Efficient unbiased variance reduction techniques for Monte Carlo simulations of radiative transfer in cloudy atmospheres: The solution

Robert Buras [a],[*],[1], Bernhard Mayer [a],[b]

[a] Meteorologisches Institut München, Ludwig-Maximilian-Universität München, Theresienstrasse 37, 80333 München, Germany
[b] Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, 82234 Wessling, Germany

ARTICLE INFO

ABSTRACT

We present five new variance reduction techniques applicable to Monte Carlo simulations of radiative transfer in the atmosphere: detector directional importance sampling, n-tuple local estimate, prediction-based splitting and Russian roulette, and circumsolar virtual importance sampling. With this set of methods it is possible to simulate remote sensing instruments accurately and quickly. In contrast to all other known techniques used to accelerate Monte Carlo simulations in cloudy atmospheres – except for two methods limited to narrow angle lidars – the presented methods do not make any approximations, and hence do not bias the result. Nevertheless, these methods converge as quickly as any of the biasing acceleration techniques, and the probability distribution of the simulation results is almost perfectly normal. The presented variance reduction techniques have been implemented into the Monte Carlo code MYSTIC ("Monte Carlo code for the physically correct tracing of photons in cloudy atmospheres") in order to validate the techniques.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Clouds and their interaction with radiation are in the center of attention in atmospheric research. Current efforts concentrate on improving the understanding of cloud physics, on providing more detailed and more accurate remote sensing observations, as well as on developing more accurate parameterizations for climate and numerical weather prediction models. With the continuously improving spatial resolution of numerical models as well as of remote sensing observations, horizontal cloud inhomogeneity and its impact on irradiance and radiance must no longer be neglected and three-dimensional radiative transfer models therefore become more and more important. Examples for their practical application are the inclusion of three-dimensional

radiative transfer in cloud resolving models [1] or the first developments of three-dimensional cloud retrievals [2]. With the increased sophistication of satellite instruments it has become important to simulate radiative transfer with high accuracy, and the simulation of radiances in presence of inhomogeneous water and ice clouds is one of the most difficult radiative transfer problems, due to the strong forward peak of the scattering phase function. While this obstacle can be removed at least for the calculation of irradiances using a delta-scaling or peak-truncation technique [3,4] it remains a challenge for radiance calculations even in the one-dimensional approximation: Radiance calculations in general require accurate treatment of the phase function as detailed below; in particular novel remote sensing techniques using e.g. the rainbow area [5] or the backscatter glory [6], or even worse, calculations of circum-solar irradiance required e.g. for concentrating solar power plants require highly accurate consideration of the phase function in one- and three-dimensional radiative transfer codes.

Monte Carlo methods have long be a means of simulating three-dimensional radiative transfer of light in cloudy

* Corresponding author. Tel.: +49 8921804440.
  E-mail addresses: robert.buras@lmu.de (R. Buras),
bernhard.mayer@lmu.de (B. Mayer).
  [1] Before at Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, 82234 Wessling, Germany.

atmospheres [7–19]. When using the local estimate technique (described e.g. by Marchuk et al. [9]) the strong forward peaks of the scattering phase function lead to rare events that contribute significantly to the total result, also known as "spikes". The accuracy of the result is then no longer determined by the number of simulated photons, but by the number of spikes having occurred, which can make the calculation unfeasible. Several solutions have been found to resolve this problem. One widely used group of solutions, the above mentioned truncation or delta-scaling methods, include the delta-M method [4,20], the delta-fit method [21,22], and delta-scaling with geometric truncation of the scattering phase function [18,23,24]. However, all of these methods introduce a bias to the result because they alter the scattering phase function. Although these methods have been driven to quite a sophisticated level [25], the magnitude of the bias cannot be estimated by the method itself and therefore it remains unknown whether the bias is small enough. The only practicable solutions without any alteration of the phase function were the variance reduction techniques proposed by Platt [10] and Noormohammadian [14]. Unfortunately, these methods only work for the special case of lidars, and are only applicable at small optical depths.

We here present a set of variance reduction methods (VRM) that allow to perform Monte Carlo simulations without having to alter the phase functions, no matter how strong the forward peak is. These methods get rid of the spikes normally encountered, and at the same time have computational convergence times of the order of those found for simulations with altered (smooth) phase functions, or even smaller. The methods are applicable for all Monte Carlo simulations aiming at calculating radiances by local estimators, including lidars. Some of these methods, in particular the detector directional importance sampling (DDIS), the n-tuple local estimate (NLE), prediction-based splitting (PBS), prediction-based Russian roulette (PBRR), and circum-solar virtual importance sampling (CS-VIS), are completely new.

The set of methods have been implemented into the Monte Carlo code MYSTIC [19] – which is part of the radiative transfer package libRadtran [26] – and run under the name Variance Reduction "Optimal Options" Method (VROOM). VROOM has already been used in various applications, see [27,28].

In Section 2 we describe the different VRM and justify why they are needed and how they solve the problems occurring. Section 3 validates VROOM and compares the computational speed between VROOM and other (biasing) acceleration techniques.

2. Variance reduction methods

For an introduction to Monte Carlo methods in the atmosphere and in particular to the well-known local estimate method we refer the reader to the pioneering work of Marchuk et al. [9], the mathematically comprehensive Evans and Marshak [17, Chapter 4.2], as well as the illustrative Mayer [19]. Absorption, surface reflection, and polarization are implemented as explained in [19,29] and
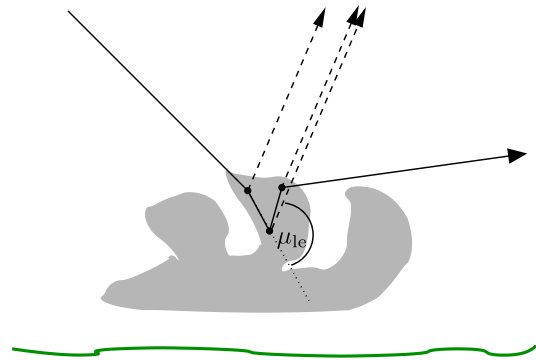
Fig. 1. Illustration of the local estimate method. The photon (solid line, scatters three times before leaving the cloud. Each time, a local estimate is performed, i.e. the probability that the photon turns towards the detector, $p(\mu_{\text{le}})$, and travels to it without being extinct, $\exp(-\tau_{\text{ext}})$, is calculated. To this end, a virtual photon is created (dashed lines) each time which is propagated to the detector in order to calculate the extinction τ_{ext}.

will be ignored in the following discussion, but generalization of VROOM is intuitive and straightforward.

2.1. A short introduction to Monte Carlo

The aim is to solve the radiative transfer equation (RTE), which is the valid approximation of the Maxwell equations in the atmosphere [30,31].

In short, in a standard Monte Carlo simulation irradiances are sampled by tracing and counting "photons",[2] thereby randomly choosing each step length and scatter direction according to pre-calculated probability density functions (PDF).

In particular, the step length until the next scattering is determined by Lambert–Beer's law:

$$p_{\text{step}}(\tau) = \exp(-\tau) \tag{1}$$

where $p_{\text{step}}(\tau)$ is the probability that the photon scatters at the scattering optical depth τ. The new direction is then sampled according to the scattering phase function

$$p(\mu) = \sum_j \beta_j p_j(\mu) / \sum_j \beta_j. \tag{2}$$

where β_j and p_j are the scattering coefficient and phase function of particle type j, respectively, and the sum is over all existing particle types. $\mu = \cos\theta$ is the cosine of the angle θ between the photon directions before and after scattering. For this discussion we will assume the phase function to be independent of the scattering azimuth ϕ. This is valid for spherical as well as for randomly oriented particles. Hence ϕ is chosen randomly from $[0,2\pi[$.

The most convenient way to calculate *radiances* for certain directions is by applying the local estimate method (LE), see also Fig. 1: At each scatter event the local estimate is calculated, that is, the probability that the photon scatters into the direction of the detector times the probability that the photon travels to the detector without being absorbed or scattered again. For a scatter event, the

---

[2] We use the term "photon" to represent an imaginary discrete amount of electromagnetic energy transported in a certain direction. It is not related to the QED photon [32].

LE is given by the weight

$$w_{\mathrm{le},i} = W_{i-1}\,p(\mu_{\mathrm{le}})\exp(-\tau_{\mathrm{ext}}) \qquad (3)$$

where W_{i-1} is the photon weight before the ith scatter, μ_{le} is the cosine of the angle θ_p between the photon direction before scattering and the direction towards the detector, and τ_{ext} is the extinction (i.e. the sum of scattering and absorption) between the scatter location and the detector. As can be proven by expanding the RTE with help of the Von-Neumann series (see e.g. [17]), adding up all local estimates gives the radiance. The total contribution of the photon to the radiance is the weight

$$w_{\mathrm{le}} = \sum_{i=1}^{n} w_{\mathrm{le},i} \qquad (4)$$

where n is the number of scatterings the photon performs.

Although the LE allows us to calculate radiances with help of Monte Carlo, it features a big disadvantage in the presence of particles with very strong forward peaks in the scattering phase functions, see e.g. Fig. 2, such as those of cloud droplets, ice crystals, and large aerosols: Whenever a photon coincidentally scatters approximately into the direction of the detector, the following scatter will produce a very large LE since $p(\mu_{\mathrm{le}} \approx 1)$ is very large (see Fig. 3). These events, which happen rarely, are called "spikes" and can slow down the statistical convergence of the Monte Carlo code by several orders of magnitude.

