

STATSSCANDLEPLOT: A NEW WAY OF MONITORING OPERATIONAL PERFORMANCE INDICATORS

Cláudia D. M. Giaquinto^{1*}, Jorge O. Trierweiler¹, Herbert Teixeira²,
Luciane F. Trierweiler¹ and Marcelo Farenzena¹

¹ Universidade Federal do Rio Grande do Sul, Departamento de Engenharia Química, Grupo de Intensificação, Modelagem, Simulação, Controle e Otimização de Processos, Porto Alegre, RS, Brasil. E-mail: claudiag@enq.ufrgs.br, ORCID: 0000-0003-4860-8819; ORCID: 0000-0002-6328-945X; ORCID: 0000-0003-4497-7334; ORCID: 0000-0001-7134-7245

² Centro de Pesquisa e Desenvolvimento Leopoldo Américo Miguez de Mello, Rio de Janeiro, RJ, Brasil.

(Submitted: August 23, 2017 ; Revised: April 9, 2018 ; Accepted: June 12, 2018)

Abstract - Operational KPIs play an extremely important role in the process industry, aiding in decision making. However, they need to be reliably calculated to be representative. The present work presents a schematic methodology for the calculation of these KPIs, including techniques of steady-state detection, denoising, error propagation and sensitivity analysis, presented, as far as it is known, in the form of a new graphical tool proposed by the authors named StatSSCandlePlot. The methodology was applied in a real case study of a gas fired boiler in which the indicator studied was its efficiency evaluated by the Stack Loss Method. From the StatSSCandlePlot it was possible to identify the trends of the indicator, the portion of each window in the steady-state, the values to be considered from the indicator and, in a complementary way, to identify the variable that most influences the variation of the indicator, through the sensitivity analysis.

Keywords: Key performance indicators; Candlestick chart; Process assessment.

INTRODUCTION

Asset management in industry is accomplished by measuring the performance of its processes. This measurement is quantified by the difference between the Key Performance Indicators (KPIs) and the pre-set reference values. KPIs serve as the basis for decision-making regarding product quality and service analysis, investment, process control, and implementation of improvements. However, some difficulties are encountered in the implementation of a measurement system, namely: availability, consistency, and reliability in obtaining the necessary information. The absence of these characteristics can lead to a system that is prone to failure and even unfeasible (Fischmann e Zilber, 1999; Müller, 2003; Nader et al., 2012).

Since KPIs play an extremely important role in the evaluation of a particular process, the data provided for their determination must be consistent so that the KPI is representative and the decisions taken on this calculated value are the closest to the optimum. Therefore, to calculate KPIs, the data need to be preprocessed, and it is also important to quantify its variations.

This article provides a systematic method for quantifying KPIs, including pre-treatment of data through noise removal and steady-state detection, statistical tools for error propagation and sensitivity analysis. In addition, a new graphic visualization tool based on candlestick graphics is proposed to facilitate the monitoring of KPIs.

The paper is divided into five sections, the first being this brief introduction. The next section includes

* Corresponding author: Cláudia D. M. Giaquinto - E-mail: claudiag@enq.ufrgs.br

background on techniques for noise suppression and steady-state detection, methods used for error propagation and sensitivity analysis, and the presentation of the candlestick graphing tool. The third section describes the methodology used. The fourth section presents a case study and in section 5 are the conclusions and suggestions for future work.

BACKGROUND

In process industries, the variables are subject to noise and errors inherent in the processes themselves and the measuring instruments. Therefore, it is extremely important to carry out the treatment of the data obtained to minimize the interference of these errors in the final values of the indicators. Among the existing data treatments, we highlight the removal of noise and steady-state detection.

Denoising and steady-state detection

In the real world, the existing signals have noise; only in ideal conditions, could it be neglected and its removal become unnecessary. Unfortunately, noise corrupts the signal and needs to be removed to proceed with data analysis and the extraction of useful information (Justen e Ramlau, 2009). In the literature, there is wide application of different techniques of noise removal in several areas, such as in the medical field with the purpose of improving the quality of ultrasound and electrocardiography data (Pratik et al.; Luo et al., 2006), geophysical area (To et al., 2009), automotive (Lee e Kim, 2015), images (Nasri e Nezamabadi-Pour, 2009) and materials (Staszewski, 2002).

With regard to methods for the removal of existing noise, there is the method of Maximum Noise Fraction (MNF) (Sun et al., 2009), the Fourier method (Justen e Ramlau, 2009; To et al., 2009) and Wavelet Shrinkage Denoising. The last is the most commonly used method, since it has presented good results, besides the possibility of allying the noise removal to the estimation of derivatives through Wavelet transform (Taswell, 2001; Luo et al., 2006; Nasri e Nezamabadi-Pour, 2009)

A wavelet is understood to be an irregular and asymmetric wave function capable of decomposing a series of data in the time and frequency domains, unlike Fourier analysis, in which the signal decomposed is only analyzed in the frequency domain. This transformation of a signal only to the frequency domain may neglect temporal information of the series, restricting Fourier analysis to series with stationary behaviors, linear problems or periodic functions (Rocha, 2008).

The Wavelet transform has two variants: Continuous (CWT) and Discrete Wavelet Transform (DWT), both of which have noise removal and differentiation

properties, and are applied for estimation of derivative calculations in several areas (Luo et al., 2006; Rocha, 2008). According to Luo et al. (2006), the DWT method has the advantage of presenting greater computational efficiency in signal noise removal and derivative estimation, when using the *à Trous* algorithm (Shensa, 1992; Wink e Roerdink, 2010).

The principle of resolution of the DWT method consists of the decomposition of the signal into low and high frequency components of the series, according to a base function (Wavelet family) and the number of decomposition levels. The high-frequency component of the series is analyzed at each level and, according to a threshold, the coefficients of this series that present values smaller than a critical value are eliminated, the most representative being kept, coming mostly from the signal and not from the noise. After the decomposition, the signal is reconstructed through the inverse transform, generating a signal with less noise when compared to the original (Luo et al., 2006; Rocha, 2008).

Eq. (1) is the relation used in the DWT method to estimate the derivative of a function $y(x)$ (Luo et al., 2006).

$$\frac{dy(x)}{dx} = \frac{y(\bar{x})^{dwt}}{K \cdot 2^{3N/2}} \quad (1)$$

where $y(x)^{dwt}$ is the discrete Wavelet transform of $y(x)$, which is calculated in relation to a Wavelet function; N is the level of decomposition of wavelets and K is a non-null parameter corresponding to the Fourier transform of a smoothing function. The factor 2^N corresponds to the dilation parameter, for which a low value results in high noise sensitivity and a high value results in a strong noise cancellation (Luo et al., 2006).

The step subsequent to noise removal is steady-state detection, which is critical for performance evaluation, process control and optimization, fault detection, and data reconciliation (Mejia et al., 2010; Korbel et al., 2014). Jiang et al. (2003) emphasize the importance of using data only when the steady-state is reached, mainly in the construction and applications of models, in order to guarantee significant results and models less prone to failure.

The method proposed by Cao and Rhinehart (1995) (CAO) is a consolidated method for the identification of the steady-state and consists in the comparison of the rate of the estimated noise variances, R , with a critical design reason, R_c , that is associated at a level of significance. If the mean of the process variable is varying, $R > R_c$, otherwise, if the mean is constant, $R \leq R_c$, the process is in steady-state. Since the adjustment of the parameters is not trivial, it can be done based on the tables and orientations reported in Cao and

Rhinehart (1997) as a function of the signal properties (Mejia et al., 2010; Korbel et al. 2014).

Another existing method is the one based on the estimate of the derivative (DER), this has its fundament in the concept of null or approximately null variation of the process variables. This proposal makes a comparative analysis of the stationarity index, denoted by $I_{diff,i}$ in Eq. (2), with a threshold value for identification of the steady-state, I_{ce} , both normalized (Mejia et al., 2010).

$$I_{diff,i} = 1 - \frac{\left| \hat{\dot{x}}_1 \right| + 10 \left| \hat{\ddot{x}}_1 \right|}{\max \left(\left| \hat{\dot{x}}_1 \right| + 10 \left| \hat{\ddot{x}}_1 \right| \right)} \quad (2)$$

Eq. (2) is based on the estimation of the first and second derivatives of the sequence after noise removal, denoted by $\hat{\dot{x}}_1$ and $\hat{\ddot{x}}_1$, respectively. Eq. (2) normalizes the sum $|\hat{\dot{x}}_1| + 10|\hat{\ddot{x}}_1|$ and returns its complementary value. After the determination of $I_{diff,i}$, its value is compared to I_{ce} and, in the case where $I_{diff,i} > I_{ce}$, the system is in the steady-state, otherwise the system is not stationary. Therefore, the greater the stringency, i.e., the higher the I_{ce} the smaller the portion that will be considered as stationary.

