Atmospheric
Measurement
Techniques

Quantification of uncertainty in aerosol optical thickness retrieval arising from aerosol microphysical model and other sources, applied to Ozone Monitoring Instrument (OMI) measurements

A. Määttä[1,2], M. Laine[1], J. Tamminen[1], and J. P. Veefkind[3,4]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]University of Helsinki, Department of Mathematics and Statistics, Helsinki, Finland
[3]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[4]Delft University of Technology, Department of Civil Engineering and Geosciences, Delft, the Netherlands

Correspondence to: A. Määttä (anu.maatta@fmi.fi)

Abstract. Satellite instruments are nowadays successfully utilised for measuring atmospheric aerosol in many applications as well as in research. Therefore, there is a growing need for rigorous error characterisation of the measurements. Here, we introduce a methodology for quantifying the uncertainty in the retrieval of aerosol optical thickness (AOT). In particular, we concentrate on two aspects: uncertainty due to aerosol microphysical model selection and uncertainty due to imperfect forward modelling. We apply the introduced methodology for aerosol optical thickness retrieval of the Ozone Monitoring Instrument (OMI) on board NASA's Earth Observing System (EOS) Aura satellite, launched in 2004. We apply statistical methodologies that improve the uncertainty estimates of the aerosol optical thickness retrieval by propagating aerosol microphysical model selection and forward model error more realistically. For the microphysical model selection problem, we utilise Bayesian model selection and model averaging methods. Gaussian processes are utilised to characterise the smooth systematic discrepancies between the measured and modelled reflectances (i.e. residuals). The spectral correlation is composed empirically by exploring a set of residuals. The operational OMI multiwavelength aerosol retrieval algorithm OMAERO is used for cloud-free, over-land pixels of the OMI instrument with the additional Bayesian model selection and model discrepancy techniques introduced here. The method and improved uncertainty characterisation is demonstrated by several examples with different aerosol properties: weakly absorbing aerosols, forest fires over Greece and Russia, and Sahara desert dust. The statistical methodology presented is general; it is not restricted to this particular satellite retrieval application.

1 Introduction

Many ongoing studies aim for a better understanding of atmospheric aerosol properties such as size distribution, type, optical properties, formation and transport. The remote sensing of atmospheric aerosols from space enables the monitoring of aerosols on both regional and global scales. Satellite measurements are widely used together with ground-based and airborne measurements to provide data for important atmospheric aerosol studies related to, for example, climate change, energy budget, air quality and cloud properties.

Atmospheric aerosols have been monitored for years from several satellite instruments including the Ozone Monitoring Instrument (OMI) (Torres et al., 2007), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Levy et al., 2010; van Donkelaar et al., 2013), the Global Ozone Monitoring Experiment-2 (GOME-2), the Multi-angle Imaging SpectroRadiometer (MISR) (Kahn et al., 2010), the (Advanced) Along-Track Scanning Radiometers ((A)ATSR) (Thomas et al., 2009; Sayer et al., 2010, 2012), the Cloud-Aerosol Lidar and Infrared Path finder (CALIPSO), the Scanning Imaging Absorption spectroMeter for Atmospheric Chartography

(SCIAMACHY), the Polarisation and Directionality of the Earth's Reflectances (POLDER) (Dubovik et al., 2011) and the Spinning Enhanced Visible and Infrared Imager (SE-VIRI) (Govaerts et al., 2010; Wagner et al., 2010). The instruments vary in terms of spatial and spectral resolution, polarisation, and viewing geometry.

The determination of aerosol properties from satellite measurements is an ill-posed inverse problem as the limited information content in the observations does not allow for complete determination of all relevant aerosol properties. Prior information, such as assumed surface conditions and selection of aerosol optical properties for pre-calculated radiative transfer results, is an essential part in the retrieval process. For the solution of the inverse problem, various assumptions and simplifications are needed.

The forward problem in aerosol retrieval is based on radiative transfer calculations which depend on various aerosol properties. Currently, these calculations are too time-consuming to be performed simultaneously with the retrieval inversion, and many operational algorithms are based on pre-calculated look-up tables (LUT) for a selection of aerosol types. The atmospheric aerosol column content above the Earth's ground pixel can be a mixture of several aerosol types, which complicates the choice of the correct aerosol type. One important reason for the disagreement between results derived from different satellite instruments for the same location and time is the difference in the algorithms and in the assumption of the underlying aerosol model (Kokhanovsky et al., 2010; Li et al., 2009; Livingston et al., 2009). This choice of an appropriate aerosol model plays a significant part in the retrieval of aerosols from satellites. An additional large source of uncertainty comes from the assumptions on surface albedo (Thomas et al., 2009; Govaerts et al., 2010; Wagner et al., 2010). This aspect will not be studied in the current paper. See the references given for the individual instruments at the beginning of this Sect. for different aerosol retrieval algorithms and attempts at uncertainty quantification. References for non-theoretical aerosol optical thickness (AOT) validation and error budgets include Kahn et al. (2010), Levy et al. (2010) and Sayer et al. (2013).

The aim of this paper is to introduce a methodology for quantifying the uncertainty in the retrieved AOT (sometimes referred to as AOD for aerosol optical depth by other authors), which is a dimensionless measure of the amount of light absorbed or scattered by the aerosols. In particular, we concentrate on two related and often overlooked aspects: uncertainty due to aerosol microphysical model selection and uncertainty due to forward model errors. We use tools from Bayesian model selection methodology to weight the aerosol microphysical models according to their goodness of fit (MacKay, 1992; Spiegelhalter et al., 2002; Robert, 2007) and combine information about the AOT over the best fitting models by averaging over the best models (Hoeting et al., 1999).

The aerosol microphysical models contain results of radiative transfer calculations for various aerosol physical properties. Systematic differences in retrieval residuals (i.e. modelled values minus observed values) indicate imperfect forward modelling. We call this source of error *model discrepancy*, following Kennedy and O'Hagan (2001). The model discrepancy itself is modelled using Gaussian processes that define the allowed deviations from modelled to observed reflectance by a suitable covariance structure that lets the residuals correlate depending on their distance in wavelength (Kennedy and O'Hagan, 2001; Rasmussen and Williams, 2006). The spectral correlation is composed empirically by exploring a set of residuals.

Here, the introduced methodology is applied to OMI measured reflectances at the top of the atmosphere (TOA), using the aerosol microphysical models of the OMI multi-wavelength aerosol algorithm OMAERO. The operational OMAERO product uses a look-up table (LUT)-based technique for the retrieval of aerosol optical properties in the ultraviolet and visible wavelength region. The multidimensional LUT contains pre-calculated aerosol microphysical models having specific optical properties such as AOT and single scattering albedo (SSA). The aerosol microphysical models represent four main types of aerosols: desert dust, biomass burning, weakly absorbing and volcanic aerosols (Torres et al., 2002, 2007; Livingston et al., 2009).

The motivation for this study was to improve the existing model selection algorithm of OMI to the benefit of the future TROPOspheric Monitoring Instrument (TROPOMI) algorithm development (Veefkind et al., 2012). The next Sect. 2 introduces the OMAERO algorithm. Section 3 describes the Bayesian model selection technique for choosing the aerosol microphysical model based on satellite observations with associated uncertainty. The characteristics for model error are determined in Sect. 4. Finally, aerosol microphysical model selection in different atmospheric cases is exemplified in Sect. 5.

2   Data and operational OMI multi-wavelength aerosol algorithm OMAERO

In this study we have used reflected solar radiation measurements from OMI on board NASA's Earth Observing System (EOS) Aura satellite, launched in July 2004. The Aura spacecraft is in polar sun-synchronous orbit at an altitude of 705 km and has a daily global coverage with 14 orbits. The OMI instrument was built in cooperation with Finland and the Netherlands. OMI is a nadir-viewing solar backscatter spectrometer, measuring in the ultraviolet (UV) and visible (VIS) regions between 270 and 500 nm. The ground pixel size is $13 \times 24\ \mathrm{km^2}$ at nadir. OMI-measured Earth radiance and solar irradiance spectra with moderate spatial resolution are used to retrieve (among others) aerosol characteristics, surface UV, cloud information and atmospheric trace gases

including ozone, NO_2, SO_2, HCHO, BrO and OClO. The retrievals are used in the studies of air quality, ozone trend and relation between atmospheric chemical composition and climate change (Levelt et al., 2006a, b; Torres et al., 2007).

The operational OMI aerosol multi-wavelength algorithm OMAERO has been developed to retrieve aerosol optical properties for cloud-free scenes using reflectance spectrum in the near UV and visible wavelength range between 331 and 500 nm. The OMAERO Level-2 data is available for public access from NASA's Goddard Space Flight Center (GSFC) Earth Sciences (GES) Data and Information Services Center (DISC) (http://disc.gsfc.nasa.gov/Aura/OMI/omaero_v003.shtml). The available data period is from 1 October 2004 to the present. The OMAERO Level-2 product provides aerosol properties including aerosol type, AOT, SSA, aerosol absorption indices and other related data (Torres et al., 2002, 2007). The principal component analysis applied to the OMAERO algorithm by Veihelmann et al. (2007) shows that OMI reflectance measurements have two to four degrees of freedom of signal.

