This is our reply to Anonymous referee #1

First we thank Anonymous Referee #1 for an extensive review of our manuscript.

As the referees comments were not numbered we will quote each of the referees comments and then give our response.

> "The manuscript discusses a very significant and too often neglected topic about cross- calibration of optical instruments used by many different institutes and universities at different locations in several countries. This type of collaboration not only results in improved data quality but also promotes good instrumentation techniques. The journey from raw data to a first-rate data product for geo-science studies is far from trivial. I strongly encourage the authors to continue on their path. Publications of this kind of work is highly recommended."

Authors response 1: We could not agree more.

> "1) I found the treatment of light standards and their tracing to international standards too ambiguous. In theory, any light source that can be traced to a known standard can be compared (intercalibrated) to another light source that is also traced to the same known standard. The concept of national and international standards then makes global comparisons possible (with error estimates)."

Authors response 2: We will attempt to clarify this in this response as well as in the revised manuscript.

This manuscript reports results of intercalibration of calibration sources for aurora and airglow from 2011 according to a method that was internationally agreed upon in the early 1980's. The method (and the associated equipment) has remained more or less unchanged since 1984 and seem to have produced acceptable results during the years.

Much of the critics from this referee is directed against the intercalibration method itself. We do not consider it responsible to suddenly, and without prior international consensus, change this method nor the equipment since this concerns long-term service to the community in the form of intercalibration of low-light calibration sources for aurora and airglow.

Yet we do acknowledge that extensions and improvements to this method will be required in a near future. To do this requires a thorough evaluation of the existing method and equipment. This manuscript is the first step in this process.

Together with the intercalibration equipment we have received an extensive set of documentation (page 94 lines 16–18). Since much of what is published concerning these matters is not very accessible today, we also attempt to make as much as possible of this documentation available to the scientific community over Internet. However, it will take a significant amount of time to go through and scan (digitize) this documentation.

This manuscript is written under two constraints: time and tradition. Time, since the results of the intercalibration workshops 2011 ought to be published before the next one. Tradition, since this is a long-term intercalibration effort spanning over 30 years or more. Therefore continuity must be preserved by not suddenly changing neither method nor equipment.

> "In addition, I was not convinced about the traceability of the prime light source (Fritz-Peak) as the text only mentions its calibrations in 1960's and late 1970's!"

Authors response 3: This manuscript concerns the series of intercalibration workshops starting in 1980. Thus traceability before 1980 is interesting and strongly desired, however, not in any way crucial for the objectives of this manuscript.

Several sources are traceable before 1980. For example calibrations of the Fritz Peak source from 1967, 1968, 1969, 1970 and 1972 appear in Torr et al. (1983) in the form of a diagram.

We hope to prepare and publish such results before 1980, but it cannot be done in the timeframe of this manuscript. There are three major reasons for this: 1) other equipment was used, in particular, different filters. 2) Many of these results only exist in the form of diagrams. Some of these are also in bad condition as they are photocopied several times (as in the case of the above reference). Retrieving the originals is difficult or impossible 3) Source naming conventions has changed over time and time consuming research is required to uniquely identify the sources.

> "2) The novelty in the manuscript is, in my opinion, limited to collecting most if not all earlier results in a form of a report. This is of significant practical value in itself, but I feel that the full potential of a comprehensive consortium of instrumentalists in multiple institutes and nations is not reached. Perhaps, the experts could agree on a common intercalibration procedure that could be recommended for all users. One should also ensure that the most critical information from the physics point of view is captured."

Authors response 4: This is correct. To suddenly change the method or equipment of a long-term intercalibration effort, agreed upon internationally by several experts just to meet the novelty requirement of a journal would be irresponsible. Yet, the method, as well as the results needs to be documented in an accessible journal in order to encourage scientific discussion about possible future improvements.

If this can't be done, these kind of efforts will cease, and we strongly doubt that this is the intention of referee #1.

To our knowledge this intercalibration series is the only one in effect in, at least in the European community (referring to the annual European optical meetings), and probably in the world (according to several U.S. participants in these conferences over the years). In the field of absolute measurements of the

aurora and airglow it is also the longest time series of intercalibrations known to us. Thus continuity must be ensured also in the future.

In order to reach a new international consensus on improved methods and equipment, the methods and equipment currently in effect must be accessible for debate by the scientific community. If this manuscript is rejected, this will probably never happen.

