Milk homogenization monitoring: fat globule size estimation from scattering spectra of milk

Annelies Postelmans[a], Ben Aernouts[b], Jeroen Jordens[c], Tom Van Gerven[c], Wouter Saeys[a*]

[a] KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

[b] KU Leuven, Department of Biostystems, Biosystems Technology Cluster, Campus Geel, Kleinhoefstraat 4, 2440 Geel, Belgium

[c] KU Leuven, Department of Chemical Engineering, Celestijnenlaan 200F, 3001 Leuven, Belgium

Abstract

The fat globule size distribution of raw milk and milk with an increasing degree of homogenization was estimated based on their bulk light scattering properties in the Vis/NIR wavelength range. The particle size distribution (PSD) was approximated as a lognormal distribution, of which the parameters were estimated simultaneously with the fat concentration. This resulted in a good agreement between the estimated PSDs and the reference PSDs obtained by laser diffraction in case of raw and strongly homogenized samples. The accuracy increased if a known fat concentration was incorporated, or when the scattering coefficient and anisotropy factor spectra were used simultaneously as input. For mildly homogenized samples, the lognormal distribution was unable to fit the bimodal PSD correctly and focused on the largest fat globules. In this case, the estimated PSDs provided still relatively accurate information on D90, D32 and the right distribution tail, which contains the largest fat globules.

Industrial relevance: The presented estimation method demonstrates the potential of bulk scattering spectra for determining the PSD and concentration of scattering particles in turbid media. Further development of this technology can lead to new solutions for spectroscopic PSD determination, allowing on-line monitoring systems for a wide range of food and non-food products.

Keywords: Particle size distribution, milk homogenization, light scattering properties

1 Introduction

The composition of milk is an important characteristic with regard to food quality and further processing into derived products. The concentrations of the main components besides water, namely fat, protein and lactose, are often optically determined based on their light absorption characteristics in the infrared wavelength region (Fox & McSweeney, 1998; Walstra, Jenness, & Badings, 1984; Lynch, Barbano, Schweisthal, & Fleming, 2006). The infrared light is preferred because of the clear absorption bands and the lower influence of light scattering on the acquired spectra. In general, milk contains two types of scattering particles suspended in the milk serum: milk fat globules and casein micelles. The average fat content in bovine milk is about 4.0% w/w (range 2.5-5.5%), while casein is on average present at about 2.6% w/w (range 1.7-3.5%) (Walstra, Wouters, & Geurts, 2005). However, not only the concentrations play a role, but also the globule size determines the physical properties of milk. Fat globules in raw milk are reported to range from 0.1 µm to 15-20 µm diameter (Fox & McSweeney, 1998). Their size shows biological variation

*Corresponding author at: KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001 Leuven, Belgium. E-mail address: wouter.saeys@kuleuven.be (W. Saeys)

Abbreviations

BOP	bulk optical properties	d_{mean}	mean diameter	UT	unscattered transmission
D...	...% percentile	LB	lower boundary	VF	volume fraction
D32	Sauter mean diameter	*par*1 & *par*2	generalized distribution parameters	μ_s	scattering coefficient
D43	DeBrouckere mean diameter	PSD	particle size distribution	μ_s'	reduced scattering coefficient
DIS	double integrating sphere	UB	upper boundary	g	anisotropy factor

with breed, age, health status of the animal etc. The fat globule size can be reduced by means of a homogenization process (Walstra, Geurts, Noomen, Jellema, & van Boekel, 1999), which stabilizes the milk against creaming and partial coalescence, and changes the viscosity (Walstra et al., 1999). Light scattering is wavelength dependent and influenced by the particle size. The smaller casein micelles, with a size ranging from 20 nm up to 400 nm (Walstra et al., 2005), particularly scatter light in the ultraviolet and visible wavelength range, while the larger fat globules have a scattering effect up to the infrared range.

To monitor the fat globule size before and after homogenization, measurements based on light spectroscopy, light scattering or electronic counting can be used (Fox & McSweeney, 1998; Bylund, 2003). These are usually bench-top instruments that report a full particle size distribution (PSD). However, they only provide measurements at regular time intervals, at which a sample has to be transferred from the production line to the lab for preparation and analysis. Most methods also require severe dilution of the small sample volumes, which can alter the sample as aggregates might break up.

Several researchers have investigated the potential of Vis/NIR spectroscopy for particle sizing purposes. Bogomolov, Melenteva, and Dahm (2013) reported diffuse transmission spectra of increasingly homogenized milk in the 400-1000 nm wavelength range and used the representative layer theory to attribute the spectral changes to a decreasing fat globule size. The next step of extracting PSD information from spectra by inverse estimation was for example done by Di Marzo, Cree, and Barbano (2016). They estimated PSD parameters such as the mean and 90% quantile based on infrared spectra using partial least squares models. However, they were not able to predict a complete PSD and such data-based models can only be used on samples very similar to the ones used for training. On the other hand, Cabassi, Profaizer, Marinoni, Rizzi, and Cattaneo (2013) made PSD estimations based on NIR transmission spectra of raw milk by assuming a Weibull distribution as PSD shape. However, their validation was limited to the estimated Sauter mean diameters (D32).

Instead of these (semi-)empirical approaches, the underlying physics of light scattering can be used to access the particle size of turbid media like milk. Mie theory provides a direct relation between the size and scattering by spherical particles, such as fat globules. Aernouts et al. (2015b) found that the bulk optical properties (BOP) of milk are strongly linked to the size distribution of the milk fat globules. These BOP include the bulk scattering coefficient μ_s, which indicates the probability of photon scattering per infinitesimal path length ($[\mu m^{-1}]$), and the anisotropy factor g, a measure for the direction of scattering ($0 =$ isotropic, $1 =$ completely forward). These two bulk scattering properties can be combined into the reduced scattering coefficient μ_s', according to the expression $\mu_s' = \mu_s \times (1-g)$. Furthermore, we have shown that the PSDs of monomodal polystyrene suspensions can be estimated from the Vis/NIR BOP extracted from double integrating spheres measurements (Postelmans, Aernouts, & Saeys, 2019). This requires a robust inversion of the Mie scattering theory. Recently, Stocker et al. (2017) succeeded in estimating the milk fat globules PSD in raw and homogenized milk from the Vis/NIR scattering and reduced scattering coefficient spectra using an inversion of Mie theory. Nevertheless, they did not explore the potential of the Vis/NIR scattering anisotropy factor, as well as a combination of the different scattering properties to further improve the PSD estimation.

