CHEMICAL ENGINEERING TRANSACTIONS
VOL. 35, 2013

A publication of
AIDIC
The Italian Association of Chemical Engineering
www.aidic.it/cet

Guest Editors: Petar Varbanov, Jiří Klemeš, Panos Seferlis, Athanasios I. Papadopoulos, Spyros Voutetakis

ISBN 978-88-95608-26-6; **ISSN** 1974-9791

DOI: 10.3303/CET1335155

Historical Process Data Based Energy Monitoring - Model Based Time-Series Segmentation to Determine Target Values

Janos Abonyi[a,*], Tibor Kulcsar[a], Miklos Balaton[b], Laszlo Nagy[b]

[a]Department of Process Engineering, University of Pannonia, P.O. Box 158, H-8201 Hungary
[b] MOL Hungarian Oil and Gas Company, Szazhalombatta, Production Excellence
janos@abonyilab.com

Energy monitoring systems calculate actual energy use, estimate energy needs at normal operation, track energy metrics, and highlight issues related to energy efficiency of process plants. Analysis of the Key Energy Indicators (KEIs) allows the comparison of operation strategies at different operating regimes. Based on the extracted knowledge realistic targets of KEI-s can be determined. The performance of data-driven targeting models depends on operating regimes determined by a complex set of process variables. Till now this modelling task is performed manually based on heuristic and subjective evaluation of the operation. We developed a goal-oriented time-series segmentation technique to automate the selection of proper dataset used for the identification of targeting models. With the proposed tool target-models for different operating regions can be automatically determined. The concept of the resulted energy monitoring system is demonstrated at Heavy Naphtha Hydrotreater and CCR Reforming Units of MOL Hungarian Oil and Gas Company.

1. Introduction

Advanced production systems expected to maximize the production and at the same time minimize cost and emission. The purpose of energy monitoring and targeting is to provide better understanding of how energy is being used. The so called energy portfolio allows the classification and prioritization of energy consumers and the derivation of target-oriented action plans towards energy and resource efficiency improvement (Thiede, et al. 2012). Energy monitoring improves energy efficiency in process plants by helping plant operators, engineers and managers to track actual and target energy consumption. Such system allows the user to ("Monitoring and targeting," 2010):

1. Detect avoidable energy waste that might otherwise remain hidden. This is waste that occurs at random because of poor control, unexpected equipment faults or human error.
2. Quantify the savings achieved by energy projects and campaigns.
3. Identify fruitful lines of investigation for energy surveys.
4. Provide feedback for staff awareness, improve budget setting and undertake benchmarking.

Monitoring is based on continuous comparison of actual and estimated energy consumption. Methods for calculating expected consumption fall into two categories. There are those based on precedent (comparison with previous periods), and activity-based methods that relate expected consumption to its driving factors. Energy efficiency is has the following four components: performance efficiency, operation efficiency, equipment efficiency, and technology efficiency (Xia and Zhang, 2010). A systematic overview of the state of the art in energy and resource efficiency increasing methods and techniques in the domain of discrete part manufacturing is given in (Duflou, et al., 2012). In this paper a structured approach, distinguishing different system scale levels, is applied: starting from a unit process focus, respectively the multi-machine, factory, multi-facility and supply chain levels are covered.

Please cite this article as: Abonyi J., Kulcsar T., Balaton M., Nagy L., 2013, Historical process data based energy monitoring - model based time-series segmentation to determine target values, Chemical Engineering Transactions, 35, 931-936 DOI:10.3303/CET1335155

Reducing energy consumption of machine tools can significantly improve the environmental performance of manufacturing systems. To achieve this, monitoring of energy consumption patterns by event stream processing techniques are applied to automate the monitoring and analysis of energy consumption (Vijayaraghavan and Dornfeld, 2010). Methods for calculating expected consumption fall into two categories. There are those based on precedent (comparison with previous periods), and activity-based methods that relate expected consumption to its driving factors. Energy consumption characteristics of machine tools are compared and the potential of using the obtained data for energy labelling of machine tools is discussed in (Behrendt et al. 2012).

Most of these developments are focused to discrete manufacturing (Kellens et al. 2013). Chemical industry is the largest energy consumer among different industrial sectors; it is responsible for around 4.7 % of the total energy consumption in Europe. We focus to this application area. We apply activity-based targeting models to calculate expected consumption reference of process units. This approach can be considered as a special software sensor (Zaouak, 2012). The applied targeting models are based on historical process data. This dataset is selected based on heuristic and subjective evaluation of the operation of the process. This procedure is time-consuming and unfortunately it does not give any hint to the user how the targets given by the resulted models should be handled. We developed a goal-oriented time-series segmentation technique to automate this procedure. With the proposed tool target-models for different operating regions can be automatically determined. The concept of the resulted historical data based energy monitoring system is demonstrated at Heavy Naphtha Hydrotreater and CCR Reforming Units of MOL Hungarian Oil and Gas Company.

