Response letter

We want to thank the reviewers for their efforts they put in the review and for the many suggestions. We tried to account for many of the points which are listed below in the text. One tricky part was the error discussion. We limited ourselves to the error sources of Polly and to those we can actually specify. The general error treatment which are needed for basically every lidar we cannot solve. We also think a more generous error discussion on all the different Polly systems would be needed. But it would take up an entire new publication (including error propagation formalisms, separation of statistical and systematic errors, Monte-Carlo approaches, and known uncertainties for all channels and systems). However this comment will be considered for the future. In general, we included more technical information and a subsection on the quality assurance methods that we often apply.

Review of amtd-8-7737-2015
EARLINET Raman Lidar PollyXT: the neXT generation by R. Engelmann et al.

The paper describes the history and setup of a series of apparently very similar lidar systems called Polly and recent technical improvements of the now called PollyXT. The paper is appropriate for publishing in AMT and will be very important as a reference for future publications of measurement results with these systems. It is well organized and well written. But in the view of this importance and in the hope that many valuable investigations will be performed and published with the help of these instruments, it is unacceptable that many details are missing (described below) which are essential to understand and to win confidence in the performance of these lidar systems.
Note 1: Because this paper will probably be used as a reference for future papers of the different groups running the systems and maybe as a technical reference by persons which newly start to work with the systems, it is very important to clearly indicate which of the individual improvements apply to which of the systems listed in Table 1.
Thank you for this comment. Yes, the paper should be thought to be a reference for the system, especially for new people who work with the system and for outsiders. We used your very helpful comments to work on this topic. However, we have to find a good compromise between a paper that introduces the system and the new ideas to others and a purely technical documentation. In general, the persons who work with the systems also have to familiarize themselves with the systems in other ways than only with this publication, e.g., they have a manual with more special characterizations and they usually have been present during the assembly of the systems. Also the systems often undergo small changes, so that by some time more special calibrations would be most likely outdated.
The presentation of a case study at the end is dispensable for this important technical work. Instead "calibration" measurements should be presented, which support the range and accuracy claims in such a way that they don't have to be repeated in follow-ups with more detailed analyses of atmospheric measurements than in the case study.
We think the measurement example is a good and common way to show what the typical products of such a system are and what the system can be used for. It helps to get an impression on the many parameters that can be derived. Therefore we would like to keep the example. We included a quality assurance section (2.4) were we show an example of a Telecover Test. Other procedures, more related to signal evaluation will be given in Baars et al. (2016). Please note that we cannot show these measurements for every single one of the instruments here but moreover only show what methods are used to ensure the data quality of Polly. Nevertheless, at the moment the idea of a follow-up paper purely on error determination and measurement accuracy with the different Polly systems emerged. Many aspects on calibration are also not only related to Polly but more general to all lidar systems.

Details:
Title: The title makes it seem like PollyXT is or will be the standard lidar system for EARLINET.
If this is not intended, EARLINET should not be mentioned in the title.

Ok, this is maybe correct. Or intention was to acknowledge the EARLINET consortium for all the fruitful collaborations. Without them Polly wouldn't exist in its current state. We changed the title to a more technical one ("*The Automated Multiwavelength Raman, Polarization, and Water-Vapor Lidar Polly-XT: The NeXT Generation*") and keep the acknowledgement to the text.

P7742 L20 *The alignment of the laser beam through the external second and third harmonic generators (SHG, THG, E2 in Fig.) was improved …*

Which effect does the improvement have on the lidar signals? If there was no effect, would it be worth mentioning it?

There is no direct improvement (for a correctly aligned Polly system) on the signal quality. However, it brings benefits to the (sometimes not so experienced) operator. We added: "...*compared to the original setup. By this upgrade it is easier to correctly realign the beam after a maintenance procedure or after an exchange of the entire laser*."

P7743 L10 *The laser-beam overlap with the receiving telescope is monitored with a triggered camera (CAM)...*

What does triggered mean?

How is the *overlap* determined by the CAM?

What is the resulting alignment accuracy?

Triggered meant synchronized to the laser trigger. The overlap function itself is not determined by the camera. But the beam pointing accuracy can be traced and monitored. We changed this sentence to: "*The stability of the laser-beam overlap with the receiving telescope is monitored with a camera (CAM) that is synchronized to the laser trigger and observes the transmitted beam at 532\,nm. It can be estimated that a laser-beam alignment error of +/-0.1 mrad can be detected in this way.*"

P7743 L21 (compare P7744 L23) *Small PMT modules of Hamamatsu (type: P10721-110) are used for all wavelengths,...*

Is it H10721-110 or P10721-110 ?

Sorry this was a mistake. The modules are called H10721P-100.

P7744 L7 *In contrast to the former systems...*

Which are the former systems? (see Note 1 above)

We added: *In contrast to the former systems (generation 1 and 2) which were described in Althausen 2009 …*

P7744 L7 *For example, the deformation of the mounting of the primary mirror is depicted in Fig. 4.*

The deformation in Fig. 4 looks rather like the one when the mirror is looking horizontally.

Sorry if the explanation was not clear enough. It is not the deformation of the glass mirror itself, but the deformation of the mirror mount. The deformation of the mount is caused by the weight of the mirror and by the weight of the mount itself. In this way, you can see the bending down of the mirror mount (as well as the inward bending of the 45° stiffeners). And from these results you can estimate the angular shift of the field of view after the installation. The mount moves by max 3 µm downwards at the top by gravity. The mount is ~320 mm wide. That accounts for an angle of ~0.01 mrad. Since during the reflection of the light the angles double we end up with a max. beam shift of 0.02 mrad. Again, this figure is just one demonstrative example on how we can use the finite-element analysis to estimate the overall stability.

P7744 L19 *A new photon-counting data acquisition hardware developed at the Max Plank Institute for Meteorology in Hamburg has been applied to all recent Polly systems.*

What means *recent Polly systems* (see Note 1 above)?

We decided to call the newest Version of Polly generation 3 (new DAQ and optical layout as shown in Figure 3. Generation 1 and 2 where described in Althausen 2009. The DAQ is also included in Table. 1.

P7744 L21 *The aim of the development was a small and low-cost data acquisition for scientific lidar measurements with high dynamic range.*

What means high dynamic range?

