International Journal of Multiphase Flow 35 (2009) 650–660

Contents lists available at ScienceDirect

International Journal of Multiphase Flow

journal homepage: www.elsevier.com/locate/ijmulflow

Probabilistic mapping of adiabatic horizontal two-phase flow by capacitance signal feature clustering

H. Canière [*], B. Bauwens, C. T'Joen, M. De Paepe

Department of Flow, Heat and Combustion Mechanics, Ghent University-UGent, St.-Pietersnieuwstraat 41, 9000, Gent, Belgium

ARTICLE INFO

Article history:
Received 22 April 2008
Received in revised form 2 March 2009
Accepted 11 March 2009
Available online 20 March 2009

Keywords:
Flow regime classification
Clustering algorithm
Regression

ABSTRACT

In order to better quantify two-phase flow regime transitions, a sensor is developed which measures the electric capacitance of two-phase flows. A large number of experiments are done with air–water flow in a 9 mm ID horizontal tube. Based on a multivariate analysis, the most suitable sensor signal parameters are selected for building a flow regime classifier. This classifier is based on a fuzzy c-means clustering algorithm together with a regression technique. The output of the algorithm is used to create a probabilistic flow regime map. A comparison between a visual classification based on high speed camera images and the outcome of the flow regime classifier shows a remarkable agreement. The flow regime transitions are further quantified and discussed based on the probabilistic information and the sensor signal characterization. Probabilistic mapping makes it possible to combine flow regime dependent correlations in the two-phase flow models for heat transfer and pressure drop with smooth and appropriately quantified transitions from one flow regime to another.

1. Introduction

When designing in-tube evaporators used in refrigeration and air-conditioning, one has to deal with the complex phenomena of two-phase flow during the phase change of the refrigerant from liquid to vapour. To accurately predict the heat transfer and pressure drop, the flow phenomena should be incorporated in the design models (Thome, 2004). Traditionally, this is achieved by classifying two-phase flows into flow regimes: stratified flow, annular flow, etc. These classifications are mainly based on visualizations (with or without use of high speed cameras). But visual observations are inherently subjective and do not provide quantitative information such as typical frequencies or local vapour concentration.

Many types of flow pattern maps indicate the occurrence of typical flow regimes in a specific coordinate system (superficial velocities or mass velocity G and vapour quality x, etc.). Attempts were made to create transition boundaries between these regimes based on theoretical assumptions (Taitel and Dukler, 1976; Barnea, 1987). But due to wide transition areas, these boundaries differ strongly between maps. More recently, this problem was approached in a probabilistic way. Rather than purely classifying a flow in a specific regime, the flow can be conceived as a combination of different flow regimes. The importance of each of the acting forces can be better described, especially in the transition areas and the chaotic flow types such as intermittent flows. Nino et al. (2003) introduced the probabilistic approach in multiport microchannels. Jassim and Newell (2006) applied probabilistic flow regime mapping to predict pressure drop and void fraction in microchannels. van Rooyen et al. (2007) used the same approach for intermittent flows during condensation in macroscale tubes. Recently, Jassim et al. (2007) obtained probabilistic two-phase flow data of R134a and R410A in single horizontal smooth, adiabatic tubes by using an automated image recognition technique. Several tubes were used with diameters ranging from 1.74 mm to 8 mm ID. Jassim (2006) developed generalized probabilistic two-phase flow regime maps from this time fraction data, which were used by Jassim et al. (2008a) for void fraction modeling and by Jassim et al. (2008b) for heat transfer modeling during condensation.

To gather quantitative and more objective information about the nature of the two-phase flows, a lot of measurement techniques were developed and successfully applied. Among them, pressure transducers as well as optical or impedance probes are the most common. Signal analysis provides valuable information on the subtle differences in flow phenomena and makes it possible to build quantifying classifiers with probabilistic data.

For this study, a capacitance sensor was developed and tested with air–water flow (Canière et al., 2007). The sensor signals were analyzed and a multivariate analysis of signal features was performed. Three signal features were chosen to build a horizontal two-phase flow classifier without any subjective visual decisions. Probabilistic data was taken using a cluster algorithm. This information quantitatively describes the width of the transition zones.

---

[*] Corresponding author. Tel.: +32 9 264 32 89; fax: +32 9 264 35 75.
E-mail address: Hugo.Caniere@UGent.be (H. Canière).

0301-9322/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijmultiphaseflow.2009.03.006

2. Experimental data

2.1. Capacitance probe and data acquisition

The capacitance probe has a concave electrode configuration. A single pair of sensing electrodes is sided by two pairs of guarding electrodes, one upstream and one downstream (Canière et al., 2007). The output of the probe is a voltage signal proportional to the electric capacitance of the two-phase mixture between the sensing electrodes. To acquire (quasi)-local two-phase flow data, the electrode width is equal to the diameter of the tube. The electronic transducer used to convert the electric capacitance to a voltage signal is based on the charge/discharge principle (Yang and Yang, 2002).

The sensor was used for horizontal air–water flow in a small diameter tube of 9 mm ID. The output voltage signal of the capacitance probe, V_{signal} is made dimensionless using Eq. (1) with V_{air} and V_{water} the voltage output of air only and water only flow.

$$V^* = \frac{V_{\text{signal}} - V_{\text{air}}}{V_{\text{water}} - V_{\text{air}}} \tag{1}$$

A series of 189 sensor signals was gathered, at different mass velocities of air (G_{air} = 0.3–50 kg/m^2 s) and water (G_{water} = 30–700 kg/m^2 s). The measurement uncertainties are described in Appendix A. Each signal was classified into one of three flow regimes: stratified flow (22 points), intermittent flow (104 points) and annular flow (63 points). By using this three-category classification a distinction can be made between stratified and non-stratified flows as well as a distinction between the ring-shaped annular flows and the more complex flow structures of intermittent flows. Stratified flow can have a smooth as well as a wavy interface. Intermittent flow can be slug flow, plug flow, elongated bubble flow, etc. All two-phase flows considered as intermittent flow have an irregular or aperiodic flow structure and are therefore grouped together. Visual classification is difficult and subjective in nature, despite the use of high speed camera images. Sensor signals analysis can provide quantitative and more objective criteria as will be shown.

2.2. Feature description

The sensor signals were gathered at a sample frequency of 1 kHz with a National Instruments DAQ system. From each signal, several statistical features were mined. A first group consists of the statistical moments of the sensor signal, i.e. the average value (AVG), the variance (M2), the skewness (M3) and the kurtosis (M4). These features determine the shape of the probability density estimation (PDE) of a signal and represent information of the signal in the amplitude domain. A second group consists of features in the frequency domain. The sensor signal is first transformed using a discrete Fourier algorithm and a power spectral density (PSD) is calculated. Based on this PSD, the contributions of different bandwidths can be added to define a frequency feature. The frequency range for gas–liquid interface phenomena is typically smaller than 100 Hz (Drahos and Cermak, 1989). Therefore, only contributions of frequencies lower than 100 Hz are considered. The frequency range from 0.5 Hz to 100 Hz is split into subranges at 5, 10, 20, 40, 60 and 80 Hz. The power spectrum contributions of the subranges (Eq. (2a)) result in a first set of seven features in the frequency domain.

$$F_k = \frac{\sum_{f=K_k}^{K_{k+1}} \mathrm{PSD}(V^*)}{\sum_{f=0.5\,\mathrm{Hz}}^{100\,\mathrm{Hz}} \mathrm{PSD}(V^*)}, \quad \text{with } K = [0.5, 5, 10, 20, 40, 60, 80, 100\,\mathrm{Hz}] \tag{2a}$$

$$F_k = \frac{\sum_{f=K_k}^{100\,\mathrm{Hz}} \mathrm{PSD}(V^*)}{\sum_{f=0.5\,\mathrm{Hz}}^{100\,\mathrm{Hz}} \mathrm{PSD}(V^*)}, \quad \text{with } K = [5, 10, 20, 40, 60, 80\,\mathrm{Hz}] \tag{2b}$$

A second set of six features in the frequency domain is determined by the power spectrum contributions starting at, respectively, 5, 10, 20, 40, 60 and 80 Hz, up to 100 Hz (Eq. (2b)). Each of these thirteen frequency features is normalized by dividing by the addition of all contributions up to 100 Hz. In total 17 signal features are used in the multivariate analysis. Remark that the sum of the components 0.5–5 Hz and 5–100 Hz is the constant one, and that both features carry the same information.