Fig. 4 shows a typical example of how the average of the radiance being calculated evolves with the number of photons in a simulation, once with local estimate, but without VRMs, and once with the VRMs presented in this paper. Clearly, only the latter case has a nice asymptotic behavior. The simulation without VRM seems to have reached 1% standard deviation after 5000 photons, but then a spike occurs which strongly increases the standard deviation. Even after 10^6 photons the standard deviation is
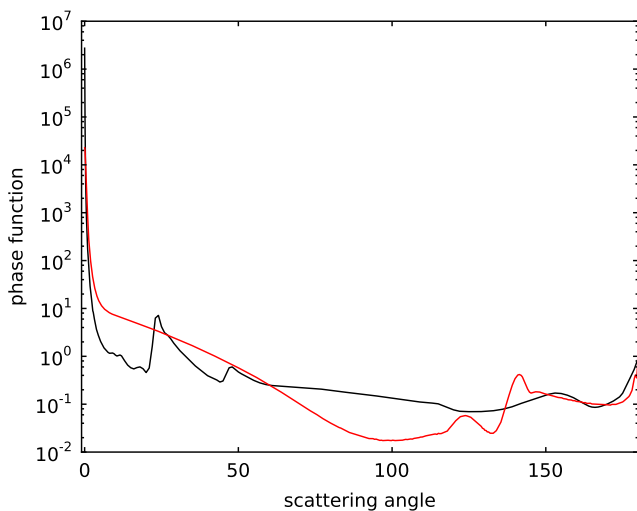
Fig. 2. Examples of scattering phase functions. The red line depicts the phase function for water droplets with effective radius $r_{\mathrm{eff}} = 16$ μm, the black line shows the phase function for ice crystals with effective radius $r_{\mathrm{eff}} = 55$ μm [33]. The two phase functions correspond to the phase functions with the strongest forward peaks occurring in the simulations shown in Section 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
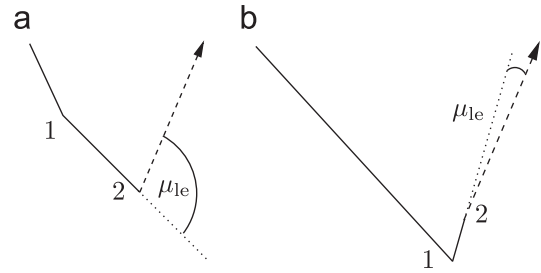
Fig. 3. (a) Example of a small local estimate in case of double scattering: At the first scattering, the photon (solid line) scatters in a direction not towards the detector; at the second scattering the probability of the photon scattering into the direction of the detector (dashed line) is small according to the phase function, hence the local estimate is small. (b) Example of a very large local estimate ("spike") in case of double scattering: At the first scattering, the photon scatters coincidentally towards the detector; at the second scattering the probability of the photon scattering into the direction of the detector is very large according to the phase function, hence the local estimate is very large.
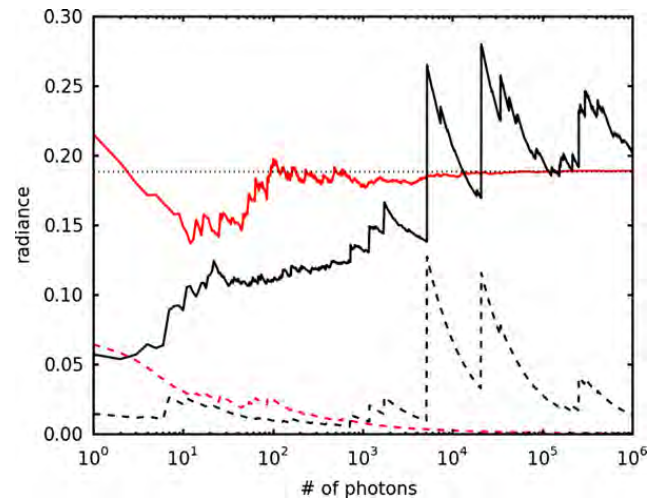
Fig. 4. Example of convergence behavior of two simulations, one without VRMs (black), and one with the VRMs of this paper (red). Shown is how the result (solid lines) improves with increasing photon number. The dashed lines depict the corresponding standard deviations. The results were taken from the simulations in geometry "S" presented in Section 3 (At $y = 18.75$ km). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

still more than 1%. This shows that the standard deviation calculated for this model cannot be trusted.

The rest of this section will describe the VRMs needed in order to obtain a nicely converging result. The goal hereby is to get rid of spikes. However, contrary to intuition, this is achieved not by reducing the probability of a spike, but by increasing this probability artificially. The result is an increased statistic of spikes, each of which contributes less to the result than it would in the absence of VRMs. If done correctly, the physics of radiative transfer is not altered, but the photon paths formerly leading to spikes can no longer be distinguished from other photon paths in their contribution to the result.

2.2. Detector directional importance sampling (DDIS)

To get rid of the spikes it is necessary to increase the probability of the spiky events. To this end, at each scatter,
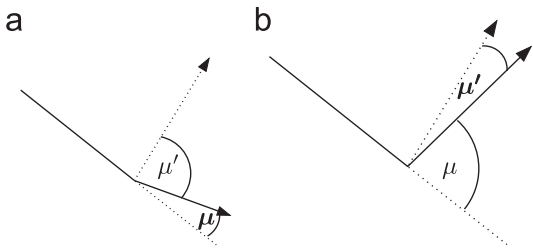
Fig. 5. Illustration of DDIS (the azimuthal angles are not shown): (a) With a probability of $(1-\varepsilon_{\mathrm{ddis}})$ we scatter the photon as usual using $p(\mu)$. The angle between the new photon direction and the direction toward the detector, μ', can be obtained geometrically and depends also on the scattering azimuth ϕ. The photon weight factor w_i is approximately $(1-\varepsilon_{\mathrm{ddis}})^{-1} \approx \mathcal{O}(1)$ because forward scattering occurred and thus $p(\mu) \gg p_{\mathrm{ddis}}(\mu')$. (b) With a probability of $\varepsilon_{\mathrm{ddis}}$ the photon is first turned towards the detector (dotted line with arrow) and then the scattering is performed using $p_{\mathrm{ddis}}(\mu')$. The resulting scattering angle μ can be obtained geometrically and depends also on ϕ'. The photon weight factor w_i is approximately $p(\mu)/(\varepsilon_{\mathrm{ddis}}p_{\mathrm{ddis}}(\mu')) \ll 1$ because the photon scattered into an unlikely direction.

the photon is scattered according to an altered phase function:

$$p^{\dagger}(\mu,\phi) = (1-\varepsilon_{\mathrm{ddis}})p(\mu) + \varepsilon_{\mathrm{ddis}}p_{\mathrm{ddis}}(\mu'(\mu,\phi)) \qquad (5)$$

where $\varepsilon_{\mathrm{ddis}} \in [0,1[$ is a free parameter to be optimized (normally of order 0.1), μ' is the cosine of the angle θ' between the direction to the detector and the photon direction after scattering, and $p_{\mathrm{ddis}}(\mu')$ is a phase function which needs to be optimized for the problem.

We can interpret the usage of Eq. (5) with help of Fig. 5: With a probability of $(1-\varepsilon_{\mathrm{ddis}})$ the new photon direction is determined as usual using the PDF $p(\mu)$, resulting in a random value for μ. However, with a probability of $\varepsilon_{\mathrm{ddis}}$ the photon is first turned exactly toward the detector ($\mu'=1$) and then "scattered" according to the PDF p_{ddis}, resulting in a random value for μ'. Then the actual scattering angle μ is a function of μ' and the corresponding azimuth ϕ' (latter is chosen randomly). The consequence is that in the presence of strongly forward-peaked phase functions we strongly enhance the number of photons scattered (approximately) into the direction of the detector, therefore the name "detector directional importance sampling" (DDIS). We will call the process of scattering the photon according to the second term in Eq. (5), see also Fig. 5b, "DDISsing", i.e. the photon is being "DDISsed". In contrast to this, a photon not being "DDISsed" is said to scatter "naturally".

Of course, this modification of physics needs to be corrected with a weight factor for the photon [34],[3]

$$w_i = p(\mu)/p^{\dagger}(\mu,\phi) \qquad (6)$$

This is the probability that the photon should move in the direction (μ,ϕ), given by $p(\mu)$, divided by the probability with which we send the photon in that same direction, given by $p^{\dagger}(\mu,\phi)$. The weight factor is multiplied to the

photon weight (initially 1) at each scatter:

$$W_i = \prod_{j=1}^{i} w_j \qquad (7)$$

Let us show that DDIS prevents spikes on the example of double scattering: Consider a photon initially moving in a direction not towards the detector, see Fig. 3b. If this photon happens to move towards the detector after scattering, the photon weight will be reduced due to the strongly enhanced probability that this direction is chosen, with respect to the probability that the direction should be chosen (see Eq. (6)), independent of whether the photon chose the direction "naturally" or by being DDISsed. The LE of the photon at the second scatter is (omitting the azimuthal dependencies)

$$\begin{aligned} w_{\mathrm{le},2} &= W_1 p(\mu_{\mathrm{le},2})\exp(-\tau) \\ &= \frac{p(\mu_1)p(\mu_{\mathrm{le},2})}{(1-\varepsilon_{\mathrm{ddis}})p(\mu_1)+\varepsilon_{\mathrm{ddis}}p_{\mathrm{ddis}}(\mu_1')}\exp(-\tau) \end{aligned} \qquad (8)$$

Note that for geometric reasons $\mu_{\mathrm{le},2} \le \mu_1'$ (the equality holds for detectors at infinite distance). Further assuming that $p(\mu_{\mathrm{le},2}) \le p(\mu_1')$ (which normally is true at least for the forward peak of the phase function), and for illustration using $\varepsilon_{\mathrm{ddis}} = 0.5$ and $p_{\mathrm{ddis}} = p$, and ignoring extinction, Eq. (8) can be approximated by

$$w_{\mathrm{le},2} \le 2\frac{p(\mu_1)p(\mu_1')}{p(\mu_1)+p(\mu_1')} \le 2\min(p(\mu_1),p(\mu_1')) \qquad (9)$$

where the second inequality can easily be proven. Hence the local estimate will never be extremely large, and spikes can no longer occur. The only case $w_{\mathrm{le},2}$ can get large is when both $p(\mu_1)$ and $p(\mu_1')$ are very large. This can only happen if both μ_1 and μ_1' are almost 1, which means that the photon is already moving toward the detector from the start. In this case, very large local estimates will be frequent, and hence cannot be considered as spikes.