2. Targeting model based energy monitoring

Precedent-based targeting models are used when expected consumption can be deduced from what was used in the corresponding day or month before. Automatic monitoring and targeting schemes attempt to compare consumption at very short intervals (e.g. minutes). For such applications precedent-based targeting models can be too simplistic. Activity-based targeting is particularly appropriate when there are clear drivers for changing energy consumption, for example, changes in production throughput ("Monitoring and targeting," 2010). These targets can be useful to reduce energy consumption through real-time comparisons of the actual energy flow vs. the targeted rates.

Activity-based energy targets are usually calculated by linear regression models,

$$\widehat{y_k} = \boldsymbol{x}_{\boldsymbol{k}}^T \boldsymbol{\theta} \ , \tag{1}$$

where the calculated output $\hat{y}_i$ is the linear combination of process variables (drivers), $\boldsymbol{x}_k = [x_{1,k}, \dots, x_{n,k}]$, where *k* represents the k-th sampling time and *n* stands for the number of process variables having significant effect to the energy consumption.

At the development of this model it is important to ensure that data are synchronised as closely as possible with the required assessment intervals. Based on a synchronized set of data $\{y_k, \boldsymbol{x}_k\}, k = 1, \dots, N$ linear least squares method can be applied to find optimal parameters of the model $\boldsymbol{\theta}$ that minimizes the $\sum (y_k - \widehat{y_k})^2$ quadratic cost function.

$$\boldsymbol{\theta} \ = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2}$$

where $\mathbf{X}$ is $n \times N$ matrix of historical process variables and $\mathbf{y}$ is an $n \times 1$ vector of measured output variable (energy consumption or efficiency measure).

When the predicted consumption $\widehat{y_k}$ is higher as the measured value y_i the technology is considered to be efficient regards to historical data. The relation $\widehat{y_k} < y_k$ suggests that the technology could work with lower energy consumption.

The selection of data to be used for targeting is the most critical task of model development since the quality of this dataset determines the quality of the monitoring system. This selection mainly focuses to excluding periods of abnormal operations such as unit shutdown/start-up, and also eliminating clear outliers. This analysis is usually done manually and based on subjective evaluation of segments of available historical process data. Furthermore, all the targets should be revisited when unit throughputs or other input conditions change significantly from the ranges of the data used for modelling.

We developed a time-series segmentation algorithm to handle these problems. The proposed algorithm automatically selects homogenous segments of operation and extracts models that effectively describe energy efficiency and consumption at different operating regions.

3. Regression based time series segmentation

The selection of appropriate data sequences can be considered as time-series segmentation. Time-series segmentation is often used to extract internally homogeneous segments from a given time series. The bottom-up algorithm is the most widely applied for off-line purposes. We developed a regression model based segmentation algorithm based on this bottom-up scheme to detect changes in multivariate time-series of model inputs (drivers) and outputs (energy consumption or efficiency measure).

A time series $T = \left\{ \boldsymbol{x}_k = [x_{1,k}, x_{2,k}, \dots, x_{n,k}]^T | 1 \le k \le N \right\}$ is a finite set of N n-dimensional samples labelled by time points $t_1, \dots, t_N$. A segment of T is a set of consecutive time points $(a,b) = \{a \le k \le b\}, \boldsymbol{x}_a, \boldsymbol{x}_{a+1}, \dots, \boldsymbol{x}_b$. The c-segmentation of time series T is a partition of T to c non - overlapping segments $\{(a_i, b_i) | 1 \le i \le c\}$, such that $a_1 = 1, b_c = N$, and $a_i = b_{i-1} + 1$. c-segmentation splits T to c disjoint time intervals by segment boundaries$s_1 < s_2 < \dots < s_c$, where $S_i(s_i, s_{i+1} - 1)$.

The goal of segmentation is to find internally homogeneous, information rich segments from a given time series. To formalize this goal, a cost function describing the internal homogeneity of individual segments should be defined. This cost function $cost(a,b)$ is defined based on the distances between actual values of the time series and values given by a simple model of the segment.