> "The introduction (page 93, line 16) does not clarify at all, why absolute measurements are becoming more important. Is this due to new analysis methods or has the use of auroral or airglow data changed recently? This is the motivation for the whole intercalibration effort and should, in my opinion, be elaborated."

Authors response 5: We will elaborate on this. "increasingly important" refers to measurements made beginning with Lord Rayleighs airglow measurement in 1930. It is still true, and indeed with present day precision high-speed imaging measurements of aurora and airglow, the importance of extensive high-quality absolute calibration efforts are often neglected. (In particular, methods and equipment for calibrating these high-quality instruments are very often insufficient, with some rare exceptions. This is, however, not in the scope of this manuscript.)

> "The concept of the unit Rayleigh is, for lack of a better word, impenetrable. The references, which seem to be the "classics" in this field, also appear slightly contradicting in their use of terminology."

Authors response 6: Despite what we or the Anonymous Referee #1 thinks about the Rayleigh unit (we probably agree), intercalibration efforts has been carried out in this unit over the years. We cannot suddenly change this now.

> "In addition to this, the authors use Ångström to "avoid con- fusion (page 94, line 3) while Baker et al. 1976 (and Baker, 1974, referenced therein) make a conscious effort to convert everything into SI-units. As the use of Ångström is officially discouraged by several international organisations, I would really prefer the authors to follow Baker's example and use SI-units for wavelengths."

Authors response 7: All earlier intercalibration results are presented in R/Å. We encourage a change, but it would not be responsible by us to suddenly present the results in other units without international discussion and consensus. In particular, suddenly changing to R/nm might lead to further confusion resulting in calibration errors by the source owners.

In our manuscript we strictly adhere to the Rayleigh defined in SI-units by Baker et al. (1976) (Eq. 1)

Despite the strong recommendation to use SI units in scientific papers, many papers are still published in cgs units. The Ångström cannot possibly be "worse" than cgs units.

Also see "http://physics.nist.gov/cuu/Units/outside.html", in particular: "Table 7. Other units outside the SI that are currently accepted for use with the SI, subject to further review" and the paragraph above it.

We conclude that changing the units of intercalibration is strongly desired but must be done carefully and in an unambiguous way. We believe that it is not responsible to suddenly change units in this work.

> "This would also be a perfect opportunity to first clarify and then confirm the common understanding of the unit Rayleigh at the collaborating institutes and universities. An unambiguous reference and calibration instructions for future would be a remarkable contribution to the science community."

Authors response 8: We strongly agree regarding the referees intention. Yet, doing this in this manuscript would apart from delaying it significantly, double its length and require changes of both the title and author list. This is a different and very important subject. Many of us will be very happy to participate in such an effort, hopefully also this referee.

> "Page 94, line 5: perhaps the sentence should clarify that optical instruments for aurora and airglow are calibrated by using the concept of column emission rate?"

Authors response 9: We will do that and also include a reference to Eq. 2 that relates apparent spectral radiant sterance to a column emission rate in Rayleighs. The concept of column emission rate is best described in Hunten (1956).

> "Page 94, line 9: is there a reference for Michael Gadsden? This is certainly not common knowledge."

Authors response 10: There is no reference for Michael Gadsden, however his work is mentioned in Torr (1983). We will insert this reference We want to mention him explicitly as he was pioneering this work in the 1960's. (For reference: Michael Gadsden was Secretary General of IAGA in the 1980s, received the IAGA long-service medal 1988 and sadly passed away in 2003.)

> "Table 1: I propose that the authors would label the calibration workshop in Kiruna and Sodankylä simply 2011a and 2011b (and change the text accordingly including caption of Table 3)."

Authors response 11: We agree and will do that.

> "Does the label "Sources" refer to number of different light sources?"

Authors response 12: Yes, we will add that information to the caption.

> "The treatment of calibration sources (page 94-96) leaves some critical items wide open. The most severe question is whether the Fritz-Peak international standard can be considered valid any more. The authors refer to calibrations performed more than 20 years ago!"

Authors response 13: This is an important issue and it was also our first question when taking over the intercalibration equipment. This is elaborated under "Error estimation" on page 97. In this section we only describe the types of participating sources and what we know about their history.

> "No proof is provided to validate the claims of stability since 1980's. Are you really sure that you can rely on the assumed light levels?"