A few words on the choice of p_{ddis}: In a realistic scene, the phase function depends on the location and we can have both ice and water clouds with different effective radii, and aerosols of different types and effective radii as well as regions with Rayleigh scattering only. Since the location, and hence the phase function of the next scatter is not known, p_{ddis} should be chosen expecting the worst cases. An intuitive choice is to define a phase function which, for each scattering angle μ, takes the maximum value of all phase functions present in the simulated scene for the given μ:

$$p_{\mathrm{ddis}}(\mu) = C_{\mathrm{ddis}}\max_i[p_i(\mu)] \quad \text{for all } \mu \in [-1,1] \qquad (10)$$

where i runs over all phase functions (e.g. for water and ice clouds of all present effective radii, and aerosol types). The normalization C_{ddis} is chosen such that $\int_{-1}^{1} p_{\mathrm{ddis}}(\mu) = 2$.

With DDIS spikes can no longer occur in optically thin atmospheres. The optimal choice for $\varepsilon_{\mathrm{ddis}}$ for e.g. double scattering is obviously 0.5 because then the natural and the spiky photon paths occur with the same frequency. However, for optically thicker atmospheres this choice gets problematic: While DDIS decreases the variance of the LEs of the next few scatters, it reduces the number of photons scattering according to the physical phase function p. For

---

[3] That this works can also be seen by looking at the so-called maximum cross-section method [35, e.g.]. For a rigorous proof see [9, p. 21–23].

instance, solar photons reflected into space by a homogeneous cloud of optical thickness τ will scatter on average $c\tau$ times, where c is a number of order of one [35,36], but after $c\tau$ scatterings only $(1-\varepsilon_{\mathrm{ddis}})^{c\tau}$ photons still follow a path unaltered by DDIS, see Fig. 6. Most photons will have left the computational domain long before, and thus large scatter orders occur very infrequently (and with very large photon weight) for large scatter orders, which therefore is a problem in case of large optical depth. A small value of $\varepsilon_{\mathrm{ddis}}$ is then desirable, which is in contradiction to the idea of DDIS. This will be addressed in the next section.
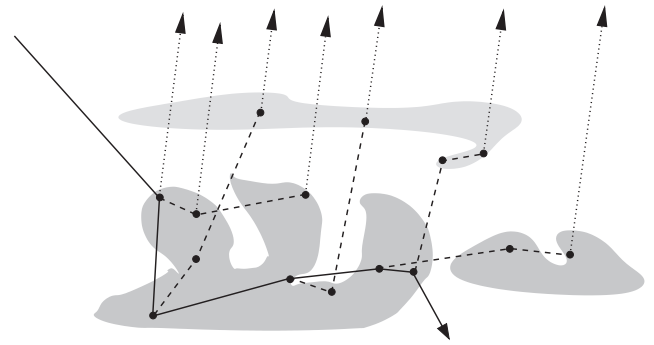
2.3. n-tuple local estimate (NLE)

The demand for the photons to move "naturally" and at the same time low variance with help of DDIS is solved with n-tuple local estimate (NLE), illustrated in Fig. 7: We simulate photons without DDIS, let us call them "mother" photons (MP). Hence the "natural" path of the photon is not altered. At each scatter of a MP, we perform an NLE: We clone the photon, i.e. we make a copy of it, called "cloned" photon (CP), which is followed for a fixed number of scatters $n_{\mathrm{scCP}}-1$, after which a LE is performed and the CP is purged. The MP is then evolved further to the next scatter.

With this method, a CP created at the n_{sc}th scattering of the MP accounts for the LE of the $n_{\mathrm{sc}}+n_{\mathrm{scCP}}$th scatter order of the MP. Of course, we have to make sure that each scatter order of the Von-Neumann series (Eq. (4)) is accounted for exactly once for each MP. To this end, single scattering is accounted for by a LE from the first scatter of the MP, and the first CP performs LEs for all scatterings up to n_{scCP} in order to account for the scattering orders 2 to n_{scCP}, see again the illustration in Fig. 7. This method has similarities with the double local estimate [9], therefore its name.

Up to here NLE is simply a computationally expensive alternative to LE, and spikes will occur in the presence of forward peaked phase functions. However, if we apply
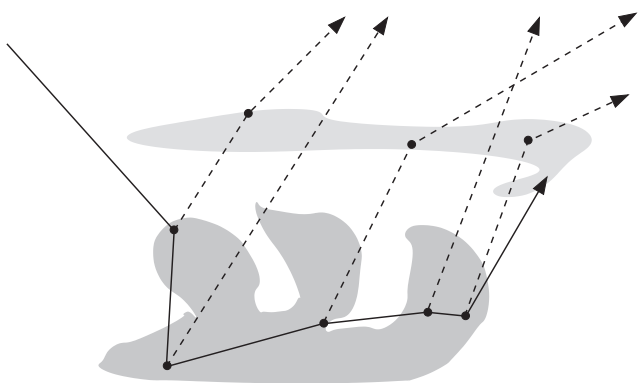
Fig. 6. Explaining the problem with DDIS in thick clouds: Without DDIS, the photon would e.g. take the "natural" path depicted by the solid line. However, if DDIS was used, then at each scatter, with a probability of $\varepsilon_{\mathrm{ddis}}$, the photon would be scattered according to $p_{\mathrm{ddis}}(\mu')$ (dashed lines). Hence the path shown by the solid line would only be followed to its end with a probability of $(1-\varepsilon_{\mathrm{ddis}})^5$ (five being the number of scatters in this example). This means that, comparing a simulation with e.g. $\varepsilon_{\mathrm{ddis}}=0.5$ with a simulation without DDIS, the former would need $(1-\varepsilon_{\mathrm{ddis}})^{-5}=32$ times more photons to get the same number of "naturally moving" photons after five scatterings.

Fig. 7. Example of 3-tuple local estimate: The "mother photon" (MP, solid line) follows its "natural" path, i.e. no DDIS is applied to it. At each scatter of the MP, a "cloned photon" (CP, dashed lines) is created. Each CP scatters twice in our example (the first scatter happens at the moment of creation of the CP), DDIS is applied to the CP, which is DDISsed with a probability of $\varepsilon_{\mathrm{ddis}}$. At the third scatter, a LE is performed (dotted lines). Note that the local estimate is also performed at the MP's first scatter, and at the first CP's second scatter. This is in order to simulate the first and second scatter orders of the Von-Neumann series, which cannot be simulated with 3-tuple LE.

DDIS with a large $\varepsilon_{\mathrm{ddis}}$ (of order 0.1) to the CP, the number of photons moving toward the detector will be enhanced greatly, and no spikes will occur.

Said in other words: On the one hand, because the MPs are allowed to scatter "the natural way", i.e. without applying DDIS, the number of photons scattering many times are not reduced in the presence of large optical depths. On the other hand, because DDIS is applied to the CP several times before the LE is performed, the probability of photons moving toward the detector is large.

We introduce two new parameters here, n_{firstCP}, which is the scatter order at which the first CP is created, and n_{LEperCP}, which is the number of local estimates performed by the CP (up to now these parameters were both 1). The Von-Neumann series (Eq. (4)) then implies that a CP is created at the scatter orders $n_{\mathrm{firstCP}}+in_{\mathrm{LEperCP}}$, $i=0,1,2,\dots$, and each CP performs LEs for the scatter orders (counted from the creation for the CP) n_{scCP} to $n_{\mathrm{scCP}}+n_{\mathrm{LEperCP}}-1$. Of course, the MP must perform LEs until the first CP is created, and the first CP performs an LE for all scatter orders. Optimizing these parameters can help reducing computational time. For better understanding, please look at Fig. 8.

2.4. Prediction-based splitting (PBS) and Russian roulette (PBRR)

Still, MPs can by chance be moving directly towards the detector when being cloned because DDIS is no longer applied to the MPs after $n_{\mathrm{firstCP}}-1$ scatter orders. The created CP is likely to scatter away from the detector direction during the n_{scCP} scatters it encounters before the LE is performed, but with a small probability it will still point towards the detector when the LE is performed without the photon weight having been reduced (note that the weight cannot be reduced with DDIS when the photon points towards the detector, because then $\mu=\mu'$, and thus $w_i=1$). This event, although rare, will produce a spike, see Fig. 9.
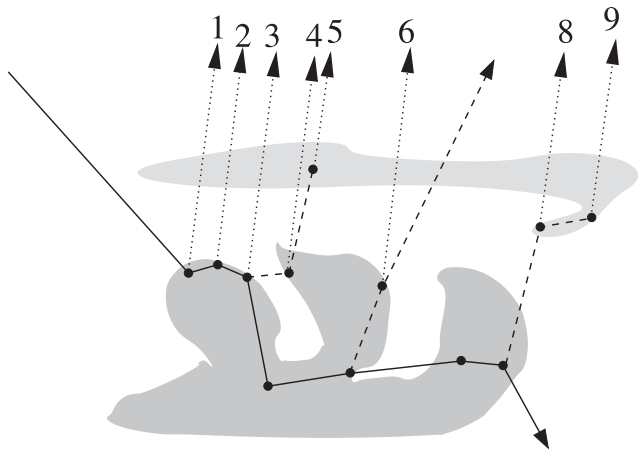
Fig. 8. Example of n-tuple local estimate using the parameter set $n_{scCP}=2$, $n_{firstCP}=3$, and $n_{LEperCP}=2$. The numbers on top denote which scatter order the LEs (dotted line) represent. The LEs of the first three scatter orders (corresponding to $n_{firstCP}=3$) are performed by the MP (solid line), the first CP (dashed lines) performs the LEs for the 4th scatter order (which is the 2nd scatter order of the CP, corresponding to $n_{scCP}=2$), and for the 5th scatter order (the CP performs two LEs, corresponding to $n_{LEperCP}=2$). The second CP is started before the fifth scattering of the MP in order to perform a double LE at the 6th scatter order. Because the CP leaves the scene before the 7th scatter order, no LE is performed. This does not introduce any bias to the simulation. Finally, the third CP is started before the seventh scatter order, and performs the 8th and 9th LE scatter orders. Note that DDIS is applied to all scatterings of CPs (i.e. Eq. (5) is used), and to the MP for the first two scatterings.
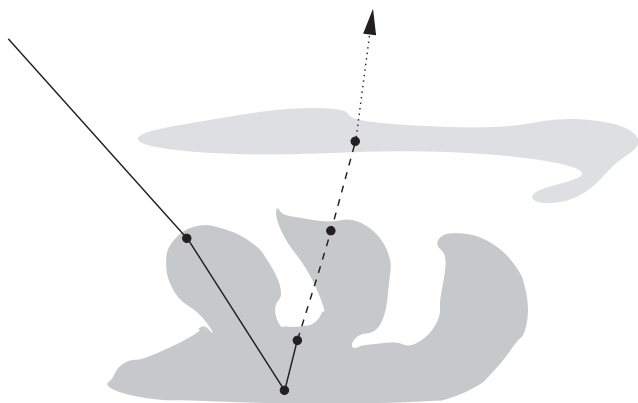

Fig. 9. Illustration of possible spike event with DDIS: The CP (dashed line) starts off in a direction approximately towards the detector. Even though DDIS is applied, the photon weight is not reduced significantly because the probability of the photon moving towards the detector is large anyway. The LE (dotted line) produces a spike. Note that this spike event is similar to the spike events discussed in Fig. 5b, but due to DDIS these spikes occur much more infrequently.

