Soft Sensor for predicting Crude Oil Distillation Side Streams using Evolving Takagi-Sugeno Fuzzy Models. Results outlined

José J. Macías-Hernández*, Plamen Angelov and Xiaowei Zhou
InfoLab21 - Lancaster University
South Drive, Lancaster University, Lancaster, LA1 4WA, UK
+44 (1524) 510391; fax: +44 (1524) 510489
email: p.angelov@lancaster.ac.uk
*Chemical Eng. Dept. - La Laguna University
Chem. Eng. Dept., University of La Laguna, Tenerife, Spain
Phone: +34 649 741084
email: jmacias@ull.es

ABSTRACT: Prediction of the properties of the crude oil distillation side streams based on statistical methods and laboratory-based analysis has been around for decades. However, there are still many problems with the existing estimators that require a development of new techniques especially for an on-line analysis of the quality of the distillation process. The nature of non-linear characteristics of the refinery process, the variety of properties to measure and control and the narrow window that normally refinery processes operates in are only some of the problems that a prediction technique should deal with in order to be useful for a practical application. There are many successful application cases that refinery units use real plant data to calibrate models. They can be used to predict quality properties of the gas oil, naphtha, kerosene and other products of a crude oil distillation tower. Some of these are distillation end points and cold properties (freeze, cloud). However, it is difficult to identify, control or compensate the dynamic process behaviour and the errors from instrumentation for an online model prediction.

The objective of this work is to report an application an a study of a novel technique for real-time modelling, namely eXtended Evolving Fuzzy Takagi-Sugeno models (xTS) for prediction and online monitoring of these properties of the refinery distillation process. The results presented here include the online prediction of Soft Sensors for distillation and In flammability of Kerosene Side Stream. The application takes data in an automatic fashion and predicts the quality of the side stream evolving its fuzzy structure and cluster parameters to represent a better behaviour of the plant. These preliminary results show the performance of this technique as an online estimator.

KEYWORDS: eTS, xTS, Takagi-Sugeno, Evolving Fuzzy Logic, Crude Oil, Soft Sensor, Distillation Cuts

PROBLEM STATEMENT

CRUDE oil distillation tower is the heart of any crude refinery because it is the process element in charge of the separation of petroleum cuts. These cuts are later being processed in other operation units in order to refine and blend the gasoline, gas oil and other commercial products. The separation is made in a distillation tower with lateral extractions.

Although there are further technological processes that deal with refining these initial properties, the specification requirements for some of them are very strict even at this initial stage of the process. Moreover, this initial separation affects the efficiency of the whole refinery process due to the fact that the yield obtained in these cuts contributes significantly to the overall refinery profit. The so called '95% ASTM' distillation curve specification limits the amount of product that can be extracted in a side stream (note that ASTM is one of the largest voluntary standards development organizations in the world [18]. The ASTM distillation curve analysis is the relation between the vapor leaving the pot and the distillated product and temperature. If extract more products from the tower one could obtain a higher end point and thus a heavier product. The economic goal of operating a distillation tower is based on obtaining the maximum amount of product with the highest quality within the specification.

A. QUALITY MONITORING BY ANALYZERS

To maintain the quality of the side stream products refiners normally sample each stream and do ASTM analysis [8], [9]. Between samples and if justified by economic benefits sometimes refiners can justify the installation of continuous analyzers to reduce the deviation from specification [10]. However, there are so many errors due to lab – analyzer discrepancies like (not limited to):
The on-line analyzer's availability is scarce.
The significant difference between the results of laboratory samples and of the analyzers that are mainly due to:
Different methods used (in the laboratory and in the on-line analyzer equipment);
Different sampling
The cost of the analyzer that reproduces the ASTM lab
Etc.
Even in the case when an analyzer is available, it is difficult to maintain a stable correlation between the results.

