Electronic Supplementary Material

Synergistic optimization framework for the process synthesis and design of biorefineries

Nikolaus I. Vollmer[1], Resul Al[2], Krist V. Gernaey[1], Gürkan Sin (✉)[1]

1 Process and Systems Engineering (PROSYS) Research Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

2 Novo Nordisk A/S, 2880 Bagsværd, Denmark

E-mail: gsi@kt.dtu.dk

Appendix

A.1 Biorefinery Unit Operations and High-Fidelity Models

As a holistic concept, a biorefinery setup can process a multitude of feedstocks into various products as fuels, chemicals, pharmaceuticals, foods, and energy; rather than the actual conversion processes, general biorefineries share the same objective of sustainable and efficient use of resources [1]. In particular, lignocellulosic biorefineries achieve this by utilizing agricultural and forestial residues or dedicated lignocellulosic crops as their raw material, making them a promising candidate for biorefinery setups due to the abundance and the low price of the feedstock [2,3]. Due to the properties of lignocellulosic biomass and the aim of developing biotechnological processes, there are several compulsory unit operations for biorefinery processes, namely a pretreatment unit, a fermentation unit, an evaporation unit, and a crystallization unit, for each of which a model is developed [4]. All mentioned units are widely used throughout industry and possess a high technology readiness level (TRL). There are further unit operations or alternatives to the four mentioned ones that have a sufficiently high TRL to be employable in a biorefinery setup; however, for the sake of the introduction of this framework, only the four mentioned units are described in detail and used in a case study for the validation of the framework. However, their use in the presented framework is not compulsory insofar as other models can equally be employed depending on the specific process design task, as the surrogate models for the framework are build independent of the underlying high-fidelity model. Furthermore, all model implementations are available through the S3O GitHub repository [5].

A.1.1 Pretreatment Model

Lignocellulosic biomass consists mainly of three fractions: the hemicellulosic fraction, the cellulosic fraction, and the lignin fraction. Classically, the conversion processes in lignocellulosic biorefineries use the monomers of the hemicellulosic and the cellulosic fraction as substrate; however, in the biomass, they are present in a polymeric form and cannot be converted directly [3]. Hence, lignocellulosic biorefineries employ a pretreatment process as the first unit operation to reduce the recalcitrance of the biomass and break down the polymeric structure of commonly one of the three fractions. It depends highly on the applied pretreatment

technology and the operating conditions, which of the fractions is primarily targeted. However, all methodologies have in common that their objective is to be very selective on the specific fraction, have a high specific monomer yield, and have a low inhibitory compound yield [3]. The number of compounds occurring in the pretreatment is high; thus, a model commonly only describes essential components [6].

A general mass balance for a component i in an open thermodynamic system is given with the following equation:

$$\frac{dM_i}{dt} = \dot{M}_{i,in} - \dot{M}_{i,out} \pm \sum_{j \in J_i} R_j. \quad (1)$$

The left side of the equation describes the change of mass of the component over time dM_i/dt with $\dot{M}_{i,in}$ being the mass flow of the component into the system, $\dot{M}_{i,out}$ being the mass flow of the compound out of the system and $\sum_{j \in J_i} R_j$ being a term that sums the set of conversion reactions J_i of the component i with R_j as the rate expression for reaction j. Conversion reactions are either production reactions that yield the respective compound as product or consumption reactions in which the component serves as substrate and is subsequently consumed.

The kinetic expression for the reaction rates used in (1) can be formulated in various ways [7,8]. For this model, we chose the description as a first-order rate expression. Exemplarily for a reaction j the rate expression R_j is formulated as follows:

$$R_j = A_j \cdot \exp\left(-\frac{E_{A,j}}{\tilde{R} \cdot T}\right) \cdot \prod_{i \in I_j} C_i^{n_i}. \quad (2)$$

Here, A_j describes the frequency factor, $E_{A,j}$ the activation energy of the reaction, $\tilde{R}$ the universal gas constant, T the reaction temperature and a term $\prod_{i \in I_j} C_i^{n_i}$ which multiplies over the set of substrate concentrations I_j for the reaction j with C_i as substrate concentration of component i and n_i as exponent describing the reaction order in which the component i participates in the reaction j. The frequency factors, the activation energies, and the reaction orders are usually estimated from experimental data to fit the model to the respective pretreatment method.

