System of Systems Model for Analysis of Biofuel Development

Mashor Housh, M.ASCE[1]; Ximing Cai, M.ASCE[2]; Tze Ling Ng, A.M.ASCE[3]; Gregory F. McIsaac[4], Yanfeng Ouyang[5]; Madhu Khanna[6]; Murugesu Sivapalan[7]; Atul K. Jain[8]; Steven Eckhoff[9]; Stephen Gasteyer[10]; Imad Al-Qadi, Dist.M.ASCE[11]; Yun Bai[12]; Mary A. Yaeger[13]; Shaochun Ma[14]; and Yang Song[15]

Abstract: This paper presents a system of systems (SoS) Biofuel model considering the interdependency among the systems involved in biofuel development, including biofuel refinery location, transportation infrastructure, agricultural production and markets, environment, and social communities. The model provides the optimal infrastructure development and land-use allocation for biofuel production in a region considering socio-economic and water quality and quantity effects. The optimal development plan quantifies economic and hydrologic outputs and specifies biofuel refinery locations and capacities, refinery operations, land allocation between biofuel and food crops, optimal shipments of products and feedstock, and transportation infrastructure. The model is formulated as a mixed integer linear program (MILP) and is solved by an algorithm developed specifically to cope with the large size of the optimization problem. In addition to the development of the SoS-Biofuel model, this paper demonstrates the functionality of the model and its ability to analyze the effects of interdependency among subsystems by applying it to a watershed in Illinois. The SoS-Biofuel model is used to investigate the effects of different biofuel polices on infrastructure needs and related environmental consequences, highlighting the interdependencies inherent in the optimal development of the entire system. **DOI:** 10.1061/(ASCE)IS.1943-555X.0000238.

Author keywords: Biofuel mandate; Boifuel development; Infrastructure; Systems of Systems; multidisciplinary modeling.

Introduction

National mandates for biofuel production stipulate new regional and national infrastructure requirements that result in several environmental concerns. This is particularly true considering the inclusion of advanced biofuel in the renewable fuel standard (RFS), in which 60 billion liters of advanced biofuels, consisting of cellulosic ethanol from corn stover, perennial grasses and woody biomass, will be blended with fossil fuel in the United States by 2022 [Energy Independence and Security Act (EISA) 2007]. The cellulosic ethanol mandate requires the construction of new cellulosic ethanol refineries, whereas current first-generation (corn-based) biofuel refineries continue to be maintained and possibly expanded to meet the ethanol demand. Land allocation decisions depend on the comparative advantages of producing conventional crops versus dedicated energy crops, and special consideration is given not only to crop yields and

[1]Postdoctoral Associate, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822; and Assistant Professor, Faculty of Management, Dept. of Natural Resources and Environmental Management, Univ. of Haifa, Haifa 31905, Israel. E-mail: mashor.housh@gmail.com

[2]Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822 (corresponding author). E-mail: xmcai@illinois.edu

[3]Assistant Professor, Dept. of Civil and Environmental Engineering, Hong Kong Univ. of Science and Technology, Clear Water Bay, Hong Kong. E-mail: tzeling@ust.hk

[4]Associate Professor Emeritus, Dept. of Natural Resources and Environmental Sciences, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: gregmcisaac@gmail.com

[5]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: yfouyang@illinois.edu

[6]Professor, Dept. of Agricultural and Consumer Economics/Energy Biosciences Institute, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: khanna1@illinois.edu

Note. This manuscript was submitted on September 11, 2013; approved on September 5, 2014; published online on October 27, 2014. Discussion period open until March 27, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342/04014050(14)/$25.00.

[7]Professor, Dept. of Geography and Geographic Information Science, Computer Applications Building, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: sivapala@illinois.edu

[8]Professor, Dept. of Atmospheric Sciences, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: jain1@illinois.edu

[9]Professor, Dept. of Agricultural and Biological Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: seckhoff@illinois.edu

[10]Assistant Professor, Dept. of Sociology, Michigan State Univ., East Lansing, MI 48824. E-mail: gasteyer@msu.edu

[11]Professor, Dept. of Civil and Environmental Engineering; and Director, Illinois Center for Transportation, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: alqadi@illinois.edu

[12]Postdoctoral Associate, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: yunbai1@illinois.edu

[13]Research Assistant, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: yaeger3@illinois.edu

[14]Graduate Student, Dept. of Agricultural and Biological Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: shaochun2004@gmail.com

[15]Graduate Student, Dept. of Atmospheric Sciences, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61822. E-mail: song81@atmos.uiuc.edu

production costs but also to proximity to biofuel refineries and transportation costs (Richard 2010).

The location and capacity of biorefineries affect the traffic-flow pattern and can induce congestion and accelerate the deterioration of highway and bridge systems, especially in some rural areas where transportation infrastructure is not designed for heavy traffic (Bai et al. 2011; Ng et al. 2011). Thus, transportation infrastructure expansion, rehabilitation, and maintenance will be key factors in determining optimal refinery location and capacity and farmland allocation to biofuel crops.

The changes in land use for biofuel crops and other agricultural inputs, such as fertilizers, will affect nutrient and sediment loads to water bodies, soil organic matter, and life-cycle carbon and nitrogen emissions. Introducing second-generation bioenergy crops is likely to increase water demand. An experimental study by McIsaac et al. (2010) in Central Illinois showed that *Miscanthus* has approximately 100 mm more annual evapotranspiration than corn and soybeans (conventional crops in the study area). Hence, large-scale planting of *Miscanthus* or similar perennial crops may alter the flow regimes at the watershed scale. However, *Miscanthus* is predicted to have positive effects on water quality because it requires less N fertilizer than corn, and produces less nitrate leaching than corn and soybeans (McIsaac et al. 2010; Ng et al. 2010). These water quality and quantity effects add complexity to planning biofuel development and associated infrastructural changes.