B. QUALITY INFERENCE MODELS

Models to inferred specifications from available plant measurements, such as temperature, flow rates, pressures, are very old (back to sixties). They are even much older than Online Analyzers. However, these models have errors from both the measurements errors and from the inadequacy of the model structure. The plant has a dynamic, non-stationary behavior, it evolves, but the model is fixed. One simplistic way to adjust this discrepancy is to recalibrate the model periodically (usually per month) [11].
Nowadays, we have a huge amount of information including measurement and laboratory sample data. These databases constitute the principal support to all statistics methods [13] to build soft sensors models (multivariate PCA and PLS) [6], [7], [12], [13], Neural Networks [11].
However, all these techniques do not solve the problem of the data pattern drift and shift or data "evolution" and this requires in practice a manual intervention, which is costly and precludes the full automation. Without recalibration, the operators have to give up the model. A strong demand exists for self-calibrating or evolving sensors [19].

C. EVOLVING STATISTICAL METHODS

An alternative is to use an adaptive soft sensor that is able to evolve its structure and to follow the data pattern, to re-train and adapt is appealing because it would save computational, time, and man-power resources, will increase the level of automation and intelligence of the manufacturing process. If use fuzzy rule-based models such an intelligent evolving sensor one would also gain the important advantage of interpretability of the results (due to the transparent nature of the fuzzy rules). The gradual evolution of the model structure (rules) will mean that a re-training of the intelligent sensor once a month or once a quarter will keep the basic fuzzy–rule base structure and will only modify (add or replace few fuzzy rules [14].
In a previous work presented we have applied a novel technique for soft sensors (Extended Evolving Takagi-Sugeno Fuzzy Model) and the results to quality prediction of crude oil cuts are very promising [22].

EXTENDED EVOLVING TAKAGI-SUGENO FUZZY MODEL (XTS) AS A BASIS FOR THE EVOLVING INTELLIGENT SENSOR

EXtended evolving Takgi-Sugeno (exTS) fuzzy modeling approach [14] is a modified version of the evolving TS fuzzy (eTS) models [20], which is a form of self-organizing transparent (linguistic) model that evolves both its structure and its parameters in a computationally efficient way based on the data stream in real-time. The evolvable intelligent sensor is considered as a set of fuzzy rules of the following form:

$$R^i: \quad \textbf{IF } (x_1 \text{ is close to } x_1^{i*}) \textbf{ AND } \ldots \textbf{ AND } (x_n \text{ is close to } x_n^{i*}) \textbf{ THEN} \quad (1)$$
$$(y^i = [1, x^T][a_0^i, a_1^i, \ldots, a_n^i]^T) \qquad i=1,2,\ldots,N$$

where R^i denotes the i^{th} fuzzy rule; N is the number of fuzzy rules; x is the input vector $x = [x_1, x_2, \ldots, x_n]^T$; *'is close to'* denotes the antecedent part of the respective fuzzy sets, $j=1,2,\ldots,n$; n is the number of measured variables

(temperatures, pressure, flow rates etc.); y^i is the local/partial output of the i^{th} fuzzy rule that is further combined to form the overall *soft* sensor prediction of the quality parameter of interest, y. $A^i = [a_0^i, a_1^i, ..., a_n^i]^T$ are its parameters.

The overall output, y represents the prediction of interest (temperature of the heavy naphtha when it evaporates 95% liquid volume (T^{hn}), temperature of the gasoil (T^{gol}). According to the Takagi-Sugeno-Kang fuzzy based reasoning paradigm [21] it can be find as a fuzzily weighted combination of the local/partial models of the type (1):

$$QP = \sum_{i=1}^{N} \lambda^i y^i \tag{2}$$

where QP is the quality parameter (T^{hn}, T^{gol}); the fuzzy weights λ^i represents a normalized level of activation of the i^{th} fuzzy rule, which can be found as:

$$\lambda^i = \frac{\prod_{j=1}^{n} \mu_j^i(x_j)}{\sum_{j=1}^{N} \prod_{j=1}^{n} \mu_j^i(x_j)} \tag{3}$$

where μ_j^i is the membership value of the j^{th} input x_j, $j=[1,n]$), to the i^{th} fuzzy set ,$i=[1,N]$. The degree of membership, which can later be labeled with meaningful linguistic labels are often defined by Gaussian type bell functions (due to the resemblance of normal distribution and good generalization properties and due to the fact that they cover the whole cardinality of the input variable - the whole range of temperatures, flow rates etc.):

$$\mu^i = e^{-\frac{\sum_{j=1}^{n} (x_j - x_j^{i*})^2}{2(\sigma_j^i)^2}} \tag{4}$$

where $(\sigma_j^i)^2$ is the spread of the membership function, which also represents the radius of the zone of influence of the cluster/rule.

