Response to reviewer #1 (R1)

italics: comments of R1

page, line, table, figure numbers etc. refer to the discussion paper unless stated otherwise

General comment

This study discusses the variability of the vertical profiles of the air vertical velocity variance observed into the convective boundary layer within a small area of about 3 km horizontal length-scale. It is based on the measurements made during 6 selected fair-weather days by 5 Doppler lidar systems installed at three different sites, which were around 3 km apart. The goal is to determine how much of the observed variability of the vertical velocity variance can be due to the small scale heterogeneity within the 3-km side triangle made by those three sites. This is an interesting issue for which the litterature does need some more answers, and which can be quite well addressed here. The manuscript is well-written, well presented, the data analysis is based on a nice dataset, and is made rigorously in several aspects (error analysis especially). However, this analysis misses several important hypotheses for the interpretation of the data to actually investigate this issue as much as it could. The starting hypothesis is put into question at the end. This was actually expected, and the analysis could be very interesting and publishable if the starting hypothesis was different and if the analysis was pushed further.
I believe that this study can be worth publishing, but only after major revision of the data analysis.

Main comments

One of the most important point here that seems to be missed in the analysis is that when the authors are considering 1h-samples of the lidar measurements (for the calculation of the variance), they are considering turbulence structures (or thermals) passing through the lidar as they are advected by the mean wind during this one hour. This corresponds to length scales that are 5 to 10 times larger than 3 km, in the case of the windspeeds observed here (Table 2). For a 5 ms−1 mean windspeed for example, it is a horizontal scale of 18 km that will be represented by the turbulent moment calculated from the one-hour sample. This sample length is very much larger than the surface heterogeneity scale that the authors are considering (about 100 m). So if the lidars were located really on the same heterogeneous area (let us say a 50 km by 50 km square of surface heterogeneity of scale 100 m like that shown within the triangle of Fig. 1), but on different fields as they are here, we would expect them to show a very similar turbulence profile, except very close to surface. (That is the reason why, when the lidars are aligned with the wind, the authors do observe very similar time series and statistics, but with a delay of a few hundred of seconds, which corresponds to the time it takes the structure to move from one site to the other with the mean wind.)

Overall, this means that:
1. the basic hypothesis that the measurements of the lidars are independent as long as they are 2 km apart cannot be right
2. the authors should consider a larger area to have an idea of the surfaces and general area that are contributing to the turbulence observed with the lidars
3. the authors should also consider that the larger the wind, the more structures they take into account in their samples, i.e. the more statistics [larger sample size] *they have into their computed turbulent moments*

Another miss is the consideration of the wind profile, and effect of wind shear. This is not at all

discussed, but it can be very important to understand the variability observed from one place to the other, and from one day to the other (wind is 8 to 12 ms−1 on some cases, which is quite moderate). The wind will increase statistics, but will also increase the shear production. The authors seem to have the possibility to estimate the wind shear close to surface and at the top of the CBL (from soundings at one site, and maybe from the lidars if VADs were made on the same selected days).

For point 2 above, and looking at the area at larger scale (see Fig. below), one can see that the triangle made by the three lidars is located in an area with heterogeneities of large scale. Especially, one can see a 42 km long forest to the southwest of the area, a large coal mine to the north of Hambach and another coal mine to the west of Wasserwerk, and also a few villages around. Depending on the wind, those surfaces will significantly contribute to the observed turbulence statistics in the experimental area, and will also potentially induce a change of wind profile (and shear production) from one site to the other. The large presence of areas of small scale crop fields like shown in the considered triangle is also obvious from this larger scale map.

For point 2 above, the authors could use the area-averaged flux as they did in their current study, but not only over the small triangle made by the three sites: what is the effective area (scale) to be considered in the area-averaged calculation of the normalizing convective scale, in order to minimize the scatter of the day-to-day (and site-to site) variability (of the profiles, or of the maximum normalize variance) ? How do the results change with increasing height ? What is the influence of the wind profile ?
(Note that ideally, an analysis similar to a surface footprint analysis would be very enriching here, but I understand it could correspond to a too large additional analysis. However, even without using a footprint-type analysis, considering various (larger) scales of the area over which the authors are calculating the area-averaged flux, and considering the effect of wind in some way, should help a lot in the understanding and improving the article.)

To sum that up, the **main comments** are
(1) that the relevant area was not appropriate (relevant area = area over which the surface sensible heat fluxes were averaged, used for calculating an averaged convective velocity scale);
(2) that other factors influencing temporal variability of vertical velocity variance, such as wind shear and stability, may exist; and
(3) that the sample size is larger on days with stronger mean wind, which should be considered.

Considering these points, we became aware that the investigation should be divided into two sections that were mixed up before: The vertical velocity variances as derived from the lidar measurements at three locations are actually investigated regarding temporal variability on the one hand and spatial variability on the other hand.

Section 4, comprising the main results, was therefore re-organized with the subsections 4.2 and 4.3, containing now the investigation of temporal and spatial variability, respectively.

What is called the “starting hypothesis” by R1 is mainly examined in new section 4.3. It was re-phrased for the revised version:
“(… T)he locations had to be close enough to be situated within the area of the given surface heterogeneity. For this configuration, the turbulence characteristics derived from the lidar measurements at the three sites should be similar within the range of statistical errors according to Lenschow et al. (1994).”

Answer to comment (1)

The main motivation to choose the investigation area of 5 km x 5 km as shown in Fig. 1 was to average all available measurements of turbulent surface fluxes over an area for which they are most representative. The turbulent surface fluxes can vary strongly even for similar land use classes, as for example for SE1 and Wasserwerk, which may be caused by variability of soil moisture or soil type. It is, thus, not clear if the surface sensible heat fluxes as measured within the area of 5 km x 5 km are really representative for the larger area of 30 km x 30 km. This also means that even a footprint analysis would not necessarily provide more representative values for scaling as long as no corresponding measurements are available.

However, as you argue, turbulence characteristics under cloud-free conditions are influenced by an upstream area which is larger than 5 km x 5 km, even if strong surface heterogeneities exist. Maronga and Raasch (2011) state that "air advected over the heterogeneities 'feels' only a mean surface heat flux that is the surface heat flux averaged along its path". Considering a time slot of one hour and assuming a mean wind speed of 4 m s^{-1}, the length of this path is about 15 km. Using a quasi-realistic setup of LES simulations, Maronga and Raasch (2011) found about 20 km. We decided to average the fluxes over an area of 30 km x 30 km with the lidar locations in its center (new section 2.3).

Answer to comment (2)

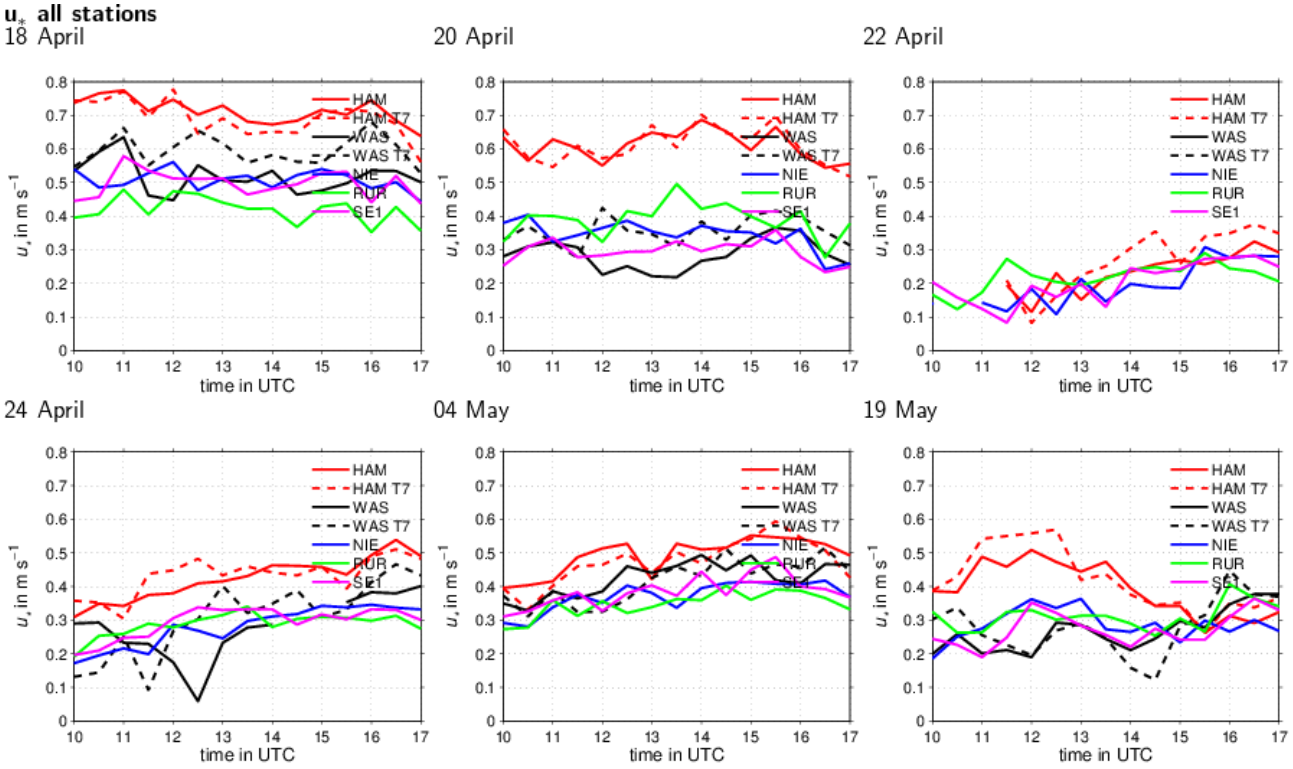

Fig. S1: Diurnal time series of friction velocity u_* for all energy balance stations as well as from turbulence masts at Hambach and Wasserwerk (HAM T7 and WAS T7, respectively).

We performed a more detailed analysis of mean wind speed, friction velocity, wind shear within the CBL, at CBL top as well as stability. According to Lenschow et al. (2012), the parameter - z_i / L was calculated, which is now given in Table 2 (daily averaged values).

Wind profiles were derived from RHI and PPI scans by the VAD algorithm for two sites, but not for the third one, because the lidar there was operated in vertical stare mode during most of the time. Moreover, the wind profiles from radiosoundings yield values above the CBL, too, so that wind shear at CBL top can be calculated, which is not always possible for VAD profiles. Therefore, we decided to use only data of radiosoundings for the evaluation of the wind profile.

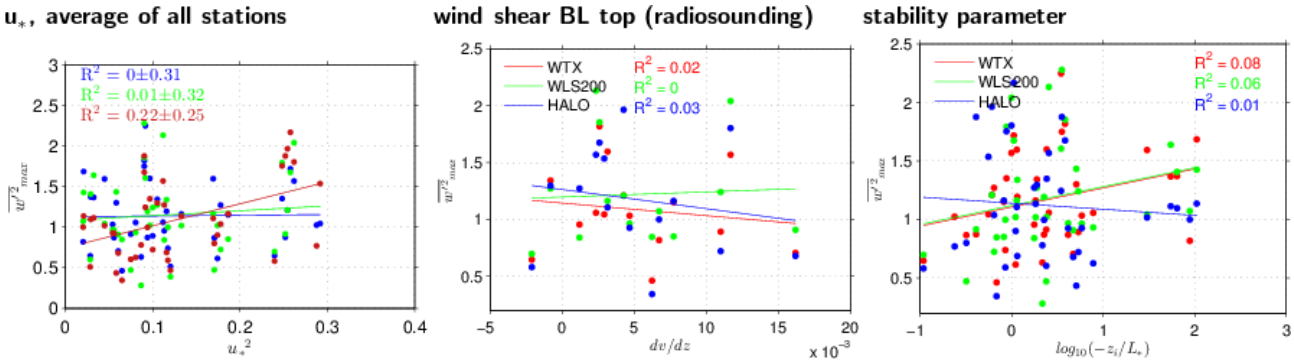

Fig. S2: Correlations of hourly values of vertical velocity variance, averaged over z_{max} ± 250 m and friction velocity u_*, wind shear at CBL top dv/dz, and the logarithm of stability parameter log(-z_i/L) for the six considered days.

Friction velocity (u_*) was taken from the energy balance stations as well as from turbulence masts at Hambach and Wasserwerk that had been installed there for verification of the energy balance stations. It is obvious that friction velocity is largest on 18 April and lowest on 22 April, which were the days with highest / lowest wind speeds (Fig. S1). On all days, friction velocity is highest at Hambach, but it is distinctly higher than at other stations on days with easterly wind (20 April and 19 May). As the measurement site called Hambach was located directly to the west of a large open-pit coal mine, this could be a hint at a possible influence on turbulence characteristics at the surface. However, this does not necessarily mean that the coal mine also influences the turbulence characteristics above the surface layer, in the CBL. We do not expect a dependency in the CBL, where buoyancy production contributes stronger to turbulence. The correlation of friction velocity and vertical velocity variance indicates that there is indeed no relation between both variables (Fig. S2). Correlations were also calculated for each day as well as for daily average values, but this did not hint at a relationship, either. Moreover, correlations were determined for the other variables (examples in Fig. S2), with the same result. The submitted version of the article discusses, thus, the possibility of these relationships only very briefly (new section 4.2.1).

Answer to comment (3)

The increased sample size on days with stronger mean wind is considered implicitly by the statistical error, see added lines in section 3.2: "On days with higher wind speed, the integral time scale and, hence, the statistical error is smaller (Table 2). By this, the dependency of sample size on the mean wind speed is considered implicitly."

Specific comments

Section 1: Introduction

• *page 18012, Abstract: The abstracts does not introduce clearly the addressed issue and main aim of the study. It does not mention where is the experiment set-up.*
The abstract was re-written.

• *page 18013, lines 12-17: It is Taylor's hypothesis which is made here, and should be mentioned. Also stationarity of the sample is assumed. The dataset shown here, with multiple measurements close to each other, gives a very nice opportunity to visit the Taylor hypothesis, and verify when it can actually be made.*
The citation was added (Taylor's hypothesis; Taylor, 1938).

• page 18014, lines 15-17, 'such that the measurements could be assumed to be independent':
The authors needs to clarify what they mean here, and also revise it as they found that they were not independent, or not always.
The statement is now "such that the lidars at the different sites did not sample the same convective cell at the same time". We assume that the turbulence cells scale with z_i so that a distance of about 3 km between each of the 3 lidars was sufficient.

• page 18014, lines 21-27, 'aims of this study': To me, the points enumerated here correspond more to the different steps of the strategy toward the aim. In any case, 'aims' or 'steps', the main goal or main issue should be expressed before those stages.
The aims were re-formulated:
"The aims of this study are to generally analyze the profiles of vertical velocity variance available from HOPE as well as to investigate their spatiotemporal variability. By investigating spatial differences of vertical velocity variance, the representativeness of point measurements of vertical turbulence profiles can be assessed."

Section 2: Overview of the measurements

• page 18015, section 2.1: A map of larger scale than that presented in Fig. 1 would be very useful. I needed it to think about the observations and analysis, and I think it is very important to have it in mind (see Figure below).
We incorporated a larger map now which is equivalent to the new "relevant area" for weighting the surface fluxes.

• page 18015, lines 18-21: 'energy balance... ' at same (Selhausen) site ?
Not exactly, see Fig. 1b.

• page 18016, line 13-14: Horizontal wind profiles from lidar VAD do not seem to be discussed and used in the study. Are they ? (for the estimate in Table 2 of the mean wind in the CBL?)
No, they are not used. The sentence was removed.

• page 18017, section 2.1.2: I would indicate here (rather than later) the fields in which the stations are installed, and describe their nearby environment.
We have considered a different order, too, but it is clearer to leave the description of the land-use classes in section 2.2 (new section 2.3) because it is needed there. Otherwise, we had to repeat it.

• page 18018, section 2.2: I was curious of watching the fluxes directly too, at least the sensible heat flux or buoyancy flux, which will be used later in the convective velocity calculation.
The time series of daily averaged sensible heat flux for all energy balance station was added (new Fig. 2b).

• page 18018, line 20: There is no clear justification of the choice of this area for the area-averaged flux. And as said before, I think the authors have a good opportunity to test the hypothesis made here for the representative flux, by making a sensitivity study to the area (size, and maybe also location) over which the averaged flux is calculated.
The area was increased according to mean wind speed and the hourly averaging period. See also the answer to main comment (1).

• page 18018, line 23-28: I am surprised why the pairs are not {Ruraue, Selhausen} and {Hambach, Wasserwerk}, which is what we deduce from Fig. 2.

Do the fluxes themselves also behave similarly among the pairs ? What do you call 'meadow' ? It seems very different from forest to me. Maybe give a few words about it. Note that needle leaf forest can have very large sensible heat flux.
The pairs are determined according to the land-use classes. Meadow and broadleaf forest were combined. Needleleaf forest had an areal fraction of less than 3 %. The Landsat image (Fig 2a) also indicates that broadleaf forest was more dominant. In the revised version, we added "not for every land-use class, an energy balance station was available". Different weighting approaches are possible and we tried to find the optimal compromise using the available data, but uncertainty due to soil moisture / soil type / more or less advanced growth of vegetation in spring will always remain.

• *page 18019, section 2.3: I guess the 5 selected days are selected among the 19 IOP days. But it is worth mentioning it (that especially means there were radiosoundings every 2 hours).*
Was the wind estimated from soundings or from another device (lidar VAD ?) ?
As said in the text, the criterion for selection was "days with mainly cloud-free CBL conditions, [on which] at least one lidar at each site was configured for *w*-measurements". To connect this to the IOP days, we added "all of these days, apart from 22 April, were also IOP days."
The wind was from the radiosoundings, which is now indicated in Table 2, too.

Section 3: Vertical velocity measurements and variance calculations

• *page 18020, lines 12-14 / page 18025, lines 3-4: It is very nice to see the combination of measurements between the two lidars, which enables you to have a cover from 50 m to the CBL top at least.*

• *page 18020, lines 15-28: Relate energy peak to scales. It is missing here in the discussion, even if it is quantitatively addressed later in the text.*
See comment below.

• *page 18021, 1-10: There are several effects which are mixed here, and the discussion is missing some points. At least four points should be considered when analysing those the spectra:*
– *The expected variation of the vertical velocity variance with height (smaller at top and bottom of the CBL)*
– *The expected variation of the wavelength of maximum vertical velocity spectral energy (as well smaller at top and bottom of the CBL)*
– *The effect of beam averaging (very small loss of energy at the smalles scales)*
– *The slopes of the inertial subrange which are found to be steeper than the -5/3 law within the CBL. And this is not only due to beam averaging (the latter has a much smaller effect), but rather to coherent structures (See Lothon et al. (2007), Lothon et al. (2009), Darbieu et al. (2014)).*
Section 3.1 was re-written in large parts. This discussion of the spectra addresses now 1) the expected variation of the wavelength of maximum and 2) the slopes of the inertial subrange.
The expected variation of the vertical velocity variance with height is not discussed here as it is discussed in relation to the variance profiles in section 4.1.

• *page 18021, 10-12, 'as the main aim of our investigation... this effect will be neglected below':*
It is also justified by its small contribution relative to the total variance.
This discussion can be found in section 3.2 in the revised version. We added "Moreover, the missing contributions are small compared to the absolute values of variance".

• *page 18021, 18-27: Is this estimate of scales done at 600 m ? at what site ?*
w at 600 m and as an average over the 3 sites, complemented in the text.

• page 18022, 15-20: If possible, give an explanation for the small difference observed (size of beam and pulse ?, ...).
"variance differences result(...) from different effective range gate lengths as well as single-pulse energies".

• page 18023 lines 20-21, page 18024 lines 1-5: Yes, this is consistent with the results of Lothon et al. (2009). They found that sometimes, a layer above the CBL with significant vertical velocity variances can be seen (from gravity waves for example, as said later in the text here).
The threshold on the aerosol backscatter was giving more robust results on Zi estimate. The numerous radiosoundings should really help on validating Zi estimates robustly here, in a systematic way.
According to a comment of reviewer #2 (R2), we also added values from method (3) in the figures. Moreover, the discussion of the different methods in section 3.3 was complemented. A systematic validation of method (3) was, however, not advantageous because this method did not yield results in all cases, mainly because the variance profiles did not always converge towards the defined threshold. We decided to take method (2), because it agreed well with method (1). Moreover, it is beyond the scope of this paper to discuss this problem in detail. Relevant literature particularly addressing this problem is given in the text.

Section 4: Spatial and temporal differences

• page 18025 lines 10-15: Profiles of skewness should be discussed more in this study. Lenschow et al. (2012) have shown profiles of higher-order moments of the vertical velocity in the CBL, and discussed them qualitatively in sheared and less sheared CBL. They show that the profiles of skewness are quite sensitive to the shear (or wind) and also to the resolution (of an LES) or spatial averaging (of observations), see figures 5 and 9 of Lenschow et al. (2012). It should be quite sensitive to the sample length and statistics (which can be related to mean wind in your study, as said before). The fact that Selhausen in Fig. 6c shows profiles of smaller skewness, and less marked change drop at the CBL, means that there are different conditions at that site, maybe in wind profile or in the 'quality' of the samples (homogeneity, stationarity).
In general in the manuscript, the effect of wind and shear is not enough taken into account.
We investigated profiles of skewness and found that on daily average, the profile from Selhausen did not deviate from the profiles at the other two sites (Fig. S3, second panel).
According to Lenschow et al. (2012), we calculated the bulk stability parameter and correlated it to all available variance profiles. However, no correlations could be found, as indicated by daily mean values in Fig. S3. Therefore, we did not extend the discussion on skewness in the article.
See also the answer to main comment (2) for a discussion on mean wind and wind shear.

