Answer to the Reviewer 2

We thank very much for a thorough and detailed review of our paper. We revised the paper according to the suggestions.

Below are the answers to the comments supplemented with a detailed list of changes in the text of the manuscript. The following convention is applied:

- standard font: referee comments;
- *italics:* reaction to the comments:

answers – red,

changes in the text - blue.

The present manuscript describes the improvement of an existing fast-response temperature sensor for airborne applications, which is not a novel topic.

At that, a more comprehensive introduction could help outlining what the research is important for, and what it is trying to achieve. For example, the applicability of findings from such sensor for improving cloud parameterizations in weather prediction etc.

In the abstract "aimed at in cloud measurements" was added.

In the introduction there is now a whole paragraph describing applications of the UFT-M:

The present study focuses on the description and performance of a specially designed version of the Ultra Fast Thermometer (UFT) (Haman et al., 1997, 2001), which was one of the key instruments used in the POST field campaign and may be used in future campaigns that involve airborne investigations of small-scale features of warm clouds. High resolution temperature measurements are needed to characterize filaments formed in the course of mixing of cloud with the environmental air. They allow to distinguish volumes undergoing active mixing from regions already mixed, characterize homogeneous and inhomogeneous mixing (see discussion in Devenish et al., 2012 and references therein), investigate a relative importance of radiative vs. evaporative cooling in clouds (see e.g. Wood , 2012), seek for effects of evaporative cooling at the interfaces between cloudy and clear air (Malinowski et al., 2008) . It can be also used, together with the other airborne data to characterize properties of different cloud layers and regions (see e.g. Haman et al., 2007; Malinowski et al., 2013). Last, not least, high-frequency temperature measurements can be used, together with the appropriate vertical velocity data in sensible heat flux estimates from slowly flying aircraft (like in Metzger et al., 2012).

The study itself is largely qualitative, and lacks rigorous quantitative assessment and proof. On p. 2096 l. 16, the phrase differ only by a marginal degree is symbolic for the mainly descriptive approach. For example, what are the propagated input uncertainty, but also accuracy, precision and resolution of the UFT-M? The use of statistical techniques (regression, variance statistics, uncertainty propagation…) would much aid the scientific significance of the paper. Also, why has only a subset of the available data been used for analysis? The exploitation of all available data could further help to improve the significance of the results.

We underline here, that UFT-M is not designed to measure temperature, but small scale temperature fluctuations in very variable conditions. In Bange (2013) the authors write: accuracy.

"To avoid extensive wind tunnel testing, most 'non–TAT' type sensors as well as radiative probes rely on in flight calibration and are, therefore, used together with a well characterized TAT probe. Under cloud free and stable conditions both instruments are compared against each other yielding a correction factor valid for these particular flight conditions..... Nevertheless, the achievable accuracy of in–flight calibration is always limited to the accuracy of the reference sensor."

The problem is that we have shown that the reference sensor is subject to wetting errors, a special kind of error in cloud measurements. We point, however that more rigorous analysis of UFT *performance is given in refereed papers of Haman (1997, 2001).*

We agree that exploitation of the all available data and use statistical techniques would help. The problem is, however, the amount and variability of the data collected. Due to the flight patterns and related variability of the records (see new Fig. 1) adoption of statistical procedures to the whole data record in order to determine properties of the sensor does not work, which is now clearly stated in the text (see answers to specific comments).

In order to partly account for this remark and to fulfil requirements of the Reviewer 1 we added discussion in the sensor description part:

A comprehensive description of accuracy, resolution and overall performance of UFT sensors can be found in (Haman et al., 1997). Below we discuss in more detail errors due to radiation and velocity fluctuations. The first is estimated from the comparison of heat fluxes due to solar heating of the sensing wire Qr and heat transfer Q between the air and the sensor at temperature difference ΔT. Q per unit length of the sensing wire can be estimated after chapter 2 of Sandborn (1972) from the formula:

$$Q = N_u \kappa \pi \Delta T \approx \kappa \pi \Delta T (0.318+0.69)(u D \nu^{-1}) \qquad (1)$$

where N_u is the Nusselt number, $\kappa = 0.025\ Wm^{-1}K^{-1}$ is the heat transfer coefficient for air, $u = 60\ ms^{-1}$ is the air velocity, $D = 2.5 \cdot 10^{-6}$ m is wire diameter, $\nu = 1.5 \cdot 10^{-5}\ m^2s^{-1}$ is the kinematic viscosity of air. This value compared to the radiative heat flux per unit length of the wire:

$$Qr = S A D \qquad (2)$$

where $S = 1300Wm^2$ is the maximum solar flux and $A=0.3$ is absorption coefficient for platinum gives the maximum temperature effect of radiation $T = 0.005$ K.
Temperature effects of dynamic pressure - changes of the airflow velocity - also affect temperature measurements (see e.g. section 2.5.1 in Bange et al. 2013). They depend on the recovery factor of the thermometer. Estimates the UFT recovery factor by Haman et al. (1997) give value of 0.6. This, together with the measurement speed of 55 ms^{-1} and fluctuations in the range of ±5 ms^{-1} result in the shift of temperature by $T = 0.2$ K and fluctuations ±0.04 K The value of shift is of secondary importance, since UFT's are designed to measure small-scale temperature fluctuations, not static temperature and are used together with calibrated low frequency temperature probe (see section Calibration and averaging).

The data processing techniques are not up to date, so the authors have to deal with residual spikes, lags among calibration variables etc.
Before writing the manuscript we tested several de-spiking techniques. In the course of preparation of the revised version of the manuscript we also tested statistical techniques suggested by the Reviewer 2, which are aimed mostly at flux measurements with eddy covariance methods. We found, that these techniques do not work with our data set (very different statistical distribution of recorded values in different parts of the flight) and lead to rejection of potentially interesting and valuable segments of records. This is described in the answers to the detailed comments below.

Specific Comments
p. 2085 l. 1: Inconsistent naming convention: modified UFT-M in the title, modernized UFT-M in the text.
Corrected.

p. 2086 l. 6: What kind of interaction with the avionic system?

An explanation is added, the appropriate sentence is now:
"Oversampling the data allowed for the effective correction of the artefacts resulting from the interference with the radio transmissions from on-board avionic systems and the thermal noise resulting from the sensor construction."

p. 2089 l. 11 ff.: The structure of the paper is not very intuitive. I suggest considering the following revision:
After discussion we decided to keep the present structure, we think that in a case of the present paper sections like "Materials and methods" are misleading.

p. 2089 l. 19 ff.: Can this behavior be described, and maybe be corrected, using the Bernoulli equation?
A description was added:
"Unfortunately, pressure fluctuations in the eddies shedding from the rod cause temperature fluctuations (i.e., aerodynamic thermal noise) of an amplitude that increases roughly with the square of the true airspeed (TAS)."

Temperature fluctuations due to vortex shedding from the rod cannot be corrected by the Bernoulli equation. Detailed information on dynamics, positions and structure of vortices would be necessary. See discussions of these effects in Haman et al., 1997, 2001 and Rosa et al 2005, (refereed in the paper).

p. 2090 l. 19 f.: Why would you place the sensing wires in stainless steel tubes? This is not apparent from Fig. 1.
We corrected the text:
The sensing wires are soldered to the tips of the Teflon-coated copper connectors placed inside stiff stainless steel tubes.

p. 2090 l. 25: What is a unit mVpp?
Milivolts peak-to-peak. The text is adjusted to avoid misunderstandings..

p. 2091 l. 4: If this temperature drift is constant, can it be characterized and corrected?
No, the drift is not constant. Sensors are hand made and each one is slightly different.

p. 2091 l. 9: special → particularly?
Corrected.

p. 2091 l. 15 ff.: As the thermometer is mounted at a fixed tilt, it cannot be parallel to the airstream at all true airspeed settings/ attack angles encountered during flight. Is the effect of such deviations negligible?
Yes. Earlier tests in the wind tunnel proved this.