The current version (V003) of operational OMAERO product uses a surface albedo climatology based on five years of OMI observations (Kleipool et al., 2008) for pixels over land. Over oceans the spectral bidirectional reflectance distribution function is calculated by means of an ocean model that accounts for wind speed and chlorophyll concentration from a climatology. The main factors that have an effect on the retrieved AOT uncertainty are the sub-pixel cloud contamination, assumed surface albedo spectrum, instrumental factors and aerosol model assumptions (Veihelmann et al., 2007; Brinksma et al., 2008; Curier et al., 2008; Livingston et al., 2009). For a thorough description of the OMAERO aerosol retrieval algorithm, the reader is referred to Torres et al. (2002, 2007) and to the OMAERO Readme document (available online, for example at http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omaero_v003.shtml).

2.1 Aerosol optical thickness retrieval

For the OMAERO algorithm, the radiative transfer calculations were done in advance for a range of aerosol physical properties and sun-satellite geometries (Torres et al., 2002, 2007). These aerosol microphysical models are divided into four main types: desert dust (DD), biomass burning (BB), weakly absorbing (WA) and volcanic aerosols (VO). The main types are divided into subtypes according to aerosol size distribution, refractive index and vertical profile, adding up to about fifty aerosol microphysical models in total. See Table 1 for the tabulated aerosol properties. LUTs. The results of the radiative transfer model calculations are stored in multidimensional LUTs. The LUTs consist of various model parameters for a set of nodal points including AOT, SSA, solar zenith angle, viewing zenith angle, relative azimuth angle, path reflectance, transmission and spherical albedo.

Table 1. The size distribution parameters and the wavelength dependent single scattering albedo (SSA) for each aerosol microphysical model in LUTs used in this study. The third digit in the model ID number, which is zero in the table, has a range of 1–3 for BB and DD models, and 1 for WA and VO models, is used for different vertical distributions. The mean particle radius, rg [micron] and the standard deviation, σ [micron], are given for both modes, m1 and m2, of bimodal distribution, and n21 is the second mode fraction of the number concentration. The SSA is given for the first and last wavelength band only.

Model	rg m1	rg m2	σ m1	σ m2	n21	SSA
WA1101	0.078	0.497	1.499	2.160	4.36e-4	1–1
WA1102	0.088	0.509	1.499	2.160	4.04e-4	1–1
WA1103	0.137	0.567	1.499	2.160	8.10e-4	1–1
WA1104	0.030	0.240	2.030	2.030	1.53e-2	1–1
WA1201	0.078	0.497	1.499	2.160	4.36e-4	0.96–0.95
WA1202	0.088	0.509	1.499	2.160	4.04e-4	0.97–0.96
WA1203	0.137	0.567	1.499	2.160	8.10e-4	0.97–0.98
WA1301	0.078	0.497	1.499	2.160	4.36e-4	0.91–0.88
WA1302	0.088	0.509	1.499	2.160	4.04e-4	0.91–0.90
WA1303	0.137	0.567	1.499	2.160	8.10e-4	0.92–0.92
BB2101	0.074	0.511	1.537	2.203	1.70e-4	0.94–0.93
BB2102	0.087	0.567	1.537	2.203	2.06e-4	0.94–0.93
BB2103	0.124	0.719	1.537	2.203	2.94e-4	0.93–0.94
BB2201	0.074	0.511	1.537	2.203	1.70e-4	0.90–0.88
BB2202	0.087	0.567	1.537	2.203	2.06e-4	0.90–0.89
BB2203	0.124	0.719	1.537	2.203	2.94e-4	0.89–0.90
BB2301	0.074	0.511	1.537	2.203	1.70e-4	0.86–0.82
BB2302	0.087	0.567	1.537	2.203	2.06e-4	0.86–0.84
BB2303	0.124	0.719	1.537	2.203	2.94e-4	0.84–0.85
DD3101	0.042	0.670	1.697	1.806	4.35e-3	0.82–0.94
DD3102	0.052	0.670	1.697	1.806	4.35e-3	0.86–0.95
DD3201	0.042	0.670	1.697	1.806	4.35e-3	0.74–0.90
DD3202	0.052	0.670	1.697	1.806	4.35e-3	0.79–0.91
VO4101	0.230	0.230	0.800	0.800	0.5	1–1

The operational OMAERO algorithm uses TOA reflectance measurements $\mathbf{R}_{obs}(\lambda)$ to find aerosol microphysical model and a value for AOT, τ, that best matches the observations. As the spectral shape of the AOT is fixed by any given aerosol microphysical model input configurations, there is only one parameter to be fitted, the AOT at reference wavelength (500 nm) $\tau = \tau(\lambda_{ref})$.

In the fitting procedure, a subset of aerosol models are preselected according to a priori knowledge of aerosol regional and seasonal distribution. The fitting is done using the least square criteria by minimising

$$\chi^2_{mod} = \sum_{i=1}^{L} \left(\frac{\mathbf{R}_{obs}(\lambda_i) - \mathbf{R}_{mod}(\tau, \lambda_i)}{\boldsymbol{\sigma}(\lambda_i)} \right)^2, \quad (1)$$

where L is the number of wavelength bands (14 in our case) between 331 and 500 nm, $\boldsymbol{\sigma}(\lambda_i)$ is the standard deviation uncertainty in the measured reflectance, which is assumed to be known, and $\mathbf{R}_{mod}(\tau, \lambda_i)$ is the reflectance from the aerosol LUT model (Torres et al., 2002, 2007). The best fitted model is selected according to lowest chi-squared value χ^2_{mod} and is used to determine the spectral AOT. In addition, a maximum

of ten models, for which the root mean square of the residual reflectance is below a given threshold value, is delivered with related AOT and SSA (Torres et al., 2007; Livingston et al., 2009).

2.2 Reflectance

The AOT $\tau = \tau(\lambda_{ref})$ is retrieved from TOA reflectance spectrum. The TOA spectral reflectance $\boldsymbol{R}_{obs}(\lambda)$ is calculated as the ratio of observed OMI Level-1b Earth radiance $E(\lambda)$ over the observed OMI Level-1b solar irradiance spectra $F(\lambda)$ by

$$\boldsymbol{R}_{obs}(\lambda) = \frac{\pi E(\lambda)}{\cos(\theta_{sun}) F(\lambda)}, \tag{2}$$

where θ_{sun} is the solar zenith angle (Levelt et al., 2006b; Torres et al., 2007).

Above a Lambertian surface the TOA reflectance $\boldsymbol{R}_{mod}(\tau, \lambda)$ for the aerosol microphysical model is calculated as

$$\boldsymbol{R}_{mod}(\lambda, \tau, \mu, \mu_0, \Delta\phi, p_s) = R_a(\lambda, \tau, \mu, \mu_0, \Delta\phi, p_s)$$
$$+ \frac{A_s(\lambda)}{1 - A_s(\lambda) s(\lambda, \tau, p_s)} T(\lambda, \tau, \mu, \mu_0, p_s), \tag{3}$$

where path reflectance R_a, transmittance T and spherical albedo s of the atmosphere as seen from below together with the parameters τ (AOT), $\Delta\phi$ (relative azimuth angle), p_s (surface pressure), μ (cosine of viewing zenith angle) and μ_0 (cosine of solar zenith angle) are in practice taken from LUT. Over land the surface reflectivity A_s is taken from the land albedo climatology. In this paper, we concentrate on over-land retrieval, only. Over-sea the reflectance model is a little more complicated as the land albedo climatology is replaced with an ocean model depending on chlorophyll and wind properties (Torres et al., 2002; Veihelmann et al., 2007).

3 Bayesian model choice

There are various sources of uncertainty affecting the accuracy of the retrieved AOT values, and the selection of an appropriate LUT for modelled reflectance calculations is only one factor. Others are related to the size of OMI pixels, sub-pixel cloud contamination, aerosol horizontal inhomogeneity, etc. One large source of uncertainty comes from the use of surface albedo climatologies. In this study, we apply Bayesian model selection tools to select the most appropriate LUT and quantify the related uncertainty. Secondly, we need to take into account the other sources of uncertainties that might cause systematic model discrepancies. This is done by characterising model discrepancy with Gaussian processes described in Sect. 4.

We want to choose an aerosol microphysical model from a set of models that provides then best explanation to the observed reflectance at each OMI pixel. As several models

might be equally good, an important task here is to be able to quantify the uncertainty coming from the model selection procedure. We will use tools from Bayesian statistical inference for model choice, model averaging and model error that naturally account for different sources of uncertainties. Bayesian analysis will provide the solution to the estimation after accounting for the uncertainties in the modelling procedure. This solution will take the form of a posterior probability density that is a measure of uncertainty in the quantity of interest.

Model selection in general is a delicate problem that can not be solved by statistical reasoning only. Theoretically speaking, for a given data set, there will be an infinite number of different models that fit the data equally well. Here, we deal with the specific problem of choosing the most suitable aerosol microphysical model from a given finite set of candidate models. We acknowledge the fact that none of the aerosol microphysical models might give adequate fit to the observations and also want to have a measure for that situation.

3.1 Bayesian parameter estimation and model comparison

Typical statistical parameter estimation is a stepwise procedure, where first a given model is fitted to the observations to get a parameter estimate and its uncertainty. Then model residuals (i.e. the difference between modelled values and observed ones) are studied to see if the assumptions about the residuals are met. This is called model diagnostics, where one typically checks any systematic features in the residuals, which signal inadequacy in the model formulation, and one also checks the form of the distribution of the residuals, which signals problems in the statistical assumptions. When we have several possible models, as in the OMI case, one can fit all the models, one by one, and see which provides the best fit according to some chosen criteria, such as minimum least squares. Gelman et al. (2003) provide a comprehensive introduction to statistical modelling in general and Bayesian inference in particular.