We added: „*The aim was to develop a small and low-cost data acquisition for lidar measurements with a signal dynamic range that is only determined by the detectors themselves. Minimum single-photon pulse lengths of are on the order of 2 ns with our PMTs. Thus the maximum count rates that*

need to be handled by the electronic design had to be above 500 MHz in order not to increase the overall dead-time effects significantly."
P7744 L26ff
Please make the paragraph better understandable for the readers which are not experts in electronics:
What means flip-flop circuit? This is a standard electronic element which can alter between two logic states. But this might be to general to be explained in the text. We added: *"a flip-flop circuit which alters its logic output level every time a photon was detected."*
What is a level transition? Changed to *"voltage-level transition".* It simply means a High to Low or Low to High transition of the digital signal.
What is the histogram? We deleted the term "histogram counter" and replaced it be "time-resolving counter" and hope this is more understandable now.
Why should the pulse discrimination level be controlled, and what is the achieved improvement or accuracy? There is no accuracy improvement if you are able to set the discrimination voltage. This is standard for every photon counter, isn't it?
What does it mean that the power supply is controlled? We meant: to be able to switch "on" or "off". We changed the text: *"Additionally, a micro controller (µC) is implemented to adjust the pulse-discrimination level and the PMT voltage, to switch the power supply, as well as to control an LED (light-emitting diode) which can be used for various test procedures of the PMT."*
What does *handling eight channels* mean?
We modified it to: *"A single counting unit is capable to control and to acquire the data of eight detection channels..."*
P7744 L26 *An optical trigger taken from laser stray light starts the histogram counter which is synthesized in an FPGA.*
What is a histogram counter and what means synthesized?
We now use "time-resolving counter". Synthesizing is the terminology of FPGA programming. In fact an FPGA does not contain a program, but just a logic circuit. In this sense it is called "synthesizing". For the sake of easier understanding we changed the term to *"...is implemented in an FPGA."*
How are the lidar signals retrieved from the histograms?
The data in a histogram, now called time-resolving counter, is the lidar signal. Maybe the term histogram was confusing as it suggests some sort of distribution…?
Does the averaging *over a predefined time period* mean the range bin or the temporal averaging of the lidar signal? It means averaged over a time period (e.g. over 30 sec⇔600 shots). Range-bin averaging is done during data evaluation. We changed it to: "…where the profiles are averaged over a predefined time period of typically 30 s."
The descriptions of the electronics are not probably not understandable by the general reader, and if not expanded, they could be skipped. However, it would be of interest which effects or improvements for the lidar signals are achieved by each of the devices compared to "usual" systems?
Our previous P7882 DAQs had several issues. First, we needed to apply preamplifiers which lengthen the photon pulse unnecessarily. Secondly, the BNC cables from the detectors to the DAQ cards in the PC were 2 m long running somewhere close to the laser and receiving signal distortions. The true cause is unclear until today, but in the first hundreds of meters of the lidar signal distortions could be observed (maybe preamplifiers, discrimination voltage reference voltage, a grounding loop, electromagnetic interference …) which depended also on the discrimination level.
Third, only maximum of 8 channels (4 P7882 cards) were possible to use simultaneously in one PC.

With the *"CNT80"* we now don't use preamplifiers and have only a few cm of shielded BNC cable with "analog" photon-counting signals. Discrimination is done right at the detector at a speed (~800 MHz) which no commercial PMT module with included discriminator circuit offers. Then all photons are already digitalized. Up to now all measured signals with the CNT80 system fit the ideal photon-counting theory (only dead-time distortion, pre-trigger signal equals high altitude signals, no observable afterpulsing).

So in fact, the only improvements are that the DAQ system works as it is supposed to work. Also having a pretrigger option on the photon counter helps to identify the zero bin better.
Another main improvement are the costs and expandability to more than 8 channels.
Neat things as only a single supply and data cable and the test LED are only beneficial in terms of comfortable operation.

We added some drawbacks of the old system at the beginning of the section and added a few details to Table 3.

P7745 L20 *When using photon counting dead-time effects can be an issue...*
Dead time effects are always an issue.
Well, if you operate the system with count rates of < 10 Mcps the dead-time effects are negligible and are not an issue which needs to be taken care of (e.g. if the dead-time error < Poisson error). But you are right, we changed the sentence to: "*When using photon counting dead-time effects need to be considered if the data acquisition operates in high count-rate regimes, e.g. for our system above 10 Mcps.*"
P7745 L23 What does *electronic transients at the discrimination level* mean?
We meant the photon signal "crossing the discrimination level". Now: "*…and will finally result in a lack of voltage-level crossings at the discrimination level*."
P7745 L25 *Measurements showed that the typical dead times with respect to the paralyzable theory are on the order of 2–3 ns (depending on the individual PMT)*
Do you mean with *individual PMT* only individual *H/P10721-110* or different types?
We meant our H10721P-100 PMT modules (together with our data acquisition) showed dead times of 2-4 ns (the value was changed now). Every individual module was a little different, but in general these values are a good estimate. But of course, different PMT types can show arbitrary different dead times.
We modified to: *"(depending on the individual H10721P-110 modules)". Also we added the statistical values of measured dead times from 32 PMT modules (3.2+/-0.6 ns, paralyzable).*
Fig. 6: Caption: *a defined light ramp …*
Why does the light ramp have to be defined?
The comparison is between the measurements with a strong and a weaker neutral density filter. The strong filter pulse is the reference, howsoever the real pulse shape is.
We meant a light ramp that is repeated and always of the same shape. We removed the word "defined" and think it should be still clear what is meant.
Fig. 6: Inset: I am astonished about the time scale in ns. If this was true, the temporal resolution of the DAQ would be in the sub-ns range.
Thanks for this hint. Of course, that was µs. We changed it.
Fig. 6: The figure is not easy to understand intuitively, because two things are combined, i.e. the measurements (black line values over blue line values of the insert) and the correction function, and because the ordinate label does not mention the low light measurement (black line; factor ~100). The black identity line is also misleading, because it is intuitively interpreted as the truth, but the correction is deviating from it. According to the figure caption (*resulting correlation*), the axis label should be something like "black line Mcps (x 100) over blue line Mcps". Or the axes are exchanged and "blue line Mcps over black line Mcps (x 100)" are plotted, with and without polynomial correction, so that the corrected blue line values agree with the identity line.
Thank you. We changed the axis label and the caption text accordingly.

P7746 L10 *From this data it was found that the paralyzable theory is not always adequate for count rates higher that 40 Mcps and differences to such a model occur.*
What does *not always* mean (sometimes yes, sometimes no?), and which differences occur? Doesn't paralyzable mean that the measured Mcps decreases again with the true Mcps increasing above a certain threshold? Either this regime is not considered here (what about the ambiguity then? What about clouds?), or the behavior of the system is closer to the non-paralyzable theory.

