Intercomparison of AOD retrievals from GAW-PFR and SKYNET sun photometer networks and the effect of calibration

Angelos Karanikolas[1,2], Natalia Kouremeti[1], Monica Campanelli[3], Victor Estellés[4,3], Masahiro Momoi[5], Gaurav Kumar[4], Stephan Nyeki[1] and Stelios Kazadzis[1]

[1] World Optical Depth Research and Calibration Centre (WORCC), Physikalisch-Meteorologisches Observatorium Davos/World Radiation Center (PMOD/WRC), Davos Dorf, 7260, Switzerland
[2] Institute for Particle Physics and Astrophysics, ETH Zurich, Zurich, 8093, Switzerland
[3] Institute of Atmospheric Sciences and Climate (ISAC), Consiglio Nazionale Delle Ricerche (CNR), Rome, 00133, Italy
[4]Earth Physics and Thermodynamics Department, Universitat de València, Valencia, 46100, Spain
[5]GRASP SAS, Lezennes, 59260, France

Correspondence to: Angelos Karanikolas (angelos.karanikolas@pmodwrc.ch)

Abstract. In this study, we assess the homogeneity of aerosol optical depth (AOD) between ~~the~~ two sun photometer networks, the Global Atmospheric Watch-Precision Filter Radiometer (GAW-PFR) and the European Skynet Radiometers network (ESR) at ~~the 2~~ both common wavelengths of their main instruments (500 nm and 870 nm). The main focus of ~~the~~ this work is ~~on~~ to ~~evaluating~~ evaluate the effect of the Improved Langley plot calibration method~~,~~ (ILP) used by SKYNET, and to investigate~~ing~~ the factors affecting its performance. We used data from three intercomparison campaigns that took place ~~in~~ during ~~the period~~ 2017 - 2021. Each campaign was organized ~~has had two phases~~ ~~in~~ at two locations (~~. One is~~ mountainous rural, ~~(~~Davos, Switzerland; ~~) and the other~~ urban, ~~(R~~Rome, Italy). Our analysis shows that ~~the AOD d~~differences in AOD due to post~~-~~processing and instrument differences are minor. The main factor leading to AOD differences is the calibration method. We found a systematic underestimation of ESR AOD compared to GAW-PFR due to underestimation of the calibration constant calculated with the ILP method compared to the calibration transfers using the PFR as a reference~~The major factor leading to AOD differences is the calibration method, where we found a systematic underestimation of AOD compared to GAW-PFR, due to an underestimation in of the ILP calibration~~. The calibration and AOD differences are smaller in Davos, where ~~at 870 nm~~ the traceability criteria are satisfied at 870 nm and ~~at 500 nm~~ the median differences are below 0.01 at 500 nm. In Rome, ~~at 500 nm~~ the AOD median differences ~~per campaign~~ at 500 nm were~~are~~ in the ~~range~~~~between~~ 0.015 - 0.~~035~~034 range. ~~In an Aattemptingt to explain the differences, we found no association between the calibration performance and the level or the variability of the aerosol properties.~~ We ~~also~~ conducted a sensitivity study, which shows that part of the difference can be potentially explained by errors in the assumed surface albedo and instrument solid view angle provided as inputs to the ILP code (based on Skyrad pack 4.2). Our findings suggest that the ILP method is mainly sensitive to the measured sky radiance. The underestimation in calibration ~~underestimation~~ is probably caused by an error ~~on~~ of the retrieved scattering ~~aerosol optical depth~~AOD (sc-AOD) through the sky radiance inversion. Using an alternative retrieval method (Skyrad MRI pack version 2) to derive sc-AOD and ~~repeat ILP calibratio~~use it to recalibrate the instruments with the ILP method~~n~~, we found no significant differences between the retrieved sc-AOD ~~n~~or a

systematic increase ofin the ILP derived calibration constant when using the MRI pack for sc-AOD inversion instead of the Skyrad 4.2between the retrieved sc-AOD nor systematic increase of the calibrations. The potential error may be a result of the model assumptions used for the sky radiance simulationsforward model assumptions. In conclusion, To conclude, the on-site calibration of sun photometers on sitehas several advantages: offers the advantage of avoiding instrument shipmentss s and data gaps can be avoided. However, it has also the disadvantage, ILP shows of a larger uncertainty and significant systematic differences compared to the traditional Langley calibration performed under low and constant AOD conditions at high high-altitude sites, . due The larger uncertainty of the ILP method can be attributed to to the uncertainties of the calibration method and the required modelling and input parameters needed for it. In the following sections, we report on results on from the AOD retrievals of several instruments in different environments using different principles approaches in their calibration methods. We also perform an investigate ion to explain the causes of the differences.

1 Introduction

Atmospheric particulate matter (aerosols) is a component of high importance in atmospheric sciences and modern environmental problems. They scatter and absorb solar radiation significantly affecting the Earth's energy budget. They also greatly assist water and ice nucleation in the atmosphere leading to the formation of clouds (Winkler & Wagner, 2022; Maloney et al., 2022). Aerosols were have been the major main driver of variations in surface solar radiation variations for several decades (Wild, 2012; et al., Correa et al., 20232024). Affecting Their influence onthe surface solar radiation can alters the exposure of organisms to biologically active radiation (Barnes et al., 2019; Bais et al., 2018) and also the efficiency of solar energy production systems capabilities (Papachristopoulou et al., 2023; Hou et al., 2022). Both their direct and indirect effects of aerosols on surface solar radiation can lead to are a significant forcings of the climate. Aerosols therefore and rrepresent a emain the source of the largeste uncertainty in the attribution of radiative forcing attribution (IPCC, 20212023).

According to the World Meteorological Organization (WMO), the most important parameter related to aerosols for Earth energy budget studies is the aerosol optical depth (AOD) (WMO, 2003). AOD describes the overall effect of the total aerosol column on the attenuation of solar radiation, and is attenuation. AOD is correlated with the total aerosol load in the atmosphere and its spectral dependence with the size of aerosolsan indicator of the total aerosol load in the atmosphere and its spectral dependence onwith the size of aerosols. AOD is calculated from direct solar irradiance (DSI) measurements by subtracting the effect of gas absorption and scattering at in the absence of clouds covering the solar disk. The main instruments used for this purpose are the sun photometers,. Sun photometerswhich measure the DSI at selected wavelengths, in whichwhere gas absorption is minimal and the AOD calculation can bise more accurate.

There are dDifferent types of sun photometers are used in several worldwide networks. There are several stations using the same type of sun photometer, which belong to an instrument network. The main sun photometer networks are the Aerosol Robotic Network (AERONET), Global Atmospheric Watch-Precision Filter Radiometer (GAW-PFR) and SKYNET.

AERONET is the largest network with more than 400 stations worldwide and uses the CIMEL sun~~-~~ and sky photometer (hereafter CIMEL) as the standard instrument (Holben et al., 1998). ~~The~~ GAW-PFR includes 15 stations mainly in remote worldwide locations. Its standard instrument is the Precision Filter Radiometer (PFR) and includes the WMO AOD reference instruments (PFR-Triad) (Kazadzis et al., 2018b). SKYNET is a multi-instrument research network divided into sub-networks and includes around 100 stations mainly in East Asia and Western Mediterranean regions. Its standard instrument for AOD observation is the PREDE-POM sun and sky radiometer (hereafter POM) (Nakajima et al., 2020). Each sub-network has developed its own calibration protocols and post-~~-~~processing algorithms independently. ~~Especially, t~~P~~wo p~~rocedures developed by two sub-networks, led by ESR~~European Sky Radiometer network (ESR)~~ and the Center for Environmental Remote Sensing (CEReS) ~~of~~ at Chiba University, are recognized as the standard in the International Skynet Committee (Nakajima et al., 2020). Due to the differences among the main networks (i.e., AERONET, GAW-PFR, SKYNET) described above, it is important to evaluate the ~~extent of~~ homogeneity between ~~the networks~~them to ensure that ~~the~~ AOD observations are comparable and ~~of~~ have a similar accuracy. For this purpose, ~~every 5 years~~ the Filter Radiometer Comparison (FRC) campaign takes place in Davos (~~,~~ Switzerland) every five years, which ~~-~~including~~es~~ ~~instruments from~~ all types of sun~~-~~ photometers (Kazadzis et al., 2023). There ~~are~~ have been several other intercomparison campaigns (Doppler et al., 2023; Mitchell & Forgan, 2003; Cacho~~tt~~rro et al.,2009; Mazzola et al. 2012; Nyeki et al., 2013; Kazadzis et al.,2018a; Gröbner et al., 2023), as well as~~but also~~ long-term comparisons between different networks (Cuevas et al, 2019; Karanikolas et al., 2022).

A necessary parameter for ~~the AOD~~ calculation of AOD ~~-~~is the DSI ~~the instrument woul~~that the instrument would measure at the top-of-the atmosphere (extraterrestrial or calibration constant). There are different ways to calibrate a sun photometer. Conventionally, they are calibrated by the standard Langley plot method (SLP) (Shaw et al., 1973) and the calibration transfer from a reference co-located instrument. An alternative method is the laboratory calibration to the international system of units (SI). Under this alternative approach, we can use satellite measurements for the top-of-the atmosphere irradiance ~~satellite measurements~~ that are also in SI units.~~It can be accomplished either by by: i) using a co-located instrument as a reference, ii) by laboratory calibration to the international system of units (SI) and use of satellite measurements for the top-of-the atmosphere, or iii)by using an indirect method to calibrate the instrument through the DSI at the ground. The cConventional ly used methods are the standard Langley plot method (SLP) (Shaw et al., 1973) and the calibration transfer from a reference instrument.~~ Recent developments show that ~~the~~ laboratory calibration can also be accurate (Gröbner ~~&~~ and Kouremeti, 2019; Kouremeti et al., 2022; Gröbner et al., 2023). Another method is the improved Langley plot method (ILP) (Tanaka et al., 1986; Campanelli et al., 2004). This is a modification of the SLP method, which accounts for AOD variations during the day in contrast to SLP that assumes ~~AOD~~ constant AOD. The assumption of constant AOD results ~~to~~ in larger errors in more polluted areas, and ~~hence~~ SLP is therefore only ~~applied~~ used ~~only in~~at high altitude locations. The aim of ILP is to calibrate instruments ~~in~~ at the station where they are normally operated, regardless of the station's location, instead of being transported to a calibration site. ~~Therefore, t~~This method therefore has several advantages: i) instrument ~~ideally brings the advantage to avoid~~ damage during ~~in the~~ transportation can be avoided, ii) there will be a minimal amount of missing data in during the calibration period, iii) low cost in the maintenance is less costly, and iv) the frequently tracking the variation of the calibration constant can be more frequently monitored. AERONET and GAW-PFR calibrate the instruments either by SLP in at Mauna Loa (, Hawaii) and Izaña (, Tenerife) or by calibration transfer from reference instruments, while SKYNET uses the ILP method.

Other than the calibration procedures, each network also uses a different post--processing and cloud--screening algorithms algorithms to derive from DSI and filter the AOD observations. One of the main differences between GAW-PFR with AERONET and SKYNET is the inclusiona correction offor absorption due to nitrogen dioxide (NO2) and water vapor (H2O) absorptionsOne of the main differences are is the inclusion of nitrogen dioxide (NO_2) and water vapor (H_2O) absorption infrom AERONET and SKYNET (Kazadzis et al., 2018a, ; Estellés et al., 2012, ; Drosoglou et al., 2023, ; Sinyuk et al., 2020),). however However, there are also differences in the way the optical depth of ozone absorption and Rayleigh scattering are calculated (Cuevas et al., 2019). The In addition, the cloud--screening algorithms also showexhibit some differences, with the SKYNET algorithm being particularly strict (Kazadzis et al., 2018a).

In order to evaluate the ILP method, GAW-PFR World Optical Depth Research and Calibration Centre (WORCC)(and European Skynet Radiometers network (ESR) have signed and a Memorantum Memorandum of Understanding (MoU) for scientific collaboration, including several intercomparison campaigns organized (Quality and Traceability of Atmospheric Aerosol Measurements or QUATRAM I, II and III). During the period 2017 - – 2021 period, a PFR was transported to Sapienza University (in Rome, Italy) once for each campaign for several weeks or months to measure AOD in parallel with one or more POMs and CIMEL (Table 1 section 2.1) instruments. AlsoIn addition, at least one POM was transported to Davos duringon 3 three different periods as well (Table 1 section 2.1), where the WMO AOD reference (PFR-Triad) and a CIMEL are operated. The POMs were calibrated using both the ILP method and calibration transfer The POMs were both calibrated both with the ILP method and by calibration transfer using with a PFR as a reference. There is already a publication under review showing calibration differences between several calibration methods (Campanelli et al., 2023).

This study aims to assess the AOD differences between GAW-PFR and ESR and the effect of the different calibration approaches.. —In addition, we investigate the extent to which different factors such as atmospheric conditions and input parameters required to perform the ILP method, contribute to the calibration differences. In the previous studies concerning intercomparison of such instruments cited earlierintercomparisons (such aseg. Kazadzis et al., 2023) the study of AOD differences was limited to the differences of AOD provided by each network. In the present study, we also separate the effect of the calibration approaches and the effect of the post-processing and instrument differences. We also include one campaign forat each location with a duration of several months, which provided a significantly larger amount of data compared to the shorter campaigns that are more frequently organized. Finally, we include a detailed analysis of the ILP calibration method in relation to the aerosol properties and its sensitivity to all required input parameters. "In additionAlso,, we investigate the extent to which different factors such as the atmospheric conditions and the input parameters required to perform ILP, contribute to the calibration and hence as a result to retrieved AOD differences.

2 Instruments, calibration ~~and~~ methods and AOD datasets

2.1 Instrumentation and locations

~~The data used are from the period 2017-2021 in two locations, Davos (Switzerland) and Rome (Italy) i~~In order to evaluate the ILP performance under different conditions, we used the sun photometer measurements from the ~~period~~ 2017-2021 period ~~in~~at two locations: Davos (Switzerland) and Rome (Italy)~~data from Davos and Rome are used for the 2017 – 2021 period~~. The ~~st~~station ~~of Davos is~~ at PMOD/WRC (1590 m a.s.l.) is close to Davos, which lies in the ~~next to a town deep in the~~ Eastern Alps~~s~~ mountain region of Switzerland~~range~~. The area has no significant local pollution. Aerosols can reach the area from other parts of Europe due to its proximity ~~with~~ to several European countries and during strong Sahara dust transport episodes. The other station ~~is in Rome a~~at Sapienza University (~~at~~ 83 m a.s.l.) is close to the centre of Rome~~, the capital city of Italy~~.

For this study, we used the sun- photometer, PFRN27 (part of the PFR reference triad), as a reference in Davos ~~(part of the PFR reference triad)~~, while in Rome we used ~~the~~ PFRN14 (2017 - 2019) and PFRN01 (2021). We also used ~~the~~ a co-located CIMEL in each campaign for AOD cross-validation. In total, we compared three POM instruments with the PFRs~~,~~: two ESR network reference (master) instruments (one of the POM masters in two different versions due to modification between QUATRAM II and III to make it suitable for lunar measurements~~one both in its initial and a later modified version~~) and one travelling standard. A summary ~~In table 1 there is a summary with~~of all instruments and~~, the used~~ datasets is shown in Table 1.

Table 1: Reference and comparison ~~The~~ instruments used ~~per~~ at each location ~~as reference and under study~~ including the time periods of the common datasets. *~~ stands for a m~~Modified version of POMCNR ~~that made it~~ suitable for lunar observations.

Location/campaign	PFR Reference~~.~~ instrument	Comparison~~.~~ Instrument(s)	Starting date	End date
DAVOS I	N27	POMVDV/CIMEL#354	09/08/2017	30/08/2017
ROME I	N14	POMVDV	18/10/2017	02/11/2017
ROME I	N14	CIMEL646	05/12/2017	27/02/2018
DAVOS II	N27	POMCNR/CIMEL#354	24/07/2018	19/10/2018
ROME II	N14	POMCNR/~~POM11~~POMSPZ/CIME#L43	02/05/2019	03/10/2019
DAVOS III	N27	POMCNR*/CIMEL#916	08/10/2021	18/10/2021
ROME III	N01	POMCNR*/CIMEL#1270	03/09/2021	20/09/2021

2.1.1 PFR

The ~~Precision Filter Radiometer~~PFR (Wehrli, 2000) is a ~~Sun~~ sun photometer that measures ~~the~~ DSI ~~in~~ at ~~4~~ four wavelengths~~.~~ ~~The channels are~~ nominally centred ~~on~~ at 368, 412, 501 and 862 nm~~. It~~, which is mounted on an independent tracking system to follow the motion of the Sun. The entrance window of the instrument is ~~covered~~ protected by~~with~~ a quartz window and ~~its~~ internal parts are fully protected from ~~the~~ outside conditions. It is filled with dry nitrogen at approximately 2 bar~~,~~ and the ~~. Its~~ internal temperature is kept constant by an active Peltier system at approximately 20ºC with an accuracy of 0.1ºC ~~by an active Peltier system~~. ~~The r~~Radiation passes through the quartz window and interference filters ~~consequently~~ to allow solar radiation from only a narrow spectral region to reach the silicon photodiode detector. ~~, which is~~ . ~~The detector is a silicon photodiode that provides voltage measurements in mV proportional to the received light~~ . ~~Their~~ The full-width-at-half-maximum (FWHM) bandwidth of the filters varies from 3 nm to 5 nm and its field-of-view angle (FOV) is approximately 2º at FWHM. ~~The four channels (filter – silicon detectors) are arranged in a grid.~~ Measurements occur ~~Every~~ every minute, when a shutter opens for 10 seconds to perform the 10 sequential measurements at each wavelength. This ~~,~~ minimiz~~ing~~es the exposure time of the filters to solar radiation, and hence their degradation. The stability of the travelling standard PFRs is validated by calibration ~~of instrument~~ before and after ~~the~~ each campaign~~s~~.

