Closed-Loop Multitarget Optimization for Discovery of New Emulsion Polymerization Recipes

Claudia Houben,[†] Nicolai Peremezhney,[†] Alexandr Zubov,[‡] Juraj Kosek,[‡] and Alexei A. Lapkin[*,†]

[†]Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, New Museums Site, Cambridge CB2 3RA, United Kingdom

[‡]Department of Chemical Engineering, University of Chemistry and Technology, Technicka 5, 166 28 Prague, Czech Republic

S *Supporting Information*

ABSTRACT: Self-optimization of chemical reactions enables faster optimization of reaction conditions or discovery of molecules with required target properties. The technology of self-optimization has been expanded to discovery of new process recipes for manufacture of complex functional products. A new machine-learning algorithm, specifically designed for multiobjective target optimization with an explicit aim to minimize the number of "expensive" experiments, guides the discovery process. This "black-box" approach assumes no a priori knowledge of chemical system and hence particularly suited to rapid development of processes to manufacture specialist low-volume, high-value products. The approach was demonstrated in discovery of process recipes for a semibatch emulsion copolymerization, targeting a specific particle size and full conversion.

Experimental platforms enabling automated search for optimal process conditions,[1−4] new chemical structures,[5] or new bioactive compounds[6] emerged very recently as a disruptive technology that has a potential to change the way chemical processes are developed across many areas of molecular sciences. The technology delivers increased throughput of a single investigator and provides a discovery platform for new entities through a directed search, as opposed to trial-and-error of high-throughput methods.

This technology was made possible through miniaturization of experiments both in batch and flow modes and the use of in-line and in situ analytics. The opportunities for directed discovery are created through coupling of automated experiments with the decision-making software.

Automated experiment systems were recently demonstrated for multistep organic syntheses[4,7−9] or discovery of unknown inorganic clusters[5] in continuous flow microreactors. This type of chemical experiments naturally lends itself to automate the decision-making process through use of design of experiments (DoE) algorithms. To achieve this automation, real time chemical sensing must be available to the decision-making algorithm. First steps regarding decision-making algorithms were exemplified on heterogeneous catalytic reactions in combination with in line GC analysis.[2,3] These approaches were recently expanded to other chemical sensing techniques, such as MS, NMR, IR, and so forth.[2,10−16]

Notably, in most published examples, simple optimization routines were successfully applied to automate the design of experiments. This is mainly due to fairly simple problems being explored thus far, with very few input variables, few measured variables, and, usually, a single optimization criteria/target. Simple optimization algorithms require large numbers of experiments to converge and thus restrict the approach to inexpensive experiments. They also do not allow to explore systems with complex multimodal output space.

We argue that without a significant advance in optimization and design of experiments algorithms the technology for self-optimization would be resigned to very simple tasks and would not allow exploration of the more interesting challenges of *discovery* within molecular sciences. Most real challenges are either experimentally expensive, have high dimensionality of input variable space, or/and have several desired features and hence require multiobjective optimization.

To exemplify this hypothesis, we turned to the problem of discovery of recipes in emulsion copolymerization. Although emulsion polymerization has been studied for many decades,[17] the process is still not fully understood. It is a complex multiphase system consisting of monomer droplets, latex particles, and continuous aqueous phase and is strongly influenced by temperature, pH, surfactant, and inhibitor concentrations and types, ratio of monomers, and so forth. Closed loop control of emulsion polymerization industrial plants requires a fully predictive model.[18−23] However, many industrial plants are run on the basis of predetermined recipes mainly due to the absence of robust physical models. Inspired by the huge impact copolymerization of styrene and butyl acrylate has in industry and research,[24−30] we studied this two-monomer system as an example of a complex process leading to a functional product.

