Author response to Anonymous Reviewer #2

Reviewer #2 comments are in black and author comments are in blue.

This is a good manuscript that is within scope and of sufficient merit for publication after minor revisions. I appreciate the level of detail to which the authors describe the calibration and uncertainty assessment process for this class of instrument. Most of my comments address vagueness or sloppiness in the description with the intent of making the work more easily 'repeatable'. These should be easy to address.

The authors thank Reviewer #2 for their thorough pass through the manuscript and recommendation for publication. We address their individual remarks with comments below. If there are duplicated remarks, we will address the first and that will hold for all others.

I have two main complaints:

Figure 10b is used to demonstrate that AirHARP and the RSP instrument agree within 1% in DoLP. Because of the plotting log-scaling, I don't agree that one could conclude that from the figure for the majority of cases for which DoLP > 0.1. My comments below expand on this and offer suggestions on how to resolve.

I think the simplified AirHARP uncertainty model in the appendix is too simplified. Furthermore, the paper doesn't show how this model was derived based on previous equations. It also does not depend on scene reflectance, which it should. More details on this below.

These two points will be addressed in detail in later areas of this document.

I do think this is an important paper and commend the authors for clearly describing an approach that has been developed over several years. It will be an important tool as the PACE/HARP2 instrument and other similar polarimeters come online.

Abstract:

Line 18: clarify if RMS is defined in units of DoLP, or is relative. Also note RMS acronym should be spelled out. My personal preference is to use the terminology in Povey et al:

Povey, A. C. and Grainger, R. G.: Known and unknown unknowns: uncertainty estimation in satellite remote sensing, Atmos. Meas. Tech., 8(11), 4699--4718 , https://doi.org/10.5194/amt-8-4699-2015, 2015.

… in which case this statement would be more like "One sigma uncertainty of 0.25"

Line 19: Spell out FOV

The root-mean-square (RMS) uncertainty here is in units of DOLP. We will make both the acronym and units explicit in-text. Reviewer #1 also remarked on the acronym definitions, so we will address all in-text definitions and make sure they are first defined before they are abbreviated.

Rest of the paper

Page 2, line 44: again here you might want to point out that 0.5% is in units of DoLP, not relative as is the case for the radiometric calibration

Page 2, lines 46-50, page 3 lines 55-60

Other SPEX references you might want to consider adding:

Rietjens, J., Campo, J., Chanumolu, A., Smit, M., Nalla, R., Fernandez, C., Dingjan, J., van Amerongen, A., and Hasekamp, O.: Expected performance and error analysis for SPEXone, a multi-angle channeled spectropolarimeter for the NASA PACE mission. in: Polarization Science and Remote Sensing IX 34 -- 47) SPIE., 2019.

van Amerongen, A., Rietjens, J., Campo, J., Dogan, E., Dingjan, J., Nalla, R., Caron, J., and Hasekamp, O.: SPEXone: A compact multi-angle polarimeter. in: Proc. SPIE 11180, International Conference on Space Optics --- ICSO 2018 , 2018.

Also note that the SPEX spectral spectral resolution is different for intensity and polarization. Polarimetric samples are every 10-40nm, depending on where this is in the spectrum.

Thank you for these citations and further information on SPEX. We will assess the papers and add them to the manuscript.

Page 3, line 65: here you refer to the instrument MSPI, above it was described as AirMSPI

Page 3, lines 71-72 AirMSPI needs to aggregate pixels to reach <0.005 DoLP. Further explored in

van Harten, G., Diner, D. J., Daugherty, B. J. S., Rheingans, B. E., Bull, M. A., Seidel, F. C., Chipman, R. A., Cairns, B., Wasilewski, A. P., and Knobelspiesse, K. D.: Calibration and validation of Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) polarization measurements, Appl. Optics, 57(16), 4499--4513 , https://doi.org/10.1364/AO.57.004499, 2018.

Knobelspiesse, K., Tan, Q., Bruegge, C., Cairns, B., Chowdhary, J., van Diedenhoven, B., Diner, D., Ferrare, R., van Harten, G., Jovanovic, V., Ottaviani, M., Redemann, J., Seidel, F., and Sinclair, K.: Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment, Appl. Optics, 58(3), 650--669 , https://doi.org/10.1364/AO.58.000650, 2019.

