Linking snowflake microstructure to multi-frequency radar observations

J. Leinonen,1,2 D. Moisseev,3 and T. Nousiainen3

Received 21 September 2012; revised 14 December 2012; accepted 4 January 2013; published 19 April 2013.

[1] Spherical or spheroidal particle shape models are commonly used to calculate numerically the radar backscattering properties of aggregate snowflakes. A more complicated and computationally intensive approach is to use detailed models of snowflake structure together with numerical scattering models that can operate on arbitrary particle shapes. Recent studies have shown that there can be significant differences between the results of these approaches. In this paper, an analytical model, based on the Rayleigh-Gans scattering theory, is formulated to explain this discrepancy in terms of the effect of discrete ice crystals that constitute the snowflake. The ice crystals cause small-scale inhomogeneities whose effects can be understood through the density autocorrelation function of the particle mass, which the Rayleigh-Gans theory connects to the function that gives the radar reflectivity as a function of frequency. The derived model is a weighted sum of two Gaussian functions. A term that corresponds to the average shape of the particle, similar to that given by the spheroidal shape model, dominates at low frequencies. At high frequencies, that term vanishes and is gradually replaced by the effect of the ice crystal monomers. The autocorrelation-based description of snowflake microstructure appears to be sufficient for multi-frequency radar studies. The link between multi-frequency radar observations and the particle microstructure can thus be used to infer particle properties from the observations.


1. Introduction

[2] Precipitation remote sensing instruments in the microwave regime are commonly used to observe falling snow. Recently, the motivation to understand the scattering properties of snowflakes has been driven by the growing interest in using millimeter-wave radars to measure snowfall remotely. Such radars are used, in particular, for global observations by satellites such as the currently orbiting CloudSat, operated by the National Aeronautics and Space Administration (NASA), Global Precipitation Measurement (GPM) under development by NASA and the Japan Aerospace Exploration Agency (JAXA), and EarthCARE, which is being developed by the European Space Agency (ESA) and JAXA.

[3] A number of recent studies using accurate computational scattering models and detailed shapes [Ishimoto, 2008; Botta et al., 2010; Petty and Huang, 2010; Tyynelä et al., 2011] have indicated that at sizes that are large compared to the wavelength, the use of homogeneous spherical or spheroidal shape models leads to a misestimation of the backscattering cross section of snowflakes. On the other hand, at the small-particle limit, such models are valid (except for the effect of particle non-sphericity, especially with regard to polarimetric variables) under the Rayleigh scattering law, and their applicability has also been successfully extended further using other theoretical methods and found valid in practice [e.g., Leinonen et al., 2011; Hogan et al., 2012]. Leinonen et al. [2012] found that the spheroid-snowflake model is sometimes, though not always, incompatible with experimental results. In the worst case, the error in the backscattering cross section can be as large as orders of magnitude.

[4] The cross section that is obtained by using the spheroidal shape model depends on the selection of the effective size of the equivalent spheroid. There are a number of different ways to determine the effective size, including the maximum diameter, the equivalent-volume diameter, and the radius of gyration [see, e.g., Hogan et al., 2000; Donovan et al., 2004; Westbrook et al., 2006]. In practice, the maximum diameter has been commonly used because it is readily measurable. However, the results by Petty and Huang [2010], Kneifel et al. [2011], and Leinonen et al. [2012] indicate that an effective radius cannot be chosen
consistently with any method if more than two wavelengths are used.

[5] Because of the complexity of the shapes of snowflakes, parametrization of their scattering properties in terms of their physical properties is needed in order to interpret remote sensing measurements. The number of particles contained in the typical volume of interest (a single radar bin) is large, and thus, it is often sufficient to consider the average scattering properties. A very large number of parameters is usually required for a complete description of an aggregate particle, but it may be possible to simplify the particle model greatly as it is enough to consider only the average backscattering.

[6] One hypothesis that has been made in various forms [e.g., Fabry and Szyrmer, 1999] is to make spheroidal shape models more realistic by accounting for the decreasing average density of the snowflake as a function of the distance from its center. This feature is not captured by the fully homogeneous sphere and spheroid models. We show in this paper that this is not sufficient at large wavelength-size ratios; instead, the small-scale inhomogeneities of the internal structure of the snowflakes must be considered in order to reproduce the backscattering cross section. The structure can be understood in terms of the density autocorrelation function. Based on the Rayleigh-Gans approximation, we demonstrate a simple parametrization of this function in terms of the particle structure and provide theoretical justification for its form.

2. Theory of Scattering by Aggregates

2.1 Rayleigh-Gans Theory

[7] The Rayleigh-Gans approximation (RGA) can be used to compute the scattering properties of a particle in the absence of interactions between parts of the particle. Following the notation of Bohren and Huffman [1983], the RGA gives the backscattering cross section (also called the radar cross section) of a particle as

\[ \sigma_b = \frac{9k^4}{4\pi |K|^2} \left( \frac{m}{\rho} \right)^2 |f|^2. \]  

(1)

The notation is as follows: \( k = 2\pi/\lambda \) is the wave number, \( \lambda \) is the wavelength, the dielectric factor \( K = (n^2 - 1)/(n^2 + 2) \) where \( n \) is the complex refractive index of the material, \( m \) is the mass of the scatterer, \( \rho \) is the density of the material, and \( f \) is the Rayleigh-Gans form factor. The latter, a complex vector in the incident beam direction (defined here as being along the \( z \)-axis), and \( \hat{e}_z \) is a unit vector to the scattering direction. We define the probability density function (PDF) of the mass distribution of a particle, \( p(R) \), such that

\[ p(R) = \begin{cases} \frac{V}{r^4}, & \text{material is found at point } R \\ 0, & \text{otherwise} \end{cases} \]  

(3)

where \( V \) is the total volume of the particle; this definition satisfies the normalization property

\[ \int_{\mathbb{R}^3} p(R) \, dR = 1. \]  

(4)

[8] The form factor contains information about the shape and the orientation of the particle, acting as a modification to the Rayleigh scattering law; at the small-particle limit, \( f \rightarrow 1 \) and the RGA converges to the Rayleigh approximation for the equi-voluminal sphere. The form factor and, consequently, the RGA are intuitively understood as the integration of the independently scattered waves from all parts of the scatterer, neglecting any electromagnetic interaction between its parts.