According to Mejia et al. (2010), after performing simulations for different methods of identifying the steady-state in different scenarios of noise and signal types, the authors concluded that the DER method has the best performance among the methods in all scenarios using two different ways of adjusting parameters, and stands out as a simple method.

The DER method, together with the DWT method, becomes a simple and efficient tool for detecting the steady-state, since the latter integrates the derivative estimation with the noise suppression of the signal, providing the necessary input data for the DER method.

Error Propagation

The process variables are subject to measurement errors and random errors, which cannot always be removed in their entirety; therefore, it becomes essential to trace their occurrences and perform the error propagation study. This study, also known as tolerance analysis, consists of a technique to examine the effect of partial tolerances of input variables on the output of the system (Cox, 1979).

Due to its high importance, the study of propagation of errors has been extensively applied in different areas such as nuclear reactors (Cox, 1979), mechanical assemblies (Chase e Parkinson, 1991; Glancy e Chase, 1999; Heijungs e Lenzen, 2014), geographic

information systems (Xue et al., 2015) and spatial explorations (Sun e Zhang, 2016).

There are several methods for error propagation in a system, namely interval analysis (Anderson e Mattson, 2012; Jiang et al., 2016), Response Surface Methodology (RSM), Fourier Amplitude Sensitivity Test (FAST) (Kuo e Uppuluri, 1983; Helton e Davis, 2003), the Monte Carlo method (MC) and Taylor series expansions (TS) (Xue et al., 2015, Sun and Zhang, 2016).

The MC method consists of one of the simplest and well-established methods for statistical analysis of tolerance through random numbers generated to simulate natural process variations. This method is flexible, accurately models non-linear effects, and does not require input and output distributions to be Gaussian. However, it requires a high computational cost, and every modification in the model and/or in the input values requires a new execution of the method, being prohibitive in iterative processes (Anderson and Mattson, 2012). Another disadvantage, according to Xue et al. (2015), is that this method does not have an analytical relationship between the uncertainties of inputs and outputs. Additional information about the Monte Carlo method can be found in Kuo and Uppuluri (1983), Helton (1993), Helton and Davis (2003), and Yan et al. (2015).

The use of the TS method for the propagation of errors is well consolidated in the literature and has as its main idea the determination of an analytical expression for the uncertainties of the outputs due to inputs, whose effects can be explained and analyzed through this relation (Xue et al., 2015).

Let y be a function of n variables as defined by Eq. (3). The first-order Taylor series approximation around the means of the input variables is given by Eq. (4), in which the partial derivatives are evaluated in the mean.

$$y = f(x_1, x_2, \dots, x_n) \quad (3)$$

$$y \approx y_1 = f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} (x_i - \bar{x}_i) \quad (4)$$

From Eq. (4), it is possible to determine the variance in y (σ_y^2) as a function of the variances in the input variables ($\sigma_{x_i}^2$), as shown in Eq. (5) (Anderson et al., 2012).

$$\sigma_y^2 \approx \sum_{i=1}^n \left[\left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 + 2 \sum_{j=i+1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \sigma_{x_i x_j}^2 \right] \quad (5)$$

The internal summation in Eq. (5) is related to the covariance between the input variables x_i and x_j .

($\sigma_{x_i x_j}^2$). This term is used when the input variables have interdependence, assuming a value equal to zero when the variables are independent.

Eq. (5) assumes that the input variables have a Normal distribution. According to Anderson et al. (2012), a correction must be inserted in Eq. (5) when these variables do not have a Gaussian distribution. In Eq. (6), we find the correction term and, in Eq. (7), the resulting relation to calculate the variance in y after the addition of the referred term.

$$N_G = \left(\frac{\partial f}{\partial x_i} \right)^2 (\mu_{1,i})^2 + 2 \sum_{j=i+1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} (\mu_{1,i}) (\mu_{1,j}) \quad (6)$$

$$\sigma_y^2 \approx \sum_{i=1}^n \left[\left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 - N_G + 2 \sum_{j=i+1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \sigma_{x_i x_j}^2 \right] \quad (7)$$

where $\mu_{1,i}$ and $\mu_{1,j}$ correspond to the first and the j -th central moment of the variable x_i , respectively.

The first order series approach offers, in general, significant results and ease in its application, being recommended for linear and polynomial functions. However, for non-linear functions the results may be less accurate, suggesting truncation of the series with higher order terms (Anderson and Mattson, 2012; Xue et al., 2015).

Aiming to increase the accuracy of the results, Xue et al. (2015) compared the variances obtained by the first, third and fifth order Taylor series expansions concerning the MC method. They noted that fifth-order expansion is the most accurate method. However, the difference is not very significant when compared to the third order expansion and presents a significant increase of the computational effort due to the number of partial derivatives that need to be calculated. Therefore, it is necessary to establish a balance of interests depending on the need of the application, because the increase of precision is accompanied by the increase of complexity, especially when the number of input variables is high. Anderson et al. (2012) show that the computational cost grows exponentially with the number of system inputs, besides the required calculations of moments of higher order and covariance between the variables, which can make the method prohibitive for complex systems.

Sensitivity Analysis

Indicators do not always have simple formulas and few parameters, so when encountering a large number of variables and complex equations, it is extremely important to know which parameters are the most sensitive, i.e., which input variables most influence the output variable. In this way, efforts to improve

the performance of a given process are concentrated and more rigorous in the most critical variables, contributing to the identification of the root cause of a possible problem, reducing expenses and focusing on continuous improvement.

The sensitivity analysis consists of a study of how the uncertainty of the model output is influenced by the uncertainties of the input of the model. Therefore, this study is complementary to the one of propagation of errors, which quantifies the uncertainty of the exit from the uncertainties of the input (Helton e Davis, 2003; Saltelli et al., 2008). This study has several applications, such as mechanical assemblies (Chase e Parkinson, 1991), food risk analysis (Christopher Frey e Patil, 2002) and biochemical and pharmacological systems (Weber e Thomas, 2005; Kent et al., 2013; Zhang et al., 2015).

In the case of mechanical assemblies, Chase and Parkinson (1991) state that the sensitivity analysis helps the designer to know which parameter variations have the most significant effect on the critical characteristics of the part, thus helping in the diagnosis of which component should focus on improving the design. From the knowledge of component sensitivity, the total variance can be reduced by reducing the variance in the most sensitive component and also reducing costs through the tolerance gap of the less sensitive component.

There are two strands of sensitivity analysis, namely: local and global. The local sensitivity analysis evaluates the changes in model output based on the individual variation of each input variable while maintaining the other variables of the domain constant, considering that the variables are independent. This approach should be used when the output is linearly related to the input parameters and has as a main limitation the fact that it does not evaluate the simultaneous changes in all parameters of the model, as well as the interactions between the parameters. In the global sensitivity analysis, all the parameters are varied simultaneously in the domain, allowing one to evaluate both the relative contribution of each parameter and of the interactions between parameters to the output model (Saltelli et al., 2008; Zhang et al., 2015).

Several methods for sensitivity analysis are available in the literature, among which we highlight the Morris method (Morris, 1991; Campolongo et al., 2007), the FAST method (Cukier et al., 1973; Saltelli et al., 1999; Zhang et al., 2001; Zhang et al., 2015), and Sobol (Sobol, 2001; Saltelli, 2002; Saltelli et al., 2010).

The Morris method is based on calculating a series of incremental rates, also called Elementary Effects, for each input variable. From the average value of these rates, the overall importance of each specific

entry is evaluated. This method is suitable when the number of variables is high and the model requires a high computational cost (Morris, 1991; Campolongo et al., 2007). Campolongo et al. (2007) emphasized the importance of this method and proposed changes in it to extend its use to systems with multiple outputs and to increase its efficiency through a new sampling strategy.

The FAST and Sobol methods are based on variance decomposition techniques to quantify the contributions of variables and of their interactions to the model. The main difference between the two methods is the algorithm used for multidimensional integration of the sensitivity indexes, which uses a sinusoidal function in the FAST method and a Monte Carlo integration in the Sobol method. Both methods do not require a linear and monotonous input-output relationship, are robust and exploit the full extent of the input set, but require high computational demand (Helton and Davis, 2003; Zhang et al., 2015).

Zhang et al. (2015) compared five commonly used methods for global sensitivity analysis, using different criteria, and concluded that the Sobol method is one of the most powerful methods, since it presents model independence, decomposition of the output variance and a high order interaction of parameters.

