

Engineering Models For Inferential Controls

By Y Zak Friedman, PhD
Principal consultant

INTRODUCTION

This paper describes a class of inferential control based on process engineering models working backwards. Normally process models start from known feed properties, and given yields and other unit conditions, they calculate product qualities. Our models start with unit conditions and yields, and aim to come up with feed properties. The models make use of the processing equipment as on-stream analyzers. Being true engineering models the inferential calculations are reliable and repeatable, plus they contain criteria for checking themselves for erroneous readings.

Why introduce inferential control in a process optimization conference? Because industry is setting itself for on-line optimization of the plant minute by minute, and the results of the optimization would be a set of control targets implemented immediately. That necessitates complete confidence in the optimization results. Do we have such a level of confidence? On-line optimization relies on on-line simulation, and confidence centers on the model's ability to duplicate process measurements. This requires first a high fidelity model, and second a very good knowledge of the unit feed properties. Fidelity can perhaps be achieved via elaborate equations and careful calibration, but our ability to assess process unit feed is limited. That imprecision has thwarted on-line optimization everywhere [1].

What will happen to the simulation results if feed properties are incorrectly assessed? Key model predictions will be different from instrument readings. At that time optimization becomes a pointless exercise. This is the context in which we address inferential calculations. Without them, there is no practical way to optimize a process in real time.

Some on-line simulations estimate feed properties from laboratory tests on the product. We find that approach unsatisfactory. Monitoring how well the unit was operated yesterday is not the same as optimizing the unit in real time.

In addition to identifying the unit feed and permitting real-time optimization, inferential models have the advantage of inferential control. These models are designed to permit control of product qualities in a constrained dynamic environment. With inferential controls in place, the process optimizer can express its results in terms of product quality targets for the inferential controller. This is much easier to do, and less error prone, than to specify only targets for flow controllers in the unit.

This paper will go through distillation examples to illustrate inferential control by engineering models. We chose distillation because on almost every unit, separation control determines the product qualities. The unit may of course include a reactor such as a coker, FCC or ethylene cracker, and there are special controls associated with any reactor, but then in each of these instances reactor effluent is separated into products by distillation, and product specifications are controlled by manipulating distillation

parameters. Thus, by addressing distillation we cover a large class of inferential problems. That is not to say that we do not have models for reaction processes. We have applied engineering models for inferential control to nearly all refinery reaction processes.

TWO PRODUCT DISTILLATION COLUMN MODEL

A good many distillation columns have controls as shown in Figure 1. This control configuration is coined "indirect mass balance with tray temperature control". The controller infers bottom purity from tray temperature, and manipulates the reboiler to keep tray temperature (a simple measure of purity) constant. Reflux policy is left for the operator, or sometimes it is controlled in ratio to the feed or distillate.

There are obvious problems with this quality controller. In addition to tray temperature, two other variables affect bottom purity: pressure and V/L (vapor to liquid) ratio. In multi component columns there are additional variables associated with composition of light-light keys and heavy-heavy keys. We need a model to predict how product quality varies with all process variables. Such a model, placed as shown in figure 2, would reset the tray temperature controller target to keep product qualities truly constant.

It turns out that the inferential knowledge for this class of distillation problems is documented in the open literature and has been there for a long time. Our method is called EDIC (Engineering based Distillation Inferential Control) model. It relies on fifty years old discoveries by Colburn [2].

$$Y_N(i)/X_B(i) = [Z(i)-1] * [KA(i)-1] / [U(i)-1] + 1 = R(i) \quad [1]$$

$$U(i) = KA(i) * (V/L) \quad [2]$$

$$Z(i) = U(i)**N \quad [3]$$

$$\sum Y_N(i) = 1 \quad [4]$$

$$\sum X_B(i) = 1 \quad [5]$$

- $X_B(i)$ = Mole fraction of component i in the bottom product.
- N = The number of theoretical trays up to and including the tray whose temperature is measured.
- $Y_N(i)$ = Mole fraction of component i in the vapor above tray N.
- $KA(i)$ = Average volatility of component i in the section below the tray of interest. It is a known function of average section temperature and pressure.
- (V/L) = Vapor to liquid mole ratio on the tray of interest.

Equation 1 is actually a set of equations, defining the ratio between tray composition and bottom composition for each component. For explaining how EDIC works we consider first

a two components problem. With two components, equations 1, 4 and 5 would be formulated as:

$$Y_N(1)/X_B(1) = R(1) \quad \text{light key} \quad [6]$$

$$Y_N(2)/X_B(2) = R(2) \quad \text{heavy key} \quad [7]$$

$$Y_N(1) + Y_N(2) = 1 \quad [8]$$

$$X_B(1) + X_B(2) = 1 \quad [9]$$

There are four equations with four unknowns: $Y_N(1)$, $Y_N(2)$, $X_B(1)$ and $X_B(2)$, and we can thus find out the bottom impurity $X_B(1)$.

Note that $R(1)$ and $R(2)$ are strong functions of tray temperature. EDIC works by setting tray temperature control target, which produces $R(1)$ and $R(2)$ such that $X_B(1)$ would be at its specified purity. Should there be a change of column pressure or V/L , the inferential controller would adjust the tray temperature to counteract the disturbance.

