Blockchain, parameterisation and automated arbitrage applied to the chemical industry

Janusz Jerzy Sikorski

Churchill College

This thesis is submitted for the
degree of Doctor of Philosophy.

2019-10-10

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Blockchain, parameterisation and automated arbitrage applied to the chemical industry

Janusz Jerzy Sikorski

This thesis considers three scenarios related to chemical industry where the concepts of eco-industrial parks (EIPs), Industry 4.0, parameterisation, blockchain and arbitrage are brought together to explore the issues of simulation speed and accuracy, machine-to-machine (M2M) communication and automated participation in financial markets.

In the first scenario, a biodiesel plant flow sheet model is analysed and parameterised. The relations between 11 inputs typical to a biodiesel plant and its energy requirements are approximated using surrogate models, of which accuracy is assessed. Additionally, the effects of dimensionality, domain size and surrogate type on the accuracy are investigated and global sensitivities of the outputs are computed using High Dimensional Model Representation (HDMR). Most surrogate models achieved at least a reasonable fit regardless of the domain size and number of dimensions. It was observed that in all cases only 4 or fewer inputs have significant influence on any of the outputs and that the interaction terms have only minor effect on any one output.

In the second scenario, applications of blockchain technology related to Industry 4.0 are explored and an example where blockchain is employed to facilitate M2M interactions and establish a M2M electricity market in the context of the chemical industry is presented. Successful implementation of two electricity producers and one electricity consumer trading with each other over a blockchain-based network is presented.

In the third scenario, an automated arbitrage spotter is developed and applied to two cases: conversion of natural gas to methanol and crude palm oil to biodiesel. The spotter is designed to search for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product. It considers cost of storage and conversion (other feedstock, steam, electricity and other utilities) derived from physical simulations of the chemical process. In a profitable scenario up to 345.17 USD per tonne of biodiesel can be earned by buying contracts for delivery of crude palm oil in September 2018 and selling contracts for delivery of biodiesel in December 2018 in a ratio of 4 to 1.

Acknowledgements

I would like to thank the following entities:

- Department of Chemical Engineering and Biotechnology in the University of Cambridge for its financial support,
- Cambridge CARES Ltd. for its financial support,
- Prof. Markus Kraft for being my supervisor,
- Dr Sebastian Mosbach for being my advisor,
- Joy Haughton for being my advisor,
- CoMo group members, past and present, with special mentions for Danielka, Ra-, Schnu-Shnoo, Minimus, -do-, DJ Yapp, Jo Chen, Andrew, -mir and Mr Lindberg,
- My family for their support,
- Zhou Li for many stimulating discussions,
- Vishvak Kanaan for being an excellent travel companion.

Preface

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration except as specified in the text. The work was performed at the University of Cambridge and at the Cambridge Centre for Advanced Research and Education in Singapore between October 2014 and April 2018. No part of this thesis has been submitted for a degree to this or any other university.

This dissertation contains approximately 42000 words and 31 figures and tables. Some of the work in this dissertation has been published:

1. J. Sikorski, O. R. Inderwildi, M. Q. Lim, S. S. Garud, J. Neukäufer and M. Kraft (2019). Enhanced Procurement and Production Strategies for Chemical Plants: Utilizing Real-Time Financial Data and Advanced Algorithms. *Industrial and Engineering Chemistry Research*, doi:doi.org/10.1021/acs.iecr.8b02925.

2. L. Zhou, M. Pan, J. Sikorski, S. S. Garud, L. K. Aditya, M. J. Kleinelanghorst, I. A. Karimi and M. Kraft (2017). Towards an ontological infrastructure for chemical process simulation and optimization in the context of eco-industrial parks. *Applied Energy*, doi:10.1016/j.apenergy.2017.05.002.

3. J. Sikorski, J. Haughton and M. Kraft (2017). Blockchain technology in the chemical industry: Machine-to-machine electricity market. *Applied Energy*, doi:10.1016/j.apenergy.2017.03.039.

4. J. Sikorski, G. Brownbridge, S. S. Garud, S. Mosbach, I. A. Karimi and M. Kraft (2016). Parameterisation of a biodiesel plant process flow sheet model. *Computers and Chemical Engineering*, doi:10.1016/j.compchemeng.2016.06.019.

Janusz Jerzy Sikorski
2019-10-10

Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

The chemical industry is in the process of adopting digital technologies. These technologies offer opportunities, including data analysis and predictive capabilities such as equipment monitoring and predictive analytics, predictive quality and improved energy management. They can aid in achieving better understanding and optimisation of its activities ranging across all different levels: from workforce schedules and individual pieces of machinery to entire plants. They can also facilitate multi-objective analysis allowing inclusion of parameters such as environmental impact in addition to the traditional profit maximisation. This is important as environmental concerns such as reducing pollutants, creating cleaner manufacturing processes or reducing carbon footprints become ever more pressing (Technavio, 2017; Klei *et al.*, 2017; Beacham, 2017).

The process of adopting digital technologies presents a number of threats and challenges. For example, the danger of cyberwarfare, which requires companies to seriously consider physical security, industrial systems security, and data systems security within their organisations. For this reason, the chemical industry is torn between accelerating the digitisation process to reap its benefits and exposing itself on one hand, and treading carefully, but missing profitable opportunities on the other, especially given that this transition is expected to take years to complete.

The process of adoption is broadly described by the notion of Industry 4.0, which aims to supply smart products using smart methods and processes that use near-real-time data in the digital integration of value chains and facilitate seamless asset lifecycle information from plant concept to decommissioning (Leeuw, 2017). It introduces many concepts relevant to taking advantage of the aforementioned opportunities, including machine-to-machine (M2M) communication, cyber-physical systems (CPSs) and the Internet of Things (IoT) (Kleinelanghorst *et al.*, 2016; Kraft and Mosbach, 2010). M2M communication refers to the ability of industrial components to communicate with each other. CPSs can monitor physical processes, create virtual copies of the physical world and make decentralised decisions. The virtual copies could be realised as surrogate models (or metamodels or parameterised models), which are approximations of experimental and/or simulation data designed to provide answers when it is too expensive to directly measure the outcome of interest. IoT is a dynamic network where physical and virtual entities have identities and attributes and use intelligent interfaces.

This thesis considers how to utilise the ideas of Industry 4.0 to bring the benefits of digital technologies to eco-industrial parks. An eco-industrial park refers to 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 (Pan *et al.*, 2015, 2016; Kastner *et al.*, 2015). An example of an already implemented digital technology is improving maintenance using sensors attached to equipment. They monitor quality and throughput continuously, while local, real-time computing provides insight and interactions that can prevent equipment failures and breakdowns. Organisations employing this technology can react quicker to arising issues resulting in a higher overall equipment effectiveness and longer mean time between failures. The aforementioned concepts and technologies such as parameterisations and blockchain can be brought together to tackle the challenges encountered by the chemical industry while trying to exploit benefits from digitisation and automation.

1.2 Research questions explored in the thesis

The aim of this thesis is to investigate how to apply selected ideas from Industry 4.0 to the chemical industry and exploit further opportunities offered by eco-industrial parks. Due to the breadth of the topic, the thesis focuses on three specific research questions and builds proof-of-concept demonstrations. Hence the following research questions are posed:

- How can a typical process flowsheet model be parameterised to give a simplified surrogate model? Can the resulting surrogate models yield sufficient accuracy? What is required to progress towards full automation of generating such surrogate models?
- How can blockchain industry be employed within the chemical industry? Can it be used to leverage new opportunities in the context of eco-industrial parks?
- Do opportunities for financial arbitration over chemical commodities exist on futures markets?

1.3 Novel elements of the thesis

This thesis presents the following **novel developments**:

- A biodiesel plant flow sheet model is successfully analysed and parameterised to produce a range of possible surrogate models. Most of the surrogate models achieved at least a reasonable fit regardless of the domain size and number of dimensions. Effects of dimensionality, domain size and surrogate type on the accuracy are investigated. Additionally, global sensitivities of the outputs are computed.

- The use of blockchain to facilitate M2M communications and establish an M2M electricity market in the context of the chemical industry is proposed. A scenario where two electricity producers and one electricity consumer are trading with each other over a blockchain-based network is successfully implemented.

- An automated arbitrage spotter is developed and demonstrated. The arbitrage spotter uses market and physical data to search for opportunities to take advantage of a price difference between multiple markets. It is applied to two scenarios: conversion of natural gas to methanol and crude palm oil to biodiesel. The programme is designed to search for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product.

1.4 How the thesis connects

The scenarios presented in this thesis are intended to be included in J-Park Simulator (JPS) (Kleinelanghorst *et al.*, 2016; Pan *et al.*, 2015). The JPS is intended to demonstrate the potential of a modelling platform that integrates cross-domain real-time data sources and models to holistically assess the impact of various contingencies and provide clear, visual information to support decision-making and option analysis in the context of the industrial complex located on Jurong Island in

Singapore. The domains covered by JPS include, but are not limited to, chemical processes, electricity grids, building information management and safety controls. The development of the JPS is a central part of the research undertaken at the Cambridge Centre for Advanced Research and Education in Singapore and it remains an ongoing research project.

This thesis focuses on three scenarios relevant to the development of JPS. The first focuses on the biodiesel plant formerly located within the park, the second on a potential M2M electricity market connecting individual machines on the island and the third on enabling process lines to participate in the financial markets.

The remainder of this thesis is structured as follows. Chapter 2 gives a general introduction and literature review of the employed concepts and modelling techniques. In Chapter 3 the case study of using parameterisation techniques to a biodiesel plant flow sheet model is presented. In Chapter 4 applications of blockchain technology related to Industry 4.0 are explored and an example where blockchain is employed to facilitate M2M interactions and establish a M2M electricity market in the context of the chemical industry is presented. In Chapter 5 a system that gathers and analyses market and physical (from simulations) data about the relevant industrial processes and advises on potential investments is presented. Chapter 6 summarises the conclusions of the thesis and suggests areas for future work.

Chapter 2

Literature review

This chapter gives a general introduction and literature review of the employed concepts and modelling techniques. The notion of Industry 4.0, and the subservient concepts of the Internet of Things, machine-to-machine communications and cyber-physical systems, are presented and how it can enable holistic modelling of complex industrial networks such as those present in eco-industrial parks. Surrogate models and blockchain technology are described as potential solutions to issues such as speed, accuracy, storage space and connectivity. Research and applications of model predictive control (MPC) are described and discussed. The interface between the chemical industry and financial markets is explored and scenario where an independent chemical plant participates in the global financial markets is presented.

2.1 Eco-Industrial Parks

Every industrial actor strives towards better understanding and, ultimately, optimisation of any and all of its activities. 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 "laws of industrial ecology" (Renner, 1947), "industrial ecology" (Armstrong and Bashshur, 1958; Barnard, 1963; Hoffman, 1971; Watanabe, 1972; Allenby, 2004, 2006), "sustainable development" (Brundtland *et al.*, 1987) and "industrial symbiosis" (Chertow, 2000). Whence crystallised the concept of eco-industrial parks, 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 (Pan *et al.*, 2015).

An early example of industrial application of industrial symbiosis is the Verbund principle created and followed by BASF since their inception in 1865 (BASF, 2018). An example of organically developed EIP is the Kalundborg industrial park, which was started in 1959 (Ehrenfeld and Gertler, 1997). Primary academic interest stems from EIPs' ability to create more sustainable industrial activities through the use of localised symbiotic relationships (Boix *et al.*, 2015; Kastner *et al.*, 2015). 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 (Cimren *et al.*, 2012; Kantor *et al.*, 2012; Keckler and Allen, 1999; Liao *et al.*, 2007; Karlsson, 2011). 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. A substantial amount of supporting information which covers not only data, but also specific knowledge about the EIP in question is required in order to understand the system and create an accurate model. In this case, the required supporting information has the following features:

- Large volume: measurement data continuously produced by sensors within the EIP needs to be stored and analysed;
- Distributed storage: data from individual plants, machines or even sensors can be stored locally, but the relevant knowledge needs to be shared globally;
- Syntax heterogeneity: engineering information can be represented in a variety of, often incompatible, forms such as text documents in natural language, mathematical models, tables and diagrams, structured worksheets;
- Semantic heterogeneity: understanding of vocabulary and relationships within information in a specific engineering domain is often implicit and thus difficult to translate to another domain.

2.2 Industry 4.0

The ideal of a holistic model of an EIP could be brought about by exploiting key concepts of Industry 4.0, a movement of manufacturing technologies towards automation and data exchange which includes cyber-physical systems, the Internet of Things and M2M communication, cloud computing and cognitive computing (Pan *et al.*, 2015; Zhang *et al.*, 2017; Hermann *et al.*, 2016). M2M communication refers to the ability of industrial components to communicate with each other. CPSs can monitor physical processes, create virtual copies of the physical world and make decentralised decisions. IoT is a dynamic network where physical and virtual entities have identities and attributes and use intelligent interfaces. Each component of an EIP could become a CPS by having a virtual copy and participating in a network collectively created with other CPSs. Such a network could

facilitate the connection and control of the technical components intelligently by sharing information and understanding that trigger actions. Four design principles of Industry 4.0 are used to identify and implement its key concepts in real-life scenarios (Hermann *et al.*, 2016). Those include interconnection (including M2M and machine-to-people communications), information transparency, technical assistance (via data analysis and visualisation and CPSs physically supporting the operators) and decentralized decisions (i.e. the ability of CPSs to make low-level decisions autonomously). Among the challenges faced during implementation of Industry 4.0 are reliability and stability, IT security issues, unclear economic benefits and need to maintain the integrity of production processes. This work focuses on the software side of Industry 4.0 and considers a number of exploratory scenarios in order to address some of those challenges: surrogate models of a biodiesel plant as means of providing CPSs with their virtual component (see Section 2.3), blockchain technology as a test of M2M interactions within an internal network (see Section 2.5) and arbitrage spotter as an example of a chemical plant simulation interacting with a commodity market in an external network and yielding a clear economic benefit (see Section 2.6).

2.3 Surrogate models

One of the critical challenges of modelling EIPs is the creation of suitably accurate, fast and lightweight virtual components of CPSs. This could be solved with surrogate models (or metamodels or parameterised models), approximations of experimental and/or simulation data designed to provide answers when it is too expensive to directly measure the outcome of interest (Forrester *et al.*, 2008). 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.* (2001) 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.* (2003).

A book by Forrester *et al.* (2008) puts the process of data sampling and generating surrogate models into engineering perspective providing numerous case studies and MATLAB code to perform 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 (2009). A review and assessment of various sampling techniques is provided by Crary (2002). Reich and Barai (1999) focuses on assessment of machine learning techniques, artificial neural networks in particular, with case studies of modelling marine propeller behaviour and corrosion data analysis. An example of surrogate models bridging models of varying fidelity is provided by Bakr *et al.* (2000) where a surrogate maps data produced by fine and coarse physical models in order to accelerate optimisation of the fine model. Surrogate models are widely employed in engineering and science for space exploration (Gough and Welch, 1994; Geyera and Schlueter, 2014), modelling (Knill *et al.*, 1999; Crary *et al.*, 2000; Chen *et al.*, 2014), sensitivity analysis (Azadi *et al.*, 2014a; Chapman *et al.*, 1994; Gough and Welch, 1994; Menz *et al.*, 2014; Jouhauda *et al.*, 2007), parameter estimation (Kastner *et al.*, 2013; Bailleul *et al.*, 2010; Braumann *et al.*, 2010a), optimisation in areas ranging from circuit design through nanoparticle synthesis to flood monitoring (Bernardo *et al.*, 1992; Aslett *et al.*, 1998; Roux and Bouchard, 2013). A number of studies addressed application of surrogates to

process flow sheet models. Caballero and Grossmann (2008) replace the computationally expensive subsystems of a flow sheet with Kriging surrogates to speed up optimisation. Hasan *et al.* (2012); First *et al.* (2014); Hasan *et al.* (2013); Nuchitprasittichai and Cremaschi (2013); Boukouvala and Ierapetritou (2013) guide sampling of an expensive rigorous model using Kriging surrogates to reduce computational time required for optimisation. Fahmi and Cremaschi (2012) optimise a design of a biodiesel production plant by replacing all subsystems in a process flow sheet model with surrogate models based around artificial neural networks (ANNs) and solving thus defined mixed-integer non-linear problem. Henao and Maravelias (2011) propose a systematic method for creating surrogate models of chemical engineering systems and arranging them into a solvable network (superstructure). The study focuses on ANNs as a base for their surrogate models and describes how a superstructure can be optimised. Kong *et al.* (2016) employ some of the concepts developed in Henao and Maravelias (2011) for design optimisation of a chemical plant with heat integration and an attached utility plant. This paper includes a case study of non-enzymatic ethanol production from biomass.

2.4 Model predictive control (MPC)

MPC is a class of control techniques of which main advantages are its use of step response data and the ability to handle hard constraints explicitly through on-line optimization. MPC comes in a number of forms including dynamic matrix control (DMC) (Cutler and Ramaker, 1980), model algorithmic control (MAC) (Rouhani and Mehra, 1982), and internal model control (IMC) (García *et al.*, 1989). MPC has a demonstrated effectiveness in industrial applications and hence is widely employed (Nunes, 2001). It is important in the context of Industry 4.0 as it is intended as or already employed as the software component of cyber-physical systems as it can handle models of varying complexity with constraints and allows for previewing information and different problem formulations (Raman *et al.*, 2014; Shi *et al.*, 2015; Kasparick and Wunder, 2018; Lucia *et al.*, 2016; Cicconi *et al.*, 2017). MPC is distinct from surrogate models as they are a multivariable control algorithm that employs an internal dynamic model of the controlled process and

an optimisation cost function while taking into account a history of past control moves (Nikolaou, 1998). Surrogate models can be used within MPC, e.g. explicit MPC uses piecewise affine functions computed offline, however they are an engineering method for approximating an input-output behaviour of a process that is difficult or expensive to measure directly.

MPC techniques stem from the linear quadratic regulator (LQR) designed by Kalman in the early 1960s (Kalman, 1960) to minimise an unconstrained quadratic objective function of states and inputs. From there the field was explored in various directions. Model Predictive Heuristic Control (later known as MAC) was presented by Richalet *et al.* (1978). Garcia and Morshedi (1986) used quadratic programming in quadratic dynamic matrix control to solve the constrained open-loop optimal control problem where the system is linear, the cost quadratic, the control and state constraints are defined by linear inequalities. Keyser and Cauwenberghe (1985) developed Extended Prediction Self–Adaptive Control which uses a constant control signal starting from the present moment while using a sub-optimal predictor. Based on that Clarke *et al.* (1987a,b) developed Generalized Predictive Control (GPC).

Another variety of MPC is the explicit MPC (eMPC) which allows fast evaluation as the solution to the considered control problem is calculated offline (Bemporad *et al.*, 2002). The solution commonly comes in the form of a piecewise affine function (PWA), a piecewise function which preserves points, straight lines and planes. In those cases, an eMPC controller needs to store coefficients of the PWA for each a control region and coefficients of the parametric representations of all the regions (García *et al.*, 2012). In order to compute the optimal solution, the region containing the current state needs to be determined and then the PWA needs to be evaluated using the PWA coefficients stored for all regions.

2.5 Blockchain technology

Section 2.1 introduced the key concepts of Industry 4.0 in the context of eco-industrial parks including cyber-physical systems, machine-to-machine commu-

nications and the Internet of Things. Implementation of those concepts in the industry could be aided by blockchain technology. For example, blockchain could be used to facilitate M2M commodity (e.g. electricity) trading. In the example of electricity traded on a wholesale market (as in, for example, USA, Australia, New Zealand, many European countries and Singapore (Green, 2008)) such a system could reduce the overhead costs of the traditional trading practice and increase speed of transaction settlements. Those costs include administration associated with billing, reconciliation, hedging contracts and purchase agreements, which may constitute a significant part of electricity price (e.g. in the UK it is 16% (Energy UK, 2015) and in Australia approximately an eighth (EnergyAustralia, 2016) depending on the place e.g. in Tasmania 12.2% (Aurora Energy Pty Ltd, 2017)). Additionally, two extensive reports on the application of blockchain technology in the energy sector by Burger *et al.* (2016); Hasse *et al.* (2016) describe potential use cases and obstacles, including legislative, that need to be overcome before the technology can be widely introduced.

Blockchain is a type of distributed, electronic database (ledger) which can hold any information (e.g. records, events, transactions) and can set rules on how this information is updated (Condos *et al.*, 2016). It continually grows as blocks (files with data e.g. transactions) are appended and linked (chained) to the previous block using a hash (the chaining is visualised in Fig. 2.1 using Bitcoin as an example). The hash is produced by running contents of the block in question through a cryptographic hash function (e.g. Bitcoin uses Secure Hash Algorithm - 256 bit, SHA-256). An ideal cryptographic hash function can easily produce a hash for any input, but it is difficult to use the hash to derive the input. Additionally, any changes in the original data should result in extensive and seemingly uncorrelated changes to the hash (Rogaway and Shrimpton, 2004; Lewis, 2015). Finally, it should be infeasible for two different inputs to result in the same hash. Using the cryptographic hashes in this manner ensures that in order to alter an entry in a past block all subsequent blocks also need to be altered (Rogaway and Shrimpton, 2004; Lewis, 2015). The ledger is validated and maintained by a network of participants (nodes) according to a predefined consensus mechanism (a set of rules allowing the network to reach a global agreement (Buterin, 2014)) so no single

centralized authority is needed. Multiple (but not necessarily all) nodes hold a full copy of the entire database.

Blockchain technology is relatively new, continues to evolve and comes in many different shapes and forms. In this work Bitcoin is used as a case study as it is the most well-known and successful implementation of blockchain technology. Bitcoin is a payment system based on a permissionless (i.e. anyone can read or write to the chain) blockchain maintained by a peer-to-peer network (P2P) (Nakamoto, 2009a). It features its native currency (bitcoin or BTC), a proof-of-work consensus mechanism (note that there exist other types of consensus mechanisms; for more see A.2), timestamped blocks not larger than 1 MB (number of transactions per block varies depending on their size), anonymity, a financial incentive to publish blocks, optional transaction fees, a cap of the total BTC supply and BTC fungibility. The blocks primarily record BTC transactions, although additional data can also be included. An example of Bitcoin's block and its contents can be viewed in Fig. 2.2 and 2.3, respectively. A transaction is a transfer of BTC from a wallet address (or addresses) to another wallet address (or addresses). For creation transactions, only a receiving wallet is required. Wallets are public representations of the public and private key pairs that are used to store and transfer coins. One or more such key pairs are generated for each participant so business can be conducted in a secure and anonymous manner. The keys are a result of an encryption method called public-private key cryptography, which uses pairs of parameters: public and private. A public key can be used to verify that a message was created by an owner of the paired private key (verification of a digital signature) and to encrypt a message such that only the aforementioned owner can decrypt.

Bitcoin employs a proof-of-work consensus mechanism where the ability to verify and publish transactions is dependent on the computing power of a node (Nakamoto, 2009a). In order to publish a block, a node is required to complete the following steps:

1. Build a candidate block using valid transactions (i.e. compatible with the rest of the chain) from among the submitted transactions.

2. Calculate a hash of the block header using SHA-256 and compare it with

the current target (a specific number of leading zeros; for more information see *Hash target* in A.1), which is imposed by Bitcoin's protocol.

3. If the hash is not correct, the nonce of the header (an arbitrary number in the header) will be repeatedly altered until a solution is found or the target is changed (which means that another node's block was added to the chain).

4. If the hash is correct, the block is broadcast to the Bitcoin network.

5. If majority of the network (weighted by computing power) accepts the block it is permanently added to the chain and the publisher is rewarded with newly created BTCs.

6. If another node's block is added to the chain, the current block will be discarded entirely, and the process needs to start all over again.

Note that in a case where multiple suitable blocks are broadcast almost simultaneously, the chain will temporarily split into two or more branches (forks) which will be pursued until one is backed by a majority of the network. Bitcoin's protocol ensures that a block is added to the chain roughly every 10 minutes (ideally 2016 blocks would be added every 1209600 seconds) by adjusting the difficulty of the hash target (Nakamoto, 2009a). However, this mechanism results in significant confirmation latency (order of tens of minutes) and can be resource exhaustive.