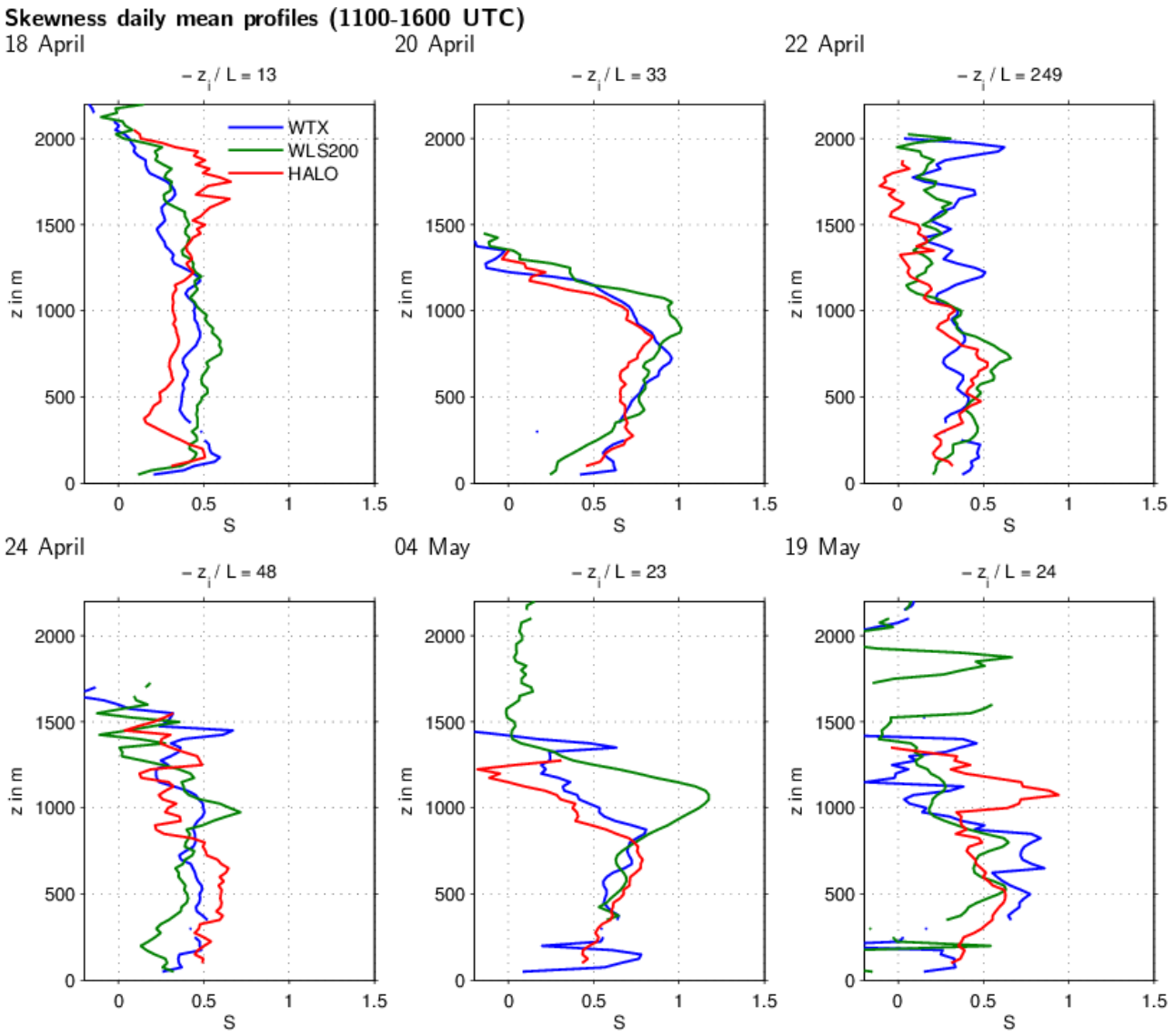

Fig. S3: Daily mean profiles of skewness S; additionally, daily averaged values of -z_i / L are given.

• *page 18026 lines 9-10: 'At Wasserwerk, the variance is slightly lower than at Hambach because less convective cells passed the site': Theoretically, if the sample are representative enough (has enough statistics and homogeneity), the moment should not depend on the number of structures that passed over the site. This might mean that the samples are not long enough. Or that this specific sample is maybe less homogeneous than others. This could also lead to larger skewness for this sample.*
Organized structures like those discussed later in the text can also lead to such kind of bias and lack or representativeness.
This part of section 4.2 was removed, but the problem is discussed in new section 4.2.3 ("Investigation of outliers", see also answer to comment on page 18030 below).
Organized structures are discussed in new section 4.3.3: "The spatial variance differences on 18 and 24 April can therefore be explained by the occurrence of organized structures of turbulence: While more convective cells travel past the Wasserwerk as well as past Hambach, subsidence in the surroundings of these cells prevails at Selhausen."

• *page 18027, section 4.3 I am not sure the discussion in 4.3.1 (starting line 12) is needed. The authors could directly address the w_* scaling issue in a whole. It seems to me that Fig. 12 is telling a lot by itself. Fig. 12b directly shows that the local scaling is not appropriate for scaling the*

maximum variance. The area-averaged scaling is more appropriate. And one question could be: can we minimize the observed scatter (due to day-to-day variability) with an optimized area-representative flux ?
The authors can also address this question with height dependency, expecting the local scaling to be potentially more and more appropriate as we get closer to the ground. (And the sonics at surface and 30 m can help on this point as well). But this might be seen only below 50 or 40 m, that is only with in situ measurements ...
And as said before, sensitivity to sampling representativeness could also be done, or sampling representativeness be taken into account in some way (for example by weighting the cases of most representativity).
This is now section 4.3.1 and it discusses "whether the detected spatial differences of *w*-variance are related to the spatial heterogeneity at the land surface which was described in Sect. 2.3. Even if local scaling could not eliminate spatial differences on average, it could reduce them for the time periods with significant spatial differences."
Sampling representativeness related to mean wind speed and the averaging period used for calculating vertical velocity variances is taken into account by the statistical error. We tested the height dependency of the scaling, but no systematic relationship (either for local or averaged scaling) could be found. The reason is that the lowest range gates already are higher than the layer where turbulence production due to wind shear dominates. Therefore, the correlation of variance and friction velocity is weak at the lowest range gates (Fig. S4), while it is not significantly different from the correlations within the CBL (cf. new Fig. 8a in the article). We also found that correlations between time series of *w* existed between the ultrasonic at 30 m and the lowest range gate of the Windcube at Hambach (40 m). The correlation was clearly weaker between ultrasonic measurements at 4 m and 30 m (not shown). This investigation was not included in the article as it does not clarify the investigated problem.
The correlations shown in new Fig. 8 changed partly compared to the first version because more time steps are considered now (1000-1700 UTC instead of 1100-1600 UTC) and because a vertical average of *w*-variance was taken instead of variance at 0.35 z_i.

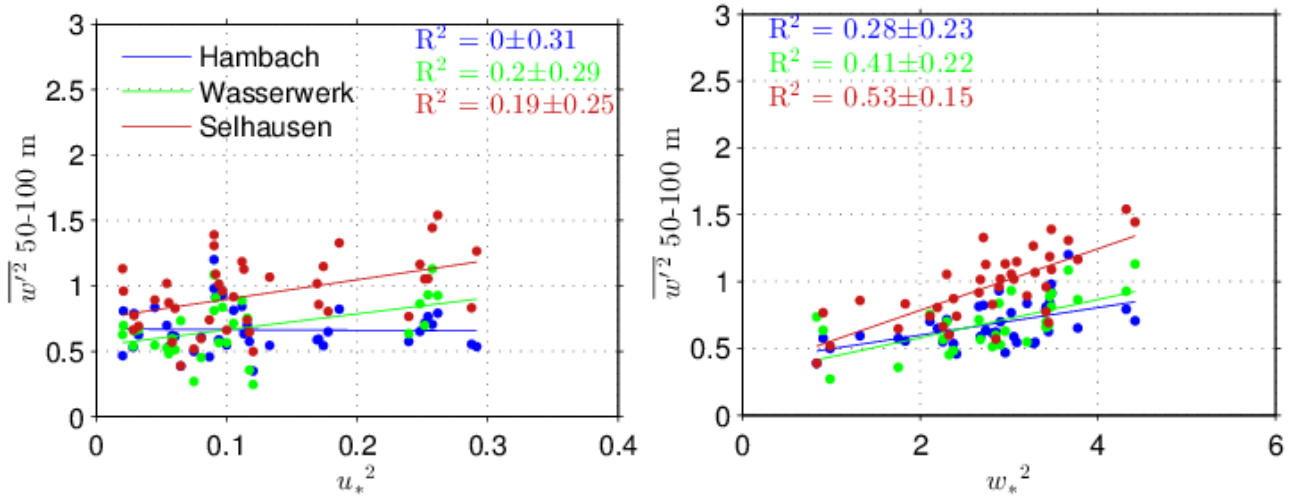

Fig. S4: Correlations between w-variance at lowest range gates and u_* and w_*, respectively (averaged scaling).

• page 18028, lines 12-14, 'it must be concluded that the heterogeneous surface conditions cannot explain the statistically significant differences of the w variances.' This is expected from the sample representativity discussed in main comments. The authors should also consider the surfaces around the area, and the wind, in their discussing the variability of the variance profiles with sites and days. For example, when the wind is south-westerly, the experimental site seems to be at the lee of a

42 km long forest area, which definitely must impact the turbulence observed (both from the buoyant and the dynamic point of view). Similarly, in north-easterly flows, Hambach is in the closest to a large coal mine, which also can impact a lot the observed statistics.
Note that this part of the article particularly addresses the spatial variability!
Westerly wind: The wind is from 250° on 18 April and from 270° on three other days, but the wind had to be from less than 230° to come from the forest. Apart from that, surface heterogeneities at a distance which is more than twice the distance between each of the three sites cannot cause the observed spatial heterogeneity, because the impact should be equal at all three sites.
Easterly wind: We agree that there could be an influence of the coal mine. In the new section 4.3.3, we added: "On 20 April, mean wind came from northeast, so that thermals traveling from Hambach to Selhausen could have been observed. However, this was not the case, and *w*-variance at both other sites differed from the one at Hambach. One possible explanation is that, on days with easterly wind, the strongest influence of the open-pit coal mine on *w*-variance occurs at Hambach."

• *page 18028, lines 19-21, 'It is assumed that the local diurnal cycle of the energy input as well as local differences from day to day can be taken into account better by local scaling than averaged one.': Isn't this contradictory to the above conclusion ? (page 18028, lines 12-14)*
The sentence in question can be found in new section 4.2.1 now, i.e. before any conclusions about local or averaged scaling are drawn.

• *page 18028, lines 23-23: Why is 19 May excluded ? I find this case is a good testimony of the analysis, with smaller heterogeneity for this case deduced from the wet ground. It should help in the analysis of the most appropriate scaling, and in the general understanding even (or especially?) if it turns out to be an outlier sometimes.*
The profiles from 19 May are only excluded in the new Figs. 7 and 8, where temporal variability is discussed, not for the discussion of spatial variability. For example, the dots with a normalized variance > 0.6 in Fig. 13c) correspond to 19 May.
As can be seen in Table 2 and Fig. 10, the variance is not smaller on 19 May than on the other days, but mean sensible heat flux and consequently w_* is smaller. This means that normalized variance is larger than on the other days, which would lead to a much larger scatter of normalized profiles in Fig. 7.

• *page 18030 line 25 to page 18031 line 6: This is an interesting discussion to be associated with the sampling issues, for the understanding of the variability of the variance profiles. But it is not clear what scales are considered here, when talking about 'variance of thermal' and 'variance of environment'. It seems that the scales considered for the moment calculation are much smaller in this conditional analysis than those considered in your analysis.*
The discussion was moved to a separate section now (new section 4.2.3). Lenschow and Stephens (1982) show a normalized variance of thermals of 0.6 at maximum and a normalized total variance of about 0.25, which is not much different from values shown in new Fig. 7. As said in the text, "a sub-sampling [as done by Lenschow and Stephens (1982)] would be beyond the scope of this investigation", but we could prove the existence of broader and / or more numerous thermals by analyzing the frequency distribution of *w*. Moreover, the difference between different time periods was statistically significant, i.e. it was not caused by insufficient representativeness of the sample.
With this, the hypothesis that significantly increased variance is caused by an increased frequency of thermals, which was in first step based on the result of Lenschow and Stephens (1982) that variance of thermals is higher than in the environment, could be confirmed.

• *page 18031, section 4.4: It is very interesting to study the influence of sample length in this study. However, it seems to me that the authors do not explore scales larger than 1 hour. As soon as the authors are averaging variances computed over 1h samples for periods of Δt larger than 1h, the scales that are represented will be still those smaller than 1 h (30 min actually). It is a 1 h filtering.*

You gain more statistics and reduce random errors, since you have 5 samples in a 5 h sample, but the sampled length scales remain the same, and are not larger than 1h (30 min). To me, that is why the curves in Fig. 13c are nicely leveling. Fig. 5 though, very interesting as well, shows that the authors can still consider intervals that are larger than 1h: Most of the days shown are quite stationary during the period from 11:00 UTC to 15:00 UTC. That is the authors could consider samples of 2, 3, and 4 hour long over this period of the day.
Filtering could still be done for all samples at a given cut off frequency, when wishing to keep smaller scales only in the computed variance (and longer samples would then increase the statistics). But I am not sure this was the goal here.
We compared variances calculated for longer averaging intervals by explicitly using these intervals with averaged variances from the hourly intervals and found only small differences. The main result (existence of statistical significant spatial variance differences) was not affected by the method used to determine the variance for longer averaging intervals.
We missed to point out that also the errors for all averaging intervals were explicitly calculated for these periods. In new section 4.3.2, we added: "the statistical error (Fig. 13b) is taken from variance calculations for explicitly larger time periods."

• *page 18033, lines 9-10: Variance is smaller and skewness is smaller as well, and with a less tilted height dependency than other cases. Hypotheses of a difference in wind profiles or in sample homogeneity should be investigated, along with the role of the long and large forest to the south-west of the experimental area.*
See answer to main comment (2) for a discussion about dependency on the wind profile and answer to main comment (3), referring to sample representativeness.
However, it is not true that profiles of skewness at Selhausen are different from those at the other sites (Fig. S3).

• *page 18034, section 4.5, title and discussion: I would not call this section 'influence of wind'. The influence of the wind is almost not considered in the study. The discussion of section 4.5 is linked with the observations of very coherent measurements between two sites that are aligned with the wind. We do expect this coherency, as well as the delay of 200 s and 400 s respectively for April 20 and April 24, given the mean wind of* $\sim 10\ ms^{-1}$ *and* $\sim 5\ ms^{-1}$ *respectively. Even if it is actually very nice to see it so well, and to be able to quantitatively explore Taylor's hypothesis. But this section is more linked with sampling issues and analysis strategy, than with the influence of the wind profile (and wind shear) itself on the observed vertical velocity statistics.*
The new section 4.3.3 is now called "Correlations of vertical velocity at different locations".

• *page 18035, lines 10-15 : This is because in case of the two sites aligned with the wind, they are sampling exactly the same air mass, one site being at the lee of the second. This is not the case when the wind is different, and especially not the case when the wind is perpendicular to the axis made by the two sites.*
This is correct, but we go a step further in hypothesizing in new section 4.3.3 that subsidence at the third site during more than two hours is related with this, so that we can assume the existence of organized structures of turbulence. This was added in new section 4.3.3: "As shown by Lenschow and Stephens (1982), the mean *w* within thermals is positive and nearly two times higher than in the environment, where it is negative. This agrees very well with the mean *w*, observed at the different locations on 24 April (Fig. 15). The spatial variance differences on 18 and 24 April can therefore be explained by the occurrence of organized structures of turbulence."

• *page 18035-18036 : I am not sure the LES is very useful here. It is not used at all for the previous questions, and especially not for the issue about surface heterogeneity and representative scaling. Nor for the study of the effect of wind. It is true that rolls occur, and that they can impact very much on our interpretation of the observed turbulence statistics. But I am not sure this limited discussion*

based on the LES at the really end of the manuscript is appropriate.
Also note that there is a possible mis-interpretation of the LES fields: when averaging over 1h, the organization seems emphasized possibly artificially because the structures seen in Fig. 16a have been advected at each time step along wind during the 1 h interval. Which can make those 'rolls' appear in Fig. 12b when averaging all of them. So I believe that the averaging is making the rolls here. The band-like structures in the instantaneous field of Fig. 12a are more reliable.
We want to show that convective cells can persist for a certain time period. Assuming "frozen" turbulence, the averaged field from the LES (created by averaging model output of *w* at every time step (1 s)) corresponds to the time series of the lidar data.
To avoid mis-interpretation, we added: "The instantaneous as well as the field averaged over one hour is given" (new section 4.3.3). We also added this in the figure caption (Fig. 16).

Section 5: Conclusions

• *page 18038, lines 12-13, 'as only days with buoyancy-driven turbulence have been chosen' : This is not quite true because this study by Lenschow et al. (2012) distinguishes the most convective cases with the least convective cases, the latter being those with stronger wind. They show differences of profiles of higher moments (including variance and skewness) between the most convective and the least convective cases, both with lidar observations and LES. Strongest wind in their case is around 8-9 ms−1, which is not as large as one case here with 12 ms^{-1}, but is still moderate. The study by Maurer et al. could actually be very complementary of that previous study (and pushing one step further), with the different suggestions made before.*
The Conclusions were revised, but the dependency of *S* on stability could not be confirmed (Fig. S3).

Figures and Tables

• *page 18048, Table 2: what site(s) is/are considered for those estimates here ? is the integral scale calculated at 600 m or at the height of maximum variance ?*
*"*same height as w'²_max" was added in the caption of the Table.

• *page 18049, Fig.1: I noticed from google-earth that the white patches in Fig. 1 are small villages. This should be specified and not ignored in the analysis. Add a larger scale map of land-use too.*
A larger map was added. The land-use is bare soil for villages, as indicated by the legend. To make this clearer, the Landsat image is shown now (Fig. 1a).

• *page 18050, Fig.2: Change one of the green colors, because the two greens are very close to each other, this is confusing. I suggest to identify the 6 days, selected for this study. I also suggest to add buoyancy or sensible heat flux, and a time series of Zi would be interesting too.*
See modification in Fig. 2; diurnal time series of z_i are shown in Fig. 5, but a determination of z_i for all days would be beyond the scope (unclear definition for days with CBL clouds, precipitation, strong instationarities, etc.).

• *page 18050 Fig.3 and page 18061 Fig. 13: I suggest to specify the location/site, rather than the lidar model in this figure, because that is what matters here. In Fig. 3a, the layer above 1000 m should be discussed in the text.*
In section 3.1, it is more the comparison of different instruments which is of interest, as for example of WLS7 and HYB at Wasserwerk as well as of HYB and WTX. It is mentioned that the HYB yields measurements above the CBL height and a technical, instrument-specific reason is given. The layer itself is described in section 4.1 when the profiles are discussed.

• *page 18059, Fig.11: Note that the variability (standard deviation) of the variance profiles is very*

similar to that observed by Lenschow et al. (2012).
See new section 4.2.1

Formal comments

• page 18014, lines 5-8 The sentence should be separated in two sentences here, for surface case and aircraft case respectively.
• page 18015, line 21: '(energy balance data)... were applied as well' To be reworded.
• page 18016, line 11: 'because lidars only partly penetrate clouds' To be reworded.
• page 18017, line 22-23: '2-hourly intervals'
Do you mean that soundings were launched every 2 hours ?
• page 18018 line 1-4: I understand that the ultrasonic and ceilometer were also installed at Hambach site. Maybe this should be more explicitly said.
• page 18018 line 7, 'using 09:00-15:00 UTC': 'averaged over the 09:00-15:00 UTC interval'.
• page 18019, line 6: Mention that Table 2 gives several characteristic variables for the 6 selected days (not only Zi and wind).
• page 18027, section 4.3.1: I suggest to give the explanation of lines 2-11 in section 3.3.

Done where applicable

References

Darbieu, C., F. Lohou, M. Lothon, J. Vil`a-Guerau de Arellano, F. Couvreux, D. Durand, D. Pino, E. G. Patton, E. Nilsson, E. Blay-Carreras, and B. Gioli: 2014, 'Turbulence vertical structure of the boundary layer during the afternoon transition'. Atmos. Chem. Phys. Discuss. 14, 32491–32533.

Lenschow, D. H., M. Lothon, S. D. Mayor, P. P. Sullivan, and G. Canut: 2012, 'A comparison of higher-order vertical velocity moments in the convective boundary layer from lidar with in situ measurements and large-eddy simulation'. Boundary-Layer Meteorol. 143, 107–123.

Lothon, M., D. H. Lenschow, and S. Mayor: 2009, 'Doppler lidar measurements of vertical velocity spectra in the convective boundary-layer'. Boundary-Layer Meteor. 132, 205–226.

Lothon, M., D. H. Lenschow, and A. Schanot: 2007, 'Status reminder report on C-130 air-motion measurements. Test of DYCOMS-II new datasets.'. NCAR/RAF internal report pp. https:www.eol.ucar.edurafProjectsDYCOMS–IIwind corrections.html.

Maronga, B. and S. Raasch (2013): Large-Eddy Simulations of Surface Heterogeneity Effects on the Convective Boundary Layer During the LITFASS-2003 Experiment, Boundary-Layer Meteorol., 148, 309-331, DOI: 10.1007/s10546-012-9748-z

Response to reviewer #2 (R2)

italics: comments of R2

page, line, table, figure numbers etc. refer to the discussion paper unless stated otherwise

General comments

This study examines the spatial heterogeneity of vertical velocity variance profiles as observed by multiple Doppler lidars. The lidars are situated at the vertices of a triangle with roughly 3km separation between each vertex, and the study area includes substantial small-scale variability in land use. The authors show that traditional local scaling techniques (i.e. using the convective velocity scale from a collocated surface station) is not effective at collapsing the profiles and that knowledge of the wind direction and the up wind fetch is important. One of the main conclusions is that multiple lidar measurements are needed in order to adequately capture the spatial variability of vertical velocity variance profiles in the convective boundary layer. Considering the widespread use of these local scaling methods, I believe the paper makes an important cautionary point. The manuscript is well-written, well organized, and the authors are particularly thorough in their error analysis. I believe the paper should be published with a few fairly minor revisions and/or clarifications. My main points are listed below.

Specific comments

Abstract
The author states that "...differences between variances at different sites were about three times higher than between those derived from measurements by different lidars at the same site." This statement doesn't make much sense to me. Given the three measurement sites, how could you possibly have differences between all three sites (1-2, 2-3, 1-3) equal to 3x the difference at one site?
According to modifications of the article that were done after considering the comments of reviewer #1 (R1), the abstract was rewritten in large parts; the sentence in question was removed.

page 18016, lines 13-16
Its not clear from this discussion how the WLS7 was operated. The author mentions the VAD mode and gives the temporal and spatial resolution for that mode, but then only briefly states that the system was operated in a vertical stare mode. Why was the VAD mode mentioned? Are the winds from the VAD mode being used in this study? Please clarify.
Wind profiles derived by VAD were not used in this investigation. The sentence concerned was removed.

Table 1
This table lists the various lidar systems and some specs. It would be useful to also include the pulse repetition frequency, the pulse integration time and the duty cycle for the vertical staring data.
We complemented the table with parameters that are known for all /most instruments. Unfortunately, not all specifications are given by all manufacturers.

page 18021 lines 6-9
The author states "...the spectra of WLS7 show some artefacts at the highest frequencies... This is presumably the signature of an aliasing effect." The WLS7 spectra as shown in Figure 4 do indeed show some peculiar behavior, and I believe the author should elaborate on the above statement.

Why is this occurring? What is the radial velocity sampling period? Is this significantly different than the averaging time? This is why I asked about the duty cycle above.
As Leosphere did not provide all technical specifications, we unfortunately cannot investigate this into detail here. We can only suspect that a filtering is done during data processing which causes this. Curiously, the peculiar behavior of the spectra occurs only for WLS7, not for WLS200.