The experimental optimization problem was designed to simulate the process of discovery of a completely new process. Any prior knowledge of the process was deliberately ignored, and only physical constraints were taken into account. This resulted in the problem including 14 input variables and two targets: full monomer conversion and a 100 nm product latex particle size. We formulated the problem of discovery as that of target optimization. An attractive strategy for target optimization in the case of multiple objectives and large number of input variables is to combine sequential experiments with a

DoE algorithm that would take into account the outcomes of previous experiments to select the next experiment. One of the most prominent algorithms for such sequential optimization is the Efficient Global Optimization (EGO) algorithm.[31] Wenzel et al., for instance, made use of the concepts of desirability[32] and virtual observations[33] to construct an algorithm capable of identifying and, with each iteration, improving on a cluster of solutions that best associate with target values. Even though the algorithm undoubtedly explores globally throughout the search, it is not designed to actively search for solutions that would allow one to gain the most information about the underlying process (i.e., solutions optimal in terms of experimental design).

The implementation of machine learning algorithms in closed-loop systems for chemical reactions or process discovery has not been reported. However, various advanced optimization algorithms had been developed for different applications, such as automatic chromatographic separation of human serum and yeast fermentation,[34] tuning the transition time of polymer reactions[35] or optimization of polymerization under power failure conditions,[36] or developing predictive models of catalytic activity using nonlinear regression of statistical models based on data mining.[37] We must emphasize that conventional design-of-experiments approaches, such as factorial design, were frequently employed to investigate emulsion polymerization about 20−30 years ago. The technique allowed to sample experimental space and identify very roughly the regions of emulsion instability[38] but could not be used for automated discovery of recipes for two main reasons, apart from the poorer state of development in analytical methods, chemo-informatics, and computing power at the time: the highly nonlinear nature of the problem and its high dimensionality result in poor prediction of infeasible (unstable) experimental regions, which in turn result in large number of failed experiments. These are overcome, in part, by the machine learning approach, which is better suited to multidimensional problems. The issue of the lack of predicting power with respect of emulsion stability remains but is approached with the use of a nonlinear multidimensional classification algorithm, vide infra.

Recently we developed an algorithm that specifically addresses the challenge of experimental sequential optimization and discovery under the constraints of expensive experiments. The resulting multi objective active learner (MOAL) algorithm utilizes both Gaussian Processes surrogate models and evolutionary algorithms. It was shown to outperform other similar algorithms in silico.[39]

We first adopted a physical model of emulsion copolymerization[18] to perform in silico target optimization driven by the MOAL algorithm. This process was run for 84 virtual experiments and discovered 18 unique recipes satisfying the target of high conversion (≥99%) and particle diameter of ca. 100 nm (within 1%); see Figure 1. This indicates that the optimization problem is highly multimodal. Throughout the optimization, scattering around the target of particle size is more pronounced than that of the target of full conversion, which is consistent with theory of emulsion polymerization.[40] The particle size is affected by many factors: transport of monomer, free radicals and surfactant to the growing particles, and partition of these reagents among the continuous aqueous phase, emulsified monomer droplets (monomer reservoir), monomer-swollen polymer particles (primary reaction loci), and oil−water interface.
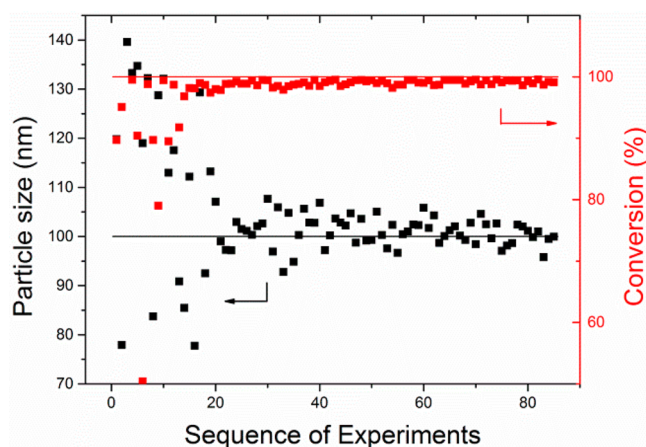

Figure 1. Results of in silico discovery of new copolymerization recipes. The two straight lines indicate the targets: particle size 100 nm and 100% conversion.

The first experiment that reached both targets was experiment number 27, excluding the 15 initial training experiments. This is quite remarkable, given the 14-dimensional input space of the problem. However, we should clarify that we cannot use this particular physical model to reliably define an optimal decision space size with respect of the overall decision space due to the validity of the model. Out of the 84 simulated recipes, only 18 reached both targets of full conversion and particle size diameter of 100 nm simultaneously. The list of the 84 simulated experiments is given in Supporting Information, Table S1, and results with respect of the optimization targets are shown in Table S2. The 84 simulated experiments do not include the 15 experiments of the initial training set.

