A benchmark for testing the accuracy and computational cost of shortwave top-of-atmosphere reflectance calculations in clear-sky aerosol-laden atmospheres

Jeronimo Escribano[1,a], Alessio Bozzo[2], Philippe Dubuisson[3], Johannes Flemming[2], Robin J. Hogan[2], Laurent C.-Labonnote[3], and Olivier Boucher[1]

[1]Institut Pierre-Simon Laplace, CNRS / Sorbonne Université, Paris, France
[2]European Centre for Medium Range Weather Forecasts, Reading, United Kingdom
[3]Laboratoire d'Optique Atmosphérique, Université de Lille / CNRS, Villeneuve d'Ascq, France
[a]now at: Barcelona Supercomputing Center (BSC), Barcelona, Spain

Correspondence: jeronimo.escribano@bsc.es

Abstract. Accurate calculations of shortwave reflectances in clear-sky aerosol-laden atmospheres are necessary for various applications in atmospheric sciences. However computational cost becomes increasingly important for some applications such as data assimilation of top-of-atmosphere reflectances in models of atmospheric composition. This study aims to provide a benchmark that can help in assessing these two requirements in combination. We describe a protocol and input data for 44,080 cases involving various solar and viewing geometries, four different surfaces (one oceanic bidirectional reflectance function and three albedo values for a Lambertian surface), eight aerosol optical depths, five wavelengths, and four aerosol types. We first consider two models relying on the discrete ordinate method: VLIDORT (in vector and scalar configurations) and DISORT (scalar configuration only). We use VLIDORT in its vector configuration as a reference model and quantify the loss of accuracy due to (i) neglecting the effect of polarisation in ~~the~~ DISORT and VLIDORT (scalar) models and (ii) decreasing the number of streams in DISORT. We further test two other models: the 6SV2 model relying on the successive orders of scattering method and FLOTSAM, a new model under development by two of the authors. Typical mean fractional errors of 2.8 and 2.4% for 6SV2 and FLOTSAM are found, respectively. Computational cost depends on the input parameters but also on the code implementation and application as some models solve the radiative transfer equations for a range of geometries while others do not. All necessary input and output data are provided as Supplementary Material as a potential resource for interested developers and users of radiative transfer models.

1   Introduction

Accurate radiative transfer calculations in the Earth's atmosphere are necessary for some applications such as remote sensing of the atmospheric and surface properties, numerical weather prediction, and climate modelling. In this study we are interested

more specifically in the calculation of radiances in the shortwave spectrum in clear-sky aerosol-laden atmospheres of the Earth for reasons explained below. Assuming perfect knowledge of the optical properties of the atmosphere and surface and a plane parallel atmosphere, solving the radiative transfer equation is a very well posed physical problem that essentially reduces to a (not-so-easy) algorithmic and numerical problem. Accurate methods have existed for a long time and testing the accuracy

5 of newly developed methods has been standard practice in the radiative transfer community for some time (de Haan et al., 1987; Stamnes et al., 1988; Spurr, 2008; Kotchenova et al., 2006; Kotchenova and Vermote, 2007; Barlakas et al., 2016; Korkin et al., 2017). Well-used state-of-the-art radiative transfer models generally show agreement within 1% (or better) under most conditions tested. There has also been a number of intercomparison or benchmarking exercises for shortwave radiance calculations largely motivated by the requirements for accurate calculations in ground-based and satellite aerosol retrievals

10 (Kotchenova et al., 2006; Kokhanovsky et al., 2010a; Emde et al., 2015, 2018). A classical benchmarking exercise consists in comparing the model results to well-known solutions for simple cases of a Rayleigh scattering atmosphere (Coulson et al., 1960; Natraj et al., 2009). However such cases are very specific and do not represent the variety of conditions met in the real atmosphere. The International Polarized Radiative Transfer (IPRT) working group of the International Radiation Commission has recently defined a set of more realistic cases (Emde et al., 2015, 2018). Another benchmark is to test the numerical

15 convergence of the scheme when the angular and/or vertical resolutions of the model are increased (de Haan et al., 1987; Ganapol, 2017). However numerical convergence may be difficult to achieve and to assess because the numerical models may experience lower precision when the resolution is increased too much, especially for models coded in single precision. Finally, there have also been benchmarking exercises for aerosol retrievals from observed clear-sky shortwave radiances, which tests simultaneously the accuracy of the radiative transfer model and the aerosol retrieval algorithm (Kokhanovsky et al., 2010b).

20 It should be noted that accounting for the polarisation of light is necessary because it affects the accuracy of the computed ~~intensity~~ radiances (e.g. Kotchenova et al., 2006) but also because some aerosol retrieval algorithms rely on the inversion of polarised radiances (e.g. Tanré et al., 2011).

Methods for computing shortwave radiances by solving the radiative transfer equation ~~that are accurate~~ are ~~also~~ often computationally expensive. In data assimilation or satellite retrieval applications for which radiances have to be computed many

25 times and on many profiles, the computational cost has always been an issue for pragmatic reasons (a numerical model that is too slow to be run on the fastest computer is virtually useless). However, developing fast or very fast radiative transfer models capable of computing radiances (and not only vertical fluxes) ~~has~~ received little attention so far. Exceptions include early attempts to use neural networks for fast and accurate radiative transfer in the longwave spectrum (Chevallier et al., 1998; Krasnopolsky and Chevallier, 2003). It may also be possible to design fast physical algorithms in the longwave spectrum be-

30 cause of the weaker scattering (as compared to the shortwave spectrum) and isotropic emission. In the shortwave spectrum, one can mention recent efforts for fast calculations of cloudy-sky visible radiances (Scheck et al., 2016). However it relies on the use of tabulated calculations (so-called look-up tables), so may only consider a limited set of possible atmospheric inputs. It is noteworthy in that respect that the performance assessment of the newly released SORD (Successive ORDers of scattering) model (Korkin et al., 2017) shows a runtime analysis for the 52 benchmark cases that were tested (their Fig. 6). There is a

35 mode in the distribution of runtimes that is below 0.01 s but also a fairly long tail with occasional runtimes between 60 and

300 s or even longer than 600 s. Computational cost of algorithms for three-dimensional radiative transfer models have been investigated by Pincus and Evans (2009) ~~and~~ are known to be expensive.

Forecasts and reanalyses of atmospheric aerosols have been produced operationally by combining numerical models and aerosol observations in data assimilation systems (Bocquet et al., 2015; Hollingsworth et al., 2008). Aerosol information

5 provided by satellite retrievals is often assimilated in aerosol models both in research (Collins et al., 2001) and operational (Benedetti et al., 2018) contexts. A widely assimilated quantity is the aerosol optical depth (AOD), a column integrated quantity of the interaction of the aerosols with the electromagnetic radiation in the atmosphere. So far, the assimilation of AOD retrievals has been successful in constraining the model variables towards the equivalent measured variables, but as the resolution, complexity and quality requirements of the products vary, possible inconsistencies between the assumptions used for the

10 satellite aerosol retrieval and the model or between the different satellite retrieval algorithms arise. Furthermore, there is some information on aerosol types in the observed reflected solar radiation that may not be included in AOD products and is therefore lost for data assimilation.

An alternative pathway to the assimilation of AOD retrieved from passive satellite measurements is the direct assimilation of these measurements. The main challenge of doing so is basically that the satellite retrieval algorithm has to be replaced by

15 a somehow equivalent procedure inside the data assimilation loop. In particular a shortwave radiative transfer model (RTM) would act as a major piece of the observation operator, and some stringent requirements on accuracy/quality, flexibility and runtime would have to be met. In the case of variational data assimilation, similar requirements would apply to the Jacobian of the RTM. The purpose of this benchmark paper is to assess the accuracy and computational cost of different RTMs to assimilate top-of-atmosphere reflectances to constrain aerosols. We will exclude the assessment of the Jacobian of the RTMs in this work.

20 Successful implementation of this approach could help to improve the performance of the current aerosol data assimilation systems by providing: i) consistency between modelled aerosol type, aerosol vertical profile, atmospheric conditions and the radiative transfer calculation, ii) potentially better characterisation of the error covariance matrices for the observations, iii) flexibility when there is a change in satellite instrument, and iv) capability of assimilating simultaneously several different sources of aerosol information. Additionally, developments in this direction could help improving the numerical weather pre-

25 diction (NWP) models by providing a functional shortwave observation operator to the data assimilation system. This latter also opens the possibility of improvements in the data assimilation of shortwave radiances for the atmospheric and surface properties (Martin et al., 2014), including cloudy atmospheres (Martin and Hasekamp, 2018). These authors show that the adjoint framework can be computationally efficient, but do not present the computation cost of the SHDOM and FSDOM models that they use.

30 To our knowledge, only a limited number of studies have been able to assimilate shortwave radiances in atmospheric composition models. Weaver et al. (2007) have assimilated the Moderate Resolution Imaging Spectroradiometer (MODIS) radiances in the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model coupled to the Arizona Radiation Transfer Code, a look-up-table-based model, showing that although the assimilation improves the model AOD, it does not outperform MODIS AOD products (Remer et al., 2005). Drury et al. (2008) were able to assimilate MODIS radiances in the GEOS-Chem

35 chemical transport model (coupled with the LIDORT radiative transfer model) with the resulting aerosol field comparing better

against the Aerosol Robotic Network (AERONET) data than MODIS (collections 4 and 5) products for a 45 day period over North America. Wang et al. (2010, 2012) have also assimilated MODIS radiances in the GEOS-Chem model (coupled with the VLIDORT model) showing improvements in the AOD computations and top-down emission estimates. Xu et al. (2013) also used GEOS-Chem and VLIDORT to estimate sources of gaseous compounds by assimilating MODIS radiances and evaluated

5   against OMI (Ozone Monitoring Instrument) SO_2 and NO_2 retrievals and Multiangle Imaging Spectro-Radiometer (MISR) and AERONET AOD.

Migliorini (2012) discusses some conditions where the assimilation of radiances is equivalent to the assimilation of the retrieval products. For this to happen, the observation operator has to be reasonably (i.e., within the errors of the variable) approximated by a linear function of the control variable, and the background information has to be chosen such that the same

10   information content of the observations is propagated in both assimilations. In our case, the first condition is usually assumed true, but the second condition is difficult to implement and to evaluate.

For operational data assimilation purposes, the computational cost of the radiance observation operator is crucial (Weng, 2007). While the computational burden of assimilating satellite radiances is expected to be larger than that of assimilating retrievals, it has not yet been specified, to our knowledge, how fast such an observation operator has to be. A rough estimate

15   could be made using the computing time of the infrared radiance calculation as a guideline. For example, in the European Centre for Medium-range Weather Forecasts (ECMWF) data assimilation system, the infrared radiance observation operator takes around 0.1 ms per profile and channel in a highly parallelised system and roughly three times more including the tangent linear and adjoint code calculations. Taking into account the current infrastructure of the Copernicus Atmospheric Monitoring Service (CAMS) at ECMWF, a first estimate of the maximum computing time for shortwave radiance calculation would be

20   approximately 10 ms per profile and channel, but ideally of the order of 1 ms (or about ten times slower than the current RTTOV infrared calculations) in this particular ECMWF/CAMS infrastructure. Likewise, there is a lack of knowledge of the required accuracy of the shortwave radiance calculations needed. A first order estimate would be an accuracy similar to the measurement or retrieval error, but this could depend on the data assimilation configuration and the measurement characteristics (in terms of multi-spectral and multi-angle viewing capability, quality of the prior including that of the surface reflectance, etc.). There may

25   also be a variety of RTMs for different atmospheric conditions (e.g., clear- or cloudy-sky) or wavelengths (ultraviolet, visible or near-infrared).