Analogously to the mass balance, an energy balance for an open thermodynamic system, in this case, the whole unit operation constitutes as follows in the form of an enthalpy balance:

$$\frac{dH}{dt} = \sum \dot{H}_{in} - \sum \dot{H}_{out} + \sum \dot{Q}_{in} - \sum \dot{Q}_{out} + \sum \dot{H}_{source} - \sum \dot{H}_{sink}. \quad (3)$$

With dH/dt being the change of enthalpy in the system, $\dot{H}_{in}$ representing all ingoing enthalpy flows due to mass flow into the system, whereas all outgoing enthalpy flows due to mass flow out of the system are described with $\dot{H}_{out}$. Ingoing and outgoing heat flows are described with $\dot{Q}_{in}$ and $\dot{Q}_{out}$ respectively, as well as possible heat sources and sinks with $\dot{H}_{source}$ and $\dot{H}_{sink}$ describing e.g. the release of enthalpy of reaction. Equation (3) in its presented form is generic, and represented terms can be omitted depending on the respective model.

A.1.2 Fermentation Model

The fermentation unit operation serves as the central reaction unit for converting the monomers from the pretreatment into the desired products. The reaction is performed by a cell factory, commonly bacteria or yeast strains, which either produce the desired product from the feedstock naturally or artificially after performing genetic and metabolic modifications on the cell factory a priori. Despite the high complexity of the metabolic network and transcriptional regulations defining the kinetic behavior, the entire cell factory can be considered as black-box, lumping the whole metabolic network into three kinetic equations in the simplest case [9]: the substrate uptake rate, a Herbert-Pirt distribution relation, and a production rate.

The substrate uptake rate q_s is constituted as follows:

$$q_s = q_{s,max} \cdot \frac{C_s}{K_s + C_s}, \tag{4}$$

where $q_{s,max}$ relates to the maximum substrate uptake rate under excess conditions, C_s being the concentration of substrate and K_s being the substrate affinity constant.

The Herbert-Pirt distribution relation describes the distribution of the substrate in the cell towards biomass growth, production of the desired component, and cell maintenance and is described in the following way:

$$q_s = a_s \cdot \mu + b_s \cdot q_p + m_s. \tag{5}$$

The biomass growth is denoted by the growth rate μ, the product formation as q_p and cell maintenance by m_s. The parameters a_s and b_s as well as the maintenance constant are specific for the employed cell factory and have to be determined experimentally.

The last kinetic equation describes the product formation for catabolic products as a function of the biomass growth for catabolic products:

$$q_p = a_p \cdot \mu + b_p. \tag{6}$$

Analogously to equation (5), parameters a_p and b_p are specific for the employed cell factory and have to be determined experimentally. Most of the mentioned products (biofuels, biochemicals, and others) are usually catabolic products; nonetheless, there are several formulations of product formation equations for anabolic products, which can be used instead of equation (6). For these, the reader is referred to additional literature [9].

In a similar fashion to the pretreatment model, also for the fermentation model, the mass and energy balances are described with equations (1) and (3) respectively. For the case of a black-box model, exactly three mass balance equations are needed for substrate, cell biomass, and product. The reaction term, which has to be considered in the substrate's mass balance (1), is the substrate uptake rate from equation (4). The reaction term for the cell biomass derives from equation (5), and the reaction term equals equation (6). All rate equations are dependent on biomass growth, so they have to be multiplied with the cell biomass concentration C_X:

$$R_j = q_j \cdot C_X \tag{7}$$

A.1.3 Evaporation Model

Due to commonly low titers in aqueous fermentation processes, unit operations for upconcentration are employed as part of the downstream processing to obtain the desired product in a sufficiently high concentration. This mainly happens with evaporation or membrane separation units, despite their high operational expenditures [10,11]. Furthermore, the evaporation unit operation mainly also serves to remove undesired compounds through the vapor phase. For the evaporation model, the mass and energy balances described with equations (1) and (3) apply. The thermodynamic equilibrium between the vapor and the liquid phase is described with vapor pressure equations for the components. Common process simulation software includes implemented evaporation models; therefore, for this paper's scope, instead of developing an own evaporation model, the one implemented in the Aspen Plus software package is utilized.

