GRASP/3MI Assessment Report – Sensitivity

Version 1.0

Michael Aspetsberger Catalysts GmbH

Oleg Dubovik GRASP SAS

GRASP-3MI-Study_D2_Assessment-Report-Sensitivity_v1 03.07.2019

Table of Contents

1 Introduction

1.1 Purpose and Scope

This document summarizes the results of the sensitivity study for the GRASP/3MI Aerosol retrieval by utilizing PARASOL/POLDER as a proxy and processing its data in a pseudo near-real-time (NRT) fashion. It lists the different configurations used, the setup of the near-real-time processing, and the validation results.

1.2 Reference Documents

RD-1 GRASP-3MI-Study_D1_Assessment-Report, EUM/CO/18/4600002063/TMa, draft version, 27.06.2018

1.3 Background of this Study

In the main project, different configurations of GRASP for 3MI had been tested for computational and documentation performance against a single 3MI synthetic orbit generated by EUMETSAT, cf. RD-1. In a separate contract, which explored the various GRASP configurations for 3MI, the scientific accuracy against the synthetic orbit was explored. It was assumed that differences in the retrieved data and the input to the test data simulation were caused by differences in the models and radiative transfer schemes of GRASP and the test data generation. In a real situation, the simulation inaccuracies were expected to be covered by the signal noise.

To address this concern, and as a way to understand the GRASP retrieval performance on 3MI, in this study the actual PARASOL/POLDER data has been processed, which is an instrument fairly comparable in terms of sensor capabilities. Given that this is actual data, including all the signal noise and calibration inaccuracies, one can expect an accurate picture of the eventual operational scenario. The PARASOL/POLDER data has been processed systematically with the different settings to have a complete understanding of the scientific retrieval quality performance.

2 Processing Setup

2.1 PARASOL Input Data

For the processing the **parasol_l1b.v03.02** dataset has been used.

In an effort to provide a representative dataset, the following test data set has been selected:

- Spatial Coverage: Full Global
- Temporal Coverage: Jun 2008 - Jul 2008 and Dec 2008 - Jan 2009
    - for Case 1 only the June 2008 period has been processed, due to processing resource requirements.

The processing was performed at the native pixel resolution of PARASOL/POLDER, which is approx. 6km nadir.

The following primary parameters were considered as retrieved products and provided for 6 wavelengths:

- Aerosol Optical Depth (AOD), Single Scattering Albedo (SSA), Bidirectional Polarized Surface Reflectance (BPDF),
- at 443, 490, 565, 670, 865, and 1020 nm.

2.2 GRASP Configuration

A total of 6 GRASP configurations have been tested. The following Table 1 gives an overview:

Table 1: Overview of the GRASP settings for each NRT Test Case

	Reference	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Aerosol	Retrieved, detailed model	Retrieved, detailed model	Retrieved, detailed model	Retrieved, optimized model	Retrieved, optimized model	Retrieved, optimized model	Retrieved, detailed model
Surface	Retrieved	Retrieved	Retrieved	Fixed a priori	Retrieved, constrained a priori	Retrieved	Retrieved, constrained a priori
Radiative Transfer optimization	performance	precision	performance	performance	performance	performance	performance
Inversion re-gime	Multi-pixel: time lat lon	Single-pixel	Multi-pixel: lat lon	Multi-pixel: lat lon	Multi-pixel: lat lon	Multi-pixel: lat lon	Multi-pixel: lat lon

Case 1-4 have been benchmarked in the main study [RD-1]. Cases 5 and 6 have been added as addtl. cases in this study given recent learnings and suggestions based on other contracts.

As outlined in the proposal, Case 5 has been introduced, which is a spatial-only multi-pixel retrieval without any surface a-priori, but with optimized aerosol models (i.e. like Case 3 but without any surface a-priori). This setting has no dependency on a climatology.

Additionally, as an extra bonus, Case 6 has been introduced, which is using surface a-priori but no aerosol modeling optimizations (i.e. like Case 2, but with surface a-priori). This setting avoids any aerosol model simplification.

All cases, 1-6, are configured for a near-real-time retrieval of a single orbit or chunk of data.

The reference configuration corresponds to the settings used for a full-mission reprocessing of PARASOL with GRASP in a multi-temporal and multi-spatial setting. The data is publicly distributed dataset via GRASP-Open[1].

The detailed configurations can be found in the appendix.

2.2.1 Adaptation of the 3MI Settings for PARASOL

The 4 cases already tested in the main study and described in [RD-1] have been adapted in the following way for the PARASOL/POLDER data:

- Removal of the additional wavelength that are not present on PARASOL: 410, 1640, 2103.
- Removal of the measurements of Q and U at 440, and 560 that are not present on PARASOL
- Adding intensity channel at 1020 that is not present in 3MI but available on PARASOL

2.3 Aerosol model

The configurations use different approaches for modeling aerosol: *detailed aerosol model* (Cases 1, 2, 3 and 6) where size distribution is modeled as superposition of 5 log-normal bins and real and imaginary part of aerosol refractive index is retrieved at every wavelength and aerosol optimized model (Cases 4, 5) where aerosol is modelled as external mixture of 5 aerosol components. The size distributions and complex refractive indices are assumed a priori and only 5 values of the component concentration are retrieved. Correspondingly, the values of AOD and SSA are modeled as total resulting values of 5 component mixture.

2.4 Surface Climatology

Some of the configurations require a surface climatology which is either used as an a-priori value (Case 4 and 6) or as a fixed surface value (Case 3).

Different options were considered to generate this climatology, both in terms of processing type and period. The following Figure 1 gives an overview.

[1] https://www.grasp-open.com/products/polder-data-release/

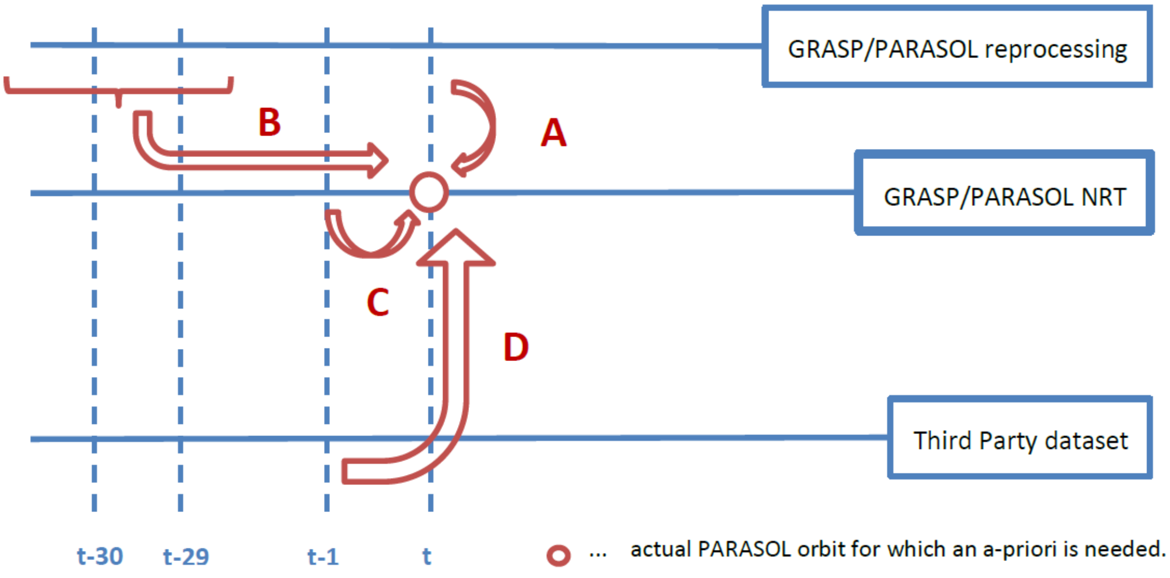

Figure 1: Overview over the various ways to generate a surface climatology.

Following the discussions with EUM in the early stages of the project, strategy B in the figure above has been selected:

- For any given orbit ***o*** in month ***m***, the surface climatology used by ***o*** correspond to the average brdf parameters of month ***m-1*** from multi-temporal and multi-spatial GRASP reprocessing of the same sensor.

