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 #2
Received and published: 8 January 2016

The authors study an interesting problem in modelling the atmospheric boundary layer by analysing the relationship between the maximum Thorpe displacement, and the Thorpe scale, which is the statistical mean of displacements. The results are based on the set of detailed observations. The authors argue that the relationship under consideration is not linear, as found previously for the atmospheric boundary layer, but powerlike, and find the corresponding power laws for the complete set of data, and separately for day and night observations.

Thank you for your opinion and the opportunity to revise our paper. I have commented below on each of the points raised by the referee.

First, the authors write, in section 4.1, that they have found two qualitatively different behaviours of Thorpe displacements. It is rather difficult to visualize these cases from the explanation. Perhaps it would be better to illustrate these behaviours with a figure.

The following text mentions that for our ABL studies, Thorpe displacements could be qualitative classified in two groups: the first group represents discrete overturns where the Thorpe displacements are always zero except in a region with an isolated *Z* patterns (usually under neutral and stable stratification conditions); the second group represents a random mix of different scale fluctuations without sharp boundaries, some having an eddylike shape similar to the larger overturns, where the Thorpe displacements rarely are zero for the whole profile (the opposite behaviour that usually happens at convective conditions). The following figures show the two Thorpe displacement groups with different behaviour. The left figure is an example of the first group, that is, an isolated overturn and the right figure is an example of the second group. From these figures it is clear that there is a different behaviour. Both graphs correspond to a campaign made 25th of September of 1995. The left figure is at 07:00 GMT (stable conditions) and the right graph is at 17:00 GMT (convective conditions).
We will probably add this figure to the revised version of the paper although this kind of figures are shown at the references cited at the paper: *López, P., Cano, J. L., Cano, D., and Tijera, M.: Thorpe method applied to planetary boundary layer data, Il Nuovo Cimento, 31C, 881–892, 2008* and *López, P., Redondo, J. M., and Cano, J. L.: Thorpe scale at the planetary boundary layer:comparison of Almaraz95 and Sables98 experiments, Complex Environmental Turbulence and Bio-Fluids Flows, Institute of Thermomechanics AS CR, Prague, 2015 (in press).*

