General

We would like to thank both reviewers for their very positive reviews and we think that we have been able to clarify all issues brought forward. Thanks to these reviews the paper has definitely become more precise in its claims and results.

Reviewer #1 (Clemens Drüe)

In our first reply, we have explained that we agree with Reviewer #1 but that our expectations were lower, hence our enthusiasm for the goodness of the results. We now explain this in the introduction:
“The original objective of the experiment was to test if air temperature differences in urban landscapes can be measured with reasonable accuracies (1°C).”

We also toned down the language in the results:
“The experiment started out with the question if corrected DTS air temperature measurements would have a reasonable accuracy to measure temperatures across urban landscapes. The results show that this is indeed the case and they were actually better than we expected at the onset (1°C). With hindsight, we would have installed a better reference temperature measurement device because we can not say if the reference was truly better than the DTS derived temperatures. Although we can not conclude this from the results, it may be possible to further reduce systematic errors, in which case this method would become valuable for vertical atmospheric soundings with DTS from balloons, quadcopters, or towers.”

We hope we also explained better why the wind measurements are not essential and only served to check for periods without wind.

Finally, we did add the error estimate for individual measurement points as suggested.

We trust that this is sufficient and that the only alternative would be to repeat the experiment with a better reference station, which would not be possible within the reply period and would also be practically not feasible. We assume the Reviewer can agree with that, as it is also supported by Reviewer #3.

Detailed comments

6288,7: "[the results...] were very good." is a too general statement. It would, however, be possible to state that they were very good, compared to something (the expected error of ... K, to a previous result of <citation>, the ... sensor, or similar).

Agreed. As explained under the main comments, it was mainly ‘very good’ with respect to our original expectations. We now state this and refer to WMO 2010 for the general consensus on the need for shielded air temperature measurements.

6288,10: "close" is, likewise, an unjustified classification that might be totally different from a different point of view. Instead, it should be stated that for example, the difference was smaller than the standard deviation of the averaged values, the accuracy of the reference device , or something similar.

Agreed. We now say: “were close to air temperature measured with a nearby shielded thermometer (RMSE of 0.61 °C).”

6288,10/11: The difference between "horizontal cables" and "vertical profiling" is not clear. To my knowledge, a recent study (Thomas et al ?) used a harp-like stack of horizontal fiber segments for vertical profiling.

Yes, that is indeed a very nice study, which also focused on night time / early dawn observations. The original thought was to use the cable along towers or under balloons/quadcopters. We generalized the statement and pointed to these specific applications:

“The temperatures were measured along horizontal cables but the same method can be applied to any atmospheric DTS measurements, especially for profile measurements along towers or with ballons and quadcopters”

6288,20: If 0.01K is the best available accuracy, what would be a typical value?

We added a general typical field value of 0.08K with reference.

6289,19: A Stevenson screen is just one type of shielding. It should be noted that artificial ventilation is indispensable for high-quality measurements (for example see WMO CIMO guide).

This remark, as well as a reference to the CIMO guide have now been included.

6289,19: Who is Gaylon Campbell?

Dr. Gaylon S. Campbell is a research scientist and engineer at Decagon, a company that develops and builds scientific measurement devices, especially for agro-meteorological applications. Previously, he has been on faculty at Washington State University for nearly 30 years. He was the person who first brought to our attention the simple scaling law that underlies this article. Not sure how to include him otherwise but it would be unethical not to give him primacy of this idea, originally developed for spherical thermometers.

6290,5: the equation should be checked against the formatting guidelines of AMT. Crosses are usually reserved for cross products and sometimes the (decimal) exponent and should be removed here.

Frankly, we are not exactly masters of LaTex but the crosses result from the \times command. Cross and dot products give the same results for scalars but to avoid confusion, we now use the \cdot command.

6290,17: 51 9’45.44"N, 42 2’39.56" is in Russia between Moscow an Saratov. It appears you mean 51 59’45.4"N 4 22’39.6"E

Thank you!

6290,12: There is no clear link or deviation from (1) to (2). The usual formula for the radiation error is T_indicated = T_air + Q / alpha_L, where Q is the net radiation at the temperature sensor surface and alpha_L the ventilation coefficient. The latter depends on the form of the sensor and ventilation speed. Could you derive (2) from such a more general expression?

