Soft sensor models: Bias updating revisited

André D. Quelhas* and José Carlos Pinto**

**Petrobras – Corporate University, Rio de Janeiro - Brazil*
(Tel: +55-21-3487-3467; e-mail:quelhas@ petrobras.com.br).
***Chemical Engineering Program, COPPE-UFRJ, Rio de Janeiro - Brazil*
(e-mail:pinto@peq.coppe.ufrj.br)

Abstract: Bias updating is a widespread adaptive procedure to allow inference models to pursue time variant features of a real world process. The aim of this work is to clarify the statistical consequences of bias updating to soft sensor estimates as well to point up the need of careful analysis of the effect of unmeasured disturbances on the true values of the variable of interest. It is shown that bias updated inferences are unbiased estimates of the true value but yields estimates whose variance are 100% larger than the ones obtained with no use of bias updating. It is suggested the use of a weighting factor to bias updating in order to balance statistical benefits and penalties. A case study of a soft sensor for weathering of LPG in oil refinery exemplifies the concepts discussed.

Keywords: Soft sensor, Bias Updating, Error Analysis, Statistics.

1. INTRODUCTION

The main goal of an industry is to operate as close as possible to the point where profit is maximum. It means that there should be no off-spec product and the lowest degree of product quality give-away should be achieved. Maximum profit is also related to the fact that the set of manipulated variables leads to lower costs by minimizing use of heat, steam, electricity, water etc..

It may be hard if not impossible to accomplish this goal. Real processes are likely to be nonlinear and highly integrated causing modeling and identification prone to errors. In addition, long term operation makes processes more susceptible to hardware upsets (corrosion, fouling, mechanical failures) and to experience environmental disturbances as well qualitative/quantitative changes in physical-chemical properties of feed streams.

Accurate knowledge of process actual model structure and parameters is essential if one intends to predict future states (for control and optimizing) or to diagnose safety risks. Unfortunately many relevant process variables are not available as frequently as desirable or even not available at all. For example, it is very common that physical-chemical properties related to quality control are measured by laboratory tests performed with a very low frequency when compared to process variables acquired by online sensors. Such process with differing sample rates for measured variables are known as multirate process (Ragahavan *et al.*, 2006).

Most of times the long period of time to be awaited before new information about low frequency variables become available is unacceptable. It is necessary to make use of some inferential knowledge based on high frequency information about the process. If a sufficiently accurate model is available, the variable of interest can be estimated from high frequency process measurements **x** as long as model structure and parameters $\boldsymbol{\alpha}$ are known:

$$\hat{y} = f(\mathbf{x}, \boldsymbol{\alpha}) + bias \qquad (1)$$

Every time a new measurement of the *true* value of y is available, an adaptive procedure can be used to adapt the inferential model. The only parameter updated through this one parameter correction is the independent coefficient in (1): $bias = y - f(\mathbf{x}, \boldsymbol{\alpha})$. This simple strategy is very common in industry as well in literature for optimizing purposes (Mercangöz and Doyle 2008; Jesus 2004; Singh 1997) or for soft sensors inferences (Sharmin *et al.* 2006; Mu *et al.* 2006; Tran *et al*. 2005).

Some questions should be posed regarding the use of inferences as (1) for anyone who has to cope with a multirate process:

- What is the best model structure f($\mathbf{x}, \boldsymbol{\alpha}$)?
- How often should bias be updated?
- How are inference errors affected by bias updating?
- What are the effects of unmeasured disturbances on inference errors?

Those questions usually receive unequal importance. A lot of effort has been spent along time to answer the first question. Models have progressively become more complex by using the mathematical weaponry of process modeling (multivariable regression, PCA, neural networks, fuzzy logic). The second question is often answered based on practical matters as availability of laboratory technicians. The last two questions are normally disregarded in spite of their huge consequences on the estimates.

The aim of this work is to pay attention to those usually forgotten questions by remembering the mathematical considerations implicit in models as (1) and answering, from a statistical point of view, what the benefits and penalties of bias updating are.