2.3. Multivariate analysis

The seventeen signal features are investigated for their ability of flow regime classification. First, a *Fisher Criterion* (Shawe-Taylor and Cristianini, 2004) was applied. A Fisher discriminant $J_{ii'}(k)$ is determined (Eq. (3)), with $\mu_i(k)$ the mean of feature k of the data points in class i and $\sigma_i(k)$ the variance of feature k of the data points in class i.

$$J_{ii'}(k) = \frac{[\mu_i(k) - \mu_{i'}(k)]^2}{\sigma_i(k) + \sigma_{i'}(k)} \tag{3}$$

The score of the Fisher Criterion for a selected feature is then the average of $J_{ii'}$ for all combination of classes i and i'. This criterion quantitatively determines whether a feature is able to separate class i from class i'. The results are shown in Table 1. The features with the highest Fisher Score are the variance (M2) and the frequency feature with bandwidth 10–100 Hz. This frequency feature is further called High Frequency Contribution Factor (HFCF) because it excludes the contribution of slow slug flow phenomena and includes the contributions from phenomena with higher frequencies such as interfacial waves. Wang et al. (1991) found typical peaks at 0–5 Hz in the PSD of slug flows from local impedance measurements of vertical air–water flow.

Secondly, a *correlation coefficient CC* (Eq. (4)) is calculated of every feature, F, with the mass velocities of air and water. cov and var are the covariance and variance respectively. The two features that correlate best with G_{water} are the AVG and the M2 (Table 2). The optimal features with G_{air} are the HFCF and the AVG. The frequency features do not correlate well with G_{water}. The 0.5–5 Hz feature and the 5–100 Hz feature correlate almost as good with G_{air} as the HFCF. But these features are strongly correlated to the HFCF.

$$CC = \frac{\mathrm{cov}(F, G_i)}{\sqrt{\mathrm{var}(F)\mathrm{var}(G_i)}}, \quad \text{with } i = \text{water or air} \tag{4}$$

From the results of both tests, AVG, M2 and HFCF are found to be most suitable for flow regime classification. When the same feature selection procedure is applied to 100 permutations of 150 signals, the same optimal features are found for 79% of the permutations. In the other cases, the 5–100 Hz feature was selected, which is strongly correlated to 10–100 Hz.

The flow regime classification potential of the selected features is visualized in Fig. 1. On the diagonal of the plot matrix, a histogram of the respective variable is shown. Every other subplot is a scatter plot of the variables at the row and column index. Annular

Table 1
Results of the Fisher Criterion on the signal features.

Feature	Fisher score	Feature (Hz)	Fisher score	Feature (Hz)	Fisher score
AVG	0.556	0.5–5	0.736	5–100	0.736
M2	**1.238**	5–10	0.114	**10–100**	**1.122**
M3	0.034	10–20	0.212	20–100	0.916
M4	0.022	20–40	1.052	40–100	0.514
		40–60	0.612	60–100	0.321
		60–80	0.361	80–100	0.219
		80–100	0.219		

Table 2
Correlation coefficients of the signal features with the mass velocities.

Feature	CC (G_{water})	CC (G_{air})
AVG	0.755	0.781
M2	0.843	0.540
M3	0.389	0.305
M4	0.489	0.119
0.5–5 Hz	0.127	0.783
5–10 Hz	0.571	0.197
10–20 Hz	0.158	0.675
20–40 Hz	0.161	0.790
40–60 Hz	0.251	0.697
60–80 Hz	0.201	0.602
80–100 Hz	0.153	0.518
5–100 Hz	0.127	0.783
10–100 Hz	0.092	0.843
20–100 Hz	0.199	0.771
40–100 Hz	0.231	0.666
60–100 Hz	0.188	0.583
80–100 Hz	0.153	0.517

flows typically have a high HFCF and low to intermediate values of AVG and M2. Stratified flows have low to intermediate values of AVG, but only low values of M2. The HFCF is spread over the full range because both smooth and wavy stratified flows are present. Intermittent flows have intermediate to high values of AVG and M2 and low to intermediate values of HFCF. Combining amplitude information (AVG, M2) and frequency information (HFCF) makes it possible to separate the flow regimes in the feature space. These features are now used to build an objective flow regime classifier and additional probabilistic flow regime information.

3. Statistical two-phase flow modeling

3.1. Fuzzy c-means clustering algorithm (Bezdek, 1981)

A clustering algorithm is an *unsupervised learning method*. The goal of such a method is to deduce properties from a dataset, without the help of a supervisor providing correct answers for each observation. In the case of two-phase flow classification, no visual decisions are needed. Clustering analysis tries to group a collection of objects into subsets or *clusters* such that those within each cluster are more closely related to one another than objects assigned to different clusters. An object is a selection of input features deduced from a sensor signal. The choice of these input features is fundamental to the clustering technique. In accordance with the multivariate analysis, the selected input features matrix I is [AVG, M2, HFCF]. The choice of a dissimilarity measure between two objects, the *distance function*, is a second important factor. By far the most common choice of the distance function is the squared or Euclidian distance D between two objects j and j' (Eq. (5)) with weight parameters w.

$$D(j,j') = w_{\mathrm{AVG}} \cdot (\mathrm{AVG}_j - \mathrm{AVG}_{j'})^2 + w_{\mathrm{M2}} \cdot (\mathrm{M2}_j - \mathrm{M2}_{j'})^2 + w_{\mathrm{HFCF}} \cdot (\mathrm{HFCF}_j - \mathrm{HFCF}_{j'})^2 \quad (5)$$

This is a weighted average of squared feature distances. Each object is iteratively assigned to one cluster based on the minimization of an objective function. Each of the weight parameters can be chosen to set the relative importance of the features upon the degree of similarity of the objects. The relative importance of each feature is proportional to its variance over the data set. Setting $w_k = 1/(2\mathrm{var}_k)$ will cause each of the features to equally influence the overall dissimilarity between pairs of objects (Appendix B.2). Variables that are more relevant in separating the clusters should of course be assigned a higher influence in defining object dissimilarity.

The fuzzy c-means clustering algorithm is a soft-clustering algorithm. This means that each data point is assigned to a cluster to some degree that is specified by a membership grade. This makes it possible to describe the boundaries between clusters in a smooth way. Since the aim of the signal clustering is finding a probabilistic description of flow regime boundaries, this soft-clustering algorithm is the preferred choice amongst other clustering algorithms like k-means clustering or hierarchical clustering. The mathematical implementation is more elaborately described in the Appendix B.1.

3.2. Data regression

The output of the clustering algorithm is a centroid, c_i, for every cluster, together with a membership grade for every object to each

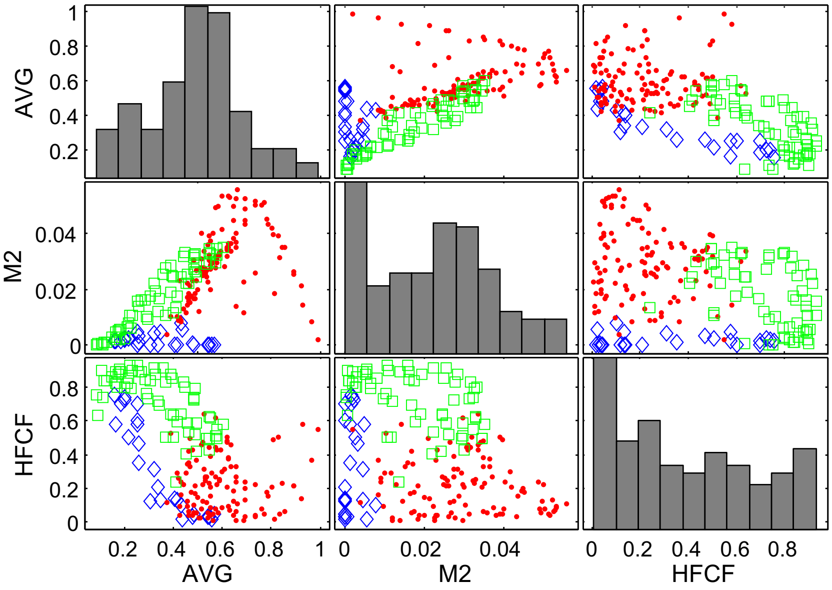

Fig. 1. Multivariate plot of signal features (blue open diamond, stratified; red filled circle, intermittent; green open square, annular). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

cluster. A centroid can be considered as typical or characteristic for the specific cluster. In this application, the two-phase flow corresponding with the cluster centroid is regarded as typical for that flow regime. The membership grade is related to the distance between the object and each centroid in feature space and can thus be considered as flow regime probabilities for that data point.