Yes, sometimes a paralyzable fit performs better, sometimes a non-paralyzable one, even for the non-ambiguity regime. The reason for this phenomenon was not further studied. But some publications also discuss a combined model [N*exp(-N t_p)/(1+N t_np)]. The study of this effect is interesting and important, but not scope of the technical paper of Polly. We just want to present the actual status and methods that we apply.

We shortly mentioned the ambiguity issue. But this effect and the need to take care to always look at raw count rates applies to all lidars, especially in clouds.

What is the corrected maximal count rate (typically)?

We did not study this yet in detail and therefore suggest to be cautious. We don't operate the systems with higher count rates than 60 Mcps (80 Mcps corrected, ~30% correction). We added: *"Although Fig.6 shows that a dead-time correction of a measured signal of 100 Mcps might be possible we try to keep the signals below 60 Mcps in order to keep the correction reasonable (max. correction factor of 1.3). Further studies might be needed to explore the full benefits from the dead-time correction scheme in terms of increasing the signal dynamic range."*

With which max. error does the polynomial fit?

That error depends on the count rate. However, we save the polynomial fit errors in our data files for later error propagation. It's also unclear up to now how the polynomial fit might influence the slope of the lidar signal. Therefore care should be taken if the polynomial correction is used for the extinction retrieval. But these investigations we would like to leave open for the data analyst, as it might be a more general and not specific to Polly.

How high is the dynamic range, which is emphasized on P7744 L21.

Well, this can be calculated straightforward. Let's assume the maximum count rate is 80 Mcps. The dark counts (can be background subtracted) are specified with 50 cps (noise = 7 cps). An average over one hour (3600s*20Hz=72000 shots) and over 150 m (20 bins = 1µs) corresponds to 72 ms of acquired photons. So at the signal maximum 5.760.000 photon events are counted. The noise photons are ~0.5. Thus the signal spans approximately 7 orders of magnitude in one hour for a range resolution of 150m (assuming no additional sky background). We added the dark counts and the max PMT count rate in Table 3.

Furthermore, it might be interesting for the readers to get statistical information about the variability among the obviously large number of PMTs investigated for the several Polly systems.

Yes, we now included: *"Measurements from 33 different PMT-modules between 0 and 60 Mcps showed that the typical dead times with respect to the paralyzable and non-paralyzable theory are 3.2+/-0.6 and 3.4+/-0.5 ns, respectively, in the range of 0 to 60 Mcps."*

P7746 L11 ... higher than ... Thank you!

P7748 L1ff How can the values below the full overlap at about 800 m (Fig. 10 at 532 nm) be obtained? Please explain or give a reference.

Can the deviations between Polly and MARTHA/Radiosonde below about 800 m be attributed to the incomplete overlap?

We added: *"A measurement close to the ground (>200 m height) is feasible because the water-vapor mixing ratio calculates from the 407 to 387 nm signal ratio and if the overlap functions of both channels are equal they cancel each other out (e.g., WandingerSpringer2005). The reason for the deviations on the order of 0.2 g/kg (16% error) of the PollyXT_IFT data below 800 m height from MARTHA and the radiosonde data remains unclear for the presented measurement. Either electronic distortions by the older P7882 data acquisition setup or different overlap functions for 387 and 407 nm might have caused the discrepancy."*

Also we now show how the Telecover Test is used to check the validity of the signal ratios in the overlap region in section 2.4.

P7748 L7 ... 0.75%. Thanks.

P7748 L13 *In contrast, the generation of second and third harmonic radiation can significantly decrease the purity of polarization and affect the accuracy of measurements.*

Please explain how this can happen?

But: isn't the opposite true, i.e. that the non-linear effect of SHG and THG is only present at a certain plane of polarization, which should result in perfect linearly polarized light - at least at 355 nm?

Yes, true. 355 nm should have the highest polarization purity. To our understanding there are several effects. Birefrigence in the SHG/THG are one reason to decrease the polarization purity. Another issue is the not-so-good alignment of the axis of the SHG/THG crystals. For the Surelite SHG/THG models that we use they are not always perpendicular to each other. No matter what truly happens our goal is to polarize all three wavelengths in the best possible way by a Glan Laser polarizer and to minimize the angle between the polarization plane and the s plane of the receiver elements.
We modified: *"In contrast, the generation of second and third harmonic radiation can significantly decrease the purity of polarization* <u>*or rotate the plane of polarization*</u> *and thus affect the accuracy of measurements. SHG and THG are both of type II in the current PollyXT setup, i.e., the radiation at 355\,nm, 532\,nm, and 1064\,nm is vertically, horizontally, and elliptically polarized, respectively, with respect to the optical board. But because of the original mounting design of the Continuum Surelite SHG and THG crystals the crystal planes are not perfectly aligned with respect to the optical board and the laser polarization planes could be rotated by a few degree. "*

P7749 L3 *Hence, the polarization impurity of the transmitted laser beam is well below 0.1 % and sufficiently low for depolarization measurements in terms of EARLINET standards.*
There are more optical components after the Glan-laser polarizer (see Fig. 3). Do they not effect the polarization purity?
How can/has that been verified?
What exactly means polarization impurity?
What does *in terms of EARLINET standards* mean here?
In fact, after submission of this manuscript it became clear that the laser beam expander might cause a degradation of polarization purity since the crystalline CaF2 lens shows stress birefringence. Further steps are the replacement of the CaF2 lens by an amorphous glass type to reduce the stress birefringence to an acceptable level. This effect is indicated from too-high linear depolarization values for Rayleigh scattering (>2% at 355nm, ~1.2% at 532 nm) while values of 0.7% are theoretically expected.
With polarization impurity we refer to any polarization states that are not linear. Either elliptical or unpolarized (a lateral mixture of different ellipticity) components.
You are right, there are no defined maximum errors in EARLINET standards. It was just a statement on typical practice. So we removed this term here.
We added: *"It was recently discovered that the beam expander (see Sec. 5.1) after the Glan-Laser polarizer can introduce further effects. Unfortunately, stress-induced birefringence was introduced during the anti-reflection coating process of the calcium fluoride lenses. The resulting birefringence pattern might again introduce circular or unpolarized laser-light components which could explain the apparent linear volume depolarization ratio for Rayleigh scattering of $>$1.5\% for the NOA and DWD systems. A replacement of the crystalline lens material by an amorphous glass type is foreseen."*

P7749 L21 ...*so that the former correction does not induce significant errors...*
and
P7750 L5 *Although the remaining bias from non-ideal beam splitters can be corrected with Eq. (2), it yields another step in the data evaluation and* <u>*can increase the overall*</u> *errors.*
Maybe I am misunderstanding something here. Corrections should remove/correct known systematic errors, not induce them.
This is true of course. But a calibrated linear depolarization profile is required in order to correct the non-ideal beam-splitter effects. And the correction values have to be known, too. And it has to be clear whether the correction was finally applied. Hence we think it is more useful to try to minimize these effects in the design phase already.
We changed to: *"so that the former correction might be even not necessary."* and *"…and can introduce new errors."*

P7750 L8 *The obtained values for the newest design of Polly are satisfactory …*
What does satisfactory mean?