What if we add a heavy-heavy key component? Equations 1, 4 and 5 will become:

$$Y_N(1)/X_B(1) = R(1) \quad \text{light key} \quad [10]$$

$$Y_N(2)/X_B(2) = R(2) \quad \text{heavy key} \quad [11]$$

$$Y_N(3)/X_B(3) = R(3) \quad \text{heavy-heavy key} \quad [12]$$

$$Y_N(1) + Y_N(2) + Y_N(3) = 1 \quad [13]$$

$$X_B(1) + X_B(2) + X_B(3) = 1 \quad [14]$$

We are facing here five equations with six variables. To solve this set we need one piece of additional information. Normally the additional information is a-priori knowledge of the ratio between components 2 and 3. If there is no knowledge of off key composition, it is possible also to infer that information. It turns out that column bottom temperature is quite sensitive to the off key composition, and equation 15 is an expression of that sensitivity.

$$\sum Y_B(i) = \sum K_B(i) * X_B(i) = 1 \quad [15]$$

$Y_B(i)$ = Mole fraction of component i in the bottom vapor.

$K_B(i)$ = Volatility of component i in the bottom. It is a known function of bottom temperature and pressure.

Now there are six equations and the set can be solved for the bottom contamination $X_B(1)$. Note that all of the equations are linear or can be made linear. The solution of this set is a simple matter of matrix inversion.

SIMPLE COLUMN MODELLING ISSUES

Several questions come to mind.

1. How would this model be calibrated? Does calibration require much laboratory effort?
2. This technique uses a steady state calculation method to handle a dynamic situation. Does it always work?
3. What about inferring both top and bottom products of the column? Can we combine equations 10 through 21 into one model?
4. What if there are more than one heavy-heavy key component? Can we use more temperature measurements to infer heavy key composition?

To begin with calibration, the inferential method being fairly precise and scientific, it needs no more than a single adjustment handle. Equation 3 requires the number of theoretical trays, and in our implementations we simply back calculate that number from several sets of column conditions and laboratory values. This procedure provides a method for validating the calibration data. We know that the number of theoretical trays is roughly 60 – 90% of that of actual trays. If the back calculated number of theoretical trays is much different – that indicates a problem with the lab test or process data.

Do we require many lab tests to calibrate the model? The reason why model calibrations often require large amounts of data is that many of the lab results are erroneous. No one likes to say that, and perhaps the situation at the reader's site is better than average, but our experience is that a percentage of lab results are invalid. Engineers take large amounts of data in the hope that the data is not biased and on average is usable. Our calibration does not require many data because there is a mechanism to check the validity of data. If a set of data leads to the calculation of unreasonable tray efficiencies, that data set can be discarded. Further, the small amount of data makes it possible to tightly supervise sample taking and lab test.

Next comes the question of a steady-state model in a dynamic environment. Flow controllers take seconds to respond while temperatures can take many minutes. How do we ensure that flow and temperature signals are in phase? In indirect mass balance configuration as in figure 2, flow changes quickly follow temperature changes. Thus, indirect mass balance configuration does not require dynamic correction of model inputs. We do however input one variable: V/L for equation 2, via a dynamic predictor. V/L is a function of reboiler heat duty (or reflux flow in figure 3), which is manipulated by the temperature controller. The predictor estimates final steady state V/L and avoids introducing noise into the inferential calculation.

The dynamic situation would be different if the basic column controls were direct mass balance, because the tray temperature loop there is slower... But this is a subject for another paper.

The next question is about modeling both product purities. Why not have tray temperature based models for the both top and bottom product purities? The bottom model would set a target for a tray temperature in the stripping section, manipulating the reboiler. The top model would set a target for a rectifying section tray of interest, manipulating reflux.

There is no theory against two models for the two products, but there are practical examples of troublesome dual temperature control applications. Any inferential model is an approximation, and operation of two models on the same column may result in conflict between them and poor control. Consider for example an error in specifying the rectifying section tray temperature target slightly too low. The rectifying controller would then increase reflux to cool the tray. That action would cool not only the rectifying section but also the stripping section. The stripping controller would respond by increasing reboiler heat duty, heating up the stripping tray, and also the rectifying tray, which would call again for a reflux increase, and so on. Thus, a mismatch between the two models can lead to high reflux and flooding, or, if in the other direction - to a low reflux. From a mass balance point of view, a mismatch between the two models traps a portion of the light key or heavy key material, not letting it get out of the column, and that accumulation would eventually flood the column.

Our preference is to have one temperature based model and one reflux model as a function of the required separation. Reflux logic can typically be determined off line via the use of a simulation tool. Such a strategy completely avoids competition between the two models.

The next question is about reading a number of column temperatures to infer multi-component off key composition. Our example model has used only two temperatures in the stripping section: at a representative tray and at the bottom. Our top section model has used rectifying tray and overhead drum temperatures. Where we had a complex off key situation we have simplified it, accepting some model drift. A case in point is the debutanizer model of figure 3. It uses drum temperature to estimate the C3 content, but assumes a constant ratio between normal butane and isobutane.

Can we use a temperature profile of the stripping section to extract information about all heavy off key components? In principle the answer is "yes", but complexity might be too large for a simplistic expansion of equations 10 through 15.

We have gone a step further in sophistication by using a standard distillation simulation program (Kesler Engineering's SPAN), to back estimate column feed composition to reproduce the actual column temperature profile. We are not yet at liberty to publish details of this approach, though have published commercial brochures [6, 7] which show this reconciliation process in concept. The details will come in a future paper.

SIMPLE DISTILLATION MODEL EXPERIENCE

We have used EDIC, either in the form of equations 10 through 15, or 16 through 21, in a large array of applications: debutanizer, depropanizer, deisopentanizer, naphtha splitter, ethane-ethylene splitter, butane-isobutane splitter, to name a few. We have also used it on complex columns such as alkylation unit isostripper for controlling the separation between butane and isobutane. The model has first been discussed in the open literature in reference [3], and now this paper provides more detail and also performance data.

