Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

On the parameterisation of process flow sheet models

Janusz J. Sikorski [1], George Brownbridge [4], Sushant S. Garud [3], Sebastian Mosbach [1], Iftekar A. Karimi [3], Markus Kraft [1,2]

released: 04 April 2016

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge, CB2 3RA
United Kingdom
E-mail: mk306@cam.ac.uk

[2] School of Chemical and Biomedical Engineering
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
E-mail: mk306@cam.ac.uk

[3] Department of Chemical and Biomolecular Engineering
National University of Singapore
4 Engineering Drive 4
Singapore 117585
E-mail: cheiak@nus.edu.sg

[4] CMCL Innovations
Sheraton House, Castle Park
Cambridge, CB3 0AX
United Kingdom
E-mail: enquiries@cmclinnovations.com

Preprint No. 166

Keywords: process flow sheet model, parameterisation, biodiesel, sensitivity analysis

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge CB2 3RA
United Kingdom

Fax: + 44 (0)1223 334796
E-Mail: c4e@cam.ac.uk
World Wide Web: http://como.cheng.cam.ac.uk/

Abstract

This paper presents results of parameterisation of typical input-output relations within process flow sheet of a biodiesel plant and assesses parameterisation accuracy. A variety of scenarios were considered: 1, 2, 6 and 11 input variables (such as feed flow rate or a heater's operating temperature) were changed simultaneously, 3 domain sizes of the input variables were considered and 2 different surrogates (polynomial and High Dimensional Model Representation (HDMR) fitting) were used. All considered outputs were heat duties of equipment within the plant. All surrogate models achieved at least a reasonable fit regardless of the domain size and number of dimensions. Global sensitivity analysis with respect to 11 inputs indicated that only 4 or fewer inputs had significant influence on any one output. Interaction terms showed only minor effects in all of the cases.

Contents

1 Introduction

Every industrial actor strives towards better understanding and, ultimately, optimisation of any and all of its activities. That applies on each level beginning with workforce schedules and individual pieces of machinery, through specific processes, ending with entire plants. Traditionally the main objectives of such an optimisation are minimising resource use and maximising profit. However, as environmental concerns become ever more pressing ecologically-focused targets such as reducing pollutants, creating cleaner manufacturing processes or reducing carbon footprints rise in prominence.

Those trends prompted significant academic and industrial interest in the concepts of "sustainable development" [19], "industrial ecology" [2, 3, 36, 69] and "industrial symbiosis" [22]. The latter concept brings together separate industries in a collective approach to competitive advantage involving physical exchange of materials, energy, water and by-products [22]. Ecological industrial development based thereon is often realised as Eco-Industrial Parks (EIPs).

An EIP is defined as an industrial park where businesses cooperate with each other and, at times, with the local community to reduce waste and pollution, efficiently share resources (such as information, materials, water, energy, infrastructure, and natural resources), and minimise environmental impact while simultaneously increasing business success [54]. An example of an EIP exists in Kalundborg, Denmark where an exchange network is centred around Asnæs Power Station, a 1500MW coal-fired power plant, and linked to the local community and several other companies [22, 27]. Sample exchanges include selling excess steam from the plant to Novo Nordisk, a pharmaceutical and enzyme manufacturer, and to Statoil power plant or using extra heat to heat local homes and a nearby fish farm. Also, one of the plant's by-products, gypsum, is purchased by a wallboard producer, helping to reduce the amount of necessary open-pit mining [29].

Primary academic interest stems from EIPs' ability to create more sustainable industrial activities through the use of localised symbiotic relationships [13]. To this date a great number of studies concerning various aspects of EIPs have been conducted. Many of them probe methods suitable for optimal design, focusing primarily on employing mathematical programming to create exchange networks of materials, water and energy connecting members of the EIP in question [23, 41, 42, 44, 48]. Utility of such designs is evaluated by monitoring environmental, social and economical impacts.

Holistic modelling of complex, highly interconnected networks is a non-trivial and expensive task, especially for EIPs which include numerous physical models of disparate processes. That is why many studies apply mathematical optimisation to simplified models of individual aspects of the parks.

The limitations of this approach may be overcome by exploiting key features of the concept of Industry 4.0 [54]: creation of virtual copies of the physical world and the ability of industrial components to communicate with each other. Those virtual copies could be surrogate models of physical models produced for a predefined range of inputs. Developing a virtual system primarily based on surrogate models would significantly reduce required computation time and storage space and allow for dynamic modelling and studies otherwise impossible to conduct. Figure 1 presents a framework of EIP modelling based on

Industry 4.0.

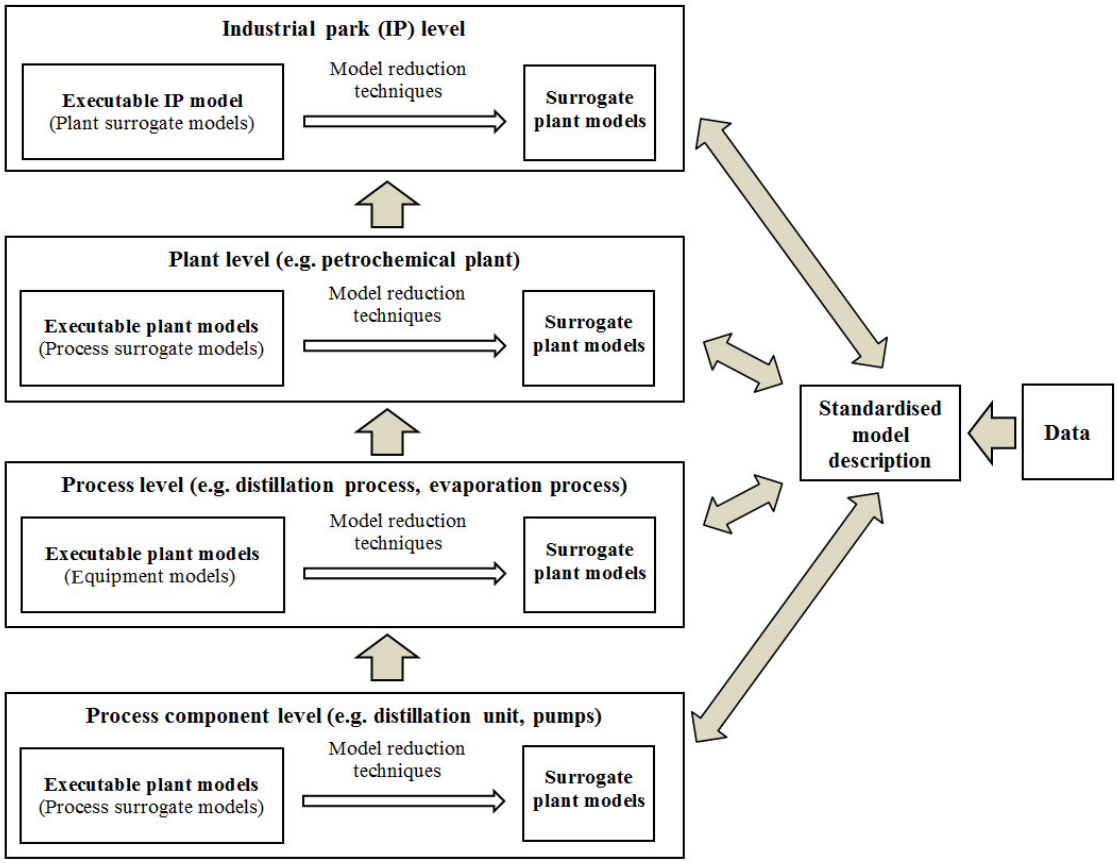


Figure 1: *Framework of EIP modelling based on Industry 4.0. Adopted from Ming Pan et al. [54].*

A surrogate model (or a metamodel) is an approximation of experimental and/or simulation data designed to provide answers when it is too expensive to directly measure the outcome of interest [32]. Two key requirements thereof are reasonable accuracy and significantly faster evaluation than the original method. The models are used to:

- explore design space of a simulation or an experiment,
- calibrate predictive codes of limited accuracy and bridging models of varying fidelity,
- account for noise or missing data,
- gain insight into nature of the input-output relationship (data mining, sensitivity analysis and parameter estimation).

Producing a surrogate model involves choosing a sampling plan (an experimental design), choosing a type of model and fitting the model to the gathered data. Numerous sampling and fitting techniques are available as documented in a number of reviews. Simpson et al.

[62] provides detailed reviews of data sampling and metamodel generation techniques, including response surfaces, kriging, Taguchi approach, artificial neural networks and inductive learning. It also discusses metrics for absolute and relative model assessment, including R^2, residual plots and root mean square error. An introduction to and analysis of linear regression with a focus on generalized linear mixed models with many examples and case studies is provided by Ruppert et al. [59].

A book by Forrester et al. [32] puts the process of data sampling and generating surrogate models into engineering perspective providing numerous case studies and MATLAB code to preform associated calculations. It discusses response surfaces, kriging, support vectors machines and radial basis functions. An in-depth review of kriging, its application and new extensions are provided by Kleijnen [45]. A review and assessment of various sampling techniques is provided by Crary [25]. Reich and Barai [57] focuses on assessment of machine learning techniques, artificial neural networks in particular, with case studies of modelling marine propeller behavior and corrosion data analysis. An example of surrogate models bridging models of varying fidelity is provided by Bakr et al. [10] where a surrogate maps data produced by fine and coarse physical models in order to accelerate optimisation of the fine model. Jin et al. [38] assesses applicability and accuracy of metamodels for optimisation under uncertainty and reports promising results noting that only a small-size analytical problem was considered. Surrogate models are widely employed in engineering and science for space exploration [34, 35], modelling [21, 26, 46], sensitivity analysis [6, 20, 35, 40, 52], parameter estimation [9, 14, 43], optimisation in areas ranging from circuit design through nanoparticle synthesis to flood monitoring [4, 12, 58]. However, none of the aforementioned applications employ surrogate models in order to describe a process flow sheet model of a typical industrial process and assess quality of the description.

The main purpose of this paper is to approximate the relations between 11 inputs typical to a biodiesel plant and its energy requirements using surrogate models and assess accuracy of the approximations. Additionally, it aims to investigate the effects of dimensionality, domain size and surrogate type on the accuracy and analyse global sensitivities of the outputs in order to identify opportunities for dimensionality reduction.

This paper is structured as follows. Section 2 describes the biodiesel plant model and its modelling environment. Section 3 presents sampling and surrogate generation techniques prodecures and software employed to perform those. Section 4 provides implementation details of the surrogate models and accuracy indices used to assess them. Section 5 presents results of the numerical analysis, while Section 6 summarizes the main findings.

2 Model

2.1 Aspen Plus V8.6

Aspen Plus is a process modelling and optimisation software used by the bulk, fine, specialty, and biochemical industries, as well as the polymer industry for the design, operation, and optimisation of safe, profitable manufacturing facilities [5]. Its capabilities include:

- optimisation of processing capacity and operating conditions,
- assessment of model accuracy,
- monitoring safety and operational issues,
- identifying energy savings opportunities and reduce greenhouse gas (GHG) emissions,
- performing economic evaluation,
- improving equipment design and performance.