[9] In the backscattering direction, \( \hat{e}_z = -\hat{e}_z \). Then the form factor is given by the Fourier transform of the mass distribution in the \( z \) direction, \( p(z) \):

\[ f = \int_{-\infty}^{\infty} \exp(-2ikz)p(z) \, dz = F[p(z)](2k), \]  

(5)

where we used the common notation for the Fourier transform of a function \( f(z) \), which is defined in this paper as

\[ F[f(z)](k) = \int_{-\infty}^{\infty} \exp(-ikz)p(z) \, dz. \]  

(6)

This elegant feature of the RGA has been pointed out and analyzed by, e.g., Sorensen [2001]; that author also presents the generalization to other scattering directions that is omitted here for the sake of brevity.

[10] Tyynelä et al. [2012] showed that the accuracy of the backscattering cross section given by the RGA is almost always within 6 dB, usually within roughly 2 dB, of the corresponding results of volume integral calculations performed with the discrete dipole approximation (DDA) method (see section 3.5 for a confirmation of these assumptions for the snowflakes used in this paper). Although these differences are not negligible and larger errors can possibly be found in cases that were not considered by Tyynelä et al. [2012], the RGA can be expected to reproduce the important features of snowflake radar backscattering, the effects of the interactions being minor corrections. This is particularly true if only qualitative analysis is required. The purpose of this paper is to understand the mechanisms and conditions that lead to the applicability (or lack thereof) of average-particle models, so the RGA is highly suitable for our purposes.

[11] Another convenient feature of the formalism given by (1) and (5) is that it is very general and independent of assumptions about the particle shape. The particle model can be a set of cubes in a regular lattice (such as with DDA calculations), a cluster of analytically defined shapes (as with cluster Mie calculations), a continuous and possibly inhomogeneous mass distribution, or even a fractal. For these, and indeed for practically any realistic ways to describe the particle, the distribution function of the particle mass is square integrable, and thus, \( f \) is well defined.

[12] We remark that there is some ambiguity in terminology between different authors using the RGA and related methods. We use \( f \) to denote the form factor, as defined in (2) and (5) according to Bohren and Huffman [1983], consistently throughout this paper. However, we also adopt the
2.2. Average-Particle and Second-Order Models

[13] As mentioned before, the number of scatterers measured simultaneously with a radar is very large, and thus, we recognize that the law of large numbers permits us to only consider \( \langle \sigma_i \rangle \), the expected value of the backscattering cross section. Considering particles of a fixed mass and material, it can be seen from (1) that to obtain \( \langle \sigma_i \rangle \), we need to know \( \langle S \rangle = \langle |f|^2 \rangle \), which is defined as

\[
\langle |f|^2 \rangle = \left( \int_{-\infty}^{\infty} \exp(-2ikz)p(z)\,dz \right)^2 = \langle |F[p(z)](2k)|^2 \rangle. \tag{7}
\]

The quantity on the right-hand side of this equation is the average power spectral density of the mass distribution. As it is well known especially in signal processing literature, the power spectral density is simply the Fourier transform of the autocorrelation function [Mitra, 2002]. As the Fourier transform is a linear operator on functions, the expected value can be moved inside the transform, and thus, \( \langle S \rangle \) as a function of \( k \) is simply related to the Fourier transform of the average density autocorrelation function \( R(z) \):

\[
\langle S(k) \rangle = \langle |f(k)|^2 \rangle = F[\langle R(z) \rangle](2k), \tag{8}
\]

where \( R(z) \) of a particle is defined in terms of its mass distribution in the beam direction, \( p(z) \), as

\[
R(z) = \int_{-\infty}^{\infty} p(z') p(z' + z)\,dz'. \tag{9}
\]

The connection between the form factor and the density autocorrelation function has been given previously in an aggregate scattering context by, for example, Sorensen [2001].

[14] For an ensemble of aggregates of various sizes and masses, an examination of (1) shows (as noted by, e.g., Westbrook et al. [2006]) that the above result can be generalized as follows: the deviation of the effective backscattering cross section from the Rayleigh assumption as a function of the wave number \( k \) can be determined from the \( m^2 \)-weighted average of the \( S(k) \) functions of the individual particles. We use the shorthand \( \langle S(k) \rangle_{m^2} \) throughout this paper to refer to this function of the particle ensemble. The inverse Fourier transform of this gives the \( m^2 \)-weighted average of the density autocorrelation functions, \( \langle R(z) \rangle_{m^2} \). This means that with knowledge of the average structure factor over the whole spectrum, it is possible to reconstruct the average density autocorrelation function perfectly. Furthermore, the radar reflectivity is proportional to \( \sigma_i k^4 \) and therefore (except for small variations in refractive index as a function of frequency) also to \( \langle S(k) \rangle_{m^2} \); this means that \( \langle S(k) \rangle_{m^2} \) gives the change of reflectivity with frequency relative to the low-frequency limit.

[15] Although radar measurements cannot determine \( \langle S(k) \rangle_{m^2} \) over the full frequency spectrum, multi-frequency radars can still be used to measure it at discrete points. Interpolating and extrapolating with the help of a proper model, it is possible to estimate \( \langle R(z) \rangle_{m^2} \); in fact, this function is the only physical property of the snowflake ensemble that can be directly retrieved using multi-frequency methods, with the assumption that the RGA is valid. Thus, we argue that physical and scattering properties of snowflakes should be related to each other through understanding how the physical properties affect the autocorrelation \( R(z) \). In the rest of this paper, we examine how the density autocorrelation function behaves in realistic snowflakes and how it can be inferred from measurements.

