Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Methods for Efficient Solutions of Spatially Explicit Biofuels Supply Chain Models

Phuc M. Tran[ab]*, Eric G. O'Neill[ab], and Christos T. Maravelias[abc]

[a] Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08540, USA
[b] DOE Great Lakes Bioenergy Research Center
[c] Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08540, USA
* Corresponding Author: pt3594@princeton.edu.

ABSTRACT

The growing size and complexity of energy system optimization models, driven by high-resolution spatial data, pose significant computational challenges. This study introduces methods to reduce model's size and improve computational efficiency while preserving solution accuracy. First, a composite-curve-based approach is proposed to aggregate granular data into larger resolutions without averaging out specific properties. Second, a general clustering method groups geographically proximate fields, replacing multiple transportation arcs with a single arc to reduce transportation-related variables. Lastly, a two-step algorithm that decomposes the supply chain design problems into two smaller, more manageable subproblems is introduced. These methods are applied to a case study of switchgrass-to-biofuels network design in eight U.S. Midwest states, demonstrating their effectiveness with realistic and detailed spatial data.

Keywords: Energy and Sustainability, Biofuels, Computation Performance, Solution Quality, Optimization.

INTRODUCTION

As a renewable energy resource, biofuels are expected to play a vital role in the transition to a low-carbon energy system, particularly in the transportation sector [1]. Although electrification has penetrated the light vehicle fleet, electrifying the heavy-duty vehicle fleet remains quite challenging [2]. In the near future, biomass-derived fuels can help alleviate the sector's reliance on fossil fuels without significantly altering the fleet [1,2].

The limitations of first-generation biofuels, mainly their competition for land use with food crops, have placed greater attention on second-generation biofuels, which are derived from ligno-cellulosic sources [3]. These crops are attractive due to their low net carbon emissions and ability to grow on non-arable lands [4]. Also, agricultural practices associated with these crops mitigate greenhouse gases (GHG) by sequestering atmospheric CO_2 in the form of soil organic carbon (SOC) [4,5]. However, the low bulk density of biomass and its dispersed availability often lead to expensive supply chain (SC) networks [6]. Hence, determining the optimal biofuels SC network is vital to making second-generation biofuels viable, both commercially and environmentally.

Detailed SC optimization models, typically in the form of mixed-integer linear programming (MILP) models, have been developed to investigate the economic and environmental performance of large-scale biofuels SC [7,8]. Recently, significant improvements have been made in identifying and characterizing potential fields for ligno-cellulosic crop growth, providing more detailed data on their locations and properties [6]. While highly granular data can improve the solution quality of SC models, they create significant computational issues, resulting in the formulation of models with large numbers of variables and constraints, mainly related to biomass procurement and transportation [7,8].

Spatial aggregation methods that group small fields into larger representative cells have been developed to address this issue. These approaches allow variables to be utilized at a coarser resolution, reducing model complexity. Nevertheless, this practice can reduce the accuracy of the aggregated data. The properties of the aggregated cells are typically the averages of fields' properties, failing to account for any variability in field-specific properties [9]. Furthermore, cells with fixed resolutions (e.g., counties or 25m x 25km) are commonly used for network reduction [8,9]. All fields within a cell are assumed to

share the same distances to facilities. Transportation arcs from these cells are assumed to originate from the cells' centers of mass, which are determined based on fields' locations and biomass availabilities within these cells. Given the dispersed locations of these fields, cells with fixed resolutions can lead to inaccurate transportation arcs. Such inaccuracies can reduce the solution quality of SC models.

In this paper, we propose two methods to reduce the large number of procurement and transportation variables while maintaining the accuracy of SC models: composite curves and location-based clustering. We also put forth a two-step algorithm that breaks large-scale problems into two smaller subproblems, which are solved more efficiently. The remainder of this paper is structured as follows: the Background section presents the necessary background for biofuels SC models and the terminology used in the paper. The Methods section outlines the major approaches and describes the models that can be created using the proposed methods. Lastly, a case study on biofuels SC in the U.S. Midwest is presented in the Case Study section.