This can be prevented by splitting CPs with expected large LE, see also Fig. 10: We define a splitting weight

$$w_{spl} = W_i p_{ddis}(\mu'_i) \tag{11}$$

which is calculated after each scatter of a CP and is an estimate of the LE (without extinction term) that the CP would produce at the next scatter (assuming $\mu_{le,i+1} = \mu'_i$, see also discussion of Eq. (8)). If w_{spl} exceeds a critical value $w_{crit,\,spl}$, the photon is split into n_{spl} identical photons with reduced weight $W_i = W_i/n_{spl}$, where $n_{spl} \simeq w_{spl}$ is an intuitive choice. Each of these split CPs (SCP) are
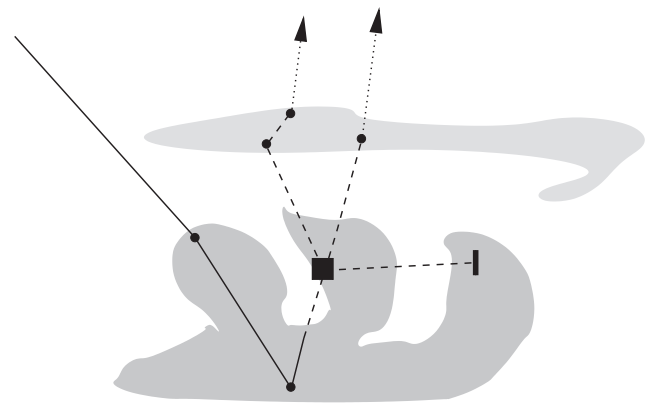
Fig. 10. Splitting and Russian roulette: If a photon becomes critical, i.e. has the potential of producing a spike, it is divided into n_{spl} identical photons (black box), in this example three. These photons obtain a weight reduced by a factor of three, hence the LE of the spike occurring for the middle photon is reduced. The rightmost split photon moves at a large angle to the detector and is unlikely to contribute much to the signal, it is destroyed by Russian roulette (bar). The left photon contributes little to the signal. Note that the advantage of splitting is to increase the frequency of getting events like the middle photon.

propagated independently of each other. An SCP should of course be allowed to be split itself at a later scatter event if necessary. It can be proven that this method limits the values of LEs to $w_{crit,spl}\varepsilon_{ddis}^{-1}C_{ddis}^{-1}$, where C_{ddis} was introduced in Eq. (10).

Note that, although PBS also reduces the variance, DDIS does so much more efficiently, but not for all cases. Splitting a photon can be interpreted as starting a small MC simulation for this special photon. Hence PBS is a very slow method, which should only be used when DDIS fails to prevent spikes.

Splitting has the disadvantage that it can significantly slow down the computational time per started photon. We will demonstrate this with two examples and present the solutions. First, if the detector is looking approximately into the direction of the photon emitter, w_{spl} gets very large values already at the first scatter, resulting in up to millions of SCPs in case of phase functions with strong forward peaks, e.g. of ice clouds. These clones are actually not needed since the probability of photons moving toward the detector is already large (we expect a strong signal!). By dividing w_{spl} by

$$w_{redspl} \equiv \max\left(1, \min_{i'=1,i-1}(p_{ddis}(\mu'_{i'}))\right) \tag{12}$$

("redspl" signifying "reduced splitting") and thus making w_{spl} smaller, this can be prevented. It is also useful to define a maximum number of SCPs per scatter, $n_{spl,\,max}$. Note that this method was found by trial and error, and is more to be understood intuitively.

Second, many CPs and SCPs have very small w_{spl}, thus it is unlikely that they will contribute significantly to the signal. These photons cost a lot of computational time without improving the result. The well-known method of Russian roulette [9] can be applied here: If w_{spl} drops below a value $w_{crit,\,rr}$ the photon is killed with a probability of $(1-w_{spl})$. The photons surviving (with probability w_{spl}) obtain an increased weight $W_i = W_i/w_{spl}$ so that the method

remains unbiased. In other words, of many photons which will only contribute little more to the result (i.e. they have a small w_{spl}), some are killed randomly, and the surviving continue with increased weight in order to account for the killed photons.

Again, it is useful to define a moderating $w_{rr, min}$, i.e. the probability of killing photons for which $w_{spl} < w_{crit,rr}$ should not be larger than $(1-w_{rr, min})$. Also, Russian roulette should be limited such that the photon weight W_i cannot get larger than one.

2.5. Circum-solar virtual importance sampling (CS-VIS)

Finally, following problem needs to be solved: If the detector is looking approximately into the direction of the photon emitter, and the optical thickness of atmospheric components with strongly peaked phase functions between emitter and detector is much smaller than one, then most photons will not contribute significantly to the signal. The few photons scattering on these critical components (e.g. ice crystals), will produce large contributions. This can be solved by artificially enhancing the scattering coefficient β. This method is based on the importance sampling described by Marchuk et al. [9] and is similar to the "maximal cross-section" method described there (see also [35]).

In particular, we enhance the scattering coefficient β in case the photon is in a grid cell containing large particles (clouds or large aerosols), the optical thickness the photon passes in this cell is smaller than 1, and $w_{redspl} > 10$. The enhanced scattering coefficient is $\beta_{IS} = 10^{-3} w_{redspl}/dl$, where dl is the step length in the grid cell, and where we limit β_{IS} to be smaller than dl^{-1} and larger than β. After moving the photon as usual with a random optical depth τ_{IS}, the local estimate is performed with a weight reduced by a factor of β/β_{IS}. Then, with a probability of β/β_{IS}, the photon scatters, i.e. changes direction. If the photon does not scatter, it moves on without changing direction. Note that the specific parametrization of this method was set by intuition.

This method does not alter the photon path, it only enhances the rate of calculating the local estimate where needed. This is why we call it "virtual" importance sampling.

2.6. Summary and set of parameters

We want to stress once more that the above methods do not bias the result, as opposed to delta-scaling methods. Also, DDIS is much more general than the similar methods invented by [10] and [14], which are only applicable for the back-scattering geometry and small optical depths, in particular for lidars. Cloning and splitting further enables us to speed up simulations where photons performing more than just a few scatters contribute significantly to the result.

The complete set of methods described above runs under the acronym VROOM (Variance Reduction "Optimal Options" Method).

The exact settings for the parameters defined above needs to be found empirically; for passive instruments we found the following set to be the most efficient:

$$\varepsilon_{ddis} = 0.1$$

$$n_{spl} = int(w_{spl}), \quad w_{crit,spl} = 3, \quad n_{spl,max} = 5000$$

$$w_{crit,rr} = 1/3, \quad w_{rr,min} = 0.2$$

$$n_{firstCP} = 4, \quad n_{scCP} = 12, \quad n_{LEperCP} = 11$$

DDIS and PBS are applied to the MP during the first $n_{firstCP}$ scatters only. They are always applied to the CPs. PBRR is always applied.

Note that for extremely forward peaked phase functions it might be necessary to increase n_{scCP}, which should be of the order of the number of scatterings needed to make the photon direction isotropic.

Finally, an efficient method to reduce the computational time of the LE without significantly reducing the convergence rate has been proposed by Iwabuchi [18]. This method has also been implemented into VROOM. In short, the calculation of the extinction term of the LE is expensive, involving the creation of a "virtual" photon that travels the path between the scatter point and the detector and sums up the total extinction in between. Therefore, a method similar to Russian roulette is applied to these "virtual" photons when they reach a critical optical depth. The method is described in detail in section 3a of [18]. However, in VROOM, a simplified version has been implemented in which the phase function (called Ψ_n in the referenced paper) is not taken into account. Also, we use $\tau_{max} = 3$.

3. Validation and timings

The above methods have been implemented into the Monte Carlo code MYSTIC [19,37]. In the following we compare results for a typical cloud scene computed with VROOM and with plain LE in order to validate VROOM. The results are also compared with results computed with a delta-scaling method in order to show the advantages of VROOM with respect to biasing acceleration methods.

