Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemical Engineering and Processing 42 (2003) 697–713
Chemical Engineering and Processing
www.elsevier.com/locate/cep

Using hybrid neural networks in scaling up an FCC model from a pilot plant to an industrial unit

G.M. Bollas [b], S. Papadokonstadakis [a], J. Michalopoulos [a], G. Arampatzis [a], A.A. Lappas [b], I.A. Vasalos [b], A. Lygeros [a,*]

[a] School of Chemical Engineering, National Technical University of Athens, Zografou Campus, Heroon Polytechniou 9, Athens, GR-157 80, Greece
[b] Chemical Process Engineering Research Institute (CPERI), Centre for Research and Technology Hellas (CERTH), P.O. Box 361, 57001 Thermi-Thessaloniki, Greece

Received 13 May 2002; received in revised form 30 September 2002; accepted 30 September 2002

Abstract

The scaling up of a pilot plant fluid catalytic cracking (FCC) model to an industrial unit with use of artificial neural networks is presented in this paper. FCC is one of the most important oil refinery processes. Due to its complexity the modeling of the FCC poses great challenge. The pilot plant model is capable of predicting the weight percent of conversion and coke yield of an FCC unit. This work is focused in determining the optimum hybrid approach, in order to improve the accuracy of the pilot plant model. Industrial data from a Greek petroleum refinery were used to develop and validate the models. The hybrid models developed are compared with the pilot plant model and a pure neural network model. The results show that the hybrid approach is able to increase the accuracy of prediction especially with data that is out of the model range. Furthermore, the hybrid models are easier to interpret and analyze.

Keywords: Fluid catalytic cracking; Process modeling; Neural networks; Hybrid modeling; Pilot to commercial unit scale up; Riser hydrodynamics; FCC kinetics

1. Introduction

Fluid catalytic cracking (FCC) is an important oil refinery process, which converts high molecular weight oils into lighter hydrocarbon products. It consists of two interconnected gas–solid fluidized bed reactors: the riser reactor, where almost all the endothermic cracking reactions and coke deposition on the catalyst occur, and the regenerator reactor, where air is used to burn off the coke deposited on the catalyst. The heat produced is carried from the regenerator to the reactor by the catalyst. Thus, in addition to reactivating the catalyst, the regenerator provides the heat required by the endothermic cracking reactions [1]. Industrial FCC units are designed to be capable of using a variety of feedstocks, including straight run distillates, atmospheric and vacuum residua and vacuum gas oils. They produce a range of products, which must adapt to seasonal, environmental and other changing demand patterns. Since FCC units are capable of converting large quantities of heavy feed into valuable lighter products, any improvement in design, operation or control can result in substantial economic benefits.

FCC pilot plant units are often used to develop accurate prediction and optimization models. The reason is that the operation of pilot plant units can be easily adapted under a wide range of conditions (feed properties, catalysts operating conditions). The main difficulty when translating the pilot scale unit observations to the commercial unit reality is to predict the scale up effects, arisen from the small geometrical features of pilot scale units.

Furthermore, due to the complexity of the industrial FCC units, it is very difficult to obtain accurate models. The complexity arises from the strong interactions between the operational variables of the reactor and the regenerator. Moreover, there is a large degree of

* Corresponding author. Tel.: +30-1-7723973; fax: +30-1-7723155.
E-mail address: alygeros@chemeng.ntua.gr (A. Lygeros).

0255-2701/03/$ - see front matter
doi:10.1016/S0255-2701(02)00206-4

uncertainty in the kinetics of the cracking reactions and catalyst deactivation by coke deposition in the riser reactor and the coke burning process in the regenerator [2].

Artificial neural networks (ANNs) are a promising alternative modeling technique. They are mathematical models, which try to simulate the brain's problem solving approach. Neural networks have been known for decades. The beginning of neural network research can be traced back to the 1940s, but until the early 1980s the research was basically theoretical and limited. The tremendous evolution of digital technology over the past two decades provided the necessary computational power in order to use neural networks in various fields. Applications of neural networks in chemical engineering appeared in the late 1980s [3]. Some published applications of ANNs in chemical engineering are: fault diagnosis in chemical plants [4], dynamic modeling of chemical processes [5], system identification and control [6], sensor data analysis [7], chemical composition analysis [8], and inferential control [9]. Quite recently there have been publications of applying neural networks in FCC modeling [10,11]. Neural networks appear to be suitable for FCC modeling because: (a) they can handle non-linear multivariable systems; (b) they are tolerant to the faults and noise of industrial data; (c) they do not require an extensive knowledge base; and (d) they can be designed and developed easily [3].

Neural networks can also be used as hybrid models. The term hybrid modeling is used to describe the incorporation of prior knowledge about the process under consideration in a neural network modeling approach. The way that this incorporation can be done depends on the form of the prior knowledge available as well as on the desired properties of the model to be created. The two main categories are the design approach, in which prior knowledge dictates the overall model structure and the training approach, which dictates the form of the weights estimation problem [12]. This paper handles only the design approach. A successful design hybrid modeling approach can lead for example to models with better generalization and extrapolation abilities compared to pure neural network models, especially when there are only a few and noisy data for the training of the neural models, which is often the case, when it comes to industrial databases. For example, if the prior knowledge is captured in a phenomenological or an empirical model, then this can lead to a lower dimensionality of the input vector used by the neural network, since some or all of the variables that are used by this model can be considered redundant for the neural network training. Consequently, neural networks with fewer weights can be applied, which is a crucial point when the data are sparse. Furthermore, smaller neural networks require less computational time for the training procedures. Another advantage of hybrid models, which arises from their internal structure, is that they can be much more easily interpreted and analyzed than simple 'black box' neural networks [12,13].

In literature there exist some efforts of hybrid modeling for the purposes of generalized on-line state estimation [13], and for fed batch bioreactors [12,14], but the scientific area of scaling up a pilot plant model to a commercial unit using hybrid modeling schemes seems not to have been explored.

In this paper a detailed kinetic–hydrodynamic model was developed for the simulation of the riser reactor of the FCC pilot plant unit, located in Chemical Process Engineering Research Institute (CPERI) in Thessaloniki, Greece [15]. The catalyst hold-up and its residence time in the reactor were calculated via a comprehensive hydrodynamic scheme and the conversion and coke yield were successfully predicted through a Blanding type [16] kinetic model [17]. The applicability of the 'pilot plant model' to the actual operation of the industrial unit of Aspropyrgos Refinery of Hellenic Petroleum S.A., the largest and most complex oil refinery in Greece, is examined. Modifications are made to the model from modeling and design perspectives. A neural model trained with data from the commercial unit is combined with this pilot plant model in various hybrid modeling schemes and compensates for the differences in unit geometry, feedstock quality and catalyst properties, between the ideal operation of the pilot plant and the reality of the commercial unit.

The structure of the paper is as follows: coupling of hydrodynamic and kinetic theories for the FCC Unit is presented in Section 2. Section 3 provides a description of neural networks and of hybrid models using neural networks. In Section 4 the pilot plant model is presented. In Section 5 there is a description of the development of the hybrid models. Section 6 presents the results of the best hybrid models and provides a comparison with the pilot plant model and a simple neural model. Concluding remarks are presented in Section 7.

2. Coupling of hydrodynamic and kinetic theories for the pilot and the commercial FCC unit

A model was developed to study the strong coupling between hydrodynamics and chemical reactions that occur in commercial and pilot FCC risers. The flow regime in the reactor determines each phase residence time and thereafter the kinetic conversion of the reaction. Thus the reaction kinetics formulation should be accomplished with a comprehensive hydrodynamic model in order to have accurate simulation of the process.

The key issue when dealing with FCC fluid dynamics is the slip velocity between the gas and solids fractions. Generally, for commercial units with large diameters it is a common practice that the gas–solids slip velocity is assumed to be close to the single particle terminal velocity and slip factors are relatively small, whereas according to the literature the slip velocity is always greater than the single particle terminal velocity [18–20]. The general idea of using the terminal velocity instead of the slip velocity is correct only in case of great gas velocities and low solids mass flow rates [18] and delivers slip factor values close to unity. In practice, the knowledge of the exact value for the slip velocity in commercial units provides more accurate results for the determination of the hydrodynamic characteristics of the process and finally for the kinetic parameters estimation (each phase residence time and the space velocity). Understanding the operation of small diameter units (such as pilot plant units), where slip effects become much more important and influence significantly the kinetic and hydrodynamic features of the cracking process, provides the knowledge for an integrated commercial unit simulation.

Regarding the kinetic aspects of the cracking process it is well accepted that the cracking reaction proceeds according to second-order rate kinetics [17,21,22]. Considering riser reactor conditions with concurrent plug flow of gas and solids phases, the final expression for the conversion of hydrocarbons during the FCC process would be of a form of Eq. (1) [21]:

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = k\phi(c)(100-x)^2, \tag{1}$$

where τ is the space time (catalyst hold-up/feed rate), x is the wt.% conversion of the hydrocarbons, c is the coke content wt.% on catalyst and $\phi(c)$ is the catalyst deactivation function expressed by Eq. (2) [23]:

$$\phi(c) = k_\mathrm{c} c^{1-1/b}. \tag{2}$$

The value of b in Eq. (2) indicates the catalyst decay constant and is found in literature to vary from 1/3 [23,24] to 1/4 [22] or even 1/6 [25]. The coke build-up rate function is supposed to be the same with the catalyst deactivation function [21,23]. Thus for the rate of coke build-up, an equation of the form [23] applies:

$$\frac{\mathrm{d}u}{\mathrm{d}\tau} = \frac{k_\mathrm{c}}{b}\left(\frac{C}{O}\right)^{1/b-1} u^{1-1/b}. \tag{3}$$

Here u is the coke yield in wt.% on fresh feed. In this work the coke build-up rate is assumed to follow the conversion correlation and the total coke produced is assumed to be equal to the catalytic coke described in Eq. (3). Thus, coke selectivity is not a function of the catalyst to oil ratio or coke on regenerated catalyst, but is only a function of temperature, catalyst and feed properties.

