Review of
The influence of idealized surface heterogeneity on virtual turbulent flux measurements

by
Frederik De Roo and Matthias Mauder

October 7, 2017

General comments:

In field experiments, it is typically assumed horizontally homogeneous and steady-state conditions, and the vertical turbulent flux of heat is measured as an estimation of the surface heat flux. However, when computing the total energy budget (including radiation and latent heat), the vertical turbulent flux of heat does not close the energy budget, indicating that one or both assumptions are wrong. In this study, it is investigated (i) the effect of surface heat flux heterogeneity on the difference between the vertical turbulent heat flux and the surface heat flux itself (the residual of the energy budget in field experiments), and (ii) the correlation between simulation variables that can be obtained in field experiments and this residual. It was found that in the presence of horizontal heterogeneity in the sensible heat flux in the order of hectometer and kilometer, there is indeed a difference (a residual) between the surface flux and the vertical turbulent flux at 50 m. Estimations located at the edge of patches with different surface heat fluxes have higher residual than the ones located at the center of patches. The effect of storage was smaller than the effects of advection and flux-divergence. A correlation between the residual and the difference in temperature between surface and turbulent flux measurement height was observed.

This work focuses on a very interesting and useful topic, and it uses an appropriate tool to investigate it (LES). However, the simulation itself is not properly discussed and the results obtained are not extrapolated to the real case, therefore contributing very little to the advancement of the knowledge in the field. In addition, the estimation of correlations between the residual and the atmospheric parameters needs to be improved. For this reason, I would like to suggest major revision including (i) a thorough discussion of the simulation results, (ii) a correction of the correlation estimations, and (iii) a proposition of a model that could be used in field experiments (or some other useful information for future field experiments), in order to improve the impact of the manuscript.

Because this is my first time thinking about this topic, I may not be able to give the best suggestions about how to present more useful results, although I will try my best. Nevertheless, I do know that there is a rich amount of information given by the LES, so it should be possible to extract a lot of interesting information from it. It would be interesting to try to obtain results that would help people from field experiments to either understand what is likely causing the closure problem, to better plan for future experiments, or to improve the closure problem.

Specific comments:

1. The LES model: despite being a well-known technique and a largely used model, I think a little more information about the PALM-LES should be given. This is a simulation of free convection, in which no mean streamwise velocity is present. Have this type of simulation been performed with PALM-LES before? If so, a citation and a brief summary of model's performance should be given. Otherwise, some description of the velocity and temperature fields should be given, including an assessment of the level of reality being represented by the model.

2. The simulations: all the information needed to reproduce the simulations exactly should be given. For example, the exact values of initial and boundary conditions of all variables, the strength of the inversion and the subsidence, etc. I'm still confused about how many simulations were run. I'm assuming it was two, one for the kilometer and another for the hectometer case. If so, the information in Tables 2 and 3 are confusing. Does the word "cases" mean "patches"? What are the ranges in ABL height and Obukhov length, are they in time or space? Are these ranges resonable?

3. p. 6, l. 21–22: "the Gauß-Ostrogradski theorem has been used to reformulate a divergence within the control volume as a surface term", please be explicit on what was done.

4. The PCA analysis: in the Methods section, a brief description of the PCA method and how to interpret its results should be given. Right now this is completely left to references, but I think I should be able to understand the technique and the plot overall without having to look in another paper. Also, a more exact description of what was done should be given, including how many and which variables were used, which equation or software, etc.

5. Simulation results: the simulation results should be presented and discussed before presenting the statistics. For example, how do the spatial fields of temperature and heat flux look like, and where do the towers rest in this field? Some time series at the tower place, to see what the towers are measuring and what are the scales of motion in it. How do they compare with the scales of heterogeneity? How much of the fluxes are resolved compared to the sub-grid scale? How realistic are these simulations? A thorough discussion of the simulation is definitely needed, as it would help to discuss the physics of the results presented later.

6. Discussion: there are three distinct physical phenomenon that could be impacting the residual: transport of the mean field, transport of the fluctuating field, and storage. What are the physics involved in

each process, how is the simulation capturing them, and how do they look like in the simulation? Is it realistic to look into the advection effect, for example, in a simulation that has no mean streamwise advection? What happened to the vertical and horizontal dispersive fluxes?

7. p. 7, l. 19: what is the *available energy*? It is the reference value in the results section, but there is no definition of it. It is only explained in the Fig. 2 caption, but it should be clear in the text too.

8. p. 7, l. 19: please make it clear if the *advection* term includes both horizontal and vertical advection.

