RESEARCH ARTICLE | MAY 08 2023

On the statistical evaluation of bubbly flows using Voronoi cells grouped in clusters with fixed cell count

L. Weber ✉ ; S. Gabriel ; A. G. Class

Physics of Fluids 35, 053311 (2023)
https://doi.org/10.1063/5.0145551

Physics of Fluids ARTICLE pubs.aip.org/aip/phf

On the statistical evaluation of bubbly flows using Voronoi cells grouped in clusters with fixed cell count

Cite as: Phys. Fluids **35**, 053311 (2023); doi: 10.1063/5.0145551
Submitted: 6 February 2023 · Accepted: 21 April 2023 ·
Published Online: 8 May 2023

View Online Export Citation CrossMark

L. Weber,[a] S. Gabriel, and A. G. Class

AFFILIATIONS

Institute for Thermal Energy Technology and Safety (ITES), Department of Mechanical Engineering, Karlsruhe Institute of Technology (KIT), Hermann-Von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[a]**Author to whom correspondence should be addressed:** lorenz.weber@kit.edu. **Tel.:** +49 721 608 28305

ABSTRACT

The extraction of statistical information from bubbly flow experiments is crucial for numerical studies. Knowledge regarding probability distributions is particularly relevant in cases where a model relying solely on the use of mean values would lead to inaccurate results. As such, existing studies have focused on evaluating spatial distributions and local histograms for the void fraction, bubble density, and bubble size. However, the traditional box-counting method, employed by various studies, results in inconsistencies when deriving these quantities, especially when the respective two-phase flow features regions with low bubble densities or density gradients. This study demonstrates the application and benefits of combining Voronoi diagrams with a constrained K-Means clustering algorithm as a method for analyzing bubbly flows. We conduct three test cases: The first two cases use synthetic snapshots with prescribed characteristics to show the influence of evaluation settings and to critically quantify the errors, and the last test uses snapshot data of a plunging-jet experiment with air entrainment. We, then, compare the identified entrainment rate and the mean void-fraction distribution with empirical values from the literature. All three test cases show good agreement with the prescribed field characteristics (synthetic snapshots) and the data from the literature (experiment). Beyond demonstrating its applicability, we also show how this method can derive local histograms more consistently. The derivation is robust throughout the domain in comparison with traditional methods. For these reasons, we conclude that this method provides good estimates of spatial distributions.

I. INTRODUCTION

Bubbly two-phase flows occur during boiling, cavitation, or air entrainment on free surfaces, among other contexts. To examine the gaseous phase, various methods have been developed.[1,2] Obtaining local information about the gaseous phase (e.g., with needle probes[1]) can be inaccurate if the flow features low bubble counts. This study investigates these flows through images, or snapshots, of the bubble distributions, which can be obtained, for instance, from high-speed cameras[3,4] (see Sec. IV) or x-ray scans.[5] These snapshots contain information about the location and size of the bubbles. The objective of this study is to extract central information on the dispersed phase, specifically to evaluate spatial distributions of the mean and standard deviation of the bubble size, bubble density, and void fraction. Furthermore, we aim to approximate local probability density functions (PDF) through histograms. Despite the various physics behind bubbles and bubble formation processes, we seek a general-purpose tool for statistical bubble evaluations in this study.

Knowledge of the probability distribution is important in numerical simulations,[6–10] where a simpler model, for example, one that exclusively uses mean values, might result in unacceptable errors. This is the case for models containing non-linear terms, such as the Rayleigh–Plesset equation[11] for cavitating flows, where the stochastic field method[12] (SFM) addresses this issue.[7,10] In the inhomogeneous MUSIG model,[6] groups of different bubble sizes are used to model their non-linear behavior. Developing closure models, like the work of Liao *et al.*,[13] requires statistical insight into experiments for validation.

The standard image processing technique for deriving the above-mentioned quantities from particle or bubble distributions is the box-counting method,[14–17] which divides the domain into subdomains (called windows or boxes) and assigns each bubble/particle to a

subdomain based on its location. Depending on the experiment, one[18] or more[16,17] observation boxes are used to study the flow. In these studies,[16–18] approximations of the bubble size PDF are obtained. Another common application of the box-count method is the analysis of preferential concentration in particle-laden flows[14,15,19] to find particle swarms. Here, the PDF of the particle density (defined as particles per box volume)[14] is utilized. Researchers have also used Voronoi diagrams[20,21] to evaluate bubble-density PDFs[5,22] in order to find bubble swarms.

When particle or bubble-density PDFs are approximated through the box-counting method, only box-averaged quantities can be assessed. Bubble-density and void-fraction information associated with single bubble occurrences is lost as a result of this averaging. This loss of knowledge is problematic, as the bubble occurrence probability distribution is of interest.

Further difficulties arise when the flow features an inhomogeneous bubble distribution due to localized effects, such as cavitation, boiling on heated surfaces, or air entrainment at jet impact points. Consequently, some subdomains of the flow may feature high bubble densities, while others contain almost none. As such, *a priori* knowledge of the relevant inter-bubble distances in the flow is necessary for defining suitable box sizes.[5,14] This is a key drawback of the box-count method.

An algorithm, which can be used to automatically decompose the flow domain, is the Voronoi diagram,[20,21] where influence regions (also called bubble-cells) are calculated around individual bubbles. Using a "cellular" model for multi-phase flows was first initiated in the early 1950s by Simha,[23] who proposed an interaction-with-nearest-neighbors model to calculate the viscosity of suspensions. In the subsequent decades, models with polyhedral shaped cells were adapted by various researchers[24,25] and applied to study dispersed flow situations. The term "Voronoi diagram" appeared later, in evaluation methods of both particle-laden[15,26–28] and bubbly[5,22] flows. One can refer to Okabe[29] for a comprehensive review of the applications of this method.

Per the definition of our approach, each Voronoi influence region contains exactly one bubble.[29] As a result, the sample size for calculating a local mean value on a single image is small (i.e., one). Grouping bubbles improves the significance of the calculated mean value. This also enables approximating local, instantaneous probability density functions of bubbles occurrences even on single snapshots (e.g., relevant if very little data are available or for transient flows). A schematic example application is depicted in Fig. 1. A similar approach is used in PDF methods for numerical simulations of reacting flows,[30] where the mean value for every point is calculated from approximately the same sample size. Our study adopts this approach by implementing a constrained K-Means clustering algorithm,[31] which places the samples (bubbles) in groups of similar size. The K-Means clustering algorithm, where K refers to the number of groups, is a simple algorithm that is commonly used in the field of data analysis[31] to create compact groups of samples. However, the constrained K-Means clustering algorithm,[31] which allows for the cluster size to be restricted, has not been applied to bubbly flows, thus far.

Combining Voronoi diagrams with constrained K-Means clustering is a promising method for evaluating bubbly flows. This method avoids the above-mentioned drawbacks of the traditional box-count method and comes with the benefits of a clear definition of the bubble influence region and clustering in groups with constrained sizes.

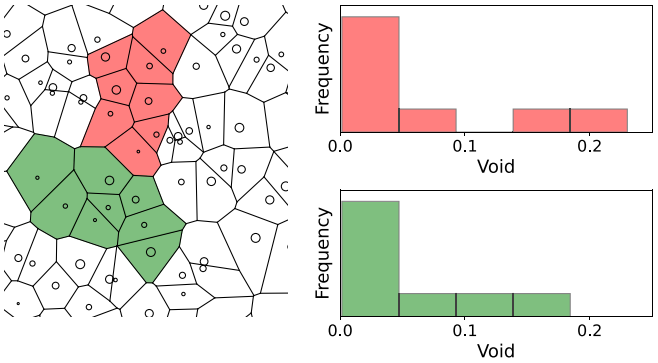

FIG. 1. Grouped Voronoi cells for a schematic bubble distribution (left) and derived approximations of local histograms (right).

This study involves three test cases to confirm the proposed method's applicability and potential benefits. The first two test cases involve generating synthetic snapshots, using prescribed equations for the bubble distributing process that aim to imitate features of a real bubble distribution. Errors can be quantified by comparing the results with the analytical characteristics underlying the true distribution process. In the third test, we apply the method to a plunging-jet experiment, which has been previously explored experimentally[32–38] and numerically.[3,39–42] In this analysis, we take existing snapshots from an earlier experiment.[4,43]

For the synthetic snapshots, we measure the quality of the results by comparing them to the analytical field. We assess deviations between the observed and expected values for the bubble size, bubble density, and void fraction. In addition, we quantify and discuss the achieved accuracy by assessing the magnitude of the systematic and statistical errors. For the plunging-jet experiment, we study the void-fraction distribution by comparing them with other studies.[3] In addition, we compare the calculated air entrainment rate with empirical correlations[33,35–37,44] and the descending bubble cone angle with other experiments.[34]

II. METHOD

Section I showed some of the shortcomings of applying the box-counting approach to bubbly flows. To overcome them, we propose an advanced method for evaluating instantaneous snapshot data. By applying ordinary Voronoi diagrams,[20,21] combined with a constrained K-Means clustering algorithm,[31] the shortcomings of the box-counting approach (see Sec. I) are addressed. The following provides a detailed description of the proposed method, including its underlying assumptions.

A. Assumptions

The following assumptions are made in the bubble evaluation process:

(a) Throughout this work, the shape of the bubbles is assumed to be spherical. Therefore, bubble sizes are characterized by their equivalent radii r. This parameter can be extracted from, e.g., shadowgraphs[4,18,43] by measuring the spatial expansion on the image.[16–18]

(b) Bubbly flows are three-dimensional phenomena; however, images of these flows are projections onto two-dimensional planes.[3,4,16,17] Analyses of the flows are conducted in two dimensions for (statistically) rotational symmetric cases or cases with one (statistically) homogeneous direction. A discussion regarding the applicability of this method for rotational symmetric flows can be found in the Appendix.

(c) For high-speed camera images taken in two dimensions, we assume that all bubbles are detected and exclude overlapping bubbles. This holds for snapshots of flows with low gas content.[3,43] Depending on the gas content, the bubble-capturing method must be adjusted accordingly. Note that overlapping bubbles affect neither the Voronoi diagrams nor the clustering algorithm, thus preserving the method's robustness.

(d) This work assumes a statistically steady flow. Flows with time-dependent mean flow characteristics are outside the scope of this paper, and therefore, are excluded. Note that the proposed method itself, however, is not limited to steady flows. For an extension to transient states, time-averaging operations need to be replaced by ensemble-averaging over repeated experiments.

(e) The images of the flow are not correlated; the time between the snapshots is long enough to eliminate dependence.[4,43] As such, bubble tracking and bubble trajectories are not considered in this study.

(f) Discontinuities in the density at the boundary between the gaseous and liquid phase are not considered. The gas content for each bubble is averaged over its influence region [see definition in Eq. (3)] with the volume V, and expressed by the void fraction α. The value α can be interpreted as the probability that a point in this region is occupied by the gaseous phase and is calculated as

$$\alpha = \frac{4}{3}\pi r^3 V^{-1}. \tag{1}$$

The local bubble density ρ is obtained from the inverse of the individual bubble-cell volume[5] V,

$$\rho = V^{-1}. \tag{2}$$

(g) The spatial extent of bubble influence regions that intersect with the boundaries is cropped by the flow domain. By trimming the regions, we ensure that their volumes are finite.[5] Otherwise, Voronoi diagrams would allow for regions of infinite size.[29]

(h) When constructing influence regions, we only take the location of the bubble center into consideration. The spatial extent of the bubbles is ignored for cell construction.[5]

B. Workflow

The method for evaluating bubbly flows can be divided into five steps: (1) pre-processing, (2) snapshot superposition, (3) Voronoi Diagram, (4) constrained K-Means clustering, and (5) post-processing. The workflow diagram is shown in Fig. 2. The first four steps are repeated for every snapshot.