BACKGROUND

A biofuels SC model incorporates a wide range of activities and a complex set of parameters [9]. Depending on the scope and objective, implementation of SC models can help determine decisions at three distinct levels: long-term strategic, medium-term tactical, and short-term operational [9]. Long-term decisions involve facility location, technology selection, and capacity installation. Medium-term decisions include materials procurement and transportation. Short-term decisions pertain to production scheduling.

This work focuses mainly on SC models that consider medium-term decisions variables, which are specifically influenced by the granularity of spatial input data. For consistency, biomass procurement refers to the harvesting of crops at a specific location, while biomass transportation denotes the delivering of this material from a location to dedicated facilities. The cells used for aggregation are denoted as "harvesting sites" (HS).

Table 1: Nomenclature

Symbol	Description
Sets/Subsets	
$f \in \mathbf{F}$	Fields
$i \in \mathbf{I}$	Compounds in the network
$j \in \mathbf{J}$	Harvesting sites
$g \in \mathbf{G} \subset \mathbf{F}\mathbf{x}\mathbf{J}$	Two-dimensional set of fields f and the HS j containing them
	Feedstock (biomass crop)
$\mathbf{I^F} \subset \mathbf{I}$	All fields contained within an HS j
$\mathbf{G_j} \subset \mathbf{G}$	
Parameters	
$\alpha_{i,f}$	Crop i productivity in field f (Mg/ha)
$\iota_{i,f}$	Land available for crop i in field f (ha)
$\lambda_i^{FEED}/\lambda_i^{MAN}$	Per-mass feedstock/management cost of feedstock i ($/Mg)
$\upsilon_i^{HARV}/\upsilon_i^{MAN}$	GHG emissions from harvesting/managing i ($MgCO_2e/Mg$)/($MgCO_2e/ha$)
$\upsilon_{i,f}^{SOC}$	SOC sequestration potential from i in field f ($MgCO_2e/ha$)
ω^{SCC}	Social cost of carbon ($/$MgCO_2e$)
Variables	
$H_{i,f}/\bar{H}_{i,j}$	Procurement of i in field f/HS j (Mg)
$N_{i,f}/\bar{N}_{i,j}$	Land use for crop i in field f/HS j (ha)
$C_{i,j}^{PRO}$	Procurement cost of crop i in field f ($)
$J_{i,f}^{PRO}$	GHG emissions from procuring crop i in field f ($MgCO_2e$)
$J_{i,j}^{SOC}$	SOC sequestration from crop i in HS j ($MgCO_2e$)
C^{TRA}	Transportation cost ($)
OBJ	Objective – total supply chain cost ($)

METHODS

We present two complexity reduction methods and a solution method to solve large-scale SC models.

Composite Curves

A composite curve depicts the relationship between two decision variables within an HS, such as land use and SOC sequestration, based on an optimized field selection order. This selection order is established using a ranking metric, which can change depending on the objective of the SC model. For instance, if the objective is to minimize land utilization, the ranking metric would prioritize fields based on their biomass productivity.

This leads to a distinctive land use versus biomass procurement curve. At first, the curve has a steep slope due to the use of highly productive fields, gradually flattening as less productive fields are incorporated. Using a composite curve instead of an average productivity value for the HS allows for a more precise representation of the field selection pattern inherent to the SC models. In practice, field-level optimization aims to prioritize the most productive fields to achieve minimal land utilization. Composite curves effectively capture this behavior, ensuring the alignment between field-level decisions and the aggregated outcomes at the HS level.

While the previous example emphasizes land use, which leads to a straightforward ranking metric, SC models often involve complex objectives. Then, it is crucial to systematically analyze how the procurement process contributes to the overall objective function. This involves identifying all the parameters, variables, and constraints pertaining to procurement, and modifying them to gauge how changes in procurement affect the

objective function. This rate of change derived from the analysis effectively serves as our ranking metric. To illustrate this approach, we present a case study from Christian County in Illinois (Figure 2), with an objective to minimize the total network cost, accounting for GHG emissions through the use of social cost of carbon (SCC).

