

Sparse Regression Approach to Modelling the Effect of Ionic Liquid Acidity in Biomass Fractionation^{*}

Suhaib Nisar^{*,**} Sarah Seidner^{**} Agnieszka Brandt-Talbot^{***}
Jason P. Hallett^{**} Benoît Chachuat^{*,**}

^{*} *The Sargent Centre for Process Systems Engineering, Imperial
College London, UK*

^{**} *Department of Chemical Engineering, Imperial College London, UK*

^{***} *Department of Chemistry, Molecular Sciences Research Hub,
Imperial College London, UK*

Abstract: Fractionation of lignocellulosic biomass is a crucial step to provide cellulose, lignin, and hemicellulose for further processing. This paper is concerned with modelling biomass fractionation using the IonoSolv process, which employs low-cost ionic liquid water mixtures, with a special focus on describing the effect of acid:base ratio of the mixture on process performance. We build on an existing semi-mechanistic modelling framework describing the solvent extraction of three main biopolymers from woody biomass for varying fractionation temperature, time, and solids loading. Since the effect of acidity is poorly understood from a mechanistic standpoint, we use sparse regression with lasso regularisation to incorporate it in the semi-mechanistic model. We investigate both polynomial and exponential functional forms and find that the latter yields more physically-consistent results. This enabled us to recalibrate the parameters of the combined semi-mechanistic and sparse data-driven models simultaneously to accurately predict the effect of varying acid:base ratio. This hybrid modelling framework opens new opportunities for further analysis and optimisation of ionic liquid-based biomass fractionation processes.

Keywords: biomass, ionic liquids, acidity, semi-mechanistic modelling, sparse regression

1. INTRODUCTION

Climate change and resource scarcity are driving the demand for sustainable alternatives to non-renewable energy sources and carbon materials such as plastics. Biomass, which refers to any renewable organic material originating from plants and animals, can be a source of both energy and materials. Lignocellulosic biomass comprises woody plant parts, including agricultural and food industry residues (Ragauskas et al., 2006; Verdía Barabá et al., 2023). It is composed of three main biopolymers: cellulose, hemicellulose, and lignin. The chemistry and spatial arrangement of the components of lignocellulosic biomass hinders its direct conversion to specific products, thus requiring a chemical transformation to maximise the availability of the biopolymers for specific applications.

The effectiveness of ionic liquids (ILs) as solvents for biomass fractionation is well established. In particular, the IonoSolv process utilises protic ionic liquid–water solutions, synthesised by the simple combination of an aqueous Brønsted acid with a Brønsted base, to fractionate

lignocellulosic biomass into a cellulose-rich pulp and a lignin- and hemicellulose-rich liquid. Solid lignin is then separated after increasing the water content of the ionic liquid solution (Brandt-Talbot et al., 2017). The cellulose pulp is washed after the pretreatment to remove the IL and residual lignin, followed by further processing to cellulosic biofuels (after hydrolysis) or biomaterials. The dissolved lignin, recovered from the IL by the addition of an anti-solvent such as water, can either be combusted for process heat or converted to other products. The IL–water mixture can be regenerated via distillation before its recycling.

IonoSolv process performance is primarily impacted by the choice of IL, biomass feedstock, fractionation conditions, process scale, along with IL-specific properties such as acidity (Brandt-Talbot et al., 2017; Weigand et al., 2017) and water content (Abouelela et al., 2023). Previous modelling efforts have mostly considered ILs with a fixed acid:base ratio and water content. The complex interactions of the ionic species in protic ILs with water complicate understanding the interplay between different lignocellulosic biomass components and the solution (Firth et al., 2024). However, the acid:base ratio has a notable impact on IonoSolv process performance. Previously in hardwoods, a mere 2% excess of acid caused severe cellulose degradation with a corresponding increase in the rate of pseudo-lignin re-deposition (Weigand et al., 2017). Such pseudo-lignin—a combination of sugar-derived humins and

^{*} Suhaib Nisar is grateful to the Department of Chemical Engineering at Imperial College London for a PhD scholarship and to the EPSRC Centre for Doctoral Training in Next Generation Synthesis & Reaction Technology for the PhD studentship under grant EP/S023232/1. Correspondence should be directed to Benoît Chachuat: b.chachuat@imperial.ac.uk

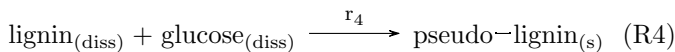
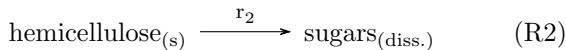
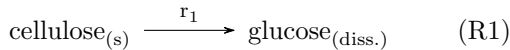
re-condensed lignin—reduces the purity of the cellulose-rich pulp and hinders the saccharification of cellulose pulps to form biofuels, limiting further process pathways. To model the effect of acidity on ionoSolv fractionation, Abouelela et al. (2023) recently used a severity factor approach with Hammett acidity—an extension of the pH scale—to correlate acidity with ionoSolv performance for varying water contents, yet for a fixed acid:base ratio.

