

A Random Forest Approach for Counting Silicone Oil Droplets and Protein Particles in Antibody Formulations Using Flow Microscopy

Miguel Saggu¹  • Ankit R. Patel¹ • Theodoro Koulis²

Received: 7 March 2016 / Accepted: 5 December 2016 / Published online: 19 December 2016
© Springer Science+Business Media New York 2016

ABSTRACT

Purpose To evaluate a random forest model that counts silicone oil droplets and non-silicone oil particles in protein formulations with large class imbalance.

Methods In this work, we present a novel approach for automated image analysis of flow microscopy data based on random forest classification enabling rapid analysis of large data sets. The random forest approach overcomes many of the limitations of traditional classification schemes derived from simple filters or regression models. In particular, the approach does not require *a priori* selection of important morphology parameters.

Results We analyzed silicone oil droplets and non-silicone oil particles observed in four model systems with protein concentrations of 20, 50 and 125 mg/mL. Filters based on random forests achieve higher classification accuracies when compared to regression based filters. Additionally, we showcase a procedure that allows for accurate counting of particles $\geq 1 \mu\text{m}$.

Conclusions Our method is generally applicable for classification and counting of different classes of particles as long as class morphologies are differentially expressed.

KEY WORDS classification · counting · flow microscopy · mAb · random forest · silicone oil

ABBREVIATIONS

CART	Classification and regression tree
ECD	Equivalent circular diameter
ESD	Equivalent spherical diameter
mAb	Monoclonal antibody
NSO	Non-silicone oil
PFS	Pre-filled syringe
SO	Silicone oil

INTRODUCTION

All protein formulations contain subvisible particles, and discrimination between proteinaceous particles and other particles is key in assessing product stability and potential risk factors (1). Regulatory agencies require manufacturers of protein therapeutics to control subvisible particles in drug products to ensure the safety and efficacy of the drug as well as to demonstrate process consistency. More recent concerns regarding the potential immunogenicity of proteinaceous aggregates have led to increased scrutiny by regulatory health authorities (1–3). Limits on particle counts are specified in pharmacopeias for particles ≥ 10 and $\geq 25 \mu\text{m}$, while particles ≥ 2 and $\geq 5 \mu\text{m}$ are typically reported for information purposes.

According to USP <788>, the preferred method for the determination of subvisible particles in biopharmaceutical formulations is light obscuration (Method 1), and in the case of samples with high viscosity or reduced clarity, membrane microscopy (Method 2) may be used. However, light obscuration is not able to distinguish between particles of different compositions, which are common in formulations manufactured in pre-filled syringes (PFS). This limitation is a major drawback in terms of particle characterization, in particular in

Electronic supplementary material The online version of this article (doi:10.1007/s11095-016-2079-x) contains supplementary material, which is available to authorized users.

✉ Miguel Saggu
saggu.miguel@gene.com

✉ Theodoro Koulis
koulis.theodoro@gene.com

¹ Late Stage Pharmaceutical Development, Genentech Inc., South San Francisco, California 94080, USA

² Nondinical Biostatistics, Genentech Inc., South San Francisco, California 94080, USA

biopharmaceutical industry where most of the interest is focused on quantitation of proteinaceous particles. Over the past few years, new technologies have been introduced into the area of particle analysis and are considered to be orthogonal techniques to light obscuration (4,5). Several of these new technologies are able to distinguish between different types of particles, thus providing additional information about product stability. For example, flow microscopy (based on different morphologies) and Archimedes microchannel resonator (based on differences in particle buoyancy) have been reviewed in the informational chapter USP <1787> (6–8).

Increasingly, pharmaceutical companies are manufacturing protein therapeutics in PFS because of their ease of handling and administration, especially in indications for which home use is desired. However, silicone oil is used as a plunger lubricant in PFS, and so it can migrate into the product in the form of droplets. Common techniques, such as light obscuration, are not able to distinguish *silicone oil* (SO) droplets from *non-silicone oil* (NSO) particles present in solution (9). In contrast, flow microscopy is able to distinguish between particles of different morphologies by taking digital images of individual particles (8,10). This technique has been successfully used to distinguish spherically shaped particles from other particles present in solution. Flow microscopy allows for the differentiation of silicone oil and/or air bubbles from other species, such as proteinaceous particles (7,11,12). State-of-the-art systems, such as the FlowCam (FluidImaging Technologies) or MFI (ProteinSimple), report up to 35 different morphology measurements per image, e.g., *aspect ratio*, *circularity*, etc. Filters based on these morphology measurements can be used to distinguish silicone oil droplets from other particles. Early studies have suggested an *aspect ratio* ≥ 0.85 to identify spherically particles, while more recent studies suggest *circularity* (Hu) ≥ 0.95 , or a combination of several parameters (7,8,11). While these single morphology filters work reasonably well for particles larger than 10 μm , their accuracies in the size range below 10 μm are usually inferior compared to regression models due to the decreased number of pixels available to obtain accurate morphology information from an image. In addition, different morphology parameters are necessary for discrimination of SO and NSO in different size regimes depending on the morphologies of the NSO particles (11). In stability studies involving protein therapeutics in PFS, the imbalance in the number of SO and NSO particles can affect count accuracies, which can be a problem for assessing the aggregation stability of a PFS product. Even with morphology filters having an accuracy of about 90%, a small prevalence of NSO often results in large errors in NSO counting, thereby providing inaccurate stability information of the product (7).

In this paper, we describe a method for constructing morphology filters that is based on the random forest, a statistical machine learning method derived from decision trees (see (13) for an introduction). In our context, random forests are used

to create predictive models that map morphology data of a given particle to a classification (SO or NSO). Random forests are easy to use and are implemented in most popular software packages. However, some of the concepts we cover in this paper can be adapted easily to other machine learning methods. In addition, we showcase a counting method called the mixture method to account for less-than-perfect classification, which often leads to over-counting of NSO particles whenever the prevalence of SO particles is large.

Filters based on the random forest were used to analyze particles in four different model systems spanning the concentration range from 20 mg/mL up to 125 mg/mL protein. All particles were generated artificially and classified in the training and validation sets as SO or NSO in the size range above 1 μm , which is at the detection limit of typical flow microscopes. Training and validation sets using particles from each model system were created. The training sets were used to build the random forest filters, and the validation sets were used to assess their classification and counting accuracies. The relative importance of various morphology parameters is also assessed for different size ranges and protein particle generation methods

MATERIALS AND METHODS

Materials

Monoclonal antibodies were produced by Genentech, Inc. (South San Francisco, California). mAb1 was formulated at 125 mg/mL and mAb2 at 20 mg/mL with buffers, stabilizers, and surfactants typical of biopharmaceutical formulations. BSA was purchased from Sigma-Aldrich at a purity $\geq 98\%$. From these materials four different model systems with artificially generated particles were created. Thermally stressed antibody samples of mAb1 and mAb2 were generated by heating 20 mg/mL of mAb in the presence of 150 mM sodium chloride to 73°C for 5 min. Stir stressed mAb1 sample was generated by stirring 20 mg/mL mAb at 500 rpm overnight. Heat stressed BSA was generated by heating 50 mg/mL BSA in PBS at pH 7.4 to 80°C under stirring at 500 rpm. The generated proteinaceous mAb1 particles were diluted into 0.22 μm filtered mAb1 sample (125 mg/mL). mAb2 particles were diluted into 0.22 μm filtered mAb2 sample (20 mg/mL). BSA particles were diluted into 0.22 μm filtered BSA sample (50 mg/mL). Protein particle concentrations in each model system were $\sim 50,000$ particles/mL $\geq 2 \mu\text{m}$.