In the xTS the spread of membership spread σ, which corresponds to the radius of influence from the cluster center in the data space, is not constant [14] as in eTS [20], but is updated time by time when the new sample are allocated to the cluster, based on the spatial density information:

$$\sigma_{jk}^l = \rho \sigma_{j(k-1)}^l + (1-\rho) S_{jk}^l; r_{j1}^l = 1 \tag{5}$$

where ρ is a leverage that weights the compatibility of the new information with the old one; S_k^i is the support (population) of i^{th} cluster at step k; z_l is the l^{th} input-output vector in i^{th} cluster, z^{i*} is the proto-type centre of i^{th} cluster.

Due to the lack of space only the basic concept of the learning methodology is outlined here. For more details, please refer to [14], [20]. The learning method combines unsupervised (in respect to the antecedent part of the model) with the supervised (in terms of the consequent parameters) learning. Note, that both phases – the unsupervised model structure identification (using on-line clustering) and the unsupervised model parameter identification (using a version of coupled recursive least squares estimation) are perfomed at the same model update stage. This type of model update-model use for prediction is typical for on-line estimation, adaptive control etc. [22].

According to the fuzzy modeling concept proposed by Tomohiro Takagi and Michio Sugeno in 1985 [21], each of the fuzzy rules of type (1) operate in certain sub-area of the input/output data space, $z = [x^T; y^T]^T$; $z \in R^{n+m}$. To identify these regions Angelov proposed to employ real-time clustering [20] thus effectively learning the antecedent part of the fuzzy rules (namely the focal point, x^{i*} and the spread, σ_j^i). Once the antecedent part of the fuzzy model is determined and fixed the parameters of the consequent part, A can be identified using fuzzily weighted recursive least squares (RLS), [21]. The overall output of the soft sensor is then given by:

$$QP = \psi^T \alpha \tag{6}$$

where $\alpha = \left[\left(A^1\right)^T, \left(A^2\right)^T, ..., \left(A^N\right)^T\right]^T$ is a vector formed by the sub-system parameters; $\psi = [\lambda^1[1,x]^T, \lambda^2[1,x]^T, ..., \lambda^N[1,x]^T]^T$ is a vector of the inputs that are weighted by the normalized activation levels of the rules, λ^i , *i=[1,N]*.

For a given sensor reading, $[x_k^T; QP_k^T]^T$ the optimal in least squares sense solution α^*_k that minimizes the following prediction error function:

$$\left(QP - \Psi^T\alpha\right)^T\left(QP - \Psi^T\alpha\right) \rightarrow \min \tag{7}$$

can be found applying fuzzily weighted RLS [20]:

$$\alpha_k^* = \alpha_{k-1}^* + C_k\psi_k(QP_k - \psi_k^T\alpha_{k-1}^*) \tag{8}$$

$$C_k = C_{k-1} - \frac{C_{k-1}\psi_k\psi_k^T C_{k-1}}{1+\psi_k^T C_{k-1}\psi_k} \tag{9}$$

where $\alpha^*_1=0$; C is a $N(n+1)\times N(n+1)$ co-variance matrix; $C_1=\Omega I$; Ω is a large positive number; I is the identity matrix; $k=2,3,...$

IMPORTANT PARAMETERS IN THIS CRUDE OIL DISTILLATION PROBLEM

A. QUALITY AND FREQUENCY OF THE DATA SAMPLING

In the previous work [22], we have being assuming that hourly averages will filter process variations and therefore the calibration with lab spot values is better. However, this is only valid is the dynamic of the process is fast enough to catch all variation within one hour. The characteristic of this work is to demonstrate that this is not always true and there are better predictions if we delay the inputs some time even taking hourly averages [11].

