

Modelling Deformation of Microgels at Multifarious Temperature using Deep Generative Model

Reza Azad¹, Pia Lenben², Yiwei Jia¹, Martin Strauch¹, Dorit Merhof¹, and Dominik Woll²

¹ Institute of Imaging and Computer Vision, RWTH Aachen University, Germany,

² Institute of Physical Chemistry, RWTH Aachen University, Landoltweg 2, D-52056 Aachen, Germany

Abstract. **Keywords:** Deep generative · Microgel · fluorescence microscopy

1 TO DO

0. Github page (the first priority)

 1. Creat 3d visualization similar to professor Dominik paper

 2. Add pia and get in contact with her

 3. Figure 2 (Pia)

1-Please read the following paper and try to write something for the Introduction, we will then revise it with Lucas from chemical department:

 ”Deformation of Microgels at SolidLiquid Interfaces Visualized in Three-Dimension”, I sent you pdf file for this paper.

2- Write algorithm in a latex format, also try to write equation for calculating the density feature

Progress:

Sample selection algorithm finished

Density feature Finished.

3- From the following paper, please paraphrase the following sections.

A- section 2, 2 Flow-based generative model (until the end)

Done

B- paraphrase section 4 and try to connect it to the previous section

Finished

C- Add one paragraph from section 5 to direct the above presentation toward the point cloud

Finished

4- For the sample selection algorithm, please provide more discussion and sample figures to show its effectiveness

done

5- For the experimental result section, please write more information you achieved

so far in each related part

6- for the surface analysis part please write information regarding the convexhull and similarity between the real and the generated data
finished

2 Introduction

Microgels are three-dimensional networks of macromolecular polymers that are shrunk to the micron range colloidal system. They have complex internal structures and are soft objects that can adaptively expand and contract in size by changing their local conformation in extremely rapid response to changes in solvent quality and external stimuli. The unique properties of microgels are mainly related to their dynamic, permeable, and adaptive deformation properties in different solvents [1], enabling them a great potential for applications such as biotechnology, drug delivery, electronics, and sensor technology. Therefore, the study of the surface behavior of μG is becoming increasingly significant.

Different experimental techniques as well as computer simulation methods have emerged to observe the shape of μG at the interface. The super-resolution fluorescence microscopy (SRFM) method provides unprecedented nanoscale views and has been applied in materials science. Due to the advantage of non-invasive imaging by SRFM, the adaptive behavior and dynamic processes of microgels stimulated by changes in external conditions can also be easily captured. In addition, this powerful method can obtain three-dimensional information on microgels with resolutions down to tens of nanometers.

Three-Dimensional Super-Resolution Fluorescence Microscopy.

Temperture Modifications. The spreading behavior of

One figure here to show different temperture modeling
 details of the dataset we have.

3 Proposed Method

Flow-based generative models are regarded as a powerful method for unsupervised learning due to their effective data generation and precise log-likelihood evaluation. The existing normalizing flow [?] contains the reversible mapping from simple latent distributions (eg. isotropic Gaussian) to complex data distributions. The mapping process can be represented as following eq.1:

$$z = f_1^{-1} \circ f_2^{-1} \circ \dots \circ f_n^{-1}(x), \quad (1)$$

where $z = z^0$ and $x = z^n$ denote the simple low-level latent variable and the high-dimension data. Then, f_i refers to an invertible transformation from z^{i-1} to z^i ($z^i \in \mathbb{R}^D$ for $i = 0, \dots, n$). The log-likelihood $\log p_X(x)$ with regards to the latent variable z can be formulated by Eq. 2

$$\log p_X(x) = \log p_Z(z) - \sum_{i=1}^n \log \left| \det \left(\frac{\partial f_i}{\partial z^{i-1}} \right) \right|. \quad (2)$$