Silicone oil stock solution was generated by spiking 2% (v/v) silicone oil (Dow Corning 360 medical fluid, 1000 cSt) into freshly filtered mAb formulation buffer or PBS. The solution was vortexed for 1 min and sonicated for an additional 10 min. An aliquot of this stock solution was spiked into the

filtered mAb and BSA solutions to a concentration of $\sim 50,000$ particles/mL $\geq 2 \mu\text{m}$.

Flow Microscopy

All samples were compared on two different flow microscopes. The first flow microscope was a MFI 5200 (ProteinSimple) equipped with a $100 \mu\text{m}$ silane coated flow cell, $5\times$ objective, monochrome camera, and peristaltic pump. The instrument was autofocused using $10 \mu\text{m}$ Duke size standards and controlled with the MFI View software version 2R3.0.1.24.2461 while MVAS version 1.3 was used for image processing.

The second flow microscope was a FlowCAM VSII (Fluid Imaging Technologies) equipped with a Sony SX90CR color camera. A $10\times$ objective and a $90 \mu\text{m}$ field-of-view cell were used for all measurements. The system was focused using high-magnification focusing beads and autofocus (Fluid Imaging Technologies). Particle detection thresholds were selected as 12 for dark pixels and 15 for light pixels. Distance to nearest neighbor was set to 2 with 3 closed-hole iterations. Images were processed using VisualSpreadsheet v3.4.11 (Fluid Imaging Technologies).

Instrument performance was verified by measuring $10 \mu\text{m}$ Count-CAL bead standards at the beginning of each session on both MFI and FlowCam systems.

Modeling Software

The random forest filters were created using *R*, a software environment for statistical computing (14), in conjunction with the *caret* package (Classification And REgression Training) (14). The *caret* package includes numerous functions that simplify and streamline the creation of predictive models. The package contains tools for splitting data (for creating training and validation sets and for *K*-fold cross-validation), model tuning (for selecting classification parameters), estimation of variable importance, as well as methods for balancing error in unbalanced data sets. Numerous predictive modeling examples using the *caret* package can be found in (13,15).

Random Forest Filters

The decision tree is a popular machine learning method that uses a set of binary rules to predict class memberships (16). These rules depend on algorithms, such as CART (17), and are readily available in many statistics software packages. These algorithms are computationally fast and do not impose strict assumptions on the data.

Filters based on decision trees have been used in multiple disciplines, such as pharmacology (18), physics (19), molecular biology (20), and medicine (20–22). In our context, the method provides a predictive filter that takes morphological information about a particle and provides a prediction for the

particle's class (NSO or SO). This predictive filter is constructed by learning from particles with known class and morphological information. The filter can then take the morphological information from an unknown particle and provide its predicted class (either NSO or SO). An example of such a decision tree is shown in Fig. 1a. The tree keeps branching until one of several predetermined stopping rules of the CART algorithm are reached (17).

Instead of a single decision tree, an ensemble of stochastically generated decision trees can be used together. This collection of decision trees is called a random forest, and it belongs to the class of supervised learning algorithms. Random forests were developed to overcome some of the limitations of decision trees, such as unstable predictions (i.e. sensitive to small perturbations of the inputs) and over-fitting (23). Because they are robust with respect to over-fitting, a large number of variables can be used as inputs, so that little or no *a priori* selection of the variables is required. In addition, random forests are computationally efficient for large datasets, and can produce very accurate results with minimal tuning. The methodology is flexible so that internal estimates obtained during model construction can be used to monitor error, accuracy, and variable importance (i.e. importance of morphologies for accurate classification). Each tree in the random forest is allowed to vote on the class for a given particle, and the number of votes for a specific class is aggregated to obtain an average vote, p , over all trees. A cutoff c is then used to classify the particle: if $p > c$ then the particle is classified as NSO, otherwise it is classified as SO. The cutoff c is treated as a tuning parameter that is determined as the forest is created, via *K*-fold cross-validation and optimization of the ROC curve (13). This choice of cutoff is robust to class imbalance as demonstrated in previous work (13).

In general, a random forest is constructed from a training set, and a test or validation set is then used to evaluate prediction performance. For our application, the particle data from our four model systems were split randomly into two sets; 80% of the particles were retained for training and the remaining 20% were used for validation. In general, a random forest should be comprised of a large number of trees in order to obtain the best overall accuracy. However, a very large number of trees will often result in small gains in accuracy and will increase computational costs (e.g., memory usage, processing time, etc.). Recently, a range between 64 and 128 trees per forest was recommended in order to obtain a good balance between classification performance and computational cost (24). In this paper, every random forest is composed of 128 trees.

Logistic S-Factor

Filters based on the random forest were compared to predictive filters based on the S-factor (11). The S-factor of a particle

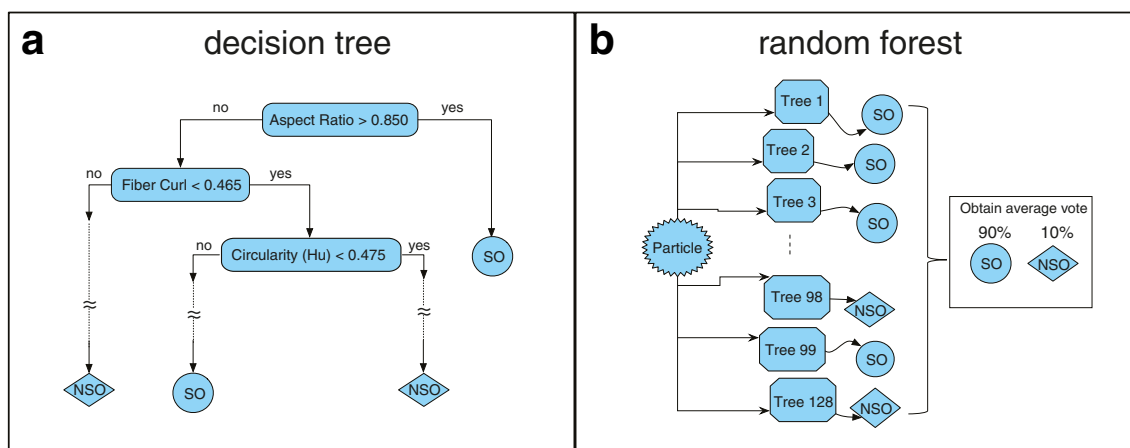


Fig. 1 (a) Example of a decision tree used in a random forest filter to separate SO from NSO (only the first three nodes are shown). Particles that satisfy the condition, e.g. aspect ratio > 0.850 , get placed in the right node, while particles that do not satisfy the condition get placed in the left node. (b) Random forest filter applied to each particle to obtain an average vote percent from all decision trees to predict whether a particle is SO or NSO. Note that the number of trees in the forest is independent of the number of morphology parameters.

is defined as the product of its *Circularity*, *Aspect Ratio*, *IntSTD*, and *IntMax*, which are morphology parameters that are derived from particle images by the MFI microscope software. However, *IntMax* is not readily available for the FlowCam, so that the S-factor is redefined for FlowCam as the product of similar morphology parameters reported by the system's software: *Circularity (Hu)*, *Aspect Ratio*, *Sigma Intensity*, and *Intensity Sum*.

The S-factor filter described in (11) works by finding a cutoff value for classification that lies between the average S-factor values of the SO and NSO classes. Manual optimization is used to determine an optimal cut-off value for classification. This classification procedure is equivalent to creating a filter based on a logistic regression with the S-factor as the sole predictor variable (13).

As with the random forest approach, the logistic regression is fitted using S-factor values from the training data. For a new particle with unknown class, the logistic regression fit provides a percentage p , similar to the vote percentage that indicates how likely the particle belongs to the NSO class. K-fold cross-validation is also used to choose the optimal threshold value c for classification via the ROC curve as with the random forest: if $p > c$ then the particle is classified as NSO, otherwise it is classified as SO. Validation data is used to compute performance measures for the logistic regression filter based on the S-factor.

