Replies to the comments if the anonymous referees

We would like to thank all referees for their careful review and the comments provided. Our replies are shown indented in italics.

Anonymous Referee #1:

This paper discusses the uncertainty factors and budget of the Cryogenic Frostpoint Hygrometer (CFH) measurements. The CFH is now widely used in both research and operational purposes for tropospheric and lower stratospheric water vapor. It is also one of the key instruments for validation of various other water vapor instruments (i.e., balloon-borne and aircraft in situ instruments, and ground-based and satellite-borne remote sensing instruments). Thus, this is a very useful paper for the CFH operators and CFH data users. The paper is well written, and I have only a few minor comments.

Introduction. It is useful to show the formula of volume mixing ratio and relative humidity from dewpoint/frostpoint temperature. With these formula, it is understood that radiosonde pressure and temperature measurements and the choice of actual expression of the Clausius-Clapeyron equation are also the important sources of uncertainties.
It would be useful to show the way how to evaluate the total uncertainty of volume mixing ratio and relative humidity with additional uncertainty from pressure and temperature measurements. These discussions might be placed in a later section.

> *A new section 2.9 has been added, which discusses the impact of pressure and temperature measurements in detail. A new Figure 10 was added to show the relative contributions of the ambient pressure uncertainty to the mixing ratio uncertainty and the contribution of the temperature uncertainty to the relative humidity uncertainty.*

Pages 4-5. Discussion on the time lag. It is pointed out here that Hasebe et al. (2013) previously discussed the time lag of CFH to be 4-10 sec in the upper troposphere to the stratosphere, being larger at higher altitudes. The results here are broadly consistent with the Hasebe paper.
Hasebe, F., Inai, Y., Shiotani, M., Fujiwara, M., Vömel, H., Nishi, N., Ogino, S.-Y., Shibata, T., Iwasaki, S., Komala, N., Peter, T., and Oltmans, S. J.: Cold trap dehydration in the Tropical Tropopause Layer characterised by SOWER chilled-mirror hygrometer network data in the Tropical Pacific, Atmos. Chem. Phys., 13, 4393-4411, doi:10.5194/acp-13-4393-2013, 2013.

> *The information in the Hasebe paper were contributed by Vömel and are justified and explained in more detail here.*

Page 10, lines 4-6. It would be useful to describe the recommended (and unrecommended) way of mirror cleaning prior to launch.

> *To be more precise we have modified the text slightly: 'Improperly cleaned mirrors, in which measurable residue remains on the mirror, may lead to very different condensate layer morphologies and thereby to unexpected instrument behaviors.'*
> *The users are trained to clean the mirror until no further increase in mirror reflectivity can be achieved. The term 'measureable' refers to this increase in reflectivity. Contamination on the mirror may result from outgassing of plastics and glues, dust, spider webs, sea salt spray near oceanic sites and other potential sources that may lead to thin film deposits on clean metal surfaces. Experienced users know how to efficiently remove these deposits during the mirror cleaning procedure, which is explained in detail in the instrument manual.*

Page 10, lines 17-18. It would be useful to describe the criteria for contamination.

> *We added the following text to the Appendix to explain the flagging criteria in more detail:*
> *"The strongest indication of contamination is a sudden increase of the water vapor mixing ratio in the upper part of the stratospheric profile, which reaches unrealistic values. In simple cases, ascent and descent measurements agree below this level; in more complicated situations, flagging of potentially contaminated data is based on experience and done conservatively. Flagging then takes into account the level of agreement between ascent and descent data, the level of contamination sources in the troposphere, deviations from a climatological mean, and information about the unwinder performance.*

In all cases, a single altitude is determined above which all data are flagged as contaminated and below which data are considered not contaminated."

Page 12, line 3. A photograph showing the ground check system as an example would be useful.

We added a schematic (Figure 11) of the reference sensor and the propeller fan inside the inlet tubes to show the arrangement of ground check and the sensor volume.

Page 12, line 20. "section 0"

Fixed.

Summary. It would be useful to show typical uncertainty values also in ppmv and %RH, by assuming typical uncertainty values of pressure and temperature measurements from recent radiosonde models.

See above.

Anonymous Referee #2:

This manuscript describes recent improvements in the understanding and treatment of errors and uncertainty in balloon sonde water vapor measurements made using the CFH chilled mirror hygrometer. The CFH is used by a number of groups worldwide for measurement of water vapor, particularly in the UTLS where water vapor concentrations are typically below those measurable with standard radiosondes. The CFH has been used as an in situ comparison for validation of satellite and lidar measurements of UTLS water vapor.