Therefore, the objective of this study is to evaluate the potential of the scattering anisotropy factor and the combination of scattering properties for estimating the fat globules size distribution in raw milk and an extensive set of milk samples with different degrees of homogenization. To improve the robustness, the estimation routine includes a procedure to obtain relevant starting points and to prevent local minima and non-converged solutions from being accepted as final PSD estimate. On top of that, it is investigated if including information on the concentration of the scatterers (the fat content), for example based on spectroscopic absorption measurements or μ_a (Aernouts, Polshin, Lammertyn, & Saeys, 2011; Aernouts, Polshin, Saeys, & Lammertyn, 2011; Aernouts et al., 2015a) can improve the accuracy of the PSD estimates. This would combine the benefits of extracting composition information from the absorption properties and particle size information from the light scattering properties.

2 Materials and methods

2.1 Milk samples and reference PSDs

A raw bulk milk sample was collected from a Belgian dairy research farm (Hooibeekhoeve, Geel). The cooled tank (at 4°C) contained the milk of 70 Holstein-Friesian dairy cows produced over two days and was regularly stirred. The fat and protein content of the milk (respectively 42.8 g/l and 33.7 g/l) was determined by the Milk Control Center Flanders with a MilkoScan FT+ (Foss, Hillerød, Denmark) according to ISO 9622:2013 (ISO, 2013). After warming up the sample to 37 °C and gently stirring, seven subsamples of 20 ml were taken. The subsamples were subjected to ultrasonic homogenization for respectively 0 s (raw milk), 30 s, 60 s, 120 s, 240 s, 480 s and 960 s. The sonication was performed using a Vibra-Cell VCX400 (400 W, 20 kHz) in combination with a CV26 converter, a 13 mm horn and a 3 mm microtip (Sonics & Materials Inc., Danbury, USA). The amplitude of the sonicator was set at 20% (80 W). The microtip was immersed about 1 cm into the milk sample, which was contained in a 50 ml conical polypropylene Falcon tube. To prevent overheating of the milk during homogenization, the samples were placed in a water bath of 24 °C for homogenization times up to 240 s, or 19 °C for longer homogenization times. Three-fold diluted samples (10 ml milk with 20 ml deionized water) were used in the optical measurements to ensure the independent scattering assumption was valid (Aernouts et al., 2015b).

The PSDs of the fat globules and casein micelles were determined using laser diffraction with a Mastersizer 3000 instrument (Malvern, UK). The milk was added drop-wise to a beaker of deionized water until a red laser obscuration of 5-9% was reached, with the rotor of the Hydro EV dispersion unit stirring at 2400 rpm. The resulting PSDs were the average of five consecutive measurements of 5 s without delay. The particle refractive index was set at $1.46 + i5 \times 10^{-5}$, as indicated in the Mastersizer software for milk fat. Particles were assumed to be spherical (Mie theory) and the general purpose analysis type was applied. The PSD of casein was identified as the first mode of the bimodal PSD of raw milk, with a clear separation from the second mode (milk fat). Subsequently, the PSD of casein was subtracted from all PSDs, and the resulting fat PSDs were converted to probability density functions.

2.2 Experimental BOP determination

To determine the BOP of the milk samples experimentally, the double integrating sphere (DIS) set-up combined with an unscattered transmission (UT) measurement path described by Aernouts et al. (2013) was used. A supercontinuum laser coupled to a monochromator illuminated the samples and allowed a sequential scan over the desired wavelength range. For more details on this set-up, the reader is referred to Aernouts et al. (2013), while more specific information on the milk measurements can be found in (Aernouts et al., 2015a, 2015b).

The milk samples were loaded into borosilicate cuvettes with a sample thickness of 0.55 mm and placed between the two integrating spheres to measure the total reflectance and total transmittance. To measure the unscattered transmittance, a 0.155 mm thick cuvette was placed in the UT path. Replicates were obtained by reloading the cuvettes five times from the same subsample. Measurements were performed in the wavelength range from 550 nm to 1350 nm in steps of 10 nm.

The acquired reflectance and transmittance spectra were passed on to the inverse adding doubling routine to calculate the BOP, as implemented by Prahl (2011). In addition to these spectra, the sample thickness and the sample refractive index were given as inputs. The refractive index of milk was calculated based on the weight fractions of the different milk components (Walstra et al., 1984) and the wavelength-dependent refractive index of water (Segelstein, 1981; Aernouts et al., 2014). Since the fat globules in raw milk are larger than 25% of the wavelength, they do not contribute to the refractive index (Walstra et al., 1984). For the homogenized samples, comparison of inverse adding-doubling results calculated based on a sample refractive index with and without a contribution of fat showed that the effect of including the fat concentration is negligible. Therefore, the milk refractive index used for all samples was calculated without contribution of milk fat.

Scattering spectra (μ_s, μ_s' and g) were smoothed by a third order Savitsky-Golay filter with a window width of 60 nm. Since the experimental BOP contained scattering contributions of both milk fat and casein micelles, the casein scattering was removed by correcting the experimental spectra according to Eq. (1), based on simulated casein spectra (see section 2.3). Eq. (1)b was obtained by substitution using the definition of g ($g = 1 - \mu_s'/\mu_s$) and the fact that Eq. (1)a is also valid for μ_s'.

$$\mu_{s,meas,fat} = \mu_{s,meas} - VF_{casein} \times \mu_{s,sim,casein} \tag{1a}$$

$$g_{meas,fat} = \left[g_{meas} \times \mu_{s,meas} - VF_{casein} \times \mu_{s,sim,casein} \times g_{sim,casein}\right] / \mu_{s,meas,fat} \tag{1b}$$

2.3 Forward simulation of BOP

The BOP (μ_s, μ_s', g) of all measured milk fat PSDs and the casein PSD were simulated using the iterative tool for polydisperse systems developed by Aernouts et al. (2014). All spectra were calculated in the 0.55-1.35 µm wavelength range with a step of 0.01 µm. A particle volume fraction (VF) of 1% was adopted, as this is in agreement with the requirements for independent scattering. In this regime, the anisotropy factor is independent of the VF, while the scattering coefficient and reduced scattering coefficient are linearly proportional to the VF. Therefore, multiplying the simulated μ_s and μ_s' spectra with the desired concentration performs a correct scaling.

The scattering by casein micelles was simulated based on the measured reference PSD of casein. The casein content was fixed at 75% w/w of the measured crude protein fraction, in accordance with the average casein fraction reported for crude protein in milk (Walstra et al., 1999). The obtained value was divided by three to match the three-fold dilution of the samples in the optical measurements. The bulk scattering spectra of the raw and homogenized milk samples were simulated using the reference PSD of the fat globules. The real part of the milk serum refractive index was obtained by adding a baseline to the refractive index of water (Segelstein, 1981), following the formula given by Walstra et al. (1999) and using the three-fold diluted average concentration of the milk components. The real and complex parts of the refractive indices of milk fat and casein were calculated based on the absorption coefficients of a mixture of milk components as described by Aernouts et al. (2015b).

