THE USE OF
FIRST-PRINCIPLES DISTILLATION INFERENCE MODELS
FOR CONTROLLING A TOLUENE / XYLENE SPLITTER

By:

Y. Zak Friedman, PhD

Petrocontrol, New York, NY, USA

Euclides Almeida Neto

Petrobras, Rio de Janeiro, RJ, Brazil

Carlos Roberto Porfirio

Petrobras, Cubatao, SP, Brazil

BACKGROUND

This paper is a part of a sequel on distillation purity inferential control models that make use of chemical engineering principles. A number of years ago we came up with inferential models for distillation columns that are based on scientific principles [3]. The modeling approach involves a short cut simulation of a section of the column, typically a bottom half of the stripping section or a top half of the rectifying section as shown in figure 1. Hence the name of the model: GDS (General distillation shortcut). GDS works by fitting a bottom (or top) column composition that would best agree with column measurements. That is fairly easy to accomplish with binary distillation, and the novelty of GDS is that by looking at a number of tray temperatures it is equipped to address multi-component distillation.

Figure 2 is an example GDS model reading two temperatures to come up with three equations, permitting composition inference of a product containing three components. The model employs three first principles: bubble point, section separation "Colburn ratios" and mass balance. GDS relies on Colburn's formulae for distillation column section performance [4], describing the ratio between tray composition and bottom (or top) composition as a function of component volatility (K value), internal reflux, and number of trays in the section. These formulae provide a way to interpret tray temperature measurements correctly, taking into account the effect of pressure, reflux and multi-component environment.

In terms of complexity, equation coefficients of figure 2 are obtained by nonlinear calculations, but the resulting equations are linear, allowing once through calculations with no convergence issues. That makes GDS very suitable for on-line closed loop work. Perhaps the most complex GDS calculation involves heat balances to estimate the internal reflux of column sections, needed for the calculation of separation parameters.

Since the initial GDS publication [3] a large number of GDS inferences have been implemented and the large volume of data permitted publication of the GDS performance [1, 2]. A 2001 paper [2] compared model prediction versus analyzer readings on a large number of distillation columns at Texaco Pembroke refinery. Figures 3 and 4 show one example from that paper, addressing a deisobutanizer column. Figure 3 is a diagram of the column, highlighting the inferentially modeled top section, whereas figure 4 shows the close agreement between analyzer and model. Experience indicated that where measurements of tray temperatures are available, and where we have reliable data for calculating internal

reflux, the inference models are excellent. When the measurement data are imprecise and we are forced to make assumptions - the models might drift over time, though they could still be useful for closed loop control.

Following the modeling work with Texaco we had the opportunity to work on a BTX (benzene, toluene, and xylene) separation unit at Petrobras RPBC refinery, where have worked on a rather demanding benzene column problem. Demanding because it has tight specifications, 200ppm on both the top and bottom products. Our February 2002 paper [1] reports the results of that experiment. Figures 5 and 6 are excerpts from that paper. Figure 5 shows the benzene column, highlighting a top section which envelopes the inference model of interest. Figure 6 illustrates the model versus lab results for a period of three months. Given measurements of tray temperature and pressure, GDS can model superfractionators quite well, and at issue was not the GDS accuracy but the inability of the laboratory to produce clean repeatable tests at the ppm level. Modeling the benzene tower was a strange experience because the model became more reliable then the lab tests used to calibrate it.

PROBLEM STATEMENT

This paper tackles another control problem in the Petrobras BTX unit: the separation of toluene from Xylene. At first glance the toluene column problem does not appear overly complicated. Columns of about 40 trays and 2:1 reflux ratio can hardly qualify as superfractionators, and the specifications of 0.25% xylene in toluene, or 0.5% toluene in xylene are easier to control than the 200ppm benzene tower specifications. On the other hand, there are circumstances that make the problem difficult. There are two toluene columns working in parallel, and the toluene lab sample points are on the combined top product streams. The bottom streams are combined and separated in the xylene column, and thus, xylene product purity is affected not only by toluene column operation but also by disturbances in the xylene column. Even that situation is manageable, except the two columns have substantially different performances; one column has 44 trays, whereas the other one is built only with 34 trays. The shorter column naturally operates with less feed and a higher reflux ratio. If that is not difficult enough, the columns do not have a pressure measurement at a convenient location, forcing us to make assumptions. All of these issues combine to make the inferential modeling of the toluene columns a difficult problem indeed.

COULD WE APPLY EMPIRICAL MODELS?

There is a school of thought that advocates the use of empirical regression to infer distillation column product purities. The authors of this paper are not generally against statistical models, especially for equipment where first principle knowledge does not exist, but in distillation there are several strikes against empirical models and we strongly believe that only first principles based modes have a chance to work. The limitations of Gaussian theory have been explained in our February 2002 paper [1], and we give below a summary of the arguments.

A) Regression requires independent inputs

Gaussian theory requires all inputs to a model to be independent. However that is not feasible in distillation because temperatures, pressures and flows are subject to mass balances, heat balances and equilibrium equations. Ignoring these relations renders the modeling process theoretically incorrect. A more correct statistical approach would involve Bayesian theory, which takes into account a-priory knowledge of dependence among input variables. Of course that a-priory knowledge brings us back to the first principles of distillation.

What then happens to models that use dependent variables? They have incorrect coefficients and hence would drift upon changes in process conditions.

B) Empirical models require large volumes of lab data

Statistical regression requires hundreds of laboratory data and that poses a problem. A fair percentage of daily lab data is biased and reliable process data is obtainable only by test-runs. There is no hope that the quantity of lab data needed for regression would come from high quality test-runs, and empirical model developers would have to rely on imprecise every-day lab data.

Can this lab data still provide the basis for a useful model? That might be possible, if the error is normally distributed without bias. Model developers often assume that those lab data that do not fit their correlation are wrong. They label those data "outliers", removing them from the original set, thus obtaining excellent fit, only to find out later that regression fitting

does not guarantee future prediction. How many outliers are acceptable? Eliminating just a few points from a large set of random data could entirely alter the regression formulae. Can we delete 1% of data from a set and still call this set reliable? What about 5%? What about the roughly 50% of the data in Figure 6?

C) **Empirical models must identify a large number of coefficients**

Scientific models incorporate model gains inherently, and the calibration procedure amount to adjusting one or two parameters. For example the model of figure 6 was calibrated by setting tray efficiency and weight in a weighted average formula. The effect of signal to noise ratio on the calibration procedure is minimal.

Empirical models on the other hand must identify at least one gain coefficient associated with every input; two or three coefficients per input for nonlinear models. Normal day-to-day operation does not provide enough movement in the data to support a multi coefficient regression and the procedure then simply models noise patterns.

D) **First principle models provide the means for checking instrument errors.**

The sister problem of erroneous lab data is erroneous instrument data. Instrument errors occur due to poor calibration, partial plugging of orifice meters, improper installation, incorrect meter range and finally also computer interface errors. Inference models that were developed from erroneous instrument data would necessarily be weak. Eventually instrument problems are identified and corrected and what would happen to a regression model then? It would have to be re-developed of course, but from what data? The old data sets with erroneous readings cannot be used. Assuming lab tests are carried out once a day, creating a new set of hundreds of lab data would take many months, during which time the APC would be off.

We conclude that before starting inferential model development it would be prudent to survey all input measurements and identify all problems. The best way to accomplish that is via first-principles knowledge; testing the readings against mass balance, heat balance and equilibrium equations. If instrument readings cannot reconcile against basic thermodynamic laws they cannot be correct. There is an obvious conflict here because people choose the

empirical approach to avoid the more time-consuming first principle models, only to find out that the use of first principles is unavoidable.

E) <u>There is no replacement for process engineering</u>

And what if the measurements set is inadequate? A key measurement is missing, or is in the wrong location?   To obtain a good model the set of measurements ought to "have the inferential information in them". A first-principles modeler would identify an insufficient set of inputs at the outset by a simple sensitivity study. He would then halt the modeling effort until the missing measurement is installed. The empirical modeler would go through model development and the problem would only be found out at the time of model validation. At that time he would have to employ first principles to analyze the problem, wait for the measurement to be installed, then months until a quantity of lab data is available.

F) <u>Ability to survive process modifications</u>

During unit turnarounds distillation columns are often modified by the way of replacing trays, cleaning condensers or reboilers, etc. Any inferential model would need to be re-calibrated upon modifications of column equipment. First-principles models might require changes of equation coefficients, but empirical models would be turned off for a period of several months until a fresh set of lab data permits model re-development from scratch.