2. MATHEMATICAL FOUNDATIONS OF BIAS UPDATING

For a steady state system, the generic mathematical relationship linking the output variable, y, and all pertinent process variables, **w**, required by fundamental physical laws, may be expressed as:

$$F(y, \mathbf{w}, \mathbf{c}) = 0 \quad (2)$$

where **w** represents the NW necessary variables to perfectly predict the unknown behavior of y given the NC constants in the vector of parameters, **c**.

Two practical reasons explain why it is unlike that any real model would incorporate the whole set of NW necessary variables. The first one is the fact that NW may be large and would conflict with science´s parsimony principle. In this sense, a less complete description would be acceptable in a trade-off for simplicity under a certain allowable tolerance. The other reason is that several of the NW variables either are not measured or are not considered relevant by the scientist due to a methodological error.

Taking these reasons under consideration one can split **w** into the subsets **x** and **z**. The first subset contains the NX measured variables that were chosen as relevant for the model. The second subset contains the remaining NZ = NW-NX variables. It contains measured and unmeasured variables that should be part of a perfect model but were set apart. The complete description of the system behavior is then expressed as:

$$F(y, \mathbf{x}, \mathbf{z}, \mathbf{c}) = 0 \quad (3)$$

In the process of justifying the possibility of a correction as proposed in (1) it is required that (3) be partially separable with respect to addition at least with respect to y. It requires that $(1/F)\partial \exp(F)/\partial y$ depends only on y (Viazminsky 2008). If this condition is satisfied one can express (3) as:

$$g(y) = F_1(\mathbf{x}, \mathbf{z}, \mathbf{c}_1) \quad (4)$$

Additionally, if the inverse function g^{-1} exists, then:

$$y = g^{-1}(F_1(\mathbf{x}, \mathbf{z}, \mathbf{c}_1)) = F_2(\mathbf{x}, \mathbf{z}, \mathbf{c}_2) \quad (5)$$

Physical knowledge or empirical insight may lead to an attempt to predict y based on measurements **x** and parameters $\boldsymbol{\alpha}$ by means of a model $f(\mathbf{x}, \boldsymbol{\alpha})$. If **z** is an empty set and the whole influence of **x** on y is taken into account by $f(\mathbf{x}, \boldsymbol{\alpha})$ we have a perfect model. Otherwise one should expect a relationship as (6), where $F_3(\mathbf{x}, \mathbf{z}, \mathbf{c}_3)$ plays the role of bias as in (1). It should be noticed that (6) is derived from (5) if $F_3(\mathbf{x}, \mathbf{z}, \mathbf{c}_3)$ is a separable function with respect to the set **z**.

$$y = f(\mathbf{x}, \boldsymbol{\alpha}) + F_3(\mathbf{x}, \mathbf{z}, \mathbf{c}_3) \quad (6)$$

The model built by the experimenter is $f(\mathbf{x}, \boldsymbol{\alpha})$. The invisible part of the true model is $F_3(\mathbf{x}, \mathbf{z}, \mathbf{c}_3)$. This term is captured by the bias term in a very common pragmatic approach assuming the form (1).

Inference structure (6) is very attractive but it is valid only if the assumptions that allowed disregarding more generalized expressions (3)-(5) are true. If not, there will be no guarantee that successive inferred values will express the true values y even if no further disturbances alter the values of the set **z**. This can be seen by comparing two simple models. One represents a model as expressed in (5) (type A model) and the other one represents the less generic model expressed in (6) (type B model), for instance:

type A true model: $y = (x+z)/x$

type B true model: $y = x + z$

It should be noticed that the type B true model in this example shows no dependence of F_3 on x. This class of true models yields the best possible performance for an adaptive experimental model as (1).