In contrast to other methods,[5,14,16,18] we jointly apply steps two to four. We discuss these steps in detail in Secs. II C–II E.

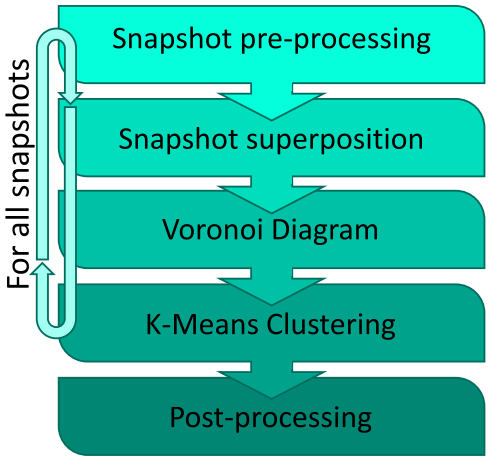

FIG. 2. Workflow of the method.

In the pre-processing step, raw measurement data (e.g., images) are processed to extract the location and size of each bubble that is, then, used in the subsequent steps. This is an important step for ensuring high data quality; however, it is not in the scope of this study, as it strongly depends on the experimental capturing technique.

In the post-processing step, the data from each snapshot are concatenated and visualized.

C. Snapshot superposition

In experimental data, the bubble count per snapshot is naturally fluctuating. Figure 3(a) depicts these large variations seen in a plunging-jet experiment.[4,43] The captured distribution reveals many snapshots with approximately ten bubbles, but also some with more than 80 bubbles. The captured distribution also demonstrates some of the challenges that arise from using the experimental data, which are given as follows:

- A fluctuating bubble count implies a varying bubble density. This influences the local spatial uncertainty in the domain since it is small for a high bubble density (locally more bubbles nearby) and high for a low bubble density (next bubble might be far away). As such, the spatial uncertainty of the captured snapshots can significantly fluctuate.
- In the case of a constant low bubble count, single snapshots can contain too little information to assess local effects and spatial distributions. In these cases, the spatial uncertainty in each snapshot may be too high for further evaluation.

To address these challenges, a snapshot superposition step is introduced before steps three and four (see Fig. 2), to obtain a constant, and sufficiently high, spatial resolution. This crucial procedure, needed for experimental data, is depicted in Fig. 3(b). Several snapshots (first row) are superpositioned (No. 1 + No. 2 and No. 3 + No. 4 in this example) in bubble sets (second row). From a number of snapshots S, we obtain S^* bubble sets (quantities marked with an asterisk refer to bubble sets in Secs. III–VI) after the superposition. Now, the second row has a higher bubble density (lower spatial uncertainty)

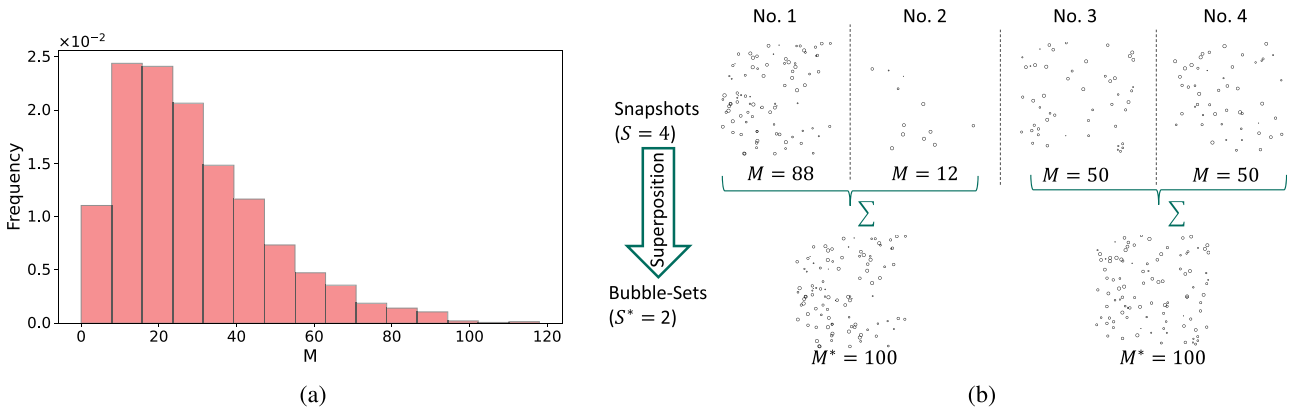

FIG. 3. (a) Statistical distribution of the bubble count per snapshot for a circular plunging-jet experiment.[4,43] (b) Schematic visualization of the snapshot superposition step, where fluctuations in the bubble count per snapshot result in varying spatial resolution of bubble data (first row). Superposition balances these fluctuations and combines them into bubble sets (second row).

with smaller fluctuations (ideally constant number of M^* bubbles per bubble set). In practice, we combine snapshots until a prescribed minimum bubble count M^*_{min} in each bubble set is reached. Note that quantities involving the influence region volume V (i.e., bubble density ρ and void fraction α) need to be corrected by the number of images that were summed in this process step. In case I (Sec. III B), we discuss this step in detail and show its validity by comparing with analytical data.

D. Voronoi diagram

With Voronoi diagrams, an influence region for every bubble in a bubble set can be calculated. This decomposes the domain into convex polyhedrons in n-dimensional space, free from overlaps or gaps.[20,21,29] For M^* bubbles with the bubble centers $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{M^*}$, the mth influence region A_m for bubble center $\mathbf{x}_m$ is defined[29] as

$$A_m = \{\mathbf{x} \,|\, \|\mathbf{x} - \mathbf{x}_m\| \leq \|\mathbf{x} - \mathbf{x}_j\| \text{ for } j \neq m, j \in \{1, 2, \ldots, M^*\}\}. \quad (3)$$

This definition of the ordinary Voronoi diagram assigns to each point the nearest bubble center based on the Euclidean distance. The resulting diagram for a random bubble distribution can be found in Fig. 4.

E. Constrained K-Means clustering

In this procedure, K cluster centers are created in the flow domain. Each bubble in a bubble set is assigned to the closest cluster center. With an iterative scheme, the original K-Means clustering algorithm minimizes the squared Euclidean distance of the samples to the cluster centroid. We employ the constrained extension of this algorithm, which allows one to set a minimum and maximum number ($C_{\min}$ and $C_{\max}$) of bubbles per cluster.[31] We set a target cluster size C_{trgt} and select one of $C_{\min} = C_{\text{trgt}}$, $C_{\max} = C_{\text{trgt}} + 1$ OR $C_{\min} = C_{\text{trgt}} - 1$, $C_{\max} = C_{\text{trgt}}$ OR $C_{\text{trgt}} = C_{\min} = C_{\max}$ for each bubble set, depending on which setting results in more clusters of size C_{trgt}. Figure 5 displays an example application of this clustering algorithm for a bubble distribution in a bounded box. Here, the bubbles are grouped in clusters with a fixed size ($C_{\text{trgt}} = C_{\min} = C_{\max} = 8$) and are colored according to their membership. By applying this algorithm, all M^* bubbles are assigned to exactly one of the K clusters. We introduce function f^{CKM}, which refers to this assignment,

$$f^{CKM} : \{1, 2, \ldots, M^*\} \rightarrow \{1, 2, \ldots, K\}. \quad (4)$$

The region A_i^C of the i-th cluster (where $i \in \{1, 2, \ldots, K\}$) is the union of a constrained amount of Voronoi regions A_m,

$$A_i^C = \bigcup_{m=1}^{M^*} \begin{cases} A_m & \text{if } f^{CKM}(m) = i, \\ \emptyset & \text{else.} \end{cases} \quad (5)$$

For an arbitrary volume-averaged quantity ϕ (e.g., α or ρ), a weight w_m is used for calculating its cluster-averaged value $\bar{\phi}$. This weight is interpreted as a contribution of the mth bubble to a point $\mathbf{x} \in A_i^C$. This weight is calculated from the bubble cell volume V_m divided by the ith cluster volume V_i^C. The weight of the mth bubble quantity ϕ_m becomes

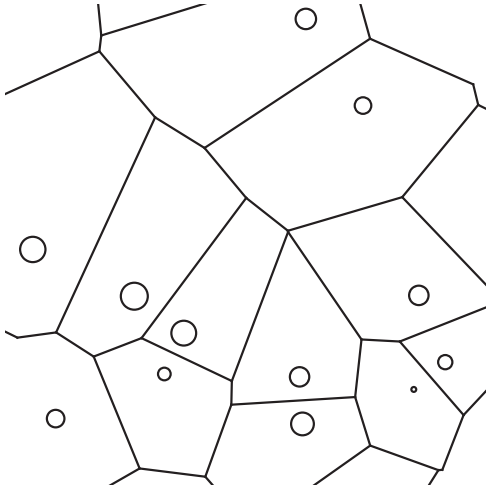

FIG. 4. Example of a Voronoi diagram for the calculation of the influence region of each bubble based on the nearest bubble center.

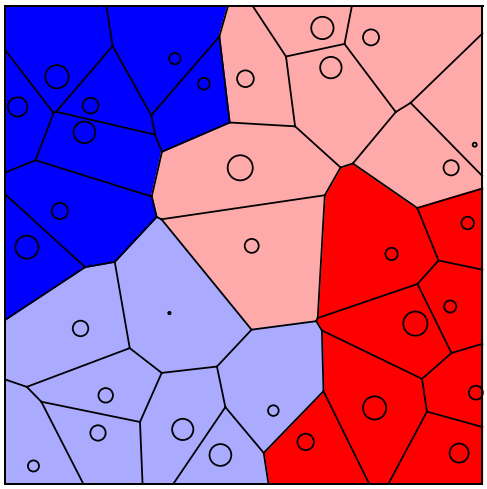

FIG. 5. 32 randomly distributed bubbles in a bounded box clustered into four clusters of exactly eight bubbles using the constrained K-Means clustering algorithm.[31]

$$w_m = \begin{cases} \dfrac{V_m}{V_i^C} & \text{if } f^{CKM}(m) = i, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The cluster-averaged value $\bar{\phi}$ is, then, calculated from

$$\bar{\phi} = \sum_{m=1}^{M^*} w_m \phi_m\,, \quad \forall \mathbf{x} \in A_i^C. \tag{7}$$

At a point $\mathbf{x}$, we obtain one cluster-averaged value $\bar{\phi}_t$ for each bubble set, where $t \in \{1, 2, \ldots, S^*\}$. The time-averaged (bubble-set-averaged) value $\langle\bar{\phi}\rangle$ is

$$\langle\bar{\phi}\rangle = \frac{1}{S^*}\sum_{t=1}^{S^*} \bar{\phi}_t. \tag{8}$$

After processing S^* bubble-sets, we assign a number of bubbles, $S^* \times (C_{trgt} - 1)$ to $S^* \times (C_{trgt} + 1)$, to each point $\mathbf{x}$. In Secs. III–VI of the paper, an over-bar ($\overline{\cdots}$) indicates a cluster-averaged value, and a value in angular brackets ($\langle\cdots\rangle$) refers to a bubble-set-averaged (time-averaged) value. We emphasize that all results of the proposed method represent volume-averaged data, due to the spatial expansion of both the clusters and the bubble influence regions.

