remote sensing

MDPI

Article

Atmospheric Correction Inter-Comparison eXercise, ACIX-III Land: An Assessment of Atmospheric Correction Processors for EnMAP and PRISMA over Land

Noelle Cremer [1,*], Kevin Alonso [2], Georgia Doxani [1], Adam Chlus [3], David R. Thompson [3], Philip Brodrick [3], Philip A. Townsend [3,4], Angelo Palombo [5], Federico Santini [5], Bo-Cai Gao [6], Feng Yin [7,8], Jorge Vicent Servera [9], Quinten Vanhellemont [10], Tobias Eckert [11], Paul Karlshöfer [11], Raquel de los Reyes [11], Weile Wang [12], Maximilian Brell [13], Aime Meygret [14], Kevin Ruddick [10], Agnieszka Bialek [15], Pieter De Vis [15] and Ferran Gascon [2]

1 SERCO SpA c/o European Space Agency (ESA), European Space Research Institute (ESRIN), Largo Galileo Galilei 1, 00044 Frascati, Italy
2 European Space Agency (ESA), European Space Research Institute (ESRIN), Largo Galileo Galilei 1, 00044 Frascati, Italy
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
4 Department of Forest and Wildlife Ecology, University of Wisconsin, Madison, WI 53706, USA
5 Institute of Methodologies for Environmental Analysis, National Research Council, Area della Ricerca di Potenza Contrada S. Loja, Zona Industriale C.P. 27, 85050 Tito Scalo, Italy
6 Remote Sensing Division, Naval Research Laboratory, Washington, DC 20375, USA
7 Department of Geography, University College London, Gower Street, London WC1E 6BT, UK
8 National Centre for Earth Observation, Space Park Leicester, Leicester LE4 5SP, UK
9 Magellium, 1 Rue Ariane, 31520 Ramonville-Saint-Agne, France
10 Operational Directorate Natural Environment, Royal Belgian Institute for Natural Sciences (RBINS), Vautierstraat 29, 1000 Brussels, Belgium
11 Imaging Spectroscopy Department, Remote Sensing Technology Institute, German Aerospace Center (DLR), Münchener Str. 20, 82234 Weßling, Germany
12 NASA Ames Research Center, Moffett Field, CA 94035, USA
13 HySpex by NEO (Norsk Elektro Optikk), Østensjøveien 34, 0667 Oslo, Norway
14 Centre National d'Etudes Spatiales, 18 Av. Edouard Belin, 31400 Toulouse, France
15 National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK
* Correspondence: noelle.cremer@ext.esa.int

Academic Editor: Carmine Serio

Received: 21 October 2025
Revised: 11 November 2025
Accepted: 12 November 2025
Published: 21 November 2025

Citation: Cremer, N.; Alonso, K.; Doxani, G.; Chlus, A.; Thompson, D.R.; Brodrick, P.; Townsend, P.A.; Palombo, A.; Santini, F.; Gao, B.-C.; et al. Atmospheric Correction Inter-Comparison eXercise, ACIX-III Land: An Assessment of Atmospheric Correction Processors for EnMAP and PRISMA over Land. *Remote Sens.* **2025**, *17*, 3790. https://doi.org/10.3390/rs17233790

Highlights

What are the main findings?

- ACIX-III Land is the first comprehensive inter-comparison of seven atmospheric processors of space-borne imaging spectroscopy missions (EnMAP and PRISMA) over land surfaces, comparing retrievals of aerosol optical depth, water vapour, and surface reflectance over 90 scenes.
- For surface reflectance retrievals, average uncertainty values range between 0.02 and 0.04. While most processors show a good performance over the test sites, some show difficulties in accurately retrieving aerosol optical depth due to a reliance on dark pixels, which were not present in all scenes.

What are the implications of the main findings?

- Although ACIX-III focuses on a limited range of land cover types and aerosol loadings for its match-ups, it serves as a rigorous and comprehensive benchmark for validating and improving atmospheric correction algorithms applied to hyperspectral data.
- This study supports users in selecting the most suitable atmospheric correction processor for their specific application needs.

Abstract

Correcting atmospheric effects on hyperspectral optical satellite scenes is paramount to ensuring the accuracy of derived bio-geophysical products. The open-access benchmark Atmospheric Correction Inter-comparison eXercise (ACIX) was first initiated in 2016 and has now been extended to provide a comprehensive assessment of atmospheric processors of space-borne imaging spectroscopy missions (EnMAP and PRISMA) over land surfaces. The exercise contains 90 scenes, covering stations of the Aerosol Robotic Network (AERONET) for assessing aerosol optical depth (AOD) and water vapour (WV) retrievals, as well as stationary networks (RadCalNet and HYPERNETS) and ad hoc campaigns for surface reflectance (SR) validation. AOD, WV, and SR retrievals were assessed using accuracy, precision, and uncertainty metrics. For AOD retrieval, processors showed a range of uncertainties, with half showing overall uncertainties of <0.1 but going up to uncertainties of almost 0.4. WV retrievals showed consistent offsets for almost all processors, with uncertainty values between 0.171 and 0.875 g/cm^2. Average uncertainties for SR retrievals depend on wavelength, processor, and sensor (uncertainties are slightly higher for PRISMA), showing average values between 0.02 and 0.04. Although results are biased towards a limited selection of ground measurements over arid regions with low AOD, this study shows a detailed analysis of similarities and differences of seven processors. This work provides critical insights for understanding the current capabilities and limitations of atmospheric correction algorithms for imaging spectroscopy, offering both a foundation for future improvements and a first practical guide to support users in selecting the most suitable processor for their application needs.

Keywords: atmospheric correction; inter-comparison; EnMAP; PRISMA; surface reflectance; aerosol optical depth; water vapour; hyperspectral; imaging spectroscopy

1. Introduction

Advances in technology have facilitated the deployment of space-borne hyperspectral imaging satellites capable of capturing contiguous spectral coverage of surface phenomena across numerous spectral bands within the optical and infrared electromagnetic spectrum. While different agencies have deployed Earth-facing spectrometers aboard the International Space Station [1,2], private ventures are actively engaged in launching hyperspectral imagers with enhanced spatial resolution [3,4]. Both the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA) are currently developing their respective suites of hyperspectral imagers, characterised by a Sun-synchronous orbit and a spatial resolution comparable to that of the Landsat and Sentinel-2 missions (NASA-SBG and ESA-CHIME), as well as more specialised imaging spectroscopy missions with a coarser spatial resolution (such as ESA-FLEX). The applications of hyperspectral imaging from space are diverse, with a particular emphasis on sustainable agriculture, food security, and the exploration of surface materials. To ensure the quality of Level 2A products delivered to users, scenes must undergo geometric and atmospheric correction. Given the contiguous and narrow spectral bands extending up to 2500 nm, hyperspectral data are subject to the influence of several atmospheric factors, including absorption by water vapour, oxygen, and carbon dioxide molecules, as well as molecular and aerosol scattering in the shorter wavelengths [5]. The currently operational sun-synchronous hyperspectral imagers, namely, the Italian Space Agency's PRISMA and the German Aerospace Center's EnMAP, provide valuable opportunities to investigate these challenges and evaluate the

performance of existing atmospheric correction processors within a community-based inter-comparison exercise.

The Atmospheric Correction Inter-comparison eXercise (ACIX), a collaborative initiative under the auspices of the Committee on Earth Observation Satellites Working Group on Calibration & Validation (CEOS WGCV), is coordinated between ESA and NASA. To address the diverse challenges posed by different atmospheric conditions, the exercise has been structured into three primary components, ACIX-Land [6,7], ACIX-Aqua [6,8], and the Cloud Masking Inter-comparison eXercise (CMIX) [6,9], focusing on land, water, and cloud scenes, respectively. Following the successful completion of previous iterations, ACIX-III Land aims to provide a comprehensive assessment of state-of-the-art atmospheric correction processors for hyperspectral data acquired over land surfaces. By adhering to community-established guidelines, this exercise facilitates a rigorous comparison of algorithms across various land cover types and atmospheric conditions, focusing on data provided by the EnMAP and PRISMA hyperspectral sensors. Through this comparative analysis, ACIX-III Land seeks to identify the strengths, weaknesses, and areas for improvement in existing algorithms. The insights gained from this exercise will be instrumental in refining atmospheric correction techniques, establishing standardised protocols for inter-comparison, and preparing the scientific community for future hyperspectral missions. Similar outcomes are anticipated from the ongoing ACIX-III Aqua and CMIX-II initiatives.

With numerous hyperspectral missions in orbit or planned, Cal/Val ground networks of hyperspectral radiometers in the visible and infrared spectra (e.g., HYPERNETS [10]) are being set up on the ground. These networks, along with data from ad hoc campaigns, can be leveraged to assess the performance of atmospheric correction (AC) processors for hyperspectral missions. This paper will expand on previous studies within the hyperspectral domain which are mostly limited by a small number of reference sites and processors involved [11–13] by looking at seven different processors over land reference data from five different networks and campaigns, evaluating the retrieval of aerosol optical depth (AOD), water vapour (WV), and surface reflectance (SR). The exercise provides a valuable framework for evaluating the quality and suitability of various in situ reference measurements. It highlights inherent limitations, such as the bias towards bare soil sites with generally low AOD values, and can thus contribute to improving the quality and selection of future Cal/Val sites.

In Section 2 of this paper, a thorough description of community-agreed-upon guidelines for the inter-comparison exercise spanning satellite datasets, reference sites, and accuracy metrics will be given. The main findings will be presented in Section 3, split into parts for both sensors (i.e., EnMAP and PRISMA) and a detailed analysis of each reference dataset. The results will be further discussed in Section 4, while the final section will present the conclusions from the inter-comparison exercise.

2. Materials and Methods

2.1. Atmospheric Correction Processors

Seven atmospheric correction processors participated in the ACIX-III Land exercise. These processors can be very roughly grouped according to their AC process. Physically based inversion approaches (ATREM, PACO, ImaACor, and MAGAC) rely on explicit radiative transfer modelling and retrieve atmospheric parameters through direct inversion or optimal estimation. Semi-empirical methods (ACOLITE/DSF and HYPER SIAC) combine precomputed radiative transfer functions with empirical or data-driven components, while statistical optimisation approaches (GeoNEX-AC) use spectral ratios and aggregated statistics (see Table 1 for a summary of their characteristics). The processors' developers responded to an open call issued by the European Space Agency to participate in the inter-

comparison exercise. While a range of processors is presented here, it is not an exhaustive list of atmospheric correction algorithms for imaging spectroscopy data (for further algorithms, refer to, e.g., [14–17]). All processors had access to the same PRISMA and EnMAP radiance scenes chosen for the exercise (see Section 2.2). Auxiliary data required for all processors, such as illumination and acquisition geometries, spectral response functions, radiometric characteristics, and a DEM for topographic correction, were available to the participants. The same DEM information was distributed to all participants, as was a common approach for the viewing angle computation and information concerning the solar irradiance model for EnMAP (see Section 2.2.2).

A series of atmospheric correction algorithms are available in ACOLITE, and for the present analysis, the Dark Spectrum Fitting (DSF) approach was used, which uses look-up tables generated with the 6SV radiative transfer model. The method relies on the identification of dark pixels across bands with high gas transmittance, assuming a small number of dark pixels exist within the scene in at least one band. Key inputs include per-pixel surface pressure derived from the Copernicus Digital Elevation Model (DEM) and ancillary meteorological data (ozone and water vapour) sourced from GMAO_MERRA2 [18] at the scene centre. AOD is retrieved by fitting the darkest 1% Top-of-Atmosphere (TOA) reflectances (in bands with gas transmittance above 85%, between 500 and 970 nm (PRISMA) or 400–2500 nm (EnMAP)) to 6SV look-up tables. The two bands yielding the lowest AOD are used to select the aerosol model (either continental or maritime) through fitting of the path reflectance to the observed dark targets. Surface reflectance is then computed under the assumption of a homogeneous surface without adjacency or topographic correction. The resulting outputs are surface reflectance for bands with high gas transmittance (>75%), without explicit correction for the bidirectional reflectance distribution function (BRDF).

The ATREM (ATmosphere REMoval Program) algorithm performs atmospheric correction by coupling radiative transfer simulations using either 6S or Improved Pseudo-Spherical Shell (IPSS), a more rigorous radiative transfer model (RTM) developed by [19]. It begins by adjusting the atmospheric model to account for surface elevation and computes solar and observational geometry from image metadata. Solar irradiances above the atmosphere are derived, and scattering effects are simulated under the assumption of no gaseous absorption. Gaseous transmittance is then introduced through a precomputed table of transmittance spectra corresponding to varying water vapour values. For each pixel in a hyperspectral image cube, both water vapour content and surface reflectance are estimated simultaneously by matching the observed spectrum to the simulated transmittance profiles. ATREM retrieves WV using a three-channel ratio algorithm, with the water absorption band centred near 825 nm and nearby atmospheric window bands centred around 800 and 825 nm. The algorithm does not explicitly include aerosol retrieval or adjacency correction. The output consists of per-pixel surface reflectance, derived with pixel-level WV retrieval and accounting for geometric and atmospheric conditions.

GeoNEX-AC (Geostationary NASA Earth Exchange—Atmospheric Correction) is a geostationary adaptation of the MAIAC (Multiangle Implementation of Atmospheric Correction) algorithm. It uses the LibRadtran as the underlying radiative transfer model. The algorithm simulates the regulations of surface BRDF (derived from external MODIS MCD19 (8-day BRDF data) on TOA radiance under various atmospheric conditions based on the Green's Function approach [20]. The Green's functions are calculated for each of the sensor's spectral bands under various illumination-view geometries, AOD and WV concentrations, and atmosphere temperature/pressure profiles. The results are formulated and archived in a set of look-up tables (LUTs) so that TOA radiance at arbitrary conditions can be efficiently and accurately interpolated from the LUTs. AOD retrieval is based on a spectral band ratio algorithm, incorporating MODIS surface BRDF products as auxiliary

information. The hyperspectral scenes are aggregated to ~1 km resolution to match the MODIS data. The algorithm calculates the band ratios between MODIS blue/red vs. SWIR (e.g., 2300 nm) bands. This ratio is used as a reference to estimate AOD from the aggregated hyperspectral images for each 1 km^2 pixel. Finally, the lower value of the median and the mean of the pixel-based AOD is chosen as the AOD estimate for the whole image. Similar to ATREM, GeoNEX-AC uses a ratio algorithm with WV absorption windows at 941 nm and 1128 nm, as well as the neighbouring bands (871 nm, 1048 nm, and 1259 nm) as reference to estimate the intensity of WV absorption. To correct the presence of surface vegetation effects on the absorption intensity, the WV estimates are multiplied by a non-linear factor (<1) calculated based on surface NDVI. The algorithm does not perform post-processing corrections for adjacency or terrain. Final outputs are surface reflectance estimates interpolated from the LUTs, with spectral and geometric effects accounted for through band-specific Green's Function simulations.

HYPER SIAC (Hyperspectral Sensor Invariant Atmospheric Correction) combines external inputs and simulated Sentinel-2 reflectances. 6S emulators serve as the underlying radiative transfer model. Key inputs include monthly cloud-free Sentinel-2 surface reflectance composites, aerosol optical depth from MCD19A2, and aerosol type from the ESA Aerosol_cci Aerosol Climatology. A Sentinel-2 TOA reflectance is simulated in the VIS range from the hyperspectral TOA radiance using the Sentinel-2 spectral response function. WV is estimated from the simulated Sentinel-2 TOA reflectance data. After solving the AOD, aerosol model and WV, a first correction of the hyperspectral TOA reflectance to surface reflectance is conducted. The estimation of WV is then improved using the Atmospheric Pre-corrected Differential Absorption (APDA) method [21] over an initial estimation of the hyperspectral surface and TOA reflectance in the spectral region between 1030 nm and 1160 nm. This improved WV is then used to correct the hyperspectral TOA reflectance to surface reflectance. In the algorithm version used for this paper, no adjacency or topographic corrections are applied for the near infrared (NIR)/short-wave infrared (SWIR) range.

The ImaACor (Image Atmospheric Correction) processor initialises a parametrisation of MODTRAN using standard profiles based on the geographic coordinates of the image centre, as well as the acquisition and illumination geometry. A WV map is generated through a cycle of MODTRAN simulations used to construct a calibration curve that relates the depth of the main WV absorption features to the actual vapour content. This calibration curve is then applied to estimate the WV amount at each pixel across the image. An average value is extracted from the per-pixel estimations. For AOD retrieval, Dark Dense Vegetation (DDV) is identified in the radiance image, and a recursive atmospheric correction is applied to the bands at 660 nm and 2100 nm. The average visibility and WV values are then used to refine the atmospheric parametrisation. To account for both adjacency and topographic effects, the atmospheric correction process combines a simplified model, i.e., excluding environmental contributions, with a complete model that includes them. The simplified correction enables the estimation of environmental functions, which are used to build a point spread function. This function is applied within the full model, allowing for the retrieval of a reflectance image corrected for both adjacency effects and terrain-induced radiometric distortions.

MAGAC (Magellium Atmospheric Correction) is a toolbox developed for hyperspectral satellite data with a flexible interface design that allows application across different sensors. The entire data processing chain relies on emulated atmospheric radiative transfer functions trained with MODTRAN6 simulations. Meteorological inputs from CAMS (reanalysis or forecast) [22] provide ozone, and an initial estimate of WV and aerosol properties at the time and location of image acquisition, while climatological data constrain the

atmospheric state. Aerosol retrieval is conducted with an optimal estimation algorithm following [17,23]. The retrieval is applied to macro-pixels of configurable size (typically 10×10 km^2), where aerosol conditions are assumed homogeneous. Multiple pixels with distinct radiometric characteristics contribute to the cost function, helping to better constrain the retrieval [24]. Besides AOD, aerosols are described by effective optical properties including Ångström exponent and asymmetry parameter, with single scattering albedo fixed from climatology. WV is retrieved pixel-wise using the APDA method [21] on the H_2O band at 940 nm. Surface reflectance is then derived empirically from TOA reflectance using pixel values of geometry and atmospheric conditions. A final post-processing step corrects for adjacency and topographic effects following [25].

PACO (Python-Based Atmospheric Correction) is a MODTRAN 5.4.0-based processor that extends the principles of ATCOR [26]. LUTs of simulated sensor radiative transfer functions are computed and binned according to sensor characterisation, observation conditions, and the atmosphere. The ozone column value is retrieved from an external source (e.g., ECMWF). The processor then classifies the pixels according to pre-defined classes, and DDV pixels are used to retrieve both the atmosphere visibility and the AOD at 550 nm. To identify dark pixels, the SWIR reflectance is calculated assuming a visibility of 23 km, excluding water pixels. Once the surface reflectance in the red and blue bands is known for each dark pixel, the at-sensor radiance for the current viewing and solar geometry is calculated as a function of the visibility using the pre-calculated LUTs. The intersection of the measured radiance with the calculated curve determines the per-pixel visibility. The AOD map is calculated from the visibility. The WV column is determined using the APDA algorithm [21] with bands at WV absorption wavelengths (generally around 820 nm, 940 nm, and/or 1130 nm). After the determination of all the LUTs bins, the BOA reflectance of each pixel is calculated by solving the radiative transfer equation. Adjacency and topographic effects are handled within the processing steps. Although PACO is part of the EnMAP L2 processing chain [27], it is not identical to the EnMAP Level 2A Processor (Atmospheric Correction over Land).

For large-scale data processing or operational business workflow, the computational efficiency of atmospheric correction processors can be a key factor. A direct comparison of processing times is challenging, as hardware and settings (e.g., parallelisation, GPU usage) differ for every participant and are known in different levels of detail. Furthermore, workflows can be adapted to operational usage (e.g., with and without terrain correction). The following summary gives an indication of computational efficiency but cannot be taken as a direct comparison. ACOLITE/DSF is written and executed in Python 3. For ACIX-III Land, the average processing time for PRISMA scenes was 79.1 (+/−18.9) seconds, and for EnMAP it was 73.1 (+/−8.9) seconds. For ATREM, scenes were processed on an Apple Mac with an Intel i9 processor without parallelisation or GPU. The average processing time per scene for both sensors was around 60 s. GeoNEX-AC was executed on the NASA Pleiades Supercomputer (SGI/HPE, Moffett Field, California, USA) (Haswell nodes, 12-core Xeon E5-2680v3 processors) with simple parallelisation by splitting grids, while no Message Passing Interface (MPI) library or GPU were used. The average processing time was around 30 min or less for both sensors. Similarly, processing took around 30 min for the scenes processed with HYPER SIAC, using a computer with one CPU, one GPU, and RAM of 40 GB. Scenes for ImaACor were processed on a 13th Gen Intel(R) Core(TM) i9-13900KF 3.00 GHz with 64.0 GB RAM without GPU or parallelisation. Processing time for one scene took between 2 h and 41 min for EnMAP and 3 h and 4 min for PRISMA scenes. These processing times were split between terrain correction (between 2 h and 39 min and 3 h and 3 min), adjacency correction (depending on AOD between 1 and 1.5 min), and the basic correction (15 s). The MAGAC developers used a Windows 10 64-bit OS, i7-4710 CPU

2.50 GHz machine with 16 GB RAM, without using parallelisation or a GPU. Average scene processing time for both sensors amounted to 20 min. For PACO, scenes were processed on an Intel(R) Xeon(R) CPU E5-2640, 504 GB RAM without GPU or parallelisation. Average processing times for were 18 min and 7 s for PRISMA and 17 min and 43 s for EnMAP.

2.2. Input and Validation Datasets

Level 1 products from two hyperspectral missions (PRISMA and EnMAP) were included in this exercise. All processors submitted results for all proposed scenes from both missions. Match-ups were selected over available validation sites for either atmospheric parameters (AERONET) or hyperspectral surface reflectance measurements (RadCalNet, HYPERNETS, Hypersense campaign, and EnMAP validation campaign). The exercise comprises 90 scenes, with 44 scenes coming from EnMAP (with acquisitions between June 2022 and June 2023) and 46 coming from PRISMA (with acquisitions between January 2021 and June 2023), respectively. While comparing data from (quasi-)synchronous overpasses from both sensors would have been preferable to directly compare both sensors under similar conditions, those overpasses were not available in the scene archives at the time of data collection. The following sections on the validation datasets describe in detail which scenes were chosen (see an overview in Figure 1).