When appropriate transformations and substitutions are made to Eq. (1) the final correlation for the mixture reaction conversion can be expressed as follows:

$$\frac{x}{100-x} = K\frac{C}{O}\, t_\mathrm{c}^b = K\frac{1}{\mathrm{WHSV}}\, t_\mathrm{c}^n, \quad n = b-1. \tag{4}$$

In Eq. (4) it is clear that the product of conversion times space velocity is an exponential function of contact time for given feed and catalyst properties and constant reactor temperature. A strategy often applied for validating the correct operation of an FCC unit, or the correctness of the hydrodynamic regime assumed, is plotting Eq. (4) on logarithmic scale.

$$\ln\left(\frac{x}{100-x}\,\mathrm{WHSV}\right) = \ln(K) + n\ln(t_\mathrm{c}). \tag{5}$$

The left hand side of Eq. (5) is a linear function of $\ln(t_\mathrm{c})$ for given catalyst, feed and constant temperature. The slope of this linear dependence would be n, or better $b-1$, the power in coking rate expression. The same correlation is assumed to apply for the coke yield and the same exponential dependence should be verified. Under this concept Eq. (5) for the coke yield would be an expression like Eq. (6):

$$\ln(\text{coke wt.\%}\cdot\mathrm{WHSV}) = \ln(k_\mathrm{c}/b) + n\ln(t_\mathrm{c}). \tag{6}$$

The development of the kinetic theory for the cracking process is identical for both the commercial and the pilot unit, since the occurring cracking reactions are exactly the same. Differences in the hydrodynamic attributes of each reactor correspond only in different kinetic parameters estimations.

3. Neural networks theory

3.1. Neural networks description

ANNs consist of a large number of simple computing elements, called nodes or neurons, which are arranged in layers. There are three types of layers: the input layer, the output layer and the hidden layer. The number of hidden layers varies from network to network and in some cases there is no hidden layer. Typically one hidden layer has been found to be sufficient in most applications [3,26].

Nodes of one layer are connected to the nodes of the next layers. A real valued number called ‘connection weight’ or simply ‘weight’ is associated with each connection. The role of the weights is to modify the signal carried from one node to the other and either enhance or diminish the influence of the specific connection.

Nodes in the input layer are not associated with any calculations. They act as distribution nodes. The outputs from the output layer represent the network's predicted outputs. The function of any node in the hidden and output layer is to receive a number of inputs from the previous layer, sum the weighted inputs plus the bias, non-linearly transform the sum via an activation function (i.e. Tanh or Logistic) and finally broadcast the output either to nodes of the next layer or to the environment [1].

There exist many network architectures [27]. In this work multi-layer perceptron (MLP) networks, with hyperbolic tangent (tanh) as activation function of hidden nodes and linear transformation as activation function of output nodes, are considered. In this type of network the nodes of one layer are fully or partially connected only to the nodes of the next layer of the network and there is no feedback of the output signals (feed-forward) [3].

The steps involved in every effort to build a functioning ANN model of an industrial process are: (a) data collection; (b) data preprocessing; (c) model selection; and (d) training and validation. The ANNs used in this study were trained using the EBP (Error-Back-Propagation) algorithm and the determination of optimum number of nodes in the hidden layer was carried out by a trial and error procedure based on cross validation. In the cross validation method, various network architectures (which are produced by changing the number of nodes in the hidden layer) are constructed. Each one of them is trained several times with different initial values of weights in order to find the combination of weights, which produces minimum output error for a validation data set. In the end, all network architectures are compared and the one with the minimum error is selected as optimum [1].

3.2. Hybrid modeling

In this paper two different hybrid-modeling schemes are implemented. They both belong to the design semi parametric approaches [12], namely they both try to correct the inaccuracy of an existing model, which is considered to contain our prior knowledge regarding to the process under consideration, by using a neural network, which is trained to compensate for this inaccuracy.

Let us assume that the phenomenological or empirical model in hand is described in general by the following functional form:

$$y = f(\vec{x},\ \vec{c}), \tag{7}$$

$$\vec{c} = g(\vec{a}), \tag{8}$$

where $\vec{x}$ is the vector of the variables that the model uses to predict the variable y, $\vec{c}$ is the vector of the constants which are included in the functional form and $\vec{a}$ is the vector expressing the assumptions made during the construction of the model.

Even if we assume that the particular functional form expresses the influence of the variable vector in every detail, there are still some limitations in the general implementation of the model, due to the assumptions made. These often influence the estimation of the constant vector (Eq. (8)). This can lead to inaccurate predictions, when the model is implemented in cases, where some of the assumptions do not any more apply and therefore the values of the constant vector are not any more appropriate. Sometimes this problem can be rather easily overcome by simply recalculating the constant vector for the new conditions. This presupposes that the relation between the constant vector and the assumptions made is either known or its functional form is predefined and all it remains is adapting the constants of this new relation, for instance by using regression analysis. But in cases, where our knowledge for this relation is very limited, the completion of the model based on Eq. (1) by a neural model is a very promising alternative.

In principle this can be done in a parallel (Figs. 1 and 2) or in a serial approach (Fig. 3). In the parallel approach the first step is to estimate the difference D_i between the prediction P_i of the model based on our prior knowledge (P-K model) and the experimental measurement M_i for all the i-measurements available:

$$D_i = M_i - P_i. \tag{9}$$

Referring to Fig. 1 this difference is the target residual for the neural model, namely the neural model is trained to calculate this residual.

The second step is to define the variables that will be used to perform this task. All or some of the available variables can be used as input variables (Variables Set-2). If we are certain that all of the information contained in the variables used in the P-K model (vector $\vec{x}$, Variables Set-1) is captured in it, then these variables are redundant for the neural network training procedure and can be left aside. An Entropy of Information analysis may be applied [28,29] to certify such a decision (see Appendix A). Furthermore, if different scenarios concerning the variables used to predict the target residual are applied, then the combination of the created

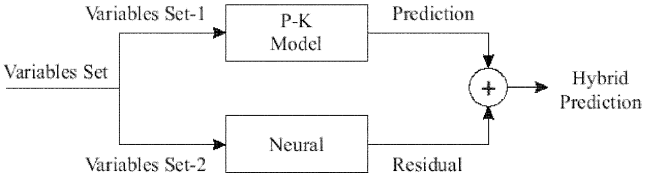

Fig. 1. A hybrid model according to the parallel design approach. The neural model is trained on the difference (residual) between the prediction of a prior knowledge model (P-K model) and the experimental prediction.

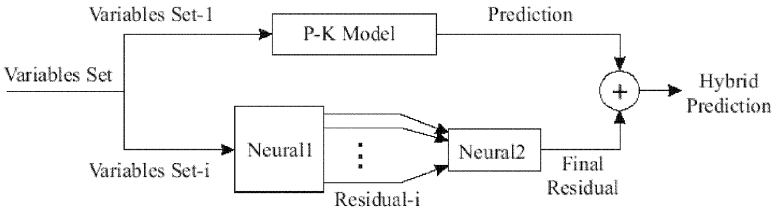

Fig. 2. A hybrid model according to the parallel approach combining more than one neural models to calculate the residual.

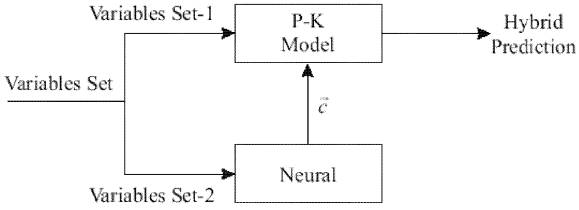

Fig. 3. A hybrid model according to the serial approach. The neural model is trained to approximate the new values of $\vec{c}$, which will be used from the P-K model to produce a more accurate prediction (hybrid prediction).

neural models is proposed instead of choosing the best of all cases (Fig. 1). To be more exact, for each one of the j-scenarios a neural model is obtained following the method described in Section 3.1. This model is able to calculate as output a j-residual ($\vec{R}_j$), which is an approximation of the true residual ($\vec{D}_j$). The vectors $\vec{R}_j$, $\vec{D}_j$ consist of all the i-measurements. These j-neural models are not only different in the variable set they are using, but also probably in the number of the neurons in the hidden layer and the number of the epochs that were used for their development. These last two training parameters depend on the algorithm used to find the optimum neural network architecture for each one of the j-scenarios. Because of the difference in the architectures of these j-neural models, it is very likely that each one tries to approximate the true residual ($\vec{D}$) from a different point of view. This makes a combination of all the cases more beneficial than choosing the best-case scenario. For the combination of neural models a neural network combiner is used. Neural network combiners perform among the best possible combiners although it complicates the procedure more than other simple linear combiners.