Moreover, regional and national biofuel targets require not only engineering infrastructure expansion but also human and social infrastructure development to improve decision making, management, and operations. It is important to consider social and community resiliency factors when determining the biofuel development plan (Magis 2008) because biofuel development is subject to social constraints—namely, social welfare and acceptance (or resistance) of local communities of the building of new biorefineries, expansion or renewal of existing infrastructures, land-use conversion, and potential negative environmental effects.

Therefore, biofuel development in the United States is likely to create unique challenges for critical lifeline infrastructure systems, including food production, energy supply, transportation, and water supply systems. Increasing ethanol production will lead not only to the expansion of biorefinery systems but also to straining portions of existing transportation, water supply and treatment, and other infrastructures, which are already aging and degrading even without the added load from ethanol production (Ng et al. 2011). Moreover, the nation's increasing dependence on biofuel crops changes the vulnerabilities of energy supply, water supply, and transportation systems to climatic and other natural and artificial factors, such as extreme weather and unexpected events. Thus, thoughtfully planned and well-maintained infrastructures are fundamental to the resilience and sustainability of the emerging biofuel economy (Wright and Brown 2007).

Given the mutual dependencies of different infrastructures and the environment, identifying the optimal development plan for the biofuel production system requires a system of systems (SoS) approach in which multiple infrastructures are simultaneously considered in one holistic framework with environmental and social constraints. The SoS is characterized by multiple-infrastructure, multidisciplinary, multiscale (spatial/temporal), multistakeholder, and multiresource phenomena. Fig. 1 shows the biofuel SoS and the interdependencies among relevant infrastructure components, which are primarily caused by physical proximity, operational interactions, competition for natural and human resources, and information communication at the local and regional scales.

Various infrastructure components operate in an environment characterized by interactions among engineering systems, the natural environment, and socio-economic systems. The interdependencies may stabilize the overall system operation; however, the damage affecting one infrastructure subsystem may result in damages in connected components, thus affecting large geographic regions and sending ripples through regional and national economies (Heller 2001). Sustaining the overall combined social and engineering system requires understanding and managing feedbacks and interrelationships among the subsystems (Fig. 1) across temporal and spatial scales. Therefore, a holistic approach is required for planning and designing sustainable and resilient infrastructure to achieve the biofuel supply target.

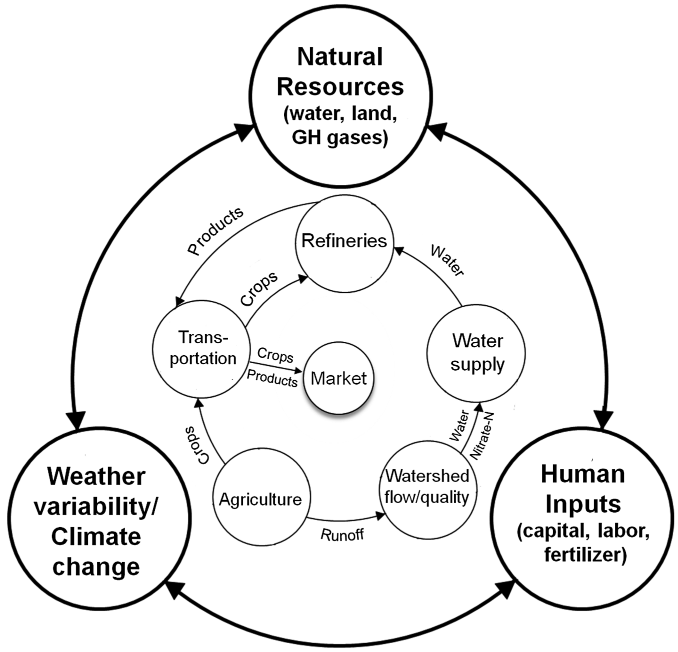

Fig. 1. Systems' interdependency and schematic representation the model

This paper presents a computational model designed to represent the SoS, which consists of several infrastructure, environmental, and social subsystems in a holistic framework. The model is useful for quantifying the effects of the subsystems' interdependencies (as demonstrated in the results), detecting bottleneck infrastructures, and shedding light on the most dominant factors that characterize the development of the biofuel system. The effect of the subsystem's interdependencies on the overall system solution is particularly analyzed in details and demonstrated in the results to show the importance of the propagation of changes from one subsystem to another.

Model Development and Model Features

Comparison with Existing Models

Current biofuel models can be classified into biorefinery location models and land allocation models. Static biorefinery location models, which are formulated for a typical year, have been presented in several studies including Eksioglu et al. (2010), Huang et al. (2010), Bai et al. (2011), and Hajibabai and Ouyang (2013). For example, Bai et al. (2011) developed a static model for a biorefinery location problem by taking into account not only the transportation cost as a function of distance but also the travel delay caused by congestion as a function of traffic volume. Hajibabai and Ouyang (2013) expanded this work by introducing the transportation network expansions to address a new tradeoff between congestion mitigation and transportation infrastructure enhancement.

The multiyear dynamic version of the refinery location problem has also been addressed in the literature. Kang et al. (2010) considered the multiyear development of a biofuel supply chain, but they considered neither congestion nor expansion of the transportation infrastructure in their model. In the aforementioned models, the biomass feedstock supply and the products' demand are given in advance as input to the model. Thus, the optimal locations of biorefineries are determined based on given supply–demand information. In contrast, Tittmann et al. (2010) solved the static biorefinery location problem by considering ethanol demand as an exogenous decision variable and feedstock supply as an endogenous decision variable, but they did not consider congestion and transportation network expansions.

