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 measure-ments 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)

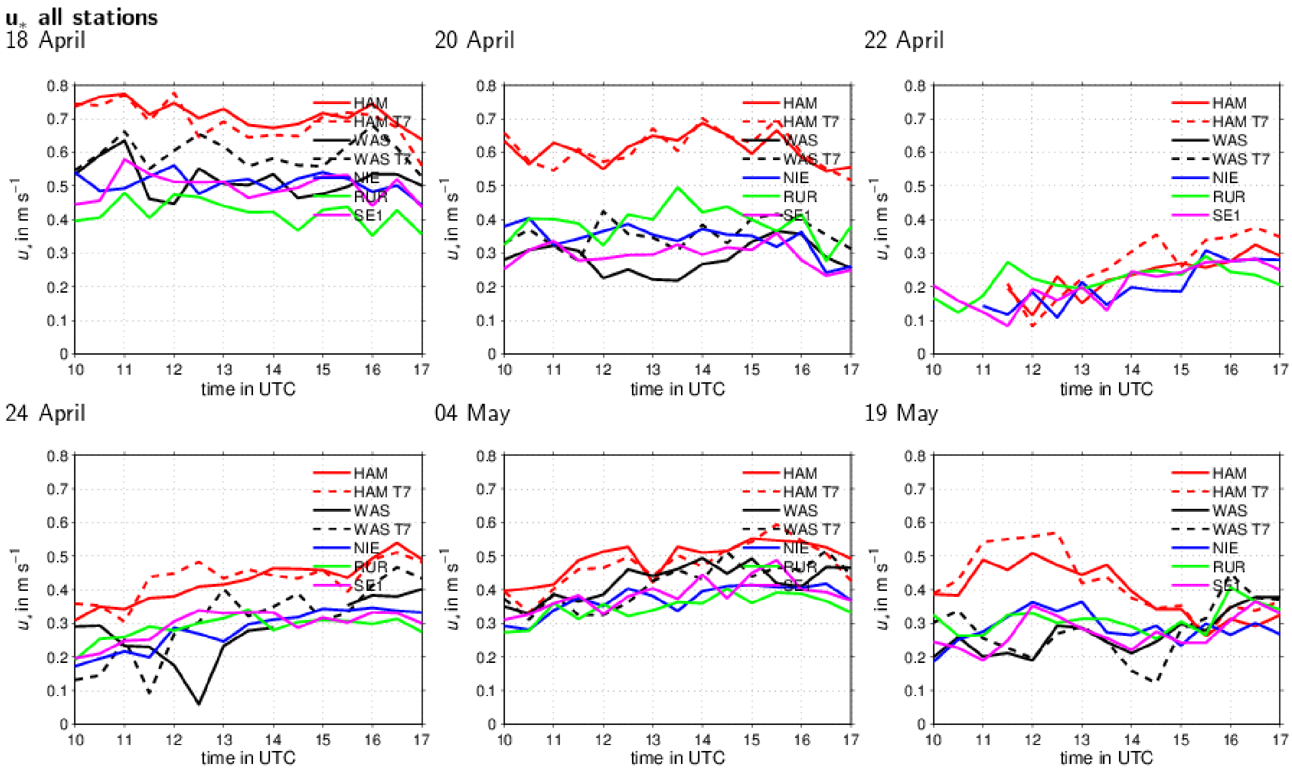

Fig. S1: Diurnal time series of friction velocity u_* for all energy balance stations as well as from turbulence masts at Hambach and Wasserwerk (HAM T7 and WAS T7, respectively).

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.