Note the difference in conclusion of these two papers. It's not particularly relevant to this paper though.

Yes, we are aware of both excellent papers and have cited them elsewhere in the manuscript. We will add the citations here as well.

Page 4, line 95 I'm not familiar with the terminology "Alternative Vision". Can you provide a reference?

We apologize for any confusion here. Alternative Vision is the reference for this statement. It is a website with a technical article on the Phillips prism design. The author was not listed on the website. See citation below:

Alternative Vision: Separation Prism Technical Data, available at: https://web.archive.org/web/20070607142209/http://www.altvision.com/color_prisms_tech_data.htm, last access: 9 August 2021.

Section 3.1: I like the differentiation between AirHARP, HARP2, HARPCubesat in the language and the use of 'HARP' as general description, but would have expected based on this that this section would use the language of HARP rather than AirHARP. Does that mean the detector specification and background correction is different for the various instruments? Perhaps so, but I think it would have been nicer to include details on all three instruments in table 1.

** Thank you for this observation! The name juggling is a bit more systematic than we realized after submission. We will address any inconsistencies throughout the manuscript in the revision.

Reviewer #2 makes a keen observation that is important to the scope of the paper. The calibration theory is shared between all HARP instruments. However, the instruments and data can be different enough that going into too much detail could distract readers from the focus of this paper. For example, AirHARP and HARP CubeSat share a similar design, but have differences in performance, uncertainty, data compression, on-board storage, and data processing due to design choices and operational environment. HARP2 is a spiritual successor to HARP CubeSat. We consider it to be a different instrument entirely because many aspects of the HARP design were optimized.

Therefore, we decided to focus this work on AirHARP and reference HARP CubeSat or HARP2 at a top level only. Finally, AirHARP is the only HARP instrument with public data so far. By focusing on AirHARP only, this paper can be used to directly connect the calibration activities to the available AirHARP L1B reflectance products. HARP2 (and potentially HARP CubeSat) will see their own calibration and data papers starting next year. HARP2 was more extensively and rigorously tested than AirHARP as well, as it is part of a major space mission with desired performance goals.

We will refer to this response throughout this document as needed with the ** indicator.

Figure 3: needs labels for (a) and (b) in the figure. Also there appears to be some sort of border irregularly surrounding the plots.

This figure will be modified to remove the irregular border and add these labels in the revision.

Page 7, line 174 – AirHARP and HARP Cubesat have an internal shutter, does this mean PACE HARP doesn't?

See ** above. HARP2 does have an internal shutter, though it is different in several ways than AirHARP and HARP CubeSat. We will revise the language here to focus on AirHARP only.

Page 7, last paragraph: considering that the CCD dark current is different in the cross track direction, would have it made more sense to have oriented the CCD 90 from this so that the vignetted areas to use for a synthetic dark calculation are more uniform?

Not necessarily. We use a dark template from the lab, which is a full frame dark capture, as a basis for the synthetic correction. If we know the relative spatial distribution of the dark frame over the entire FPA, we could use any region of a vignetted detector to scale the template by temperature – as long as those two areas have a similar dark response. Therefore, there is no preferential orientation of the detector for the dark correction.

Page 8, line 198-199: "with a port fraction <5% and sphere multiplier >10 (PerkinElmer)" uses terminology that may not be familiar to all readers – please define port fraction and sphere multiplier.

We will add to this paragraph that the port fraction of an integrating sphere is the ratio between the internal surface area and the area of open holes and apertures. The sphere multiplier characterizes how much the reflective surface and port fraction of the sphere contributes to increasing the light output relative to the source. We will also add the PerkinElmer citation, which we accidentally omitted from the reference list.

Page 8, line 206: I'm confused by what appears to be a link: andor.oxinist.com, which is actually a reference.

We apologize for the confusion. We will change this citation to "Oxford Instruments Andor". Like the Alternative Vision reference, it is a technical article with no lead author and no publication year given.

