Reply to the reviewers

Frederik De Roo and Matthias Mauder

March 2018

We thank the reviewers again for their careful inspection of the manuscript and for their remarks and suggestions. We replaced u^* by u_* (because of the source we refer to). As in the previous round, an italic script indicates our response to the reviewer, and Roman script indicates our changes and additions in the manuscript. For clarity our answers are colored in blue, reviewers' comments in black. In red we have highlighted some paragraphs that we removed from the manuscript, as suggested by the reviewer.

1 Reviewer 1

1.1 Specific comments

1. PCA analysis: you still have not mentioned how many and which variables were used, and if there was any data preparation for the analysis (such as normalization). The manuscript needs to provide all the information necessary to reproduce the work.

   We added the names of the variables explicitly (in the methods we had previously written "The correlation matrix contains exactly the same variables that are plotted in the biplot."). Normalization and centering is understood for a correlation biplot, which we mentioned in the methods, but we now repeat it in the results section as well to be as clear as possible. We also repeat it in the figure caption.

   In the methods: The correlation matrix contains exactly the same variables that are plotted in the biplot: besides EBR, we selected the friction velocity (momentum flux) u^*, the boundary-layer height z_i, the temperature difference between the surface and the tower measurement $T_s - T$, the (total) advection by the mean flow, and the horizontal flux-divergence, with the latter two normalized by the surface flux. This leads to a six-dimensional data space.

   In the results: The biplots are based on the PCA explained in section 2.3, and the data variables are first centered and normalized by their standard deviation.

   In the caption of figure 8: For the PCA behind these correlation biplot, we only use the six variables shown in the plot. For a correlation biplot the variables are centered around their mean and normalized by their standard deviation. The EBR

is used as the supervised variable. The abscissa is the first principal component (PC1) from the PCA, the ordinate is the unit vector perpendicular to PC1 and lying in the plane spanned by EBR and PC1, i.e. it is the vector given by the difference between EBR and its projection on PC1, including subsequent normalization.

2. PCA results: I personally cannot extract any conclusive information from this section. I understand that the biplot can be done even if the variance explained is only 60%, but I still think this is not the best choice of analysis. The fact that the third PC is almost as important as the second tells me that this biplot is a poor choice, as conclusions could change significantly if the third dimension was included. Considering that there are so many options of statistical analysis, I dont understand why using this one. All the conclusions I mentioned above do not depend on the PCA analysis, so it could be removed.

   We follow the suggestion of the reviewer and we have removed the PCA section and biplot, and we abbreviated the discussion on the correlation analysis.

   Abridged discussion: (Correlations with the energy balance ratio)

   We investigate the possible connection between the energy balance ratio, the different flux contributions, and variables such as friction velocity and boundary-layer height. We performed a linear correlation analysis between these variables and the energy balance ratio. We made one restriction on the data set, which is to limit the boundary-layer depth to values larger than 1 km, thereby excluding about 8% of the data, in order to obtain a better representation of the boundary-layer depth (when the boundary-layer depths smaller than 1 km are included, the correlation deteriorates).

   We found that friction velocity and boundary-layer depth cluster are well-correlated with each other, but not with EBR. Although we might have supposed that higher boundary-layer heights will arise if patches are present with vigorous surface heating, however we found that u_* decreased with stronger surface heterogeneity. Closer analysis reveals that the highest boundary layer heights are obtained when the heterogeneity amplitudes are smaller and the domain is more homogeneous. Hence the former clustering can be explained because in our scenario with varying heterogeneity amplitudes the highest boundary-layer height and larger u_* are both obtained for smaller heterogeneity amplitudes. Though advection and flux-divergence correlate well with EBR, they cannot be measured independently and therefore cannot be used as independent predictors.In the literature (e.g. Stoy et al 2013, Eder et al 2015a) a correlation between friction velocity and energy balance closure has been found: a high friction velocity leads to a smaller residual. Typically, a higher friction velocity is correlated to smaller atmospheric instability, and hence roll-like convection instead of cellular convection. Maronga and Raasch (2013) found that boundary-layer rolls "smear out" the surface heterogeneity, leading to an effective surface that looks less heterogeneous, which has been related to a higher EBR (Mauder et al 2007, Stoy et al 2013). Therefore a possible cause for the present low correlation of u_*

with the EBR, could be our range of the stability parameter. For the free convective cases considered here the stability parameter lies below the range where the friction velocity has a high correlation with EBR.

The linear correlation analysis shows that the simulated EBR does not linearly depend on easily measured characteristics. As we have learned from Fig. 5, there can be a good fit between the parameter-averages of two variables, e.g. normalized flux-divergence and energy balance ratio average, despite the fact that the individual data points do not correlate as well. This highlights the importance of testing parameterizations for the energy balance closure problem on the level of a data ensemble, instead of parameterizing on the level of the individual hourly measurements.

Abstract: Finally, we seek correlators for the energy balance ratio and the energy residual in the simulations. The correlation with the friction velocity is less pronounced than previously found, but this is likely due to our concentration on effectively strongly to freely convective conditions.

Conclusions: We did not find a high correlation between the friction velocity and energy balance ratio, but this could be due to the limited range of u_* as we have investigated free convection.

Before taking the analysis out, we investigated it more deeply by means of a so-called supervised correlation analysis (and supervised biplot) which concentrates on the exact correlation with a supervised variable (in our case EBR). However, the supervised biplot does not yield additional insights and therefore we exclude it from the revised manuscript.

Removed from 2.3

For investigating the response of the virtual tower measurements to the changes in the parameters, and for investigating the correlation between the measured variables, a principle component analysis (PCA) is applied. PCA relies on the singular value decomposition (SVD) of the data matrix, which consists of the data points for each of the data variables. Through SVD, the data matrix decomposes into the matrix of the left eigenvectors, a diagonal matrix with the singular values and a matrix with the right eigenvectors. The singular values are ordered by their magnitude, because the square of each singular value is the variance of the data explained by its corresponding eigenvectors. Hence the first eigenvectors with the largest singular values represent the principal components that explain the largest fractions of the data variability. We will present many of our results in correlation biplots introduced by Gabriel (1971,1978). Correlation biplots offer a picture of the relationship between the interdependent variables that make up the data matrix through the PCA method. First of all, for a correlation biplot, each variable is centered around its mean and normalized by its standard deviation. On the normalized data matrix PCA is applied. The data variables are projected into the subspace spanned by the first principal components and then the vectors of the projection within this subspace are plotted in a 2D (or a 3D) biplot, when the two (or three) largest principal components are

chosen. These principal components express the (orthogonal) directions in which the data has the most variability. In a correlation biplot, the inner product between the variable vectors (hence, the product of their length and the cosine of the angle) directly measures their correlation. The scree plots related to the biplots express how much variance is captured by the principal components by plotting the (relative) variance explained by the principal components. The variance explained is a measure for the goodness of the fit. The better the variable can be explained by the first two principal components, the longer the length of the vector of the variable in a two-dimensional correlation biplot, which is at most unity, indicated by the unit circle. For a pedagogical description of biplots see e.g. Greenacre (2010).

As the N-dimensional biplot shows how well the variables are explained by the first N principal components, standard correlation biplots can be used to investigate the clustering of certain variables: when the arrows of the variables lie close to each other (resp. opposite), their (anti-)correlation is high and consequently these variables predict each other relatively well, so the full data can be described in terms of fewer variables. Simultaneously, biplots can be used to investigate the correlation between two variables, or between one variable of particular interest and the other variables, in which case a supervised biplot is even more adequate. Hence, we have chosen to present two-dimensional correlation biplots with a supervised variable, that is, we plot the projection of the variable vectors into the subspace spanned by the first principal component ($PC1$) and the supervised variable — otherwise, for a standard correlation biplot, the variable vectors are projected in the subspace spanned by the first two principal components. By construction the supervised variable still has unit length in the biplot, as it is already completely in this plane and its length was already normalized during the first step of the PCA analysis. In this supervised biplot, the inner product between the two-dimensional projection of any variable with the supervised variable is the exact expression of the correlation between these the variable and the supervised variable. This exact representation does not hold for correlations among any pair of the variables except for the supervised variable. For these pairs of variables the approximation holds up till the extent that the first principal components capture sufficient variability of the data as shown in the scree plot.

In practice, we are most interested in the energy balance ratio (7) and therefore we choose the EBR as the supervised variable. Its correlation (corr) with another variable is given by their inner product, i.e. the product of the length of the variable arrow and the cosine of the angle of the variable arrow with the EBR arrow in the supervised biplot:

$$\mathrm{corr}(EBR, V) = \frac{\mathrm{Cov}(EBR, V)}{\mathrm{Std}(EBR)\mathrm{Std}(V)} = E\hat{B}R \cdot \hat{V} = \mathbf{EBR} \cdot \mathbf{V} = |\mathbf{V}| \cos \upsilon, \quad (1)$$

with $\hat{V}$ the (normalized) variable vector in the full data space of the PCA, $\mathbf{V}$ its projection into the plotted two-dimensional subspace, $|V|$ the arrow length in the

plot and v the angle in the plotted subspace between the variable and **EBR**. Due to the choice of EBR as the supervised variable and the construction of the supervised biplot in the subspace spanned by $PC1$ and **EBR** it follows that

$$\mathbf{EBR} = \mathbf{E}\hat{\mathbf{B}}\mathbf{R}\,. \tag{2}$$

For producing the biplots we have made use of Python3, combining our own routines with standard packages. The correlation matrix contains exactly the same variables that are plotted in the biplot: besides EBR, we selected the friction velocity (momentum flux) u_*, the boundary-layer height z_i, the temperature difference between the surface and the tower measurement $T_s - T$, the (total) advection by the mean flow, and the horizontal flux-divergence, with the latter two normalized by the surface flux. This leads to a six-dimensional data space.

Removed from 3.4:

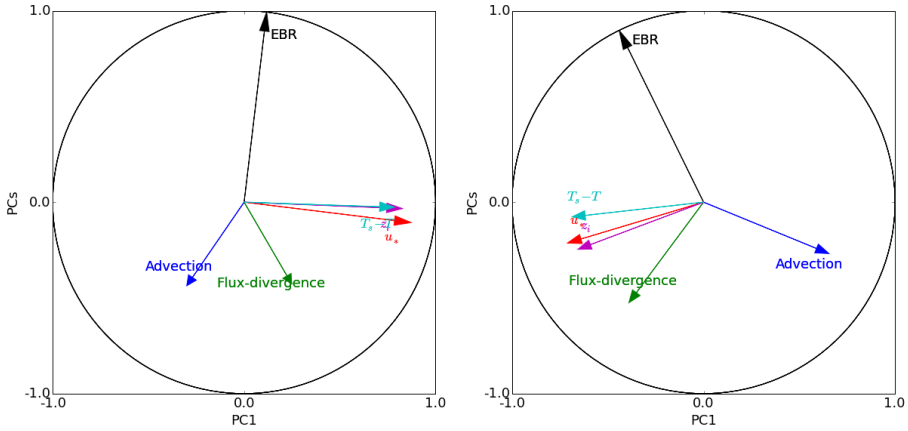
We now turn our attention to the possible connection between the energy balance ratio, the different flux contributions, and other measured variables like the friction velocity. We investigate the correlations for the data sets of Tables 2–3 in separate correlation biplots (Fig. 8) with the EBR as supervised variable. The biplots are based on the PCA explained in section 2.3, and the data variables are first centered and normalized by their standard deviation. We made one restriction on the data set, which is to limit the boundary-layer depth to values larger than 1 km, thereby excluding about 8% of the data, in order to obtain a better representation of the boundary-layer depth within the biplots (when the boundary-layer depths smaller than 1 km are included, the correlation deteriorates). Within the subspace spanned by the first two principal components, the temperature difference between surface and measurement height has the highest correlation to the EBR, however, for the full 6D variable space there is not such a good correlation between EBR and temperature difference, as can be seen in the supervised correlation biplot, where the angle between the EBR and temperature difference tends towards 90 degrees. The temperature difference between the surface and the measurement height can be considered as proportional to the temperature gradient in the surface layer. Correlating EBR with the temperature gradient from the surface with respect to the center of the mixed layer did not improve the correlation.

From the scree plots in Fig. 9 that depict the fraction of explained variance, it can be seen that there is still some scatter in the data and the explanatory strength of the two most important principal components is not very large (a combined 61% and 63% respectively). This is also the reason why we have limited the number of variables that are highlighted in the biplots to only 6 (the ones shown in the biplot) as more variables always introduced additional variance, and the explanatory power of the most important principal components decreased further. The third component has almost the same fraction as the second component. For the supervised biplot, the supervised variable mainly lies along in PC3 with a fraction of PC2 as well.

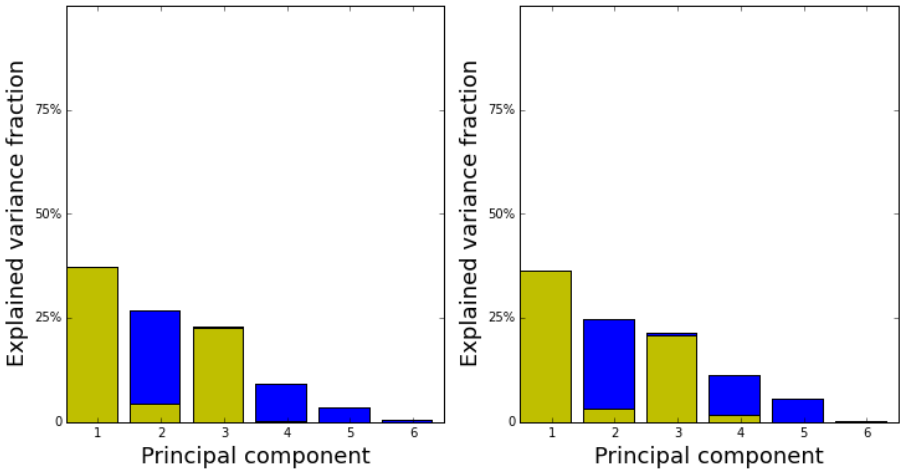
Despite the variance explained of around 60%, plotting the correlation biplot is still possible, it remains a graphical representation of the correlation between the variable vectors, even though it shows that the variance in those variables cannot be simply explained by the first two principal components. In addition, the supervised correlation biplot shows the correlations with EBR exactly, independent of the explained variance. Furthermore, the linear correlation analysis in the biplots is still useful in a few other respects: for one, it shows that EBR does not linearly depend on easily measured characteristics. For another, the biplots represent the correlation between the individual (hourly) data points, whereas the regression lines in Figs. 5 and 7 were obtained from averaging those individual data points over similar tower locations. It is somewhat puzzling that the individual hourly data points leading to the biplots do not show such a high correlation.