The void fraction α [see Eq. (1)] depends strongly on the definition of the influence region [i.e., Eq. (3)]. Therefore, we introduce an alternative void fraction definition $\tilde{\alpha}$ that associates a cluster-averaged volume (cluster volume divided by cluster size) to each bubble. This equation for the m-th bubble reads

$$\tilde{\alpha}_m = \frac{4}{3}\pi r_m^3 \frac{1}{V_i^C}\sum_{m=1}^{M^*} \begin{cases} 1 & \text{if } f^{CKM}(m) = i, \\ 0 & \text{else,} \end{cases} \quad \forall A_m \in A_i^C. \tag{9}$$

In the latter equation, the effect of the choice of the influence region definition decreases with increasing C_{trgt}. This is in contrast to Eq. (1) and clearly shows the relation of $\tilde{\alpha}$ to the box-counting definition (if the box and the cluster contain the same amount of bubbles).

III. CASE I: SYNTHETIC BUBBLE DATA

We use synthetic bubble data for the first stage of test cases, as it is possible to identify the exact statistical characteristics of such bubble distributions. The bubble count M is a constant for every generated snapshot. This simplification is used to test the superposition step of Sec. II C (refer to Fig. 2). Different settings are applied, and the results are compared to prescribed values.

A. Description

A two-dimensional bubble distribution with depth H is considered. Domain length H is chosen to be the characteristic length scale. For each generated snapshot, M bubbles are distributed in the domain. The distribution process is governed by a prescribed bubble-density expectation $\mu_\rho(x, y)$. We choose μ_ρ to have the spatial distribution,

$$\begin{aligned} \mu_\rho(x,y) &= \mu_M f_1(x) \times f_2(y) \\ &= \mu_M \left[\frac{192}{71}\frac{y}{H} - \frac{100}{71}\left(\frac{y}{H}\right)^3\right] \\ &\quad \times \left[\frac{6.0162}{\sqrt{2\pi}}\exp\left(-\frac{9}{2}\left(2\frac{x}{H} - 1\right)^2\right)\right]. \end{aligned} \tag{10}$$

In Eq. (10), μ_M represents the expectation of the bubble count M per snapshot. The distribution of Eq. (10) is shown in Fig. 6.

The mth bubble is distributed in the domain space using the inverse transform sampling method.[45] We define the functions $F_1(x)$ and $F_2(y)$, where

$$\frac{dF_1(x)}{dx} = f_1(x), \quad \frac{dF_2(y)}{dy} = f_2(y), \quad F_1(0) = F_2(0) = 0,$$

and

$$F_1(1) = F_2(1) = 1. \tag{11}$$

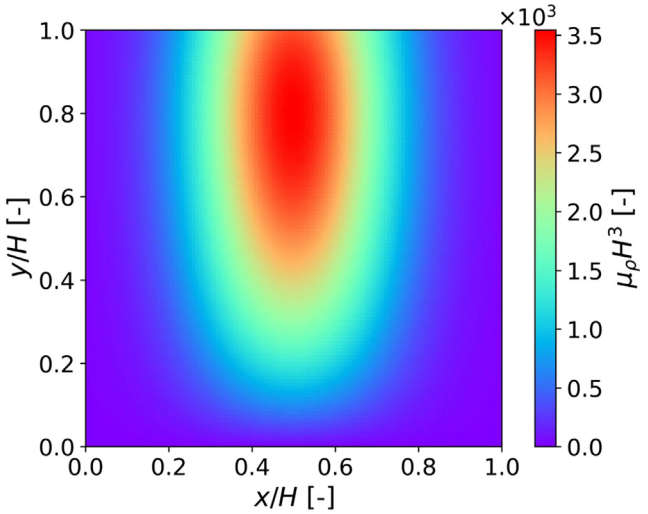

FIG. 6. Plot of $\mu_\rho(x, y)$ from Eq. (10).

The functions $F_1(x)$ and $F_2(y)$ are invertible. Their inverse is subsequently referred to as F_1^{-1} and F_2^{-1}. Now an uniform continuous distribution in the interval $[0, 1]$ is used. For the mth bubble, we draw independently for the variables $\zeta_{1,m}$ and $\zeta_{2,m}$ a sample from that distribution. Location $\mathbf{x}_m$ is obtained from

$$\mathbf{x}_m = \begin{pmatrix} F_1^{-1}(\zeta_{1,m}) \\ F_2^{-1}(\zeta_{2,m}) \end{pmatrix}. \tag{12}$$

At location $\mathbf{x}_m = \begin{pmatrix} x_m \\ y_m \end{pmatrix}$, the prescribed properties $\mu_r(x, y)$ and $\sigma_r(x, y)$ are taken into account when generating the bubble radius r_m. The expectation $\mu_r(x, y)$ and standard deviation $\sigma_r(x, y)$ of the bubble radius are given by

$$\mu_r(x, y) = 0.006 + 0.002\left(\frac{y}{H} - 1\right) \tag{13}$$

and

$$\sigma_r(x, y) = 0.000\,95\sqrt{\left(\frac{x}{H} - \frac{1}{2}\right)^2 + \left(\frac{y}{H} - 1\right)^2}, \tag{14}$$

respectively. Although the distribution of $\mu_r(x, y)$, $\sigma_r(x, y)$, and $\mu_\rho(x, y)$ is arbitrary, they aim to simulate features of a real bubbly flow. The standard deviation prescribed by $\sigma_r(x, y)$ could be interpreted as an increase in the bubble size variety due to coalescence, and breakup, away from an imaginary entrainment point, as seen in Fig. 8. The plot of $\mu_r(x, y)$ is depicted in Fig. 7. The expected bubble radius decreases with decreasing y/H in the used artificial data; this could be the result of buoyancy effects when gravity acts in the negative y-direction. Note that various interpretations regarding the physical meaning of Figs. 6–8 are possible and that these distributions do not aim to be in accordance with the experimental results.

In this test case, the generation of r_m is governed by a Maxwell distribution PDF,[46]

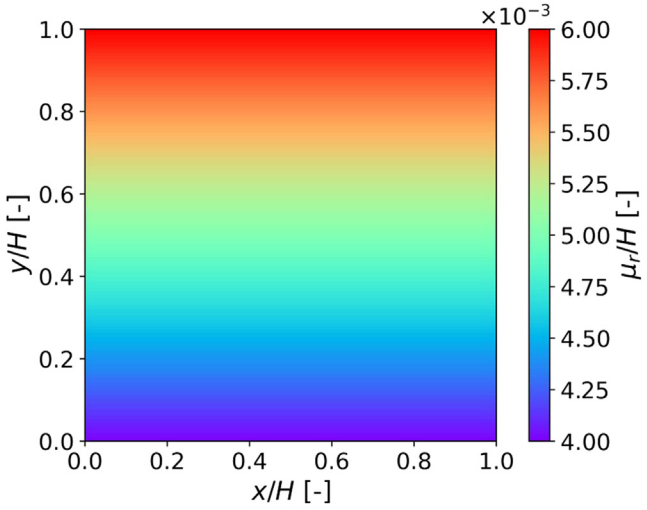

FIG. 7. Plot of $\mu_r(x, y)$ from Eq. (13).

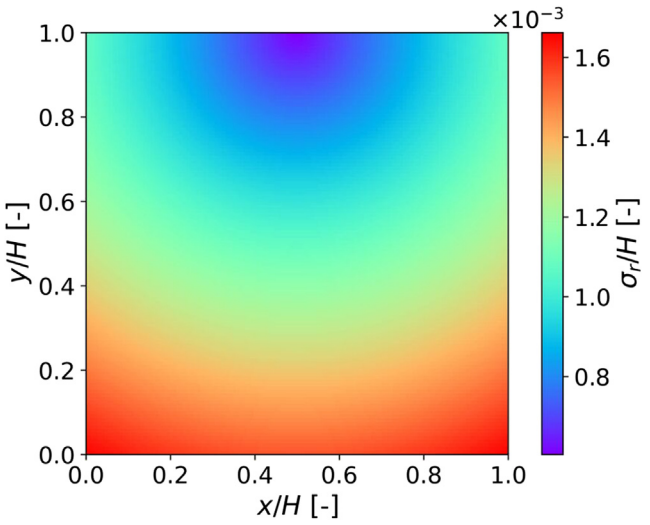

FIG. 8. Plot of $\sigma_r(x, y)$ from Eq. (14).

$$f^*(r) = \sqrt{\frac{2}{\pi}}\frac{r^2}{C_2^3}e^{-\frac{r^2}{2C_2^2}}, \quad r \in [0, \infty), \tag{15}$$

where

$$C_2 = \sigma_r\sqrt{\frac{\pi}{3\pi - 8}}. \tag{16}$$

To prescribe μ_r, we modify $f^*(r)$ by shifting the PDF by a parameter C_1 on the r-axis. The shifted PDF $f(r)$ is

$$f(r) = \sqrt{\frac{2}{\pi}}\frac{(r - C_1)^2}{C_2^3}e^{-\frac{(r-C_1)^2}{2C_2^2}}, \quad r \in [C_1, \infty), \tag{17}$$

where

$$C_1 = \mu_r - 2C_2\sqrt{\frac{2}{\pi}}, \quad C_1 > 0. \tag{18}$$

A typical snapshot resulting from the above equations is shown in Fig. 9. After distributing the bubbles, we employ the method described in Sec. II. The results can be compared to the fields prescribed in Eqs. (10), (13), and (14). In order to compare the calculated mean void fraction with its expectation, we first need the expected bubble volume μ_{V_r}, obtained from

$$\begin{aligned}\mu_{V_r} &= \int_{C_1}^{\infty}\frac{4\pi r^3}{3}f(r)dr \\ &= \frac{4\sqrt{\pi}}{3C_2}\left(2^{\frac{7}{2}}C_2^4 + 9\sqrt{\pi}C_1C_2^3 + 3\cdot 2^{\frac{3}{2}}C_1^2C_2^2 + \sqrt{\pi}C_1^3C_2\right).\end{aligned} \tag{19}$$

With the expected bubble density μ_ρ, we can calculate the expectation for the void fraction μ_α [refer to Eq. (1)],

$$\mu_\alpha = \mu_\rho\mu_{V_r}, \tag{20}$$

which is shown in Fig. 10.

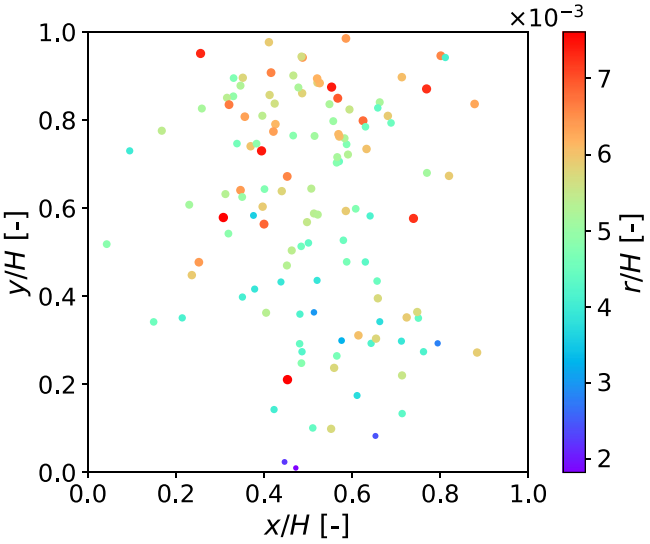

FIG. 9. Example of a synthetically generated snapshot of $M = 128$ bubbles, with properties described in Eqs. (10), (13), and (14).

