ANSWERS TO

Interactive comment on "Study of the overturning length scales at the Spanish planetary boundary layer" by P. López and J. L. Cano

Anonymous Referee #1
Received and published: 5 January 2016

We would like to thank the referee for him/her useful comments towards the improvement of our manuscript.

The paper is addressing a very interesting topic that can have a deep consequences in modelling the ABL. The problem is well exposed and carefully documented by chosen references. The obtained results seem to bring a little more complexity to the problem by obtaining power-law, rather than linear relations between the considered length scales. Also the day-time versus nocturnal period separated statistics seems to be a reasonable approach.

Thank you for your opinion.

When it comes to results, the only thing that puzzles me are the breaks in time series of measured data. These breaks were explained, but I was wondering how these gaps in data could have affected the results and conclusions. It means, would we get somehow different results with complete data, or inversely, would the other authors get different results if they will also have such gaps in data?

There is not data between aproximately 12:00-15:00 hours (as figures 1 and 2 show) because the data were registered mainly by a tethered balloon which needs to be watched and monitored to prevent its breakage (for example, the balloon must be collected if there is a storm or wind suddenly appears).

We consider that these gaps in data would not affect our conclusions, that is, they would be the same if we were able to measure 24 hours a day. There are physical explanations which would support our hypothesis. We consider that at the 12:00-15:00 h time interval, the overturns could be generated by one or several convective burst with different scales (due to the effects of solar heating and the meteorological conditions). These convective situation could make several mixing events which could superimpose and could make greater overturns. Based on two-dimensional visualizations of temperature data, Keller and Van Atta conjectured that overturns could be generated by a localized vertical advection of well-mixed lumps of fluid past their equilibrium position and subsequnt displacement of stable density fronts (Keller, K. H. and Van Atta, C.: An experimental investigation into the vertical temperature structure of homogeneous stratified shear turbulence, J. Fluid Mech. 425, 1-29, 2000). Moreover, it has been studied that the close proximity of adjacent overturns allows them to merge and to generate larger-scale overturns (Diamessis, P. J and Nomura, K. K.: The structure and dynamics of overturns in stably stratified homogeneous turbulence, J. Fluid Mech, 499, 197-229, 2004).

As a consequence, it is possible to expect that the corresponding maximum Thorpe displacement $(d_T)_{max}$ and the Thorpe scale, L_T, would be greater. As a consequence, it would be possible to get features having an ‘eddylike’ shape similar, some a random mix of different-scale fluctuations without sharp boundaries as in the following figures.

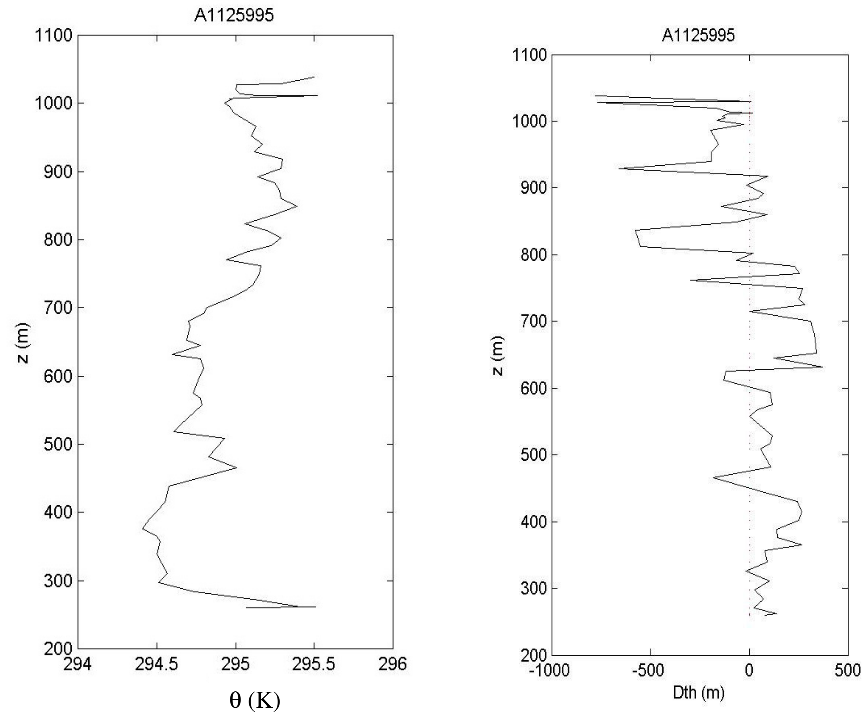

Figure 1. The real potential temperature profile (left curve) and the corresponding Thorpe displacements profile (right curve) corresponding to 11:00 GMT.