E.g. when processing any orbit in June 2008, the climatology would correspond to the average of May 2008 from a processing covering March-May 2008, etc.

Using a same-sensor dataset is essential to avoid having to inter- or extrapolate from different wavelengths, bandwidths, and response functions. The one-month period is suitably long to provide values for the full geographic extent and cloud situations. By taking the climatology from a multi-temporal processing instead of simply averaging the near-real-time, single-orbit products it yields extremely accurate values for the retrieval and ensures the “anchoring” of the retrieval results.

Generating the climatology from a re-processed dataset does add a burden to the overall near-real-time processing setup. However, given that this re-processing is not time critical it can happen as a background processing and it will not have any impact on the near-real-time performance. While not tested in this study, it is expected that a 1-week latency is fine. Furthermore, it has to be reiterated, that the climatology is only relevant for 3 out of the 6 tested cases.

3 Validation

The processed PARASOL data obtained with developed NRT approach were analyzed by comparing with reference data, i.e. data originally generated with established post-processing approach, where PARASOL data were inverted using complete Multi-Pixel concept processing 3 months or more of PARASOL data under constraints on the surface reflectance time variability. Also, the aerosol retrieved from satellite were validated with AEONET data.

3.1 Approach

The evaluation of the obtained results was mainly based on comparisons of the images of the retrieved parameters with the images of the reference retrievals and between the images obtained using different NRT Cases. First, the simple comparisons of maps were done in order to evaluate spatial consistency of the global distribution of obtained results. Second, the point by point correlations of different results were generated and analyzed. This analysis was done for selected days within the considered period.

The above evaluation is set for identifying possible discrepancies and differences between the different approaches. In principle, it is expected that satisfactory quality of the NRT retrieval can be confirmed if NRT retrievals do not exhibit significant differences with the reference retrieval. At the same time, the reference data also have certain accuracy limits and there is also the possibility that some features obtained in NRT retrieval can differ from the reference result but be closer to reality. Therefore, in addition to intercomparisons of satellite retrievals the validation against AERONET data was performed. Specifically, the data retrieved from PARASOL during 4 months (June - July 2008 and December 2008 -January 2009) with NRT approach using Case 2 - 6 were validated against all available AERONET data. It should be noted that NRT retrieval Case 1 was not used for processing of 4 months of the data, nonetheless the accuracy of this approach was evaluated by comparing with AERONET data for the retrieval samples of smaller size. In addition, the results of validation of NRT data were compared with results of PARASOL reference results validation against the same AERONET data.

3.2 Results

3.2.1 Inter-comparison with Reference Data

The detailed comparisons of the products maps were done for all data obtained by NRT retrievals with different settings. Figure 2, Figure 3, and Figure 4 provide the illustrations for AOD, SSA, DHR for last day of the May - June 2008 processing. It is expected that the most differences are to be more clearly manifested in the end of processing since the time flow of the data by NRT processing is the longest.

The general conclusion of multiple comparisons suggests that images produced by different approaches for all parameters are rather consistent and do not show any evident contradiction in spatial distribution of different aerosol parameters. Normally some difference can be noted for the situation where information content about some of atmospheric parameters is limited. For example, in situations with low aerosol loading (small AOD) the retrieval of detailed parameters like SSA and AE (Angstrom Exponent) is very challenging, correspondingly divergence in results obtained by different methods for such situation is quite probable. Similarly, retrieval of all aerosol parameters over very bright surfaces is less certain when over dark waters and green vegetation. On the other hand, for some parameters such as land surface reflectance (see Figure 4) higher image consistency is expected, because land reflectance is often high and dominates satellite signal.

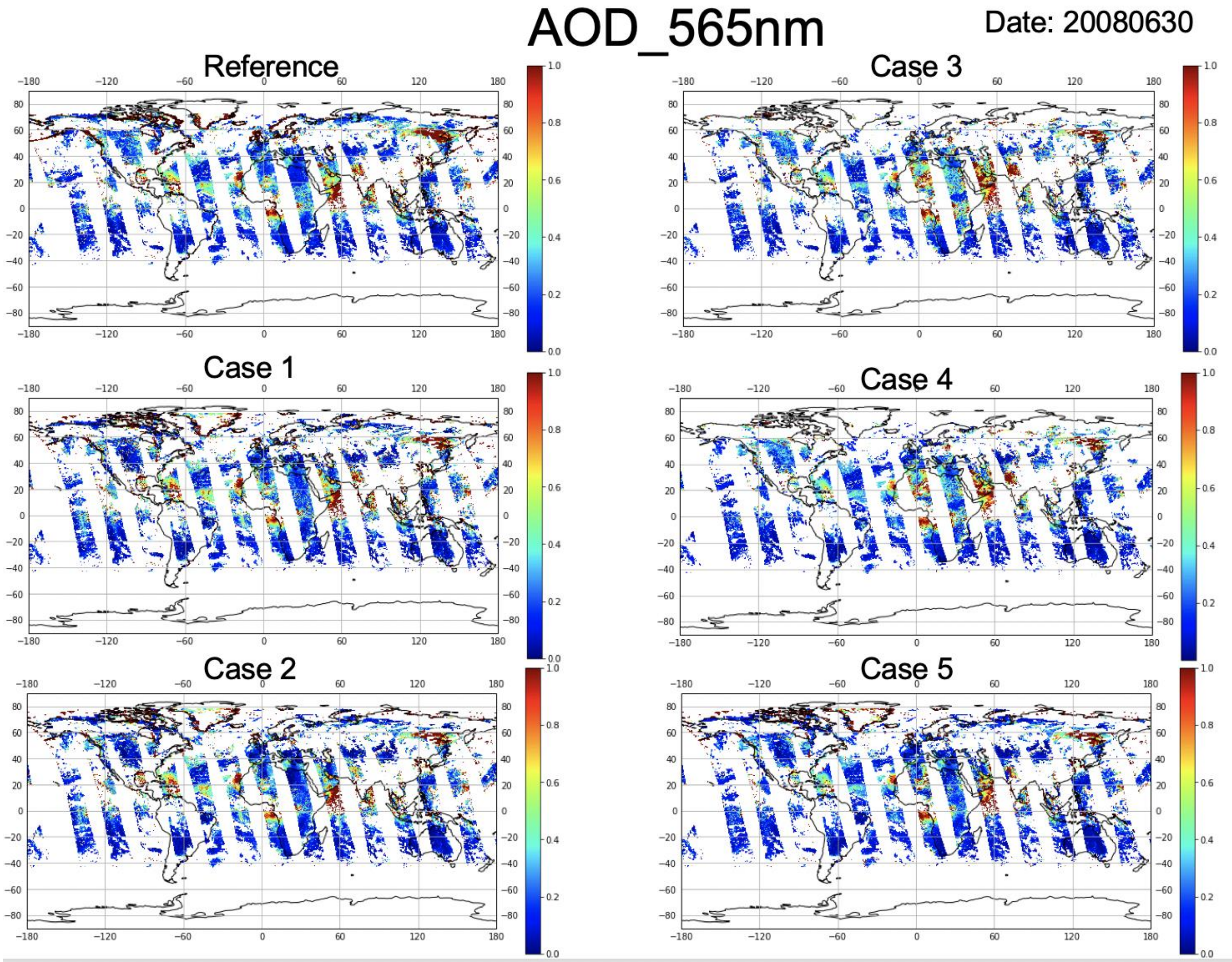

Figure 2: The global AOD(565) retrieved for June 30 2008 using reference POLDER processing and NRT approach using Case-1 - 5 settings.