Figure 4 shows a trend of model versus lab reading of C5 in LPG on a debutanizer. This is a test period lasting about five weeks. After three weeks of test the lab determined that the inferential calculation was reasonable and reduced the frequency of tests to once a week. Operators then started to view the inferential trend, and control of the column improved. All of the data in figure 4 represent open loop operation though. We do not have an equivalent closed loop comparison because shortly after closing the loop the laboratory stopped taking this sample altogether. (This is a difficult test because of LPG sample volatility).

FRACTIONATOR CUTPOINT MODELS

A multiple product fractionation column is shown in figure 5. We chose this example for three reasons. First there are fractionators of this kind in every crude unit, FCC, coker and other refinery units. Second there is much industrial experience with inferential control of these units by the GCC (Generalized Cut Control) and other models. Third, there are large differences in product prices and cutting the products correctly is an important economical consideration. The challenge here is to infer certain events, such as a change of feed to the fractionator and take action to counteract disturbances. The change of feed may be a result of a crude switch, coker drum switch, FCC feedstock switch, etc.

Fractionators are very different from the distillation columns of figures 2 or 3. The feed is either crude oil or other wide boiling mixture of components. Being a mixture of many compounds, fractionator feeds are characterized by their boiling curve indicating temperature versus percent boiled. The fractionation goal is to separate the feed into several fractions, which are also characterized by boiling curves. There are several kinds of boiling curves depending on the tests to produce them, and there are approximate API procedures for converting from one to the other.

Most fractionator models, GCC included, work on the TBP (true boiling point curve) principle. This curve is convenient because it describes ideal fractionation of the feed. An example TBP curve is given in figure 6. The heavy line is for total crude (or other column feed) whereas the intermittent lines show how the feed is cut into three products: naphtha, kerosene and diesel. The fourth cut is called overflash and is not a real product, but material to be refluxed back down with the bottoms. In GCC nomenclature we call the cuts between the products:

- + NCP (Naphtha cutpoint, between naphtha and kerosene)
- + KCP (Kerosene cutpoint, between kerosene and diesel)
- + DCP (Diesel cutpoint between diesel and bottoms)
- + FCP (Flash zone cutpoint between overflash and bottoms)

The thin lines of figure 6 show typical product TBP curves. If we had ideal fractionation, product TBP curves would coincide with the crude curve. The heavy and light “tails” on product curves are due to imperfect fractionation and boiling range intermixing.

We have published at least two papers [4, 5] about the engineering principles of GCC. In essence GCC relates column measurements to points on an EFV (Equilibrium flash vaporization) boiling curve of specific streams. Then it converts that information to feed

TBP points. The procedure recreates a feed TBP curve from process measurements. From the TBP curve it is possible to determine product cuts and properties. This current paper will give a summary of how the model works, explain key features of the model, and demonstrate the model inferential performance.

- Naphtha cutpoint calculation.
Naphtha is the column overhead product. Its cut point is inferred from column top temperature after correction for partial pressure. The top temperature is indicative of overhead product EFV endpoint (or dewpoint).
- Flash zone cutpoint calculation.
The column flash zone temperature is indicative of flash zone vapor EFV endpoint. GCC applies standard API methods to first correct flash zone temperature measurement for partial pressure and then convert from EFV to cutpoint TBP. The conversion technique is the same as the one for Naphtha cutpoint calculation.
- Crude TBP slope calculation.
The model assumes a straight line boiling curve between Naphtha and Flash zone cutpoints. Nonlinearities can be calibrated, but this is not usually done for crude columns as most crudes do have a straight line boiling curve. For FCC, visbreaker or coker fractionators nonlinear shaping of the TBP curve relies on intermediate temperature measurements such as sidestream draw temperatures.
- Sidestream distillation qualities.
With TBP curve reconstructed, front and back cutpoint of all sidestreams become known as shown in figure 5. All distillation qualities are then calculated as a function of both front and back cuts. To the extent that internal reflux variations have a large influence, it is also taken into account. For example:
$$\text{Diesel 90\% point} = K1 * KCP + K2 * DCP + K3 * \text{Diesel} / \text{IR} \quad [22]$$

K1, K2, K3 = Known coefficients.
Diesel / IR = Ratio of diesel draw to internal reflux below the diesel draw tray.
- Cold properties.
Qualities such as freeze and cloud are a function not only of cutpoints but also of aromatic content. Aromatic content varies from feed to feed, and there is no good way to estimate it from simple measurements of temperature, pressure or flow. Precise inference of cold properties requires a gravity analyzer on the sidestream in question.
- Overflash calculation.
The model includes inferential calculation of overflash from temperature data. This inferential calculation is important because overflash flow meters usually have reliability and range problems.

SELF INFLICTED DISTURBANCES

All fractionator inferential control models must face a certain “chicken and egg” dilemma, associated with the use of product flow measurements as model inputs. When in closed

loop, this creates a strange situation. The model on one hand sets product flow targets, while on the other hand it inputs measurements flows controlled to these targets. Such models read their own outputs as disturbances and mislead themselves. That is where steady state models fail to handle a dynamic situation. The temperature dynamics are much slower than flow dynamics, and the model, after manipulating flows, sees transient data and interprets that as a change of feed. The previous section has mentioned a similar problem with regards to our model of equations 10-15 in a direct mass balance configuration. The fractionator of figure 5, has all its sidestreams on direct mass balance control, and any closed loop model must deal with that problem.

