HEFAT2008
6th International Conference on Heat Transfer, Fluid Mechanics and Thermodynamics
30 June to 2 July 2008
Pretoria, South Africa
Paper number: CH2

TOWARDS OBJECTIVE FLOW PATTERN MAPPING WITH THE K-MEANS CLUSTERING ALGORITHM

Canière H.*, T'Joen C. and De Paepe M.

Department of Flow, Heat and Combustion Mechanics, Ghent University - UGent,
St.-Pietersnieuwstraat 41, Gent, 9000, Belgium,
*E-mail: Hugo.Caniere@UGent.be

ABSTRACT

Two-phase flow regime prediction is of great importance for designing evaporators and condensers because the influence of the heat transfer coefficients is strongly related to the flow regimes. These flow regimes are often presented using flow pattern maps. As most flow pattern maps are based on visual observation, they lack objectively defined flow regime transition criteria. In order to add flow characteristics to the transitions boundaries, a sensor was developed which measures the capacitance of the two-phase flow. Due to the difference in dielectric constant of liquid and vapour and the dependency on the capacitance to the internal distribution of liquid and vapour in the cross-section of the tube, the sensor is able to characterize two-phase flow regimes. A large number of experiments was done with air-water flow. The setup was able to cover three main flow regimes for horizontal flow in a 9 mm tube, namely stratified, annular and intermittent flow. A multivariate analysis was performed to find characteristic signal parameters. The average signal value, together with the variance and a high frequency contribution factor were found suitable. These parameters were used as input features for the k-means clustering method, which groups the sensor data into a given number of classes. The influence of the weight parameters of the features was mapped, as well as the influence of the distance function. A comparison between the visual classification based on high speed camera images and the cluster classification shows a remarkable agreement.

NOMENCLATURE

B		Intersection boundary
c		Centroid
D, d	[-]	Distance
G	[kg/m²s]	Mass velocity
I		Input feature matrix
NC		Number of clusters
P, p	[-]	Probability
V	V	Voltage
V^*	[-]	Dimensionless voltage
w		Weight parameter
x	[-]	Vapour quality

Subscripts

i,i'	Cluster index
j,j'	Data point index
k	Feature index

INTRODUCTION

Upon the design of 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. Traditionally, this is achieved by classifying two-phase flows into flow regimes: stratified flow, annular flow, etc. These classifications are mainly based on visualisations (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 G-x, etc.). Attempts were made to create transition boundaries between these regimes based on theoretical assumptions [1-2]. 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. [3] introduced the probabilistic approach in multiport microchannels. Jassim and Newell [4] applied probabilistic flow regime mapping to predict pressure drop and void fraction in microchannels. van Rooyen et al. [5] used the same approach for intermittent flows during condensation in macroscale tubes.

Signal analysis can learn a lot about the subtle differences in flow phenomena and makes it possible to build quantifying classifiers or probabilistic data. In this study, a capacitance sensor was developed and tested with air-water flow. The sensor signals were analysed and three signal features were

chosen to build up a horizontal two-phase flow classifier without any subjective visual decisions. Probabilistic data was added to better describe the transition zones.

EXPERIMENTAL DATA

Capacitance Probe [6]

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. To acquire (quasi)-local two-phase flow data, the electrode width is equal to the diameter of the tube. The output of the probe is a voltage signal proportional to the capacitance of the momentary two-phase mixture between the sensing electrodes. The electronic transducer used to convert the electric capacitance to a voltage signal is based on the charge/discharge principle.

Experimental data acquisition

The sensor was used for horizontal air-water flow in a small diameter tube of 9 mm I.D. The output voltage signal of the capacitance probe is made dimensionless according to Eq. (1).

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

A series of 202 sensor signals was gathered, changing the mass velocities of air and water. Each signal was classified into one of three flow regimes: stratified flow (25 points), intermittent flow (116 points) and annular flow (61 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. Therefore sensor signals analysis can provide more quantitative and objective criteria.

Feature selection

From each sensor signal, several statistical parameters were mined. A first group consists of the statistical moments of the sensor signal, i.e. the average value (M1), the variance (M2), the skewness (M3) and the kurtosis (M4). These parameters determine the shape of the probability density function (PDF) of a signal and represent information of the signal in the amplitude domain. A second group consists of parameters from the frequency domain. The sensor signal is first transformed using a fast Fourier algorithm and a power spectral distribution (PSD) is calculated. Based on this PSD, the contributions of different bandwidths can be added to define a frequency parameter.

