Reply to the Interactive comment on "Quantifying the value of redundant measurements at GRUAN sites" by F. Madonna et al.

Anonymous Referee #1

> *The authors of the manuscript "Quantifying the value of redundant measurements at GRUAN sites" by F. Madonna et al. have checked the received review of their manuscript and they would like to acknowledge the anonymous reviewer for his/her detailed work on the manuscript and the useful comments. In the following, the authors report their general comments as well as well a detailed reply to the reviewer (in italic).*

---

*This manuscript describes several metrics from information theory for quantifying the uncertainty and redundancy of observations of the same atmospheric variable, and it applies these metrics to up*per air water vapor observations from several GCOS Reference Upper Air Network (GRUAN) sites. I learned a lot from reading this paper and feel it introduces some potentially important concepts to the atmospheric measurement and climate research communities. Although the paper is a little hard to follow at times, I feel it would be acceptable for publication in AMT after the comments below are addressed.

General comments

1. Acronyms. There are too many new acronyms, which make the paper difficult to read. They include a set for instrument types, a set for GRUAN station names, and a set for statistical measures. I suggest limiting acronyms to the statistical measures only and spelling out all other words. That will both make the key concepts (the statistical measures) stand out and make for smoother reading.

- *According to the reviewer's suggestion, in the new version of the manuscript, the acronyms are used only for statistical measures and for those instruments that are usually reported in literature with the their acronym (e.g. GPS, MWR), while the acronyms for the other instruments have been spelt out.*

2. Interpretation of concepts from information theory. In discussion of mutual correlation (MC) and distance (D), the authors should recognize that these concepts are probably new to many readers. The descriptions and interpretation are generally good, but adding some illustrative examples, using simple data distributions (not GRUAN data), would help explain how to interpret MC and D. Please consider illustrating MC and D by plotting from some invented datasets, with known means, variances, ranges, biases, etc., plotting those (maybe both as series and as scatterplots) then computing and interpreting MC and D. Some points worth illustrating include: linear vs non-

linear redundancy (6333/7-9); whether 0<D<1, as one might guess from Figure 4; relationship between the information theory variables and more familiar statistical measures.

- *The authors thank the reviewer for this comment. In order to show the reader the real advantages of using entropy and MC with respect to standard deviation and correlation and to make the reader more familiar with the concepts of information theory, the authors added a new figure (Fig. 3) showing the traditional Taylor diagram obtained for the GPS IWV time series collected at Lindenberg (already shown in Figure 2): the values of the Taylor diagram are obtained from the original series by adding 5, 10, 15, 20, and 40 outliers to the IWV probability density function. The value of the standard deviation in the diagram obtained from the original time series is reported as observed. This is compared, following previous studies, with a modified Taylor diagram (Correa and Lindstrom, 2012) obtained by replacing the standard deviation with the entropy and ρ with MC. Correlation and MC have been calculated with respect to an underlying Gaussian distribution, fit in with the data.*

3. Data detrending (6337/13) and vertical averaging (6339/10). The motivation for removing linear trends in the datasets is not explained, and it seems possible that the results might depend on this data processing choice, particularly if the trends in different sets of observations are unequal. Either an explanation for this choice is needed, or the data should not be detrended. Similarly it does not seem kosher to perform vertical averaging of the profile data before computing normalized entropy (H/log(n)) or MC. The resulting smoothing must change the numerical values of H/log(n) and MC, doesn't it?

- *The data detrending is performed in order to minimize the effects of the heterogeneous non-stationarity (e.g. spurious biases or calibration drifts) affecting different instruments in a different way over the time that might affect the results and induce any misinterpretation. The aim of the presented study is to assess the random component of the uncertainty budget affecting a measurement time series, assuming that biases and calibration issues could be strongly minimized by the implementation a rigorous quality assurance and best practice program. Anyhow, in the presented datasets, the trend provided very minor effect and even if considered it would not change significantly the results.*
Regarding the smoothing of the datasets reported in Figure 5 (now (Figure 6), this was uniformly performed on all the three datasets from lidar, AERI and sondes with the aim to degrade the effective resolution of the higher resolution instruments (lidar and sondes) to match the effective resolution of AERI. The inversion applied to radiances are usually the solution of a an ill-posed problem: depending on the inversion and the number of information and constraints available for the retrieval, the effective resolution is typically of 100-500 m in the PBL and rapidly degrades to 1-2 km in the mid troposphere. Therefore, although it is true that smoothing can slightly alter the absolute values of the entropies, it is also true that comparing data over the same vertical resolution is more important

to avoid misinterpretations of the data provided by different instruments and measurement techniques.
*In addition, the authors want to point out that they have removed the sentence "**Entropies have been smoothed to obtain an effective vertical resolution of 60 m**" (line…) in the new version of the manuscript because in Figure 6 (now Figure 7) the data are shown at their effective raw resolution.*

4. GRUAN uncertainty estimates. A major hallmark of "GRUAN data products" is that quantitative and complete uncertainty estimates accompany all observations. I recognize that some of the observations used in this investigation have not yet become GRUAN data products, but I think some have (e.g., column water vapor from GPS). I'd like to see the uncertainty of observations plotted along with the measured values, to judge whether other observations fall within the GRUAN uncertainty estimates, in Figure 2, for example.

