A REVERSIBLE JUMP MCMC ALGORITHM FOR PARTICLE SIZE INVERSION IN MULTIANGLE DYNAMIC LIGHT SCATTERING

A. Boualem and M. Jabloun and P. Ravier

Univ. Orléans, PRISME, 12 rue de Blois,
45067 Orléans, France

M. Naiim and A. Jalocha

CILAS, 8 Avenue Buffon,
45063 Orléans, France

ABSTRACT

The inverse problem of estimating the Particle Size Distribution (PSD) from Multiangle Dynamic Light Scattering measurements (MDLS) is considered using a Bayesian inference approach. We propose to model the multimodal PSD as a normal mixture with an unknown number of components (modes or peaks). In order to achieve the estimation of these variable dimension parameters, a Bayesian inference approach is used and solved by the Reversible Jump Markov Chain Monte Carlo sampler (RJMCMC). The efficiency and robustness of the method proposed are demonstrated using simulated and experimental data. Estimated PSDs are close to the original distributions for synthetic data. Moreover an improvement of the resolution is noticed compared to the Clementi method [1].

***Index Terms*—** Particle Size Distribution, Multiangle Dynamic Light Scattering, Inverse Problem, Bayesian Inference, MCMC, Reversible Jump.

1. INTRODUCTION

The determination of the size distribution of particles dispersed in a liquid in the range of sub-micrometers and nanometers is nowadays performed by using the dynamic light scattering (DLS) technique [2]. The increasing success of DLS comes from the fact that it is easy to handle, non-destructive, fast, and it requires no calibration process [3–5]. Several instruments are commercially available, among them the Nano DS from CILAS.

The DLS technique is based on the analysis of the temporal fluctuations of light scattered by the illuminated particles at a given angle. The particle size distribution (PSD) is retrieved by inverting the time autocorrelation function (ACF) of these fluctuations. This problem is a highly ill-posed inverse problem where small noise in data leads to large changes in the estimated PSD. Several methods have been proposed to retrieve the PSD from DLS data such as the cumulants method [6], CONTIN [7], truncated singular-value decomposition [8], non-negative least squares [9] and maximum entropy [10]. In general, satisfying results are achieved for monomodal PSDs. The main drawback of these methods is the poor capacity to discriminate peaks of multimodal PSDs. Indeed, the best results are obtained for populations with comparable intensity contributions and spaced at least by a factor 2 in diameter. Moreover, these methods are very sensitive to noise and suffer from lack of robustness. Multiangle dynamic light scattering (MDLS) allows getting more information about the studied sample by processing the whole DLS data acquired at different angles. Several works have demonstrated that MDLS provides more robust, reproducible and accurate PSD estimation than single-angle DLS, especially for polydisperse and/or multimodal samples [11–13].

In the last years, statistical techniques based on Bayesian inference became useful tools for solving DLS inverse problem [1, 14, 15]. In [1], the Bayesian method was applied to the Z-average (harmonic intensity averaged) diameters in order to retrieve the PSD, these Z-average diameters being first determined using the cumulants method. However, the errors on the Z-average diameters propagate for multimodal samples, which leads to large errors in the estimated PSD.

In the present paper, we propose to directly estimate the PSD from MDLS measurements. The main idea is to describe each component (mode) of the PSD by a mean, a standard deviation and a weight. Thus, the PSD is modelled as a Gaussian mixture. The number of the mixture components is assumed to be unknown, which allows flexibility in the proposed method. To estimate this unknown number of components and their parameters, we propose to use a Bayesian inference approach and we derive the *posterior* probability density function (PDF) of interest. Because of the complexity and high dimension of this *posterior* PDF, closed-form expressions of Bayesian estimators can not be derived. Furthermore, the dimension of the parameters vector depends on the unknown number of components. Therefore, a Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm [16] is used to generate samples distributed according to the *posterior* PDF of interest. This dimension matching strategy allows moves between parameter spaces with different dimensions which is clearly relevant for the proposed PSD model. We demonstrate the efficiency and robustness of the

This work has been supported in the frame of the Nano+ project by the FUI and the Région Centre and Orléans Innovation network.

proposed method by applying it to simulated noisy data for monomodal and bimodal PSDs and experimental data and we compare our results with Clementi's ones [1]. We note that the model parameters vector using the proposed approach is of small size compared to that of the Clementi method.

The paper is organized as follows. Sections 2 and 3 detail the DLS theory and the proposed Bayesian inversion method with the RJMCMC algorithm, respectively. In section 4, the estimation results from simulated and experimental data are presented and compared with those obtained with the Clementi method [1]. Finally, Section 5 concludes and discusses work in progress.