How computational speed is achieved depends on the desired application. Some methods, such as the discrete ordinate method or the Monte Carlo method (e.g., Emde et al., 2015, 2018), can compute radiances for a set of viewing geometries (given one sun geometry) in a single simulation. Other methods, such as the successive orders of scattering (e.g., Lenoble et al., 2007),

30   require one simulation for each viewing geometry (again given one sun geometry). The latter is not necessarily a disadvantage for applications where a different atmospheric profile of optical properties is associated with every viewing geometry. When the problem to be solved requires a lot of tasks to be processed in parallel and frequent thread synchronisation, as is the case in many three-dimensional atmospheric models, it is the slowest tasks that matter. Therefore the tail of the runtime distribution matters as much, if not more, than the average or median runtime across atmospheric conditions and viewing geometries. In the

35   specific case of aerosol radiance assimilation in NWP models, it may nevertheless be possible to leave out the conditions and

geometries that take too long, either by eliminating them in advance or by stopping them on the fly during the first iteration if the corresponding tasks exceed a threshold time. In the end, there is likely to be some trade off to be accepted between accuracy and computational cost of the radiance calculations.

Keeping this introduction in mind, the objective of this study is to provide a benchmarking tool for evaluating shortwave
5   radiances in the clear-sky aerosol-laden atmosphere under a wide range of aerosol conditions. The benchmarking is designed to assess both the accuracy and computational cost of radiative transfer models. Although data assimilation of aerosol radiances is the key motivation for this work, we report computational cost both for single and multiple geometries to make the scope more general. We provide all necessary input and output data for this benchmarking as Supplementary Material in NetCDF and ACSII format so that it becomes an available resource for model developers and potential model users. The protocol is
10 described in the next section. It is followed by a test of the protocol in different RTM and some conclusions.

2 Benchmark protocol

The detailed benchmark protocol is provided as Supplementary Material. Only the main features are summarised here. We consider five wavelengths spanning the shortwave spectrum: 470, 550, 660, 865 and 1024 nm. These wavelengths correspond to typical wavelengths in shortwave atmospheric windows used by satellite instruments such as MODIS, POLarization and Di-
15 rectionality of the Earth's Reflectances (POLDER) / Polarisation et Anisotropie des Réflectances au sommet de l'Atmosphère, couplées avec un Satellite d'Observation emportant un Lidar (PARASOL), Visible Infrared Imaging Radiometer Suite (VIIRS), Along Track Scanning Radiometer (ATSR) and Advanced Along Track Scanning Radiometer (AATSR). They offer a range of conditions because molecular scattering and the impact of polarisation are more important at the shorter wavelengths.

2.1 Vertical profiles

20 The molecular and aerosol profiles are described below. It should be noted that we assume the atmosphere to be plane-parallel, i.e. there is no correction applied for sphericity of the Earth's atmosphere.

2.1.1 Molecular profile

We use a midlatitude summer atmospheric profile (McClatchey et al., 1971) and provide the layer altitude (km), pressure (hPa), temperature (K), relative humidity (%), water vapour mixing ratio (ppmv), ozone mixing ratio (ppmv), humid and dry
25 air density (g/cm^3) on 50 levels ranging from $z = 0$ to $z_{max} = 120$ km. The molecular scattering and absorption optical depths are provided for the 49 corresponding layers from the surface to the top-of-atmosphere using the radiative transfer code GAME (Dubuisson et al., 2004, 2006). In particular, the absorption optical depths have been calculated with GAME for the five selected wavelengths by averaging the gaseous absorption (for H_2O and O_3) in each layer from the k-distribution parameterisation used in the GAME code. Note that gaseous absorption is weak at the selected wavelengths so this approximation remains valid.
30 The molecular depolarisation ratios are taken from Bodhaine et al. (1999), with values of 0.0288567, 0.0283241, 0.0279199, 0.0275716 and 0.0274591 for the five selected wavelengths. There is no obligation to use the 49 layers provided in the protocol.

Indeed it could be interesting to perform the RTM calculations on a reduced vertical grid to decrease the computational cost. In case a different vertical grid is used, we ask that the vertically-integrated molecular scattering and absorption optical depths be conserved.

2.1.2 Aerosol profile

5   We define the vertical profile of the aerosol extinction coefficient (σ^{aer}) as an exponential function with a height scale H of 2 km, normalized to unity when integrated between $z = 0$ and $z = z_{\mathrm{max}} = 120$ km, that is:

$$\sigma^{\mathrm{aer}}(z) = \frac{e^{-z/H}}{H\left(1 - e^{-z_{\mathrm{max}}/H}\right)}. \tag{1}$$

The aerosol optical depth of a layer i, τ_i, is computed as:

$$\tau_i = \int_{\underline{z_i}}^{\overline{z_i}} \sigma^{\mathrm{aer}}(z)\,\mathrm{d}z \tag{2}$$

10   where Δz_i is the layer thickness, that is, the difference between the altitudes of the layer top ($\overline{z_i}$) and the layer bottom ($\underline{z_i}$). The average aerosol extinction coefficient for layer i then satisfies:

$$\sigma_i^{\mathrm{aer}} = \tau_i / \Delta z_i. \tag{3}$$

We provide σ_i^{aer} and τ_i for the 49 layers defined in our atmospheric profile, but these quantities can be easily recomputed for any vertical grid. These values are normalised to an AOD of 1, so they have to be multiplied by the actual AOD of each case.

15 ## 2.2 Aerosol types

Four aerosol types have been defined. Aerosol size distributions are assumed to be log-normal with the parameters shown in Table 1. Refractive indices are taken from the literature (see Table 1) and the aerosols are assumed to be internally mixed, with a volume-weighting rule. The four aerosol types defined in this benchmark have been chosen to represent a variety of size and optical properties observed in the atmosphere. Two fine-mode aerosols are defined: industrial scattering (aer_is) is sulphate

20 aerosol with a relative humidity of 50%, while industrial absorbing (aer_ia) is a mixture of the OPAC (Hess et al., 1998) aerosol type INSO (insoluble), SOOT (soot) and WASO (water soluble, at 50% of relative humidity). The industrial absorbing aerosol is similar to an organic matter aerosol type. We also consider a monomodal dust aerosol and a bimodal sea-salt aerosol (aer_ss).

2.3 Aerosol optical properties

Aerosol optical properties were computed by the authors using their own Mie routines and results were cross-checked for

25 consistency. The Mie routine used to generate the input data was rewritten in Fortran based on the implementation of Toon and Ackerman (1981) and was compiled with the Intel Fortran compiler in quadruple precision. The calculations are monochromatic (i.e., no integration was performed on the wavelength). The aerosol size distribution is integrated over the size range

reported for each type in Table 1 with a Gaussian quadrature of 10,000 points in logarithm of the radius. The number of terms used to compute the sums of the a_n and b_n coefficients is approximated as a function of the Mie parameter ($x = 2\pi r/\lambda$) following Wiscombe (1980).

For each aerosol type, we provide the aerosol optical depths at the 440, 670, 865 and 1024 nm wavelengths corresponding
5 to the selected AOD at 550 nm. The aerosol extinction coefficient (although not necessary to conduct the RTM calculations), single scattering albedo and asymmetry parameter are also given.

The phase matrix is computed and provided at 50,000 points evenly spaced in scattering angle from 0 to 180°. As the particles are assumed to be spherical, only the S_{11}, S_{12}, S_{33} and S_{34} elements of the phase matrix are non-zero and provided. We also decompose the phase matrix in generalised spherical functions (de Rooij and van der Stap, 1984; de Haan et al.,
10 1987) as needed for the vector version of the discrete ordinate method. For this, the phase matrix is integrated with a 8,000 points Gaussian quadrature over the cosine of the scattering angle and we provide the first 750 moments of that decomposition for the six commonly-used elements (α_1, α_2, α_3, α_4, β_1, β_2). For completeness we also provide the corresponding Legendre coefficients (by definition the lth Legendre coefficient equals the lth Legendre moment multiplied by $2l + 1$). Thus we provide all necessary information on aerosol optical properties for usual RTM so that no further processing of the aerosol optical
15 properties should be required. Of course, the RTM that neglect polarisation should only consider the S_{11} term of the phase matrix or the α_1 Legendre moments.

The accurate reconstruction of the original phase function from the Legendre decomposition (S_{11}) cannot be ensured if a too low number of Legendre moments is used to approximate it. As for the vertical resolution, there is no obligation for this benchmarking exercise to represent the exact phase function and it may be interesting to approximate it to strike a balance
20 between accuracy and computational cost. The minimum number of Legendre moments for which the approximated phase function is positive depends on the aerosol type and the wavelength (basically on how large the forward peak is). For the industrial scattering and industrial absorbing aerosols, a minimum of 4 to 8 moments is required (depending on the wavelength). For dust aerosol, the minimum is between 14 moments (at 1024 nm) and 40 moments (at 470 nm), while for sea salt, the minimum is between 144 moments (at 1024 nm) and 348 moments (at 470 nm).
25 Figure 1 shows the phase function and the linear depolarisation ratio for the four aerosol types considered in this benchmark.

2.4 Surface reflectance

Two surface reflectance models are selected, a Lambertian and an oceanic bidirectional reflectance distribution function (BRDF) model. For the Lambertian model, three surface albedos were chosen, namely 0, 0.05 and 0.1. The spectral depen-
30 dence of the surface reflectance is ignored, i.e., the albedo is assumed the same for the five wavelengths considered. The second surface reflectance model is the oceanic glitter model from Mishchenko and Travis (1997). This model is suited for both full vector radiative transfer models (i.e., including polarisation) and for the scalar approximations.

Wind speed is set to 10 m s^{-1} but the Mishchenko and Travis (1997) code assumes an isotropic Gaussian distribution of (oceanic) surface slopes, so the wind orientation is not taken into account. The salinity is set to 34.3‰ and water refractive

Aerosol type		Industrial Scattering	Industrial Absorbing		Sea-Salt		Dust
Mixture input		sulphate (50% RH)	WASO (50% RH)	INSO	SOOT	fine	coarse	———
N particles		1	1200	0.1	8300	70	3	1
ρ_{wet} (g cm^{-3})		1.264	1.437	2	1	2.16	2.16	2.61
r_min (μm)		0.00611	0.005	0.005	0.005	0.03	0.03	0.03
r_max (μm)		24.44	25	25	25	20	20	20
r_mod (μm)		0.0259	0.0262	0.471	0.0118	0.1992	1.992	0.29
σ_g		2.24	2.24	2.51	2.0	1.9	2.0	2.0
Ref Index (real)	470 nm	1.435	1.439	1.530	1.750	1.357	1.357	1.530
	550 nm	1.431	1.437	1.530	1.750	1.354	1.354	1.530
	670 nm	1.427	1.436	1.530	1.750	1.351	1.351	1.530
	865 nm	1.422	1.430	1.520	1.750	1.347	1.347	1.520
	1024 nm	1.418	1.428	1.514	1.760	1.345	1.345	1.515
Ref Index (imag)	470 nm	5.184 E-08	2.649 E-03	8.000 E-03	4.530 E-01	3.519 E-09	3.519 E-09	6.690 E-03
	550 nm	5.367 E-08	3.179 E-03	8.000 E-03	4.400 E-01	2.980 E-09	2.980 E-09	5.200 E-03
	670 nm	7.491 E-08	3.709 E-03	8.000 E-03	4.340 E-01	2.668 E-08	2.668 E-08	4.040 E-03
	865 nm	6.722 E-07	6.331 E-03	8.000 E-03	4.333 E-01	3.895 E-06	3.895 E-06	4.300 E-03
	1024 nm	2.539 E-06	8.391 E-03	8.000 E-03	4.410 E-01	2.353 E-05	2.353 E-05	4.500 E-03

Table 1. Aerosol physical and optical properties for the different aerosol types considered in the benchmark. σ_g is the geometric standard deviation. The refractive indices are linearly interpolated from the OPAC database (Hess et al., 1998). Dust refractive index are linearly interpolated from Woodward (2001).

indices are interpolated, for each wavelength, from the data provided in Hale and Querry (1973) as inputs to the Mishchenko and Travis (1997) routine (available at https://www.giss.nasa.gov/staff/mmishchenko/brf/). The effect of shadowing is also taken into account.

The convention for the azimuthal angle used in the Mishchenko and Travis (1997) routine is the opposite of the convention that we use below to define geometries, so the coupling between the atmospheric RTM and the oceanic code has to be done carefully.

