CyberGIS-enabled decision support platform for biomass supply chain optimization

Tao Lin, Shaowen Wang, Luis F. Rodríguez, Hao Hu, and Yan Liu

Author	Contact Information
Tao Lin Corresponding Author	University of Illinois at Urbana-Champaign CyberGIS Center for Advanced Digital and Spatial Studies Department of Agricultural & Biological Engineering 3102B NCSA Building 1205 W. Clark St. Urbana, IL 61801 Phone: +1-217-819-2999 Fax: +1-217-244-0323 taolin1@illinois.edu
Shaowen Wang	University of Illinois at Urbana-Champaign CyberGIS Center for Advanced Digital and Spatial Studies Department of Geography and Geographic Information Science CyberInfrastructure and Geospatial Information Laboratory Room 265, Computing Applications Building 605 East Springfield Avenue Champaign, IL 61820 shaowen@illinois.edu
Luis F. Rodríguez	University of Illinois at Urbana-Champaign Department of Agricultural & Biological Engineering 376C Agricultural Engineering Sciences Building, MC-644 1304 West Pennsylvania Avenue Urbana, IL 61801 lfr@illinois.edu
Hao Hu	University of Illinois at Urbana-Champaign Department of Geography and Geographic Information Science CyberInfrastructure and Geospatial Information Laboratory Room 267, Computing Applications Building 605 East Springfield Avenue Champaign, IL 61820 haohu3@illinois.edu
Yan Liu	University of Illinois at Urbana-Champaign CyberGIS Center for Advanced Digital and Spatial Studies National Center for Supercomputing Applications (NCSA) 3102D NCSA Building 1205 W. Clark St. Urbana, IL 61801 yanliu@illinois.edu

Abstract. *Biomass supply chain optimization aims to facilitate large-scale production of biofuels by improving the efficiency and effectiveness of biomass feedstock provision. Most existing models are not web based, limited by the accessibility for real-world applications. A CyberGIS-enabled biomass supply chain decision support platform was developed to improve model accessibility and computational performance. The platform includes four major components: BioScope optimization model, GISolve middleware, high-performance cyberinfrastructure, and an interactive web interface. The workflow and functions of each component are provided to illustrate the development and usage of the platform. Case studies and associated system performance have been evaluated to demonstrate the utility of the CyberGIS-enabled decision support platform. Through leveraging cyberinfrastructure resources and interactive web-based interface, the platform enables solving complex biomass supply chain optimization problems. The improved computational performance could provide responsive decision support for group-based applications.*

Keywords. Biomass supply chain, cyberinfrastructure, CyberGIS, decision support, GIS, optimization

1. Introduction

Biofuels are considered an important component of renewable energy, which could improve energy independence, reduce greenhouse gas emissions, and spur rural economics. Dedicated energy crops (i.e. Miscanthus and switchgrass), agricultural residuals (i.e. corn stover and wheat straw), and wood residuals are considered as potential cellulosic based biomass feedstock. Given its direct impact on agricultural land use, the search for beneficial biofuels should consider energy, environmental, and food security perspectives (Tilman et al., 2009). The land use changes (Havlik et al., 2011), biodiversity (Groom et al., 2008), and water usage (De Fraiture et al., 2008) have been considered major factors for environmental impact of biofuel production. A sustainable large-scale biofuel production, therefore, should integrate engineering/technology development, economic, and environmental considerations.

A feature of biomass-biofuel system is that biomass feedstock is reproducible, distributed, and has low energy and bulk density as compared to petroleum based energy. To enable large-scale biofuel production, a typical biomass supply chain system includes three major stages, including biomass suppliers (farms), centralized storage and preprocessing (CSP) facilities, and biorefineries (Figure 1). Biomass feedstocks are collected and baled from farm fields and are transported by trucks to a CSP facility for storage and preprocessing operations including size reduction and densification. The preprocessed biomass feedstocks are further transported either by trucks or rail cars to biorefineries for ethanol production. How to design supply chain configurations and supply and transport biomass efficiently and effectively is one of major challenges for large-scale biofuels production.

Biomass supply chain optimization facilitates the development of large-scale biomass provision systems, from biomass supply to biofuel production. Recently, several biomass supply

chain optimization and simulation models have been developed to minimize biomass production costs by determining optimal supply chain network designs (Panichelli and Gnanasounou, 2008; Kim et al., 2011; Delzeit et al., 2012; Lin et al., 2013) as well as biomass production and delivery schedules (Shastri et al., 2010; Zhang et al., 2013). Most existing optimization models have been developed using the mixed integer linear programming, which requires significant computational resources for large-scale problem solving. Several studies have been proposed to develop and apply approximation methods to improve computational efficiency and scalability, including Lagrangean relaxation (Fisher, 1985), Lagrangean decomposition (Chen and Pinto, 2008), and bi-level decomposition (You et al., 2010). However, the reduction of computational time by using approximation methods occurs at the expense of result quality. Moreover, most optimization models are developed based on sequential computing and for single user access, not for online sharing, which limits the model accessibility.

Figure 1: A typical three-stage biomass supply chain system includes biomass supplies (farms), centralized storage and preprocessing (CSP) facilities, and biorefineries.

Visualization is another major challenge for interpreting and understanding results of complex spatial decision-making problems. Biomass production is spatially explicit, as influenced by weather and soil differences. Therefore, spatial visualization facilitates the understanding of biomass supply chain analysis. Several studies have applied Geographic Information Systems (GIS) to manage biomass availability (Fiorese and Guariso, 2010) and production costs (Khanna et al., 2008; Jain et al., 2010), and visualize numerical results from optimization models on maps (Tittmann et al., 2010).

