The authors thank Referee #2 for useful comments and suggestion to improve the paper. Below we answer to the comments point-by-point. The Referee comments are in **bold**. The pages and lines reported correspond to the manuscript under discussion.

Response to Anonymous Referee #2

This paper presents an analysis of aerosol type retrieval and uncertainty quantification from OMI data. The method is based on Bayesian inference approach. The aerosol types are used in forward calculation of OMI spectra and compare that with the measured one. The differences are then used to create the probability density function to estimate the uncertainty in retrieval of AOD. With this approach, this retrieval and its uncertainty can be assessed in probability terms.
The paper is interesting with sound math. For it to be published and attract wider readability, it needs significant revisions, especially in many places where Maatta et al's paper is referred.

We thank the Referee #2 for reviewing our manuscript.
We agree that we have frequently referred to the paper Määttä et al. (2014) for the theoretical details. In that paper the theoretical background and construction of the method is presented in detail and we did not want to repeat it in this manuscript.

As suggested by the Referee #2, we have now included more description of the method in the revised version and took into account the detailed Referee comments listed below. We have also included an Appendix document, as supplement to the comments, for describing our computational implementation of the method.

Detailed comments are

1. what is the key difference in method between this paper and Maatta et al? Is it simply that Maatta et al. didn't analyze the retrieval uncertainty (p. 11, line 5)?
The method is the same in both papers and applied to OMI measurements. Both papers also analyze the retrieval uncertainty.
The difference is that in the paper Määttä et al. the test cases examined the method at single OMI pixels using two ways: with and without the included model discrepancy term. Whereas in this manuscript the method includes the model discrepancy term and is applied to more comprehensive data set.
In addition this manuscript considers more the presentation and description of the uncertainty due to the model selection.

We have now rephrases the first part of Section 5 (Discussion and Conclusions) (p11, lines 4-9) to express more clearly the difference between the papers.

2. equation 1. To compute reflectance, one needs to know path reflectance that in turn is related to aerosol optical depth. the same is true for

transmittance. Please explain how the calculation in equation 1 is implemented? what are the inputs and from where?

The path reflectance $R_a(\lambda,\tau,\mu,\mu_0,\Delta\phi,p_s)$ and transmittance $T(\lambda,\tau,\mu,\mu_0,p_s)$ are both related to τ (i.e. AOD). They, as well as spherical albedo $s(\lambda,\tau,p_s)$, are taken from the associated multi-dimensional model table LUT by interpolating between LUT contained nodal point values of τ, $\Delta\phi$, p_s , μ and μ_0 . This is explained in the manuscript (page 5 lines 1-7).

As input data we use wavelength bands λ and sun-satellite geometry data included in OMI data ($\Delta\phi$, p_s, μ and μ_0). Please see Section 2.1. and e.g. Torres et al., (2002) for more information about content of OMI LUTs.

We added in the revised manuscript the sentence (page 5): "The sun-satellite geometry data $\Delta\varphi$, p_s, μ and $\mu0$ are included in the OMI Level 1B data."

3. equation 2. Is observation error kept constant for each wavelength in this case?

The observation error $\varepsilon_{obs}(\lambda)$ is assumed to be Gaussian distributed with zero mean and variance $\sigma^2_{obs}(\lambda)$. The variance is not constant as the standard deviation is calculated by $\sigma_{obs}(\lambda) = R_{obs}(\lambda)/SNR$ where we set SNR=500.

To make this clear we added $\varepsilon_{obs}(\lambda) \sim N(0, \sigma^2_{obs}(\lambda))$ and $\sigma_{obs}(\lambda) = R_{obs}(\lambda)/SNR$ in the revised manuscript in Section 3.1.

4. page 5. line 25, "we constructed the covariance function empirically by using the wavelength distance dependent correlation structure of the residuals (See Maata et al 2014 details)". This sentence is very difficult to understand. The paper should standalone by itself.

We have now clarified this sentence in the revised manuscript. We also changed the "covariance function" to "covariance matrix" for simplicity.

However, the theoretical details are left to the reference paper Määttä et al. (2014).

In hope of clarifying the process for constructing the covariance matrix C, we have changed the sentence to:

"The covariance matrix C was constructed by means of an empirical semivariogram when the variances of the residual differences were calculated for each wavelength pairs with the distance d. Next, the theoretical Gaussian variogram model was fitted to these empirical semivariogram values. The outcome of this analysis were the values for parameters that defines the model discrepancy covariance matrix C (see Määttä et al. (2014) for details)."