Previously, we developed a semi-mechanistic model to describe ionoSolv fractionation and demonstrated it for the grassy biomass *Miscanthus* with the IL triethylammonium hydrogen sulfate [TEA][HSO₄] at varying temperature, time, and solids loading (Nisar et al., 2024). However, this model assumed a fixed acid:base ratio and water content. Herein, we build on this framework and incorporate sparse data-driven elements to describe the effect of acidity on biomass fractionation with protic ILs. We use sparse regression with lasso regularisation to isolate the model terms necessary to accurately predict the effect of acid:base ratio. In the rest of the paper, the model structure and sparse regression methodology will first be outlined (Sec. 2), followed by the presentation and discussion of the model training and predictions for varying acid:base ratio (Sec. 3).

2. BACKGROUND AND METHODOLOGY

2.1 Semi-Mechanistic Model Structure

The simple reaction network R1–R4 was proposed by Nisar et al. (2024) to describe the main reactions occurring during lignocellulosic biomass fractionation.



Mass balances (Eqn. 1) were enforced in terms of the mass fraction m_j of each species j relative to the initial total biomass solids, with $\nu_{i,j}$ the stoichiometric coefficient of species j in reaction i .

$$\frac{dm_j}{dt} = \sum_{i=1}^4 \nu_{i,j} r_i(T, S, \mathbf{m}) \quad (1)$$

The reaction kinetics were expressed as functions of the temperature T (K), the solids loading S (%) representing the solids-to-solvent mass ratio, and the mass fractions of reactants, with \mathcal{R}_i the set of reactants in reaction i . An Arrhenius power kinetic expression was employed for reactions R2–R4 (Eqn. 3), while a Haldane kinetic expression was used for reaction R1 (Eqn. 2). The parameters θ_i^0 and θ_i^1 model the effect of temperature, with reference temperature $T_{\text{ref}} = 298.15$ K; the parameters α_i and $\beta_{i,j}$, the solids loading and reaction order effects; and the parameters κ_1 and κ'_1 , the saturation and inhibition regimes at intermediate and high cellulose concentration $S \times m_{\text{cell}}$, respectively.

$$r_1 = \exp\left(\theta_i^0 + \theta_i^1 \left(1 - \frac{T_{\text{ref}}}{T}\right)\right) \frac{S m_{\text{cell}}}{\kappa_1 + S m_{\text{cell}} + \frac{(S m_{\text{cell}})^2}{\kappa'_1}} \quad (2)$$

$$r_i = \exp\left(\theta_i^0 + \theta_i^1 \left(1 - \frac{T_{\text{ref}}}{T}\right)\right) S^{\alpha_i} \prod_{j \in \mathcal{R}_i} m_j^{\beta_{i,j}}, \quad i = 2 \dots 4 \quad (3)$$

The kinetic parameters in the model were calibrated based on pulp compositions consisting of residual cellulose, hemicellulose and lignin measurements. Since the wet-lab compositional analysis protocol (Sluiter et al., 2010) measures the lignin content gravimetrically, it cannot distinguish native lignin and pseudo-lignin—encompassing condensed (re-precipitated) lignin and sugar-derived humins. Therefore, when calculating residuals for lignin, the components lignin and pseudo-lignin were combined. The wet-lab protocol is also unable to distinguish between glucose derived from cellulose or hemicellulose, but as most glucose originates from cellulose, the term glucan is understood as cellulose hereon.

2.2 Sparse Regression of Acid:Base Correction Terms

The dynamic model (1)–(3) describes the combined effect of temperature, time and solids loading for a fixed acid:base ratio around $1 \pm 0.5\%$.¹ Further investigations have revealed that, without acidity correction, even a small (c. 2–3%) addition of either excess acid or base could induce significant model prediction mismatch, echoing the conclusions from other studies (Weigand et al., 2017). Therefore, extending the model (1)–(3) to describe the effect of acid:base ratio is paramount to improving its robustness and predictive capability. Previous attempts at modelling the effect of acidity in ionoSolv fractionation involved modifying empirical expressions developed for hydrothermal systems (Abouelela et al., 2023), yet with limited success. Instead, the approach adopted herein seeks to identify multiplicative correction factors for the reaction rates r_i (Eqns. 2 & 3) when the acid:base ratio ($A:B$) deviates from unity.

