# PARAMETER ESTIMATION FOR MULTIPLE SCATTERING PROCESS ON THE SPHERE

*Florent Chatelain,*

Gipsa-lab, DIS
University of Grenoble,
Grenoble Campus, France.

*Nicolas Le Bihan*[*]*, Jonathan H. Manton*

The University of Melbourne
EEE Department
Parkville 3001, Australia.

## ABSTRACT

This paper considers the problem of parameter estimation for multiple scattering process on the sphere. Using harmonic analysis, a Fourier expansion of the *pdf* of the process is obtained. Based on the Fourier coefficient statistics, we consider the problem of estimating the parameter of the process using an Approximate Bayesian Computation (ABC) approach. Simulations show the ability of the proposed approach for the density estimation of intensity and concentration parameters for the von Mises Fisher multiple scattering process.

***Index Terms—*** Multiple scattering process on the sphere, Harmonic analysis, Method of Moments, Bayesian Inference, Approximate Bayesian Computation, von Mises-Fisher distribution.

## 1. INTRODUCTION

The description of multiple scattering using random processes is well known in Physics, especially in speckle and backscattering description. In many cases, scattering description consider spatial disorder information and describe a scattering medium using random fields [1]. In this paper, we consider the case of *forward* multiple scattering of a vector valued variable (direction of propagation). The random process of interest here thus describes the time evolution of a unit vector, *i.e.* an element of the unit sphere $\mathbb{S}^2$ in three dimensional space. The presented results can be formulated for any dimension $p$, *i.e.* for elements of the hyperspheres $\mathbb{S}^{p-1}$ in $\mathbb{R}^p$, but focus is made on the case $p = 3$.

Estimation for such processes has been investigated using non-parametric methods in [2, 3], where authors used the Henyey-Greenstein distribution to model the scatterers effect. The decompounding technique was also applied to geometric phase analysis using an EM approach in [4]. Recently, the use of von Mises Fisher distribution has also been proposed in [5] to provide approximation of the density of the multiple scattering process and perform estimation of the scattering parameters.

In this paper, we propose to use Approximate Bayesian Computation (ABC) to perform inference on the multiple scattering process parameters. This method is compared to the Generalized Method of Moments (GMM) estimator thanks to an adapted version of the Population Monte Carlo (PMC) algorithm. Simulations illustrate the interest of using ABC approach for the von Mises-Fisher multiple scattering process characterization in different scenarios.

[*]N. Le Bihan's research was supported by the ERA, European Union, through the International Outgoing Fellowship (IOF GeoSToSip 326176) program of the 7th PCRD

## 2. MULTIPLE SCATTERING PROCESS ON $\mathbb{S}^2$

The purpose of the multiple scattering process model is to describe the distribution of the output direction of propagation, denoted $\boldsymbol{x}_t \in \mathbb{S}^2$, when a particle/wave has traveled through a random medium for a time $t$. Such medium is considered to be made of a homogeneous matrix with small circular and isotropic inclusions, *i.e.* scatterers. The number of scatterers is unknown and the number of them encountered by the wave/particle during a time $t$ is modeled by a homogeneous Poisson counting process $N_t$ with parameter $\lambda t$. The parameter $\lambda$ is related to the mean free path $\eta$ of the random media by $\eta = c/\lambda$ where $c$ is the celerity in the medium.

### 2.1. Isotropic random walk on $\mathbb{S}^2$

Consider an initial vector $\boldsymbol{x}_0 = \boldsymbol{\mu} \in \mathbb{S}^2$, subject to $n$ consecutive random isotropic rotations. The generated random walk is the set of $n$ vectors $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n \in \mathbb{S}^2$ with independent steps $\boldsymbol{x}_{k-1} \to \boldsymbol{x}_k$ for all $k \geq 1$. The isotropic assumption of the random walk involves that the conditional *pdf*

$$f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = g_{k,k-1}(\boldsymbol{x}_{k-1}^T\boldsymbol{x}_k) \tag{1}$$

is *symmetrical*, *i.e.* only function of the cosine $\boldsymbol{x}_{k-1}^T\boldsymbol{x}_k$.