Regarding *newest design*, please see Note 1 above and bring Table 4 in compliance with the text.
P7750 L8 ...*a difference...below 10%....*
10% of what?
As a rule of thumb, satisfactory means that the Ri values are close to 1 with a range of 0.85 to 1.15 (15% difference for p and s components) in order to keep the error on the backscatter coefficient below 5% (Mattis 2009). If the data analyst in the end uses the correction or not is beyond our scope for now. So we just can mention that this correction should be applied. We included all systems where the Ri values are known in Table 4 and the text was adopted accordingly.
"The obtained values for the newest design of PollyXT (DWD and NOA) are satisfactory and show a sensitivity ratio of each channel for parallel and cross-polarized light between 0.85 and 1.15 as suggested by Mattis (2009). Table 4 gives an overview on the actual values of Ri and their measurement uncertainties. For the systems with Ri values outside this margin the correction with Eq.2 should generally be considered."

P7750 L11ff The described calibration procedure has an error, which is not specified here, and which influences Eqs. 2, 3 and 4, resulting in overall errors of the lidar signal and of the linear depolarization ratio. Furthermore (P7751 L7ff), the absolute calibration also has an error. Fig. 9 shows a variability of the calibration constant C, but it is not clear whether the variability stems from random measurement errors or from variations of the calibration constant. The latter case would be an argument that the calibration has to be performed regularly and that the actual error of C can be decreased compared to the shown variability. However, the error of C together with the errors of the receiver transmission ratios contribute to the error of the linear depolarization ratio in Eq. 3.
This is correct. As for every measurement there is an uncertainty. Table 4 also specifies the uncertainties for some of the Ri values where they are known. And of course, by error propagation it is possible and also needed to calculate the errors for Eq. 2-4. These calculations also have been performed but due to the straightforward calculation and the lengthiness the error formulas are omitted in this manuscript.
We added: *"The determined R_i values also have an uncertainty which needs to be considered for error propagation of the volume linear depolarization ratio (see Eq. 3)."*
On P7747 L25 it is emphasized that the linear depolarization ratio has to be measured with high accuracy in order to enable the analyses described in that chapter.
Not only therefore it is essential that the errors mentioned above and their influence on the final products are quantified and their determination described in sufficient detail.
Measurements of depolarization with high accuracy are not easy. Several papers are being published momentarily about calibration procedures and uncertainties. These approaches might even be superior to the method shown here because all optical elements are taken into account. However, the topic of depolarization measurements with Polly could be a publication for itself. Here we want to primarily present the technical concepts that we follow to improve the depolarization measurements. We extended the description of the lamp calibration a bit.
P7753 L14 ...*the height of complete overlap is not as essential as the equality of the overlap function for the separate detection channels...*
Please explain or give references.
The sentence starts with: "…for quantities that are determined from signal ratios…". Whenever a signal ratio is used (O*P1 / O*P2) the overlap ratio is 1. We think that fact doesn't require a citation? It is only important that the overlap functions are equal. Of course, at some minimum height were the overlap function is close to zero the signal intensities become lower and the error of the ratio increases. P7753 L4 *From EARLINET work-shops in the past years it emerged that such a compromise to cover measurements in the entire troposphere with only one receiver telescope is almost impossible.*
Where is the compromise?
We mean the compromise between designing a *"fast"* overlap function for low altitudes and a *"slow"* overlap function to measure signals in the tropopause for calibration. The fixed signal dynamic range

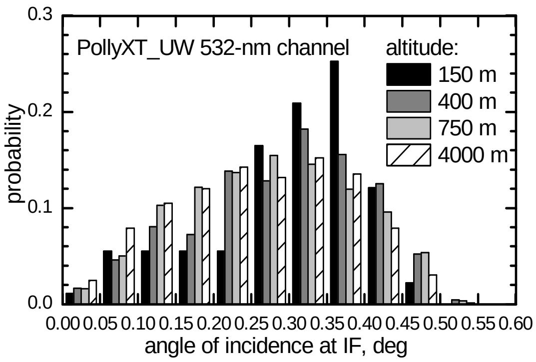
always demands such a compromise. Measurements from 50m to 20km height already demand a signal dynamic range of 5 orders of magnitude only based on geometrical considerations (not including variable backscattering and extinction).

P7753 L16 *Therefore, extreme care was taken to guarantee that the optical paths behind the field stop are similar and the spread of incident angles of the radiation is less than 1° at the interference filters.*

The first statement is just qualitative. What does *similar* mean, and which are the quantitative consequences with respect to the products of the signal ratios?

The second statement (1°) is also not related to its effects on the signal products (see comment to P7754 L10 below).

Usually the acceptance angles of narrow-bandwidth filters are on the order of 2°. This was also calculated, e.g. for RAMSES (Reichardt, et al. 2012). Their filters bandwidth is even a bit smaller (0.22 nm). We found that the incident angles for our system are below 0.6° (see graph, ZEMAX calculation by Raytracing). Thus we simply assume a height-independent transmission.



We changed the text a bit: "*Therefore, the optical paths behind the field stop were designed in such a way that the spread of incident angles of the radiation is below 0.6° at the interference filters so that a height-independent transmission is assured (e.g., Reichardt, 2012).*"

P7753 L28 The NA of the fiber should be mentioned.

The NA=0.22. It is now mentioned.

P7754 L6 *The scrambler consists of a sapphire ball lens (personal communication, I. Serikov, 2012) of 2 mm diameter...*

To what does *personal communication* refer? Do you mean *proposed by Serikov* or similar?

Yes. He proposed it during a coffee break. We changed it accordingly.

P7754 L10 *Figure 10 shows the simulated and experimentally determined overlap functions for the far-range and near-range receiving telescopes.*

How was the simulation done?

We included: "*The simulation was performed with help of the software ZEMAX. The optical geometry of the receiver was implemented in the ZEMAX model. The height-depended location and size of a laser-illuminated source plane was calculated and the overlap function was determined by Raytracing from this plane towards the detector. Special coating effects (at mirrors, lenses, and interference filters) and the effect on spatial non-homogeneous photo cathodes of the PMTs were neglected for the simulation.*"

Are the experimental overlap values derived according to Eq. 3 of Wandinger and Ansmann (2002)?