It indicates that obtaining the optimal evaluation metrics of the flow-based model requires simultaneous operations to optimizing f_i^{-1} and $\log \left| \det \left(\frac{\partial f_i}{\partial z^{i-1}} \right) \right|$. The trained model can generate complex data x by performing efficiently the following sampling steps on the lower-order latent variable z :

$$z \sim P_z z \quad (3)$$

$$x = f_n \circ f_{n-1} \circ \cdots \circ f_1(z). \quad (4)$$

Compared to the discrete normalizing flow model that employs limited architectures, the continuous normalizing flow (CNF) proposed by [?] can alleviate the restrictions of model selection, as the objective function can be calculated by performing Hutchinson's estimator [?]. Unlike the simple discrete latent variable z in the aforementioned method, the latent variable $z(t_0)$ is considered to vary with time:

$$z(t_0) = z(t_1) + \int_{t_1}^{t_0} f(z(t), t) dt \quad (5)$$

The variation of the log density follows the equation of instantaneous change of the variables.

$$\log p(z(t_1)) = \log p(z(t_0)) - \int_{t_0}^{t_1} \text{Tr} \left(\frac{\partial f(z(t), t)}{\partial z(t)} \right) dt, \quad (6)$$

where $f(z(t), t)$, $z(t_0)$ and $z(t_1)$ correspond to $\frac{dz(t)}{dt}$, z and x in the normalizing flow respectively.

Due to the limitation that the normalizing flows are only valid to generate the network output with the same dimensions as the input, [?] designed SoftFlow that enables generating realistic samples from the training on high-dimensional manifolds. They estimate a conditional distribution of the perturbed data to reduce the discrepancy between the dimensions of the data and the target latent variable. Specifically, they first select a random value c_i from the uniform distribution $\text{unif}[a, b]$ and sample a noise signal v_i from the distribution $N(0, \Sigma_i)$. The perturbed data x'_i is obtained from the summation of v_i and x_i . The addition of v_i makes x'_i not restricted to a low-dimensional manifold and enables networks to be trained properly on the data space of the entire dimensions. Compared to the Eq. 2, the objective function is expressed as the following equation:

$$\log p_X(x'_i | c_i) = \log p_Z(z) - \log \left| \det \left(\frac{\partial f(z | c_i)}{\partial z} \right) \right|, \quad (7)$$

where $f(\cdot | c_i)$ is the flow transformation from the latent variable z to x'_i . The flow networks can be trained to learn the dependencies between the shape of data distributions and the noise distributions.

$$z_{sp} \ p_Z(z) \quad (8)$$

$$x_{sp} = f(z_{sp} | c_{sp}) \quad (9)$$

According to Eq.8 and Eq.9, a realistic sample x_{sp} can be synthesized by reducing the value of c_{sp} . The following dynamics can further extend the method to any CNF:

$$\frac{dz(t)}{dt} = f(z(t), t, c_i), \quad z(t0) = z, \quad z(t1) = x_i + v_i \quad (10)$$

SoftFlow built on the previous flow-based models is proposed to mitigate the issue of input-output homogeneity by introducing high-dimensional noise variables that enable the network to estimate the distribution of perturbed data conditional on the noise parameters.

SoftPointFlow

Point clouds, which can be acquired by various range-scanning devices, such as LiDARs, have attracted attention in several research fields due to their sparse data types, the tractability of geometric operations, and the effective representation of ensemble details. PointFlow [?] is proposed to implement the task of point cloud generation. However, the performance of PointFlow [?] to generate fine-grained point cloud structure is constrained by the dimensional mismatch of normalizing flows in estimating the density on low-dimensional manifolds. SoftPointFlow is a proposed solution to address this issue by incorporating SoftFlow methods into PointFlow. The overall architecture of SoftPointFlow is composed of two normalizing flow networks, PriorFlow and DecoderFlow which model a two-level hierarchical distribution of shapes and points following the PointFlow training framework, as demonstrated in Fig.2. A point set X_{set} containing M points is first encoded by the same encoder as PointFlow to a latent variable S employing reparameterization [?]. The likelihood of S can be estimated by PriorFlow, a Glow-like architecture to learn the prior for S . Eq.11 presents the perturbed data x_i :

$$x'_i = x_i + v_i, \quad v_i \sim \mathcal{N}(0, c_i^2 I), \quad c_i \sim \text{unif}[a, b]. \quad (11)$$

