Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment (PODEX)

Kirk Knobelspiesse[1*], **Qian Tan**[2,3], **Carol Bruegge**[4], **Brian Cairns**[5], **Jacek Chowdhary**[5,6], **Bastiaan van Diedenhoven**[5,6], **David Diner**[4], **Richard Ferrare**[7], **Gerard van Harten**[4], **Veljko Jovanovic**[4], **Matteo Ottaviani**[5,8], **Jens Redemann**[9], **Felix Seidel**[4], and **Kenneth Sinclair**[5,6]

[1] *NASA Goddard Space Flight Center, Greenbelt, MD, USA*
[2] *NASA Ames Research Center, Moffett Field, CA, USA*
[3] *Bay Area Environmental Research Institute, Petaluma, CA, USA*
[4] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA*
[5] *NASA Goddard Institute for Space Studies, New York, NY, USA*
[6] *Columbia University, New York, NY, USA*
[7] *NASA Langley Research Center, Hampton, VA, USA*
[8] *SciSpaceLLC, Bethesda, MD, USA*
[9] *University of Oklahoma, Norman, OK, USA*
[*] *Corresponding author: kirk.d.knobelspiesse@nasa.gov*

Compiled October 31, 2018

In early 2013, three airborne polarimeters were flown on the high altitude NASA ER-2 aircraft in California for the Polarimeter Definition Experiment (PODEX). PODEX supported the pre-formulation NASA Aerosol-Cloud-Ecosystem (ACE) mission, which calls for an imaging polarimeter in polar orbit (among other instruments) for the remote sensing of aerosols, oceans and clouds. Several polarimeter concepts exist as airborne prototypes, some of which were deployed during PODEX as a capabilities test. Two of those instruments to date have successfully produced Level 1 (georegistered, calibrated radiance and polarization) data from that campaign: the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) and the Research Scanning Polarimeter (RSP). We compared georegistered observations of a variety of scene types by these instruments to test if Level 1 products agree within stated uncertainties. Initial comparisons found radiometric agreement, but polarimetric biases beyond measurement uncertainties. After subsequent updates to calibration, georegistration, and the measurement uncertainty models, observations from the instruments now largely agree within stated uncertainties. However, the 470nm reflectance channels have a roughly +6% bias of AirMSPI relative to RSP, beyond expected measurement uncertainties. We also find that observations of dark (ocean) scenes, where polarimetric uncertainty is expected to be largest, do not agree within stated polarimetric uncertainties. Otherwise, AirMSPI and RSP observations are consistent within measurement uncertainty expectations, providing credibility for subsequent creation of Level 2 (geophysical product) data from these instruments, and comparison thereof. The techniques used in this work can also form a methodological basis for other intercomparisons, such as of the data gathered during the recent Aerosol Characterization from Polarimeter and Lidar (ACEPOL) field campaign, carried out in October and November of 2017 with four polarimeters (including AirMSPI and RSP).

http://dx.doi.org/10.1364/ao.XX.XXXXXX

---

1. INTRODUCTION

Aerosols and their interactions with clouds are a large source of uncertainty in our understanding of climate ([1]), due, in part, to inadequate global observations. While single/fixed viewing angle multispectral imagers such as the Moderate Resolution Imaging Spectroradiometer (MODIS, [2]) have brought us into the modern era of aerosol remote sensing, instruments capable of measuring more information (for example additional viewing angles and sensitivity to polarization) about a scene are required to meet the needs of long term global aerosol monitoring ([3–5]). Multi-angle polarimeters are desirable for such purposes because they add multiple views and polarization sensitivity, which when combined with spectral information, can be used to determine aerosol quantity, size distribution, and complex refractive index, (e.g. [6–8]) and cloud phase, particle size distribution, ice crystal shape, and optical depth (e.g. [9–15]).

For these reasons, the 2007 Earth Science Decadal Survey from the US National Research Council identified the study of aerosols, clouds and ecosystems and their interactions as an imperative field of study within earth system science and recommended the creation of a NASA mission devoted to Aerosol, Cloud and Ecosystem (ACE) observations to address this challenge. The Earth Science Decadal Survey determined an imaging, multi-angle polarimeter is needed to achieve these goals ([16]). NASA commissioned a science working group to help formulate science objectives, and to explore the current and potential future capabilities of such measurement systems ([17]). Passive imaging polarimetry encompasses a diverse set of optical analysis techniques, which vary greatly in their respective complexity, spatial resolution, observation geometry, spectral capability, and measurement uncertainty ([18]). For this reason, NASA also funded an airborne field experiment that deployed several multi-angle polarimeter prototypes on the high altitude ER-2 aircraft. This field campaign, called the POlarimeter Definition Experiment (PODEX), was intended to further polarimeter design by comparing instrument observations, and the aerosol and cloud optical properties derived from such observations. The scope of this paper is the former: an assessment of Level 1 (calibrated and georegistered radiance and polarization) observations. Several publications thus far have produced Level 2 (geophysical parameters derived from Level 1 data) results ([11, 15, 19–27]), but direct instrument comparisons at Level 2 are limited to liquid cloud properties during the NASA ORACLES (ObseRvations of Aerosols above CLouds and their intEractionS) field campaign ([28]). If geophysical product differences are indeed found between instruments, a Level 1 evaluation such as this will be necessary to help identify if the cause is inherent instrument observation differences, or the ability of retrieval algorithms.

PODEX was conducted during January and February, 2013, from the Armstrong Flight Research Center (AFRC) in Palmdale, California. Three polarimeters were deployed on the NASA ER-2 (tail number 809) aircraft: the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI, [22]), the Research Scanning Polarimeter (RSP, [29, 30]), and the Passive Aerosol and Cloud Suite (PACS). Additional sensors on the ER-2 included the Autonomous Modular Sensor (AMS), which provided multiwavelength calibrated radiances and cloud products generated using MODIS algorithms, the Cloud Physics Lidar (CPL, [31]), which provided real-time and post flight aerosol and cloud backscatter profiles to locate and identify aerosol and cloud layers, and the Solar Spectral Flux Radiometer (SSFR, [32]), which provided spectrally resolved shortwave upwelling and downwelling irradiance measurements. The ER-2 flights conducted during PODEX were coordinated with airborne and ground-based measurements acquired during the third deployment of the Deriving Information on Surface conditions from Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) mission. DISCOVER-AQ used the NASA P-3 and NASA Langley Research Center (LaRC) King Air aircraft to study air quality over the California San Joaquin Valley during this period. The NASA P-3 aircraft was equipped with several in situ sensors that measured trace gas concentrations and aerosol optical (scattering, absorption) and microphysical (size, composition) properties. In particular, the PACS group have also developed the Polarized Imaging Nephelometer (the PI-Neph, [33]) for the detailed measurement of the P11 and P12 elements of the scattering matrix of the aerosol particles, which can be directly compared to the polarimetric retrievals of the PACS, AirMSPI and RSP sensors (Dolgos and Martins 2014). The King Air deployed the LaRC High Spectral Resolution Lidar-2 (HSRL-2, [34]), which is a prototype of the multiwavelength lidar called for by ACE to provide layer-resolved retrievals of aerosol optical and microphysical retrievals. The Distributed Regional Aerosol Gridded Observation Network (DRAGON) of AERONET ([35], [36]) sun-sky photometers was also deployed in the southern part of the San Joaquin Valley and provided measurements of aerosol optical depth (AOD) and retrievals of column averaged aerosol optical and microphysical properties.

During PODEX, the ER-2 acquired 49 hours of science data during 10 flights between January 14 and February 6, 2013. The flights were designed so that the polarimeters acquired data over bright (desert, snow, clouds) and dark (ocean) scenes, during light and moderate aerosol loading conditions in maritime, rural and urban regions, and over fog, stratus, stratocumulus, and cirrus clouds. Data were also acquired over the calibration targets located at Rosamond Dry Lake, Ivanpah, and Railroad Valley (dry lakes in California and Nevada). The flights over the San Joaquin Valley contained several legs above the DRAGON AERONET sensors and were coordinated with the DISCOVER-AQ aircraft to obtain simultaneous measurements of aerosol optical and microphysical properties. DISCOVER-AQ also conducted flights over the ocean to support the PODEX flights. The PODEX flights were executed as planned, with the exception of the flight on January 28 when RSP lost operation of the SWIR (Short Wave Infrared) bands due to operator error. This also prevented the operation of these SWIR bands on subsequent PODEX flights. Post mission repairs and calibration showed that the visible channels were not affected.

2. INSTRUMENTS

Three polarimeters were deployed on the NASA ER-2 for the PODEX field campaign. Two of those instruments (AirMSPI and RSP) ultimately produced calibrated and georegistered radiometric data that were archived at their respective data repositories (see Data Availability section). The third instrument, PACS, suffered detector nonlinearity problems, and was not able to archive data. In this paper, we are therefore restricted to a comparison of AirMSPI and RSP observations. Table 1 provides an overview of the characteristics of these instruments.

A. Units

Although polarimeters use different techniques to measure the polarization state, they all can produce a common set of observable quantities. In this analysis, we use reflectances, defined as

$$R_I = \frac{I\pi r^2}{F_o cos(\theta_s)}; \quad R_Q = \frac{Q\pi r^2}{F_o cos(\theta_s)}; \quad R_U = \frac{U\pi r^2}{F_o cos(\theta_s)} \tag{1}$$

where I, Q and U are the first three elements of the Stokes polarization vector observed by the instrument. Thus, R_I indicates 'total' reflectance, equivalent to what non-polarimetric

Table 1. Instrument characteristics for the polarimeters used during the PODEX field campaign. Note AirMSPI has two targeting modes. The "Sweep" (abbreviated *Sw*) mode scans a wide along-track field of view, and provides multiangular views of extended scenes in which the different view angles are not spatially co-registered. The "Step and Stare" (abbreviated *SaS*) mode observes a shorter along-track target at multiple spatially co-registered viewing angles, typically 9 or 15 during PODEX. Both modes are mapped to different ground spatial resolutions, and have subsequently different measurement uncertainties. Spatial resolutions are reported for nadir views from a typical ER-2 aircraft altitude (19km). **Boldface** for band centers indicates polarization sensitivity

Instrument	Polarimetric analysis method	Imager?	Polarimetric uncertainty	Number of view angles per pixel	Nadir ground resolution	Band center wavelengths, nm	observations per pixel
AirMSPI	Photoelastic modulation	yes	*Sw*: 0.3% *SaS*: 0.5%	*Sw*: 1 *SaS*: 9-15	*Sw*: 25m *SaS*: 10m	355, 380, 445, **470**, 555, **660**, **865**, 935	up to 210
PACS	Philips prism, linear polarizers	yes	unknown	65	37m	**470**, **550** , **670**, **766**, **870**	975
RSP	Wollaston prisms	no	0.15%	155	277m	**410**, **470**, **555**, **670**, **865**, **960**, **1590**, **1880**, **2250**	4,185

radiometers are intended to observe, while R_Q and R_U indicate the direction (with respect to a predefined reference plane) and magnitude of linear polarization. Reflectance units are chosen to normalize observations by solar position and spectral variability in solar irradiance from the observed I, Q and U (which have units of $Wm^{-2}sr$). F_o is the average annual exo-atmospheric irradiance (Wm^{-2}), r is the earth-solar distance normalized by the annual mean value, and π relates radiant intensity to irradiance.

R_Q and R_U can be combined into a reference frame insensitive unit, namely the polarized reflectance: $R_p = \sqrt{R_Q^2 + R_U^2}$. One can then compute the fraction of light that is linearly polarized, the Degree of Linear Polarization ($DoLP$):

$$DoLP = \frac{R_p}{R_I} = \frac{\sqrt{R_Q^2 + R_U^2}}{R_I} = \frac{\sqrt{Q^2 + U^2}}{I}. \quad \textbf{(2)}$$

Thus, $DoLP$ is the fraction of linearly polarized to total intensity, and is unitless. This is the unit used to define required accuracy (0.005, alternatively 0.5% in reflectance units) of a polarimetric imager in the NASA ACE 2010 report ([17]). It also has the advantage of removing the impact of polarimetric reference plane in the comparison, which are highly sensitive to uncertainties in aircraft and sensor attitude. Furthermore, all three instruments are designed to meet uncertainty thresholds expressed in $DoLP$, and have (with the exception of PACS) corresponding uncertainty models for this parameter.

B. The Airborne Multiangle SpectroPolarimetric Imager (AirMSPI)

The AirMSPI is a gimbaled polarimetric imager that is a precursor to the NASA Earth Venture class Multi-Angle Imager for Aerosols (MAIA) instrument, currently in development. It uses dual photoelastic modulators (PEMs) to analyze the polarimetric state, and is thus sensitive to the first three elements of the Stokes polarization vector (I, Q and U). The temporal modulation approach encodes both Q and I (or U and I) on the same detector pixel, such that the ratios Q/I and U/I and hence $DoLP$ are insensitive to pixel-to-pixel response differences. AirMSPI has eight spectral bandpass channels with centers between 355nm and 935nm. Three of those channels (470, 660 and 865nm) have polarization sensitivity, while the others are sensitive to intensity alone. Since AirMSPI views the scene at one viewing angle at a time, multiple view angle observations are made by the use of a gimbal. For this reason, AirMSPI does not continuously observe beneath the aircraft with multiple view angles. Instead, it employs two targeting modes. "Sweep" (abbreviated by us as *Sw*) continuously scans along a wide forward and aft field of view (±67°), and provides a single view of each ground pixel. "Step and Stare" (abbreviated by us as *SaS*) makes repeated observations centered about a specified location. *Sw* mode observations are gridded to a 25m pixel size on the ground, while *SaS* observations are gridded to 10m. The PODEX field campaign, which was the first scientific deployment of AirMSPI ([37]), employed both targeting modes. *SaS* targeting mode observations had between 9 and 15 viewing angles, depending on the scene. Since multi-angle observations are likely to be the primary targeting mode used in the retrieval of atmospheric properties ([26, 27]), we only compare RSP observations to AirMSPI *SaS* observations, and for those we compare the observation image containing the nadir view.

The complete polarimetric response of AirMSPI is calibrated at the Jet Propulsion Laboratory (JPL) using a rotating wire grid polarizer, and validated with ±0.001 uncertainty using a number of well-characterized partial polarizers with different $DoLP$ values between 0 and 1. Additionally, an onboard polarized validation light source is viewed throughout the flight to calibrate thermal sensitivity of the PEMs for a roughly 0.001 residual impact on the $DoLP$ uncertainty (depending on $DoLP$ and wavelength). This calibration does not require a priori knowledge of the validator output $DoLP$ and its stability, which are therefore not designed to be reliable. However, as a worst-case limit for the AirMSPI performance, the $DoLP$ as measured as a byproduct of the validator views is stable to within roughly 0.001 in flight, whereas the systematic difference between flight and validator views on the ground is 0.0003 and 0.005 at 470 and 660 nm, respectively (results at 865 nm strongly indicate that the observed variations are dominated by instability of the validator DoLP). Further AirMSPI details can be found in [38], which also includes a comparison between AirMSPI and RSP. That comparison makes use of 112 AirMSPI images over land at multiple view angles, acquired during the PODEX and Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS) field campaigns. The observed root-mean-squared DoLP differences ranged between 0.003 and 0.006, while the regression slopes deviated from unity by values up to 0.024. In contrast to the current study, interpretation of the per-pixel differences in terms of the instrument uncertainty models was beyond the scope of that paper.