Various land-allocation economic models have been developed to estimate the implications of biofuels on the agricultural and fuel sectors. For instance, the forest and agricultural sector optimization model (FASOM) is used by the U.S. Environmental Protection Agency (USEPA) to simulate the effect of the RFS mandate [Energy Information Administration (EIA) 2010]. FASOM is a multiyear model for determining land allocation and the price of gasoline and ethanol endogenously. Chen et al. (2011) developed the biofuel and environmental policy analysis model (BEPAM), which is a multiyear, multimarket equilibrium model for the fuel, agriculture, and livestock sectors. In BEPAM, prices are determined endogenously and decisions regarding land allocation and practices for producing row crops and perennial crops are made based on the spatial resolution of crop reporting districts.

These land-allocation economic models focus on land allocation decisions and do not consider the infrastructure aspects, such as refinery location and capacity, transportation investment, or traffic congestion. Recently, Chen and Önal (2012a) extended BEPAM to include biorefinery location. However, the model does not consider traffic congestion, transportation infrastructure expansion, water flow, or water quality effects.

This paper proposes a model that integrates a variety of system aspects that are typically modeled separately and demonstrates the interdependencies among these complex systems. The proposed model is holistic and includes a number of distinct features. First, rather than treating feedstock supply locations and the amount of feedstock as external inputs, the model handles the location of biorefineries and land allocation among various feedstock simultaneously; therefore, feedstock supply is determined endogenously by the model based on the spatial and temporal interdependency between the feedstock supply and the biorefinery capacities. Second, the proposed model is dynamic; it considers the intertemporal dependency among refinery facilities, feedstock production, and the supporting infrastructures. As such, the model allows for multistage biorefineries' capacity expansion corresponding to the increase in the biofuel mandates over the planning horizon.

Third, the model encompasses traffic congestion and transportation infrastructure expansion/renewal decisions. Fourth, it incorporates environmental aspects at the watershed scale—namely, surface water flows and nitrate loads in the subwatersheds—into the planning model. The various infrastructures in the system of systems are traditionally described at different temporal and spatial scales/resolutions (e.g., annually for agricultural systems, monthly/daily for water quantity, and hourly for transportation systems).

Interdisciplinary Approach

A SoS approach integrates data, experience, perspectives, and concepts from more than one discipline. The integrated model described in this paper was developed under an interdisciplinary research project involving multidisciplinary engineering, economics, sociology, and natural sciences such as hydrology and atmospheric science.

Researchers from these areas worked jointly on the conceptual development and implementation of the integrated model. The models and input and output data sets from the working groups of the subsystem were used to develop the interrelations among the subsystems, which were justified by the entire team and were eventually used to construct the integrated model. The model was used to address infrastructural support for biofuel development, which provided a platform for researchers from different groups to explore various planning scenarios and policy options. Results from the integrated model were also fed back to the subsystem models for more detailed simulation and discipline-specific study. One of the challenges for such interdisciplinary study is that each discipline uses its own terminology, but intergroup discussions greatly helped to overcome this barrier.

User Interface

Large-scale models can overwhelm users with input requirements and output possibilities. Because of the multidisciplinary nature of the model, communication is needed for developing models that involve multiple areas among the different groups from different areas of expertise. To facilitate communication, a user-friendly graphic user interface (GUI) was developed using a geographic information system (GIS) package. Model users can input the model parameters through the GUI, and tabular and graphic-based reports can be automatically generated by the GUI. The GUI incorporates a spreadsheet input tool for defining the system layout, components, and parameters; it automatically generates the input data for a MATLAB code that formulates the optimization problem, solves it using a specially developed solution algorithm that takes advantage of the problem structure (see the section on the model solution), and generates the solution report. The solution report uses *ArcGIS* to create GIS maps that describe the optimal development of the system along with the planning horizon, and provides spreadsheets with detailed tabular information. The result reports help users to examine the information obtained from the solution provided by the integrated model efficiently.

Model Components

A schematic representation of the SoS model is depicted in Fig. 1. The agriculture system produces crops for food markets and refineries. The type of crops determines the runoff and water quality in the watershed streams. The feedstock is transported through a transportation network to food markets (e.g., export nodes and local markets) and to biofuel refineries. After extracting the amount of water needed for the ethanol production process, refineries use the delivered raw materials to produce ethanol and dried distillers grains with solubles (DDGS). DDGS is the protein and fiber remaining from the conversion of corn to ethanol; it is used as a secondary commodity to feed different types of livestock. Ethanol and DDGS are transported through the same transportation system to their respective demand zones.

The land allocation between food, first-generation, and second-generation biomass crops affects the optimal location, size, and operation of biorefineries. The optimal location of refineries depends on the proximity to both raw material supply regions and ethanol and DDGS consumers. These two factors are negatively correlated because the feedstock supply is abundant in rural areas, where ethanol demand is limited. Although the cost of transporting raw materials is greater than that of transporting refinery products, the right balance between the two is not trivial.

The capital cost of cellulosic refineries is expected to be higher than first-generation refineries. This may lead to fewer and more centralized high-capacity cellulosic refineries. However, feedstock transportation costs will cause the tradeoff between capital and operational costs given that cellulosic feedstock is bulkier than corn and therefore requires higher transportation cost.

The location and capacity of a biorefinery affect the traffic flow patterns, create congestion, and accelerate transportation infrastructure deterioration, especially when refineries are located in rural areas where the transportation infrastructure may not be designed for heavy traffic. Land-use change, expansion of refineries, and expansion of transportation infrastructures face environmental and social constraints, which will affect the acceptance/resistance of local communities to proposed changes.