Another reason to use hourly averages was the availability and data compression characteristics.

B. LABORATORY SPOT VALUES

In the Plant database there is in some cases laboratory information with very low periodicity. This makes that the number of samples used to calibrate the models very limited, and therefore, the data error can propagate easily to the model. In addition, the laboratory samples are taken only once or twice per day and not exactly in the time that is indicated on the label of the bottle.

C. CRUDE DISTILLATION TOWER

The plant that has been used for this study is a crude distillation tower with a design capacity of 80000 bbl/d of Arabian crude. It has 47 valve trays with a diameter of 5.2 m in its top section and of 2.4 m in its bottom section. This tower has two cold pumps around, kerosene oil (KNO), gas oil (GOL) and a hot pump around GOP (washout). It has five side streams from top to bottom these include:

- Heavy Naphtha (HN)
- Kerosene (KN)
- Light Gas oil (LGO)
- Medium Gas oil (MGO)
- Heavy Gas oil (HGO).
- Atmospheric Residue (RES).

It has a vapor side stripper for each side stream and a bottom vapor injection.

D. THE DATA SET

The data used in this study represents the analysis from the laboratory and instrument data from a crude unit operation. The input data are instrument readings and the output data or the data we need to predict are the sample results from laboratory tests.
The laboratory samples are taken on a daily basis, seven days a week. In this work we are reading the 24 hourly averages a day and not only the hourly average when the lab sample was taken. The total period studied is from 1/1/2006 to 19/972006. The dataset includes unit emergencies, shutdowns, process and instrument malfunction and even lab samples errors if any. There is no filtered applied to the dataset.

Input data

The input data are temperatures, flow rates, and pressures of the main crude tower. Steam injection is included to side stream strippers, unit throughput and crude density.

Output data

The aim is to predict:
1) Temperature of the heavy Naphtha when it evaporates 95% liquid volume, ASTM D86-04b, Thn, oC. Std. deviation of the original data 6.2ºC
2) Temperature of the GOL when it evaporates 85% liquid, ASTM D86-04b, Tgol, oC. Std. Dev. 9.6ºC

E. PRIOR KNOWLEDGE

In our previous work [22], we have used the previous engineering knowledge of the main variables that affects the quality of the analysis. In this work we are not using this but applying a methodology to extract the main contributors to the model. Just as a reference we copy the main theoretical contributions expected in the plant.
The existing knowledge of the tower operation indicates that:
A) Temperature of the heavy Naphtha when it evaporates 95% liquid volume in a distillation tower mainly depends on:
- The pressure of the tower, p, kg/cm2g
- The amount of the product taking off, P, %
- The density of the crude, d, g/l
- Temperature of the column overhead, Tco, oC
- Temperature of the naphtha extraction, Tne, oC

B) Temperature of the GOL when it evaporates 85% liquid, TGOL
- The pressure of the tower, p, kg/cm2g
- Amount of product taking off, P, %
- Density of the crude, r, g/l
- Temperature of the column overhead, Tco, oC
- Steam introduced in GOM stripper, SgG, kg/h,
- Temperature of the GOL Extraction, TGOL, oC
- Temperature of the Kerosene Extraction, Tke, oC
- Temperature of the Naphtha Extraction, Tne, oC

For a distillation analysis, the most importance variables normally are extraction temperatures (naphtha, kerosene, and gas oil), corrected by overhead pressure input variable.
For the Abel analysis, the most important variable is the steam ratio to kerosene extraction. Inflammability depends on the lighter part of the cut and this can be removed using steam. It should also be emphasized that the input data is highly collinear.

PERFORMANCE INDICATOR USED

The performance to identified and select variables used was based on the standard deviation over the absolute errors between real outputs and estimations during the modeling phase (90 lab samples (around 2 months lab sample data):

$$e = \sqrt{(\frac{1}{N_{val}-1}) \sum_{i=1}^{N_{val}-1} (\varepsilon_i - \bar{\varepsilon})^2} \quad (10)$$

Where e is the measurement; Nval is the number of samples during the calibration period; ε denotes the absolute error.

