Neural Networks Tools
for Improving Tacite Hydrodynamic Simulation
of Multiphase Flow Behavior in Pipelines

I. Rey-Fabret[1], R. Sankar[1], E. Duret[1], E. Heintze[1] et V. Henriot[1]

1 Institut français du pétrole, 1 et 4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex - France
e-mails : isabelle.rey-fabret@ifp.fr - emmanuel.duret@ifp.fr - eric.heintze@ifp.fr - veronique.henriot@ifp.fr

Résumé — **Les réseaux de neurones pour l'amélioration de la simulation hydrodynamique du comportement des écoulements polyphasiques dans les conduites** — Les simulateurs d'écoulements multiphasiques sont généralement utilisés pour dimensionner le schéma de production. L'un des problèmes rencontrés est de prédire de façon exacte une chute de pression et l'arrêt de circulation. Ceci peut être résolu grâce à l'utilisation d'un schéma numérique précis et un comportement thermodynamique approprié lié à un modèle hydrodynamique adapté et robuste.

Dans le code Tacite, développé par l'*IFP*, un modèle hydrodynamique mécanique a été développé. Ce modèle est capable de prédire le régime d'écoulement, les vitesses de phases et les chutes locales de pression pour tous profils et tous diamètres. Il contient les lois de fermeture fondées sur les régimes d'écoulements. Ce modèle mécanique a été validé sur différentes bases de données.

Les deux limitations de ce type de modèle hydrodynamique peuvent être les perturbations mathématiques (continuité et dérivabilité ne sont pas toujours garanties) et le temps de calcul. Ceci peut être gênant lorsque l'on utilise un schéma numérique précis requérant des calculs de dérivées, et dans l'objectif du temps réel.

Cet article présente une démarche s'appuyant sur les réseaux de neurones pour remplacer le module hydrodynamique du modèle diphasique selon deux objectifs :

– éliminer les problèmes de discontinuité durant les calculs hydrodynamiques ;

– réduire significativement le temps de calcul.

Cette méthode a été testée sur des données expérimentales et simulées. Les résultats présentés prouvent la pertinence de cette approche.

Mots-clés : écoulements polyphasiques, réseaux de neurones, hydrodynamique.

Abstract — *Neural Networks Tools for Improving Tacite Hydrodynamic Simulation of Multiphase Flow Behavior in Pipelines* — *Transient multiphase flow simulators are generally used to dimension the production scheme. One of the problems encountered is to predict accurately the pressure drop and the liquid holdup. This can be solved using an accurate numerical scheme and an appropriate thermodynamic behavior linked to an accurate and robust hydrodynamic model.*

In the Tacite Code, developed by IFP, *a mechanistic hydrodynamic model has been developed. This model is able to predict the flow regime, the phase velocities and the local pressure drop for all slopes and all diameters. It contains closure laws based on flow regimes. This mechanistic model has been validated against various data banks.*

The two limitations of such an hydrodynamic model may be its mathematical disturbance (continuity, derivability are not always guaranteed) and the time consuming. This can be troublesome when using an accurate numerical scheme that requires derivative computation and for real time purposes.

This paper presents a neural network based approach to efficiently replace the hydrodynamic module in the two phase model with the following two objectives:

– to avoid discontinuity problems during hydrodynamic computations;

– to reduce significantly computational time.

This method was tested with experimental and simulated data. The results given in this paper prove the relevancy of this approach.

Keywords: multiphase flow, neural networks, hydrodynamic.

NOTATIONS

V_M barycentric mixture velocity (m/s)
DV difference between gas velocity and liquid velocity
β fraction of separated flow
V_G gas velocity
V_L liquid velocity.

INTRODUCTION

The Tacite compositional Code [1] is able to simulate the behavior of transient multiphase flow in pipeline with process equipment such as controllers, valves, separators or lateral injectors. This code is composed of three main parts: the thermodynamic module describes the thermodynamic state of the system and calculates the fluid physical properties, the hydrodynamic module mainly indicates the local pressure drop and phases velocities, and a precise and robust numerical scheme gives each time a solution [2, 3].

