FACULTY OF TECHNOLOGY

INDIRECT MONITORING OF ENERGY EFFICIENCY IN A SIMULATED CHEMICAL PROCESS

Teemu Pätsi

DEGREE PROGRAMME IN ENVIRONMENTAL ENGINEERING

Master's thesis

July 2021

ABSTRACT

Indirect monitoring of energy efficiency in a simulated chemical process

Teemu Pätsi

University of Oulu, Degree Programme in Environmental Engineering

Master's thesis 2021, 53 pp. + 1 Appendix

Supervisor(s) at the university:

Prof. Mika Ruusunen, D.Sc. Markku Ohenoja, D.Sc. Petri Österberg, D.Sc. Tero Vuolio

Energy efficiency is an important part of chemical process sustainability. Wasted energy contributes significantly to process costs and overall emissions. Therefore, contributions to improving energy efficiency in chemical processes are of value. The main objective of this thesis is the exploration of indirect energy efficiency monitoring methods and their compilation into a generalized framework. As part of the proposed framework, data-based modelling methods were explored and used to identify a model for estimating energy efficiency in a simulated process. The proposed framework can act as a potential tool in different practical applications with energy efficiency improvements as an objective.

As a simulated test process for this thesis, the Tennessee Eastman process was utilized. This process is widely used in research, especially regarding fault diagnosis and control design. The process includes slow dynamics and nonlinearity, providing interesting challenges for research. Even though the process has been studied extensively, the energy efficiency aspect of the process has not been taken into account in research.

The results of the thesis show that data-based models provide sufficient accuracy for real-time estimation of energy efficiency for the Tennessee Eastman process. Parts of the proposed framework were tested with the explored methods, but some areas were beyond the scope of this thesis. As such, further research, for example prediction of the energy efficiency horizon, fault diagnosis and advanced process control, could be beneficial.

Keywords: energy efficiency, Tennessee Eastman, digitalization, soft sensor

TIIVISTELMÄ

Energiatehokkuuden epäsuora monitorointi simuloidussa kemiallisessa prosessissa

Teemu Pätsi

Oulun yliopisto, Ympäristötekniikan tutkinto-ohjelma

Diplomityö 2021, 53 s. + 1 liite

Työn ohjaajat yliopistolla:

Prof. Mika Ruusunen, TkT Markku Ohenoja, TkT Petri Österberg, TkT Tero Vuolio

Energiatehokkuus on tärkeä osa kemiallisen teollisuuden kestävyyttä. Energian käytön tehottomuus näkyy merkittävästi kasvavina prosessikustannuksina ja kokonaispäästöinä. Toimet energiatehokkuuden nostamiseksi ovat siksi merkityksellisiä. Diplomityön päätavoitteena on erilaisten epäsuorien energiatehokkuuden seurantamenetelmien tutkiminen ja niiden kokoaminen yleistettävään menetelmäkehykseen. Datapohjaisia mallinnusmenetelmiä tutkitaan osana esitettyä kehystä, ja niitä hyödynnetään energiatehokkuutta arvioivan mallin muodostuksessa. Esitetty menetelmäkehys voi toimia mahdollisena työkaluna erilaisissa käyttökohteissa, joissa energiatehokkuuden parantaminen on päämääränä.

Tutkittavana kohteena diplomityössä käytettiin simuloitua Tennessee Eastman prosessimallia. Vaikka prosessia on tutkittu laajasti, energiatehokkuuden tarkempi tarkastelu on jäänyt vajaaksi. Simuloitua prosessidataa hyödynnettiin tässä työssä prosessin energiatehokkuuden mallipohjaisen arvion muodostuksessa. Työssä analysoitiin myös mallinnuksen luotettavuuteen vaikuttavia tekijöitä, kuten opetusdatan rajallisuutta ja siitä seuraavaa mallin ekstrapolointia.

Diplomityön tulokset osoittavat, että Tennessee Eastman prosessin energiatehokkuuden reaaliaikainen arviointi datapohjaisilla menetelmillä onnistuu riittävällä tarkkuudella. Esitetyn menetelmäkehyksen osia testattiin tutkituilla menetelmillä, mutta jotkin alueet jäivät työn ulkopuolelle. Tulevaisuuden mahdollisiin tutkimusalueisiin kuuluukin energiatehokkuuden ennustaminen, vikadiagnostiikka ja niitä yhdistävä kehittynyt prosessisäätö.

Asiasanat: energiatehokkuus, Tennessee Eastman, digitalisaatio, epäsuora mittaus

Preface

This thesis was related to work done in project ‘Operational eXcellence by Integrating Learned information into AcTionable Expertise (OXILATE)’ in Environmental and Chemical Engineering Group at University of Oulu.

I would like to thank all my supervisors for their excellent assistance and patience during the period of my thesis. D.Sc. Markku Ohenoja and D.Sc Petri Österberg have provided clear and helpful advice, in addition, D.Sc. Tero Vuolio provided a wide variety of expertise regarding data-based modelling in the form of scripts and explanations for the methods, and Prof. Mika Ruusunen provided a wider vision for the scope of the study. Most of all, I would like to thank my family for their support during my studies.

Kajaani, 13.7.2021

Teemu Pätsi

TABLE OF CONTENTS

NOTATIONS AND ABBREVIATIONS

EE	Energy efficiency
β	Coefficient parameter
c_i	Concentration of component i
$c_{p,i}$	Heat capacity of i
CW	Cooling water
d	Euclidean distance
D_{KL}	Kullback–Leibler divergence
ε	Residual
$\mathbf{F}$	Scaled principal component matrix
$\boldsymbol{\gamma}$	Coefficient vector
$\dot{m}_i$	Mass flow of i
MAE	Mean absolute error
$MAPE$	Mean absolute percentage error
MLR	Multiple linear regression
MSE	Mean squared error
ρ	Density
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares
$\dot{Q}_i$	Flow of heat from i
R	Correlation coefficient
$RMSE$	Root mean squared error
s	Histogram intersection
SEC_j	Specific energy consumption for product j
T_i	Temperature of i
TE	Tennessee Eastman
$\bar{V}_i$	Volumetric flow of stream i
VIF	Variation inflation index

1 INTRODUCTION

Chemical processes are very energy intensive with a high volume of production. Thus, even small improvements to energy efficiency decrease the amount of consumed energy significantly. In addition, the consumed energy covers a large fraction of the overall production costs, making energy efficiency improvement a worthy investment. Energy consumption reduction is also important for achieving sustainability and reducing emissions.

The objective of this thesis is to explore methods for monitoring energy efficiency in an energy-intensive, multi-stage chemical process. Monitoring would allow for displaying the current and past status of operations to make necessary adjustments to minimize otherwise wasted energy. Data-based modelling methods could be utilized to estimate the current and future energy efficiency of a process instead of expensive or slower hardware measurements. The information provided by the model could then be refined and used to guide the process operators. In addition, fault diagnosis and optimal production planning could be performed by comparing the observed energy efficiency with the modelled energy efficiency horizon. For this aim, a model-based monitoring framework for energy efficiency is proposed.

As the test problem for the framework, a benchmark simulator of a complex chemical process is utilized. The simulated data obtained is used to model energy efficiency. The studied process, namely Tennessee Eastman, is a chemical process that is widely used as a test case in academia. Even though the process has been extensively explored in research, the resulting debate still lacks the energy efficiency aspect.

To develop the monitoring framework, a literature review was performed regarding energy monitoring applications and indirect monitoring methods, which is presented in Chapter 2. The test case for the thesis, namely the Tennessee Eastman benchmark model, is introduced in Chapter 3. Then, the proposed framework is synthesized from the explored methods, constructed into a flowsheet for energy efficiency monitoring, and possible applications for the framework are explored in Chapter 4. The indirect energy efficiency model is developed and analysed in Chapter 5, followed by conclusions drawn from the results presented in Chapter 6. Finally, the thesis is summarized in Chapter 7.

2 ENERGY EFFICIENCY

Energy efficiency is an important and timely topic in process engineering. The chemical process industry often involves very high energy amounts and production volumes, and energy losses can add up to a substantial portion of process costs, besides the obvious impact on the environmental footprint. These processes have several successive sub-processes requiring energy, often in the form of heat, for example distilling and reactor heating, or electricity, such as pressurizing compressors and pumps. Inefficient energy usage can also be seen to increase the amount of greenhouse gas emissions from processes (Akdag and Yıldırım 2020). Optimizing energy efficiency is therefore regarded as a smart action for both the business and the ecological side of the industry. In addition, governmental regulations, for example in Finland, require large companies to review the energy usage of their energy applications at least every four years. (FINLEX 2014)

2.1 Definition

Energy efficiency (EE) can be defined as the ratio of power and the production rate. In this study, energy efficiency is defined as follows:

$$EE = \frac{P}{\dot{m}}, \tag{1}$$

where P (J/s) is the power consumption of the process, comprising the energy required for the process and any residual energy present within the process, and $\dot{m}$ is the production rate of the end product (kg/s) minus the product loss caused by sub-standard product quality. The objective in optimizing energy efficiency is to minimize energy usage and maximize the obtained product within the specification range. Losses within the process also reduce the overall energy efficiency, for example leaks and spills, insufficient product quality, and inefficient energy utilization. For example, control strategy (Nigitz et al. 2020), process design (Oh et al. 2018) and operator decisions (Chen et al. 2021) are all factors that affect the energy efficiency of a process.

2.2 Intelligent monitoring

In the early 1980s, a complete set of measurements was required for energy management (Kaya and Keyes 1980). Development in energy efficiency monitoring has been substantial in this regard; intelligence can be integrated into energy efficiency monitoring by implementing soft sensors, namely data-based models or digital twins. In many industrial applications, direct monitoring of energy usage might be infeasible due to lack of measurements or the complex nature of the process, favouring computational methods and data fusion instead. Existing process measurements might allow for data-based models for estimating energy efficiency indirectly. Being able to predict the energy efficiency in a finite time horizon may further allow the dynamic optimization of the process. Improving product quality is another way to improve energy efficiency by reducing the amount of discarded product due to insufficient quality and reducing the need for recycle feeds (Luan et al. 2018). Some methods that have been implemented and studied for energy efficiency monitoring in process and manufacturing industries are explored in the following sections.

2.2.1 Reported methods

The energy efficiency and utilization of intelligent monitoring methods in production processes has been of great interest. Recently, energy efficiency has gained attention especially in the context of smart manufacturing and Industry 4.0 (Tesch da Silva et al. 2020). A noticeable trend in research beyond mere optimizing of energy usage is exploring techniques of information management and reporting for higher levels of management. Accurate energy usage monitoring allows for improved energy management (Monjurul Hasan and Trianni 2020). Chemical processes in particular require robust process monitoring methods because of safety regulations, and usually quite strict product specifications (Farsang et al. 2014).

Bauerdick et al. (2017) present a software framework for energy monitoring in machine tool manufacturing. Data acquired from sensors is aggregated and used for analysing energy usage. It was found that adjusting the start-up times for different machines allowed reduction of the maximum load peak on the energy grid. Bauerdick et al. (2017) also studied fault detection by analysing energy consumption and finding abnormalities in the

power demand. The fault scenario is identified by reduced power demand during machining process when compared to a reference work piece.

Gontarz et al. (2015) implemented a soft-sensor approach for monitoring energy efficiency in a machine tool. Measurements are gathered from the machine with internal sensors, and necessary external sensors are implemented to calibrate and verify simulated models for the energy usage behaviour of machine tool components. The gathered data can be evaluated and parameterized. The energetic behaviour of machine tools can be identified and further optimized on the component level.

Qin et al. (2017) introduced an Internet-of-Things-based framework for energy consumption. In their framework, energy consumption analysis is performed in a layered order. In the first layer data from the target system is extracted with sensors and components. The collected data is fed into the second layer, where gathered data is compared with the total energy consumption and attributes that relate to the total energy consumption. Any new attributes with relations to either of these are marked and the data related to that attribute is stored. The next layer displays the obtained data in different implementations. System energy consumption and prediction for energy use can be used to guide operators. Different analyses and reports on the system life cycle and energy sustainability can also be obtained from this layer.

Tan et al. (2017) introduced a framework for Internet-of-Things-enabled, real-time monitoring of energy efficiency. This approach allows for data analysis techniques to identify abnormal energy usage by monitoring energy consumption and the amount of product obtained. Process efficiency is described by specific energy consumption, defined by the amount of energy used per product. Current performance is then compared to historical performance to find relative efficiency, and potential gaps are identified for energy efficiency improvement.

A monitoring framework for estimating expected boiler efficiency is described in Nikula et al. (2016). The framework presented by the authors makes use of Multiple Linear Regression (MLR) for estimating energy efficiency. The MLR model was obtained using variables chosen by ranking, using mutual information and Shannon’s entropy. Process efficiency is estimated for each process state and deviations from the expected value are monitored and can be displayed for process operators and supervisors. The estimated

process efficiency is obtained by finding the maximum historical boiler efficiency for each process state. If the actual boiler efficiency is higher than expected, new efficiency values are assigned for the process states in real time, making the model highly adaptable. If the actual efficiency is lower than estimated, process states leading up to that point are diagnosed and variables with values outside the operational range during the diagnosed period are identified.

A variety of tools and methods can be employed for estimating energy usage and efficiency. A literary review presented in Narciso and Martins (2020) concluded that neural networks are commonly used for energy efficiency and consumption forecasts (see e.g. Geng et al. 2017; Zhang et al. 2018). Other frequently observed methods are regression-based (Golkarnarenji et al. 2018) and fuzzy models (Geng et al. 2018). The number of published papers on machine learning in industrial energy efficiency has increased lately, proving the possibility for increased insight into energy efficiency monitoring using intelligent methods (Narciso and Martins 2020).

2.2.2 Commercial solutions

Companies are always interested in increasing profits and reducing energy usage is an important factor for achieving this goal. Many companies have noticed the possibility for business in this area and have started working on commercial solutions for monitoring and improving energy efficiency. Hence, a variety of commercial products already exist, two examples of which are described below.

STRUCTese, developed by Bayer MaterialScience, is an energy efficiency management system. Current energy efficiency is analysed and potential improvements for energy savings are identified and evaluated. A major element of STRUCTese is the energy loss cascade, which provides easy reporting of energy usage for plant managers. Theoretical energy usage and measured energy losses are plotted and displayed, showing where more focus could be beneficial for improving overall energy efficiency. (Drumm et al. 2013)

ABB has developed Energy Manager, an application for reducing energy costs and the carbon footprint. Energy Manager optimizes energy operations by reading data from different process systems. Monitoring and reporting can be used to identify areas of

improvement by comparing actual energy efficiency to targets. Energy consumption is predicted, allowing for energy scheduling and load planning. (ABB 2019)

3 TENNESSEE EASTMAN PROCESS

The Tennessee Eastman (TE) process has been used as a case study for identifying a model for energy efficiency estimation in this work. The Tennessee Eastman process model was originally developed as a test case for the process control academic community by Downs and Vogel (1993). The process is a multistage, multi-product chemical process which is well suited for benchmarking. The model is based on an actual industrial chemical process, but accurate details have been left out to protect the proprietary nature of the process.