Page 10, line 260: Wouldn't it have been simpler to just represent this as integration time instead of 'integration lines'? It is confusing at first glance because larger values mean opposite things. For example the statement "Sensor B saturates earlier than Sensors A or C" is confusing because the lowest 'integration line' corresponds to the longest integration time, if I understand correctly.

Yes, we will make this change in the revision. Integration lines represent the integer values we used in the HARP instrument firmware to set integration times for each detector, and in hindsight, they are not as intuitive as the actual integration times in a plot like this. Thanks again for the suggestion. And your understanding is correct – larger integration lines correspond to shorter integration times.

Section 3.3: If possible, add a sentence or two explaining the likelihood that doubling the integration time is identical to doubling the sphere intensity. This test, which varies exposure time but not intensity, relies on this.

Yes, this is true and a core tenet of the non-linearity part of the pipeline. We will add this to the text is some fashion. Beyond this, the non-linear correction allows us to modify the detector integration time on the fly during a field campaign or space mission, with the knowledge that the detector response will be predictable.

Page 11, line 264: how is the 'linear region' defined at 3000 ADU? I think I understand what you are doing in Figure 5b that expresses this, but feel like you need another sentence or two to explain how you're going from fig 5a to fig 5b

The linear region upper bound at 3000 ADU is the location where the fit residuals are no longer randomly distributed around zero. The bound itself depends on the resolution of the data and is empirically defined. We will add more detail on this in the text.

Equation 4: I was confused by this equation until I realized that the DNflat on the left hand side of the equation is different from that on the right hand side. The left hand side is the value of DNflat from the linear *fit* to the data, not the actual data. If my understanding is correct, I suggest you use some notation to indicate that they are different things.

We apologize for the confusion here. The entire left side of the Eq. 4 is the fit in Figure 5b. DN_corr is the linear fit from Fig 5a and DN_flat is the counts data from Fig 5a. We will separate the left side into a separate equation DN_fit = DN_corr – DN_flat, and have the current Eq.(4) be DN_fit = the polynomial function.

Page 16, line 396: Can you clarify if the polarimetric coefficients are determined for all pixel bins, or just the nadir bin?

The polarimetric coefficients shown in this work are determined for the nadir bin, and we leverage the telecentric optical train/flatfield to spread those coefficients to other FOVs. We will clarify this in the text of this section.

Page 16, line 398: do you mean 'deviate' rather than 'derivate'?

Page 17, line 412: you're referencing equation 5, but do you mean equation 10 instead?

Page 17 line 413: is sigma_c_ij really from Table 2a rather than 2b?

Yes, this is a typo. Sigma_C_ij refers to the Table 2a. We will address this in the revision.

Page 18, line 440: I suspect that you're referring to a different equation here than eq 5, suggest you verify all references to equations are correct

We will address the above comments to Page 16, 17, and 18 in the revision.

Page 18, line 442: I thought you were referring to "Detector A, B, C" rather than "Sensor 1, 2, 3". Is there a difference? If so, explain, if not, please make the terminology consistent

Detector [A, B, C] and Sensor [1, 2, 3] mean different things. Detector [A, B, C] refer to the physical FPAs behind the each port of the prism. Sensor [1, 2, 3] refers to the light path through the instrument that corresponds to Detector [A, B, C]. This is defined in Section 2, line 108. It is a nuanced definition and we will address any inconsistencies in the text in the revision.

In this section, the SRF is calculated at the Sensor level because there is a possibility that the SRF in Sensor 1 is not the same as Sensor 2 or 3 (as we see in the 440nm). The actual Detectors play a role in the SRF, but only from the perspective of quantum efficiency (shown in Table 1). The system SRF is a combination of the detector QE, stripe filter, prism response, and multi-bandpass filter.

Page 20, line 484: "the illumination of the nine lamps is linear" – does this mean that each lamp adds linearly to the total illumination?

Yes, and we will add a statement about this in the revision.

Page 21 line 501 and equation 17: If gamma doesn't mean the same thing as in equation 11, I suggest using a different parameter. If it does, please note this.