The radiometric calibration of AirMSPI is established through both laboratory and vicarious experiments, for pixel-relative response determination and to set the channel-average absolute scale, respectively. The latter methodology makes use of in-situ surface and atmospheric measurements to compute the sensor-incident radiance. Traceability to Systeme Internationale (SI) units is through a vendor-calibrated Spectralon reflectance standard, plus a database of the exo-atmospheric irradiance ([39]). The laboratory work at JPL makes use of a 1.6m (65 inch) barium sulfate integrating sphere. The sphere provides uniform radiance across the swath, thus is best used for the "flat-field" (i.e. pixel-relative) response determination. Its output is established through reference to a Gooch and Housego vendor calibrated 6" radiance output standard. Although the absolute output radiance from the large sphere is not used in the final AirMSPI coefficient delivery, the channel-average responses of AirMSPI between the laboratory and vicarious calibrations agree within 5% for the 470, 660 and 865 nm bands used in this paper. Because the commercial sphere standard has no reported uncertainties other than at 550 nm, the vicarious calibration results are selected over the laboratory for the absolute scale. This radiometric scale, as established through the vicarious calibration experiment, has also been validated by cross-comparison with AVIRIS and MODIS-Terra. Agreement with these is within 3% for the 470-865 nm spectral range.

C. The Research Scanning Polarimeter (RSP)

The RSP is a multi-angle polarimeter that continuously scans in the aircraft along-track direction, providing 152 views of each ground location at angles between 60° forward and aft of nadir when not vignetted by aircraft structure. It uses six bore sighted telescopes and a polarization compensated scan mirror assembly to observe the scene at nine visible and SWIR wavelength spectral channels (see Table 1). The six telescopes are arranged in three pairs, with one telescope in each pair measuring the Stokes vector elements I and Q in three spectral bands. The other telescope measures I and U in the same three spectral bands, providing redundant intensity measurements for each spectral band. In each telescope, the Stokes vector elements I and Q (or I and U) are determined by using a Wollaston prism to split the incoming intensity into polarimetrically orthogonal components measured by pairs of detectors. In fact, it is not possible to ensure that the pairs of telescopes are perfectly oriented to measure I, Q and U, so the exact orientation of the Wollaston polarizers and any cross talk between orthogonal polarization states needs to be characterized prior to field missions. The polarimetric response of RSP was calibrated in the laboratory at NASA Ames Research Center (ARC) prior to the PODEX field campaign using a rotating wire grid polarizer that has previously been calibrated against a Glan-Taylor polarizer at NASA Goddard Space Flight Center. The relative orientation of the Wollaston prisms used to measure Q and U is determined to within 0.01° and cross-talk is characterized such that uncertainties in $DoLP$ caused by this are less than 0.0005, with stability in the magnitude of cross-talk over the last decade being roughly 0.001. The polarimetric calibration and characterization is validated against a tilted glass plate that provides an intermediate value of DoLP ([30]). As noted above, orthogonal polarization components are measured by pairs of detectors and untracked variations in the relative gain of the detectors could therefore create false polarization. An inflight calibrator is therefore used to provide a reference with near zero polarization to allow the relative gain of the detectors observing orthogonal polarizations to be tracked continuously in flight. The inflight calibrator consists of viewing the nadir scene via a relay mirror and a depolarizer and is observed on every scan (every 0.84 seconds). It should be noted that while the relative gain of the detectors measuring orthogonal polarizations is tracked continuously, it is not only stable from scan to scan, but also over many years with variations of less than 0.1% over 4.5 years (see Table 2).

In order to obtain accurate radiometric observations from a sensor it is necessary that the dark, or zero, radiance value, linearity, and radiometric gain are accurately known. The dark values for the RSP are determined using nine views of a dark reference after the signal chain has been reset. The uncertainties in the determination of dark level are negligibly small and, in all cases, less than 10^{-5} in units of normalized radiance. The RSP also has a dark measurement prior to the signal chain reset that is used to track drifts in the offset within a scan. The scan period is 0.84 seconds, and the amount of drift is currently negligible with a worst-case drift of less than 0.05 digital numbers (DN) and typical drifts of 0.005 DN that can only be detected by averaging over long periods (i.e. several hours). The radiometric gain of RSP for PODEX was established through laboratory calibration using a 30" high-output integrating sphere source at NASA ARC that has its radiance scale traceable to the standards for radiance and irradiance established by NIST. The calibration uses between four and seven different radiance levels to establish the linearity of the detectors and consistency of the radiometric performance over a wide dynamic range. In addition, views into the sphere both normal to its aperture and at ±4° from normal are used to assess the potential effects of heterogeneity on the radiometric calibration, which are non-negligible at very low light levels and are included in the assessed uncertainty of the calibration. In order to calculate the normalized radiance, a solar spectrum is required (Equation 1) and for that purpose the spectrum from [40] is used, which allows for historical variations in solar irradiance, including those over the solar cycle, to be incorporated. The solar spectrally weighted irradiance used in calculating the normalized radiances is included in all RSP L1B data files. The RSP is operated at close to room temperature (20°C) on the NASA ER-2 aircraft through the use of thermal insulation and heaters, and the calibration established in the laboratory at around 20°C is therefore expected to be valid in flight. Over the last 4.5 years, radiometric calibrations of the RSP in the laboratory at NASA ARC have been stable to within 1% (see Table 3). Radiometric comparisons with AVIRIS on 31 March 2014 show differences in radiometric gain of 2.5, 4.5 and 1% for the 470, 670 and 865 nm bands of interest for this paper ([41]).

In summary, the RSP is able to determine linear polarization very accurately (< 0.2%, see Appendix B for full model) and is an accurate and stable radiometer. The RSP served as the airborne prototype for the Aerosol Polarimetry Sensor (APS) on the NASA Glory Mission ([4], [42], [43]). While Glory did not successfully enter orbit due to a launch failure, a pair of RSP

Table 2. RSP relative gain coefficients of the detectors measuring orthogonal polarizations in telescopes measuring I and Q (K1) and in telescopes measuring I and U (K2).

Date	K1(470nm)	K1(670nm)	K1(865nm)	K2(470nm)	K2(670nm)	K2(865nm)
5/20/12	0.97704	0.98149	1.00105	0.97742	1.01994	0.98632
5/8/13	0.97599	0.98108	1.00157	0.97551	1.02034	0.98500
5/2/16	0.97511	0.98116	1.00117	0.97724	1.02141	0.98498
12/15/16	0.97516	0.98115	1.00094	0.97772	1.02126	0.98499
Mean	0.97583	0.98122	1.00118	0.97697	1.02074	0.98532
Standard Deviation	0.09%	0.02%	0.03%	0.10%	0.07%	0.07%

Table 3. RSP radiometric calibration coefficients in Digital Numbers pre- and post-PODEX and since that field campaign for telescopes measuring I and Q (Cal1) and telescopes measuring I and U (Cal2).

Date	Cal1(470nm)	Cal1(670nm)	Cal1(865nm)	Cal2(470nm)	Cal2(670nm)	Cal2(865nm)
5/20/12	17254.9	18106.5	20287.6	17426.2	19328.2	19100.6
5/8/13	17138.5	18013.8	20133.1	17303.6	19226.2	18971.3
5/2/16	17189.9	18122.5	20347.0	17340.8	19307.6	19164.9
12/15/16	16909.7	18124.2	20431.7	17064.8	19318.3	19250.6
Mean	17123.2	18091.8	20299.8	17283.9	19295.1	19121.8
Standard Deviation	150.1	52.5	125.9	154.8	46.7	117.7
Relative Standard Deviation	0.9%	0.3%	0.6%	0.9%	0.2%	0.6%

instruments (denoted RSP1 and RSP2, the latter was used in PODEX) continue to make robust observations and have been deployed on over 25 field missions in the last twenty years. As an along-track scanner, RSP does not fulfill the requirement for an imaging polarimeter for ACE. However, its heritage, accuracy, and comprehensive measurements characteristics mean that it is well suited for observations of clouds ([10–12, 19], [15, 21], [44]), aerosols ([45–47], [48, 49], [24, 25], [20], [50]), the ocean ([51–54]) and snow ([23]).

D. Measurement uncertainty

Radiometric and polarimetric measurement uncertainty has been characterized for both AirMSPI and RSP. These uncertainty models, summarized in Appendix A and B, are functions of scene specific parameters, such as solar zenith angle, and observation reflectance and DoLP. The appendices are based upon literature describing characterization, calibration and measurement uncertainty for AirMSPI ([38]) and RSP ([55], [29], [41], [50]). Since the latter is the airborne prototype of the APS, formulation of the RSP uncertainty model is similar to what is described for Glory/APS ([42], [43]). As described above, AirMSPI is calibrated using facilities at JPL, including a vicarious adjustment from reflectances using overflights of uniform surfaces that have been characterized on the ground. RSP2 is calibrated at the NASA Ames Research Center, whose facilities trace to the calibration facility at the NASA Goddard Space Flight Center (to which RSP1 is calibrated).

Figure 1 shows the uncertainties of the AirMSPI and RSP instruments. Except for very small reflectances, the relative reflectance uncertainty is constant at 5% for AirMSPI and 3% for RSP. $DoLP$ uncertainty, on the other hand, decreases as reflectance increases. Thus, we would expect polarimetric uncertainty to be better for bright cloud scenes than for dark oceans, for example. We have also indicated two sets of AirMSPI uncertainties. Dotted lines indicate the measurement uncertainty at the 10m gridded SaS spatial resolution, while dashed lines represent the AirMSPI uncertainty when averaged to the much larger RSP footprint (see Section B). Reflectance uncertainty is primarily dominated by calibration and other systematic effects, so spatial resolution has almost no impact. Thus, SaS and RSP footprint AirMSPI uncertainties overlay in the leftmost plot in Figure 1. $DoLP$ uncertainty is dominated by shot noise, and can therefore be reduced by averaging (accounted for by the m and n parameters in equation 5). For the SaS native spatial resolution, AirMSPI $DoLP$ uncertainty is larger than that of RSP, and above the 0.5% requirement in ACE Mission study report ([17]) for all but the brightest scenes. When averaged to RSP spatial resolution, AirMSPI $DoLP$ uncertainty meets ACE requirements, becomes less than that of RSP, and is primarily defined by systematic uncertainty. Note also that the ACE Mission study report calls for a 3% reflectance measurement uncertainty. However, reflectance uncertainty is primarily driven by the ground calibration facility and not instrument design (see Appendix A and B). Meeting the ACE requirement for reflectance is best determined by assessments of calibration facilities.

Retrieval algorithms that utilize polarimeters such as AirMSPI and RSP must incorporate data with dramatically varying uncertainties profiles. Reflectance uncertainties are largely driven by systematic uncertainties from calibration, in the 3 to 5% range. $DoLP$, on the other hand, can be an order of magnitude smaller, and has a much larger fractional contribution of random noise (such as shot noise) to the overall uncertainty.

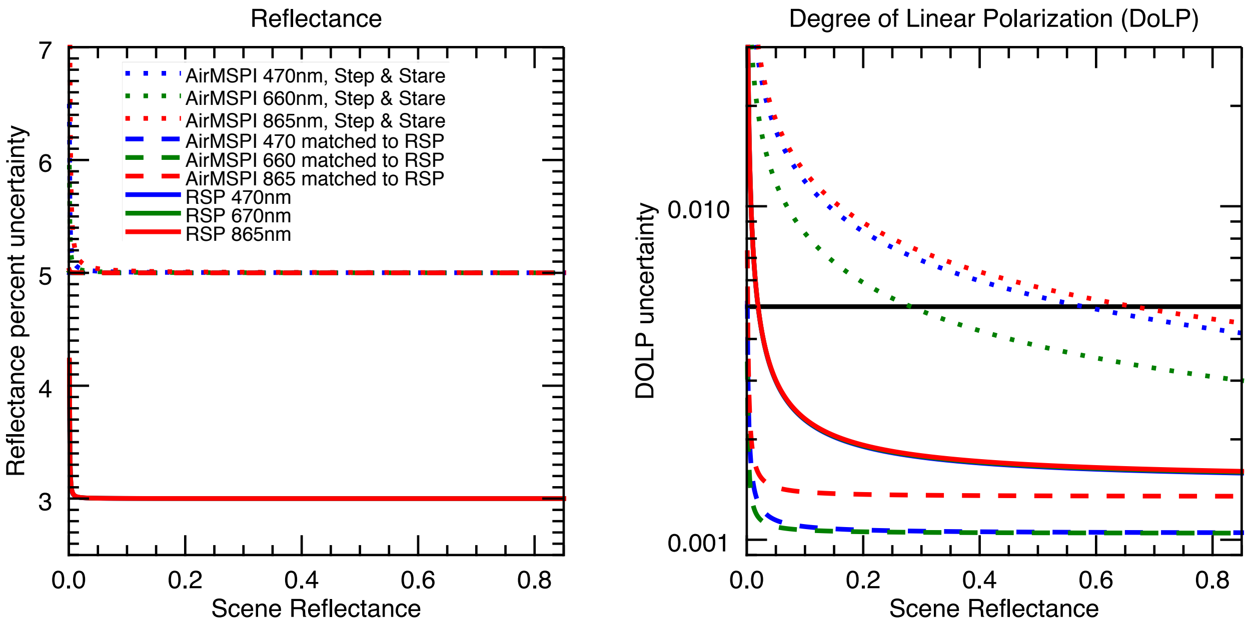

Fig. 1. AirMSPI (dashed and dotted) and RSP (solid) uncertainty expectations for reflectance (left, in percent uncertainty) and DoLP (right). Values were calculated for a solar zenith angle of 45° and DoLP=0.3 and a variety of scene reflectances (abscissa). The uncertainty model in Appendix A provides the means to determine AirMSPI measurement uncertainty at the desired spatial resolution, which is relevant to DoLP uncertainty. Thus, we have presented DoLP uncertainties for observations for the 10m gridded AirMSPI *Step and Stare* mode (dotted lines) and for AirMSPI observations downsampled to the larger RSP footprint (dashed lines). AirMSPI reflectance uncertainty is dominated by systematic error, so averaging provides no beneficial reduction in measurement uncertainty. On the other hand, DoLP uncertainty is dominated by random errors, which are reduced by spatial averaging, hence the reduction in DoLP uncertainty for downsampled AirMSPI compared to *Step and Stare* mode. The solid black line at right indicates the Decadal Survey ACE requirement for polarimetric uncertainty, 0.5%.

Figure 2 shows the fractional contribution of random noise to total uncertainty for both instruments at 660 and 670nm. For $DoLP$, scene reflectance determines the fraction of total uncertainty due to random noise, such that dark scenes have higher uncertainty and a larger contribution from random noise than bright scenes. As we shall see in subsequent sections, the agreement between instruments is different in bright and dark scenes, and this difference may be due to the nature (and accuracy) of these uncertainty models.