2.5 Geometries and configurations

We have selected 8 cases for aerosol loadings in the atmosphere, with AOD at 550 nm ranging between 0 (no aerosol) to 2 (high aerosol loading). These values refer to the 550 nm wavelength and the corresponding AODs can be estimated for the other wavelengths depending on the aerosol types. A table with the equivalences in AOD, for each aerosol type and wavelength is also provided in the benchmark protocol.
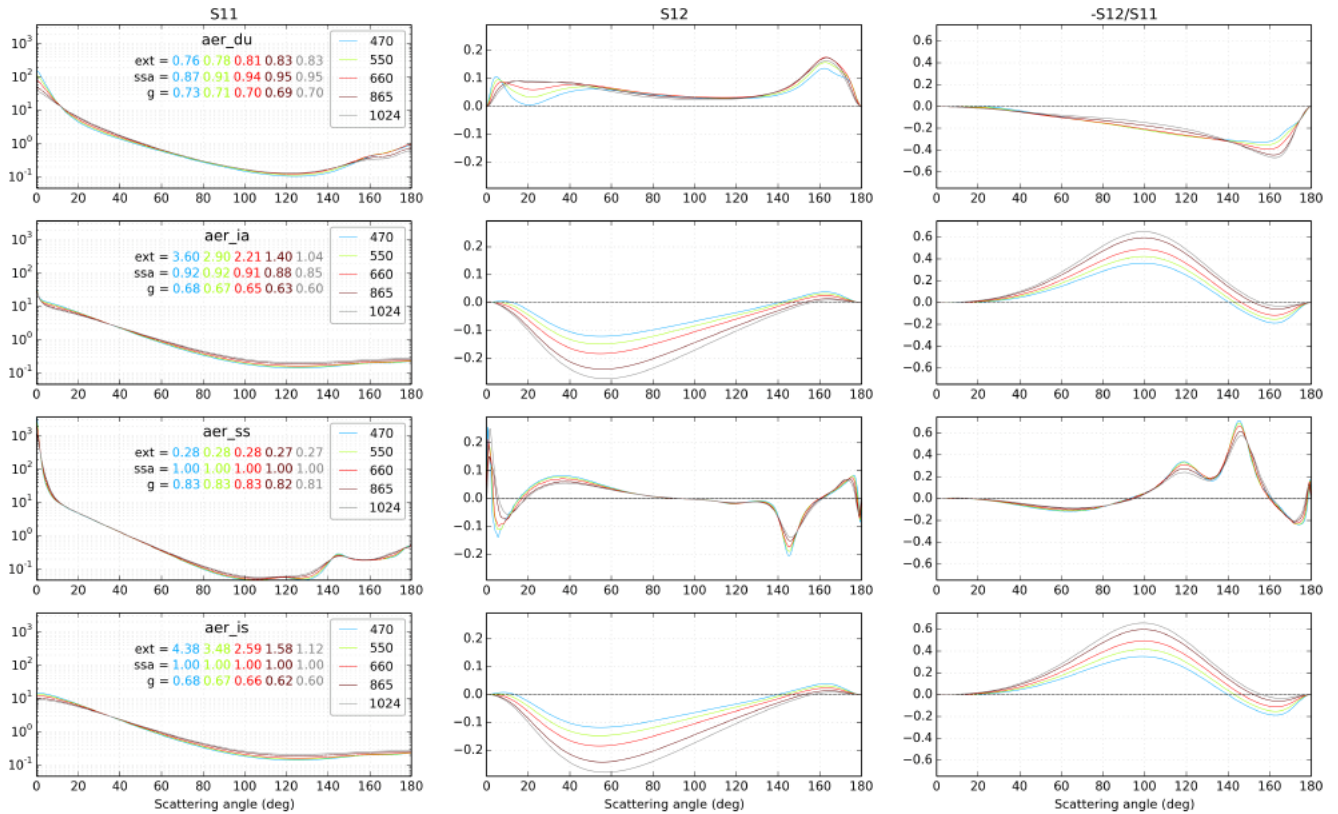
Figure 1. First (i.e., S_{11}, left panels), second (i.e., S_{12}, middle panels) element of the phase matrix and linear depolarisation ratio (i.e., $-S_{12}/S_{11}$) (right panels) as a function of the scattering angle (°) for the four aerosol types (rows), where *aer_du* stands for dust, *aer_ia* for industrial absorbing, *aer_ss* for sea-salt and *aer_is* for industrial scattering. The colour code corresponds to the five wavelengths (470, 550, 660, 865 and 1024 nm). The extinction coefficient (ext, in m^2 g^{-1}), single scattering albedo (ssa) and asymmetry parameter (g) are also indicated in each panel. The phase function is normalised to 2 (i.e., $\int_0^\pi S_{11}(\Theta) \sin(\Theta) \, d\Theta = 2$).

We have further selected four solar and four viewing zenith angles, θ_s and θ_v, from 0° (zenith) up to 60°. Given the provided surface reflectance models, and the assumption of spherical (hence randomly-oriented) aerosols in the atmosphere, the radiance at the top of the atmosphere is only dependent on the difference between the incident and reflected azimuthal angle, and not on each value taken independently. We have chosen five azimuthal angles between $\phi_r = 0°$ (backscatter when solar and viewing zenith angles are equal) and $\phi_r = 180°$ (forward scattering when solar and viewing zenith angles are equal). The different cases are summarised in Table 2. Removing the redundancies in aerosol types for AOD=0 and azimuthal angle for $\theta_s = \theta_v = 0°$, this results in 11,020 cases for the BRDF surface and 33,060 cases for the Lambertian surfaces (see Table 3), which gives a grand total of 44,080 cases.

Geometrical conditions	
Solar zenith angles, θ_s	0°, 20°, 40°, 60°
Solar viewing angles, θ_v	0°, 20°, 40°, 60°
Azimuthal angles, ϕ_r	0°, 45°, 90°, 135°, 180°
Environmental conditions	
Wavelengths	470, 550, 660, 865 and 1024 nm
Aerosol optical depth at 550 nm	0, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 2
Surface	Oceanic BRDF and Lambertian reflectance (0.0, 0.05, 0.1)
Aerosol profile	Exponential profile with height scale of 2 km
Atmospheric profile	Midlatitude summer
Aerosol types	sea-salt (aer_ss), dust (aer_du), industrial scattering (aer_is), in-dustrial absorbing (aer_ia)

Table 2. Summary of the proposed geometrical and environmental conditions used in the simulations.

2.6 Measured variables

We aim to provide a standardised dataset to evaluate the possibility of assimilating satellite radiances in operational forecasts of atmospheric composition. Thus, two variables are important to estimate for each case: the simulated reflectance (or radiance) that would be observed by the satellite, and a "sound" estimate of the radiative transfer code runtime.

5   We will describe the accuracy of the models in terms of the reflectance, ρ_I, of the intensity (i.e., the first Stoke component), which is related to the radiance, I, through the relationship:

$$\rho_I = \frac{\pi I}{I_s \cos(\theta_s)} \tag{4}$$

where I_s is the incident solar flux at the top of the atmosphere (conveniently taken equal to one for reflectance calculations), and θ_s is the solar zenith angle.

10   The runtime may be difficult to estimate and compare between different models. We aim to measure the runtime of the computations excluding input/output operations, as typical applications would embed the RTM code in a larger operational system. The runtime of the RTM should be measured in a single-thread configuration. Ideally, the simulations should be performed on the same core of the computer and with the same system load conditions (except for the RTM) as other threads on the operating system (in particular with a high input/output load) could hamper the correct timing of the model. Multiple
15   repetitions of the computations can be performed to get better estimates.

Given a set of atmospheric and aerosol conditions, some models can compute, almost without extra cost, the radiances for several viewing geometries, while other models require multiple simulations to achieve the same. For this protocol, this would introduce a difference of a factor of circa 80 (the number of viewing geometries) in the reported runtimes. Depending on the

application, this feature of some models may or may not be relevant. For this reason, we present here the runtimes of the tested models in two configurations: one taking advantage of the simultaneous output for multiple viewing geometries, and another one without this feature. In the second configuration (called *ind* in Section 3), each simulation outputs only one geometry while in the first configuration (called *mult* in Section 3), it outputs as many values as possible within the specifications of the

5 protocol. The number of multiple outputs per simulation will be detailed in the next Section.

3 Models

3.1 Reference model: VLIDORT

The VLIDORT model is an independent implementation of the discrete ordinate method based on earlier work by Siewert (2000a, b). It solves the radiative transfer equation by using the three elements (3×3) approximation of the Stokes vector

10 (Spurr, 2008). A detailed user's guide is available (Spurr, 2014). One interesting feature of this model is that it also provides Jacobians with respect to the model inputs (not used in this study). In this model, it is possible to choose how many elements are computed in the Stokes vector. We will refer to the *vector* runs when the 3×3 approximation of the RT problem is used; and we will refer to the *scalar* runs when only the first element of the Stokes vector is used in the computations (i.e., when polarisation of electromagnetic radiation is not accounted for). The protocol vertical resolution with 49 layers is used. It should

15 be noted that the discrete ordinate method shows a numerical instability for zenith viewing angle θ_v of 0° or a relative azimuthal angle ϕ_r of 90°. These angles are therefore prescribed at $\theta_v = 0.0001°$ and $\phi_r = 89.98°$ instead to circumvent this problem. We use the model with 32 streams and the TMS correction (Nakajima and Tanaka, 1988) is activated. VLIDORT vector is our reference model.

3.2 DISORT

20 The DISORT model (Stamnes et al., 1988) is a flexible and widely used RTM. It solves the scalar approximation of the radiative transfer equation in a plane parallel atmosphere by using the discrete ordinate method. In this study, we use the DISORT code, version 4, available from http://lllab.phy.stevens.edu/disort/, and we have coupled the oceanic BRDF following the code developer instructions. We use the model with 32 streams and the TMS correction (Nakajima and Tanaka, 1988) is activated. It should be noted that most of this code is written in single precision, but important routines (as the computation of

25 the Gaussian nodes) are written in double precision. We use 32 streams in our main configuration but sensitivity results for 4, 8, 16, 32, 64, 96 and 128 streams are also presented below. The protocol vertical resolution with 49 layers is used. Similar to VLIDORT, the DISORT code has numerical instabilities which manifest themselves when the solar and viewing zenith angles (θ_s and θ_v) are equal. In such cases, we perturb the cosine of the viewing zenith angle by 8×10^{-5}.

Given a solar zenith angle, the DISORT model can provide outputs for multiple viewing geometries. In this study, the number

30 of geometries for the *mult* timing configuration is the number of ϕ_r cases multiplied by the number of θ_v cases (i.e., $4 \times 5 = 20$).

Although we know the DISORT model to be too slow for online data assimilation of aerosol radiances, we find it useful to document here the trade-off between accuracy and computing time as a function of the number of streams used in the calculations. Figure 2 shows both the Mean Fractional Error and the average computing time for 4, 8, 16, 32, 64, 96 and 128 streams. The Mean Fractional Error is shown relative to our reference VLIDORT (vector) and is of the order of 1 or 1.5%

5   for the Lambertian and oceanic BRDF surface cases, respectively, as soon as the number of streams exceeds 8. Most of that discrepancy is due to DISORT neglecting polarisation as the Mean Fractional Error against VLIDORT (scalar) is significantly smaller: <0.1% for the Lambertian surface case and a <1% for the surface BRDF case.

There is a big gain in accuracy when increasing the number of streams from 4 to 8 and little further gain in accuracy beyond 16 streams. A deterioration in accuracy is observed beyond 96 streams in the case of the BRDF surface, which may be due to

10   the DISORT model being coded in single precision float. While the dependence of model accuracy on the number of streams is well known (Stamnes and Swanson, 1981), we find it useful to confront the loss of accuracy with the gain in computational cost. Computing time increases rapidly with the number of streams (faster than the cube of the number of streams beyond 16 streams). Using 16 streams might, therefore, be an interesting trade-off between accuracy and computational cost.