2.4 Estimation of PSD and VF

Milk fat PSDs were estimated from the μ_s, μ_s' and g spectra with a procedure similar to those described by Postelmans, Aernouts, and Saeys (2018) and Postelmans et al. (2019). The PSDs were approximated by a lognormal probability density function since this type performed best when fitting lognormal, normal and weibull distributions directly to the measured milk fat PSDs (results not shown). These directly fitted lognormal distributions were also used to set the parameter boundaries of the constrained optimization in the PSD estimation routine. Lower boundaries were set at 70% of the minimal fitted distribution parameter values, while upper boundaries were set at 130% of the maximal values. This resulted in the range -2.70 to 0.86 for lognormal distribution parameter μ and 0.29 to 1.89 for parameter σ. VF was limited by a minimum of zero fat content and a maximum of 3% v/v fat in three-fold diluted milk, given that the fat content in undiluted milk will rarely exceed 5.5% w/w (6.21% v/v) (Walstra et al., 1999). During optimization, all parameters were scaled to the range of 0.5-1.5 to reduce possible effects of differences in magnitude.

The PSD estimation routine consists of three steps (Postelmans et al., 2019): (1) defining the starting points of the optimization, (2) the optimization step in which the calculated scattering spectrum of the PSD estimate is iteratively updated to match the experimental scattering spectrum, and (3) a selection procedure on the solutions to retrieve the final PSD estimate.

(1) In order to determine a set of limited but relevant starting points, the cost based on normalized spectra (in Eq. (2) shown for μ_s) was evaluated for a grid of 75x75 points, equidistantly distributed in the scaled distribution parameter space. Local minima were detected using Matlab's 'imregionalmin' function (Image Processing Toolbox, Matlab R2016b, The Mathworks Inc.,

Massachusetts, USA). In case of μ_s and μ_s', the ratio of the mean input spectrum and the mean of the spectrum calculated for the local minima is a rough estimate for VF. This value had to be within the VF boundaries, otherwise this combination of PSD parameters was discarded as starting point. For g, the ratio of spectral means had to be in the 0.75-1.25 range, since a maximal multiplicative baseline of ±25% was tolerated. If more than ten local minima remained, only the ten with the lowest cost were retained as starting points. In case of estimation on μ_s and μ_s', an initial VF value was provided by the above-mentioned ratio of spectral means.

$$min \log_{10} \left[\sum_{i=1}^{N_\lambda} \left(\frac{\frac{\mu_{s,i}}{mean(\mu_s)} - \frac{\widehat{\mu_{s,i}}(par1,par2)}{mean(\widehat{\mu_s})}}{\frac{\mu_{s,i}}{mean(\mu_s)}} \right)^2 \right] \tag{2}$$

(2) The PSD estimation routine used the 'patternsearch' algorithm, a non-gradient based optimizer, as implemented in the Global Optimization Toolbox of Matlab R2016b (The Mathworks Inc., Massachusetts, USA; Conn, Gould, & Toint, 1997). In case of μ_s or μ_s', the PSD and VF were estimated simultaneously using the cost in Eq. (3). For PSD estimates based on g, only PSD parameters $par1$ and $par2$ remain in Eq. (3), since g is concentration independent in the assumed independent scattering regime.

$$min \log_{10} \left[\sum_{i=1}^{N_\lambda} \left(\frac{\mu_{s,i} - \widehat{\mu_{s,i}}(par1, par2, VF)}{\mu_{s,i}} \right)^2 \right] \tag{3}$$

(3) A selection procedure was applied to the solutions found for the different starting points. First, all non-converged end points were discarded, as well as solutions that reached one or more of the parameter boundaries. In case of g, the ratio of the mean input spectrum and the mean calculated spectrum had to be between 0.75-1.25, since a multiplicative baseline error of maximal 25% on the measurements was tolerated. From the remaining end points, only those with a cost value within 2.5% of the lowest remaining cost were considered. If they formed one group, i.e. absolute difference between scaled distribution not more than 0.05, the solution with the minimal cost was considered as final PSD estimate. If not, the estimated PSDs were considered non-unique and no final estimate was selected.

2.5 Estimation of PSD and VF on μ_s and g simultaneously

PSD and VF estimates were also made using μ_s and g spectra simultaneously. The distribution parameter boundaries and the steps of the optimization routine remained identical to those described in section 2.4. Only the cost functions were replaced to include both μ_s and g, respectively Eq. (2) by Eq. (4) for the grid calculation based on normalized spectra, and Eq. (3) by Eq. (5) in the optimization. In order to attribute an equal weight to μ_s and g, a cost function based on the ratio of the sum of squared errors (SSE) to the total squared errors (SST) was preferred over one based on the relative difference between the spectra.

$$\begin{aligned} &min \left[\frac{SSE_{\mu_s,norm.}}{SST_{\mu_s,norm.}} + \frac{SSE_{g,norm.}}{SST_{g,norm.}} \right] \\ &= min \left[\frac{\sum_{i=1}^{N_\lambda} (\frac{\mu_{s,i}}{mean(\mu_s)} - \frac{\widehat{\mu_{s,i}}(par1,par2)}{mean(\widehat{\mu_s})})^2}{\sum_{i=1}^{N_\lambda} (\frac{\mu_{s,i}}{mean(\mu_s)} - mean(\frac{\mu_s}{mean(\mu_s)}))^2} + \frac{\sum_{i=1}^{N_\lambda} (\frac{g_i}{mean(g)} - \frac{\widehat{g}_i(par1,par2)}{mean(\widehat{g})})^2}{\sum_{i=1}^{N_\lambda} (\frac{g_i}{mean(g)} - mean(\frac{g}{mean(g)}))^2} \right] \end{aligned} \tag{4}$$

$$min \left[\frac{SSE_{\mu_s}}{SST_{\mu_s}} + \frac{SSE_g}{SST_g} \right] = min \left[\frac{\sum_{i=1}^{N_\lambda} (\mu_{s,i} - \widehat{\mu_{s,i}}(par1, par2, VF))^2}{\sum_{i=1}^{N_\lambda} (\mu_{s,i} - mean(\mu_s))^2} + \frac{\sum_{i=1}^{N_\lambda} (g_i - \widehat{g}_i(par1, par2))^2}{\sum_{i=1}^{N_\lambda} (g_i - mean(g))^2} \right] \tag{5}$$

2.6 PSD estimation with fixed VF

All PSD estimations were repeated with the VF value fixed at the three-fold dilution of the fat content determined in the reference analysis. This way, a degree of freedom is eliminated from the optimization routine, leaving only the distribution parameters to be determined. For estimations based on μ_s and μ_s', the cost function used for both the starting point determination and for the optimization is based on non-normalized spectra, cf. Eq. (3) and Eq. (5), although VF is a fixed value instead of a parameter to be estimated.

