Supporting Information

A data-driven approach to generate pseudo-reaction networks for thermal conversion of Athabasca bitumen

Kaushik Sivaramakrishnan, *[+] Anjana Puliyanda, [+] Arno de Klerk, Vinay Prasad
[+] Equal contribution

Department of Chemical and Materials Engineering, University of Alberta, Edmonton AB

*Email: sivarama@ualberta.ca
Phone: +1-(780) 200-2049

1. Introduction

Detailed information on extraction of concentration and spectral profiles of the liquid products obtained at different experimental conditions of thermal conversion by self-modeling curve resolution (SMCR) is provided in the manuscript. However, some sections do not require that all figures, plots and tables be supplied in the manuscript itself, at the same time not causing difficulty for the readers in relating to the global aim of the study. These additional details are given in this Supporting Information document.

2. Experimental

All experimental details are provided in the main manuscript.

3. Methods and parameters used

3.1 FTIR data available

All the data regarding the FTIR spectra of the liquid products from thermal conversion at different temperatures and residence times is provided in the manuscript itself.

3.2 Pre-processed and residual data for temperatures of 420 ˚C, 400 ˚C, 380 ˚C, 300 ˚C

The FTIR spectra of liquid samples obtained after thermal conversion at 350 ºC after baseline correction and SG filtering are provided in the manuscript. The respective plots along with the residual obtained from smoothing and the raw data for the other 4 temperatures are given in Figure S1, Figure S2, Figure S3 and Figure S4.

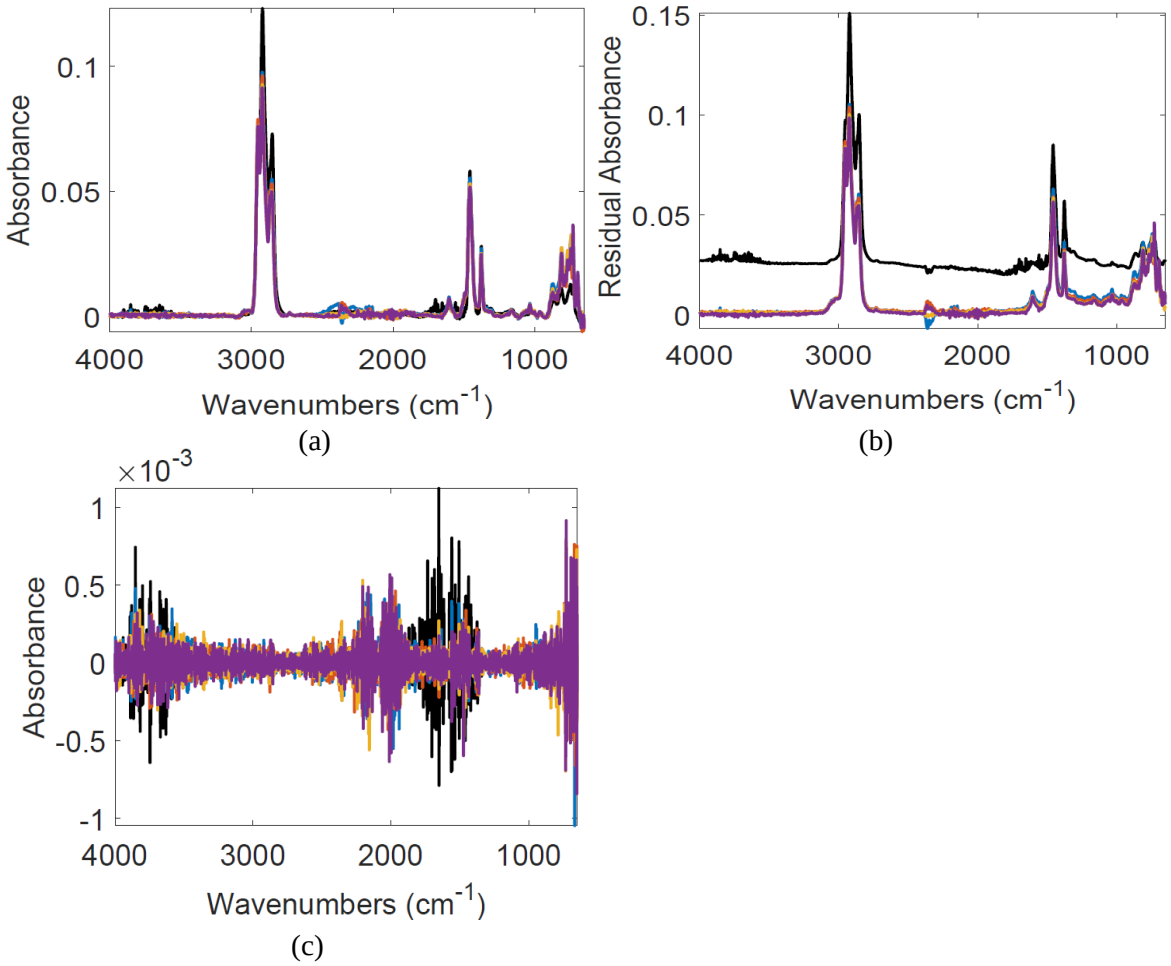


Figure S1. Plots of: (a) Baseline corrected and smoothed data; (b) the raw FTIR spectra of the liquid products from thermal conversion of Athabasca bitumen at 420 ºC; (c) residual after smoothing.

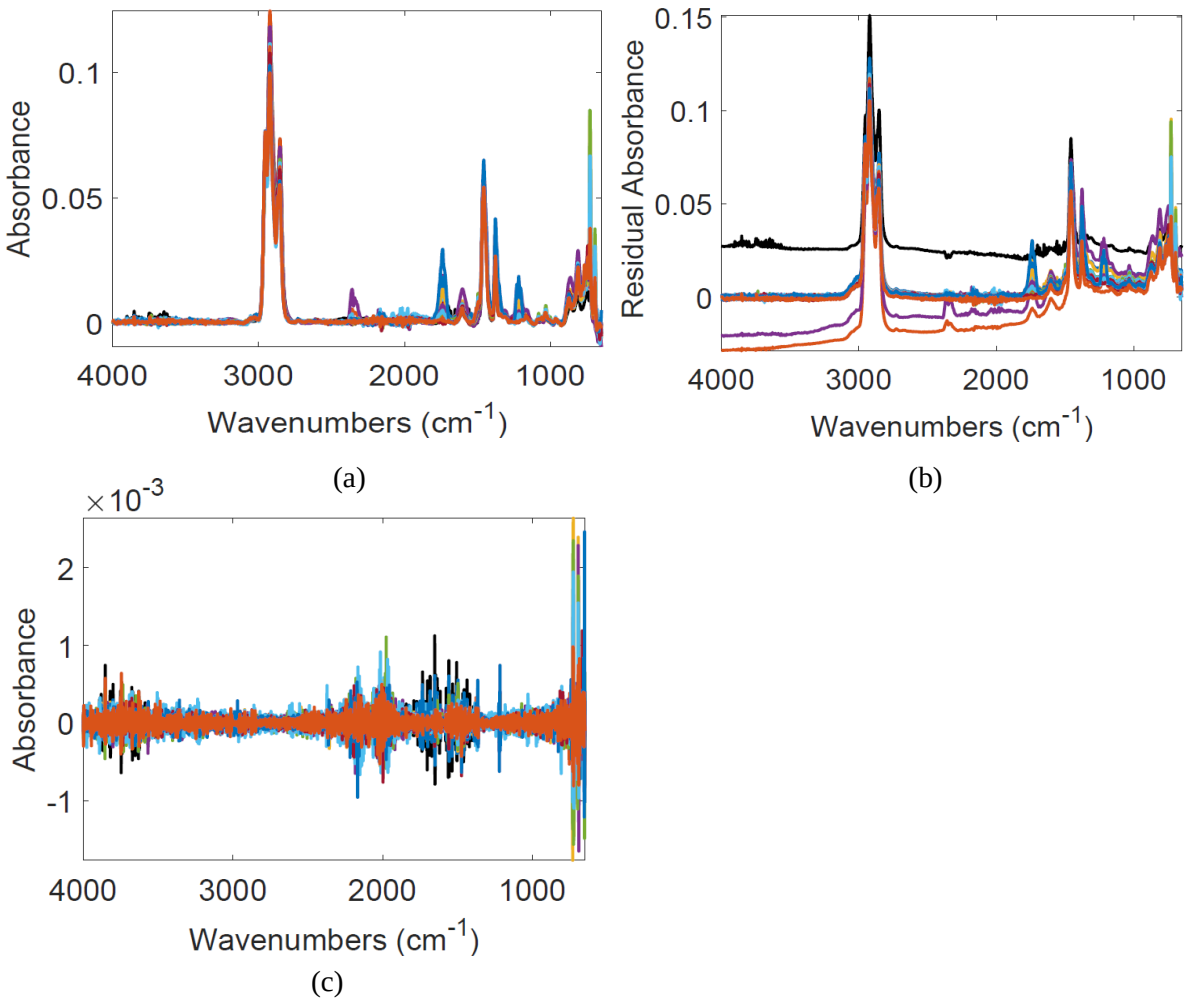


Figure S2. Plots of: (a) Baseline corrected and smoothed data; (b) the raw FTIR spectra of the liquid products from thermal conversion of Athabasca bitumen at 400 ºC; (c) residual after smoothing.