The first step of the Sobol method algorithm is to define the parameter limits, referring to the interval that includes all the values that the parameters can assume in the analysis. The execution of the method starts with the generation of sets of input variables using the low-discrepancy Sobol sequence, characterized by generating more uniform numbers (Zhang et al., 2015). The generated sets are then executed in the model for further determination of the Sobol sensitivity indexes, which provide quantitative information regarding the individual sensitivity of the parameters and also concerning the sensitivity due to the interactions between the parameters, as follows.

Let y be a function described in Eq. (3), the variance in y , $\text{Var}(y)$, can be decomposed as follows in Eq. (8) (Helton and Davis, 2003; Zhang et al., 2015).

$$\text{Var}(y) = \sum_{i=1}^n \text{Var}_i + \sum_{i<j} \text{Var}_{ij} + \sum_{i<j<k} \text{Var}_{ijk} + \dots + \text{Var}_{1,2,\dots,k} \quad (8)$$

where Var_i corresponds to the part of $\text{Var}(y)$ due only to x_i , Var_{ij} is the part of $\text{Var}(y)$ due to the interaction of x_i and x_j , and so on.

From Eq. (8), the sensitivity indices of Sobol, S , as presented in Eqs. (9), (10) and (11), can be determined (Helton and Davis, 2003; Zhang et al., 2015).

$$S_i = \frac{\text{Var}_i}{\text{Var}(y)} \quad (9)$$

$$S_{ij} = \frac{\text{Var}_{ij}}{\text{Var}(y)} \quad (10)$$

$$S_{Ti} = S_i + S_{ij_{i \neq j}} + \dots + S_{1\dots i\dots n} \quad (11)$$

where S_i provides the first-order contribution of x_i to the output variable, S_{ij} the second order contribution of the interaction between the parameters x_i and x_j for the output variable, and finally S_{Ti} is the total order sensitivity index, which quantifies the total contribution of x_i and the interactions of x_i with the other variables to the output of the model. The higher the value of these indices, the greater the influence on the output of the model.

Candlestick Plot

A very popular chart in the financial stock market among traders, especially in tracking currencies, is the candlestick chart. Due to the high price variation, it is important to know the best time to buy or sell stocks aiming at a high return on investment (Lee e Jo, 1999; Kurita, 2014). Fig. 1 shows an example of a candlestick for the dollar with fictitious values.

In Fig. 1, each candlestick represents one day and provides four informations on the values assumed by

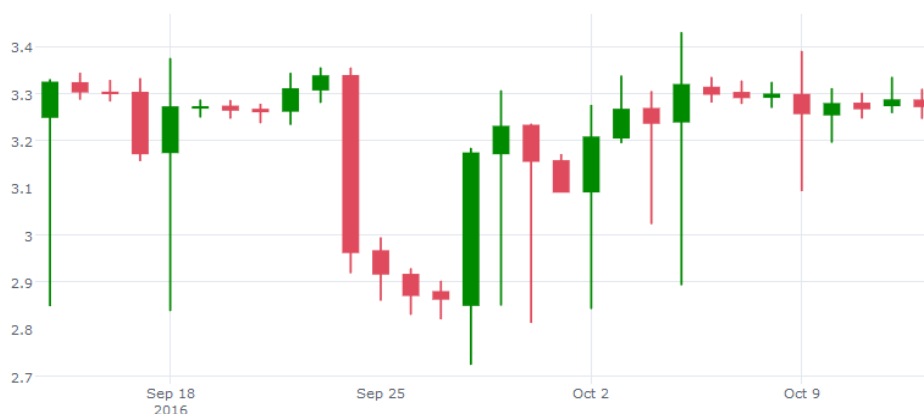


Figure 1. Dollar candlestick chart (fictitious values).

the dollar, namely: opening (O), high (H), low (L) and closing (C). The extreme values of the candlestick correspond to the minimum and maximum values. The body of the candlestick is given by the interval between the opening and closing values, and its color is related to the trend of the currency on that day, as shown in Fig. 2.

According to Fig. 2, if the dollar has a tendency to grow (closing price higher than the opening price), it is represented with the green color, otherwise, when the dollar closes the day in low (opening price lower than the closing price), the body is red. In addition, in this chart, there are two possible sequences for each candlestick: opening-maximum-minimum-closing or opening-minimum-maximum-closing.

From the candlestick chart, investors can detect patterns that can be used to predict short-

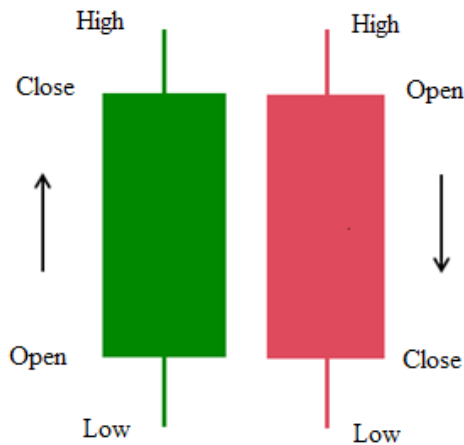


Figure 2. Candlestick interpretation.

term movements in market prices. The difference between the maximum and the minimum suggests the degree of trend and intraday volatility in price movements. Between the minimum and the closing, one has the direction in which the prices move until the closing time of the day. These relationships may contain information correlated with the central price movement from the present day to the next day, and may contain evidence of the trend between days (Kurita, 2014).

Marshall et al. (2006) developed independent and rigorous analyses of business strategies using candlestick charts to aid in the interpretation of results. This work contains a detailed description of simple (daily) candlesticks and patterns observed in an analysis containing 11-year data from the Dow Jones Industrial Index (DJIA), identifying continuation or reversal candlesticks, indicating, respectively, the predominance or the change of tendency, being able to show variations of bullish and bearish.

The use of the candlestick graphical tool is of paramount importance in helping to make decisions regarding the stock market. The application or extension of this graph to be used with other KPIs may provide additional information about their behavior, aiding in the diagnosis of problems.

METHODOLOGY

Next, in this chapter, the steps followed for the development of the methodology used in the present work will be detailed for quantification and monitoring

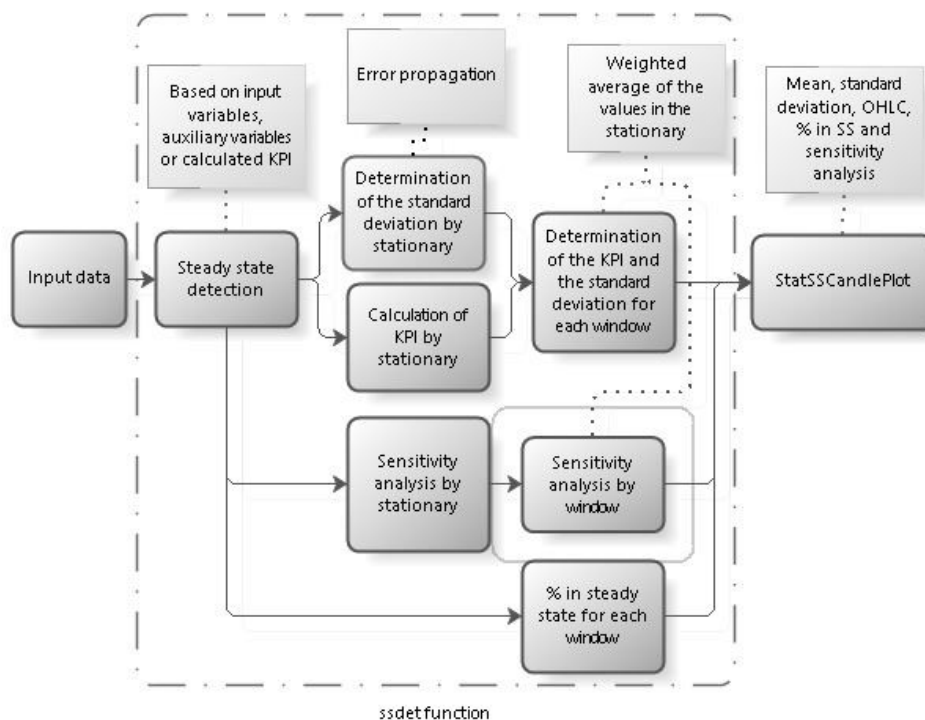


Figure 3. Flowchart representative of the methodology used in this work.

of KPIs. In Fig. 3, a flowchart representative of this methodology is given.