<u>THE TOLUENE COLUMNS</u>

Figure 7 illustrates the available measurements around the toluene columns. N08 is the taller of the two, having 44 trays, with feed on tray 23. N09 only has 34 trays and the feed comes in on tray 18.   Both columns have rectifying section tray temperature measurements that could be used by the GDS model: tray 14 of N08 and tray 10 of N09. But that is not true for pressure measurements; both columns have top and bottom pressure points and the only way to estimate the pressure at the temperature points is by interpolation. We will see that lack of ability to precisely estimate the pressure stood in the way of obtaining accurate models.

The stripping sections of N08 and N09 are not sufficiently instrumented to support a GDS model. In theory that is not a problem because we did not intend to apply two temperature based models in closed loop on the same column. As explained in previous papers [1,2], it is

better to apply one temperature based and one internal reflux based models, obtaining the internal reflux model by regression. But in practice the lack of stripping section measurements does present a problem. Our normal procedure calls for first developing two proper GDS models for each column, one for rectifying and one for stripping, and later convert one of them by regression into an internal reflux model. The lack of stripping section model on the toluene columns has forced us to a less accurate one step regression, directly against lab data.

The advantage of going through a two-stage process: develop a GDS model and then regress it into an internal reflux model, is in the treatment of lab data. GDS being a first principles model requires calibration of one or two coefficients, and hence is not very sensitive to erroneous lab data. Direct use of lab data in the regression bring us back to the problems of regression models described above. We can avoid dependent inputs but not the other problems associated with low fidelity lab data.

RECTIFYING SECTION MODEL

Our rectifying model relies first on a calculation of the xylene content at a pivot point: tray 14 of N08 or tray 10 of N09. The pivot point concentration of xylene turned out in the order of 20%, fairly easily identified by equilibrium calculations. Once that composition is known, the model goes on to estimate the xylene contamination of the top toluene product from a Colburn relation between the pivot point and the side-draw point.

Obtaining a pair of temperature - pressure measurements at the pivot point was problematic. The model expresses the missing pivot tray pressure as a weighted average of the top and bottom pressures, and we have calibrated the weighting coefficient to obtain the best model fit.

A second dilemma was the common lab sample point for the two columns. Operators do try to manipulate the two columns in an "identical way" but with the columns themselves not being identical, what is the meaning of "identical way"?

Being unable to obtain individual column data we pretended that the lab data was individual and tuned the columns assuming they were producing identical product purities. There were two calibration handles: tray efficiency and pressure interpolation coefficient. One would

expect such a simplistic approach to encounter calibration problems but to our surprise we could calibrate both columns using reasonable tray efficiencies and pressure coefficients. Figure 8 shows the N08 model versus <u>combined</u> lab data. The number of theoretical trays in the top section that gave the best fit was 13, as opposed to 14 actual trays. The pressure interpolation coefficient that gave the best fit was: [0.81* (top pressure) + 0.19 * (bottom pressure)]. Figure 9 shows the N09 calibrated model versus <u>combined</u> lab data. The number of theoretical trays in the top section that gave the best fit was 9.7, as opposed to 10 actual trays. The pressure interpolation coefficient that gave the best fit was: [0.8 * (top pressure) + 0.2 * (bottom pressure)]. Amazing fit that to this day we do not know how much of it was shear luck.

STRIPPING SECTION MODEL

Encouraged by the rectifying section results of figures 8 and 9, we proceeded with bottom models for N08 and N09. To develop the models we regressed the combined lab data against the following model inputs.

➢ Top purity as per the top model (without using the top lab result). For the same V/L the bottom purity should decline when the top purity improves.

➢ Bottom pressure compensated temperature. This temperature is an indication of C9+ material in the bottom.

➢ Stripper V/L. Higher V/L should improve the bottom purity.

➢ Toluene to feed ratio. Higher content of toluene in the feed makes the bottom separation more difficult.

Figures 10 and 11 show the bottom models of columns N08 and N09 respectively. If we discard some obviously incorrect lab results the fit is excellent, almost too good to be true, considering again that we have made the unlikely assumption that these two different columns produce the same purity products. Attempting to find faults with the models we have compared the model coefficients in table 1, however table 1 only confirms the validity of the two bottom models by illustrating how similar the two models are.

Table 1. Bottom model coefficients

Columns	LN(Top purity)	Bottom pressure	Stripping	Toluene/Feed

		compensated temperature	V/L	ratio
N08	-0.10286	-0.27409	-13.3405	+3.63968
N09	-0.09059	-0.33122	-14.7761	+4.24454

CORRECTED N08 MODELS

In spite of our success we were reluctant to permit closed loop control based on inferential models whose basic assumption could be wrong. After a discussion with the laboratory personnel they had agreed to temporarily conduct separate tests on column N08. We then found out that our top model did not change much as a result of the dedicated column test. The best number of trays remained 13 and the pressure interpolation coefficient has changed from 0.81 to 0.91. This individual column model is shown in Figure 12.

CLOSED LOOP PERFORMANCE

Encouraged by the open loop model performance we decided to close the loop. The lab has agreed that at least in the first few months they would test the top streams of each column separately. The bottom streams continued to be combined and taken to a downstream column N11 to separate the xylene. The lab sample for testing toluene in xylene is on the top stream of N11.

Figures 13, 14 and 15 trend the closed loop performance of inference model plus controller. The three figures start with a month of open loop period, followed by a month of closed loop period. Figures 13 and 14 show the contamination of xylene in toluene for the two columns, whereas figure 15 trends the combined toluene in xylene. In all three cases the figures demonstrate significant reduction in variability.

CONCLUSIONS

We have introduced GDS; a first-principles inferential technique for multi component distillation that is much better than empirical models. GDS had been tested successfully for refinery distillation processes, a previous paper has described superfractionation of benzene separation to 100 ppm purity level. Now this paper has tackled a difficult distillation

situation with two parallel columns, insufficient lab testing and missing measurements. Our toluene columns did not have pressure measurements in correct locations, and we had to apply interpolations, introducing error into the inference prediction. Nevertheless, the paper has shown that even under these difficult conditions the GDS technique is successful to the point that its closed loop performance is satisfactory.

Having said that, we did recommend that at the first opportunity Petrobras install pressure measurements at the correct locations.

LITERATURE CITED

1. Friedman, Y. Z., Neto, E. A., Porfirio, C. R., "First-principles distillation inference models for product quality prediction". Hydrocarbon Processing Journal, February 2002.
2. Friedman, Y. Z., Reedy, G. T., "Experience With GDS – A First-principles Inferential Model for Distillation Columns". Petroleum Technology Quarterly, fall 2001.
3. Friedman, Y. Z., "Simulation Based Inferential Control". AIChE spring meeting, 1995.
4. Colburn, A. P., Ind. Eng. Chem. 33, 459, 1941