However, this simulator is also able to give precisely the flow regime during time at each point of a pipeline. But there are two limitations with such an hydrodynamic model: continuity, derivability are not always guaranteed and the computing time is too excessive. This can be troublesome for real time purposes or when using an accurate numerical scheme that required derivative computation.

This paper presents a system based on artificial neural network as an alternative to this hydrodynamic model used for two phases flow configurations in transient cases. Main objectives of this approach are to always assure continuity between different flow regimes and to significantly reduce computational time. Moreover this model has to be compatible with the computational environment of the Tacite model.

The first part of this paper gives a short presentation of the Tacite hydrodynamic model and neural networks principles. In the second part, we describe the structure of the neural network system developed for the hydrodynamic estimation and we explain technical choices for making use of it. Simulations are finally presented to allow comparison between the existing hydrodynamic model and the neural network system.

1 TACITE HYDRODYNAMIC MODEL

In this section, we describe the hydrodynamic model considered as a reference for Tacite development: this model takes into account flows regimes and the existing knowledge in hydrodynamic area. There are three main flow regimes considered with Tacite:

– in stratified flow, a liquid layer flows beneath the gas stream. The phases interact through the interface which may be smooth or wavy;

– in dispersed flow, the gas is assumed to be distributed in the form of spherical bubbles, and the liquid phase is continuous;

– in the intermittent flow, liquid slugs containing dispersed bubbles alternate with gas pockets where the flow is essentially stratified.

Finally, main outputs of this model are DV, the difference between gas velocity and liquid velocity, and β, the fraction of separated flow; the others follow directly from these values.

This model is strongly nonlinear, because of the nature of described phenomena. Then, sometimes, it can not assure continuity and derivability of the solutions. This implies long time computation. In order to accelerate obviously time computation required for the estimation of hydrodynamic flow characteristics and for the numeric scheme resolution, it seems necessary to assure DV continuity and derivability and to estimate β without help of a choice module. Indeed, the Tacite hydrodynamic module solves two implicit functions and calculates by iterations hydrodynamic flow characteristics; an explicit C^{∞} function would give faster these results.

2 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are a statistical approach of data processing, based on training methods [4-6]. Initially this approach was used for automatic classification. Today, its ability to approximate complicated systems behavior allows to consider it also as a modelling tool. This approach is then

dedicated to facilitate the resolution of problems which were previously difficult or impossible to solve. They are often characterized by their nonlinearity, and the large number of their entries.

In their concept, artificial neural networks are a simplified simulation of biological neural networks. They are structured in many layers of neurons (*Fig. 1*), through which the information signal travels to be processed.

– As input, each neuron receives signals sent by other neurons. It then applies on them a generally nonlinear function, named *activation function*, in order to produce its output.

– Neurons are connected between them to transmit information. Each link is associated with a weight to modulate the importance of each information part in the neural network by multiplying the signal transmitted. Then, the neuron inputs are weighted sums of connected neuron outputs (*Fig. 2*).

The development of a neural network consists in the determination of its structure and its optimum weights. To realize this optimization, the network has to be trained with help from a database. We can then distinguish two main categories of artificial neural networks: the unsupervised one [7] and the supervised one.

– The first one is usually used for classification, because of their capacity of self-organization; no feedback information about output results are given to help the network in its data analysis.

– With the supervised method, the network development is realized with help from input and output information, to determine optimum link weights.

Whatever the nature of neural network, this approach requires an important database. This database is divided in three portions, each one representing the behavior of the system in all its definition domain, respectively dedicated for neural network training, calibration and generalization.

The training set is used to develop the network. During this phase of work, different algorithms can be chosen, depending of the neural network nature; they are used to calculate optimum link weights to estimate the desired outputs.

The calibration database helps to identify the good moment to stop the training process. Indeed, when the training is too long, the network is overtrained; it memorizes the examples proposed by the training database, but its outputs are not representative of system response in all its definition domain. It is unable to interpolate the system behavior between two examples extracted from training database. Then the calibration database is used during the training process to evaluate the generalization capacity of the network. When the difference between desired outputs and calculated outputs for this database increases, the training process must be stopped.