3 Results and discussion

3.1 Reference PSDs

The PSD of casein was identified as the first mode of the bimodal PSD of raw milk and is shown in Fig. 1a. The casein micelles are clearly smaller than the fat globules in raw milk, but the higher the degree of homogenization, the more their PSDs overlap. The measured peak for casein shows a high similarity with the casein PSD reported by Aernouts et al. (2015b). However, both are clearly underestimating the average size of casein micelles that lies around 150-200 nm according to literature (Walstra et al., 1999; C. de Kruif, 1998; C. G. K. de Kruif & Huppertz, 2012). Stocker et al. (2017) measured a mean diameter of 189 nm for casein in commercial skim milk, but it is unclear whether this PSD obtained by dynamic light scattering refers to an intensity or volume based PSD. By fitting μ_s and μ_s' spectra based on Mie theory, they also estimated a casein PSD with a mean diameter of 211 nm. The discrepancy between our results and those reported by other researchers was most likely caused by the restriction of laser diffraction measurements to assume only one type of particles. More specific, only one particle refractive index could be defined, even if the sample is known to contain multiple particle types, as was the case here. Since milk fat globules were the scatterers of interest, the refractive index of milk fat was used. Because of this, the size of the casein micelles was calculated with a too low refractive index, resulting in a casein peak shifted to smaller sizes.

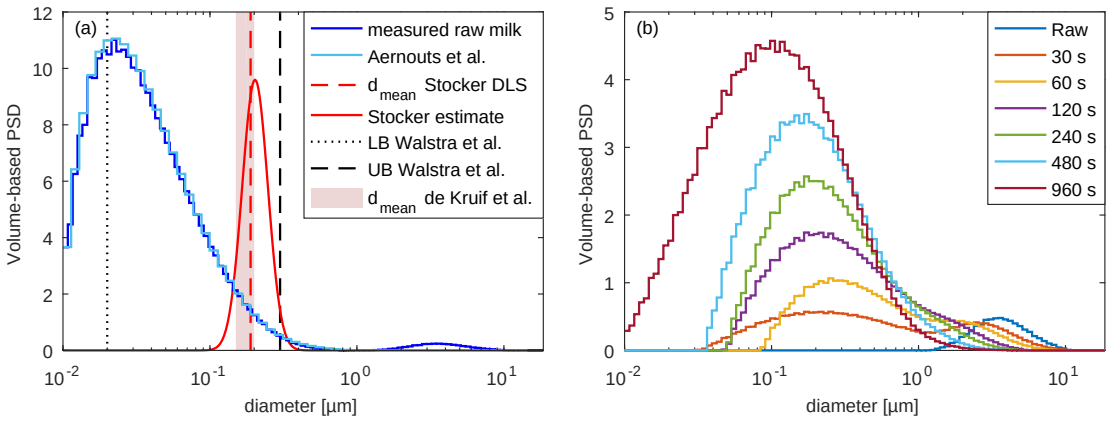

Figure 1: a) Measured PSD of raw milk compared to the casein PSD of Aernouts et al. (2015b), Stocker et al. (2017), Walstra et al. (1999), C. de Kruif (1998) and C. G. K. de Kruif and Huppertz (2012). b) Measured PSDs of milk fat (contribution of casein removed). LB = lower boundary, UB = upper boundary.

The measured PSDs for the milk fat globules, the scatterers of interest, are shown in Fig. 1b. They were obtained by subtracting the casein PSD in Fig. 1a, from all sample PSDs before converting them to probability density functions. The fat globule size distribution in raw milk is monomodal with the peak around 3.5 µm diameter. During homogenization, a second peak of smaller particles around 0.2 µm in diameter appears. The longer the homogenization time, the

more the relative importance of this new peak increases, rather than shifting the original peak to smaller particle sizes.

The PSD of fat globules in the unhomogenized raw milk sample was found to be monomodal, just as the example shown in the work of Cabassi et al. (2013). On the other hand, Jhanwar and Ward (2014) obtained a bimodal PSD for whole milk, even if casein was removed before the PSD measurement. Stocker et al. (2017) also noticed this in their PSD measurements, but they concluded that it was an artifact and assumed a monomodal PSD for the raw fat globules.

3.2 Measured & simulated BOP

The mean bulk scattering spectra calculated from DIS and UT measurements (after casein correction) are shown in Fig. 2. The μ_s spectra were well reproducible with a low noise level, while g and μ_s' were more susceptible to noise. This is possibly due to low total reflectance values in the DIS measurements and therefore a low signal-to-noise ratio. Moreover, such errors on the total reflection spectra could also be the cause of the baseline mismatch between measured and simulated g and μ_s' spectra in case of the 'raw', '30 s' and '60 s' samples. A slight underestimation in μ_s', in combination with a correct μ_s (less depending on DIS measurements), leads to an overestimated g through the relation $g = 1 - \mu_s'/\mu_s$. Inaccuracies on the sample or particle refractive index can also contribute to such baseline effects (Postelmans et al., 2018), as an approximate formula for the refractive index of milk and milk serum was used.

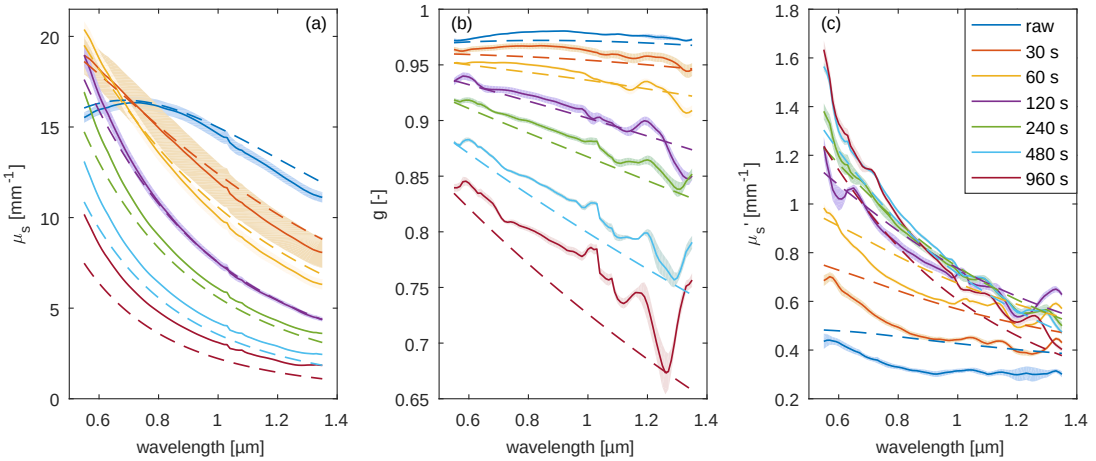

Figure 2: Mean experimental spectra with casein correction ± standard deviation for a) μ_s, b) g and c) μ_s'. Dashed lines indicate the corresponding spectra simulated based on the reference PSDs.