METHODOLOGY AND MODELING RESULTS

The experiments were conducted with the real data 24 hourly averages/day from 1/1/2006 to 19/9/2006 (aprox. 10 months, 6171 lines of data) including shutdowns, instruments malfunction, etc.

1) For each variable to model we have taken 90 data where we do know the value we need to predict (lab samples). Sometimes the sample is taken one or twice a day and therefore, the period is around two or three months. The take the calibration error as a guide to select best variable

2) For each variable Y, we use one X at the time and calibrate the fuzzy model xTS to obtain calibration error. The X variable is each instrumentation variable available on line in the plant. To identify the dynamics of the instrumentation data is delay from 0 to 3 hours and make one model for each shift (four models each pair, 4*33*4=528).

3) Then, all models are sorted and for each variable problem we have taken the best 5 or 6 X variables and Delays that gives the better performance.

4) Taken this small group of variables, we have calibrated again the xTS mode, but now with the correct delay.

5) This model is then simulated online by predicting from the end of calibration period over every hour even if no lab samples is available (no feedback) until a new lab data comes in. In that moment, the model evolves and continues predicting with no lab samples until next. This sequence repeats to the end of the period.

6) We have make models in the same way but only with the variables that our engineering knowledge recommends and compared results (the delays are in the optimum as well)

RESULTS

Hereafter, it's shown some figures of the results obtained from automatic selection criteria and engineering criteria based on previous knowledge. For all cases studied its repeat that the previous selection of variables just based on engineering criteria is not enough to select the main variables due to those variables my or may not have enough variability. As it shown for Naphtha and GOL, the results are very completive each other.

Table I:. Variables selected with main contribution criteria for ASTM D86, 95%NAPTHA

Variables	Delay, (hours)
GOM stripping steam	2
Reflux temperature	1
Tower Bottoms temperature	1
KNO side stream temperature	0
GOL side stream temperature	0
Naphtha extraction/Feed ratio	1

Table II:. Variables selected with main contribution criteria for ASTM D86, 85%GOL

Variables	Delay, (hours)
Reflux temperature	2
Temperature of GOL side stream	1
Temperature of GOP side stream	2
GOM extraction/Feed ratio	1

RESULS FOR ASTM D86, 95%NAPTHA

Automatic model variable selection

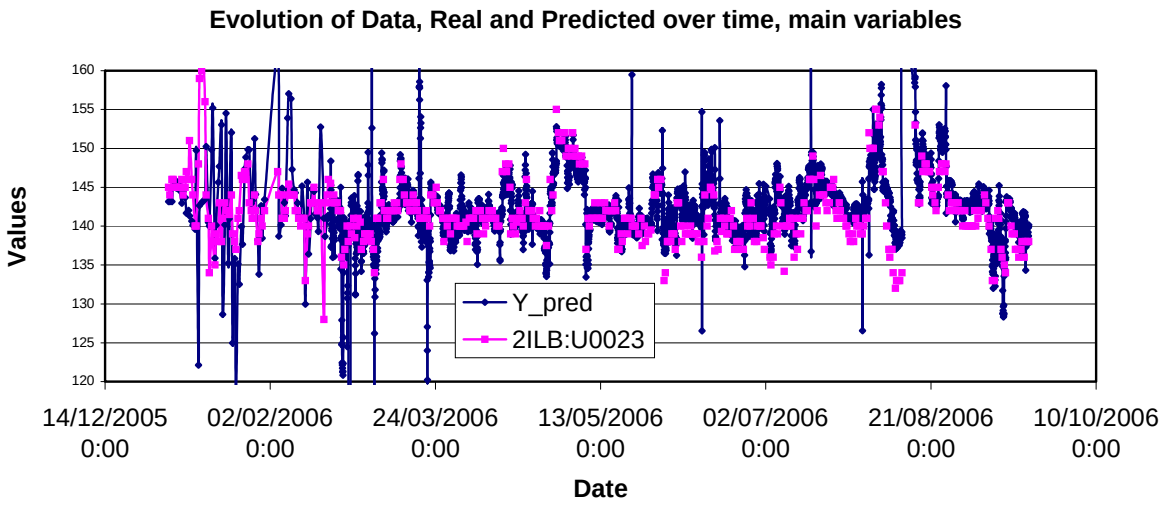

Figure 1: ASTM D86, 95%BPNAPHTHA. Automatic model variable selection.