The model adopts DecoderFlow which utilizes an autoregressive function for flow transformation (AR Layer) to compute the conditional likelihood given S and C for the perturbed point x_i . The final objective function is shown as Eq.12.

$$\begin{aligned} \mathcal{L}(X_{set}; \theta, \psi, \phi) &= \mathbb{E}_{q\phi(S|X_{set})} [\log p_\theta(X_{set}|S) + \log p_\psi(S) - \log q_\phi(S|X_{set})] \\ &\approx \mathbb{E}_{q\phi(S|X_{set})} \left[\sum_{i=1}^M \left(\mathbb{E}_{c_i \sim \text{unif}[a, b]} \left[\log p(z_i) - \log \left| \det \left(\frac{\partial g_\theta(z_i|S, c_i)}{\partial z_i} \right) \right| \right] \right) \right. \\ &\quad \left. + \log p(Z) - \log \left| \det \left(\frac{\partial f_\psi(Z)}{\partial Z} \right) \right| + H[q_\phi(S|X_{set})], \right] \end{aligned} \quad (12)$$

where $Z = f_\psi^{-1}(S)$, $z_i = g_\theta^{-1}(x'_i|S, c_i)$ and H denotes the entropy.

3.1 Sample Selection Criteria

Algorithm 1 The Sample Selection Algorithm

Input: $S_1 \dots S_N$ (All the samples of a temperature)
Output: $S_1 \dots S_N$ (Ranked Samples)

$SumCorr_1 \dots SumCorr_N$
(Sum of the correlation values between a sample and all the other samples)

```

function SELECT( $S[ ]$ )
    for  $i = 1$  to  $N$  do
         $SumCorr_i \leftarrow 0$ 
        for  $j = 1$  to  $N$  do
            Correlation matrix  $R_{(i,j)} \leftarrow S_i$  and  $S_j$ 
             $CorrValue_{(i,j)} \leftarrow R_{(i,j)}$   $\triangleright$  Extract the correlation value
             $SumCorr_i \leftarrow SumCorr_i + CorrValue_{(i,j)}$ 
        end for
    end for
    Rank the samples in descending way w.r.t.  $SumCorr_i$ 
    Select the top 10 samples.
end function
```

The key objection of the sample selection algorithm is to select the most representative samples at a specific temperature. As shown in Algorithm 1, the density distribution of each sample is first transferred to a feature vector. Based on that, we calculate the sum of the correlations between a sample and all other feature vectors of samples respectively. The algorithm selects the top ten samples with the highest sum of correlation with other samples for the training dataset of each temperature to improve the performance of the generator model.

At $21^\circ C$, the density of localizations of realistic data should be rather constant throughout the microgel, as shown in Fig.1. We can observe fewer white pixels for Sample 21C-16 and Sample 21C-11 in the center part which means that proper localizations were detected in the corresponding toroid in the central region. The density distributions of these two samples are consistent with the performance of the real data and therefore included in the training dataset. However, points of the Sample 21C-10 and Sample 21C-2 are more spread out in location with more white pixels.

Furthermore, the central localization density of real microgel point cloud data shows a gradually increasing trend from $21^\circ C$ to $35^\circ C$. The overall size of the microgels should decrease with respect to the temperatures. For example, Samples 35C-20 and 35C-17 demonstrate a more pronounced increase in density in the center compared to Sample 35C-18 and Sample 35C-2.

At $43^\circ C$ and $53^\circ C$, the entire realistic microgel should become further compact. The localization density in the core remains significantly higher than that in the shell. The selected samples (eg. 43C-3 43C-14 53C-10 53C-11) are more representative than the dropped samples (eg. 43C-10 43C-13 53C-6 53C-18).

