Progress in Scale Modeling, an International Journal

Volume 2 | Issue 1 Article 1

2021

Data-driven tools guided by first-principles for scale modeling

Sadegh Poozesh
Tuskegee university, sadegh.poozesh@gmail.com

Recommended Citation

Poozesh, Sadegh (2021) "Data-driven tools guided by first-principles for scale modeling," *Progress in Scale Modeling, an International Journal*: Vol. 2 : Iss. 1 , Article 1.
DOI: https://doi.org/10.13023/psmij.2021.02-01-01
Available at: https://uknowledge.uky.edu/psmij/vol2/iss1/1

Abstract

For decades, traditional scale-modeling techniques have been relying on first-principles models (FPMs). FPMs have been used to find non-dimensional numbers (PIs) and identify normalized underlying forces and energies behind the phenomenon in focus. The two main challenges with FPM-based PIs extraction are finding the relevant PIs and proper correlations between PIs. The emergence and surge of data-driven modeling (DDM) provide a new opportunity to leverage experimental data in model development across scales/plants. In this paper, first, the two mentioned issues in PIs development will be elaborated to reveal the gap, and second, a new insight into scale modeling and similarity concepts will be presented. Then, to showcase the presented framework for a two-fluid spray nozzle case study, DDM techniques will be synergized with FPMs to obtain a robust relationship between relevant properties of the model spray and atomization parameters.

Keywords

Data-driven model; scale modeling; PI numbers; system commercialization; spray technology; two-fluid nozzle

Cover Page Footnote

I would like to acknowledge insightful comments and suggestions given by Dr. Kozo Saito.

Progress in Scale Modeling, an International Journal, Vol. 2, Issue 1 (2021) Article 02-01-01, pp. 1–6
https://doi.org/10.13023/psmij.2021.02-01-01

Data-driven tools guided by first-principles for scale modeling

Sadegh Poozesh

Mechanical Engineering Department, Tuskegee University, 516 University Avenue, Tuskegee, AL 36088, USA

E-mail: sadegh.poozesh@gmail.com

Received October 18, 2020, Accepted December 16, 2020

Abstract

For decades, traditional scale-modeling techniques have been relying on first-principles models (FPMs). FPMs have been used to find non-dimensional numbers (PIs) and identify normalized underlying forces and energies behind the phenomenon in focus. The two main challenges with FPM-based PIs extraction are finding the relevant PIs and proper correlations between PIs. The emergence and surge of data-driven modeling (DDM) provide a new opportunity to leverage experimental data in model development across scales/plants. In this paper, first, the two mentioned issues in PIs development will be elaborated to reveal the gap, and second, a new insight into scale modeling and similarity concepts will be presented. Then, to showcase the presented framework for a two-fluid spray nozzle case study, DDM techniques will be synergized with FPMs to obtain a robust relationship between relevant properties of the model spray and atomization parameters.

Keywords: Data-driven model; Scale modeling; PI numbers; System commercialization; Spray technology; Two-fluid nozzle

Introduction

Similarity exists when characteristics of one system can be associated with the corresponding characteristics of another system by a simple conversion factor, called the scale factor, which is a non-dimensional parameter [1]. Good scale modeling which is interpreted as finding the right scale factors is integral to process, unit scale-up, and cross-scale analysis. The scale factors are often being the core around which design space is determined at higher scales [2]. Chronologically, scale modeling has been guided by empiricism, theory, and most recent computational simulation techniques [1]. According to Saito [3], science and engineering have not explored a viable assumption making method which can be universally applied to practical problems. That may be because the assumption making process requires an inductive process that is rather subjective in nature and therefore does not fit well to natural science and engineering, which exclude human factors and human thinking process [4]. This equation-based approach has limitation. That limitation becomes more evident in scale modeling, since scale modeling requires both science and art [4]. To further emphasize this point, it may be interesting to see Einstein's quote on the role of human intuition on the development of physics: "Thus the supreme task of the physicist is the discovery of the most general elementary laws from which the world-picture can be deduced logically. But there is no logical way to the discovery of these elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind the appearance, and this Einfühlung is developed by experience" [5]. Einstein's claim can be applied to scale modeling, e.g., the role of human factors (intuition, feeling and experience) on a sound assumption making process." according to Saito's preface in [4], which focuses on the role of kufu.