The software was used to simulate the process described in Section 2.2.

2.2 Biodiesel plant simulation

The process flow sheet model under investigation includes initial stages of a biodiesel production line, namely a reaction step and a separation step, with auxiliary equipment as seen in Figure 2. The final fuel, fatty acid methyl ester, is produced via trans-esterification pathway where triglycerides react with methanol to form methyl ester and glycerin in the presence of an alkaline catalyst. The flow sheet was based on an existing plant designed by Lurgi GmbH. It consists of the following elements: a continuously stirred tank reactor (CSTR), a flash drum, a decanter, 3 heaters and 11 material streams. In the process tripalmitine oil is reacted with methanol in the CSTR to produce glycerol and methyl-palmitate (biodiesel) and then passed through a flash drum and a decanter to separate excess methanol and glycerol. The simulation is solved for steady-state operation and produces a wide variety of chemical and physical information ranging from throughput to heat duties of individual equipment.

In this study surrogate models were used to describe relations between chosen inputs and outputs occurring in the process flow sheet model. The choice of variables aimed to study effects of inputs typical for chemical plants on energy consumption as it is desired to study interactions between chemical and electrical models in the future. Three domain sizes of the input variables were considered in order to assess their effect on the parametrisation accuracy. The variables' names, domain and preferred operating conditions are listed in Tables 1 and 2. Plots of heat duties of various equipment against molar flow of tripalmitin oil can be seen in Figure 3.

Table 1: *Input variables.*

Name	Lower bounds	Upper bounds	Operating point
Molar flow of tripalmitine oil (kmol/hr)	20, 22.5, 25	40, 37.5, 35	30
Temperature of tripalmitine oil (^{o}C)	20, 22.5, 25	40, 37.5, 35	30
Operating temperature of CSTR 10D01 (^{o}C)	44, 49, 54	64	60
Volume of CSTR 10D01 (m^3)	40, 43, 45	50, 49, 47	45
Operating temperature of flash drum 10D02 (^{o}C)	80, 82.5, 85	100, 97.5, 95	90
Operating temperature of heater 10E01 (^{o}C)	60, 62.5, 65	80, 77.5, 75	70
Molar flow of methanol (kmol/hr)	150, 160, 170	210, 200, 190	180
Temperature of methanol (^{o}C)	20, 22.5, 25	40, 37.5, 35	30
Operating temperature of decanter 10D02D (^{o}C)	20, 22.5, 25	40, 37.5, 35	30
Operating temperature of heater 10E02 (^{o}C)	80, 82.5, 85	100, 97.5, 95	90
Operating temperature of heater 10E03 (^{o}C)	60, 62.5, 65	80, 77.5, 75	70

Table 2: *Output variables.*

Name
Heat duty of heater 10E01 (MW)
Heat duty of heater 10E02 (MW)
Heat duty of heater 10E03 (MW)
Heat duty of reactor 10D01 (MW)
Heat duty of flash drum 10D02 (MW)
Heat duty of decanter 10D02D (MW)

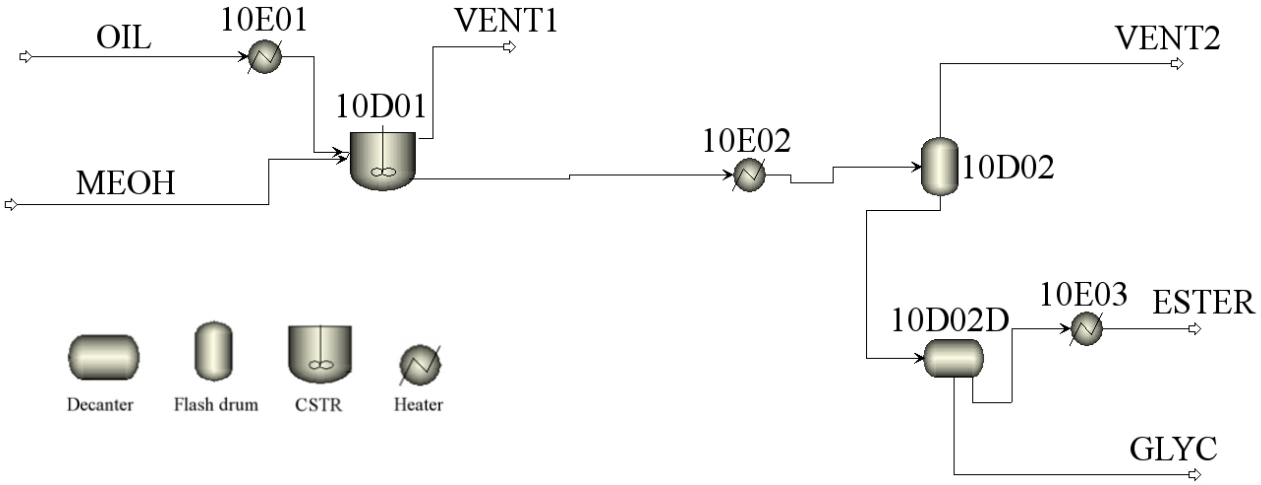
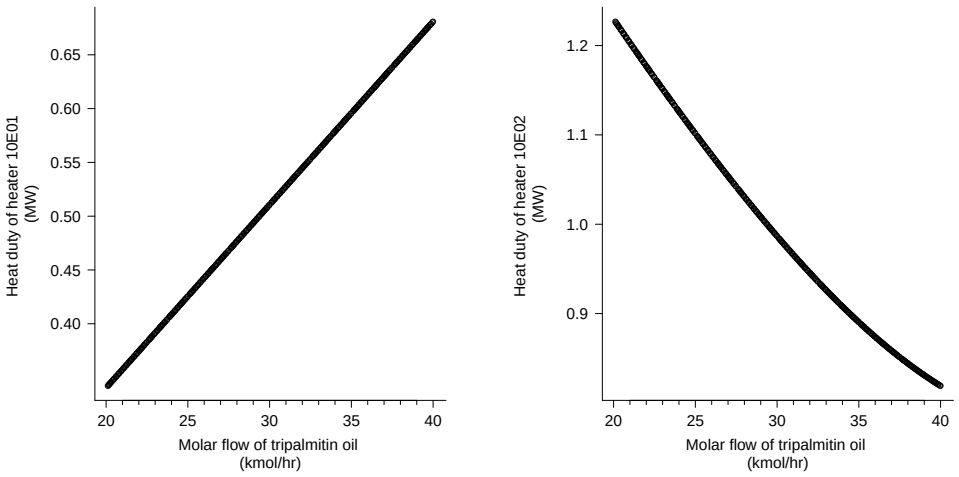
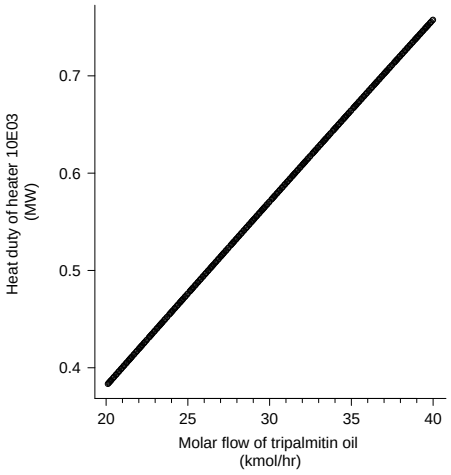


Figure 2: *Aspen Plus flowsheet representing the biodiesel production line.*

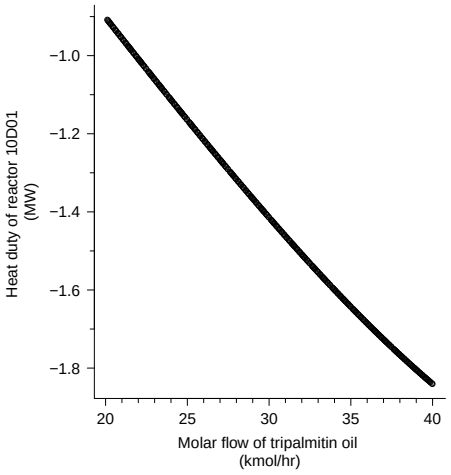


(a) Heater 10E01

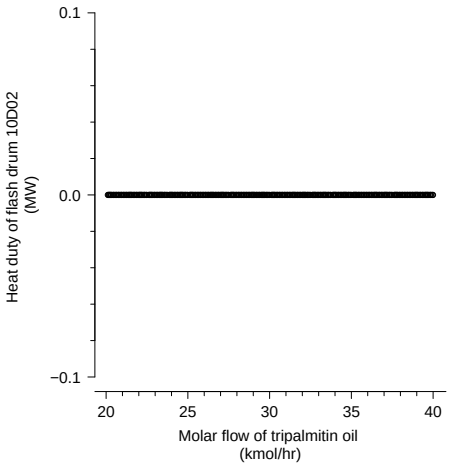
(b) Heater 10E02



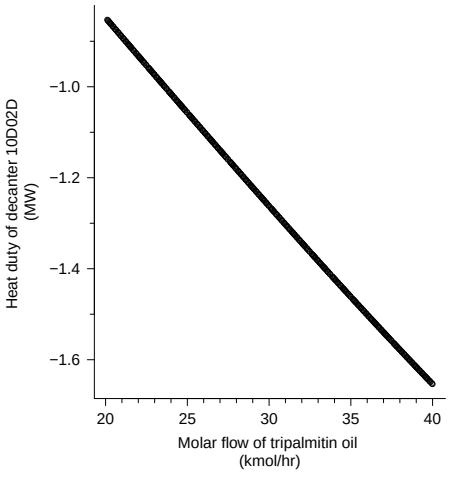
(c) Heater 10E03



(d) Reactor 10D01



(e) Flash drum 10D02



(f) Decanter 10D02D

Figure 3: *Plots of heat duties of various equipment against molar flow of tripalmitin oil.*

3 Parameterisation

3.1 Model Development Suite

Model Development Suite (MoDS) [24] is an advanced software tool designed to analyse black-box models (e.g. executables, batch scripts). It includes a broad range of tools such as data-driven modelling, multi-objective optimisation, generation of surrogate models, data standardisation and visualisation, global parameter estimation [14–16, 49–52, 60, 61], uncertainty propagation [7, 18], global and local sensitivity analysis [6, 66, 67], and intelligent design of experiments [1, 8, 17, 30, 63, 70]. It was used to sample data, produce surrogate models and compute global sensitivities.

Sobol sequence, a quasi-random low discrepancy sampling method, is employed for sampling data and polynomial fitting and High Dimensional Model Representation (HDMR) fitting are used to generate surrogate models. A brief description of each is included, respectively, in Sections 3.4, 4.1 and 4.2.

3.2 MoDS-Aspen Plus interface - Component Object Model (COM)

The data collection and parameterization process of a model can be automated using MoDS provided an executable file capable of reading an input file, running the considered model and producing an output file (input and output files need to have either .csv or .xml format).