3.1 Role as a benchmark

The TE process has been used as a benchmark problem in many studies focusing on control design, optimization and fault diagnosis. Process nonlinearity and constraints related to the model have proved to be an interesting challenge for control design (Ricker 1993; Ricker and Lee 1995a; Ricker and Lee 1995b). As an example, Jockenhövel et al. (2003) focused on the real-time dynamic optimization of the TE process with nonlinear programming. On the other hand, Golshan et al. (2005) minimized the process costs in real time with an optimization algorithm based on sequential quadratic programming method and first-order linear filters which adjust controller setpoints gradually in order to keep the process under control. A linear model predictive control strategy for the TE process was implemented in Jämsä (2018). A control strategy developed in Larsson et al. (2001) focuses on process optimization with self-optimizing control methods and emphasizing the importance of variable selection for control design.

3.2 Process description

The process produces two products (G, H) from four reactants (A, C, D, E) in a reactor. A by-product, F, is also produced by two additional reactions. The actual components are kept hidden to protect the proprietary nature of the process. A non-volatile catalyst dissolved in the liquid phase accelerates the gas phase reactions. The reactions are irreversible and exothermic, requiring constant cooling. The reactions are as follows:

$A(g)+C(g)+D(g) \rightarrow G(liq)$	Product 1	(2)
$A(g)+C(g)+E(g) \rightarrow H(liq)$	Product 2	(3)
$A(g)+E(g) \rightarrow F(liq)$	By-product	(4)
$3D(g) \rightarrow 2F(liq)$	By-product	(5)

The above reactions (2)–(5) are a function of temperature, following the Arrhenius equation. The reaction for producing G is more sensitive to temperature due to a higher activation energy. The reactions can be considered as first-order in terms of reactant concentrations. There are five main unit operations in the process. These are described after the process diagram, shown in Figure 1.

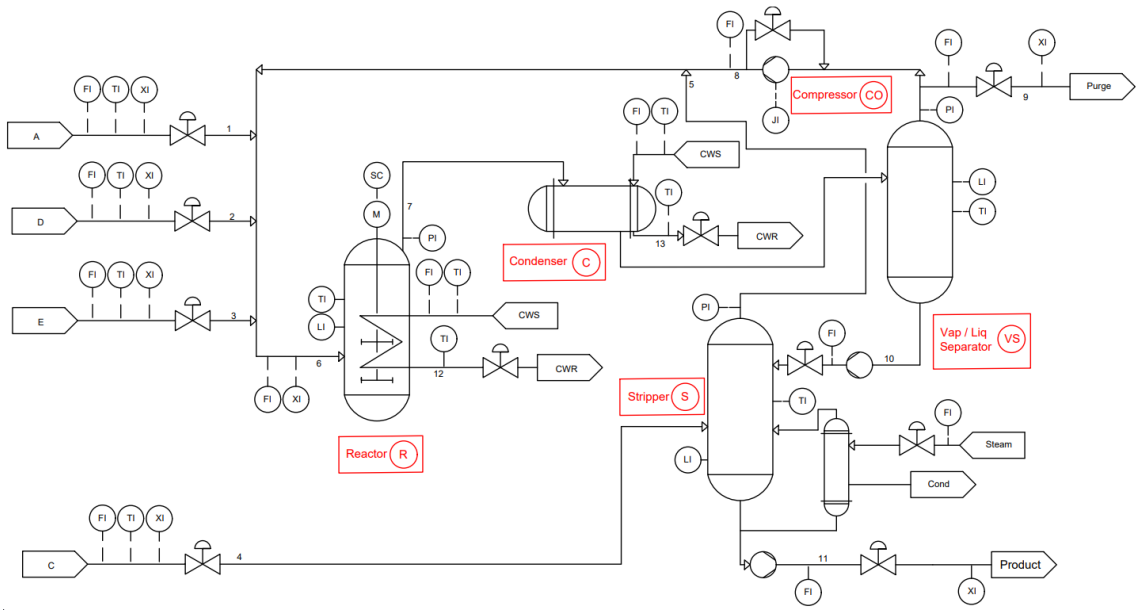

Figure 1. Process and instrument diagram of the Tennessee Eastman process (adapted from Bathelt et al. 2015; Downs and Vogel 1993).

Reactant gases are fed with feed streams A, D, and E to the reactor (R) where liquid products are produced. The temperature of the reactor can be controlled by the cooling water circulation. The gaseous reactor products are fed into a condenser (C), leaving the liquid phase in the reactor. A water-cooling system reduces the temperature in the condenser, condensing the gases into liquids. The non-condensed gaseous reactants are recycled back to the reactor with a vapour-liquid separator (VS). The pressure drop in the

stream is compensated with a compressor (CO) to maintain a constant pressure in the reactor. The inert and by-product components are mainly purged with the vapour from the vapour-liquid separator to avoid build-up in the system. The liquid products go into a stripper (S) where the remaining reactants are separated from the products and fed back into the reactor. (Downs and Vogel 1993)

Downs and Vogel (1993) describe six modes of operation for the process at different product ratios in the final product stream. The base case, mode 1, produces products G and H at a ratio of 50/50 at a desired production rate. The process modes are described in more detail in Downs and Vogel (1993). The TE process also involves process constraints, primarily to protect the equipment. If the process conditions go out of control, the process is automatically shut down. (Downs and Vogel 1993)

The process has 12 controlled and 41 measured variables. Of these measurements, 22 are continuous, while 19 of the measurements are delayed and have a different measurement frequency, caused by sampling and analysis of the streams. 'Revision of the Tennessee Eastman Process Model' by Bathelt et al. (2015) adds 32 additional measurement points for the process. In the process model, the process measurements have zero-mean white noise applied to them. All the variables previously mentioned, including the process constraints, are shown in Tables 5–7 in Appendix 1.

3.3 Energy efficiency

Like most chemical processes, the Tennessee Eastman process also includes energy-intensive sub-processes. Monitoring and optimizing these energy applications can allow for more economic operation. Even though the TE process has been widely used as a research subject, the energy efficiency aspect is still missing. The energy efficiency in the TE process can be defined as described in the following sections.

The Tennessee Eastman process involves great amounts of energy being transferred. The exothermic reaction releases a lot of energy, which is captured in cooling water. Process modification to allow utilization of this heat would result in higher energy efficiency. A constant steam flow in the stripper and the recycling and pressurization of gaseous components require a significant amount of energy. For optimal energy efficiency, the

energy consumed by the heating of the steam flow and the energy consumed by the compressor are to be minimized while maintaining desired output stream concentrations and keeping the process within its constraints.

In this work, the energy efficiency of the TE process is calculated according to Equation (1) presented in Section 2.1, where specifically the energy used by the steam flow in the stripper and the compressor workload in the recycle flow are considered. It is assumed that the efficiencies of the compressor and reboiler for the stripper steam flow are constant. It can also be assumed that the energy losses in this case are insignificant from the energy efficiency monitoring point of view. In general, due to exothermic reactions and the need for cooling, the TE process may include several heat losses, but these are excluded from this thesis. The amount of product is obtained from the product stream. A minor amount of product loss occurs in the purge stream, approximately 0.5% of the amount obtained in the product stream. Hence, the effect of improving energy efficiency in the TE process by reducing production losses is also negligible and excluded from this work.

In previous work by Jockenhövel et al. (2003), the original TE process model was modified by adding energy balances for the process, as described in Equations (6)–(8).

$$\dot{Q}_r = \dot{m}_{CW,r} c_{p,CW} \left(T_{CW,r,out} - T_{CW,r,in}\right), \tag{6}$$

$$\dot{Q}_c = \dot{m}_{CW,c} c_{p,CW} \left(T_{CW,c,out} - T_{CW,c,in}\right), \tag{7}$$

$$\dot{Q}_{str} = 2258{,}717 \frac{kJ}{kg} \dot{m}_{steam}, \tag{8}$$

where $\dot{m}_{CW,r}$ is the mass flow of the cooling water circuit for the reactor and $\dot{m}_{CW,c}$ is the same for the condenser. The temperatures of the cooling water flow, T_{CW}, are used for calculating the heat flow from the reactor ($\dot{Q}_r$) and condenser ($\dot{Q}_c$). However, the possibility of utilizing this energy is excluded from the process description and only the consumed energy is assessed for energy efficiency monitoring. The saturated steam flow to the stripper ($\dot{m}_{steam}$) is assumed to be condensed completely at a constant temperature and the energy added to the stripper by the steam flow ($\dot{Q}_{str}$) is estimated using Equation (8). The energy added by steam is obtained in respect to the steam flow and the stripper

heat duty in the base case mentioned in Downs and Vogel (1993). (Jockenhövel et al. 2003)

Additionally, the mass amount for the specific product considered must be obtained from the measured product stream volumetric flow. This can be performed using the following Equations (9) and (10):

$$\dot{m}_i = y_i \dot{m}_{11}, \tag{9}$$

$$\dot{m}_{11} = \bar{\rho}_{11} \bar{V}_{11}, \tag{10}$$

where product mass flows $\dot{m}_i$ can be obtained using Equation (9), where y_i is the mass fraction of product i in the product stream with total mass flow $\dot{m}_{11}$. The mass flow of the product stream can be calculated from Equation (10) with the average density $\bar{\rho}$ and the volumetric flow $\bar{V}$ of the product stream.

3.4 Simulation model

'Revision of the Tennessee Eastman Process Model' in Bathelt et al. (2015) provides additional measurements for monitoring the simulation. The added measurements, listed in Tables 8 and 9 of Appendix 1, also include delay- and disturbance-free measurements for process stream concentrations, process conditions inside the reactor and production costs. These additions allow for monitoring and supervision of the simulation and the internal states of the process. The modified process model also removes dependencies on the solver and time increment, increases the speed of the simulation and added outputs for monitoring the process disturbances. (Bathelt et al. 2015)

In addition, the revised model allows the use of variable-step integration methods in simulation environments. In this work, the provided Simulink model was used, available from the Tennessee Eastman Challenge Archive by N.L. Ricker. The process model is written in C programming language and integrated into Simulink. (Ricker 2015)

The Simulink model includes separate model files for the mode 1 and mode 3 operations described by Downs and Vogel (1993). Also included in the simulation package is the self-optimizing control strategy introduced in Larsson et al. (2001). In this work, the mode

1 simulation model was applied. The model requires inputs for the setpoints and disturbances, which are shown in Tables 5 and 11 of Appendix 1. The outputs provided by the model are listed in Tables 6, 8 and 9 of Appendix 1. Disturbance scenarios for the simulator can be controlled with an input vector, with coded values for enabled and disabled disturbances, using Boolean flags. These disturbances can be used for simulating fault scenarios for process diagnosis (Downs and Vogel 1993), but fault detection is beyond the focus of this work.

4 INDIRECT MONITORING METHOD

The objective of this chapter is to propose a general framework for energy efficiency monitoring. The approaches discovered in the literature are synthesized and compiled to form a monitoring framework. The framework is described and portrayed as a flowchart in Section 4.1 (Figure 2). Later, data processing and modelling methods are explored as parts of the framework.

4.1 Framework

Process data must be obtained to develop a data-based energy efficiency estimator. The process data needs to be extensive enough to capture phenomena in different process states to guarantee a sufficiently accurate model. The data may be obtained from historical data – or for more accurate results, experimental process control may be performed to define the accurate operation range for the model (Sargent 2010). Process knowledge is used to find the energy applications within the process. For estimation, an energy efficiency metric needs to be formulated. The metric could be based on describing the ratio between the product obtained and energy used within the process (Tan et al. 2017). Existing energy usage measurements are examined and in the case of insufficient measurements, indirect monitoring is a possibility (Lin et al. 2007).

By finding variables which influence the process energy efficiency, a model can be obtained for estimating energy efficiency from the chosen variables. An energy efficiency index describing the current energy efficiency in the process is estimated from direct or indirect measurements. If the available data is insufficient for estimation, it may be possible to include additional measurements in the process, as indicated in Figure 2.

Further development could enable a predictive model for process energy efficiency, providing a horizon for future operations (see the lower right corner of Figure 2). Energy efficiency prediction can be further used for process control strategy and overall plant energy management. Advanced process control can be used to control the process for automatic energy efficiency optimization. Combining multiple predictive models from different processes can allow optimization of overall energy usage load and reduction in the spikes in power demand.

Process operators can use energy efficiency indicators to evaluate the state of the process and decide on necessary actions. If a lower-than-expected energy efficiency is identified by the operators, they can troubleshoot possible causes for the reduction in efficiency. If the reason for worse efficiency is a failure in the process, the fault might be found early, and in the best case an otherwise inevitable or unexpected process shutdown could be prevented.

Energy efficiency reporting (Figure 2) can be done to record energy operations within a selected time period. Reports might enable the discovery of the sub-processes with the weakest energy efficiency. This can be used for planning maintenance and further process design, for example. In addition, reporting could provide more detailed information for process knowledge, for example by identifying process states and performed control practices with the best energy efficiency.

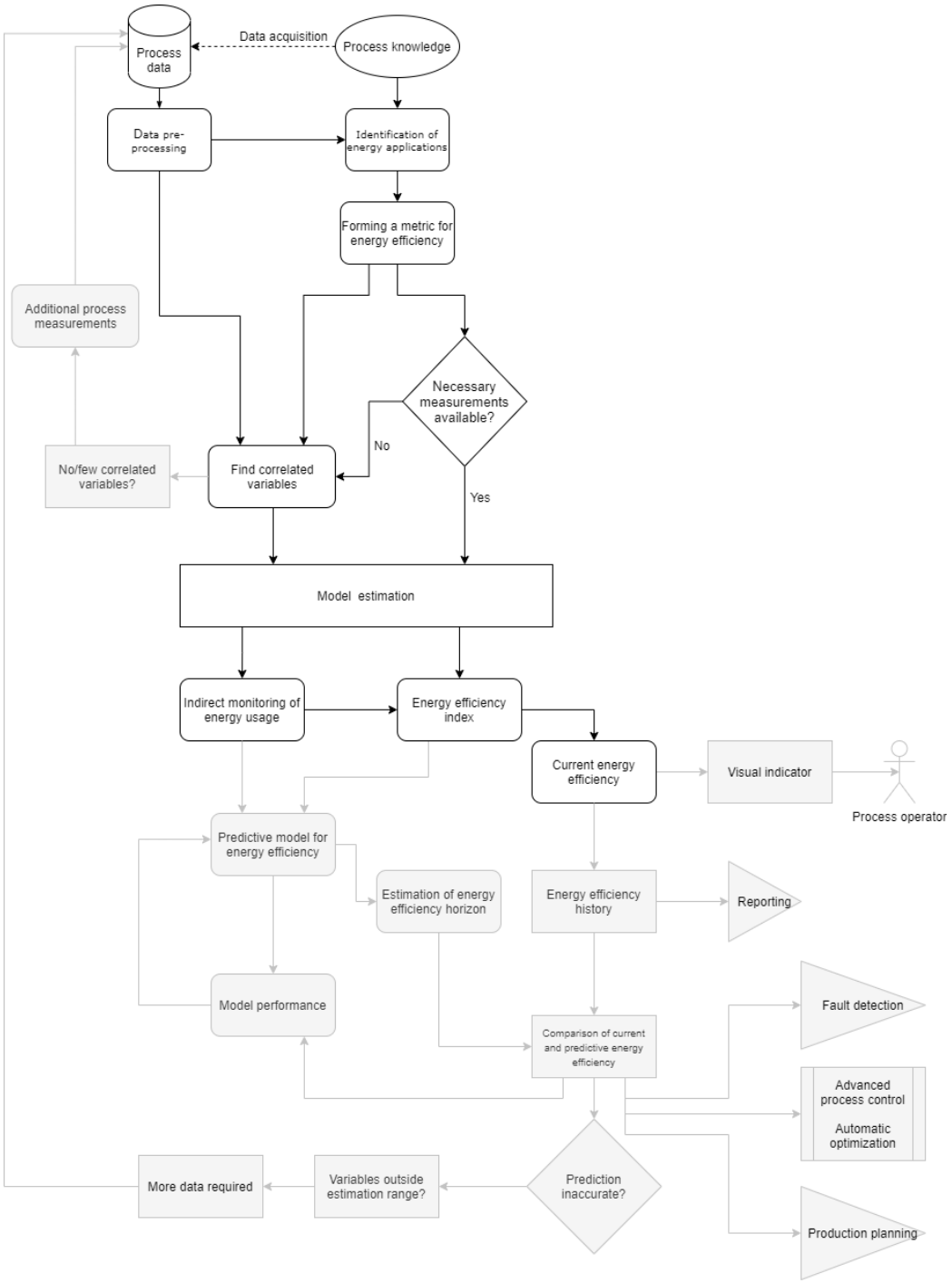

Figure 2. Flowchart of the proposed generalized process framework. Areas explored in this thesis are in the white boxes and areas for possible further research are in the boxes with grey backgrounds.

