Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant

Iñigo Monedero [a], Félix Biscarri [a], Carlos León [a], Juan I. Guerrero [a], Rocio González [b], Luis Pérez-Lombard [b]

[a] School of Computer Science and Engineering, Electronic Technology Department, Av. Reina Mercedes S/N, 41012 Seville, Spain
[b] School of Industrial Engineering, Department of Energy Engineering, ESI Camino de los Descubrimientos S/N, 41092 Seville, Spain

Keywords:

Petrochemical plant
Expert system
Data mining
Decision system
Neural network
Crude oil distillation
Cost optimization

A B S T R A C T

The energy efficiency of industrial plants is an important issue in any type of business but particularly in the chemical industry. Not only is it important in order to reduce costs, but also it is necessary even more as a means of reducing the amount of fuel that gets wasted, thereby improving productivity, ensuring better product quality, and generally increasing profits. This article describes a decision system developed for optimizing the energy efficiency of a petrochemical plant. The system has been developed after a data mining process of the parameters registered in the past. The designed system carries out an optimization process of the energy efficiency of the plant based on a combined algorithm that uses the following for obtaining a solution: On the one hand, the energy efficiency of the operation points occurred in the past and, on the other hand, a module of two neural networks to obtain new interpolated operation points. Besides, the work includes a previous discriminant analysis of the variables of the plant in order to select the parameters most important in the plant and to study the behavior of the energy efficiency index. This study also helped ensure an optimal training of the neural networks. The robustness of the system as well as its satisfactory results in the testing process (an average rise in the energy efficiency of around 7%, reaching, in some cases, up to 45%) have encouraged a consulting company (ALIATIS) to implement and to integrate the decision system as a pilot software in an SCADA.

1. Introduction

The applications of expert systems are rapidly increasing in the industry. Such applications are very effective in situations when the domain expert is not available (Shiau, 2011). There are diverse problems which need to be solved in the real world and they are difficult to solve by the expert at the moment of carrying out his work. Thus, the expert systems, and specifically the decision systems, become prolific in many fields (Liao, 2005). On the other hand, data mining (Köksala, Batmazb, & Testikc, 2011), or the step of extracting knowledge from the databases, is a discipline intimately related to expert system and which makes it possible to extract the necessary knowledge to design them.

In chemical industry, one of the complex problems for the control of which a computational intelligent approach is amenable, is a crude oil distillation unit. In a crude distillation process, the first objective is to perform an entire process optimization including high production rate with a required product quality by searching an optimal operating condition of the operating variables. In the previous decade, there was considerable research concerning the optimization of crude distillation process. In Seo, Oh, and Lee (2000), the optimal feed location on both the main column and stabilizer is obtained by solving rigorous "a priori" models and mixed integer nonlinear programming. The sensitivity to small variations in feed composition is studied in Dave, Dabhiya, Satyadev, Ganguly, and Saraf (2003). Julka et al. propose in a two-part paper (Julka, Karimi, & Srinivasan, 2002; Julka, Srinivasan, & Karimi, 2002) a unified framework for modeling, monitoring and management of supply chain from crude selection and purchase to crude refining. In addition to analytical non-linear models, computational intelligence techniques such as neural networks (Liau, Yang, & Tsai, 2004) and genetic algorithms (Motlaghi, Jalali, & Ahmadabadi, 2008) are used for the same purpose. In particular, neural networks have been used for modeling and estimation of processes in petrochemical and refineries (Falla et al., 2006; Shirvani, Zahedi, & Bashiri, 2010; Zahedi, Parvizian & Rahimi, 2010).

The scope of present study is concerned with a part of the crude oil distillation called the platforming unit. It is constituted of two subunits: the catalytic reforming or reaction unit and the distillation unit or train distillation. The decision system is focused on optimizing the production rate of the distillation unit which is the most important zone of the platforming unit since it is the one that concentrates the consumption of the plant.

At present, research is not focused only in the rise of the production rate (Jarullah, Mujtaba, & Wood, 2011; Meidanshahi,

* Corresponding author.
E-mail address: imonedero@us.es (I. Monedero).

Bahmanpour, Iranshahi, & Rahimpour, 2011) but also in the improvement of product quality (Iranshahi, Bahmanpour, Paymooni, Rahimpour, & Shariati, 2011; Rahimpour, Vakili, Paymooni, Iranshahi, & Paymooni, 2011). In this sense, classical applications of linear control theories on the distillation unit are widely available in the literature (Jabbar & Alatiqi, 1997). Also nonlinear state estimation research (Jana, Samanta, & Ganguly, 2009) and optimal planning strategy research (Kuo & Chang, 2008) are available. The main objective of these papers was to remove impurities in the distillate (i.e., C_5^+ in the debutanizer column) and maintain the minimum possible amount of product (butane) in the bottom residual fuel oil to maximize the yield of the product.