   - *It is important to note that not all the considered station are routinely providing the uncertainties related to each instrument. However, (to help the reader in the interpretation of the results), typical uncertainties affecting the considered measurements are mentioned: radiosonde water vapor mixing ratio profiles have a relative uncertainty typically lower than 6 % from the surface till 15 km a.g.l., though the uncertainty might change depending on the measurements conditions (more details in Dirkseen et al., 2014); Raman lidar relative random uncertainty increases with height and keeps less than 25 % at 7-8 km a.g.l. plus a calibration error typically within 5-10 % affecting the entire profile; the uncertainty on the integrated water vapor content achievable with microwave radiometers and profilers is strongly depending on the retrieval types, but it is typically within about ±0.07 cm; the GPS uncertainty on the integrated water vapor content is typically within about ±0.15 cm (first results from GRUAN comparisons with CFH).*
   Moreover, to avoid any ambiguity the authors has contacted one of the members of the GRUAN Task Team, who is dealing with GPS/GNSS data (Dr. Galina Dick, GFZ Potsdam). Her reply is here reported "Concerning the uncertainty of the GPS derived IWV, this is one of the discussion points in the GRUAN GNSS…….
   Concerning the status of GPS processing for GRUAN:GFZ is working together with GRAUN LC at Lindenberg to maintain the data flow between GFZ and Lead Centre (LC) in Lundeberg. The long term data archive will be maintained by LC, data conversion and processing will be done by GFZ. Both, LC and GFZ, installed FTP server to exchange GNSS raw and RINEX data, metadata and products. Stations Lindenberg and Ny Alesund are first two sites for data flow. GFZ plan to do near-real-time processing, also reprocessing is in plan".
   According to Task Team report published on the GRUAN website (http://www.dwd.de/bvbw/generator/DWDWWW/Content/Projekte/Gruan/Downloads/ICM6__2014/doc__42__ttreport__gnsspw,templateId=raw,property=publicationFile.pdf/doc_42_ttreport_gnss-pw.pdf).

Data flow through NCDC portal is expected by 1. Jun. 2015.
Finally, no data are available yet on the NCDC GRUAN archive (checked 11 August 2014).

5. Figures. The figures are generally good, but a few need to be reconsidered and redesigned. In Fig. 4, the grouping of histogram bars is problematic for Potenza, in part because of the way the x axis is designed and labeled. Figure 7 is not effective in communicating the way normalized conditional entropy changes with grouping of measurement methods. Some other way of displaying this information is needed, although I don't have a good suggestion for an alternative graph.

- *Figure 4 and Figure 7 have been improved according to the reviewer's suggestion.*

6. Conclusions. I'd suggest including some concluding statements about the merit of using information theory, in general, and specific statistics from information theory, in particular, to quantify measurement redundancy.

- *The last paragraph is now written as follows "As a whole the concepts of entropy and mutual correlation demonstrate their potential if used as metrics for quantifying random uncertainty and redundancy in time series of atmospheric observations. The examples discussed in this work support the use of the mutual correlation as a more general concept than other linear metrics for the study of redundant measurements. Moreover, the analysis based on the entropy, MC and conditional entropy can be used for a preliminary feasibility study of the effective advantages obtained in using retrieval algorithms integrating measurements provided by different observation platforms, ground-based or satellite, both for direct measurements (e.g., radiances) and retrieved products (e.g., temperature, water vapor content, aerosol optical depth)."*

Specific comment and suggestions (by page and line)

7. 6327 Avoid using an acronym in the title.

- *Ok*

8. 6328/5 Specify what humidity parameters are used and whether they are column-integrated or profile parameters. Are these variables measured directly or calculated from some other observed quantity?

- *The water vapor mixing ratio and the relative (humidity?) are measured directly with soundings and Raman lidar (using a calibration procedure). The precipitable water vapor is estimated from radiances measured by microwave radiometers and from the refraction in the propagation of a signal in the atmosphere by GPS.*

9. 6328/10 The 8% number is meaningless without information about the variable in question and the typical uncertainty of a given measurement..

- *The abstract has been corrected including that this is referred to the comparison of the integrated water vapor.*

10. 6328/12 How can one instrument be considered to have the "highest redundancy?" Doesn't redundancy depend on the existence of at least two instruments?