Figure 1. Typical GDS configurations

Petrocontrol
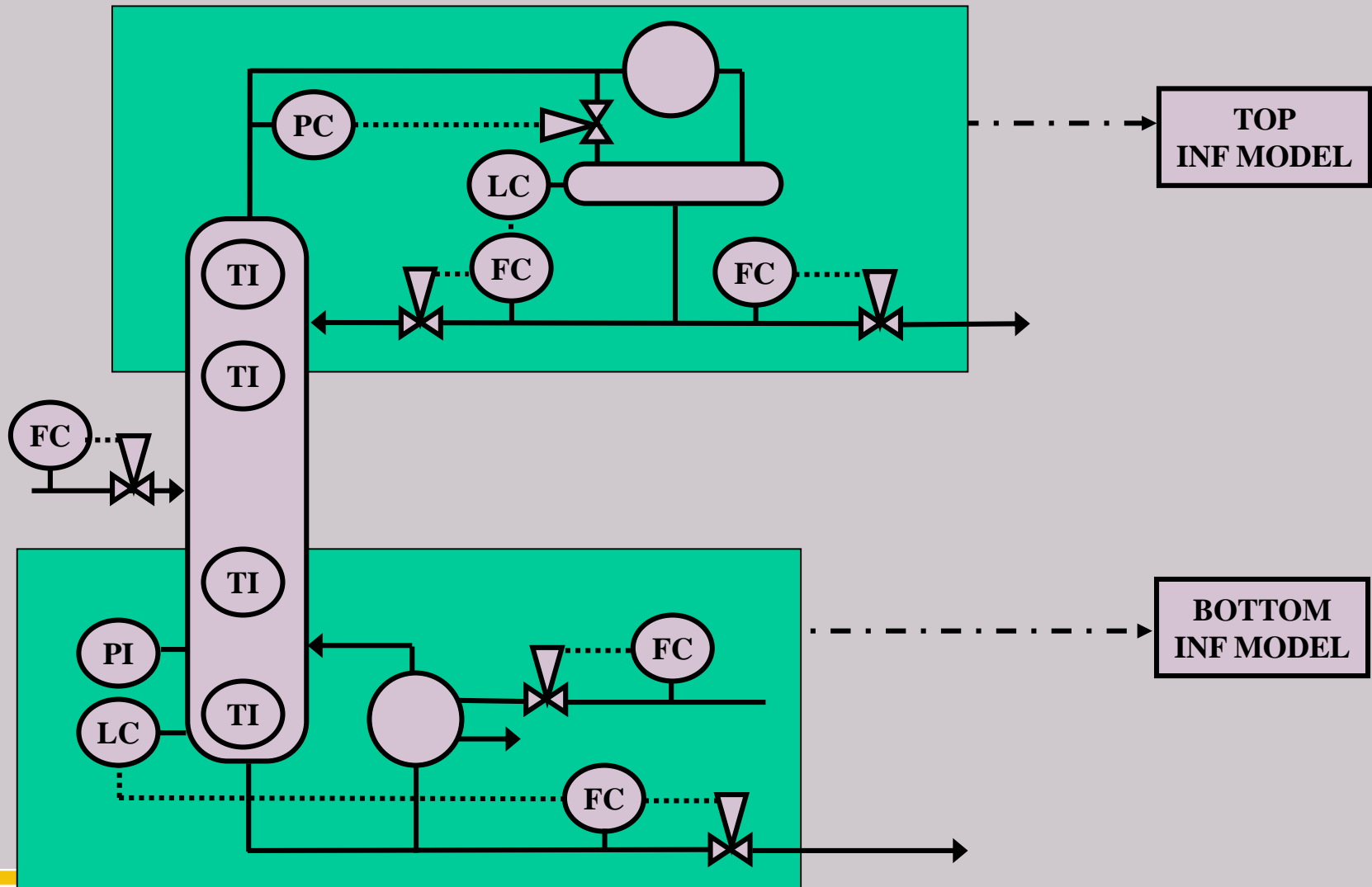
Figure 2. Example bottom model

Three bottom composition unknowns:
Benzene, toluene, xylene -- XB1, XB2, XB3
Three linear equations

- **Bottom bubble point equation**

  $\Sigma \,(KBi * XBi) = 1$ $\qquad$ **KBi = YBi/XBi at bottom conditions**

- **Section separation equation (Colburn ratios)**

  $\Sigma \,(Ri * XBi) = 1$ $\qquad$ **Ri = YT7i/XBi = F(i, T7, P, L/V)**

  $\qquad\qquad\qquad\qquad\qquad$ **T7 = tray 7 temperature**

  $\qquad\qquad\qquad\qquad\qquad$ **Y7i = vapor composition tray 7**

- **Mass balance equation**

  $\Sigma \,(XBi) = 1$

Figure 3. DeIsobutanizer

Petrocontrol
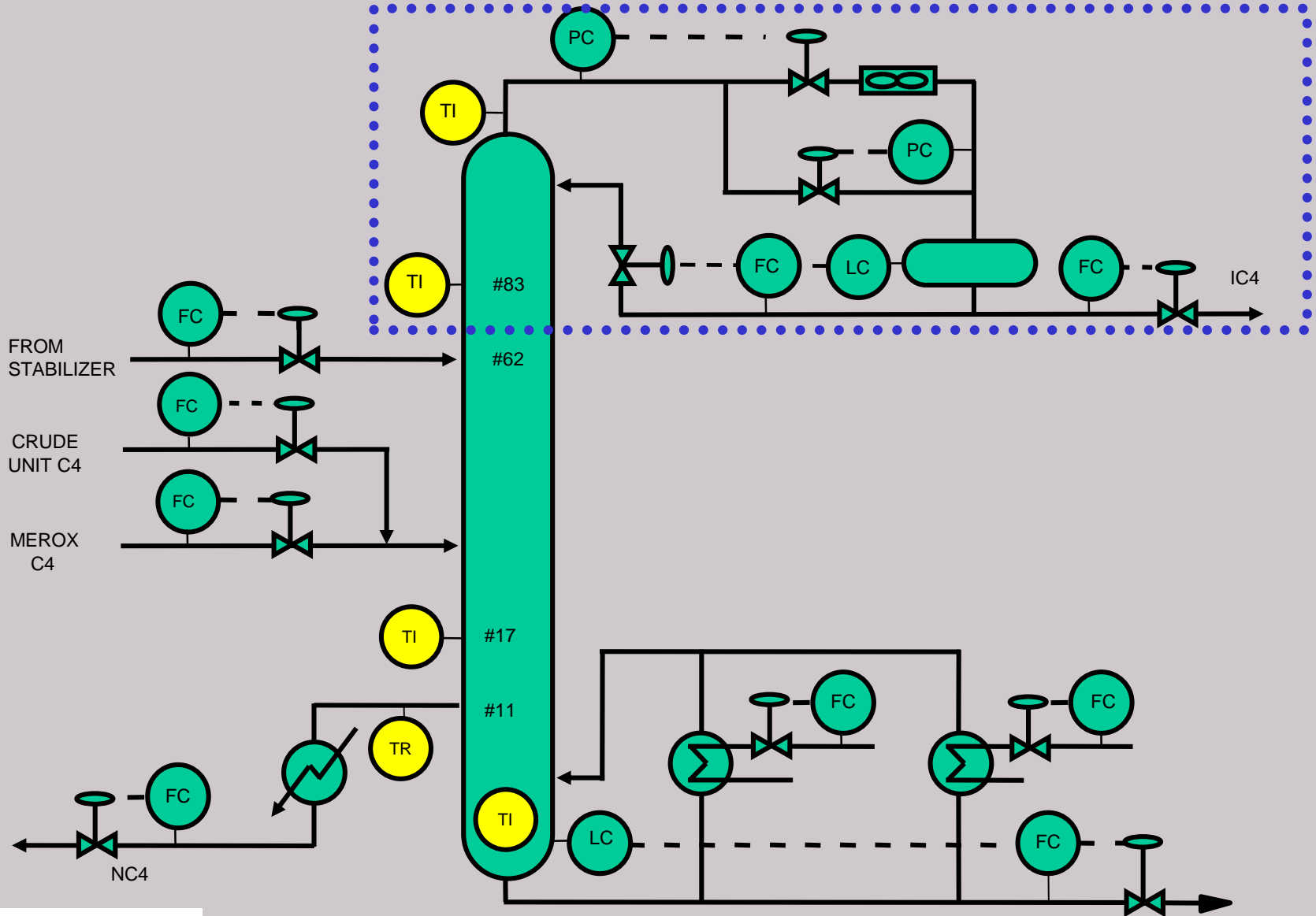
Figure 4. DIB top NC4 inference