For the purpose of this study a script written in Python 3.4 was used to manipulate the Aspen Plus simulation via Microsoft Component Object Model (COM) interface. COM is a platform-independent, binary-interface standard enabling creation of objects and communication between them [53]. COM object (also known as COM component) is defined as a piece of compiled code that provides a service to the rest of the system. That can be a script, an instance of a program e.g. an Aspen Plus simulation. A primary feature of this architecture is the fact that COM components access each other through interface pointers, rather than directly. It provides a number of functions applicable to all components. Any additional functions need to be provided by the object or the user, in both cases via a library associated with the object. In this project COM interface is primarily used to launch, explore data structures, access data entries and solve models simulated within Aspen Plus.

3.3 Data harvest and surrogate generation

Data collection, processing and visualisation were done using MoDS and custom-made Python 3.4 and R 3.2.2 scripts. The process of producing a surrogate of existing models involves the following steps: generation of input data, reception of output data from the studied model and, when both data sets are complete, scanning for and excluding erroneous data points and executing a parametrisation algorithm. The first two steps are critical to ensure high accuracy of the surrogate model and hence a sufficient number of

points and a suitable sampling method are required to satisfactorily describe the input-output relation for a given number of independent variables and operating range. In this study the following procedure was used:

1. A Sobol sequence was used to generate input data for user-specified variables within the process flow sheet model.
2. Model's input data was altered according to the generated input data.
3. The simulation was evaluated with the new inputs.
4. MoDS retrieved values of user-specified outputs.
5. Data was scanned for errors and corrected.
6. Polynomial and HDMR fitting were used to generate surrogate models describing the relation between inputs and outputs.

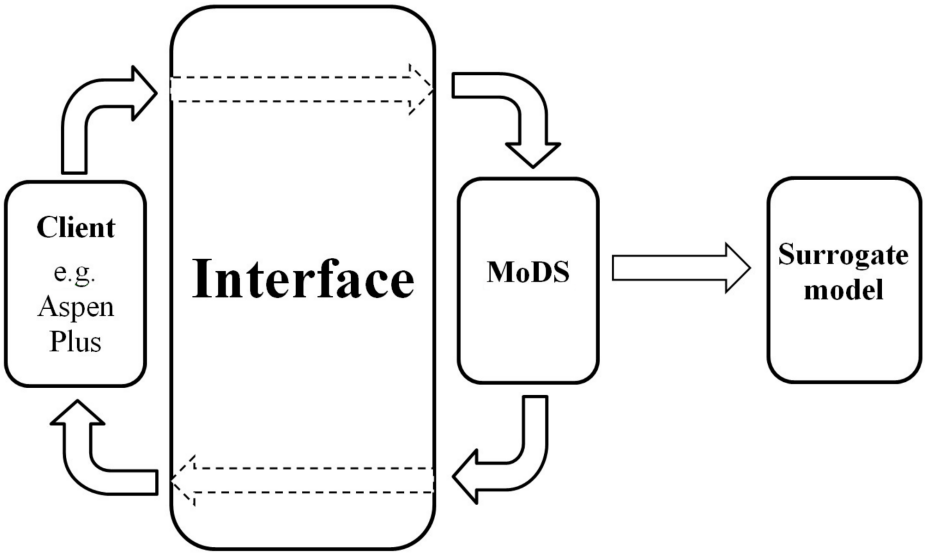


Figure 4: *Model Development Suite work flow.*

The workflow of MoDS is visualized in Figure 4. A variety of scenarios were considered: 1, 2, 6 and 11 input variables were changed simultaneously, 3 different domain sizes of the input variables were considered and 2 different surrogate generation methods (polynomial and HDMR fitting) were used. To ensure that there is always sufficient number of points required to generate a surrogate, each simulation produced 400 points per input variable (prior to error exclusion). They were used for fitting surrogates and calculating R^2 and $\bar{R}^2$. Depending on the case, erroneous points made up to 1% of all points. They arose due to convergence and stability issues within Aspen Plus. Additionally, test sets of points (100 points per dimension) were generated for calculating Root-Mean-Square Deviation (RMSD) and residuals (see Section 4.3 for further description). In this study three domain sizes of the input variables were considered in order to assess their effect on the parameterisation accuracy. The domain bounds of input variables during simulations and initial steady state values are summarised in Table 1.

3.4 Sampling

Data points were generated using Sobol sequences, a type of quasi-random, low-discrepancy sequences. Low discrepancy of points in such a sequence means that their proportion falling into an arbitrary set is approximately proportional to the measure of the set. This property is true on average, but not necessarily for specific samples. Their ability to cover considered domain quickly and evenly gives them advantages over purely random numbers. Also, in contrast to deterministic sequences, they do not require a predefined number of samples and their coverage improves continually as more data points are added. Sobol sequences uses a base of two to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension [64]. The MoDS implementation of a Sobol sequence generator follows the description of Joe and Kuo [39].

4 Implementation

4.1 Polynomial response surfaces

Polynomial response surfaces are a subset of response surface methodology, a group of mathematical and statistical techniques designed to facilitate empirical model building [55]. Polynomials of a predefined degree are optimized to describe an unknown relation between independent variables (input variables) and responses (output variables). Input and output data sets are obtained via series of tests, an experiment, in which the input variables are modified in order to study the changes in the output responses. As the number of adjustable coefficients in a polynomial surrogate increases combinatorially with its order and number of variables so does the minimum number of data points required to produce it. Hence applying high-order polynomials to problems with many inputs may lead to overfitting and hence poorer predictive power. Generally, overfitting occurs when a model describes features specific to the data set on which it is trained such as random error or noise. For deterministic computer experiments those are not an issue, but an overfitted model will suffer from having an exaggerated set of coefficients providing no intuitive insight into nature of the relationship under consideration and from introducing irrelevant nonlinearity.

General linear least-squares fit

When fitting polynomial of a given order k to a data set the objective function to be minimised is the weighted sum of the squares of the differences between data and model. This analysis assumes N data values $y^{(1)}, \ldots, y^{(N)}$ obtained at the points $x^{(1)}, \ldots, x^{(N)}$, and statistical weights $W^{(1)}, \ldots, W^{(N)}$ are given. Coefficients of the polynomial are given by

$$\beta^* = \underset{\beta}{\operatorname{argmin}} \Phi(\beta)$$

with

$$\Phi(\beta) = \sum_{i=1}^{N} W^{(i)} \left[y^{(i)} - f_\beta\left(x^{(i)}\right) \right]^2$$

The polynomial f_β is given by

$$f_\beta(x) = \sum_{|p| \le k} \beta_p x^p.$$

The necessary condition $\frac{\partial \Phi}{\partial \beta_q} = 0$ for any multi-index 1 with $|q| \le k$ for stationary points of Φ then becomes

$$\begin{aligned} 0 = \frac{\partial}{\partial \beta_q} \Phi(\beta) &= 2 \sum_{i=1}^{N} W^{(i)} \left[y^{(i)} - f_\beta\left(x^{(i)}\right) \right] \nabla_\beta f_\beta\left(x^{(i)}\right) \\ &= 2 \sum_{i=1}^{N} W^{(i)} \left[y^{(i)} - f_\beta\left(x^{(i)}\right) \right] \nabla_\beta \sum_{|p| \le k} \beta_p \left(x^{(i)}\right)^p). \end{aligned}$$

Looking at the q^{th} component of the gradient the following equation is obtained

$$\begin{aligned} 0 \overset{!}{=} \frac{\partial \Phi}{\partial \beta_q} &= 2\sum_{i=1}^{N} W^{(i)} \left[y^{(i)} - f_\beta\left(x^{(i)}\right) \right] \frac{\partial}{\partial \beta_q} \sum_{|p|\leq k} \beta_p g_p\left(x^{(i)}\right) \\ &= 2\sum_{i=1}^{N} W^{(i)} \left[y^{(i)} - \sum_{|p|\leq k} \beta_p g_p\left(x^{(i)}\right) \right] g_q\left(x^{(i)}\right). \end{aligned}$$

Rearranging yields

$$\begin{aligned} \sum_{i=1}^{N} W^{(i)} y^{(i)} g_q\left(x^{(i)}\right) &= \sum_{i=1}^{N} W^{(i)} \sum_{|p|\leq k} \beta_p g_p\left(x^{(i)}\right) g_q\left(x^{(i)}\right) \\ &= \sum_{|p|\leq k} \beta_p \left[\sum_{i=1}^{N} W^{(i)} g_p\left(x^{(i)}\right) g_q\left(x^{(i)}\right) \right]. \end{aligned} \tag{1}$$

In case of a polynomial, this linear system of equations, called *normal equations*, consists of $\binom{n+k}{k}$ equations for as many unknown coefficients β.

4.2 High Dimensional Model Representation

High Dimensional Model Representation (HDMR) is a finite expansion for a given multivariable function as described by Rabitz and Alış [56], Sobol [65]. Its allows readily extracting global sensitivities with respect to the independent variables by calculating them from the coefficients of a HDMR surrogate. Also, it needs to be noted that the number of parameters within HDMR fit increases far slower than within polynomial fit when high-dimensional problems are considered.

In HDMR representation the output function y is decomposed into a sum of functions that only depend on subsets of the input variables such that:

$$y = f(x) = f_0 + \sum_{i=1}^{N_x} f_i(x_i) + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} f_{ij}(x_i, x_j) + \ldots + f_{12\ldots N_x}(x_1, x_2, \ldots, x_{N_x})$$

where N_x is the number of input parameters, i and j index the input parameters, and f_0 is the mean value of $f(x)$. The expansion given above has a finite number of terms and exactly represents $f(x)$, however for most practical applications terms containing functions of more than two input parameters can often be ignored due to their negligible contributions compared to the lower order terms [47, 56]. Hence for most models or data the truncated approximation

$$y \approx f(x) = f_0 + \sum_{i=1}^{N_x} f_i(x_i) + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} f_{ij}(x_i, x_j)$$

is sufficient. An efficient method of evaluating each of these terms is to approximate the functions $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ with analytic functions, $\phi_k(x_i)$, [47]. For data produced

using random and quasi-random sampling these functions are related by:

$$f_0 = \overline{f}\,, \tag{2a}$$

$$f_i(x_i) = \sum_{k=1}^{M} \alpha_{i,k}\phi_k(x_i)\,, \tag{2b}$$

$$f_{ij}(x_i,x_j) = \sum_{k=1}^{M'}\sum_{l=k+1}^{M'} \beta_{ij,kl}\phi_k(x_i)\,\phi_l(x_j)\,. \tag{2c}$$

The functions, $\phi_k(x_i)$ are orthonormal obeying,

$$\int \phi_k(x_i)\,\mathrm{d}x_i = 0 \tag{3a}$$

$$\int \phi_k(x_i)\,\phi_l(x_i)\,\mathrm{d}x_i = \delta_{kl}\,. \tag{3b}$$

This leads the following equations for the coefficients:

$$f_0 = \int f(x)\mathrm{d}x\,, \tag{4a}$$

$$\alpha_{i,k} = \int f(x)\phi_k(x_i)\,\mathrm{d}x\,, \tag{4b}$$

$$\beta_{ij,kl} = \int f(x)\phi_k(x_i)\,\phi_l(x_j)\,\mathrm{d}x\,, \tag{4c}$$

The separation of the contributions from each individual input parameter and each combination of parameters makes the process of calculating the global sensitivities almost trivial. It has been described by Rabitz and Alış [56] that the contribution of each term in (2), $\sigma^2_{\bar{y},i}$ and $\sigma^2_{\bar{y},ij}$, to the variance of the output parameter can be related to the total variance by

$$\sigma^2_{\bar{y}} = \sum_{i=1}^{N_x}\int_{-1}^{1} f_i^2(x_i)\,\mathrm{d}x_i + \sum_{i=1}^{N_x}\sum_{j=i+1}^{N_x}\int_{-1}^{1}\int_{-1}^{1} f_{ij}^2(x_i,x_j)\,\mathrm{d}x_i\mathrm{d}x_j \tag{5a}$$

$$= \sum_{i=1}^{N_x}\sigma^2_{\bar{y},i} + \sum_{i=1}^{N_x}\sum_{j=i+1}^{N_x}\sigma^2_{\bar{y},ij}\,. \tag{5b}$$

The sensitivities, S_i and S_{ij}, can then be calculated by dividing by the total variance $\sigma^2_{\bar{y}}$ to get

$$S_i = \frac{\sigma^2_{\bar{y},i}}{\sigma^2_{\bar{y}}} \quad \text{and} \quad S_{ij} = \frac{\sigma^2_{\bar{y},ij}}{\sigma^2_{\bar{y}}}\,. \tag{6}$$

Global sensitivity analysis explores the parameter space and provides robust sensitivity measures throughout the region of interested even in the presence of nonlinearity and parameter interactions. In nonlinear cases, derivative-based local sensitivity analysis can give a false impression of sensitivity [68].

