

## **A workflow for the sustainable development of closure models for bubbly flows**

Hänsch, S.; Evdokimov, I.; Schlegel, F.; Lucas, D.;

Originally published:

June 2021

**Chemical Engineering Science 244(2021), 116807**

DOI: <https://doi.org/10.1016/j.ces.2021.116807>

Perma-Link to Publication Repository of HZDR:

<https://www.hzdr.de/publications/Publ-32323>

Release of the secondary publication  
on the basis of the German Copyright Law § 38 Section 4.

CC BY-NC-ND

# A workflow for the sustainable development of closure models for bubbly flows

S. Hänsch<sup>a</sup>, I. Evdokimov, F. Schlegel, D. Lucas

*Helmholtz-Zentrum Dresden-Rossendorf e.V., Institute of Fluid Dynamics,  
01314 Dresden, Germany*

<sup>a</sup>*Corresponding author: s.haensch@hzdr.de*

---

## Abstract

Many years of research in developing closure models for polydisperse bubbly flows have produced a plethora of empirical and semi-empirical models. The continuous development and analysis of such models requires their constant validation with the steadily increasing number of validation cases in the literature.

In this paper we present a pipeline for the fully-automated analysis of OpenFOAM simulations using the Snakemake workflow management system. The pipeline is applied to an extensive collection of well-established validation cases for bubbly flows and allows the fast and efficient production of large amounts of results that are summarized in well-structured reports. An optional post-processing step introduces a fuzzy logic controller developed for the detailed analysis of these results by quantifying the agreement of the simulation with the available experimental data. It is demonstrated how such quantification enables the systematic evaluation of new closure models and contributes to a more sustainable model development.

*Keywords:*

Baseline, bubbly flow, OpenFOAM, workflow, artificial intelligence

---

## 1. Introduction

Understanding and predicting multiphase flows is crucial for the development of reliable, safe and efficient devices in the nuclear, chemical, energy and oil-and-gas industry. Computational fluid dynamics (CFD) has proven to be an indispensable tool that adds to a better understanding of multiphase flows on all relevant scales, from single bubbles to the flow in large industrial components. Much effort was focused on achieving predictive capabilities of CFD simulations during the last decades. Today, the Eulerian-Eulerian two-fluid methodology is widely used for analysing multiphase flows on the industrial component-scale. Low computational cost and great flexibility in terms of possible flow geometries have made it the state-of-the-art approach for multiphase flow analysis in the industry. The reliability of results depends on the availability of accurate closure models and their correct application. Although good predictions can often be obtained for single cases, there is an ongoing effort to reduce the empiricism of models in favour of more mechanistic formulations for a better accuracy and generality of these models.

Decades of developing new models for Eulerian-Eulerian CFD have produced a vast amount of literature on the subject, and a huge number of both models and validation cases [1, 2, 3]. For each required Euler-Euler closure a multitude of empirical or semi-empirical models exists in the literature from which the CFD engineer is free to choose. This accumulation of models makes it increasingly difficult to give best practice guidelines for the simulation of bubbly flows. Evaluating new models and comparing them against the plethora of available alternatives becomes an increasingly tedious endeavour.

However, the huge amount of data to be found in the literature offers the opportunity to apply workflow tools for a faster and more efficient processing of CFD data. In other fields of research, such as bio-informatics [4] and RNA sequencing in particular [5], the development of workflows has already proven to be fundamental for gaining a deeper understanding of large data sets. The enormous amount of CFD data produced from existing validation cases and the increasing number of closure models make a fast and efficient tool to process all this data a crucial part of the Euler-Euler model development in the future. Furthermore, such automation allows us to explore the possibility of including data science and artificial intelligence tools in the analysis and evaluation process of new Euler-Euler models.

37

38 In Section 2 of this paper we describe the current status of Eulerian-  
39 Eulerian CFD modeling of multiphase flows. We will review the conven-  
40 tional way of developing and testing new models and highlight the issues  
41 that arise from following such an approach. We then present the alternative  
42 ‘baseline’ strategy [6] developed at Helmholtz-Zentrum Dresden-Rossendorf  
43 (HZDR), which aims at converging towards a fixed set of universal closure  
44 models, so-called baseline models. An overview of the current collection of  
45 validation cases for baseline model testing available at HZDR is given. In  
46 Section 3 we then present the Snakemake workflow, which was developed  
47 for the fully-automated pre-processing, running and post-processing of the  
48 extensive amount of OpenFOAM cases in our collection. Section 4 will apply  
49 fuzzy logic, a subset of artificial intelligence, for the evaluation of simulation  
50 results. We will describe the parts of a fuzzy logic controller, designed to  
51 infer from two error metrics an output metric that allows us to compare the  
52 performance of new Euler-Euler models with the baseline model set. Results  
53 of a demonstration case are then presented in Section 5 where we test a new  
54 lift force correlation against the current baseline model. We describe the per-  
55 formance of the workflow and present validation graphs of selected cases to  
56 demonstrate the fuzzy logic evaluation. An overview of fuzzy logic results for  
57 the entire case collection aims to demonstrate the potential of such artificial  
58 intelligence tools for a sustainable Euler-Euler model development. Conclu-  
59 sions and an outlook to where our future work is going is given in Section  
60 6.

## 61 2. Background

### 62 2.1. Conventional Euler-Euler model development

63 The description of two-phase flows with the Eulerian-Eulerian methodol-  
64 ogy requires many closure models. Closure models need to describe all the  
65 detailed flow phenomena, which the usually coarse computational grid can-  
66 not resolve. For polydisperse bubbly flows a dozen such closures is needed:  
67 interfacial force closures, i.e. drag, shear lift, turbulent dispersion, virtual  
68 mass and wall lubrication forces [7], turbulence closures for the shear and  
69 bubble-induced turbulence of the liquid phase [8], and closures describing  
70 the polydispersity of bubbles, i.e. a population balance model, and models  
71 for both coalescence and breakup events [9]. A lot of different model options  
72 for each of these closures have emerged in the literature, each of which comes  
73 with a multitude of adjustable model parameters.

74 Table 1 lists a selection of such models suitable for the modeling of bubbly  
75 flows, part of which are provided by OpenFOAM [10] and by the *HZDR*  
76 *Multiphase Addon for OpenFOAM* [11]. Listed are five drag models, three lift  
77 models, three models for turbulent dispersion, two for virtual mass and four  
78 for wall lubrication, three different turbulence models for the liquid and three  
79 different models for bubble-induced turbulence, two different approaches for  
80 population balance modeling, five different coalescence and four different  
81 binary breakup models. These 34 options alone give a total of 129600 possible  
82 model combinations. This number becomes even more extensive considering  
83 all the models in the literature not yet included in the HZDR model library.  
84 Numerous model parameters inherent in every sub-model further add to the  
85 complexity. The description of polydisperse bubbly flows using the Euler-  
86 Euler approach is inarguably based on a highly complex set of intricately-  
87 linked models making the interpretation of CFD results extremely difficult.

Table 1: Closure models for polydisperse bubbly flow available with the *HZDR Multiphase Addon for OpenFOAM* [11]. The current baseline model set is indicated in bold text.

Required closures		Model options	Parameter
Interfacial forces	Drag	<b>Ishii and Zuber [12]</b>	-
		Schiller and Naumann [13]	-
		Tomiyama et al. [14]	-
		Tomiyama et al. [15]	-
		Tomiyama et al. [16]	-
Shear lift	<b>Tomiyama et al. [17]</b>	Legendre and Magnaudet [18]	-
		Moraga et al. [19]	-
		<b>Burns et al. [20]</b>	1
Turbulent dispersion	<b>Gosman et al. [21]</b>	Gosman et al. [21]	1
		Lopez de Bertodano [22]	1
		<b>Crowe et al. [23]</b>	1
Virtual mass	<b>Lamb [24]</b>	Lamb [24]	-
		<b>Hosokawa et al. [25]</b>	-
Wall lubrication	<b>Antal et al. [26]</b>	Antal et al. [26]	2
		Frank [27]	3
		Tomiyama [28]	1
		<b>k-omega-SST [29]</b>	12
Turbulence	Shear-induced	k-epsilon [30]	6
		k-omega [31]	5
		<b>Ma et al. [32]</b>	-
Bubble-induced	<b>Rzehak and Krepper [33]</b>	Rzehak and Krepper [33]	-
		Sato and Sadatomi [34]	1
		<b>Class method [35]</b>	-
Poly-dispersity	Population balance	Ishii et al. [36]	-
	Coalescence	<b>Liao et al. [9]</b>	9
Coulaloglou and Tavlarides [37]		2	
Lehr et al. [38]		2	
Luo [39]		2	
Prince and Blanch [40]		3	
Breakup	<b>Liao et al. [9]</b>	Lehr et al. [38]	4
		Laakkonen et al. [41]	-
		Laakkonen et al. [41]	4
		Luo and Svendsen [42]	3

88 There is no consensus about the use of any of those closure models. In the  
89 literature usually a different set of models is chosen by each research group  
90 [43, 1, 7, 44]. Model constants tend to be tuned for each specific validation  
91 case. Despite a large number of publications on the model development  
92 for the last two decades little progress is seen towards Eulerian-Eulerian  
93 simulations with reliable predictive abilities [6].

## 94 2.2. The ‘Baseline’ Strategy

95 The aim of the baseline strategy at HZDR is to arrive at a single univer-  
96 sal set of Eulerian-Eulerian two-fluid models that can predict bubbly flows  
97 in any flow configuration. Baseline models should offer a certain range of  
98 applicability regarding the local flow characteristics. The baseline model set  
99 is fixed, as are the model constants. For different flow situations the contri-  
100 bution of each of the phenomena listed in Table 1 will of course vary, but the  
101 baseline model set should predict such changes without any further tuning.  
102 As described by Lucas et al. [6] for the sake of generality the models chosen  
103 for each phenomena should preferably be mechanistic models that are based  
104 on local flow conditions rather than empirical models derived from observa-  
105 tions made in a specific experiment. The baseline model set should produce  
106 convincing results for a large range of different validation cases in order to  
107 prove its predictive abilities for future cases.

108 Table 1 lists the current set of baseline models for polydisperse bubbly  
109 flow at HZDR in bold letters. Find the detailed equations for these models  
110 in Liao et al. [8]. Over the years the models listed here have proven to  
111 repeatedly produce good results for a variety of cases [9, 7, 8].