The generalization database contains signals that the network never processed. It is used only when a network is chosen, its structure determined and optimum weights estimated. Then this base allows to test its generalization ability: the estimation error given by the difference between desired outputs and the same outputs calculated by the network represents an evaluation tool for the estimation of its performances.
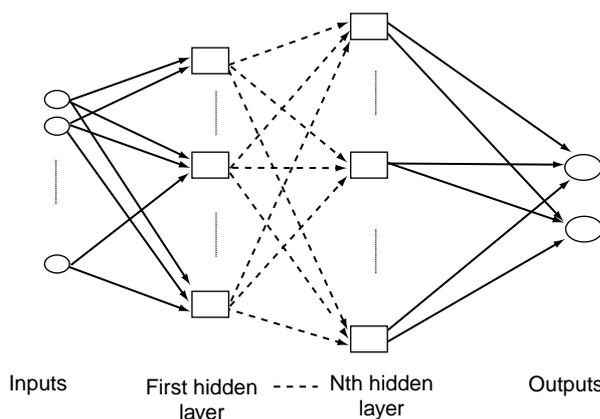

Figure 1

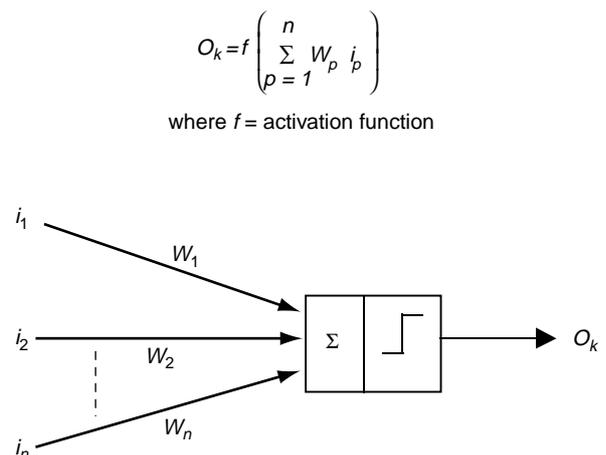General scheme of neural network structure.

$$O_k = f\left(\sum_{p=1}^{n} W_p \, i_p\right)$$

where f = activation function

Figure 2

Neuron process.

3 HYDRODYNAMIC MODELLING USING ARTIFICIAL NEURAL NETWORKS

In this section, we present the solutions chosen to generate a model based on artificial neural networks able to estimate at each time the flow regime in the case of two phases flows. Technical choices are described, which take into account environmental constraints.

3.1 Databases

3.1.1 Description

One must keep in mind that the reference model to develop is the Tacite hydrodynamic model. For this reason, a database was created, according to the physical domain of inputs variations and corresponding outputs calculated by Tacite. This database can be as big as needed for neural networks development. Furthermore, generalization performances of a neural network are particularly better when the training database gives a good representation of outputs variations. Thus, *DV* outputs contained in the database dedicated for Tacite hydrodynamic modelling are uniformly distributed on a given interval.

Moreover, beside this synthetic database, an experimental one is also available. It is the Boussens database [8] which was used for the validation of our reference hydrodynamic model.

3.1.2 How to Use these Data?

These databases allow to elaborate neural networks system and to optimize it. They are also useful to evaluate the system behavior facing new unknown data. In the case of networks elaboration, a first package of synthetic data is used for neural networks training and a second one helps to determine overtraining.

In order to validate the system with unknown data, two other groups of data are processed:
- The first one is composed of the last unused data extracted from the synthetic database (near 36 000 points). This makes it possible to compare the outputs given by the neural network system and Tacite hydrodynamic module in the case of inputs and outputs never processed in elaboration phases. These comparisons must prove validity and efficiency of technical choices.
- The second one is the Boussens database. It represents particular experimental cases. The use of this database aims at showing the neural network system behavior in case of field data.