Figure 1. Location of the validation sites over which PRISMA and EnMAP scenes were acquired. The corresponding validation network can be identified by the label's colour.

2.2.1. PRISMA Input Dataset

PRISMA is an optical remote sensing mission with both a panchromatic and a hyperspectral push broom scanning payload. With a ground sampling distance (GSD) of 30 m, it has 66 bands in the visible and near infrared (VNIR) (63 were used in the exercise) and 171 bands in the SWIR, with an absolute radiometric accuracy of better than 5%. The mission products include both an L1 TOA radiometrically and geometrically calibrated radiance image, as well as an L2 geolocated and geocoded atmospherically

corrected reflectance product [28]. The data were downloaded under the ASI licence by the ACIX coordinators and shared with the defined affiliate users, i.e., the ACIX participants. As noted by others [29], the PRISMA geolocation accuracy stated as up to 200 m varies between scenes or even within a single scene. A high geolocation accuracy is needed to accurately match validation sites and satellite pixels; therefore, per-pixel coordinates were adjusted using an image matching algorithm to reduce spatial errors, and datasets were projected into WGS84/UTM at 30 m spatial resolution through nearest neighbour interpolation, where the Copernicus Digital Elevation Model (DEM) was used to match the sensor grid. The processed dataset is named L1G. Beyond projection to UTM, no other changes were made to the radiance datasets, and they retain their original band order and scaling factors. As sensor geometry data are only provided within the PRISMA L2 product, several new attributes were added to the L1G dataset, as follows: (1) the solar geometry (deg), calculated at mean scene acquisition time; (2) the sensor geometry (deg), calculated using satellite positional data provided with the L1 dataset; (3) the surface elevation (meters) per pixel, created using the Copernicus DEM; and (4) a map info string, describing projection, pixel size, and corner coordinates. This L1G product was distributed in its original scene size to participants as an HDF file in order to facilitate processing time and enable a fair comparison (see Table 2).

2.2.2. EnMAP Input Dataset

EnMAP is a German imaging spectroscopy mission operating within the optical spectral range in 224 bands. Within the exercise, 88 VNIR and 154 SWIR bands were used. The dual-spectrometer instrument delivers data with a spatial resolution of 30 m and high spectral and radiometric stability and accuracy [45,46]. Data are provided through the DLR EOWEB® GeoPortal under a stated License Agreement [47]. EnMAP products generated with processor versions higher than 01.03.00 were provided to profit from the improved co-registration error between VNIR and SWIR bands. However, as some participants remarked during the exercise, some residual alignment errors between the VNIR and SWIR spectrometers remain, resulting in slight anomalies in the WV retrieval. Data were used from the start of the mission in April 2022. The data were distributed with the Copernicus DEM at 30 m to ensure the provision of a common DEM for all participants, as well as the common Fontenla 2011 solar irradiance model [48]. At the time of the exercise, information on the viewing zenith (VZA) and azimuth angles (VAA) was not provided within the EnMAP metadata; therefore, a common formula to compute the angles was provided in the form of a Python script by the DLR Ground Segment. Data were provided in their original scene size to participants in a zip folder containing different TIFF files, including the VNIR and SWIR radiance files, as well as quality masks and metadata (see Table 3).

Table 1. Overview of all the processors and their characteristics for ACIX-III Land.

	■ ACOLITE/DSF	■ ATREM	■ GeoNEX-AC	■ HYPER SIAC	■ ImaACor	■ MAGAC	■ PACO
RTM	6S	Improved Pseudo-Spherical Shell (IPSS) [19]	GeoNEX/LibRadtran	6S Emulators	MODTRAN-4/6	MODTRAN6 emulators [30,31]	MODTRAN 5.4
AOD	A single AOD is estimated for the scene using the 1 percentile darkest pixels	No AOD retrievals over land surfaces, but possible retrieval of AOD and aerosol model over water surfaces	Spectral ratio-based; prior information of surface reflectance was obtained from MODIS 8-day BRDF products (MCD19)	SIAC	DDV-based procedure	Multi-pixel optimal estimation, inspired by ISOFIT [17] and the multi-pixel approach from [24]	DDV + Red/NIR dark surfaces
Aerosol Model	Continental and Maritime (6S)	Assumed a Shettle & Fenn 1979 [32] Rural Aerosol Model having an optical depth of 0.1 at 550 nm and a relative humidity of 70%	OPAC https://geisa.aeris-data.fr/opac/, accessed on 20 May 2025	ESA Aerosol_cci Aerosol Model	MODTRAN standard on the basis of image scene (based on Shettle & Fenn 1979 [32])	Aerosol optical properties: single scattering albedo (fixed from climatology), Ångström exponent, and asymmetry parameter	Rural (MODTRAN, based on Shettle & Fenn 1979 [32])
WV	Auxiliary data: GMAO_MERRA2 [18] WV	Three-band ratio algorithm using bands near 800 nm, 825 nm, and 850 nm	ATREM, slightly revised	APDA [21]	MODTRAN-based calibration curve from water absorption features	APDA [21]	APDA [21]
Adjacency Effects	No	No	No	No	Yes	Yes	Yes
Terrain Correction	No	Mean surface elevation	No	No	Yes	Yes	Yes

Table 1. *Cont.*

	■ ACOLITE/DSF	■ ATREM	■ GeoNEX-AC	■ HYPER SIAC	■ ImaACor	■ MAGAC	■ PACO
Specific Auxiliary Data	GMAO_MERRA2 [18] ozone and WV data are used, and the Copernicus DEM is used for a pressure estimate	-	MODIS MCD19 (8-day BRDF)	ESA Aerosol_cci Aerosol Climatology, Sentinel 2 (S2) monthly clean image composite, TOMS and OMI Merged Ozone Data, MCD19 AOD	Optional: Flying zenith and azimuth; DEM for off-nadir and topographic correction	Climatology (MACv3, [33]); CAMS forecasting/reanalysis (EAC4) [22]; instrument noise model parameters; atmospheric RTM emulator object; surface reflectance spectral library (USGS, ECOSTRESS, and ECOSIS)	LST and ozone column: ECMWF, Biome: WWF Terrestrial Ecoregions, Shoreline (optional): GSHHG
Quality Flags	No	No	Yes	Yes	No	Yes	Yes
Version	8 May 2024	4.0	1.0	V0.0.1	6.55	v1.0	1.0.1
Average Processing Times Per Scene *	76 s	60 s	30 min	30 min	15 s (basic) | 75 s (adjacency) | 2 h 51 min (terrain)	20 min	18 min
Licence	GNU General Public License V3 [34]	No licence	GNU General Public License (LibRadtran)	GNU General Public License V3	Free in beta version for validation purposes	TBD, probably Apache v2.0	Proprietary (DLR)
Organisation	Royal Belgian Institute of Natural Sciences (RBINS)	United States Naval Research Laboratory (NRL)	NASA Ames Research Center	University College London (UCL); National Centre for Earth Observation (NCEO)	National Research Council (CNR-IMAA)	Magellium	German Aerospace Center (DLR)
Main Reference	[35]	[5,36–38]	[39]	[40]	[41–43]	-	[44]

* This comparison is based on the processing times of different machines. Please refer to Section 2.1 for a detailed description of processing times and hardware used.

Table 2. L1G geo-corrected PRISMA data structure for ACIX-III.

PRS_L1G_STD_OFFL_SCENE_ID.hdf5			
HDFEOS	ADDITIONAL	FILE_ATTRIBUTES	
SWATHS	PRS_L1_HCO	Data Fields	SWIR_Cube VNIR_Cube
		Geolocation Fields	Latitude Longitude
		Geometric Fields	Sensor_Azimuth_Angle Sensor_Zenith_Angle Solar_Azimuth_Angle Solar_Zenith_Angle
		Terrain Fields	DEM
HDFEOS INFORMATION			
Info			
KDP_AUX			

Table 3. EnMAP data structure for ACIX-III.

ENMAP01____L1C_*SCENE_ID*
ENMAP01____L1C_*SCENE_ID*-DEM.TIF
ENMAP01____L1C_*SCENE_ID*-HISTORY.XML
ENMAP01____L1C_*SCENE_ID*-METADATA.XML
ENMAP01____L1C_*SCENE_ID*-QL_PIXELMASK.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_CIRRUS.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_CLASSES.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_CLOUD.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_CLOUDSHADOW.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_HAZE.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_SNOW.TIF
ENMAP01____L1C_*SCENE_ID*-QL_QUALITY_TESTFLAGS.TIF
ENMAP01____L1C_*SCENE_ID*-QL_SWIR.TIF
ENMAP01____L1C_*SCENE_ID*-QL_VNIR.TIF
ENMAP01____L1C_*SCENE_ID*-SPECTRAL_IMAGE.TIF

2.2.3. AERONET Validation Dataset and Inter-comparison

AERONET (AErosol RObotic NETwork) is a network of ground-based remote sensing aerosol measurement sites, jointly established by NASA and PHOTONS. The network has provided long-term and continuous measurements of AOD and WV properties for over twenty years and has been extensively used for previous ACIX Land and Aqua iterations [49,50]. The retrieval of AOD is a crucial step for any atmospheric correction algorithm, as the presence of aerosols can absorb or scatter the solar radiation reaching an optical sensor on a satellite platform [51]. A diverse range of climate zones and land cover types was selected based on the availability of AERONET measurements and coincident satellite acquisitions. The selection was limited both by the availability of tasked hyperspectral satellite scenes and the disposal of an AERONET measurement within 30 min (±15 min) from the satellite overpass. Furthermore, the choice was driven by the availability of both PRISMA and EnMAP acquisitions on the same station to ease comparison. For PRISMA, a range of sensor zenith acquisition angles spanning 0 to 18° was selected to assess the impact of the acquisition angle. This resulted in a total number of 18 AERONET stations (see Table 4). Estimated AOD and WV values from each processor were compared to the Level 1.5 (cloud-screened) AERONET measurements of the corresponding variable, as this was the highest AERONET processing level available for all stations. For AOD, comparisons were performed at 550 nm, as this is the standard measurement agreed upon

in previous exercises. AERONET AOD was therefore derived by interpolating with the 440 to 870 nm Ångström exponent. To match the measurement to the satellite overpass, the measurements are interpolated within the 30 min time window.

Table 4. The study sites based on the location of AERONET stations.

AERONET Sites	Lat/Lon (°)	Country	Match-Up EnMAP	Match-Up PRISMA
AOE_Baotou	40.9/109.6	China	20 July 2022, 24 October 2022	15 June 2021
Bangkok	13.7/100.5	Thailand	5 March 2023	5 January 2023
Birdsville	−25.9/139.3	Australia	—	18 May 2022
GSFC	39.0/−76.8	USA	7 November 2022	20 September 2022, 19 March 2023
Learmonth	−22.2/114.1	Australia	21 October 2022	28 January 2021, 25 April 2021
Migal	33.2/35.6	Israel	11 June 2022	14 July 2022
NASA_Ames	37.4/−122.1	USA	15 February 2023	3 May 2022, 21 December 2022
NEON_SJER	37.1/−119.7	USA	18 February 2023	4 November 2021
Rome Tor Vergata	41.8/12.6	Italy	—	2 July 2022
SEDE_BOKER	30.9/34.8	Israel	11 June 2022	10 January 2022, 3 June 2022
Sioux_Falls	43.7/−96.6	USA	10 August 2022	22 November 2022
TGF_Tskukuba	36.1/140.1	Japan	3 August 2022	17 February 2021
USDA_ALARC	33.1/−112.0	USA	2 December 2022	20 March 2021, 19 November 2022

2.2.4. RadCalNet Validation Dataset

The Radiometric Calibration Network (RadCalNet, https://www.radcalnet.org/, accessed on 20 May 2025) consists of four stations, which are located in Gobabeb (Namibia), Railroad Valley Playa (USA), La Crau (France), and Baotou (China). Each site is equipped with automated instrumentation to continuously measure surface reflectance with characterised and peer-reviewed uncertainties. At La Crau, an automatic photometric station, the RObotic Station for Atmosphere and Surface (ROSAS), monitors the optical properties of the atmosphere and derives surface reflectance measurements by capturing the upwelling radiance in 12 narrow bands from 380 to 1650 nm. Additionally, a modelled bidirectional reflectance distribution function is available for the site. The La Crau site is managed by the French Centre for Space Studies (CNES) and has a sunny, dry Mediterranean climate with sparse, low vegetation and the presence of pebbles. Changes due to the vegetation cycle are monitored [52,53]. The Railroad Valley Playa site in Nevada currently relies on four ground-viewing multispectral radiometers (GVRs). The GVRs measure in seven bands in the visible and VNIR and one band in the SWIR channels, and are operated by the University of Arizona. Situated in a clay-based playa, the surface conditions are mostly stable, with occasional changes due to rain or snowfall [52]. The site at Gobabeb (Namibia) mirrors the technical setup of the La Crau site and is jointly operated by ESA and CNES. It is situated in a desert environment with a high level of spatial uniformity and clear skies [52,54]. The Baotou site (China) is operated by the Chinese Academy of

Sciences and consists of two separate sites, an artificial site and a natural sandy site, further referred to as Baotou Sand. Due to the sensor GSD requirements, only the Baotou Sand site was considered in this exercise. At the sandy site, two automatic spectrometers measure ground-reflected radiance in the spectral region of 380 to 1080 nm. Baotou is dominated by a semi-arid climate at an altitude of 1270 m and 300 mm average precipitation [52,55].

The spatial homogeneity index (SHI) around the sites is defined as the ratio of the standard deviation to the mean reflectance, calculated over a rolling window of size $L \times L$:

$$\mathrm{SHI} = \frac{\sigma(\rho)}{\bar{\rho}} \tag{1}$$

where $\sigma(\rho)$ and $\bar{\rho}$ represent the standard deviation and mean reflectance within the window, respectively. The SHI for RadCalNet sites is given in increments of 100 m × 100 m, 500 m × 500 m and 1 km × 1 km for all sites except Baotou Sand, as the site extent is measured at 350 m × 350 m. Within these increments, the SHI is given as 3 to 5 % for La Crau, 0.4 to 1.5 % for Railroad Valley Playa, 3 to 5 % for Gobabeb, and 3 % for Baotou Sand. For the dates used in the exercise, aerosol loading was generally low over all RadCalNet sites, with AOD values (given as mean ± standard deviation) of 0.106 ± 0.025 for La Crau, 0.037 ± 0.016 for Railroad Valley Playa, and 0.126 ± 0.101 for Gobabeb. For the single Baotou Sand site, AOD values were measured at 0.147. WV is slightly more varied, with $1.490 \pm 1.160\,\mathrm{g/cm^2}$ for La Crau, $0.766 \pm 0.368\,\mathrm{g/cm^2}$ for Railroad Valley Playa, $1.171 \pm 0.562\,\mathrm{g/cm^2}$ for Gobabeb, and $1.304\,\mathrm{g/cm^2}$ for Baotou Sand.

RadCalNet data are generally provided as a nadir surface reflectance product. For this exercise, CNES preprocessed the data for its sites (La Crau and Namibia) to provide the spectral reflectance in the sensor viewing geometry. For the sites that are collecting the surface reflectance data at multispectral wavelengths, hyperspectral data at 10 nm resolution is derived by combination with representative hyperspectral spectra taken at the sites [52]. Measurement uncertainties are provided within the products downloaded from the RadCalNet Portal [56], while the BRDF-adjusted data supplied by CNES were expanded with the typical error budget for the given site taken from the Uncertainty Analysis Statements. Given the availability of PRISMA and EnMAP scenes, RadCalNet sites were chosen to be included in the exercise. For EnMAP, acquisitions were available for all RadCalNet stations, while there was no matching acquisition over Baotou Sand for PRISMA in the surveyed time period. Adjustment to the sensor viewing angle was performed for the sites of La Crau and Gobabeb. More information on each site is listed in Table 5.

Table 5. The study sites based on the location of RadCalNet stations.

RadCalNet Sites	Lat/Lon (°)	Country	BRDF Correction	Match-Up EnMAP	Match-Up PRISMA
Railroad Valley Playa	38.5/−115.7	USA	No	9 July 2022, 1 September 2022, 25 June 2023	3 February 2021, 4 March 2021, 30 May 2021, 28 June 2021, 25 December 2022
La Crau	43.6/4.9	France	Yes	15 July 2023	21 February 2021, 10 February 2023

Table 5. *Cont.*

RadCalNet Sites	Lat/Lon (°)	Country	BRDF Correction	Match-Up EnMAP	Match-Up PRISMA
Gobabeb	−23.6/15.1	Namibia	Yes	12 June 2022, 13 July 2022, 17 July 2022, 21 July 2022, 2 October 2022, 6 October 2022, 21 October 2022	8 April 2021, 4 July 2021, 29 September 2021, 28 October 2021, 25 December 2021, 14 April 2022, 11 June 2022, 10 July 2022, 5 October 2022, 3 November 2022
Baotou Sand	40.9/109.6	USA	No	16 July 2022	—

2.2.5. HYPERNETS Validation Dataset

HYPERNETS is a recent federated network of hyperspectral ground measurements over water (WATERHYPERNET) and land (LANDHYPERNET), using a newly developed autonomous spectroradiometer system, the HYPSTAR® (RSware OÜ, Tõravere, Estonia) . This instrument has a range from the visible to the near infrared (380 to 1020 nm), with an additional short-wave-infrared module for land sites which extends the range to 1680 nm. In the VNIR, the spectral sampling is 0.5 nm with a spectral resolution full-width half-maximum (FWHM) of 3 nm, while the sampling and FWHM for the SWIR are 3 nm and 10 nm, respectively. The HYPSTAR® is equipped with an automatic pointing system which acquires data over various viewing angles and is consistently processed with a suite of algorithms called the hypernet_processor [10,57]. Each measurement comes with related systematic and random uncertainty estimations, which were included in the exercise. Possible match-ups were identified in cooperation with the HYPERNETS coordinators. LANDHYPERNET data for Gobabeb and Wytham Woods were provided by the National Physics Laboratory (NPL) close to satellite overpass times, as L2B data [57] with additional quality checks by the team. HYPERNETS data are in the early development stage, especially concerning the vegetated sites. The Gobabeb HYPERNETS site exhibits minimal daily variation in surface cover and weather, with stable atmospheric conditions and low precipitation. Located 650 m from the RadCalNet Gobabeb calibration site, the HYPERNETS site lies on a gravel plain adjacent to a dry riverbed separating it from a neighbouring dune sea. Due to the proximity to the RadCalNet site and their analysis, spatial homogeneity between 3 and 5% SHI (spatial homogeneity index) can be assumed. Wytham Woods (United Kingdom) has a long history of scientific research and is managed by the University of Oxford. It is a semi-natural woodland in a protected area, with mostly deciduous trees. The HYPSTAR® instrument is installed on the flux tower at a height of 28 m, with regular, daily data collection at varying viewing zenith and azimuth angles. The site is dominated by the vegetation cycle and therefore highly variable depending on the season and varying degrees of illumination and shadowing. The spatial homogeneity was assessed with airborne hyperspectral, indicating spatial homogeneity over a 100 m × 100 m area. Reflectance at the HYPSTAR® location closely matches the Region of Interest (ROI) average, with differences below 0.01 nm in the visible and 0.01–0.04 nm in the 800–1400 nm range. The Spectral Relative Absolute Error (RAE) is below 10% above 500 nm, confirming representativeness [58]. The measurement protocol for LANDHYPERNET sites is designed to optimise viewing geometries during satellite

overpasses and throughout the day to capture site-specific BRDF characteristics. BRDF modelling at several LANDHYPERNET sites is not currently implemented, but it is being investigated to enhance outlier detection and enable interpolation over missing angles. Previous findings show that the vegetated, heterogeneous Wytham Woods site exhibits greater variability across viewing zenith angles than the homogeneous, near-Lambertian Gobabeb site [57]. Aerosol loading for the HYPERNETS Gobabeb scenes used in the exercise is slightly higher than for the RadCalNet Gobabeb site, with a mean and standard deviation of 0.193 ± 0.148, while WV is comparable at $1.038 \pm 0.369\,g/cm^2$. The Wytham Woods scenes have an AOD of 0.271 and WV of $0.822\,g/cm^2$. Please refer to Table 6 for more information on the selected HYPERNETS sites and satellite match-ups.

Table 6. The study sites based on the location of HYPERNETS stations.

HYPERNETS Sites	Lat/Lon (°)	Country	Match-Up EnMAP	Match-Up PRISMA
Gobabeb	−23.6/15.1	Namibia	16 June 2022, 13 July 2022, 1 September 2022, 9 September 2022, 6 October 2022, 21 October 2022	10 June 2022, 5 October 2022, 3 November 2022
Wytham Woods	52.5/−1.3	UK	—	24 March 2022

2.2.6. Hypersense Campaign Validation Dataset

The Hypersense campaign, also called the ESA CHIME & SBG 2021 mission, was flown between May and July 2021 as a joint effort between ESA, NASA/JPL, and UZH. Measurements were timed at PRISMA overpasses [59]. While an AVIRIS-NG was operated by NASA/JPL, simultaneous ground measurements were taken with an ASD Fieldspec (Braccagni, Italy). Measurements followed an agreed-upon methodology [60] and SGCP Field Campaign Protocol, documenting the target details, photos, cloud cover, acquisition geometry, time, and setup. Analytical Spectral Device (ASD) and white reference measurements were made available and processed by the ACIX coordinators and cover the VNIR and SWIR spectral range. The standard deviation of the repeated measurements was included as the uncertainty value of the measurements. All measurements from the campaign were provided in 1 nm resolution between 350 and 2500 nm. AOD and WV measurements are not available for these sites.