So the outputs of these j-neural models are then used as inputs to a new neural network, which is trained to achieve a better approximation of the true residual ($\vec{D}$). Depending on the number of scenarios applied, some or all of them may be combined. The selection can also be supported by an Entropy of Information analysis [29].

In both cases of this parallel design approach (Figs. 1 and 2) the residual ($\vec{R}$), which is calculated by the neural part of the hybrid model, is added to the prediction of the P-K model ($\vec{P}$) to give the final prediction ($\vec{F}$) of the hybrid model:

$$\vec{F} = \vec{P} + \vec{R}. \tag{10}$$

All the vectors referred in Eq. (10) have the i-measurements as arguments.

In the serial design approach the neural model part is used to calculate the new values of the constant vector and more specifically to approximate the function (g) in Eq. (2) (Fig. 3). It is assumed that in Eq. (2) the vector of the assumptions ($\vec{a}$) refers to variables, which were kept constant or were ignored during the construction of the P-K model and therefore their influence has not been adequately evaluated. Because of variation of these variables in the process under consideration, the constant vector ($\vec{c}$) needs to be re-estimated. According to this, the first step in this approach will be to solve Eq. (1) and obtain the appropriate values of the constant vector ($\vec{c}_{ii}$) by replacing the actual experimental values in the variables y, $\vec{x}$ for all the i-measurements. These are the target values, which the neural model will try to approximate. The second step is similar to the one made for the parallel approach, that is the selection of the variables used as input variables for the neural network during the training procedure (Variables Set-2 referring to Fig. 3). Once again different scenarios can be used and the Entropy of Information analysis can be a valuable tool for the selection of variables as well as for the choice of scenarios that will be combined to produce the final prediction of ($\vec{c}$).

4. The pilot plant model formulation—application to the commercial unit

The model developed in this study was based on the experiments performed in CPERI FCC pilot plant unit. An analytical description of the CPERI FCC pilot plant and the formulation of the model developed to simulate it can be found in literature [15,17]. For the scale-up of the pilot model to the actual commercial conditions (different geometry, higher flow rates, variable feed and catalyst properties etc.) adjustments were applied to the model for the better representation of the industrial process.

The main difference between the pilot and the commercial unit arises from the differences in geometry. In small diameter FCC risers gas and solids rise with different velocities, corresponding to different residence

times for each phase. This 'slip phenomenon' can be described by the 'slip factor' concept used to ascribe the back-mixing in the riser reactor. The slip factor stands for t_s/t_g, or u_g/u_s [19,20] with g and s representing the gas and solids phase, respectively. The slip factor is a feature rarely reported in the scientific literature and empirical correlations for the determination of slip effects in CFB risers are not often published. A widely used empirical correlation for the slip factor estimation in circulating fluidized bed risers was proposed by Patience et al. [19]. They suggested that the 'slip phenomenon' is independent of the gas properties and solids characteristics at gas velocities much greater than the terminal velocity, but is a function of the riser diameter, gas volumetric flow and solids terminal velocity. Later, Pugsley and Berruti [30] noted that the correlation of Patience et al. [19] over-predicts the average solids holdup and proposed an improved expression (Eq. (11)):

$$y = 1 + \frac{5.6}{Fr^2} + 0.47 Fr_t^{0.41}. \tag{11}$$

In Eq. (11) Fr and Fr_t are the Froude numbers for the superficial gas and terminal velocity, respectively and y the slip factor.

The comparison of the slip factor dependence on superficial gas velocity as it is estimated from Eq. (11), for the fully developed flow region of the pilot and the commercial unit is shown in Fig. 4. The inverse proportion of slip factor with superficial gas velocity is obvious in Fig. 4 and is consistent to Patience et al. [19] and Pugsley and Berruti [30] observations.

With the average slip factor known, the average reactor voidage can be easily calculated. Commonly, for commercial risers, by knowing the average reactor voidage, the total pressure drop can be directly estimated, since the static head of solids is the dominant pressure gradient. However, including every pressure gradient in the pressure balance analysis should provide more accurate description of the actual flow regime for

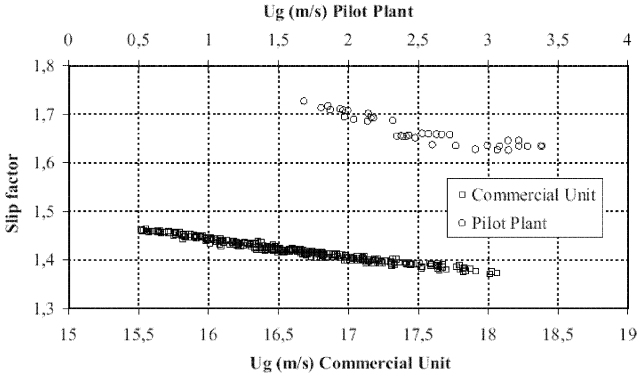

Fig. 4. Slip factor dependence on superficial gas velocity for the commercial and the pilot unit (fully developed flow region).

both the pilot and the commercial unit. For this analysis all pressure gradients must be taken into account and the following expression is valid [17]:

$$\Delta P = \Delta P_{fg} + \Delta P_{fs} + \Delta P_{acc} + \varepsilon \rho_g g \Delta z + (1-\varepsilon)\rho_p g \Delta z, \tag{12}$$

where ΔP_{fg} is the gas-wall frictional pressure drop; ΔP_{fs}, solids-wall frictional pressure drop; ΔP_{acc}, pressure drop due to solids acceleration.

In the commercial unit the pressure drop measurements are used for the estimation of the dense zone height, where the flow regime is assumed to follow the 'Dense Suspension Upflow' reported by Grace et al. [31]. For this dense region the feedstock is assumed unconverted (in terms of molar expansion) and the total volumetric flow and the superficial velocity are significantly lower. The height of this dense region is computed around 5% of the total riser height.

Applying the 'pilot plant model' to the commercial unit after implementing the adjustments described above was a complicated task. The commercial unit operates under non-steady feed and catalyst properties and the consistency of the kinetic model can not be verified just by plotting Eq. (5). The reactor temperature varies within 980 and 995 °F, so the constant temperature hypothesis used for the pilot unit model development is incorrect in the case of the commercial unit. Changes in feedstock and catalyst supply correspond to different pre-exponential factors in the Arrhenius type formulation of the K constant in Eq. (5) and differences in temperature correspond to different values of K. In order to examine the consistency of the kinetic model applied to the commercial unit Eq. (5) should be reordered to the form of Eq. (13):

$$\ln\left(\frac{x}{100-x}\,\mathrm{WHSV}\right) - \ln(k_o \exp(-E/RT)) = n \ln(t_c), \tag{13}$$

where the pre-exponential factor k_o and the activation energy E of the cracking reaction should be estimated via the hybrid scale-up procedure proposed. For the implementation of the model to the commercial unit the values of activation energy E and the catalyst decay constant n were estimated 24.800 Btu/lbmol/R and 0.72, respectively (Fig. 5).

Similarly for the coke yield prediction Eq. (6) can be rewritten for the commercial unit:

$$\ln(\text{coke wt.\%}\cdot \mathrm{WHSV}) - \ln(k_{co} \exp(-E/RT)) = n \ln(t_c). \tag{14}$$

The applicability of the pilot plant model to the commercial data for the coke yield prediction appears excellent. The coke build-up decay constant for the commercial unit is now calculated at the value of −1

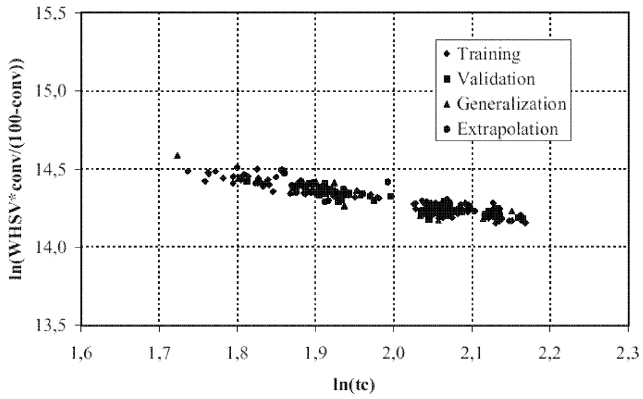

Fig. 5. Experimental data fit with model results for kinetic conversion (commercial unit).

and the activation energy for the coke production is 1660 Btu/lbmol/R (Fig. 6).

The agreement of the pilot and commercial units appears quite satisfactory and the proposed hydrodynamic scheme represents both units correctly. The scatter of the commercial model outputs should be refined via a neural network combinational procedure, in order to invoke the influence of variable feed and catalyst properties and scale-up factors for the transition from the 'pilot plant model' to the industrial reality.

5. The neural scale-up (hybrid model development)

In this paper the hybrid modeling schemes mentioned in Section 3 were used for the needs of the scaling up of the pilot plant model already described in Section 2. This model takes the place of the P-K model in Figs. 1–3. It is able to predict the conversion (wt.%) of the FCC process using some of the operational variables, which are the vector $\vec{x}$ referring to the analysis made in Section 3. The assumptions (vector $\vec{a}$) of the model are that the feed and catalyst properties of the unit are constant.