3.1. Setup

In order to validate VROOM, several simulations were performed on a scene containing a thunderstorm cloud, illustrated in Fig. 11. The clouds were taken from the large eddy simulation described in [24] after 270 min of simulation time. The scene has an extension of 64 km × 64 km, a horizontal resolution of 250 m, and a vertical resolution of 200 m. It contains ice clouds with effective radii between 1 and 56.3 μm, as well as water clouds with effective radii between 4.9 and 15.9 μm. The maximum optical depths are 407 for the ice clouds and 742 for the water clouds. The ice cloud optical properties were parameterized according to [33,38,39]. The water cloud optical properties were calculated using the Mie theory for gamma size distributions:

$$n = n_0 r^\alpha \exp\left(-\frac{(\alpha+3)r}{r_{eff}}\right) \tag{13}$$

with $\alpha = 7$, corresponding to an effective variance of 0.1; r_{eff} is the effective radius. As aerosol we used a parameterization by [40], the settings correspond to a rural type aerosol
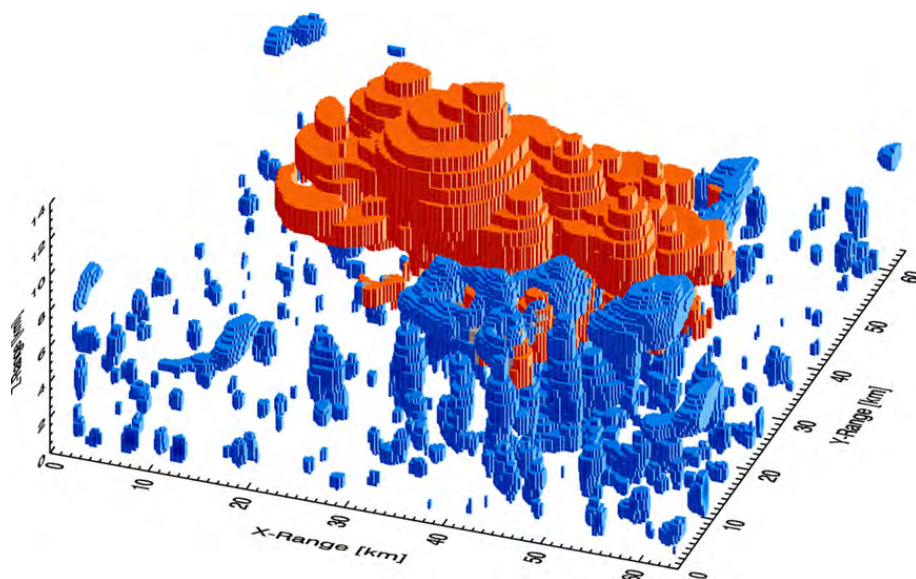
Fig. 11. 3D illustration of the cloud scene. Red boxes are filled with ice clouds, blue boxes with water clouds. (Zinner, private communication). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the boundary layer, background aerosol above 2 km, spring–summer conditions and a visibility of 50 km. The aerosol scattering was parameterized by a Henyey–Greenstein phase function. The surface was assumed to reflect Lambertian with an albedo of 0.1. A standard mid-latitude summer atmosphere was used [41].

All simulations were performed at a wavelength of 500 nm, with the sun at a solar zenith angle of 30°, being 30° west of the southern direction. In the first set of simulations, denoted "S" for satellite, the radiance at the top of atmosphere was calculated for a detector, looking at an angle 10° from nadir, and from the north-east. In the second set, denoted "C" for circum-solar, the radiance at the surface, looking towards the south-west at a detector zenith angle of 30° and an azimuth of 30.2°, i.e. looking 0.1° past the center of the sun, was calculated in order to validate the circum-solar radiation. The sun was taken to be a point source of light. All radiances were averaged over pixels of 250×250 m^2 size. The radiances were simulated applying the backward Monte Carlo method. For each pixel, 10^4 photons were used. The resulting diffuse radiance fields can be seen in Figs. 12 and 13 (Models S_VROOM and C_VROOM, respectively, see below). Because of limitations in computational time the statistical test simulations were performed for only one line of pixels, see the red lines in Figs. 12 and 13.

For each geometry, three types of simulations were tested: Type "PLAIN" was performed with LE as the only VRM, type "DELTA" with LE and a delta-scaling method based on geometric truncation of the scattering phase function, and type "VROOM" with the methods described in this paper. The delta-scaling method works as follows: scattering angles with $\mu > 0.99$ are no longer allowed, and the scattering coefficient is rescaled, such that effectively scattering angles with $\mu > 0.99$ now correspond to non-scattering.

Each simulation was run 1000 times in order to evaluate the statistics of the methods. Since the two PLAIN simulations converge very slowly, they were run 5000 times each.
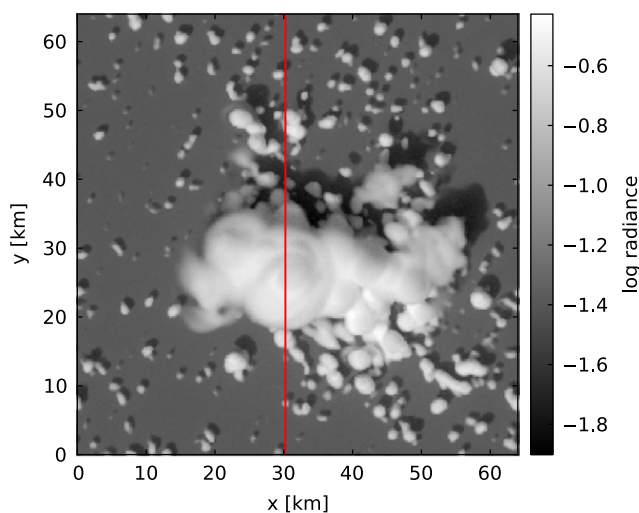
Fig. 12. Radiance fields as a result of the simulation S_VROOM. The plot can be interpreted as a satellite image of the scene. The red line depicts the line of pixels for which the statistical tests were performed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Timings and biases

Table 1 lists the pixel-averaged CPU time for a simulation with 10^4 photons, run at a CPU speed of 2.3 GHz, $\langle t_{CPU} \rangle_j$, as well as the pixel-averaged standard deviation $\langle \sigma \rangle_j$, where the standard deviation of pixel j for a single simulation is estimated using (for a derivation of the equation see Appendix A)

$$\sigma(j) \simeq \sqrt{\frac{1}{N} \sum_{i=1}^{N} [\sigma_i^2(j) + x_i^2(j)] - x^2(j)} \tag{14}$$

Here, $x_i(j)$ is the radiance calculated by simulation run i for pixel j, $\sigma_i(j)$ is the corresponding standard deviation and N
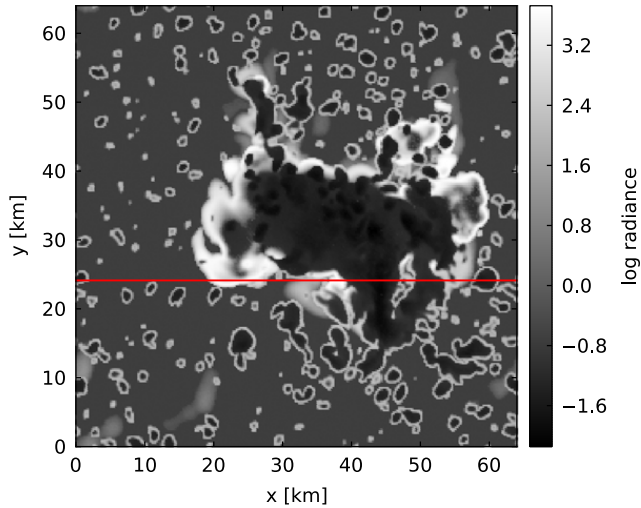
Fig. 13. Radiance fields as a result of the simulation C_VROOM. The image depicts the diffuse radiance hitting the ground at an angle of 0.1° from the sun center. Hence the radiance is very low below the thick thunderstorm cloud, but strongly enhanced at the thin edges of the clouds. The red line depicts the line of pixels for which the statistical tests were performed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Convergence speed of the different simulations.

Model	$\langle \sigma/x \rangle_j$ (%)	$\langle t_{\mathrm{CPU}} \rangle_j$ (s)	$\langle t_{1\%} \rangle_j$ (s)	$\langle \delta \rangle_j \pm \sigma_{\langle \delta \rangle}$ (0.01 %)
S_PLAIN	51.5	29.1	1.7×10^5	-9 ± 6
S_DELTA	1.86	13.2	49.6	10 ± 0.5
S_VROOM	1.88	65.6	211	–
C_PLAIN	53.2	48.7	2.1×10^6	4 ± 10
C_DELTA	2.11	23.3	216	-4200 ± 0.6
C_VROOM	3.00	18.6	1039	–
C_VROOM*	2.75	18.5	373	–

All numbers are averages over all 256 pixels, denoted by the brackets $\langle \rangle_j$. $\langle \sigma/x \rangle_j$ is the relative standard deviation, $\langle t_{\mathrm{CPU}} \rangle_j$ the computational time, both correspond to simulations with 10^4 photons per pixel. $\langle t_{1\%} \rangle_j$ is the estimated computational time to reach a relative standard deviation of one percent. $\langle \delta \rangle_j$ is the systematic bias introduced by the different simulation methods with respect to VROOM. Finally, $\sigma_{\langle \delta \rangle}$ denotes the precision with which the bias could be calculated. Because VROOM is taken as reference value, no bias is shown. C_VROOM* is the same as C_VROOM, but the pixel with longest convergence time at $x=39.25$ km was omitted in the calculation.

is the number of simulation runs. $x(j)$ is the photon-averaged radiance of pixel j:

$$x(j) = \frac{1}{N} \sum_{i=1}^{N} x_i(j) \qquad (15)$$

illustrated in Figs. 14a and 15a.