Finally, the monitoring framework (Figure 2) would need to include methods for assessing the model performance continuously for more accurate energy efficiency estimation. By recognizing the estimation drift, previously unobserved process states can be identified and, with adaptation, accurately captured in the model.

4.2 Modelling

Different data processing and modelling methods must be utilized for the framework. In the following sub-sections, some of the possible methods are explored for data pre-processing, data division, variable selection, model identification and analysis of the model performance.

4.2.1 Data pre-processing

Measurement data is often incomplete, including possible measurement drift, false measurements and noise (Moatar et al. 2001). These inaccuracies can affect the model performance and need to be taken into account. One possibility for reducing the noise in measurement data is to take the average of multiple data points as a new data point. This is called downsampling by block averaging. This can also reduce the computation time of the model selection algorithms as the size of the data matrix is reduced. A significant fraction of the obtained process measurements might be false, resulting in global or contextual outliers. Removal or replacement of these outliers is necessary to avoid modelling false phenomena. (García et al. 2015)

Data can also include delays when comparing the variables with the output target vector. Finding the optimal delays for the model estimation is important. Different measurements can also have different delays, increasing the difficulty in finding the optimal delays for each variable. Delay estimation can be done by finding the amount of delay which best explains the target. This can be done using various methods, for example maximizing correlation or finding the delay which yields the maximum amount of mutual information (Kraskov et al. 2004) between the delayed variable vector and the target vector (Moddemeijer 1989). Chen et al. (2015) described an approach for estimating time delay in an irregularly sampled data set, utilizing a separable nonlinear least-squares method. (Chen et al. 2015; Wang et al. 2020)

4.2.2 Data division

Data must be split into separate, mutually exclusive sets for training and testing (also known as external validation) of the model, which is commonly known as the hold-out method (Kohavi, 1995). Possible pre-processing for the data must be performed before

the data split to avoid including information from the training data in the test data set. The split must be planned out to include sufficient data for both modelling steps while also including a representative variable range for training the model. If the model input space variables drift outside the trained range within the test set, model extrapolation occurs, referred to in the literature as covariate shift (Moreno-Torres et al. 2012). This could result in decreased model performance.

Depending on the training method used in the modelling, training data could be split further. An internal validation set that is drawn from the training set for measuring the model error can be used for model training. This is necessary to reduce the risk of overfitting the model by stopping the training (Ying 2019), or regulating the number of input variables (Baumann 2003; Vuolio et al. 2020). Alternatively, k-fold cross-validation can be used for the model training. In k-fold cross-validation, the input data must be split into k number of folds, whereby the cross-validation is performed with each fold acting as the internal validation set for the training. (Kohavi 1995)

4.2.3 Variable selection

Variable selection must be performed for choosing the model input variables. Unnecessary variables can be omitted from the model. Including irrelevant or redundant variables might result in model overfitting and thus reduce the model performance. Possible collinearities can also be eliminated from the model inputs with variable selection, for example by including a variation inflation factor (VIF) index as a selection criterion (James et al. 2013). Another possibility is to use variable filtering based on the input space correlations (Guyon and Elisseeff 2003). Alternatively, partial least squares (PLS) or principal component regression could be used to eliminate the collinearities, as these use a projected input space in model estimation. (Dormann et al. 2013; Massy 1965; Wold et al. 1984).

There are multiple ways to proceed with the variable selection. The most obvious one is exhaustive search, meaning that all possible combinations are evaluated. The exhaustive search can also be referred to as the best subset selection. However, due to exponential algorithm complexity ($O(2^n)$), where n is the number of input variable candidates, it often appears impractical in real-life applications. (Guyon and Elisseeff 2003)

Manual selection of the variables relies on domain knowledge for selecting the relevant variables. In forward selection, the variables are added one by one to the model and discarded if the cross-validation statistics do not increase. The main disadvantage with the forward selection method is the possibility of staying at a local optimum and ignoring some possible selection combinations. Sorting the variables in order of importance could improve the algorithm. Another method would be backward elimination, which removes variables one at a time and the testing of a chosen criterion. Variations of these methods are also possible, for example adding multiple variables and removing them one by one, keeping the variables which improve the model performance. (Heinze et al. 2018; Vuolio 2021; Xu and Zhang 2001)

Variables can also be selected using a genetic algorithm, which explores possible variable combinations using random crossing and mutation. In genetic algorithms, the solution candidates are presented as a population of individuals, which are crossed with each other and mutated randomly. Genetic algorithms can be effective in avoiding local optima because they explore multiple objective function points in parallel. Due to the random nature of the method, multiple iterations of the selection process might be necessary to find the actual optimal solution. (Jarvis and Goodacre 2005; Vuolio 2021)

4.2.4 Model identification

The model is trained with the chosen variables within the training data set. In the case of multiple input variables, multiple linear regression (MLR) can be used to estimate a single output variable with the input variables, as expressed in the following Equation (11),

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon, \tag{11}$$

where y is the observed output, obtained from the bias β_0 and coefficients $\beta_1, \beta_2 \ldots, \beta_n$ and the residual ε. For the MLR model, the global optimum of the objective function with respect to the parameter space can be found by obtaining the analytical solution of the objective function. (Andrews 1974; Uyanık and Güler 2013)

Principal component analysis (PCA) can be used for developing a principal component regression (PCR) model, which avoids the problem of collinear variables present in the

data by reducing the input data dimensionality with PCA. The method can be constructed as Equation (12),

$$y = \boldsymbol{\gamma}\mathbf{F} + \varepsilon, \tag{12}$$

where $\boldsymbol{\gamma}$ is the vector of coefficients and $\mathbf{F}$ is the constructed scaled principal component matrix, utilizing PCA. However, PCR is dependent on the input data dimensions, requiring the re-estimation of the model if additional variables have to be introduced in the model input space. (Kawano et al. 2018; Massy 1965)

Additionally, neural networks can be utilized for estimating outputs with multiple input signals. The input variable selection and training data for the neural network facilitates a variety of approaches. (Narciso and Martins 2020)

4.2.5 Model performance and validation

The identified model must be evaluated to verify its accuracy and applicability. The model performance is dependent on various factors, for example on its training process; if the model is overfitted for the training data, poor generalization capability can be expected. To assess the model performance realistically, an independent data set needs to be used in which the outputs of the final model are compared to the observed output in the independent test data. The results can be analysed visually and with figures of merit, namely error statistics, some of which are described below. (Vuolio 2021)

In model performance evaluation the residuals are analysed, namely the differences between the modelled and observed values. The residuals can be used for calculating several statistics, including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE) and root mean squared error (RMSE). These values describe the deviation of the estimation from the observed value. The correlation coefficient (R) value measures the linear correlation between the estimated and observed values for the output. The standard deviation of the estimation error describes the variation observed in the residuals, and it can be used to define a confidence interval for the estimated values. (Botchkarev 2018)

5 MODELLING RESULTS AND DISCUSSION

In the following chapter, model changes and simulation are discussed. The simulated process data is generated, and the presented modelling methods are utilized for identifying a model for estimating energy efficiency in the TE process. Finally, the model performance is evaluated, and the results are further discussed.

For this study, the effects of disturbances were not considered, so the disturbance flags of TE model were disabled. Instead, setpoint changes for the selected manipulated variables were used for the simulation scenario. These setpoint changes were necessary for simulating the TE process model in different process states and thus describing the energy efficiency in varying conditions. From the provided simulation package, the model with the base case operation conditions for the process was chosen for the subsequent simulations (mode 1 in Downs and Vogel (1993)).

5.1 Data acquisition

5.1.1 Configuration of simulated process for data generation

In this work, the actual energy efficiency of the TE process was defined according to Equation (1) in Section 2.1 by comparing the amount of energy used by the compressor and steam flow with the amount of final products G and H produced. Energy use by the steam flow is estimated with Equation (8), as explained in Section 3.3, while compressor workload is a measured process variable, including some white noise. The amount of G and H produced can be obtained by calculating the mass flow of the products in the product stream. The simulator does not calculate energy usage and product mass flows by default, and thus the studying of energy efficiency demands additional calculations.

To obtain reference data, namely values for energy efficiency, the following information is required:

- Product stream concentrations,
- Product mass flows,
- Stripper steam flow,
- Amount of energy transferred between steam and stripper, and
- Compressor workload for recycle flow pressurization.

For the product stream concentrations, the values available in the simulation model were used. Product mass flows were calculated by multiplying the mass fractions of product stream (G or H) by the total mass flow of the stream. The mass fractions were directly obtained from the continuous monitoring outputs for molar fractions of the product stream, which were added to the model in Bathelt et al. (2015). The total mass flow was calculated by multiplying the volumetric flow measurement and the average density of the stream (see Equation 10), which was calculated from the molar fraction measurements and liquid densities for the components listed in Table 10 of Appendix 1. The steam mass flow rate is also monitored, and hence is available in the simulation model. Energy added to the stripper by steam is calculated with Equation (8), using the measured steam mass flow and assuming that the steam condenses completely at a constant temperature. The model outputs for product stream composition use the delay- and disturbance-free values added to the model in Bathelt et al. (2015) to create the reference value for the next steps. The energy efficiency estimated with the soft sensor model in this work is, however, calculated from simulated process outputs, some of which include noise and delay.

5.1.2 Setpoints and data ranges

For generating simulated process data, the process setpoints were adjusted to cause variance. Keeping the process in control is necessary in the process, so the setpoint changes needed to be kept reasonable. The base case for data comparison was obtained from the steady-state mode 1 operation described in Downs and Vogel (1993). Inputs and their ranges for the simulation study were chosen using first-order finite difference-based sensitivity analysis (Saltelli et al. 2000) and values reported in the literature (see Table 2). The applied sensitivity index $z_{i,j}$ was calculated as follows:

$$z_{i,j} = \frac{\Delta SEC_j}{\Delta x_i}, \tag{13}$$

where Δx_i is the performed setpoint change for input variable i and ΔSEC_j is the observed change in energy consumption for product j. The energy consumption per amount of product in the product stream, specific energy consumption (SEC), was adapted for the process from the energy efficiency Equation (1), and calculated for each product G and H, then used as criteria for simulation input selection. The specific energy consumption sensitivity to certain setpoint changes was calculated with Equation (13) and presented in Table 1. The variables with the most significant effect on the specific energy consumption were selected to be changed in the simulation scenario. The selected variables were the production rate, stripper level, mole fraction of product G in product stream, mole fractions of A and C in the reactor feed stream and reactor temperature.

Table 1. Variable ranges chosen for setpoint sensitivity analysis and sensitivities of the calculated specific energy consumptions.

Setpoint	Nominal value	Minimum	Maximum	$\lvert z_{i,G}\rvert$ $\lvert z_{i,H}\rvert$ (Maximum) x10000	$\lvert z_{i,G}\rvert$ $\lvert z_{i,H}\rvert$ (Minimum) x10000
Production rate (m^3/h)	22.949	21.802	24.096	76.84 143.4	76.91 124.01
Stripper level (%)	50	25	87.5	33.87 0.23	0.04 0.02
Separator level (%)	50	25	70	1.97 2.60	0.10 0.46
Reactor level (%)	75	71.25	78.75	0.54 9.55	1.12 4.00
Reactor pressure (kPa)	2705	2569.75	2975.5	0.12 0.04	0.03 0.69
Mole fraction G in product	53.724	51.038	80.586	126.39 703.48	196.74 254.78
yA (Mole fraction A in reactor feed)	54.956*	90% of nominal value	120% of nominal value	11.25 16.77	12.22 29.78
yAC (Mole fraction A & C in reactor	32.188+ 26.383	90% of nominal value	110% of nominal value	18.59 28.58	16.82 54.99
Reactor temperature (°C)	120.4	118.59	132.44	27.75 38.79	5.28 12.68
Recycle valve (%)	22.21	0	39.98	8.16 15.23	0.94 3.49
Steam valve position (%)	47.45	0	100	3.43 4.00	5.03 6.27
Agitator setting (%)	50	0	100	0.07 0.25	0.10 0.22

* The molar fraction setpoint of reactor feed stream A used for the simulator model is obtained as follows:

$$\frac{100\cdot(A\ mol-\%)}{(A\ mol-\%)+(C\ mol-\%)} = \frac{100\cdot 32.188}{32.188+26.383}. \quad (14)$$

Data ranges for process simulation were chosen mainly by referencing the values used in the sensitivity analysis for keeping the process in control. The process description in Downs and Vogel (1993) gave examples for production rate and G mole fraction setpoint changes. Tran and Georgakis (2018) used similar ranges for production rate, stripper level and G mole fraction. The chosen reactor temperature range had a similar range to that in Tran and Georgakis (2018), but was adjusted to fit the base case used in this study. The mole fraction of G in the product was chosen within a small range to keep the process within the base case mode. Adjusting the G mole fraction also had the largest impact on specific energy consumption, having a major effect with a small change. The ranges for the chosen setpoints are shown in the following Table 2.

Table 2. Chosen ranges for setpoint changes compared to values suggested in (Downs and Vogel 1993; Tran and Georgakis 2018).

Setpoint	Nominal value	Minimum	Maximum	Range suggested in Tran and Georgakis (2018)	Setpoint changes suggested in Downs and Vogel (1993)
Production rate (m^3/h)	22.949	20.5	24	[20.5 24]	-15% (14228 to 12094 kg/h)
Stripper level (%)	50	40	60	[40 60]	-
Mole fraction G in product (mol %)	53.724	51	57	[0.51 0.57]	50 G/50 H to 40 G/60 H
yA (Mole fraction A in reactor feed)	54.956	90% of nominal value	120% of nominal value	*	-
yAC (Mole fraction A+C in reactor feed)	32.188+26.383	90% of nominal value	110% of nominal value	*	-
Reactor temperature (°C)	120.4	118	125	[119.5 126.5]	-

* Ranges for yA and yAC differ in Tran and Georgakis (2018) compared to the configured simulation model for this study.

The value for the variables regarding reactor input stream composition (yA and yAC) was obtained differently in the simulation model used in this work than that described in Tran

and Georgakis (2018). As such, data ranges for these variables were chosen from the performed sensitivity analysis, where a stable control for the simulator model was found.