Data-driven modeling (DDM) offers intuition-based analysis just the way an expert in a particular field, after exposure to a giant pool of trial and errors and input and output data, can intuitively guess the most important driving forces underlying a phenomenon. The new emerging tool that has been enabled by sensor technology and robust computational tools can bring new opportunities in scale modeling by relaxing some of the main current restrictions in similarity concepts.

One traditional approach in scale modeling is using the non-dimensional parameters (PIs) that are determined based on the physics behind the understudied

Progress in Scale Modeling
AN INTERNATIONAL JOURNAL
https://uknowledge.uky.edu/psmij/

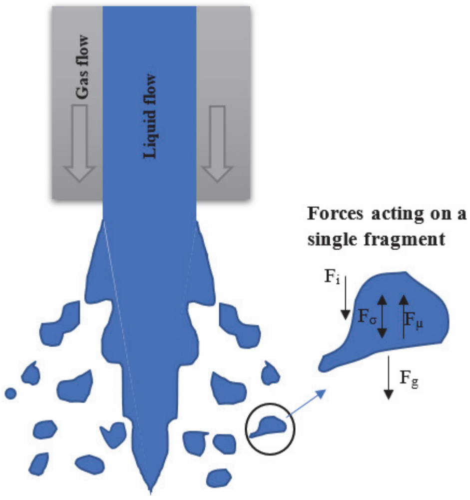

Fig. 1. A schematic of atomization in a two-fluid atomizer and a zoom-in view of a liquid fragment with acting forces on it. *F* stands for force, while *i*, σ, μ, and *g* represent inertia, surface tension, viscosity, and gravity, respectively.

problem, and/or interacting forces and energies driving occurring phenomena. For highly coupled problems with many parameters (and often non-linearly interwoven), it is typically the discretion of the researcher to determine the most appropriate PIs. Scale factors or PIs can also be related to the system outcome, or variable(s) embodying system performance. In this area of research, finding firm mathematical correlations between PIs and system outputs has been tedious [1]. The form of correlations (the type of equations), and the specific applicable PIs are still the main bottleneck. The former is normally found by physical analysis combined with heuristic assessments of the data.

With progress made in predictive capabilities in modeling of complex, non-linear, multivariate systems, DDMs (which serve as an umbrella covering machine learning) can be used to address the observed challenges. Yet, caution must be practiced in merely using DDMs for system identification, especially for scaling. Such models are uninterpretable and usually are not reproducible from one scale to the other [6]. Therefore, it is imperative to integrate physics-based models with DDMs to lift their limitations. What follows is the application of this hybrid method to explore atomization for a two-fluid nozzle, in particular to relate atomization conditions and formulation to the ensued droplet size distribution. Starting with finding PIs, and after extracting relevant PIs, they will be fed to a Support Vector Machines (SVM) platform to construct a mathematical model.

Physics-based modeling

In a two-fluid type gun, the air at high pressure is ballasted on the liquid film (filament) and after several breakup stages, droplets at a wide distribution are generated. The atomization process is dictated by the balance between several forces. Other factors, such as the mass flow rates of the liquid and the gas (alternatively, often captured in terms of the relative velocity between the gas and the liquid) promote the atomization process through enhancing shear forces against viscous forces and acting as a resistance to the atomization process. Surface tension forces, depending on the breakup stage (the instabilities involved and the extent of each) can either promote or resist disintegration and stripping of the droplets. When the Rayleigh-Plateau instabilities become important, capillary-induced pinch-off promotes atomization [7]. The main atomization is driving by the Kelvin–Helmholtz instability due to the high shear forces in two-fluid atomization.