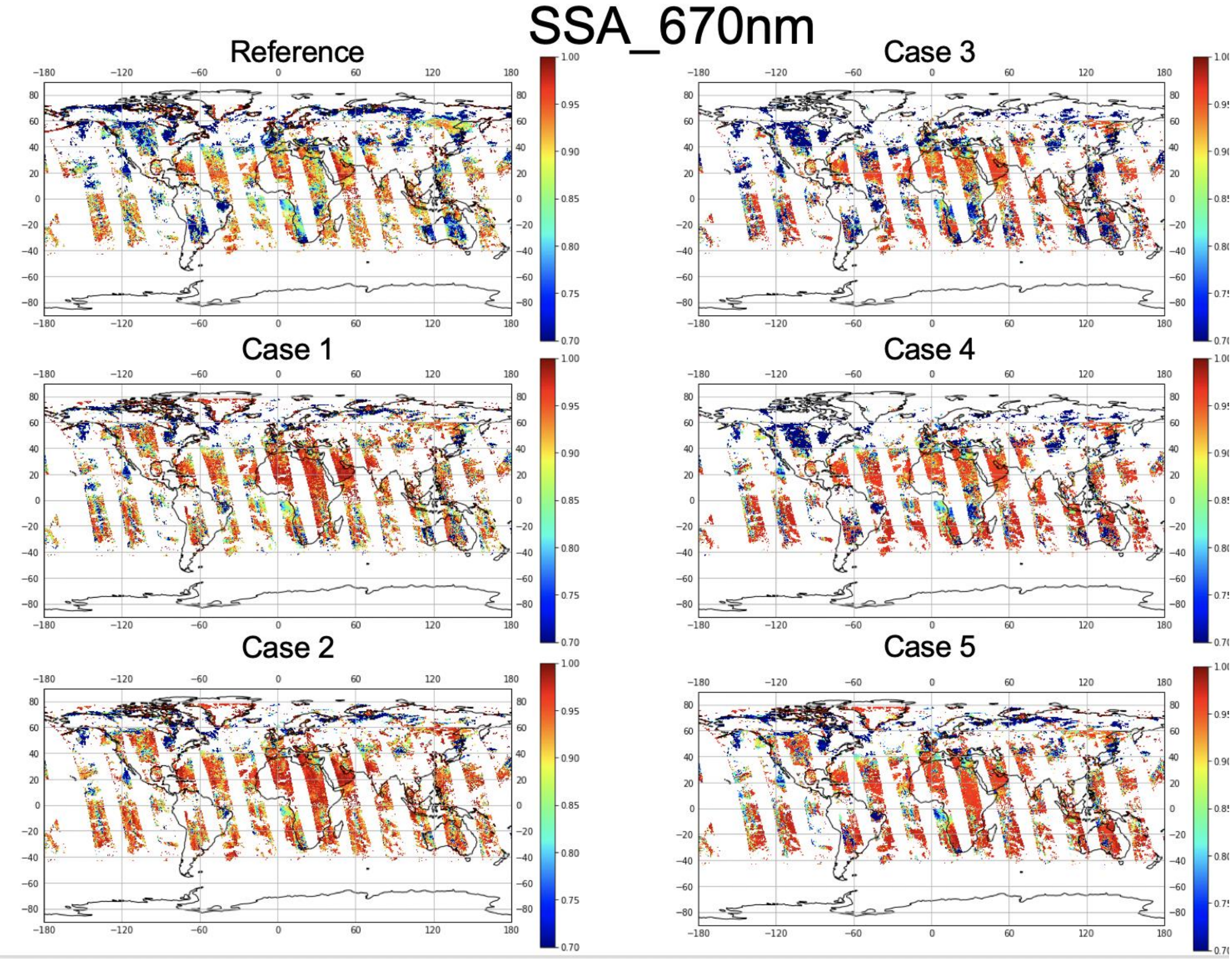

Figure 3: The global SSA(670) retrieved for June 30 2008 using reference POLDER processing and NRT approach using Case-1 - 5 settings.

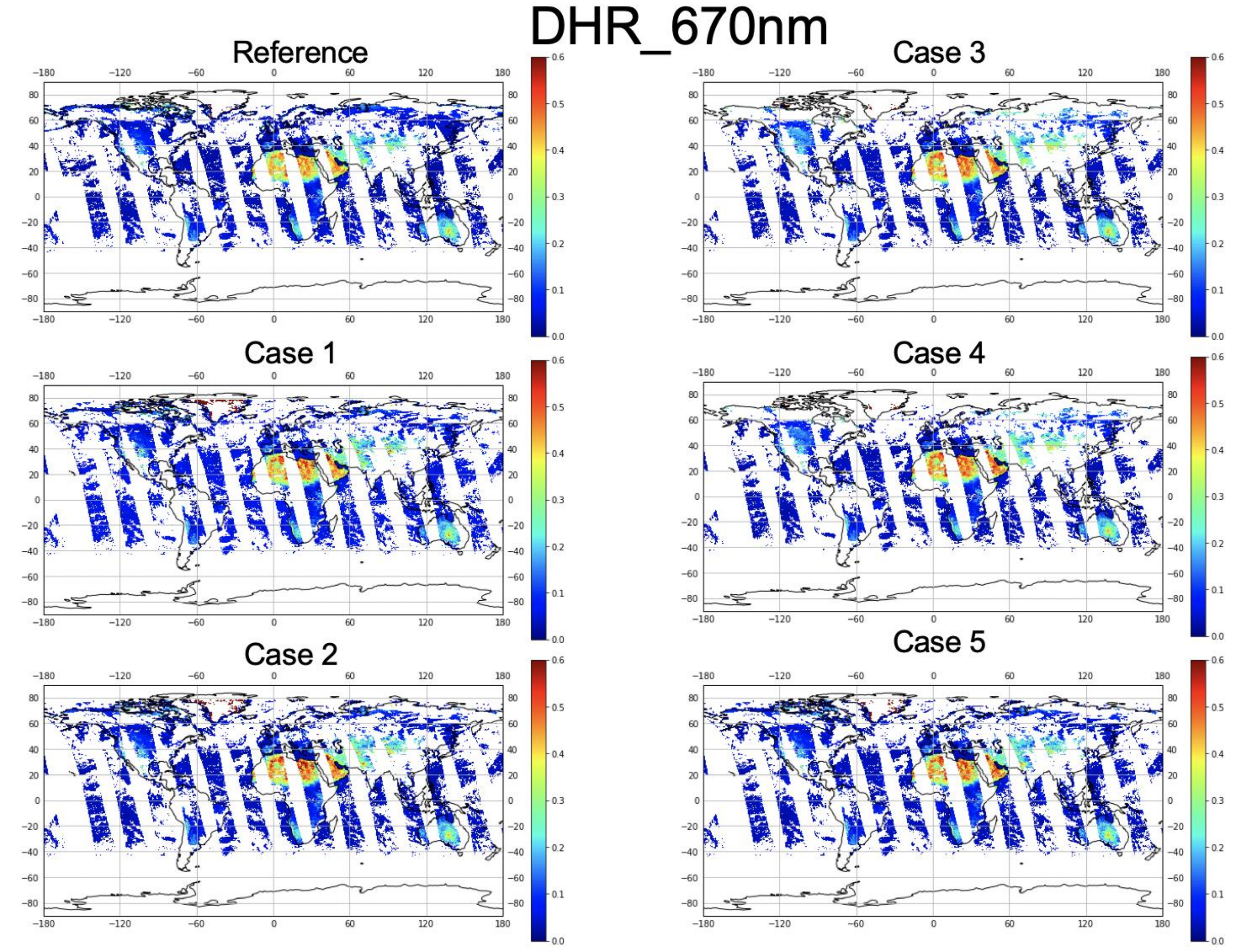

Figure 4: The global Direct Hemispheric Reflectance DHR(670) retrieved for June 30 2008 using reference POLDER processing and NRT approach using Case 1 - 5 settings.

In addition to qualitative comparisons of maps, the point-by-point correlation of images have been done. Figure 5, Figure 6, and Figure 7 provide the illustrations of such correlation for AOD, SSA, DHR for 30 June 2008. The figures illustrate correlations of NRT with Case 1-6 settings against products from reference PARASOL processing. The best correlation seems to be for the Case-6 processing, that was generally a Case recommended by the main project on the 3MI-NRT enhanced algorithm development. This can be explained by the fact that similar aerosol model was used in Case-3 surface reflectance was constrained to a priori values adapted from monthly averages of reference PARASOL processing. At the same time these monthly averages did not contained the surface properties daily variability. This fact also explains smaller correlation of Case-3 with the reference PARASOL processing. Also, smaller correlation of Case 4 and 5 with reference PARASOL processing can be partially explained by the differences in used approaches for modeling aerosol. Apparently, as will be seen in the next section, some features provided by Case 4 and 5 retrieval agree better with AERONET.