As explained in [6], symmetrical *pdf*s on $\mathbb{S}^2$ are elements of $L^1(SO(2)\backslash SO(3)/SO(2), \mathbb{R})$, with the additional constraint that $\int_{\mathbb{S}^2} f(\boldsymbol{x})d\boldsymbol{x} = 1$. The double coset $SO(2)\backslash SO(3)/SO(2)$ is the orbit of the group action of $SO(2) \times SO(2)$ on elements of $SO(3)$ with left and right action of $SO(2)$. In the sequel, we make use of the notation $\mathcal{H}$ for the double coset $SO(2)\backslash SO(3)/SO(2)$, *i.e.* $\mathcal{H} \cong SO(2)\backslash SO(3)/SO(2)$. Functions belonging to $L^1(\mathcal{H}, \mathbb{R})$ are known to be parametrized by the (co)lattitude with respect to the axis left invariant, and thus can be thought as functions $g(\boldsymbol{x}^T\boldsymbol{\mu}) = g(\cos\theta_{\boldsymbol{x}})$ where $\boldsymbol{\mu}$ is the axis left invariant by the action of $SO(2)$.

A *pdf* $f(\boldsymbol{x}; \boldsymbol{\mu})$ on $\mathbb{S}^2$, symmetrical with respect to $\boldsymbol{\mu}$, has a characteristic function given by

$$\widehat{f}_\ell = E\left[P_\ell(\cos\theta_{\boldsymbol{x}})\right] = E\left[P_\ell(\boldsymbol{\mu}^T\boldsymbol{x})\right] \tag{2}$$

$$= \int_{\mathbb{S}^2} f(\boldsymbol{x}; \boldsymbol{\mu}) P_\ell(\boldsymbol{\mu}^T\boldsymbol{x}) d\boldsymbol{x} \tag{3}$$

with $P_\ell$ the Legendre polynomial of order $\ell \geq 0$. The Legendre polynomial moments $|\widehat{f}_\ell| \leq 1$, for all $\ell \geq 0$, are the Fourier coefficients of the Fourier/Legendre expansion of $f$

$$f(\boldsymbol{x}; \boldsymbol{\mu}) = \sum_{\ell \geq 0} \frac{(2\ell+1)}{4\pi} \widehat{f}_\ell P_\ell(\cos\theta_{\boldsymbol{x}}). \tag{4}$$

The convolution product in $L^1(\mathcal{H},\mathbb{R})$ is inherited from the one in $L^1(SO(3),\mathbb{R})$ [6], and takes the form

$$(f \star_{\boldsymbol{\mu}} g)(\boldsymbol{x}) = \int_{\mathbb{S}^2} f(\boldsymbol{x}^T\boldsymbol{y}) g(\boldsymbol{y}^T\boldsymbol{\mu}) d\boldsymbol{y}, \tag{5}$$

where $\star_{\boldsymbol{\mu}}$ indicates that the (co)latitudes are measured with respect to $\boldsymbol{\mu}$. This convolution product fulfills the following properties:

- i) *Stability*:
  For $f, g \in L^1(\mathcal{H},\mathbb{R})$, then:
  $$(f \star_{\boldsymbol{\mu}} g)(\boldsymbol{x}) = (f \star_{\boldsymbol{\mu}} g)(\boldsymbol{\mu}^T\boldsymbol{x})$$
  which means that $f \star_{\boldsymbol{\mu}} g \in L^1(\mathcal{H},\mathbb{R})$.
- ii) *Fourier product*:
  Given $f, g \in L^1(\mathcal{H},\mathbb{R})$ and their Legendre polynomial moments $\widehat{f}_\ell$ and $\widehat{g}_\ell$, then their convolution product has the moments:
  $$\widehat{(f \star_{\boldsymbol{\mu}} g)}_\ell = \widehat{f}_\ell\, \widehat{g}_\ell$$

Thanks to the stability of the convolution product, the expression of the *pdf* of the $n$-step random walk takes a remarkable form.