From the measurement points, an estimation of the cluster membership grades is obtained for the 189 points in the G_{water}–G_{air} plane. To extrapolate the scattered data to continuous data in the full plane, a regression technique is needed that creates a regression surface for every cluster. The probabilities are strongly nonlinear in the G_{water}–G_{air} plane, so a linear regression is not sufficient. A polynomial multivariate regression PR is compared with ε-Support Vector Regression (SVR) and ν-SVR using radial basis function kernels (Appendix C), (Schölkopf and Smola, 2002; Chang and Lin, 2001). ε-SVR and ν-SVR gave similar results, so only ε-SVR is further considered. Standard settings of the parameters in the regression technique ($\varepsilon = 0.1$, $\sigma = 1/k = 1/3$ and $C = 1$) are compared with those obtained from a grid search evaluated by fivefold cross-validation. The latter reveal a detailed structure of the probabilities in the G_{water}–G_{air} plane. This structure is not surprising because of the complex connections between the three features and the air and water mass velocities. In this study, the focus is on the global structure of the inlet signals. Therefore standard settings are further used. In the case of multivariate polynomial regression, the degree of the polynomials is limited to two for the same reason. After regression of all probability surfaces, normalization is performed to obtain fractions in terms of percentage in the end.

The final step in building the flow regime classifier is applying a maximum probability criterion to the probability surfaces to track the transition boundaries in the G_{water}–G_{air} plane. This means that the flow regime boundaries are found at the intersection of the two surfaces with the highest probability.

3.3. Test case results

A test case was run which divides the full data set in three clusters. The Euclidian distance function was used and the selected input feature matrix was $I = [\mathrm{AVG}, \mathrm{M2}, \mathrm{HFCF}]$ with all weight parameters set to one. Fig. 2 shows the data points in feature space after clustering into three clusters. Remark that there is no clear distinction between the different clusters. A hard clustering algorithm will therefore not give useful results. The data is directed in three main directions. Each cluster centroid is directed into one of these directions.

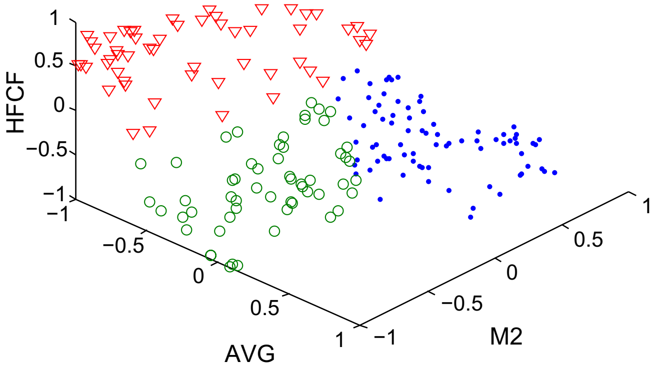

Fig. 2. Data clusters in features space from the test case.

Fig. 3 shows the clusters of the test case in a G_{air}–G_{water} map. The solid lines are the boundaries found at the intersection of the regression surfaces. The contours are the normalized probabilities after regression and the scatter data indicate the division of the data points into the three clusters. The clustering algorithm divides the data points into clearly separable zones of the map and the regression technique is able to track the boundaries between the clusters in the mass velocity plane. Only a few data points are located at the wrong side of the boundaries because of small errors due to the smoothing characteristic of the regression. With SVR, the root mean squared errors for each of the three probabilities are 0.17718, 0.10837 and 0.13514. This is a little better than the second order PR: 0.23108, 0.14513 and 0.16842.

3.4. Model parameters

3.4.1. Input features

The input features provide the cluster algorithm with the necessary flow regime data and therefore have a major influence up on the final probabilistic map. In Fig. 4, the result of the data processing technique is shown when individual signal features are applied. It is very unlikely that a single parameter can separate signals into three meaningful classes. Therefore a 2-cluster classification was executed.

The algorithm divides the signal data in two groups with a diagonal boundary when only AVG is applied. The AVG is a measure of the void fraction and is therefore influenced by both air and water mass fluxes. Using this parameter, the model can separate signals with corresponding low void fractions from signals with corresponding high void fractions. When time-averaged void fraction is an important parameter for the desired flow classification, the corresponding weight parameter should be set high.

Using only the M2 as input feature a more vertical boundary is found at intermediate water mass flux (60–200 kg/m^2 s). From the multivariate plot in Fig. 1, it is clear that the variance has the po-

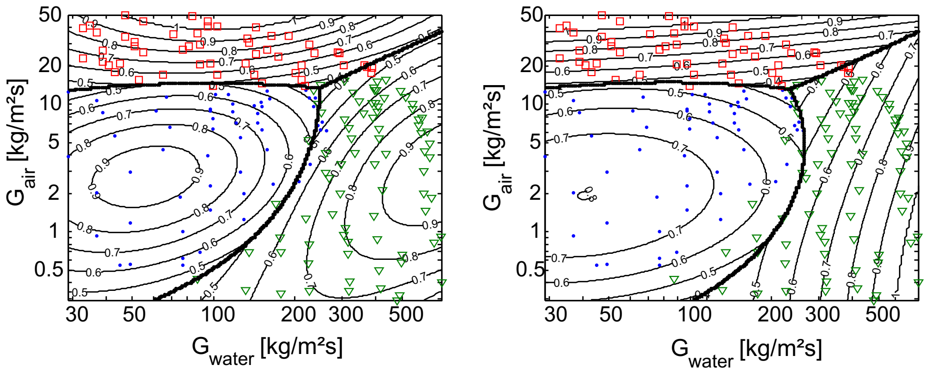

Fig. 3. Probability map with cluster points (blue closed circle, cluster 1; red open square, cluster 2; green open down triangle cluster 3) and regression surface intersection boundaries (left) SVR (right) 2nd order PR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

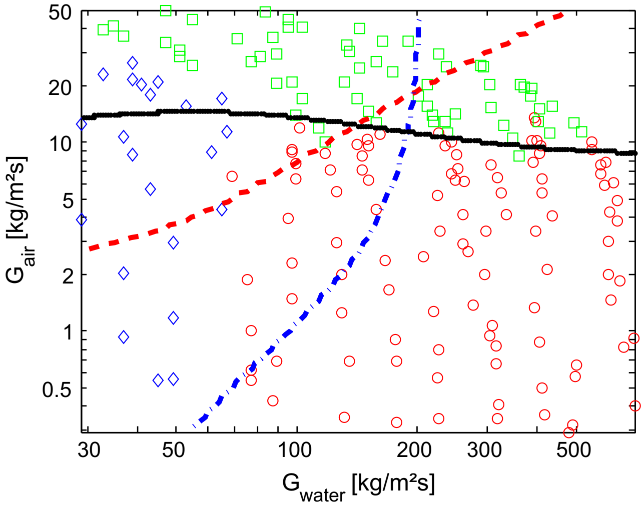

Fig. 4. Influence of input features: our visual classification data (blue open diamond, stratified; red open circle, intermittent; green open square, annular) and SVR cluster boundaries of individual features (red dashed line, AVG; blue dash-dot line, M2; black solid line, HFCF). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

tential to separate stratified flows from intermittent flows. However, the boundary found by the algorithm does not fully coincide with the visual stratified-intermittent flow transition, but is located at higher mass fluxes. When a single water slug occurs every few seconds, a second but small peak in the PDE of the corresponding signal will appear at high V^* (see Fig. 9). Because this peak is small, it has only a limited influence on M2. Therefore flows with water slugs occurring with low frequency are visually classified as intermittent flow. But according to the algorithm, they show more similarity with pure stratified flow, when only M2 is used as parameter.

Applying the HFCF as input feature a quasi-horizontal boundary is found around $G_{air} = 10$ kg/m^2 s. This line coincides with the visual transition from annular to intermittent flow. The HFCF can also separate smooth stratified flows from stratified-wavy flows. As during the visual classification no distinction was made between these two flow types, the boundary tears across the stratified flow area.

3.4.2. Weight parameters

To compare the influence of more than one feature the weights should be set equal to $w_k = 1/(2\text{var}_k)$ (see Appendix B.2). The data set was first column-wise normalized so each feature is mapped to a $[-1,1]$ space. Then, the variances and the corresponding weight parameters for equal influence of each feature were calculated (Table 3). Using these weight parameters, the influence of the AVG is enhanced and the influence of HCFC is diminished compared to the test case. According to the input feature analysis, this should result in more diagonal boundaries. The new cluster centroids are not anymore directed in three distinct direction but are now almost in line. The result of using equally weighted input features is shown in Fig. 5.