Ideally, the sensitivity of the droplet size on process conditions can be deduced from the relative magnitude of the involved factors/forces. For example, increasing the difference between gas and liquid velocity at the nozzle outlet, either by lowering the liquid flow rate or raising the air pressure, results in smaller droplets.

There have been various approaches to describe drop size distribution based on atomization conditions including instability analysis, maximizing entropy generation, and other empirical studies [8–11]. One approach to this end is to develop correlations between these forces and the associated measured droplet size. This may be accomplished using non-dimensional analysis capturing both the physical properties of the feed solution and operation conditions to droplet size distribution at an appropriate distance downstream of the nozzle tip. This approach captures the critical operational conditions and liquid feed physicochemical properties into dimensional parameters that describe the underlying physics governing the breakup phenomena [12]. This approach is meant to compare the relative magnitude of the significant forces. These forces are shown in Fig. 1. Note that since gravity for a micron size droplet is minor, therefore gravity force and associated *Fr* number is ignored in this study [13, 10]. The Weber number (*We*) compares the dynamic gas pressure acting on the liquid sheet to the liquid capillary pressure of the liquid sheet [14] as follows.

$$We = \frac{v_g^2 \rho_g d_L}{\sigma} \tag{1}$$

Here, v_g, ρ_g, d_L and σ are respectively gas velocity, gas density, liquid orifice diameter, and liquid surface tension. We chose gas velocity in the numerator, as the gas velocity magnitude is much higher than the liquid velocity for a two-fluid nozzle (in the order of 10-100-fold). When *We* is increased, the atomization mechanism becomes more intense, and the resulting droplets are smaller [15]. When a gas stream is directed onto a liquid surface, produced are oscillation and waves on the surface of the liquid to promote the fragmentation process of bulk liquid. This phenomenon is captured in the numerator of Eq. (1). The resisting force to these oscillations is liquid capillary force—captured in the denominator of Eq. (1).

The other dimensionless number that relates gas inertia to the liquid viscous force is the Reynolds number (*Re*) [16, 17]. While working with both *We* and *Re* numbers, to isolate inertia effects, another PI, the Ohnesorge number (*Oh*), can be derived via combining *We* and *Re*, $Oh = We^{0.5}/Re$. This dimensionless number is purely dependent on feed solution properties and geometry of liquid orifice. The *Oh* number is defined as follows.

$$Oh = \frac{\mu}{\sqrt{\rho_l \sigma d_L}} \tag{2}$$

where, the liquid density is represented by the variable ρ_l. In addition to the involved forces, the information on the mass flow rate of each phase is needed to analyze the dynamics of the atomization system and assess the magnitude of shear forces. The gas-to-liquid ratio, $GLR=\dot{m}_g/\dot{m}_l$, is defined as the ratio between air and liquid mass flow rates and has often been cited as a dimensionless quantity that captures the atomization operating conditions [18]. The next step uses this pool of resulting PIs (predictors), regardless of their importance in the final relationship, to arrive at a generalized physically based data driven model for describing droplet size. Traditional physics-based scale modeling requires sound assumptions based on a careful interpretation and observation of the phenomena. Furthermore, several criteria have to be considered to avoid conflicting scaling law predictions [3, 4]. Yet, via the new hybrid method, these assumptions can be relaxed and the DDM automatically decides which PIs are most important for a specific problem.

The Sauter mean diameter (*SMD* or *D* [3, 2]) that is the diameter of a sphere with the same volume to surface area ratio of the entire spray is used to represent droplet size for the generated spray [19]. Multiple physics-based empirically driven mathematical correlations have been proposed to link droplet size and operation conditions. One of the most notable one that predicts *SMD* as a function of PIs is offered by Groom et al. [20] (*Cs* are constant):

$$SMD = d_L C_1 . \left[\frac{We}{(1+GLR)^2}\right]^{C_2} (1 + C_3 . Oh) \tag{3}$$

In the Groom study, a commercial two-fluid nozzle consisting of an external liquid capillary outlet with a co-flow gas stream at comparatively high relative velocity was used. The experiments were performed with water and aqueous glycerol solutions with a viscosity ranging from 1-100 mPa.s and an atomizing air pressure ranging from 50–300 kPa [20], the conditions similar to the other operational conditions. The explanation of *SMD* behavior in relation with PIs in Eq. (3) was provided elsewhere [20, 10], and not repeated here.