Some of the 18 recipes were tested in the laboratory. Results of the validation experiments are shown in Supporting Information, Table S3. This revealed that the assumption of the monomer-starved emulsion polymerization was not satisfied in practice and the physical model of the process did not adequately represent the actual polymerization process. However, high conversion was obtained in three experimentally tested recipes. The second target, particle size of 100 nm in diameter, was not obtained in any recipe predicted by the in silico optimization. However, two recipes produced particle sizes of 135 and 178 nm, which where the closest values to the target.

To perform sequential discovery of new recipes without a priori process knowledge, we included a classification algorithm, which learns infeasible process conditions, in this case corresponding to reactor fouling due to coagulation and precipitation. The MOAL model was trained with only five experiments to explore the decision space. Two early experiments (Nos. 2 and 5) were infeasible due to reactor fouling. None of the initial five experiments reached the set targets. It was expected that, due to the low number of the initial set of experiments, the algorithm would require more iterations to discover successful recipes. Only one experiment failed during the optimization run. The failure was caused by too-high reaction temperature and the consecutive reactor fouling. However, small number of failed experiments suggests the successful implementation of the classification algorithm. Results of sequential process discovery are shown in Figure 2, and the complete set of obtained recipes is shown in Supporting Information, Table S4.
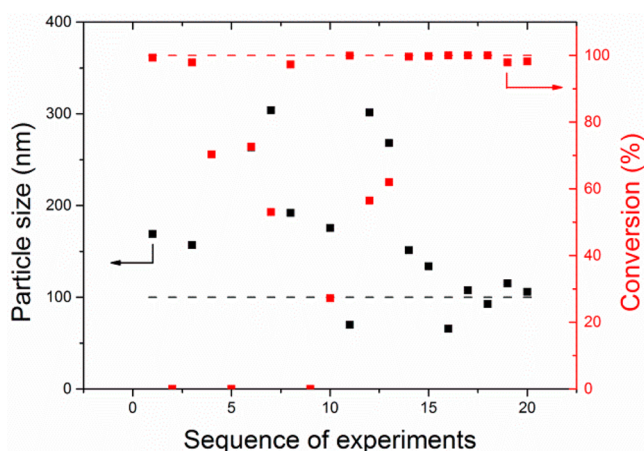
Figure 2. Results of the MOAL optimization of emulsion copolymerization carried out in the laboratory. The dotted lines indicate the targets to be reached.

The first experiment that reached both targets simultaneously was experiment 17 (including the five experiments from the initial training set). In this case we accepted 10% error in particle size and 5% error in conversion. As expected, following the results obtained from the simulations, the more difficult target to reach was the particle diameter. Despite the fact that obtaining particles with a certain size is a rather complex physical process, the sequential discovery processes reached both targets from iteration 17 onward. The optimization was stopped after 20 experiments.

One experiment reached only 27% conversion with the reaction time of 163 min. This is the lowest conversion achieved from all of the feasible experiments. The reason for the low conversion could be the high solid content (60% wt) and/or the low amount of the initiator solution (0.5 mL, 0.13 g mL^{-1}). The other five experiments with low conversion ($x <$ 75%) had in general low amount of initiator fed into the reactor.

A high amount of solid content can lead to coagulation and fouling, so it was not surprising that the experiments which failed or did not reach the targets had high solid content.

Although in this discovery process all 14 input variables were treated as independent, it is known from polymerization theory that some of those must be correlated. However, in the experimental results obtained, no pairwise correlations between input variables were observed (see Supporting Information, Figures S4−S6). Most likely this is due to the very small number of experimental points and a highly multimodal nature of the problem.

Comparing the successful in silico recipes with the recipes performed in the experimental system the main differences are (i) lower temperature for the laboratory experiments, (ii) in general higher ratio of initiator to monomer in the in silico recipes and (iii) all in silico recipes were considered feasible.