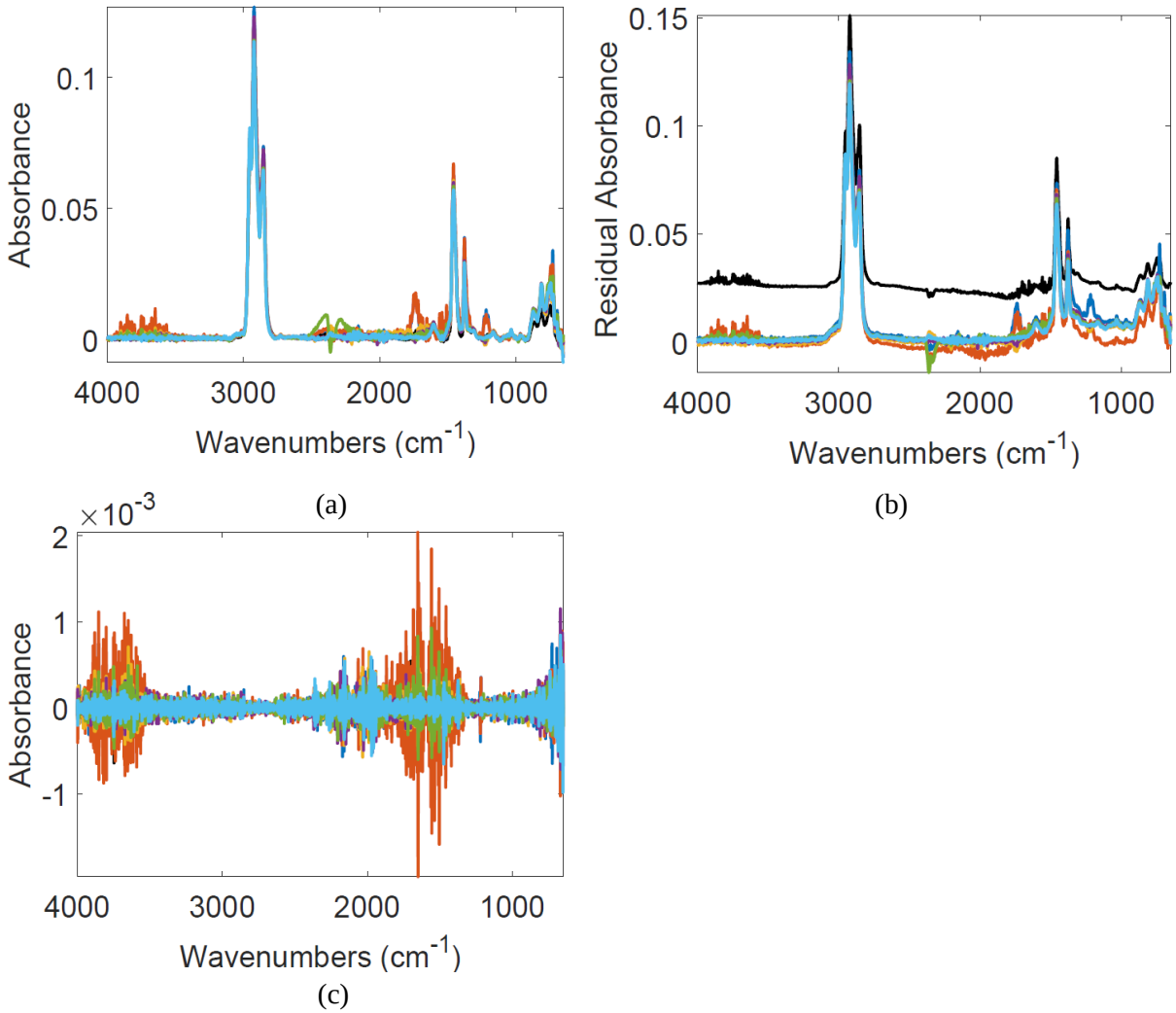


Figure S3. Plots of: (a) Baseline corrected and smoothed data; (b) the raw FTIR spectra of the liquid products from thermal conversion of Athabasca bitumen at 380 ºC; (c) residual after smoothing.

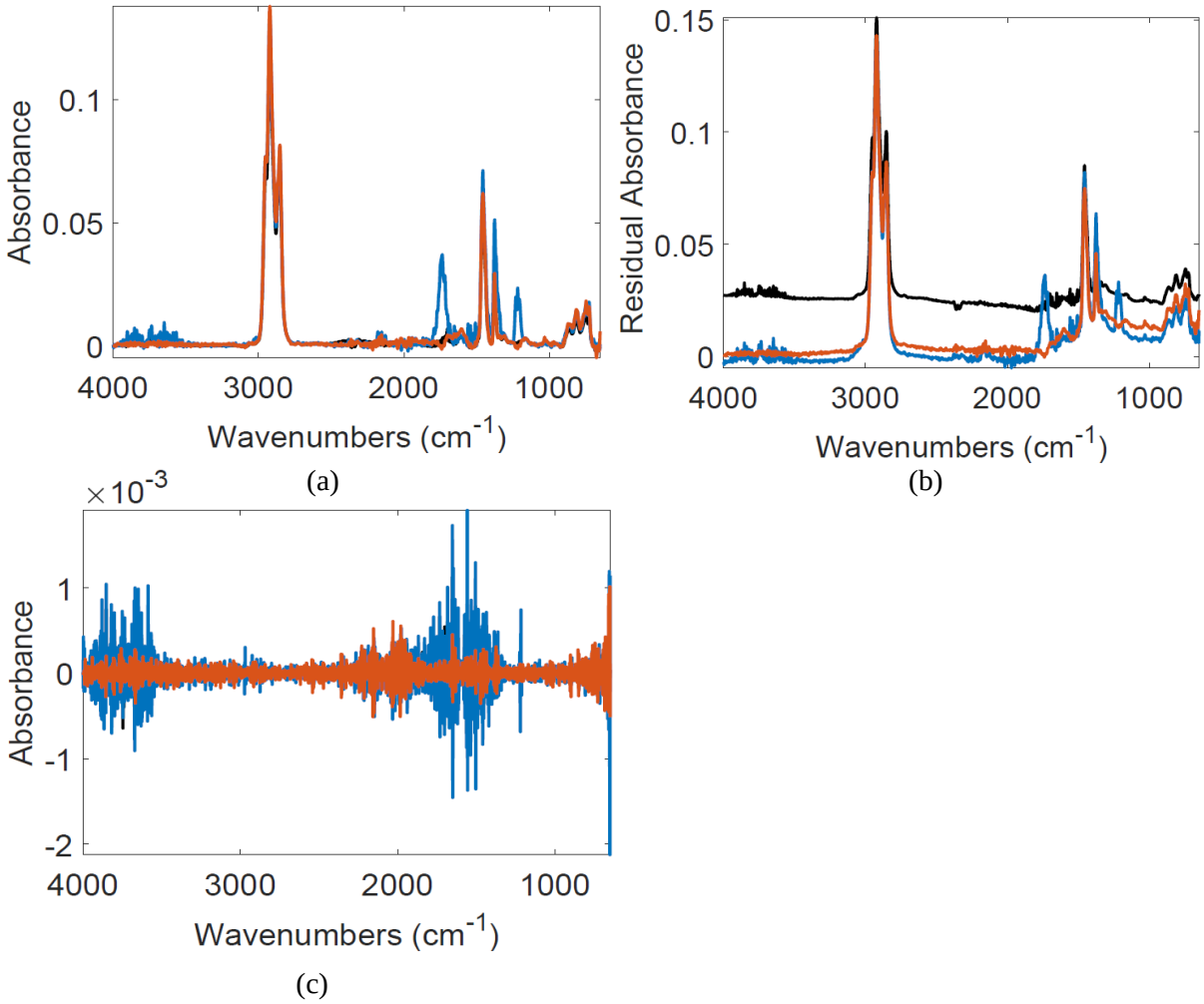


Figure S4. Plots of: (a) Baseline corrected and smoothed data; (b) the raw FTIR spectra of the liquid products from thermal conversion of Athabasca bitumen at 300 ºC; (c) residual after smoothing.

3.3 SMCR-ALS and SMCR-ALS-PSO methods

To deal with some of the limitations of MCR like rotational and intensity ambiguities, datasets from different runs and techniques are combined together into a single data matrix. A row-wise combination is performed when the same batch of experiments is monitored by different sets of techniques like FTIR, NMR, ESR, etc. The parent equation is illustrated in equation S1. A column-wise matrix is obtained when multiple batches of experiments conducted at different experimental conditions are monitored by the same technique. This is given in equation S2.

$$[D_1\ D_2\ D_3 \ldots D_n] = C[S_1\ S_2\ S_3 \ldots S_n]^T + [E_1\ E_2\ E_3 \ldots . E_n] \tag{S1}$$

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} S^T \tag{S2}$$

Intensity ambiguity is represented by:

$$D = (Ck)(S^T \frac{1}{k}) \quad \text{(S3)}$$

where k is a scalar.

Rotational ambiguity is given in equation S4 by:

$$D = (CT)(T^{-1}S^T) + E \quad \text{(S4)}$$

where T is a non-singular invertible matrix that multiplies with C and whose inverse multiplies with S. There are infinite possibilities for T in the absence of other constraints.

The ALS-optimization algorithm and the accompanying constraints is described in the manuscript. The respective equations of the alternative minimization of the Frobenius norm of the residual are given below:

$$\min_{S \geq 0}(\left|\left|D - CS^T\right|\right|^2) \quad \text{(S5)}$$

$$\min_{C \geq 0}(\left|\left|D - CS^T\right|\right|^2) \quad \text{(S6)}$$

Table S1 gives some of the common strategies of choosing the inertia weight parameter for velocity updating in PSO.

Table S1. Common strategies for inertia weight employed in the PSO literature.

Type of strategy	Remarks
Constant [1]	A value between 0.7 – 1 shows lower error but larger number of iterations for convergence
Random [2]	Increases convergence in early stages of PSO; Gives faster overall convergence
Linearly decreasing [3]	Decreasing values in the range 0.9 – 0.4 are employed but risk of local optimum exists; Gives low error
Global-local best inertia weight [4]	Falls in between constant and random inertia weight strategies; takes global and local best particle positions into consideration but gives large error

As mentioned in the manuscript, *'fmincon'* was used to further carry out a local search for the PSO-optimized concentration profiles inside the ALS loop. The next few paragraphs discuss two algorithms used by *'fmincon'* for the optimization process in further detail. These are the 'Sequential Quadratic Programming' algorithm and the 'Interior Point' algorithm. First, a nonlinear unconstrained minimization problem of a general nature is explained, followed by the algorithms for the constrained optimization.

a. Unconstrained minimization:

Consider a scalar function $f(x)$ whose minimum point and the corresponding value needs to be found. Most algorithms are based on building trust regions around the neighborhood (N) for a simplified version q of f. [5] The trust region sub-problem is expressed in equation S7 as:

$$\min_s q(s), s \in N \tag{S7}$$

where s is a sample step that assists in updating the present position if $f(x+s) < f(x)$.