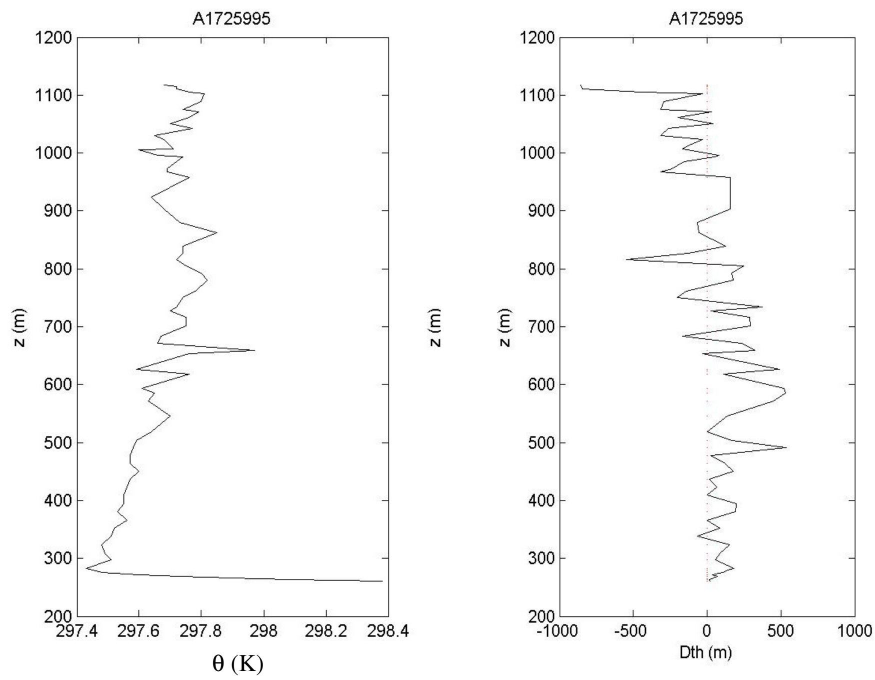

Figure 2. The real potential temperature profile (left curve) and the corresponding Thorpe displacements profile (right curve) corresponding to 17:00 GMT.

The data correspond to the same field campaign made at 25TH September of 1995. We represent the potential temperature profile and the Thorpe displacements profile. Figure 1 shows the behavour at 11:00 GMT, when the convective effects would start. Figure 2 corresponds to 17:00 GMT, when the convective effects would be more developed. We clearly observe the random mix of different-scale fluctuations without sharp boundaries with a vertical extent of the order of 500 m at 11:00 GMT, and greater at 17:00 GMT (of the order of 1000 m). These mentioned fluctuations act as external intermittency that refers instead to the intermittency of the occurrence and variability among different turbulent events (which could generate overturns as sporadic convective processes o baroclinic instabilities).

Finally, we present a new figure (figure 3) which represents the potential temperature profile and the Thorpe displacements profile at 07:00 GMT (without convective effects). We observe a clear *z*-shape overturn that has sharp boundaries with displacement fluctuations of a size comparable to the size of the disturbance itself in the interior, that is, with intense mixing inside (Dillon, T. M., 1982: Vertical Overturns: A Comparison of Thorpe and Ozmidov Length Scales, J. Geophysical Research, 87, C12, 9601-9613). This typical large overturning eddies have sharp upper and lower boundaries with intense mixing inside. This kind of overturn could be probably generated by random breaking of internal waves or Kelvin-Helmholtz instabilities. We also observe that this overturn is not so greater (about 40 m) as the ones of fiures 1 and 2.