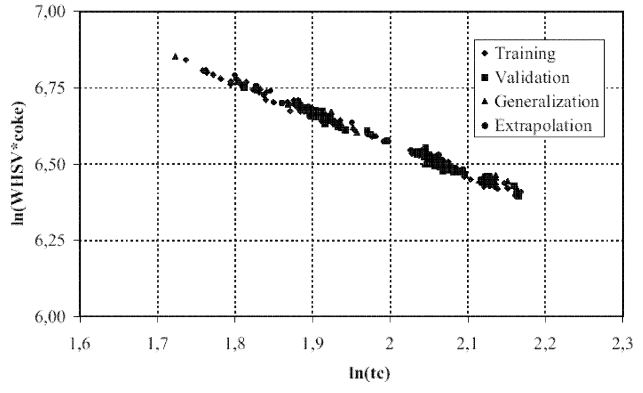

Fig. 6. Experimental data fit with model results for coke (commercial unit).

These assumptions are included in the estimation of the constant k_o (Eq. (13)), that is the ($\vec{c}$) vector in our case. When we apply this model to the industrial database, a difference between the predictions of the model and the experimental measurements is being remarked. Our hypothesis is that the main reason for this difference is the variation in the feed and catalyst properties as well as the scale up factors.

The variables that are included in the database are shown in Table 1a. For the needs of the hybrid modeling some additional 'variables' must be constructed. Specifically for the parallel design approach the input variables that concern feed or catalyst properties are shown to the neural network as differences (DV_j), which are defined in Eq. (15):

$$DV_j = V_j - V_{pj}. \tag{15}$$

where V_j is the actual value of the property j in the industrial database, V_{pj} is the constant value of the property j during the construction of the pilot plant model and DV_j is the difference between these two values. Furthermore, the neural part tries to approximate the difference between the experimental conversion and the conversion calculated from the pilot plant model based on some or all of the above input variables. For this to be done these differences must also be calculated for all the runs available (see Eq. (9)). The set of the target output variables of the neural network consists of these differences.

In the serial approach the neural part of the hybrid has the task to calculate the values of the constant k_o (Eq. (13)) for every run available. To be more precise it is trained to calculate the $\ln k_o$ value. Because of the change in feed and catalyst properties in every run k_o has not anymore a constant value as it was assumed during the development of the pilot plant model, where the feed and catalyst properties were kept constant. The values of the $\ln k_o$, which the neural part must approximate were calculated using for all the variables included in Eq. (13) the respective experimental values in the database. In the serial design approach no transformation like the one made in Eq. (15) is performed for the input variables used by the neural part.

As we have described in Section 3 the next important step for the training of the neural part of the hybrids is to define which input variables will be used. Referring to the pilot plant model, we can separate the variables shown in Table 1b in three categories: variables that were kept constant during the construction of the pilot plant model but they are included in the model (Variables 1–4), variables that were kept constant during the construction of the pilot plant model but they are not included in it (Variables 10–13) and variables that were included in the pilot plant model and their variation is believed to be satisfactorily explained by it (Variables 5–9).

Table 1a
The variables that were used in each one of the hybrid models

No.	Variables	Pilot plant	PHybrid1	PHybrid2	PHybrid3	PHybrid4	SHybrid1	SHybrid2
1	Specific gravity	U	U	NU	U	U	U	U
2	MeABP	U	U	NU	U	U	U	U
3	APS	U	U	NU	NU	NU	U	U
4	ABD	U	U	NU	NU	NU	U	U
5	CCR	U	NU	NU	U	NU	NU	U
6	Reactor temperature	U	NU	NU	NU	NU	NU	NU
7	Reactor pressure	U	NU	NU	NU	NU	NU	NU
8	Feed rate	U	NU	NU	U	NU	NU	U
9	Riser inject steam	U	NU	NU	NU	NU	NU	NU
10	Sulfur	NU	U	U	NU	U	U	U
11	Basic N2	NU	U	U	NU	U	U	U
12	RI	NU	U	U	U	U	U	U
13	MAT	NU	U	U	U	U	U	U

U means that the variable is used in the model and NU that the variable was not used.

For the needs of the parallel design approach we have built the four scenarios, which are presented in Table 1b as far as the neural part of the hybrid is concerned. Specifically:

- *PHybrid1* uses only those variables, which are referring to feed and catalyst properties. It is assumed that the pilot plant model does not totally capture the influence of the variables 1–4, although they are included in it (see Section 2), since the effect of feed properties and catalyst quality is included in the pilot plant model only from their hydrodynamic perspective and their actual influence on the cracking reaction may be underestimated, when apply the pilot plant model to the commercial data. So they are used together with the variables 10–13 to predict the residual.
- *PHybrid2* uses only the variables 10–13. It assumes that the pilot plant model satisfactorily expresses the influence of the variables 1–4.
- *PHybrid3* uses the Entropy of Information Analysis (results of this analysis are presented in Appendix A) considering all the 13 variables as possible input variables for the neural model. The six best variables that this analysis proposes are selected (Table A2(a)). Two of the operational variables (feed rate and CCR) are included in these six variables. This can have the following theoretical basis. Inaccuracies in the application of the empirical equation of Pugsley and Berruti [30] to the commercial unit and in the pressure balance formulation, could influence the prediction of slip effects on the commercial unit. Thus the flow rates of gas–oil and catalyst were included in the hybrid model, in order to examine if better predictions are accomplished.

Table 1b
The limits of the available variables in the training and the extrapolation set

Variable	Units	Training set		Extrapolation set	
		Min	Max	Min	Max
Specific gravity (SG)		0.900	0.921	0.901	0.917
Mean average boiling point (MeABP)	°C	437.8	469.4	438.3	463.3
Average particle size (APS)	microns	72.0	84.0	73.8	83.0
Apparent bulk density (ABD)	G/ml	0.87	0.94	0.88	0.93
Catalyst circulation rate (CCR)	m^3/h	18.3	21.5	17.8	22.0
Reactor temperature	°C	529.5	535.9	528.8	536.0
Reactor pressure	kp/cm^2	2.2	2.5	2.2	2.5
Feed rate	tn/h	239.9	280.5	235.8	283.6
Riser inject steam	kg/h	2852	3300	2898	3322
Sulfur (S)	wt.%	0.29	1.85	0.32	1.46
Basic N2	wppm	104	381	112	241
Refractive index (RI)		1.48331	1.49180	1.48367	1.49060
Micro activity test (MAT)		66	75	66	73
Conversion	wt.%	69.06	77.81	70.53	77.81

- *PHybrid4* uses the Entropy of Information Analysis considering this time only the variables referring to feed and catalyst properties. The six best variables that this analysis proposes are selected (Table A2(b)).

These four scenarios are used for the combination procedure described in Section 3 (Fig. 1). The hybrid model that is produced according to this scheme will be referred as PHybrid5 from now on.

For the needs of the serial approach we have considered the following two scenarios as far as the neural part of the hybrid is concerned (see Table 1a):

- SHybrid1 uses the same variables set as PHybrid1.
- SHybrid2 uses additionally the two operating variables that were indicated by the Entropy of Information Analysis, namely the feed rate and the CCR.

Finally a hybrid model identical to SHybrid2 was created to predict the coke yield. This model will be further named CokeSHybrid in this paper. It uses the same variables that the SHybrid2 uses for the neural part and also has the same architecture. The difference for the neural part is that it now tries to approximate the $\ln k_{co}$ value, which is now referred to Eq. (14). This equation using the new adapted values of $\ln k_{co}$ for every run predicts the coke yield. The data that were used for the coke yield concern the same runs that were mentioned above as well as their partition in the four data sets. This model was created to test the applicability of the method in another important parameter of the FCC unit such as the coke yield, which is however oscillating in a much smaller range compared to the conversion of the unit.

The data sets for the models development was based upon industrial data provided by the Aspropyrgos Refinery of Hellenic Petroleum S.A. (Athens). The data set was collected every 1 day for a period of 15 months. In selecting data for model building, however, it is important to ensure that it represents normal operating states to avoid spurious predictions from unusual conditions. So, blocks of data corresponding to process faults were excluded from the study. Also, outliers that may have been caused by some measurement errors were removed. A simple outlier detecting method was followed, where any observation that differ more that three standard deviations from the mean is removed from the set [1]. As a result, a set of 308 observations, representative of various operating conditions and a broad range of the input variables, were used for the development of the models.

The available 308 runs were separated in four data sets. The training and the validation set (178 and 50 runs, respectively) were used for the needs of the training algorithm (Section 3). The generalization set (50 runs) is used for the testing of the models, when all the variables interpolate. The extrapolation set (30 runs) is used for the testing of the models, when some of the operational variables extrapolate. We have let only the operational variables to extrapolate for two reasons. The first is that their effect is supposed to be mainly captured by the pilot plant model and we wanted to see how far this is helping the extrapolation abilities of the hybrid model, in which this prior knowledge is included. The second is that if we had let more variables to extrapolate that would lead to a greater extrapolation set and the other sets would have to be reduced, that could seriously affect the training procedure of the neural networks. The maximum and minimum values of all the variables for all the sets are shown in Table 1b.

After this partition of the sets, the values of the variables were normalized. Both for the input and output variables a linear normalization based on the training set was used. After the normalization all the variables have values belonging to the interval $[-1, 1]$.