Assuming that: 1) experiments to identify the inference $f(x, \alpha)$ were carried out under controlled conditions in order to keep z at a constant value z_0 in both cases and 2) perfect model identification led to inferences with the same mathematical structure than true models:

type A inferred model: $\tilde{y} = (x + z_0)/x$

type B inferred model: $\tilde{y} = x + z_0$

If the inferred models were parameterized by means of proper statistical criticism both inferred models will adequately represent the behavior of the variable of interest. However, as time passes, it is possible that z assumes values different of the one kept controlled along identification phase. So, if z assumes the value z_1 and $x=x_1$ at the moment of correction in both cases, according to the bias updating routine:

type A true value: $y_1 = (x_1+z_1)/x_1$,

type A inferred value: $\tilde{y}_1 = (x_1 + z_0)/x_1$

$\Rightarrow$ bias = $y_1 - \tilde{y}_1$ = $(z_1 - z_0)/x_1$

corrected inference: $\hat{y}_1$ = $(x + z_0)/x + ((z_1 - z_0)/x_1)$

type B true value: $y_1 = x_1 + z_1$,

type B inferred value: $\tilde{y}_1 = x_1 + z_0$

$\Rightarrow$ bias = $y_1 - \tilde{y}_1 = z_1 - z_0$, corrected inference: $\hat{y}_1 = x + (z_1 - z_0)$

It is clear that, after bias correction, inferences derived from type B models will produce results as close to the truth as they were before the change of z value as long as this variable is kept constant from this change on. On the other hand, inferences derived from type A models will not behave this way because accuracy of the corrected inference will be affected not only by further changes of z value but also by additional changes in the x value because the nonlinear behavior is not captured by a single point correction.

3. BIAS UPDATING PROCEDURE

In order to describe the behavior of predictions of the value *y* along time it is interesting to write inference model to allow time course to be taken into account:

$$\tilde{\mathbf{y}} = f(\mathbf{X}, \boldsymbol{\alpha}), \quad \mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \dots \\ \mathbf{x_{NS}} \end{bmatrix}, \quad \mathbf{x_i} = \begin{bmatrix} x_{i1} & x_{i2} & x_{i3} & \dots & x_{i,NX} \end{bmatrix},$$

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \dots \\ \tilde{y}_{NS} \end{bmatrix} \qquad (7)$$

where NS is the number of time samples of the process variable signals.

Corrected values of *y* along time are obtained from bias updating according to:

$$\hat{\mathbf{y}} = \tilde{\mathbf{y}} + \mathbf{bias} \qquad (8)$$

where the array of time values of bias is built according to:

$bias_1 = 0$

$bias_k = \left(y_k^m - \tilde{y}_k\right)s_k + \left(1 - s_k\right)bias_{k-1}$

leading to:

$$\mathbf{bias} = \begin{bmatrix} 0 \\ \left(y_2^m - y_2\right)s_2 + \left(1 - s_2\right)bias_1 \\ \left(y_3^m - y_3\right)s_3 + \left(1 - s_3\right)bias_2 \\ \dots \\ \left(y_{NS}^m - y_{NS}\right)s_{NS} + \left(1 - s_{NS}\right)bias_{NS-1} \end{bmatrix} \qquad (9)$$

Vector $\mathbf{y^m}$ contains measurements of the true values $\mathbf{y}$ sampled with period at a lower rate, T_{meas}, than primary variables of the model. Vector **s** is a binary set that indicates when true values $\hat{y}$ are available:

$$\mathbf{y^m} = \left[[0]_{1x(Tmeas-1)} \;\; y_{Tmeas} \;\; [0]_{1x(Tmeas-1)} \;\; y_{2Tmeas} \;\dots\right]^T$$

$$\mathbf{s} = \left[[0]_{1x(Tmeas-1)} \;\; 1 \;\; [0]_{1x(Tmeas-1)} \;\; 1 \;\dots\right]^T$$

4. STATISTICAL IMPACT OF BIAS UPDATING

Although equations (7-9) indicate the *modus operandi* of inference correction, it is not clear how our expectation about error values is affected. Since corrections are made at a low frequency the duration of their benefits will be affected by the probability of occurrence of new disturbances before a new gold standard measurement is ready, thus making possible another correction. A reasonable question would be: what benefits are obtained with periodic bias updating comparing with no bias correction at all?