It should be noted that different number of points on the correlation plots can be explained by the fact that for selecting the data of high quality the data were filtered by the value of the residual. As a result, the Cases such as Case-3 retrieving less parameters provided the poorer fitting of observations. Therefore, if such processing will be used for generation of the products the criteria of quality screening should be adjusted and optimized.

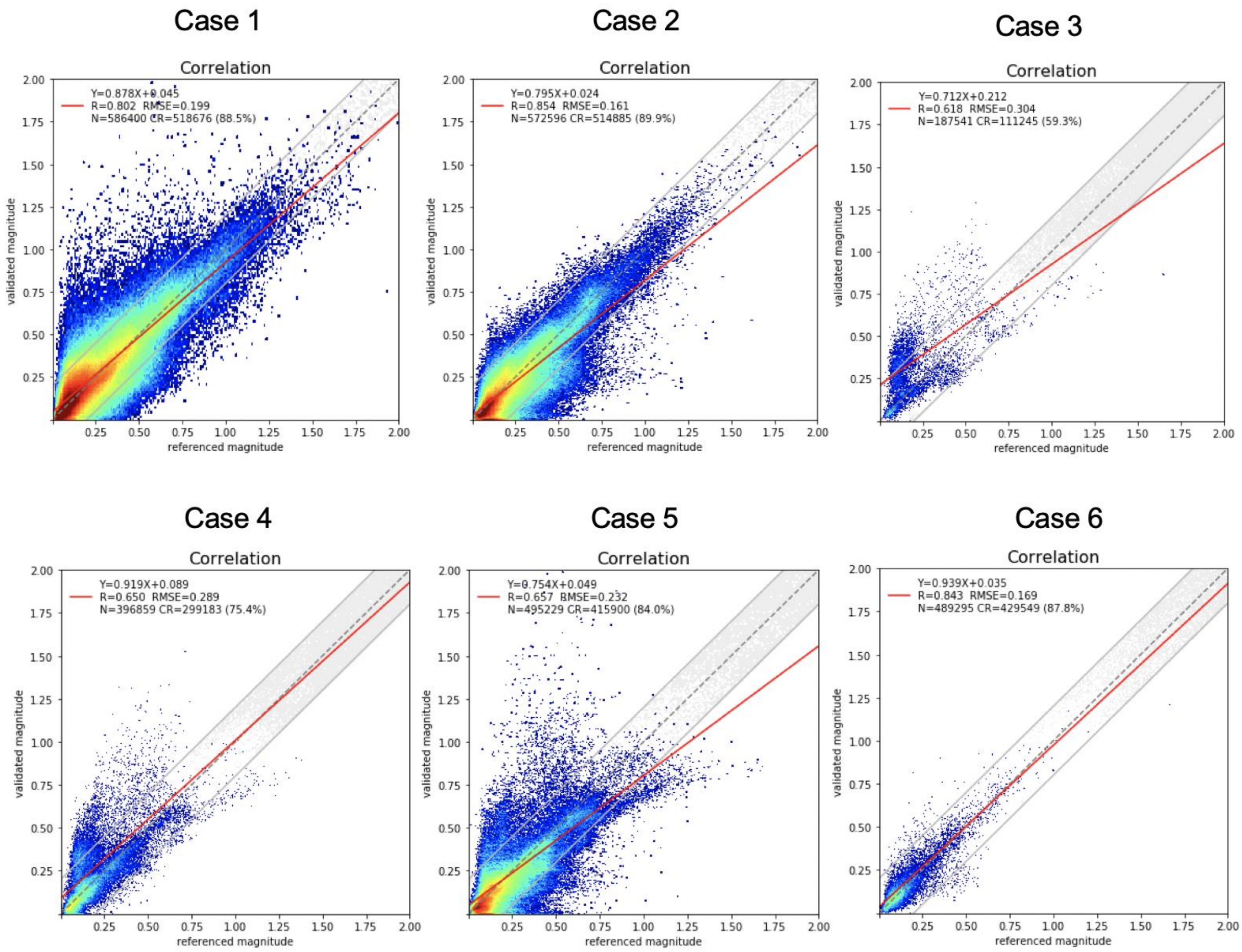

Figure 5: The correlation of AOD (564) retrieved with NRT approach using Case 1 - 6 settings against values obtained with reference POLDER processing

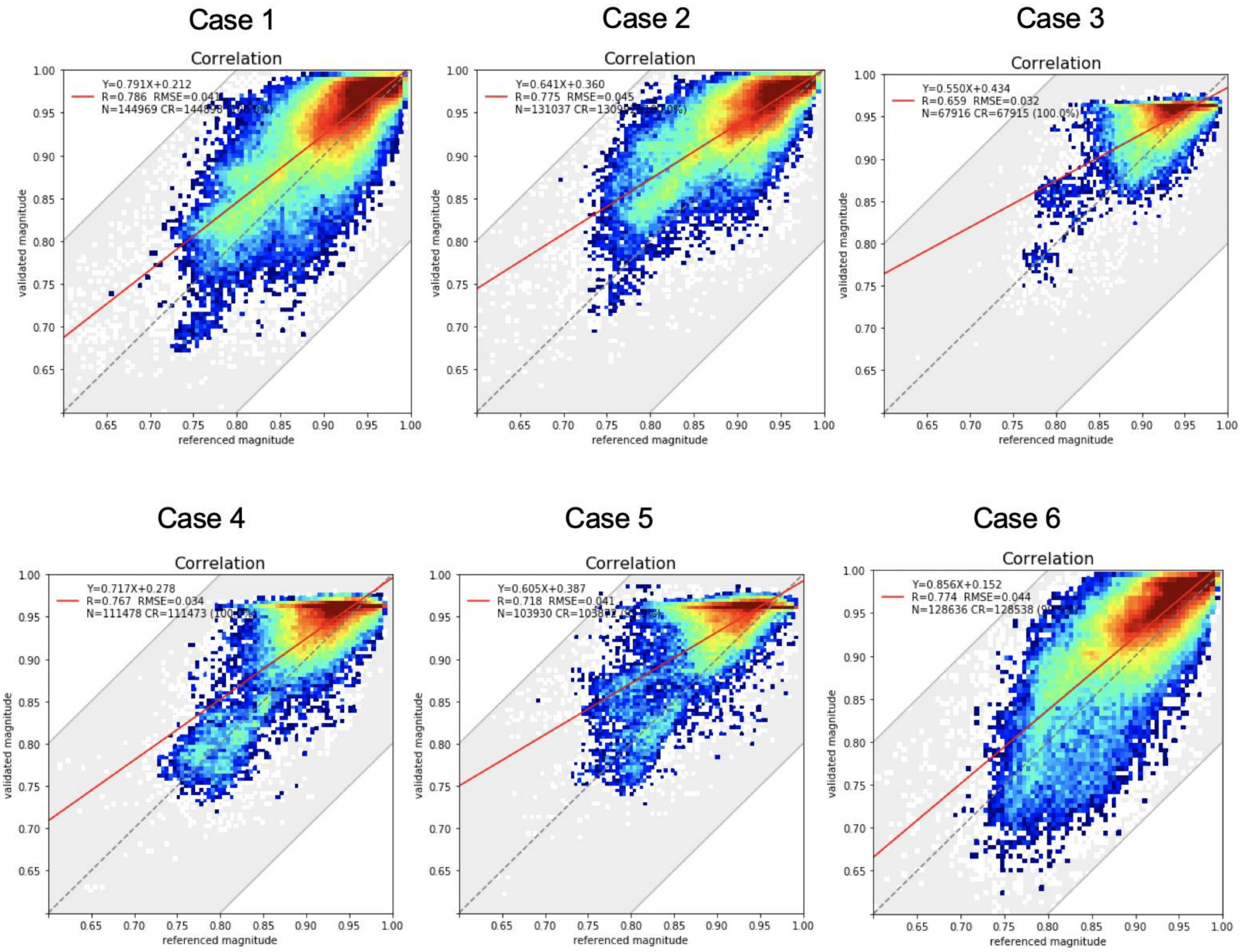

Figure 6: The correlation of SSA (670) retrieved with NRT approach using Case 1 - 6 settings against values obtained with reference POLDER processing