The aim of our work is to perform a plant energy process optimization, including an adequate production rate with the required product quality but minimizing operating cost (fuel consumption in boilers) through a data mining approach. Several research endeavors have treated consumption analysis as a knowledge discovery problem using intelligence techniques (Li, Bowers, & Schnier, 2010). In De Silva, Yu, Alahakoon, and Holmes (2011), the authors proposed an interesting Incremental Summarization and Pattern Characterization (ISPC) framework for data mining, intelligent analysis and prediction of energy consumption based on electricity meter readings. Both forms of learning, supervised and unsupervised, have been adopted in these studies (Hippert, Pedreira, & Souza, 2001; Metaxiotis, Kagiannas, Askounis, & Psarras, 2003). In Hippert et al. (2001), the unsupervised learning based on the SOM algorithm for the three tasks, namely classification, filtering and identification, of customer load pattern is proposed. Clustering has also been successful in industry applications of data stream mining, such as in Iglesias, Angelov, Ledezma, and Sanchis (2011). The intelligent control algorithms applied to the control of combustion processes have produced satisfactory results and show a great potential for growth. Previous research has shown that boiler efficiency can be optimized with data-mining approaches (Miyayama et al., 1991 and Ogilvie, Swidenbank, & Hogg, 1998). In Kusiak and Song (2006), authors proposed an optimization with clustering-derived centroids. In Song and Kusiak (2007), authors develop a data mining approach for optimizing the combustion efficiency of an electric-utility boiler subject to industrial operating constraints. Latest cited papers offer interesting researches about single boilers. These studies encouraged the authors of the current paper to offer a mining approach to optimize the efficiency of a complete distillation plant, minimizing the operating and economical constraints.

Due to a close monitoring in real-time of the process is, in practice, rarely available, only information collected into an historical database and the data mining software tools were used. The expert's performance is hidden in the collected dataset. This valuable knowledge feeds the proposed decision support system framework. It is not necessary that the global plant control model be reconfigured; the expert's information can simply be extracted. The question that emerges is: Is it possible to extract expert information from the limited amount of data collected in the historical database, searching in past data optimal cost operating conditions? And is it possible to improve energy efficiency by the estimation of new operating condition with a decision system software tool? In this work, we present a decision system, designed through a data mining process, based on an algorithm which integrates a module of neural networks. Besides, a pilot commercial software with the system already integrated is also presented.

2. The refinery platforming unit process

Refineries are composed of several operating units that are used to separate fractions, improve the quality of these fractions and increase the production of higher valued products like gasoline, jet fuel, diesel oil and home heating oil. The function of the refinery is to separate the crude oil into many kinds of petroleum products. This paper pays special attention to platforming unit. This unit is constituted of two basic units: The catalytic reforming or reaction unit and the distillation unit or train distillation.

The catalytic reforming of naphtha is an important refining process that seeks to improve the octane number of gasoline due to a conversion to paraffins and naphthenes in aromatics. The feed to the naphtha reformer is a crude oil fraction from the refinery crude unit with a boiling range between 100 and 180 °C. This process is adiabatically carried out at high temperatures, building up gasoline with a high octane number, LPG, in three reformers: hydrogen, fuel gas and coke. The coke deposits on the spent catalyst surface causing its deactivation. To recover its activation, the catalyst with coke is regenerated after certain running time.

In the first reactor, the major reactions such as dehydrogenation of naphthenes are endothermic and very fast, causing a very sharp temperature drop. For this reason, this process is designed using a set of multiple reactors. Heaters between the reactors allow an adequate reaction temperature level to maintain the catalyst operation.

This process is performed in two different distillation columns (Fig. 1). The separator liquid and a stream, called aromatic LPG from the external platforming unit, feed off the first column, the debutanizer column. This column fractionates the input into two basic products: butane, to the top of the column and a high hydrocarbon flow, also called platformer, to the bottom of the column. Platformer feeds off the debenzenizer, the second distillation unit. Its goal is to obtain a light aromatic flow to the top free to the high hydrocarbon. This stream is fed off the third distillation column that produces benzene and toluene. Benzene and toluene are the important products to the plant. The products are sent to the Morphylane Unit are stored up or sent to the other units of the refinery.

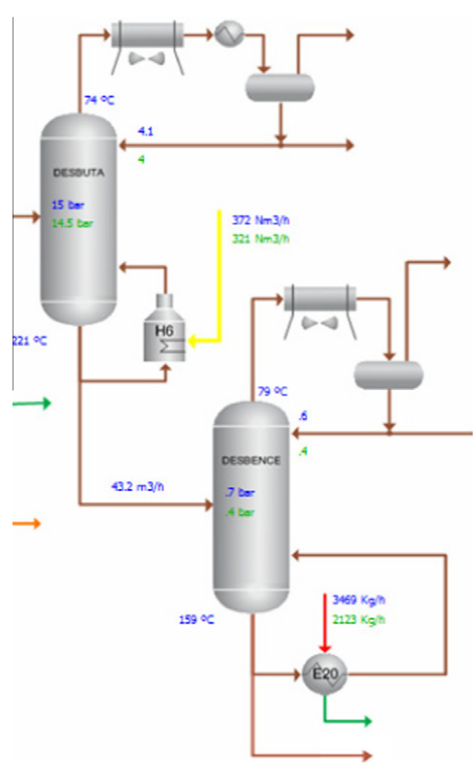

Fig. 1. Flow diagram of the distillation unit.

As the platforming unit is one of the critical and important unit operations for the petroleum industry, the goal is to achieve a well-controlled and stable system, high production rate and product quality as well as low operating cost for the economic consideration. For this reason, the attention has been paid to this unit to improve product rate, efficiency and quality assurance in petroleum industry in recent years.

3. Data preprocessing and discriminant analysis

Any data mining process requires data preprocessing where the data are analyzed, filtered, and formatted (Maimon & Rokach, 2010 and Han & Kamber, 2001). Thus, with regard to the typology of the data recording of the plant, the frequency of the register is hourly, but the quality of the product is analyzed only once a day. The time interval of the samples of data that we had was from January 2009 to May 2010. So, the total sample contained 12149 records corresponding to the set of variables of the plant in an hourly register.