Spatial decision support systems (SDSS) have been established for various application in renewable energy development, ranging from biomass resource evaluation (Frombo et al., 2009; Zambelli et al., 2012) to multi-criteria infrastructure planning (Aydin et al., 2010; Perimenis et al., 2011). However, most of these SDSS are not web-based, not feasible for group-based application. To improve model accessibility and result visualization and sharing, few web-based spatial decision support platforms have been reported in literature for renewable energy development, mainly focusing on strategic planning of wind farms (Simao et al., 2009; Mari et al., 2011). However, these two platforms are criteria evaluation based planning, without any optimization modelling that requires significant computing resources of processing power and memory.

Furthermore, how to enable multiple users to access the application simultaneously and provide sufficient computational resources to complete their job submissions are other challenges in terms of model accessibility and throughput performance. It is expected that model accessibility and throughput as well as visualization and sharing of modeling results are key limiting factors for large-group decision-making applications. Cyberinfrastructure (CI) is designed to integrate computing systems, data storage systems, advanced instruments, and

visualization environments together by software and high performance networks to facilitate complex problem solving (Stewart et al., 2010). GIS based on advanced CI – defined as CyberGIS – provides a seamless integration of CI, GIS, and spatial modeling analysis, which is becoming important to enable complex problem solving, model accessibility, and visualization capabilities (Wang et al., 2012). The CyberGIS approach to the integration of computing, data, and visualization capabilities represents an innovative CI environment. The CyberGIS software environment has three interrelated pillars: 1) CyberGIS Gateway for a large number users to access online CyberGIS capabilities, 2) CyberGIS Toolkit for distributing open source and scalable modules, and 3) GISolve middleware for integrating CI and CyberGIS capabilities (Liu et al., 2015; Wang et al., 2013). The GISolve middleware is an important element of CyberGIS software, which is composed of service-oriented GIS components, spatial middleware, a suite of parallel and distributed GIS algorithms, and a set of user-interface and collaboration services (Wang, 2010). The development of GISolve middleware is critical to support distributed system management that can utilize multiple computing resources without physical boundaries and also support heterogeneous system management for different operating systems and resource management tools. GISolve has been applied to various research areas, including for example analysis of climate change impact (Wang and Zhu, 2008) and ecological modeling (Tang et al., 2011), and public health (Padmanabhan et al., 2014).

In order to improve model accessibility and computational performance for complex optimization problem solving, there is a need to develop an integrated system 1) to enable complex optimization modelling through the support of high-performance distributed computing infrastructure; 2) to integrate optimization tools with CyberGIS to improve model accessibility and understanding of spatial relationships embedded in modeling results through intuitive online

visualization. The proposed CyberGIS-enabled decision support platform is designed to achieve the following user-centered designs: 1) allow users to evaluate the system performance based on their analysis scope; 2) manage historical scenario analyses; 3) provide spatial visualization tools for better result interpretation; 4) improve computing efficiency by reducing computing time for both single job and multiple job submissions.

The objective of this study is to describe how to integrate optimization based modelling with the CyberGIS platform to enable complex spatial decision making for biomass feedstock provisions. The content of this study is organized as follows: Section Two describes the components of the BioScope model. Section Three discusses the existing limitations for BioScope model application. Section Four describes the components and workflow of the CyberGIS-enabled BioScope model application. Section Five presents the case studies and performance evaluation of the CyberGIS-BioScope application. The summary of advantages, limitations, and future work for the platform application is presented in Section Six.

2. Overview of BioScope Model

The BioScope model was developed to minimize annual biomass-ethanol production costs by optimizing long-term strategic decisions (Lin et al., 2013). The model was applied to analyze a three-stage biomass supply chain system that includes biomass suppliers (farms), centralized storage and preprocessing (CSP) facilities, and biorefineries. The key feature of the model is its integration with detailed spatial information to quantify the key tradeoffs between economies of scales for bioprocessing facilities and transportation costs for various biomass supply chain configurations. The objective of the optimization model is to minimize annual biomass-ethanol production costs (Z) that are comprised of four costs: biomass purchase costs (C_B), transportation related costs (C_T), CSP site related costs (C_S), and biorefinery related costs (C_E) (Eq. 1).

$$\text{Minimize } Z = C_B + C_T + C_S + C_E \quad (1)$$

The BioScope model incorporates there major groups of input data, including spatial, agricultural, and engineering/technology related data (Figure 2). The spatial data includes candidate facility locations for each stage of the system and the shortest transportation distances among these candidate locations. The agricultural related data includes crop yield (Mg per hectare), production costs (\$ per Mg of biomass), and cropland area used for biomass production (hectare). The engineering/technology data includes transportation costs, and operating and capital costs for different bioprocessing technologies. The key decisions include the optimal number, location, and capacity of biomass supply sites, centralized storage and preprocessing sites, and biorefineries and optimal biomass flow patterns within the system. The detailed illustration of nomenclature, constraint equations, and objective functions were reported in Lin et al. (2013).

Figure 2: The components of the BioScope model and their data flow. The rectangles represent source data (biomass harvesting, preprocessing, and storage, land opportunity costs, county level farm size and cropland area, county level biomass yield, road networks). The rounded rectangles represent model outputs.