The raw input data, as can be seen in Fig. 3, were processed in the *ssdet* function, in which they were initially divided into windows and subjected to steady-state detection. After determining the intervals that are in the steady-state in each window, the percentage of each window in the stationary was determined, as well as the KPI value, the standard deviation through the propagation of errors and the sensitivity indexes for each stationary. To determine the values per window, we calculated the weighted mean values in the KPI stationary, standard deviation and sensitivity indexes. As shown in Fig. 3, these data served as the basis for the construction of the StatSSCandlePlot, the graphical KPI monitoring tool proposed in this study.

The steps outlined briefly above are detailed below.

Ssdet function

The *ssdet* function was programmed in Python 3 and its input parameters are presented in Eq. (12).

$$\text{ssdet} = f(n_{\text{var}}, \text{data}, w, x_{\text{est}}, I_{\text{ee}}, N, \text{wavfunc}) \quad (12)$$

where n_{var} is the number of input variables, *data* is a DataFrame composed of the values of input variables, auxiliary variables and the calculated KPI at each point, *w* is the window size, x_{est} is the criterion used for the detection of steady-state, I_{ee} is the design stationary threshold (discussed in Section 2.1), *N* is the decomposition level of Wavelets, and *wavfunc* is the Wavelet function chosen.

The first step of the *ssdet* function is the division of the data into windows, according to the number *w* reported by the operator. After this division, the data are subjected to steady-state detection, as described below.

Discrete Wavelet Transform (DWT) method and steady-state detection

The method chosen for the steady-state detection was the DER method with the derivatives estimated by the DWT method, which is also responsible for the removal of noise from the data. Initially, the smoothed derivatives were estimated by Eq. (1) and then Eq. (2) was used to determine the stationarity index, $I_{\text{diff},j}$, which was later compared to I_{ee} to verify the state of the window in question. Four input parameters of the *ssdet* function are directly related to these methods, namely: x_{est} , I_{ee} , *N* and *wavfunc*.

The x_{est} parameter of the *ssdet* function provides the stationary detection criterion, i.e. the operator can choose one or more variables to be the criterion, which can be composed of input variables, auxiliary variables or the KPI itself. The auxiliary variables

are those that are not part of the KPI expression, but can be extremely important for the process. In case of choosing only one variable, the detection is based on this and only the portions where it is in the stationary state are considered for subsequent steps. In the case of more than one variable, only the stretches in which all the variables have in common the stationary state are considered for the later analyses.

The I_{ee} is a value between 0 and 1 provided by the operator and is related directly to the accuracy of steady-state detection. In the case of parameter *N*, a decomposition at a very high level can result in an opposite effect, canceling loudly the noises and de-characterizing the signal, so it is necessary to adjust it with caution. In case of $x_{\text{est}} > 1$, values of I_{ee} and *N* corresponding to each variable can be chosen.

Finally, the parameter *wavfunc* determines which Wavelet function is to be used in the estimation of derivatives and noise removal. In this case, the operator can choose between two functions, namely: Haar or quadratic spline. In case of $x_{\text{est}} > 1$, different functions can also be chosen for each variable.

From these parameters, the method was executed and for each window the stretches in steady-state were determined. Then, the percentages of the windows in steady-state (% SS) were calculated, providing a notion of stability of the data in each window. The expression used for % SS is given in Eq. (13).

$$\%SS = \frac{\text{n}^\circ \text{ of data in SS}}{\text{n}^\circ \text{ of data in window}} \cdot 100 \quad (13)$$

Assuming the intervals in the EE to be considered for each window, the study of the propagation of errors for the determination of KPI and standard deviation of stationary states and the window was performed.

Error Propagation using a First-Order Taylor Series

For the study of the propagation of errors, the first-order Taylor series expansion method with the correction of non-normality and the covariance calculation, as presented above, was used to determine the standard deviation per stationary state. Considering a KPI with *n* input variables whose window *i* has *k* stationaries, the sequence of calculations used until the determination of the standard deviation of the window *i* was determined as shown in Fig. 4.

In Fig. 4, from the input variables corresponding to the stationary and the KPI equation for the determination of the derivatives, Eq. (7) was used to calculate the standard deviation of each stationary, $\sigma_{st,j}$ ($j \in [0, k]$). The variance in the window *i*, $\sigma_{win,i}^2$, was determined from the weighted average of variances per stationary, whose weight, $p_{est,j}$, was determined according to Eq. (14).

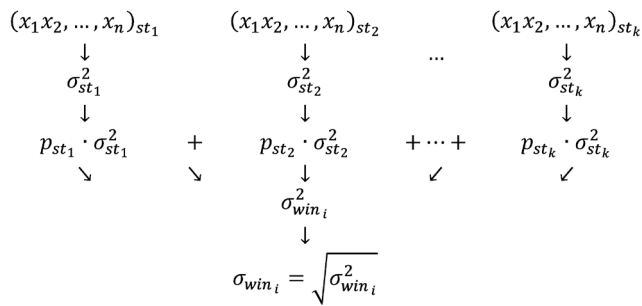


Figure 4. Scheme used to determine the standard deviation of the window i .

$$p_{est_j} = \frac{t_{est_j}}{\sum_{j=1}^k t_{est_j}} \quad (14)$$

where: t_{est_j} corresponds to the time period of the stationary j .

To calculate the average KPI of the window, the KPI was initially determined for each stationary j , according to Eq. (15).

$$KPI_{est_j} = f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)_{est_j} \quad (15)$$

From the values obtained from Eq. (15) for all stationary states, the scheme shown in Fig. 4 and Eq. (14) were used to determine the average indicator of the window i , KPI_{win_i} .

Sensitivity Analysis

The Sobol method, described above, was used to perform the sensitivity analysis. This method is available in the Python Sensitivity Analysis Library (SALib) (Herman e Usher, 2017) and has been executed according to the following steps:

(1) Definition of the problem: creation of a dictionary containing the number of KPI input variables, their respective notations and their lower and upper limits;

(2) Generation of the sample inputs from the model: the KPI input variables were generated through the Saltelli extension to the Sobol sequence, saltelli.sample in Python, which has as inputs the defined problem, the number of samples to be generated, N_A , and a boolean variable for the calculation of second order indexes (Eq. (10)). For a KPI with D variables, this function generates $N_A \cdot (D + 2)$ rows, if the second order indexes are not computed, and $N_A \cdot (2D + 2)$ rows if the second order indexes are calculated. According to Zhang et al. (2015), for a low complexity model, $N_A = 1,000$ samples is sufficient; for more complex models with about 20 parameters, at least $N_A = 100,000$ samples;

(3) KPI calculation: KPI was calculated at each point in the matrix resulting from the saltelli.sample function;

(4) Calculation of the sensitivity indexes per stationary: the sobol.analyze function was used to carry out the sensitivity analysis. From the defined problem and calculated KPI values, this function returns the sensitivity indexes and their respective trusts;

(5) Calculation of the sensitivity indexes per window: from the total order sensitivity indexes calculated in step (4) for each stationary, S_{Tstj} , its weighted average was calculated to determine the indexes of each variable per window, S_{Txwinj} , similar to Fig. 4.

StatSSCandlePlot

StatSSCandlePlot is a new approach to KPI tracking, integrating data from statistical analysis, steady-state detection and the candlestick graph. This graphical tool provides guidelines on the trend of the indicator through the candlestick pattern, presents consistent values of KPI using error propagation statistics coupled with the study of the stability of the indicator by the detection of steady-state, in addition to providing an initial diagnosis of the root cause of the variation of the indicator through the sensitivity analysis; all the presented interactively with Plotly by Python (Plotly Technologies Inc, 2015).

StatSSCandlePlot contains the following information:

(1) Candlestick: information about all values assumed by the KPI in the window, highlighting the minimum and maximum, as well as the opening and closing values;

(2) Steady-state percentage (scatter plot): indicative of the dynamicity of the indicator in each window;

(3) Mean and standard deviation (line graph): based on the propagation of errors, indicate the reliable values of the KPI and their respective deviation, corresponding to three deviations from the mean;

(4) Global sensitivity indexes (bar graph): Root cause analysis of KPI variation indicating the influence of the variables in a quantitative way.

CASE STUDY

After the development of the methodology for calculating and monitoring KPIs, it was applied in a case study. The studied system consists of a gas fired steam generator, whose general scheme is shown in Fig. 5.