2. DLS THEORY

DLS measurements involve the analysis of the time ACF of scattered light performed by a digital correlator. The normalized intensity ACF $g_\theta^{(2)}(\tau)$ measured is related to the normalized electric field ACF $g_\theta^{(1)}(\tau)$ by the Siegert relation [3]

$$g_\theta^{(2)}(\tau) = 1 + \beta |g_\theta^{(1)}(\tau)|^2, \tag{1}$$

where τ is the time delay. $\beta(< 1)$ is an instrumental factor.

For a polydisperse sample, $g_\theta^{(1)}(\tau)$ is shown to be the summation of the normalized intensity-weighted PSD $h(D)$ with exponential decay functions [3], where D is the hydrodynamic diameter of the spherical particles

$$g_\theta^{(1)}(\tau) = \int_0^\infty h(D) \exp(-\frac{\Gamma_{0,\theta}}{D}\tau)\, dD. \tag{2}$$

The value $\Gamma_{0,\theta} = \frac{16\pi n^2 \sin^2(\theta/2) k_B T}{3\lambda_0^2 \eta}$ depends on the scattering angle θ and on other experimental conditions (Boltzmann constant k_B, absolute temperature T, wavelength of the incident light in vacuum λ_0, refractive index n and viscosity η of the medium).

Since the scattered intensity has an angular dependence, the intensity-weighted PSD $h(D)$ are angle-dependent too. Cancelling this angular dependence is achieved by considering the number-weighted PSD $f(D)$ which are angle-independent. The relation is given by [1, 17]

$$g_\theta^{(1)}(\tau) = \frac{1}{I_\theta} \int_0^\infty f(D) C_{I,\theta}(D) \exp(-\frac{\Gamma_{0,\theta}}{D}\tau) dD, \tag{3}$$

where $C_{I,\theta}(D)$ represents the fraction calculated through the Mie theory [17] of light intensity scattered at θ by a single spherical particle of diameter D. $I_\theta = \int_0^\infty f(D) C_{I,\theta}(D) dD$ is a proportionality constant ensuring $g_\theta^{(1)}(0) = 1$.

In the present paper, we assume for the sake of simplicity that the unknown size distribution can be described as a normal mixture distribution with

$$f(D) = \sum_{i=1}^{k} \frac{w_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(D-\mu_i)^2}{2\sigma_i^2}\right), \tag{4}$$

where w_i, μ_i and σ_i are respectively the weight, the mean and the standard deviation of the ith normal component. The number of components, k, is supposed unknown. For identifiability, we adopt the following labelling: $\mu_1 < \mu_2 < \cdots < \mu_k$. We denote $\boldsymbol{w}^{(k)} = [w_1, \cdots, w_k]^T$, $\boldsymbol{\mu}^{(k)} = [\mu_1, \cdots, \mu_k]^T$ and $\boldsymbol{\sigma}^{(k)} = [\sigma_1, \cdots, \sigma_k]^T$. ^{T} denotes the transpose operator.

The aim of the next section is to estimate the unknown number of components k and the parameters $\boldsymbol{w}^{(k)}$, $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$ using a Bayesian inference approach.

3. PROPOSED BAYESIAN INVERSION METHOD WITH RJMCMC

3.1. Observation model

In practice, the MDLS measurements are acquired at different scattering angles $\{\theta_r,\ r = 1, \ldots, R\}$, R is the total number of scattering angles. For each angle θ_r, $g_{\theta_r}^{(2)}(\tau)$ is measured at discrete time delays $\{\tau_j,\ j = 1, \ldots, M_r\}$, M_r being the total number of points at angle θ_r. Let us consider the following model of the MDLS observations

$$\begin{aligned} \tilde{g}_{\theta_1}^{(2)}(\tau_j) &= g_{\theta_1}^{(2)}(\tau_j) + w_1(j), \quad j = 1, \cdots, M_1 \ , \\ &\vdots \\ \tilde{g}_{\theta_r}^{(2)}(\tau_j) &= g_{\theta_r}^{(2)}(\tau_j) + w_r(j), \quad j = 1, \cdots, M_r \ , \\ &\vdots \\ \tilde{g}_{\theta_R}^{(2)}(\tau_j) &= g_{\theta_R}^{(2)}(\tau_j) + w_R(j), \quad j = 1, \cdots, M_R \ , \end{aligned} \tag{5}$$