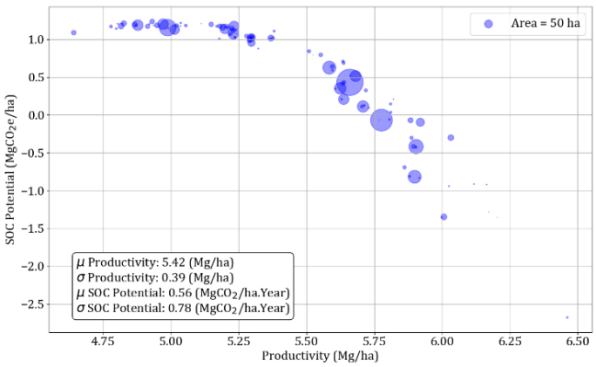

Figure 2. Properties of fields within Christian county for switchgrass

We first identify the contributions that biomass procurement makes to the overall network cost: procurement cost, $C_{i,f}^{PRO}$. We then identify all relevant constraints:

$$H_{i,f} = \alpha_{i,f} N_{i,f} \quad (1)$$

$$N_{i,f} \leq \iota_{i,f} \quad (2)$$

$$J_{i,f}^{PRO} = \upsilon_i^{HARV} H_{i,f} + \upsilon_i^{MAN} N_{i,f} - \upsilon_{i,f}^{SOC} N_{i,f} \quad (3)$$

$$C_{i,f}^{PRO} = \lambda_i^{FEED} H_{i,f} + \lambda_i^{MAN} N_{i,f} + \omega^{SCC} J_{i,f}^{PRO} \quad (4)$$

where Equation (1) relates land use decisions to biomass procurement decisions for each field. Equation (2) enforces an upper bound on land use. Equation (3) tracks the GHG emissions from biomass procurement, and Equation (4) calculates procurement costs. To determine how procurement costs scale with biomass procurement, we aim to convert all variables on the right-hand side of Equation (4) into $H_{i,f}$. We can make use of Equation (1):

$$N_{i,f} = \frac{H_{i,f}}{\alpha_{i,f}} \quad (5)$$

and substitute this expression into Equations (3) and (4):

$$C_{i,f}^{PRO} = \left[\lambda_i^{FEED} + \frac{\lambda_i^{MAN}}{\alpha_{i,f}} + \omega^{SCC}\left(\upsilon_i^{HARV} + \frac{\upsilon_i^{MAN}}{\alpha_{i,f}} - \frac{\upsilon_{i,f}^{SOC}}{\alpha_{i,f}}\right)\right] H_{i,f} \quad (6)$$

The rate of change of procurement cost per unit biomass procurement, described by the term in the square brackets of Equation 6, serves as the ranking metric. This ensures that the most cost-effective fields are prioritized. To aggregate field-specific variables, we introduce the following HS-level variables:

$$J_{i,j}^{SOC} = \sum_{f \in \mathbf{G}_j} \upsilon_{i,f}^{SOC} N_{i,f} \quad (7)$$

$$\bar{H}_{i,j} = \sum_{f \in \mathbf{G}_j} H_{i,f} \quad (8)$$

$$\bar{N}_{i,j} = \sum_{f \in \mathbf{G}_j} N_{i,f} \quad (9)$$

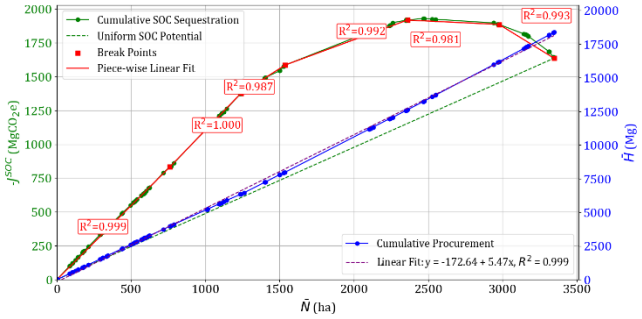

Figure 3. $\bar{N} - J^{SOC}$ and $\bar{N} - \bar{H}$ composite curves for Christian county with ω^{SCC} = \$65/MgCO₂e

Figure 3 displays a non-linear relationship between $\bar{N}_{i,j}$and $J_{i,j}^{SOC}$, and a linear one between $\bar{N}_{i,j}$ and $\bar{H}_{i,j}$. It is evident that we can use an average value in cases where the data has a narrow spread in values. However, composite curves are needed for data with high variability. In this example, composite curves are used to model $J_{i,j}^{SOC}$, while $\bar{H}_{i,j}$ can be assumed to scale linearly with $\bar{N}_{i,j}$.