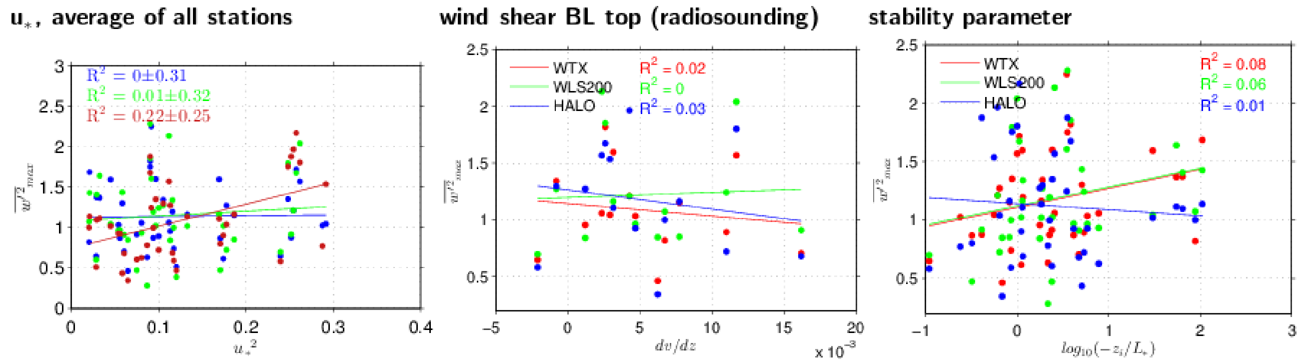

Fig. S2: Correlations of hourly values of vertical velocity variance, averaged over z_{max} ± 250 m and friction velocity u_*, wind shear at CBL top dv/dz, and the logarithm of stability parameter log(-z_i/L) for the six considered days.

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).

<u>Answer to comment (3)</u>
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.”

<u>***Specific comments***</u>

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.