15   ## 3.3   6SV2

The 6SV2 model is a radiative transfer model based on the successive order of scattering (SOS) method (Vermote et al., 1997). It solves the radiative transfer equation by using the three elements approximation (3×3) of the Stokes vector. The accuracy of the code against previous benchmark and Monte Carlo codes has been reported to be better than 1% (Kotchenova et al., 2006; Kotchenova and Vermote, 2007). We have used version 2.1 of the model. Two configurations are available: a standard accuracy

20   configuration that consists of 83 Gaussian points for the phase function, 30 vertical levels, 25 zenith angles and 181 azimuth angles, and a high accuracy configuration that consists of 121 Gaussian points for the phase function, 50 vertical levels, 75 zenith angles and 181 azimuth angles. We will present results only from the standard accuracy, which is 2 to 70 faster than the high accuracy configuration, but will comment briefly on the high accuracy configuration in Section 4.2.

Molecular scattering and absorption vertical profiles are embedded in the code. Both are internally computed according to

25   the atmospheric profile as part of "core" SOS routines. It is therefore not appropriate to prescribe our own molecular scattering and absorption profiles, so we use the 6SV2 defaults instead. The difference in the total molecular scattering optical depth is less than 0.26% with respect to the files provided in this protocol. We have coupled the BRDF calculations to the 6SV2 model but it should be noted that the atmosphere-surface coupling of the oceanic BRDF model in 6SV2 does not use the full surface reflection matrix but only the first column of it (i.e., the column that relates reflected I with incoming I, Q and U).

30   The 6SV2 model can only compute one viewing geometry in each model call (i.e., the *ind* configuration is exactly the same as the *mult* configuration).
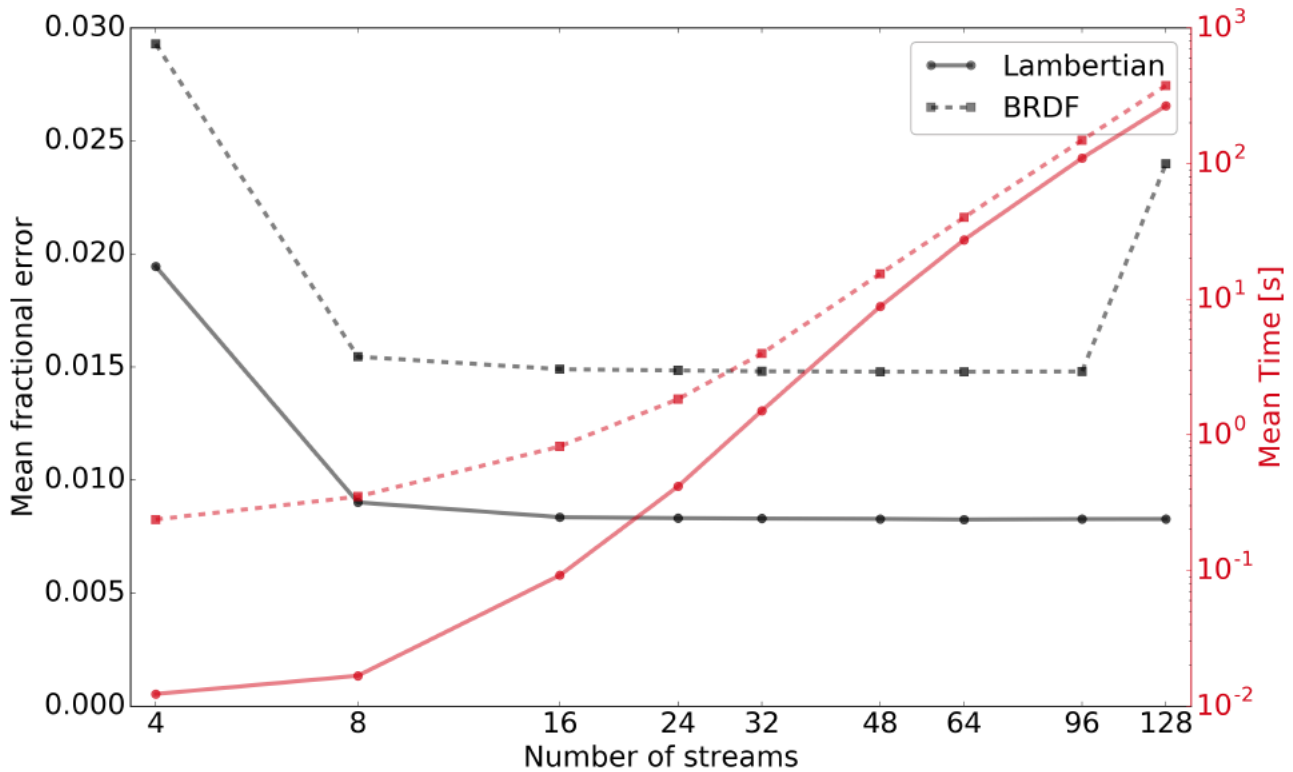
Figure 2. Accuracy (in black) and computing times (in red) for the DISORT model as a function of the number of streams used. The Lambertian and oceanic BRDF surface cases are shown with solid and dashed lines, respectively. The accuracy is shown in terms of Mean Fractional Error against VLIDORT (vector). The computing times are an average for 20 geometries and were estimated on a processor AMD Opteron 6378, 2.4GHz. Please note the logarithmic scales for the number of streams used in the DISORT calculations (x-axis) and the timings (right vertical axis).

3.4 FLOTSAM

We also test in this study a preliminary version of the Forward-Lobe Two-Stream Radiance Model (FLOTSAM), which is being developed at the European Centre for Medium-Range Weather Forecasts (ECMWF). Unlike look-up-table-based solar RTM, FLOTSAM permits an arbitrary layering of different types of scatterers, including clouds, aerosols and molecules. The model has been designed to be fast enough to be used in iterative assimilation and retrieval schemes. It exploits the fact that particle phase functions typically contain a "narrow forward lobe" associated with scattering angles of less than a few degrees, and a "wide forward lobe" associated with scattering angles of the order of 15°. This enables coupled ordinary differential equations to be written down for radiation in the quasi-direct solar beam (including the narrow forward lobe) and a wider beam of radiation propagating away from the Sun due to scattering by the wide forward lobe. Additionally, the upwelling and downwelling diffuse fluxes are computed using the two-stream approximation. FLOTSAM may, therefore, be thought of as the solar analogue of the infrared Two-Stream Source Function technique of Toon et al. (1989). To facilitate the use of

FLOTSAM in variational assimilation and retrieval schemes, it has been coded in C++ using version 2.0 of the Adept library (Hogan, 2014, 2017), which enables the Jacobian to be easily computed via Adept's fast automatic differentiation capability. FLOTSAM has been incorporated into the Cloud, Aerosol and Precipitation from Multiple Instruments using a Variational Technique (CAPTIVATE) retrieval scheme, which will provide operational products from the forthcoming EarthCARE satellite
5 (Illingworth et al., 2015).

FLOTSAM is designed to be used in iterative schemes in which multiple calculations are performed with different profiles of optical properties but for the same viewing geometry. This is implemented by allowing the user to perform the set-up calculations once for a particular solar zenith angle, viewing zenith angle and viewing azimuth angle, and then to re-use the information in subsequent radiance calculations for different optical profiles. We assess the computational cost for both the *ind*
10 and the *mult* configurations, with the latter consisting of $4 \times 4 \times 5 \times 8 = 640$ cases involving 80 geometry set-up calculations and each separate geometry being used with 8 different optical depths.

4   Results

4.1   Methods and statistics

The protocol has been tested for the 5 model versions described above: VLIDORT (vector), VLIDORT (scalar), DISORT,
15 6SV2 and FLOTSAM. As explained in Section 2.6, the accuracy is quantified in terms of monochromatic reflectances at the top of the atmosphere and the accuracy is computed against the vector version of the VLIDORT model unless stated otherwise.

With as many as 44,080 cases considered in this study, we have to find original ways to visualize the results as explained below. Along with the average reflectance computed for each model, the following statistical measures have also been computed:

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (m_i - r_i)^2}$$

$$\text{Mean Fractional Bias (MFB)} = \frac{2}{N} \sum_{i=1}^{N} \frac{m_i - r_i}{m_i + r_i}$$

$$\text{Mean Fractional Error (MFE)} = \frac{2}{N} \sum_{i=1}^{N} \left| \frac{m_i - r_i}{m_i + r_i} \right|$$

where N is the number of experiments, m_i is the reflectance of the tested model and r_i is the reflectance of the reference
25 model for case i. By definition, the Mean Fractional Bias can range between -2 and 2, with a best score of zero, while the Mean Fractional Errors can range between 0 and 2, with a best score of zero. Both values are multiplied by 100 with respect to the definitions above in the Tables presented below. We also state the percentages of cases where the relative error of the reflectances is above defined relative error thresholds. These diagnostics, which could be detailed for particular subsets of our 44,080 cases, can help to decide if and under which conditions a particular RTM could be useful, according to the required
30 accuracy of the reflectance calculations.

Runtimes of the models are presented in an absolute measure (seconds), and no reference is needed. Finally, we will comment on possible trade-offs between accuracy and computing time.

4.2 Accuracy

Figure 3 shows histograms of the simulated reflectances of the VLIDORT (vector) model at the top of the atmosphere for all or various subsets of the 44,080 cases. These reflectances form the reference simulations for this study. The width of the histograms (or *violins*) in Figure 3 are proportional to the number of cases in each bin of the reflectance (y-axis) discretization. The multiple panels (and multiple violins) are computed with all the simulations, but filtered by the value indicated on the x-axis. Therefore, they can be interpreted as an analogous to marginal density probability functions of the set of simulations. Figure 3 shows a wide range of reflectances (from near 0 to almost 0.5) and known dependencies of the reflectance with respect to some of the input variables such as the surface albedo (Lambertian case), the AOD or the wavelength in this range of the spectrum.

A first comparison of the relative error is shown in Figure 4 for two scalar models (with otherwise the same degree of sophistication as the reference model). Figure 4 shows the histograms of the relative errors, $(m-r)/r$, with m being the tested model and r the reference model (the vector version of VLIDORT). It can be seen that for both models, DISORT and the scalar version of VLIDORT, the errors are mostly below 5%. They are centred on zero for the Lambertian surface, but a negative bias is observed for the BRDF case. Comparison of scalar VLIDORT against vector VLIDORT provides a direct measure of the relative errors due to neglecting polarisation in the model. As expected, we observe larger relative errors for the shorter wavelengths (MFE = 1.8% at 470 nm; MFE = 0.5% at 1024 nm), and for the accumulation-mode aerosols (industrial scattering and industrial absorption have MFE ≈ 1.2%) in contrast to coarse-mode aerosols (dust and sea-salt have MFE ≈ 0.6%). The oceanic BRDF function used in this study also enhances the errors of the scalar versions as these neglect the polarising effect of the surface. Larger relative errors are shown for the dark surface case (Lambertian surface albedo equal to zero have MFE ≈ 1.2%) because the reference reflectances are smaller than for their 0.05 and 0.1 Lambertian counterparts (with MFE ≈ 0.7% and MFE ≈ 0.5%, respectively). The error statistics of the VLIDORT (scalar) and DISORT models are very similar, with noticeable differences only for the BRDF case (Figure 4 and Table 3). The scalar approximation may thus be appropriate for some combinations of wavelengths and surface types but not for others. We further note, that for both types of surfaces, the scalar VLIDORT simulations outperform DISORT simulations when they are compared against the vector version of VLIDORT (which is not surprising), but in both cases the overall Mean Fractional Error is larger than 0.7 % and the Mean Fractional Bias is negative.

Figure 5 and Table 4 show the relative errors of the other two models used in this study, 6SV2 and FLOTSAM, against VLIDORT (vector). As the FLOTSAM model does not include polarisation, it is not expected to outperform the scalar approximations exemplified in the DISORT and VLIDORT (scalar) comparisons. However, this is the case for the 470 nm cases, as can be seen from the corresponding violin plot of Figure 5. For 6SV2, the relative error shows a mode around zero, while for FLOTSAM, the shapes of the violin plots are more similar to those of the scalar approximations shown in Figure 4. The reason for the unexpected dependence of the 6SV2 model errors with the wavelength is not completely clear, but it could be

related to the technical difficulties in prescribing user-specific molecular scattering and absorption profiles in the 6SV2 code (see Sections 2.1.1 and 3.3). In contrast to Figure 4, Figure 5 shows larger relative errors for sea salt aerosols, for both models. While larger errors in Figure 4 were related to the polarisation/depolarisation effects from aerosol particles (as shown in the right column of Figure 1), errors of 6SV2 and FLOTSAM could be related to the more pronounced forward scattering peak of

5   the sea salt aerosols. There is no clear dependence of the relative errors with respect to the viewing and solar geometries but, as for Figure 4, larger errors are observed for larger AOD and for the surface BRDF case.