To effectively approximate $\bar{N} - J^{SOC}$ composite curves while maintaining linearity in our model, we employ piecewise linearization. We develop a technique which seeks to maximize the linear fitting coefficients of determination R^2 by discretizing the $\bar{N} - J^{SOC}$ curve into consecutive linear segments. We also limit the number of segments in each curve to a maximum of six.

Location-based Clustering

We propose using location-based clustering to define HS that better capture the geographical locations of fields for network reduction. The clustering algorithm of choice in this study is the *k*-means algorithm. The *k*-means algorithm partitions a dataset (fields) into a user-defined number *K* of distinct clusters (HS) based on selected properties (longitude and latitude):

$$min \sum_{j=1}^{K} \sum_{x \in \boldsymbol{C}_j} \left|\left|x - \mu_j\right|\right|_2^2 \quad (7)$$

where *K* is the user-defined number of HS; x contains the longitude and latitude of a field; $\boldsymbol{C}_j$ is the set of fields assigned to HS j; μ_i contains the longitude and latitude of the centroid of HS j. The algorithm minimizes the within-HS variance by finding the optimal HS centroids and their member fields at the same time. With spatial coordinates as input properties, fields in the same *k*-means HS are spatially proximate, making sure that transportation arcs originating from these centroids are representative of their member fields' locations.

Because the objective function is non-convex, the algorithm is run several times with different initializations,

and the best solution is selected. The use of the *k*-means clustering for network reduction terminates the reliance on Geographical Information System (GIS) software to process geospatial data. Also, the number of HS can be easily adjusted through the parameter *K*, offering flexibility and scalability.

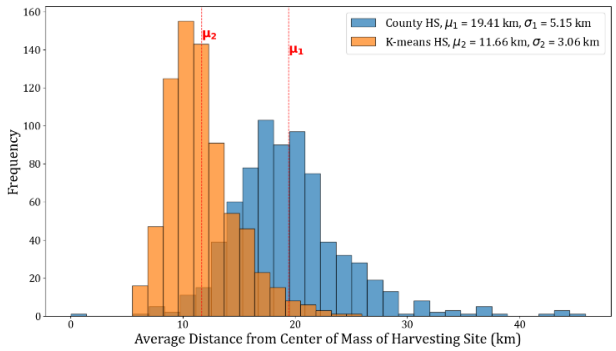

Figure 4. Distribution of mean distances between fields and the center of mass/centroid of their HS

To demonstrate, we apply the *k*-means algorithm to our case study, creating *K* = 734 HS. We select this *K* value because the fields in our dataset span across 734 counties in the study area. Figure 4 shows the statistics of the best solutions out of 10 *k*-means runs different initialization. As shown, the mean distances within *k*-means generated HS are more narrowly distributed than those of county-level HS.

Two-step Algorithm

To balance computational efficiency with spatial granularity, we introduce a two-step algorithm that effectively solves large-scale network design models with high spatial resolution. In Step 1, an MILP model that incorporates at least one of the proposed complexity reduction methods is solved to attain a "coarse" solution of the large-scale model. In Step 2, this solution is refined to a higher resolution. We develop a linear programming (LP) model that takes in the HS-level procurement decisions from the Step 1 solution as parameters and re-optimizes these decisions to a field level. This second step uses field-originating transportation arcs, enabling more precise calculation of transportation costs, which can then be used to re-evaluate the objective function values found in Step 1.

The Step 1 model can have different formulations based on the methods employed, while the Step 2 model always stays unchanged. We refer to all Step 1 models by the letter M. Superscripts denote the resolution of the biomass procurement variables: field-level (*F*) or HS-level (*HS*). Subscripts describe the complexity reduction methods used. For network reduction, we can use fixed-resolution (-) or *k*-means HS (*K*). Composite curves (CC) can also be used.

Four Step 1 models are considered in this study, each having a unique combination of the proposed methods. The baseline model, $\mathrm{M}^{F}_{-,-}$, has a field resolution and uses fixed-resolution (county-level) HS for network reduction. Models $\mathrm{M}^{HS}_{-,CC}$ and $\mathrm{M}^{F}_{K,-}$ use composite curves and *k*-means HS, respectively. Finally, model $\mathrm{M}^{HS}_{K,CC}$ combines both complexity reduction methods.