112 Figure 1 illustrates the general process of updating this fixed set of base-  
113 line models in case a new model has been developed or found in the literature.  
114 In a first step the new model needs to be analysed regarding the model re-  
115 quirements of the baseline model set. Various features need to be checked,  
116 such as a physical and mechanistic basis of the model, its generality or its  
117 derivation from advanced experiments. If the advantages of the new model  
118 become apparent, it will then be tested in a next step. This part of the  
119 process is essential as it aims to demonstrate an overall improvement of the  
120 CFD predictions for a collection of cases. This case collection contains an  
121 extensive selection of validation cases, which will be described in the next  
122 section. Generally, a new model will replace the current baseline model if an  
123 improvement for the majority of our validation cases can be demonstrated.

124 During the continuous model development we try to identify deficiencies of  
 125 models, in which we should invest our efforts for further improvement.

126 As illustrated on the right hand side of the flow diagram in Figure 1 a  
 127 model update can be considered even if overall case results do not improve.  
 128 The advantages of a new model then need to outweigh the deterioration of  
 129 results. The coupling of physical phenomena is complex making a straight-  
 130 forward model evaluation impossible at times. Worse results obtained with  
 131 a better model can be caused by the interdependence of the current set of  
 132 baseline models, where potential shortcomings of one model are possibly  
 133 compensated by the tuning of another.

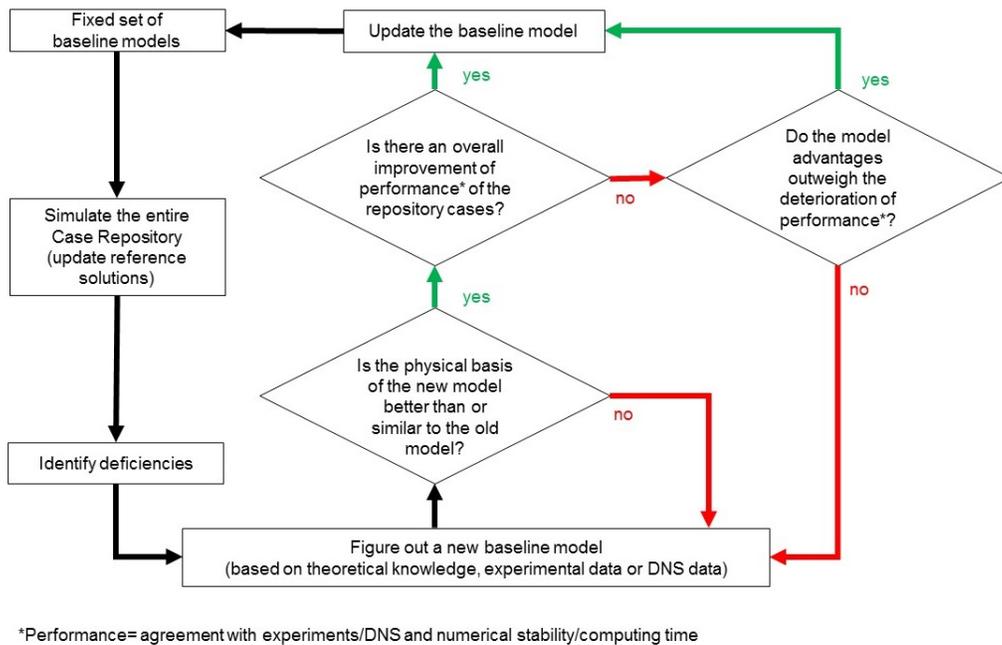


Figure 1: Schematic of the process employed to update baseline models.

### 134 2.3. The current collection of cases at HZDR

135 As described above, the process of testing new models involves the simu-  
 136 lation of a large number of validation cases. The case collection at HZDR  
 137 currently contains 56 pipe and bubble column cases, with the case number  
 138 continuously growing. Cases differ regarding their geometry, gas injection,  
 139 bubble sizes, water quality etc. Keywords were introduced to categorise the

140 cases and their sub-cases as illustrated in Table 2. The keywords illustrated  
141 are not complete, but rather form a starting point to label existing cases.  
142 More cases are to be added in the future in order to diversify the collection  
143 and its keywords, as well as to simply increase the amount of validation data.

144 Case setups are standardised using the best-practice guidelines given in  
145 Lucas et al. [45] and simulated using the *HZDR Multiphase Addon for Open-*  
146 *FOAM*, which was customized and extended for the various model develop-  
147 ments at HZDR. All cases are under Git version control and managed in  
148 GitLab [46]. GitLab Continuous Integration and Continuous Development  
149 (CI/CD) mechanisms allow an efficient maintenance of the case collection,  
150 as well as its extension by new cases.

Table 2: Overview of cases with all active keywords.

Cases		Geo-metry		Gas in-jection		Gas phases		Bubble size change		Bubble sizes		Water quality			State		Devel-opment		
		Pipe	Bubble column	Bottom	Wall	one-gas	multiple-gas	Fixed	Density-change	Coalescence-breakup	Monodisperse	Polydispersse	Filtered	Tap water	Contaminated	Steady-state	Transient	Fully developed	Developing
Wang et al. (1987)	W1	X		X		X		X		X		X			X		X		
	W2	X		X		X		X		X		X			X		X		
	W3	X		X		X		X		X		X			X		X		
Liu and Bankoff (1993)	LB1	X		X		X		X		X		X			X		X		
	LB2	X		X		X		X		X		X			X		X		
Liu (1998)	L11A	X		X		X		X		X			X		X		X		
	L21B	X		X		X		X		X			X		X		X		
	L21C	X		X		X		X		X			X		X		X		
	L22A	X		X		X		X		X			X		X		X		
Pfleger (1999)		X	X			X			X	X		X			X			X	
Deen (2001)		X	X		X		X		X				X		X			X	
Hibiki (2001)	H11	X		X		X		X		X		X			X		X		
	H12	X		X		X		X		X		X			X		X		
	H13	X		X		X		X		X		X			X		X		
	H31	X		X		X		X		X		X			X		X		
	H32	X		X		X		X		X		X			X		X		
	H33	X		X		X		X		X		X			X		X		
Shawkat (2008)	S21	X		X		X		X		X		X			X		X		
	S23	X		X		X		X		X		X			X		X		
	S31	X		X		X		X		X		X			X		X		
	S33	X		X		X		X		X		X			X		X		
Hosokawa (2009)	H11	X		X		X		X		X			X		X		X		
	H12	X		X		X		X		X			X		X		X		
	H21	X		X		X		X		X			X		X		X		
	H22	X		X		X		X		X			X		X		X		
Akbar et al. (2012)	13		X	X			X			X		X				X		X	
	3		X	X		X		X		X		X				X		X	
Hosokawa and Tomiyama (2013)	H1	X		X		X		X		X			X		X		X		
	H2	X		X		X		X		X			X		X		X		
	H3	X		X		X		X		X			X		X		X		
	H4	X		X		X		X		X			X		X		X		
Kim (2016)	K1	X		X		X		X		X			X		X		X		
	K2	X		X		X		X		X			X		X		X		
	K3	X		X		X		X		X			X		X		X		
	K4	X		X		X		X		X			X		X		X		
Lucas et al. (2005)	fd 19	X		X		X		X			X	X			X		X		
	fd 37	X		X		X		X			X	X			X		X		
	fd 39	X		X		X		X			X	X			X		X		
	fd 41	X		X		X		X			X	X			X		X		
	fd 59	X		X		X		X			X	X			X		X		
	fd 61	X		X		X		X			X	X			X		X		
	fd 63	X		X		X		X			X	X			X		X		
	fd 85	X		X		X		X			X	X			X		X		
	d 19	X		X		X			X	X		X	X		X			X	
	d 37	X		X		X			X	X		X	X		X			X	
	d 39	X		X		X			X	X		X	X		X			X	
	d 41	X		X		X			X	X		X	X		X			X	
	d 59	X		X		X			X	X		X	X		X			X	
	d 61	X		X		X			X	X		X	X		X			X	
	d 63	X		X		X			X	X		X	X		X			X	
	d 85	X		X		X			X	X		X	X		X			X	
	Lucas et al. (2010)	30	X			X	X			X		X	X			X			X
		40	X			X	X			X		X	X			X			X
41		X			X	X			X		X	X			X			X	
42		X			X	X			X		X	X			X			X	
52		X			X	X			X		X	X			X			X	

151 *2.4. The challenge of evaluating new models*

152 Each case in the case collection produces results that are compared to  
153 validation data, i.e. experimental data. Some fully-developed cases produce  
154 steady-state results, for other transient cases the results need to be time-  
155 averaged. Typical graphs for analyzing results are gas void fraction profiles,  
156 liquid and gas velocity profiles, sometimes profiles of turbulent quantities,  
157 such as the turbulent kinetic energy. Transient cases produce the respective  
158 mean values of the same quantities.

159 In case a new model is tested the computed results additionally need  
160 to be compared with the old baseline reference solution. For each of the  
161 plots various curves need to be compared: the computed result using the  
162 new model, the reference result obtained with the current baseline model set  
163 and the measured data for validation. Examples of such graphs are given  
164 in Figure 2. The reason why this kind of evaluation is a time-consuming  
165 challenge is twofold:

166 First, the sheer number of results to evaluate is extensive and hard to  
167 handle. It is difficult to demonstrate an overall improvement for the total  
168 number of currently 56 cases, each of which produces three to five graphs.  
169 Assuming only three graphs for each case, a total of 168 plots has to be  
170 evaluated. These plots are typically scattered across case folders. The issue  
171 becomes even more pressing with an increasing number of cases and plots,  
172 which for baseline development purposes is intended. Therefore, a **workflow**  
173 for an automated production of such graphs and their neat presentation in a  
174 single report is developed, as will be presented in Section 3.

175 Second, the information in some graphs can be vague and the decision  
176 if there is an improvement of results can become rather subjective. Plotted  
177 results need to be evaluated and compared, which is typically done during  
178 long and difficult discussions among experts. For a better qualitative judge-  
179 ment and quantification of the performance of new models we will therefore  
180 introduce a **fuzzy logic approach** in Section 4 that is meant to aid the  
181 decision-making process.

### 182 3. Automation of cases in a workflow

183 Workflow engines are designed to automate the successive execution of  
184 commands and applications for large-scale data analyses. The generation of  
185 data via workflows follows so-called FAIR principles [47] for a good manage-  
186 ment of scientific data, which should be transparent and reproducible. Out  
187 of a large number of potential workflow systems (such as [48] etc.) for the  
188 work presented here we chose Snakemake and developed a workflow, which  
189 will be described in the next sections.

#### 190 3.1. The Snakemake library

191 Snakemake is a general purpose Python library originally developed for  
192 applications in the field of bio-informatics [4]. The library allows to build  
193 complex algorithms that process large data-sets using scientific software, and  
194 manages the resulting output.