In fact, bias updating and no updating schemes are extreme points of a continuous range of possible single point corrections. Considering the weight parameter $\varphi \in \Re$, $\varphi \subset [0\ 1]$, the time values $\bar{\mathbf{y}}$ are a weighted mean of bias corrected values (8) and values from the original inference model (7):

$$\bar{\mathbf{y}} = \varphi\hat{\mathbf{y}} + (1 - \varphi)\tilde{\mathbf{y}} \qquad (10)$$

If samples of true values are taken with period T_{meas} the n^{th} element suffers the effects of the last bias updating made at sample $i = \mathrm{int}(n/T_{meas})T_{meas}$, where int(x) retains the integer part of the floating point real number x:

$$\overline{y_n} = \varphi\left(\tilde{y}_n + bias_n\right) + (1 - \varphi)\,\tilde{y}_n \qquad (11)$$

$$\overline{y_n} = \varphi\left(\tilde{y}_n + y_i - \tilde{y}_i\right) + (1 - \varphi)\,\tilde{y}_n = \tilde{y}_n + \varphi\left(y_i - \tilde{y}_i\right) \qquad (12)$$

Inference error at the n^{th} element will be:

$$\varepsilon_n = y_n - \overline{y_n} = y_n - \tilde{y}_n - \varphi y_i + \varphi \tilde{y}_i \tag{13}$$

Since n^{th} and i^{th} elements of the true values come from the same sample space as well n^{th} and i^{th} elements of the inferred values, their statistical moments are the same, i.e., $E[y_n] = E[y_i]$ and $E[\tilde{y}_n] = E[\tilde{y}_i]$. Dropping indexes to simplify notation, it is possible to say that the expected error value is:

$$E[\varepsilon] = (1-\varphi)E[y] + (\varphi-1)E[\tilde{y}] \tag{14}$$

$$if \begin{cases} \varphi = 1, & E[\varepsilon] = \varepsilon_{\min} = 0 \\ \varphi = 0, & E[\varepsilon] = \varepsilon_{\max} = E[y] - E[\tilde{y}] \end{cases}$$

It can be seen that the bias update scheme expressed in (8) ($\varphi = 1$) guarantees mean error value of zero if length of $\mathbf{y}$ tends to infinity. If no correction is made ($\varphi = 0$), long term error mean depends on the ability of model $f(\mathbf{X}, \boldsymbol{\alpha})$ to be an unbiased estimate of the true value. It is also possible to investigate the dependence of error variance on the choice of φ. From (13) it is possible to write:

$$\operatorname{var}(\varepsilon_n) = \operatorname{var}(y_n - \tilde{y}_n - \varphi y_i + \varphi \tilde{y}_i) \tag{15}$$

$$\begin{aligned} \operatorname{var}(\varepsilon_n) = {} & \operatorname{var}(y_n) + \operatorname{var}(\tilde{y}_n) + \varphi^2 \operatorname{var}(y_i) + \varphi^2 \operatorname{var}(\tilde{y}_i) \\ & - 2\operatorname{cov}(y_n, \tilde{y}_n) + 2\varphi \operatorname{cov}(y_n, \tilde{y}_i) - 2\varphi \operatorname{cov}(y_i, \tilde{y}_n) \\ & - 2\varphi^2 \operatorname{cov}(y_i, \tilde{y}_i) \end{aligned} \tag{16}$$

For the same reason explained above $\operatorname{var}(y_n) = \operatorname{var}(y_i)$ and $\operatorname{var}(\tilde{y}_n) = \operatorname{var}(\tilde{y}_i)$, making it more convenient to drop subscripts and simplify (16):

$$\begin{aligned} \operatorname{var}(\varepsilon) = {} & \operatorname{var}(y) + \operatorname{var}(\tilde{y}) + \varphi^2(\operatorname{var}(y) + \operatorname{var}(\tilde{y})) - 2\operatorname{cov}(y, \tilde{y}) \\ & - 2\varphi^2 \operatorname{cov}(y, \tilde{y}) \end{aligned} \tag{17}$$