**Theorem 1.** *Given an isotropic $n$-step random walk on $\mathbb{S}^2$, the* pdf *of $\boldsymbol{x}_n \in \mathbb{S}^2$ is the $n$-fold convolution in $L^1(\mathcal{H},\mathbb{R})$, where $SO(2)$ is the rotation subgroup such that $\boldsymbol{\mu}$ is left invariant. It reads*

$$f(\boldsymbol{x}_n;\boldsymbol{\mu}) = (g_{n,n-1} \star_{\boldsymbol{\mu}} \cdots \star_{\boldsymbol{\mu}} g_{1,0})\ (\boldsymbol{x}_n), \tag{6}$$

*where $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = g_{k,k-1}(\boldsymbol{x}_k^T\boldsymbol{x}_{k-1})$ can be identified as the conditional* pdf *of $\boldsymbol{x}_k$ given $\boldsymbol{x}_{k-1}$.*

*Proof.* The proof is conducted by induction. Since $\boldsymbol{x}_0 = \boldsymbol{\mu}$ is a deterministic vector, the *pdf* of $\boldsymbol{x}_1$ is $f(\boldsymbol{x}_1;\boldsymbol{\mu}) = g_{1,0}(\boldsymbol{\mu}^T\boldsymbol{x}_1)$ and belongs to $L^1(\mathcal{H},\mathbb{R})$. Thus the base case holds for $n = 1$. Assume now that the pdf of $\boldsymbol{x}_{k-1}$, for $k > 1$, is symmetrical with respect to $\boldsymbol{\mu}$, *i.e.* $f(\boldsymbol{x}_{k-1};\boldsymbol{\mu}) = g_{k-1}(\boldsymbol{\mu}^T\boldsymbol{x}_{k-1}) \in L^1(\mathcal{H},\mathbb{R})$, and is given by the following $(k-1)$-fold convolution: $f_{k-1}(\boldsymbol{x}_{k-1};\boldsymbol{\mu}) = (g_{k-1,k-2} \star_{\boldsymbol{\mu}} \cdots \star_{\boldsymbol{\mu}} g_{1,0})\ (\boldsymbol{x}_{k-1})$. Due to the isotropic assumption, the conditional *pdf* $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = g_{k,k-1}(\boldsymbol{x}_k^T\boldsymbol{x}_{k-1})$ also belongs to $L^1(\mathcal{H},\mathbb{R})$. Moreover, this conditional *pdf* allows us to express the density of $\boldsymbol{x}_k$ as

$$f(\boldsymbol{x}_k;\boldsymbol{\mu}) = \int_{\mathbb{S}^{p-1}} g_{k,k-1}(\boldsymbol{x}_k^T\boldsymbol{x}_{k-1}) g_{k-1}(\boldsymbol{\mu}^T\boldsymbol{x}_{k-1};) d\boldsymbol{x}_{k-1}.$$

According to (5), we recognize the following convolution on the double coset: $f(\boldsymbol{x}_k;\boldsymbol{\mu}) = (g_{k,k-1} \star_{\boldsymbol{\mu}} g_{k-1})(\boldsymbol{x}_k)$. Thus $f(\boldsymbol{x}_k;\boldsymbol{\mu})$ is also symmetrical about $\boldsymbol{\mu}$ according to property i). As $g_{k-1}$ is assumed to be a $k-1$-fold convolution, it comes finally by associativity that $f(\boldsymbol{x}_k;\boldsymbol{\mu}) = (g_{k,k-1} \star_{\boldsymbol{\mu}} \cdots \star_{\boldsymbol{\mu}} g_{1,0})\ (\boldsymbol{x}_k)$, and the inductive step holds. □

The Fourier expansion of the *pdf* of the isotropic random walk can be obtained thanks to convolution property ii).

**Corollary 2.** *For all $n \geq 1$, the Fourier expansion* pdf *of the $n$-step direction $\boldsymbol{x}_n \in \mathbb{S}^{p-1}$ reads*

$$f(\boldsymbol{x}_n;\boldsymbol{\mu}) = \sum_{\ell\geq 0} \frac{(2\ell+1)}{4\pi} \widehat{f}_\ell^{\otimes n} P_\ell(\boldsymbol{\mu}^T\boldsymbol{x}), \tag{7}$$

*where in the case where all the steps are identically distributed, the Fourier coefficients $\widehat{f}_\ell^{\otimes n}$ are given by*

$$\widehat{f}_\ell^{\otimes n} = (\widehat{g}_\ell)^n \tag{8}$$

*with $\widehat{g}_\ell = E[P_\ell(\boldsymbol{\mu}^T\boldsymbol{x}_1)]$ the Fourier coefficient of the distribution that governs a random walk step.*

*Proof.* It comes directly from the convolution expression (6) and the Fourier product property. □

