Investigation of Global Sensitivity Analysis on Operating Revenue and CO$_2$ Emissions from Crude Oil Distillation using Two-Stage Uncertainty Quantification Method

Le Quang Minh, Pham Luu Trung Duong, Moonyong Lee*

School of Chemical Engineering, Yeungnam University, Gyeongsan, Republic of Korea 712-749
mynlee@yu.ac.kr

This paper presents a global sensitivity analysis based strategy for uncertainty quantification of the operating revenue and the carbon dioxide emissions in a crude oil distillation column. Aspen HYSYS is used for rigorous simulation of the column. A two-stage approach, which is executed by an interface between MATLAB and HYSYS, is implemented in this study. A multiplicative dimensional reduction method is applied to first identify which factors have the most influential to the model outputs, e.g. the operating revenue and the CO2 emissions. In the second stage, the uncertainty quantification regressed by Gaussian prediction regression method is exploited. As a result, there is a good agreement between the predicted results by the Gaussian prediction regression method with that of the conventional Quasi-Monte Carlo approach, which shows that the computational efforts was reduced significantly compared to the conventional study (about less than 10 times). Furthermore, the Kernel points estimated by the Gaussian prediction regression method are generated to highlight influential factors identified in the first stage.

1. Introduction

A crude distillation unit (CDU) is among the largest energy-consuming processes owing to the large amount of energy consumption and high operating temperature (Kim, 2017). Previous studies recommended the employment of rigorous, simplified, and statistical/empirical models for the design and optimization of crude distillation units (Enríquez-Gutiérrez et al., 2015; More et al., 2010; Ochoa-Estopier et al., 2014). Nevertheless, uncertainties are sometimes correlated in the process design because of the connected process and various random factors. The complex nature of CDUs, including their interactions with the associated heat recovery network and the large number of degrees of freedom, renders their optimization a highly challenging task.

The goal of an uncertainty quantification (UQ) is to assess the effect of the input uncertainties on the model output and consequently, on the system performance (Doostan & Owhadi, 2011). For considering the unknown factors, a global sensitivity analysis (SA), which is a widely used technique for selecting representative parameters from mathematical models (Saltelli, 2002; Sobol', 2001), is required to reduce their quantity and select the ones to be incorporated. Monte Carlo (Webster et al., 1996) and Quasi Monte Carlo (QMC) methods are representative probabilistic approaches for the propagation of uncertainties in the model inputs to its output (Abubakar et al., 2015). However, notwithstanding the simplicity of their implementation, the mean convergence rate was estimated at high number of simulations; this renders MC/QMC-based approaches computationally expensive, and they are used only as the final option. To overcome the problem, a convenient-to-evaluate function may be used to replace the original computational model (Schobi et al., 2015). Duong et al. (2016) studied the problem of UQ/SA of chemical processes using the standard polynomial chaos for systems with a small number of random inputs. Using an efficient-to-evaluate surrogate model, Celse et al.(2015) studied the influence of each input on the construction of the model of the oil hydro-treatment process. Furthermore, to address a system with a moderate/large number of uncertain parameters, Duong et al. (2016) and Minh et al. (2017) recently developed a compressed sensing polynomial chaos technology and an efficient two-stage polynomial chaos approach to overcome the

computational limitation. In terms of our understanding, there is no report of the global SA that includes a discussion on the uncertain effects of process inputs on the process performance for the conventional CDU with large number of uncertainties. Thus, this work is expected to describe the uncertain effects of the process inputs on the process interests.

2. Methodologies

2.1 Multiplicative dimensional reduction method (MDRM)

The readers could find the main content of this method in the literature (Minh et al., 2017).

2.2 Gaussian prediction regression (GPR)

We assume that m (< n) important inputs ($\xi_m = \left(\xi_{i_1}, ..., \xi_{i_m} \right)$) were detected in the previous step. Let us

make p simulations and collect the following data set $\left\{ \xi_m^{(i)}, y^{(i)} = M(\xi_m^{(i)}) \right\}_{i=1}^{p}$.