At the extreme points of φ:

$$if \begin{cases} \varphi = 1, & \operatorname{var}(\varepsilon) = v\varepsilon_{\max} = 2\operatorname{var}(y) + 2\operatorname{var}(\tilde{y}) - 4\operatorname{cov}(y, \tilde{y}) \\ \varphi = 0, & \operatorname{var}(\varepsilon) = v\varepsilon_{\min} = \operatorname{var}(y) + \operatorname{var}(\tilde{y}) - 2\operatorname{cov}(y, \tilde{y}) \end{cases}$$

With respect to the error variance the progressive updating ($\varphi = 1$) doubles the value obtained when no correction is made ($\varphi = 0$). Confronting this result with the expected value of the error one can see that bias updating is associated with an expectation of unbiased mean value of estimates but it also causes a 100% increase in error variance. There would be a choice of φ to cope with these consequences? In order to answer this question it is necessary to create a single objective function that combines both effects.

As an example, a possible choice for such function could be $\psi = E[\varepsilon] + \operatorname{var}(\varepsilon)$, choosing φ that minimizes its value. However this function is too dependent of the problem specificities and units of measurement. In fact even the choice the objective function depends on the problem to be solved and on the needs of the plant personnel in order to fulfill several goals related to the industrial process.

Taking this into consideration, it is suggested a very simple objective function, derived from the previous one. It represents an attempt to equalize the importance of the effects of φ regarding each statistical moment. Such function could assume the normalized form:

$$\psi = E[\varepsilon]_{norm} + \operatorname{var}(\varepsilon)_{norm} \tag{18}$$

where

$$E[\varepsilon]_{norm} = \frac{E[\varepsilon] - \varepsilon_{\min}}{\varepsilon_{\max} - \varepsilon_{\min}} \tag{19}$$

and

$$\operatorname{var}(\varepsilon)_{norm} = \frac{\operatorname{var}(\varepsilon) - v\varepsilon_{\min}}{v\varepsilon_{\max} - v\varepsilon_{\min}} \tag{20}$$

Substituting (19-20) in (18):

$$\begin{aligned} \psi = {} & \frac{(1-\varphi)E[y] + (\varphi-1)E[\tilde{y}]}{E[\hat{y}] - E[y]} \\ & + \frac{\varphi^2(\operatorname{var}(y) + \operatorname{var}(\tilde{y})) - 2\varphi^2 \operatorname{cov}(y, \tilde{y})}{\operatorname{var}(y) + \operatorname{var}(\tilde{y}) - 2\operatorname{cov}(y, \tilde{y})} \end{aligned} \tag{21}$$

The choice of φ is made in order to minimize ψ and is represented by the solution of:

$$\frac{\partial \psi}{\partial \varphi} = 1 - \varphi - \varphi^2 = 0 \;\Rightarrow\; \varphi = 1/2 \tag{22}$$

It is interesting to see how formalism of (21) and (22) conducts to a common sense value of ½ for the weighting factor in this case.

5. CASE STUDY

In this section it will be shown the statistical features of bias updating in a soft sensor to be implemented in an oil refinery. The process unity at study is a FCC debutanizer showed in figure 1. In order to improve quality control of liquefied petroleum gas (LPG) it is desirable to have online information about the relative amount of molecules with more than four carbon atoms present on LPG stream. A laboratory or field test usually carried out a few times a day measures weathering of LPG, expressed in temperature units, which is correlated to the ratio of heavier molecules.

An empirical mathematical model of LPG weathering based on NX = 3 process variables feeds the model predictive control of the process unity with inferred values along time as in (7).

For the purposes of this work, actual behavior of the unity is represented by data from customized process simulation software. The discrete mathematical space of operating scenarios has its basis formed by the N_{inp} = 4 process simulation input parameters as shown in figure 1.