In this test case, we generate 256 snapshots, each containing 128 bubbles. For an adequate spatial resolution, we set a requirement of K = 128 clusters for each bubble set. The designed experiment has the parameters listed in Table I. We repeat the analysis of the snapshots three times, each time applying one of the three settings. Parameters vary from setting 1, which does not involve any clustering (cluster size $C_{trgt} = C_{min} = C_{max} = 1$) or snapshot superposition (256 bubble-sets) to setting 3 where 64 snapshots are superpositioned for each bubble set.

B. Results

Figure 11 shows the resulting void-fraction profile at $y/H = 0.85$ for each setting. All three settings of Table I result in similar profiles,

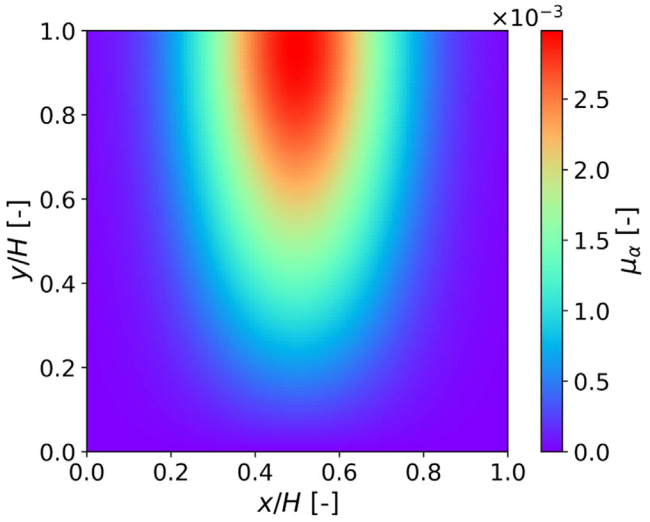

FIG. 10. Plot of $\mu_\alpha(x, y)$ from Eq. (20).

TABLE I. Parameter settings for case I.

Name	K	M^*_{min}	C_{trgt}	S^*
Setting 1	128	128	1	256
Setting 2	128	1024	8	32
Setting 3	128	8192	64	4

considering the overall deviations from the analytic profile [see Sec. III A and Eq. (20)]. As per the method, 256 bubbles are assigned to each point in space after processing the bubble sets ($C_{trgt} \times S^*$), regardless of the chosen setting. At the same time, the spatial resolution is equal for all three settings because K is unchanged. As a result, we conclude that each setting provides a statistically equal evaluation.

The number of bubbles processed simultaneously increases with snapshot superposition. Due to the complexity of the evaluation algorithm, the overall runtime required for each bubble set also increases. We found that our evaluation's complexity is approximately of the order $\mathcal{O}(M^{*2})$. If snapshots are superpositioned before the subsequent steps (see Fig. 2), the total number of bubble sets (for a fixed number of snapshots) is reduced (refer to the fifth column of Table I). As a result, the total time required to process all snapshots increases approximately linearly with the number of bubbles in a set. This behavior is shown in Fig. 12.

For a consistent quality of results for experimental data, we recommend conducting the snapshot superposition step prior to the Voronoi diagram and clustering steps (see Fig. 2). In addition, snapshot superposition should be kept to a minimum due to the longer computation run-time required for evaluations with increasing M^*.

IV. CASE II: SYNTHETIC BUBBLE DATA

In the second test case, the bubble count M is fluctuating [similar to experimental data, refer to Fig. 3(a)] with the distribution for each synthetically generated snapshot being set to the same as that of Sec. III. The analytically prescribed fields quantify the method's error. We, then, analyze the systematic error, discuss the statistical error, quantify deviations from the prescribed fields, and finally, extract local histograms.

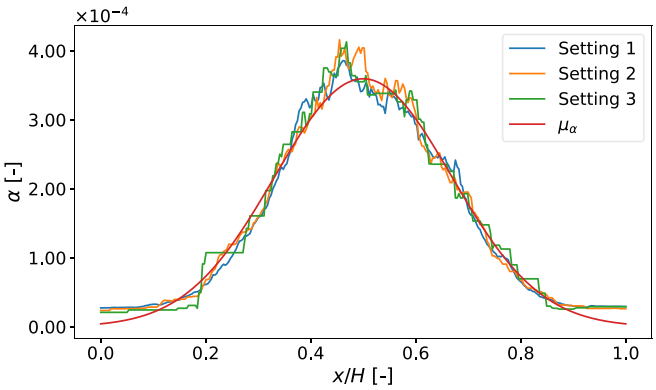

FIG. 11. Void-fraction profiles at $y/H = 0.85$ for the three settings listed in Table I.

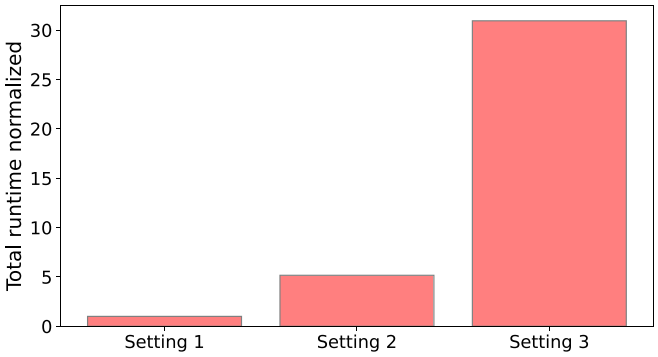

FIG. 12. Total time required for processing all test case snapshots, normalized by the time required for setting 1 (refer to Table I).

A. Description

We generate $S = 150$ snapshots for this test calculation. The bubble count per snapshot M for each snapshot is obtained by a Gaussian-shaped distribution ($\mu_M = 1024$ and $\sigma_M = 250$). Figure 13 shows the histogram of the bubble count per snapshot used for this test case. The cluster size is set as $C_{\text{trgt}} = 16$. In the snapshot superposition step, we superposition images until a bubble count of $M^*_{\text{min}} = 16 \times 128$ is reached.

B. Results

The snapshots are processed according to the workflow described in Fig. 2. Figure 14 shows the void-fraction field for one bubble set following the Voronoi diagram step. The cluster-averaged void fraction $\bar{\alpha}$, generated after employing the constrained K-Means clustering algorithm, is shown in Fig. 15.

1. Systematic errors

There are two systematic errors. One source of the error is the relative distance between the cluster centroid and the mean location of the bubble centers. The distance between these locations is visualized by the error vector $\vec{\varepsilon}_{err}$ in Fig. 16. The true spatial mean location of information (bubble occurrences) is located at the big red dot (Fig. 16). In contrast, the apparent location of this information is located at the big green dot as a result of the spatial expansion of the cluster. This error vector varies for each snapshot and each cluster.

In situations with a bubble-density gradient, the mean error vector is obtained after averaging multiple snapshots, which causes a spatial mislocation of information. This snapshot-averaged error is shown in Figs. 17(a) and 17(b). The length [Fig. 17(a)] of the error vector $\vec{\varepsilon}_{\text{err}}$ grows larger as it nears the domain boundaries and scales with the cluster size (c). The direction of $\vec{\varepsilon}_{\text{err}}$ is visualized by the angle γ between $\vec{\varepsilon}_{\text{err}}$ and the x-axis. Figure 17(b) displays the bubble-set-averaged angle $\langle\gamma\rangle$ and demonstrates that $\langle\gamma\rangle$ is driven by the direction of the gradient vector $\vec{\nabla}\rho$ of the bubble density ρ.

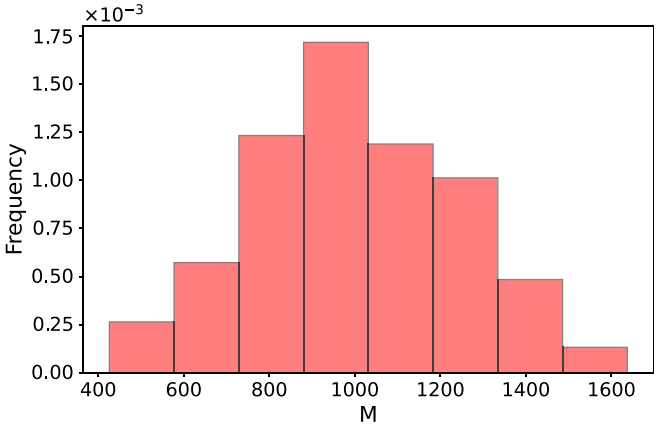

FIG. 13. Histogram of the bubble count M per snapshot used in test case II.

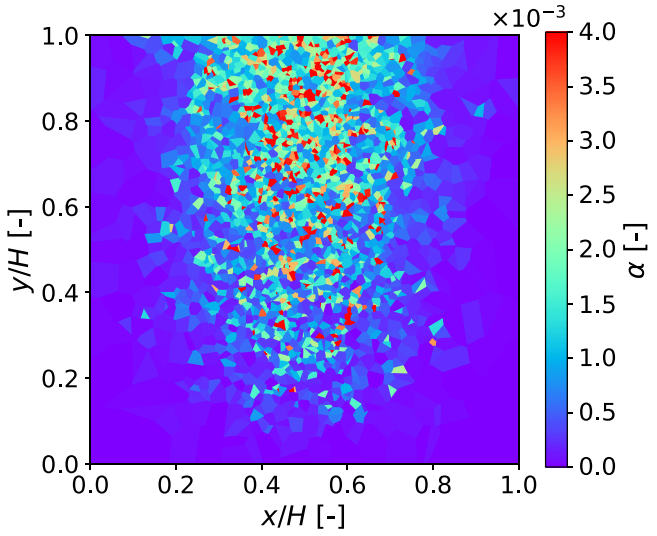

FIG. 14. Example void fraction α following the Voronoi diagram step for test case II (see Fig. 2).

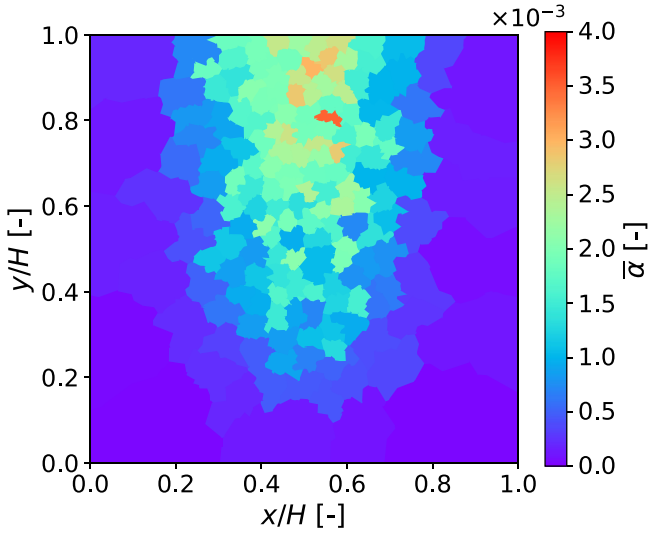

FIG. 15. Example cluster-averaged void fraction $\bar{\alpha}$ for case II.