The statistical parameters of the data points were investigated for their ability of flow regime identification. The average value (AVG) and the variance (M2) of the signal amplitude together with a high frequency contribution factor (HFCF) were found to be most suitable for this purpose. The HFCF is the addition of the power spectrum contributions starting from 10Hz up to 100Hz divided by the addition of all contributions up to 250Hz. Annular 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.

DATA PROCESSING

The three sensor signal features (AVG, M2 and HFCF) are used to build an objective flow regime classifier and additional probabilistic flow regime information. Any visual training input is excluded, in contrast with other statistical models where visual input is required to determine signals of typical flow regimes. The k-means algorithm groups the data into subgroups of similar signals.

K-means clustering algorithm [7]

The K-means clustering algorithm is an *unsupervised learning* or *learning without a teacher* method. The goal of such a method is to deduce properties from a dataset, without the help of a supervisor or teacher providing correct answers for each observation. In the case of two-phase flow identification, 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. Fundamental to the clustering technique is the choice of the dissimilarity measure between two objects, often called *distance function* and the choice of the input features. In this application the input features are the sensor signal parameters AVG, M2 and HFCF. By far the most common choice of the feature distance is the *Squared* or *Euclidian* distance D(j,j') between two objects:

$$D(j,j') = w_{AVG} \cdot (AVG_j - AVG_{j'})^2 + w_{M2} \cdot (M2_j - M2_{j'})^2 + w_{HFCF} \cdot (HFCF_j - HFCF_{j'})^2$$
$$\text{with } w_{AVG} + w_{M2} + w_{HFCF} = 1 \tag{2}$$

This is a weighted average of the feature distances. Each observation is iteratively assigned to one cluster based on a minimization of a loss function. Each of the weight parameters, w_k, 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 [7]. Setting $w_k=1/(2\cdot var_k)$ will cause each of the features to equally influence the overall dissimilarity between pairs of objects. Although it may seem reasonable to do so, 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 of course be assigned a higher influence in defining object dissimilarity.

Generalizing data and creation of transition boundaries

The output of the clustering algorithm is a centroid c_i, for every cluster, together with the distance d_{ij} from each data point to every centroid. The data set is thus divided into the chosen number of clusters, based on the minimum distance to centroids. These distances can now be used to create probabilistic data.

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. This makes the distance in the feature space, from a data point to a centroid, d_{ij}, a measure of similarity between the flow regime of that data point and that of the centroid. When using Eq. (3), the distances are transformed into flow regime probabilities, p_{ij}. The sum of the probabilities over the number of clusters is of course 100% and the centroids it selves has 100% probability of the corresponding flow regime.

$$p_{ij} = \frac{1/d_{ij}}{\sum_{i=1}^{NC} 1/d_{ij}}, \text{ with NC the number of clusters} \qquad (3)$$

The distance data, as well as the probability data is scattered data. To obtain continuous data in a flow pattern map, the Support Vector Regression (SVR) technique [8-9] is applied. This regression technique creates a regression surface, P_i, for every cluster through the scattered distance data. Such a surface is visualized in Figure 1. The regression errors were minimized through a grid search of the SVR parameters.

The final step in defining flow regime transition boundaries is applying a maximum probability criterion to the SVR surfaces. This means that the flow regime boundaries are found at the intersection of the 2 surfaces with the highest probability.

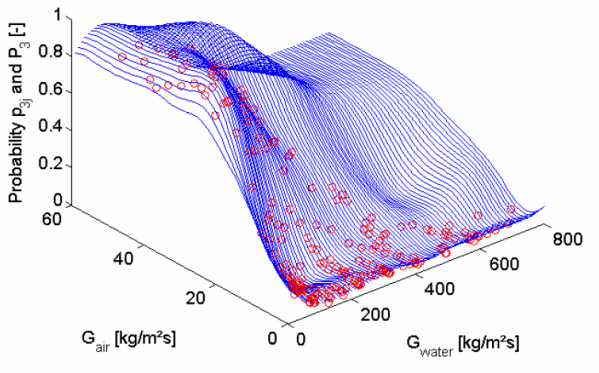

Figure 1 SVR surface with scatter probability data

In Figure 2, a flow chart is shown of the full data processing procedure. The probe signals and corresponding mass fluxes are fed from the database. A feature selection is done to extract the main flow characteristics from the signals into three features. The clustering algorithm divides the data set into clusters, finds cluster specific centroids c_i and calculates the distances d_{ij} between the centroids and all data points. At probability level the distances are transformed into probabilities p_{ij}. A regression is made to generalize the probabilities to smooth functions $P_i(G_{water},G_{air})$. Finally at classification level, boundaries $B_{ii'}$ are found at the intersections of the probability functions P_i. These boundaries separate zones on a G_{water}-G_{air} map which can be compared to visual flow pattern classification or theoretical prediction methods.