What is a good way to deal with the dynamic problem? Conventional approaches slow down the flow inputs by filtering. Slowing down the flow signals allows time for temperatures to be more in phase with flows, to the point of no longer confusing the model. The fly in this ointment is that the slowed down model has lost its ability to deal with quick disturbances such as crude switch or coker drum switch. Worse yet, during major disturbances fractionators typically operate off mass balance, accumulating material on trays and downcomers. Simple input of flow measurements would mislead even an open loop model.

GCC applies a different solution. Instead of measuring the side product flows it estimates the flow of vapor from flash zone into the column from heat balance. In terms of figure 6, it estimates the yields $Y_0 + Y_1 + Y_2 + Y_{OF}$, not as they go out of the column as products, but as they come into the column as vapor. Now the model no longer inputs its own manipulated variables and there is no need to slow down any input.

From a timing point of view, heat balance is nearly perfectly in phase with feed volatility. As the feed changes volatility, more vapor enters the column, and then cooling load must increase immediately. Otherwise a pressure relief valve would open or temperatures would go up (and that would reflect in the heat balance). Indeed, this method has demonstrated success with handling crude switches and other disturbances. Even when operation is off material balance, the model predicts the eventual product qualities, should yields remain at current setting. This prediction precedes the actual quality change, thus permitting ample time for the controller to respond to the prediction and correct the yields.

Having solved the difficulty of a model feeding on its own output, GCC faces another problem. Heat balance is not a perfect device and there might be a discrepancy between heat balance and mass balance. We accepted and even welcomed a short term discrepancy, but a steady state discrepancy would bias our inferential model. GCC assumes that in the long run mass balance is more accurate than heat balance (there are heat balance constants such as heat of evaporation or heat capacity which are only approximately known). To force steady state agreement between heat and mass balance, GCC slowly manipulate a heat loss term.

In addition to eliminating inferential bias, it turned out that the heat loss handle gives us information about how healthy the plant data is. If heat loss is negative, or if larger than, say 3% of the total cooling load, the quality of data is suspect and one or more flow meters probably need calibration.

FRACTIONATION MODEL EXPERIENCE

GCC has been implemented in about 25 locations, and it is now a mature technology. We have chosen data from one of these sites to demonstrate how the inferential model works. The specific unit has a distillation analyzer on kerosene and a cloud analyzer on diesel. Plotting the analyzer readings with time shows how well the controllers have worked. Further, the analyzers are configured in closed loop internal model type control, where the internal model controller changes a model bias in the inferential calculation. Trending of this bias over time indicates the inferential calculation quality. It helps test how well control would have been without the analyzers.

Figure 7 shows how the model has coped with a crude switch disturbance on the unit. Figure 7A trends the inferred crude TBP curve slope in the sidestream boiling range: how many degrees it takes to increase evaporation by one percent. A higher slope indicates that the switch was from light to heavier feed. The column went from 60% distillates to 45%, quite a substantial change. Figure 7B trends the reading of a kerosene (first sidestream) 95% boiling analyzer, and its model bias. Figure 7C trends the diesel cloud reading and model bias.

According to the inferential model of figure 7A, feed change took about an hour; a reasonable time period because of mixing in heat exchangers and in the desalter. It is interesting to note that the crude switch was nearly over before the analyzers noticed any disturbance. That is because the analyzer sample points are downstream of the side strippers and heat exchangers, and then there are sampling loop delays. A one hour dead time in crude unit analyzer responses is common. In this specific experiment the time delay permits us to separately analyze whether the timing of inferential control was correct, because no feedback action was taken until the analyzers detected a deviation. The peak deviation of quality from target was small, and that confirms the good timing of our heat balance calculation. The kerosene distillation deviation peaked at 4°F and cloud at 6°F.

The kerosene distillation model bias showed some dynamic errors, but overall was well behaved and the steady state bias was little changed from one crude to another. This supports our claim that GCC can infer distillation properties well, and that it can work through major disturbances.

Trending of the diesel cloud bias relays a different cloud control scenario. There is no gravity analyzer on the diesel stream and thus no good way to estimate the aromatic content. Cloud is estimated solely as a function of boiling range. Figure 7C shows that while control of diesel cloud was not bad, the bias has changed by about 10°F from one crude to the next. With the help of a cloud analyzer we could eliminate the steady state error but not the dynamic one. To predict the change of bias and eliminate the dynamic error we would need a gravity analyzer.

CONCLUSIONS

We have shown some examples of a modeling technique which is based on fundamental physical principles. These models are powerful in several ways:

1. Use engineering variables, such as heat duty or partial pressure. These are easily understood by process engineers.
2. Non linear, physical laws models, whose range of validity is greater than that of regression models.
3. Guaranteed no conflict between plant data and model results. That is a necessary condition for real time optimization.
4. All calculations are based on process measurements, not on operator inputs.

LITERATURE CITED

1. Friedman, Y. Z., "What's Wrong with Unit Closed Loop Optimization". Hydrocarbon Processing Journal, October 1995.
2. Colburn, A. P., Ind. Eng. Chem. 33, 459, 1941.
3. Friedman, Y. Z., "Simulation Based Inferential Control". AIChE spring meeting, 1995.
4. Friedman, Y. Z., "Control of Crude Fractionator Product Qualities During Feedstock Changes by Use of a Simplified Heat Balance". American Control Conference, 1985.
5. Friedman, Y. Z., "Crude Unit Advanced Control Experience". Hydrocarbon processing Journal, February 1994.
6. Kesler, M. G. and Friedman, Y. Z., "CPPM" and "MINFRAC", two commercial brochures issued by Kesler Engineering, 1995.