Measurements in Braccagni (Italy) were made by a team from the Institute of Methodologies for Environmental Analysis at the Italian National Research Council (CNR IMAA) and the University of Tuscia in an agricultural area over four distinct fields, covering lightly wet bare soil without vegetation cover, further referred to as bare_wet_soil; a summer wheat crop, dry with yellow ears of 70 to 80 cm height, referred to as summer_wheat; a dry, bare, and light soil without vegetation cover, further called bare_dry_soil; and a brownish bare, dry soil without vegetation cover, referred to as bare_reddish_soil. Spatial homogeneity was not assessed during the campaign, but measurements were taken in approximately homogeneous agricultural fields, varying in size between 300 m × 600 m (bare_wet_soil) and 450 m × 600 m (summer_wheat). These fields are generally bordered by fields in diverse states of cultivation.

Near Camarena (Spain), a team from the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) measured three different targets in an agricultural area: a first target over bare ploughed soil without vegetation cover (BareSoil1), a second target over bare ploughed soil without vegetation cover (BareSoil2), and a field of dry vegetation, cut dry after harvest with a remaining vegetation crop of 5 to 10 cm height

and crop cuttings on the field (DryVeg). Again, spatial homogeneity was not explicitly assessed. Measurements were taken in the middle of elongated agricultural fields, surrounded by field at a different state of cultivation. These fields varied in length (200 to 2000 m) and width (30 to 50 m).

Near Rio Tinto (Spain), measurements were performed by a team from the Estación Biológica de Doñana (EBD - CSIC) in a mining pond. For both measurements, the target is of a homogeneous mineralogical composition made up mostly of quartz as the main mineral and chlorite–smectite and jarosite as accessory phases. The area in the pond was assessed and deemed spatially homogeneous in an area of 100 m × 100 m around the measurement point. Dates and location for each site are shown in Table 7.

Table 7. The study sites based on the location of Hypersense Campaign stations.

Hypersense Campaign Sites	Lat/Lon (°)	Country	Match-Up PRISMA
Camarena	40.0/−4.1	Spain	30 June 2021
Rio Tinto	37.9/−6.6	Spain	25 June 2021
Braccagni	42.8/11.1	Italy	4 June 2021

2.2.7. EnMAP Campaign Validation Dataset

The EnMAP Product Validation team at Helmholtz Centre for Geosciences (GFZ) collected data for a validation campaign for EnMAP in 2022 and 2023 with cooperating site teams [61] and a consistent measurement protocol [62]. All sites were assessed to have a highly spatially and spectrally homogeneous area of least 90 m × 90 m, as well as to exhibit temporal stability, flat terrain, and minimal shadowing. The sites were chosen to be free from nearby large water bodies and adjacency effects and composed of stable surface materials unaffected by measurement activities [62]. These data were shared with the ACIX coordination team. ASD measurements were taken in the VNIR and SWIR spectral regions, covering all relevant spectral EnMAP bands. The data were provided already spectrally resampled to the EnMAP spectral range and resolution, between 418 and 2445 nm. Indications of measurement uncertainty were not available for these measurements. These sites include Pinnacles (Australia), a spatially uniform desert site with high reflectance values [63]; Amiaz Plain and Makhtesh Ramon (Israel), two homogeneous flat desert sites [64]; Black Rock Playa (USA), a sizeable and homogeneous playa site [65]; TXL Airfield Berlin (Germany), a sealed airfield surface at the airport of Berlin measured by GFZ; Nansen Ice Shelf (Antarctica), a homogeneously snow covered flat area with a strong influence of katabatic wind [66]; and additional measurements from Railroad Valley Playa as part of RadCalNet described in Section 2.2.4. AOD and WV measurements are not available for these sites. All campaign sites and dates are shown in Table 8.

Table 8. The study sites based on the location of EnMAP Validation (EnVal) Campaign stations.

EnVal Campaign Sites	Lat/Lon (°)	Country	Match-Up EnMAP
Antarctica	−74.7/163.4	Antarctica	11 December 2022
Berlin	52.6/13.3	Germany	24 July 2022
Makhtesh Ramon	30.6/34.8	Israel	8 July 2022
Amiaz Plain	31.1/35.4	Israel	23 August 2022
Black Rock Playa	40.9/−118.9	USA	28 June 2022, 29 July 2022
Railroad Valley Playa	38.5/−115.7	USA	21 July 2022
Pinnacles	−30.6/115.2	Australia	20 August 2022, 6 December 2022, 5 March 2023

2.3. Inter-comparison and Validation Methodology

2.3.1. Inter-comparison and Validation of Aerosol Optical Depth and WV

As with previous ACIX implementations, the inter-comparison analysis for AOD and WV was performed on a 9 km × 9 km image subset, which was centred on the AERONET station. This area was selected to contain a sufficient number of pixels at a 30 m sensor resolution and provide spatially homogeneous atmospheric conditions [7].

AERONET AOD and WV data were processed and matched to satellite retrievals over the image subset as described in Section 2.2.3. For scenes where RadCalNet surface reflectance ground measurements were used for the SR inter-comparison, the corresponding AOD and WV measurements were downloaded and added to the exercise. To better place the results in a wider context, reference specifications for both AOD and WV were added to the plots. For AOD, the empirical uncertainty of MODIS AOD retrievals [67], defined as

$$\Delta AOD_{550,Specification} = 0.15 \times AOD_{550,Reference} + 0.05 \quad (2)$$

was used. For WV, the specification

$$\Delta WV_{Specification} = 0.1 \times WV_{Reference} + 0.2\,g/cm^2 \quad (3)$$

was adapted from the empirical uncertainty associated with Sentinel-2 WV retrievals [25]. An average AOD and WV value over the 9 km × 9 km subset was calculated for each processor and scene and then compared to the corresponding AERONET/RadCalNet measurement. For each processor, the two retrieved variables were plotted against reference values for all sites, and statistical metrics were calculated (see Section 2.3.3 for details). As there was a strong bias towards low AOD in the dataset, a linear fit could not be assumed, and the AOD retrievals were assessed with accuracy, precision, and uncertainty (APU) metrics per 0.05 validation bin. For WV, a scatterplot was used to depict the relationship between estimated and reference value, using precision, uncertainty, and R^2 as metrics.

2.3.2. Inter-comparison and Validation of Surface Reflectance

All surface reflectance ground measurements were spectrally resampled to the spectral resolution of the satellite sensors by applying a spectral convolution with the sensor's response function. The only exception was for the BRDF-adjusted RadCalNet measurements from Gobabeb and La Crau supplied by CNES, as these were already delivered as a spectrally resampled product. For the comparison, the four closest pixels to the station location were selected and their reflectance values retrieved. This number of pixels was chosen based on recommendations by site operators and a sensitivity analysis. For a small number of sites, the number of pixels was reduced due to heterogeneous terrain (for a discussion on this issue, please refer to Section 4.1). The estimated reflectances for each processor were then averaged over each target area, and the mean and standard deviation were compared to the validation measurement and its uncertainty. This study focuses on the comparison of each processor to the corresponding in-situ measurement. For more information on direct processor by processor comparisons, please refer to the Appendix A.

The Spectral Angle Mapper (SAM) metric was used to give an indication of the similarity between retrieval and reference across the whole spectrum. For the general assessment, APU plots were used to compare the processors' performances. To contextualise the SR retrieval, a reference specification for APU metrics was chosen as

$$\Delta\rho_{Specification} = 0.05 \times \rho_{Reference} + 0.005, \quad (4)$$

as used in previous ACIX iterations [6,7,68].

2.3.3. Inter-comparison Metrics

The metrics used to assess AOD, WV, and SR are shown in more detail in this section. The SAM computes the angle between the spectra, which is indicative of their similarity

$$\mathrm{SAM}(s_1, s_2) = \arccos\left(\frac{s_1 \cdot s_2}{||s_1|| \cdot ||s_2||}\right). \tag{5}$$

Here, s_1 and s_2 refer to two different spectra, i.e., the retrieved and the reference spectrum. SAM is computed for each processor and site.

The accuracy (A), precision (P), and uncertainty (U) metrics assess the processors' performance in retrieving AOD, WV, and SR. The APU framework was chosen for this inter-comparison as it provides a multidimensional assessment of model performance by evaluating how close predictions are to the truth (accuracy), how repeatable they are (precision), and how far predictions are from the truth on average (uncertainty). Conventional metrics (e.g., Root Mean Squared Error (RMSE), bias) summarise total error in a single value, while APU gives a more comprehensive and diagnostic measure of algorithm performance. They are computed starting with the difference between the retrieved value y_i and the reference $y_{\mathrm{REF},i}$ for observation i, i.e., $\Delta_{\mathrm{REF},i} = y_i - y_{\mathrm{REF},i}$. Then APU metrics are calculated based on Equations (6) to (8):

$$\mathrm{Accuracy}(A) : A = \frac{1}{n}\sum_{i=1}^{n} \Delta_{\mathrm{REF},i}, \tag{6}$$

$$\mathrm{Precision}(P) : P = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (\Delta_{\mathrm{REF},i} - A)^2}, \tag{7}$$

$$\mathrm{Uncertainty}(U) : U = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \Delta_{\mathrm{REF},i}^2}. \tag{8}$$

The APU metrics are given in the respective units, i.e., reflectance [0 . . . 1] or % for SR, g/cm^2 for WV, and unitless for AOD.

Following these definitions, the accuracy [Equation (7)] is equivalent to the bias, while the uncertainty [Equation (8)] is equivalent to the RMSE, respectively, defined as

$$\mathrm{BIAS} = \frac{1}{n}\sum_{i=1}^{n} (y_i - y_{\mathrm{REF},i}), \tag{9}$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - y_{\mathrm{REF},i})^2}. \tag{10}$$

For the assessment of WV, Pearson's correlation coefficient R is defined as follows:

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{11}$$

where x_i and y_i are the individual data points for the retrieved value over area i and the predicted value over area i, respectively. $\bar{x}_i$ and $\bar{y}_i$ are the means of variables x_i and y_i. The predicted value here refers to the value predicted by the linear regression.

2.3.4. CEM-PAL Processing Environment

The Copernicus Expansion Mission-Platform Algorithm Laboratory (CEM-PAL) is an open-source scientific computing environment, developed to support the evolution and prototyping of L2 products and algorithms for Copernicus Expansion Missions. It is an online cloud platform infrastructure, which provides storage and computing resources

using open-source processing software. It is a collaborative environment, where users can collaborate on code, connect to repositories, and create shareable dataset catalogues. Different environments can be set up for code development and data analysis [69]. All input data used in ACIX-III (i.e., satellite products, validation datasets, participant submissions) were stored on the workspace and analysed on the same platform. In this way, the established environment and shared resources could be exploited and easily shared within the coordination team. The ACIX-III Land catalogue and validation scripts can further be used within the CEM-PAL to support CHIME L2A development and other missions (e.g., SBG, LSTM).

3. Results

3.1. AOD Validation over AERONET/RadCalNet Sites

The AOD retrieval is assessed over PRISMA and EnMAP scenes to achieve a statistically significant number of samples, yielding 78 samples across AERONET and RadCalNet sites. All processors except ATREM submitted AOD values to assess for the exercise. For ATREM, surface spectral reflectances, aerosol models, and optical depths can, in principle, be retrieved simultaneously with a spectrum-matching technique. However, because of the low signal-to-noise ratios in both PRISMA and EnMAP data, AOD retrievals were not submitted to the exercise. The PACO processor flags its aerosol retrievals, resulting in only 22 samples included in this comparison, as all other AOD values were not retrieved from the DDV method and a fallback to dark surfaces was used. Most processors make use of a variation of the DDV algorithm to retrieve AODs, as seen in Table 1. For a description of each processor's specific approach to AOD retrieval, please refer to Section 2.1.

As shown in Table 9, HYPER SIAC achieves the overall best performance, with the best metrics in accuracy, uncertainty, and the highest percentage of uncertainty within the specification. MAGAC (with the best precision) and PACO perform similarly well. ACOLITE/DSF shows the highest overestimation of values. More detailed results in the form APU plots are shown in Figure 2. Results are binned to 0.05 AOD intervals with the number of reference measurements in each bin illustrated by a histogram (measurement counts shown on the left y-axis). Accuracy (blue), precision (green), and uncertainty (orange) values are depicted for each bin and calculated as described in Section 2.3.1. The reference specification is shown in the light brown shaded area. The processors retrieve AOD with an accuracy between 0.039 and 0.371 (median 0.089), a precision between 0.017 and 0.079 (median 0.051), and an uncertainty between 0.059 and 0.382 (median 0.120). Depending on the processor, between 2.6% and 35.9% (median 30%) of uncertainty values fall within the specification limits. HYPER SIAC (0.059), PACO (0.080), and MAGAC (0.093) show the lowest uncertainty values, with ACOLITE/DSF (0.382) showing the biggest discrepancy between retrieved and measured AOD. For ACOLITE/DSF, there is no fallback mode for the aerosol retrieval so, for image acquisitions over bright desert environments with a lack of DDV or water pixels, the aerosols are strongly overestimated. GeoNEX-AC utilises a fallback value for a significant number of scenes when retrieving AOD, which contributes to a slightly less accurate performance. Generally, the percentage of uncertainty values within the specifications is quite low. However, from the APU graphs in Figure 2, two relevant observations arise. Firstly, uncertainty values are close to the upper limit of the reference specification for a number of bins, especially for HYPER SIAC and PACO retrievals. Secondly, the majority of retrievals are between 0 and 0.15 AOD, where the specification limits are very small. For HYPER SIAC, MAGAC, and PACO, the retrieval is within or close to the specification limit for low (<0.2) and high AODs (>0.4), with slightly higher uncertainties for the 0.2–0.25 bin (PACO) and the 0.25–0.3 AOD bin (HYPER SIAC, MAGAC). For GeoNEX-AC and ImaACor, uncertainties increase for AODs higher than 0.2.

ACOLITE/DSF shows a generally strong overestimation of values. HYPER SIAC, ImaACor, MAGAC, and PACO show an underestimation of the highest retrieved AODs, while all other AODs are generally overestimated.

Table 9. Averaged accuracy (A), precision (P), and uncertainty (U) (dimensionless) of AOD retrievals compared to ground measurements for each processor. U in specs % refers to the percentage of uncertainty scores within the reference specifications. For each metric, the best overall performance is shown in bold. ([a] PACO has overall best precision but only retrieved AOD for 22 scenes.)

	ACOLITE	GeoNEX-AC	HYPER SIAC	ImaACor	MAGAC	PACO
A	0.371	0.119	**0.039**	0.103	0.071	0.075
P	0.078	0.062	0.039	0.079	**0.037**	0.017 [a]
U	0.382	0.157	**0.059**	0.146	0.093	0.080
U in specs %	2.5	15.4	**36.0**	28.2	33.3	31.8

3.2. Water Vapour Validation over AERONET/RadCalNet Sites

For the evaluation of WV retrieval, six processors were compared. Their WV retrievals are assessed against AERONET and RadCalNet measurements. ACOLITE/DSF does not retrieve WV but uses ancillary data derived from GMAO_MERRA2 [18] instead; therefore, it is not included in the comparison. Processors generally use ratio algorithms or the APDA method [21]. Please refer to Section 2.1 for more details on each processor.

The results are shown in Figure 3. The bigger scatterplot on the left side zooms in on the WV measurements between 0.2 and 3 g/cm^2, while the smaller plot on the bottom shows the full range of measurements (between 0 and 6 g/cm^2). The error metrics are indicated in the top-right corner, while an error histogram is shown in the right middle. Within the scatterplots, the dashed brown line indicates the 1:1 line, representing a perfect correspondence between the estimated and reference values. The light brown area shows the reference specification (3), adapted from the empirical uncertainty associated with Sentinel-2 WV retrievals [25]. Generally, the results (see Table 10) are consistent with R^2 values higher than 0.9, except for ImaACor with an R^2 values of 0.667, which reflects the more widely dispersed outliers. ATREM, MAGAC, and PACO show similar results with uncertainty between 0.267 and 0.399 g/cm^2, with an accuracy throughout the measurements between 0.137 and 0.289 g/cm^2 and a positive offset. GeoNEX-AC is the only sensor showing a strong negative offset, with consistently underestimated values and an accuracy of -0.355 g/cm^2. Although GeoNEX-AC uses an approach adapted from the ATREM processor, the retrievals are overestimated in the case of ATREM and underestimated for GeoNEX-AC. HYPER SIAC shows a very well-performing retrieval, with low accuracy (-0.028 g/cm^2) and low uncertainty (0.171 g/cm^2).

3.3. Surface Reflectance Validation for EnMAP

For each processor, Figure 4 shows the APU results averaged over all sites included in the exercise for the EnMAP sensor. The plots show the accuracy (blue), precision (green), and uncertainty (orange). In these graphs, all wavelengths submitted by the processors are displayed; hence, there might be slight differences between the graphs, as not all processors output the same wavelengths. All wavelengths submitted by each processor are averaged over 10 nm. The reference specification (4) [68], averaged over all reference sites, is displayed in the brown shaded line and lies between ± 2 and 3 % across the spectrum. For most processors, the average accuracy, precision, and uncertainty stay within the reference specifications, with uncertainty and precision often being situated at the upper limit of the specification. Outliers occur around the water absorption features at 1450 nm

and 1950 nm, as well as around major water absorption features at approximately 950 nm and 1200 nm [70] for the processors that output a reflectance value for those wavelengths.

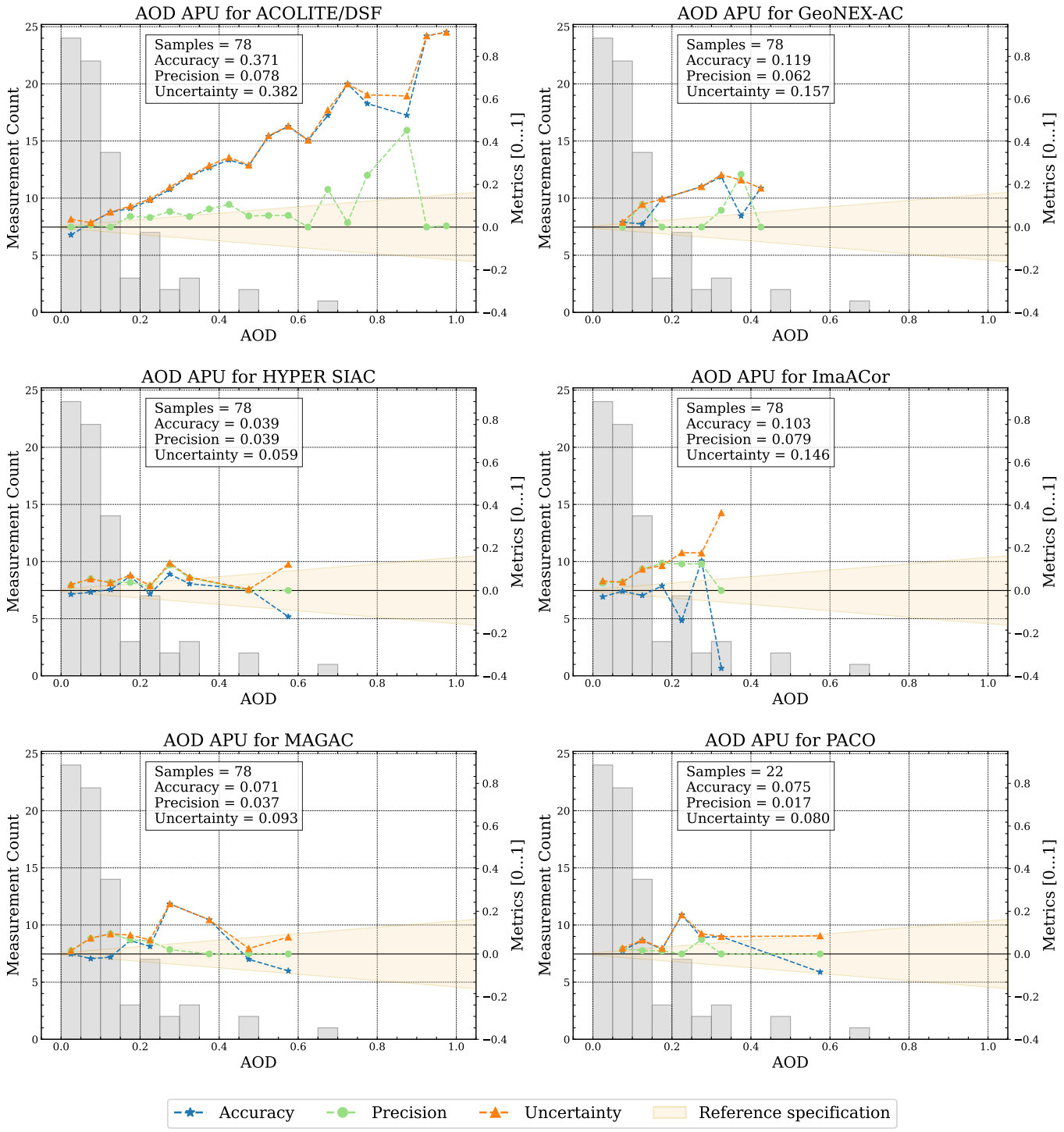

Figure 2. APU plots for AOD retrieved over AERONET and RadCalNet stations for both EnMAP and PRISMA sensors, shown with 0.05 value bins. The reference specification refers to (2).

Spectral regions strongly affected by water vapour absorption exhibit near-saturation of H_2O absorption, where TOA radiances approach zero. Consequently, atmospheric correction in these bands yields unrealistically large relative errors due to divisions by values close to zero. Generally, the performance for SWIR bands (>1900 nm) is a bit weaker for all sensors compared to the shorter wavelengths, with the metrics falling slightly above

(or below) the specification area. For some processors, such as ACOLITE/DSF, ATREM, and HYPER SIAC, a larger underestimation (up to 10% accuracy) can be observed for the end of the spectrum (wavelengths bigger than 2300 nm). HYPER SIAC was updated in response to these findings, mitigating the underestimation issues in the affected spectral regions. With this update, the hyperspectral surface reflectance over the NIR and SWIR spectral range is added as a post-processing step, as the processor's RTM (6S) is not able to simulate the gas absorption effects at the fine spectral resolution required for hyperspectral image corrections. However, this updated version is not included in the current comparison.