Polynomial Correction terms. Our initial attempt considered reaction rate corrections as $r_i(1 + \varepsilon_i)$, where the polynomial functions ε_i vanish when $A:B = 1$ (Eqn. 4), and so the corrected reaction rates equal those from the original model (1)–(3). Following the report of a marked increase in the effect of acidity with temperature by Weigand et al. (2017), we accounted for the combined effect of $A:B$ and temperature T (°C). Specifically, the polynomials—which are compactly written using multi-index notation in Eqn. (4)—operate on normalised acid:base and temperature variables $(\overline{A:B}, \overline{T})$; Eqn. 5). With third-order correction terms, the polynomials comprise 10 monomials and the regression problems seeks to estimate the corresponding 10 coefficients $\phi_{i,\kappa}$.

¹ The acid:base ratio of hydrogen sulfate-based ionic liquids such as [TEA][HSO₄] is determined through titration, which can also introduce uncertainty into the measurements.

$$\varepsilon_i(A:B, T) = (A:B - 1) \sum_{\substack{\boldsymbol{\kappa} \in \mathbb{N}^2 \\ \kappa_1 + \kappa_2 \leq 3}} \phi_{i, \boldsymbol{\kappa}} \overline{A:B}^{\kappa_1} \overline{T}^{\kappa_2} \quad (4)$$

$$\text{with: } \overline{A:B} = \frac{A:B - 0.95}{0.1}, \quad \overline{T} = \frac{T - 25}{170 - 25}. \quad (5)$$

Exponential Correction Terms. The functional expressions used to model the effect of severity on hydrothermal and dilute acid systems often employ exponential dependencies (Malaret et al., 2020). Our second attempt at correcting the reaction rates, therefore, was in the form of $r_i \exp(\varepsilon_i)$, with the same polynomial functions—hence the same number of coefficients $\phi_{i, \boldsymbol{\kappa}}$ —as in Eqn. (4). Here again, the corrected reaction rates equal those from the original model (1)–(3) when $A:B = 1$.

Regression Problem Formulation. A general formulation of the regression problem is summarised in Eqns. (6)–(7), for a collection of N biomass fraction measurements $\hat{m}_{j_k}^{(k)}$, with $j_k \in \{\text{cellulose, hemicellulose, lignin}\}$, under given conditions $(\hat{T}^{(k)}, \hat{t}^{(k)}, \hat{S}^{(k)}, \hat{A:B}^{(k)})$ of temperature, pretreatment time, solids loading and acidity.

$$\min_{\phi_{i, \boldsymbol{\kappa}}, \delta_{i, \boldsymbol{\kappa}}} \left\{ \begin{array}{l} \sum_{k=1}^N \left[\hat{m}_{j_k}^{(k)} - m_{j_k} \left(\hat{T}^{(k)}, \hat{t}^{(k)}, \hat{S}^{(k)}, \hat{A:B}^{(k)} \right) \right]^2 \\ + \lambda \sum_{i=1}^4 \sum_{\substack{\boldsymbol{\kappa} \in \mathbb{N}^2 \\ \kappa_1 + \kappa_2 \leq 3}} \delta_{i, \boldsymbol{\kappa}} \end{array} \right\} \quad (6)$$

$$\text{s.t. } -\delta_{i, \boldsymbol{\kappa}} \leq \phi_{i, \boldsymbol{\kappa}} \leq \delta_{i, \boldsymbol{\kappa}}, \quad \forall \boldsymbol{\kappa} \in \mathbb{N}^2 : \kappa_1 + \kappa_2 \leq 3, \\ i = 1 \dots 4 \quad (7)$$

The first term in the cost function (6) is a standard (un-weighted) least-squares estimator to minimise the residuals between the experimental data and model predictions. The second term is a lasso (ℓ_1) regularisation, for driving the coefficients $\phi_{i, \boldsymbol{\kappa}}$ to zero, and thereby promote sparsity. Increasing the regularisation weight $\lambda \geq 0$ decreases the risk of over-fitting since fewer monomials comprise the polynomial function; but it also increases the residual error and, therefore, the risk of under-fitting. By and large, selecting an appropriate regularisation weight λ is essential to training a model with good generalisation capability, and so we determined its value by means of a simple trial-and-error approach. Note also that the auxiliary variables $\delta_{i, \boldsymbol{\kappa}}$ were introduced to reformulate the absolute values in ℓ_1 regularisation as extra constraints (7) and make the regression problem differentiable.