4.2.1 Basis functions

Polynomials, including Lagrange polynomials [11], orthonormal polynomials, cubic B splines, and ordinary polynomials [47], are commonly used as basis functions for HDMR construction.

In MoDS, Legendre polynomials, $P_m(x)$, are used as the basis functions, $\phi(x)$. They are normalised according to

$$\int_{-1}^{1} P_m(x)P_n(x)\,dx = \frac{2}{2n+1}\delta_{mn}\,, \tag{7}$$

to satisfy (3b). The polynomials are generated at runtime according to Bonnet's recursion formula

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)\,, \tag{8}$$

where $P_0(x) = 1$ and $P_1(x) = x$. This means that maximum polynomial order, M^*, can be set to an arbitrary natural number. Additionally, maximum interaction order, M'^*, needs to be set to either 1 or 2.

4.2.2 Automatic order selection

Accuracy improvement due to each new term is assessed by calculating $\bar{R}^2$ value (9) and comparing it against a predefined minimum value $\bar{R}^{2*}$ (0.99999), before continuing on to the next one. If a term's contribution is smaller than the threshold, the term is discarded. The algorithm terminates once maximum polynomial orders M^* and M'^* are reached. This is a new approach different from generally used least square method described by Ziehn and Tomlin [71]. In MoDS implementation the fitting algorithm computes successive terms of a polynomial fit and discards one deemed unnecessary. It has several advantages over employment of a raw polynomial including reduction of data processing, computational complexity and number of optimisable parameters, which greatly helps dealing with high-dimensional problems.

$$\begin{aligned}\bar{R}^2 =& 1-(1-R^2)\frac{N-1}{N-p-1}\\ =& 1-\frac{\sum_{i=1}^{N}(y^{(i)}-\bar{y})^2}{\sum_{i=1}^{N}(y^{(i)}-f^{(i)})^2}\frac{N-1}{N-p-1}\,,\end{aligned} \tag{9}$$

where $y^{(i)}$ is an i^{th} data point, $f^{(i)}$ is an i^{th} model predicted value, $\bar{y}$ is the empirical mean of data points, N is the number of data points, p is the number of adjustable parameters and $i = 1,2,..,N$.

The use of $\bar{R}^2$ rather than R^2 acts as a partial safeguard against over fitting. All of the functions f_i have the same polynomial order, M^*, and the f_{ij} are all of order M'^*. Also, it is assumed that the magnitude of the coefficients decreases as the order of the basis function increases. Whilst this is valid in many situations it may not always be applicable.

4.3 Accuracy measures

There exist various accuracy measures applicable to surrogate models, but there is no single, all-encompassing index. For that reason a number of methods was used including R^2, $\bar{R}^2$, Root-Mean-Squared-Deviation(RMSD) and residual plots. The indices are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y^{(i)} - \bar{y})^2}{\sum_{i=1}^{N}(y^{(i)} - f^{(i)})^2}$$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{N}{N-p}$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(y^{(i)} - f^{(i)})^2}{N}}$$

$$e^{(i)} = y^{(i)} - f^{(i)}$$

where $y^{(i)}$ is the i^{th} data point, $f^{(i)}$ is an i^{th} model predicted value, $\bar{y}$ is the empirical mean of data points, N is the number of data points, p is the number of adjustable parameters, $e^{(i)}$ refers to residual for i^{th} data point and $i = 1, 2, .., N$. The first three measures are single number indices thus more convenient, but less informative than residual plots.

R^2 (coefficient of determination) is a measure indicating fit of a statistical model to data [28]. In essence, it compares the discrepancies between the predicted data and actual data with the discrepancies between the arithmetic average and actual data.

$\bar{R}^2$ is R^2, as described above, corrected for the number of fitted parameters relative to the number of data points. This measure cannot be greater than R^2(for $N > p$) and it decreases as $N \to p$ indicating that the model overfits the data.

RMSD is the sample standard deviation of the differences between predicted values and observed values [37]. It is a good metric for comparing predictive power of different models for a particular variable (but not between the variables due to scale dependency).

5 Numerical experiments

5.1 Polynomial versus HDMR

$\bar{R}^2$ values were produced using the training set and are used to assess fit of the surrogates to the training data (data sampled from Aspen Plus used for model fitting), while RMSD and residual plots were produced using the test set (data sampled from Aspen Plus used for testing, but not model fitting). Values sampled from entire domain of the input variables were used unless specified otherwise. Plots comparing surrogate types include polynomial fits of order 1 through 5 (labelled as P1 through P5) and HDMR fits with various constraints. Label H1 corresponds to a 1^{st} order fit, H2a to a 2^{nd} order without interactions, H2b to 2^{nd} order with interactions and H10 to 10^{th} order with 2^{nd} order interactions. Note that HDMR fits may consist of terms with powers lower than specified, but in such a case it will be explicitly mentioned.

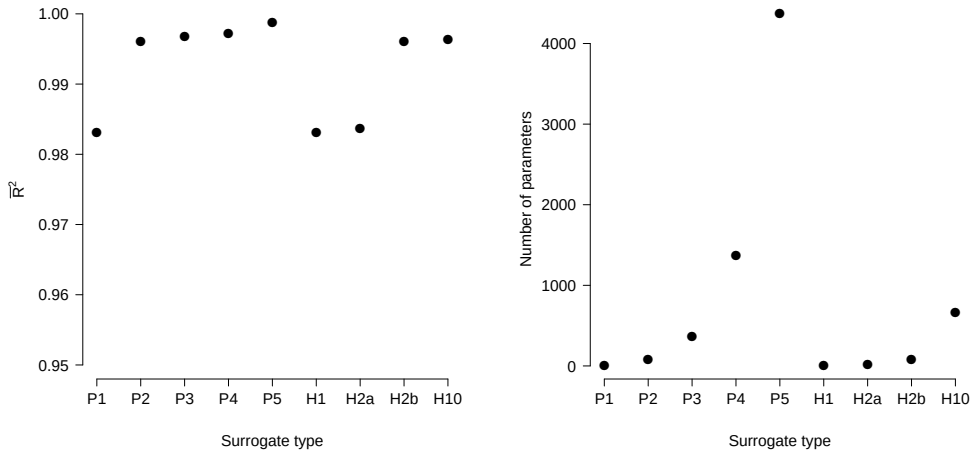
A number of different behaviours were observed in the study. Most surrogate models achieved at least a reasonable fit regardless of the domain size, number of dimensions and according to $\bar{R}^2$ and RMSD. Neither R^2 nor $\bar{R}^2$ can be used to effectively differentiate between the models as most achieve values in excess of 0.98 (for an example see Figure 5(a)). However, there is noticeable increase in $\bar{R}^2$ due to 2nd order interaction terms (P1 to P2 and H2a to H2b). Also, it needs to be noted that the number of parameters within HDMR fit increases far slower than within polynomial fit when high-dimensional problems are considered. Even the most extensive HDMR fit H10 had far fewer parameters than polynomial fits of order > 3, as seen on plot 5(b).

RMSD provides a reasonable measure for comparing accuracy of models, as seen in Figure 6. Plots 6(a) and 6(b) suggest that polynomial fit of order 3 and HDMR fit H2b (marked by green squares) minimise RMSD and hence are the best fit for the duty of reactor 10D01 with respect to all 11 inputs. The aforementioned plots (marked by orange triangles) also show that increasing order of polynomial fit lead to poorer predictive powers, most likely due to overfitting the training data. Similarly, HDMR fit H10 produces larger RMSD values than H2b. It can be seen that adding interaction (H2a to H2b) effect noticeably decreases RMSD in HDMR fitting.

Plots 6(c) and 6(d) show how RMSD changes as the domain size of inputs increases. The former plot (for 5^{th} order polynomial fit) shows an exponential increase, while the latter (for HDMR fit H10) shows decrease of RMSD from smallest to intermediate size and sharp increase from intermediate to largest size.

Residual plots are the most informative form of error measurement as they show the error size and distribution helping to understand whether the fit captures the true nature of the data. In most cases data does not seem to follow a polynomial relation resulting in non-random distribution of the residuals. Figures 8 and 9 present residual plots for 11-dimensional surrogates of heat duties of reactor 10D01 and heater 10E03. Comparison of plots in Figures 8 and 7 shows that for output produced by surrogates with multiple input variables the non-random features are much more difficult to make out. Magnitude of the residuals in most cases is relatively small indicating strong predictive powers of the fits. Comparing plots 8(c) and 7(c) reveals that performance of polynomial fit of order 5 drops

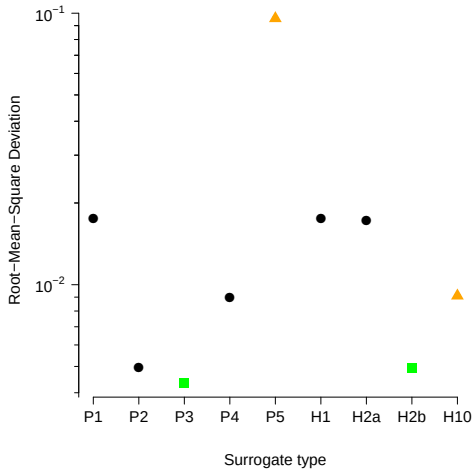
from being the best model to the worst. Plots 8(b) and 8(d) show that even though HDMR fit H10 produced a higher RMSD, its residual plot is as good as seemingly better P3 fit. Those also confirm that P3 seems to be one of the best fits. Plot 8(c) confirms that P5 fit exhibits relatively low accuracy, even worse than that of a simple linear fit (see plot 8(a)).