195 As described by Evdokimov et al. [49] the Snakemake library provides  
196 several features that make it a suitable framework for managing the extensive  
197 amount of CFD simulations we are interested in here, and their subsequent  
198 analysis:

- 199 • The syntax of Snakemake offers great flexibility in terms of the scripting  
200 languages that jobs can be composed of, which allows individual case-  
201 by-case setups.
- 202 • The library comes with interfaces to common cluster schedulers re-  
203 quired for high-performance computing.
- 204 • Snakemake was built for parallization enabling us to spawn parallel case  
205 jobs on separate nodes of a cluster to maximize speed. Furthermore,  
206 it can adapt the number of jobs in execution to the allocated workflow  
207 resources.
- 208 • Its modularity allows to rerun downstream analysis and re-process sub-  
209 sets of jobs without the need to rerun the entire pipeline.
- 210 • The post-processing output is aggregated and put into a single place  
211 with the possibility to include results of hundreds of cases.
- 212 • On the low-level a case-by-case approach allows to easily split work  
213 among researchers in a team, and

- 214 • On the top-level common CI/CD software design practices are applica-  
215 ble and allow the consistent integration of the library into our existing  
216 case repository.

### 217 3.2. Workflow overview

218 The details of the workflow on both the case- and the top-level, are de-  
219 scribed and illustrated in Evdokimov et al. [49]. The following section will  
220 briefly summarize the main steps and workflow specifications relevant for our  
221 baseline model development.

222 The workflow consists of the following three steps reflecting a *bottom-up*  
223 method of determining job execution:

- 224 1. During a *configuration step* an algorithm searches in sub-folders for  
225 **Snakefiles** that define the specifications of the pre-processing, solv-  
226 ing and post-processing job for each individual case. **Snakefiles** are  
227 designed in such a way that they allow the independent workflow ex-  
228 ecution of a selection of cases. Larger cases that include extensive  
229 parameter-variations (e.g. Lucas et al. [50] with hundreds of systematic  
230 case setup variations) are setup via templates building file structures  
231 and sub-cases. Once the *configuration* step has completed, the estab-  
232 lished file structure is ready for launching the workflow.
- 233 2. The workflow then runs the selected cases during the *solution step*.  
234 Workflow reports containing the plots of all the cases are produced  
235 using the report-generating feature embedded in the Snakemake library.
- 236 3. A third optional *post-processing* step produces a case-by-case overview  
237 of the agreement of the computed results with the validation data. This  
238 step is particularly important for the testing of new baseline models and  
239 incorporates the fuzzy logic system described in Section 4. Computed  
240 results are not only compared to the validation data, but also to the  
241 reference solution obtained by applying the baseline model set.

### 242 3.3. Workflow reports

243 Snakemake reports are generated in the form of static web-pages, which  
244 are easily deployed and shared on a web-server. CI/CD mechanisms for  
245 the workflow execution allow deploying reports onto web-servers as soon as  
246 a certain job has finished making results instantly accessible to the whole  
247 research team. A demonstration of the final workflow report via screenshots  
248 is presented in Evdokimov et al. [49]. A side panel lists all cases simulated,

249 which upon selection will show the corresponding validation plots on the  
250 main page for inspection. For further illustration purposes we refer to the  
251 demonstration report provided by the Snakemake library on a public web-  
252 page [51]. This report gives an impression of the typical report structure for  
253 results.

254 The Snakemake workflow gives us an efficient tool to produce a large  
255 amount of plots and to represent them in a well-structured report. The  
256 evaluation of these reports for the baseline model development remains the  
257 task of the user, but can be supported by artificial intelligence tools, such as  
258 the fuzzy logic controller described in the next section.

259 **4. Evaluation of results**

260 The decision whether or not a computed *simulation* result is better than  
 261 the *baseline* reference solution can be quite easy to make, such as in the plot  
 262 illustrated on the left hand side of Figure 2, or it can be less obvious, such  
 263 as the example on the right hand side.

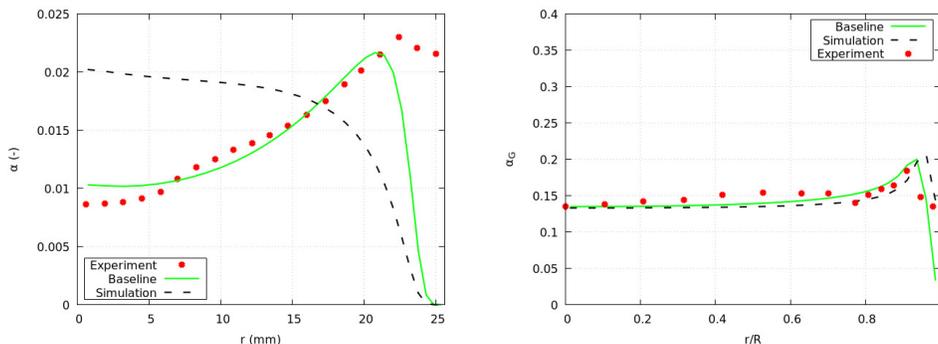


Figure 2: Gas void fraction profiles of fullyDeveloped\_39 (left) and case L11A (right).

264 Fuzzy logic is a subset of artificial intelligence, which can accommodate  
 265 the imprecision of the real world and support such a decision-making process.  
 266 By using linguistic variables fuzzy logic can deal with imprecise, vague and  
 267 uncertain information and approximate human reasoning.

268 The `scikit-fuzzy` library ([52], [53]) allows the integration of such fuzzy  
 269 logic tools into our workflow, where it can be utilized for the testing of new  
 270 models. From the system design point of view the library takes responsibility  
 271 for the correct and robust implementation of the underlying algebra and the  
 272 numerical algorithms, while on the top level we investigate metrics and their  
 273 value boundaries suitable to our multiphase CFD problems. In essence, we  
 274 propose a *fuzzy logic controller* for the purpose of evaluating and comparing  
 275 CFD results. In the next sections we briefly describe how this controller  
 276 works.

277 *4.1. The fuzzy logic system*

278 Fuzzy logic uses a specific terminology, which is briefly introduced here  
 279 for a better understanding of the following sections. When a value is re-  
 280 ferred to as being crisp this means it is explicit and concise. *Fuzzification*  
 281 is the process of decomposing such crisp values into a spectrum of different

282 linguistic categories, referred to as fuzzy sets. A typical fuzzy logic system  
 283 is illustrated in Figure 3. The system consumes and *fuzzifies* several crisp  
 284 inputs, which means numerical floating-point or integer values are *mapped*  
 285 onto linguistic sets via *membership functions*. These *membership functions*  
 286 are normalized and might be considered as weight-factors of the expert sys-  
 287 tem. *Inference logic* combines the numerous *fuzzy input sets* via a set of  
 288 rules to produce *fuzzy output sets*, which themselves represent linguistic sets.  
 289 Finally, these output sets are summed up, *defuzzified* and mapped onto a  
 290 single crisp output value.

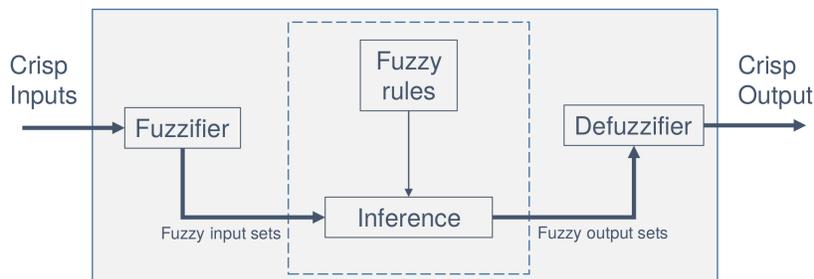


Figure 3: Schematic of a fuzzy logic system.

291 The inputs specified for the fuzzy system demonstrated here are two error  
 292 metrics that will be described in the next section. One of the advantages of  
 293 the fuzzy logic approach lies in its modularity allowing us to extend this  
 294 list of fuzzy input variables in the future. In our case the desired output  
 295 of the fuzzy logic system is a value between 0 and 1 quantifying how good  
 296 the computed CFD result agrees with the measured validation data. We will  
 297 refer to this crisp output as **goodness-value**  $G$  of the CFD prediction.

#### 298 4.2. The crisp inputs: error metrics

299 The crisp inputs for our fuzzy logic system are specific error metrics that  
 300 describe the similarity of two curves: our computed result, such as the gas  
 301 void fraction profile, and the corresponding experimental data for validation.  
 302 Various metrics could be used, but for demonstration purposes the number  
 303 here is limited to two, which complement each other. The error metrics below  
 304 are defined using the example of evaluating sorted sets of gas void fraction  
 305 data. Hereby, the sorted set represents the radial profile in case of round

306 geometries or the lateral profile in case of rectangular columns. Metrics for  
 307 the other validation fields are computed accordingly.

308 As a quantitative measure the **Mean relative error MRE** is defined  
 309 as:

$$\text{MRE} = \frac{1}{N} \sum_{n=1}^N \frac{|\alpha_{n,sim} - \alpha_{n,exp}|}{\bar{\alpha}_{exp}} \quad (1)$$

310 with  $N$  being the number of values, the computed values of the gas void  
 311 fraction  $\alpha_{sim}$  and the experimentally measured ones  $\alpha_{exp}$ , and  $\bar{\alpha}$  the arith-  
 312 metic average. The MRE describes the quantitative distance between two  
 313 curves along the ordinate. It is normalized by the mean experimental value  
 314 for the entire profile in order to avoid a heavier penalty of discrepancies for  
 315 data points with smaller experimental values. Possible values for the MRE  
 316 range from 0 to  $\infty$ . We therefore limit values to a maximum of 1. Note that  
 317 a case with a  $\text{MRE} > 1$  therefore will not receive a heavier penalty than cases  
 318 with a  $\text{MRE} = 1$ . Most cases in our collection, however, fall into the range  
 319 of  $0 < \text{MRE} < 1$ .