Computational Procedure. We proceeded with the solution of three regression problems. In the first two problems, the kinetic model parameters $(\theta_i^0, \theta_i^1, \alpha_i, \beta_{i, j}, \kappa_1, \kappa_1')$ in (1)–(3) were kept fixed to those estimated at $A:B = 1$ (Nisar et al., 2024). In all three regression problems, the separable structure of the reaction network R1–R4 allowed us to solve three decoupled regression sub-problems, for the cellulose reaction R1, the hemicellulose reaction R2, and the two lignin reactions R3–R4. The latter sub-problem used the optimal parameters of reaction R1 since reaction R4 involves glucose, whose production rate is decoupled from its subsequent reaction rate.

In the first two problems, we considered polynomial and exponential corrections as $r_i(1 + \varepsilon_i)$ and $r_i \exp(\varepsilon_i)$, respectively, and fitted all 10 coefficients $\phi_{i, \boldsymbol{\kappa}}$ in each reaction $i = 1 \dots 4$. We increased the regularisation weight $\lambda \in [10^{-5}, 1]$ in the lasso regularisation to promote sparsity, while monitoring the fit to the available experimental data and the number of non-zero coefficients in the correction terms. This allowed us to select the minimal number of monomials in the exponential corrections of all four reactions R1–R4. The better fitting approach between the polynomial and exponential corrections was then employed in the final regression problem.

In the final problem, both the kinetic parameters in Eqns. (2)–(3) and the coefficients in the acidity correction terms using the sparse polynomials in Table 1 were re-estimated simultaneously. The kinetic parameter values were allowed to vary within 10% of their original values to compensate for slight deviations around $A:B = 1$ due to experimental errors. The three sub-problems for cellulose, hemicellulose and lignin were again addressed separately.

In our Python implementation, the constrained regression problems were solved using the code SLSQP interfaced from SciPy (Virtanen et al., 2020). Since a finite difference-based approach was unable to provide reliable and reproducible solutions of the sparse regression problems, we used the library CRONOS (v4.0)² for numerical integration and sensitivity analysis of ODE systems. CRONOS provides an interface with the solver CVODES of the library SUNDIALS (v7.1.1)³ and relies on the library MC++ (v4.0)⁴ to evaluate and differentiate expression trees.

3. RESULTS AND DISCUSSION

The system considered was the grass *Miscanthus* with the protic ionic liquid [TEA][HSO₄] at varying temperature, pretreatment time, solids loading, and acid:base ratio. The data used included both previously published (Brandt-Talbot et al., 2017) and newly collected experimental data.

3.1 Sparse Correction Results

Figure 1 illustrates the variation in the prediction accuracy and number of non-zero terms in the polynomial and exponential correction functions of each reaction with increasing lasso weight. There is an increase in model sparsity with increasing lasso weight, with parameter values gradually being driven to zero. The black dotted lines in each subplot mark the chosen lasso weights for each approach, yielding the optimal balance between model accuracy and sparsity. Recall that every correction function ε_i has a dependence on the acid:base ratio through the $(A:B - 1)$ multiplier in Eqn. (4). For the glucan (i) and lignin (iii) models, the exponential correction slightly outperforms the polynomial approach with generally a lower number of parameters, while the exponential approach clearly performs better for hemicellulose (ii). The sparse correction trajectories (black dotted line) were then visualised over the experimental data range to further compare both approaches.

