Predicting phase inversion in agitated dispersions with machine learning algorithms

J. M. Maffi[1,2], D. A. Estenoz[3,4]

[1]Instituto Tecnológico de Buenos Aires (ITBA), Departamento de Ingeniería Química, Av. Madero 399, C.P. C1106ACD, Buenos Aires, Argentina.

[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

[3]Instituto de Desarrollo Tecnológico para la Industria Química, INTEC (Universidad Nacional del Litoral - CONICET), Güemes 3450, C.P. 3000, Santa Fe, Argentina.

[4]Facultad de Ingeniería Química, FIQ (Universidad Nacional del Litoral - CONICET), Santiago del Estero 2829, C.P. 3000, Santa Fe, Argentina.

Corresponding author e-mail: destenoz@santafe-conicet.gov.ar,

Abstract

In agitated systems, the phase inversion (PI) phenomenon – the mechanism by which a dispersed phase becomes the continuous one – has been studied extensively in an empirical manner and few models have been put forward through the years. The underlying physics are still to be fully understood. In this work, the experimental evidence published in literature is used to train machine learning models that may infer the inherent rules that lead to a given dispersion type (O/W or W/O), as well as predict the value of the dispersed phase volume fraction at the edge of the inversion point. Decision trees, bagged decision trees, support-vector machines and multiple perceptrons are implemented and compared. Results show that it is possible to infer an ensemble of physical rules that explain why a given dispersion is O/W or W/O, where a strong "turbulence constraint" is identified. The intuitive rule that PI occurs at 50% dispersed

phase almost never holds. Moreover, neural networks have shown a better performance at predicting the PI point than the other algorithms tested. Finally, a theoretical study is performed in an effort to produce a phase inversion map with the relevant operating variables. This study showed a strong non-linear effect of the impeller-to-vessel size ratio, and an asymmetrical behavior of the interfacial tension on the phase inversion points.

Keywords: phase behavior, machine learning, neural network, liquid-liquid dispersions

Nomenclature

Greek letters		
γ	Interfacial tension	N/m
μ	Viscosity	Pa·s
ρ	Density	kg/m^3
σ	Surface tension	N/m
ϕ	Phase volume fraction	--
Latin letters		
N	Stirring speed	min^{-1}
D/T	Impeller-to-vessel diameter ratio	--
D	Impeller diameter	m
Subscripts		
c	Continuous phase	
d	Dispersed phase	
o	Oil phase	
w	Aqueous phase	

1. Introduction

In any stirred system containing a dispersed phase (dispersion or emulsion) the set of operating conditions may be such that, at a given phase composition, the dispersed phase becomes the continuous one and vice versa, a phenomenon called *phase inversion* (PI) (Norato et al. 1998). This particular moment can either be desired or prevented, depending on the application. For example, in the transport of heavy oil, water may be added in a ratio such that it may become the continuous phase as it flows through the

pipeline, so that the dispersion viscosity may drop substantially, improving the transport energy costs (Arirachakaran et al. 1989).

The dispersed phase volume fraction at which PI occurs is usually dubbed *the PI point*, and the effect of the operating variables on its value has been studied empirically for several decades (recent reviews by Kumar et al. (2015) or Perazzo et al. (2015) summarize them). Yet, the comprehensive understanding of its dynamics is still a challenge. When the composition of a given dispersion is changed by adding dispersed phase at a constant stirring speed, the coalescence rate of the discrete droplets initially increases; this may be compensated by an increase in particle break-up rate – due to a variety of possible mechanisms (Liao and Lucas 2009) – and a new state is reached where the balance of break-up and coalescence velocities results in a new particle size distribution (Castellano et al. 2018). If, on the contrary, the increase in coalescence frequency cannot be compensated by particle rupture, then – after a transition period – phase inversion will occur. This imbalance between coalescence and break-up is a widely adopted mechanism to explain PI in agitated systems, though is not the only one (Tidhar et al. 1986; Yeh et al. 1964; Yeo et al. 2002).

For a given agitation rate, in traditional O/W stirred systems there exists a range of phase compositions for which the dispersion structure is not uniquely determined (i.e., it may be either O/W or W/O depending on the methodology applied or the initial conditions), which is known as the *ambivalent range* (Ghotli et al. 2013). **Figure *1*** shows an example of a typical ambivalent range as a function of the organic phase volume fraction (ϕ_o) where three zones are clearly identified: zone A refers to organic-continuous dispersions (W/O), zone B to water-continuous (O/W) and zone C to the non-determined, ambivalent range.

[Figure 1 near here]

The upper bound of this range is the locus of organic phase volume fractions above which water can no longer be the continuous phase, while it is the opposite for the lower bound.

Few mathematical models have been developed to predict the PI point, the most interesting being those by Hu et al. (2005, 2006) or Brauner and Ullmann (2002) for traditional O/W systems (the former for stirred tanks and the latter for pipe flow), and perhaps a simple rheological model for polymer blends (Mekhilef and Verhoogt 1996). The drawback of these models is that they need to make simplifications to reduce the physical complexity of the multivariate, nonlinear phenomenon.

The main obstacle that arises when trying to understand the effect of each variable on the PI point is that the experimental conditions vary significantly from author to author, so that isolating the weight of each variable is quite complex. For this objective, empirical correlations using dimensionless numbers (like Reynolds, Weber or Froude) have the advantage of being simple to implement (with traditional regression techniques). However, in multivariate and strongly nonlinear phenomena, like the one described in this work, these techniques are not quite suitable given the many possible functionalities would have to be tested. For this reason, a different approach is essayed in this work.

Generalizing implicit behavior that is difficult to understand at first hand is one of the most interesting features of *machine learning* models (Lee et al. 2018; Raju and Cooney 1998). As well as empirical correlations, these models can also handle multivariate problems and, even if not as straightforward, their implementation is simple enough to make them interesting candidates for the PI point prediction. Their main advantage is that these algorithms use a subset of the empirical data to "reveal" their underlying structure and outperform the prediction capabilities of traditional regression techniques.

In this paper, the experimental results of phase inversion measurements in agitated dispersions present in literature have been summarized and fed to four different models – from heuristic to black-box – in an effort to contribute to the comprehensive understanding of the inversion phenomenon. To this end, two types of problems were treated: a **classification** and a **regression** one. In the first kind, the goal is to predict which phase would be continuous at a given composition and operating condition; in the second one, to predict the dispersed phase volume fraction at which PI would occur for a given operating condition.

The artificial intelligence methods that are suited for these objectives are *supervised learning* algorithms, which use a portion of the database to adjust its internal parameters (known as a "training stage") and gain generalization capabilities, which may be validated with the remainder of the data (a "test stage"). The algorithms implemented in this work are described in Section 3. Their main advantages include a fairly simple implementation and a reasonable performance in regression problems (Caruana and Niculescu-Mizil 2006). In classification problems, it is better to avoid a "class imbalance" (a feature of the database by which there are far more registries from a given class than from another), since these algorithms many not handle them efficiently and special modifications may be required (Nguyen, Bouzerdoum, and Phung 2008). Depending on the model, overfitting could be an issue to be aware of.

Other machine learning models, which are not trained to minimize a prediction error, are the *unsupervised learning* algorithms. Among others, these methods include hierarchical clustering, K-means, hidden Markov models, self-organizing (or Kohonen) maps, etc. These are mostly used for clustering purposes (in which there are no classes or values to be predicted but rather groups into which data is to be separated) and are therefore left out of this study, although self-organizing maps may be certainly used in

classification problems. A comparative study between supervised and unsupervised learning was conducted by Sathya and Abraham (2013).

2. Relevant operating variables

Determining the phase composition at which PI occurs is not straightforward, since it depends on an ensemble of physical properties (phase density, phase viscosity, interfacial tension) (Norato et al. 1998), on the geometry used (vessel size, agitation speed, impeller type and size, number of baffles, their materials) (Groeneweg et al. 1998) and even on the initial conditions and the procedure chosen (Bouchama et al. 2003). A brief discussion on the effect of these variables on the PI point is offered in what follows. A comprehensive review by the authors is available (Maffi, Meira, and Estenoz 2020).

2.1. Viscosity ratio

As a general rule, observed empirically by several authors (Selker and Sleicher 1965; Hu et al. 2005; Colmanetti et al. 2018), the tendency to remain dispersed increases with phase viscosity. The damping of particle coalescence is usually a given explanation for this behavior (Bazhlekov, Chesters, and Van De Vosse 2000). In literature, this ratio is usually defined as the ratio of the dispersed to the continuous phase viscosity. In this work, in order to avoid changing the definition for each predictive model, it will be defined as $r = {\mu_o}/{\mu_w}$. Several simplified correlations have been suggested to predict the PI curves as a function of this variable only, especially in polymeric systems (Yeh, Haynie, and Moses 1964; Jordhamo, Manson, and Sperling 1986; Kitayama, Keskkula, and Paul 2001).

The effect of viscosity is not only observed as the ratio between the properties of each phase. The absolute value of a phase viscosity is also important since it plays a fundamental role in the break-up and coalescence frequencies (Alopaeus et al. 2002).

2.2. Phase densities

Perhaps the least studied of variables that have an effect on the PI point. Some authors explain that its role is only noteworthy at low stirring speeds, when large phase density differences require stronger stirring speeds to sustain the dispersion (McClarey and Mansoori 1978; Norato, Tsouris, and Tavlarides 1998). Other workers suggest that a large difference should favor PI since it increases the local relative velocity of the droplets and thus increases the shear stress to which the system is subject to, which in turns favors breakage and increases the surface area available to coalescence (Rodger, Trice, and Rushton 1956; S. Kumar, Kumar, and Gandhi 1991).

