

On-line Data Processing and Product Properties Prediction for Crude Distillation Units

Shrikant Bhat, Tirtha Chatterjee, Deoki. N. Saraf
Process Control Laboratory
Department of Chemical Engineering
IIT-Kanpur, 208016, India

Prepared for presentation at
AICHE 2003 Spring National meeting New Orleans, Louisiana March 30 – April 3, 2003

Using Information Technology for Increased Profitability and Productivity
Predictive IT for Increased Productivity and Profitability

Copyright © Shrikant Bhat, Tirtha Chatterjee, Deoki. N. Saraf

Abstract

This paper presents a model that uses the crude true boiling point (TBP) curve and other routinely made measurements on crude distillation unit (CDU) for product properties prediction. On-line data are first checked for existence of steady state using a technique developed by Cao and Rhinehart and modified in the present study. The performance of this technique depends upon three filter constants which have been optimized to reduce both Type I and Type II errors and at the same time facilitate early detection of both steady as well as unsteady states. After checking for steady state, the algorithm makes use of past measurements and Kalman filter to detect the presence of gross error if any and to estimate its magnitude. The filter simultaneously carries out data reconciliation minimizing the random errors using both the spatial and temporal redundancies associated with the measurements.

The preprocessed data are used for product properties prediction. Top distillate, side-stripper (SS) product draw plate temperatures and flash zone temperature are corrected for partial pressure of hydrocarbons which represent equilibrium flash vaporization (EFV) temperatures of the same products. These EFVs are converted to product TBPs and are superimposed on the crude TBP curve which enables estimation of missing TBP end-points. The terminal product TBPs are joined by straight lines to estimate the intermediate TBP temperatures. However, since these TBP curves are usually 'S'-shaped with most of the curvature being present at the two ends, the end points are shifted vertically towards the crude TBP curve using a shifting technique before straight line joining. These TBP temperatures along with other available information are correlated with the desired product properties.

Several properties have been predicted using the above procedure which include product densities, Flash Points, Reid Vapor Pressure (RVP), Freeze Point, Pour Point etc. which can be repeated every minute. The soft sensor has been validated using data from two different refineries. Finally the entire algorithm including preprocessing has been applied to 24 hours on-line data and results presented.

Introduction

Petroleum industry is one of the most prolific and dynamic industries of modern civilization. Because of a highly competitive market and stringent environmental laws, strict quality control of refinery products is a must. Crude distillation unit (CDU) is one through which entire crude entering a refinery must be processed. So a close monitoring and control of CDU product properties will help us in controlling the properties of final refinery products.

Crude oil as well as all the CDU products are complex mixtures of hydrocarbons, so it is not convenient to characterize them in terms of individual components. Moreover, it is often sufficient to characterize refinery products in terms of certain gross properties such as Reid Vapor Pressure for volatile products, Flash Point for light distillate, Pour Point for heavy distillate etc. Continuous control of the unit demands that these properties should be measured on-line so that it can be effectively controlled through a feedback mechanism. Unfortunately no suitable on-line hardware sensors are available and laboratory measurement procedures are tedious and time consuming. Cut-point temperatures (Horn, 1980) are generally used to control atmospheric towers which do not account for changes in the composition and hence can not control product properties. Alternatively, mathematical models (referred to as soft sensors) are used to estimate product properties which use only easily measurable secondary variables as inputs.

On-line data, used by the soft sensor, actually represent a dynamic system. But most of the refinery simulators use steady state model of the CDU for monitoring including product property estimation. Therefore, steady state identification (SSI) is the starting step for on-line data processing. Once the existence of steady state is ensured, the next step is detection of presence of gross error arising from instrument failure, measurement bias, presence of leaks etc. The objective of gross error detection is not only to identify its presence but also to estimate its magnitude so that the correction can be applied and further data processing carried out till fault is rectified and the instrument is recalibrated. The final step is data reconciliation. It aims to eliminate random noise associated with the measurements so that they satisfy process constraints such as material and energy balances.

The property prediction methodology makes use of crude true boiling point (TBP) curve (available from crude assay) and error-free steady state data such as operating flowrates, temperatures and pressures. Based on draw plate temperatures of different products the algorithm calculates six equilibrium flash vaporization (EFV) temperatures. These EFVs are then converted to TBPs which, when superimposed on crude TBP curves, allow estimation of six more product TBP temperatures. Then TBP curves are approximated by straight lines for ease of interpolation. The product TBP temperatures are related to various product properties either directly or after conversion to ASTM (American Society for Testing Materials) temperatures. Thus without making the simulation over the entire column, product properties are determined in an easy and instant way.

Literature Review

There are various techniques available for steady state identification. Cao and Rhinehart (1995), discussed modification of the F-test type of statistics (Crow et al., 1955) so that the data can be treated sequentially for steady state identification without the need to select a time window which is required in most of the other methods (Narasimhan et al., 1986 & 1987; Akelman, 1994; Loar, 1994). The modification involved incorporating exponentially weighted moving average filter to calculate average

and the variance by two different methods. For this method to be effective online, the filter constants must be chosen judiciously and optimally and this was one of the objectives of the present study.

Both gross error detection (GED) and data reconciliation (DR) are the issues addressed simultaneously in chemical engineering literature. Data reconciliation, with the data containing only random errors and subjected to balance constraints, by means of constrained least squared minimization procedure was first introduced by Kuhn and Davidson (1961). Presence of gross errors vitiates the data reconciliation exercise and hence, it is necessary to eliminate gross errors before attempting data reconciliation. Of various methods developed for gross error detection, the generalized likelihood ratio (GLR) (Narasimhan and Mah, 1987) method not only detects the existence but estimates its magnitude which makes it very useful. Biegler and Tjoa (1991) presented a simultaneous strategy for data reconciliation and gross error identification. Principle Component Analysis (Crowe and Tong, 1995) causes reduction in the dimensions of the problem and is able to detect subtle gross errors which otherwise remain undetected by other methods.

All the above methods exploit the spatial redundancy available with the system to carry out DR and GED. However, for CDU, where there are more than one streams entering and leaving the unit, available spatial redundancy can not be effectively made use of and most of the existing methods fail. This calls for use of temporal redundancy for data reconciliation. In this work we apply the approach as applied to quasi-steady states using Kalman filter by Stanley and Mah (1977) and couple it with a simple comparison strategy making use of the previously reconciled data.

As we mentioned earlier, multicomponent distillation models, based on equilibrium stage relations, assume the existence of steady state to simulate the crude distillation unit (Boston and Sullivan, 1974; Russel 1983; Kumar et al., 2001). For a mixture of C components Boston and Sullivan (1974) developed an algorithm which was later modified by Russel (1983). Kumar et al., (2001) developed a model using (C+3) iteration variables similar to that by Russel but the choice of variables makes it more robust and efficient. This model will be referred to as the ‘Simulator’ in latter sections. In order to be able to use the steady state model for on-line application it is necessary to tune the model frequently using on-line operating data (Dave et al., 2003). The tuned simulator provides product TBP temperatures and densities which are then empirically correlated to various product properties. But for on-line tuning, the authors suggested use of genetic algorithm, a random search technique, or sequential quadratic programming (SQP), a non-linear optimization technique. Either of the methods is time consuming and hence property prediction can not be repeated faster than every 10 to 12 minutes which makes it unacceptably slow for feedback control.

Processing of On-line Operating Data

As discussed above processing of operating data comprises of three operations: identification of steady state, gross error detection and data reconciliation. These three steps are discussed below.

Steady State Identification

The method of Cao and Rhinehart (1995) involves calculation of filtered value of the online measurements and variance calculated by two different approaches; one using filtered mean square

deviation from the previous filtered value, $v_{f,i}^2$, and the other using filtered mean square difference of successive data, $d_{f,i}^2$.