The challenge is to define q and the trust region N. Expressing q in terms of the first two terms of the Taylor's expansion, the quadratic programming problem comes down to solving the equation:

$$\min \frac{1}{2} s^T H s + s^T g \text{ for } ||Ds|| \leq \Delta \tag{S8}$$

In equation S8, D is the diagonal scaling matrix, Δ is a positive tolerance level for the constraint and can be adjusted according to whether the updated value of f meets the inequality condition or not, H is the square matrix of second derivatives of f (Hessian) and g is the gradient of f. A number of approaches to solve this equation are given in the literature. [6,7] All these algorithms require rigorous calculations of eigenvalues but it is easier to solve using the definition of a sub-space s that forms a boundary for the trust region. s is constructed in the 2-D space as a combination of the gradient direction (s_1) and the Newton direction (s_2), which is the solution to the following equation

$$H.s_2 = -g \tag{S9}$$

The solution to equation S9, which is a system of linear equations, is given by the preconditioned conjugate gradient (CG) method whose output direction, p is used to build the sub-space. The key step in solving unconstrained optimization problems is determining the 2-D sub-space. It is chosen such that global convergence is achieved through the steepest descent direction while local convergence is accomplished through the Newton step. Nonlinear least squares and linear least squares solutions also work on similar principles of trust regions and 2-D sub-space.

b. Constrained minimization:

Two common constraints for these kinds of problems are linear equality and box constraints. The linear equality constrained problems are solved considering an initial point that satisfies the equality $Ax_0 = b$, where A and b are known. A matrix system is created to calculate s and is elaborated by Coleman and Verma. [8] Box constraints consist of lower and upper bounds and a scaled Newton step evolving from the Karush-Kuhn-Tucker (KKT) conditions is considered to find the sub-space for solving the problem. [9] The solution also comprises of a reflection step that delineates the step size.

c. Algorithms used by *'fmincon'*

Active set algorithm:

This is a medium-scale algorithm where full matrices are generated and complex linear algebra is used to solve the constrained equations. They were based on the conversion of the constrained problem into an unconstrained one by the use of a penalty function. The KKT conditions are necessary and sufficient for optimality when both the objective function and the constraints are convex. The KKT conditions of the quadratic programming problem are given as:

$$\nabla f(x_s) + \sum_{i=1}^{m} \lambda_i . \nabla \mathrm{G_i}(x_s) = 0$$
$$\lambda_\mathrm{i} . \mathrm{G_i}(x_s) = 0 \text{ and } \lambda_i \geq 0 \quad \text{(S10)}$$

where λ_i are the Lagrange multipliers that take positive values only and serve as a link between the objective and constraint functions. The solution revolves around finding the Lagrange multipliers for each data point.

Sequential Quadratic Programming (SQP) algorithm:

'*fmincon'* utilizes SQP methods frequently to solve the constrained optimization problems. The principle of SQP rests on creating quadratic programming sub-problems at each loop iteration. [10] It is analogous to the active-set algorithm explained in the previous section and instead of a Newton step used for the unconstrained optimization (equation S9), a quasi-Newton updating procedure is used for dealing with the Hessian matrix (H). Detailed reviews of the method are available in various texts in the literature. [11,12]

The solution of the quadratic sub-problem is used to form a search direction for the variable x as:

$$x_{k+1} = x_k + \alpha_k d_k \quad \text{(S11)}$$

Here, d_k is the search direction and α_k is the step length parameter obtained by line search. It helps the solution to progress toward the function minimum by decreasing the value of the objective function. Schittkowski [12] also opined that the advantage of utilizing the SQP method is that it makes the constrained optimization converge faster than an unconstrained problem due to a fixed search area and α_k. The SQP algorithm has 4 major steps:

(i) Updating the Hessian (H_k) of the Lagrangian formulation

The Lagrangian formulation of the quadratic problem is given by the following equation:

$$L(\lambda, x) = f(x) + \sum \lambda_i . g_i(x) \quad \text{(S12)}$$

A quasi-Newton approximation of $H(L(\lambda, x))$ is conducted at each iteration. In order to track the convergence path in MATLAB, the 'Display' option can be set to 'iter'. When this is done, messages such as 'Hessian modified', 'infeasible' are displayed that indicate that the extent of nonlinearity is high.

(ii) Solution of the QP sub-problem

The solution of this problem is executed by the active-set method described in the previous section. It is also called a projection method. This involves primarily two steps: estimating a feasible starting point and then generating a number of points that remain active throughout the iterations and subsequently converge to the final solution. The active points lead to the search direction (d_k in equation S11) that is

present on the boundaries of the given constraints. This search direction facilitates the calculation of the new point of x in the search space (equation S11). d_k is usually obtained through a linear combination of a vector that is orthogonal to the active points.

Two directional choices are available for α_k during the line search procedure. One is the direct step along d_k that would lead to the optimum of $f(x)$ considering the active point set and thus, the solution of the QP sub-problem. If this does not occur, further iterations are required to reach the solution. The condition of positive Lagrange multipliers needs to be satisfied, otherwise the equality constraint is violated and the data point corresponding to this violation is removed from the algorithm.

(iii) Finding the starting point

This can be done by finding an x that satisfies the equality constraint in the QP sub-problem. A system of linear equations needs to be solved to obtain the initial point. The initial search direction can be obtained by substituting d_k for s in equation S9.

(iv) Merit function and step length

A merit function proposed by Han [13] is used and a penalty parameter was introduced by Powell. [14] The merit function is similar to the Lagrangian function L but has more parameters. The penalty parameter distinguishes between constraints having smaller and larger gradients and penalizes the smaller gradients more. The step length parameter, as discussed before, reduces the merit function value.

From the implementation viewpoint, the algorithm in MATLAB allows for failed steps in the case of a bogus value for the objective function. During the running of the algorithm, lesser memory and time is consumed as compared to the active-set strategy though both are medium-scale algorithms. In addition, in the case of some nonlinear constraints being violated, SQP calculates a second order approximation for the constraints and proceeds with the iteration, though it sacrifices convergence speed.

Interior Point Algorithm:

This is the default algorithm adopted by MATLAB for the *'fmincon'* function. A detailed description of this method is given by Waltz et al. [15] and only the two important steps of the solution process are described in this section. The main objective function is split into constituent small-scale optimization problems given by equation S13:

$$\min_{x,s} f(x,s) = \min f(x,s) - \mu \sum \ln(s_i) \tag{S13}$$

where s_i are the slack variables and μ is a positive parameter that controls the barrier function $\sum \ln(s_i)$

The purpose of the approximate problem is the conversion of inequality constraints to equality constraints to make it easier for problem solving. Equation S13 can be solved by taking either of the following 2 steps: direct step or a CG step. The KKT conditions are applied to the QP and the obtained system of equations are tried to be solved by linear approximation. This is the first and default step attempted by the algorithm. The CG step comes into play when the objective functions fails to remain convex at any iteration. In either case, a merit function that combines the objective function and the constraints is required to be decreased in value as much as possible. The algorithm can deal with constraint violations when a particular point x_j returns an unreal value for the constraint function. In this situation, the step length is modified to a shorter value and the iteration is continued.

In the direct step, matrix factorization gives information about the Hessian. If the Hessian is not positive definite, the algorithm attempts to solve the system of equations using the CG method. Similar to the unconstrained minimization, CG utilizes a trust region to create a sub-space for the solution to the QP

problem. As with other cases, Lagrangian multipliers are obtained from solving KKT condition equations to obtain the solution for the interior point algorithm. Unlike SQP, interior point algorithm is a large-scale algorithm that does not store or generate full sized matrices and thus, lesser space is used and is the preferred approach for computer programming.

4. Results and Discussion

4.1 Rank determination of each sub-matrix

Figure S5 gives the plots of residuals obtained after performing SVD on the 400 ºC data set choosing 2 and 4 components while the manuscript gives the residual plot for SVD performed with optimal 3 components. The ROD, SD, residual after performing SVD with 3 components and the scree plots for data sets at the other 4 temperatures (300 ºC, 350 ºC, 380 ºC, 420 ºC) are given in Figure S6, Figure S7, Figure S8 and Figure S9 respectively.

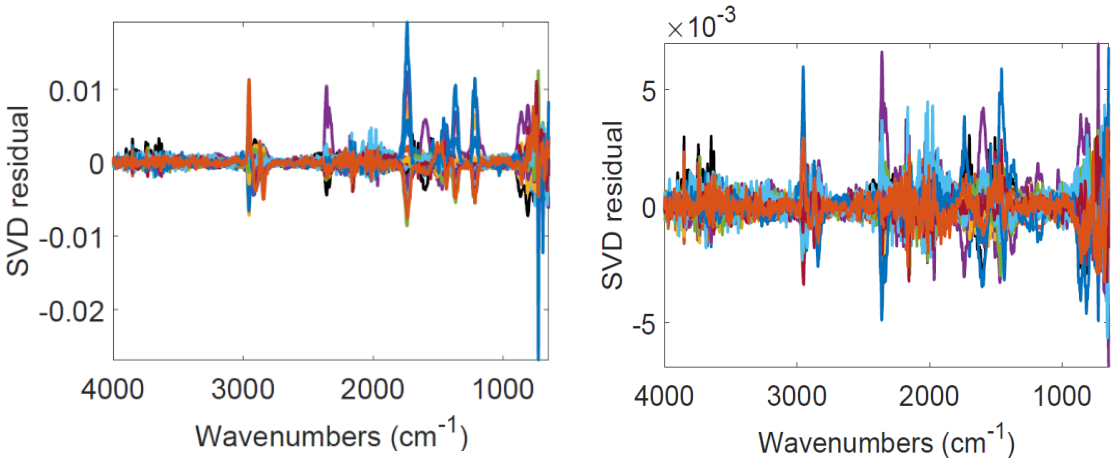


Figure S5. Residuals obtained after performing SVD on the 400 ºC data set considering: (a) 2 components and (b) 4 components.