3. Model Experiments

3.1. Aggregate Snowflakes

[16] For analysis of snowflakes, model particles were generated with an iterative algorithm based closely on that by Westbrook et al. [2006] and Westbrook [2008]. The algorithm begins with an initial population of \( N \) randomly oriented monomers built from small volume elements. Since snowflakes aggregate primarily because of differences in their vertical velocities, the collision probabilities of particle pairs are computed on each iteration from the aerodynamic cross section and the mass of the particles, and according to these probabilities, a pair to collide is randomly selected. The larger particle in the pair impacts the smaller one directly from above at a random position on the horizontal plane; the particles stick together rigidly at first impact (the collisions are not perfectly efficient in nature [Pruppacher and Klett, 1997], but this can be neglected as the purpose of the algorithm is merely to build an aggregate from a given set of monomers). The resulting particle is oriented randomly and returned into the population; the process is repeated until only one particle remains.

[17] The monomers used in this study were dendrite ice crystals generated using the algorithm of Reiter [2005]. Their thickness was determined using an empirical relation given by Pruppacher and Klett [1997].

[18] For testing under different conditions, we generated five data sets consisting of aggregates generated using varying sizes and numbers of crystal monomers. The first three data sets consist of aggregates made up of monodisperse crystal monomers, each 0.5 mm in diameter. Although the crystals in natural snowflakes are not monodisperse, these data sets were included in order to carry out a simplified analysis of the microstructure, so that the essential features of the structural model can be demonstrated without the complicating factor of variable monomer size. The other data sets consist of aggregates where the monomer maximum diameter was sampled from an exponential distribution with a rate parameter \( \Lambda = 1 \text{ mm}^{-1} \), truncated at small and large sizes such that the maximum diameter of the monomers is in the range 0.2 mm < \( D_{\text{max}} < 3 \text{ mm} \). In one of the polydisperse data sets (PMIX), a mixture of differently sized aggregates was used. The properties of the data sets, along with the abbreviations used for them later in this section, are summarized in Table 1. The number of aggregates per data set varies because for some data sets, a larger
Table 1. An Overview of the Data Sets Used in Section 3. The Rightmost Columns Show the Average Maximum Dimension Along the z Axis \(D_{\text{max}}\) and the Average Mass \(m\)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Monomer PSD</th>
<th>Monom./aggregate</th>
<th>Num. aggregates</th>
<th>(D_{\text{max}}) (mm)</th>
<th>(m) (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5</td>
<td>Monodisp.</td>
<td>5</td>
<td>50</td>
<td>0.83</td>
<td>0.011</td>
</tr>
<tr>
<td>M50</td>
<td>Monodisp.</td>
<td>50</td>
<td>200</td>
<td>2.4</td>
<td>0.11</td>
</tr>
<tr>
<td>M300</td>
<td>Monodisp.</td>
<td>300</td>
<td>50</td>
<td>5.7</td>
<td>0.66</td>
</tr>
<tr>
<td>P100</td>
<td>Exponential</td>
<td>100</td>
<td>150</td>
<td>9.3</td>
<td>1.5</td>
</tr>
<tr>
<td>PMIX</td>
<td>Exponential</td>
<td>35/100/300</td>
<td>50/25/10</td>
<td>7.6</td>
<td>1.3</td>
</tr>
</tbody>
</table>

number of aggregates was generated to test the convergence properties better.

### 3.2. Average Mass Distribution

[19] The physical properties of a single snowflake are shown in Figure 1. It can be seen from Figure 1b that the single-particle mass distribution has no simple form. For better insights about the distribution of the typical positions of the monomers in the aggregates, we can instead examine the average mass distribution of a number of particles. In Figure 2, we show the probability distribution function (PDF) of the mass distribution (around the center of mass) of the data set M50. It is clear that the average mass PDF corresponds with a high accuracy to the normal (Gaussian) distribution, even though the mass distribution for the individual particles can be very different from it.

[20] Aggregate formation is a complex random process. Thus, the normal mass distribution is perhaps not surprising considering that according to the central limit theorem of probability theory [Jaynes, 2003], the PDF of the sum of random variables tends toward normal as the number of random variables increases. We posit, consistently with this result, that at the limit of a large number of aggregates, the average three-dimensional distribution of the positions of the monomers within an aggregate follows a three-variable normal distribution. The marginal distributions of a multivariate normal distribution are also normal, reproducing the above result.

[21] In Figure 2, we also show a mass distribution that corresponds to a homogeneous sphere with the same standard deviation as the normal distribution. It is a reasonably good approximation of the correct distribution, which shows that one way to justify the spheres or spheroids as models of aggregate particles is to consider them as models of the average mass distribution. For this reason, we shall refer to the Gaussian, spherical, and spheroidal shape models together as “average-particle models”.

### 3.3. Modeling the Density Autocorrelation Function

[22] In Figure 3a, we show the average density autocorrelation function \(\rho(z)_{\text{ave}}\) for the model snowflakes of data set M50, normalized such that \(\rho(0)_{\text{ave}} = 1\). Figure 3b depicts

![Figure 1.](image-url)  

**Figure 1.** An example of a single snowflake from the data set M300. (a) A silhouette of the snowflake. (b) The PDF of the above snowflake mass along the z axis. (c) The density autocorrelation function corresponding to the PDF of Figure 1b.

![Figure 2.](image-url)  

**Figure 2.** The distribution of mass around the center of mass for the aggregates of the data set M50. Gray bars: a normalized histogram of the average mass distribution of all aggregates in the data set; stepped purple line: a normalized histogram of the mass distribution of a single aggregate; dashed red line: normal distribution fit; dotted black line: spherical mass distribution with the same standard deviation as the normal distribution.
The corresponding \( \langle S(k) \rangle_m \) function, derived by taking the discrete Fourier transform (DFT) and given as a function of frequency \( v \) for easier interpretation \( (v = c k/2\pi \) where \( c \) is the speed of light). The functions determined computationally from the data are shown as the thick gray line. In Figure 3b, we also show, for various commonly used frequencies, dots corresponding to the \( m^2 \)-weighted average of \( S \) as calculated directly using (2). This verifies that the results of the indirect determination of \( S \) (by computing the autocorrelation function and taking the DFT) match those of the direct computation.