2.3. Interfacial tension

Few studies aiming to isolate the effect of interfacial tension on PI are available. Some authors argue that a system with lower interfacial tension should be less likely to invert (Clarke and Sawistowski 1978; Kumar, Kumar, and Gandhi 1991), thus widening the span of the ambivalent region. However, the results found by Reeve and Godfrey (2002) challenge that idea: they indicate that the system with lower interfacial tension finds it easier to invert from O/W to W/O but harder in the opposite direction. The theoretical model derived by Hu et al. (2005) based on a population balance in a two-region vessel, agrees with those results.

A minimization of free surface energy could be expected at the PI point, since this would reflect the natural need of the system to invert. However, by measuring interfacial area, some investigators found that a minimization of the interfacial energy

happened only when inverting from W/O to O/W but not in the opposite case (Clarke and Sawistowski 1978; Luhning and Sawistowski 1971). Consequently, they postulate that phase energy minimization is not a criterion for phase inversion. In turn, Norato, Tsouris, and Tavlarides (1998) propose that a decrease in interfacial tension would promote drop breakage and increase film drainage times, which would not favor phase inversion.

It is possible that decreasing interfacial tension may either delay or promote PI, depending on the system. Increasing thermodynamic compatibility would favor the transition to the inverted system; yet, the increase in particle breakage rate may stabilize the dispersion and hinder the inversion.

2.4. Stirring speed

A general trend that is satisfied by all dispersions is the asymptotic value of phase volume fraction at the inversion point with increasing stirring speed. This has been observed consistently in several experiments (Kumar, Kumar, and Gandhi 1991; Reeve and Godfrey 2002; Deshpande and Kumar 2003; Quinn and Sigloh 1963) in both batch and flow vessels. This asymptotic behavior means that, in very turbulent conditions, there exists a controlling mechanism for phase inversion.

However, increasing stirring speed may have different effects depending on the system, as observed empirically by different authors (Arashmid and Jeffreys 1980; Deshpande and Kumar 2003).

2.5. Vessel geometry and material

Impeller design, its position and inclination, the number and size of the baffles, the vessel diameter and its material have all been reported to present an effect on the phase inversion point (Kato, Nakayama, and Kawasaki 1991; Guilinger, Grislingas, and Erga

1988). A thorough research on these variables was conducted by Deshpande and Kumar (2003). The problem when analyzing these variables is that their effect is not clear. A wide range of effects are reported, sometimes with opposing results.

2.6. Selected variables

Table A. 1, shown in the Supplementary Material, gathers the experimental results of both the upper and lower bounds of the ambivalent ranges found in literature for stirred O/W systems. After analyzing the physical and geometrical conditions used in each investigation, the following set of variables was chosen to be used in this work:

[Table 1 near here]

The choice of viscosity ratio may seem unnecessary since it may be built from each phase viscosity. However, in the implementation of the regression algorithms, if the ratio is deemed an important variable, then it must be fed as such. This viscosity ratio has been extensively studied and considered an important variable on PI (Selker and Sleicher 1965), but the viscosities of each phase (separately) play a major role on the break-up and coalescence rates (Liao and Lucas 2009, 2010); then, the ratio alone does not suffice. Regarding interfacial tension, the values shown in **Table A. 1** correspond to either the ones informed in the authors' work or – in their absence – to the ones predicted by the Girifalco and Good (1957) equation:

$$\gamma_{12} = \sigma_1 + \sigma_2 - 2\frac{4\sqrt[3]{V_1 V_2}}{\sqrt{\sqrt[3]{V_1} + \sqrt[3]{V_2}}}\sqrt{\sigma_1 \sigma_2} \quad (1)$$

where V_1 and V_2 are the molar volumes of phase 1 and 2 respectively. This equation has the advantage of being simple enough to be implemented with the available information and has been found to provide good results in the simple O/W systems of this study (Becher 1987).

The effect of temperature is indirectly observed through its effect on the different physical properties. If correlations between temperature and phase densities, phase viscosities and interfacial tensions were available, this variable could indeed be used in the models.

D/T is chosen as one of the most important geometrical variables because its effect on the energy dissipation rate has long been reported (Ghotli et al. 2013), and because other parameters are usually not informed by the authors. Moreover, gathered data refer only to stainless-steel impellers, ruling out impeller material as a variable.

3. Algorithms

Four techniques were selected to predict the experimental data presented in **Table A. 1**, namely: a) decision tree, b) bagged decision trees, c) support-vector machine and d) feed-forward neural network. A brief overview is of each algorithm is presented in what follows.

3.1. Decision trees

A decision tree is a classic heuristic model that may be used for both regression and classification problems. It is a simple model that may yield good performances with little training effort while providing the set of *if-then* rules inferred from the dataset. The complexity of the tree may be adjusted (usually through a "complexity parameter"): long trees with a large number of branches may improve the classification or regression accuracy, but almost certainly result in a loss of generalization capability, since each branch turns to a "particular case" rather than a "general rule" (overfitting). There exists an optimum tree size that provides a good balance between accuracy and generalization.

The main drawback of a decision tree is its high sensitivity to the training set (high variance) (Olaru and Wehenkel 2003). To deal with this issue, *bootstrap aggregating* (bagging) trees were developed. Essentially, this algorithm consists of a number of decision trees created in a way that each one has its own train/test subset of the formal training set (Breiman 1996). Then, each tree is trained independently of the rest and, when presented the formal test set, the outcome is the average of the prediction of each tree (for regression problems) or the majority vote (for classification goals). This compensates the variance problem but complicates the heuristic nature of the decision tree, by which interesting useful rules may be inferred. However, some *rule-extraction* algorithms exist (Deng 2019; Iqbal 2012) that may explain the general rules found by the ensemble of trees.

3.2. Support-vector machines

The support-vector machine (SVM) is a supervised learning algorithm that may be implemented for both regression and classification goals. In the first case, it is designed to find the smoothest function whose images lie within a user-defined margin (ε) around the data points:

$$f(x) = \boldsymbol{w}^T \langle \boldsymbol{\psi}_{(x_i)}, \boldsymbol{\psi}_{(x)} \rangle + b \qquad (2)$$

where $\boldsymbol{w}^T$ is a vector of weights, $\boldsymbol{\psi}_{(x)}$ is the non-linear mapping of the input space to a higher dimensional space ($\boldsymbol{\psi}_{(x_i)}$ refers to the map applied to the elements in the **training** set), b is the bias, and $\langle \cdot, \cdot \rangle$ stands for the dot product.

To this end, a certain number of training samples that lie outside the desired image region are used as pivot points (support vectors) to determine the function that may accommodate the majority of the predicted values inside the ε margin region. This

results in a constrained optimization problem usually expressed as in equations *(3*) and *(4*) and solved using Lagrange multipliers.

$$\max\begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha^*{}_i)\left(\alpha_j-\alpha^*{}_j\right)\langle\boldsymbol{\psi}_{(\boldsymbol{x_i})},\boldsymbol{\psi}_{(\boldsymbol{x_j})}\rangle \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha^*{}_i)+\sum_{i=1}^{n}y_i(\alpha_i+\alpha^*{}_i) \end{cases} \tag{3}$$

Subject to

$$\sum_{i=1}^{n}(\alpha_i-\alpha^*{}_i)=0 \tag{4}$$

where $\alpha_i, \alpha^*{}_i$ are the Lagrange multipliers and y_i is the i-th element of the target vector. $\alpha_i, \alpha^*{}_i \in [0, C]$, with C being a regularization parameter that determines the trade-off between the flatness of the function $f(x)$ and error deviation (which is why this model is usually referred to as ε-insensitive SVR).

The non-linear map $\boldsymbol{\psi}_{(\boldsymbol{x})}$ need not be known, since only the dot product $\langle\boldsymbol{\psi}_{(\boldsymbol{x_i})},\boldsymbol{\psi}_{(\boldsymbol{x})}\rangle$ is actually computed. Then, kernel functions $k_{(\boldsymbol{x},\boldsymbol{x_i})}=\langle\boldsymbol{\psi}_{(\boldsymbol{x_i})},\boldsymbol{\psi}_{(\boldsymbol{x})}\rangle$ are used instead, the most frequent being quadratic, cubic or Gaussian.

In classification problems, the SVM is designed to find a hyperplane that divides the dataset in two groups, each containing "only" the set of points belonging to a given class. The distance between the plane and the nearest sample points of each set is a margin that may be adjusted, and those points lying on the margin are the support-vectors. The algorithm is also implemented as a constrained optimization problem, similar to the one described previously for the regression case, where a classification score is defined to represent the distance between an input and the hyperplane defined.

3.3. Neural networks

Several types of neural networks have been developed over the years (Prieto et al. 2016; Schmidhuber 2015). In this work, the neural network model chosen is the standard feed-forward multilayer perceptron, with one or more hidden layers. The idea behind this algorithm is to either classify or predict function values by using a matrix of weights (which are tuned to minimize prediction errors) and an *activation function*, which is any form of a sigmoid function that may mimic an "on-off" response in a continuous way. This enables the model to behave like the brain neurons, in the sense that each input variable (stimulus) produces a different effect (response) on each neuron, ultimately resulting in a combined response that produces the output.

Figure 2 shows a scheme of a traditional perceptron with three layers: the input layer, receiving the input variables, a hidden layer, where the effect of each input is combined and treated with the chosen activation function, and the output layer, which combines the effect of every neuron. In order to shift the result from the (0,0) coordinate, bias neurons are also introduced, usually before and after the hidden layer.