5.1.3 Simulated data

A sufficiently long simulation time is required to guarantee the capture of as many phenomena as possible. Because of the slow process dynamics, many hours of simulation were required. Fortunately, simulating 100 hours of the process took approximately 1.25 minutes in real time. The simulation is demonstrated in Figure 3, where measurements for one of the selected variables for the generation of simulated data are presented.

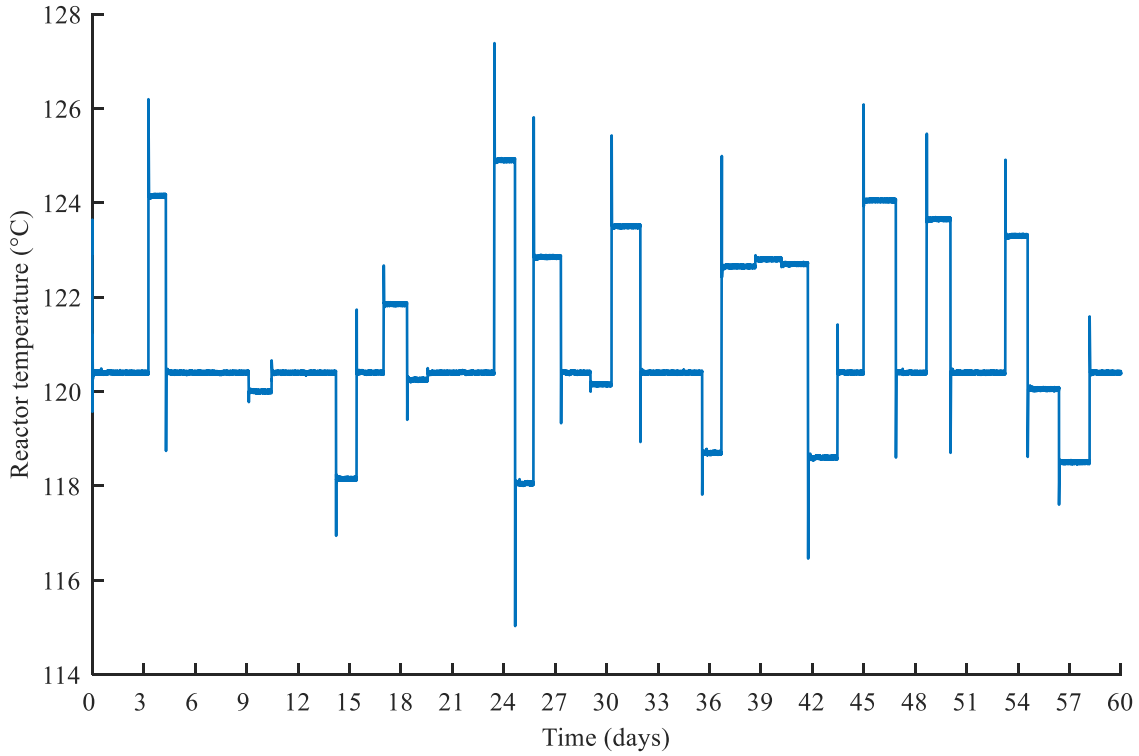

Figure 3. Measured variable 6 (reactor temperature) during the simulation period. Note that the observed variable was one of the controlled variables.

The details for the generation of process data from the TE simulator were formulated as follows. A random number generator seed was set constant for the simulations for repeatability between simulations. To obtain a presentative data set of the process, the total simulation time was chosen to be two months (1440 hours). Setpoint changes were set to take place randomly between every 24 and 48 simulated hours, to simulate routine process operation. The random numbers were drawn from a uniform probability

distribution with the information given in Table 2. According to the slow dynamics, more frequent step changes would not guarantee reaching the steady state in the process before the next step change would be initiated. One to four variables were randomly selected to be adjusted at each step, using uniformly distributed random integers. When performing the next step change, the previously adjusted variables were set back to their nominal values to keep the process within the control range. The chosen variables were then set new random setpoints from an even distribution within the previously chosen control range.

The simulated data was recorded every five seconds in simulation time, and consequently the obtained data points for the measured variables were saved in a matrix of 1036801x41 in size. Specific energy consumptions for products G and H were calculated and saved in a matrix of 1036801x2 in size.

5.2 Data-based model selection

5.2.1 Data pre-processing

Before any modelling was carried out, the process data was pre-processed to remove any abnormalities. The first 1000 data points were removed from the data due to the chaotic transient nature at the start of the simulation, which were present even without any setpoint adjustments. The removed 1000 data points corresponded to 83.33 minutes of simulation time, which was approximately 0.10% of the simulation period. However, because of the synthetic nature of the data, there was no measurement drift, which often takes place in a real process (Pou and Leblond 2019).

The process data has white noise introduced to the measurements, so downsampling (Section 4.2) was carried out to reduce randomness in the data. At the same time the data matrix row dimension was reduced, allowing faster computation in the model selection phase. In this case, downsampling was executed using an average of eight data point sized blocks as the new data point, reducing the size of the measurement data matrix to 129475x41.

5.2.2 Data splitting

Before the first modelling steps, the obtained data was split into training and testing sets using the hold-out method (Section 4.2). Only the training set was used in the model selection, and the test set was left for subsequent analysis of the model performance. In this case, the remaining 30% of the data was left out for testing. Data splitting is demonstrated in Figure 4, where the red vertical line depicts the data split between the training and test sets for measured variable 19.

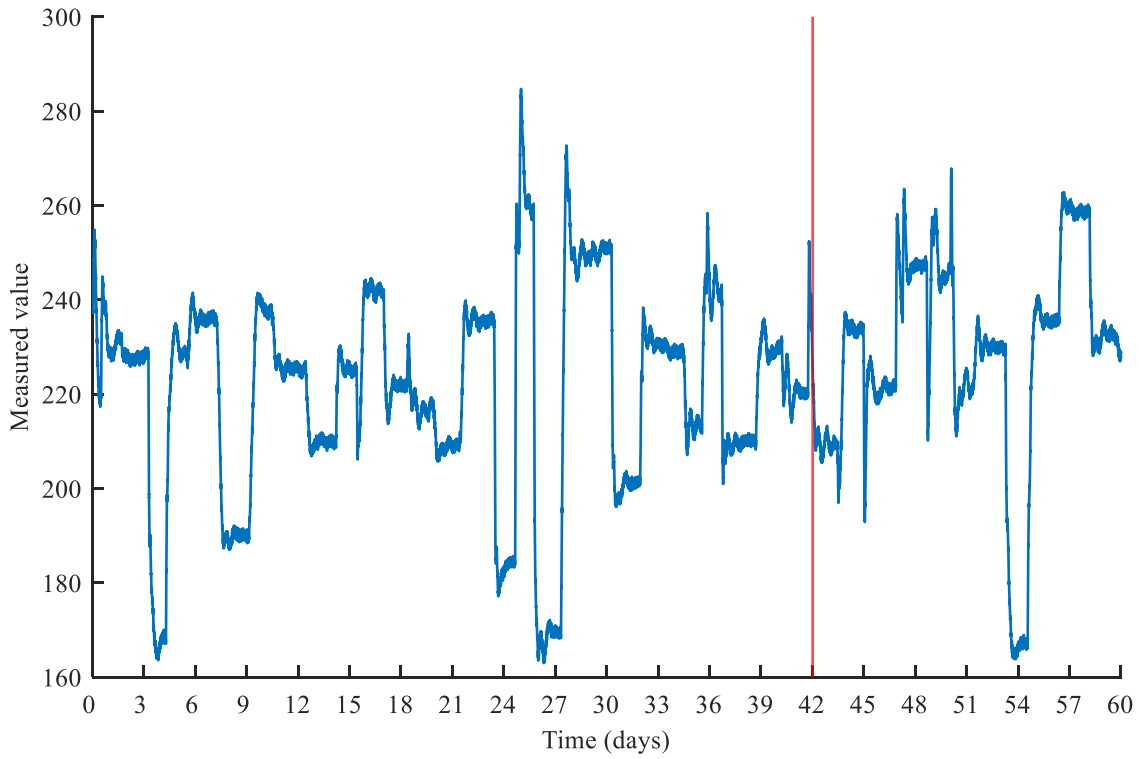

Figure 4. Data division into training and test sets for measured variable 19.

5.2.3 Variable selection and model identification

For model variable selection, the simulated data was delayed by a chosen maximum amount (120 data points were chosen). Increasing the maximum delay amount removes some information from the delayed vector, due to the addition of zero values at the beginning of the vector. In this case, the selected 120 data points correspond to about 1.33 hours of simulation. From the delayed set of vectors for each variable individually, the delay value indicating the highest mutual information (Section 4.2) with the energy efficiency target was coded with Boolean flags to be utilized in the variable selection

algorithm. The optimally delayed data was then split into folds for sequential k-fold cross-validation (Section 4.2) to identify the model parameters. For the cross-validation, 30 folds were utilized.

A forward selection method (Section 4.2) was applied in selection of the model's variables. Mutual information with the training data set output was utilized to rank the variables in descending order before the forward selection. The variables were added to the model, which was trained and evaluated with cross-validation. As an evaluation metric for cross-validation, the mean of the cross-validation errors was used. The selected variables and delays were presented in a binary-coded vector to filter the data matrix **X** before the final model estimation. As shown in Figure 5, the final model outputs and the reference values are compared with the training data set when estimating the energy efficiency of production regarding product G.

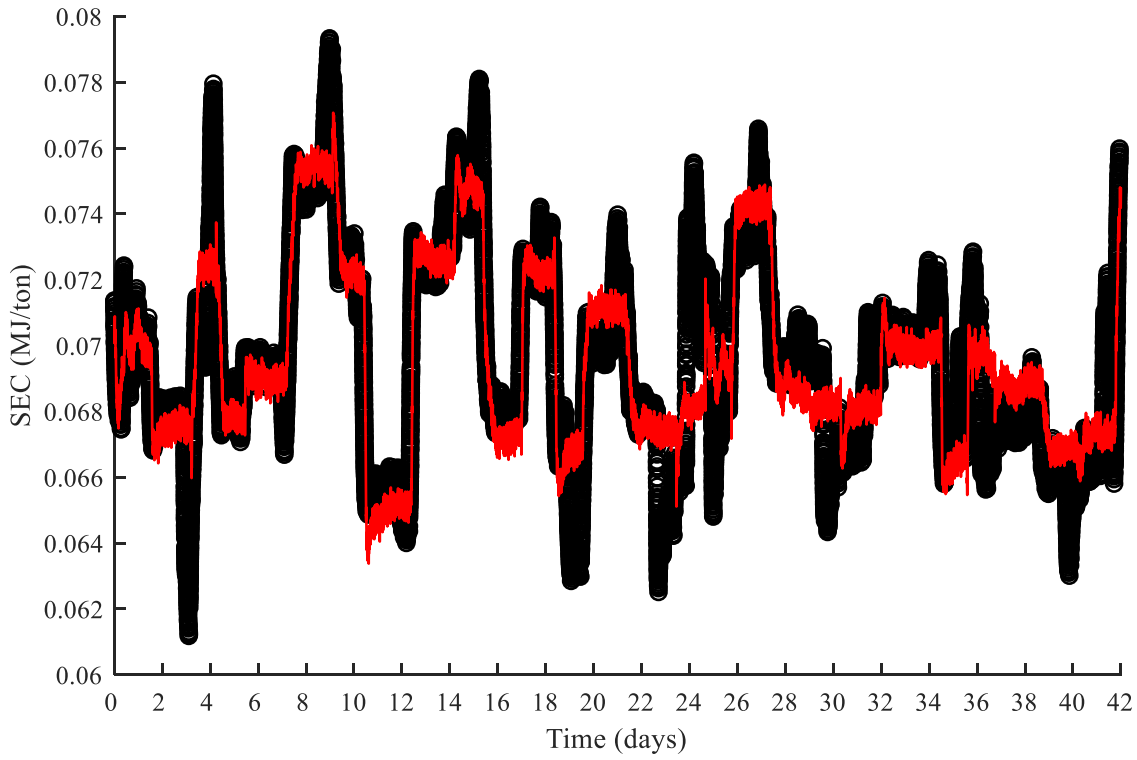

Figure 5. Model outputs in red compared to the reference energy efficiency outputs in black regarding the amount of product G with the training data set.

The obtained final model was then tested with the independent test data. As above for the training data, the model outputs and the reference values are compared with the test data set in Figure 6.

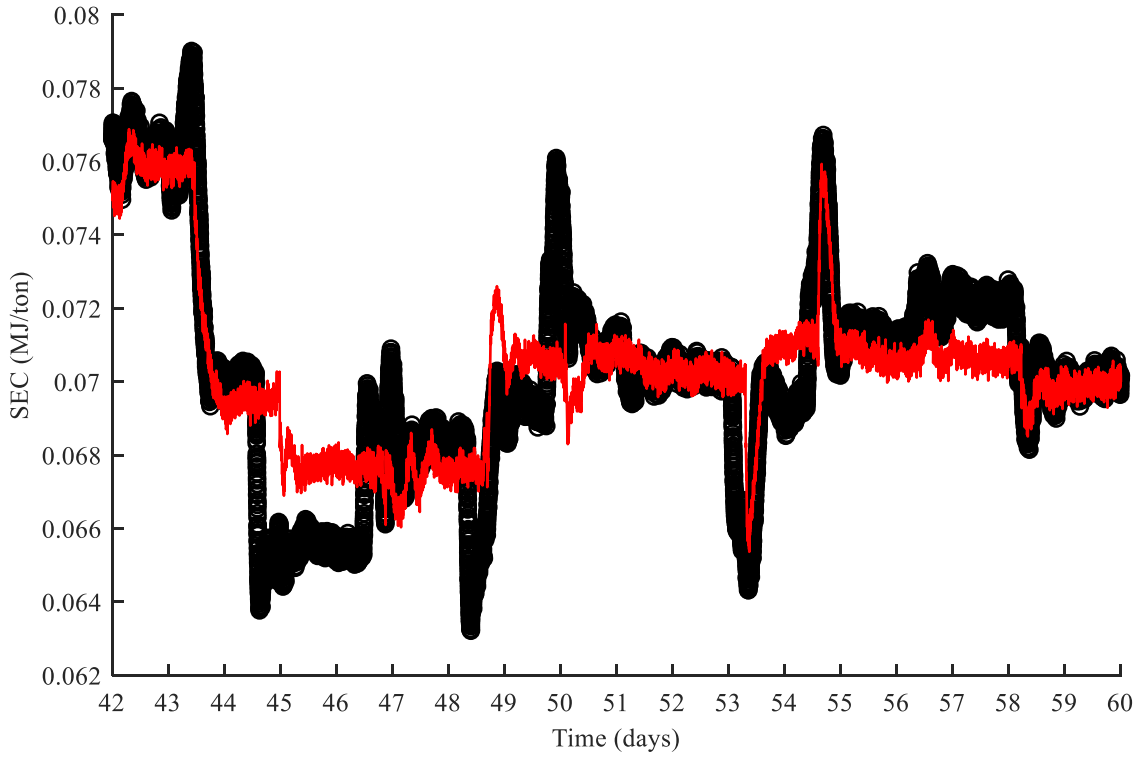

Figure 6. Model outputs in red compared to the observed energy efficiency outputs in black regarding the amount of product G with the test data set.

Next, the procedure was repeated to estimate the energy efficiency in the case of product H. The model outputs are compared to the training set when estimating the production energy efficiency regarding product H. The results are shown in Figure 7.