Counting Methods

We consider multiple counting methods in this paper. One obvious method (referred to as the *Classification* method) involves using a predictive filter, such as a random forest or logistic regression, to predict the classes of particles in a new test set and then counting the predicted classes. However, if

the overall prevalence in a test set is heavily skewed towards the majority SO class, NSO counts will be overestimated, even if the filter achieves a high degree of accuracy. This simple scheme gives accurate class counts only when prediction is perfect.

The problem of quantifying class counts using less than perfect predictive filters (a common occurrence for real-world applications) has received some attention recently, notably in (25,26). In this paper, we show how the *Mixture* method described by Forman (25) can be used to obtain more reliable counts of SO and NSO particles.

The following four counting methods are considered in this paper:

- *Classification*: a predictive filter (e.g., random forest or logistic regression) is used to predict the class membership (SO or NSO) of particles in a test sample; the number of NSO particles is the sum of the predicted NSO cases.
- *Mixture*: the distributions D_{NSO} and D_{SO} of vote percentages among the NSO and SO classes are estimated from the training set. Estimation of the vote distributions is accomplished during training of a predictive filter via K-fold cross-validation. The observed distribution of votes D_{Test} of a test set is regarded as a mixture of the distributions D_{NSO} and D_{SO} :

$$D_{Test} = \alpha D_{NSO} + (1 - \alpha) D_{SO}$$

The Probability-Probability plot method described in (25) is used to estimate the mixture parameter α . The estimate of α multiplied by the total number of particles in the test set is used as the estimate of the number of NSO particles in the test set.

- *Circularity* or *Circularity (Hu)*: Particles with *Circularity* or *Circularity (Hu)* < 0.95 are classified as NSO; the number of NSO particles is the sum of the predicted NSO cases

- *Aspect Ratio*: Particles with Aspect Ratio < 0.85 are classified as NSO; the number of NSO particles is the sum of the predicted NSO cases

All of the counting methods listed above are compared with respect to the true count of NSO in a given test sample.

Evaluation of Classification Accuracies and Counting Errors

The classification accuracy of a predictive filter was assessed by computing the confusion matrix on an independent validation set. The confusion matrix is a tabulation of the true and false predictions for each particle class (see Figure S1). Classification accuracy was determined as the percentage of true predictions for both SO and NSO particles (diagonal elements of the confusion matrix) while the percentage of misclassification was described by the percentage of false predictions for both SO and NSO particles (off-diagonal elements).

Counting errors were assessed by comparing the predicted number to the true values for both SO and NSO particles. The absolute error for counting is defined as:

$$\%error[(N)SO\ counting] = \left| \frac{(N)SO\ particles\ [predicted] - (N)SO\ particles\ [truth]}{(N)SO\ particles\ [truth]} \right|$$

RESULTS AND DISCUSSION

Artificially generated protein particles were prepared and spiked into formulations of mAb1, mAb2, and BSA at concentrations 20, 125 and 50 mg/mL, respectively. These different formulations were used to account for differences in refractive indices between protein particles and surrounding solution as well as different morphologies of protein particles. Typically this concentration range can change the refractive index of the solution from 1.33 up to >1.36 thereby changing the sensitivity of detection of translucent particles, which relies on the optical contrast between the particle and the solution (27). In particular, proteinaceous particles and SO droplets have a refractive index that is usually ≤ 1.4 thereby hampering reliable detection in higher concentration formulations (refractive index of 1.4046 in case of Dow Corning 360 Medical Fluid, 1000 cSt). In addition, differences in particle morphologies can have an impact on classification accuracy using current methods of classification.

In this study, we compare performance of the two most commonly used flow microscopes in biopharmaceutical industry, FlowCam VSII (color) and MFI 5200. The same samples from all model systems were measured on both instruments, allowing for a direct comparison of performance. The measured data sets were split into size bins of 1 μm for particles

larger than 1 μm up to 10 μm . In each size bin, a training set consisting of artificially generated data was used to create a predictive filter (80% of the data) and a validation set (20% of the data) was used to assess classification and counting accuracy. For all model systems we examined the following parameters in each size bin:

- Variable importance for classification (Fig. 3)
- Classification accuracy for SO and NSO particles (Fig. 5)
- Counting accuracy for SO and NSO particles (Fig. 7)

It should be pointed out here that there are several differences between the flow microscopes used in this study that can impact a fair comparison. The cameras used in this study are different, i.e. MFI uses a monochrome camera while the FlowCam has a color camera. Another difference is the used magnification, which is 5 \times for the MFI and 10 \times for the FlowCam. In addition, the FlowCam does not correct for refraction effects for smaller particles, which over-sizes them as shown in a recent publication (28). Finally, the FlowCam does allow for more user optimization of the binarization settings for data acquisition. Despite these differences we think that a comparison between both instruments is useful.

Morphology Parameters Obtained by Flow Microscopy

Flow microscopy takes images of individual particles and provides their morphology parameters. For our application, we used both FlowCam and MFI for flow imaging. The FlowCam is capable of reporting up to 35 different particle morphology parameters, while the MFI is capable of reporting 10 morphology parameters; these are grouped into three classes:

1. Basic shape parameters,
2. Advanced morphology parameters,
3. Gray scale and color measurements.

A listing of all the morphology parameters provided by the FlowCam and MFI instruments is given in Table S1, and detailed information on each parameter is available on each manufacturer's website. It should be noted that several of the morphology parameters are not independent of each other, e.g. compactness is inversely proportional to circularity.

SO droplets are spherically and can be distinguished from other particles that are not spherically. Previous studies have reported the use of single morphology parameters for classification of SO and NSO. For example, particles with *aspect ratio* ≥ 0.85 (ratio of width/length) are classified as SO. These single parameter filters work well for particles $\geq 10\ \mu\text{m}$, but have low accuracy for smaller sizes. The main reason for the poor performance is that the morphologies vary in importance for correct classification in the different size ranges; this is

partly due to an increase of measurement error in the morphologies of smaller particles (where the flow imaging software cannot easily determine whether a pixel of an image is part of a particle or part of the background). In addition, morphologies of protein particles are variable and depend on several factors such as stress or storage condition. These differences in morphologies are important to pick the best morphology parameters for classification. Thus, it is unlikely that only a couple of morphology parameters can be used together for accurate classification across all size ranges. In more recent work, a combination of multiple morphology parameters using regression models was applied to classify SO and NSO particles. In general it was found that, in addition to *aspect ratio* and *circularity* (Hu), the following morphology parameters are important for classification of SO:

- *IntMean* (intensity mean of all pixels of the particle),
- *IntSTD* (intensity standard deviation between higher and lower intensity values within a particle),
- *IntMin* (intensity of the darkest pixel of the particle), and
- *IntMax* (intensity of the brightest pixel of the particle).

Note that FlowCAM and MFI report different morphology parameters of a particle, e.g. FlowCam does not report intensity minimum and intensity maximum, which hampers transfer of regression models between instruments. The complexity of morphology parameters necessary for particle classification requires a thorough statistical evaluation about their importance in different size regimes and this is still lacking in the current literature. Examples of factors that will influence parameter importance for classification are magnification of the objective used, sample illumination intensity, camera type (monochrome *vs.* color), and flow cell dimensions. In general, regression-based filters require *a priori* knowledge about parameter importance because the parameters included in the model must be selected ahead of time and the relative importance of each parameter must be mathematically defined before analysis while random forest filters are able to rank parameter importance for classification without *a priori* knowledge and thus do not give predefined weight to each parameter (see below).