Removed from the conclusions

However, we found an unexpected correlation between the energy balance ratio and the difference between sonic temperature and surface temperature. Both can be measured readily by typical micro-meteorological stations. Therefore, this difference could be a promising predictor for a potential energy balance closure correction. Nevertheless, the virtual measurement height remains a critical issue and further investigations are needed with more realistic measurements heights to confirm such a relationship.



Supervised correlation biplots for the kilometer scale heterogeneity (left panel) and the hectometer scale heterogeneity (right panel) with EBR as the supervised variable. For the PCA behind these correlation biplots, we only use the six variables shown in the plot. For a correlation biplot the variables are centered around their mean and normalized by their standard deviation. The EBR is used as the supervised variable. The abscissa is the first principal component (PC1) from the PCA, the ordinate is the unit vector perpendicular to PC1 and lying in the plane spanned by EBR and PC1, i.e. it is the vector given by the difference between EBR and its projection on PC1, including subsequent normalization.



Scree plots of the biplot analyses of Fig. 8 with all 6 principal components. Left panel: order kilometer length scale. Right panel: order hectometer length scale. As far as the explained variance is concerned, both orders of length scale behave very similarly. The fraction of the explained variance that is visible in the supervised biplot is colored in yellow. For a standard two-dimensional biplot, only PC1 and PC2 would contribute to the explained variance in the biplot, but in the supervised biplot we project onto the plane in which the supervised variable lies. In this case the supervised variable is mainly aligned along PC3 with some contribution from PC2 (and PC4 for the hectometer scale), and these fractions build the explained variance of the supervised biplot in addition to PC1.

3. Simulation results: if I understood correctly, you varied the amplitude and the length scale of the heterogeneity resulting in 144 different simulations, but you combined all of them together in your statistics, looking at tower location as the independent variable. This is the reason for the large error bars in Figures 3 and 5, correct? So, why didn't you look into how these two variables (amplitude and length scale) affected your results? Why perform so many simulations? Would the results and conclusions be different if you looked into a smaller set of amplitude and length scale?

   We investigated the amplitude variations in more detail as well, which was the original reason for the size of the set of amplitude variations. The EBR (and other statistics) didn't show a clear pattern in function of the amplitudes (A_x and A_y) and therefore we didn't include it in the article. We now add a comment. As to the size of the set, if we drop amplitudes 0.2 and 0.4 (thinning the set without changing its range) the results are indeed very similar, some data points shift maybe $1-2\%$ in value, but nothing qualitatively different. So with hindsight we could indeed have done with a few simulations less. Nevertheless for analyzing the statistics for the suite of one length scale, since we had these simulations available, we use their variation (in the simulations with different amplitudes) as well.

   Added in 3.2.: We also analyzed the variation of EBR in function of the surface amplitudes (A_x and A_y) but didn't find any clear dependence there.

   We focus on the length scale's order of magnitude, not on the exact length. As the flow is inherently three-dimensional and the surface two-dimensional, the length scale of the surface pattern cannot be exactly captured with a single number (a surface with discs would behave slightly differently from a surface with squares as we have here) so we concentrate on the order of magnitude.

   Added in 2.1.: One suite is focused on kilometer scale heterogeneity, the other on hectometer scale heterogeneity. As the surface heterogeneity is two-dimensional, the length scale of the surface pattern cannot be exactly captured by a single numbers and therefore we concentrate on the order of magnitude of the length scale, and not on the exact length, thus comprising 4 combinations of length scales (L_x and L_y) within the suite of kilometer scale (resp. hectometer scale) heterogeneity.

   The error bars indeed show the spread. We had written in the article: "The error bars on the normalized fluxes denote the spread on the virtual measurements of each tower with respect to the suite. The spread is naturally quite large since at each tower, different amplitudes for the surface heat flux pattern are considered."

4. Is the homogeneous control run also lumped together with everything else in Figures 3-6? Some comparison between homogeneous and heterogeneous cases would be important too.

   We had actually lumped it in because we have a case that satisfies $A_x = A_y = 0$ which could be considered as the "limit of vanishing surface heterogeneity". However, one

could equally well argue that it should be left out. As we will now introduce a direct comparison with the homogeneous results, we decided to update the heterogeneous plots as well, by taking this case with $A_x = A_y = 0$ out. Our results are practically unchanged, as this only forms a small fraction (1/36) of the suite, so we do not need to update the text.

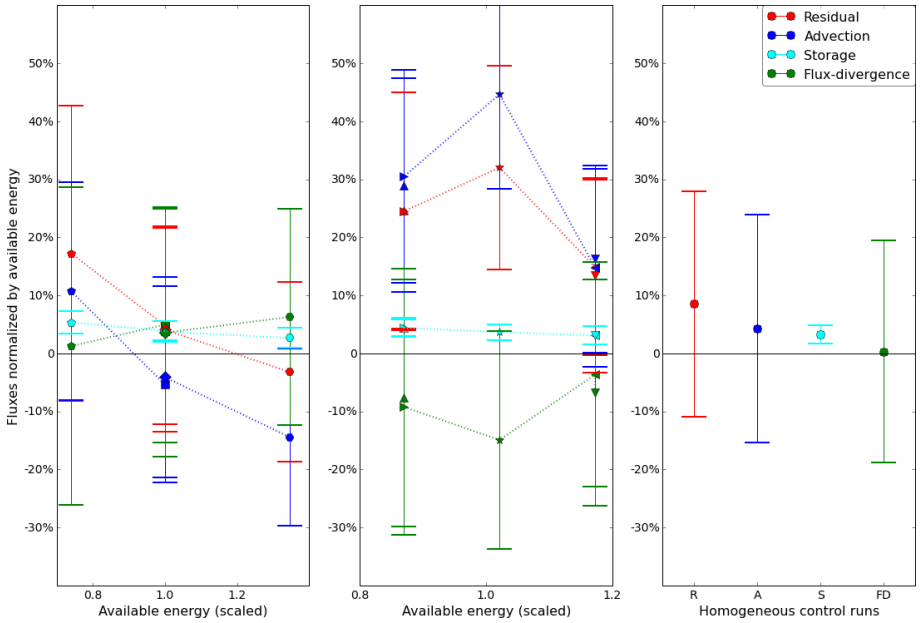
In 3.2.: That is, we averaged over the data with different time stamps and also over all cases within the suite corresponding to the kilometer length scale: this entails $(6 \times 6 - 1)$ variations of the surface flux amplitudes (we do not count the case where both amplitudes are zero, $A_x = A_y = 0$, as this is a homogeneous run) multiplied by 2×2 variations of the heterogeneity length as expressed in table 2.

In order to make a direct comparison between the suite and the homogeneous runs, in Figs. 4 and 6 we have added an additional panel for the homogeneous runs. These include 4 runs with each 9 towers, though all of those are naturally in the same environment. These 36 cases for the homogeneous average are then better comparable with the 35 amplitude variations for each of the heterogeneous towers (though the latter have another factor of 4 length scale variations). We changed the marker for the tower in the coolest patch into a pentagon because we wanted to reserve the circle for the homogeneous runs (we updated all Figures involved).

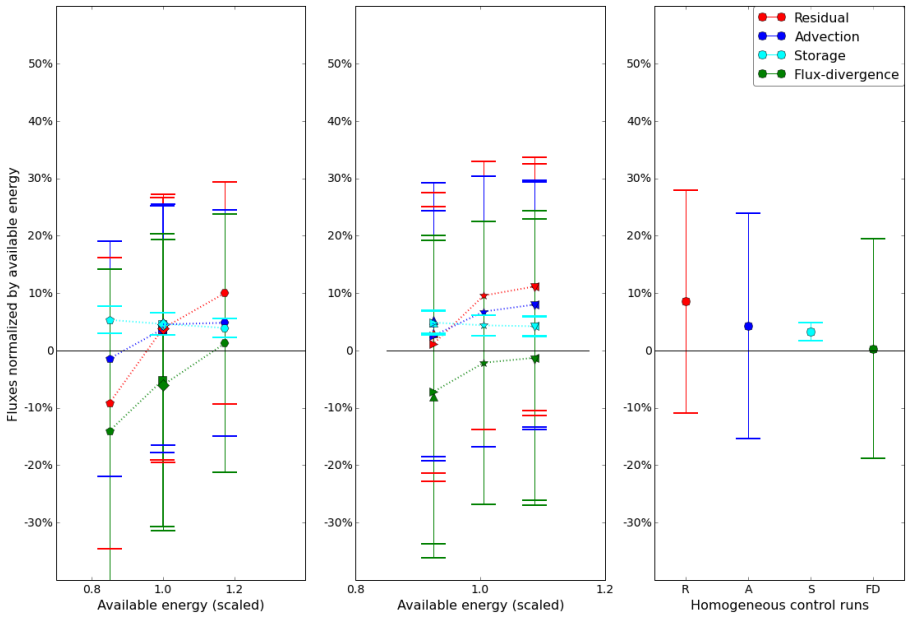
Added in 3.2:

In the right panel we show the data from four homogeneous control runs (with data extraction window and data selection in the same manner as for the heterogeneous runs). Each of these simulations has nine towers as well, but now all towers have the same surface properties. The mean residual (underclosure) of the homogeneous control runs is around 10%, less than for the heterogeneous cases but not negligible. There is significant spread on the results, but the residual is mainly composed of advection and storage. Compared to the towers at the edges (middle panel), which are locally heterogeneous, the homogeneous case is clearly different. Compared to the towers at the centers of the patches (left panel) the homogeneous case has a different average but the difference is still within the spread. It is remarkable that flux-divergence is very small in the homogeneous case, in contrast to the heterogeneous terrain. The negligible flux-divergence for a homogeneous site was also apparent in the desert site of Eder et al (2015).

Updated figures:



Control volume fluxes as a function of available energy (scaled by the median value) for kilometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note that we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 2. The error bars denote the spread over the different cases of surface heterogeneity within the suite of kilometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.



Control volume fluxes as a function of available energy (scaled by the median value) for hectometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 2. The error bars denote the spread over the different cases of surface heterogeneity within the suite of hectometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.

Added in 3.3:

Again, in the right panel we show the data from four homogeneous control runs. However, except for the flux-divergence, the tower responses in heterogeneous terrain of hectometer scale heterogeneity look similar to the homogeneous runs.

5. Figures 3 and 5: "That is, we averaged over the data with different time stamps and also over all cases corresponding to the intermediate length scales formed by the choice of surface flux amplitude and length scale." I don't understand this sentence. What intermediate length scales? What choice of surface flux amplitude and length scale? I understood that each data point is a combination of 144 cases, all at the same location independent of the value of surface flux at that location. "The towers are ordered according to the available energy at their location, for our model setup the available energy is equal to the surface flux." you mean "ordered" in the x-axis? Is it the relative or absolute value of the available energy at their location? The use of the "scaled available energy" as the abscissa is to represent the relative location of updraft/downdraft? I am assuming this from the discussion in paragraph 2 of section 3.2, but it is not really clear. Please clarify how the plot was constructed, and justify the choice of variable for the abscissa.

   In retrospect, we agree that our formulation isn't easy to read. We've rewritten the figure caption. With intermediate length scales we meant, for example, that we have heterogeneity lengths of 1.5 and 3 kilometers within the suite of kilometer scale heterogeneity.

   In 3.2.: That is, we averaged over the data with different time stamps and also over all cases within the suite corresponding to the kilometer length scale: this entails $(6\times6-1)$ variations of the surface flux amplitudes (we do not count the homogeneous case $A_x = A_y = 0$) multiplied by 2×2 variations of the heterogeneity length as expressed in table 2.

   Added in the figure caption: The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 2. The error bars denote the spread over the different cases of surface heterogeneity within the suite of hectometer scale surface heterogeneity.

   As we want to group towers of the same tower type together, we have to take into account the cases with different surface amplitudes. For this reason we scale the available energy at the tower by the mean available energy of the towers for that case. This maps towers of the same tower type at the same location of the abscissa. We added this to the figure captions.

   Added in the figure caption: . The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface

amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts).

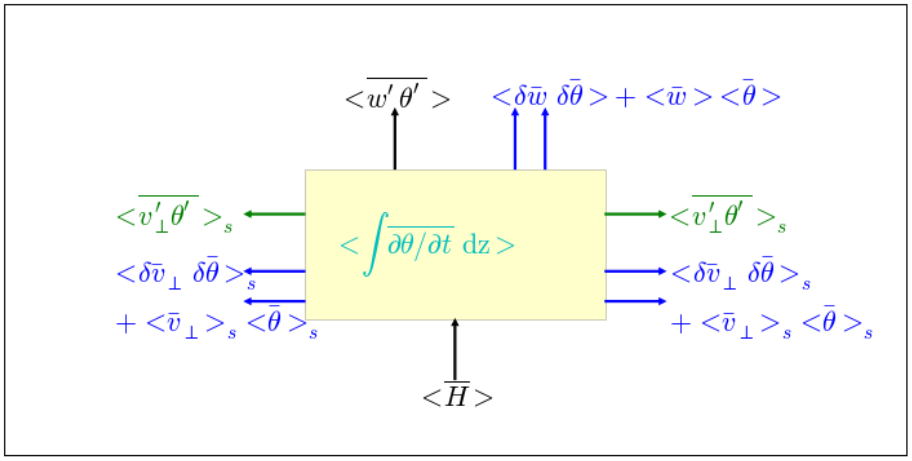
1.2 Suggestions

1. Simulation results: I believe it would be useful to see how the variables of the simulation, for example, velocity, temperature, heat flux, advection, flux divergence, look like as a function of height and how do they vary as a function of horizontal directions due to the heterogeneity. I believe that seeing and discussing the physics of the flow and how these variables behave in the spatial domain could help us visualize what is happening physically. Performing LES is a great opportunity to look into these things that we dont have in field experiments. This is a minor suggestion that I believe would help to interpret the statistics better.

   The turbulent flux, advection and flux-divergence were in this suite only calculated for the designed control volumes so unfortunately we cannot show their horizontal and vertical dependence here (they depend on spatial and temporal correlations and these are expensive to store). We included a horizontal cross-section for the velocity in section 3.1 for some specific simulations, which gives the basic structure of the flow.

2. Equation (3): a sketch of the fluxes and a figure showing where in the 50 by 50m box each term is being calculated would be useful. Another way to avoid copywrite issues is to make a figure yourself. Just a suggestion.

   We added a sketch of the fluxes in the control volume (Fig. 1):



Graphical representation of (1). The control volume is colored in yellow, with horizontal flux-divergence in green, the advection terms in blue, and the storage flux in cyan. The surface flux and the measured turbulent flux are both in black. For clarity the lateral dimension perpendicular to the cross-section is not shown. The direction of the arrows indicate a positive contribution.

1.3 Technical remarks

- Could you add the ranges of surface flux and u* in Tables 2 and 3?

Ranges have been added.