[Figure 2 near here]

The neural weights may initialize at random and then tuned with different gradient-descent algorithms, trying to minimize the prediction errors.

4. Implementation and results

In the implementation of the previously discussed algorithms, the dataset was divided randomly as follows: 70% for training, 10% for validation and 20% for testing. These subsets were kept the same throughout the procedure, so that comparing the results between each algorithm is done with the same testing set. This is equivalent to feed the models with new (unseen) experimental data. The decision trees were coded in R while

the SVM and neural networks were programed in Matlab R2019a. A simplified, step-by-step pseudocode of each algorithm is provided in Supplementary Material.

4.1. Classification

In order to build the classification problem, the table of experimental data found in the Supplementary Material had to be filtered, since only a fraction of the entries share exactly the *same* operating conditions for both O/W→W/O and W/O→O/W inversions (for example, some authors only studied the lower bound of the ambivalent range, or used different stirring speeds between the upper and the lower curves). For those data points where both inversions have been studied at the same conditions, a clear identification of three classes is possible for any phase composition, *i.e.*: "oil continuous", "water continuous" and "not determined". Then, for this subset, which only contains the upper and lower bounds of the ambivalent range, any number of extra points may be added artificially (above, below, and between each curve), to create a new database that includes the phase behavior for each case, thus yielding the classification problem. Upon inspection, this filtered subset shared the same impeller type, number of baffles and equipment material; therefore, they were left out as input variables.

With this subset, 30 extra entries were added (10 for each class), by creating random artificial points above, inside and below the ambivalent range for each operating condition. Results will be discussed and compared in terms of the confusion matrices

4.1.1. Decision tree

A decision tree was trained to predict the phase behavior at a given condition, with a default complexity parameter (CP) of 0.01 using the rpart library in R (Therneau and Atkinson 2018). An overall accuracy of 81.% was obtained with the testing subset, with

the confusion matrix shown in ***Table 3***. A schematic figure of the tree, with its if-then rules is depicted in ***Figure 3***. The classes were named O (for oil-continuous), W (for water continuous) an ND (for non-determined). The percentages shown in each box represent the fraction of data points from the test set that ended in each branch. Interestingly, with this complexity parameter, only three variables were used in the construction of the tree: ϕ_o, μ_o and D/T. This is so because the algorithm that creates the decision tree calculates the weight and entropy of each variable and builds a tree following the *if-then* rules that arise from the dataset. In this case, the algorithm found that ϕ_o, μ_o and D/T were the only important variables for the desired complexity.

[Figure 3 near here]

Upon optimization of the complexity parameter (which is a measure of how many branches, leaves and levels a tree may develop), a trade-off between accuracy and complexity (and strong overfitting) must be obtained. The goal is to arrive to a decision tree that may explain the problem as accurately as possible but with simple rules inferred by the most relevant variables. After analyzing trees with different complexity levels, the one shown in ***Figure 4*** was found to be the optimum case, with an overall accuracy of 85.7%. Previously disregarded variables are here used in the construction of the tree: ρ_o and N. This is expected after having increased the complexity of the tree. An improvement in both sensitivity and sensibility of the class "water continuous (W)" is observed.

[Figure 4 near here]

The interesting feature of this second tree is that one may infer the following rules to determine whether the aqueous or the organic phase would be continuous. The intuitive rule (the phase whose volume fraction is above 0.5) **almost never** holds directly.

1. If the organic phase volume fraction is above 70%, it is very likely (87% of cases studied) to be the continuous phase regardless of the operating conditions.
2. The organic phase shall be continuous following the intuitive rule (ϕ_o>0.5) only if the size of the impeller is at most 40% the size of the vessel. This may imply that a rather inefficient mixing (small impeller) is related to the inability of the dispersed phase to become continuous, perhaps due to lack of turbulence which produces few occluded droplets and deters phase inversion at low holdups.
3. In contrast, if the organic phase is the one in higher proportion but larger impellers are used (D/T>0.4), then it is **almost never** the continuous phase. The only case possible for it to be continuous is to have a very large density difference with the aqueous phase (notice the ρ_o>1398 kg/m^3 condition). This again may reflect the same effect described in the previous rule: the increase in turbulence provided by a bigger impeller – which would promote breakage, coalescence and occlusion formation – may be countered by a very high momentum of the continuous phase, thus disabling the aqueous phase to become continuous. In all other cases, either the aqueous phase is continuous, or the result is undetermined. This effect caused by a large density difference was not thoroughly discussed by the original authors (Arashmid and Jeffreys 1980; Quinn and Sigloh 1963).
4. If the aqueous phase volume fraction is above 76% then it will be the continuous phase almost without question. There is only one exception that yielded a "not determined" case (notice the 0.4>$\phi_o \geq$ 0.045 branch) which, upon verification of the dataset, refers to a very particular mixture of kerosene/water with a very low interfacial tension (≈10 mN/m) by Arashmid and Jeffreys (1980) and a small impeller (D/T<0.4). This means that the effect of poor mixing was countered with a very low interfacial tension, which enhanced particle break-up and coalescence to a

point that the only way to *ensure* an aqueous continuous phase is to have an extremely high fraction of it. A similar system was also studied by Quinn and Sigloh (1963) and, in their analysis, the authors arrived to am analogous conclusion, as they observed O/W/O structures in a dispersion that should have been water-continuous but was not. Leaving this exception aside, then all dispersions with at least 76% aqueous phase will be O/W.

5. Finally, at an intermediate range of aqueous phase composition ($0.76 > \phi_w \geq 0.6$) the organic phase will almost surely be dispersed. If its viscosity is high enough, the probability of getting a "not determined" result somewhat increases, but an oil-continuous phase is never assured. More so, at high stirring speeds the aqueous phase is very likely to be the continuous one, since the attenuation of coalescence of viscous oil drops will deter the formation of W/O dispersions.

4.1.2. Bootstrap aggregation (bagging)

The former inferred set of rules may not be the only one, since different decision trees may be obtained varying the train/test subsets. For this reason, a bagging technique was implemented with 300 trees (upon verification of the error evolution, this number was found to be optimum). All variables were included in the analysis and a rule-extraction technique following Deng (2019) was applied. ***Table 2*** summarizes some of the rules found by this algorithm across the ensemble of trees. A total of 37 rules were extracted, but only the ones with no prediction error and simple enough not to be considered "special cases" are shown here.

[Table 2 near here]

In general, the same pattern is observed in this set of rules compared to the ones obtained with one tree (some are even repeated, which is a good sign for the tree

presented in the previous section): being in a greater proportion is not a sufficient condition to be the continuous phase. For instance, the first rule cited in the previous section has been extended to 100% accuracy by adding two constraints: low interfacial tension and high stirring speed (2nd condition in ***Table 2***). As a general rule, with a volume fraction between 50% and 75%, there is a "turbulence constraint" needed to ensure that a given phase will be surely continuous. Otherwise, it would depend on how the dispersion was prepared and thus end up inside the ambivalent range.

A clear advantage of the ensemble model is that it outperforms one single decision tree in its classification goal (overall accuracy of 93.5%), as observed from the confusion matrix shown in ***Table 3***.

[Table 3 near here]

4.1.3. Support-vector machine and neural network

The last two techniques considered in this work are not heuristic. Then, the analysis here offered is only related to their performance in their classification goal. For the SVM models, three kernels (apart from the simple linear one) were tested, since the problem is suspected to be strongly nonlinear and thus a transformation is required to generate a hyperplane such that the three classes become separable. The functions in question were quadratic, cubic and a radial basis (Gaussian). Their confusion matrices are reproduced in ***Table 4***.

[Table 4 near here]

The best performance is offered by the cubic kernel, with an overall classification accuracy of 92.18%, which is still slightly below the performance of the bagging model.

Regarding the artificial neural network, four cases were compared changing the number of neurons of the hidden layer. The general rule of thumb states that this number should be between the number of predictors and the size of the output vector; *i.e.*, in this case between 3 and 9. Thus, networks with 3 to 10 neurons were tested to assess the quality of this rule. A scaled conjugate gradient was used as a training algorithm, which was performed 20 times to account for the randomness in the initialization of weights.

Confusion matrices are shown in ***Table 5***: the net with 7 neurons performed slightly better than the rest, but almost no difference is observed with the 5 or 10-neuron net.

[Table 5 near here]

Comparing the classification performances of all four implemented techniques, it would seem that the bagging model resulted in the most accurate one, followed closely by the cubic SVM, with the extra advantage of being able to put forward a set of physical rules behind the data. This is an interesting result since, in the past few years, this technique has left ground for the more sophisticated models, like artificial neural nets.

4.2. Regression

All four techniques compared for the classification case can be modified to behave as regression models, in this case aiming to predict the dispersed phase volume fraction at which phase inversion would occur. Since, as explained in Section 4.1, some of the entries of **Table A. 1** correspond only to one of the two possible inversions, two different models had to be developed for each technique: one to predict the organic phase volume fraction at which the dispersion becomes oil-continuous (ϕ_o), and another one for the opposite inversion case (ϕ_w). The following paragraphs describe the results obtained with each model. In all cases, the performance is measured with the root mean squared error (RMSE).

4.2.1. Decision (and bagged) trees

The number of trees in the bagging models was set to 300, as in the classification problem. Results are show in **Figure *5*** and **Table *6***, where a comparison with the predictions by only one decision tree is offered. As observed, a slightly better performance is obtained with the ensemble model for the O/W → W/O inversion but no improvement was gained for the opposite case.