In addition to the seven hybrid models presented above a typical 'black box' neural model using all the variables was created, so that the results from the hybrid modeling approaches could be compared to a more typical method. This neural model is referred from now on as simple neural (SN).

During the training procedure of each neural model it was paid attention to the fact that all the neural networks must have less than 50 weights, including the weights of the biases. This limitation arises from the number of available industrial runs. Specifically, we have tried to experiment with 'data to parameters ratios' greater than 3.5. This has defined the upper limit for the number of weights in the hidden layer during the training procedure. Only the case of one hidden layer was tried, since it is proved that one hidden layer is enough for approximating continuous functions [26]. In all cases we have achieved ratios greater than 4 for the 'winner' neural models except from the case of the SN, which is due to the fact of the 13 input variables (Table 2). Furthermore, no neural model was trained for more than 20,000 epochs.

The structure of each of the eight neural models that were trained using this procedure is presented in Table 2.

Table 2
The structure of the 'winner' neural network models

Neural model	Structure	Weights	Ratio
Simple neural	13–3–1	46	3.9
PHybrid1	8–4–1	41	4.3
PHybrid2	4–7–1	42	4.2
PHybrid3	6–5–1	41	4.3
PHybrid4	6–5–1	41	4.3
PHybrid5	4–5–1	31	5.7
SHybrid1	8–4–1	41	4.3
SHybrid2	10–3–1	37	4.8

In the column named 'structure' the first number indicates the number of the input neurons, which is predefined by the choice of variables used, the second number indicates the number of neurons in the hidden layer for the 'winner' neural model and the third number the number of neurons in the output layer, which is always 1. This is so, because the target variable was in the SN approach the conversion itself, in the PHybrid1 to PHybrid5 approaches the difference between the experimental conversion and the predicted from the pilot plant model and finally in the SHybrid1 and SHybrid2 approaches the constant $\ln k_o$ of the pilot plant model.

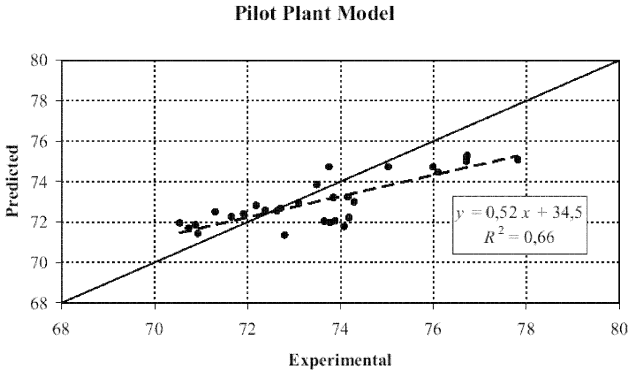

Fig. 8. Scatterplot of measured against predicted values of conversion using the pilot plant model (extrapolation set).

6. Results and discussion

In this section we describe at first the results that all the models give when they attempt to predict the conversion of the commercial unit. At the end of this section we also verify the applicability of the hybrid approach to the prediction of the coke yield of the unit.

We use the generalization and the extrapolation set for the comparison between the models, because these two sets have in no way affected the training procedure of the neural parts in the hybrid models. Therefore, in these two sets the true abilities of the models are shown.

In all the figures the x-axis has the experimental values and the y-axis the predictions. We use two lines to show the success of the prediction. The one is the line with equation $y = ax + b$ with $a = 1$ and $b = 0$, on which all the data of an ideal model should lay. The other line is the line that best fits on the data of the scatter plot and it is obtained with regression analysis based on the minimization of the squared errors.

The correlation factor of this line is also presented (R^2). The closer to 1 this factor is and the closer the coefficients of the line to 1 and 0, respectively, are the better the model is.

In Figs. 7 and 8 we present the results for the pilot plant model. The results for the other two sets (training and validation set) are similar, since this model has not in any way used the commercial data for its development. These results are considered satisfactory if one takes into account the constancy of the pilot plant experimental runs and the multivariation of an actual commercial unit operation. This fact supports the assumption that this model can be successfully used as the prior knowledge model, which with the assistance of a neural part will result to an integrated hybrid model with excellent results.

In Figs. 9 and 10 the results of the simple neural model are presented. Compared to the pilot plant model the Simple Neural Network approached gives better results as expected, since it is trained with data of the specific commercial unit.

The results for the best of the four scenarios used according to the parallel design approach are presented in the Figs. 11 and 12. These figures verify the statement made before about the excellence of the results of the hybrid modeling. Much more if we take into account that we have to deal with real industrial and therefore 'noisy' data. In Figs. 13 and 14 the results for the best of the two scenarios used in the serial design approach are presented. The two approaches do not seem to have many differences as far as the accuracy is concerned, although the parallel design approach seems to general-

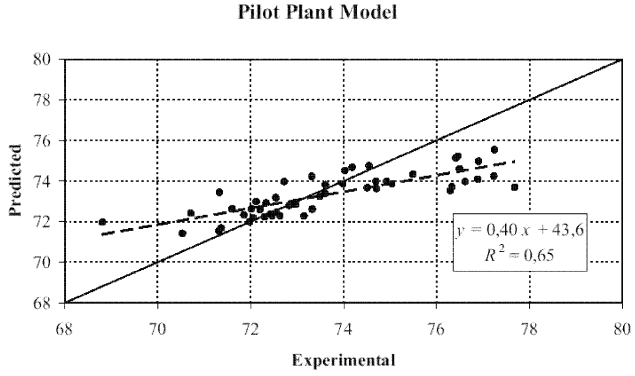

Fig. 7. Scatterplot of measured against predicted values of conversion using the pilot plant model (generalization set).

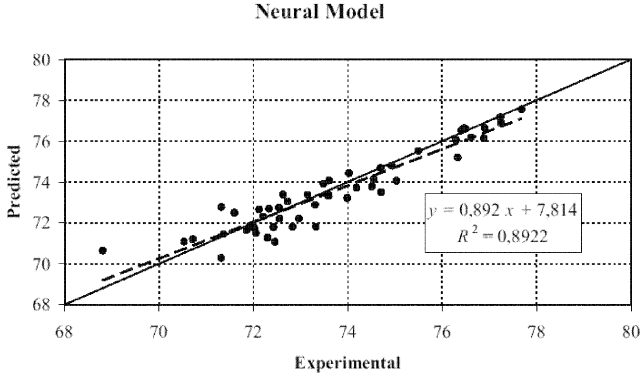

Fig. 9. Scatterplot of measured against predicted values of conversion using the simple neural model (generalization set).

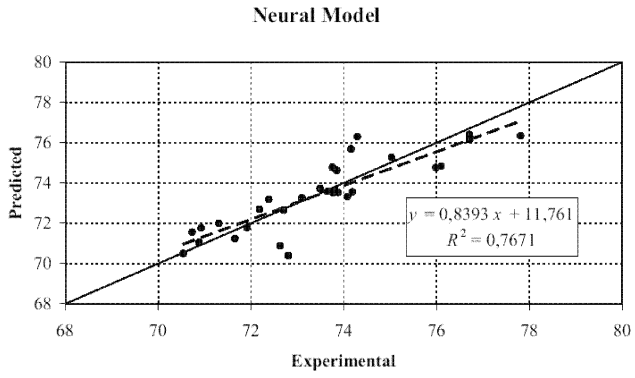

Fig. 10. Scatterplot of measured against predicted values of conversion using the simple neural model (extrapolation set).

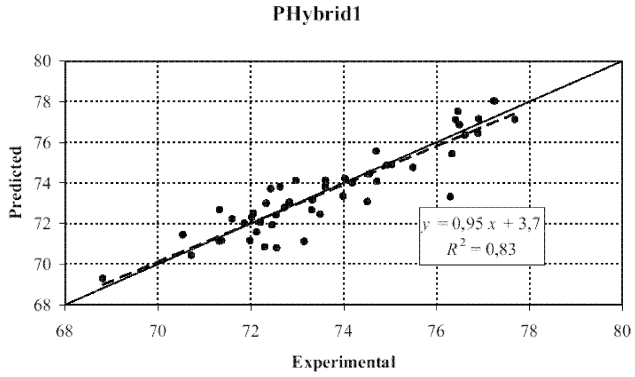

Fig. 11. Scatterplot of measured against predicted values of conversion using the best of the parallel hybrid models (generalization set).

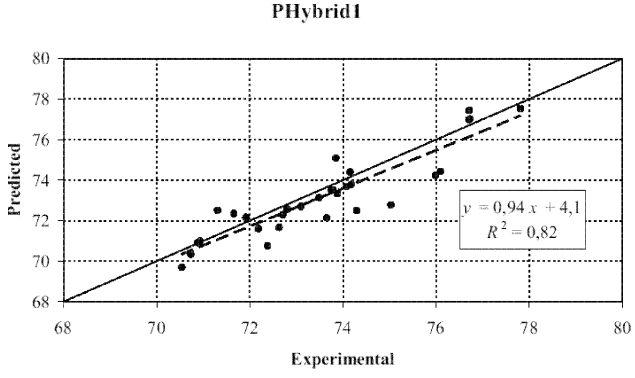

Fig. 12. Scatterplot of measured against predicted values of conversion using the best of the parallel hybrid models (extrapolation set).

ize a little better as it can be seen in the coefficients of the 'best fit' line.

