Journal of Quantitative Spectroscopy & Radiative Transfer 201 (2017) 115–127

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Applying machine learning methods for characterization of hexagonal prisms from their 2D scattering patterns – an investigation using modelled scattering data

Emmanuel Oluwatobi Salawu [a,b,c], Evelyn Hesse [d,*], Chris Stopford [d], Neil Davey [c], Yi Sun [c,*]

[a] TIGP Bioinformatics Program, Academia Sinica, Taipei, Taiwan
[b] Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan
[c] School of Computer Science, University of Hertfordshire, Hatfield, Herts, UK
[d] School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield, Herts, UK

ARTICLE INFO

Article history:
Received 23 March 2017
Revised 4 July 2017
Accepted 4 July 2017
Available online 5 July 2017

Keywords:
Machine learning
Scattering pattern
Hexagonal prisms
Ice crystals
Size
Aspect ratio
Ray tracing with diffraction on facets
Zernike moments

ABSTRACT

Better understanding and characterization of cloud particles, whose properties and distributions affect climate and weather, are essential for the understanding of present climate and climate change. Since imaging cloud probes have limitations of optical resolution, especially for small particles (with diameter < 25 µm), instruments like the Small Ice Detector (SID) probes, which capture high-resolution spatial light scattering patterns from individual particles down to 1 µm in size, have been developed. In this work, we have proposed a method using Machine Learning techniques to estimate simulated particles' orientation-averaged projected sizes (PAD) and aspect ratio from their 2D scattering patterns. The two-dimensional light scattering patterns (2DLSP) of hexagonal prisms are computed using the Ray Tracing with Diffraction on Facets (RTDF) model. The 2DLSP cover the same angular range as the SID probes. We generated 2DLSP for 162 hexagonal prisms at 133 orientations for each. In a first step, the 2DLSP were transformed into rotation-invariant Zernike moments (ZMs), which are particularly suitable for analyses of pattern symmetry. Then we used ZMs, summed intensities, and root mean square contrast as inputs to the advanced Machine Learning methods. We created one random forests classifier for predicting prism orientation, 133 orientation-specific (OS) support vector classification models for predicting the prism aspect-ratios, 133 OS support vector regression models for estimating prism sizes, and another 133 OS Support Vector Regression (SVR) models for estimating the size PADs. We have achieved a high accuracy of 0.99 in predicting prism aspect ratios, and a low value of normalized mean square error of 0.004 for estimating the particle's size and size PADs.

1. Introduction

Cloud feedbacks are a large source of uncertainty in climate models [1]. In particular, there are uncertainties about the radiative forcing of clouds containing ice crystals, especially cirrus. Whether cirrus clouds warm or cool the Earth's surface depends on ice crystal morphology [2]. Reducing this uncertainty requires detailed in situ characterization of cloud particles. Cloud probes based on imaging techniques, such as the Cloud Particle Imager (CPI) have limitations of optical resolution when dealing with particles smaller than about 25 µm [3]. These limitations do not apply to instruments detecting light scattering patterns of cloud particles, like the Small Ice Detector (SID) developed at the University of Hertfordshire [4].

However, to obtain particle information from the scattering patterns requires solving the inverse scattering problem. This is often facilitated by previous knowledge of how scattering properties of small particles depend on the particle size parameter, morphology, relative refractive index and orientation. This knowledge may be the cumulative result of investigating many specific cases combined with interpolation or extrapolation of particle characteristics not covered by existing theoretical or experimental data [5]. In order for such databases to be created, either computational models which can determine the intensity of light scattered by a known particle into a given angular range, or experimental data are required. Scattering by spheres can be described by Lorenz-Mie theory. For computation of non-spherical particles, numerically ex-

* Corresponding authors.
E-mail addresses: emmanuel@gapp.nthu.edu.tw (E.O. Salawu), e.hesse@herts.ac.uk (E. Hesse), c.stopford@herts.ac.uk (C. Stopford), n.davey@herts.ac.uk (N. Davey), comrys@herts.ac.uk (Y. Sun).

http://dx.doi.org/10.1016/j.jqsrt.2017.07.001
0022-4073/

act methods like T-matrix [6,7] or Discrete Dipole Approximation (DDA) [8,9] can be applied, however they are currently computationally very expensive and are therefore restricted to small particle sizes. This leaves a size parameter range that is covered neither by exact methods nor by geometric optics. Approximate methods, such as the geometric optics approximation or physical optics, have to be used for scatterers much larger than the wavelength. In the classical geometric optics approximation, scattered light is divided into two parts, firstly light reflected or transmitted by the scatterer, and secondly externally diffracted light [e.g. 10]. Diffraction of reflected and refracted light is neglected, resulting in singular intensity peaks. Therefore, this method is not applicable for interpreting measured 2D scattering patterns. Improved methods including diffraction of the ray-tracing component have been presented e.g. in [11–15], and a volume integral method in [15,16]. Ulanowski et al. [17] retrieved the size of particles with rough and complex surfaces from two-dimensional scattering patterns by investigating the median surface area of intensity peaks. The RTDF model [14] has been used to generate a database of 2D light scattering patterns (2DLSPs) [18] investigated here.