The model copolymerization system is well-studied and is produced commercially.[24,29,41] Therefore, we can compare the known industrial recipe (Supporting Information) with those discovered through the sequential experimental optimization. A number of differences between the recipes can be highlighted. The total feeding time in the discovered recipes is much shorter than in the commercial one, but the total reaction times are similar. In the last four discovered recipes the amount of the initiator changed, but generally speaking, the discovered recipes

still have a higher concentration of initiator than the recipe used in industry. Similarly, for the solid content, the solid content changed from 18% (Exp. 17) to 44% (Exp. 20). A higher solid content is preferential in industry. However, experiment 20 is close to the solid content used in the commercial recipe. There are also differences between the discovered recipes and the recipe used in the industry in the amount of surfactant and the ratio of initiator to monomer. The discovered recipes contained more surfactant and initiator. Consequently, the discovered recipes are likely to be more expensive than the recipe used in industry. This implies that the multitarget optimization should also include cost target.

■ CONCLUSION

In conclusion, we demonstrated for the first time the discovery of a new process, assuming little a priori knowledge of the chemical system. The developed approach is generic and allows multitarget optimization, automated learning of feasible process conditions, and minimizing the number of experiments. The approach is particularly suited to systems with large dimensionality of input variables and multimodal optimization surface. The approach was used to discover new recipes for the required targets of high conversion and particle diameter, without any prior knowledge of emulsion polymerization process. The system was capable to discover, after 27 experiments in silico and after 17 experiments in closed-loop experiments, a recipe to obtain a product with the desired property in a process with the required conversion.

■ EXPERIMENTAL SECTION

Materials. Styrene (ST, Aldrich, 98%), butyl acrylate (BA, Sigma-Aldrich, 99%), sodium dodecyl sulfate (SDBS, Sigma, 99%), *tert*-dodecylmercaptan (CTA, Sigma 98.5%), and sodium persulfate (NaPS, Aldrich) were all used as received. Super-Q water was used in all experiments. Polystyrene seed particles were produced in house.

Methods. The particle size was measured off line by dynamic light scattering (DLS, Malvern Zetasizer Nano-ZS). Conversion was monitored by GC-FID with an internal standard (toluene).

Reaction Setup. A reaction system, schematically shown in Figure S3 (Supporting Information) included a 0.5 L double-jacketed reactor with a four-blade turbine impeller, two syringe (Harvard PHD Ultra), and one metering (Eldex Optos) pumps. The pumps and the heating unit were controlled through a dedicated LabVIEW program. The information transfer, namely, the next recipe suggested by the MOAL algorithm, from MATLAB to LabVIEW was established via DataHub middleware.

MOAL Algorithm. A detailed description of the MOAL algorithm can be found elsewhere.[42,43] For this work the algorithm was adapted to the problem of target value optimization with unknown constrains to account for infeasible process conditions resulting in experimental failures. In the specific case of emulsion copolymerization stable and unstable solutions were labeled as feasible (1) and infeasible (−1), respectively. As a result, the next recipe to perform is chosen from the part/s of the decision space that are predicted (by the classification model) to consist of feasible recipes only. The highly dimensional decision space of 14 variables was chosen to allow the discovery of new recipes for the target of high conversion and particle size of 100 nm. The 14 variables were

relaxed as much as possible to allow the discovery of new recipes with the MOAL algorithm. However, only physical constraints were made to obtain feasible recipes, as for the amount of surfactant and initiator the water solubility was taken into account, or in another example for the reaction temperature, the activation temperature of the initiator and the boiling point of water were taken into account. The algorithm was implemented in Matlab.

In Silico Optimization. A first-principles model of emulsion copolymerization of styrene and butyl acrylate in a semibatch processes[18] was used for in silico optimization and discovery. Both models were implemented in Matlab.

Experimental Closed-Loop Optimization. Each experiment was carried out in a similar way. First, the four solutions were prepared: starting reactants, initiator solution (feed 1), surfactant solution (feed 2), and monomer solution (feed 3). The solutions were then degassed for an hour. After the starting reactants were heated up to the required reaction temperature, the feeding started. The stirrer speed was 400 rpm in all experiments. The statistical model was programmed in such a way that feeding rate of the initiator solution could be changed into feeding time 1.1 and feeding time 1.2, whereas feeding rate of the monomer and surfactant solutions where constant over the entire feeding time. The reaction was carried out under nitrogen atmosphere. Analysis of particle size and conversion was performed off line, and manual experimental setup and reactor cleaning were performed.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.oprd.5b00210.