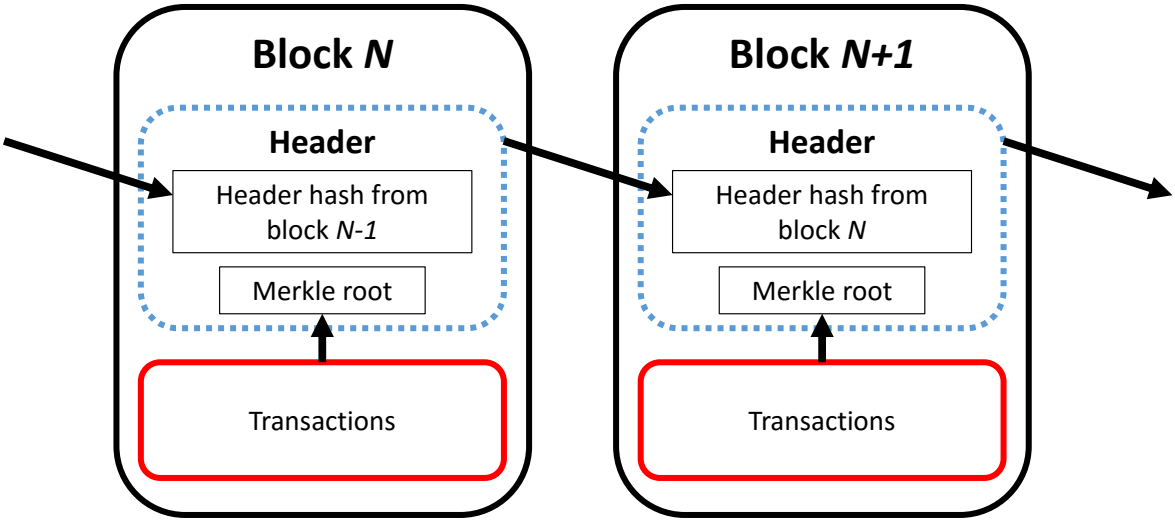


Figure 2.1: *Chaining of the Bitcoin blocks (adapted from Nakamoto (2009a)). Note that merkle root is a hash based upon all transactions in a block (for more details see entry "Merkle tree" in A.1).*

Block #438995

Summary	
Number Of Transactions	1527
Output Total	12,726.25791485 BTC
Estimated Transaction Volume	965.70420589 BTC
Transaction Fees	0.41251118 BTC
Height	438995 (Main Chain)
Timestamp	2016-11-15 08:26:59
Received Time	2016-11-15 08:26:59
Relayed By	GBMiners
Difficulty	254,620,187,304.06
Bits	402936180
Size	749.146 KB
Version	536870912
Nonce	173727158
Block Reward	12.5 BTC

Hashes	
Hash	000000000000000000042b6b0a7bfa7f43b648167a5c0547d6b3a676e23d4c226
Previous Block	000000000000000000023522e6372109dd28f97a86eef46a54cf15199c005e6a85
Next Block(s)	000000000000000000257616ab8ed39e0e7bca056d13faf1652075a34e335e7ed
Merkle Root	4a6be460a67526fd576788cf737713ae614003989706a0675e9d946ec58ca030

Network Propagation (Click To View)

Figure 2.2: *General information about Bitcoin block no. 438995 (Nakamoto, 2016).*

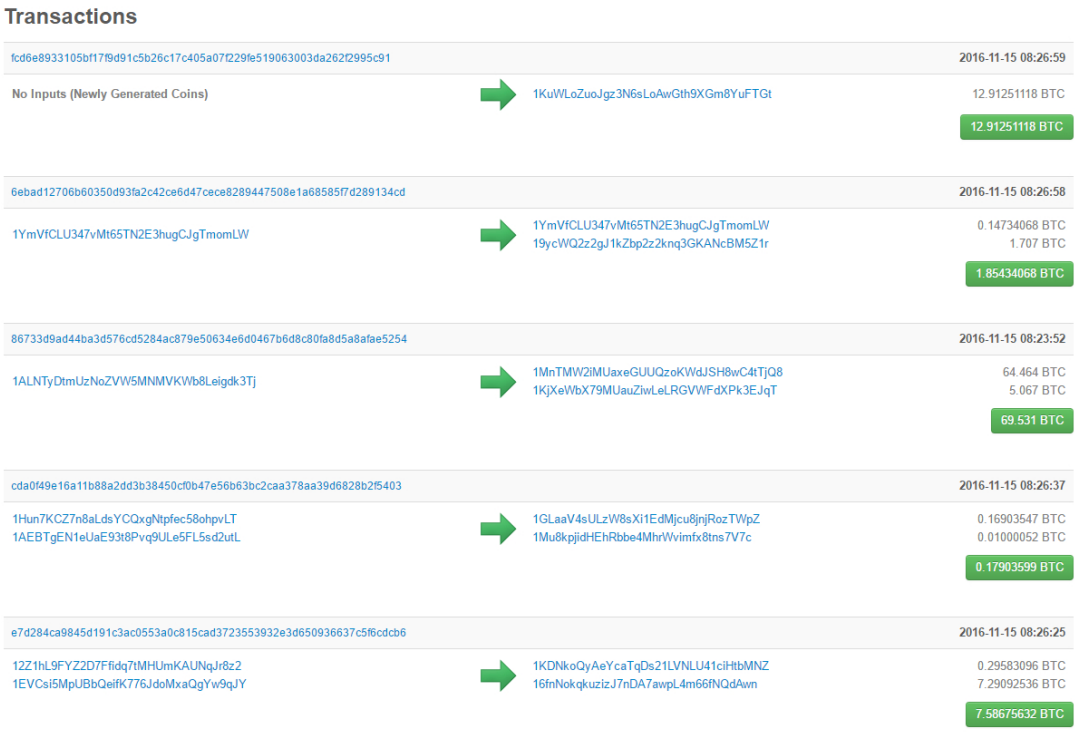


Figure 2.3: *Sample of transactions from Bitcoin block no. 438995 (Nakamoto, 2016).*

Explored areas and applications

Blockchain is yet to be fully explored in the academic literature, particularly in relation to the chemical industries. The review of the literature that informed this section therefore included technical reports, industrial and governmental position papers and news articles, which provide better access to the latest work in several areas.

The findings were divided into the following areas:

Explored areas

Security and privacy

Wasted resources and usability

Applications

Record-keeping and contract enforcement

The Internet of Things

Security and privacy

Security and privacy are among the core issues of blockchain technology as applied to digital currencies and at the same time the most explored areas. The main issues of security include a possibility of 51% attack, which involves attackers collectively controlling majority of the network, scams (e.g. Ponzi scams, mining scams, scam wallet, fraudulent exchanges) and distributed denial-of-service (DDoS) attacks on exchanges and mining pools. A degree of privacy is introduced as every participant may use one or more anonymous wallets. However, it is still possible to uncover information on the wallet owners. For example, Koshy *et al.* (2014) managed to map a subset of Bitcoin addresses to IP addresses by monitoring and analysing transaction traffic.

A comprehensive review by Yli-Huumo *et al.* (2016) found that the majority of research publications are concerned with this area. These issues are also addressed

by Koblitz and Menezes (2016) who describe two solutions to the problem of creating a digital currency with the advantages of physical cash, namely an elliptic-curve-based version of a construction provided by Brands (1993) and Bitcoin. A detailed description of the mathematics and necessary protocols (setup, signature, withdrawal, payment, deposit and double-spending prevention) for currency systems based on cryptographic hash functions is provided. Applications specifically addressing the issues include CoinParty (Ziegeldorf *et al.*, 2015), CoinShuffle (Decker and Wattenhofer, 2014), Zerocash (Sasson *et al.*, 2014) and Enigma (Sasson *et al.*, 2014). Zerocash is a ledger-based digital currency which allows user identities, transaction amounts and account balances to be hidden from public view, but still with the ability to quickly and efficiently facilitate transactions (not exclusively financial). Enigma combines blockchain and off-blockchain data storage to construct a personal data management platform focused on privacy.

Wasted resources and usability

Maintaining the most popular blockchain network consumes significant amounts of energy on calculations which have no meaning other than the maintenance. According to O'Dwyer and Malone (2014) in 2014 the power used for Bitcoin mining was comparable to Ireland's electricity consumption. Furthermore, increasing accessibility of blockchain technology (e.g. via more user-friendly application programming interface, API) should increase its exposure to areas other than technical computer science and thus help to alleviate the problem of wasted resources and many others.

The review by Yli-Huumo *et al.* (2016) identified eight papers focused on the problems of wasted resources and usability (four each). The applications aimed at improving Bitcoin's usability include BitConeView (Kishigami *et al.*, 2015) and BitIodine (Spagnuolo *et al.*, 2014). English *et al.* (2016) demonstrate how Semantic Web and blockchain technology can enhance each other: the former could facilitate implementation of the latter for several novel applications (e.g. Industry 4.0 platforms for online education or for supply chain management), while the latter could contribute towards the realization of a more robust Semantic Web (for a definition see A.1). An ontology for capturing data within a blockchain was created in order to increase usability of the technology, to facilitate a shared

understanding of this technology between humans and to enable interlinking with other Linked Data (for a definition see A.1) to conduct formal reasoning and inference. A number of consensus mechanisms were developed which are not primarily based on performing intensive computations and typically enjoy lower electricity consumption for a similar blockchain network. Those include the following mechanisms: proof-of-stake (Vukolić, 2016; O'Dair *et al.*, 2016; Goodman, 2014), deposit-based (Zamfir, 2015; Sompolinsky and Zohar, 2013), Byzantine agreement (Higgins, 2015; Vukolić, 2016; Castro and Liskov, 1999; Maziéres, 2015) and a rotation scheme (Greenspan, 2015). Goodman (2014) and Greenspan (2015) employ those concepts in their projects, respectively, Tezos and MultiChain. The first is a generic and self-amending crypto-ledger employing proof-of-stake consensus mechanism. The second is an off-the-shelf platform for the creation and deployment of private blockchains aiming to facilitate easy deployment of blockchain in the organisations of the financial sector.

Application to record-keeping and contract enforcement

Keeping and creation of records and enforcement of contracts are among the most promising applications of blockchain technology across a wide range of industries from finance to construction. In the context of Industry 4.0 such capability could facilitate logging and sharing data (e.g. emissions) and advanced M2M trading (e.g. bonds).

Watanabe *et al.* (2015) presents a blockchain-based system for confirming contractor consent and archiving the contractual documents. Cardeira (2015) argues that employment of the blockchain technology might remedy the major problems of construction industry, namely timing and guarantee of payments, via smart contracts (for a definition see A.1). Smart contracts would ensure that sufficient funds are available to finance the project and that everyone is paid in a timely manner. The governmental report by Condos *et al.* (2016) assesses the opportunities and risks of blockchain technology from the perspective of the American state of Vermont. It is identified that a valid blockchain could be a reliable way of confirming the party submitting a record, the time and date of its submission, and the contents of the record at the time of submission. The final conclusion states that currently the costs and challenges associated with the technology for

Vermont's public recordkeeping outweigh the identifiable benefits. Korpela *et al.* (2017) explore the potential effects of the technology on the supply chain management across a number of industries. Holotiuk *et al.* (2017); Lindman *et al.* (2017) analyse the, potentially disruptive, effects of cryptocurrencies and blockchain on the payments industry. O'Dair *et al.* (2016) discuss various applications in the music industry, including a networked copyright database, efficient royalty payment system and provision of access to alternative funding sources for artists. Organisations using blockchain in the music industry include Bittunes (Edhouse, 2013), Dot Blockchain Music (Rogers, 2016) and Mycelia (Heap, 2016). In the local infrastructure field, a number of projects have adopted blockchain technology to enable residents to choose where to buy renewable energy from (their neighbours or others) and to support communities in keeping energy resources local, reducing dissipation and increasing micro- and macro-grid efficiency. Those include GridSingularity (GridSingularity, 2016; Nguyen, 2016), LO3 (LO3 Energy, 2016) and SolarCoin (SolarCoin, 2016), as described by a number of technology news (Rutkin, 2016; Lacey, 2016; Lilic, 2015b,a).

Furthermore, a number of applications were found in finance including: chain.com (Ludwin, 2014) (deployment of blockchain networks); Augur (Peterson, 2015) (prediction trading); Everledger (Kemp, 2015) (certification of precious gemstones); Stroj (Wilkinson and Quinn, 2015) (sharing service for internet bandwidth and spare disk space); Namecoin (Kraft, 2011)(an open-source Internet infrastructure such as DNS and identities).

Application to the Internet of Things

Employment of blockchain technology for the purpose of introducing transactional functionality to the IoT has been addressed by a number of ideas and applications including:

- IoTcoin (Zhang and Wen, 2015) - a currency based on BTC intended to facilitate proof of ownership and exchanges of IoT commodities (e.g. sensor data or smart property).
- Community currency (Vandervort *et al.*, 2015) - a proposed crypto-currency issued by a non-government entity to serve the economic or social interests

of a group of people.

- Enigma (Zyskind *et al.*, 2015) - whilst primarily a blockchain-based platform for personal data protection, an assessment by Atzori (2016) deems it a suitable solution for the issue of privacy in the IoT.

- IOTA (Buntix, 2014; Atzori, 2016) - a crypto-currency developed for the IoT and M2M economy based on Tangle, a blockchain "without blocks" (i.e. each transaction is confirmed separately).

- ADEPT (Autonomous Decentralized Peer-to-Peer Telemetry) (Panikkar *et al.*, 2015; Atzori, 2016) - an architecture designed for a dynamic democracy of objects connected to a universal digital ledger, which provides users with secure identification and authentication.

- Filament (Clift-Jennings, 2016) - a technological framework developed to enable devices to hold unique identities on a public ledger and to discover, communicate and interact with each other in an autonomous and distributed manner.

2.6 Commodity futures market and the chemical industry

Most organisations participating in industrial parks are profit-seeking companies. Sometimes investment and energy-saving opportunities present themselves but cannot be exploited due to significant risks and absence of relevant data or a physical model with sufficiently strong predictive capabilities. For example, a chemical plant, with spare production capacity, could make additional earnings by exploiting the price differences between its feedstock and product on the commodity markets. However, straying from the established throughput without an accurate simulation may lead to decreased product quality or even damaging equipment.

Industry 4.0 introduces many concepts relevant to taking advantage of the aforementioned opportunities, including M2M communication, cyber-physical systems

and the Internet of Things (Kleinelanghorst *et al.*, 2016; Kraft and Mosbach, 2010). A system could be established that gathers and analyses market (local and global) and physical (from sensors and simulations) data about the relevant industrial processes and advises on potential investments and energy savings. It is conceivable that in the future an entire production plant or even an entire industrial park would autonomously seek and fulfil such opportunities.

Furthermore, exploring and exploiting such possibilities within an industrial park may encourage closer cooperation of participants' plants leading to a transformation into an eco-industrial park. An EIP is 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 minimize environmental impact while simultaneously increasing business success (Pan *et al.*, 2015, 2016; Kastner *et al.*, 2015).

Numerous studies have been written on trading commodities and their futures on an exchange, for example, Garcia and Leuthold (2004); Szakmary *et al.* (2010); Fung and Hsieh (2001); Campa (1994). A number of publications considers the interactions between the production and the commodity markets. Smith and Stulz (1985) examine the reasons why firms hedge and what risks do they choose to hedge and develop a theory of the hedging behaviour of value-maximizing corporations. Bjorgan *et al.* (1999) study the issues of financial risk management in the energy sector and explores impact of financial contracts (such as futures) for scheduling policies of the companies in the industry. The paper by Tanlapco *et al.* (2002) examines risk-minimizing hedging strategies using futures contracts in an electricity market and finds that the use of electricity futures contracts is superior to using other related futures contracts such as crude oil. Spinler *et al.* (2003) present a theoretical analysis of options contracts for physical delivery which shows how spot market price risk, demand and cost risk can be shared between buyer and seller. They conclude that such a contingency may complement financial risk management instruments capital-intensive firms such as those in the chemical industry. Ding *et al.* (2007) explore integrated operational and financial hedging decisions faced by a global firm. The study found that finan-

cial hedging strategy ties closely to, and can have both quantitative and qualitative impact on, the firm's operational strategy. The publication by Kannegiesser *et al.* (2009) presents a planning model for coordinating sales and supply decisions for commodities in a chemical industry. The impact of elasticities, variable raw material consumption rates and price uncertainties on planned profit and volumes are demonstrated.

A number of studies develop and apply modelling techniques in order to aid companies with financial risk management. Ryu (2006) discusses two multi-period planning strategies used to minimise negative impact of varying conditions on a company's profitability. One is to modify the external condition, e.g. demands, and the other is to expose explicit constraints limiting capacity expansion. Park *et al.* (2010) and Ji *et al.* (2015) present financial risk management methods in the petrochemical industry. The former demonstrates a two-stage stochastic programming framework for operational planning and financial risk management of a refinery. The method optimised the contract sizes (long-term, spot and futures) and the plant's operational plan. The latter presents a one-stage stochastic programming model for the integration of operational hedging and financial hedging strategy in the crude oil procurement process subject to oil price fluctuation. This approach uses Conditional Value-at-Risk as the risk measure and considers futures contracts, put options and call options during optimisation. A publication by Longinidis *et al.* (2015) presents a supply chain network design model that yields the optimal configuration under a variety of exchange rate realizations and integrates operational hedging actions that mitigate exchange rate risk. Kwon *et al.* (2017) describe and demonstrate a two-stage programming framework for maximizing profit of a petrochemical company. The two stages are hedge trading (minimising costs of raw materials) and production planning (maximising sales profits). The procedure simultaneously optimises timing, amount, and price of raw materials and the strategies for facility operation and product sales. An example of a Korean petrochemical company is used to present that this approach can improve profitability.

Chapter 3

Surrogate models of a biodiesel plant

This chapter investigates the extent to which surrogate models can be used to parameterise typical input-output relations in a process flow sheet model. The investigation is performed in the context of a biodiesel plant on Jurong Island. It is shown that accurate surrogate models can be generated and can be used as replacements for the detailed flow sheet model in the context of cyber-physical systems that require rapid and stable model execution. A variety of scenarios were considered: 1, 2, 6 and 11 input variables were changed simultaneously, 3 domain sizes of the input variables were considered and 2 different surrogates (polynomial and 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. The chapter concludes with a discussion of continuing the presented work.

3.1 Methodology

Aspen Plus (a process modelling and optimisation software), MoDS (a software tool designed to analyse black-box models) and custom-made Python 3.4 (a programming language for general-purpose programming) and R 3.2.2 (a software tool for statistical computing and graphics) scripts were employed together in order to produce and analyse surrogate models of the biodiesel plant simulation (described in detail in 5.1.1). In this study the following procedure was used:

1. MoDS generated a Sobol sequence (an example of quasi-random, space-filling sequences) of input data for user-specified variables within the process flow sheet model.
2. Python script altered model's input data, evaluate the simulation and produce a file with user-specified output values.
3. MoDS retrieved values of user-specified outputs from the file.
4. MoDS scanned data for errors and corrected them (e.g. a reactor reaching unrealistic temperatures).
5. MoDS generated polynomial and HDMR surrogates that map inputs to outputs. Only a subset of the data was used in this step, while the rest was used for analysis of the surrogates' accuracy.
6. R was used for postprocessing.

This workflow is visualized in Figure 3.1. More detail is available in the rest of the chapter.

3.1.1 Codes and data

This work can be reproduced by following this workflow and using the instructions and code provided under the following link: https://doi.org/10.17863/CAM.33680.

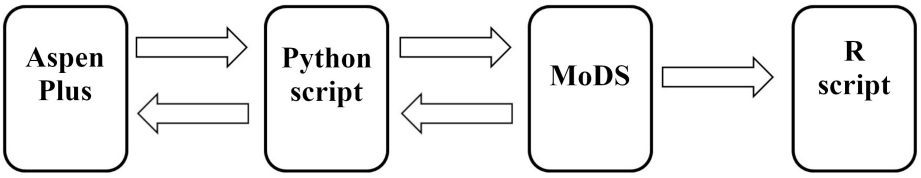


Figure 3.1: *Study workflow for Chapter 3.*

3.2 Description of the biodiesel plant model

3.2.1 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 3.2. The final fuel, fatty acid methyl ester, is produced via trans-esterification pathway where triglycerides react with methanol to form methyl ester and glycerine 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 tripalmitin 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 3.1 and 3.2. Plots of heat duties of various equipment against molar flow of tripalmitin oil can be seen in Figure 3.3.

Table 3.1: *Biodiesel surrogate: Input variables.*

Name	Lower bounds	Upper bounds	Operating point
Molar flow of tripalmitin oil (kmol/hr)	20, 22.5, 25	40, 37.5, 35	30
Temperature of tripalmitin oil (^{o}C)	20, 22.5, 25	40, 37.5, 35	30
Operating temperature of CSTR 10D01 (^{o}C)	44, 49, 54	64, 64, 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 3.2: *Biodiesel surrogate: 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)

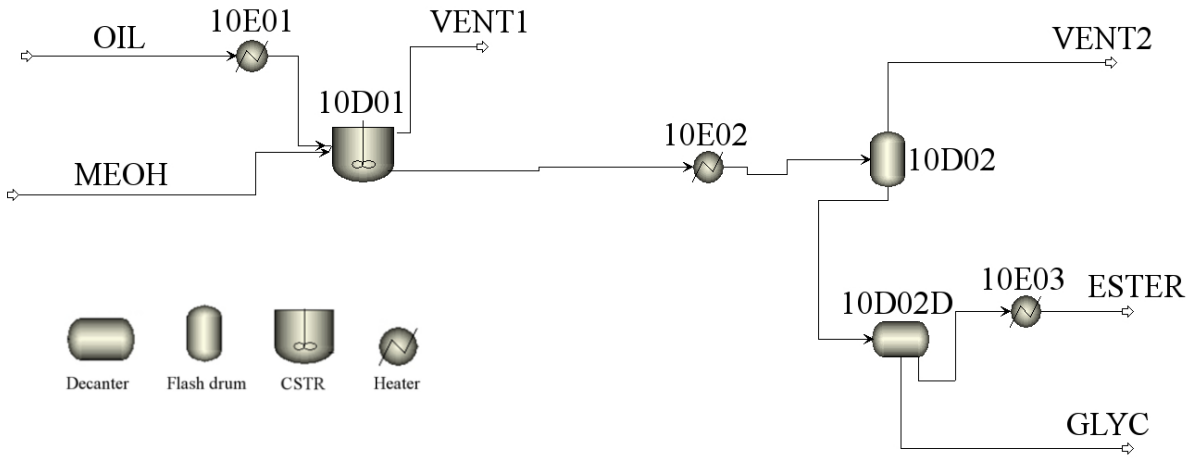


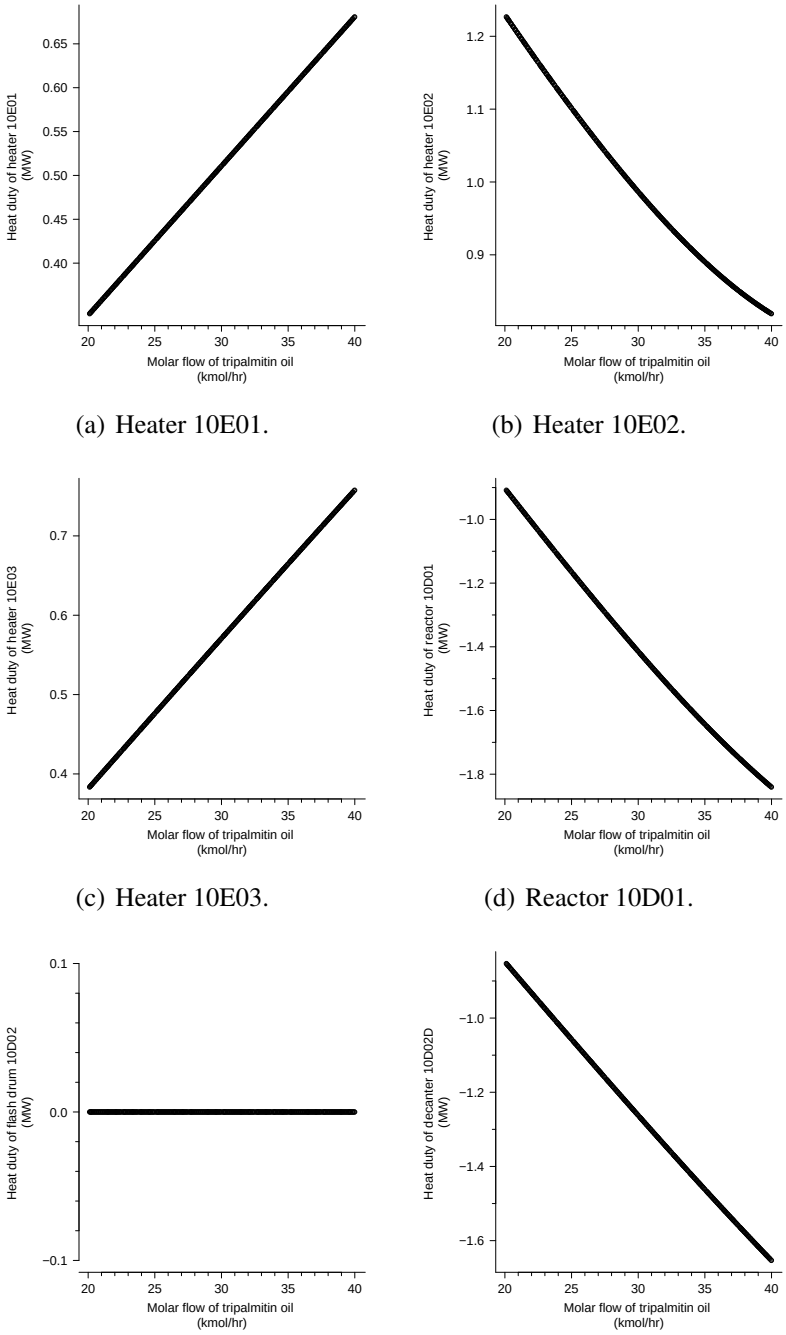
Figure 3.2: *Graphical representation of the process flow sheet model of a biodiesel production line.*

3.2.2 Software - Aspen Plus V8.6

Aspen Plus V8.6 (AspenTech, 2015) 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. 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 5.1.1.