² <https://github.com/omega-icl/cronos>

³ <https://github.com/LLNL/sundials/tree/main>

⁴ <https://github.com/omega-icl/mcpp>

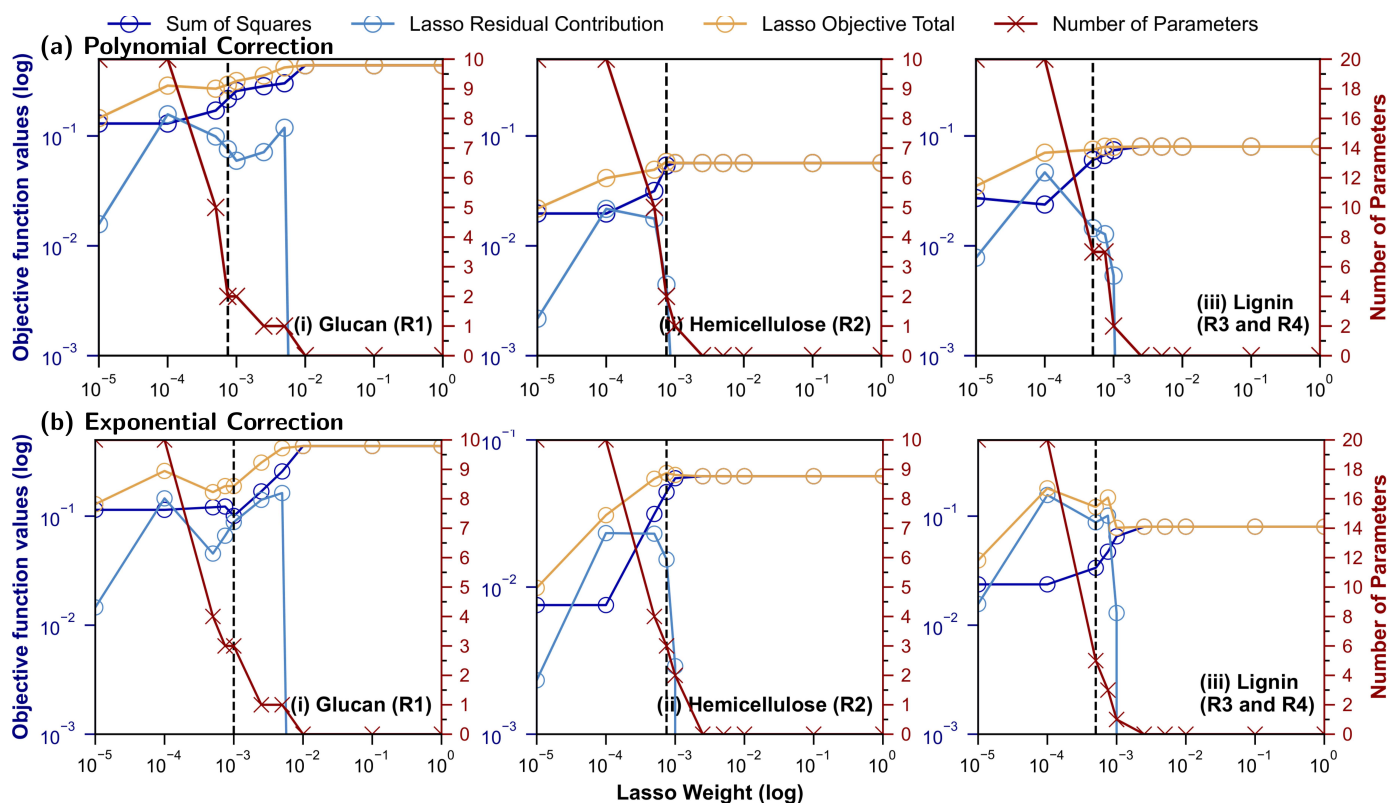


Fig. 1. Change in model accuracy and sparsity for (a) polynomial and (b) exponential functions with increasing lasso regularisation.

3.2 Sparse Correction Trajectories

Figure 2(a) illustrates the fitted sparse correction functions $\epsilon_i(A:B, T)$ for the four reactions R1–R4 for the polynomial approach. The function values broadly increase with both acid:base ratio and temperature—albeit less noticeably with temperature—but the overall correction value does become negative when extrapolated to lower acid:base ratios for hemicellulose (R2). The muted temperature effects likely originate from the Arrhenius temperature dependence in the kinetic expressions, which then only requires an interaction between temperature and acid:base ratio under extremely severe fractionation conditions at higher acid:base ratios. While data-driven polynomial models are relatively simple to implement, they can predict non-physical behaviours, such as the increase in cellulose hydrolysis rate at acid:base ratios below 1 or negative reaction rates. Thus, while the use of polynomials provides a useful guide to the underlying dynamics, they do not sufficiently elucidate the more nuanced features in the system and do not yield physically consistent models.

Figure 2(b) illustrates the fitted sparse correction functions $\epsilon_i(A:B, T)$ for the four reactions R1–R4, this time embedded within an exponential function. As in Figure 2(a), the correction values increase with acid:base ratio, suggesting a catalytic increase in the reaction rate with acid concentration. For the sugar reactions (R1–R2), the expressions capture the interaction between temperature and acid:base ratio under extremely severe fractionation conditions at higher acid:base ratios. These sparse corrections are also always non-negative due to the exponential function, avoiding non-physical situations of negative re-

action rates. These trends are broadly in line with prior experimental knowledge. While cellulose and hemicellulose hydrolysis are definitely catalysed under acidic conditions (Carvalho and Colodette, 2017), lignin hydrolysis still occurs under alkaline conditions in processes such as Kraft pulping (Taherzadeh and Karimi, 2008). When the acid:base ratio becomes low enough for the overall system to become basic, therefore, the lignin hydrolysis rate may again start to increase due to a change in the hydrolysis mechanism. However, it should be noted that the overall system remains acidic at acid:base ratios slightly less than 1 here. Since each ammonium cation can react with a single proton, while each sulphuric acid anion contributes two protons, there is still an excess of protons in the system for $A:B < 1$. Further investigations are thus needed in regions of lower acid:base ratio to clarify whether lower acid concentrations will contribute to a change in mechanism of lignin hydrolysis.