Variables selected based on engineering criteria

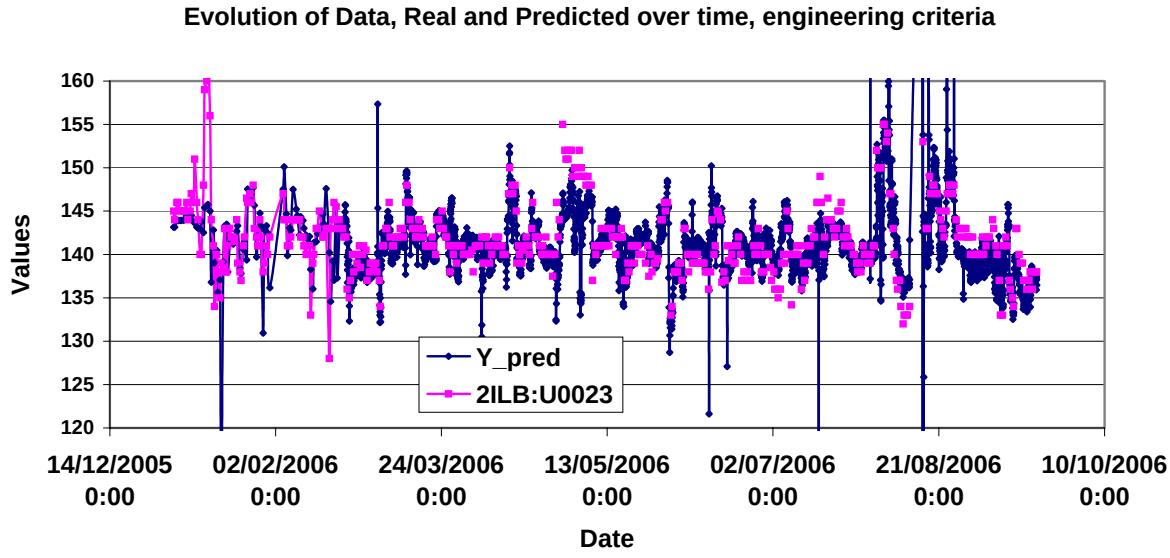

Figure 2: ASTM D86, 95%BPNAPHTHA.Variables selected based on engineering criteria.