Therefore, it is necessary to develop an integrated model that considers the interdependencies among the interlinked subsystems, i.e., system of systems (SoS), for the adequate evaluation of the optimal biofuel development plan and its effects. The model was developed as an annual model and was later expanded into a multiyear model. The annual model served as the building block for the multiyear model. The state variables corresponding to the amounts of perennial crops and infrastructure capacities were used to link the years. The annual model is an integrated model of SoS, which considers the interlinked subsystems of the biofuel development. The subsystems of the integrated model are shown in Fig. 1. The objective of the model is to maximize the profit from the entire system—without regard to relative individual profits—under physical, technological, operational, and environmental constraints from all subsystems. In the annual model, the effect of crop selection and facility location decisions on the watershed subsystem is represented by a monthly time step to capture the intra-annual seasonality, and the effect on the transportation subsystem is represented by peak-hour traffic to account for congestion conditions.

Land-Use Subsystem

The land-use subsystem models the allocation of land between two conventional annual crops: corn and soybeans, and an energy crop, *Miscanthus*. It also allows for the harvesting of corn stover as a biofuel feedstock. It is assumed that the owner of a land parcel can choose crop i, where $i = m$ (*Miscanthus*), $i = c$ (corn), and $i = s$ (soybeans), respectively. A land parcel level spatial disaggregation is used, in which a given region (e.g., watershed) is divided into N_l land parcels. Accordingly, the land parcel data vectors in $\mathbb{R}^{N_l}$ are introduced as described subsequently. A is the area of a land parcel, whereas CC_i^t and CY_i^t represent annual crop production cost and crop yield per unit area, respectively. Because of the perennial nature of *Miscanthus*, which has a productive lifetime of 15–20 years, significant establishment costs, and a low yield in the first two years, the yield and cost of *Miscanthus* are age-dependent. Both yield and production cost are assumed to stabilize after the third year; thus, *Miscanthus* age 1 is defined by $i = m_1$, *Miscanthus* age 2 by $i = m_2$, and *Miscanthus* age >2 by $i = m_3$. Given this notation, *Miscanthus* production cost consists of two parts: CC_m^t is the annual crop production cost per unit area of *Miscanthus* for all ages, and CC_i^t $i = \{m_1, m_2, m_3\}$ is the establishment cost per unit area for *Miscanthus*, which is age-dependent. $i = m$ refers to total *Miscanthus*, irrespective of age.

Corn stover is also considered as a raw material for second-generation refineries. The maximum harvested yield and the production cost of corn stover per unit of area are CY_{cs}^t and CC_{cs}^t, respectively. To facilitate the notation, two sets are defined: $I \equiv \{c, cs, s, m_1, m_2, m_3\}$ where *Miscanthus* is age-dependent, and $I_1 \equiv \{c, cs, s, m\}$ where all *Miscanthus*, irrespective of age, is included. Based on the data mentioned previously, the model predicts the fractions of each land parcel, defined as vectors $X_i^t \forall i \in I$, which will be allocated to different activities. The total fraction of land covered with *Miscanthus* is given in Eq. (1)

$$X_m^t = X_{m_1}^t + X_{m_2}^t + X_{m_3}^t \quad \forall\ t \tag{1}$$

Eq. (2) represents the total annual cost of *Miscanthus* production, which consists of production cost (i.e., machinery, mowing, harvesting, baling) and establishment cost in the first two years, which covers the costs of rhizomes and land preparation to establish the crops

$$K_{\text{land}}^t = \sum_{i \in I} (CC_i^t)^T \text{diag}(A) \cdot X_i^t + (CC_m^t)^T \text{diag}(A) \cdot X_m^t \quad \forall\ t \tag{2}$$

where diag$(\cdot)$ = diagonal matrix; and $(\cdot)^T$ = transpose operator.

Given the crop yields and the agricultural area in each land parcel, fixed input-output Leontief production functions are used for row crops and *Miscanthus* production. The land parcels feedstock supply is defined in Eq. (3)

$$S_i^t = \text{diag}(A) \cdot \text{diag}(CY_i^t) \cdot X_i^t \quad \forall\ i \in I \quad \forall\ t \tag{3}$$

The total amount of *Miscanthus* is given in Eq. (4)

$$S_m^t = S_{m_1}^t + S_{m_2}^t + S_{m_3}^t \quad \forall\ t \tag{4}$$

In modeling the multiyear dynamics of *Miscanthus*, crops that are 2 years or older exist only as a continuation of the current crop. Furthermore, *Miscanthus*-planted areas may be converted for growing conventional crops. This former is represented in Eqs. (5) and (6)

$$X_{m_2}^t \le X_{m_1}^{t-1} \quad \forall t \tag{5}$$

$$X_{m_3}^t \le X_{m_3}^{t-1} + X_{m_2}^{t-1} \quad \forall\ t \tag{6}$$

In Eq. (7) the total amount of land used for all agricultural production activities cannot exceed the available agricultural land

$$X_m^t + X_c^t + X_s^t \le \alpha_a \quad \forall\ t \tag{7}$$

where α_a = vector of the fraction of agriculture land in each land parcel.

Corn stover is considered a raw material for cellulosic refineries. Because it is a byproduct of cultivating corn, corn stover can only be harvested from land planted under corn, as in Eq. (8)

$$X_{cs}^t \le \text{diag}(\alpha_{cs}) \cdot X_c^t \quad \forall\ t \tag{8}$$

where α_{cs} = vector of the maximum fraction of corn area from which corn stover is harvested for ethanol production.

To avoid extreme monocultures, in which land parcels in specific regions are dedicated to the most profitable crop type, the historical crop mix pattern is used for each land parcel to constrain the optimal solution to more realistic land allocations (Chen and Önal 2012b). Because the model includes a new second-generation bioenergy crop, deviation from the historically observed crop mix pattern is allowed as shown in Eq. (9)

$$CP - \alpha_{dv} \le X_s^t \cdot \mathrm{diag}(X_c^t)^{-1} \le CP + \alpha_{dv} \qquad \forall\ t \tag{9}$$

where CP and α_{dv} = vectors of the observed crop mix pattern (i.e., the ratio between soybeans and corn land) and allowed deviation, respectively.