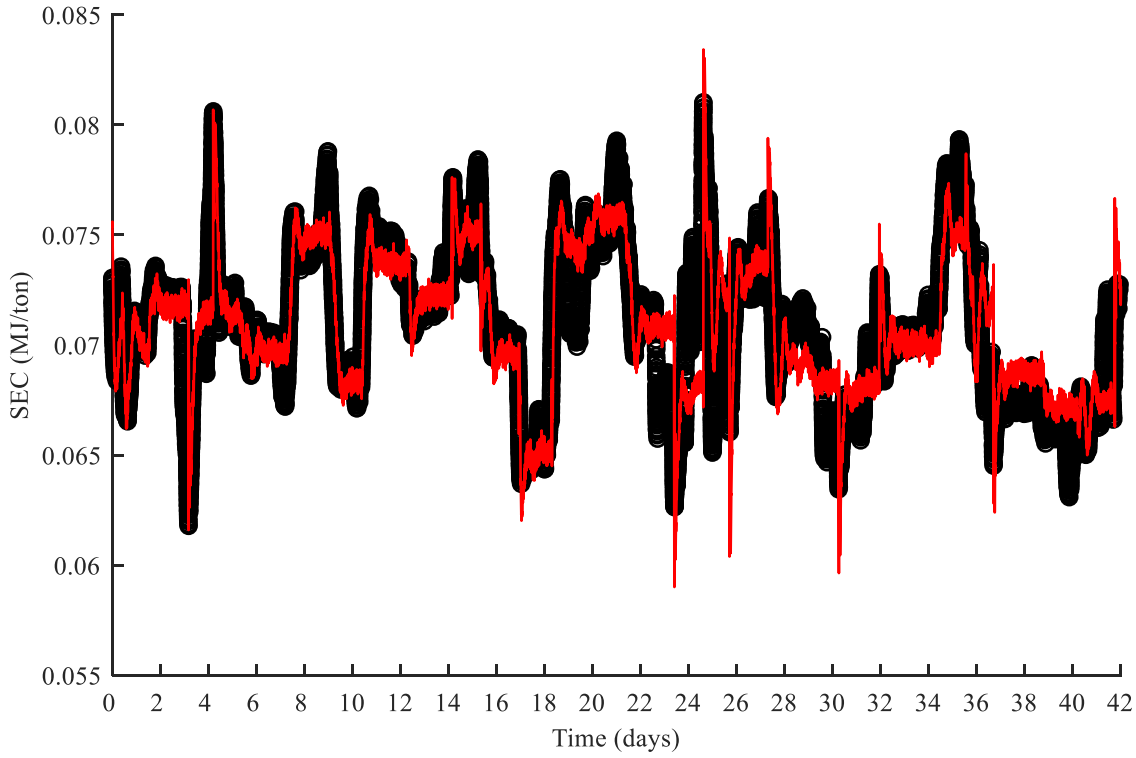

Figure 7. Model outputs in red compared to the observed energy efficiency outputs in black regarding the amount of product H with the training data set.

The obtained model was then tested with the independent test data set. As shown in Figure 8, as in the case with product G, a comparison was performed with the test data set for product H.

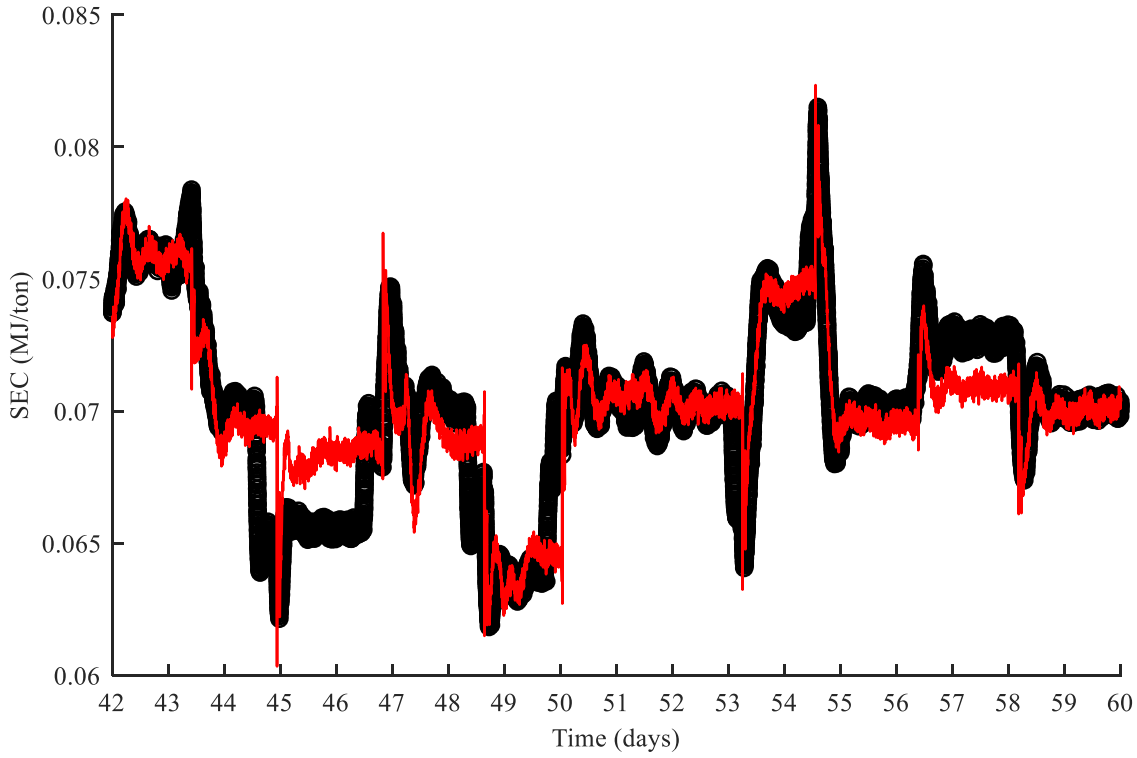

Figure 8. Model outputs in red compared to the observed energy efficiency outputs in black regarding the amount of product H with the test data set.

The variables selected for the models with the forward selection algorithm are shown in Table 3. The identified MLR model for the energy efficiency contains 18 and 21 variables that were selected for product G and product H, respectively. It should be noted that the chosen variables were selected from a set of 41 variables, each with 120 possible delays. From these possibilities, the delay with the highest performance was chosen for each variable individually.

Table 3. Variables selected for the identified models.

	Selected variables
Product G	1 6 8 9 11 12 16 17 19 20 21 23 25 26 32 34 35 36
Product H	1 7 8 10 11 13 14 16 17 20 21 22 23 25 26 27 28 29 31 34 38

5.2.4 Model evaluation

Model performance can be evaluated with a variety of statistics, as previously mentioned in Section 4.2.3. In this case, figures of merit in terms of RMSE, 2σ (two times standard

deviation), MAPE and R-value are presented in Table 4. Similar metrics for both models are observed, and the results suggest that the models have reasonable accuracy.

Table 4. Figures of merit for the modelling results.

Product Data set	G Training	G Test	H Training	H Test
RMSE (kJ/metric ton)	1.5	1.6	1.5	1.6
2σ (kJ/metric ton)	3.0	3.2	3.1	3.2
MAPE (%)	1.6	1.7	1.6	1.6
R	0.87	0.85	0.89	0.89

The calculated 2σ from the training data can be used to calculate a 95.4% confidence interval for the models. The confidence interval for product component G is presented in Figure 9.

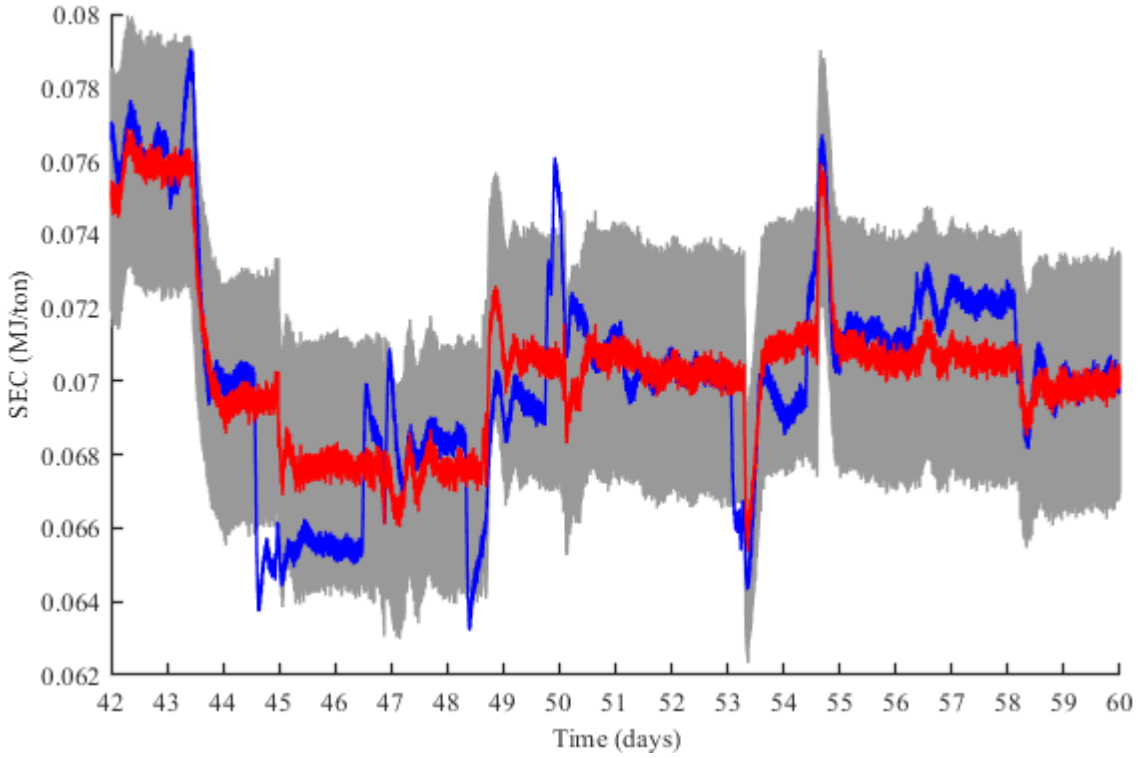

Figure 9. Plot of the 95.4% confidence interval for the energy efficiency estimation for output component G with the test data. The estimation (in red) and confidence interval defined for it (in grey) are compared to the observed output (in blue).

The model outputs are shown in red and compared to the reference outputs in blue. The 95.4% confidence interval for the model is shown in grey. Similarly to the case of product G, the confidence interval for the model regarding product H is presented in Figure 10.

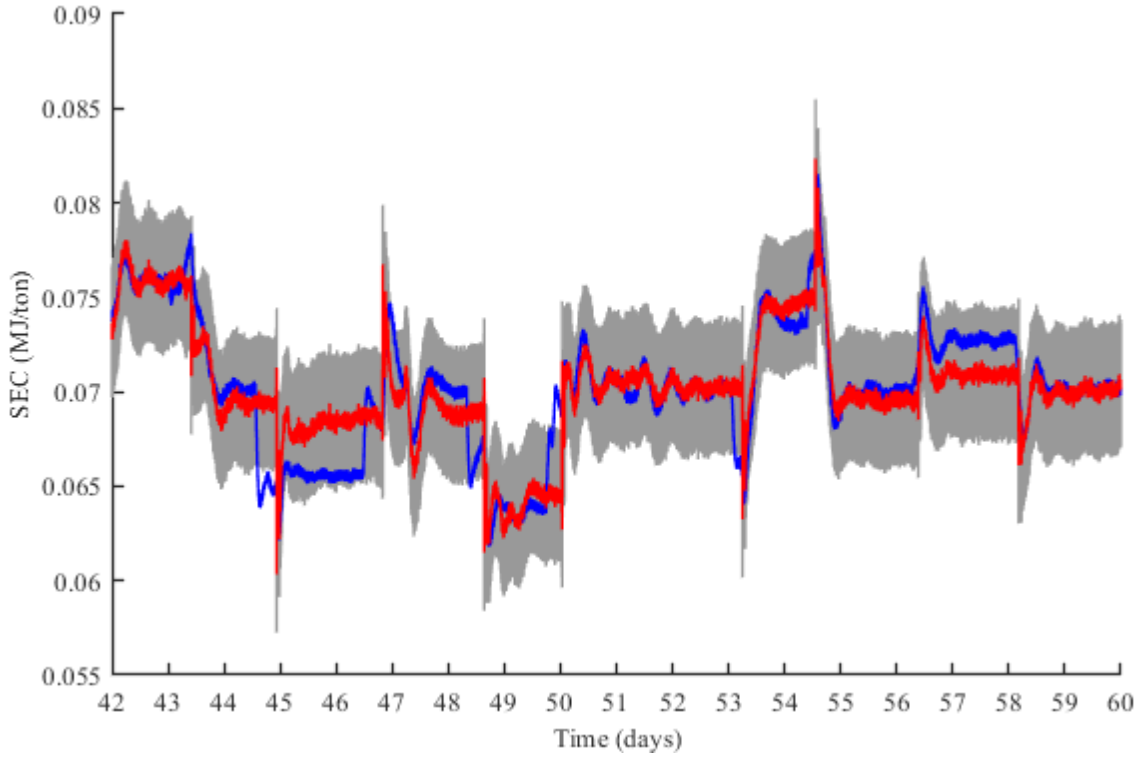

Figure 10. Plot of the 95.4% confidence interval for the energy efficiency estimation for output component H with the test data. The estimation (in red) and confidence interval defined for it (in grey) are compared to the observed output (in blue).

5.3 Discussion

According to the figures of merit shown in Table 4, both models performed sufficiently well, having an R-value of 0.85 and 0.89 for the independent test data set. For the same data, RMSE values of 1.6 kJ/metric ton were obtained for both of the models, which can be considered adequate for the purpose of estimating the energy efficiency for this process. The estimates are within ±3.2 kJ/metric ton of the observed energy efficiency with 95.4% confidence, as demonstrated in Figures 9 and 10 respectively.

However, some areas of inaccuracy can be observed in Figures 6 and 8. In particular, for days 45–47 and 56–58 for both of the product components, G and H. These areas also reside near the limits of the 95.4% confidence interval shown in Figures 9 and 10.

Transient areas of energy efficiency also have some inaccuracies in the estimation, for example during day 47 in the case of product G. Therefore, a more detailed analysis of the model is needed.

Data similarity between the training and test data sets can be compared for example with histogram intersection s (Patacchiola 2016), the Euclidean distance d (Cha 2008) and Kullback–Leibler divergence D_{KL} (Kullback and Leibler 1951). The differences between the training and test data for variable 35 are presented in Figure 11, together with the data similarity metrics.