In the first step of preprocessing, 120 outliers corresponding to the days where the plant did not have a usual operation were filtered and deleted from the training sample. Besides, the resulting sample was filtered on the basis of a quality requirement. Concretely, the limit of the level of impurities in the distillate needs to satisfy two rules:

- The percentage of benzene (C_5^+ components) should be less than 1% at the bottom of the debutanizer column. From the sample, this requirement deleted the entries corresponding to 17 days.
- The percentage of toluene should be less than 10% at the bottom of the debenzenizer column. From the sample, this requirement deleted 125 days.

Under these conditions, 2747 registers were filtered. Thus, after the total preprocessing, the sample was reduced from 12149 to 9282 records.

In order to carry out the data preprocessing as well as for developing the models, we used SPSS Modeler (by SPSS Inc., an IBM company). This software is very powerful and extended in the field of data mining. The data processing in SPSS Modeler is done through the use of nodes that are connected together to form a stream frame. Besides, the software includes libraries of artificial intelligence tools (such as neural networks or bayesian networks).

Once the preprocessing is carried out, first of all, we generate a discriminant analysis (Han & Kamber, 2001; Maimon & Rokach, 2010). This type of analysis is used for classification and prediction. The procedure tries to predict, on the basis of one or more predictor, or independent variables, whether an individual or any other subject can be placed in a particular category of a categorical-dependent variable. Our aim with this analysis was to study the influence of the variables of the zone in the energy efficiency grade as well as to quantify the importance of each of these variables.

Before applying the discriminant analysis, we filtered among all the variables of the zone those more important ones that would be used as input parameters of the analysis. For this purpose, we generated the p-value-based Pearson chi-square tests for independence of the target and the predictor without indicating the strength or direction of any existing relationship, and this is suitable for framework purposes. As result of this test, we selected the variables of the zone with highest grade of importance (>0.9). Thus, from the list of variables of the zone (Appendix), eleven attributes were selected (Table 1) to carry out the discriminant analysis.

On the other hand, an objective of the optimization model, which was an indicator of the energy efficient, was generated and added to each register of the sample. Thus, the Energy Efficient Indicator (EEI) was built as the measure of the distillation zone consumption (whose consumption is closely related to the total

Table 1
Attribute selection.

Variable	Importance
PTFINFLOW	1.0
DEBINFLOW	1.0
TPRDLTPPV567	1.0
CSMFGPPH6	1.0
RELRFWDBT	1.0
CSMSTMPRSDES	1.0
GASDESFLOW	1.0
FLWLPGLIQ	1.0
RELRFWDBC	1.0
TPROOM	1.0
PRSOPEDBC	0.99

consumption of the plant, as it is the most important zone of the plant and it uses all the variables from the reaction zone) with regard to the plant input flow. It was defined as follows:

$$EEI = (CSMFGPPH6 * 0.7 * 0.667 + CSMSTMPRSDES * 0.026)/PTFINFLOW \quad (1)$$

This indicator quantifies the consumption of the debenzenizer and the debutanizer. It calculates the cost (in Euros) of the debutanizer (Fuel Gas) and debenzenizer (Vapor on Average) consumption. It is as follows:

Fuel Gas Density: 0.7 kg/m^3
Calorific Fuel Gas: 8.98 kW/m^3
Fuel Gas Cost: 667 euros/ton
Vapor on Average Cost: 26 euros/ton

Once the inputs are selected, a canonical discriminant analysis carries out a discretization and so separates the five classes of EEI through a linear combination of selected attributes. The model evaluation is performed first using ten-fold cross validation in the training sample. Later, a new validation by means of the testing sample is done. This kind of evaluation was selected to train the algorithms using the entire testing data set and obtain a more precise model. This will not only increase the computational effort but also improves the model's capacity for generating different data sets. The evaluation is performed by splitting the initial sample into 10 sub-samples in order to fill the consumption range. The model is trained using 9/10 of the data set and tested with the 1/10 left. This is performed 10 times on different training sets, and finally, the 10 estimated errors are averaged to yield an overall error estimate.

The result of the discriminant analysis is two canonical functions (named Function1 and Function2). Function1 covers 90.8% of the variance, and Function2 covers an additional 9.2%. So, and

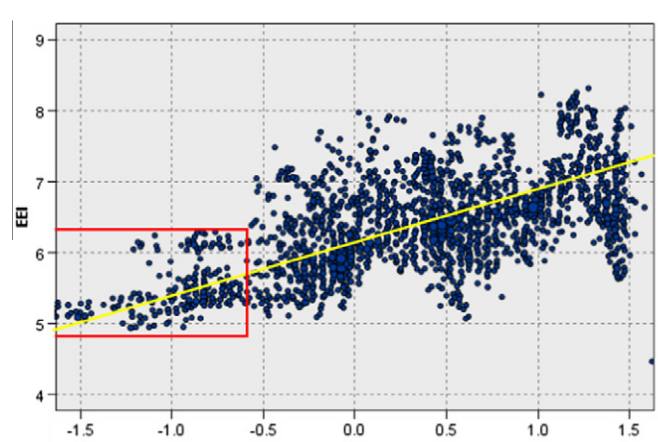

Fig. 2. Trend between Function1 and EEI.

in view of this difference in percentage, we would use only Function1 as a guide of EEI. Fig. 2 shows the trend existing between Function1 and the EEI.