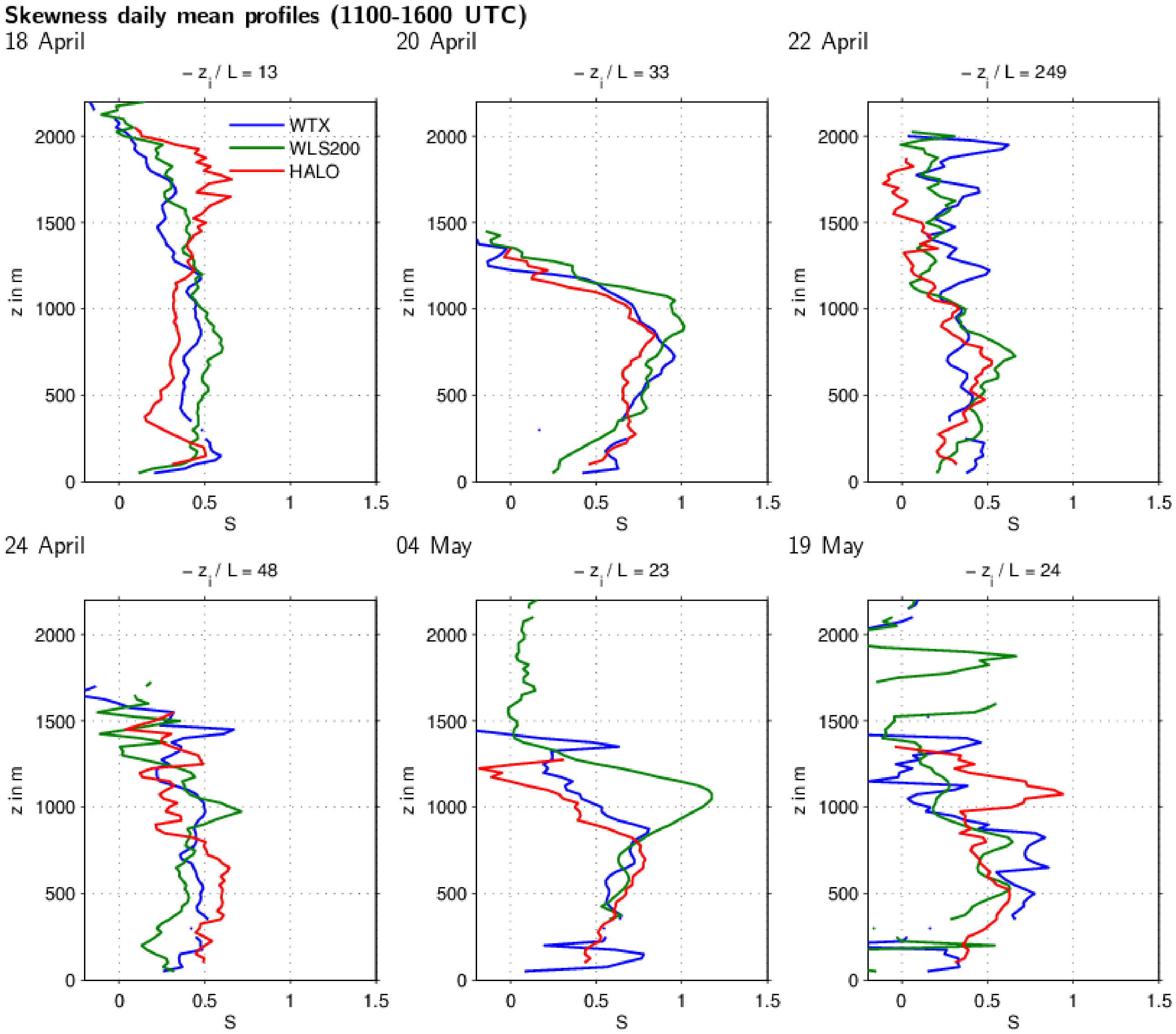

Fig. S3: Daily mean profiles of skewness S; additionally, daily averaged values of -z_i / L are given.

• *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.

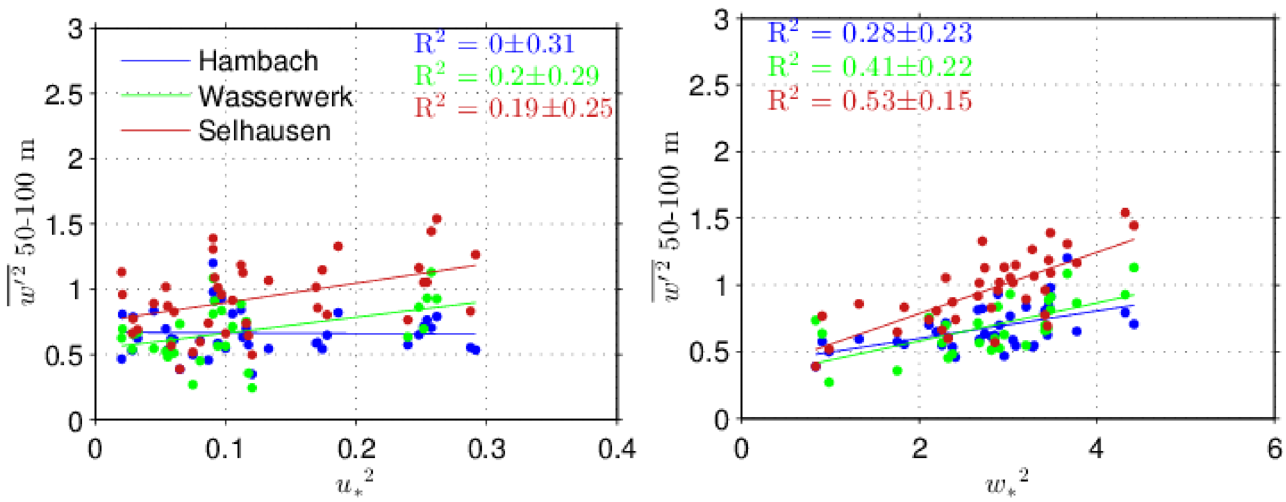

Fig. S4: Correlations between w-variance at lowest range gates and u_* and w_*, respectively (averaged scaling).

• *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 ~ 10 ms^{-1} and ~ 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