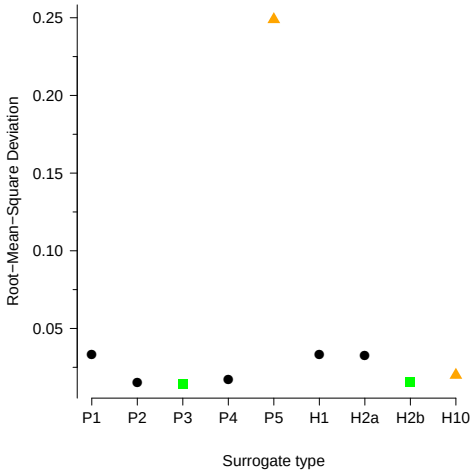
(a) Plot of $\bar{R}^2$ for the considered surrogates.

(b) Plot of number of parameters for the considered surrogates.

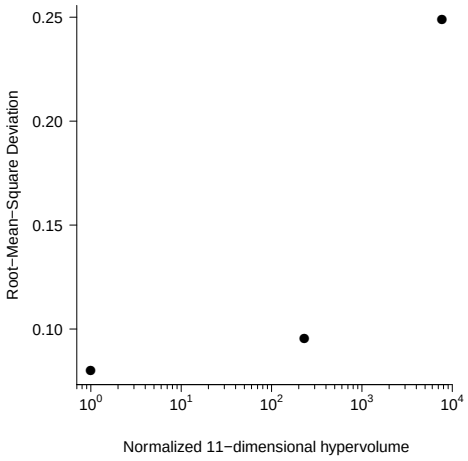
Figure 5: *Plots of RMSD and number of parameters for the considered surrogates produced for heat duty of reactor 10D01 with respect to all 11 inputs. Labels P1 through P5 correspond to polynomial fits of order 1 through 5. Label H1 corresponds to a 1^{st} order fit, H2a to a 2^{nd} order without interactions, H2b to 2^{nd} order with interactions and H10 to 10^{th} order with 2^{nd} order interactions.*



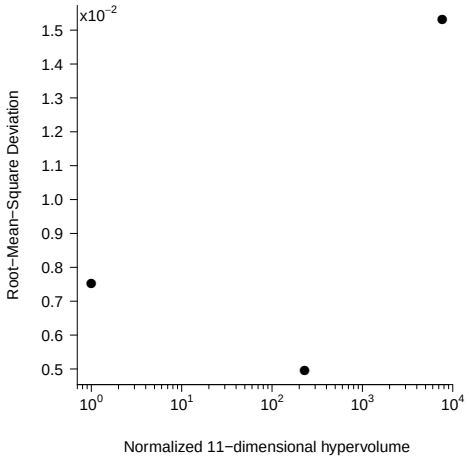
(a) RMSD for the considered surrogates for medium domain size.



(b) RMSD for the considered surrogates.

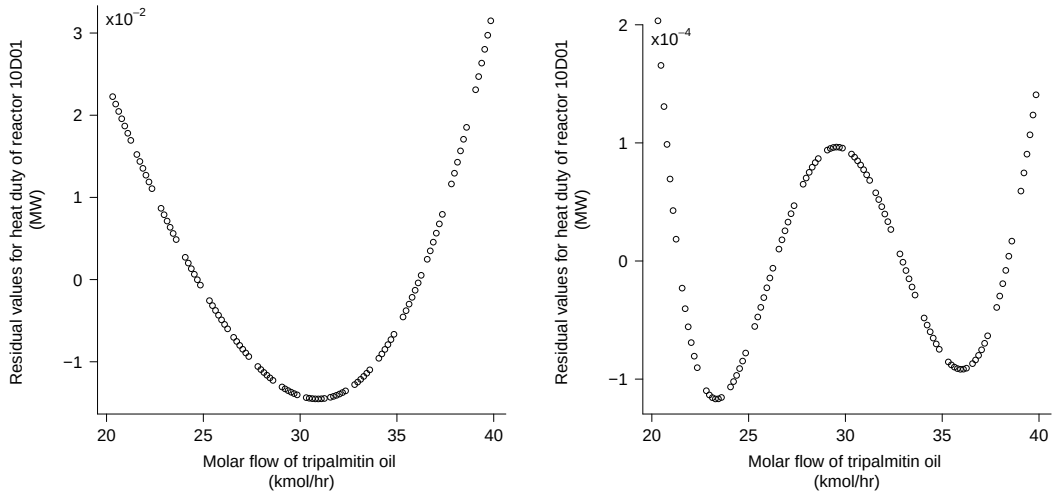


(c) RMSD against domain sizes for polynomial fit of order 5 (for boundaries see Table 1).



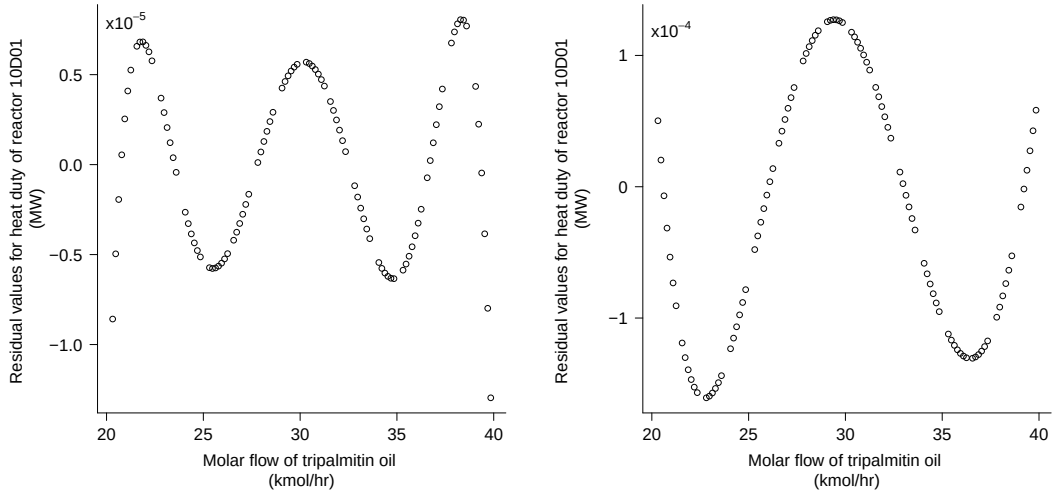
(d) RMSD against domain sizes for HDMR fit (for boundaries see Table 1).

Figure 6: *Plots of RMSD for the considered surrogates and domain sizes produced for heat duty of reactor 10D01 with respect to all 11 inputs. Labels P1 through P5 correspond to polynomial fits of order 1 through 5. Label H1 corresponds to a 1^{st} order fit, H2a to a 2^{nd} order without interactions, H2b to 2^{nd} order with interactions and H10 to 10^{th} order with 2^{nd} order interactions. Green squares indicate models (one per type) with lowest RMSD, while red triangles indicate models (one per type) with suffering most from overfitting.*



(a) Plot of residuals for 1^{st} order polynomial fit.

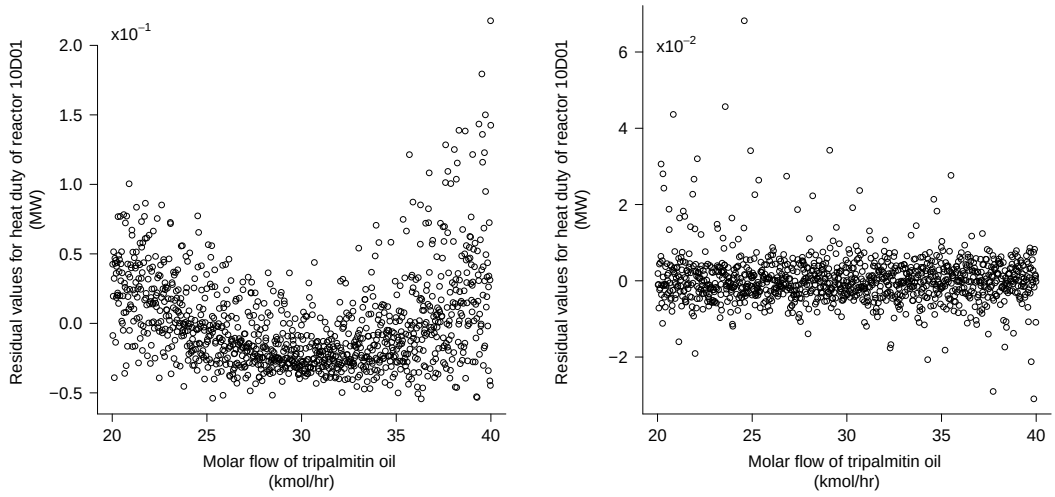
(b) Plot of residuals for 3^{rd} order polynomial fit.



(c) Plot of residuals for 5^{th} order polynomial fit.

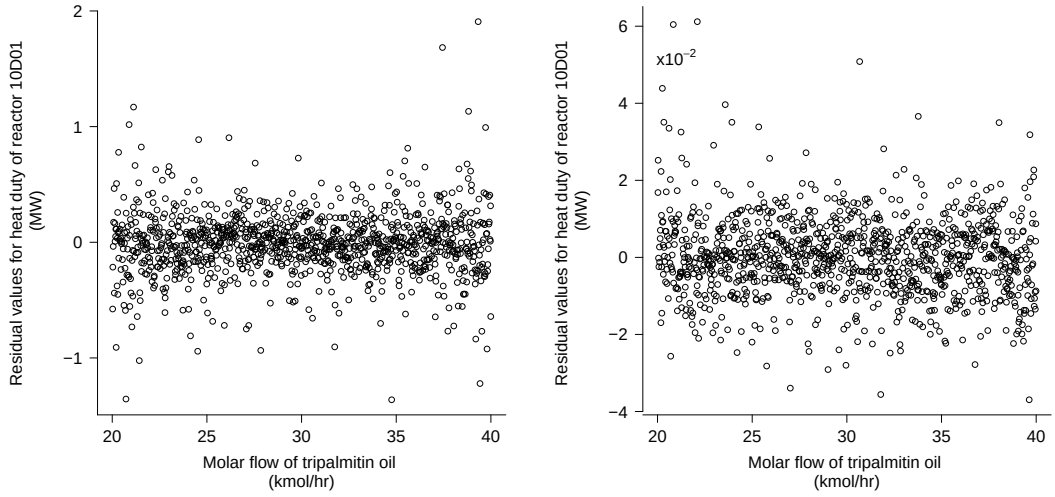
(d) Plot of residuals for HDMR fit H10 (3^{rd} order polynomial).

Figure 7: *Plot of residuals against molar flow of tripalmitin oil for heat duty of reactor 10D01 produced for 1 input.*



(a) Plot of residuals for 1^{st} order polynomial fit.

