_2 variations, Nature, 404, 171–174, 2000.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1988.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C., Delille, B., Bates, N., and de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, Deep-Sea Res. Pt. II, 56, 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009.

Ulden, A. and Wieringa, J.: Atmospheric boundary layer research at Cabauw, Bound.-Lay. Meteorol., 78, 39–69, doi:10.1007/BF00122486, 1996.

Vermeulen, A. T., Hensen, A., Popa, M. E., van den Bulk, W. C. M., and Jongejan, P. A. C.: Greenhouse gas observations from Cabauw Tall Tower (1992–2010), Atmos. Meas. Tech., 4, 617–644, doi:10.5194/amt-4-617-2011, 2011.

Vogelezang, D. and Holtslag, A.: Evaluation and model impacts of alternative boundary-layer height formulations, Bound.-Lay. Meteorol., 81, 245–269, 1996.

Wackernagel, H.: Multivariate Geostatistics, An Introduction with Applications, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1995.

Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, Physica D, 230, 1–16, 2007.