Data-driven modeling

The experimental setup consists of two main components: a RTS 5114 Malvern Spraytec (Malvern Instruments, Worcestershire, UK), and a two-fluid atomizing nozzle. The complete setup and further details can be found in the previous studies [11, 10]. Full cone sprays were produced with an external mixing nozzle with liquid orifice diameters, d_L, of 0.7 and 0.5 mm, while having a cap diameter of 1.5 mm. Other operation conditions are liquid flowrate (20–80cc/min), and gas pressure (50–250kPa), which was then converted into gas flowrate. These conditions rendered a *SMD* range of 2–100 µm. After collecting data for multiple feed solutions with a wide physicochemical properties and atomization conditions, data were converted to PIs for further analysis. The PI data from 440 experimental trials are gathered

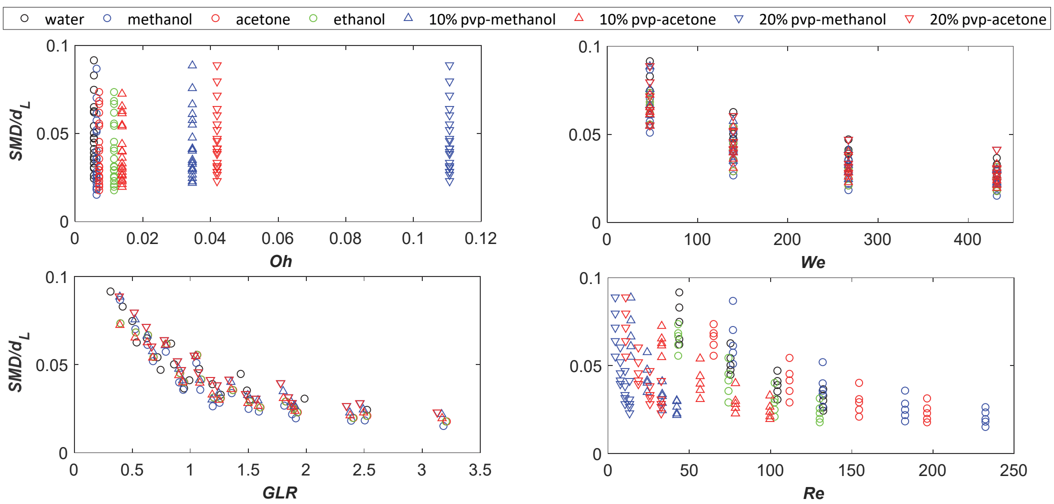

Fig. 2. Normalized droplet size (SMD/d_L) as a function of non-dimensional parameters (PIs).

Table 1. Coefficients in Eq. (5).

b_1	−0.12544	b_7	2.5731
b_2	0.1798	b_8	0.29245
b_3	−0.16362	b_9	1.0319
b_4	−36.041	b_{10}	0.053379
b_5	36.951	b_{11}	0.43261
b_6	0.0041449	b_{12}	2.8161

(can be requested by emailing the author). These data are administered to extract a predictor of the type $SMD/d_L = f(We, Oh, Re, GLR)$ which would contain the most important/relevant PIs.

Fig. 2 provides the scattered distribution of the PIs and the relationships between them and the normalized *SMD*, for a portion of the data pool. Based on this plot, *GLR* and *We* show clear reversal relationships with normalized *SMD*. Additionally, the *Oh* can be used as a firm classifier for embodying feed selection impacts. The data for *Re* show no clear trend.