p. 2091 l. 18: How can the flow in the boundary layer around an aircraft be undisturbed?
The sensor is placed outside the surface layer, there are certain distortions of the flow in the position of the sensor but no turbulent disturbances. A suitable explanation is added in the text

p. 2092 l. 12: middle → centre?
Middle removed.

p. 2093 l. 4 ff.: The proposed spike detection algorithm is unable to capture spike events that affect successive (and not just single) observations. Also, in Sect. 3.2 "residual spikes" are mentioned, and in Fig. 6 (bottom panel) it appears that there is a recurring spike at 0.2 K every 0.005 s. Is there any reason why established spike detection algorithms based on sliding windows (e.g., Hojstrup, 1993; Vickers and Mahrt, 1997) are not used? Moreover, the use of distribution statistics (e.g., median and median absolute deviation) would aid the robustness of such method (e.g., Mauder et al., 2013; Metzger et al., 2012; Papale et al., 2006), and replacement through linear interpolation helps minimizing the effect on power spectra.

The goal of the spike removal was not minimizing their influence on power spectra, but on 1000Hz averaged data submitted to the POST database. The spikes were effect of electromagnetic interference with the digital radio signal transmission. The protocol of transmission was such that the spikes were single events, sometimes in groups. For the purpose of the POST database the proposed linear interpolation of spikes of the amplitude exceeding the maximum amplitude of the other artefacts (vortex shedding) seems satisfactory.
Nevertheless we tested suggested spike-removal algorithms (equations 1-4 in Papale et al., 2006, Mauder et al., 2013, and found that the variability of observed temperature structures in various segments of the record (free tropsphere, EIL, cloud top region) is such, that algorithms based on Median of Absolute Deviations (MAD) lead to false of the significant segments of the data. Suggested by the authors manual verification of data series is not possible, considering 10GB records from each flight. We tried to modify the approach, applying MAD's to shorter segments of the data series but found the method impractical. After these trials we are even more convinced that the purpose of the POST database the proposed linear interpolation of spikes of the amplitude exceeding the maximum amplitude of the other artefacts (vortex shedding) seems satisfactory. We agree that more sophisticated algorithms suggested by the reviewer can help in analysis of the statistically uniform segments of 20kHz records.
The appropriate section of the text was rewritten:

"We tested several error-correction algorithms, e.g. based on Median of Absolute Deviations (MAD) as in Papale et al. (2006) and Mauder et al. (2013). We found, that statistics of temperature fluctuations in various segments of the record, (e.g. free troposphere, temperature inversion, cloud top region) are very different. MAD-based algorithms lead to rejection of many good and interesting data segments, even when applied to e.g. high-pass filtered data in order to remove large scale temperature fluctuations. Finally, since spikes were typically single-point events, a simple detection algorithm produced satisfactory results"

p. 2093 l. 21 ff.: The sensitivity of the sensor is likely to change with aging of the wire. Has a periodical re-calibration/validation been performed?
As mentioned in the text, each sensor was used typically in one, maximum in two consecutive flights Calibrations against UCI Rosemount were performed in each flight.

p. 2093 l. 27 ff.: In particular for calibration purposes any lags should be corrected. This can be easily achieved through maximizing the lagged correlation between two measurements and subsequent shifting (e.g., Rebmann et al., 2012).
As discussed in the text, this is not a linear lag resulting from temporal shift of the signal, but the effect of different response of the sensors.

p. 2094 l. 8 ff.: Records from all three sensors are representative of the same sampling volume. For this statement to hold it would require mentioning the exact 3-D displacements among sensors. Even for displacements below 1 m at 25 m s−1 true airspeed, sensors are not representative of the same air volume in that conversions via the ideal gas law (e.g. calculation of dry mole fraction) do not hold true.
This is an approximate statement referring to the typical measurement speed of the CIRPAS Twin Otter aircraft, which cannot fly as slow as 25m/s.

p. 2094 l. 28 ff.: Can the "signature" of the Rosemount housing wetting be quantified?
No.