In Figs. 15 and 16 we present the results for the parallel design approach when the best models for each one of the four scenarios are combined. Once again the accuracy of this hybrid is excellent. Compared to the best of the four scenarios this method seems to give more stable results, since its 'best fit' line has similar coefficients but with greater correlation factor (0.91–0.83).

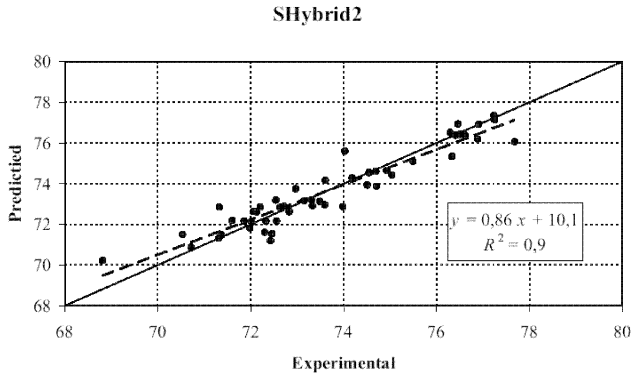

Fig. 13. Scatterplot of measured against predicted values of conversion using the best of the serial hybrid models (generalization set).

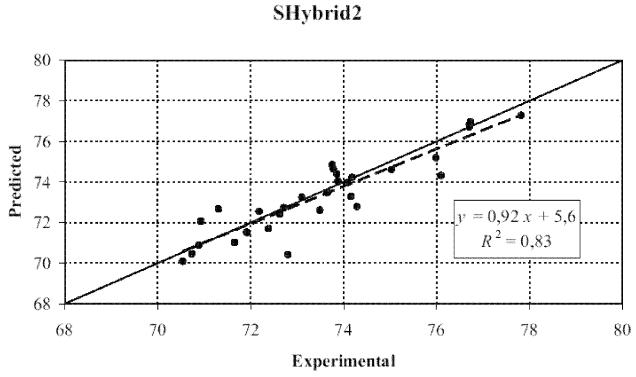

Fig. 14. Scatterplot of measured against predicted values of conversion using the best of the serial hybrid models (extrapolation set).

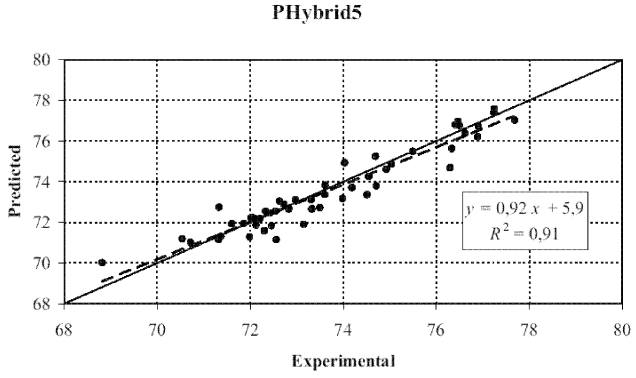

Fig. 15. Scatterplot of measured against predicted values of conversion using the combination of the parallel hybrid models (generalization set).

Comparing the results of the standard 'black box' neural network approach to the best hybrid modeling approaches we can see that the accuracy of the standard neural model is similar to the hybrids in the generalization set, but as expected it starts getting worse in the extrapolation set. On the other hand, the hybrids seem to be more stable in the results they give for the two sets. This indicates a trend, which would be clearer, if we were able to implement a more intense extrapolation.

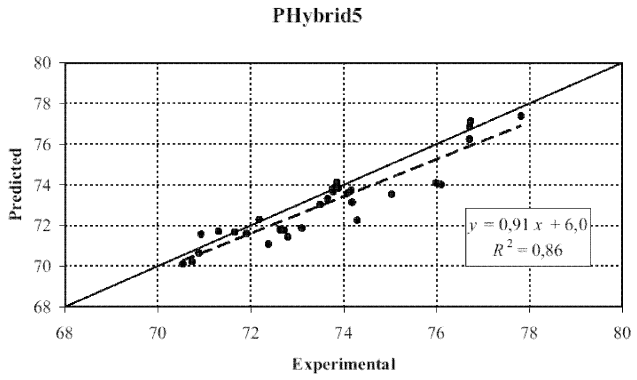

Fig. 16. Scatterplot of measured against predicted values of conversion using the combination of the parallel hybrid models (extrapolation set).

Unfortunately, this is a limitation arising from the fact that we are working with industrial data.

In Tables 3–6 some of the most commonly used statistical pointers are presented for all the sets used and for all the models created. In these tables the R^2 and A, B are the correlation factor and the coefficients concerning the 'best fit' line as described above. MSE is the Mean Square Error, ARE is the Average Relative Error (%) and MaxRE is the Maximum Relative Error (%) (Table 1a).

As far as the different hybrids in the parallel design approach are concerned, we can see in these tables that although the approaches using the Entropy of Information analysis for defining the input variables did not result to more accurate hybrids, they did give us the chance to have more than one scenarios, which after their combination resulted to a better model. Since the serial and the design approaches do not seem to have many differences in their final results, it can be assumed that this combination scheme proposed could also be applicable in a serial design approach.

Furthermore, all the hybrids have an ARE less than 1% in the generalization set and the best of them have the same level of accuracy in the extrapolation set too, which is a significant improvement compared to using only the pilot plant model. An exception to this behavior is the PHybrid1 model, which seems to suffer from overfitting problems (it gives a MSE of 0.1 and a ARE of 0.3 on training data but a MSE of 0.8 and ARE of 0.9 on the generalization data). However, the good generalization performance of the PHybrid5 model (which combines the four hybrid models, including PHybrid1) shows that the combiner successfully eliminates the disadvantages of the PHybrid1 model. From this point of view the effort of scaling up this pilot plant model to the commercial unit using the concept of hybrid modeling has been proved successful.

The differences between the hybrids are few, but the overall best model seems to be the one that combines the four scenarios in the parallel design approach. We must point out the fact, that among the best models seem to be one model (namely the SHybrid2), which uses two of the operational variables (feed rate and CCR) to fit the pilot plant model on the industrial data. This can be an indication that the hydrodynamic description of sleep effects in the commercial unit and the assumptions used in the commercial model formulation do not describe the actual industrial reality with the desired accuracy.

We must also point out the fact that the superiority of the hybrid models compared to a simple 'black box' neural network approach should not be only restricted to the level of the accuracy of the predictions, but also the aspect of interpretability and stability of the models should also play an important role.

At last in 17–20 the coke yield prediction for the pilot plant model and one of the best hybrid models (SHybrid2), described in Section 5, are presented. The predictions of the pilot plant model (Figs. 17 and 18) were extremely good, if one takes into account the variability of feed and catalyst properties of the industrial data. The hybrid approach corrected the prediction average error by 0.3% for the generalization and extrapolation sets (Figs. 19 and 20). Considering the very low oscillation of the coke yield of the commercial unit and the accuracy of the pilot plant model prediction, the improvement of the hybrid approach appears

Table 3
Results of models for training data set

	Training					
	MSE	R^2 (trendline)	A (trendline)	B (trendline)	ARE (%)	MaxRE (%)
Pilot plant	2.1	0.62	0.37	45.9	1.5	4.7
Simple neural	0.3	0.92	0.92	5.5	0.6	2.3
PHybrid1	0.1	0.98	0.97	2.0	0.3	1.3
PHybrid2	0.5	0.88	0.88	8.9	0.7	4.0
PHybrid3	0.5	0.88	0.86	10.6	0.8	4.4
PHybrid4	0.4	0.91	0.89	8.4	0.7	3.2
PHybrid5	0.2	0.96	0.96	2.9	0.4	1.8
SHybrid1	0.2	0.96	0.94	4.1	0.5	1.8
SHybrid2	0.2	0.94	0.93	5.1	0.5	1.9

Table 4
Results of models for validation data set

	Validation					
	MSE	R^2 (trendline)	A (trendline)	B (trendline)	ARE (%)	MaxRE (%)
Pilot plant	1.1	0.60	0.42	42.3	1.1	3.9
Simple neural	0.4	0.87	0.85	10.8	0.7	2.2
PHybrid1	0.3	0.90	0.97	2.5	0.5	1.9
PHybrid2	0.4	0.85	0.86	10.6	0.7	1.8
PHybrid3	0.4	0.85	0.84	11.7	0.7	1.7
PHybrid4	0.3	0.87	0.86	9.9	0.6	2.1
PHybrid5	0.1	0.95	0.95	4.0	0.4	1.3
SHybrid1	0.3	0.88	0.90	7.4	0.6	2.0
SHybrid2	0.3	0.88	0.83	12.5	0.6	2.3

Table 5
Results of models for generalization data set

	Generalization					
	MSE	R^2 (trendline)	A (trendline)	B (trendline)	ARE (%)	MaxRE (%)
Pilot plant	2.0	0.65	0.40	43.6	1.4	5.1
Simple neural	0.5	0.89	0.89	7.8	0.7	2.7
PHybrid1	0.8	0.83	0.95	3.7	0.9	3.9
PHybrid2	0.7	0.83	0.86	10.2	0.9	2.9
PHybrid3	0.7	0.83	0.82	13.0	0.8	3.8
PHybrid4	0.6	0.87	0.80	14.5	0.8	3.8
PHybrid5	0.4	0.91	0.92	5.9	0.7	2.1
SHybrid1	0.6	0.87	0.88	9.0	0.8	2.5
SHybrid2	0.4	0.90	0.86	10.1	0.7	2.2

satisfactory. The fact that the behavior of this model yield is similar compared to the model predicting conversion, is an indication that the prediction of the other product yields (gasoline, LPG, etc.) can also be realized in the same way. Therefore, one of the tasks planned for the future is to develop and test such models, when commercial data will be available.