Authors response 14: We think that this discussion belongs under "Error estimation" on page 97. We do not claim that the Fritz Peak source is stable, we assume it. (p.97 lines 8-9). The ratios presented in Table 3 supports this assumption. If any of the participating radioactive C14 light-standards are becoming unstable with time, it would be very unlikely that they would be unstable in exactly the same way.

Furthermore: the preliminary results of the intercalibration at UNIS with totally different equipment also supports this assumption. We elaborate more on this later on in this response. It would of course have been even better if we could have brought the C14 sources to UNIS. Unfortunately this was impossible due to flight-safety regulations.

A very old typed report (1960's?) that we have not referenced in the manuscript because we only have found fractions of it so far concerns "Light calibration by C14 activated light standards from U.S. Radium Corp." (such as Y275). Applying this information to the IRF-UJO-Y275 yields 262.65 R/Å at 5600 Å. This is to be compared to 2011b calibration that gave 261 R/Å at 5573 Å. This is preliminary, to be confirmed but gives some indication on the stability of C14 sources. The same report states that the accuracy of these sources are around 10% according to their manufacturer. This is also in agreement with our findings. We hope to manage to dig out the full report, so it can be referenced and used but this cannot be done in the time frame of this manuscript.

> "If there is no proof available such as recent and actual measurements, you have only obtained a relative calibration of light sources."

Authors response 15: This mainly is a relative calibration (inter calibration) of participating sources against each other including and using the Fritz Peak international source as reference. If the assumptions about the Fritz Peak source holds, it is also an absolute calibration. As discussed above and later on, the Fritz Peak source is still probably stable enough within at least 10% or so (refer also to Table 3).

> "Traceability to a valid national or international standard is to be confirmed. An intercalibration without an absolute reference is still a highly desirable result, but the authors should address the traceability to an international standard with high priority."

Authors response 16: To repeat the traceability measurement of Torr and Espy (1981) is strongly desired. This can't be done as long as no suitable calibration facility exists within convenient distance for ground transport on the Scandinavian or European mainland.

Yet, we do have carried out validation measurements linking the Fritz-Peak source directly to the FMI integrating sphere in Sodankylä, Finland, as well as indirectly both to the calibration facility at UNIS, Svalbard and to calibrations of one FMI MIRACLE EMCCD-imager carried out by T.S. Trondsen, Keo Scientific in Canada. This is described in the manuscript page. 97-99 starting at line 23. The results of these validation efforts produced large amounts of data that requires careful analysis and will be published later.

The preliminary results of the intercalibration at UNIS (Table 4) indicates agreement within 20% for relevant wavelengths.

> "Has any of the source owners used their respective national standards to gauge the light output of their equipment?"

Authors response 17: No. Inquires has been made to National Bureaus of Standards (NBS) in Scandinavian countries, but this has not been possible for various reasons. However, see the previous response regarding other validation attempts. Both the certified Tungsten lamp at UNIS, Svalbard and Tungsten lamps of the FMI integrating sphere in Sodankylä are traceable to NBS sources.

> "If this is done properly at each institute and university, you should be able to predict the outcome of the intercalibration effort. This would then provide a very useful validation point in addition to recording historical metadata that will assist the scientists in using the final data products."

Authors response 18: We agree completely, but this has not been possible yet.

> "It would be of great benefit for readers to provide the measured (relative) spectra of all sources."

Authors response 19: We again agree (we state this on page 98 lines 11–13). This was done at UNIS, Svalbard for three sources (including their certified lamp). These spectras will be published in the official report of that intercalibration effort. For the other sources this was not possible due to lack of suitable instrumentation on the mainland. We hope to remedy this soon (hopefully already during the next intercalibration workshop), but not in the scope of this manuscript.

> "The spectral response and quantum efficiency of the reference photometer should also be provided. In my opinion, this would help in interpreting and would also support discussing the results, especially in the case of the LED-based sources."

Authors response 20: We agree. Copies of data sheets with this information exists in the documentation of the intercalibration photometer. To prepare these plots for publication is not possible in a reasonable amount of time. It is our intention to scan these documents and make them available over Internet. Quantum efficiency drops off towards the red part of the spectra, this is also seen in the signal to noise ratio of the measurements and explains why filter position 8 (6707 Å) is not used. This is mentioned on page 97 lines 18–22.

> "Page 95, line 22: In my opinion, regular calibration of LED sources can and do provide long-term stability."