- *The abstract has been modified including the sentence "Comparisons of time series of IWV content from ground-based remote sensing instruments with in situ soundings showed that microwave radiometers have the highest redundancy with the IWV time series measured by radiosondes and therefore the highest potential to reduce the random uncertainty of the radiosondes time series.".*

11. 6328/16-15 The fact that data from one instrument are used to "condition" data from another seems more of a problem than a benefit. The abstract refers to the "best reduction of random uncertainty" but that feels a bit like cheating to me.

- *The use of conditional probability is not "cheating". It is the way typically used in the retrievals proposed by several communities to integrate measurements from several sensors to improve the accuracy in the estimation of atmospheric parameters. These retrievals are based on the Bayes' theorem. Examples are Lohenert et al., 2004, Simpson et al., 1999; Hewison et al., 2006. Bayes' theorem is largely used in modeling and measurements communities. A paper largely describing Bayes' theorem is Rodgers (2000). Of course, when conditioning a measurement with another, care should be taken that each of them is taken at the best of its possibility and with an extensive characterization of the uncertainty budget.*

12. 6328/22-23 This first assertion in the Introduction is highly debatable. Understanding processes can be advanced through both theoretical and observational approaches, and among observational approaches it has long been the case that field experiments tend to favor a suite of measurements of different, related parameters, not a redundant set of measurements of one or more parameters.

- *In the new version of the manuscript, the sentence has been modified in the following way: "The use of redundant measurements is considered the best approach to reduce uncertainty of an atmospheric variable."*

13. 6329/3-4 This sentence, if pulled out from the manuscript, could be criticized as ridiculous. In common language, if something is redundant, it is

probably not needed at all, because the need is already being met in some other way. So it provides no added value and would be considered instead a waste of resources.

- *To avoid misinterpretations, the sentence is now "Without doubt, redundant measurements provide added value and the advantages are related to …….".*

14. 6329/10-11 I don't understand this bullet

- *Actually this bullet is somehow confusing and it has been removed.*

15. 6329/19 Is the Thorne et al. 2013 a citable, peer-reviewed, easily accessed reference? If not, use another reference that gives an overview of GRUAN, for example Seidel et al. (2008, BAMS).

- *It is citable. However the authors added also Seidel et al. that is more complete and appropriate.*

16. 6329/20 The use of the adverb "soon" to describe the network expansion is both vague and optimistic.

- *Replaced with "….. aimed at supporting 30-40 stations.".*

17. 6329/24 Avoid using a web-based brochure as a reference.

- *Ok, now there is a reference for this (Seidel et al., 2009).*

18. 6330/7-8 and 6331/15-18 These two sentences seem somewhat contradictory, offering different views of ways in which measurements are compared. Consider consolidating these statements as part of a more general discussion of traditional, parametric methods of approaching the task of comparing datasets.

- *The authors do not feel that the two sentences are contradictory. Anyhow, the comparison of the two Taylor diagrams included in section 3 (see comment #2), should also address this issue and allow the reader to have a much clearer idea of the differences between linear and non linear metrics, of their meaning and their different potentials.*

19. 6330/13 Is "This study" the present study or the one just mentioned by Fasso et al?

- *Replaced by "the presented study"*

20. 6330/18 Should “correlate the value with” be changed to “relate the value to”? The latter suggests a description approach, while the former suggests a quantitative one.

- *Since the concept behind the sentence is quantitative, the authors have decided to keep the original sentence.*

21. 6331/5 None of the five sites have been certified as GRUAN sites, and it seems unlikely that at least one of them will be. Consider using language such as “candidate GRUAN sites” or “sites currently affiliated with GRUAN (but not yet certified)”.

- *Ok.*

22. 6331/20 I don’t understand the meaning of “of the freedom in selection of an event”.

- *The authors removed this sentence though typically reported on statistical books to avoid misunderstandings.*

23. 6332/5 What is the antecedent of “these”? Is it H and sigma? Would it be clearer to say “they”?

- *Ok.*

24. 6332/17-18 Consider saying more directly that MC is a more general measure than rho, because it does not assume linear or even monotonic correlation.

- *Ok, the concept has been clarified in the new version of the manuscript according to the reviewer’s suggestions.*

25. 6332/25 Should “information” be replaced with “correlation”?

- *Yes, though in information theory this is practically equivalent.*

26. 6333/18 Either state the triangle equality or remove this mention of it, but do not assume the reader is familiar with it or that the connection to D will be obvious.

- *Triangle equality is now explained in the text as “given X, Y, Z, the sum of D of any two of the considered variables must be greater than or equal to the value of D for the remaining variable”*

27. 6338/20 This sentence is confusing. Why bring up variance and correlation in this discussion of MC? Consider keeping all the discussion of the

advantages of MC over more conventional, parametric methods in the beginning of this section?