7. Conclusions

In this effort we have presented nine models concerning the prediction of the conversion of a FCC unit. One of them was created based on the experiments carried on a pilot plant, another one was created based only on the industrial data using a typical 'black box' neural network approach and the rest models were hybrid models. The hybrids consist of two parts: the pilot plant model part, which gives the basis for the prediction and the

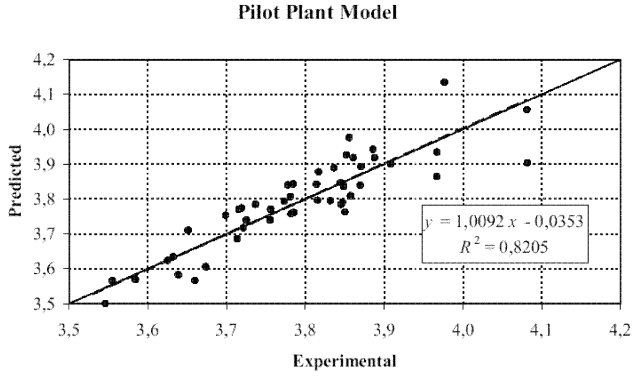

Fig. 17. Scatterplot of measured against predicted values of coke yield using the pilot plant model (generalization set).

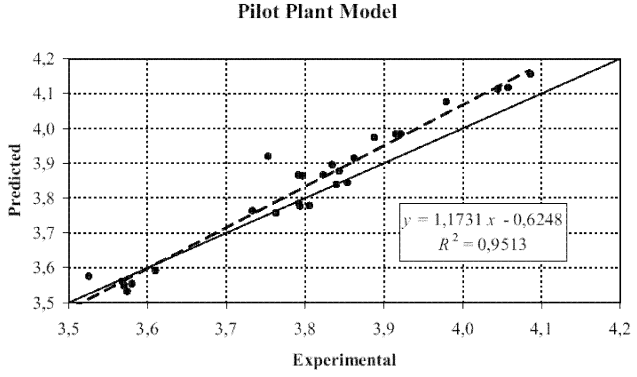

Fig. 18. Scatterplot of measured against predicted values of coke yield using the pilot plant model (extrapolation set).

Table 6
Results of models for extrapolation data set

	Extrapolation					
	MSE	R^2 (trendline)	A (trendline)	B (trendline)	ARE (%)	MaxRE (%)
Pilot plant	3.3	0.66	0.52	34.5	1.5	3.5
Simple neural	0.9	0.77	0.84	11.8	1.0	3.3
PHybrid1	1.8	0.82	0.94	4.1	1.0	3.0
PHybrid2	3.5	0.70	0.79	14.9	1.5	3.4
PHybrid3	2.9	0.79	0.90	6.2	1.3	4.5
PHybrid4	1.6	0.82	0.85	10.4	0.9	2.7
PHybrid5	1.6	0.86	0.91	6.0	0.9	2.7
SHybrid1	1.7	0.75	1.10	−7.5	1.4	3.9
SHybrid2	0.7	0.83	0.92	5.6	0.8	3.3

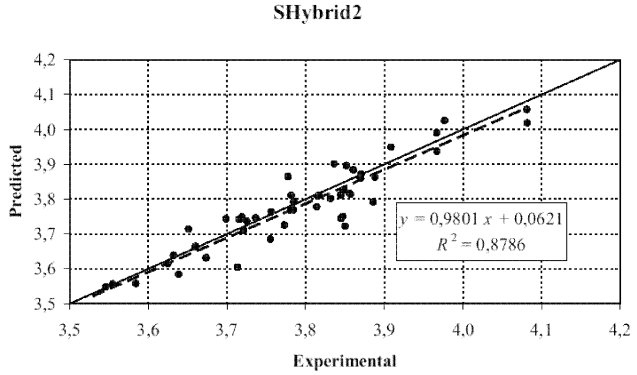

Fig. 19. Scatterplot of measured against predicted values of coke yield using the best of the serial hybrid models (generalization set).

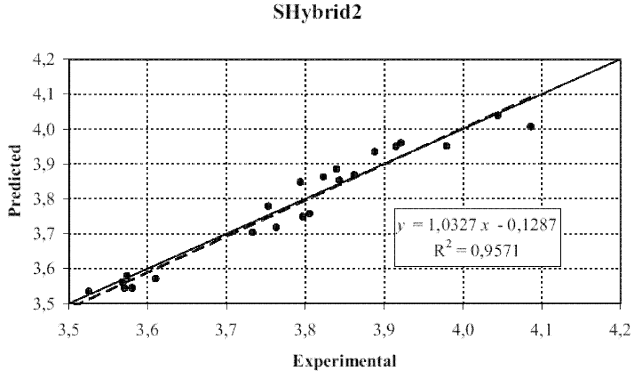

Fig. 20. Scatterplot of measured against predicted values of coke yield using the best of the serial hybrid models (extrapolation set).

neural model part, which tries to refine this prediction, so that the influence of variable feed and catalyst properties as well as scale up factors are taken into account. These two parts were combined in various hybrid-modeling schemes resulting to seven different hybrids.

The results of all the hybrids have shown that the method proposed here has been proven successful. The already satisfactory predictions of the pilot plant model are clearly refined since the hybrid models predict the conversion of the commercial unit with less than 1% average relative error, namely they have reached the limitations of the experimental error. Furthermore, the comparison between the hybrids has shown that particularly promising is the technique, which combines various scenarios as far as the neural part of the hybrid model is concerned.

The comparison between the standard 'black box' neural network approach and the hybrid modeling approach has shown that the best of the hybrid scenarios generalize slightly better and extrapolate clearly better. The reason for this is the pilot plant model part of the hybrids. The superiority of the hybrids for the extrapolation properties is assumed to become clearer, as the extrapolation becomes more intense. Our industrial database did not permit such an intense extrapolation, but this trend has become clear, even under the mild extrapolation conditions that were applied.

Finally, it is important to be noticed that the advantages of the hybrid models compared to the 'black box' neural networks do not only concern the accuracy of the predictions but also the interpretability of the model, which is a crucial point when the created model is used for the study of the behavior of the process and for optimization purposes.

Appendix A: Entropy of information analysis

Let X be a random variable with a set of s possible outcomes $\{x_1, x_2, \ldots, x_s\}$. If N is the number of cases in a data set that we have for this variable, and N_i is the number of cases, in which $X = x_i$, then the probability of

this event is calculated by:

$$p_i = P(X = x_i) = \frac{N_i}{N}, \tag{A1}$$

and the entropy $H(X)$ of the variable X is defined by:

$$H(X) = \begin{cases} \sum_i p_i \ln(1/p_i) & \text{if } p_i \neq 0 \\ 0 & \text{if } p_i = 0 \end{cases}. \tag{A2}$$

The conditional probability p_{ij}, namely the probability of X taking the value x_i while simultaneously another variable Y is taking the value y_j, is defined by:

$$p_{ij} = P(X = x_i,\ Y = y_j) = \frac{N_{ij}}{N}, \tag{A3}$$

where N_{ij} is the number of cases, for which the condition in the parenthesis is true. The conditional entropy of Y when X is given, is defined by:

$$H(Y|X) = \sum_{i,j} p_{ij} \ln\left(\frac{p_i}{p_{ij}}\right). \tag{A4}$$

It is obvious from the analysis above that in cases of continuous variables there is a need to divide the input domain of each variable into a finite number of intervals, in which the variables are considered to have a constant value. So we can convert a continuous variable to a discrete one and apply the analysis made above. In our efforts this finite number of intervals was defined by the accuracy of the measurements for each variable. The number n of intervals also defines the maximum value of entropy, as it is shown in the following equation:

$$H(X)_{\max} = \ln(n). \tag{A5}$$

The closer that the actual value of entropy is to its maximum value, the better the representation of this variable in the data set. Consequently, the model based in this data set (the neural model in our case) will have more possibilities for a good interpolation. The ratio of the entropy to its maximum possible value expressed as percentage (representation %) is shown in Table A1 for all the variables used in this study.

When the conditional entropy criterion is being implemented, so that the best subset of variables is chosen, we try to choose these variables, which minimize the conditional entropy or maximize the following ratio ($U(\vec{X})$):

$$U(\vec{X}) = 100\frac{H(Y) - H(Y|\vec{X})}{H(Y)}, \tag{A6}$$

where $H(Y)$ is the entropy of information calculated in Eq. (A2) for the variable Y, that is the variable which the model tries to predict, and $H(Y|\vec{X})$ is the conditional entropy calculated in Eq. (A4), with $\vec{X}$ having as arguments the variables that are used to predict the Y variable.