Pages 18023 and 18024
The author discusses three different methods for estimating boundary layer height, with method 3 being based on the velocity variance profiles from the lidars. The author points out that the sonde and backscatter method are merely proxies for method 3 (and I would tend to agree). But then at the top of page 18024 they essentially dismiss this method and say "methods (1) and (2) showed good agreement." This implies to me that method (3) did not show good agreement. I believe the authors should show the results of method 3 in Figure 5, and provide a more thorough discussion of the differences.
We provided the results from method (3) and complemented the corresponding text passages (section 3.3): "For the six days investigated here, the methods agree well for most time steps around noon (dashed lines and black dots in Fig. 5). Mainly before 1100 and after 1500 UTC, method (3) yields lower values of z_i than method (2). The reason is that especially method (2) tends to detect the cap of the residual layer, which is not the case for method (3). However, the threshold value of method (3) is not applicable to all of the profiles here. For several time steps, the decrease of variance with height is weak and the variance does not reach the defined threshold, so that zi cannot be determined by method (3). In contrast to method (1), method (2) also provides values for periods when no radiosoundings are available. Therefore, zi values derived by method (2) are used for the following calculations. Correlating all zi values from method (1) against values derived by method (2) from different lidars shows that zi values derived from backscatter data of WTX at Hambach fit best."

Page 18023 lines 23,24
The author states "...can presumably be attributed to the existence of gravity waves..." Since the authors offer no firm evidence, it is best not to "presume." It would be better to say something like "may be caused by gravity waves in the capping inversion layer."
done

Page 18036 line 26
The author states that "Different methods to derive zi agreed well." This conflicts with the statement on page 18024 (see comment above).
adapted according to modifications in the text (section 3.3), see comment above.

Page 18039 line 23
The author should be more explicit about the type error they are referring to. In this case it is the variance of the noise that is being referred to. The author might say "...the noise variance is equal to the difference..."
done

Page 18040 line 11
The author should be more explicit here and say "...it can be seen that the systematic error..."
done

Page 18040 line 16-17
The author should be more explicit here and say "...the random error can be approximated as..."
done

Manuscript prepared for Atmos. Chem. Phys.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 October 2015

Observed ~~spatial~~ spatiotemporal variability of boundary-layer turbulence over flat, heterogeneous terrain

V. Maurer[1], N. Kalthoff[1], A. Wieser[1], M. Kohler[1], M. Mauder[2], and L. Gantner[1]

[1]Institut für Meteorologie und Klimaforschung (IMK-TRO), Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany
[2]Institut für Meteorologie und Klimaforschung (IMK-IFU), Karlsruher Institut für Technologie (KIT), Garmisch-Partenkirchen, Germany

Correspondence to: V. Maurer (vera.maurer@kit.edu)

Abstract. In spring 2013, extensive measurements with multiple Doppler lidar systems were performed. The instruments were arranged in a triangle with edge lengths of about 3 km in a moderately flat, agriculturally used terrain in northwestern Germany. For six mostly cloud-free convective days, vertical velocity variance profiles were ~~compared for the three locations. On the average over all considered cases, differences between variances at different sites were about three times higher than between those derived from measurements by different lidars at the same site. For all investigated averaging periods between 10 minutes and 4 hours, the differences were not significant on the average when considering the~~ calculated. Weighted-averaged surface fluxes proved to be more appropriate than data from individual sites for scaling the variance profiles; but even then, the scatter of profiles was mostly larger than the statistical error. ~~However, statistically significant spatial differences were found in several individual cases. These~~ The scatter could not be explained by ~~the existing surface heterogeneity. In some cases, nearby energy balance stations provided surface fluxes that were not suitable for scaling the variance profiles. Weighted-averaged values proved to be more applicable, but even then, the scaled profiles showed a large scatter for each location. Therefore, it must be assumed that the intensity of turbulence is not always well-determined by the local heat supply at the Earth's surface~~ mean wind speed or stability, whereas time periods with significantly increased variance contained broader or more thermals. Periods with an elevated maximum of the variance profiles could also be related to broad thermals. Moreover, statistically significant spatial differences of variance were found. They were not influenced by the existing surface heterogeneity. Instead, ~~a certain dependency of turbulence characteristics on mean wind speed and direction was found: Thermals were detected that travelled from one site to the other with the mean wind~~ thermals were preserved between two sites when the travel time was shorter than the large-eddy turnover

time. At the same time, no thermals passed for more than two hours at a third site that was located perpendicular to the mean wind direction in relation to the first two sites. ~~Subsidence~~ Organized structures of turbulence with subsidence prevailing in the surroundings of thermals ~~advected with the mean wind~~ can thus partly explain significant spatial variance differences existing for several hours. Therefore, the representativeness of individual variance profiles derived from measurements at a single site cannot be assumed.

1 Introduction

The vertical velocity variance, $\overline{w'^2}$, is one of the relevant parameters describing the turbulent structure of the convective boundary layer (CBL). Measurements of $\overline{w'^2}$ have been analyzed for several decades (e.g. Wyngaard et al., 1971; Panofsky and Mazzola, 1971; Kaimal et al., 1976; Young, 1988). Most of these early investigations were based on aircraft observations. Later, radar wind profiler (e.g., Eymard and Weill, 1988; Angevine et al., 1994; Eng et al., 2003) and more recently, Doppler lidar measurements (e.g. Lothon et al., 2009; Hogan et al., 2009; Ansmann et al., 2010; Lenschow et al., 2012) became available for studying vertical velocity characteristics in the CBL. Both in-situ aircraft measurements and ground-based remote sensing have advantages and disadvantages: ~~As~~ as aircraft observations are expensive, data are usually available for a small number of flight levels only. The measurements must cover a certain distance, i.e. flight legs must be long enough, to meet the requirements of ~~turbulent statistics (Lenschow and Stankov, 1986; Lenschow et al., 1994) ,~~ turbulence statistics (Lenschow and Stankov, 1986; Lenschow et al., 1994) so that the turbulence characteristics on the different levels are not available simultaneously. Ground-based remote sensing observations provide ~~turbulent~~ turbulence statistics on different levels at the same time for time periods of typically one hour or even longer. However, even if it is assumed that temporal and spatial integration are comparable, i.e. that time can be transformed into space via the mean wind speed (Taylor's hypothesis; Taylor, 1938) , lidar measurements are representative of a restricted region only.

In the part of the CBL where buoyant production dominates over the shear production of turbulent kinetic energy, turbulent mixing is supposed to be driven mainly by the heat supply at the Earth's surface. Deardorff (1970a) proposed that for situations with sufficient thermal instability, vertical velocity fluctuations could be scaled by the convective velocity w_*. Warner (1972), Willis and Deardorff (1974) and Caughey and Palmer (1979) were among the first to present scaled variance profiles, based on laboratory experiments as well as aircraft measurements performed over mainly homogeneous terrain. Large eddy simulations (LES) confirmed the empirical profiles (e.g. Deardorff, 1974; Moeng, 1984; Hadfield et al., 1991). Different fit functions were proposed by Kaimal et al. (1976), Lenschow et al. (1980), ~~Sorbjan (1988) or Sorbjan (1989)~~ or Sorbjan (1988, 1989) , which reveal a considerable uncertainty. Hogan et al. (2009), e.g., found that scaled variance profiles derived from

lidar measurements at one particular site displayed a case-to-case variability that was about as large as the scatter of the fit functions given by Lenschow et al. (1980) and Sorbjan (1986), which had been derived from aircraft measurements. Hence, the uncertainty or representativeness of point measurements is very relevant and becomes even more important for heterogeneous terrain.

Different studies addressed the representativeness of point measurements of ~~the surface energy balance (e.g. Mahrt, 1998; Steinfeld et al., 2007) and~~ turbulent surface fluxes (e.g. Mahrt, 1998; Steinfeld et al., 2007) . Others examined sampling errors made by aircraft measurements (e.g. Lenschow and Stankov, 1986; Schröter et al., 2000). Lenschow et al. (1994) considered general statistical errors, including the sampling error, that should be taken into account when calculating turbulence statistics. To our knowledge, no investigation specifically addressed the statistical errors made for simultaneously performed point measurements of vertical turbulence profiles.

During the High Definition Clouds and Precipitation for Climate Prediction (HD(CP)^2) observational prototype experiment (HOPE) performed in April and May 2013 in the Lower Rhine region in Germany, Doppler lidars were deployed in a triangle in an agriculturally used, moderately flat terrain (Fig. 1). The length of about 3 km of the three edges had been chosen such that the ~~measurements could be assumed to be independent. As it was suppossed that most processes in the CBL scale with its depth (e.g. Deardorff, 1970a; Willis and Deardorff, 1974) , the lengths~~ lidars at the different sites did not sample the same convective cell at the same time. Hence, the distance between the lidar sites had to be larger than the diameter of the convective cells which are assumed to scale with the CBL depth of 1–2 km (e.g. Deardorff, 1970a; Willis and Deardorff, 1974) . On the other hand, the locations had to be close enough to be situated within the area of the given surface heterogeneity. For this configuration, the turbulence characteristics derived from the lidar measurements at the three sites should be similar within the range of statistical errors according to Lenschow et al. (1994) .

The aims of this study ~~were 1) to compare the~~ are to generally analyze the profiles of vertical velocity variance ~~profiles at the three sites, i.e. to investigate the spatial variabilityof CBL turbulence and with this, to assess~~ available from HOPE as well as to investigate their spatiotemporal variability. By investigating spatial differences of vertical velocity variance, the representativeness of point measurements ~~over patchy terrain; 2) to analyze the conditions of the time periods with statistically significant spatial variance differences in more detail; 3) to investigate the effect of w_*-scaling on the spread of profiles~~ of ~~$\overline{w'^2}$ by using spatially averaged values of w_* or those derived from stations near to the lidar locations; and 4) to determine the impact of the averaging time on spatial variance differences.~~ vertical turbulence profiles can be assessed. The paper is structured as follows: In the next section, the observations and the measurement setup are described. Section 3 presents ~~the~~ analyses of the vertical velocity measurements and gives an overview of the computation of the vertical velocity variances and considered errors~~and~~ . It also includes considerations regarding the normalization procedure. In ~~Section~~ Sect. 4, ~~the main results are presented and discussed. This comprises the spatial comparison of the variances, the investigation of surface and~~

atmospheric conditions during periods with statistically significant differences of the variances, and the discussion of scaled vertical velocity variances are described separately for the different sites as well as compared for the three sites, and possible influencing factors are discussed. Finally, section Sect. 5 summarizes the main findings.

2 Overview of the measurements

2.1 Measurement site and instruments

The HOPE measurement area was located near Forschungszentrum J10001000Jülich, in the north of a low mountain range (Eifel), with two larger open-pit coal mines (up to 10 km wide) and several smaller wooded areas in the vicinity . (Fig. 1a). All instruments considered here were located within an agriculturally used area near the villages of Hambach and Niederzier (Fig. 1 1b). The diagonals of the individual fields with various crops are roughly between 100 m and 500 m. The Landsat image of April 2013 (Fig. 1a) shows that a part of the crop fields was already covered by vegetation while others were still bare.

As part of HOPE, the Karlsruhe advanced mobile observation platform KITcube (Kalthoff et al., 2013) was installed. Most of the KITcube instrumentation was operated at Hambach (50.897°° N / 6.464°° E, 110 m m.s.l.). Additionally, instruments were installed at a second site, called Wasserwerk (50.891°° N / 6.430°° E, 96 m m.s.l.), 2.6 km west of Hambach. For this study, Doppler lidar data from a site near Selhausen (50.869°° N / 6.451°° E, 105 m m.s.l.) and energy balance data from nearby eddy-covariance stations (Graf et al., 2010) of the Terrestrial Network of Observatories (TERENO; Zacharias et al., 2011) were applied used as well. The instruments whose data are used here are briefly described below.

2.1.1 Doppler lidars at three sites

At Hambach, a 1.6-µm heterodyne Doppler lidar (WindTracer "WTX" with an Er:YAG laser, Lockheed Martin Coherent Technologies, Inc.) was deployed. The lidar measures the radial wind velocity via the Doppler shift of radiation scattered at aerosol particles. It can be operated with different scan patterns. Mean horizontal wind speed profiles can be calculated with the VAD algorithm (Browning and Wexler, 1968). Applying the vertical stare mode as for this investigation yields vertical velocity w with a time resolution of 1 s from about 375 m above ground level (AGLa.g.l.) to the top of the boundary layer and partly above, depending on the aerosol concentration as well as on the measurement setup. Technically, a higher data rate of 10Hz Hz would be possible, but a temporal resolution of 1 Hz is considered the optimal setting for the vertical stare mode, as it ensures higher signal-to-noise ratios by longer averaging. The effective range-gate resolution is about 60 m (Träumner et al., 2011). The measurements are mainly restricted to the cloud-free atmosphere, because lidars only partly penetrate the radiation emitted by the lidar is attenuated within clouds. In order to cover the

range between the top of the surface layer and the lowest measurement heights of WTX, a Doppler lidar (WLS7-V2, Leosphere, hereafter called WLS7) with a wavelength of 1.5 µm was used. ~~Applying the VAD mode yields the wind profile from~~ This instrument is capable to measure radial velocity at distances between 40 ~~m AGL up to about~~ and 400~~m AGL with a temporal resolution of 1.6 s to 10 min and a vertical~~ m with a range resolution of 20 m. As for WTX, operation of the system in the vertical stare mode allows for the direct detection of vertical velocity. In combination with the WindTracer WTX at Hambach, a full vertical coverage of vertical velocity from the top of the surface layer up into the entrainment zone results.

Two Doppler lidars (a~~2~~ 2 µm lidar called WindTracer "HYB" with a Tm:LuAG laser / Lockheed Martin CT, and WLS200 / Leosphere) were operated at ~~the~~ Wasserwerk. Apart from the different laser transmitters, the HYB has similar system settings as the WTX. The Doppler lidar at Selhausen, the third site, was a Stream Line manufactured by HALO Photonics Ltd. (Pearson et al., 2009, hereafter called HALO), which measures with a range-gate length of 18 m (Eder et al., 2015). In contrast to the WindTracer systems having a laser pulse of high energy, the HALO and the WLS200 operate in a "low-pulse energy / high-pulse rate mode" and they can resolve the lowest hundreds of meters ~~AGL~~ a.g.l.. An overview of the lidar instruments at the different locations is also given in ~~Tab.~~ Table 1. The variability of the threshold of signal-to-noise ratio taken for filtering noisy data for the different instruments is also related to the different technical specifications. The measurement frequency of 1~~Hz~~ Hz was the same for all Doppler lidars and the measurement settings were chosen such that vertical velocity data were available at intervals of 25 m for the WindTracer systems as well as for the systems from Leosphere. For the WindTracer systems, this setup causes an overlap of the effective range gates. The data of HALO were interpolated to the same heights.

As all heights used in this study will be in m~~AGL~~ a.g.l., we will omit the adjunct "~~AGL~~a.g.l." in the following sections.

2.1.2 Energy balance stations

The energy balance stations measure solar and reflected irradiance, long-wave incoming and outgoing radiation, soil heat, sensible heat, latent heat, and momentum fluxes. For the turbulent fluxes, temperature, humidity, and wind speed are measured with an ultrasonic anemometer/thermometer and a fast infrared hygrometer at a height of 4 m. All turbulent fluxes used in this study were calculated for time intervals of 30~~minutes~~ min using the eddy-covariance software package TK3.11 of Mauder and Foken (2011) and Mauder et al. (2013). Altogether, data of five energy balance stations were used: ~~Two~~ two energy balance stations of KITcube that were co-located with the lidar instruments at Hambach and at the Wasserwerk site and three TERENO stations at Niederzier, Selhausen, and Ruraue (Fig.~~1~~ 1b).

2.1.3 Additional instruments at Hambach

To obtain vertical profiles of temperature, humidity, wind speed, and wind direction, the KITcube radiosonde system (DFM-09, Graw) was operated at Hambach. On 18 days selected as intensive operation periods (IOPs), radiosondes were launched ~~at 2-hourly intervals~~every two hours. On all other days, launches were done at least at 11~~UTC~~ :00 UTC and 23~~UTC~~:00 UTC. A microwave radiometer (HATPRO, Radiometer Physics GmbH) was also operated at Hambach. The instrument detects thermal radiation emitted by atmospheric components. From these data, for example time series of integrated water vapor (IWV) can be derived with high accuracy (Pospichal and Crewell, 2007). An additional ultrasonic anemometer was installed on a mobile tower and measured wind components and virtual temperature at a height of 30 m. Finally, a ceilometer (CHM 15k, Jenoptic) measured cloud-base heights.

2.2 ~~Turbulent surface fluxes~~

~~An overview of the daily averaged Bowen ratios (ratio of daily averaged sensible heat flux to daily averaged latent heat flux, using 09–15 UTC) indicates that the values were very high (up to 4) for some stations until 6 May 2013, but below one at all stations after that date (Fig. 2a). The Bowen ratio was below one at Selhausen and Ruraue during all the time so that spatial heterogeneity within the respective area of about 5 x 5 km^2 existed in April until early May. The rain gauge measurements at the Wasserwerk (Fig. 2b) reveal that there was much less rainfall during this period than after 6 May. From the land-surface point of view, the whole measurement period may be divided into a drier period with considerable spatial heterogeneity and a wetter period with less heterogeneity. Similar differences of Bowen ratio between a wet and a dry period were found during the field experiment LITFASS-2003, which also took place in an area dominated by agricultural land use (Beyrich and Mengelkamp, 2006) . In order to derive spatially representative values of sensible heat flux, an average of flux measurements was calculated by weighting each station with the fraction of the respective land-use class in the considered area (Fig. 1). The land-use map was available at 15 x 15 m^2 horizontal resolution and the land-use classes were combined to the following three categories: Bare soil, crops, and meadow / forest, with fractions of 70.2, 22.8, and 7.1, respectively. As the growth of vegetation was not yet advanced in spring 2013, the fluxes of Niederzier were considered to be representative of bare soil, even though the station was located in a field of sugar beets. The stations Selhausen and Wasserwerk (crops) were assigned to the second land-use class and Ruraue and Hambach (meadow) to the third one. As the fraction of bare soil was by far the highest and the Bowen ratio at Niederzier was high during the drier period, the weighted-averaged flux also is higher during the drier than during the wetter period (Fig. 2a).~~

2.2 Selected days

On six days with mainly cloud-free CBL conditions, at least one lidar at each site was configured for w-measurements: 18, 20, 22, and 24 April as well as 04 and 19~~May.~~ May. All of these days, apart from 22 April, were also IOP days. Here, the variance profiles for ~~all of these~~ the six days were analyzed. From the radiosoundings, mean CBL conditions were estimated ~~for these six days (Tab.~~ (included in Table 2): On four of the six days, the main regime was governed by westerly to southwesterly flow. On 20 April, the mean wind direction was from northeast and on 19 May, it varied between northeast to north in the CBL, while it was from the east directly above the CBL. Wind speed was low on 22 April (4 m s^{-1}), high on 18 April (12 m s^{-1}), and moderate on the remaining days. The Obukhov length L (Monin and Obukhov, 1954) was calculated from averaged values of available energy balance measurements of kinematic sensible heat flux at the surface and friction velocity. As expected in the CBL, L was negative for all days. According to mean wind speeds, its absolute value was highest on 18 April and lowest on 22 April, indicating that turbulence production by wind shear may have been more important on 18 April than on the other days.

As indicated by microwave radiometer measurements, the IWV was moderately high on most days and much higher on 24 April. Incoming shortwave radiation, as measured by a pyranometer network operated by TROPOS (Leipzig), naturally increased from 18 April to 19 May. At the same time, the spatial standard deviation of incoming radiation, in combination with ceilometer data and cloud camera images, revealed the existence of some CBL clouds on 18~~April, of cirrus clouds at about 8 km on 24 April,~~ April and of altocumulus clouds at about 5 km on 19~~May.~~ May. Cirrus clouds at about 8 km existed on 24 April, but they did not affect incoming radiation. In comparison to the other days, ~~however,~~ the maximum sensible heat flux was reduced on 19~~May because this was the only day falling into the wetter period~~ May. The height of the capping inversion of the CBL was also lowest on 19 May, while it was highest on 18~~April.~~ April. As indicated by the temporal evolution of temperature profiles of the radiosondes for 18 April (not shown), a ~~neutrally stratified residual layer was present between 800 m and 1400 m at 09 UTC above the mixed layer.~~ When the boundary layer grew into this neutral layer, its height increased abruptly from 700 m at 09~~UTC~~ :00 UTC to 1600 m at 11~~UTC~~:00 UTC. This also may have contributed to the formation of some boundary-layer clouds on this day as the sudden mixing throughout the deepened CBL led to a cooling of the former residual layer.

2.3 Turbulent surface fluxes

An overview of the daily averaged Bowen ratios (ratio of sensible heat flux to latent heat flux, both averaged over 09:00–15:00 UTC) indicates that the values were very high (up to 4) for some stations until 6 May 2013, but below one at all stations after that date (Fig. 2a). Accordingly, values of daily averaged sensible heat flux were highest (up to 220 W m^{-2}) until 6 May (Fig. 2b).

The Bowen ratio was below one at Selhausen and Ruraue during all the time so that spatial heterogeneity within the respective area of about $5\,\mathrm{km} \times 5\,\mathrm{km}$ existed in April until early May. The rain gauge measurements at Wasserwerk (Fig. 2c) reveal that there was much less rainfall during this period than after 6 May. From the land-surface point of view, the whole measurement period may be divided into a drier period with considerable spatial heterogeneity and a wetter period with less heterogeneity. Similar differences of Bowen ratio between a wet and a dry period were found during the field experiment LITFASS-2003, which also took place in an area dominated by agricultural land use (Beyrich and Mengelkamp, 2006) .

In order to derive spatially representative values of sensible heat flux, an average of flux measurements was calculated by weighting each station with the fraction of the respective land-use class in an area of $30\,\mathrm{km} \times 30\,\mathrm{km}$ (50.7511–51.0209 ° N / 6.2366–6.6654 ° E) with the locations of the lidar instruments in its center. This corresponds to an upstream distance from the measurement sites which the airflow of $4\,\mathrm{m\,s^{-1}}$ passes during one hour. The land-use map was available at a horizontal resolution of $15\,\mathrm{m} \times 15\,\mathrm{m}$. Not for every land-use class, an energy balance station was available, so that the land-use classes were combined to the following three classes: (1) Bare soil / coniferous forest, (2) crops, and (3) meadow / broadleaf forest, with fractions of 31.6%, 50.9%, and 12.3%, respectively. As the growth of the sugar beets at Niederzier was not yet advanced in spring 2013, the fluxes were considered to be representative of bare soil (class 1), even though the station was located in a field of sugar beets. The stations Selhausen and Wasserwerk (crops) were assigned to class (2) and Ruraue and Hambach (meadow) to class (3). The weighted-averaged flux as well as the Bowen ratio are shown in Fig. 2.