Figure 3 shows the results of a test case, which divides the full data set in three clusters. The Euclidian distance function was used and the selected input feature matrix was I = [AVG, M2, HFCF] with all weight parameters set to one. The solid

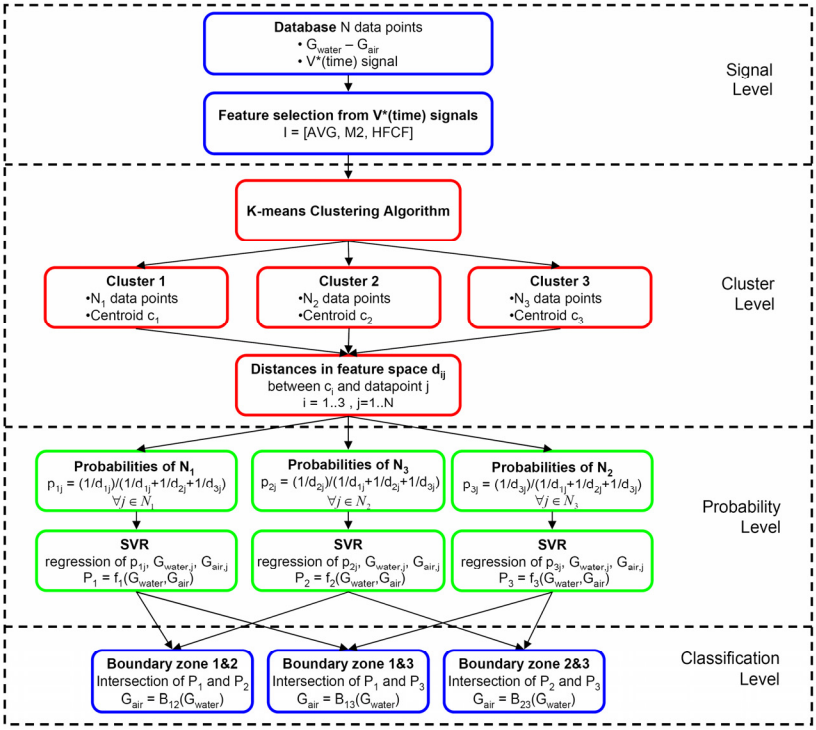

Figure 2 Flowchart of the data processing

lines are the boundaries $B_{ii'}$. Only a few data points are located at the wrong site of the boundaries because of the smoothing characteristic of the SVR regression.

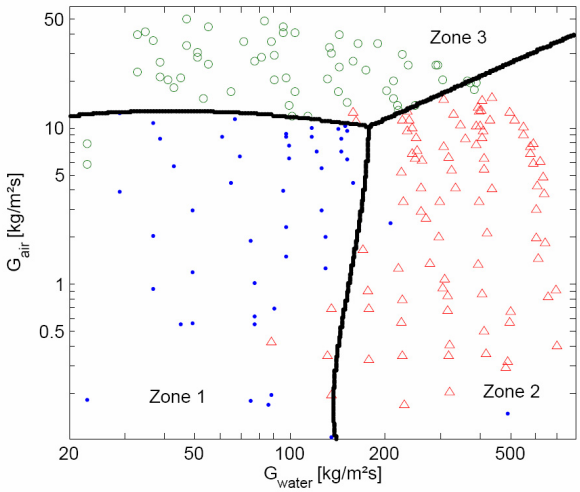

Figure 3 Clustermap with cluster classification data (• = cluster 1 – Δ = cluster 2 – o = cluster 3) and regression surface intersection boundaries

RESULTS

Comparison of cluster classification with visual classification and Taitel-Dukler map