[Table 6 near here]
[Figure 5 near here]

4.2.2. Support vector regression

The ε-insensitive SVR model was implemented setting the cost factors (C) and the margin of errors (ε) as shown in ***Table 7***. A 10-fold cross-validation was performed on the training set before running the model, to avoid overfitting. Kernel functions and their scale parameters used are shown in ***Table 8***, as well as their performance measured by the RMSE. Predicted phase volume fractions are shown in ***Figure 6***. The quadratic kernel showed very good fit with both prediction sets (ϕ_o and ϕ_w), outperforming even the classic radial basis function. The fact that the kernel that gave better results is different from the classification problem is not surprising since the implementation algorithm and the dataset are different (as explained in Sections 3.2 and 4.1 respectively).

[Table 7 near here]
[Table 8 near here]
[Figure 6 near here]

Like in the classification problem, the bagging model slightly outperforms the SVM technique for both inversion curves.

4.2.3. Neural networks

A multiple perceptron with the traditional Levenberg-Marquardt backpropagation algorithm (Lv et al. 2018) was implemented, using one or two hidden layers. In all cases, the activation function chosen was $f(x) = \frac{2}{1+e^{-2x}} - 1$. The effect on the number of neurons on each layer was analyzed to find an optimum network size. Due to the random initialization of synaptic weights, each net was trained and assessed 20 times and the best network (based on its performance with the test subset) was chosen for each case.

The performance of each net is summarized in ***Table 9*** and ***Figure 7***, where an improvement is observed compared to the SVR and bagging models. Prediction of the PI point does not vary substantially with the number of neurons. Standard deviation of errors lie around 0.009 for the O/W→W/O inversion point and around 0.004 for the opposite case.

[Table 9 near here]
[Figure 7 near here]

A second hidden layer was added to test for any performance improvement, which is usual among nonlinear problems. However, no significant change in the RMSE was observed compared to the one-layer case.

[Table 10 near here]

Therefore, adopting a network with one hidden layer with 20 neurons for the prediction of each PI point seems reasonable.

5. Theoretical predictions of the ambivalent range

Considering the good results offered by the neural net with 15 neurons in the hidden layer, this model was used to analyze the results of several hypothetical scenarios. The goal is to build a phase inversion map with the most relevant operating and geometrical variables. A base case was chosen as a representative sample of the dataset, fixing: ρ_o = 800 kg/m^3, ρ_w = 1000 kg/m^3, μ_o = 0.8 cP, μ_w= 1 cP (r=0.8), γ= 30 mN/m, D/T = 0.5, Rushton turbine, 4 baffles, stainless-steel vessel.

Figure 8 shows the evolution of the ambivalent range when subject to variations in phase viscosity ratio (r), interfacial tension, phase density difference (ρ_o-ρ_w) and D/T, as predicted by the model. Some interesting remarks arise from these theoretical predictions by the net:

1. Increasing the phase viscosity ratio – μ_o/μ_w as previously defined – widens significantly the ambivalent range, especially in the lower section of stirring speeds, which is in line with the empirical rule find by several workers (Colmanetti et al. 2018; Hu et al. 2005; Selker and Sleicher 1965) that states that "the tendency to remain dispersed increases with viscosity". Moreover, the curvature of both bounds changes significantly: for higher viscosity ratios, it appears to be easier for the W/O dispersions to invert to O/W (lower bound) when increasing the agitation rate, which is the opposite tendency at low viscosity ratios. But this is not surprising since the ambivalent ranges are plotted as a function of the organic phase volume fraction, ϕ_o, and not a generic "dispersed phase". Then, if the dispersed phase is more viscous than the continuous phase, it will need higher agitation rates to be able to invert.
2. At very low interfacial tensions, the span of the ambivalent range is quite wide, especially at high stirring speeds. This is possibly explained by the fact that low

interfacial tensions yield almost plane curvatures and so favor the formation of co-continuous systems (Chuesiang et al. 2018) (with continuous phase entrapment), so that the dynamics of drop break-up and coalescence may favor either type of dispersion. This effect is particularly enhanced at high energy dissipation rates, where the occlusion formation is frequent (thus increasing the effective phase volume fraction), especially if the dispersed phase is the aqueous one (Kumar 1996). At low rates, the W/O structure is apparently preferred.

3. Increasing the interfacial tension hampers the previous effect and thus O/W structures become more likely (oil drops are far less likely to entrap water droplets, as observed by Pacek et al. (1994)), especially if the energy input is not high enough to compensate this effect.
4. A surprising effect is observed regarding the phase density difference: it is not a symmetric property, in the sense that two systems owning the same absolute difference do not invert at the same composition. Phase viscosity difference could be responsible for this behavior, but simulated results indicate otherwise, as observed in **Figure *9***, which shows the simulated inversion curves of two dispersions (O/W and W/O) with the same density difference and viscosity ratio. A clear difference is observed between both curves, meaning that having the phase density and viscosity differences does not imply the same inversion point. Increasing ρ_o shortens the span of the ambivalent range, implying that both structures are favored. The fact that this does not happen when increasing ρ_w may be due to the difference in the dynamics of the break-up and coalescence processes (and perhaps to the ability to form multiple dispersions).
5. The impeller-to-vessel size ratio is strongly nonlinear and, as it appears, not monotone. This parameter has a formidable effect on the mixing efficiency, through

the energy dissipation rate. The break-up vs coalescence imbalance is greatly affected by the flow patterns throughout the vessel, thus producing different dispersion structures.

[Figure 8 near here]
[Figure 9 near here]

6. Concluding remarks and perspectives

The phase inversion ambivalent curves present in literature were simulated with different approaches, from heuristic to black-box machine learning models. A decision tree with a manageable complexity level was presented, allowing to decide whether the aqueous or the organic phase will be the continuous phase under a given condition (only valid for stainless-steel, four-baffled vessels equipped with a Rushton turbine). The *if-then* rules inferred by the model were analyzed thoroughly and a physical explanation for the PI phenomenon is offered. These rules showed that the intuitive explanation for PI (reaching a dispersed phase fraction above 50%) almost never holds. A strong "turbulence constraint" was identified as a requirement for inversion: if not efficiently mixed, phase inversion does not seem to occur. More sophisticated classification techniques, like the support vector classifier and the multiple perceptron, which provide no heuristic rules, did not offer a differential improvement in the classification goal.

Among the regression techniques, the neural network multiple perceptron with 10 to 20 neurons in its hidden layer showed the best results in terms of its fit to the experimental data. The number of hidden layers was found not to be an important parameter of the model. This powerful tool was used to study the theoretical phase inversion paths for different simulated systems. This study showed that the ambivalent range is strongly affected by the impeller-to-vessel size ratio, in non-linear non-monotonic ways. This is

certainly due to the different flow patterns that are developed when varying the impeller size, which exert different energy levels to the fluid and therefore change the break-up and coalescence imbalance. Furthermore, simulations showed how decreasing interfacial tension may strongly affect the W/O → O/W inversion path but not affect the opposite inversion as importantly, which may be related to the possibility of water drops entrapping oil phase. Varying the viscosity ratio and the phase density difference showed tendencies that are in line with what is observed in literature.

This model may continue to be perfected as new experimental findings are fed. It would be of interest to use it under particularly different conditions where phase inversion takes place, like polymer-polymer blends or polymer-solvent mixtures (for example, in the high-impact polystyrene manufacture process), where phase viscosities are out of range compared to traditional O/W dispersions. Furthermore, it could be extended to O/W/surfactant systems, but it would require to fully characterize the structure of the surface-active components, as both its concentration and its hydrophilic-lipophilic balance (HLB) – or surface affinity difference (SAD) – have a major role in the PI point (Sajjadi et al. 2003). If enough information is gathered and used to train the neural network model, then a much complete, comprehensive idea of the phase inversion phenomenon will be achieved, which may have a major impact on several Chemical Engineering fields.

7. Acknowledgements

The authors are grateful to Mrs. Mercedes Alais, Mrs. Catalina Swinnen and Mrs. Paloma Aqueveque Cuetos for their contributions to this work, to CONICET for the doctoral fellowship granted to Juan M. Maffi, and to ITBA for its support.

References

Afshar Ghotli, R., A. A. A. Raman, S. Ibrahim, and S. Baroutian. 2013. "Liquid-Liquid Mixing in Stirred Vessels: A Review." *Chemical Engineering Communications* 200 (5): 595–627. https://doi.org/10.1080/00986445.2012.717313.

Alopaeus, Ville, Jukka Koskinen, Kari I Keskinen, and Joakim Majander. 2002. "Simulation of the Population Balances for Liquid-Liquid Systems in a Nonideal Stirred Tank. Part 2-Parameter Fitting and the Use of the Multiblock Model for Dense Dispersions." *Chemical Engineering Science* 57 (10): 1815–25. https://doi.org/10.1016/S0009-2509(02)00067-2.

Arashmid, M, and V Jeffreys. 1980. "Analysis of the Phase Inversion Characteristics of Liquid-Liquid Dispersions." *AIChE Journal* 26 (1): 51–55.

Arirachakaran, S., K. D. Oglesby, M. S. Malinowsky, O. Shoham, and J. P. Brill. 1989. "An Analysis of Oil/Water Flow Phenomena in Horizontal Pipes." *SPE Production Operations Symposium* 18836: 155. https://doi.org/10.2118/18836-ms.

Bazhlekov, I. B., A. K. Chesters, and F. N. Van De Vosse. 2000. "The Effect of the Dispersed to Continuous-Phase Viscosity Ratio on Film Drainage between Interacting Drops." *International Journal of Multiphase Flow* 26 (3): 445–66. https://doi.org/10.1016/S0301-9322(99)00032-4.