The filtered value $Z_{f,i}$ of the measurement value Z_i is given as

$$Z_{f,i} = L_1 Z_i + (1 - L_1) Z_{f,i-1} \quad (1)$$

The filtered square deviation of the measurement value from the previous filtered value is given by

$$v_{f,i}^2 = L_2 (Z_i - Z_{f,i-1})^2 + (1 - L_2) v_{f,i-1}^2 \quad (2)$$

while the filtered square difference of successive data is given by

$$d_{f,i}^2 = L_3 (Z_i - Z_{i-1})^2 + (1 - L_3) d_{f,i-1}^2 \quad (3)$$

The R-statistics, which is used to ascertain existence of steady state, is defined as

$$R = \frac{(2 - L_1) v_{f,i}^2}{d_{f,i}^2} \quad (4)$$

where L_1 is the filter constant for filtering incoming data, while L_2 and L_3 are for the variances by two different methods.

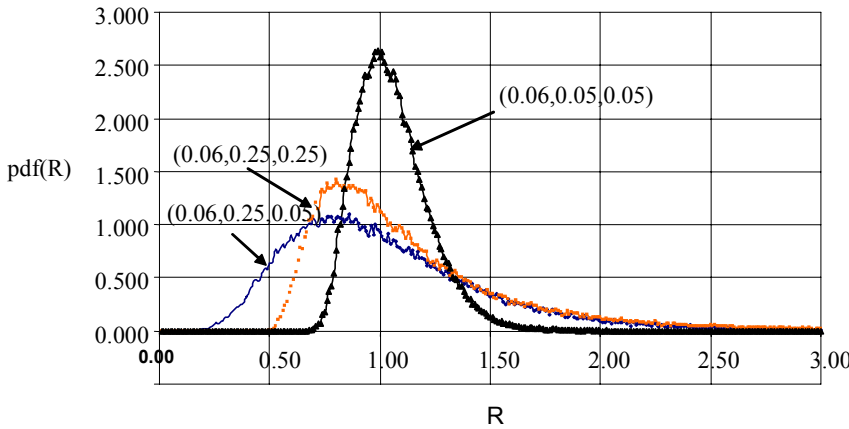


Figure 1. The probability distribution function (pdf) of R for equal and unequal sets of L_2 and L_3

R-value is compared with the critical value of the R-statistic. If R value is greater than R_{crit} , the process is considered to be not at steady state at an associated level of significance. In this scheme, Cao and Rhinehart proposed L_2 and L_3 , which are the filter constants for the two variances, to be selected equal. Now if we weigh the two variances unequally by keeping lower value of L_3 and higher value of L_2 , it is observed that we are able to reduce Type I and Type II errors as well as detection lag. Fig. 1 shows the probability distribution function of R for equal and unequal L_2 and L_3 with the same L_1 . As seen in this figure unequal values of L_2 and L_3 only increase the spread of the R-statistics both below and above $R=1$.

The important factors which gauge the effectiveness of the steady state identifier are the extents of Type I and Type II errors, and the delay in the steady as well as unsteady state detection. Type I errors can be contained by selecting suitable R_{crit} for a given set of filter constants (Cao and Rhinehart, 1995). In the present study Type I error was fixed at 1% level of significance. The selection of R_{crit} for a given set of filter constants also fixes the extent of Type II errors and both the steady and unsteady state detection lags for a given slope of ramp. An important aspect of this analysis is that the Type II error is

calculated based on a ramp change continuously increasing in magnitude within a definite range, say from 100 to 200°C for a temperature value. Two factors contribute to the Type II errors; one corresponding to the delay in unsteady state detection and the second due to false indication of steady state. If we choose to have detection of unsteady state within steps in which the true value of the process changes by 2.5 standard deviations (σ), and select the filter constants in appropriate manner, Type II error contribution can be predominantly due to delay in detection of unsteady state. Beyond 2.5σ , it is expected that the Type II errors will be predominantly because of false detection of steady state. This criterion is used in setting allowable Type II errors to 15 cases out of 1000 to cross 2.5σ limit and the filter constants are chosen accordingly. This also helps in setting the minimum slope of the ramp to be detected as unsteady state so that any change to the extent smaller than this is considered as steady state.

Optimization of filter constants: It may be pointed out here that in absence of any known analytical or empirical relationship between the filter constants and the two types of errors and delay times, it is not possible to use standard optimization techniques to find the optimal values of L_1 , L_2 and L_3 . Optimization through enumeration following one at a time approach has, therefore, been adopted. After understanding the significance and relative importance of each of these filter constants, the following procedure was evolved.

- 1) Select a value of L_3 (say 0.01).
- 2) Select a value of L_2 (say 0.05).
- 3) For these values of L_2 and L_3 , start with a low value of L_1 (say 0.02), and calculate R_{crit} , Type II errors as well steady state detection lag.
- 4) Increase L_1 and repeat calculations of step 3 until allowable Type II error limit is crossed.
- 5) Increment L_2 by 0.05 and then return to step 3. Go on incrementing L_2 till the local minimum in terms of earlier detection of steady state is obtained corresponding to given value of L_3 .
- 6) Increment L_3 and return to step 2 till global minimum is obtained.

Selection of measurements for SSI: For any process unit it is difficult to get a single measurement which would reflect all sorts of unsteady state effects. Therefore it is necessary to select more than one measurement as unsteady state markers. If any one of these measurements shows transient behavior then unsteady state can be inferred for that unit. However, to avoid the possibility that this measurement is corrupted with gross error leading to false conclusion, a second variable should preferably be chosen for which the sensor is in physical proximity of the first. The chances that both these measurements will have gross error are small. In case of CDU, it is possible to have local disturbances (unsteady conditions) which may not be reflected in other parts of the unit. To overcome this problem the temperatures of seven different stages distributed throughout the entire length of the column were selected for SSI for the column under study.

Gross Error Detection and Data Reconciliation

Almost all the techniques in the gross error detection (GED) and data reconciliation (DR) literature are based on spatial redundancy of measurements. In the case of CDU, the spatial redundancy is available only with respect to the incoming feed and the six outgoing products. The circulating reflux flows as well as all the temperatures do not form the part of spatially redundant system. Under these circumstances, the temporal redundancy available with these data can be exploited to some advantage.

Data reconciliation using Kalman filter: There are a total of 12 flow measurements made in the CDU under consideration. These flow measurements can be put through Kalman filter as suggested by Stanley and Mah (1977). Through the analysis given in this reference, the Kalman filter equations can be formulated for the 11 state variables and the 12 flow measurements. In this case the process noise covariance matrix Q is treated as filter design parameter and can be chosen as diagonal for ease of computation and simplicity. The measurement noise can also be assumed to be uncorrelated and the variance-covariance matrix M can be taken to be diagonal. It is necessary to adjust the M/Q ratio to allow for the drift in the measurements, if any, occurring during steady state regime. The maximum extent of the drift is usually of the order of 0.05 to 0.1 % per minute.

Gross error detection: The gross error detection for all these cases can be applied based on simple comparison strategy. To start with, we assume that there is no gross error present in any measurement, which can be assured through calibration of sensors at the very initial stages. Each time it is observed whether the next measurement value is within 6 standard deviation limits. If it is within the limit, the measurement is sent for further data reconciliation exercise. If the limit is violated, then a gross error is identified and the magnitude of the gross error is taken as the difference between the measured and the reference value. The reconciliation exercise is carried out with the previous reconciled value for this measurement. If the next measurement is more than 6 standard deviations away from the previous measurement value, a still developing gross error is expected and the difference between the reference value (the previous reconciled value) and this measurement value is considered as magnitude of gross error and used as the correction. If the measurement value remains within 6 standard deviation limit of the previous measurement value, then the average of these two measurements is taken and the difference between this averaged value and the reference value is taken as magnitude of gross error. The averaging is continued if measurements till 10 sampling intervals are found within 6 standard deviation limit from the previous measurement. Then this average value is used as the correction for gross error till any other gross error occurs or the instrument bias is corrected. Gross errors of the magnitude of 10 standard deviations can be detected reasonably accurately by this analysis. In case of CDU, a leak at any point except at the exit product streams would lead to unsteady state and hence would be taken care of during SSI exercise. The leaks associated with the outgoing product streams would be identified as measurement bias only.