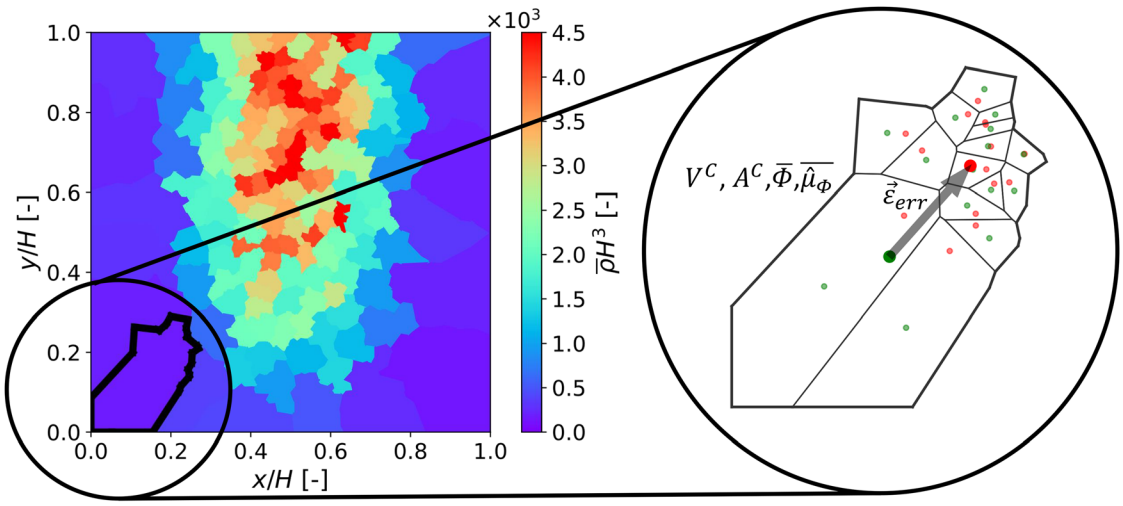

FIG. 16. Visualization of the error vector $\vec{\varepsilon}_{err}$ (light gray arrow) for an example cluster of 15 bubbles. The error vector connects the cluster centroid (big green dot) based on the Voronoi centroids (small green dots) with the location of the mean bubble center (big red dot). The bubble centers are represented by small red dots. The clusters are colored based on their cluster-averaged bubble density (on the left).

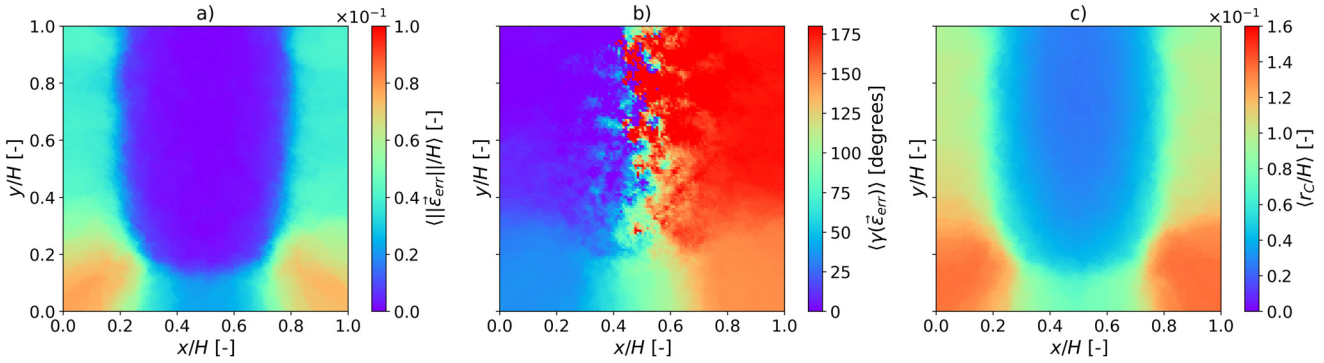

FIG. 17. Visualization of the averaged systematic errors. (a) Mean length of the error vector $\vec{\varepsilon}_{\text{err}}$, (b) angle $\langle\gamma\rangle$ between $\vec{\varepsilon}_{\text{err}}$ and the x-axis, and (c) mean equivalent cluster radius $\langle r_C\rangle$.

The second systematic error arises from the low spatial resolution, which implies uncertainty in the location of the information. In Fig. 18, we show an example cross section at $y/H = 0.85$ of the cluster- and bubble-set-averaged void fraction $\langle\bar{\alpha}\rangle$. At nine locations, the averaged uncertainty in the x-location, obtained from the mean expansion of the clusters, is shown. The spatial uncertainty increases toward the boundaries of the domain due to lower bubble density.

With the proposed method, volume-averaged values ϕ_m (from the mth Voronoi cell) are derived as approximations of local prescribed quantities $\mu_\phi(x, y)$. We define $\hat{\mu}_\phi^m$ as the volume-averaged expectation that is consistently estimated by a Voronoi region A_m,

$$\hat{\mu}_\phi^m = \frac{1}{V_m}\int_{V_m} \mu_\phi(x, y)dV. \tag{21}$$

For every point $\mathbf{x}$, we calculate a cluster-averaged value $\bar{\phi}$ and a cluster-averaged expectation $\overline{\hat{\mu}_\phi}$ (see Fig. 16). As a result, the larger

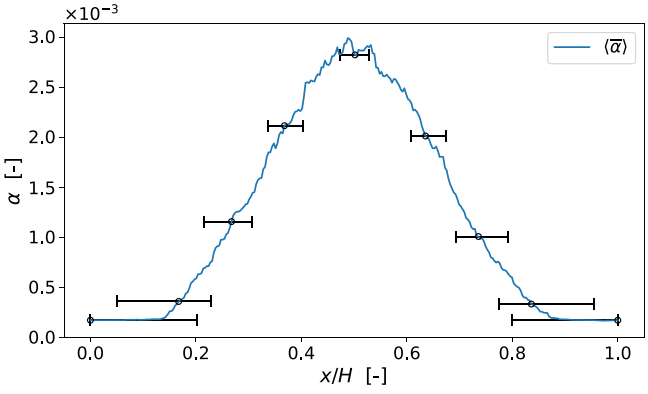

FIG. 18. Profile of cluster- and time-averaged void fraction $\langle\bar{\alpha}\rangle$ at $y/H = 0.85$ with nine error bars indicating the time-averaged spatial uncertainty.

the cluster volume V^C, the lower the approximation accuracy of $\mu_\phi(x,y)$. The difference between $\langle\overline{\hat{\mu}_\phi}\rangle$ and $\mu_\phi(x,y)$ is depicted in Fig. 19. While the deviation at $x/H = 0.5$ is negligible, we see a significant difference at $x/H = 0$ and $x/H = 1$.

We introduce an equivalent, average cluster radius $\langle r_C\rangle = (\langle V^C\rangle\pi^{-1})^{0.5}$ as a measure of the uncertainty. A visualization of this radius can be found in Fig. 17(c). To reduce the two errors, the number of clusters K in each bubble set can be increased (i.e., keeping M^*_{min} and decreasing C_{trgt} or increasing M^*_{min} and keeping C_{trgt}).

The systematic nature of these errors could be addressed through correction algorithms; however, this is beyond the scope of this work.

2. Statistical errors

After processing each bubble set, any given point in the flow domain is assigned a total of $N \in \{S^* \times (C_{\text{trgt}} - 1)\ldots S^* \times (C_{\text{trgt}} + 1)\}$ bubbles. The observed mean $\langle\bar{\phi}\rangle$ of an arbitrary quantity ϕ can be calculated from the observations $\phi_1, \phi_2, \ldots, \phi_N$ using the weights $w_1, w_2, \ldots, w_N$ [see Eq. (6)], where $\sum_{i=1}^{N} w_i = 1$. The value $\langle\bar{\phi}\rangle$ is an estimation of the cluster- and time-averaged expectation of $\langle\overline{\hat{\mu}_\phi}\rangle$. According to the law of large numbers,[47] the error $|\langle\bar{\phi}\rangle - \langle\overline{\hat{\mu}_\phi}\rangle|$ almost surely converges to zero as N approaches infinity. However, for a finite set of N samples (bubbles), a statistical error remains given that $\langle\overline{\hat{\mu}_\phi}\rangle$ is in general unknown. With the weights w_i and the samples ϕ_i, an interval is calculated, where we suspect the expectation $\langle\overline{\hat{\mu}_\phi}\rangle$ to lay with a probability $(100\% - \alpha\%)$. We obtain this confidence interval[48,49] from

$$\langle\bar{\phi}\rangle \pm \frac{t_{\alpha/2,(N-1)}\sqrt{\sum_{i=1}^{N} w_i(\phi_i - \langle\bar{\phi}\rangle)^2}}{\sqrt{(N-1)\left(N^N\prod_{i=1}^{N} w_i\right)^{1/(N-1)}}}. \tag{22}$$

Here, $t_{\alpha/2,(N-1)}$ refers to the t-value of the student-t distribution with $N - 1$ degrees of freedom. As seen in Eq. (22), the confidence interval decreases with increasing sample size N. Figure 20 displays the confidence interval for the mean void fraction $\langle\bar{\alpha}\rangle$. The confidence

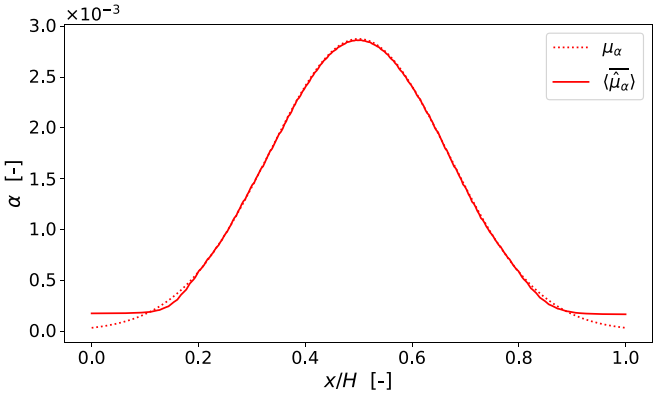

FIG. 19. Profile of the cluster- and time-averaged void-fraction expectation $\langle\overline{\mu_\alpha}\rangle$ at $y/H = 0.85$ compared to $\mu_\alpha(x,y)$.

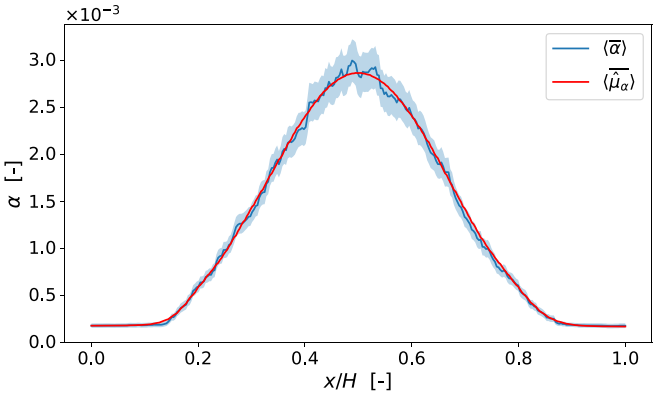

FIG. 20. 99% confidence interval (light blue region) for the calculated mean void fraction $\langle\bar{\alpha}\rangle$ (solid blue line) and the averaged expectation $\langle\overline{\hat{\mu}_\alpha}\rangle$ (solid red line).

interval covers the region with 99% confidence and matches the expectation $\langle\overline{\hat{\mu}_\alpha}\rangle$.