Synthetic error statistics are shown in Table 4. Both the 6SV2 and FLOTSAM models have comparable Mean Fractional Errors, but these are larger than those of DISORT and VLIDORT (scalar). Some difference concerning the type of surface can be appreciated with 6SV2 performing better in the case of Lambertian surfaces. It should be noted that using the high

10   accuracy configuration of 6SV2 instead of the standard accuracy configuration reduces the Mean Fractional Error from 1.5 to 1.1% for the Lambertian surface case, but does not improve this score for the BRDF surface case. The Mean Fractional Bias is negative, and it is of similar magnitude for both models. However, 6SV2 presents larger biases over the two types of surfaces (and the sign of the bias differs for the BRDF and Lambertian surfaces). In contrast, FLOTSAM shows virtually no bias for the Lambertian surface cases and a smaller bias than 6SV2 for the BRDF cases. An important difference with the scalar models is

15   the presence of a heavy tail in the relative error distribution, which can be inferred from the last rows of Table 4.

4.3   Timing

20   In this Section, we present statistics of computing times for three of the models used in this study. All the computing times presented below were estimated on an Intel i5-4690, 3.5 GHz workstation running Linux (kernel 2.6.32), GNU C++ and Fortran compiler version 4.4.7, on a single processor and with exclusive use of the workstation. We have compiled the models using their default compilation flags, ie, "-O" for 6SV2; "-O3" and "-march=native" for FLOTSAM and "-O3" for DISORT.

We start this section with the 6SV2 model which can only be used in *ind* configuration. Figure 6 shows the violin plots of the

25   computing time of the 6SV2 model that spans as much as two orders of magnitude. Clearly, the BRDF cases are more expensive than the Lambertian surface cases, with most computing times ranging between ca. 0.7 and 4 s for the former. This appears to be due to multiple calls to the BRDF routine. It may be possible to optimise this routine or recode it in a more efficient way, but we have not attempted to do so. We also note an increase in computing time with increasing solar zenith angle (θ_s) and AOD, and with decreasing wavelength. This is consistent with a larger number of scattering orders being necessary in such cases.

30   Cases for coarse mode aerosols (especially sea salt aerosols) are more expensive than cases for accumulation mode aerosols, with some calculations reaching 5 s for sea salt aerosols. We do not have an explanation as to why the cases for industrial absorbing aerosols are more expensive than those for industrial scattering aerosols.

	All		BRDF		Lambertian	
	DISORT	VLIDORT (scalar)	DISORT	VLIDORT (scalar)	DISORT	VLIDORT (scalar)
Reference (VLIDORT vector) Mean	0.114		0.108		0.116	
Model Mean	0.114	0.114	0.108	0.108	0.116	0.116
RMSE	0.002	0.002	0.003	0.002	0.002	0.002
Mean Fractional Bias (%)	-0.44	-0.42	-0.82	-0.89	-0.31	-0.26
Mean Fractional Error (%)	0.99	0.91	1.48	1.29	0.83	0.78
N	44,080	44,080	11,020	11,020	33,060	33,060
% of exps. with errors > 0.5%	55.09	51.34	68.43	61.79	50.64	47.86
% of exps. with errors > 1%	35.27	34.47	48.97	47.71	30.70	30.05
% of exps. with errors > 2%	14.63	14.71	25.95	27.04	10.85	10.60
% of exps. with errors > 2.5%	9.09	9.08	17.18	18.12	6.39	6.07
% of exps. with errors > 5%	0.77	0.28	2.18	0.60	0.30	0.17
% of exps. with errors > 7.5%	0.29	0.00	0.97	0.00	0.07	0.00
% of exps. with errors > 10%	0.18	0.00	0.68	0.00	0.02	0.00

Table 3. Statistical measures of DISORT and scalar VLIDORT reflectances against vector VLIDORT (first row). Errors of the last rows are defined as the absolute value of the relative error, using the VLIDORT (vector) reflectance as the reference. VLIDORT is the configuration with 32 streams and the TMS correction activated. N refers to the number of cases.

As discussed above, the DISORT and FLOTSAM models offer efficiency savings when multiple profiles are computed together. In the case of DISORT, multiple viewing geometries can be computed for one solar zenith angle. In the case of FLOTSAM, a smaller saving is that the geometry set-up calculations may be computed once for a given solar/viewing geometry and then re-used for multiple atmospheric profiles with different optical properties. The runtime of these models is, in general, not sensitive to the viewing geometry, wavelength, aerosol type or AOD. However, there are some differences between the two types of surface reflectance. Table 5 shows a summary of the measured runtimes for the three tested models. Computing times are indicated for the two measurement procedures (*ind* and *mult*) and for the two types of surface reflectance (BRDF and Lambertian). In the case of FLOTSAM, we also report the "one-off set-up" cost, which in the case of the BRDF involves creating a four-dimensional look-up table. We stress that this is invoked only once per call to the executable, so in an operational data assimilation system performing many millions of radiance calculations it would increase the executable runtime by only a very small fraction. The one-off set-up costs are not shown for the 6SV2 and DISORT models as they are not isolated from the rest of the computations.

For the 6SV2 model, values of Table 5 are in concordance with Figure 6 with runtimes of the order of seconds. The DISORT model average runtime is smaller than that of 6SV2. Similarly to 6SV2, there is an important dependence of the DISORT

	All		BRDF		Lambertian	
	6SV2	FLOTSAM	6SV2	FLOTSAM	6SV2	FLOTSAM
Reference (VLIDORT vector) Mean	0.114		0.108		0.116	
Model Mean	0.114	0.114	0.102	0.107	0.117	0.116
RMSE	0.007	0.008	0.014	0.012	0.003	0.006
Mean Fractional Bias (%)	-0.58	-0.49	-5.60	-2.01	1.09	0.02
Mean Fractional Error (%)	2.78	2.41	6.60	4.01	1.51	1.88
N	44,080	44,080	11,020	11,020	33,060	33,060
% of exps. with errors > 0.5%	71.30	80.21	82.56	89.86	67.54	77.00
% of exps. with errors > 1%	45.84	63.70	71.08	79.76	37.42	58.35
% of exps. with errors > 2%	30.04	39.91	54.44	61.91	21.91	32.58
% of exps. with errors > 2.5%	25.04	32.02	48.57	53.80	17.20	24.76
% of exps. with errors > 5%	14.01	11.89	34.50	26.05	7.18	7.17
% of exps. with errors > 7.5%	8.25	4.70	26.00	11.59	2.34	2.40
% of exps. with errors > 10%	5.63	1.97	20.20	5.56	0.77	0.77

Table 4. Same as Table 3 but for the 6SV2 and FLOTSAM models.

runtime on the type of surface considered. The extra cost of computing 20 profiles, instead of one profile, can be estimated as the difference between the *mult* and *ind* cases. For DISORT, this additional cost is about 0.07 s for the Lambertian case and about 0.22 s for the BRDF case, which correspond to 25% and 18% of the total *mult* runtime. This would benefit an application that requires simultaneous calculation of many viewing geometries. The spread of DISORT runtimes indicated in Table 5 arises

5 because DISORT efficiently decreases the number of internal computations for symmetric geometries (solar zenith angle equal to zero). For the remaining cases of Table 5, there is almost no spread in the DISORT runtime.

Table 5 shows the runtimes of the FLOTSAM model. Firstly, it is noted that the standard deviation of the reported values is small (around one order of magnitude less than the values). This is because FLOTSAM performs the same number of operations for every case, thus there is no added value of showing histograms of runtimes. For this model, it can be seen that

10 the one-off set-up cost of the BRDF is a little under 3 s, but in the *ind* case, subsequent radiance calculations are 3–4 orders of magnitude faster than the other two models tested, and the BRDF calculation is only slightly slower. If many profiles have to be computed with the same observation geometry, then the geometry set-up cost is invoked only once and the average cost per profile is reduced. This is illustrated in the *mult* case in which 640 profiles are computed, which involves 80 different geometries. Excluding the model set-up time, the computation of a single profile takes around 0.3 to 0.4 ms in the *ind* case,

15 and around $(99\,\text{ms})/640 = 0.16\,\text{ms}$ in the *mult* case.

		FLOTSAM	6SV2	DISORT
one-off set-up	Lambertian	2.87×10^{-5} (1.99×10^{-6})		
	BRDF	2.76 (5.28×10^{-3})		
ind	Lambertian	3.05×10^{-4} (1.84×10^{-6})	1.06 (0.87)	0.22 (0.18)
	BRDF	3.70×10^{-4} (1.49×10^{-5})	1.90 (0.95)	0.99 (0.18)
mult	Lambertian	9.90×10^{-2} (6.23×10^{-4})		0.29 (0.16)
	BRDF	9.96×10^{-2} (5.63×10^{-4})		1.21 (0.16)

Table 5. Average runtimes in seconds. The standard deviation is shown in parenthesis. The *mult* runtimes account for 640 profiles for FLOTSAM, 20 profiles one profile per model call for 6SV2. The one-off set-up costs of FLOTSAM are invoked only once per instance of the executable, no matter how many radiances are subsequently calculated, so are the same for the *ind* and *mult* experiments.

5   Summary and conclusions

We have presented in this study a comprehensive benchmarking methodology and protocol, with an unprecedented number of cases combining different viewing geometries, aerosol optical depths, aerosol types, wavelengths and surface types. The VLIDORT model in its vector configuration (i.e., considering polarisation) is used the reference model. Preliminary results

5   in terms of accuracy and computing time are presented ~~here~~ for three models: the well-known models DISORT and 6SV2 on one side, and FLOTSAM, a new fast model under development by two of the authors of this study. We have chosen the VLIDORT model (in vector configuration) as the reference for this study. All models perform better when using Lambertian surface reflectance than when using an oceanic BRDF. All the models perform better under low AOD, and the scalar models have lower accuracies at shorter wavelengths. For the Lambertian surface cases, the Mean Fractional Errors of the models are

10   0.8, 1.5 and 1.9% for DISORT, 6SV2 and FLOTSAM, respectively. The BRDF cases show a larger Mean Fractional Errors with 1.5, 6.6 and 4% for DISORT, 6SV2 and FLOTSAM, respectively.

The DISORT and 6SV2 models show comparable computing time, between tenths of a second and seconds, but DISORT can provide the solution at multiple viewing geometries for each atmospheric condition, while 6SV2 cannot. FLOTSAM is very fast, with computing times of much less than a millisecond per profile. However it should be noted that this model is

15   still under development and its accuracy could improve further. Moreover, all models present a tail of cases with larger errors and/or computing times. It could be interesting to characterise these better to investigate if screening out those cases can be an option in data assimilation applications. Also, a fast tangent linear and adjoint computation is required for the integration of a RTM in a variational data assimilation system.

The protocol, input and output data are available as a potential resource for interested developers and users of RTM willing

20   to benchmark their models.

Code and data availability. The detailed protocol is available in PDF format. All inputs necessary to the benchmark are available as ASCII text files. The surface BRDF is available as a Fortran routine from the M. Mishchenko website https://www.giss.nasa.gov/staff/mmishchenko/brf/. The VLIDORT reference results are available in NetCDF format.

5   *Author contributions.* JE designed the benchmarking exercise, selected the aerosol models, performed the 6SV2 and DISORT simulations, processed the results and drafted the manuscript. PD prepared the vertical molecular profiles. OB, JE, LL and AB performed the Mie calculations. LL performed the VLIDORT simulations. RH designed the FLOTSAM model and performed the corresponding simulations, while AB added FLOTSAM's BRDF capability. All authors contributed to the analysis and the manuscript.

  Competing interests. The authors do not declare any competing interest.