Property Prediction Methodology

Property prediction methodology is mainly comprised of three parts. First is crude TBP reconciliation which is used in off-line mode. Extensive laboratory data are required for this reconciliation technique but it is immensely important for accurate properties prediction. Second one is product ASTM temperature calculation using routinely measured operating conditions (different flow rates, temperatures and pressures) and reconciled crude TBP curve. Thirdly, product ASTM temperatures and other available information are correlated to various product properties using empirical relations.

Crude TBP Reconciliation

True Boiling Point (TBP) curve is one of the most significant characteristic features of the feedstock. It decides the amounts of various fractionation products available from the crude as well as the composition and properties of these products. So accuracy of property prediction largely depends on the accuracy of the TBP curve used. Generally TBP data of a pure crude is available from its crude assay

which may not represent the crude being processed at a later time. The deviation may arise due to various reasons such as blending of different crudes, contamination of one crude with another in the storage tanks or even due to stratification. Feed TBP reconciliation is a useful method to modify the crude TBP curve so that it truly represents the crude being processed at that time. It is formulated as an optimization problem where one tries to adjust the crude TBP such that the simulator calculated product ASTM temperatures match closely with those measured in the laboratory. It is expected that the crude TBP curve, so reconciled using a large number of laboratory measured ASTM temperatures of different products, will be closer to the real feed being processed at that time as compared to the initial TBP curve available from the crude assay. The success of the property prediction depends largely on the success of the TBP reconciliation. So it is an integral part of the property prediction package. The TBP reconciliation problem can be mathematically expressed as:

Minimize

$$I = \sum (\text{Product TBP}_{\text{Measured}} - \text{Product TBP}_{\text{Calculated}})_i^2$$

Feed TBP

subject to:

$$\text{model equations and } \text{TBP}_{j \text{ volume\% of crude}} < \text{TBP}_{i \text{ volume\% of crude}} \quad \text{for } i > j \quad (5)$$

The product TBPs included in the objective function are those of all the distillate products. Sequential quadratic programming (SQP) was used for the minimization problem. Since laboratory measured product ASTM temperatures are required for reconciliation it is undertaken only occasionally in an off-line mode.

Product ASTM Temperature Calculation

In the present study an instant, on-line method is developed to estimate the product ASTM temperatures using the operating conditions of the CDU and the reconciled crude TBP curve. The present method starts with six operating temperatures in the main column. These are the temperatures at a) distillate draw tray (or, top tray), b) flash zone and c) side-stream column draws (four in number as there are four side-strippers present in the test unit). All these temperatures correspond to the hydrocarbon partial pressures prevailing in the column at appropriate locations. Distillate (the top product) is withdrawn in the vapor phase whereas all side-stream products are withdrawn in the liquid phase. So the top plate temperature actually denotes the equilibrium flash vaporization (EFV) dew point of the distillate at the corresponding hydrocarbon partial pressure. Similarly all side-stream column draw temperatures as well as flash zone temperature denote the EFV bubble points of the un-stripped products at the corresponding hydrocarbon partial pressures. However, EFV and TBP curves are available only at atmospheric pressure; hence, these temperatures have to be corrected for the pressure difference.

The partial pressure of the hydrocarbon vapors in the overhead, p_{top} , is given by Watkins (1981) as:

$$p_{\text{top}} = P_{\text{top}} [(F_{\text{ref}} + F_{\text{dl}}) / (F_{\text{o}} + S_{\text{T}})] \quad (6)$$

where F_{ref} , F_{dl} and F_{o} are the molar flow rates of the top plate reflux, liquid distillate and overhead vapor leaving the top of the column respectively. S_{T} is the molar flow rate of the total steam input to the column. It comprised of the steam input to the bottom of the main column as well as to the bottom of all the side-strippers. P_{top} is the total pressure at the top of the column. The partial pressure of the hydrocarbon product vapors at the flash zone, p_{fz} , is given by:

$$p_{fz} = P_{fz} [(F_L + F_{dl}) / (F_L + F_{dl} + F_{dv} + S_R)] \quad (7)$$

where F_L is the molar flow rate of the total liquid side-stripper products along with the overflash that are in vapor phase in the flash zone. F_{dv} is the molar flow rate of the vapor distillate product. S_R is the molar flow rate of the steam input to the bottom of the main column. P_{fz} is the total pressure at the flash zone. The partial pressure of the hydrocarbon vapors above the i th tray (from which k th side-stripper stream is withdrawn), $p_{ss,k}$, is given by:

$$p_{ss,k} = P_{ss,k} [(L_{r, (i-1)}) / (L_{r, (i-1)} + V_{ss,i} - F_{ss, (k-1)})] \quad (8)$$

where $L_{r, (i-1)}$ is the hydrocarbon liquid reflux flow to the i th (k th side-stripper stream column draw) tray (i.e. liquid flow from the tray above the side-stream draw tray). $V_{ss,i}$ is the molar flow rate of hydrocarbon vapor and steam leaving the i th tray. $F_{ss, (k-1)}$, the molar flow rate of the product drawn from the $(k-1)$ th side-stripper, is not included in the hydrocarbon vapor partial pressure calculation since it is near its critical temperature as it leaves the draw tray under consideration and hence, assumed to have no effect on partial pressure (Watkins, 1981). The tray temperatures measured in the plant are corrected to atmospheric pressure using Clasius Clayperon equation.

$$d (\ln P) / d (1/T) = K \quad (9)$$

where K is a constant for the hydrocarbon stream under consideration. The vapor pressure of hydrocarbon stream is calculated using an analytical model developed by Lee and Kesler (1980).

Calculation of Product TBPs: The EFV temperatures as calculated above are then converted to TBP temperatures using the correlation developed by Riazi and Daubert (1986). The six TBP temperatures, converted from EFV temperatures, correspond to top distillate final boiling point (FBP), special cut naphtha (SCN) initial boiling point (IBP), kerosene IBP, light gas oil (LGO) IBP, heavy gas oil (HGO) IBP and flash zone temperature. These temperatures, when superimposed on the feed TBP curve at their respective cumulative volume percent distilled, form the cut points of the different products.

Feed TBP temperatures at cumulative volume percent distilled of different products are obtained from the crude TBP curve. These temperatures are termed as ‘*Mid temperatures*’ (see Fig. 2). It is assumed that at the cut points the feed TBP points equally divide the TBP overlaps. i.e. the difference between the heavier product’s IBP and the mid temperature (at the cumulative volume percent distilled of the lighter product) is equal to the difference between the mid temperature and the lighter product’s FBP. For example, the difference between LGO IBP and the mid temperature calculated at the cumulative volume percent kerosene distilled is equal to the difference between that same mid temperature and kerosene FBP. Using the above method final boiling points of all the side stripper products are calculated. To calculate the FBP of HGO, difference between flash zone temperature and the mid temperature calculated at the cumulative volume percent of HGO is considered.

Around two volume percent of crude is usually collected as liquefied petroleum gas (LPG). LPG is an overhead product from the stabilizer. In the laboratory generally properties of the stabilizer bottom are reported. So feed TBP temperature at two volume percent distilled is calculated and that temperature is assigned as the IBP of the distillate (overhead product). Thus using feed TBP curve and six TBP temperatures calculated from the operating conditions, six other TBP temperatures are calculated.