3 Vertical velocity measurements and variance calculations

3.1 Characteristics of vertical velocity data

As an example, vertical velocity measurements from ~~11–13 UTC~~ 11:00–13:00 UTC on 20 April at the three sites are shown for comparison (Fig. 3). Up- and downdrafts with a maximum vertical velocity of more than 2.5 $\mathrm{m\,s^{-1}}$, which are typical of ~~convective boundary layers~~CBLs, were observed at all sites. The thermals lasted for several minutes and rose up to 1200 m during this time period. The isolines of potential temperature of 283–285 ~~K of potential temperature~~ K (Fig. 3b) display the height of the inversion layer at 1200 m, which also agrees with the measurement heights of the lidars WTX and HALO (~~Figs.~~Fig. 3b and c). The HYB yielded measurements up to 1500 m (Fig. 3a)~~. This is due to the averaging of a higher number of laser pulses so that the signal-to-noise ratio is still above the selected threshold (Tab. 1) at heights~~ , i.e. also above the inversion layer, where the aerosol concentration ~~is~~ was much lower. This is presumably due an improved performance of HYB after a refurbishment of the laser transceiver shortly before HOPE. It can also be seen that the w-measurements of the WLS7 and WLS200 for the lowest 400 m are ~~subjectively~~ qualitatively

consistent with the measurements above (~~Figs.~~Fig. 3a and b).

For a first analysis of the time series, ~~power density spectra of w were calculated~~ spectra of energy density S were calculated for w at different heights for the ~~instruments~~ lidars at Hambach (Fig. 4a). Additionally, the spectrum of w-measurements by an ultrasonic on a 30-m tower is given and can be compared with those of WLS7 at the lowest range gate (60 m). ~~At low frequencies of about 10^{-3} Hz to 10^{-2}~~ Generally, turbulence spectra are characterized by a peak at a certain frequency or wavelength ($f_{p,w}$ or $\lambda_{p,w}$, respectively), which yields the largest spectral contribution to the variance, and by a slope in the inertial subrange ($fS \propto f^{-2/3}$). According to Kaimal et al. (1976) , $f_{p,w}$ is height dependent for spectra of vertical velocity. This dependency is strongest near the surface and weaker or even disappearing in the CBL. In the selected example, $f_{p,w}$ is about $3 \cdot 10^{-3}$ to $5 \cdot 10^{-3}$ Hz (i.e. time periods ~~of about 2–15 min) , the energy is highest for all range gates except for the lowest one. Maximum turbulent energy for the considered time series can be found at 400 m height, as indicated by the integral spectrum~~ are about 3–6 min) for range gates of 200 m and higher (Fig.~~4b~~ 4a). At the lowest given range gate (60 m) and for the ultrasonic measurement, maximum ~~energy~~ variance is shifted towards higher frequencies (~~0.01 Hz to 0.1~~ 10^{-2} Hz to 10^{-1} Hz) and smaller time periods (10 s to 2 min), respectively. ~~Besides, as discussed by Frehlich et al. (1998) or Brugger et al. (2015) for example, it is obvious that in the inertial subrange, the energy of the lidar spectradecreases faster~~ The different values of $f_{p,w}$ in the frequency range of $3 \cdot 10^{-3}$ Hz to 10^{-1} Hz are also well visible in the integral spectra (Fig. 4b). As the integral of the spectral energy density over all frequencies is equal to the total variance, the integral spectra also illustrate that vertical motions in this frequency range contribute to more than 50% of the total variance. Moreover, they indicate that the maximum total variance for the considered time series can be found at 400 m height.

The inertial subrange can also be discerned in the given spectra, but the slope is steeper than the theoretical one of $-2/3$ ~~-slope, i.e. for frequencies higher than about 0.1 Hz for~~ for the WTX (400 m and above) ~~and about 0.3 Hzfor WLS7 (60 m and 200 m).~~ at frequencies higher than about 0.1 Hz. This effect is also discussed by Frehlich et al. (1998) or Brugger et al. (2015) , for example: Even if the measurement frequencies of 1 Hz of the ~~two lidar systems~~ lidar system would be high enough to register fluctuations ~~on these scales~~of frequencies larger than 0.1 Hz, the sampling frequency is restricted due to the spatial averaging of the lidar pulses. ~~This affects the absolute values of~~ The frequency of 0.1 Hz corresponds to the physical range gate resolution of the instrument ($\Delta r \approx |\boldsymbol{v}| f^{-1}$, with $|\boldsymbol{v}| \approx 8$ m s^{-1} on 20 April). Moreover, Darbieu et al. (2015) also found steeper slopes in spectra derived from aircraft measurements. Lothon et al. (2009) and Darbieu et al. (2015) assume that a steeper slope could also be caused by asymmetric convective structures, i.e. by anisotropy of the w-field. A steeper slope in the inertial subrange also affects the total variance, as can be seen in the integral spectra: ~~While the~~ The contribution to the total ~~energy still~~ variance increases up to ~~the highest frequency resolved by the measurement frequency for the WLS7 spectra, this is not the case~~

~~for the WTX spectra. The total energy or variance, respectively, is therefore higher at 200 m than at 900 m.~~ frequencies of 0.1 Hz only.

Apart from that, the spectra of WLS7 show some artefacts at the highest frequencies, which were also observed by ~~Carbajo Fuertes et al. (2014) , for example~~Cañadillas et al. (2011) . This is presumably the signature of an aliasing effect~~. As the effect of missing variance contributions at high frequencies due to the spatial averaging is found for all lidar systems and as the main aim of our investigation is an intercomparison of lidar measurements at different locations, this effect will be neglected below.~~ , but the reason for this cannot be clarified in detail here, as not all the necessary technical specifications are communicated by the manufacturer.

~~The integral spectra indicate that the largest contributions to the variance (> 50) lie in the frequency range of $2 \cdot 10^{-3}$ Hz to 10^{-2} Hz (Fig. 4b).~~ Based on the spectra, ~~the peak frequency~~ $f_{p,w}$ ~~lying within this range~~ was estimated for all days. From $f_{p,w}$, the time and length scales ($T_{p,w}$ and $\lambda_{p,w}$), on which the turbulent energy contained in the vertical motions is highest, were calculated. They ~~varied~~ vary for the considered days between ~~five and eight minutes~~ 5 and 8 min or 2–2.7 km, respectively (~~Tab.~~ Table 2). The values of $T_{p,w}$ are smaller on days with higher wind speeds, because turbulence elements are advected faster past the location of the measurement. Additionally, scales can be estimated from the autocorrelation function of w ~~: It~~ in the CBL (w at 600 m was chosen here): The autocorrelation becomes negative at a certain time interval and will have a second maximum (and further maxima), if a dominant periodic fluctuation exists. The interval, at which the second maximum can be discerned does then correspond to the repetition frequency of the up- or downward motions. This repetition frequency often corresponds to ~~the peak wavelength~~ $f_{p,w}$. The values are slightly larger than those estimated from the spectra, but they confirm that ~~the peak wavelength is about 3 km (Tab.~~ $\lambda_{p,w}$ is about 2–3 km at the three sites on average (Table 2). This means that the energy-containing length scale of the turbulent motions in the CBL was much larger during HOPE than the length scale of the surface heterogeneity, which is several 100 m at the maximum.

~~Instead of~~ Additional to the calculation via the integrated spectrum, ~~the~~ $\overline{w'^2}$ ~~profiles were~~ was determined directly from the time series. For a validation of both computation methods, the hourly variances for all considered instruments and all six days calculated by both methods were compared for the 600-m range gates and were found to be in good agreement, with a mean relative deviation of 3%.

3.2 Errors considered for variance calculations

As in Träumner et al. (2011), the variances were corrected for uncorrelated random noise using a technique proposed by Lenschow et al. (2000). Additionally, the statistical error was considered as described by Lenschow et al. (1994). This method is based on the separation of the random and the systematic error (App. A). On days with higher wind speed, the integral time scale and, hence, the statistical error is smaller (Table 2). By this, the dependency of sample size on the mean wind speed

is considered implicitly.

Even if the signal noise is considered, we cannot be sure that different instruments can provide identical measurements, especially if they are from different manufacturers and are based on different technical principles like HALO and WLS200 compared to the WindTracer systems. Therefore, both WLS200 and HYB were operated at ~~the~~ Wasserwerk in the vertical stare mode on 20, 22, and 24 April, so that the w-measurements of the two lidar systems could be compared directly. The cross correlation function between the two w time series on 20 April was calculated for measurement heights between 400 m and 1000 m (not shown). The highest correlations (> 0.8) can be found between 600 m and 800 m. As for the autorcorrelation functions, an oscillation between positive and negative values is observed for increasing time lags, symmetrically for positive and negative ones. For 18 and 22 April, the maximum correlations are 0.88 and 0.95, respectively. This means that the two measurements were not perfectly the same on all days, but sufficiently well correlated to possibly yield similar statistics. The variance differences resulting from ~~the different devices~~ different effective range gate lengths as well as single-pulse energies will be taken into account for the spatial comparisons in ~~Sec.~~Sect. 4.3.2.

Finally, another error that may have an influence is the missing ~~contribution to turbulent energy~~ variance contribution in the higher frequency part of the spectrum due to the vertical averaging of the lidar measurements. ~~As explained above, this~~ This error will be neglected here, as it would lead to higher variances at all stations and not change the spatial differences. Moreover, the missing contributions are small compared to the absolute values of variance.

3.3 ~~Scaling of variance profiles~~Scales and scaling parameters

According to Lenschow et al. (1980) and Sorbjan (1989), vertical profiles of $\overline{w'^2}$ can be normalized and best fitted by

$$\frac{\overline{w'^2}}{{w_*}^2} = 1.8\left(\frac{z}{z_i}\right)^{2/3}\left(1-0.8\frac{z}{z_i}\right)^2 \text{ and } \frac{\overline{w'^2}}{{w_*}^2} = 1.17\left(\frac{z}{z_i}\right)^{2/3}\left(1-\frac{z}{z_i}\right)^{2/3}, \quad (1)$$

respectively. The convective velocity scale is defined as

$$w_* = \left(z_i \frac{g}{\overline{\theta_{v,0}}} \overline{w'\theta_v'}\Big|_0\right)^{1/3}, \quad (2)$$

with the CBL height z_i, the gravitational acceleration g, the temporal mean of virtual potential temperature at the surface $\overline{\theta_{v,0}}$, and the kinematic sensible heat flux at the surface, $\overline{w'\theta_v'}\Big|_0$. For the sensible heat flux, ~~either~~ the weighted-averaged heat flux (see ~~Sec.2) or~~ Sect. 2.3) as well as the fluxes measured by the energy balance stations next to the lidar instruments can be used here. To distinguish between both scaling approaches of the variance values, they will be called averaged and local scaling, respectively, in the following investigation.

For z-axis scaling as well as to calculate w_*, the CBL height has to be determined. At least, three different methods are in use, depending on the available measurement systems (cf. Emeis et al.,

2008; Träumner et al., 2011, and references therein): (1) Determining the CBL capping inversion from radiosonde profiles, (2) estimating the top of the aerosol layer from lidar backscatter data, and (3) calculating the top of CBL convection from profiles of the vertical velocity variance. While the first two methods can be regarded as proxies for the CBL depth, the third method is a direct one. Tucker et al. (2009) systematically investigated the determination of the ~~CBL height~~ z_i using variance profiles and found that a threshold value to which the variance decreases was the best objective criterion. Träumner et al. (2011) determined this threshold value for the HYB for several field campaings and found that a value of $0.16\ \mathrm{m^2\,s^{-2}}$ gave the best results.

For the six days investigated here, the methods agree well for most time steps around noon (dashed lines and black dots in Fig. 5). Mainly before 1100 and after 15:00 UTC, method (3) yields lower values of z_i than method (2). The reason is that especially method (2) tends to detect the cap of the residual layer, which is not the case for method (3). However, the threshold value ~~was~~ of method (3) is not applicable to all of the profiles here. ~~With this in mind and as the~~ For several time steps, the decrease of variance with height is weak and the variance does not reach the defined threshold, so that z_i ~~values derived by methods~~ cannot be determined by method (3). In contrast to method (1~~) and 2) showed a good agreement (dashed lines and black dots in Fig. 5~~), method (2) ~~was chosen, because it~~ also provides values for periods when no radiosoundings are available. Therefore, z_i values derived by method (2) are used for the following calculations. Correlating all z_i values from method (1) with values derived by method (2) from different lidars shows that z_i values derived from backscatter data of WTX at Hambach fit best.

The values of w_* resulting from using z_i determined by method (2) and the weighted-averaged fluxes are also given in Fig.~~5 (grey~~ 5 (gray lines). A comparison of diurnal maximum values of $\overline{w'^2}$ and w_* is included in ~~Tab.~~ Table 2. From w_*, a convective time scale $t_* = z_i \,/\, w_*$ can be derived that describes how long it takes to transport an air parcel from the ground to the top of the CBL. Therefore, t_* is also known as large-eddy turnover time. Comparing t_* and T (Table 2), it is obvious that the large-eddy turnover time is on all days larger than the energy-containing time scale of the turbulence elements, T, which depends on their advection past the location of measurement (Sect. 3.1). This means that the turbulence elements do not change substantially during the time it takes them to pass the lidar.

4 Spatial and temporal differences of vertical velocity variances

4.1 Profiles of variance and skewness: examples for 20 April

Examples of profiles of w-variance calculated for four instruments at the three locations are shown in Fig. 6. The given times always indicate the end of the averaging period of one hour. As described by Deardorff (1974) or Lenschow et al. (1980), the variance profiles display a maximum at a height of about one third of the convective boundary layer (the top of the CBL is between 1000 m and

1400 m on 20 April, Fig. 5) and a decrease above. The profiles in Fig. 6 are not normalized so that the diurnal evolution ~~may~~ can be observed: Variances are small at 10~~UTC~~ :00 UTC (12~~LT~~:00 LT), increase to maximum values at about ~~12–14 UTC~~ 12:00–14:00 UTC and decrease subsequently. Above a local minimum indicating the top of the CBL, an increase of variance can be seen in several profiles (e.g. ~~13–16 UTC~~ 13:00–16:00 UTC profiles of HYB at about 1500 m, Fig. 6a). These higher values lie in and above the capping inversion of the CBL (Fig. 3b) and ~~can presumably be attributed to the existence of gravity waves there~~may be caused by gravity waves in the capping inversion and a stable layer above the CBL.

As already shown by the comparison of vertical velocity measurements of the smaller WLS7 and of WTX (Fig. 3), the combined variance profiles fit well at the transition height from one instrument to the other (Fig. 6d). The maximum variance is sometimes located at low heights that are not covered by HYB or WTX (for example, at 11~~UTC~~ :00 UTC in Fig. 6c), indicating the usefulness of the combination of different lidar systems with complementary ranges. The variance profiles derived from the measurements of HYB and WLS200 (~~Figs.~~Fig. 6a and b) do not agree in all details, as indicated by the calculated cross correlations, but the profiles are much more similar to each other than to the profiles from the other two sites in ~~termes~~ terms of structure, temporal evolution, and absolute values.

Additionally, profiles of skewness $\left(\overline{w'^3} \,/\, \overline{w'^2}^{3/2}\right)$ ~~were~~ are analyzed (Fig. 6). Positive skewness is usually expected in the CBL and means strong, narrow updrafts and weaker, broader downdrafts. On 20 April, values of skewness are positive within the CBL. They confirm the existence of a well-mixed boundary layer, as they illustrate a net upward transport of variance (according to the variance budget equation of Stull, 1988) and with this, of turbulent energy. This means that the turbulent energy is mainly created at the surface, i.e. by buoyancy.

4.2 Scaling of variance profiles

4.2.1 Overview of all scaled variance profiles

Diurnal variability of w-variance is obvious on 20 April (Fig. 6). This temporal variability should be eliminated by scaling with w_*, assuming that the temporal variability of the w-variance depends mainly on the strength of buoyancy. It is expected that the scaled profiles are similar within the range of uncertainty indicated by the statistical error. Differences of the Bowen ratio point to a large spatial heterogeneity (Sect. 2.3). Hence, at an individual location, the diurnal cycle of the energy input as well as differences from day to day may be taken into account better by local scaling than by the averaged one (see Sect. 3.3 for the definition of the scaling approaches). Therefore, also the question is addressed whether the spread of the profiles at each individual location is smaller for the locally scaled profiles. On 19 May, which is the only day falling into the wetter period with less surface heterogeneity, lower Bowen ratio and consequently, lower w_* is observed at all stations (see

Sect. 2.3, Figs. 2 and 5). This day is excluded from the analysis of the scaled profiles.

There were two energy balance stations were located near Selhausen: The energy balance station of Niederzier was about 1 km north of Selhausen which may be relatively far away, but the land-use class was the same as at the lidar location. The station called SE1 was closer, but the land-use class there differed and the flux was very low, even lower than at Ruraue (Fig. 2b), which was located in a meadow close to a river. Both are used for local scaling of the variance profiles from Selhausen. As Niederzier is a bare-soil station with relatively high sensible heat fluxes (Fig. 2b), i.e. a high Bowen ratio, and SE1 is characterized by a low Bowen ratio, large differences are found between the two normalizations: The maximum values of mean normalized variance are 0.32 and 0.79, respectively (Fig. 7g and h). For the averaged scaling, by contrast, the maximum value of the mean scaled variance at Selhausen is 0.42 (Fig. 7c), which is closer to the mean values of $\overline{w'^2}/w_*^2$ at Hambach and Wasserwerk (0.45 and 0.46, respectively, Fig. 7a and b). This means that in comparison to the scaled variances at the other locations, the surface sensible heat flux at Niederzier is too high and SE1 too low with respect to the observed CBL turbulence at Selhausen. The mean variance profiles at all locations display a vertical behavior that is similar to the profile of Lenschow et al. (1980, Fig. 7d) , with a maximum in the lower half of the CBL, but not exactly at $0.35\,z_i$. The difference between standard deviation of all profiles and the mean normalized statistical error signifies their temporal variability which is not explained by variability of buoyancy. At Hambach and Selhausen, the standard deviation is higher than the statistical error at all heights, most distinctly between 0.2 and $0.6\,z_i$. The mean relative differences between error and standard deviation, vertically averaged, lie between 5% (Fig. 7f) and 36% (Fig. 7h). At Wasserwerk, the difference is small, especially for local scaling (Fig. 7f). This indicates either that turbulence at Wasserwerk is strongly influenced by nearby surface conditions or that the nearby surface conditions represent the larger-scale upstream conditions very well.

In a similar investigation, Lenschow et al. (2000) found a difference of 10% between error and standard deviation. They explained it by dependency on wind shear or stability, represented by $-z_i/L$. However, a dependency of w-variance on $-z_i/L$ cannot be found here, neither on friction velocity nor on values of wind shear at the CBL top, as derived from radiosoundings.

4.2.2 Correlation of variance and convective velocity scale

In a next step, correlation coefficients are determined between the w-variance values averaged between 0.25 and $0.60\,z_i$ ($\overline{w'^2}_{ave}$) and w_*^2. As in Sect. 4.2.1, values of w_* for both averaged (1) and local scaling (2) are applied. By vertical averaging of w-variances, the height dependency of the maximum is eliminated. In case (1), the squared correlation coefficient R^2 is 0.45 for Hambach and 0.50 for Wasserwerk; in case (2), the correlation is slightly higher than in case (1) for Hambach ($R^2 = 0.49$) and considerably higher for Wasserwerk ($R^2 = 0.72$). For Selhausen, R^2 is 0.46 in case (1) and lower in case (2) when using the fluxes from Selhausen or Niederzier ($R^2 = 0.28$ or 0.34,

respectively). This means that the local scaling is not preferable for Selhausen. For Hambach, local scaling is only slightly better than averaged scaling, but local scaling is clearly better for Wasserwerk. For the given sample sizes, the correlations are all significantly higher than zero when considering a confidence interval on a 95% level. However, only for Wasserwerk using local scaling, the explained variance (concerning the temporal evolution of $\overline{w'^2}_{ave}$, hereafter called "temporal variance" to avoid ambiguity) is significantly higher than 50%. In contrast, for Selhausen using local scaling with SE1, the explained temporal variance is not significantly higher than 10%, indicating that this scaling is not suitable.

Deardorff (1970b) and Deardorff (1974) showed that $\overline{w'^2}(0.35z_i) = aw_*^2$ and found values of a between 0.37 and 0.44, derived from both numerical experiments and different observations. Here, R^2 is 0.34–0.39 for the averaged w_*^2 values and 0.30–0.43 for the local ones, which agrees tolerably well with values found before. For Wasserwerk and the local scaling, a is 0.43, i.e. at the upper limit of values given in literature.

The implication of the correlations found here is that it is hard to find the specific site in a region with heterogeneous surface fluxes which represents the whole upstream conditions relevant for the turbulence in the CBL. Therefore, it is preferable to apply a weighted-averaged flux for scaling. A possible explanation why the correlation for local scaling (Wasserwerk) is higher than for averaged scaling is the uncertainty of the spatial averaging procedure and with this, of averaged scaling, due to the combination of different land-use classes as well as the choice of the considered area (Sect. 2.3).

4.2.3 Investigation of outliers

The findings show that temporal variability of w-variance cannot be completely eliminated by scaling and that the remaining variability cannot be explained by wind shear or stability. Therefore, individual profiles with particularly high values of $\overline{w'^2}/w_*^2$ are examined in detail. The largest outliers from Wasserwerk, which has the smallest portion of unexplained temporal variance, are selected (Fig. 7b and f, respectively). They occur at 12:00 UTC on 20 and 24 April. Each of the two profiles is compared to a profile from the respective day which is more similar to the mean (Fig. 7). Radiosone profiles indicate no strong diurnal change in wind speed or direction on these two days (not shown). The comparison, including error bars, indicates that $\overline{w'^2}/w_*^2$ is significantly higher for the selected time periods than usual (Fig. 9ai and bi). If longer time periods are chosen, differences decrease, but the statistical error decreases likewise so that they are still significant.

A hypothesis for high values of $\overline{w'^2}/w_*^2$ is the occurrence of more numerous or stronger thermals. Lenschow and Stephens (1980) developed a method for a sub-sampling of thermals from the time series of w and Lenschow and Stephens (1982) showed that the variance of thermals is 2–2.5 times higher than for the environment, depending on the method of calculation (the ratio is higher when the mean velocity of the sub-samples is subtracted before calculating the variance). As a sub-sampling

would be beyond the scope of this investigation, the frequency distributions of the respective time series are investigated (Fig. 9). As variance is equal to the second central moment of a probability distribution, larger variance signifies a broader and flatter distribution by definition. The frequency distribution for 20 April, 11:00–12:00 UTC reveals that there is a higher frequency of updrafts than between 14:00–15:00 UTC as well as stronger downdrafts (Fig. 9aiv). When the frequency distribution is considered as a function of height (Fig. 9aii), it can be shown that this behavior can be observed between 200 and 900 m, i.e. distributed over a large part of the CBL (z_i is between 1300 and 1400 m on this day). On 24 April, the maximum of $\overline{w'^2}/w_*^2$ at 12:00 UTC is elevated compared to the one at 10:00 UTC (Fig. 9bi), while z_i is the same (about 1350 m) for both periods. In contrast to 20 April, higher variance is caused by a higher frequency of strong updrafts only, not by stronger downdrafts (Fig. 9biv). Moreover, the differences between the frequency distributions occur mainly at heights between 400 and 800 m, i.e. they are vertically more confined to the layer where $\overline{w'^2}/w_*^2$ is actually higher.