Both instrument uncertainty models depend slightly on scene $DoLP$ for the determination of $DoLP$ uncertainty (see Equation 11 in Appendix A and Equation 17 in Appendix B). In Figures 1 and 2, the reference $DoLP$ was chosen as 0.3. In most cases, different $DoLP$ values had very minimal impact on total DoLP uncertainty. However, because of the larger assessed uncertainty associated with the in-flight stability monitor for AirMSPI at 865nm, $DoLP$ uncertainty for that channel has a moderate dependence on scene $DoLP$. For example, $DoLP(865)$ uncertainty for an unpolarized scene is roughly 0.001, while for a highly polarized scene ($DoLP = 0.5$) the uncertainty is nearly double at 0.0019. This is relevant for AirMSPI data that have been averaged to match the RSP footprint. For higher spatial resolution scenes (such as the *Step and Stare* mode data), scene $DoLP$ is less relevant because random uncertainty dominates total uncertainty. Since the expression of scene $DoLP$ on the uncertainty model affects the systematic uncertainty component, the fraction of random to total uncertainty (Figure 2) decreases slightly with increasing $DoLP$.

Uncertainty model fidelity is required for successful retrieval of geophysical parameters from the observations. For example, [50] uses a functionally similar model to that in Appendix B in a cost function for the optimal estimation aerosol and ocean property retrieval algorithm. [27] does something similar for AirMSPI, where the uncertainty model is a component of a weighting function for the retrieval of aerosol and land surface properties. For our purposes, the uncertainty model is required to identify how close we expect an AirMSPI / RSP measurement pair to be. As we shall see, the scene dependent nature of the instrument uncertainty models limits the value of a direct regression between measurement pairs, rather, comparisons must be made in the context of expectations of measurement uncertainty. In this sense, we are validating agreement considering both the uncertainty models and the observations.

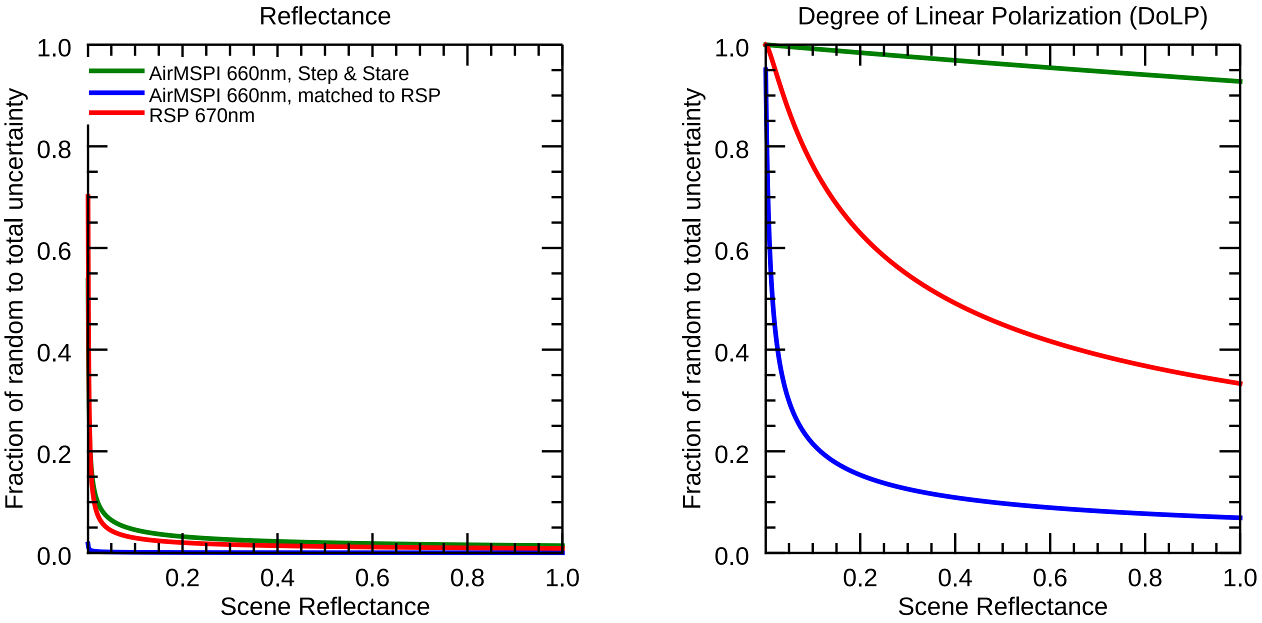

Fig. 2. The fraction of the AirMSPI and RSP uncertainty models, at 660 and 670nm, respectively, contributed by random noise. The remainder is due to systematic uncertainties such as those that arise from calibration. Thus, reflectances observed by both instruments are dominated by systematic uncertainties, while $DoLP$ uncertainties have a much larger contribution from random uncertainties, the degree to which depends on the scene reflectance. Note that these figures were calculated for a scene with $DoLP = 0.3$. The fraction of random to total polarimetric uncertainty decreases slightly as scene $DoLP$ increases.

E. Instrument spectral response

AirMSPI and RSP share three channels with polarization sensitivity, roughly centered at 470nm, 660 or 670nm, and 865nm. Normalized spectral responses are shown in Figure 3. The RSP Full Width at Half Minimum (FWHM) is about half as wide as that of AirMSPI, so the RSP spectral responses fall nearly entirely within that of AirMSPI. We do not expect the types of scenes employed in our intercomparison (clouds, cloud free oceans, desert and snow) to have strong spectral gradients within the spectral response functions of the instruments, so we compare AirMSPI and RSP observations without adjustment. As we shall see, the 660/670nm channels, which have the greatest potential for spectral differences due to slight water vapor absorption and difference between the spectral response band centers, have the best agreement of all three channels.

3. INTERCOMPARISON METHODOLOGY

The AirMSPI and RSP instruments have different spatial resolutions and measurement geometries, and these differences must be accounted for in a point-to-point comparison of the data. In order to identify gridded AirMSPI data within the larger RSP footprint, and to process those data for comparison, we use the following steps.

1. Identify RSP observations of sufficient quality as described below for comparison, and aggregate the data such that the multi-angle observations are defined with respect to the ground elevation or ocean surface. We screen to only include data that have not been flagged with instrumental errors, that were made during straight and level ER-2 flight, and are not at boundaries between surface types or atmospheric conditions.

2. For each RSP observation, identify the latitude, longitude, and view zenith angle (θ_v) closest to nadir.

3. Build a spatial mask (see Figure 4) in AirMSPI coordinate space centered about the RSP observation latitude and longitude. This mask has its maximum value in a circle with radius 138.5m, the projection of the RSP 14 mrad Instantaneous Field of View (IFOV) at the ground from an average ER-2 altitude of 19.7km. This IFOV moves the equivalent of 14 mrad due to scanner rotation during capture, resulting in a 'smear' of sensitivity in the along-track direction. This is represented as the oval of linearly decreasing value whose long axis is aligned with the flight track. In practice, three spatial masks, one for each spectral channel, are created.

4. Apply the spatial mask to the nadir view AirMSPI data to determine the averaged R_I and $DoLP$. Additionally, apply the mask to AirMSPI geometry fields to identify a mean θ_v within the RSP footprint.

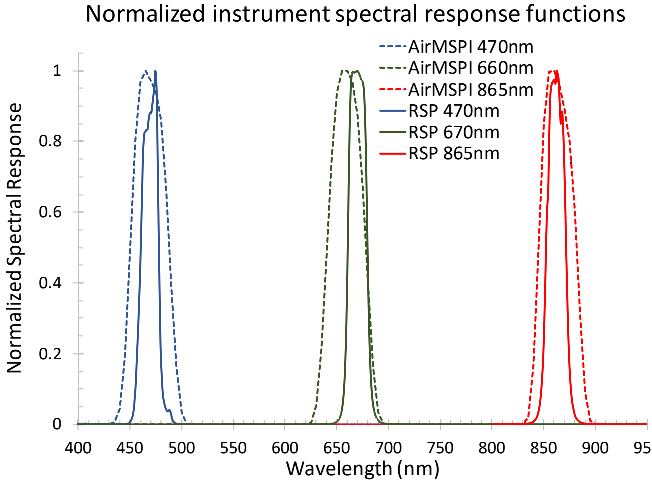

Fig. 3. Spectral responses of the common polarization sensitive channels of the AirMSPI and RSP instruments.

5. Select the RSP R_I and $DoLP$ whose θ_v is closest to that of the mean AirMSPI θ_v within the footprint.

6. Calculate the expected measurement uncertainty for these AirMSPI and RSP values using models in Appendix A and Appendix B.

This procedure will find the georegistered match between RSP and the nadir centered AirMSPI observations, provided knowledge of measurement geometry and sensor IFOV are correct. While significant effort has been put into ensuring this is the case (e.g. [41], [22]), this category of error is separate from the radiometric accuracy we are attempting to verify. For this reason we have been extremely conservative in our selection of comparison data (see Section A). We should also note that the RSP coordinates associated with the selection in step 5 could be slightly different than those identified for the nadir view in step 2. Steps 2 through 5 could be iterated to refine this process. We initially attempted this, but found the amount of adjustment was too small to justify the additional complexity.

Analysis is limited to (near) nadir observations for several reasons. A large concern was the ability to accurately geolocate AirMSPI and RSP observations. This geolocation becomes more difficult as the geometry moves from nadir, so we were afraid that might contribute to the overall assessment of sensor agreement. Additionally, the degree to which geolocation contributes to intercomparison, if it does, would vary with viewing angle, confusing any conclusions that could be drawn from such analysis. Furthermore, observations can not be considered uniquely independent if they include nadir and off nadir views of the same location, and such independence is the basis of some of the statistical techniques we describe later. Thus, considering that the uncertainty models do not include explicitly include viewing angle in their characterization, we decided that the intercomparison should be performed simply for the nadir observations.

A. Scene selection

Spatial homogeneity, in order to minimize geolocation uncertainty, was the primary selection criterion. A total of 636 RSP observations within eleven AirMSPI nadir files were chosen for analysis. Three categories of scenes were chosen, including cloudless, spatially homogeneous, land targets (such as Rosamond dry lake or snow fields), cloudless oceans, and uniform marine stratocumulus clouds. Marine stratocumulus clouds are characterized by high R_I and low $DoLP$, cloudless oceans by low R_I and moderate to high $DoLP$, while the land scenes we chose had moderate to high R_I and moderate $DoLP$. We thus span a wide range of R_I and $DoLP$ values and associated uncertainties. Table 4 contains the mean values of R_I and $DoLP$ for the different scenes, while Appendix D has details about the individual files used in this analysis.

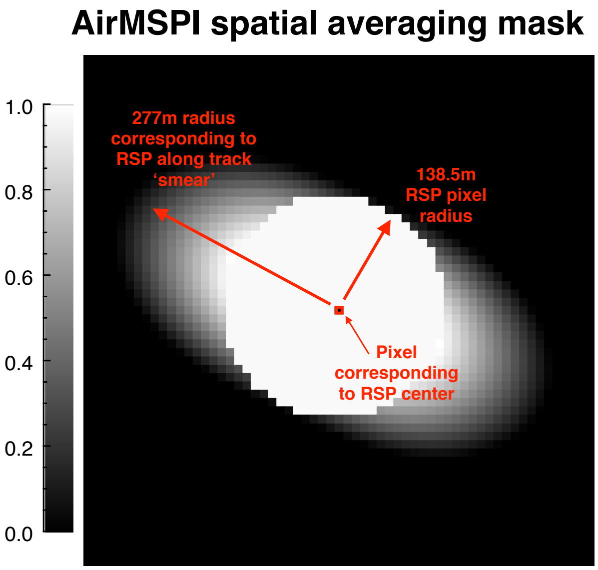

Fig. 4. The spatial mask used to average AirMSPI data for comparison to RSP. From the AirMSPI pixel corresponding to the RSP coordinates, a radius of 138.5m is drawn in the cross track direction. This corresponds to the 14 mrad IFOV of RSP at a mean ER-2 altitude of 19.7km. There is an additional 14 mrad 'smear' in the along-track direction, corresponding to scanner rotation in the time required to make a measurement. The mask value is unitless, and is used in a multiplicative manner to weight the AirMSPI observation.

B. Georegistration

Significant effort by both instrument teams was put into georegistration. Connection of each instrument's reference frame to that of the aircraft Inertial Navigation Systems (INS) is dependent upon the location of the instrument on the aircraft. The AirMSPI was located in the nose of the ER-2, while the RSP was located in a wing pod. The latter is subject to wing flex, which varies RSP roll angle throughout a flight as a function of fuel load. In the first iteration of AirMSPI and RSP comparison, RSP georegistration was verified using shoreline crossing flight tracks as a reference. Based upon the experience and input from other ER-2 instrument teams, a strategy for handling wing flex was devised as is described in Appendix C, which is what is used in the currently archived (version 2) RSP data and in this analysis. A description of AirMSPI georegistration can be found in [56].

An example of a matched AirMSPI and RSP scene ("Cloud E" in Appendix D) for a marine stratocumulus cloud is shown in Figure 5. AirMSPI imagery is overlaid with magenta circles

Table 4. Comparison scene characteristics, including mean R_I and $DoLP$ values.

	All	**Land**	**Cloud**	**Ocean**
Number of observations	636	133	322	181
mean AirMSPI R_I(470nm)	0.387	0.316	0.573	0.110
mean RSP R_I(470nm)	0.364	0.303	0.535	0.104
mean AirMSPI DoLP(470nm)	0.182	0.158	0.088	0.368
mean RSP DoLP(470nm)	0.186	0.160	0.089	0.380
mean AirMSPI R_I(660nm)	0.352	0.349	0.533	0.032
mean RSP R_I(660nm)	0.351	0.355	0.529	0.031
mean AirMSPI DoLP(670nm)	0.149	0.052	0.046	0.403
mean RSP DoLP(670nm)	0.151	0.051	0.045	0.414
mean AirMSPI R_I(865nm)	0.352	0.356	0.539	0.015
mean RSP R_I(865nm)	0.361	0.374	0.549	0.015
mean AirMSPI DoLP(865nm)	0.137	0.033	0.034	0.395
mean RSP DoLP(865nm)	0.143	0.032	0.033	0.419

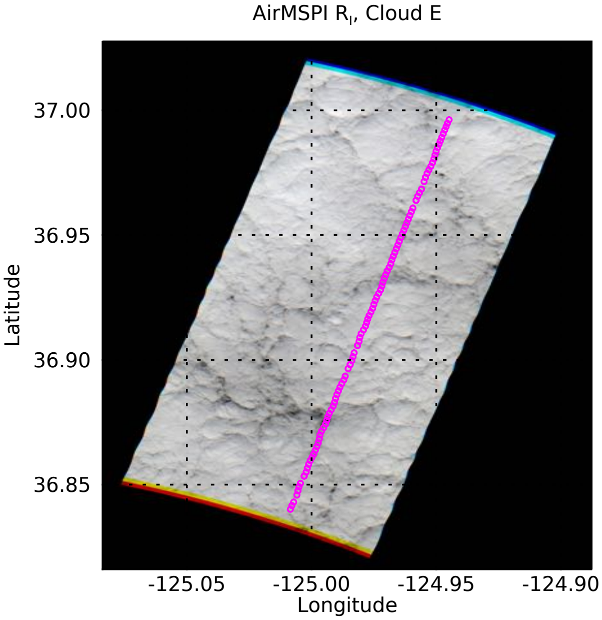

Fig. 5. Example comparison scene ("Cloud E"), of marine stratocumulus clouds over the North Pacific Ocean (center coordinates 36.9189°N, 124.976°W) on February 3rd, 2013. Imagery is from AirMSPI, magenta circles indicate the RSP sample locations. Flight direction was from North to South.

indicating RSP observation locations. The corresponding data, plotted with respect to latitude, is shown in Figure 6. R_I clearly has the greatest spatial variability, and features such as the dip around a latitude of 36.875°N correspond to features in the imagery. As expected for (white) clouds, reflectances at all three wavelengths have similar values. R_p and $DoLP$, on the other hand, have larger values for the shortest wavelengths, expressing increased Rayleigh scattering. R_p and $DoLP$ also have less spatial variability.