In order to compare the convergence speed of the different methods it is better to look at the computational time it takes to reach a relative standard deviation of 1%. Hence, Table 1 also contains the pixel-average of this normalized computational time, defined by

$$t_{1\%}(j) \simeq t_{\mathrm{simulation}}(j) \left(\frac{\sigma(j)}{0.01 x(j)} \right)^2 \qquad (16)$$
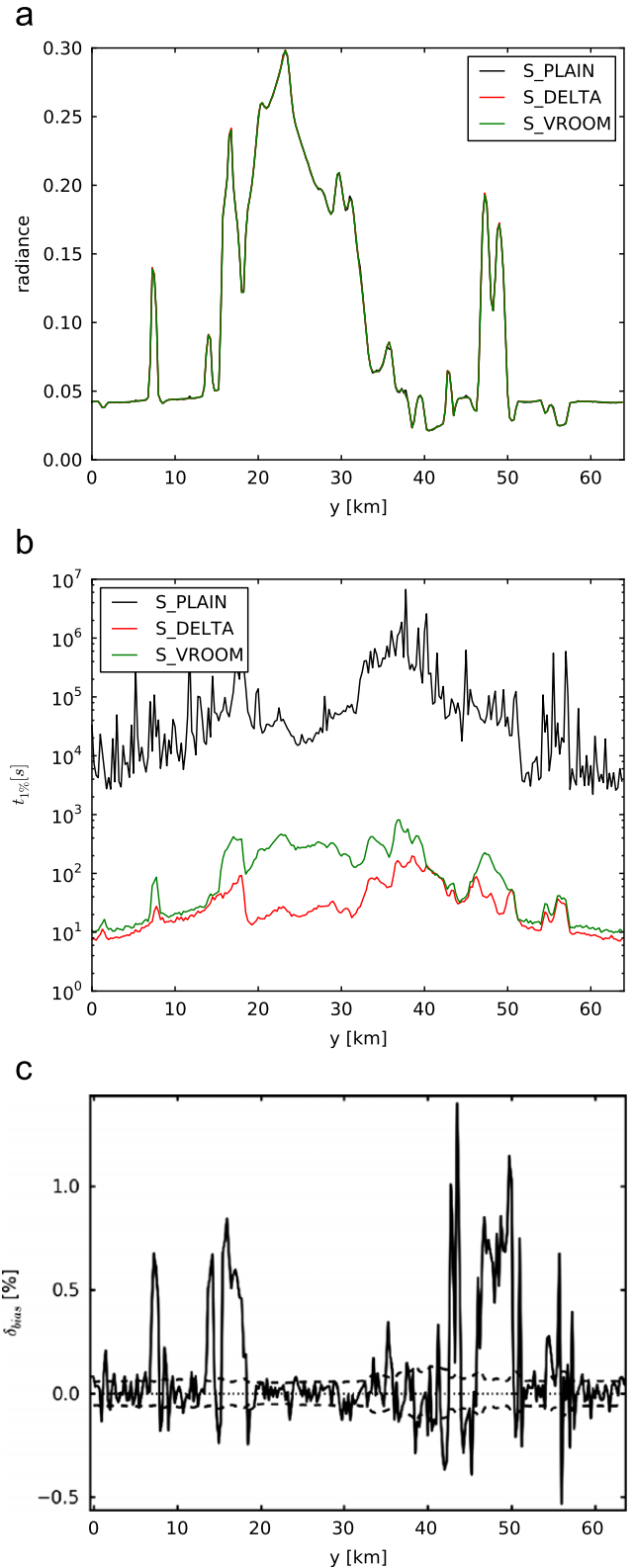
Fig. 14. Results for the simulations with satellite geometry ("S"). (a) Calculated radiances by all simulations. All three lines lie almost on top of each other. (b) Time it takes to achieve one percent precision for each pixel individually. (c) Bias of simulation S_DELTA with respect to S_VROOM. The dashed lines depict the standard deviation of the bias.

This is a good approximation in case the result of the Monte Carlo simulation is normally distributed. The results are illustrated in Figs. 14b and 15b.
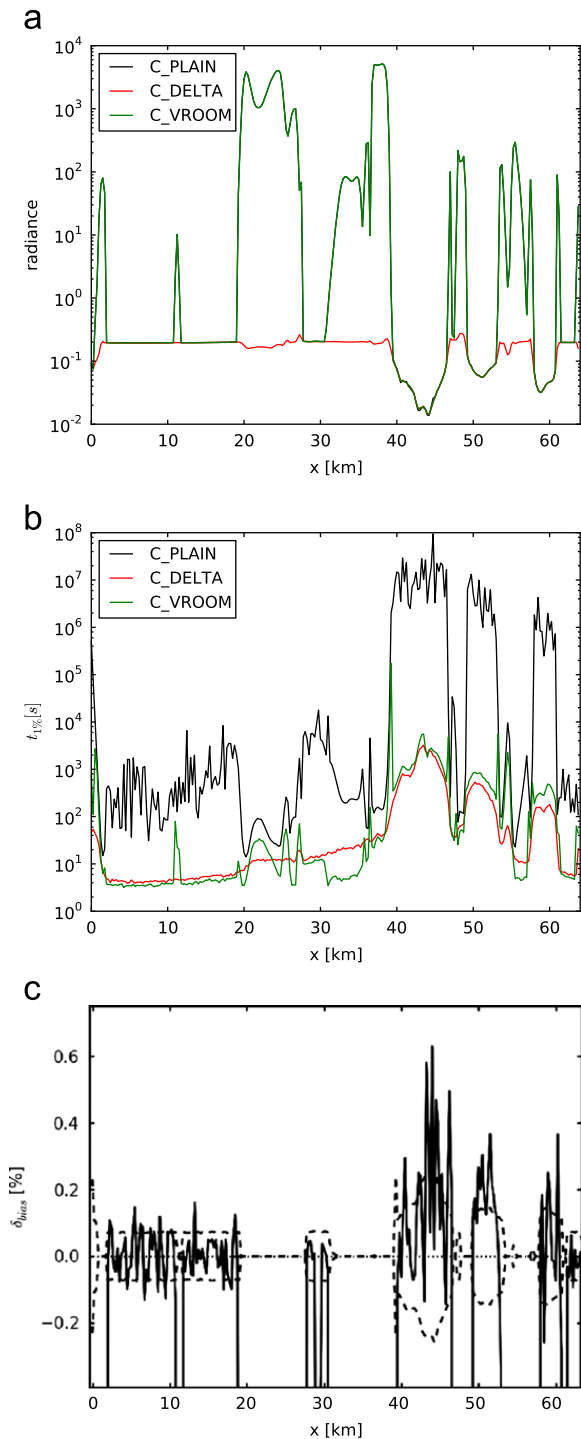
Fig. 15. Results for the simulations looking towards the sun ("C"). (a) Calculated radiances by all simulations. The line for "C_PLAIN" lies below the green line. (b) Time it takes to achieve one percent precision for each pixel individually. (c) Bias of simulation C_DELTA with respect to C_VROOM. The dashed lines depict the standard deviation of the bias. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, we should have a look at the bias introduced by the acceleration techniques. To this end, we look at the bias of each pixel j, defined as

$$\delta(j) = \frac{x(j) - x_{\text{REF}}(j)}{x_{\text{REF}}(j)} \qquad (17)$$

where x_{REF} is the result from the reference run. The corresponding standard deviation is

$$\sigma_{\langle \delta \rangle}(j) \approx \frac{x(j)}{x_{\text{REF}}(j)} \sqrt{\left(\frac{\sigma(j)}{x(j)}\right)^2 + \left(\frac{\sigma_{\text{REF}}(j)}{x_{\text{REF}}(j)}\right)^2} \qquad (18)$$

Because the PLAIN simulations have too large variance, they cannot be used as reference, which is why we used the VROOM simulations as reference. In Figs. 14c and 15c we only compare DELTA with VROOM, assuming VROOM to be unbiased.

We can also evaluate the systematic bias, i.e. the bias (Eq. (17)) averaged over all computed pixels. The systematic bias and the corresponding standard deviation can be seen in Table 1 for both the PLAIN and the DELTA simulations, with VROOM as reference model.

From the numbers for PLAIN simulations in Table 1 it is obvious that some kind of variance reduction is needed in order to achieve convergence in a feasible time. Looking at the satellite geometry ("S"), we see that on average DELTA converges about four times faster than VROOM. However, as can be seen in Fig. 14b, DELTA can be up to 18 times faster in convergence, especially when looking at thick clouds. Still, if we have a look at Fig. 14c, we see that the bias made by DELTA can be up to 1.4%, in particular near strongly inhomogeneous clouds. Although this does not seem to be a large number, larger biases are likely to be obtained in other geometries (see e.g. [25]). This suggests that VROOM should be used when one is interested in unbiased radiances.

Looking in the direction of the sun ("C"), the advantages of VROOM become even more pronounced. Except for very few exceptions (the pixel at $x=39.25$ km, which will be discussed at the end of Section 3.3) the convergence times of VROOM and DELTA are of similar order (see Fig. 15b and Table 1). The bias made by DELTA, however, can be several orders of magnitude, as expected, because the forward peak of the scattering phase function has been cut off due to delta-scaling. Hence, when calculating circum-solar radiances, VROOM is the only option.

Finally, the systematic bias between the PLAIN and the VROOM simulations are in both geometries of the order of the standard deviation (see Table 1), i.e. is probably caused by Monte Carlo noise. We can conclude that a possible systematic bias introduced by VROOM is smaller than 0.1%.

3.3. Spikiness

The main reason for developing these VRMs was to reduce spikes. This can best be analyzed if we check how close the different MC simulations are to a normal distribution of the results. This can be illustrated in a q–q-plot, or quantile–quantile-plot. The plots that will be presented here can be derived as follows:

Let us define a normalized deviation of the radiance of pixel j obtained during simulation run i from its average radiance:

$$q_i(j) \equiv \frac{x_i(j) - x(j)}{\sigma(j)} \qquad (19)$$

For the normal, or Gauss, distribution, the probability distribution function is given by

$$p_{gauss}(q) = c\exp(-q^2) \tag{20}$$

and the cumulative distribution function (CDF) is

$$P_{gauss}(q) = \int_{-\infty}^{q} dq' p_{gauss}(q') \tag{21}$$

where c is a normalization factor defined by $P_{gauss}(\infty) = 1$. The CDF of the data can by calculated using

$$P_{data}(q) = \frac{1}{N} \sum_{i=1}^{N} \Xi[q - q_i] \tag{22}$$

where N is the number of simulation runs, and $\Xi[x]$ is the Heaviside step function, being 1 for $x > 0$ and 0 for $x < 0$.
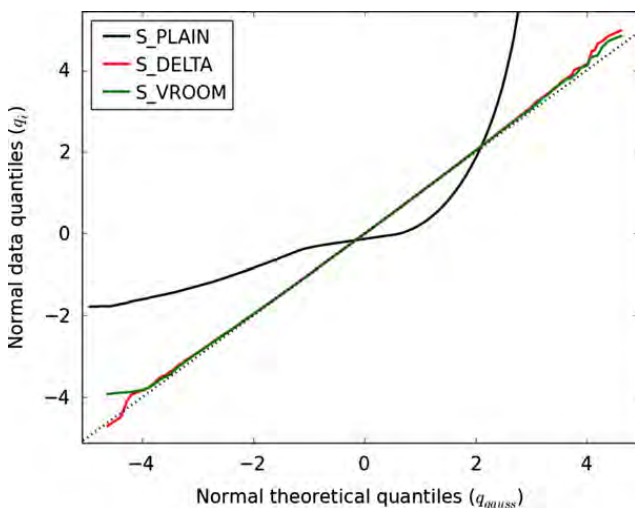
If we invert Eq. (22) to $q_i(P_{data})$, and set $P_{data} = P_{gauss}$ (q_{gauss}), we obtain $q_i(q_{gauss})$, which defines the q–q-plot. $q_i(q_{gauss})$ can be interpreted such that the normalized deviation from the average, q_i, occurs with a probability that corresponds to the probability with which one would expect a deviation q_{gauss} in case of a normal distribution. Hence, ideally we would like the VRM to obtain $q_i(q_{gauss}) = q_{gauss}$.