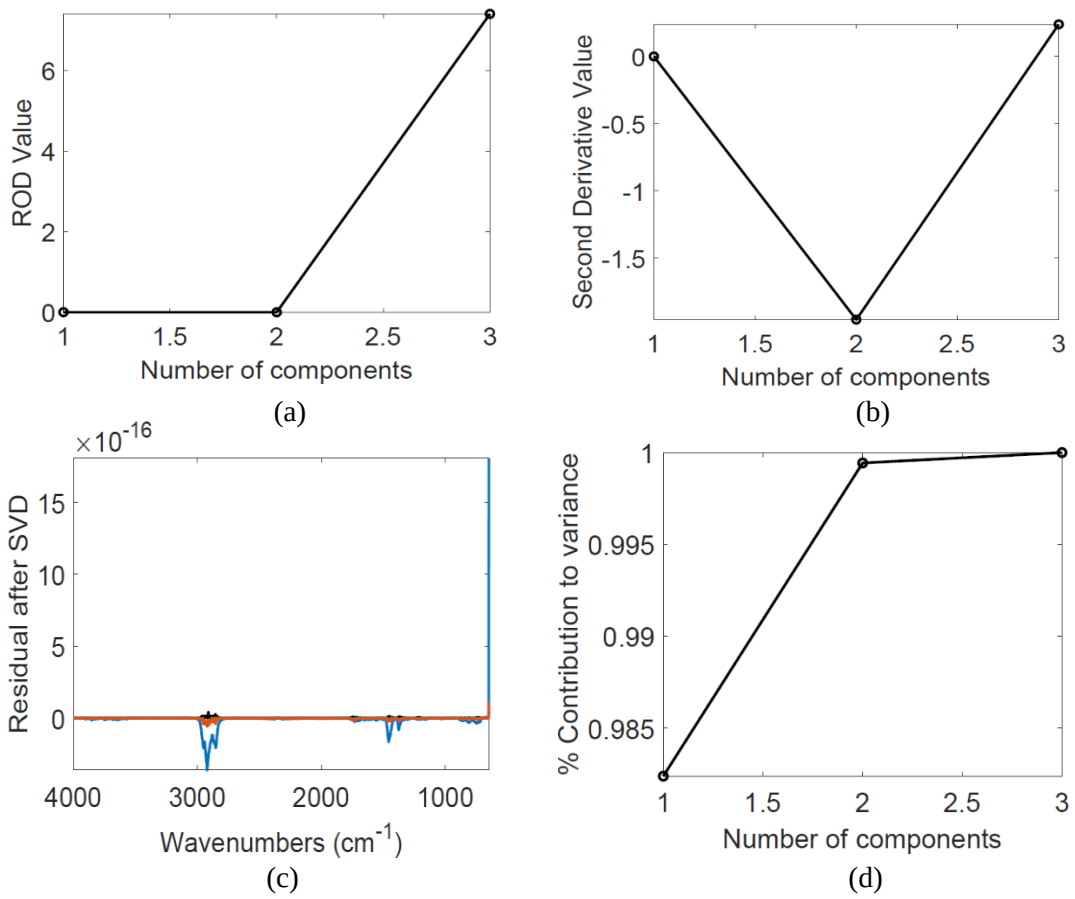


Figure S6. Plots for (a) ROD with respect to each component; (b) SD with respect to each component; (c) Residual after performing SVD considering 3 components on the FTIR data set for all 1738 wavenumbers; (d) Percentage contribution to the variance explained by the eigenvalues corresponding to each component in the system. These results correspond to data obtained at 300 ºC.

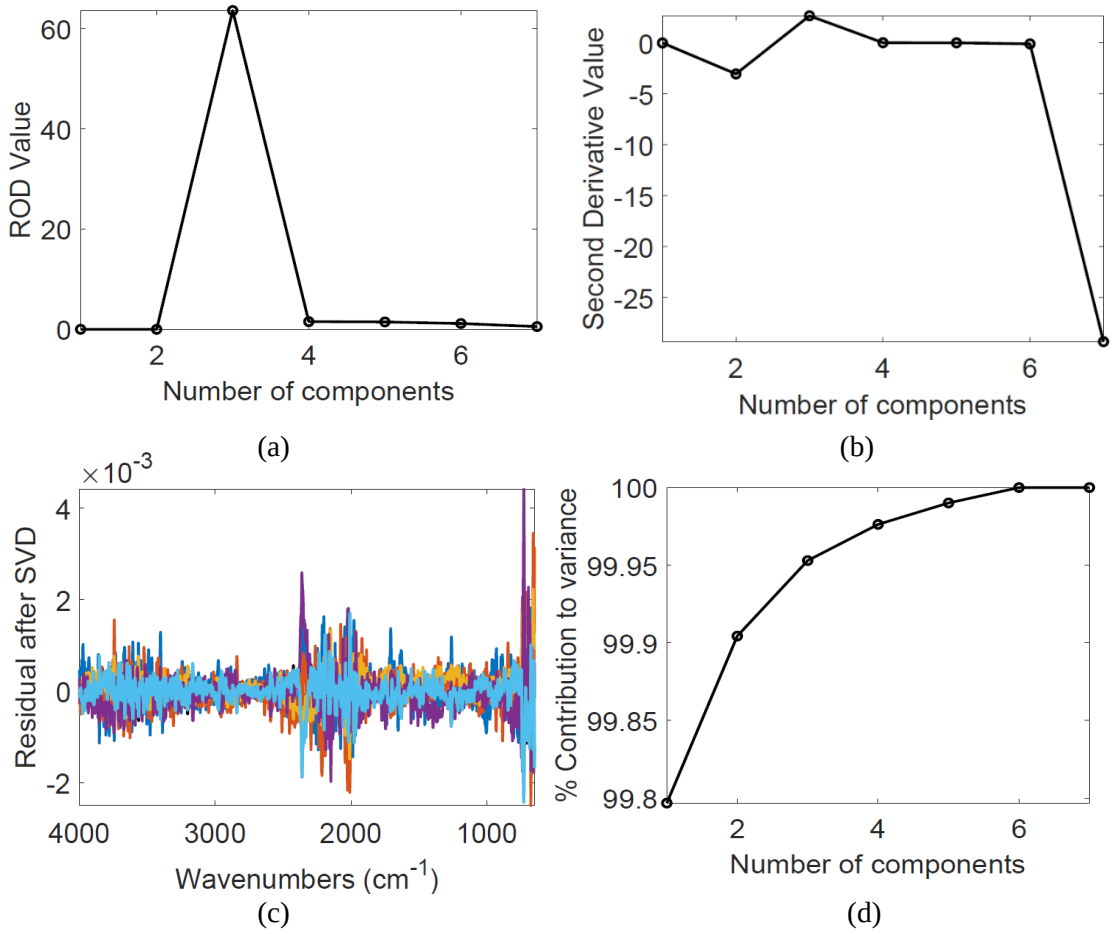


Figure S7. Plots for (a) ROD with respect to each component; (b) SD with respect to each component; (c) Residual after performing SVD considering 3 components on the FTIR data set for all 1738 wavenumbers; (d) Percentage contribution to the variance explained by the eigenvalues corresponding to each component in the system. These results correspond to data obtained at 350 ºC.