320 As a qualitative measure the **Pearson correlation coefficient PCC** is  
 321 defined as the covariance of the two gas void fraction data sets divided by  
 322 the product of their standard deviation:

$$\text{PCC} = \frac{\sum_{n=1}^N (\alpha_{n,sim} - \bar{\alpha}_{sim})(\alpha_{n,exp} - \bar{\alpha}_{exp})}{\sqrt{\sum_{n=1}^N (\alpha_{n,sim} - \bar{\alpha}_{sim})^2} \sqrt{\sum_{n=1}^N (\alpha_{n,exp} - \bar{\alpha}_{exp})^2}} \quad (2)$$

323 The Pearson correlation coefficient determines the similarity of the two  
 324 data sets, i.e. the shapes of the two curves. The range of the Pearson  
 325 coefficient is  $-1 < \text{PCC} < 1$ . Examples of PCC values are illustrated in  
 326 Figure 4, where simulation profiles are compared with experimental data. A  
 327 negative coefficient indicates a negative linear correlation and a low similarity  
 328 between curves, such as in the case of a predicted wall peak for a measured  
 329 core peak profile. A positive coefficient represents a positive linear correlation  
 330 of the data set meaning a high similarity of shapes, as illustrated for the  
 331 almost matching profiles on the right hand side.

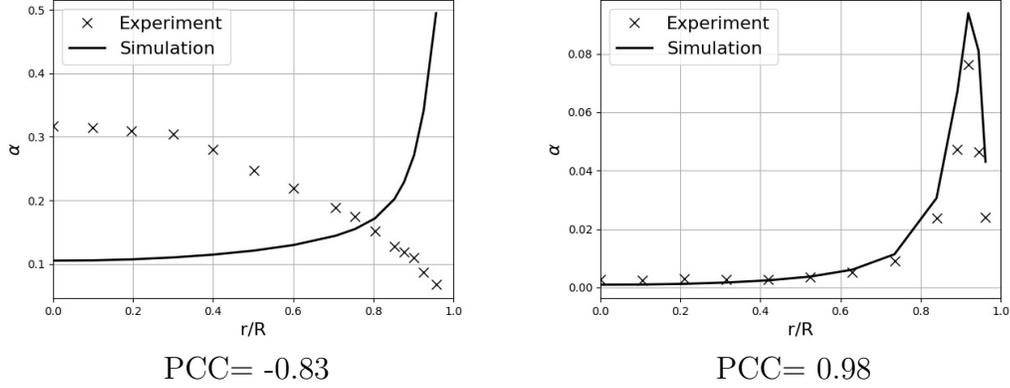


Figure 4: Pearson correlation coefficients computed for different data sets of gas void fraction profiles.

332 The PCC criteria is not sensitive to shifts of the profiles along the or-  
 333 dinate, but sensitive to shifts along the abscissa. Thus, it complements the  
 334 MRE metric, which is mainly sensitive to shifts along the ordinate.

### 335 4.3. Membership functions

336 A fuzzy set of linguistic terms is assigned to each input and output metric  
 337 as illustrated in Table 3. The process of turning crisp variables into linguistic  
 338 sets is termed *fuzzification*.

Table 3: Linguistic sets assigned to each variable.

Variable	Linguistic sets	Parameters a,b,c
MRE	"low"	0, 0, 0.3
	"medium"	0, 0.3, 1.0
	"high"	0.3, 1.0, 1.0
PCC	"high"	0.4, 1.0, 1.0
	"medium"	-1.0, 0.4, 1.0
	"low"	-1.0, -1.0, 0.4
G	"perfect"	0.7, 1.0, 1.0
	"good"	0.5, 0.7, 0.9
	"tolerable"	0.3, 0.5, 0.7
	"bad"	0.1, 0.3, 0.5
	"defect"	0.0, 0.0, 0.3

339 Membership functions  $\mu_i(x)$  tie the input and output variables  $x$  for each  
 340 linguistic set item  $i$  of the above table to a number, which can be understood  
 341 as an impact value. Piece-wise *triangle* functions are used:

$$\mu_i(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ \frac{c-x}{c-b} & b \leq x \leq c \\ 0 & c \leq x \end{cases} \quad (3)$$

342 with parameters  $a$ ,  $b$  and  $c$  specified separately for each item of the set  
 343 as listed in Table 3. The triangle-shaped membership functions assigned to  
 344 the linguistic terms listed in Table 3 are illustrated in Figure 5.

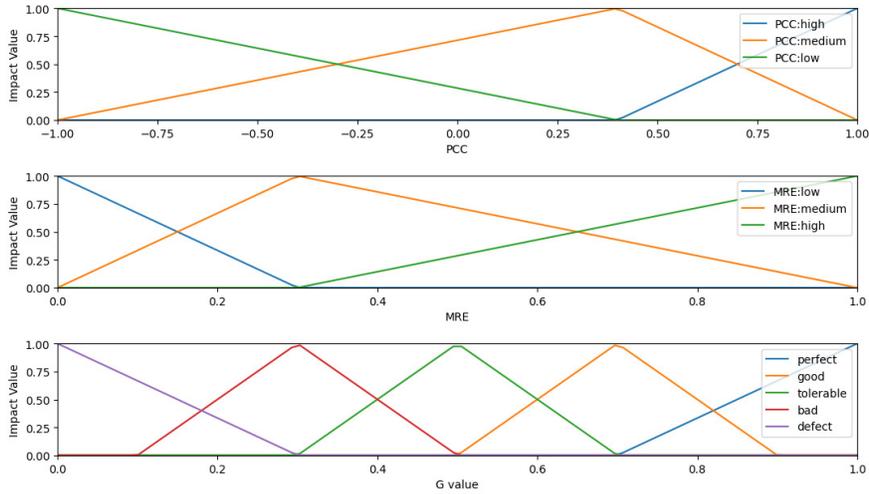


Figure 5: Membership functions for the fuzzy input and output sets. The  $x$  axis represents the crisp value space, the  $y$  value represents the membership function output value for each linguistic term.

345 For a specific metric value the membership function output represents  
 346 a fuzzy degree of membership in the qualifying linguistic set. Thus, value  
 347 intervals are constrained to a uniform range from zero to one for all input  
 348 and output metrics, and the fuzzy set is *normalized*. For a crisp value  $x$  a  
 349 fuzzy set now is defined as a pair:

$$A(x) = (i, \mu_i(x)) \quad (4)$$

350 where item  $i$  is a linguistic term and  $\mu_i(x)$  the membership function ex-  
351 pressing the degree of membership in that term.

#### 352 4.4. Fuzzy rule base

353 Operations can be defined on the various fuzzy sets by means of their  
354 membership function. This allows us to prescribe an intuitive rule base,  
355 which is easy to interpret.

356 The current workflow evaluation consists of the following five rules:

- 357 1. If PCC:low AND MRE:high  $\rightarrow$  Defect G
- 358 2. If (PCC:low AND MRE:medium) OR (PCC:medium AND MRE:high)  $\rightarrow$   
359 Bad G
- 360 3. If (PCC:medium AND MRE:medium)  
361 OR (PCC:high AND MRE:high)  
362 OR (PCC:low AND MRE:low)  $\rightarrow$  Tolerable G
- 363 4. If (PCC:high AND MRE:medium) OR (PCC:medium AND MRE:low)  $\rightarrow$   
364 Good G
- 365 5. If PCC:high AND MRE:low  $\rightarrow$  Perfect G

366 A simplified visual representation for the target "Goodness" metric  $G$  is  
367 illustrated in Figure 6.

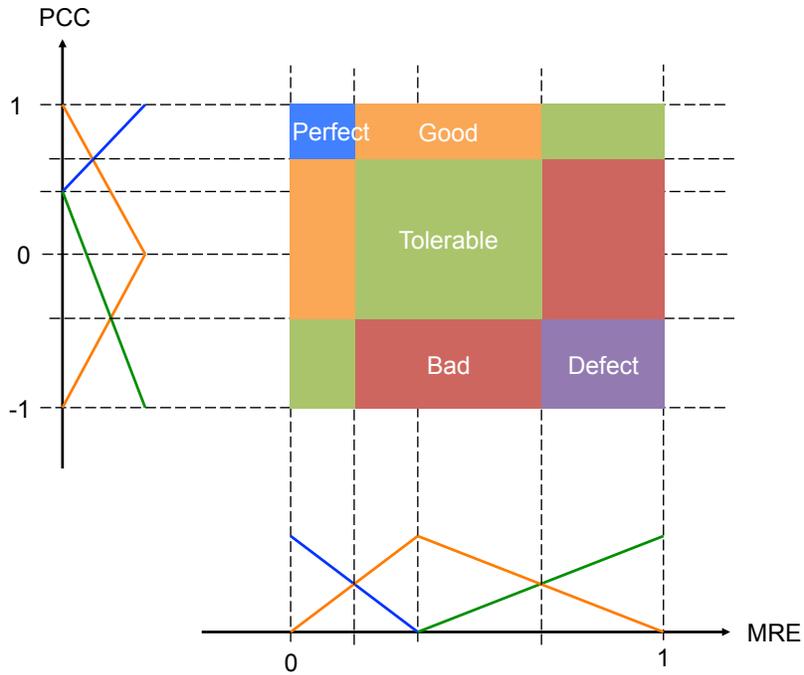


Figure 6: A schematic view of the relationship between the fuzzy input metrics and the output metric  $G$ .

368 Fuzzy operators are applied to obtain one value representing the result  
 369 of each rule. Logical *min/max* operators correspond to the AND/OR rules  
 370 mentioned above [54]. Input for these operators are the membership values  
 371 of the fuzzified input variables, and the output is a single value.

372 An example is illustrated in Figure 7 for a case with  $PCC = 0.49$  and  
 373  $MRE = 0.36$ . A single fuzzy rule, rule 4 for the fuzzy set *Good* of the  
 374  $G$  metric, is exemplified. The condition (PCC:high AND MRE:medium) OR  
 375 (PCC:medium AND MRE:low) can be translated into  $\max(\min(0.15, 0.91),$   
 376  $\min(0.85, 0)) = 0.15$ . This output is used for the implication of the fuzzy  
 377 rule as described in the next section.

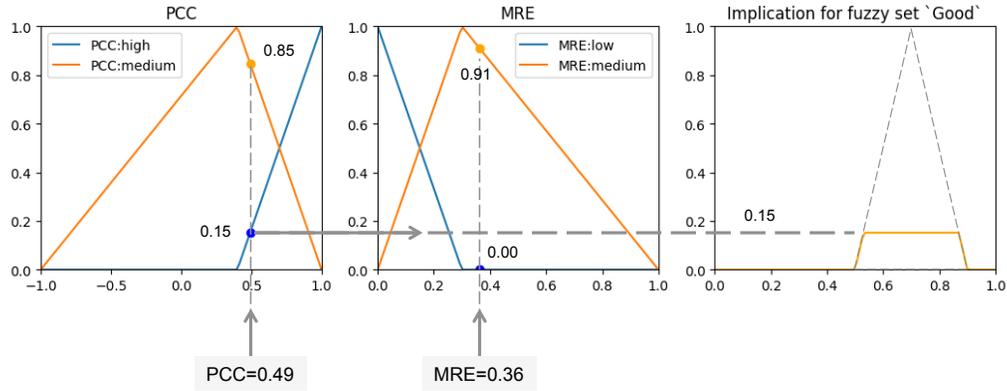


Figure 7: Application of rule 4 via fuzzy operators for an example of two input metrics.

