This paper discusses the calibration pipeline for the polarization resolving cameras of the spectrometer of the Munich Aerosol Scanner (specMACS). The authors discuss the instrument itself, then delve into geometric calibration of the image frame, dark characterization, non-linearity, spectral response, polarization calibration of an enclosing window and optical assembly, flatfielding, and absolute radiometric response. The authors close with a discussion of overall uncertainty and a measurement-model intercomparison over sunglint measured during a recent field campaign.

This paper comes after Portge et al. (2023), which demonstrated the polarimetric cloud retrieval capabilities of the same specMACS cameras over marine and popcorn cumulus cloud fields. Polarimetric remote sensing is a hot topic in the climate community right now. Papers that demonstrate polarimeter instrument calibration (as well as their science) will continue to be relevant, interesting, and useful to AMT readers. I recommend publication with minor/optional revisions.

<u>In-line comments:</u>

Line 85: I understand that the Equation 1 is the form given in Hansen and Travis (1974), though the Sony sensor allows for a more comprehensive calculation using all four angles (shown in Lane et al. 2022):

$$I = \frac{1}{2}(I_0 + I_{90} + I_{45} + I_{135})$$

This form is used later in the paper (as a normalization during polarization calibration, line 313), but it isn't clear to me if the actual intensity measurement (Stokes I) is calculated this way throughout the entire paper. Either way, please harmonize the definition of I across the paper. Also, the typical convention of V in Eq. (1) is flipped from what is shown (right – left).

Section 3. I recommend to add a figure to this section. A visual of the chessboard calibration from the perspective of the specMACS sensor would help a lot with the interpretation here.

Line 193: Figure 3 shows that polLL and polLR have systematic differences in the forward and aft sides of the dark frame. Even at a ~2.5 counts spread, this structure is important to capture, and could be relatively easy to apply in post-processing on image data. At 30000DN, I agree it will not make much of a difference to use a single value or adopt a spatial dark map for correction. However, I imagine part of the interest in specMACS data goes beyond clouds – possibly to science retrievals of aerosol, land, ocean, and free atmosphere properties. Many of these targets will not have 30000DN signals. For example, open ocean is extremely dark in RGB and can go to <5% in DOLP off-glint, like in Figure 12b. At these low light levels, a few counts could be important. Also, the later sections discuss the many ways that the calibration could be improved – using the spatial field of the dark (and scaling the dark counts relative to any measurement temperature) could be a step in this direction. I would consider including this in the calibration pipeline instead of a single value for the dark.

Figure 6. It is challenging to differentiate the curves in each figure due to the overlap and large scale. This could be stronger as a residual plot (i.e. $\widetilde{S_o}$ - S_o), as a function exposure time for all pixels shown. Also, please make the points larger.

Line 215: Relative to the detector spec, are these non-linearities reasonable?

Figure 7. Is there is any new information in (a) and (c) that isn't already in (b) and (d)? If not, I recommend to only show (b) and (d).

Figure 8. Though the smaller peaks in blue @ 650nm and in red @ 550nm are typical of some Bayer filter designs, how is this addressed in the radiometric calibration? This could be important for cross-talk considerations, and may have some influence on the error analysis in Table 3.

Line 278 + Line 294: This may not hold for the entire specMACS FOV, though. The Lane et al. (2022) study prioritized pixels near the image center and predicted higher errors in focus and polarization measurement at large AOI. Since a single-camera specMACS FOV is decently large (~45 deg nadir-to-aft), and the Cinegon lens does not seem to be telecentric (from the spec), there will likely be AOI-related differences in the transfer matrix at wider angles. I am glad this is recognized by the authors in the discussion towards the end of the paper, but I would reword these statements to differentiate specMACS from the Lane et al. (2022) study a bit more here.

Line 320: I strongly recommend to add a figure that visually explains these three reference frames. It is difficult to reconcile them from the text alone and the following paragraphs require the reader to fully understand each one.

Line 359: The reconstruction error on I of 10^{-13}% is incredibly small versus the error on Q. Even with normalized intensities, I would still expect to see a reconstruction error in the ballpark of what is reported for Q. This suggests to me that the derivation of the transfer matrix is weighting the I inputs more strongly than Q or U. Can you give more details on how that value was derived?

Line 365: Why is it useful to know that the specMACS pol cameras would be between 3 and 5% biased, if they were used uncalibrated while imaging a target with DOLP = 1? Most scientists will never use uncalibrated specMACS data – maybe this is a marker of how close the instrument is to an ideal calibration already? Either way, I suggest changing this to how much error we could expect to see in a calibrated specMACS DOLP measurement (or defer this to Table 3 – see comment below).

Line 390: Systematic and spatial differences between model and measurement in Figures 10 and 11 on the order of 2-6% are quite large for a flatfield residual. This may impact science retrievals done in specific pixel regions – was there any reason not to trust the spatial distribution of the LIS field outright? Integrating spheres should be excellent spatial sources for flatfield.

The other way to approach this could be to step the specMACS field of view across the LIS aperture while taking images. This would place the LIS aperture in different locations of the FOV and fully cover the FPA in a “composite” flatfield over all images taken. Was a test like this considered? I am not requesting extra work, but for this section, I would add more details about why a model was preferred despite significant spatial residuals in Figures 10 and 11.

Line 435: I recommend including a table that lists the sigma errors for each of the terms in Eq. (22), for each wavelength – or if some are functions, give the functional form. Much of this data is already given throughout the paper, but a summary table is preferable.

Table 3. Can you also provide the uncertainty for DOLP? This is a benchmark used to gauge the overall polarization accuracy of a multi-angle polarimeter. This may take further propagation of Eq. (22), but it is also important to show (especially relative to typical atmospheric signals).