Artificially Generated Protein Particles and Silicone oil Droplets

Four model systems containing artificially generated protein particles and SO droplets are used to test classification and counting accuracy of the random forest filters. Representative particle images taken on both FlowCam and MFI 5200 are shown in Fig. 2. It can be seen that morphologies and particle opaqueness are different in each system. The mAb1 particles generated by stirring are the most opaque particles while the BSA particles generated by heat are the most translucent ones.

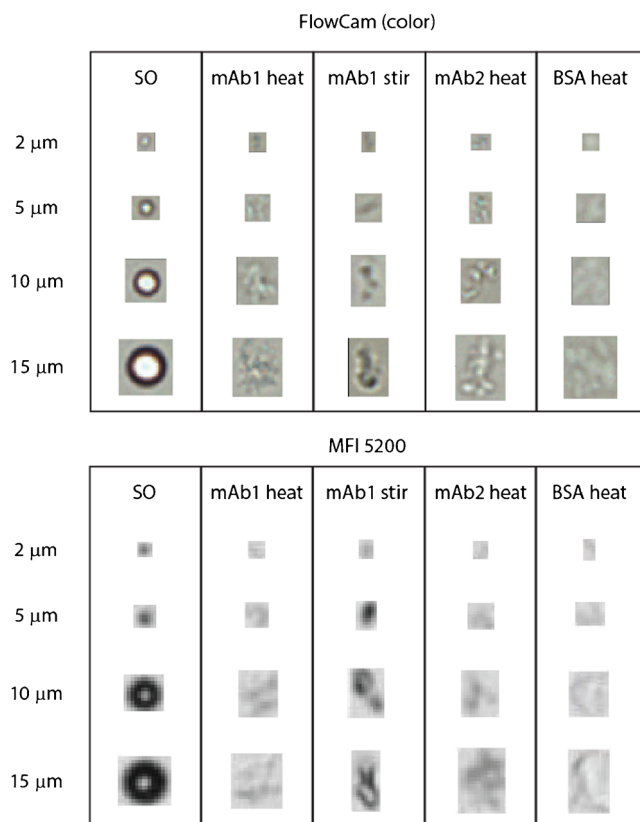


Fig. 2 FlowCam (color) and MFI images of representative particles found in the artificially generated samples.

mAb1 and mAb2 particles generated by heat have an intermediate translucency. Similar morphologies using comparable stress conditions have been reported previously (7,29). Qualitatively, the parameters in this study encompass the general appearance of the stressed particles shown in literature. Quantitatively, parameters such as grey scale parameters (e.g. transparency, edge gradient, sigma intensity, etc.) span a range for the particles investigated in this work.

The corresponding particle size distributions are shown in Fig. S2. These size distributions also demonstrate different cases of imbalanced data sets, e.g. the mAb1 (stir) sample has a lot of very small particles and very little large particles resulting in a significant imbalance of SO and NSO particles above 5 μm while the mAb2 (heat) sample has significantly more NSO particles above 10 μm. The BSA (heat), mAb1 (heat) and mAb2 (heat) are more balanced below 10 μm. These differences can significantly impact counting accuracy when the predictive filter is imperfect (see below).

Building a Random Forest Classification Filter

The random forest approach uses a training set that contains morphology information on particles with known identity. As with all supervised learning algorithms, the better the training set the more accurate the prediction, which is both the

strength and weakness of the methodology. Ideally, the goal is to build an accurate filter by using a representative (i.e. very large) and balanced dataset (comparable amounts of SO and NSO). We create filters using morphology information obtained from flow microscopy images of SO and NSO particles for each model system. In our work we compare MFI5200 and FlowCam (color). In general, the image quality of FlowCam is superior compared to the MFI 5200 (10), but a thorough comparison of filter performance between instruments is still lacking in the literature.

First, we created a database containing defined SO and NSO particles for each model system (two different mAbs as well as BSA). Each database contains morphology information of particles $\geq 1 \mu\text{m}$, and details for all size bins are shown in Tables S2 and S3.

A random forest filter was created for each size bin, using training particles from the corresponding size bin (80% of the particle data set). The accuracies of the classification and counting methods were evaluated using the remaining 20% of the particles as an independent validation set. It is

important to point out that accuracy should always be assessed on an independent data set that was not used to create the predictive filter. If we assess accuracy on the training set (initial 80% of data), we obtain $>99.9\%$ accuracy for classification and counting for all particles $\geq 1 \mu\text{m}$ in all model systems.

Importance of Morphology Parameters in Different Size Ranges

Each random forest filter is able to evaluate the importance of morphologies for accurate classification in each size bin. Importance for a given morphological parameter is defined as the mean decrease in accuracy of the random forest filter whenever that parameter is omitted from the training data.

The plots in Fig. 3 display the importance of each morphology parameter measured with the MFI by size bin for the four different model systems. It should be stressed here that variable importance varies by size bin, as well as by model system.

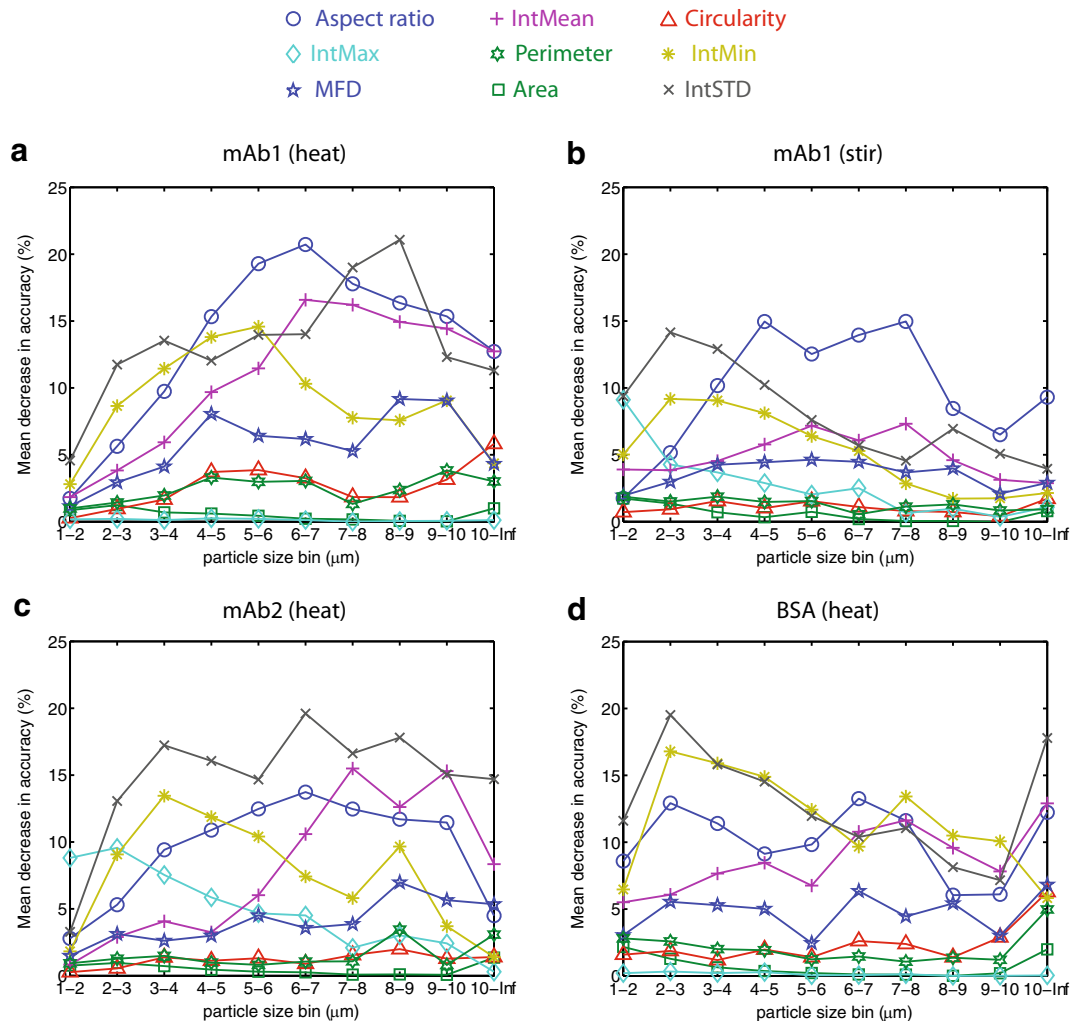


Fig. 3 Relative importance of different morphology parameters for correct classification by MFI using random forest filters in the different size ranges: (a) mAb1 (heat) (b) mAb1 (stir) (c) mAb2 (heat) and (d) BSA (heat).