In Fig. 2, the increasing underestimation of simulated μ_s spectra compared to the experimental casein corrected spectra suggests there might be a systematic effect on top of the measurement errors. For samples with homogenization times ranging from 240 s to 960 s, the baseline mismatch clearly aggravates in both μ_s and g. It is therefore thought to be related to the homogenization process itself. In raw milk, the fat globule membrane mainly consists of phospholipids and proteins, and has an average thickness around 15 nm (Walstra et al., 1999). Upon homogenization, the total globule surface area increases due to the newly formed small fat globules. To cover the increased fat-milk plasma interface, the original membrane material is completed with adhering casein and serum proteins (Strawbridge, Ray, Hallett, Tosh, & Dalgleish, 1995). Walstra et al. (1999) reported an average protein load per surface area of 10 mg/m^2. Casein (sub)micelles are preferentially adsorbed over serum proteins and make up about 93% of the proteins in the new surface layer, with a preference for the largest micelles (Walstra et al., 1999). Therefore, the concentration of free casein micelles also decreases.

The discrepancy between simulated and experimental scattering spectra, especially for the milk samples that were homogenized for a longer time, suggests that the full complexity of the homogenization process is not captured in the simulations. Both the casein coating and the decrease in free casein micelles would result in an increase of the scattering coefficient and the anisotropy factor, bringing the simulated and experimental profiles closer together (results not shown). However, as information on the exact refractive indices of casein and milk fat, the fat globule membrane thickness and the volume fraction of free casein is not available, it was not possible to take this full complexity into account.

3.3 PSD estimation on μ_s

The PSDs estimated on simulated and experimental μ_s spectra are plotted together in Fig. 3 (& Fig. S-1). The estimations on simulated spectra provide the 'ideal' case, since the simulations were noise-free and the same refractive indices were used for the forward simulation and the inverse estimation. In all cases shown in red, the VF was estimated simultaneously with the distribution parameters. The estimated VF values are discussed separately in section 3.7.

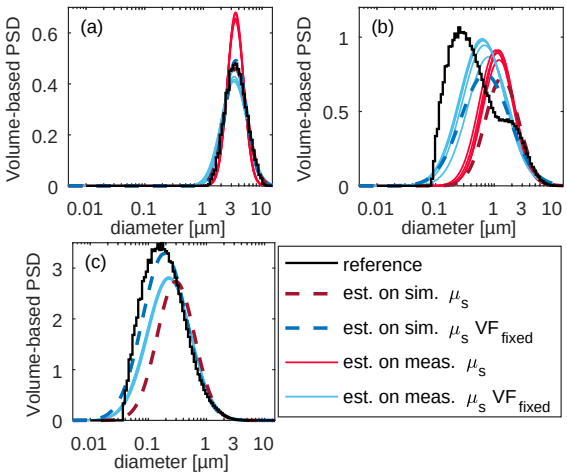

Figure 3: PSDs of milk fat globules estimated on simulated and experimental μ_s spectra, with estimated or fixed VF. a) 'raw', b) '60 s', c) '480 s'.

The PSD of the 'raw' milk sample (Fig. 3a) was well estimated based on μ_s spectra, as the lognormal distribution properly fits the shape of the measured PSD. A lognormal PSD was also used by Stocker et al. (2017), while Cabassi et al. (2013) preferred a Weibull distribution. For homogenized samples, the estimated PSD is mainly dominated by the larger fat globules (Fig. 3b) since their scattering is more pronounced than this of the smaller particles. A lower sensitivity to submicron particles due to the similar shape of their μ_s spectra was already observed when estimating PSDs (Postelmans et al., 2019). For the most intensively homogenized samples ('480 s' and '960 s'), even no valid estimates were retained by the selection procedure. Stocker et al. (2017) reported similar difficulties: a small difference in mean particle size can have a large effect on μ_s, but the estimation algorithm may attribute it to a change in particle concentration.

If, however, the VF was incorporated as a fixed value (1.56% v/v, three-fold dilution of the reference analysis) instead of being estimated, it resulted in valid PSD estimates for all samples. Moreover, the underestimation of the small globule fraction reduced (Fig. 3 blue lines), because errors in scattering level due to an incorrect particle size could no longer be compensated for by adapting the particle concentration. Cabassi et al. (2013) also used a fixed fat concentration when estimating PSDs, although their estimation routine included an additional correction factor besides the two distribution parameters. Stocker et al. (2017) estimated the concentration of milk fat simultaneously with the PSD parameters, but no reference analysis was done to confirm

the accuracy of their VF estimates. In practical applications, the VF of fat globules could be determined based on absorption spectroscopy or the bulk absorption coefficient μ_a.

3.4 PSD estimation on g

The PSDs estimated on g spectra are shown in Fig. 4 (& Fig. S-2). In case of the 'raw' sample, no valid estimate was obtained based on experimental spectra as all replicates reached the lower boundary for σ. The PSDs of homogenized samples are more accurate: the higher the degree of homogenization, the more the distribution peak shifts to smaller sizes. Furthermore, the distributions are wider and approximate the left distribution tail more accurately. Nevertheless, PSDs estimated on experimental spectra tend to be less wide than their respective counterpart estimated on simulations, although the difference decreased with increasing degree of homogenization.

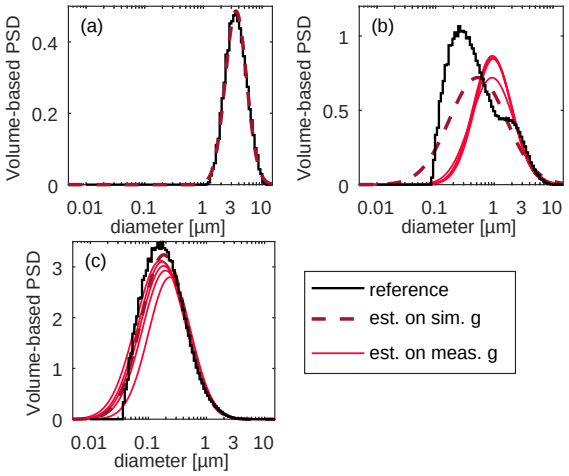

Figure 4: PSDs of milk fat globules estimated on simulated and experimental g. a) 'raw', b) '60 s', c) '480 s'.