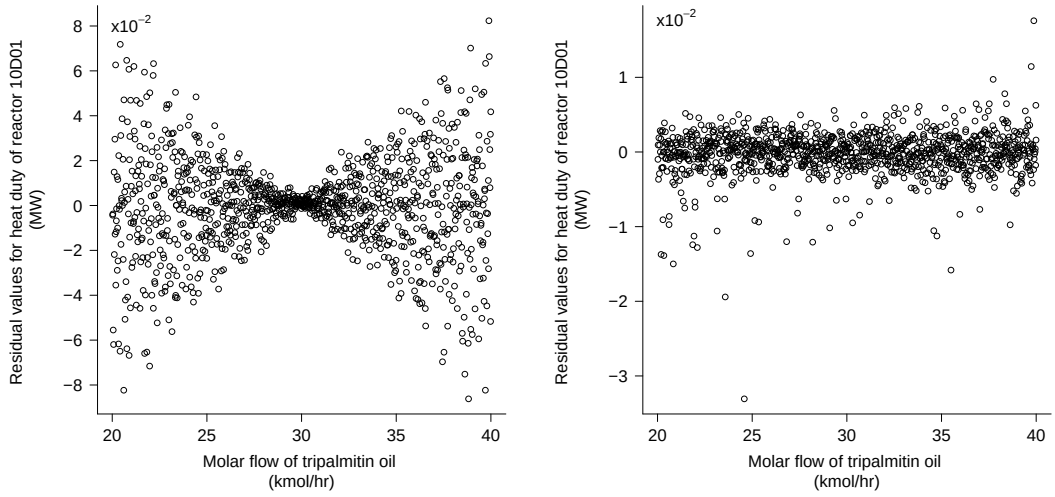
(b) Plot of residuals for 3^{rd} order polynomial fit.



(c) Plot of residuals for 5^{th} order polynomial fit.

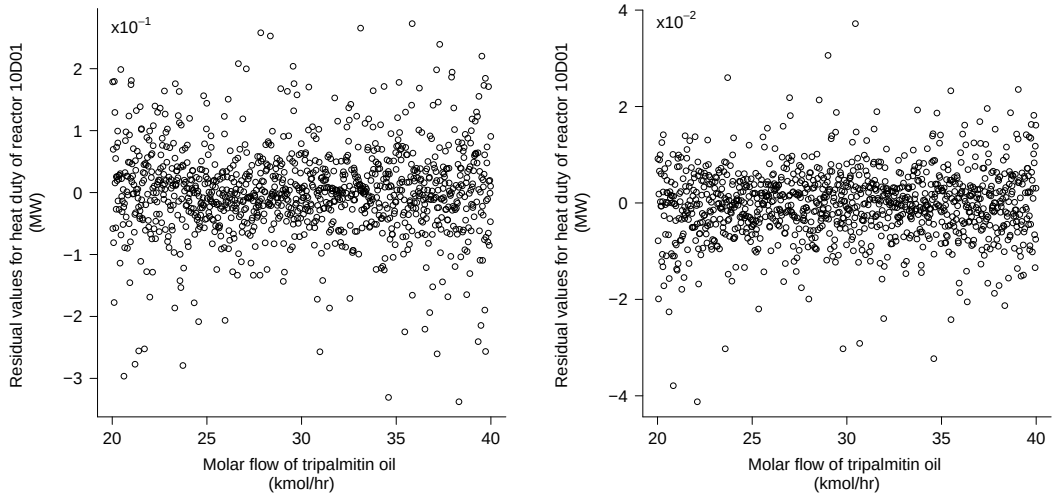
(d) Plot of residuals for HDMR fit H10.

Figure 8: *Plot of residuals against molar flow of tripalmitin oil for heat duty of reactor 10D01 produced for 11 inputs.*



(a) Plot of residuals for 1^{st} order polynomial fit.

(b) Plot of residuals for 3^{rd} order polynomial fit.



(c) Plot of residuals for 5^{th} order polynomial fit.

(d) Plot of residuals for HDMR fit H10.

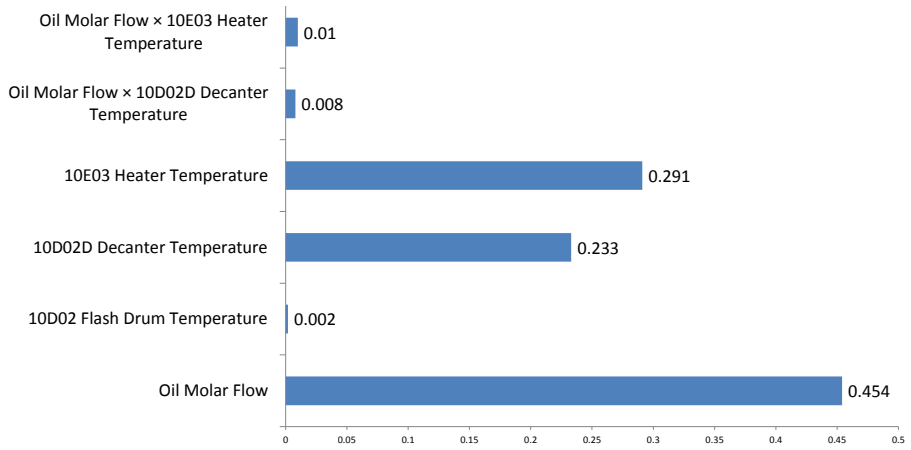
Figure 9: *Plot of residuals against molar flow of tripalmitin oil for heat duty of heater 10E03 produced for 11 inputs.*

5.2 Global sensitivity

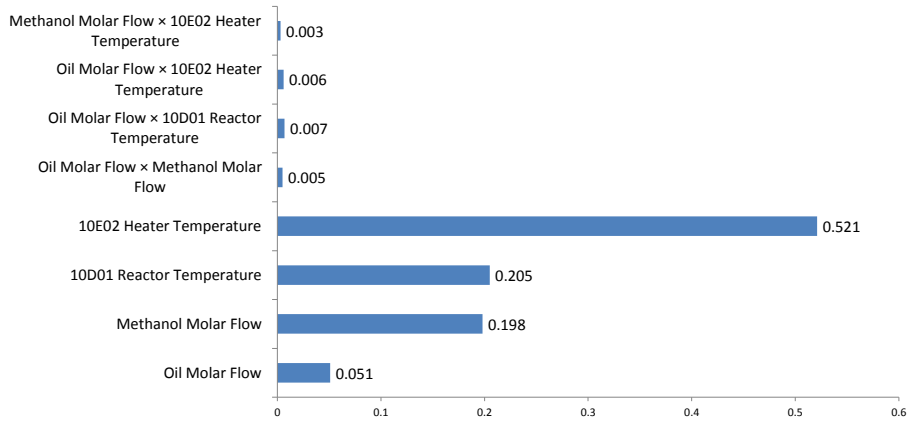
Global sensitivities of the heat duties of all equipment under consideration with respect to the 11 inputs produced by HDMR fitted over the entire domain are summarised in Figures 10 and 11. It can be seen that in all cases only 4 or fewer inputs have significant influence on a given output. Interaction terms have no significant effect in all of the cases. Heat duty of each device is significantly affected by its own operating temperature and operating temperature of a heating device directly upstream (given such exists). While molar flow of oil, main feedstock of the process, has significant effect on all heat duties (except that of the flash drum), molar flow of methanol only affects heat duty of heater 10E02. This is because heat capacity of oil is around 100 higher than that of methanol (1665.0 $\frac{J}{mol*K}$ and 79.5 $\frac{J}{mol*K}$ [31, 33]) and only in the flash drum there is significantly more methanol than oil.

Heat duty of heater 10E01 is primarily affected by its operating temperature and molar flow and temperature of incoming oil. Heat duty of heater 10E02 is primarily affected by its operating temperature, operating temperature of reactor 10D01 and molar flows of oil and methanol. Heat duty of heater 10E03 is primarily affected by its operating temperature, operating temperature of decanter 10D02D and molar flows of oil. Heat duty of reactor 10D01 is primarily affected by its operating temperature, operating temperature of heater 10E01 and molar flows of oil. Heat duty of flash drum 10D02 is primarily affected by its operating temperature and operating temperature of heater 10E02. Heat duty of decanter 10D02D is primarily affected by its operating temperature, operating temperature of flash drum 10D02 and molar flows of oil. Terms and variables not mentioned here have a negligible global sensitivity with respect to those outputs.

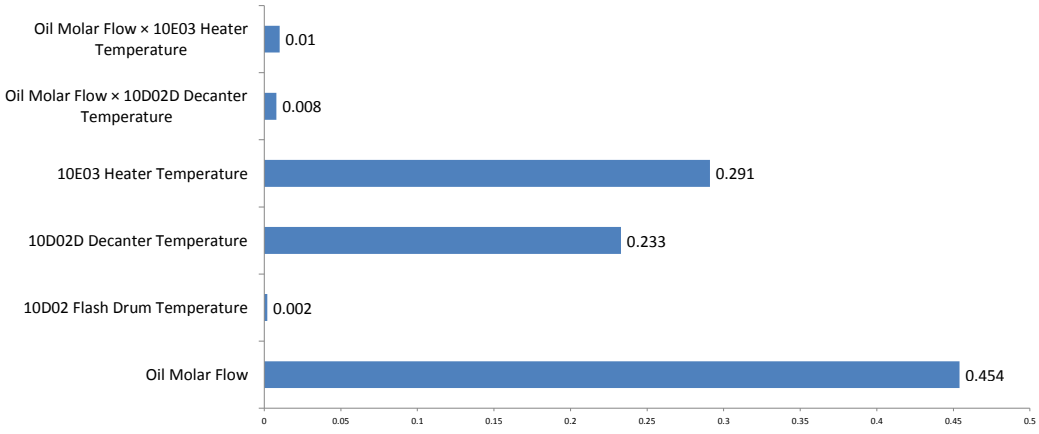
These observations show that when performing multi-dimensional analysis of heat duties within the system many terms in the surrogate models can be ignored due to insignificant influence. Thus calculation complexity and computational expense can be greatly reduced. Additionally, it shows which inputs are important when heat duties of the equipment needs to be controlled.



(a) 10E01 Heater - Heat Duty.

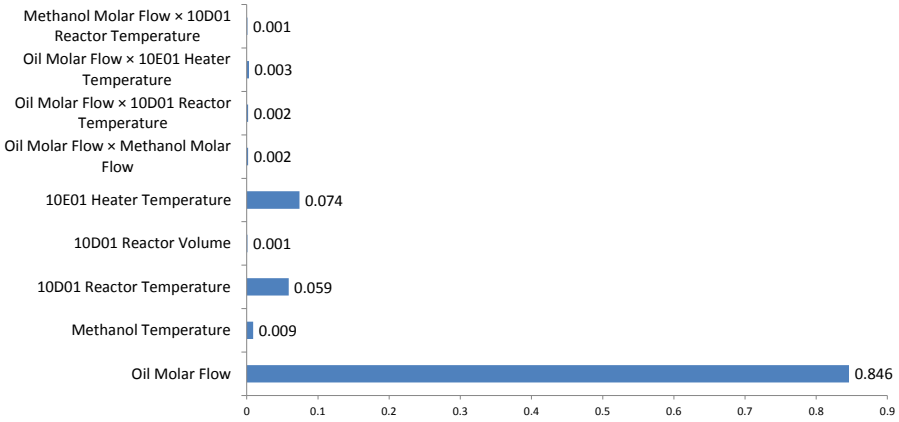


(b) 10E02 Heater - Heat Duty.