Farmers adhering to sound risk-management practices will not allocate all of their lands to *Miscanthus*, which is a relatively new crop. To simulate this behavior, the model restricts the land allocated to perennial grasses in each land parcel to a maximum fraction *Miscanthus* $X_m^{\max}$, as shown in Eq. (10)

$$X_m^t \le X_m^{\max} \quad \forall\ t \tag{10}$$

Eqs. (1)–(10) describe the land-use system for biomass production, which represents the constraints on land availability and land allocations between the first and second generation of biofuel crops.

Biorefinery Subsystem

The locations of biorefineries and the land-use patterns are interdependent. The comparative advantage of growing energy crops in a specific region depends on the region's proximity to ethanol production facilities. A new refinery may induce farmers to switch from conventional crops to cellulosic energy crops or to sell corn as biofuel feedstock instead of selling it to food markets; whereas, land use may attract biorefinery investors to build in the proximity of feedstock supply to minimize costs. Thus, agricultural land-use allocations and biorefinery locations should be solved simultaneously using a unified model.

In the refinery subsystem N_r, potential refinery locations are defined in the modeled region. Each of the potential locations can hold a first-generation biorefinery, a second-generation biorefinery, or both. First-generation biorefineries are fed by corn grain ($i = c$) as a feedstock and produce both ethanol and DDGS as products $j = \{e, gs\} \equiv J$. Second-generation biorefineries (i.e., cellulosic refineries) are fed by *Miscanthus* and corn stover ($i = \{m, cs\}$) and produce ethanol ($j = e$).

Consequently, two vectors of decision variables, C_1^t and C_2^t, are defined in $\mathbb{R}^{N_r}$ as the refineries' capacity, corresponding to first and second-generation biorefineries. The capacities are time-dependent as both building and expanding the refineries over the multiyear planning horizon are considered. This is necessary to reflect the expanding demand of biofuel products and the time-dependent supply of the perennial crops. Therefore, building or expanding a refinery in a given year depends on the system's needs throughout the planning horizon. However, the optimal operation plan of the refineries may require below-capacity operation in certain years (e.g., as a result of a draught in certain year). The refineries' feedstock demand is defined as the decision vectors $DR_i^t \in \mathbb{R}^{N_r}\ i = \{c, m, \mathrm{cs}\}$.

The model is designed to include both capital and operation costs of the biorefineries along the planning horizon. The refineries' capital and operation costs are influenced by economies of scale. The capital cost of building a new refinery plant includes a fixed investment that does not depend on the capacity and a variable cost that is associated with the designed capacity. The operation cost accounts for expenses that result from the refinery's process, which does not include the feedstock purchasing cost. The operation cost consists of a variable operation cost that depends on the amount of processed raw material (e.g., energy cost, labor cost) and a fixed cost that is not greatly influenced by changes in production level (e.g., maintenance cost). The capital and the operation costs have been formulated with fixed and linear variable costs. The variable cost depends on the activity level, whereas the fixed cost is constant. Thus, the model prefers the lower capital cost of the centralized refinery, which reflects the effect of economy of scale.

Because of the discontinuous terms associated with the fixed capital costs, two vectors of binary decision variables, $ZI_{1,2}^t \in \mathbb{R}^{N_r}$, are defined. These binary variables are used to track the increase in the capacity of refineries (first and second-generation refineries), and take values of 1 if there is capacity expansion and 0 otherwise. Consequently, the capital cost associated with the refineries expansion in a given year is given in Eq. (11)

$$K_{\mathrm{Capital}}^t = \sum_{k=1}^{2} [(FC_k^t)^T \cdot ZI_k^t + (VC_k^t)^T \cdot (C_k^t - C_k^{t-1})] \quad \forall\ t \tag{11}$$

where FC_k^t, VC_k^t are vectors of fixed and variable capital cost, respectively.

Unlike the capital cost in which the fixed cost is encountered only when a capacity expansion occurs, in the operation cost, Eq. (12), the fixed cost is encountered when the refinery exists. Hence, vectors of binary decision variables $ZE_{1,2}^t \in \mathbb{R}^{N_r}$ correspond to existing refineries and take values of 1 if the refinery exists and 0 otherwise

$$K_{\mathrm{Operation}}^t = \sum_{k=1}^{2} [(FO_k^t)^T \cdot ZE_k^t + (VO_k^t)^T \cdot DR_k^t] \quad \forall\ t \tag{12}$$

where $DR_1^t = DR_c^t$, $DR_2^t = DR_m^t + DR_{cs}^t$; FO_k^t and VO_k^t = vectors of fixed and variable operation costs, respectively; and $DR_{1,2}^t$ = vectors of first and second-generation refineries' feedstock demands.