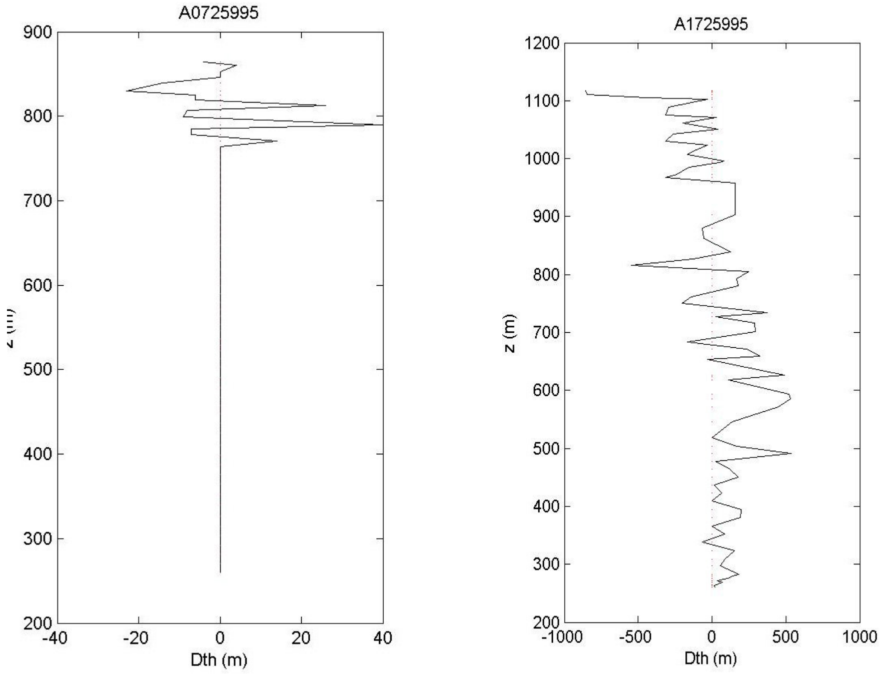


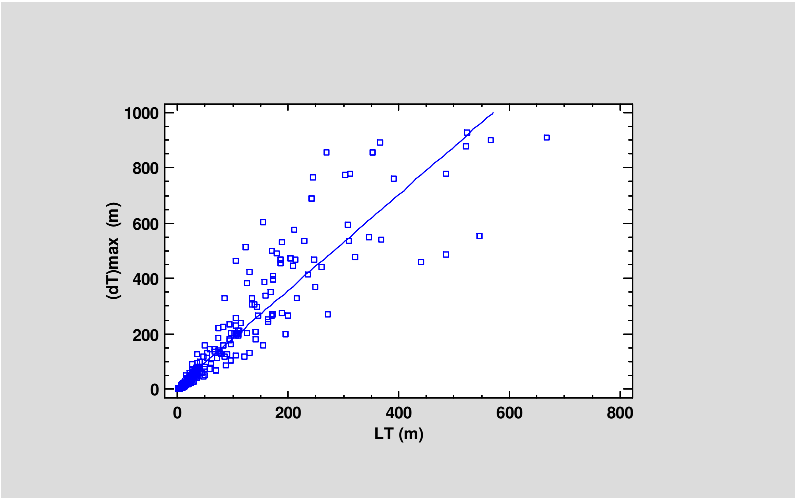
Figure 1. Left curve**,** Thorpe displacements profile with an isolated patch corresponding to 07:00 GMT. Right curve, Thorpe displacements profile with a random mix of fluctuations corresponding to 17:00 GMT.

Second, it does appear from figures 3-5 that a power law fits the data better than a linear one. However, it would be better to explain it clearly in statistical terms. [....] I can trust that they have chosen the best exponent for the powerlike fit, but how much better it is, in statistical terms, than a linear fit with a certain coefficient, similar to what has been found by other authors?

To explain in statistical terms that the the power law fits the data better than a linear one, we have realized a new statistical study. We have made a simple regression procedure to construct a statistical model describing the dependence of $|(d_T)_{max}|$ on L_T considering the different situations, i. e., the daytime data (figure 4) and the nighttime data sets (figure 5). The new study for the whole data (figure 3) is described in the following comment.

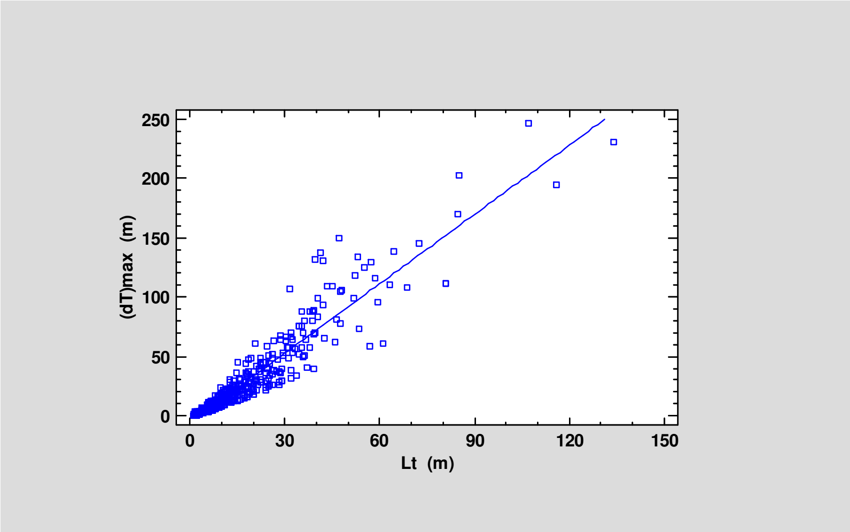
First, we analyze the behaviour of the daytime data set (figure 4). The linear model was fit using least squares and tests were run to determine the statistical significance of this model. Our results show that the estimated linear model is $|(d_T)_{max}|=10.794+1.732\ L_T$. The analysis of variance, which tests the statistical significance of the fitted model, indicates that a significant relationship of the form specified exists between $|(d_T)_{max}|$ and L_T (because the p-value is less than 0.05). In the daytime sample data, the linear model is significant but the *R-squared* –or determination coefficient- which represents the percentage of the variability in $|(d_T)_{max}|$ which has been explained by the fitted regression model is 84.3%. The regression has accounted for about 84% of the variability in the maximum Thorpe displacements measurements. The remaining 16% is attributable to deviations around the line, which may be due to other factors, for example, to a failure of the linear model to fit the data adequately.