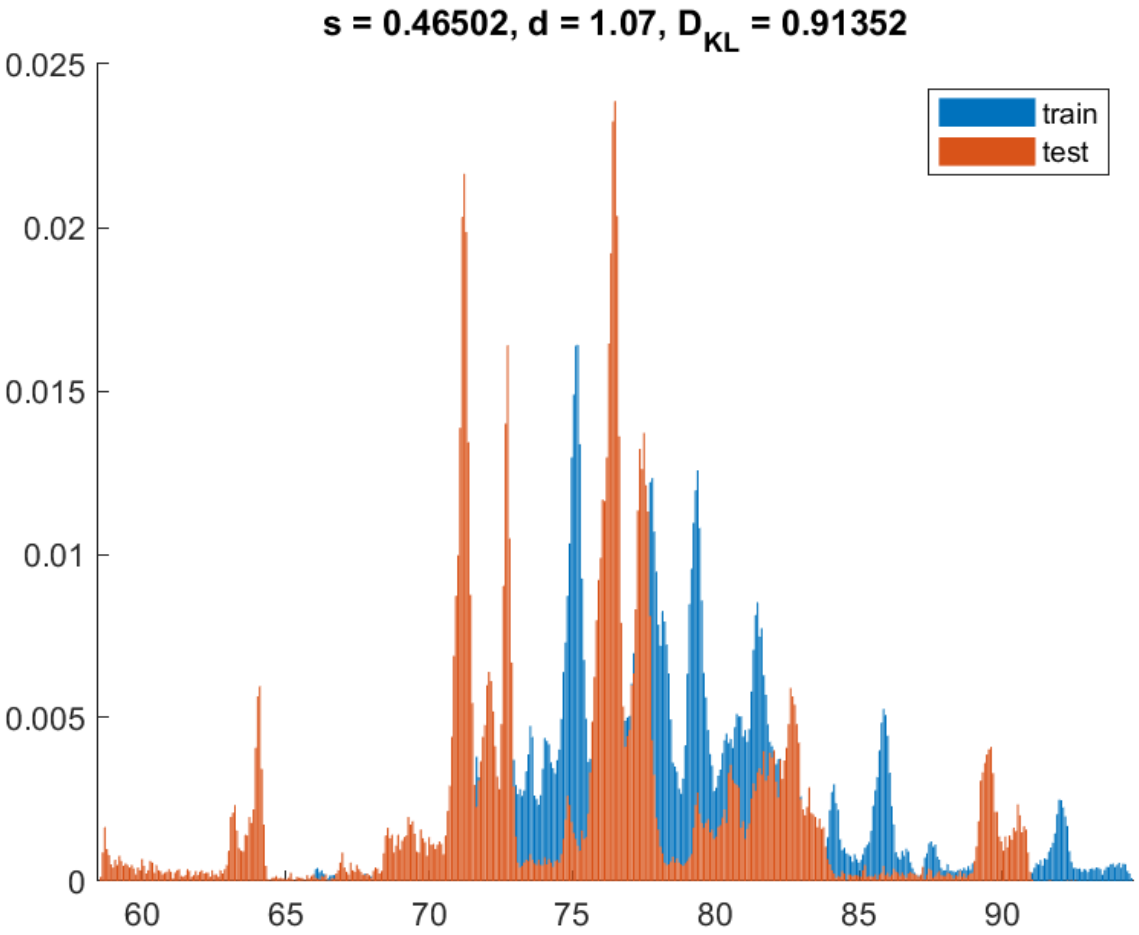

Figure 11. Histogram of the training and test data sets for variable 35.

The histogram intersection compares the areas of the data sets, obtaining a value of one with identical data. The Euclidean distance and Kullback–Leibler divergence gain a value of zero with identical data sets. As seen in Figure 11, with significant difference in the data sets, the histogram intersection has a value of 0.46502, a Euclidean distance of 1.07,

and a Kullback–Leibler divergence of 0.91352. As demonstrated for variable 35, the data differences and data similarity metrics for variable 1 are presented in Figure 12.

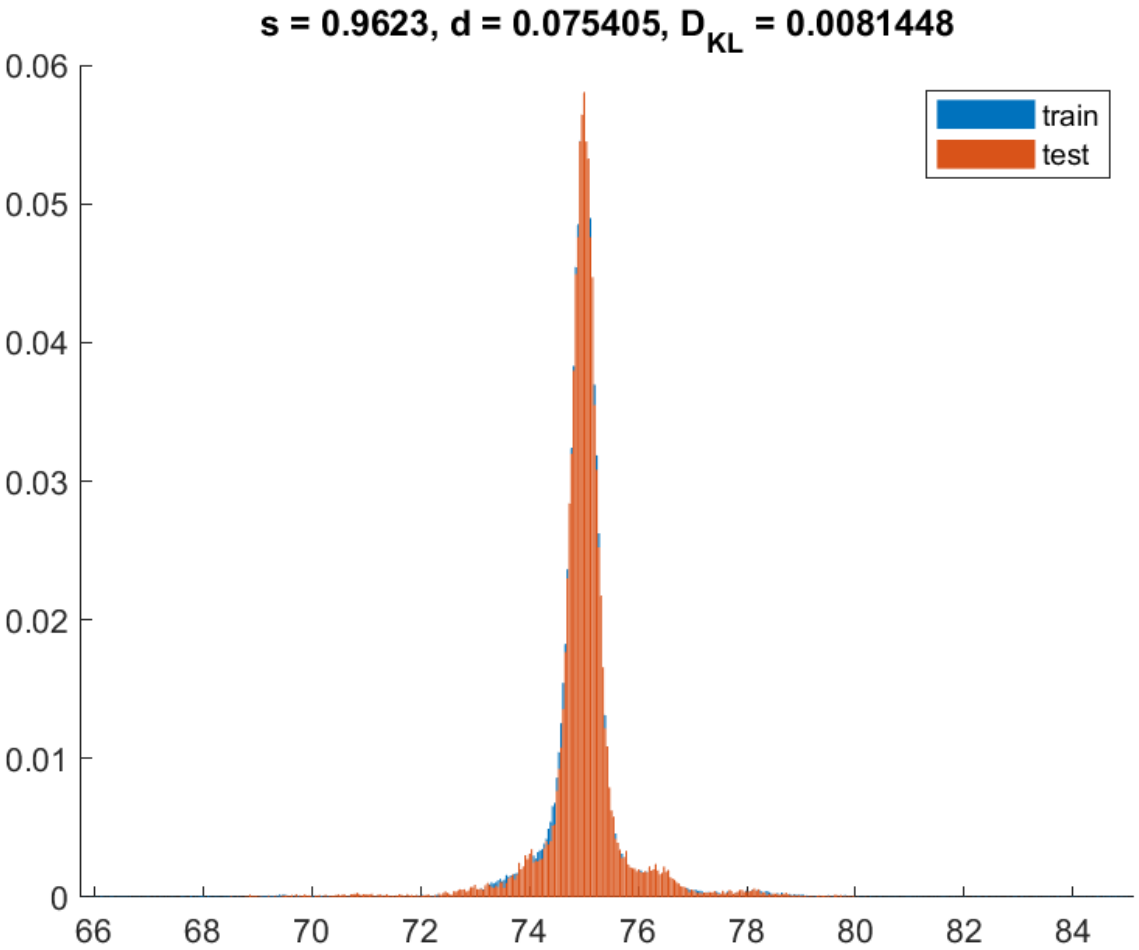

Figure 12. Histogram of the training and test data sets for variable 1.

The data shown in Figure 12 is nearly identical in the training and test data sets, displaying values close to the ideal values for the data similarity metrics.

The data for the modelling was gathered using random setpoint changes rather than with systematic experimental design, with the goal of simulating a normally operating process. Maximum ranges for the variables may not therefore be taken into account in the training data due to the random nature of the setpoint changes. The effect of the performed data split is demonstrated for variable number 20 in Figure 13.

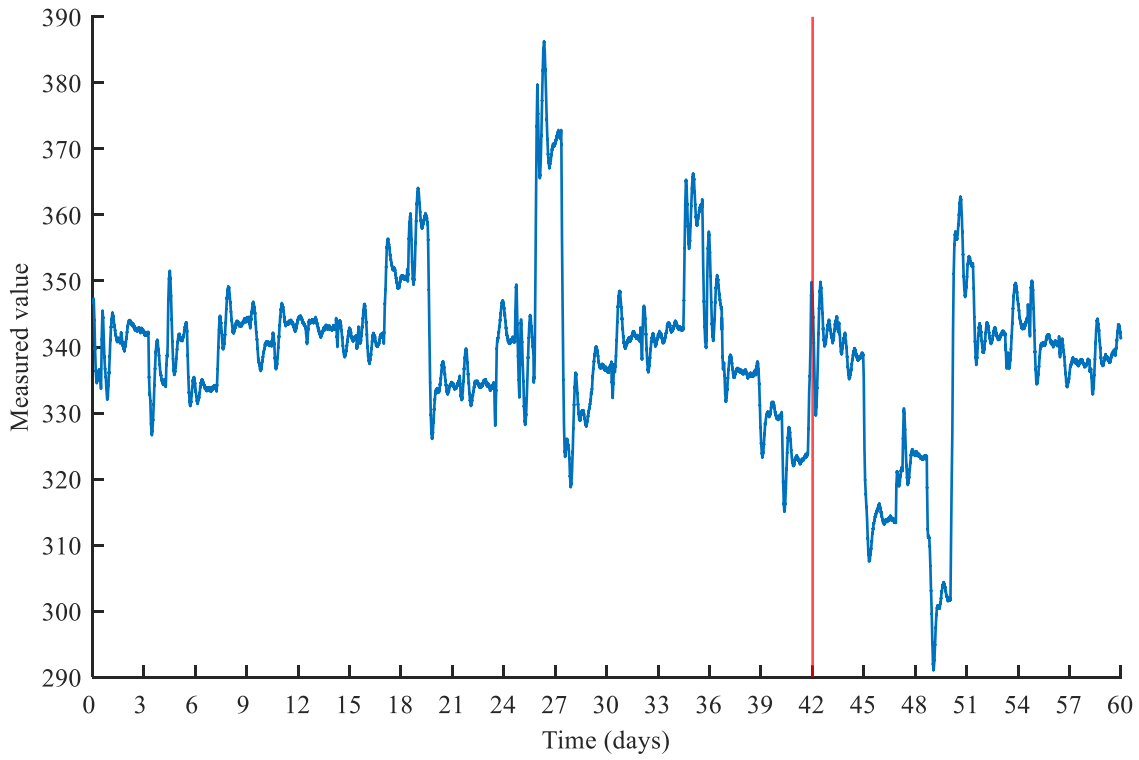

Figure 13. Data distribution for training and testing for variable number 20.

As can be observed in Figures 11 and 13, the test data set includes variables outside the training data range for variables 20 and 35. Similar behaviour for other variables were observed, with some areas of the test data set reaching beyond the range within the training data. Hence the observed inaccuracy can be explained by model extrapolation rather than overfitting of the model.

Discrepancy in the variables can also be observed by monitoring the minimum and maximum values of each variable. For example, in the training data set, the measured variable 20 has a minimum value of 315.1, while in the test data set the minimum value is 291.0. This can also be observed in Figure 13, where the test data set has significant time periods with values below the training data set.

For a slow process, such as the TE process, a large amount of data is required to capture all the possible variable areas for training an accurate estimation model. The gathering of data needs to be continued to improve the accuracy of the model. By planning systematic experimentation for this purpose, most of the phenomena within a process could be captured accurately, but this may be difficult with slow and expensive chemical processes.

As noted in Table 3, the models for products G and H included slightly different selected variables. However, some common variables are present in each model, namely:

- Feed rate,
- Reactor level,
- Product separator temperature,
- Stripper pressure,
- Stripper underflow (product stream),
- Compressor workload,
- Reactor cooling water outlet temperature,
- Mole percentage of components A, C, D in reactor feed, and
- Mole percentage of component F in product stream.

It can then be assumed that the shared variables for the identified models are due to the common factors between the product components, mainly regarding reaction rates and the workload of the compressor and reboiler of the stripper steam flow. The variables not shared between the models can be attributed to the different reaction dynamics for the product components; the reaction for G was mentioned as being more sensitive to temperature due to a higher activation energy (Downs and Vogel 1993).

To develop the energy efficiency metric in the TE process, it was assumed that losses in product, feed and energy were not significant. The heat energy required by the stripper was obtained based on the heat duty mentioned in the original process description, with the assumption that the steam flow fully condenses at a constant temperature. If the assumption turns out to be inaccurate, the energy addition via the steam flow will have been miscalculated. Due to the conservation of energy, less energy is transferred to the stripper if the steam flow only condenses partially. However, if the steam flow does fully condense, but a constant temperature is not kept, more energy is transferred to the stripper.

6 CONCLUSIONS

In this study, as part of the proposed framework, a data-based energy efficiency estimator was created for the Tennessee Eastman process. The process was simulated for 60 days, and the gathered data was used for data-based models. The models identified for energy efficiency regarding two product components of the process performed sufficiently well. The developed model could be improved further by utilizing more advanced methods for variable selection, delay estimation and modelling. Nonlinear estimation and time series models could also be considered in order to obtain higher estimation accuracies. For this thesis, however, a multiple linear regression approach was used for modelling, and forward selection was used for variable selection, starting with the variables ranked highest according to shared mutual information. As shown in Narciso and Martins (2020) (Section 4.2), methods such as affinity propagation, PCA, PCR and PLS can be used to identify relevant data for energy efficiency. However, the observed deficiency in the current data might cause inaccuracies, even with more advanced methods.

Another approach for further development of the estimation accuracy could be clustering. This method can be used to identify and classify different process operation points. Different models can then be developed for the different identified scenarios, allowing more accurate modelling results within each operating point. (Srinivasan et al. 2004)

According to the TE model, energy released in the exothermic reactions is captured as heat in the cooling water circuits. The energy can be monitored by comparing the temperature difference between the input and output cooling streams in the reactor and condenser. Increased energy efficiency could be obtained by using this energy for some process units within the TE process by utilizing common energy integration methods such as Pinch analysis (Kemp 2007). Heating the stripper steam flow with the captured reaction heat may reduce, or even eliminate, the need for added energy to the stripper, improving energy efficiency. In the context of a complete chemical plant, the obtained heat can be transferred to different sections, and excess energy from other parts of the plant may be utilized in the TE process.

The TE process has been widely used in research, mainly for topics such as fault detection and control design. The energy efficiency aspect of the TE process has been disregarded

in previous modelling, although similar accuracies for estimates of the general process states have been observed in related research (Sheta et al. 2019). The complex process proved to be suitable for energy efficiency monitoring with data-based models, adding to the wide variety of study areas the TE process is applicable to. Predictive modelling for energy efficiency and advanced process control are possible approaches for further study regarding this topic. Fault detection by observing decreased energy efficiency is also a potential research topic, as indicated in the suggested framework.

The proposed framework can be considered as a tool for refining energy efficiency operations for other processes. By utilizing the discussed methods, energy usage could thus be monitored and controlled, enabling reduced emissions and process costs. With the addition of predictive models, the estimation of a future time horizon for energy efficiency would be possible, facilitating potential plantwide energy management and production planning. The prediction can be compared with the real-time (soft sensor) estimation of energy efficiency, and deviations from the prediction can be analysed. The predictive model can be updated with new information to further improve prediction accuracy and to reduce the extrapolation need of the model. Possible fault scenarios can also be identified from unexpected behaviour in energy usage.

7 SUMMARY

Current practices for monitoring energy efficiency with indirect methods were examined in this thesis. Energy efficiency can be predicted with data-based models, utilizing direct or indirect measurements. Some of the observed methods were included in the proposed framework for energy efficiency monitoring and could be used as a potential tool for management of energy usage and development of energy efficiency estimators. The explored methods and procedures can be adapted and utilized in different practical applications. As part of the framework, general possibilities for improving the energy efficiency of a process were considered, such as operator decision support with visual indicators, fault diagnosis, advanced process control and production planning.

Parts of the suggested framework were tested with a simulated process, namely the Tennessee Eastman model. This process has been widely used in research as a test case, providing a complex chemical process for a wide variety of fault detection and control practices. The energy efficiency aspect of the Tennessee Eastman process had, however, not previously been widely studied. Minor additions were introduced to the provided Tennessee Eastman model for calculating an energy efficiency metric.