3.5 PSD estimation on μ_s'

In general, there was less variation in the level of μ_s' spectra compared to μ_s and the spectra showed similar noise as g. This negatively affected the PSD estimation, since only five valid PSD estimates were obtained when estimated simultaneously with the VF (Fig. 5 red lines, Fig. S-3). Similarly, problems with the distribution width estimated based on μ_s' were already reported for polystyrene particle suspensions (Postelmans et al., 2018, 2019).

Fixing VF drastically increased the number and quality of the valid PSD estimates, namely all samples ranging from 60 s to 960 s homogenization (Fig. 5b-c, blue lines). Stocker et al. (2017) also fixed the VF of fat while estimating a bimodal lognormal distribution on μ_s'. However, the effect of incorporating a known VF cannot be investigated based on their data since no PSD estimates with VF estimation were reported by them, nor any reference PSD measurements.

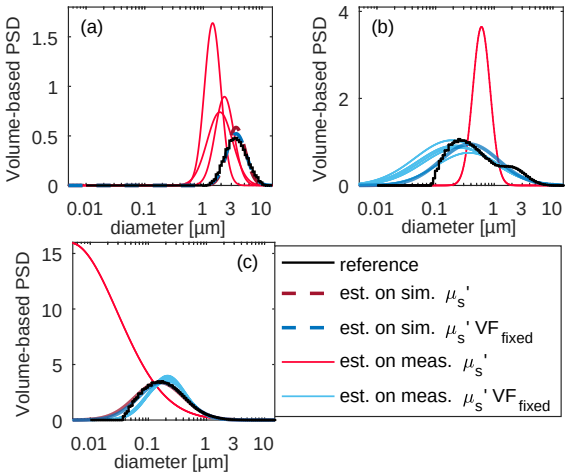

Figure 5: PSDs of milk fat globules estimated on simulated and experimental μ_s' spectra, with estimated or fixed VF. a) 'raw', b) '60 s', c) '480 s'.

3.6 PSD estimation more robust by combining μ_s and g in cost

When looking at the PSDs estimated on μ_s, the input spectrum is fitted relatively accurate. However, the g spectrum of these estimated PSDs often does not match well with the corresponding g spectrum. On the contrary, the PSDs estimated on g, have most often a well-matching normalized μ_s spectrum, but no VF estimation could be made. Therefore, it was investigated if PSDs estimated on a combination of μ_s and g would inherit 'the best of both': the PSD estimates on g with a VF estimate on μ_s.

Using both scattering spectra as input produces a valid PSD and VF estimate for all samples except raw milk, as can be seen in Fig. 6 (red lines, Fig. S-4). The estimates for raw milk stranded at the lower boundary of distribution parameter σ in an attempt to fit both the level and shape of the experimental spectra, just like the estimates on solely g. The estimated PSDs are generally wider than those estimated on μ_s and resemble more those estimated from the g spectra. The effect of using a fixed VF in estimations on a combination of μ_s and g is rather small, as shown by the blue lines in Fig. 6. Combining μ_s and g provided the highest number of valid PSD estimates, with the smallest difference in estimated distribution parameters with or without a fixed VF value.

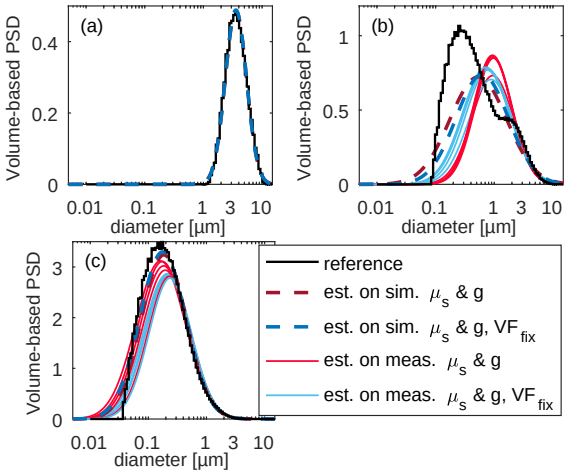

Figure 6: PSDs of milk fat globules estimated on a combination of μ_s and g (simulated or experimental), with estimated or fixed VF. a) 'raw', b) '60 s', c) '480 s'.

Therefore, if the concentration of scatterers is known, it is beneficial for the PSD estimation to fix the VF value and only estimate the distribution parameters. If VF is unknown, a simultaneous estimation of the PSD parameters and VF based on μ_s and g spectra provides a good alternative.

3.7 Estimated milk fat volume fractions

Figure 7 provides an overview of all estimated VF values. Firstly, there is a clear trend in VF values estimated on simulated μ_s spectra (red dots in Fig. 7a). As the PSD estimate for the raw milk sample was accurate, the accompanying VF is also relatively correct. Since the PSD estimation routine focuses on the large fat globules for bimodal and asymmetric PSDs, the VF is underestimated to compensate for the higher scattering level of larger particles. For higher degrees of homogenization, PSDs become again more monomodal and the estimated PSDs are more accurate, while the estimation of VF stabilizes or even improves. The same general trend is present in case of experimental μ_s spectra (blue dots in Fig. 7a). Fig. 7b only presents a limited set of VF estimates, of which one is even close to being discarded, because the majority of the PSD estimates on μ_s' were marked as invalid.

The relatively good PSD estimates based on μ_s and g simultaneously were accompanied by relatively accurate VF estimates. Fig. 7c shows that the estimated VF values for the '30 s', '60 s' and '120 s' samples based on experimental spectra are rather constant, with a small underestimation compared to the reference. The most intensively homogenized samples ('480 s' and '960 s') on the other hand, have overestimated VF values, most likely caused by underestimation of the left distribution tail without an accompanying overestimation of the fraction large particles.

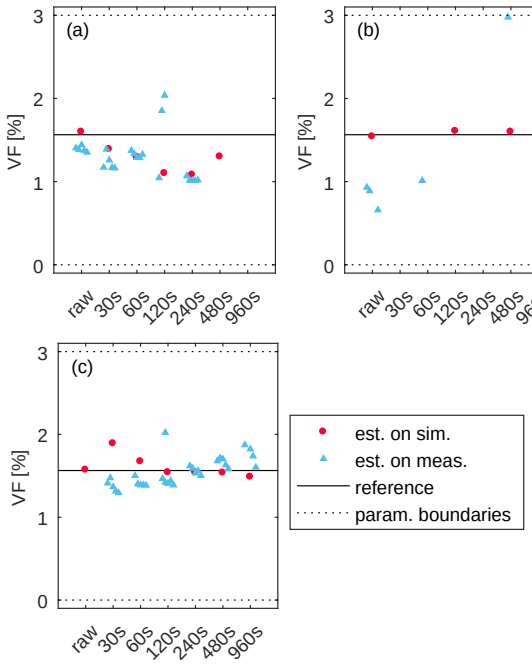

Figure 7: Volume fractions (VF) of milk fat estimated based on a) μ_s, b) μ_s', and c) of μ_s and g simultaneously.