For mAb1 (heat) particles in the larger size bins the top three parameters are *aspect ratio*, *IntSTD* and *IntMean*. When one of these parameters is omitted from the training data, accuracy decreases by at least 10% (Fig. 3a). At smaller particle size bins, *IntMin* becomes more important, while most morphologies decrease in importance. For the smallest size bin, no parameter has variable importance greater than 5%.

The situation is different for BSA (heat) particles where the same morphology parameters remain important throughout the size bins, indicating that the random forest filter will perform better under this model system. Comparing these findings with the other two model systems shows that different protein particle morphologies require different morphology filters for accurate classification due to the varying parameter across all size bins and model systems. One interesting observation for all systems is that circularity has very low variable importance ($\sim 5\%$), indicating that a very good random forest filter can be constructed with the remaining eight parameters.

The corresponding importance plots for data obtained with FlowCam (color) can be found in Figure S3. For clarity only the top five most important parameters are shown. The remaining 30 morphology parameters have very low variable importance (less than 5% in all size bins) indicating that they are not useful for classification. Similar to MFI, parameter importance varies by size bin and model system. However, a different set of parameters is important for classification under FlowCam. In addition to *aspect ratio* and *sigma intensity*, the new parameters include *edge gradient*, *circularity (Hu)*, *circle fit*, and *transparency*. Despite increased image quality with FlowCam, variable importance is much smaller for individual morphologies. This is an indication that morphologies under FlowCam are less differentially expressed between SO and NSO particles.

In general, the varying importance among morphology parameters emphasizes the strength of the random forest approach. The regression-based filters require *a priori* knowledge about parameter importance, while the random forest filters automatically determine parameter importance for classification.

Classification Accuracy of the Random Forest Filter

The random forest uses a particle's information from all morphologies, and outputs a new measure: the vote percentage for the NSO class. This can be regarded as a new composite feature that can be used to separate particles into SO or NSO. Three density histograms¹ of NSO vote percentages separated by the known particle class are displayed in Fig. 4. In the ideal case, the vote percentage histograms should be clearly separated, with the histogram for SO clustered at 0 and the one for NSO clustered at

¹ The density histogram is a smoothed version of the relative histogram such that the entire area of the histogram equals 1.

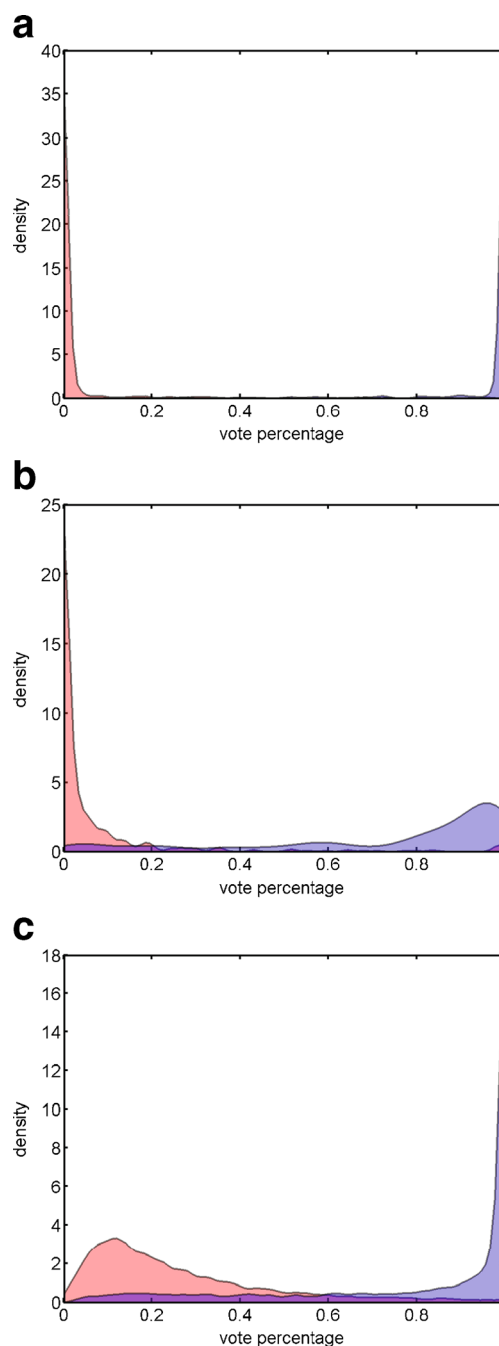


Fig. 4 Separation of the new observable “vote percentage” of SO (blue) and NSO (red) particles using the random forest filter based on BSA heat particles: **(a)** particles in largest size bin $\geq 10 \mu\text{m}$ (ideal separation), **(b)** particles in size bin $5\text{--}6 \mu\text{m}$ (little overlap), **(c)** particles in smallest size bin $1\text{--}2 \mu\text{m}$ (more overlap and misclassification).

1. This separation occurs for particles $\geq 10 \mu\text{m}$ and is indicative of a predictive filter with very high accuracy (Fig. 4a). In the smaller size bins, the separation is still good, but there is more overlap between the SO and NSO vote histograms; this overlap results in a higher misclassification rate. In the smallest size bin $1\text{--}2 \mu\text{m}$, this overlap is more pronounced, resulting in a misclassification rate of more than 10% in SO particles and nearly 20% in NSO particles (see below).

Confusion matrices on the independent validation sets are computed in order to evaluate the classification accuracies of the random forest filters (see [Materials and Methods](#) and [Fig. S1](#)). The classification accuracies for each model system using both MFI and FlowCam (color) per individual size bin are displayed in [Fig. 5](#). Classification accuracy is strongly dependent on protein particle morphology, as well as instrument type. For particles $>5\ \mu\text{m}$, MFI produces filters with classification accuracies $>90\%$ for both SO and NSO particles. Performance below $5\ \mu\text{m}$ depends strongly on protein particle morphologies, e.g., mAb1 (heat) classification accuracies are as low as 60% for SO and 68% for NSO in the smallest size bin, but classification accuracies of BSA (heat) remains fairly constant at around 88% for SO overall and down to 80% for NSO in the smallest size bin. This finding is interesting as the BSA (heat) sample has the most translucent NSO particles, which are easier to separate from SO particles than NSO particles in the other three model systems.

Random forest filters based on MFI outperform filters based on FlowCam (color) in every model system. At least for our model systems, MFI filters have accuracies $5\text{--}10\%$

higher compared to FlowCam (color) filters in most size bins, except for smaller particle sizes below $3\text{--}4\ \mu\text{m}$. For the three mAb systems analyzed on FlowCam there is a constant decrease in classification accuracy for particle sizes below $10\ \mu\text{m}$ in contrast to MFI classification, which remains fairly constant down to $5\ \mu\text{m}$. Interestingly, accuracies of FlowCam filters are fairly constant for BSA heat particles with $>90\%$ for SO particles and $>85\%$ for NSO particles across all size ranges. The difference in classification accuracy is most likely due to the use of the color camera, which is known to create edge artifacts that are more problematic for smaller particles as well as known over-sizing effects of smaller particles in the FlowCam (28).