Thus, while high values of normalized variance for the profile at 15:00 UTC on 20 April are caused by strong up- and downdrafts, they are actually caused by broader and / or more numerous thermals on 24 April. This agrees with the results of Lenschow and Stephens (1982) that the variance of thermals is higher. However, it is not possible to explain these thermals by corresponding higher surface sensible heat fluxes and, thus, why $\overline{w'^2}/w_*^2$ is higher than on average.

Due to the elevated maximum, the profile for 12:00 UTC on 24 April corresponds better to the symmetrical profile of Sorbjan (1989, Fig. 7d) . Caughey and Palmer (1979) , e.g., also discuss the variability of heights of the variance maxima reported by different authors. The height-dependent frequency distribution shown here suggests that the elevated maximum is caused by strong thermals rising up to a certain height. LES of van Heerwaarden et al. (2014) also support the finding that an elevated maximum of variance is related to particularly strong plumes.

4.3 Spatial ~~comparison~~ differences of vertical velocity variances

The main finding of the investigation of scaled profiles is that averaged scaling was preferable, i.e. that the same scaling could be used for the three locations. This implies that also the absolute values of variance should be similar at the three locations. However, unexplained temporal variance is found even for the "best" scaling. The question is now if there is also a spatial variability of w-variance.

One noticeable difference between the hourly variance profiles at the three locations on 20 April (Fig. 6) is the diurnal cycle: While maximum variance occurs at 12~~UTC~~ :00 UTC at Wasserwerk and Selhausen, it ~~occurred~~ occurs at 14~~UTC~~ :00 UTC at Hambach. ~~The question is whether this is statistically significant or not.~~ To investigate this spatial difference, the height of maximum variance, z_{max}, ~~was~~ is determined for all days and all hourly variance profiles. It ~~was~~ is encountered between 0.1 z_i and 0.5 z_i. A maximum variance $\overline{w'^2}_{max}$ ~~was~~ is then calculated by vertical averaging of each profile over a height range of $z_{max} \pm 250$ m. The statistical errors ~~were~~ are determined for the same

height range. The time series of $\overline{w'^2}_{max}$ for the three locations are shown in Fig. 10. The difference of $\overline{w'^2}_{max}$ between Wasserwerk and Hambach on 20 April for the ~~12 UTC~~ 12:00 UTC period is not significant when considering the statistical error~~. However, for some~~ , but it is significant for the 14:00 UTC period. For other time periods, as for example for 11~~UTC and 15–16 UTC~~ :00, 15:00, and 16:00 UTC on 18 April, 11~~UTC and 14 UTC on 20 April or~~ :00 UTC and 12~~UTC~~ :00 UTC on 24 April, significant differences between the individual locations ~~can be discerned. For 20 April, 13–14 UTC for example, differences are also obvious in the time series of vertical velocity (Fig. ??): At Hambach, where the variance is the highest, about 5–6 periods with convective cells can be distinguished. As the peak energy resides at the lowest frequencies (Fig. 4), as it is associated with the largest turbulence elements, the high variance at Hambach for this hour is attributable to the multiple occurrence of convective cells. At Wasserwerk, the variance is slightly lower than at Hambach because less convective cells passed the site. At Selhausen, the variance is smallest and the least convective cells occurred. Obviously, the spatial variance differences are attributable to the different numbers of convective cells at the three sites that are only about 3 km apart. For the three selected periods with significant differences , we will investigate now whether these differences can be explained by surface conditions~~are also evident. In the following sections, different reasons that could cause significant differences are explored.

4.4 ~~Influence of surface conditions~~

4.3.1 Influence of the surface energy balance

~~Spatial differences of the state of the CBL may be caused by spatially heterogeneous surface conditions.~~ For the days investigated here, positive values of skewness confirm that the strength of CBL turbulence is dominated by surface-based buoyancy-driven convection (examplarily shown for 20 April in Fig. 6). ~~The spatial heterogeneity of the buoyancy flux at the surface may be considered by scaling the variance profiles with w_*^2.~~ Therefore, it is investigated now whether the detected spatial differences of w-variance are related to the spatial heterogeneity at the land surface which was described in Sect. 2.3. Even if local scaling could not eliminate spatial differences on average, it could reduce them for the time periods with significant spatial differences.

Generally, surface heterogeneity as observed during the drier period (Fig. 2) may be caused by heterogeneous surface characteristics such as land use and soil moisture, which influence the partitioning of available energy into sensible and latent heat. On the other hand, heterogeneity also can result from the available energy itself, which can be modified strongly by the occurrence of clouds. As shown in ~~Sec.~~Sect. 2.2, clouds actually ~~occurred on three~~ influenced incoming radiation on two of the six selected days.

The spatial heterogeneity of the buoyancy flux at the surface, including the influence of spatially

heterogeneous cloud cover, may be considered by scaling the variance profiles with w_*^2 (local scaling). For Selhausen, Niederzier is chosen as it provides better correlations than SE1 (Fig. 8).

4.3.2 ~~Scaled profiles: selected periods~~

For the three selected time periods on 18, 20, and 24 April when spatial differences were observed, scaled profiles with the corresponding error bars are given in Fig.~~11. For each site, the surface sensible heat flux from a nearby energy balance station was used for calculating w_* and, thus, for scaling. Two energy balance stations were located near Selhausen: The energy balance station of Niederzier was about 1 km north of Selhausen which may be relatively far away, but the land-use class was the same as at the lidar location. The station called SE1 was closer, but the land-use class there differed and the flux was very low, even lower than at Ruraue (Figs. 2 and 12), which was located in a meadow close to a river. Hence, Niederzier was chosen for scaling the variances of Selhausen.~~ 11.

For all time periods, at least two profiles still show statistically significant differences after applying the ~~w_*-scaling~~local scaling. For 18 April, 15~~UTC~~:00 UTC (Fig. 11a), the difference between Hambach and Wasserwerk becomes even ~~more obvious~~ stronger than without scaling. This means that the spatial differences cannot be explained by the surface heterogeneity. The reason becomes obvious when looking at the net radiation and surface sensible heat flux for the three selected time periods (Fig. 12):

On 18 April at 15~~UTC~~:00 UTC, the w-variance is the highest at Selhausen and lower at Hambach as well as at Wasserwerk (Fig. 10). If local sensible heat fluxes were responsible for the spatial differences of ~~the turbulence between 14–15 UTC~~CBL turbulence between 14:00–15:00 UTC, the spatial flux differences would be similar. However, the flux is highest at Hambach (Fig. 12) so that the scaled variance was the lowest. At Niederzier, the flux is slightly lower and much lower at Wasserwerk. Consequently, the differences of the sensible heat flux ~~can not~~ cannot explain the variance differences. Moreover, net radiation (Fig. 12) shows that some clouds ~~occur~~ occurred on this day and from cloud camera images, it is known that also boundary-layer clouds ~~are present~~ were present between 13:00 and 15:00 UTC. These clouds do not cause ~~considerably~~ considerable temporal variation in the sensible heat flux data, but they can certainly influence the variance profiles (e.g. Neggers et al., 2003).

On 20 April, 14~~UTC (Fig. 11b)~~:00 UTC, the variance is highest at Hambach and lower at Wasserwerk ~~as well as at~~ and Selhausen (Fig. 10). However, the surface sensible heat flux is equally high at the three locations ~~.~~ (Fig. 12). At the same time, the net radiation shows little spatial variability (< 20 W m^{-2})~~at this time~~. Thus, the surface forcing does not display large differences between the three locations, which explains why a scaling using the fluxes from the nearby stations does not remove the spatial differences of variances ~~.~~ (Fig. 11b).

On 24 April, 12~~UTC (Fig. 11c)~~:00 UTC, the variance at Selhausen is significantly lower than at

Hambach and Wasserwerk (Fig. 10) but again, the spatial differences between the fluxes ~~can not~~ cannot explain this difference ~~.~~ (Fig. 12). The flux is highest at Niederzier ~~(Fig. 12)~~ so that the scaled variance profile for Selhausen becomes very low compared to the scaled profiles at the other two locations ~~.~~ (Fig. 11c).

~~As regards the selected examples~~Therefore, it must be concluded that the heterogeneous surface conditions ~~can not~~ cannot explain the statistically significant spatial differences of the w-variances. This is consistent with the finding from Sect. 4.2.3 that significantly increased values of the w-variance within the diurnal cycle cannot be eliminated by scaling, either.

4.3.2 ~~Scaled profiles: whole data set~~

~~To investigate the impact of the scaling on all available profiles in a systematic way, all profiles were normalized by both averaged scaling (using weighted-averaged fluxes for w_*) and local scaling (using fluxes of nearby stations as in Sec. **??**). It is assumed that the local diurnal cycle of the energy input as well as local differences from day to day can be taken into account better by local scaling than by the averaged one. Therefore, also the question is addressed whether the spread between the profiles at each individual location is smaller for the locally scaled profiles. 19 May, which is the only day falling into the wetter period with less surface heterogeneity (see Sec. 2.3), is excluded from the following analysis. In Sec. **??**, Niederzier had been chosen before for scaling because of the land-use class, but proved not to be completely suitable. A different land-use class is given at SE1, but the station is closer. Consequently, both were used for scaling the variance profiles from Selhausen (Figs.7g and h). As Niederzier is a bare-soil station with relatively high sensible heat fluxes (Fig. 12), i.e. a high Bowen ratio, and SE1 is characterized by a low Bowen ratio, large differences are found between the two normalizations: The maximum values of mean normalized variance are 0.31 and 0.70, respectively. For the averaged scaling, by contrast, the maximum value of the mean scaled variance at Selhausen is 0.35 (Fig. 7c), which is more similar to the mean values of $\overline{w'^2}/w_*^2$ at Hambach and Wasserwerk (0.38 and 0.37, respectively, Figs. 7a and b). This means that in comparison to the scaled variances at the other locations, the surface sensible heat flux at Niederzier is too high for scaling the variances from Selhausen and SE1 too low with respect to the observed CBL turbulence. The mean profiles using averaged scaling (thick lines in Figs. 7a–c) display a vertical behavior that is similar to the profile of Lenschow et al. (1980, Fig. 7d) , but with a lower maximum ($\overline{w'^2}/w_*^2 < 0.4$). The similarity of these three mean profiles implies that also the mean profiles of w-variances (without scaling) are similar at the three locations. For WTX at Hambach (Figs. 7a and e), the difference between averaged and local scaling is very small for both mean values (mean maximum of $\overline{w'^2}/w_*^2 = 0.39$ for local scaling) as well as the scatter of the profiles. This means that the energy balance station at this site provides values which are representative of the considered domain. For HYB (Figs. 7b and f), the locally scaled profiles exhibit a smaller scatter than those generated by averaged scaling. To investigate the dependence~~

of the spread between the profiles on the scaling method, correlation coefficients were determined: 1) Between the w-variance values at $0.35\,z_i$ ($\overline{w'^2}_{0.35}$) and w_*^2, calculated with weighted-averaged fluxes (averaged scaling, Fig. 8a), and 2) between $\overline{w'^2}_{0.35}$ and w_*^2, calculated with the respective fluxes used for local scaling (Fig. 8b). In case of 1), the squared correlation coefficient R^2 is 0.25 for Hambach and 0.32 for Wasserwerk; in case of 2), the correlation is slightly higher than in case 1) for Hambach ($R^2 = 0.28$) and strongly for Wasserwerk ($R^2 = 0.58$). For Selhausen, the squared correlation coefficient is higher in case 1) than in case 2) when using the fluxes from Niederzier or Selhausen (0.21 or 0.29, respectively, compared to 0.33 for case 1). This means that the local scaling is not preferable for Selhausen when using either of the available energy balance measurements. For Hambach, local scaling is only slightly better than averaged scaling, but local scaling is clearly better for Wasserwerk. The correlations are all significant according to a t-test on a 1-level, except for $R^2 = 0.21$ (Fig. 8b). However, the values also indicate that the explained variances (concerning the temporal evolution of the $\overline{w'^2}_{0.35}$, hereafter called "temporal variance" to avoid ambiguity) are about 30in all cases but in one. Deardorff (1970b) and Deardorff (1974) showed that $\overline{w'^2}_{0.35} = \alpha w_*^2$ and found values of α between 0.37 and 0.44 which were derived from both numerical experiments and different observations. Here, α is 0.32–0.44 for the averaged w_*^2 values and 0.36–0.47 for the local ones. This variation is not negligible and, in combination with high portions of unexplained temporal variances, it implies that either the intensity of turbulence in the CBL also depends on parameters other than the heat supply at the Earth's surface or that the uncertainty of the calculated scaling parameters is too large. Spatial variability of z_i may be larger than assumed so that the values of w_* using z_i derived from variance measurements at Hambach are not valid for Wasserwerk and Selhausen. However, the normalized variance profiles mainly display a minimum at $z/z_i = 1$ (Fig. 7). For HALO, the profiles break off at $z/z_i \approx 0.95$, which may indicate a lower z_i at the Selhausen. The value of $\overline{w'^2}/w_*^2$ would be less than 4higher in this case, i.e. the uncertainty in z_i does not explain the variability of α and the temporal variance. Besides, some of the profiles with particularly high values of $\overline{w'^2}/w_*^2$ display a maximum at a height which is considerably above the average one at about $0.35\,z_i$. They are more similar to the profile of Sorbjan (1989, Fig. 7d) . Caughey and Palmer (1979) also discuss the variability of heights of the variance maxima given by different authors. One assumption is that this is caused by strong thermals rising up to a certain height. Lenschow and Stephens (1980) developed a method for a sub-sampling of thermals from the time series of w and Lenschow and Stephens (1982) showed that the variance of thermals is 2–2.5 times higher than for the environment, depending on the method of calculation (the ratio is higher when the mean velocity of the sub-samples is subtracted before calculating the variance). Inspection of the time series of w for periods corresponding to the profiles with elevated w-variance maxima (Fig. 7) reveals that these often contain strong convective cells. LES of van Heerwaarden et al. (2014) also support the finding that an elevated maximum of variance is related to particularly strong plumes. Hence, it is concluded that local scaling, i. e. using

~~the surface sensible heat flux from a single nearby station for the calculation of w_*, can lead to errors, especially when small-scale heterogeneity of the surface fluxes exists. This is reflected here by the large difference of the Bowen ratio between the two energy balance measurements of SE1 and Niederzier, which are less than 1.5 km apart. On the other hand the local scaling for one station (Wasserwerk) results in a much higher correlation than all other combinations. This means that it is possible for a single station to provide fluxes that are representative of the area influencing the CBL turbulence. Nevertheless, it can generally be assumed that the radius of influence and, thus, the area of representative w_* upstream of the measurements is several kilometers, depending on the mean wind and the convective time scale. When multiple energy balance measurements cannot be used, the representativeness of a single flux measurement site for scaling should be considered very carefully.~~

4.4 ~~Influence of averaging periods and measurement uncertainties~~

4.3.2 Influence of averaging periods and measurement uncertainties

The variance profiles considered so far were determined using hourly averaging periods. We now want to investigate how strongly the spatial differences are dependent on the length of the applied averaging periods. For this reason, the differences between $\overline{w'^2}_{max}$ values at different locations ~~were~~ are calculated for different averaging periods Δt. For the computation of variances for $\Delta t > 1$ h, the non-stationarity of the CBL, especially due to increasing z_i in the morning, has to be considered. For this, $\overline{w'^2}_{max}$ values ~~were~~ are first determined for the hourly averaging periods and then averaged to retrieve $\overline{w'^2}_{max}$ for longer averaging periods. In contrast, the statistical error (Fig. 13b) is taken from variance calculations for explicitly larger time periods. After that, relative deviations (absolute difference normalized by the mean value) ~~were~~ are calculated for each time step and each instrument combination. ~~Finally, relative daviations were averaged for each day.~~ The resulting mean relative differences are given as an average ~~for~~ of all considered six days (Fig. 13a). For the three days when simultaneous w-measurements by HYB and WLS200 at ~~the Wasserwerk were~~ Wasserwerk are available (20, 22, and 24 April), the relative difference between these two measurements at the same site ~~was~~ is calculated as well. This gives a good estimate for the uncertainty that exists due to the comparison of measurements by instruments that are based on different technologies or made by different manufacturers (instrument uncertainty).

The daily mean relative deviation for HYB and WLS200 is less than 0.1 for $\Delta t = 1$ h and about 0.05 for longer averaging periods. For the other instrument combinations, it is about 0.5 for $\Delta t = 10$ min and decreases to about 0.2 for $\Delta t = 3$ h. For $\Delta t > 3$ h, it does not clearly decrease further. The mean normalized statistical error for $\Delta t = 3$ h is about 0.1 (Fig. 13b), so that the relative deviation is about twice the error. This means that the spatial differences between the variances are not statistically significant on the average, at least if the instrument-to-instrument uncertainty is considered. However,

this does not exclude the possibility of individual periods with significant spatial differences existing~~, as shown in Sec. 4.3~~; the diurnal time series of $\overline{w'^2}_{max}$ with the corresponding error bars ~~were~~ are also compared for larger Δt and the significant differences for the periods concerned ~~remained~~ remain (not shown). At the same time, a mean relative deviation of about 0.2 for $\Delta t = 3$ h means that the mean error that has to be expected when calculating variances from point measurements is about 10% minus the instrument uncertainty of about 2% (a factor of 0.5 is taken into account to derive the uncertainty of a single instrument from the calculated deviation); in other words, a point measurement is – on the average – spatially representative with an uncertainty of less than 10% when a measurement period of three hours is covered. This agrees with the statistical error of Lenschow et al. (1994) that was derived by theoretical considerations.

As the absolute difference does not provide any evidence of possible biases between the instrument measurements, absolute values of $\overline{w'^2}_{max}$ / w_*^2 are compared in Fig. 13c. The variances ~~were~~ are normalized by w_*^2 (averaged scaling) to retrieve comparable values for the different days. ~~As for the normalized profiles (Fig. 7), the values mainly range from 0.25 to 0.5.~~ While on the average they are as high at ~~the~~ Wasserwerk (HYB and WLS200) as at Hambach, most values are below the 1-1 diagonal for HALO. This explains why the relative difference is higher between HALO and both other instruments than between HYB and WTX (Fig. 13a). Nevertheless, there is no clear explanation why the variance is systematically smaller at Selhausen than 3 km north of this location. The sensible heat flux of SE1 ~~mostly~~ is quite low most of the time, but as shown in ~~Sec.~~Sect. 4.3.1, it is not representative of the surroundings of the HALO site. Finally, to compare the daily differences, the absolute differences between the lidars ~~were~~ are normalized by w_*^2 (Fig. 13d). The comparison reveals that on three days (18, 20, and 22 April), the deviations are largest between HALO and WTX and on one day between HALO and HYB (24 April). On 4 May, which is ~~the most perfect~~ closest to a perfectly cloud-free day, ~~all~~ the differences are smallest and on 19 May, which is a day with several mid-level clouds, they are largest. 19 May is the only day that falls into the wetter period with the Bowen ratio being low for all stations. ~~Due to this~~Therefore, scaling with w_*^2 (using a small sensible heat flux) results in higher values than for the other days. The variation of the differences from day to day can, hence, partly be explained by the occurrence of clouds and ~~therefore, by~~ by the resulting differences of the incoming radiation (~~Tab.~~ Table 2).

~~For this section, we~~ We finally conclude that the spatial differences on the average are as large as the statistical error ~~, which is~~ derived from theory, ~~and that this is~~ independent of the averaging period. The instrument uncertainty can be estimated to about 2% and the mean error is about 10% for an averaging period of three hours.

4.4 ~~Influence of the mean wind~~

4.3.3 Correlations of vertical velocity at different locations

~~Finally, we want to investigate the impact of the mean wind on spatial differences of the w-variance, especially for periods when surface heterogeneities do not explain the differences and when differences do not disappear, even if the averaging interval amounts to several hours.~~ For two of the three time periods investigated in ~~Sec.?? Sect. 4.3.1~~ (on 18 and 24 April), the mean wind ~~is from~~ direction is west to southwest. On both days, it is noticeable that the diurnal time series of $\overline{w'^2}_{max}$ at Wasserwerk and Hambach are very similar~~(Fig. 10), while it~~ , while the time series is different at Selhausen (~~see also Fig.13d~~Fig. 10). As the variances are similar, it can be expected that also the time series of w at Wasserwerk and Hambach exhibit a certain similarity. To investigate this, the cross correlation function of the two time series of w ~~was~~ is determined (Fig. 14).

As the convective time scale t_* ~~, also referred to as large-eddy turnover time,~~ is of the order of 10~~minutes~~ min and the travel time for the given distances between the lidar locations of about 3 km is between 4 and 12~~minutes~~ min, convective cells can be preserved between two locations at least on days with relatively strong mean wind. ~~This means that the original assumption that the w measurements were independent as long as they were more than 2 km apart turned out to be not valid for some days.~~ The day with the ~~strongest mean wind was~~ highest mean wind speed is 18 April; in the ~~mean~~ westerly flow, the WTX at Hambach is located downstream of WLS200 at ~~the~~ Wasserwerk. The cross correlation function between WLS200 and WTX in fact reveals a distinct maximum of correlation at a time lag of 200 s (Fig. 14a). The maximum correlation of 0.44 is found at heights between 500 m and 900 m. When shifting the time series of w' at 600 m for WTX backwards by 200 s compared to that of WLS200, the two time series agree very well (Fig. 14a). That means that the larger convective cells are advected from ~~the~~ Wasserwerk to the Hambach site without substantial changing (Taylor's hypothesis), which explains the similarity of the two time series of w and $\overline{w'^2}_{max}$.

In contrast to 18 April, the mean wind direction on 20 April is northeast. On this day, large differences of $\overline{w'^2}_{max}$ are observed between Hambach and Wasserwerk in the afternoon. The cross correlation function ~~(not shown)~~ also shows very low correlations (< 0.1; not shown).

On 24 April, the mean wind ~~again is from~~ direction again is southwest, but weaker than on 18 April. A maximum of the cross correlation function between WLS200 and WTX can also be discerned (Fig. 14b), but it is only 0.27. Nevertheless, the two time series (WTX shifted by 400 s) at 700 m agree ~~again~~ very well, at least after ~~1145 UTC~~11:45 UTC. At the same time, the cross correlation mainly gives negative values, if it is calculated between the time series of vertical velocity for Selhausen and Hambach or between Selhausen and ~~the~~ Wasserwerk (not shown).