Authors response 21: We state only that "Not much is yet known about the long-term stability of the LED sources." Intercalibration data for the same LED source does not appear twice in recent intercalibration workshops before 2011. Other types of sources have intercalibration records of 10–20 years (or more).

> "For a commercial example, the Instrument Systems have a product (ACS-530 Calibration LED) that can be traced to national standards."

Authors response 22: This does not prove that it is stable for 20 years or more. Only time can tell the truth.

> "Obviously, one needs established practices for regular calibrations, which is something that should be considered in any case for absolute measurements."

Authors response 23: We do not see the relevance of this comment. The procedure described in this manuscript is the established practice for intercalibrations by the European optical community. It is based on similar efforts in the U.S. by M. Torr according to Lauche and Barke (1985) This is a long-term commitment and the methods should not be suddenly changed without international consensus. Continuity must be preserved! As discussed earlier we agree that new practices are needed, but until these are agreed upon, should we really discard all previous efforts and experiences?

> "This approach also makes it possible to replace worn-out equipment as the new light source can still be traced to the same international reference standard."

Authors response 24: Does this imply that this referee want the scientific community to hand over the responsibility of intercalibration efforts to commercial companies?

> "A schematic overview would greatly clarify the intercalibration setup details. Perhaps, this could be combined with the photo of the calibration photometer (Fig 1)."

Authors response 25: We will add a schematic and a block-diagram.

> "What is the purpose of a multimeter (pag 96, line 13), if you are already using a frequency counter for recording the PMT-counts?"

Authors response 26: The multimeter indicates the filter position to the person outside the darkroom. This is used to confirm that the person inside the darkroom selected the right filter.

> "The details of the filter bandwidths were hidden in Fig 2. One can assume that you are using bandpass filters with a FWHM from 1.5nm to 4.1nm. No transmittance curves were provided: as we are discussing comparisons, could you perhaps comment on the significance of the lack of this information?"

Authors response 27: This is correct. The intercalibration photometer documentation contains detailed information about the filters. Providing transmittance curves would add eight more plots to the manuscript and require analog chart-recorder plots to be digitized. This cannot be done in a reasonable amount of time. As stated earlier we hope to make all this information available as soon as time permits.

> "Also, how does the quantum efficiency of the PMT vary in this wavelength range?"

Authors response 28: This comment is repeated and already answered above, see Author response #20.

> "I find that recording only three samples for each filter position is not sufficient for estimating the precision. Obviously, this issue cannot be rectified for this manuscript in a reasonable timeframe, but I strongly suggest the authors to increase the number of samples in future."

Authors response 29: A precision frequency counter was used, setting a long gate time (5s). Three such measurements (samples) are averaged. Maybe we should use the word "measurements" instead of samples? This is as prescribed by the intercalibration procedure and has produced acceptable results before. Raw data is available on Internet.

For example: The dark-current corresponds to a typical count-rate of around 20 Hz. Three such measurements corresponds to an effective gate time of 15 s. This corresponds to 300 counts. We cannot see that this is "not sufficient"

> "For this manuscript, I would like to see a recommended number of samples with scientific justification based on desired error margins, confidence level etc."

Authors response 30: This method has been in effect since 1984. Nothing in previous measurements or our raw data indicates that three samples are insufficient. The raw data is on Internet. Starting in 2011 the instrument readings are input to a spreadsheet that immediately calculates the average and standard deviation of the samples as well as the final result. In this way errors caused by the tedious nature of this manual procedure will quickly be discovered. To automate this is strongly desired.

> "The filter position 8 is "traditionally discarded": does increasing the number of samples still produce a poor signal-to-noise ratio?"

Authors response 31: Yes. A low count-rate is a low count-rate no matter how many times it is sampled. Most likely this is related to a combination of the following explanations: 1) too low quantum-efficiency of the PMT for this wavelength, 2) low output, and also lack of absolute calibration data for this wavelength, 3) A broken filter? The only way to find a definitive answer is to take the calibration photometer apart and check the components. Both we and our predecessors have been strongly discouraged to do this without a backup instrument.

> "How is the intercalibration result calculated (page 96, line 26)?"

Authors response 32: This is rather trivial but we will add the equation to the revised manuscript. This is also easily available on Internet, see web-page at page 97, line 5–6, look in the spreadsheets..

> "Is this procedure something that all authors could consider the "standard procedure" for intercalibrating their optical instruments in future?"