Knowing both, the IBP and the FBP of a product, a straight line approximation is considered to represent the product TBP curve (see Fig. 2). But, since the product TBPs as well as product ASTMs have an ‘S’ type of curve, the straight line joining of IBP and FBP gives a poor approximation of the

intermediate volume percent product TBPs. So these points are shifted vertically towards the crude TBP curve. This is performed to provide a better linear approximation of the product TBPs. The magnitude of

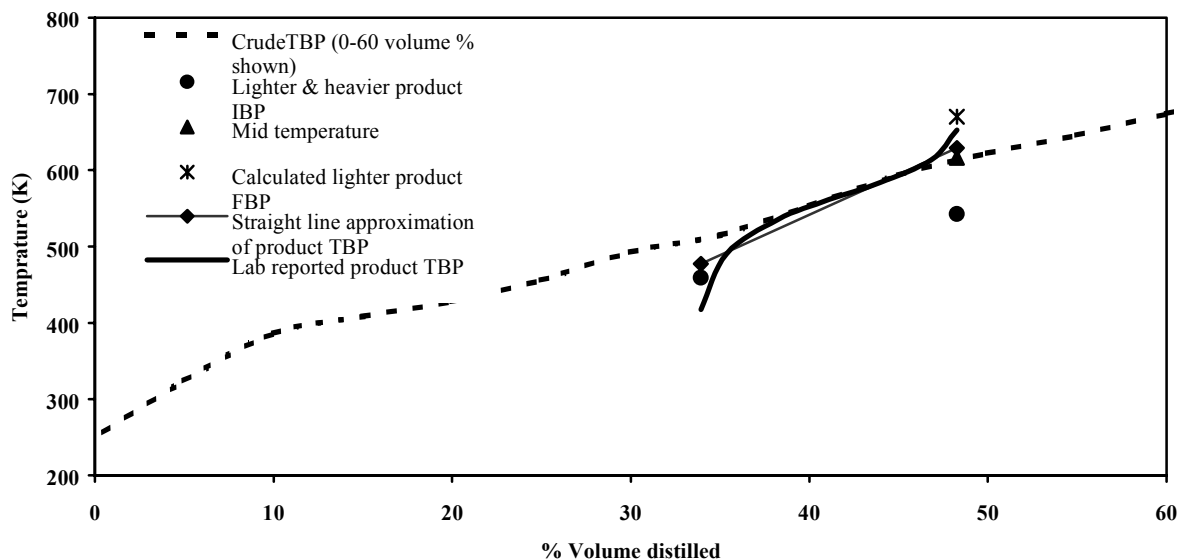


Figure 2: A graphical description of the method

shift (shifting constant) is obtained through a procedure (Chatterjee and Saraf, 2002) using the laboratory measured product ASTM temperatures. Since these measurements are available only once in a while, for a particular crude these shifting constants are stored and used until laboratory measured product ASTMs are available for that crude again.

Once IBPs and FBPs for all the products are modified using shifting constants, 10%, 30%, 50%, 70% and 90% product TBP temperatures are extracted from the linear relation and subsequently converted into corresponding product ASTM temperatures using Daubert's correlation (Daubert, 1994).

Property Prediction Package

The main aim of the present study is to make an on-line estimation of the product properties repetitively and at short time intervals of the order of about one minute. This facilitates the use of this information in feedback loop to control the operation of the CDU for strict product quality control. So more instantaneous is the property prediction, quicker and more effective is the control action.

Several important properties of petroleum products such as specific gravity, Reid Vapor Pressure (RVP), Flash Point, Pour Point and Freeze Point have been correlated using product TBP/ASTM temperatures and other available data. Presently specific gravity is predicted for all the distillate products. Flash point is predicted for all the side-stream products and pour point is predicted for light gas oil only. RVP is predicted only for the top distillate. As an example of the correlations consider Flash Point prediction. Flash Points (T_F) for Kerosene, LGO and HGO have been correlated as follows:

$$1/T_F = 0.076204 - 4.17015/T_{ASTM10\%} - 0.01043 \ln(T_{ASTM10\%}) + 0.000257 \ln(SG_{prod}) \quad (10)$$

where $T_{ASTM 10\%}$ is product 10 volume % ASTM temperature in K. SG_{prod} is product specific gravity. Flash Point calculated using the above correlation is in K. Though for the crudes tested the above

correlation works satisfactorily the authors believe that introduction of some other parameters such as composition of hydrocarbon in the product would improve the Flash Point prediction substantially which is discussed briefly in later section.

Results and Discussion

Steady State Identification

For the trial case in which the temperature is changing from 100 to 200 °C, linearly with a slope of 0.25 °C/min, the trend of optimum search for average of 2 and 4 successive data is reported in Table 1. 1000 such trials were tested and average of these 1000 trials is reported. Each trial consisted of 4000 steps (15 sec each) - the first 400 steps representing a constant value 100 °C with associated variance of 0.5 and autocorrelation between successive data (generated through first order autoregressive process) to be 0.05. Then there are 1600 steps generated with the slope of 0.0625 per step with the additional noise corresponding to variance 0.5. From steps 2000 to 4000 we have data representing a constant value of 200 °C with the variance of the

Table 1. A comparison of delay times for equal and unequal values of L_2 and L_3 for average of 2 and 4 readings. (autocorrelation between every 15 sec data at 0.05)

	$L_2 \neq L_3$					$L_2 = L_3$				
Average	L_1	L_2	L_3	US to SS	SS to US	L_1	L_2	L_3	US to SS	SS to US
2	0.06	0.35	0.05	1017 (1000)	209 (200)	0.052	0.2	0.2	1029 (1000)	209 (200)
<i>Slope = 0.125 / step</i>					<i>Lag 8.5 min</i>					
						<i>Lag 14.5 min</i>				
4	0.16	0.5	0.04	505 (500)	105 (100)	0.15	0.24	0.24	513 (500)	105 (100)
<i>Slope = 0.25 / step</i>					<i>Lag 5 min</i>					
						<i>Lag 13 min</i>				

associated noise at 0.5 and autocorrelation 0.05. For simulation based on the average of the two data in succession, the first steady state is from 1st to 200th step. The unsteady state is from 200th to 1000th step after which next steady state follows till 2000th step. For average of 4 successive data, the first steady state is upto 100th step after which unsteady state follows till 500th step. The second steady state is upto 1000th step. In this case the maximum unsteady state detection lag was allowed to be equal to number of steps in which the true value changes by 2.5 times standard deviation. For the change occurring with the given slope of 0.25 °C/min, for average of 2 and 4 successive data, the unsteady state detection occurs at 209th step (4.5 min lag) and 105th step (5 min lag) respectively for both the equal and unequal L_2, L_3 . The steady state detection, however, occurs earlier for unequal L_2, L_3 in both the cases as shown in the table.

Validation of the Property Prediction Methodology

We present some results to validate the properties prediction methodology. The algorithm was tested with several data sets involving seven different types of crudes for which plant operating data from two different refineries as well as laboratory measured properties were available. Data sets are classified into two categories. One is ‘*training set*’ and the other is ‘*test set*’. *Training sets* are those for which laboratory measured detailed product properties and product ASTM temperature profiles were available. These sets were employed to reconcile the crude TBP as well as to generate shifting constants. For a particular type of crude, reconciled crude TBP and shifting constants, generated from the *training sets*, were used to predict product properties for *test sets*.

