Reviewer3.

This paper presents a synergistic approach for aerosol property retrieval from multiple satellites with the GRASP algorithm, called SYREMIS/GRASP. The intent is to combine the different types of information content into a coherent product that merges both LEO and GEO observations. This is a laudable goal and exists within a framework of the GRASP algorithm which has been developing this capability.

I do unfortunately have significant concerns about the fundamental approach of SYREMIS/GRASP, specifically the lack of direct accounting for measurement and model uncertainty, and the ad-hoc basis for the smoothness criteria. Furthermore, the approach was not explained in sufficient detail to be reproducible. I found myself struggling to understand how the 'weighting' parameters were derived, and what exactly was performed during retrieval setup optimization.

I do not believe the manuscript successfully makes the case that the figures and other results support the conclusions. Often the analysis and figures are poorly conceived, such as inappropriate histogram bin widths in figure 4, overuse of scatterplots which do not clearly indicate comparison skill, and statistical metrics that are calculated without analysis of what those values mean. The number of figures diminishes the impact. I counted 130 panels among 19 figures. In a peer-reviewed publication only the salient points should be reported. I think often figures were included without considering if they represent an appropriate analysis given the amount of data or other matters (such as panels e and f in figure 6).

Response:

The criticism of the review was accounted for in the revised manuscript. In general, we emphasized that the SYREMIS/GRASP approach is based on fundamental principles of the GRASP algorithm. Clear references to the fundamental basis of the GRASP algorithm were provided for readers who are not familiar with the LSM multi-term concept. We also provided a discussion on how this fundamental basis is used in the synergetic approach. In particular, the discussion of the measurements preparation for the synergy, forward models, instrument "weighting", and a priori constraints in the SYREMIS/GRASP approach was introduced in Sections 2.1-2.4. Section 2.4 describes in detail the tests that were used to test and select the optimized approach.

In this manuscript, we provide comprehensive considerations of the SYREMIS/GRASP results. In particular:

1. Section 3.1 is devoted to the consideration of the consistencies in aerosol and surface retrieval from the different instruments in the LEO+LEO synergy.
2. Section 3.2 shows the advantages of the LEO+LEO synergy over the single instrument retrieval
3. Section 3.3 considers LEO+GEO synergy
4. Section 3.4 shows preliminary results of the inter-comparison in a global scale.

Such comprehensive consideration requires a considerable number of figures and tables for all considered aerosol parameters (AOD, AE, SSA). We think that the number of figures is adequate for such an analysis. Nevertheless, we agree with the reviewer that more discussion and analysis should be added. Therefore, we have revised the presentation of figures, included tables to summarize the key statistical metrics and considerably extended the discussion and the analysis all over the manuscript.

Then there is the issue of scope and purpose. The conclusion states briefly that the high temporal resolution results could be used for “air quality studies, for monitoring aerosol transport, aerosol-cloud interaction etc.” I found myself wondering why aerosol data assimilation is not used instead. This has been done for years (one quick example: Yumimoto, et al 2016. https://doi.org/10.1002/2016GL069298). The nice thing about assimilation is that it should represent the correlation between parameters well. It is certainly more sophisticated than the selection of smoothness parameters in GRASP, the values of which I find difficult to connect to actual spatial/temporal variability. If there is some other purpose than the sort of studies one might do with a model that assimilates satellite data, it should be described.

Response:

The SYREMIS multi-instrument synergetic retrieval approach converts the merged L1 measurements from multiple satellites into a synergetic L2 aerosol and surface product. As it was presented in “Introduction”, the purpose of the SYREMIS GRASP approach is to improve L2 output both in terms of the extended aerosol/surface properties and better temporal resolution. Certainly, this L2 output can be used for different purposes, including assimilation into models (see for example, Garrigues, S., Remy, S., Chimot, J., et al., Atmos. Chem. Phys., 22, 14657–14692, https://doi.org/10.5194/acp-22-14657-2022, 2022). Nevertheless, this is not the only goal (and even not the main goal) of the satellite retrieval. More information can be extracted from the satellite measurements (physical, chemical, temporal, etc.), the wider range of applications of the satellite L2 product can be: from the monitoring of the state of the atmosphere and surface to the investigation of physical/chemical processes in the atmosphere. As it was described in the “Introduction”, the main product from most satellites now is AOD. In this paper, we provide a synergetic SYREMIS approach that allows overcoming a number of the existing limitations. The approach is based on the GRASP algorithm (Dubovik et al., 2011, 2021a) with already demonstrated possibilities to account for the temporal and spatial relations between aerosol/surface characteristics. The key elements of the GRASP algorithm are smoothness constraints. Compared to assumptions (based on limited observations) used in traditional algorithms such as the Dark Target algorithm (Levy 2016), the GRASP smoothness constraints allow more natural separation of surface and atmospheric signals based on their inherent differences in the spatial, temporal, and spectral variabilities. A recently developed merged satellite aerosol product suite XAERDT (https://amt.copernicus.org/articles/17/5455/2024/) has a similar purpose and is based on the separate processing of six different satellite measurements using the Dark Target algorithm. A manuscript is under preparation, describing the detailed comparison between SYREMIS and XAERDT products. We observed unique advantages from the SYREMIS product, which will be the subject of the studies.

The selection of smoothness parameters in GRASP is based on rigorous retrieval tests (see new Section 2.4) over representative global sites. These tests minimize the observation and model uncertainties in GRASP, and the resulting smoothness constraints possess solid physical meanings related to the spatial/temporal/spectral variabilities of the retrieved parameters.

Finally, grammar in the manuscript needs help. Several times I found myself unsure at to what was intended to be communicated.
The authors of this publication have produced excellent work in the past, and I believe they are able to do so with this manuscript. However, it requires major revision before I think it is ready for publication.

Response:

The manuscript was reviewed, and the grammar was improved.

Specific comments:

Page 1, paragraph 1: Spell out SYREMIS

Response:

SYREMIS (SYnergetic REtrieval from multi-MISsion instruments) is introduced

Page 3, some HARP2 and SPEXone references:

Fu, G., Rietjens, J., Laasner, R., van der Schaaf, L., van Hees, R., Yuan, Z., van Diedenhoven, B., Hannadige, N., Landgraf, J., Smit, M., Knobelspiesse, K., Cairns, B., Gao, M., Franz, B., Werdell, J., and Hasekamp, O.: Aerosol Retrievals From SPEXone on the NASA PACE Mission: First Results and Validation, Geophysical Research Letters, 52(4), e2024GL113525 , https://doi.org/https://doi.org/10.1029/2024GL113525, 2025.

Hasekamp, O. P., Fu, G., Rusli, S. P., Wu, L., Noia, A. D., aan de Brugh, J., Landgraf, J., Smit, J. M., Rietjens, J., and van Amerongen, A.: Aerosol measurements by SPEXone on the NASA PACE mission: expected retrieval capabilities, J. Quant. Spectrosc. Ra., 227, 170 - 184, https://doi.org/https://doi.org/10.1016/j.jqsrt.2019.02.006, 2019.

Werdell, P. J., Franz, B., Poulin, C., Allen, J., Cairns, B., Caplan, S., Cetinić, I., Craig, S., Gao, M., Hasekamp, O., Ibrahim, A., Knobelspiesse, K., Mannino, A., Martins, J. V., McKinna, L., Meister, G., Patt, F., Proctor, C., Rajapakshe, C., Ramos, I. S.,
Rietjens, J., Sayer, A., and Sirk, E.: Life after launch: a snapshot of the first six months of NASA's plankton, aerosol, cloud, ocean ecosystem (PACE) mission. in: Sensors, Systems, and Next-Generation Satellites XXVIII 131920E) SPIE., 2024.

McBride, B. A., Sienkiewicz, N., Xu, X., Puthukkudy, A., Fernandez-Borda, R., and Martins, J. V.: In-flight characterization of the Hyper-Angular Rainbow Polarimeter (HARP2) on the NASA PACE mission. in: Sensors, Systems, and Next-Generation Satellites XXVIII 131920H) SPIE., 2024.