2 Reviewer 2

The authors addressed many of my concerns, including a correction of an equation. However, the second read of the manuscript was not satisfying and I will try to give a constructive reasoning below.

2.1 Comments on the author's reply to the first revision

1. Major comment 1. I commented that the authors should be careful when integrating of large volumes (here 50m height) and not considering the logarithmic temperature profile. They answered that the logarithmic profile only used as boundary condition and thus affects the region from 0 to 5m above ground. I disagree on this point. The logarithmic profile is supposed to be valid in the entire surface layer. With a boundary layer height of 1 km as in the present simulations, the surface layer extents up to say 100 m, whatever it assumed as boundary condition. Please show me the temperature profile between 0 m and 50 m to prove that its shape can be approximated by a linear function.

   - *We apologize for our misunderstanding. The logarithmic profile obtained from MOST is only used as the boundary condition. This is indeed a different topic than the shape of the potential temperature in the surface layer.*
   - *Nevertheless, we do not need to assume that the shape of the potential temperature in the surface layer can be approximated by a linear function (we expand on this below).*
   - *Furthermore, in an implicit LES the numerical integration over the grid boxes is straightforward. We explain the method that we use below, and we have added a paragraph to the methods section to make it clear for the reader as well.*

   Going into more details, we want to clarify our integration method:

   (a) *With "approximated by a linear function", we presume that the reviewer is referring to the application of the trapezoidal rule in numerical integration, where for each sampling interval (in our case each grid cell) of the total numerical integration interval, the function is approached by a piecewise linear function. When the approximated function does not vary wildly within the sampling interval, which is the case for a logarithm (this condition does not have to be satisfied for the entire integration interval at once) errors cancel each other out to a large extent and become smaller when the sampling interval is made smaller.*

   (b) *We actually make use of the "midpoint rule" (interpolation function is piecewise constant) and we have 5 grid points up to 50 meters. Again, numerical errors cancel each other out (partly) and disappear when the sampling intervals are made small. How small the interval has to be, depends on the behaviour of the*

function that is approximated. For the storage term, only the resolved potential temperature has to be used (there is no sub-grid contribution here, they exist only for the fluxes in the LES). Furthermore the time change over 1 hour is averaged out again, which makes it less susceptible to turbulent motion. We only need to make assumptions on the regularity of the potential temperature. Of course, the numerical integration error will be smaller in the case we can assume the shape is closer to piecewise constant, but we do not need to assume a linear shape over the whole integration range in order to apply numerical midpoint integration and we can certainly interpolate the logarithm between z_0 and 50 meters on five grid levels, especially because a logarithm for large values of the argument (z/z_0) is hardly changing and hence for integration purposes the piecewise constant approximation is really a good approximation for large values of the argument $(z \gg z_0 = 0.1\,\mathrm{m})$, because $\lim_{x\to\infty} \frac{d}{dx}\log(x) \to 0$. Furthermore, any remaining numerical error can only be a fraction of an already small term (the storage term does not contribute significantly). Hence we only need to assume that the function can be approximated by a piecewise constant function over each of the sampling intervals (this is considerably more flexible than the linear function over the whole integration interval that the reviewer is referring to).

(c) *In an LES such as PALM that uses "implicit filtering", the representation of the resolved quantities has a so-called "finite volume" character: it is by construction assumed that the scalar in the grid box is already the spatial mean of the grid box, therefore the midpoint rule (which otherwise assumes piecewise constant interpolation function) is the most appropriate, because the LES computed $\theta[k]$ is not $\theta(z = z_k)$ but already $\int_{z_k-dz}^{z_k+dz} \theta(z)dz$, with z_k the height of the grid point k, dz the grid spacing and θ the potential temperature. Therefore, to remain self-consistent with the LES, midpoint integration is the natural choice, and yields the desired result for an implicit LES.*

In the methods:

For the integration of the temperature in the storage term we apply numerical integration with the midpoint rule, which assumes a piecewise constant interpolation function. PALM uses implicit filtering, where it is by construction assumed that the prognostic variable within the grid cell is the volumetric mean of the variable over the domain of the grid cell, therefore the midpoint rule is the most appropriate, because by definition the LES computed $\theta[k]$ is not $\theta(z = z_k)$ but instead

$$\theta[k] = \int_{z_k-dz}^{z_k+dz} \theta(z)dz\,, \tag{3}$$

with z_k the height of the grid point k, dz the grid spacing and θ the potential temperature, and we have suppressed the indices ij for clarity. In this way, the summation of the LES computed discrete profile values is defined to be equal to the

integration of the continuous profile,

$$\sum_{k=1}^{K} \theta[k] = \int_{0}^{z_m} \theta(z) dz \,, \tag{4}$$

with the measurement height $z_m = z_K + dz$.

2. Major comment 2. The authors state that they performed homogeneous control runs, but actually, I do not see any data analysis for these runs. Hence, my comment is still not addressed. I still want to see that the residual is zero for this case! And if not, it might indicate that the time-averaging was not long enough (see below).

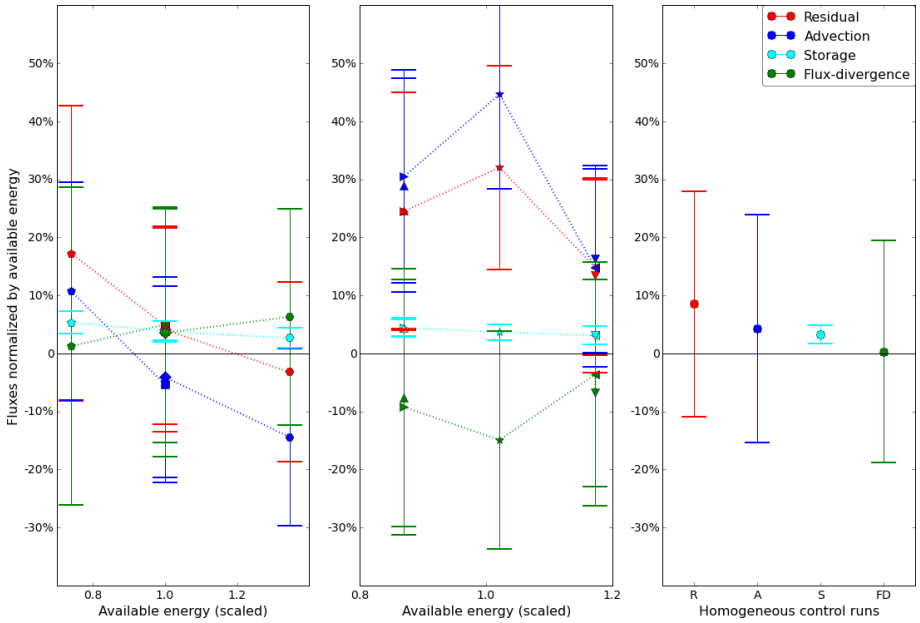
   We added the homogeneous control runs to Figs. 4 and 6 and added some lines to the discussion. We are working along the same lines as the LES literature, as previous studies have found non-zero residuals for homogeneous terrain as well, as mentioned in our introduction (Kanda et al. 2004, Steinfeld et al. 2007, Huang et al., 2008). These studies showed that turbulent organized structures develop also over an ideal homogeneous surface. They are often quasi-stationary and do not propagate with the mean wind, i.e. they violate the ergodic hypothesis. Therefore, the transport associated with these can inherently not be captured by measurement at a certain point. Only a spatial covariance would be able to do so.

   Added in 3.1.:

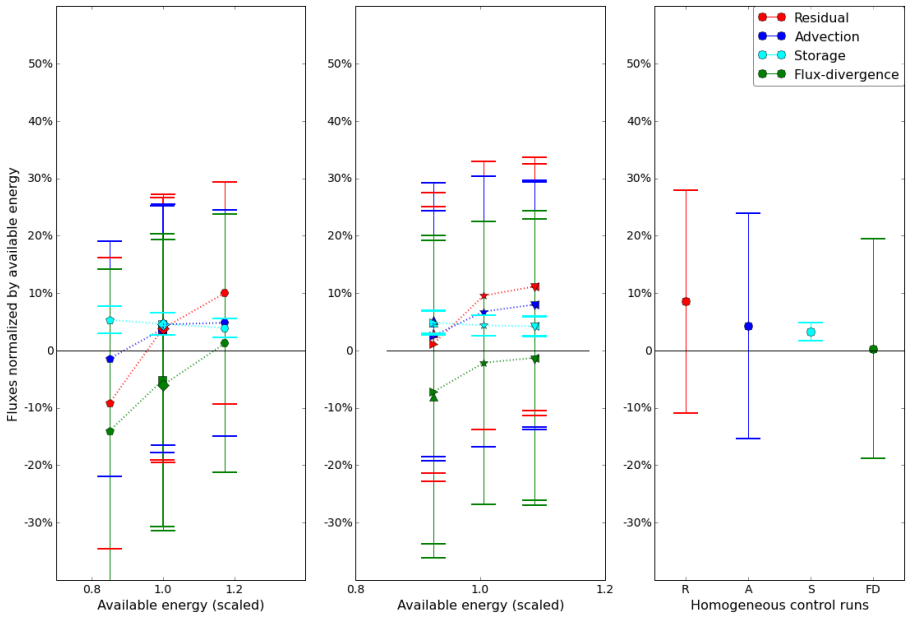
   Due to the absence of a background wind, significant circulation patterns can emerge in the homogeneous case as well. With even longer averaging times a zero mean can be achieved for idealized simulations in homogeneous terrain, but in a real atmospheric boundary-layer this is not possible due to non-stationarity on those timescales.

   Added in 3.2: In the right panel, we show the data from four homogeneous control runs (with data extraction window and data selection in the same manner as for the heterogeneous runs). Each of these simulations has nine towers as well, but now all towers have the same surface properties. The mean residual (underclosure) of the homogeneous control runs is around 10%, less than for the heterogeneous cases but not negligible. There is significant spread on the results, but the residual is mainly composed of advection and storage. Compared to the towers at the edges (middle panel), which are locally heterogeneous, the homogeneous case is clearly different. Compared to the towers at the centers of the patches (left panel) the homogeneous case has a different average but the difference is still within the spread. It is remarkable that flux-divergence is very small in the homogeneous case, in contrast to the heterogeneous terrain. The negligible flux-divergence for a homogeneous site was also apparent in the desert site of Eder et al (2015).

Updated figures:



Control volume fluxes as a function of available energy (scaled by the median value) for kilometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note that we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 1. The error bars denote the spread over the different cases of surface heterogeneity within the suite of kilometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.



Control volume fluxes as a function of available energy (scaled by the median value) for hectometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 1. The error bars denote the spread over the different cases of surface heterogeneity within the suite of hectometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.

Added in 3.3:

Again, in the right panel we show the data from four homogeneous control runs. However, except for the flux-divergence, the tower responses in heterogeneous terrain of hectometer scale heterogeneity look similar to the homogeneous runs.

3. Minor comment 6 and 14. The authors state in their answer that the heterogeneity creates a non-zero friction velocity. Now, strictly speaking the friction velocity is a parameter based on MOST and which does not (strictly speaking) exist when the flow is heterogeneous. I am aware that it is used nevertheless and in LES models it is even used LOCALLY (another violation of MOST). Unfortunately, I have the feeling that the authors use these terms rather loosely and I would be happy to have a more precise use and/or additional comments in the manuscript. In summary: friction velocity is only defined properly when a) the flow is homogeneous and b) when it is defined based on a mean profile. It is zero for free convection. And yes, when it is measured it is ¿ 0 because of the local shear created by thermals, but that is partly because of the violations explained above.

   We consider the friction velocity as obtained through the momentum flux (velocity covariances). We have incorporated the remarks of the reviewer into the manuscript and clarified our usage of friction velocity:

   In section 1.3: as we derive the friction velocity from the kinematic momentum flux (τ_0/ρ), in the same manner as how it is applied in standard eddy-covariance measurements (e.g. Kaimal and Finnigan, 1994):

$$u_*^2 = \tau_0/\rho = \left(\overline{u'w'}^2 + \overline{v'w'}^2\right)^{1/2} . \tag{5}$$

   This definition of friction velocity by the momentum flux is found in general fluid mechanics as well (e.g. Landau and Lifschitz 1959). However, only in homogeneous flow, the friction velocity makes sense as a scaling parameter in Monin-Obukhov similarity theory. Therefore, we want to stress that when the friction velocity is derived from the mean velocity gradient, this is only valid in homogeneous flow. For conditions of free convection in homogeneous terrain the friction velocity derived from the mean velocity is clearly zero (even though free convection flow is locally inhomogeneous). As we do not have homogeneous flow in our study of heterogeneous terrain, we will make use of the momentum flux (5) to derive the friction velocity.

   We also add the comment about the lower boundary in an LES:

   In the methods: It is worth to note that the application of MOST at the first grid point in an LES, is done locally and based on the instantaneous velocity.

4. Language 10. I still do not like the sentence "we plot" because the plot was already made. You do not plot something while I am reading your manuscript. You can say "we show" or something similar instead.

   All instances replaced by "we show".

2.2 Comments on the revised manuscript

1. Eq. 1 and 2: "H" is used in both with different meaning. Even worse, in Eq. 2 it is not explained that is shall be the heat flux, so the reader is forced to believe it to be the Heaviside function.

   We agree that we should avoid the confusion. For clarity, we renamed the Heaviside function $\mathcal{H}$ and we added "surface heat flux H" explicitly, in order not to have to rely on the context.

2. P8 L19. So at which height is the flow sufficiently resolved? You use 50 m without justification.

   The 50 m height is relative to the grid size. We request that the ratio of the subgrid-scale flux to the total flux at that height to be less than 1 %.

   Added in the methods: Demanding that the subgrid-scale flux does not exceed 1 % of the resolved flux, we place our virtual flux measurement at 50 m height.

3. Fig. 2: This data is at 50 m height, correct? This information is missing in the caption. Then, I am very much surprised that the homogeneous control run displays horizontal variations that large (0.5 m/s is not negligible). You report that you averaged over 4 h of time and ideally one expects zero variance in the resulting fields. The fact that there is so much variance suggests that the averaging time is way from being sufficiently long. This then affects all other runs as well and questions the results in general.

   The homogeneous run exhibits turbulent organized structures, which are especially apparent because of the absence of the background wind (they become less pronounced with background wind because the circulation pattern is smeared out, even when rolls don't appear). This is in accordance with literature results for unstable flows over a homogeneous surface (e.g. Kanda et al 2004). At the level an individual realization, the LES predicts that there are still turbulence structures in the homogeneous case for free convection. One can indeed remove structures in the simulations by longer averaging times or by considering ensemble runs. For a true ABL with diurnal evolution, the longer averaging time is however not a possibility and in reality ensemble runs are out of the question too. We do not agree that this affects the heterogeneous simulations because there the circulation is expected to persist due to the heterogeneity. Furthermore, because the averaging over the different cases (35×4) within each suite can be considered as a very large ensemble run, the additional averaging over the suite removes this type of random turbulence.