3.2 Generative Model

- Generating samples with an arbitrary number of points.
- Predicting Microgel behavior in a given temperature.
- .

Figure proposed method here

4 Experimental Results

Implementation details

Experiment setting, low and high sample generation process, adding small noise in each iteration and conditioning with the tempreture value

4.1 Metrics

Defining a metrics to match the generated samples and the real samples

4.2 Quantitative Results

cumulative mass function figure here

4.3 Qualitative Results

Density figure here

Algorithm 2 The localization Density Calculation Algorithm

Input: $P_1 \dots P_N$ (All the localizations within a microgel sample)

Output: $Dens_{(1,1)} \dots Dens_{(D,H)}$
(Localization Densities at $D \cdot H$ toroids divided from the volume)

```

function DENSITY( $P[]$ )
    for  $d = 1$  to  $D$  do
        for  $h = 1$  to  $H$  do
             $Vol_{(d,h)} \leftarrow ((d \cdot d_{dist})^2 - ((d - 1) \cdot d_{dist})^2) \pi h_{dist}$ 
        end for
    end for
    for  $i = 1$  to  $N$  do
        if  $P_i$  is located in  $Toroid(d, h)$  then
             $Count_{(d,h)} \leftarrow Count_{(d,h)} + 1$ 
        else
            Exclude the localization  $P_i$  out of range.
        end if
    end for
     $Dens_{(d,h)} \leftarrow \frac{Count_{(d,h)}}{Vol_{(d,h)}}$ 
end function
```

Fig. ?? The microgel localization densities at a specific temperature has the corresponding feature representation. We qualitatively analyze the similarity between the generated microgel point cloud and the real microgel point cloud by calculating and comparing the localization densities of both, as shown in Alg.2.

We place the point cloud in a 3-dimensional cylindrical coordinate system and divide the z axis and distance from cylinder axis equidistantly to obtain the rectangles of the same area. Then we take each rectangle with the same area but different distances from the symmetry axis as a cross section of the toroid and calculate the volume of each toroid. Then we count the number of data points within each toroid and divide it by the volume of the corresponding toroid to obtain the density distribution information.

Fig. 4 illustrates the location density distributions of the real data and the generated data from low-high methods where the densities are plotted versus a relative z-position and the distance from the symmetry axis. As the temperature increases, we can observe a tendency for the sample to become progressively narrower in the horizontal direction. The core part located in the middle part of the microgel has more significant density values. This demonstrates the high similarity between the real and generated data. The consistency between the two rows indicates the promising performance of our low-high generation method.

4.4 Surface Analysis

We need to show the surface distribution with two method, Convex hull and sphere fitting algorithm

for the surface analysis part please write information regarding the convex hull and similarity between the real and the generated data

Convex Hull The convex hull of a point cloud is defined as the smallest convex set that contains all points within a specific confidence value. It can be applied to compare shape similarity between the real and the generated data.

As shown in Fig. 5, the convex hull keeps shrinking with increasing temperature, indicating that the point cloud of the generated data becomes more compact. The trend of size change of the generated samples is consistent with the trend of the real data. Furthermore, the length of convex vectors continues to decrease as the temperature rises, showing that points are getting closer to within the radius and have a higher density in the center part. The density variation of the generated data has a similar trend to the real data.

We also investigate the shape features of the original and generated point clouds in the XZ-Plane to comprehensively compare the similarity of the shape variation among elevated temperatures. Fig 6 demonstrates the convex hulls of the point clouds in the XZ plane at different temperatures. It can be observed that a more pronounced decrease in the length of convex vector in the sectors from $-45^{\circ}C$ to $45^{\circ}C$ and from $135^{\circ}C$ to $225^{\circ}C$.

Moreover, the overall shape gradually changes from a circle-like shape to an ellipse-like shape. The shape change of the XZ-Plane projection corresponds to the trend in the actual data.