Table 10. Averaged precision (P), uncertainty (U) (in g/cm^2), and R^2 of WV retrievals compared to ground measurements for each processor. For each metric, the best overall performance is shown in bold.

	ATREM	GeoNEX-AC	HYPER SIAC	ImaACor	MAGAC	PACO
A in g/cm^2	0.196	−0.355	**−0.028**	0.499	0.288	0.137
U in g/cm^2	0.338	0.525	**0.170**	0.875	0.399	0.267
R^2	0.946	0.936	**0.973**	0.667	0.927	0.949

While accuracy values are mostly within the reference specification bounds, uncertainty and precision are slightly above the bounds for ACOLITE/DSF. Strong outliers are present around 950 nm and at the end of the spectrum. For ATREM, accuracy, precision, and uncertainty values lie within and at the upper edges of the reference specification, with the exception of the very short (<500 nm) and long (>2300 nm) wavelengths and some outliers around 1150 nm and the mentioned water absorption bands. The retrieval is generally smooth and consistent within wavelengths. Reflectances are generally slightly underestimated until 1500 nm and overestimated from 1500 nm to 2300 nm. GeoNEX-AC's retrieval differs slightly from that of the other processors. Accuracies are mostly within the reference specification bounds, with overestimations for the shorter wavelengths until 1100 nm, underestimations between 1100 nm and 2250 nm, and underestimations for wavelengths > 2300 nm. Uncertainty and precision are more strongly out of the specification for short wavelengths (<600 nm), between 750 nm and 1000 nm, and then again for longer wavelengths > 2000 nm. With the exception of some outliers at water absorption features around 1750 nm and 1950 nm, the metrics stay below 4%, and this is also the case for the longer wavelengths. HYPER SIAC shows very smooth retrievals, with most metrics being within the reference specification throughout the spectrum. Wavelengths at the water absorption bands are not retrieved by the processor. Until 1450 nm, retrievals are slightly underestimated. They are slightly overestimated between 1500 nm and 1700 nm. Larger underestimations occur in two spectral ranges, first between 2000 nm and 2100 nm and then more pronouncedly between 2300 nm and 2500 nm. Retrievals for ImaACor show accuracies mostly within the reference specification except for very short (<500 nm) and long (>2300 nm) wavelengths. Uncertainty and precision are at the upper bounds of the specification and also slightly above the bounds for very short and long wavelengths. Outliers can be seen at water absorption features around 950 nm, 1450 nm, 1750 nm, and 1950nm. MAGAC shows a very similar pattern, with most metrics within the reference specification. There are some outliers at around 450 nm, 760 nm, 820 nm, and 2050 nm, as well as the water absorption bands. The metrics are slightly beyond the specifications for wavelengths longer than 2300 nm. Excluding very short (<500 nm) and long (>2300 nm) wavelengths, the retrievals for PACO also show accuracies mostly within the reference specification, and uncertainty and precision are at the upper bounds of the specification. Outliers can be seen around 760 nm and at the water absorption features around 950 nm, 1450 nm, 1750 nm, and 1950 nm.

A summary of uncertainty results can be seen in Figure 5. Higher uncertainties for ACOLITE/DSF throughout the spectrum, and for GeoNEX-AC in the VNIR become apparent, as well as the overall higher uncertainties in the higher wavelengths > 1400 nm and especially >2000 nm.

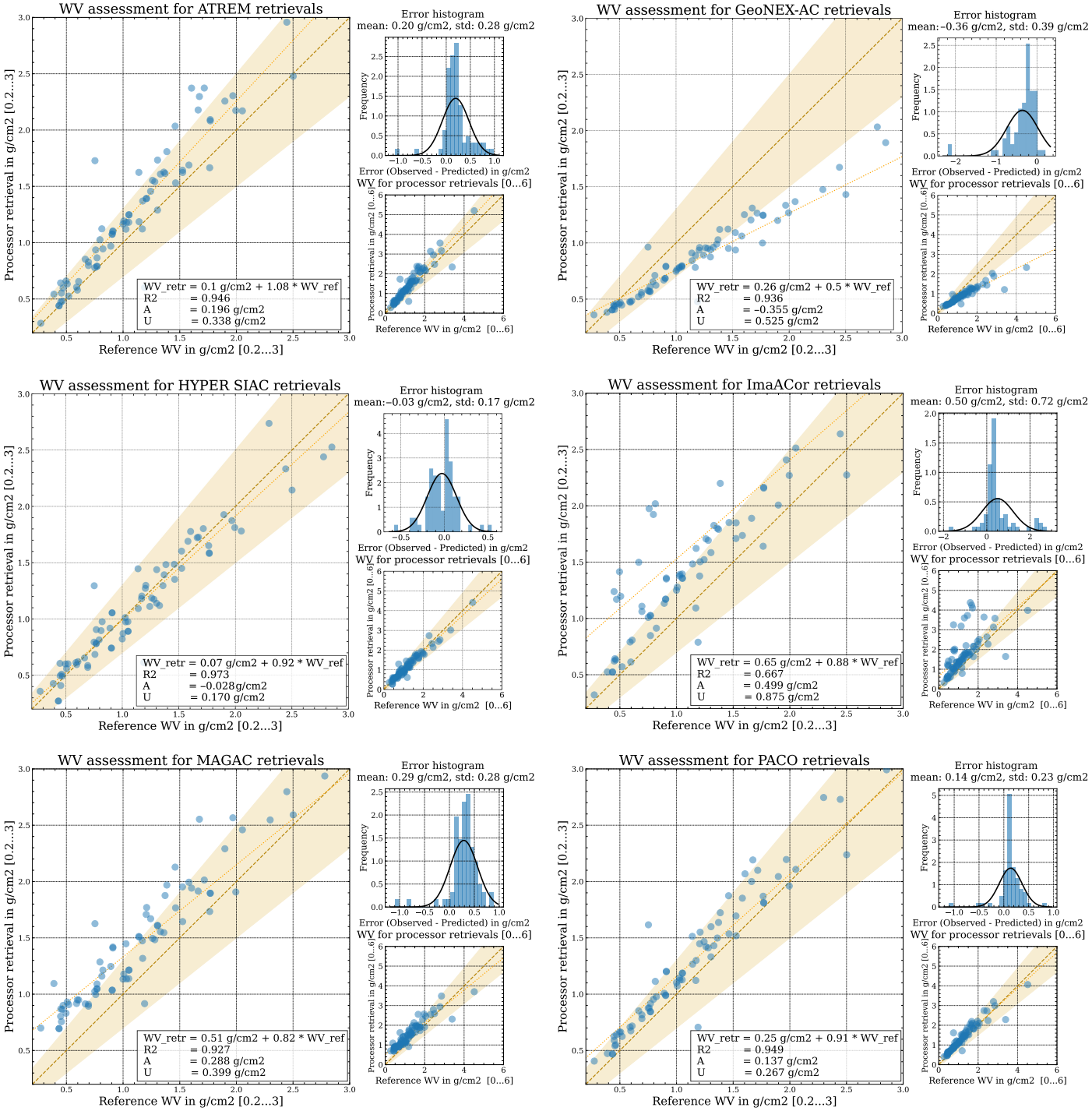

Figure 3. Scatterplots (with ranges 0.2 to 3 g/cm^2 for the bigger plot on the left and 0 to 6 g/cm^2 for the smaller plot on the lower right) and error histogram (on the upper right) of WV retrieval and WV reference values from AERONET/RadCalNet. The dashed line represents the 1:1 agreement, while the shaded brown area shows the uncertainty limits specifications, i.e., (3). The dotted orange line shows the regression line.

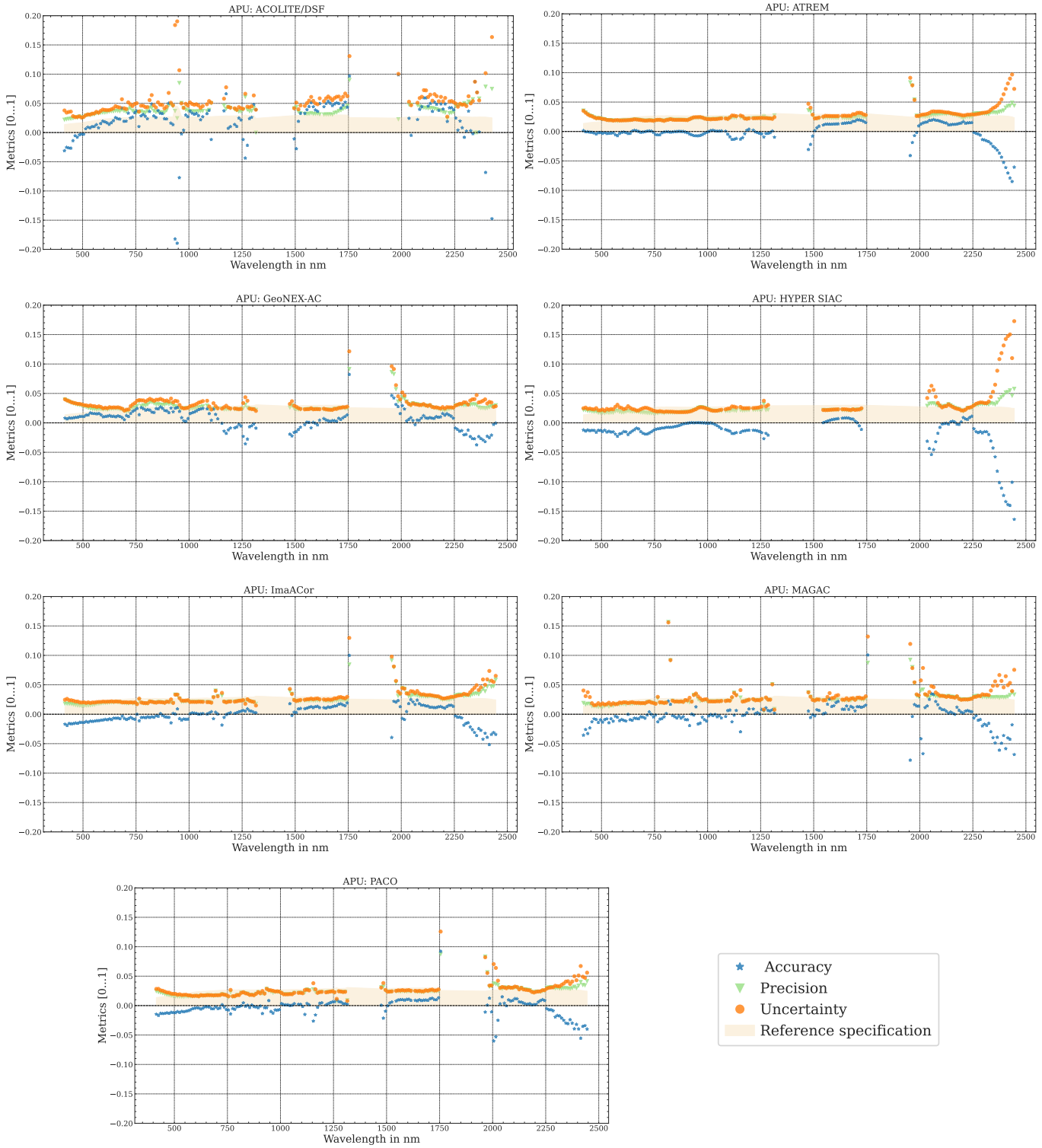

Figure 4. Accuracy, uncertainty, and precision (APU) plots for EnMAP SR retrieved over RadCalNet and HYPERNETS stations, as well as for the EnMAP validation campaign. All submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4).

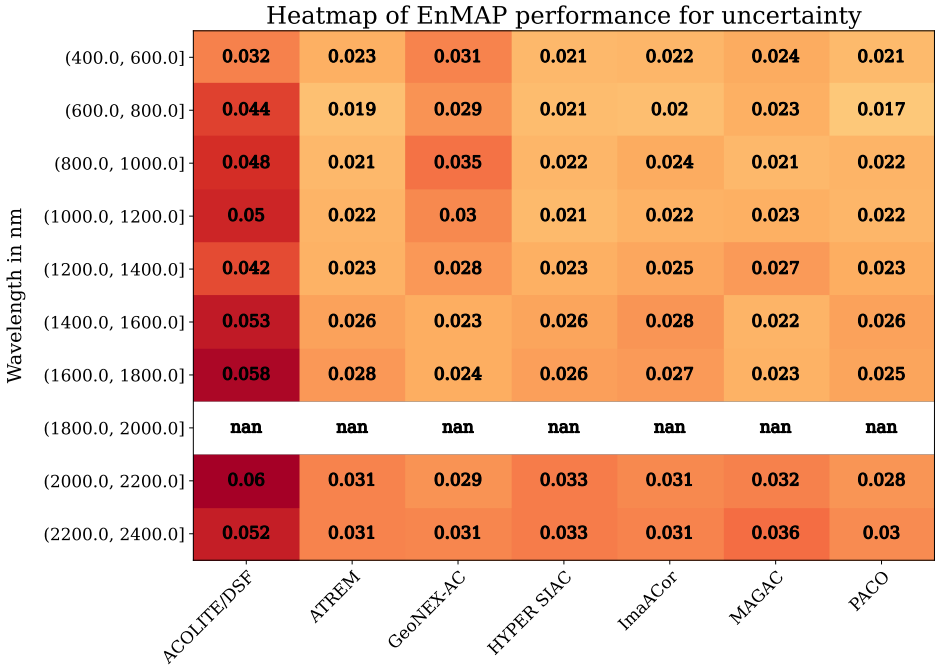

Figure 5. Heatmap of average uncertainties over all EnMAP sites for all processors. Only bands common to all processors (excluding water absorption regions) are included in the computation of the uncertainty. A darker colour corresponds to a higher uncertainty value.

3.3.1. Surface Reflectance Validation Using RadCalNet Measurements for EnMAP

For EnMAP, all RadCalNet sites were covered in the exercise (see Table 5). Figure 6 shows an average uncertainty over all scenes taken over the individual sites for all processors in one graph. The uncertainty values are binned to 10 nm, and only wavelengths submitted by all processors are shown in these results. The mean over all processors (corrected for outliers) is shown with the dashed black line, indicating which processors perform better or worse than average. The light brown area indicates the reference specification (4).

For Baotou Sand, the uncertainty values throughout the spectrum roughly follow the same trajectory for all processors, with peaks around 450 nm, 600 nm, 775 nm, and 875 nm and corresponding valleys at 500 nm, 700 nm, and 850 nm. The mean uncertainty oscillates between 1% and 3% and lies mostly 0.5% to 1.5% above the reference specification. HYPER SIAC and PACO lie mostly within the specification, while GeoNEX-AC shows uncertainties about 1 pp (percentage points) higher than all other sensors throughout the spectrum. Gobabeb shows good performances for all processors, with a larger uncertainty for ACOLITE/DSF. The average performance is within the reference requirements of below 3% (with exception of the very short wavelengths < 475 nm) and so is the individual processor performance. Uncertainties are especially low for ATREM (up to 1300 nm) and HYPER SIAC. ACOLITE/DSF shows uncertainties between 2% and 6%. Performances for La Crau are mostly within the reference specification until 1300 nm and below 1% uncertainty for ATREM, ImaACor, and HYPER SIAC, while the average is above the specification for longer wavelengths. For the longer wavelengths, the uncertainty is up to 5% for the processors ATREM and ACOLITE/DSF. The results over Railroad Valley Playa show a similar pattern, with mean values below or only slightly above the reference specification for the VNIR and above the boundary for wavelengths > 1300 nm. Similarly to Gobabeb, ACOLITE/DSF shows higher uncertainty with values up to 10%. GeoNEX-AC also shows slightly higher values (up to 4%) in the VNIR but is within specification

between 1500 and 1750 nm. All other processors are around or below 2% uncertainty for the spectrum below 1300 nm and between 2% and 4% for the longer wavelengths.

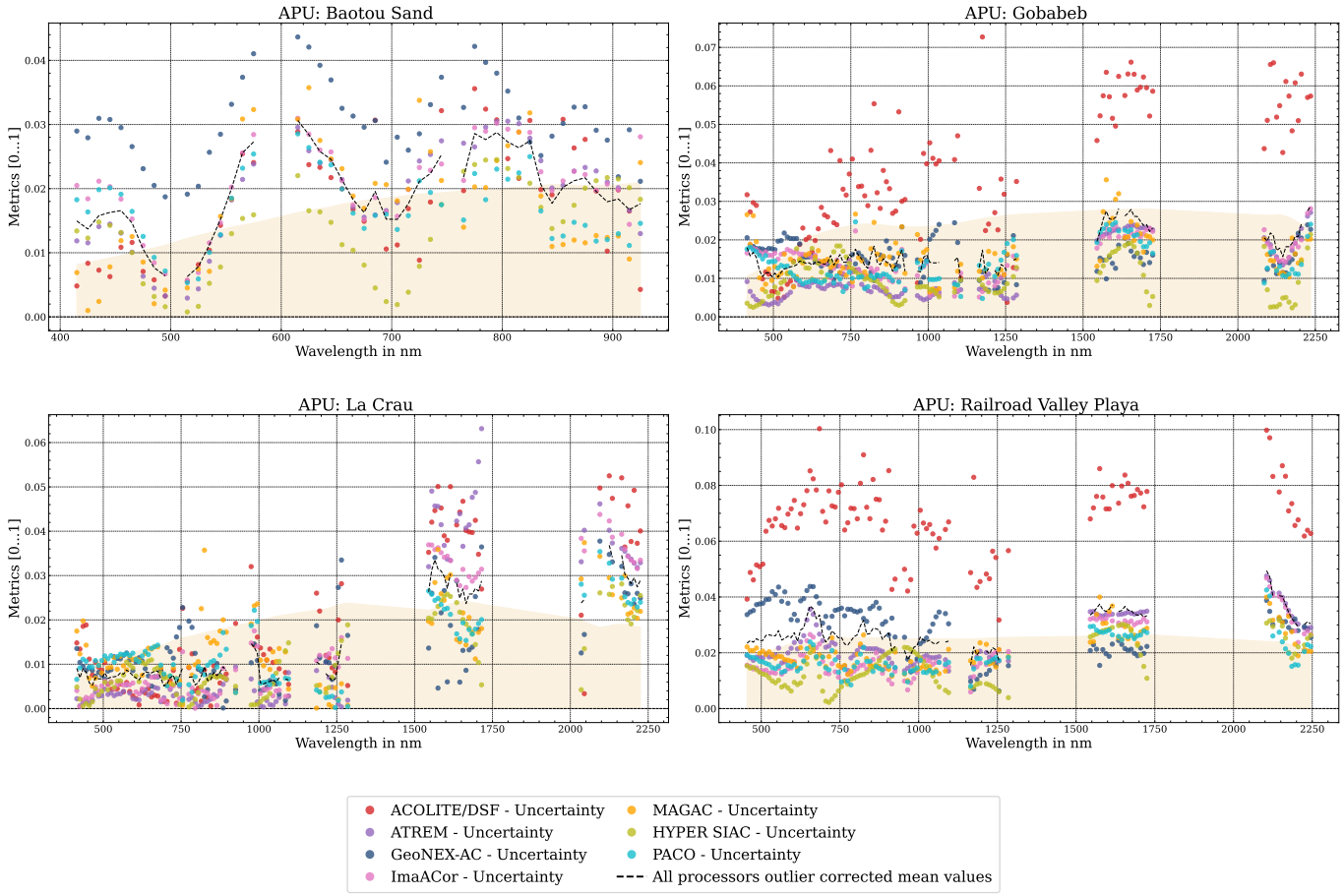

Figure 6. Uncertainty plots for EnMAP SR retrieved over RadCalNet sites. Common submitted wavelengths shown for each sensor, binned to 10 nm. The reference specification in the brown area refers to (4).

3.3.2. Surface Reflectance Validation Using HYPERNETS Measurements for EnMAP

There is one site within the HYPERNETS network with an EnMAP overpass, which is the Gobabeb site (see Table 6). Figure 7 shows that except for the very short wavelengths (<450 nm), the average uncertainty is within the reference specification of 1 to 3 %. Depending on the wavelength, ATREM, GeoNEX-AC, HYPER SIAC, and ImaACor show very good performances with uncertainty values partly under 1%. ACOLITE/DSF lies within the specification boundary except for some wavelength regions where the uncertainty reached 3–4%. A salient difference between the results over the Gobabeb site of the RadCalNet network (see Figure 6) and the HYPERNETS station is in fact the performance of the ACOLITE/DSF processor, which shows a much lower uncertainty over the HYPERNETS site.

When looking in more detail at the three EnMAP overpasses which cover both RadCalNet and HYPERNETS measurements (13 July 2022, 6 October 2022, 21 October 2022), ACOLITE/DSF showed better performance for HYPERNETS, especially beyond 1200 nm, with RadCalNet and HYPERNET uncertainties at a maximum of 8% and 4%, respectively. ATREM, GeoNEX-AC, ImaACor, HYPER SIAC, and PACO all demonstrated similar performance within requirements over these three scenes, with minor wavelength-dependent variations, often showing lower uncertainties for Hypernet data between 1500 and 1700 nm. MAGAC exhibited no clear or systematic performance pattern across wavelengths.

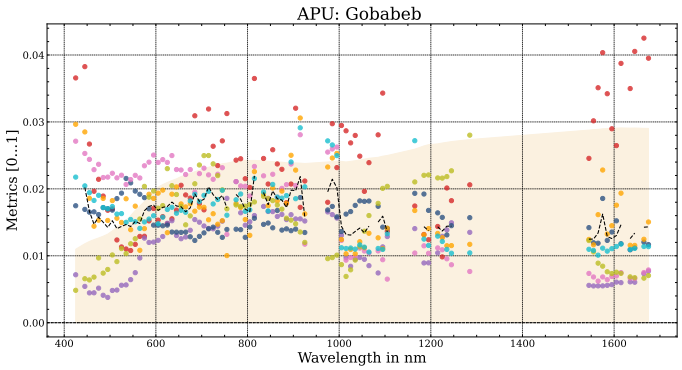

Figure 7. Uncertainty plots for EnMAP SR retrieved over the HYPERNET Gobabeb site. Common submitted wavelengths shown for each processor, binned to 10 nm. The reference specification in the brown area refers to (4). For the legend, please refer to Figure 6.