All three parameters show a slight shift in latitude, where the RSP data (dashed lines) are biased northward of AirMSPI. While this bias has minimal impact on the comparison statistics shown later, we found that a slight adjustment, corresponding to an aftwards RSP pitch of 0.65°, improved all eleven comparison scenes. The ground distance of such a shift is less than half the RSP IFOV radius. Figure 7 shows the result of this shift, which has been applied to all the scenes compared in this analysis. We should note that such a shift is not an unequivocal indication of incorrect knowledge of RSP instrument pitch, but simply a systematic difference between the geolocation of the two instruments. Figures 6 and 7 also contain information about the differences between the instruments, examined in more detail in subsequent sections (such as the examination of bias for the $R_I(470nm)$ channel noted in Tables 6 and 7). Note that AirMSPI $R_I(470nm)$ is generally largest for that instrument's spectral channels, followed by $R_I(865nm)$ then $R_I(660nm)$. The spectral order is different for RSP, where channels have very similar values but are largest at $R_I(865nm)$, followed by $R_I(670nm)$ then $R_I(470nm)$.

C. Statistical considerations

Once we have created the paired AirMSPI and RSP measurements and their corresponding uncertainty values, we must address our hypothesis that instruments agree within stated measurement uncertainties. It is common in the remote sensing community to compare observations by creating measurement vs. measurement scatterplots, calculating a correlation coefficient, and by determining a linear regression to the data (e.g. [57], [58], [59], [60], [61], [38], although many of these publications acknowledge the limitations of this approach). Such methods indicate the amount of association between a pair of measurements, but not the amount of agreement, and not if such measurements can be used interchangeably ([62]). This is in part because this type of analysis is affected by the range of compared values, and because calculations may be based upon numerical distribution assumptions, which have not been tested. Correlation coefficients measure the amount of linear association between measurements, but not equality. For our purposes, these assessment techniques also do not incorporate variable measurement uncertainty, which is the case for both AirMSPI and RSP instruments.

For comparison purposes, we show traditional scatterplots and calculate correlation coefficients and linear regressions, but extend our analysis to the "Limits of Agreement" (LOA) technique developed in the medical statistics literature ([62], [63], [64], [65], [66], [67]). These techniques, and the so-called Bland-Altman plot, are based upon the need to estimate both error and bias between paired measurements, and a desire to visualize this in a more illustrative manner than a scatterplot. The first step of this technique is to calculate and plot the difference of paired measurements against their means (ie. $(x_{\mathrm{RSP}} + x_{\mathrm{AirMSPI}})/2$ vs. $x_{\mathrm{RSP}} - x_{\mathrm{AirMSPI}}$). In this form, bias and difference are more obviously differentiated than in a scatterplot. The next step is to determine the bias and error (scatter) of this relationship. The LOA technique relies upon independence of $x_{\mathrm{RSP}} - x_{\mathrm{AirMSPI}}$ against $(x_{\mathrm{RSP}} + x_{\mathrm{AirMSPI}})/2$. This can be examined with a statistical test

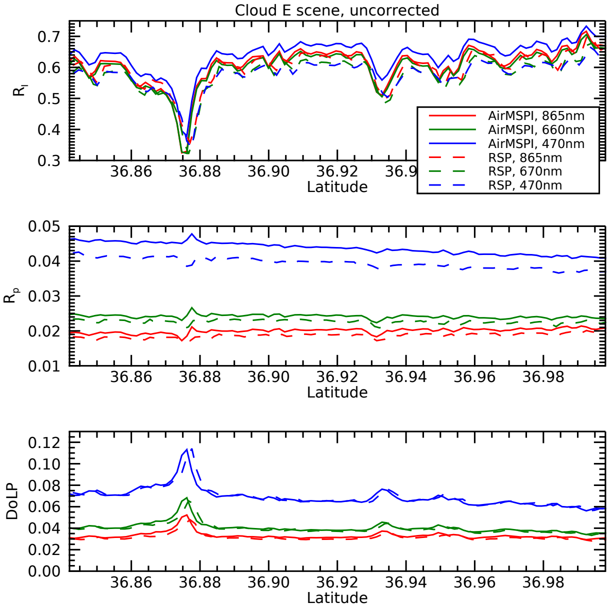

Fig. 6. Example AirMSPI and RSP flight line observations, for marine stratocumulus clouds over the North Pacific Ocean (center coordinates 36.9189°N, 124.976°W) on February 3rd, 2013.

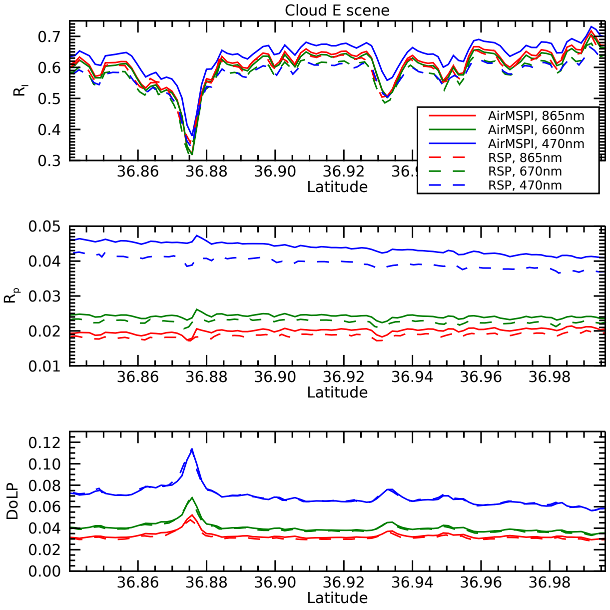

Fig. 7. Example AirMSPI and RSP flight line observations, for marine stratocumulus clouds over the North Pacific Ocean (center coordinates 36.9189°N, 124.976°W) on February 3rd, 2013. Unlike Figure 6, RSP data have been shifted equivalent to a mounting pitch of 0.65° aftwards, which appears to correspond to a better match between the instruments.

of the correlation coefficient. Once independence is established, we can calculate the mean difference, $\bar{d}$ and standard deviation of the difference, s. If these differences are normally distributed, then we can expect that 95% of differences should fall within the following LOA:

$$\begin{aligned}\text{upper limit} &= \bar{d} + 1.96s \\ \text{lower limit} &= \bar{d} - 1.96s.\end{aligned} \tag{3}$$

We can also calculate confidence intervals for the LOA. The $\bar{d}$ standard error is $\sqrt{s^2/n}$, while for the limits above it is roughly $\sqrt{3s^2/n}$. Using a Student's t value of 1.96, the confidence intervals are as presented in Table 5.

An added complication for our analysis is that we expect measurement uncertainties to vary with R_I and $DoLP$. Since we have models for that uncertainty (Appendix A and B), we can validate those models as part of our assessment by transforming the difference between paired measurements (d) by normalizing by measurement uncertainty, which we denote with D:

$$D_i = \frac{d_i}{(\sigma^2_{\mathrm{RSP},i} + \sigma^2_{\mathrm{AirMSPI},i})^{1/2}} = \frac{x_{\mathrm{RSP},i} - x_{\mathrm{AirMSPI},i}}{(\sigma^2_{\mathrm{RSP},i} + \sigma^2_{\mathrm{AirMSPI},i})^{1/2}} \tag{4}$$

In this manner we can compare the expectation of measurement uncertainty to what exists in the comparison data. For example, a D value of 1 indicates that the RSP observation is larger than that of AirMSPI by the sum (in quadrature) of their paired uncertainties. If the measurements agree to stated uncertainties, the LOA should be 1.96 and -1.96 for data normalized by measurement uncertainty. As we shall see, we find this to be the case (or nearly so) for most channels, with some exceptions.

Based on this approach, we perform the following statistical analysis steps:

1. Calculate the Pearson correlation coefficient, and linear regression slope and Y-intercept. Create a scatterplot of AirMSPI to RSP observations. While this is insufficient to test agreement between the observations, it can be used to place the comparison in the context of other analyses.
2. Calculate the normalized difference between observations (D, equation 4), and plot versus mean AirMSPI/RSP observation values to create a Bland-Altman plot.
3. Test if the normalized differences between observations are independent of observation magnitude. If they are not independent, narrow the scope of the analysis by excluding data, if possible, or splitting the analysis into components (which we do for bright and dark scenes).
4. Test the distribution of D for normality.
5. Calculate the LOA as described above, and compare to expectations (lower and upper limits of -1.96 and 1.96, respectively).

If items 4 and 5 meet expectations, then we can say that the instruments agree within stated uncertainties for that channel. Given the possibility of incorrect georegistration contributing to instrument differences, which is not included in the measurement uncertainty models, we also note the cases that come very close to passing the LOA test. If, however, we can not find an independent relationship between observations (step 3) and/or find a normal distribution for D (step 4), we use an additional metric, which is to calculate the percentage of values for D that are greater than 1.96 or less than -1.96. Normally distributed observations that agree within stated uncertainties should have

Table 5. Limits of Agreement (LOA) confidence intervals.

upper limit:	$\bar{d} + 1.96s - (1.96\sqrt{3s^2/n})$	$\leq$	$\bar{d} + 1.96s$	$\leq$	$\bar{d} + 1.96s + (1.96\sqrt{3s^2/n})$
bias:	$\bar{d} - (1.96\sqrt{s^2/n})$	$\leq$	$\bar{d}$	$\leq$	$\bar{d} + (1.96\sqrt{s^2/n})$
lower limit:	$\bar{d} - 1.96s - (1.96\sqrt{3s^2/n})$	$\leq$	$\bar{d} - 1.96s$	$\leq$	$\bar{d} - 1.96s + (1.96\sqrt{3s^2/n})$

less than 5% outside the 1.96 envelope, so this number is an alternate indicator of agreement.

4. RESULTS

Figure 8 contains the scatterplots of AirMSPI to RSP R_I and $DoLP$ presented in this analysis, while Table 6 has the corresponding correlation coefficients and linear regression parameters. The $DoLP$ scatterplot can be compared to the AirMSPI and RSP comparison in Figure 7, top right panel, in [38]. Similar to [38], we find very high correlation coefficients (>0.99), which one should expect for instruments observing the same scene. Our comparison finds slightly greater deviations from 1:1 linear regressions for $DoLP$. The range of the comparison values for some channels differs as well, as [38] does not show $DoLP(865nm)$ greater than about 0.25 for PODEX, while we have $DoLP(865nm)$ as large as 0.45 ($DoLP$ ranges are more similar for the 660/670nm channels, while [38] finds a wider range for 470nm). This indicates different data were used for the comparison, although this is impossible to verify without more descriptive scene detail than are contained in [38]. This also reinforces the assertion by [62], such that the sources of differences in our respective analyses could merely be due to differences in the range of the observations, and not the true population characteristics.

The Bland-Altman plot in Figure 9 offers a much clearer picture of the relationship between the observations. For R_I, the distribution of D appears independent of $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$. Correlation coefficients (line 4. in Table 6) meet or only slightly exceed the critical value for no correlation. For $R_I(470nm)$, there appears to be an AirMSPI > RSP bias equivalent to more than one sigma uncertainty difference. $DoLP$, on the other hand, has a significant correlation between D and $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$. The nature of this relationship is such that bright R_I, low $DoLP$, scenes (land and clouds) have little to no bias, while dark R_I, high $DoLP$ (coastal ocean) scenes have significant RSP > AirMSPI bias. For this reason, we cannot continue with the LOA analysis without further data refinement. In this sense, the distinction between scene types is important. Ocean scenes have very small R_I (see Table 4), and are thus expected to have greater $DoLP$ uncertainty in both the AirMSPI and RSP uncertainty models (Appendix A and B). This increase is due to the contribution of shot noise to total uncertainty, which is otherwise (for larger reflectances associated with land and cloud scenes) dominated by systematic polarimetric characterization and calibration uncertainties. This is the context in which we separately analyze ocean scenes, whose uncertainty is dominated by shot noise, and land or cloud scenes, whose uncertainty is dominated by systematic characterization and calibration errors. As we shall see, the former has a significant RSP > AirMSPI $DoLP$ bias, well beyond measurement uncertainties, while the latter's $DoLP$ largely agrees within stated measurement uncertainties.

We first compare bright land and cloud scenes. Table 7 contains statistics for this restricted comparison dataset. Correlation between D and $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$ for both R_I and $DoLP$ has largely been removed (although it slightly exceeds the critical value for some channels). Distributions of D also come close to or pass a Kolmogorov-Smirnov test for normal distribution. Because of this, we proceed to calculate the bias ($\bar{D}$) and LOA. In addition to Table 7, these are shown graphically as the colored solid ($\bar{D}$) and dashed (LOA) lines in Figure 10. This analysis reveals three significant points:

For bright scenes:

1. AirMSPI and RSP channels at 660/670nm and 865nm mainly agree within stated measurement uncertainties when corrected for small amounts of bias, as indicated by the LOA close to ±1.96.

2. AirMSPI $R_I(470nm)$ is biased higher than that of RSP by a degree greater than the sum of their measurement uncertainties (corresponding to 6.3% in reflectance units). LOA when corrected for bias are otherwise within ±1.96 expectations.

3. While $\bar{D}$ is not large, the LOA for $DoLP(470nm)$ exceed expectations.

As a simple comparison between a pair of instruments, this analysis cannot indicate which of the instruments is the source of the bias identified for $R_I(470nm)$, and it may indeed be due to both. [41] compared RSP to the well calibrated Landsat 8 Operational Land Imager (OLI) using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to transfer across spectral band differences. They found a roughly AVIRIS > RSP bias of about 2.5% at $R_I(470nm)$. However, this channel was most subject to spectral transfer related uncertainties in the analysis technique. Additionally, the AirMSPI team encountered difficulties matching the $R_I(470nm)$ channel in optimal estimation algorithms using the old version V003 data with a MISR-based vicarious calibration in July 2012 (see [22]), and made a calibration adjustment of similar magnitude based on the vicarious calibration in May 2013 as described in Section B. In any case, AirMSPI and RSP are calibrated with respect to different radiometric sphere facilities, so this difference may not be solved without further investigation, including comparisons of the sphere facilities themselves.