The same statistical analysis was made using a power law fit and we got that the *R*-squared coefficient is 96.76%, that is the power law fit accounts for about 98% of the variability in the maximum Thorpe displacement $|(d_T)_{max}|$ as a function of the Thorpe scale, L_T. We conclude that both models, the power law fit and the linear one, are statistically significant but the power law fit has a better determination coefficient, that, is, it accounts better for the variability in the maximum Thorpe displacements measurements. Therefore, we consider that the power law fit is the best model for the daytime data set. We also present the graph of the daytime data and the linear fitted model. The plot includes the line of best fit. This plot clearly shows that the daytime data does not follow a linear model.



Second, we analyze the behaviour of the nighttime data set (figure 5). The statistical analysis was repeated: the linear model was fit using least squares and tests were run to determine the statistical significance of this model. Our results show that the estimated linear model is $|(d_T)_{max}|=-5.571+1.947\, L_T$. Again, the p-value of the analysis of variance is less than 0.05 and indicates that a significant relationship of the form specified exists between $|(d_T)_{max}|$ and L_T for the nighttime data set. The *R-squared* coefficient is 90.11%. The regression has accounted for about 90% of the variability in the maximum Thorpe displacements measurements. The remaining 10% is attributable to other factors (may be the linear model does not fit the data adequately). The same statistical analysis was made using a power law fit and we got that the *R*-squared coefficient is 95.89%, that is the power law fit accounts for about 96% of the variability in the maximum Thorpe displacement $|(d_T)_{max}|$ as a function of the Thorpe scale, L_T.

In the same way as in the previous case, the power law fit accounts better for the variability in the maximum Thorpe displacements data and we consider it is also the best model for the nighttime data set. We also present the graph of the nighttime data and the line of best fit. This plot clearly shows that the nighttime data does not follow a linear model.



For example, most (although not all) data in figure 3 appear to fit rather well to a linear dependence.

The figure 3 represents the data in logarithmic scale, and it is clear that they fit well to a linear relation. The following figure shows the same data of figure 3 (of the paper), but on a linear scale. We observe that the data do not fit so well to a linear relation (mainly due to the behaviour of the greatest values of the maximum Thorpe displacement and the Thorpe scale).

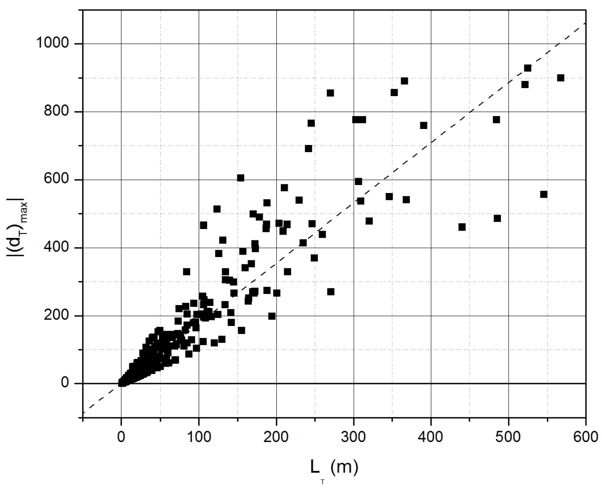
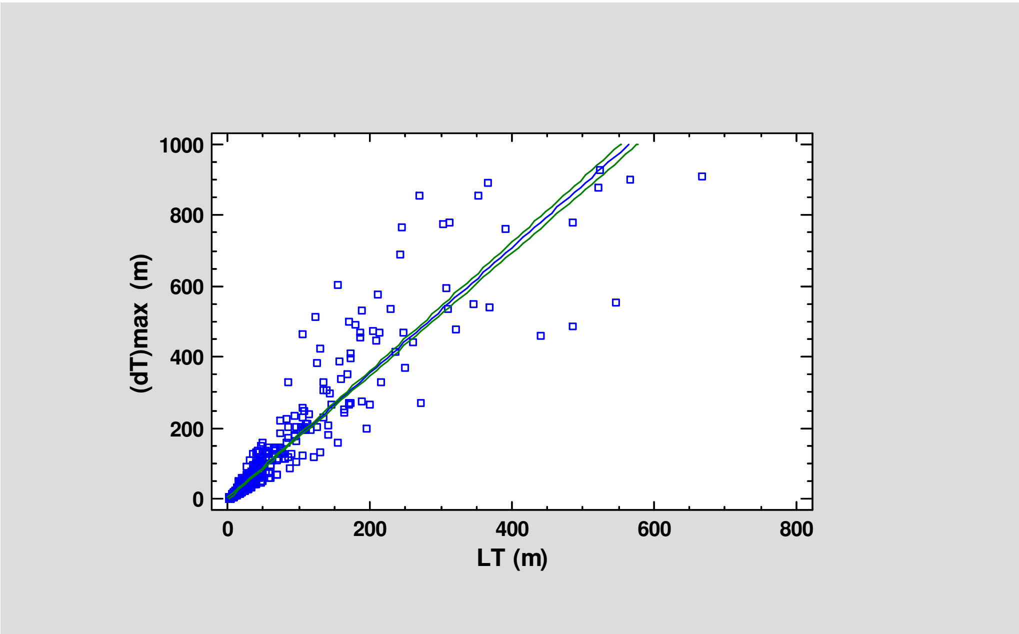