To obtain narrower confidence intervals, one can increase the number of snapshots S, and thus, the total sample size N per point in space.

3. Choice of method settings

The magnitude and influences of the local confidence interval and its spatial uncertainty are discussed in the previous paragraphs. Based on the resulting uncertainties, a decision on the final method settings can be made. An unacceptable spatial uncertainty can be addressed by increasing M^*_{min} while admitting a longer runtime (see Sec. III B). Decreasing C_{trgt} has a positive effect on spatial uncertainty, but decreases the number of assigned bubbles per point in space, and thus widens the confidence interval. The two parameters are chosen arbitrary for synthetic bubble data but need to be adjusted iteratively for experimental data to find an acceptable compromise of computational cost, spatial uncertainty, and statistical significance. Recommendations for these parameters are generally challenging to make, as they strongly depend on the application. Additionally, the decision on the acceptance of the error compromise is driven by the purpose of the results. For example, parameters optimized for calculating higher-order moments where a larger count of samples (bubbles) per point is required might be inappropriate for evaluating a mean value's spatial gradient.

4. Deviations of the fields

Next, we assess the deviation of the calculated approximations for $\mu_r(x,y)$, $\sigma_r(x,y)$, and $\mu_\alpha(x,y)$, compared to the analytically prescribed fields (see Sec. III A). These deviations in the domain are visualized in Fig. 21. The deviation from $\mu_r(x,y)$ is in the range of zero to five percent and can be viewed as rather small; however, we found larger deviations up to 15% from $\sigma_r(x,y)$. The largest errors are found in Fig. 21(c), where we compare our results with the prescribed expectation of the void fraction $\mu_\alpha(x,y)$. Here, the regions close to the boundaries exhibit large deviations (>20%) also coinciding with the regions characterized by high systematic error (see Fig. 17).

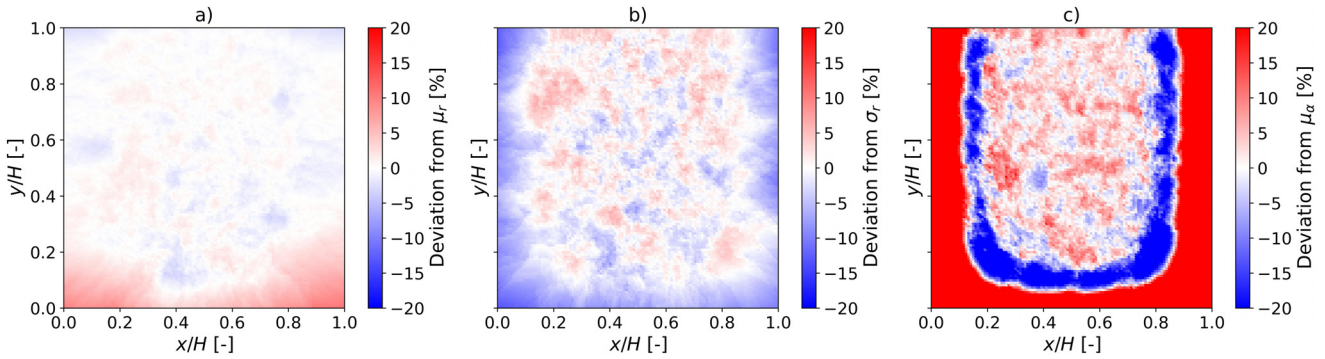

FIG. 21. Error plots of (a) the deviation from the expected bubble radius μ_r, (b) the standard deviation for the bubble radius σ_r, and (c) the deviation from the expected void fraction μ_α. All deviations are presented as percentages.

5. Local histogram

The proposed method allows local histograms for quantities related to single bubble occurrences to be generated. These approximations of local PDFs each have a similar sample size N, independent of their location in the domain. This important feature is demonstrated in Fig. 22, where we use the void fraction α to create a local histogram of the occurrence frequencies of each α-value, using $x/H = 0.3$ and $y/H = 0.3$ as an example location in the domain.

V. CASE III: EXPERIMENTAL BUBBLE DATA

In this section, we examine two-dimensional snapshot data, taken from an existing experiment,[4,43] to investigate a circular plunging-jet with air entrainment. Case III aims to apply the proposed method to real experimental data and confirm its results by comparing them to the literature.[3,33–37,44] For that purpose, we extract the void-fraction distribution and the air entrainment rate. Table II summarizes the relevant parameters of this experimental study.

A. Description

The shadowgraphs obtained from the experiment[4,43] are processed as follows. The schematic setup can be found in Fig. 23; here, the location of the detection region for the shadowgraphs is visualized.

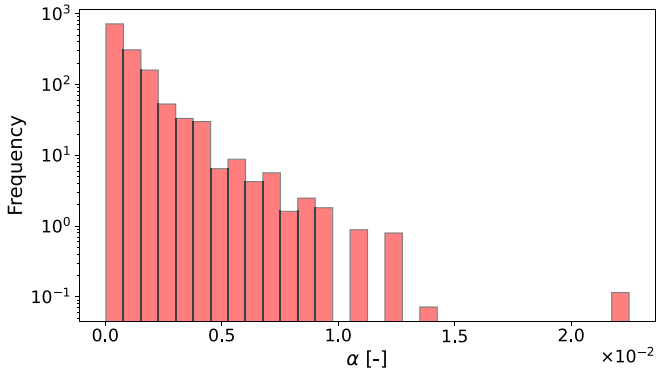

FIG. 22. Normalized histogram of void-fraction occurrences gathered over $S = 100$ snapshots at location $\frac{x}{H} = 0.3$ and $\frac{y}{H} = 0.3$.

Raw shadowgraph images (e.g., Fig. 24) are used to extract the bubble location, equivalent radius, and bubble velocity.[43] The circular plunging-jet bubble distribution is, on a statistical average, a rotational-symmetric, three-dimensional flow. With shadowgraph imaging, the results are captured in two dimensions. A correction method[50] (refer to Interferometry[51] for an example application) is applied to the apparent void-fraction field (in 2D) to yield the true void fraction (as a function of the centerline distance R). The applicability of this method for evaluating void fractions in bubbly flows can be found in Appendix. During the experiment, 3600 snapshots were taken, each with an average of 48 bubbles [PDF of the bubbles per snapshot are found in Fig. 3(a)]. The cluster size is $C_{trgt} = 4$, while a minimum bubble count of $M^*_{min} = 220 \times 4$ is set for each bubble set. More detailed information on the experiment has been previously published.[4,43]

B. Results

In the following, we discuss the resulting void-fraction distribution and air entrainment rate.

1. Void-fraction distribution

The results of all of the bubble sets are averaged, allowing the mean void-fraction field to be calculated from the two-dimensional

TABLE II. Parameters for the plunging-jet experiment.[43]

Description	Variable	Value	Unit
Nozzle diameter	d_0	6.0	mm
Jet velocity at nozzle	v_0	1.76	m/s
Jet diameter at impact point	d_1	4.90	mm
Jet velocity at impact point	v_1	2.64	m/s
Jet length	L	197	mm
Fluid surface tension	$\sigma_{Fl.}$	0.0723	N/m
Fluid density	$\rho_{Fl.}$	997.3	kg/m^3
Jet fluid volume flux	$Q_{Fl.}$	0.0497	dm^3/s
Dynamic viscosity	$\mu_{Fl.}$	940×10^{-6}	Pa·s
Jet turbulence level[33,43]	Tu	1–1.5	%

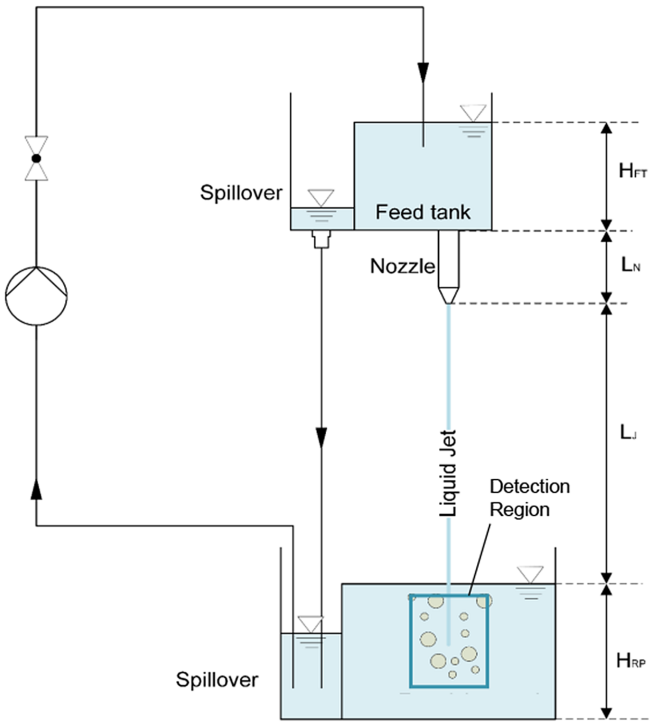

FIG. 23. Visualization of the plunging-jet experimental facility.[4,43]

bubble data. In Fig. 25, the distribution of $\langle\bar{\alpha}\rangle(x,z)$ is displayed. In this experiment, the void fraction is concentrated at the lower end of the descending bubble core; this was also found by Qu *et al.*[3] The field of $\langle\bar{\alpha}\rangle(x,z)$ exhibits local minima and maxima in the mean void-fraction field. Our experience has shown that removing the local extrema generates better final results. In the presented case, a kernel smoothing method is used, with the smoothing radius set to 8 mm.

At this point, the correction method[50] used for Interferometry[51] is applied to the smoothed void-fraction profile by using the assumption of a statistically rotational symmetric flow (see study in Appendix). The correction method calculates the true void fraction $\langle\bar{\alpha}\rangle'(z,R)$ from the apparent void fraction $\langle\bar{\alpha}\rangle(x,z)$ that was obtained from the shadowgraphs. Figure 26 shows the corrected void-fraction profiles for five different depths in the pool.