FIGURE 1
INDIRECT MASS BALANCE CONTROL
WITH TRAY TEMPERATURE FEEDBACK

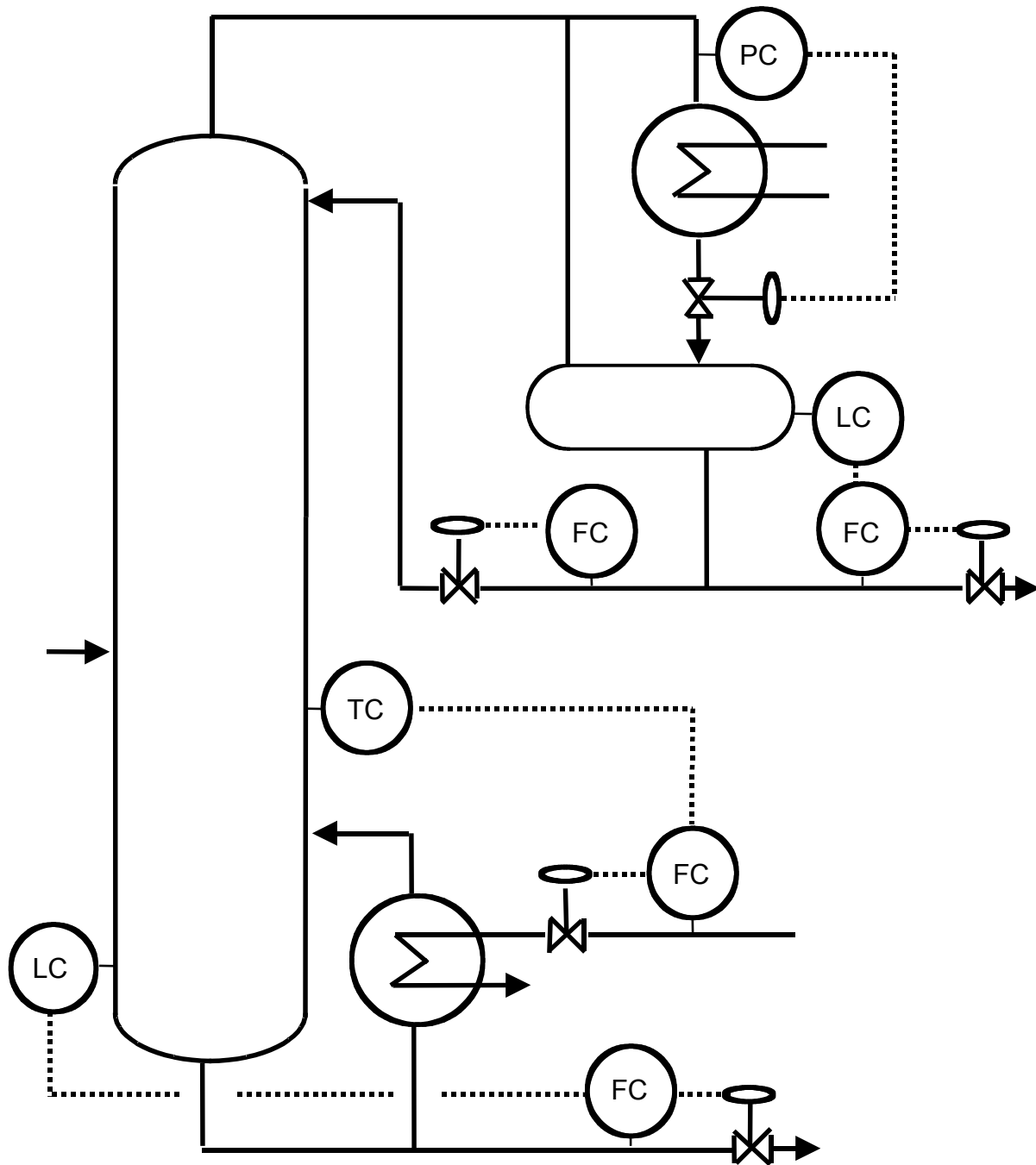


FIGURE 2

INDIRECT MASS BALANCE CONTROL
WITH INFERENCE CONTROL MODEL