   Added in the figure caption: ($z = 50\,\mathrm{m}$)

4. P13 L 28: "one magnitude smaller" than what?

   One magnitude smaller than the length scale considered in the preceding sentence (there the kilometer scale is mentioned).

Changed into: "For surface heterogeneity of hectometer scale".

5. General: you state you performed many simulations (288 in total), but I gather that the reader is only confronted with data from six runs. Is this correct or did I understand something wrong here? I was expecting at least some analysis for all runs otherwise you should drop them.

   This is a misunderstanding, we use all the simulations and consider the average over the suite, in order to find out the average effect of surface heterogeneity of a certain scale. We now stress it even more explicitly.

   In 3.2: This is the average of the simulation output belonging to the suite of kilometer scale heterogeneity. In this manner, we investigate the average effect of surface heterogeneity of kilometer scale.

6. Again, I do not understand the biplots. You make some effort to describe them but I would have to read other literature to understand them. I do not think that this is educationally good. Maybe there are more capable readers than me, but if you want to reach a larger audience why not include a very short and easy tutorial on how to read these plots.

   On suggestion of reviewer 1, we have removed the section about the PCA and the biplots altogether. We only include a section in the discussion with the main findings of this linear correlation analysis without referring to biplots.

7. General: Maybe it's just me, but I find the manuscript hard to read. That is also partly because of all the terms like "energy balance ratio" and "available energy (scaled)". Why not use the formula terms in the plots instead?! It would make reading so much easier.

   We believe that energy balance ratio is a better quantity to plot than the turbulent flux, because the former is normalized (and made dimensionless) by the surface flux, and allows better comparison because the different patches have different surface fluxes. For more clarity we already prepare the reader in the methods in addition to our explanation in the results. Moreover, energy balance ratio is common terminology in the literature about the energy balance closure problem (e.g. Stoy et al., 2013 etc.)

   Added in the methods: The energy balance ratio (EBR) of the control volume, which represents the amount of closure of the eddy-covariance measurement with respect to the true surface flux, is given by

$$\mathrm{EBR} = \frac{\langle \overline{w'\theta'} \rangle}{\langle \overline{H} \rangle} \,. \tag{6}$$

The influence of idealized surface heterogeneity on virtual turbulent flux measurements

Frederik De Roo[1] and Matthias Mauder[1,2]

[1] Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Kreuzeckbahnstrasse 19, 82467 Garmisch-Partenkirchen, Germany
[2]Institute of Geography and Geoecology (IfGG), Karlsruhe Institute of Technology (KIT), Kaiserstrasse 12, 76131 Karlsruhe, Germany

Correspondence to: Frederik De Roo (frederik.deroo@kit.edu)

Abstract.

The imbalance of the surface energy budget in eddy-covariance measurements is still an unsolved problem. A possible cause is the presence of land surface heterogeneity, which affects the boundary-layer turbulence. To investigate the impact of surface variables on the partitioning of the energy budget of flux measurements in the surface layer under convective conditions, we set up a systematic parameter study by means of large-eddy simulation. For the study we use a virtual control volume approach, which allows the determination of advection by the mean flow, flux-divergence and storage terms of the energy budget at the virtual measurement site, in addition to the standard turbulent flux. We focus on the heterogeneity of the surface fluxes and keep the topography flat. The surface fluxes vary locally in intensity and these patches have different length scales. Intensity and length scales can vary for the two horizontal dimensions but follow an idealized chessboard pattern. Our main focus lies on surface heterogeneity of the kilometer scale, and one order of magnitude smaller. For these two length scales, we investigate the average response of the fluxes at a number of virtual towers, when varying the heterogeneity length within the length scale and when varying the contrast between the different patches. For each simulation, virtual measurement towers were positioned at functionally different positions (e.g. downdraft region, updraft region, at border between domains, etc.). As the storage term is always small, the non-closure is given by the sum of the advection by the mean flow and the flux-divergence. Remarkably, the missing flux can be described by either the advection by the mean flow or the flux-divergence separately, because the latter two have a high correlation with each other. For kilometer scale heterogeneity, we notice a clear dependence of the updrafts and downdrafts on the surface heterogeneity, and likewise, we also see a dependence of the energy partitioning on the tower location. For the hectometer scale we do not notice such a clear dependence. Finally, we seek correlators for the energy balance ratio in the simulations. The correlation with the friction velocity is less pronounced than previously found, but this is likely due to our concentration on effectively strongly to freely convective conditions.

1 Introduction

1.1 The role of landscape heterogeneity in the energy balance closure problem

The interpretation of the turbulent fluxes of latent and sensible heat at the Earth's surface still suffers from the unresolved energy balance closure problem of the eddy covariance (EC) measurement technique. That is, the measured turbulent fluxes do not close the available energy at the earth's surface (e.g., Foken, 2008; Leuning et al., 2012). There is an ongoing debate whether the missing energy can perhaps be solely described by additional missing terms related to energy conversion and storage or that the imbalance is a consequence of measurement errors in the velocity measurement due to flow distortion from the sonic anemometer pins. With respect to flow distortion, Horst et al. (2015) quoted an error of maximal 5%, but Kochendorfer et al. (2012) and Frank et al. (2013) claimed an error up to 15%. In response to the 15% error, one of us (Mauder, 2013) has pointed out some counterevidence, and a recent modeling study by Huq et al. (2017) on flow distortion did not find evidence for such large errors either. In short, it is unlikely that the previously mentioned issues can explain the fact that very different sites around the world often exhibit an imbalance of more than 20% (e.g., Wilson et al., 2002; Hendricks-Franssen et al., 2010; Stoy et al., 2013).

In fact, the studies by Mauder et al. (2007) and Stoy et al. (2013) have shown that a common property among sites that do not close the energy balance, is a more pronounced surface heterogeneity on the landscape-scale. This motivates us to investigate the energy balance closure problem in the context of landscape heterogeneity. Moreover, Stoy et al. (2013) also found a good correlation between the friction velocity (u_*) and the energy balance closure. This result was reproduced by Eder et al. (2015b) by means of a study combining Doppler wind LiDAR and EC tower data. The same correlation has also been noticed in a recent year-long large-eddy simulation (LES) by Schalkwijk et al. (2016) and in an idealized LES study by Inagaki et al. (2006). In addition, the study of Eder et al. (2015b) could relate the energy balance residual to the mean gradients in the lower boundary-layer, thereby providing more evidence for the connection between the energy imbalance and the presence of quasi-stationary structures in the boundary layer. These circulations typically arise in heterogeneous terrain but may also develop over a completely homogeneous surface to a lesser extent, depending on the atmospheric stability regime, due to self-organization. Persistent updrafts and downdrafts tied to the landscape heterogeneity have been found e.g. by Mauder et al. (2010) during the 2008 Ottawa field campaign. In the case of cellular convection in heterogeneous terrain the distinction between the primary and the secondary circulation becomes blurred, when the convection cells are tied to the landscape heterogeneity.

1.2 The influence of landscape heterogeneity on the boundary-layer structure

The influence of heterogeneous landscapes on properties of the atmospheric boundary-layer has already been investigated for a few decades with numerical models, primarily large-eddy simulation. We will summarize a few results that are relevant to the non-closure of the energy balance. Avissar and Chen (1993) obtained significant mesoscale fluxes tied to the terrain heterogeneity. These mesoscale fluxes are carried by the vertical wind of the meso-scale circulations, however, they are not present at the ground level. Raupach and Finnigan (1995) also found that surface heterogeneity induces boundary-layer motions, nevertheless the area-averaged properties, including the fluxes, were not significantly influenced by the heterogeneity or

the circulation. At the first glance, both statements appear in conflict with a generic influence of the landscape heterogeneity around a measurement site on the energy balance closure.

On the other hand, Shen and Leclerc (1995) found that the horizontally averaged variances and covariances were influenced by land surface heterogeneity with scales smaller than the boundary-layer depth. This was also confirmed by Raasch and Harbusch (2001). This apparent contradiction can be explained by the fact the resolution of these models was coarse due to computational restrictions at that time, which has a few implications. Firstly, from continuity we indeed expect no vertical meso-scale transport by advection with the mean flow at the lowest grid point representing the lower surface, since w=0 due to the rigid no-slip boundary, but horizontal flux-divergence plays a role, too. Secondly, we should keep in mind that areally averaging over sufficiently large distances represents a form of spatial filtering due to the coarse resolution. Steinfeld et al. (2007) argued that a spatial filtering method will yield energy balance closure, whereas single-tower temporal averaging of the sensible heat flux signal in heterogeneous domain suffers from low-frequency contributions due to the shifted co-spectrum.

In summary, the previously mentioned studies showed that landscape hetereogeneity can induce mesoscale motions in the boundary-layer, especially for heterogeneity of length scales larger than the boundary-layer height. By using a large-eddy simulation model coupled to a land-surface-scheme, Patton et al. (2005) investigated striplike heterogeneities between 2 and 30 km. They found that the heterogeneities with length scales of 4 to 9 km were the most influential in altering the structure of the boundary-layer. A similar coupled model approach was used by Brunsell et al. (2011) to study three heterogeneity scales (approximately $10^{-1}z_i$, z_i, 10 z_i, with z_i the boundary-layer height). They found that only in the surface layer the length scale of the heterogeneity affected the spectral signature of the turbulent heat fluxes, and signals appeared blended in the mixed layer. Still, for the heterogeneity length of $10z_i$, secondary circulations arising from surface heterogeneity that extend through the whole boundary-layer were found. Furthermore, Brunsell et al. (2011) found that the partitioning between latent and sensible heat was affected by the scale of heterogeneity as the simulations for the intermediate scales led to a higher Bowen ratio. Since the intermediate scales (of scale z_i) appear more heterogeneous than the small or the large scales, this points toward the dominant influence of the sensible heat flux. Charuchittipan et al. (2014) also suggested to ascribe a larger fraction of the residual to the sensible heat flux than to the latent heat flux. The influence of synthetic surface heterogeneity on the Bowen ratio was also investigated by Friedrich et al. (2000) who found a non-linear response of the aggregated Bowen-ratio on the underlying land-surface distribution. Bünzli and Schmid (1998) investigated idealized heterogeneity by means of a two-dimensional $E - \epsilon$ model and found good correspondence with an analytical averaging scheme based on the context of a numerical blending height.

Although the above findings indicate that surface heterogeneity at scales of boundary-layer depth and larger can couple to the full boundary layer, surface heterogeneity at scales considerably smaller than the boundary-layer height appears to be blended, as observed by Raupach and Finnigan (1995). Furthermore, Avissar and Schmidt (1998) found that under a mild background wind, the influence of surface heterogeneity is quickly destroyed, in accordance with the findings of Hechtel et al. (1990). However, Maronga and Raasch (2013), who performed LES simulations for the response of the convective boundary layer in realistic heterogeneous terrain, advised that sufficient time and ensemble averaging is needed to extract the heterogeneity-induced signal, and they concluded that the upstream surface conditions can still influence the boundary-layer

properties under light winds. Albertson and Parlange (1999) showed that blending of the surface heterogeneity appears even under convective conditions, except for very large heterogeneities. However, Suehring and Raasch (2013) suggest that the blending of the surface follows from insufficient averaging. Therefore an apparent blending does not necessarily imply that small-scale surface heterogeneity could not have an influence on the energy budget at the surface. However, if smaller scales are indeed completely blended in the mixed layer and therefore do not lead to circulations that involve the full boundary-layer, then we cannot expect non-surface layer properties (say, bulk gradients in the mixed layer or entrainment parameters) to correlate well with the energy balance residual. Though, even in the blended case small scale heterogeneity could still influence the surface energy budget through motions in the surface layer when the latter survive half-hour averaging. Indeed, for suburban terrain Schmid et al. (1990) noted significant differences in energy balance ratios at scales of $10^2 - 10^3\,\mathrm{m}$, presumably due to micro-advection between the patches of different surface type.

1.3 Scope of this paper

Acknowledging the connection between the energy imbalance and quasi-stationary flow on the one hand, and quasi-stationary flow and surface heterogeneity on the other hand, we will investigate the effect of surface heterogeneity on the energy balance closure problem in this work. To this end, we will study a series of synthetic idealized landscapes that consist of a chess-board pattern of surface fluxes with different amplitude and different wavelength in the x and the y direction. We will quantify the average influence on virtual tower data, and perform principal component analysis to link the energy balance ratio with surface characteristics, boundary-layer properties and turbulence statistics. To disentangle the influence of the surface heterogeneity from that of the meteorology, we will focus on a set-up of free convection without a synoptic wind (which will effectively lead to strongly to freely convective conditions diagnosed by the virtual towers). As hinted to in Brunsell et al. (2011), in heterogeneous terrain the sensible heat flux appears more important for the imbalance at the intermediate length scales considered in their work, and we shall therefore focus on simulations that are practically dry (we have added a very small moisture flux). In addition, as both the lack of closure and the strength of the circulations are most pronounced for strongly convective conditions, we will likewise focus on (effectively) strongly unstable conditions to free convection, with the unstability parameter $-z/L$ ranging from 1 to 5000. The $-z/L$ is different from ∞ because the convective conditions lead to cellular circulation patterns, which locally induce a friction velocity at the surface, and due to its positiveness, there will also be a horizontally averaged u_* different from zero, as we derive the friction velocity from the kinematic momentum flux (τ_0/ρ), in the same manner as how it is applied in standard eddy-covariance measurements (e.g. Kaimal and Finnigan, 1994):

$$u_*^2 = \tau_0/\rho = \left(\overline{u'w'}^2 + \overline{v'w'}^2\right)^{1/2}. \tag{1}$$

This definition of friction velocity by the momentum flux is found in general fluid mechanics as well (e.g. Landau and Lifschitz 1959). However, only in homogeneous flow, the friction velocity makes sense as a scaling parameter in Monin-Obukhov similarity theory. Therefore, we want to stress that when the friction velocity is derived from the mean velocity gradient, this is only valid in homogeneous flow. For conditions of free convection in homogeneous terrain the friction velocity derived from the mean velocity is clearly zero (even though free convection flow is locally inhomogeneous). As we do not have homogeneous

flow in our study of heterogeneous terrain, we will make use of the momentum flux (1) to derive the friction velocity. From the perspective of the tower measurement, by eddy-covariance measurements alone it cannot be distinguished if a measured u_* follows from the wind aloft or locally from the convection-driven circulation. In addition, the circulation locally leads to advective terms that can influence the energy balance closure: e.g. near an updraft there will be horizontal convergence in the flow field. Even in homogeneous terrain these advective terms can lead to a non-closure of the surface energy budget (e.g. Kanda et al., 2004). Despite the issues related to blending, we will focus on heterogeneity of length scales between $10^2 - 10^3\,\mathrm{m}$ since for these scales the energy imbalance is most pronounced. The intermediate scales of O($10^3\,\mathrm{m}$) are of the order of the boundary-layer depth under typical convective conditions for mid-latitude afternoons, whereas the smaller scales of O($10^2\,\mathrm{m}$) are of the order of the surface-layer height. To keep the terminology more general than typical convection for mid-latitude afternoons, we will refer to them as heterogeneity of kilometer scale and hectometer scale, however. According to the classification of Orlanski (1975) these length scales are at the lower end of the meso-gamma-scale and at the upper end of the micro-alpha-scale, respectively.