The BioScope model is a mixed integer linear programming model that was developed on the GAMS platform using the CPLEX solver (GAMS, 2013). The BioScope model is a general mathematical programming model that is developed to quantify and optimize a biomass supply chain system for different crops in various geospatial areas using different transportation modes. By incorporating required input data based on the analysis scope, the model can optimize system configurations from county, state, regional, to national scales. The model complexity, however, increases significantly with the analysis scope. GIS was used to store, manage, and retrieve geospatial information, including county-level biomass availability, biomass production costs, and transportation distances between facilities in the supply chain, and to visualize the optimized supply chain configuration on maps.

3. Challenges of Existing BioScope Model Application

Data collection, mathematic modeling and analysis, and result visualization comprise the three major steps of the current workflow of the BioScope model (Figure 3). The input data are saved into multiple common-separated value (csv) files that can be read by the model to instantiate parameters and constraints for analyses. The numerical results from the model will be exported into csv files. ArcGIS software (ESRI, 2013) will be further applied to correlate numerical results with the corresponding spatial information and visualize the results on maps.

Figure 3: The workflow of the BioScope model

The major challenge of the current model workflow is that the linkages among data collection, modeling and analysis, and result visualization have not been well integrated, which requires certain knowledge and efforts on data transferring, processing, and spatial visualization. Without a seamless integrated workflow, any responsive decision support cannot be achieved.

In addition to the lack of an integrated workflow for problem solving and result visualization, the application of the Bioscope model has three major limitations: 1) significant computing time and memory requirement for optimization problem solving, 2) limited model accessibility, and 3) lack of high throughput computational capability for scenario analyses by a large group of users.

The BioScope model is based on mixed integer programming, which is classified as an *NP*-Hard problem. *NP*-Hard problems cannot be solved in polynomial time based on the existing algorithms. Significant computational challenges could be foreseen for complex problem solving. Large-scale analysis increases the computational intensity of model runs dramatically due to the increase of problem sizes. How to design and implement scalable computing approaches to exploit advanced cyberinfrastructure (CI) resources would be critical to reduce computational challenges.

The source code of the BioScope model currently can only be executed through the GAMS software. Therefore, the application of the model requires the installation of the GAMS software and ability with GAMS programming. Lack of graphic user interface on the GAMS platform further constraints the application accessibility of the model.

Our previous study showed that an optimal biomass supply chain configuration is highly dependent on several input parameters, such as cropland usage rate, biomass yield rate, and transportation cost rate (Lin et al., 2013). In order to provide a comprehensive decision support

for a new biomass supply chain configuration, many scenario analyses should be conducted to evaluate the system performance under various conditions. The uncertainty analysis centered on the BioScope model may generate a large number of computing jobs, each conducting a complete computation process of a series of spatial processing and optimization problem solving. Currently, the model is operated on a single desktop computer where scenario analyses can only be conducted sequentially. The problem solving time for a comprehensive decision support case study would be highly dependent on the number of scenarios under consideration.

Complex decision-making is often an interactive process between decision makers and model computation, which requires results to be generated within a short amount of time (e.g., in minutes instead of hours or days). Implementation of a high-throughput problem solving workflow is effective to reduce computational time of scenario analyses of BioScope modelling, and this will be even more important to enable large group-based applications.

4. Development of CyberGIS-enabled BioScope Application

4.1. The structure of proposed systems integration approach

To facilitate large-scale complex problem solving and improve model accessibility and throughput, a CyberGIS-enabled Bioscope model application for complex spatial decision support has been developed. The application includes four major components: 1) a high performance cyberinfrastructure for complex problem solving, 2) the BioScope optimization model deployed on the cyberinfrastructure; 3) GISolve middleware that includes a set of Open Service APIs and web server PHP functions; and 4) an interactive web-based application user interface on the CyberGIS Gateway (Wang and Liu, 2009) (Figure 4).

A web-based user interface, CyberGIS Gateway, was developed to support user management, input data submission, job status updating, and geospatial visualization. The design of the web interface provides dynamic and interactive functions based on Web 2.0 and HTML5. With its connection to the GISolve middleware, the interface manages to submit user defined input parameters and files, retrieve the message of problem solving status in a real time, and provide data and result visualization requests.

Figure 4: The abstract diagram of a three-tier platform integration of web interface, middleware, and cyberinfrastructure. The BioScope model is deployed on cyberinfrastructure to leverage high-performance and high-throughput computing resources.

To better integrate web client requests with cyberinfrastructure, GISolve middleware manages the complexity of accessing low-level operating software and standards, security, and communication protocols to facilitate data interactions, communications, and sharing (Wang et al., 2005). The design of GISolve is based on an open service integration framework that supports three integration levels across the toolkit, cyberinfrastructure, and web interface. The GISolve manages data and software functions as web services and provides programming interfaces for service developers and consumers. This framework provides a unified approach to streamlining the encapsulation of CyberGIS data and functions and transparent access to the cyberinfrastructure (CI) for high-performance computation and large-scale data handling. Functions in the framework are established using representational state transfer (REST) based web services with JavaScript Object Notation (JSON) as the format of service request and response messages based on the GISolve open-service APIs. The CI provides powerful computational capabilities for large complex problem solving and data storage and handling. Supported by a queuing and resource management system, the CI provides not only high-performance computing, exploiting large numbers of processors through specialized, high-speed interconnections, but also high throughput computing, allocating multiple job submissions to available computational resources simultaneously.