### 2.2. Characteristic function of the multiple scattering process on $\mathbb{S}^2$

The multiple scattering process $\boldsymbol{x}_t \in \mathbb{S}^2$ describes the evolution of the direction of propagation during a propagatsion in a random medium. In the time lapse $[0,t]$, the number of scattering events is given by $\mathbb{P}[N_t = n] = e^{-\lambda t}(\lambda t)^n/n!$, and the density of $\boldsymbol{x}_t$ takes the form:

$$f(\boldsymbol{x}_t;\boldsymbol{\mu}) = e^{-\lambda t}\delta_{\boldsymbol{\mu}}(\boldsymbol{x}_t) + \sum_{n\geq 1} \frac{e^{-\lambda t}(\lambda t)^n}{n!} f^{\otimes n}(\boldsymbol{x}_t;\boldsymbol{\mu}) \tag{9}$$

where $\delta_{\boldsymbol{\mu}}(\boldsymbol{x}_t)$ denotes a mass located in the original direction $\boldsymbol{\mu} \equiv \boldsymbol{x}_0 \in \mathbb{S}^2$, and $f^{\otimes n}(\boldsymbol{x}_t;\boldsymbol{\mu})$ denotes the $n$-step random walk *pdf* with original direction $\boldsymbol{\mu}$. The first term (mass) in (9) is known as the Beer-Lambert law describing absorption in a medium.

The absolutely continuous part of $f(\boldsymbol{x}_t;\boldsymbol{\mu})$ is denoted $f^{\otimes>0} = c_0 h^{\otimes>0}(\boldsymbol{x}_t;\boldsymbol{\mu})$ where $c_0 = \left(1 - e^{-\lambda t}\right)^{-1}$ is the normalizing constant of the truncated distribution corresponding to the event $N_t > 0$ and $h^{\otimes>0}$ is the unnormalized density

$$h^{\otimes>0}(\boldsymbol{x}_t;\boldsymbol{\mu}) = \sum_{n\geq 1} \frac{e^{-\lambda t}(\lambda t)^n}{n!} f^{\otimes n}(\boldsymbol{x}_t;\boldsymbol{\mu}). \tag{10}$$

Based on its Fourier expansion

$$h^{\otimes>0}(\boldsymbol{x}_t;\boldsymbol{\mu}) = \sum_{\ell\geq 0} \frac{(2\ell+1)}{4\pi} \widehat{h^{\otimes>0}}_\ell P_\ell(\boldsymbol{\mu}^T\boldsymbol{x}_t), \tag{11}$$

it is finally possible to obtain the Legendre moments of the multiple scattering process.

**Proposition 1.** *The Legendre moments of the multiple scattering process express as*

$$\widehat{f}_\ell = E\left[P_\ell(\boldsymbol{\mu}^T\boldsymbol{x}_t)\right] = \exp\left(\lambda t\left(\widehat{g}_\ell - 1\right)\right), \tag{12}$$

*where the coefficients $\widehat{g}_\ell$ are the Fourier coefficients of the isotropic and identically distributed random steps.*

*Proof.* Based on the Fourier coefficients of $n$-fold *pdf*, orthogonality of Legendre polynomials, and the contribution of the mass in $\boldsymbol{\mu}$ which reduces to $e^{-\lambda t}$ since $P_\ell(\boldsymbol{\mu}^T\boldsymbol{\mu}) = P_\ell(1) = 1$, the result comes by straightforward calculation. □

Among the possible distributions for the isotropic random steps, we consider the von Mises-Fisher (vMF) on $\mathbb{S}^2$, denoted $M_3(\boldsymbol{\mu},\kappa)$, due to its importance in directional statistics [7]. This distribution is defined by the *pdf*

$$f(\boldsymbol{x};\boldsymbol{\mu},\kappa) = \frac{\kappa}{4\pi\sinh(\kappa)} e^{\kappa\boldsymbol{x}^T\boldsymbol{\mu}} \tag{13}$$

for $\boldsymbol{x} \in \mathbb{S}^2$. The distribution is rotationally symmetric with respect to its mean direction $\boldsymbol{\mu} \in \mathbb{S}^2$ and $\kappa \geq 0$ is the concentration parameter. The larger the value of $\kappa$, the more concentrated is the distribution about the mean direction $\boldsymbol{\mu}$. Conversely, when $\kappa = 0$ the distribution reduces to the uniform distribution on $\mathbb{S}^2$. It is also known [7] that the Fourier coefficients of the vMF distribution take the form

$$\widehat{g}_\ell(\kappa) = E\left[P_\ell(\boldsymbol{x}^T\boldsymbol{\mu})\right] = \frac{I_{\ell+1/2}(\kappa)}{I_{1/2}(\kappa)} \tag{14}$$

for $\kappa > 0$ and $\ell \geq 0$. A multiple scattering process with vMF distribution for the random steps will be called a vMF multiple scattering process.