3.4.3. Number of clusters

When increasing the number of clusters, the clustering method still groups the measurement points in separable areas of the flow map and the intersection of the regression surfaces agree well with this grouping. This can be an interesting characteristic for applications where more than three characteristic groups are advisable.

Table 3
Variances and weight parameters by feature.

Feature	Variance	$w_k = 1/(2\text{var}_k)$
AVG	0.171	2.919
M2	0.285	1.753
HFCF	0.383	1.305

4. Discussion

4.1. Coupling with heat transfer and pressure drop modeling

It is commonly accepted that flow regime mapping should be included into two-phase flow modeling of heat transfer and pressure drop. The probability map technique can be used for this purpose (Jassim et al., 2007, 2008a, 2008b; Jassim, 2006). Rather than purely classifying a certain flow into a single flow regime, flow regime specific correlations can be combined in a probabilistic way with smooth transition from one flow regime to another:

$$\text{Prediction} = \sum_{i=1}^{NC} P_i(G_{air}, G_{water}) \cdot \text{FSC}_i(G_{air}, G_{water}) \tag{6}$$

with FSC_i the flow regime specific correlation and P_i the corresponding probability. For optimizing the results to a given application, the weight parameters of the algorithm can be fitted to the data set. The number of categories or flow regimes (NC) can also be chosen according to the application and/or the fitting results. The resulting probabilities and flow regime boundaries will of course differ with different input parameters corresponding with different applications.

4.2. Comparison with the visual classification and the Taitel–Dukler map

In Fig. 6, a comparison is made between the boundaries found in the test case and a visual classification (into stratified-intermittent-annular flow) based on high speed camera images. The boundaries found by the algorithm divide the flow map into three zones, i.e. zone 1 in the lower left corner, zone 2 in the lower right corner and zone 3 in the upper part of the flow map. Zone 3 contains most of the observed annular flows and the stratified-wavy

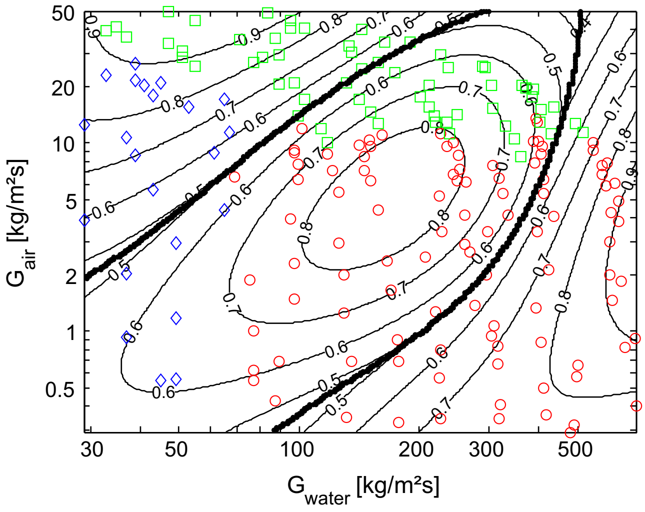

Fig. 5. Probabilistic flow regime map with equally weighted input features: cluster boundaries (solid), probabilities (contour) and our visual classification data (blue open diamond, stratified; red open circle, intermittent; green open square, annular). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

flows which are classified as stratified flow. Zone 1 contains all smooth stratified flows and some intermittent flows with water mass velocities lower than 200 kg/m^2 s. Zone 2 covers intermittent flows at higher mass velocities as well as some annular flows. The boundary between zone 1 and zone 2 occurs at higher water mass fluxes compared to the transition from stratified to intermittent flow. However, during our visual classification of stratified flows no liquid slugs were allowed at all. This was very strictly imposed even when only one slug occurred every few seconds. The cluster algorithm does not penalize that strict on slugs at very low frequency for the settings used.

The clustering flow map is also compared to the theoretical flow map of Taitel and Dukler (1976) (and modifications by Barnea et al. (1983) and Barnea (1987)). Transition lines are drawn for a 9 mm tube at atmospheric conditions of air and water (line A = stratified/non-stratified boundary, line B = modified intermittent-stratified boundary (Barnea et al., 1983), line C = smooth stratified/stratified-wavy boundary and line D = annular-intermittent boundary). There is a remarkable agreement in location and orientation of the cluster boundary between zone 2 and zone 3 and the Taitel–Dukler boundary D. The boundary between zone 1 and zone 3 converges with the Taitel–Dukler boundary A at low water mass fluxes. The stratified-intermittent transition does differ with Taitel–Dukler boundary.

If a better agreement with the visual classification is desired, more flow regimes should be taken into account. Then, the algorithm can try to separate different sub regimes in the intermittent flow area, as well as make a distinction between smooth stratified and stratified-wavy flows. The weight parameters can also be set to optimize the agreement. But this was not an intension in this work.

4.3. Discussion of the probabilistic flow regime map for horizontal two-phase flow

In Fig. 7, a probabilistic flow map is presented applicable to air–water flow in horizontal smooth tubes of 9 mm ID at mass velocities ranging between $G_{air} = 0.3–50$ kg/m^2 s and $G_{water} = 30–700$ kg/m^2 s under near atmospheric conditions. Extrapolation beyond the mass velocity ranges and to other tube diameters is not advisable. The settings of the test case were used to create this map. The equations of the probabilities are given in Appendix D.

Some flow regimes and corresponding sensor signals are further discussed. For every flow regime, the data point nearest to the cluster centroid was selected. These points can be considered as typical for the corresponding flow regime. The corresponding labels are S for stratified, I for intermittent and A for annular flow. For each of the transitions, a data point was selected as well. These points are located on a straight line between the typical signals and close to the cluster boundaries of the test case. The mass fluxes, signal features and flow regime probabilities are listed in Table 4.

For every data point, three images of the flow were selected to illustrate the typical flow phenomena. These images were captured with a high speed camera at 260 fps. The dimensionless voltage of the capacitance sensor is plotted vs. time as well as the probability density estimation (PDE) and power spectral density (PSD) of the V^* signal. To better visualize the characteristics of the sensor signals, the scale of the ordinate was not set equal in every plot. One should keep this in mind when comparing different plots.

In the stratified flow signal (Fig. 8), an almost constant voltage signal is typical. This corresponds with a constant liquid level that can vary very slowly. The resulting PDE has a sharp peak and so a very small variance M2. There is no important frequency content, which makes the HFCF minimal.

When following a straight line on the flow map towards the data point of intermittent flow, the water mass flux increases. Liquids slugs appear which momentarily fill the entire cross section of the tube with water. The sensor signal of the data point SI at the transition boundary (Fig. 9) clearly indicates the presences of short liquid slugs. In the PDE, a second but small peak appears at V^* equal to one and the PSD shows an important low frequency content (<10 Hz). This signal would definitely be classified as slug flow. According to the probabilities indicated in Table 4, this flow has as much in common with the signal of stratified flow as with the signal of intermittent flow (Fig. 10). The intermittent flow I has a higher water mass flux. A typical double peak is visible in the PDE. The most important contribution in the PSD has frequencies lower than 10 Hz. In this chaotic type of flow, aerated liquid slugs are present as well as air plugs, bubbles of all shapes and diameter and geometrically very complex interface structures.