Once a biorefinery is built at a given location and in a given year, downsizing the capacity of facilities is not allowed, as given in Eq. (13). Thus, the model assumes that it remains operational throughout the planning horizon. However, as discussed previously, the production amount may vary over time and fall below the facility's design capacity

$$C_k^t \ge C_k^{t-1} \quad \forall\ k = 1, 2 \quad \forall\ t \tag{13}$$

Eq. (14) ensures the desired behavior of the zero/one pattern in vectors $ZI_{1,2}^t$ and $ZE_{1,2}^t$ (in which the value of one is assigned whenever the activity level is nonzero); a refinery may only increase its capacity if $ZI_{1,2}^t$ is one

$$\Delta \mathrm{Cap}_{\min,k}^t \cdot ZI_k^t \le C_k^t - C_k^{t-1} \le \Delta \mathrm{Cap}_{\max,k}^t \cdot ZI_k^t \quad \forall\ k = 1, 2 \quad \forall\ t \tag{14}$$

where $\Delta \mathrm{Cap}_{\min,k}^t$ and $\Delta \mathrm{Cap}_{\max,k}^t$ are the minimum and maximum-capacity expansion allowed in the case of refinery upgrade. Therefore, a positive increase in the capacity may take place only if the refinery is expanded. Furthermore, the expansion cannot fall below a minimum and exceed a maximum-capacity expansion. Similarly, in Eq. (15), a refinery may have a positive capacity if $ZE_{1,2}^t$ is one

$$0 \le C_k^t \le \mathrm{Cap}_{\max,k}^t \cdot ZE_k^t \quad \forall\ k = 1, 2 \quad \forall\ t \tag{15}$$

where $\mathrm{Cap}_{\max,k}^t$ = maximum allowed capacity. If the capacity C_k^t is positive, then constraint Eq. (13) will keep the capacity positive in the subsequent year C_k^{t+1}, which in turn implies that $ZE_{1,2}^{t+1}$ are ones. Thus, if a refinery exists in a given year, it will remain operational throughout the planning horizon.

Fixed input-output Leontief production functions are used for the biorefinery feedstock processing. Given the production per unit mass of feedstock, the first-generation refineries' ethanol and DDGS supply is defined in Eq. (16)

$$S_{1,j}^t = PY_{j,c}^t \cdot DR_c^t \quad \forall\ j \in J \quad \forall\ t \tag{16}$$

where $PY_{e,c}^t$ and $PY_{gs,c}^t$ = ethanol and DDGS yield per unit mass of corn, respectively. The second-generation refineries' ethanol supply is defined in Eq. (17)

$$S_{2,e}^t = PY_{e,m}^t \cdot DR_m^t + PY_{e,cs}^t \cdot DR_{cs}^t \quad \forall\ t \tag{17}$$

where $PY_{e,m}^t, PY_{e,cs}^t$ = ethanol yield per unit mass of *Miscanthus* and corn stover, respectively. Furthermore, the amount of ethanol supplied by each refinery cannot exceed its processing capacity, as in Eq. (18)

$$S_{k,e}^t \le C_k^t \quad \forall\ k = 1,2 \quad \forall\ t \tag{18}$$

Water withdrawal from refineries depends on the amount of raw material processed by the refineries. The yearly water amount required by the refineries is given in Eq. (19)

$$QR^t = \sum_i WC_i \cdot DR_i^t \qquad i = \{c, m, cs\} \quad \forall\ t \tag{19}$$

where WC_i = per-unit feedstock water consumption.

Eqs. (11)–(19) describe the type and capacity of biorefineries and the relations between this and other subsystems such as the biomass production and water supply systems.

Transportation Subsystem

In the transportation subsystem, both the routine operation (shipment routing) and the development of the transportation network are considered in the planning horizon.

Routine Operation

In the operational component, the model considers the routing of crop shipments for food and biorefineries' processing. Once the feedstock is converted into ethanol and DDGS, these products are transported to their respective demand zones. In the transportation subsystem, N_d potential demand zones (e.g., markets, consumers, or export nodes) are defined. The demands for theses demand zones are decision variables, represented in the form of vectors $D_i^t,\ D_j^t \in \mathbb{R}^{N_d},\ j = \{e, gs\},\ i = \{c, s\}$. The location of ethanol demand nodes is given exogenously, but terminal blending is not explicitly considered. Moreover, it is assumed that the demand is not constrained by the blend wall.

The land parcels, transportation infrastructure, refineries, and demand zones must satisfy mass-flow conservation for all crops and products in all years. The network can be represented as a directed graph consisting of N nodes connected by M directed edges. Each edge generally represents a roadway link between two nodes. If the direction of travel is not restricted, each link is represented by two edges, one in each direction. The topology of the network is represented by the connectivity matrix G, where $G \in \mathbb{R}^{N\times M}$ has a row for each node and a column for each edge. The nonzero elements in each row are $+1$ and -1 for incoming and outgoing edges, respectively. The first subset of rows in G corresponds to the N_l land parcels nodes, whereas the last subsets of rows correspond to the N_r refinery nodes and the N_d demand nodes, respectively. For each crop $i = \{c, cs, s, m\} \equiv I_1$ and product $j = \{e, gs\} \equiv J$ in year t, the system of linear equations in Eqs. (20) and (20) ensures mass-flow conservation at the network nodes

$$G \cdot f_i^t = [-S_i^t, 0, DR_i^t, D_i^t] \quad \forall\ i \in I_1 \quad \forall\ t \tag{20}$$

$$G \cdot f_j^t = [0, -S_j^t, D_j^t] \quad \forall\ j \in J \quad \forall\ t \tag{21}$$

where f_i^t and f_j^t = vectors containing the shipment flow on each link.

The shipment cost varies across commodities because of density differences between the bulky feedstock and the final biofuel products. For example, *Miscanthus* is much bulkier than corn and ethanol, and therefore imposes a much higher traffic load and, consequently, a higher average transportation operation and infrastructure cost per ton. The total operational transportation cost (i.e., the shipment cost) is considered linear with the shipment quantity/flow. The linearity assumption is reasonable because biofuel is shipped mostly during nonpeak hours of the day (Foulds 1976). The yearly shipment cost is given in Eq. (22)

$$K_{\text{Shipment}}^t = \sum_{i\in I_1} (TC_i^t)^T \cdot f_i^t + \sum_{j\in J} (TC_j^t)^T \cdot f_j^t \quad \forall\ t \tag{22}$$

where $TC_i^t \in \mathbb{R}^M$ = unit transportation cost on each link of the transportation network.