## 3. APPROXIMATE BAYESIAN COMPUTATION

For a known initial direction $\boldsymbol{\mu} = \boldsymbol{x}_0$, the vMF multiple scattering process is governed by the parameter vector $\theta = (\kappa, \lambda)$, where $\kappa$ is the concentration parameter associated with each step, and $\lambda$ is the intensity parameter of the Poisson process that models the occurrence of scattering events. Note that without loss of generality, we can assume that the time lapse is $t = 1$, thus $\lambda \equiv \lambda t$ stands for the mean number of scattering events, and $\boldsymbol{x}$ denotes a realization of the vMF multiple scattering process.

This section addresses now the problem of the inference of the parameter vector $\theta$ from the observations of a dataset $D = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)})$ that consists of $n$ independent observations of the vMF multiple scattering process. This problem is challenging since there is no simple closed form expression for the multiple scattering distribution. Thus, the computation of the likelihood based on the Fourier expansion (11) raises a too dramatic computational issue.

Approximate Bayesian computation (ABC) is a popular likelihood-free method to perform Bayesian inference in situations where the likelihood function is either intractable, or too expensive to calculate. In the Bayesian paradigm, $\theta$ is assumed to be random and its posterior distribution of interest reads

$$p(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{p(D)}, \tag{15}$$

where $L(D|\theta)$ is the likelihood of the dataset $D$, $\pi(\theta)$ is the prior distribution and $p(D) = \int L(D|\theta)\pi(\theta)d\theta$ is a normalizing constant.

The standard ABC algorithm is a special case of a rejection method. The parameters are generated from the prior, and a dataset $D'$ is then simulated for these parameters. The parameter proposal is accepted if the simulated dataset $D'$ is almost identical to the observed sample $D$. This yields to Algorithm 1 where $\rho(D'^{(m)}, D)$

**Algorithm 1:** Rejection ABC

**for** $m = 1 \ldots N$ **do**
- Draw $\theta^{(m)} \sim \pi(\cdot)$
- Simulate $D'^{(m)}$ from the model with parameter $\theta^{(m)}$
- Accept $\theta^{(m)}$ if $\rho(D'^{(m)}, D) \leq \epsilon$

is a discrepancy function that measures the similarity between the simulated dataset $D^{(m)}$ and the observed one $D$, and $\epsilon > 0$ is a tolerance parameter. This algorithm gives independent draws from the posterior distribution $p(\theta|\rho(D', D) \leq \epsilon)$ which approximates the true posterior $p(\theta|D)$ [8]. Note that when $\epsilon \to 0$, this algorithm becomes exact. Therefore, this tolerance $\epsilon$ controls a trade-off between the acceptance rate of the rejection algorithm, i.e. the computational cost to obtain a given sample size of accepted proposals $\theta^{(m)}$, and the accuracy of the approximated posterior. In practice, the tolerance parameter $\epsilon$ is determined as a quantile on observed discrepancy values $\rho(D'^{(1)}, D), \ldots, \rho(D'^{(N)}, D)$. Basically $\epsilon$ is set to the $1\%$ or $0.1\%$ quantile when the number of proposals is large enough, e.g. $N = 10^5$ or $N = 10^6$.

### 3.1. Summary statistics

When the dataset $D = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)})$ is high dimensional, which occurs in our case when the sample size $n$ is large, the rejection algorithm suffer from the curse of dimensionality to compare efficiently the dataset with the simulated one [9]. In such cases, the comparison between $D'$ and $D$ is usually performed using summary statistics to project $D$ onto a lower dimensional space. In fact, using a representative enough summary statistic should still produce a good approximation to the posterior distribution for small values of the tolerance parameter $\epsilon$.