In this application, GAMS based BioScope model has been implemented in the CI and integrated with the GISolve middle and computing resource management tools. High-throughput computing capabilities have been enabled to support multiple jobs running simultaneously with the support of computing resource management. Other GAMS based model or even different optimization solvers can be installed and integrated with the GISolve middleware and cyberinfrastructure to facilitate various optimization problem solving.

4.2. Workflow of an integrated platform development

The integration of web interface, middleware, model, and cyberinfrastructure requires a systematic workflow approach. A typical workflow includes three major steps: job submission, job status query, and results management, where the realization of each step requires an integration across the three layers of the system (Figure 5).

Figure 5: A workflow diagram of the integrated platform.

The proposed BioScope application was implemented by integrating the following components: a dynamic CyberGIS Gateway web interface was developed using Sencha Ext JS, a JavaScript based web platform development framework. AJAX was applied to create asynchronous web applications that communicate with a server in the background, without interfering with the current stage of the Web page. In the current AJAX implementation, JSON format was used for the interchange of data. OpenLayers, an open source JavaScript library for displaying map data, and Keyhole Markup Language (KML), an XML notation for expressing geographic annotation, were applied to visualize geospatial related input and output information on web-based maps.

To operate the system, the user first needs to register to the system to get an appropriate user privilege for application access. Once the user logs in to the BioScope application online environment, the web-based biomass resource management interface will be presented to illustrate the biomass production performance (Figure 6). The interface provides county-level data on biomass yield, biomass production cost, and cropland area.

Figure 6: The user interface of the CyberGIS-enabled decision support platform.

In the job panel of the web interface, the user can view previously submitted job information (Figure 7). By clicking the “New Job” tab, the user can create a new analysis scenario. A new job submission consists of four major stages, including the selection of candidate biomass supply locations, CSP locations, and biorefinery locations, and the identification of the key input parameters. The detailed information of job submission will be illustrated in the case studies section. Furthermore, users can click the “Help” Tab (Figure 7) to view a wiki-based help document to get familiar with the detailed usage of CyberGIS-BioScope application.

The screenshot shows the 'BioScope Wizard' interface with three tabs: 'tao'jobs', 'New Job', and 'Help'. Below the tabs are 'Delete' and 'Refresh' buttons. A table displays the following job submissions:

Name	Time created	Status
IL-2milliongallon	2014/8/11 14:38:24
my_bioscope	2014/8/11 12:28:27
test2	2014/8/10 18:20:11
test3	2014/8/9 15:19:48
test2	2014/8/9 9:43:26
test11	2014/8/8 15:29:32
david1	2014/3/3 14:31:17

Figure 7: The job panel of the CyberGIS-enabled decision support platform.

After completing the data submission, GISolve middleware will retrieve these data and notify the user of successful job submission. The selected input data and files will then be converted and packaged as public URLs. The middleware will further submit a job request to the CI through a JSON format that includes the data source URLs. Once receiving the job

submission request, the computing platform retrieves all the necessary data and files from the web server based on the URLs provided in the JSON message.

Once the job is submitted, the interface automatically changes to a job status panel, where the user can view the current job status. The current system provides seven different status to users (Figure 8). Each status is visualized as an animation picture on the interface for better understanding. The status is updated in a real time manner through the linkage between interface, middleware, and CI. Furthermore, in the job status panel, users can also review the status of all job submissions (Figure 7).

Icon	Status	Icon	Status
Analysis submitted for computation		Results created
Analysis waiting to be executed		Results ready for visualization
Analysis running	No Icon	Computation problem
Computation done		

Figure 8: The description of status icon for visualizing the current job status.

GISolve middleware schedules jobs based on the first-in first-out principle while sending status requests to CI resources to check job running status. Once a job is completed on CI, GISolve automatically transfers the results from CI and processes the numerical results for geospatial visualization. Currently, optimization results are exported into csv files. By referencing to geospatial data, the selected supply counties, CSP, and bioreifnery locations and biomass transportaton patterns are converted into geospatial formats such as Keyhole Markup Language (KML). The CyberGIS gateway web interface supports the visualization of these KML files by integrating OpenLayers libraries on web browsers.

Users can visualize the results from previously completed jobs by double clicking the selected job name. Each completed job has five key result layers including optimized biomass supply counties, CSP facilities, biorefinery facilities, and biomass transportation patterns from supply counties to CSPs and from CSPs to biorefineries. Users can check the box next to the layers on the web interface to see the optimization model results on a web-based map. The web-based map visualization is interactive, where detailed information such as facility capacity and transportation flow amount can be displayed in pop-up windows.

5. Case studies of the BioScope application on the CyberGIS platform

5.1. User experience of the BioScope application

Based on the seamless integration of web-interface, middleware, cyberinfrastructure, and BioScope modelling, users can easily get access to the BioScope application for case studies. The BioScope model requires seven key input files that include candidate locations of biomass supply counties, centralized storage and preprocessing facilities, and biorefineries, county-level cropland area, biomass yield, biomass production costs, and transportation distances. Furthermore, biomass demand rate, cropland usage rate, biomass yield rate, biomass-ethanol conversion rate, and transportation cost rate are five key parameters defined for each scenario analysis.

The existing platform has already incorporated county-level data of *Miscanthus* yield (Miguez et al., 2012), cropland area (USDA, 2011), and *Miscanthus* production costs across the contiguous United States, including 48 states that are composed of 3109 counties (Figure 6).

These data can be displayed on the web-based map by selecting different layers. The road transportation distances among each county seat have also been provided in the platform.