For *training set* the product ASTM temperature match with the laboratory reported data is quite good which is to be expected in view of the crude TBP curve being reconciled using laboratory measured product ASTM temperatures. In most of the cases the average deviation between the experimental and predicted data is $\pm 1^{\circ}$ C (not shown). But a few larger deviations are encountered for HGO ASTM temperatures. The reason behind this might be that a relatively small amount of total crude (less than 10 volume %) is drawn as HGO and moreover the portion of crude corresponding to HGO is quite steep. So our straight line approximation for real

Table 2: Comparison of model predicted product properties with measured values

Properties	Training set		Test set #1		Test set #2	
	Lab	Model	Lab	Model	Lab	Model
SRN/UN						
Specific gravity	0.6859	0.6919	0.6854	0.6917	0.6826	0.6906
RVP(KPa)	64.0	60.5	65.7	61.2	61.0	66.4
SCN/HN						
Specific gravity	0.7576	0.7556	0.757	0.7546	0.7563	0.752
Flash Point(deg C)	N.R.	17.4	N.R.	16.4	N.R.	13.2
ATF/SK						
Specific gravity	0.7913	0.7906	0.7912	0.7906	0.7905	0.7859
Flash Point(deg C)	40.0	40.9	39.5	40.9	38.5	39.8
Freeze Point(deg C)	-61.0	-60.3	-58.0	-60.3	-61.0	-62.6
LGO						
Specific gravity	0.8405	0.8392	0.8413	0.8395	0.8396	0.8367
Flash Point (deg C)	60.0	60.4	62.0	60.7	60.0	59.4
Pour Point (deg C)	-6.0	-5.9	-6.0	-6.0	-12.0	-10.5
HGO						
Specific gravity	0.8785	0.8788	0.8795	0.8802	0.8807	0.8731
Flash Point (deg C)	114.0	124.0	121.0	125.4	118.0	115.3

N.R. Not reported

life ‘S’ shaped HGO ASTM curve may fail. For the *test sets*, the average deviation between the experimental and the predicted ASTM temperatures is in the range of $\pm 5^{\circ}$ C. Higher average deviation

for *test sets* compared to *training sets* may be expected since any measurement errors in the *training set* are likely to lead to higher deviations in predictions in the *test sets*. Using the predicted ASTM temperatures with the property prediction package the product properties were predicted. Results for only one crude have been included here because of space limitations. The computed product properties are compared with the laboratory measured values in Table 2 for one *training* and two *test sets*.

The predicted properties show a reasonable match with the corresponding laboratory measured values. Product specific gravity prediction works well and predicts up to two decimal places successfully in most of the cases. The predicted Reid vapor pressure (RVP) matches satisfactorily with the experimental data for most of the cases. The marginal deviation may be due to experimental errors. Flash Point prediction also falls within $\pm 2^{\circ}$ C of the experimental data. Though for not a single case experimental Flash Point data of SCN is reported, the model predicted SCN Flash Point falls in the range commonly encountered in the literature. The predicted Freeze Point values for ATF rarely deviated more than $\pm 2^{\circ}$ C.

It is important to mention here that raw operating data were used for properties prediction for both training as well as test sets. The refinery does not have any on-line data processing and since operating data corresponding to only time instants when samples were withdrawn for laboratory measurements were made available, we could not apply our above discussed methodology either. As will be seen latter, data preprocessing can significantly improve the performance of the soft sensors.

Table 3: Evaluating product properties prediction correlations

Properties	Specific gravity		Flash Point		Pour Point	
Range	0.6826 – 0.8928		33.0 – 134.0 ⁰ C		-12.0 – 39.0 ⁰ C	
No. of data points used	62		96		15	
Correlations	Absolute deviation		Absolute deviation (⁰ C)		Absolute deviation (⁰ C)	
	Avg.	Max.	Avg.	Max.	Avg.	Max.
Present	0.0032	0.0373	1.575	7.9	2.206	8.8
Riazi-Daubert (1986)	0.0169	0.0435	-	-	-	-
Riazi-Daubert (1987)	-	-	9.155	35.7	8.52	24.1
Chakrabarty (1997)	-	-	-	-	14.88	27.6

$$\text{Absolute deviation} = |\text{Predicted property} - \text{experimental property}|$$

$$\text{Average deviation} = (\sum \text{absolute deviation}) / (\text{no. of data points})$$

Table 3 shows the statistics of the properties prediction correlations developed in this study. Also included are the results of other correlations for comparison. Though the methodology of predicting product properties has worked satisfactorily for all the crudes tested, the authors feel that still much

scope is left for introduction of nature of crude as a parameter in prediction of various properties. As we know that the distribution of saturated and unsaturated hydrocarbons differs from crude to crude, using that information one may improve the prediction capability, especially for Flash Point and Pour Point. For example, we found the introduction of wt % aromatics present in the kerosene improved its Flash Point prediction significantly (not shown here). For prediction of product ASTM profiles where linear approximation fails for HGO, a scheme with multiple straight lines instead of a single straight line may also be investigated.

Case Study

Crude distillation unit operating data from an existing refinery at 5-minute interval were available for a period of 24 hrs. For steady state identification we got a fresh set of filter constants for the temperature range changing from 100 to 200⁰C with the slope of the ramp to be detected at 1.25⁰C / step (5 min) and the maximum variance associated with the data at 1 as was observed from the raw data. The optimum values of the filter constants were obtained at 0.18, 0.45 and 0.05 for L₁, L₂ and L₃ respectively. In order that the detection of the unsteady state be occurring within 2.5 standard deviations (delay equivalent to two steps), the steady state detection is too much delayed. So the delay equivalent to 3 steps (3.25 standard deviations) was chosen. For equal L₂L₃ case even the allowable steady state delay of 3.25 standard deviations was found to cause undue delay in steady state detection.

The raw data under consideration showed a non-random cyclic nature giving a look of high degree of autocorrelation which would normally be interpreted as unsteady state (Type I errors) by the algorithm. However, it is most unlikely that such autocorrelation is real. To overcome this problem simulated random noise was introduced by averaging every two successive readings and adding random

Table 4. Unsteady state detection steps for the case study.

Step	Temperatures showing unsteady state	Cause
42	SCN and Kero draw temperature	Change in Kero and LGO CR flows.
65	Kero and LGO draw temperatures	Change in HGO CR flow
96-102	SCN and Kero draw temperatures	Change in Top CR and LR flow
154-176	Column top, SCN, Kero, LGO, HGO draw temperatures	Change in Top CR flow
198-210	All the temperatures mainly column top and SCN and Kero draw	Sudden change in Top CR and UN flows
241-249	SCN, Kero, LGO & HGO draw temperature	Change in SCN and LGO draw flows

noise component equivalent to 80% of the standard deviation value. This helped in eliminating excessive Type I errors as well as captured the potential unsteady states. Although there might remain some stray