Becher, Paul, ed. 1987. *Encyclopedia of Emulsion Technology Vol. 3*. 1st ed. New York: Marcel Dekker.

Bouchama, F., G. A. Van Aken, A. J.E. E Autin, and G. J.M. M Koper. 2003. "On the Mechanism of Catastrophic Phase Inversion in Emulsions." *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 231 (1–3): 11–17. https://doi.org/10.1016/j.colsurfa.2003.08.011.

Brauner, Neima, and Amos Ullmann. 2002. "Modeling of Phase Inversion Phenomenon in Two-Phase Pipe Flows." *International Journal of Multiphase Flow* 28 (7): 1177–1204. https://doi.org/10.1016/S0301-9322(02)00017-4.

Breiman, Leo. 1996. "Bagging Predictors." *Machine Learning* 24 (2): 123–40. https://doi.org/10.1007/bf00058655.

Caruana, R., and A. Niculescu-Mizil. 2006. "An Empirical Comparison of Supervised Learning Algorithms." In *Proceedings of the 23rd International Conference on Machine Learning*, 161–68. Pittsburgh. https://doi.org/10.1145/1143844.1143865.

Castellano, Simone, Nida Sheibat-Othman, Daniele Marchisio, Antonio Buffo, and Sophie Charton. 2018. "Description of Droplet Coalescence and Breakup in Emulsions through a Homogeneous Population Balance Model." *Chemical Engineering Journal* 354 (May): 1197–1207. https://doi.org/10.1016/j.cej.2018.07.176.

Chuesiang, Piyanan, Ubonrat Siripatrawan, Romanee Sanguandeekul, Lynne McLandsborough, and David Julian McClements. 2018. "Optimization of Cinnamon Oil Nanoemulsions Using Phase Inversion Temperature Method: Impact of Oil Phase Composition and Surfactant Concentration." *Journal of Colloid and Interface Science* 514: 208–16. https://doi.org/10.1016/j.jcis.2017.11.084.

Clarke, S, and H Sawistowski. 1978. “Phase Inversion of Stirred Liquid/Liquid Dispersions under Mass Transfer Conditions.” *Chemical Engineering Research Design* 56: 50–55.

Colmanetti, Alex Roger Almeida, Marcelo Souza de Castro, Marcel Cavallini Barbosa, and Oscar Mauricio Hernandez Rodriguez. 2018. “Phase Inversion Phenomena in Vertical Three-Phase Flow: Experimental Study on the Influence of Fluids Viscosity, Duct Geometry and Gas Flow Rate.” *Chemical Engineering Science* 189: 245–59. https://doi.org/10.1016/j.ces.2018.05.050.

Deng, Houtao. 2019. “Interpreting Tree Ensembles with InTrees.” *International Journal of Data Science and Analytics* 7 (4): 277–87. https://doi.org/10.1007/s41060-018-0144-8.

Deshpande, K. B., and S. Kumar. 2003. “A New Characteristic of Liquid-Liquid Systems - Inversion Holdup of Intensely Agitated Dispersions.” *Chemical Engineering Science* 58 (16): 3829–35. https://doi.org/10.1016/S0009-2509(03)00235-5.

Girifalco, L, and R J Good. 1957. “A Theory for Estimation of Surface and Interfacial Energies.” *Journal of Physical Chemistry* 61 (7): 904–9. https://doi.org/10.1021/j150553a013.

Groeneweg, F., W.G.M. Agterof, P. Jaeger, J.J.M. Janssen, J.A. Wieringa, and J.K. Klahn. 1998. “On the Mechanism of the Inversion of Emulsions.” *Chemical Engineering Research and Design* 76 (1): 55–63. https://doi.org/10.1205/026387698524596.

Guilinger, T.R., A.K. Grislingas, and O. Erga. 1988. “Phase Inversion Behavior of Water-Kerosene Dispersions.” *Industrial & Engineering Chemistry Research* 27: 978–82. https://doi.org/10.1021/ie00078a015.

Heidari, Ali Asghar, Hossam Faris, Ibrahim Aljarah, and Seyedali Mirjalili. 2018. “An Efficient Hybrid Multilayer Perceptron Neural Network with Grasshopper Optimization.” *Soft Computing*. https://doi.org/10.1007/s00500-018-3424-2.

Hu, Bin, Panagiota Angeli, Omar K. Matar, and Geoffrey F. Hewitt. 2005. “Prediction of Phase Inversion in Agitated Vessels Using a Two-Region Model.” *Chemical Engineering Science* 60 (13): 3487–95. https://doi.org/10.1016/j.ces.2005.02.002.

Hu, Bin, Omar K. Matar, Geoffrey F. Hewitt, and Panagiota Angeli. 2006. “Population Balance Modelling of Phase Inversion in Liquid-Liquid Pipeline Flows.” *Chemical Engineering Science* 61 (15): 4994–97. https://doi.org/10.1016/j.ces.2006.03.053.

Iqbal, Md Ridwan Al. 2012. “Rule Extraction from Ensemble Methods Using Aggregated Decision Trees.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7664 LNCS (PART 2): 599–607. https://doi.org/10.1007/978-3-642-34481-7_73.

Jordhamo, G. M., J. a. Manson, and L. H. Sperling. 1986. “Phase Continuity and Inversion in Polymer Blends and Simultaneous Interpenetrating Networks.” *Polymer Engineering and Science* 26 (8): 517–24. https://doi.org/10.1002/pen.760260802.

Kato, Satoru, Eiichi Nakayama, and Junjiro Kawasaki. 1991. “Types of Dispersion in Agitated Liquid-liquid Systems.” *The Canadian Journal of Chemical Engineering* 69 (1): 222–27. https://doi.org/10.1002/cjce.5450690126.

Kitayama, N, H Keskkula, and Donald R. Paul. 2001. “Reactive Compatibilization of Nylon 6/Styrene–Acrylonitrile Copolymer Blends Part 3. Tensile Stress-Strain Behavior.” *Polymer* 42 (8): 3751–59. https://doi.org/10.1016/S0032-3861(00)00695-9.

Kumar, Ankit, Shigeng Li, Chieh Min Cheng, and Daeyeon Lee. 2015. “Recent Developments in Phase Inversion Emulsification.” *Industrial and Engineering Chemistry Research* 54 (34): 8375–96. https://doi.org/10.1021/acs.iecr.5b01122.

Kumar, S. 1996. “On Phase Inversion Characteristics of Stirred Dispersions.” *Chemical Engineering Science* 51 (5): 831–34. https://doi.org/10.1016/0009-2509(96)90025-1.

Kumar, S., R. Kumar, and K. S. Gandhi. 1991. “Influence of the Wetting Characteristics of the Impeller on Phase Inversion.” *Chemical Engineering Science* 46 (9): 2365–67. https://doi.org/10.1016/0009-2509(91)85136-L.

Lee, Jay H., Joohyun Shin, and Matthew J. Realff. 2018. “Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field.” *Computers and Chemical Engineering* 114: 111–21. https://doi.org/10.1016/j.compchemeng.2017.10.008.

Liao, Yixiang, and Dirk Lucas. 2009. “A Literature Review of Theoretical Models for Drop and Bubble Breakup in Turbulent Dispersions.” *Chemical Engineering Science* 64 (15): 3389–3406. https://doi.org/10.1016/j.ces.2009.04.026.

———. 2010. “A Literature Review on Mechanisms and Models for the Coalescence Process of Fluid Particles.” *Chemical Engineering Science* 65 (10): 2851–64. https://doi.org/10.1016/j.ces.2010.02.020.

Luhning, R. W., and H. Sawistowski. 1971. “Phase Inversion in Stirred Liquid-Liquid Systems.” In *Proceedings International Solvent Extraction Conference*, 873–87.

Lv, Chen, Yang Xing, Junzhi Zhang, Xiaoxiang Na, Yutong Li, Teng Liu, Dongpu Cao, and Fei Yue Wang. 2018. “Levenberg-Marquardt Backpropagation Training of Multilayer Neural Networks for State Estimation of a Safety-Critical Cyber-Physical System.” *IEEE Transactions on Industrial Informatics* 14 (8): 3436–46. https://doi.org/10.1109/TII.2017.2777460.

Maffi, J. M., G. R. Meira, and D. A. Estenoz. 2020 (in press). “Mechanisms and Conditions That Affect Phase Inversion Processes. A Review.” *The Canadian Journal of Chemical Engineering*.

McClarey, M. J., and G. A. Mansoori. 1978. “Factors Affecting the Phase Inversion of Dispersed Immiscible Liquid-Liquid Mixtures.” *AIChE Symposium Series* 74 (173): 134–39.

Mekhilef, Nafaa, and Henk Verhoogt. 1996. “Phase Inversion and Dual-Phase Continuity in Polymer Blends: Theoretical Predictions and Experimental Results.” *Polymer* 37 (18): 4069–77. https://doi.org/10.1016/0032-3861(96)00254-6.

Nguyen, Giang H., Abdesselam Bouzerdoum, and Son L. Phung. 2008. “A Supervised

Learning Approach for Imbalanced Data Sets." In *Proceedings of the 19th International Conference on Pattern Recognition*. Tampa. https://doi.org/10.1109/icpr.2008.4761278.

Norato, M., C. Tsouris, and L. Tavlarides. 1998. "Phase Inversion Studies in Liquid-Liquid Dispersions." *The Canadian Journal of Chemical Engineering* 76: 486–94.

Olaru, Cristina, and Louis Wehenkel. 2003. "A Complete Fuzzy Decision Tree Technique." *Fuzzy Sets and Systems* 138 (2): 221–54. https://doi.org/10.1016/S0165-0114(03)00089-7.