Machine Learning is ideally suited for estimating a range of properties of scattering particles from linked known parameters. Baran and Newman [19] have demonstrated the application of principal component analysis to estimate cloud ice water content and environmental temperature from bulk integral optical properties of ice cloud particles. Here, we wish to derive single particle properties from their 2DLSP, which are in general quite complex. Radial Basis Function (RBF) neural networks have been applied to solve the inverse light scattering problem for spheres [20]. In order to discriminate potentially hazardous respirable fibres, such as asbestos, Kaye et al. [21] have used experimental data to train RBF neural networks. Genuer et al. [22] applied machine learning on scatterograms of microcolonies for discriminating bacteria and yeasts at an early stage of growth. For this, they projected the patterns on either the Zernike orthogonal basis or a Fourier-Bessel function basis. The radial function of the latter was found to be more useful for analysis of the patterns, which consisted largely of concentric rings.

The work presented here aims to investigate the applicability of advanced machine learning methods to solve the inverse problem for hexagonal prisms. They are a useful test case, since virtually all the ice on Earth's surface and in its atmosphere is of a hexagonal crystalline structure. Due to their symmetry, scattering patterns of hexagonal prisms are quite different from the largely concentric patterns discussed in [22]. At this initial stage we disregard complex crystals and any surface roughness. Since roughness has found to be important [2], this will restrict the direct applicability of the results obtained in this study, but we aim to demonstrate a proof of concept here.

For any particle except a sphere the 2DLSP will depend on its orientation with respect to the incident laser beam. However, for a highly symmetric particle like a hexagonal prism there will be a continuous range of particle orientations which result in 2DLSP's differing only by a rotation around the direction of incident light (see Section 2). Such orientations should be attributed to *one single representative orientation*. This means that the representations of 2DLSPs which are the inputs of each computational model should be rotationally invariant. Zernike polynomials, which were originally derived to assist the characterization of imperfections of concave mirrors by analyzing their diffraction pattern [23] and have been used for other surface reflectance applications [24,25] seem a suitable approach to achieve this. Building on Zernike polynomials and the general theory of moments, Teague [26] derived the Zernike moments, in which the Zernike polynomials have been used as the basis functions for the moments, and applied them to visual pattern recognition [27]. This method has been widely applied in pattern recognition [28–31].

The aim of this work is two-fold: 1) to investigate if machine learning methods can be applied for estimating characteristics of small particles from 2DLSP using the example of 2DLSP of hexagonal prisms computed using RTDF; 2) to investigate if Zernike moments are suitable for representing 2DLSPs, and to know the range of the moments that should be used.

We employ machine learning methods, such as Random Forests Classification (RFC) [33], Support Vector Classification (SVC) and the Support Vector Regression (SVR) methods, to estimate characteristics of small particles based on sets of representations including Zernike moments, summed intensities and root mean square contrasts from each 2DLSP image. The hexagonal prism properties we wish to estimate from a 2DLSP are aspect ratio, projected size and orientation averaged projected size (generally, particle projected size is orientation dependent). By projected size we mean the diameter of a circle with the same area as the projected cross section of the particle. The orientation-averaged projected size is described by the diameter of a circle with the same area as the orientation averaged projected cross section of the particle (PAD denotes average projected area diameter). In our previous work [34], we have developed a method combining a Feed Forward Multi-Layer Perceptron neural network with Bayesian regularization back-propagation and rotation invariance with Fast Fourier Transform. However, the model could not deal with orientation averaged projected size, and the model could not predict the size of very small particles (between 3 and 10 µm size) with the same precision as it did for the larger particles.

The rest of paper is organized as follows. In Section 2, we introduce the dataset used in this work. In Section 3, we describe our method and give details on approaches we have applied in this work. We evaluate our approach and show experimental results in Section 4. We conclude our paper in Section 5 by discussing the potential use of our approach.

2. Dataset

Atmospheric ice crystals are typically of an intermediate size range that can neither be covered by exact electromagnetic methods like T-matrix, because they are currently computationally very expensive, nor by classical geometric optics. Here, we used the Ray Tracing with Diffraction on Facets (RTDF) model [14], which is a hybrid method combining ray-tracing with a physical optics approximation for diffraction, for computing the 2DLSP of hexagonal prisms. The 2DLSP are azimuthally resolved phase functions P_{11} [32], i.e. the intensity has been normalised to 4π over the complete scattering sphere. In general, incident light can be described by its Stokes vector $[I_i Q_i U_i V_i]^{\mathrm{T}}$. As a result of the scattering event, it is transformed into the Stokes vector $[I_s Q_s U_s V_s]^{\mathrm{T}}$ of the scattered field. This corresponds to multiplying $[I_i Q_i U_i V_i]^{\mathrm{T}}$ with the 4×4 phase matrix, $\mathbf{P}(\theta,\varphi)$ of the particle (Eq. (1)).

$$\begin{pmatrix} I_s \\ Q_s \\ U_s \\ V_s \end{pmatrix} = \begin{pmatrix} P_{11} P_{12} P_{13} P_{14} \\ P_{21} P_{22} P_{23} P_{24} \\ P_{31} P_{32} P_{33} P_{34} \\ P_{41} P_{42} P_{43} P_{44} \end{pmatrix} \begin{pmatrix} I_i \\ Q_i \\ U_i \\ V_i \end{pmatrix} \tag{1}$$

Unpolarised light of unit intensity has the Stokes vector $[1\ 0\ 0\ 0]^{\mathrm{T}}$. This results in an intensity $I_s = P_{11}$ of the scattered light. The laser in the SID instrument emits circularly polarised light to minimize polarization-dependent variations in the captured particle scattering patterns [5]. Circularly polarised light of unit intensity has the Stokes vector $[1\ 0\ 0\ \pm\ 1]^{\mathrm{T}}$. The sign of the V parameter will be positive for right-handed and negative for left-handed light. This results in an intensity $I_s = P_{11} \pm P_{14}$ of the scattered light. The phase matrix element P_{14}, which is also called circular