NC4 in top
Model

NC4 in top
Analyzer

ONE MONTH

PETROBRAS

Petrocontrol

4
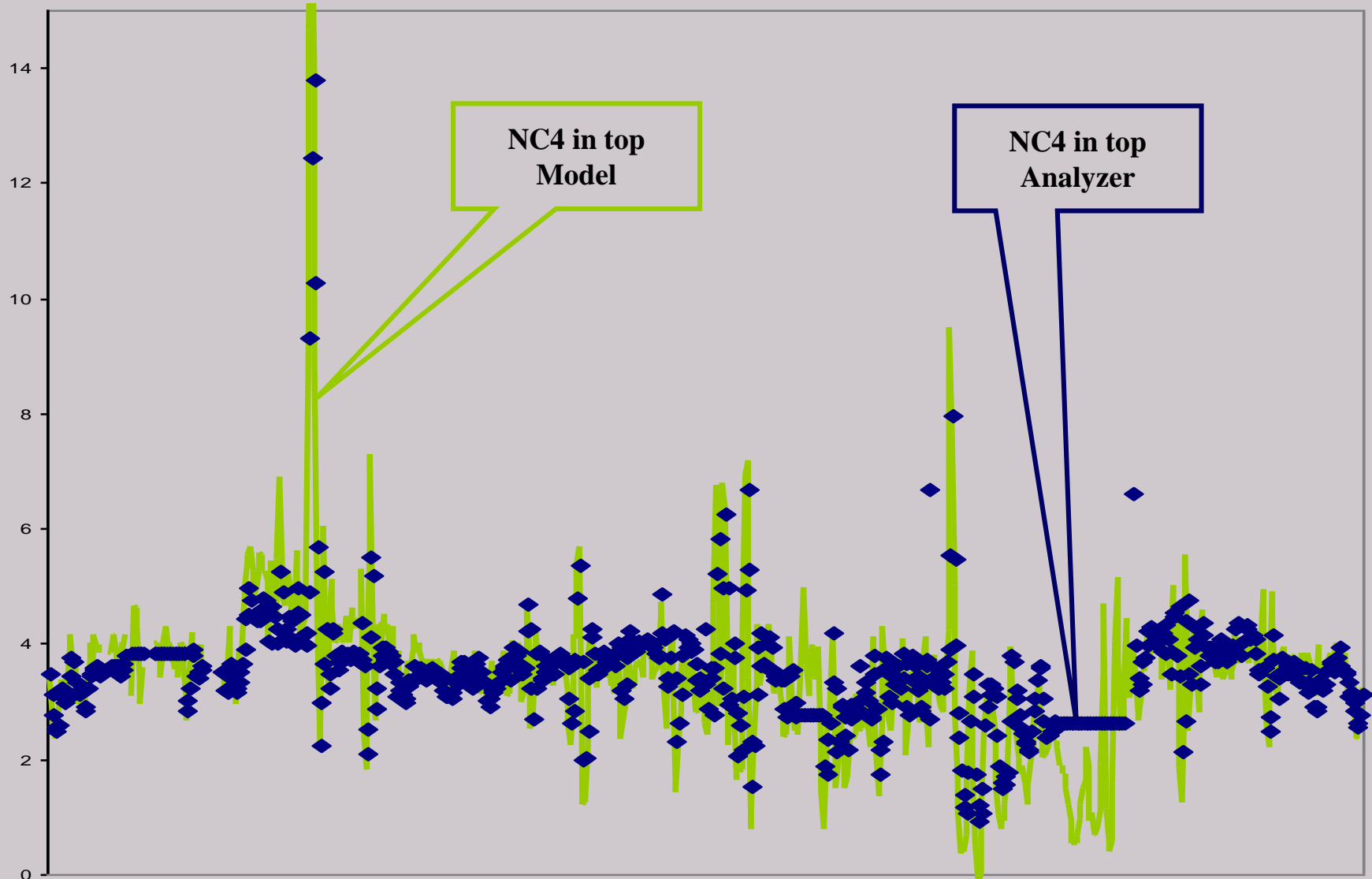
Figure 5. Benzene column

Petrocontrol
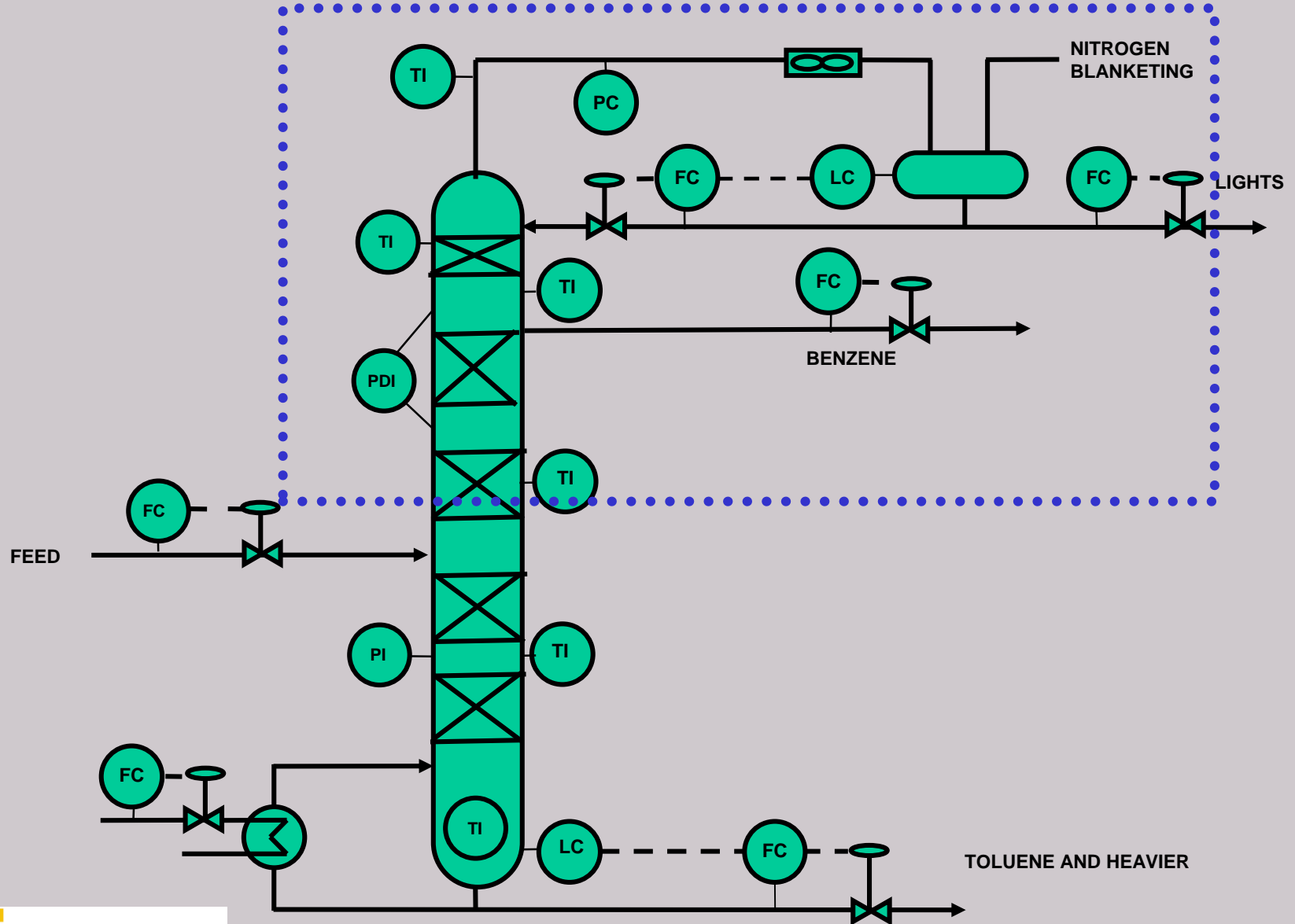
Figure 6. Toluene in Benzene