(a) Heater 10E01.

(b) Heater 10E02.

(c) Heater 10E03.

(d) Reactor 10D01.

(e) Flash drum 10D02.

(f) Decanter 10D02D.

Figure 3.3: *Plots of heat duties of various equipment against molar flow of tripalmitin oil.*

3.3 Parameterisation of the biodiesel plant model

3.3.1 Software - Model Development Suite

Model Development Suite (MoDS) (CMCL Innovations, 2015) is a software tool designed to analyse black-box models (e.g. executables, batch scripts). It includes a range of tools such as data-driven modelling, multi-objective optimisation, generation of surrogate models, data standardisation and visualisation, global parameter estimation (Braumann *et al.*, 2010a,b; Man *et al.*, 2010; Braumann *et al.*, 2011; Shekar *et al.*, 2012b; Menz *et al.*, 2012; Shekar *et al.*, 2012a; Menz and Kraft, 2013; Menz *et al.*, 2014), uncertainty propagation (Azadi *et al.*, 2014b; Brownbridge *et al.*, 2014), global and local sensitivity analysis (Vikhansky and Kraft, 2004, 2006; Azadi *et al.*, 2014a), and intelligent design of experiments (Azadi *et al.*, 2015; Yapp *et al.*, 2016). 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 HDMR fitting are used to generate surrogate models. A brief description of each is included, respectively, in Sections 3.3.4, 3.4.1 and 3.4.2.

3.3.2 MoDS-Aspen Plus interface - Component Object Model

The data collection and parametrisation 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 binary-interface standard enabling creation of objects and communication between them (Microsoft, 2015). 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.

3.3.3 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.

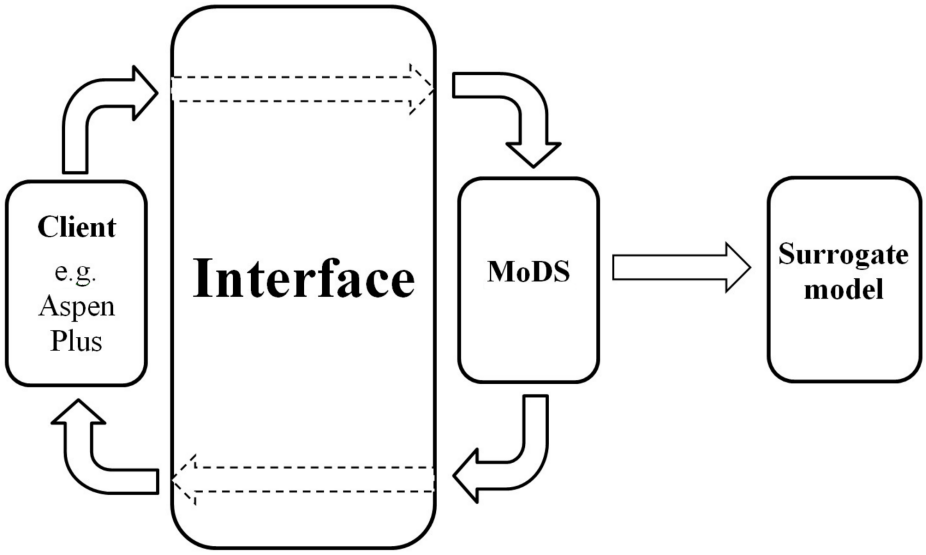


Figure 3.4: *Model Development Suite workflow.*

The workflow of MoDS is visualized in Figure 3.4. 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 3.4.3 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 3.1.

3.3.4 Sampling of the data used for parameterisation

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 use a base of two to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension (Sobol, 1967). The MoDS implementation of a Sobol sequence generator follows the description of Joe and Kuo (2008).

3.4 Implementation of the response surface models

3.4.1 Polynomial response surfaces model

Polynomial response surfaces are a subset of response surface methodology, a group of mathematical and statistical techniques designed to facilitate empirical model building (Myers *et al.*, 2009). 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

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

In order to simplify the notation, *multi-indices* are employed. For example, if p is a multi-index of order l, that means $p \in \mathbb{N}_0^l$, where $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. Then,

$$|p| := \sum_{i=1}^{l} p_i.$$

The independent variable is denoted by x and it is assumed that $x \in \mathbb{R}^n$. A polynomial in x is then a sum of terms of the form

$$x_1^{p_1} x_2^{p_2} \ldots x_n^{p_n},$$

which can be abbreviated to x^p and is of order $|p|$. Thus, the polynomial f_β can be written as

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

where the βs denote the coefficients of the individual terms and k corresponds to the polynomial order.

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

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

Rearranging yields

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

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

3.4.2 High Dimensional Model Representation (HDMR) response surface model

High Dimensional Model Representation (HDMR) is a finite expansion for a given multivariable function as described by Sobol (1990); Rabitz and Alış (1999). It allows for 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:

$$y = f(x) = f_0 + \sum_{i=1}^{N_x} f_i(x_i) + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots N_x}(x_1, x_2, \ldots, x_{N_x}) \quad (3.1)$$

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 the terms containing functions of more than two input parameters are ignored resulting in the following truncated approximation:

$$y \approx f(x) = f_0 + \sum_{i=1}^{N_x} f_i(x_i) + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} f_{ij}(x_i, x_j) \quad (3.2)$$

This approximation relies on the assumption that terms beyond second-order are negligible for many practical applications (including most non-biological processes in the chemical industry). This is supported by the fact that in statistics often only the input covariances play a significant role (Li *et al.*, 2002; Rabitz and Alış, 1999). However, HDMR may not be applicable to cases where high-order terms are important such as wave-body interactions (Molin *et al.*, 2014), modelling of ecological and biological systems (Billick and Case, 1994; Sanchez-Gorostiaga *et al.*, 2018) and arbitrary mathematical functions. Note that equation 3.1 is always an exact representation.

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, $\phi_k(x_i)$, (Li *et al.*, 2002). For data produced using random and quasi-random sampling these functions are related by:

$$f_0 = \overline{f}\,, \tag{3.3a}$$

$$f_i(x_i) = \sum_{k=1}^{M} \alpha_{i,k}\phi_k(x_i)\,, \tag{3.3b}$$

$$f_{ij}(x_i,x_j) = \sum_{k=1}^{M'} \sum_{l=k+1}^{M'} \beta_{ij,kl}\phi_k(x_i)\,\phi_l(x_j)\,. \tag{3.3c}$$

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

$$\int \phi_k(x_i)\,\mathrm{d}x_i = 0 \tag{3.4a}$$

$$\int \phi_k(x_i)\,\phi_l(x_i)\,\mathrm{d}x_i = \delta_{kl}\,. \tag{3.4b}$$

This leads the following equations for the coefficients:

$$f_0 = \int f(x)\mathrm{d}x\,, \tag{3.5a}$$

$$\alpha_{i,k} = \int f(x)\phi_k(x_i)\,\mathrm{d}x\,, \tag{3.5b}$$

$$\beta_{ij,kl} = \int f(x)\phi_k(x_i)\,\phi_l(x_j)\,\mathrm{d}x\,, \tag{3.5c}$$

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 (1999) that the contribution of each term in (3.2), $\sigma^2_{\bar{y},i}$ and $\sigma^2_{\bar{y},ij}$, to the variance of the output parameter can be related to the total variance by

$$\sigma_{\bar{y}}^2 = \sum_{i=1}^{N_x} \int_{-1}^{1} f_i^2(x_i)\,\mathrm{d}x_i + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} \int_{-1}^{1} \int_{-1}^{1} f_{ij}^2(x_i, x_j)\,\mathrm{d}x_i \mathrm{d}x_j \tag{3.6a}$$

$$= \sum_{i=1}^{N_x} \sigma_{\bar{y},i}^2 + \sum_{i=1}^{N_x} \sum_{j=i+1}^{N_x} \sigma_{\bar{y},ij}^2 . \tag{3.6b}$$

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

$$S_i = \frac{\sigma_{\bar{y},i}^2}{\sigma_{\bar{y}}^2} \quad \text{and} \quad S_{ij} = \frac{\sigma_{\bar{y},ij}^2}{\sigma_{\bar{y}}^2} . \tag{3.7}$$

Global sensitivity analysis explores the parameter space and provides robust sensitivity measures throughout the region of interest even in the presence of nonlinearity and parameter interactions. In nonlinear cases, derivative-based local sensitivity analysis can give a false impression of sensitivity (Wainwright *et al.*, 2014).

Basis functions

Polynomials, including Lagrange polynomials (Baran and Bieniasz, 2015), orthonormal polynomials, cubic B splines, and ordinary polynomials (Li *et al.*, 2002), 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

$$\int_{-1}^{1} P_m(x) P_n(x)\,\mathrm{d}x = \frac{2}{2n+1}\delta_{mn} , \tag{3.8}$$

to satisfy (3.4b). The polynomials are generated at runtime according to Bonnet's recursion formula

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) , \tag{3.9}$$

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.

Automatic order selection

Accuracy improvement due to each new term is assessed by calculating R^2 value and comparing it against a predefined minimum value $R^{2*} = 0.00001$, 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. 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. 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.

3.4.3 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 were used including R^2, $\bar{R}^2$, Root-Mean-Squared-Deviation (RMSD) and residual plots. The indices are defined as follows:

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

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

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

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

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 (Draper and Smith, 1998). In essence, it compares the discrepancies between the predicted data and actual data with the discrepancies between the arithmetic average and actual data.

$\bar{R}^2$ (adjusted 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 (Hyndman and Koehler, 2006). 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).

3.5 Numerical experiments

3.5.1 Comparison of the surrogate models - 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 the process flow sheet model used for parameterisation), while RMSD and residual plots were produced using the test set (data sampled from process flow sheet model used for testing, but not parameterisation). 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.

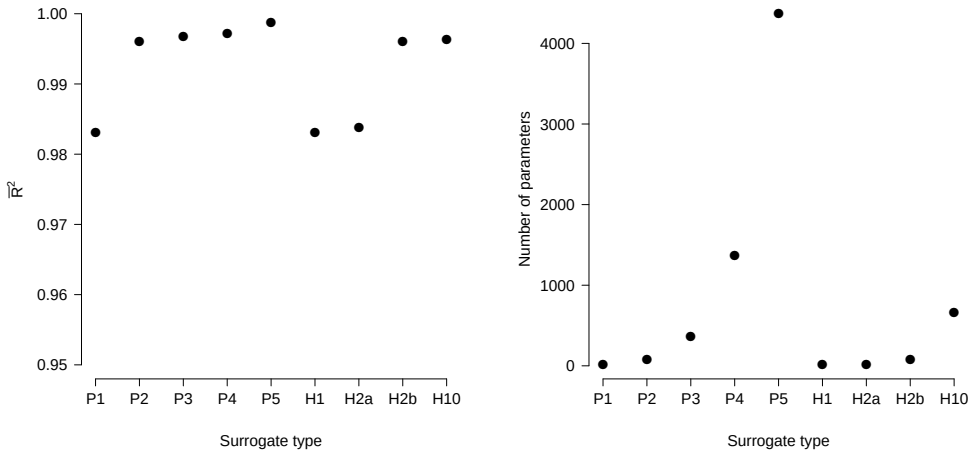
A number of different behaviours were observed in the study. Most surrogate models achieved at least a reasonable fit regardless of the domain size, number of dimensions and according to $\bar{R}^2$ and RMSD. Neither R^2 nor $\bar{R}^2$ can be used to effectively differentiate between the models as most achieve values in excess of 0.98 (for an example see Figure 3.5(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 3.5(b).

RMSD provides a reasonable measure for comparing accuracy of models, as seen in Figure 3.6. Plots 3.6(a) and 3.6(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 3.6(c) and 3.6(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 polyno-

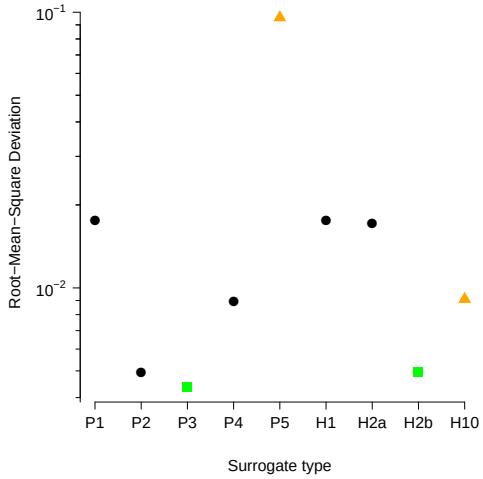
mial relation resulting in non-random distribution of the residuals. Figures 3.8 and 3.9 present residual plots for 11-dimensional surrogates of heat duties of reactor 10D01 and heater 10E03. Comparison of plots in Figures 3.8 and 3.7 shows that for output produced by surrogates with multiple input variables the non-random features are much more difficult to identify. Magnitude of the residuals in most cases is relatively small indicating strong predictive powers of the fits. Comparing plots 3.7(c) and 3.8(c) reveals that performance of polynomial fit of order 5 drops from being the best model to the worst. Plots 3.8(b) and 3.8(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 3.8(c) confirms that P5 fit exhibits relatively low accuracy, even worse than that of a simple linear fit (see plot 3.8(a)).



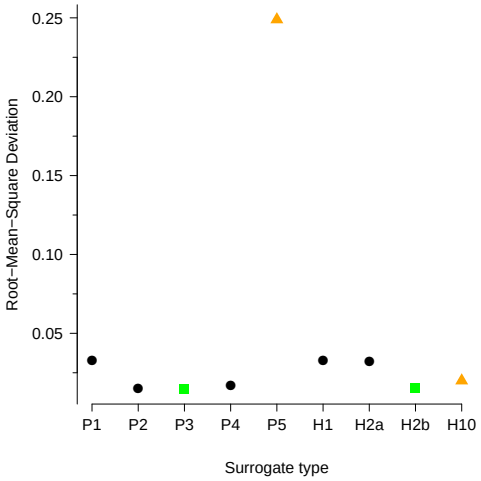
(a) Plot of $\bar{R}^2$ for the considered surrogates.

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

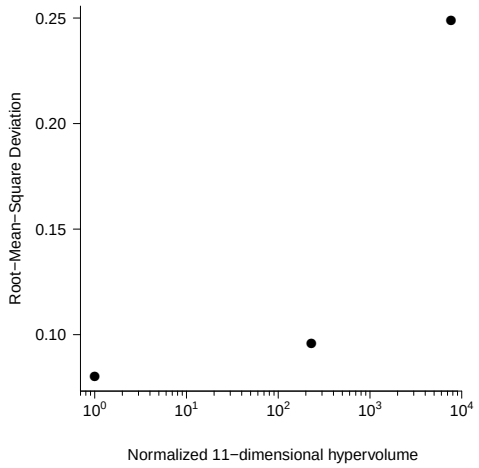
Figure 3.5: *Plots of RMSD and number of parameters for the considered surrogates produced for heat duty of reactor 10D01 with respect to all 11 inputs. Labels P1 through P5 correspond to polynomial fits of order 1 through 5. Label H1 corresponds to a* 1^{st} *order fit, H2a to a* 2^{nd} *order without interactions, H2b to* 2^{nd} *order with interactions and H10 to* 10^{th} *order with* 2^{nd} *order interactions.*



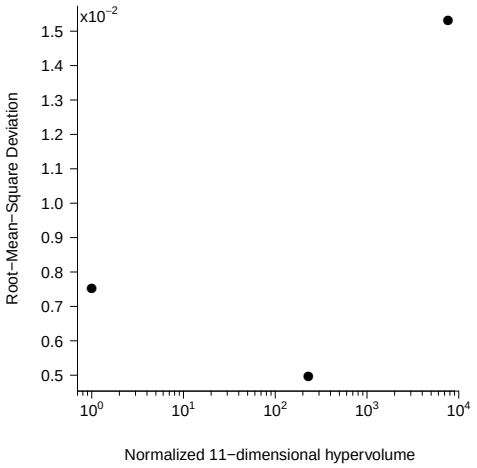
(a) RMSD for the considered surrogates for medium domain size.



(b) RMSD for the considered surrogates.

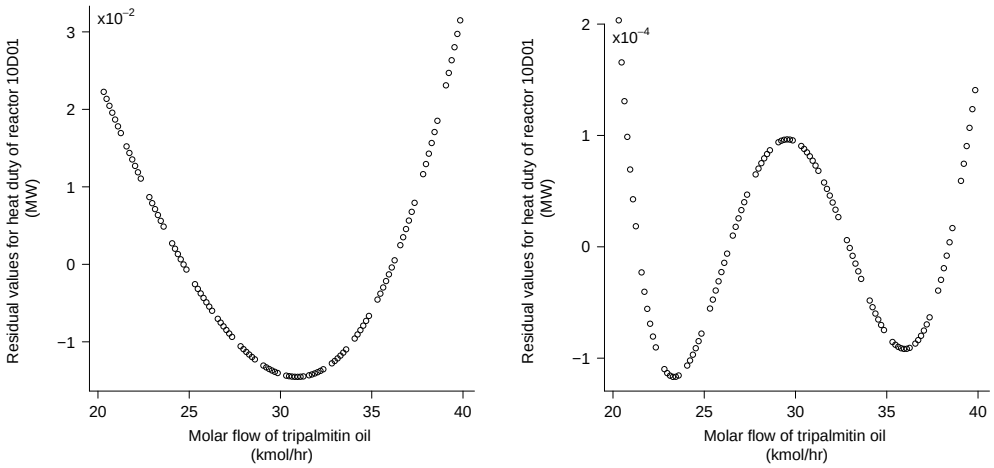


(c) RMSD against domain sizes for polynomial fit of order 5 (for boundaries see Table 3.1).

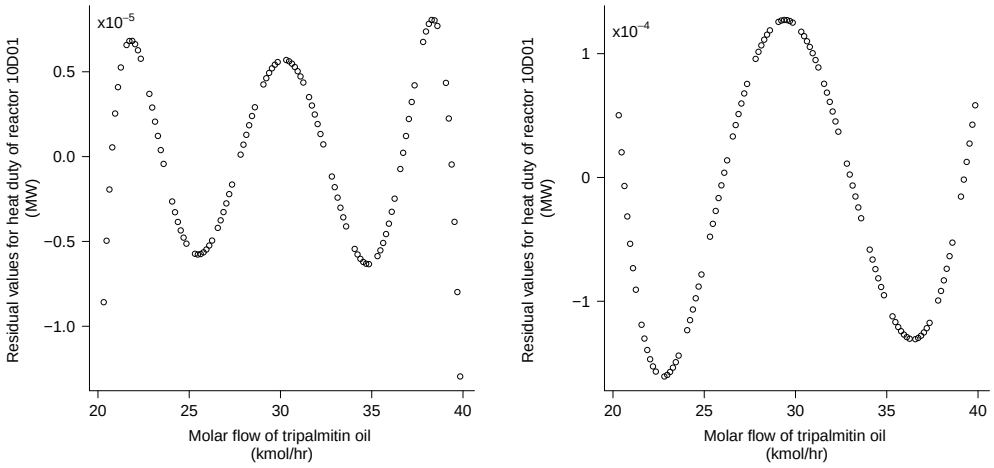


(d) RMSD against domain sizes for HDMR fit (for boundaries see Table 3.1).

Figure 3.6: *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 orange triangles indicate models (one per type) with suffering most from overfitting.*

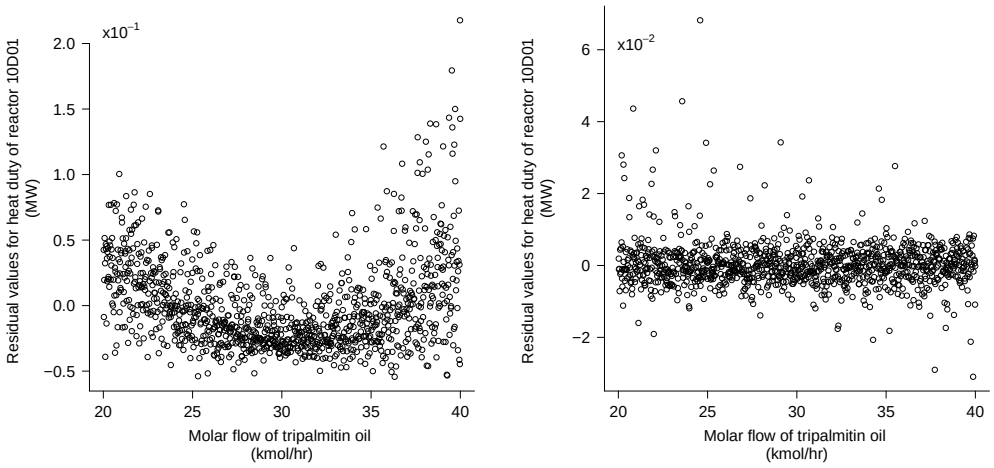


(a) Plot of residuals for 1^{st} order polynomial fit. (b) Plot of residuals for 3^{rd} order polynomial fit.

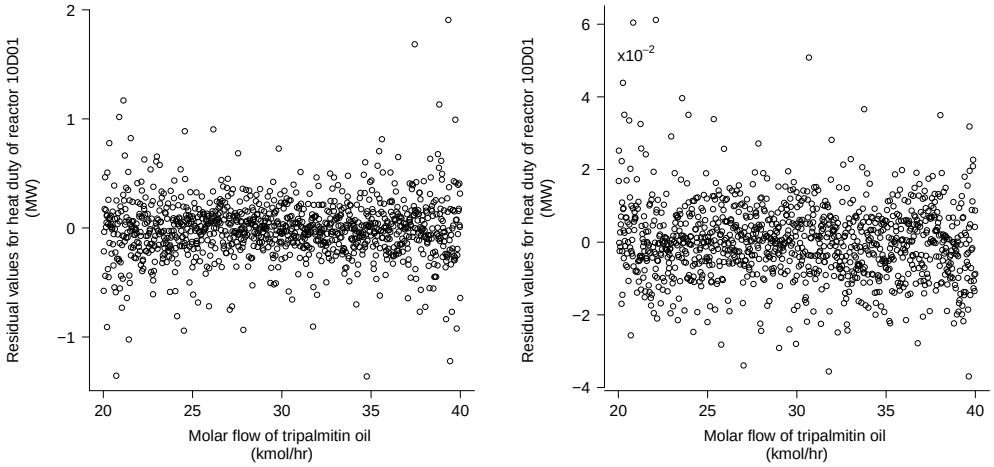


(c) Plot of residuals for 5^{th} order polynomial fit. (d) Plot of residuals for HDMR fit H10 (3^{rd} order polynomial).

Figure 3.7: *Plot of residuals against molar flow of tripalmitin oil for heat duty of reactor 10D01 produced for 1 input.*

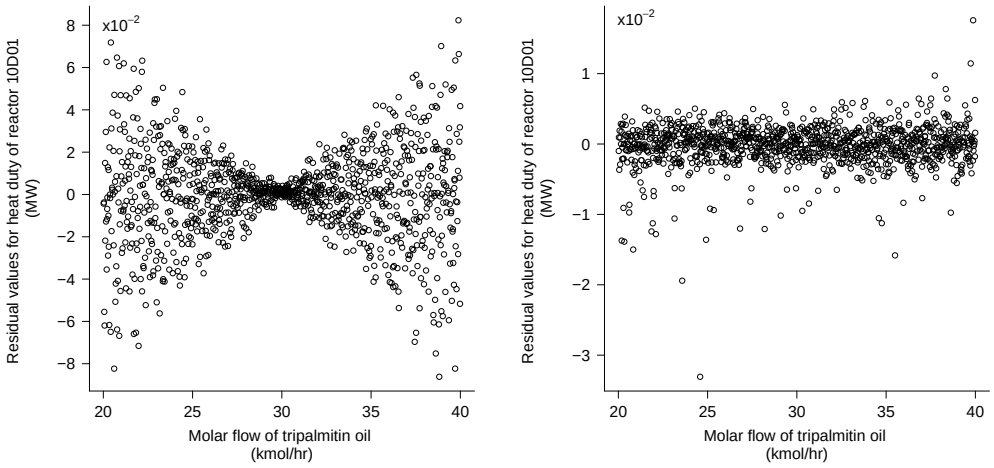


(a) Plot of residuals for 1^{st} order polynomial fit. (b) Plot of residuals for 3^{rd} order polynomial fit.