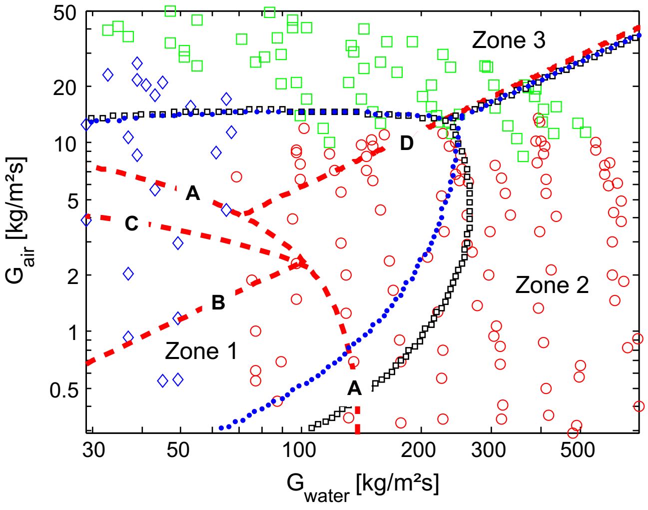

Fig. 6. Flow map with cluster boundaries (blue filled circle, SVR; black open square, 2nd order PR), Taitel–Dukler boundaries (dashed) and our visual classification data (blue open diamond, stratified; red open circle, intermittent; green open square, annular). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

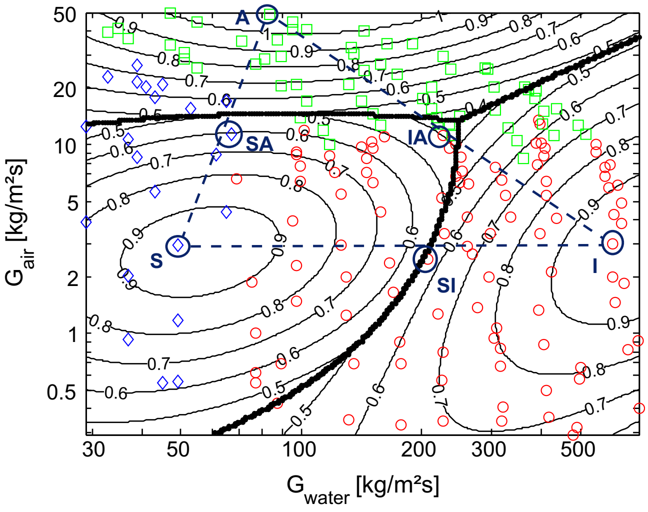

Fig. 7. Probabilistic flow regime map: SVR cluster boundaries (solid), probabilities P (contour) and our visual classification data (blue open diamond, stratified; red open circle, intermittent; green open square, annular). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 4
Selected signals, signal features and probabilities.

Flow regime	Label	G_{water}	G_{air}	AVG	M2	HFCF	p_S	p_I	p_A
Stratified	S	49.04	2.96	0.479	4.66e−4	0.022	0.944	0.05	0.006
Intermittent	I	601.78	2.99	0.643	4.44e−2	0.241	0.061	0.939	0
Annular	A	83.33	48.91	0.147	2.72e−3	0.858	0	0	1
S-I transition	SI	208.41	2.47	0.497	1.98e−2	0.182	0.482	0.518	0
I-A transition	IA	226.56	11.28	0.476	2.28e−2	0.431	0.397	0.328	0.275
S-A transition	SA	67.20	11.43	0.321	4.41e−3	0.311	0.606	0	0.394

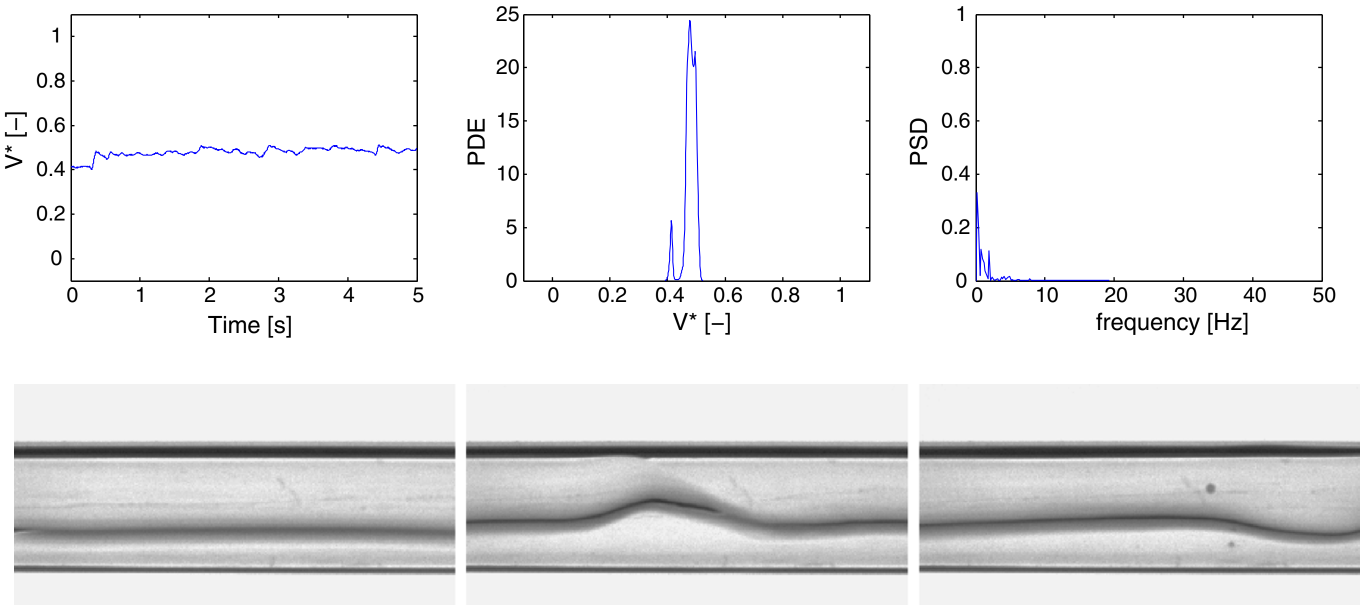

Fig. 8. Stratified flow characterization.

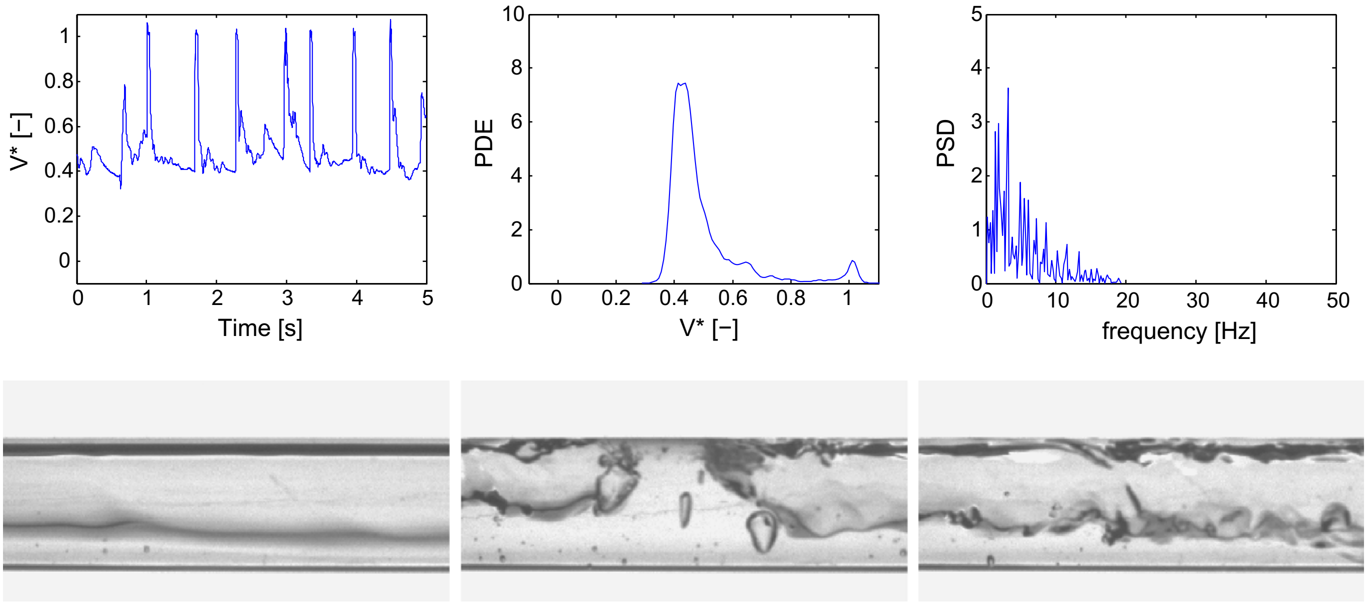

Fig. 9. Stratified-intermittent transitional flow characterization.

Following the trajectory towards annular flow, the liquid bridges gradually break up and a central air core develops (Fig. 11). The second peak in the PDE disappears and the HFCF increases. This data point is located very closely to the intersection of the three boundaries. Purely classifying this flow into one of the typical flow regimes does not make a lot of sense in that area. This probabilistic technique makes it possible to appropriately combine all typical flow regime characteristics.