2. Air entrainment rate

We consider the drift of the void $\alpha_m \cdot \vec{v}_{B,m}$ for every bubble influence region, where $\vec{v}_{B,m}$ is the velocity of the mth bubble. We calculate the corrected, bubble-set- and cluster-averaged void drift $\langle\overline{\alpha\vec{v}_B}\rangle'(z,R)$ analogous to the void fraction as described above. The volumetric void flux through an imaginary area A_{cone} of the descending bubble cone in the xy-plane is used to calculate the entrainment rate,

$$\frac{Q_{\text{Air}}}{Q_{Fl.}} = \frac{1}{Q_{Fl.}}\int\int_{A_{\text{cone}}} \langle\overline{\alpha\vec{v}_B}\rangle'(z,R)\vec{n}_z dA = \frac{1}{Q_{Fl.}}\int_0^{R_{\text{cone}}} 2\pi R\langle\overline{\alpha\vec{v}_B}\rangle'(z,R)\vec{n}_z dR, \quad (23)$$

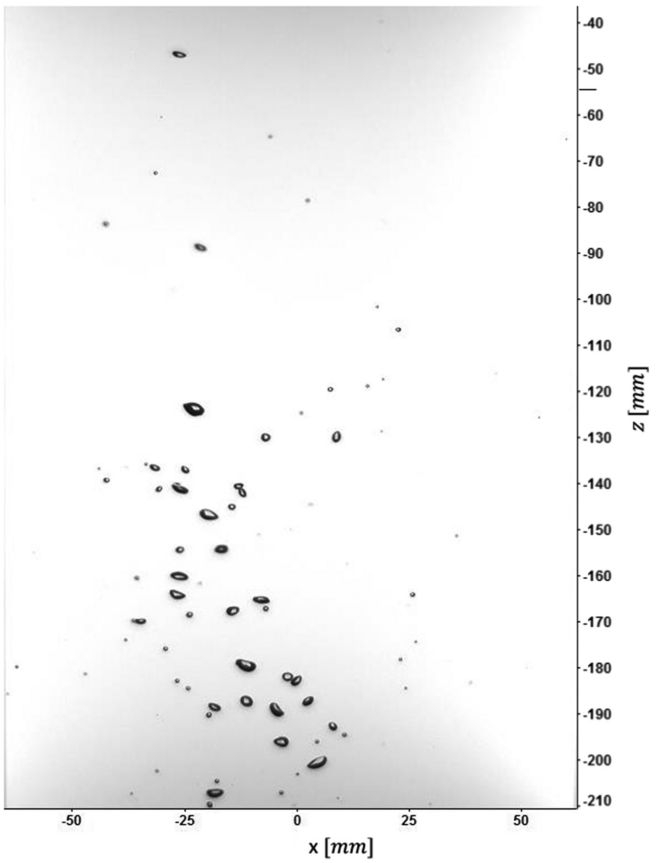

FIG. 24. Unprocessed shadowgraph of a bubble distribution obtained from the plunging-jet experiment.[4,43]

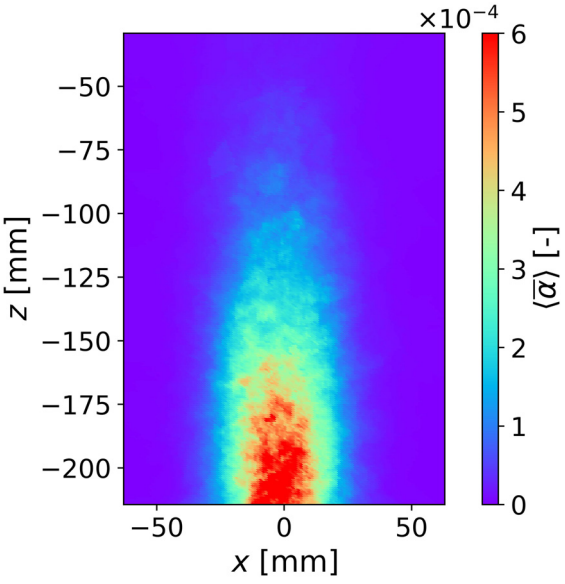

FIG. 25. Distribution of $\langle\bar{\alpha}\rangle(x,z)$ after averaging over snapshots of the plunging-jet experiment.

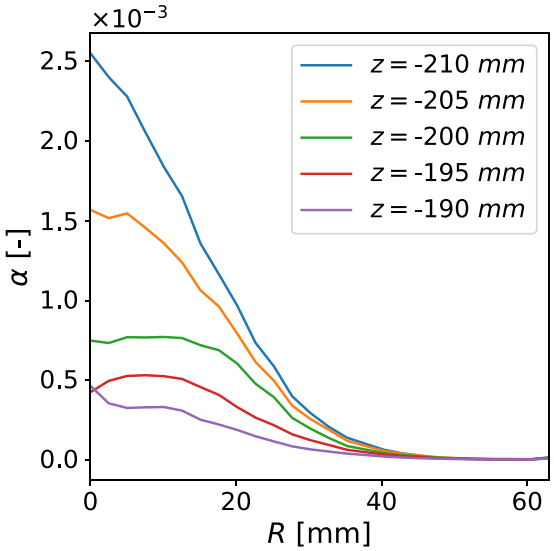

FIG. 26. Cross section at various pool depths with corrected mean void-fraction profiles.

where R_{cone} is the limit of the descending bubble cone. Our experiment[43] exhibits a cone angle of ≈11°, which matches the angles found by Ervine and Falvey.[34] For z = 100 mm, we obtain an entrainment rate of $\frac{Q_{Air}}{Q_{Fl.}} \approx 0.16$. The ratio $\frac{Q_{Air}}{Q_{Fl.}}$ is slightly lower than the value predicted by the empirical correlation of El Hammoumi and Davoust,[36] which is based on similar experimental settings. Using their correlation[36] gives

$$\frac{Q_{Air}}{Q_{Fl.}} = \frac{1}{Q_{Fl.}} 3.851^{-5} L^{0.885} {d_0}^{0.193} {v_0}^{2.23} \approx 0.24. \quad (24)$$

Taking Chanson's book[37] diagram for entrainment rates of circular plunging jets, we find that the extracted air entrainment rate is underestimated. From this empirical diagram, we extract a predicted entrainment rate of ≈0.13 for our experiment. The value from Eq. (23) falls between the values predicted by El Hammoumi and Davoust[36] and Chanson.[37] Other correlations[33,35,44] overestimated our extracted air entrainment rate.

VI. CONCLUSIONS

We use Voronoi diagrams to calculate an influence region for every bubble. Unlike the box-count method (see Sec. I), no *a priori* knowledge of the mean bubble distance is needed to define a suitable box size. This is a significant benefit of using the proposed method.

The constrained K-Means clustering algorithm allows bubbles to be grouped into clusters with constrained sample sizes. As a result, each point in the flow domain is assigned a similar number of bubbles, regardless of its location in the flow domain. In contrast to other methods, we can create local histograms with approximately constant amounts of samples. Therefore, we conclude that we can obtain a robust evaluation of the flow even in regions where the bubble density is low (sparsely occupied regions). This capability represents a key difference from other methods that capture only local information of the gaseous phase (e.g., needle probes).

An advantage of the strict definition of the influence region in our method is that every bubble has an associated cell volume V. This means that no box-averaged volume has to be used for the bubble density ρ or the void fraction α; therefore, our method achieves a more consistent derivation of histograms for these values.

Furthermore, we demonstrate that the proposed method achieves a constant and high spatial resolution by utilizing a snapshot superposition step. This superposition step is important for experimental data, as it smooths fluctuations in the bubble count per snapshot. A drawback of the superposition is that it prolongs the runtime of the algorithm, as shown in Sec. III B.

The inherent errors of the proposed method are critically assessed in Sec. IV B. We show that the systematic error can be reduced by increasing the number of clusters K and that increasing the number of snapshots S narrows the confidence interval of the statistical error. Therefore, these errors show this favorable property of convergence as we increase the resolution and amount of information.

The proposed method provides good approximations for spatial distributions of mean values and standard deviations. This was confirmed by comparing the results to the analytical expectations and standard deviations of the synthetic snapshot data as well as to literature data in Sec. V B. Now, in subsequent studies, the results obtained from the proposed method can be used to reveal a wide range of fascinating physics.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of F. Buettner and W. Heiler, from the Multiphase Flow (MPF) research group at the Institute of Thermal Energy Technology and Safety (at Karlsruhe Institute of Technology) for providing the plunging-jet experimental data. The support of the Framatome Professional School (FPS) is acknowledged. In addition, we are grateful for the comments of J. Yanez, C. Tazzeo, and P. Marthaler, which improved the quality of the manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Lorenz Weber: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Stephan Gabriel:** Resources (equal); Supervision (supporting); Writing – review & editing (equal). **Andreas G. Class:** Conceptualization (equal); Funding acquisition (lead); Methodology (supporting); Project administration (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: 3D DATA PROCESSED IN 2D

This section discusses the application of a correction method[50] for void-fraction analysis in bubbly flows. This method was used to

correct two-dimensional density measurements from a rotational symmetric flow in the area of Interferometry.[51] This method assumes rings of statistically constant values ϕ'_i (see Fig. 27). An observer (ray in Fig. 27) will measure a weighted mean ϕ. The weights are given by lengths L_i that the ray is passing through the ith ring. This mean value ϕ is, then, calculated by

$$\phi = \frac{\sum L_i \phi'_i}{\sum L_i}. \tag{A1}$$

We carry out a test with a synthetic bubble distribution with known spatial expectations for the void-fraction field μ_α. The statistically rotational symmetric distribution is generated through the inverse transform sampling method.[45] The domain has the dimensions $H \times H \times H$. Here, we generate 40 snapshots with bubbles distributed in three dimensions. An example snapshot can be seen in Fig. 28. In each snapshot, the bubble count M is fluctuating, as described in Sec. IV ($\mu_M = 1024$, $\sigma_M = 250$). For the test, the bubble snapshots are processed in two- and three-dimensional environments and the results are compared. For both calculations, a spatial kernel smoother is used, with a kernel radius of $0.05H$.

1. 2D Evaluation

For the two-dimensional evaluation, we project the bubble distribution of each snapshot onto the xz-plane. The proposed workflow (Fig. 2) is, then, applied in two dimensions. The resulting profile is, then, transformed, using the correction method.[50,51] We select $C_{\text{trgt}} = 8$, $M^*_{\text{min}} = 128 \times 8$ for this evaluation.

2. 3D Evaluation

The method described in Sec. II has no limitations in terms of the spatial dimensionality of the bubble data. As such, the proposed method is directly applied to the three-dimensional snapshots. The cluster count K is adjusted so that the equivalent mean cluster radius is the same in two and three dimensions. Here, $C_{\text{trgt}} = 2$ and $M^*_{\text{min}} = 966 \times 2$.

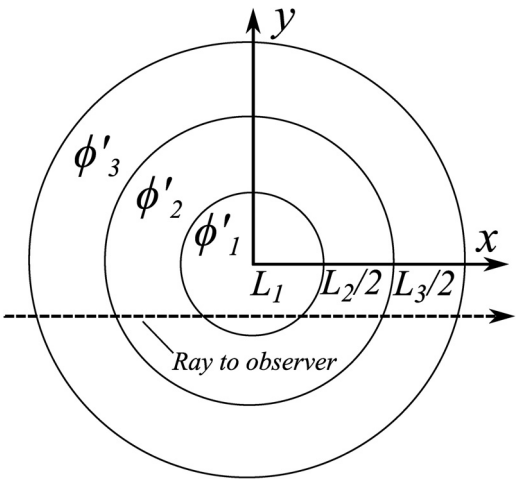

FIG. 27. Rotational symmetric densities (here in two dimensions) measured in one dimension by an observer ray. Rings of the same density ϕ'_i are formed around the center. Lengths L_i refer to distance of the ray traveling through the ith ring.

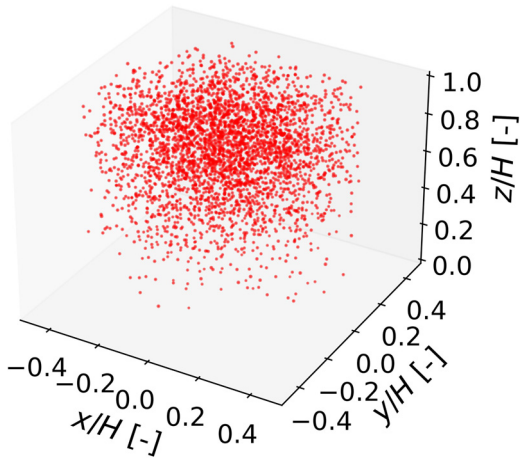

FIG. 28. Example snapshot for three-dimensional bubble distribution with $M = 1024$ in a cube with lengths H.