3.8 General discussion

The presented PSD estimates were based on μ_s, g and μ_s' spectra, with the most accurate results if the concentration of the fat globules was fixed, or when μ_s and g were used simultaneously as input. From a practical point of view, μ_s' would be most easy to implement as it requires solely DIS measurements or spatially resolved reflectance spectroscopy. Watté, Aernouts, Van Beers,

Postelmans, and Saeys (2016) et al. already reported the *in silico* optimization of a spatially resolved reflectance sensor design for determining the BOP of milk. However, μ_s' was more noisy than μ_s in this dataset and already resulted in larger errors in the estimated distribution width (Postelmans et al., 2019). Nevertheless, relatively accurate PSD estimates were obtained for homogenized samples if VF information was included. Stocker et al. (2017) used μ_s' spectra to estimate bimodal PSDs of milk, but no reference measurements of the samples' PSDs are available to check the quality of these estimates. The PSD estimates for the 'raw' samples based on μ_s can be compared to those of Cabassi et al. (2013), who used corrected NIR absorption spectra to obtain PSD information. The D32 and span (D80-D20) of PSD estimates presented here were less correct if VF was estimated (RMSE of respectively 0.195 and 1.013 for our study compared to 0.110 and 0.47 for Cabassi et al.). However, the accuracy improved to a similar level if a fixed VF was used (RMSE_{D32} of 0.131, $\text{RMSE}_{\text{D80-D20}}$ of 0.472).

The consistent underestimation of the fraction small fat globules, observed for all BOP types, is caused by the relatively limited scattering by submicron particles compared to the scattering by larger particles. On top of that, it might also be related to the baseline mismatch between experimental and simulated spectra (Fig. 2). Identifying and reducing the cause of the mismatch might help to decrease this phenomenon. In case of homogenized milk, this may imply an adaptive casein correction that takes into account the full complexity of the homogenization process and the effects of a, most likely, increased particle refractive index on the scattering properties. Enhancing the sensitivity to the submicron size globules can also be obtained by including shorter wavelengths in the spectra, since the smaller the particle, the more its scattering peak shifts to smaller wavelengths. Michels, Foschum, and Kienle (2008) studied the BOP of different types of soy bean oil emulsions, and the start of the scattering peak in μ_s appeared around 0.4 µm wavelength for the Lipovenoes 10% and the ClinOleic 20% samples (largest particle diameter around 0.55 µm).

Despite the issue of underestimating the left tail of the measured PSDs, the right tail is fitted quite well, even for the bimodal and asymmetric PSDs of homogenized samples. Therefore, it contains valuable information on the largest particles in the samples, even if the rest of the distribution is not estimated perfectly. Fig. S-5 provides an overview of the D90 and D32 of the estimated and reference PSDs. It shows that the estimated values for samples homogenized for 120 s or longer are consistent between the replicates and close to the reference values if a fixed VF was used. Furthermore, Table S-1 includes the mean difference between measured and reference values (D50, D90, D32, D43) listed per sample. The standard deviation on these values for milk samples homogenized for 120 s or longer are small, indicating a good reproducibility. Di Marzo et al. (2016) investigated if these PSD parameters could be predicted accurately to monitor the performance of the homogenizer inside a MIR analyser (Milkocan) and alert if it is not working properly (D90 > 1.7 µm) (Smith, Barbano, Lynch, & Fleming, 1995). Imposing an upper limit on particle size rather than specifications on the complete PSD would be applicable in milk homogenization since large fat globules have the largest impact on creaming properties. As the presented PSD estimation routine provides good results for the estimation of D90, it has large potential to be used for such purposes.

Although the monomodal distributions were fitted quite well and relatively accurate D90 and D32 values were obtained, there is a clear issue of bimodality. For the bimodal and asymmetric homogenized PSDs, e.g. samples '30 s' and '60 s', a combination of two lognormal distributions would be more suitable than a single one. This 'bimodal lognormal distribution' was already used by Stocker et al. (2017) to estimate the PSD of homogenized milk samples based on μ_s and μ_s'. Since the use of μ_s and g simultaneously as input for PSD estimation gave the most promising results, a next step could be to estimate such a bimodal PSD based on these two scattering spectra.

4 Conclusion

The potential of estimating milk fat PSDs based on wavelength dependent light scattering properties for monitoring the homogenization process of milk was investigated. Therefore, the bulk

optical properties of raw milk samples with an increasing degree of ultrasonic homogenization were experimentally determined by means of double integrating sphere and unscattered transmission measurements. The bulk scattering spectra of μ_s, g, μ_s' or μ_s and g simultaneously were used as input for the estimation of lognormal PSD parameters and the volume fraction of milk fat globules. Estimated PSDs were compared to reference PSDs obtained with laser diffraction.

If the volume fraction and PSD parameters were estimated simultaneously, PSD estimates based on measured μ_s spectra focused on the largest particles, especially in the mildly homogenized samples (bimodal distributions). For strongly homogenized samples, no valid estimates could be made due to the inability of the algorithm to distinguish between a small change in particle size and a change in VF. Furthermore, measured μ_s' spectra produced practically no valid estimates since the optimizer stranded on one of the distribution parameter boundaries. Overall, estimates based on a combination of μ_s and g proved to be most robust as valid estimates were produced for all samples except raw milk.

A second set of PSDs was estimated with the VF fixed at the reference value instead of estimating the VF. Including this information in the estimation routine drastically improved the number and accuracy of the PSD estimates, especially in case of μ_s'. PSDs estimated on μ_s, g or both still retained the tendency of underestimating the number of small particles, but not as severe as in case of a simultaneously estimated VF.

Overall, the single lognormal distribution was not able to fit bimodal PSDs and focussed on the largest fat globules. Nevertheless, the good fits for the right distribution tail provided relatively accurate information on the D90 of the samples, and on the D32 for samples homogenized for 120 s or more. Therefore, the presented estimation routine could be a useful tool for monitoring specifications on the largest particles. Moreover, the good PSD estimates on a combination of μ_s and g spectra invite to estimate a weighted combination of two lognormal distributions based on these input spectra in order to improve the accuracy for bimodal PSDs.

Acknowledgments

This work was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders [IWT grant 141687] and the Research Foundation Flanders [FWO grant 12K3916N]

Declaration of interest: none.