Pacek, A. W., A. W. Nienow, and I. P T Moore. 1994. "On the Structure of Turbulent Liquid-Liquid Dispersed Flows in an Agitated Vessel." *Chemical Engineering Science* 49 (20): 3485–98. https://doi.org/10.1016/S0009-2509(94)85027-5.

Perazzo, A., V. Preziosi, and S. Guido. 2015. "Phase Inversion Emulsification: Current Understanding and Applications." *Advances in Colloid and Interface Science* 222: 581–99. https://doi.org/10.1016/j.cis.2015.01.001.

Prieto, Alberto, Beatriz Prieto, Eva Martinez Ortigosa, Eduardo Ros, Francisco Pelayo, Julio Ortega, and Ignacio Rojas. 2016. "Neural Networks: An Overview of Early Research, Current Frameworks and New Challenges." *Neurocomputing* 214: 242–68. https://doi.org/10.1016/j.neucom.2016.06.014.

Quinn, J. A., and D. B. Sigloh. 1963. "Phase Inversion in the Mixing of Immiscible Liquids." *The Canadian Journal of Chemical Engineering*, 15–18.

Raju, Gokaraju K., and Charles L. Cooney. 1998. "Active Learning from Process Data." *AIChE Journal* 44 (10): 2199–2211. https://doi.org/10.1002/aic.690441009.

Reeve, R N, and J C Godfrey. 2002. "Phase Inversion During Liquid – Liquid Mixing in Continuous Flow, Pump – Mix, Agitated Tanks." *Chemical Engineering Research and Design* 80 (8): 864–71.

Rodger, W. A., V. G. Trice, and J. H. Rushton. 1956. "Effect of Fluid Motion on Interfacial Area of Dispersions." *Chemical Engineering Progress* 52.

Sajjadi, S., F. Jahanzad, M. Yianneskis, and B. W Brooks. 2003. "Phase Inversion in Abnormal O/W/O Emulsions. 2. Effect of Surfactant Hydrophilic−Lipophilic Balance." *Industrial & Engineering Chemistry Research* 42 (15): 3571–77. https://doi.org/10.1021/ie021044e.

Sathya, R., and A. Abraham. 2013. "Comparison of Supervised and Unsupervised Learning Algorithms for Patter Classification." *International Journal of Advanced Research in Artificial Intelligence* 2 (2): 34–38.

Schmidhuber, Jürgen. 2015. "Deep Learning in Neural Networks: An Overview." *Neural Networks* 61: 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Selker, A H, and C A Sleicher. 1965. "Factors Affecting Which Phase Will Disperse When Immiscible Liquids Are Stirred Together." *The Canadian Journal of Chemical Engineering* 43 (6): 298–301.

Therneau, T., and B. Atkinson. 2018. "Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-13." https://cran.r-project.org/package=rpart.

Tidhar, M., J. C. Merchuk, A. N. Sembira, and D. Wolf. 1986. "Characteristics of a

Motionless Mixer for Dispersion of Immiscible Fluids- II. Phase Inversion of Liquid-Liquid Systems." *Chemical Engineering Science* 41 (3): 457–62. https://doi.org/10.1016/0009-2509(86)87027-0.

Yeh, George C., Fred H Haynie, and R. E X A. Moses. 1964. "Phase-Volume Relationship at the Point of Phase Inversion in Liquid Dispersions." *AIChE Journal* 10 (2): 260–65.

Yeo, Leslie Y., Omar K. Matar, E S P de Ortiz, Geoffrey F. Hewitt, E. S. Perez De Ortiz, and Geoffrey F. Hewitt. 2002. "Simulation Studies of Phase Inversion in Agitated Vessels Using a Monte Carlo Technique." *Journal of Colloid and Interface Science* 248 (2): 443–54. https://doi.org/10.1006/jcis.8159.

Physical properties	Geometrical parameters
• phase densities (ρ_o and ρ_w) • phase viscosities (μ_o and μ_w) and their ratio ($r = {\mu_o}/{\mu_w}$) • interfacial tension (γ)	• number of baffles • impeller-to-vessel size ratio (D/T) • type of impeller • vessel material

Table 1. Variables chosen to build the models.

	Condition	Predicted continuous phase	Frequency (%)
1	ϕ_w>0.75 and D/T>0.4	W	17.5
2	ϕ_o>0.74, γ<50 mN/m and N>700 rpm	O	11.1
3	ϕ_w>0.6 and μ_o<1.17 cP	W	6.1
4	ϕ_o>0.55 and D/T<0.4	O	3.1
5	ϕ_o>0.5 and ρ_o>1398	O	1.5
6	ϕ_o>0.63 and D/T>0.5	O	1.5
7	ϕ_w >0.5, r>1.2 and N>1274 rpm	W	1

Table 2. Inferred rules from the ensemble of trees.

		Reference								
		One tree – CP = 0.01			**One tree – CP = 0.007**			**300 bagged trees**		
Prediction		N/D	O	W	N/D	O	W	N/D	O	W
	N/D	74	7	25	73	5	12	88	3	6
	O	19	91	0	19	93	0	8	95	0
	W	5	0	73	6	0	86	2	0	92

Table 3. Confusion matrices for decision trees pruned at different complexity parameters and a bagging of 300 trees. Test subset is the same for all.

		Reference											
		Linear SVM			**Quadratic SVM**			**Cubic SVM**			**Gaussian SVM**		
Prediction		N/D	O	W	N/D	O	W	N/D	O	W	N/D	O	W
	N/D	81	10	12	88	6	12	85	5	5	88	9	8
	O	10	88	1	7	92	0	10	93	0	7	89	0
	W	7	0	85	3	0	86	3	0	93	3	0	90

Table 4.Confusion matrices for the support vector machine models with different kernels.

		Reference											
		3 neurons			5 neurons			7 neurons			10 neurons		
Prediction		N/D	O	W	N/D	O	W	N/D	O	W	N/D	O	W
	N/D	75	16	11	86	5	13	83	9	14	81	7	6
	O	13	82	0	8	93	0	8	89	0	11	91	0
	W	10	0	87	4	0	85	7	0	84	6	0	92

Table 5. Confusion matrices for the neural networks with different hidden layer sizes.

RMSE of	One decision tree	300 bagged trees
ϕ_w	0.0657	0.0712
ϕ_o	0.0637	0.0564

Table 6. Performance of each decision tree model.

Prediction of	C	ε
ϕ_w	0.1490	0.01490
ϕ_o	0.1291	0.01291

Table 7. Parameters used in SVR models.

Kernel Type	Kernel Function	Scale Parameter	RMSE	
			ϕ_o	ϕ_w
Linear	$k_{(x_i,x_j)} = \frac{\langle x_i, x_j \rangle}{\lambda^2}$	$\lambda = 0.4135$	0.0713	0.0858
Quadratic	$k_{(x_i,x_j)} = \left(1 + \frac{\langle x_i, x_j \rangle}{\lambda^2}\right)^2$	$\lambda = 0.3495$	0.0678	0.0570
Cubic	$k_{(x_i,x_j)} = \left(1 + \frac{\langle x_i, x_j \rangle}{\lambda^2}\right)^3$	$\lambda = 0.4703$	0.0821	0.0695
Gaussian	$k_{(x_i,x_j)} = e^{-\frac{\|x_i - x_j\|^2}{2\lambda^2}}$	$\lambda = 0.9$	0.0630	0.0717

Table 8. Kernel functions tested, their scale parameters and the performance of each model as per the RMSE.

Number of neurons	RMSE	
	ϕ_o	ϕ_w
5	0.0556	0.0591
10	0.0537	0.0595
15	0.0532	0.0577
20	0.0533	0.0590

Table 9. Performance of each neural network.

	Neurons in 1st layer	Neurons in 2nd layer			
		2	4	6	8
ϕ_o	5	0.0554	0.0542	0.0537	0.0543
	10	0.0542	0.0551	0.0539	0.0539
	15	0.0569	0.0537	0.0531	0.0534
	20	0.0541	0.0537	0.0533	0.0539
ϕ_w	5	0.0585	0.0573	0.0586	0.0601
	10	0.0610	0.0536	0.0596	0.0569
	15	0.0590	0.0591	0.0590	0.0587
	20	0.0581	0.0610	0.0583	0.0592

Table 10. RMSE of the two-layered neural networks with the test set.

Figure captions

Figure 1. Ambivalent range for different pairs of liquids: toluene(O)-water(W)(Arashmid and Jeffreys 1980) (solid lines), CCl_4(O)-water(W)(Arashmid and Jeffreys 1980) (dashed lines) and heptane(O)-acetonitrile(W)(S. Kumar 1996) (dotted lines).

Figure 2. Scheme of a feedforward multiple perceptron (taken from Heidari et al.(Heidari et al. 2018)).

Figure 3. Decision tree to decide whether a given phase will be continuous.

Figure 4. Optimum tree, pruned at a cp of 0.007.

Figure 5. Prediction of PI points as per the decision tree models. Dashed lines represent a margin of ± 0.03, which is considered a suitable error gap.

Figure 6. Performance of different kernel functions used in SVR algorithms. Test set is the same for all kernels. Dashed lines represent a margin of $\pm$ 0.03, which is considered a suitable error gap.

Figure 7. Prediction of each neural network with one hidden layer and varying number of neurons. Test subset is the same for all and the same for the other models in this work. Dashed lines represent a margin of ± 0.03, which is considered a suitable error gap.

Figure 8. PI hypothetical map as predicted by the neural network model. Phase density differences are in kg/m^3 and interfacial tensions in mN/m.