5 Conclusion

References

1. Karg, M., Pich, A., Hellweg, T., Hoare, T., Lyon, L.A., Crassous, J., Suzuki, D., Gumerov, R.A., Schneider, S., Potemkin, I.I., et al.: Nanogels and microgels: From model colloids to applications, recent developments, and future trends. *Langmuir* **35**(19), 6231–6255 (2019)

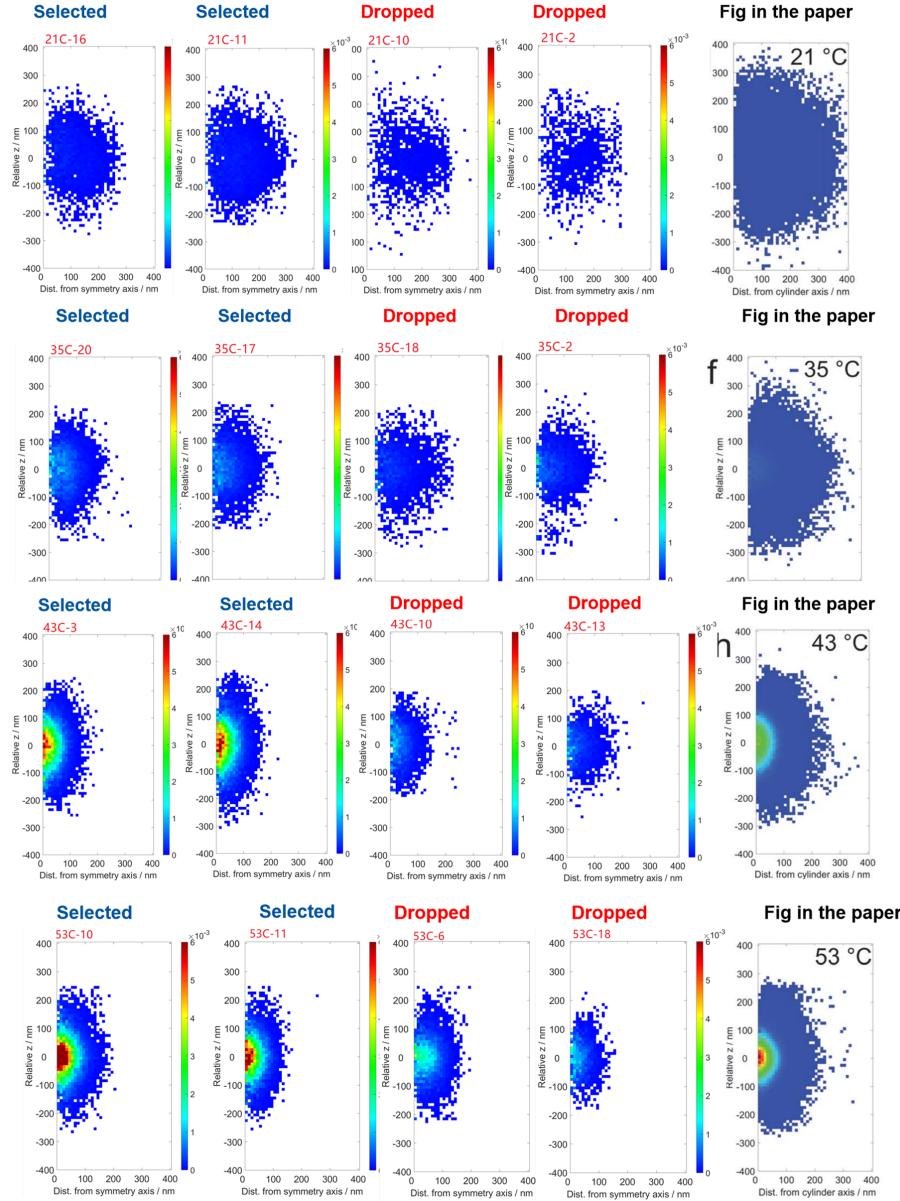


Fig. 1: Illustration of the density distributions for the selected or dropped samples at 4 temperatures: 21°C, 35°C, 43°C and 53°C. The "Selected" and "Dropped" denote the samples are added to or discarded from the training dataset respectively. We selected 4 temperatures from a total of 6 to better show the effectiveness of our sample selection algorithm.