2.1.2 PREDE-POM

The PREDE-POM (Estelles et al., 2012; Prede Co. Ltd., Japan: https://prede.com/english/skyradio.html) is a sun~~-~~ and sky radiometer with a 2-axis stepping motor as a tracking system to perform both direct sun and ~~–~~diffuse sky irradiance observations. The step is 0.0036º per pulse. There are ~~2~~ two major versions of the instrument ~~containing~~ with different wavelengths. POM-01 measures direct solar irradiance and diffuse sky irradiance at ~~7~~ seven wavelengths centred at 315, 400, 500, 675, 870, 940 and 1020 nm. POM-02 is an extended version measuring at 315, 340, 380, 400, 500, 675, 870, 940, 1020, 1627 and 2200 nm. In both cases, the FWHM bandwidth is 2 - 10 nm depending on the channel. The wavelengths are isolated using filters mounted on a filter wheel and the detector is a silicon photodiode except for the case of wavelengths above 1600 nm ~~of~~ in the POM-02, which are measured by an InGaAs photodiode. The FOV of the instrument is approximately 1º~~,~~ and ~~. It~~ includes a temperature control system to ~~keep~~ maintain an internal ~~the~~ temperature ~~at~~ of 30ºC, a 4-element silicon Sun sensor~~,~~ and a rain sensor. In this study, we used a standard POM-01 instrument, while ~~and~~ the rest were ~~a~~ modified POM-01 versions to measure at 340 nm instead of 315 nm.

2.1.3 CIMEL

The CIMEL ~~Sun~~sun~~-~~ and sky photometer (Giles et al., 2019) is an instrument including a 2-axis robotic tracking system. This tracking system allows it to perform direct sun and sky scans in order to measure either DSI ~~and~~ or diffuse sky radiance. There are different versions measuring at different wavelengths. The smallest wavelength is 340 nm and the largest 1640 nm, although for some versions it is 1020 nm. The number of wavelengths is up to 10. In this study, we used CIMELs with at least ~~8~~ eight interference filters centred at 340, 380, 440, 500, 675, 870, 940, and 1020 nm. The ~~bandwidth has a full width at half maximum (~~FWHM~~)~~ ~~of~~ is 10 nm, except for 340~~ and~~, 380 and 1640 nm~~,~~ ~~(which have~~ 2, 4 and ~~4~~ 25 nm ~~FWHM~~, respectively). A silicon detector is used to ~~To~~ measure the radiation~~, it includes a silicon detector~~. ~~The f~~Filters are mounted on a filter~~-~~wheel that moves every second to switch to a different wavelength until all channels are measured in a measurement sequence. The ~~measurement~~ sequence is then repeated ~~3~~ three times within 30 seconds to provide triplet observations. The instrument has a FOV of 1.2°. It also has ~~also~~ a four-quadrant detector, which detects the point of ~~the~~ maximum solar radiation intensity, enabling it to correctly point~~so it can point correctly~~ to the Sun before the measurement sequence starts. ~~The~~ AERONET AOD data are publicly available at ~~3~~ three levels (1.0, 1.5 and 2.0). In this study, we only used ~~only~~ level 2.0, which included cloud~~-~~screening, the final calibration and quality assurance.

2.2 Calibration methods

We used ~~2~~ two different calibration methods to calculate the extraterrestrial constant of the POMs. The ~~Improved Langley Plot method (~~ILP method~~)~~ and a calibration transfer using a PFR as reference.

2.2.1 Improved Langley Plot

The ILP method (Campanelli et al., 2004; Nakajima et al., 2020; Campanelli et al., 2023) is a modification of the conventionally used SLP. The basic principle in both methods is to use the solar radiation measured at the ground during at least a half~~-~~day and the Beer-Lambert-Bouguer law:

$$I = I_0 e^{-m\tau} \quad (1)$$

where *I* is the DSI measured at the ground, I_0 the calibration constant (solar irradiance at the ~~top~~ top-~~of~~ of-~~the~~ the-atmosphere in the units of the instrument), *m* the air mass coefficient, and *τ* the total optical depth of the atmosphere. The solar irradiance is measured in~~at~~ the instrument's units as the SLP and ILP methods do not require conversion to units of W/m^2. The total optical depth is the sum of the scattering and absorption optical depths of the atmospheric constituents.

For the case of ~~Under~~ no clouds, in front of the solar disk:

$$\tau = \tau_R + \tau_g + \tau_a \quad (2)$$

where τ_R the Rayleigh scattering optical depth, τ_g the gas absorption optical depth, and τ_a the extinction aerosol optical depth. Eq. (1) can be written as:

$$\ln I = \ln I_0 + m\tau \quad (3a)$$

or

$$\ln I = \ln I_0 - m\tau_R - m_g\tau_g - m_a\tau_a \quad (3b)$$

The value of τ_R is calculated using the atmospheric pressure. The value of τ_g is calculated from the total column of gases absorbing at a certain wavelength. The values of mg and ma are the air masses corresponding to gases and aerosols,

Formatted: Font: Italic

Formatted: Subscript

Formatted: Subscript

respectivelyThe value of τ_R is calculated using the Knowing the aatmospheric pressure, and τg from we can calculate τR and tthe total column of gases absorbing at a certain wavelength we can calculate τg. The values of *m*g and *m*a are the air masses corresponding to gases and aerosols.

The SLP method uses Eq. (3a) and. bBy measuring the DSI during the day at several known air masses, we can perform a linear fitting procedure, to the pairs of *m* and *I* values. The intercept of the fitted line is the natural logarithm of the calibration constant. This method assumesing that the total optical depth of the atmosphere is constant for at least several hours (slope of the linear fit). However, the optical depth can vary which does not happen under real atmospheric in real conditions. At wavelengths where gas absorption is minor or the gases that absorb radiation show no rapid variability, AOD dominates the total optical depth. The SLP methodat sun photometers (which sun photometers use carefully selected wavelengths to avoid strong absorptions) is applicable with high accuracy in at high altitude locations where the AOD is usually very low and its fluctuations do not have a significant effect on the total optical depth in over timescales of a few hours. On the other hand, the SLP method cannot be used in theat aerosol sites with aerosol polluted sitesion (Shaw et al., 1983; Toledano et al.,2018). In order to avoid the shipment of instruments to such locations and, to increase the frequency of calibration and to monitoring its their status, we require a method that is usable at the any type of station where the instrument is operated. The ILP method was developed for this purpose. Instead of using Eq. (3a), we can use a modified version of Eq. (3b) is used, which is now described.

Considering tThe Rayleigh scattering and gas absorption optical depths can be calculatedknown, so AOD τ_a. is the only required parameter to be retrieved before we calculatederiving the calibration constant. In the ILP method, instead of AOD τ_a, tthe used paramparameter eter is the scattering aerosol optical depth (τ_{sc}sc-AOD) is used as a parameter. If ω is the single scattering albedo (SSA), τ_a the AOD and τ_{sc} the sc-AOD then $\tau_{sc} = \omega\tau_a$, and which leads to Eq. (3b) taking takes the form:

$$lnI + m\tau_R + m\tau_g = lnI_0 - m\frac{\tau_{sc}}{\omega} \quad (4)$$

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Assuming $y = lnI + m\tau_R + m\tau_g$ and $x = m\tau_{sc}$ we get obtain a straight line $y = ax + b$ where the slope is $a = -\frac{1}{\omega}$ and $b = lnI_0$.

Therefore, calculating τ_{sc} for several times during the day, we can apply a linear fitting to all pairs of x and y values and calculate the calibration constant. This method takes into account the variability of the AOD but assumes constant SSA during the measurement period instead. Therefore, large variability of SSA can affect the accuracy of the method.

The estimation of τ_{sc} is possible through inversion modelling (by Skyrad pack code version 4.2 in our case) applied to the angular distribution of normalized sky radiance (NSR) (Eq. 5) observed in almucantar geometry at scattering angles up to 30°. The NSR is defined in Eq. 5:

$$R(\theta) = \frac{E(\theta)}{m\Omega I} \quad (5)$$

where *E* is the measured diffuse sky irradiance, θ the scattering angle, *m* the air mass, Ω the solid view angle (SVA) of the instrument and *I* the direct solar irradiance.

The model estimates sc-AOD the τ_{sc} and the aerosol phase function by retrieving the size distribution with an a-priori refractive index. To model the radiative transfer and to retrieve τ_{sc}sc-AOD, the surface albedo (SA), the total ozone column (TOC) and the surface pressure (P) are also required as inputs.

The Skyrad code also derives also ω SSA and therefore τ_aAOD, but it is not used in the ILP calibration. However, it is used for a screening criterion as all values corresponding to AOD $\tau_a \geq$ 0.4 are rejected before the final calibration.

2.2.2 Calibration transfer and AOD calculation

To evaluate ILP, we calibrated the POMs using a PFR as a reference for the campaign. Awe begin by assuming thate two co-located instruments (a PFR and a POM) that measure DSI at the same wavelength. If I_1 is the DSI at the ground measured with a PFR, I_2 is the DSI measured with a POM at the same time, I_{01} the PFR calibration constant and I_{02} the POM calibration constant then the irradiances satisfy the following equationTo evaluate ILP, we calibrated the POMs using a PFR as a reference for each case. For measurements of DSI from co-located instruments at the same wavelength with I_1 being the DSI at the ground measured with a from PFR, I_2 is the DSI measured with a from POM at the same time, I_{01} the PFR calibration constant, and I_{02} the POM calibration constant:

$$\frac{I_1(\lambda,t)}{I_{2(\lambda,t)}} = \frac{I_{01(\lambda)}}{I_{02(\lambda)}} \qquad (6a)$$

The POM calibration constant is:

$$I_{02} = I_{01}\frac{I_2}{I_1} \qquad (6b)$$

Therefore, we used the raw signal ratio of the instruments for measurements with a maximum of 30 sec time difference and the known calibration of the PFR to calculate the calibration for the POM. The calibration constants and raw signals are in the instrument units (different for each instrument).each instrument measures and Theyand were corrected for the an Earth-Sun distance of differences by shifting everything to 1 A.U.

The signal ratios were cloud-screened with the PFR AOD cloud-screening algorithm (Kazadzis et al. 2018a) and visually filtered visually ffor outliers and days with erroneous measurements. Due to a diurnal variation of the signal ratios, we only used only data from between 9 - 13 UTC. We also excluded all days with fewer than 20 measurements in this day time interval and calculated a point point-to-point calibration for the restremaining data. We checked whether the two standard deviations (σ) of all points during each day fell below or were equivalent to 0.5% of the daily median calibration. If the 2σ were above 0.5% of the daily calibration, we repeatedly removed all points outside the 2σ range until the day satisfiesd this criterion. In case,If the remaining points of that day fell below 20 during this procedure, the day was rejectedWe removed all point calibrations outside 2 two standard deviations of the points during each day in a loop until 2 two standard deviations fall fell below or were equivalent to equal 0to 0.5% of the daily median calibration. If the remaining points wereare below 20, the day is was rejected. Finally, we further examined the point calibrations and their corresponding AOD further to reject any remaining days with erroneous calibrations. From the quality assured datasets, we calculated the POM daily median calibration and their monthly average (since ESR calculates monthly the calibration with ILP on a monthly basis).

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

To calculate the AOD, we used the following procedure (used by ESR): For the first month of each campaign, we used the monthly calibration constant for all measurements of the month. For the rest of the months, we assumed that the monthly calibrations correspond to the last day of each month at 12:00 UTC. For measurements between two monthly calibrations, we used linear interpolation to calculate the calibration at the time of the measurement. The interpolation is only based only on these two consecutive monthly calibrationsWe assumed that the monthly calibrations corresponded to the last day of each month at 12:00 UTC. For measurements between 2 two monthly calibrations, we used linear interpolation to calculate the calibration at the time of the measurement. For the first month of each campaign, we used the monthly calibration constant for all measurements of the month. We only used only 2two wavelengths, (500 and 870 nm), as they are directly comparable between the instruments. The actual wavelength of each instrument may vary. The first channel has the same nominal value for both instruments (500 nm) and the difference of the actual central wavelengths may vary by less than 1 - 2 nm. For the second channel the nominal wavelength of the PFR is 862 nm, while for the POM it is 870 nm. However, theDespite the difference of 8 nm in wavelength, Rayleigh and Mie scattering are weaker for at longer wavelengths so the effect of approximately 8 nmthe difference is less significant at in this spectral region.

2.3 Intercomparison

2.3.1 Measurement intercomparison

In order to assess the effect of calibration differences on AOD, we compare the AOD of POMs retrieved from different calibrations at 500 nm and 870 nm. There are two AOD datasets for each POM: the original AOD provided by ESR and the AOD calculated from the calibration transfer. The twoBoth sets of monthly calibrations used and their differences are present shown in the supplement table, S1. These 2 two AOD datasets also differ also as the algorithms to calculate on AAOD calculation algorithmwere different (Kazadzis et al., 2018a). The ESR algorithm provides ascalculates AOD, at for a given moment, based on the average of three consecutive measurements in one minute. In the calibration transfer-red based dataset, we use the AOD from the raw signals corresponding to individual measurements. AlsoIn addition, the second dataset has no correction for the nitrogen dioxide (NO_2), while SKYNET takes NO_2 into account. Finally, there are differences regarding the pressure and ozone column values. We screened the data for clouds according to the GAW-PFR algorithm. The reference AOD in all cases is the PFR AOD.

We added the co-located CIMEL instruments in the comparison as a third independent instrument taking advantage of the long-term experience on of meaurementsthe observation of AOD differences between AERONET and GAW-PFR (Kazadzis et al., 2018a, ; Cuevas et al., 2019, ; Karanikolas et al., 2022). The CIMEL data were cloud- screened by the AERONET algorithm, and we further screened them according to GAW-PFR algorithm.

AsWe use as indicators of the AOD differences, we use the median difference, the standard deviation of the differences, and their 5th and 95th percentiles. According to the World Meteorological Organization (WMO), instruments are considered traceable when at least 95% of the AOD differences are within specific limits (WMO/GAW, 2005) given by the following Eq. 7:

$$\lim = \pm(0.005 \pm 0.01/m) \quad (7)$$

where *m* is the air mass coefficient. Therefore, another indicator we used for the comparison is the percentage of data within the WMO limits.

2.4 Investigation on of potential ILP error sources

As the findings presented in Campanelli et al. (, 2023) and section 3.1 of the present study showed systematically negative differences between the ILP calibration and PFR- based calibration transfers that are always larger in Rome, we investigate several potential causes. Initially, we explore whether the aerosol properties between the twoboth locations show any systematic difference in terms of value and variability. We also assess the sensitivity of the ILP method to the pre-assigned values of six input parameters: solid view angle (SVA), surface pressure (P), total ozone column (TOC), surface albedo (SA), and the real and the imaginary part of the aerosol refractive index (RRI and IRI). Finally, we investigate whether the AOD, sc-AOD and SSA retrieved from the inversion modelling can provide evidence that may lead to an explanation of the observed differences. In the sections below, we describe the methodology of the three aforementioned parts of thethese investigations.

2.4.1 Aerosol properties

There are tThree parameters are discussed which we included in this section. AOD, SSA and the Angström Exponent (AE). According to Nakajima et al. (, 2020) the level of AOD affects the ILP performance. Also, the ILP method uses a pre-assigned refractive index value and assumes a stable SSA (which is connected with IRI) during the half- day the ILP is performed (Eq. 4). Therefore, the SSA value and variability may affect the calibration. Due to the above, we assess whether there is an association of the levels or the variability of AOD and SSA with the differences between ILP and the calibration transfer-based calibrations. For the AOD, we used the PFR dataset. For the SSA, the AERONET level 1.5 retrievals, due to lack of data availability of the quality assured level 2.0. Because of the still limited SSA dataset and the larger uncertainty (compared to level 2.0), we also added used the AE from the PFR in the investigation. AE is related to the size of aerosols, and a. A change to in AE reflects a change to in aerosol composition that may affect IRI and SSA as well. For the AOD and AE, we used only data corresponding to the half- days used for ILP calibrations. In aAdditionnally, we removed all points corresponding to AOD ≥ 0.4 at 500 nm and air masses ≥ 3, according to the screening criteria of the ILP method. For the SSA, we used all data during the campaign months of the campaigns except values < 0.1those corresponding to AOD at 440 nm, <0.1 and a very small number of outliers. Since ESR provides monthly calibrations, we used the monthly median values as indicator of the AOD, SSA and AE average levels. Each monthly median is the median of the daily medians. As indicators of the variability during the ILP method, we use the discrepancies between the monthly medians of the daily 5th, 20th, 80th and 95th percentiles.

2.4.2 Sensitivity of ILP on with respect to input parameters

As the ILP calibration requires the instrument solid view angle (SVA), the surface pressure (P), the total ozone column (TOC), the surface albedo (SA), the real and the imaginary part of aerosol refractive index (RRI and IRI) as inputs, we examine to what extent they affect the ILP calibration.

The Skyrad 4.2 code use pPre-selected user values by the user for each of the last 5 parameters (P, TOC, SA, RRI and IRI) can be entered into the Skyrad 4.2 code. Surface pressure depending on the altitude of the station is provided bcalculated using y the Eq. 8:

$$P = P_0 e^{-0.0001184h} \quad (8)$$

where P is the pressure in atm, P_0=1 atm and *h* the altitude in meters. TOC is fixed to 300 DU for both Davos and Rome. SA is fixed to 0.1 (forat non-polar regions such as like the ones those in the present study), RRI is set to 1.5 and IRI to 0.005 for all wavelengths (340, 400, 500, 675, 870 and 1020 nm).