A GPR model is a generalization of the multivariate Gaussian random variables to an infinite dimension. It is defined as

$$y(\xi_m) = h(\xi_m)^T \beta + f(\xi_m) \tag{1}$$

where $f(\xi_m)$ is a zero mean Gaussian random process with the covariance function $k(\cdot, \cdot)$, and $h(\xi_m)$

is a set of basic functions or a trend which transforms the original input space into a new feature vector $h(\xi_m)$

in R^m, a vector of basis coefficients. That is, the GPR is a statistical model that exhibits an instant response of the form

$$P(y^{(i)} \mid f(\xi_m^{(i)}), \xi_m^{(i)}) \approx N(y^{(i)} \mid h(\xi_m^{(i)})^T \beta + f(\xi_m^{(i)}), \sigma^2) \tag{2}$$

where $\mathcal{N}\,()$ is a normal distribution in which the polynomial functions of the inputs can be normally used as the basis functions. It is noted that the latent function, f , is introduced for each observation. Various covariance functions are described in the literature (Kocijan, 2016; Rasmussen & Williams, 2005). In this study, the automatic relevance determination (ARD) squared exponential is used, which is defined as follows:

$$k(\xi_m^{(i)}, \xi_m^{(j)} \mid \boldsymbol{\theta}) = \theta_f \exp(-\frac{1}{2} \sum_{l=1}^{m} (\xi_l^{(i)} - \xi_l^{(j)})^2 / \theta_l) \tag{3}$$

where $\boldsymbol{\theta} \in R^{m+1} = (\theta_f, \theta_1, ..., \theta_m)$ is a vector of hyper-parameters. It is expected that similar values of inputs should exhibit similar responses. Thus, the covariance function can be used to characterize the similarity between two input vectors $\xi_m^{(i)}, \xi_m^{(j)}$.

Therefore, the GPR model is defined by the coefficient vectors $\boldsymbol{\beta}$, hyper parameters $\boldsymbol{\theta}$, and measurement noise variance, σ . The values for these parameters will be updated and learnt from the given data by maximizing the logarithm likelihood expressed in Eq. (15):

$$\log P(y \mid \xi_m, \beta, \theta, \sigma)$$
$$= \frac{1}{2}(y - H\beta)^T \left[K(\xi_m, \xi_m' \mid \theta) + \sigma^2 I_p \right]^{-1} (y - H\beta) - \frac{p}{2} \log(2\pi) - \frac{1}{2} \log \left| K(\xi_m, \xi_m' \mid \theta) + \sigma^2 I_p \right| \tag{4}$$

where I_p is an identity matrix of size p, and

$$X = \begin{pmatrix} (\xi_m^{(1)})^T \\ (\xi_m^{(2)})^T \\ ... \\ (\xi_m^{(P)})^T \end{pmatrix}, \ H = \begin{pmatrix} h(\xi_m^{(1)})^T \\ h(\xi_m^{(2)})^T \\ ... \\ h(\xi_m^{(P)})^T \end{pmatrix}$$

$$\mathbf{y} = \begin{pmatrix} y^{(1)} \\ ... \\ y^{(P)} \end{pmatrix}$$

$$K(\xi_m, \xi_m) = \begin{bmatrix} k(\xi_m^{(1)}, \xi_m^{(1)}) & ... & k(\xi_m^{(1)}, \xi_m^{(P)}) \\ ... & ... & ... \\ k(\xi_m^{(P)}, \xi_m^{(1)}) & ... & k(\mathbf{x}^{(P)}, \xi_m^{(P)}) \end{bmatrix}$$

(5)

where $H_{p \times r}$ is a matrix of values of the basic vector at training points, r is the number of basic function, and K is the covariance matrix at the training points.

Regarding the estimated optimal parameters $\widehat{\beta}, \widehat{\theta}, \widehat{\sigma}$, which are obtained by the maximization of the logarithm likelihood function, the density function of the response at a new data point ξ_m^* is

$$N(h(\xi_m^*)\widehat{\beta} + \mu, \widehat{\sigma}^2 + \Sigma)$$

(6)

where

$$\mu = K(\xi_m^{*T}, \xi_m \mid \widehat{\theta})(K(\xi_m, \xi_m \mid \widehat{\theta}) + \widehat{\sigma} I_p)^{-1}(y - H\beta)$$

(7)

$$K(\xi_m^{*T}, \xi_m \mid \widehat{\theta}) = \left(k(\xi_m^*, \xi_m^{(1)}), ..., k(\xi_m^*, \xi_m^{(n)}) \right)$$

and

$$\mathring{a} = k(\xi_m^*, \xi_m^*)$$
$$- K(\xi_m^{*T}, \xi_m \mid \widehat{\theta})(K(\xi_m, \xi_m \mid \widehat{\theta}) + \widehat{\sigma} I_p)^{-1} K(\xi_m, x^{*T} \mid \widehat{\theta})$$

(8)

The mean value calculated in Eq(8) is used as the prediction for the new data, whereas the variance Σ is an indicator of the uncertainty of the prediction. One can use Eq(7) as a surrogate to sample large numbers of output responses and to estimate density functions and statistics of the output interest.