Petrocontrol
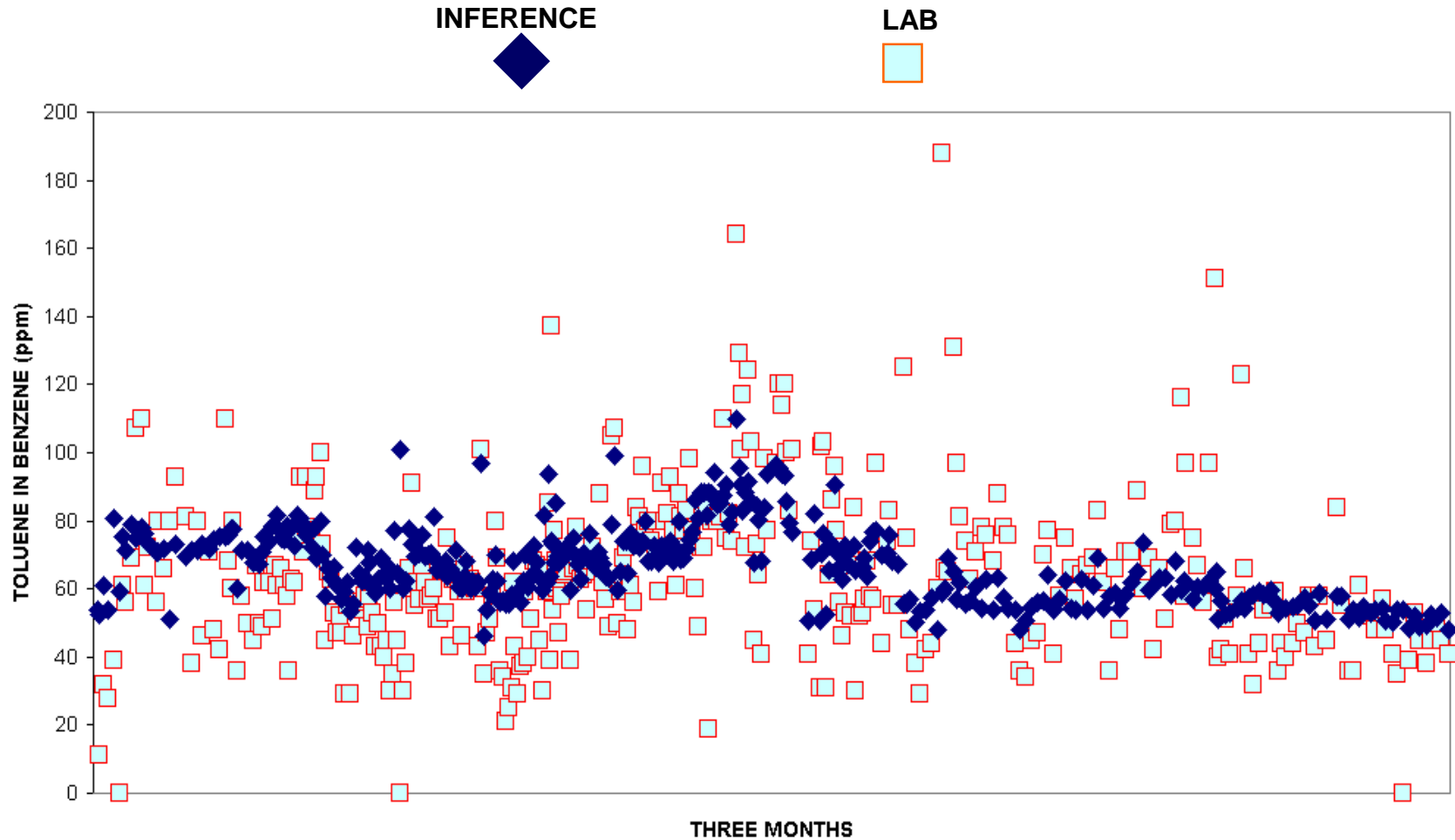
Figure 7. Toluene columns: N08, N09

Petrocontrol
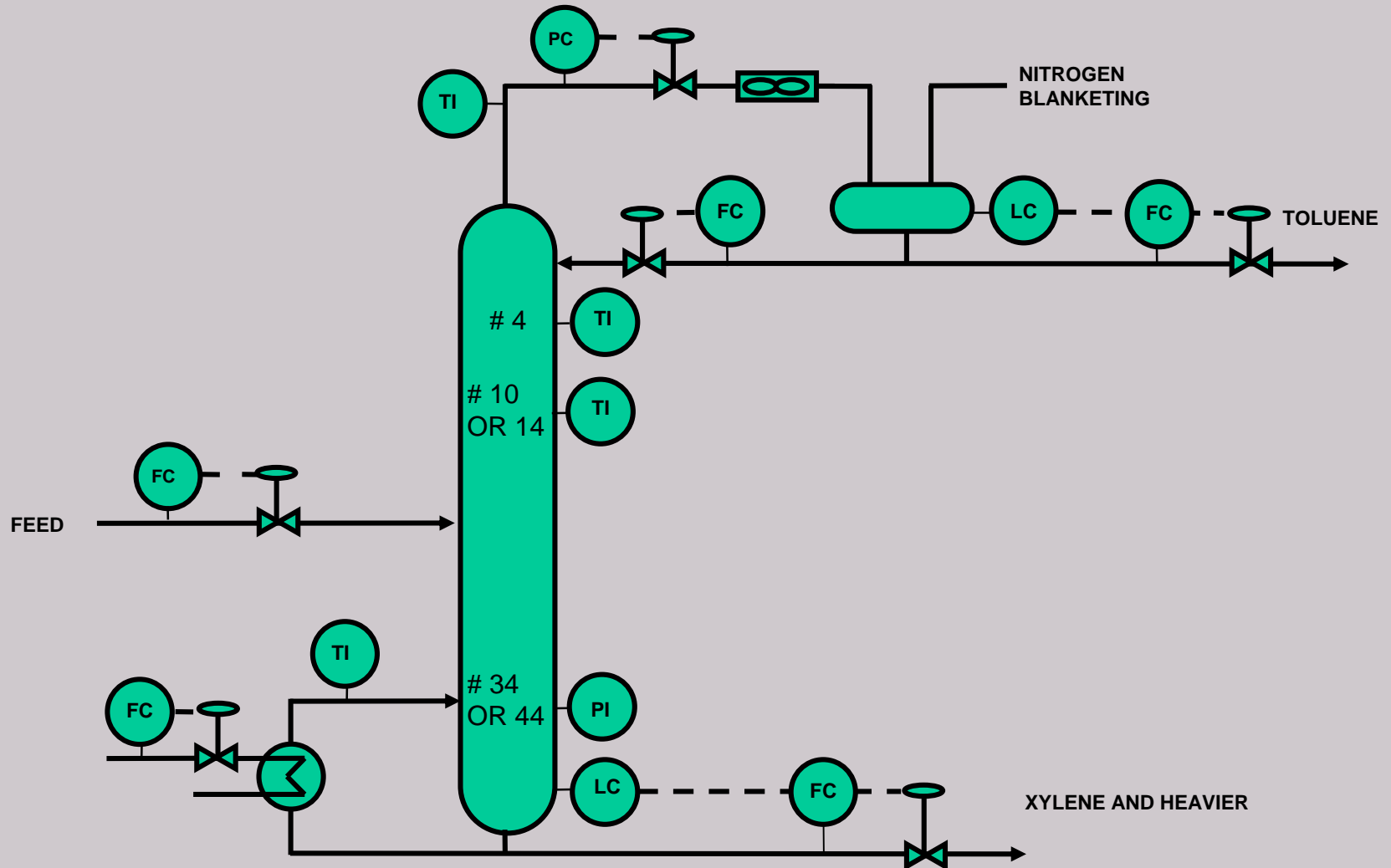
Figure 8. N08 Xylene in Toluene

Inference versus combined lab open loop
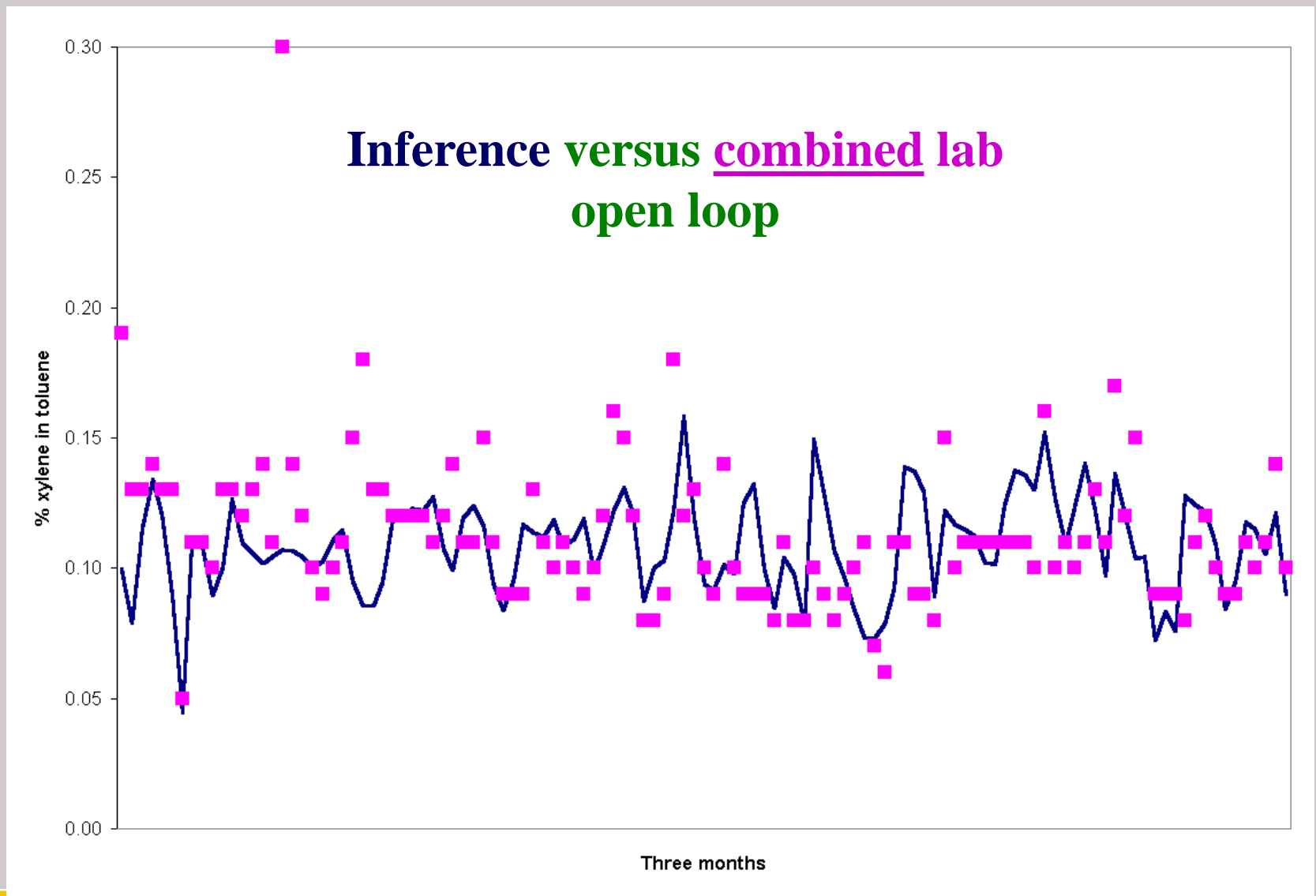
Figure 9. N09 Xylene in Toluene