In Fig. 5, the gas is injected through the burners into the furnace, where the combustion occurs. In the furnace, the gases come in contact with the tubes of the evaporating section, where the heat is transferred to the fluid inside the tubes. The preheated water enters

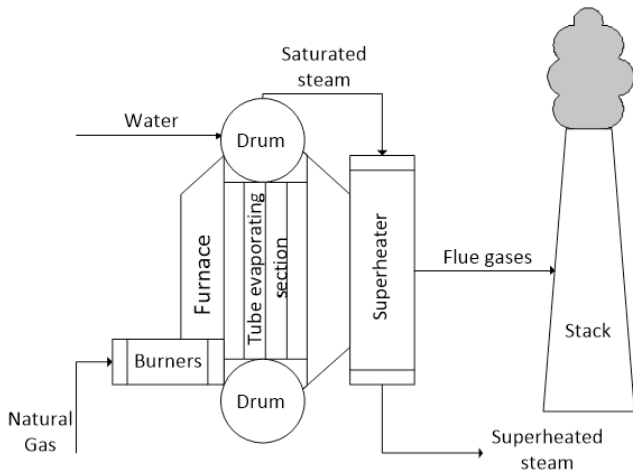


Figure 5. Representative gas-fired steam generator scheme.

the upper drum and descends through the tube bundle to the lower drum. The mixture of water and steam returns to the upper drum through the tube walls of the furnace. The saturated steam from the upper drum enters into a superheater where it will be reheated until it reaches the superheated steam temperature. The flue gases from the furnace are discharged into the atmosphere through the stack.

Industrially, it is important to know if the boiler is well operated, i.e., operating efficiently, since a minimum consumption of gas is desired for a greater amount of steam generated. For this, a boiler efficiency (BE) KPI can be used. One of the methods of calculating boiler efficiency is the Stack Loss Method, which estimates the BE from the losses in the stack according to PTC 4.1 (American Society of Mechanical Engineers, 1964; Asme, 1964), as shown in Eq. (16).

$$BE (\%) = 100\% - L_{DG} - L_{WG} - 1\% \quad (16)$$

In Eq. (16), L_{DG} and L_{WG} consider the dry and wet losses in the stack, respectively, and 1% is relative to other losses, such as radiation and surface convection.

The California Air Resources Board (2011) considered a natural gas composed of 95% of methane, 2% of ethane, 1% of propane, and 2% of nitrogen, with a density of 0.0445 lb/ft³ and a high heat value of 22,983.189 BTU/lb, and determined the formulae of L_{DG} and L_{WG} , as shown from Eq. (17) to Eq. (19).

$$DG = 14.7365 \cdot \left(\frac{O_{2\%}}{21\% - O_{2\%}} \right) + 15.371 \quad (17)$$

$$L_{DG} (\%) = 0.001044 \cdot DG \cdot (T_{stack} - 70) \quad (18)$$

$$L_{WG} (\%) = 9.482 \cdot 0.004351 \cdot T_{stack} \quad (19)$$

where: DG is the dry gas per volume of fuel, $O_{2\%}$ is the volumetric fraction of O_2 in the flue gases, and T_{stack} is the temperature in the stack [°F]. Additional details of boiler efficiency can be found in the PTC 4 standard (Asme, 2008) and in Natural Resources Canada (2018).

As can be observed from Eq. (16) to Eq. (19), the BE obtained from the Stack Loss Method depends only on the percentage of O_2 in the flue gases, and the stack temperature. So, this KPI has two independent variables.

Data for these two variables were collected every minute from a boiler operating in a Brazilian refinery during 26 consecutive days, and can be seen in Figures 6 and 7. The Boiler Efficiency was calculated, and the result can be seen in Figure 8.

According to Fig. 8, the BE was calculated every minute during the studied period. However, some

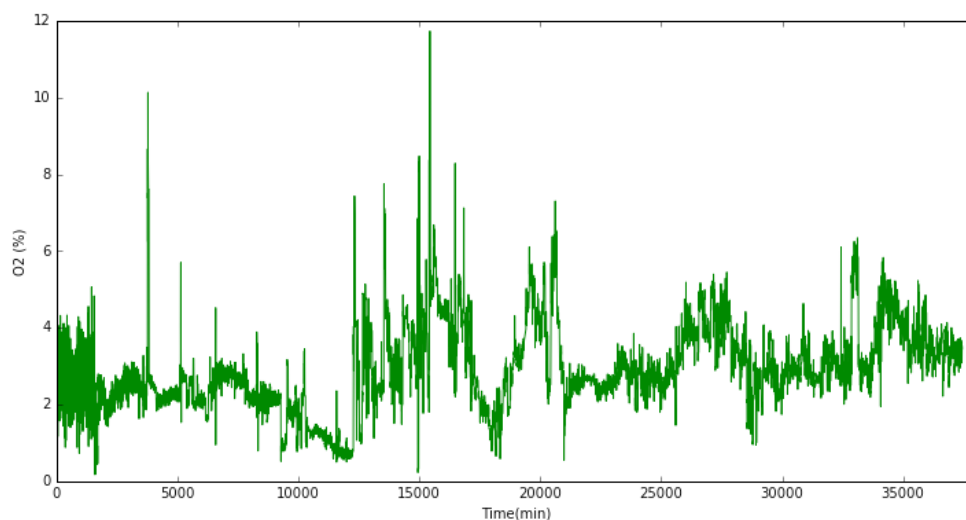


Figure 6. Volumetric fraction of O_2 in flue gases.

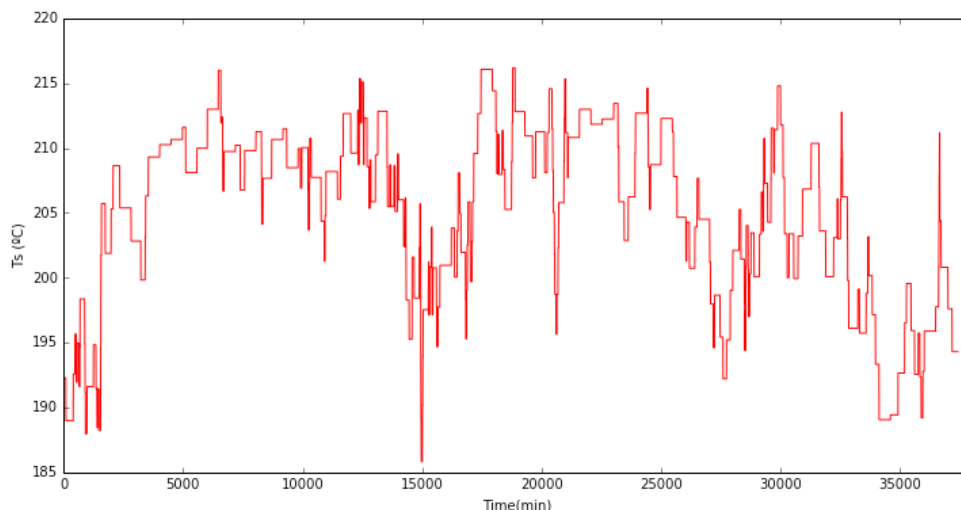


Figure 7. The temperature of flue gases in Stack.

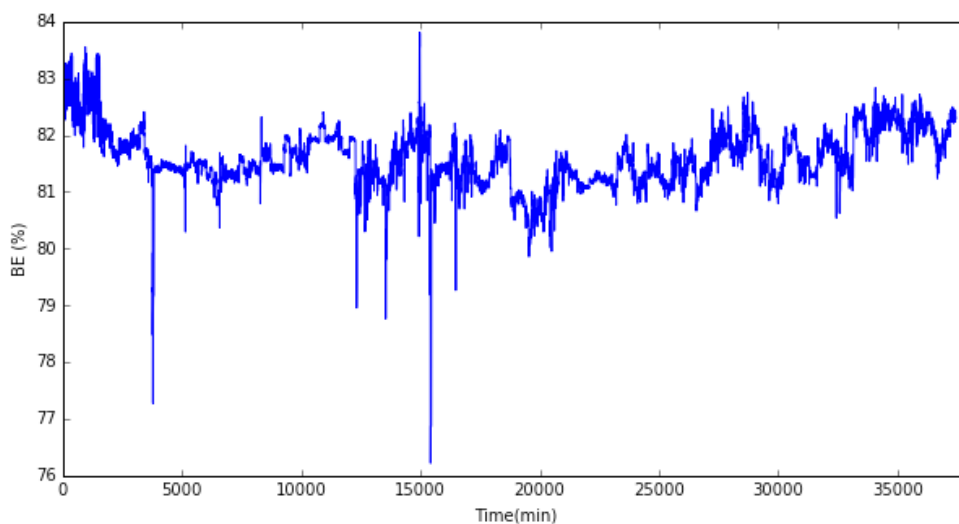


Figure 8. Calculated Boiler Efficiency according to eq. (16).

questions arise regarding the determination and visualization of this indicator:

(i) in the presence of dynamic data, how to determine a static KPI?