[23] At short distances \((\leq 1.0 \text{ mm})\), the autocorrelation function seems to fall off linearly as a function of distance. At longer distances, the slope of the function becomes less steep, gradually approaching zero and reaching it at the maximum snowflake diameter of M50, 4.7 mm. Closer examination reveals that at distances shorter than approximately 0.4 mm, the shape deviates slightly from linear, giving a narrower peak than a purely linear relationship would. The main features of \( \langle S(k) \rangle_m \) are a Gaussian shape at frequencies up to \( v \approx 65 \text{ GHz} \) (note the logarithmic scale in Figure 3b) and a gradual widening beyond that. In the following analysis, we construct a model that explains the features of \( \langle R(z) \rangle_m \) and, consequently, of \( \langle S(k) \rangle_m \).

[24] As a starting point, we use the assumption that the density autocorrelation function is a zero-centered Gaussian, as it would be if the mass distribution was normal for each individual aggregate. This function is given by

\[
R_{\text{avg}}(z) = \frac{1}{\sqrt{4\pi (\sigma_2^2 + \sigma_3^2)}} \exp \left( -\frac{z^2}{4(\sigma_2^2 + \sigma_3^2)} \right) \tag{10}
\]

where \( \sigma_m \) is the standard deviation of the monomer mass distribution and \( \sigma_a \) is the standard deviation of the monomer positions (see appendix A). This function is plotted as a dashed line in Figure 3. However, the Fourier transform of a Gaussian function is also a Gaussian [Jaynes, 2003], which corresponds to a parabola in Figure 3b. The autocorrelation function computed from the model aggregates is clearly not parabolic, and it can be seen that the Gaussian of Figure 3b falls off rapidly after approximately 90 GHz and quickly starts to deviate drastically from the computed value. Additionally, this assumption does not reproduce the shape of the function at short range: a Gaussian function would be wider around \( z = 0 \) than the \( \langle R(z) \rangle_m \) derived from synthetic data in Figure 3a.

[25] The discrepancy at high frequencies can be resolved by taking the effect of the monomers into account. As a simplistic model, one can consider an aggregate as a collection of \( N_e \) monomers whose mass distribution is normal with a standard deviation \( \sigma_m \) and whose positions within the aggregate are independent and identically distributed, also according to the normal distribution. Then the expected \( R(z) \) is as a weighted sum of two Gaussian functions (see appendix B). \( N_e \) should be interpreted as an “effective” number of particles and is not equal to \( N \), as the positions of the monomers are not truly independent as we assume in the derivation: nearly equally validly, the aggregate could be considered to consist of \( N/2 \) two-monomer clusters, \( N/3 \) three-monomer clusters, and so on. For the same reason, the size of the monomers needs to be corrected with a factor \( b \) to give an effective monomer size. Our model autocorrelation function is then

\[
\langle R(z) \rangle = \frac{N_e^2 - N_e}{N_e^2} \frac{1}{\sqrt{4\pi (b\sigma_m)^2 + \sigma_a^2}} \exp \left( -\frac{z^2}{4((b\sigma_m)^2 + \sigma_a^2)} \right) + \frac{N_e}{N_e \sqrt{4\pi(b\sigma_m)^2}} \exp \left( -\frac{z^2}{4(b\sigma_m)^2} \right) \tag{11}
\]

This model modifies the average-particle model of (10) with a correction that is relevant at high frequencies; at low frequencies, the average-particle term dominates.
The correction term corresponds to the effect of the inhomogeneities that the monomers introduce.

The linear behavior at \( z = 0 \) can be reproduced by using a truncated Gaussian. We approximate the average autocorrelation corresponding to a truncated normal mass distribution as \( G(z) \), a Gaussian multiplied by a triangular function:

\[
G(z) = C \exp \left( -\frac{z}{4\sigma^2} \right) \text{tri} \left( \frac{z}{D} \right)
\]  

where \( \sigma \) is the standard deviation of the mass distribution, \( a \) is a correction factor that is included to compensate for effects introduced by, e.g., truncation of the Gaussian, \( D \) is the maximum dimension of the snowflake in the \( z \) direction, and \( C \) is a normalization constant that ensures that the function is a PDF. The variables \( \sigma \) and \( D \) are both measures of the size of the particle. By analogy to the pure Gaussian model, the following autocorrelation function can be justified:

\[
\langle R(z) \rangle_{m^2} \approx \frac{N_e^2 - N_e}{N_e^2} G(z)_{(\sigma,D,a)} + \frac{N_e}{N_e^2} G(z)_{(\sigma_m,D,a)}
\]

Simply choosing representative \( \sigma \) and \( D \) is not sufficient to reproduce the shape at longer distances: for instance, using the average \( \langle D \rangle = 2.5 \text{ mm} \) of the particle ensemble would force \( G(z) = 0 \) at distances longer than that; this is obviously not correct. Although the particles were constructed with identical parameters of the aggregation code, the size of the particles is highly variable, \( D \) ranging from 1.5 mm to 4.7 mm. The length measures are also highly correlated: the correlation coefficient of \( \sigma \) and \( D \) was \( \rho = 0.90 \) with \( D = 3.3\sigma \) on average. In order to account for the joint distribution properly, we computed this function for each \( \sigma,D \) pair and took an \( m^2 \)-weighted average of the functions (this roughly amounts to an integration over a size distribution). The parameters \( b \) and \( N_e \) were determined by nonlinear least-squares fitting such that the model matched \( \langle S(k) \rangle_{m^2} \) as well as possible. The correction factor \( a \) is another free variable, but in practice, we found \( a = 2.0 \) to work very well for every tested case. There may be an underlying theoretical reason for this, but we did not investigate further in the current study. The parameters \( \sigma_m \) and \( D_m \) were sampled from the monomers using the same method. For the model of (11), the same approach was used, except for ignoring the maximum dimensions \( D \) and \( D_m \).