Inference versus combined lab open loop
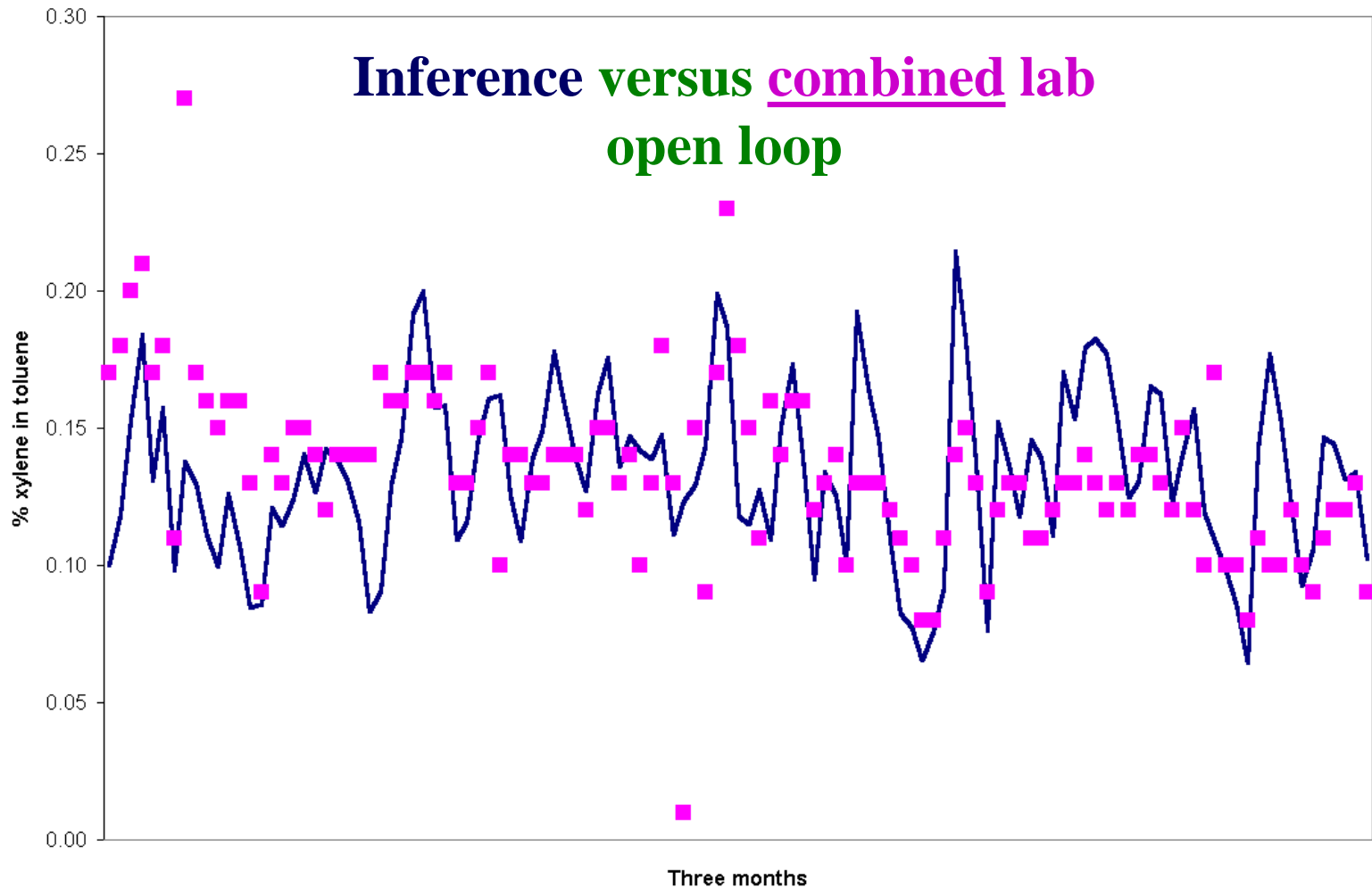
Figure 10. N08 Toluene in Xylene
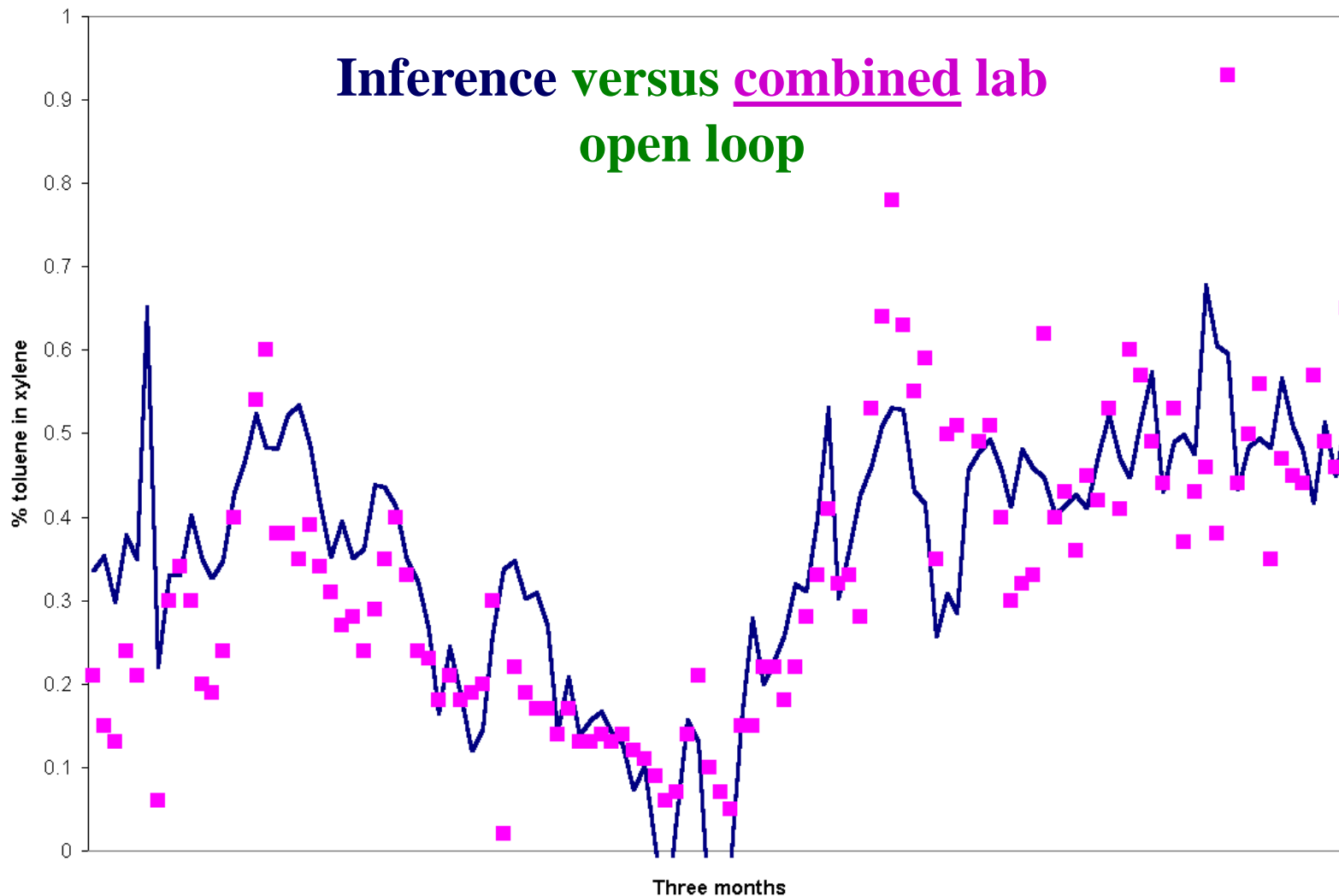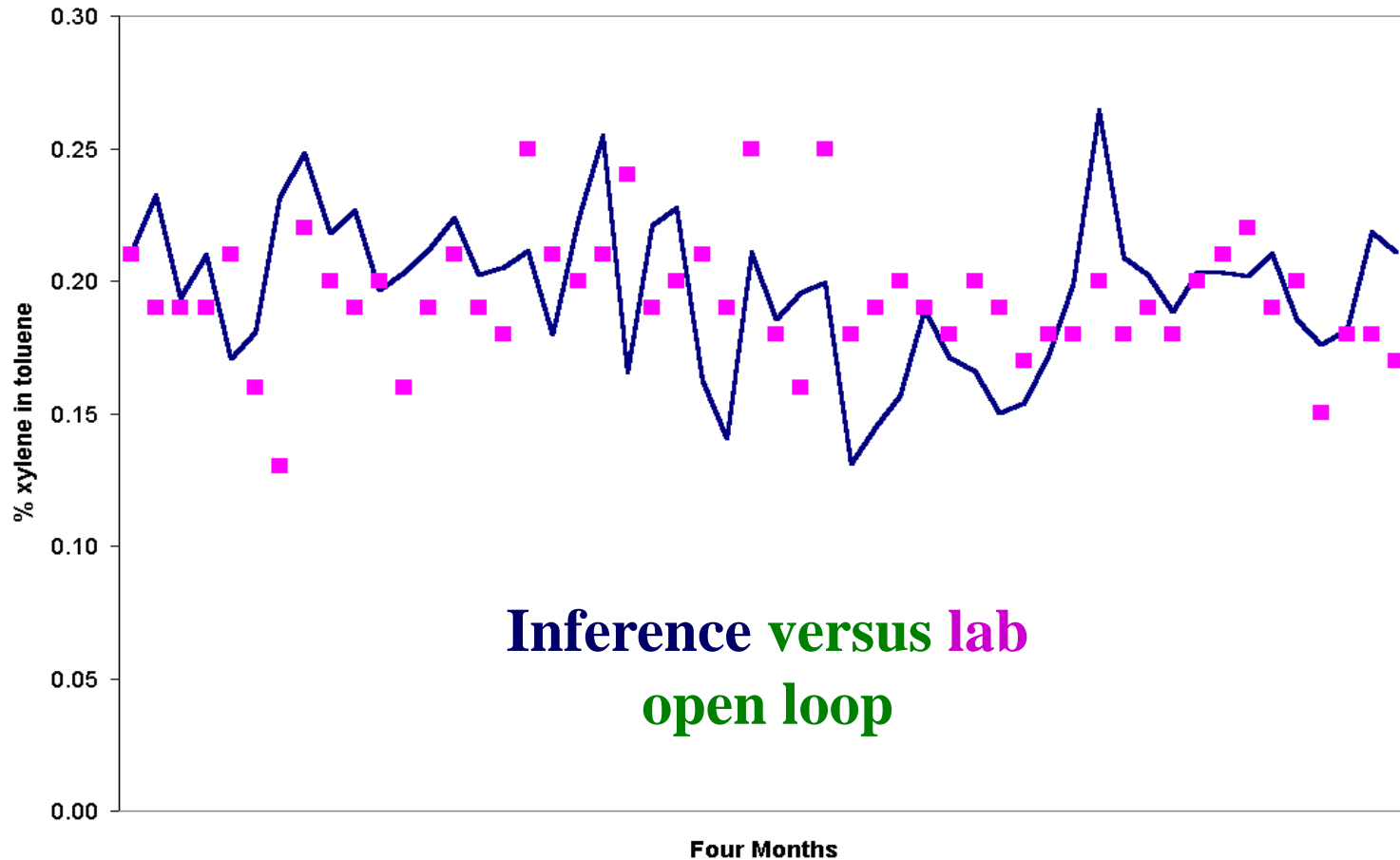
Figure 11. N09 Toluene in Xylene