378 *4.5. Implication and defuzzification*

379 An implication method turns the single output value into the consequence  
 380 of the fuzzy rule. The consequent fuzzy sets are represented by the member-  
 381 ship functions of the output sets. Input for the implication method is the  
 382 value outputs given by each of the antecedent rules, such as rule 4 above.

383 Mamdani-type inference was applied for the implication process as it is an  
 384 intuitive method, well-suited to human input, allowing an interpretable rule  
 385 base and with widespread acceptance [54]. It works with the *min* implication  
 386 operator to infer the output functions, as it is illustrated for rule 4 on the right  
 387 hand side of Figure 7. The five fuzzy rules produce the output functions for  
 388 the five potential sets of the goodness variable. Figure 8 illustrates the whole  
 389 inference process for the above example with  $PCC = 0.49$  and  $MRE = 0.36$ .  
 390 In the example only three out of five output sets qualify during the inference  
 391 step.

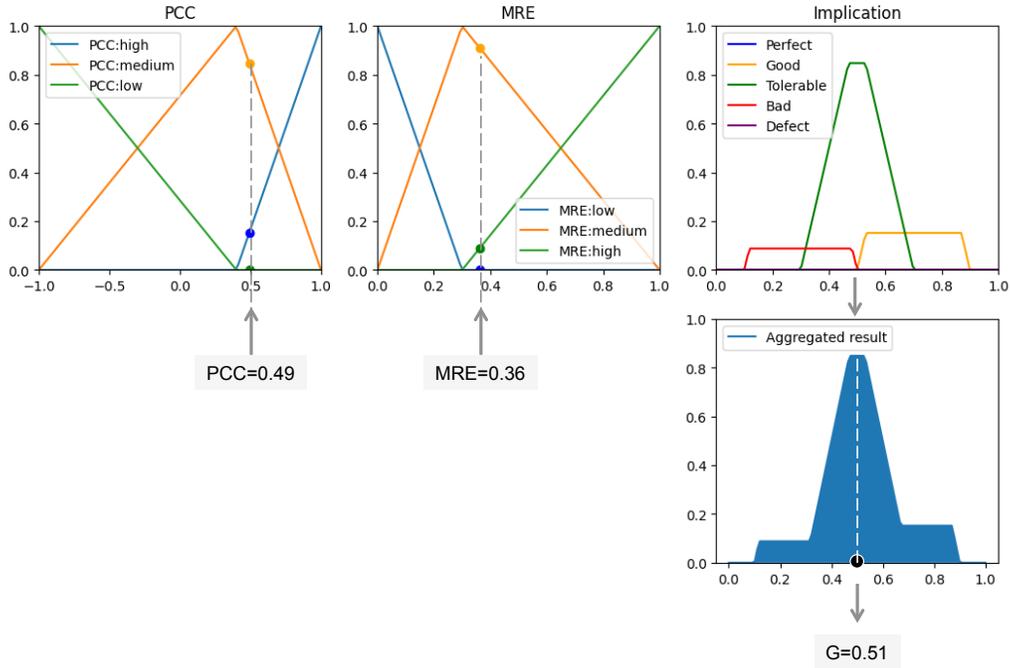


Figure 8: Inference system including all rules for an example of two input metrics.

392 The consequent functions of all rules then need to be aggregated, which  
 393 is done by a simple *max* function. In a last step the aggregated function is  
 394 defuzzified. The `scikit-fuzzy` library allows to choose from several meth-  
 395 ods suggesting different trade-offs between the smoothness of the response  
 396 surface and the robustness of the final defuzzification step. In the model  
 397 evaluation attempted here we tried to track the smallest variations to allow  
 398 a continuous differentiation of results. Therefore, our chosen defuzzification  
 399 method relies on area measurements of the aggregated result. Defuzzification  
 400 is performed by the *bisector* method, which finds the vertical line that di-  
 401 divides the aggregated result into two sub-regions of equal area. The response  
 402 surface of our current two-input, one-output, five-rule fuzzy logic system is  
 403 illustrated in Figure 9.

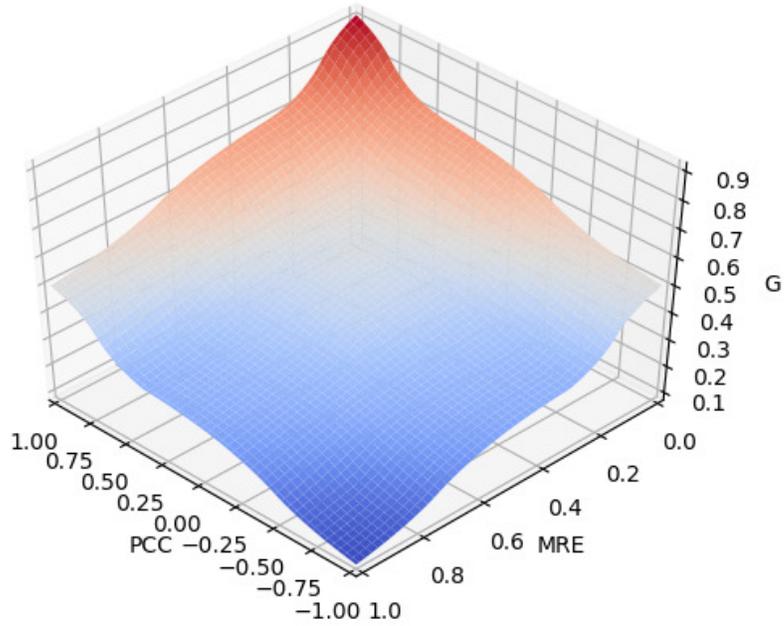


Figure 9: Response surface of the fuzzy logic system.

404 *4.6. Evaluation for the case collection*

405 The fuzzy logic approach offers a convenient framework for an evaluation  
 406 algorithm, which could be easily repurposed, extended and redefined. The  
 407 alternative usage of an equivalent multidimensional target function requires  
 408 the strict definition of potentially controversial input values fitting a specific  
 409 output metric. Fuzzy logic makes it easy to reach a consensus amongst re-  
 410 searchers by using relations between linguistic sets derived from these values.  
 411 Errors divided into a *low*, *medium* and *high* subset are easily agreed on, but  
 412 certain values fitting these sets are not.

413 As described above for each case in the case collection several fields are  
 414 investigated for the validation of the CFD results, e.g. gas void fractions or  
 415 velocity profiles. A goodness value  $G$  is computed for each of those individual  
 416 validation fields and is meant to describe how good the predicted profile  
 417 agrees with the experimental data.

418 The average over all these individual goodness values for the available  
419 validation fields produces a single value for each case. This single value rep-  
420 represents the overall goodness of the CFD prediction for the case investigated.  
421 In the future different weightings for the various validation fields could be  
422 considered.

423 When testing a new model the above procedure is done twice, for the  
424 results using the new model and for the reference solution obtained with the  
425 baseline model set. An improvement or deterioration of the CFD results for a  
426 specific case can then be expressed via the difference in the overall goodness  
427 of the computed and the baseline result.

428 **5. Demonstration case: Testing a new lift force correlation**

429 To demonstrate the efficiency of the baseline workflow and its fuzzy logic  
 430 tool, this section will now investigate results when testing a new lift force  
 431 correlation. We will refer to the results obtained with the Tomiyama correla-  
 432 tion [17] as *baseline* results, and the results with a new correlation for the lift  
 433 coefficient proposed by Hessenkemper et al. [55] as *computed* results. This  
 434 section does not aim to validate the new correlation, but to demonstrate the  
 435 benefits and future potential of the developed workflow tools in assisting the  
 436 baseline model development.

437 We will give an impression of the performance and reporting features of  
 438 the workflow when analysing a certain selection of cases. The evaluation of  
 439 results via fuzzy logic will then be verified for a selected case. Finally, an  
 440 overview of the overall goodness evaluation for all cases is presented.

441 *5.1. Workflow performance*

442 For the following analysis we restrict our selection of repository cases to  
 443 the ones that have already been published elsewhere, as listed in Table 4. The  
 444 individual case setups in OpenFOAM are described in the references given,  
 445 and are to be found in the *HZDR Multiphase Case Collection for OpenFOAM*  
 446 [56]. Including all sub-cases the total number of cases investigated here is 36.  
 447 The simulation of each individual sub-case demands three to eight processors  
 448 and an execution time of up to two hours.

Table 4: List of cases for the workflow demonstration.

Experimental reference	Cases	Setup	CPUs/Case
Hosokawa and Tomiyama [57]	4	[58]	4
Liu [59]	4	[58], [60]	4
Shawkat et al. [61]	4	[60]	4
Hosokawa and Tomiyama [62]	4	[60], [63]	8
Kim et al. [64]	4	[63]	8
Lucas et al. [50]	16	[65]	3

449 The Snakemake workflow processes all these cases with a total execution  
 450 time of  $\sim 5$ h (the entire case collection takes  $\sim 10$ h). A total of 236 plots  
 451 (baseline relevant and auxiliary) is generated automatically. In the plots  
 452 relevant for the baseline evaluation the computed results are compared to

453 both the experimental data and the reference solution. Examples of such  
454 plots are illustrated in Figure 2. The fuzzy logic evaluation of these results  
455 during the optional post-processing step is done with an execution time of  
456 only four minutes.

## 457 5.2. Verification of the fuzzy logic output

458 For a single case this section will now analyse the computed results and  
459 compare them with the error metrics and the goodness output computed by  
460 the fuzzy logic system. Figure 10 illustrates the validation plots produced  
461 for exemplary case L11A of Liu [59], and associated values of error metrics  
462 and the  $G$  output metric.

463 The predicted gas void fraction profiles show a generally good shape,  
464 both the baseline and the computed profile. This generally good agreement  
465 with the experiment is well represented by high values of the  $PCC$ . The  
466 result with the Hessenkemper lift appears slightly deteriorated due to a shift  
467 of the wall peak towards the wall, which is well captured by a drop of the  
468 respective  $PCC$ . There are no significant changes in  $MRE$ . The drop in  
469 overall goodness and a negative  $\Delta G$  reflect the subjective evaluation of the  
470 gas void fraction profiles.