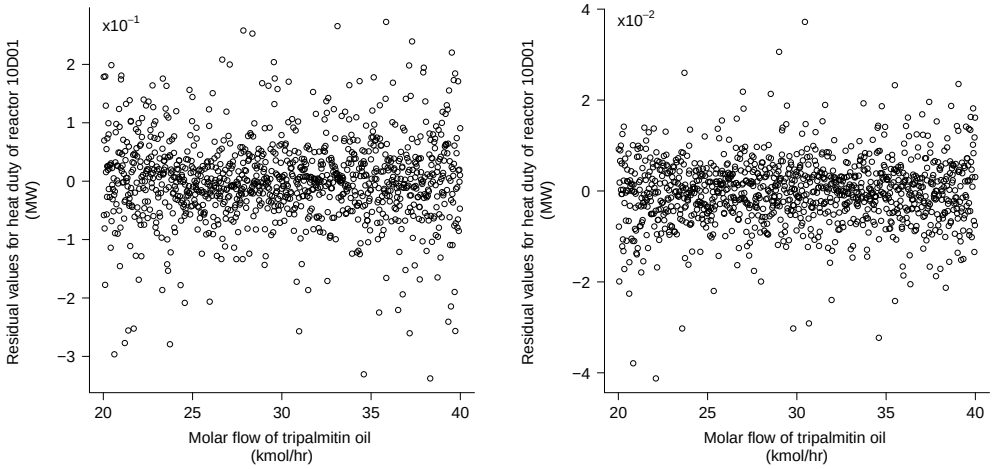


(c) Plot of residuals for 5^{th} order polynomial fit. (d) Plot of residuals for HDMR fit H10.

Figure 3.8: *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.

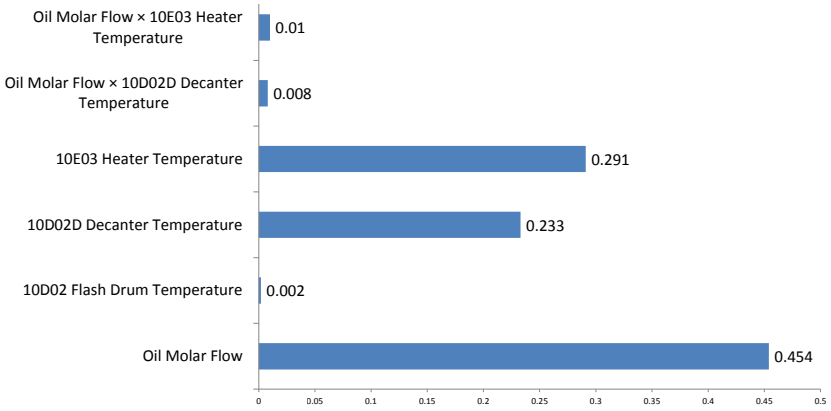
Figure 3.9: *Plot of residuals against molar flow of tripalmitin oil for heat duty of heater 10E03 produced for 11 inputs.*

3.5.2 Global sensitivity analysis

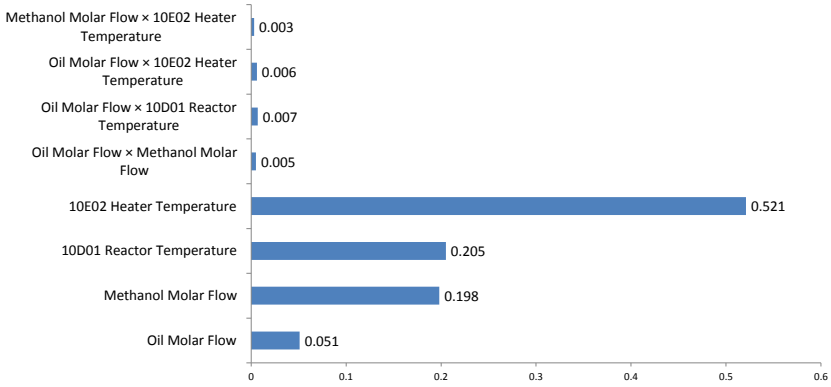
Global sensitivities of the heat duties of all equipment under consideration with respect to the 11 inputs produced by HDMR fitted over the entire domain are summarised in Figures 3.10 and 3.11. It can be seen that in all cases only 4 or fewer inputs have significant influence on a given output. Additionally, interaction terms have only minor effect on any one output. 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 J/mol/K (Filatov and Afanas'ev, 1992) and 79.5 J/mol/K (Freedman *et al.*, 1989)) 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 mostly affected by its operating temperature, operating temperature of reactor 10D01 and molar flow of oil and methanol. Heat duty of heater 10E03 is primarily affected by its operating temperature, operating temperature of decanter 10D02D and molar flow of oil. Heat duty of reactor 10D01 is primarily affected by its operating temperature, operating temperature of heater 10E01 and molar flow 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 flow of oil. Global sensitivities with respect to terms and variables not mentioned here were negligible.

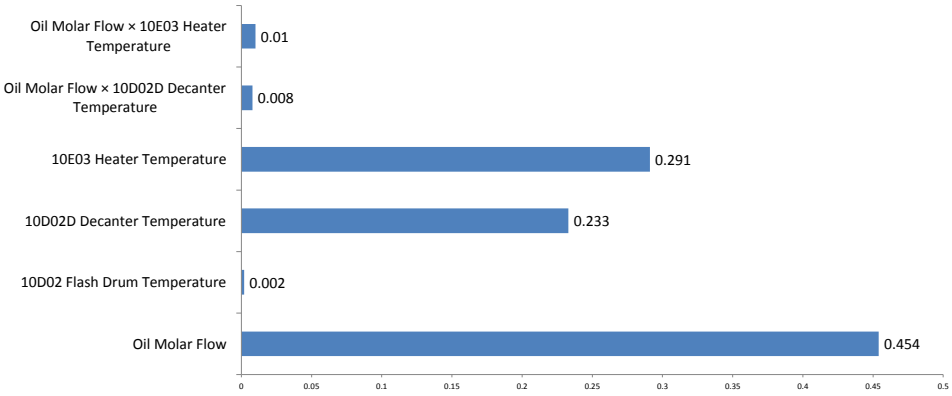
These observations show that when performing multi-dimensional analysis of heat duties within the system many terms in the surrogate models can be ignored due to insignificant influence. Thus, calculation complexity and computational expense can be greatly reduced. Additionally, it shows which inputs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



(a) 10E01 Heater - Heat Duty.

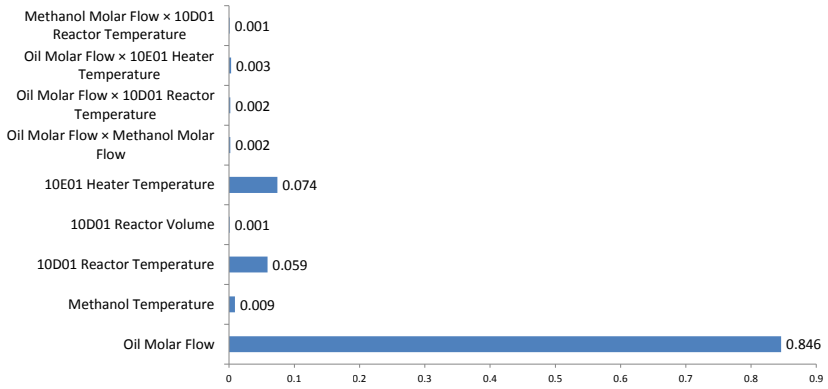


(b) 10E02 Heater - Heat Duty.

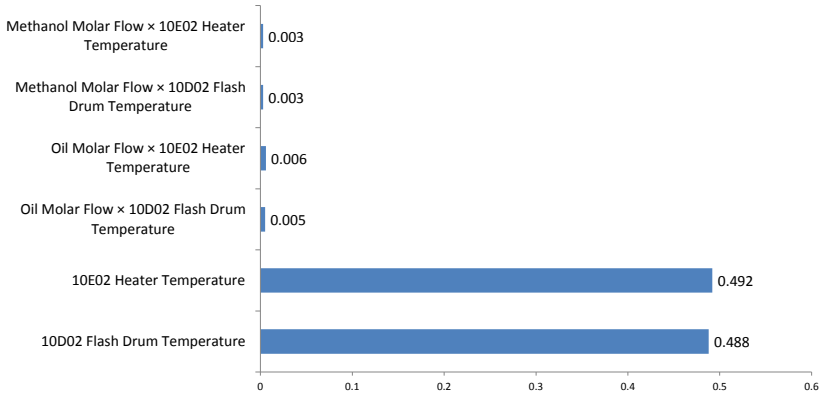


(c) 10E03 Heater - Heat Duty.

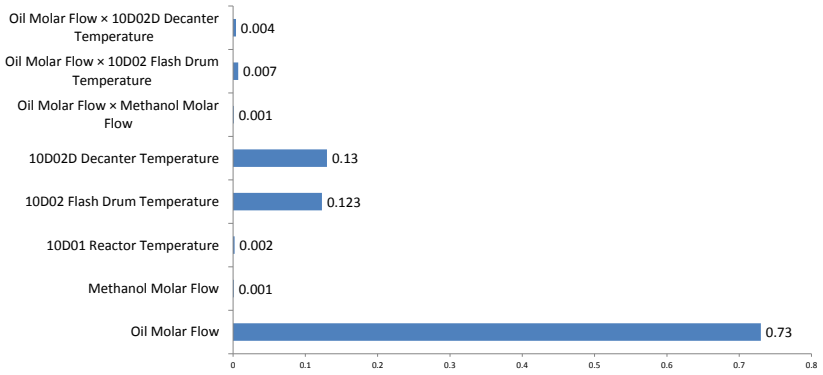
Figure 3.10: *Global sensitivities produced by 11-dimensional HDMR fit over the entire domain.*



(a) 10D01 Reactor - Heat Duty.



(b) 10D02 Flash Drum - Heat Duty.



(c) 10D02D Decanter - Heat Duty.

Figure 3.11: *Global sensitivities produced by 11-dimensional HDMR fit over the entire domain.*

3.6 Future considerations

In this chapter it was shown how to create accurate surrogates of a model typical for chemical industry using polynomial response surfaces and High Dimensional Model Representation response surfaces. A number of research questions that are beyond the scope of the current work remain to be investigated in order to develop the technology demonstrated in this chapter to the point where it can be deployed within an autonomous Industry 4.0 environment. A list of examples is given below.

- **What restrictions exist on the behaviour and/or complexity of the chemical model?**

  The current chapter focused on a typical, but somewhat simple, chemical process flowsheet model. It was observed that the numerical behaviour of the model caused some issues for the generation of the surrogate model. In this particular case, this was overcome by identifying and ignoring a (small) problematic subset of the model results. Would this get worse with more complex models and is it a behaviour that all surrogate models are likely to need to deal with? Probably. There are many reasons why a model could display such behaviour. For example, there may be genuine bifurcations in the solution of the underlying model equations. Examples of this include reaction problems with strong heat release (for example a flame) that have a "hot" solution and a "cold" solution. In this example, the behaviour is a reflection of the underlying physics. It is not a deficiency in the model. In other cases, the converse may be true. For example, it could be that there are issues with a numerical method that cause it to return erroneous results. For example, suppose that the underlying physics requires the identification of a global minimum, and that we search for this minimum using a method that requires an initial guess. If we have a poor initial guess, is it possible that the method may not converge. Hopefully this problem would be easy to identify. Alternatively, perhaps the numerical method finds a local minimum that nevertheless falls within its convergence criteria, causing it to terminate and return a result based on the local minimum. This type of

error could be much more subtle and therefore much harder to identify.

- **How should complexity be measured? Can we identify a relationship between complexity and accuracy?**

  How should the complexity of a process of a chemical process model be measured? For example, should it be measured based on the physical characteristics of the process? The number of unit operations, process streams, phases etc. Or should it be measured in terms of a mathematical description of the process. For example, the number of process parameters, inputs and outputs, the state space of the inputs and outputs, whether or not the response is expected to be smooth, steady and/or continuous? Is it possible to define a measure of complexity that can be used to make statements about the expected accuracy of a (given) surrogate method? Does the accuracy change smoothly, or does it suddenly drop off once you reach a certain complexity?

- **What is the best choice of surrogate method for a given situation?**

  Given a mathematical problem, is it possible a priori to identify good (and bad) choices of surrogate models? This is an active research question. For example, see the work by Garud *et al.* (2018) and references therein. Are there specific types of surrogates that are often good choices for chemical processes? Or is it the case that the optimal choice of surrogate differs for different parts of a chemical process? If so, and also considering the likely benefits of lower complexity, would it be better to model different parts of a process using different surrogates, which could then be combined into an overall model? Could this be taken a step further and used to define a library of ready-made surrogate models for standard items of equipment (pumps, separators etc), instances of which could be trained to model small parts of a given process flowsheet and then combined to create an overall surrogate for the overall process?

- **How could the processes of fitting a surrogate model be automated?**

  In addition to requiring answers and solutions to many of the issues discussed above, a number of other capabilities need to be developed in order

to realise a system that could automate the fitting of surrogate models. For example, the system would need to be able to communicate and interact with the underlying process models, both to extract information used to decide what type of surrogate should be used, and to run the models to retrieve data as part of the training process. Likewise, it would need to be able to communicate and interact with the software used to fit to surrogate. Active research is ongoing to address specific questions within these areas. For example, see the work by Xiaochi Zhou (2019), which presents progress towards developing a framework that supports the deployment of autonomous interacting computational agents. In the proposed framework, each model would be seen by the system as an agent that could be asked to perform tasks as and when required.

3.7 Chapter summary

This chapter presents results of parameterisation of typical input-output relations within process flow sheet of a biodiesel plant and assesses parameterisation accuracy. 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 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). 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.

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 > 3. 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 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, but minor accuracy improvement in terms of $\bar{R}^2$ and RMSD. It was observed that non-random features in residual plots are much more difficult to identify 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 have significant influence on a given output. Interaction terms have only minor effect on any one output. 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 need to be controlled.

It was shown that the presented approach can be effectively used to parameterise a model typical for chemical industry. Such surrogate models can be used as a replacement of sufficient accuracy in modelling platforms which need to rely on models of physical processes, but also need to be fast and stable during execution.

Chapter 4

Blockchain-enabled M2M electricity market

This chapter investigates how blockchain technology might be coupled to cyber-physical systems to leverage new opportunities in the context of eco-industrial parks by enabling automated trading of resources. The investigation is performed in the context of using blockchain to facilitate M2M interactions and establish a M2M electricity market in the chemical industry. The presented scenario includes two electricity producers and one electricity consumer trading with each other over a blockchain. All participants are supplied with realistic data produced by process flow sheet models. This work contributes a proof-of-concept implementation of the scenario. It was shown that the blockchain technology can be employed in the chemical industry for the purposes of allowing virtual representations of real equipment trading with each other. Extending this idea to Industry 4.0 and cyber-physical systems would allow real equipment to interact with each other autonomously in a decentralised network. The chapter concludes with a discussion of continuing the presented work.

4.1 Methodology

Aspen Plus (a process modelling and optimisation software), Multichain (a software package for the creation and deployment of private blockchains), Oracle VM Virtual Box (a virtualization software) and custom-made Python 3.4 (a programming language for general-purpose programming) scripts were employed together in order to produce an example in which blockchain technology is employed to facilitate M2M interactions and simulate a M2M electricity market. In this study the following procedure was used:

1. Multichain was used to establish a blockchain between the virtual machines and allow them to post, view and execute transactions on a continuous basis.
2. Every 5 seconds electricity supply and demand data was updated using Aspen Plus simulations (natural gas is burnt to provide supply and steam is compressed to provide demand) on a Windows 10 machine.
3. A Python script on Windows 10 used the file system to pass the generated data to three Fedora 24 virtual machines, each representing either an electricity consumer or producer.
4. Python scripts on the Fedora machines were used to read the data from Aspen Plus simulations and interact with the blockchain.

This workflow is visualized in Figure 4.1. More detail is available in the rest of the chapter.

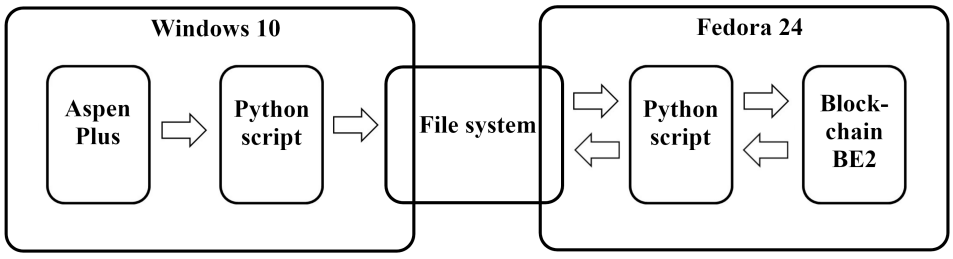


Figure 4.1: *Study workflow for Chapter 4.*

4.1.1 Codes and data

This work can be reproduced by following this workflow and using the instructions and code provided under the following link: https://doi.org/10.17863/CAM.33680.

4.2 Design and implementation

This section presents an example in which blockchain technology is employed to facilitate M2M interactions and establish a M2M electricity market in the context of the chemical industry and the IoT. Electricity is a convenient example as its transfer is near-instantaneous (as are the corresponding blockchain transactions), but in principle any other commodity (e.g. steam, natural gas, coal) could be used here. However, the likelihood of a discrepancy between the blockchain record and reality is more likely for commodities which require significant delivery time.

This scenario consists of two electricity producers, producing approx. 750 MW, and one electricity consumer, consuming approx. 23 MW, which trade with each other over a blockchain. The producers publish exchange offers of energy (in kWh) for currency (in USD) in a data stream, which functions as a publishing board. Energy produced in a 5-second interval is split into batches of 100 kWh (and one, smaller remainder batch) and priced according to a predefined policy (one producer randomly chooses a price between 0.01 and 0.2 USD per kWh and the other between 0.1 and 0.3 USD per kWh). The consumer reads the offers, analyses them and attempts to satisfy its energy demand at a minimum cost. Its energy demand is updated every 5-second interval. When an offer is accepted it is executed as an atomic exchange (i.e. two simultaneous transactions are executed, and both must either succeed together or fail together). The scenario is visualised conceptually in Fig. 4.2. It is envisaged that the machines participating in this system would each be equipped with a computer containing their digital representation enabling them to interact with a blockchain and provide relevant sensor data. Here, the physical machines are replaced with physical simulations of industrial processes in Aspen Plus (AP) (AspenTech, 2015).

Table 4.1: *General information about chain BE2.*

MultiChain version	1.0 alpha 25
Protocol version	10006
Chain name	BE2
Mining diversity	0.75
Names of native assets	USD, kWh
Number of participants	3

Each producer is simulated by a process in which natural gas is burnt to produce energy, see Fig. 4.3. It consists of a methane stream which is throttled in a valve from 15 bar to 1.3 bar and fed into a Gibbs reactor where it combusts in air, which is pressurised to 1.3 bar in a compressor and then fed into the reactor. Energy is harvested from the exhaust stream using a heat exchanger. Each consumer is simulated by a compressor increasing steam pressure, see Fig. 4.4. It pressurises steam from 1 bar to 21 bar. It is assumed that consumer can be readily driven using the energy produced by the producers.

The example was implemented on a Windows 10 machine hosting three Fedora 24 (Fedora Project, 2016) virtual machines using Oracle VM Virtual Box. These used MultiChain (Greenspan, 2015) to establish a blockchain (named BE2) and AP to simulate industrial processes. It is important to note that the employment of the final setup was preceded by, in additional to a standard literature review, an extensive search for a suitable application (primarily by consultations and interviews) and appropriate software (using Bitcoin blockchain or creating of an Ethereum side-chain was considered) was conducted.

The machines running Fedora represent the producers and the consumer on chain BE2 and receive data from the AP simulations. General information about blockchain BE2 can be found in Table 4.1. All data transfers outside the blockchain, interpretation and analysis of the posted offers, electricity pricing and automation were facilitated using scripts written in Python 3.5.

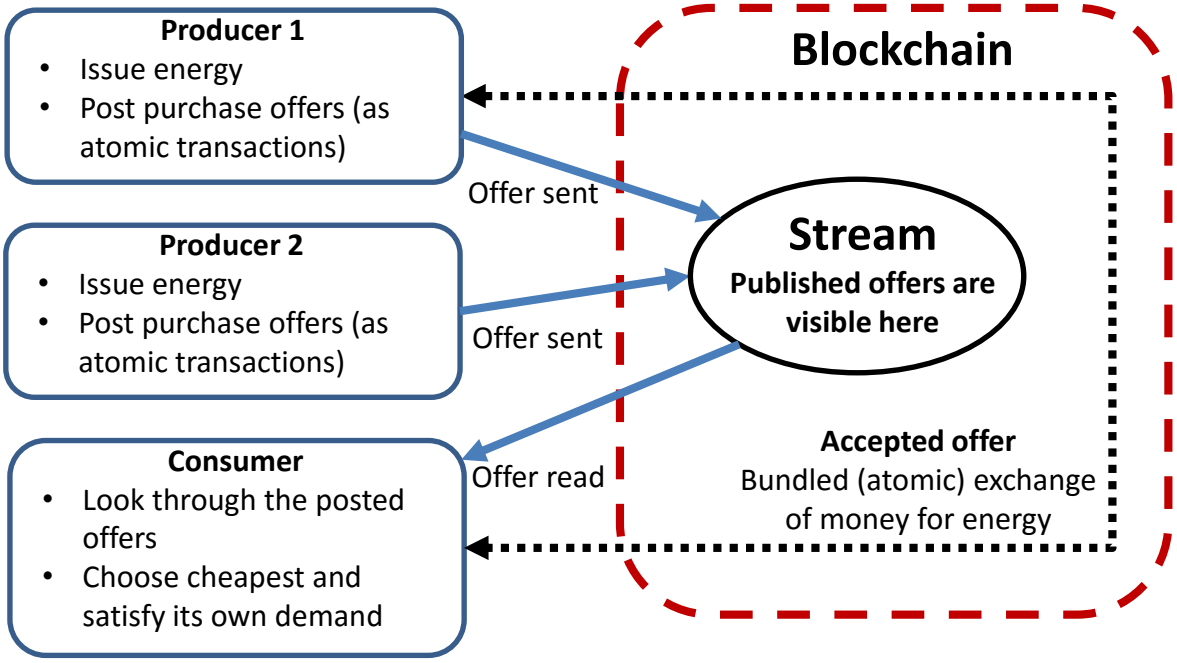


Figure 4.2: *Visual presentation of energy producers and a consumer participating in an electricity market over a blockchain.*

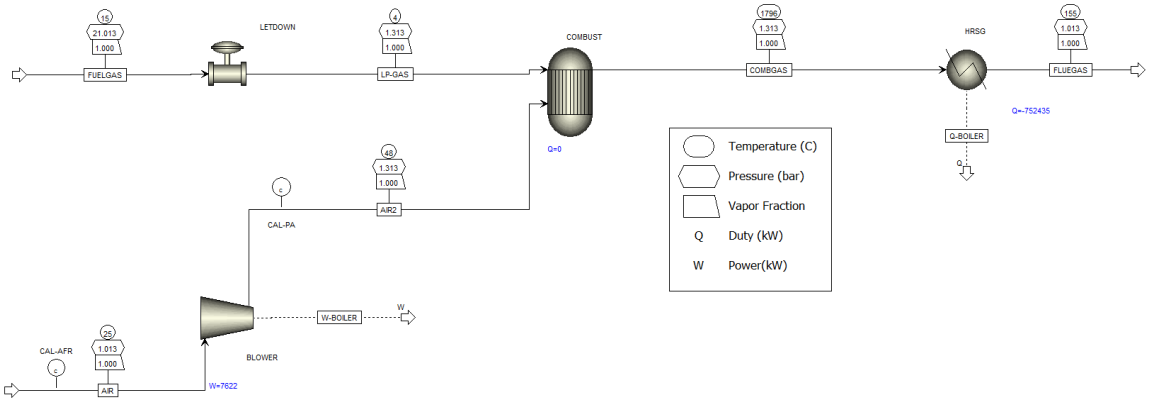


Figure 4.3: *An Aspen Plus (AspenTech, 2015) simulation in which natural gas is burnt.*

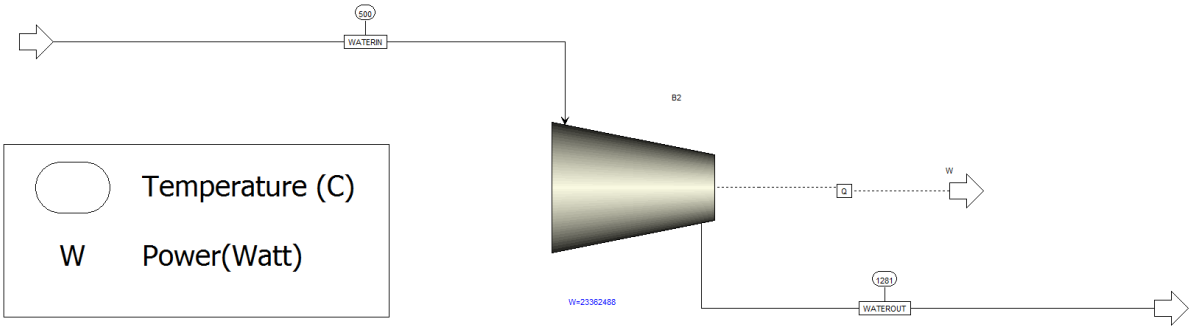


Figure 4.4: *An Aspen Plus (AspenTech, 2015) simulation in which an electricity-driven compressor increases steam pressure.*

4.3 Software used in the implementation

Multichain

MultiChain is a software package, in development, designed as an off-the-shelf platform for the creation and deployment of private blockchains, where certain degree of trust between the participants is possible. In this implementation the primary features of BE2 include round robin consensus mechanism, and native assets (here, a digital currency created on top of the chain's native currency; see *Assets* in A.1). In round robin consensus mechanism, the set of miners is limited to known entities which take turns in publishing blocks. The strictness of the rotation scheme is controlled using a parameter called mining diversity ($0 \leqslant$ mining diversity $\leqslant 1$). This parameter defines the minimum proportion of permitted miners needed to control the network. 0.75 is a recommended value (Greenspan, 2015), as high values are safer, but a value too close to 1 can cause the blockchain to freeze up if some miners become inactive. In the case that the network splits temporarily (e.g. due to communications failure) resulting in a fork, the branch with the longer chain will be adopted.