The overlap function was in this case determined by the iterative approach mentioned in Wandinger and Ansmann, 2002. This is now given in the text.

It would be nice to see in Fig. 10 a bit more than 900 m height, and is it possible to use a different colour for the far and near range curves in the overlap plot?

As the overlap function is in general a smooth function, the curves could be smoothed to better show the agreement at the higher ranges, and/or another measurement could be used with longer averaging time and lower noise.

Our experience shows that the overlap function should never be smoothed as smoothing will affect the slope of the sometimes very steep overlap function. At least if the overlap function is used to correct the data. At the altitude where the overlap function reaches 1 it is usually kept constant, not to add further noise.
The plot was extended to 1200 m and different colors were used for the measured near and far range overlap functions. The discrepancy between the simulated and measured overlap function above 600 m height is briefly discussed.

Removed due to the corrected review: ~~The basic assumption for the determination of the overlap function according to Wandinger and Ansmann (2002) is that the overlap functions for the Raman and elastic channels are the same. While for 532 nm the overlap function of the large telescope retrieved with the Wandinger and Ansmann (2002) method can be checked with the 532 nm near range telescope (it would be nice to see this in Figure 10), this is not possible for 355 nm and 1064 nm. But Figure 11 shows backscatter coefficients at 355 and 1064 nm down to about 300 m, and linear depolarization ratios at 532 and 355 nm down to about 100 m.~~
~~How is this possible with a field of view of 1 mrad for the large telescope?~~
~~According to Table 2 different interference filter bandwidths are used with 1 nm bandwidth for the elastic and 0.3 nm for the Raman channels. Furthermore, the spectral transmission of dichroic beam-splitters with steep filter edges is sensitive to the incidence angle, and also polarizing beam splitters can exhibit such an angle sensitivity. On P7753 L18 the maximal incidence angle in the parallel beam path is given as 1° for the large telescope. Is this the value for the nominal 1 mrad field of view? Estimating the distance of the telescope and laser axes with 250 mm, the backscattered laser light from 100 m range has an incidence angle at the telescope of 0.25 m / 100 m = 2.5 mrad. With the telescope magnification of Ftelescope / Fcollimator = 900 mm / 50 mm = 18 (with estimated focal length F of the collimating lens of 50 mm) we get an incidence angle of 2.5 mrad * 18 = 2.9° at the interference filters in the receiver optics. Depending on the laser alignment it could be more.~~

This shows that it is important to somehow verify the assumption of identical overlap functions for the signal ratio products – and to include more information about the receiving optics in Table 2.
We now include the information about the collimator lens pair as well as about the focusing lenses.

For example, if the overlap function is wrong by 5% and the backscatter ratio is 2, the retrieved aerosol backscatter coefficient using the overlap correction is already wrong by about 10%.
How is the overlap function for 1064 nm achieved? At this wavelength no signal ratio or Raman method can be applied.
In summary: the overlap function and the way how it is achieved should be presented for each channel for which it is actually used, together with its uncertainties. But perhaps there are other ways to verify certain signal ratio products in the overlap regime.
For all products (except extinction) signal ratios are used. These we can check with either coaligned measurements with other lidar systems or by the Telecover Test as now shown in section 2.4. This is in general also possible for the 1064 and 607 nm signal ratio in order to analyze the errors involved. The overlap function might then only be needed for the extinction calculation. And because of the uncertainties that can arise we implemented the near-range channels for the lowest 1 km height. Of course it would be preferable to also include near-range channel at 1064 nm which is open for the future. Up to now we included already two additional near-range channels at 355 and 387 nm in the OCEANET lidar.
The way the overlap function is finally used, has to be decided by the data evaluator. In fact, most often for Klett and Raman extinction retrievals we don't even perform overlap correction and are just satisfied with a minimum height of 800 m. The new near-range channels are then used to extend those profiles further down. But all these topics are more related to data evaluation than to the technical aspects presented in this paper.

Table 2 is incomplete:
Are all laser pulse energies (not power!) the values after the THG?
Thank you, it was changed (energy after THG). The requested information below was included.
Which SHG and THG crystals are employed? ->now included
Laser beam radius and divergence? ->now included
Is the beam divergence FWHM? -> full width at 86% energy, as specified by continuum, now included
Telescope secondary mirror diameter, distance from primary, and obscuration? ->now included
Telescope axis distance from emitted laser beam axis? ->240 mm, now included
Is the field of view FW? ->now included
Diameter of collimation lenses and focal length? ->the lens types are now given, Thorlabs standard
P7756 L24 *The system combines latest EARLINET lidar quality standards in a stand-alone design.*
Similar: P7749 L4: Here we removed the EARLINET reference as there are no real standards defined.
Which are these standards? Can you give a reference?
We included several EARLINET related references on quality standards in the section 2.4 now.
Figure 11 shows error bars for all products without any comments or explanations what they mean and how they are achieved.
We now include in the caption: "*The error bars for the water vapor, the extinction, and backscatter coefficients include the signal-noise and for the latter one also the Rayleigh-calibration errors. Those errors propagate to lidar ratio and Angström exponent. The error bars of the depolarization ratio additionally include the calibration error of C, the uncertainties of Ri, and for the particle depolarization also the uncertainty of the backscatter coefficient*."

Correction of review C2813 of MS No. amt-2015-184 (doi:10.5194/amtd-8-7737-2015)
EARLINET Raman Lidar PollyXT: the neXT generation
by R. Engelmann et al.
When discussing the overlap correction and its possible effects on products of signal ratios, my mind had obviously been trapped for some time in the problems of the receiver optics, and my criticism is partly wrong – on the one hand. On the other hand I forgot to mention the real problem.
While the mentioned problems with the receiver optics do exist in principle, the angular distribution of the light beam in the receiving optics is actually limited by the field stop of the telescope with 1 mrad field of view. If the field stop, the collimating lens, and the following optical elements are well aligned, the small 1° divergence of the light beam resulting from the 1 mrad field of view of the telescope is reasonable, and considerable range dependent signal distortions or significant differences between the signals of signal ratios are not to be expected.
Therefore the sentences in the original review text (highlighted below) can be deleted as shown.
P7754 L10 *Figure 10 shows the simulated and experimentally determined overlap functions for the far-range and near-range receiving telescopes.*
However, Fig. 10 shows that the overlap function is close to zero below 200 m, but Fig. 11 and 7 show values down to about 100 m for the linear depolarization ratio and even lower for the water vapor mixing ratio. It should be discussed how trustworthy these values are.
This remark is of course correct. We removed these "invalid" data points.
The problem in the overlap range I forgot to mention is that the cross-sectional intensity distribution of the emitted laser beam is different for different emission wavelengths. This is an effect of the SHG and THG, because the SHG depletes the intensity of the 1064 nm where 532 nm is generated, and the THG depletes 1064 nm and 532 nm where 355 nm is generated.
Because the telescope images the laser beam around the telescope focus, the field stop truncates the beam at different wavelengths differently, which results in different overlap functions. Therefore it is not possible to transfer the overlap function from one emission wavelength to another, and it should be explained how accurate the overlap function for the 1064 nm backscatter coefficient can be achieved, for which no Raman channel or near range telescope exists.
You are right!!! This effect might be an interesting point to consider. One could for example try to simulate the overlap function in ZEMAX with different laser intensity distributions and examine the