(ii) what does it take for the indicator to become a representative value rather than just numbers?

(iii) assuming that the manager wants to know the KPI from the previous month, how can it be represented: a monthly average? a table with daily values? a chart with time series like the one in Fig. 7?

(iv) going further: what is the most influential variable in the KPI variation?

Only looking at Figs. 6 to 8, it is not possible to infer precisely which variable is more responsible for the BE variation. Facing this problem and aiming to extract valuable information from these data, the methodology of quantifying and monitoring KPIs proposed in this paper it was applied.

The parameters of the steady-state detection were manually adjusted, and the following values were used

for the following case study, x_{est} being the KPI, that is, only the part where the KPI is in steady-state will be considered for the window, $I_{\text{ce}} = 0.7$, $N = 4$, and the quadratic spline as wavelet function. In the sensitivity analysis method, $N_A = 1000$, due to the low number of input variables. Fig. 9 shows the StatSSCandlePlot for the BE.

Firstly, in Fig. 9, it can be observed that the data were grouped in daily windows without losing information about their behavior in this period by using the candlestick plot, which indicates the tendency of increasing or decreasing of the indicator. Since data were collected every minute, it would be confusing and meaningless to interpret the whole time series, because looking at a one minute change does not bring substantial information of a KPI evaluated in a period of hours and days, so, it is important to cluster data in windows.

To determine mean and confidence intervals, a data pre-processing was made, as the process data



Figure 9. StatSSCandlePlot for the boiler efficiency.

was dynamic and a static KPI was applied. So, only intervals where the BE was in the steady-state were considered in each window for calculating these variables, and its percentages are also displayed in the chart as SS (values in the secondary axis) to have an idea of stationarity in each window. The mean and the standard deviation were determined using error propagation and by weighting the stationaries that occurred in each window aiming to return the most representative BE value for each window and its respective uncertainty.

Additionally, in Fig. 9, there is a reference line which is the minimum accepted value for the BE inside the company (which may also be a limit established by law depending on the KPI, for example, the maximum allowed concentration of sulfur in diesel). The means of all windows are above this limit, but there is still margin to improve the boiler efficiency, which can reach values close to 90% (ASME, 2008).

Looking more closely at the candlesticks in Fig. 9, it can be observed that in the windows 3 and 9, the closing value of BE was lower than the opening value. However, the closure was considerably higher than the minimum, indicating a recovery trend. The minimum value of these windows appeared not to be common in these periods since they do not move the average down from the reference. A similar interpretation can be made for window 11, but in this case, the BE reached a maximum noticeably higher than the

opening value, showing a recovery with the closing much higher than the minimum value. This wide range of values assumed in the window was responsible for a longer confidence interval in the graph. Other bullish expressive candlesticks are the windows 15 and 24, where values closed about 1% above the opening value. Window 17 shows a bearish candlestick where the closing value was slightly above the opening value, but the BE reached a maximum considerable bigger than the closing, indicating that the BE increase was not sustained during that day.

Still in Fig. 10, windows 3, 9 and 12 are the ones that draw the most attention and the first question asked is about which is the input variable responsible for these significant variations in the process. To answer this question, the result of the sensitivity analysis is presented in Fig. 10.

In Fig. 10, it is observed that, on all days the stack temperature is the variable with the highest total sensitivity index. Therefore, the observed variation of the BE is mainly due to stack temperature fluctuation, with the $O_2\%$ being more significant in some windows (3, 4, 9, and 11, for example). Given this information, the focus for the improvement of the indicator is the T_{stack} . This initial diagnostic gives a first look to the problem of the BE variation, giving a clue of the root cause, that might be a problem related to the stack temperature.

The steady-state percentage (SS in Fig. 9) is valuable information to know how much of the window

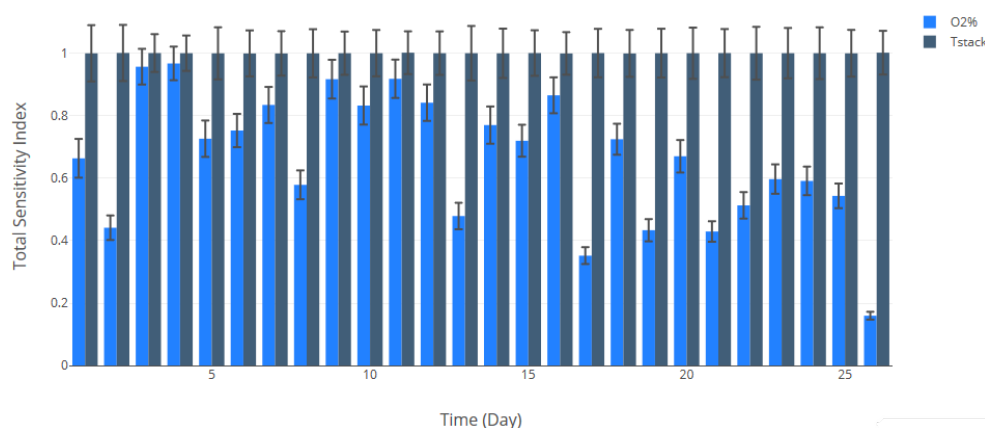


Figure 10. BE sensitivity analysis.

is considered for the KPI calculation, the standard deviation, and the sensitivity indexes. Evaluating the windows 3 and 17, which have the highest and the lowest fraction in the steady-state, we obtain Fig. 11.

In Fig. 11, the state was multiplied by 75 for better visualization only, assuming a zero value when the SS is not identified and 75 when it is in SS. The window 3 has 96.7 % of the data in SS, and the window 17 has 70.3%. With the threshold of 0.7, the method identified the change in the indicator and characterized this transition as non-stationary in window 3. In window 17, it is observed that the values of the BE fluctuate around an average value, without abrupt changes of the values. In this case, the SS detection is more sensitive, reducing the portion of the window in steady-state. Increasing this threshold to 0.9 and decreasing to 0.5 for these same windows results in Fig. 12 and Fig. 13, respectively.

When Fig. 12 is compared to the results in Fig. 11, it can be seen that the method became more sensitive and rigorous, decreasing even more the parcel of the window in SS; the opposite is observed in Fig. 13, where the SS percentage increased. Table 1 lists the values of SS (%) for the windows 3 and 17, varying the I_{cc} .

Table 1. Percentage of the window in SS according to I_{cc} for windows 3 and 17.

| I_{cc} | SS (%) | |
|----------|-----------|-----------|
| | Windows 3 | Window 17 |
| 0.5 | 97.7 | 90.0 |
| 0.6 | 97.4 | 82.6 |
| 0.7 | 96.7 | 70.3 |
| 0.8 | 96.0 | 57.6 |
| 0.9 | 93.1 | 38.5 |

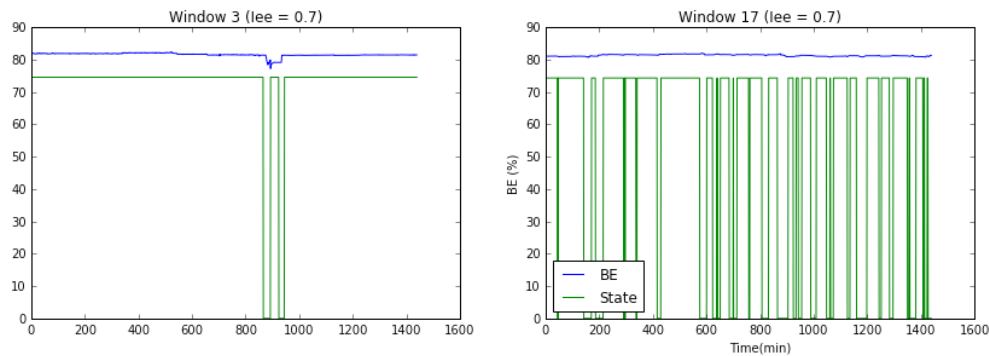


Figure 11. Steady-state ID graphs for BE with $I_{cc} = 0.7$.