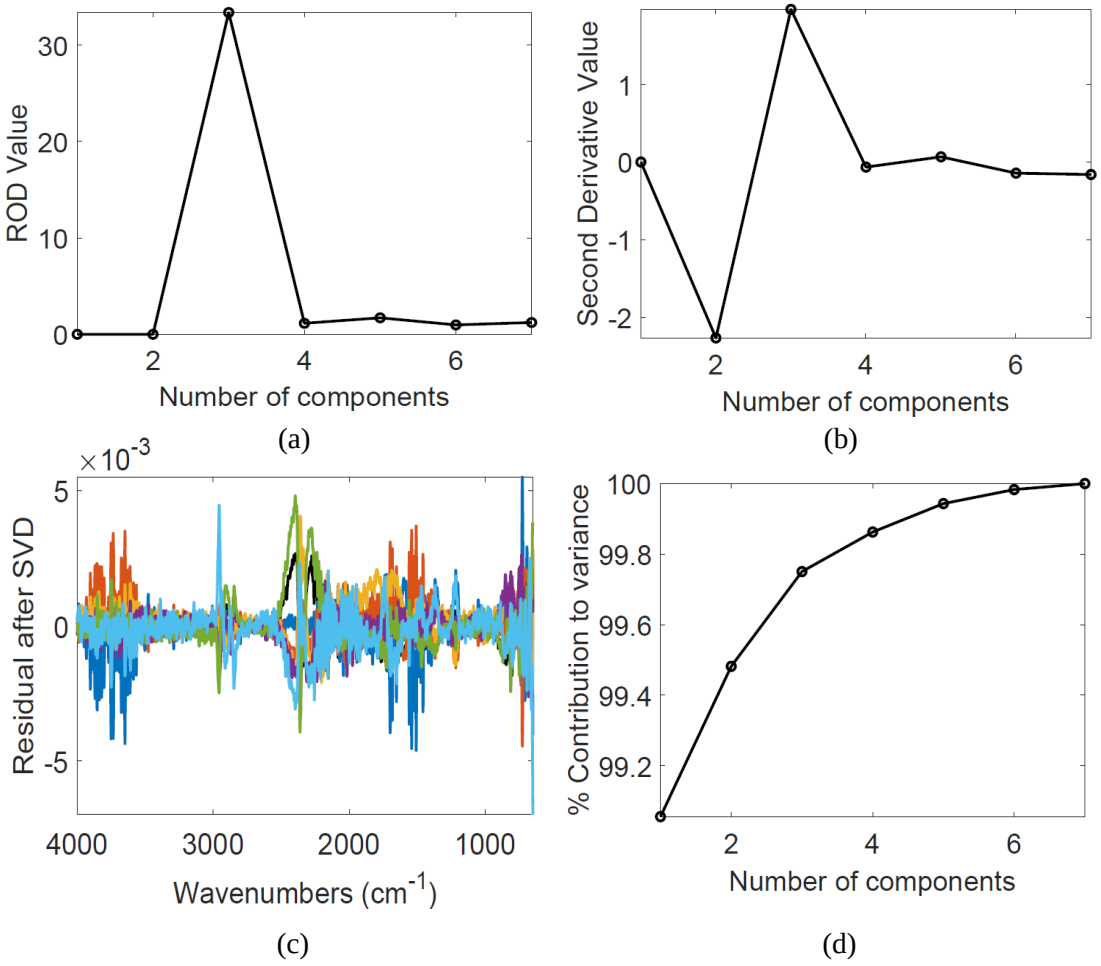


Figure S8. Plots for (a) ROD with respect to each component; (b) SD with respect to each component; (c) Residual after performing SVD considering 3 components on the FTIR data set for all 1738 wavenumbers; (d) Percentage contribution to the variance explained by the eigenvalues corresponding to each component in the system. These results correspond to data obtained at 380 ºC.

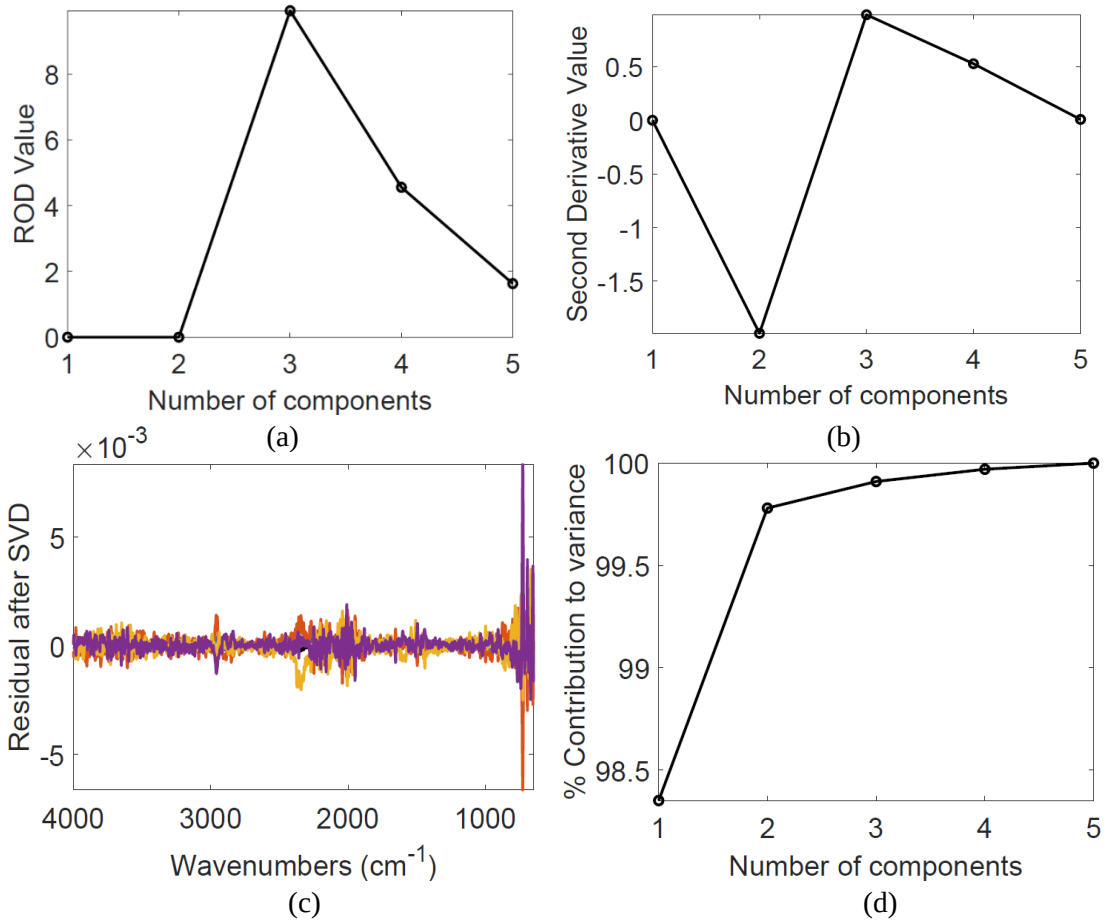


Figure S9. Plots for (a) ROD with respect to each component; (b) SD with respect to each component; (c) Residual after performing SVD considering 3 components on the FTIR data set for all 1738 wavenumbers; (d) Percentage contribution to the variance explained by the eigenvalues corresponding to each component in the system. These results correspond to data obtained at 420 ºC.

The values of performance indicators (LOF and R^2) for SVD with 2, 3 and 4 pseudo-components are given in Table S2.

Table S2. LOF and R^2 values (% contribution to variance) on reconstruction of the original matrix after performing SVD for the datasets at 300 ºC, 350 ºC, 380 ºC and 420 ºC.

	300 ºC		**350 ºC**			**380 ºC**			**420 ºC**		
No. of components	**2**	**3**	**2**	**3**	**4**	**2**	**3**	**4**	**2**	**3**	**4**
LOF	2.38	8.27E-14	3.09	2.17	1.54	7.20	4.99	3.71	4.72	2.93	1.83
R^2	99.94	100	99.90	99.95	99.97	99.48	99.75	99.86	99.78	99.91	99.96

4.2 Initial concentration estimates

The initial estimates of concentration profiles at 300 ºC are given in Figure S10.

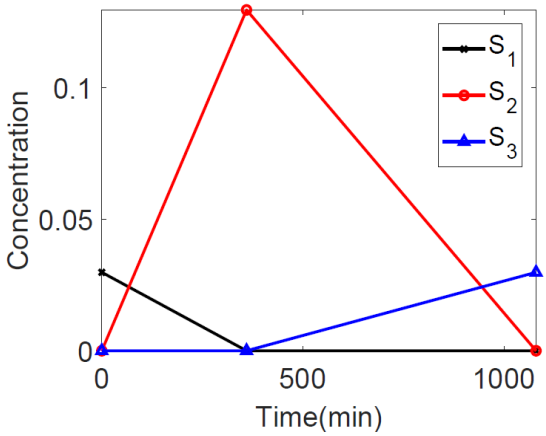


Figure S10. Initial concentration estimates for S_1, S_2 and S_3 at 300 ºC.

4.3 ALS-optimized profiles and spectra-derived quantitative parameters

The residuals obtained after subtracting the ALS-reproduced matrix from the original matrix for datasets at all temperatures are given in Figure S11.

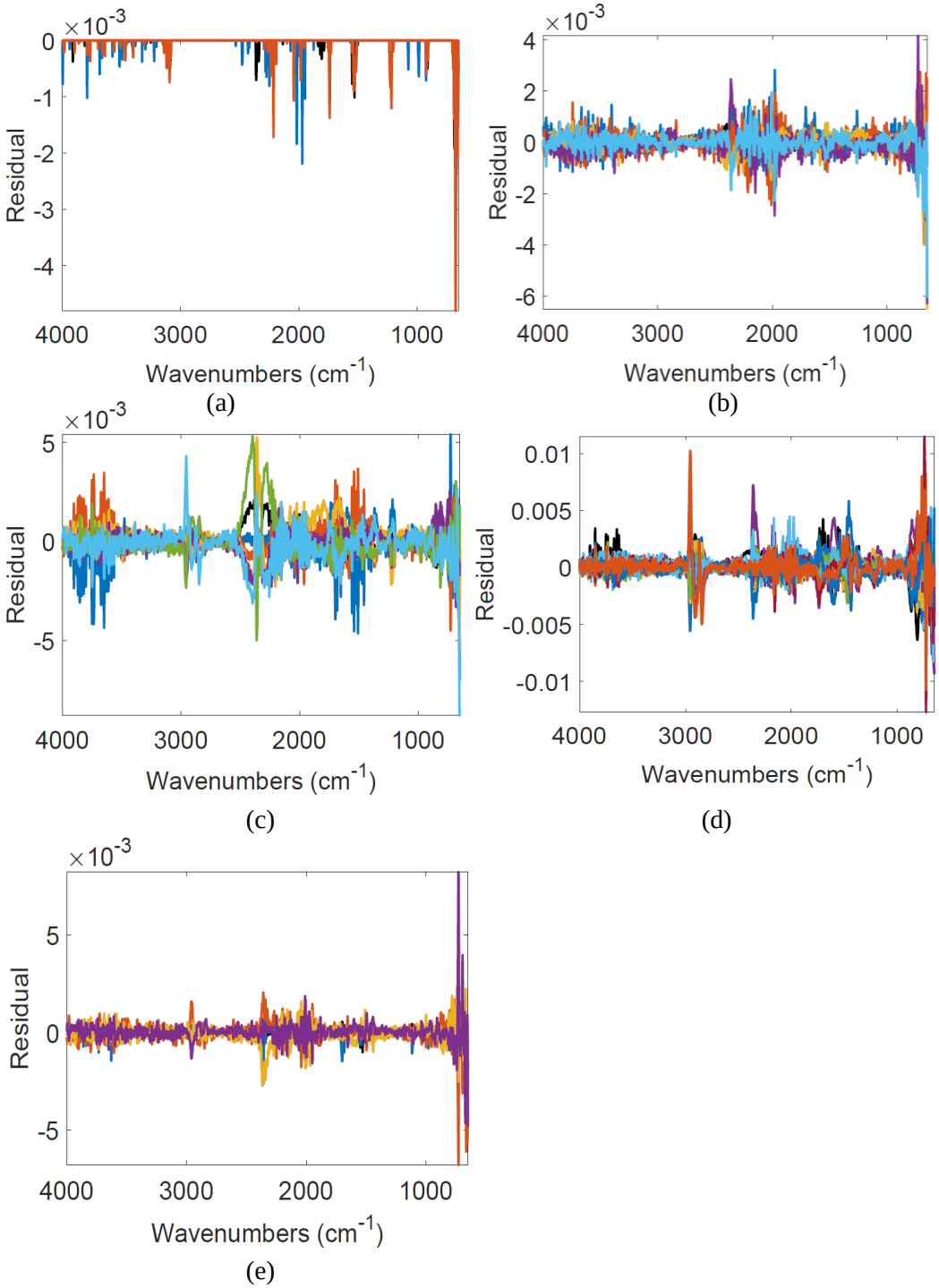


Figure S11. ALS residuals for datasets obtained at: (a) 300 ºC; (b) 350 ºC; (c) 380 ºC; (d) 400 ºC; (e) 420 ºC.