Figs. 16 and 17 show the q–q-plots for the two geometries simulated, including all types of simulations. In order to include all pixels at once into the plot, we used, instead of Eq. (22),
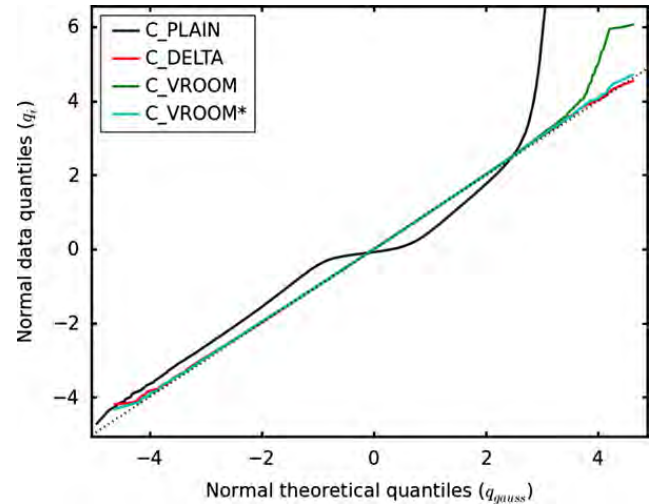
$$P_{data}(q) = \frac{1}{NM} \sum_{j=1}^{M} \sum_{i=1}^{N} \Xi[q - q_i(j)] \tag{23}$$

where M is the number of pixels, i, e. $M = 256$. In case all pixels are normally distributed, this sorting of all $q_i(j)$ will also lead to a normal distribution, whereas deviations from normal distribution will analogously be visible.
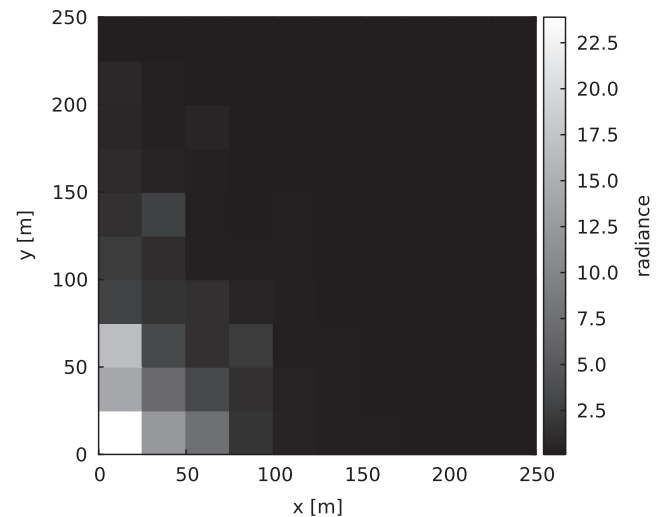
Clearly, in Fig. 16, the run without VRM produces spikes, visible in the strong deviation from a normal distribution at

Fig. 17. q–q-plot for all simulations looking towards the sun ("C"). The special line C_VROOM* is identical to C_VROOM, except that the results from the pixels $x = 39.25$ km and $x = 0$ km were removed.
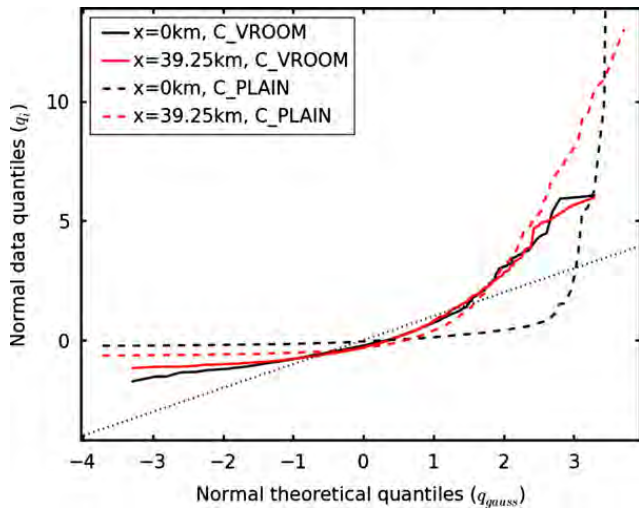

Fig. 18. High resolution picture of pixel $x = 39.25$ km.

large q_{gauss}. For S_VROOM and S_DELTA, however, the distribution of results is very close to the normal distribution so that we can argue that both VROOM and DELTA do not produce spikes.

The same is valid for C_DELTA, see Fig. 17. For C_VROOM, however, we see a moderate deviation from gaussianity for large q_{gauss}. However, if we omit the pixels at $x = 0$ and 39.25 km, this deviation completely vanishes (C_VROOM* in Fig. 17).

Fig. 18 shows a high resolution simulation of the pixel at $x = 39.25$ km, where the pixel has been subdivided into 10×10 sub-pixels. The lower left sub-pixel contributes about 10% of the total signal, and the 20 brightest sub-pixels contribute 86% of the total signal. Since the photons are started evenly distributed over the full pixel, this means that only few photons will contribute significantly to the signal. The slow convergence in case of the pixel at $x = 39.25$ km is hence due to the strong optical heterogeneity of the pixel, i.e. that it is situated below a thick cloud, but at the border of the

Fig. 16. q–q-plot for all simulations with the satellite geometry ("S"). The interpretation of this plot is that the deviation from the average $q_i\sigma$ occurs with a probability which would be expected for a deviation $q_{gauss}\sigma$. Hence the dashed line corresponds to the normal distribution, and the closer a line lies to the dashed line, the more Gaussian the method is. Lines deviating toward very large values for large q_{gauss} denote methods with spikes.

Fig. 19. q–q-plot of the two problematic pixels, calculated with and without VROOM. The line for $x=0$ km without variance reduction (black dashed) reaches values of 68.

cloud where a small part of the pixel is illuminated by the sun through a thin cloud layer. The situation is similar for the pixel at $x=0$ km.

This example shows that, even with all the tricks used in VROOM, it can still not be guaranteed that the result will converge quickly. However, as can be seen in Fig. 19, VROOM does lead to a roughly Gaussian distribution even in the worst cases. Hence, the standard deviation that can be calculated for a given simulation can be trusted and allows to estimate the required computational time.

Recall Fig. 4, where it was shown that the PLAIN simulation seems to converge after 5000 photons, and then suddenly an unexpected spike lets the standard deviation increase dramatically. The sceptic reader would argue that the same could be true for the VROOM calculations, but that the simulations were simply not run long enough in order to encounter a spike. Yet, while hundreds of spikes were encountered in the PLAIN calculations (0.1% of the data points result in $q_i > 10$), not a single one was created with VROOM calculations, even though 5×10^9 photons were simulated. Although this is no proof of the absence of spikes when using VROOM, it does give a good feeling.

4. Conclusions

We have developed a set of variance reduction methods (VRM) for performing Monte Carlo simulations of radiances in the presence of clouds that does not bias the result, but at the same time converges in reasonable time even for phase functions with extreme forward peak. The VRMs needed for this are detector directional importance sampling, n-tuple local estimate, and prediction-based splitting, as well as prediction-based Russian roulette and circum-solar virtual importance sampling.

The major advantage is that this set of VRMs prevents the Monte Carlo simulation to produce spikes, i.e. statistically rare but large contributions to the result. These spikes make an estimation of the statistical uncertainty

impossible, since before the first spike pops up the calculated standard deviation will be too small, and with the occurrence of the first spike will the standard deviation jump to a very large value.

The new VRMs were implemented into the Monte Carlo code MYSTIC [19] – which is part of the radiative transfer package libRadtran [26] – and run under the name Variance Reduction "Optimal Options" Method (VROOM). It was shown that VROOM is only slightly slower than delta-scaling. The parameter set defined for VROOM can possibly be improved in order to achieve faster convergence. However, it is unlikely that VROOM can be made much faster than delta-scaling, since the latter has to deal with about half as many scatters as VROOM and is thus naturally faster in computational time.

For calculations of radiances in interesting directions such as the rainbow or circum-solar radiation an unbiased variance reduction method such as VROOM is absolutely indispensable. But also for other geometries it is becoming more and more important to use an unbiased method in order to obtain correct radiances to a precision of better than 1%.

It needs to be stated that VROOM does converge slowly for few special conditions, e.g. when looking into the sun through the edge of a thin cloud. These problems can only be solved by analyzing the geometric peculiarities of the scene in advance of the simulation, a task which we found too tedious to automate. For the few cases where the convergence is slow it is best to use good old brute force and increase the number of photons in the simulation.

Although we have described VROOM only in the context of scalar RTE without absorption, reflection, or polarisation, we would like to stress once more that VROOM does work together with these. Absorption and reflection were actually considered in the simulations presented in this paper, and VROOM with polarization has been validated in [27].