For those different validation tests, performances have to be evaluated:
- Concerning *DV* estimation, a comparison between the (DV/V_M) ratio estimated by the system and that contained

in the synthetic database is done. Indeed, this ratio is of prime necessity for process implemented in other Tacite modules. The compatibility between neural networks system and the environment represented by Tacite is then measurable thanks to this ratio. The same comparison is performed for Boussens data.
- Evaluations of the performances are more complex with β estimation: regression and classification capacities of the system have to be proved. Then, a confusion table is calculated: it gives the percentage of well classification of β, for each class of β, obtained when using either neural networks or Tacite module. Furthermore, in the case of $0 < \beta < 1$, the regression ability of the system is evaluated by comparing β given by Tacite with β processed by networks.

3.2 Neural Networks

DV and β can be separated as outputs of two different systems. Indeed, they are not correlated, and *DV* continuity and derivability are not imposed in the case of β estimation.

As inputs and outputs are well known for each *DV* and β estimation, the supervised neural network methods seem to fit the best.

Taking into account these considerations, *DV* estimation is realized thanks to Multi Layer Perceptron (MLP) neural networks [4-5].

The system dedicated to *DV* estimation is described in Figure 3. It consists of three elements:
- normalization: as the magnitude orders of the 10 inputs can highly differ between them, it is necessary to normalize them;
- the MLP neural networks link the 10 inputs with *DV* by a nonlinear transformation;
- after MLP process, *DV* varies in [0;1] interval. Then a denormalization function is required to obtain its real value.

In the case of β estimation, MLP neural networks are used for classification and regression. The system developed for this study is able to give the nature of flow regime in the case of one or two phases. As for DV estimation, inputs have to be normalized. But there is no need for β denormalization: β value always varies between 0 and 1 (*Fig. 4*).

Training phases are realized thanks to the training part of the synthetic database. The second part of those data is also used, to determine the overtraining threshold. For each neural network to be realized, different structures are tested, whose differences are the number of hidden layers, the number of neurons in each layer, the activation function, and the cost function to minimize as estimation of the distance between the objective (Tacite reference hydrodynamic module outputs) and neural network ones.

Figure 3

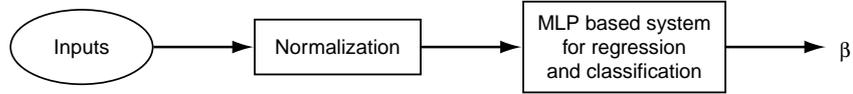Estimation of the difference between gas velocity and liquid velocity.

Figure 4

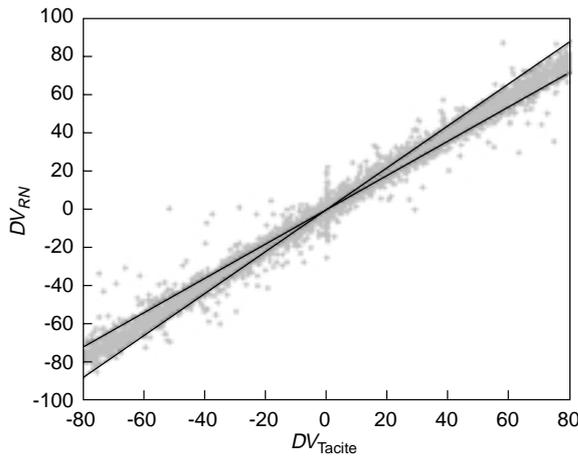Estimation of the fraction of separated flow.

Figure 5

The difference *DV* between gas velocity and liquid velocity estimated by the neural networks system as function of *DV* given by Tacite.
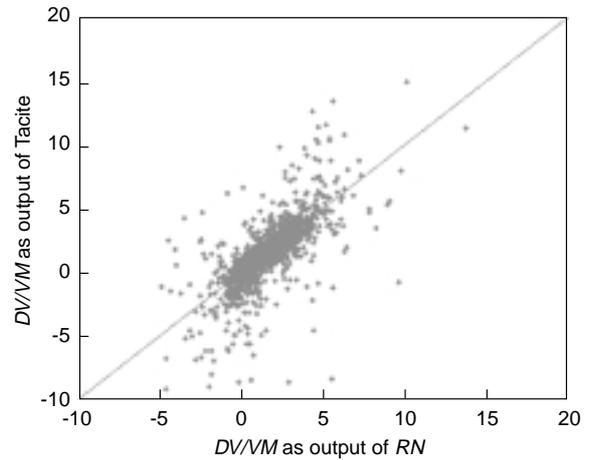

Figure 6

The ratio between the difference *DV* between gas velocity and liquid velocity estimated by the neural network system and the barycentric mixture velocity, *VM*, as function of the same ratio given by Tacite.