Using the normalized variables, the discriminant analysis also offers a structure matrix that allows building the discriminant functions from discriminating variables, without using the canonical form. From now on, the *N*_prefix indicates a normalized attribute. Variables are ordered by the absolute value of their importance within Function1. Using the normalized variables, and by weighing up the high percentage of variance covered by Function1, the plant energy efficiency will be improved by means of the new attribute defined in (2).

$$\begin{aligned} F = {} & \text{N\_PTFINFLOW} * (-0.343) + \text{N\_CSMSTMPRSDES} * (0.081) \\ & + \text{N\_CSMFGPPH6} * (-0.072) + \text{N\_CSMFGPPH3} * (0.03) \\ & + \text{N\_TPROOM} * (0.064) + \text{N\_DSDGASRCL} * (0.062) \\ & + \text{N\_FLWLPGLIQ} * (-0.054) + \text{N\_PRSOPEDBC} * (0.023) \\ & + \text{N\_RELRFWDBT} * (0.09) + \text{N\_CSMFGPPH4} * (-0.05) \\ & + \text{N\_GASDESFLOW} * (-0.196) + \text{N\_RELRFWDBC} \\ & * (-0.055) + \text{N\_PRSSEPSEPPPV8} * (-0.236) \end{aligned} \tag{2}$$

Thus, this function marks the way that follows the operating points registered for the plant. As observed in Fig. 2, low values of F (specifically below −0.6) guarantee, in a high percentage, low consumption (an EEI that is between 4.8 and 6.2) with regard to the platforming input flow. This resulting function F would be used as a guide and an input parameter in the optimization algorithm based on neural networks (Han & Kamber, 2001; Maimon & Rokach, 2010).

4. Decision system to optimize the plant

The solution that was chosen to perform the optimization algorithm was based on a combined decision system of searching for a path for optimizing the energy index in the plant. Thus, the model combines a search based on the historical data of the working environment of the plant, and an artificial neural network module for additional interpolations of new work environments. We chose this solution taking advantage of the real data we had at the plant, with the objective that our model reached a realistic smooth way from a working point with high EEI to another point with an optimized EEI. Thus, the design of the decision system consisted of an algorithm (or calculus followed for the optimization of a specific operation point), a neural network module (which consists of two interconnected networks), and a graphical environment (with which to visualize the results of the optimization process).

4.1. Algorithmic

The steps carried out in the system for energy optimization are as follows:

1. Reading and storage of historical variables registered in the distillation unit.
2. Selection of the working point of the plant that is necessary to optimize (Each working point implies a set of values in the variables and, therefore, certain conditions of work at the plant).
3. Distinction between variables that plant operators can control and those variables that cannot be controlled. The list of variables with their controllability is indicated in Table 2.
4. Selection from the sample those operation points whose environmental conditions satisfy those of the point selected for optimization. Thus, this process filters those registers whose values are out of a maximum percentage in which the values of the variables can be moved in each one of the iterations. These conditions that result in percentages would be configured by the operator based on their knowledge of the variables of the plant. In this regard, we distinguished between the conditions of the not controllable variables (those conditions for the starting point of work that one wants to optimize and which shall not be violated throughout the optimization process from that point) and the conditions set for the controllable variables (valid for the point that is currently being processed).
5. Activation of the optimization process. This step is carried out by a loop that in each iteration searches for a point (from the set of points of the sample) which meets the environmental conditions marked (described in step 4 of the algorithm) that has the lowest energy. Thus, each one of these iterations improves the energy conditions since the previous point and shows the operator of the plant how (that new values must be set for the controllable variables of the plant) to carry out that improvement.
6. At the time it is not possible to improve the EEI of the current point, meeting the conditions of iteration, the previous loop stop. At this point, it is possible to use a neural network module to improve the energy index. This module generates, by means an interpolation process, a new estimated working point, fulfilling the conditions set for improving the EEI variables at that point (This improvement consists of a little shift of both the EEI as the function F looking for better energy efficiency for that operation point).
7. Once a new point is generated by the neural network module, the operator can shoot again the optimization process to search, from this new estimated point generated by interpolation, for historical points that improve the energy index to meet the conditions for the variables configured. Thus, points 5 and 6 can be executed by the user iteratively, until that operator of the plan reaches the desired improvement in the EEI.

Table 2
Controllability of the variables of the zone.

Variable	Controllability
PTFINFLOW	Not controllable
DEBINFLOW	Not controllable
CSMFGPPH5	Controllable
CSMSTMPRSDES	Controllable
CSMFGPPH6	Controllable
CSMFGPPH3	Controllable
TPROOM	Not controllable
DSDGASRCL	Not controllable
TPRDLTPPV567	Not controllable
FLWLPGLIQ	Controllable
PRSOPEDBC	Controllable
RELRFWDBT	Controllable
CSMFGPPH4	Controllable
GASDESFLOW	Controllable
RELRFWDBC	Controllable
PRSSEPSEPPPV8	Controllable

4.2. Neural network module

The structure of the neural network module for the distillation zone consists of two networks that are applied sequentially. The inputs and outputs of these networks are as follows:

– The first network has seven inputs: the set of uncontrollable variables (N_DSDGASRCL, N_PTFINFLOW, N_DEBINFLOW, N_TPROOM, and N_TPRDLTPPV567) and the parameters EEI and FUNCTION1. The outputs of the network are the five most important parameters implied in function F (with greater weight in this function): N_GASDESFLOW, N_CSMFGPPH6, N_CSMSTMPRSDES, and N_PRSSEPSEPPPV8 N_RELRFWDBT.