The mean wind direction may thus be ~~a possible~~ one explanation why differences between the variances at Wasserwerk and ~~at~~ Hambach are found on 20 April, but not on 18 and 24 April (Fig. 10), although similar surface conditions exist on all of these days: The diurnal cycles of variances are similar at the two sites when the mean wind is parallel to their connecting axis, but different otherwise. For the time periods when the correlation between the two sites is high, the correlation

between the third site and each of the two is low. It is remarkable that on 24 April, when convective cells are advected past Wasserwerk and Hambach without substantial ~~changings~~changing, the mean vertical velocity (Fig. 15) is positive at Wasserwerk between ~~11–12 UTC~~ 11:00–12:00 UTC (more than 1 $\mathrm{m\,s^{-1}}$) and negative at Selhausen (~~11–13 UTC~~11:00–13:00 UTC, i.e. even for two hours). We hypothesize that, while many cells are observed on the northern axis, less occur about 3 km further south due to the subsidence in the surroundings of the cells. This assumption is confirmed by model simulations for 24 April with the Consortium for Small-scale Modeling (COSMO) model in LES mode. They were performed on a grid with 100 m horizontal resolution using a 3D-turbulence parameterization by Herzog et al. (2002). Model analyses of the operational model COSMO-DE (Baldauf et al., 2011) provided atmospheric initial and boundary conditions. The vertical velocity as calculated by the model is shown on a horizontal cross ~~sections~~ section at 600 m (Fig. 16). The instantaneous as well as the field averaged over one hour is given. About 1–1.5 km ~~the~~ south and north of the regions where the mean vertical velocity is positive on the hourly average, which is caused by convective cells advected with the mean wind, subsidence prevails. As shown by Lenschow and Stephens (1982), the mean w within thermals is positive and nearly two times higher than in the environment, where it is negative. This agrees very well with the mean w observed at the different locations on 24 April (Fig. 15). The spatial variance differences on 18 and 24 April can therefore be explained by the occurrence of ~~thermals~~organized structures of turbulence: While more convective cells travel past the Wasserwerk as well as past Hambach, ~~less occur near~~ subsidence in the surroundings of these cells prevails at Selhausen. This structure is presumably the signature of horizontal rolls that develop during conditions of combined surface heating and strong winds (Stull, 1988, Ch. 11.2), as was observed by Brown (1970) or Kropfli and Kohn (1978).

On 20 April, mean wind comes from northeast, so that thermals traveling from Hambach to Selhausen may be observed. However, this is not the case, and w-variance at both other sites differs from the one at Hambach (Fig. 10). One possible explanation is that, on days with easterly wind, the strongest influence of the open-pit coal mine on w-variance occurs at Hambach.

5 Summary and conclusions

During the HOPE campaign, multiple Doppler lidars were operated simultaneously at three different sites in the vertical stare mode to retrieve temporally high-resolved vertical velocity measurements. For this study, profiles of vertical velocity variance ~~profiles~~ were derived for the three sites to investigate the ~~spatial~~ spatiotemporal heterogeneity of turbulence in the cloud-free CBL. The aims were to ~~compare, in a first step,~~ analyze temporal variability as well as scaling of variance profiles and to compare the variance profiles for the different sites~~and to examine how large~~. It was investigated if spatial differences were ~~and, in a second step, to investigate if these differences were~~ statistically significant and if they depended on surface conditions, atmospheric conditions or on the averaging

intervals.

The investigated area was characterized by patchy agricultural land use. The typical size of the crop fields was of the order of 100 m. The eight weeks of the measurement period were divided into a drier period (mid-April to 6 May) and a wetter one (starting on 7 May). ~~It was found that the~~ The Bowen ratio varied between 0.5 and 4 during the drier period, while it was < 1 at all stations during the wetter period. Five of the six ~~selected days~~ days selected for this study fell into the drier period. Boundary-layer mixing was strong on all of the selected days and the height of the CBL was between 1.2 km and 2 km. Different methods to derive z_i ~~agreed well.~~ (radiosonde profiles, aerosol backscatter, w-variance profiles) agreed well in most cases. Only when a residual layer was present above the CBL, lower values were derived from w-variance profiles than by the other methods. Finally, z_i values from the aerosol backscatter were used, because it was the only method that yielded values for all time steps. On three of the days, clouds occurred~~and~~ , but the diurnal cycle of incoming radiation was only slightly affected on 18 April, when some boundary-layer clouds ~~occurred~~ were present and on 19 May, when mid-level clouds were observed. ~~Some cirrus clouds occurred~~ There were cirrus clouds on 24~~April, but they~~ April which did not perceptibly reduce incoming radiation. Moderate westerly wind dominated on most days; on 18 April, the mean horizontal wind was stronger than on the other days and it came from northeast on 20 April and 19 May.

The combination of smaller and larger Doppler lidars with complementary measurements at different ~~range gates and heights above ground~~ heights proved to be beneficial for the investigations. For the calculation of higher-order moments of w~~as measured by lidars~~, different aspects were considered: (1) The random noise of the signal ("uncorrelated noise") was removed, (2) the lack of spectral contribution to the total energy caused by spatial averaging of the lidar measurement was neglected, and (3) the statistical errors (systematic and sampling error according to Lenschow et al., 1994) that appear due to the spatial and temporal sub-sampling were ~~determined~~provided. Moreover, as measurements by lidar instruments from different manufacturers were compared here, also the instrument-dependent differences were calculated.

~~We found spatial differences of vertical velocity variances that were statistically significant. To investigate whether these differences were generated by heterogeneous surface conditions, scaling with the convective velocity w_* was applied.~~ For the scaling of the w-variance profiles, representative surface fluxes ~~are needed. It is assumed that the relevant area for these has a side length of $t_*|\boldsymbol{v}| \approx 3$–$5$ km. This means that a sensible heat flux that is representative of the whole area and with this, a spatially representative w_*, should be most suitable for scaling. However, using the same~~ were needed. The relevant length scale was estimated according to $\Delta t \cdot |\boldsymbol{v}|$, which is about 15 km for an averaging interval of one hour and a mean wind speed of 4 m s^{-1}. Weighted-averaged values of w_* ~~for all locations, only the temporal variability of the variances can be eliminated. Spatial differences can only be reduced by using different values of~~ were derived for an area of 30×30 km with the lidars in its center. Additionally, w_* ~~for each location for scaling the variances. Both scaling~~

methods were applied and the results imply that the spatial differences of the w-variances can not be explained by the heterogeneity of the surface conditions. Moreoverwas calculated using fluxes from the individual energy balance stations near the three sites. On average, scaled profiles for the whole data set showed large variations at the at the three locations agreed well with those shown by Willis and Deardorff (1974) , Caughey and Palmer (1979) or Lenschow et al. (1980) . However, they showed large scatter at individual locations, which indicates that the local hourly heat supply is not the only factor influencing the and the standard deviation was larger than the statistical error in most cases. The relative difference between both was between 5% and 34%. In a similar investigation, Lenschow et al. (2012) found a relative difference of about 10%. Evaluating the correlations between w_* and vertically averaged values of w-variance during the respective time interval in all cases. Apart from that, it was found that in some cases, the nearby energy balance stations could not provide representative surface fluxes so that the -variance, it turned out that the choice of the energy balance station that provides sensible heat fluxes was crucial for local scaling. The correlations varied between $R^2 = 0.28$ and 0.72, i.e. local scaling could be completely inappropriate to describe the upstream conditions determing the CBL turbulence at the lidar site. Therefore, the use of weighted-averaged fluxes for the calculation of scaling variables was preferable in these cases. Only at one location was the temporal variability is preferable for scaling (R^2 between 0.46 and 0.50). Unexplained temporal variance of w-variance well could not be related to the variability temporal variability of wind shear, mean wind speed or of the Obukhov length. Thus, time series of w were analyzed for two cases which contributed significantly to the unexplained temporal variance. In one case, the high variance was caused by an increased relative frequency of strong up- and downdrafts, and in the other one by broader or a larger number of thermals. Broad and strong thermals may additionally have caused the elevated maximum of w_* using the flux of the co-located energy balance station, with the correlation coefficient being 58. Lenschow et al. (2012) analyzed variance profiles for one location. As expected, the scatter around their mean value decreased after scaling, but a certain variation between values of 0.2 and 0.6 (for the vertical maximum) also remained. By removing the statistical error, they estimated that the real case-to-case variability was about 10and attributed it to the atmospheric stability which can be determined via the Obukhov length. However, the atmospheric stability can w-variance.

Statistically significant spatial differences were found by comparing vertically averaged values of vertical velocity variance at the three sites. They also occurred on days with westerly to southwesterly wind, when the influence of a large open-pit coal mine in the surroundings was presumably low. To investigate whether these differences were generated by heterogeneous surface conditions, local scaling was applied. The results implied that the heterogeneity of the surface conditions could not be the main factor causing the case-to-case variability in this investigation, as only days with buoyancy-driven turbulence have been chosen. reason. Secondly, the influence of different averaging intervals on the spatial differences of w-variance were was analyzed. Relative deviations of

w-variances between all instruments averaged over all days as well as statistical errors decreased strongly with increasing averaging intervals. Thus, mean relative deviations were about as large as the relative statistical errors ~~that can be derived from theoretical considerations (Lenschow et al., 1994)~~ for all averaging intervals. ~~Mean relative deviations as well as errors decreased strongly with increasing averaging intervals~~On the other hand, relative deviations of variances at different sites were about three times higher than between those derived from measurements by different lidars at the same site. Postulating that the uncertainty of a point measurement should not be larger than 10%, measurement periods of at least 3 hours ~~– or hourly measurement periods of three instruments at different locations –~~ are necessary. ~~However, the uncertainty does not decrease much further for longer averaging intervals. Moreover, daily averages revealed that mean deviations were larger for days with a small number of clouds than on days with no clouds.~~

Finally, a detailed analysis of periods with significant spatial differences of w-variance provided some insight into possible reasons: It was found that a varying degree of correlation between vertical velocity fluctuations existed for two locations on ~~a~~ an east-west axis~~, depending on the mean wind speed~~. On 18 April, a day with stronger ~~southwesterly~~ west-southwesterly wind, and on 24 April with moderate wind speed, the travel time was smaller than the large-eddy turnover time. On these days, fluctuations and variances were similar at the two locations, while the correlation of both with fluctuations at the third location about 2.5 km further south was low. ~~It could be shown that for example on 24 April~~Simultaneously, the mean vertical velocity ~~at the third location was negative~~ was positive at the first two sites, while it was negative at the third location for a time period of two hours~~, while it was positive at the other sites~~. The reason is that several convective cells travelled past the ~~two northern~~ first two sites, while subsidence prevailed at the third site during the whole 2-hour period~~, which also~~. The presence of organized structures of turbulence, which is also confirmed by LES, explains why spatial variance differences ~~do~~ existed and did not disappear even for averaging periods of more than three hours. ~~This is confirmed by LES.~~

Based on these findings, the following conclusions can be drawn: (1) The representativeness of single-column turbulence characteristics as observed by Doppler lidars is not necessarily given, even if long time periods are available (with the maximum possible length of the time period being the whole part of day with an existing CBL); (2) local scaling with w_* is possible but should only be considered, if the representativeness of ~~a single~~ an individual energy balance station for a larger area is ~~given; and~~ proven; and (3) ~~it is recommended to register turbulence profiles at more than one location – if Doppler lidar measurements are performed – to take the spatial variability of turbulence into account, which can depend on the relative location of the measurements compared to the mean wind direction~~organized structures of turbulence in the CBL such as horizontal rolls aligned with the mean wind may be the reason for statistically significant spatial differences of vertical velocity variances.

Appendix A: Error Statistics

A1 Uncorrelated Noise

The so-called "uncorrelated noise" defined by Lenschow et al. (2000) is based on the assumption that the measurement signal is "contaminated by uncorrelated random noise". By definition, it is uncorrelated from the signal and the respective error can, thus, be removed from the calculated variance. According to equation (8) from Lenschow et al. (2000), the uncorrelated-noise error is equal to the difference between the first and zero lag of the autocovariance function.

A2 Systematic Error

According to Lenschow et al. (1994), the statistical error can be separated into the systematic and the random error (see App. A3). The systematic error is caused by the fact that the variance $\overline{w'^2}$ derived from the measurement is, strictly speaking, a time average $\overline{w'^2}^t$, which is not equal to the ensemble average $\overline{w'^2}^{t,\boldsymbol{x}}$. With these definitions, equation (14) from Lenschow et al. (1994) is

$$\frac{\overline{w'^2}^{t,\boldsymbol{x}}}{\overline{w'^2}^t} \approx 1 - 2\frac{\widetilde{T}}{\Delta t}, \tag{A1}$$

with the averaging time Δt and the integral time scale $\widetilde{T}$ (see App. B). The absolute value of the systematic error can, thus, be calculated as

$$|\overline{w'^2}^{t,\boldsymbol{x}} - \overline{w'^2}^t| = \overline{w'^2}^t \cdot 2\frac{\widetilde{T}}{\Delta t}. \tag{A2}$$

From this, it can be seen that the ~~error decreases for increasing averaging periods and increases with the integral time scale~~systematic error increases for increasing integral time scales, decreasing averaging periods as well as with the variance itself.

A3 Random Error

The random or sampling error takes into account that the length of the measured time series is not unlimited and that "random" time slots may differ. Lenschow et al. (1994) show that, using the error variance σ_2^2 for the second moment, the random error can be approximated to

$$\sigma_2 = \overline{w'^2}^t \cdot \sqrt{2\frac{\widetilde{T}}{\Delta t}}. \tag{A3}$$

The ratio of the systematic to the random error can, thus, be determined as $\sqrt{2\frac{\widetilde{T}}{\Delta t}}$. For the commonly used averaging time of 1 h and a typical integral time scale of about 50 s (in this study, which agrees with numbers from Lothon et al., 2006, for example), this expression amounts to 0.17. This means that in this case, the random error is more than five times higher than the systematic error. Only for distinctly larger integral time scales, i.e. $\widetilde{T} \geq 450$ s, does the systematic error become higher than the random error for the 1-h averaging period.

Appendix B: Integral Time Scale

Going back to Lumley and Panofsky (1964), the integral time scale is defined as the integral of the autocorrelation function R. Here, it was calculated as the integral between lag zero determined by extrapolation (Lenschow et al., 2000) and the first zero-crossing of R.

Acknowledgements. This work was funded by the Federal Ministry of Education and Research in Germany (BMBF) under the research program "High Definition Clouds and Precipitation for Climate Prediction – HD(CP)^2" (FKZ: 01LK1212F). We want to thank Marius Schmidt from Forschungszentrum Jülich for providing surface fluxes from the energy balance stations that were operated within the framework of TERENO. Furthermore, Mauro Sulis and Prabhakar ~~Shreshtha~~ Shrestha from University of Bonn provided the land-use data.
Katja Träumner and the Young Investigator Group "Coherent structures" contributed to the operation of the Doppler lidars during the HOPE experiment and made their analysis tools available, including functions for error calculations. We are also grateful to the whole IMK team for their efforts in deploying KITcube.

References

Angevine, W. M., Doviak, R. J., and Sorbjan, Z.: Remote sensing of vertical velocity variance and surface heat flux in a convective boundary layer, J. Appl. Meteor., 33, 977–983, 1994.

Ansmann, A., Fruntke, J., and Engelmann, R.: Updraft and downdraft characterization with Doppler lidar: cloud-free versus cumuli-topped mixed layer, Atmos. Chem. Phys., 10, 7845–7858, doi:10.5194/acp-10-7845-2010, 2010.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities, Mon. Weather Rev., 139, 3887–3905, 2011.

Beyrich, F. and Mengelkamp, H.-T.: Evaporation over a heterogeneous land surface: EVA_GRIPS and the LITFASS-2003 experiment – an overview, Boundary-Layer Meteorol., 121, 5–32, doi:10.1007/s10546-006-9079-z, 2006.

Brown, R. A.: A secondary flow model for the planetary boundary layer, J. Atmos. Sci., 27, 742–757, 1970.

Browning, K. and Wexler, R.: The determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteor., 7, 105–113, 1968.

Brugger, P., Träumner, K., and Stawiarski, C.: Evaluation of a procedure to correct spatial averaging in turbulence statistics from a Doppler lidar by comparing time series with an ultrasonic anemometer, submitted, 2015.

Cañadillas, B., Westerhellweg, A., and Neumann, T.: Testing the performance of a ground-based wind LiDAR system: One year intercomparison at the offshore platform FINO1, DEWI Mag, 38, 58–64, 2011.

Carbajo Fuertes, F., Iungo, G. V., and Porté-Agel, F.: 3D turbulence measurements using three synchronous wind LiDARs: validation against sonic anemometry, J. Atmos. Oceanic Technol., 31, 1549–1556, doi:10.1175/JTECH-D-13-00206.1, 2014.

Caughey, S. J. and Palmer, S. G.: Some aspects of turbulence structure through the depth of the convective boundary layer, Q. J. R. Meteorol. Soc., 105, 811–827, doi:10.1002/qj.49710544606, 1979.

Darbieu, C., Lohou, F., Lothon, M., Vilà-Guerau de Arellano, J., Couvreux, F., Durand, P., Pino, D., Patton, E. G., Nilsson, E., Blay-Carreras, E., and Gioli, B.: Turbulence vertical structure of the boundary layer during the afternoon transition, Atmos. Chem. Phys., 15, 10 071–10 086, doi:10.5194/acp-15-10071-2015, 2015.

Deardorff, J. W.: Preliminary results from numerical integrations of the unstable planetary boundary layer, J. Atmos. Sci., 27, 1209–1211, 1970a.

Deardorff, J. W.: Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection, J. Atmos. Sci., 27, 1211–1213, 1970b.

Deardorff, J. W.: Three-dimensional numerical study of the height and mean structure of a heated planetary boundary layer, Boundary-Layer Meteorol., 7, 81–106, doi:10.1007/BF00224974, 1974.

Eder, F., Schmidt, M., Damian, T., Träumner, K., and Mauder, M.: Mesoscale eddies affect near-surface turbulent exchange: Evidence from lidar and tower measurements, J. Appl. Meteor., 54, 189–206, doi:10.1175/JAMC-D-14-0140.1, 2015.

Emeis, S., Schäfer, K., and Münkel, C.: Surface-based remote sensing of the mixing-layer height – a review, Meteorol. Z., 17, 621–630, doi:10.1127/0941-2948/2008/0312, 2008.

Eng, K., Coulter, R. L., and Brutsaert, W.: Vertical velocity variance in the mixed layer from radar wind profilers, J. Hydrol. Eng., 8, 301–307, 2003.

Eymard, L. and Weill, A.: Dual Doppler radar investigation of the tropical convective boundary layer, J. Atmos. Sci., 45, 853–864, 1988.

Frehlich, R., Hannon, S., and Henderson, S.: Coherent Doppler lidar measurements of wind field statistics, Boundary-Layer Meteorol., 86, 233–256, doi:10.1023/A:1000676021745, 1998.

Graf, A., Schüttemeyer, D., Geiß, H., Knaps, A., Möllmann-Coers, M., Schween, J. H., Kollet, S., Neininger, B., Herbst, M., and Vereecken, H.: Boundedness of turbulent temperature probability distributions, and their relation to the vertical profile in the convective boundary layer, Boundary-Layer Meteorol., 134, 459–486, 2010.

Hadfield, M., Cotton, W., and Pielke, R.: Large-eddy simulations of thermally forced circulations in the convective boundary layer. Part I: A small-scale circulation with zero wind, Boundary-Layer Meteorol., 57, 79–114, doi:10.1007/BF00119714, 1991.

Herzog, H.-J., Vogel, G., and Schubert, U.: LLM – a nonhydrostatic model applied to high-resolving simulations of turbulent fluxes over heterogeneous terrain, Theor. Appl. Climatol., 73, 67–86, doi:10.1007/s00704-002-0694-4, 2002.

Hogan, R. J., Grant, A. L., Illingworth, A. J., Pearson, G. N., and O'Connor, E. J.: Vertical velocity variance and skewness in clear and cloud-topped boundary layers as revealed by Doppler lidar, Q. J. R. Meteorol. Soc., 135, 635–643, 2009.

Kaimal, J. C., Wyngaard, J. C., Haugen, D. A., Coté, O. R., Izumi, Y., Caughey, S. J., and Readings, C. J.: Turbulence structure in the convective boundary layer, J. Atmos. Sci., 33, 2152–2169, 1976.

Kalthoff, N., Adler, B., Wieser, A., Kohler, M., Träumner, K., Handwerker, J., Corsmeier, U., Khodayar, S., Lambert, D., Kopmann, A., Kunka, N., Dick, G., Ramatschi, M., Wickert, J., and Kottmeier, C.: KITcube – a mobile observation platform for convection studies deployed during HyMeX, Meteorol. Z., 22, 633–647, doi:10.1127/0941-2948/2013/0542, 2013.

Kropfli, R. and Kohn, N.: Persistent horizontal rolls in the urban mixed layer as revealed by dual-Doppler radar, J. Appl. Meteor., 17, 669–676, 1978.

Lenschow, D. H. and Stankov, B. B.: Length scales in the convective boundary layer, J. Atmos. Sci., 43, 1198–1209, 1986.

Lenschow, D. H. and Stephens, P. L.: The role of thermals in the convective boundary layer, Boundary-Layer Meteorol., 19, 509–532, doi:10.1007/BF00122351, 1980.

Lenschow, D. H. and Stephens, P. L.: Mean vertical velocity and turbulence intensity inside and outside thermals, Atmos. Environ., 16, 761 – 764, doi:10.1016/0004-6981(82)90393-6, 1982.

Lenschow, D. H., Wyngaard, J. C., and Pennell, W. T.: Mean-field and second-moment budgets in a baroclinic, convective boundary layer, J. Atmos. Sci., 37, 1313–1326, 1980.

Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics?, J. Atmos. Oceanic Technol., 11, 661–673, 1994.

Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring second-through fourth-order moments in noisy data, J. Atmos. Oceanic Technol., 17, 1330–1347, 2000.

Lenschow, D. H., Lothon, M., Mayor, S. D., Sullivan, P. P., and Canut, G.: A comparison of higher-order vertical velocity moments in the convective boundary layer from lidar with in situ measurements and large-eddy simulation, Boundary-Layer Meteorol., 143, 107–123, 2012.

Lothon, M., Lenschow, D. H., and Mayor, S. D.: Coherence and scale of vertical velocity in the convective boundary layer from a Doppler lidar, Boundary-Layer Meteorol., 121, 521–536, doi:10.1007/s10546-006-9077-1, 2006.

Lothon, M., Lenschow, D. H., and Mayor, S. D.: Doppler lidar measurements of vertical velocity spectra in the convective planetary boundary layer, Boundary-Layer Meteorol., 132, 205–226, 2009.

Lumley, J. L. and Panofsky, H. A.: The structure of atmospheric turbulence, Interscience monographs and texts in physics and astronomy ; 12, Interscience Publ., New York, NY [u.a.], 239 pp., 1964.

Mahrt, L.: Flux sampling errors for aircraft and towers, J. Atmos. Oceanic Technol., 15, 416–429, 1998.

Mauder, M. and Foken, T.: Documentation and instruction manual of the eddy-covariance software package TK3, Abteilung Mikrometeorologie: Arbeitsergebnisse 46, ISSN 1614-8924, 60 pp, 2011.

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H. P., Schmidt, M., and Steinbrecher, R.: A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, Agric. Forest Meteorol., 169, 122 – 135, doi:10.1016/j.agrformet.2012.09.006, 2013.

Moeng, C.-H.: A large-eddy-simulation model for the study of planetary boundary-layer turbulence, J. Atmos. Sci., 41, 2052–2062, 1984.

Monin, A. S. and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Contrib. Geophys. Inst. Acad. Sci. USSR, 151, 163–187, 1954.