The next step in the development of the MC code MYSTIC will be to develop a Lidar simulator with using VROOM.

Acknowledgments

Appendix A. Proof of Eq. (14)

The squared standard deviation is defined as

$$\sigma^2 \simeq \frac{1}{M} \sum_{k=1}^{M} x_k^2 - \left(\frac{1}{M} \sum_{k=1}^{M} x_k \right)^2 \qquad (A.1)$$

where $M=NL$ is the total number of photons simulated, N is the number of simulations, L is the number of photons per

simulation, and x_k is the contribution of photon k to the result. We can write Eq. (A.1) as

$$\sigma^2 \simeq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{L}\sum_{j=1}^{L}x_{i,j}^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{L}\sum_{j=1}^{L}x_{i,j}\right)^2 \qquad (A.2)$$

where $x_{i,j}=x_k$ is the contribution of the jth photon of the ith simulation to the result, i.e. $k=(i-1)L+j$. For a single simulation, we have

$$x_i = \frac{1}{L}\sum_{j=1}^{L}x_{i,j} \qquad (A.3)$$

and

$$\sigma_i^2 \simeq \frac{1}{L}\sum_{j=1}^{L}x_{i,j}^2 - \left(\frac{1}{L}\sum_{j=1}^{L}x_{i,j}\right)^2 \qquad (A.4)$$

Reformulation of Eq. (A.4) leads to

$$\frac{1}{L}\sum_{j=1}^{L}x_{i,j}^2 \simeq \sigma_i^2 + \left(\frac{1}{L}\sum_{j=1}^{L}x_{i,j}\right)^2 \qquad (A.5)$$

Inserting Eqs. (A.3) and (A.5) into Eq. (A.2) leads to

$$\sigma^2 \simeq \frac{1}{N}\sum_{i=1}^{N}[\sigma_i^2 + x_i^2] - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2 \qquad (A.6)$$

Appendix B. Acronyms

CS-VIS	circum-solar virtual importance sampling
CP	clone photon
DDIS	detector directional importance sampling
LE	local estimate
MP	mother photon
NLE	n-tuple local estimate
PBRR	prediction-based Russian roulette
PBS	prediction-based splitting
SCP	split clone photon
VRM	variance reduction method
VROOM	variance reduction "optimal options" method

References

[1] Wapler K, Mayer B. A fast method for the three-dimensional calculation of surface irradiance within a cloud resolving model. J Appl Meteorol and Climatology 2008;47:3061–71.

[2] Zinner T, Mayer B, Schröder M. Determination of 3D cloud structures from high resolution radiance data. J Geophys Res 2006;111(D8).

[3] Joseph J, Wiscombe W, Weinman J. The Delta–Eddington approximation for radiative flux transfer. J Atmos Sci 1976;33:2452–9.

[4] Wiscombe WJ. The delta-M method: rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. J Atmos Sci 1977;34:1408–21.

[5] Breon FM, Goloub P. Cloud droplet effective radius from spaceborne polarization measurements. Geophys Res Lett 1998;25(11):1879–82.

[6] Mayer B, Schröder M, Preusker R, Schüller L. Remote sensing of water cloud droplet size distributions using the backscatter glory: a case study. Atmos Chem Phys 2004;4:1255–63.

[7] Wendling P. Albedo and reflected radiance of horizontally inhomogeneous clouds. J Atmos Sci 1977;34:642–50.

[8] Davies R. The effect of finite geometry on the three-dimensional transfer of solar irradiance in clouds. J Atmos Sci 1978;35:1712–25.

[9] Marchuk GI, Mikhailov GA, Nazaraliev MA. The Monte Carlo methods in atmospheric optics. Berlin: Springer Series in Optical Sciences; 1980.

[10] Platt CMR. Remote sounding of high clouds. III: Monte Carlo calculations of multiple-scattered lidar returns. J Atmos Sci 1981;38:156–67.

[11] O'Brien D. Accelerated quasi Monte Carlo integration of the radiative transfer equation. J Quant Spectrosc Radiat Transf 1991;48:41–59.

[12] Bruscaglioni P, Ismaelli A, Zaccanti G. Monte-Carlo calculations of lidar returns: procedure and results. Appl Phys B 1995;60:325–9.

[13] Starkov A, Noormohammadian M, Oppel U. A stochastic model and a variance-reduction Monte-Carlo method for the calculation of light transport. Appl Phys B 1995;60:335–40.

[14] Noormohammadian M. Varianzreduzierende monte-carlo methoden zur berechnung mehrfach gestreuter lidar rücksignale. PhD thesis, Ludwig-Maximilians-Universität München, Germany; 1996.

[15] Barker HW, Goldstein RK, Stevens DE. Monte Carlo simulation of solar reflectances for cloudy atmospheres. J Atmos Sci 2003;60:1881–94.

[16] Cahalan R, Oreopoulos L, Marshak A, Evans K, Davis A, Pincus R, et al. The international intercomparison of 3D radiation codes (I3RC): bringing together the most advanced radiative transfer tools for cloudy atmospheres. Bull Am Meteorological Soc 2005;86(9):1275–93.

[17] Marshak A, Davis A. 3D radiative transfer in cloudy atmospheres. Springer; 2005 (ISBN-13 978-3-540-23958-1).

[18] Iwabuchi H. Efficient Monte Carlo methods for radiative transfer modeling. J Atmos Sci 2006;63(9):2324–39.

[19] Mayer B. Radiative transfer in the cloudy atmosphere. Euro Phys J Conf 2009;1:75–99.

[20] Nakajima T, Tanaka M. Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation. J Quant Spectrosc Radiat Transf 1988;40(1):51–69.

[21] Hu YX, Stamnes K. Climate sensitivity to cloud optical properties. Tellus B Chem Phys Meteorol 2000;52B:81–93.

[22] Ding S, Xie Y, Yang P, Weng F, Liu Q, Baum B, et al. Estimates of radiation over clouds and dust aerosols: optimized number of terms in phase function expansion. J Quant Spectrosc Radiat Transf 2009;110(13):1190–8.

[23] Antyufeev VS. Solution of the generalized transport equation with a peak-shaped indicatrix by the monte carlo method. Russ J Numerical Anal Math Model 1996;11(2):113–37.

[24] Zinner T, Marshak A, Lang S, Martins JV, Mayer B. Remote sensing of cloud sides of deep convection: towards a three-dimensional retrieval of cloud particle size profiles. Atmos Chem Phys 2008;8(16):4741–57.

[25] Iwabuchi H, Suzuki T. Fast and accurate radiance calculations using truncation approximation for anisotropic scattering phase functions. J Quant Spectrosc Radiat Transf 2009;110(17):1926–39.

[26] Mayer B, Kylling A. Technical note: the libradtran software package for radiative transfer calculations—description and examples of use. Atmos Chem Phys 2005;5:1855–77.

[27] Kokhanovsky AA, Budak VP, Cornet C, Duan M, Emde C, Katsev IL, et al. Benchmark results in vector atmospheric radiative transfer. J Quant Spectrosc Radiat Transf 2010;111(12–13):1931–46.

[28] Kalesse H, Schmidt S, Buras R, Wendisch M, Mayer B, Emde C. et al., Spatial heterogeneity and ice crystal shape: effects on remote sensing of ice cloud optical thickness and effective crystal radius—a case study. J Geophys Res, submitted.

[29] Emde C, Buras R, Mayer B, Blumthaler M. The impact of aerosols on polarized sky radiance, validation, and applications: model development. Atmos Chem Phys 2010;10:383–96.

[30] Mishchenko MI, Travis LD, Lacis AA. Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering. Cambridge University Press; 2006.

[31] Mishchenko MI. Multiple scattering, radiative transfer, and weak localization in discrete random media: unified microphysical approach. Rev Geophys 2008;46:RG2003.

[32] Mishchenko MI. Gustav mie and the fundamental concept of electromagnetic scattering by particles: a perspective. J Quant Spectrosc Radiat Transf 2009;110(14–16):1210–22 XI Conference on Electromagnetic and Light Scattering by Non-Spherical Particles: 2008.

[33] Baum B, Heymsfield A, Yang P, Bedka S. Bulk scattering models for the remote sensing of ice clouds. Part 1: Microphysical data and models. J Appl Meteorol 2005;44:1885–95.

[34] House LL, Avery LW. The Monte Carlo technique applied to radiative transfer. J Quant Spectrosc Radiat Transf 1969;9(12):1579–91.

[35] Marshak A, Davis A, Wiscombe W, Cahalan R. Radiative smoothing in fractal clouds. J Geophys Res 1995;100(D12):26247–61.

[36] Mayer B, Kylling A, Madronich S, Seckmeyer G. Enhanced absorption of UV radiation due to multiple scattering in clouds: experimental

evidence and theoretical explanation. J Geophys Res 1998;103(D23): 31241–54.

[37] Emde C, Mayer B. Simulation of solar radiation during a total solar eclipse: a challenge for radiative transfer. Atmos Chem Phys 2007;7: 2259–70.

[38] Baum B, Yang P, Heymsfield A, Platnick S, King M, Hu YX, et al. Bulk scattering models for the remote sensing of ice clouds. Part 2: narrowband models. J Appl Meteorol 2005;44:1896–911.

[39] Baum B, Yang P, Nasiri S, Heidinger A, Heymsfield A, Li J. Bulk scattering properties for the remote sensing of ice clouds. Part 3: high resolution spectral models from 100 to 3250 cm^{-1}. J Appl Meteorol 2007;46:423–34.

[40] Shettle E. Models of aerosols, clouds and precipitation for atmospheric propagation studies. In: Atmospheric propagation in the uv, visible, ir and mm-region and related system aspects. No. 454, AGARD Conference Proceedings; 1989. p. 15–32.

[41] Anderson G, Clough S, Kneizys F, Chetwynd J, Shettle E. AFGL atmospheric constituent profiles 0–120 km. Technical Report AFGL-TR-86-0110, Air Force Geophys. Lab., Hanscom Air Force Base, Bedford, Mass.; 1986.