3.3.3. Surface Reflectance Validation Using EnMAP Validation Measurements

The results for the seven sites from the EnMAP validation campaign (see Table 8) are shown in Figure 8. For the Amiaz Plain desert site, performances are within reference requirements for roughly two parts of the spectrum, i.e., 500–750 nm and 2000–2300 nm. Outside of these areas, uncertainties go up as high as 8%, especially for GeoNEX-AC (between 500 nm and 1100 nm) and MAGAC (between 1100 and 1750 nm). Depending on the wavelength region, ACOLITE/DSF, HYPER SIAC, and GeoNEX-AC (>1100 nm) perform better than the average performance across all processors. The Antarctica site is the only polar site within the exercise. Overall, the performance from most processors (and therefore the average uncertainty) is within the reference specification. In particular, ImaACor, ACOLITE, and MAGAC show a consistently good performance across the spectrum. ATREM and PACO have some higher uncertainties between 450 nm and 550 nm, while HYPER SIAC shows some deviation around 1000 nm. GeoNEX-AC has higher uncertainties (up to 16%) throughout the spectrum and does not map well to the polar reflectance reference. For Berlin, all processors have uncertainties between 4% and 10% but are comparable across the spectrum, with about a 2% to 3% difference between processors. Generally, processors performing best in the shorter wavelengths (<950 nm), such as GeoNEX-AC and PACO, and have higher uncertainties in the longer wavelengths. The same can be said the other way around (e.g., ATREM, ACOLITE/DSF). None of the processors is within the reference specification. The retrievals over Makhtesh Ramon show smooth results within the specification below 3% uncertainty for most processors until 2000 nm. Retrievals between 2000 nm and 2300 nm show slightly higher uncertainties (2% to 5%) for most processors. MAGAC has some higher uncertainties in the very short bands, GeoNEX-AC's uncertainties are slightly higher between 500 nm and 1300 nm, and ACOLITE/DSF has higher uncertainties up to 14% throughout the spectrum. Over Pinnacles, most processors are within the reference specification throughout the spectrum. In particular, PACO and MAGAC, as well as ATREM, show a good performance with uncertainties mostly <2%. Some higher outliers (up to 6 %) are visible for HYPER SIAC (between 500 nm and 750 nm, as well as around 2050 nm) and GeoNEX-AC (between 750 nm and 900 nm), as well as sporadically for ACOLITE/DSF. The retrievals over Railroad Valley Playa are mostly within reference specification for retrievals < 1300 nm and mostly beyond the specification > 1500 nm. HYPER SIAC and PACO show a good performance, while ACOLITE/DSF shows slightly higher uncertainties up to 8%. GeoNEX-AC has uncertainties up to 4% until 1100 nm but performs well in comparison to other sensors for wavelengths > 1500 nm.

Figure 8. Uncertainty plots for EnMAP SR retrieved over EnMAP validation campaign sites. Common submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4). For the legend, please refer to Figure 6.

3.4. Surface Reflectance Validation for PRISMA

APU plots over all sites when processing PRISMA data with each processor are shown in Figure 9. Again, the average metrics (accuracy (blue), precision (green), and uncertainty (orange)) are shown for each wavelength (averaged over 10 nm) provided, depending on the processors. The average reference specification over all sites is shown in the brown area, between ±2 and 3% across the spectrum. The average retrieved surface reflectance uncertainty and precision are generally slightly above the reference specification (and within the specification for accuracy) for most processors (HYPER SIAC, ImaACor, MAGAC, and PACO) in the VNIR part of the spectrum. All processors show results outside of the reference specifications for wavelengths > 2250 nm. The processors that provide results at the end of the spectrum ∼2500 nm (ACOLITE/DSF, MAGAC, PACO, ImaACor) show extreme outliers there (outliers go up to 0.30 and are cut for better visualisation). The sharp divide in metrics observed around the 2250 nm line and the larger deviation from the specifications compared to EnMAP suggest an issue with PRISMA calibration

affecting most processors. Outliers around the water absorption features (970 nm, 1200 nm, 1450 nm, and 1950 nm), as described already for EnMAP, are also present for the PRISMA sensor but are generally less pronounced in magnitude.

For ACOLITE/DSF, all metrics are mostly above the reference specification. The overestimation for PRISMA is more pronounced than for EnMAP, with uncertainties roughly between 5% and 10% throughout the spectrum. ATREM shows a performance outside the reference specification, with uncertainty values between 2% and 6%, largely driven by an overestimation of reflectances on agricultural sites, which are not included in the EnMAP exercise. There are some outliers around 2000 nm and at the end of the spectrum. The retrievals until 1450 nm are about 2% lower in uncertainty than those >1450 nm. GeoNEX-AC's retrieval is slightly less smooth when compared to the other processors. The metrics are slightly higher, though comparable to those sites obtained with the EnMAP sensor. Throughout the spectrum, the uncertainty ranges between 2% and 4.5%. There are no big performance differences throughout the spectrum, with a general overestimation of values. An underestimation of values occurs between 2300 nm and 2400 nm. HYPER SIAC retrievals with PRISMA are very comparable to those with EnMAP, with very smooth retrievals and uncertainty and precision at the upper limit of the reference specification until 1750 nm. Again, wavelengths at the known water absorption bands are not retrieved by the processor. The areas of strong underestimation (2000–2100 nm and 2300–2500 nm) identified over EnMAP sites are equally present in the PRISMA retrievals. The retrievals of the ImaACor processor show good performance, with uncertainties slightly above the specification boundary throughout the spectrum. There are small outliers at water absorption bands and some higher uncertainty values around 2000 nm, but overall, the performance is very consistent throughout the spectrum. The underestimation of values > 2250 nm observed for the EnMAP sensor is not present. The APU graph for MAGAC is very similar to that of ImaACor, and like the ImaACor processor, the underestimation of SWIR values at the end of the spectrum is not as pronounced as over the EnMAP sites. The retrieval is quite smooth and consistent across the spectrum. Also, PACO is comparable with ImaACor and MAGAC, with similar APU metrics across the spectrum. The mentioned outliers, especially around 2000 nm are lower, and the retrieval is smooth.

A summary of uncertainty results can be seen in Figure 10. Again, higher uncertainties for ACOLITE/DSF throughout the spectrum can be seen, as well as slightly higher values for ATREM (NIR/SWIR) as well as MAGAC, PACO, and HYPER SIAC between 1200 and 1400 nm. Generally, there are higher uncertainties for all processors in the SWIR. The comparison of uncertainty estimates between EnMAP and PRISMA across all processors shows that EnMAP consistently exhibits lower uncertainty values, ranging from 0.017 to 0.06, whereas PRISMA values are generally higher, reaching up to 0.09. For the ACOLITE/DSF processor, EnMAP exhibits uncertainty values up to 0.06, while PRISMA values go up to 0.09. Similar patterns are observed for ATREM at a lower scale, with EnMAP uncertainties mostly between 0.02 and 0.03, while PRISMA values reach up to almost 0.05. The GeoNEX-AC and ImaACor processors also show lower uncertainty in EnMAP (around 0.02–0.035) but only slightly higher uncertainties for PRISMA (up to 0.045). For MAGAC, HYPER SIAC, and PACO, EnMAP maintains uncertainty values generally below 0.03, whereas PRISMA uncertainties are one and a half times higher, approaching 0.045.

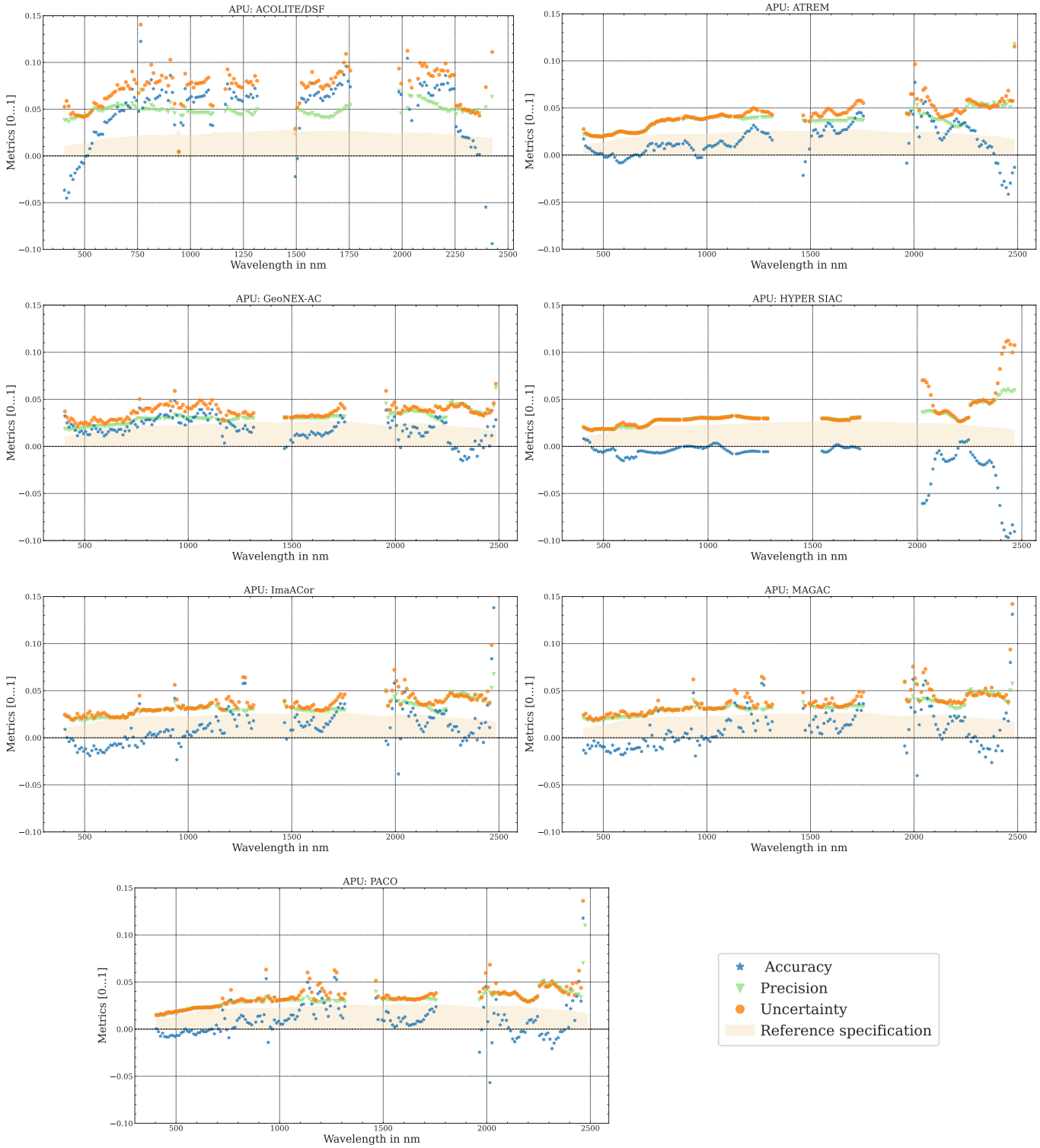

Figure 9. Accuracy, uncertainty, and precision (APU) plots for PRISMA SR retrieved over RadCalNet and HYPERNETS stations, as well as for the HYPERSENSE validation campaign. All submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4). Note that outliers at the end of the spectrum > 0.15 are cut for visualisation.

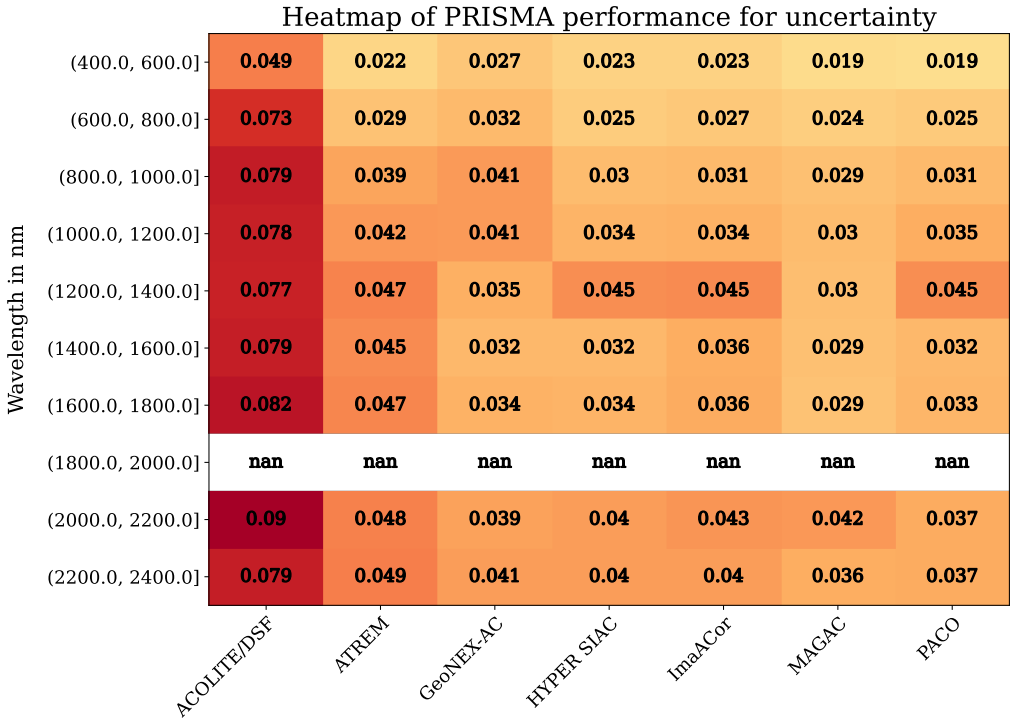

Figure 10. Heatmap of average uncertainties over all PRISMA sites for all processors. Only bands common to all processors (excluding water absorption regions) are included in the computation of the uncertainty. A darker colour corresponds to a higher uncertainty value.

3.4.1. Surface Reflectance Validation Using RadCalNet Measurements for PRISMA

Three stations from RadCalNet were included for PRISMA scenes in the exercise (see Table 5); there was no overpass available for the Baotou Sand site. Again, the mean uncertainty of all processors (screened for outliers) is indicated with the dashed black line, while the reference specification is shown in the light brown area (see Figure 11).

For Gobabeb, all processors except ACOLITE/DSF and GeoNEX-AC perform well within the specification boundary (around 2% uncertainty). GeoNEX-AC is slightly above that (with up to 4% uncertainty) for wavelengths < 1300 nm. ACOLITE/DSF shows higher uncertainties (up to 12%) throughout the spectrum. Compared to the EnMAP retrievals over RadCalNet/Gobabeb, the uncertainties for ACOLITE/DSF are about 5 pp higher. HYPER SIAC performs very well throughout the spectrum, ATREM until 1100 nm. For La Crau, the average uncertainty is slightly beyond the reference specification across the spectrum, within 1% and 5%. Outliers are prevalent especially around 450 nm, 950 nm, and 1550 nm. ACOLITE/DSF, PACO, MAGAC, and GeoNEX-AC perform well, generally below the average, while HYPER SIAC has uncertainties slightly higher than the rest of the processors.

The results over Railroad Valley Playa are comparable to those of Gobabeb, with general performance around 2% within the reference specification until 2000 nm. PACO stays within the specification for the whole spectrum. As with Gobabeb, GeoNEX-AC is slightly beyond that with 4% uncertainty until 1100 nm and higher uncertainties (up to 14%) for ACOLITE/DSF. Again, the ACOLITE/DSF uncertainties are about 4 pp higher when compared to the retrievals over Railroad Valley Playa for EnMAP.

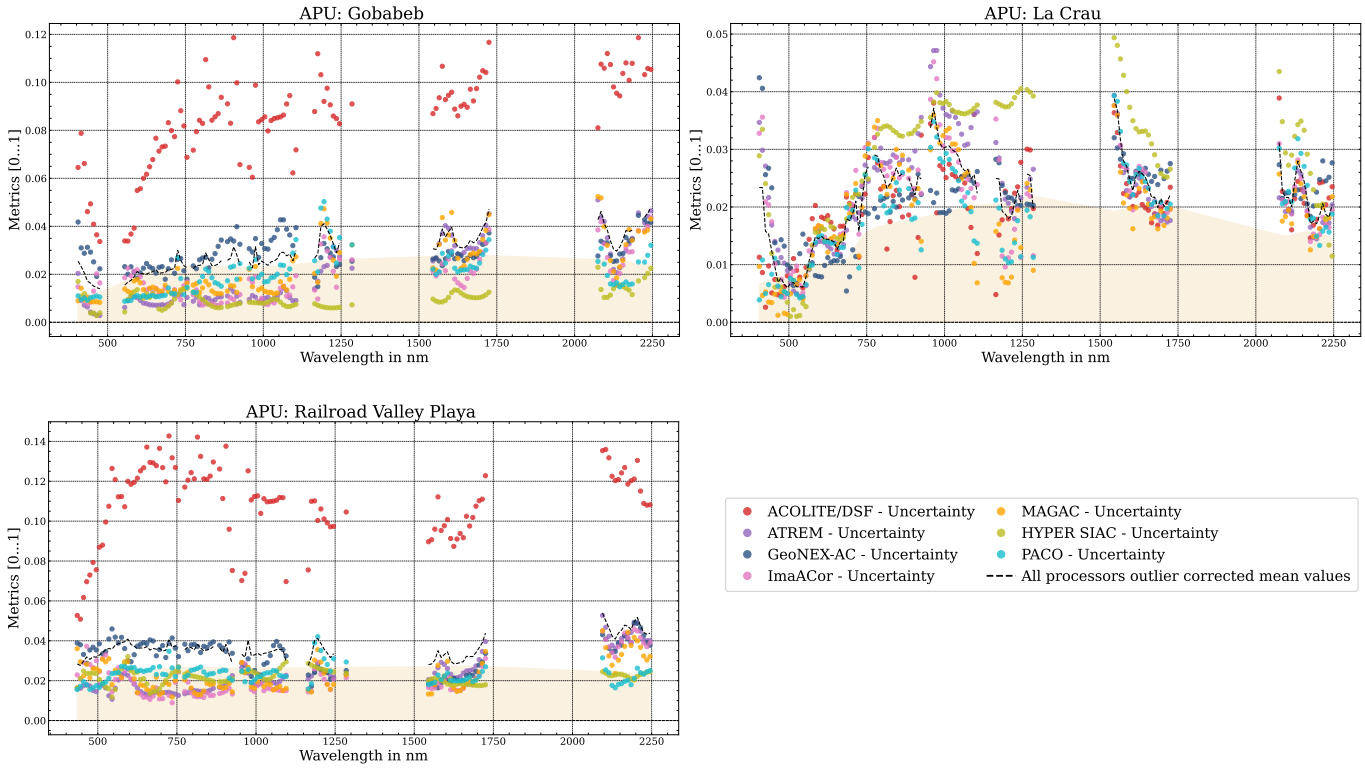

Figure 11. Uncertainty plots for PRISMA SR retrieved over RadCalNet sites. Common submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4).

3.4.2. Surface Reflectance Validation Using HYPERNETS Measurements for PRISMA

For HYPERNETS, there are two stations with PRISMA overpasses, Gobabeb and Wytham Woods (see Table 6), shown in Figure 12. The results over Gobabeb are very similar to the PRISMA RadCalNet results over Gobabeb, with a good performance from all sensors within the reference specification and a larger uncertainty for ACOLITE/DSF across the spectrum. At 8%, this uncertainty is lower than over the RadCalNet site. Additionally, the GeoNEX-AC results are better than over the RadCalNet site (see Figure 11) and are within the specification boundary for the whole spectrum, except the very short wavelengths. When compared to the EnMAP results over the HYPERNETS Gobabeb site (see Figure 6), most processors perform similarly, but the uncertainty associated with ACOLITE/DSF is again much more pronounced when processing PRISMA data. For the Wytham Woods site, the average uncertainty is 1% to 3% above the reference specification. There is a good performance from different processors over different parts of the spectrum: ACOLITE/DSF (especially between 550 nm and 750 nm), ATREM (except for the short wavelengths < 900 nm), and ImaACor and HYPER SIAC mainly throughout the whole spectrum. MAGAC and PACO have higher uncertainties (up to 7%) for wavelengths > 800 nm. Uncertainties for GeoNEX-AC go up to 5% between 800 and 1100 nm.

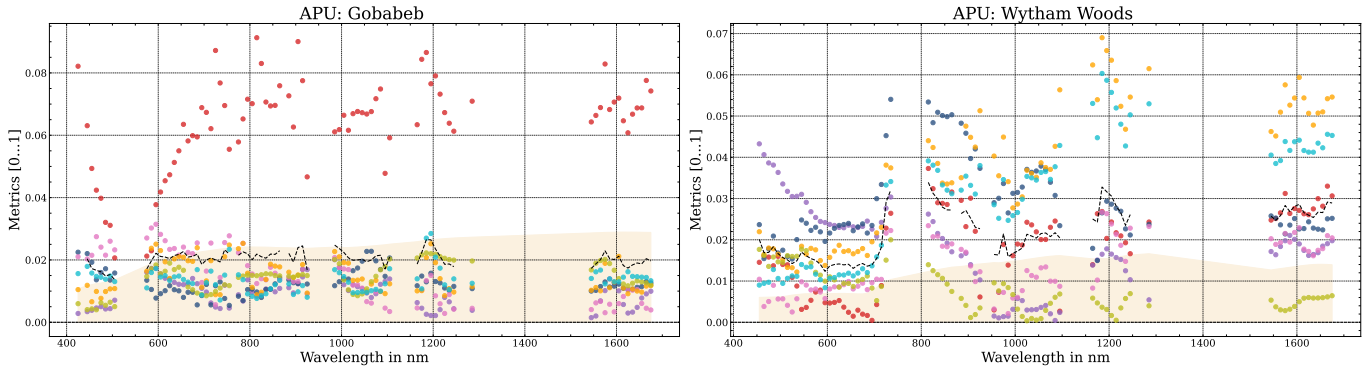

Figure 12. Uncertainty plots for PRISMA SR retrieved over HYPERNETS sites. Common submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4). For the legend, please refer to Figure 11.