where $\tilde{g}_{\theta_r}^{(2)}(\tau_j)$ is the measured time ACF at θ_r and $g_{\theta_r}^{(2)}(\tau_j)$ is the noise-free time ACF related to the PSD (or to the parameters k, $\boldsymbol{w}^{(k)}$, $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$) by (1), (3) and (4). The additive noise $w_r(j)$ are assumed to be independent (white) and normally distributed, with zero mean and variance $\sigma_{n,r}^2$ at the angle θ_r. We consider the vector $\tilde{\mathbf{g}}^{(2)} = [\tilde{\mathbf{g}}_1^{(2)^T}, \ldots, \tilde{\mathbf{g}}_R^{(2)^T}]^T$ where $\tilde{\mathbf{g}}_r^{(2)} = [\tilde{g}_{\theta_r}^{(2)}(\tau_1), \ldots, \tilde{g}_{\theta_r}^{(2)}(\tau_{M_r})]^T$ for $r = 1, \ldots, R$. We denote $\boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}} = [\sigma_{n,1}^2, \cdots, \sigma_{n,R}^2]^T$.

3.2. Bayesian Model

In the Bayesian inference approach proposed for MDLS inverse problem, all quantities incorporated in the mathematical model ((4) and (5)) are considered as random variables with probability densities. The joint *posterior* PDF of variables k, $\boldsymbol{w}^{(k)}$, $\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\sigma}^{(k)}$ and $\boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}}$ can be written as

$$p\left(k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}} | \tilde{\mathbf{g}}^{(2)}\right) = \frac{p\left(\tilde{\mathbf{g}}^{(2)} | k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}}\right) p(k) p(\boldsymbol{w}^{(k)}|k) p(\boldsymbol{\mu}^{(k)}|k) p(\boldsymbol{\sigma}^{(k)}|k) p(\boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}})}{p\left(\tilde{\mathbf{g}}^{(2)}\right)}, \tag{6}$$

where $p\left(\tilde{\mathbf{g}}^{(2)} | k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}}\right)$ is the likelihood function and $p(k)$ is the *prior* probability density on k. $p(\boldsymbol{w}^{(k)}|k)$, $p(\boldsymbol{\mu}^{(k)}|k)$ and $p(\boldsymbol{\sigma}^{(k)}|k)$ are the *prior* probability densities conditioned to k of the parameters $\boldsymbol{w}^{(k)}$, $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$, respectively. $p(\boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}})$ expresses *prior* information about the noisy measurements. $p\left(\tilde{\mathbf{g}}^{(2)}\right)$ is a normalizing constant.

By taking into account the noise independence assumption and assuming the independence between the measurements, the joint distribution (6) simplifies to give

$$p\left(k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}} | \tilde{\mathbf{g}}^{(2)}\right) \propto p(k) p(\boldsymbol{w}^{(k)}|k) p(\boldsymbol{\mu}^{(k)}|k) p(\boldsymbol{\sigma}^{(k)}|k) \times \prod_{r=1}^{R} \left[p(\sigma_{n,r}^2) p\left(\tilde{\mathbf{g}}_r^{(2)} | k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \sigma_{n,r}^2\right)\right]. \tag{7}$$

The white Gaussian assumption of noise (zero mean and variance $\sigma_{n,r}^2$) allows to write the likelihood function for each θ_r

$$p\left(\tilde{\mathbf{g}}_r^{(2)}|k,\boldsymbol{w}^{(k)},\boldsymbol{\mu}^{(k)},\boldsymbol{\sigma}^{(k)},\sigma_{n,r}^2\right)=\left(\sqrt{2\pi}\sigma_{n,r}\right)^{-M_r}\exp\left(-\frac{\chi_r^2}{2\sigma_{n,r}^2}\right), \tag{8}$$

with $\chi_r^2=\sum_{j=1}^{M_r}\left(\tilde{g}_{\theta_r}^{(2)}(\tau_j)-g_{\theta_r}^{(2)}(\tau_j)\right)^2$.

We now detail the chosen *prior* distributions. For k, we use a uniform *prior* discrete distribution between 1 and a specified maximum number of components k_{max}. The used *prior* distributions on $\boldsymbol{w}^{(k)}$, $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$ are such that all the parameters are drawn independently: $w_i\sim\mathcal{U}(0,1)$, $\mu_i\sim\mathcal{U}(0,\mu_{max})$ and $p(\sigma_i)=\frac{1}{\ln(\sigma_{max}/\sigma_{min})\sigma_i}$ with $(\sigma_{min}\leq\sigma_i\leq\sigma_{max})$ and 0 otherwise. Then, the resulting *prior* probability densities conditioned to k are

$$p(\boldsymbol{\mu}^{(k)}|k)=\begin{cases}\left[\frac{1}{\mu_{max}}\right]^k, & 0<\boldsymbol{\mu}^{(k)}\leq\mu_{max},\\ 0 & \text{otherwise},\end{cases} \tag{9}$$

$$p(\boldsymbol{\sigma}^{(k)}|k)=\begin{cases}\prod_{i=1}^{k}\frac{1}{\ln(\sigma_{max}/\sigma_{min})\sigma_i}, & \sigma_{min}\leq\boldsymbol{\sigma}^{(k)}\leq\sigma_{max},\\ 0 & \text{otherwise}.\end{cases} \tag{10}$$

The used *prior* for the noise variance is chosen to be the Jeffrey's *prior* [18], $p(\sigma_{n,r}^2)\propto\frac{1}{\sigma_{n,r}^2}$ for $r=1,\ldots,R$.