Great catch, these are different parameters. We will use an epsilon in Eq. 17 instead.

Page 21 line 504 "compatible with zero within 3 sigma" what does this mean? Does it mean 'equivalent to zero'? I presume so, but I don't know what within 3 sigma means in this context.

We do not show this plot in the paper, but the radiometric calibration is a linear fit between the calibrated radiances of the NIST-tracable integrating sphere and the relative intensity (Eq. 17 terms in parentheses) measurements of the sphere light by AirHARP. The calculated uncertainty on each fit parameter (k and gamma) give an estimation of how well the parameter describes the data. The gamma term is equivalent to zero within 3 times its fit uncertainty. We will remove the "sigma" definition and explicitly describe this in the revision.

Page 23, line 560: references characterization in “late 2022” Is this a typo, or if this already happened please rephrase (and note where the error model is documented)

Thanks for the catch here, and this shows how long this paper was in development! We will remove this sentence altogether for reasons in ** above. However, this test was done for HARP2 and researchers in our group are currently assessing the data. The results for ** reasons are beyond the scope of this AirHARP work. The AirHARP error model has not been documented publicly. We will address the reviewer #2 concerns on the error model in later sections below.

Page 24, lines 571-9: I think you reference them earlier, but this would be a good place to put references for the LMOS and ACEPOL field campaigns as well.

Yes, this makes sense. We will add these citations here in the revision.

Page 24, line 591: does this require a priori information on wind speed / surface roughness?

Not for the cross-comparison shown in this work. We know the target is sunglint from visual context in the AirHARP data and the geometry of the observation, but we are only interested in the closeness of the AirHARP-RSP angular matching over the same targets.

Page 25, line 600: perhaps some citations for the RSP instrument here, specifically about calibration?

We will add van Harten et al. (2018), Knobelspiesse et al. (2019), Smit et al. (2019), and Cairns et al. (1999) citations, which are a good spread of calibration work using RSP over the years.

Page 25, line 610: is the RSP footprint a circle, or an oval? There is the IFOV, but an integration time that makes it like an oval. The Knobelspiesse 2019 paper you cite has a model for this as well as McCorkel et al 2016. It would potentially make your fits slightly better.

McCorkel, J., Cairns, B., and Wasilewski, A.: Imager-to-radiometer in-flight cross calibration: RSP radiometric comparison with airborne and satellite sensors, Atmos. Meas. Tech., 9(3), 955--962 , https://doi.org/10.5194/amt-9-955-2016, 2016.

The true RSP footprint is an oval due to bi-directional smear, as discussed in both papers. An important thing to note though is both papers are intercomparison papers first. Our paper only uses the AirHARP/RSP intercomparison as a top-level demonstration of our nadir calibration spread to all FOVs. Therefore, we simplified the intercomparison here in several ways. One of those simplifications is the RSP footprint as a superpixel square. This said, converting this to an oval as in Knobelspiesse et al. (2019) and McCorkel et al. (2016) is a reasonable suggestion by reviewer #2, and would not require a major revision. This may improve the Figures 9 and 10 in lieu of their later comments and has basis in the literature.

Page 27, line 663: only Fig 10b appears to be log-scaled

This was done to emphasize the comparison at the low end of the DOLP, which is part of the focus of this paper. Polarimetric uncertainty and intercomparisons in this DOLP region are of considerable interest to the climate community, as modern aerosol and ocean color property retrievals may rely on accurate measurements at low DOLP.

Figure 10 refers to “Figure 32” which isn’t a part of the paper. I presume you mean Fig 9

Yes, we will correct this in the revision.

Figure 10 (right): This figure doesn’t convince me that RSP and AirHARP agree within 1% for the log scaled DoLP. Because of overplotting, it is impossible to tell if the difference is beyond 1% for DoLP > 0.1 (possibly majority of observations based on Fig 9). It would have been better to plot the difference between RSP and AirHARP. It would have been even better to consider how that difference compares to the paired RSP and AirHARP uncertainty estimates. Validation of the uncertainty model is what is most important for algorithms which will use that model in retrieval. The models are not fixed: for RSP, the uncertainty varies with DoLP and reflectance, for AirHARP the simplified model shows it varies with DoLP (see later comment about the AirHARP simplified model). Furthermore, you cite the papers (Knobelspiesse et al. 2019, Smit et al. 2019, van Harten et al. 2018) that I’m not sure support the assertion that the difference is ‘reasonable’ since they relate to other instruments besides AirHARP. However, they do provide examples of more in depth analysis on this.