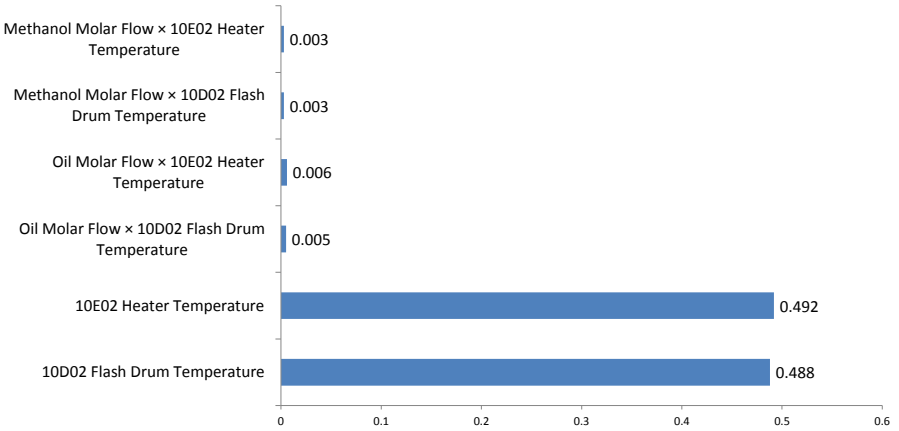


(c) 10E03 Heater - Heat Duty.

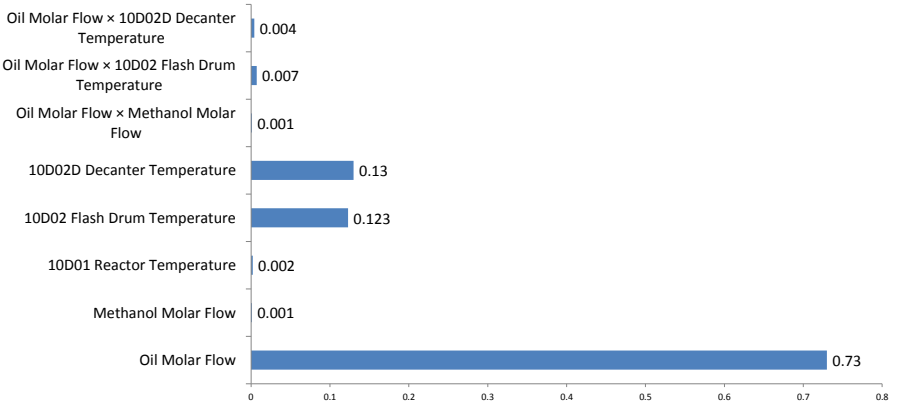
Figure 10: *Global sensitivities produced by 11-dimensional HDMR fit over the entire domain.*



(a) 10D01 Reactor - Heat Duty.



(b) 10D02 Flash Drum - Heat Duty.



(c) 10D02D Decanter - Heat Duty.

Figure 11: *Global sensitivities produced by 11-dimensional HDMR fit over the entire domain.*

6 Conclusions

This paper studies the possibility of describing Aspen Plus simulation of a biodiesel plant with surrogate models and assess their quality. The model under investigation includes a reaction and separation steps with auxiliary equipment and was solved for steady-state operation. Thus produced data was used to generate surrogate models describing relations between chosen inputs and outputs. A variety of scenarios was considered: 1, 2, 6 and 11 input variables were changed simultaneously, 3 different domain sizes of the input variables were considered and 2 different surrogate generation methods (polynomial and HDMR fitting). Each simulation produced 400 points per input variable used for fitting and calculating R^2 and $\bar{R}^2$. Test sets of points (100 points per dimension) were generated for calculating RMSD and residuals. Furthermore, a new automatic order selection method was implemented in HDMR generation with $\bar{R}^2$ used as an optimisation criterion.

A number of different behaviours were observed in the study. Most surrogates achieved at least a reasonable fit regardless of the domain size, number of dimensions and according to $\bar{R}^2$ and RMSD. Neither R^2 nor $\bar{R}^2$ could be used to effectively differentiate between the models as most achieve values in excess of 0.98. Also, it needs to be noted that the number of parameters within HDMR fit increases far slower than within polynomial fit when high-dimensional problems are considered. The most extensive HDMR fit (H10) had far fewer parameters than polynomial fits of order > 4. RMSD provides a reasonable measure for comparing accuracy of models. Fits P3 and H2b minimised RMSD and hence are the best fit for the duty of reactor 10D01 with respect to all 11 inputs. Increasing order of polynomial fit above 3 lead to poorer predictive powers, most likely due to overfitting the training data. RMSD increases exponentially for polynomial fits as the domain size of inputs increases. For fit H10 RMSD decreases from smallest to intermediate size and sharply increases from intermediate to largest size. Inclusion of 2nd order interaction terms accounted for a noticeable accuracy increase in terms of $\bar{R}^2$ and RMSD. It was observed that non-random features in residual plots are much more difficult to make out when multiple inputs were considered. Higher order polynomial fits may not be suitable for describing high dimensional, chemical data. For example, performance of polynomial fit of order 5 drops from being the best model to the worst as dimensionality increases from 1 to 11.

Global sensitivities of the heat duties of all equipment under consideration with respect to the 11 inputs were produced by HDMR fitted over the entire domain. It was observed that in all cases only 4 or fewer inputs had significant influence on a given output. Interaction terms had no significant effect in all of the cases. Heat duty of each device is significantly affected by its own operating temperature and operating temperature of a heating device directly upstream (given such exists). While molar flow of oil, main feedstock of the process, has significant effect on all heat duties (except that of the flash drum), molar flow of methanol only affects heat duty of heater 10E02. These observations show that when performing multi-dimensional analysis of heat duties within the system many terms in the surrogate models can be ignored due to insignificant influence. Thus calculation complexity and computational expense can be greatly reduced. Additionally, it shows which inputs are important when heat duties of the equipment needs to be controlled.

In the future a more complex chemical model should be considered as the simulation

used in this study was relatively simple. For example a number of interconnected models forming a feedback loop necessitating coupling surrogate models and solving them simultaneously. In order to further the goal of modelling eco-industrial parks chemical and electrical models and their interactions should be considered.

Acknowledgements

This project is funded by the National Research Foundation (NRF), Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] A. M. Aldawood, S. Mosbach, M. Kraft, and A. A. Amer. Multi-objective optimization of a kinetics-based HCCI model using engine data. Technical Report 2011-01-1783, Society of Automotive Engineers International, 2011.

[2] B. Allenby. Clean production in context: an information infrastructure perspective. *Journal of Cleaner Production*, 12:833–839, 2004. doi:10.1016/j.jclepro.2004.02.010.

[3] B. Allenby. The ontologies of industrial ecology? *Progress in Industrial Ecology - An International Journal*, 1-2:28–40, 2006. doi:10.1504/PIE.2006.010039.

[4] R. Aslett, R. J. Buck, S. G. Duvall, J. Sacks, and W. J. Welch. Circuit optimization via sequential computer experiments: design of an output buffer. *Journal of the Royal Statistical Society: Series C*, 47:31–48, 1998. doi:10.1111/1467-9876.00096.

[5] AspenTech. aspentech - aspen plus v8.6, 2015. URL `http://www.aspentech.com/products/engineering/aspen-plus/`. Date accessed: 12.01.2016.

[6] P. Azadi, G. Brownbridge, S. Mosbach, O. R. Inderwildi, and M. Kraft. Production of biorenewable hydrogen and syngas via algae gasification: A sensitivity analysis. *Energy Procedia*, 61:2767–2770, 2014. doi:10.1016/j.egypro.2014.12.302.

[7] P. Azadi, G. Brownbridge, S. Mosbach, A. J. Smallbone, A. Bhave, O. R. Inderwildi, and M. Kraft. The carbon footprint and non-renewable energy demand of algae-derived biodiesel. *Applied Energy*, 113:1632–1644, 2014. doi:10.1016/j.apenergy.2013.09.027.

[8] P. Azadi, G. Brownbridge, I. Kemp, S. Mosbach, J. S. Dennis, and M. Kraft. Microkinetic modeling of fischer-tropsch synthesis over cobalt catalysts. *ChemCatChem*, 7:137–143, 2015. doi:10.1002/cctc.201402662.

[9] I. F. Bailleul, P. L. Man, and M. Kraft. A stochastic algorithm for parametric sensitivity in Smoluchowski's coagulation equation. *Society for Industrial and Applied Mathematics Journal on Numerical Analysis*, 48:1064–1086, 2010. doi:10.1137/090758234.

[10] M. H. Bakr, J. W. Bandler, K. Madsen, and J. Soandergaard. Review of the space mapping approach to engineering optimization and modeling. *Optimization and Engineering*, 1:241–276, 2000. doi:10.1023/A:1010000106286.

[11] M. Baran and L. Bieniasz. Experiments with an adaptive multicut-HDMR map generation for slowly varying continuous multivariate functions. *Applied Mathematics and Computation*, 258:206–219, 2015. doi:10.1016/j.amc.2015.02.007.

[12] M. C. Bernardo, R. Buck, L. Liu, W. A. Nazaret, J. Sacks, and W. J. Welch. Integrated circuit design optimization using a sequential strategy. *IEEE Transactions on Computer-Aided Design*, 11:361–372, 1992. doi:10.1109/43.124423.

[13] M. Boix, L. Montastruc, C. Azzaro-Pantel, and S. Domenech. Optimization methods applied to the design of eco-industrial parks: a literature review. *Journal of Clean Production*, 87:303–317, 2015. doi:10.1016/j.jclepro.2014.09.032.

[14] A. Braumann, M. Kraft, and P. Mort. Parameter estimation in a multidimensional granulation model. *Powder Technology*, 197:196–210, 2010. doi:10.1016/j.powtec.2009.09.014.

[15] A. Braumann, P. L. Man, and M. Kraft. Statistical approximation of the inverse problem in multivariate population balance modeling. *Industrial and Engineering Chemistry Research*, 49:428–438, 2010. doi:10.1021/ie901230u.

[16] A. Braumann, P. Man, and M. Kraft. The inverse problem in granulation modeling-two different statistical approaches. *AIChE Journal*, 57:3105–3121, 2011. doi:10.1002/aic.12526.

[17] G. Brownbridge, A. J. Smallbone, W. Phadungsukanan, M. Kraft, and B. Johansson. Automated IC engine model development with uncertainty propagation. Technical Report 2011-01-0237, 2011.

[18] G. Brownbridge, P. Azadi, A. J. Smallbone, A. Bhave, B. J. Taylor, and M. Kraft. The future viability of algae-derived biodiesel under economic and technical uncertainties. *Bioresource Technology*, 151:166–173, 2014. doi:10.1016/j.biortech.2013.10.062.

[19] G. Brundtland, M. Khalid, S. Agnelli, S. Al-Athel, B. Chidzero, and L. Fadika. *Our Common Future: The World Commission on Environment and Development*. Oxford University Press, Oxford, 1987.