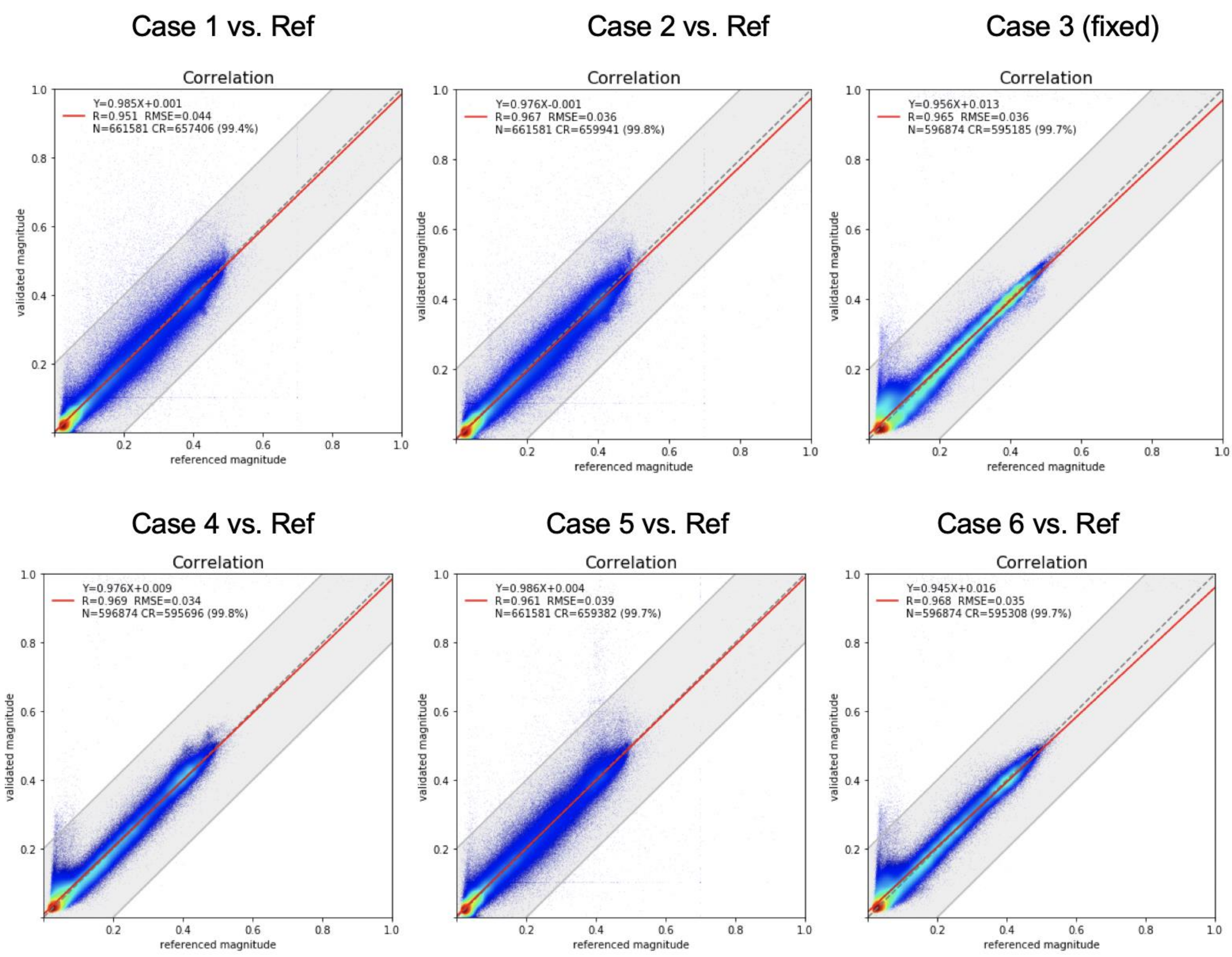

Figure 7: The correlation of DHR (670) retrieved with NRT approach using Case 1 - 6 settings against values obtained with reference POLDER processing

3.2.2 Inter-comparison with AERONET

Figure 8 and Figure 9 show correlation of the data retrieved from PARASOL during 4 months (June - July 2008 and December 2008 -January 2009) with NRT approach using Case 2 - 6 against all available AERONET data. The correlations are shown for AOD and AE. In should be noted that the correlations were also produced for SSA, however the number of available SSA's in AERONET data was too low for convincing illustrations.

As can be seen from correlations plots, in principle, all approaches provide rather robust and reliable results that are comparable with reference processing. The results are especially positive for AOD, where Cases 4-6 are comparable to the correlation of reference data. Moreover, NRT retrieval using Case 4 and 5 settings show somewhat better slope and clearly smaller bias at low AOD (general bias for Cases 4 and 5 is ~ 0.015 while it is ~ 0.1 for the reference case. For AE the results, NRT retrievals are slightly less performant than reference retrieval. At the same time, most of the approaches provide reasonably good correlations and the Case - 4 shows close correlation to the reference case. Moreover, this observation can change for different set of sampled data. For example, if only data of June 2008 used (see Figure 10) the correlation for AE retrieved by some NRT approaches are comparable or even higher than for reference case.

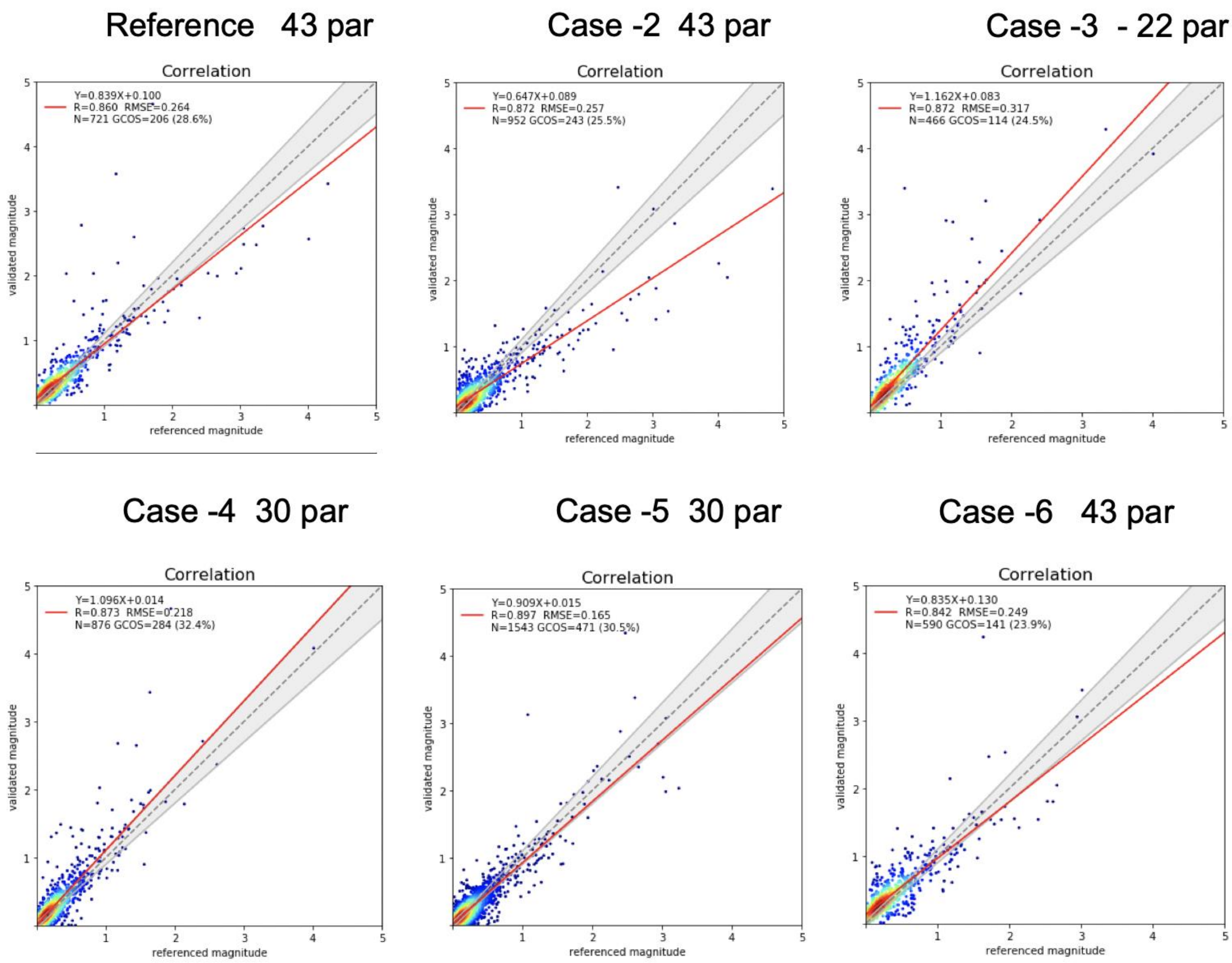

Figure 8: The correlation of AOD (440) retrieved during 4 months (June - July 2008 and December 2008 -January 2009 with reference POLDER processing and with NRT approach using Case 2 - 6 settings against available AERONET data.