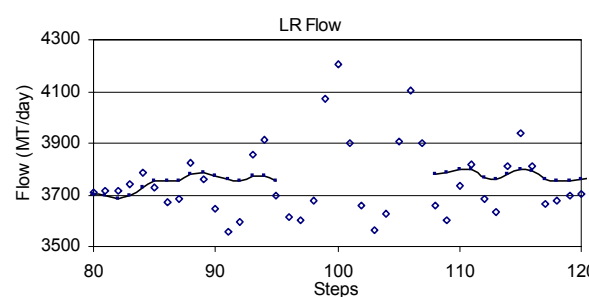
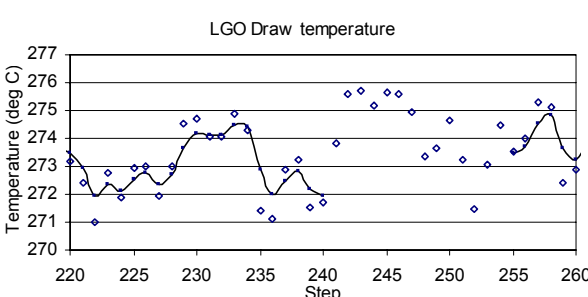
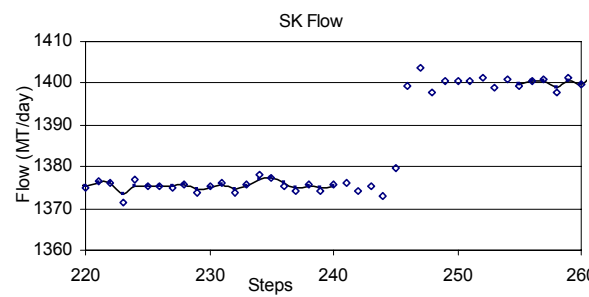
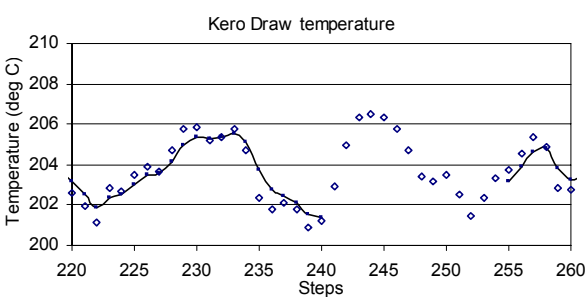
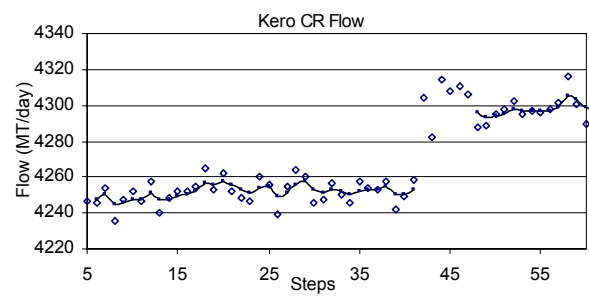
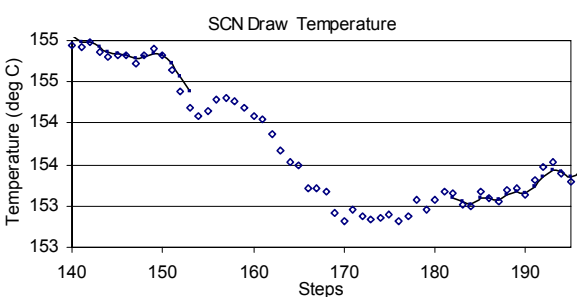
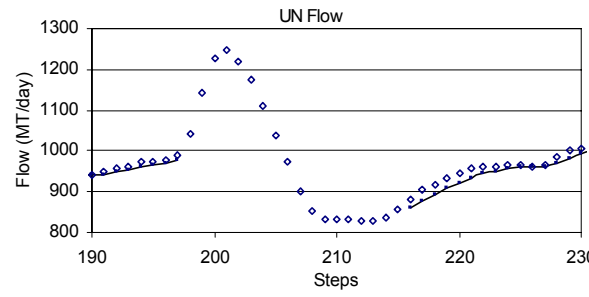
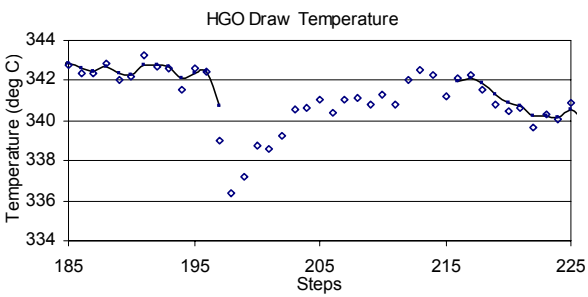
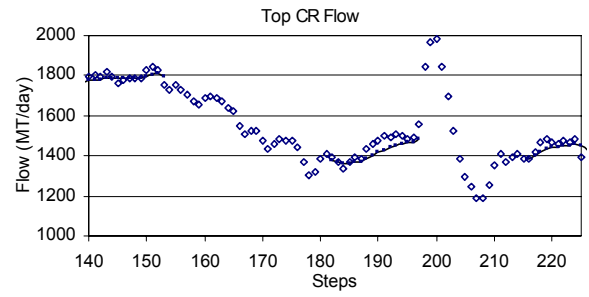
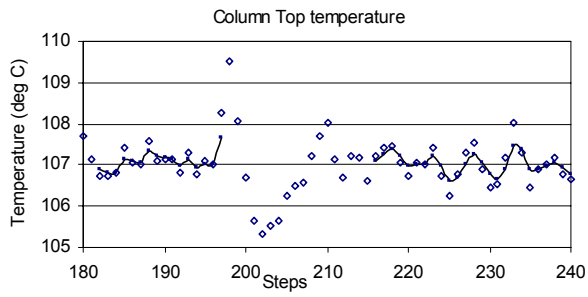


Figure 3. Raw and processed temperature and flow profiles

observations causing false indication of unsteady state, but getting simultaneously two adjacent temperature values showing unsteady states was less probable.

Table 4 shows the steps at which unsteady state was detected and the corresponding cause behind the unsteady state detection as seen from raw data. After every steady state attainment, the first 5 readings were averaged to provide the true value of the data to be used as reference value for further reconciliation. For reconciliation of temperatures, exponentially weighted moving average filter with filter constant at 0.5 was chosen so as to eliminate the measurement noise and at the same time follow the drift if any.

The data showed that the variance associated with the measurements varied from a very low value for certain product flows to a very high value for crude and LR flows. Based on the variance the flows were divided in three categories and a different R/Q ratio used for each. The reconciled values along with the raw data for various flows are shown in Figure 3. The open diamonds represent the raw data at 5 min intervals whereas the continuous curves joining dots shows the steady state data after reconciliation. The gaps in the curves represent the unsteady state and the delay times in detecting the return of steady states.

Each steady state data set, after getting reconciled, was used by the property prediction package to estimate product properties. The crude under consideration was Kuwait crude. Firstly crude TBP was reconciled by using laboratory measured product ASTM temperatures and converting to corresponding TBP temperatures. This data set formed the training set. Figure 4 shows the crude TBP before and after reconciliation. The reconciled TBP curve deviates from the original by utmost a few degrees only but this small change is important in causing a substantial improvement in the product ASTM match and hence in the properties. This training set was also used for generating the shifting constants which were subsequently used for rest of the data (test sets).