4.4 PSO-optimized concentration and spectral profiles

4.4.1 Results at 300 ºC

The concentration and spectral profiles when the ALS-PSO algorithm was used to resolve the FTIR spectra obtained at 300 ºC for Athabasca bitumen is given in Figure S12. The residual when the reproduced matrix from the ALS-PSO-resolved profiles is subtracted from the original data matrix is also provided in this figure (Figure S12b). Discussion on the differences of these profiles with respect to ALS-optimized results in terms of resolution quality and convergence speed is provided in the manuscript.

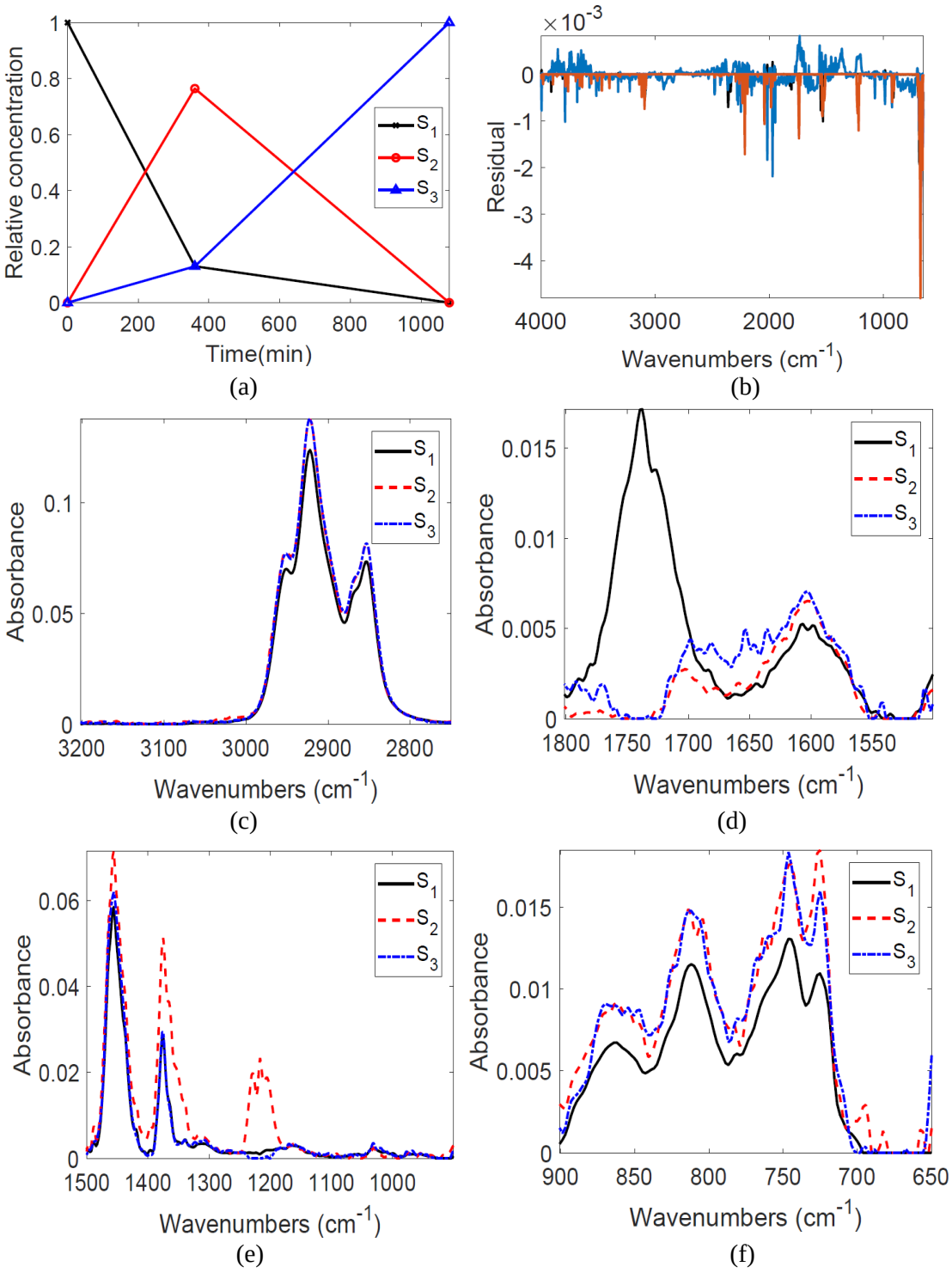


Figure S12. Results of SMCR-ALS-PSO applied to FTIR spectra of liquid products from thermal conversion of Athabasca bitumen at 300 ºC. The profiles are arranged as: (a) concentration vs. reaction time for the three pseudo-components; (b) residual plot; and resolved spectra for each pseudo-component shown as absorbance vs. wavenumber in the ranges: (c) 3200 – 2750 cm^{-1}; (d) 1800 – 1500 cm^{-1}; (e) 1500 – 900 cm^{-1}; (f) 900 – 650 cm^{-1}.

4.4.2 Results at 350 ºC

The concentration and spectral profiles when the ALS-PSO algorithm was used to resolve the FTIR spectra obtained at 350 ºC for Athabasca bitumen are given in Figure S13. The residual when the reproduced matrix from the ALS-PSO-resolved profiles is subtracted from the original data matrix is also provided in this figure. Discussion on the differences of these profiles from the ALS-optimized results in terms of resolution quality and convergence speed is given in the manuscript.

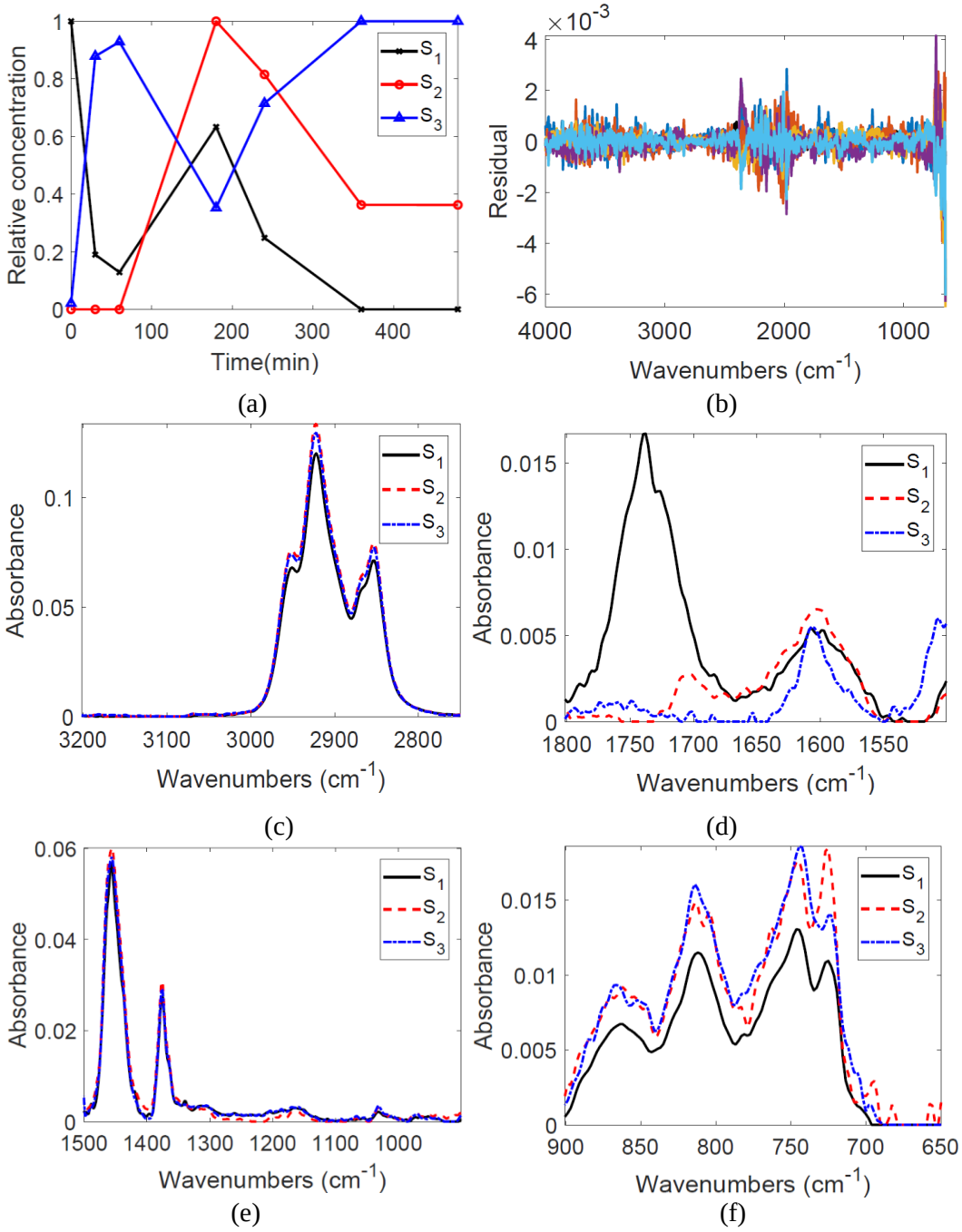


Figure S13. Results of SMCR-ALS-PSO applied to FTIR spectra of liquid products from thermal conversion of Athabasca bitumen at 350 ºC. The profiles are arranged as: (a) concentration vs. reaction time for the three pseudo-components; (b) residual plot; and resolved spectra for each pseudo-component shown as absorbance vs. wavenumber in the ranges: (c) 3200 – 2750 cm^{-1}; (d) 1800 – 1500 cm^{-1}; (e) 1500 – 900 cm^{-1}; (f) 900 – 650 cm^{-1}.