We recall and outline the Bayesian parameter estimation and model selection in the current framework of finding the posterior distribution of the AOT parameter τ using the OMAERO algorithm. The posterior distribution for the uncertainty in τ after observing $\boldsymbol{R}_{obs}$ is given by Bayes' formula

$$p(\tau|\boldsymbol{R}_{obs}, m) = \frac{p(\boldsymbol{R}_{obs}|\tau, m) p(\tau|m)}{p(\boldsymbol{R}_{obs}|m)}, \tag{4}$$

where the likelihood $p(\boldsymbol{R}_{obs}|\tau, m)$ and the prior distribution $p(\tau|m)$ depend on the aerosol microphysical model m. This will give us valid posterior inferences about τ given the observed and modelled reflectance, prior distribution on τ and assuming that m is the correct model. As the posterior $p(\tau|\boldsymbol{R}_{obs}, m)$ is a probability distribution and the denominator $p(\boldsymbol{R}_{obs}|m)$ does not depend on τ, the latter must be a

constant that normalises the numerator

$$p\left(\boldsymbol{R}_{\mathrm{obs}}|m\right) = \int p\left(\boldsymbol{R}_{\mathrm{obs}}|\tau, m\right) p(\tau|m)\,\mathrm{d}\tau. \qquad (5)$$

For model selection this constant has an important use. It is the probability of observing $\boldsymbol{R}_{\mathrm{obs}}$ given the model m. This value is sometimes called evidence (e.g. Gershenfeld, 1999) and it is needed in the computation of model posterior probabilities. Basically, we could select the aerosol microphysical model that has the largest evidence with respect to the observations. There are some caveats on using the evidence for model choice pointed out in statistical literature (Robert, 2007). In this particular case, where a one-element vector is fitted with a selection of possible models, we find basic Bayesian model selection very useful, provided that we can account for the model error as done in Sect. 4. In general, evaluating the evidence from Eq. (5) requires calculating the integral, which is difficult for unknowns of dimensions larger that, say, three. In our case, the dimension of τ is one, which makes this calculation relatively straight forward by numerical quadrature.

The least square criteria in Eq. (1) has a direct counterpart within Bayesian inference as it appears in exactly the same form in the likelihood function for Gaussian observation error

$$p\left(\boldsymbol{R}_{\mathrm{obs}}|\tau, m\right) \propto$$
$$\exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{\boldsymbol{R}_{\mathrm{obs}}\left(\lambda_i\right) - \boldsymbol{R}_{\mathrm{mod}}\left(\tau, \lambda_i\right)}{\boldsymbol{\sigma}\left(\lambda_i\right)}\right)^2\right), \qquad (6)$$

where we assume the measurement noise standard deviations $\boldsymbol{\sigma}(\lambda)$ to be known. If we assume an uninformative prior for τ, i.e. $p(\tau|m) = 1$, the least squares estimate, maximum likelihood estimate (MLE) and Bayesian maximum a posteriori (MAP) estimate are all equal.

To compare models, we use a method based on the posterior model probabilities. For a model m and measurements $\boldsymbol{R}_{\mathrm{obs}}$ we use Bayes' theorem again to obtain

$$p\left(m|\boldsymbol{R}_{\mathrm{obs}}\right) = \frac{p(\boldsymbol{R}_{\mathrm{obs}}|m)\,p(m)}{p(\boldsymbol{R}_{\mathrm{obs}})}, \qquad (7)$$

where $p(m)$ is the prior probability that model m is the correct one. This formula describes the probability of model m, assuming that the measurements have been generated from this model. The evidence term from Eq. (5) appears here as the marginal likelihood $p(\boldsymbol{R}_{\mathrm{obs}}|m)$ of observed data within model m. The denominator is again a normalising constant defined as the sum over all the models considered:

$$p\left(\boldsymbol{R}_{\mathrm{obs}}\right) = \sum_i p\left(\boldsymbol{R}_{\mathrm{obs}}|m_i\right) p\left(m_i\right). \qquad (8)$$

As we are going to deal with a relatively small number of different models, this term is easily calculated, provided we can calculate the individual evidence values.

In case when, a priori, all models are equally likely, the model comparison and calculation of relative weights for each model simplifies to calculating the relative evidences:

$$p\left(m|\boldsymbol{R}_{\mathrm{obs}}\right) = \frac{p(\boldsymbol{R}_{\mathrm{obs}}|m)}{\sum_i p(\boldsymbol{R}_{\mathrm{obs}}|m_i)}. \qquad (9)$$

Consequently, in this case the model with the highest evidence is the best among the models involved. We can compare models to see if one is clearly the best with respect to other models, or if there are several that are equally plausible. Note, however, that having the largest evidence among a set of models does not guarantee a good, or even adequate, fit in itself.

An important aspect in Bayesian analysis, the specification of prior distributions, has not been discussed so far. As we are mainly performing a feasibility and method development study, we have used rather conventional choices. For each individual model fit, the prior distribution for AOT parameter τ was set to log-Gaussian with mean value 2 and 700 % standard deviation. This ensured the positivity of the estimated AOT values and was only weakly informative in all of the test cases. For the model choice, uniform prior was used for $p(m)$, i.e. all the models were a priori equally likely.

3.2 Bayesian model averaging

In practice, several aerosol microphysical models can provide equally good explanations for the measurements and the particular one with the highest evidence may simply have been obtained by chance. If there is uncertainty in the model selection, it should be accounted for in the inference about the quantity of interest. A Bayesian model averaging technique (Hoeting et al., 1999; Robert, 2007) enables the shared inference about an unknown appearing in several alternative models.
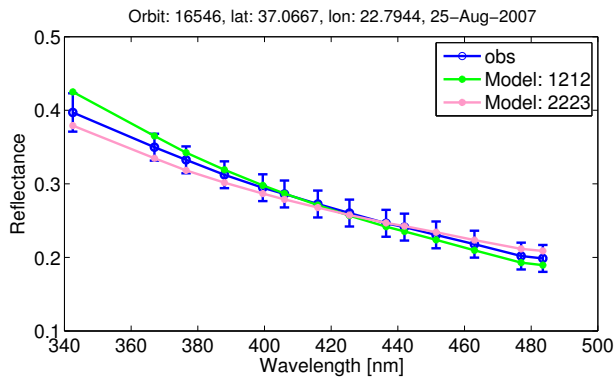
The Bayesian model averaging uses combined posterior distribution defined by weighting the individual posteriors by their evidence-based weights:

$$p_{\mathrm{avg}}\left(\tau|\boldsymbol{R}_{\mathrm{obs}}\right) = \sum_{i=1}^{n} p\left(\tau|\boldsymbol{R}_{\mathrm{obs}}, m_i\right) p\left(m_i|\boldsymbol{R}_{\mathrm{obs}}\right). \qquad (10)$$

If different models give rise to different values for the unknown, then the uncertainty in the averaged posterior distribution $p_{\mathrm{avg}}(\tau|\boldsymbol{R}_{\mathrm{obs}})$ can be larger than it is with any single model. This means that the uncertainty in model selection has been incorporated into the result (Hoeting et al., 1999; Robert, 2007).

4 Model discrepancy

In Fig. 1, the reflectance spectrum from one OMI pixel is shown together with two fitted OMAERO models. The models represent two different aerosol main types, weakly

Fig. 1. Two aerosol models that fit the observed reflectance equally well.

Fig. 2. The spectral differences between the observed reflectance and the reflectance for the best fit models. Each colour represents a set of residuals for pixels from a selected orbit. The residuals correspond to different atmospheric situations: dust storm in Sahara (green), wildfires in Russia (red) and when weakly absorbing models dominate (blue).

absorbing (model "1212") and biomass burning (model "2223"). They both fit the observed reflectance equally well and the two modelled curves show similar, though opposite deviations from the observed curve. Both models can explain the observations within the individual observation error-bar uncertainties, but there is significant systematic bias. Next, we want to characterise this forward modelling uncertainty, which is called model discrepancy (Kennedy and O'Hagan, 2001). It contains all sources of uncertainties not directly due to the measurement noise, such as those related to surface albedo, LUT interpolation and aerosol microphysical modelling.

To account for the model discrepancy, we use an additional error term $\eta(\lambda)$ and write the general model equation as

$$\boldsymbol{R}_{\mathrm{obs}}(\lambda) = \boldsymbol{R}_{\mathrm{mod}}(\tau, \lambda) + \eta(\lambda) + \epsilon_{\mathrm{obs}}(\lambda). \tag{11}$$

As before, we assume that the spectral measurement uncertainty due to instrument noise is known and Gaussian

$$\epsilon_{\mathrm{obs}}(\lambda) \sim N\left(0, \sigma^2(\lambda)\right). \tag{12}$$