The participants are approved for publishing blocks as follows:

1. Any permission changes defined by transactions in the current block are applied;
2. The current number of permitted miners is calculated;
3. The number of miners is multiplied by mining diversity and rounded up to get spacing;
4. If any of the spacing-1 blocks were mined by the current miner, the block is invalid.

The scheme enjoys the following advantages over a centralised database:

- Each participant has full control over its assets via their ownership of private key(s);

- Distributed control prevents an individual or a small group from unilaterally deciding which transactions are valid or will be confirmed;

- More robust as access and validation of transactions will continue even if a server malfunctions (i.e. no single point of failure).

Aspen Plus V8.8

Aspen Plus V8.8 (AspenTech, 2017b) is a process modelling and optimisation software used by the bulk, fine, specialty, and biochemical industries. More details of the software are provided in Section 3.2.2. It was used to simulate a process in which natural gas is burnt to produce energy, see Fig. 4.3, acting as an energy producer on the blockchain network and a compressor increasing steam pressure, see Fig. 4.4, acting as an energy consumer on the blockchain network.

Python 3.5.0

Python (Python Software Foundation, 2017) is an interpreted programming language for general-purpose programming. In this study it was used to automate the interactions between the Multichain blockchain and the Aspen Plus simulations. Aspen Plus was accessed via COM interface (more in Section 3.3 of Chapter 3), while Multichain using a Savoir 1.0.6 package (Python Software Foundation, 2018) adapted for Python 3.

Oracle VM VirtualBox

VirtualBox (Oracle Corporation, 2018) is a virtualization software for x86-based systems. It can be installed on Windows, Linux, Mac OS X and Solaris x86 computers. It allows creating and running multiple virtual machines, running different operating systems, on the same computer at the same time. It was employed to host Fedora machines.

4.3.1 Description of a typical trade

This section presents a sequence of events which leads to successful trade between participants of blockchain BE2.

A typical trade proceeds as follows:

1. The producer nodes prepare and publish exchange offers of kWh for USD in the stream "elec-market-open". The preparations require the producers to lock a sufficient amount of energy asset and encode details of the exchange.
2. The consumer node looks for the offers related to each publisher and decodes them.
3. The consumer compares the offers and chooses the one which minimises the energy cost.
4. The consumer prepares a transaction matching the chosen offer by locking sufficient funds and appending the chosen offer with payment details. It then encodes this and submits the accepted exchange to the chain.
5. Finally, the consumer verifies that the transaction was validated by the chain.

4.4 Future considerations

In this chapter it was shown that the blockchain technology can be employed in the chemical industry to allow virtual representations of real equipment to trade with each other to improve the efficiency of the process. The work was performed in the context of eco-industrial parks and the JPS project and it is anticipated that it will be continued by progressing towards the vision of fully automated industry. Employing blockchains in conjunction with JPS will enable the application of the findings from the current example to larger networks and different types of commodities (e.g. steam, water, carbon credits), the implementation of more complex pricing models, balancing of the positions of customers and producers

(at the moment the market is purely producer-driven) and the introduction of more complex trade deals using smart contracts, which would allow to further automate the considered system and thus increase its efficiency. From the chemical engineering perspective, the intention is to use a greater variety of models and to introduce dynamic behaviour (e.g. a simulation of a process line during start-up and shut-down) in order to investigate how does the technology perform in different environments. Blockchain technology may have the potential to change the engineering industry by facilitating the transition to and functioning of Industry 4.0. When applying this technology to further problems it is important to assess whether it is suitable for a given scenario by considering its benefits and limitations. Benefits include:

- Increased trust - trust in the protocols replaces trust in verification conducted by humans,
- Immutability and integrity of the stored data - disputes between participating parties can be easily resolved,
- Disintermediation - removes need for a central authority within the network,
- Automation of transactions between participating parties,
- Cost reductions due to automation and disintermediation,
- Increased processing speed for certain scenarios compared to a centralised system.

The limitations of blockchain technology include:

- High latency - the mechanism of block addition is computationally expensive (applicable to the proof-of-work consensus mechanism),
- High costs - the entire transaction history needs to be stored on multiple nodes and the mechanism of block addition is computationally expensive in the case of the proof-of-work consensus mechanism,

- Public-private key cryptography technology - it provides no recourse if a party publishes or misplaces their private key,
- Immutability - makes amending the transaction history difficult,
- Lack of privacy - many nodes in the network contain the entire transaction history.

These considerations raise a number of open questions that it is recommended should be investigated before implementing the type of trading scheme considering in this work. A list of examples is given below.

- **How can challenges of scalability be handled to facilitate transactions within the environment of Internet of Things?** Internet of Things entails allowing thousands (possibly many more) of devices to rapidly communicate and conduct transactions with each other. Due to the costs associated with processing transactions blockchain technology may not be suitable for high frequency operations, however finding the point at which the technology becomes unsuitable could guide further research. Overall whether blockchain technology will be able to facilitate those interactions hinges on overcoming scalability issues including high latency, constantly increasing storage requirements and energy-expensive process of block addition.
- **What kind of consensus mechanism is optimal for employment in the context of chemical industry?** What characteristics (for example, level of permission or distribution of mining power) should a consensus mechanism have in order to serve well the needs of chemical industry? How do those characteristics interact with each other? This work used a round robin consensus mechanism as it was designed with private blockchains, where certain degree of trust between the participants is possible, in mind. This is a reasonable method for companies which are engaged in a cooperative venture or for intra-company transactions. However, for applications to a more competitive environment a different one may be more suitable. For example, a deposit-based mechanism, where the participants are required to

register a security deposit in order to be allowed to produce blocks, could effectively discourage any attempt at misbehaviour.

- **How will introduction of blockchain technology into chemical industry impact the existing business models?** What kind of (if any) new business models and forms of employment may arise in this situation? What magnitude of benefit should be expected from using this approach? Given the advantage of disintermediation provided by blockchain technology implies that intermediary services may need to significantly redesign their business models in order to stay relevant. Furthermore, companies using blockchain for internal purposes may find it beneficial to restructure in order to use this technology more effectively.

- **Could the presented idea be applied to other commodities?** This scenario used electricity as its delivery and reception are effectively instantaneous. For the same reason applying it to exchange of virtual entities (such as in CO_2 emissions trading schemes) would be viable. However, trading commodities with significant exchange times may result in discrepancies between the blockchain records and the real state of affairs. For that reason, such exchanges may need an additional mechanism designed to realign the records and reality.

4.5 Chapter summary

This chapter demonstrates that it is possible to successfully employ the blockchain technology to facilitate M2M interactions and establish a M2M electricity market in the context of the chemical industry via the IoT. The presented scenario includes two electricity producers and one electricity consumer trading with each other over a blockchain. The producers publish exchange offers of energy (in kWh) for currency (in USD) in a data stream. The consumer reads the offers, analyses them and attempts to satisfy its energy demand at a minimum cost. When an offer is accepted it is executed as an atomic exchange. This work contributes a proof-of-concept implementation of the described scenario and its technical de-

tails. Furthermore, all participants are supplied with realistic data produced by process flow sheet models of industrial equipment.

It was shown that the blockchain technology can be employed in the chemical industry for the purposes of allowing virtual representations of real equipment trading with each other. Extending this idea to Industry 4.0 and cyber-physical systems would allow real equipment to interact with each other autonomously in a decentralised network.

Chapter 5

Automated arbitrage spotter

This chapter investigates how the cyber-physical system developed in previous chapters can be applied to find and exploit opportunities for financial arbitration over chemical commodities on futures markets. It presents the implementation of an automated arbitrage spotter powered by market and physical data applied to two scenarios: conversion of natural gas to methanol and crude palm oil to biodiesel. It was shown that models of chemical plants can find and assess opportunities for financial arbitration over chemical commodities on futures markets in an automated manner by calculating their total costs per unit of product and comparing it with the online market prices. Extending this idea to Industry 4.0 it shows that a cyber-physical systems operating a chemical plant could participate in and interact with a trading network. Analysis conducted using market data from 28.05.2017 in the former scenario indicates that no trade should be made. In the latter case up to 345.17 USD per tonne of biodiesel can be earned by buying September 2018 reagent delivery contracts and selling December 2018 product delivery contracts. The chapter concludes with a discussion of continuing the presented work.

5.1 Methodology

Aspen Plus (a process modelling and optimisation software) and custom-made Python 3.5 (a programming language for general-purpose programming) scripts were employed together in order to produce an automated arbitrage spotter for a chemical plant converting a reagent into a product which searches for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product. In this study the following procedure was used:

1. Python script was used to read process data (such as product mass flow, reagent mass flow and consumption of utilities) from Aspen Plus simulations modelling a chemical process and download futures market prices from an appropriate online source (such as a commodity exchange).

2. Python script was used to consider the futures market prices of reagent and the product, costs of storage, transport and conversion (steam, electricity and other utilities) and produce an investment recommendation.

This workflow is visualized in Figure 5.1. More detail is available in the rest of the chapter.

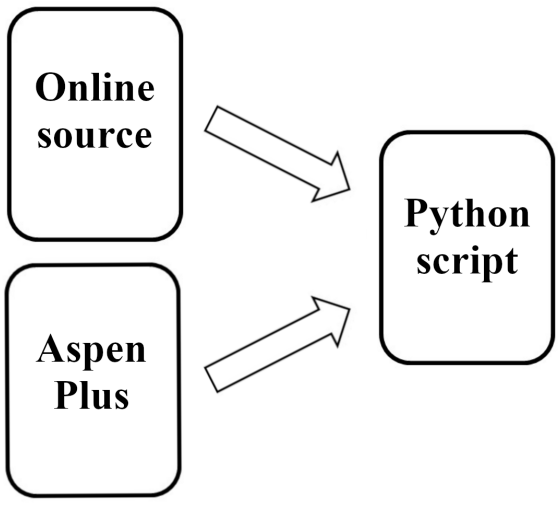


Figure 5.1: *Study workflow in Chapter 5.*

5.1.1 Codes and data

This work can be reproduced by following this workflow and using the instructions and code provided under the following link: https://doi.org/10.17863/CAM.33680.

5.2 Design and implementation

This section describes the design and implementation of the automated arbitrage spotter for a chemical plant converting a reagent into a product. The spotter searches for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product. It considers costs of storage, transport and conversion (steam, electricity and other utilities) and produces an investment recommendation. It is assumed that the plant is located in Singapore and operates with a long-term production contract, but its throughput could be marginally increased.

The implementation involves the following steps:

1. downloading market data from an appropriate online source (the considered scenarios include the Chicago Mercantile Exchange (CME Group, 2017b) and Zhengzhou Commodity Exchange (ZCE, 2017)),

2. performing a feasibility analysis based on the downloaded data and physical data provided by a simulation of the chemical plant under consideration (the included scenarios employ Aspen Plus (AspenTech, 2017b) and Aspen HYSYS (AspenTech, 2017a)).

Both steps are executed by scripts written in Python 3.5.0 (Python Software Foundation, 2017). Market data is downloaded on a daily basis as it is updated at the same rate. The second step determines potential profit to be made on delivering a single unit of the product. The required quantities of the reagent contracts, utilities (steam, water, heating fuel and electricity), transport to and from the plant and storage volume are calculated and costed. The utility, transport and storage prices

were determined based on specifications of a biodiesel plant designed by Lurgi GmbH in 2007 and the available literature. Necessary currency conversions were done with data from Oanda (2017); XE.com Inc. (2017), while Coinnews Media Group LLC (2017) was used to adjust for inflation.

The programme solves the following equation for all contracts in the correct chronological order (*i.e.* delivery of the product must be scheduled at least a month after delivery of the reagent) in order to determine profit per unit of the product:

$$\text{Profit} = P_p - T_p - (S_p \oplus (S_r \times R)) \times d - R \times (P_r - T_r) - U \tag{5.1}$$

where subscripts p and r refer to the product and the reagent respectively, P to price, T to transport cost, S to storage cost, R to units of the reagent per unit of the product, d to the storage duration and U to the cost of utilities per unit of the product. Note that $\oplus$ indicates that the programme assumes either the product or the reagent is stored for the entire time between reception of latter and delivery of the former.

Finally, the software will return the most profitable contract and storage schedule combination.

5.3 Software used in the implementation

Aspen Plus V8.8

Aspen Plus V8.8 (AspenTech, 2017b) is a process modelling and optimisation software used by the bulk, fine, specialty, and biochemical industries. More details of the software are provided in Section 3.2.2 of Chapter 3.

Aspen HYSYS V8.8

Aspen HYSYS (AspenTech, 2017a) is a process modeling tool used by the petrochemical industry for process simulation and process optimization in design and

operations. Its uses include, but are not limited to:

- designing various components in a petrochemical plant,
- detecting abnormal operating conditions,
- modeling steady state and dynamic processes.

Python 3.5.0

Python (Python Software Foundation, 2017) is an interpreted programming language for general-purpose programming. In this study it was used to download online data, perform calculations and connect to Aspen Plus and Aspen HYSYS via COM interface. More details of the interface are provided in Section 3.3 of Chapter 3.

5.4 Results and discussion

5.4.1 Description of the crude palm oil-to-biodiesel conversion process

Biodiesel plant simulation

A process producing 24.334 tonnes per hour of biodiesel was modelled in Aspen Plus V8.8 using the UNIFAC-DMD property method and steam tables for calculations involving pure water (Gmehling *et al.*, 2002). The flow sheet model includes two reaction stages (modelled using continuously stirred tank reactors - CSTRs), a separation stage, a methanol recycle loop, a gas-fuelled steam generation section and auxiliary equipment, see Fig. 5.2. The final fuel, fatty acid methyl ester, is produced via trans-esterification pathway where triglycerides react with methanol to form methyl ester and glycerine in the presence of an alkaline catalyst. The flow sheet was based on an existing plant designed by Lurgi GmbH. In the process tripalmitin oil is reacted with methanol in the CSTRs to produce glycerol and

Table 5.1: *Details of the futures contracts (CME Group, 2017a,c).*

	Crude palm oil	**Biodiesel**
Exchange	Chicago Mercantile Exchange	Chicago Mercantile Exchange
Contract Unit	25 t	100 t
Currency	USD	USD
Delivery location	Malaysia	Southern China

methyl palmitate (biodiesel), then passed through a flash drum and a decanter and washed with water to separate the remaining methanol and glycerol. Finally, dry air is used to remove water from the product. 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. It is assumed that crude palm oil can be directly fed into this processing line.

Financial analysis

This scenario involves the following steps:

1. downloading market data from the website of Chicago Mercantile Exchange (CME Group, 2017c,a),

2. performing a feasibility analysis based on the downloaded data and physical data provided by Aspen Plus (AspenTech, 2017b) model of a chemical plant converting crude palm oil into biodiesel, see Fig. 5.4.

It is important to note that both futures markets deal with financially settled contracts, but it is assumed that their prices are an accurate approximation of the physically settled ones. For transportation purposes it is assumed that the delivery location of crude palm oil is in Malaysia and of biodiesel in southern China. The contract details can be seen in Table 5.1. A plot of the prices can be viewed in Fig. 5.3.

The required quantities of crude palm oil contracts, utilities (steam, water, heating fuel and electricity), transport to and from the plant and storage volume are calculated and costed per tonne of biodiesel. The prices were determined based

on the plant specifications and the literature. The summary of prices can be found in Table 5.2. To calculate the profit the programme solves eqn. 5.1 outlined in Section 5.2.

The best contract and storage schedule combination is chosen based on this information using an exhaustive search. The result of the analysis using the market data from 28.05.2017 is 345.17 USD per tonne of biodiesel can be made. This can be achieved by buying contracts for delivery of crude palm oil in September 2018 and selling contracts for delivery of biodiesel in December 2018 in a ratio of 4 to 1. The largest cost contribution comes from transportation. Note that it is assumed that all crude palm oil was converted instantaneously (relatively to the timescales involved). Duration of storing the reagent and the product is irrelevant as their storage costs are the same.

This result indicates that there may be a realistic scope for increasing profitability of a chemical plant by exploiting the opportunities across different commodity markets. It is also shown that this can be achieved in an automated manner.

Table 5.2: *Storage, transport and utility prices.*

Storage		
Crude palm oil	0.2075	USD per month-tonne
Biodiesel	0.2075	USD per month-tonne
Transport		
Crude palm oil (Fadhil, 2013)	5.0	USD per tonne
Biodiesel (Fadhil, 2013)	40.0	USD per tonne
Utilities		
Electricity (EMA, 2017)	0.0000386	SGD per kJ
Fuel gas	0.0000093	USD per kJ
Steam (low pressure)	0.01264	USD per kg
Steam (high pressure)	0.00349	USD per kg
Cooling water	0.00174	USD per kg
Process water	0.00033	USD per kg

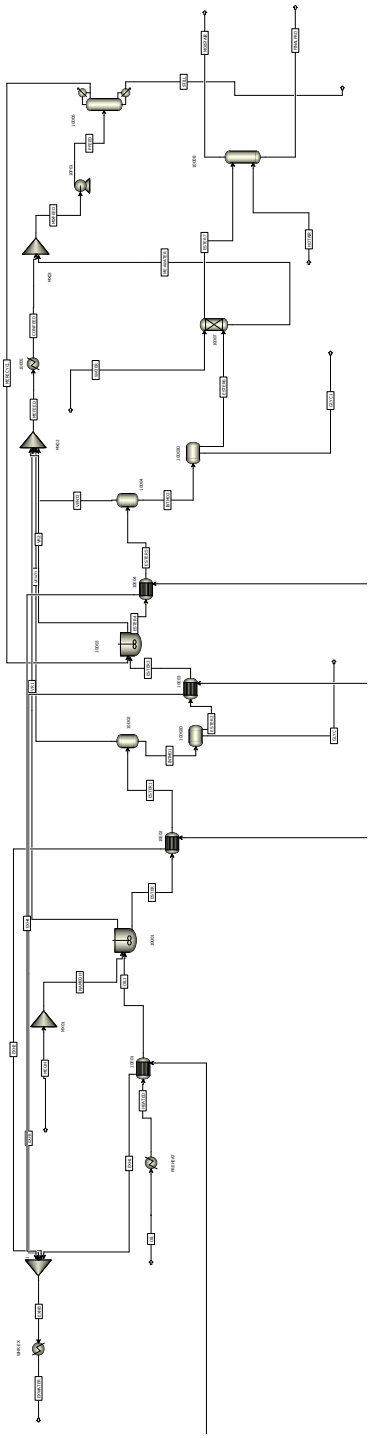


Figure 5.2: *Aspen Plus model of a crude palm oil-to-biodiesel plant (steam generation equipment not shown). Note: while the text in the figure appears very small, the figure is encoded in a vector format so that it can be read clearly by zooming in when reading the electronic version of the thesis.*

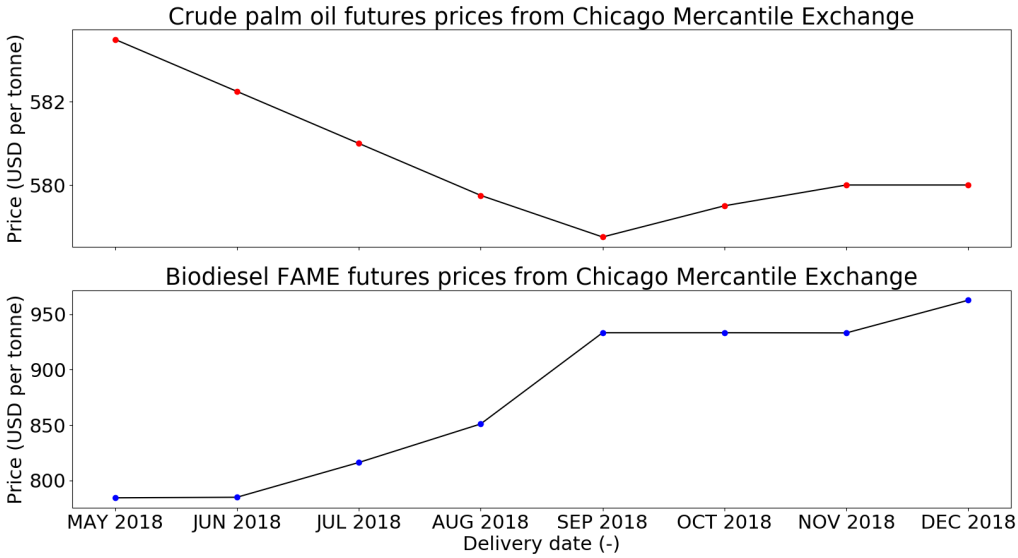


Figure 5.3: *Plot of the crude palm oil and biodiesel futures prices from 28.05.2017. For example, a contract for future delivery of crude palm oil in September 2018 can be bought (or sold) at approx. 584 USD per tonne.*

5.4.2 Description of the natural gas-to-methanol conversion process

Methanol plant simulation

A process producing around 56.791 tonnes per hour (1085 mmBTU per hour) of methanol was modelled in Aspen HYSYS V8.8 using the Peng-Robinson property method (Peng and Robinson, 1976). The flow sheet model can be divided into two parts: natural gas-to-syngas conversion and syngas-to-methanol conversion. Overall, they include two reaction steps (modelled using plug flow reactors - PFRs), three separation steps, three recycle loops, a water top-up section and auxiliary equipment, see Fig. 5.4. Natural gas (mostly composed of methane, but also ethane and propane) is stripped of any pollutants (such as sulphur compounds) and then converted into syngas (a mixture of carbon monoxide and hydrogen) via steam reforming. Water is removed from the mixture, which is then converted into methanol in a PFR. Lastly, the product is purified up to the desired specifications. 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.

Financial analysis

This scenario involves the following steps:

1. downloading market data from the websites of Chicago Mercantile Exchange (CME Group, 2017b) and Zhengzhou Commodity Exchange (ZCE, 2017) (the contract details can be seen in Table 5.3, while a plot of the prices can be viewed in Fig. 5.5),

2. performing a feasibility analysis based on the downloaded data and physical data provided by Aspen HYSYS (AspenTech, 2017a) model of a chemical plant converting natural gas into methanol, see Fig. 5.4.

The required quantities of natural gas contracts, utilities (steam, water, heating

Table 5.3: *Details of the futures contracts (ZCE, 2017; CME Group, 2017b).*

	Natural gas	**Methanol**
Exchange	Chicago Mercantile Exchange	Zhengzhou Commodity Exchange
Contract Unit	10 mmBTU	1 tonne (approx. 19.1 mmBTU)
Currency	USD	CNY
Delivery location	Henry Hub pipeline	Zhengzhou Exchange

fuel and electricity), transport to and from the plant and storage volume are calculated and costed per mmBTU of methanol. The prices of utilities, except electricity, and methanol storage were determined based on specifications of the biodiesel plant from Section 5.4.1, while the remaining ones from the literature. The summary of prices can be found in Table 5.4. To calculate the profit the programme solves eqn. 5.1 outlined in Section 5.2.