So, my recommendation is to redo figure 10b to show the difference between RSP and AirHARP scaled by the squared sum of the uncertainty estimate for the given DoLP and Reflectance value. For one sigma uncertainties 67% should agree within those bounds. You could also check with 2 sigmas / 95%.

A 1:1 comparison was chosen for this paper instead of a direct difference because of the paper scope. We wanted to emphasize the overall comparison and not the angle-to-angle differences that are front-and-center in a direct difference plot, especially with only four data cases.

However, revisiting this analysis is a reasonable request from reviewer #2. If we were to show a difference plot in this paper as suggested, we would prefer to include more ACEPOL datasets than the four cases in the current version. We suspect this is a larger revision. It would avoid overemphasizing any granule-specific or angle-to-angle differences in the intercomparison though. Because reviewer #2 only suggested minor revisions, we will weigh the effort vs. benefit of redoing this part of the paper and perform any reasonable re-analysis.

If it is found that this amounts to a larger revision, we still agree that the interpretation of Figure 10 as noted by Reviewer #2 could be stronger. We will assess the readability, clarity, and overplotting effect of these plots. One solution in this direction could be to separate the plots by wavelength (as well as re-analyze with the oval RSP footprint and uncertainty model below). This would reduce the impact of point crowding.

Thank you for these suggestions.

Appendix A, line 760

You show a 'simplified' AirHARP error model which is what an algorithm developer would use. Presumably this comes from equation 19 and 14, but those are not expressed in terms of reflectance or DoLP. This appendix should show the full model for reflectance and dolp, and then the simplified model as you show.

Additionally, I am concerned the simplified model you show is too simplified. The DoLP uncertainty does not depend on shot noise or reflectance, and I find that that drives performance for scenes over the ocean. The (Knobelspiesse et al. 2019, Smit et al. 2019, van Harten et al. 2018) papers you cite all focus to some extent on this. Of course, the binning of AirHARP compared to RSP will drive down shot noise. Random uncertainty is reduced by sqrt(n), so I am also confused why you have the 1/B term in equations 26 and 27 rather than 1/sqrt(B) or similar.

Finally, I would have expected some decreasing performance from the center for where the calibration is performed to other pixels for which it is extended. Is this expressed in the calibration coefficients? Knowledge on this characteristic if it is indeed different would also be useful for algorithm developers who can account for this in retrieval algorithms.

Since the submission of this paper in May 2023, we have advanced the error model for AirHARP, though it is still currently unpublished. Adding a more involved error model based on the earlier equations is also a reasonable request from reviewer #2, given the utility in retrieval studies and public availability of AirHARP data. This will be adjusted in the appendix in the revision, and address reviewer #2's other concerns here.

As far as off-nadir performance goes, we can see from Figure 4c that the flatfield "flats" its generating dataset across the FOV within 0.5%. This distribution is comprised of random noise. The flatfield and telecentric physics are the vehicle that allows the nadir calibration spread to other FOVs. Therefore, our current lab analysis suggests that there is negligible off-nadir degradation in response. This may change in the field data due to interpolation during data processing and variable FOV effects like lens condensation or pitch surfing mentioned in Section 4. These impacts have not been fully characterized as of this work. Further studies are underway on these effects. Section 4 intercomparisons demonstrate the off-nadir performance in the field from a data perspective, which also suggests the performance is comparable at nadir up to 35 degrees in VZA. Adding more ACEPOL datasets to the comparisons in Figure 10 may improve the lab and field assessments of off-nadir performance as well.

Forthcoming papers will go into detail about off-nadir performance from a laboratory perspective for HARP2. Similar studies were not done at the same level for AirHARP for time or resource reasons.