The base generalized function considered for optimization that encompasses all possible terms formed by combining PIs is as follows,

$$\begin{aligned}\frac{SMD}{d_L} = \sum b_i . We^{b_j} . Re^{b_k} . Oh^{b_m} . GLR^{b_l} \\ = b_1 + b_2 We^{b_3} + b_4 Re^{b_5} \\ + \cdots + b_{10} We^{b_{11}} . Re^{b_{12}} \\ + \cdots \\ + b_{28} We^{b_{29}} . Re^{b_{30}} . Oh^{b_{31}}\end{aligned} \quad (4)$$

Eq. (4) was used as a base function for the latter optimization and fitting problem. In the next step of the study, we fed the data to a data-driven modeling tool, and then trained the model to minimize the error between actual data and predicted ones. The machine learning tools offered in MATLAB by fitrsvm function as an optimizer are applied here together with a least-squares support vector machine (LS-SVM) optimization model used for objective-function minimization via quadratic programming [21, 22]. SVM-based techniques are simple in their theoretical background and very powerful in finding predictors for multivariable relationships [23]. Especially when seeking for a generalized correlation, compared to neural networks, SVM has the advantage of yielding a global model that is capable of efficiently dealing with high dimensional input vectors [23]. Based on the data given to LS-SVM and optimization of the problem, which is interpreted as minimizing error while avoiding insignificant terms in the generalized Eq. (4), the following simplified equation can be derived after further mathematical operations,

$$\frac{SMD}{d_L} = b_1 + b_2 \left(\frac{We^{b_3}}{(b_4 + b_5 GLR^{b_6})^{b_7}} \right)^{b_8} (b_9 + b_{10} Oh^{b11})^{b_{12}} \quad (5)$$

with coefficients given in Table 1.

The experimental data, the theoretically predicted data, and the difference between them are shown in Fig. 3. Before administering the experimental data were divided into training and validation datasets. About 20% of the data were used for validation. The predictor was assessed according to diverse error metrics, also referred to as fitness functions (e.g., mean absolute error, mean square error, or R-squared). Mean absolute error for the final fit presented in Eq. (5) is 0.001456 and R-squared is 0.98.

Comparing the obtained equation via LS-SVM, Eq. (5) and the offered equation by Groom et al.[20] and Poozesh, et al. [8] for a similar two-fluid gun, it is clear that the physics-based features in Eq. (5) are preserved

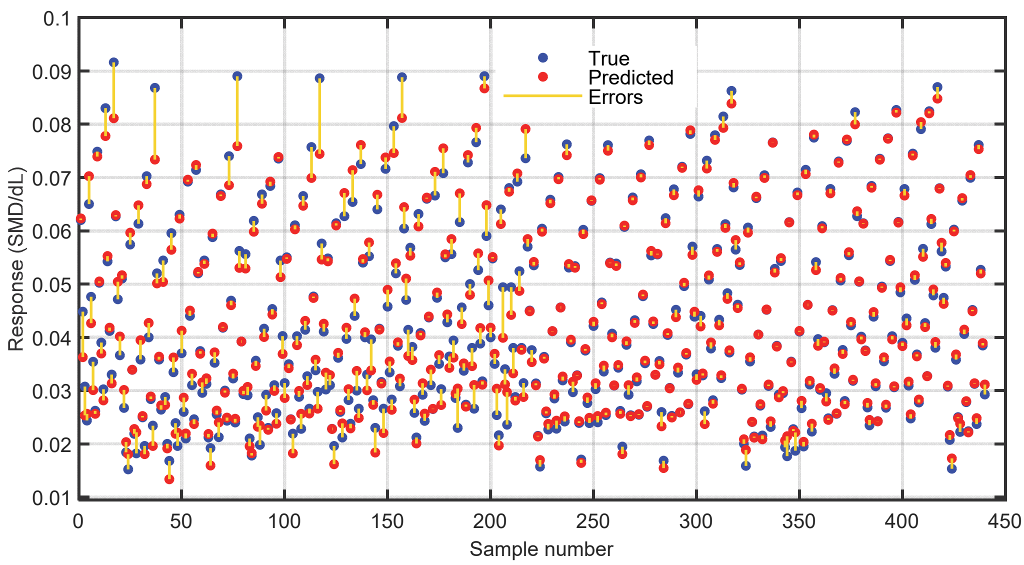

Fig. 3. Response plot of normalized *SMD* for predicted data via SVM, the actual administered data, and the error between the two.