We wish to build a statistical model for the remaining model discrepancy term $\eta(\lambda)$. In order to see how this discrepancy behaves, we studied residuals of model fits, i.e. the differences between the observed reflectances and the modelled reflectances,

$$\boldsymbol{R}_{\mathrm{res}}(\lambda) = \boldsymbol{R}_{\mathrm{obs}}(\lambda) - \boldsymbol{R}_{\mathrm{mod}}(\tau, \lambda) \tag{13}$$

at wavelengths λ. The modelled reflectances were calculated from aerosol microphysical models that were the most appropriate according to the operational OMAERO product. Figure 2 shows residuals representing three different atmospheric situations: dust storm in Sahara (green), wildfires in Russia (red) and when weakly absorbing models dominate (blue). We found that the residuals have typically very similar systematic behaviour that could be modelled by a suitable correlation structure. By using standard tools from spatial statistics, we estimate this correlation structure and use it to build a model for the model error.
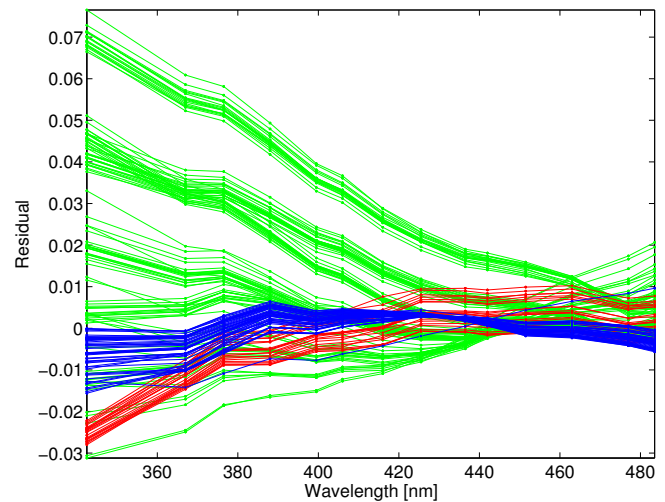
4.1 Gaussian process

Following Kennedy and O'Hagan (2001), we use a Gaussian process approach to describe the model discrepancy term $\eta(\lambda)$ that originates from the difference between the aerosol-model-generated reflectance and the observations. Gaussian process is a stochastic process for which every finite set of its realisations has a joint Gaussian distribution (Rasmussen and Williams, 2006). It is a theoretical tool that provides a general and flexible framework for constructing the discrepancy model $\eta(\lambda)$. As we only deal with finite representations, we can, in practice, work with random variables and covariance matrices.

A Gaussian process is defined by its mean and covariance functions, and the essential part in the implementation is the determination and parameter estimation related to the covariance function. Here, the model discrepancy will be a zero mean Gaussian process $\eta(\lambda) \sim \mathrm{GP}(0, C)$, where the covariance function C quantifies the correlation properties of the discrepancy. As there typically is no direct data available about the covariance, one proceeds by assuming a certain parameterised functional form. Following Banerjee et al. (2004), we derived the covariance function C using a Gaussian variogram model. The covariance depends only on the wavelength distance $|\lambda_i - \lambda_j|$ and is defined as

$$C\left(\lambda_i, \lambda_j\right) = \tag{14}$$
$$\begin{cases} \sigma_1^2 \exp\left(-\left(\lambda_i - \lambda_j\right)^2 / l^2\right), & \lambda_i \neq \lambda_j \\ \sigma_0^2 + \sigma_1^2, & \lambda_i = \lambda_j \end{cases},$$

where l is the correlation length parameterising the distance between two wavelengths where the residuals are still correlated. Parameter σ_0^2 represents non-spectral diagonal variance and σ_1^2 corresponds to spectral variance. These three parameters σ_0^2, σ_1^2 and l are the essential characteristics of the covariance function to be determined. In the next section we show how we estimated the covariance function empirically from wavelength-dependent correlation structure of residuals of model fits.

After the model discrepancy term has been estimated, the theoretical covariance function is used to form the corresponding covariance matrix $\mathbf{C}$, defined for the range of wavelength bands of the observations. Then it can be incorporated into the likelihood function (Eq. 6) as an additional error covariance:

$$p\left(\boldsymbol{R}_{\mathrm{obs}}|\tau, m\right) \propto$$

$$\exp\left(-\frac{1}{2}\,\boldsymbol{R}_{\mathrm{res}}^{T}\left(\mathbf{C} + \mathrm{diag}\left(\sigma^2(\boldsymbol{\lambda})\right)\right)^{-1}\boldsymbol{R}_{\mathrm{res}}\right), \qquad (15)$$

where $\boldsymbol{R}_{\mathrm{res}}$ is the residual of model fit (Eq. 13). The joint covariance matrix in Eq. (15) consists now of two elements: $\mathbf{C}$ is the covariance matrix for model discrepancy (Eq. 14) and $\mathrm{diag}(\sigma^2(\boldsymbol{\lambda}))$ is the diagonal matrix having measurement error variances $\sigma^2(\boldsymbol{\lambda})$ as its diagonal elements.

By choosing a suitable representation for model error covariance matrix $\mathbf{C}$, we allow a smooth departure from the model to the observed reflectance. The covariance function parameters define this allowed smoothness. As a consequence we achieve a more realistic, although wider, uncertainty estimation of AOT.
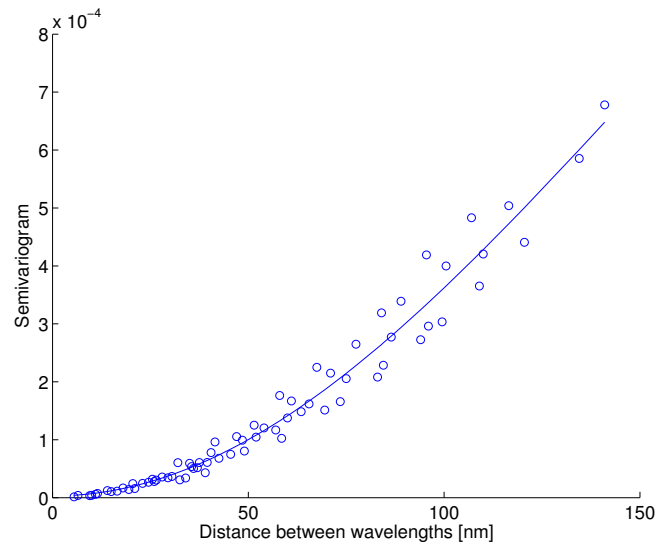
4.2 Empirical semivariogram

The wavelength-dependent correlation structure of the residuals can been estimated by means of an empirical semivariogram. The relationship between theoretical variogram models and the covariance functions of the Gaussian process provides a way to determine the covariance function of model discrepancy (Banerjee et al., 2004). The empirical semivariogram for particular distance d between wavelengths λ_i and λ_j is given as

$$\gamma(d) = \frac{1}{2}\frac{1}{n(d)}\sum_{d=|\lambda_i-\lambda_j|}^{n(d)}\left(\boldsymbol{R}_{\mathrm{res}}\left(\lambda_i\right) - \boldsymbol{R}_{\mathrm{res}}\left(\lambda_j\right)\right)^2, \qquad (16)$$

where $n(d)$ is the number of pairs of wavelengths with the same distance d. In the formula for the particular distance d, the sum of squared residual differences is taken over the set of wavelength pairs with that distance d. The variance of the difference between residuals at any two wavelengths depends only on the wavelength distance.

We have calculated the empirical semivariogram (Eq. 16) for the ensemble of residuals from different orbits. The empirical semivariogram at different wavelength distances d is

Fig. 3. The variance of the residual differences versus wavelength distance is expressed by the empirical semivariogram (circles) and the fitted Gaussian parametric semivariogram model (solid line).

plotted as circles in Fig. 3. This Fig. shows the wavelength-dependent correlation structure of the residual differences. The residuals are similar at nearby wavelengths while the variance of residual differences increases for those wavelength pairs that are further apart.

Next we estimate the parameters of a theoretical parametric semivariogram model that fits the empirical semivariogram. In the literature there are several predefined parametric forms of semivariogram (Banerjee et al., 2004). The commonly used Gaussian variogram model used here is given as

$$\gamma(d) = \begin{cases} \sigma_0^2 + \sigma_1^2\left[1 - \exp\left(-\left(\frac{d^2}{l^2}\right)\right)\right], & \text{if } d > 0 \\ 0, & \text{otherwise} \end{cases}, \quad (17)$$

where $d = |\lambda_i - \lambda_j|$ is the particular distance between wavelengths. In spatial statistics, parameter σ_0^2 is called a nugget, $\sigma_0^2 + \sigma_1^2$ is called a sill and σ_1^2 is a partial sill (Banerjee et al., 2004). The correlation length l defines a scale for the distance between wavelengths where the residuals are still correlated. Parameters l, σ_0^2, and σ_1^2 are tuning parameters of the variogram model that exactly correspond to those of the covariance function in Eq. (14). The fitted Gaussian semivariogram model is plotted in Fig. 3 as a solid curve.

To illustrate the covariance function parameters, Fig. 4 shows how the averaged posterior probability (Eq. 10) changes when the correlation length l in the covariance function (Eq. 14) is increased from 20 to 200. The averaged posterior probability of τ is the weighted mean of the posteriors within the best models. Between any two wavelength bands at the distance of appointed correlation length, the modelled reflectance is allowed to smoothly diverge from the measured reflectance, instead of a close fit at intervening wavelength
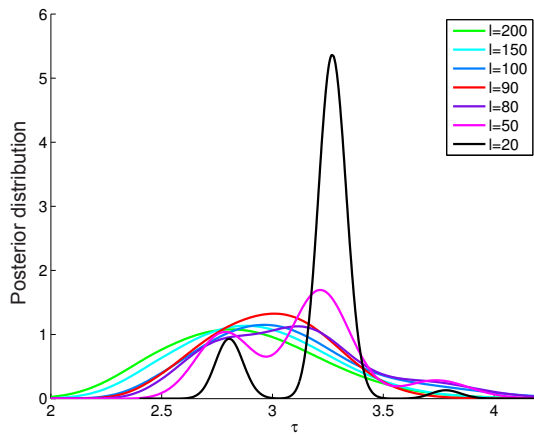
Fig. 4. The effect of correlation length l, defining the model error covariance matrix, on the posterior probability distribution. The posterior is the weighted average on individual posteriors, each fitted with the same model discrepancy term. Each coloured curve corresponds to the averaged posterior obtained with a given value of the correlation length.

bands. That is, the higher the value of correlation length, the smoother the modelled spectral reflectance is allowed to deviate from the measurements. This is related to the higher uncertainty from model discrepancy that increases the uncertainty in the AOT retrieval in our case.