The job submission panel consists of four major steps that include the selection of candidate biomass supply locations, candidate CSP locations, candidate biorefinery locations, and the identification of key model input parameters. The platform provides an interactive web-based map for users to identify the analysis region by selecting candidate locations. The selection can be accomplished by a rectangle box selection or by clicking the candidate county directly from the map. The selected areas for analysis will be highlighted on the web-based map (Figure 9). Once the analysis region is selected, the associated data relate to the selected counties (i.e. biomass yield, biomass production cost, and cropland area) will be displayed in the table on the left-side window. The users have the authority to review the data provided by the platform and change the value based on their own understanding.

Figure 9: An interactive data and parameter panel for analysis region selection. The areas highlighted in blue have been selected as the candidate biomass supply areas. The left panel provides the attribute information (biomass yield, biomass production cost, and cropland area) for these selected counties.

In this case study, we chose to implement the BioScope application to analyze a Miscanthus-ethanol supply chain in Illinois. 102 counties in Illinois have been selected as candidate locations for biomass supplies, centralized storage and preprocessing facilities (CSP), and biorefineries. In the current application, we aggregated different biomass suppliers within each county to its county seat as a representation, given its better access to the transportation infrastructure. The shortest transportation distances among these candidate locations have been provided from the platform.

Furthermore, the input parameters can be determined by users based on their own case designs. The annual ethanol production demand sets the total biorefinery production capacity while different conversion technologies generate various ethanol conversion rates. Considering the ethanol demand and conversion rate, the annual biomass demand can be finalized. Cropland usage rate means how much percentage of the cropland in each county is allocated to growing biomass. Biomass yield rate is the percentage of the projected biomass yield data by Míguez et al. (2012). In this case study, we chose to analyze a system with an annual ethanol demand of 45,400,000 L y⁻¹ (12,000,000 gal y⁻¹) when cropland usage rate is 5% and biomass yield rate is 100%. The default values of biomass-ethanol conversion rate (80 Mg ha⁻¹) and transportation cost rate (100%) have been selected. The mixed integer linear programming problem for Illinois case study has a total of 13,064 constraint equations and 26,625 decision variables, including 816 binary variables.

The results showed that the best supply chain configuration in this scenario requires eight counties in southern Illinois as biomass suppliers, five centralized storage and preprocessing (CSP) facilities, and one centralized biorefinery (Figure 10). Each colored area represents the biomass supply counties sharing the same CSP facility. That being said, the five CSP facilities

have five different colored regions to source biomass feedstocks. In the current scenario, three green-colored counties will ship biomass to the CSP facility that is located in the same colored region. The single centralized biorefinery is located in the middle of the proposed supply area, the same location as one CSP facility, as a result of the savings of transportation costs.

Figure 10: Results visualization of a case study in Illinois by CyberGIS-BioScope application. The colored areas represent biomass supply counties where each colored area represents the supply counties sharing the same CSP facility. The green icons represent the proposed locations of CSP facilities, and the red icon represents the proposed location of the biorefinery facility. The red lines represent the biomass transportation routes from suppliers to CSP facilities, and the blue lines represent the routes from CSP facilities to biorefineries.

The capital investment costs of such a large biorefinery establishment are \$793 million dollars, where the annualized biorefinery capital related costs account for \$109 million dollars. The unit biomass-ethanol production costs are \$0.68 per liter of ethanol produced, where the biorefinery related costs account for almost half of the costs, at \$0.34 L⁻¹. The significant costs of biorefinery establishment corresponding to the optimal supply chain design, which prefers to build a centralized facility to achieve the economies of scale. A pie chart of cost breakdown is

provided in the right column panel (Figure 10). Further detailed information has also provided in the right panel to display the results on transportation flow patterns and facility capacities.

5.2. Scenario analysis of the Illinois case studies

A comprehensive decision support of biomass supply chain optimization should be provided with the consideration of multiple scenario analyses, given the uncertainty of biomass supply and demand. Different decision makers would consider different scenarios based on their own understanding of the system, such as the estimation of cropland usage and biomass yield. Our decision support platform aims to support multi-user applications. Leveraging the advanced cyberinfrastructure resources, Monte-Carlo based scenario and sensitivity analysis can be implemented effectively.

This analysis aims to quantify the sensitivity analysis of biomass supply chain optimization in Illinois through considering ten levels of annual ethanol demand ranged from 333,000,000 to 605,600,000 L y⁻¹, three levels of cropland usage rate at 4%, 5%, and 6%, and two levels of Miscanthus yield rate at 80% and 100% of the projected yield. The default values of biomass-ethanol conversion rate (302.8 L Mg⁻¹) and transportation cost rate have been applied in these scenarios. A total of 60 scenarios have been identified for this case study.

The results showed that different scenarios require various optimal biomass-ethanol supply chain configurations. All these optimal results from 60 scenarios chose to build a single centralized biorefinery to achieve the economies of scale. To support that biorefinery operation, the optimal Miscanthus-ethanol system requires various plans of supply chain configurations in terms of numbers and locations of biomass supply counties and CSP facilities. The optimal number of biomass supply counties range from 5 to 19 counties, while the number of CSP facilities range from 1 to 5 in various scenarios (Figure 11). The number of biomass supply

counties required is largely dependent on the changes of biomass demand. High biomass demand increases the required biomass supply areas. The changes of CSP facilities depend on the changes of both supply and demand ends, because the purpose of building CSP facilities is to provide a buffer to balance the tradeoffs between economic scales and transportation costs. With high biomass demand and low biomass yield and cropland usage rate, the system prefers to build more CSP facilities. These CSP facilities have been established in a distributed pattern to support centralized biorefinery operations.