5. eq. 3. Where does this equation come from? how is measurement error variance computed?

Equation (3) is the likelihood function that describes the distribution of the observations given the model and is dependent on the residuals. The likelihood has that form (Eq. 3) since we assume it follows a multivariate Gaussian distribution with non-diagonal covariance matrix C+diag($\sigma^2_{obs}(\lambda)$). Here C is the model discrepancy covariance matrix.

We added the following sentence in the revised manuscript (page 5 line 28):
"We assume that the likelihood function describing the distribution of the observations given the model follows a Gaussian distribution."

The measurement error variance $\sigma^2_{obs}(\lambda)$ is computed as described above (the comment 3), i.e. $\sigma^2_{obs}(\lambda) = (R_{obs}(\lambda)/SNR)^2$ where we have used SNR = 500.

6. eq. 4. It is not clear how p(tau|m) is constructed. "In the present case, the estimation and model selection procedure seeks the solution for a one-dimensional parameters tau, and the calculations will be fairly straightforward by numerical quadrature. The posterior distribution calculation is presented in the more detail in Maata et al 2014". Again, this reviewer doesn't understand this.

The prior $p(\tau|m)$, i.e. the prior distribution for τ depending on the aerosol microphysical model m, is constructed in our study by assuming that it follows a log-normal distribution with mean value, say 2. This confirms that $p(\tau|m)$ can take only positive real values and thus ensures that estimated AOD is positive.

We have now added in the revised manuscript the sentence (page 6 line 7):
"We assumed that the prior $p(\tau|m)$ follows a log-normal distribution in order to ensure that the estimated AOD is positive."

This sentence (page 6 lines 8-10) was unclear and we have now rephrased the text in the revised manuscript as:
"In our case, the model selection procedure seeks the solution for one parameter τ and then the calculation of posterior distribution is fairly straightforward by numerical quadrature. The calculation of the posterior distribution is presented in more detail in Määttä et al. (2014)."

Please, note that we have also included an Appendix document, as supplement to the comments, for describing our computational implementation of the method.

7. P6, L11-15. How the evidence is computed? This reviewer doesn't understand this paragraph. Later again, Maata et al 2014 is cited, generating a pause in text flow.

The evidence $p(R_{obs}|m)$ is calculated by numerical integration
$p(R_{obs}|m) = \int p(R_{obs}|\tau,m)\, p(\tau|m)\, d\tau$.

We added this formula (page 6 line 11) and the sentence reads now
"The denominator $p(R_{obs}|m) = \int p(R_{obs}|\tau, m)\, p(\tau|m) d\tau$ in Eq. (4) is the probability of the observed reflectance R_{obs} assuming the model m is the correct one."

8. Overall, the method presented here lacks materials for readers to comprehend. An specific example will be helpful to illustrate how all these equations are implemented.

We have now included Appendix document, as supplement to the Referee #2 comments, for describing computational implementation of the method.

9. The results part also lack validation or inter-comaprison with MODIS AOD. Does the method help to interpret the inter-comparison?

The inter-comparison with MODIS, or with other satellite retrievals, is outside scope of this paper.

This is an interesting question. The method presented can help to interpret the results from inter-comparison if the uncertainty is determined and characterized in a way it is comparable. But in principle, maybe in the future, the method described can give additional benefit for the inter-comparison.

10. Introduction part needs to include couple of references that reflect the research activities in U.S.

(1) p2, L15. The following paper used AOD to constrain the emissions as well.

Wang, J. et al., 2012. Top-down estimate of dust emissions through integration of MODIS and MISR aerosol retrievals with the GEOS-Chem adjoint model, Geophys. Res. Lett. L08802.

Xu et al., 2013. Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with Multi-sensor (OMI, MISR) data, J. Geophys. Res. At- mos., 118, 6396-6413.

We have now added the reference papers Wang et al. (2012) and Xu et al. (2013). We also rephrased the sentence in p2 line 15 as

”Furthermore, the satellite based data can be combined with numerical models when estimating aerosol emission fluxes (Huneeus et al.,2012) or spatially constraining amount of aerosol emissions (Wang et al., 2012; Xu et al., 2013).”

(2) P2, L 17-27. while LUT is widely used in operational retrieval algorithm, several research algorithm used aerosol properties from chemistry transport models. This point should be mentioned here.