Previous investigations with LES on the energy budget had been limited to more regular terrain with at least one homogeneous dimension, see the works of e.g. Kanda et al. (2004), Inagaki et al. (2006), Steinfeld et al. (2007), or Huang et al. (2008). Typically, the storage term was subtracted from the surface flux and only the vertical components of the energy balance were considered: i.e. the turbulent flux and a meso-scale flux (i.e. vertical advection) arising from turbulent organized structures (TOS) or heterogeneity-induced meso-scale motions (TMC). On the contrary, we will also analyze the contribution of the storage flux to the energy imbalance explicitly. Furthermore, the results presented there hold for the domain-averaged imbalance, and the method used is limited to heterogeneous terrain with at least one homogeneous dimension. In this work, however, we can extend the analysis of the energy budget to a full budget of the turbulent fluxes, by including additional terms stemming from horizontal advection by the mean flow. We take full account of all horizontal and vertical energy balance components with a so-called control volume approach, as in Finnigan et al. (2003), Wang (2010), and Eder et al. (2015a). As such, a study of two-dimensional heterogeneous domains becomes possible.

Let us stress again the research questions of this paper. The first aim is to investigate the average influence on virtual flux measurements of land surface heterogeneity in the form of a variable surface heat flux, for a given length scale of the heterogeneity. We focus on length scales of the order of kilometer, and also on length scales of the order of hectometers. The second aim is to correlate the simulated energy balance ratio to various observables that can be obtained from the simulation output and that are also measurable in a realistic setting.

2 Methods

2.1 Simulation set-up

For our simulations we have made use of the LES model PALM (Maronga et al., 2015). More precisely, we ran our simulations with PALM version 3.9. PALM resolves the turbulence down to the scale of the grid spacing, all turbulence below is parameterized by implicit filtering. The closure model in PALM is a so-called 1.5-order closure scheme, where the equations for the

resolved velocities and scalars are derived by implicit filtering over each grid box of the turbulent Navier-Stokes equations, and where an additional prognostic equation for the turbulent kinetic energy is solved. The turbulent kinetic energy in PALM (the sum of the variance of the subgrid-scale velocities) allows to model the energetic content of the subgrid-scale motions, and because it is related to spatial filtering it should not be confused with the typical turbulent kinetic energy in eddy-covariance measurements related to the averaging of a time series. Of course, the latter can be approximated by the resolved kinetic energy in PALM plus the subgrid-scale turbulent kinetic energy. Finally, the Reynolds fluxes that appear in PALM's filtered equations (the spatial covariances of the subgrid-scale quantities) are parameterized by a flux-gradient approach involving the resolved gradient and a diffusivity coefficient that depends on the before-mentioned turbulent kinetic energy, the grid spacing and the height above the lower surface. However, at the first gridpoint above the surface, Monin-Obukhov similarity theory is applied and therefore the turbulence there is completely parameterized. It is worth to note that the application of MOST at the first grid point in an LES, is done locally and based on the instantaneous velocity.

Relevant parameters of the simulation setup are summarized in Table 1, the grid spacing is $10\,\mathrm{m}$ in all three dimensions and the domain size is 6×6 square kilometers in the horizontal, and $2.4\,\mathrm{km}$ in the vertical. Demanding that the subgrid-scale flux does not exceed $1\,\%$ of the resolved flux, we will place our virtual flux measurements at $50\,\mathrm{m}$ height. The boundary conditions of the simulations are periodic in the lateral dimensions. For the velocity we have Dirichlet conditions at the bottom (i.e. rigid no-slip conditions) with zero vertical and horizontal wind. At the top the horizontal velocity is commonly set to the geostrophic wind and the vertical velocity is zero. However, we have turned the geostrophic wind off (this is a homogeneous horizontal pressure gradient): $(u_g, v_g) = (0, 0)$. Nevertheless, due to the differences in surface heating, local pressure gradients will still develop. For potential temperature and humidity we have Neumann conditions at the lower boundary (given by the surface fluxes) and also at the top boundary (where the flux is given by the lapse rate at initialization). The domain is initialized with constant profiles for the velocity (equal to the geostrophic wind for x and y and zero for the vertical velocity). The initial profiles are homogeneous in x and y and for potential temperature (θ) it reads

$$\theta(z) = 300\,\mathrm{K} - 0.01\,\mathrm{K/m} \times (z - 1\,\mathrm{km}) \times \mathcal{H}(z - 1\,\mathrm{km}) \tag{2}$$

with $\mathcal{H}(\cdot)$ the Heaviside function. The top of the domain is situated within a stable inversion layer, which prevents that the turbulence within the boundary-layer is influenced by the vertical domain size. In the lateral dimensions the domain is about 3 to 5 times the boundary-layer depth. For the vertical velocity we have added an a very small subsidence term (leading to a vertical pressure gradient in the equations) for heights above 1 km to counteract the destabilizing influence of the surface heat flux, with the subsidence velocity $w_s = -0.00003\,(z - 1\,\mathrm{km})\,\mathrm{s}^{-1}$ for all simulations. The data are extracted for four hours after two hours of spin-up time. For each hour a data point is collected by averaging over virtual measurements sampled at every second. As our focus lies on the influence of the surface characteristics, we concentrate in the present study on the wind circulations purely generated by the surface heat flux, without complicating the analysis with additional synoptic drivers such as e.g. a geostrophic wind.

We ran two suites of simulations, one suite with 144 simulated cases focusing on surface heterogeneity of the kilometer scale (Table 2), and another suite with 144 simulated cases focusing on surface heterogeneity of the hectometer scale (Table 3).

The simulations are driven by a spatially variable surface sensible heat flux, the variation of which is controlled by a few parameters. More precisely, the surface sensible heat flux H at each surface point (x, y) is determined by

$$H(x,y) = (1 + A_x \, \mathbb{1}(x/L_x)) \, (1 + A_y \, \mathbb{1}(y/L_y)) \, H_0 \,, \tag{3}$$

where $\mathbb{1}$ is a antisymmetric periodic function with period equal to 2, and alternating between -1 and 1, i.e.

$$\mathbb{1}(x) = \sin(\pi x)/\left|\sin(\pi x)\right| \,. \tag{4}$$

The amplitudes of the two-dimensional surface heat flux is given by A_x and A_y and the periods by L_x and L_y. H_0 is the average surface heat flux. In Fig. 2 we show an example of a synthetic surface heat flux as in (3) creating eight patches on the surface with four different values for the surface sensible heat flux. The number of patches depends on the length scale of the heterogeneity.

The main aim of this parameter study is to find out the response of virtual towers in heterogeneous terrain of a certain length scale with variable surface parameters, and for this reason we create two suites of simulations where each simulated case has another combination of the surface parameters. The surface parameters are the length scales L_x and L_y and the amplitudes A_x and A_y. One suite is focused on kilometer scale heterogeneity, the other on hectometer scale heterogeneity. As the surface heterogeneity is two-dimensional, the length scale of the surface pattern cannot be exactly captured by a single number and therefore we concentrate on the order of magnitude of the length scale, and not on the exact length, thus comprising 4 combinations of length scales (L_x and L_y) within the suite of kilometer scale (resp. hectometer scale) heterogeneity. For determining the average behavior under the varying surface fluxes within the suite, no weighing is applied to a particular configuration of the parameters, all amplitudes and length scales under consideration are treated equally. In Tables 2–3 we have summarized the range of the parameters that determine the landscape heterogeneity for each simulated cases within that suite (two suites of 144 simulated cases). The range of the Obukhov length and boundary-layer height expresses the variation of these quantities over the range of the parameter space spanned by the cases of the suite.

2.2 Control volume approach

Within the domain, we have positioned nine virtual control volumes. These control volumes are located at functionally different positions with respect to the surface heterogeneity, as can be seen in Figure 2. Four of them are located at the centers of the patches, four others are located on the borders between the patches, and one is located at the crossing of the four patches. The four at the center are positioned in a site that is homogeneous at the site scale, but heterogeneous at the landscape level. The virtual towers that are located at the borders of the patches are positioned at a site which is not homogeneous at the site level. For every control volume around a virtual tower the size is 5×5 grid points in the horizontal and 5 grid points in the vertical, representing a cube of $(50\,\mathrm{m})^3$. The limits of the control volume are set on the staggered vector grid. The implementation of the energy balance calculation for the control volumes follows the method described in Eder et al. (2015a), which incorporates the approach suggested by Wang (2010). We briefly summarize the main equation, obtained in two steps; first by spatially

averaging over the control volume, and then by additional temporally averaging over 1 hour intervals:

$$\left\langle \overline{H} \right\rangle = \left\langle \overline{w'\theta'} \right\rangle + \sum_{s=1}^{4} \left\langle \overline{\mathbf{v}'_{\perp}\theta'} \right\rangle_s + \left\langle \bar{w} \right\rangle \left\langle \bar{\theta} \right\rangle + \sum_{s=1}^{4} \left\langle \bar{\mathbf{v}}_{\perp} \right\rangle_s \left\langle \bar{\theta} \right\rangle_s + \left\langle \delta\bar{w}\,\delta\bar{\theta} \right\rangle + \sum_{s=1}^{4} \left\langle \delta\bar{\mathbf{v}}_{\perp}\,\delta\bar{\theta} \right\rangle_s + \left\langle \int \frac{\overline{\partial\theta}}{\partial t} \mathrm{dz} \right\rangle . \tag{5}$$

Here H denotes the surface heat flux, x, y and z are the Cartesian coordinates, w the wind component in z direction, θ the potential temperature, $\mathbf{v}_{\perp}$ the velocity vector perpendicular to the lateral faces in the xz- or yz-planes, which are indicated by "s" during the summation over the 4 lateral faces. The angular brackets indicate the spatial average over a face of the cube, either lateral ("s"), top or ground surface, and the δ are the corresponding spatial fluctuations. An overbar indicates a temporal average and the primes the corresponding temporal fluctuations. The term on the left-hand side of the equation is the "true" surface heat flux, whereas the terms of the right-hand side denote the eddy-covariance flux at the top of the control volume, the horizontal flux divergence, the vertical and horizontal advection by the mean flow, the vertical and horizontal dispersive fluxes (Belcher et al., 2012) and the storage of θ in the control volume. The terms of the above formula are clarified in Fig. 1. A positive sign for the directional fluxes means that they point outward of the control volume. The surface flux, however, is considered positive when the flow is from the surface to the atmosphere. Where possible, the Gauß-Ostrogradski theorem[1] has been used to reformulate a divergence within the control volume as a surface term. Due to the choice of a cuboid aligned with the coordinate system for the control volume, the control volume energy balance (5) simplifies further because only the velocity component perpendicular to the faces remain. The energy balance ratio (EBR) of the control volume, which represents the amount of closure of the eddy-covariance measurement with respect to the true surface flux, is given by

$$\mathrm{EBR} = \frac{\left\langle \overline{w'\theta'} \right\rangle}{\left\langle \overline{H} \right\rangle} . \tag{7}$$

From a control volume point of view the net fluxes through the faces are what balances the storage term inside the volume, and in this manner advection effects are automatically included in the energy balance of the volume. Of course, in analogy with measurements, the fluctuations at the top face yield the "virtually measured" turbulent heat flux: first the temporal correlations are calculated, then a spatial average over the upper face of the volume is calculated. The latter average improves the statistical significance of the virtual measurement. Although the subgrid fluxes become small at the height of the control volume, we nevertheless include the vertical component of the subgrid flux into the turbulent heat flux. In this manner we can also capture the highest-frequency correlations. Real data from measurement towers is usually sampled up to $10-50\,\mathrm{Hz}$, whereas for computational efficiency our simulation advances with a time step of one second, i.e. our simulated data is obtained at 1 Hz.

[1]The Gauß-Ostrogradski theorem or "divergence theorem" is a special case of the Stokes-Cartan theorem in differential geometry. For our purposes, we also restrict ourselves to three-dimensional space. We consider a compact volume V with a piecewise smooth boundary S. If $\mathbf{F}$ is a continuously differentiable vector field defined on a neighborhood of V, then:

$$\int_V (\nabla \cdot \mathbf{F})\,\mathrm{dV} = \oint_S \mathbf{F} \cdot \mathbf{dS} . \tag{6}$$

The left side is a volume integral of the divergence of the vector field $\mathbf{F}$ over the volume V, with dV the volume element, and the right side is the surface integral over the boundary of the volume V. $\mathbf{dS}$ is the outward pointing unit normal field of the boundary $S = \partial V$ multiplied by the surface element. For our purposes we take $\mathbf{F} = \mathbf{v}\,\theta$ and V is the control volume described in the text.

A higher sampling frequency would not resolve the turbulence better, as the resolution of the latter is limited by the grid spacing. The part of the total turbulent flux that is not captured directly by the resolved turbulent flux by 1-Hertz sampling is transported by the subgrid turbulent flux. For the advective components we have made a distinction between advection due to the mean flow versus advection due to the horizontal flux-divergence. In complex terrain we do not know a well-defined choice of reference for the base temperature, in contrast to the base temperature in homogeneous terrain that appeared in Webb et al. (1980). Therefore we have avoided introducing a base temperature altogether by adding up the advection by the mean flow components, this means that our advection term is the sum of the horizontal and vertical advection by the mean flow. The virtual measurement height is quite high, but this is due to the vertical resolution and the need for sufficient grid points in the vertical direction to suppress the influence of the subgrid-fluxes whence the turbulence becomes sufficiently resolved. For the integration of the temperature in the storage term we apply numerical integration with the midpoint rule, which assumes a piecewise constant interpolation function. PALM uses implicit filtering, where it is by construction assumed that the prognostic variable within the grid cell is the volumetric mean of the variable over the domain of the grid cell, therefore the midpoint rule is the most appropriate, because by definition the LES computed $\theta[k]$ is not $\theta(z = z_k)$ but instead

$$\theta[k] = \int_{z_k - dz}^{z_k + dz} \theta(z) dz \,, \tag{8}$$

with z_k the height of the grid point k, dz the grid spacing and θ the potential temperature, and we have suppressed the indices ij for clarity. In this way, the summation of the LES computed discrete profile values is defined to be equal to the integration of the continuous profile,

$$\sum_{k=1}^{K} \theta[k] = \int_{0}^{z_m} \theta(z) dz \,, \tag{9}$$

with the measurement height $z_m = z_K + dz$.