3.4.3. Surface Reflectance Validation Using Hypersense Campaign

Figure 13 shows the results over the HYPERSENSE campaign sites (see Table 7). For the sites of Braccagni and Camarena, the different sub-sites are shown separately, as they consist of very different land cover types (e.g., dry vegetation vs. wet soil). For Rio Tinto, results are averaged over both sites, as the surface type is the same.

The three Braccagni sites (bare dry soil, bare wet soil, and summer wheat) show a similar performance of processors when compared to each other. Over those three sites, ATREM shows the highest uncertainties across the spectrum (up to 12% for the soil sites and 7% for the dry vegetation), while MAGAC, PACO, and HYPER SIAC have the lowest uncertainty. Those processors are within the reference specification for the vegetation sites across the spectrum, while they are mostly beyond it (save for the extreme ends of the spectrum) for the soil sites. For the bare dry soil site, the average uncertainty lies between 2% and 6%, while it goes up to 8% for the bare wet soil site. For the latter, a strong increase in uncertainty for all processors around the red edge mark at 750 nm can be observed, suggesting the presence of vegetation in the pixel. Performance for all processors is best on the summer wheat site, with average uncertainties only about 1% to 2% beyond the reference specification. For the Camarena soil sites (BareSoil1 and Bare Soil2), very different results can be observed. For BareSoil1, ACOLITE/DSF shows a very good performance, and it is the only processor whose results lie within the reference specification (for wavelengths > 750 nm). PACO and MAGAC show the highest uncertainties between 10% and 12%, while the other processors lie in between those extremes, with uncertainties between 4% and 9%. For the site BareSoil2, the performance of processors is inverted, with MAGAC and PACO, as well as HYPER SIAC, having the lowest uncertainty, and ACOLITE/DSF at the higher end with up to 12% uncertainty. The average lies between 2% and 8%. For the dry vegetation site, again ACOLITE/DSF shows higher uncertainties up to 12%. Generally, most processors are within the reference specification for wavelengths < 1100 nm. Then the dispersal is higher for longer wavelengths.

All processors show similar patterns, with dips in uncertainty around 1650 nm and 2000 nm. PACO and HYPER SIAC perform especially well. For the Rio Tinto site, the processor performance is wavelength-dependent. ACOLITE/DSF performs well and within the reference specification for wavelengths > 500 nm. GeoNEX-AC lies at the higher end of the specification, except for higher uncertainties between 750 nm and 1100 nm up to 6%. PACO performs well but has higher uncertainties up to 8% for wavelengths > 2000 nm. ImaA-Cor and ATREM are also well within the specification boundary and only slightly above for wavelengths > 2000 nm. HYPER SIAC and MAGAC perform well for shorter wave-

lengths but have higher than average uncertainties for wavelengths > 1300 nm. The overall performance is thus better for the VNIR spectral region.

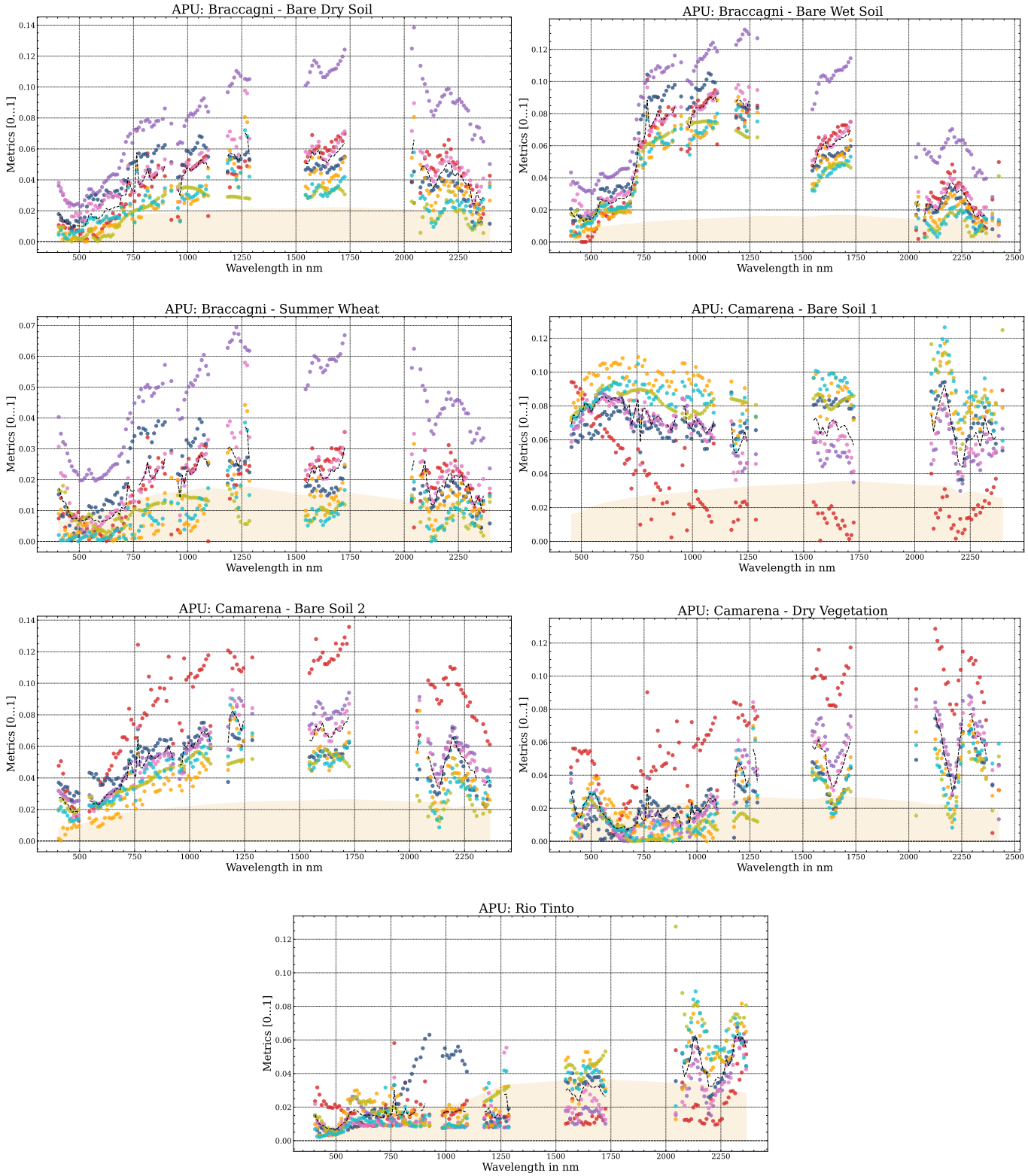

Figure 13. Uncertainty plots for PRISMA SR retrieved over Hypersense campaign sites. Common submitted wavelengths shown for each sensor binned to 10 nm. The reference specification in the brown area refers to (4). For the legend, please refer to Figure 11.

4. Discussion

4.1. Validity In Situ Data

Validating atmospheric correction across diverse land covers and atmospheric conditions requires careful consideration of site heterogeneity and sampling strategies. Several sites used in the exercise (e.g., the HYPERSENSE campaign, Wytham Woods, and several EnMAP Validation sites) can be considered heterogeneous. Taking the site at the Tegel airport in Berlin as an example, none of the processors is within the reference specification. This strongly suggests an issue with the representativeness of the in situ data at this site, possibly related to the chosen pixel site or adjacency effects. Valid statistical samples cannot be obtained by measuring only one pixel, but increasing pixels reduces in-site representativeness. Results for all processors over the sites at Braccagni and Camarena (see Figure 13) show higher uncertainty, especially around the vegetation red edge (700 nm). This suggests the presence of mixed pixels or the impact of vegetation due to adjacency effects that are not fully corrected. While these sites might be less suitable for validation than homogeneous desert sites, showing the results here is important for two reasons. First, the exercise demonstrates validation results across various land cover types and atmospheric conditions, comparing processor performance in different scenarios. Second, operational processors need to perform atmospheric correction also over challenging sites. Seeing how they perform over heterogeneous areas can benefit developers and users working on agricultural land, urban areas, or forests. In response, ongoing efforts aim to extend spatially representative (vegetated) sites for hyperspectral validation, covering homogeneous areas spanning multiple pixels. The LANDHYPERNET network is expanding and has added several sites (a mixed forest in Estonia and two agricultural sites, one in Belgium and one in Italy) after the completion of this exercise. Furthermore, ESA's Cal/Val Park project plans on building a comprehensive validation site, featuring five large radiometric targets made of different materials (artificial, such as asphalt and gravel, as well as natural targets such as bare soil and natural vegetation) spanning a wide range of brightness levels and spectral profiles [71]. Sites like these might significantly improve the range of surfaces that can be reliably validated within exercises such as ACIX.

Proximity of sites from different networks, such as Gobabeb within RadCalNet and HYPERNETS, provides unique research opportunities. As described in the Results section, processor performance can vary considerably between these sites, though this observation is based on average values across all scenes. Especially for PRISMA, more sites are available for RadCalNet, which span the different seasons and potentially show a more balanced average of atmospheric conditions. For HYPERNETS, two of the three selected days have high AOD > 0.2. But also, same-day acquisitions from the two networks show differences. This might be related to the different processing steps as explained in Sections 2.2.4 and 2.2.5: the Gobabeb RadCalNet site was post-processed to adjust the BRDF for each scene, while for HYPERNETS, the closest matching viewing angle mirroring the satellite sensor viewing angle was chosen. While not the focus of this more general review paper, further work on comparing the two sites and the impact of measurement protocols could improve measurement best practices in general.

4.2. Bias of In Situ Data

The composition of reference datasets influences the reliability of atmospheric correction assessments, highlighting biases due to limited land cover and atmospheric variability. Although the exercise aims for an unbiased assessment of atmospheric correction processors, the imbalance in reference data must be considered when interpreting results (Section 3). For hyperspectral validation, match-ups were available from two networks (i.e., RadCalNet and HYPERNETS). The same is true for the ad hoc campaigns used for

validation. Most match-ups originate from RadCalNet sites, primarily homogeneous bright deserts. The lack of dark pixels within the scene hampers the performance of several algorithms for their aerosol retrieval. This is especially evident in results from the ACOLITE/DSF processor. These dark-pixel-dependent processors typically overestimate the AOD of the scene and, subsequently, also the retrieved surface reflectance. Another challenge is the absence of BRDF correction to adjust references for viewing geometry. This BRDF correction is only available for two out of four RadCalNet sites. At the time of the exercise, access to HYPERNET sites was limited. Again, most scenes come from the homogeneous bright desert site Gobabeb. While HYPERNET is an important step towards a hyperspectral reference dataset measured from a variety of angles, the SWIR above 1700 nm is not currently measured. Furthermore, match-ups over vegetation sites such as Wytham Woods were rare due to often overcast conditions over the site and do not span the full vegetation cycle. Other potential vegetation sites were either not producing data at the time of data collection or were not well-suited for satellite validation. As mentioned, ad hoc campaigns diversify land cover types but face issues such as missing measurement uncertainties, BRDF characterisation, and location inaccuracies. Like the limited land cover diversity, scene AOD and WV distributions skew towards low values (see Figure 14). Most scenes have an AOD < 0.2 and a WV < 2.0 g/cm^2.

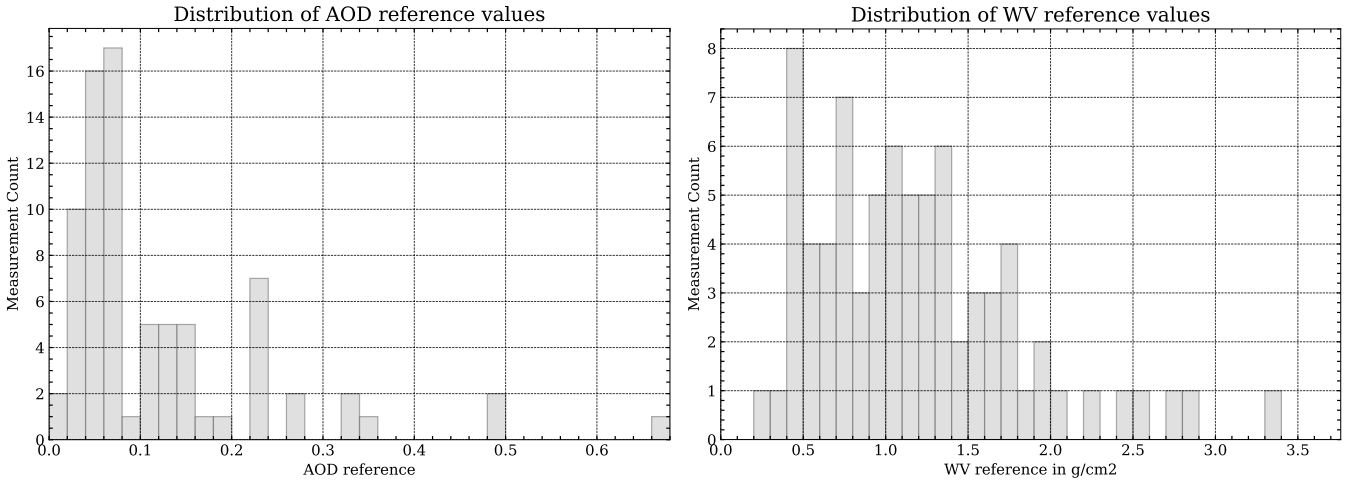

Figure 14. Histograms of reference scenes and their associated AOD (**left**) and WV (**right**) values.

When looking at the overall results (see Figure 4 for EnMAP and Figure 9 for PRISMA), these are dominated by bright homogeneous desert surfaces with low AOD/WV. ACOLITE/DSF shows the largest overestimation of SR retrievals among the processors, which is caused by the large subset of bright desert sites in the exercise and obscures the processor's better performance over different land cover types (see Antarctica site for PRISMA in Figure 8 or Rio Tinto and Camarena–BareSoil1 for PRISMA in Figure 13). More generally, the performance metrics may not directly translate to more heterogeneous or challenging environments. In several cases, a priori information, such as atmospheric state variables or BRDF characterisation, contributed to the atmospheric correction process and may have improved the realism of the retrieved surface reflectance. Under more complex atmospheric conditions (e.g., elevated aerosol load, high humidity, or rugged terrain), larger deviations can be expected, particularly for processors relying on empirical or climatological assumptions (see Section 2.1 and Table 1 for an overview on these assumption and ancillary data used). To mitigate bias from land cover, seasonality, aerosol loading, and WV, we recommend incorporating synthetic scenes, as in prior ACIX implementations. In this way, a performance analysis over more challenging areas (isolated dark objects, inland water bodies, tilted terrain) can furthermore illustrate the differences between processors. Inter-

comparison exercises in that domain, e.g., the Radiation Transfer Model Intercomparison for Atmosphere (RAMI4ATM) [72], can be a basis on which to choose an independent model for those simulations.

4.3. Performance Results for Each Processor

For ACOLITE/DSF, the results generally indicate a poor SR estimation, although good accuracy is achieved over specific scenes (e.g., Antarctica, Camarena Bare Soil 1, Rio Tinto, Wytham Woods). This suggests that while the processor and associated LUTs produced with 6SV can retrieve high-quality SR estimates, there are challenges with the direct estimation of AOD from the darkest targets and over strong WV absorption bands (see Section 4.6). For the bright scenes, even for the darkest pixels in the scene, residual land reflectance—either direct or due to adjacency effects—elevates the AOD estimates. Bright desert scenes, such as Gobabeb, represent a typical and significant proportion of sites in this exercise. Therefore, this pattern illustrates a typical source of error in intermediate products: direct estimation of AOD from the darkest targets is less reliable in bright environments, and strong water vapour (WV) absorption bands further complicate the retrieval. Elevated AOD estimates in turn propagate into overestimation of SR. Errors in WV retrieval can similarly affect SR, particularly in spectral regions with strong absorption, creating systematic deviations. While this does not affect all processors, it is an important error source for processors relying on DDVs for AOD retrieval (e.g., ACOLITE/DSF, PACO, ImaACor). DSF was originally developed for processing coastal and inland water targets, where a sufficient number of dark targets is typically available to enable a reasonable direct AOD estimate. These dark targets are generally found over water, i.e., pixels where water is highly absorbing and no signal from below the surface is detectable, either in the VIS (due to high CDOM absorption) or in the NIR/SWIR (due to high pure-water absorption). Alternative targets include dense vegetation, shadows, snow, or ice, where at least one spectral band gives a good estimate of the AOD. When such targets are absent, directly fitting the path reflectance to the darkest pixels results in a significant overestimation of AOD from non-zero target reflectance. The scatterplots in Figure 15 illustrate this relationship between absolute errors in AOD and SR retrieval across different processors. The left panel shows ACOLITE/DSF, with a relatively strong correlation (R^2 = 0.625) (see Table 11) between AOD and SR errors, indicating that inaccuracies in AOD retrieval significantly affect SR retrieval performance. In contrast, the right panel shows results for other processors—GeoNEX-AC, HYPER SIAC, ImaACor, and MAGAC (not enough data were available for ATREM and PACO for this comparison)—which exhibit considerably lower R^2 values (ranging from 0.02 to 0.21), suggesting a weaker or negligible dependence of SR error on AOD error. The DSF is unsuitable for many of the bright reflectance targets in this study and cannot be recommended for general application. Using an alternative method for extracting AOD over bright surfaces could substantially enhance SR retrieval fidelity for these scenes. Using a fallback or ancillary AOD value brings the results in line with other processors and confirms that the AOD estimation is the main issue with ACOLITE/DSF processing.

Based on the results of this study and the observed substantial overestimation of AOD in bright scenes, ACOLITE was extended to incorporate ancillary aerosol data from the GMAO_MERRA2 reanalysis product [18] instead of relying on the DSF estimation. Using ancillary aerosol information significantly improves retrieval uncertainty and greatly reduces the SR overestimation in bright scenes. Consequently, ACOLITE outputs processed with ancillary aerosol information demonstrate a much closer agreement with other processors, reducing the overall SR uncertainty from approximately 5% to 3% for EnMAP and from 8% to 3% for PRISMA.

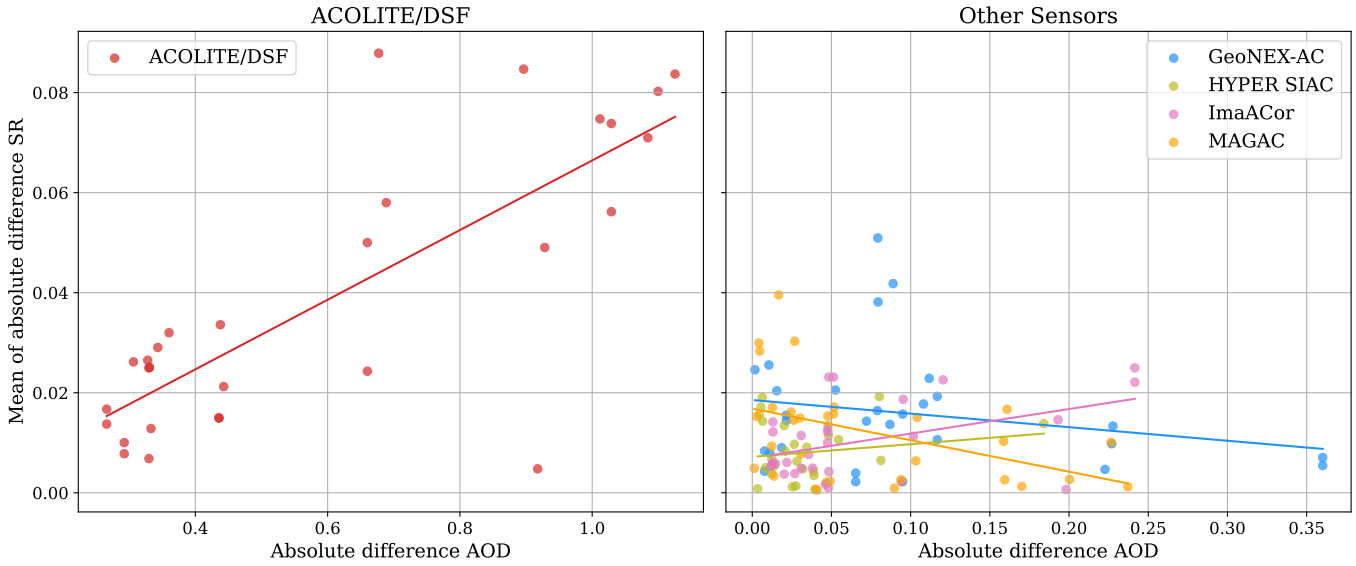

Figure 15. Scatterplots comparing absolute error in AOD to absolute error in SR retrieval. ACOLITE/DSF on the (**left**), other processors on the (**right**).

Table 11. R^2 of error in AOD to error in SR retrieval.

	ACOLITE	GeoNEX-AC	HYPER SIAC	ImaACor	MAGAC
R^2	0.625	0.049	0.024	0.201	0.205

ATREM demonstrates robust and reliable performance, particularly for the EnMAP sensor, which exhibits low uncertainties across the entire spectral range. In contrast, PRISMA results show slightly higher uncertainties, especially over agricultural scenes and primarily beyond the 1500 nm wavelength. PRISMA data processing was performed without the use of pre-stored spectral libraries; after standard atmospheric correction, spectral smoothing was applied, utilising only the 1140 nm WV absorption band for vapour retrieval. Conversely, EnMAP processing also employed spectral smoothing but leveraged both the 825 nm and 1140 nm WV bands, which may contribute to the lower uncertainties observed in EnMAP data. Notably, challenges persist over strong WV absorption bands, which have been attributed to limitations of the 6S radiative transfer model implemented within ATREM (see Section 4.6).