Using the Gaussian mixture model of (11) resulted in least-squares estimates \( N_e = 49 \) and \( b = 1.12 \) for the snowflakes of M50; this function is shown as the dash-dot line. The Gaussian-triangular mixture model (13) resulted in \( N_e = 142 \) and \( b = 1.00 \); the corresponding estimate is shown as the thin solid line. Both functions reproduce \( \langle R(z) \rangle_{m^2} \) and \( \langle S(k) \rangle_{m^2} \) quite accurately.
[29] The same models were applied on the data set M300 as shown in Figure 4. For this data set, we obtained $N_e = 40$ and $b = 3.88$ for the Gaussian mixture model (11) and $N_e = 235$ and $b = 3.20$ for the triangular-Gaussian mixture (13). For P100 (Figure 5), the convergence of the functions to the averages was not as good and thus there is some noise left in the structure factor. We achieved a good fit to data with $N_e = 50$, $b = 3.66$ for (11) and $N_e = 42$, $b = 5.89$ for (13).

[30] To simulate what might happen in the usual situation where particles of various sizes are contained in a radar volume, we applied the model to PMIX (Figure 6). The same model also seems to be applicable in this case, with $N_e = 45$, $b = 6.68$ for (11) and $N_e = 92$, $b = 7.57$ for (13). It is encouraging that there appears to be no need to increase the number of free parameters when differently sized aggregates are present.

[31] Finally, to demonstrate a case where the average-particle models are approximately applicable for all commonly used radar frequencies, we used the same approach for the data set M5 (Figure 9). In this case, the estimated parameters were $N_e = 2$, $b = 2.48$ for (11) and $N_e = 3$, $b = 1.50$ for (13).

[32] In practice, we found $b$ and $N_e$ to be difficult to decouple. Similar results were obtained with considerable increases of $b$ if $N_e$ was decreased to compensate, or vice versa. This is perhaps not surprising considering the ambiguity of interpreting these parameters and the ad hoc
nature of their introduction. Less intuitively, values of \( N_c \) larger than \( N \) were found in a few cases. It is not clear at present if there is a physical reason for this or if it is an artifact of the fitting procedure.

### 3.4. Implications for Multi-Frequency Radar Measurements

[33] Determining the physical properties of snowflakes from radar measurements is an important practical measurement problem. Multi-frequency measurements can measure \( \langle S(k) \rangle_m \) at discrete points, from which the whole function can be interpolated, preferably with the aid of a suitable model. Unfortunately, the shape of the \( \langle S(k) \rangle_m \) function is only partially predictable and can exhibit variation outside the range of measurements. The highest frequencies that would be required for full reconstruction would often be higher than the current frequencies used in cloud radars, around 95 GHz at maximum, although the frequency range needed in practice would depend on the size of the snowflakes and the monomers. Attenuation constraints limit the usability of radars heavily at higher frequencies.

[34] From the number of free parameters of the \( \langle S(k) \rangle_m \) function, one can estimate the minimum number of colocated radar measurements required to reconstruct the function. Because different masses and sizes are present in a radar volume, the size or mass distribution requires 2–3 parameters, including the number concentration. Additionally, a mass-dimensional relationship of (typically) two parameters is needed. Those quantities are also needed separately for the effective monomers in order to model their effects properly. Then apparently the number of free parameters can be quite large, perhaps 8–10 depending on how many simplifications can be made. This is more than is practically ever simultaneously available from real measurements (although it may be possible to obtain more information from polarimetric and Doppler data), but since the \( \langle S(k) \rangle_m \) functions in Figures 3–6 do not seem to be particularly complex in shape, we consider it unlikely that the number of parameters required to represent them would be that high. It is not evident how this function should be constructed using fewer parameters; we think that this problem should be treated thoroughly in a later work. As an observation from our current results, however, we note that at low frequencies, the form of \( \langle S(k) \rangle_m \) follows a Gaussian function even for ensembles of differently sized aggregates. This allows a rough estimation of the representative particle size from low-frequency measurements, as demonstrated in Figure 7. This finding is also consistent with Guinier’s law, which states that at low frequencies, for a constant \( c \), \( |f(k)|^2 \approx 1 - ck^2 \) are the first two terms of the series expansion of a Gaussian function.

### 3.5. Comparison to Exact Models

[35] The performance of RGA calculations on snowflakes was examined in detail by Tyynelä et al. [2012]. The snowflake models used in that study were, however, slightly different from ours, and thus, we made our own comparisons to ensure that the RGA is also valid for computing the backscattering from our snowflakes. As a reference, we used DDA calculations performed using the ADDA software [Yurkin and Hoekstra, 2011], as well as various models that are exact for analytical shapes.

[36] First, we tested the RGA against a sphere model, where each sphere was assigned the same mass and radius of gyration as the corresponding snowflake. The backscattering cross section was computed using Mie scattering, which is exact for spheres. The electromagnetic properties of the snowflake were modeled using the Maxwell-Garnett effective medium approximation (EMA) computed from the snow density, using ice as the inclusions and air as the matrix (which is probably justified at low ice densities, characteristic of snowflakes). Second, due to the popularity of layered spheres as snowflake models [e.g., Fabry and Szyrmer, 1999; Szyrmer et al., 2012], we used a two-layered sphere, again conserving the mass and the radius of gyration and optimizing the remaining parameters such that the density reproduces that of the corresponding Gaussian model as accurately as possible. For a Gaussian particle with a radius of gyration \( R_g \), this yields the radius of the inner layer as \( a = 1.08R_g \), the radius of the outer layer as \( b = 1.36R_g \), and a density ratio of 4.91 between the inner and outer layers. The layered model corresponded to Model 4 of Fabry and Szyrmer [1999]; the inner layer density was low enough to require the use of the same EMA (Maxwell-Garnett with ice as inclusions) for both layers. Third, because the Gaussian model was earlier found to be the most realistic average-particle model, we also compared to that, but because the inhomogeneity of the Gaussian model makes it intractable for Mie solvers, we used the RGA solution for Gaussian models instead.