RESULS FOR ASTM D86, 85%GOL

Automatic model variable selection

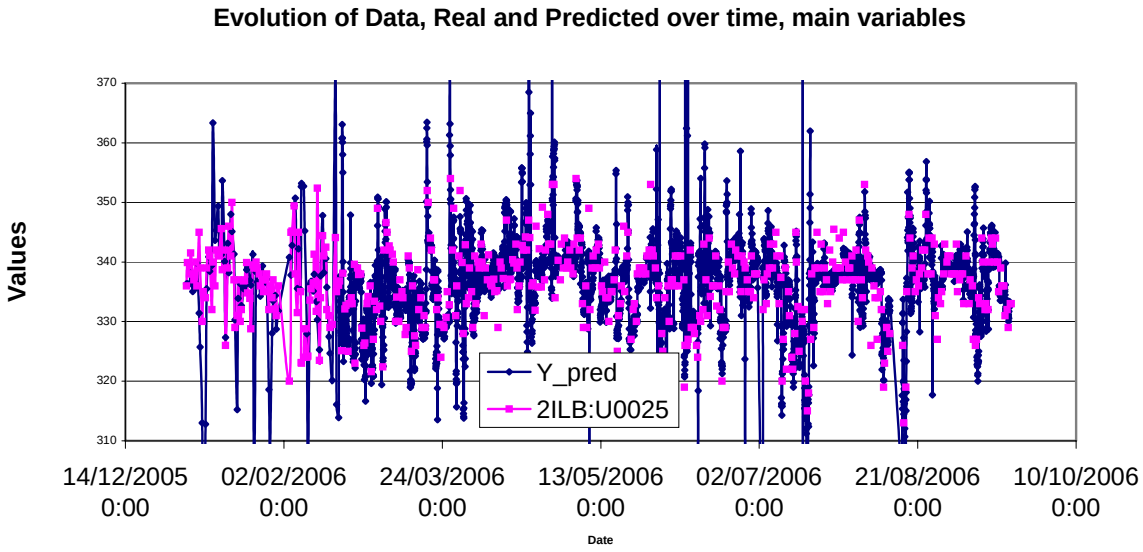

Figure 3: ASTM D86, 85%BPGOL. Automatic model variable selection.

Variables selected based on engineering criteria

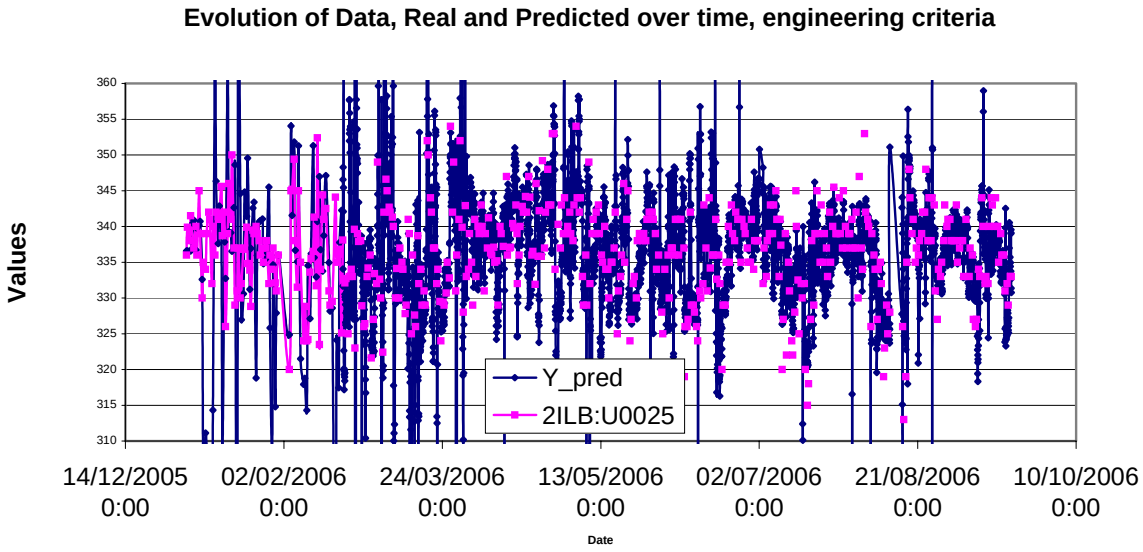

Figure 4: ASTM D86, 85%BPGOL.Variables selected based on engineering criteria.

CONCLUSIONS

The preliminary results obtained in [22] are confirmed here. The results obtained with this tool are stable and probably can go online.

The results obtained demonstrate that dynamic behavior suppose in [22] was wrong and the data have to be delay more than one hour and in some cases. This delay is not constant and depends on the variable. For instance, density of the crude delays until three hours to get in the side streams quality. Other variables have smaller delays (one or two hours). The results obtained demonstrate also that engineering criteria is not the unique criteria that give a stable result but also a methodology can help. However, finding the inner relations between the data that explain the results is more difficult with this kind of selection.

The procedure used in this contribution that evolves and predicts have being found very convenient to solve problems of instrument and laboratory drift over time that were unsolved with other approaches and software tools without recalibration.

Future research will be directed towards practical implementation of this technique in an online estimator in closed loop. This estimator is going to be implemented online in one of the CEPSA oil refineries, in Tenerife, Spain.

