Dear Dr. Xu,

First of all, we would like to thank you for the positive and very interesting comments that certainly are very helpful to improve the manuscript. Please, find the detailed answers below. The original text of your comments is highlighted in blue color.

By performing truth-in-truth-out simulations, the paper by Herrera et al. did a careful work on assessing GRASP retrieval errors. The topic of study is important and technical routine was well presented. Following an introduction of the error estimate model, examples of error estimate for ground based observation including AERONET and elastic lidar were demonstrated through 40 figures and in-depth analysws in the paper. The main conclusion matches their numerical results and technical approach is well displayed and explained. I have the following comments for the authors to consider:

1. The imposition of a priori constraints (e.g. the smoothness constraints) in GRASP algorithm can very effectively mitigate the subjection of optimization towards locally optimized solutions. However, such a possibility of getting locally optimized solutions still exist when the parameter space gets large and/or the measurement uncertainties increase. In addition, the some modeling assumption of aerosol microphysics (e.g. type, shape, aerosol size components, …) and atmosphere structure (e.g. plane-parallel atmosphere) behind the retrieval can bring in certain levels of errors as well. If I understand correctly the authors' paper, these error sources are not attempted to be analzyed in the present work. So the authors might want add some caveat information to this regard either in the introduction or sumary of the paper.

We fully agree that there many retrieval assumptions as those on aerosol microphysics, atmosphere structure and used a priori constrains, etc. affect the accuracy of retrieval results. This is common situation in all retrieval algorithms. At the same time, those uncertainties belong to so-called systematic errors that do not have random character. Commonly, in the retrieval approaches aimed to provide statistically optimum solution, the systematic errors are considered to be negligible and the optimization is realized only for random error component. This is a case for GRASP algorithm as it was stated in this and other key GRASP paper. Certainly, one can never assure with 100% certainty that all systematic errors are truly negligible. Therefore, usually the used forward model is to be tested extensively. In these regards, our studies are based on the results of quite large number previous studies including Dubovik et al. (2000, 2002, 2006), Dubovik and King (2000), Sinyuk et al. (2007, 2020), Torres et al. (2017), etc. All those studies are resulted in the fact that all main biases were addressed rather adequately (e.g., non-sphericity) and quantified. Therefore, at present we can state that there is a rather solid certainty that there no clear significant systematic errors in forward model of GRASP and AERONET. There is also clear understanding of inevitable presence of operational systematic biases as those related calibration of the instruments. The values of these uncertainties are well known on the quantitative level and, therefore, the effect of these bias is analyzed extensively in the paper. Finally, we realize that some unexpected biases can inevitably appear in the retrieval and there is no guaranteed approach for detecting all biases, except the fact that some biases (and especially strong ones) can manifest themselves via misfit of measurements or misfit of a priori constraints. Therefore, the value of misfit (that is also includes effect of random errors) is rigorously incorporated in the used methodology for estimating the dynamic uncertainties.

2. As illustrated via the box plots (Figs.6 and 13-19), the comparison of GRASP estimate errors of various aerosol properties against the actual errors indicated general agreement, which is impressive ! Interestingly, there are some systematic differences. It is quite commendable that the authors are looking into the impact of correlation of retrieval errors. I'm curious whether there are other possible causes behind these systematic differences. For example: the authors made an interesting observation - "The results of the statistical tests with randomly generated noise showed that GRASP error estimates in most cases are comparable or exceed the actual errors by the 20 to 30% and therefore can be safely used for assuring uncertainties of actual retrieval products." Could the overestimate of the error due to the chance of double counting the effect of random error in both "C_{delta_a_ran}" and "a_{bias}*a_{bias}" terms via Eq.(22) ?

Actually, there is no "double counting" of random errors in Eq.(22), because the second term is zero in case if no systematic errors present. The most probable reason for overestimating the actual errors is the fact that the used equations are defined in linear approximation that tends overestimate the actual errors if inverted functions are strongly non-linear. This is especially true in a case when one retrieved a very large number of parameters, in such situation the actual covariance matrices practically can't not be explicitly derived (in fact, the covariance matrices can be estimated but it is rather tedious and challenging task) and sophisticated solvers are used that not fully transparent, i.e. full control of all uncertainties is hardly possible in practice. It also can be noted, the "actual errors" produced in our studies are generated using solvers that may in some situation diminish the error in synthetic experiments. This is why, we consider our results as satisfactory at present, while we certainly plan to continue the efforts in improving them.

3. As described by the authors in Table 1, the authors assume almucantar geometry in their simulation. I wonder whether there is any dependence of the accuracy of GRASP error estimate model on other observation geometries (e.g. principal-plane scan) ?

In current paper only measurement in the almucantar geometry were analyzed because these are the most popular measurements and the analysis of only these geometries is already quite extensive. Nonetheless, the applicability of the error estimate formalism has general character and the developed formalism should be equally applicable for PPL measurements. On the basis of known experiences, it is possible to expect in general the measurements errors may be a bit higher for retrieved aerosol parameters from PPL measurements. Indeed, based on the result of a detailed comparative analysis of PPL and ALM retrievals performed by Torres et al. (2014) some differences in error tendencies can be expected. For example, the PPL geometry is generally more sensitive to measurements errors (e.g., influence of pointing errors) and also to retrieval assumptions (as aerosol vertical distribution).

4. The abstract and the conclusion of the paper may be shortened by summarizing the main work and finding. For example, the first two paragraphs can be moved to the introduction part of the paper.

In principle we agree that the abstract is somewhat long. However, after making some unsuccessful efforts to reduce it following you suggestion, we have decided to keep the abstract as it is. The main rational for that decision was related to fact that the paper itself is rather

long, and shortening the abstract would make it more difficult for the readers to get sufficient intro to all aspects considered in the paper