In this paper the definition of $\vec{X}$ was done by the following algorithm:

1) Calculate all the $U(\vec{X}_1)$ for $\vec{X}_1$ consisting each time of 1 variable
2) Choose this $\vec{X}_1$, which gives the maximum value of $U(\vec{X}_1)$
3) Calculate all the $U(\vec{X}_2)$ for $\vec{X}_2$ consisting each time of two variables one of which is the variable already chosen in step 2.
4) Choose this $\vec{X}_2$ which gives the maximum value of $U(\vec{X}_2)$
5) Go on with triplets and so on until $U(\vec{X})$ sufficiently great.

Tables A2(a) and A2(b) present the results obtained when the algorithm mentioned above was implemented: (a) considering all the variables as possible candidates for $\vec{X}$; and (b) considering only the variables relative to

Table A1
Ratio of the entropy to its maximum possible value for all the variables

Variables	Representation (%) ($H(X)/H(X)_{\max}$)
Specific gravity	67.1
MeABP	47.0
APS	80.0
ABD	80.0
CCR	95.4
Reactor temperature	87.0
Reactor pressure	87.7
Feed rate	68.9
Riser inject steam	61.4
Sulfur	76.3
Basic N2	58.9
RI	71.9
MAT	88.3
Conversion	71.4

Table A2(a)
Calculated U ratios using all the variables as candidates

Variables	U (%)
RI	11.5
MeBP	69.4
SG	92.9
Feed rate	97.3
MAT	98.6
CCR	99.5
Basic N2	99.7
Sulfur	99.7
APS	99.7
ABD	99.7
Riser inject steam	99.7
Rector temperature	99.7
Reactor pressure	99.7

Table A2(b)
Calculated U ratios using only the feed and catalyst properties variables as candidates

Variables	U (%)
RI	11.5
MeBP	69.4
SG	92.9
Sulfur	96.9
Basic N2	98.2
MAT	99.1
APS	99.4
ABD	99.4

feed and catalyst properties as possible candidates for $\vec{X}$, respectively.

Appendix B: Nomenclature

c	coke yield wt.% on feed
C/O	catalyst to oil ratio
d_p	catalyst particles mean diameter
D	riser reactor diameter
f_{gw}	gas-wall Fanning friction coefficient
f_{sw}	solids-wall friction coefficient
G_s	solids mass flux
K	catalytic reaction kinetic constant
k_o	catalytic reaction pre-exponential factor
E	catalytic reaction activation energy
Q_g	gas volumetric flow
Re	Reynolds number for gas phase
Re_p	Reynolds number for solids phase
t_c	catalyst–oil contact time
u_o	superficial gas velocity
u_t	single particle terminal velocity
WHSV	weighted hourly space velocity
x	conversion wt.% on feed
ΔP_{fg}	gas-wall frictional pressure drop
ΔP_{fs}	solids-wall frictional pressure drop
ΔP_{acc}	pressure drop due to solids acceleration
Δz	riser reactor height

Greek symbols

ε	reactor void fraction
μ_g	average gas viscosity
ρ_g	average gas density (reaction vapor mixture and nitrogen)
ρ_p	catalyst particle density

Subscripts

g	gas phase
s	solids phase
p	particle

Table symbols

MSE	Mean Square Error
R^2	R-squared of linear regression trendline
A	the A factor of linear regression trendline ($A \cdot x+B$)
B	the B factor of linear regression trendline ($A \cdot x+B$)
ARE	Absolute Relative Error
MaxRE	Maximum Relative Error

References

[1] J. Michalopoulos, S. Papadokonstantakis, G. Arampatzis, A. Lygeros, Modelling of an industrial fluid catalytic cracking unit using neural networks, Chem. Eng. Res. Des. 79A (2001) 137–142.
[2] G. Christensen, M.R. Apelian, K.J. Hickey, S.B. Jaffe, Future directions in modeling the FCC process: an emphasis on product quality, Chem. Eng. Sci. 54 (1999) 2753–2764.
[3] S.S. Tambe, B.D. Kulkarni, P.B. Desphande, Elements of Artificial Neural Networks, Simulation & Advanced Controls, India, 1996.
[4] B. Lennox, P. Rutherford, G.A. Montague, C. Haughin, Case study investigating the application of neural networks for process modeling and condition monitoring, Comput. Chem. Eng. 22 (11) (1998) 1573–1579.
[5] A.M. Shaw, F.J. Doyle, III, J.S. Schwaber, A dynamic neural approach to nonlinear process modeling, Comput. Chem. Eng. 21 (4) (1997) 371–385.
[6] J.M. Zamarreno, P. Vega, Identification and predictive control of a melter unit used in the sugar industry, Artif. Intell. Eng. 11 (4) (1997) 365–373.
[7] M.A. Kramer, Autoassociative neural networks, Comput. Chem. Eng. 16 (1992) 313.
[8] Z. Weixiang, C. Dezhao, H. Shangxu, Potential function based neural networks and its application to the classification of complex chemical patterns, Comput. Chem. 25 (2) (1998) 385–391.
[9] C. DiMassimo, G. Montague, M.J. Willis, M.T. Tham, A.J. Morris, Towards improved penicillin fermentation via artificial neural networks, Comput. Chem. Eng. 16 (4) (1992) 382.
[10] C. McGreavy, M.L. Lu, X.Z. Wang, E.K.T. Kam, Characterization of the behavior and product distribution in fluid catalytic cracking using neural networks, Chem. Eng. Sci. 49 (24A) (1994) 4717–4724.
[11] A.A. Alaradi, S. Rohani, Identification and control of a riser-type FCC unit using neural networks, Comput. Chem. Eng. 26 (2002) 401–421.
[12] Michael L. Thompson, Mark A. Kramer, Modeling chemical processes using prior knowledge and neural networks, AIChE J. 40 (8) (1994) 1328–1340.
[13] J.A. Wilson, L.F.M. Zorzetto, A generalized approach to process state estimation using hybrid artificial neural network/mechanistic models, Comput. Chem. Eng. 21 (9) (1997) 951–963.
[14] Dimitris C. Psychogios, Lyle H. Ungar, A hybrid neural network-first principles approach to process modeling, AIChE J. 38 (10) (1992) 1499–1511.
[15] I.A. Vasalos, A.A. Lappas, D.K. Iatridis, S.S. Voutetakis, Proceedings of Circulating Fluidized Bed Technology V, Beijing, China, 1996.

[16] F.H. Blanding, Reaction rates in the catalytic cracking of petroleum, Ind. Eng. Chem. 45 (1953) 1186.
[17] G.M. Bollas, I.A. Vasalos, A.A. Lappas, D. Iatridis, Modeling small diameter FCC risers. A hydrodynamic and kinetic approach, Ind. Eng. Chem. Res. 41 (2002) 5410–5419.
[18] A. Rautiainen, S.G. Stewart, V. Poikolainen, P. Sarkomaa, An experimental study of vertical pneumatic conveying, Powder Technol. 104 (1999) 139.
[19] G.S. Patience, J. Chaouki, F. Berruti, R. Wong, Scaling considerations for circulating fluidized bed risers, Powder Technol. 72 (1992) 31.
[20] M. Matsen, Fluidization Technology, vol. 2, J.D.L. Keairns, Hemisphere, 1976 p. 135.
[21] Avidan A. Amos, Reuel Shinnar, Development of catalytic cracking technology. A lesson in chemical reactor design, Ind. Eng. Chem. Res. 29 (1990) 931.
[22] V.W. Weekman, Jr., D.M. Nace, Kinetics of catalytic cracking selectivity in fixed, moving, and fluid bed reactors, AIChE J. 16 (3) (1970) 397.
[23] F.J. Krambeck, Continuous Mixtures in Fluid Catalytic Cracking and Extensions. Mobil Workshop on Chemical Reaction in Complex Mixtures, Van Nostrand Reinhold, New York, 1991.
[24] Arnon Arbel, Zupeng Haung, Irven H. Rinard, Reuel Shinnar, Dynamic and control of fluidized catalytic crackers. 1. Modeling of the current generation of FCC's, Ind. Eng. Chem. Res. 34 (1995) 1228.
[25] Cesar Vergel-Hernandez, Integrated approach for the analysis of FCC evaluation results, Appl. Catal. A 220 (2001) 265.
[26] S. Ramasamy, P.B. Desphande, G.E. Paxton, R.P. Hajare, Consider neural networks for process identification, Hydrocarbon Processing (1995) 59–62.
[27] D.T. Pham, An introduction to artificial neural networks, in: A.B. Bulsari (Ed.), Neural Networks for Chemical Engineering (Chapter 1, pp. 1–19), Elsevier, Amsterdam, 1995.
[28] Dasaratha V. Sridhar, Eric B. Bartlett, Richard C. Seagrave, Information theoretic subset selection for neural network models, Comput. Chem. Eng. 22 (4/5) (1998) 613–626.
[29] D.V. Sridhar, E.B. Bartlett, R.C. Seagrave, An information theoretic approach for combining neural network process models, Neural Networks 12 (1999) 915–926.
[30] S.T. Pugsley, F. Berruti, A predictive hydrodynamic model for circulating fluidized bed risers, Powder Technol. 89 (1996) 57.
[31] J.R. Grace, A.S. Issangya, Bai Dingrong, Bi Hsiaotao, Zhu Jingxu, Situating the high-density circulating fluidized bed, AIChE J. 45 (10) (1999) 2108.