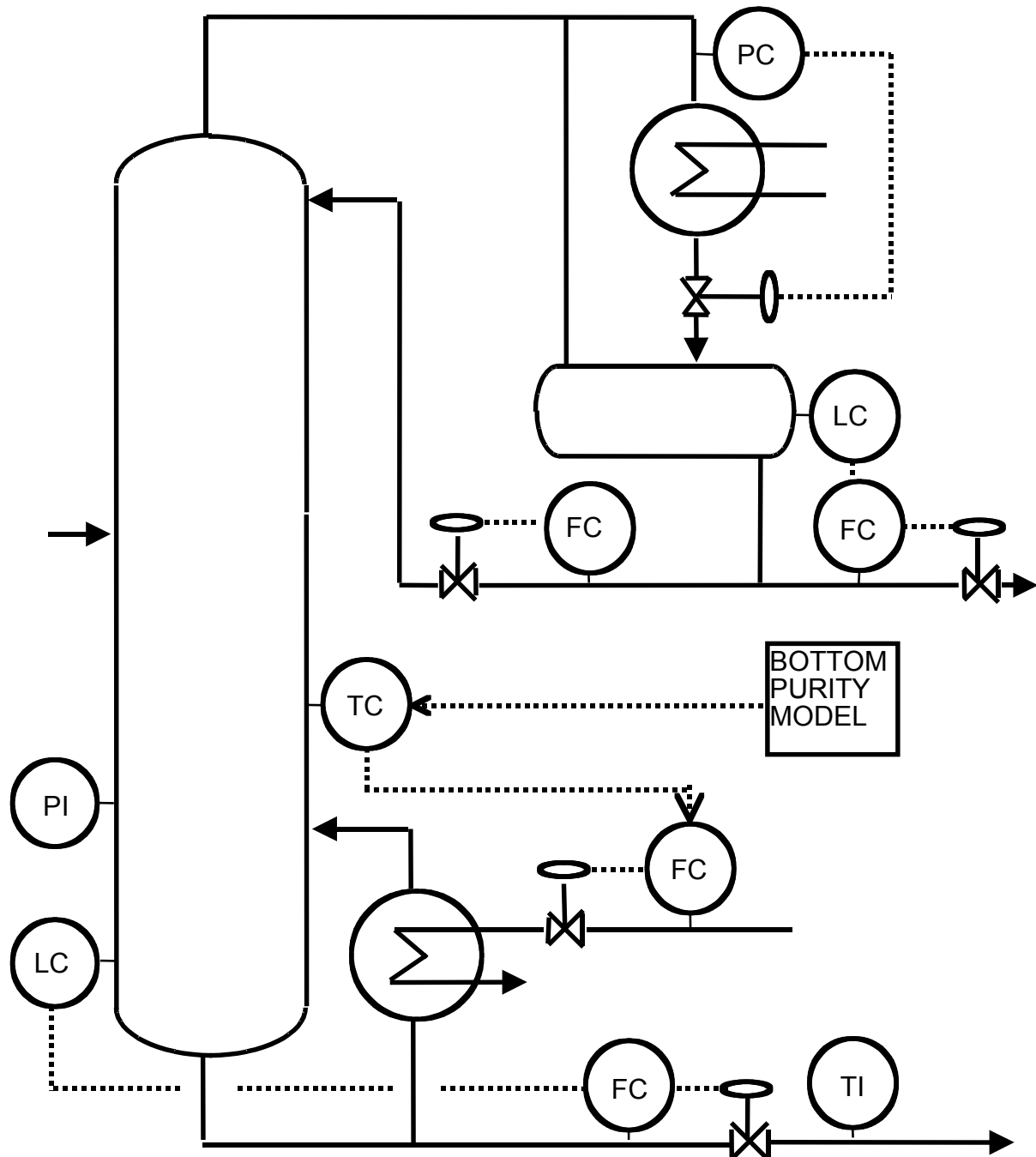


FIGURE 4

INFERRENTIAL CALCULATION VS. LAB
C5 % IN DEBUTANIZER TOP PRODUCT

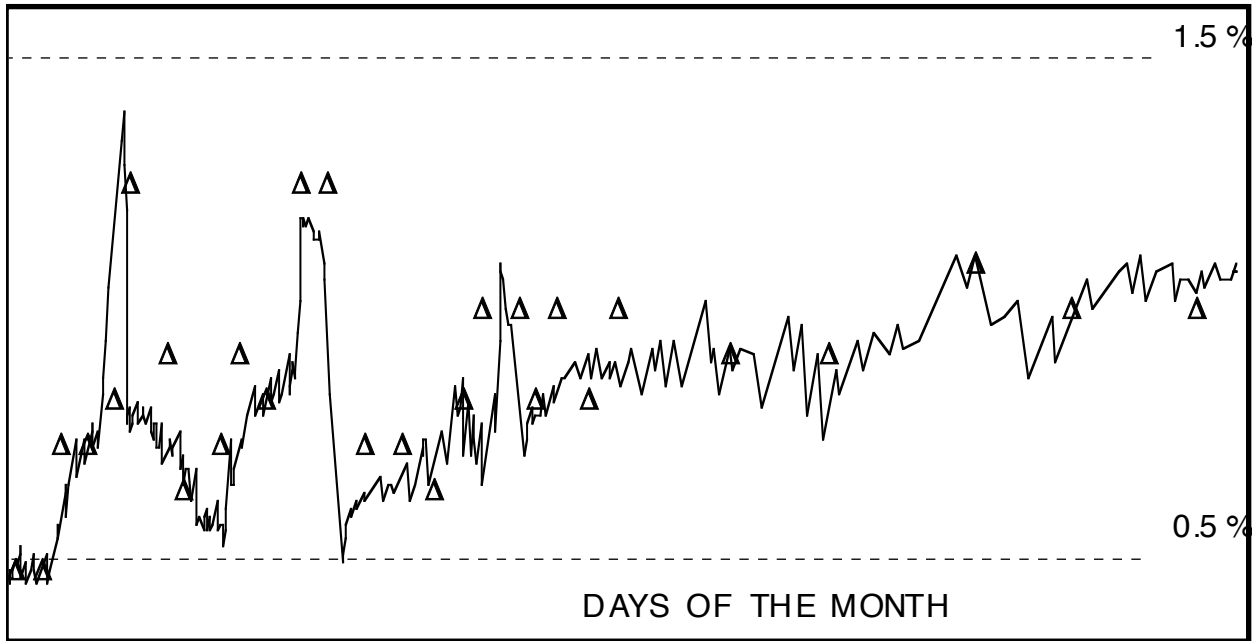


FIGURE 5