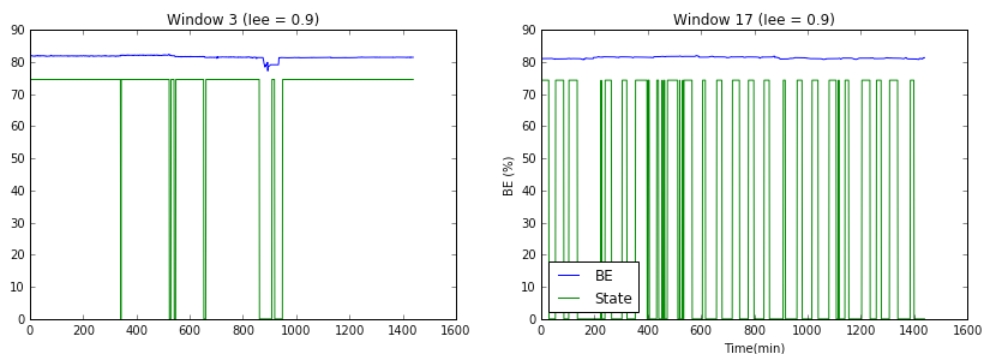


Figure 12. Steady-state ID graphs for BE with $I_{cc} = 0.9$.

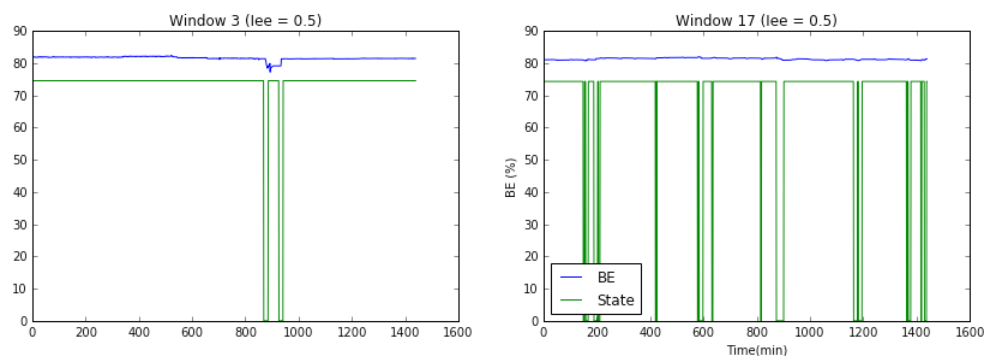


Figure 13. Steady-state ID graphs for BE with $I_{cc} = 0.5$.

As can be seen in Table 1, the window 17 is penalized severely by the threshold increase. Therefore, since the data can present different behaviors in different windows, the best choice is not always the most rigorous. Based on this analysis and the behaviour of the data, the intermediate I_{ee} value of 0.7 was evaluated as suitable to represent the windows, avoiding that some windows would not have a representative value by increasing or decreasing the threshold considerably, resulting in an excessive or missing rigorousness.

CONCLUSIONS

A case study of a gas fired boiler was made, in which the KPI studied was the boiler efficiency, determined by the Stack Loss Method that uses only the stack temperature and the percentage of oxygen in the flue gases to estimate the boiler efficiency. In possession of data collected from a real process in a Brazilian refinery, results of this KPI were compared considering a simple time series graph without data preprocessing and the StatSSCandlePlot. The StatSSCandlePlot, using the division of data in windows of the candlestick standard, of the mean, standard deviation, and percentages of the steady-state window, provided information on indicator trends in each window, mean values to be considered for window evaluation, and the dynamicity of the windows. In addition, the graph complementary to the StatSSCandlePlot with the sensitivity analysis provided an initial diagnosis regarding the main cause of variation of the indicator, which in this case is the stack temperature.

StatSSCandlePlot is a powerful tool for quantifying and monitoring KPIs that includes a data preprocessing step, statistical analysis, data visualization tool and root cause diagnosis which, coupled with correct interpretations, positively assists in decisions about process changes.

NOMENCLATURE

| | |
|---------------|---|
| DG | dry gas per volume of fuel |
| $I_{diff,i}$ | stationarity index |
| I_{ee} | threshold value for identification of steady-state |
| K | non-null parameter corresponding to the Fourier transform of a smoothing function |
| L_{DG} | dry losses in the stack |
| L_{WG} | wet losses in the stack |
| N | level of decomposition of wavelets |
| N_A | number of samples to be generated |
| N_G | correction term for a non-gaussian distribution |
| n_{var} , D | number of input variables |

| | |
|---------------------|--|
| $O_{2\%}$ | volumetric fraction of oxygen in the flue gases |
| p_{stj} | weight of the stationary j |
| R | rate of the estimated noise variances |
| R_C | critical design reason |
| S_i | first-order sensitivity index |
| S_{ij} | second-order sensitivity index |
| S_{Ti} | total order sensitivity index |
| T_{stack} | temperature in the stack (in °F) |
| t_{stj} | time period of the stationary j |
| $Var_{1,2,\dots,k}$ | part of $Var(y)$ due to the interaction of x_1, x_2, \dots , and x_k |
| \hat{x}_i | first derivative after denoising |
| \ddot{x}_i | second derivatives after denoising |
| \bar{x}_i | mean of a variable |
| $y(\bar{x})^{dwt}$ | discrete Wavelet transform of $y(x)$ |
| x_{est} | criterion used for steady-state detection |

Greek letter

| | |
|----------------------|-------------------------------|
| $\sigma_{x_i}^2$ | variance in an input variable |
| $\sigma_{x_i x_i}^2$ | covariance between the inputs |
| σ_y^2 | variance in a function y |

ACKNOWLEDGMENTS

The authors would like to thank the financial support by the National Petroleum, Natural gas and Biofuels Agency (ANP) and Petróleo Brasileiro S.A. – PETROBRAS.

REFERENCES

- American Society of Mechanical Engineers. ASME PTC 4.1: Steam Generating Units 1964.
- Anderson, T. V., and Mattson, C. A., Propagating Skewness and Kurtosis Through Engineering Models for Low-Cost, Meaningful, Nondeterministic Design. *Journal of Mechanical Design*, 134(10), 100911-100911 (2012). Available on: < <http://dx.doi.org/10.1115/1.4007389> >.
- Anderson, T. V., Mattson, C. A., Larson, B. J. and Fullwood, D. T., Efficient Propagation of Error Through System Models for Functions Common in Engineering. *Journal of Mechanical Design*, 134(1), 014501-014501-6 (2012). Available on: < <http://dx.doi.org/10.1115/1.4005444> >.
- ASME. Steam Generating Units: The American Society of Mechanical Engineers. ASME PTC 4.1 (1964).
- ASME. Fired Steam Generators: Performance Test Codes. New York: The American Society of Mechanical Engineers. ASME PTC 4, p. 294 (2008).
- California Air Resources Board. ASME PTC 4: Indirect Method: Stack Loss Method. (2011). Available on: < https://www.arb.ca.gov/cc/capandtrade/allowanceallocation/boiler_efficiency_calc.pdf >.