Finally, we have the following expression for the joint *posterior* PDF

$$p\left(k,\boldsymbol{w}^{(k)},\boldsymbol{\mu}^{(k)},\boldsymbol{\sigma}^{(k)},\boldsymbol{\sigma_n^2}|\tilde{\mathbf{g}}^{(2)}\right)\propto\frac{\prod_{r=1}^{R}\frac{1}{\sigma_{n,r}^{(M_r+2)}}\exp\left(-\frac{\chi_r^2}{2\sigma_{n,r}^2}\right)}{k_{max}\prod_{i=1}^{k}\left[\mu_{max}\ln\left(\frac{\sigma_{max}}{\sigma_{min}}\right)\sigma_i\right]}. \tag{11}$$

To cancel the nuisance parameters, we marginalize the joint *posterior* PDF with respect to $\boldsymbol{\sigma_n^2}$. As a result, we have (see Appendix A)

$$p\left(k,\boldsymbol{w}^{(k)},\boldsymbol{\mu}^{(k)},\boldsymbol{\sigma}^{(k)}|\tilde{\mathbf{g}}^{(2)}\right)\propto\frac{\prod_{r=1}^{R}\left[\chi_r^2\right]^{-\frac{M_r}{2}}}{k_{max}\prod_{i=1}^{k}\left[\mu_{max}\ln\left(\frac{\sigma_{max}}{\sigma_{min}}\right)\sigma_i\right]}. \tag{12}$$

Since this *posterior* PDF is highly multivariate and known up to a multiplicative constant, we should use a Markov chain Monte Carlo (MCMC) sampler. Since we deal with a varying dimension problem where classical Metropolis-Hastings MCMC methods are not valid, we use the approach termed reversible jump MCMC algorithm elaborated in [16]. In the next sub section, we detail the RJMCMC algorithm used to generate samples from the target *posterior* PDF (12).

3.3. Reversible Jump MCMC Algorithm

We propose to use a similar RJMCMC algorithm to the one introduced in [19] with little modifications. One sweep in the proposed algorithm consists of the following moves:

(a) Update the weights vector $\boldsymbol{w}^{(k)}$;

(b) Update the means $\boldsymbol{\mu}^{(k)}$ and the standard deviations $\boldsymbol{\sigma}^{(k)}$;

(c) Split one mixture component into two, or merge two into one.

In moves (a) and (b), there is no change of k. We then use classical MCMC samplers. For updating the weights, we use the multi-stage Gibbs sampler [18]. The weight of a component i is updated as follow. We generate a random $u\sim\mathcal{N}(0,\sigma_w^2)$ and new candidates are set as $w_i^*=w_i+u$ and $w_{i+1}^*=w_{i+1}-u$. The new weights w_i^* and w_{i+1}^* are accepted with the probability $\min\left(1,\prod_{r=1}^{R}\left[\chi_r^{2*}/\chi_r^2\right]^{-\frac{M_r}{2}}\right)$. This step is repeated for $i=1,...,k-1$. The parameters $\boldsymbol{\mu}^{(k)}$ and $\boldsymbol{\sigma}^{(k)}$ are updated in one step by using the Metropolis-Hastings algorithm [18]. The used instrumental densities $q(\boldsymbol{\mu}^{(k)*}|\boldsymbol{\mu}^{(k)})$ and $q(\boldsymbol{\sigma}^{(k)*}|\boldsymbol{\sigma}^{(k)})$ are chosen as Gaussian distributions with $\mathcal{N}(\boldsymbol{\mu}^{(k)},\sigma_\mu^2)$ and $\mathcal{N}(\boldsymbol{\sigma}^{(k)},\sigma_\sigma^2)$ respectively. $\boldsymbol{\mu}^{(k)*}$ and $\boldsymbol{\sigma}^{(k)*}$ are accepted with the probability $\min\left(1,\frac{p(\boldsymbol{\mu}^{(k)*}|k)p(\boldsymbol{\sigma}^{(k)*}|k)}{p(\boldsymbol{\mu}^{(k)}|k)p(\boldsymbol{\sigma}^{(k)}|k)}\prod_{r=1}^{R}\left[\chi_r^{2*}/\chi_r^2\right]^{-\frac{M_r}{2}}\right)$.