p. 2096 l. 1 ff.: See comment to p. 2093 l. 4 ff. Simply averaging over spikes is not a sound procedure,especially as capable spike detection algorithms are readily available (also see Goring and Nikora, 2002).
Spikes in our data series are "single point" events effecting from digital radio emissions.. Polynomial filling of the gaps, as suggested by Goring and Nikora, 2002 is not applicable.

p. 2096 l. 3 ff., p. 2098 l. 22 f.: What is the calculation basis for these uncertainty estimates? Absolute error, accuracy, precision, median absolute deviation…?
There are new elements in section 1 and in description of Fig. 10 with more quantitative analysis of sensor errors and precision of averaged data.

p. 2096 l. 23, p. 2098 l. 5: TO10: This descriptor is meaningless without an overview (table of similar) of all available research flights.
In p. 2086 l. 23-25 we refer to publications and open database with information on all flights.

p. 2097 l. 20: The time scale corresponds to this scale in the upper panel. I do not understand. Maybe a graphical depiction would help?
Frames showing expanded segments were added to Panels 1 and 2. Time axis was changed to the relative time.

p. 2097 l. 25: Temporal resolution?
Description added.

p. 2097 l. 27 ff.: It would be good to provide a table of 3-D sensor separations, which would also serve other use cases of the published/online available data
Explanation and references to the detailed sketches are added in the text.

p. 2098 l. 1: What are larger sizes?
Rewritten, now:
"Thus, the correlations of temperature and LWC in cloudy volumes on scales larger than ~0.5m can be investigated."

p. 2098 l. 5: S s−1, p. 2098 l. 26: samples per second… – why not simply using Hertz (Hz)?
This was introduced to distinguish between the sampling frequency and frequency in Fourier spectrum.
Also, the units S s−1 and kS s−1 have never been formally introduced in the text.
They have bee introduced in p. 2092 l. 3 (kS s-1) and p. 2093 l. 14 (S s-1).

p. 2097 *(should be 2098)* l. 10: The power law decay with −5/3 slope applies to (i) homogeneous and isotropic turbulence,
and (ii) eddy sizes much smaller than the integral scale but much larger than the Kolmogorov viscous dissipation scales. Can (i) be justified for cloud crossings, and (ii) for frequencies up to 500 Hz, as shown in Fig. 5?
All data shown here were collected in turbulent environment. Range of wavelengths spans from a few hundreds of meters (corresponding to the distance between time ticks in Fig.5 and to the integral scale of turbulence) to 10 cm which is few tens of Kolmogorov scales.

Does the spectral flattening above 100 indicate that the dissipation scales are reached?
No. Dissipation scale is revealed by steeper spectra. What we see that there are increased

temperature fluctuations in these scales.

Figs. 4, 8, 10: Information is provided in the figure legend, so single figure panels do not require an additional title.
We decided to leave the titles providing redundant information to the readers.

Fig. 7: Figure legend is missing.
We have it in our printout from AMT web page. Nevertheless the legend is improved:
Temperature fluctuations from two sensing wires of UFT-M. Temperature record from UFT-M2 was shifted up for comparison with that from UFT-M1. Colour code corresponds to that in Fig.6.

Fig. 9:
What are Sc, EIL? Abbreviations should be introduced.
Top panel: …in the course of descending…;
Center and bottom panels: the colors are different from the legend in the top panel. What is the meaning of the different colors?

Figure description was improved according to the suggestions:
Temperature fluctuations (red) recorded during the descent into the stratocumulus cloud. The smooth signal at 0 – 9 s is identified as recorded inside the free troposphere; that at 9 – 21.2 s as the temperature fluctuations in the inversion capping the cloud deck (no significant LWC, blue)); and that at 21.2 – 30 s as the temperature fluctuations at the cloud top and inside the cloud. The two lower panels present selected blow-ups of temperature records to illustrate sensor performance. Colour code on these panels correspond to that in Fig. 6.

The Bottom panel: The averaged values appear time lagged. Has a leading window been used during the averaging process, or is this also a result from not lag-correcting the data?
We used a centered window. No more sophisticated method was necessary, there were no other 1000s/S (5.5cm resolution) data for a direct comparison.