Given different scenarios, the biomass-ethanol production costs in Illinois vary from \$0.68 to \$0.73 per liter of ethanol production (Figure 12). To meet a certain ethanol production demand, the lower cropland usage rate and Miscanthus yield, the larger biomass sourcing area required, and therefore the higher biomass transportation and ethanol production costs.

Figure 11: The histogram of selected optimal biomass supply counties and centralized storage and preprocessing (CSP) facilities of 60 scenario analyses for Illinois

Figure 12: The histogram of biomass-ethanol production costs of 60 scenario analyses in Illinois

5.3. Computational performance evaluations of scenario analyses

How to leverage available computing capabilities to efficiently support decision making for a group of users is critical for any decision support platform development. To illustrate the differences of computing performance between a desktop application and CyberGIS-enabled application, the system evaluation was conducted by adjusting the number of available computing cores from 1 to 128 to solve the same 60 scenario analyses described previously for Illinois case studies. All 60 jobs would be executed sequentially if there is only one core available, which requires 4411 seconds for the total computing time. With more available computing cores, multiple scenario analyses could be conducted simultaneously, rather than solving the problems sequentially. The computing time of these 60 scenario analyses could be reduced significantly to 191 seconds by allocating more than 64 cores (Figure 13). Once the number of available cores exceeds the number of submitted jobs, 60 in the current test, all the

jobs can be executed simultaneously. The overall turnaround time for these 60 jobs was minimized to the execution time of the longest job.

Figure 13: Computational performance evaluation for 60 scenario analyses by the BioScope model application on the CyberGIS platform.

The experiment described above shows that even given sufficient computing cores, the total computing time for a group of job submissions, therefore, is constrained by the most difficult scenario analysis. The computing time of each individual scenario analysis ranged from seven seconds, when ethanol demand is 333,000,000 L y⁻¹ at 6% cropland usage rate and 100% Miscanthus yield rate, to 191 seconds when ethanol demand is 605,600,000 L y⁻¹ at 4% cropland usage rate and 80% Miscanthus yield rate. The complexity of the problem solving is increased with higher Miscanthus demand and lower cropland usage and Miscanthus yield rate. If the most complex scenario was submitted by 60 different users simultaneously, one user might have to wait more than 10,000 seconds to retrieve the results if these jobs were handled sequentially.

By leveraging the multi-threaded computing capability for a single problem-solving, the execution time of a single job can be further reduced. The CPLEX solver recently provided a multi-threaded approach to improve computational efficiency for complex problem solving

(GAMS, 2013). Experiments showed that four-threaded computing could reduce the computing time for the most difficult scenario to 95 seconds. When we allocated 128 available cores for the computing of the whole set of 60 jobs, each job execution using four threads, the system reduces not only the single job computing time, but also the total computing time for these 60 different jobs to 121 seconds. It is expected that if there exist sufficient computing cores, multi-threaded computing would further reduce the overall turnaround time for a group of job submissions.

Supported by the high performance cyberinfrastructure, CyberGIS decision support platform could solve multiple jobs simultaneously, enabling decision support for group-based application responsively. The computing capability of the CyberGIS platform can be scaled up by incorporating more computing nodes to resolve the computing challenges in terms of complex problem solving and a large number of job submissions. It is suggested the integration of high-performance cyberinfrastructure and optimization problem solving is critical to improving the computing performance of multiple scenario analyses, especially for strategic planning issues which requires considerations of multiple aspects and involves different decision makers. Given the capabilities of solving complex optimization problems responsively and visualizing the results on web-based maps, CyberGIS platform provides the possibility for future web-based collaborative research development and results sharing.

5.4. User defined case study analysis

The current version of CyberGIS-BioScope application provides a user-friendly interface to allow decision makers to select their analysis scope across the contiguous United States. The analysis scope ranges from several counties within a state boundary to regional scale that includes multiple states, based on the interest of analysis. Given the analysis scope set up by the users, the decision support application could provide responsive optimization solutions at various

planning scales. An example is provided to demonstrate a multi-state biomass-biofuel supply chain analysis. The optimized supply chain configuration includes 66 supply counties from three states, 12 CSP facilities, and two biorefineries (Figure 14). The platform currently provides county-level data for Miscanthus. The provided data on biomass costs and yield can be also modified by users based on their own understanding and knowledge. Other biomass feedstocks such as agricultural and forest residuals will be incorporated into the platform database in a near future. The platform provides the uploading function to enable users to upload their own data sources for their analysis.

Figure 14: A case study of multi-state biomass-biofuel supply chain optimization.

6. Platform Advantages, Limitations, and Future Work

The CyberGIS-BioScope application integrates BioScope optimization modeling and CyberGIS environment to provide decision support on biomass-to-biofuel supply chain management. It can be applied to analyze the system with different biomass feedstocks using different transportation

modes in different regions. The key advantages of the CyberGIS-BioScope application are three-fold: 1) support complex problem solving, especially for a large-scale biomass supply chain optimization, 2) provide an integrative and interactive approach for users to optimize a proposed supply chain system and visualize the optimal results on the web-maps; and 3) exploit a relatively easy way to utilize advanced computing resources for complex problem solving and group-based applications. The developed application provides an interactive and intuitive interface that supports complex problem solving for users without any requirement or knowledge on web-interface programming or high-performance cyberinfrastructure development.