CASE STUDY

We use the four developed models to analyze a switchgrass-to-biofuels SC. The area of study includes the 8 states in the US Midwest: Ohio, Indiana, Iowa, Illinois, Michigan, Minnesota, Missouri, and Wisconsin.

All models are implemented for multiple instances. Specifiaclly, ω^{SCC} values are varied from \$0/MgCO$_2$e to \$25/MgCO$_2$e, with \$5/MgCO$_2$e increments. We also have three levels of biofuels demand, denoted as β: 500M gallons of gasoline equivalent (GGE), 800 GGE, and 1000M GGE.

We are given (i) the realistic, spatially explicit data on the biomass productivity and SOC potential of each field; (ii) potential biorefinery and preprocessing depot locations and their technology options; (iii) operating and transportation cost parameters. We also have the state-dependent prices and GHG impacts of grid electricity. Model models yield the optimal (i) fields for procurement and land utilization; (ii) facility locations, as well as their technologies and capacities (iii) transportation routing and production planning. The objective function is the minimization of total SC cost, including environmental factors, in net-present dollar amount.

All models in this study are formulated on GAMS Studio 45.7.0 and implemented with Gurobi 10.0.1.

Effects of Composite Curves

We look into how J^{SOC} values change between Step 1 and Step 2 solutions to investigate the effects of composite curves, as they are used to approximate this variable. Small differences in J^{SOC} are observed between the two steps for $\mathrm{M}^{\mathrm{HS}}_{-,CC}$ in Figure 5a, showing that composite curves are able to capture the behavior of sequestration on a field level. There are, however, notable differences in J^{SOC} values for both steps among models. This is due to the use of average productivity values for the HS in $\mathrm{M}^{HS}_{-,CC}$, which leads to different land use decisions, and thus different sequestration patterns in the models' solutions.

Despite this, $\mathrm{M}^{HS}_{-,CC}$ has great performance, as shown in Figure 6. The y-axis shows the fraction of all instances that are solved to optimality. The multiplier ε represents how much longer it takes for each model to solve with respect to the fastest one, for that instance. The performance profile of $\mathrm{M}^{HS}_{-,CC}$ starts at y=0.39, indicating that the model is the most efficient for 39% of the instances. The

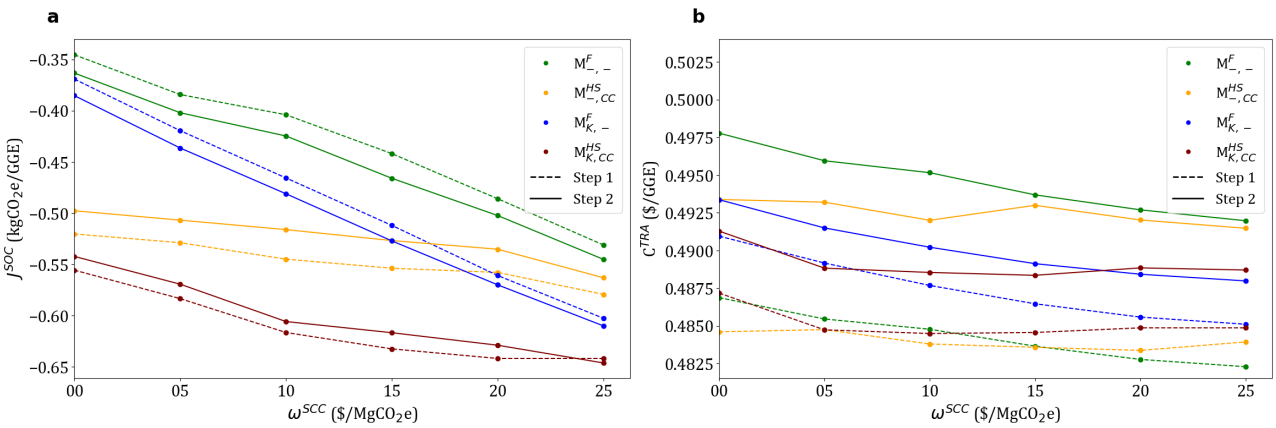

Figure 5: Results from the case study for all models, with β = 500M GGE and varying ω^{SCC}. a) SOC sequestration, b) transportation cost, normalized to a per GGE basis.

profile also reaches y=1.0 within a small range of ε, meaning that, for the instances that $M^{HS}_{-,CC}$ is not most efficient model, its run times are not significantly longer than those of the fastest model. Overall, composite curves allow us to aggregated field-level variables into HS-level ones in model $M^{HS}_{-,CC}$ without severely degrading solution quality, leading to improved run times compared to models with field-level variables.