Table 1 summarises the non-zero monomials at the optimal lasso weights shown previously. Both glucan (R1) and hemicellulose (R2) hydrolysis correction functions include temperature dependences to describe the interaction with acid:base ratio at high severities. In contrast, the lignin hydrolysis (R3) correction function has a single cubic relationship in acid:base ratio but no temperature dependence. The joint estimation of the lignin hydrolysis and pseudo-lignin re-deposition complicated the dynamic understanding of each individual reaction. The pseudo-lignin re-deposition (R4) reaction required the largest number of parameters, reflecting the limited mechanistic understanding of this complex reaction. The temperature dependence encodes the promoted formation of pseudo-

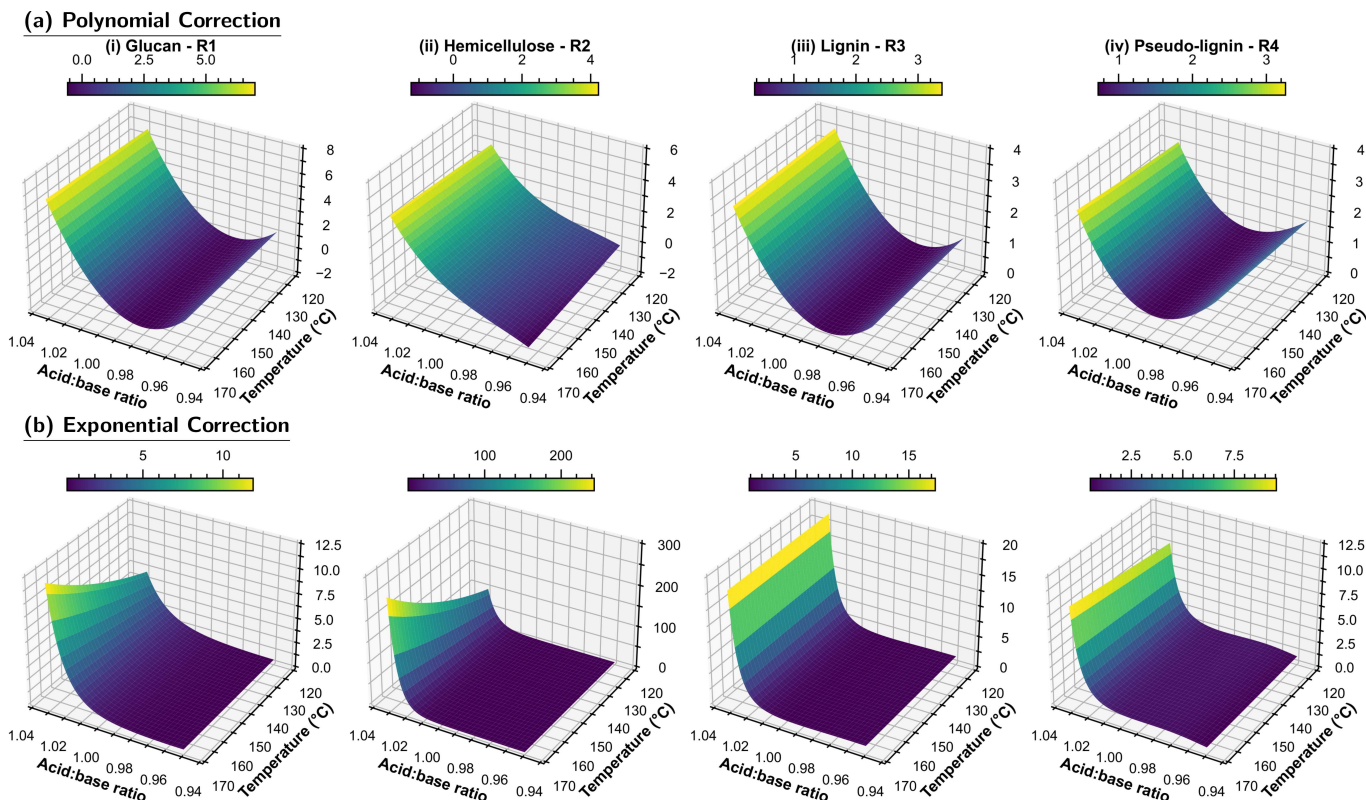


Fig. 2. (a) Polynomial and (b) Exponential sparse correction trajectories for reactions R1–R4.

lignin under the severe conditions of very high temperature and acid:base ratio.

Table 1. Selected sparse correction functions from lasso regression.