Based on the presented case studies, the computational performance can be improved significantly for Monte-Carlo based scenario analyses or sensitivity analyses for different decision makers. It is expected, however, that analyses occurring with scopes beyond the state level would significantly increase the computing complexity, given the nature of *NP*-hard for mixed integer linear programming problems. The expected computing complexity would require an advanced cyberinfrastructure to support complex problem solving and further group-based scalable decision making.

Data availability, however, is a major bottleneck for the current decision support platform, given the complexity of data required to execute the optimization model. Our existing platform provides county-level *Miscanthus* information in the continental U.S. Further collaborations with other research groups can be established to analyze different feedstocks in different areas around the world. Some potentially interesting work could include the supply chain optimization regarding how to better utilize crop residuals to produce ethanol or electricity in Brazil, China, and other countries. The type, format, and resolution of data, either vector (i.e. county) or raster (i.e. grid) based, in other countries, need to be considered in data uploading and

management development. The established CyberGIS-BioScope platform provides tremendous potential to enable this type of collaboration and provide interactive decision support.

Uncertainty analysis is a major component for any decision support system. For BioScope model application, changes of biomass demand and yield and cropland usage rates could affect the optimal supply chain decisions significantly. Providing functional modules to consider the sources of uncertainty in model inputs would improve the decision support of biomass feedstock provision management. Uncertainty analysis usually leads to a collection of results from different settings or scenarios. How to provide a web-based interactive module to summarize the results from a group of sensitivity, scenario, or uncertainty analyses, while providing the capability to visualize these results effectively on maps is essential for the future development of spatial decision support platform.

In addition to cost-benefit analysis, the platform will incorporate environmental analysis to evaluate the sustainability performance. Energy consumption, water footprint, and greenhouse gas (GHG) emissions are the three key environmental indicators representing the biomass-biofuels supply chain, from biomass production to transportation to biofuel conversion. Environmental analysis will be incorporated into the CyberGIS-BioScope application for multi-objective optimization toward sustainable biofuel development.

7. Conclusions

A CyberGIS platform for biomass supply chain decision support has been developed to improve optimization model accessibility and achieve high computational performance of complex problem solving. The platform has four major components, including the BioScope optimization model, GISolve middleware, hybrid cyberinfrastructure, and an interactive CyberGIS Gateway user interface. The workflow and functions of each component have been presented to illustrate

the development and usage of the platform. Case studies of biomass-ethanol supply chain have showed that the optimal biomass supply chain configuration varies at different scenarios, highly dependent on the biomass demand, yield, and cropland usage rate. The results of computational performance evaluations have demonstrated that the CyberGIS platform can support interactive decision-making for a group of users on complex problem solving of biomass supply chain management.

8. Acknowledgements

This work was support in part by the National Science Foundation (NSF) under Grant Number OCI-1047916 and the Energy Biosciences Institute. We thank the anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript.

9. Software Availability

The decision support platform described in this work can be accessed through CyberGIS Gateway website at <http://gateway.cigi.illinois.edu>. User registration is required to access the BioScope model application.

10. References

- Aydin, N. Y., E. Kentel, and S. Duzgun. 2010. GIS-based environmental assessment of wind energy systems for spatial planning: A case study from Western Turkey. *Renewable and Sustainable Energy Reviews* 14(1): 364-373.
- Chen, P. and J. M. Pinto. 2008. Lagrangean-based techniques for the supply chain management of flexible process networks. *Computers & Chemical Engineering* 32(11): 2505-2528.

- Delzeit, R., B. Wolfgang, and K. Holm-Muller. 2012. Modelling regional input markets with numerous processing plants: The case of green maize for biogas production in Germany. *Environmental Modelling & Software* 32: 74-84.
- De Fraiture, C., M. Giordano, and Y. Liao. 2008. Biofuels and implications for agricultural water use: blue impacts of green energy. *Water Policy* 10: 67-81.
- ESRI. 2013. ArcGIS Desktop, Release 10. Redlands, CA.: Environmental Systems Research Institute.
- Fiorese, G. and G. Guariso. 2010. A GIS-based approach to evaluate biomass potential from energy crops at regional scale. *Environmental Modelling & Software* 25(6): 702-711.
- Fisher, M. L. 1985. An applications oriented guide to Lagrangian relaxation. *Interfaces* 15(2): 10-21.
- Frombo, F., R. Minciardi, M. Robba, F. Rosso, and R. Sacile. 2009. Planning woody biomass logistics for energy production: A strategic decision model. *Biomass and Bioenergy* 33(3): 372-383.
- GAMS. 2013. GAMS for windows, Version 23.7. Washington, D.C.: GAMS Development Corporation.
- Groom, M. J., E. M. Gray, and P. A. Townsend. 2008. Biofuels and biodiversity: principles for creating better policies for biofuel production. *Conservation biology* 22(3): 602-609.
- Havlík, P., U. A. Schneider, E. Schmid, H. Böttcher, S. Fritz, R. Skalský, K. Aoki et al. 2011. Global land-use implications of first and second generation biofuel targets. *Energy Policy* 39(10): 5690-5702.