5   Results

The aerosol microphysical model selection, model averaging and model discrepancy modelling are demonstrated here by four examples representing different atmospheric aerosol situations where we expect different dominant main aerosol types. In the examples, we have tested the method using two cases: without the model discrepancy term being included (Eq. 6) and with the model discrepancy included (Eq. 15). Table 2 lists the examples with the appropriate information. The selected pixels are cloud free and over land, and Fig. 9 has MODIS true-colour images for the cases.

Our work is based on the OMI multiwavelength algorithm OMAERO (Torres et al., 2002) introduced in Sect. 2. We used spectral measurements from 14 wavelength bands: 342.5, 367.0, 376.5, 388.0, 399.5, 406.0, 416.0, 425.5, 436.5, 442.0, 451.5, 463.0, 477.0 and 483.5 nm. For practical reasons, there were some differences in our experimental retrieval algorithm compared to the operational OMAERO. We took the surface reflectivity at a given location and date from the database based on Total Ozone Mapping Spectrometer (TOMS) and MODIS data, whereas, for cases over land, the current OMAERO product (version V003) uses the surface albedo climatology based on OMI measurements spanning five years. Also, instead of using the exact operational algorithm for the measurement noise standard deviation $\sigma(\lambda)$,

Table 2. Orbits, dates and locations of example cases.

Orbit	Date	Latitude (deg)	Longitude (deg)
o16415	16 Aug 2007	37.088	22.906
o16546	25 Aug 2007	37.067	22.794
o32258	8 Aug 2010	55.335	36.878
o35754	5 Apr 2011	30.120	13.790
o10153	12 Jun 2006	region	region

we used a simpler SNR (signal-to-noise ratio)-based estimate, with $\mathrm{SNR} = 500$ and $\sigma(\lambda) = R_{\mathrm{obs}}(\lambda)/\mathrm{SNR}$. We used fifty OMAERO aerosol microphysical model LUTs and the modelled TOA reflectance R_{mod} was calculated as in Eq. (3).

The size of the covariance matrix $\mathbf{C}$ in Eq. (15) depends on the number of wavelength bands involved. In our case the dimension of $\mathbf{C}$ is 14×14, which is quite moderate for the matrix operations needed. The empirical semivariogram model described in Sect. 4.2 was used to estimate the parameters defining the covariance matrix $\mathbf{C}$ as $l = 90$, $\sigma_0^2 = 1 \times 10^{-6}$ and $\sigma_1^2 = 0.0004$.

As we are mainly performing a feasibility and method development study, we have used rather conventional choices for prior distributions. We have used only weakly informative priors in all of the test cases. For each individual model fit the prior distribution for AOT parameter τ was set to log-Gaussian with mean value 2 and 700 % standard deviation. This ensured the positivity of the estimated AOT values. For the model choice, uniform prior was used for $p(m)$, i.e. all the models were a priori equally likely. In the example below, we first include all the 50 aerosol microphysical models. Then, up to ten models with a total posterior probability at least 80 % are selected for further analysis.

5.1   Greece forest fires, 2007

During summer 2007 there were massive forest fires in many parts of Greece (Kaskaoutis et al., 2011). We considered two days, approximately at the same location in the Peloponnese, namely, 16 and 25 August 2007 (Table 2). The latter date represents the time when the fires were in their most disastrous phase in that area.

Figure 5 shows observed and modelled reflectances on the left column for 16 August 2007. The observed reflectance is marked by blue dots and the measurement uncertainty as error-bars for 2σ standard error. The posterior distributions of τ on the right-hand column describe the uncertainty of the retrieved AOT, assuming that the associated aerosol microphysical model is correct. The legend shows the relative posterior probability percentage values for each of the aerosol microphysical models involved. The upper row represents the results of model comparison and AOT estimation when the model discrepancy has not been involved. The five most likely models are of weakly absorbing (models with "1" as the first digit) and biomass-burning types ("2" as the
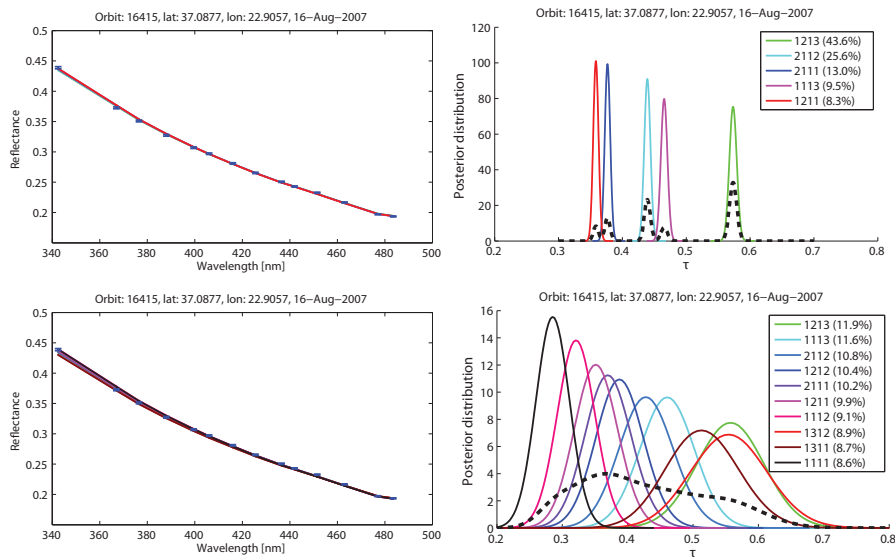
Fig. 5. Greece, 16 August 2007. Upper row: the best five models when model discrepancy is not included. Bottom row: the best ten models when model discrepancy is included. Observed and modelled reflectances on the left, and posterior probability distributions for the AOT parameter τ on the right. The reflectance observations are marked by blue dots and error-bars corresponding to $2 \times$ standard error uncertainty. The modelled reflectance curves on the left match the colours of the individual posterior distributions on the right, although overlaying each other. On the right, the dashed black curve is the averaged posterior distribution over the best models that account for at least 80 % of the total posterior weights of all the models.
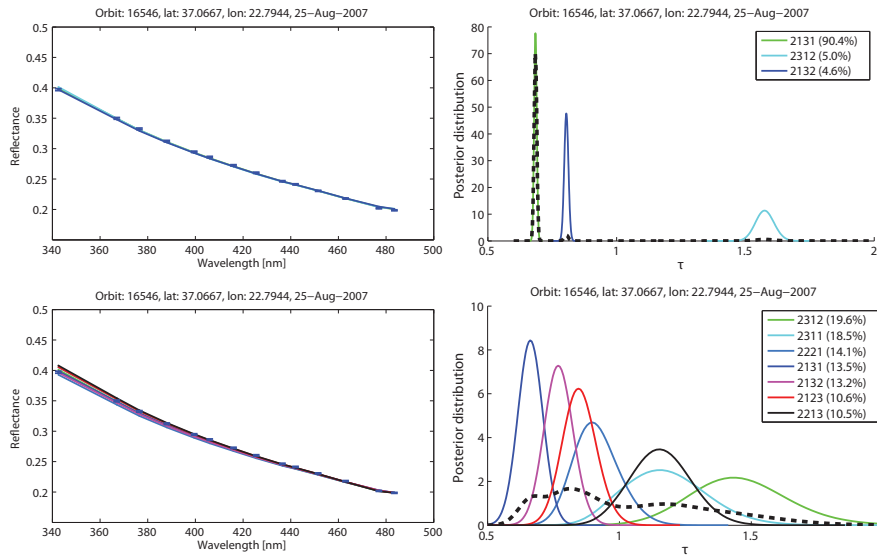
Fig. 6. Greece, 25 August 2007. Upper row: the best three models when the model discrepancy is not included. Bottom row: the best seven models when the model discrepancy is included. See Fig. 5 for more explanation.

first digit). The model "1213" has the strongest contribution to the posterior distribution. The averaged posterior distribution (Eq. 10), plotted as a thick dashed black line, has spread over the posteriors of τ within these five models. The sharp peaked and narrow posterior probabilities indicate low uncertainty of retrieved AOT τ. We suspect that this posterior underestimates the true uncertainty. The lower row shows the results when the model discrepancy has been acknowledged in the fitting procedure. Now there are ten models almost as likely in the averaged posterior distribution of τ. It appears that the uncertainty averaged over models is very wide when the model discrepancy is involved. Also, the single posterior
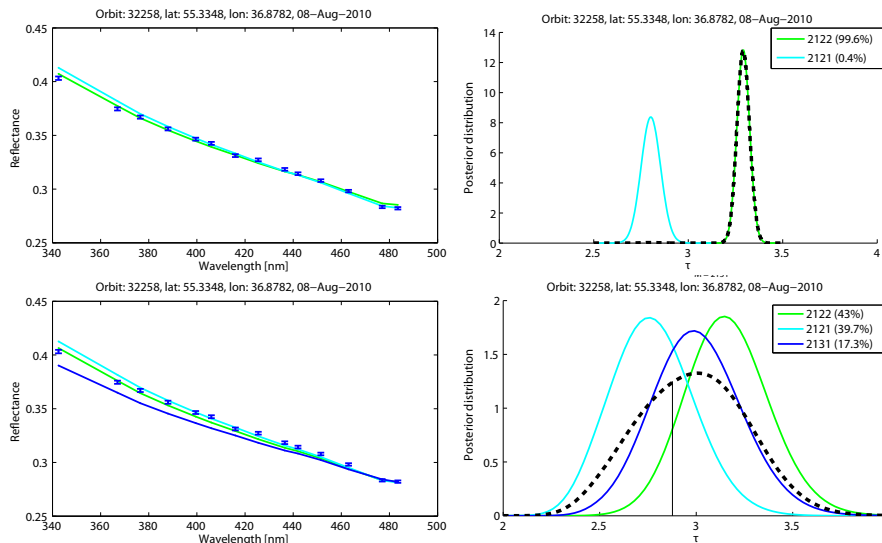
Fig. 7. Wild fires in Moscow, 2010. Upper row: the best two models when the model discrepancy is not included. Bottom row: the best three models when the model discrepancy is included. Also see Fig. 5. The AOT_500 from AERONET is near the mode, i.e. the maximum point, of the averaged posterior as can be seen from the black vertical line (bottom right panel).

distributions of τ within models are clearly broader in this case.