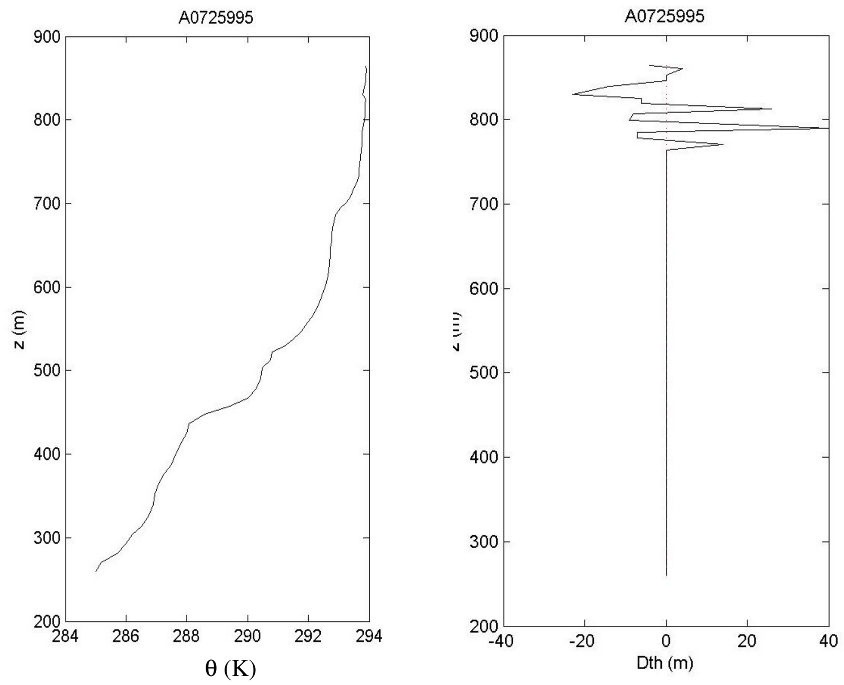

Figure 3. The real potential temperature profile (left curve) and the Thorpe displacements profile (right curve) corresponding to 07:00 GMT.

Finally, there are other reasons. From Figures 1 and 2 (paper figures), it follows that 12:00-15:00 hours missing data should correspond to the greater values of the maximum Thorpe displacements and the Thorpe scales. That is, these missing data should not have small values of $(d_T)_{max}$ and L_T under convective conditions and the typical meteorological situation of this area. Therefore, these missing (no measured) values would be shown in the right part of the graphics (only in figures 3 and 4). Simultaneously, if we were able to measure 24 hours a day, the sample size would be greater and, therefore, the reliability of our results will improve. The reason is statiscal because one of the ways to get an improvement of the hypothesis test power is to increase the sample size. As a consequence, our conclusions would be reinforced.

From technical point of view, I don't like the figures at the end of the paper (which makes it harder to read), but this is probably just the manuscript style, not a choice of authors.

Yes, it is true.

There are few misspelled words in the text, which is easily fixable in the final version of the paper. A little annoying for me was also the use of expressions "P value is ...", "R-squared coefficient is...", "F test for ...", which is probably some common notation use by someone in certain branches of statistics, but for a technical (physical) paper, these terms (and notation) should be explained or rather properly referenced.

We agree (we have used the typical statistical notation) and we are going to describe these terms properly in the revised version of the paper.

The *p*-value helps us to determine the significance of the results when we perform a hypothesis test which is used to test the validity of a claim that is made about a population. This claim that's on trial, in essence, is called the null hypothesis. The alternative hypothesis is the one we would believe if the null hypothesis is concluded to be untrue. The *p*-value is defined as the probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming that the null hypothesis is true. We use a *p*-value (always between 0 and 1) to weigh the strength of the evidence. A small *p*-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis. A small *p*-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis.

The *R* coefficient or linear correlation coefficient is a normalized measurement of how two variables are linearly related. It represents the correlation coefficient of two variables. If the correlation coefficient is close to 1, it would indicate that the variables are positively linearly related. The R-squared coefficient is called the determination coefficient which represents the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph. The coefficient of determination is a measure of how well the regression line represents the data.

As it was mentioned at the paper, it is necessary to do a multiple regression analysis. The comparison of regression lines procedure is designed to compare the regression lines relating *y* and *x* at two or more levels of a categorical factor. Tests are performed to determine whether there are significant differences between the intercepts and the slopes at the different levels of that factor.

Comparing two regression lines is the simplest model of covariance analysis. It uses the independent variable *x* as covariate and dependent variable *y* as outcome in a 2 group analysis of variance (decomposition of the variability of the dependent variable *y* into a model sum of squares and a residual or error sum of squares). Of particular interest is the F-test on the model line which tests the statistical significance of the fitted model. A small p-value (less than 0.05) indicates that a significant relationship of the form specified exists between *y* and *x*. The F-test is any statistical test in which the statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.