10   *Acknowledgements.* The authors acknowledge support from the Copernicus Atmosphere Monitoring Service implemented by ECMWF on behalf of the European Commission. The authors thank Didier Tanré for his interesting comments to the draft of the manuscript. The authors would also like to thank the 6SV2 development team for their help in resolving small issues with the model. Computing was partly performed on the ESPRI data and computing centre of the Institut Pierre-Simon Laplace.

References

Barlakas, V., Macke, A., and Wendisch, M.: SPARTA – Solver for Polarized Atmospheric Radiative Transfer Applications: Introduction and application to Saharan dust fields, Journal of Quantitative Spectroscopy and Radiative Transfer, 178, 77–92, https://doi.org/10.1016/j.jqsrt.2016.02.019, 2016.

5 Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Rémy, S., Basart, S., Boucher, O., Brooks, I. M., Menut, L., Mona, L., Laj, P., Pappalardo, G., Wiedensohler, A., Baklanov, A., Brooks, M., Colarco, P. R., Cuevas, E., da Silva, A., Escribano, J., Flemming, J., Huneeus, N., Jorba, O., Kazadzis, S., Kinne, S., Popp, T., Quinn, P. K., Sekiyama, T. T., Tanaka, T., and Terradellas, E.: Status and future of numerical atmospheric aerosol prediction with a focus on data requirements, Atmospheric Chemistry and Physics, 18, 10 615–10 643, https://doi.org/10.5194/acp-18-10615-2018, https://www.atmos-chem-phys.net/18/10615/2018/, 2018.

10 Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmospheric Chemistry and Physics, 15, 5325–5358, https://doi.org/10.5194/acp-15-5325-2015, 2015.

Bodhaine, B. A., Wood, N. B., Dutton, E. G., and Slusser, J. R.: On Rayleigh optical depth calculations, Journal of Atmospheric and Oceanic
15 Technology, 16, 1854–1861, https://doi.org/10.1175/1520-0426(1999)016<1854:ORODC>2.0.CO;2, 1999.

Chevallier, F., Chéruy, F., Scott, N. A., and Chédin, A.: A neural network approach for a fast and accurate computation of a longwave radiative budget, Journal of Applied Meteorology, 37, 1385–1397, https://doi.org/10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2, 1998.

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., Lamarque, J., and Zender, C. S.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, Journal of Geophysical Research: Atmospheres, 106,
20 7313–7336, https://doi.org/10.1029/2000JD900507, 2001.

Coulson, K., Dave, J. V., and Sekera, Z.: Tables related to radiation emerging from a planetary atmosphere with Rayleigh scattering, Berkeley, U.S., University of California Press, 1960.

de Haan, J. F., Bosma, P. B., and Hovenier, J. W.: The adding method for multiple scattering calculations of polarized light, Astronomy and Astrophysics, 183, 371–391, 1987.

25 de Rooij, W. A. and van der Stap, C. C. A. H.: Expansion of Mie scattering matrices in generalized spherical functions, Astronomy and Astrophysics, 131, 237–248, 1984.

Drury, E., Jacob, D. J., Wang, J., Spurr, R. J. D., and Chance, K.: Improved algorithm for MODIS satellite retrievals of aerosol optical depths over western North America, Journal of Geophysical Research: Atmospheres, 113, D16 204, https://doi.org/10.1029/2007JD009573, d16204, 2008.

30 Dubuisson, P., Dessailly, D., Vesperini, M., and Frouin, R.: Water vapor retrieval over ocean using near-infrared radiometry, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2004JD004516, 2004.

Dubuisson, P., Roger, J. C., Mallet, M., and Dubovik, O.: A Code to Compute the Direct Solar Radiative Forcing: Application to Anthropogenic Aerosols during the ESCOMPTE Experiment, pp. 127–130, A. Deepak Publishing, Hampton, Va, 2006.

Emde, C., Barlakas, V., Cornet, C., Evans, F., Korkin, S., Ota, Y., Labonnote, L. C., Lyapustin, A., Macke, A., Mayer, B., and Wendisch, M.:
35 IPRT polarized radiative transfer model intercomparison project – Phase A, Journal of Quantitative Spectroscopy and Radiative Transfer, 164, 8–36, https://doi.org/10.1016/j.jqsrt.2015.05.007, 2015.

Emde, C., Barlakas, V., Cornet, C., Evans, F., Wang, Z., Labonotte, L. C., Macke, A., Mayer, B., and Wendisch, M.: IPRT polarized radiative transfer model intercomparison project – Three-dimensional test cases (phase B), Journal of Quantitative Spectroscopy and Radiative Transfer, 209, 19–44, https://doi.org/10.1016/j.jqsrt.2018.01.024, 2018.

Ganapol, B.: A 1D radiative transfer benchmark with polarization via doubling and adding, Journal of Quantitative Spectroscopy and Radiative Transfer, 201, 236–250, https://doi.org/10.1016/j.jqsrt.2017.06.028, 2017.

Hale, G. M. and Querry, M. R.: Optical Constants of Water in the 200-nm to 200-μm Wavelength Region, Appl. Opt., 12, 555–563, https://doi.org/10.1364/AO.12.000555, 1973.

Hess, M., Koepke, P., and Schult, I.: Optical Properties of Aerosols and Clouds: The Software Package OPAC, Bulletin of the American Meteorological Society, 79, 831–844, https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

Hogan, R. J.: Fast reverse-mode automatic differentiation using expression templates in C++, ACM Trans. Math. Softw., 40, 26:1–26:16, 2014.

Hogan, R. J.: Adept 2.0: a combined automatic differentiation and array library for C++, https://doi.org/10.5281/zenodo.1004730, 2017.

Hollingsworth, A., Engelen, R. J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J. W., Morcrette, J.-J., Rayner, P., Peuch, V.-H., Rouil, L., Schultz, M. G., Simmons, A. J., and The Gems Consortium: Toward a monitoring and forecasting system for atmospheric composition, Bulletin of the American Meteorological Society, 89, 1147–1164, https://doi.org/10.1175/2008BAMS2355.1, 2008.

Illingworth, A. J., Barker, H. W., Beljaars, A., Chepfer, H., Delanoë, J., Domenech, C., Donovan, D. P., Fukuda, S., Hirakata, M., Hogan, R. J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T. Y., Nishizawa, T., Ohno, Y., Okamoto, H., Oki, R., Sato, K., Satoh, M., Wandinger, U., and Wehr, T.: The EarthCARE Satellite: the next step forward in global measurements of clouds, aerosols, precipitation and radiation, Bulletin of the American Meteorological Society, 96, 1311–1332, 2015.

Kokhanovsky, A. A., Budak, V. P., Cornet, C., Duan, M., Emde, C., Katsev, I. L., Klyukov, D. A., Korkin, S. V., C-Labonnote, L., Mayer, B., Min, Q., Nakajima, T., Ota, Y., Prikhach, A. S., Rozanov, V. V., Yokota, T., and Zege, E. P.: Benchmark results in vector atmospheric radiative transfer, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 1931–1946, https://doi.org/10.1016/j.jqsrt.2010.03.005, 2010a.

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 polarization characteristics of reflected light, Atmospheric Measurement Techniques, 3, 909–932, https://doi.org/10.5194/amt-3-909-2010, 2010b.

Korkin, S., Lyapustin, A., Sinyuk, A., Holben, B., and Kokhanovsky, A.: Vector radiative transfer code SORD: Performance analysis and quick start guide, Journal of Quantitative Spectroscopy and Radiative Transfer, 200, 295–310, https://doi.org/10.1016/j.jqsrt.2017.04.035, 2017.

Kotchenova, S. Y. and Vermote, E. F.: Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces, Applied Optics, 46, 4455–4464, https://doi.org/10.1364/AO.46.004455, 2007.

Kotchenova, S. Y., Vermote, E. F., Matarrese, R., and Frank J. Klemm, J.: Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance, Applied Optics, 45, 6762–6774, https://doi.org/10.1364/AO.45.006762, 2006.

Krasnopolsky, V. M. and Chevallier, F.: Some neural network applications in environmental sciences. Part II: Advancing computational efficiency of environmental numerical models, Neural Networks, 16, 335–348, https://doi.org/10.1016/S0893-6080(03)00026-1, 2003.

Lenoble, J., Herman, M., Deuzé, J., Lafrance, B., Santer, R., and Tanré, D.: A successive order of scattering code for solving the vector equation of transfer in the earth's atmosphere with aerosols, Journal of Quantitative Spectroscopy and Radiative Transfer, 107, 479–507, https://doi.org/10.1016/j.jqsrt.2007.03.010, 2007.

Martin, W., Cairns, B., and Bal, G.: Adjoint methods for adjusting three-dimensional atmosphere and surface properties to fit multi-angle/multi-pixel polarimetric measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 144, 68–85, https://doi.org/10.1016/j.jqsrt.2014.03.030, 2014.

Martin, W. G. and Hasekamp, O. P.: A demonstration of adjoint methods for multi-dimensional remote sensing of the atmosphere and surface, Journal of Quantitative Spectroscopy and Radiative Transfer, 204, 215–231, https://doi.org/10.1016/j.jqsrt.2017.09.031, 2018.

McClatchey, R., Fenn, R., Volz, F., and Garing, J.: Optical Properties of the Atmosphere (Revised), Tech. rep., Air Force Cambridge Research Laboratories Hascom AFB, MA, 1971.

Migliorini, S.: On the equivalence between radiance and retrieval assimilation, Monthly Weather Review, 140, 258–265, https://doi.org/10.1175/MWR-D-10-05047.1, 2012.

Mishchenko, M. I. and Travis, L. D.: Satellite retrieval of aerosol properties over the ocean using polarization as well as intensity of reflected sunlight, Journal of Geophysical Research: Atmospheres, 102, 16 989–17 013, https://doi.org/10.1029/96JD02425, 1997.

Nakajima, T. and Tanaka, M.: Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation, Journal of Quantitative Spectroscopy and Radiative Transfer, 40, 51 – 69, https://doi.org/https://doi.org/10.1016/0022-4073(88)90031-3, 1988.

Natraj, V., Li, K.-F., and Yung, Y. L.: Rayleigh scattering in planetary atmospheres: Corrected tables through accurate computation of X and Y functions, The Astrophysical Journal, 691, 1909–1920, 2009.

Pincus, R. and Evans, K. F.: Computational cost and accuracy in calculating three-dimensional radiative transfer: Results for new implementations of Monte Carlo and SHDOM, Journal of the Atmospheric Sciences, 66, 3131–3146, https://doi.org/10.1175/2009JAS3137.1, 2009.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS aerosol algorithm, products, and validation, Journal of the Atmospheric Sciences, 62, 947–973, https://doi.org/10.1175/JAS3385.1, 2005.

Scheck, L., Frèrebeau, P., Buras-Schnell, R., and Mayer, B.: A fast radiative transfer method for the simulation of visible satellite imagery, Journal of Quantitative Spectroscopy and Radiative Transfer, 175, 54–67, https://doi.org/10.1016/j.jqsrt.2016.02.008, 2016.

Siewert, C.: A concise and accurate solution to Chandrasekhar's basic problem in radiative transfer, Journal of Quantitative Spectroscopy and Radiative Transfer, 64, 109–130, https://doi.org/10.1016/S0022-4073(98)00144-7, 2000a.

Siewert, C.: A discrete-ordinates solution for radiative-transfer models that include polarization effects, Journal of Quantitative Spectroscopy and Radiative Transfer, 64, 227–254, https://doi.org/10.1016/S0022-4073(99)00006-0, 2000b.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, pp. 229–275, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-48546-9_7, 2008.

Spurr, R.: VLIDORT User's Guide, version 2.7, Tech. rep., RT Solutions, Inc., 2014.

Stamnes, K. and Swanson, R. A.: A new look at the Discrete Ordinate Method for radiative transfer calculations in anisotropically scattering atmospheres, Journal of the Atmospheric Sciences, 38, 387–399, https://doi.org/10.1175/1520-0469(1981)038<0387:ANLATD>2.0.CO;2, 1981.

Stamnes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K.: Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media, Appl. Opt., 27, 2502–2509, https://doi.org/10.1364/AO.27.002502, 1988.