3. Case study - crude distillation unit (CDU)

Figure 1 demonstrates a rigorous CDU model as a substitute for the actual plant (HYSYS, 2007). The tower consists 29 stages in the main shell and three side columns associated with three pump-arounds for stripping the middle products. Furthermore, to separate the light components from the liquid phase, a pre-flash separator is used to enhance the separation efficiency of the tower (Errico et al., 2009). After leaving the pre-flash separator, the crude oil enters the CDU furnace (see Figure 1). A significant amount of the main stripping steam is included at the bottom, which serves to strip any residue and prevent the excessive thermal cracking of crude oil because of the high temperature at the flash zone. A heat source (Q-trim) is introduced on this tray to "trim" the feed temperature as needed for the initial solution trials and facilitate a converged solution. Other light fractions of the crude were retrieved by side strippers at different locations using steam or reboilers. The presence of pump-arounds is an additional feature in the CDU. These units permit the reduction of the condenser cooling requirements by reducing the vapor flow in the column and allowing for heat recovery. The residue from the column contains valuable hydrocarbons, which is typically further separated in a vacuum distillation column (not discussed in this study).
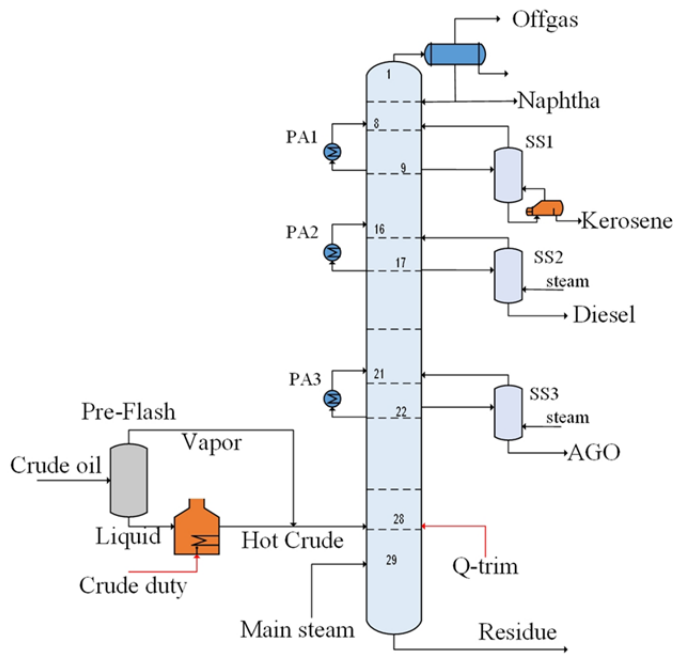
Figure 1: Process flow diagram of CDU.

4. Results and discussion

According to the MDRM, a set of 160 Gaussian quadrature nodes was generated from the orthogonal polynomial toolbox (Gautschi, 2004) passed to HYSYS. A list of sixteen variables could be found in Minh et al. (2018). The operating revenue (Osuolale & Zhang, 2017) and the CO_2 emissions (Gadalla et al., 2006) were calculated as the outputs in HYSYS spreadsheet cells and used for the global SA. The Sobol's sensitivity indices were derived in step 2.1, which indicates the influential factors for the two outputs. For the operating revenue, three random variables such as the **flow rates of the crude oil, naphtha, and diesel products** were detected as being critical. Other random variables become non-influential factors that can be omitted in the next stage. The probability of the CO_2 emissions is affected by the variation of the **flow rates of the diesel steam, AGO steam, and the main stripping steam**.

The effective detection of the non-influential inputs from the SA step enables one to simplify the model. Two simulation sets of size 500 were considered for training the GPR (Clifton et al., 2012) for the output. Because precise estimates of the process outputs are not accessible, the results of the proposed method were compared with those of the QMC method with a sufficient number of simulations. For an accurate estimation of the probability for the QMC method, the number of samples was selected from Chernoff bound following Tempo et al. (2012). Figure 2a compares the density functions of the operating revenue obtained by the GPR (with 500 training data points) and by the QMC methods (10 000 simulations) with three influential random inputs (flow rates of crude oil, naphtha, and diesel). Furthermore, these density functions were compared with that by the QMC method with all the sixteen random inputs (using the 10 000 simulations from Halton sequence). Similarly, as observed in Figure 2b, the density functions of the CO_2 emissions were obtained using the GPR and the QMC methods with three critical random variables, namely, the flow rates of diesel steam, AGO steam, and the main stripping steam, which were then compared with those by the QMC method with all sixteen random inputs. Table 5 lists the statistical properties and computational time achieved by the proposed and QMC methods for the economic revenue/operating costs. The computational time of the proposed method includes the computational time for running 160 simulations for SA permitted by the MDRM and for constructing two surrogate GPR models with 500 quadrature nodes for UQ. The QMC/MC methods require a large number of simulations (approximately 10 000) to estimate the expected values, variances, and densities accurately; hence, they are computationally expensive. It was observed by comparing Figures 5 and 6 that the proposed (MDRM + GPR) method can achieve acceptable results with only a small computational cost in comparison with the conventional QMC method.