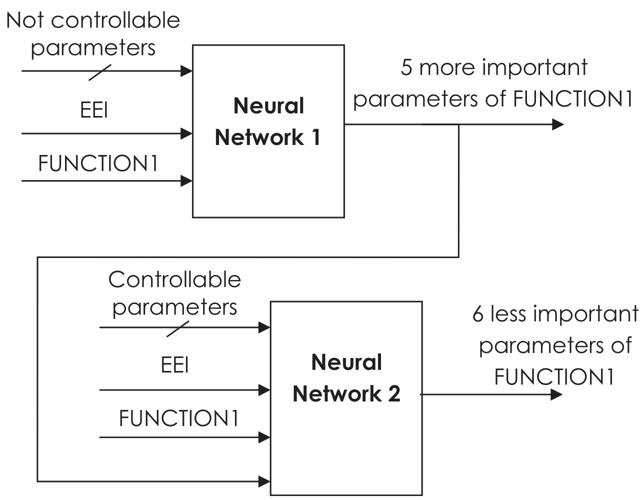

Fig. 3. Structure of the neural network module of the distillation zone.

- The second network has 12 inputs: the set of uncontrollable variables (N_DSDGASRCL, N_PTFINFLOW, N_DEBINFLOW, N_TPROOM, and N_TPRDLTPPV567), EEI and function F as well as the variables used as outputs in the previous network: N_GASDESFLOW, N_CSMFGPPH6, N_CSMSTMPRSDES, N_RELRFWDBT and N_PRSSEPSEPPPV8. The outputs of this network are the 6 parameters of less importance in FUNCTION1: N_CSMFGPPH3, N_CSMFGPPH4, N_CSMFGPPH5, N_RELRFWDBC, N_PRSOPEDBC and FLWLPGLIQ.

The scheme of this neural networks module is shown in Fig. 3.

It is possible to observe both two neural networks that are applied consecutively (since the outputs of the first network are used as inputs of the second one). A scheme of two neural networks working in serial was designed in order to give greater importance and to get a better adjustment in the first network (that predict the most important input parameters of the EEI).

On the other hand, the structures of the neural networks were typical of a backpropagation network, and they had a single hidden layer with 12 neurons in the first network and 15 neurons in the second one. These structures were optimal in order to avoid overtraining as well as overfitting, and they were reached after testing numerous different structures. For the training process, we used 80% of the operation points as training patterns and 20% as validation patterns. The results obtained in these trainings were, respectively, 96.5% and 96.05%.

4.3. Graphical environment

In order to test the model of optimization of the plant, we developed a graphical environment that works on Microsoft Windows, which makes it possible to visualize the results of the decision system described in Section 4.1. This environment is shown in Fig. 4.

The environment as well as the algorithm of the combined system was programmed in C++ (extracting from IBM Modeler the code relative to the different models generated) with the objective of programming it as a software pilot and making easier its final integration in an SCADA.

This environment includes the following working areas in the main window:

- A graph that one can observe, for each of the various operating points of the historic, its EEI (Energy_Index) and function F.
- The value of different parameters for the current operating points (distinguishing between controllable and uncontrollable variables), as well as the configuration of the coefficient or rate of change for each iteration of the loop. In the screen of Fig. 4, these coefficients were configured by the operator of the plant inside a typical range.
- The configuration of the ratio of improvement for the module based on neural networks (on the *X*-axis corresponding to Function1 offset and on the *Y*-axis corresponding to EEI and a ratio by which these values are multiplied).
- A historic text showing the evolution of different variables along the process of iterations of the algorithm that optimize the energy efficiency.
- A set of buttons that carry out the execution and test the combined model for a particular operating point.

5. Results

The objective of our work was to optimize the energy consumption of the plant (quantified as EEI) for any operation point in which the plant could operate. Thus, in order to validate and quantify the results, an additional algorithm for measuring the improvement was obtained in our model in each one of the operating

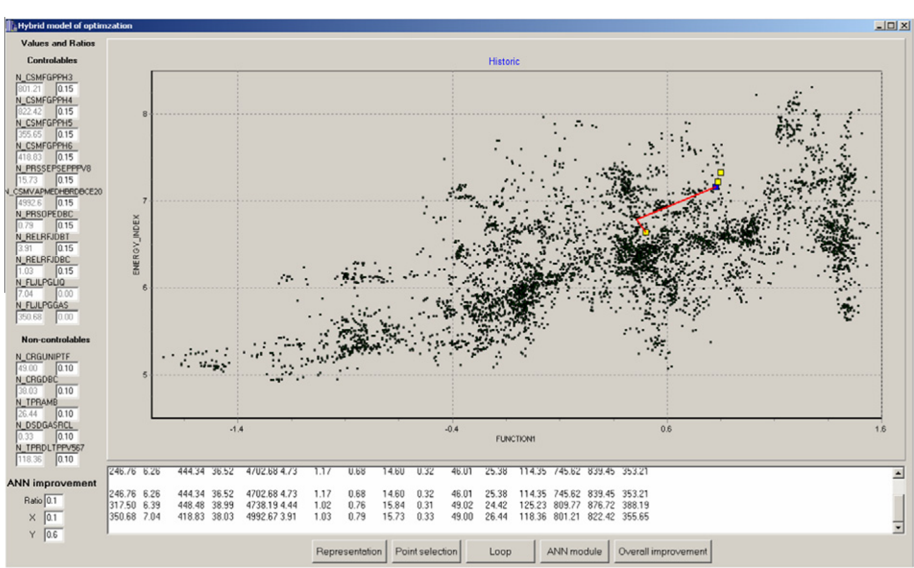

Fig. 4. Graphical environment developed to test the system.