The best contract and storage schedule combination is chosen based on this information using an exhaustive search. The analysis using the market data from 28.05.2017 recommends that no trade should be made in order to avoid making a loss. The least loss of approximately 3690 USD per mmBTU of methanol can be achieved by buying contracts for delivery of natural gas in June 2017 and selling contracts for delivery of methanol in July 2017 in a ratio of 3 to 1. The largest cost contribution comes from the utilities. Note that it is assumed that all natural gas was instantaneously (relatively to the timescales involved) converted on arrival and methanol was stored until delivery date.

This results indicates that direct arbitrage with natural gas may not be possible since the markets are efficient (*i.e.* market prices reflect all relevant information) and the natural gas prices in Asia due to the so called "Asian premium" (*i.e.* additional transport and storage costs). Additionally, it is cheaper to produce methanol at the source of natural gas as the former is significantly cheaper to transport than the latter.

Table 5.4: *Storage, transport and utility prices.*

Storage		
Natural gas (FERC, 2004)	0.121	USD per month-mmBTU
Methanol	0.01086	USD per month-mmBTU
Transport		
Natural gas (EYGM, 2014)	approx. 7	USD per mmBTU
Methanol (Aitken, 2016)	1.200	USD per mmBTU
Utilities		
Electricity (EMA, 2017)	0.0000386	SGD per kJ
Fuel gas	0.0000093	USD per kJ
Steam (low pressure)	0.01264	USD per kg
Steam (high pressure)	0.00349	USD per kg
Cooling water	0.00174	USD per kg
Process water	0.00033	USD per kg

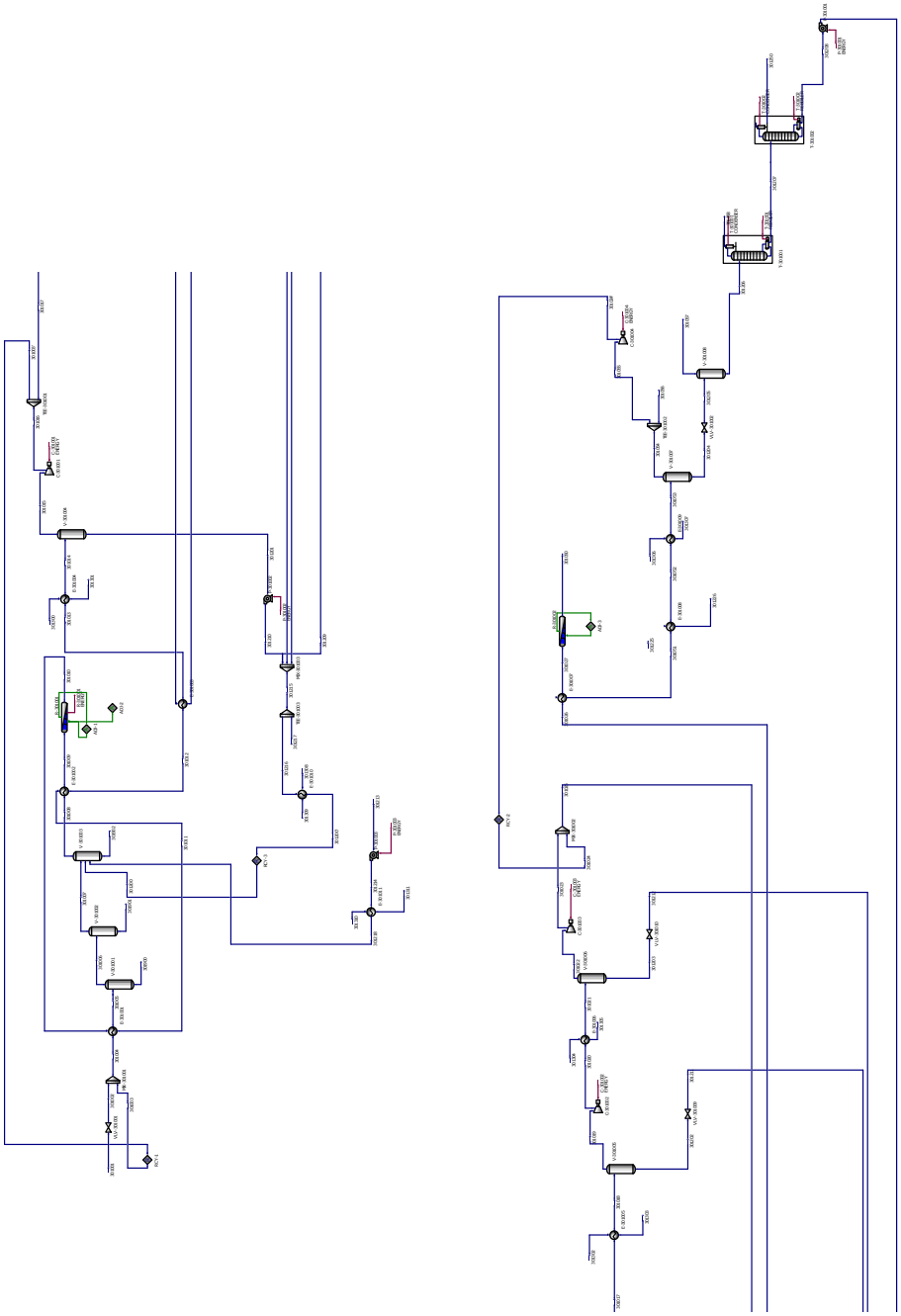


Figure 5.4: *Aspen HYSYS model of a natural gas-to-methanol plant. The upper section converts natural gas into syngas, while the lower syngas into methanol. Note: while the text in the figure appears very small, the figure is encoded in a vector format so that it can be read clearly by zooming in when reading the electronic version of the thesis.*

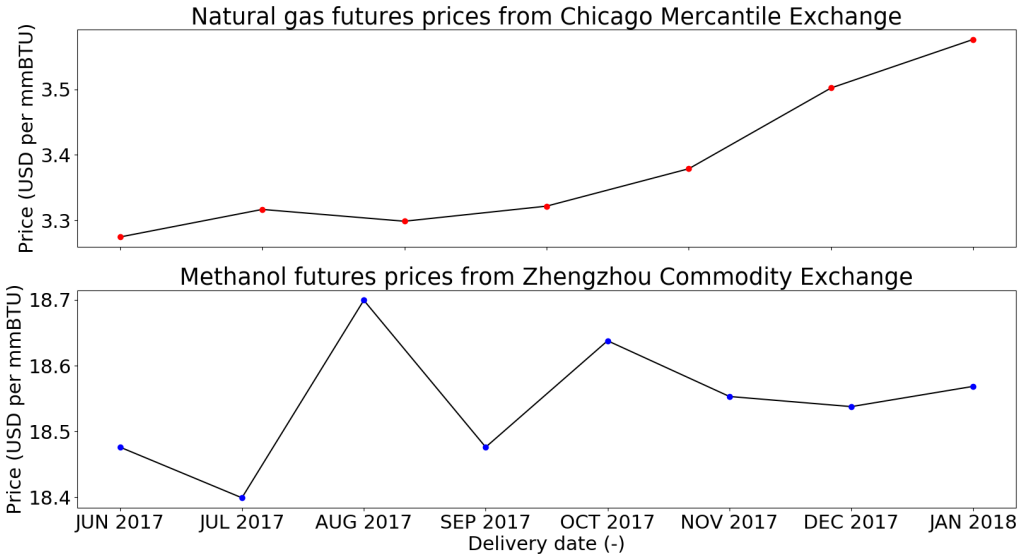


Figure 5.5: *Plot of the natural gas and methanol futures prices from 28.05.2017. For example, a contract for future delivery of methanol in Aug 2017 can be bought (or sold) at approx. 18.7 USD per mmBTU.*

5.5 Future considerations

In this chapter it was shown that models of chemical plants can find and assess opportunities for financial arbitration over chemical commodities on futures markets in an automated manner by calculating their total costs per unit of product (including transport, storage, reagents and utilities) and comparing it with the online market prices. Employing this approach in conjunction with JPS, blockchain technology and parameterisation of suitable plant models could allow simulation of financial markets with more complex trade deals. Additionally, there exist many research questions that are beyond the scope of the current work but should be investigated. A list of examples is given below.

- **What model characteristics are desirable for this kind of application?** This work presents a scenario where commodity contracts from an existing market are accepted and completed with a minimum of a month between reception and delivery. Given this temporal resolution the accuracy of the considered model is significantly more important that the evaluation speed. If a market, where physical exchanges are made more frequently, is entered speed becomes more important. However, unless the time between reception and delivery falls into the order of hours or shorter (e.g. in the electricity market), accuracy will always be paramount. Using surrogate models would most likely be only warranted to substitute computationally expensive and/or unstable models. It is important to remember that a surrogate model cannot be more accurate than the model on which it is based so the base model needs to have a minimum level of accuracy.

- **What kind of chemical industries could participate in a commodity market as presented?** Throughout this work it is assumed that the plant operates with a long-term production contract, but its throughput could be marginally increased. By using the spare production capacity, the plant could make additional earnings by exploiting the price differences between its feedstock and product on the commodity market. Chemical processes which can be flexible with their production throughput will be more suitable to apply the scheme than ones which are not.

- **How would commodity markets be affected by wide participation of automated plants?** What kind of behaviours would such plants need to exhibit to be able to effectively participate? How would a market composed of models of chemical plants behave? Would the optimal plant behaviour change based on the type of commodities involved? While the presented scenario involves a model of chemical plants participating in an existing market, given a sufficiently large number of automated chemical plants they could have a significant impact on the markets that the participate in or even create their own market exchanges.

- **How would existing business models be impacted by common participation of automated plants in commodity markets?** Would new kinds of business models arise in this situation? One of the advantages of the presented system is the disintermediation between the considered chemical plant and the target commodity markets. Given that the currently existing intermediaries would be pressured to rethink their business models. Additionally, it could encourage plants with spare production capacity to temporarily increase their production in order to participate in commodity markets. For example, a plant involving electrochemical processes could operate using grid electricity whenever it is profitable.

5.6 Chapter summary

This chapter presents an implementation of an automated arbitrage spotter powered by market and physical data applied to two scenarios: conversion of natural gas to methanol and crude palm oil to biodiesel. The programme searches for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product. It considers cost of storage and conversion (other feedstock, steam, electricity and other utilities) derived from physical simulations of the chemical process. It is assumed that the plant is located in Singapore and operates with a long-term production contract.

Two scenarios considered are natural gas-to-methanol conversion and crude palm

oil-to-biodiesel conversion. Analysis conducted using the market data from 28.05.2017 in the former scenario no trade should be made in order to avoid making a loss. However, in the latter case up to 345.17 USD per tonne of biodiesel can be earned by buying contracts for delivery of crude palm oil in September 2018 and selling contracts for delivery of biodiesel in December 2018 in a ratio of 4 to 1. Duration of storing the reagent and the product is irrelevant as their storage costs are the same. It is shown that there may be a realistic scope for increasing profitability of a chemical plant by exploiting the opportunities across different commodity markets in an automated manner. Additionally, the results suggest that direct arbitrage with natural gas may not be possible as the markets are efficient and transporting natural gas tends to be more expensive than methanol.

It was shown that models of chemical plants can find and assess opportunities for financial arbitration over chemical commodities on futures markets in an automated manner by calculating their total costs per unit of product (including transport, storage, reagents and utilities) and comparing it with the online market prices. Extending this idea to Industry 4.0 it shows that a cyber-physical systems operating a chemical plant could participate in and interact with a trading network.

Chapter 6

Conclusions

6.1 Conclusions of the presented work

All of the scenarios were designed as parts of J-Park Simulator, a holistic, modelling platform based on the industrial park located on Jurong Island in Singapore. The first focused on the biodiesel plant formerly located within the park, the second on a potential M2M electricity market connecting individual machines on the island and the third on enabling process lines to participate in the financial markets.

Surrogate models of a biodiesel plant

Model Development Suite was used to produce various surrogates of an Aspen Plus biodiesel plant flow sheet model using polynomial-based surrogate models. The approximations relate 11 inputs typical to the biodiesel plant and their accuracy is assessed. The impact of dimensionality, domain size and parameterisation type on the accuracy are explored. Additionally, global sensitivities of the outputs are calculated using HDMR. It is found that most surrogate models achieved at least a reasonable fit regardless of the domain size and number of dimensions. Furthermore, it was observed that in all cases only 4 or fewer inputs have significant influence on any of the outputs and that the interaction terms have only minor

effect on any one output. This suggests that this application can be parameterised with even simpler models i.e. ones ignoring interaction terms altogether.

It was shown that this approach can be effectively used to parameterise a model typical for chemical industry. Thus, they can be used as an accurate replacement in modelling platforms which need to rely on models of physical processes, but also need to be fast and stable during execution.

Blockchain-enabled M2M electricity market

The potential of applying blockchain technology in the chemical industry is explored and analysed. A scenario employing blockchain to establish and facilitate a M2M electricity market in the context of the chemical industry is presented and implemented. Two electricity producers and one electricity consumer continuously trade with each other over a blockchain-based network. The participants are automated process flow sheet models of industrial equipment supplied with realistic data. It was shown that the blockchain technology can be employed in the chemical industry for the purposes of allowing virtual representations of real equipment trading with each other. Extending this idea to Industry 4.0 and cyber-physical systems would allow real equipment to interact with each other autonomously in a decentralised network.

Automated arbitrage spotter

An automated arbitrage spotter, supplied with market and physical data, is designed and implemented. The software searches for opportunities to make additional profit by analysing the futures market prices for both the reagent and the product in the cases of conversion of natural gas to methanol and crude palm oil to biodiesel. It considers cost of storage and conversion (other feedstock, steam, electricity and other utilities) derived from physical simulations of the chemical processes. No profitable scenario was found for the first chemical process, while for the second up to 345.17 USD per tonne of biodiesel can be earned by buying contracts for delivery of crude palm oil in September 2018 and selling contracts

for delivery of biodiesel in December 2018 in a ratio of 4 to 1. It is assumed that the plant is located in Singapore and operates with a long-term production contract. It was shown that models of chemical plants can find and assess opportunities for financial arbitration over chemical commodities on futures markets in an automated manner by calculating their total costs per unit of product (including transport, storage, reagents and utilities) and comparing it with the online market prices. Extending this idea to Industry 4.0 it shows that a cyber-physical systems operating a chemical plant could participate in and interact with a trading network.

6.2 Suggestions for future work

Interoperability

Significant work in the future can be done in the area of interactions between the concepts explored in this thesis. For example, behaviour of a network where surrogate models of chemical plants interact with each other and the external world on a blockchain-based forum could be investigated. An ontology-based context model for the network could be developed and employed within such a network. It could provide participating agents with the context of their execution potentially leading to improved performance and ability to engage in more complex interactions.

Surrogate models of a biodiesel plant

Suggestions for investigations continuing the work described in Chapter 3, where the process of generating accurate surrogates of a model typical for chemical industry using polynomial response surfaces and HDMR response surfaces was investigated, include:

- study more complex chemical models e.g. a number of interconnected models forming a feedback loop necessitating coupling surrogate models and solving them simultaneously to understand the extent to which these methodologies can be applied to real-world systems,
- investigate existing restrictions on the behaviour and complexity of the chemical model in order to better understand the boundaries within which it should be parameterised,
- study measures of model complexity as well as the relationship between complexity of chemical models and accuracy of the derived surrogate model in order to uncover how this relationship varies across types of chemical models and parameterisation techniques,

- identify best choice of a parametrisation method given a particular scenario in the context of the chemical industry in order to determine which technique(s) is best for a given class of cases,

- develop a capability to effectively automate the process of fitting surrogate models so that it can deployed within an autonomous Industry 4.0 environment.

Blockchain-enabled M2M electricity market

Recommendations for studies continuing the work presented in Chapter 4, where it was shown that the blockchain technology can be employed in the chemical industry to allow virtual representations of real equipment to trade with each other to improve the efficiency of the process, include:

- apply blockchain technology to scenarios involving larger networks, more complex pricing models, more complex trade deals, greater variety of chemical models and different types of commodities (e.g. steam, water, carbon credits) to understand the extent to which the technology can be extended to real-world systems,

- address the challenges of scalability within the environment of Internet of Things so that the presented application can deployed within an autonomous Industry 4.0 environment,

- identify best choice of a consensus mechanism given a particular scenario in the context of the chemical industry in order to determine which mechanism(s) is best for a given class of cases,

- study the impact of blockchain technology on the existing business models in the chemical industry in order to gain understanding into the change that this technology may effect on its environment.

Automated arbitrage spotter

Suggestions for studies continuing the work presented in Chapter 5, where it was shown that models of chemical plants can find and assess opportunities for financial arbitration over chemical commodities on futures markets in an automated manner, include:

- utilise the presented approach in conjunction with JPS, blockchain technology and parameterisation of suitable plant models to allow simulation of financial markets to understand the extent to which these technologies can be applied to real-world systems,
- study what model characteristics are desirable for this kind of application and which chemical industries could likely participate in a commodity market as presented in order to better understand the boundaries within which it should be applied,
- study the effects of wide participation of automated plants on commodity markets and the impact of such participation on the existing business models in the chemical industry in order to gain understanding into the change that this application may effect on its environment.

Appendix A

Key concepts and definitions of the blockchain technology

A.1 Definitions

51% attack - an attempt by one or more participants with collective majority control of a network (e.g. by hash rate or stake) to revise transaction history and/or prevent new transactions from being confirmed.

Assets - an entity (e.g. currency, commodity) created by sending additional data in transactions of a chain's native currency.

Block - a file in which data (e.g. transactions, events) are recorded.

Blockchain - a distributed, electronic database which can hold any information (records, events, transactions, etc.) and can set rules on how information is updated (Condos *et al.*, 2016). It continually grows as discrete chunks (blocks) are appended and linked (chained) to the previous block using the hash of its content. It also records every change made in its history so in order to alter a past entry all subsequent blocks also need to be altered. It is authenticated and maintained through a distributed network of participants (nodes) according to a predefined consensus mechanism (Condos *et al.*, 2016).

Byzantine Generals Problem (as described by Lamport *et al.* (1982)) - an agreement problem in which a group of generals, each commanding a portion of the Byzantine army, encircle a city. These generals wish to formulate a plan for attacking the city. In the most basic form, they need to decide whether to attack or retreat. It is vital that every general agrees on a common decision, for a half-hearted attack by a few generals would become a rout and be worse than a coordinated attack or a coordinated retreat. The problem is complicated by the presence of traitorous generals who may not only cast a vote for a suboptimal strategy, but also do so selectively (i.e. different answers sent to different people). This is analogous to a number of nodes participating in a blockchain attempting to arrive at a global consensus whilst using unreliable communication and under threat of some participants malfunctioning or being malicious.

Consensus mechanism - a set of state transition rules enabling an economic set (among which the rights to conduct the transition are distributed) to perform secure update of the state (Buterin, 2014). Bitcoin users are an example of the aforementioned economic set. For further description and examples see A.2.

Crypto-currency - a digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank (The Oxford Dictionary, 2016).

Cryptographic hash function - a type of hash functions (see below) suitable for use in cryptography e.g. Bitcoin uses SHA-256 (Secure Hash Algorithm - 256 bit) (Rogaway and Shrimpton, 2004; Lewis, 2015). An ideal cryptographic hash function exhibits the following traits:

- a hash can be easily produced for any message;
- it is difficult to derive the original data from its hash;
- any changes in the original data result in the hash changing so extensively that the new hash value appears uncorrelated with the old hash value;
- it is infeasible for two different inputs to result in the same hash.

Fork - the event of a blockchain splitting into two or more chains. A fork can occur when two or more miners publish a valid block at roughly the same time, as a part of an attack (e.g. 51% attack) or when a blockchain protocol change is attempted (such a fork is "hard" if all users are required to upgrade, otherwise it is "soft") (Nakamoto, 2009b).

Hash function - any function that can be used to map data of arbitrary size to data of fixed size (Rogaway and Shrimpton, 2004; Lewis, 2015).

Header hash - a hash of the information contained in a block's header which is used to link the block with the next one. In the case of Bitcoin blocks it contains the blockchain version number, the header hash of the previous block, the merkle root (see *Merkle tree* below) of all transactions in the block, the current time and the current difficulty (see *Hash target* below).

Hash target - a set of acceptance criteria imposed on a block's header hash (see *Header hash* above) by the protocol of a blockchain. In the case of Bitcoin, the target is an upper bound on the hash's value.

Internet of Things (IoT) - dynamic, global network infrastructure that can self-configure using standards and interoperable protocols where physical and virtual things have identities, attributes, and personalities, use intelligent interfaces, and can seamlessly integrate into the network (Atzori, 2016).

Industry 4.0 - is characterised by the ability of industrial components to communicate with each other. It includes cyber-physical systems, the Internet of Things and cloud computing (Kleinelanghorst *et al.*, 2016; Hermann *et al.*, 2016).

Linked Data - a method of publishing structured data so that it can be interlinked and become more useful through semantic queries (Bizer *et al.*, 2009).

Merkle tree - a tree constructed by pairing data (e.g. in the Bitcoin system it usually refers to transactions), then hashing the pairs, then pairing and hashing the results until a single hash remains, the merkle root (Nakamoto, 2009c).

Mining - the process of verifying transactions and publishing blocks. The exact procedure varies widely depending on a particular blockchain implementation. In Bitcoin's case miners compete to solve a mathematical puzzle that requires the

consumption of computing power (Vukolić, 2016). Once the puzzle is solved, the new block of transactions is accepted by the network and committed to the blockchain. The miner is rewarded with newly generated coins. For further description and examples see A.2.

Node - any device which is part of a network and has a unique network address. In the context of blockchain and crypto-currencies it refers to a wallet software such as the Bitcoin client application.

Nonce - an arbitrary number that may only be used once. In the case of Bitcoin, it is a part of block's header and mining nodes repeatedly adjust the number in order to meet the target imposed on header hashes.

"Nothing at stake" problem - a shortfall experience by blockchain using a proof-of-stake consensus mechanisms where block generators have nothing to lose by voting for multiple blockchain histories leading to consensus never resolving (Poelstra, 2016).

Peer-to-peer (P2P) network - a network of nodes (peers) directly connected with each other. The system relies on the peers, who have equal standing within the network, sharing at least as many resources as they consume.

Permissioned blockchain - a blockchain whose use is restricted to known, vetted participants (O'Dair *et al.*, 2016).

Permissionless blockchain - a blockchain that is accessible to anyone who wishes to use it (O'Dair *et al.*, 2016).

Private blockchain - a blockchain that limits read access to particular users (O'Dair *et al.*, 2016).

Public blockchain - a blockchain that grants read access and ability to create transactions to all users (O'Dair *et al.*, 2016).

Public-private key cryptography - a class of encryption methods that uses pairs of keys (e.g. a pair of two special numbers): public and private. A public key can be used to verify that a message was created by an owner of the paired private key (verification of a digital signature) and to encrypt a message such that only the aforementioned owner can decrypt.

Smart contract - a contractual agreement built on computer protocols, whose terms are executed automatically (O'Dair *et al.*, 2016).

Semantic Web - an extension of the Web through standards by the World Wide Web Consortium (W3C) in order to promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF) (Nakamoto, 2013).

Transaction - a transfer of a digital asset from an address (or addresses) to another address (or addresses) (O'Dair *et al.*, 2016).

Wallet - a public representation of the public and private key pairs that are used to store and transfer coins.