Furthermore, the Kernel scales obtained by the trained GPR model for the operating revenue were {**0.0304**, 0.1202, and **0.0028**}. According the ARDs mechanism, the GPR kernel scales imply that the random inputs of the flow rate of the crude oil and the diesel product are relevant (important) for the operating revenue.

Regarding the result of the GPR, the higher revenue was achieved associated with the increased crude oil flow rate and a high proportion of light product (such as diesel) in the feed. Meanwhile, the Kernel scale of the trained GPR model for the CO_2 emissions were {0.7723, 0.5650, and **0.0475**}, which indicates that the GPR identifies the flowrate of the main stripping steam as the most important input.

Table 1: Statistical data and simulation times obtained by MDRM-GPR and QMC methods

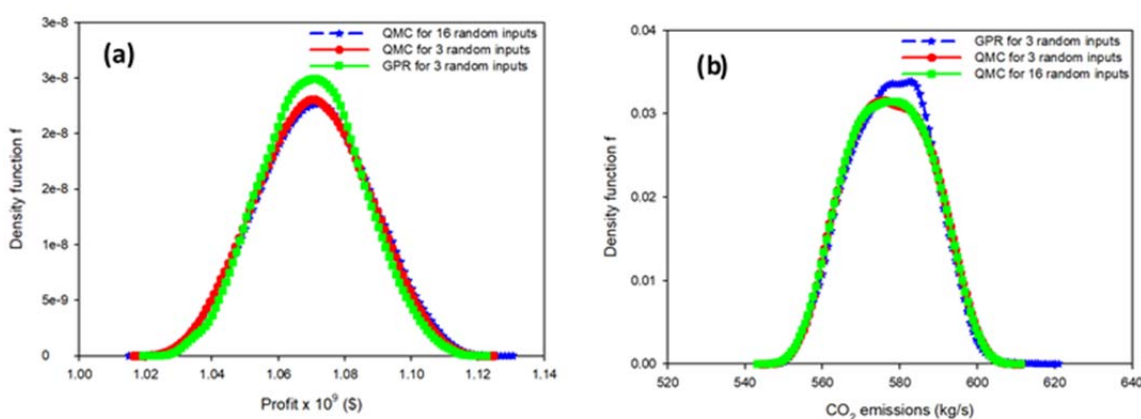| Methods	No. of simulations	Runtime (sec.)	Mean μ (Operating revenue)	Mean μ (CO_2 emissions)
QMC (sixteen variables)	10,000	15,721.4	1.073e9	577.27
QMC (three variables)	10,000	14,756.8	1.071e9	577.79
Proposed method (Operating revenue)	660	1,603.1	1.071e9	-
Proposed method (CO_2 emissions)	660	770.4	-	583.21

Figure 2: Density profiles of: (a) Operating revenue; and (b): CO2 emissions obtained by GPR and QMC methods.

5. Conclusions

The purposes of this work focused on UQ and SA for testing the model robustness against uncertainties in the CDU: obtaining a better understanding of the effects of inputs on the output, determining the key inputs that mainly influence the uncertainty of model outputs (such as the operating revenue and the CO_2 emissions), and achieving a model simplification. HYSYS was used to obtain a rigorous simulation result. To handle a large number of uncertainties with the computational limitation, a two-stage SA approach was presented efficiently because the SA indices were first estimated by the M-DRM to screen the non-influential parameters and the GPR method was then applied to the UQ based on the simplified model from the previous step. There is a precise agreement with those of the conventional QMC method, and it is superior to the popular QMC approaches primarily in terms of the computational efforts. The results indicates that the operating net revenue is mostly dominated by the change of flow rates of crude oil and light fractionation (e.g. diesel product); whereas, the flow rate of main steam for stripping any residue significantly affected the CO_2 emissions. The proposed approach is expected to offer an effective way to handle both SA and UQ for other complex chemical processes associated with a large number of uncertainties. In the future work, another study of SA to focus on how to choose a blend of crude oils for the revenue and CO_2 emissions to imitates a real refinery optimization process, and therefore provides a more realistic simulation regarding the practical application.