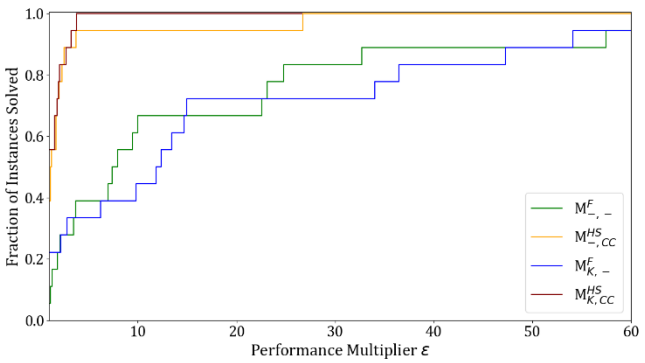

Figure 6. Performance chart of all models, with 18 total instances

Effects of Location-based Clustering

For all instances, Figure 5b showcases that $M^{F}_{K,-}$ has significantly lower C^{TRA} differences between Steps 1 and 2 than those of $M^{F}_{-,-}$ and $M^{HS}_{-,CC}$, which use county-level HS. These lower C^{TRA} differences influence the objective in a meaningful way. For most instances, although $M^{F}_{K,-}$ finds worse OBJ values than $M^{F}_{-,-}$ does, Step 2 solutions then reveal that $M^{F}_{K,-}$, in reality, attains better objective function values when C^{TRA} changes are accounted for. This can lead to differences in optimal configurations. One such instance is where ω^{SCC}= 15/MgCO$_2$e and β = 800M GGE. Here, $M^{F}_{K,-}$ finds an optimal configuration that contains one more biorefinery than that of $M^{F}_{-,-}$ (circled in red on Figure 7). This result emphasizes the effects of using location-based clustering to improve the quality and accuracy of model solutions.

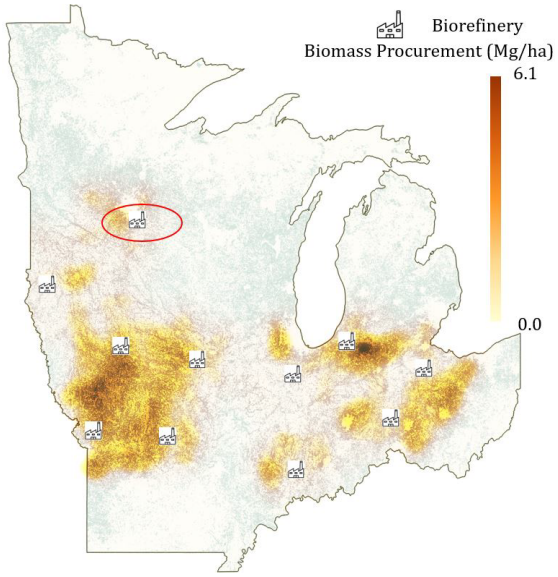

Figure 7. Optimal configuration of $M^{F}_{K,-}$ for instance ω^{SCC} = $15/MgCO$_2$e and β = 800M GGE. All refineries have pyrolysis with hydrogen purchase as the techonology

Combined Effects

Figure 5a shows that $M^{HS}_{K,CC}$ has small changes in J^{SOC} between Steps 1 and 2, illustrating the ability to capture within-HS procurement patterns of composite curves. Figure 5b exhibits similarly small C^{TRA} differences between $M^{HS}_{K,CC}$ and $M^{F}_{K,-}$, highlighting the importance of *k*-means clustering. Also, $M^{HS}_{K,CC}$ displays short run times that are characteristic of composite curves. Figure 6 shows comparably fast performance between $M^{HS}_{K,CC}$ and $M^{HS}_{-,CC}$. In fact, $M^{HS}_{K,CC}$ slightly outperforms $M^{HS}_{-,CC}$. It is evident that $M^{HS}_{K,CC}$ inherits the characteristics of both

complexity reduction methods. The model maintains the field-level accuracy while using HS-level variables and having improved run times through the use of composite curves. Moreover, $M_{K,CC}^{HS}$ reduces transportation cost changes between Steps 1 and 2 through the use of *k*-means clustering. The combination of both complexity reduction methods enables $M_{K,CC}^{HS}$ to find high-quality solutions within a short amount of time.