The biorefineries' products and corn and soybeans are shipped to potential N_d demand nodes, which represent markets, consumers, or export nodes. The profit from selling these commodities is based on a fixed price given at each demand node for each type of commodity. Because the model is designed to work on the watershed level, a significant spatial variation in the commodities' price is unlikely to occur, and the supply of products from the biofuel system is unlikely to affect the prevailing market-equilibrium prices. Therefore, when determining the optimum resource allocation and infrastructure investment, the model incorporates fixed prices for ethanol, DDGS, corn, and soybeans as exogenous parameters. The revenue in each year is defined in Eq. (23)

$$\text{Re}^t = \sum_{i\in I_1} (P_i^t)^T \cdot D_i^t + \sum_{j\in J} (P_j^t)^T \cdot D_j^t \quad \forall\ t \tag{23}$$

where $P_i^t \in \mathbb{R}^M$ = price vector containing the price for all demand nodes.

Capacity Expansion

The biorefineries are expected to result in a high traffic volume on the transportation system, which could cause additional congestion and delays and could accelerate transportation infrastructure deterioration. Congestion in roadway links with high background traffic demand (i.e., public traffic) results in high transportation cost and community resistance. Moreover, routing the biofuel shipments through congested transportation links may hinder the efficiency of the supply chain. Capacity analysis is performed at the peak hour in terms of traffic volume, the most critical period of the traffic system. For that purpose, the peak hour traffic volume, TV^t, for each link of the transportation system is defined in Eq. (24)

$$TV^t = \sum_{i\in I_1} \text{diag}(CE_i^t) \cdot f_i^t + \sum_{j\in J} \text{diag}(CE_j^t) \cdot f_j^t + B^t \quad \forall\ t \tag{24}$$

where $CE_i^t \in \mathbb{R}^M$ = conversion rate of feedstock flow to passenger car equivalent (PCE) per hour in each link; and B^t = peak hour background traffic on the links.

The biofuel/biomass shipment demand is prorated uniformly over time to avoid overtime operation and storage requirements that may be required if peak hours are avoided completely. As such, a portion of shipments is conducted in the peak hours, during which the level of service (LOS) may be violated. To ensure the feasibility of biofuel development, it might be worthwhile to invest in transportation infrastructure. In fact, the capital investment associated with transportation infrastructure expansion and/or renewal may be considered part of the biofuel industry's development plan, e.g., through private-public partnerships (Unnikrishnan et al. 2009). The investment in transportation infrastructure is directly interlinked with the location of biorefineries; therefore, it is necessary to address the expansion of the transportation network in an integrated framework that encompasses the biorefinery location and the shipment routing (Bai et al. 2011; Hajibabai and Ouyang 2013).

The increase in transportation network capacity, considering its effects on travel time and congestion, is known in the transportation literature as the network design problem (NDP). In the NDP, improvements or additions to links to the network are sought to minimize the summation of infrastructure investment and traffic congestion. This study considers the system-optimal continuous NDP (CNDP). For details on the classification of the NDP, the reader is referred to the Supplemental Data. In classical CNDP, the travel cost is taken as a nonlinear convex function of the traffic volume, reflecting the fact that the travel time increases quickly in congested conditions. This nonlinear relation on the travel time adds significant complexity to the CNDP and requires assigning a monetary value to the travel time, which may vary across industry and public users (LeBlanc 1979).

The approach suggested by Davis and Saderson (2002) is adopted to account for congestion, which does not explicitly require nonlinear travel costs or conversion between travel time and travel cost. In this approach, the ratio of traffic volume and capacity is used as an indicator of the LOS (Roess et al. 1998). Instead of directly computing the travel costs under congestion for biofuel shipments and the general public, performance guarantee constraints are imposed so that the LOS does not degrade below a certain threshold on all network links. The LOS requirement is enforced by the linear constraint in Eq. (25)

$$TV^t - \mathrm{diag}(LC^t) \cdot LOS^t_{\max} \leq 0 \quad \forall\ t \tag{25}$$

where LC^t, $LOS^t_{\max} \in \mathbb{R}^M$ = vectors containing the transportation link capacity and maximum allowed LOS, respectively. In Eq. (26), the transportation capacity is allowed to be increased over time to account for the increasing traffic volume

$$LC^t = LC^{t-1} + \Delta LC^t \quad \forall\ t \tag{26}$$

where $\Delta LC^t \in \mathbb{R}^M$ = capacity increment.

In Eq. (27), the yearly capacity expansion cost is considered to be a linear function of the capacity increase (Unnikrishnan et al. 2009)

$$K_{\text{expansion}} = (EC^t)^T \cdot \Delta LC^t \quad \forall\ t \tag{27}$$

where $EC^t \in R^M$ = per-unit capacity expansion cost. The relationships described in the transportation subsystem connect the transportation system development (capacity, traffic congestion, and costs) to biofuel production, refinery, and fuel shipment to demand sites.

Environmental Subsystem

The integrated model includes a watershed module to represent the effects of land allocation and refinery water extraction decisions on the water flow and nitrate load in the streams. The watershed is modeled as a network of flow, in which the nodes represent subwatershed outlets and the linkages are defined as the one-directional flow paths between the subwatersheds. Similar to the transportation network, the water/nitrate flow network can be represented as a directed graph. However, because of the special spatial relationship between the subwatersheds, this graph consists of a *tree graph*, which contains no cycles.