471 The profiles for the turbulent kinetic energy are examples of curves with  
472 distinctively different shapes compared to the experimental data (wall-peak  
473 vs. core-peak). This judgement is well captured by very low values for the  
474  $PCC$ , as well as very high values for the  $MRE$ . Comparing the baseline  
475 result to the new model no significant changes are observed. Consequently,  
476 the  $\Delta G$  computed by the fuzzy logic system is relatively small.

477 The profiles for the liquid velocity show a good agreement with the exper-  
478 iment in terms of the shape. The difference to the experimental data slightly  
479 increases with the new model, as correctly captured by the  $MRE$  metric.

480 All the resulting  $G$  values of the validation fields are averaged to produce  
481 one value for the overall goodness of the CFD predictions for the case. The  
482 difference in results between the two lift force models is expressed by the  
483 difference  $\Delta G$ . For the L11A case the new lift model slightly decreases the  
484 overall  $G$  and deteriorates results, which corresponds to our judgement for  
485 the plots in Figure 10.

Validation plots	Metrics	Baseline	Simulation	$\Delta G$
	MRE PCC <b>G</b>	0.089 0.701 <b>0.713</b>	0.091 0.349 <b>0.651</b>	  <b>-0.062</b>
	MRE PCC <b>G</b>	1.064 -0.933 <b>0.095</b>	0.962 -0.973 <b>0.099</b>	  <b>0.004</b>
	MRE PCC <b>G</b>	0.066 0.793 <b>0.770</b>	0.082 0.683 <b>0.713</b>	  <b>-0.057</b>
<b>Overall G</b>		<b>0.526</b>	<b>0.487</b>	<b>-0.039</b>

Figure 10: Goodness evaluation for the validation fields of case L11A.

486 *5.3. Overview of all case evaluations*

487 The above evaluation of results expressed via the  $G$  metric is performed  
488 for each of the cases under investigation. Figure 11 illustrates the overall  
489 goodness values of all cases, represented as bars ranging from zero to one.  
490 Numbers next to each case indicate in green the cases which have improved,

491 and in red the deteriorated ones, such as L11A on the top. A  $\Delta G$  value  
 492 averaged over all cases aims to quantify the general trend in results. Thus,  
 493 the plot allows to obtain a first overview of the general performance of a new  
 494 model.

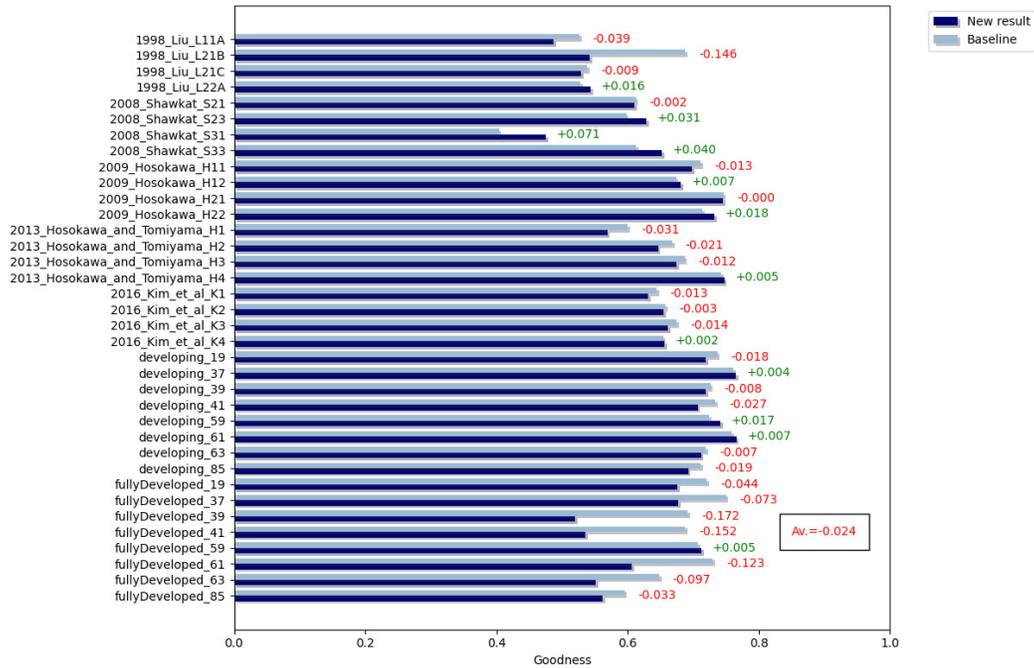


Figure 11: Plots for the overall goodness of all 36 demonstration cases. Light blue bars represent the reference solution obtained with the baseline lift model, deep blue bars the solution obtained with the new lift model. Numbers next to each case indicate the difference in overall goodness, with red indicating a deterioration and green an improvement of results.

495 Similar plots are produced for the individual validation fields, such as the  
 496 gas void fraction plot presented in Figure 12. The goodness values are based  
 497 on the computed mean relative errors and Pearson correlation coefficients  
 498 illustrated in Figure 13.

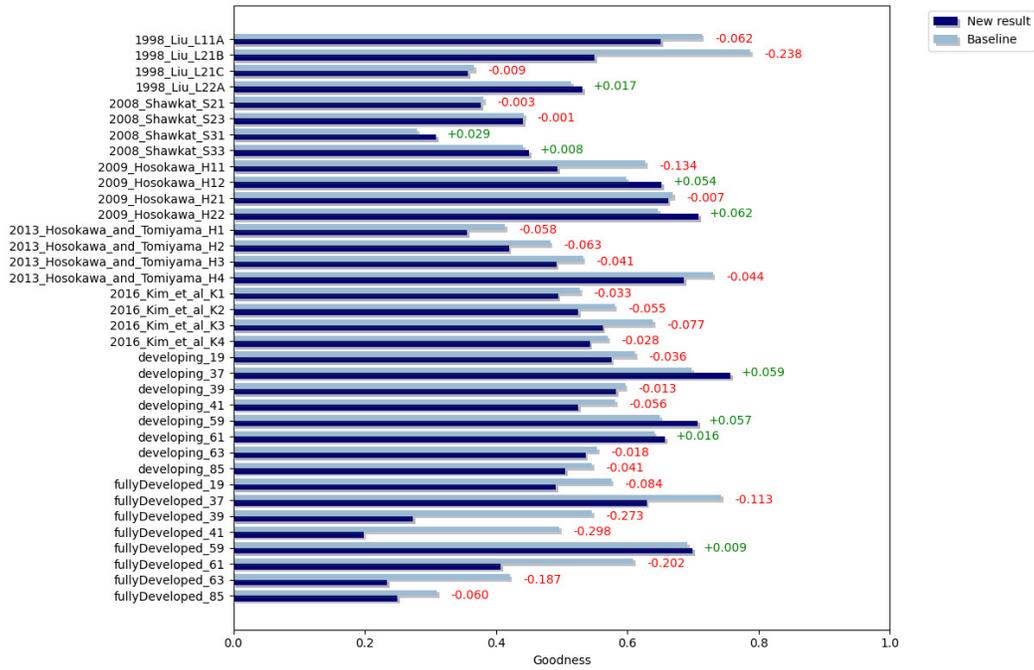


Figure 12: Goodness output for the gas void fraction prediction of all 36 demonstration cases.

499 Plots visualizing the change in goodness for each case, such as the ones  
500 above, have proven to be a useful tool for analysing the performance of a new  
501 model. Not only do they provide a neat case-by-case overview of results,  
502 but also help pointing the researcher to the cases most affected by a new  
503 model. Figure 12 indicates a slight change of gas void fraction results for  
504 case L11A, but a more substantial change for case fullyDeveloped\_39. This  
505 is confirmed by the corresponding plots in Figure 2. Table 5 illustrates the  
506 detailed input and output metrics for both cases, with the quality of CFD  
507 predictions adequately quantified by the fuzzy logic system.

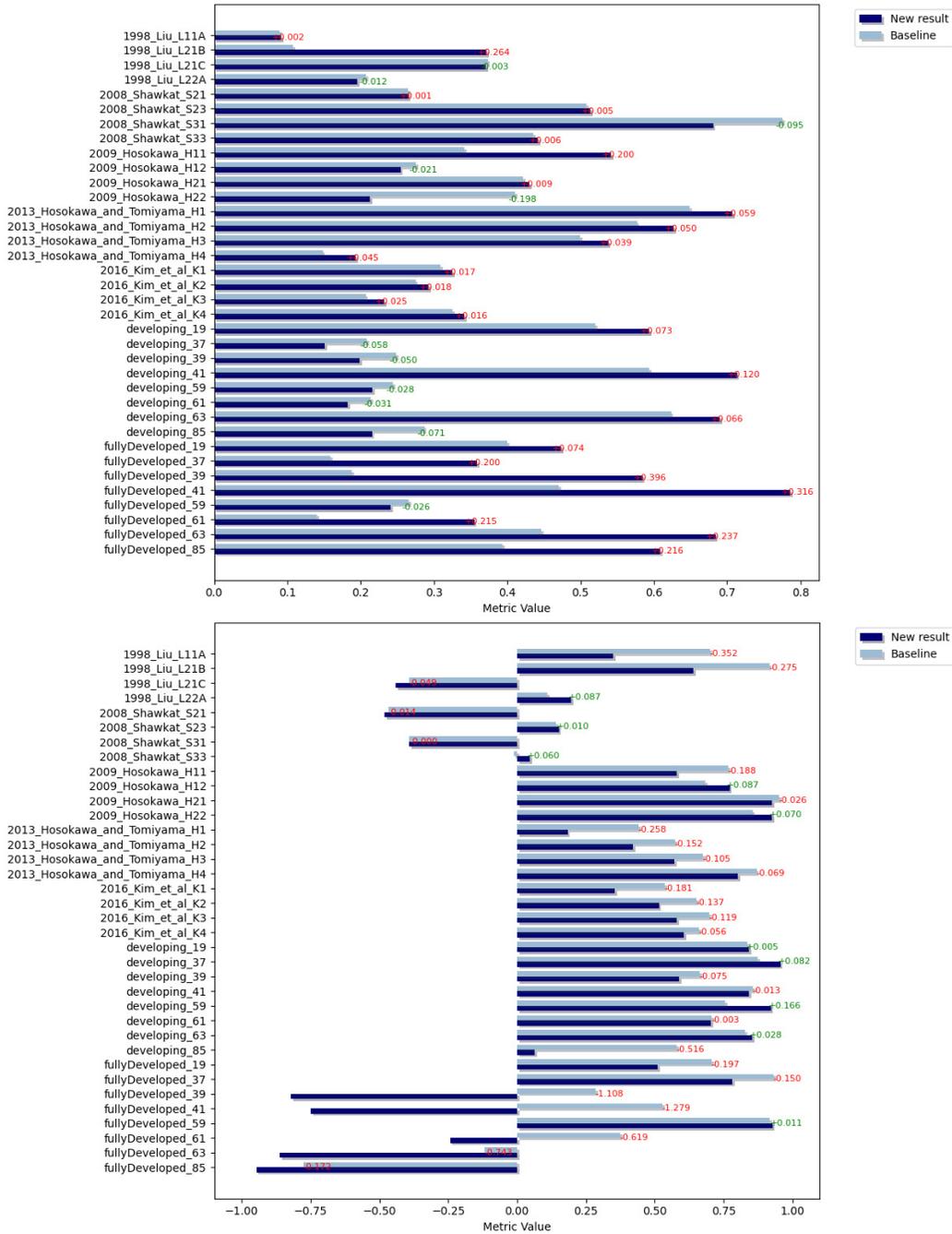
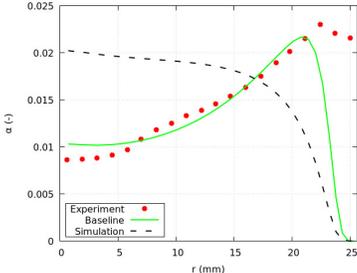
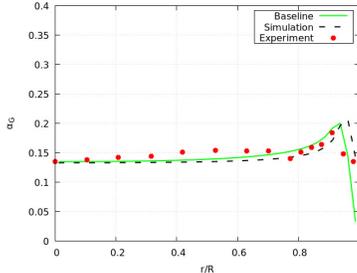