Reaction	Sparse correction
R1	$\phi_{1,[0,0]} + \phi_{1,[1,0]} \overline{A:B} + \phi_{1,[1,1]} \overline{A:B} \overline{T}$
R2	$\phi_{2,[0,0]} + \phi_{2,[1,0]} \overline{A:B} + \phi_{2,[0,1]} \overline{T}$
R3	$\phi_{3,[3,0]} \overline{A:B}^3$
R4	$\phi_{4,[1,0]} \overline{A:B} + \phi_{4,[0,1]} \overline{T} + \phi_{4,[2,0]} \overline{A:B}^2 + \phi_{4,[3,0]} \overline{A:B}^3$

3.3 Full Hybrid Model Recalibration

Figure 3 shows the performance of the hybrid model after its kinetic parameters and acidity correction coefficients were jointly recalibrated, here for two different temperatures (150 and 170 °C) and three acid:base ratios (around 0.98, 1, and 1.02). For glucan and hemicellulose, the model predictions were overall in good agreement with the experimental data at all temperatures and acid:base ratios. The rate of increase with acidity is more severe at higher temperatures, confirming the need of temperature in the correction polynomial expressions. Interestingly, a 2% excess of acid at 150 °C leads to a similar drop in glucan and hemicellulose content to that observed at 170 °C with an acid:base ratio of 1. This highlights the reactivity improvement from the addition of acid in this system, as well as the need to accurately predict its effect to create models that are robust to process disturbances.

For lignin, the situation is complicated by the need for both reactions R3–R4 to describe the entire behaviour. At

both temperatures, the model predictions were in good agreement with the experimental data for an acid:base ratio greater than 1, as it was at 170 °C for an acid:base ratio around 1. However, there was a slight under-estimation of lignin content at the other conditions shown, especially when the lignin content approached its minimum. As glucan hydrolysis slows down under less acidic conditions, there is less dissolved glucose available to react to form pseudo-lignin. Therefore, the model under-estimates the pseudo-lignin formation rate, and therefore the lignin content itself. This highlights a possible limitation of using these simple models for lignin, which has poorly understood dissolution and re-deposition mechanisms. In agreement with previous conclusions, while this approach provides useful intuitions into the complex underlying system, further experiments are required to understand the lignin and pseudo-lignin reaction mechanisms under less acidic conditions.

4. CONCLUSIONS

An existing semi-mechanistic model describing the iono-Solv fractionation of woody biomass for varying temperature, pretreatment time and solids loading was extended to encompass the effect of acidity, a critical but poorly-understood factor in ionic liquid-based systems. A sparse regression framework using lasso regularisation was first deployed, comparing both polynomial and exponential correction functions, with the results favouring the latter. These sparse correction functions were then re-trained with the original semi-mechanistic kinetic model. While the hybrid model showed good predictive capability for glucan and hemicellulose, it slightly under-estimated the