On 25 August, all the best models are of the biomass-burning type (see Fig. 6). Again, when the model discrepancy is not included, the uncertainty shown in the graph on the upper right-hand panel gives the impression of low uncertainty of retrieved AOT. In addition, there is clearly only one best model according to the relative posterior probability. When the model discrepancy is included (Fig. 6, bottom panels), there are seven models almost as likely. This can also be seen by the mean posterior curve when the support is spread over the most likely seven models. When comparing the results of these two days, the aerosol load is larger on the latter day, leading to different aerosol microphysical models chosen and higher AOT estimates.

5.2   Russian wildfires, 2010

There were several wildfires in the western part of Russia from the end of July until August 2010 (Mei et al., 2011; Mielonen et al., 2011). The sample ground pixel from during the day on 8 August 2010 is located near Moscow (Table 2). Figure 7 shows the reflectances and AOT estimates of the best fitted models when the model discrepancy is not included (upper row) and when the model discrepancy is included (lower row). In both cases the two best fitted aerosol microphysical models are the same biomass-burning-type models. When the model discrepancy is not included, the model "2122" is clearly the most likely and the second best model does not have much weight. Because of this, the averaged posterior (dashed black line) covers the posterior curve of τ within model "2122" completely. When fitting the

model to the measured reflectance and acknowledging the model discrepancy term, the ranking between the best two models is not so clear anymore. The posterior distributions of τ under the best models are broad and they now overlap.

There is a ground-based AOT measurement site in Moscow, Moscow_MSU_MO (55° N, 37° E), operated within the Aerosol Robotic Network (AERONET). The Level-2 (Smirnov et al., 2000) AOT at 500 nm from AERONET is 2.88. This ground-based AOT value lies almost in the middle of the possible AOT range (Fig. 7).

5.3   Sahara sandstorm, 2011

In April 2011 there were strong Sahara dust storms (Preißler et al., 2011). At that time, favourable weather conditions transported the dust a long way across the North Atlantic and Europe. We consider here the date of 5 April 2011 (Table 2). The best fitted aerosol microphysical models are of the desert dust type (Fig. 8). With or without model discrepancy, the same two best models show the largest evidence. Also, the best model, "3212", exhibits almost the same relative evidence in both cases, as can be seen from the relative posterior model probability percentage values in the legend boxes. However, when the model discrepancy term is included (lower row) the posterior curves indicate higher uncertainty in the retrieved AOT τ value. The reflectance curves (left column) show visible systematic errors in both of the models. The inclusion of the model discrepancy shifts both posterior curves to the right and widens the uncertainty (right column).
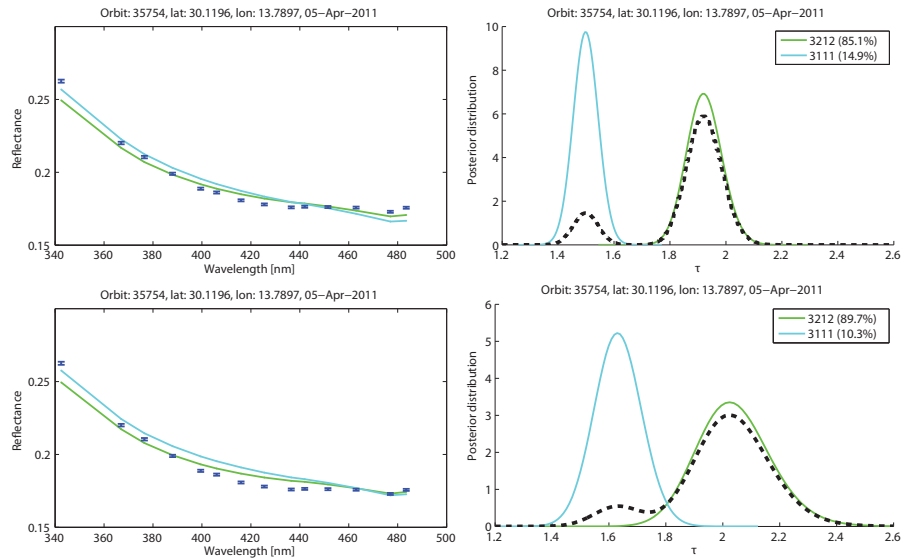
Fig. 8. Sahara dust storm, 2011. Upper row: the best two models when the model discrepancy is not included. Bottom row: the best two models when the model discrepancy is included. See Fig. 5 for more explanation.

5.4 Western Europe 2006

In addition to the single pixel studies above, we have conducted a limited study for a larger set of pixels. We consider one summer day, 12 June 2006, over western Europe with small aerosol amounts. Figure 10 shows the results from operational OMAERO using the best aerosol microphysical model. It also shows results from the new algorithm that uses model averaging and accounts for model discrepancy in a set of 14×7 pixels that also collocate with 3 AERONET ground-based measurements. For the operational algorithm, the model chosen as the best was weakly absorbing aerosol microphysical model number 1211, 1212, or 1213 in all the cases (see Table 1). The grey vertical lines show results with 95 % uncertainty regions from the proposed new algorithm. We see that the variability between pixels in the OMAERO results corresponds very well to the uncertainty attributed to each individual pixel in the new algorithm, suggesting more realistic uncertainty characterisation. We assume here that the variability in OMAERO in the neighbouring pixels is mostly due to the uncertainty in model selection and not in the actual variability of the AOT. In the pixels furthest to the left, that is, in column number 10, the agreement is not as good. Here, we are at the edge of the cloud-masked area. In the three OMAERO values marked with stars the fit did not pass the operational goodness of fit threshold. Also, in many pixels where the operational algorithm does not produce any results, the proposed algorithm seems to give reasonable values.

The grey dots over the vertical lines mark the mode of the posterior distribution (i.e. the maximum a posteriori estimate). These values are typically lower than the operational OMAERO means. However, as the posterior distributions are very asymmetric and skewed to the right, the mean of the posterior would be closer to the OMAERO value. AERONET has three sites in the area of interest that provide data for that day. These sites are Paris (48.52° N, 2.19° E), Fontaineblau (48.24° N, 2.40° E), and Mainz (49.59° N, 8.18° E). The Level-2 AOT_500 from AERONET are marked by red dots in the panel and we show the daily average AOT values, which are 0.16 for Paris, 0.15 for Fontaineblau, and 0.10 for Mainz (http://aeronet.gsfc.nasa.gov/). The AERONET values are in reasonable agreement.

6 Discussion and conclusions

Our aim was to study the additional retrieval uncertainty originating from the need to select an LUT-based aerosol microphysical model from a set of pre-calculated models. We utilised Bayesian statistical methodologies that are general in scope and applicable to a wide range of similar problems. As a particular application example we used operational OMI reflectance measurements from NASA's Aura satellite and modified operational OMAERO aerosol algorithm to estimate AOT parameter. In OMI, the amount of information in the measurements is known to be too small to accurately select the correct aerosol type. Also, in practice there may be several models that explain the observations equally well within the uncertainties of the sensor and forward model.

The use of Bayesian statistical inference provides a unified approach to the quantification of uncertainties originating from the model choice and from parameter estimation. Here, Bayes' formula is applied twice: first, when defining the
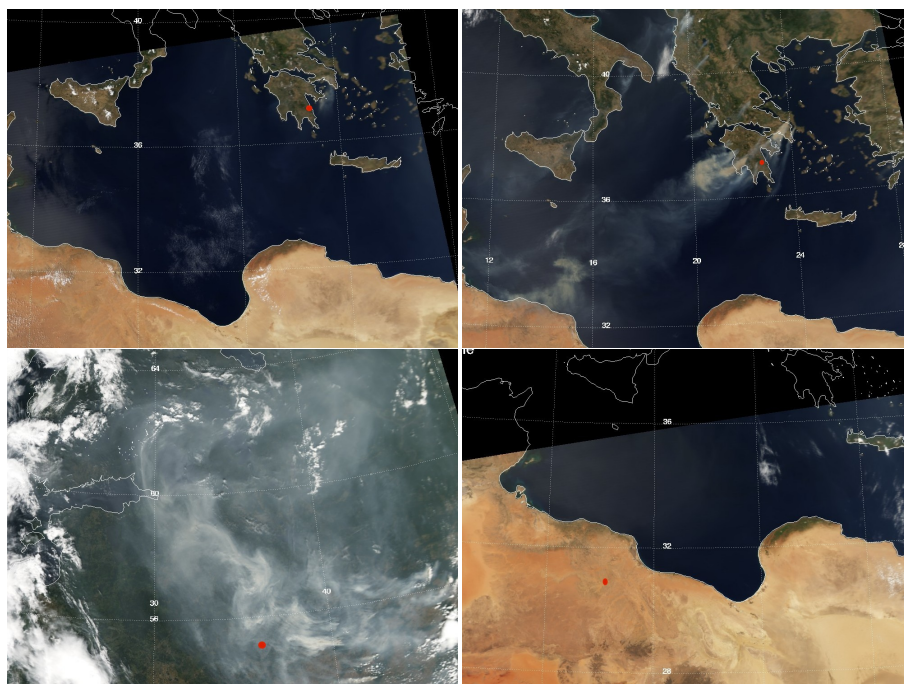
Fig. 9. MODIS true-colour images for the four example cases described in Sects. 5.1–5.3. Top left is Greece, 16 August 2007; top right is Greece, 25 August 2007; bottom left wildfires are in Moscow, 2010; bottom right is Sahara sandstorm, 2011. The approximate centre location of the OMI pixel is marked with small red dot.

posterior distribution of unknown AOT within each aerosol microphysical model, and second, when comparing these models to select the most appropriate aerosol microphysical model. In our particular case there is only one unknown parameter (the AOT) and the actual statistical calculations are rather simple. The obtained posterior probability weights of the models are used to build an averaged model that accounts for the uncertainty in the selection procedure.