Neggers, R. A. J., Duynkerke, P. G., and Rodts, S. M. A.: Shallow cumulus convection: A validation of large-eddy simulation against aircraft and Landsat observations, Q. J. R. Meteorol. Soc., 129, 2671–2696, doi:10.1256/qj.02.93, 2003.

Panofsky, H. A. and Mazzola, C.: Variances and spectra of vertical velocity just above the surface layer, Boundary-Layer Meteorol., 2, 30–37, 1971.

Pearson, G., Davies, F., and Collier, C.: An analysis of the performance of the UFAM pulsed Doppler lidar for observing the boundary layer, J. Atmos. Oceanic Technol., 26, 240–250, doi:10.1175/2008JTECHA1128.1, 2009.

Pospichal, B. and Crewell, S.: Boundary layer observations in West Africa using a novel microwave radiometer, Meteorol. Z., 16, 513–523, doi:10.1127/0941-2948/2007/0228, 2007.

Schröter, M., Bange, J., and Raasch, S.: Simulated Airborne Flux Measurements in a LES generated Convective Boundary Layer, Boundary-Layer Meteorol., 95, 437–456, doi:10.1023/A:1002649322001, 2000.

Sorbjan, Z.: On similarity in the atmospheric boundary layer, Boundary-Layer Meteorol., 34, 377–397, doi:10.1007/BF00120989, 1986.

Sorbjan, Z.: Local similarity in the convective boundary layer (CBL), Boundary-Layer Meteorol., 45, 237–250, doi:10.1007/BF01066672, 1988.

Sorbjan, Z.: Structure of the atmospheric boundary layer, Prentice Hall, Englewood Cliffs, New Jersey, 317 pp., 1989.

Steinfeld, G., Letzel, M., Raasch, S., Kanda, M., and Inagaki, A.: Spatial representativeness of single tower measurements and the imbalance problem with eddy-covariance fluxes: results of a large-eddy simulation study, Boundary-Layer Meteorol., 123, 77–98, doi:10.1007/s10546-006-9133-x, 2007.

Stull, R. B.: An introduction to boundary layer meteorology, Atmospheric sciences library ; 13, Kluwer, Dordrecht, Netherlands [u.a.], 666 pp., 1988.

Taylor, G. I.: The spectrum of turbulence, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 164, pp. 476–490, The Royal Society, 1938.

Träumner, K., Kottmeier, C., Corsmeier, U., and Wieser, A.: Convective boundary-layer entrainment: Short review and progress using Doppler lidar, Boundary-Layer Meteorol., 141, 369–391, doi:10.1007/s10546-011-9657-6, 2011.

Tucker, S. C., Senff, C. J., Weickmann, A. M., Brewer, W. A., Banta, R. M., Sandberg, S. P., Law, D. C., and Hardesty, R. M.: Doppler Lidar Estimation of Mixing Height Using Turbulence, Shear, and Aerosol Profiles, J. Atmos. Oceanic Technol., 26, 673–688, doi:10.1175/2008JTECHA1157.1, 2009.

van Heerwaarden, C. C., Mellado, J. P., and De Lozar, A.: Scaling laws for the heterogeneously heated free convective boundary layer, J. Atmos. Sci., 71, 3975–4000, doi:10.1175/JAS-D-13-0383.1, 2014.

Warner, J.: The structure and intensity of turbulence in air over the sea, Q. J. R. Meteorol. Soc., 98, 175–186, doi:10.1002/qj.49709841514, 1972.

Willis, G. E. and Deardorff, J. W.: A laboratory model of the unstable planetary boundary layer, J. Atmos. Sci., 31, 1297–1307, 1974.

Wyngaard, J. C., Coté, O. R., and Izumi, Y.: Local free convection, similarity, and the budgets of shear stress and heat flux, J. Atmos. Sci., 28, 1171–1182, 1971.

Young, G. S.: Turbulence Structure of the Convective Boundary Layer. Part I. Variability of Normalized Turbulence Statistics, J. Atmos. Sci., 45, 719–726, 1988.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A network of terrestrial environmental observatories in Germany, Vadose Zone J., 10, 955–973, doi:10.2136/vzj2010.0139, 2011.

Table 1. Overview of lidar instruments at the three sites, with abbreviations used in the text, measurement range r for the vertical stare mode, range-gate length Δr, and applied threshold of signal-to-noise ratio (SNR; w measurements with SNR below the threshold were not used in this study); *n/c* for "not communicated", * according to Pearson et al. (2009) .

lidar	HYB	WLS200	HALO	WTX	WLS7
location	Wasserwerk	Wasserwerk	Selhausen	Hambach	Hambach
specification	WindTracer	WINDCUBE ~~200S~~200s	Stream Line	WindTracer	~~WindCube~~ WINDCUBE v2
manufacturer	Lockheed Martin CT	Leosphere	Halo Photonics	Lockheed Martin CT	Leosphere
laser wavelength in nm	2023	1543	≈ 1500	1617	1543
r in m~~AGL~~ a.g.l.	350 – above CBL top	50 – CBL top	60 – CBL top	350 – CBL top	40 – 290
Δr in m	≈ 60	25	18	≈ 60	25
SNR threshold in dBZ	-8	-26	-16	-8	-22
pulse repetition frequency in kHz	0.50	*n/c*	15	0.75	30
sampling rate in MHz	250	*n/c*	30*	250	250

Table 2. Overview of characteristic mean values and scales for all considered days (spatially averaged for surface measurements and turbulence characteristics, Hambach for other variables): Diurnal maximum of surface sensible heat flux H_0 and of boundary-layer height z_i; daily mean values of integrated water vapor IWV, of spatial mean and standard deviation of incoming shortwave radiation $Q_{SW,in}$, of mean boundary-layer wind speed $|\boldsymbol{v}|$, and of wind direction; diurnal maximum of convective velocity scale w_* and , of corresponding convective time scale t_* ; , and diurnal mean of $\overline{w'^2}_{max}$ and of integral time scale $\widetilde{T}$ (instrument mean)Obukhov length L; estimated peak wavelength of turbulence spectra in 600 m height (10–17 UTC10:00–17:00 UTC), $\lambda_{p,w} = |\boldsymbol{v}| f_{p,w}^{-1}$ (using Taylor's hypothesis), with time period $T = f_{p,w}^{-1}$scale $T_{p,w} = f_{p,w}^{-1}$, and the period T of the autocorrelation function with corresponding wavelength λ (denoted as *n/a* when no estimation was possible); diurnal mean of $\overline{w'^2}_{max}$ and of integral time scale $\widetilde{T}$ (same height as $\overline{w'^2}_{max}$).

	18/04	20/04	22/04	24/04	04/05	19/05
$\|\boldsymbol{v}\|$ in m s^{-1} s^{-1}	12	8	4	5	8	5
wind dir. in °	250	45	270	270	270	0-90
IWV in kg m^{-2} m^{-2}	12	8	8	20	10	10
$Q_{SW,in}$ in W m^{-2} m^{-2}	460	490	510	520	560	580
$\sigma(Q_{SW,in})$ in W m^{-2} m^{-2}	100	60	30	60	30	90
H_0 in W m^{-2} m^{-2}	200	210	180	180	200	90
z_i in m	2030	1350	1900	1330	1280	1250
w_* in m s^{-1}	2.10	1.92	1.86	1.70	1.82	1.45
T t_* in min	15	11	16	12	11	15
$-L$ in m	125	34	5	30	51	37
peak of spectra:						
$T_{p,w}$ in min	*n/a*	5.5	8	8	5	*n/a*
$\lambda_{p,w}$ in km	*n/a*	2.7	2	2.5	2	*n/a*
period of autocorrelation function:						
T in min	*n/a*	6–10	15	10	*n/a*	6
λ in km	*n/a*	2.8–4.4	3.6	3	*n/a*	1.8
$\overline{w'^2}_{max}$ in m^{-2}s^{-2} m^{-2} s^{-2}	1.65	1.55	1.2	0.95	1.1	1.05
w_* in ms^{-1} 2.241.98 2.02 1.82 1.93 1.39 t_* in min 15 11 16 12 11 15$\widetilde{T}$ in s	40	47	55	56	40	45

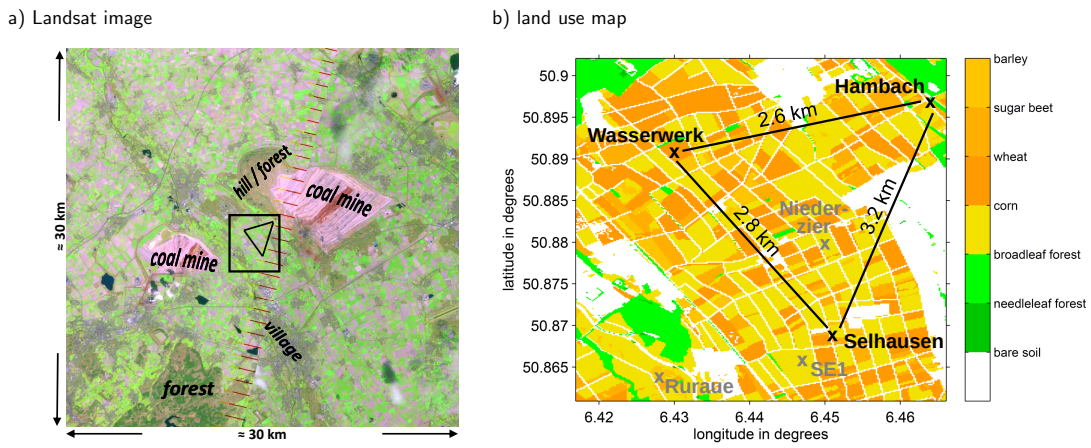

Figure 1. (a) Landsat false color image (composite of infrared and visible bands) from 24 April 2013 (available from the U.S. Geological Survey at landsatlook.usgs.gov); the black rectangle denotes the sector shown in (b); additionally, characteristic topographic features are marked; (b) land-use classification with positions of energy balance stations and lidars at Hambach and Wasserwerk and of the lidar at Selhausen (black crosses) as well as of the TERENO energy balance stations at Ruraue, near Selhausen (SE1), and at Niederzier (gray crosses); black lines denote the relative lidar locations.

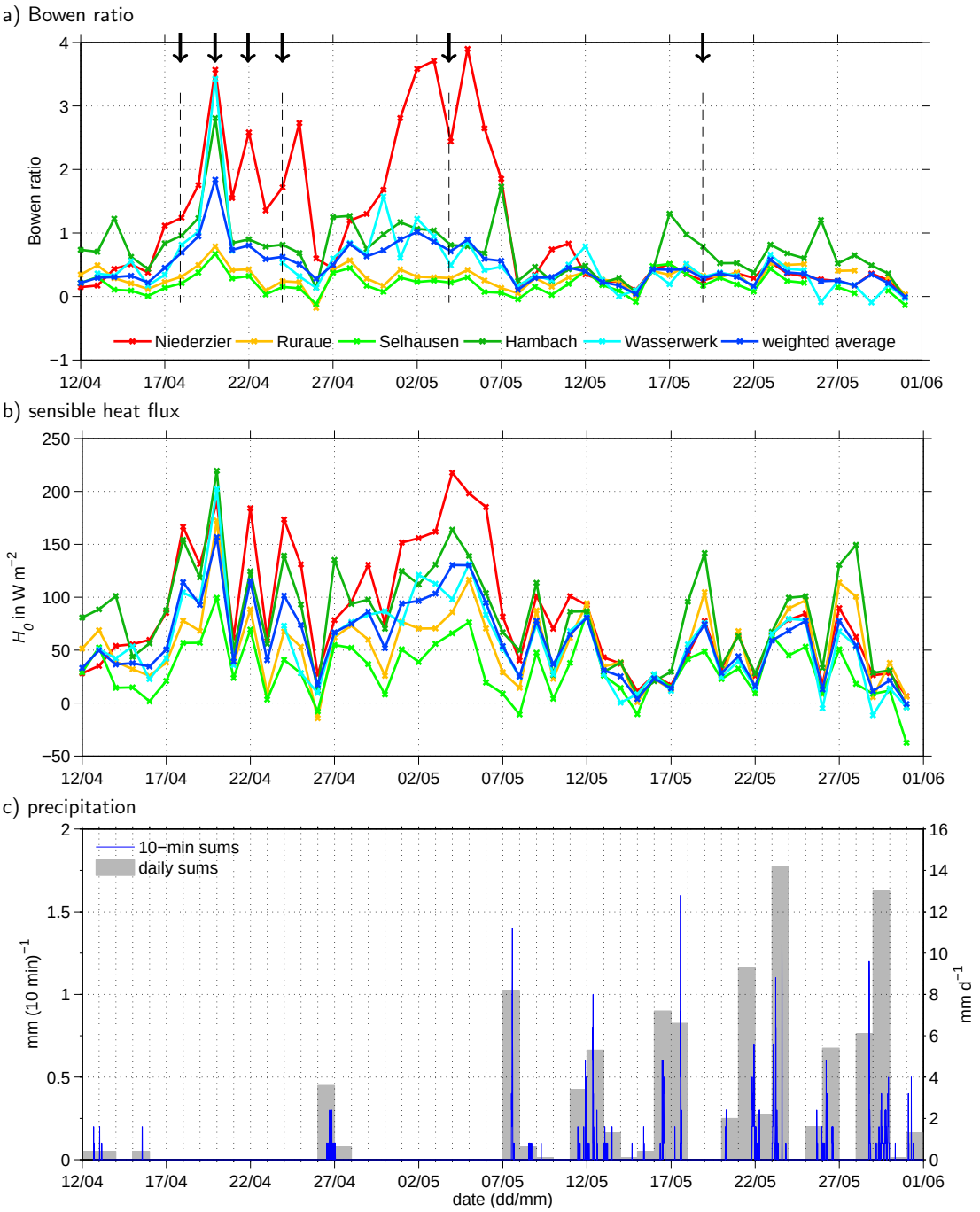

Figure 2. (a) Bowen ratio for all energy balance stations as well as for the weighted-averaged fluxes (weighted with the area fraction of each land-use class), calculated from daily averaged values of surface fluxes for 09:00–15:00 UTC; black arrows denote the selected days; (b) sensible heat fluxes H_0 as used for calculation of Bowen ratio in (a); (c) precipitation from rain gauge measurements at Wasserwerk.

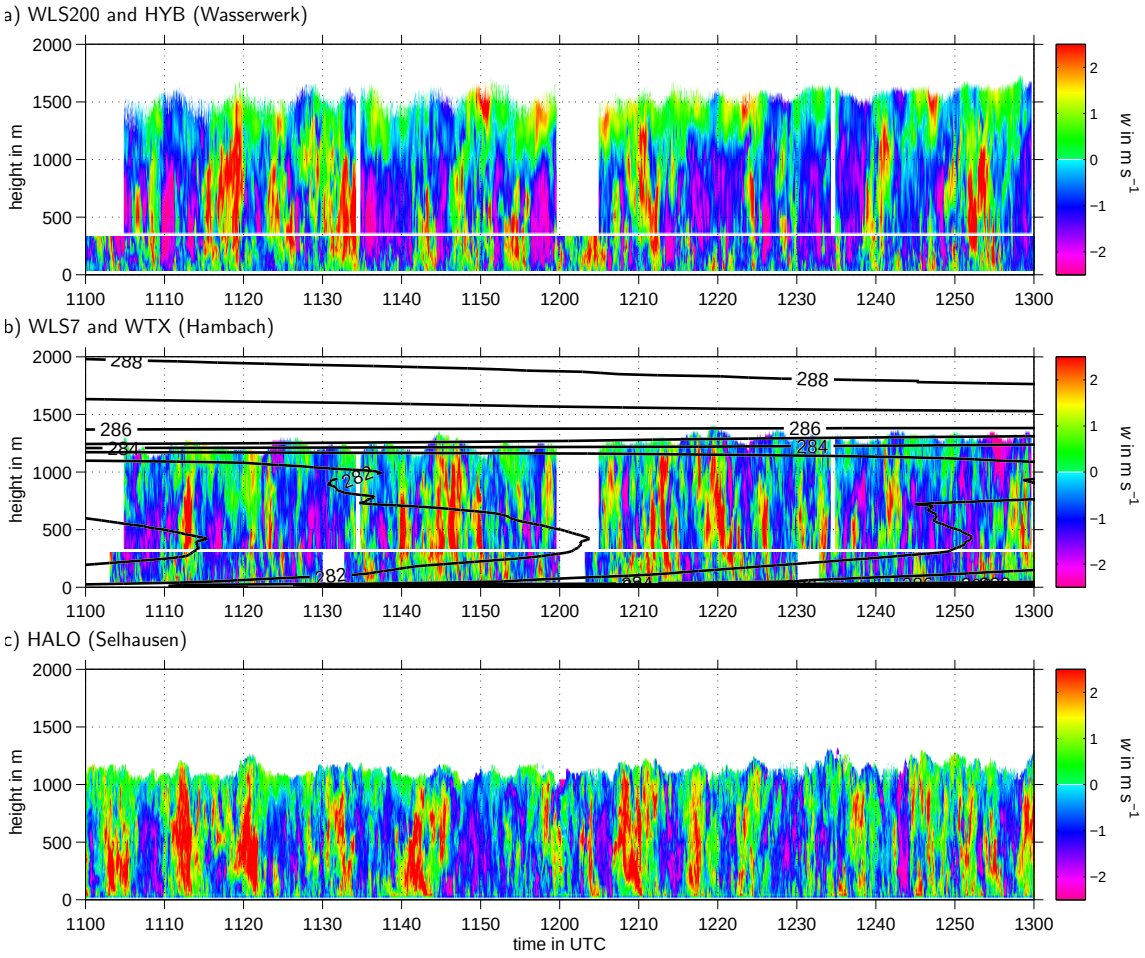

Figure 3. Vertical velocity as observed by Doppler lidars at three different locations on 20 April 2013 (11:00–13:00 UTC) with isolines of potential temperature (in K) in (b) as derived from radiosoundings.

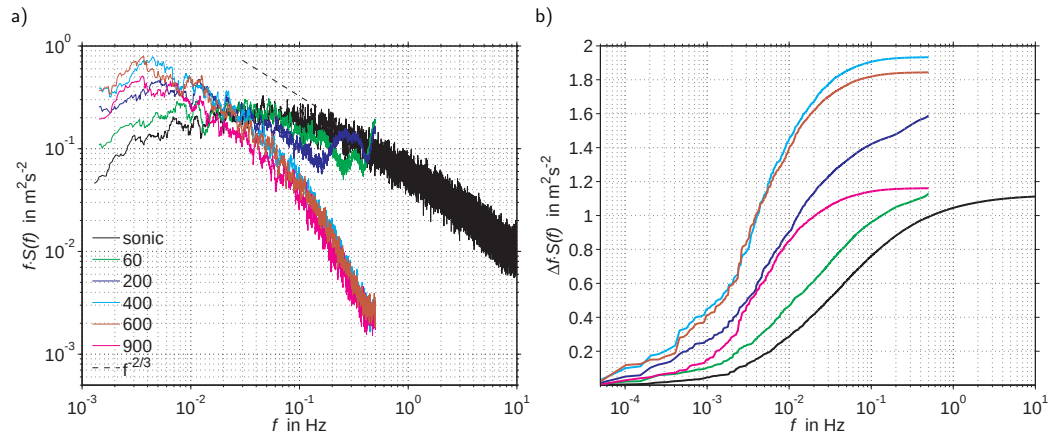

Figure 4. a) Energy density (S) spectra of w at Hambach on 20 April 2013, 09:00-15:00 UTC, from an ultrasonic at 30 m, WLS7 (60 m and 200 m), and WTX (400 m, 600 m, 900 m); additionally, the theoretical slope in the inertial subrange is given; (b) As in (a), but accumulated curves to illustrate the contributions of different frequencies to the variance.

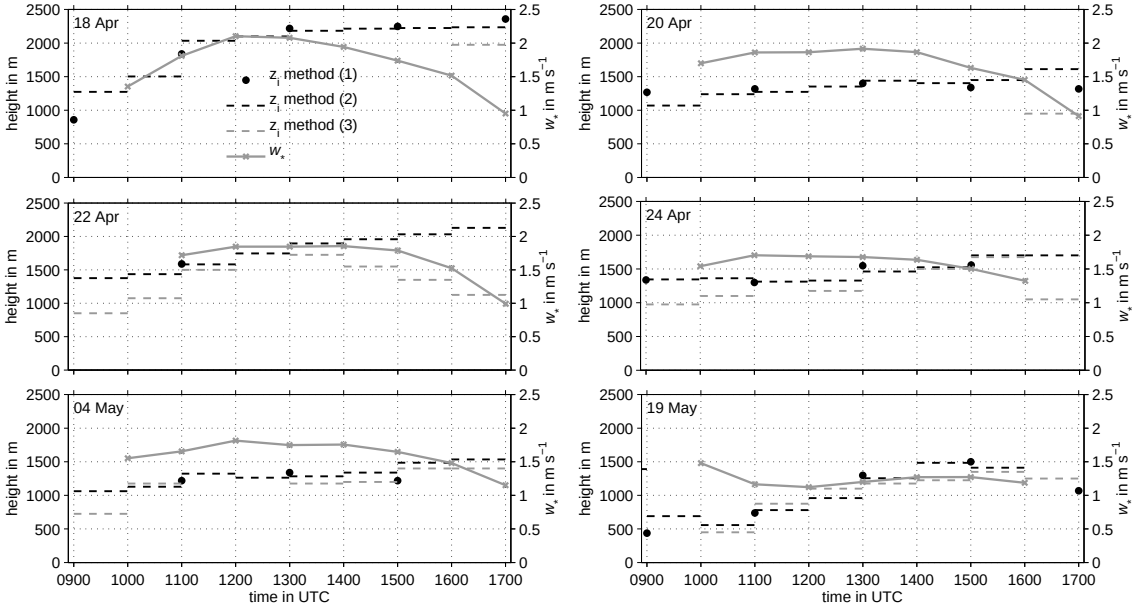

Figure 5. CBL heights derived from radiosoundings (maximum temperature gradient = inversion; method (1)), from lidar backscatter data (WTX; method (2)) as well as from a variance threshold (method (3)) for all considered cloud-free days; additionally, the convective velocity scale w_* (determined using weighted-averaged values of sensible heat flux) is given.