Response:

Many thanks for the suggested references. They were introduced into the manuscript.

Page 3 line 86 – condition ‘v’ says retrieval should be based on an ‘advanced inversion approach’ which is not defined.

Response:

We reformulated what we mean by “advanced inversion approach” as the one that satisfies the conditions (v) and (vi):

(v) the retrieval should be based on the flexible forward models adaptable to the information content of the measurements;

(vi) the retrieval should be able to account for diverse measurements with, possibly, different calibration accuracy, different spectral and spatial resolution;

An advanced approach should also accounts for observation and model uncertainty, which does not seem the case in this paper. I like Maahn et al 2020 because it lays out the reasoning for this, and Povey 2015 and Sayer 2020's take on measurement uncertainty. I do not believe one can honestly combine data synergistically without accounting for measurement uncertainty – how else can a retrieval algorithm reconcile biases or inconsistencies between the measurements? I know you used a 'weighting' parameter, but this doesn't appear to be based upon an understanding of measurement uncertainty (I am a little unsure what was actually done with the weighting parameter, more on that later). Additionally, it is not clear to me if the output product has a prognostic error estimate, which seems like it would be important given the different sources of data.

Maahn, M., Turner, D. D., Löhnert, U., Posselt, D. J., Ebell, K., Mace, G. G., and Comstock, J. M.: Optimal Estimation Retrievals and Their Uncertainties: What Every Atmospheric Scientist Should Know, Bulletin of the American Meteorological Society, 101(9), E1512 - E1523 , https://doi.org/10.1175/BAMS-D-19-0027.1, 2020.

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

Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., Patadia, F., Popp, T., Povey, A. C., Stebel, K., and Witek, M. L.: A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing, Atmos. Meas. Tech., 13(2), 373--404 , https://doi.org/10.5194/amt-13-373-2020, 2020.

Response:

We don't agree with the reviewer regarding his statements about "observation uncertainties, which does not seem the case in this paper". In the SYREMIS/GRASP approach, they are accounted for with the standard deviation of the uncertainties which are associated with the standard deviation of the measurement fit (Dubovik et al., 2011, 2021a) and contribute to the instrument weighting. Therefore, the weights (importance of the measurements) are driven by known (or assumed) standard deviations of uncertainties in each data set: the smaller the standard deviation of measurement fitting is required in the retrieval, the bigger "weight" of such measurements can be set up in the synergy. In the frame of the GRASP Multi-Term LSM concept, the detailed description of the data weighting Multi-Term LSM concept can be found in (Dubovik et al. 2004, Dubovik et al. 2021a), and brief discussion is also provided in Sections 2.2, 2.4, and Appendix A (Eqs.(3A)-(12A)) of the revised manuscript.

Table 1 It would be nice to add the hyperspectral resolution for TROPOMI

Response:

Thank you for pointing this out. The spectral resolutions are added in Table 1.

Page 7 section 2.1. It is a little unclear what exactly is being done with spectral 'harmonization'. Is it as simple as just adding all spectral channels to the measurement vector? Or is something more being done? I feel like this step should have radiometric harmonization as well, ie removing biases between measurements.

Response:

Speaking about spectral "harmonization," we assumed the spectral band selection in synergy from different instruments. This is corrected in the text to avoid misunderstanding. Checking for biases is part of the optimization remote sensing test (ii) described in Section 2.4 in the revised manuscript. Here, we apply the GROSAT/GRASP approach, performing AERONET/satellite synergetic retrieval, and compare the derived surface products. Depending on the result of the GROSAT/GRASP test, the decision about instrument "bias" is taken (Section 2.2 and 2.4, Appendix A).