The aerosol microphysical model represents some aerosol type with certain size distribution, refractive index and aerosol layer height, and is an approximation of the reality which seldom matches the simplifying assumptions used in model calculation. This causes additional uncertainty in the retrieval. This model discrepancy is taken into consideration by applying a Gaussian process model to explain the characteristics for this model error. The covariance function defining the model discrepancy is estimated empirically from an ensemble of residuals of fits. Adding this model discrepancy term to measurement errors in the fitting procedure will allow wider deviation for the forward model from the observed spectral reflectance.

The applied characterisation of model discrepancy is just one example of different possible ways to explain the systematic uncertainties in forward modelling. The Gaussian process approach allows the modelled reflectances to have smooth deviations from the observed reflectances, and in our studies it was able to account for the typical systematic features in the model residual. Once having estimated it, one

global model error covariance matrix was used in our case for all the test cases we considered. If needed, it would be possible to set up a table of model error covariance parameters depending, for instance, on geographical distribution climatology of models, or even to estimate error model parameters individually for each orbit, etc. Instead of using observed deviances, one way to study model error would be by doing radiative transfer simulations for some fixed atmospheric states and then estimating the model deviations for these situations.

In our examples, all the available aerosol microphysical models were equally probable a priori. Because of the limited information in the measured reflectance, the prior selection of aerosol models for certain locations and times would be necessary, in practice. Prior information about the background aerosol conditions is important, especially, in situations where the amount of aerosols is small, as the different models would be indistinguishable based on the observed reflectance only. In practice these prior weights could be based on aerosol distribution climatologies.

An interesting question is how large the uncertainty caused by the choice of aerosol microphysical model is, compared to other forward model errors and to the measurement uncertainty. Based on our limited experiments, we can not give a definite answer. If we look at the uncertainties of AOT retrievals from individual models based solely on the assumed observation noise, we see that this within-model uncertainty is, in general, significantly smaller than the variability between those models that fit the same observations
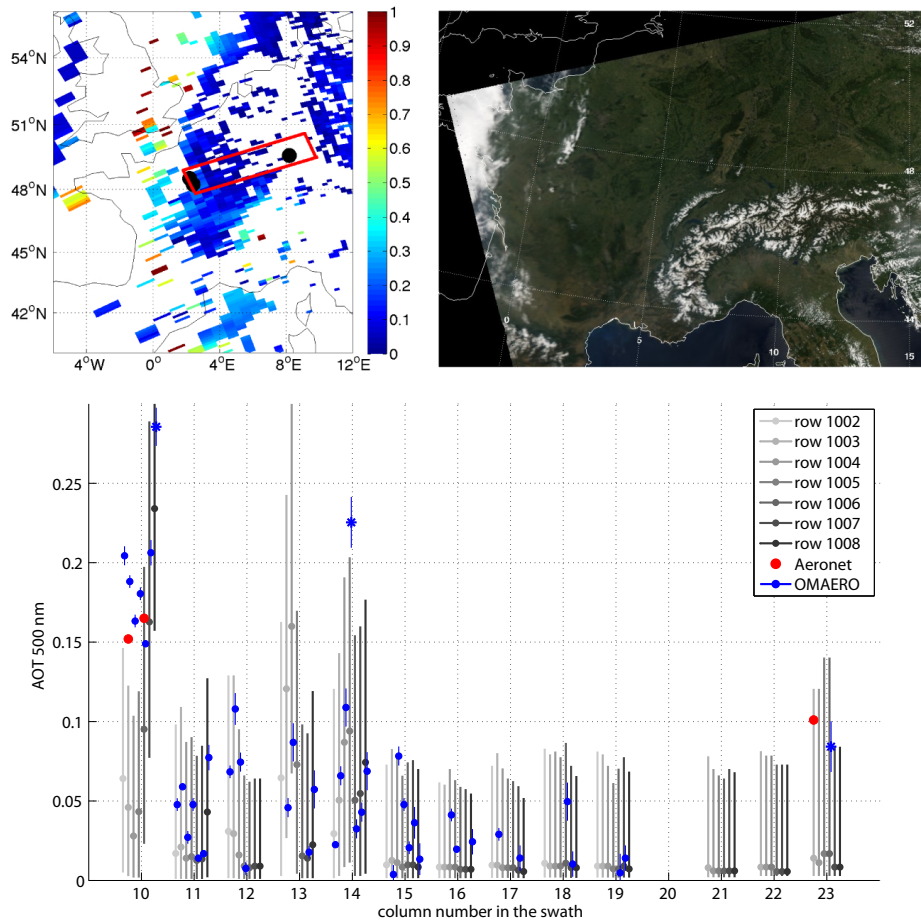
Fig. 10. The top left panel shows AOT from the official OMAERO product. The data are from one orbit (o10153), 12 June 2006. The top right panel shows a MODIS true-colour image from the same day. The lower panel shows AOT values and their uncertainties for a selection of pixels, marked as a red rectangle at the top left. The pixel indexes vary from columns 10 to 23 and rows 1002 to 1008. Each column of seven pixels has been grouped together along the horizontal axis. The vertical axis shows AOT at 500 nm reference wavelength. Not all individual pixels were available, due to cloud contamination. The vertical lines with different shades of grey correspond to the rows of pixels in the satellite orbit and show the results of the retrieval algorithm using model averaging and accounting for discrepancy. The grey dot is the mode of the posterior distribution and the line spans 95% of the posterior probability mass. The blue dot is the best fit model from the operational OMAERO algorithm (also shown on the top left panel) with 2 times standard error uncertainty indicated by the blue line. Three OMAERO values, marked with a star, did not pass the operational goodness of fit threshold. The three red dots show AOT values from collocated AERONET stations, whose locations are shown as black dots in the top left panel. See the text for a discussion of the results.

equally well. The between-model uncertainty reflects the uncertainty that arises from using discrete LUT-based approximate aerosol microphysical model, but it can also include other sources of uncertainties as the model-choice-related uncertainties probably correlate with other forward model approximation errors. The model discrepancy term tries to account for typical non-modelled systematic features in residuals by using a statistical approach. Also, averaging over different models and allowing for statistical model discrepancy will hopefully account for most of the uncertainties in the AOT retrieval. Judging very roughly from the example cases, e.g. Figs. 5–8, we could conclude that uncertainty from the discrete model choice can be from 2 up to 10

times larger that the uncertainty from the assumed measurement noise and this model choice uncertainty may contribute (or can be used to account for) at least one half, and typically even more, of the total uncertainty budget. This issue is also discussed with the multi-pixel example in Sect. 5.4 where we see that our wider uncertainty regions covered the nearby pixel variability in the official product. Overall, this indicates that the results obtained from using just one individual aerosol microphysical LUT model can vastly underestimate the overall uncertainty.

We see potential for this proposed method. If empirical studies of the model discrepancy can be conducted, this information can be included in the retrieval estimation. The

use of Bayesian model selection methodology as described here depends on calculation of the full posterior probability distribution of the quantity of interest, here the AOT, instead of point estimates and uncertainty standard deviations. However, in the OMAERO case, and probably in many similar cases, the numerical calculations needed can be done quite efficiently. The additional computations over the existing non-linear least squares fit include replacing the sum of squares formula with one that has a non-diagonal model error covariance matrix and doing numerical integration for the density function. Another numerical integration would be needed if the minimum mean squares model comparison criteria are replaced with model evidence calculations. In the AOT retrieval case, both these are one dimensional and computationally easy. For a higher-dimensional parameter, some approximation of these calculation would be needed, however. Also, we would like to note that there are similar methods based on a non-diagonal model error covariance matrix that have been applied to aerosol retrievals, for example by using optimal estimation techniques (see, e.g., Thomas et al., 2009; Govaerts et al., 2010).

The methodology introduced is generic and potentially applicable to other instruments, too. For example, the MODIS AOT retrieval algorithm is based on finding the best combination from a set of predefined coarse and fine modes (over ocean) and coarse-dominated and fine-dominated aerosol microphysical models (over land) (Levy et al., 2009). The methodology introduced here could potentially also be applied to MODIS retrieval to take into account the uncertainty in the microphysical model selection.