Figure 11 and Table 8 contain an analysis of ocean scenes, which have low R_I and moderate to high $DoLP$. Random errors such as shot noise contribute to total uncertainty to a much larger degree than for land and cloud scenes, and uncertainty overall is larger (see Figure 1 and 2). For example, the fraction of random to total uncertainty for RSP $DoLP(670nm)$ is 92%, and for AirMSPI $DoLP(660nm)$ is 37%, compared to 46% and 10%, respectively , for the bright scenes. In this dataset we did not find an independent relationship between D and $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$ (except for $DoLP(865nm)$), meaning that, in a strict sense, LOA should not be calculated. Furthermore, the distributions of D did not pass Kolmogorov-Smirnov tests for normal distribution. This means we could not separate bias and scatter in the relationship between AirMSPI and RSP observations of dark oceans. A complete examination of this relationship

Table 6. Statistics for full comparison dataset, n=636. Differences are calculated as RSP - AirMSPI. The critical value for 4., the correlation of normalized difference (D) with respect to $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$, is calculated to be 0.077 based upon the sample size.

	R$_I$(470nm)	**R$_I$(660/670nm)**	**R$_I$(865nm)**	**DoLP(470nm)**	**DoLP(660/670nm)**	**DoLP(865nm)**
1. Pearson correlation	0.996	0.996	0.996	0.999	0.999	0.999
2. Linear regression slope	0.931	0.992	1.014	1.041	1.034	1.071
3. Linear regression Y-intercept	0.003	0.002	0.004	-0.003	-0.003	-0.004
4. Difference correlation	-0.089	0.297	0.103	0.788	0.836	0.853
5. Bias (Norm. difference, $\bar{D}$)	-1.47	-0.25	0.55	1.53	0.16	0.37

Table 7. Statistics for comparison dataset excluding coastal ocean scenes, n=455. Differences are calculated as RSP - AirMSPI. The critical value for 4., the correlation of normalized difference (D), with respect to $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$ is 0.092. Confidence intervals for $\bar{D}$ (5.) are roughly ±0.08 for all bands, and for Limits of Agreement (LOA) (6., 7.) roughly ±0.14.

	R$_I$(470nm)	**R$_I$(660/670nm)**	**R$_I$(865nm)**	**DoLP(470nm)**	**DoLP(660/670nm)**	**DoLP(865nm)**
1. Pearson correlation	0.990	0.985	0.979	0.998	0.992	0.978
2. Linear regression slope	0.922	0.963	0.964	1.014	0.959	0.989
3. Linear regression Y-intercept	0.008	0.017	0.030	-0.001	0.001	-0.001
4. Difference correlation	-0.001	-0.131	-0.176	0.201	-0.235	0.063
5. Bias (Norm. difference, $\bar{D}$)	-1.53	0.03	0.68	0.41	-0.67	-0.75
6. Upper Limit of Agreement	0.47	2.32	3.04	3.68	1.15	1.24
7. Lower Limit of Agreement	-3.53	-2.26	-1.69	-2.87	-2.49	-2.75
8. Percentage of $\lvert D \rvert > 1.96$	35.6%	8.4%	13.0%	15.6%	7.5%	8.4%

Table 8. Statistics for comparison dataset with only coastal ocean scenes, n=181. Differences are calculated as RSP - AirMSPI. The critical value for 4., the correlation of normalized difference (D), with respect to $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$ is 0.144. Only D for $DoLP(865nm)$ can be considered independent of $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$, so $\bar{D}$ and Limits of Agreement (LOA) were calculated for this band alone. The confidence intervals for $\bar{D}$ (5.) is ±0.10 and for Limits of Agreement (LOA) (6., 7.) roughly ±0.18.

	R$_I$(470nm)	**R$_I$(660/670nm)**	**R$_I$(865nm)**	**DoLP(470nm)**	**DoLP(660/670nm)**	**DoLP(865nm)**
1. Pearson correlation	0.999	0.999	0.999	0.999	0.999	0.996
2. Linear regression slope	0.953	1.004	1.062	1.060	1.043	1.067
3. Linear regression Y-intercept	-0.001	-0.001	-0.001	-0.010	-0.007	-0.002
4. Difference correlation	0.745	0.820	0.558	0.877	0.706	0.113
5. Bias (Norm. difference, $\bar{D}$)						3.19
6. Upper Limit of Agreement						4.90
7. Lower Limit of Agreement						1.47
8. Percentage of $\lvert D \rvert > 1.96$	0.0%	0.0%	0.6%	87.8%	64.1%	91.2%

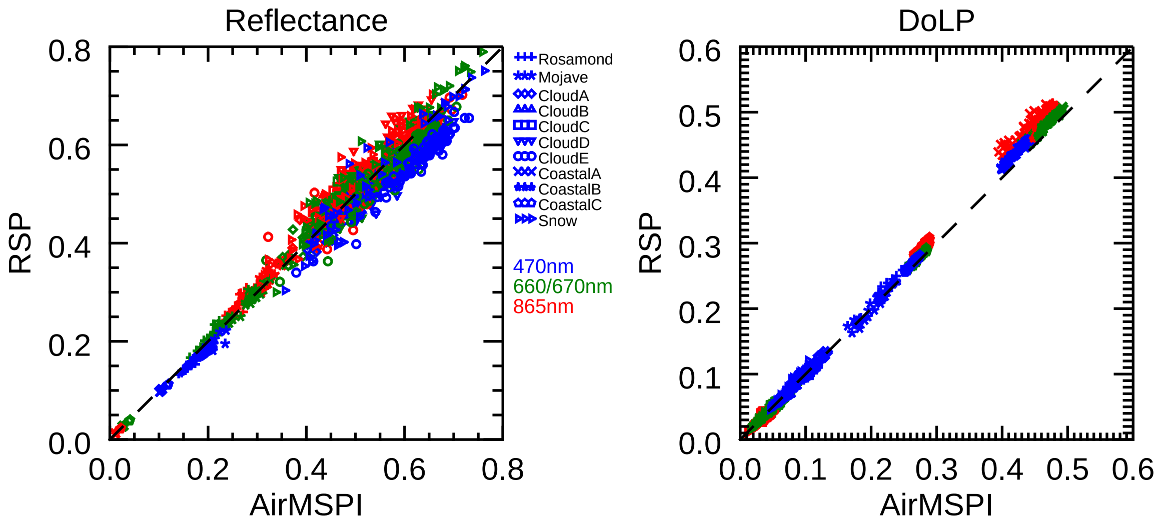

Fig. 8. Scatterplots of AirMSPI vs. RSP reflectance (R_I, left) and Degree of Linear Polarization ($DoLP$, right). Data from the 470nm channel are shown in blue, 660/670nm channel in green, and 865nm channel in red. While a linear, highly correlated relationship between the instruments is obvious, this is not sufficient to prove our hypothesis that instruments agree within stated uncertainties. This is in part because the R_I uncertainty is expressed as a percentage of the measured value, and because $DoLP$ uncertainty is defined with respect to R_I (see Figure 1). Thus, the expected amount of agreement varies throughout the figure. Furthermore, as noted in [62], scatterplots and the associated correlation coefficients are as much an expression of the range of the observed values as they are of the agreement, and are subject to other difficulties that obscure the ability to determine if the measurements can be used interchangeably.

requires more data than were collected during the PODEX field campaign. That said, it is clear from Figure 11 that, while R_I performs well within expectations (less than 1% $|D| > 1.96$), there is a significant RSP > AirMSPI bias for $DoLP$. For 865nm, the only channel for which D can be considered independent of $(x_{\mathrm{RSP}} + x_{\mathrm{AirMSPI}})/2$, a the bias was more than triple the sum of the uncertainties and on average equal to 0.0245 in units of DoLP. We have no explanation for this lack of agreement, although its presence only for dark scenes dominated by random errors may be a clue. Without a correction, comparison of Level 2 (geophysical) parameters from AirMSPI and RSP observations over the ocean must be made with caution, as it will be difficult to identify if differences reside in the Level 1 observations (which we assess) or in the strategy and algorithms used for Level 2 product determination. Sources of other comparison data over the oceans could include the SEAC4RS and ACEPOL field campaigns, which will be the subject of future analysis. The ACEPOL field campaign also collected data with the aircraft heading in the solar principal plane (on October 23rd, 2017), meaning that $DoLP$ values approaching 1.0 were observed.

5. CONCLUSIONS

The intent of this analysis is to verify that Level 1 observations (geolocated radiances and polarization) from the AirMSPI and RSP airborne polarimeters agree within stated measurement uncertainties. If it exists, such agreement would support the hypothesis that the instruments are behaving as expected, and that their measurement uncertainty models are appropriate. The latter is important, since uncertainty models are a key component of the algorithms that determine Level 2 (geophysical) products, which must properly weigh observations that can have either high and low measurement uncertainty (e.g. R_I and $DoLP$). Knowledge of measurement uncertainty can also be passed on to estimates of Level 2 product uncertainty, so comparison at that level requires validation that Level 1 observations agree. That said, both uncertainty models (for RSP and for AirMSPI when spatially averaged to the RSP footprint size) are significantly smaller than Decadal Survey ACE polarimetric uncertainty requirements, so small lack of agreement between sensors does not indicate that ACE requirements can not be met. As described in the AirMSPI uncertainty model in Appendix A and displayed in Figure 1, the native AirMSPI Step and Stare resolution uncertainty meets ACE requirements for only the brightest scenes (such as clouds). The instrument meets requirements at scene reflectances ≥ 0.02 by averaging an 8x8 array of pixels.

We do not have knowledge of true scene R_I and $DoLP$, but we can verify if measurements agree within stated uncertainties. While the possibility exists that the instruments diverge from geophysical reality but agree with each other, we consider this to be unlikely given that the instruments have radically different designs and calibration methods. We can not attribute

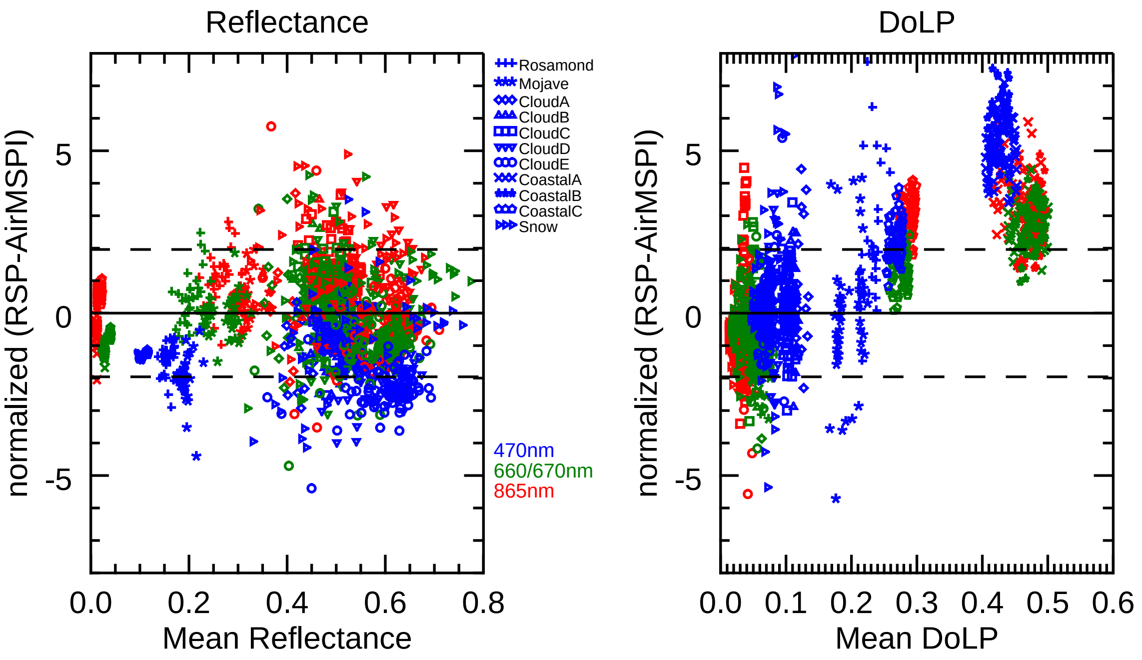

Fig. 9. Uncertainty normalized RSP-AirMSPI difference (D), plotted with respect to $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$, for R_I (left) and $DoLP$ (right). Normalization is performed according to equation 4, and is done to account for variable measurement uncertainty. Color denotes comparison channel, such that 470nm is blue, 660/670nm is green, and 865nm is red. Symbols indicate the comparison scene (see Appendix D), while horizontal and dashed solid lines indicate $D = 0$ and $D \pm 1.96$, respectively. Roughly 95% of comparisons should fall within the dashed lines. $DoLP$ shows a significant positive correlation, so Limits of Agreement (LOA) were not calculated. This figure expresses the same data as 8, but in a manner that better visualizes the the relationship between observations and incorporates variable measurement uncertainty.

divergences to one instrument or the other, but simply identify if paired sets of measurements agree as expected. In performing this analysis, we go beyond the traditional comparison techniques used in the aerosol remote sensing community (scatterplots and calculation of correlation coefficients and linear regressions). These techniques have been shown to be related to, but not proof of, agreement between measurements. In our case they are also incompatible with measurement uncertainty models which depend upon R_I and $DoLP$. Instead, we use the technique of Bland-Altman plots and calculation of mean bias and LOA in use by the medical statistics community (e.g. [62]) to assess agreement between observations. Our addition to the techniques of [62] is to normalize the difference between a pair of observations by the quadratic sum of the measurement uncertainties for those values, thus transforming the measurement differences into a space defined by those uncertainties in a clear manner (i.e., a value of 1 indicates a 1 sigma difference). This technique may be useful for other comparison activities, such as from observations made by the recent ACEPOL field campaign.

The Bland-Altman technique calls for differentiation between mean bias ($\bar{D}$) and scatter (described by the LOA), which is verified by testing for independence between D and $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$. We found that this independence did not exist for the dataset as whole, but could be partially verified if the data were split into groups: one defined by bright and weakly polarized scenes over land and clouds, the other by dark and moderate to highly polarized scenes over the ocean.

The bright land and cloud group has systematic calibration and characterization uncertainty as its primary source of total uncertainty. With some exceptions, both R_I and $DoLP$ agree within stated uncertainties. This is a distinct improvement over an earlier analysis with the first version of the data, for which R_I agreed, but $DoLP$ did not. We did notice a lack of agreement for the 470nm channel. For R_I, this was exhibited as an AirMSPI > RSP bias of $\bar{D} = -1.53$, corresponding to 6.3% in reflectance units. As noted previously, either instrument (or both) could be the source of this bias. For $DoLP(470nm)$, we noticed LOA slightly larger than expectations, but little bias, indicating that the bias in R_I does not seem to impact $DoLP$. This result indicates the need for a cross comparison of the (different) calibration facilities used for the instruments.

The total uncertainty of the dark ocean group is dominated by random uncertainty (such as shot noise). Even with this more narrowly defined group, it was difficult to find independence between D and $(x_{\text{RSP}} + x_{\text{AirMSPI}})/2$. For this reason, $\bar{D}$ and LOA could only be calculated for one channel, $DoLP(865nm)$. For this channel, a significant RSP > AirMSPI bias was found ($\hat{D} = 3.19$, corresponding to 0.0245 in DoLP on average). Visually (see Figure 11), it is clear that the other $DoLP$ wavelengths are similarly affected. However, all three channels of R_I fall entirely within the $-1.96 < D < 1.96$ range. For this reason, we believe these channels agree within stated uncertainties, although a bias at 470nm, similar to that found in the other group, is also present.