Figure 2. Absolut value of the maximum Thorpe displacement vs. Thorpe scale for all field campaigns. The representation is in linear scale.

As mentioned in the previous comment, we have realized a new statistical study to know if the linear fit is better or not than the power law fit. We made a simple regression procedure to construct a statistical model describing the dependence of $|(d_T)_{max}|$ on L_T. The linear model was fit using least squares and tests were run to determine the statistical significance of the model. Our results show that the estimated linear model is $|(d_T)_{max}|=0.218+1.771\ L_T$. The analysis of variance indicates that a significant relationship of the form specified exists between $|(d_T)_{max}|$ and L_T (because the p-value is less than 0.05). The percentage of the variability in $|(d_T)_{max}|$ which has been explained by the fitted regression model is 87.9% which is the value of the determination coefficient. For the campaign data, the regression has accounted for about 88% of the variability in the maximum Thorpe displacements measurements. The remaining 12% is attributable to deviations around the line, which may be due to other factors, for example, to a failure of the linear model to fit the data adequately.

The same statistical analysis was made using a power law fit and we got that the *R*-squared coefficient is 96.95%, that is the power law fit accounts for about 97% of the variability in the maximum Thorpe displacement $|(d_T)_{max}|$ as a function of the Thorpe scale, L_T.

We also present the graph of the data (as before) and the fitted model plotted with confidence limits. The plot includes the line of best fit and the confidence intervals for the mean response which describe how well the location of the line has been estimated given the available data sample.



As a conclusion, we conclude that both models, the linear fit and the power law one, are statistically significant but the power law fit has a better determination coefficient, that, is, it accounts better for the variability in the maximum Thorpe displacements measurements. Therefore, we consider that the linear regression is not the best model.

The authors need to make sure that all statistical concepts they use (P value, F test, etc) are properly defined.

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.

Also, the use of the term "linear" needs to be more consistent in the paper. [….] I guess the authors use the term "linearity" interchangeably in algebraic and statistical sense, which is rather confusing.

We agree and we have revised the text to clarify the statistical sense of the "linear" term when itis mentioned.

The authors write "We observe that the linear relationship $|(d_T)_{max}| = L_T$ proposed by other authors... "; but the other authors have proposed, in particular, a linear relationship with a ratio $|(d_T)_{max}=L_T$ which is different from 1.

Yes, it is true. It is an unfortunate phrase that has no relation to the context of the paragraph and creates confusion. It is also a bad explanation because it seems that the relation $|(d_T)_{max}|=L_T$ is that others authors have deduced and this is not true. We used $|(d_T)_{max}=L_T$ as the perfect relationship, a pattern or reference but it is not true nor necessary. Therefore, we have decided to remove this phrase and all the comments related to the relationship $|(d_T)_{max}=L_T$ in the text, in the figures and figure captions.

It is not clear why the light grey line in figure 5 represents the linear fit, as stated in the caption, while it is clearly a powerlike function, in logarithmic coordinates.

We have revised the figure and we have redone the calculations. Furthermore, the new figure 5 is more understandable because we use other shades of gray and we have eliminated the relationship $|(d_T)_{max}=L_T$, which is not essential.