The SVA is derived with the disk scan method, an on-site calibration procedure (Nakajima et al., 2020; Campanelli et al., 2023). To investigate the effect of these input files, we performed a set of ILP calibrations under different conditions in 3 three sub-studies. For this section, we only used only data from QUATRAM II as it is was the longest campaign.

In the first sub-study, we focus separately on each a-priori parameter of the ILP calibration. We keep aAll other parameters are left at their in their original values and change only the parameter under study . The goal is to recalculate the ILP calibrations for the local station conditions of the station. Therefore, for each parameter under study, we select a value based on observations in at the measurement site. Specifically, TOC and P are present in the PFR data. TOC is taken from the OMI overpass (aura_omi_l2ovp_omto3_v8.5 https://acd-ext.gsfc.nasa.gov/anonftp/toms/omi/data/overpass/) and P was measured by a Setra barometer (uncertainty of less than 10 mbar). The refractive index parts (RRI and IRI) are available from datasets of the AERONET almucantar scans datasets (only at 440, 675, 870 and 1020 nm). SA is also taken from the AERONET datasets in at the same wavelengths. Over land, this and over land originates from a Li-Ross bidirectional reflectance distribution function (BRDF) model (Lucht & Roujean 2000) based on MODIS (or Moderate Resolution Imaging Spectroradiometer) satellite observations (Sun et al., 2017). For the rest of the wavelengths (340, 400, 500 and 940 nm), we had to select values based on the existing wavelengths either by interpolation and extrapolation ((we used linear) (RRI, IRI) or by using a separate criterion (SA). The SA selection is based on its the observed SA value and the its spectral dependence of the SA in the IGBP library from the LibRadtran package (Emde et al., 2016). The SVA is provided by ESR.

For each parameter, we used three different values to calculate three different ILP calibration constants. We calculated one ILP calibration using the median (RRI, IRI) or the mean (TOC, P and SA) value during all the months of the three QUATRAM campaigns. The other two calibrations correspond to values equivalent to the value one standard deviation above and below each average. For the SVA, we used the values provided by ESR for the first ILP calibration. The other two values are based on the maximum difference observed between ESR SVA and other SVA calibration methods for POMs presented in Campanelli et al. (, 2023). In the supplement (sections Sections 3 - 5 and tables Tables S4 - S6), we present all the values used for the six input parameters.

In the second sub-study, we alter the values of all parameters simultaneously except SVA (we used the value provided by ESR). Again, tThe goal is again tto adapt the input parameters to the site conditions. We calculated the ILP calibration in for two separate cases:

a) Average case: 1 one calibration per month using the monthly average values used in the first sub-study for all five parameters under testing (RRI, IRI, P, TOC and SA).

b) 'Selected' case: 1 one calibration per month. Here we selected one of the three values used in the first sub-study for the same five parameters. The selected values are those of the three that lead to a larger calibration constant. We picked only 1 one month per location for this case. The values of the input parameters used for this second sub-study are shown in the supplement section Section 6.

In the third sub-study, we tested the IRI, SA and SVA for a more extensive number of values (seven fixed values regardless of the location) to assess the behaviour of the calibration. For IRI and SA, the selection includes is based on the three values of the first sub-study, the 5th - 95th percentiles of the observations and the minimum/maximum values appeared. We also added semi-arbitrary values between the observed and two extreme values (one very small and one very large) to test the performance of the method at over a wider range of inputs. For the SVA, we use values based on the differences between the different SVA calibration procedures appearing in Campanelli et al. (, 2023). The actual values for each parameter are in the supplement, section Section 10, table Table S11.

2.4.3 Investigation on of the aerosol optical depth retrievals from sky radiance

Since the ILP method is performed using a linear fit of the logarithm of DSI with respect to the product of air mass coefficient and scattering aerosol optical depth (sc-AOD) (Eq. 4), errors on the retrieval of the sc-AOD will transfer errors to the calibration. Since there is no reference dataset available for the sc-AOD, we tried to indirectly investigate potential errors using any available data.

The Skyrad code retrieves both sc-AOD and SSA through inversion modelling and calculates the corresponding AOD as additional information. Therefore, initially we compare that the AOD dataset with the PFR AOD for potential differences. However, systematic underestimation or overestimation on of both the sc-AOD and SSA retrievals can result in opposite opposing errors to the corresponding AOD that cancel each other. Due to the limitations of the AERONET SSA dataset (lack of level 2.0 data and limited retrievals per day), we cannot evaluate the SSA retrieved by Skyrad 4.2 with confidence. Also, part of the SSA difference between the AERONET product and the output of Skyrad code for ILP calibration may be attributed to the fixed refractive index and the different scattering angles in the almucantar geometry used for the sky radiance measurements (ILP uses only forward scattering having a maximum angle of 30° degrees).

Another indirect method to investigate the effect of the sc-AOD retrievals on the calibration performance is to use a different inversion model to retrieve sc-AOD and re-calibrate the instrument with ILP. We For this purpose, wetherefore used the inversion model, Skyrad pack MRI version 2 (Kudo et al, 2021). MRI allows the modelling of non-spherical particles in contrast to Skyrad pack 4.2 retrievals. It also introduced introduces stability constraints at on the edges of the size distribution as well as other to be stable and different smoothness constraints (see Kudo et al., 2021 provide for a detailed description). As mentioned in the same study, the MRI method is more accurate at high AOD. Under low AOD conditions in Davos, a noticeable portion of data showed large retrieval errors and unrealistic sc-AOD/AOD values. However, in both locations there were was sufficient data at both locations to recalculate the ILP calibration, and hence it was and we applied it to the data of the the QUATRAM II data.

We also investigated whether the variability of the SSA corresponding to the Skyrad 4.2 and MRI retrieval shows any association with the calibration differences.

All retrieved AOD, sc-AOD and SSA data are screened according to the ILP criteria: keeping only data corresponding to AOD at 500 nm < 0.4 and air mass < 3.

3 IntercomparisonResults

3.1 Methodology

In order to assess the effect of calibration differences on AOD, we compare the AOD of POMs retrieved from different calibrations at 500 nm and 870 nm. There are two AOD datasets for each POM: the original AOD provided by ESR and the AOD calculated from the calibration transfer. Both sets of monthly calibrations used and their differences are shown in the supplement table, S1. These two AOD datasets also differ as the algorithms to calculate AOD were different (Kazadzis et al., 2018a). The ESR algorithm calculates AOD, at a given moment, based on the average of three consecutive measurements in one minute. In the calibration transfer-based dataset, we use the AOD from the raw signals corresponding to individual measurements. In addition, the second dataset has no correction for nitrogen dioxide (NO_2), while SKYNET takes NO_2 into account. Finally, there are differences regarding the pressure and ozone column values. We screened the data for clouds according to the GAW-PFR algorithm. The reference AOD in all cases is the PFR AOD.

We added the co-located CIMEL instruments in the comparison as a third independent instrument taking advantage of the long-term experience of measurements between AERONET and GAW-PFR (Kazadzis et al., 2018a; Cuevas et al., 2019; Karanikolas et al., 2022). The CIMEL data were cloud-screened by the AERONET algorithm (Smirnov et al., 2000; Giles et al., 2019), and wethen further screened them according to the GAW-PFR algorithm (Kazadzis et al., 2018b).

As indicators of the AOD differences, we use the median difference, the standard deviation of the differences, and their 5th and 95th percentiles. According to the World Meteorological Organization (WMO), instruments are considered traceable when at least 95% of the AOD differences are within specific limits (WMO/GAW, 2005) given by Eq. 7:

$$\lim = \pm(0.005 \pm 0.01/m) \quad (7)$$

where *m* is the air mass coefficient. Therefore, another indicator we used for the comparison is the percentage of data within the WMO limits is another indicator we used for the comparison.

3.2 Results

In this section we present the main findings of the study. First, we show the AOD differences between the CIMEL or POM using different calibrations and the reference PFR. Then we present the stability and uncertainties of the used calibrations. ~~Finally, we attempt to investigate the causes of the observed differences through the methodology described in section 2.4.~~

3.~~1~~2.1 AOD intercomparison

There ~~are~~ were three campaigns per location and we present the AOD differences between the PFR and POMs or CIMEL. In Fig. 1, we show the median AOD differences and standard deviation (box size), as well as ~~and~~ the 5th and 95th percentiles of the differences (error bars). A noticeable feature ~~Something evident~~ is that the ESR AOD calculated with the ILP method is systematically lower than the PFR AOD. In Davos, the median differences are between -0.006 and -0.01 at 500 nm and 0.000 to -0.005 at 870 nm. In Rome, the median differences range from approximately -0.014 to -0.034 at 500 nm with the vast majority of differences ~~below~~ <-0.01. At 870 nm QUATRAM I in Rome shows a median difference of -0.005 and the other ~~cases~~ campaigns ~~below~~ <-0.01. For QUATRAM II in Rome, which was the longest campaign and the one with the largest differences ~~at~~ of the POM master (POMCNR), we included a second POM (~~POM11~~POMSPZ). This ~~, which s~~shows a performance similar to the POM master (POMCNR*) ~~of~~ during QUATRAM III in Rome.

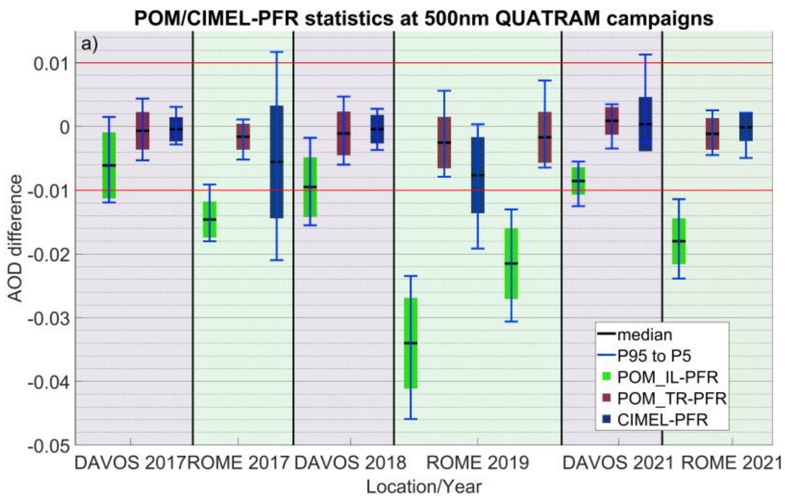
When using a PFR calibration transfer ~~from PFR t~~to recalculate the AOD for POMs, the absolute median differences are ~~below~~ <0.005 for all cases. ~~The median difference remains mostly negative Most of the times in the case of the calibration transfer the median difference remains negative, but there are exceptions.~~ The CIMEL-PFR comparison shows similar results with all median AOD differences below 0.01. ~~Also~~In addition, the majority of the 5th - 95th percentiles for either CIMEL-PFR or POM-PFR using the calibration transfer are within 0.01.

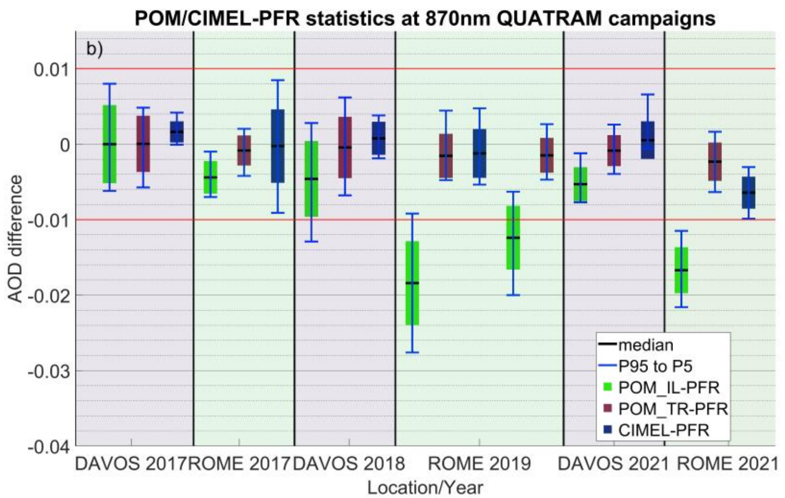
Regarding the WMO traceability criteria, the data within the, WMO limits for POM AOD with an ILP calibration are below 95% for all cases at 500 nm ~~and~~ as well as for QUATRAM II and III in Rome at 870 nm (~~table~~ Table 2). However, there is a large deviation between ~~the two~~both locations: ~~where~~ while at 500 nm the percentage in Davos is above 60%, it is ~~while in Rome belo~~below 4% in Rome. Using the calibration transfer to calculate POM AOD, ~~in all cases there are more than~~> 98% of data are within the WMO limits (~~table~~ Table 2). The CIMEL-PFR comparison (~~table~~ Table 3) also shows percentages mainly above 98%. Exceptions are ~~the~~ QUATRAM I and II ~~phases~~ in Rome at 500 nm and QUATRAM I in Rome ~~phase~~ at 870 nm. All ~~cases~~ CIMEL-PFR comparisons show ~~have~~ at least ~60% of differences within the WMO limits.

Recalculating the AOD with the same post-processing algorithm and for the same instrument (once for each POM) for b~~the two~~oth POM calibrations (ILP and calibration transfer), we can more clearly observe ~~purely~~ the effect of just the calibration on AOD. In ~~that~~ this case, the median AOD difference ~~in that case~~ is similar to the difference between the original POM and PFR datasets shown by the green boxes in Fig. 1~~ green boxes~~. The results of the comparison, showing the calibration effect along with the ~~'original'~~ "original" differences, are in the supplement in Fig. S1 (~~section~~ Section 1). The median AOD differences attributed to the calibration, deviate from the ~~'"original'~~ original" AOD differences by ~~less than~~< 0.003, except for ~~aside from~~ three cases. It is approximately 0.005 for ~~the~~ QUATRAM III in Rome ~~phase~~ at 500 nm and in Davos ~~phase~~ at 870 nm. A larger value of ~~It is almost~~ 0.01 was observed for ~~the~~ QUATRAM II in Rome ~~phase~~ at 500 nm for ~~of o~~only one of the POMs (POM_CNR). These deviations are not systematically larger or smaller than the ~~"'original'~~ "original" at 870 nm, but they are smaller for most campaigns at 500 nm.

The variability of AOD differences ~~in the case of~~for the comparison between ~~the two~~both recalculated POM AOD datasets (which just show ~~purely~~ the calibration effect), is a result of the dependence of the calibration effect on the air mass. Therefore, it depends on the magnitude of the calibration difference, its month-to-month variability and the air mass distribution present ~~on~~ in the data.

These results suggest that the overall contribution of the post-processing algorithm and instrument differences between the networks, result ~~to~~in AOD differences that are within the PFR AOD retrieval uncertainty~~These results suggest that the post post-processing algorithm and instrument technical differences between the networks are a source of only random AOD differences within the retrieval uncertainty~~. ~~In the case of~~For ESR, the calibration method ~~difference~~ dominates the overall AOD difference.


POM/CIMEL-PFR statistics at 500nm QUATRAM campaigns
a)
AOD difference
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
median
P95 to P5
POM_IL-PFR
POM_TR-PFR
CIMEL-PFR
DAVOS 2017
ROME 2017
DAVOS 2018
ROME 2019
DAVOS 2021
ROME 2021
Location/Year


POM/CIMEL-PFR statistics at 870nm QUATRAM campaigns
b)
AOD difference
0.01
0
-0.01
-0.02
-0.03
-0.04
median
P95 to P5
POM_IL-PFR
POM_TR-PFR
CIMEL-PFR
DAVOS 2017
ROME 2017
DAVOS 2018
ROME 2019
DAVOS 2021
ROME 2021
Location/Year

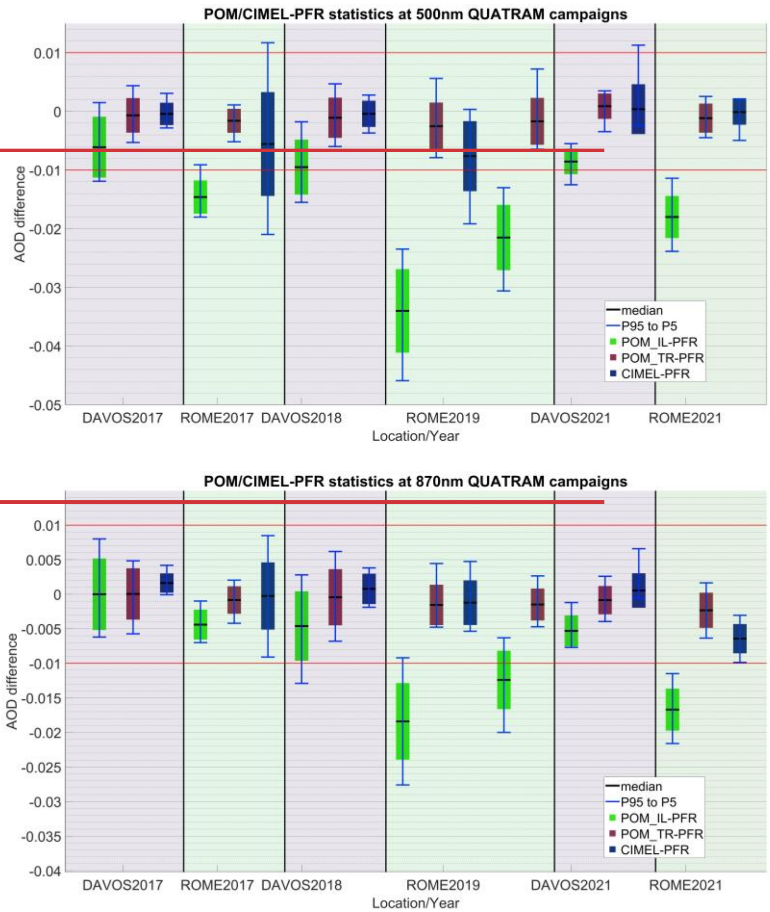


Figure 1: Box plot of the statistics of AOD differences~~' statistics~~ for all instrument comparisons ~~during~~ for both ~~phases~~ locations of the ~~3~~ three QUATRAM campaigns. ~~Top panel~~a): 500 nm. ~~Bottom panel~~b): 870 nm. The black line is the median difference, the size of the boxes denotes the distance between the median and the standard deviation, while the error bars show the 5th and the 95th percentile of the AOD differences. In all cases the PFR AOD is the reference instrument. The green boxes correspond to the differences between the original AOD from POMs and the reference. The red boxes correspond the POM AOD calculated with the calibration retrieved with transfer from the PFR. The blue boxes correspond to the differences between CIMEL and PFR. For the Rome 2019 campaign, we compare ~~2~~ two different POMs with the same PFR (left POM~~_~~CNR and right ~~POM11~~POMSPZ). ~~Top: 500 nm. Bottom: 870 nm.~~

Table 2: The percentage of AOD differences within WMO limits for the comparison between PFRs and POMs. IL refers to the original POM AOD retrieved using the ILP calibration method and TR to the calibration transfer-based AOD.