Our motivation was to improve the model choice process by acknowledging uncertainties from model selection together with measurement uncertainty as well as the aerosol microphysical model discrepancy, by taking advantage of statistical methodologies. We have demonstrated that by a relatively simple additional calculation we can improve the existing OMAERO algorithm to include model selection uncertainty into the retrieval uncertainty estimates. To further quantify the benefits of these additional calculations would need more refined validation and comparisons of different aerosol retrieval products, both satellite and ground based. However, we feel that our study has already demonstrated the importance and the added value of more careful characterisation of modelling uncertainties.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E.: Hierarchical Modeling and Analysis for Spatial Data, Chapman and Hall/CRC Press, Boca Raton, 2004.

Brinksma, E. J., Pinardi, G., Volten, H., Braak, R., Richter, A., Schönhardt, A., van Roozendael, M., Fayt, C., Hermans, C., Dirksen, R. J., Vlemmix, T., Berkhout, A. J. C., Swart, D. P. J., Oetjen, H., Wittrock, F., Wagner, T., Ibrahim, O. W., de Leeuw, G., Moerman, M., Curier, R. L., Celarier, E. A., Cede, A., Knap, W. H., Veefkind, J. P., Eskes, H. J., Allaart, M., Rothe, P., Piters, A. J. M., and Levelt, P. F.: The 2005 and 2006 DAN-DELIONS NO2 and aerosol intercomparison campaigns, J. Geophys. Res., 113, D16S46, doi:10.1029/2007JD008808, 2008.

Curier, R. L., Veefkind, J. P., Braak, R., Veihelmann, B., Torres, O., and de Leeuw, G.: Retrieval of aerosol optical properties from OMI radiances using a multiwavelength algorithm: Application to western Europe, J. Geophys. Res., 113, D17S90, doi:10.1029/2007JD008738, 2008.

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmos. Meas. Tech., 4, 975–1018, doi:10.5194/amt-4-975-2011, 2011.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, 2nd Edn., Chapman & Hall, 2003.

Gershenfeld, N.:The Nature of Mathematical Modeling, Cambridge University Press, 1999.

Govaerts, Y. M., Wagner, S., Lattanzio, A., and Watts, P.: Joint retrieval of surface reflectance and aerosol optical depth from MSG/SEVIRI observations with an optimal estimation approach: 1. Theory, J. Geophys. Res., 115, D02203, doi:10.1029/2009JD011779, 2010.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian Model averaging: A Tutorial, Statist. Sci., 14, 382–417, 1999.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., and Holben, B. N.: Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res., 115, D23209, doi:10.1029/2010JD014601, 2010.

Kaskaoutis, D. G., Kharol, S. K., Sifakis, N., Nastos, P. T., Sharma, A. R., Badarinath, K. V. S., and Kambezidis, H. D.: Satellite monitoring of the biomass-burning aerosols during the wildfires of August 2007 in Greece: Climate implications, Atmos. Environ., 45, 716–726, 2011.

Kennedy, M. C. and O'Hagan, A.: Bayesian Calibration of Computer Models, J. Roy. Stat. Soc. B, 63, 425–464, 2001.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt , P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113, D18308, doi:10.1029/2008JD010290, 2008.

Kokhanovsky, A. A., Deuzé, J. L., Diner, D. J., Dubovik, O., Ducos, F., Emde, C., Garay, M. J., Grainger, R. G., Heckel, A., Herman, M., Katsev, I. L., Keller, J., Levy, R., North, P. R. J., Prikhach, A. S., Rozanov, V. V., Sayer, A. M., Ota, Y., Tanré, D., Thomas, G. E., and Zege, E. P.: The inter-comparison of major satellite aerosol retrieval algorithms using simulated intensity and polar-

ization characteristics of reflected light, Atmos. Meas. Tech., 3, 909–932, doi:10.5194/amt-3-909-2010, 2010.

Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., van den Oord, G. H. J., Bhartia, P. K., Tamminen, J., de Haan, J. F., and Veefkind, J. P.: Science Objectives of the Ozone Monitoring Instrument, IEEE T. Geosci. Remote, 44, 1199–1208, 2006a.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE T. Geosci. Remote, 44, 1093–1101, 2006b.

Levy, R. C., Remer, L. A., Tanré, D., Mattoo, S., and Kaufman, Y. J.: Algorithm for remote sensing of tropospheric aerosol over dark targets from MODIS: Collections 005 and 051: Revision 2, Feb 2009, MODIS Algorithm Theoretical Basis Document, 2009.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmos. Chem. Phys., 10, 10399–10420, doi:10.5194/acp-10-10399-2010, 2010.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., and Maring, H.: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, Ann. Geophys., 27, 2755–2770, doi:10.5194/angeo-27-2755-2009, 2009.

Livingston, J. M., Redemann, J., Russell, P. B., Torres, O., Veihelmann, B., Veefkind, P., Braak, R., Smirnov, A., Remer, L., Bergstrom, R. W., Coddington, O., Schmidt, K. S., Pilewskie, P., Johnson, R., and Zhang, Q.: Comparison of aerosol optical depths from the Ozone Monitoring Instrument (OMI) on Aura with results from airborne sunphotometry, other space and ground measurements during MILAGRO/INTEX-B, Atmos. Chem. Phys., 9, 6743–6765, doi:10.5194/acp-9-6743-2009, 2009.

MacKay, D. J. C.: Bayesian interpolation, Neural Comput., 4, 415–447, doi:10.1162/neco.1992.4.3.415, 1991.

Mei, L., Xue, Y., de Leeuw, G., Guang, J., Wang, Y., Li, Y., Xu, H., Yang, L., Hou, T., He, X., Wu, C., Dong, J., and Chen, Z.: Integration of remote sensing data and surface observations to estimate the impact of the Russian wildfires over Europe and Asia during August 2010, Biogeosciences, 8, 3771–3791, doi:10.5194/bg-8-3771-2011, 2011.

Mielonen, T., Portin, H., Komppula, M., Leskinen, A., Tamminen, J., Ialongo, I., Hakkarainen, J., Lehtinen, K. E. J., and Arola, A.: Biomass burning aerosols observed in Eastern Finland during the Russian wildfires in summer 2010 – Part 2: Remote sensing, Atmos. Environ., 47, 279–287, doi:10.1016/j.atmosenv.2011.07.016, 2012.

Preißler, J., Wagner, F., Pereira, S. N., and Guerrero-Rascado, J. L.: Multi-instrumental observation of an exceptionally strong Saharan dust outbreak over Portugal, J. Geophys. Res., 116, D24204, doi:10.1029/2011JD016527, 2011.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, available at: http://www.GaussianProcess.org/gpml (last access: 11 April 2014), The MITT Press, 2006.

Robert, C. P.: The Bayesian Choice, From Decision-Theoretic Foundations to Computational Implementation, 2nd Edn., Springer, New York, 2007.

Sayer, A. M., Thomas, G. E., and Grainger, R. G.: A sea surface reflectance model for (A)ATSR, and application to aerosol re-

trievals, Atmos. Meas. Tech., 3, 813–838, doi:10.5194/amt-3-813-2010, 2010.

Sayer, A. M., Thomas, G. E., Grainger, R. G., Carboni, E., Poulsen, C., and Siddans, R.: Use of MODIS-derived surface reflectance data in the ORAC-AATSR aerosol retrieval algorithm: Impact of differences between sensor spectral response functions, Remote Sens. Environ., 116, 177–188, doi:10.1016/j.rse.2011.02.029, 2012.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., and Jeong, M.-J.: Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data, J. Geophys. Res. Atmos., 118, 7864–7872, doi:10.1002/jgrd.50600, 2013.

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud-screening and quality control algorithms for the AERONET database, Remote Sens. Environ., 73, 337–349, 2000.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A.: Bayesian measures of model complexity and fit, J. Roy. Stat. Soc. B, 64, 583–639, 2002.

Thomas, G. E., Poulsen, C. A., Sayer, A. M., Marsh, S. H., Dean, S. M., Carboni, E., Siddans, R., Grainger, R. G., and Lawrence, B. N.: The GRAPE aerosol retrieval algorithm, Atmos. Meas. Tech., 2, 679–701, doi:10.5194/amt-2-679-2009, 2009.

Torres, O., Decae, R., Veefkind, J. P., and de Leeuw, G.: OMI Aerosol Retrieval Algorithm, in: OMI Algorithm Theoretical Basis Document: Clouds, Aerosols, and Surface UV Irradiance, 3, version 2, OMI-ATBD-03, edited by: Stammes, P., NASA Goddard Space Flight Center, Greenbelt, Md, 47–71, 2002.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overwiew, J. Geophys. Res., 112, D24S47, doi:10.1029/2007JD008809, 2007.

van Donkelaar, A., Martin, R. V., Spurr, R. J. D. , Drury, E., Remer, L. A., Levy, R. C., and Wang, J.: Optimal estimation for global ground-level fine particulate matter concentrations, J. Geophys. Res. Atmos., 118, 5621–5636, doi:10.1002/jgrd.50479, 2013.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sens. Environ., 120, 70–83, doi:/10.1016/j.rse.2011.09.027, 2012.

Veihelmann, B., Levelt, P. F., Stammes, P., and Veefkind, J. P.: Simulation study of the aerosol information content in OMI spectral reflectance measurements, Atmos. Chem. Phys., 7, 3115–3127, doi:10.5194/acp-7-3115-2007, 2007.

Wagner, S. C., Govaerts, Y. M., and Lattanzio, A.: Joint retrieval of surface reflectance and aerosol optical depth from MSG/SEVIRI observations with an optimal estimation approach: 2. Implementation and evaluation, J. Geophys. Res., 115, D02204, doi:10.1029/2009JD011780, 2010.