INFERENCE CONTROL OF A CRUDE FRACTIONATOR

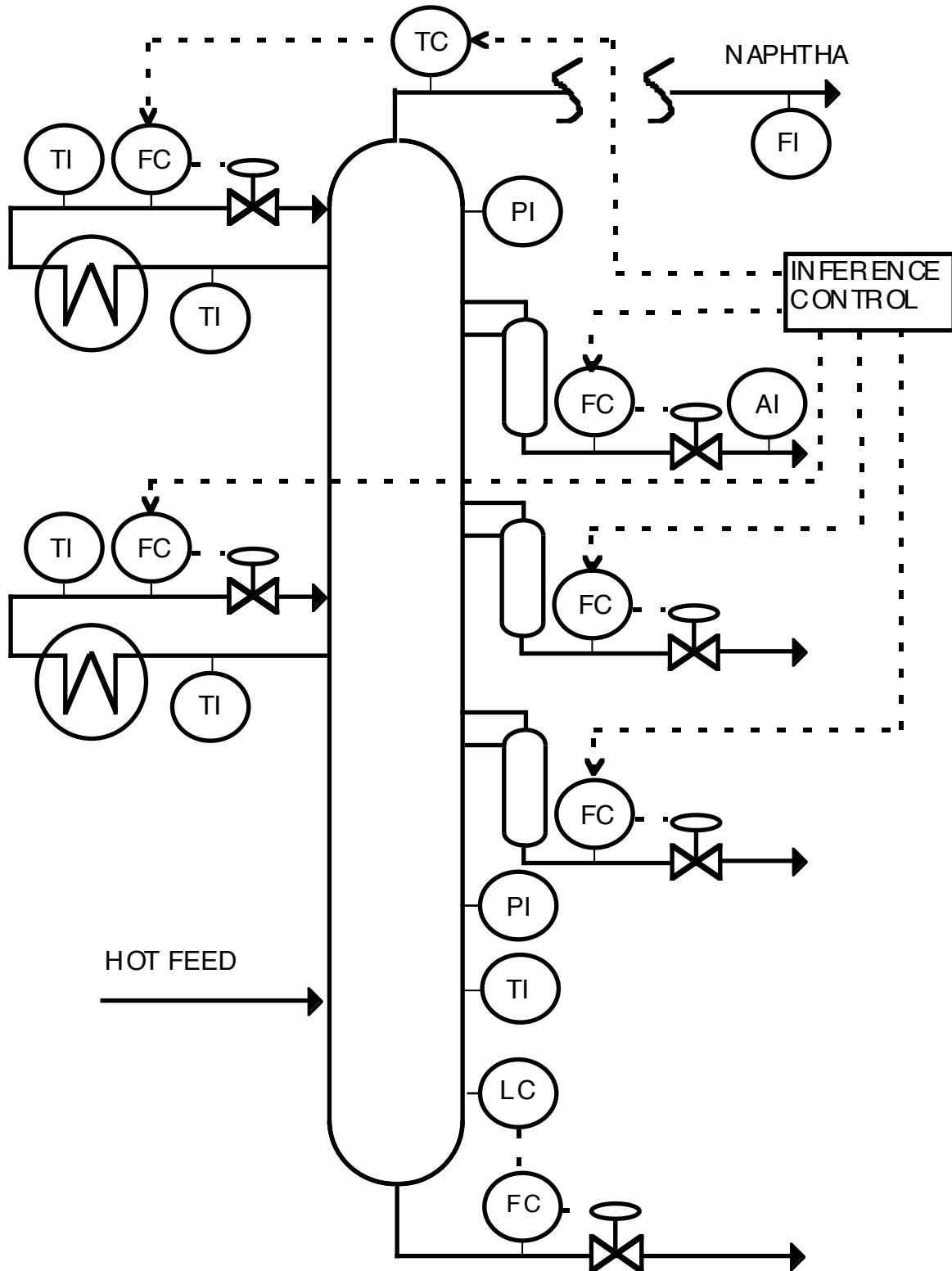


FIGURE 6

TYPICAL CRUDE TBP AND PRODUCT TBP CURVES

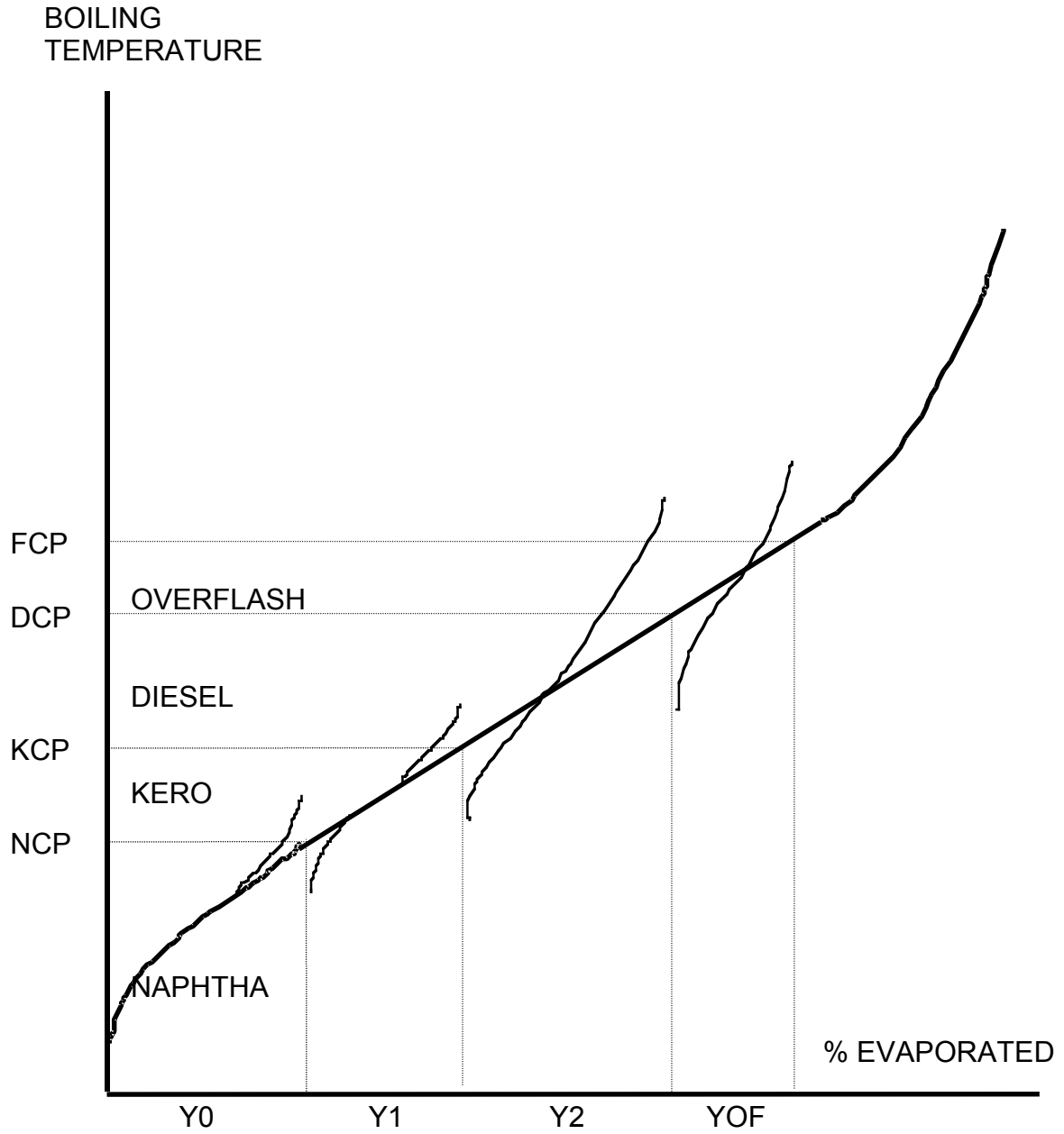


FIGURE 7