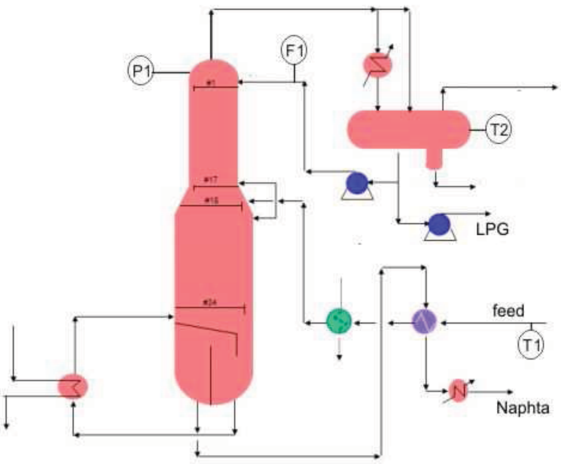

Figure 1 – FCC debutanizer. Process variables used as inputs for the process simulator: P1, F1, T1, T2.

The subregion of operation considered for analysis was the regular mesh **S** (N_{sc} x N_{inp}) of equally spaced points around nominal condition of operation. This region of operation induces the subregion **χ** (N_{sc} x NX) of the input variables of the empirical model of weathering. For simulation of long term operation a string of scenarios, $\mathbf{S}^{str}$ (L_{sc} x N_{inp}), representing the time course of conditions of operation, was assembled:

$$ind_i \sim \text{Unif}(1, N_{sc});\ \ ind_i \in \mathbb{N}\ ;\ i = [1\ 2\ ...\ L_{sc}]$$

Each choice ind_i is a uniformly distributed random variable that indicates where, in the subregion **S**, is the i^{th} line of $\mathbf{S}^{str}$ and, as consequence, maps **X** (L_{sc} x NX) as in (7):

$$\begin{aligned} &\mathbf{S}(ind_i, \mathbf{j}) \to \chi(ind_i, \mathbf{k}) \\ &\mathbf{S}^{str} = \mathbf{S}(\mathbf{ind}, \mathbf{j}) \to \mathbf{X} = \chi(\mathbf{ind}, \mathbf{k}) \\ &\mathbf{j} = [1\ \ 2\ \ ...\ \ N_{inp}],\ \mathbf{k} = [1\ \ 2\ \ ...\ \ NX] \end{aligned} \tag{22}$$

Since quality of the feed is a major unmeasured disturbance the set of variables **z** is represented by the ratio of the slope of the true boiling point curve of the actual feed related to the one at nominal operating condition. If feed stream may have three different compositions symmetrically disturbed:

$$\mathbf{r} = [0.95\ \ 1\ \ 1.05]$$

$$ind_i \sim \text{Unif}(1, 3);\ \ ind_i \in \mathbb{N};\ i = [1\ 2\ ...\ L_{sc}]$$

$$\mathbf{z} = \mathbf{r}(\mathbf{ind}) \tag{23}$$

In order to allow a better understanding of the different effects observed in the results there will be considered two cases of study. In the more generic case A it is supposed that the set of variables **z** is represented by (23) and that the inference model is the actual one used in industrial practice. Case B will also take disturbances as in (23) into account but it is supposed that the inference model was perfectly modeled in the absence of disturbances. It is perfect in the sense that all the effects of the model input variables perfectly propagate to the output variable. In other words, at $z=z_{nominal}$, $F_3 = F_3(z,c3)$ as in (6) and the inference is correct for any value of **x**.

As it can be seen in figure 2, in both cases bias updating procedure yields an expected mean value of zero although values show less dispersion when no bias correction is used.

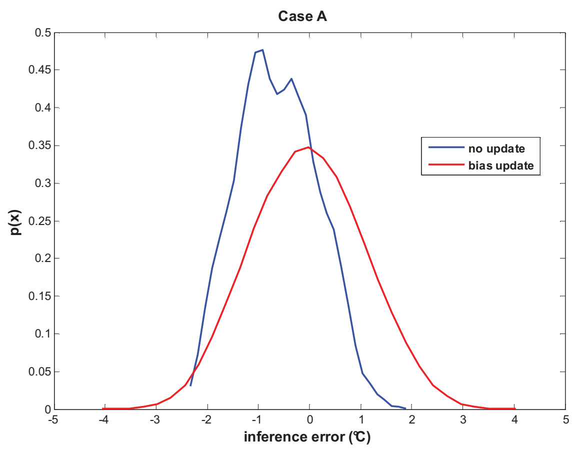

Figure 2 – Estimated probability density function of inferred weathering values for case A.