Going towards fully developed annular flow (Fig. 12), the peak in the PDE moves to lower V^* or higher void fraction. The peak gets

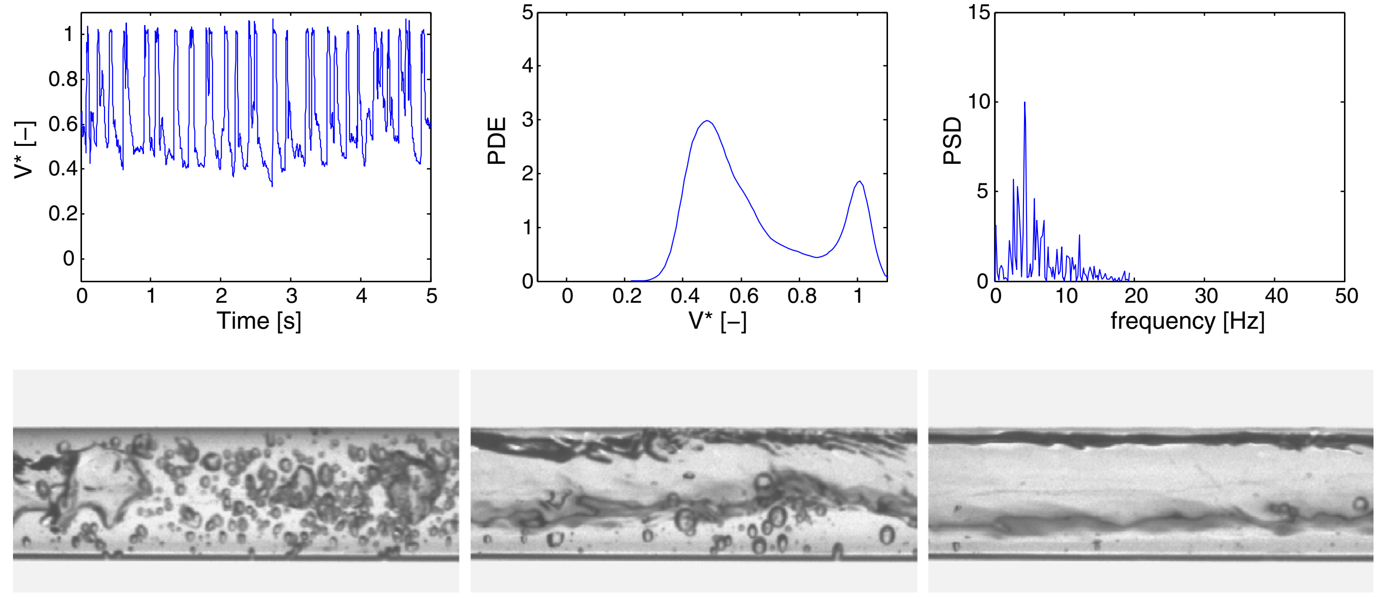

Fig. 10. Intermittent flow characterization.

sharper which results in a lower M2 value. The HFCF is higher than all other signals because of the frequencies of the disturbances on the ring-shaped interface. Finally going back to stratified flow, a wavy type of flow (Fig. 13) is detected at the stratified-annular transition boundary. The average signal value increases, as well as the M2 value. Due to the smaller air mass flux, the liquid does not reaches the top of the tube, but the air–water interface is still not smooth. The HFCF gradually diminishes back towards almost zero for stratified flows.

This discussion proves that the sensor can track the typical flow phenomena of different flow regimes, characterized by the corresponding time signal, PDE and PSD. The signal feature selection in the multivariate analysis was successful. The flow phenomena propagated well into the features. This is reflected in the outcome of the clustering algorithm which pointed out typical flow regimes in the flow map and quantitatively tracks the transition boundaries without any visual intervention.

5. Conclusions

In this study the signals of a capacitance sensor were analyzed and processed to build a horizontal two-phase flow classifier without any subjective visual decisions. From a multivariate analysis, the average, the variance and a high frequency contribution factor of the capacitance signals were found most suitable for separating flow regimes.

The c-means clustering algorithm was applied together with a regression technique to create objective flow regime transition boundaries and deduce probabilistic flow regime data. This data makes it possible to combine flow regime dependent correlations in the two-phase flow models for heat transfer and pressure drop with smooth and appropriately quantified transitions from one flow regime to another. The influence of the different model parameters on the algorithm outcome is mapped and discussed.

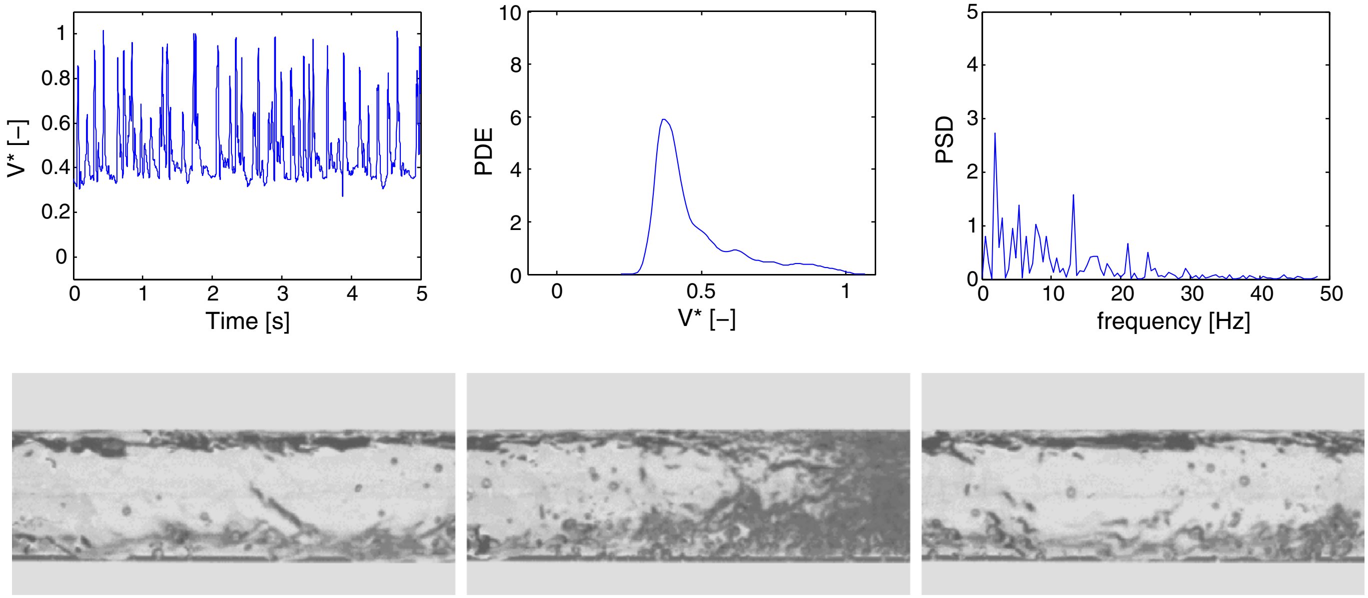

Fig. 11. Intermittent-annular transitional flow characterization.

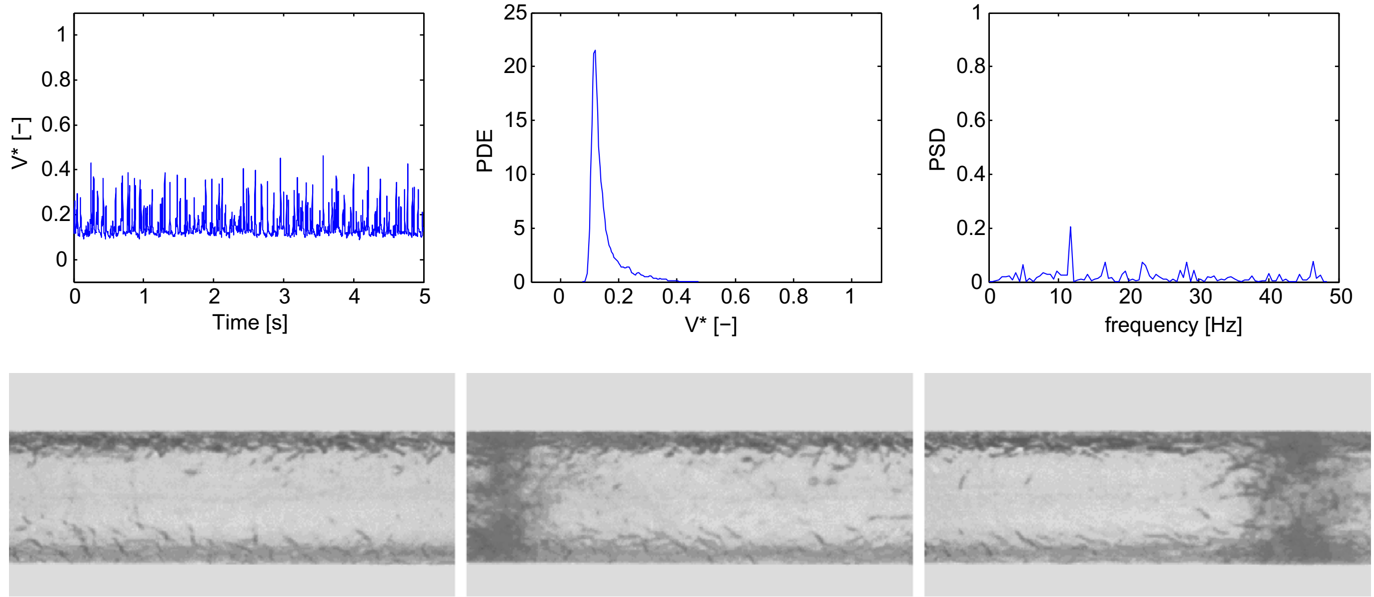

Fig. 12. Annular flow characterization.