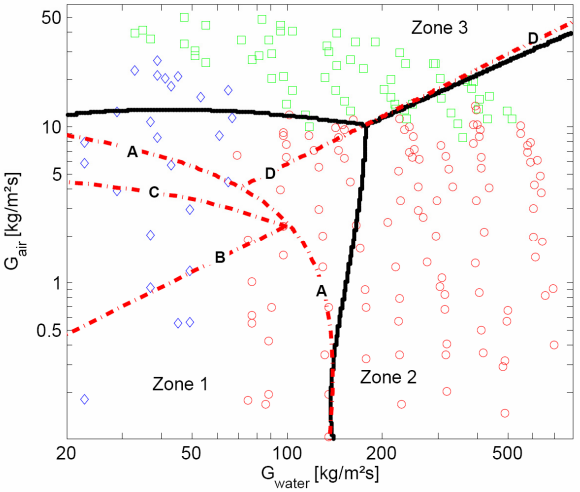

Figure 4 Flow Map with cluster boundaries (solid), Taitel-Dukler boundaries (dash-dot) and visual classification data (◊ = stratified – o = intermittent – □ = annular)

In Figure 4, the cluster boundaries of the test case are compared with the visual classification data. There is a good agreement between the traditional visual classification (into stratified-intermittent-annular flow) and the boundaries found by the clustering algorithm based on the sensor signal parameters. Zone 3 almost fully coincides with the annular flow regime area. The boundary between zone 1 and zone 2 is similar in shape to the transition from stratified to intermittent flow but occurs at higher water mass fluxes. However, during the visual classification of stratified flows no slugs were allowed at all. This was very strictly imposed even when there occurred only one slug every few seconds. The cluster algorithm does not penalize that strict on slugs at very low frequency. This may explain the discrepancy in location of the boundary. The clustering flow map is also compared with the theoretical flow map of Taitel and Dukler [1-2,10]. 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 [10], 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 boundaries and Taitel-Dukler boundaries. The boundary between zone 2 and zone 3 fully coincides with the theoretical boundary between annular flow and intermittent flow. This proves the proper selection of the input features.

Probabilistic map for horizontal two-phase flow

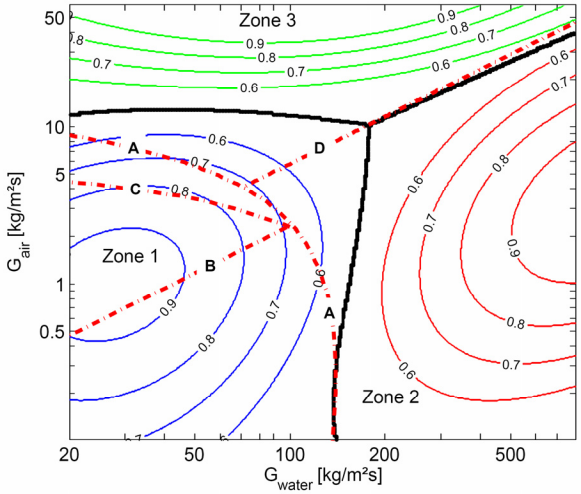

Figure 5 Probabilistic flow regime map: cluster boundaries (solid), probabilities P (contour) and Taitel-Dukler boundaries (dash-dot)

In Figure 5, the probability surfaces and cluster boundaries are shown in a contour plot together with the Taitel-Dukler boundaries. The centres of the flow regimes are clearly visible. The iso-probability lines quantify the width of the transition zones. A more gradual transition exists between zone 1 and zone 2 compared to the transition between zone 1 and zone 3. At the crossing of the three intersection lines a wide area exist where the two-phase flows are very chaotic. These flows have characteristics of all other types of flow. Purely classifying the 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.

Influence of 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 Figure 6, the

result of the data processing technique is shown when individual input 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.

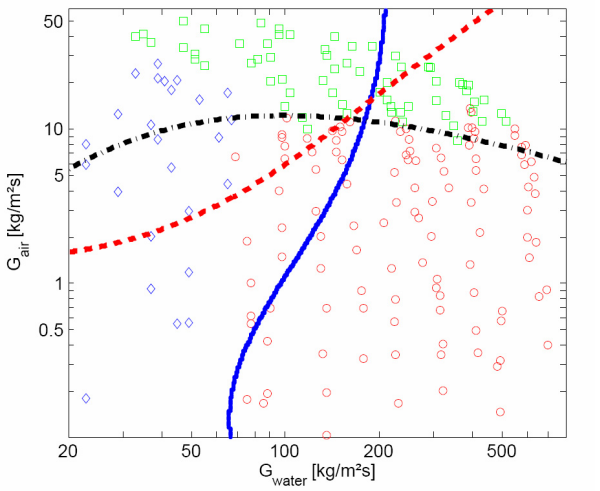

Figure 6 Influence of input features: visual scatter data and cluster boundaries of individual features
(◊ = stratified – o = intermittent – □ = annular)
(**dashed --** = AVG, **solid** = M2, **dash-dot** = HFCF)