Inference **versus** **combined** lab open loop

PETROBRAS

Petrocontrol
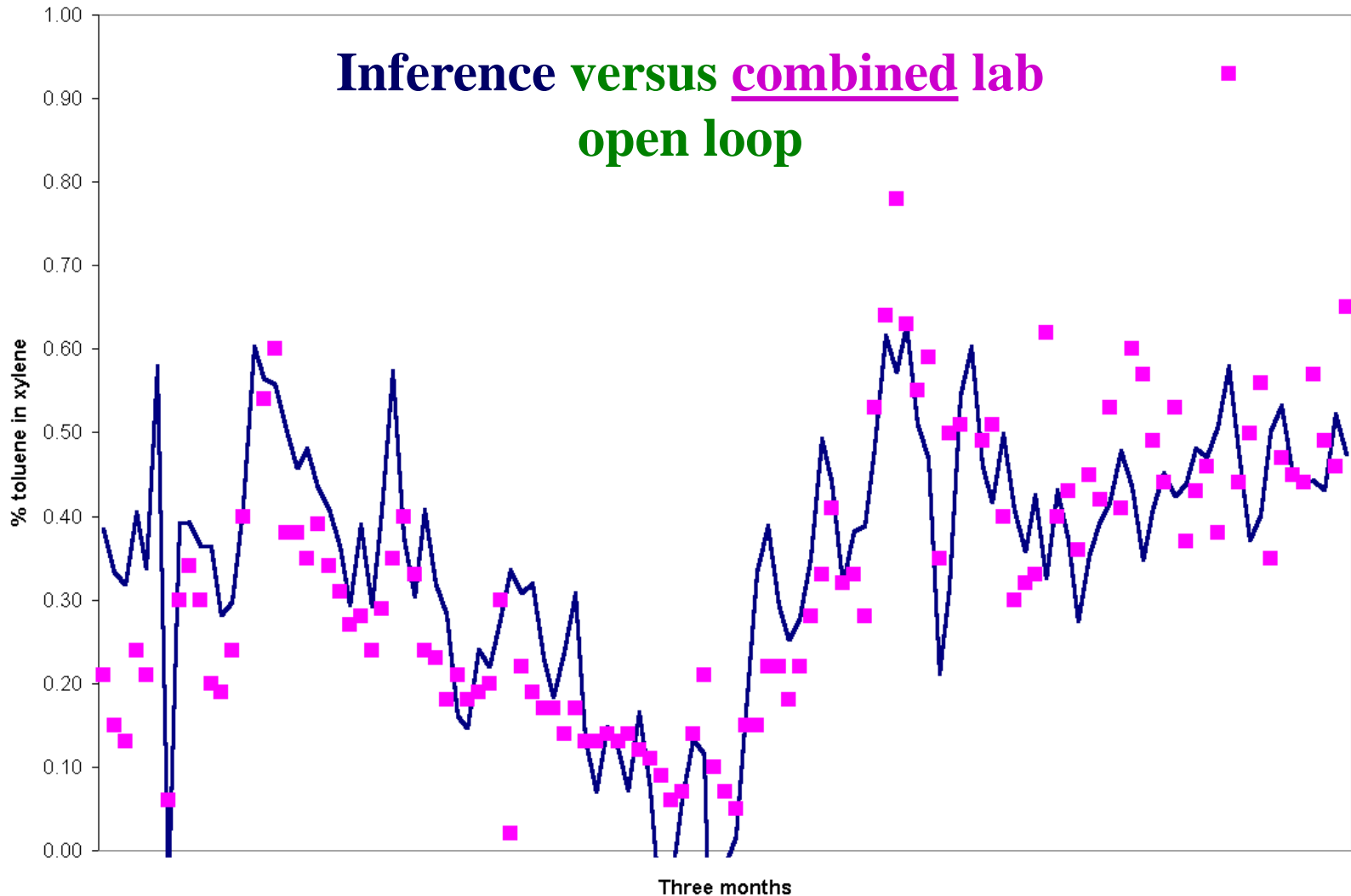
Figure 12. N08 Xylene in Toluene (individual column measurement)