Comparison to a Regression Model Containing Four Parameters

We compare the random forest filters to filters based on the logistic S-factor (see [Materials and Methods](#) for more details). The corresponding classification accuracies for all model

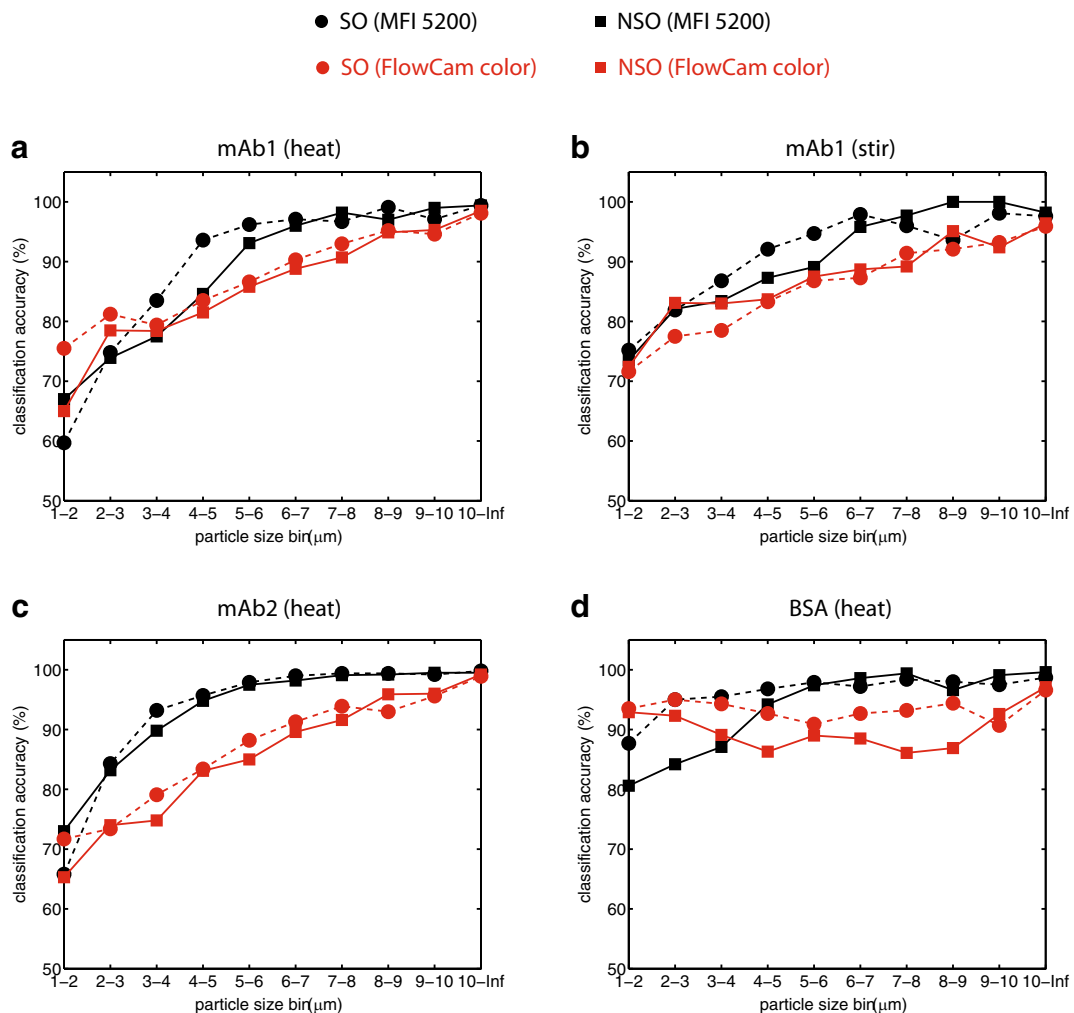


Fig. 5 Comparison of the random forest predictive filters to classify SO droplets and NSO particles across the four model systems: **(a)** mAb1 (heat) **(b)** mAb1 (stir) **(c)** mAb2 (heat) and **(d)** BSA (heat).

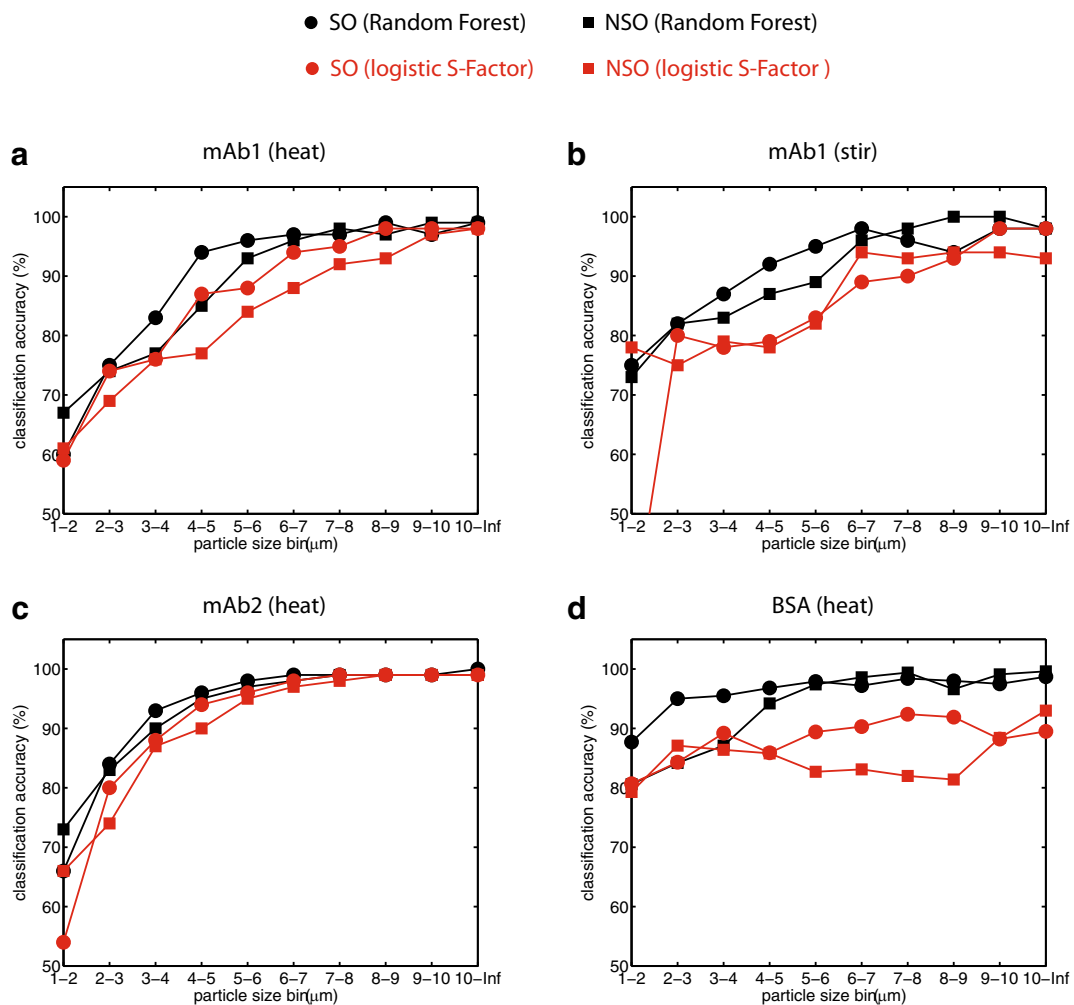


Fig. 6 Comparison of the random forest and logistic S-Factor filters based on MFI datasets: **(a)** mAb1 (heat) **(b)** mAb1 (stir) **(c)** mAb2 (heat) and **(d)** BSA (heat).

systems are shown in Fig. 6 (the original S-Factor plots are shown in Fig. S4). In general, the random forest fits outperform the logistic S-factor filters in every model system. The smallest difference in classification accuracy is for mAb 2 (heat), most likely due to the fact that the parameters *IntSTD*, *aspect ratio* and *IntMax* used in the S-factor belong to the five most important morphology parameters for classification in the random forest filter (see Fig. 3).