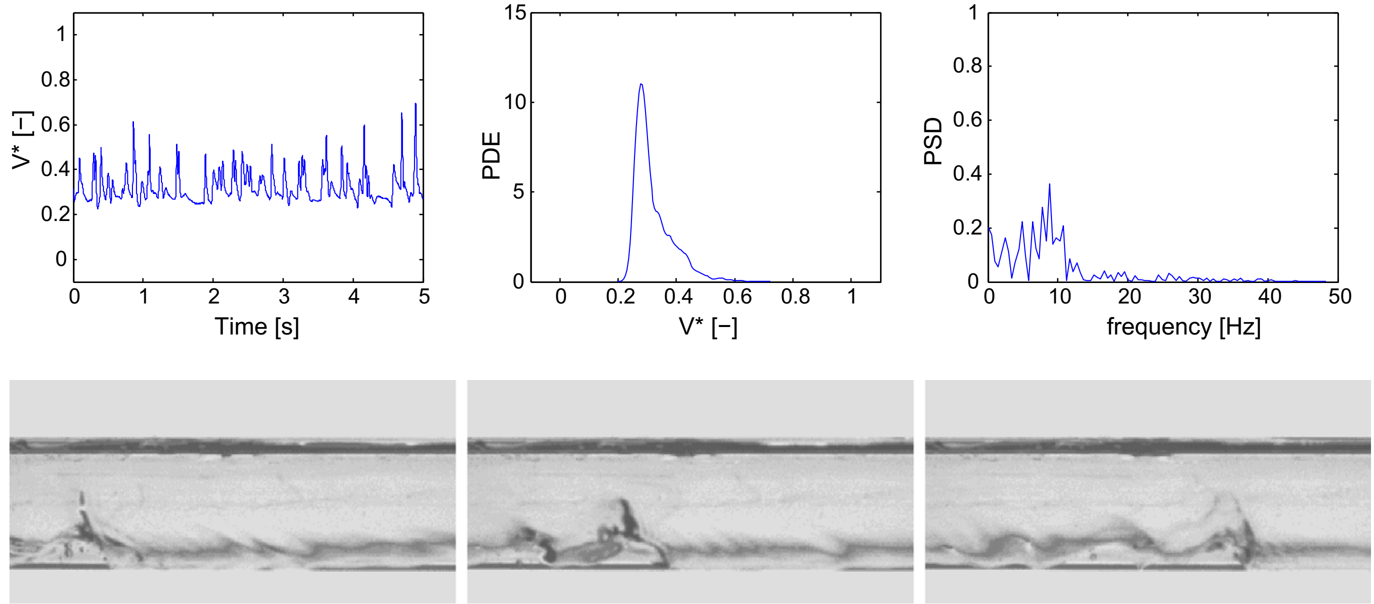

Fig. 13. Stratified-annular transitional flow characterization.

A probabilistic flow map for horizontal air–water flow in a 9 mm tube (G_{air} = 0.3–50 kg/m^2 s and G_{water} = 30–700 kg/m^2 s) is presented to quantify the flow transitions between stratified, intermittent and annular flow. These transition areas were further characterized and discussed with sensor signals and high speed camera images.

Acknowledgements

The authors would like to express gratitude to the BOF fund (B/06634) of the Ghent University – UGent which provided support for this study.

Appendix A. Measurement uncertainty analysis

A.1. Capacitance measurement

The capacitance transducer circuit was calibrated against fixed capacitances (with uncertainties of 0.5% of value) in the range of 0–35 pF. A linear relationship of the capacitance with respect to the output voltage is found. The coefficient of correlation is 0.995. The uncertainty in the voltage measurements is 1.6% of full scale. The accuracy of the DAQ system has a negligible effect upon the measurements.

A.2. Operating conditions

The accuracy of the air mass flow meter calibrated by the manufacturer is 2% of reading. With an uncertainty of 100 μm on the tube diameter of 9 mm, and air flow rates between 2×10^{-6} and 6×10^{-3} kg/s, the uncertainty on G_{air} is 3%. The water flow meter was calibrated with a stopwatch-weighing technique using a precise balance and a large tank. The uncertainty on the measurements of the calibration curve is ±0.002 kg/s. The uncertainty of G_{water} is 4.6% f.s. Temperature is measured at inlet of the sensor with a type-K thermocouple with an accuracy of 0.1 °C. Pressure is measured at the inlet with a relative pressure device up to 6 bar with ±0.05% accuracy.

Appendix B. Concepts of the fuzzy c-means clustering algorithm

B.1. Fuzzy c-means clustering algorithm (Bezdek, 1981)

Consider a data set of N measurement points of K features: x_{kj} with $k = 1,2,\ldots,K$ and $j = 1,2,\ldots,N$. In this application the features are the sensor signal parameters AVG, M2 and HFCF. A distance or dissimilarity $d_k(x_{kj},x_{kj'})$ is first defined between values of the kth feature. The distance between two objects j and j' is then:

$$D(x_j, x_{j'}) = \sum_{k=1}^{K} w_k \cdot d_k(x_{kj}, x_{kj'}) \tag{B.1}$$

This is a weighted average of the feature distances. By far the most common choice of the feature distance is the *Squared* or *Euclidian* distance:

$$d_k(x_{kj}, x_{kj'}) = (x_{kj} - x_{kj'})^2 \tag{B.2}$$

The user chooses a number of clusters (NC). The fuzzy c-means clustering algorithm then starts with initial guesses for the centers of each cluster, c_i. Initial cluster fractions u_{ij} are assigned to each data point in such a way that

$$\sum_{i=1}^{\mathrm{NC}} u_{ij} = 1 \tag{B.3}$$

The algorithm minimizes an objective function J_m (Eq. (B.4)) based on the distance between a data point x_j and a cluster centroid c_i:

$$J_m = \sum_{j=1}^{N} \sum_{i=1}^{\mathrm{NC}} u_{ij}^m \|x_j - c_i\|^2, \quad 1 \leqslant m \leqslant \infty \tag{B.4}$$

The parameter m (chosen at the default value 2 in this work) determines the smoothness of the cluster transitions. When m approaches 1, the cluster boundaries are sharp, when m approaches infinity, u_{ij} becomes constant in a cluster. The values of u_{ij} and c_i are iterated from an initial value until convergence using Eqs. (B.5) and (B.6).

$$u_{ij} = \frac{1}{\sum_{k=1}^{K} \left(\frac{\|x_j - c_i\|}{\|x_j - c_k\|} \right)^{\frac{2}{m-1}}} \tag{B.5}$$

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^m \cdot x_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{B.6}$$

B.2. Weight parameters (Hastie et al., 2001)

The input matrix of the cluster algorithm contains three sensor signal parameters. For each of them, a weight parameter w_k can be chosen to set the relative importance of the features upon the degree of similarity of the objects. It is important to realize that setting the weight w_k to the same value for each variable (every k) does not necessarily give all features equal influence. The influence of the kth feature, X_k, on the object dissimilarity $D(x_j,x_{j'})$ depends upon its relative contribution to the average object dissimilarity measure over all pairs of observations in the data set

$$\bar{D} = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{j'=1}^{N} D(x_j, x_{j'}) = \sum_{k=1}^{K} w_k \cdot \bar{d}_k \tag{B.7}$$

with

$$\bar{d}_k = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{j'=1}^{N} d_k(x_{kj}, x_{kj'}) \tag{B.8}$$

being the average dissimilarity on the kth attribute. Thus, the relative influence of the kth variable is $w_k \cdot \bar{d}_k$ and setting $w_k \propto 1/\bar{d}_k$ would give all attributes equal influence in characterizing overall dissimilarity between objects.