Authors response 33: Again: This is a standard procedure agreed upon in 1984 for intercalibrating calibration light-sources. We strictly adhere to this method until an agreement regarding a new method is reached. This is a long-term service to the community, that is now inherited for the second time. The first thing to do is not to change it. The calibration light sources are then used in the calibration procedures of the respective source-owners instruments.

This procedure is of course not suitable for calibration of other optical instruments. In particular imagers requires extensive and complicated calibration procedures as compared to source intercalibration.

> "Page 97, line 5:" I find it recommendable that the authors provide the raw data for the science community. However, I would like to see more details about what they consider "unreliable data", which was removed for this study."

Authors response 34: We refer to the empty slots in Table 2. Data with lower spectral column emission ratio than 1 R/Å was discarded. Looking at the raw-data it is easily seen why (poor SNR).

> "This is related to my question about number of recorded samples for each filter."

Authors response 35: This has been tried and for any reasonable number of measurements, no improvement is seen as compared to the precribed average of three measurements.

> "In my opinion, you should include numeric error estimates. Very likely, it is sufficient to provide a single figure or an estimated error range that is common to all measurents. This would then support the discussion in the conclusions."

Authors response 36: Traditionally the ratios of Table 3 have been the provided error estimations. With a few exceptions the errors are less then about +/- 10% for wavelengths where the sources have a significant output. Even where the sources have poor output typically errors are typically less than +/- 20%. This is clearly acceptable error levels for absolute calibrations. We will update the revised manuscript to reflect this. If the referee knows about intercalibrations spanning well over 20 years with lower errors, we are happy to receive this information.

> "Figure 3: In my opinion, if the spectra of all sources could be provided, the data in this figure could be presented to show the long-term variation i.e. the year in X-axis (= data from Table 3)."

Authors response 37: We will consider to add such plots, at least for some selected wavelengths.

> "As the corresponding author kindly offers the archived data to the scientific community, it should also be possible to provide the error estimates for each sample point."

Authors response 38: As discussed above, the error estimation for this series of intercalibrations has been in the form of ratios to earlier measurements. Other error information are not available in the documentation of earlier calibration workshops. The independent validation efforts described on page 97–98 and in this response preliminary suggests that any systematic errors are most likely less than about 20%. Furthermore, the light-standards themselves are stated to have an accuracy of about 10% according to their manufacturers.

> "It is interesting to note that, e.g., the IRF-Lauche-lamp ratios to earlier intercalibrations has first increased and then decreased. Can you offer any reasons?"

Authors response 39: These variations are most likely random fluctuations due to various measurement errors. They are mostly within 10% which is the stated accuracy of the C14 light-standards. When comparing to the 2011a workshop one should bear in mind that this was the first intercalibration carried

out by the new intercalibration team. Our overall impression is that the errors are relatively small over such a long time-span given the old and manual nature of the measurement procedure.

> "Table 3: the layout and especially the label of the second column are not clear. Please re-format."

Authors response 40: We agree and will fix this error.

> "Also, I'm not sure whether it would be more useful to choose one of the lamps (which should be the one that can be most accurately traced to the international one) as the reference point (i.e. ratio 1 in case of workshop 2011b)."

Authors response 41: We are somewhat confused about this comment, please clarify. The reference point is 2011b. Ratios are to previous measurements (e.g. 2011a, 2007, 2006, 2001 and so on). In addition a ratio of 2011b to the mean value of all measurements is calculated. There is only one Tungsten lamp in Table 3, the other sources are radioactive C14 sources.

> "As mentioned earlier, the measure spectra of all sources should be provided."

Authors response 42: See author response #19.

> "Concerning the PGI Chernouss-38AM source: do I really interpret it correctly that there is no power (current) regulation for the LED system when used with batteries? This simply sounds unbelievable from electrical engineering point of view, but if there really is a difference with a regulated power source, then I strongly suggest revising the electronic control of the LED source and making sure there is sufficient regulation also when battery operated."

Authors response 43: Of course there is some form of current regulation. There is however, no indication if the batteries are good enough. The measured spectra of this source also displays two sharp peaks as would be expected from a source with multiple LEDs. As we state this is under investigation and it will take at least one more intercalibration workshop to get a definitive answer.

In general we do not dismantle or pass any judgment on participating sources. The sources are the responsibilities of their respective owners.