The subject matter is highly appropriate for Atmospheric Measurement Techniques and overall the manuscript is well organized and clearly written. The manuscript will be of interest and useful to the atmospheric science community engaged in studying UTLS water vapor and its role in climate since the CFH is an instrument that has the potential to contribute significantly to a long-term record of UTLS water vapor changes to study interannual variation and identify trends.

General comments: The authors state at the beginning of the introduction that "Cryogenic frostpoint hygrometers are widely considered as reference instruments" and then more generally in section 2.2 that "Frostpoint and dewpoint hygrometers. ... are not calibrated against water vapor standards and are considered water vapor standards". The authors, however, also state in the introduction "not all frostpoint or dewpoint hygrometers are equivalent, and some understanding of the technical realization is needed to properly interpret the reported frostpoint temperature and to be able to estimate the measurement uncertainty." Many laboratory frostpoint instruments are certified by comparison with metrological water vapor standards, which verifies that the reported mirror temperature does accurately represent the equilibrium saturation temperature– more than just a traceable calibration of the thermistor itself. This seems to be part of the motivation for the "manufacturer-independent ground check" that has been instituted.

At some of the metrology labs high quality frostpoint instruments are even used as primary standards for water vapor at low concentrations. Laboratory instruments can and should be re-calibrated against primary or secondary standards at regular intervals. This cannot be done in the strict sense for disposable instruments and the proposed ground check is the next best operationally feasible approach. Another part of this proposal is simply based on bad experience (e.g. section 2.4). In fact the last sentence of the manuscript recommends a manufacturer independent ground check for all disposable sounding instruments.

The opening sentence uses the term 'reference' in the loose sense often used in the community as substitute for 'best available'. The paper then becomes more specific with the interpretation of 'reference' (see next comment) and the ground check is one component for that.

The authors invoke the framework of Immler et al. (2010) in the discussion of "reference", but perhaps for the reader a more complete definition of how the term is used here would be helpful, especially given the extensive discussion of data filtering.

We have added some text to summarize the essential points from Immler et al and point out, which of these are being discussed in the paper. The last paragraphs of the introduction have been changed: "These instruments have been used in a large number of studies of upper tropospheric and stratospheric water vapor (e.g. Vömel et al., 2007b; Hasebe et al., 2007; Fujiwara et al., 2010; Selkirk et al., 2010; Shibata et al., 2010). Although they are recognized by many as reference instruments, we refer to the rigorous definition of what constitutes a reference observation given recently by Immler et *al. (2010). This paper defines reference quality atmospheric observations as such, which are based on traceability, a detailed analysis of the uncertainty budget, and a detailed knowledge of the calibration procedures and data processing algorithms, which are required for determining the uncertainty of each individual data point.*

The present paper discusses the measurement uncertainties of the CFH within this framework and describes the advanced processing algorithms, calibration record, a systematic error correction, and traceability. The work leading up to this paper has resulted in some instrument improvements over the work presented by Vömel et al. (2007a), which are discussed here. The sequence of processing and data quality control steps from raw data to final data product is described in Appendix A."

The brief description of the principle and operation of the CFH (with reference to the more complete description in Vömel et al. (2007)) should be expanded somewhat and care taken to be precise, such as the mirror is actively illuminated and the control signal is a decrease in the light reflected by the mirror due to diffuse scattering by the mirror condensate. This could be accomplished with only slight modifications and additions to the text.

We have slightly modified the instrument description, most importantly including the mirror temperature measurement, which had been absent. The paragraph now reads:

"Here we focus on the Cryogenic Frostpoint Hygrometer (CFH), which has been described in detail elsewhere (Vömel et al., 2007a). This instrument uses the chilled-mirror principle, in which the water condensate is formed on a small temperature controlled mirror, which is exposed to ambient air flowing across the mirror. An optical detector senses the condensate by measuring the amount of light that is reflected off the mirror and a digital controller regulates the temperature of the mirror in order to maintain a constant reflectivity of the condensate covered mirror surface. To the extent that the reflectivity is constant, the condensate on the mirror is assumed to be in equilibrium with the gas phase. The temperature of the mirror is measured using a small individually calibrated thermistor. Under the condition of equilibrium it is considered to be equal to the ambient dewpoint or frostpoint temperature, depending on whether the condensate phase is liquid or ice."