FIGURE 7A

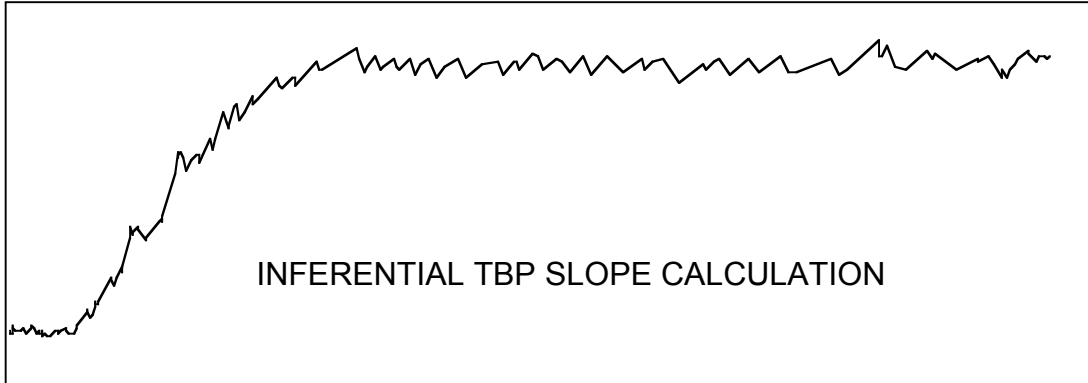


FIGURE 7B

KEROSENE 95% ANALYZER READING

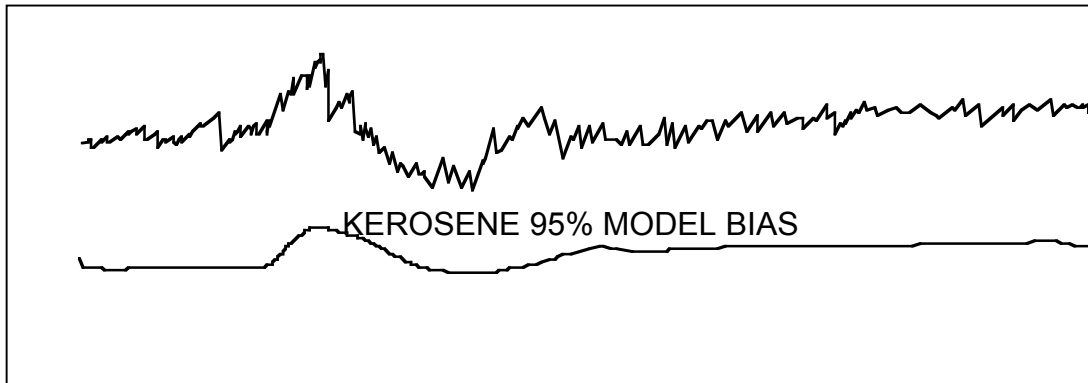


FIGURE 7C

DIESEL CLOUD ANALYZER READING