3 Results and discussion

3.1 Circulation patterns in heterogeneous terrain

We start our analysis with a discussion of the location of the updrafts and downdrafts in heterogeneous terrain. For this purpose, we concentrate on a few specific cases, more precisely $A_x = A_y = 0.3$ and all four heterogeneity lengths (with $L_x = L_y$). We will take the mean vertical velocity as the simplest proxy for circulation patterns in the boundary layer. In Fig. 3 we show the time-averaged vertical velocity at the height of the control volumes (50 m). We stress that the structures at 50 m extend into the mixed layer above where the absolute velocities become larger (not shown). The reason for the additional time average (over the complete virtual measurement interval of 4 hours) of the already hourly mean data is to remove the drift of the turbulent structures. Due to the absence of a background wind, significant circulation patterns can emerge in the homogeneous case as well. With even longer averaging times a zero mean can be achieved for idealized simulations in homogeneous terrain, but in a real atmospheric boundary-layer this is not possible due to non-stationarity on those timescales.

We notice that for the heterogeneity lengths of O(km), the motions within the mixed-layer clearly reflect the surface pattern, with updrafts concentrated above the hotter patches and downdrafts above the lower patches in the 3-km heterogeneity length, and a little offset in case of the 1.5-km heterogeneity length. However, the structure of the convective turbulence for both kilometer scale are clearly different from homogeneous control run, where typical cellular convection patterns arise (Schmidt and Schumann, 1989), though the hectometer scales are qualitatively rather similar to the homogeneous run. The latter could be a consequence of the blending height. Investigating the heterogeneity lengths of O(hm) with more horizontal detail for the time-averaged w, we do not see clear updrafts or downdrafts tied to the surface heterogeneity. However, in this respect it could be interesting to note that some of the hourly mean vertical velocity (without additional time-average) for the O(hm) appear better related the surface structure. Similar results appear for weaker amplitudes and also when A_x is different from A_y, in which case the dominant pattern is visible along the direction with the larger amplitude (not shown). We can conclude that circulations are tied to the landscape heterogeneity when it is of O(km). For the O(hm) such a correspondence is unclear. However, the latter could be related to the "coarse" grid resolution and the distance from the ground. Indeed, Mauder et al. (2010) found persistent updraft and downdraft regions during the 2008 Ottawa field campaign.

On the topic of circulations driven by a surface conditions that are by design freely convective, we investigate how the domain average of u_* is influenced by the surface heterogeneity. The ratio between the surface flux at the hottest patch and the surface flux at the coolest patch is given by:

$$r = (1 + A_x + A_y + A_x \cdot A_y) \times (1 - A_x - A_y + A_x \cdot A_y)^{-1} \,. \tag{10}$$

The horizontal mean of the friction velocity scales very well with the natural logarithm of this ratio,

$$u_* = -0.046 \ln(r) + 0.384 \,, \, R^2 = 0.85 \tag{11}$$

The remaining spread in u_* does not result from the time stamp or the heterogeneity length scale. The monotonous decrease of u_* in function of the heterogeneity ratio shows that for more homogeneous terrain we will obtain a slightly larger domain averaged u_*.

3.2 Virtual tower measurements for landscape heterogeneity of kilometer scale

In Fig. 4, we look at the response of the towers with respect to their location, corresponding to the simulations summarized in Table 2.This is the average of the simulation output belonging to the suite of kilometer scale heterogeneity. In this manner, we investigate the average effect of surface heterogeneity of kilometer scale. The towers are ordered according to the available energy at their location, for our model setup the available energy is equal to the surface flux. For each tower we have plotted the energy balance residual (available energy minus the turbulent flux), the advection component from the mean flow, the flux-divergence and the storage flux, all normalized by the available energy at the respective tower, with the plot on the left collecting the towers located in the centers of the patches and the plot on the right collecting the towers located at the borders of the patches. The normalized turbulent flux is effectively the energy balance ratio (EBR), but we show the non-closure ($1 - \mathrm{EBR}$), i.e. the normalized energy balance residual, as the latter's magnitude is of the same size as the remaining components. The

normalized fluxes in Fig. 4 are also averaged for all the available data points of the respective tower. That is, we averaged over the data with different time stamps and also over all cases within the suite corresponding to the kilometer length scale: this entails $(6 \times 6 - 1)$ variations of the surface flux amplitudes (we do not count the case where both amplitudes are zero, $A_x = A_y = 0$, as this is a homogeneous run) multiplied by 2×2 variations of the heterogeneity length, as expressed in table 2. The error bars on the normalized fluxes denote the spread on the virtual measurements of each tower with respect to the suite. The spread is naturally quite large since at each tower, different amplitudes for the surface heat flux pattern are considered.

To analyze the tower response in more detail, we have separated the towers at the centers (left panel) from those at the borders (central panel). We notice that most towers show the typical underestimation of the energy balance (i.e., positive energy balance residual), except for the tower located at the warmest spot where there is an updraft. In fact, the closed energy balance for the tower in the warm patch is similar to a result in Eder et al. (2015a) where the energy balance was closed for the site with a pronounced updraft. The residual clearly depends on the location of the tower: towers located at the centers of the patches are located in a more homogeneous environment and they exhibit remarkably smaller residuals, as expected. Towers at the borders have up to 10% more imbalance than the adjacent towers in the center. The tower on the corner of the four patches has the lowest mean closure of only 69%. For towers located at the centers, it is evident that the tower sites are locally homogeneous but there is still a clear imbalance. As a consistency check we note that the similar towers (the two towers in the center of the patches with same surface heating; the two sets of two towers on the borders between patches of similar surface heating) behave similarly. We present some arguments why the regions with updrafts have better closure. Banerjee et al. (2017) investigated the dependence of the aerodynamic resistance on the atmospheric stability for homogeneous terrain. As a consequence a surface with a higher surface heat flux is more efficient in transporting away the surface flux. Therefore, one hypothesis is that when a patch with higher surface flux is coupled to a patch with lower surface flux in heterogeneous terrain, the patch with the higher surface flux transports part of the surface flux of the patch with the lower surface flux, due to its higher efficiency, leading to a net advection of sensible heat from the downdraft region to the updraft region. Another hypothesis is that the shape of the cellular convection cells matters: the updrafts cover a smaller area than the downdrafts. Therefore, as the turbulence structures move across the towers, above a region with preferential updrafts, the likelihood of sampling both the updrafts and downdrafts is higher than above a region with preferential downdrafts.

In the right panel, we show the data from four homogeneous control runs (with data extraction window and data selection in the same manner as for the heterogeneous runs). Each of these simulations has nine towers as well, but now all towers have the same surface properties. The mean residual (underclosure) of the homogeneous control runs is around 10%, less than for the heterogeneous cases but not negligible. There is significant spread on the results, but the residual is mainly composed of advection and storage. Compared to the towers at the edges (middle panel), which are locally heterogeneous, the homogeneous case is clearly different. Compared to the towers at the centers of the patches (left panel) the homogeneous case has a different average but the difference is still within the spread. It is remarkable that flux-divergence is very small in the homogeneous case, in contrast to the heterogeneous terrain. The negligible flux-divergence for a homogeneous site was also apparent in the desert site of Eder et al. (2015a).

As the residual is formed by the sum of advection by the mean flow, storage and flux-divergence, we now turn our attention to these flux components. It turns out that primarily the advection by the mean flow determines the different residuals, but that the flux-divergence has to be taken into account as well for the full picture. In addition, the storage flux also plays a role, but its signature is independent on the location of the tower, and it is always small, which is understandable for our type of surface conditions: there is only a storage flux due to the heating of the air inside the control volumes. For different towers the allocation of the residual to advection by the mean flow versus flux-divergence varies. At first the behavior of the flux-divergence appears irregular. Let us however take a closer look in Fig. 5, where the flux-divergence and advection by the mean flow, resp. are plotted against the energy balance ratio. As in Fig. 4 flux-divergence and advection are normalized by the available energy (i.e. the surface flux in our settings). In the left panel of Fig. 5 we note that the normalized flux-divergence correlates rather well to the normalized turbulent flux, when we look at their average behavior at each tower. For the individual data points the correlation is nevertheless scattered (not plotted). It is somewhat remarkable that both the towers at the center and those at the borders exhibit a similar average behavior. Indeed, the linear regression is very satisfactory when fitting the B-type towers and the C-type towers together. We could have made two separate fits, one for each tower type as in Figure 4, but with only 3 or 4 towers of different functionality a linear regression through those 3 or 4 points would carry less meaning than considering all 9 virtual towers together. If we repeat this linear regression for the advection by the mean flow versus the energy balance ratio we see that the linear correlation fits even better (Fig. 5 right panel) but that it has opposite slope. We had expected that the sum of both components correlates very well with the energy balance ratio, since the storage is small and constant, but it is an interesting result that also flux-divergence and advection separately correlate well with energy balance ratio and consequently also with each other.

Finally, we want to remark that, due to computational constraints, the virtual measurement height in our simulations lies at 50 meters, which is an order of magnitude larger than the typical tower height over short vegetation with comparable surface roughness. This means that our findings for virtual EC towers cannot be directly transferred to real eddy-covariance towers. Other LES studies of the energy balance closure point towards a larger imbalance at higher z-levels, e.g. Steinfeld et al. (2007), Huang et al. (2008), and Schalkwijk et al. (2016). It remains an open question if we can scale the measurement height (as long as it is in the constant flux layer) with the boundary-layer depth and the scale of the heterogeneity. We also analyzed the variation of EBR in function of the surface amplitudes (A_x and A_y) but didn't find any clear dependence there.

3.3 Virtual tower measurements for landscape heterogeneity of hectometer scale

In Fig. 6 we repeat the foregoing analysis for the landscape heterogeneity of hectometer scale, with the parameters in the suite now corresponding to those of table 3. The difference between the towers is much less pronounced here compared to the kilometer scale. Furthermore, the towers in the center of the patches even behave in the opposite manner when the kilometer and hectometer scales are compared. Indeed, for the hectometer scales the cooler patches have a smaller residual, hence better energy balance closure, up to even a mean over-closure for the tower in the coolest patch, whereas the energy balance at the hottest patch is not closed. Another example of the opposite behavior is shown by the flux-divergence. In Fig. 6 it is positively correlated with the normalized residual, and in Fig. 7 we notice that the flux-divergence is now indeed anti-correlated with the

EBR. The advection by the mean flow is again anti-correlated with the EBR, as it was for the kilometer scale. The storage is again roughly constant for all towers. The likely cause for the different behavior between the two scales of heterogeneity would be the blending of the hectometer landscape heterogeneity due to the virtual tower heights of 50 meter. For the surface heterogeneity of O(10^2 m) the flux footprint of each of the towers can cover several of the surface patches, regardless of the type of tower. Again, in the right panel of Fig. 6 we show the data from four homogeneous control runs. However, except for the flux-divergence, the tower responses in heterogeneous terrain of hectometer scale heterogeneity look similar to the homogeneous runs.

3.4 Correlations with the energy balance ratio

We investigate the possible connection between the energy balance ratio, the different flux contributions, and variables such as friction velocity and boundary-layer height. We performed a linear correlation analysis between these variables and the energy balance ratio. We made one restriction on the data set, which is to limit the boundary-layer depth to values larger than 1 km, thereby excluding about 8% of the data, in order to obtain a better representation of the boundary-layer depth (when the boundary-layer depths smaller than 1 km are included, the correlation deteriorates).

We found that friction velocity and boundary-layer depth cluster are well-correlated with each other, but not with EBR. Although we might have supposed that higher boundary-layer heights will arise if patches are present with vigorous surface heating, however we found that u_* decreased with stronger surface heterogeneity. Closer analysis reveals that the highest boundary layer heights are obtained when the heterogeneity amplitudes are smaller and the domain is more homogeneous. Hence the former clustering can be explained because in our scenario with varying heterogeneity amplitudes the highest boundary-layer height and larger u_* are both obtained for smaller heterogeneity amplitudes. Though advection and flux-divergence correlate well with EBR, they cannot be measured independently and therefore cannot be used as independent predictors.In the literature (e.g. Stoy et al., 2013; Eder et al., 2015b) a correlation between friction velocity and energy balance closure has been found: a high friction velocity leads to a smaller residual. Typically, a higher friction velocity is correlated to smaller atmospheric instability, and hence roll-like convection instead of cellular convection. Maronga and Raasch (2013) found that boundary-layer rolls "smear out" the surface heterogeneity, leading to an effective surface that looks less heterogeneous, which has been related to a higher EBR (Mauder et al., 2007; Stoy et al., 2013). Therefore a possible cause for the present low correlation of u_* with the EBR, could be our range of the stability parameter. For the free convective cases considered here the stability parameter lies below the range where the friction velocity has a high correlation with EBR.

The linear correlation analysis shows that the simulated EBR does not linearly depend on easily measured characteristics. As we have learned from Fig. 5, there can be a good fit between the parameter-averages of two variables, e.g. normalized flux-divergence and energy balance ratio average, despite the fact that the individual data points do not correlate as well. This highlights the importance of testing parameterizations for the energy balance closure problem on the level of a data ensemble, instead of parameterizing on the level of the individual hourly measurements.

4 Conclusions

In this work, we have investigated the effect of idealized surface heterogeneity on the components of the surface energy budget measured at virtual measurement towers, by means of large-eddy simulation. By means of a control volume approach, we have decomposed the modeled surface energy budget to highlight its partitioning, and we have shown that the modeled energy balance ratio exhibits values that are found in field experiments. In addition, this approach allows us to investigate the energy balance closure in two-dimensional complex terrain. We have found that for surface heterogeneity with length scale of order kilometer, there is a clear relation between the energy budget components and the location of the tower with respect to the patches of surface heterogeneity. For surface heterogeneity of hectometer scale, the response of the different towers appears to depend to a lesser extent on their respective location. Towers located at the borders between patches with different surface heat flux have worse closure than towers located in the center of a patch. Although storage terms are not negligible, the size of the residual depends mostly on the advection and flux-divergence terms. Remarkably, flux-divergence and advection by the mean flow correlate separately very well with the energy balance ratio, which implies that the EBR can be explained by the advection or flux-divergence only, as the latter two are well correlated among themselves. For the kilometer scale heterogeneities, advection by the mean flow and flux-divergence behave opposite, while they are positively correlated for hectometer scale heterogeneities. We did not find a high correlation between the friction velocity and energy balance ratio, but this could be due to the limited range of u_* as we have investigated free convection.