Depending on the nature of the regression and of the classification, a structure is finally chosen for each part of the neural networks system, to optimize *DV* and β estimation. This optimization represents an important work which determines the results quality.

3.3 Results

3.3.1 Validation of *DV* Estimation

Figures 5 and 6 show neural networks generalization performances for *DV* regression in the case of synthetic database. They represent synthetic database values as function of estimated values, respectively for *DV* and (DV/V_M) ratio. Qualitatively, only a few points are far from

the first bisector. Table 1 gives statistical quantities to quantitatively compare the (DV/V_M) ratio contained in the synthetic database (*SD*) with the same ratio calculated by the neural networks system. This table shows that the system is able to reproduce the statistical behavior of the Tacite module. Moreover, one can define the relative error as:

$$E(\text{Tacite}, NN) = \frac{1}{K}\sum_{i=1}^{K}\left[\frac{\left|\left(\dfrac{DV}{V_M}\right)_{i\,\text{Tacite}} - \left(\dfrac{DV}{V_M}\right)_{i\,NN}\right|}{\text{Max}\left(1,\left|\left(\dfrac{DV}{V_M}\right)_{i\,\text{Tacite}}\right|\right)}\right]$$

where *NN* = neural network and *K* = total number of data.

TABLE 1

Statistics concerning the ratio between the difference *DV*
between gas velocity and liquid velocity
and the barycentric mixture velocity, V_M,
in the case of synthetic database

	Mean	Standard deviation
Synthetic database	0.69	0.63
Neural networks estimations	0.7	0.55

TABLE 2

Statistics concerning the ratio between the difference *DV*
between gas velocity and liquid velocity estimated
by the neural network system and the barycentric mixture velocity,
V_M, in the case of synthetic database

	Mean	Standard deviation
Boussens database	1.7	4.33
Neural networks estimations	1.3	4.01

When applied to the synthetic database and its estimation by neural networks, we obtain:

$$E(\text{Tacite}, NN)_{SD} = 0.08$$

This relative error is low. Figure 7 shows the error *E* distribution for each estimation: more than 85% of unknown examples presented to the neural networks system are estimated with less than 10% of relative error. These results confirm that the system developed comes up to modelling expectations: C^∞ property of *DV* response obtained with neural network method doesn't generate big relative errors.

In the case of Boussens measurements, the results are good too. Table 2 gives statistical results corresponding to the same ratio estimation, which are well estimated too. However, these results are less satisfactory than the previous ones, because of the nature of the noised measurements used to calculate $(DV/V_M)_{\text{Boussens}}$ ratio.
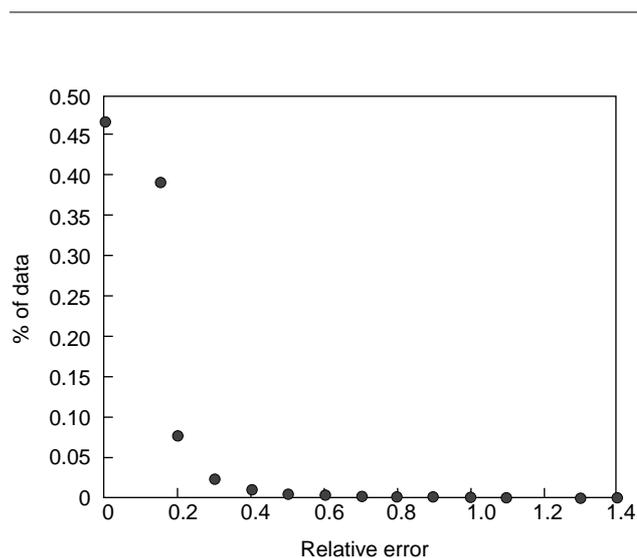
TABLE 3