Location	**Instrument**	**Year**	**Number of measurements**	**WMO limits % IL**		**WMO limits % TR**	
				500 nm	**870 nm**	**500 nm**	**870 nm**
DAVOS I	POMVDV	2017	1929	84.3~~4~~	95.2~~3~~	99.7~~4~~	98.~~65~~7
DAVOS II	POMCNR	2018	6604	63.5~~1~~	89.1~~3~~	99.0~~3~~	98.2~~1~~
DAVOS III	POMCNR*	2021	1516	72.1	99.~~47~~5	100.0~~0~~	100.0~~0~~
ROME I	POMVDV	2017	507	3.~~16~~2	99.0~~1~~	98.6~~2~~	100.0~~0~~
ROME II	POMCNR	2019	3903	0.0~~0~~	11.~~48~~5	~~99~~100.~~95~~0	~~99~~100.~~95~~0
ROME II	~~POM11~~POMSPZ	2019	6079	2.~~66~~7	44.~~56~~6	99.1~~0~~	100.0~~0~~
ROME III	POMCNR*	2021	904/908	~~2~~3.~~99~~0	1.3~~2~~	100.0~~0~~	100.0~~0~~

Table 3: The percentage of AOD differences within WMO limits for the comparison between PFRs and CIMELs.

Location	**Instrument**	**Year**	**Number of measurements**	**WMO limits %**	
				500 nm	**870 nm**
DAVOS I	CIMEL#354	2017	614	99.8~~4~~	99.8~~4~~
DAVOS II	CIMEL#354	2018	1127	99.~~38~~4	99.~~47~~5
DAVOS III	CIMEL#916	2021	271	100.0~~0~~	100.0~~0~~
ROME I	CIMEL#646	2017/2018	117	59.8~~3~~	90.6~~0~~
ROME II	CIMEL#43	2019	2278	75.2~~0~~	100.~~0~~0
ROME III	CIMEL#1270	2021	243/253	100.0~~0~~	98.8~~1~~

3.~~1~~2.2 Calibration stability and uncertainties

In the previous section, we showed that the major source of AOD differences was due to differences in the PFR and POM calibration methods~~between the PFRs and POMs is the calibration method difference~~. The calibration differences between the ILP method and the PFR-based transfer can be found in the supplement ~~table~~ Table S1 (~~section~~ Section 1). The values in the supplement show some minor differences compared to Campanelli et al. (~~,~~ 2023) for some months mainly due differences in the selected days. The difference is larger for August 2018 in Davos. During this month we observed an abrupt shift of daily calibrations early in the month. ~~So~~Hence, we removed the days before the shift ~~since~~as the monthly calibration is attributed to the end of the month when retrieving the AOD~~The values in the supplement show some minor differences~~

compared to Campanelli et al. (, 2023) for some months, mainly due to differences in the day selection that are larger for August 2018 in Davos (where we observed an abrupt calibration shift during the month and removed the days before the shift as the monthly calibration is attributed to the end of the month when retrieving AOD). In this section, we present discuss the stability and the uncertainties of the different calibrations.

The ILP calibrations show either positive or negative fluctuations for consecutive months at the same location lying in the 0.17-2.3% range with a median absolute value of 0.55% and a standard deviation of 0.87%. These calibration fluctuations can be either attributed either to changes in the instruments' response or the random component –of the ILP method uncertaintyThe ILP calibrations show either positive and or negative fluctuations for consecutive months atin the same location, lying between in the 0.17 - 2.3% range with a median absolute value of 0.55% and a standard deviation of 0.87%. It can be attributed both to changes in the instruments and the random uncertainty of the ILP method. The coefficient of variation (CV%) of the daily ILP calibrations per month (Campanelli et al., 2023 Table 2a) is an estimate of the ILP monthly calibration uncertainties. CV% is the percentage of the standard deviation of daily calibration constants during the month divided by the monthly calibration constant. The CV% for the ILP calibrations used in this study lies in the 0.18%-2.87% range at 500 and 870 nm.An estimation of the uncertainty magnitude is evident in the coefficient of variation (CV%) of the daily ILP calibrations per month (Campanelli et al., 2023; preprint, table Table 2a), which lies in the are between 0.18% - 2.87% range at 500 and 870 nm.

The PFR calibration differences between consecutive calibrations are in the between 0.00 - 0.45% range at 500 and 870 nm (supplement table Table S3). All , with all calibrations have having an uncertainty below 0.4% (supplement table Table S2). The PFR-–based calibration transfers of POMs show fluctuations for consecutive months in at the same locations in the between 0.00 - 1.72% range with a median absolute value of 0.19% and a standard deviation of 0.56%. The uncertainties of the calibration transfers calculated as the combination of the PFR calibration uncertainty σ_{PFR} and the standard deviation of the daily calibrations σ_d are calculated as:

$$\sigma_{TR} = \sqrt{\sigma_{PFR}^2 + \sigma_d^2}$$

(9)

Applying Eq. 9 shows that theThe calibration transfer uncertainties are in the between 0.27% - 0.8% range (supplement table Table S2).

The month-to-month variability of the ILP method and transfer-based calibrations do not coincide. This is reflected in the month-to-month variability of the calibration differences between the twoboth methods, which is in the 0.01%-1.93% range. Their median absolute value is 0.55% and their standard deviation 0.96%"The fluctuations of ILP and transfer-based calibrations do not coincide, which is reflected in the month-to-month fluctuations of their difference, in the being 0.01% - 1.93% range with a median absolute value of 0.55% and standard deviation 0.96%.

However, not all calibration fluctuations can be explained by the presented uncertainties in the present section. A particularly interesting case is the calibration change from July to August 2019 in Rome for POMCNR at 870 nm. The CV% of the ILP calibrations of these two months is below 0.5% (Campanelli et al., 2023), while their calibration difference is

1.3%. The calibration transfers from the PFR for the same months differ by only ~~by~~ 0.2% providing no evidence of changes in the instrument response. The same months show an ILP calibration change above 2% for ~~POM11~~POMSPZ, with the calibration transfers differing by 0.3%. At 500 nm for the same months, the ILP differences are above 1%, while the calibration transfer differences are 0%. Therefore, the ILP differences between these two months are attributable to the overall uncertainty of the ILP calibration.

4 Investigation of potential ILP error sources

As the findings presented in Campanelli et al. (2023) showed systematically negative differences between the ILP calibration and PFR-based calibration transfers that are always larger in Rome compared to Davos, we investigate several potential causes. Initially, we explore whether the aerosol properties between both locations show any systematic difference in terms of value and variability. We also assess the sensitivity of the ILP method to the pre-assigned values of six input parameters: SVA, P, TOC, SA, and the real and the imaginary part of the aerosol refractive index (RRI and IRI). Finally, we investigate whether the AOD, sc-AOD and SSA retrieved from the inversion modelling can provide evidence that may lead to an explanation of the observed differences.

4.1 Aerosol properties

4.1.1 Methodology

Three parameters are discussed in this section, namely AOD, SSA and the Angström Exponent (AE). According to Nakajima et al. (2020) the level of AOD affects the ILP performance. Also, the ILP method uses a pre-assigned refractive index value and assumes a stable SSA (which is connected with IRI) during the half-day the ILP is performed (Eq. 4). Therefore, the SSA value and variability may affect the calibration. Due to the above, we assess whether there is an association of the levels or the variability of AOD and SSA with the differences between the ILP method and the calibration transfer-based calibrations. For the AOD, we used the PFR dataset. For the SSA we used the AERONET level 1.5 retrievals, due to lack of data availability of the quality assured level 2.0. Because of the limited SSA dataset and the larger uncertainty (compared to level 2.0), we also used the AE from the PFR in the investigation. AE is related to the size of aerosols, and a change in AE reflects a change in aerosol composition that may affect IRI and SSA. For the AOD and AE, we used only data corresponding to the half-days used for ILP calibrations. In addition, we removed all points corresponding to AOD ≥ 0.4 at 500 nm and air masses ≥ 3, according to the screening criteria of the ILP method. For the SSA, we used all data during the campaign months except values < 0.1 corresponding to AOD at 440 nm, and a very small number of outliers. Since ESR provides monthly calibrations, we used the monthly median values as indicators of the AOD, SSA and AE average levels. Each monthly median is the median of the daily medians. As indicators of the variability during the ILP method, we use the discrepancies between the monthly medians of the daily 5th, 20th, 80th and 95th percentiles.

34.1.2 Investigation on of calibration differencesResults

As shown in section 3.1.1, the ESR dataset shows a systematic AOD underestimation compared to GAW-PFR and AERONET due to an underestimation in the calibration from the ILP method. However, this calibration difference varies significantly between the twbotho locations and from month month-to to-month. Using the methods described in section 2.4, we attempted to explain why this underestimation happens and why it is systematically larger for Rome.

3.2.1 Aerosol properties

Here we investigate whether there is any systematic difference between Davos and Rome on with respect to AOD, SSA and AE values or variability that could potentially be associated with the larger calibration differences in Rome for all months. We use AOD and AE from the PFR data during the half/full days of the ILP calibrations. and SSA is from the AERONET data during the QUATRAM campaigns. We used monthly medians median statistics as the average level and monthly medians of the daily percentiles (5th, 20th, 80th and 95th) as a variability indicator as described in section 24.41.1.

34.21.12.1 Aerosol Optical Depth

Here we presentFig. 2 shows the PFR AOD values for all months of the campaigns in at both locations.. The results are visible shown in Fig. 2, where the green boxes correspond to 500 nm and the red to 862 nm. For most months it is evident that the AOD is higher and more variable in Rome, but there are exceptions, such as for the like QUATRAM I (DAV17/ROM17) campaign. Also, we can see that the highest AOD corresponds to QUATRAM III in Rome (ROM21) (Fig. 2), while the largest calibration and AOD differences between PFR and POM were in QUATRAM II (ROM19) (Fig. 1, Table S1). Both AOD values and AOD variability vary within at the same location and between the twoboth from month month-to to-month, showing no consistency between AOD (Fig. 1) and calibration differences (supplement, table Table S1).

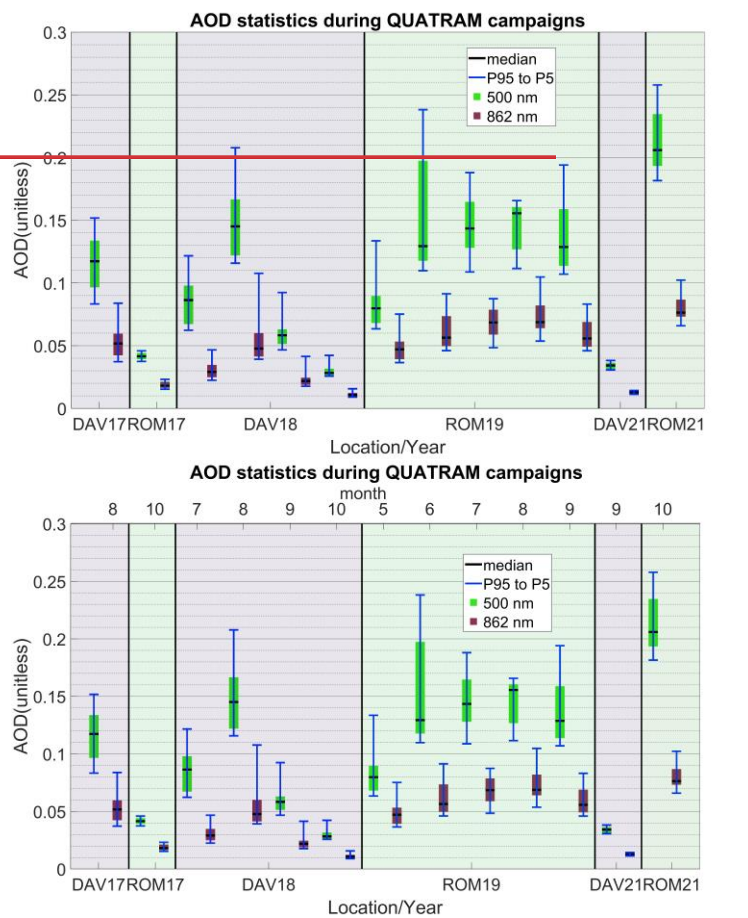


Figure 2: ~~The~~ PFR AOD statistics for all months for~~of~~ all campaigns. The green boxes correspond to 500 nm and the red to 862 nm with each pair being one month. ~~The extent of the each box shows the median of the 20th and 80th percentiles per day and the error bars the median of the 5th and 95th percentiles per day.~~ Each box ~~is~~ represents ~~1~~ one month of the campaign.

34.21.12.2 Single Scattering Albedo

The ILP method assumes a constant SSA as the inverse slope of the linear fit (section 2.2.1) and uses an a-priori refractive index (selected by the operator). These assumptions potentially reduce the accuracy of the method. Here we present the SSA values provided by AERONET and their variability during the campaign months of the campaigns (Fig. 3) at 440 nm (green) and 870 nm (red).ILP assumes a constant SSA as the inverse slope the linear fit (section 2.2.1) and the refractive index pre-assigned to specific value , which potentially reduces the accuracy of the method. Here we present the AERONET SSA values and variability between the months of the campaigns (Fig. 3) at 440 nm (green) and 870 nm (red). For the Davos 2018 campaign, there are three months instead of four as there was a lack of data during the first month (July 2018,) since because the campaign started towards the end of the month. In gGenerally, neither isno a systematic difference between the twoboth locations is is evident nor is there an association between the calibration and AOD differences, even for the same location. In Rome, the largest SSA variability corresponds to QUATRAM I (ROM17) (Fig. 3), in whichwhere we observed the smallest calibration and AOD differences during the Rome phasescampaigns (Fig. 1, Table S1). Similarly, in Davos the largest variability in is during QUATRAM III (DAV21) in Davos, which also exceeds the Rome SSA variability. However, we did not observe larger differences between ILP and the calibration transfer in Davos during QUATRAM III (DAV21) compared to QUATRAM II (DAV18). In terms of median SSA, depending on the month, either Rome or Davos may have larger SSA. The fluctuations of SSA do not seem to significantly affect the calibration differences. However, we acknowledge that the limitations of the SSA dataset (section 2.4.14.1.1) limit the confidence in theof conclusions.

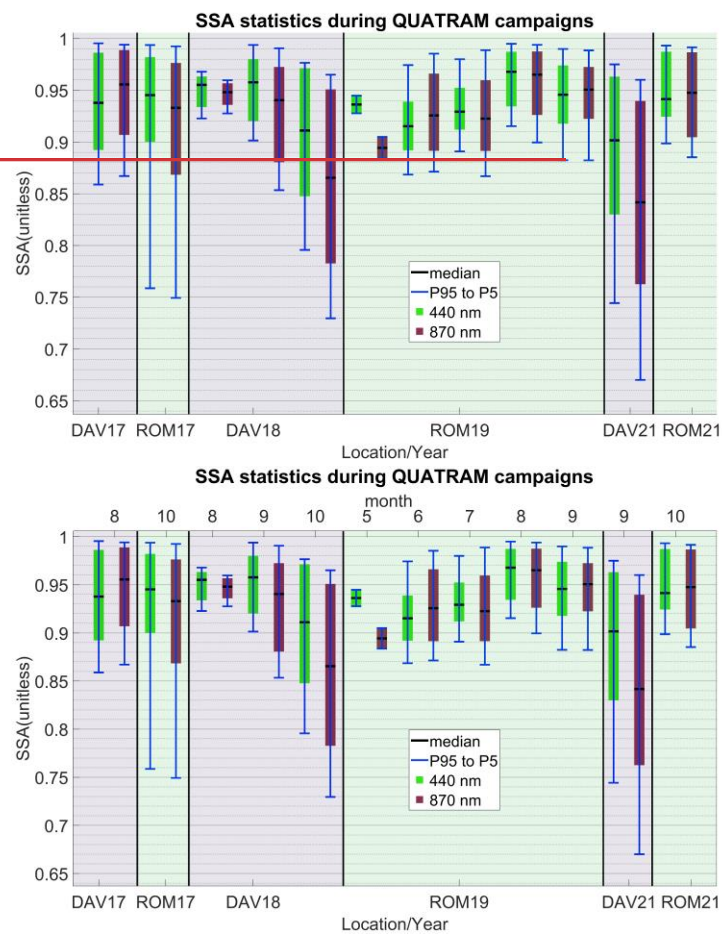
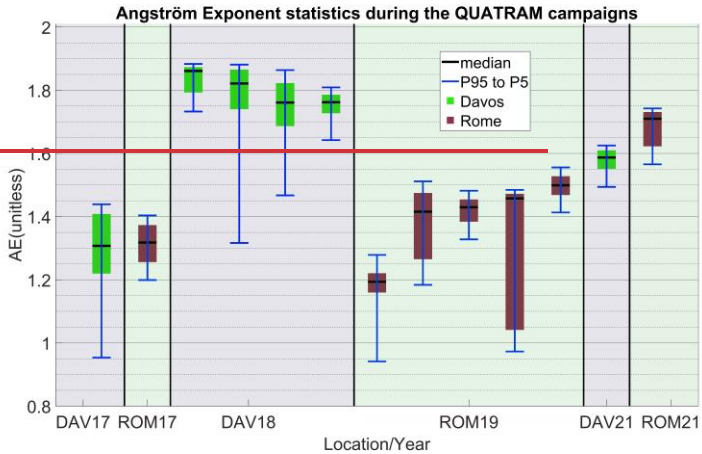


Figure 3: The AERONET SSA statistics for all months and for of all campaigns. The green boxes correspond to 440 and the red to 870 nm with each pair being one month. The extent of the box shows the median of the 20th and 80th percentiles per day and the error bars the median of the 5th and 95th percentiles per day. Each box is represents 1 one month of the campaign. In QUATRAM II (DAV18), the first month of the campaign (July) is missing due to lack of data.