Appendix A: Example of the heterogeneity length scale of a field site

Even though the focus of this study is on virtual flux measurements, we can look at an example of a real EC measurement site to make a qualitative comparison of these virtual tower measurements with real tower measurements. In a first approximation, the heterogeneity of the landscape around a measurement site can be characterized by the dominant length scale of a suitable surface variable. In Eder et al. (2014), the dominant length scales corresponding to a few sites belonging to the TERENO measurement network (Zacharias et al., 2011) were computed from the Fourier spectrum of the surface roughness. The site of them with the least pronounced topography, the site Fendt, has an effective length scale close to 3 km and a mean EBR of 0.77, which is a typical value for the energy balance ratio (Stoy et al., 2013). The location of the measurement tower in Fendt would correspond to a tower of the central type, and due to its location in the meadow with lower albedo than the forest or the small built-up area, we would assign it to the central tower of the cooler patch. However, the Fourier spectrum of the sensible heat flux may differ from that of the surface roughness. Moreover, the Fourier spectrum for the TERENO site in Fendt exhibits an additional local maximum in its Fourier spectrum of the surface roughness, at 600 m (Eder, pers. comm.). Additionally, it should be noted that even a simplified version of the landscape heterogeneity of Fendt would appear primarily striplike, in contrast the synthetic chessboard pattern here. The EC tower of Fendt is located in a large north-south oriented meadow which is flanked by two forests further away to the west and the east. Despite these apparent differences between our idealised simulations and the real situation at the Fendt site, the EBR of 0.77 is comparable to the EBR of the virtual towers investigated here for the kilometer heterogeneity.

Competing interests. There are no competing interests.

Acknowledgements. This work was conducted within the Helmholtz Young Investigators Group "Capturing all relevant scales of biosphere-atmosphere exchange — the enigmatic energy balance closure problem", which is funded by the Helmholtz-Association through the President's Initiative and Networking Fund, and by KIT. We thank the PALM group at Leibniz University Hannover for their open-source PALM code and their support.

References

Albertson, J. and Parlange, M.: Natural integration of scalar fluxes from complex terrain, Adv Water Res, 23, 239–252, 1999.

Avissar, R. and Chen, F.: Development and analysis of prognostic equations for mesoscale kinetic energy and mesoscale (subgrid-scale) fluxes for large scale atmospheric models, J Atmos Sci, 50, 3751–3774, 1993.

Avissar, R. and Schmidt, T.: An Evaluation of the Scale at which Ground-Surface Heat Flux Patchiness Affects the Convective Boundary Layer Using Large-Eddy Simulations, J Atmos Sci, 55, 2666–2689, 1998.

Banerjee, T., De Roo, F., and Mauder, M.: Explaining the convector effect in canopy turbulence by means of a large eddy simulation, Hydrol Earth Syst Sci, 21, 2987–3000, 2017.

Belcher, S., Harman, I., and Finnigan, J.: The Wind in the Willows: Flows in Forest Canopies in Complex Terrain, Annu. Rev. Fluid Mech., 44, 479–504, 2012.

Brunsell, N., Mechem, D., and Anderson, M.: Surface heterogeneity impacts on boundary layer dynamics via energy balance partitioning, Atmos Chem Phys, 11, 3403–3416, 2011.

Bünzli, D. and Schmid, H.: The influence of surface texture on regionally aggregated evaporation and energy partitioning, J Atmos Sci, 55, 961–972, 1998.

Charuchittipan, D., Babel, W., Mauder, M., Leps, J., and Foken, T.: Extension of the averaging time in eddy-covariance measurements and its effect on the energy balance closure, Bound-Lay Meteorol, 152, 303–327, 2014.

Eder, F., De Roo, F., Kohnert, K., Desjardins, R., Schmid, H., and Mauder, M.: Evaluation of Two Energy Balance Closure Parametrizations, Bound-Lay Meteorol, 151, 195–219, 2014.

Eder, F., De Roo, F., Rotenberg, E., Yakir, D., Schmid, H., and Mauder, M.: Secondary circulations at a solitary forest surrounded by semi-arid shrubland and its impact on eddy-covariance measurements, Agr Forest Meteorol, 211-212, 115–127, 2015a.

Eder, F., Schmidt, M., Damian, T., Träumner, K., and Mauder, M.: Mesoscale Eddies Affect Near-Surface Turbulence Exchange: Evidence from Lidar and Tower Measurements, J Appl Meteorol Clim, 54, 189–206, 2015b.

Finnigan, J. J., Clement, R., Malhi, Y., Leuning, R., and Cleugh, H. A.: A re-evaluation of long-term flux measurement techniques part 1: Averaging and coordinate rotation, Bound-Lay Meteorol, 107, 1–48, 2003.

Foken, T.: The energy balance closure problem: an overview, Ecol Appl, 18, 1351–1367, 2008.

Frank, J. M., Massman, W. J., and Ewers, B. E.: Underestimates of sensible heat flux due to vertical velocity measurement errors in non-orthogonal sonic anemometers, Agr. Forest Meteorol., 171, 72–81, 2013.

Friedrich, K., Mölders, N., and Tetzlaff, G.: On the influence of surface heterogeneity on the Bowen-ratio: A theoretical case study, Theor Appl Clim, 65, 181–196, 2000.

Hechtel, L., Moeng, C., and Stull, R.: The effects of nonhomogeneous surface fluxes on the convective boundary layer: a case study using large-eddy simulation, J Atmos Sci, 47, 1721–1741, 1990.

Hendricks-Franssen, H., Stöckli, R., Lehner, I., Rotenberg, E., and Seneviratne, S.: Energy balance closure of eddy-covariance data: a multisite analysis for European FLUXNET stations, Agr Forest Meteorol, 150, 1553–1567, 2010.

Horst, T., Semmer, S., and MacLean, G.: Correction of a non-orthogonal and three-component sonic anemometer for flow distortion by transducer shadowing, Bound-Lay Meteorol, 155, 371–395, 2015.

Huang, J., Lee, X., and Patton, E.: A modelling study of flux imbalance and the influence of entrainment in the convective boundary layer, Bound-Lay Meteorol, 127, 273–292, 2008.

Huq, S., De Roo, F., Foken, T., and Mauder, M.: Evaluation of probe-induced flow distortion of Campbell CSAT3 sonic anemometers by numerical simulation, Boundary-Lay Meteorology, in press, doi:10.1007/s10546-017-0264-z, 2017.

Inagaki, A., Letzel, M. O., Raasch, S., and Kanda, M.: Impact of surface heterogeneity on energy imbalance, J Meteorol Soc Jpn, 84, 187–198, 2006.

Kaimal, J. C. and Finnigan, J. J.: Atmospheric boundary-layer flows: their structure and measurement, Oxford University Press, 1994.

Kanda, M., Inagaki, A., Letzel, M. O., Raasch, S., and Watanabe, T.: LES study of the energy imbalance problem with eddy covariance fluxes, Bound-Lay Meteorol, 110, 381–404, 2004.

Kochendorfer, J., Meyers, T. P., Frank, J., Massman, W. J., and Heuer, M. W.: How well can we measure the vertical wind speed? Implications for fluxes of energy and mass, Bound-Lay Meteorol, 145, 383–398, 2012.

Leuning, R., van Gorsel, E., Massman, W. J., and Isaac, P. R.: Reflections on the surface energy imbalance problem, Agr. Forest Meteorol., 156, 2012.

Maronga, B. and Raasch, S.: Large-eddy simulations of surface heterogeneity effects on the convective boundary layer during the LITFASS-2003 experiment, Bound-Lay Meteorol, 146, 17–44, 2013.

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., Ketelsen, K., Letzel, M., Sühring, M., and Raasch, S.: The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives, Geosci. Model Dev., 8, 1539–1637, 2015.

Mauder, M.: A Comment on "How well can we measure the vertical wind speed? Implications for fluxes of energy and mass" by Kochendorfer et al, Bound-Lay Meteorol, 147, 329–335, 2013.

Mauder, M., Desjardins, R., and MacPherson, I.: Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem, J Geophys Res, 112, D13 112, 2007.

Mauder, M., Desjardins, R., Pattey, E., and Worth, D.: An Attempt to Close the Daytime Surface Energy Balance Using Spatially-Averaged Flux Measurements, Boundary-Layer Meteorol, 136, 175–191, 2010.

Orlanski, I.: A rational subdivision of scales for atmospheric processes, Bull. Am. Meteorol. Soc., 56, 527–530, 1975.

Patton, E., Sullivan, P., and Moeng, C.: The influence of idealized heterogeneity on wet and dry planetary boundary layers coupled to the land surface, J Atmos Sci, 62, 2078–2097, 2005.

Raasch, S. and Harbusch, G.: An analysis of secondary circulations and their effects caused by small-scale surface inhomogeneities using large-eddy simulation, Boundary-Layer Meteorol, 101, 31–59, 2001.

Raupach, M. R. and Finnigan, J. J.: Scale issues in boundary-layer meteorology: surface energy balances in heterogeneous terrain, Hydr Proc, 9, 589–612, 1995.

Schalkwijk, J., Jonker, H. J. J., and Siebesma, A. P.: An Investigation of the Eddy-Covariance Flux Imbalance in a Year-Long Large-Eddy Simulation of the Weather at Cabauw, Boundary-Layer Meteorol, 160, 17–39, 2016.

Schmid, H., Cleugh, H., Grimmond, C., and Oke, T.: Spatial variability of energy fluxes in suburban terrain, Boundary-Layer Meteorol, 54, 249–276, 1990.

Schmidt, H. and Schumann, U.: Coherent structure of the convective boundary layer derived from large-eddy simulations, J Fluid Mech, 200, 511–562, 1989.

Shen, S. and Leclerc, M.: How large must surface inhomogeneities be before they influence the convective boundary layer structure? A case study, Q J R Meteorol Soc, 121, 1209–1228, 1995.

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, 2007.

Stoy, P., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, J., Merbold, L., Montagnani, L., Papale, D., and Reichstein, M.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape-scale heterogeneity, Agr. Forest Meteorol., 171-172, 137–152, 2013.

Suehring, M. and Raasch, S.: Heterogeneity-Induced Heat-Flux Patterns in the Convective Boundary Layer: Can they be Detected from Observations and is There a Blending Height?—A Large-Eddy Simulation Study for the LITFASS-2003 Experiment, Boundary-Layer Meteorol, 148, 309–333, 2013.

Wang, W.: The influence of topography on single-tower-based carbon flux measurements under unstable conditions: a modeling perspective, Theor Appl Clim, 99, 125–138, 2010.

Webb, E. K., Pearman, G. I., and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, Quart. J. R. Met. Soc., 106, 85–100, 1980.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, R., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., and Verma, S.: Energy balance closure at FLUXNET sites, Agr Forest Meteorol, 113, 223–243, 2002.

Zacharias, S., Bogena, H., Samaniego, S., 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., Papen, H., Priesack, E., Schmid, H., Steinbrecher, R., Rosenbaum, U., Teutsch, F., and Vereecken, H.: A Network of Terrestrial Environmental Observatories in Germany, Vadose Zo J, 10, 955, 2011.

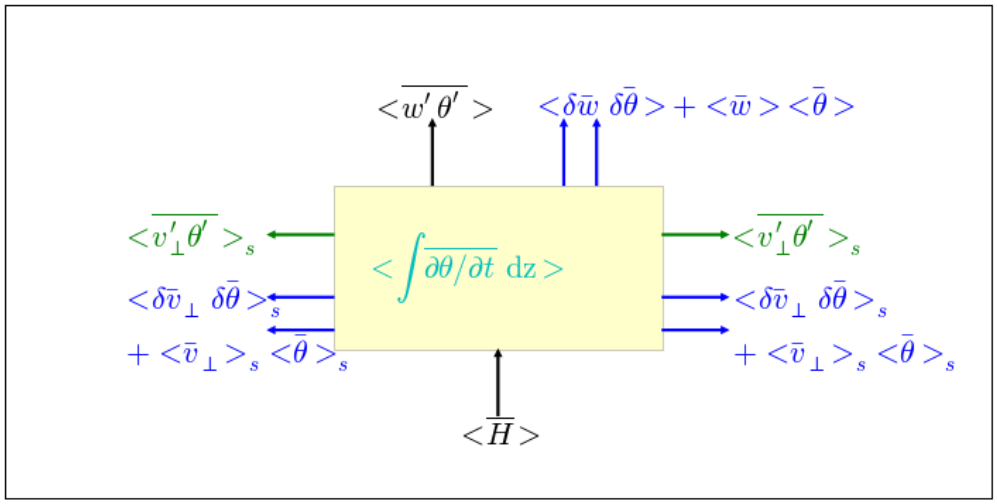


Figure 1. Graphical representation of (5). The control volume is colored in yellow, with horizontal flux-divergence in green, the advection terms in blue, and the storage flux in cyan. The surface flux and the measured turbulent flux are both in black. For clarity the lateral dimension perpendicular to the cross-section is not shown. The direction of the arrows indicate a positive contribution.

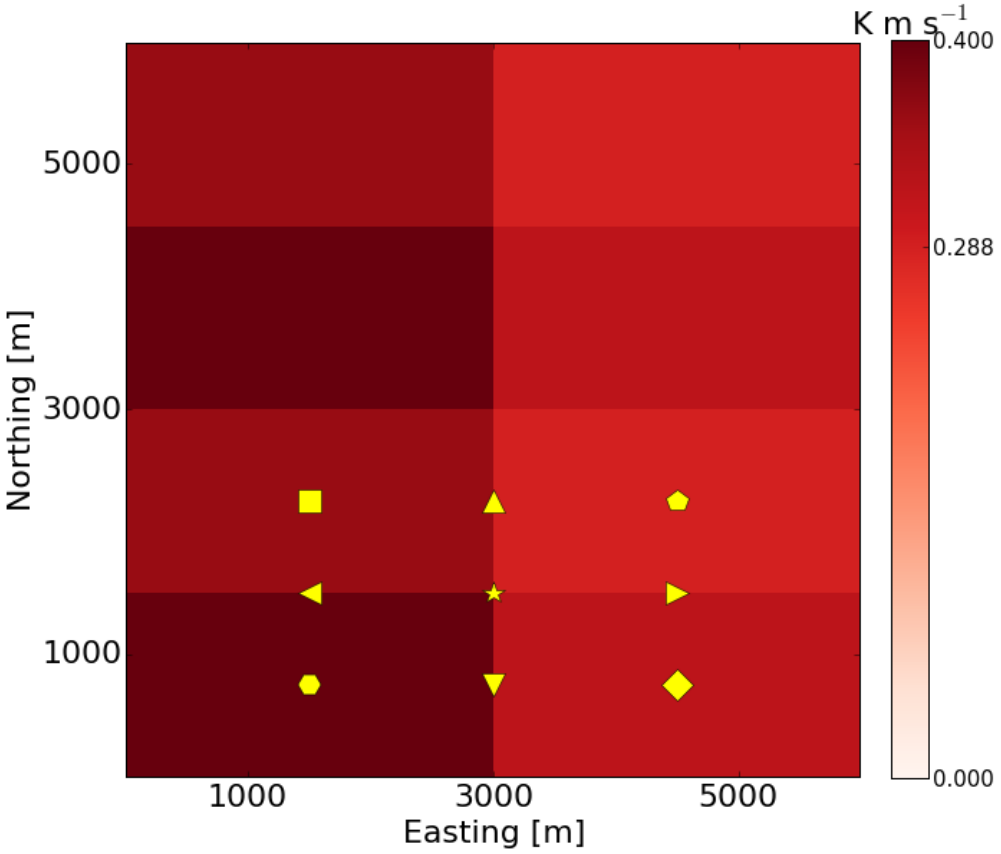


Figure 2. Fixed location of the virtual towers for the kilometer scale heterogeneity. The surface heat flux pattern of this example corresponds to $L_x = 3000\,\mathrm{m}$, $L_y = 1500\,\mathrm{m}$. Please note that all the control volumes have the same shape of $5 \times 5 \times 5$ grid points, the symbols are only to distinguish the different types of towers. For the hectometer scale heterogeneity, the towers are located at the similar positions in or in between the patches, only the patches are smaller. The towers fall into two classes: those located at the center of the patches and those located at the borders.