[37] In Figure 8, we show the backscattering cross section as a function of frequency for the data sets M50 (Figure 8a) and P100 (Figure 8b), calculated using the above-mentioned models. For Figure 8b, only a 20-aggregate subset of P100 was used due to the high computational requirements of the DDA with such large targets. The agreement of the DDA and the RGA is quite good, especially when compared to their difference to the average-particle models. There is no significant difference in the performance of the single- and dual-layered sphere models. On the other hand, the backscattering cross section from the Gaussian model vanishes rapidly to extremely small values. At high frequencies, the difference of the average-particle backscattering cross sections and those from the DDA is much larger than the difference of the DDA and the RGA. From this, we can conclude that the use of the RGA is valid for the purposes of this paper and that the difference of average-particle and detailed snowflake models is primarily due to the difference in the complexity of the shapes, with electromagnetic interactions (which are neglected by the RGA but modeled by the DDA and Mie methods) playing a comparatively minor role.

### 4. Discussion

[38] The description of scattering from aggregate snowflakes presented in this paper provides a physical basis for why, and when, the average-particle models of snowflakes are able (or unable) to describe backscattering from snowflakes. If the average-particle model is used, it is assumed that the mass density functions of similar-sized particles average out. We have demonstrated why it is the density autocorrelation function, and not the mass density, that is averaged when a large number of particles are considered together. The difference of the two approaches can
be quite dramatic, especially when the structure factor is close to zero on average, but exhibits variability between individual particles.

[39] It can be summarized from the detailed treatment of the above sections that the following approach can be used to calculate the average backscattering cross section of an ensemble of simultaneously measured snowflakes:

[40] 1. Determine density autocorrelation functions \( R(z) \) of each snowflake using (9).

[41] 2. Calculate the mass-squared weighted average of the autocorrelation functions, \( \langle R(z) \rangle_{w^2} \).

[42] 3. Determine the mass-squared weighted average structure factor \( \langle S(k) \rangle_{w^2} \) as a function of the wave number \( k \) by taking the Fourier transform of \( \langle R(z) \rangle_{w^2} \) according to (8).

[43] 4. Calculate the average backscattering cross section as a function of the wave number as

\[
\sigma_b(k) = \frac{9k^4}{4\pi|K|^2} \left( \frac{m^2}{\rho} \right) \langle S(k) \rangle_{w^2}.
\]  

(14)

[44] We formulated a particle model based on the Gaussian function, which, with its tailed structure, was found in section 3.2 to be a more realistic model of the average mass distribution than a spherical model. The Gaussian function also has a number of extremely convenient properties, among others the central limit theorem, being its own Fourier transform, and the convolution of two Gaussian models producing another Gaussian. These properties not only support the model theoretically but also allow the derivation of analytical results that facilitate the understanding of the observed features of the density autocorrelation function and the structure factor. However, the same general reasoning about the difference of autocorrelation-based and average-particle models can be expected to apply to the spherical and spheroidal particle models even if the details are different. The departure from the average-particle models is due to small-scale inhomogeneities (the monomer crystals) that are not accounted for in the Gaussian or spherical descriptions. Although the decrease in the structure factor with frequency exhibited by the spherical models is not as drastic as that of the Gaussian model, and thus, the spheres seemingly perform better at high frequencies (see Figure 8); this is attributable mainly to a mathematical coincidence rather than better modeling of the particle microphysics by the spherical models.

[45] Other authors [e.g., Schmitt and Heymsfield, 2010] have treated the inhomogeneities in aggregate snowflakes by using fractal geometry. The connection between the fractal dimension and the density autocorrelation function has been given by, for example, Sorensen [2001]. We have not used the fractal description explicitly in this paper for a number of reasons. Most importantly, the results obtained by constructing the density autocorrelation function statistically are more quantitative and yield more insight into the role of monomers as the inhomogeneities that cause the failure of the average-particle model at high frequencies. Additionally, describing the particles with a single fractal dimension ignores the fact that because of the different growth processes of the aggregates and the monomers, different length scales may exhibit different fractal behavior: at high frequencies, it is the fractal dimension of the monomers that is relevant (and different high-frequency behavior would be expected for monomers with a different fractal dimension). Furthermore, when using the fractal description for particles of limited size, the variation in average-particle density with distance from the center must be accounted for by using a cut-off function [Sorensen, 2001]. In the description presented here, these limited-size effects arise naturally.

[46] Our approach can also be compared with that of Westbrook et al. [2006], who constructed a fit between two known extremal behaviors of backscattering. Their model matched the directly calculated values of the structure factor well, but the parameters of their polynomial fit lack physical context, which was sought with the model presented here; even though empirical parameters are also found in our model, they are formulated as adjustments to a physically
based model instead. It is interesting to note that an approximately Gaussian shape of $\langle S(k) \rangle_{W}$ is apparent in their Figure 5.

[47] The derivation of the density autocorrelation function shown in appendix B shows that the effects of monomers emerge as a correction to the average-particle model. This correction is relevant at short distances and, correspondingly, at high frequencies. This also shows that the average-particle models can be expected to be accurate at low frequencies, whereas they are likely to fail at high frequencies where the monomers become relevant. As the Gaussian function is a better average-particle description than the sphere, but diverges extremely rapidly above some given frequency, we suggest that it could, in some cases, be used as an indicator for the applicability of average-particle models: if the values given by the Gaussian model are implausibly small at a given frequency, the spherical model can also be expected to be inaccurate at that frequency.