resulting differences. But (without being proofed yet) we would guess it only plays a role in the steep slope of the overlap function at the beginning. Looking at Fig. 10 maybe up to 300-400 m in height. Of course the effect depends on the ratio of the laser beam divergence and the field of view which is in our case as large as 5. So relatively quick, the overlap function should be dominated only by the distance of the image plane and the focal plane of the telescope rather than by the laser intensity distribution. Nevertheless, this effect should be calculated! But again, this is a general issue and not specific to the Polly systems.

Up to now, our primary quality test is the comparison of a new Polly system with the other systems in Leipzig. In this way it is much easier to identify alignment or design problems. This procedure is usually done among different EARLINET test campaigns.

Also with the telecover test the signal ratios are determined (as shown in the new section 2.4) for each telescope quadrant separately. If these four ratios agree it is also be a sign that the overlap functions of both channels are similar. Down to 350 m height this was found to be true within a relative margin of 20% for e.g. the PollyXT_NOA system.

Interactive comment on "EARLINET Raman Lidar Anonymous Referee #2
Received and published: 7 September 2015
"EARLINET Raman Lidar Polly.." by Engelmann et al, AMTD 7737, 2015

The authors report on the latest development of a mobile lidar system, POLLY XT that is used for measurements of particle backscatter and extinction coefficients and depolarization ratios at 2-3 wavelengths. In addition, water vapor is measured. The system is compact and comparably light in weight, which allows for deployment to remote regions on earth. The authors present details of the system and put the system in comparison to previous versions of the system, the prototype of which was developed around 15 years ago.

This research group has the potential of developing what they claim could become an AERONET-similar network of ground-based, well calibrated, standardized remote sensing instruments (lidar). As in the case of AERONET the same data analysis procedure are used, hardware is well characterized and calibrated, (near) real-time display of the data products is offered. Helping other research teams with issues in case their instruments fail or provide low quality profiles (AERONET is offering the same service) is another asset. In that sense POLLY could also be a candidate for complementing the AERONET network at some aerosol supersites.

In general the paper is well organized, informative and well written.

Thank you for the effort of reviewing the manuscript and for these general comments. These are exactly the points that keep us going with our work. However the systems have not been commercialized up to now. These systems are still considered to be research lidars with a tendency towards standardization. All systems (except NIER and CGE) were constructed in terms of research collaboration agreements.

We will answer the specific points in detail below.

General comments: The paper is acceptable for publication, but I ask for some mandatory changes which mainly concern improvements of the technical part, i.e. the description of hardware and data acquisition and some additional information on the hardware and calibration. Furthermore, the measurement example at the end of the manuscript would be an ideal platform to show how calibration and other technical improvements improve the quality of the derived optical parameters. Otherwise the example section is more like a "instrument indeed works" section, and I believe that the instrument is capable of providing the data products shows in the manuscript. The interesting part however would be to show in how far the quality of the data products been improved with the new Polly system. This part will also be important for the follow-up paper, as in that case the authors may be able to refer to this present paper when it comes to presenting the data of many different Polly systems that took data over the past decade.

We tried to give more details about certain technical aspects in the new version. As it often occurs for lidar measurements not all error sources are fully understood and sometimes hard to quantify. In that sense it seems a bit difficult to quantitatively show what the improvements were on the measured data. Many improvements aimed at standardization. But the new DAQ unit probably has the biggest influence on the data quality. Firstly, it has a much lower dead time on the order of 2-3 ns while with the old system we experienced dead times on the order of 10-15 ns. Secondly, it behaves much better especially shortly after laser beam emission and thus the signals close to the lidar are much better to use. Before it was not easy to distinguish if distortion effects were caused by a probably misaligned overlap or by electromagnetic interference of the DAQ with the laser.

We now decided to name the latest Pollys (New data acquisition and 3D construction) Generation 3. This might help to identify the systems better. Also Table 1 has been updated to include the differences among the systems better.

The measurement example at the end is thought to give a summary on the parameters that can be derived and on the approximate errors that are expected.

The follow up paper by Baars indeed contains a nice table at which locations and with what measurement configuration all systems have been used in the past.

In that sense the link between the technical part of the paper and the short experimental part has potential for improvements, not in the sense of scientific interpretation of the results, but in the sense of convincing the reader that the high-end technical improvements transfer into trustworthy, high-end optical parameters which are among the goals of EARLINET: “quality assured data products.” The authors should explain in more detail the calibrations that are needed before deployment to field site.
We now included an additional section 2.4 which primarily focusses on the Telecover Test, a standard method to check the system alignment and approximate errors resulting from the alignment.
Are these calibrations (and their parameters) really stable, also in view of the various environmental conditions (very hot to very cold air)?
The cabinet is equipped with heaters and an air condition system and internal temperature sensors. For most installations, the internal temperature was kept within 20 and 25°C. We are constantly working on improving the regulation in order to keep the temperatures even more stable. But you are right, more testing might be required in the future.
How are conversion efficiencies of the SHG and THG affected by external temperature changes?
SHG and THG are heated to about 70°C with a remotely adjustable temperature controller (now mentioned in the text). Up to now we did not experience any temperature instability with these crystals. However, if the laser beam changes pointing direction (depends on internal laser temperature) then an adjustment of SHG and THG temperatures might be needed.
Is the cabinet of the lidar completely sealed off (isolated) in terms of outside temperatures?
The cabinet walls are double layered. The latest version is indeed isolated within the double walls.
You mention the overlap: it would be helpful to see if the overlap is the same for all channels, and achieving the same overlap function for all channels is a challenge. The approach as shown by Wandinger and Ansmann (2002) cannot be done in the case of the backscatter at 1064 nm. You only derive the overlap function at 532 nm? What makes you confident that the same overlap can be applied at 355 nm? Are the overlap functions solely based on simulations carried out during the design phase of Polly?
As already mentioned at the first review, we do not often use the overlap function for correction. And of course, if we use it we try to measure the overlap functions for UV and VIS separately. In terms of the 1064 issue we can only rely on the Telecover Tests up to now, and also on a constant Angström exponent in a well-mixed PBL. And for the determination of the backscatter coefficient we sometimes use the ratio 1064/607, to derive the Backscatter coefficient further down to the ground (3 wavelength Raman method). However caution has to be taken, because of the uncertainties from the unknown Angstrom exponents.