References

Aernouts, B., Polshin, E., Lammertyn, J., & Saeys, W. (2011). Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: reflectance or transmittance? *Journal of Dairy Science*, *94*(11), 5315–5329. doi: 10.3168/jds.2011-4354

Aernouts, B., Polshin, E., Saeys, W., & Lammertyn, J. (2011). Mid-infrared spectrometry of milk for dairy metabolomics: A comparison of two sampling techniques and effect of homogenization. *Analytica Chimica Acta*, *705*(1-2), 88–97. doi: 10.1016/j.aca.2011.04.018

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Lammertyn, J., & Saeys, W. (2015a). Visible and near-infrared bulk optical properties of raw milk. *Journal of Dairy Science*, *98*(10), 6727–38. doi: 10.3168/jds.2015-9630

Aernouts, B., Van Beers, R., Watté, R., Huybrechts, T., Jordens, J., Vermeulen, D., ... Saeys, W. (2015b). Effect of ultrasonic homogenization on the Vis/NIR bulk optical properties of milk. *Colloids and Surfaces B: Biointerfaces*, *126*(17), 510–519. doi: 10.1016/j.colsurfb.2015.01.004

Aernouts, B., Watté, R., Van Beers, R., Delport, F., Merchiers, M., De Block, J., ... Saeys, W. (2014). Flexible tool for simulating the bulk optical properties of polydisperse spherical particles in an absorbing host: experimental validation. *Optics express*, *22*(17), 20223–20238. doi: 10.1364/OE.22.020223

Aernouts, B., Zamora-Rojas, E., Van Beers, R., Watté, R., Wang, L., Tsuta, M., ... Saeys, W. (2013). Supercontinuum laser based optical characterization of Intralipid® phantoms in the 500-2250 nm range. *Optics Express*, *21*(26), 32450–32467. doi: 10.1364/OE.21.032450

Bogomolov, A., Melenteva, A., & Dahm, D. J. (2013). Fat globule size effect on visible and shortwave near infrared spectra of milk. *Journal of Near Infrared Spectroscopy*, *21*(5), 435–440.

Bylund, G. (2003). *Dairy processing handbook*. Lund: Tetra Pak Processing Systems AB.

Cabassi, G., Profaizer, M., Marinoni, L., Rizzi, N., & Cattaneo, T. M. P. (2013). Estimation of fat globule size distribution in milk using an inverse light scattering model in the near infrared region. *Journal of Near Infrared Spectroscopy*, *21*(5), 359–373. doi: 10.1255/jnirs.1070

Conn, A. R., Gould, N. I. M., & Toint, P. L. (1997). A globally convergent augmented lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation*, *66*(217), 261—288.

de Kruif, C. (1998). Supra-aggregates of casein micelles as a prelude to coagulation. *Journal of Dairy Science*, *81*(11), 3019–3028.

de Kruif, C. G. K., & Huppertz, T. (2012). Casein micelles: size distribution in milks from individual cows. *Journal of agricultural and food chemistry*, *60*(18), 4649–4655.

Di Marzo, L., Cree, P., & Barbano, D. M. (2016). Prediction of fat globule particle size in homogenized milk using Fourier transform mid-infrared spectra. *Journal of Dairy Science*, *99*(11), 8549–8560. doi: 10.3168/jds.2016-11284

Fox, P., & McSweeney, P. (1998). *Dairy chemistry and biochemistry*. London: Blackie Academic & Professional.

ISO. (2013). *Milk and liquid milk products - Guidelines for the application of mid-infrared spectrometry. Page 14 in International Standard ISO 9622:2013/IDF 141:2013. International Dairy Federation.*

Jhanwar, A., & Ward, R. E. (2014). Particle size distribution and lipid composition of skim milk lipid material. *International Dairy Journal*, *36*(2), 110–117. doi: 10.1016/j.idairyj.2014.01.010

Lynch, J., Barbano, D., Schweisthal, M., & Fleming, J. (2006). Precalibration evaluation procedures for mid-infrared milk analyzers. *Journal of Dairy Science*, *89*(7), 2761–2774. doi: 10.3168/jds.S0022-0302(06)72353-0

Michels, R., Foschum, F., & Kienle, A. (2008). Optical properties of fat emulsions. *Optics Express*, *16*(8), 5907–5925. doi: 10.1364/OE.16.005907

Postelmans, A., Aernouts, B., & Saeys, W. (2018). Estimation of particle size distributions from bulk scattering spectra: sensitivity to distribution type and spectral noise. *Optics Express*, *26*(12), 15015–15038. doi: 10.1364/OE.26.015015

Postelmans, A., Aernouts, B., & Saeys, W. (2019). Estimation of particle size distribution from bulk scattering spectra: validation on monomodal suspensions. *Analytical chemistry*, *91*(15), 10040–10048. doi: 10.1021/acs.analchem.9b01913

Prahl, S. A. (2011). *Everything I think you should know about inverse adding doubling.* Retrieved from `https://omlc.org/software/iad/manual.pdf` (Accessed 14 October 2019)

Segelstein, D. J. (1981). *The complex refractive index of water* (Doctoral dissertation, University of Missouri-Kansas City). doi: -

Smith, E., Barbano, D., Lynch, J., & Fleming, J. (1995). Infrared analysis of milk: effect of homogenizer and optical filter selection on apparent homogenization efficiency and repeatability. *Journal of AOAC International*, *78*(5), 1225–1233. doi: -

Stocker, S., Foschum, F., Krauter, P., Bergmann, F., Hohmann, A., Scalfi Happ, C., & Kienle, A. (2017). Broadband optical properties of milk. *Applied Spectroscopy*, *71*(5), 951–962. doi: 10.1177/0003702816666289

Strawbridge, K. B., Ray, E., Hallett, F. R., Tosh, S. M., & Dalgleish, D. G. (1995). Measurement of particle size distributions in milk homogenized by a microfluidizer: estimation of populations of particles with radii less than 100 nm. *Journal of Colloid and Interface Science*, *171*, 392–398. doi: 10.1006/jcis.1995.1195

Walstra, P., Geurts, T., Noomen, A., Jellema, A., & van Boekel, M. (1999). *Dairy technology: principles of milk properties and processes*. New York: Marcel Dekker Inc.

Walstra, P., Jenness, R., & Badings, H. (1984). *Dairy chemistry and physics*. New York: John Wiley & Sons.

Walstra, P., Wouters, J., & Geurts, T. (2005). *Dairy science and technology*. Boca Raton: CRC Press.

Watté, R., Aernouts, B., Van Beers, R., Postelmans, A., & Saeys, W. (2016). Computational optimization of the configuration of a spatially resolved spectroscopy sensor for milk analysis. *Analytica Chimica Acta*, *917*, 53–63. doi: 10.1016/j.aca.2016.02.041