The focus of the manuscript is on the uncertainty related to the PID control loop, but for context and to present a complete picture, a brief discussion of the parameterization used to relate the saturation temperature to partial pressure, and the reliance on co-measured temperature and pressure to determine mixing ratio and relative humidity and the contribution from these to the overall measurement uncertainty would be helpful.

See reply to the same comment from reviewer 1.

Another comment is related to the ordering of Figures 9 and 10–it seems they should be reversed to match the order in which they are discussed in the text.

Corrected.

Specific comments and suggestions on manuscript text:
Minor grammatical: In many instances, commas are missing following introductory clauses.

We went through the entire manuscript and improved the use of commas.

Page 1, line 20: "considered as" could be changed to "considered to be"–but see comment on "reference" above.

Corrected.

Page 1, line 27: "grows nor shrinks"-more specifically, its scattering is maintained at a constant value.

While the reviewer is correct that the scattering is maintained at a constant value, we prefer to use the current wording. Condensate grows at temperatures below the equilibrium temperature and shrinks above. This is easier to relate to the frostpoint or dewpoint temperature.

Page 2, line 32: "regulate the bulk reflectivity"–it isn't the condensate that is providing the reflectivity, but decreasing it due to diffuse scattering.

The reviewer is correct that the instrument measures the reflectivity of the mirror with diffuse scattering by the condensate layer. However, combining both in the term 'bulk reflectivity of the condensate layer' may be easier to visualize and focuses on the important role the condensate layer plays. We have left the current expression unchanged.

Page 3, line 19: "in a poorly behaving"

Corrected

Page 6, line 29: recalibrated over what time period?

The calibrations were run since 2004, which has been noted in the text.

Page 6, line 30: “is less than 0.02 K”–with a couple of exceptions.

Added the words “(with a few exceptions)”

Page 8, line 15: Highest RH in 1 s, or within some averaging/smoothing period? Figure 1R would return a higher RH than Figure 1L, even if the central values were identical.

The maximum of the filtered RH was selected, not the raw data. The word ‘filtered’ was added to this sentence to make this difference clear.

Page 9, line 29: “does not significantly change throughout the time of a typical sounding”–was this tested under large dynamical changes in H2O such as might be encountered in the atmosphere? It seems likely to hold as long as the mirror condensate does not experience significant changes in scattering, i.e. error signal remains small at all times.

This statement is based on the frost layer observations during the first AquaVIT campaign. In some of the experiments the frost layer was left alone for several hours, while the water vapor concentration in the chamber was changed significantly. We could observe small changes in the appearance of the frost layer, which had no impact on the behavior of the PID controller or on the comparison with the other instruments. In a typical sounding, the frost layer is re-formed in the upper troposphere and is detected for at most 60 more minutes during ascent. The AquaVIT experience translates directly to the frost layer control of the CFH in the upper troposphere and stratosphere during ascent.

At the balloon ceiling, frost layer control becomes undefined for a short period of time, when the airflow stops and reverses. We could not study how this flow reversal impacts the frost layer. Since measurements are usually used on ascent, this issue is irrelevant for the CFH.

Page 10, line 16: “measurable” contamination is almost never observed in the troposphere.

*Added the word ‘*measurable’*.*

Page12, line 13: “the CFH agrees with this reference to better than 0.1 K”–CFHs are one- or few-use devices not a singular instrument, and based on figure 9 I would say “individual CFHs (typically) agree with this reference to within 0.2 K” or similar. The “typically” would cover the two instances where the difference was observed to exceed 0.2 K.

Agreement to within 0.1 K refers to the standard deviation of the scatter of all instruments used. We try to avoid the use of maximum deviation, which is more difficult to establish for a population of instruments, when only a small sample is available. The estimated one standard deviation uncertainty of 0.1 K based on the limited sample seems to be a better statistical statement. One might claim that this choice of definition makes the instrument appear slightly better than it is. On the other hand, the scatter combines the uncertainty of the CFH and that of the polymer sensor based reference. If we assume that the uncertainty contributions are evenly split between the two instruments, then the uncertainty of the CFH would be only half of what we claim. Since this is only semantics, we have left the 0.1 K as is and added ‘to within 0.1 K at one standard deviation’. This wording avoids having to discuss outliers, which may exist in a nearly Gaussian distribution.