Confusion table for β classification in the case of general database

Data	Estimation			
	C1 = {β = 0}	C2 = {0 < β < 1}	C3 = {β = 1}	
C1 = {β = 0}	**87.3%**	10.7%	2%	100%
C2 = {0 < β < 1}	3.1%	**81.5%**	15.4%	100%
C3 = {β = 1}	0.2%	4.8%	**95%**	100%

3.3.2 Validation of β Estimation

As explained previously, performances evaluation of β estimation is composed of two parts: β classification and β regression in]0;1[interval. A confusion table helps to show classification ability of the system; a representation of β_{Tacite} as function of β_{NN} is dedicated to regression performances.

In terms of classification, Table 3 shows that the system can classify very efficiently unknown data to determine the nature of flow regime (dispersed = C1; intermittent = C2; stratified = C3). Indeed, the results concerning well classification are respectively 87,3% for dispersed regime (β = 0), 81.5% for intermittent one (0 < β < 1) and 95% for stratified one (β = 1). The worst result is obtained for β ∈]0;1[, due to difficulties to classify the values at the border of the other classes.

Concerning β regression in the case of an intermittent flow regime, Figure 8 shows how the system is able to evaluate β value. Even if those results are not as accurate as the ones obtained for *DV* regression, their precision is adequate: β influence is less than *DV* one for the computation of numerical flux and derivatives.

All previous results are globally well and validate the process chosen in this study. Then, a general system elaborated with help from a synthetic database including all physical intervals of inputs variations can be generated to give good estimations of main hydrodynamic quantities, with a important reduction of computation time: the factor of reduction is nearly 15 (mean of nearly 36 000 calculations of synthetic data). This method can be improved by elaborating specific models, linked to particular production fields. The next paragraph explains this process in the case of Boussens field.

Figure 7

Distribution of relative error for the ratio between the difference *DV* between gas velocity and liquid velocity and the barycentric mixture velocity, V_M, for neural network/Tacite comparison.
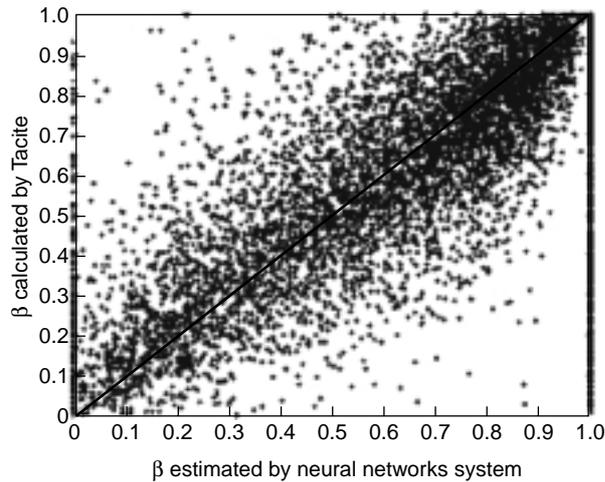
Figure 8

Regression of the fraction of separated flow in the case of $0 < \beta < 1$ synthetic database.

3.3.3 Validation of the Method for Boussens Field

As explained in the previous part about *DV* estimation, inputs variations are different for Boussens field than for synthetic database. It is then possible to generate a synthetic database dedicated to this case by changing variations intervals of inputs definitions. The system optimization is based on this specific database, to take into account field specificity. The system structure (number of hidden layers, number of neurons for each layer, activation function, etc.) are adapted to field characteristics. Then, the estimation results are remarkably improved. Indeed, we can for example compare confusion tables previously obtained in general cases with the one stemmed from specific optimization (*Table 4*): results are well and confirm a saving of precision.