~~3~~4.~~2~~1.~~1~~2.3 Angström Exponent

Due to the limitations of the SSA dataset (section ~~2.4.1~~4.1.1), we ~~added~~ included a comparison of the AE medians and variability during the campaigns as an additional indicator of aerosol composition. ~~The results are in Fig. 4 with green corresponding to Davos and red to Rome.~~ During QUATRAM I (DAV17/ROM17) ~~the two~~both locations have ~~very~~ similar median AE, but Davos shows the largest variability. During QUATRAM II (DAV18/ROM19) the AE in Davos is the largest, while the variability varies significantly between the months. Similarly, during QUATRAM II in Rome, AE is lower and each variability largely ~~depending~~ depends on the month. Finally, during QUATRAM III (DAV21/ROM21), Rome shows the largest AE and variability. Again, there is ~~no~~ neither a systematic difference between ~~the two~~both locations nor an association of calibration differences and AE ~~within~~ at the same location.



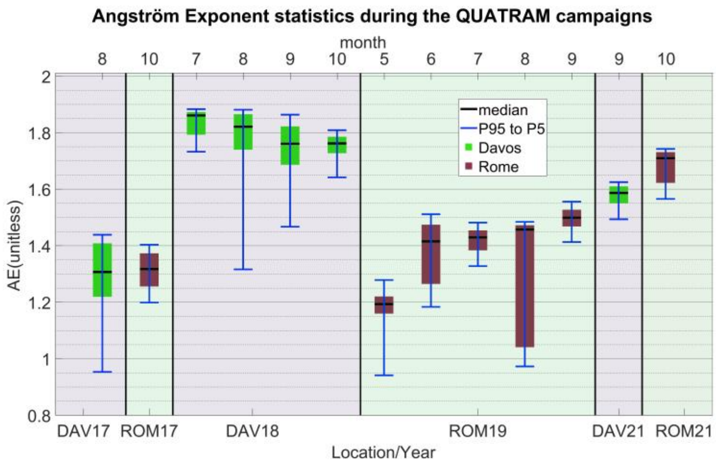


Figure 4: The PFR AE statistics for all months and for~~of~~ all campaigns. The green boxes correspond to Davos and the red to Rome. ~~The extent of the box shows the median of the 20th and 80th percentiles per day and the error bars the median of the 5th and 95th percentiles per day.~~ Each box ~~is~~ represents ~~1~~ one month of the campaign.

4.2 Sensitivity of the ILP method with respect to input parameters

As the ILP calibration requires the instrument SVA, P, TOC, the SA, the real and the imaginary part of aerosol refractive index (RRI and IRI) as inputs, we examine to what extent they affect the ILP calibration.

Pre-selected user-values for each of the last five parameters (P, TOC, SA, RRI and IRI) can be entered into the Skyrad 4.2 code. Surface pressure depends on the altitude of the station and is calculated using Eq. 8:

$$P = P_0 e^{-0.0001184h} \qquad (8)$$

where *P* is the pressure in atm, P_0=1 atm and *h* the altitude in meters. TOC is fixed to 300 DU for both Davos and Rome. SA is fixed to 0.1 (for non-polar regions such as those in the present study), RRI is set to 1.5 and IRI to 0.005 for all wavelengths (340, 400, 500, 675, 870 and 1020 nm).

SVA is derived with the disk scan method, an on-site calibration procedure (Nakajima et al., 2020; Campanelli et al., 2023). To investigate the effect of the aforementioned input parameters, we performed a set of ILP calibrations under different conditions in three sub-studies. For these sub-studies, we only used data from QUATRAM II as it was the longest campaign.

4.2.1 Sub-study 1: ILP Test based on local observations: one variable parameter per case

4.2.1.1 Sub-study 1: Methodology

In the first sub-study, we focus separately on each a-priori parameter of the ILP calibration. All other parameters are left at their original values except for one that is variable The goal is to recalculate the ILP calibrations for the local station conditions. Therefore, for each parameter under study, we select a value based on observations at the measurement site. Specifically, TOC and P are present in the PFR data. The TOC used in the PFR algorithm ~~is~~ corresponds to the OMI satellite product (aura_omi_l2ovp_omto3_v8.5 https://acd-ext.gsfc.nasa.gov/anonftp/toms/omi/data/overpass/) and P was measured with a Setra barometer (uncertainty of less than 10 mbar). The refractive index values (RRI and IRI) are available from datasets of the AERONET almucantar scans (only at 440, 675, 870 and 1020 nm). SA is also taken from the AERONET datasets at the same wavelengths. Over land, this originates from a Li-Ross bidirectional reflectance distribution function (BRDF) model (Lucht & Roujean 2000) based on MODIS (or Moderate Resolution Imaging Spectroradiometer) satellite observations (Sun et al., 2017). For the rest of the wavelengths (340, 400 and 500 nm) we had to select values based on the existing wavelengths. ~~In the case of~~ For RRI and IRI we used linear interpolation and extrapolation to estimate their values at those three missing wavelengths. The SA selection at 340, 400 and 500 nm is based on its observed values and its spectral dependence in the IGBP library from the LibRadtran package (Emde et al., 2016). SVA is provided by ESR.

For each parameter, we used three different values to calculate three different ILP calibration constants. We calculated one ILP calibration using the median (RRI, IRI) or the mean (TOC, P and SA) value during all the months of the three QUATRAM campaigns. The other two calibrations correspond to values equivalent to one standard deviation above and below each average. For SVA, we used the values provided by ESR for the first ILP calibration. The other two values are based on the maximum difference observed between ESR SVA and other SVA calibration methods for POMs presented in

Formatted: Font: Italic

Formatted: Font: Italic

Campanelli et al. (2023). In the supplement (Sections 3 - 5 and Tables S4 - S6), we present all the values used for the six input parameters.

3.2.2 Sensitivity of ILP on input parameters

As the available aerosol conditions during the campaigns cannot be explained by the underestimation of ILP show no indication of an explanation to the ILP underestimation and the differences between locations, we attempted to investigate the causes through a sensitivity study of the ILP. The latter ILP uses six parameters as inputs: Real part of refractive index (RRI), Imaginary part of refractive index (IRI), Surface albedo (SA), Total Ozone Colum (TOC), Surface Pressure (P) and Solid View Angle (SVA). The first five are pre-selected and the last is provided by an in-situ calibration method. Therefore, there are discrepancies between the real atmospheric conditions under which the ILP is performed and the selected values.

34.2.21.1 2 ILP Test based on local observations: one variable parameter per case Sub-study 1: Results

Here we present results of the ILP calibration using different values for the input parameters of Skyrad 4.2. The selection is described in section 2.4.24.2.1.1.

The RRI average observations from AERONET were very closesimilar to the pre-assigned input of Skyrad pack 4.2 (1.5 for all wavelengths) and, while the standard deviation was small, . so Hence, we used the averageoriginal, minimum and maximum values (1.33, 1.5 and 1.6). The calibration difference due to this change in the RRI were was in the between 0.00 - 0.21% range.

For the surface pressure (P), we used the values 0.8, 0.83 and 0.85 atm for Davos, while and 0.97, 1 and 1.02 atm for Rome (the middle value is the one used originally values for ILP wereas 0.83 and 1). Most differences were below 0.05%. During one month at 870 nm, we obtained the a maximum difference of 0.2% (July 2019 in Rome where the maximum sensitivity at RRI was also presentoccurred).

For total ozone column (TOC), we used for both locations 260, 300 and 400 DU for both locations, which again resulted again in differences were below of up to 0.0543% except for July 2019 at 870 nm. The comparisons for RRI, P and TOC are available in more detail in the supplement (Tables S8 - S10).

Due to the small sensitivity at of these three parameters, we do not include a more detailed analysis. However on them, but, the comparisons are available in the supplement (sections and tables Tables S8 - S10). For the imaginary part of refractive index (IRI), surface albedo (SA) and solid view angleSVA, we observed cases of larger sensitivity.

In the Figs. 5 - 7,. we can see the calibration differences between the ILP runs and the calibration transfer from PFR for different conditions. The results correspond to the first sub-study described in section 2.4.2 where we study each parameter separately according to the observations of at each site. The results correspond to all months of QUATRAM II.

For IRI, SA and SVA, we show the ILP calibration differences in Figs. 5 – 7. For the majority of the cases, the calibration differences due to IRI are smaller than 0.5% (Fig. 5). For specific months (August 2018-Davos and July 2019-Rome), it isthey are 1% or higher. However, a calibration difference between ILP and the calibration transfer of 1% in Davos and 2.5% in Rome at 500 nm and above 1.5% in Rome at 870 nm remains even for those particular months.

Using ~~the~~ SA from AERONET noticeably reduces the calibration difference ~~noticeably~~ (Fig. 6) at 500 nm for most months ~~in~~ at both locations, but the effect can explain a calibration difference of approximately up to 0.75% (September 2019, Rome), while the calibration differences in Rome are in the ~~between~~ 2.5 - 3.5% range (~~table~~ Table S1 supplement). ~~In the case of~~For SVA, there are also noticeable differences of 0.5 - 1% from the central value (Fig. 7). SVA, like IRI, also shows a ~~also~~ particularly high sensitivity during the second month (August 2018, Davos). The central SVA value corresponds to identical ~~all~~ input parameters with respect to ~~as~~ the original calibration, and therefore we expect the magenta line (original) in ~~fig~~Fig. 7 and the blue line (central SVA) to be identical. Some differences below 0.1% are probably present~~t~~ ~~probably~~ in most months due to the useage of different compilers and versions of the Skyrad pack 4.2. However, for September 2019 in Rome at 500 nm they differ by up to 0.5% and for August 2018 in Davos at 870 nm by ~~above~~ >1%. This may be a result of computational instability ~~or modifications in the Skyrad pack 4.2 screening criteria for the selection of data to perform the ILP since the time the instruments were initially calibrated~~. For the ~~rest of the~~other months, such differences are below 0.1%.

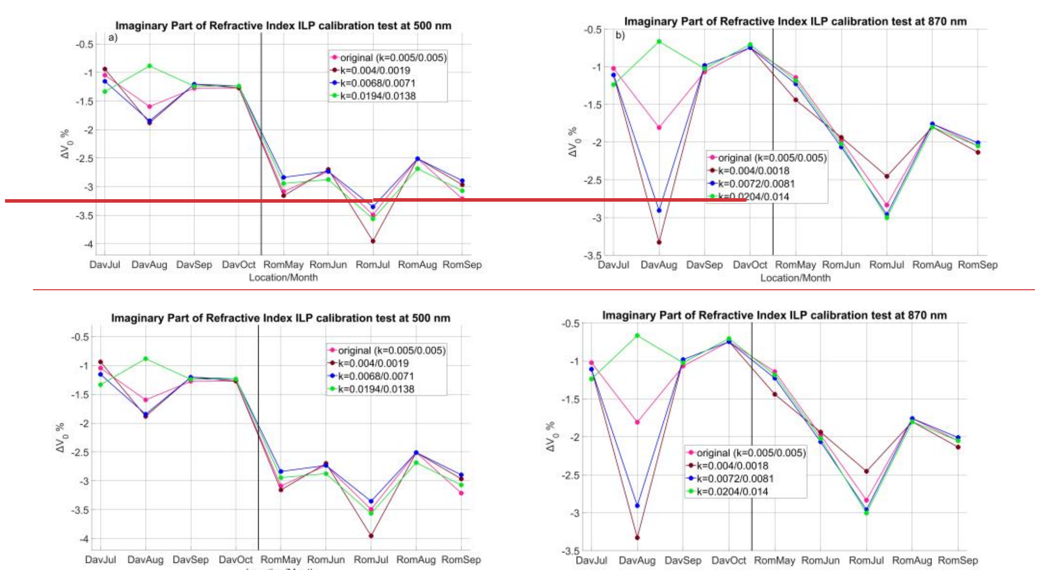
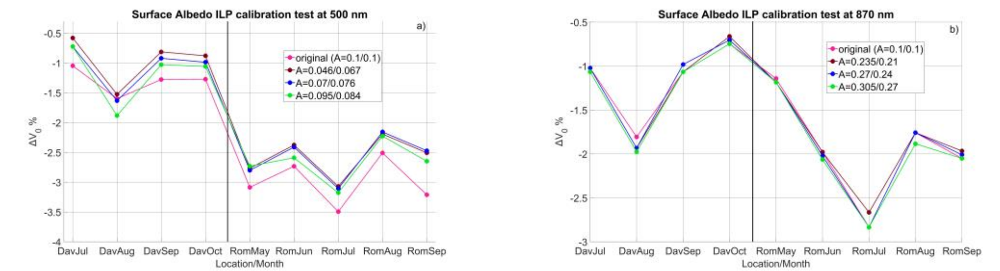


Figure 5: The ~~%~~ percentage differences between the IL calibration and calibration transfer for POMCNR during the QUATRAM II months using different values of the imaginary refractive index (original calibration, median k and median±std). ~~Left panel~~a): 500 nm. ~~Right panel~~b) ~~and righ~~:~~t~~ 870 nm. ~~The l~~Left ~~of the side from the~~ black vertical line corresponds to the Davos calibrations and right ~~the right side~~corresponds to Rome. ~~The black vertical line separated separates the Davos and Rome months (July to October 2018 and May to September 2019).~~



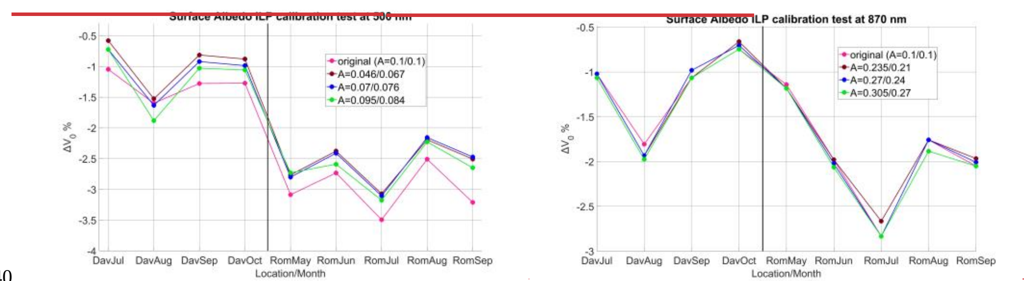
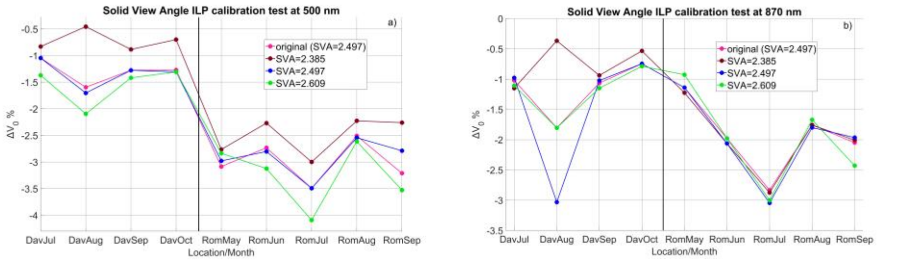


Figure 6: The percentages%d difference between the IL calibration and calibration transfer for POMCNR during the QUATRAM II months using different values of surface albedo (original calibration, median A and median±std). Left panela): 500 nm. Right panelb): and right 870 nm. The lLeft side of the from the black vertical line corresponds to the Davos calibrations and the right corresponds side to Rome. The black vertical line separated separates the Davos and Rome months (July to October 2018 and May to September 2019).