9. p. 7, l. 29–31: "We notice that most towers show the typical underestimation of the energy balance, except for the tower located at the warmest spot where there is an updraft." How do you know it is an updraft? Is it always an updraft? On average? It is mentioned in the abstract that updraft and downdraft positions were chosen for the towers, but this is not showed or discussed in the paper. Do constant regions of updraft/downdraft exist in the simulation or in reality?

10. p. 7, l. 30–31: please explain better physically the causes of a negative/positive residual, and why there is a negative residual where there is an updraft.

11. p. 8, l. 13–14: "In the left panel of Fig. 4 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." What does "normalized" mean? If it is the "normalized by the available energy at their respective location" mentioned in the caption of Fig. 2, it should be mentioned in the text too.

12. p. 8, l. 22–23: "flux-divergence and advection separately correlate well with energy balance ratio and consequently also with each other". What does that mean physically? Does it makes sense that these two processes are correlated in the simulation? What is the implication of this for the imbalance observed in the real case?

13. p. 8, l. 24–26: "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." Why not compare the residual observed in the simulation as a function of height, including the lower points? If I understood correctly, the residual term is estimated at a given height, so it does not need the control volume approach. If so, you can see if the results at 50 m are similar at 10 m, and if the conclusions could be extrapolated. Without this extrapolation, I see little usefulness in the conclusions obtained here.

14. p. 8, l. 32–33: "the towers in the center of the patches even behave in the opposite manner when the kilometer and hectometer scales are compared". This needs to be better investigated and discussed physically. What can be causing this?

15. Discussion: I think it is important to discuss the differences when there is a small residual due to low values of all other fluxes, or due to their canceling effect. What is likely to be happening in field experiments?

16. p. 9, l. 4–7: "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." I agree this is likely the cause of difference, but how does this make the relation between flux-divergence and EBR opposite, or the relation between the residual and the surface flux opposite?

17. PCA results: from the very little I know about PCA analysis, I think the results obtained here are not useful. I believe it is useful to look at correlation biplots when most of the variance are explained by the first two PC's (something around 90%), not the 60% found here. As explained by Greenacre (2010) (your own reference on the topic), PCA correlation biplots are useful when there is a clear separation in the scree plot between the first two PC's and the rest (called *elbow*), which is definitely not the case here. In your data, the third PC is almost as important as the second, and it is not taken into account. In addition, the EBR (the variable you want to explain) is the one with less representation by the first two PC's among the variables in the correlation biplot, being much less than 50% representation in the hectometer case. I don't know which analysis should be done instead, but I think PCA is not the one.

18. Conclusions: ignoring the results from the PCA, the only conclusion from this work is that this type of heterogeneity in the surface flux generates a difference between the turbulent flux measured at 50 m and the surface flux itself. This is an interesting information, but given the potential of these simulations, I believe that much more can be obtained. I would like to see some results and conclusions that could help improve or understand the closure problem in field experiments. The idea of finding parameters measurable in the field that correlate well with the residual is a good approach. But another technique should be used to find the right parameters. After finding them, I suggest that a model for the residual should be developed and tested against the LES results. Naturally, in reality things are not equal to the LES, but this would gives us a place to start. When combined with a thorough discussion of the simulation itself, it would be easier to extrapolate the results and conclusions to the real case. Without it, I don't see any significant contribution to the field.

Technical corrections:

- Abstract: there is too much introduction information in this abstract. It is also clear from it that there is no significant contribution from this work. After improving the results and conclusion, the abstract should focus more on them.

- p. 4, l. 17: “we have added a very small moisture flux” why? why not make it zero?
- Figure 1: missing unit of color plot (surface heat flux)
- Equation (3): a sketch of the fluxes and a figure showing where in the 50 by 50 m box each term is being calculated would be usedful.
- p. 7, l. 30: include text: “most towers show the typical underestimation of the energy balance **(i.e., positive energy balance residual)**”
- p. 8, l. 12: what is “resp.”?
- p. 8, l. 12, 13: it should be Fig. 3 instead of Fig. 4
- p. 8, l. 20: opposite slope? They look the same to me...
- p. 8, l. 31: since you are comparing Fig. 2 with Fig. 4, why not make them one single plot? It makes it easier to compare.
- Fig. 3 and Fig. 5 are equal, one of them is wrong (I guess it is Fig. 3, based on the text)
- p. 9, l. 2: Fig. 4 should be Fig. 3

References

Greenacre, M. (2010). *Biplots in Practice*. fundacion BVBA.