Using only the M2 as input feature, a rather vertical boundary is found occurring at intermediate water mass flux (70-200 kg/m²s). From the multivariate analysis used for the signal feature selection, it became clear that the variance has potential to separate stratified flows from intermittent flows. The boundary found here does not fully coincide with the visual stratified-intermittent flow transition, but is located at higher mass fluxes. A single slug every few seconds causes a second but small peak in the PDF at high V*. Because the second peak is small, it only has a limited influence on M2. Therefore some slow slug flows are clustered together with the stratified flows. Applying the HFCF as input feature, a quasi-horizontal boundary is found. This line coincides with the visual transition from annular to intermittent flow. The HFCF can also separate 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.

Influence of weight parameters

To compare the influence of more than one feature when using the Euclidian distance function, the weights should be set equal to $w_k=1/(2\cdot var_k)$. 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 1).

Using these weight parameters, the influence of the AVG is enhanced. According to the input feature analysis, this should result in more diagonal boundaries. The result of using equally weighted input features is shown in Figure 7. The expected trend is confirmed. The similarity with the Taitel-Dukler boundaries however is not improved, on the contrary.

Table 1 Variances and weight parameters by feature

Feature	Variance	$w_k=1/(2\cdot var_k)$
AVG	0.1726	2.897
M2	0.3922	1.275
HFCF	0.2834	1.764

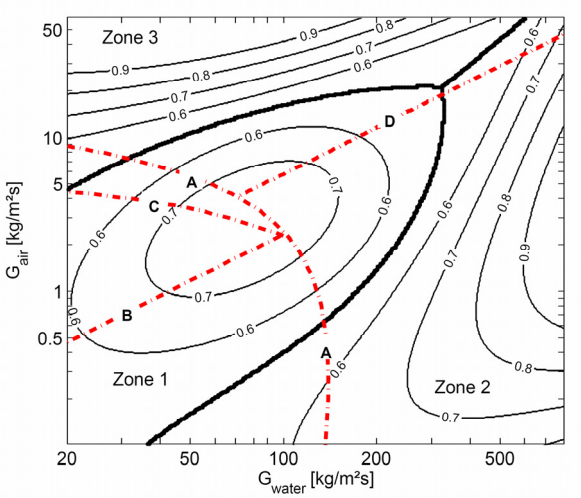

Figure 7 Probabilistic flow regime map with equally weighted input features cluster boundaries (solid), probabilities P (contour) and Taitel-Dukler boundaries (dash-dot)

Influence of distance function and number of clusters

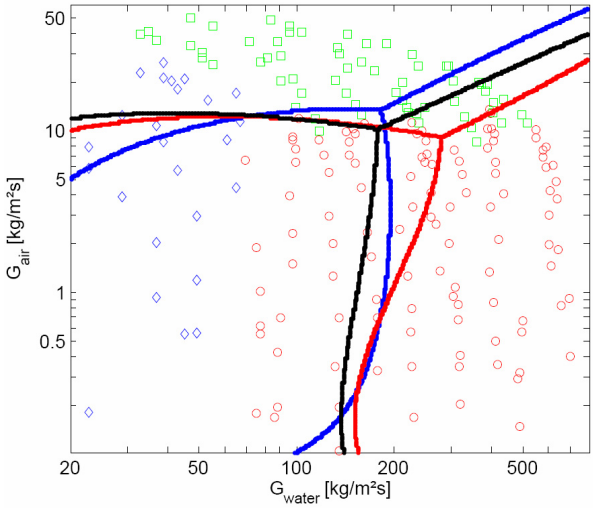

Figure 8 Flow Map with visual classification data
(◊ = stratified – o = intermittent – □ = annular)
and cluster boundaries of different distance functions
(**black** = Euclidian; **blue** = cityblock; **red** = cosine)

In the clustering theory, the choice of the degree of similarity between object is important. In Figure 8, a comparison is shown between the commonly used Euclidian distance function and with two other functions: the city block distance and the cosine distance. The input matrix was the normalized feature matrix without weight parameters. The boundaries found with the two other distance functions are comparable with the boundaries found with the Euclidian distance function. Therefore, only the Euclidian distance function will be further used. 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.