In the move (c), we make a random choice between a split or a merge move. In split move, a randomly chosen component i^* of the mixture is replaced by two new components labelled i_1 and i_2. This makes a jump from model with k components ($3k$ parameters) to model with $(k+1)$ components ($3(k+1)$ parameters). The other components are kept unchanged. The parameters of the new components are created using the same proposal proposed in [18]. To satisfy the dimension matching condition [16], we generate the auxiliary variables $u_1,u_2,u_3\sim\mathcal{U}(0,1)$, and take

$$\begin{aligned}w_{i_1}&=u_1w_{i^*}, & w_{i_2}&=(1-u_1)w_{i^*},\\ \mu_{i_1}&=u_2\mu_{i^*}, & \mu_{i_2}&=\frac{w_{i^*}-u_2w_{i_1}}{w_{i^*}-w_{i_1}}\mu_{i^*},\\ \sigma_{i_1}&=u_3\sigma_{i^*}, & \sigma_{i_2}&=\frac{w_{i^*}-u_3w_{i_1}}{w_{i^*}-w_{i_1}}\sigma_{i^*}.\end{aligned} \tag{13}$$

The reverse move (merge) consists in the combination of two components (i_1,i_2), randomly chosen and adjacent ($\mu_{i_1}<\mu_{i_2}$ with no other μ_j in $[\mu_{i_1},\mu_{i_2}]$), and replaced by a new component labelled i^*, reducing k by 1. The other components are kept unchanged. Since split and merge moves must form a reversible pair (a bijection) [16], then we have

$$\begin{aligned}w_{i^*}&=w_{i_1}+w_{i_2},\\ \mu_{i^*}&=\frac{w_{i_1}}{w_{i^*}}\mu_{i_1}+\frac{w_{i_2}}{w_{i^*}}\mu_{i_2},\\ \sigma_{i^*}&=\frac{w_{i_1}}{w_{i^*}}\sigma_{i_1}+\frac{w_{i_2}}{w_{i^*}}\sigma_{i_2}.\end{aligned} \tag{14}$$

The acceptance probabilities for split (from k to $k+1$) and merge (from $k+1$ to k) moves are $\min(1,A)$ and $\min\left(1,A^{-1}\right)$ respectively, where

$$A=\frac{\pi_{(k+1)k}\,p\left(k+1,\boldsymbol{w}^{(k+1)},\boldsymbol{\mu}^{(k+1)},\boldsymbol{\sigma}^{(k+1)}|\tilde{\mathbf{g}}^{(2)}\right)}{\pi_{k(k+1)}\,p\left(k,\boldsymbol{w}^{(k)},\boldsymbol{\mu}^{(k)},\boldsymbol{\sigma}^{(k)}|\tilde{\mathbf{g}}^{(2)}\right)}|J|. \tag{15}$$

$\pi_{k(k+1)}$ is the move probability from a model with k components to a model with $k+1$ components. Respect to $\pi_{k(k+1)}+\pi_{k(k-1)}=1$, we choose $\pi_{1(2)}=1$, $\pi_{k_{max}(k_{max}-1)}=1$ and otherwise $\pi_{k(k+1)}=0.5$. $|J|$ is the absolute value of the determinant of split transform Jacobian given by

$$|J|=\left|\frac{\partial(\boldsymbol{w}^{(k+1)},\boldsymbol{\mu}^{(k+1)},\boldsymbol{\sigma}^{(k+1)})}{\partial(\boldsymbol{w}^{(k)},\boldsymbol{\mu}^{(k)},\boldsymbol{\sigma}^{(k)},u_1,u_2,u_3)}\right|=\frac{w_{i^*}^3\mu_{i^*}\sigma_{i^*}}{(1-u_1)^2}. \tag{16}$$

In order to derive estimates of the PSD mixture parameters from the generated samples, we propose to use the maximum a *posteriori* estimator for k, $\hat{k} = max(p(k|\tilde{\mathbf{g}}^{(2)}))$ and the *posterior* expectation conditioned to $k = \hat{k}$ for the other parameters, $\hat{\boldsymbol{w}} = E[p(\boldsymbol{w}|k = \hat{k}, \tilde{\mathbf{g}}^{(2)})]$, $\hat{\boldsymbol{\mu}} = E[p(\boldsymbol{\mu}|k = \hat{k}, \tilde{\mathbf{g}}^{(2)})]$ and $\hat{\boldsymbol{\sigma}} = E[p(\boldsymbol{\sigma}|k = \hat{k}, \tilde{\mathbf{g}}^{(2)})]$.