> Definition of input variables and their constraints, a full list of in silico and experimental optimization recipes, graphs showing correlations between targets and some input variables, description of experimental setup and details of an industrial recipe which the obtained results are compared with. The published recipes provide all underlying research data to comply with EPSRC research data policy. (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: aal35@cam.ac.uk. Fax: + 44 1223 334796.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Bourne, R. A.; Skilton, R. A.; Parrott, A. J.; Irvine, D. J.; Poliakoff, M. *Org. Process Res. Dev.* **2011**, *15*, 932−938.
(2) Skilton, R. A.; Parrott, A. J.; George, M. W.; Poliakoff, M.; Bourne, R. A. *Appl. Spectrosc.* **2013**, *67*, 1127−1131.
(3) McMullen, J. P.; Stone, M. T.; Buchwald, S. L.; Jensen, K. F. *Angew. Chem., Int. Ed.* **2010**, *49*, 7076−7080.
(4) Smith, C. J.; Nikbin, N.; Ley, S. V.; Lange, H.; Baxendale, I. R. *Org. Biomol. Chem.* **2011**, *9*, 1938−1947.
(5) Richmond, C. J.; Miras, H. N.; de la Oliva, A. R.; Zang, H.; Sans, V.; Paramonov, L.; Makatsoris, C.; Inglis, R.; Brechin, E. K.; Long, D.-L.; Cronin, L. *Nat. Chem.* **2012**, *4*, 1037−1043.
(6) Czechtizky, W.; Dedio, J.; Desai, B.; Dixon, K.; Farrant, E.; Feng, Q.; Morgan, T.; Parry, D. M.; Ramjee, M. K.; Selway, C. N.; Schmidt, T.; Tarver, G. J.; Wright, A. G. *ACS Med. Chem. Lett.* **2013**, *4*, 768−772.
(7) Ingham, R. J.; Battilocchio, C.; Hawkins, J. M.; Ley, S. V. *Beilstein J. Org. Chem.* **2014**, *10*, 641−652.
(8) Lange, H.; Carter, C. F.; Hopkin, M. D.; Burke, A.; Goode, J. G.; Baxendale, I. R.; Ley, S. V. *Chem. Sci.* **2011**, *2*, 765.
(9) Guetzoyan, L.; Ingham, R. J.; Nikbin, N.; Rossignol, J.; Wolling, M.; Baumert, M.; Burgess-Brown, N.; Strain-Damerell, C. M.; Shrestha, L.; Brennan, P. E.; Fedorov, O.; Knapp, S.; Ley, S. V. *MedChemComm* **2014**, *5*, 540.
(10) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. *Science* **2015**, *347*, 49−53.
(11) Danieli, E.; Perlo, J.; Blümich, B.; Casanova, F. *Angew. Chem., Int. Ed.* **2010**, *49*, 4133−4135.
(12) Sans, V.; Porwol, L.; Dragone, V.; Cronin, L. *Chem. Sci.* **2015**, *6*, 1258−1264.
(13) Santos, A. F.; Silva, F. M.; Lenzi, M. K.; Pinto, J. C. *Polym.-Plast. Technol. Eng.* **2005**, *44*, 1−61.
(14) Vieira, R. A. M.; Sayer, C.; Lima, E. L.; Pinto, J. C. *Ind. Eng. Chem. Res.* **2002**, *41*, 2915−2930.
(15) Van den Brink, M.; Pepers, M.; van Herk, A. M.; German, A. L. *Polym. React. Eng.* **2001**, *9*, 101−133.
(16) Van den Brink, M.; German, A. L.; van Herk, A. M. *Process Control Qual.* **1999**, *11*, 265−275.
(17) Dinsmore, R. P. Synthetic Rubber and Method of Making It. U.S. Patent 1732795A, 1929.
(18) Zubov, A.; Pokorny, J.; Kosek, J. *Chem. Eng. J.* **2012**, *207−208*, 414−420.
(19) Mauntz, W. A Contribution to Observation and Time-optimal Control of Emulsion Co-Polymerization Reactions. Dissertation, TU Dortmund, 2010.
(20) Arora, S. Modeling and Control of Emulsion Polymerization Processes with Evaporative Cooling. Dissertation, TU Dortmund, 2008.
(21) Gesthuisen, R. Prozessvisualiserung von Polymerizationsprozessen mittels nichtlinearer Zustandsschätzung. Dissertation, TU Dortmund, 2001.
(22) Krämer, S. Heat Balance Calorimetry and Multirate State Estimation Applied to Semi-Batch Emulsion Copolymerization to Achieve Optimal Control. Dissertation, TU Dortmund, 2005.
(23) Pelz, K.; Brandt, H.; Finkler, T. F.; Engell, S. In *8th IFAC Symposium on Advanced Control of Chemical Processes*; Furama Riverfront, Singapore, 2012.
(24) Hagiopol, C. Styrene-acrylate copolymer composition suitable for surface size. U.S. Patent 6734232 B2, 2004.
(25) Santos, A. M.; Coutinho, F. M. B. *Polym. Bull.* **1993**, *30*, 407−414.
(26) Cruz-Rivera, A.; Rios-Guerrero, L.; Monnet, C.; Schlund, B.; Guillot, J.; Pichot, C. *Polymer* **1989**, *30*, 1872−1882.
(27) Chrástová, V.; Ďuračková, S.; Mrenica, J.; Černáková, L. *Chem. Papers* **1999**, *53*, 140−144.
(28) Moffat, K. A.; Pontes, F. M.; Paine, A. J.; McAneney, B.; Puri, P. Styrene-butyl acrylate toner resins with excellent gloss and fix properties. U.S. Patent 5462828, 1995.
(29) Lee, J.; Kim, S.; Yon, K. Preparation method of toner having micro radius. U.S. Patent 271970 A1, 2005.
(30) Paine, A. J.; Pontes, F. M.; Moffat, K. A. Starve fed emulsion polymerization process. U.S. Patent 5444140, 1995.
(31) Jones, D. R.; Schonlau, M.; Welch, W. J. *J. Glob. Optim.* **1998**, *13*, 455−492.