Figure 9. PI curves for systems having the same density difference (ρ_c-ρ_d=100 kg/m^3) and viscosity ratio (μ_c/μ_d=1.25). Solid line refers to the inversion from O/W to W/O, while dashed line is the opposite inversion.

Figure 1

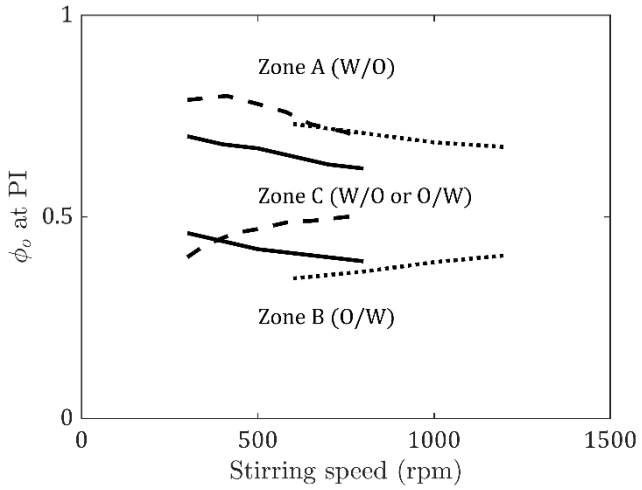

Figure 2

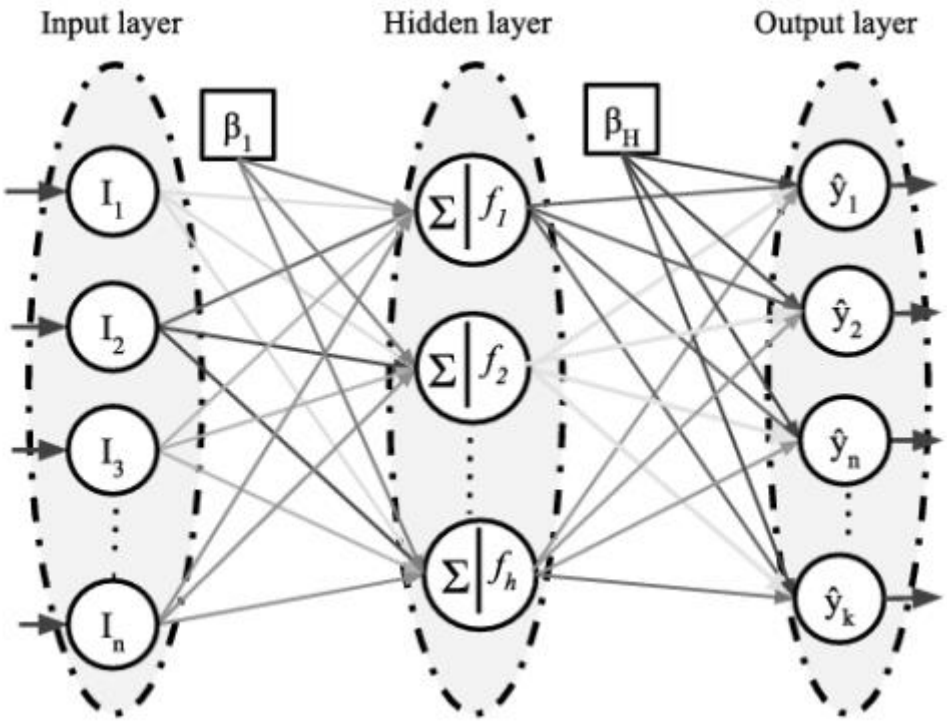