4. RESULTS AND DISCUSSIONS

In this section, we examine the proposed RJMCMC algorithm for estimating monomodal and bimodal PSDs through simulated and experimental MDLS data. For all simulated and real data, we used latexes with refractive index 1.59, dispersed in pure water (refractive index 1.33 and viscosity $\eta = 0.89$ mPa.s). We used a vertically-polarized laser of wavelength $\lambda_0 = 638$ nm. The temperature was stabilised at 298.15 K. The algorithm was run with the following settings, $k_{max} = 5$, $\mu_{max} = 2000$ nm, $\sigma_{min} = 0.2$ nm, $\sigma_{max} = 200$ nm, $\sigma_w = 0.05$, $\sigma_\mu = 10$ and $\sigma_\sigma = 1$. The parameters estimates are extracted from 400000 sweeps after a burn-in period of 100000 sweeps. Estimation results are compared with those obtained with the Clementi method [1].

4.1. Simulated MDLS data

MDLS data were simulated as follow. First, the noise-free time ACFs $g^{(2)}_{\theta_r}(\tau_j)$ were simulated from the corresponding PSD using (1), (3) and (4). Then, the noisy time ACFs $\tilde{g}^{(2)}_{\theta_r}(\tau_j)$ were simulated by adding a white Gaussian noise (5) with $\sigma_{n,r} = 0.001$ for all the scattering angles.

The RJMCMC algorithm is first tested with a monomodal Gaussian PSD with $\mu = 200$ nm and $\sigma = 20$ nm. The scattering angles are 30°, 60°, 90°, 120° and 150°. Performing Monte Carlo simulations with 50 independent runs of noise gave the following results. For 47 runs (94%), RJMCMC algorithm has successfully estimated the correct number of modes ($\hat{k} = 1$). For those runs, the obtained statistics of the (mean; standard deviation) estimates are (199.9 ± 0.3 nm; 20.1 ± 0.4 nm) with RJMCMC algorithm vs (196.1 ± 3.4 nm; 26.8 ± 4.4 nm) with Clementi method. These results show higher accuracy for estimating monomodal Gaussian PSD with the RJMCMC algorithm than with the Clementi method.

To validate these results for multimodal PSDs, RJMCMC algorithm is also tested with a bimodal PSD Gaussian mixture with the values $w_1 = 0.6$, $\mu_1 = 800$ nm and $\sigma_1 = 10$ nm for the first mode and $w_2 = 0.4$, $\mu_2 = 1000$ nm and $\sigma_2 = 25$ nm for the second mode. The scattering angles are 30°, 60°, 80°, 90°, 100°, 120° and 150°. The estimation results of one noise run are shown in Figure 1. The *posterior* distribution of k depicted in Figure 1(a) gives $\hat{k} = 2$ which is the true number of modes. The paramters estimates from the generated samples (Figures 1(b), 1(c) and 1(d)) are $\hat{w}_1 = 0.6$, $\hat{\mu}_1 = 799.5$ nm, $\hat{\sigma}_1 = 8.7$ nm, $\hat{w}_2 = 0.4$, $\hat{\mu}_2 = 1001.4$ nm and $\hat{\sigma}_2 = 20.6$ nm. The reconstructed PSD using (4) from

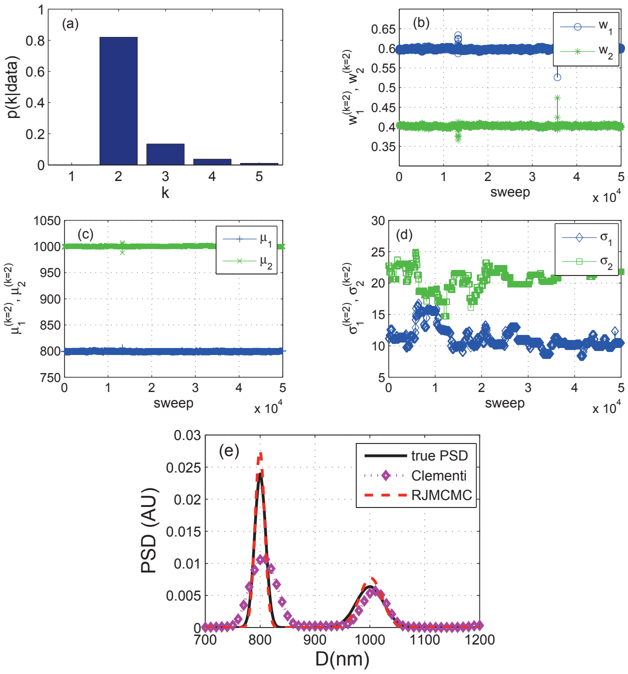

Fig. 1. Estimation results using RJMCMC algorithm from simulated MDLS data of the bimodal PSD. (a) Estimated *posterior* distribution of k. (b), (c) and (d) last 50000 sweeps, conditioning on $k = 2$ of the weights, means and standard deviations respectively. (e) Comparison between the true and estimated PSD with the RJMCMC algorithm and the Clementi method [1].

these estimated parameters is shown in Figure 1(e) and compared with the one estimated with the Clementi method. RJMCMC algorithm gave a better PSD estimate than the Clementi method.