HYPER SIAC performs best in the estimation of the two main atmospheric state parameters (AOD and WV). While some of the methods used overlap with those used by other processors in this study (e.g., the APDA method for WV retrieval), including the simulated Sentinel-2 TOA reflectance in a preliminary step might lead to more accurate results. The algorithm also provides low uncertainty in the surface reflectance over a large part of spectral range compared to other processors evaluated in this study. Similarly to ATREM, its performance drops over strong WV absorption regions due to the potential 6S model error (see Section 4.6). This issue was resolved by replacing the 6S model with libRadtran; however, the updated version is not included in the present comparison.

GeoNEX-AC maintains a stable mid-range uncertainty across both sensors, reflecting consistent but not exceptional performance compared to other processors. GeoNEX-AC was developed to process high temporal (i.e., 10 min) resolution geostationary satellite observations. Its main algorithm requires multi-angle information from the varying sun-target-satellite geometry during the day to help retrieve the AOD of the atmosphere column. However, such information is not available from the polar orbiting hyperspectral observations. Therefore, the GeoNEX-AC algorithm falls back to the traditional approach

that depends on the use of spectral band ratios, estimated from corresponding MODIS products, to perform atmospheric correction. The performance of the algorithm is thus likely affected by spectral and spatio-temporal discrepancies between the hyperspectral sensors and the broadband MODIS observations. An alternative inversion approach is under development, which aims to combine GeoNEX-AC radiative transfer look-up-tables with the ISOFIT code, leveraging its optimal estimation specifically formalised in performing hyperspectral atmospheric correction.

ImaACor demonstrates strong and stable performance with low uncertainty values across both sensors, indicating that it is among the most reliable processors evaluated. Some issues with regard to WV and AOD retrieval can, however, be noted. Regarding the WV absorption bands, ImaACor does not proceed on a pixel-by-pixel basis. It rather applies the mean value of the atmospheric WV map across the entire image, since additional processing would provide only limited benefit in the other bands. While this approach is efficient, it could lead to less accurate results, particularly in scenes containing vegetation, where water stored in leaves can bias the mean estimate. Using the reference location (RadCalNet or HYPERNETS) or excluding vegetation through a mask based on the Normalised Difference Vegetation Index (NDVI) before computing the WV mean value could contribute to reducing uncertainty. As for AOD retrieval, ImaACor faces issues especially in arid areas, as it relies on the identification of a sufficient number of DDVs in the scene for its atmospheric characterisation. A further limitation arises from the reliance on standard MODTRAN models. A more specific atmospheric profile characterisation could be particularly beneficial at shorter wavelengths. Beyond the need to improve consistency between images and reference measurements, further improvements in ImaACor could be achieved by reducing some of the approximations—currently introduced to decrease processing time—in the computation of the atmospheric scattering function.

MAGAC's overall performance in ACIX-III is solid and comparable to other processors, particularly for AOD retrievals, where results are generally robust but show some degradation in the moderate AOD range (0.25–0.4). This happens especially over Gobabeb test sites, likely due to the combined effect of poorly constrained retrievals driven by climatological assumptions and suboptimal first-guess values from meteorological inputs. WV retrievals demonstrate a good correlation (R^2), but the presence of a systematic positive bias ($\sim$0.29 g/cm^2) is indicative of errors introduced by the current refinement step within the aerosol retrieval process, which uses WV as a support variable in the state vector. This mechanism should be revisited to assess whether the baseline differential absorption approach is more reliable. Surface reflectance retrievals for EnMAP and PRISMA are broadly consistent, yet distinct outliers appear in strong gaseous absorption bands (760, 840, 2050 nm), largely attributable to spectral convolution effects (e.g., ISRF modelling, smile) and inaccuracies in surface pressure due to the use of unsmoothed DEM data. The latter can also be considered a contributing factor in the observed increased errors in the blue spectral region, where Rayleigh scattering is sensitive to errors in surface pressure. Taken together, these issues highlight how specific algorithmic choices in MAGAC—the reliance on climatology and meteorological priors for aerosols, the WV refinement coupling, and the handling of topography and spectral response—directly influence its performance, with site- and sensor-specific effects visible in the validation results.

PACO ranks as a reliable processor with low uncertainty values on both sensors, suggesting a solid and dependable performance. Nevertheless, errors in PACO's SR retrieval performance can be traced back to various sources. As for atmosphere characterisation, PACO bases its AOD characterisation on spectral features typical of DDV in absence of additional dependencies. Therefore, the SR is systematically affected below 750 nm for arid sites. In ACIX-III, only single tile scenes were provided. However, in operational use,

PACO can obtain information such as DDV from neighbouring tiles to enhance the overall retrieval. Moreover, a constant CO_2 concentration of 420 ppm is used in the radiative transfer simulation. Deviation from this average might cause differences in the CO_2 absorption region between 1950 and 2050 nm. After the characterisation of the atmosphere, PACO performs a strictly band-wise AC without any mixing of neighbouring bands or modelling of the spectra. This preserves the full spectral characteristic of the input radiance data but makes the SR more susceptible to band-to-band variations and absorptions features. Multiple artefacts, such as the boundaries of the WV absorption regions and the absorption region at 1150 nm can be partially attributed to this design. These areas are generally challenging to correct due to the low signal level and susceptibility to inaccuracies in the WV estimation. The oxygen A-band near 760 nm is also affected. This assumption of a constant CO_2 concentration and application of a strictly band-wise AC are also part of MAGAC, and the effects described are likely also found in the results of that processor. As for EnMAP, the separate VNIR and SWIR detectors of the instrument result in minor radiance variation in the region between 900 and 1000nm which are transferred to the SR. PACO processes all bands in the input radiance data, even the ones close to challenging absorption regions. This must be taken into account when comparing the performance to other processors.

4.4. Differences in Results over PRISMA and EnMAP

During the exercise, several key differences and challenges between the EnMAP and PRISMA sensors were identified that affect data quality and retrieval performance. Especially with regard to the distributed PRISMA L1G product, several issues were raised. Notably, the DEM included in the dataset exhibits striping patterns. These artefacts are likely caused by nearest neighbour resampling of the Copernicus DEM used to align it with the sensor grid. If not properly filtered, such spatial artefacts could negatively impact images during terrain correction. To mitigate this, participants were advised to apply smoothing filters to the DEM prior to performing terrain correction.

Although PRISMA geolocation was largely corrected, some scenes still exhibited a 2–3 pixel offset, further resulting in misalignment with the DEM. For these scenes, the spatial accuracy was insufficient for reliable terrain correction. This geolocation offset complicates the validation of PRISMA data and likely contributes to the higher uncertainties observed across all sites, as a precise overlap of validation pixels and in situ measurement cannot be guaranteed. Additionally, PRISMA data exhibit spectral smile, particularly in the first 10 spectral bands. This spectral distortion can affect haze estimation for certain processing algorithms that rely on the initial blue bands to identify haze. Specifically, the uncertainty in AOD retrieval nearly triples from EnMAP (0.35) to PRISMA (0.92). Other processing algorithms do not demonstrate such a pronounced difference between EnMAP and PRISMA sensors. For example, MAGAC, HYPER SIAC, and ImaACor show negligible differences, while GeoNEX-AC achieves even better AOD retrieval accuracy for PRISMA.

The discrepancy between EnMAP and PRISMA in AOD—and consequently SR—retrievals for ACOLITE/DSF is caused by a differences in the default configuration, which use different spectral ranges for AOD retrieval. For PRISMA, a shorter spectral range in the VIS to NIR is employed (500–970 nm), whereas EnMAP uses the full range from VIS to SWIR (400–2500 nm). The characteristics of the targets within the scene influence the direct estimation of AOD by ACOLITE/DSF, generally resulting in higher AOD retrievals within the PRISMA wavelength range. The PRISMA wavelength range was used to improve retrieval performance over coastal and inland water sites [73], but it clearly gives worse results over land targets when compared to the default EnMAP configuration.

4.5. Imaging Spectroscopy Analysis

Compared to previous ACIX iterations, which dealt with multispectral data, ACIX-III looks at a greatly increased number of data points per scene. However, in the direct comparison between all sensors, some level of detail is lost, either due to the omission of bands that are not part of all processors' outputs or because of a certain need to average results to make comparisons possible. Outliers affected by atmospheric absorption bands can have a strong impact on the overall performance metrics of single processors and/or scenes and do not necessarily reflect the overall retrieval capability of a processor. Those parts of the spectrum should therefore be assessed independently. To more easily evaluate a processor's performance, the requirements used in ACIX-II were adapted for ACIX-III. These specification requirements can be smoothly applied to AOD and WV, but pose more issues for the SR retrievals, as they were developed with multispectral sensors, especially Sentinel-2, in mind.

4.6. 6S Modelling for Hyperspectral Data

ACOLITE/DSF, ATREM, and HYPER SIAC make use of 6S as their RTM. The participants have observed that the 6S code, while widely adopted for multispectral remote sensing applications, exhibits significant limitations when applied to hyperspectral data over strong water vapour absorption region. At the core of the issue they identify its use of the Malkmus narrow-band model for simulating gas transmittance, which does not reliably capture the transition from weak to strong absorption features. This results in inaccurate spectral shapes, particularly for WV absorption bands centred around 940 nm and 1140 nm, when compared with observed EnMAP spectra under both dry and moist atmospheric conditions. These discrepancies, though negligible in multispectral contexts, can substantially degrade hyperspectral retrievals. For example, earlier versions of the 5S and 6S codes were observed to omit or mishandle certain gas species—such as CH_4 near 2300 nm—leading to misinterpretation of absorption features as surface mineral signatures. Additional limitations are suspected to exist in the approximation of spherical albedo, particularly in deep absorption bands where interpolation is based on unaffected spectral regions, potentially introducing further errors. Given the widespread use of 6S in the community, it is essential to raise awareness of these issues and to encourage more accurate modelling frameworks tailored to the demands of hyperspectral atmospheric correction.

5. Conclusions

As with the previous two ACIX iterations, ACIX-III Land was jointly organised by ESA and NASA within the framework of CEOS WGCV. For the first time, ACIX Land included data from PRISMA and EnMAP hyperspectral missions. Seven processors participated in a comparison of AOD, WV, and SR retrievals over 90 satellite scenes. AOD and WV were assessed over a variety of AERONET and RadCalNet sites located in North America, Europe, Southern Africa, Asia, and Australia. Half of the processors had average uncertainties of <0.1 for the AOD retrieval, but maximally 35% of retrievals were within the specification requirements. WV retrievals showed good results with some consistent offsets, showing uncertainties between 0.17 and 0.88 g/cm^2. Average uncertainties for SR retrievals depended on wavelength, processor, and sensor (uncertainties are slightly higher for PRISMA and at longer wavelengths), with average uncertainties between 0.02 and 0.04. While most processors showed a good performance over the test sites, some processors had issues with the correct AOD retrieval due to a reliance on dark pixels, which were not present in all scenes. Although ACIX-III focusses on a limited range of land cover types and aerosol loading for its match-ups, it serves as a rigorous and comprehensive benchmark for validating and advancing AC algorithms applied to hyperspectral data. Looking

ahead, a potential ACIX-IV might adopt an even broader and more ambitious scope by incorporating a wider diversity of in situ scenes, particularly as hyperspectral networks such as HYPERNETS expand and become operational. Additionally, the integration of synthetic scenes generated via an independent RTM could provide a valuable complement to these observations, offering a wider scope of land cover, seasonality, aerosol loading, and WV. This comprehensive approach will be essential to drive the next generation of algorithm development and ensure robust performance across varied conditions.

Author Contributions: N.C.: Conceptualisation, writing—original draft, software, formal analysis, visualisation, validation. K.A.: Conceptualisation, writing—review and editing, methodology, software, formal analysis, visualisation. G.D.: Conceptualisation, writing—review and editing, methodology. A.C.: Conceptualisation, writing—review and editing, methodology, software. D.R.T.: Conceptualisation, writing—review and editing, methodology. P.B.: Conceptualisation, writing—review and editing, methodology. P.A.T.: Conceptualisation, methodology. A.P.: Conceptualisation, writing—review and editing, methodology, software. F.S.: Conceptualisation, writing—review and editing, methodology, software. B.-C.G.: Conceptualisation, writing—review and editing, methodology, software. F.Y.: Conceptualisation, writing—review and editing, methodology, software. J.V.S.: Conceptualisation, writing—review and editing, methodology, software. Q.V.: Conceptualisation, writing—review and editing, methodology, software. T.E.: Conceptualisation, writing—review and editing, methodology, software. P.K.: Software. R.d.l.R.: Conceptualisation, writing—review and editing, methodology, software. W.W.: Conceptualisation, writing—review and editing, methodology, software. M.B.: Data curation. A.M.: Data curation. K.R.: Data curation. A.B.: Data curation. P.D.V.: Data curation. F.G.: Conceptualisation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: ACIX-III Land was carried out as part of the ESA-funded Quality Assurance for Earth Observation (QA4EO-2) framework contract.

Data Availability Statement: Restrictions apply to the availability of EnMAP and PRISMA data. Data were obtained from the DLR EOWEB® GeoPortal under a stated License Agreement and the PRISMA data portal under PRISMA Data Policy and are available on the respective portals with the permission of the EnMAP and PRISMA ground segment, respectively. The full ACIX-III results, including graphs, can be accessed openly on the CEOS Cal/Val Portal. The code used for the comparative analysis is hosted on the ESA CEM-PAL environment and accessible to users upon registration to the workspace.

Acknowledgments: We thank all the PI(s) and Co-I(s) and their staff for establishing and maintaining the AERONET, RadCalNet, and HYPERNETS sites used in this investigation. We acknowledge CNES, ESA, and NPL for maintaining the RadCalNet stations involved in the study. The HYPERNETS data was created in the context of the H2020/HYPERNETS project funded by the European Union's Horizon 2020 research and innovation programme and the HYPERNET-POP project funded by ESA. We acknowledge RBINS and NPL for maintaining the HYPERNETS sites. We further thank the teams of GFZ for access to the EnMAP Validation campaign data and ESA for access to the Hypersense campaign data. Special thanks to Aime Meygret and their team in CNES for processing the corresponding RadCalNet measurements.

Conflicts of Interest: Authors Noelle Cremer and Georgia Doxani were employed by the company Serco. Author Jorge Vicent Servera was employed by the company Magellium. Author Maximilian Brell was employed by the company HySpex by NEO. The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Figures A1 and A2 show scatterplots (for PRISMA and EnMAP, respectively) intercomparing all processors and the reference, with the wavelength on the horizontal axis and R^2 on the vertical axis. For a single wavelength, R^2 is computed between two processors

over all scenes. This step is repeated for every wavelength separately. Wavelengths with more than two retrieved surface reflectance values are kept. R^2 with respect to each wavelength is plotted for every pair for processors (and the reference). This graph allows for a more in-depth, direct comparison between processors.

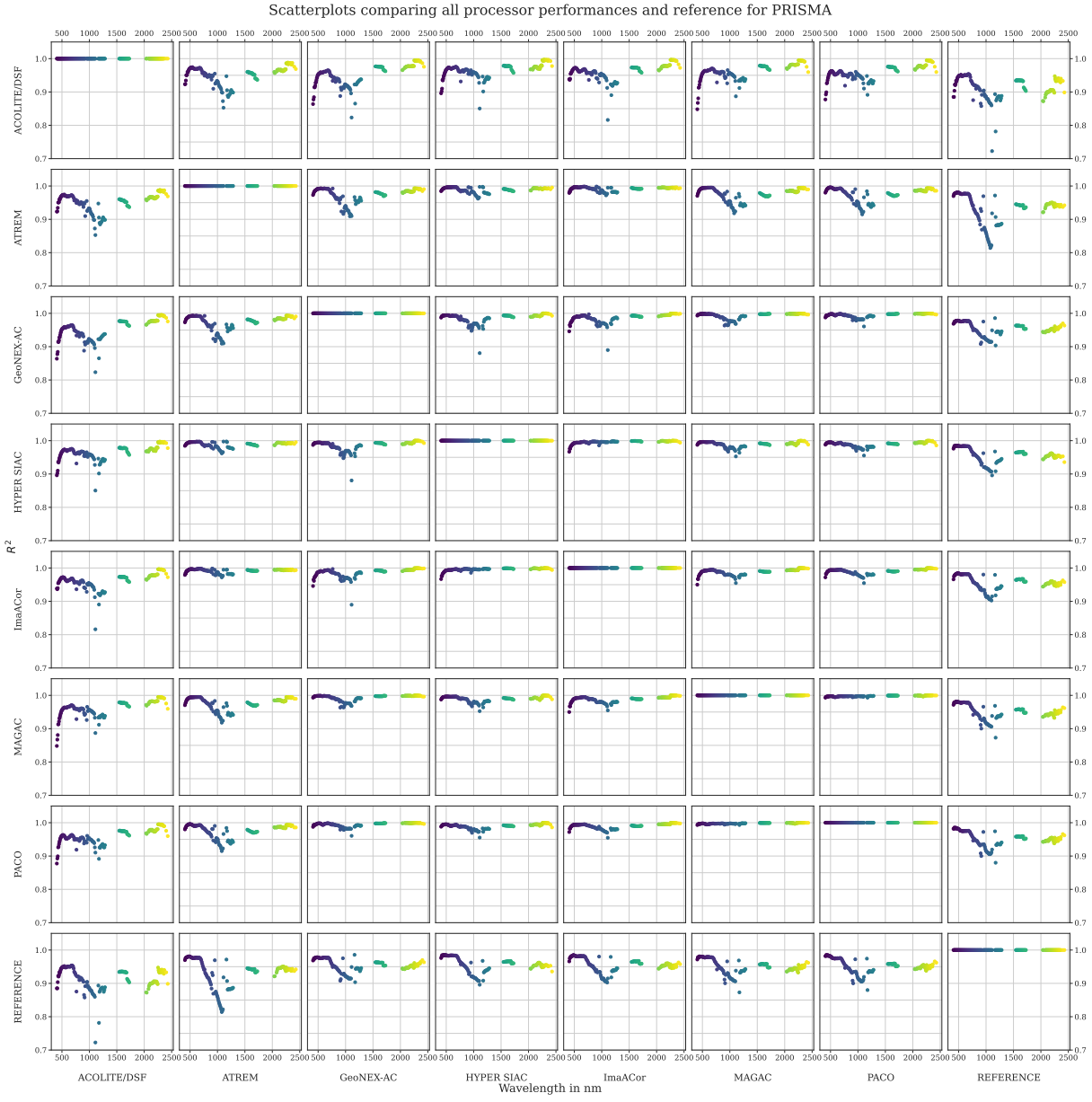

Figure A1. Scatterplot of R^2 per wavelength of all processors and the reference for PRISMA scenes. The colour scale from blue to yellow refers to the wavelengths from the VIS to the SWIR.

For both PRISMA and EnMAP, all processors show similar correlations with respect to the reference data, with a slightly lower overall R^2 value for ACOLITE/DSF, as discussed already in the Results and Discussion sections of this paper. ACOLITE/DSF also shows the strongest differences when directly compared to the other processors. However, some processors also show remarkably similar results over both PRISMA and EnMAP scenes. For PRISMA, those include MAGAC/PACO and ImaACor/HYPER SIAC. Over the included EnMAP scenes, performance with respect to the reference data are more uniform, with very similar results for MAGAC, PACO, ATREM, ImaACor, and HYPER SIAC. For PRISMA, processors show a consistent broad dip around 1000 nm in which the processors align the least. EnMAP shows larger inconsistencies at the end of the spectrum.

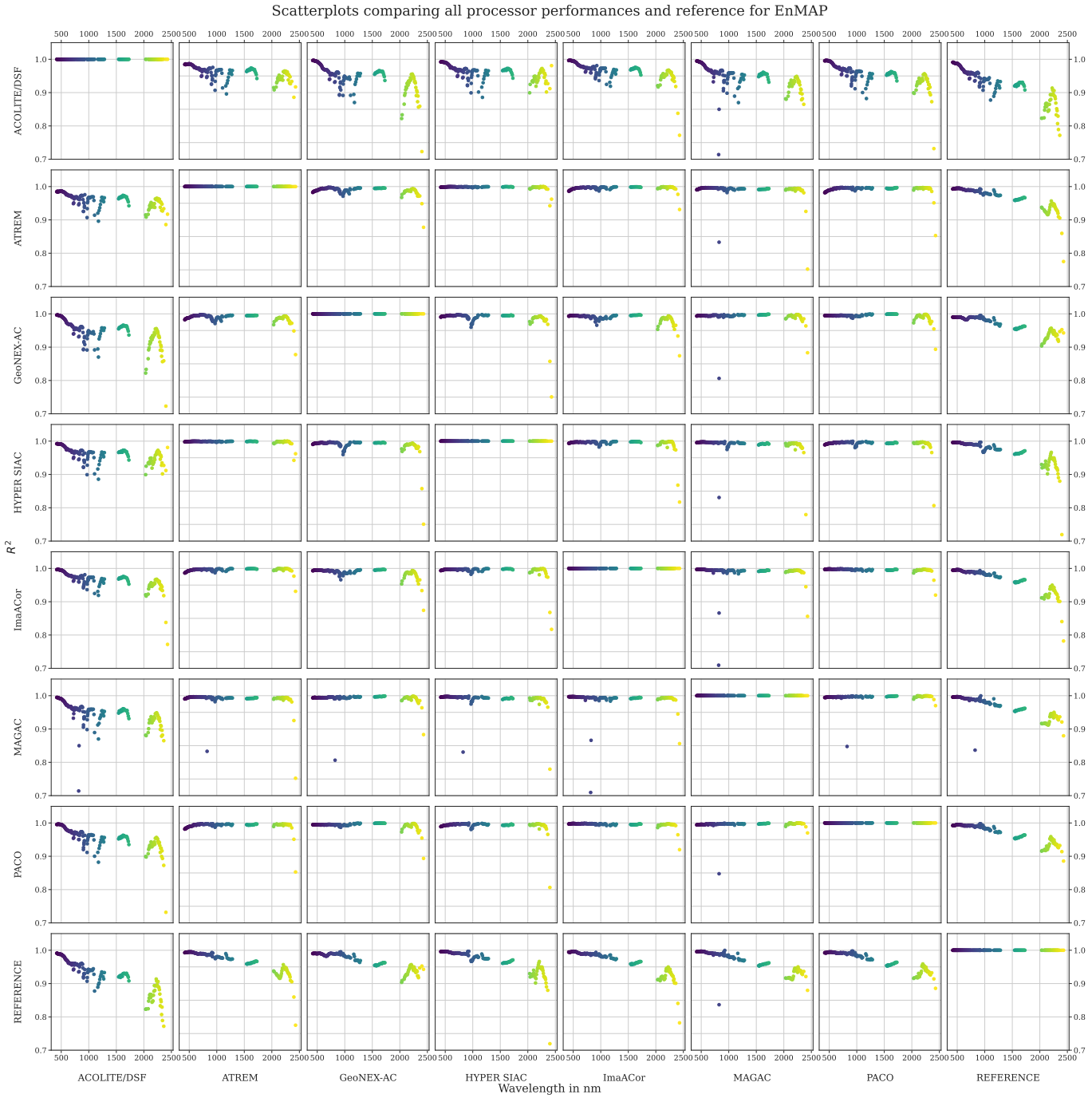

Figure A2. Scatterplot of R^2 per wavelength of all processors and the reference for EnMAP scenes. The colour scale from blue to yellow refers to the wavelengths from the VIS to the SWIR.