(32) Harrington, E. C. J. *Ind. Qual. Control* **1965**, *10*, 494−498.

(33) Cox, D. D.; John, S. In *Multidisciplinary Design Optimization: State of the Art*; Alexandrov, N., Hussaini, M. Y., Eds.; SIAM Philadelphia: Philadelphia, 1997; pp 315−329.

(34) O'Hagan, S.; Dunn, W. B.; Brown, M.; Knowles, J. D.; Kell, D. B. *Anal. Chem.* **2005**, *77*, 290−303.

(35) Lee, M. H.; Han, C.; Chang, K. S. *Comput. Chem. Eng.* **1997**, *21*, S1037−S1042.

(36) Sangwai, J. S.; Bhat, S. A.; Saraf, D. N.; Gupta, S. K. *Chem. Eng. Sci.* **2007**, *62*, 2790−2802.

(37) Omata, K. *Ind. Eng. Chem. Res.* **2011**, *50*, 10948−10954.

(38) Schork, F. J.; Ray, W. H. *J. Appl. Polym. Sci.* **1987**, *34*, 1259−1276.

(39) Peremezhney, N.; Hines, E.; Lapkin, A.; Connaughton, C. *Eng. Optim.* **2014**, *46*, 1593−1607.

(40) Chern, C. S. *Prog. Polym. Sci.* **2006**, *31*, 443−486.

(41) Hua, H. Semi-continuous emulsion copolymerization of styrene - butyl acrylate with methacrylic acid: structure-property relationship. University of Ottawa, Thesis, 2009.

(42) Peremezhney, N.; Connaughton, C.; Unali, G.; Hines, E.; Lapkin, A. A. *Chem. Eng. Res. Des.* **2012**, *90*, 2179−2185.

(43) Peremezhney, N. Chemical product/process design and optimization: development of novel techniques and integration of bio-feedstocks, Univeristy of Warwick, Thesis, 2013.