#### *3.1.1. Method of Moments*

We first recall some important results on parameter estimation based on method of moment methods in a frequentist framework. Consider a function $h(\cdot) : \mathbb{R}^p \to \mathbb{R}^L$ and the size $L$ summary statistics defined as

$$\boldsymbol{s}_n = \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{x}^{(i)}),$$

and denote as

$$f(\theta) \equiv E[\boldsymbol{s}_n] = E\left[h(\boldsymbol{x}^{(1)})\right],$$
$$C(\theta) \equiv n\,\mathrm{cov}\,(\boldsymbol{s}_n) = \mathrm{cov}\left(h(\boldsymbol{x}^{(1)})\right).$$

An important result in the framework of method of moments is that the following positive definite matrix

$$B(\theta) = \left(F(\theta)C(\theta)^{-1}F(\theta)^T\right)^{-1}, \tag{16}$$

where $F(\theta)$ is the Jacobian matrix of the function $f(\theta)$, provides a lower bound on the asymptotic mean squared errors of all estimators $\hat{\theta}$ constructed from the summary statistics $\boldsymbol{s}_n$

$$\lim_{n\to+\infty} nE\left[(\hat{\theta} - \theta)^2\right] \geq B(\theta),$$

where $A \geq B$ means that $A - B$ is positive semidefinite [10, 11]. Moreover, this asymptotic lower bound is attained for the generalized method of moments estimator defined by minimizing the following non-linear least squares criterion

$$\hat{\theta}_{\text{GMM}} = \arg\min_t \left(f(t) - \boldsymbol{s}_n\right)^T C(t)^{-1} \left(f(t) - \boldsymbol{s}_n\right). \tag{17}$$

#### *3.1.2. Choice of the summary statistics*

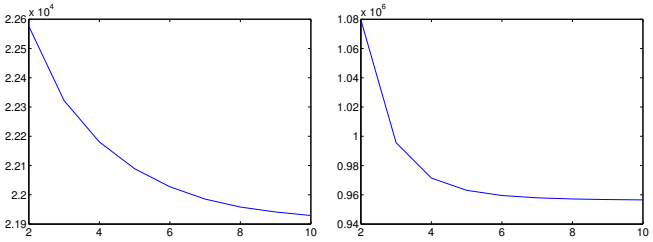
Based on the characteristic function of the distribution of $\boldsymbol{x}$, the $L$ first empirical Legendre polynomials seem to be a good choice to obtain a representative summary statistics. It leads to consider the following function

$$h(\boldsymbol{x}) = \left(P_1(\boldsymbol{\mu}^T\boldsymbol{x}), \ldots, P_L(\boldsymbol{\mu}^T\boldsymbol{x})\right),$$

while the summary statistic becomes $\boldsymbol{s}_n = \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{x}^{(i)})$. Note that the mean value $f(\theta)$ of $\boldsymbol{s}_n$ is just the vector of the $L$ first Fourier coefficients defined in (3), whose expression is obtained by combining (12) and (14) for the vMF multiple scattering process.

Moreover its covariance $C(\theta)$ can also be calculated analytically. In fact, the product of two Legendre polynomials with same argument expresses as a linear combination of Legendre polynomials (the interested reader is invited to see [12] for the analytical formula). As a consequence, the asymptotic lower bound $B(\theta)$ can be

computed for any value of $L$ and can be used to determine the trade-off between the efficiency and the dimensionality of the summary statistics. The Frobenius norm of the lower bound $B(\theta)$ is depicted in Fig. 1 for different values of the statistic size $L$, and for two values of the process parameter vector $\theta = (\kappa, \lambda)$. As expected, the asymptotic MSEs decrease when the number of Legendre polynomial moments increases. However, a good trade-off can be achieved when $L = 5$. Finally, by mimicking the non-linear least squares cri-