We apologizes we do not have better knowledge of the research algorithms that use chemistry transport models for aerosol properties. And we do not have a reference to this research work either.

But in order to mention the use of climate models in the retrievals we have added the following sentence (p2, line 23):

“The aerosol properties in the LUTs can be based on observations or combination of observations and climate models (Holzer-Popp et al., 2013).”

A Appendix

A.1 Computational implementation of the method

This Appendix presents a pseudo-code for implementation of a method applied in manuscript Kauppi et al. (2017) and introduced in paper Määttä et al. (2014) step-by-step for a one Ozone Monitoring Instrument (OMI) pixel. The method is based on Bayesian inference approach.

OMI Data:

- The observed top-of-the-atmosphere (TOA) spectral reflectance $\vec{R}_{\text{obs}}(\lambda)$ at selected wavelength bands $\lambda = (\lambda_1, \ldots, \lambda_n)$ calculated from the OMI Level 1B VIS and UV radiances and Level 1B Solar irradiance data
- The measurement error variances $\sigma^2_{\text{obs}}(\lambda)$, $\lambda = (\lambda_1, \ldots, \lambda_n)$
- The set of Look-up-tables (LUTs) containing pre-calculated aerosol microphysical models (e.g. hdf5 files)

Outcome:

- Posterior distribution $p(\tau|\vec{R}_{\text{obs}}, m)$ of τ (i.e. AOD) given as a discrete set of values for τ in the range of [0,$\tau_{\max}$]. The posterior distribution is evaluated for each selected best fitting model (maximum of 10) and stored in a table.
- Averaged posterior distribution $p_{\text{avg}}(\tau|\vec{R}_{\text{obs}})$ given as a discrete set of values for τ in the range of [0,$\tau_{\max}$] and stored in a table.
- Point estimate for AOD at 500 nm determined as maximum a posteriori (MAP) estimate, i.e. mode of the averaged posterior distribution

We use a symbol τ for AOD in the formulas. The modeled reflectance $\vec{R}_{\text{mod}}(\tau, \lambda)$ depends on τ and is calculated by interpolation between nodal values of LUT while fitted to the measured reflectance $\vec{R}_{\text{obs}}$ in order to find τ that minimizes

$$\chi^2_{\text{mod}}(\tau) = \vec{R}_{\text{res}}(\lambda)^T \left(\mathbf{C} + \text{diag}(\sigma^2_{\text{obs}}(\lambda))\right)^{-1} \vec{R}_{\text{res}}(\lambda). \tag{1}$$

Here $\vec{R}_{\text{res}}(\lambda) = \vec{R}_{\text{obs}}(\lambda) - \vec{R}_{\text{mod}}(\tau, \lambda)$ is the residual of model fit. This is done for each aerosol microphysical model in turn. In the formula $\sigma^2_{\text{obs}}(\lambda)$ are the measurement error variances and $\mathbf{C}$ is non-diagonal covariance matrix for model discrepancy (i.e. forward modelling uncertainty). In our experiment we calculated the elements of the covariance matrix $\mathbf{C}$ for wavelength pair λ_i and λ_j as

$$\mathbf{C}_{i,j} = \sigma_1^2 \exp\left(-\frac{1}{2}\left(\lambda_i - \lambda_j\right)^2 / l^2\right) + \sigma_0^2 \tag{2}$$

where parameter l is a correlation length, parameter σ_0^2 is non-spectral (i.e. non-spatial) diagonal variance and σ_1^2 is spectral (i.e. spatial) variance. We like to note that our used parameter values are specific for this study with OMI data and have been empirically evaluated. These parameter values were estimated from an ensemble of the residuals, i.e. the differences between the observed and modeled reflectances, as described in the paper Määttä et al. (2014). Here we used $l = 90$ nm and for σ_0^2 and σ_1^2 we used values of 1% and 2% of the observed reflectance, respectively.

By Bayes' formula the posterior distribution for τ within the model m and given the observed reflectance $\vec{R}_{\text{obs}}$ is

$$p(\tau|\vec{R}_{\text{obs}}, m) = \frac{p(\vec{R}_{\text{obs}}|\tau,\, m)\, p(\tau|m)}{p(\vec{R}_{\text{obs}}|m)}. \tag{3}$$

In this case we have one unknown τ (i.e. AOD at 500 nm) and the full posterior distribution is calculated as described below.