Tanré, D., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., and Waquet, F.: Remote sensing of aerosols by using polarized, directional and spectral measurements within the A–Train: the PARASOL mission, Atmospheric Measurement Techniques, 4, 1383–1395, https://doi.org/10.5194/amt-4-1383-2011, 2011.

Toon, O. B. and Ackerman, T. P.: Algorithms for the calculation of scattering by stratified spheres, Applied Optics, 20, 3657–3660, https://doi.org/10.1364/AO.20.003657, 1981.

Toon, O. B., McKay, C. P., Ackerman, T. P., and Santhanam, K.: Rapid calculation of radiative heating rates and photodissociation rates in inhomogeneous multiple scattering atmospheres, Journal of Geophysical Research, 94, 16 287–16 301, 1989.

Vermote, E., Tanré, D., Deuzé, J.-L., Herman, M., and Morcrette, J.-J.: Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview, IEEE Transactions on Geoscience and Remote Sensing, 35, 675–686, 1997.

Wang, J., Xu, X., Spurr, R., Wang, Y., and Drury, E.: Improved algorithm for MODIS satellite retrievals of aerosol optical thickness over land in dusty atmosphere: Implications for air quality monitoring in China, Remote Sensing of Environment, 114, 2575–2583, https://doi.org/10.1016/j.rse.2010.05.034, 2010.

Wang, J., Xu, X., Henze, D. K., Zeng, J., Ji, Q., Tsay, S.-C., and Huang, J.: Top-down estimate of dust emissions through integration of MODIS and MISR aerosol retrievals with the GEOS-Chem adjoint model, Geophysical Research Letters, 39, L08 802, https://doi.org/10.1029/2012GL051136, 2012.

Weaver, C., da Silva, A., Chin, M., Ginoux, P., Dubovik, O., Flittner, D., Zia, A., Remer, L., Holben, B., and Gregg, W.: Direct insertion of MODIS radiances in a global aerosol transport model, Journal of the Atmospheric Sciences, 64, 808–827, https://doi.org/10.1175/JAS3838.1, 2007.

Weng, F.: Advances in Radiative Transfer Modeling in Support of Satellite Data Assimilation, Journal of the Atmospheric Sciences, 64, 3799–3807, https://doi.org/10.1175/2007JAS2112.1, 2007.

Wiscombe, W. J.: Improved Mie scattering algorithms, Applied Optics, 19, 1505–1509, https://doi.org/10.1364/AO.19.001505, 1980.

Woodward, S.: Modeling the atmospheric life cycle and radiative impact of mineral dust in the Hadley Centre climate model, Journal of Geophysical Research: Atmospheres, 106, 18 155–18 166, https://doi.org/10.1029/2000JD900795, 2001.