A.1.4 Crystallization Model

Crystallization as a unit operation is commonly employed for recovering solid products out of a liquid phase. Especially for solid products, the isolation and purification are classically done with a crystallization unit as the product size and shape are good to control [12]. In the crystallization process, the thermodynamic driving force is an excess concentration of the product above the respective saturation concentration under the given operational conditions. The process is controlled either by the temperature, which relates to either evaporation or cooling crystallization. Alternatively, an anti-solvent crystallization utilizes a decrease in solubility by adding a second solvent with low solubility for the compound [12]. Like all the other unit operations, this unit operation's mass and energy balances can be described with equations (1) and (3). The kinetics of the formation of crystals can be described with a population balance equation as the following:

$$\frac{\partial n}{\partial t} + \frac{\partial (G \cdot n)}{\partial L} = B - D, \tag{8}$$

where n describes the particle number density, G the crystal growth rate, B the crystal birth rate, D the crystal decay rate, and L the crystal length. Commonly, all rates and the underlying solubility curves are estimated from experimental data in order to quantify the driving force and model the process. There exist several ways of solving a population balance, e.g., by the method of moments or the method of classes, to yield a kinetic expression that describes the mass transfer between the liquid and solid phase of the solute described with equation (1).

A.2 Uncertainty and Sensitivity Analysis

All of the models introduced in section A.1 have in common that they are based on assumptions to a certain extend. Moreover, all of them rely on experimental data for their calibration to the system they are supposed to predict. Both present a significant source of uncertainty for the model prediction, especially when coupling models in a flowsheet simulation. These errors can propagate and accumulate to an extent that deteriorates the models' predictive quality. Hence, it is crucial to assess these models' robustness by analyzing how uncertainties in the model input propagate to the model output and how the uncertainty in the model output can be apportioned to the model input [13]. The former analysis is called uncertainty analysis, whereas the latter being

complementary to the prior is called sensitivity analysis. The purpose of performing uncertainty analysis is to assess the model's robustness and quantify the error propagation due to the estimated model parameters or uncertain input or design variables. The purpose of performing a sensitivity analysis is more diverse: It reaches from a general design space exploration with the underlying model to identifying sensitive input variables or parameters to select a subset of variables for optimization problems and identify crucial parameters to estimate model parameters other tasks.

A.2.1 Monte Carlo-Based Uncertainty Analysis

In this framework, the assessment is performed with a Monte Carlo-based method [14]. Therefore, its theory is described briefly:

Monte Carlo methods utilize probability statistics and random numbers [15]. In the case of the uncertainty analysis, the model input space is sampled randomly with a sufficiently high number of random samples [16]. For each of the samples, a simulation is performed in order to calculate the model output. Based on these model outputs, the output uncertainty of the model can be quantified. In practice, this is performed in the following way:

Firstly, ranges and uncertainty distributions are defined for all model parameters m that are considered uncertain. Secondly, a sufficiently high sample number N is chosen, and Latin Hypercube Sampling (LHS) is performed for all the uncertain inputs. The result is a matrix $X_{N\times m}$ over the input space. For each sample n, a model simulation is performed to create the respective output space matrix $Y_{N\times k}$ for all considered outputs k. Based on the output calculations, statistical performance parameters as mean value, standard deviations, and percentiles can be calculated.

A.2.2 Variance-Based Sensitivity Analysis

The complementary analysis to the uncertainty analysis is a sensitivity analysis, aiming at apportioning the output uncertainty on the different inputs [17]. The here applied methodology is based on a variance-based sensitivity measure. This method calculates the model's total variance based on Monte Carlo simulations and is usually referred to as Sobol's sensitivity method [18].

Saltelli et al. (2010) describe the methodology as follows: For independent inputs of a model $y = f(\theta_i)$, the variance of the model output $V(y)$ can be partitioned in the following way:

$$V(y) = \sum_{i}^{k} V_i + \sum_{i}\sum_{k} V_{ij} + \ldots + V_{123\ldots k}, \tag{9}$$

with the variance V_i as:

$$V_i = \int f(\theta_i)^2 \, d\theta_i. \tag{10}$$

Each term serves as a measure of sensitivity of the respective input(s) in equation (9). Applying the law of total variance:

$$\mathrm{V}(y) = V\big(E(y|\theta_i)\big) + E\big(V(y|\theta_i)\big), \tag{11}$$

for each term in (9) yields measures for the first-order sensitivity index S_i and the total sensitivity index S_{Ti}:

$$S_i = V\big(E(y|\theta_i)\big)/\mathrm{V}(y), \tag{12}$$

$$S_{Ti} = E\big(V(y|\theta_i)\big)/\mathrm{V}(y). \tag{13}$$

The numerical calculation of both sensitivity indices is performed by a Monte Carlo-based procedure: As a first step, Sobol sampling is performed, and two sampling matrices A and B are generated. From those, two mixed matrices A_B^i and B_A^i are generated, where column i from the one matrix is replaced by the same column of the respective other matrix, and all other columns are kept. For all four sampling matrices, the model outputs are calculated. The respective sensitivity measures as in (2.12) and (2.13) can be calculated with those. Applying the method for the first-order sensitivities, these are the following:

$$S_i = V(y) - \frac{1}{2N}\sum_{j=1}^{N}\big(y_B(j) - y_{ABi}(j)\big)^2, \tag{14}$$

and the total sensitivities are calculated as follows:

$$S_i = \frac{1}{2N}\sum_{j=1}^{N}(y_A(j) - y_{ABi}(j))^{\,2}. \tag{15}$$

The interpretation of both is also twofold: The first order sensitivity explains this parameter's single effect, indicating the expected reduction in the output variance if this particular parameter could be fixed. Hence, the total sensitivity is the expected variance if all the parameters except the respective one could be fixed [19].

1. Cherubini F. The biorefinery concept: using biomass instead of oil for producing energy and chemicals. Energy Conversion and Management, 2010, 51(7): 1412–1421
2. Hassan S S, Williams G A, Jaiswal A K. Lignocellulosic biorefineries in Europe: current state and prospects. Trends in Biotechnology, 2019, 37(3): 231–234
3. Galbe M, Wallberg O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials. Biotechnology for Biofuels, 2019, 12(1): 1–26
4. Chaturvedi T, Torres A I, Stephanopoulos G, Thomsen M H, Schmidt J E. Developing process designs for biorefineries-definitions, categories, and unit operations. Energies, 2020, 13(6): 1493
5. S3O GitHub Repository. 2021, 10.5281/zenodo.5017353
6. Rasmussen H, Sørensen H R, Meyer A S. Formation of degradation compounds from lignocellulosic biomass in the biorefinery: sugar reaction mechanisms. Carbohydrate Research, 2014, 385: 45–57
7. Prunescu R M, Blanke M, Jakobsen J G, Sin G. Dynamic modeling and validation of a biomass hydrothermal pretreatment process—a demonstration scale study. AIChE Journal,

2015, 61(12): 4235–4250
8. Shen J, Wyman C E. A novel mechanism and kinetic model to explain enhanced xylose yields from dilute sulfuric acid compared to hydrothermal pretreatment of corn stover. Bioresource Technology, 2011, 102(19): 9111–9120
9. Smolke C D, ed. The Metabolic Pathway Engineering Handbook: Fundamentals. Boca Raton: CRC Press, 2009
10. Kolfschoten R C, Bruins M E, Sanders J P M. Opportunities for small-scale biorefinery for production of sugar and ethanol in the Netherlands. Biofuels, Bioproducts and Biorefining, 2014, 8(4): 475–486
11. Huang H J, Ramaswamy S, Tschirner U W, Ramarao B V. A review of separation technologies in current and future biorefineries. Separation and Purification Technology, 2008, 62(1): 1–21
12. Kirwan D J, Orella C J. Crystallization in the pharmaceutical and bioprocessing industries. In: Handbook of Industrial Crystallization. Elsevier, 2002, 249–266
13. Sin G, Gernaey K V., Lantz A E. Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. Biotechnology Progress, 2009, 25(4): 1043–1053
14. Coleman H W, Steele W G. Experimentation, Validation, and Uncertainty Analysis for Engineers. 3rd ed. New Jersey, USA: John Wiley & Sons, Inc., 2009
15. Hammersley J M, Handscomb D C. Monte Carlo Methods. Springer Netherlands, 1964
16. Helton J C. Treatment of uncertainty in performance assessments for complex systems. Risk Analysis, 1994, 14(4): 483–511
17. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Cariboni J, Gatelli D，Saisana M, Tarantola S. Global Sensitivity Analysis. The Primer. Chichester, UK: John Wiley & Sons, Ltd., 2008
18. Sobol I M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 2001, 55(1-3): 271–280
19. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications, 2010, 181(2): 259–270