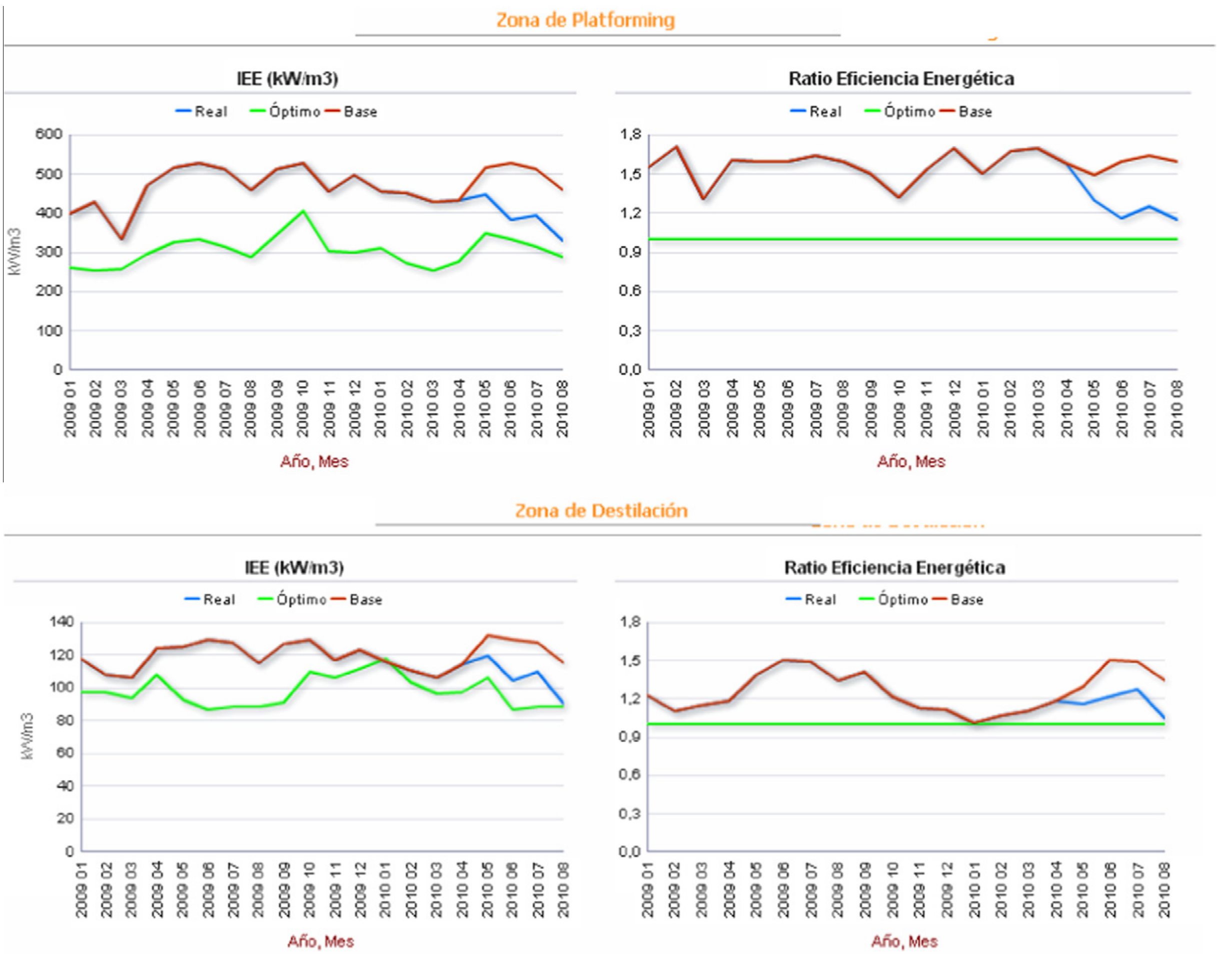

Fig. 5. SCADA environment by ALIATIS with the system integrated.

points of the past. The aim of this algorithm was to calculate the improvement in the EEI for every operation point of the plant registered in the past.

This measurement algorithm (which could be fired by means of the button 'overall improvement' in the graphical environment) carries out the following steps:

1. For each operating point of the historical:
   1.1. Execution of the optimization algorithm (described in Section 4.1).
   1.2. Fire of the module of neural networks.
   1.3. New execution of the optimization algorithm (once added to the list of operation points, the point is generated with the module of neural networks).
   1.4. Calculation of the percentage of improvement in the EEI for that point.
2. Calculation and register (in the historic box) of the improvement of the average of the percentages of all points.