The largest difference in classification accuracy can be seen for the BSA (heat) particles, where it is consistently 5–15% higher for the random forest filters. This is again in agreement with parameter importance in Fig. 3 where importance remains fairly constant across all size ranges and *IntMax* is not important for classification at all (<1% decrease accuracy in all size bins). Also, in this system, there is a higher misclassification rate for NSO particles compared to SO particles.

Counting Accuracy

Classification and counting are two different, but related problems. For example, consider the two datasets in Fig. S1.

Assuming a 90% classification accuracy and a balanced dataset in this example results in misclassification of 10 particles as SO and 10 particles as NSO. Summing up the SO and NSO particles gives correct values for counting since the errors cancel each other out. PFS typically have highly unbalanced datasets such that the number of SO particles exceeds the number of NSO particles by a large margin (e.g. in Fig. S2b for particles $\geq 5 \mu\text{m}$). For the case with a larger prevalence of SO particles, the errors do not cancel out, resulting in significant over-counting of NSO particles (Fig. S1b).

Ultimately, flow microscopy is a counting assay, so that accurate counting is desired. To obtain accurate counts, we implemented the mixture method to account for unbalanced prevalences in the data (see **Materials and Methods** for details). This is illustrated in Fig. 7 where we compare counting errors of all filters for the four different model systems. For balanced datasets, such as BSA (heat) and mAb1 (heat), counting errors are below 10% for particle sizes down to 3 μm . Below this point, the single morphology filters (aspect ratio or circularity) result in larger counting errors. In contrast, the random forest filters coupled with the mixture method

- Random forest mixture method
- * Logistic S-Factor
- Random forest
- + Logistic S-Factor mixture method
- ◆ Circularity < 0.95 for NSO
- ▲ Aspect ratio < 0.85 for NSO

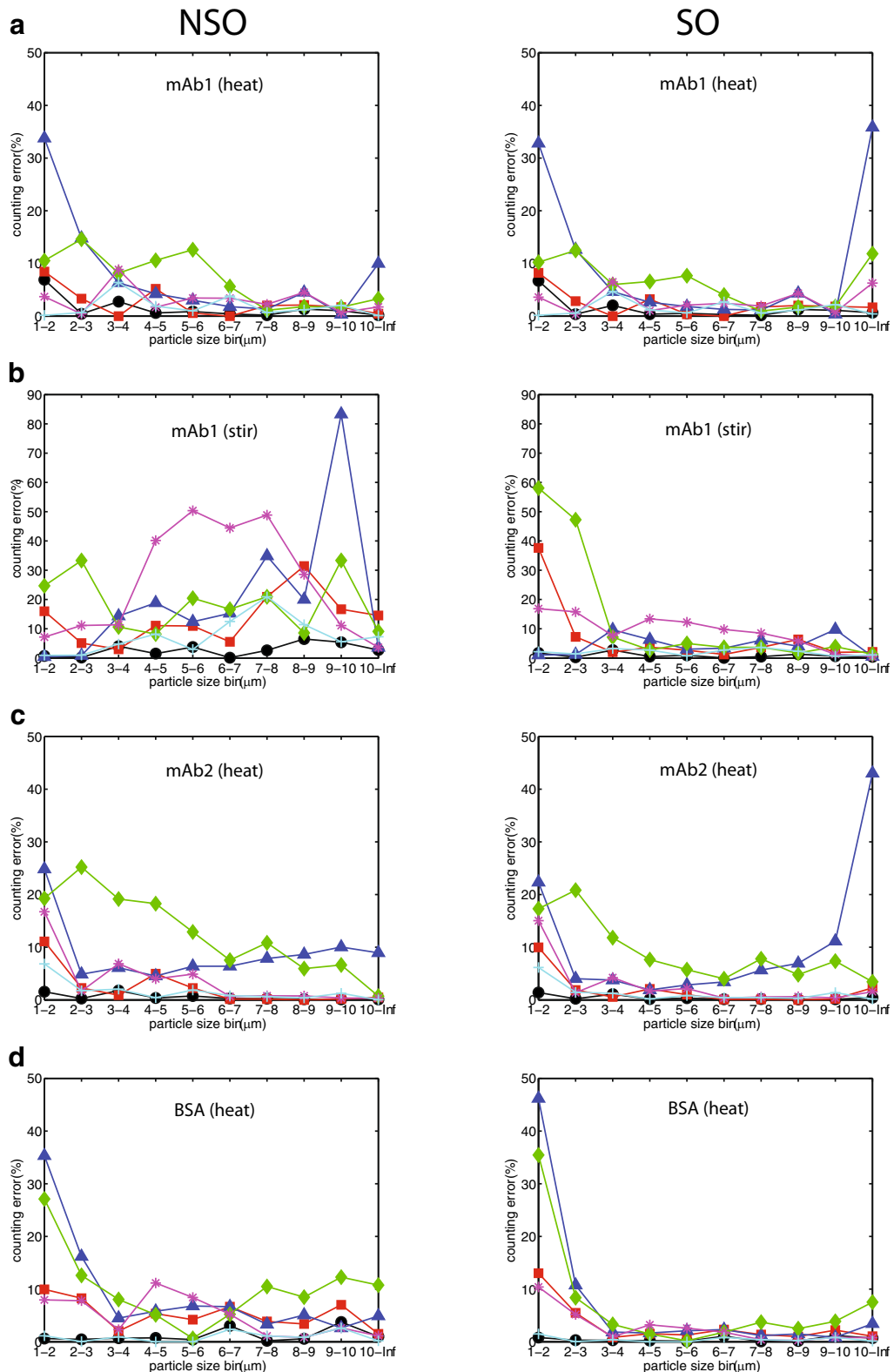


Fig. 7 Counting accuracy of the different predictive filters using the MFI datasets: (a) mAb1 (heat) (b) mAb1 (stir) (c) mAb2 (heat) and (d) BSA (heat).

have counting errors below 5% for all molecules and size bins except for the smallest size bin of mAb1 (heat).

Counting errors using the logistic S-factor are variable depending on the model system. For the balanced datasets of mAb1 (heat) and BSA (heat) error rates are below 10% in all size bins. For mAb1 (stir), the amount of SO particles is much higher than NSO particles above 5 μm , which leads to over-counting of NSO particles. While filters based on the logistic S-factor perform reasonably well for the three other model systems, they tend to over-count NSO particles the most, resulting in counting errors of up to 50% for particles $\geq 5 \mu\text{m}$.

In general, the mixture method can be used with any predictive filter and is not limited to the random forest. To demonstrate this we coupled the fits based on the logistic S-factor with mixture method. Applying the mixture method gives similar counting errors compared to the random forest filters in all model systems. In the case of mAb1 (stir), counting errors fall below 20%.

We analyzed counting accuracies described above (see Figure S5). Since classification accuracy is higher for MFI based filters, they will also be associated with higher counting accuracy. This is apparent in the unbalanced mAb1 (stir) sample under FlowCam, where counting errors for NSO particles can be nearly 200% under the Classification method, even for particles $>9 \mu\text{m}$.