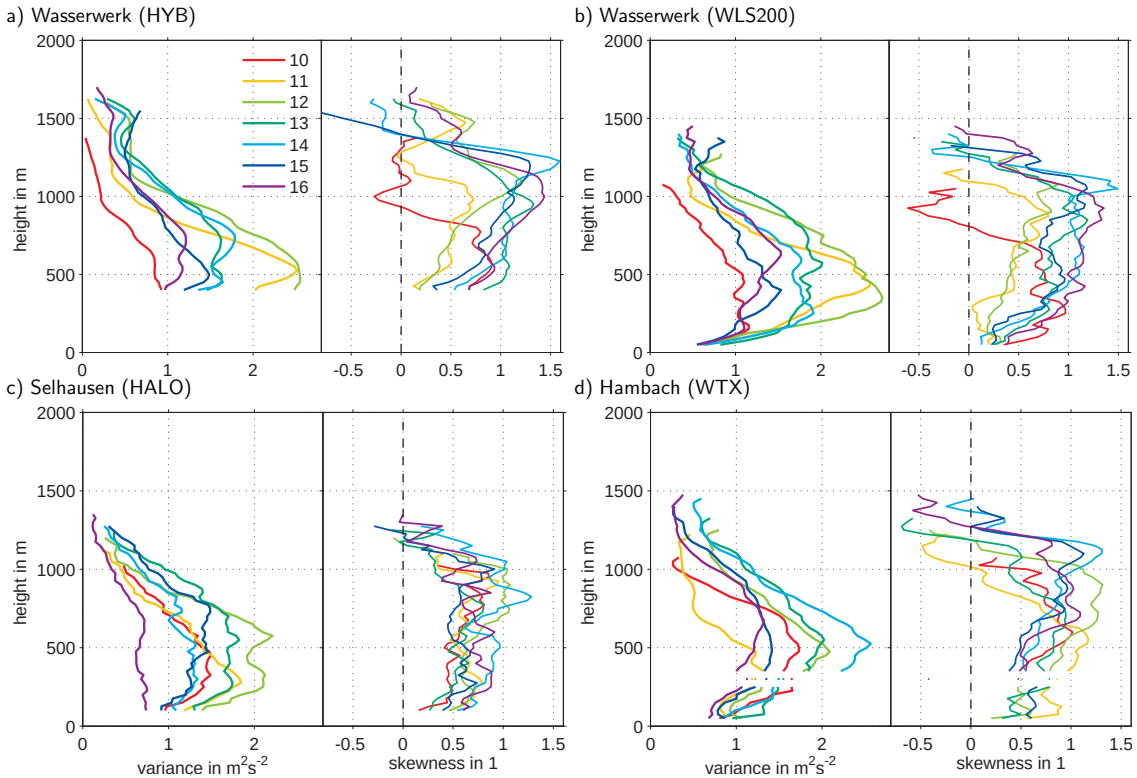

Figure 6. Vertical profiles of hourly vertical velocity variance and skewness from lidar measurements at the three locations for 10:00–17:00 UTC on 20 April 2013; the legend labels in (a) refer to the end in UTC for each averaging period of 60 min.

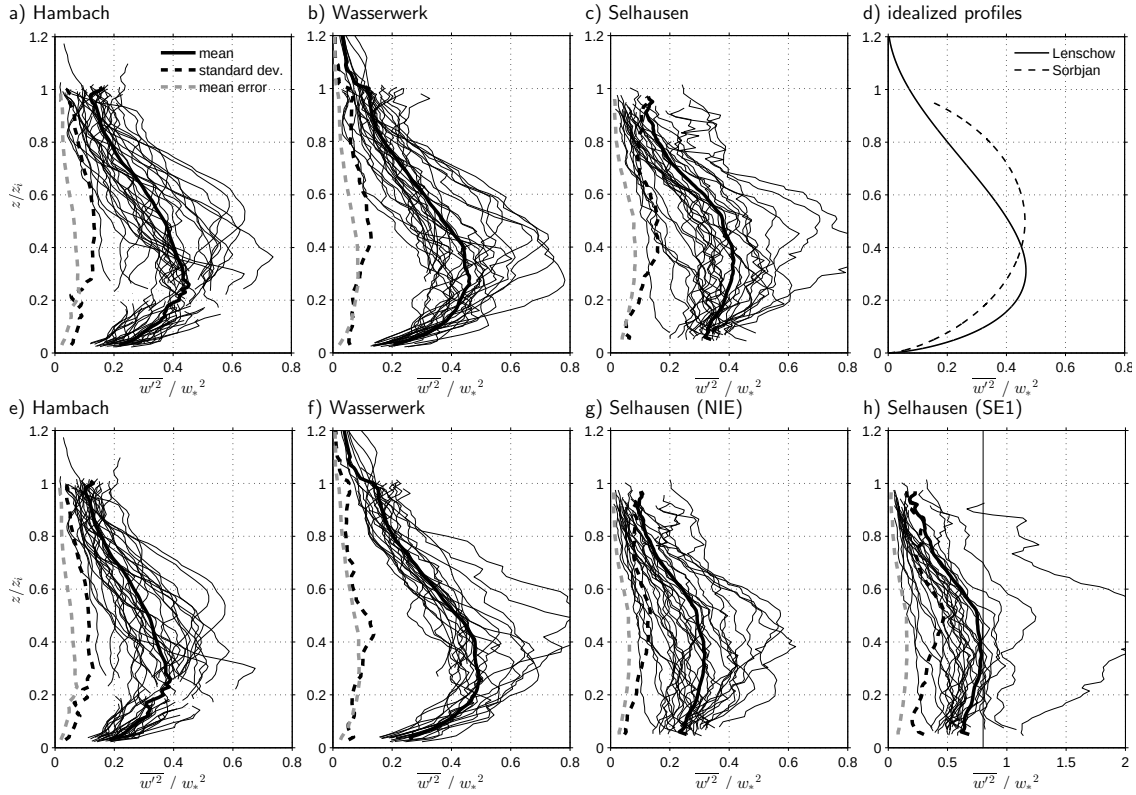

Figure 7. Normalized hourly variance profiles for 18, 20, 22, 24 April and 04 May (11:00–16:00 UTC) with mean profile, standard deviation and mean normalized statistical error (legend in a), using averaged (a, b, c) and local scaling (e, f, g, h) for each location; different energy balance stations were used for scaling the profiles of Selhausen in (g and h); in (d), the idealized profiles according to Eq. 1 are given.

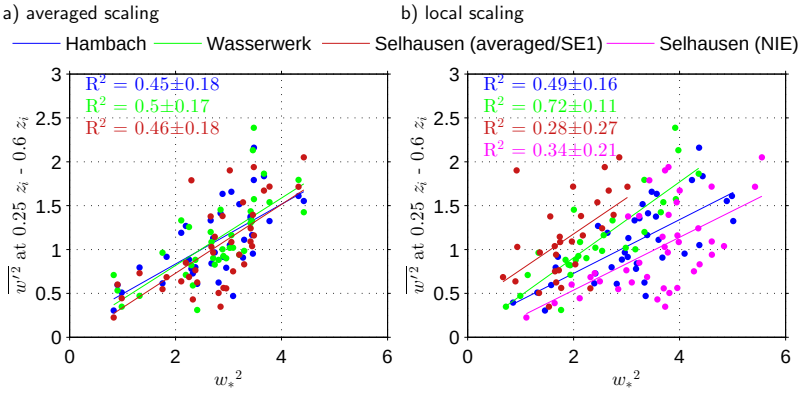

Figure 8. Correlations of vertical velocity variance averaged over 0.25 to 0.60 z_i and w_*^2, calculated using the weighted-averaged fluxes (a) and fluxes of nearby stations (b) for all time steps as in Fig. 7 but for 10:00–17:00 UTC, with lines of best fit from linear regression, squared correlation coefficients R^2 and confidence interval at the 95% level.

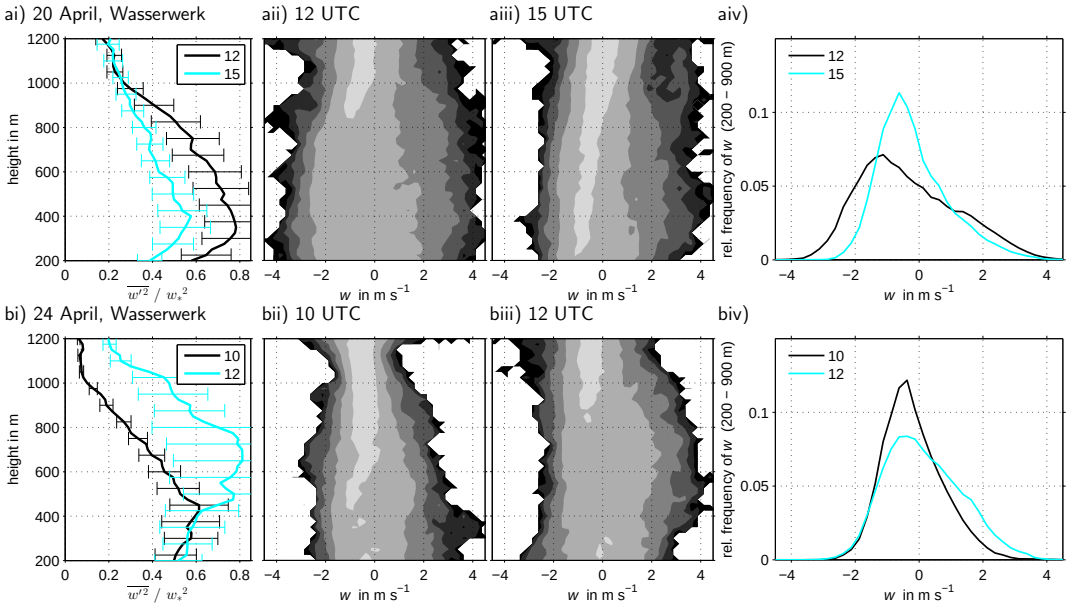

Figure 9. Normalized variance profiles with error bars (statistical error according to Lenschow et al., 1994) ; for each (ai) and (bi), two time steps were selected from from Fig. 7(b) and (f), respectively; for each time step, frequency distributions are given as a function of height (aii, aiii, bii, and biii, gray shading with steps proportional to logarithm of relative frequency, higher values for lighter shadings) and as distributions over a range of heights (aiv and biv).

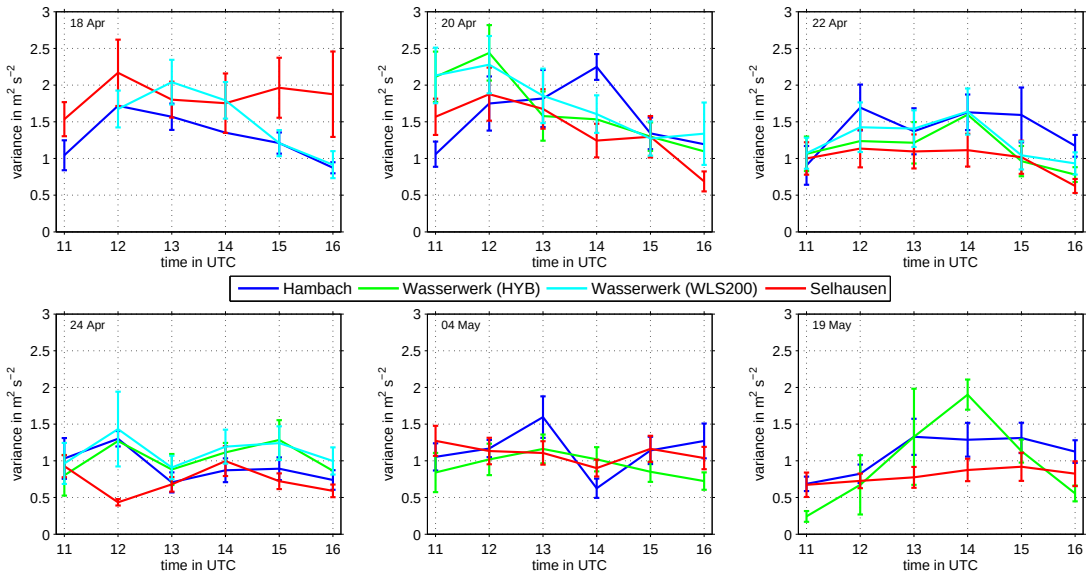

Figure 10. Vertical velocity variances (hourly profiles averaged over $z_{max} \pm 250$ m) at the three locations with error bars displaying the statistical error according to Lenschow et al. (1994) for all six days (different panels).

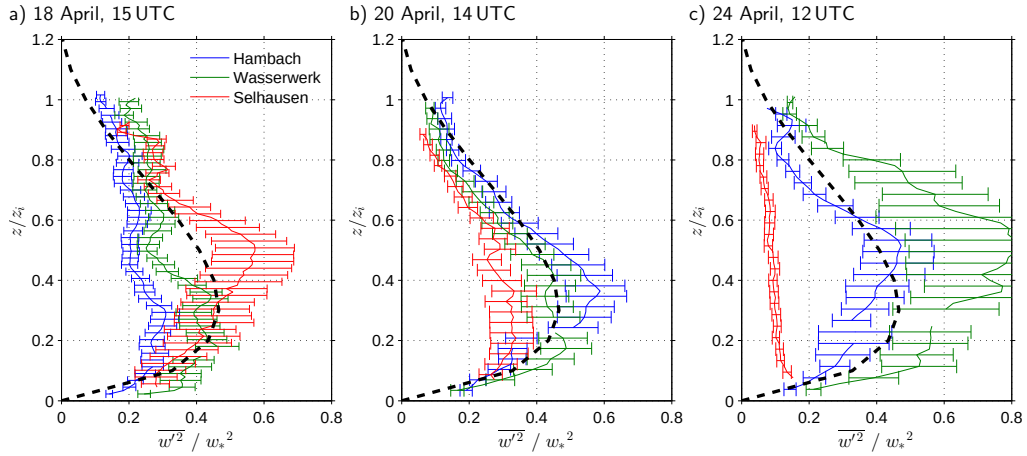

Figure 11. Normalized variance profiles with error bars (statistical error according to Lenschow et al., 1994) for three time periods (local scaling); the black dashed line corresponds to the fit of Lenschow et al. (1980) , Eq. 1.

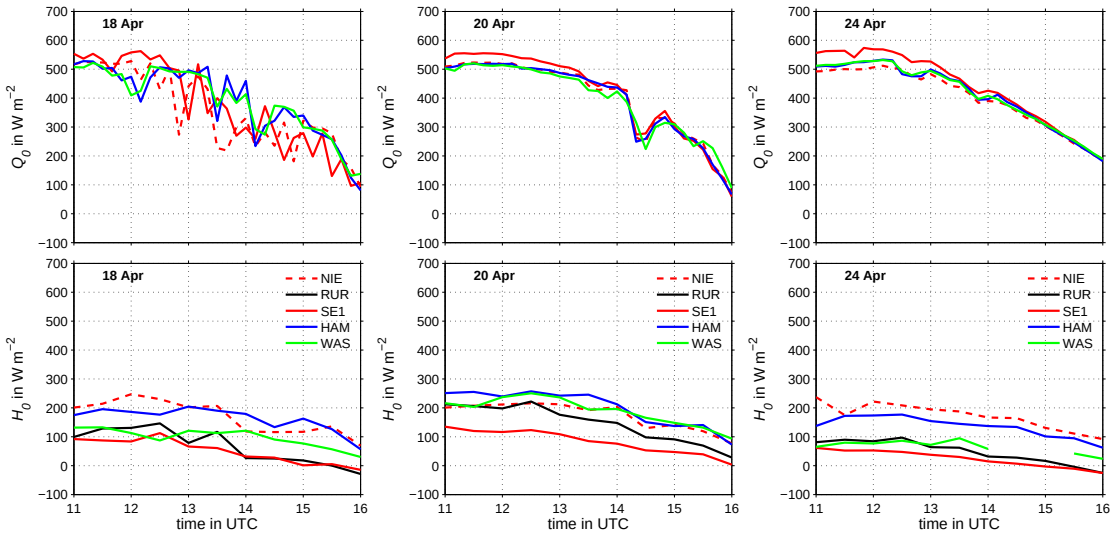

Figure 12. Net radiation (Q_0, upper row) and surface sensible heat flux (H_0) at the five energy balance stations (NIE - Niederzier; RUR - Ruraue; SE1 - Selhausen; HAM - Hambach; WAS - Wasserwerk, cf. Fig. 1) for three days with significant spatial differences of vertical velocity variances.

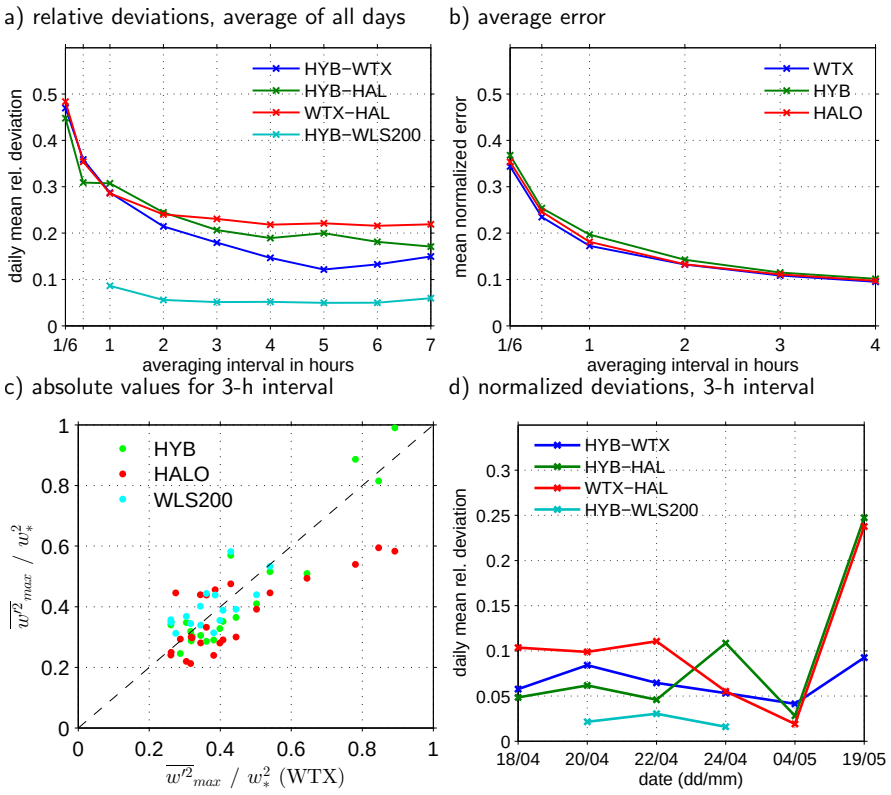

Figure 13. Relative deviations between $\overline{w'^2}_{max}$ time series of each two lidars, averaged daily and over all days (a) and statistical error for each instrument, normalized with the respective $\overline{w'^2}_{max}$ time series (b), given as a function of the averaging interval used for the calculation of the variance profiles; absolute values of $\overline{w'^2}_{max}$ / w_*^2 for the 3-h averaging interval for HYB, HALO, and WLS200 as a function of $\overline{w'^2}_{max}$ / w_*^2 for WTX (c); deviation normalized with w_*^2 for 3-h averaging interval for each day (d).

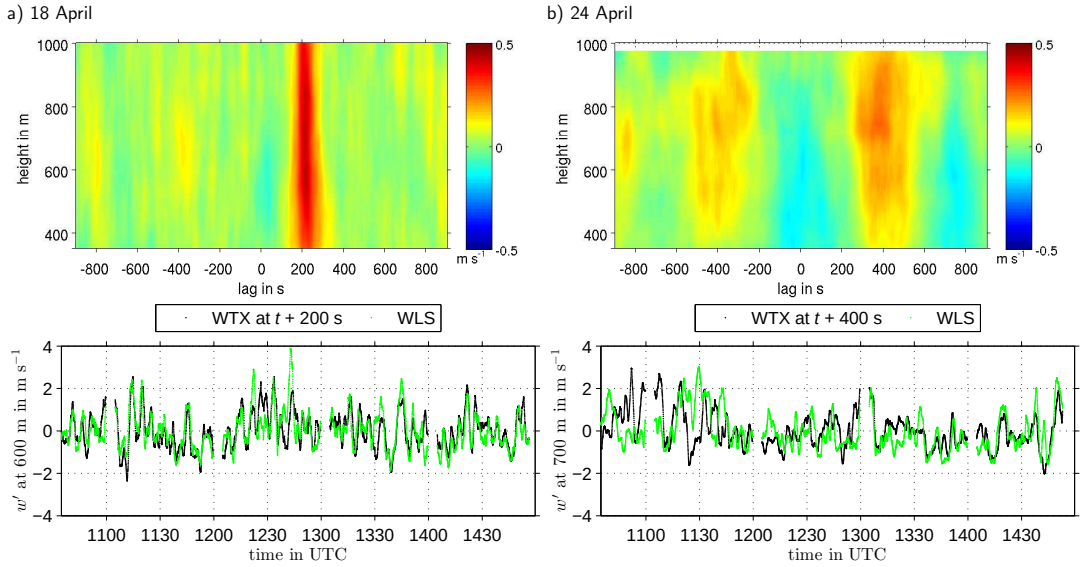

Figure 14. Cross correlation functions between w' time series (10:30–15:00 UTC) at Hambach and Wasserwerk (WTX and WLS200, respectively) for all range gates between 380 m and 1000 m (upper row) and w' time series ($\pm$50-s running average) for both lidars at one range gate (lower row) on 18 April (a) and 24 April 2013 (b).

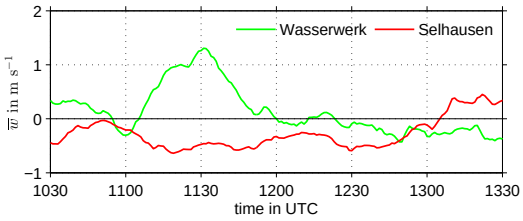

Figure 15. Mean vertical velocity (running average of 60 min) at 700 m ($\pm$1 range gate) at Wasserwerk and Selhausen on 24 April.

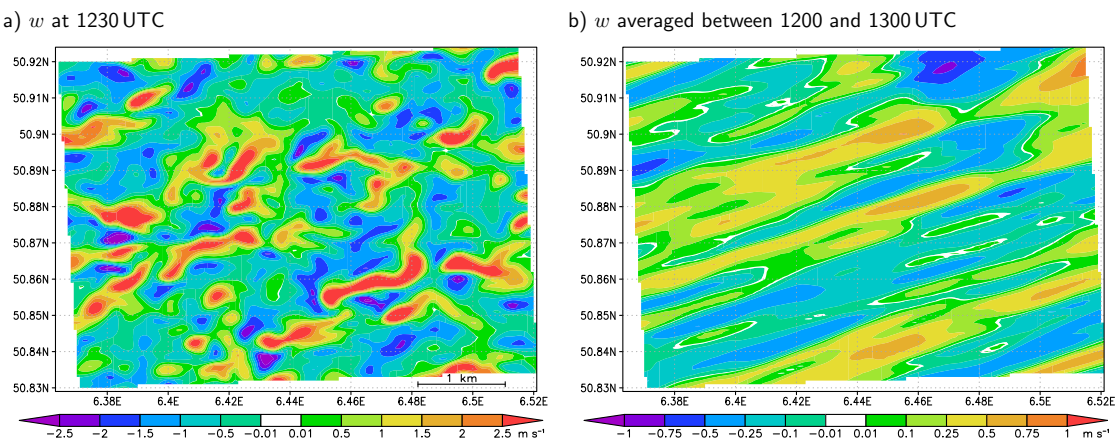

Figure 16. Vertical velocity at 600 m on 24 April 2013 from LES model output: (a) instantaneous, (b) averaged field.