- Campolongo, F., Cariboni, J. and Saltelli, A., An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10), 1509-1518 (2007). Available on: < <http://www.sciencedirect.com/science/article/pii/S1364815206002805> >.
- Cao, S. and Rhinehart, R. R., An efficient method for on-line identification of steady-state. *Journal of Process Control*, 5(6), 363-374 (1995). Available on: < <http://www.sciencedirect.com/science/article/pii/095915249500009F> >.
- Cao, S. and Rhinehart, R. R., Critical values for a steady-state identifier. *Journal of Process Control*, 7(2), 149-152 (1997). Available on: < <http://www.sciencedirect.com/science/article/pii/S0959152496000261> >.
- Chase, K. W. and Parkinson, A. R., A survey of research in the application of tolerance analysis to the design of mechanical assemblies. *Research in Engineering Design*, 3(1), 23-37 (1991). Available on: < <http://dx.doi.org/10.1007/BF01580066> >.
- Christopher Frey, H. and Patil, S. R., Identification and Review of Sensitivity Analysis Methods. *Risk Analysis*, 22(3), 553-578 (2002). Available on: < <http://dx.doi.org/10.1111/0272-4332.00039> >.
- Cox, N. D., Tolerance analysis by computer. *Journal of Quality Technology*, 11(2), 80-87 (1979). Available on: < <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0018455687&partnerID=40&md5=0d1b953fe4f03037666fc8d896fd139c> >.
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. and Schaibly, J. H., Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8), 3873-3878 (1973). Available on: < <http://dx.doi.org/10.1063/1.1680571> >.
- Fischmann, A. A. and Zilber, M. A., Utilização de Indicadores de Desempenho como Instrumento de Suporte à Gestão Estratégica. *Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração*. Rio de Janeiro (1999).
- Glancy, C. G. and Chase, K. W., A Second-Order Method for Assembly Tolerance Analysis. *ASME Design Engineering Technical Conferences*. Las Vegas, 1999.
- Heijungs, R. and Lenzen, M., Error propagation methods for LCA—a comparison. *The International Journal of Life Cycle Assessment*, 19(7), 1445-1461 (2014). Available on: < <http://dx.doi.org/10.1007/s11367-014-0751-0> >.
- Helton, J. C. and Davis, F. J., Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1), 23-69 (2003). Available on: < <http://www.sciencedirect.com/science/article/pii/S0951832003000589> >.
- Herman, J. and Usher, W., SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software*, 2(9) (2017).
- Jiang, C., Fu, C. M., Ni, B. Y. and Han, X. Interval arithmetic operations for uncertainty analysis with correlated interval variables. *Acta Mechanica Sinica*, 32(4), 743-752 (2016). Available on: < <http://dx.doi.org/10.1007/s10409-015-0525-3> >.
- Jiang, T., Chen, B., He, X. and Stuart, P., Application of steady-state detection method based on wavelet transform. *Computers & Chemical Engineering*, 27(4), 569-578 (2003). Available on: < <http://www.sciencedirect.com/science/article/pii/S0098135402002351> >.
- Justen, L. and Ramjau, R., A general framework for soft-shrinkage with applications to blind deconvolution and wavelet denoising. *Applied and Computational Harmonic Analysis*, 26(1), 43-63 (2009). Available on: < <http://www.sciencedirect.com/science/article/pii/S1063520308000183> >.
- Kent, E., Neumann, S., Kummer, U. and Mendes, P., What Can We Learn from Global Sensitivity Analysis of Biochemical Systems? *PLOS ONE*, 8(11), 79244 (2013). Available on: < <https://doi.org/10.1371/journal.pone.0079244> >.
- Korbel, M., Bellec, S., Jiang, T. and Stuart, P., Steady-state identification for on-line data reconciliation based on wavelet transform and filtering. *Computers & Chemical Engineering*, 63, 206-218 (2014). Available on: < <http://www.sciencedirect.com/science/article/pii/S0098135414000350> >.
- Kuo, W. and Uppuluri, V. R. R., A review of error propagation analysis in systems. *Microelectronics Reliability*, 23(2), 235-248 (1983). Available on: < <http://www.sciencedirect.com/science/article/pii/0026271483903311> >.
- Kurita, T., Dynamic characteristics of the daily yen-dollar exchange rate. *Research in International Business and Finance*, 30, 72-82 (2014). Available on: < <http://www.sciencedirect.com/science/article/pii/S0275531913000263> >.
- Lee, K. H. and Jo, G. S., Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, 16(4), 357-364 (1999). Available on: < <http://www.sciencedirect.com/science/article/pii/S0957417499000111> >.
- Lee, S. and Kim, J., Discrete wavelet transform-based denoising technique for advanced state-of-charge estimator of a lithium-ion battery in electric vehicles. *Energy*, 83, 462-473 (2015). Available on: < <http://www.sciencedirect.com/science/article/pii/S0360544215002017> >.
- Luo, J., Bai, J. and Shao, J., Application of the wavelet transform on axial strain calculation in ultrasound elastography. *Progress in Natural Science*. China: China Academic Journal Electronic Publishing House (2006).

- Marshall, B. R., Young, M. R. and Rose, L. C. Candlestick technical trading strategies: Can they create value for investors? *Journal of Banking & Finance*, 30(8), 2303-2323 (2006). Available on: < <http://www.sciencedirect.com/science/article/pii/S0378426605002116> >.
- Mejia, R. I. G., Duarte, M. B. and Trierweiler, J. O. Avaliação do Desempenho e Ajuste Automático de Métodos de Identificação de Estado Estacionário. XVIII Congresso Brasileiro de Engenharia Química. Foz do Iguaçu (2010).
- Morris, M. D., Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161-174 (1991). Available on: < <http://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804> >.
- Müller, C. J., Modelo de Gestão Integrando Planejamento Estratégico, Sistemas de Avaliação de Desempenho e Gerenciamento de Processos (MEIO - Modelo de Estratégia, Indicadores e Operações). Ph.D. Thesis, Universidade Federal do Rio Grande do Sul (2003).
- Nader, B., Tomi, G. D. and Passos, A. O. Indicadores-chave de desempenho e a gestão integrada da mineração. *Rem: Revista Escola de Minas*, 65, 537-542 (2012). Available on: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0370-44672012000400015&nrm=iso >.
- Nasri, M. and Nezamabadi-Pour, H. Image denoising in the wavelet domain using a new adaptive thresholding function. *Neurocomputing*, 72(4), 1012-1025 (2009). Available on: < <http://www.sciencedirect.com/science/article/pii/S0925231208002488> >.
- Natural Resources Canada. Stack Losses for Natural Gas. Canada, (2018). Available on: < <http://www.nrcan.gc.ca/energy/efficiency/industry/technical-info/tools/boilers/5437> >.
- Plotly Technologies Inc. Collaborative data science. Montréal, QC (2015). Available on: < <https://plot.ly> >.
- Pratik, Pradhan, G. and Shahnawazuddin, S., Denoising of ECG signal by non-local estimation of approximation coefficients in DWT. *Biocybernetics and Biomedical Engineering*. (2017). Available on: < <http://www.sciencedirect.com/science/article/pii/S0208521617301122> >.
- Rocha, V. B. D., Uma abordagem de Wavelets aplicada à combinação de previsões: uma análise teórica e experimental. 155 (2008). Universidade Federal do Paraná, Curitiba.
- Saltelli, A., Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2), 280-297 (2002). Available on: < <http://www.sciencedirect.com/science/article/pii/S0010465502002801> >.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S., Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259-270 (2010). Available on: < <http://www.sciencedirect.com/science/article/pii/S0010465509003087> >.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., Front Matter. In: (Ed.). *Global Sensitivity Analysis. The Primer*: John Wiley & Sons, Ltd (2008).
- Saltelli, A., Tarantola, S. and Chan, K. P. S. A., Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1), 39-56 (1999). Available on: < <http://www.jstor.org/stable/1270993> >.
- Shensa, M. J., The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, 40(10), 2464-2482 (1992).
- Sobol, I. M., Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1), 271-280 (2001). Available on: < <http://www.sciencedirect.com/science/article/pii/S0378475400002706> >.
- Staszewski, W. J., Intelligent signal processing for damage detection in composite materials. *Composites Science and Technology*, 62(7), 941-950 (2002). Available on: < <http://www.sciencedirect.com/science/article/pii/S0266353802000088> >.
- Sun, L., Rieger, J., Hinrichs, H., Maximum noise fraction (MNF) transformation to remove ballistocardiographic artifacts in EEG signals recorded during fMRI scanning. *NeuroImage*, 46(1), 144-153 (2009). Available on: < <http://www.sciencedirect.com/science/article/pii/S1053811909000937> >.
- Sun, Y., Zhang, J., Error propagation characteristic analysis of halo orbit based on condition number of state transition matrix. *Astrophysics and Space Science*, 361(9), 297 (2016). Available on: < <http://dx.doi.org/10.1007/s10509-016-2856-3> >.
- Taswell, C., Experiments in Wavelet Shrinkage Denoising. *Journal of Computational Methods in Science and Engineering*. 1-12 (2001).
- To, A. C., Moore, J. R. and Glaser, S. D. Wavelet denoising techniques with applications to experimental geophysical data. *Signal Processing*, 89(2), 144-160 (2009). Available on: < <http://www.sciencedirect.com/science/article/pii/S0165168408002442> >.
- Weber, A. and Thomas, R. *Key Performance Indicators: Measuring and Managing the Maintenance Function*. Ontario (2005).

- Wink, A. M. and Roerdink, J. B. T. M., Polyphase decompositions and shift-invariant discrete wavelet transforms in the frequency domain. *Signal Processing*, 90(6), 1779-1787 (2010). Available on: < <http://www.sciencedirect.com/science/article/pii/S016516840900485X> >.
- Xue, J., Leung, Y. and Ma, J. H., High-order Taylor series expansion methods for error propagation in geographic information systems. *Journal of Geographical Systems*, 17(2), 187-206 (2015). Available on: < <http://dx.doi.org/10.1007/s10109-014-0207-x> >.
- Zhang, S., Zheng, J., Wenliang, G., Zhang, H., Xun, H. and Hong, G., Application of spline wavelet transform in differential of electroanalytical signal. *Chinese Science Bulletin*, 46(7), 550 (2001). Available on: < http://engine.scichina.com/publisher/Science_China_Press/journal/Chinese_Science_Bulletin/46/7/10.1007/BF02900406 >.
- Zhang, X. Y., Trame, M. N., Lesko, L. J. and Schmidt, S., Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT: Pharmacometrics & Systems Pharmacology*, 4(2), 69-79 (2015). Available on: < <http://dx.doi.org/10.1002/psp4.6> >.