The order in which the equations were presented was indeed confusing. (2) was not derived from (1) but follows directly from the underlying scaling law. We have changed the order, which makes things a bit clearer, and made it explicit that the equation follows from extrapolation to a zero diameter. In a way, it could be derived from the radiation error formula by including a 1/sqrt(diameter) term in alpha_L. By doing that for two diameters with Q’s effect on temperature

and other terms in alpha_L equal, the result is the same. We could add that but it would probably add confusion instead of reducing it.

6290,25: What does "fused together" mean?

Technically, “spliced together” is better, so we use that word now, but it involves the alignment of the cores within the actual glass fiber and melting them together with a fiber optics splicer. It is a typical part of any DTS set-up. Not sure if it makes sense to provide more detail in the text.

6291,1:"measured" -> "operated" ?

Is indeed better, and changed accordingly.

6291,1st para: the water temperature was measured how? What was the accuracy of this measurement? Was the water temperature constant? What was the stdv of the water temperature?

We added: “Bath temperatures were measured with the two PT100 thermometers that came with the HALO unit, which have a reported accuracy of 0.1 K.”

The baths do not have to be of constant temperature as long as the temperature is known for the period over which one wants to calibrate. So for each time step, we use the measured bath temperatures. We added: “For each measurement period, the bath temperatures from that same measurement period were used for the calibration.”

6291,9/10: How many values were averaged in total? You should use the stdv of the measurements to estimate the random error of the individual DTS measurements.

We have added this to the results: “The standard deviations of individual measurements within the stretches were 0.29°C for 3.0 mm white, 0.28°C for 1.6 mm white, 0.39°C for 3.0 mm black, and 0.35°C for 1.6 mm black. The standard deviation of the average of the 62 points would then be between 0.04°C and 0.05°C. It should be noted that the accuracy of the instrument is about 0.02°C for a point measurement with perfect calibration and that similar set-ups have given field accuracies of 0.1°C (Hausner et al., 2011). So the measured variation is likely to be also caused by real temperature differences along the cable, caused by uneven heating/cooling.”

6291,14: If the reference exhibits a radiative error, it is trivial that the values are close to another sensor that has a similar radiative error. Was the Temp/RF sensor shielded from solar radiation and artificially ventilated? Or was ist alt least shaded by a naturally ventilated shield (e.g. HOBO RS3)? If not artificially ventilated: Did you estimate and correct its radiative error? The quality of the reference has to be thoroughly checked and its accuracy should be well defined.

We have added: “The sensor was placed inside a HOBO RS1 radiation shield with only natural ventilation. No radiative error correction was applied so the actual accuracy so the reference temperature sensor accuracy was worse than the reported 0.13°C.”

In the conclusions we have further commented upon the lack of accuracy and the consequences for this experiments, as explained under the major comments.

6291,16-19: The wind is typically measured at 10m above ground at airports. The DTS fibers were at 1m above ground. Furthermore, the roughness lengths z_0 at both sites seem to be different. The CORINE-database says "airport" => 0.05m and TUD campus = "Discontinuous

urban fabric" => 0.6m (see e.g. http://www.eea.europa.eu/data-andmaps/ explore-interactive-maps/corine-landcover-2006)

As explained under major comments in the first reply, the wind measurements are not critical as they only serve to check if forced advection conditions are true. We added: “There can be important differences between the wind measured in Rotterdam and the wind at our site but these data only served as a check to see if there were no periods without wind (<0.1 m/s).”

6292,1st para: Does the difference T_air-T_fibre scale with the net radiation (or the incoming solar radiation, provided that the cloud fraction did not change much)? Compare my remark on (2).

There clearly is a relation, as one would assume, also given the differences between black and white and night/day but it is difficult to build a radiation balance for the cable without further information on emissivity, albedo, and the arrival of radiation at the cable both directly and indirectly. The derivation of (2) (now (1)) does not follow this pathway as explained above.

6291,18: Even 0.61K difference is not really a close match between two thermometers. The statement should be replaced by a conclusions which is quantitatively backed by the results (see remarks on the abstract).