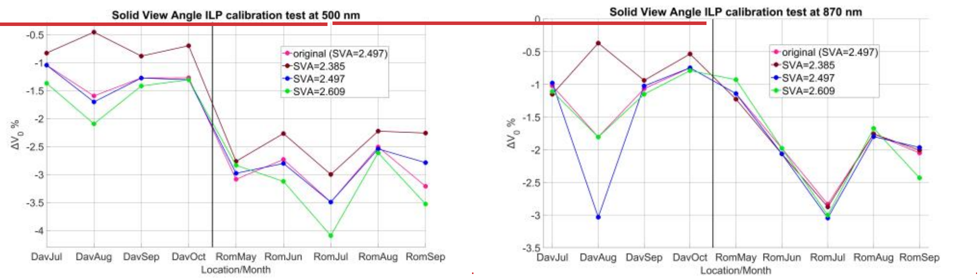


Figure 7: The % difference between ILP calibration and transferred calibration ~~transfer~~ for POMCNR during the QUATRAM II months using different values of ~~solid view angle~~SVA (original calibration, runs with the provided SVA and SVA± fixed deviation). ~~Left panel~~a): 500 nm~~. Right and panel~~b):~~right~~ 870 nm. ~~The l~~Left ~~s~~of the ~~ide from the~~ black vertical line corresponds to the Davos calibrations and ~~the r~~right ~~side~~ corresponds to Rome. ~~The black vertical line separated separates the Davos and Rome months (July to October 2018 and May to September 2019).~~

~~3~~4.2.2.~~2~~ Sub-study 2: ILP Test based on local observations: all parameters as variables

4.2.2.1 Sub-study 2: Methodology

In the second sub-study, we alter the values of all parameters simultaneously except SVA (we used the value provided by ESR). Again, the goal is to adapt the input parameters to the site conditions. We calculated the ILP calibration for two separate cases:

a) Average case: one calibration per month using the monthly average values used in the first sub-study for all five parameters under test (RRI, IRI, P, TOC and SA).

b) ~~"~~"Selected~~"~~" case: one calibration per month. Here we selected one of the three values used in the first sub-study for the same five parameters. The selected values are those ~~of the~~ three that lead to a larger calibration constant. We picked only one month per location for this case. The values of the input parameters used for this second sub-study are shown in the supplement Section 6.

4.2.2.2 Sub-study 2: Results

In ~~the~~ this section, we present the results of the second sub-study described in section ~~2.~~4.2.2.1 ~~Here t~~There are two ~~cases of~~ calibration cases that we tested in the whole QUATRAM II campaign.

The results ~~on~~ in ~~the t~~Table 4 show changes <~~for the average case less than~~ 0.5% for the average case ~~changes~~ with the exception of ~~in the case of~~ August 2018 in Davos. D~~, d~~ue to the large sensitivity in the IRI, the calibration changed by >~~more than~~ 1%.

Under the ~~‘‘~~“selected”~~’’~~ case (selected conditions for all parameters that increase the ILP calibration), there is a larger increase of the calibration in Davos ~~and~~ compared to Rome at both wavelengths (~~table~~ Table 4), but ~~. A~~all differences are below 1%.

Table 4: The ~~%~~percentage difference between the original ILP and ~~transferred~~ calibration transfers~~s~~ minus the ~~%~~percentage difference between the ILP method, for ~~under~~ selected conditions and the ~~transferred~~ calibration transfer~~constant d~~. Positive values indicate a smaller difference between the ILP calibration and calibration transfers compared to the differences of the original calibrations.

Instrument	Location	Year	Month	ΔV_0 % 500 nm	ΔV_0 % 870 nm
Average case					
POMCNR	DAVOS	2018	7	0.25	-0.09
POMCNR	DAVOS	2018	8	0.14	-1.27
POMCNR	DAVOS	2018	9	0.36	0.08
POMCNR	DAVOS	2018	10	0.29	0.08
POMCNR	ROME	2019	5	0.46	-0.09
POMCNR	ROME	2019	6	0.36	-0.26
POMCNR	ROME	2019	7	-0.14	-0.13
POMCNR	ROME	2019	8	0.32	-0.04
POMCNR	ROME	2019	9	0.46	0.00
~~‘‘Selected’~~ “Selected” case					
POMCNR	DAVOS	2018	9	0.89	0.34
POMCNR	ROME	2019	8	0.60	0.13

34.2.23.3 Sub-study 3: ILP sensitivity tests

4.2.3.1 Sub-study 3: Methodology

In the third sub-study, we tested the IRI, SA and SVA for a more extensive number of values (seven fixed values regardless of the location) to assess the behaviour of the calibration. For IRI and SA, the selection is based on the three values of the first sub-study, the 5th - 95th percentiles of the observations and the minimum/maximum values. We also added semi-arbitrary values between the observed and two extreme values (one very small and one very large) to test the performance of the method over a wider range of inputs. For SVA, we use values based on the differences between the different SVA calibration procedures appearing in Campanelli et al. (2023). The actual values for each parameter are in the supplement, Section 10, Table S11.

4.2.3.2 Sub-study 3: Results

In this section, we present the results of the third sub-study described in section 2.44.2.23.1, where we only test IRI, SA and SVA for seven values over a larger range. We only selected only one month per location, avoiding the August 2018 and July 201997 due to the behaviour presented in the previous two sections 4.2.1.2 and 4.2.2.2. In the fFigures 8 - 10 show the are the results per for each parameter.

Changing only the IRI (while it is below <0.05) shows that the ILP changes by <less than 0.25% for both wavelengths and locations (Fig. 8) and IRI below 0.05 . Increasing IRI to larger >0.05 or even either to other rare or and unrealistic values has no effect on the calibration. Therefore, ILP IRI appears to be either have a affect it significantly or a small effect on the ILP calibration, or very little depending on the month.

Changing only the SA (Fig. 9), shows (Fig. 9) a monotonic, but non-linear dependence of the ILP calibration, where larger SA leads to a smaller calibration constant. At 870 nm, there is a maximum calibration constant at SA = 0.04 with approximately 0.07 - 0.08 being the average values from AERONET and 0.1 the values used by ESR. At 500 nm, the difference between the ILP calibrations in Davos and Rome are also smaller at lower SA, also are reducing at lower SA showing that ILP in Rome is affected to a larger extent by the SA value at 500 nm. However, even when using an SA value as low as 0.02, the remaining calibration difference between the calibration transfer and ILP at 500 nm is approximately 2% in Rome and 0.7% for Davos. At 870 nm the difference is at least 0.95% for Davos and 1.7% for Rome for all SA values used as input. At 870 nm the difference remains at least for 0.95% for Davos and 1.7% for Rome.

Finally, in the case offor SVA (Fig. 10), there is a monotonic decreasing dependency of the calibration constant and SVA, at 500 nm, while some fluctuations occur at 870 nm. The minimum calibration difference at 500 nm is approximately 0.58% for Davos and 1.7% for Rome, while at 870 nm, results are 0.78% for Davos and 1.6% for Rome.

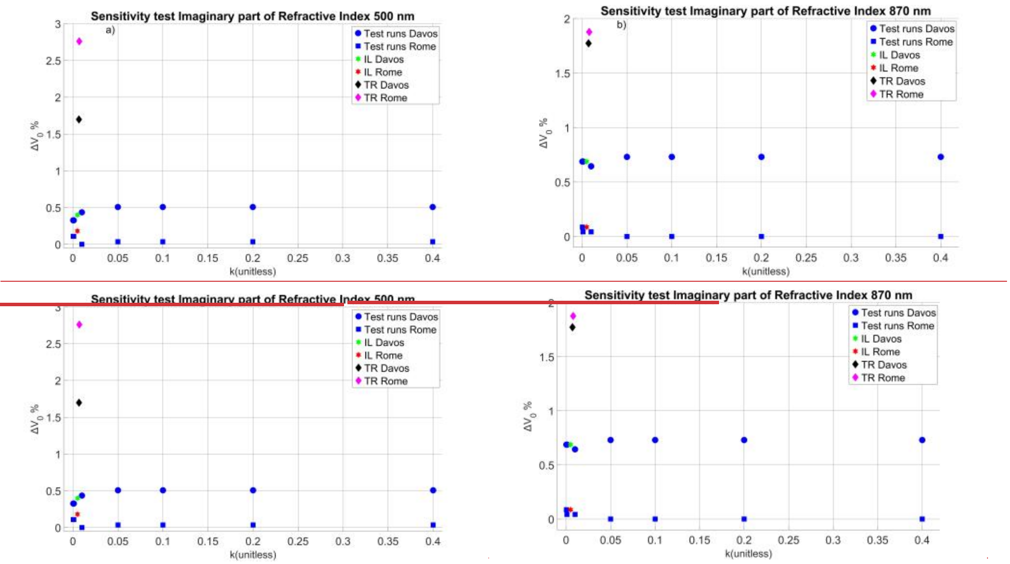
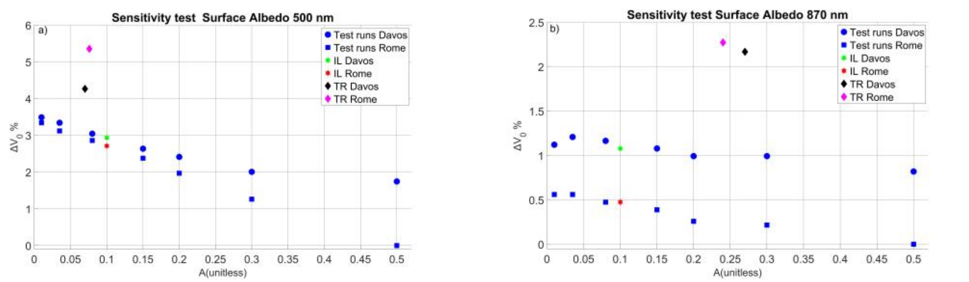


Figure 8: Sensitivity test of the IL calibration with respect to ~~on~~ the imaginary refractive index at 500 (~~left~~a) nm and 870 nm (~~right~~b). The vertical axis shows the ~~%~~ percentage difference of each calibration from the selected zero value. ~~As~~ For the ~~0~~latter, we selected the lowest calibration constant of the sensitivity tests present in each graph. The blue squares correspond to ~~Rome~~ sensitivity runs at Rome, the blue circles to Davos, the stars to the original ILP calibration transfer and the diamonds to the calibration constants ~~from transfer~~ with a PFR as reference.



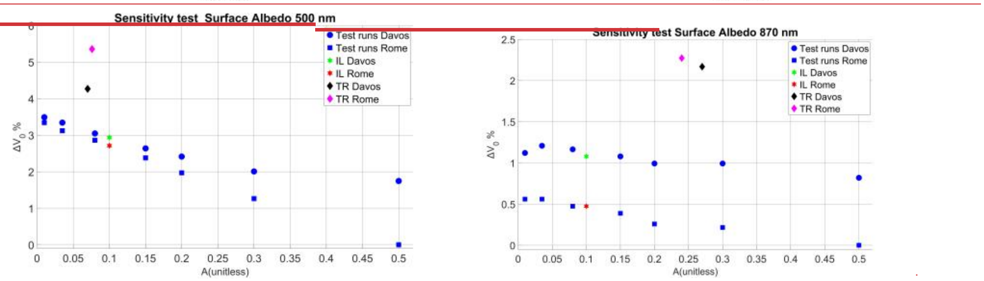


Figure 9: Sensitivity test of the IL calibration ~~on~~ with respect to the imaginary refractive index at 500 (~~left~~a) nm and 870 nm (~~right~~b). The vertical axis shows the ~~%~~ percentage difference of each calibration from the selected zero value. For the latter~~As 0~~, we selected the lowest calibration constant of the sensitivity tests present in each graph. The blue squares correspond to ~~Rome~~ sensitivity runs at Rome, the blue circles to Davos, the stars to the original ILP calibration and the diamonds to the calibration transfer constants ~~from transfer~~ with a PFR as reference.

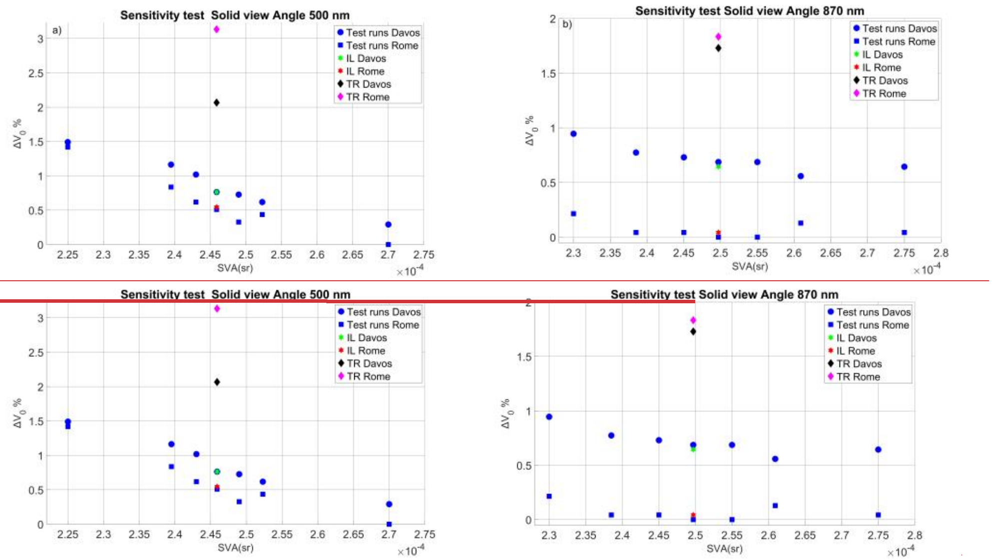


Figure 10: Sensitivity test of the IL calibration ~~on~~ with respect to the imaginary refractive index at 500 ~~(left~~a) nm and 870 nm (~~right~~b). The vertical axis shows the % difference of each calibration from the selected zero value. ~~As 0~~For the latter, we selected the lowest calibration constant of the sensitivity tests present in each graph. The blue squares correspond to ~~Rome s~~sensitivity runs at Rome, the blue circles to Davos, the stars to the original ILP calibration and the diamonds to the calibration transfer constants ~~from transfer~~ with a PFR as reference.

~~3.2~~4.3 Investigation ~~on~~ of the ~~aerosol optical depth~~AOD retrievals from sky radiance

4.3.1 Methodology

Since the ILP method is performed using a linear fit of the logarithm of DSI with respect to the product of air mass coefficient and ~~scattering aerosol optical depth (~~sc-AOD~~)~~ (Eq. 4), errors from the retrieval of sc-AOD will transfer errors to the calibration. Since there is no reference dataset available for sc-AOD, we tried to indirectly investigate potential errors using available data.

The Skyrad code retrieves both sc-AOD and SSA through inversion modelling and calculates the corresponding AOD as additional information. Therefore, we initially compare the AOD dataset with the PFR AOD for potential differences. However, systematic underestimation or overestimation of both the sc-AOD and SSA retrievals can result in opposing errors to the corresponding AOD that cancel each other. Due to the limitations of the AERONET SSA dataset (lack of level 2.0 data and limited number of retrievals per day), we cannot evaluate the SSA retrieved by Skyrad 4.2 with confidence. Also, part of the SSA difference between the AERONET product and the output of the Skyrad code for the ILP calibration may be attributed to the fixed refractive index and the different scattering angles in the almucantar geometry used for the sky radiance measurements (ILP uses only forward scattering having a maximum angle of 30°).

Another indirect method to investigate the effect of the sc-AOD retrievals on the calibration performance is to use a different inversion model to retrieve sc-AOD and to re-calibrate the instrument with the ILP method. We therefore used the inversion model, Skyrad pack MRI version 2 (Kudo et al, 2021). MRI allows the modelling of non-spherical particles in contrast to Skyrad pack 4.2 retrievals. It also introduces stability constraints on the edges of the size distribution as well as other smoothness constraints (see Kudo et al., 2021 for a detailed description). As mentioned in Kudo et al., 2021, the MRI method is more accurate at high AOD. Under low AOD conditions in Davos, a noticeable portion of data showed large retrieval errors and unrealistic sc-AOD/AOD values. However, there was sufficient data at both locations to recalculate the ILP calibration, and hence it was applied to the QUATRAM II data.

We also investigated whether the variability of the SSA corresponding to the Skyrad 4.2 and MRI retrieval shows any association with the calibration differences. All retrieved AOD, sc-AOD and SSA data retrieved by MRI are screened according to the ILP criteria: keeping only data corresponding to AOD at 500 nm < 0.4 and air mass < 3.

4.3.2 Results

As discussed in section 3.~~1~~2.2, the ILP method can have significant random uncertainty as individual ILPs for half~~-~~days can lead~~ing~~ to different average monthly values ~~that are averaged monthly~~. However, the vast majority of daily calibration constants are lower than the calibration transfers from PFR and most of them by >~~more than~~ 0.5 - 1% (~~table~~ Table 5) for both locations and wavelengths. ~~This shows the significance of the systematic bias.~~ One way to ~~get~~ obtain such biased results, is a systematic underestimation in ~~the~~ sc-AOD ~~provided~~ by the inversion of NSR or an underestimation of sc-AOD in the small air masses and overestimation in ~~the~~ large air masses.

Table 5: The percentage of daily ILP calibration constants below the corresponding monthly calibration transfer (column 4), below a ~~the~~ calibration transfer value $\%\Delta V_0 <= -0.5\%$~~at least 0.5%~~ (column 5) and $\%\Delta V_0 <= -1\%$~~at least 1%~~ (column 6). The rows correspond to the days used for the final ILP monthly calibrations for each location ~~under~~ and for all campaigns at ~~1~~ a single wavelength.

Wavelength (nm)	Location	Number of days	$\%\Delta V_0<0$	$\%\Delta V_0<=-0.5\%$	$\%\Delta V_0<=-1\%$
500	DAVOS	45	95.56	91.11	73.33
500	ROME	112	100.00	100.00	98.02
870	DAVOS	38	94.74	86.84	52.63
870	ROME	101	97.03	96.04	93.07

In this section, we investigate the effect of the sc-AOD retrieval through inversion ~~of~~ ~~on~~of the ILP calibration. ~~Performance~~ As there were two inversion algorithms available, we compare the calibration and the sc-AOD calculated by Skyrad pack 4.2 with the calibration and sc-AOD from Skyrad MRI.