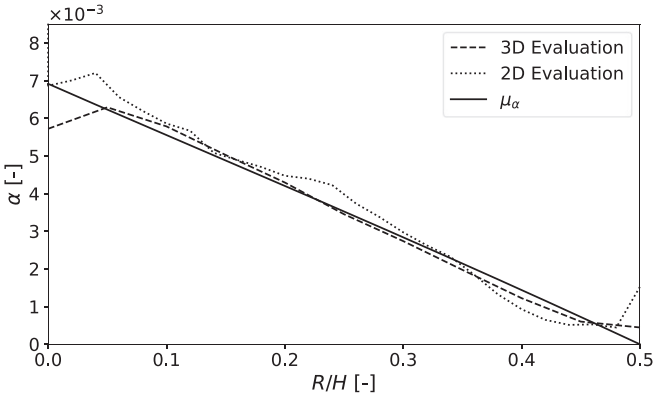

FIG. 29. Comparison of void-fraction profiles at $z/H = 0.85$ calculated in three dimensions (dashed line) and with a correction method[51] in two dimensions (dotted line). The analytic expectation of μ_α is depicted by the solid line.

Figure 29 compares the void-fraction profiles at $z/H = 0.85$ as a function of the centerline distance $R = \sqrt{x^2 + y^2}$. The two- and three-dimensional analyses with the analytic expectation of μ_α are depicted. Both evaluations provide a good approximation of μ_α. While the three-dimensional analysis is more accurate, the two-dimensional methods exhibit stronger fluctuations around the expectation μ_α. These stronger fluctuations can be explained by the accumulation of errors toward the centerline.[51]

REFERENCES

[1]F. Hernandez-Alvarado, S. Kleinbart, D. V. Kalaga, S. Banerjee, J. B. Joshi, and M. Kawaji, "Comparison of void fraction measurements using different techniques in two-phase flow bubble column reactors," Int. J. Multiphase Flow **102**, 119–129 (2018).

[2]A. R. Gardenghi, E. D. S. Filho, D. G. Chagas, G. Scagnolatto, R. M. Oliveira, and C. B. Tibiriçá, "Overview of void fraction measurement techniques,

databases and correlations for two-phase flow in small diameter channels," Fluids **5**, 216 (2020).
[3]X. L. Qu, L. Khezzar, D. Danciu, M. Labois, and D. Lakehal, "Characterization of plunging liquid jets: A combined experimental and numerical investigation," Int. J. Multiphase Flow **37**, 722–731 (2011).
[4]G. Albrecht, W. Heiler, F. Büttner, and S. Gabriel, "Experimental investigation of air entrainment by a vertical jet plunging into a liquid pool," in 10th International Conference on Multiphase Flow, Rio de Janeiro, BR, May 19–24 (2019).
[5]Y. Lau, K. Müller, S. Azizi, and M. Schubert, "Voronoï analysis of bubbly flows via ultrafast X-ray tomographic imaging," Exp. Fluids **57**, 35 (2016).
[6]E. Krepper, D. Lucas, T. Frank, H.-M. Prasser, and P. J. Zwart, "The inhomogeneous MUSIG model for the simulation of polydispersed flows," Nucl. Eng. Des. **238**, 1690–1702 (2008).
[7]J. X. P. Dumond, "Fluidic diode for critical two-phase flows," Ph.D. thesis [Karlsruher Institut für Technologie (KIT), 2012].
[8]S. Navarro-Martinez, "Large eddy simulation of spray atomization with a probability density function method," Int. J. Multiphase Flow **63**, 11–22 (2014).
[9]B. Chen and M. Oevermann, "LES investigation of ECN spray G2 with an Eulerian stochastic field cavitation model," in SAE Technical Paper, 2018.
[10]M. R. Raquet, "Stochastische Euler-Euler-PDF-methodik für das zwei-fluid-modell und anwendung auf kavitierende strömungen der automobilindustrie," Ph.D. thesis [Karlsruher Institut Für Technologie (KIT), 2019].
[11]D. M. Wang and D. Greif, "Progress in modeling injector cavitating flows with a multi-fluid method," in Fluids Engineering Division Summer Meeting, 2006.
[12]L. Valiño, "A field Monte Carlo formulation for calculating the probability density function of a single scalar in a turbulent flow," Flow, Turbul. Combust. **60**, 157–172 (1998).
[13]Y. Liao, R. Rzehak, D. Lucas, and E. Krepper, "Baseline closure model for dispersed bubbly flow: Bubble coalescence and breakup," Chem. Eng. Sci. **122**, 336–349 (2015).
[14]A. Aliseda, A. Cartellier, F. Hainaux, and J. C. Lasheras, "Effect of preferential concentration on the settling velocity of heavy particles in homogeneous isotropic turbulence," J. Fluid Mech. **468**, 77–105 (2002).
[15]R. Monchaux, M. Bourgoin, and A. Cartellier, "Analyzing preferential concentration and clustering of inertial particles in turbulence," Int. J. Multiphase Flow **40**, 1–18 (2012).
[16]Y. Lau, K. T. Sujatha, M. Gaeini, N. Deen, and J. Kuipers, "Experimental study of the bubble size distribution in a pseudo-2D bubble column," Chem. Eng. Sci. **98**, 203–211 (2013).
[17]W. Zhao, H. Wang, R. Bai, W. Wei, and H. Wang, "Bubble characteristics and turbulent dissipation rate in horizontal bubbly pipe flow," AIP Adv. **11**, 025125 (2021).
[18]D. Laupsien, C. Le Men, A. Cockx, and A. Liné, "Image processing for bubble morphology characteristics in diluted bubble swarms," Phys. Fluids **31**, 053306 (2019).
[19]L. Villafañe-Roca, M. Esmaily-Moghadam, A. Banko, and J. Eaton, "A robust method for quantification of preferential concentration from finite number of particles," *Center for Turbulence Research: Annual Research Briefs* (Stanford University, 2016).
[20]G. Voronoi, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites," J. Reine Angew. Math. **1908**, 97–102.
[21]G. Voronoi, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs," J. Reine Angew. Math. **1908**, 198–287.
[22]R. Kipping, M. Wagner, and U. Hampel, "On inter-bubble distances and bubble clustering in bubbly flows: An experimental study," Chem. Eng. J. **431**, 133486 (2022).
[23]R. Simha, "A treatment of the viscosity of concentrated suspensions," J. Appl. Phys. **23**, 1020–1024 (1952).
[24]R. Nigmatulin, "Spatial averaging in the mechanics of heterogeneous and dispersed systems," Int. J. Multiphase Flow **5**, 353–385 (1979).
[25]G. S. Arnold, D. A. Drew, and R. T. Lahey, Jr., "Derivation of constitutive equations for interfacial force and Reynolds stress for a suspension of spheres using ensemble cell averaging," Chem. Eng. Commun. **86**, 43–54 (1989).
[26]A. G. Kidanemariam, C. Chan-Braun, T. Doychev, and M. Uhlmann, "Direct numerical simulation of horizontal open channel flow with finite-size, heavy particles at low solid volume fraction," New J. Phys. **15**, 025031 (2013).
[27]R. Monchaux, M. Bourgoin, and A. Cartellier, "Preferential concentration of heavy particles: A Voronoï analysis," Phys. Fluids **22**, 103304 (2010).
[28]Y. Tagawa, J. M. Mercado, V. N. Prakash, E. Calzavarini, C. Sun, and D. Lohse, "Three-dimensional Lagrangian Voronoï analysis for clustering of particles and bubbles in turbulence," J. Fluid Mech. **693**, 201–215 (2012).
[29]A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* (John Wiley & Sons, Inc., New York, 1992).
[30]D. Haworth, "Progress in probability density function methods for turbulent reacting flows," Prog. Energy Combust. Sci. **36**, 168–259 (2010).
[31]K. Bennett, P. Bradley, and A. Demiriz, "Constrained K-Means clustering," Technical Report No. MSR-TR-2000-65 (Microsoft Research, 2000).
[32]E. Van de Sande, "Air entrainment by plunging water jets," Ph.D. thesis (TU Delft, 1974).
[33]E. McKeogh and D. Ervine, "Air entrainment rate and diffusion pattern of plunging liquid jets," Chem. Eng. Sci. **36**, 1161–1172 (1981).
[34]D. A. Ervine and H. T. Falvey, "Behaviour of turbulent water jets in the atmosphere and in plunge pools," Proc. Inst. Civ. Eng. **83**, 295–314 (1987).
[35]A. K. Biń, "Gas entrainment by plunging liquid jets," Chem. Eng. Sci. **48**, 3585–3630 (1993).
[36]M. El Hammoumi and L. Davoust, "Measurements of air entrainment by vertical plunging liquid jets," Exp. Fluids **32**, 624 (2002).
[37]H. Chanson, "17—Interaction between flowing water and free surfaces: Self-aeration and air entrainment," in *Environmental Hydraulics of Open Channel Flows*, edited by H. Chanson (Butterworth-Heinemann, Oxford, 2004), pp. 348–398.
[38]D. Melzer, "Experimental investigations on impinging liquid jets with gas entrainment," Ph.D. thesis (Technische Universität Dresden, 2018).
[39]J. Ma, A. A. Oberai, D. A. Drew, and R. T. Lahey, Jr., "A two-way coupled polydispersed two-fluid model for the simulation of air entrainment beneath a plunging liquid jet," J. Fluids Eng. **134**, 101304 (2012).
[40]F. Zidouni Kendil, D. V. Danciu, M. Schmidtke, A. Bousbia Salah, D. Lucas, E. Krepper, and A. Mataoui, "Flow field assessment under a plunging liquid jet," Prog. Nucl. Energy **56**, 100–110 (2012).
[41]A. Bahadar, "Volume of fluid computations of gas entrainment and void fraction for plunging liquid jets to aerate wastewater," ChemEngineering **4**, 56 (2020).
[42]Z. Yin, Q. Jia, Y. Li, Y. Wang, and D. Yang, "Computational study of a vertical plunging jet into still water," Water **10**, 989 (2018).
[43]F. Büttner, "Experimental investigation of air entrainment by a vertical jet plunging into a liquid pool," Master thesis [Karlsruhe Institute of Technology, Department of Mechanical Engineering, Institute for Thermal Energy Technology and Safety (ITES), 2018].
[44]S. Miwa, T. Moribe, K. Tsutstumi, and T. Hibiki, "Experimental investigation of air entrainment by vertical plunging liquid jet," Chem. Eng. Sci. **181**, 251–263 (2018).
[45]L. Devroye, "General principles in random variate generation," in *Non-Uniform Random Variate Generation* (Springer New York, New York, 1986), pp. 27–82.
[46]M. V. Sadovskii, *Statistical Physics* (De Gruyter, Berlin/Boston, 2019).
[47]N. Etemadi, "Convergence of weighted averages of random variables revisited," Proc. Am. Math. Soc. **134**, 2739–2744 (2006).
[48]A. L. Rukhin, "Conservative confidence intervals based on weighted means statistics," Stat. Probab. Lett. **77**, 853–861 (2007).
[49]A. L. Rukhin, "Weighted means statistics in interlaboratory studies," Metrologia **46**, 323 (2009).
[50]N. Abel, "Auflösung einer mechanischen aufgabe." J. Reine Angew. Math. **1826**, 153–157.
[51]W. Merzkirch, *Flow Visualization*, 2nd ed. (Academic Press, San Diego, 1987), pp. 1–13.