and complemented in data-driven Eq. (5). Even though, not all coefficients across the two equations are same/close, having $b_9 \sim 1.0$ and $|b_4| \sim b_5$ makes Eq. (5) very similar to Eq. (3). Furthermore, general arrangements of PIs in the equation and their reversal or direct correlations to normalized SMD, shows that Eq. (5) not only retains the physics but presents a more generalized version of the correlations presented in the relevant literature [10, 20, 24]. Lastly, focusing on Eq. (4) that includes *Re* number and its absence in Eq. (5) shows that this hybrid technique is able to remove minor and unnecessary PIs/terms that would have been compensated by other PIs such as *Oh* and *We* and their associated terms.

As a result, this example clearly demonstrates that the solution provided in this study can address the earlier mentioned current issues with traditional scale-up methods. All PIs, that may or may not play a role in the phenomena under consideration, may unintentionally be integrated without regards to their importance. If this happens, the simplification process which is the heart of scale modeling [3, 4, 17] may be compromised. The presented hybrid framework will decide the faith of each PI, whether or not each can be a part of the final expression. In the illustrated case study, despite incorporation of *Re*, the final expression leaves *We* and *Oh* to account for dependency of normalized *SMD* on *Re*. Besides, the new framework can also address the correlation/formula type challenge, as it demands a generalized function and eventually finds the best formula based on the data behavior—discovering physical relationships within administered data.

Conclusion

With the surge in data-driven modeling (DDM) in almost all aspects of engineering, the possibility of coupling these tools with traditional first-principle models (FPM) for highly demanding scale modeling deserves a closer look. Therefore, this paper is the first attempt to address the current challenges with FPM-based scale modeling, and then present a hybrid framework to leverage DDM to remediate the observed issues. To corroborate the idea with a practical and widely used technology, the framework was examined with a two-fluid spray nozzle case study, where DDM techniques were synergized with FPMs to obtain a robust predictor model. After finding non-dimensional parameters (PIs), they were administered to a Support Vector Machine data driven technique to map the data pool on a base generalized model. This machine learning optimization tool could produce a robust physical model containing only the most significant terms, relating normalized droplet size with atomization PIs. Finally, the extracted predictor model showed similarities with the proposed physics-based models in the literature. As accurate data-driven modeling techniques demand a large dataset, future improvement may be needed to address how to effectively extract data on two-fluid guns in relevant open literature and present a more inclusive predictor.

References

[1] Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E., Van Emmerik, T., Uijlenhoet, R., Achieng, K., Franz, T. E., Woods, R., "Scaling, similarity, and the fourth paradigm for hydrology," Hydrology and Earth System Sciences 21: 3701–3713, 2017.

[2] Poozesh, S., Bilgili, E., "Scale-up of pharmaceutical spray drying using scale-up rules: A review," International Journal of Pharmaceutics 562: 271–292, 2019.

[3] Saito, K., Progress in Scale Modeling: Summary of the First International Symposium on Scale Modeling (ISSM I in 1988) and Selected Papers from Subsequent Symposia (ISSM II in 1997 through ISSM V in 2006), Springer Science & Business Media, 2008.

[4] Emori, R., Saito, K., Sekimoto, K., Scale Models in Engineering (Mokei Jikken no Riron to Ohyou), Gihodo Publishing Co., 2000.

[5] Planck, M., Where is Science Going? Pickle Partners Publishing, 2017.

