Response to review by Dr. Piskozub:

The review is in general very positive. We greatly appreciate Dr. Piskozub's thoughtful comments that helped improve the manuscript. We trust that all suggestions have been addressed accordingly in a revised manuscript. In the following, we give a point-by-point reply to the comments of Dr. Piskozub:

1) I was surprised by the procedure of considering all height level permutations in order to determine the parameters of the logarithmic profile (first paragraph of Section 4.4). I do not see why it would be better than simply finding a best-fit for all the levels as we did. It seems to me we both use the same amount of information making the outcome equivalent. Am I wrong? I believe a comment on the reason of using the procedure would improve the paper allowing future users of the gradient method to do an educated choice between the variants.

Indeed, both procedures lead to identical results with respect to the estimated flux values. This is true only if we force the fitting line to go through zero, which means the following: For $\ln(z2/z1) = 0$, i.e. $z2 = z1$ (a measurement at a certain height $z1 = z2$), $T2 - T1 = 0$, i.e. $T2 = T1$ (the measurement result at that height is identical). In fact, if we were to use all possible combinations of data pairs, i.e. each height difference is considered twice with opposite signs (e.g. $T3 - T1$ and $T1 - T3$), the regression line would always go through zero for reasons of symmetry. However, we chose to consider the difference of the same two data points only once (neglecting the difference of the same two data points with opposite sign) and force the fitting line through zero. The main benefit of this procedure becomes obvious in Fig. 7a, where two diverging trends indicate that the data point at the lowest height level may be inconsistent with the logarithmic model. As mentioned in the manuscript, with this representation we gain insight into the variability of the data with respect to the logarithmic behavior. In the revised manuscript, we add a paragraph at the end of section 4.3 to clarify that our results are in agreement with the more standard approach (cf. reply to reviewer 2).

2) The authors of the reviewed paper had the advantage of using eddy correlation at the same time as the gradient method. This allowed not only for the comparison of calculated fluxes but also made it possible to estimate independently friction velocity. At the time we made the measurements described in Petelski & Piskozub 2006, we did not have yet the possibility. Still we believed already then that simultaneous measurements with the gradient and eddy correlation methods could help establish whether the vor Karman constant is applicable also to particle flux (it's value was empirically established for heat fluxes and therefore its application for particle fluxes should be also checked experimentally). This was discussed in the Andreas comment to our paper and in our reply to it. We had seen some hints that the counterpart to van Karman constant for particle fluxes (let me call it Petelski constant) could be closer to 1.0 than 0.4.
Would the authors care to comment whether their data can help constrain its value? I do not insist on including such a discussion in the manuscript (although I would not mind that). Commenting in the reply to this review would be enough if the authors do not feel their data could help constrain the Petelski constant in any meaningful way.

The discussion about the applicability of the von Karman constant in aerosol flux-profile relationships (Petelski and Piskozub, 2006; Andreas, 2007; Petelski and Piskozub, 2007) is interesting, however, we do not think that the measurements presented here can contribute to this discussion and help constrain its value. Clearly, a comparison of profile-derived and eddy covariance fluxes is a valuable effort, but since our data coverage is very limited, unfortunately, we cannot make a strong statement about this discussion based on our data.

3) The third thing I would like to comment is using the statistical tests in the null hypothesis. First of all the phrase "the probabilities of acceptance of the null hypothesis" (line 10 in Section 4.1) is wrong. We never accept the null hypothesis. In fact we test how improbable it is to obtain our research hypothesis by accident, assuming the null hypothesis is true. If it's improbable enough (below the rather arbitrary threshold of 5% probability) we say we "rejected" the null hypothesis. However if the probability over the threshold we still do not accept the null hypothesis (as we never tested it in any way). We just say our research hypothesis "is not statistically significant".

We fully agree with the reviewer and changed the phrasing in section 4.1 as well as in Tables 1 and 2 accordingly.

However my comment goes further. I suggest not using the null hypothesis rejection and significance level analysis at all. The literature proposing this has long tradition. Cohen (1994) already said that after "4 decades of severe criticism, the ritual of null hypothesis significance testing - mechanical dichotomous decisions around a sacred .05 criterion - still persists." This methodology is criticized not only for the arbitrary threshold (the list of complains is too long to repeat here). Hunter (1997) in a paper which title itself tell it all argues that "The significance test as currently used is a disaster. Whereas most researchers falsely believe that the significance test has an error rate of 5%, empirical studies show the average error rate across psychology is 60% - 12 times higher than researchers think it to be". That is one of the reasons why Armstrong (2007) stated "I was unable to find empirical evidence to support the use of significance tests under any conditions" while Hubbard and Lindsay (2008) concluded "it is bad enough for researchers to misuse a measure that is useful: But it strains credulity to do so when that measure is seriously flawed in itself. And this paper has demonstrated - from a multitude of perspectives - that the p value is just that". Gigerenzer et al. (2004) actually compared using this methodology to rituals: "Elements of social rituals include (a) the repetition of the same action, (b) a focus on special numbers or colors, (c) fears about serious sanctions for rule violations, and (d) wishful thinking and delusions that virtually eliminate critical thinking [..]. The null ritual has each of these four characteristics: a repetitive sequence, a fixation on the 5% level, fear of sanctions by editors or advisers, and wishful thinking about the outcome (the p-value) combined with a lack of courage to ask questions". To make it worse tests show that even 80% of scholars teaching statistics do not understand what significance testing actually means (Haller & Kraus 2002). Most of the above examples of rejecting the "null ritual" come from social sciences and psychology. However at least two papers voiced the same concerns in the field of atmospheric (Nicholls 2000) and climate science (Ambaum 2010). However one may say: "OK, but what is the alternative?" There is more than one. Ambaum (2010) suggests Bayesian analysis which may be the future but the scientific world may not yet be ready for it (at least I'm not). The other proposition (one of the advices of Nicholls 2000) is using confidence intervals. This also is not a new proposal, Gardner and Altman proposed it in 1986 and later wrote a whole *book promoting this approach (Altman et al. 2005). In the case of the reviewed manuscript, the confidence interval approach would call for checking how many standard deviations ("sigmas") the values are from each other. If the distributions are normal, two sigmas correspondent to a 95% confidence interval, which actually implies what people expect from a 5% significance. I believe all the data presented in the paper would pass the 2 sigmas test. You may be surprised but I do not insist on implementing this suggestion. It is a matter of philosophy and I do not believe in coercion with respect to this matter, rather evangelizing (which I exactly what I did above).*

We carefully considered the reviewer's suggestion and decided to keep the null hypothesis significance testing as presented in the manuscript. In fact, we use this test as an independent analysis in addition to the 95 % confidence interval test which is shown in Figs. 3c and 5c. Both tests agree and imply significant differences of adjacent means except for data points 4 and 5 as shown in Tab. 2 and Fig. 5c. Again, it should be emphasized that the significance of the differences is not a prerequisite for the validity of a profile. It is simply an indication that the differences between two heights could be resolved by our measurements.

There are some purely technical matters I would like also to mention:
data (abstract, line 5) is usually treated as plural of "datum" so I would prefer "were" to "was"

We corrected the manuscript accordingly. (abstract, l.5, and in addition on page 12, l.5, l.8, page 13, l.20, l.21, l.25, l.27, page 23, l.4, l.9)

height sensor "pointing normally" toward the ground (Section 2.1 line 34). I would prefer "pointing vertically"

We revised the manuscript accordingly.

Thank you very much for your time and effort!