Page 8, line 198-199. The method of weighting is explained as "realized with application of the different requirements on the standard deviation of measurements fitting for the different spectral bands". This seems like an important description of what weighting is, but I don't understand the language. "Standard deviation" is mentioned several times but I have no idea what this is at standard deviation of? My closest guess is that it has something to do with minimizing the difference between observations and AERONET data. If that is the basis for deriving these weights it needs to be described in far more detail, since the specifics of which AERONET data were used could drive your results. Also – what does it mean to 'exchange measurements between weighting groups'? This is poorly explained.

Ultimately, I cannot say with any confidence that I understand how you are weighting the instruments.

Response:

Indeed, the "weighting" refers to the importance of the different sets of measurements or a priori data in the global fitting of all input data by the forward models. In brief, in the frame of Multi-Term LSM (Dubovik et al., 2011, 20021a), the "weights" corresponds to Lagrange multipliers that are defined and the squared ratios of standard deviations of the standard deviation of uncertainties in each data set to the standard deviation of uncertainties in selected ("main) data set (i.e. in the radiances of the first satellite). Therefore, determining or assuming the standard deviation of uncertainties in each data set allows for defining the importance of the considered measurements and a priori data. The quantitative description of the "weighting" is described in Sections 2.2, 2.4, and Appendix A of the revised manuscript with reference to (Dubovik et al., 2011, 20021a) for more details..

Page 9, table 3 (and text). In some cases, you defined the temporal threshold in terms of hours, to which I presume means the associated parameter is held constant in that time period. Does this mean that beyond the time period there is no constraint at all? I am also attempting to reconcile this with the numerical smoothness constraints which are also provided. Additionally, I struggle to connect those values with physical reality – where do they come from? How do you justify the choice of values in the 'relaxed' case? Shouldn't these be based on some analysis of aerosol

temporal and spatial variability, such as Alexandrov et al 2004 or Shinozuka et al 2010 (or something more recent).
Alexandrov, M. D., Marshak, A., Cairns, B., Lacis, A. A., and Carlson, B. E.: Scaling Properties of Aerosol Optical Thickness Retrieved from Ground-Based Measurements, J. Atmos. Sci., 61(9), 1024--1039 , 2004.
Shinozuka, Y., Redemann, J., Livingston, J., Russell, P., Clarke, A., Howell, S., Freitag, S., O'Neill, N., Reid, E., Johnson, R., and others: Airborne observation of aerosol optical depth during ARCTAS: vertical profiles, inter-comparison, fine-mode fraction and horizontal variability, Atmos. Chem. Phys. Discuss., 10, 18315-18363 , 2010.

Response:

The developed retrieval relies on a priori constraints of the temporal variability of retrieved aerosol and surface reflectance parameters. These constraints are introduced as an a priori estimate of first derivatives (approximated by the finite differences) with respect to time. The means of these estimates were assumed as normally distributed values with zero means and standard deviations assigned based on the known variability of each parameter. Actually, the use of the limitations on finite differences when the limits are applied to differences between the parameters divided by the time period when they were observed is a very convenient way to ensure flexibility in applying time constants. Indeed, such constraints allow much larger variability for parameters corresponding to more distant in time observations compared to the parameters from observations that are very close in time. However, in cases when observations are nearly simultaneous or very close in time, the finite differences may have very large values that can produce an imbalance in practical fitting. To avoid such difficulties, we used the temporal threshold, which considerably constrains the temporal variability within a specified period of time (threshold). Full description can be found in (Dubovik et al., 2011, 20021a). It should be noted that in Section 2.4, we emphasize that the values provided in Table 4 are suggested based on rather general considerations of how each parameter can vary in time (how high the temporal derivatives can be), and in practice, the exact choice of the parameters is usually defined based on sensitivity studies and retrieval trials with real data. Moreover, these values are not really unique since quite similar retrieval performance of aerosol and surface properties can be obtained within a certain range of the constraints. Nevertheless, the chosen parameters ("weights") are generally within an optimal range in the sense that they adequately reflect the tendencies in temporal dependencies of aerosol and surface properties adapted to the information content in the LEO+LEO and LEO+GEO synergies. For example, in the case of LEO+LEO synergies, the time constraints on the retrieved aerosol properties are nearly irrelevant, while for the case of LEO+GEO synergies, such constraints are very efficient and need to be carefully defined.