These results are the product of exceptional effort by the

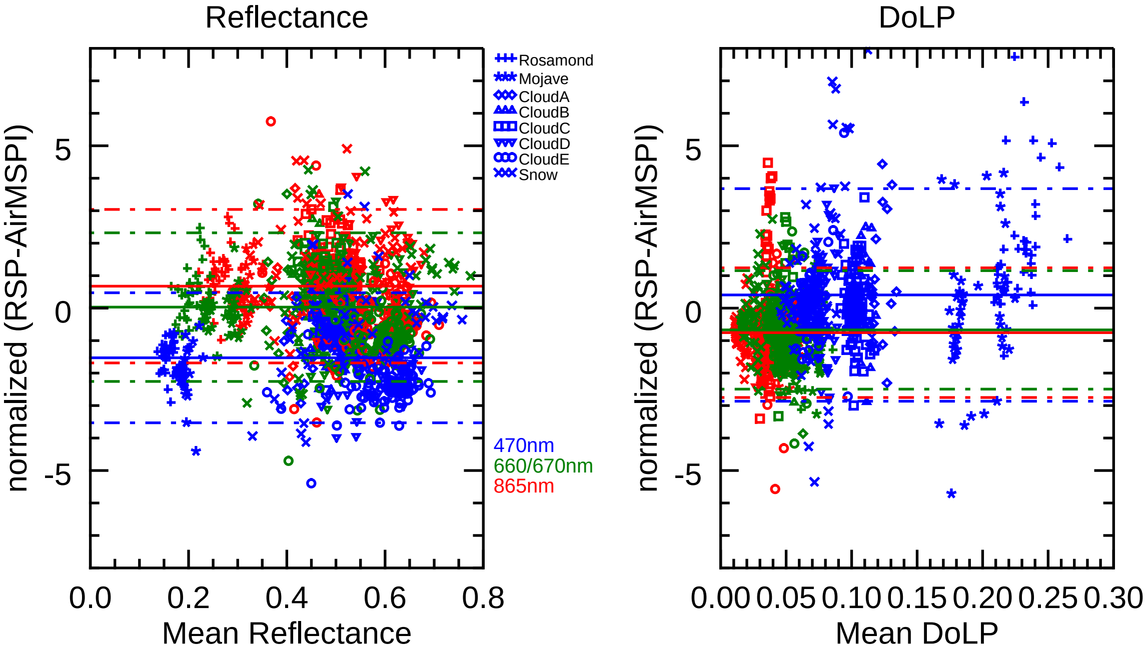

Fig. 10. The same as Figure 9, but with low R_I, high $DoLP$, coastal scenes removed. Without these scenes correlation is minimal, so Limits of Agreement (LOA) are calculated and shown with solid and dashed lines. Limits are within the ±1.96 range, indicating agreement within stated uncertainties, for all channels except 470nm. $R_I(470nm)$ show a significant AirMSPI > RSP bias roughly equivalent to 6%, while $DoLP(470nm)$ has a somewhat larger than required scatter.

instrument teams to properly calibrate, georegister, and otherwise process their data, and represent a significant improvement over comparisons with the first versions of the data products. Both teams revisited their data and uncertainty models, to make continual improvement, with a goal of meeting the ACE polarimetric uncertainty of 0.005 (see black line in Figure 1). As noted above, we validate uncertainty models for bright scenes that are far better than the ACE threshold, but find differences between dark scenes that exceed the uncertainty models. These differences do not mean that instruments cannot meet ACE requirements, but that the uncertainty models that define them do not seem appropriate for $DoLP$ for dark scenes. They do indicate agreement, and therefore meet ACE polarimetric requirements, for bright scenes. We should also note that spatial averaging of AirMSPI data has an important impact on measurement uncertainty. For most scene brightnesses, the model for AirMSPI polarimetric $f$measurement uncertainty exceeds the ACE requirements at the native Step and Stare spatial resolution, but can be met after spatial averaging. This means that at high spatial resolution, AirMSPI $DoLP$ is dominated by random sources of uncertainty, which have been reduced from our analysis due to the need for spatial averaging. The exception is for the low reflectance, ocean scenes, where instruments did not agree within stated uncertainties.

The lack of agreement over dark, moderate to highly polarizing ocean scenes is an indicator that one or both instruments has either a bias or is underestimating polarimetric measurement uncertainty for such scenes. Given that dark ocean polarimetric observations are dominated by random, not systematic, sources of uncertainty (see Figure 2), underestimation of random measurement uncertainty (noise) is a likely scenario. This points to the need for a reassessment of that component of the instrument measurement uncertainty models. The bias present in Figure 11 also indicates a need for laboratory cross comparison of the instruments. Uncorrected, these differences can have several consequences. Retrieval algorithms that use expectations of measurement uncertainty may rely too heavily on polarimetric data (compared to reflectance), and may underestimate the uncertainty in the retrieved geophysical product. Comparison of geophysical products between the two instruments must also consider that sources of a difference could be instrumental, in addition to due to the differences in retrieval algorithms or available information content.

An initial comparison between the AirMSPI and RSP instruments was performed in [38] for the purposes of validating the AirMSPI instrument. The root-mean-square $DoLP$ differences determined in that paper generally agreed with instrument measurement uncertainties. This analysis went further, by comparing in a manner that incorporates pixel level uncertainties and uses the statistically rigorous Bland-Altman technique. This lead to a perhaps more nuanced understanding of the relationship between the instruments. While in a general sense the instruments still agree within stated measurement uncertainties (and within ACE polarimetric uncertainty requirements) we uncovered a radiometric bias at 470nm and polarimetric differences beyond measurement uncertainties for dark ocean scenes, neither of which was found in [38].

Finally, this paper presents a methodology for determining properly weighted paired AirMSPI and RSP observations, and their corresponding uncertainties. Details of these efforts are

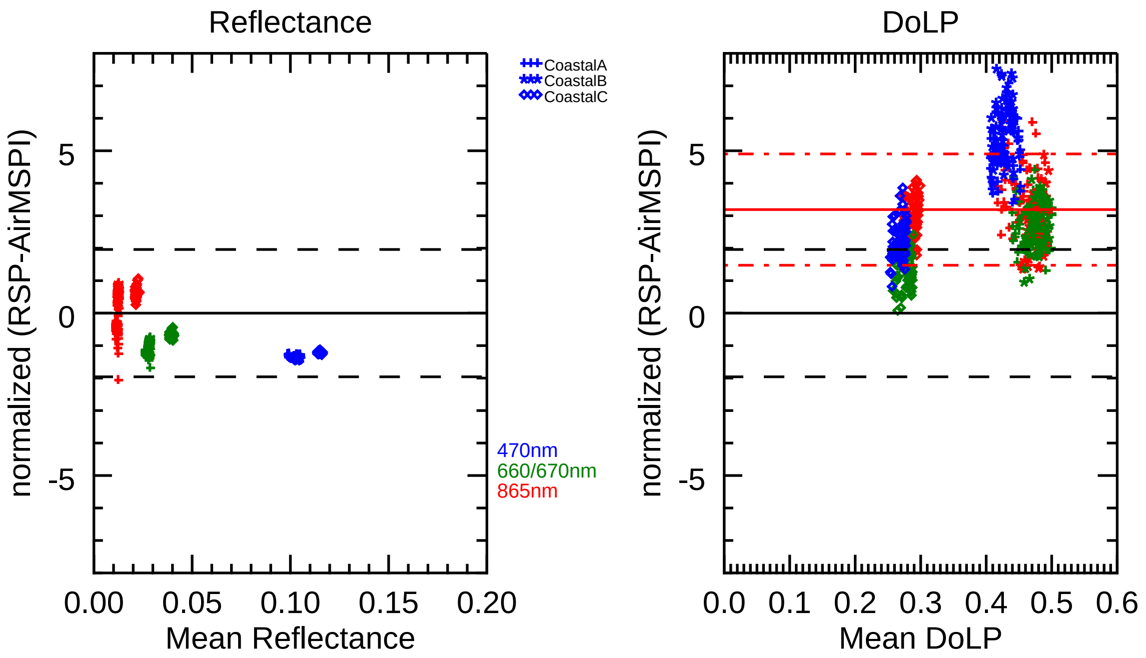

Fig. 11. The same as Figure 9, but with low R_I, high $DoLP$, coastal scenes only. Except for $DoLP(865nm)$, there is significant correlation between D and $(x_{\mathrm{RSP}} + x_{\mathrm{AirMSPI}})/2$, so we cannot calculate Limits of Agreement (LOA).

described in the appendices. Appendix A is the AirMSPI uncertainty model, while Appendix B is the same for RSP. Appendix C describes efforts to account for flex in the ER-2 aircraft wing, which potentially affects georegistration. A list of comparison scenes is in Appendix D. This information may be a resource for subsequent comparative analysis.

6. DATA AVAILABILITY

Calibrated, geolocated data Level 1 from AirMSPI and RSP are available to the public at the locations described in Table 9. These locations also include ancillary data and products of Level 2 analysis. Individual filenames are included in Table 12.

7. SUPPLEMENTAL MATERIAL

A. Appendix A: AirMSPI uncertainty model

More details about the derivation of this uncertainty model can be found in [22] and [68].

Signal to noise ratio

Both the radiometric and polarimetric uncertainty calculations require determining the signal-to-noise ratio (SNR). This is given by:

$$SNR = \frac{S(lmn)^{1/2}}{(1.25S + r^2 f)^{1/2}} \quad \textbf{(5)}$$

where
S = signal in electrons
l = the number of rows read out from the AirMSPI detector array in a given channel. Since the rows are co-registered in ground data processing, this has no effect on spatial resolution. The instrument is currently configured such that l = 1 in all channels.
m = number of cross-track pixels to be averaged during image analysis.
n = number of along-track pixels to be averaged during image analysis.
r = read noise = 9 electrons
f = number of subframes in each image frame = 23.

Signal, S, is calculated from the following formula:

$$S = \frac{1.408 \times 10^{18} \xi \eta \mu_s R_I \Delta\lambda}{\lambda^4 \left[\exp \frac{2489.7}{\lambda} - 1\right]} \quad \textbf{(6)}$$

where
ξ = optical throughput (dimensionless)
η = detector quantum efficiency (electrons/photon)
R_I = intensity bidirectional reflectance factor (BRF)
μ_s = cosine of the solar zenith angle (dimensionless)
λ = band center wavelength in nanometers
$\Delta\lambda$ = bandpass in nanometers.

Note that the number of cross-track (m) and along-track (n) pixels are defined at the sensor, and are not equivalent to the number of pixels in ground projected Sw or SaS data. The number of at-sensor pixels contained in a ground projected pixel is a function of viewing geometry. For the SaS mode, the number of at-sensor pixels per ground projected pixels, N, can be calculated as:

$$N = 1.45 \frac{1 - \sin^2(\alpha - \beta)\sin^2(\gamma)}{\sqrt{1 + \cos^2(\alpha - \beta)\tan^2(\gamma)}} \quad \textbf{(7)}$$

Table 9. AirMSPI and RSP data availability

AirMSPI, ellipsoid projected (cloud and ocean scenes):

DOI 10.5067/AIRCRAFT/AIRMSPI/PODEX/RADIANCE/ELLIPSOID_V005

AirMSPI, terrain projected (land scenes):

DOI 10.5067/AIRCRAFT/AIRMSPI/PODEX/RADIANCE/TERRAIN_V005

RSP, all data

https://data.giss.nasa.gov/pub/rsp/PODEX

where
α = aircraft heading
β = viewing azimuth angle
γ = viewing zenith angle.

Radiometric uncertainty The uncertainty in top-of-atmosphere equivalent reflectance is the result of radiometric calibration uncertainty and random noise. Letting C be the radiometric calibration uncertainty, expressed as a fraction of total signal, then the fractional uncertainty is:

$$\sigma_{R_I}^2 = \mu_s^2 R_I^2 \left(C^2 + SNR^{-2}\right) \tag{8}$$

SNR is calculated from Equation 5 and C is currently estimated at 5%. With further refinement, it is expected to improve to roughly 3%.

DoLP uncertainty As mentioned above, the ground projected Sw or SaS data are gridded such that multiple at-sensor pixels may be averaged. Further averaging is required to meet the ACE $DoLP$ uncertainty requirement. Averaging pixels results in a proportionate decrease in spatial resolution.

The uncertainty in the DOLP due to random noise is given by

$$\sigma_{DoLP}(noise) = s/SNR \tag{9}$$

where s is the noise sensitivity factor resulting from the PEM demodulation approach. The values of s in the AirMSPI polarization bands are given in Table 10. Polarimetric calibration in the optical lab leaves a systematic uncertainty of 0.001 at any DOLP for all three polarimetric bands ($\sigma_{DoLP}(calibration) = 0.001$). The uncertainty associated with the in-flight stability of the polarization modulator is

$$\begin{aligned} \sigma_{DoLP}(in-flight, 470, 660nm) &= 0.001DoLP \\ \sigma_{DoLP}(in-flight, 865nm) &= 0.003DoLP \end{aligned} \tag{10}$$

Thus, the total DoLP uncertainty is the root-sum-squared for the three aforementioned uncertainties, according to

$$\begin{aligned} \sigma_{DoLP}(470, 660nm)^2 &= (s/SNR)^2 + 0.001^2 + 0.001DoLP^2 \\ \sigma_{DoLP}(865nm)^2 &= (s/SNR)^2 + 0.001^2 + 0.003DoLP^2. \end{aligned} \tag{11}$$

B. Appendix B: RSP uncertainty model

An early model for multi-angle polarimeters similar to the RSP was included in [5]. This model splits sources of uncertainty into those that are caused by systematic biases, primarily calibration related, and those due to measurement noise. The differentiation is that the latter can be reduced by averaging, while the former cannot. Definitions and units for the parameters that define the RSP uncertainty model are defined in Table 11.

The **reflectance uncertainty model** provides the absolute uncertainty in units of reflectance.

Uncertainty due to noise:

$$\sigma_{R_I}^2(noise) = \left(\frac{r^2\sigma'_{floor}}{\mu_s}\right)^2 + \frac{a' R_I r^2}{2\mu_s} \tag{12}$$

Uncertainty due to calibration:

$$\sigma_{R_I}^2(calibration) = \frac{\sigma_{InK}^2}{16} R_P^2 + \sigma_{\alpha_c}^2 R_I^2 \tag{13}$$

Total reflectance uncertainty:

$$\sigma_{R_I}^2 = \left(\frac{r^2\sigma'_{floor}}{\mu_s}\right)^2 + \frac{a' R_I r^2}{2\mu_s} + \frac{\sigma_{InK}^2}{16} R_P^2 + \sigma_{\alpha_c}^2 R_I^2 \tag{14}$$

The **DoLP uncertainty model** provides the absolute uncertainty in the units of DoLP (percentages).