REFERENCES

[1] Angelov, P. P., *Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems*, Springer-Verlag, Heidelberg: Germany, 2002, ISBN 3-7908-1457-1

[2] Bhat, S., Chatterjee, D., Saraf, N.: On line data processing and product properties prediction for crude distillation units. AIChe 2003 Spring National Meeting, New Orleans, Luisiana, 2003

[3] Barsamian, A. , Macías, J: Neural Network-based Inferential Property Predictors. Hydrocarbon Processing, vol. 77, No. 10 (1998)

[4] Delgado, S, Macías, J.: Control of Reid Vapor Pressure with inferential estimators. VII Congreso Mediterráneo de Ingeniería Química, Barcelona, 1996

[5] Geladi, P.; Kowalski, B.R., An example of 2 block Predictive Partial Least Squares Regression with Simulated data, Analytica Chimica Acta, 185, 19-32, 1986

[6] Geladi, P.; Kowalski, B.R., Partial Least Squares Regression: a tutorial, Analytica Chimica Acta, 185, 1-17,1986

[7] Kaspar, M. H. :"Dynamic PLS modelling for Process Control", Chemical Engineering Science , 48 , 3447-3461 . (1993).

[8] Kaspar, M. H.; Ray, W.H. :"Chemometric Methods for Process Monitoring and High Performance Controller Design", AIChE J. , 38(10) , 1593-1608 . (1992).

[9] Kaspar, M. H.; Ray, W.H., Chemometric Methods for Process Monitoring and High Performance Controller Design, AIChE J., 38(10), 1593-1608, 1992

[10] Lambert, S.: Using Advanced data mining to improve process control and energy management. NPCS, New Practice, Case Study 128, www.energy-efficiency.gov.uk, February 2002

[11] Macias, J.: “Aplicación de la Técnica de Calibración Multivariante PLS al Control y Modelado de Procesos”, PhD Thesis, University of La Laguna, (1998).

[12] Macías J., Feliu J. A.: Dynamic Study of Inferential Sensor (NN) in quality prediction of crude oil distillation tower side streams. 11th European Symposium on Computer Aided Process Engineering, ESCAPE 11, Kolding, Denmark, May 2001. ISBN 0-444-50709-4

[13] Rodriguez, M.; Rubio, J.J., Jimenez, A.: Naphta Cut Point Inference using Neural Networks. ttp:\\www.diquima.upm.es/˜mrod/eann7.html, Universidad Politécnica de Madrid

[14] Angelov, P., X. Zhou, "Evolving Fuzzy Systems from Data Streams in Real-Time" *Proceedings of 2nd Intern. Symposium on Evolving Fuzzy Systems, EFS06*, 7-9 September 2006, Lake District, UK, IEEE Press, pp.29-35.

[15] Angelov, P.**,** An Approach for Fuzzy Rule-base Adaptation using On-line Clustering, *International Journal of Approximate Reasoning*, vol. 35 (3), March 2004, pp. 275-289.

[16] Chiu, S. L., "Fuzzy model identification based on cluster estimation," *Journal of Intelligent and Fuzzy Systems*, vol.2, pp. 267-278, 1994.

[17] Ljung L. (1987) *System Identification: Theory for the User*, New Jersey, USA: Prentice-Hall.

[18] http://www.astm.org.

[19] Nature Inspired Self-Healing Soft Sensors for Process Industry, task Force to NiSiS, available at http://www.nisis.risk-technologies.com/ (iybsetnz2wnqll55qij2qfbj)/filedown.aspx?File=58

[20] Angelov P., D. Filev, An Approach to On-line Identification of Takagi-Sugeno Fuzzy Models, *IEEE Transactions on System, Man, and Cybernetics, part B - Cybernetics*, vol.34, No1, 2004, pp.484-498.

[21] Takagi, T., M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", *IEEE Trans. on Syst., Man & Cybernetics,* vol. 15, pp. 116-132, 1985.

[22] Macias, J., Angelov, P., Zhou, Y., "Predicting Quality of the Crude Oil Distillation using Evolving Takagi-Sugeno Fuzzy Models", " *Proceedings of 2nd Intern. Symposium on Evolving Fuzzy Systems, EFS06*, 7-9 September 2006, Lake District, UK, IEEE Press, ISBN 0-7803-9719-3, pp.214-220, *'best industry-related paper' award*