Performing Monte Carlo simulations with 50 independent runs of noise gave the following results. For 46 runs (92%), RJMCMC algorithm has successfully estimated the correct number of modes ($\hat{k} = 2$). For those runs, statistics of the mixture parameters estimates are reported on Table 1. The results show a higher accuracy for the parameters estimated with the RJMCMC algorithm than with the Clementi method. For the standard deviation parameters that are usually difficult to estimate, RJMCMC algorithm gave good estimates, but with less precision compared to the other parameters.

	Mode 1			Mode 2		
	w_1 (%)	μ_1 (nm)	σ_1 (nm)	w_2 (%)	μ_2 (nm)	σ_2 (nm)
expected	60	800	10	40	1000	25
RJMCMC	60 ± 0.5	800.0 ± 1.1	8.2 ± 3.6	40 ± 0.5	1000 ± 2.6	23.2 ± 4.0
Clementi	65 ± 0.7	807.2 ± 0.8	25.2 ± 0.8	35 ± 0.7	1008.8 ± 1.9	37.8 ± 2.0

Table 1. Statistics (mean±std) of estimation results of bimodal PSD Gaussian mixture, comparison between RJMCMC algorithm and Clementi method (Statistics obtained using 46 independent Monte Carlo noise runs).

4.2. Experimental MDLS data

The experimental MDLS data were acquired using the Nano DS equipment from CILAS. The studied sample was a bimodal mixture of polystyrene latex spheres with nominal diameters (standard deviation) of 800 nm (5 nm) and 1000 nm

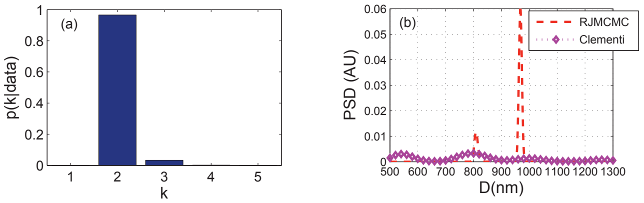

Fig. 2. Estimation results using RJMCMC algorithm for real MDLS data of bimodal PSD (800 nm and 1000 nm). (a) Estimated *posterior* distribution of k. (b) Comparison between the estimated PSDs with the RJMCMC algorithm and the Clementi method [1].

(10 nm). The ACFs were measured at the scattering angles 65°, 70°, 80°, 90°, 95°, 100° and 115°. The estimation results are shown in Figure 2. The estimated number of modes from the *posterior* distribution of k depicted in Figure 2(a) $\hat{k} = 2$ is in good agreement with the announced number of modes. The parameters estimates from the generated samples are $\hat{w}_1 = 0.19$, $\hat{\mu}_1 = 809$ nm, $\hat{\sigma}_1 = 6.4$ nm, $\hat{w}_2 = 0.81$, $\hat{\mu}_2 = 968$ nm and $\hat{\sigma}_2 = 5.4$ nm. Compared to the PSD estimated with the Clementi method, the reconstructed PSD (Figure 2(b)) from these estimated parameters seems to be the closest to the expected one.

5. CONCLUSIONS

A new approach for estimating multimodal PSDs from MDLS measurements is proposed. The PSD is modelled as a Gaussian mixture with an unknown number of modes. The estimation of the model parameters is achieved using a Bayesian inference approach solved by a RJMCMC algorithm without any prior knowledge about the number of modes. Analysis of simulated and experimental data has shown higher accuracy and robustness for estimating multimodal PSDs compared to the Clementi method [1]. In future works, limitations of the proposed method for the estimation of asymmetric PSDs and its resolution in terms of PSD peaks ratio will be determined.