Uncertainty due to noise:

$$\begin{aligned} &\sigma_{DoLP}^2(noise) = \\ &4\left(1 + \frac{DoLP^2}{2}\right)\left(\frac{r^2\sigma'_{floor}}{\mu_s R_I}\right)^2 + 2\left(1 - \frac{DoLP^2}{2}\right)\frac{a' r^2}{\mu_s R_I} \end{aligned} \tag{15}$$

Uncertainty due to calibration:

$$\begin{aligned} &\sigma_{DoLP}^2(calibration) = \\ &\frac{\sigma_{InK}^2}{2}\left[1 - DoLP^2 + \frac{DoLP^4}{2}\left(1 - 0.5\sin^2 4\chi\right)\right] + \sigma_{In\alpha}^2 DoLP^2 \end{aligned} \tag{16}$$

note χ is the polarization azimuth angle.

Total uncertainty:

$$\begin{aligned} &\sigma_{DoLP}^2 = \\ &4\left(1 + \frac{DoLP^2}{2}\right)\left(\frac{r^2\sigma'_{floor}}{\mu_s R_I}\right)^2 + 2\left(1 - \frac{DoLP^2}{2}\right)\frac{a' r^2}{\mu_s R_I} \\ &+ \frac{\sigma_{InK}^2}{2}\left[1 - DoLP^2 + \frac{DoLP^4}{2}\left(1 - 0.5\sin^2 4\chi\right)\right] + \sigma_{In\alpha}^2 DoLP^2 \end{aligned} \tag{17}$$

Table 10. Spectrally dependent AirMSPI uncertainty model parameters

Wavelength, λ (nm)	355	380	445	470	555	660	865	935
$\Delta\lambda$	30	32	36	37	31	42	39	48
ξ	0.806	0.710	0.551	0.516	0.641	0.605	0.602	0.607
η	0.12	0.19	0.35	0.40	0.43	0.35	0.13	0.08
s				4.37		3.61	2.96	

Table 11. RSP uncertainty model parameters. Some coefficient values are different for the pair of instruments (RSP1 and RSP2). RSP2 was used during PODEX, so the uncertainty model for those coefficients was used in this study.

Parameter	RSP1 Value	RSP2 Value	Description
r	1.0	1.0	Solar distance, in astronomical units (AU)
σ'_{floor}	2.0×10^{-5}	2.0×10^{-5}	Detector noise floor, with scaling for normalized radiance
a'	1.0×10^{-7}	1.0×10^{-7}	Shot noise parameter, with scaling for normalized radiance
σ_{lnK}	0.005	0.002	Relative gain coefficient characterization uncertainty
σ_{α_c}	0.015	0.03	Absolute radiometric characterization uncertainty
$\sigma_{ln\alpha}$	0.002	0.002	Polarimetric characterization uncertainty
μ_s			Cosine of the Solar Zenith Angle
R_I			Intensity reflectance
R_P			Polarized reflectance
$DoLP$			Degree of Linear Polarization, $DoLP = R_P/R_I$

The **polarized reflectance uncertainty model**, like that of reflectance and DoLP, is absolute, and in the units of reflectance. Uncertainty due to noise:

$$\sigma^2_{R_P}(noise) = 4\left(\frac{r^2\sigma'_{floor}}{\mu_s}\right)^2 + 2\frac{a'r^2R_I}{\mu_s} \qquad \textbf{(18)}$$

Uncertainty due to calibration:

$$\sigma^2_{R_P}(calibration) = \frac{\sigma^2_{lnK}}{2}R_I^2 + (\sigma^2_{ln\alpha} + \sigma^2_{\alpha_c})R_P^2 \qquad \textbf{(19)}$$

Total uncertainty:

$$\sigma^2_{R_P} = 4\left(\frac{r^2\sigma'_{floor}}{\mu_s}\right)^2 + 2\frac{a'r^2R_I}{\mu_s} + \frac{\sigma^2_{lnK}}{2}R_I^2 + (\sigma^2_{ln\alpha} + \sigma^2_{\alpha_c})R_P^2 \qquad \textbf{(20)}$$

C. Appendix C: RSP mounting and wing flex corrections

RSP does not include an Inertial Measurement Unit (IMU) and instead uses information from the IMU mounted in the ER-2 body to determine the roll, pitch and yaw of the instrument during flight. However, some corrections need to be applied to the ER-2 IMU measurements because 1) the angles of the mounting of RSP in the ER-2 wing pod relative to the aircraft body are not determined absolutely, and 2) the wings tend to flex progressively less during flight as fuel load is depleted. These adjustments depend on which ER-2 pod is used, the load, version of the ER-2 and possibly other factors.

To account for the wing flex for the PODEX flights, IMU measurements provided by the enhanced MODIS Airborne Simulator (eMAS) that was mounted on the left wing of the ER-2 during the SEAC^{4}RS campaign are used. The difference in roll measured by the ER-2 body IMU and the eMAS IMU is interpreted as caused by the angular flex of the wing. Wing flex information for all flights during SEACR^{4}RS are evaluated as a function of flight time normalized by the total flight duration, and fitted using a quadratic function. The coefficients of this function are $a_0 = 2.155°$, $a_1 = -1.122°$, $a_2 = 0.4695°$. The assumption is that the fuel loads at the beginning and end of the flights are roughly the same for all flights. Since RSP was mounted on the right wing during PODEX, the wing flex model is mirrored (i.e. $a_x \rightarrow -a_x$) assuming that the wing flex is the same in both wings, which is likely justified since ER-2 operations requires that the instrument payload and fuel in both wings is balanced. Note also that the right wing pod on the ER-2 contained RSP and CPL during both SEAC^{4}RS and PODEX, and thus the wing flex for both campaigns is assumed similar.

In addition to the wing flex, the IMU measurements need to be adjusted for the roll, pitch and yaw angles of the instrument mounting in the ER-2 wing pod. These angles were determined by a combination of on ground measurements and geophysical feature matching using high spatial resolution spectral imagery of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument from flights during the HyspIRI mission when RSP was again mounted in the right wing pod. The pitch, yaw, and effective roll were determined to be 8.35°, -1.07° and 1.1°, respectively. Note that this roll is in addition to the roll caused by the wing flex.

D. Appendix D: Comparison scenes

This analysis was performed for 11 scenes observed during PODEX between January 14th and February 6, 2013. A total 576 individual footprints were compared. 55% of those were cloud, 21% cloud free ocean, and 23% cloud free land scenes.

Table 12. Details of comparison scenes. Note the "CloudC" scene was used for validation of cloud retrievals (case 1) in [11].

Short Name	Type	# footprints	Coordinates	Date	UTC (RSP start)	RSP start	RSP end
Rosamond	Cloud free land	37	34.8266°N, 118.476°W	2013/01/31	20:57	34.8542°N	34.7977°N
AirMSPI: AirMSPI_ER2_GRP_TERRAIN_20130131_210159Z_CA-Rosamond_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130131T205704Z_V002-20170315T202217Z.h5							
Mojave	Cloud free land	47	34.8689°N, 118.079°W	2013/01/31	21:09	34.8330°N	34.9050°N
AirMSPI: AirMSPI_ER2_GRP_TERRAIN_20130131_211538Z_CA-Mojave_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130131T210946Z_V002-20170315T202510Z.h5							
CloudA	Cloudy	24	36.8245°N, 122.477°W	2013/02/03	17:56	36.8168°N	36.8317°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130203_181210Z_NorthPacificOcean-37N122W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130203T175613Z_V002-20170315T211227Z.h5							
CloudB	Cloudy	24	37.8947°N, 124.627°W	2013/02/03	18:17	37.8844°N	37.9039°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130203_183045Z_NorthPacificOcean-38N125W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130203T181701Z_V002-20170315T211447Z.h5							
CloudC	Cloudy	106	37.5503°N, 125.250°W	2013/02/03	18:45	37.6241°N	37.4762°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130203_185005Z_NorthPacificOcean-38N125W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130203T184505Z_V002-20170315T212022Z.h5							
CloudD	Cloudy	62	38.3300°N, 124.874°W	2013/02/03	19:19	38.2817°N	38.3770°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130203_194304Z_NorthPacificOcean-38N125W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130203T191950Z_V002-20170315T211847Z.h5							
CloudE	Cloudy	106	36.9189°N, 124.976°W	2013/02/03	20:31	36.8401°N	36.9962°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130203_203602Z_NorthPacificOcean-37N125W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130203T203103Z_V002-20170315T212653Z.h5							
CoastalA	Cloud free ocean	60	32.8326°N, 117.344°W	2013/01/14	21:02	32.8527°N	32.8120°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130114_210928Z_CA-SanDiegoCounty_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130114T210241Z_V002-20170315T162343Z.h5							
CoastalB	Cloud free ocean	61	32.2821°N, 118.566°W	2013/01/14	21:02	32.3034°N	32.2602°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130114_212034Z_NorthPacificOcean-32N119W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130114T210241Z_V002-20170315T162343Z.h5							
CoastalC	Cloud free ocean	60	33.5793°N, 118.103°W	2013/02/06	19:35	33.6081°N	33.5499°N
AirMSPI: AirMSPI_ER2_GRP_ELLIPSOID_20130206_194400Z_NorthPacificOcean-34N118W_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130206T193554Z_V002-20170315T222056Z.h5							
Snow	Snow	49	39.7441°N, 120.301°W	2013/01/31	20:06	39.7819°N	39.7061°N
AirMSPI: AirMSPI_ER2_GRP_TERRAIN_20130131_201432Z_CA-Loyalton_000N_F01_V005.hdf							
RSP: RSP2-ER2_L1B-RSPGEOL1B-GeolocatedRadiances_20130131T200659Z_V002-20170315T202019Z.h5							

8. ACKNOWLEDGEMENTS, AUTHOR CONTRIBUTIONS AND FUNDING INFORMATION

This analysis was lead by the first author (Knobelspiesse) with significant support from second author Tan, and with continual review and assessment by the AirMSPI and RSP instrument teams (the remaining authors, excluding Ferrare and Redemann). Data and its preparation, characterization and uncertainty model creation was performed by the AirMSPI and RSP instrument teams. Coauthors Ferrare and Redemann provided additional advice, review and support. Portions of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

The PODEX field campaign, including deployment and development of instruments and this analysis, was supported by the NASA ACE mission, managed by Hal Maring and Arlindo da Silva. AirMSPI data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. We would also like to acknowledge the large team involved in the successful conduct of the PODEX field campaign, including the ER-2 aircraft pilots, crew and management, the PODEX field campaign team, and the engineers who ensured the proper operation of the instruments and other supporting systems.

REFERENCES

1. IPCC, *Climate Change 2013 - The Physical Science Basis : Contribution of the Working Group I to the Fifth Assessment Report of the IPCC* (Cambridge University Press, New York, NY, 2013).
2. L. A. Remer, R. G. Kleidman, R. C. Levy, Y. Kaufman, D. Tanré, S. Mattoo, J. V. Martins, C. Ichoku, I. Koren, H. Yu, and B. Holben, “Global aerosol climatology from the modis satellite sensors,” J. Geophys. Res. **113** (2008).
3. M. Mishchenko, B. Cairns, J. Hansen, L. Travis, R. Burg, Y. Kaufman, J. Vanderlei Martins, and E. Shettle, “Monitoring of aerosol forcing of climate from space: analysis of measurement requirements,” J. Quant. Spectrosc. Radiat. Transf. **88**, 149–161 (2004).
4. M. Mishchenko, B. Cairns, G. Kopp, C. Schueler, B. Fafaul, J. Hansen, R. Hooker, T. Itchkawich, H. Maring, and L. Travis, “Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory mission,” Bull. Amer. Meteorol. Soc **88**, 687–691 (2007).
5. K. Knobelspiesse, B. Cairns, M. Mishchenko, J. Chowdhary, K. Tsigaridis, B. van Diedenhoven, W. Martin, M. Ottaviani, and M. Alexandrov, “Analysis of fine-mode aerosol retrieval capabilities by different passive remote sensing instrument designs,” Opt. Express **20**, 21457–21484 (2012).
6. O. Hasekamp and J. Landgraf, “Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements,” Appl. Opt. **46**, 3332–3344 (2007).
7. O. Dubovik, M. Herman, A. Holdak, T. Lapyonok, D. Tanré, J. L. Deuzé, F. Ducos, A. Sinyuk, and A. Lopatin, “Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multiangle polarimetric satellite observations,” Atmospheric Meas. Tech. **4**, 975–1018 (2011).
8. A. Kokhanovsky, A. Davis, B. Cairns, O. Dubovik, O. Hasekamp, I. Sano, S. Mukai, V. Rozanov, P. Litvinov, T. Lapyonok, I. Kolomiets, Y. Oberemok, S. Savenkov, W. Martin, A. Wasilewski, A. di Noia, A. Stap, J. Rietjens, F. Xu, V. Natraj, M. Duan, T. Cheng, and R. Munro, “Space-based remote sensing of atmospheric aerosols: The multi-angle spectro-polarimetric frontier,” Earth-Science Rev. **145**, 85–116 (2015).
9. P. Goloub, M. Herman, H. Chepfer, J. Riedi, G. Brogniez, P. Couvert, and G. Sèze, “Cloud thermodynamical phase classification from the POLDER spaceborne instrument,” J. Geophys. Res. **105**, 14747–14760 (2000).
10. M. Alexandrov, B. Cairns, and M. Mishchenko, “Rainbow fourier transform,” J. Quant. Spectrosc. Radiat. Transf. **113**, 2521–2535 (2012).
11. M. D. Alexandrov, B. Cairns, A. P. Wasilewski, A. S. Ackerman, M. J. McGill, J. E. Yorks, D. L. Hlavka, S. E. Platnick, G. T. Arnold, B. van Diedenhoven, J. Chowdhary, M. Ottaviani, and K. D. Knobelspiesse, “Liquid water cloud properties during the polarimeter definition experiment (podex),” Remote. Sens. Environ. **169**, 20 – 36 (2015).
12. M. D. Alexandrov, B. Cairns, C. Emde, A. S. Ackerman, and B. van Diedenhoven, “Accuracy assessments of cloud droplet size retrievals from polarized reflectance measurements by the research scanning polarimeter,” Remote. Sens. Environ. **125**, 92 – 111 (2012).
13. B. van Diedenhoven, B. Cairns, I. V. Geogdzhayev, A. M. Fridlind, A. S. Ackerman, P. Yang, and B. A. Baum, “Remote sensing of ice crystal asymmetry parameter using multi-directional polarization measurements; part 1: Methodology and evaluation with simulated measurements,” Atmospheric Meas. Tech. **5**, 2361–2374 (2012).
14. B. van Diedenhoven, A. M. Fridlind, B. Cairns, and A. S. Ackerman, “Variation of ice crystal size, shape, and asymmetry parameter in tops of tropical deep convective clouds,” J. Geophys. Res. Atmospheres **119**, 11,809–11,825 (2014).
15. B. van Diedenhoven, A. M. Fridlind, B. Cairns, A. S. Ackerman, and J. Yorks, “Vertical variation of ice particle size in convective cloud tops,” Geophys. Res. Lett. **43**, 4586–4593 (2016).
16. National Research Council, https://doi.org/10.17226/11820, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond* (2007).
17. ACE Science Working Group, https://acemission.gsfc.nasa.gov, *Aerosol, Cloud and Ecosystems (ACE) Proposed Satellite Mission, Study Report* (2010).
18. J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, “Review of passive imaging polarimetry for remote sensing applications,” Appl. Opt. **45**, 5453–5469 (2006).
19. M. D. Alexandrov, B. Cairns, B. van Diedenhoven, A. S. Ackerman, A. P. Wasilewski, M. J. McGill, J. E. Yorks, D. L. Hlavka, S. E. Platnick, and G. T. Arnold, “Polarized view of supercooled liquid water clouds,” Remote. Sens. Environ. **181**, 96 – 110 (2016).
20. A. Di Noia, O. P. Hasekamp, L. Wu, B. van Diedenhoven, B. Cairns, and J. E. Yorks, “Combined neural network/phillips-tikhonov approach to aerosol retrievals over land from the nasa research scanning polarimeter,” Atmospheric Meas. Tech. **10**, 4235–4252 (2017).
21. B. van Diedenhoven, B. Cairns, A. M. Fridlind, A. S. Ackerman, and T. J. Garrett, “Remote sensing of ice crystal asymmetry parameter using multi-directional polarization measurements; part 2: Application to the research scanning polarimeter,” Atmospheric Chem. Phys. **13**, 3185–3203 (2013).
22. D. J. Diner, F. Xu, M. J. Garay, J. V. Martonchik, B. E. Rheingans, S. Geier, A. Davis, B. R. Hancock, V. M. Jovanovic, M. A. Bull, K. Capraro, R. A. Chipman, and S. C. McClain, “The airborne multiangle spectropolarimetric imager (airmspi) a new tool for aerosol and cloud remote sensing,” Atmospheric Meas. Tech. **6**, 2007–2025 (2013).
23. M. Ottaviani, B. van Diedenhoven, and B. Cairns, “Photopolarimetric retrievals of snow properties,” The Cryosphere **9**, 1933–1942 (2015).
24. L. Wu, O. Hasekamp, B. van Diedenhoven, and B. Cairns, “Aerosol retrieval from multiangle, multispectral photopolarimetric measurements: importance of spectral range and angular resolution,” Atmospheric Meas. Tech. **8**, 2625–2638 (2015).
25. L. Wu, O. Hasekamp, B. Diedenhoven, B. Cairns, J. E. Yorks, and J. Chowdhary, “Passive remote sensing of aerosol layer height using near-uv multiangle polarization measurements,” Geophys. Res. Lett. **43**, 8783–8790 (2016).
26. F. Xu, O. Dubovik, P.-W. Zhai, D. J. Diner, O. V. Kalashnikova, F. C. Seidel, P. Litvinov, A. Bovchaliuk, M. J. Garay, G. van Harten, and A. B. Davis, “Joint retrieval of aerosol and water-leaving radiance from multispectral, multi-angular and polarimetric measurements over ocean,” Atmospheric Meas. Tech. **9**, 2877–2907 (2016).
27. F. Xu, G. van Harten, D. J. Diner, O. V. Kalashnikova, F. C. Seidel, C. J. Bruegge, and O. Dubovik, “Coupled retrieval of aerosol properties and land surface reflection using the airborne multiangle spectropolarimetric imager,” J. Geophys. Res. Atmospheres **122**, 7004–7026 (2017). 2017JD026776.
28. F. Xu, G. van Harten, D. J. Diner, A. B. Davis, F. C. Seidel, B. Rheingans,