Thus, once calculated with the previous algorithm, for the improvement that had been obtained in the system for each one of the operation points registered for the plant, the average improvement of energy efficiency reached was 6.33%. This result is very good taking into account that a lot of operating points were in the range of [5,6] in their EEI and, therefore, the scope of improvement was less. Thus, once analyzed the results, we could observe as the improvement for those operating points with EEI higher than 7 were 9.85% (some operating points reached an improvement of 45%). Besides, these results largely depend on the chosen adjust of the percentages of variation permitted for the controllable and not controllable variables (which have been percentages quite restrictive for our tests).

Having validated the results of the developed system, this was implemented and integrated in an SCADA by an engineering software company (ALIATIS). Thus, the commercial application shows in real time the present state of EEI of the plant as well as the optimum EEI calculated from the designed system. In the SCADA, the operator is informed of how this optimum EEI could be reached (by means of the regulation of the parameters that they can control). Thus, this application also includes the present proposed saving cost and the decision system save report, by an EEI operation screen. The results obtained once they carried out the simulation in the SCADA environment with the system since they were integrated are shown in Fig. 5 (the company is in Spain and, for this reason, the environment is a Spanish one). In these results, an average difference of 15% between the evolution of EEI once the system is integrated (marked in the graphic as 'real') and the evolution of EEI without the system (marked in the graphic as 'base') is observed. Besides, the cost information can be shown for the entire plant and also for every part (heaters) of the plant. Thus, the evolution of attributes and the savings from the heaters are displayed in diverse graphics and tables.

6. Conclusions

The present work describes the design and development of a combined decision system based on a module of neural networks for the optimization of the consumption of a petrochemical plant.

The algorithm, which is based on the system, uses the information relative to the parameters registered for the plant help with a neural network module for optimizing its future operation points. Thus, this algorithm helps the operator to take decisions to improve the energy efficiency of the plant.

On the other hand, a bibliographical revision of works with the same objective has been carried out. We have checked, as the use of the neural network is not everything that should be extended in the chemical industry for our particular purpose. This kind of neural structure is demonstrated as having an excellent behavior in a similar type of interpolation problems when it counts with historical data of the plant and a great number of input variables.

The system implements a kernel based on two backpropagation neural networks. The main contribution of our work is how to combine the historic data of a plant with a neural network structure for generating new interpolated operation points and so to generate a decision system for the operator. There are two advantages of this model:

- It is a system generated on real conditions of operation. Thus, the use of an interpolation algorithm as neural networks is only for linking the operation point in the present with operation points that had already happened in the past. This fact ensures results that are not only theoretical but also eminently practical.
- It is system that can be constantly improved. This is due to the fact that the new conditions and operation points can be used to train and adjust the neural networks.

Besides, the work presented in this article has been implemented and integrated in an SCADA by a consulting company. Thus, the results obtained so far are considered satisfactory taking into account the limitation of the available data for the plant. In fact, the system attains an average improvement of around 15% for the plant, which is very significant from the previous company process. Besides, these results will be improvable in the future by means of a refinement of the developed neural networks.

Appendix A. Main attributes in the catalytic reforming

PTFINFLOW (m^3/h): The platforming input flow.
DEBINFLOW (m^3/h): The debenzenizer input flow.
WAIPTF (°C): The variable that measures the catalyst deterioration.
TPROUT_PPV5 (°C): The output PPV5 reactor's temperature.
TPROUT_PPV6 (°C): The output PPV6 reactor's temperature.
TPRIPPPV7 (°C): The input PPV7 reactor's temperature.
TPRIPPV5 (°C): The input PPV5 reactor's temperature.
TPRIPPV6 (°C): The input PPV6 reactor's temperature.
TPRIPPV7 (°C): The input PPV7 reactor's temperature.
GASDESFLOW (Nm^3/h): Gas fraction of the desbutanizer's top flow.
FLWLPGLIQ (m^3/h): Liquid fraction of the desbutanizer's top flow.
CSMSTMPRSDES: Medium pressure steam consumption in the desbenzenizer's reboiler.
CSMFGPPH3 (m^3/h): The fuel gas PPH3 heater's consumption.
CSMFGPPH4 (m^3/h): The fuel gas PPH4 heater's consumption.
CSMFGPPH5 (m^3/h): The fuel gas PPH5 heater's consumption.
CSMFGPPH6 (m^3/h): The fuel gas PPH6 heater's consumption.
PRSSEPSEPPPPV8 (bar): The PPV8 product separator pressure.
RELRFWDBT: Reflux ratio: Total of desbutanizer's top stream/liquid return to the column.
RELRFWDBC: Reflux ratio: Total of desbenzenizer's top stream/liquid return to the column.
PRSOPEDBC (bar): Debenzenizer's pressure operation.
DSDGASRCL ($Kg/(N * m^3)$): The recycle gas density. This variable maintains P PPV 8 brought under control.
TPRDLTPPV567 (°C): The temperature increase between the three reactors.
TPROOM (°C): The room temperature.

References

Dave, D. J., Dabhiya, M. Z., Satyadev, S. V. K., Ganguly, S., & Saraf, D. N. (2003). Online tuning of a steady state crude distillation unit model for real time applications. *Journal of Process Control* 13(3), pp. 267–282.

De Silva, D., Yu, X., Alahakoon, D., & Holmes, G. (2011). A data mining framework for electricity consumption analysis from meter data. *IEEE Transactions on Industrial Informatics, 7*(3), 399–407.

Falla, F., Larini, C., Leroux, G., Quina, F., Moro, L., & Nascimento, C. (2006). Characterization of crude petroleum by NIR. *Journal of Petroleum Science and Engineering, 51*(1-2), 127–137.