Estimated probability density function for case B (fig. 3) shows additional features. In this case it is clear that, with no update, the inference will be correct every time $z=z_{nominal}$ whatever the **x** values. The two triangular areas under the blue line around the central peak in figure 3 are originated when $z= z_{nominal}\pm\Delta z$. It should be noticed that the fact that those areas are not as thin as the central peak is due to the dependence of F3 (6) on **x**. When bias update is implemented, two more regions appear as well all regions become flatter. It is because bias expected values will be the result of the difference of all possible two random samples respectively chosen from the sample space of the non corrected inferred values and from the sample space of true values. These bias values will be summed to the inferred ones creating the oscillations of the red line at extreme inference errors observed in figure 3.

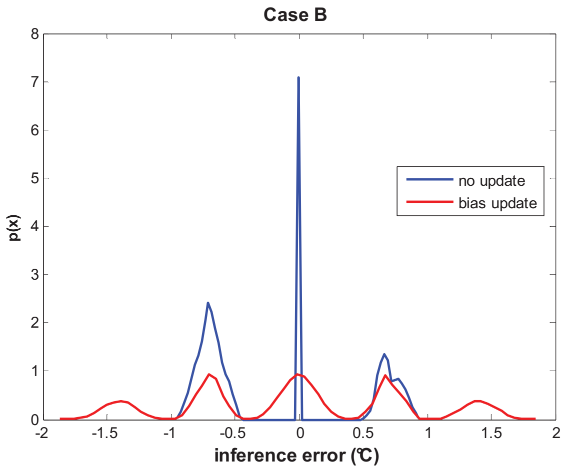

Figure 3 – Estimated probability density function of inferred weathering values for case B.

6. CONCLUSIONS

This work addressed the problem of continuous time monitoring in processes with differing sample rates for measured variables. Bias updating is a common adaptive procedure to periodically correct soft sensor models estimates. It was shown that this strategy is associated with long term zero mean error but at very high cost of 100% increase in variance of estimates. Our intention was to shown that a procedure to implement periodical parameter update should be problem-specific. It means to take into account statistical impact on estimates based on prior knowledge of probability density of disturbances as well error magnitude of soft sensor estimates.

ACKNOWLEDGMENT

The authors are grateful to Eng. Cristina Neves Passos, from Petrobras, for providing simulation files from FCC plant that were used as templates for our work.

REFERENCES

Jesús, R.A.J. (2004), *Modelación y Optimización del Mezclado de Petróleo Crudo con Redes Neuronales*, Master thesis, Centro de Investigación y de Estudos Avanzados del Instituto Politécnico Nacional de Mexico.

Mercangöz, M., Doyle, F.J. (2008), Real-time optimization of the pulp mill benchmark problem, *Computers and Chemical Engineering*, 32, 789-804.

Mu, S., Zeng, Y., Liu, R., Wu, P., Su, H., Chu, J. (2006), Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process, *Journal of Process Control*, 16, 557-566

Ragahavan, H.; Tanrigala, A.K.; Gopaluni, B.; Shah, S. (2006), Identification of chemical processes with irregular output sampling, *Control Engineering Practice*, 14, 467-480.

Sharmin, R., Sundararaj, U., Shah, S., Griend, L.V., Sun, Y. (2006), Inferential sensors for estimation of polymer quality parameters - Industrial application of a PLS-based soft sensor for a LDPE plant, *Chemical Engineering Science*, 61, 6372-6384.

Singh, A. (1997). *Modeling and Model Updating in the Real-Time Optimization of Gasoline Blending*, Master thesis, University of Toronto.

Tran, M., Varvarezos, D.K., Nasir, M. (2005), The importance of first-principles, model-based steady-state gain calculations in model predictive control - a refinery case study, *Control Engineering Practice*, 13, 1369-1382.

Viazminsky, C.P. (2008). Necessary and sufficient conditions for a function to be separable. *Applied Mathematics and Computation*, 204, 658-670.