In the practical section, a real-time energy efficiency estimation for the Tennessee Eastman process was obtained through a multiple linear regression model. The model was identified from simulated data, from a period of 60 days. Model performance was then validated with an independent test set, separated from the data at the beginning. The models for estimating the energy efficiency regarding the two final products of the process performed sufficiently well with R values of 0.85 and 0.89, respectively. As the results obtained from the modelling showed, estimation of the energy efficiency within the Tennessee Eastman process with indirect measurements and a data-based approach is feasible.

Real-time estimation of energy efficiency was performed as part of the monitoring framework. However, exploration of some areas of the proposed framework were excluded from this thesis. Therefore, further research in the other areas of the framework, such as predictive modelling of the energy efficiency horizon, advanced process control and fault diagnosis could be beneficial.

REFERENCES

ABB, 2019. Energy Manager [WWW Document]. Available: https://search.abb.com/library/Download.aspx?DocumentID=9AKK107565&LanguageCode=en&DocumentPartId=&Action=Launch [accessed 9.5.2021].

Akdag, S., Yıldırım, H., 2020. Toward a sustainable mitigation approach of energy efficiency to greenhouse gas emissions in the European countries. Heliyon 6, e03396. https://doi.org/10.1016/j.heliyon.2020.e03396

Andrews, D.F., 1974. A Robust Method for Multiple Linear Regression. Technometrics 16, 523–531. https://doi.org/10.1080/00401706.1974.10489233

Bathelt, A., Ricker, N.L., Jelali, M., 2015. Revision of the Tennessee Eastman Process Model. IFAC-Pap. 48, 309–314. https://doi.org/10.1016/j.ifacol.2015.08.199

Bauerdick, C.J.H., Helfert, M., Menz, B., Abele, E., 2017. A Common Software Framework for Energy Data Based Monitoring and Controlling for Machine Power Peak Reduction and Workpiece Quality Improvements. Procedia CIRP 61, 359–364. https://doi.org/10.1016/j.procir.2016.11.226

Baumann, K., 2003. Cross-validation as the objective function for variable-selection techniques. TrAC Trends Anal. Chem. 22, 395–406. https://doi.org/10.1016/S0165-9936(03)00607-1

Botchkarev, A., 2018. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. ArXiv. https://doi.org/10.28945/4184

Cha, S.-H., 2008. Taxonomy of Nominal Type Histogram Distance Measures. MATH 08 6.

Chen, F., Garnier, H., Gilson, M., 2015. Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. J. Process Control 25, 19–27. https://doi.org/10.1016/j.jprocont.2014.10.003

Chen, Y.-Z., Li, Y.-G., Tsoutsanis, E., Newby, M., Zhao, X.-D., 2021. Techno-economic evaluation and optimization of CCGT power Plant: A multi-criteria decision support system. Energy Convers. Manag. 237, 114107. https://doi.org/10.1016/j.enconman.2021.114107

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. Comput. Chem. Eng. 17, 245–255. https://doi.org/10.1016/0098-1354(93)80018-I

Drumm, C., Busch, J., Dietrich, W., Eickmans, J., Jupke, A., 2013. STRUCTese® – Energy efficiency management for the process industry. Chem. Eng. Process. Process Intensif. 67, 99–110. https://doi.org/10.1016/j.cep.2012.09.009

Farsang B., Nemeth S., Abonyi J., 2014. Synergy between data reconciliation and principal component analysis in energy monitoring. Chem. Eng. Trans. 39, 721–726. https://doi.org/10.3303/CET1439121

FINLEX, 2014. FINLEX ® - Ajantasainen lainsäädäntö: Energiatehokkuuslaki 1429/2014 [WWW Document]. Available: https://www.finlex.fi/fi/laki/ajantasa/2014/20141429 [accessed 19.5.2021].

García, S., Luengo, J., Herrera, F., 2015. Data Preprocessing in Data Mining, Intelligent Systems Reference Library. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-10247-4

Geng, Z., Bai, J., Jiang, D., Han, Y., 2018. Energy structure analysis and energy saving of complex chemical industries: A novel fuzzy interpretative structural model. Appl. Therm. Eng. 142, 433–443. https://doi.org/10.1016/j.applthermaleng.2018.07.030

Geng, Z., Chen, J., Han, Y., 2017. Energy Efficiency Prediction Based on PCA-FRBF Model: A Case Study of Ethylene Industries. IEEE Trans. Syst. Man. Cybern. Syst. 47, 1763–1773. https://doi.org/10.1109/TSMC.2016.2523936

Golkarnarenji, G., Naebe, M., Badii, K., Milani, A.S., Jazar, R.N., Khayyam, H., 2018. Support vector regression modelling and optimization of energy consumption in carbon fiber production line. Comput. Chem. Eng. 109, 276–288. https://doi.org/10.1016/j.compchemeng.2017.11.020

Golshan, M., boozarjomehry, R.B., Pishvaie, M.R., 2005. A new approach to real time optimization of the Tennessee Eastman challenge problem. Chem. Eng. J. 112, 33–44. https://doi.org/10.1016/j.cej.2005.06.005

Gontarz, A.M., Hampl, D., Weiss, L., Wegener, K., 2015. Resource Consumption Monitoring in Manufacturing Environments. Procedia CIRP 26, 264–269. https://doi.org/10.1016/j.procir.2014.07.098

Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. J. Mach. Learn. Res. 1157–1182.

Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection – A review and recommendations for the practicing statistician. Biom. J. 60, 431–449. https://doi.org/10.1002/bimj.201700067

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

Jämsä, N., 2018. Model predictive control for the Tennessee Eastman process. Aalto-yliopisto.

Jarvis, R.M., Goodacre, R., 2005. Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. Bioinformatics 21, 860–868. https://doi.org/10.1093/bioinformatics/bti102

Jockenhövel, T., Biegler, L.T., Wächter, A., 2003. Dynamic optimization of the Tennessee Eastman process using the OptControlCentre. Comput. Chem. Eng. 27, 1513–1531. https://doi.org/10.1016/S0098-1354(03)00113-3

Kawano, S., Fujisawa, H., Takada, T., Shiroishi, T., 2018. Sparse principal component regression for generalized linear models. Comput. Stat. Data Anal. 124, 180–196. https://doi.org/10.1016/j.csda.2018.03.008

Kaya, A., Keyes, M.A., 1980. Energy Management Technology in Pulp, Paper and Allied Industries. IFAC Proc. Vol. 13, 609–622. https://doi.org/10.1016/S1474-6670(17)64986-3

Kemp, I.C., 2007. 2 - Key concepts of pinch analysis, in: Kemp, I.C. (Ed.), Pinch Analysis and Process Integration (Second Edition). Butterworth-Heinemann, Oxford, pp. 15–40. https://doi.org/10.1016/B978-075068260-2.50007-9

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Ijcai 14, 1137–1145.

Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Phys. Rev. E 69, 066138. https://doi.org/10.1103/PhysRevE.69.066138

Kullback, S., Leibler, R.A., 1951. On Information and Sufficiency. Ann. Math. Stat. 22, 79–86. https://doi.org/10.1214/aoms/1177729694

Larsson, T., Hestetun, K., Hovland, E., Skogestad, S., 2001. Self-Optimizing Control of a Large-Scale Plant: The Tennessee Eastman Process. Ind. Eng. Chem. Res. 40, 4889–4901. https://doi.org/10.1021/ie000586y

Lin, B., Recke, B., Knudsen, J.K.H., Jørgensen, S.B., 2007. A systematic approach for soft sensor development. Comput. Chem. Eng., ESCAPE-15 31, 419–425. https://doi.org/10.1016/j.compchemeng.2006.05.030

Luan, X., Zhang, S., Li, J., Mendis, G., Zhao, F., Sutherland, J.W., 2018. Trade-off analysis of tool wear, machining quality and energy efficiency of alloy cast iron milling process. Procedia Manuf. 26, 383–393. https://doi.org/10.1016/j.promfg.2018.07.046

Massy, W.F., 1965. Principal Components Regression in Exploratory Statistical Research. J. Am. Stat. Assoc. 60, 234–256. https://doi.org/10.2307/2283149

Moatar, F., Miquel, J., Poirel, A., 2001. A quality-control method for physical and chemical monitoring data. Application to dissolved oxygen levels in the river Loire (France). J. Hydrol. 252, 25–36. https://doi.org/10.1016/S0022-1694(01)00439-5

Moddemeijer, R., 1989. On estimation of entropy and mutual information of continuous distributions. Signal Process. 16, 233–248. https://doi.org/10.1016/0165-1684(89)90132-1

Monjurul Hasan, A.S.M., Trianni, A., 2020. Energy Management: Sustainable Approach Towards Industry 4.0, in: 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Presented at the 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, Singapore, Singapore, pp. 537–541. https://doi.org/10.1109/IEEM45057.2020.9309939

Moreno-Torres, J.G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N.V., Herrera, F., 2012. A unifying view on dataset shift in classification. Pattern Recognit. 45, 521–530. https://doi.org/10.1016/j.patcog.2011.06.019

Narciso, D.A.C., Martins, F.G., 2020. Application of machine learning tools for energy efficiency in industry: A review. Energy Rep. 6, 1181–1199. https://doi.org/10.1016/j.egyr.2020.04.035

Nigitz, T., Gölles, M., Aichernig, C., Schneider, S., Hofbauer, H., Horn, M., 2020. Increased efficiency of dual fluidized bed plants via a novel control strategy. Biomass Bioenergy 141, 105688. https://doi.org/10.1016/j.biombioe.2020.105688

Oh, S.-Y., Yun, S., Kim, J.-K., 2018. Process integration and design for maximizing energy efficiency of a coal-fired power plant integrated with amine-based CO2 capture process. Appl. Energy 216, 311–322. https://doi.org/10.1016/j.apenergy.2018.02.100

Patacchiola, M., 2016. The Simplest Classifier: Histogram Comparison [WWW Document]. Mpatacchiola's Blog. Available: https://mpatacchiola.github.io/blog/2016/11/12/the-simplest-classifier-histogram-intersection.html [accessed 1.7.2021].

Pou, J.-M., Leblond, L., 2019. Evaluate and quantify the drift of a measuring instrument, in: Gazal, S. (Ed.), 19th International Congress of Metrology (CIM2019). Presented at the 19th International Congress of Metrology (CIM2019), EDP Sciences, Paris, France, p. 12004. https://doi.org/10.1051/metrology/201912004

Ricker, N.L., 2015. Tennessee Eastman Challenge Archive [WWW Document]. Available: https://depts.washington.edu/control/LARRY/TE/download.html#Basic_TE_Code [accessed 9.5.2021].

Ricker, N.L., 1993. Model predictive control of a continuous, nonlinear, two-phase reactor. J. Process Control 3, 109–123. https://doi.org/10.1016/0959-1524(93)80006-W

Ricker, N.L., Lee, J.H., 1995a. Nonlinear model predictive control of the Tennessee Eastman challenge process. Comput. Chem. Eng. 19, 961–981. https://doi.org/10.1016/0098-1354(94)00105-W

Ricker, N.L., Lee, J.H., 1995b. Nonlinear modeling and state estimation for the Tennessee Eastman challenge process. Comput. Chem. Eng. 19, 983–1005. https://doi.org/10.1016/0098-1354(94)00113-3

Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity Analysis as an Ingredient of Modeling. Stat. Sci. 15, 377–395.

Sargent, R.G., 2010. Verification and validation of simulation models, in: Proceedings of the 2010 Winter Simulation Conference. Presented at the Proceedings of the 2010 Winter Simulation Conference, pp. 166–183. https://doi.org/10.1109/WSC.2010.5679166

Sheta, A., Braik, M., Al-Hiary, H., 2019. Modeling the Tennessee Eastman chemical process reactor using bio-inspired feedforward neural network (BI-FF-NN). Int. J. Adv. Manuf. Technol. 103, 1359–1380. https://doi.org/10.1007/s00170-019-03621-5

Srinivasan, R., Wang, C., Ho, W.K., Lim, K.W., 2004. Dynamic Principal Component Analysis Based Methodology for Clustering Process States in Agile Chemical Plants. Ind. Eng. Chem. Res. 43, 2123–2139. https://doi.org/10.1021/ie034051r

Tan, Y.S., Ng, Y.T., Low, J.S.C., 2017. Internet-of-Things Enabled Real-time Monitoring of Energy Efficiency on Manufacturing Shop Floors. Procedia CIRP 61, 376–381. https://doi.org/10.1016/j.procir.2016.11.242

Tesch da Silva, F.S., da Costa, C.A., Paredes Crovato, C.D., da Rosa Righi, R., 2020. Looking at energy through the lens of Industry 4.0: A systematic literature review of concerns and challenges. Comput. Ind. Eng. 143, 106426. https://doi.org/10.1016/j.cie.2020.106426

Tran, A.P., Georgakis, C., 2018. On the estimation of high-dimensional surrogate models of steady-state of plant-wide processes characteristics. Comput. Chem. Eng. 116, 56–68. https://doi.org/10.1016/j.compchemeng.2018.02.014

Uyanık, G.K., Güler, N., 2013. A Study on Multiple Linear Regression Analysis. Procedia - Soc. Behav. Sci. 106, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Vuolio, T., 2021. Model-based identification and analysis of hot metal desulphurisation. Oulun Yliop. 148.

Vuolio, T., Visuri, V.-V., Sorsa, A., Ollila, S., Fabritius, T., 2020. Application of a genetic algorithm based model selection algorithm for identification of carbide-based hot metal desulfurization. Appl. Soft Comput. 92, 11. https://doi.org/10.1016/j.asoc.2020.106330

Wang, X., Zuo, J., Zhang, L., Su, J., 2020. System modeling oriented time-delay estimation. ISA Trans. 98, 149–160. https://doi.org/10.1016/j.isatra.2019.08.048

Wold, S., Ruhe, A., Wold, H., Dunn, III, W.J., 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. SIAM J. Sci. Stat. Comput. 5, 735–743. https://doi.org/10.1137/0905052

Xu, L., Zhang, W.-J., 2001. Comparison of different methods for variable selection. Anal. Chim. Acta, 7th International Conference on Chemometrics and Analytical Chemistry Antwerp, Belgium, 16-20 October 2000 446, 475–481. https://doi.org/10.1016/S0003-2670(01)01271-5

Ying, X., 2019. An Overview of Overfitting and its Solutions. J. Phys. Conf. Ser. 1168, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Zhang, X.-H., Zhu, Q.-X., He, Y.-L., Xu, Y., 2018. A novel robust ensemble model integrated extreme learning machine with multi-activation functions for energy modeling and analysis: Application to petrochemical industry. Energy 162, 593–602. https://doi.org/10.1016/j.energy.2018.08.069

Appendix 1. Tables for Tennessee Eastman model.

Table 5. Process manipulated variables (Downs and Vogel 1993).

Variable name	Variable number	Units
D feed flow (stream 2)	1	kg/h
E feed flow (stream 3)	2	kg/h
A feed flow (stream 1)	3	1000 kg/h
A and C feed flow (stream 4)	4	1000 kg/h
Compressor recycle valve	5	%
Purge valve (stream 9)	6	%
Separator pot liquid flow (stream 10)	7	m^3/h
Stripper liquid product flow (stream 11)	8	m^3/h
Stripper steam valve	9	%
Reactor cooling water flow	10	m^3/h
Condenser cooling water flow	11	m^3/h
Agitator speed	12	rpm

Table 6. Process measured variables (Bathelt et al. 2015; Downs and Vogel 1993).