Han, J., & Kamber, M. (Eds.), 2001. Data Mining. Concepts and Techniques, 1st ed. Morgan Kaufmann.

Hippert, H., Pedreira, C., & Souza, R. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems, 16*(1), 44–55.

Iglesias, J., Angelov, P., Ledezma, A., & Sanchis, A. (2011). Creating evolving user behavior profiles automatically. *IEEE Transactions on Knowledge and Data Engineering, 9*(99), 1.

Iranshahi, D., Bahmanpour, A. M., Paymooni, K., Rahimpour, M. R., & Shariati, A. (2011). Simultaneous hydrogen and aromatics enhancement by obtaining optimum temperature profile and hydrogen removal in naphtha reforming process; a novel theoretical study. *International Journal of Hydrogen Energy, 36*(14), 8316–8326.

Jabbar, N. A., & Alatiqi, I. (1997). Inferential-feedforward control of petroleum fractionators: A PNA approach. *Computational Chemical Engineering, 21*, 255–262.

Jana, A. K., Samanta, A. N., & Ganguly, S. (2009). Nonlinear state estimation and control of a refinery debutanizer column. *Computers & Chemical Engineering, 33*(9), 1484–1490.

Jarullah, A. T., Mujtaba, I. M., & Wood, A. S. (2011). Enhancement of productivity of distillate fractions by crude oil hydro-treatment: Development of kinetic model for the hydro-treating process. *Computer Aided Chemical, 29*, 261–265.

Julka, N., Srinivasan, R., & Karimi, I. (2002). Agent-based supply chain management1: Framework. *Computers & Chemical Engineering, 26*(12), 1755–1769.

Julka, N., Karimi, I., & Srinivasan, R. (2002). Agent-based supply chain management2: A refinery application. *Computers & Chemical Engineering, 26*(12), 1771–1781.

Köksala, G., Batmazb, I., & Testikc, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications, 38*(10), 13448–13467.

Kuo, T. H., & Chang, C. T. (2008). Optimal planning strategy for the supply chains of light aromatic compounds in petrochemical industries. *Computers & Chemical Engineering, 32*(6), 1147–1166.

Kusiak, A., & Song, Z. (2006). Combustion efficiency optimization and virtual testing: A data-mining approach. *IEEE Transactions on Industrial Informatics, 2*(3), 176–184.

Li, X., Bowers, C., & Schnier, T. (2010). Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics, 57*(11), 3639–3644.

Liao, S. H. (2005). Expert system methodologies and applications–a decade review from 1995 to 2004. *Expert Systems with Applications, 28*(1), 93–103.

Liau, L. C., Yang, T. C., & Tsai, M. T. (2004). Expert system of a crude oil distillation unit for process optimization using neural networks. *Expert Systems with Applications, 26*(2), 247–255.

Maimon, O., & Rokach, L. (Eds.), (2010). Data Mining and Knowledge Discovery Handbook, 2nd ed. Springer, US.

Meidanshahi, V., Bahmanpour, A. M., Iranshahi, D., & Rahimpour, M. R. (2011). Theoretical investigation of aromatics production enhancement in thermal coupling of naphtha reforming and hydrodealkylation of toluene. *Chemical Engineering & Processing, 50*(9), 893–903.

Metaxiotis, K., Kagiannas, A., Askounis, D., & Psarras, J. (2003). Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher. *Energy Conversion and Management, 44*(9), 1525–1534.

Miyayama, T., Tanaka, S., Miyatake, T., Umeki. T, Miyamoto, Y., Nishino, K. & Harada E. (1991). A combustion control support expert system for a coal-fired boiler. In *Proceedings of the international conference on industrial electronics, control and instrumentation IECON'91*, 1991, Vol. 2, pp. 1513–1516.

Motlaghi, S., Jalali, F., & Ahmadabadi, M. N. (2008). An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework. *Expert Systems with Applications, 35*(4), 1540–1545.

Ogilvie, T., Swidenbank, E., & Hogg, B. (1998). Use of data mining techniques in the performance monitoring and optimisation of a thermal power plant. Knowledge Discovery and Data Mining (1998/434). In *IEE Colloquium on 7/1-7/4*.

Rahimpour, M. R., Vakili, R., Paymooni, K., Iranshahi, D., & Paymooni, K. (2011). A novel integrated, thermally coupled fluidized bed configuration for catalytic naphtha reforming to enhance. *International Journal of Hydrogen Energy*, *36*, 2979–2991.

Seo, J. W., Oh, M., & Lee, T. H. (2000). Design optimization of a crude oil distillation process. *Chemical Engineering & Technology*, *23*(2), 157–164.

Shiau, W. L. (2011). A profile of information systems research published in expert systems with applications from 1995 to 2008. *Expert Systems with Applications*, *38*(4), 3999–4005.

Shirvani, Y., Zahedi, G., & Bashiri, M. (2010). Estimation of sour natural gas water content. *Journal of Petroleum Science and Engineering*, *73*, 156–160.

Song, Z., & Kusiak, A. (2007). Constraint-based control of boiler efficiency: A data-mining approach. *IEEE Transactions on Industrial Informatics*, *3*(1), 73–83.

Zahedi, G., Parvizian, F., & Rahimi, M. R. (2010). An expert model for estimation of distillation sieve tray efficiency based on artificial neural network approach. *Journal of Applied Sciences*, *10*(12), 1076–1082.