Figure 3

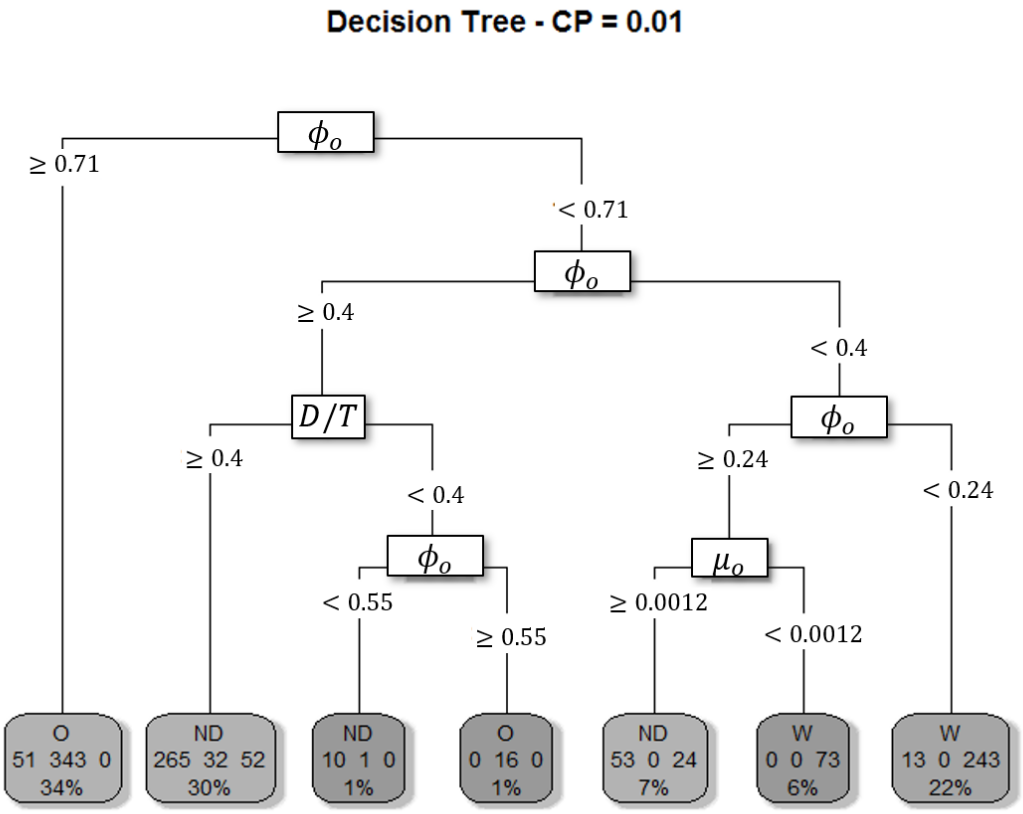

Figure 4

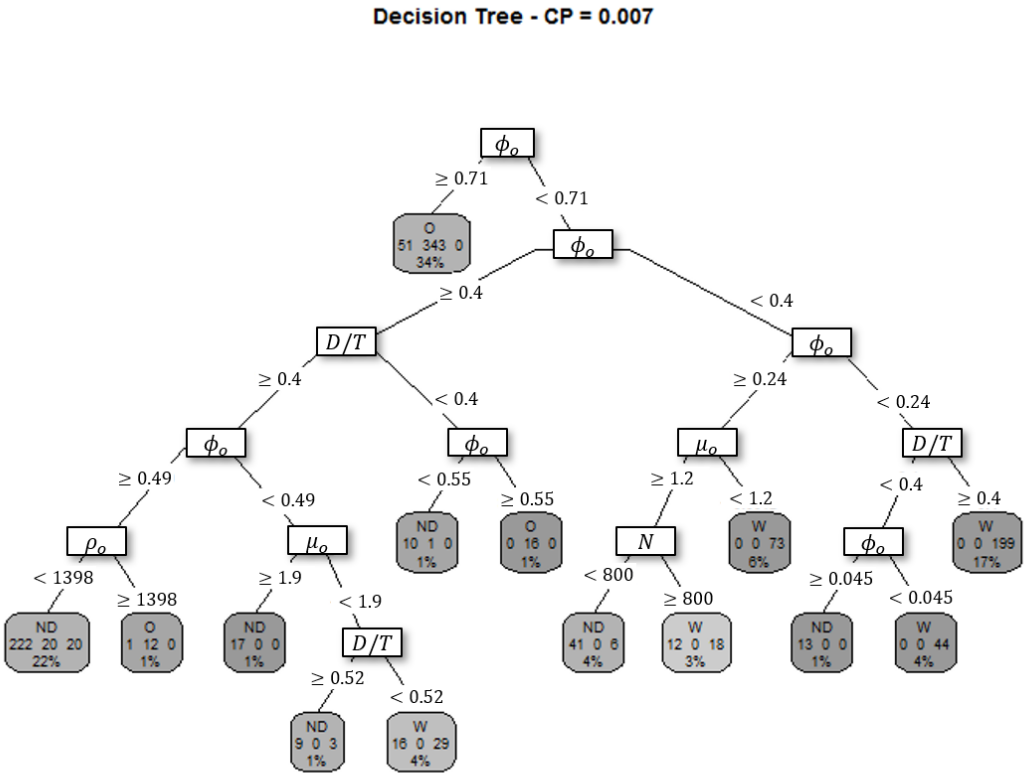

Figure 5

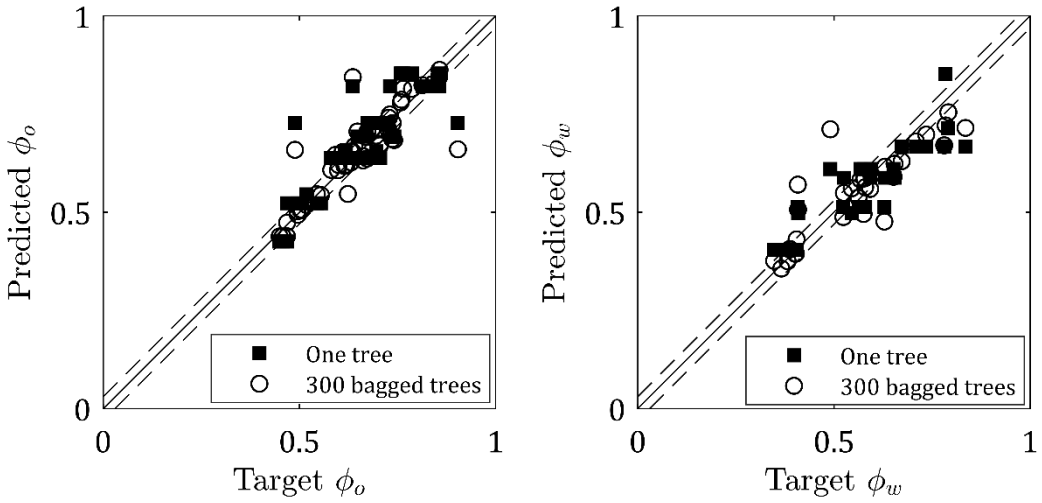

Figure 6

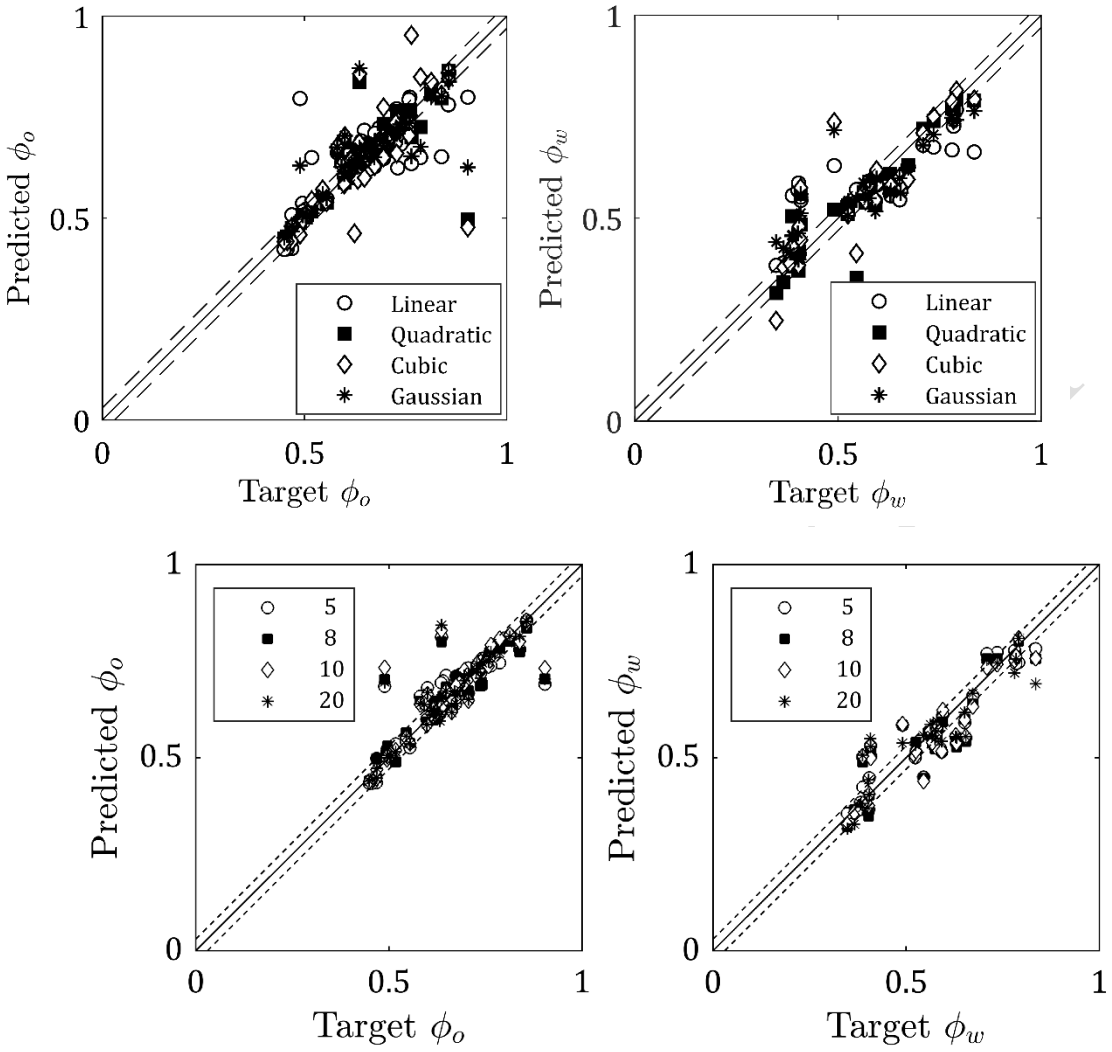

Figure 7

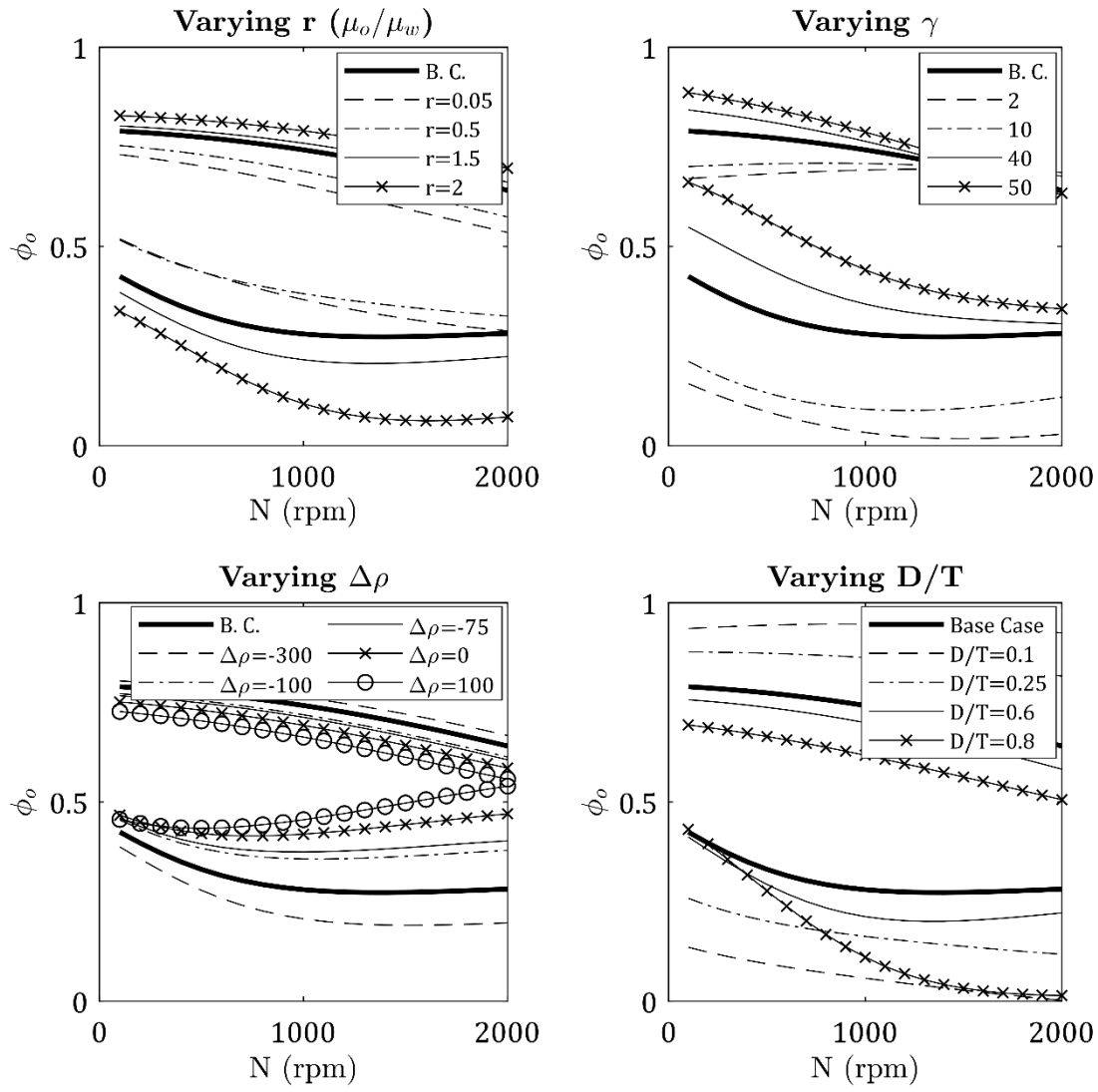

Figure 8

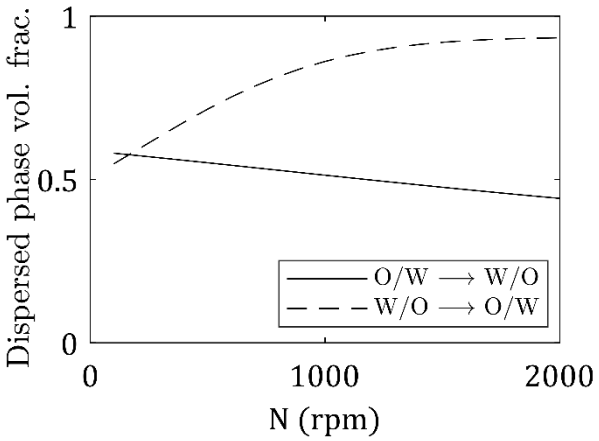

Figure 9