Applicability and generalization of the method in future work

The probability map technique can now be used for combining heat transfer or pressure drop correlations of different flow regimes in a probabilistic way. In this paper, the probabilistic method is applied to only three main flow regimes at adiabatic conditions of air-water. The probabilistic map can be expanded to other flow regimes that are not covered in the currently used database, such as bubbly flow, since the number of clusters (and so the number of flow regimes) can be chosen. For optimizing the results, the weight parameters of the algorithm can be fitted to the given data set.

Two-phase flow in evaporators and condensers occurs under diabatic conditions. In diabatic flow maps, additional flow regimes occur such as dryout and mist flow. Typical diabatic flow maps are those by Kattan et al. [11] and updates. More recently Cheng et al. [12] developed a diabatic map for CO_2 which intrinsically relates various flow regimes to the corresponding heat transfer mechanisms and therefore is also a kind of objective flow map. Crucial for successfully applying the method under other conditions, the sensor signals must contain typical and detectable changes in flow regime characteristics. Incorporating other flow regimes means that the sensor signals of these additional regimes must be separable from each of the already incorporated flow regimes. If the sensor fails to do so, the algorithm will not be able to cluster the data correctly. The typical flow regime characteristics should be reflected in the selected signal features that are provided as input to the k-means clustering algorithm. Provided this condition is achieved, the method itself is not limited to any working condition or application specific flow regime. So the method can be applied and generalized for diabatic conditions, different fluids and different tube diameters (even microchannels).

In future work, this method will be applied to diabatic refrigerant flows to add probabilistic data to the current diabatic two-phase flow maps and combine flow regime specific correlations in a probabilistic way.

CONCLUSIONS

In this study the signals of a capacitance sensor were analysed and processed to build a horizontal two-phase flow classifier without any subjective visual decisions. The k-means clustering algorithm was applied together with the Support Vector Regression technique to create flow regime transition boundaries. As input features, the AVG, M2 and HFCF were used and a three category classification was chosen. A remarkable agreement with visual observations and Taitel-Dukler boundaries was found. A probabilistic flow map was constructed to be able to quantify the transition areas. This makes it possible to combine flow regime dependent correlations in a probabilistic way in the two-phase flow models for heat transfer and pressure drop.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Taitel, Y. and Dukler, A.E., Model for Predicting Flow Regime Transitions in Horizontal and near Horizontal Gas-Liquid Flow, *Aiche Journal*, Vol. 22, 1976, pp. 47-55

[2] Barnea, D., A Unified Model for Predicting Flow-Pattern Transitions for the Whole Range of Pipe Inclinations, *International Journal of Multiphase Flow*, Vol. 13, 1987, pp. 1-12

[3] Nino, V.G., Hrnjak, P.S. and Newell, T.A., Two-phase flow visualization of R134A in a multiport microchannel tube, *Heat Transfer Engineering*, Vol. 24, 2003, pp. 41-52

[4] Jassim, E.W. and Newell, T.A., Prediction of two-phase pressure drop and void fraction in microchannels using probabilistic flow regime mapping, *International Journal of Heat and Mass Transfer*, Vol. 49, 2006, pp. 2446-2457

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

[6] Canière, H., T'Joen, C., Willockx, A., De Paepe, M., Christians, M., van Rooyen, E., Liebenberg, L. and Meyer, J.P., Horizontal two-phase flow characterization for small diameter tubes with a capacitance sensor, *Measurement Science & Technology*, Vol. 18, 2007, pp. 2898-2906

[7] Hastie, T., Tibshirani, R. and Friedman, J., The elements of statistical learning: Data mining, inference and prediction. 2001, Springer, New York, USA

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

[9] Chang, C.C. and Lin, C.J., LIBSVM: a library for support vector machines. 2001, Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[10] Barnea, D., Luninski, Y. and Taitel, Y., Flow Pattern in Horizontal and Vertical 2 Phase Flow in Small Diameter Pipes, *Canadian Journal of Chemical Engineering*, Vol. 61, 1983, pp. 617-620

[11] Kattan, N., Thome, J.R. and Favrat, D., Flow boiling in horizontal tubes: Part 1 – Development of a Diabatic Two-Phase Flow Pattern Map, *ASME – Journal of Heat Transfer*, Vol. 120, 1998, pp.140-165

[12] Cheng , L.X., Ribatski, G., Wojtan, L. and Thome, J.R., New flow boiling heat transfer model and flow pattern map for carbon dioxide evaporating inside horizontal tubes, *International Journal of Heat and Mass Transfer*, Vol. 49, 2006, pp 4082-4094