The posterior is evaluated at a dense grid, e.g. at 200 points, of τ values, basically in the range of $[0,\, \tau_{\max}]$. The maximum allowed $\tau_{\max}$ is determined by the model LUT.

We calculated the likelihood as

$$p(\vec{R}_{\text{obs}}|\tau,\, m) = c \exp(-\frac{1}{2} * \chi^2_{\text{mod}}(\tau)), \tag{4}$$

where $\chi^2_{\text{mod}}(\tau)$ is calculated from Eq. 1 for the set of τ values in the range of $[0,\, \tau_{\max}]$. The constant c ensures that the probability distribution is properly defined and it is the same for all the models m.

We assumed that a prior distribution $p(\tau|m)$ for τ within aerosol microphysical model m follows a log-normal distribution

$$p(\tau|m) \propto \log N(\tau_0, \sigma_\tau^2). \tag{5}$$

This confirms that $p(\tau|m)$ can take only positive real values and ensures that AOD is positive. We set mean value $\tau_0 = 2$ for the log-normal distribution.

We calculated the normalizing constant (or scaled factor) of the posterior numerically as

$$p(\vec{R}_{\text{obs}}|m) = c \int p(\tau|m) * \exp(-\frac{1}{2} * \chi^2_{\text{mod}}(\tau))d\tau. \tag{6}$$

Consequently, we have now calculated all the elements of the posterior distribution for τ (Eq. 3).

In our study we call $p(\vec{R}_{\text{obs}}|m)$ as the model evidence that is used to make the model selection. We select models with the highest evidence value until the cumulative sum of the selected models' evidences pass the value 0.8 or the number of chosen models is 10.

Next we calculate relative evidence for model m_i with respect to the other models selected above (max 10) by

$$p(m_i|\vec{R}_{\text{obs}}) = \frac{p(\vec{R}_{\text{obs}}|m_i)}{\sum_j(\vec{R}_{\text{obs}}|m_j)}. \tag{7}$$

These relative evidence values are used to compare models among the set of selected best fitting models.

The averaged posterior distribution over the selected best models m_i is calculated as

$$p_{\text{avg}}\left(\tau|\vec{R}_{\text{obs}}\right) = \sum_{i=1}^{n} p(\tau|\vec{R}_{\text{obs}},\, m_i)\, p(m_i|\vec{R}_{\text{obs}}), \tag{8}$$

where n is the number of models.

We accept the solution for the pixel if the threshold value $\chi^2 \leq 2$ calculated by following modified chi-squared formula

$$\chi^2 = \frac{1}{n-1}\vec{R}_{\text{res}}(\lambda)^{\text{T}}\left(\mathbf{C} + \text{diag}(\sigma^2(\lambda))\right)^{-1}\vec{R}_{\text{res}}(\lambda). \tag{9}$$

We do this test only for the best model.

As a summary, we do the following for model selection, calculation of posterior distributions and getting MAP estimate of AOD:

1. fit each model from LUT (i.e. $\vec{R}_{\text{mod}}(\tau,\, \lambda)$) in turn to the measured reflectance $\vec{R}_{\text{obs}}(\lambda)$
2. for each model, find τ that minimizes $\chi^2_{\text{mod}}(\tau)$ (Eq. 1)
3. for each model, calculate posterior distribution $p(\tau|\vec{R}_{\text{obs}}, m)$ (Eq. 3)
4. use model evidence (Eq. 6) to select max 10 best models
5. calculate the relative evidence (Eq. 7) for each model among the selected best models. Actually, we first carry out steps 2.-3. once more for the selected best models and then calculate the relative evidences.

6. calculate the averaged posterior distribution (Eq. 8) and get point estimate for AOD, i.e. MAP estimate

7. finally, do the goodness-of-fit test (Eq. 9)

References

Määttä, A., Laine, M., Tamminen, J., and Veefkind, J. P.: Quantification of uncertainty in aerosol optical thickness retrieval arising from aerosol microphysical model and other sources, applied to Ozone Monitoring Instrument (OMI) measurements, Atmos. Meas. Tech., 7, doi:10.5194/amt-7-1185-2014, 2014.

Kauppi, A., Kolmonen, P., Laine, M., and Tamminen, J.:Aerosol type retrieval and uncertainty quantification from OMI data, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-47, 2017.