[20] W. L. Chapman, W. J. Welch, K. P. Bowman, J. Sacks, and J. E. Walsh. Arctic ice variability: model sensitivities and a multidecadal simulation. *Journal of Geophysical Research*, 99:919Ű935, 1994. doi:10.1029/93JC02564.

[21] N. Chen, K. Wang, C. Xiao, and J. Gong. A heterogeneous sensor web node meta-model for the management of a flood monitoring system. *Environmental Modelling & Software*, 54:222–237, 2014. doi:10.1029/93JC02564.

[22] M. Chertow. Industrial symbiosis: literature and taxonomy. *Annual Review of Energy and Environment*, 25:313–337, 2000. doi:10.1146/annurev.energy.25.1.313.

[23] E. Cimren, J. Fiksel, M. Posner, and K. Sikdar. Material flow optimization in byproduct synergy networks. *Journal of Industrial Ecology*, 15:315–332, 2012. doi:10.1111/j.1530-9290.2010.00310.x.

[24] CMCL Innovations. MoDS (model development suite), 2015. URL `http://www.cmclinnovations.com/mods/`. Date accessed: 12.01.2015.

[25] S. B. Crary. Design of computer experiments for metamodel generation. *Analog Integrated Circuits and Signal Processing*, 32:7–16, 2002. doi:10.1023/A:1016063422605.

[26] S. B. Crary, P. Cousseau, D. Armstrong, D. M. Woodcock, E. H. Mok, O. Dubochet, P. Lerch, and P. Renaud. Optimal design of computer experiments for metamodel generation using I-OPT. *Computer Modeling in Engineering and Sciences*, 1:127–140, 2000. doi:10.3970/cmes.2000.001.127.

[27] P. Desrochers. Cities and industrial symbiosis: Some historical perspectives and policy implications. *Journal of Industrial Ecology*, 5:29–44, 2001. doi:10.1162/10881980160084024.

[28] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Interscience, 3rd edition, 1998.

[29] J. Ehrenfeld and N. Gertler. Industrial ecology in practice. the evolution of interdependence at kalundborg. *Journal of Industrial Ecology*, 1:67–79, 1997. doi:10.1162/jiec.1997.1.1.67.

[30] J. E. Etheridge, S. Mosbach, M. Kraft, H.Wu, and N.Collings. A fast detailed-chemistry modelling approach for simulating the SI-HCCI transition. Technical Report 2010-01-1241, Society of Automotive Engineers International, 2010.

[31] V. Filatov and V. Afanas'ev. Differential heat-flux calorimeter. *Khimiya i Tekhnoliya Vody*, 35:97–100, 1992.

[32] A. Forrester, A. Sobester, and A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 1st edition, 2008.

[33] B. Freedman, M. Bagby, and H. Khoury. Correlation of heats of combustion with empirical formulas for fatty alcohols. *Journal of the American Oil Chemists' Society*, 66:595–596, 1989. doi:10.1007/BF02885455.

[34] P. Geyera and A. Schlueter. Automated metamodel generation for design space exploration and decision-making - a novel method supporting performance-oriented building design and retrofitting. *Applied Energy*, 119:537–556, 2014. doi:10.1016/j.apenergy.2013.12.064.

[35] W. A. Gough and W. J. Welch. Parameter space exploration of an ocean general circulation model using anisopycnal mixing parameterization. *Journal of Marine Research*, 52:773–796, 1994. doi:10.1357/0022240943076911.

[36] C. Hoffman. The industrial ecology of small and intermediate-sized technical companies: Implications for regional economic development. Technical report, Texas University, USA, 1971.

[37] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688, 2006. doi:10.1016/j.ijforecast.2006.03.001.

[38] R. Jin, X. Du, and W. Chen. The use of metamodeling techniques for optimization under uncertainty. *Structural and Multidisciplinary Optimization*, 25:99–116, 2003. doi:10.1007/s00158-002-0277-0.

[39] S. Joe and F. Y. Kuo. Notes on generating sobol' sequences. Technical report, University of New South Wales, 2008.

[40] J. C. Jouhauda, P. Sagautb, M. Montagnaca, and J. Laurenceaua. A surrogate-model based multidisciplinary shape optimization method with application to a 2D subsonic airfoil. *Computers & Fluids*, 36:520–529, 2007. doi:10.1016/j.compfluid.2006.04.001.

[41] I. Kantor, M. Fowler, and A. Elkamel. Optimized production of hydrogen in an eco-park network accounting for life-cycle emissions and profit. *International Journal of Hydrogen Energy*, 37:5347–5359, 2012. doi:10.1016/j.ijhydene.2011.08.084.

[42] M. Karlsson. The MIND method: a decision support for optimization of industrial energy systems - principles and case studies. *Applied Energy*, 88:577–589, 2011. doi:10.1016/j.apenergy.2010.08.021.

[43] C. A. Kastner, A. Braumann, P. L. W. Man, S. Mosbach, G. Brownbridge, J. Akroyd, M. Kraft, and C. Himawan. Bayesian parameter estimation for a jet-milling model using Metropolis-Hastings and Wang-Landau sampling. *Chemical Engineering Science*, 89:244–257, 2013. doi:10.1016/j.ces.2012.11.027.

[44] S. Keckler and D. Allen. Material reuse modeling: a case study of water reuse in an industrial park. *Journal of Industrial Ecology*, 2:79–92, 1999. doi:10.1162/jiec.1998.2.4.79.

[45] J. P. C. Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192:707–716, 2009. doi:10.1016/j.ejor.2007.10.013.

[46] D. L. Knill, A. A. Giunta, C. A. Baker, B. Grossman, W. H. Mason, R. T. Haftka, and L. T. Watson. Response surface models combining linear and euler aerodynamics for supersonic transport design. *Journal of Aircraft*, 36:75–86, 1999. doi:10.2514/2.2415.

[47] G. Li, S.-W. Wang, and H. Rabitz. Practical approaches to construct RS-HDMR component functions. *The Journal of Physical Chemistry A*, 106:8721–8733, 2002. doi:10.1021/jp014567t.

[48] Z. W. Liao, J. T. Wu, B. B. Jiang, J. D. Wang, and Y. R. Yang. Design methodology for flexible multiple plant water networks. *Industrial & Engineering Chemistry Research*, 46:4954–4963, 2007. doi:10.1021/ie061299i.

[49] P. L. Man, A. Braumann, and M. Kraft. Resolving conflicting parameter estimates in multivariate population balance models. *Industrial and Engineering Chemistry Research*, 65:4038–4045, 2010. doi:10.1016/j.ces.2010.03.042.

[50] W. J. Menz and M. Kraft. A new model for silicon nanoparticle synthesis. *Combustion and Flame*, 160:947–958, 2013. doi:10.1016/j.combustflame.2013.01.014.

[51] W. J. Menz, S. Shekar, G. Brownbridge, S. Mosbach, R. Körmer, W. Peukert, and M. Kraft. Synthesis of silicon nanoparticles with a narrow size distribution: A theoretical study. *Journal of Aerosol Science*, 44:46–61, 2012. doi:10.1016/j.jaerosci.2011.10.005.

[52] W. J. Menz, G. Brownbridge, , and M. Kraft. Global sensitivity analysis of a model for silicon nanoparticle synthesis. *Journal of Aerosol Science*, 76:188–199, 2014. doi:10.1016/j.jaerosci.2014.06.011.

[53] Microsoft. COM technical overview, 2015. URL `https://msdn.microsoft.com/en-us/windows/desktop`. Date accessed: 02.06.2015.

[54] J. S. Ming Pan, C. A. Kastner, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft. Applying industry 4.0 to the jurong island eco-industrial park. *Energy Procedia*, 150, 2015. doi:10.1016/j.egypro.2015.07.313.

[55] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley-Blackwell, 3rd edition, 2009.

[56] H. Rabitz and Ö. F. Alış. General foundations of high-dimensional model representations. *Journal Of Mathematical Chemistry*, 25:197–233, 1999. doi:10.1023/A:1019188517934.

[57] Y. Reich and S. Barai. Evaluating machine learning models for engineering problems. *Artificial Intelligence in Engineering*, 13:257–272, 1999. doi:10.1016/S0954-1810(98)00021-1.

[58] E. Roux and P. Bouchard. Kriging metamodel global optimization of clinching joining processes accounting for ductile damage. *Journal of Materials Processing Technology*, 213:1038–1047, 2013. doi:10.1016/j.jmatprotec.2013.01.018.

[59] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*.

[60] S. Shekar, M. Sander, R. Shaw, A. J. Smith, A. Braumann, and M. Kraft. Modelling the flame synthesis of silica nanoparticles from tetraethoxysilane. *Chemical Engineering Science*, 70:54–66, 2012. doi:10.1016/j.ces.2011.06.010.

[61] S. Shekar, A. J. Smith, A. Braumann, P. L. Man, and M. Kraft. A multidimensional population balance model to describe the aerosol synthesis of silica nanoparticles. *Journal of Aerosol Science*, 44:93–98, 2012. doi:10.1016/j.jaerosci.2011.09.004.

[62] T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17:129–150, 2001. doi:10.1007/PL00007198.

[63] A. J. Smallbone, A. Bhave, A. Braumann, M. Kraft, A. Dris, and R. McDavid. Moving toward establishing more robust and systematic model development for IC engines using process informatics. Technical Report 2010-01-0152, Society of Automotive Engineers International, 2010.

[64] I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7:86–112, 1967. doi:10.1016/0041-5553(67)90144-9.

[65] I. M. Sobol. Sensitivity estimation for Nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2:112–118, 1990.

[66] A. Vikhansky and M. Kraft. A Monte Carlo methods for identification and sensitivity analysis of coagulation processes. *Journal of Computational Physics*, 200:50–59, 2004. doi:10.1016/j.jcp.2004.03.006.

[67] A. Vikhansky and M. Kraft. Two methods for sensitivity analysis of coagulation processes in population balances by a Monte Carlo method. *Chemical Engineering Science*, 61:4966–4972, 2006. doi:10.1016/j.ces.2006.03.009.

[68] H. M. Wainwright, S. Finsterle, Y. Jung, Q. Zhou, and J. T. Birkholzer. Making sense of global sensitivity analyses. *Computers and Geosciences*, 65:84–94, 2014. doi:10.1016/j.cageo.2013.06.006.

[69] C. Watanabe. Industrial ecology: Introduction of ecology into industrial policy. Technical report, Ministry of International Trade and Industry (MITI), Tokyo, 1972.

[70] E. K. Y. Yapp, R. I. A. Patterson, J. Akroyd, S. Mosbach, E. M. Adkins, J. H. Miller, and M. Kraft. Numerical simulation and parametric sensitivity study of optical band gap in a laminar co-flow ethylene diffusion flame. *Combustion and Flame*, 2016. doi:10.1016/j.combustflame.2016.01.033.

[71] T. Ziehn and A. Tomlin. Global sensitivity analysis of a 3D street canyon model - Part I: the development of high dimensional model representations. *Atmospheric Environment*, 42:1857–1873, 2008. doi:10.1016/j.atmosenv.2007.11.018.