M. Tosca, M. D. Alexandrov, B. Cairns, R. A. Ferrare, S. P. Burton, M. A. Fenn, C. A. Hostetler, R. Wood, and J. Redemann, “Coupled retrieval of liquid water cloud and above cloud aerosol properties using the airborne multiangle spectropolarimetric imager (airmspi),” J. Geophys. Res. Atmospheres **123**, 3175–3204 (2018).

29. B. Cairns, E. E. Russell, J. D. LaVeigne, and P. M. Tennant, “Research scanning polarimeter and airborne usage for remote sensing of aerosols,” in *Proceedings of SPIE,* , vol. 5158 (2003), pp. 33–44.
30. B. Cairns, E. E. Russell, and L. D. Travis, “Research scanning polarimeter: calibration and ground-based measurements,” in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation,* (International Society for Optics and Photonics, 1999), pp. 186–196.
31. M. McGill, D. Hlavka, W. Hart, V. S. Scott, J. Spinhirne, and B. Schmid, “Cloud physics lidar: instrument description and initial measurement results,” Appl. Opt. **41**, 3725–3734 (2002).
32. P. Pilewskie, J. Pommier, R. Bergstrom, W. Gore, S. Howard, M. Rabbette, B. Schmid, P. Hobbs, and S. Tsay, “Solar spectral radiative forcing during the southern african regional science initiative,” J. Geophys. Res **108**, 8486 (2003).
33. W. R. Espinosa, L. A. Remer, O. Dubovik, L. Ziemba, A. Beyersdorf, D. Orozco, G. Schuster, T. Lapyonok, D. Fuertes, and J. V. Martins, “Retrievals of aerosol optical and microphysical properties from imaging polar nephelometer scattering measurements,” Atmospheric Meas. Tech. **10**, 811–824 (2017).
34. J. W. Hair, C. A. Hostetler, A. L. Cook, D. B. Harper, R. A. Ferrare, T. L. Mack, W. Welch, L. R. Izquierdo, and F. E. Hovis, “Airborne high spectral resolution lidar for profiling aerosol optical properties,” Appl. Opt. **47**, 6734–6752 (2008).
35. B. Holben, T. Eck, I. Slutsker, D. Tanre, J. Buis, A. Setzer, E. Vermote, J. Reagan, Y. Kaufman, T. Nakajima *et al.*, “Aeronet—a federated instrument network and data archive for aerosol characterization,” Remote. sensing environment **66**, 1–16 (1998).
36. B. N. Holben, J. Kim, I. Sano, S. Mukai, T. F. Eck, D. M. Giles, J. S. Schafer, A. Sinyuk, I. Slutsker, A. Smirnov, M. Sorokin, B. E. Anderson, H. Che, M. Choi, J. H. Crawford, R. A. Ferrare, M. J. Garay, U. Jeong, M. Kim, W. Kim, N. Knox, Z. Li, H. S. Lim, Y. Liu, H. Maring, M. Nakata, K. E. Pickering, S. Piketh, J. Redemann, J. S. Reid, S. Salinas, S. Seo, F. Tan, S. N. Tripathi, O. B. Toon, and Q. Xiao, “An overview of mesoscale aerosol processes, comparisons, and validation studies from dragon networks,” Atmospheric Chem. Phys. **18**, 655–671 (2018).
37. D. J. Diner, M. J. Garay, O. V. Kalashnikova, B. E. Rheingans, S. Geier, M. A. Bull, V. M. Jovanovic, F. Xu, C. J. Bruegge, A. Davis *et al.*, “Airborne multiangle spectropolarimetric imager (airmspi) observations over california during nasa's polarimeter definition experiment (podex),” in *SPIE Optical Engineering+ Applications,* (International Society for Optics and Photonics, 2013), pp. 88730B–88730B.
38. G. van Harten, D. J. Diner, B. J. S. Daugherty, B. E. Rheingans, M. A. Bull, F. C. Seidel, R. A. Chipman, B. Cairns, A. P. Wasilewski, and K. D. Knobelspiesse, “Calibration and validation of airborne multiangle spectropolarimetric imager (airmspi) polarization measurements,” Appl. Opt. **57**, 4499–4513 (2018).
39. C. Wehrli, “Extraterrestrial solar spectrum,” Tech. Rep. WRC Publication No. 615, World Radiation Center (WRC), Davos-Dorf, Switzerland (1985).
40. J. Lean, “Evolution of the Sun's spectral irradiance since the Maunder Minimum,” Geophys. Res. Lett. **27**, 2425–2428 (2000).
41. J. McCorkel, B. Cairns, and A. Wasilewski, “Imager-to-radiometer inflight cross calibration: Rsp radiometric comparison with airborne and satellite sensors,” Atmospheric Meas. Tech. **9**, 955–962 (2016).
42. R. Peralta, C. Nardell, B. Cairns, E. Russell, L. Travis, M. Mishchenko, B. Fafaul, and R. Hooker, “Aerosol polarimetry sensor for the glory mission,” in *Proceedings of SPIE,* , vol. 6786 (2007), p. 67865L.
43. S. Persh, Y. Shaham, O. Benami, B. Cairns, M. Mishchenko, J. Hein, and B. Fafaul, “Ground performance measurements of the glory aerosol polarimetry sensor,” in *Proceedings of SPIE- The International Society for Optical Engineering,* , vol. 7807 (SPIE, P. O. BOX 10 Bellingham WA 98227-0010 USA, 2010).
44. K. Sinclair, B. van Diedenhoven, B. Cairns, J. Yorks, A. Wasilewski, and M. McGill, “Remote sensing of multiple cloud layer heights using multi-angular measurements,” Atmos. Meas. Tech. **10**, 2361–2375 (2017).
45. J. Chowdhary, B. Cairns, M. Mishchenko, and L. Travis, “Retrieval of aerosol properties over the ocean using multispectral and multiangle photopolarimetric measurements from the research scanning polarimeter,” Geophys. Res. Lett. **28**, 243–246 (2001).
46. J. Chowdhary, B. Cairns, and L. Travis, “Case studies of aerosol retrievals over the ocean from multiangle, multispectral photopolarimetric remote sensing data,” J. Atmos. Sci. **59**, 383–397 (2002).
47. J. Chowdhary, B. Cairns, F. Waquet, K. Knobelspiesse, M. Ottaviani, J. Redemann, L. Travis, and M. Mishchenko, “Sensitivity of multiangle, multispectral polarimetric remote sensing over open oceans to water-leaving radiance: Analyses of rsp data acquired during the milagro campaign,” Remote. Sens. Environ. **118**, 284–308 (2012).
48. K. Knobelspiesse, B. Cairns, M. Ottaviani, R. Ferrare, J. Hair, C. Hostetler, M. Obland, R. Rogers, J. Redemann, Y. Shinozuka, A. Clarke, S. Freitag, S. Howell, V. Kapustin, and C. McNaughton, “Combined retrievals of boreal forest fire aerosol properties with a polarimeter and lidar,” Atmospheric Chem. Phys. **11**, 7045–7067 (2011).
49. K. Knobelspiesse, B. Cairns, J. Redemann, R. W. Bergstrom, and A. Stohl, “Simultaneous retrieval of aerosol and cloud properties during the milagro field campaign,” Atmospheric Chem. Phys. **11**, 6245–6263 (2011).
50. S. Stamnes, C. Hostetler, R. Ferrare, S. Burton, X. Liu, J. Hair, Y. Hu, A. Wasilewski, W. Martin, B. van Diedenhoven, J. Chowdhary, I. Cetinić, L. K. Berg, K. Stamnes, and B. Cairns, “Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the mapp algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products,” Appl. Opt. **57**, 2394–2413 (2018).
51. J. Chowdhary, B. Cairns, and L. Travis, “Contribution of water-leaving radiances to multiangle, multispectral polarimetric observations over the open ocean: bio-optical model results for case 1 waters,” Appl. Opt. **45**, 5542–5567 (2006).
52. J. Chowdhary, B. Cairns, M. Mishchenko, and L. Travis, “Using multiangle multispectral photo-polarimetry of the NASA Glory mission to constrain optical properties of aerosols and clouds: results from four field experiments,” Proc. SPIE **5978**, 1–12 (2005).
53. J. Chowdhary, B. Cairns, M. Mishchenko, P. Hobbs, G. Cota, J. Redemann, K. Rutledge, B. Holben, and E. Russell, “Retrieval of aerosol scattering and absorption properties from photopolarimetric observations over the ocean during the clams experiment,” J. Atmos. Sci. **62**, 1093–1117 (2005).
54. M. Ottaviani, B. Cairns, J. Chowdhary, B. Van Diedenhoven, K. Knobelspiesse, C. Hostetler, R. Ferrare, S. Burton, J. Hair, M. Obland, and R. Rogers, “Polarimetric retrievals of surface and cirrus clouds properties in the region affected by the deepwater horizon oil spill,” Remote. Sens. Environ. **121**, 389–403 (2012).
55. B. Cairns, L. Travis, and E. Russell, “An analysis of polarization: ground based upward looking and aircraft/satellite based downward looking measurements,” in *Satellite Remote Sensing of Clouds and the Atmosphere II,* (SPIE, 1997).
56. D. J. Diner, M. J. Garay, C. J. Bruegge, F. C. Seidel, M. A. Bull, V. M. Jovanovic, I. N. Tkatcheva, B. E. Rheingans, and G. van Harten, “Airmspi data quality statement: Podex campaign,” Tech. Rep. JPL D-97927, Jet Propulsion Laboratory, https://eosweb.larc.nasa.gov/project/airmspi/quality_summaries/AirMSPI-Data_Quality_PODEX_V005.pdf (2017).
57. R. Levy, L. Remer, D. Tanré, Y. Kaufman, C. Ichoku, B. Holben, J. Livingston, P. Russell, and H. Maring, “Evaluation of the moderate-resolution imaging spectroradiometer (modis) retrievals of dust aerosol over the ocean during pride,” J. Geophys. Res **108** (2003).
58. K. Knobelspiesse, C. Pietras, G. Fargion, M. Wang, R. Frouin, M. Miller, A. Subramaniam, and W. Balch, “Maritime aerosol optical thickness measured by handheld sun photometers,” Remote. Sens. Environ. **93**, 87–106 (2004).

59. J. Martonchik, D. Diner, R. Kahn, B. Gaitley, and B. Holben, “Comparison of misr and aeronet aerosol optical depths over desert sites,” Geophys. Res. Lett **31**, L16102 (2004).
60. L. Remer, Y. Kaufman, D. Tanré, S. Mattoo, D. Chu, J. Martins, R. Li, C. Ichoku, R. Levy, R. Kleidman, T. Eck, E. Vermote, and B. Holben, “The MODIS Aerosol Algorithm, Products, and Validation,” J. Atmos. Sci. **62**, 947–973 (2005).
61. J. Redemann, Q. Zhang, J. Livingston, P. Russell, Y. Shinozuka, A. Clarke, R. Johnson, and R. Levy, “Testing aerosol properties in modis collection 4 and 5 using airborne sunphotometer observations in intex-b/milagro,” Atmos. Chem. Phys. **9**, 8159–8172 (2009).
62. D. G. Altman and J. M. Bland, “Measurement in medicine: the analysis of method comparison studies,” The Stat. pp. 307–317 (1983).
63. J. M. Bland and D. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” The Lancet **327**, 307–310 (1986).
64. J. M. Bland and D. G. Altman, “Comparing methods of measurement: why plotting difference against standard method is misleading,” The Lancet **346**, 1085–1087 (1995).
65. J. M. Bland and D. G. Altman, “Measuring agreement in method comparison studies,” Stat. methods medical research **8**, 135–160 (1999).
66. J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” Int. J. Nurs. Stud. **47**, 931–936 (2010).
67. E. Olofsen, A. Dahan, G. Borsboom, and G. Drummond, “Improvements in the application and reporting of advanced bland–altman methods of comparison,” J. clinical monitoring computing **29**, 127–139 (2015).
68. D. J. Diner, A. Davis, B. Hancock, G. Gutt, R. A. Chipman, and B. Cairns, “Dual-photoelastic-modulator-based polarimetric imaging concept for aerosol remote sensing,” Appl. Opt. **46**, 8428–8445 (2007).