More details on specifications of the photon counting system would be helpful.
We included some more details in Tab. 3. But we wanted only to give some overview of the DAQ system. More details can probably be obtained by the developer Holger Linné (MPI,Hamburg).

How does the laser behave under various humidity and temperature conditions? The photos that show field sites of the instrument obviously show the challenges involved. In space technology environmental chambers are used to test the behavior of the equipment.
Note: The size of this instrument would allow for developing such an environmental chamber at TROPOS and simulate the various environmental conditions under which the system needs to operate.
You are right. The climatic conditions are often a major challenge. This topic is also not fully resolved, yet. If the lidar will become a commercial product one day such environmental tests need to be performed. Also expert people who know about climate control should be involved. This was however not possible up to now at TROPOS. We only gained the experience that the systems run somewhat more stable in a concealed (e.g. container sized) housing than in the outdoor cabinets. Of course, the bigger air volume in such a container can be better controlled. On the other hand, for

field campaigns and transport via airfreight it is more feasible to use the small outdoor cabinet. In this case overlap readjustments (can be done remotely) have to be performed more often.
That comments brings me back to calibration and stability of the system design. A few more plots and tables detailing the variability of important parameters, like expansion or retraction of the carbon-fiber parts, laser performance, alignment of the optics, stability of the overlap function would help convincing the reader a) why this instrument in fact could become a standard instrument for the research-based lidar community and b) why such a system may be more suitable than commercially available instruments for setting up a worldwide network of multiwavelength Raman lidars.
Of course, we do not want to claim to have better systems than commercially available. Every system could probably be built by a company, but TROPOS follows the strategy to construct the systems with the partners together to have best practice knowledge-transfer so that the operator of the lidar can know any detail of the system. So Polly lidars are in that way very different from a commercial system as all knowledge is provided and shared and solution for critical issues are made together. In fact, if scientists only chose to buy a system, there are companies available offering also advanced lidar systems, but this is not the purpose of a research institute as TROPOS is. The characteristics you desire are of course of high interest, and we always try to improve the methods and develop new calibration schemes, but also haven't been able fully evaluate all the mentioned details and thus cannot present them yet. These are task which are still subject to research and the new EARLINET lidar calibration center in Bucharest which is set up in the frame of ACTRIS-2 is perfect to deal which such specific issues.

What material is the primary mirror?
Pyrex, now included in Table 2.
Can the distortions of the frame be reduced by choosing lighter mirror material?
The distortions are not an issue. The receiver field of view is 1 mrad, and 0.02 mrads caused by the deformation are below the laser pointing stability. The graph shall show how FEM calculations were used to improve and check e.g. the mirror mount.
What are the dead time corrections beyond 40Mcps? Is this dead time correction (figure 6) done for all detectors? What is the uncertainty of the 5-th order fit?
We haven't studied all aspects of the dead time correction so we usually only operate the detectors up to 40-60 Mcps so that we can safely correct the dead time effects (max correction by a factor of 1.3). Most likely higher count rates would be possible to correct, too, but for now it feels "uncomfortable". The dead-time-correction polynomials were determined for all H10721P-100 PMTs and hence the correction is done for every channel. Fig. 6 just shows one example. On one occasion we also tested the 1064 nm PMT with the presented method. Usually the 1064 nm PMTs are only operated up to 20 Mcps were the effects are still very small until more measurements have been performed.
In general I am missing uncertainty bars of the calibration constants.
We included more uncertainty bars for the Ri values of thy systems. The uncertainty of the depol calibration constant C is mainly given by the signal error (now noted in the figure caption). An extended error analysis on depolarization measurements is given recently by Belegante et al, 2016.
Please show them for all channels and/or show a table that summarizes the polynomial regression (fit) constants. I am sure these parameters will be checked during maintenance cycles of the instruments.
The dead-time measurement can only be performed in the laboratory and has therefore only done once per system and channel. We now included the polynomial parameters for the shown measurement. All other coefficients are stored in the individual system configurations and in the data files. We think they are probably not so interesting for the reader. But, we now included the dead-times and their standard deviations for 32 different PMT modules according to the paralyzable and non-paralyzable theory.
The paper is mainly about the instrument, thus more tables with technical details of the detectors, photon counting, mirrors and optics in terms of numbers would be helpful.

We now included some more specific parts in Table 2. Mostly we mentioned the part names from which all parameters are available online at the manufacturers.
A table that shows the various improvements and modifications that were made between the prototype and this next generation Polly would be quite instructive and would put the text in a better context, e.g., calibration stability, detector sensitivity, counting efficiency, overlap stability, new procedures for instrument characterization that were not done for the previous versions of Polly? The history of Polly in that way could be clearly followed which is important for future work.
We improved Table 1 a bit. The procedures for the instrument characterization are mentioned in the text and time by time are performed on every instrument. The detector sensitivity is given from the manufacturers of the PMTs. This sensitivity is however not of major importance as we use ND filters in front of every channel (except water vapor) to reduce the signal to a photon-counting level. The only calibration that we regularly perform is the depol calibration. The value and error of C is given in the text. Overlap stability is also not very easy to describe. These parameters always depend on the specific lidar system (e.g. the laser stability is different for every single laser) and on the climatic conditions and can thus be different for all systems. What we can estimate up to now is that the beam pointing within the field of view (1 mrad) is stable maybe on a level of 0.2-0.3 mrad in the worst case. This uncertainty we try to overcome with the near-range channels. Maybe for the future we could also think of an active beam stabilization.