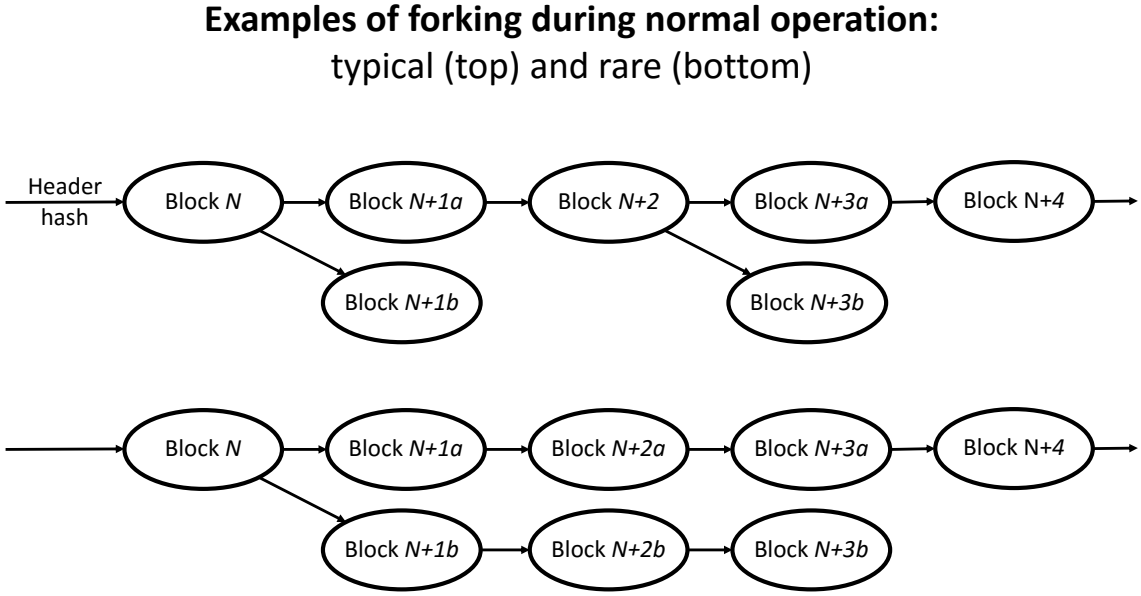


Figure A.1: *Diagram demonstrating temporary blockchain forks (Nakamoto, 2009a).*

A.2 Consensus mechanisms

This section briefly describes various kinds of consensus mechanisms (see definition in A.1) used for blockchain implementations. The following mechanisms are discussed:

- Proof-of-work
- Proof-of-stake
- Deposit-based
- Byzantine agreement (PBFT)
- Rotation scheme

Proof-of-work

Proof-of-work is a method of achieving network consensus where the ability to verify and publish transactions is dependent on the computing power of the miner (O'Dair *et al.*, 2016; Buterin, 2014). The details have already been explained in Section 2.5.

Proof-of-stake

Proof-of-stake is a consensus mechanism in which the ability to verify and publish blocks depends on the "stake" (e.g. amount of the native currency) already possessed (O'Dair *et al.*, 2016). Verification is performed by the nodes with the largest stake in the network as its correct operation is in their best interest, e.g. mining is easier for those who can show they control a large amount of the blockchain's native currency.

Publishing blocks proceeds as follows (Vukolić, 2016):

1. A participant needs to "lock" (e.g. deposit, spend) a number of coins in order to be allowed to publish blocks;

2. The participant needs to generate a block with a valid hash (as in the proof-of-work system, but the more coins consumed, the easier the search for a valid hash);

3. The block is published and validated by other participants.

The system benefits over the proof-of-work mechanism from reduced energy consumption, immunity from hardware centralization and reduced risk of any one member acquiring the controlling stake as its cost might be higher than the cost of acquiring significant mining power. However, it suffers from the "nothing at stake" problem (described in A.1).

Deposit-based

A deposit-based consensus protocol requires the participants to register a security deposit in order to serve the consensus by producing blocks (Zamfir, 2015). In the case of Ethereum, a chain selection rule called GHOST (Greedy Heaviest Observed Sub Tree) serves as an arbitrator governing the security deposits (Sompolinsky and Zohar, 2013). If a node validates a transaction that GHOST considers invalid, the node loses its deposit and forfeits the privilege of participating in the consensus process. This directly solves the "nothing-at-stake" problem (described in A.1). This system benefits from strong convergence of history (i.e. every block would either be fully abandoned or fully adopted) and strengthened immutability as blocks that are not in the main chain remain on the record (Sompolinsky and Zohar, 2013). Ethereum is expected to introduce a deposit-based consensus protocol called Casper (Zamfir, 2015).

Byzantine agreement

Byzantine agreement, also known as Practical Byzantine Fault Tolerance (PBFT), type consensus mechanisms are based on a solution to the Byzantine Generals Problem (described in A.1). In this case each node generates a private-public key pair and publishes the public key. Messages from other nodes, which are concerned with issues requiring the network agreement, passing through the node

are signed by the node to verify their format. Once enough identical responses are recorded, the consensus about the issue in question is reached. This protocol is suitable for low-latency storage system and digital asset-based platforms that do not require a large data throughput, but need many transactions (Vukolić, 2016; Castro and Liskov, 1999). One of the platforms using it is Hyperledger.

This method does not require any hashing power (hence enjoys reduced energy usage), provides fast and efficient consensus convergence and decouples trust from resource ownership making it possible for the small to keep the powerful honest. However, the system needs to be set up by a central authority or over a course of closed negotiations and all parties have to agree on the exact list of participants (Maziéres, 2015).

Federated Byzantine Agreement (FBA) is a type of PBFT, but it enjoys an open membership scheme (Maziéres, 2015), where all nodes know other nodes and can consider some to be important. Whenever a transaction needs to be verified, any given node waits for the vast majority of the nodes it considers important to agree with each other. At the same time, the important participants do not agree to the transaction until the participants they consider important agree as well. Eventually, sufficiently large part of the network accepts the transaction making it infeasible for an attacker to make any changes. The FBA system relies on small sets of trusted parties which would consist of the nodes that built their trust level over time through good behaviour (Maziéres, 2015). A platform called Stellar employs this scheme (Higgins, 2015).

Round robin

For private blockchains, where certain degree of trust between the participants is possible, the network consensus can be achieved without difficult computations. In the case of MultiChain (Greenspan, 2015) the set of miners is limited to known entities which take turns in publishing blocks. The strictness of the rotation scheme is controlled using a parameter called mining diversity ($0 \leqslant$ mining diversity $\leqslant 1$). This parameter defines the minimum proportion of permitted miners needed to control the network. 0.75 is a recommended value (Greenspan,

2015), as high values are safer, but a value too close to 1 can cause the blockchain to freeze up if some miners become inactive. In the case that the network splits temporarily (e.g. due to communications failure) resulting in a fork, the branch with the longer chain will be adopted.

The participants are approved for publishing blocks as follows:

1. Any permission changes defined by transactions in the current block are applied;
2. The current number of permitted miners is calculated;
3. The number of miners is multiplied by mining diversity and rounded up to get spacing;
4. If any of the spacing-1 blocks were mined by the current miner, the block is invalid.

The scheme enjoys the following advantages over a centralised database:

- Each participant has full control over its assets via their ownership of private key(s);
- Distributed control prevents an individual or a small group from unilaterally deciding which transactions are valid or will be confirmed;
- More robust as access and validation of transactions will continue even if a server malfunctions (i.e. no single point of failure).

Glossary

A Cyber Physical System (CPS) is a physical mechanism controlled or monitored by software where both aspects are deeply intertwined. It is usually connected to a network e.g. a control system of an electrical distribution network. Pages: 2, 9, 10

An Eco-Industrial Park (EIP) is 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 minimize environmental impact while simultaneously increasing business success (Boix *et al.*, 2015; Ehrenfeld and Gertler, 1997). Pages: iv, 8–10, 24

Industry 4.0 is characterised by the ability of industrial components to communicate with each other. It includes cyber-physical systems, the Internet of Things and cloud computing (Hermann *et al.*, 2016). Pages: iv, 2, 3, 5, 7, 9, 10, 12, 13, 20, 21, 23, 53, 57, 66, 69, 71, 87, 90, 91, 93

Machine-to-Machine (M2M) communication refers to the ability of industrial components to directly communicate with each other. Pages: 2, 4, 9, 23

The Internet of Things (IoT) is a dynamic, global network infrastructure that can self-configure using standards and interoperable protocols where physical and virtual things have identities, attributes, and personalities, use intelligent interfaces, and can seamlessly integrate into the network (Atzori, 2016). Pages: 2, 9, 22, 23, 59, 68

Nomenclature

Upper-case Roman

M^*	Maximum polynomial order
M'^*	Maximum interaction order
N	Size of a set (where $N \in \mathbb{N}$)
N	Number of data points
N_x	The number of input parameters, x
P_m	Legendre polynomials of order m (where $m \in \mathbb{N}$)
P_p	Price of a product
P_r	Price of a reagent
R	Units of the reagent per unit of the product
R^2	Coefficient of determination - a measure indicating fit of a statistical model to data
$\bar{R}^2$	R^2 corrected for the number of fitted parameters relative to the number of data points
R^{2*}	R^2 threshold for inclusion of surrogate terms
S_i	Sensitivity with respect to the ith input variable
S_{ij}	Sensitivity with respect to the interaction term of the ith and jth input variables
S_p	Storage cost of a product
S_r	Storage cost of a reagent

T_p	Transport cost of a product
T_r	Transport cost of a reagent
U	Cost of utilities per unit of the product
W	Statistical weights in the general linear least-square fit method
$W^{(1)}$	First statistical weight
$W^{(i)}$	ith statistical weight
$W^{(N)}$	Nth statistical weight

Lower-case Roman

argmin	A function that returns the arguments that minimise a given function
d	Storage duration for product or reagent
$e^{(i)}$	Residual error for i^{th} data point
f	A function
$f(x)$	A function in x
f_0	Mean value of $f(x)$
$\overline{f}$	Mean value of $f(x)$
$f^{(i)}$	i^{th} model predicted value
f_β	Polynomial in x with coefficient β
$f_i(x_i)$	Function describing surrogate contribution of the ith input variable (excluding the constant term and interaction terms)
f_{ij}	Function describing surrogate contribution of the interactions of the ith and jth input variables
$f_{12...N_x}(x_1, x_2, ..., x_{N_x})$	Function describing surrogate contribution of the interactions of N_x input variables
i	Index for ordering parameters
j	Index for ordering parameters
k	Polynomial order
p	Number of adjustable parameters
p	A multi-index

x	An independent variable
x	Polynomial
x^p	Polynomial in a set of variables x of order p
x_1^p	Polynomial in a single variable x_1 of order p
$x^{(1)}$	First input
$x^{(i)}$	ith input
$x^{(N)}$	Nth input
y	Set of outputs
$y^{(i)}$	ith output data point
$y^{(1)}$	First output
$y^{(i)}$	ith output
$y^{(N)}$	Nth output
$\overline{y}$	Mean value of y

Upper-case Greek

Φ	A function defining coefficients of a polynomial in the general linear least-square fit method

Lower-case Greek

$\alpha_{i,k}$	Coefficient used for evaluating an approximation of HDMR
β	Coefficients of terms in a polynomial
β^*	Optimised polynomial coefficients
$\beta_{ij,kl}$	Coefficient used for evaluating an approximation of HDMR
δ	Kronecker delta
ϕ	An analytic function used for evaluating an approximation of HDMR
σ	Standard deviation
σ^2	Variance

$\sigma^2_{\bar{y},i}$	Contribution of the *i*th input variable to the total variance in the output variable (excluding interaction terms)
$\sigma^2_{\bar{y},ij}$	Contribution of the interactions of the *i*th and *j*th input variables to the total variance in the output variable
$\sigma^2_{\bar{y}}$	Total variance in the output variable

Symbols

10D01	Continuously stirred tank reactor as seen on Fig. 3.2
10D02	Flush drum as seen on Fig. 3.2
10D02D	Decanter as seen on Fig. 3.2
10E01	Heater as seen on Fig. 3.2
10E02	Heater as seen on Fig. 3.2
10E03	Heater as seen on Fig. 3.2
H1	Denotes a 1^{st} order HDMR fit
H2a	Denotes a 2^{nd} order HDMR fit without interactions
H2b	Denotes a 2^{nd} order HDMR fit with interactions
H10	Denotes a 10^{th} order HDMR fit with 2^{nd} order interactions
P1	Denotes a 1^{st} order polynomial fit
P2	Denotes a 2^{nd} order polynomial fit
P3	Denotes a 3^{rd} order polynomial fit
P4	Denotes a 4^{th} order polynomial fit
P5	Denotes a 5^{th} order polynomial fit
$\mathbb{N}$	Set of natural numbers
$\mathbb{N}_0$	Set of natural numbers including 0
$\mathbb{N}_0^l$	l-dimensional set of natural numbers including 0
$\mathbb{R}^n$	Real coordinate space of n dimensions
$\oplus$	Exclusive disjunction
$\int$	Integration symbol
$\binom{n+k}{k}$	Binomial coefficient $(n+k)!/k!n!$

$\frac{\partial \Phi}{\partial \beta_q}$	Partial derivative of function Φ with respect to β_q

Abbreviations

AH	Aspen HYSYS
ANN	Artificial neural network
AP	Aspen Plus
API	Application programming interface
BTC	Bitcoin
CARES	Cambridge Centre for Advanced Research and Education in Singapore
CME	Chicago Mercantile Exchange
COM	Component Object Model
CoMo	Computational Modelling
CPS	Cyber-physical system
CSTR	Continuously stirred tank reactor
DDoS	Distributed denial-of-service
DMC	Dynamic matrix control
EIP	Eco-industrial park
eMPC	explicit Model Predictive Controller
FBA	Federated Byzantine Agreement
GHG	Greenhouse gas
GPC	Generalized Predictive Control
HDMR	High Dimensional Mathematical Representation
IMC	Internal model control
IoT	Internet of Things
JPS	J-Park Simulator
LQR	Linear quadratic regulator
M2M	Machine-to-Machine
MAC	Model algorithmic control

MoDS	Model Development Suite
MPC	Model Predictive Controller
P2P	Peer-to-peer
PBFT	Practical Byzantine Fault Tolerance
PFR	Plug flow reactor
PWA	Piecewise affine function
RDF	Resource Description Framework
RMSD	Root-mean-square deviation
SHA	Secure Hash Algorithm
W3C	World Wide Web Consortium
ZCE	Zhengzhou Commodity Exchange

Bibliography

Aitken, B. (2016). Methanex Investor Presentation. Date accessed: 29.05.2017. URL https://www.methanex.com/sites/default/files/investor/MEOH%20Presentation%20-%20June%202016_0.pdf.

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

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

Armstrong, L., Bashshur, R. (1958). Ecological Patterns and Value Orientations in Lebanon. *Public Opinion Quarterly* **22** (3), 406–415. doi:10.1086/266808.

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

AspenTech (2015). aspentech - Aspen Plus V8.6. Date accessed: 12.01.2016. URL http://www.aspentech.com/products/engineering/aspen-plus/.

AspenTech (2017a). aspentech - Aspen HYSYS V8.8. Date accessed: 29.05.2017. URL http://www.aspentech.com/products/aspen-hysys/.

AspenTech (2017b). aspentech - Aspen Plus V8.8. Date accessed: 29.05.2017. URL http://www.aspentech.com/products/Aspen-Plus/V88/.

Atzori, M. (May 2016). Blockchain-Based Architectures for the Internet of Things: A Survey. Date accessed: 24.10.2016. URL https://ssrn.com/abstract=2846810.

Aurora Energy Pty Ltd (2017). Electricity cost breakdown. Date accessed: 02.02.2017. URL https://www.auroraenergy.com.au/your-home/bills-and-payments/your-bill-explained/electricity-cost-breakdown#generation.

Azadi, P., Brownbridge, G., Kemp, I., Mosbach, S., Dennis, J. S., Kraft, M. (2015). Microkinetic Modeling of Fischer-Tropsch Synthesis over Cobalt Catalysts. *ChemCatChem* **7**, 137–143. doi:10.1002/cctc.201402662.

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

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

Bailleul, I. F., Man, P. L., Kraft, M. (2010). A Stochastic Algorithm for Parametric Sensitivity in Smoluchowski's Coagulation Equation. *Society for Industrial and Applied Mathematics Journal on Numerical Analysis* **48**, 1064–1086. doi:10.1137/090758234.

Bakr, M. H., Bandler, J. W., Madsen, K., Soandergaard, J. (2000). Review of the Space Mapping Approach to Engineering Optimization and Modeling. *Optimization and Engineering* **1**, 241–276. doi:10.1023/A:1010000106286.

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

Barnard, F. (1963). Education for management conceived as a study of industrial ecology. *The Vocational Aspect of Education* **15** (30), 22–26. doi:10.1080/03057876380000041.

BASF (2018). Circular economy: A visionary roadmap? Date accessed: 05.08.2018. URL https://www.basf.com/za/en/we-create-chemistry/creating-chemistry-magazine/resources-environment-and-climate/circular-economy.html.

Beacham, W. (2017). Time is ripe for chemical industry to grasp digitisation - Accenture. Date accessed: 14.03.2018. URL https://goo.gl/GrAr4a.

Bemporad, A., Morari, M., Dua, V., Pistikopoulos, E. N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica* **38** (1), 3 – 20. doi:10.1016/S0005-1098(01)00174-1.

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

Billick, I., Case, T. J. (1994). Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected? *Ecology* **75** (6), 1530–1543. URL http://www.jstor.org/stable/1939614.

Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems* **5** (3), 1–22. doi:10.4018/jswis.2009081901.

Bjorgan, R., Chen-Ching Liu, Lawarree, J. (1999). Financial risk management in a competitive electricity market. *IEEE Transactions on Power Systems* **14** (4), 1285–1291. doi:10.1109/59.801886.

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

Boukouvala, F., Ierapetritou, M. G. (2013). Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing. *Journal of Pharmaceutical Innovation* **8** (2), 131–145. doi:10.1007/s12247-013-9154-1.

Brands, S. (1993). An Efficient Off-line Electronic Cash System Based On The Representation Problem. Tech. rep. doi:10.1.1.41.8533.

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

Braumann, A., Man, P., Kraft, M. (2011). The Inverse Problem in Granulation Modeling - Two Different Statistical Approaches. *AIChE Journal* **57**, 3105–3121. doi:10.1002/aic.12526.

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

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

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

Buntix, J. P. (November 2014). IOTA: Internet of Things Without the Blockchain? Date accessed: 14.11.2016. URL http://bitcoinist.net/iota-internet-things-without-blockchain/.

Burger, C., Kuhlmann, A., Richard, P., Weinmann, J. (November 2016). Blockchain in the Energy Transition: a survey among decision-makers in the Germany energy industry. Tech. rep. URL https://www.dena.de/fileadmin/dena/Dokumente/Meldungen/dena_ESMT_Studie_blockchain_englisch.pdf.

Buterin, V. (November 2014). Consensus Mechanisms used in Blockchain. Date accessed: 18.10.2016. URL https://blog.ethereum.org/2014/11/25/proof-stake-learned-love-weak-subjectivity/.

Caballero, J. A., Grossmann, I. E. (October 2008). An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal* **54** (10), 2633–2650. doi:10.1002/aic.11579. URL http://onlinelibrary.wiley.com/doi/10.1002/aic.11579/full.

Campa, J. M. (1994). Multinational Investment under Uncertainty in the Chemical Processing Industries. *Journal of International Business Studies* **25** (3), 557–578. URL http://www.jstor.org/stable/155359.

Cardeira, H. (March 2015). Smart contracts and their applications to the construction industry. In *New Perspectives in Construction Law Conference, 2015*.

Castro, M., Liskov, B. (1999). Practical Byzantine fault tolerance. In *Proceedings of the Third Symposium on Operating Systems Design and Implementation*, OSDI '99, pp. 173–186. USENIX Association, Berkeley, CA, USA, date accessed: 09.12.2016. URL http://dl.acm.org/citation.cfm?id=296806.296824.

Chapman, W. L., Welch, W. J., Bowman, K. P., Sacks, J., Walsh, J. E. (1994). Arctic ice variability: Model sensitivities and a multidecadal simulation. *Journal of Geophysical Research* **99**, 919–935. doi:10.1029/93JC02564.

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

Chertow, M. (2000). Industrial symbiosis: Literature and taxonomy. *Annual Review of Energy and Environment* **25**, 313–337. doi:10.1146/annurev.energy.25.1.313.

Cicconi, P., Russo, A. C., Germani, M., Prist, M., Pallotta, E., Monteriù, A. (Sept 2017). Cyber-physical system integration for industry 4.0: Modelling and simulation of an induction heating process for aluminium-steel molds in footwear soles manufacturing. In *2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)*, pp. 1–6. doi:10.1109/RTSI.2017.8065972.

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

Clarke, D., Mohtadi, C., Tuffs, P. (1987a). Generalized predictive control — Part I. The basic algorithm. *Automatica* **23** (2), 137 – 148. doi:10.1016/0005-1098(87)90087-2.

Clarke, D., Mohtadi, C., Tuffs, P. (1987b). Generalized Predictive Control — Part II Extensions and interpretations. *Automatica* **23** (2), 149 – 160. doi:10.1016/0005-1098(87)90088-4.

Clift-Jennings, A. (November 2016). Foundations for the Next Economic Revolution: Distributed Exchange and the Internet of Things. Date accessed: 14.11.2016. URL https://filament.com/assets/downloads/Filament%20Foundations.pdf.

CMCL Innovations (2015). Model Development Suite (MoDS). Date accessed: 12.01.2015. URL http://www.cmclinnovations.com/mods/.

CME Group (2017a). FAME 0 Biodiesel FOB Rdam (Argus) (RED Compliant) futures quotes. Date accessed: 29.05.2017. URL http://www.cmegroup.com/trading/energy/refined-products/fame-0-argus-biodiesel-fob-rdam-red-compliant-swap-futures.html.

CME Group (2017b). Henry Hub Natural Gas Futures Quotes. Date accessed: 29.05.2017. URL http://www.cmegroup.com/trading/energy/natural-gas/natural-gas.html.

CME Group (2017c). USD Malaysian Crude Palm Oil Calendar Futures Quotes. Date accessed: 29.05.2017. URL http://www.cmegroup.com/trading/agricultural/grain-and-oilseed/usd-malaysian-crude-palm-oil-calendar.html?optionProductId=8075.

Coinnews Media Group LLC (2017). US Inflation Calculator. Date accessed: 29.05.2017. URL http://www.usinflationcalculator.com/.

Condos, J., Sorrell, W. H., Donegan, S. L. (January 2016). Blockchain Technology: Opportunities and Risks. Tech. rep., State of Vermont, USA. URL http://inpluslab.sysu.edu.cn/files/Paper/Summary/Blockchain_Technology_Principles_And_Applications.pdf.

Crary, S. B. (2002). Design of Computer Experiments for Metamodel Generation. *Analog Integrated Circuits and Signal Processing* **32**, 7–16. doi:10.1023/A:1016063422605.

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

Cutler, C., Ramaker, B. (1980). Dynamic Matrix Control - A Computer Control Algorithm. In *Automatic Control Conference, San Francisco, CA*.

Decker, C., Wattenhofer, R. (2014). Bitcoin Transaction Malleability and MtGox. In Kutyłowski, M., Vaidya, J. (Eds.), *Computer Security - ESORICS 2014: 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II*, pp. 313–326. Springer International Publishing. doi:10.1007/978-3-319-11212-1_18.

Ding, Q., Dong, L., Kouvelis, P. (June 2007). On the Integration of Production and Financial Hedging Decisions in Global Markets. *Operations Research* **55** (3), 470–489. doi:10.1287/opre.1070.0364.

Draper, N. R., Smith, H. (1998). *Applied Regression Analysis*, 3rd edition. Wiley-Interscience.

Edhouse, S. (April 2013). bittunes.org: An Independent Digital Music Market. Date accessed: 14.11.2016. URL http://www.bittunes.org/.

Ehrenfeld, J., Gertler, N. (1997). Industrial Ecology in Practice. The Evolution of Interdependence at Kalundborg. *Journal of Industrial Ecology* **1**, 67–79. doi:10.1162/jiec.1997.1.1.67.

EMA (2017). Electricity Tariffs. Date accessed: 29.05.2017. URL https://www.ema.gov.sg/Non_Residential_Programmes_Electricity_Tariffs.aspx.

Energy UK (2015). Energy bill breakdown. Date accessed: 02.02.2017. URL http://www.energy-uk.org.uk/customers/about-your-energy-bill/the-breakdown-of-an-energy-bill.html.

EnergyAustralia (2016). Understand energy prices & rates. Date accessed: 02.02.2017. URL https://www.energyaustralia.com.au/residential/electricity-and-gas/understanding-plans/understanding-prices-and-rates.

English, M., Auer, S., Domingue, J. (May 2016). Block Chain Technologies & The Semantic Web: A Framework for Symbiotic Development. Tech. rep., University of Bonn, Germany. URL http://cscubs.cs.uni-bonn.de/2016/proceedings/paper-10.pdf.

EYGM (2014). Competing for LNG demand: the pricing structure debate. Date accessed: 29.05.2017. URL https://webforms.ey.com/Publication/vwLUAssets/Competing-for-LNG-demand/$FILE/Competing-for-LNG-demand-pricing-structure-debate.pdf.

Fadhil, M. (2013). SE Asia-to-China palm oil freight rates may fall on weak demand. Date accessed: 29.05.2017. URL https://goo.gl/UUNe28.

Fahmi, I., Cremaschi, S. (November 2012). Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Computers & Chemical Engineering* **46**, 105–123. doi:10.1016/j.compchemeng.2012.06.006. URL http://www.sciencedirect.com/science/article/pii/S0098135412001822.

Fedora Project (2016). Fedora 24 Workstation. Date accessed: 25.11.2016. URL https://archive.fedoraproject.org/pub/fedora/linux/releases/24/Workstation/.

FERC (2004). Current state of and issues concerning underground natural gas storage. Date accessed: 29.05.2017. URL https://www.ferc.gov/EventCalendar/Files/20041020081349-final-gs-report.pdf.