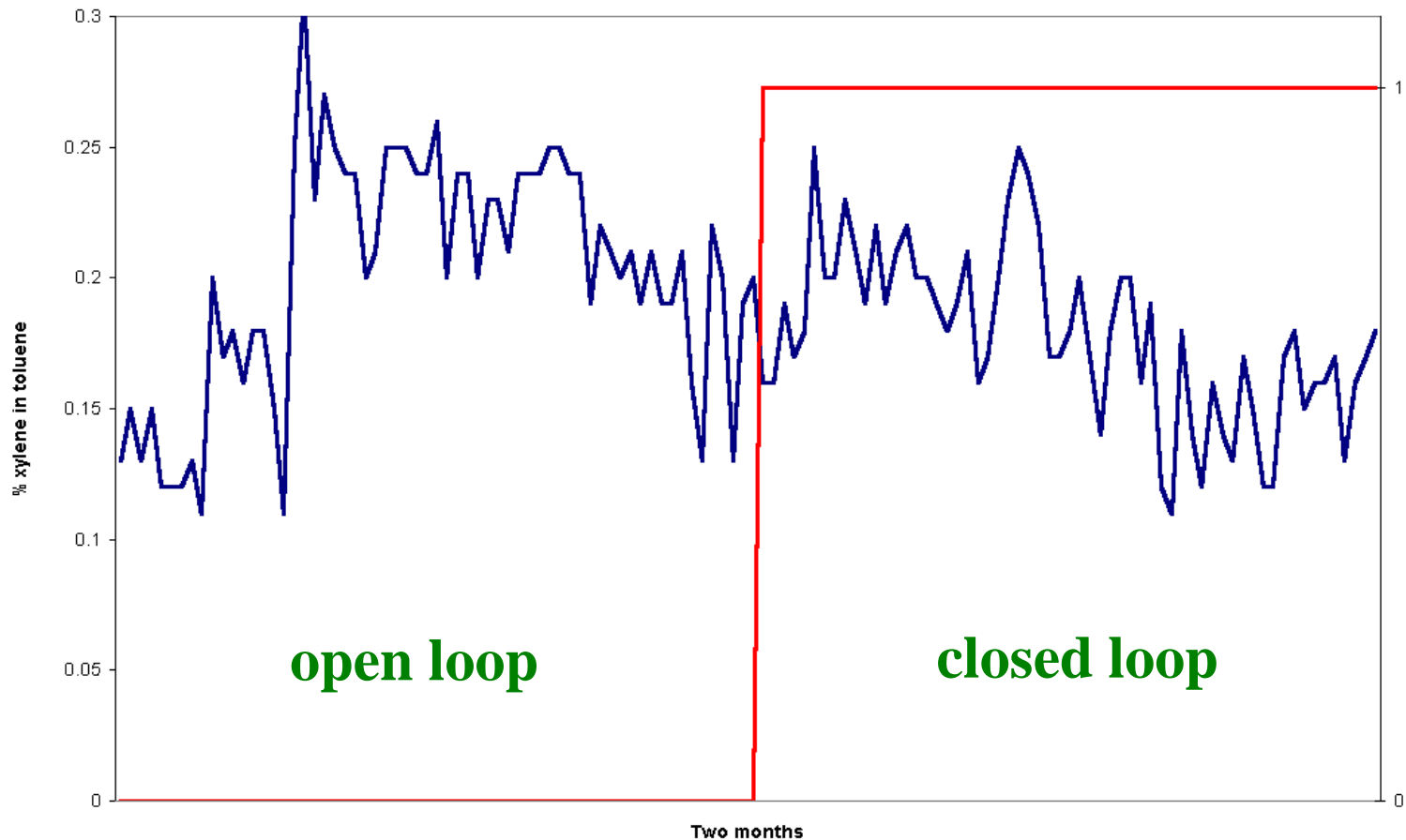Inference versus lab
open loop

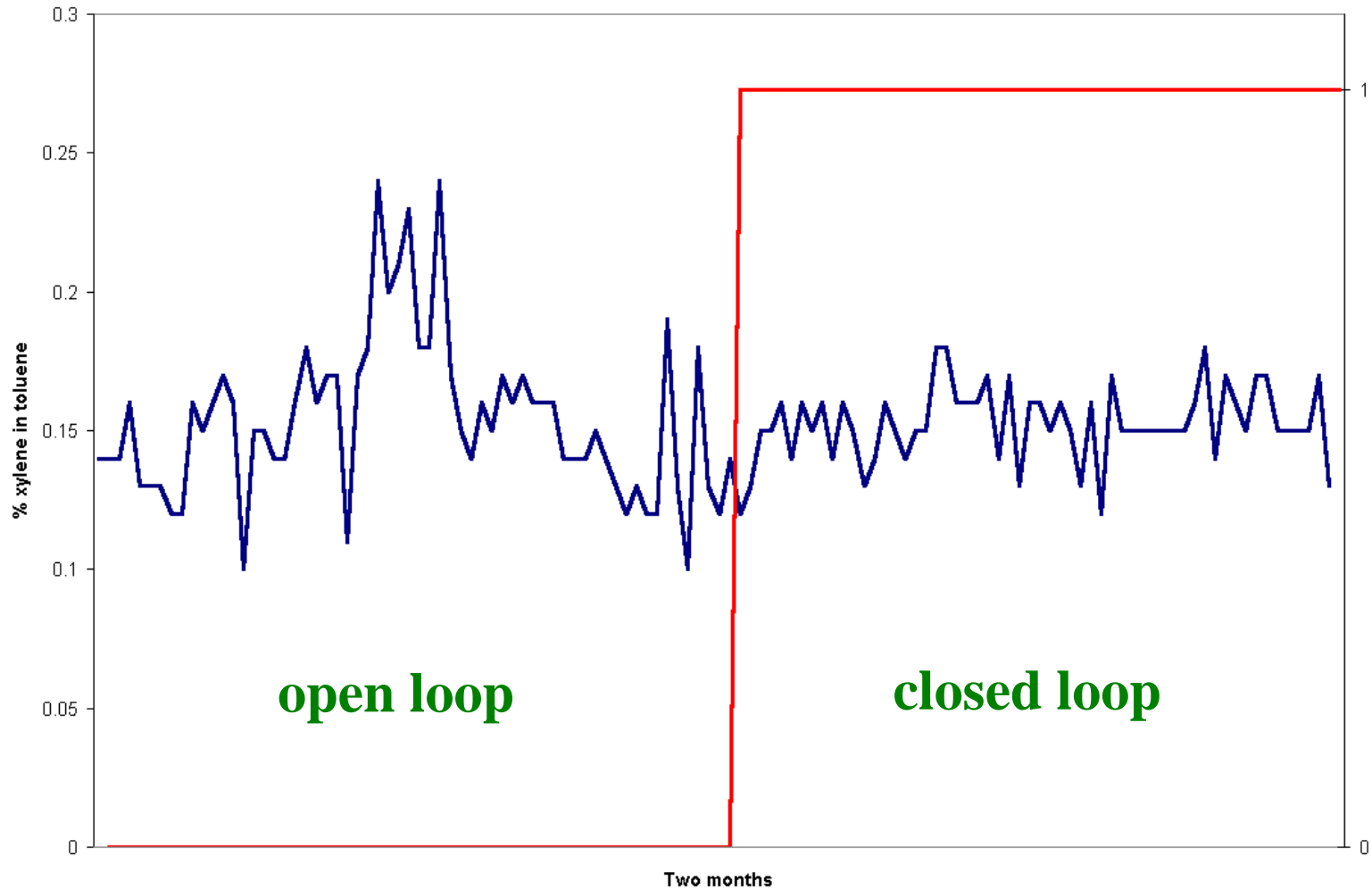# Figure 13. N08 Xylene in Toluene

Figure 14. N09 Xylene in Toluene

open loop   closed loop

Figure 15. N11 Toluene in Xylene

Petrocontrol