In the case of K features and the Euclidian distance function, Eq. (B.1) becomes

$$D(x_j, x_{j'}) = \sum_{k=1}^{K} w_k \cdot (x_{kj} - x_{kj'})^2 \tag{B.9}$$

In this case Eq. (B.8) becomes

$$\bar{d}_k = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{j'=1}^{N} (x_{kj} - x_{kj'})^2 = 2 \cdot \mathrm{var}_k \tag{B.10}$$

where var_k is the sample estimate of var(X_k). Thus the relative importance of each such variable is proportional to its variance over the data set. Setting $w_k = 1/\bar{d}_k$ for all features will cause each one of them to equally influence the overall dissimilarity between pairs of objects $(x_j,x_{j'})$. Although this may seem reasonable, it can be highly counterproductive as well. If the goal is to divide the data into groups of similar objects, all features may not contribute equally to the notion of dissimilarity between objects. Some feature value differences may reflect greater actual object dissimilarity. Variables that are more relevant in separating the groups should be assigned a higher influence in defining object dissimilarity.

Appendix C. Support vector regression

Consider a datavector x_j with corresponding values y_j. When applying a linear regression technique, a straight line is searched for under the condition that all data points must be as close as possible to that line. With multiple features, an optimal hyperplane (Eq. (C.1)) is looked for (with $\langle w,x \rangle$ the dot-product of w and x) (see Fig. C.1).

$$f(x) = \langle w, x \rangle + b \tag{C.1}$$

The deviation of y from $f(x)$ of new samples x needs to be as small as possible. This is achieved by optimizing the flatness, which can be expressed as $||w||^2$. On the other hand, a sufficient accuracy is also necessary by trying to keep all data points within a margin ε of the hyperplane. This causes the optimization problem, with tuning parameters ε and C:

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1} (\zeta_i + \zeta_i^*) \tag{C.2a}$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leqslant \varepsilon + \zeta_i \\ \langle w, x_i \rangle + b - y_i \leqslant \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \tag{C.2b}$$

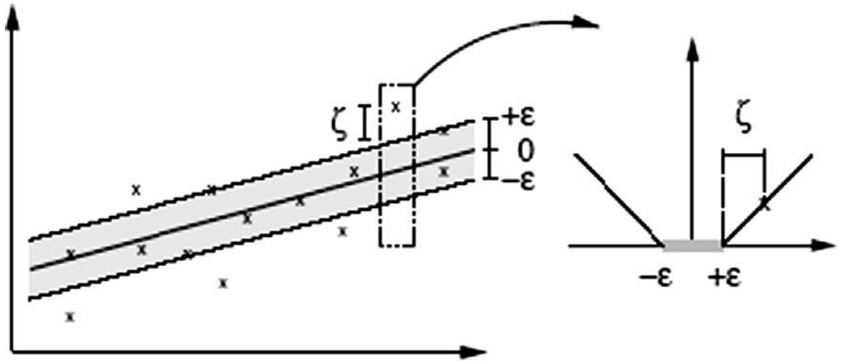

Fig. C.1. Linear support vector regression.

The optimal solution can be written as:

$$f(x) = \sum_{i=1} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{C.3}$$

In general, α and α^* can be uniquely determined. The algorithm and the result only uses scalar products. When the data samples are projected to another space $\varphi(x)$, then only $\langle x_1, x_2 \rangle$ should be replaced by $\langle \varphi(x_1), \varphi(x_2) \rangle = K(x_1, x_2)$. In this way, non-linear function can easily be fitted. The standard choice of kernel K in this work is a radial basis function:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \tag{C.4}$$

Appendix D. Probability functions

Eqs. (D.1)–(D.3) are the probability surfaces obtained with a second order polynomial regression and under test case conditions. These surfaces differ slightly from the SVR surfaces which are used in this paper. The SVR surfaces are a better regression of the data and are therefore preferred, but cannot easily be written in an analytical form. Eqs. (D.1)–(D.3) are applicable for air–water flows in a 9 mm tube with 30 kg/m^2 s < G_{water} < 700 kg/m^2 s and 0.3 kg/m^2 s < G_{air} < 50 kg/m^2 s.

$$X_1 = 0.388 \ln(G_{\text{air}}) - 0.523 \tag{D.1a}$$

$$X_2 = 0.627 \ln(G_{\text{water}}) - 3.109 \tag{D.1b}$$

$$\begin{aligned} Y_1 &= 0.652 - 0.199X_1 - 0.313X_2 - 0.55X_1^2 - 0.221X_2^2 + 0.123X_1X_2 \\ Y_2 &= 0.315 - 0.215X_1 + 0.382X_2 - 0.082X_1^2 + 0.261X_2^2 - 0.020X_1X_2 \\ Y_3 &= 0.033 + 0.414X_1 - 0.070X_2 + 0.635X_1^2 - 0.039X_2^2 - 0.103X_1X_2 \end{aligned} \tag{D.2}$$

$$P_{\text{strat}} = \frac{Y_1}{Y_1 + Y_2 + Y_3}, \quad P_{\text{int}} = \frac{Y_2}{Y_1 + Y_2 + Y_3}, \quad P_{\text{ann}} = \frac{Y_3}{Y_1 + Y_2 + Y_3} \tag{D.3}$$

References

Barnea, D., 1987. A unified model for predicting flow-pattern transitions for the whole range of pipe inclinations. Int. J. Multiphase Flow 13, 1–12.

Barnea, D., Luninski, Y., Taitel, Y., 1983. Flow pattern in horizontal and vertical 2 phase flow in small diameter pipes. Can. J. Chem. Eng. 61, 617–620.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Canière, H., T'Joen, C., Willockx, A., De Paepe, M., Christians, M., van Rooyen, E., Liebenberg, L., Meyer, J., 2007. Horizontal two-phase flow characterization for small diameter tubes with a capacitance sensor. Measur. Sci. Technol. 18, 2898–2906.

Chang, C., Lin, C., 2001. LIBSVM: a library for support vector machines. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Drahos, J., Cermak, J., 1989. Diagnostics of gas–liquid flow patterns in chemical engineering systems. Chem. Eng. Process. 26, 147–164.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York.

Jassim, E., 2006. Probabilistic flow regime map modeling of two-phase flow, Ph.D. Thesis, Urbana-Champaign, University of Illinois, IL, USA.

Jassim, E., Newell, T., 2006. Prediction of two-phase pressure drop and void fraction in microchannels using probabilistic flow regime mapping. Int. J. Heat Mass Transfer 49, 2446–2457.

Jassim, E., Newell, T., Chato, J., 2007. Probabilistic determination of two-phase flow regimes in horizontal tubes utilizing an automated image recognition technique. Exp. Fluids 42, 563–573.

Jassim, E., Newell, T., Chato, J., 2008a. Prediction of refrigerant void fraction in horizontal tubes using probabilistic flow regime maps. Exp. Therm. Fluid Sci. 32, 1141–1155.

Jassim, E., Newell, T., Chato, J., 2008b. Prediction of two-phase condensation in horizontal tubes using probabilistic flow regime maps. Int. J. Heat Mass Transfer 51, 485–496.

Nino, V., Hrnjak, P., Newell, T., 2003. Two-phase flow visualization of R134A in a multiport microchannel tube. Heat Transfer Eng. 24, 41–52.

Schölkopf, B., Smola, A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel methods for pattern analysis. Cambridge University Press, Cambridge, UK.

Taitel, Y., Dukler, A., 1976. Model for predicting flow regime transitions in horizontal and near horizontal gas–liquid flow. AIChE J. 22, 47–55.

Thome, J.R., 2004. Two-phase heat transfer using no-phase flow models? Heat Transfer Eng. 25, 1–2.

van Rooyen, E., Christians, M., Liebenberg, L., Meyer, J., 2007. Optical measurement technique for predicting time-fractions in two-phase flow. In: Proceedings of 5th International Conference on Heat Transfer, Fluid Mechanics, and Thermodynamics (HEFAT), Sun City, South Africa, 1st–4th July.

Wang, Y.W., Pei, B.S., Lin, W.K., 1991. Verification of using a single void fraction sensor to identify 2-phase flow patterns. Nucl. Technol. 95, 87–94.

Yang, S.X., Yang, W.Q., 2002. A portable stray-immune capacitance meter. Rev. Sci. Instrum. 73, 1958–1961.