Depolarization ratio measurements are particularly critical, and the authors spend some time explaining the procedure of calibrating the channels and how the automated routines work. A plot showing the calibration numbers and how an offset from these values affects the final data products could be put in a section called "sensitivity study and error analysis". In fact, the uncertainties that are involved in the calibrations and their dependence on various environmental factors like (fast) temperature changes or extreme temperature regimes might be included in the sample measurement that is shown at the end of this paper. Although the paper is for the most part about the technology of the instrument the measurement example could serve as a platform to guide the reader through the different parts of the calibrations, the photon counting efficiency, dead time corrections, etc,. In that sense the authors may want to reconsider the way they present error bars in the measurement example. I assume the error bars are done in the "usual", i.e. "traditional" way of averaging the signals in time and space, carrying out calibration in the molecular atmosphere, lidar ratios need to be assumed in the case of the backscatter at 1064 nm, etc . It would be helpful to see how the error bars become larger (smaller?) if the technical components (calibration numbers) of the instrument and the data acquisition properties are taken into account of. In fact this is an important task in EARLINET, i.e. what do the error bars of the profiles in fact mean, what are they made up of (experimental error versus "technical, hardware-driven", i.e., instrument error), what is the statistical law that determines the error bars (Poisson, Gauss, log(?), random... ?). Polly could provide a first approach to this more sophisticated error analysis.
You are right. There are a lot of error sources to be considered. We indeed followed the "traditional" way of error calculation. We included errors that we can specify and propagated them to the result (Poisson error, error of the depol. calibration constant C, errors of Ri values, errors of the backscatter coefficient (part. depol.). Possibly we could use Polly for a further analysis. But this would be an entirely separated publication on data analysis schemes. And we also agree, EARLINET should take up the task to define, how errors are treated (EARLINET lidar calibration center). But again, we think this is a more general lidar issue rather than specific to Polly and therefore should not be included in this publication.

Figure 7: what is the reason for the significantly different water vapor mixing ratios below 600 m between the radiosonde (and MARTHA) and Polly? What is the sensitivity of Polly with respect to measurements above 6 km height? Is there a systematic offset or statistical noise?
In fact, Polly was never considered a water vapor lidar. The laser energy is somewhat small at 355 nm. However, adding this channel gave a nice addition to the system and the profiles, especially in

conjunction with microwave-radiometer measurements these data are extremely helpful (eg. Foth et al., 2015).
We added: *"A measurement close to the ground (>200 m height) is feasible because the water-vapor mixing ratio calculates from the 407 to 387 nm signal ratio and if the overlap functions of both channels are equal they cancel each other out (e.g., WandingerSpringer2005). The reason for the deviations on the order of 0.2 g/kg (16% error) of the PollyXT_IFT data below 800 m height from MARTHA and the radiosonde data remains unclear for the presented measurement. Either electronic distortions by the older P7882 data acquisition setup or different overlap functions for 387 and 407 nm might have caused the discrepancy."*
Measurements above 6 km are not shown because the water vapor mixing ratio is smaller than the statistical error. Of course, depending on smoothing and temporal averaging the sensitivity is variable. We haven't seen a systematic offset because for all our tests the statistical error was larger at greater heights.

Are the neutral density filter (e.g. in figure 9) calibrated as well? Do they pose a significant source of uncertainty in the calibration procedure?
The ND filters are not calibrated individually as keeping track of them would be almost impossible. We perform the depol calibration regularly including all receiver optics and PMT sensitivity (except the lidar roof window and the secondary telescope mirror, which have to be assumed fix). Therefore changes of the ND filters are reflected in Fig 10.
Figure 10 shows the overlap function. Was the overlap function measured several times during the ship cruise? What was the scatter of the experimental overlap function? Is the "true" curve (simulated) the one obtained during the design phase on the computer? There is no uncertainty bar of the simulated overlap function. In view of the various parameters that determine the overlap function this uncertainty would need to be included in the presentation of the final data products.
The overlap function can unfortunately only be measured if the aerosol conditions are extremely clean. Otherwise the assumption of a lidar ratio results in large errors. Thus the overlap function was only measured once at the ship cruise.
The simulated overlap function was calculated by raytracing with ZEMAX. We did not perform an extensive analysis on manufacturing margins of specific parts, alignment errors, and the influence on the overlap function. With this figure we wanted to show how the overlap functions look in general for Polly and how simulation and measurement agree. The specific function for each system has to be determined individually. Another approach is to use the near-range channels to study the overlap function of the far-range channels. Now with even a 4 channel N/R receiver this can be done for the UV and VIS wavelengths. But the analysis is not so extended up to now to include it in this publication. Maybe in the mentioned future publication on error analysis this would fit better.

Figure11: what is the reason that the backscatter at 1064 is so much lower than the backscatter at 355 and 532 nm (neutral spectral behavior)? What is the reason that extinction shows wavelength dependence in the center of the aerosol plume, but the 1064 nm backscatter is off? Is this an instrument effect or an aerosol-type effect? If it is an instrument effect it should be explained in this paper.
This is clearly an aerosol type effect. The dust particles are in the size range of 1 um and thus at higher wavelengths the spectral slope is not neutral anymore. This effect was shown several times during e.g. the SAMUM I and II campaigns (e.g. Tesche et al. 2011).
What do you mean by "predefined adjustment apertures"? What happens if they change their properties?
These apertures are fixed in relation to the transmitter board. The most variable part is the laser beam direction, e.g., after a laser maintenance or readjustment. In such cases the apertures help to realign the laser beam quickly.
Specify the EARLINET guidelines (7738, line 13) as the reader may not be familiar with them and will not find these numbers (guidelines) in the EARLINET publications. Specify in a table how close Polly is in achieving these "guidelines" or even exceeds them.

We included some comments on quality assurance in Sec 2.4 which might be more helpful now. Up to now the EARLINET guidelines include system comparisons, annual Telecover Tests, and Rayleigh fits. In general we perform those tests more often, especially after transportation.

Please avoid using qualitative terms like “rather”, “appropriate (e.g., 7742, line 6), “sufficient”, etc.
We tried to straighten this out by providing numbers or references if possible.
It is a technical paper, so please provide as many numbers as possible and put them in the context of older Polly versions (in a table).
The older systems were described by Althausen et al. 2009. We think we really shouldn’t repeat these specifications. However, we now tried to provide the most necessary (and to us available) information of the new (generation 3) systems.
“Smart” (7741, line 12) though I understand what you man to say by this, but is not the right word in this context. Expressions as for example “sophisticated”, “thought-out”, “mature”, “elaborate”, “meticulously thought through design...,” maybe be better choices.
Thank you, we have changed these words accordingly.

Some language editing at a later stage of the manuscript preparation could further improve the readability of the paper. Words and grammar are often used in the German context rather than the English context. Sentences lack in verbs in some spots of the manuscript.
We tried to go over the manuscript several times to smoothen the text. We’re sorry for the German context use of grammar that is most likely still left in some occasions.