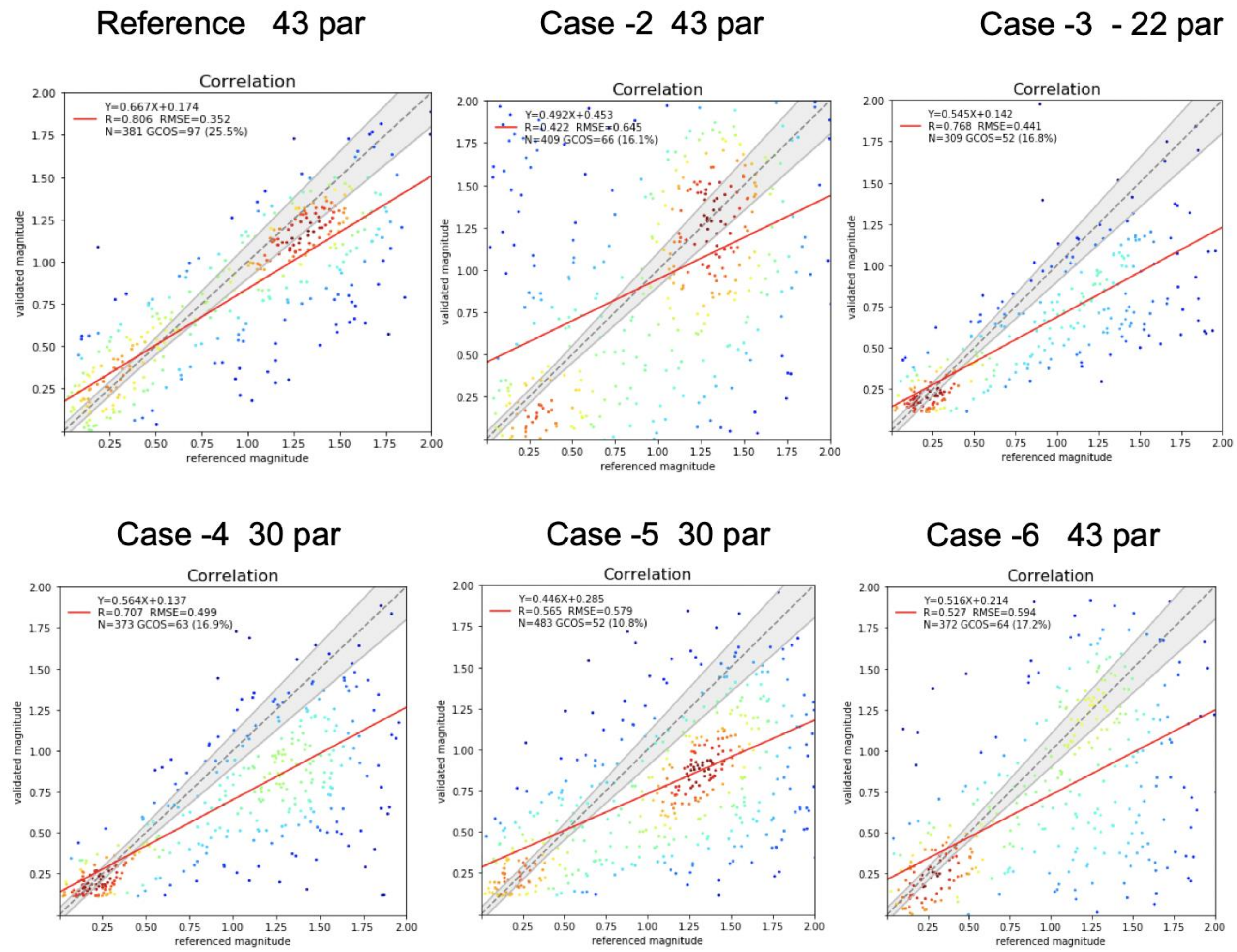

Figure 9: The correlation of Angstrom exponent (440/870) retrieved during 4 months (June - July 2008 and December 2008 -January 2009 with reference POLDER processing and with NRT approach using Case-2 - 6 settings available AERONET data.

As noted above the NRT retrieval Case 1 was not used for processing of 4 months of the data, nonetheless the accuracy of this approach was evaluated by comparing with AERONET data for the retrieval samples of smaller size. The evaluation of NRT retrieval Case 1 was quite important since in Case 1 settings used the highest accuracy in RT calculations. Therefore, it could be expected much higher accuracy for Case -1 results. Figure 10 shows the results of validation of NRT retrieval obtained for 1 month June 2008 with results of PARASOL reference results validation against the same AERONET data. As can be seen from the illustrations, the NRT Case-1 retrievals do not show clear superiority over other NRT retrieval. For example, correlation of NRT Case-4 and 5 retrievals with AERONET show better slope and smaller bias than NRT Case-1 retrievals. Thus, we can conclude that there is no clear advantage of using more precise and much more time-consuming RT calculation than optimized RT calculations used in Cases 2-6. It is also interesting to note that NRT Case-5 retrieval in all plots of correlations show very solid (one of the best) correlation with AERONET for AOD. This is an interesting observation taking into account that in NRT Case-5 no a priori assumptions were used for surface reflectance.