[48] The same derivation also demonstrates a physical explanation as to why dual-frequency measurements can retrieve the size of aggregate snowflakes: the shape of the structure factor as a function of frequency is dependent only on the size of the aggregates at low frequencies; the monomers are only relevant at higher frequencies. The particle size is determined from the $m^2$-weighted average density autocorrelation function $\langle R(z) \rangle_{m^2}$; this results in a similar “representative size” measure as the mass-weighted mean diameter (although it is not obvious how exactly these two measures are related to each other). On the other hand, the mass-dimensional ($m$-$D$) relationship only affects $\langle R(z) \rangle_{m}$ by determining the weighting of differently sized snowflakes from the ensemble. This suggests that the effect of the $m$-$D$ relationship on $\langle R(z) \rangle_{m}$ is difficult to discern, and thus, the coefficients of the $m$-$D$ relationship may not be directly retrievable from multi-frequency measurements.

[49] It is noteworthy that the large dynamic range of the average structure factor $\langle S(k) \rangle_{m}$ means that the effects of the monomers are essential at high frequencies, even though they are almost unnoticeable in $\langle R(z) \rangle_{m}$. This unfortunately means that $\langle R(z) \rangle_{m}$ would have to be determined with extremely high precision in order to reconstruct $\langle S(k) \rangle_{m}$ properly, requiring measurements of the mass PDF at a spatial resolution on the order of the spatial scale of the monomer microstructure. Conversely, there is a positive implication for multi-frequency measurements: $\langle R(z) \rangle_{m^2}$, and thus, the retrieved particle size is not very sensitive to errors in the estimate of $\langle S(k) \rangle_{m^2}$, as could be seen in Figure 7.

[50] Since the sizes of the snowflakes we used were fairly realistic, we are able to make some predictions regarding the applicability of average-particle models. In all cases that we tested, the C- and K-Band measurements are in the aggregate regime, which corresponds to the average-particle description. For cases other than that of Figure 9, the W-band measurements, and with larger particles also the K-Band measurements, were in the transition or monomer regimes. While the commonly used spherical and spheroidal models may not give correct results when measuring large snowflakes with millimeter-wavelength radars (K-Band and higher frequencies), they are applicable at lower frequencies. It should also be emphasized that our usage of the term “lower frequencies” extends well beyond the Rayleigh regime.

5. Summary

[51] In this paper, we have constructed a model that explains the main features of the radar backscattering properties of aggregate snowflakes. Our model is based on treating the aggregates as collections of monomers that are distributed within the aggregates according to the normal distribution, which we found to be realistic in the average case. The model is based on the Rayleigh-Gans approximation that, while not exactly correct, has been found to be quite accurate for snowflakes and allows for much more straightforward mathematical analysis of the connection between microphysical and backscattering properties of the aggregates, facilitating the understanding of the scattering properties at various frequencies.

[52] Most importantly, the new description addresses the question: when and why are spherical or spheroidal snowflake shape models applicable for calculating the radar backscattering properties of aggregate snowflakes? We approached the question by comparing the density autocorrelation functions resulting from these particle models with those computed for realistic aggregates. Under the Rayleigh-Gans approximation, it can be shown that...
the backscattering intensity of multiple simultaneously observed scatterers is equal to that of a single particle whose density autocorrelation function is the mass-squared weighted average of the density autocorrelation functions of the individual particles. Therefore, it is justified to define the “effective” scatterer in terms of the autocorrelation.

[S] Spheres or spheroids are commonly used as simplified shape models, but we observed that the average mass distribution of aggregate snowflakes is very close to the Gaussian distribution. Therefore, we also included a Gaussian mass distribution in our analysis, using the term “average-particle model” for these types of particle shape models together. The Gaussian distribution may be more realistic as a mass distribution model, but more importantly, it allowed us to construct a simple model of an aggregate as a collection of a finite number of discrete monomers. Computing the autocorrelation function, and the resulting backscattering cross section, showed that the Gaussian average-particle model emerges as a low-frequency approximation to the aggregate model. At high frequencies, the differences between the two approaches can be very large. The difference of “low” and “high” frequencies depends on the size of the snowflake, but the mass-squared weighting of density autocorrelation functions of different simultaneously observed aggregates means that large snowflakes, if present, should not be ignored even if the number of small snowflakes is much higher.

[S4] Computation of the density autocorrelation function from our aggregate model assumptions results in a weighted sum of two terms. The first term is equal to the Gaussian average-particle model; the weight of this term is generally much larger than that of the second, but the term diminishes rapidly as the frequency increases. The second term corresponds to the monomers; its smaller weight means that it is irrelevant at low frequencies, but this Gaussian is wider in the frequency domain, and thus remains significant after the effect of the first, narrower, term vanishes.

[S5] When we compute the autocorrelation function given by the model for each aggregate in a data set of 50 or more, and then take the mass-squared weighted average of the contributions, the directly calculated autocorrelation function is reproduced well, as is the corresponding function that gives the structure factor (the square of the absolute value of the Rayleigh-Gans form factor) as a function of frequency. As the clustering of monomers violates our assumption that their positions are independent, we introduced additional parameters to the model that correct for this effect; these are derived through numerical optimization. We also considered an optional correction that includes particle edges, with the effect of the first, narrower, term vanishes.

[S6] This paper has built a framework that illuminates a number of features of radar backscattering from snowflakes but also raises a number of questions and goals for future study. The main weakness of the presented model is its dependence on numerically optimized parameters. The determination of these (or some equivalent parameters) from the physical properties of the particle is perhaps the most important goal of future work. The implications for retrieving the particle microphysical properties from measurements, either by multi-frequency radars or in situ instruments, should also be investigated more thoroughly.

### Appendix A: Gaussian Average-Particle Model

[S7] We denote a Gaussian PDF with mean $\mu$ and variance $\sigma^2$ as $N(\mu, \sigma^2)$. We shall use the result that the convolution of two Gaussian PDFs $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is itself a Gaussian PDF, $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ [Jaynes, 2003].