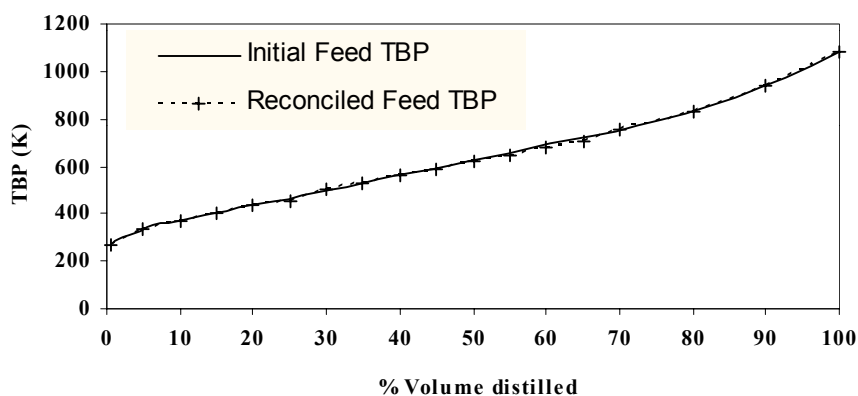
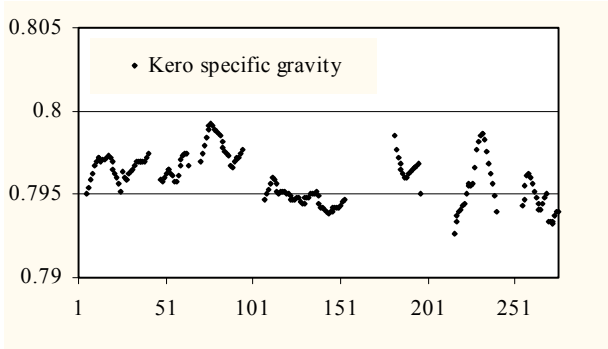
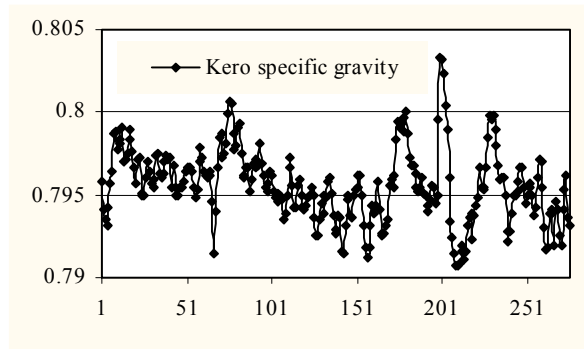
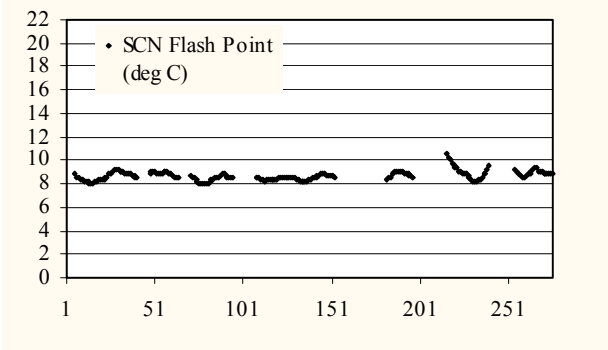
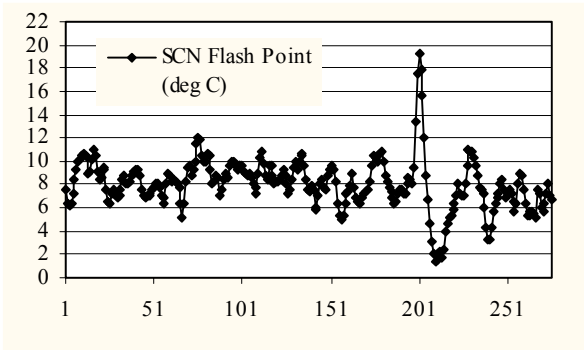
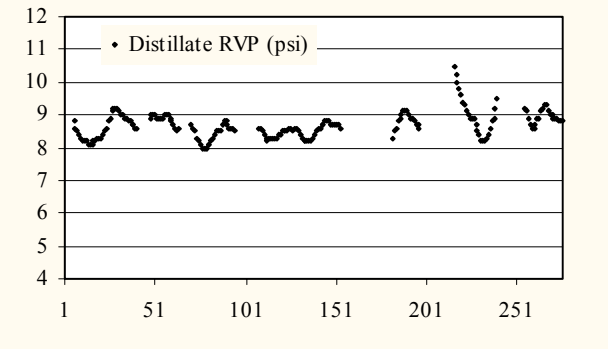
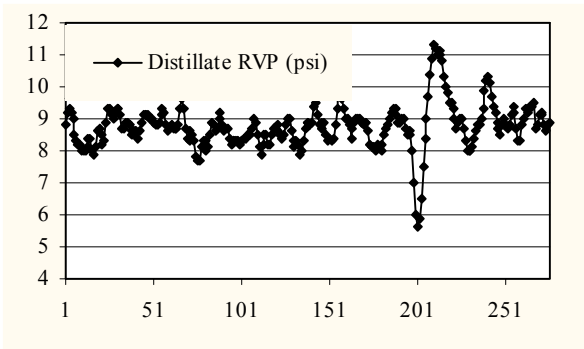
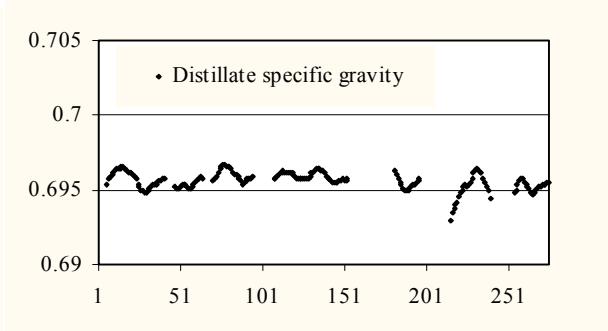
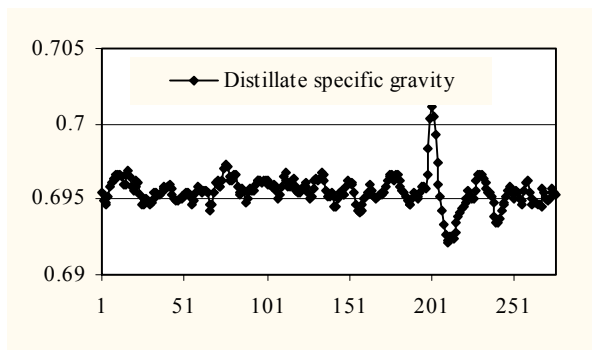


Figure 4: TBP reconciliation for Kuwait crude

As all on-line data were available to us in a file we ran our property prediction algorithm twice for each operating condition: once before preprocessing of the operating data and once after it. Though the present package predicts 12 properties (see Table 2) for 5 different CDU products, due to lack of space we only present 8 of these properties in Figure 5. The x-coordinate in all the plots of Figure 5 is time (in steps of 5 minutes) and y-coordinate represents different properties. The plots on the left column show variation of product properties over 24 hours if no preprocessing of data was carried out. The plots on the right show the same properties using preprocessed data. The gaps on the right hand plots denote the