(a) $\kappa = 100, \lambda = 2$ (b) $\kappa = 100, \lambda = 10$

**Fig. 1**. Frobenius norm of the asymptotic lower bound $B(\theta)$ vs number $L$ of Legendre polynomial moments for different values of $\theta$

terion (17) which is asymptotically optimal, the discrepancy function used in the ABC algorithm is defined as

$$\rho(D', D) \equiv \rho(\boldsymbol{s}'_n, \boldsymbol{s}_n) = \left(\boldsymbol{s}'_n - \boldsymbol{s}_n\right)^T C(\theta)^{-1} \left(\boldsymbol{s}'_n - \boldsymbol{s}_n\right), \quad (18)$$

where $\theta$ is the proposed parameter, while $\boldsymbol{s}'_n$ is the summary statistic obtained for the simulated dataset with parameter $\theta$. This leads to a kind of Bayesian generalized method of moments. However this methods differs from previous works, see for instance [13, 14] and references therein, where the posterior is obtained based on some normal distribution approximation. Such normal approximations are not required here. Moreover, the main advantage with respect to method of moments is that the ABC algorithm allows us to estimate the (approximate) posterior distribution of the parameters. For instance, this yields directly to credible intervals on the parameters to be inferred.

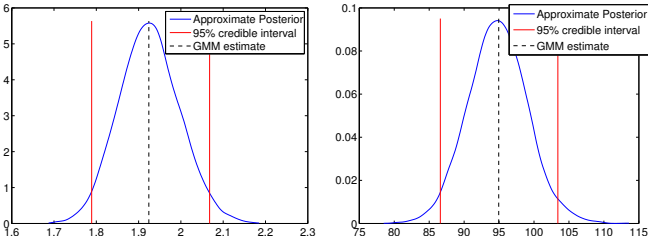
### 3.2. Adapted PMC algorithm

In practice, the standard rejection algorithm is very demanding computationally as it requires a large number of runs to sample accurately the parameter space. Several improved versions have been proposed in the literature to achieve considerable reduction of the computational burden. One possible improvement consists in performing some local correction to improve the accuracy of the sampled parameters [9]. Some methods embeds the simulation of the summary statistics in a Markov chain Monte Carlo algorithm [15]. A last family consists in adaptive algorithms to sample the posterior with sequential Monte-Carlo methods [16, 17, 8]. In this work, we adopt the population Monte-Carlo (PMC) one proposed in [8] and adapted in [18] to sequentially approximate the posterior. The outline of this algorithm is as follows. In the first step, the parameters are sampled form the prior as in the rejection Algorithm 1. In the other steps, the parameters are proposed based on the weighted parameters accepted in the previous step. The new tolerance level $\epsilon$ is updated to a quantile $\alpha$ of the discrepancy values accepted in the previous step. The algorithm stops when the acceptance ratio of the current step is below a given threshold $r_{\text{min}}$. In the simulations these parameters are set to $\alpha = 0.1$ and $r_{\text{min}} = 0.001$.

## 4. SIMULATIONS

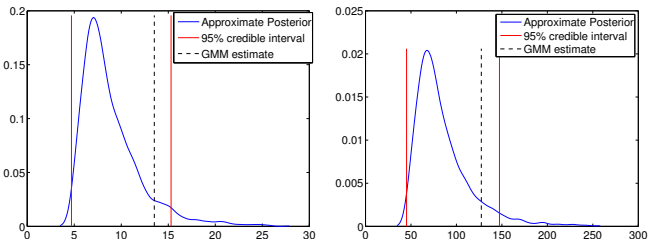
Several simulations have been conducted on synthetic dataset in order to evaluate the interest of the proposed method. In these simulations, the dataset are composed of $n = 1000$ samples $\boldsymbol{x}$, and the summary statistic is composed of the $L = 5$ first Legendre polynomial empirical moments as discussed in section 3.1.2. As no prior information is assumed to be known in the general case, a vague uniform prior is chosen $\pi(\theta) \propto \mathbb{1}_{I_\kappa \times I_\lambda}(\kappa, \lambda)$, where $I_\kappa = [10^1, 10^3]$ and $I_\lambda = [10^{-1}, 10^2]$.

Fig. 2 shows the marginal posterior density estimates based on $N_a = 2200$ samples provided by the adapted ABC-PMC algorithm. The dataset is generated for the parameter values $\kappa = 100$ and $\lambda = 2$. In this case, the approximate posterior distribution seems to be in agreement with the true value of the parameters, which belongs for instance to the 95% highest posterior density (HPD) credible interval. Moreover, one can see here that this credible interval is well centered around the frequentist GMM estimator introduced in (17). Fig. 3 shows the simulation for a larger value of $\lambda$ ($\lambda = 10$). This is a more difficult inference problem. In fact, when the mean number of scattering events increases, the distribution of the observation is less concentrated. In the limit case, the model is not identifiable. As expected, the marginal posteriors have a larger dispersion. Moreover one can see that these posteriors are quite asymmetric. In particular, the frequentist GMM estimator is not centered anymore into the HPD credible intervals. This emphasizes the interest of the proposed ABC based algorithm.