Figure 13: Computed error metrics for the gas void fraction prediction of all 36 demonstration cases. Top: Mean relative error. Bottom: Pearson correlation coefficient.

Table 5: Comparison of the computed fuzzy logic metrics for the gas void fraction profiles in Figure 2.

	<b>fd_39</b>	<b>L11A</b>
		
Baseline		
MRE	0.188	0.089
PCC	0.287	0.701
<b>G</b>	<b>0.546</b>	<b>0.713</b>
New model		
MRE	0.585	0.091
PCC	-0.822	0.349
<b>G</b>	<b>0.273</b>	<b>0.651</b>
$\Delta G$	<b>-0.273</b>	<b>-0.062</b>

508 The presented fuzzy logic plots above are a powerful tool assisting in  
509 the model evaluation and comparison for the baseline model development  
510 using a large number of cases. Note, however, that results will not cease  
511 to require the careful interpretation by the experienced researcher taking  
512 into account the underlying physical phenomena. An obvious limitation of  
513 the current validation strategy is that bubbly flow cases typically require a  
514 set of interdependent closures rather than isolated, individual models. This  
515 means the computed error metrics and corresponding goodness evaluation  
516 is only able to demonstrate the accuracy of all closure models combined.  
517 How well each individual closure model performs remains a task of careful  
518 interpretation by the investigator.

519 For a more detailed analysis of a new closure model the researcher can  
520 facilitate the keywords presented in Table 2, which are assigned to each sub-

521 case. Via keyword selection during the configuration step of the workflow  
522 the model analysis can be tailored to specific cases of interest with certain  
523 features.

524 The fuzzy logic analysis demonstrated here constitutes a first step towards  
525 a more automated model evaluation. In the future a more refined control of  
526 the results is possible by introducing alternative and/or additional metrics.  
527 Furthermore, core and near-wall regions could be analysed separately adopt-  
528 ing different error metrics and fuzzy sets for these individual flow regions  
529 if needed. The optimal use of fuzzy sets and metric combinations will be  
530 determined by the expert community and will evolve by gaining experience  
531 over time.

## 532 6. Conclusions

533 The development of closure models that can predict the behaviour of  
534 multiphase flows continues to be a challenge due to the increasing number of  
535 available model options, and their intricately-linked interactions. The base-  
536 line strategy offers a way towards a more sustainable model development. By  
537 setting the goal of establishing a fixed set of models and related model con-  
538 stants it is turning away from case-specific tuning, and instead gives priority  
539 to the generality and predictive capabilities of models. The practical imple-  
540 mentation of such a strategy involves the continuous validation of a large  
541 number of cases available in the literature, which continues to grow. Within  
542 the last few years of baseline research the HZDR team has accumulated a  
543 diverse set of such validation cases and built an extensive case collection.

544 We have presented here a Snakemake workflow that now allows the flexible  
545 and fully-automated pre-processing, simulation and post-processing of all  
546 these OpenFOAM cases at scale. The automated production of hundreds of  
547 validation plots and their representation in a single, easily accessible web-  
548 report allow an efficient analysis of our CFD results and their comparison to  
549 the experimental validation data.

550 Such advanced case management enables us to integrate artificial intelli-  
551 gence tools that assist in the analysis and evaluation of the extensive amount  
552 of results. The work presented here introduces a fuzzy logic controller quan-  
553 tifying the agreement of computed CFD results with the validation data.  
554 By comparing new results with a reference solution obtained with the base-  
555 line model set this fuzzy logic tool allows the systematic evaluation of new  
556 Euler-Euler models. The testing of a new lift force correlation demonstrates  
557 the potential of such an approach with comprehensive plots illustrating neat  
558 case-by-case comparisons of the accuracy of the CFD predictions.

559 The further development of the fuzzy logic evaluation and its extension by  
560 more input metrics is ongoing. A particular focus will be on metrics describ-  
561 ing the convergence behaviour and runtime performance of cases to obtain  
562 an alternative criteria by which to compare various Euler-Euler models. Fur-  
563 thermore, the uncertainty of the validation data should be included into the  
564 evaluation. The application of fuzzy logic is just one of many potential ways  
565 of including data science tools into the baseline model development. Another  
566 valuable area of further work is the integration of case-specific keywords into  
567 the model evaluation for further categorisation, which suits various machine  
568 learning algorithms.

## 569 Acknowledgements

570 It is a pleasure to acknowledge the helpful discussions with Dr. rer.  
571 nat. P. Steinbach from the Department of Computational Sciences at HZDR  
572 regarding the Snakemake library. This work was partly supported by the  
573 "Helmholtz European Partnering" program in the project "Crossing borders  
574 and scales - an interdisciplinary approach (CROSSING)".

## 575 References

- 576 [1] R. Sugrue, B. Magolan, N. Lubchenko, E. Baglietto, Assessment of a  
577 simplified set of momentum closure relations for low volume fraction  
578 regimes in STAR-CCM+ and OpenFOAM, *Annals of Nuclear Energy*  
579 110 (2017) 79 – 87.
- 580 [2] G. Besagni, F. Inzoli, T. Ziegenhein, Two-phase bubble columns: A  
581 comprehensive review, *ChemEngineering* 2 (2018).
- 582 [3] M. Colombo, R. Rzehak, M. Fairweather, Y. Liao, D. Lucas, Bench-  
583 marking of CFD modelling closures for two-phase turbulent bubbly  
584 flows, in: NURETH-18.
- 585 [4] J. Köster, S. Rahmann, Snakemake - a scalable bioinformatics workflow  
586 engine, *Bioinformatics* 28 (2012) 2520–2522.
- 587 [5] M. Cornwell, et al., VIPER: Visualization Pipeline for RNA-seq, a  
588 Snakemake workflow for efficient and complete RNA-seq analysis, *BMC*  
589 *Bioinformatics* 19 (2018).
- 590 [6] D. Lucas, R. Rzehak, E. Krepper, T. Ziegenhein, Y. Liao, S. Kriebitzsch,  
591 P. Apanasevich, A strategy for the qualification of multi-fluid approaches  
592 for nuclear reactor safety, *Nuclear Engineering and Design* 299 (2016)  
593 2–11. CFD4NRS-5.
- 594 [7] R. Rzehak, T. Ziegenhein, S. Kriebitzsch, E. Krepper, D. Lucas, Unified  
595 modeling of bubbly flows in pipes, bubble columns, and airlift columns,  
596 *Chemical Engineering Science* 157 (2017) 147–158.
- 597 [8] Y. Liao, T. Ma, E. Krepper, D. Lucas, J. Fröhlich, Application of a novel  
598 model for bubble-induced turbulence to bubbly flows in containers and  
599 vertical pipes, *Chemical Engineering Science* 202 (2019) 55–69.

- 600 [9] Y. Liao, R. Rzehak, D. Lucas, E. Krepper, Baseline closure model  
601 for dispersed bubbly flow: Bubble coalescence and breakup, *Chemical*  
602 *Engineering Science* 122 (2015) 336 – 349.
- 603 [10] Openfoam-dev Github, [https://github.com/OpenFOAM/](https://github.com/OpenFOAM/OpenFOAM-dev)  
604 [OpenFOAM-dev](https://github.com/OpenFOAM/OpenFOAM-dev), 2020.
- 605 [11] F. Schlegel, M. Draw, I. Evdokimov, S. Hänsch, H. Khan, R. Lehnigk,  
606 R. Meller, G. Petelin, M. Tekavčič, HZDR Multiphase Addon for Open-  
607 FOAM, <https://rodare.hzdr.de/record/768>, 2021.
- 608 [12] M. Ishii, N. Zuber, Drag coefficient and relative velocity in bubbly,  
609 droplet or particulate flows, *AIChE Journal* 25 (1979) 843–855.
- 610 [13] L. Schiller, Z. Naumann, A drag coefficient correlation, *Zeitschrift des*  
611 *Vereins Deutscher Ingenieure* 77 (1935) 318–320.
- 612 [14] A. Tomiyama, I. Kataoka, T. Sakaguchi, Drag coefficients of bubbles:  
613 1st Report, Drag coefficients of a single bubble in a stagnant liquid,  
614 *Transactions of the Japan Society of Mechanical Engineers Series B* 61  
615 (1995) 2357–2364.
- 616 [15] A. Tomiyama, G. Celata, S. Hosokawa, S. Yoshida, Terminal velocity of  
617 single bubbles in surface tension force dominant regime, *International*  
618 *Journal of Multiphase Flow* 28 (2002) 1497–1519.
- 619 [16] A. Tomiyama, I. Kataoka, I. Zun, T. Sakaguchi, Drag coefficients of  
620 single bubbles under normal and micro gravity conditions, *JSME In-*  
621 *ternational Journal Series B Fluids and Thermal Engineering* 41 (1998)  
622 472–479.
- 623 [17] A. Tomiyama, H. Tamai, I. Zun, S. Hosokawa, Transverse migration of  
624 single bubbles in simple shear flows, *Chemical Engineering Science* 57  
625 (2002) 1849 – 1858.
- 626 [18] D. Legendre, J. Magnaudet, The lift force on a spherical bubble in a  
627 viscous linear shear flow, *Journal of Fluid Mechanics* 368 (1998) 81–126.
- 628 [19] F. J. Moraga, F. J. Bonetto, R. Lahey, Lateral forces on spheres in  
629 turbulent uniform shear flow, *International Journal of Multiphase Flow*  
630 25 (1999) 1321–1372.