This system was tested with Boussens data, where some data are calculated with help from measurements. In this case, β classification results are given by Table 5. Taking into account that it is very difficult to distinguish β values near 0 or equal to 0 (respectively near 1 or equal to 1) by measurements, β is generally well estimated. But some values near 0 are attributed to the second class, and there is the same phenomenon near 1. This explains the percent equal to 42.4% and 28.6% of the confusion table. Moreover, values associated with these bad classified points are estimated near 0 (respectively near 1) with β regression. Then, their bad classification is directly linked with measurements precision. In conclusion, even if those results are not as excellent as synthetic ones, they give generally a good β estimation and show that the precision obtained with a process of specific neural network system optimization is useful to make the system robust with noise measurements.

TABLE 4

Confusion table for β classification in the case of specific database

Data	Estimation			
	C1 = {$\beta = 0$}	C2 = {$0 < \beta < 1$}	C3 = {$\beta = 1$}	
C1 = {$\beta = 0$}	**99.49%**	0.5%	0.01%	100%
C2 = {$0 < \beta < 1$}	1%	**98.3%**	0.7%	100%
C3 = {$\beta = 1$}	0.05%	2.9%	**97%**	100%

TABLE 5

Confusion table for β classification in the case of Boussens database

Data	Estimation			
	C1 = {$\beta = 0$}	C2 = {$0 < \beta < 1$}	C3 = {$\beta = 1$}	
C1 = {$\beta = 0$}	**57.6%**	42.4%	0%	100%
C2 = {$0 < \beta < 1$}	0%	**85.6%**	14.4%	100%
C3 = {$\beta = 1$}	0.3%	28.6%	**71.1%**	100%

CONCLUSION

The aim of this study was to model effluents hydrodynamic behavior in pipelines thanks to derivable solutions, even at flow regime transitions. Modelling constraint was to be compatible with the physical environment of the existing model, in order to replace it directly. Then, the modelling reference was the Tacite hydrodynamic module and development validations had to be realized by comparing system results with Tacite ones.

The main parameters to characterize hydrodynamic behavior are the difference of velocity between gas and liquid named *DV*, and β, the separated flow fraction. This paper shows that those parameters are well estimated using neural networks methodology. Moreover, the efficiency of this development is all the more proved because computing time is significantly reduced: the saving of time is nearly 15.

Furthermore, this method shows that synthetic databases are more useful and efficient for such developments than real ones, because of their ability to take into account all possible real cases. But we have seen that if such a process can be used generally, the neural networks model can be adapted to specific field applications in order to be more precise and then to give excellent performances. In conclusion, this method is efficient, fast to process and adaptable to each case of production field, to obtain still better precision. This model is then a new tool to follow along time the flow behavior in a production field.

REFERENCES

1 Pauchon, C.L. (1994) Tacite: A Transient Tool for Multiphase Pipeline and Well Simulation. *SPE Annual Technical Conference and Exhibition.*

2 Masella, J.M., Tran, Q.H., Ferre, D. and Pauchon, C. (1998) Transient Simulation of Two-Phase Flows in Pipes. *International Journal of Multiphase Flow*, **24**, 739-755.

3 Faille, I., Heintze, E. (1999) A Rough Finite Volume Scheme for Modelling Two-Phase Flow in a Pipeline. *Computers & Fluids,* **28,** 213-341.

4 Rivals, I., Personnaz, L. and Dreyfus, G. (1995) Modélisation, classification et commande par réseaux de neurones: principes fondamentaux, méthodologie de conception et illustrations industrielles, in *Les réseaux de neurones pour la modélisation et la commande de procédés*, Lavoisier Tec et Doc.

5 Lippmann, R.P. (1987) An Introduction to Computing with Neural Nets*, in IEEE ASSP Magazine*, avril 1987, 4-22.

6 Dreyfus, G. (1998) Les réseaux de neurones, in *Mécanique industrielle et matériaux*, **51**.

7 Frayssinet, D. (2000) Utilisation des réseaux de neurones en traitement des données de diagraphies : prédiction en reconstitution de faciès lithologiques. *Mémoire ingénieur,* CNAM Paris.

8 Corteville, J. (1983) Expérimentation des écoulements diphasiques en conduites pétrolières : boucle d'essais de Boussens, in *Revue de l'Institut français du pétrole*, **38**, 143-151.