(a) $\lambda$ (true value $\lambda = 2$) (b) $\kappa$ (true value $\kappa = 100$)

**Fig. 2**. Density estimates (blue line) based on the ABC-PMC samples for the vMF multiple scattering process parameters $\theta = (\lambda, \kappa)$. True values: $\kappa = 100$, $\lambda = 2$. Red vertical lines: 95% HPD credible intervals. Black dotted vertical line: GMM estimate.




(a) $\lambda$ (true value $\lambda = 10$) (b) $\kappa$ (true value $\kappa = 100$)

**Fig. 3**. Density estimates (blue line) based on the ABC-PMC samples for the vMF multiple scattering process parameters $\theta = (\lambda, \kappa)$. True values: $\kappa = 100$, $\lambda = 10$. Red vertical lines: 95% HPD credible intervals. Black dotted vertical line: GMM estimate.

## 5. REFERENCES


[1] E. Jakeman and K.D. Ridley, *Modeling fluctuations in scattered waves*, Taylor and Francis, 2006.

[2] N. Le Bihan and L. Margerin, "Nonparametric estimation of the heterogeneity of a random medium using compound poisson process modeling of wave multiple scattering," *Physical Review E*, vol. 80, pp. 016601, 2009.

[3] S. Said, C. Lageman, N. Le Bihan, and J.H. Manton, "Decompounding on compact lie groups," *IEEE Transactions on Information theory*, vol. 56, no. 6, pp. 2766 – 2777, 2010.

[4] J. Boulanger, N. Le Bihan, and V. Rossetto, "Stochastic description of geometric phase for polarized waves in random media," *J. Phys. A: Math. Theor.*, vol. 46, no. 035203, 2013.

[5] F. Chatelain and N. Le Bihan, "Von mises-fisher approximation of multiple scattering process on the hypersphere," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada*, 2013.

[6] H. Dym and H.P. McKean, *Fourier series and integrals*, Academic Press, 1972.

[7] K.V. Mardia and P.E. Jupp, *Directional statistics*, John Wiley & Sons Ltd, 1999.

[8] Mark A. Beaumont, Jean-Marie Cornuet, Jean-Michel Marin, and Christian P. Robert, "Adaptive approximate bayesian computation," *Biometrika*, pp. 1–9, 2009.

[9] M.G.B. Blum and O. François, "Non-linear regression models for approximate bayesian computation," *Statistics and Computing*, vol. 20, no. 1, pp. 63–73, 2010.

[10] F. Chatelain, A. Ferrari, and J-Y Tourneret, "Parameter estimation for multivariate mixed poisson distributions," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings.*, May 2006, vol. 3.

[11] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *International Journal of adaptive control and signal processing*, vol. 3, pp. 191–229, 1989.

[12] J.C. Adams, "On the expression for the product of any two legendre's coefficient by means a series of legendre's coefficients," *Proc. Roy. Soc*, pp. 63–71, 1878.

[13] A. Zellner, J. Tobias, and H. K. Ryu, "Bayesian method of moments (bmom) analysis of parametric and semiparametric regression models," *South African Statistical Journal*, vol. 31, pp. 41–69, 1997.

[14] G. Yin, "Bayesian generalized method of moments," *Bayesian Analysis*, vol. 4, no. 2, pp. 191–207, 06 2009.

[15] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavar, "Markov chain monte carlo without likelihoods," *Proc. Natl. Acad. Sci.*, vol. 100, no. 26, pp. 15324–15328, 2003.

[16] S.A. Sisson, Y. Fan, and M. Tanaka, "Sequential monte carlo without likelihoods," *Proc. Natl. Acad. Sci.*, vol. 104, pp. 1760–1765, 2007.

[17] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, "Inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, vol. 6, no. 31, pp. 187–202, 2009.

[18] M. Lenormand, F. Jabot, and G. Deffuant, "Adaptive approximate bayesian computation for complex models," *Computational Statistics*, vol. 28, no. 6, pp. 2777–2796, 2013.