Page 12, line 20: “section 0” should be “section 2.4”

Corrected.

Page12, line 26: While truly random errors will not produce incorrect long-term trends, they do affect the ability to detect trends and are therefore quite important in terms of understanding the long-term behavior of the variable in question.

The reviewer is absolutely correct. We have changed to wording to “Random errors in the measurements are less likely to impact long-term trends, but strongly impact the ability to detect long term trends. Changes in systematic errors, on the other hand, may impact long-term trends to the extent of the change directly.”

Page 17, Figure 1: “responses”

Corrected.

Page 18, Figure 2: “Water vapor profile from a CFH sonde launched at Lindenberg”. Cause of the high/low deviations in the descent profile following time-lag correction?

Any timelag correction will artificially amplify fast signals and in particular noise, which needs to be filtered out again. These faster signals are a result of the PID controller and the behavior of the condensate layer and are difficult to characterize. The high low deviations in Figure 2 are a result of this improper filtering and one of the reasons, why a time lag correction is not done in data processing.

Page 21, Figure 5: As noted above, time span over which the calibration runs were conducted?

The calibrations were run since 2004, which has been noted in the text.

Anonymous Referee #3:

This is an excellent paper on the CFH measurement uncertainties. CFH is a reliable, light weight instrument, which provides accurate measurements of water vapor in the upper troposphere and lower stratosphere. The instrument is widely used to study UTLS processes, to provide reference observations for satellite and lidar measurement validation and also for comparisons with radiosonde measurements. The paper describes a method to estimate uncertainty of the CFH water vapor profiles. Regarding some older versions of the CFH, the authors have studied a small bias found in the lower troposphere. They also provide a method to correct the bias. Since such biases may occur in the future, the authors recommend the use of an additional ground check prior to the launch. The authors find that the ground check introduced in 2014 confirms that the systematic error of the instrument is less than 0.1 K and no drift can be seen during the recent time period.

General comments: In the paper the CFH is characterized as a disposable instrument.
It would be interesting to learn if recovered instruments could be flown to reduce the cost of such measurements.

Instruments can be reused if their condition after recovery is acceptable.

Vertical range of the measurements by the CFH instrument depends on the possible contamination due to outgassing from the balloon, the parachute, the load line or the intake tube of the instrument. The contamination affected data are flagged during the data post-processing. It would be useful to include some more detailed discussion on how to perform the flagging. It is possible that in some cases it would be difficult to separate real variability from a suspected contamination.

See response to the same comment by reviewer 1. We added a short description to criteria used in the flagging.

The reviewer is correct that there will always be ambiguous cases, where true variability cannot be distinguished from artificial contamination. To reduce the risk that contaminated data are used in scientific studies, flagging is usually done quite conservatively.

Specific comments and suggestions: Page 2, line 15. Add references for papers published after 2010, which have made use of the CFH measurements.

Corrected

Page 8, line 14. Add locations, where these 1022 soundings were made.

The soundings were launched globally at 35 different sites. This number has been added to the manuscript.

Page 9, line 22. Figure numbered 9 should probably be before Figure 10 in the text.

Corrected

Page 11, from line 14. Missing commas in several sentences.

We went through the entire manuscript and improved the use of commas.

Page 11, line 14- 15. ” For tropospheric and stratospheric observations the CFH is a fast-responding instrument and lag issues are not suspected”. Does this mean that time lag is not impacting the measurements during balloon ascent?

Correct. In the lower troposphere the response time of the instrument is believed to be faster than 1 s. In the stratosphere it may increase to on the order of 10 s. At 5 m/s this would correspond to 5 m in the troposphere and on the order of 50 m in the stratosphere. All studies we are familiar with focus on larger layer averages and in that sense time lag are not suspected to impact measurements during ascent.

Page 12, line 13. From the figure it looks like the dots are within 0.2 K?

The dotted line at 0.2 K is shown as visual guide for the approximate max/min of the data. The uncertainty of the mean difference is given as one standard deviation following the convention throughout the paper and is indeed 0.1 K.

Page 12, line 20. “section 0”, replace with “2.4.”?

Fixed.

Page 24, the figure 8 caption. Should it read for example “Distribution of the correction as a function of altitude for all soundings, where a bias is suspected”?

The caption to Figure 8 was changed to “Distribution of the thermistor installation correction as a function of altitude for all soundings, where a bias is suspected.”