Figures 2 and 3 (although these comments apply in a similar nature to many other figures). What is one, in a broad sense, supposed to understand from these six plots? The text says 'one can observe essential improvement' from them. I strongly disagree. All six look very very similar. Perhaps the numerical statistical metrics written on each plot, but these are barely described. Which metric should we use? What specifically has improved from one plot to the other?

Response:

Thank you for pointing out this issue. We have comprehensively revised the presentation of figures in the manuscript, and include tables to summarize the key statistical metrics for intercomparison. A table (Table A1) is used to compare Figures 2 and 3 (before and after harmonization). For example, the fulfilment of GCOS requirement is improved from 34% to 48%, and bias for low AOD improved from +0.06 to +0.01.

Table A1. Summary of all-instrument AOD accuracy statistics for the two tests ("Before", "After") in Figures 2 and 3.

	GCOS(%)	Bias (AOD)	Bias (AOD<0.2)	R	RMSE	slope	intercept
Figxx(a) ("Before")	34%	0.04	0.06	0.89	0.16	0.88	0.08
Figxx(b) ("After")	48%	-0.02	0.01	0.90	0.14	0.88	0.02

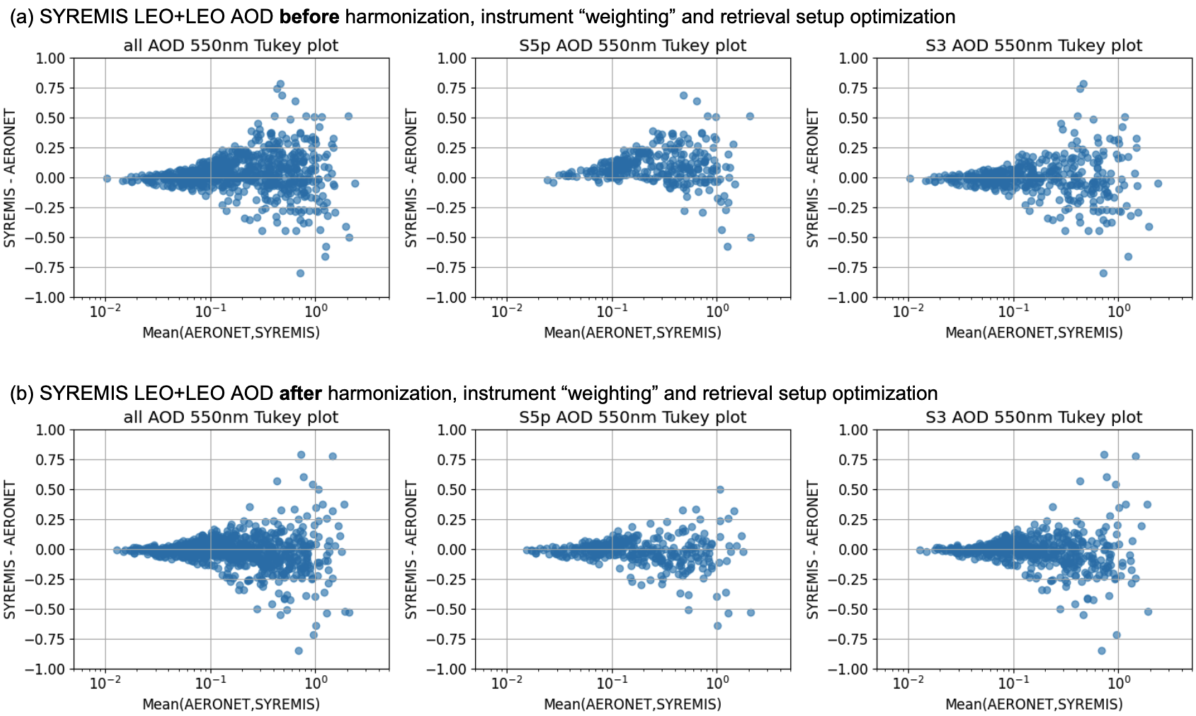
I realize that many algorithm developers in our community use scatterplots such as these to illustrate the success (or otherwise) of a given algorithm. The truth is that they are not appropriate, and figures 2 and 3 are a very good example of why. For starters, you are representing a parameter which is lognormally distributed on axes that are not, and the maximum value of the range is far larger than the majority of the data. So, you have most of the data represented in a tiny corner of the plot. It is impossible to see differences.