In summary, the random forest filters combined with the mixture method had the smallest counting errors over all model systems (below 10% for the NSO particles) compared to all other classification and counting methods. Application of the mixture method for both the random forest filters and the logistic S-factor filters increase counting accuracies across all size bins.

While this work was done with stressed model systems, the classification and counting methodology demonstrated here can be applied to unstressed samples. It is recommended that the training set be created using real-world samples to ensure that representative particles are used for training. The use of real-world samples for training requires manual classification of the training set. Manual classification can typically be done consistently for particles $>5 \mu\text{m}$ by trained analysts, but may not be possible for particles smaller in size. This is a general limitation for real-world samples, which can have different particle morphologies than artificially generated particles.

CONCLUSION

We have developed a novel approach to classify and count SO and NSO particles in biopharmaceutical formulations. Our random forest method does not require *a priori* knowledge of parameter importance and achieves high classification accuracies, which was determined on particles from independent validation sets. We used morphology information of SO and

NSO particles in four different model systems for training of the statistical algorithm. This is done to account for variations in optical contrast as well as particle morphologies.

Accurate quantification of particle counts is performed using a mixture model method in which particles are classified using a random forest, which in turn is used to count particles in the NSO and SO classes. The mixture method can be combined with any prediction model to deliver high counting accuracies.

In our analysis we observed that classification accuracy of the MFI is higher than for the FlowCam (color) in all model systems. This is somewhat unexpected as image quality is superior and might be due to several reasons such as the color camera, which is known to cause fuzzy particle edges as well as software settings that can influence how efficient the binarization is. However, implementation of the mixture method in conjunction with the random forest model results in high counting accuracy for both MFI and FlowCam indicating that the mixture method can be used to improve counting with imperfect prediction models. A future study should evaluate performance of the FlowCam with monochrome camera, which was not available in our study.

In conclusion, our methodology is generally applicable to quantify counts of NSO particles in biopharmaceutical formulations. The methodology is not limited to quantification of just SO and NSO particles and can be used to classify and count different types of particles as long as their morphology parameter distributions are sufficiently different. A possible future application may be differentiation of different types of foreign particles and potentially stressed protein particles.

ACKNOWLEDGMENTS AND DISCLOSURES

The authors would like to acknowledge Greg Downing, Mark Hu and Thomas Scherer for providing samples and valuable discussions. Daniel Coleman and Barthelemy Demeule are acknowledged for helpful discussions and reviewing the manuscript.

REFERENCES

1. Singh SK, Afonina N, Awwad M, Bechtold-Peters K, Blue JT, Chou D, *et al.* An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. *J Pharm Sci.* 2010;99(8):3302–21.
2. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G, *et al.* Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. *J Pharm Sci.* 2009;98(4):1201–5.
3. Rosenberg A. Effects of protein aggregates: an immunologic perspective. *AAPS J.* 2006;8(3):E501–7.
4. Narhi LO, Jiang YJ, Cao S, Benedek K, Shnek D. A critical review of analytical methods for subvisible and visible particles. *Curr Pharm Biotechnol.* 2009;10(4):373–81.

5. Zölls S, Tantipolphan R, Wiggenhorn M, Winter G, Jiskoot W, Friess W, *et al.* Particles in therapeutic protein formulations, part 1: overview of analytical methods. *J Pharm Sci.* 2012;101(3):914–35.
6. Patel AR, Lau D, Liu J. Quantification and characterization of micrometer and submicrometer subvisible particles in protein therapeutics by use of a suspended microchannel resonator. *Anal Chem.* 2012;84(15):6833–40.
7. Weinbuch D, Zölls S, Wiggenhorn M, Friess W, Winter G, Jiskoot W, *et al.* Micro-flow imaging and resonant mass measurement (archimedes) – complementary methods to quantitatively differentiate protein particles and silicone oil droplets. *J Pharm Sci.* 2013;102(7):2152–65.
8. Sharma D, King D, Oma P, Merchant C. Micro-flow imaging: flow microscopy applied to sub-visible particulate analysis in protein formulations. *AAPS J.* 2010;12(3):455–64.
9. Demeule B, Messick S, Shire SJ, Liu J. Characterization of particles in protein solutions: reaching the limits of current technologies. *AAPS J.* 2010;12(4):708–15.
10. Zölls S, Weinbuch D, Wiggenhorn M, Winter G, Friess W, Jiskoot W, *et al.* Flow imaging microscopy for protein particle analysis—a comparative evaluation of four different analytical instruments. *AAPS J.* 2013;15(4):1200–11.
11. Strehl R, Rombach-Riegraf V, Diez M, Egodage K, Bluemel M, Jeschke M, *et al.* Discrimination between silicone oil droplets and protein aggregates in biopharmaceuticals: a novel multiparametric image filter for sub-visible particles in microflow imaging analysis. *Pharm Res.* 2012;29(2):594–602.
12. Huang CT, Sharma D, Oma P, Krishnamurthy R. Quantitation of protein particles in parenteral solutions using micro-flow imaging. *J Pharm Sci.* 2009;98(9):3058–71.
13. Kuhn M, Johnson K. *Applied predictive modeling*: Springer; 2013.
14. Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem and Luca Scrucca. *caret: Classification and Regression Training*. R package <http://CRAN.R-project.org/package=caret>. 2015.
15. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008;28(5):26.
16. Maimon O, Rokach L. *Data mining with decision trees: theory and applications*. USA: World Scientific Publishing; 2012.
17. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*: Taylor & Francis; 1984.
18. Dago KT, Luthringer R, Lengellé R, Rinaudo G, Macher JP. Statistical decision tree: a tool for studying pharmaco-EEG effects of CNS-active drugs. *Neuropsychobiology.* 1994;29(2):91–6.
19. Bowser-Chao D, Dzialo DL. Comparison of the use of binary decision trees and neural networks in top-quark detection. *Phys Rev D.* 1993;47(5):1900–5.
20. Salzberg S. Locating protein coding regions in human DNA using a decision tree algorithm. *J Comp Biol.* 1995;2(3):473–85.
21. Kokol P, Mernik M, Završnik J, Kancler K, Malčić I. Decision trees based on automatic learning and their use in cardiology. *J Med Syst.* 1994;18(4):201–6.
22. Falconer JA, Naughton BJ, Dunlop DD, Roth EJ, Strasser DC, Sinacore JM. Predicting stroke inpatient rehabilitation outcome using a classification tree approach. *Arch Phys Med Rehabil.* 1994;75(6):619–25.
23. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
24. Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest? In: Perner P, editor. *Machine learning and data mining in pattern recognition: 8th international conference, MLDM 2012, Berlin, Germany, July 13–20, 2012 proceedings*. Berlin: Springer Berlin Heidelberg; 2012. p. 154–68.
25. Forman G. Counting positives accurately despite inaccurate classification. *Machine Learning: ECML 2005*: Springer; 2005. p. 564–75.
26. Milli L, Monreale A, Rossetti G, Giannotti F, Pedreschi D, Sebastiani F, editors. *Quantification trees*. Data Mining (ICDM), 2013 I.E. 13th International Conference on; 2013: IEEE.
27. Zölls S, Gregoritz M, Tantipolphan R, Wiggenhorn M, Winter G, Friess W, *et al.* How subvisible particles become invisible—relevance of the refractive index for protein particle analysis. *J Pharm Sci.* 2013;102(5):1434–46.
28. Ripple D, Hu Z. Correcting the relative bias of light obscuration and flow imaging particle counters. *Pharm Res.* 2015;1–20.
29. Joubert MK, Luo Q, Nashed-Samuel Y, Wypych J, Narhi LO. Classification and characterization of therapeutic antibody aggregates. *J Biol Chem.* 2011;286(28):25118–33.