[6] Zhang, D., Del Rio-Chanona, E. A., Petsagkourakis, P., Wagner, J., "Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization," Biotechnology and Bioengineering 116: 2919–2930, 2019.

[7] Joseph, D. D., Renardy, Y. Y., Fundamentals of Two-Fluid Dynamics: Part I: Mathematical Theory and Applications, Vol. 3, Springer Science & Business Media, 2013.

[8] Li, X., Li, M., Fu, H., "Modeling the initial droplet size distribution in sprays based on the maximization of entropy generation," Atomization and Sprays 15: 295–322, 2005.

[9] Teng, H., Kinishita, C., Masutani, S., "Prediction of droplet size from the breakup of cylindrical liquid jets," International Journal of Multiphase Flow 21: 129–136, 1995.

[10] Poozesh, S., Setiawan, N., Akafuah, N. K., Saito, K., Marsac, P. J., "Assessment of predictive models for characterizing the atomization process in a spray dryer's bi-fluid nozzle," Chemical Engineering Science 180: 42–51, 2018.

[11] Poozesh, S., Grib, S. W., Renfro, M. W., Marsac, P. J., "Near-field dynamics of high-speed spray dryer coannular two fluid nozzle: effects of operational conditions and formulations," Powder Technology 333: 439–448, 2018.

[12] Hassan, R., Loubiere, K., Legrand, J., Delaplace, G., "A consistent dimensional analysis of gas-liquid mass transfer in an aerated stirred tank containing purely viscous fluids with shear-thinning

properties," Chemical Engineering Journal 184: 42–56, 2012.

[13] Poozesh, S., Saito, K., Akafuah, N. K., Graña-Otero, J., "Comprehensive examination of a new mechanism to produce small droplets in drop-on-demand inkjet technology," Applied Physics A 122: 110, 2016.

[14] Hede, P. D., Bach, P., Jensen, A. D., "Two-fluid spray atomisation and pneumatic nozzles for fluid bed coating/agglomeration purposes: A review," Chemical Engineering Science 63: 3821–3842, 2008.

[15] Chigier, N., Farago, Z., "Morphological classification of disintegration of round liquid jets in a coaxial air stream," Atomization and Sprays 2: 137–153, 1992.

[16] Poozesh, S., Akafuah, N., Saito, K., "New criteria for filament breakup in droplet-on-demand inkjet printing using volume of fluid (VOF) method," Korean Journal of Chemical Engineering 33: 775–781, 2016.

[17] Saito, K., Williams, F. A., "Scale modeling in the age of high-speed computation," in: Progress in Scale Modeling, Volume II, Springer, 2015, pp 1-18.

[18] Lefebvre, A. H., "Airblast atomization," Progress in Energy and Combustion Science 6: 233–261, 1980.

[19] Filkova, I., Mujumdar, A. S., "Industrial spray drying systems," Handbook of Industrial Drying 1: 263–308, 1995.

[20] Groom, S., Schaldach, G., Ulmer, M., Walzel, P., Berndt, H., "Adaptation of a new pneumatic nebulizer for sample introduction in ICP spectrometry," Journal of Analytical Atomic Spectrometry 20: 169–175, 2005.

[21] Gunn, S. R., "Support vector machines for classification and regression," ISIS Technical Report 14: 5–16, 1998.

[22] Pelckmans, K., Suykens, J. A., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B., Vandewalle, J., "LS-SVMlab: A Matlab/C Toolbox for Least Squares Support Vector Machines," Tutorial KULeuven-ESAT Leuven 142, 2002.

[23] Balabin, R. M., Lomakina, E. I., "Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," Analyst 136: 1703–1712, 2011.

[24] Poozesh, S., Akafuah, N. K., Campbell, H. R., Bashiri, F., Saito, K., "Experimental and mathematical tools to predict droplet size and velocity distribution for a two-fluid nozzle," Fluids 5: 231, 2020.