Filatov, V., Afanas'ev, V. (1992). Differential heat-flux calorimeter. *Khimiya i Tekhnoliya Vody* **35**, 97–100.

First, E. L., Hasan, M. M. F., Floudas, C. A. (May 2014). Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. *AIChE Journal* **60** (5), 1767–1785. doi:10.1002/aic.14441. URL http://onlinelibrary.wiley.com/doi/10.1002/aic.14441/full.

Forrester, A., Sobester, A., Keane, A. (2008). *Engineering Design via Surrogate Modelling: A Practical Guide*, 1st edition. Wiley.

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

Fung, W., Hsieh, D. A. (April 2001). The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers. *Review of Financial Studies* **14** (2), 313–341. doi:10.1093/rfs/14.2.313.

García, C. E., Prett, D. M., Morari, M. (1989). Model predictive control: Theory and practice — A survey. *Automatica* **25** (3), 335 – 348. doi:10.1016/0005-1098(89)90002-2.

García, M. R., Vilas, C., Santos, L. O., Alonso, A. A. (2012). A robust multimodel predictive controller for distributed parameter systems. *Journal of Process Control* **22** (1), 60 – 71. doi:10.1016/j.jprocont.2011.10.008.

Garcia, C., Morshedi, A. (1986). Quadratic Programming Solution of Dynamic Matrix Control (QDMC). *Chemical Engineering Communication* **46**, 73–87.

Garcia, P., Leuthold, R. M. (September 2004). A selected review of agricultural commodity futures and options markets. *European Review of Agriculture Economics* **31** (3), 235–272. doi:10.1093/erae/31.3.235.

Garud, S. S., Karimi, I. A., Kraft, M. (2018). LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. *Computers and Chemical Engineering* **119**, 352 – 370. doi:10.1016/j.compchemeng.2018.09.008.

Geyera, P., Schlueter, A. (2014). 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. doi:10.1016/j.apenergy.2013.12.064.

Gmehling, J., Wittig, R., Lohmann, J., Joh, R. (2002). A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Industrial & Engineering Chemistry Research* **41** (6), 1678–1688. doi:10.1021/ie0108043.

Goodman, L. M. (September 2014). Tezos - a self-amending crypto-ledger. Date accessed: 22.10.2016. URL https://www.tezos.com/pdf/white_paper.pdf.

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

Green, R. J. (2008). Electricity Wholesale Markets: Designs Now and in a Low-carbon Future. *The Energy Journal* **0** (Special I), 95–124. URL https://ideas.repec.org/a/aen/journl/dn-se-a06.html.

Greenspan, G. (July 2015). MultiChain Private Blockchain. Date accessed: 22.10.2016. URL http://www.multichain.com/download/MultiChain-White-Paper.pdf.

GridSingularity (2016). About us. Date accessed: 25.10.2016. URL http://gridsingularity.com/.

Hasan, M. M. F., Baliban, R. C., Elia, J. A., Floudas, C. A. (2012). Modeling, simulation, and optimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 2. Pressure swing adsorption and vacuum swing adsorption processes. *Industrial and Engineering Chemistry Research* **51** (48), 15665–15682. doi:10.1021/ie301572n.

Hasan, M. M. F., First, E. L., Floudas, C. a. (2013). Cost-effective CO_2 capture based on in silico screening of zeolites and process optimization. *Physical Chemistry Chemical Physics: PCCP* **15**, 17601–18. doi:10.1039/c3cp53627k. URL http://www.ncbi.nlm.nih.gov/pubmed/24037279.

Hasse, F., von Perfall, A., Hillebrand, T., Smole, E., Lay, L., Charlet, M. (June 2016). Blockchain - an opportunity for energy producers and consumers? Date accessed: 20.10.2016. URL http://www.pwc.com/gx/en/industries/assets/pwc-blockchain-opportunity-for-energy-producers-and-consumers.pdf.

Heap, I. (July 2016). Mycelia For Music. Date accessed: 14.11.2016. URL https://myceliaformusic.org/.

Henao, C. A., Maravelias, C. T. (May 2011). Surrogate-based superstructure optimization framework. *AIChE Journal* **57** (5), 1216–1232. doi:10.1002/aic.12341.

Hermann, M., Pentek, T., Otto, B. (January 2016). Design Principles for Industrie 4.0 Scenarios. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 3928–3937. doi:10.1109/HICSS.2016.488.

Higgins, S. (April 2015). Jed McCaleb Talks Stellar's New Protocol for Consensus. Date accessed: 19.10.2016. URL http://www.coindesk.com/stellar-founder-jed-mccaleb-new-protocol/.

Hoffman, C. (1971). The Industrial Ecology of Small and Intermediate-sized Technical Companies: Implications for Regional Economic Development. Tech. Rep. 197411, Texas University, USA.

Holotiuk, F., Pisani, F., Moormann, J. (2017). The Impact of Blockchain Technology on Business Models in the Payments Industry. URL http://aisel.aisnet.org/wi2017/track09/paper/6.

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

Ji, X., Huang, S., Grossmann, I. E. (2015). Integrated Operational and Financial Hedging for Risk Management in Crude Oil Procurement. *Industrial & Engineering Chemistry Research* **54** (37), 9191–9201. doi:10.1021/acs.iecr.5b00903.

Joe, S., Kuo, F. Y. (2008). Constructing Sobol Sequences with Better Two-Dimensional Projections. *SIAM Journal on Scientific Computing* **30**, 2635–2654. doi:10.1137/070709359.

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

Kalman, R. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* **82** (1), 35–45. doi:10.1115/1.3662552.

Kannegiesser, M., Günther, H.-O., van Beek, P., Grunow, M., Habla, C. (jan 2009). Value chain management for commodities: a case study from the chemical industry. *OR Spectrum* **31** (1), 63–93. doi:10.1007/s00291-008-0124-9.

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

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

Kasparick, M., Wunder, G. (September 2018). Stable Wireless Network Control Under Service Constraints. *IEEE Transactions on Control of Network Systems* **5** (3), 946–956. doi:10.1109/TCNS.2017.2659624.

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

Kastner, C. A., Lau, R., Kraft, M. (2015). Quantitative tools for cultivating symbiosis in industrial parks; a literature review. *Applied Energy* **155**, 599 – 612. doi:10.1016/j.apenergy.2015.05.037.

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

Kemp, L. (July 2015). Everledger. Date accessed: 14.11.2016.
URL https://www.everledger.io/.

Keyser, R. D., Cauwenberghe, A. V. (1985). Extended Prediction Self-Adaptive Control. *IFAC Proceedings Volumes* **18** (5), 1255 – 1260. doi:10.1016/S1474-6670(17)60736-5.

Kishigami, J., Fujimura, S., Watanabe, H., Nakadaira, A., Akutsu, A. (August 2015). The Blockchain-Based Digital Content Distribution System. In *Big Data and Cloud Computing (BDCloud), 2015 IEEE Fifth International Conference on*, pp. 187–190. doi:10.1109/BDCloud.2015.60.

Klei, A., Moder, M., Winkler, G. (2017). Digital in chemicals: From technology to impact. Date accessed: 14.03.2018. URL https://www.mckinsey.com/industries/chemicals/our-insights/digital-in-chemicals-from-technology-to-impact.

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

Kleinelanghorst, M. J., Zhou, L., Sikorski, J., Shyh, E. F. Y., Aditya, K., Mosbach, S., Karimi, I., Lau, R., Garud, S., Zhang, C., Pan, M., Moirangthem, J., Sun, Y., Chhabra, P., Nurul, K., Yong, D., Sng, Y. R., Amaratunga, G., Maciejowski, J., Ong, H., Panda, S., Kraft, M. (2016). J-Park Simulator: Roadmap to Smart Eco-Industrial Parks. Tech. Rep. 174, Computational Modelling Group, University of Cambridge, date accessed: 09.12.2016. URL https://como.cheng.cam.ac.uk/index.php?Page=Preprints&No=174.

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

Koblitz, N., Menezes, A. J. (Jan. 2016). Cryptocash, Cryptocurrencies, and Cryptocontracts. *Des. Codes Cryptography* **78** (1), 87–102. doi:10.1007/s10623-015-0148-5.

Kong, L., Murat Sen, S., Henao, C. A., Dumesic, J. A., Maravelias, C. T. (2016). A superstructure-based framework for simultaneous process synthesis, heat integration, and utility plant design. *Computers and Chemical Engineering* doi:10.1016/j.compchemeng.2016.02.013.

Korpela, K., Hallikas, J., Dahlberg, T. (January 2017). Digital Supply Chain Transformation toward Blockchain Integration. In *Hawaii International Conference on System Sciences (HICSS) 2017)*. URL http://hdl.handle.net/10125/41666.

Koshy, P., Koshy, D., McDaniel, P. (2014). *An Analysis of Anonymity in Bitcoin Using P2P Network Traffic*, pp. 469–485. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-662-45472-5_30.

Kraft, D. (April 2011). Namecoin. Date accessed: 14.11.2016. URL https://namecoin.org/.

Kraft, M., Mosbach, S. (2010). The future of computational modelling in reaction engineering. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **368** (1924), 3633–3644. doi:10.1098/rsta.2010.0124.

Kwon, H., Tak, K., Cho, J. H., Kim, J., Moon, I. (2017). Integrated Decision Support Model for Hedge Trading and Production Planning in the Petrochemical Industry. *Industrial & Engineering Chemistry Research* **56** (5), 1267–1277. doi:10.1021/acs.iecr.6b03527.

Lacey, S. (February 2016). The Energy Blockchain: How Bitcoin Could Be a Catalyst for the Distributed Grid. Date accessed: 24.10.2016. URL https://goo.gl/tgQYgF.

Lamport, L., Shostak, R., Pease, M. (Jul. 1982). The Byzantine Generals Problem. *ACM Transactions on Programming Languages and Systems* **4** (3), 382–401. doi:10.1145/357172.357176.

Leeuw, V. (March 2017). Industrie 4.0 in the Chemical Industry - Covestro's View. Date accessed: 09.02.2018. URL https://industrial-iot.com/2017/03/industrie-4-0-chemical-industry/.

Lewis, A. (September 2015). A gentle introduction to blockchain technology. Date accessed: 18.10.2016. URL https://bitsonblocks.net/2015/09/09/a-gentle-introduction-to-blockchain-technology/.

Li, G., Wang, S.-W., Rabitz, H. (2002). Practical Approaches To Construct RS-HDMR Component Functions. *The Journal of Physical Chemistry A* **106**, 8721–8733. doi:10.1021/jp014567t.

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

Lilic, J. (November 2015a). Ethereum Enabled Community Energy Market Sharing Economy Phase 1: Renewable Energy Certificates (RECs), Microgrids, Smart Meters and Ethereum. Date accessed: 24.10.2016. URL https://goo.gl/WcgB9I.

Lilic, J. (November 2015b). Ethereum Enabled Community Energy Market Sharing Economy Phase 2: Here's How ConsenSys is Building The TransActive Grid. Date accessed: 24.10.2016. URL https://goo.gl/z5gjXa.

Lindman, J., Tuunainen, V. K., Rossi, M. (January 2017). Opportunities and Risks of Blockchain Technologies - A Research Agenda. In *Hawaii International Conference on System Sciences (HICSS) 2017)*. URL http://hdl.handle.net/10125/41338.

LO3 Energy (2016). Projects. Date accessed: 25.10.2016.
URL http://lo3energy.com/projects/.

Longinidis, P., Georgiadis, M. C., Kozanidis, G. (2015). Integrating Operational Hedging of Exchange Rate Risk in the Optimal Design of Global Supply Chain Networks. *Industrial & Engineering Chemistry Research* **54** (24), 6311–6325. doi:10.1021/acs.iecr.5b00349.

Lucia, S., Kögel, M., Zometa, P., Quevedo, D. E., Findeisen, R. (2016). Predictive control, embedded cyberphysical systems and systems of systems – A perspective. *Annual Reviews in Control* **41**, 193 – 207. doi:10.1016/j.arcontrol.2016.04.002.

Ludwin, A. (January 2014). Chain. Date accessed: 14.11.2016.
URL https://chain.com/.

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

Maziéres, D. (April 2015). The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus. Date accessed: 20.10.2016.
URL https://www.stellar.org/papers/stellar-consensus-protocol.pdf.

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

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

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

Microsoft (2015). COM technical overview. Date accessed: 02.06.2015. URL https://msdn.microsoft.com/en-us/windows/desktop.

Molin, B., Kimmoun, O., Remy, F., Chatjigeorgiou, I. (2014). Third-order effects in wave–body interaction. *European Journal of Mechanics - B/Fluids* **47**, 132 – 144, Enok Palm Memorial Volume. doi:/10.1016/j.euromechflu.2013.11.009.

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

Nakamoto, S. (2009a). Bitcoin Developer Guide. Date accessed: 17.11.2016. URL https://bitcoin.org/en/developer-guide.

Nakamoto, S. (2009b). Fork, Accidental Fork. Date accessed: 26.11.2016. URL https://bitcoin.org/en/glossary/fork.

Nakamoto, S. (2009c). Merkle Tree. Date accessed: 26.11.2016. URL https://bitcoin.org/en/glossary/merkle-tree.

Nakamoto, S. (2013). W3C SEMANTIC WEB ACTIVITY. Date accessed: 26.11.2016. URL https://www.w3.org/2001/sw/.

Nakamoto, S. (2016). Block #438995. Date accessed: 17.11.2016. URL https://blockchain.info/block/000000000000000000042b6b0a7bfa7f43b648167a5c0547d6b3a676e23d4c226.

Nguyen, C. (March 2016). An indie, off-the-grid, blockchain-traded solar power market comes to Brooklyn. Date accessed: 07.03.2017. URL https://goo.gl/shpfWb.

Nikolaou, M. (1998). Model Predictive Controllers: A Critical Synthesis of Theory and Industrial Needs.

Nuchitprasittichai, A., Cremaschi, S. (July 2013). Optimization of CO_2 Capture Process with Aqueous Amines-A Comparison of Two Simulation-Optimization Approaches. *Industrial & Engineering Chemistry Research* **52** (30), 10236–10243. doi:10.1021/ie3029366.

Nunes, G. C. (2001). Design And Analysis Of Multivariable Predictive Control Applied To An Oil-Water-Gas Separator: A Polynomial Approach. Ph.D. thesis, University of Florida.

Oanda (2017). Currency converter. Date accessed: 29.05.2017. URL https://www.oanda.com/currency/converter/.

O'Dair, M., Beaven, Z., Neilson, D., Osborne, R., Pacifico, P. (July 2016). Music On The Blockchain. Tech. Rep. 1, Blockchain For Creative Industries Research Cluster, Middlesex University, UK.

O'Dwyer, K. J., Malone, D. (June 2014). Bitcoin mining and its energy footprint. In *25th IET Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, pp. 280–285. doi:10.1049/cp.2014.0699.

Oracle Corporation (2018). Oracle VM VirtualBox 5.1. Date accessed: 09.11.2018. URL https://www.oracle.com/technetwork/server-storage/virtualbox/overview/index.html.

Pan, M., Sikorski, J., Akroyd, J., Mosbach, S., Lau, R., Kraft, M. (2016). Design technologies for eco-industrial parks: From unit operations to processes, plants and industrial networks. *Applied Energy* **175**, 305 – 323. doi:10.1016/j.apenergy.2016.05.019.

Pan, M., Sikorski, J., Kastner, C. A., Akroyd, J., Mosbach, S., Lau, R., Kraft, M. (2015). Applying Industry 4.0 to the Jurong Island Eco-industrial Park. *Energy Procedia* **150** doi:10.1016/j.egypro.2015.07.313.

Panikkar, S., Nair, S., Brody, P., Pureswaran, V. (January 2015). ADEPT: An IoT Practitioner Perspective. Tech. rep., IBM. URL https://goo.gl/3WhUpm.

Park, J., Park, S., Yun, C., Kim, Y. (2010). Integrated Model for Financial Risk Management in Refinery Planning. *Industrial & Engineering Chemistry Research* **49** (1), 374–380. doi:10.1021/ie9000713.

Peng, D.-Y., Robinson, D. B. (1976). A New Two-Constant Equation of State. *Industrial & Engineering Chemistry Fundamentals* **15** (1), 59–64. doi:10.1021/i160057a011.

Peterson, J. (October 2015). Augur. Date accessed: 14.11.2016. URL https://www.augur.net/.

Poelstra, A. (May 2016). A Treatise on Altcoins. Date accessed: 19.10.2016. URL https://download.wpsoftware.net/bitcoin/alts.pdf.

Python Software Foundation (2017). Python 3.5.0. Date accessed: 29.05.2017. URL https://www.python.org/downloads/release/python-350/.

Python Software Foundation (2018). Savoir 1.0.6. Date accessed: 09.11.2018. URL https://pypi.org/project/Savoir/.

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

Raman, V., Maasoumy, M., Donzé, A. (2014). Model Predictive Control from Signal Temporal Logic Specifications: A Case Study. In *Proceedings of the 4th ACM SIGBED International Workshop on Design, Modeling, and Evaluation of Cyber-Physical Systems*, CyPhy '14, pp. 52–55. ACM, New York, NY, USA. doi:10.1145/2593458.2593472.

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

Renner, G. T. (1947). Geography of Industrial Localization. *Economic Geography* **23** (3), 167–189. URL http://www.jstor.org/stable/141510.

Richalet, J., Rault, A., Testud, J., Papon, J. (1978). Model predictive heuristic control: Applications to industrial processes. *Automatica* **14** (5), 413 – 428. doi:10.1016/0005-1098(78)90001-8.

Rogaway, P., Shrimpton, T. (2004). *Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance*. Springer Berlin Heidelberg.

Rogers, B. (July 2016). Dot Blockchain Music. Date accessed: 14.11.2016. URL http://dotblockchainmusic.com/.

Rouhani, R., Mehra, R. K. (1982). Model algorithmic control (MAC); basic theoretical properties. *Automatica* **18** (4), 401 – 414. doi:10.1016/0005-1098(82)90069-3.

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

Ruppert, D., Wand, M. P., Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.

Rutkin, A. (2016). Blockchain aids solar sales. *New Scientist* **231** (3088), 22. doi:10.1016/S0262-4079(16)31558-5.

Ryu, J.-H. (2006). Multiperiod Planning Strategies with Simultaneous Consideration of Demand Fluctuations and Capacity Expansion. *Industrial & Engineering Chemistry Research* **45** (19), 6622–6625. doi:10.1021/ie060601k.

Sanchez-Gorostiaga, A., Bajić, D., Osborne, M. L., Poyatos, J. F., Sanchez, A. (2018). High-order interactions dominate the functional landscape of microbial consortia. *bioRxiv* doi:10.1101/333534.

Sasson, E. B., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., Virza, M. (May 2014). Zerocash: Decentralized Anonymous Payments from Bitcoin. In *2014 IEEE Symposium on Security and Privacy*, pp. 459–474. doi:10.1109/SP.2014.36.

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

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

Shi, X., Li, Y., Cao, Y., Tan, Y., Xu, Z., Wen, M. (November 2015). Model predictive control considering cyber-physical system to dampen low frequency oscillation of interconnected power systems. In *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pp. 1–5. doi:10.1109/APPEEC.2015.7380996.

Simpson, T. W., Peplinski, J. D., Koch, P. N., Allen, J. K. (2001). Metamodels for Computer-based Engineering Design: Survey and recommendations. *Engineering with Computers* **17**, 129–150. doi:10.1007/PL00007198.

Smith, C. W., Stulz, R. M. (December 1985). The Determinants of Firms' Hedging Policies. *The Journal of Financial and Quantitative Analysis* **20** (4), 391. doi:10.2307/2330757.

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

Sobol, I. M. (1990). Sensitivity Estimation for Nonlinear Mathematical Models. *Matematicheskoe Modelirovanie* **2**, 112–118.

SolarCoin (2016). SolarCoin. Date accessed: 25.10.2016. URL http://solarcoin.org/en/front-page/.

Sompolinsky, Y., Zohar, A. (December 2013). Accelerating Bitcoin's Transaction Processing: Fast Money Grows on Trees, Not Chains. Tech. rep., Hebrew University of Jerusalem.

Spagnuolo, M., Maggi, F., Zanero, S. (2014). Bitiodine: Extracting intelligence from the bitcoin network. In *International Conference on Financial Cryptography and Data Security*, pp. 457–468. Springer Berlin Heidelberg.

Spinler, S., Huchzermeier, A., Kleindorfer, P. (2003). Risk hedging via options contracts for physical delivery. *OR Spectrum* **25** (3), 379–395. doi:10.1007/S00291-003-0128-4.

Szakmary, A. C., Shen, Q., Sharma, S. C. (February 2010). Trend-following trading strategies in commodity futures: A re-examination. *Journal of Banking & Finance* **34** (2), 409–426. doi:10.1016/j.jbankfin.2009.08.004.

Tanlapco, E., Lawarree, J., Liu, C. C. (July 2002). Hedging with Futures Contracts in a Deregulated Electricity Industry. *IEEE Power Engineering Review* **22** (7), 54–54. doi:10.1109/MPER.2002.4312381.

Technavio (2017). The Evolving Landscape of Digitalization in the Chemicals Industry. Date accessed: 14.03.2018. URL https://www.technavio.com/blog/evolving-landscape-digitalization-chemicals-industry.

The Oxford Dictionary (2016). *The Oxford Dictionary*. Oxford University Press, date accessed: 26.11.2016. URL https://en.oxforddictionaries.com/definition/cryptocurrency.

Vandervort, D., Gaucas, D., Jacques, R. S. (2015). *Issues in Designing a Bitcoin-like Community Currency*, pp. 78–91. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-662-48051-9_6.

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

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

Vukolić, M. (2016). The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication. In Camenisch, J., Kesdoğan, D. (Eds.), *Open Problems in Network Security: IFIP WG 11.4 International Workshop, iNetSec 2015, Zurich, Switzerland, October 29, 2015, Revised Selected Papers*, pp. 112–125. Springer International Publishing. doi:10.1007/978-3-319-39028-4_9.

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

Watanabe, C. (1972). Industrial ecology: Introduction of Ecology into Industrial Policy. Tech. rep., Ministry of International Trade and Industry (MITI), Tokyo.

Watanabe, H., Fujimura, S., Nakadaira, A., Miyazaki, Y., Akutsu, A., Kishigami, J. J. (October 2015). Blockchain contract: A complete consensus using blockchain. In *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, pp. 577–578. doi:10.1109/GCCE.2015.7398721.

Wilkinson, S., Quinn, J. (January 2015). Stroj. Date accessed: 14.11.2016. URL https://storj.io/.

XE.com Inc. (2017). Currency converter. Date accessed: 29.05.2017. URL http://www.xe.com/currencyconverter/.

Xiaochi Zhou, Andreas Eibeck, M. Q. L. N. K. M. K. (2019). An agent composition framework for the J-Park Simulator - a knowledge graph for the process industry. Tech. Rep. 227, University of Cambridge.

Yapp, E. K., Patterson, R. I., Akroyd, J., Mosbach, S., Adkins, E. M., Miller, J. H., Kraft, M. (2016). Numerical simulation and parametric sensitivity study of optical band gap in a laminar co-flow ethylene diffusion flame. *Combustion and Flame* **167**, 320 – 334. doi:10.1016/j.combustflame.2016.01.033.

Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K. (October 2016). Where Is Current Research on Blockchain Technology? - A Systematic Review. *PLoS ONE* **11** (10), 1–27. doi:10.1371/journal.pone.0163477.

Zamfir, V. (August 2015). Introducing Casper "the Friendly Ghost". Date accessed: 18.10.2016. URL https://blog.ethereum.org/2015/08/01/introducing-casper-friendly-ghost/.

ZCE (2017). Methanol contract daily data. Date accessed: 29.05.2017. URL http://english.czce.com.cn/enportal/DFSStaticFiles/Future/EnglishFutureQuotesMA.htm.

Zhang, C., Zhou, L., Pan, M., Romagnoli, A., Kraft, M. (2017). Towards Intelligent Thermal Energy Management of Eco-industrial Park through Ontology-based Approach. *Energy Procedia* **105**, 3295–3300. doi:10.1016/j.egypro.2017.03.745.

Zhang, Y., Wen, J. (February 2015). An IoT electric business model based on the protocol of bitcoin. In *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on*, pp. 184–191. doi:10.1109/ICIN.2015.7073830.

Ziegeldorf, J. H., Grossmann, F., Henze, M., Inden, N., Wehrle, K. (2015). CoinParty: Secure Multi-Party Mixing of Bitcoins. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, CODASPY '15, pp. 75–86. ACM, New York, NY, USA. doi:10.1145/2699026.2699100.

Zyskind, G., Nathan, O., Pentland, A. (May 2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *Security and Privacy Workshops (SPW), 2015 IEEE*, pp. 180–184. doi:10.1109/SPW.2015.27.

Citation index