Variable name	Variable number	Units	Frequency	Delay
A feed (stream 1)	1	kscmh	-	-
D feed (stream 2)	2	kg/h	-	-
E feed (stream 3)	3	kg/h	-	-
A and C feed (stream 4)	4	kscmh	-	-
Recycle flow (stream 8)	5	kscmh	-	-
Reactor feed rate (stream 6)	6	kscmh	-	-
Reactor pressure	7	kPa gauge	-	-
Reactor level	8	%	-	-
Reactor temperature	9	°C	-	-
Purge rate (stream 9)	10	kscmh	-	-
Product separator temperature	11	°C	-	-
Product separator level	12	%	-	-
Product separator pressure	13	kPa gauge	-	-
Product separator underflow (stream 10)	14	m^3/h	-	-
Stripper level	15	%	-	-
Stripper pressure	16	kPa gauge	-	-
Stripper underflow (stream 11)	17	m^3/h	-	-
Stripper temperature	18	°C	-	-
Stripper steam flow	19	kg/h	-	-
Compressor workload	20	kW	-	-
Reactor cooling water outlet temperature	21	°C	-	-
Condenser cooling water outlet temperature	22	°C	-	-
Component A in stream 6 (Reactor feed)	23	mol %	0.1 h	0.1 h
Component B in stream 6	24	mol %	0.1 h	0.1 h
Component C in stream 6	25	mol %	0.1 h	0.1 h
Component D in stream 6	26	mol %	0.1 h	0.1 h
Component E in stream 6	27	mol %	0.1 h	0.1 h

Appendix 1. Tables for Tennessee Eastman model.

Component F in stream 6	28	mol %	0.1 h	0.1 h
Component A in stream 9 (Purge)	29	mol %	0.1 h	0.1 h
Component B in stream 9	30	mol %	0.1 h	0.1 h
Component C in stream 9	31	mol %	0.1 h	0.1 h
Component D in stream 9	32	mol %	0.1 h	0.1 h
Component E in stream 9	33	mol %	0.1 h	0.1 h
Component F in stream 9	34	mol %	0.1 h	0.1 h
Component G in stream 9	35	mol %	0.1 h	0.1 h
Component H in stream 9	36	mol %	0.1 h	0.1 h
Component D in stream 11 (Product)	37	mol %	0.25 h	0.25 h
Component E in stream 11	38	mol %	0.25 h	0.25 h
Component F in stream 11	39	mol %	0.25 h	0.25 h
Component G in stream 11	40	mol %	0.25 h	0.25 h
Component H in stream 11	41	mol %	0.25 h	0.25 h
Temperature in stream 1	42	°C	-	-
Temperature in stream 2	43	°C	-	-
Temperature in stream 3	44	°C	-	-
Temperature in stream 4	45	°C	-	-
Reactor cooling water inlet temperature	46	°C	-	-
Reactor cooling water flow	47	m^3/h	-	-
Condenser cooling water inlet temperature	48	°C	-	-
Condenser cooling water flow	49	m^3/h	-	-
Component A in stream 1	50	mol %	-	-
Component B in stream 1	51	mol %	-	-
Component C in stream 1	52	mol %	-	-
Component D in stream 1	53	mol %	-	-
Component E in stream 1	54	mol %	-	-
Component F in stream 1	55	mol %	-	-
Component A in stream 2	56	mol %	-	-
Component B in stream 2	57	mol %	-	-
Component C in stream 2	58	mol %	-	-
Component D in stream 2	59	mol %	-	-
Component E in stream 2	60	mol %	-	-
Component F in stream 2	61	mol %	-	-
Component A in stream 3	62	mol %	-	-
Component B in stream 3	63	mol %	-	-
Component C in stream 3	64	mol %	-	-
Component D in stream 3	65	mol %	-	-
Component E in stream 3	66	mol %	-	-
Component F in stream 3	67	mol %	-	-
Component A in stream 4	68	mol %	-	-
Component B in stream 4	69	mol %	-	-
Component C in stream 4	70	mol %	-	-
Component D in stream 4	71	mol %	-	-
Component E in stream 4	72	mol %	-	-
Component F in stream 4	73	mol %	-	-

Appendix 1. Tables for Tennessee Eastman model.

Table 7. Process constraints (Downs and Vogel 1993).

Process variable	Normal operation		Shut down limit	
	Low limit	High limit	Low limit	High limit
Reactor pressure	-	2895 kPa	-	3000 kPa
Reactor level	50% (11.8 m^3)	100% (21.3 m^3)	2.0 m^3	24.0 m^3
Reactor temperature	-	150°C	-	175°C
Product separator level	30% (3.3 m^3)	100% (9.0 m^3)	1.0 m^3	12.0 m^3
Stripper base level	30% (3.5 m^3)	100% (6.6 m^3)	1.0 m^3	8.0 m^3

Table 8. Process monitoring outputs (Bathelt et al. 2015).

Description	Number	Unit
Substance conversion rate (A)	1	kmol/h
Substance conversion rate (B)	2	kmol/h
Substance conversion rate (C)	3	kmol/h
Substance conversion rate (D)	4	kmol/h
Substance conversion rate (E)	5	kmol/h
Substance conversion rate (F)	6	kmol/h
Substance conversion rate (G)	7	kmol/h
Substance conversion rate (H)	8	kmol/h
Partial pressure of component A	9	kPa abs
Partial pressure of component B	10	kPa abs
Partial pressure of component C	11	kPa abs
Partial pressure of component D	12	kPa abs
Partial pressure of component E	13	kPa abs
Partial pressure of component F	14	kPa abs
Partial pressure of component G	15	kPa abs
Delay-free and disturbance-free measurements of reactor feed analysis (A)	16	mol %
Delay-free and disturbance-free measurements of reactor feed analysis (B)	17	mol %
Delay-free and disturbance-free measurements of reactor feed analysis (C)	18	mol %
Delay-free and disturbance-free measurements of reactor feed analysis (D)	19	mol %
Delay-free and disturbance-free measurements of reactor feed analysis (E)	20	mol %
Delay-free and disturbance-free measurements of reactor feed analysis (F)	21	mol %
Delay-free and disturbance-free measurements of purge gas analysis (A)	22	mol %
Delay-free and disturbance-free measurements of purge gas analysis (B)	23	mol %
Delay-free and disturbance-free measurements of purge gas analysis (C)	24	mol %
Delay-free and disturbance-free measurements of purge gas analysis (D)	25	mol %
Delay-free and disturbance-free measurements of purge gas analysis (E)	26	mol %
Delay-free and disturbance-free measurements of purge gas analysis (F)	27	mol %

Appendix 1. Tables for Tennessee Eastman model.

Delay-free and disturbance-free measurements of purge gas analysis (G)	28	mol %
Delay-free and disturbance-free measurements of purge gas analysis (H)	29	mol %
Delay-free and disturbance-free measurements of product analysis (D)	30	mol %
Delay-free and disturbance-free measurements of product analysis (E)	31	mol %
Delay-free and disturbance-free measurements of product analysis (F)	32	mol %
Delay-free and disturbance-free measurements of product analysis (G)	33	mol %
Delay-free and disturbance-free measurements of product analysis (H)	34	mol %
Delay-free and disturbance-free measurements of feed A analysis (A)	35	mol %
Delay-free and disturbance-free measurements of feed A analysis (B)	36	mol %
Delay-free and disturbance-free measurements of feed A analysis (C)	37	mol %
Delay-free and disturbance-free measurements of feed A analysis (D)	38	mol %
Delay-free and disturbance-free measurements of feed A analysis (E)	39	mol %
Delay-free and disturbance-free measurements of feed A analysis (F)	40	mol %
Delay-free and disturbance-free measurements of feed D analysis (A)	41	mol %
Delay-free and disturbance-free measurements of feed D analysis (B)	42	mol %
Delay-free and disturbance-free measurements of feed D analysis (C)	43	mol %
Delay-free and disturbance-free measurements of feed D analysis (D)	44	mol %
Delay-free and disturbance-free measurements of feed D analysis (E)	45	mol %
Delay-free and disturbance-free measurements of feed D analysis (F)	46	mol %
Delay-free and disturbance-free measurements of feed E analysis (A)	47	mol %
Delay-free and disturbance-free measurements of feed E analysis (B)	48	mol %
Delay-free and disturbance-free measurements of feed E analysis (C)	49	mol %
Delay-free and disturbance-free measurements of feed E analysis (D)	50	mol %
Delay-free and disturbance-free measurements of feed E analysis (E)	51	mol %
Delay-free and disturbance-free measurements of feed E analysis (F)	52	mol %
Delay-free and disturbance-free measurements of feed C analysis (A)	53	mol %
Delay-free and disturbance-free measurements of feed C analysis (B)	54	mol %
Delay-free and disturbance-free measurements of feed C analysis (C)	55	mol %
Delay-free and disturbance-free measurements of feed C analysis (D)	56	mol %
Delay-free and disturbance-free measurements of feed C analysis (E)	57	mol %
Delay-free and disturbance-free measurements of feed C analysis (F)	58	mol %
Production costs related to product amount based on measurements	59	ct/(kmol product)
Production costs related to product amount based on disturbance-free process values	60	ct/(kmol product)
Production costs related to time based on measurements	61	$/h
Production costs related to time based on disturbance-free process values	62	$/h

Appendix 1. Tables for Tennessee Eastman model.

Table 9. Concentrations within process streams (Bathelt et al. 2015).

Description	Number	Unit	Description	Number	Unit
Component A in stream 2	1	mol %	Component A in stream 7	49	mol %
Component B in stream 2	2	mol %	Component B in stream 7	50	mol %
Component C in stream 2	3	mol %	Component C in stream 7	51	mol %
Component D in stream 2	4	mol %	Component D in stream 7	52	mol %
Component E in stream 2	5	mol %	Component E in stream 7	53	mol %
Component F in stream 2	6	mol %	Component F in stream 7	54	mol %
Component G in stream 2	7	mol %	Component G in stream 7	55	mol %
Component H in stream 2	8	mol %	Component H in stream 7	56	mol %
Component A in stream 3	9	mol %	Component A in stream 8	57	mol %
Component B in stream 3	10	mol %	Component B in stream 8	58	mol %
Component C in stream 3	11	mol %	Component C in stream 8	59	mol %
Component D in stream 3	12	mol %	Component D in stream 8	60	mol %
Component E in stream 3	13	mol %	Component E in stream 8	61	mol %
Component F in stream 3	14	mol %	Component F in stream 8	62	mol %
Component G in stream 3	15	mol %	Component G in stream 8	63	mol %
Component H in stream 3	16	mol %	Component H in stream 8	64	mol %
Component A in stream 1	17	mol %	Component A in stream 9	65	mol %
Component B in stream 1	18	mol %	Component B in stream 9	66	mol %
Component C in stream 1	19	mol %	Component C in stream 9	67	mol %
Component D in stream 1	20	mol %	Component D in stream 9	68	mol %
Component E in stream 1	21	mol %	Component E in stream 9	69	mol %
Component F in stream 1	22	mol %	Component F in stream 9	70	mol %
Component G in stream 1	23	mol %	Component G in stream 9	71	mol %
Component H in stream 1	24	mol %	Component H in stream 9	72	mol %
Component A in stream 4	25	mol %	Component A in stream 10	73	mol %
Component B in stream 4	26	mol %	Component B in stream 10	74	mol %
Component C in stream 4	27	mol %	Component C in stream 10	75	mol %
Component D in stream 4	28	mol %	Component D in stream 10	76	mol %
Component E in stream 4	29	mol %	Component E in stream 10	77	mol %
Component F in stream 4	30	mol %	Component F in stream 10	78	mol %
Component G in stream 4	31	mol %	Component G in stream 10	79	mol %
Component H in stream 4	32	mol %	Component H in stream 10	80	mol %
Component A in stream 5	33	mol %	Component A in stripper sump feed	81	mol %
Component B in stream 5	34	mol %	Component B in stripper sump feed	82	mol %
Component C in stream 5	35	mol %	Component C in stripper sump feed	83	mol %
Component D in stream 5	36	mol %	Component D in stripper sump feed	84	mol %
Component E in stream 5	37	mol %	Component E in stripper sump feed	85	mol %
Component F in stream 5	38	mol %	Component F in stripper sump feed	86	mol %
Component G in stream 5	39	mol %	Component G in stripper sump feed	87	mol %
Component H in stream 5	40	mol %	Component H in stripper sump feed	88	mol %
Component A in stream 6	41	mol %	Component A in stream 11	89	mol %
Component B in stream 6	42	mol %	Component B in stream 11	90	mol %
Component C in stream 6	43	mol %	Component C in stream 11	91	mol %
Component D in stream 6	44	mol %	Component D in stream 11	92	mol %
Component E in stream 6	45	mol %	Component E in stream 11	93	mol %
Component F in stream 6	46	mol %	Component F in stream 11	94	mol %
Component G in stream 6	47	mol %	Component G in stream 11	95	mol %
Component H in stream 6	48	mol %	Component H in stream 11	96	mol %

Table 10. Process component properties (Downs and Vogel 1993).

Component	Molecular weight	Liquid density (kg/m³)	Liquid heat capacity (kJ/(kg °C))	Vapour heat capacity (kJ/(kg °C))	Heat of vaporization (kJ/kg)
A	2	-	-	14.6	-
B	25.4	-	-	2.04	-
C	28	-	-	1.05	-
D	32	299	7.66	1.85	202
E	46	365	4.17	1.87	372
F	48	328	4.45	2.02	372
G	62	612	2.55	0.712	523
H	76	617	2.45	0.628	486

Table 11. Simulator model disturbances (Bathelt et al. 2015; Downs and Vogel 1993; Ricker 2015).

Number	Type	Disturbed value
1	Step	A/C ratio of stream 4, B composition constant
2	Step	B composition of stream 4, A/C ratio constant
3	Step	D feed (stream 2) temperature
4	Step	Cooling water inlet temperature of reactor
5	Step	Cooling water inlet temperature of separator
6	Step	A feed loss (stream 1)
7	Step	C header pressure loss (stream 4)
8	Random	A/B/C composition of stream 4
9	Random	D feed (stream 2) temperature
10	Random	C feed (stream 4) temperature
11	Random	Cooling water inlet temperature of reactor
12	Random	Cooling water inlet temperature of separator
13	Drift	Reaction kinetics
14	Stiction	Cooling water outlet valve of reactor
15	Stiction	Cooling water outlet valve of separator
16	Random	(unknown); deviations of heat transfer within stripper (heat exchanger)
17	Random	(unknown); deviations of heat transfer within reactor
18	Random	(unknown); deviations of heat transfer within condenser
19	Stiction	(unknown); recycle valve of compressor, underflow separator (stream 10), underflow stripper (stream 11) and steam valve stripper
20	Random	(unknown)

Appendix 1. Tables for Tennessee Eastman model.

21	Random	A feed (stream 1) temperature
22	Random	E feed (stream 3) temperature
23	Random	A feed (stream 1) pressure (= flow)
24	Random	D feed (stream 2) pressure (= flow)
25	Random	E feed (stream 3) pressure (= flow)
26	Random	A & C feed (stream 4) pressure (= flow)
27	Random	Cooling water pressure (= flow) of reactor
28	Random	Cooling water pressure (= flow) of condenser