Furthermore you have plotted a linear regression to the data (why?) and there is an unexplained grey shaded areas which I presume are GCOS boundaries. The parameters of the linear fit, as well as the R2 value, are meaningless to explain what you are attempting to show, which is how well the GRASP retrieved AOD can represent the AERONET AOD.

Here's how I would do this: use a mean bias plot (also known as a Tukey or Bland-Altman plot). Consider the data as pairs of corresponding GRASP and AERONET AOD. On the x-axis, plot the mean of each pair (AOD_grasp + AOD_aeronet)/2. Use a log scale for this axis. For the y-axis, plot the bias AOD_grasp-AOD_aeronet. Use a linear scale for this axis. This will expand the plotted area of interest and make it clear if there is a bias or any scale dependence. The y axis scatter will express differences in the unit that matters. Among the statistic metrics, I think the percentage fitting within GCOS thresholds is best (since they scale with AOD), but this should be explained, including with what you expect the values to be.

Response:

Thanks for the suggestions! Yes, the gray shaded area in the figure represents the GCOS uncertainty boundaries. We have added this clarification to the figure caption for better understanding. To illustrate the scale dependence of mean bias before and after harmonization, instrument weighting, and retrieval setup optimization, the Tukey plot is presented below. Overall, the revised manuscript presents scatter plots and tables that compare the main statistical metrics more clearly than in the initial submission.



Tukey plot corresponding to fig.4. Each panel corresponds to the scatter plot and histogram in the same relative location in Fig.4. In a Tukey plot, the x-axis (log-scaled) is the mean of each SYREMIS AOD and AERONET AOD pair, y-axis is SYREMIS AOD minus AERONET AOD for the same pair of data.

Page 13, paragraph 1 – similar to above: the results are described as 'high quality'. What is your threshold for 'high quality'? Which parameters matter, and what do you expect them to be?

Response:

Thanks very much for the suggestions! In the revised manuscript, we have included tables throughout to present and compare the validation statistics more clearly and systematically. As shown in Table 6, the high-quality AOD retrievals from SYREMIS achieve approximately 54% within the GCOS uncertainty requirement over land, with a correlation coefficient (R) of ~0.89. In contrast, the S5P/TROPOMI single-instrument retrievals show ~48% within GCOS and an R value of ~0.85.

Figures 4-11 – are all these figures necessary? What are we showing with the TROPOMI or OLCI extracts? Can this be demonstrated with less figures? My above comments for the scatterplots apply. The histograms are good, but the bin size should be adjusted for the number of parameters – for example the 'green' high optical depth case is not meaningfully presented, and this applies to many other cases too. Some of the plots don't have enough data to be meaningful (ie fig 6 e and f).

Response:

Thanks very much for the suggestions! The size bin of each histogram has been adjusted in the revised manuscript. And we revised the presentation of figures and used consistent descriptions of the TROPOMI or OLCI extracts from SYREMIS synergy, for example, SYREMIS LEO+LEO: S5p/TROPOMI, SYREMIS LEO+LEO:

S3A/OLCI, in contrast with GRASP single instrument results: GRASP/TROPOMI and GRASP/OLCI.

Note in the updated validation plot above, the number of datapoints N in each plot is different compared to the number of datapoints in the original Figs. 2 and 3 in the first submission of the manuscript; this is because the updated plots were created with updated AERONET Level 2 products. The latest access date to AERONET is 2025 July 18, which is about 2 years after the creation of the original plots in the first version of the manuscript. "