CONCLUSION

We investigated the use of two complexity reduction methods and a solution method, with the goal to increase the performance as well as accuracy of large-scale, spatially explicit SC models in this paper. We first introduced a composite-curves based approach, which creates characteristic curves for HS-level decision variables according to a specific ranking metric for fields. We also presented a framework to determine this ranking metric. Second, we utilized location-based network reduction, which reduces transportation-related variables by assuming the same transportation arc from multiple fields to dedicated facilities. The accurate clustering of fields for network reduction was achieved via the *k*-means algorithm, leading to more realistic transportation arcs. Lastly, a two-step algorithm was designed to decompose large-scale SC models into two more manageable sub-problems. These sub-problems can be solved sequentially and more efficiently. Lastly, we put forth a switchgrass-to-biofuels SC case study in the U.S. Midwest to probe how the proposed complexity reduction methods influence the performance and solution quality of models. Our results suggested that composite curves effectively improve the performance of models while successfully capturing the within-HS fields selection pattern, and *k*- means clustering improves models' accuracy in establishing representative transportation arcs for groups of fields. The combination of both complexity reduction methods led to better objective function value as well as more efficient performance. These ideas can be utilized in SC models that deal with large spatial datasets in order to increase their performance without trading off the accuracy of solutions.

DIGITAL SUPPLEMENTARY MATERIAL

Additional description of proposed methods, formulation of Step 1 and Step 2 models, parameters and results of the case study can be found at:
https://psecommunity.org/LAPSE:2025.0029

ACKNOWLEDGEMENTS

This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-SC0018409.

REFERENCES

1. Demirbas A. Importance of biodiesel as transportation fuel. Energy Policy 2007;35:4661–70. https://doi.org/10.1016/j.enpol.2007.04.003
2. Geissler CH, Ryu J, and Maravelias CT. The future of biofuels in the United States transportation sector. Renewable and Sustainable Energy Reviews 2024;192:114276. https://doi.org/10.1016/j.rser.2023.114276
3. Naik SN, Goud VV, Rout PK, and Dalai AK. Production of first and second generation biofuels: A comprehensive review. Renewable and Sustainable Energy Reviews 2010;14:578–97. https://doi.org/10.1016/j.rser.2009.10.003
4. Cai X, Zhang X, and Wang D. Land availability for biofuel production. Environmental science & technology 2011;45:334–9. https://doi.org/10.1021/es103338e
5. Martinez-Feria R and Basso B. Predicting soil carbon changes in switchgrass grown on marginal lands under climate change and adaptation strategies. GCB Bioenergy 2020;12:742–55. https://doi.org/10.1111/gcbb.12726
6. Mielenz JR. Ethanol production from biomass: technology and commercialization status. Current Opinion in Microbiology 2001;4:324–9. https://doi.org/10.1016/S1369-5274(00)00211-3
7. He-Lambert L, English BC, Lambert DM, et al. Determining a geographic high resolution supply chain network for a large scale biofuel industry. Applied Energy 2018;218:266–81.
8. O'Neill EG, Martinez-Feria RA, Basso B, and Maravelias CT. Integrated spatially explicit landscape and cellulosic biofuel supply chain optimization under biomass yield uncertainty. Computers & Chemical Engineering 2022;160:107724.https://doi.org/10.1016/j.compchemeng.2022.107724
9. Ng RT, Kurniawan D, Wang H, Mariska B, Wu W, and Maravelias CT. Integrated framework for designing spatially explicit biofuel supply chains. Applied Energy 2018;216:116–31. https://doi.org/10.1016/j.apenergy.2018.02.077