A. APPENDIX

The marginalization step is given by

$$p\left(k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)} | \tilde{\mathbf{g}}^{(2)}\right) = \int_0^\infty \cdots \int_0^\infty p\left(k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)}, \boldsymbol{\sigma}_{\boldsymbol{n}}^{\boldsymbol{2}} | \tilde{\mathbf{g}}^{(2)}\right) d\sigma^2_{n,1} \cdots d\sigma^2_{n,R}. \tag{17}$$

Therefore, we can write

$$p\left(k, \boldsymbol{w}^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\sigma}^{(k)} | \tilde{\mathbf{g}}^{(2)}\right) \propto \frac{\prod_{r=1}^{R} \int_0^\infty \frac{1}{\sigma_{n,r}^{(M_r+2)}} \exp\left(-\frac{\chi_r^2}{2\sigma^2_{n,r}}\right) d\sigma^2_{n,r}}{k_{max} \prod_{i=1}^{k} \left[\mu_{max} \ln\left(\frac{\sigma_{max}}{\sigma_{min}}\right) \sigma_i\right]}. \tag{18}$$

By making the variable change $x = 1/\sigma^2_{n,r}$, we have

$$\int_0^\infty \frac{1}{\sigma_{n,r}^{M_r+2}} \exp\left(\frac{-\chi_r^2}{2\sigma^2_{n,r}}\right) d\sigma^2_{n,r} = \int_0^\infty x^{\frac{M_r}{2}-1} \exp\left(\frac{-\chi_r^2 x}{2}\right) dx = \Gamma\left(\frac{M_r}{2}\right) \left[\frac{\chi_r^2}{2}\right]^{-\frac{M_r}{2}}. \tag{19}$$

By replacing (19) in (18), we get the expression of (12).

REFERENCES

[1] L.A. Clementi, J.R. Vega, L.M. Gugliotta, and H.R.B. Orlande, "A bayesian inversion method for estimating the particle size distribution of latexes from multiangle dynamic light scattering measurements," *Chemometrics and Intelligent Laboratory Systems*, vol. 107, pp. 165–173, 2011.

[2] International Standard ISO22412, "Particle size analysis - dynamic light scattering (dls)," 2008.

[3] R. Xu, *Particle Characterization: Light Scattering Methods*, Particle Technology Series. Springer, 2000.

[4] B.J. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, Dover Books on Physics Series. Dover Publications, 2000.

[5] W. Brown, *Dynamic light scattering: the method and some applications*, Monographs on the physics and chemistry of materials. Clarendon Press, 1993.

[6] D.E. Koppel, "Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants," *The Journal of Chemichal Physics*, vol. 57, no. 11, pp. 4814–4820, 1972.

[7] S.W. Provencher, "A constrained regularization method for inverting data represented by linear algebraic or integral equations," *Computer Physcis Communications*, vol. 27, pp. 213–227, 1982.

[8] Y. Wang, J. Shen, G. Zheng, Z. Dou, and Z. Li, "Analysis of non-negative tikhonov and truncated singular value decomposition regularization inversion in pcs," *Theoretical and Applied Information Technology*, vol. 48, pp. 303–310, 2013.

[9] I.D. Morrison, E.F. Grabowski, and C.A. Herb, "Improved techniques for particle size determination by quasi-elastic light scattering," *Langmuir*, vol. 1, no. 4, pp. 496–501, 1985.

[10] S.-L. Nyeo and B. Chu, "Maximum-entropy analysis of photon correlation spectroscopy data," *Macromolecules*, vol. 22, no. 10, pp. 3998–4009, 1989.

[11] G. Bryant and J.C. Thomas, "Improved particle size distribution measurements using multiangle dynamic light scattering," *Langmuir*, vol. 11, no. 7, pp. 2480–2485, 1995.

[12] G. Bryant, C. Abeynayake, and J.C. Thomas, "Improved particle size distribution measurements using multiangle dynamic light scattering. 2. refinements and applications," *Langmuir*, vol. 12, no. 26, pp. 6224–6228, 1996.

[13] C. De Vos, L. Deriemaeker, and R. Finsy, "Quantitative assessment of the conditioning of the inversion of quasi-elastic and static light scattering data for particle size distributions," *Langmuir*, vol. 12, no. 11, pp. 2630–2636, 1996.

[14] M. Iqbal, "On photon correlation measurements of colloidal size distributions using bayesian strategies," *Journal of Computational and Applied Mathematics*, vol. 126, pp. 77–89, 2000.

[15] S.-L. Nyeo and R. R. Ansari, "Sparse bayesian learning for the laplace transform inversion in dynamic light scattering," *Journal of Computational and Applied Mathematics*, vol. 235, no. 8, pp. 2861–2872, 2011.

[16] P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[17] C.F. Bohren and D.R. Huffman, *Absorption and scattering of light by small particles*, Wiley science paperback series. Wiley, 1983.

[18] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer, 2004.

[19] S. Richardson and P. J. Green, "On bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 4, pp. 731–792, 1997.