The AOD from Skyrad 4.2 is retrieved through the inverted sc-AOD ~~it~~ and may show similar errors. Since we do not have a sc-AOD reference dataset, we compared the Skyrad AOD with ~~the~~ PFR AOD.

The ~~AOD~~ difference between the AOD retrieved from the Skyrad pack 4.2 using almucantar scans of POM and the PFRs show a systematic underestimation as expected, ~~aside from~~except for the comparison at 870 nm for Davos (~~table~~ Table 6). The differences are also higher in Rome than in Davos. However, the median differences are significantly smaller than ~~the ones~~those corresponding to the ESR direct sun AOD product compared to the same PFRs and the percentage of differences within the ~~higher~~ WMO limits is higher ~~higher~~. The AOD differences ~~are~~ also increas~~ing~~e ~~for~~ with smaller air masses in Rome, but not in Davos. For air masses below 1.5, the median AOD difference is -0.012/-0.004 at 500/870 nm in Rome and 0.000/0.001 at 500/870 nm in Davos. For air masses above 2, the median AOD difference is -0.005/-0.000 at 500/870 nm in Rome and -0.003 /0.000 at 500/870 nm in Davos. More details including linear fitting of the air mass dependencies are available in the supplement section 12 ~~table~~ Table S15.

Table 6: The statistics of the differences between ~~the~~ AOD from Skyrad pack 4.2 using POM almucantar scans and ~~the~~ AOD from PFR. The results correspond to all QUATRAM campaigns ~~for~~ at each location. The time difference threshold is 30 seconds. ~~Starting from the third column, we show the median of all AOD differences, the percentage of differences within the WMO limits, the 5th and the 95th percentiles of AOD differences and the total number of measurements compared per location.~~

Location	**wavelength**	**median difference**	**WMO limits %**	**P5th**	**P95th**	**Number of measurements**
DAVOS	500	-0.002	82.91	-0.014	0.015	1129
DAVOS	870	0.000	97.25	-0.004	0.007	1129
ROME	500	-0.009	64.09	-0.027	0.007	1231
ROME	870	-0.003	92.85	-0.012	0.009	1231

Using the sc-AOD from MRI as an input to the ILP method instead of ~~the~~ Skyrad 4.2 in Davos 2018 and Rome 2019, we obtained different calibration constants for each month, but there is no consistent improvement (~~table~~ Table 7). At 500 nm, ~~6~~ six out of ~~9~~ nine months show a calibration closer to the calibration transfers by between 0.29 to -0.96% (negative differences in Table 7), while at 870 nm the calibration constant is larger for ~~increased~~ only ~~for 3~~ three months (0.04 - 1.39%). ~~The sc-AOD median differences are negative at 500 nm and positive at 870 nm, which is in accordance with the~~

~~sign of the calibration differences for most cases.~~ However, ~~they~~ the AOD median differences are very small (up to 0.002) and there is no consistency between sc-AOD and calibration differences (~~table~~ Table 6). Due to the fact that the datasets are different, there is also a different selection of individual sc-AOD inversions and days that pass~~ing~~ the criteria for the final ILP calibration. The combination of using randomly different sc-AOD~~,~~ points and ~~half~~ half-day selections, results ~~to~~ in the observed calibration differences ~~observed~~ that are mainly ~~below~~ <~~1~~1%. Such random differences are similar to the magnitude of ILP CV% values (defined in section 3.2.2)~~–~~ in Campanelli et al. (~~,~~2023).

Table 7: The ~~%~~ percentage differences between the original ILP calibrations and the ILP calibrations using sc-AOD inverted by Skyrad MRI (columns 3 and 4) and the median differences of the corresponding sc-AOD (columns 6 - 7).

Year	**Month**	**ΔV_0 % 500 nm**	**ΔV_0 % 870 nm**	**Median Δsc-AOD 500 nm**	**Median Δsc-AOD 870 nm**	**Number of sc-AOD measurements**
2018	7	0.40	0.17	-0.002	0.000	194
2018	8	-0.54	2.16	-0.002	0.001	404
2018	9	-0.96	-0.64	-0.002	0.000	332
2018	10	-0.54	-1.39	-0.002	0.000	184
2019	5	-0.44	0.17	-0.002	0.001	238
2019	6	-0.29	-0.04	-0.001	0.002	1215
2019	7	0.33	0.22	-0.001	0.001	1178
2019	8	0.11	0.13	-0.001	0.001	1123
2019	9	-0.51	0.26	-0.001	0.001	680

The ratio of the provided sc-AOD and AOD in the ILP output allows us to calculate the corresponding SSA. The number of available QUATRAM II common measurements between Skyrad 4.2 and MRI is 1114 for Davos and 4434 for Rome. ~~In the case of~~For ILP retrieved SSA from both Skyrad 4.2 and MRI, we ~~can see~~ mainly observe a ~~l~~larger median in Davos (0.952/0.926 for 500/870 nm from Skyrad 4.2 and 0.959/0.939 from MRI) compared to Rome (0.934/0.917 from Skyrad 4.2 and 0.942/0.927 from MRI). The monthly values are in the supplement ~~table~~ Table S12. The difference between the 80th and 20th percentiles of the SSA is overall ~~is~~ larger in Rome at 500 nm (0.03/0.02 from Skyrad 4.2 at 500/870 nm and 0.025/0.015 from MRI) and larger in Davos at 870 nm (0.021/0.029 from Skyrad 4.2 nm and 0.014/0.02 from MRI). However, there are ~~month~~ month-~~to~~ to-month variations. In the supplement ~~table~~ Table S13, we show the monthly medians of the daily differences between the 80th and 20th percentiles. Depending on the month, either Rome or Davos shows a larger variability. ~~The number of available QUATRAM II common measurements is 1114 for Davos and 4434 for Rome.~~

~~4~~ 5 Discussion

In section 3.~~1~~2.1, we compared the AOD between several PFRs and POMs ~~in~~ at two locations with different characteristics (Davos and Rome) ~~under~~ using different POM calibration methods ~~of the POM.~~. Using the original POM AOD (calculated after ILP calibration of the POMs), we found that the POMs ~~provide~~ systematically ~~provide~~gave lower AOD values than the PFRs up to the 0.034 range at 500 nm and 0.018 at 870 nm (median difference). This systematic difference is larger in Rome. Using calibration transfers with the PFR as a reference to re-calibrate the POMs, we achieved excellent agreement showing that the differences between the post~~-~~processing algorithms of the networks and the technical characteristics only have a ~~only~~ minor effect on AOD differences. The major cause of AOD differences was the calibration method. The calibration differences per campaign were approximately 0.7 - 1.6% in Davos and 1.6 - 3.5% in Rome at 500 nm, and 0.2 - 1.8% in Davos and 1 - 3.4 % in Rome at 870 nm (supplement ~~table~~ Table S1). The AOD differences per campaign were approximately 0.~~005~~006 - 0.01 in Davos and 0.015 - 0.~~035~~034 in Rome at 500 nm, and 0 - 0.005 in Davos and 0.005 - 0.017 in Rome at 870 nm (section 3.~~1~~2.1).

We also compared the AOD between the reference PFR and the co-located CIMEL for each case for cross-validation. All median AOD differences between CIMEL and PFR were ~~below~~ < 0.01 and the traceability criteria ~~are~~ were satisfied with the exception of ~~the Rome phase in~~ the QUATRAM I campaign in Rome and at ~~the~~ 500 nm ~~of~~ for ~~the Rome phase in~~ the QUATRAM II campaign, also in Rome. The generally good agreement between PFR and CIMEL is consistent with the small differences of the CIMEL and PFR~~-~~based calibration transfers in Campanelli et al. (~~,~~2023).

Regarding the PFR calibrations, the uncertainty is lower as shown in section 3.~~1~~2.2. The PFRN01 and PFRN14 sun photometers used ~~for the~~in Rome ~~phases~~ showed good calibration stability before and after their shipment~~s~~ (section 3.~~1~~2.2). ~~The~~ PFRN27 used in ~~the~~ Davos ~~phases~~ as a reference, was ~~for the whole 2017-2021 period~~ present in Davos as part of the PFR reference triad for the whole of the 2017 – 2021 period. ~~Also~~In addition, it is used in a long-term comparison study with AERONET (Karanikolas et al., 2022), and has ~~-~~shown~~ings~~ very good agreement with a co-located CIMEL during the ~~in the period~~ 2007 ~~-~~ – 2019 period.

In an ~~A~~attempt~~ting~~ to explain the observed calibration differences, we investigated whether ~~the two~~both stations show some systematic difference during the campaigns in terms of the values or variability in aerosol properties~~' values or variability~~ that could explain the different calibration performance. The available datasets of AOD, SSA and AE showed no such association. However, the AERONET SSA dataset has important limitations ~~of~~ with regard to data availability and accuracy as explained in section ~~2.~~4.1.1. One explanation could be that the values or the variability of SSA and AE affect the calibration proportionally to the AOD ~~levels~~values. However, we cannot identify such an association ~~as well~~ from our results (details in Figs. 2 - 4 and supplement ~~table~~ Table S1). ~~In~~ For example, in Davos, the last two months of QUATRAM II (~~9~~September-~~10/~~October 2018) show similar calibration differences between the ILP method and calibration transfers under different conditions for ~~in~~ all ~~3~~ three parameters (AOD, SSA and AE). ~~AlsoIn addition, AOD at 500 nm was above 0.1 in QUATRAM I (8/2017) and below 0.05 the AOD at 500 nm is above 0.1, while in in QUATRAM III (10/2021) below 0.05,~~

but the calibration difference is was smaller in QUATRAM I. Similarly, iIn Rome at during QUATRAM II, the first month (5/2019) shows simultaneously exhibits the lowest AOD and SSA variability in at both wavelengths. At 500 nm, the second and fourth months (6 and 8/2019) show a smaller calibration difference, while AOD is higher and all three parameters are more variable. Tthe third month (7/July 2019) shows the largest calibration difference under similar AOD and SSA conditions with June6 and 8/August 2019, but lower AE variability.

We also conducted a sensitivity analysis of the ILP method under different conditions with respect to on its six input parameters: (Real part of refractive index (RRI), Imaginary part of refractive index (IRI), Surface albedo (SA), Total Ozone Colum (TOC), Surface Pressure (P) and Solid View Angle (SVA)). SVA and SA errors can explain part of the underestimation in the ILP calibration underestimation. Regarding IRI, the ILP calibration showed very little sensitivity duringof most months (which is consistent with the study in Campanelli et al., 2004), but was very large for specific months. and IRI valuesThis showing showed some evidence of model instabilities under certain conditions and combinations of NSR and IRI values.- RRI, TOC and P showed no evidence of a significant effect. However, tTo conclude, the largest most part of the calibration differences remained unexplained.

By comparing the retrieved AOD from the Skyrad code (using NSR) with PFR AOD, we can identify an underestimation, mainly in Rome, although smaller than the AOD retrieved from direct sun scans and the ILP calibration. However, the ILP calibration uses sc-AOD instead of AOD for the calibration. A stronger underestimation of sc-AOD compared to AOD or dependence of the sc-AOD error with the air mass can explain the calibration difference. Such underestimation may not be not fully visible in the AOD dataset due to a systematic error in the ILP inverted SSA that reduces the AOD error.

Using an alternative inversion model (Skyrad MRI) to retrieve sc-AOD, we found no significant systematic differences of sc-AOD compared to Skyrad 4.2. The ILP calibration using MRI had positive and negative differences from the original oneILP, mainly by less than 1%. Such differences can be attributed to the different selection of data and random differences of sc-AOD between the 2both models. Under both models, we found no consistency between the SSA variability corresponding to the provided sc-AOD/AOD. The AERONET median SSA is higher in Davos (0.02), however, the difference is within the uncertainty of the inversions and corresponds to different scattering angles. Also, the high SSA uncertainties and the mainly low sensitivity of the ILP calibration with respect to the imaginary part of the refractive indexIRI further limit further the significance of this finding.

Another issue related to the ILP calibration is its random uncertainty. Despite the clear systematic bias we observed compared to the calibration transfers, the random fluctuations uncertainty component remains significant. In section 3.12.2 we showed that there can be both fluctuations for consecutive monthsa month-to-month variability of the calibration constant and estimated random uncertainty componentsies of the ILP calibration above 1%. The lack of coincidence between the month-to-month variability of ILP and transfer-based calibrations suggests that indeed we cannot indeed attribute these month-to-month variabilityfluctuations of ILP calibrations to instabilities of the instruments. The calibration transfers showed smaller uncertainty and larger stability apart from large shifts during specific months. The PFR calibrations are more stable and have smaller uncertainties than the calibration transfers, so we cannot attribute the calibration transfer fluctuations to changes in the PFR response. However, as described in section 3.~~1~~2.2, we cannot attribute all fluctuations in ILP calibrations ~~fluctuations~~ to their CV% value ~~of the ILP calibrations and changes in the instruments, but rather to the overall ILP uncertainty~~. A potential source of uncertainty (or bias) is the linearity of the fit during the ILP calibration. The currently used ~~linear fitting~~ standard error threshold of the linear fit may allow a discrepancy from ~~the~~ linear behaviour that is large enough to cause uncertainties at the observed level. More research is needed to further clarify the matter.

The calibration underestimation observed by the ILP calibration compared to the calibration transfers is probably a result of errors in the sc-AOD retrievals. As the ILP method shows sensitivity, mainly to the provided ~~normalized sky radiance~~ ~~(~~NSR~~)~~, the retrieval errors are probably a result of assumptions in the forward model that simulates the NSR. The effect is amplified in Rome compared to Davos. A known constant difference between ~~the two~~both locations is the altitude. As Davos is higher ~~(~~by ~~~~~about 1500 m~~)~~, the atmospheric pressure is constantly lower leading to a reduced Rayleigh scattering optical depth, which contributes towards a reduced DSI and decreased multiple radiation scattering. Therefore, the NSR dependence with the scattering angle can be systematically different between ~~the two~~both locations for any given SZA. In that case, the forward ILP model~~ of ILP~~ may ~~simulate~~ less accurately simulate the effect of ~~the~~ multiple scattering in Rome or the increased multiple scattering there may amplify the errors of the simulations. More research is required to investigate whether the source of the larger calibration differences in Rome ~~.~~is indeed due to the lower altitude of the station in Rome, ~~station~~ and to what extent it can be generalized for other sites.

Significant improvement ~~seems to~~may be possible using the Cross Improved Langley Plot (XILP) (Nakajima et al., 2020; Campanelli et al., 2023) ~~instead~~, which seems to lead ~~in~~ to smaller biases. XILP performs ILP with the axes reversed, but also includes different criteria for the selection of data used for the final linear fit and the days considered as valid. However, XILP also showed a few cases ~~of~~ with large differences (or even larger than ILP) compared to the calibration transfer. Therefore, more research is required to assess the XILP sensitivity in the sc-AOD, inputs parameters and whether it can lead to long-term traceability of AOD regardless of the location and ~~the~~ conditions.

~~5~~ 6 Conclusions

In this study, we assess AOD differences between GAW-PFR and ESR instruments and investigate their causes. We used data ~~of~~ from three intercomparison campaigns, each with two ~~phases~~locations~~.~~: ~~eachOne phase was in~~location Davos, a mountainous area and ~~one i~~the other is~~n~~ Rome, a low altitude urban area. A ~~Comparison~~ comparison of different pairs of PFR and POM instruments showed that the traceability criteria are satisfied at 870 nm in Davos for all campaigns and for Rome in one campaign. ~~At 500 nm they~~Criteria are not satisfied at 500 nm, but ~~in Davos~~ the differences in Davos are smaller and below the AOD standard uncertainty (median AOD difference below 0.01). Our analysis shows that the contribution of the instrument and post~~-~~processing differences to the AOD differences is minor. The major cause is the different calibration methods. We concluded that the ILP calibration method used by ESR results ~~to~~ in a systematic underestimation of the calibration constant and as a result, an underestimation in ~~the~~ AOD as well, compared to GAW-PFR and AERONET measurements. Our investigation ~~on~~ of the causes showed that part of the difference (mainly at 500 nm) can be explained by potential errors in ~~the surface albedo~~SA and the instrument ~~solid view angle~~SVA used as input for the ILP calibration. However, the largest part of the difference cannot be attributed to errors in the input parameters~~. It~~ but can be explained by errors in the sc-AOD retrieval, which is required to perform the ILP method. The error is probably a result of the forward model assumptions. A potential explanation could be related to the way the model handles multiple scattering, which probably amplifies the error ~~in~~ at lower altitude sites. This work is a demonstration of the limitations and challenges of the ILP ~~‘~~"on-~~site’~~ site" calibration procedure for sun photometers. The present study and Campanelli et al. (~~,~~ 2023) offer a starting point for future research aimed at a better ~~to their further~~ understanding with ~~towards~~ more general conclusions and potential improvements.

Code availability. The ~~used~~ version of SKYRAD 4.2 code package used in this study is available through communication with the authors.

Data availability. The CIMEL AOD data are available from https://aeronet.gsfc.nasa.gov/

The PFR and POM raw signals and AOD data are available ~~through communication with~~by contacting the authors.

Author contribution. AK analysed the data and wrote the paper with contributions from the co-authors. AK and SK conceptualized the study. NK and SK contributed to the PFR sun photometer data provision. NK assisted with the CIMEL and PFR sun photometer data selection. MC and VE contributed to the POM sun and sky radiometer data provision. MC, MM and GK contributed to the SKYRAD 4.2 pack code provision and assisted with its operation. GK contributed with the SKYRAD pack MRI output. SN assisted with the editing. All authors were involved in the interpretation of the results and reviewing the paper.

Competing interests. The authors declare that they have no conflict of interest.