Agreed. We have scaled down the expectations in abstract and introduction and now state in the conclusions: “The corrected temperatures matched the temperature measurements of the reference station with a RMSE of 0.38°C. The reference measurement took place without forced ventilation so the RMSE of the reference station could easily account for half the RMSE.”

6291,21-: I cannot agree with the conclusion that the accuracy is "reasonably good" for vertical profiling. A random error of 0.61K corresponds to an error of 0.9K for a temperature difference or - if I assume a vertical distance of adjacent measurements of 2m (fiber section length) - an error of 0.42K/m for the vertical gradient. For 10m it would be 0.08K, which is still much larger than the adiabatic lapse rate. Hence, such vertical profiles would be not very helpful. If, however, most of the RSME would stem from systematic (but not resolved in this study) errors and the random error of the DTS is much smaller (see my comment on the averaging) this statement might be supported by the measurements. But this has yet to be proven.

The reviewer is correct and we have adjusted the language accordingly. The final paragraph now reads: “The experiment started out with the question if corrected DTS air temperature measurements would have a reasonable accuracy to measure temperatures across urban landscapes. The results show that this is indeed the case and they were actually better than we expected at the onset (1°C). With hindsight, we would have installed a better reference temperature measurement device because we can not say if the reference was truly better than the DTS derived temperatures. Although we can not conclude this from the results, it may be possible to further reduce systematic errors, in which case this method would become valuable for vertical atmospheric soundings with DTS from balloons, quadcopters, or towers.”

Reviewer #3

Main comment

d) You stretched 750m of (different) cable with 3 splices and several spools for the calibration baths. Moreover, the Halo is not the non-plus-ultra of the instruments. You said that you measured in single-ended mode. I expect a quite high loss due to splices and distance, in particular for the thin white cable, the last segment. Did you check that? Did you also measure in single-ended but from “end to start” (from thin white to thick black) to compare and, in case, correct the temperatures? Or was the particular calibration that you applied able to compensate for the losses along the way? I think this is a crucial point that must be better explained. This also links to the reference temperature in the water baths. Please add more information (how did you monitor it? Fluctuations? Did you calibrate once or every measurements?) since the measured fiber temperature heavily depend on that.

Good point, although 750 m turned out to be still quite ok and not really that far so the signals were still strong enough. We did not shoot from the other side, although that would have been better. We checked for losses but given our set up, the splices were not a problem because there were baths between the measurement stretches and the splices. We adjusted each stretch following Hausner et al. 2011. We added: “The setup of baths and splices was such that the cable from each stretch went directly through the two baths without first passing through a splice, thereby avoiding step losses within stretches. The signal was checked to ensure sufficient strength, especially towards the end of the cable.”

We also added more information about the measurement of temperature in the baths and the continuous calibration (we did indeed calibrate for each measurement period of five minutes):

“Bath temperatures were measured with the two PT100 thermometers that came with the HALO unit, which have a reported accuracy of 0.1 K. Calibration of the fiber optic cable was based on the method described by Hausner (2011). For each measurement period, the bath temperatures from that same measurement period were used for the calibration.”

Technical comments

I agree with the technical comments posted by the first reviewer, therefore I will not repeat them. The authors will fix them properly. In particular, since the journal mainly focuses on the Measurements Techniques, more technical details, accuracies, comparisons with references, statistical indexes (rather than “good”, “close”, etc.) are preferable and they would strengthen the message. In the abstract, you should give more emphasis to the significant decrease of RMSE due to the correction applied rather than the r2. I think that is the most impressive result.

Done.

Beside that, please cut the vertical axis in Figure 2 to 30 C for the 4 panels. This would increase a bit the temperature traces and would be consistent with Figure 3.

Done.

In the conclusions: what do you finally think is the best solution for atmospheric air temperature measurements? White thin cable plus the correction? Or at the end the correction acts so well that it does not matter the color and the thickness?

Interesting question. We can conclude that for accurate measurements the correction is essential. Also, white gives lower standard deviation so is probably preferable. We now conclude: “In that case, the correction using different diameters is essential. White would be preferable over black, given the lower standard deviation within stretches.”