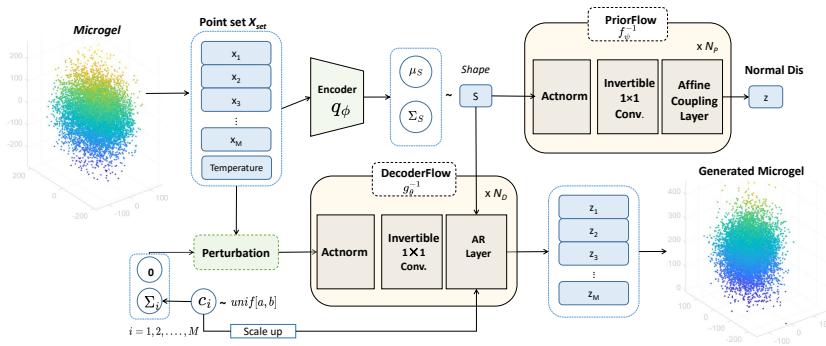


Fig. 2: General overview of the proposed deep generative model to mimic the Microgel behaviour on a giving temperature.

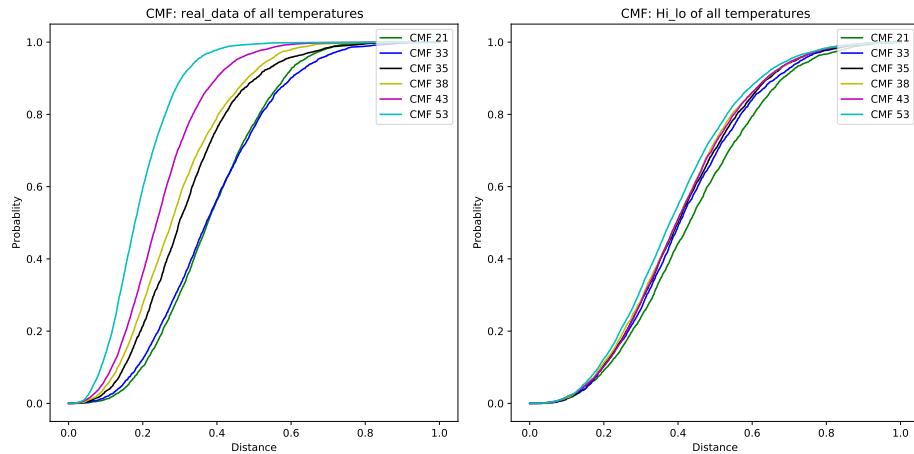


Fig. 3: Illustration of cmf calculated from the real data and the generated data from the low-high experiments.