- Jain, A. K., M. Khanna, M. Erickson, and H. Huang. 2010. An integrated biogeochemical and economic analysis of bioenergy crops in the Midwestern United States. *GCB Bioenergy* 2(5): 217-234.
- Khanna, M., B. Dhungana, and J. Clifton-Brown. 2008. Costs of producing miscanthus and switchgrass for bioenergy in Illinois. *Biomass and Bioenergy* 32(6): 482-493.
- Kim, J., M. J. Realff, J. H. Lee, C. Whittaker, and L. Furtner. 2011. Design of biomass processing network for biofuel production using an MILP model. *Biomass and Bioenergy* 35(2): 853-871.
- Lin, T., L. F. Rodríguez, Y. N. Shastri, A. C. Hansen, and K. C. Ting. 2013. GIS-enabled biomass-ethanol supply chain optimization: Model development and Miscanthus application. *Biofuels, Bioproducts and Biorefining* 7(3): 314-333.
- Liu, Y.Y., Padmanabhan, A., and Wang, S. 2015. CyberGIS Gateway for enabling data-rich geospatial research and education. *Concurrency and Computation: Practice and Experience*, 27(2): 395-407.
- Mari, R., L. Bottai, C. Busillo, F. Calatrini, B. Gozzini, and G. Gualtieri. 2011. A GIS-based interactive web decision support system for planning wind farms in Tuscany (Italy). *Renewable Energy* 36(2): 754-763.
- Miguez, F. E., Maughan, M., Bollero, G. A., and Long, S. P. 2012. Modeling spatial and dynamic variation in growth, yield, and yield stability of the bioenergy crops *Miscanthus* × *giganteus* and *Panicum virgatum* across the conterminous United States. *GCB Bioenergy* 4(5): 509-520.
- Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K., and Liu, Y.Y. 2014. FluMapper: a cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and Computation: Practice and Experience*, 26(13): 2253–2265.

- Panichelli, L. and E. Gnansounou. 2008. GIS-based approach for defining bioenergy facilities location: A case study in Northern Spain based on marginal delivery costs and resources competition between facilities. *Biomass and Bioenergy* 32(4): 289-300.
- Perimenis, A., H. Walimwipi, S. Zinoviev, F. Müller-Langer, and S. Miertus. 2011. Development of a decision support tool for the assessment of biofuels. *Energy Policy* 39(3): 1782-1793.
- Shastri, Y. N., A. C. Hansen, L. F. Rodríguez, and K. C. Ting. 2010. Optimization of miscanthus harvesting and handling as an energy crop: BioFeed model application. *Biological Engineering* 3(2): 37-69.
- Simao, A., P. J. Densham, and M. Haklay. 2009. Web-based GIS for collaborative planning and public participation: An application to the strategic planning of wind farm sites. *Journal of Environmental Management* 90(6): 2027-2040.
- Stewart, C. A., S. Simms, B. Plale, M. Link, D. Y. Hancock, and G. C. Fox. 2010. What is cyberinfrastructure. In *Proceedings of the 38th annual fall conference on SIGUCCS*, 37-44. ACM.
- Tang, W., Wang, S., Bennett, D.A., and Liu, Y. 2011. Agent-based modeling within a cyberinfrastructure environment: a service-oriented computing approach. *International Journal of Geographical Information Science*, 25 (9): 1323–1346.
- Tilman, D., R. Socolow, J. A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala et al. 2009. Beneficial biofuels—the food, energy, and environment trilemma. *Science* 325(5938): 270-271.
- Tittmann, P., N. Parker, Q. Hart, and B. Jenkins. 2010. A spatially explicit techno-economic model of bioenergy and biofuels production in California. *Journal of Transport Geography* 18(6): 715-728.

- Wang, S. 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers* 100(3): 535-557.
- Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., and Nyerges, T. L. 2013. CyberGIS software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science*, 27(11): 2122-2145.
- Wang, S., M. P. Armstrong, J. Ni, and Y. Liu. 2005. GISolve: A Grid-based Problem Solving Environment for Computationally Intensive Geographic Information Analysis. In *Proceedings of the 14th International Symposium on High Performance Distributed Computing (HPDC-14) – Challenges of Large Applications in Distributed Environments (CLADE) Workshop*, 3-12. Research Triangle Park, North Carolina, USA: IEEE Press.
- Wang, S., and Y. Liu. 2009. TeraGrid GIScience Gateway: Bridging Cyberinfrastructure and GIScience. *International Journal of Geographical Information Science* 23 (5): 631-656.
- Wang, S., N. R. Wilkins-Diehr, and T. L. Nyerges. 2012. CyberGIS – Toward Synergistic Advancement of Cyberinfrastructure and GIScience: A Workshop Summary. *Journal of Spatial Information Science* 4: 125-148.
- Wang, S. and X. Zhu. 2008. Coupling cyberinfrastructure and geographic information systems to empower ecological and environmental research. *Bioscience* 58(2): 94-95.
- You, F., I. E. Grossmann, and J. M. Wassick. 2010. Multisite capacity, production, and distribution planning with reactor modifications: MILP model, bilevel decomposition algorithm versus Lagrangean decomposition scheme. *Industrial & Engineering Chemistry Research* 50(9): 4831-4849.

- Zambelli, P., C. Lora, R. Spinelli, C. Tattoni, A. Vitti, P. Zatelli, and M. Ciolli. A GIS decision support system for regional forest management to assess biomass availability for renewable energy production. *Environmental Modelling & Software* 38: 203-213.
- Zhang, J., A. Osmani, I. Awudu, and V. Gonela. 2013. An integrated optimization model for switchgrass-based bioethanol supply chain. *Applied Energy* 102: 1205-1217.