- *The authors changed the section according to the reviewer's comment.*

28. 6334/3 What is the axiom of information theory?

- *The authors removed this concept though not needed there and largely described in the reported reference at the end of the paragraph.*

29. 6334/9 Be clear that you are addressing water vapor observations only, not other parameters measured by radiosondes or other instruments.

- *The authors think that this is well described at line 12 ". This study focused on the investigation of atmospheric water vapor measurements, both profiling and columnar".*

30. 6334/13 At least some of the instruments do not sample the "complete column". Their vertical ranges are limited. This should be stated explicitly, and described quantitatively, because it is a source of non-redundancy of the measurements.

- *This is mainly the case for the Raman lidar, though this is well discussed in section 2.2.*

31. 6334/21-23 I don't think this statement is true. Other GRUAN data (e.g., from Lindenberg) are flowing into the GRUAN archive. Maybe I'm missing the point here.

- *Unless the authors are not aware of a different location for the GRUAN data, no different data than radiosondes are available yet on the NCDC GRUAN archive (checked on 11 August 2014).*

32. 6335/4 What is meant by "passive" instruments. Aren't radiosondes also passive, in that they don't send signals out as part of their measurement method?

- *In remote sensing, passive is used for those instrument that do not make use of sources but use the sun as a source, looking mainly at the process of absorption and scattering of the solar radiation occurring in the atmosphere. Radiosondes are usually considered in-situ sensors.*

33. 6335/27-28 This last sentence seems unconnected to the rest of the paragraph.

- *The authors changed the sentence as follows: "This was done to suppress the bias component of the time series uncertainty and to ensure that the reported entropies are related only to the random uncertainty.".*

34. 6336/9 What do you mean by "selected by stations"?

- *The sentence was confusing, so the authors removed it and clarified this concept at the beginning of the section 2.2*

35. 6336/23 Considering replacing "Starting at 25 bins" with "Between 25 and 100 bins".

- Ok.

36. 6337/9 Does solar radiation affect humidity observations (or only temperature)? If so, specify which instruments suffer from this source of bias.

- *In the text of the manuscript, the authors meant to describe the radiosonde humidity sensor's sensitivity to solar radiation heating. Other instruments are affected by the solar radiation but in a different way, e.g. the vertical range covered during daytime by a lidar is limited by the solar background that strongly decreases the achievable signal to noise ratio. But this limit is not relevant for the investigation of the measurements reported in the paper that only referred to nighttime conditions.*

37. 6338/18-26 This discussion of D raises some questions that should be clarified here. Does the value of D always range from 0 to 1? Is "redundancy" a function of D, and D alone, as suggested. What value of D (or what other quantitative measure) is typically used, or is used in this paper, to judge that techniques "show good redundancy"?

- *This is a not conventional discussion because this is, as far as we know, the first use of information theory concepts to deal with ground based measurements. This is the reason why the authors introduced section 3.5 where they provide a criteria with the value of D below which measurements are considered redundant. This is strongly dependent on the maximum measurement uncertainty required for a certain application.*

38. 6339/19 Change "entropies retrieved" to "entropies computed" or "entropies estimated".

- *Ok.*

39. 6339/23 Insert "only" before "20".

- *Ok.*

40. 6339/29 The entropy values don't seem so very similar to me, particularly near the ground. A qualitative term like "similar" should either be avoided or supported with quantitative results. See also 6340/14.

- *The authors thanks the reviewer for this comment and, in the new version of manuscript, the discrepancies among the entropy profile are described in a more quantitative way.*

41. 6340/23 Consider changing "reported" to "shown" or "illustrated", since you are discussing a figure.

- *Ok.*

42. 6341/24 Should "normalized over" be changed to "normalized by"?

- *Right, "normalized by".*

43. 6341/29 I'm not sure it is fair to say that MC is "more accurate" than linear mutual correlation (LMC). They are different, and MC may be more appropriate and more general, but both measures accurately measure what they are intended to measure.

- *Ok, the sentence is now "This example supports the use of MC as a more general concept than the LMC for quantifying the value of redundant measurements".*

44. 6342/4 Explain Taylor's diagrams.

- *The authors reported in the text the following sentence "Taylor diagrams provide a concise statistical summary of the similarity between two patterns, quantified in terms of their correlation, their centered root-mean-square difference and the amplitude of their variations (represented by their standard deviations). These diagrams are especially useful in evaluating multiple aspects of complex models or in gauging the relative skill of many different models or measurement techniques.". This has been put in the section 3 where also two Taylor diagrams are commented (shown in Figure 3).*