Xu, X., Wang, J., Henze, D. K., Qu, W., and Kopacz, M.: Constraints on aerosol sources using GEOS–Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data, Journal of Geophysical Research: Atmospheres, 118, 6396–6413, https://doi.org/10.1002/jgrd.50515, 2013.
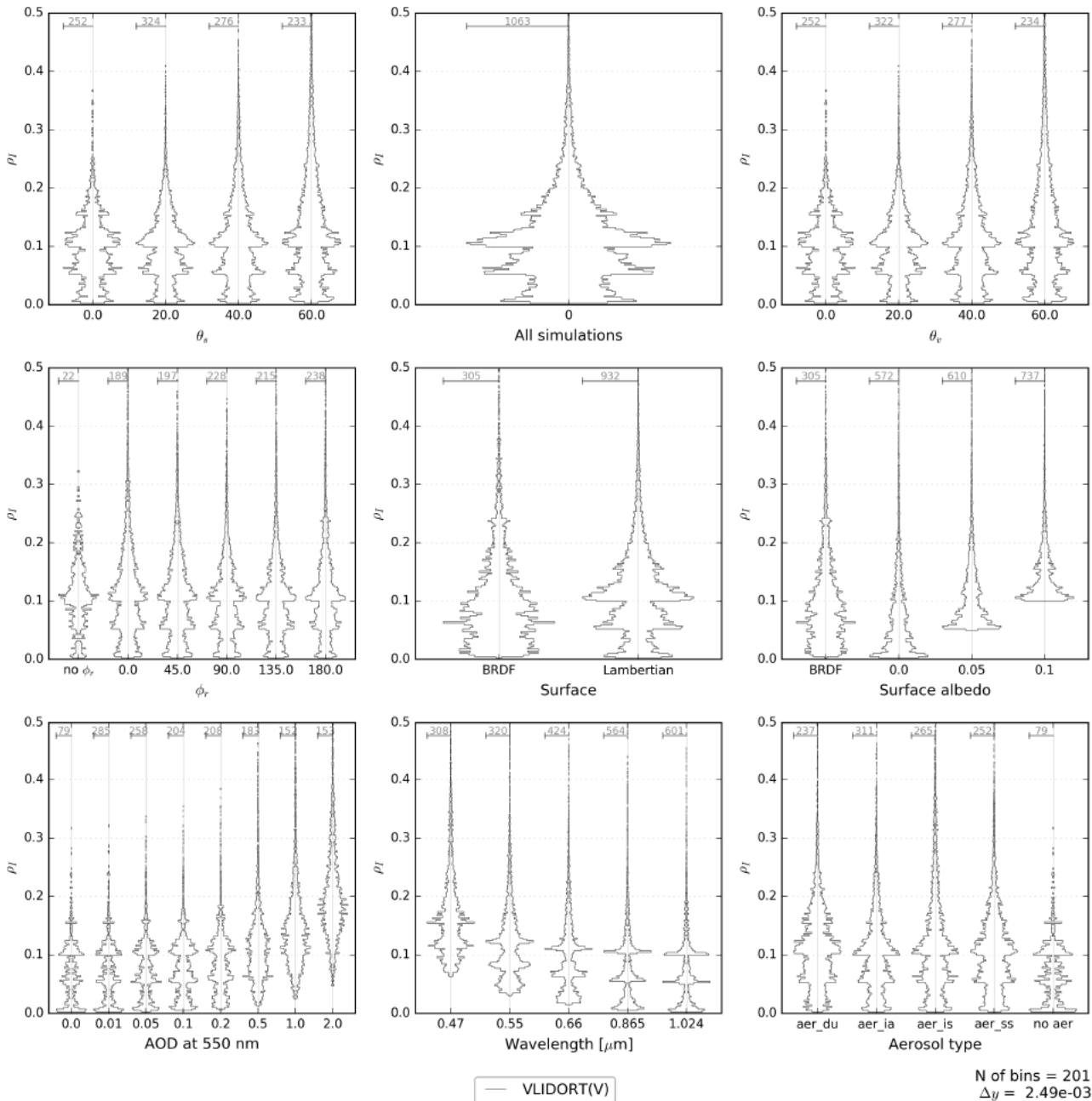
Figure 3. Histograms of the computed reflectances for the VLIDORT model. The top-middle panel is a histogram of all considered cases, while the other panels classify the cases according to the value of θ_s (top-left), θ_v (top-right), $\phi_r = \phi_v - \phi_s$ (middle-left, "no ϕ_r" corresponds to the cases where $\theta_s = \theta_v = 0.0$), surface type: BRDF or Lambertian (middle-middle), surface albedo value (middle-right), 550 nm aerosol optical depth (bottom-left), wavelength (in μm, bottom-middle) and aerosol type ("no aer" corresponds to AOD=0, bottom-right). The width of the classes is indicated in the bottom-right ($\Delta y = 2.49 \times 10^{-3}$), and the maximum number of cases in a class is provided for each histogram. The histograms are presented as "violin plots" with rotated density plot on each side. Note that the scale used on the x-axis is different for each violin plot.
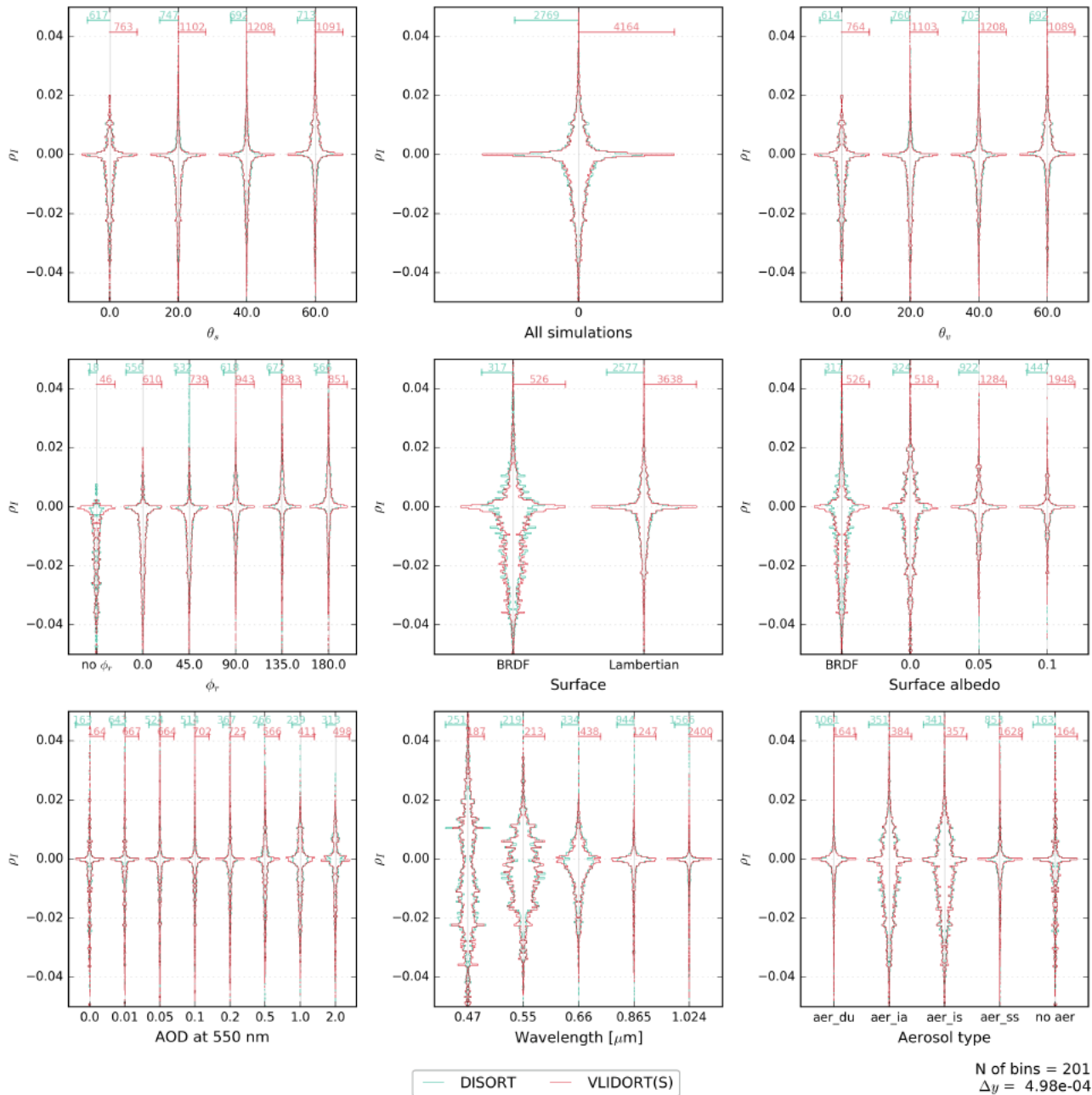
Figure 4. Same as Fig. 3 but for the relative error of the DISORT and VLIDORT (scalar) models against the vector version of VLIDORT. Please note the $\pm 5\%$ error range on the y-axis. The number of cases outside the range is indicated in Table 3. The width of the classes is $\Delta y = 4.98 \times 10^{-4}$. Note that the scale used on the x-axis is different for each violin plot but is the same for the two models within a violin plot.
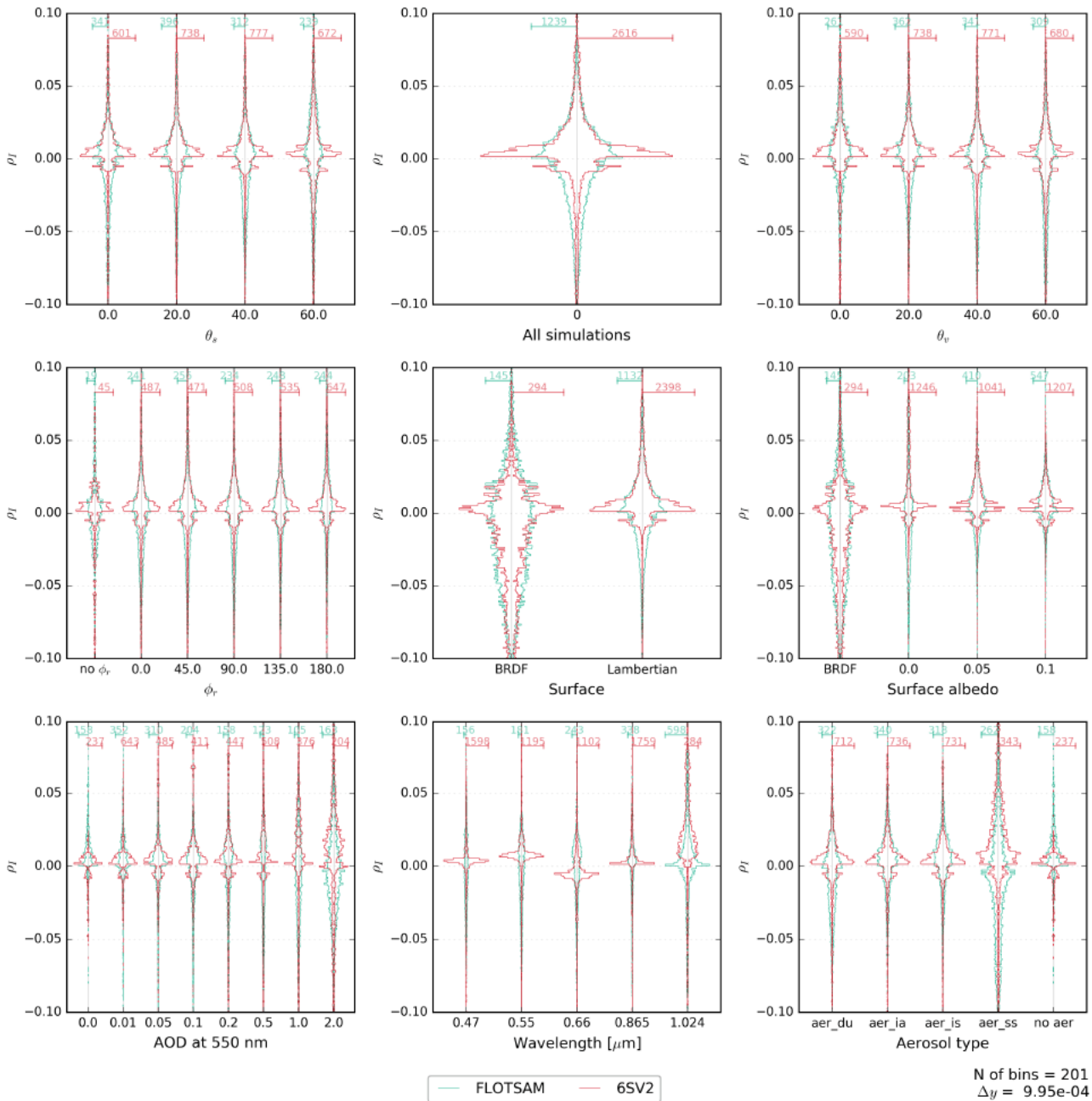
Figure 5. Same as Fig. 3 but for the relative error of the FLOTSAM and 6SV2 models against the vector version of VLIDORT. Please note the ±10% relative error range of the y-axis. The width of the classes is $\Delta y = 9.95 \times 10^{-4}$. The number of cases outside the range is indicated in Table 4. Note that the scale used on the x-axis is different for each violin plot but is the same for the two models within a violin plot.
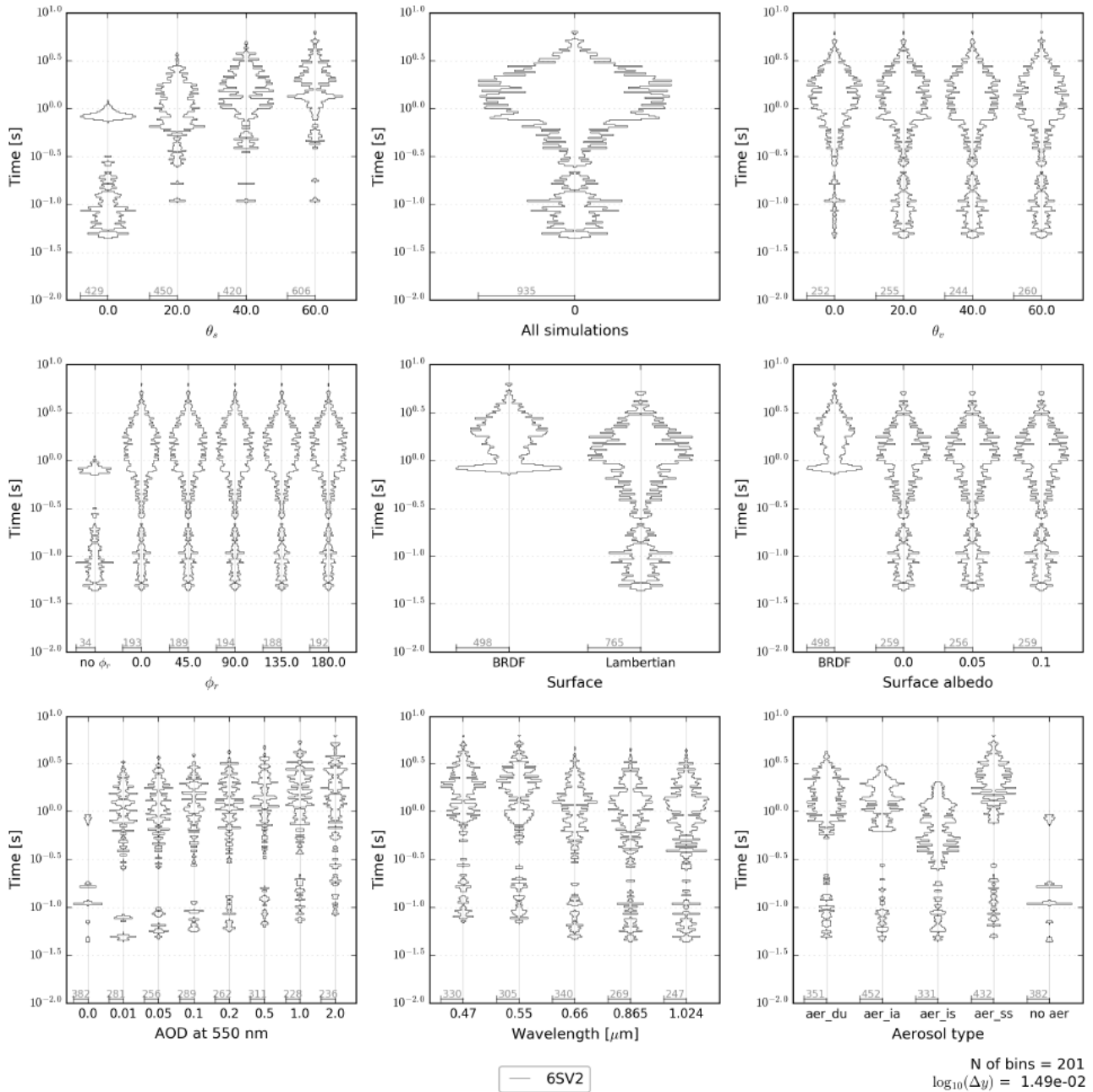
Figure 6. Same as Fig. 3 but for the runtime of the 6SV2 model. Please note the logarithmic scale on the y-axis.
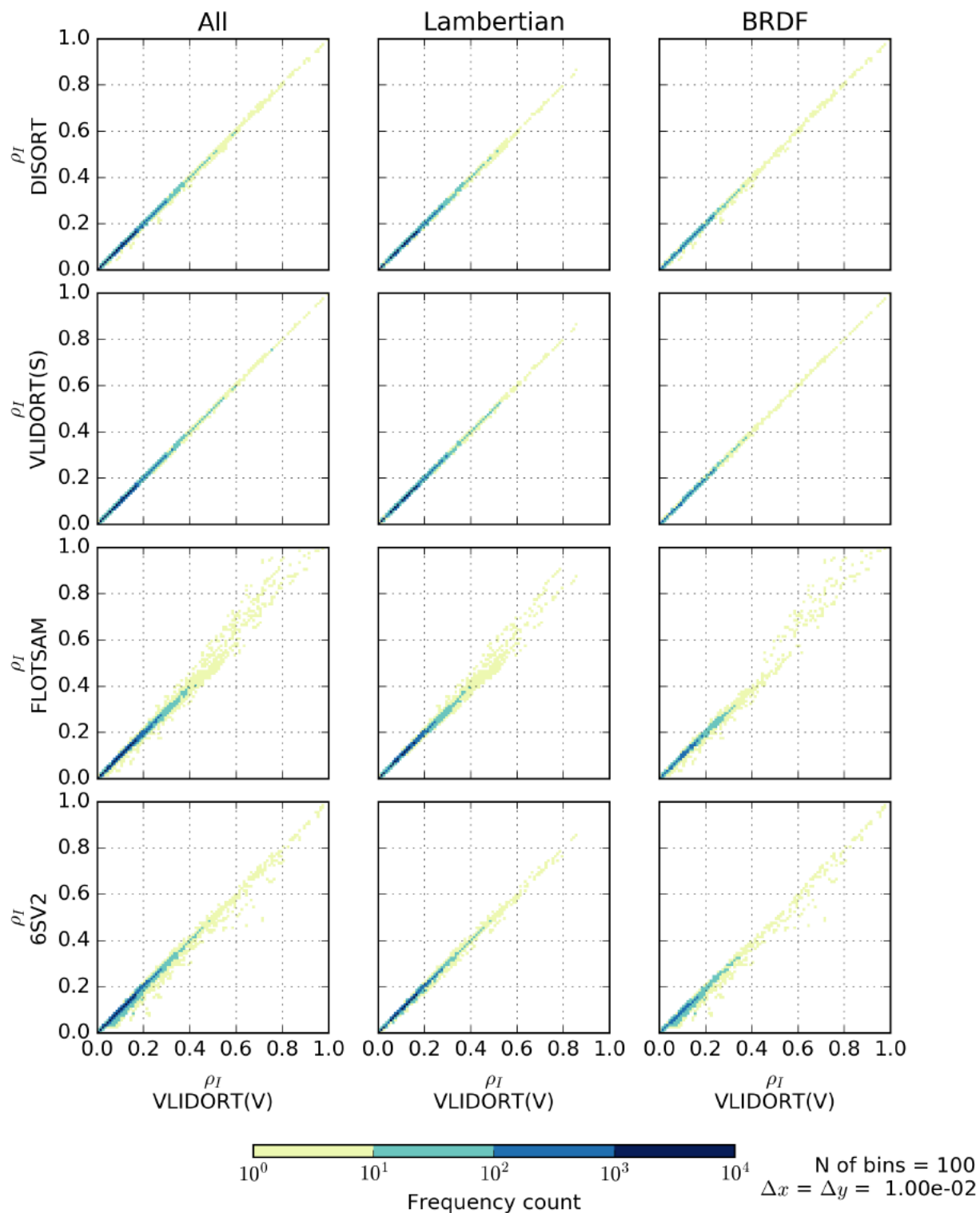
Figure A1. Two-dimensional density plots of the simulated reflectances for the 6SV2, FLOTSAM, DISORT and VLIDORT (scalar) models versus those of the VLIDORT (vector) model. The plots on the left column mix all 44,080 cases, while those in the middle and right columns are for the Lambertian and BRDF surface cases, respectively.