Financial support. This research has been supported by COST (European Cooperation in Science and Technology) under the HARMONIA (International network for harmonization of atmospheric aerosol retrievals from ground-based photometers), action CA21119.~~.~~

Acknowledgments. Angelos Karanikolas has been supported by the European Metrology Program for Innovation and Research (EMPIR) within the joint research project EMPIR 19ENV04 MAPP "Metrology for aerosol optical properties". ~~The~~ EMPIR is jointly funded by the EMPIR participating countries within EURAMET and the European Union.

Stelios Kazadzis would like to acknowledge the ACTRIS Switzerland project funded by the Swiss State Secretariat for Education, Research and Innovation.

The participation of Gaurav Kumar has been ~~also~~ supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund through project PID2022-138730OB-I00, and Santiago Grisolia program fellowship GRISOLIAP/2021/048.

References

Bais, A. F., Lucas, R. M., Bornman, J. F., Williamson, C. E., Sulzberger, B., Austin, A. T., Wilson, S. R., Andrady, A. L., Bernhard, G., McKenzie, R. L., Aucamp, P. J., Madronich, S., Neale, R. E., Yazar, S., Young, A. R., de Gruijl, F. R., Norval, M., Takizawa, Y., Barnes, P. W., Robson, T. M., Robinson, S. A., Bailaré, C. L., Flint, S. D., Neale, P. J., Hylander, S., Rose, K. C., Wängberg, S.-Å., Hader, D.-P., Worrest, R. C., Zepp, R. G., Paul, N. D., Cory, R. M., Solomon, K. R., Longstreth, J., Pandey, K. K., Redhwi, H. H., Torikai, A., and Heikkilä, A. M.: Environmental effects of ozone depletion, UV radiation and interactions with climate change: UNEP Environmental Effects Assessment Panel, update 2017, ~~Photochem Photobiol Sci~~Photoch. Photobio. Sci., 17, 127–179, https://doi.org/10.1039/c7pp90043k, 2018.

Barnes, P. W., Williamson, C. E., Lucas, R. M., Robinson, S. A., Madronich, S., Paul, N. D., Bornman, J. F., Bais, A. F., Sulzberger, B., Wilson, S. R., Andrady, A. L., McKenzie, R. L., Neale, P. J., Austin, A. T., Bernhard, G. H., Solomon, K. R., Neale, R. E., Young, P. J., Norval, M., Rhodes, L. E., Hylander, S., Rose, K. C., Longstreth, J., Aucamp, P. J., Ballaré, C. L., Cory, R. M., Flint, S. D., de Gruijl, F. R., Häder, D.-P., Heikkilä, A. M., Jansen, M. A. K., Pandey, K. K., Robson, T. M., Sinclair, C. A., Wängberg, S.-Å., Worrest, R. C., Yazar, S., Young, A. R., and Zepp, R. G.: Ozone depletion, ultraviolet radiation, climate change and prospects for a sustainable future, ~~Nat Sustain~~Nature Sustainability, 2, 569–579, https://doi.org/10.1038/s41893-019-0314-2, 2019

Cachorro, V. E., Berjon, A., Toledano, C., Mogo, S. N., Prats, A. M., De Frutos, J. Vilaplana, M., Vilaplana, J. M., Sorribas, M., De La Morena, B. A., Gröbner, J., ~~-~~Laulainenet, N.: ~~"~~Detailed aerosol optical depth intercomparison between Brewer and Li-Cor 1800 spectroradiometers and a Cimel sun photometer~~,~~." ~~Journal of Atmospheric and Oceanic Technology~~J. Atmos. Ocean. Tech., 26, no. 8: 1558-1571, 2009.

Campanelli, M., Nakajima, T., and Olivieri, B.: Determination of the solar calibration constant for a sun-sky radiometer: proposal of an in-situ procedure, ~~Appl. Opt.~~Appl. Optics, AO, 43, 651–659, https://doi.org/10.1364/AO.43.000651, 2004.

Campanelli, M., Estelles, V., Kumar, G., Nakajima, T., Momoi, M., Gröbner, J., Kazadzis, S., Kouremeti, N., Karanikolas, A., Barreto, A., Nevas, S., Schwind, K., Schneider, P., Harju, I., Kärhä, P., Diémoz, H., Kudo, R., Uchiyama, A., Yamazaki, A., Iannarelli, A. M., Mevi, G., Di Bernardino, A., and Casadio, S.: Evaluation of “on-site” calibration procedures for sun-sky photometers, ~~Atmospheric~~ Atmos. ~~Measurement~~ Meas. ~~Techniques~~ Tech. ~~Discussions~~D., 1–26, https://doi.org/10.5194/amt-2023-165, 2023.

Correa, L. F., Folini, D., Chtirkova, B. and Wild, M.: Causes for ~~decadal~~ Decadal ~~trends~~ Trends in ~~surface~~ Surface ~~solar~~ Solar ~~radiation~~ Radiation in the Alpine ~~region~~ Region in the 1981-2020 ~~period~~Period, J. Geophys. Res.-Atmos., 129(9),~~Authorea Preprints~~ https://doi.org/10.1029/2023JD039998, ~~2023~~2024.

Cuevas, E., Romero-Campos, P. M., Kouremeti, N., Kazadzis, S., Räisänen, P., García, R. D., Barreto, A., Guirado-Fuentes, C., Ramos, R., Toledano, C., Almansa, F., and Gröbner, J.: Aerosol optical depth comparison between GAW-PFR and AERONET-Cimel radiometers from long-term (2005–2015) 1-min synchronous measurements, Atmos. Meas. Tech.~~Atmospheric Measurement Techniques~~, 12, 4309–4337, https://doi.org/10.5194/amt-12-4309-2019, 2019.

Doppler, L., Akriti Masoom, A., and Karanikolas, A.: Create a list of existing and foreseen campaigns or experiments needed for night and day aerosol measurements and report on the data collection and analysis of the data/measurements, https://harmonia-cost.eu/wp-content/uploads/2023/10/COST_CA21119_HARMONIA-WG1_Deliverable-D11.pdf, 2023.

Drosoglou, T., Raptis, I. P., Valeri, M., Casadio, S., Barnaba, F., Herreras-Giralda, M., Lopatin, A., Dubovik, O., Brizzi, G., Niro, F., Campanelli, M., and Kazadzis, S.: Evaluating the effects of columnar NO2 on the accuracy of aerosol optical properties retrievals, Atmos. Meas. Tech., 16, 2989–3014, https://doi.org/10.5194/amt-16-2989-2023, 2023.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), ~~Geoscientific Model Development~~Geosci. Model. Dev., 9, 1647–1672, https://doi.org/10.5194/gmd-9-1647-2016, 2016.

Estellés, V., Campanelli, M., Smyth, T. J., Utrillas, M. P., and Martínez-Lozano, J. A.: Evaluation of the new ESR network software for the retrieval of direct sun products from CIMEL CE318 and PREDE POM01 sun-sky radiometers, Atmos. Chem. Phys., 12, 11619–11630, https://doi.org/10.5194/acp-12-11619-2012, 2012.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: 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., 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, 2019.

Gröbner, J., Kouremeti, N., Hülsen, G., Zuber, R., Ribnitzky, M., Nevas, S., Sperfeld, P., Schwind, K., Schneider, P., Kazadzis, S., Barreto, Á., Gardiner, T., Mottungan, K., Medland, D., and Coleman, M.: Spectral aerosol optical depth from SI-traceable spectral solar irradiance measurements, Atmos. Meas. Tech.~~Atmospheric Measurement Techniques~~, 16, 4667–4680, https://doi.org/10.5194/amt-16-4667-2023, 2023.

Gröbner, J., & Kouremeti, N.: The Precision Solar Spectroradiometer (PSR) for direct solar irradiance measurements~~.~~, ~~Solar Energy~~Sol. Energy, 185, 199-210, https://doi.org/10.1016/j.solener.2019.04.060, 2019.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of EnvironmentRemote Sens. Environ., 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, 1998.

Hou, X., Papachristopoulou, K., Saint-Drenan, Y.-M., and Kazadzis, S.: Solar Radiation Nowcasting Using a Markov Chain Multi-Model Approach, Energies, 15, 2996, https://doi.org/10.3390/en15092996, 2022.

IPCC: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Lee, H. and Romero, J. (eds.)]. IPCC, Geneva, Switzerland, 184 pp., doi: 10.59327/IPCC/AR6-9789291691647, 2023IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp. https://doi.org/10.1017/9781009157896, 2021.

Karanikolas, A., Kouremeti, N., Gröbner, J., Egli, L., and Kazadzis, S.: Sensitivity of aerosol optical depth trends using long-term measurements of different sun photometers, Atmos. Meas. Tech.Atmospheric Measurement Techniques, 15, 5667–5680, https://doi.org/10.5194/amt-15-5667-2022, 2022.

Kazadzis, S., Kouremeti, N., Diémoz, H., Gröbner, J., Forgan, B. W., Campanelli, M., Estellés, V., Lantz, K., Michalsky, J., Carlund, T., Cuevas, E., Toledano, C., Becker, R., Nyeki, S., Kosmopoulos, P. G., Tatsiankou, V., Vuilleumier, L., Denn, F. M., Ohkawara, N., Ijima, O., Goloub, P., Raptis, P. I., Milner, M., Behrens, K., Barreto, A., Martucci, G., Hall, E., Wendell, J., Fabbri, B. E., and Wehrli, C.: Results from the Fourth WMO Filter Radiometer Comparison for aerosol optical depth measurements, Atmos. Chem. Phys.Atmospheric Chemistry and Physics, 18, 3185–3201, https://doi.org/10.5194/acp-18-3185-2018, 2018a.

Kazadzis, S., Kouremeti, N., Nyeki, S., Gröbner, J., and Wehrli, C.: The World Optical Depth Research and Calibration Center (WORCC) quality assurance and quality control of GAW-PFR AOD measurements, Geoscientific Instrumentation, Methods and Data SystemsGeosci. Instrum. Meth., 7, 39–53, https://doi.org/10.5194/gi-7-39-2018, 2018b.

Kazadzis, S., Kouremeti, N., and. Gröbner, J.: Fifth WMO Filter Radiometer Comparison (FRC-V) 27 September to 25 October 2021, Davos, Switzerland, WMO GAW report 280, https://library.wmo.int/records/item/66263-fifth-wmo-filter-radiometer-comparison-frc-v?offset=5, (last access: 1 February 2024), 2023.

Kouremeti, N., Nevas, S., Kazadzis, S., Gröbner, J., Schneider, P., & Schwind, K. M.: SI-traceable solar irradiance measurements for aerosol optical depth retrieval, Metrologia, 59(4), 044001, 10.1088/1681-7575/ac6cbb, 2022.

Lucht, W., & Roujean, J. L.: Consideration in parametric modelling of BRDF and albedo from multi-angular satellite sensors observations., Remote Sensing Reviews, 18, 343-379, https://doi.org/10.1080/02757250009532395, 2000.

Maloney, C., Toon, B., Bardeen, C., Yu, P., Froyd, K., Kay, J., and Woods, S.: The Balance Between Heterogeneous and Homogeneous Nucleation of Ice Clouds Using CAM5/CARMA, ~~Journal of Geophysical Research: Atmospheres~~J. Geophys. Res.-Atmos., 127, e2021JD035540, https://doi.org/10.1029/2021JD035540, 2022.

Mazzola, M., , Stone, R.S., Herber, A., Tomasi, C., Lupi, A., Vitale V., Lanconelli, C., Toledano, C., Cachorro V.E., O'Neill, N.T., Shiobara, M., Aaltonen, V., Stebel, K., Zielinski, T., Petelski, T., Ortiz de Galisteo, J.P., Torres, B., Berjon, A., Goloub, P., Li, Z., Blarel, L., Abboud, I., Cuevas, E., Stock, M., Schulz, K., H., Virkkul, A.; ~~"~~Evaluation of sun photometer capabilities for retrievals of aerosol optical depth at high latitudes: The POLAR-AOD intercomparison campaigns,~~."~~ ~~Atmospheric environment~~Atmos. Environ., 52, 4-17, 2012.

Mitchell, R. M.; Forgan, B. W.: Aerosol Measurement in the Australian Outback: Intercomparison of Sun Photometers. J. Atmos. Ocean. Tech.~~Journal of Atmospheric and Oceanic Technolog~~, 20 (1), 54–66. https://doi.org/10.1175/1520-0426(2003)020<0054:AMITAO>2.0.CO;2, 2003.

Nakajima, T., Campanelli, M., Che, H., Estellés, V., Irie, H., Kim, S.-W., Kim, J., Liu, D., Nishizawa, T., Pandithurai, G., Soni, V. K., Thana, B., Tugjsurn, N.-U., Aoki, K., Go, S., Hashimoto, M., Higurashi, A., Kazadzis, S., Khatri, P., Kouremeti, N., Kudo, R., Marenco, F., Momoi, M., Ningombam, S. S., Ryder, C. L., Uchiyama, A., and Yamazaki, A.: An overview of and issues with sky radiometer technology and SKYNET, ~~Atmospheric~~ Atmos. ~~Measurement~~ Meas. ~~Techniques~~Tech., 13, 4195–4218, https://doi.org/10.5194/amt-13-4195-2020, 2020.

Nyeki, S., Gröbner, J., Wehrli, C.: Ground-based aerosol optical depth inter-comparison campaigns at European EUSAAR super-sites, AIP Conference Proceedings. Vol. 1531. No. 1. American Institute of Physics, 2013.

Sinyuk, A., Holben, B. N., Eck, T. F., Giles, D. M., Slutsker, I., Korkin, S., Schafer, J. S., Smirnov, A., Sorokin, M., and Lyapustin, A.: The AERONET Version 3 aerosol retrieval algorithm, associated uncertainties and comparisons to Version 2, Atmos. Meas. Tech., 13, 3375–3411, https://doi.org/10.5194/amt-13-3375-2020, 2020.

Smirnov, A., Holben, B., N., Eck, T., F., Dubovik, O., Slutsker, I.: Cloud-screening and quality control algorithms for the AERONET database, ~~Remote sensing of environment~~Remote Sens. Environ., 73.3, 337-349, 2000

Sun, Q., Wang, Z., Li, Z., Erb, A., and Schaaf, C. L. B.: Evaluation of the global MODIS 30 arc-second spatially and temporally complete snow-free land surface albedo and reflectance anisotropy dataset,~~.~~ ~~Intl. Journal of Applied Earth Observation & Geoinformation~~Int. J. Appl. Earth Obs., 58, 36-49, 10.1016/j.jag.2017.01.011, 2017.

Papachristopoulou, K., Fountoulakis, I., Bais, A. F., Psiloglou, B. E., Papadimitriou, N., Raptis, I.-P., Kazantzidis, A., Kontoes, C., Hatzaki, M., and Kazadzis, S.: Effects of clouds and aerosols on downwelling surface solar irradiance nowcasting and sort-term forecasting, ~~Atmospheric~~ Atmos. ~~Measurement~~ Meas. ~~Techniques~~ Tech. Discussions, 1–31, https://doi.org/10.5194/amt-2023-110, 2023.

Shaw, G. E., Reagan, J. A., and Herman, B. M.: Investigations of Atmospheric Extinction Using Direct Solar Radiation Measurements Made with a Multiple Wavelength Radiometer, ~~Journal of Applied Meteorology and Climatology~~J. Appl. Meteorol. Clim., 12, 374–380, https://doi.org/10.1175/1520-0450(1973)012<0374:IOAEUD>2.0.CO;2, 1973.

Shaw, G. E.: Sun photometry, Bull. Am. Meteorol. Soc., 64, 4–10, 1983.

Tanaka, M., Nakajima, T., and Shiobara, M.: Calibration of a sun-photometer by simultaneous measurements of direct-solar and circumsolar radiations, Appl. ~~Optics~~Optics, 25, 1170–1176, 1986.

Toledano, C., González, R., Fuertes, D., Cuevas, E., Eck, T. F., Kazadzis, S., Kouremeti, N., Gröbner, J., Goloub, P., Blarel, L., Román, R., Barreto, Á., Berjón, A., Holben, B. N., and Cachorro, V. E.: Assessment of Sun photometer Langley calibration at the high-elevation sites Mauna Loa and Izaña, Atmos. Chem. Phys., 18, 14555–14567, https://doi.org/10.5194/acp-18-14555-2018, 2018.

Wehrli, C.: Calibrations of filter radiometers for determination of atmospheric optical depth, Metrologia, 37, 419, https://doi.org/10.1088/0026-1394/37/5/16, 2000.

Wild, M.: Enlightening Global Dimming and Brightening, ~~Bulletin of the American Meteorological Society~~B. Am. Metereol. Soc., 93, 27–37, https://doi.org/10.1175/BAMS-D-11-00074.1, 2012.

Winkler, P. M. and Wagner, P. E.: Characterization techniques for heterogeneous nucleation from the gas phase, ~~Journal of Aerosol Science~~J. Aerosol. Sci., 159, 105875, https://doi.org/10.1016/j.jaerosci.2021.105875, 2022.

WMO: Aerosol measurement procedures, guidelines and recommendations, GAW Report 153, WMO/TD-No 1178, https://library.wmo.int/opac/index.php?lvl=notice_display&id=11085#.WpqIOOdG1PY (last access: 4 October 2022), 2003.

WMO/GAW: Experts workshop on a global surface-based network for long term observations of column aerosol optical properties, Davos 2004, GAW Report 162, WMO/TD-No 1287, edited by: Baltensperger, U., Barries, L., and Wehrli, C., https://library.wmo.int/opac/index.php?lvl=notice_display&id=11094{#}.WpqIledG1PY (last access: 4 October 2022), 2005.