References

1. Green, R.O.; Mahowald, N.; Ung, C.; Thompson, D.R.; Bator, L.; Bennet, M.; Bernas, M.; Blackway, N.; Bradley, C.; Cha, J.; et al. The Earth Surface Mineral Dust Source Investigation: An Earth Science Imaging Spectroscopy Mission. In Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2020; pp. 1–15. [CrossRef]
2. Eckardt, A.; Horack, J.; Lehmann, F.; Krutz, D.; Drescher, J.; Whorton, M.; Soutullo, M. DESIS (DLR Earth Sensing Imaging Spectrometer for the ISS-MUSES platform). In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 6–31 July 2015; pp. 1457–1459. [CrossRef]
3. PlanetLabs. First Light Images from Tanager-1 Hyperspectral Satellite. 2024. Available online: https://www.planet.com/pulse/first-light-images-from-tanager-1-hyperspectral-satellite (accessed on 20 May 2025).
4. OrbitalSidekick. OSK Debuts Initial Hyperspectral Insights from GHOSt Constellation. 2023. Available online: https://www.orbitalsidekick.com/news-blog/osk-debuts-initial-hyperspectral-insights-from-ghost (accessed on 20 May 2025).

5. Gao, B.C.; Montes, M.J.; Davis, C.O.; Goetz, A.F. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sens. Environ.* **2009**, *113*, S17–S24 . [CrossRef]
6. Doxani, G.; Vermote, E.; Roger, J.C.; Gascon, F.; Adriaensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F.; et al. Atmospheric Correction Inter-Comparison Exercise. *Remote Sens.* **2018**, *10*, 352. [CrossRef] [PubMed]
7. Doxani, G.; Vermote, E.F.; Roger, J.C.; Skakun, S.; Gascon, F.; Collison, A.; De Keukelaere, L.; Desjardins, C.; Frantz, D.; Hagolle, O.; et al. Atmospheric Correction Inter-comparison eXercise, ACIX-II Land: An assessment of atmospheric correction processors for Landsat 8 and Sentinel-2 over land. *Remote Sens. Environ.* **2023**, *285*, 113412. [CrossRef]
8. Pahlevan, N.; Mangin, A.; Balasubramanian, S.V.; Smith, B.; Alikas, K.; Arai, K.; Barbosa, C.; Bélanger, S.; Binding, C.; Bresciani, M.; et al. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sens. Environ.* **2021**, *258*, 112366. [CrossRef]
9. Skakun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Frantz, D.; Gascon, F.; Gómez-Chova, L.; Hagolle, O.; et al. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990. [CrossRef]
10. Ruddick, K.G.; Bialek, A.; Brando, V.E.; De Vis, P.; Dogliotti, A.I.; Doxaran, D.; Goryl, P.; Goyens, C.; Kuusk, J.; Spengler, D.; et al. HYPERNETS: A network of automated hyperspectral radiometers to validate water and land surface reflectance (380–1680 nm) from all satellite missions. *Front. Remote Sens.* **2024**, *5*, 1372085. [CrossRef]
11. de Los Reyes, R.; Langheinrich, M.; Alonso, K.; Bachmann, M.; Carmona, E.; Gerasch, B.; Holzwarth, S.; Marshall, D.; Müller, R.; Pato, M.; et al. Atmospheric Correction of DESIS and EnMAP Hyperspectral Data: Validation of L2a Products. In Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 1034–1037. [CrossRef]
12. de los Reyes, R.; Langheinrich, M.; Alonso, K.; Carmona, E.; Gerasch, B.; Holzwarth, S.; Marshall Ingram, D.; Mueller, R.; Pato, M.; Pflug, B.; et al. Inter-comparison of PACO BOA surface reflectance between multi- and hyperspectral sensors: EnMAP overpasses with DESIS, Sentinel-2, Landsat and Calval sites. In Proceedings of the LVPE 2023, Rome, Italy, 12–14 June 2023. Available online: https://elib.dlr.de/199142/ (accessed 20 May 2025).
13. De Vis, P.; Howes, A.; Vanhellemont, Q.; Bialek, A.; Morris, H.; Sinclair, M.; Ruddick, K. Feasibility of satellite vicarious calibration using HYPERNETS surface reflectances from Gobabeb and Princess Elisabeth Antarctica sites. *Front. Remote Sens.* **2024**, *5*, 1323998. [CrossRef]
14. Cooley, T.; Anderson, G.; Felde, G.; Hoke, M.; Ratkowski, A.; Chetwynd, J.; Gardner, J.; Adler-Golden, S.; Matthew, M.; Berk, A.; et al. FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. *IEEE Int. Geosci. Remote Sens. Symp.* **2002**, *3*, 1414–1418. [CrossRef]
15. Poutier, L.; Miesch, C.; Lenot, X.; Achard, V.; Boucher, Y. COMANCHE and COCHISE: Two reciprocal atmospheric codes for hyperspectral remote sensing. In *2002 AVIRIS Earth Science and Applications Workshop Proceedings*; Jet Propulsion Laboratory: Pasadena, CA, USA, 2002; pp. 1059–0889.
16. Stefan, A.; Sindy, S.; De Keukelaere, L.; Van De Kerchove, R.; Knaeps, E. Atmospheric Correction Icor and Integration in Operational Workflows. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 3524–3526. [CrossRef]
17. Thompson, D.R.; Natraj, V.; Green, R.O.; Helmlinger, M.C.; Gao, B.C.; Eastwood, M.L. Optimal estimation for imaging spectrometer atmospheric correction. *Remote Sens. Environ.* **2018**, *216*, 355–373. [CrossRef]
18. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [CrossRef]
19. Zhai, P.W.; Hu, Y. An improved pseudo spherical shell algorithm for vector radiative transfer. *J. Quant. Spectrosc. Radiat. Transfe* **2022**, *282*, 108132. [CrossRef]
20. Lyapustin, A.; Knyazikhin, Y. Green's function method for the radiative transfer problem. I. Homogeneous non-Lambertian surface. *Appl. Opt.* **2001**, *40*, 3495–3501. [CrossRef]
21. Schläpfer, D.; Borel, C.C.; Keller, J.; Itten, K.I. Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. *Remote Sens. Environ.* **1998**, *65*, 353–366. [CrossRef]
22. Inness, A.; Ades, M.; Agustí-Panareda, A.; Barré, J.; Benedictow, A.; Blechschmidt, A.M.; Dominguez, J.J.; Engelen, R.; Eskes, H.; Flemming, J.; et al. The CAMS reanalysis of atmospheric composition. *Atmos. Chem. Phys.* **2019**, *19*, 3515–3556. [CrossRef]
23. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding*; World Scientific: Singapore, 2000. [CrossRef]
24. Guanter, L.; González-Sanpedro, M.D.C.; and, J.M. A method for the atmospheric correction of ENVISAT/MERIS data over land targets. *Int. J. Remote Sens.* **2007**, *28*, 709–728. [CrossRef]
25. Pflug, B.; Louis, J.; Debaecker, V.; Mueller-Wilm, U.; Quang, C.; Gascon, F.; Boccia, V. Next updates of atmospheric correction processor Sen2Cor. In *Image and Signal Processing for Remote Sensing XXVI*; Bruzzone, L., Bovolo, F., Santi, E., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2020; Volume 11533, p. 1153304. [CrossRef]

26. Richter, R.; Schläpfer, D. Atmospheric and Topographic Correction (ATCOR Theoretical Background Document). DLR-IB 564-03. 2019. Available online: https://rese-apps.com/pdf/atcor_atbd.pdf (accessed 14 October 2025).
27. de Los Reyes, R.; Langheinrich, M.; Bachmann, M. EnMAP ground segment-level 2A Processor (atmospheric correction over land) ATBD. *EN-PCV-TN-6007* **2023**, *2*. Available online: https://www.enmap.org/data/doc/EN-PCV-TN-6007_Level_2A_Processor_Atmospheric_Correction_Land.pdf (accessed on 20 May 2025).
28. Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. Prisma: The Italian Hyperspectral Mission. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 175–178. [CrossRef]
29. De Luca, G.; Carotenuto, F.; Genesio, L.; Pepe, M.; Toscano, P.; Boschetti, M.; Miglietta, F.; Gioli, B. Improving PRISMA hyperspectral spatial resolution and geolocation by using Sentinel-2: Development and test of an operational procedure in urban and rural areas. *ISPRS J. Photogramm. Remote Sens.* **2024**, *215*, 112–135. [CrossRef]
30. Vicent, J.; Verrelst, J.; Sabater, N.; Alonso, L.; Rivera-Caicedo, J.P.; Martino, L.; Muñoz-Marí, J.; Moreno, J. Comparative analysis of atmospheric radiative transfer models using the Atmospheric Look-up table Generator (ALG) toolbox (version 2.0). *Geosci. Model Dev.* **2020**, *13*, 1945–1957. [CrossRef] [PubMed]
31. Vicent Servera, J.; Martino, L.; Verrelst, J.; Camps-Valls, G. Multifidelity Gaussian Process Emulation for Atmospheric Radiative Transfer Models. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–10. [CrossRef]
32. Shettle, E.; Fenn, R. Models for the Aerosols of the Lower Atmosphere and the Effects of Humidity Variations on their Optical Properties. *Environ. Res.* **1979**, 94, 676 . Available online: https://web.gps.caltech.edu/~vijay/Papers/Aerosol/SF79-Aerosol-Models-part1of4.PDF (accessed on 20 May 2025).
33. Kinne, S. Aerosol radiative effects with MACv2. *Atmos. Chem. Phys.* **2019**, *19*, 10919–10959. [CrossRef]
34. Vanhellemont, Q. ACOLITE: Generic Atmospheric Correction Module. 2021. Available online: https://github.com/acolite/acolite (accessed on 10 October 2025).
35. Vanhellemont, Q.; Ruddick, K. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote Sens. Environ.* **2018**, *216*, 586–597. [CrossRef]
36. Gao, B.C.; Heidebrecht, K.B.; Goetz, A.F. Derivation of scaled surface reflectances from AVIRIS data. *Remote Sens. Environ.* **1993**, *44*, 165–178 . [CrossRef]
37. Gao, B.C.; Davis, C.O. Development of a line-by-line-based atmosphere removal algorithm for airborne and spaceborne imaging spectrometers. In *Imaging Spectrometry III*; Descour, M.R., Shen, S.S., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 1997; Volume 3118, pp. 132–141. [CrossRef]
38. Gao, B.C.; Montes, M.J.; Ahmad, Z.; Davis, C.O. Atmospheric correction algorithm for hyperspectral remote sensing of ocean color from space. *Appl. Opt.* **2000**, *39*, 887–896. [CrossRef]
39. Wang, W.; Wang, Y.; Lyapustin, A.; Hashimoto, H.; Park, T.; Michaelis, A.; Nemani, R. A Novel Atmospheric Correction Algorithm to Exploit the Diurnal Variability in Hypertemporal Geostationary Observations. *Remote Sens.* **2022**, *14*, 964. [CrossRef]
40. Yin, F.; Lewis, P.E.; Gómez-Dans, J.L. Bayesian atmospheric correction over land: Sentinel-2/MSI and Landsat 8/OLI. *Geosci. Model Dev.* **2022**, *15*, 7933–7976. [CrossRef]
41. Santini, F.; Palombo, A. Physically Based Approach for Combined Atmospheric and Topographic Corrections. *Remote Sens.* **2019**, *11*, 1218. [CrossRef]
42. Palombo, A.; Santini, F. ImaACor: A Physically Based Tool for Combined Atmospheric and Topographic Corrections of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 76. [CrossRef]
43. Santini, F.; Palombo, A. Impact of Topographic Correction on PRISMA Sentinel 2 and Landsat 8 Images. *Remote Sens.* **2022**, *14*, 3903. [CrossRef]
44. de los Reyes, R.; Langheinrich, M.; Schwind, P.; Richter, R.; Pflug, B.; Bachmann, M.; Müller, R.; Carmona, E.; Zekoll, V.; Reinartz, P. PACO: Python-Based Atmospheric Correction. *Sensors* **2020**, *20*, 1428. [CrossRef]
45. Storch, T.; Honold, H.P.; Chabrillat, S.; Habermeyer, M.; Tucker, P.; Brell, M.; Ohndorf, A.; Wirth, K.; Betz, M.; Kuchler, M.; et al. The EnMAP imaging spectroscopy mission towards operations. *Remote Sens. Environ.* **2023**, *294*, 113632. [CrossRef]
46. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sens.* **2015**, *7*, 8830–8857. [CrossRef]
47. EnMAP.org. EnMAP: Data and Access. 2025. Available online: https://www.enmap.org/data_access/ (accessed on 20 May 2025).
48. Fontenla, J.M.; Harder, J.; Livingston, W.; Snow, M.; Woods, T. High-resolution solar spectral irradiance from extreme ultraviolet to far infrared. *J. Geophys. Res. Atmos.* **2011**, *116*, D20108. [CrossRef]
49. Holben, B.; Eck, T.; Slutsker, I.; Tanré, D.; Buis, J.; Setzer, A.; Vermote, E.; Reagan, J.; Kaufman, Y.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [CrossRef]

50. Giles, D.; Sinyuk, A.; Sorokin, M.; Schafer, J.; Smirnov, A.; Slutsker, I.; Eck, T.; Holben, B.; Lewis, J.; Campbell, J.; et al. Advancements in the Aerosol Robotic Network (AERONET) Version 3 Database—Automated Near Real-Time Quality Control Algorithm with Improved Cloud Screening for Sun Photometer Aerosol Optical Depth (AOD) Measurements. *Atmos. Meas. Tech. Discuss.* **2019**, *12*, 169–209. [CrossRef]
51. Liang, S.; Wang, J. Chapter 4—Atmospheric correction of optical imagery. In *Advanced Remote Sensing*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2020; pp. 131–156. [CrossRef]
52. Bouvet, M.; Thome, K.; Berthelot, B.; Bialek, A.; Czapla-Myers, J.; Fox, N.P.; Goryl, P.; Henry, P.; Ma, L.; Marcq, S.; et al. RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range. *Remote Sens.* **2019**, *11*, 2401. [CrossRef]
53. Meygret, A.; Santer, R.P.; Berthelot, B. ROSAS: A robotic station for atmosphere and surface characterization dedicated to on-orbit calibration. *SPIE Proc.* **2011**, *8153*, 815311. [CrossRef]
54. Marcq, S.; Meygret, A.; Bouvet, M.; Fox, N.; Greenwell, C.; Scott, B.; Berthelot, B.; Besson, B.; Guilleminot, N.; Damiri, B. New Radcalnet Site at Gobabeb, Namibia: Installation of the Instrumentation and First Satellite Calibration Results. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6444–6447. [CrossRef]
55. Ma, L.; Zhao, Y.; Woolliams, E.R.; Dai, C.; Wang, N.; Liu, Y.; Li, L.; Wang, X.; Gao, C.; Li, C.; et al. Uncertainty Analysis for RadCalNet Instrumented Test Sites Using the Baotou Sites BTCN and BSCN as Examples. *Remote Sens.* **2020**, *12*, 1696. [CrossRef]
56. radcalnet.org. Welcome to the Radiometric Calibration Network Portal. 2025. Available online: https://www.radcalnet.org/ (accessed on 20 May 2025).
57. De Vis, P.; Goyens, C.; Hunt, S.; Vanhellemont, Q.; Ruddick, K.; Bialek, A. Generating hyperspectral reference measurements for surface reflectance from the LANDHYPERNET and WATERHYPERNET networks. *Front. Remote Sens.* **2024**, *5*, 1347230. [CrossRef]
58. Morris, H.; Sinclair, M.; Vis, P.D.; Bialek, A. Utilising LANDHYPERNET data products over a deciduous broadleaf forest to validate Sentinel-2 and Landsat surface reflectance products. *Front. Remote Sens.* **2024**, *5*, 1322760. [CrossRef]
59. ares-observatory.ch. ESA CHIME and SBG 2021 Mission Information, 2025. Available online: https://ares-observatory.ch/esa_chime_mission_2021/ (accessed on 20 May 2025).
60. Malthus, T.; Ong, C.; Lau, I.; Fearns, P.; Byrne, G.; Thankappan, M.; Chisholm, L.; Suarez Barranco, L.; Clarke, K.; Scarth, P.; et al. *A Community Approach to the Standardised Validation of Surface Reflectance Data—A Technical Handbook to Support the Collection of Field Reflectance Data*; CSIRO: Canberra, Australia, 2019. [CrossRef]
61. Brell, M.; Guanter, L.; Segl, K.; Chabrillat, S.; Scheffler, D.; Soppa, M.; Bohn, N.; Gorroño, J.; Kokhanovsky, A.; Bracher, A.; et al. Assessment of Enmap Data Quality Through Global Product Validation Activities. *Prepr. Submitt. Remote Sens. Environ.* **2024**, *in press*. [CrossRef]
62. Brell, M.; Milewski, R.; Neumann, C.; Ong, C.; Lau, I.; Hank, T.; Förster, S.; Chabrillat, S. *EnMAP Field Guide—In-Situ Measurements for Validation Purposes. V.1.3.*; EnMAP Field Guides Technical Report; GFZ Data Services: Potsdam, Germany, 2023. [CrossRef]
63. Ong, C.; Caccetta, M.; Lau, I.; Ong, L.; Middleton, E. Compositional characterisation of the pinnacles vicarious calibration site. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 3059–3062. [CrossRef]
64. Heller Pearlshtien, D.; Ben-Dor, E. Calval Evaluation of Desis Products in Amiaz Plain and Makhtesh Ramon Test Sites, Southern Israel. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *46W1*, 13–21. [CrossRef]
65. Thompson, D.R.; Green, R.O.; Bradley, C.; Brodrick, P.G.; Mahowald, N.; Dor, E.B.; Bennett, M.; Bernas, M.; Carmon, N.; Chadwick, K.D.; et al. On-orbit calibration and performance of the EMIT imaging spectrometer. *Remote Sens. Environ.* **2024**, *303*, 113986. [CrossRef]
66. Di Mauro, B.; Cogliati, S.; Bohn, N.; Traversa, G.; Garzonio, R.; Tagliabue, G.; Bramati, G.; Cremonese, E.; Julitta, T.; Guanter, L.; et al. Evaluation of PRISMA Products Over Snow in the Alps and Antarctica. *Earth Space Sci.* **2024**, *11*, e2023EA003482. [CrossRef]
67. Remer, L.A.; Tanre, D.; Kaufman, Y.K.; Levy, R.; Matoo, S. Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS: Collection 5, Product ID: MOD04/MYD04, 2009. Available online: https://modis.gsfc.nasa.gov/data/atbd/atbd_mod02.pdf (accessed on 20 May 2025).
68. Vermote, E.F.; Kotchenova, S. Atmospheric correction for the monitoring of land surfaces. *J. Geophys. Res. Atmos.* **2008**, *113*, D23S90. [CrossRef]
69. Alonso, K.; Cremer, N.; Boccia, V.; Brodrick, P.G.; Chlus, A.; Doxani, G.; Gascon, F.; Niemeijer, S.; Thompson, D.R.; Townsend, P.; et al. Integration of ACIX-III Land Atmospheric Correction Inter-comparison eXercise within the Copernicus Expansion Mission Product Algorithm Laboratory to Support Surface Reflectance Cal/Val. In Proceedings of the EGU General Assembly Conference, Vienna, Austria, 14–19 April 2024; p. 20397. [CrossRef]
70. Clevers, J.; Kooistra, L.; Schaepman, M. Using spectral information from the NIR water absorption features for the retrieval of canopy water content. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 388–397. [CrossRef]

71. Niro, F.; Colombo, R.; Cogliati, S.; Genesio, L.; Migiletta, F.; Kuusk, J.; Saunier, S.; Berthelot, B.; Origo, N.; Kocaman, S.; et al. Cal/Val Park: Fostering innovation and international cooperation in the Cal/Val domain. In Proceedings of the AGU Fall Meeting, Washington, DC, USA, 9–13 December 2024; Volume 2024, pp. B13G–162.
72. Joint Research Centre (JRC). RAdiation Transfer Model Intercomparison for Atmosphere (RAMI4ATM) Web Portal. 2025. Available online: https://rami-benchmark.jrc.ec.europa.eu (accessed on 20 May 2025).
73. Braga, F.; Fabbretto, A.; Vanhellemont, Q.; Bresciani, M.; Giardino, C.; Scarpa, G.M.; Manfè, G.; Concha, J.A.; Brando, V.E. Assessment of PRISMA water reflectance using autonomous hyperspectral radiometry. *ISPRS J. Photogramm. Remote Sens.* **2022**, *192*, 99–114. [CrossRef]