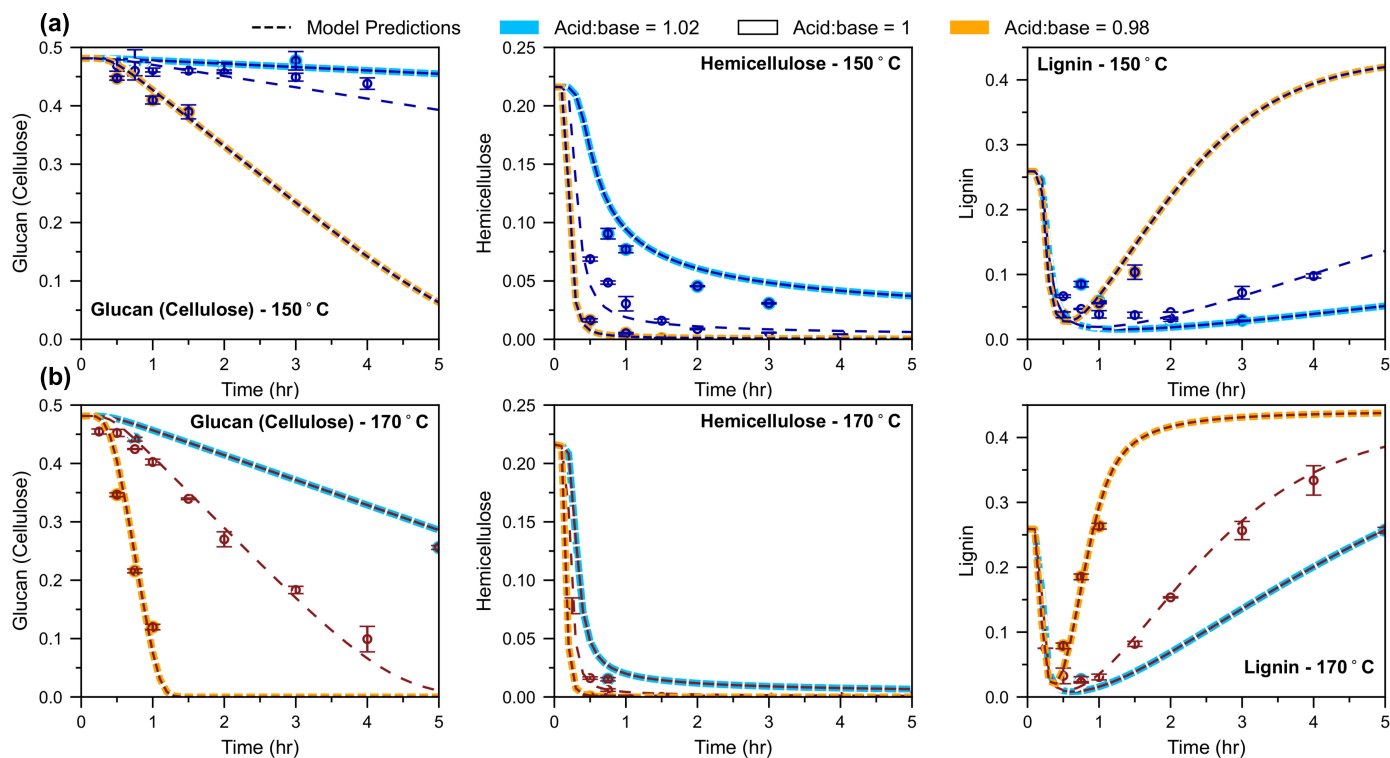


Fig. 3. Model predictions under different acid:base ratio conditions at (a) 150 °C and (b) 170 °C.

lignin content at lower acid:base ratios. The complex nature of lignin dissolution and re-deposition mechanisms warrants further study into the mechanisms of these processes under lower acidity conditions.

REFERENCES

- Abouelela, A.R., Nakasu, P.Y.S., and Hallett, J.P. (2023). Influence of Pretreatment Severity Factor and Hammett Acidity on Softwood Fractionation by an Acidic Protic Ionic Liquid. *ACS Sustainable Chemistry & Engineering*, 11(6), 2404–2415.
- Brandt-Talbot, A., Gschwend, F.J., Fennell, P.S., Lammens, T.M., Tan, B., Weale, J., and Hallett, J.P. (2017). An economically viable ionic liquid for the fractionation of lignocellulosic biomass. *Green Chemistry*, 19(13), 3078–3102.
- Carvalho, D.M.D. and Colodette, J.L. (2017). Comparative study of acid hydrolysis of lignin and polysaccharides in biomasses. *BioResources*, 12(4), 6907–6923.
- Firth, A.E.J., Nakasu, P.Y.S., Hallett, J.P., and Matthews, R.P. (2024). Exploiting Cation Structure and Water Content in Modulating the Acidity of Ammonium Hydrogen Sulfate Protic Ionic Liquids. *The Journal of Physical Chemistry Letters*, 15(9), 2311–2318.
- Malaret, F., Gschwend, F.J., Lopes, J.M., Tu, W.C., and Hallett, J.P. (2020). Eucalyptus red grandis pretreatment with protic ionic liquids: effect of severity and influence of sub/super-critical CO₂ atmosphere on pretreatment performance. *RSC Advances*, 10(27), 16050–16060.
- Nisar, S., Brandt-Talbot, A., Hallett, J.P., and Chachuat, B. (2024). Semi-mechanistic modelling of ionic liquid-based biomass fractionation. *Computer Aided Chemical Engineering*, 53, 2527–2532.
- Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R., and Tschaplinski, T. (2006). The Path Forward for Biofuels and Biomaterials. *Science*, 311(5760), 484–489.
- Suiter, J.B., Ruiz, R.O., Scarlata, C.J., Sluiter, A.D., and Templeton, D.W. (2010). Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. *Journal of Agricultural & Food Chemistry*, 58(16), 9043–9053.
- Taherzadeh, M.J. and Karimi, K. (2008). Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review. *International Journal of Molecular Sciences*, 9(9), 1621–1651.
- Verdía Barbará, P., Abouelela Rafat, A., Hallett, J.P., and Brandt-Talbot, A. (2023). Purifying cellulose from major waste streams using ionic liquids and deep eutectic solvents. *Current Opinion in Green & Sustainable Chemistry*, 41, 100783.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., and Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.
- Weigand, L., Mostame, S., Brandt-Talbot, A., Welton, T., and Hallett, J.P. (2017). Effect of pretreatment severity on the cellulose and lignin isolated from *Salix* using ionic liquid pretreatment. *Faraday Discussions*, 202, 331–349.