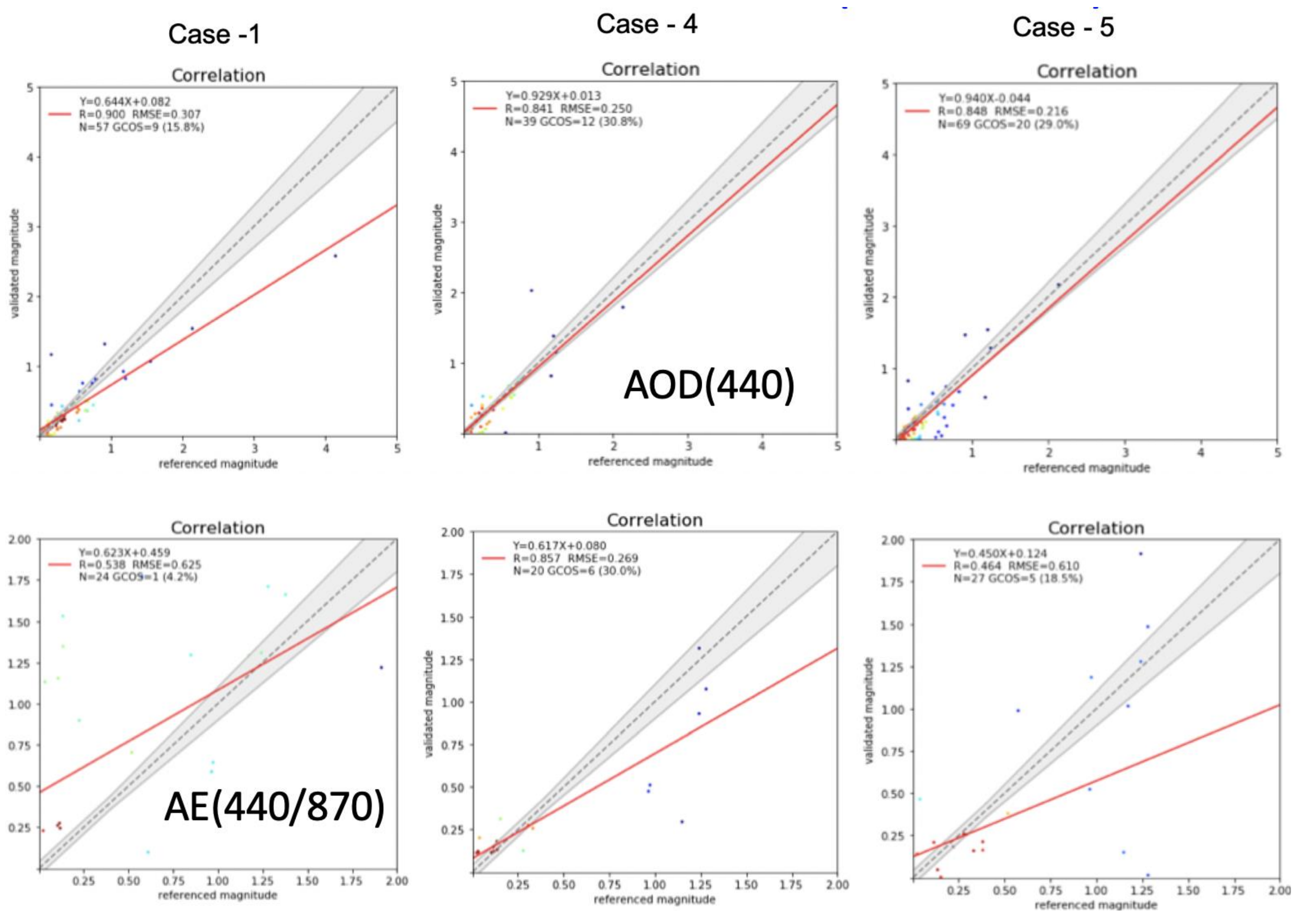

Figure 10: The correlation of AOD(440) - upper panel and Angstrom exponent (440/870) - lower panel retrieved during June 2008 with reference POLDER processing and with NRT approach using Case-4 and 5 settings available AERONET data.

4 Performance Statistics

In the main project, the various settings have been benchmarked extensively to generate performance metrics. The results for those tests are described in [RD-1] and are re-iterated below in Table 2. The processing hardware has been comparable to the one used in the main study[2].

Table 2: The 3MI Processing Statistics based on the benchmarking of the synthetic 3MI test dataset

Test Case	min	max	mean	median	5th quantile	95th quantile
test1	0.814	19.063	3.302	3.069	1.628	5.691
test2	0.053	5.391	0.334	0.350	0.224	0.406
test3	0.022	0.819	0.142	0.145	0.078	0.200
test4	0.034	1.608	0.234	0.234	0.159	0.301

When running the PARASOL/POLDER processing **similar - yet not directly comparable -** timing statistics were collected. While the 3MI pixel timings were compiled only from one individual orbit, they were generated in isolation, with no additional processing ongoing. This was not the case for the PARASOL/POLDER processing: Given the size and resource requirements, this was scheduled as a batch job on the cluster, with other tasks running concurrently and with over-scheduling, to focus on utilization rather than performance. **Hence, while the absolute values for the PARASOL/POLDER cannot be used to make any predictions on the performance, the relative difference between the different cases does have moderate significance.**

PARASOL/POLDER statistics have been compiled for the full processing and for 4 reference orbits:

- 2008_08_28/P3L1TBG1086037MD
- 2008_08_29/P3L1TBG1086051MD
- 2008_08_30/P3L1TBG1086066MD
- 2008_08_31/P3L1TBG1086080MD

Table 3 shows the benchmark statistics for the 4 reference orbits. There are no benchmarks available for case 6 as it had been included only in the final stages of the project. The benchmarks have been done without any intentional overscheduling, yet they were not performed in isolation and the times include the data transfer from the distributed file system.

[2] Intel(R) Xeon(R) CPU E5-2650 v4, 2 sockets with 12 physical cores each, 2.2 GHz, 8x 32 GB DDR4 (total 256 GB) per node, multiple nodes in use

Table 3: The timing statistics for 4 PARASOL/POLDER orbits

Test Case	core-seconds per pixel				Total core-sec per orbit	Total average core-sec per pixel	Minutes per orbit with 100 cores
	min	max	mean	median			
test1	1.08	40.38	4.29	3.54	825,089.12	3.98	137.51
test2	0.13	10.30	0.71	0.54	118,136.65	0.57	19.69
test3	0.11	6.25	0.57	0.40	83,046.89	0.44	13.84
test4	0.10	7.26	0.59	0.42	86,651.19	0.46	14.44
test5	0.11	7.02	0.65	0.47	110,095.83	0.53	18.35

Overall, when comparing the PARASOL/POLDER median statistics to the 3MI median statistics, the relative ranking in terms of processing time is preserved: Case 3 is fastest, followed by 4, (5), and 2, while Case 1 is by far the slowest. The difference amongst Test 2-4 is much smaller than it was in the 3MI tests, but when considering a fixed 0.2 sec overhead per pixel, the numbers are comparable again.

Figure 11 shows a boxplot for the timing statistics from the batch processing of the full test data set. As noted above, given the batch processing nature, and focus on utilization rather than timeliness, the numbers can only be used for relative not absolute assessments. The approx. 2x increase compared to the benchmarking most likely stems from an overscheduling of the processing resources (i.e. hyper-threads vs. cores).

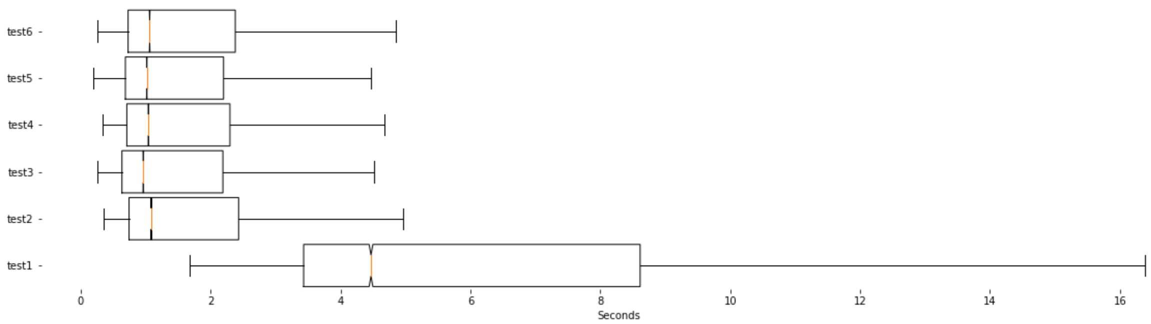

Figure 11: A boxplot for all 6 tests, showing the 1st and 3rd quartile (i.e. 25th and 75th percentile) as box, with the median (50th Percentile) as orange line in the middle, and whiskers to symbolize the 5th and 95th percentile respectively.

Figure 12 shows a density histogram for the various configurations. As shown in the boxplot above, it is directly obvious that Case 1 has not only a much higher time per pixel but also a greater spread. All other cases have fairly comparable times and spread.

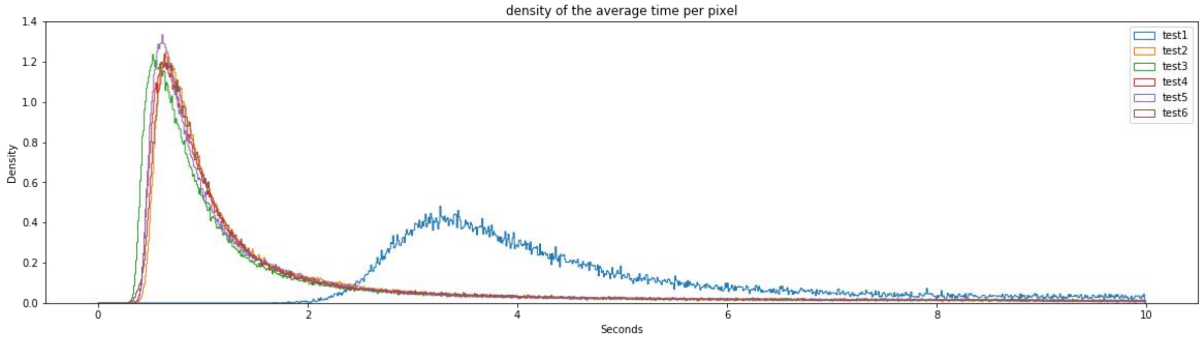

Figure 12: A density histogram showing the average time per pixel for all 6 cases, with case 1 being the clear outlier with a much greater value.