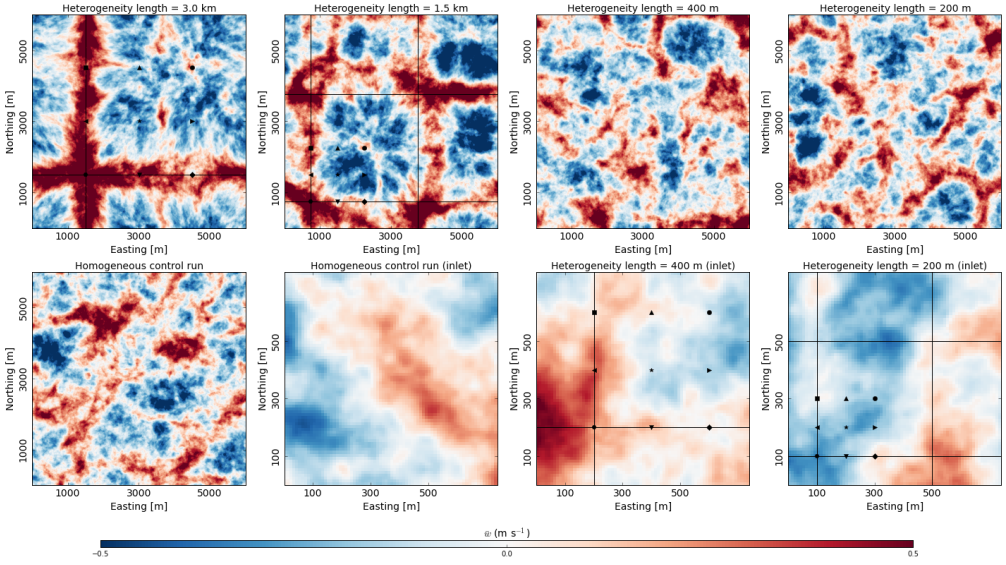


Figure 3. Analysis of the circulation patterns induced but the surface heterogeneity by means of the vertical velocity (w) averaged over the 4-hr data output, including a homogeneous control run. The results are for a particular surface amplitude of $A_x = A_y = 0.3$ and with $L_x = L_y$ ($z = 50\,\mathrm{m}$). For reference the tower locations are indicated as well, as is the center of the "hot" patches by means of a black line. The plots of the whole domain for O(hm) show their reminiscence with the homogeneous control run. For the O(hm) heterogeneity we show an inlet around the towers, because the correspondence with the surface heterogeneity is otherwise hard to visualize due to the smallness of the heterogeneity length.

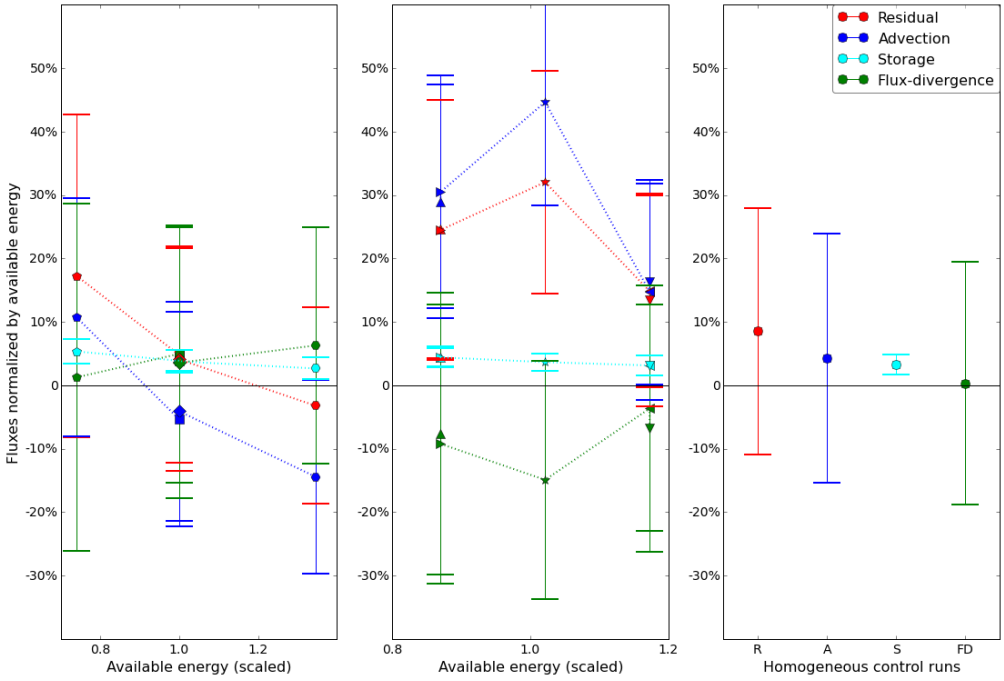


Figure 4. Control volume fluxes as a function of available energy (scaled by the median value) for kilometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note that we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 1. The error bars denote the spread over the different cases of surface heterogeneity within the suite of kilometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.

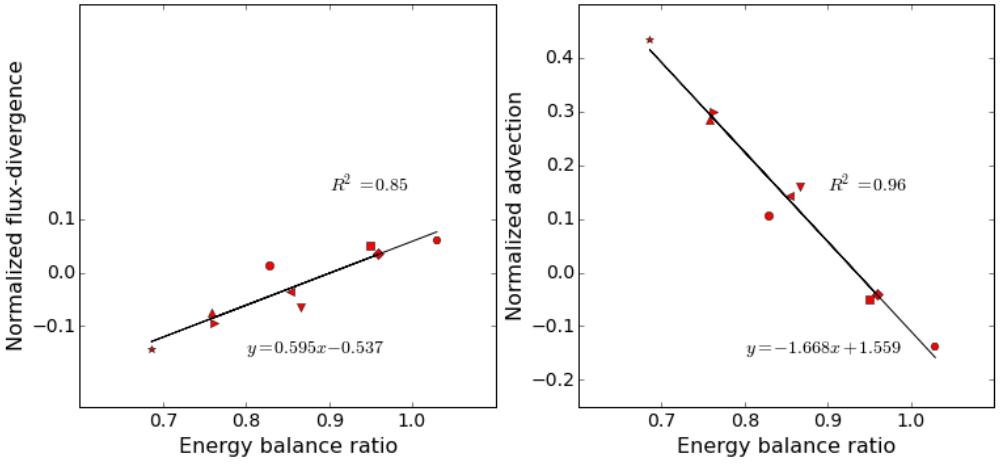


Figure 5. Correlation between flux-divergence and EBR for kilometer scale heterogeneity (left panel); correlation between advection and EBR for kilometer scale heterogeneity (right panel)

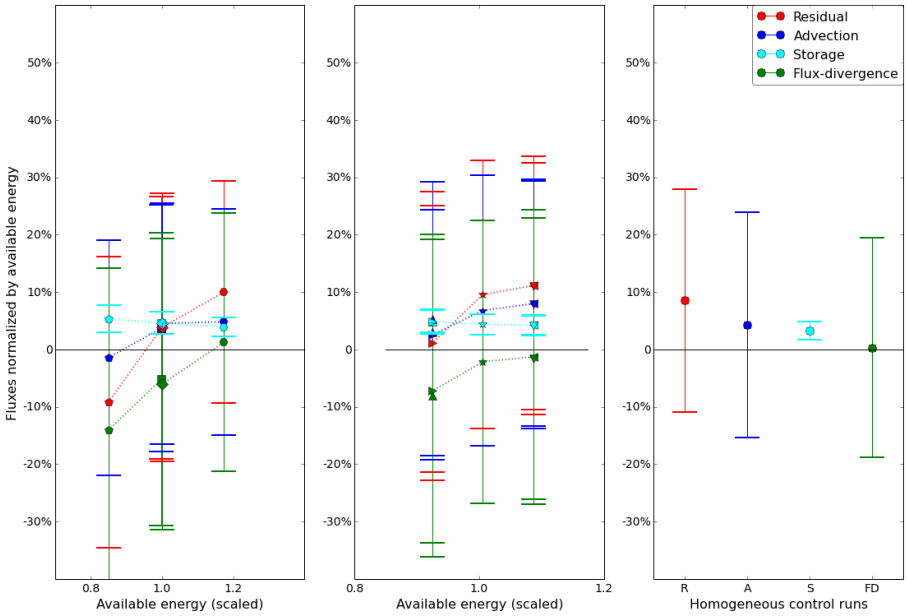


Figure 6. Control volume fluxes as a function of available energy (scaled by the median value) for hectometer scale landscape heterogeneity. The fluxes are normalized by the available energy at their respective location, in our setup this means normalization by the surface flux. Please note we have plotted the non-closure (normalized energy balance residual) instead of the energy balance ratio EBR (normalized turbulent flux). The left panel shows the towers at the centers of the patches, the middle panel the towers at the edges of the patches, and the right panel the results for the homogeneous control runs. For the tower symbols, see Fig. 1. The error bars denote the spread over the different cases of surface heterogeneity within the suite of hectometer scale surface heterogeneity. The abscissa is the available energy at the tower, but scaled by the mean available energy of the nine towers for that case. In this way, we can group the towers by tower type, also for the cases with different surface amplitudes. Thus, the low values represent the towers located at the cooler patches (downdrafts), the high values the towers located at the hotter patches (updrafts). See text for further discussion.

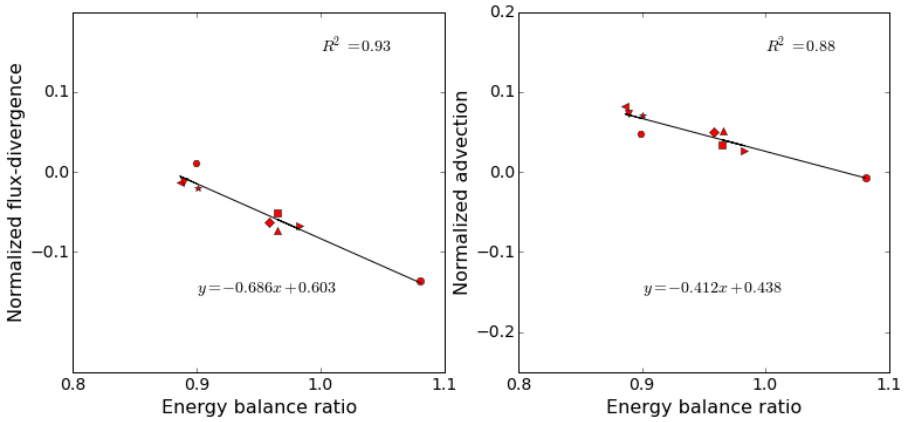


Figure 7. Correlation between flux-divergence and EBR for hectometer scale heterogeneity (left panel); correlation between advection and EBR for hectometer scale heterogeneity (right panel)

Table 1. Parameters of the LES configuration

Quantity	Unit	Value
Number of grid points	(-)	600 x 600 x 240
Spatial resolution (dx, dy, dz)	(m)	10.0, 10.0, 10.0
Domain size	(m^3)	6,000 x 6,000 x 2,400
Temporal resolution	(s)	1.0
Spin-up time	(s)	7,200.0
Data capture	(s)	7,200.0 - 21,600.0
Averaging interval	(s)	3,600.0
Size of the control volume	(m^3)	50.0 x 50.0 x 50.0
Approximate walltime for one simulation	(core-hours)	5500
Roughness length	(m)	0.1
Surface moisture flux	(kg kg^{-1} m s^{-1})	5e^{-6}

Table 2. Parameters of the simulations within the suite focusing on the landscape heterogeneity at kilometer scale

Strong to free convection		$6 \times 6 \times 2 \times 2 = 144$ cases
Average surface heat flux (H_0)	(K m s^{-1})	0.25
Amplitude x (A_x)	(K m s^{-1})	0.0 ; 0.1 ; 0.2 ; 0.3 ; 0.4 ; 0.5
Amplitude y (A_y)	(K m s^{-1})	0.0 ; 0.1 ; 0.2 ; 0.3 ; 0.4 ; 0.5
Length scale x (L_x)	(m)	1,500.0 ; 3,000.0
Length scale y (L_y)	(m)	1,500.0 ; 3,000.0
Surface flux range	(K m s^{-1})	From 0.0625 to 0.5625
u_* (momentum flux)	(m s^{-1})	From 0.071 to 0.69
Boundary-layer height	(km)	From 1.4 to 2.2
Obukhov length	(m)	From -36.1 to -0.04 (average -3.96)

Table 3. Parameters of the simulations within the suite focusing on the landscape heterogeneity at hectometer scale

Strong to free convection		$6 \times 6 \times 2 \times 2 = 144$ cases
Average surface heat flux (H_0)	(K m s^{-1})	0.25
Amplitude x (A_x)	(K m s^{-1})	0.0 ; 0.1 ; 0.2 ; 0.3 ; 0.4 ; 0.5
Amplitude y (A_y)	(K m s^{-1})	0.0 ; 0.1 ; 0.2 ; 0.3 ; 0.4 ; 0.5
Length scale x (L_x)	(m)	200.0 ; 400.0
Length scale y (L_y)	(m)	200.0 ; 400.0
Surface flux range	(K m s^{-1})	From 0.0625 to 0.5625
u_* (momentum flux)	(m s^{-1})	From 0.052 to 0.74
Boundary-layer height	(km)	From 1.5 to 2.2
Obukhov length	(m)	From -53.8 to -0.01 (average -5.75)