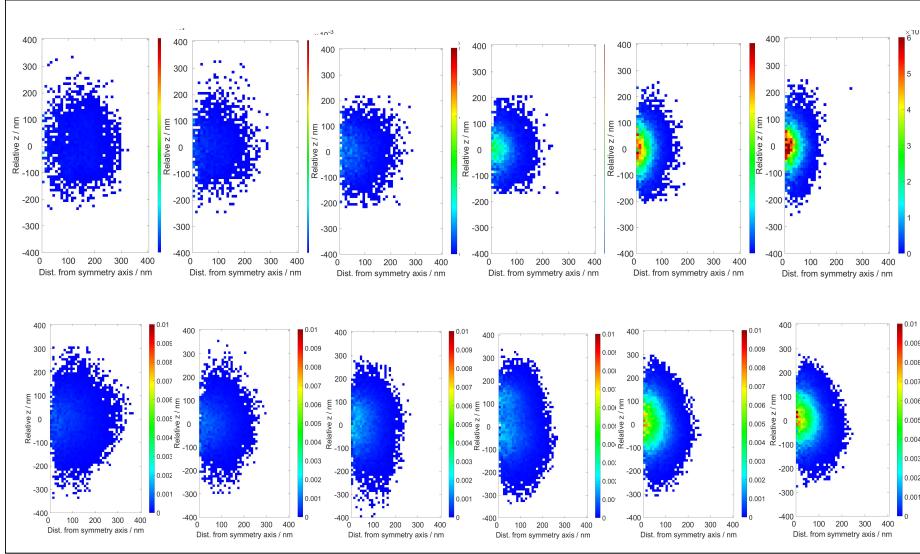


Fig. 4: The comparison of the density distributions between the real data and the generated data from low-high experiments. The upper row demonstrates the real data and the lower row presents the generated ones.

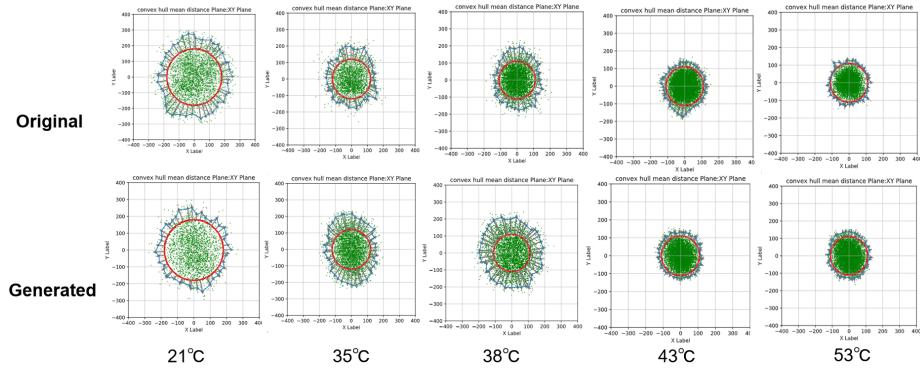


Fig. 5: Demonstration of convex hulls of samples in XY-plane. Red reference circles are added to the convex hulls for a more intuitive comparison of size. Most of the sample points are located in the circles. The radius at 21°C , 35°C , 38°C , 43°C , 53°C respectively: 180, 150, 120, 110, 110; The convex vectors denoted by the blue arrows spread from the radius to the point with a maximum distance within the 0.95 confidence value.

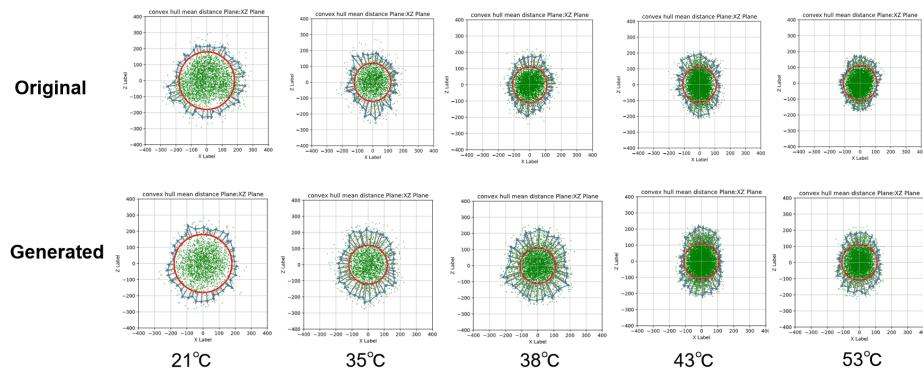


Fig. 6: Demonstration of convex hulls of samples in XZ-plane. Red reference circles are added to the convex hulls for a more intuitive comparison of size. Most of the sample points are located in the circles. The radius at 21°C , 35°C , 38°C , 43°C , 53°C respectively: 180, 150, 120, 110, 110; The convex vectors denoted by the blue arrows spread from the radius to the point with a maximum distance within the 0.95 confidence value.