- 631 [20] A. Burns, T. Frank, I. Hamill, J.-M. Shi, The Favre averaged drag  
632 model for turbulent dispersion in Eulerian multi-phase flows, in: 5th  
633 International Conference on Multiphase Flow.
- 634 [21] A. D. Gosman, C. Lekakou, S. Politis, R. I. Issa, M. K. Looney, Mul-  
635 tidimensional modeling of turbulent two-phase flows in stirred vessels,  
636 AICHE Journal 38 (1992) 1946–1956.
- 637 [22] M. Lopez de Bertodano, Turbulent bubbly two-phase flow in a triangular  
638 duct, Ph.D. thesis, Rensselaer Polytechnic Institution, 1992.
- 639 [23] C. Crowe, J. Schwarzkopf, M. M. Sommerfeld, Y. Tsuji, Multiphase  
640 Flows with Droplets and Particles, 2011.
- 641 [24] H. Lamb, Hydrodynamics, 1993.
- 642 [25] S. Hosokawa, A. Tomiyama, S. Misaki, T. Hamada, Lateral migration  
643 of single bubbles due to the presence of wall., in: ASME 2002 Joint  
644 US-European Fluids Engineering Division Conference.
- 645 [26] S. P. Antal, R. Lahey Jr, J. Flaherty, Analysis of phase distribution in  
646 fully developed laminar bubbly two-phase flow, International Journal of  
647 Multiphase Flow 17 (1991) 635–652.
- 648 [27] T. Frank, Advances in computational fluid dynamics (CFD) of 3-  
649 dimensional gas-liquid multiphase flows, in: NAFEMS Seminar: Simu-  
650 lation of Complex Flows (CFD)-Applications and Trends, pp. 1–18.
- 651 [28] A. Tomiyama, Struggle with computational bubble dynamics, Multi-  
652 phase Science and Technology 10 (1998) 369–405.
- 653 [29] F. Menter, M. Kuntz, R. Langtry, Ten years of industrial experience  
654 with the SST turbulence model, Turbulence, Heat and Mass Transfer 4  
655 (2003) 625 – 632.
- 656 [30] B. Launder, D. Spalding, The numerical computation of turbulent flows,  
657 Computer methods in applied mechanics and engineering 3 (1974) 269–  
658 289.
- 659 [31] D. Wilcox, Turbulence modeling for CFD, volume 2, 1998.

- 660 [32] T. Ma, C. Santarelli, T. Ziegenhein, D. Lucas, J. Fröhlich, Direct nu-  
661 merical simulation-based Reynolds-averaged closure for bubble-induced  
662 turbulence, *Physical Review Fluids* 2 (2017).
- 663 [33] R. Rzehak, E. Krepper, CFD modeling of bubble-induced turbulence,  
664 *International Journal of Multiphase Flow* 55 (2013) 138–155.
- 665 [34] Y. Sato, M. Sadatomi, Momentum and heat transfer in two-phase bubble  
666 flow - I, Theory, *International Journal of Multiphase Flow* 7 (1981)  
667 167–177.
- 668 [35] Y. Liao, R. Oertel, S. Kriebitzsch, F. Schlegel, D. Lucas, A discrete  
669 population balance equation for binary breakage, *Int J Numer Meth*  
670 *Fluids* (2018) 1–14.
- 671 [36] M. Ishii, S. Kim, J. Kelly, Development of interfacial area transport  
672 equation., *Nuclear Engineering and Technology* 37 (2005) 525–536.
- 673 [37] C. Coualaloglou, L. Tavlarides, Description of interaction processes in ag-  
674 itated liquid-liquid dispersions, *Chemical Engineering Science* 32 (1977)  
675 1289–1297.
- 676 [38] F. Lehr, M. Millies, D. Mewes, Bubble-size distributions and flow fields  
677 in bubble columns, *AIChE Journal* 48 (2002) 2426–2443.
- 678 [39] H. Luo, Coalescence, breakup and liquid circulation in bubble column  
679 reactors, Ph.D. thesis, Department of Chemical Engineering, The Nor-  
680 wegian Institute of Technology, Trondheim, Norway, 1993.
- 681 [40] M. Prince, H. Blanch, Bubble coalescence and breakup in air-sparged  
682 bubble columns, *AIChE journal* 36 (1990) 1485–1499.
- 683 [41] M. Laakkonen, V. Alopaeus, J. Aittamaa, Validation of bubble break-  
684 age, coalescence and mass transfer models for gas-liquid dispersion in  
685 agitated vessel, *Chemical Engineering Science* 61 (2006) 218–228.
- 686 [42] H. Luo, H. Svendsen, Theoretical model for drop and bubble breakup  
687 in turbulent dispersions, *AIChE Journal* 42 (1996) 1225–1233.
- 688 [43] M. Colombo, M. Fairweather, RANS simulation of bubble coalescence  
689 and break-up in bubbly two-phase flows, *Chemical Engineering Science*  
690 146 (2016) 207 – 225.

- 691 [44] M. Colombo, R. Rzehak, M. Fairweather, Y. Liao, D. Lucas, Bench-  
692 marking of computational fluid dynamic models for bubbly flows, Nu-  
693 clear Engineering and Design 375 (2021) 111075.
- 694 [45] D. Lucas, Y. Liao, R. Rzehak, E. Krepper, Towards best practice  
695 guidelines for Euler-Euler simulations of poly-disperse bubbly flows, in:  
696 NURETH-18.
- 697 [46] Agile delivery with GitLab, [https://about.gitlab.com/solutions/  
698 agile-delivery/](https://about.gitlab.com/solutions/agile-delivery/), 2020.
- 699 [47] M. D. Wilkinson, et al., The FAIR guiding principles for scientific data  
700 management and stewardship, Scientific data 3 (2016).
- 701 [48] Python workflow automation tool for time-series, [https://github.  
702 com/KIT-IAI/pyWATTS](https://github.com/KIT-IAI/pyWATTS), 2020.
- 703 [49] I. Evdokimov, S. Hänsch, F. Schlegel, Scalable Workflows for Open-  
704 FOAM Evaluation, in: IPS RAS 2020.
- 705 [50] D. Lucas, E. Krepper, H.-M. Prasser, Development of co-current airwa-  
706 ter flow in a vertical pipe, International Journal of Multiphase Flow 31  
707 (2005) 1304 – 1328.
- 708 [51] J. Köster, Snakemake report, [https://koesterlab.github.io/  
709 resources/report.html](https://koesterlab.github.io/resources/report.html), 2020.
- 710 [52] Github repository scikit-fuzzy, [https://github.com/scikit-fuzzy/  
711 scikit-fuzzy](https://github.com/scikit-fuzzy/scikit-fuzzy), 2020.
- 712 [53] J. Warner, et al., scikit-fuzzy/scikit-fuzzy: Scikit-fuzzy 0.4.2, [https:  
713 //zenodo.org/record/3541384](https://zenodo.org/record/3541384), 2019.
- 714 [54] S. S. Izquierdo, L. R. Izquierdo, Mamdani fuzzy systems for modelling  
715 and simulation: A critical assessment, Journal of Artificial Societies and  
716 Social Simulation 21 (2018) 2.
- 717 [55] H. Hessenkemper, T. Ziegenhein, R. Rzehak, D. Lucas, A. Tomiyama,  
718 Lift force coefficient of ellipsoidal single bubbles in water, International  
719 Journal of Multiphase Flow 138 (2021) 103587.

- 720 [56] S. Hänsch, M. Draw, I. Evdokimov, H. Khan, B. Krull, R. Lehnigk,  
721 Y. Liao, H. Lyu, R. Meller, F. Schlegel, M. Tekavčič, HZDR Multiphase  
722 Case Collection for OpenFOAM, [https://doi.org/10.14278/rodare.](https://doi.org/10.14278/rodare.812)  
723 812, 2021.
- 724 [57] S. Hosokawa, A. Tomiyama, Multi-fluid simulation of turbulent bubbly  
725 pipe flows, *Chemical Engineering Science* 64 (2009) 5308 – 5318.
- 726 [58] R. Rzehak, Y. Liao, R. Meller, F. Schlegel, R. Lehnigk, D. Lucas, Radial  
727 pressure forces in Euler-Euler simulations of turbulent bubbly pipe flows,  
728 *Nuclear Engineering and Design* 374 (2021) 111079.
- 729 [59] T. Liu, The role of bubble size on liquid phase turbulent structure in  
730 two-phase bubbly flow, in: 3rd International Conference on Multiphase  
731 Flow (ICMF98).
- 732 [60] S. Kriebitzsch, R. Rzehak, Baseline model for bubbly flows: Simulation  
733 of monodisperse flow in pipes of different diameters, *Fluids* 1 (2016) 29.
- 734 [61] M. Shawkat, C. Ching, M. Shoukri, Bubble and liquid turbulence char-  
735 acteristics of bubbly flow in a large diameter vertical pipe, *International*  
736 *Journal of Multiphase Flow* 34 (2008) 767 – 785.
- 737 [62] S. Hosokawa, A. Tomiyama, Bubble-induced pseudo turbulence in lam-  
738 inar pipe flows, *International Journal of Heat and Fluid Flow* 40 (2013)  
739 97 – 105.
- 740 [63] Y. Liao, K. Upadhyay, F. Schlegel, Eulerian-Eulerian two-fluid model for  
741 laminar bubbly pipe flows: Validation of the baseline model, *Computers*  
742 *& Fluids* 202 (2020) 104496.
- 743 [64] M. Kim, J. Lee, H. Park, Study of bubble-induced turbulence in upward  
744 laminar bubbly pipe flows measured with a two-phase particle image  
745 velocimetry, *Exp Fluids* 57 (2016) 1432–1114.
- 746 [65] R. Lehnigk, W. Bainbridge, Y. Liao, D. Lucas, T. Niemi, J. Peltola,  
747 F. Schlegel, An open-source population balance modeling framework for  
748 the simulation of polydisperse multiphase flows based on the method of  
749 classes, *AIChE Journal* (submitted).