When looking at the performance in detail, one can see a correlation between the processing chunk size and the time per pixel, see Figure 13 and Figure 14: the larger the processing chunk, the smaller the time per pixel generally gets. This can be explained by the constant overhead of loading data and setting up the internal retrieval, which is then added proportionally to a pixel's processing time. Hence, the more pixel in a chunk, the smaller the overhead.

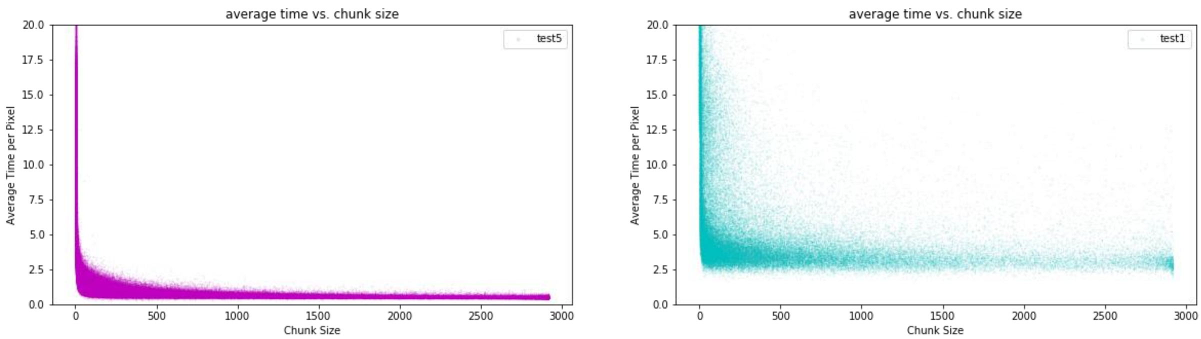

Figure 13: A plot of average time per pixel in a chunk and the actual chunk size for case 5 (left, representative of cases 2-6) and case 1 (right)

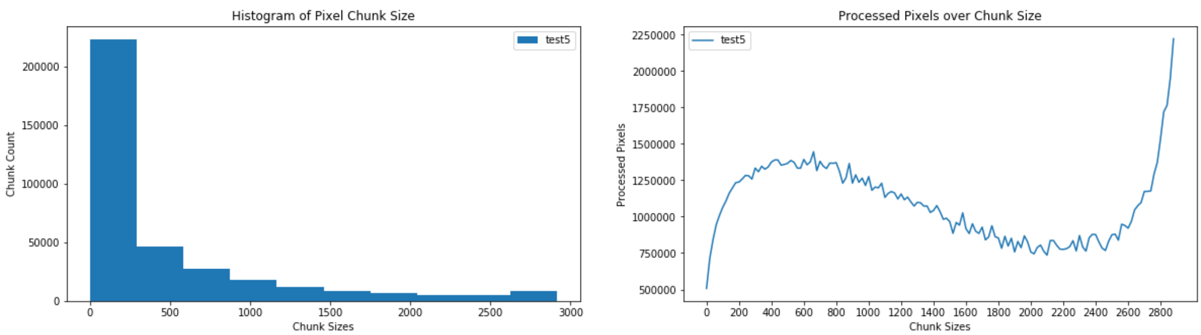

Figure 14: A histogram showing the number of chunk sizes (left) and the number of pixels processed in a given chunk (right)

The reasons for varying chunk sizes are (a) observation characteristics (i.e. simply not covered by the swath), (b) meteorological conditions (cloud cover), and (c) to some extent illumination conditions (solar zenith). The chunks had been selected based on a fixed, rectangular pattern on the PARASOL/POLDER L1c grid, see Figure 15, which is a heritage of the typical multi-temporal processing scheme. Given that the temporal layer for a near-real-time scenario is not necessary, one could chunk on the actual pixel grid, rather than the fixed L1c grid, and, therefore, have larger processing chunks due to avoiding (a).

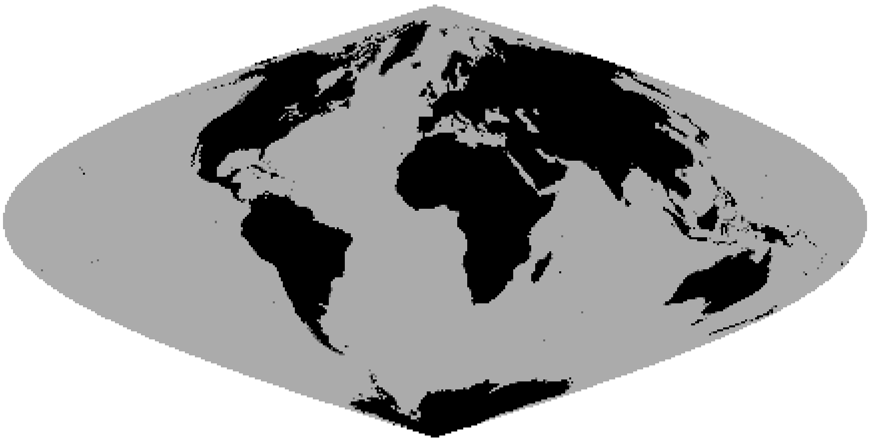

Figure 15: The PARASOL/POLDER data is on a fixed sinusoidal grid[3].

While the learning from this processing here cannot be applied directly to 3MI, due to different processing modes (PARASOL/POLDER was on a batch basis, whereas 3MI will be stream processing) it is in anyway suggested to have at least chunks with 100 pixels in size.

Overall, the following Table 4 shows the overall processing statistics for the full test dataset and all configurations:

Table 4: The re-processing statistics for the full PARASOL/POLDER dataset.

	median per case	**total over all cases**
Total re-processing time in core-sec	99,961,741.28	630,021,827.74
Total pixels re-processed	130,314,298	751,074,928
Output product size	109 GB per month	2410 GB
Total processing wall-time	approx. 2 days	approx. 1 month

[3] Parasol Level-1 Product Data Format and User Manual, F.-M. Bréon et al., Ed. 1 - Rev. 4, Dec 13th , 2016, http://www.icare.univ-lille1.fr/projects_data/parasol/docs/Parasol_Level-1_format_latest.pdf

5 Conclusion

Based on the efforts done within project and described above, the following conclusions are drawn:

- All NRT retrieval settings produce feasible results in terms of accuracy while in terms of performance some important differences are evident.
- Globally, Cases 4 and 6 can be recommended as the best setting for 3MI NRT operational processing.
- At the same time, Case 5 also shows very solid results without using any a priori constraints for surface reflectance. Therefore Case 5 can also be considered as a candidate for operational 3MI NTR processing.

In terms of quality:

- All NTR retrieval approaches provide comparable images of surface reflectance and main aerosol parameters such as AOD, AE and SSA;
- Case 4 and 6 settings provide the best correlation with available AERONET data that are overall comparable with quality of reference retrievals for AOD and slightly less performant for AE.
- Case 5, which did not rely on any a priori constraints about surface reflectance, also provided solid non-biased correlation

In terms of compute resources:

- The processing of PARASOL/POLDER data in this project has not been setup to provide reliable benchmark metrics. The study has shown the performance in a batch processing scenario, and, furthermore, all timings were performed with a dataset different to 3MI and a processing facility different to the future EPS-SG PDAP. The only reliable timing metric can, therefore, be the relative difference between the various settings.
- All NTR retrieval approaches with Case 2 to 6 settings provide comparable performance with a median time per PARASOL/POLDER polder pixel ranging from 0.4 to 0.5 core-seconds for processing data off the archive. This roughly translate to a 10-20 min processing time of one orbit using 100 cores for computation. Considering a more effective in-memory or at least node-local data processing, the Case 2 to 6 retrieval scenarios are fully acceptable for 3MI NRT processing in terms of performance.
- The Case 1 settings, which use the most accurate RT calculations, require about 10 times more computation time compared to the Case 2 to 6 retrievals. At the same time, Case 4, 5 and 6 retrievals provide the retrieval of comparable quality with Case 1. Thus, Case 1 settings are not recommended for further consideration for 3MI NRT processing.