Usually, the average of the variables is used as a simplest model of the segment. In this case the sum of variances defines the $cost(a,b)$ cost of the segment,

$$cost_i(a_i, b_i) = \frac{1}{b_i - a_i + 1} \sum_{k=a_i}^{b_i} \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2, \boldsymbol{v}_i = \frac{1}{b_i - a_i + 1} \sum_{k=a_i}^{b_i} \boldsymbol{x}_k. \tag{3}$$

Segmentation algorithms simultaneously determine the parameters of the models used to approximate the behavior of the system in the segments, and the a_i, b_i borders of the segments by minimizing the sum of the costs of the individual segments:

$$cost_T^c = \sum_{i=1}^{c} cost_i \tag{4}$$

Our special application requires a goal oriented cost function and a goal oriented segment model. All segments have the same model structure but the functional relationship is described by different parameters:

$$\widehat{y_k} = \boldsymbol{x}_k^T \boldsymbol{\theta}_i\ , \tag{5}$$

$$cost_i(a_i, b_i) = \frac{1}{b_i - a_i + 1} \sum_{k=a_i}^{b_i} (y_k - \hat{y}_k)^2 = \frac{1}{b_i - a_i + 1} \sum_{k=a_i}^{b_i} \left(y_k - \boldsymbol{x}_k^T \boldsymbol{\theta}_i\right)^2 \tag{6}$$

where the $\boldsymbol{\theta_i}$ parameters of the segments are determined based on a dataset of the segment,

$$\boldsymbol{\theta_i} = \left(\mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \mathbf{X}_i^T \mathbf{y}_i \tag{7}$$

$$\mathbf{X}_i = \left[\boldsymbol{x}_{a_i}, \dots, \boldsymbol{x}_{b_i}\right]^T, \mathbf{y}_i = \left[y_{a_i}, \dots, y_{b_i}\right]^T.$$

In data mining, the bottom-up algorithm has been used extensively to support a variety of time series data mining tasks for off-line analysis of process data. The algorithm starts with large number of segments and merges neighbouring segments having low merging costs, $t_i(a_i, b_{i+1})$. Several stopping criteria can be defined, e.g. the minimal error of the model (minimal allowed cost of the segment) or the minimal number of segments. When the pair of adjacent segments $cost_i$ and $cost_{i+1}$ are merged, the cost of merging the new segment with its right neighbour and the cost of merging the $cost_{i-1}$ segment with its new larger neighbour is calculated.

The algorithm automatically finds the segment borders and the related model parameters, so the user only has to evaluate the result of segmentations and analyse the differences among the model parameters. We implemented the algorithm in MATLAB.

The code is downloadable from (Abonyilab, 2013).

4. Results and discussion

The concept of the resulted historical data based energy monitoring system is demonstrated at Heavy Naphtha Hydrotreater and CCR Reforming Units of MOL Hungarian Oil and Gas Company. The fuel gas consumption of a furnace is analysed as demonstrating example

The furnace is operated by fuel gas from the refinery's fuel gas network. Fuel gas quality and flow are measured. The fuel gas consumption target was calculated based on one-year historical data. We assume that the range of this dataset is wide enough, so it covers operation rages with high and low energy consumptions. The following drivers of fuel gas consumption were identified based on the analysis of the technology and the data: total feed, inlet temperature, density of fuel gas, and ambient temperature.

We applied least squares regression to obtain the parameters of this "Global" model (see Table 1). Figue. 1 compares targeted (predicted) energy flows and actual data. This correlation diagram helps to qualify the energy efficiency of the technology. When all the data is taken into account the target model estimates the average energy consumption. This means, when the estimation of the target model is bigger than the measured value the process operates well. Otherwise it should be checked, what is the reason of a higher consumption than expected from the current operating parameters. For sophisticated decision support the confidence of the model should also be taken into account. In this case for an estimated target value, $\hat{y}$, $[\hat{y} + \delta(y) \ldots \hat{y} + \delta(y)]$ bounds related to a given confidence level α can be calculated.

The confidence interval calculation is based on Student distribution with $n-2$ freedom factor. Where α is the confidence level (probability).

$$\hat{y}^* = \hat{y} \pm t_{1-\frac{\alpha}{2}} * s'_{\hat{y}_e} = \hat{y} \pm \delta(\hat{y}_e),\;\; P(y + \Delta y_e - \delta(y) < \hat{y} < y + \Delta y_e + \delta(y)) = 1 - \alpha \tag{8}$$

$$s'_{\widehat{y}_e} = \sigma_e^2 \sqrt{\boldsymbol{x}_k^T (\mathbf{X}^T\mathbf{X})^{-1} \boldsymbol{x}_k},\; s_e = \sqrt{\frac{\sum (y-\hat{y})^2}{N-2}} \tag{9}$$