References

Abubakar U., Sriramula, S., Renton, N. C., 2015, Reliability of complex chemical engineering processes. Computers & Chemical Engineering, 74, 1-14.

Celse B., Costa, V., Wahl, F., Verstraete, J. J., 2015, Dealing with uncertainties: Sensitivity analysis of vacuum gas oil hydrotreatment. Chemical Engineering Journal, 278, 469-478.

Clifton L., Clifton, D. A., Pimentel, M. A. F., Watkinson, P. J., & Tarassenko, L., 2012, Gaussian process regression in vital-sign early warning systems. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 6161-6164.

Doostan A., & Owhadi, H., 2011, A non-adapted sparse approximation of PDEs with stochastic inputs. Journal of Computational Physics, 230, 3015-3034.

Duong P. L. T., Ali, W., Kwok, E., Lee, M., 2016, Uncertainty quantification and global sensitivity analysis of complex chemical process using a generalized polynomial chaos approach. Computers & Chemical Engineering, 90, 23-30.

Duong P. L. T., Minh, L. Q., Pham, T. N., Goncalves, J., Kwok, E., Lee, M., 2016, Uncertainty quantification and global sensitivity analysis of complex chemical processes with a large number of input parameters using compressive polynomial chaos. Chemical Engineering Research and Design, 115, 204-213.

Enríquez-Gutiérrez V. M., Jobson, M., Ochoa-Estopier, L. M., Smith, R., 2015, Retrofit of heat-integrated crude oil distillation columns. Chemical Engineering Research and Design, 99, 185-198.

Errico M., Tola, G., Mascia, M., 2009, Energy saving in a crude distillation unit by a preflash implementation. Applied Thermal Engineering, 29, 1642-1647.

Gadalla M., Olujić, Ž., Jobson, M., Smith, R., 2006, Estimation and reduction of CO_2 emissions from crude oil distillation units. Energy, 31, 2398-2408.

Gautschi W., 2004, Orthogonal Polynomials: Computation and Approximation: Clarendon Press.

HYSYS, 2007, Aspen HYSYS customization guide. *Aspen Technology Inc.*

Kim Y. H., 2017, An Energy-Efficient Crude Distillation Unit with a Prefractionator. Chemical Engineering & Technology, 40, 588-597.

Kocijan J., 2016, Modelling and Control of Dynamic Systems Using Gaussian Process Models. Cham: Springer

Minh L. Q., Duong, P. L. T., Goncalves, J., Kwok, E., Lee, M., 2017, A two-stage approach of multiplicative dimensional reduction and polynomial chaos for global sensitivity analysis and uncertainty quantification with a large number of process uncertainties. Journal of the Taiwan Institute of Chemical Engineers, 78, 254-264.

Minh L. Q., Duong, P. L. T., Lee, M., 2018, Global sensitivity analysis and uncertainty quantification of crude distillation unit using surrogate model based on Gaussian process regression. Industrial and Engineering Chemistry Research, 57, 5035-5044.

More R. K., Bulasara, V. K., Uppaluri, R., Banjara, V. R., 2010, Optimization of crude distillation system using aspen plus: Effect of binary feed selection on grass-root design. Chemical Engineering Research and Design, 88, 121-134.

Ochoa-Estopier L. M., Jobson, M., Smith, R., 2014, The use of reduced models for design and optimisation of heat-integrated crude oil distillation systems. Energy, 75, 5-13.

Osuolale F. N., Zhang, J., 2017, Thermodynamic optimization of atmospheric distillation unit. Computers & Chemical Engineering, 103, 201-209.

Rasmussen C. E., Williams, C. K. I., 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning): The MIT Press.

Saltelli A., 2002, Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications, 145, 280-297.

Schobi R., Sudret, B., Wiart, J., 2015, Polynomial-chaos-based Kriging. International Journal of Uncertainty Quantification, 5, 171-193.

Sobol' I. M., 2001, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55, 271-280.

Tempo R., Calafiore, G., Dabbene, F., 2012, Randomized algorithms for analysis and control of uncertain systems: with applications: Springer Science & Business Media.

Webster M. D., Tatang, M. A., & McRae, G. J., 1996, Application of the Probabilistic Collocation Method for an Uncertainty Analysis of a Simple Ocean Model. Report 4 (MIT Joint Program on the Science and Policy of Global Change).