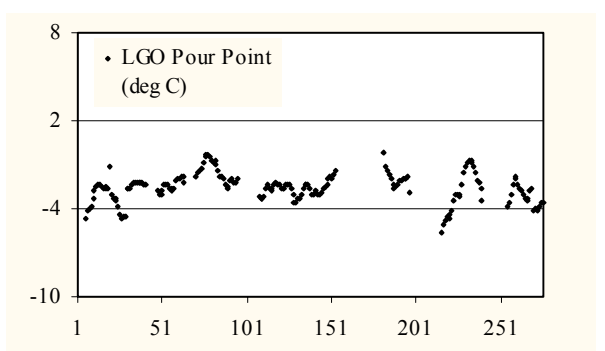
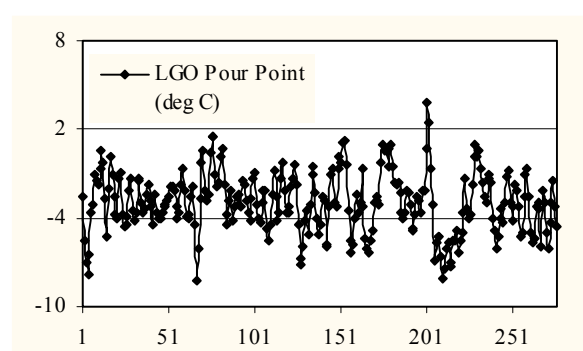
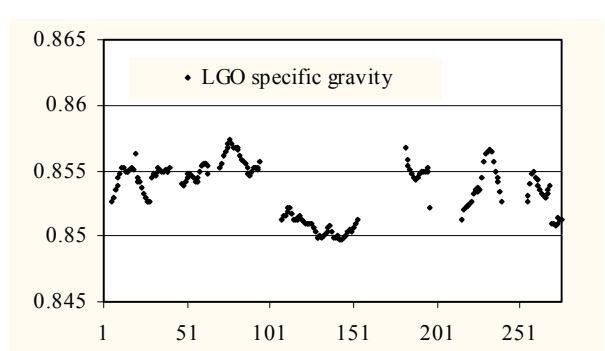
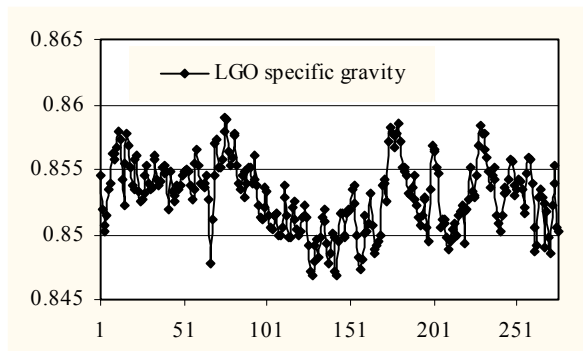
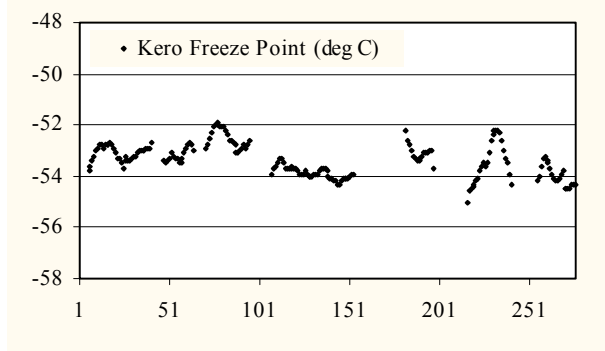
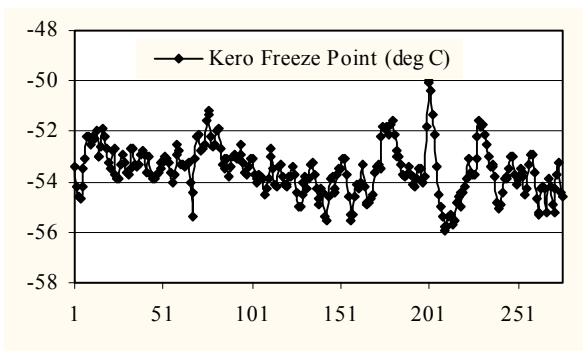
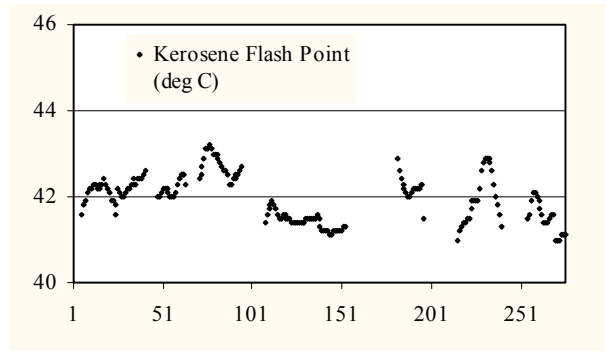
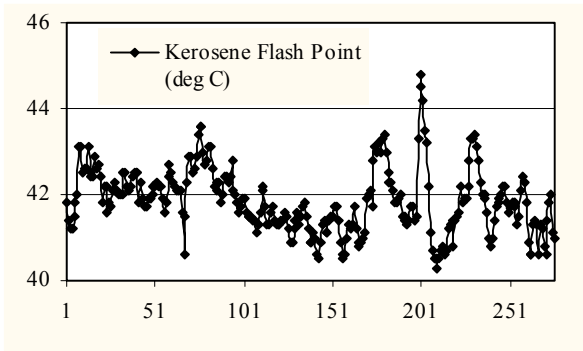


Figure 5: CDU product properties before and after data processing

existence of unsteady states during which the properties prediction algorithm remains idle. It is evident from the figure that though the properties follow the same trend using data with or without any preprocessing the variance of product properties are substantially reduced when preprocessing is undertaken. For example, Kero flash point has a range of 40.2 – 45°C using raw data which is reduced to 41-43.2°C.

Conclusion

An on-line data preprocessing algorithm has been developed and tested successfully using real raw data collected from a crude distillation unit in a refinery. The algorithm first checks for the existence of steady state and then reconciles the data after correcting for any gross errors if present. The CDU product properties methodology developed in this study allows repetitive prediction of several properties every minute. The accuracy of prediction is quite good and improves substantially with the use of preprocessed data.

References

- Alekman, S.L. Control for the Process Industries, Putman Publications, Chicago, IL., Vol VII, No. 11, November **1994**, 62.
- Biegler, L.T.; Tjoa, B. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Comput. Chem. Engg.* **1991**, 15, 679-690.
- Boston, J.F.; Sullivan, S.L.JR. A new class of solution methods for multicomponent, multistage separation process, *Can. J. Chem. Eng.* **1974**, 52, 52-63.
- Cao, S.; Rhinehart, R.R. An efficient method for on-line identification of steady state, *J. Process Control.* **1995**, 5(6), 363-374.
- Chakrabarti, S. M.Tech. Dissertation, Indian Institute of Tech., Kanpur, **1997**.
- Chatterjee, T; Saraf, D.N. Online estimation of product properties for crude distillation unit, submitted to *J. Process Control* (October, 2002).
- Crow, E.L., Davis, F.A.; Maxfield, M.W. Statistics Manual, Dover Publications, New York, NY, **1955**, 63.
- Crowe, C.M.; Tong, H. Detection of gross errors in data reconciliation by principal component analysis, *AIChE J.* **1995**, 41, 1712-1722.
- Daubert, T.E. Petroleum fraction distillation interconversions, *Hydrocarbon Process.* Sept. **1994**, 75-78.
- Dave, D. J.; Dabhiya, M.Z.; Satyadev, S.V.K.; Ganguly, S.; Saraf, D.N. Online tuning of a steady state crude distillation unit model for real time applications. *J. Process Control*, **2003**, 13(3), 267-282.

- Horn, L.D.V. Crude unit computer control – How good is it? *Hydrocarbon Process.* **1980**, 59.4, 145-148.
- Kuhn, D.R.; Davidson, H.; Computer control. II: Mathematics of control, *Chem. Eng. Prog.* **1961**, 57, 44–47.
- Kumar, V.; Sharma, A.; Roy Chowdhury, I.; Ganguly, S.; Saraf, D.N. A crude distillation unit model suitable for online applications. *Fuel Process. Technol.* **2001**, 73, 1-21.
- Lee, B.I.; Kesler, M.G.; Improve vapor pressures prediction, *Hydrocarbon Process.* July **1980**, 163-167.
- Loar, J. Control for the Process Industries, Putman Publications, Chicago, IL., Vol VII, No. 11, November **1994**, 62.
- Narasimhan, S.; Kao, C.N.; Mah, R.S.H. Detecting changes of steady state using mathematical theory of evidence, *AIChE J.* **1987**, 33(11), 1930-1933.
- Narasimhan, S.; Mah, R.S.H. Generalized likelihood ratio method for gross error detection, *AIChE J.*, **1987**, 33, 1514-1521.
- Narasimhan, S.; Mah, R.S.H.; Tamhane, A.C.; Woodward, J.W.; Hale, J.C. A composite statistical test for detecting changes of steady states, *AIChE J.* **1986**, 32(9), 1409-1418.
- Riazi, M.R.; Daubert, T.E. Analytical correlations interconvert distillation curve types, *Oil Gas J.* Aug. 25, **1986**, 51-57.
- Riazi, M.R.; Daubert, T.E. Predicting flash and pour points, *Hydrocarbon Process*, Sept. **1987**, 81-83
- Russel, R.A. A flexible and reliable method solves single tower and crude distillation problems, *Chem. Eng.* **1983**, Oct, 53-59.
- Stanley, G.M.; Mah, R.S.H. Estimation of flows and temperatures in process networks, *AIChE J.*, **1977**, 23(5), 642-650.
- Watkins, R.N. Petroleum Refinery Distillation, 2nd edition, *Gulf Publishing Company*, Houston, **1981**