4.4.3 Results at 380 ºC

Figure S14 provides the ALS-PSO-resolved concentration and spectral profiles for the 380 ºC dataset. The residual plot when the reproduced matrix is subtracted from the original data matrix is also provided in the figure.

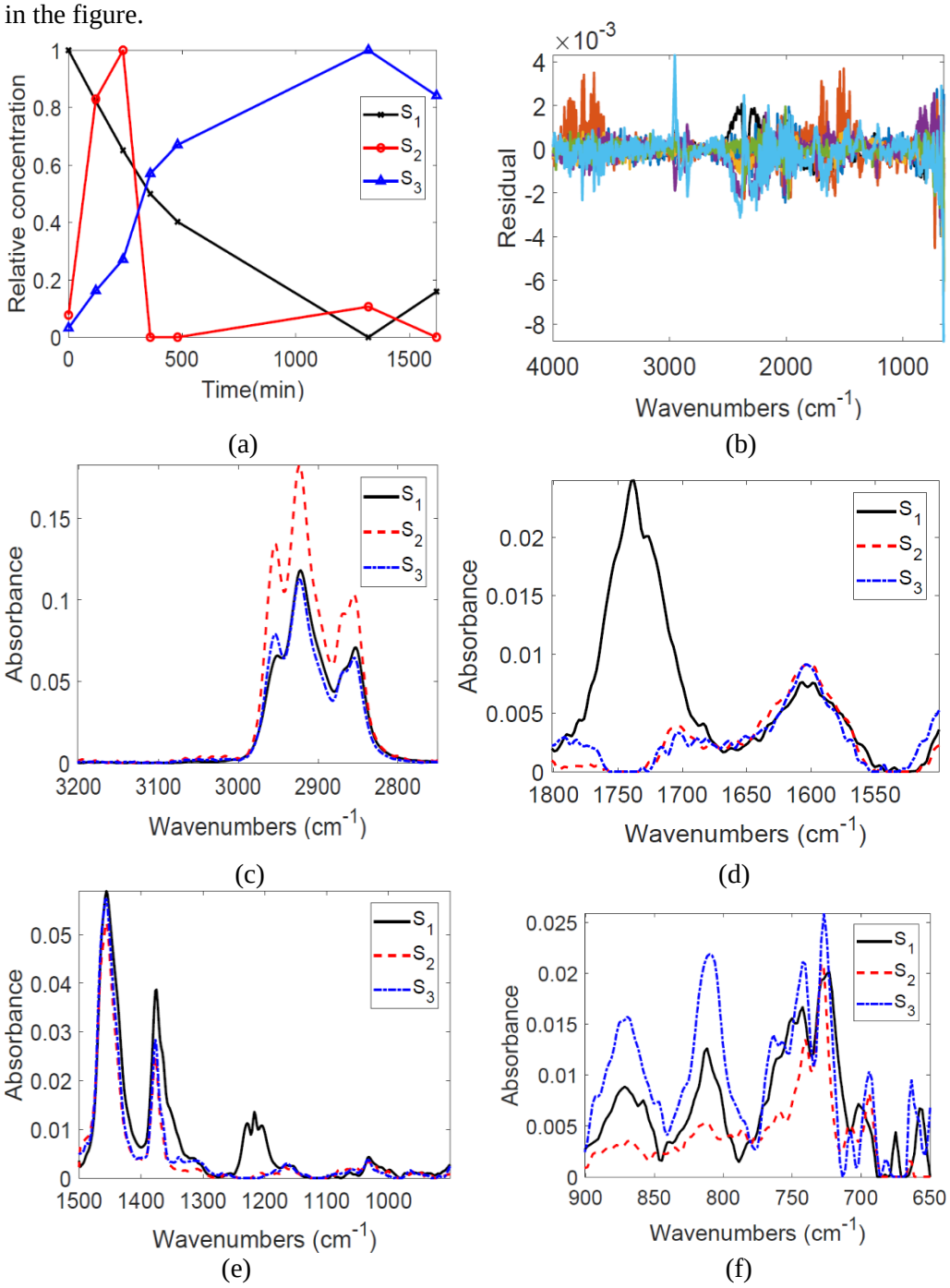


Figure S14. Results of SMCR-ALS-PSO applied to FTIR spectra of liquid products from thermal conversion of Athabasca bitumen at 380 ºC. The profiles are arranged as: (a) concentration vs. reaction time for the three pseudo-components; (b) residual plot; and resolved spectra for each pseudo-component shown as absorbance vs. wavenumber in the ranges: (c) 3200 – 2750 cm^{-1}; (d) 1800 – 1500 cm^{-1}; (e) 1500 – 900 cm^{-1}; (f) 900 – 650 cm^{-1}.

4.4.4 Results at 400 ºC

Figure S15 gives the ALS-PSO resolved final profiles for the dataset obtained at 400 ºC. The residual plot when the reproduced matrix is subtracted from the original data matrix is also provided in the figure.

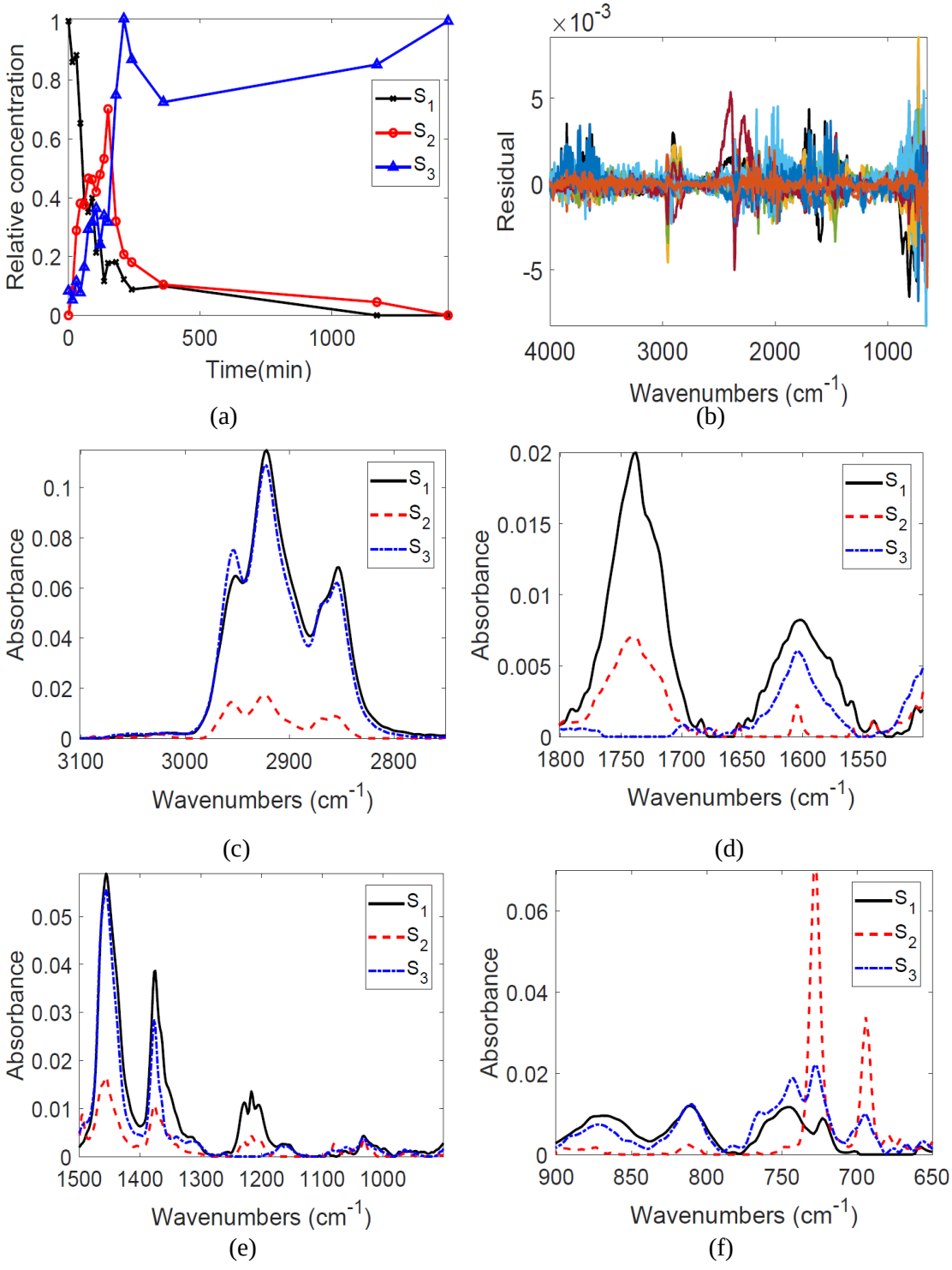


Figure S15. Results of SMCR-ALS-PSO applied to FTIR spectra of liquid products from thermal conversion of Athabasca bitumen at 400 ºC. The profiles are arranged as: (a) concentration vs. reaction time for the three pseudo-components; (b) residual plot; and resolved spectra for each pseudo-component shows as absorbance vs. wavenumber in the ranges: (c) 3200 – 2750 cm^{-1}; (d) 1800 – 1500 cm^{-1}; (e) 1500 – 900 cm^{-1}; (f) 900 – 650 cm^{-1}.