[S8] Assume that an aggregate particle is made up of $N$ monomers whose locations along the beam axis are distributed according to $N(0, \sigma^2_m)$. For simplicity, we assume that the monomer positions are independent and that the mass within each monomer is distributed according to identical normal mass distributions $N(0, \sigma^2_m)$. The PDF of the average mass density of the aggregate particles is the cross correlation of these two distributions. For symmetric mass distributions, this is equal to the convolution, which is $N(0, \sigma^2_m + \sigma^2_m)$. Thus, the average mass density PDF is

$$p_{\text{avg}}(z) = \frac{1}{\sqrt{2\pi (\sigma_m^2 + \sigma_m^2)}} \exp \left( -\frac{z^2}{2(\sigma_m^2 + \sigma_m^2)} \right). \quad (A1)$$

The density autocorrelation function is computed as the cross correlation of $p_{\text{avg}}(z)$ with itself. Again, the distribution is symmetric so we can directly write

$$R_{\text{avg}}(z) = (p_{\text{avg}}(z) \otimes p_{\text{avg}}(z)) = \frac{1}{\sqrt{4\pi(\sigma_m^2 + \sigma_m^2)}} \exp \left( -\frac{z^2}{4(\sigma_m^2 + \sigma_m^2)} \right) \quad (A2)$$

where “$\otimes$” denotes convolution.

### Appendix B: Average Autocorrelation for Gaussian Monomer Distribution

[S9] Here we shall calculate the average autocorrelation function for the particles described in appendix A. We shall use the theorems given above, and additionally, the fact that if two random variables are distributed according to $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, their difference is distributed according to $N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$ (this results directly from the rule for the convolution and the symmetry of the PDFs).

[S10] The mass density PDF for a single particle is a normalized sum of the mass density PDFs of the monomers:

$$p(z) = \frac{1}{N} \sum_{i=1}^{N} \exp \left( -\frac{z-z_i}{2\sigma_m^2} \right) \quad (B1)$$

The autocorrelation for this mass density PDF is

$$p(z) \ast p(z) = p(-z) \otimes p(z). \quad (B2)$$

We can compute the convolution pairwise as

$$R(z) = p(z) \otimes p(z) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi \sigma_m^2}} \exp \left( -\frac{(z-z_i)^2}{2\sigma_m^2} \right) \otimes \frac{1}{\sqrt{2\pi \sigma_m^2}} \exp \left( \frac{(z-z_j)^2}{2\sigma_m^2} \right)$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \exp \left( -\frac{(z-z_i)^2 + (z-z_j)^2}{4\sigma_m^2} \right) \quad (B3)$$

where we use the rule of the convolution of Gaussian models. We can divide the sum over $i$ and $j$ in two parts: $i = j$ and
In the former case, the monomers are the same. In the latter case, their positions are independent, and we have

\[ R(z) = \frac{1}{N^2} \frac{1}{4\pi \sigma_m^2} \left( \sum_{i=1}^{N} \sum_{j=i+1}^{N} \exp \left( -\frac{(z - (z_i - z_j))^2}{4\sigma_m^2} \right) \right) \]

\[ + \frac{1}{N^2} \frac{1}{4\pi \sigma_m^2} \left( \sum_{i=1}^{N} \exp \left( -\frac{z_i^2}{4\sigma_m^2} \right) \right) \]  

(B4)

Taking the average over possible locations of particles \(i\) and \(j\), and recognizing that \(d_{ij} = z_j - z_i\) is the distance between two monomers, we can write

\[ \langle R(z) \rangle = \frac{1}{N^2} \frac{1}{4\pi \sigma_m^2} \left( (N^2 - N) \exp \left( -\frac{(z - d)^2}{4\sigma_m^2} \right) \right) \]

\[ + N \exp \left( -\frac{z^2}{4\sigma_m^2} \right) \]  

(B5)

The average on the right-hand side can be computed from the normal difference distribution. As the monomer positions \(z_i\) are distributed according to \(N(0, \sigma_m^2)\), the average over possible positions is

\[ \frac{1}{\sqrt{4\pi \sigma_m^2}} \left( \exp \left( -\frac{z^2}{4\sigma_m^2} \right) \right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi \sigma_m^2}} \exp \left( -\frac{z^2}{4\sigma_m^2} \right) \frac{1}{\sqrt{4\pi \sigma_m^2}} \exp \left( -\frac{d^2}{4\sigma_m^2} \right) \, dd \]

\[ = \frac{1}{\sqrt{4\pi \sigma_m^2 \sigma_n^2}} \exp \left( -\frac{z^2}{4\sigma_m^2} \right) \right) \frac{1}{\sqrt{4\pi \sigma_n^2}} \exp \left( -\frac{z^2}{4\sigma_n^2} \right) \]

\[ = \frac{1}{\sqrt{4\pi \left( \sigma_m^2 + \sigma_n^2 \right)^2}} \exp \left( -\frac{z^2}{4(\sigma_m^2 + \sigma_n^2)} \right) \]  

(B6)

Then the resulting average autocorrelation function is

\[ \langle R(z) \rangle = \frac{N^2 - N}{N^2} \frac{1}{\sqrt{4\pi \left( \sigma_m^2 + \sigma_n^2 \right)^2}} \exp \left( -\frac{z^2}{4(\sigma_m^2 + \sigma_n^2)} \right) \]

\[ + \frac{N}{N^2} \frac{1}{4\pi \sigma_m^2} \exp \left( -\frac{z^2}{4\sigma_m^2} \right) \]  

(B7)

that is a weighted mixture of Gaussian. For similar applications of the RGA on stochastically shaped particles, see Muinonen [1996, 2000].

[61] Acknowledgments. This work was supported by the Academy of Finland (contracts 125180, 128255, and 128328). We also thank Karri Muinonen for his helpful comments.

References