Table 1: Parameter values in global model and in the best two segments

Model	Total feed	Inlet temperature	Density of fuel gas	Ambient temperature	bias
Global	0.6332	-0.7888	-0.3184	-0.0471	0.9946
Segment 4	0.5730	0.0214	-0.3074	-0.1729	0.4450
Segment 17	0.6388	-0.4766	-0.3794	0.0195	0.7518

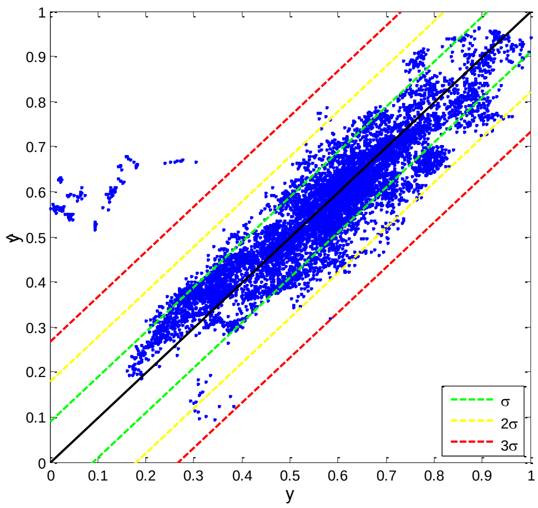

Figure 1: Correlation diagram of measured and predicted energy consumption. The black line shows the ideal prediction; the dashed lines show Q levels based on standard deviation (σ) of prediction error ($y - \hat{y}$). Colored lines belong to $1\sigma, 2\sigma, 3\sigma$ Q levels. These lines can be used as tuning aggressive or conservative behavior of the monitoring system. We scaled the data since the nominal values of process variables are confidential.

Fig. 1 and Eq. 8 and 9 illustrate the accuracy of the model (the standard deviation of the estimation error) has significant affect to the applicability of the monitoring system. The tuning (how aggressive or conservative will be the model) is realised by shifting the predicted output based on the variance of the modelling error, and the confidence bound is also based on this measure. To reduce this variance and increase model accuracy abnormal consumption patterns has to be filtered out from the pool of data and models related to different operating regimes should be defined.

Instead of manual selection of these operating periods and manual removing outliers we applied the proposed time-series segmentation algorithm (see Figure 2). The comparison of the parameters of the local models can give useful information about how the impact of the drivers changing from one operating regime to another. The performances of these local models are shown in Figure 3.

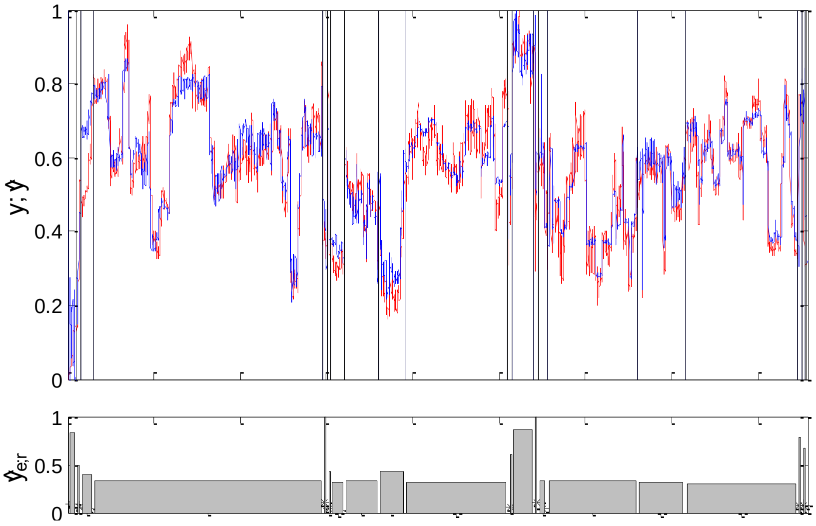

Figure 2: Results of the segmentation. Estimated and measured energy consumption values are shown. The bottom subfigure shows the average mean square errors of the models. Based on the analysis of the results the most important operating modes of the process can be automatically determined. The parameters of the models related to the two most relevant segments are given in Table 1.