4.4.5 Results at 420 ºC

Figure S16 provides the concentration and spectral profiles for the ALS-PSO optimized profiles including the residual obtained when the reproduced data matrix is subtracted from the original matrix.

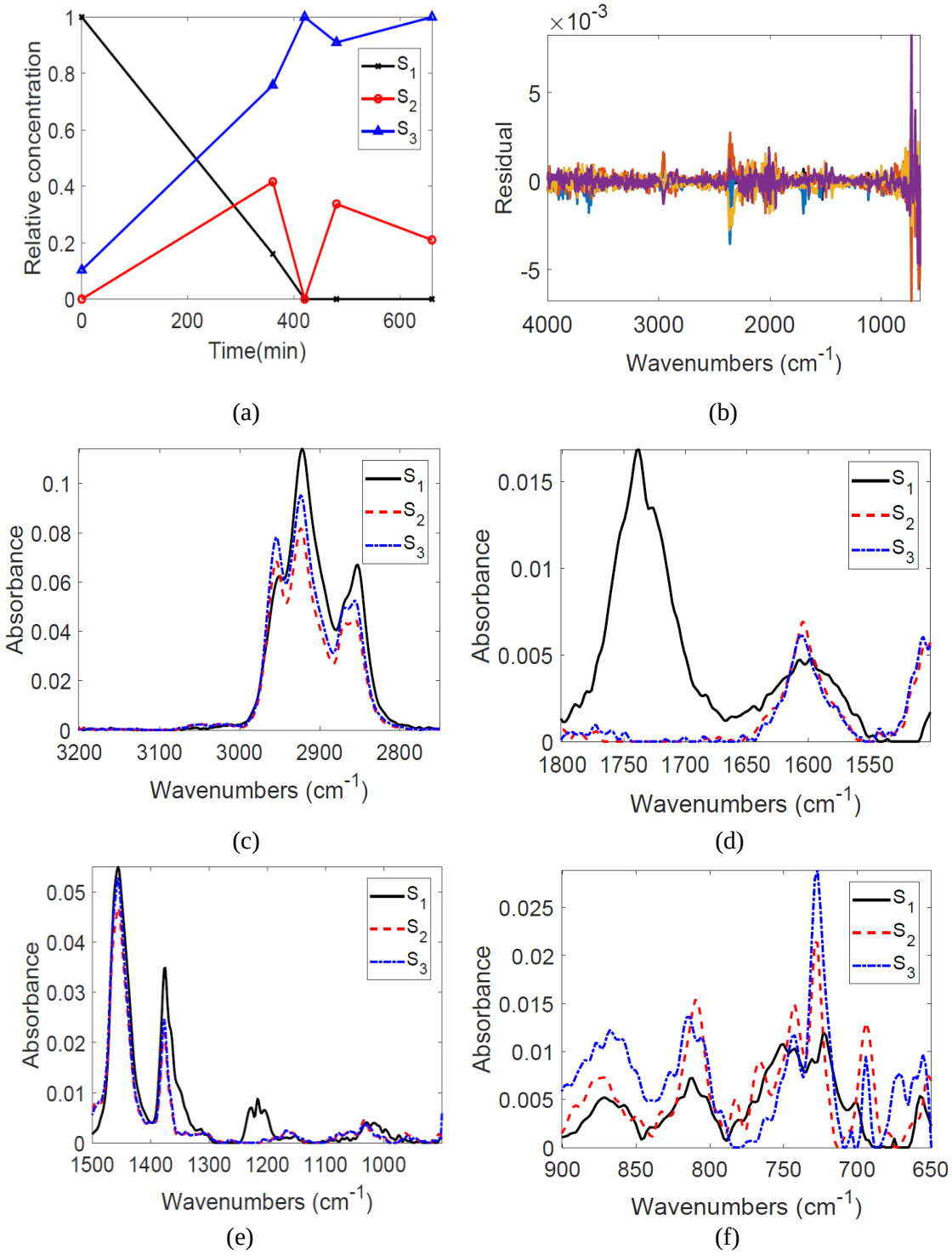


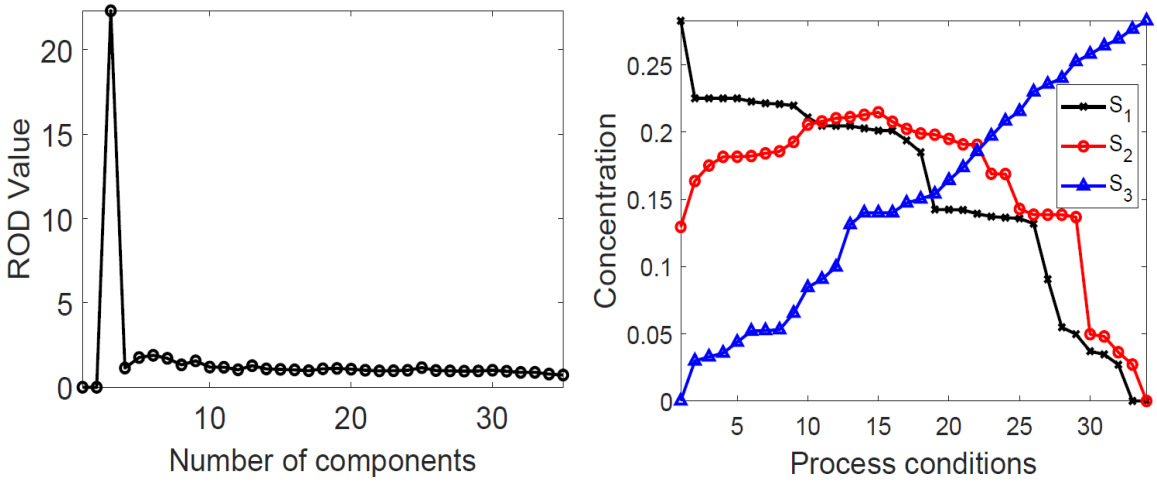
Figure S16. Results of SMCR-ALS-PSO applied to FTIR spectra of liquid products from thermal conversion of Athabasca bitumen at 420 ºC. The profiles are arranged as: (a) concentration vs. reaction time for the three pseudo-components; (b) residual plot; and resolved spectra for each pseudo-component shows as absorbance vs. wavenumber in the ranges: (c) 3200 – 2750 cm^{-1}; (d) 1800 – 1500 cm^{-1}; (e) 1500 – 900 cm^{-1}; (f) 900 – 650 cm^{-1}.

4.4.6 Comparison of ALS and ALS-PSO methods

The results and corresponding discussion of this section are provided in the manuscript itself.

4.5 ALS-optimized profiles for the global model

Figure S17 provides the plots for the ROD and initial concentration estimates obtained through EFA for the 35 samples when the augmented matrix consisting of all temperatures and respective reaction times was used for SMCR analysis.



(a) (b)

Figure S17. Plots of: (a) ROD vs. number of components and (b) initial estimates of concentration obtained through EFA for the 35 samples at various process conditions used in the SMCR-ALS global model.

References

1. Shi Y, Eberhart R. A modified particle swarm optimizer. In: 1998 IEEE International Conference on Evolutionary Computation Proceedings IEEE World Congress on Computational Intelligence (Cat No98TH8360). Anchorage, AK: IEEE; 2002. p. 69–73.
2. Eberhart RC, Yuhui Shi. Tracking and optimizing dynamic systems with particle swarms. In: Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat No01TH8546). Seoul: IEEE; 2002. p. 94–100.
3. Xin J, Chen G, Hai Y. A Particle Swarm Optimizer with Multi-stage Linearly-Decreasing Inertia Weight. In: 2009 International Joint Conference on Computational Sciences and Optimization. Sanya, China: IEEE; 2009. p. 505–8.
4. Arumugam MS, Rao MVC. On the performance of the particle swarm optimization algorithm with various inertia weight variants for computing optimal control of a class of hybrid systems. Discret Dyn Nat Soc. 2006;2006:1–17.
5. Moré JJ, Sorensen DC. Computing a Trust Region Step. SIAM J Sci Stat Comput. 1983 Sep;4:553–72.
6. Steihaug T. The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. SIAM J Numer Anal. 1983 Jun;20:626–37.
7. Byrd RH, Schnabel RB, Shultz GA. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math Program. 1988 Jan;40–40:247–63.
8. Coleman TF, Verma A. A Preconditioned Conjugate Gradient Approach to Linear Equality Constrained Minimization. Comput Optim Appl. 2001;20:61–72.

9. Wu HC. The Karush–Kuhn–Tucker optimality conditions in multiobjective programming problems with interval-valued objective functions. Eur J Oper Res. 2008;196:49–60.
10. Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. Least Squares Support Vector Machines. River Edge, NJ: World Scientific; 2002.
11. Fletcher R. Practical Methods of Optimization. New York: John Wiley & Sons; 1987.
12. Schittkowski K. NLPQL: A fortran subroutine solving constrained nonlinear programming problems. Ann Oper Res. 1986 Jun;5:485–500.
13. Han SP. A globally convergent method for nonlinear programming. J Optim Theory Appl. 1977 Jul;22:297–309.
14. Powell MJ. A fast algorithm for nonlinearly constrained optimization calculations. In: Numerical Analysis. Berlin, Heidelberg: Springer; 1978. p. 144–57.
15. Waltz RA, Morales JL, Nocedal J, Orban D. An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math Program. 2006 Jul 25;107:391–408.