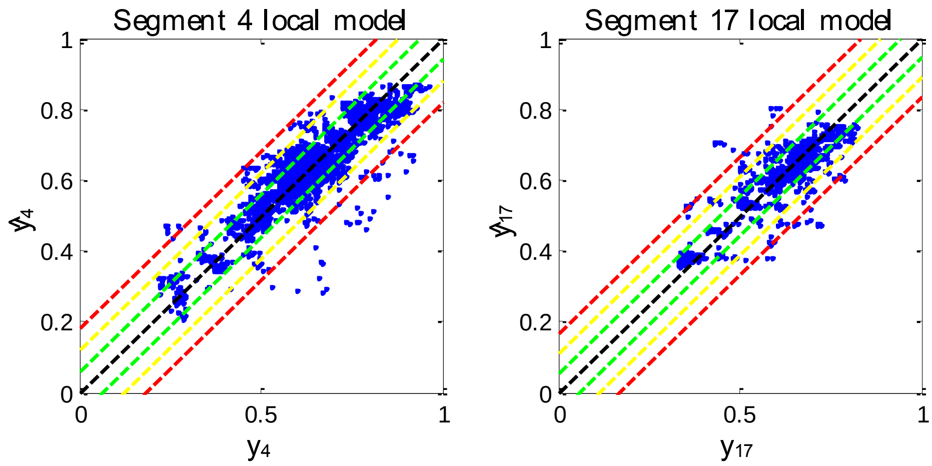

Figure 3: Performances of the local target models in their segments. As can be seen the models has much smaller variance, so they can provide much more accurate and relevant information for energy monitoring.

5. Conclusions

Energy monitoring improves energy efficiency in process plants by helping plant operators, engineers and managers to track actual and target energy. Energy monitoring is based on the comparison of Key Energy Indicators (KEIs) and their target vales. These targets depend on operating regimes determined by a complex set of process variables. We developed advanced data-driven modelling techniques to support on-line targeting. We showed that the quality of the targeting models is mostly influenced by the quality of historical process data used for the identification of the parameters of the targeting model. Till now the selection of proper training data is performed manually based on heuristic and subjective evaluation of the operation of the process. We developed a goal-oriented time-series segmentation technique automate this procedure. With the proposed tool target-models for different operating regions can be automatically determined. The presented case study shows the applicability of the proposed methodology, we were able to extract set of accurate models and set of operating regimes showing different impacts of the drivers of energy consumption and efficiency. Once the proposed scheme has been set up, building and analysis of targeting models is routine operation and should be neither time-consuming nor complex procedure.
We also showed that deviations from the target should be investigated based on the analysis of the variance of the model error. Based on this analysis the aggressive or conservative behaviour of energy monitoring can be easily tuned. Further analysis will focus on the detailed comparison of the operating regimes and models.

Acknowledgement

This publication/research has been supported by the European Union and the Hungarian Republic through the projects TÁMOP-4.2.2.A-11/1/KONV-2012-0071 and TÁMOP-4.1.1.C-12/1/KONV-2012-0017.

References

Abonyilab, <www.abonyilab.com> accessed 14.07.2013

Behrendt, T., Zein, A., Min, S., 2012, Development of an energy consumption monitoring procedure for machine tools. CIRP Annals - Manufacturing Technology, 61(1), 43–46. doi:10.1016/j.cirp.2012.03.103, 2012

Duflou, J. R., Sutherland, J. W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M.. 2012, Towards energy and resource efficient manufacturing: A processes and systems approach. CIRP Annals - Manufacturing Technology, 61(2), 587–609. doi:10.1016/j.cirp.2012.05.002, 2012

Kellens, K., Dewulf, W., Lauwers, B., Kruth, J.-P., Duflou, J. R., 2012, Environmental Impact Reduction in Discrete Manufacturing: Examples for Non-Conventional Processes. Procedia CIRP, 6(Lci), 27–34. doi:10.1016/j.procir.2013.03.003, 2012

Monitoring and Targeting. 2010, doi:10.4324/9780203349021_chapter_10, 2010

Thiede, S., Bogdanski, G., Herrmann, C., 2012, A Systematic Method for Increasing the Energy and Resource Efficiency in Manufacturing Companies. Procedia CIRP, 2, 28–33. doi:10.1016/j.procir.2012.05.034, 2012

Vijayaraghavan, A., Dornfeld, D., 2010, Automated energy monitoring of machine tools, CIRP Annals - Manufacturing Technology, 59(1), 21–24. doi:10.1016/j.cirp.2010.03.042, 2010

Xia, X., Zhang, J., 2010, Energy efficiency and control systems-from a POET perspective. Methodologies and Technology for Energy Efficiency. <www.ee.up.ac.za/main/_media/en/postgrad/subjects/ees732/2010 portugalpaper124.pdf> Accessed 14.07.2013

Zaouak O, Daoud B. Fages M, Fanlo J. Aubert, B., 2012, High performance cost effective miniature sensor for continuous network monitoring of h2s, Chemical Engineering Transactions, 30, 325-330, 2012