The topology of the subwatersheds' network is represented by the connectivity matrix W, in which $W \in \mathbb{R}^{N_w \times N_w}$ has a row for each node and a column for each linkage between the nodes. Additionally, the last node and the last column in W correspond to the watershed outlet and the flow at the outlet, respectively. The runoff contribution (water and nitrate) of each subwatershed is determined based on the land cover and the water use in that subwatershed. Thus, the decisions of other subsystems such as land allocation in the land-use subsystem and water use in the refinery subsystem define the sources and the sinks of the network.

Land-use decisions are taken in 10×10 km land parcels, whereas delimitation of the subwatersheds is determined based on the topographic characteristics of the land. Thus, it is necessary to convert land-allocation decisions in the land-use subsystem into subwatershed-land allocation. A subwatershed may contain more than one land parcel and may also contain partial land parcels. Thus, the overlapping between the land parcels and the subwatershed should be based on a spatial analysis of the modeled region. Given N_w subwatersheds, the overlapping matrix could be defined as $OV \in \mathbb{R}^{N_w \times N_l}$, in which each row sums to one and corresponds to the fractions of the overlapping area between the subwatersheds and the land parcels. Thus, the land allocation in the subwatersheds is given in Eqs. (28)–(31) as

$$Y^t_i = OV \cdot X^t_i \quad \forall\ i = \{m, c, s\} \quad \forall\ t \tag{28}$$

where Y^t_i = fraction of land with crop i in the subwatersheds.

As indicated in the land-use subsystem, only a fraction α_a of land parcel is considered agricultural land

$$Y_u = OV \cdot (1 - \alpha_a) \tag{29}$$

$$Y^t_f = 1 - \sum_i Y^t_i \qquad i = \{m, c, s, u\} \quad \forall\ t \tag{30}$$

where Y_u = fraction of nonagricultural land in the subwatersheds; and Y^t_f = fraction of the fallow land in the subwatersheds.

The runoff from each land parcel is a function of the land cover defined previously. Given the per-unit area water runoff (or water yield) for each subwatershed, the fixed input-output Leontief production functions are used to determine the total runoff contribution from the subwatershed. To capture the intra-annual seasonality, the time resolution of the runoff is lowered to a monthly scale. Thus, different production functions are used for each month, and the runoff contribution is defined in Eq. (31) as

$$\Delta Q^{t,\tau} = \sum_i \mathrm{diag}(WY^{t,\tau}_i) \cdot Y^t_i \qquad i = \{m, c, s, u, f\} \quad \forall\ t \quad \forall\ \tau \tag{31}$$

where $WY^{t,\tau}_i \in \mathbb{R}^{N_w}$ = runoff contribution when the entire subwatershed has the land cover i. Similarly, Eq. (32) defines the nitrate-N load runoff as

$$\Delta N^{t,\tau} = \sum_i \mathrm{diag}(NY^{t,\tau}_i) \cdot Y^t_i \qquad i = \{m, c, s, u, f\} \quad \forall\ t \quad \forall\ \tau \tag{32}$$

where $NY^{t,\tau}_i \in \mathbb{R}^{N_w}$ = nitrate-N runoff contribution when the entire subwatershed has the land cover i.

A subset of the subwatersheds can have water storage facilities (i.e., reservoirs). Given N_s reservoirs, the volume of the reservoirs' water along the planning horizon can be represented by Eq. (33) as

$$V^{(t,\tau)+1} = V^{t,\tau} + \Delta V^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{33}$$

where $V^{t,\tau} \in \mathbb{R}^{N_s}$ = water volume in the reservoirs; and $\Delta V^{t,\tau} \in \mathbb{R}^{N_s}$ = change in the storage. Eq. (34) defines how maximum and minimum constraints on storage volume are imposed to reflect policy and operational limits

$$V_{\min}^{t,\tau} \leq V^{t,\tau} \leq V_{\max}^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{34}$$

The mass conservation should be maintained in the network flow of the watershed. The linear equation systems in Eq. (35) ensure water-mass conservation in the subwatershed's network nodes

$$W \cdot Q^{t,\tau} = -\Delta Q^{t,\tau} + WR \cdot QR^{t}/12 + WV \cdot \Delta V^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{35}$$

where $WR \in \mathbb{R}^{N_w \times N_r}$ has a row for each subwatershed and a column for each potential refinery. The nonzero elements in each row have the value of 1 corresponding to refineries that withdraw water from the subwatershed. Similarly, $WV \in \mathbb{R}^{N_w \times N_s}$ has a row for each subwatershed and a column for each reservoir and nonzero element of 1 corresponding to the reservoir in the subwatershed.

The mass balance conservation should also hold for the nitrate mass propagating throughout the subwatersheds, as in Eq. (36)

$$W \cdot N^{t,\tau} = -\Delta N^{t,\tau} + WV \cdot \Delta VN^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{36}$$

where $\Delta VN^{t,\tau}$ = nitrate mass loss caused by denitrification in each reservoir (David et al. 2006).

Environmental policies are addressed by imposing streamflow and nitrate load constraints in Eqs. (37) and (38). These constraints can be specified by location (per subwatershed) and time (monthly) in the watershed to reflect the spatial and temporal variation of the regulations

$$Q^{t,\tau} \geq Q_{\min}^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{37}$$

$$N^{t,\tau} \leq N_{\max}^{t,\tau} \quad \forall\ t \quad \forall\ \tau \tag{38}$$

where $Q_{\min}^{t,\tau}$, $N_{\max}^{t,\tau} \in \mathbb{R}^{N_w}$ = minimum flow and the maximum nitrate of the subwatersheds.

In summary, the environmental system—including water quantity and quality and water supply to the refinery—is depicted at the watershed scale, with linkages to the biomass production system and refinery system, and to water resources management policies.

Social Subsystem

The biofuel development plan is also subject to social constraints, which are related to local communities' acceptance/resistance to biorefinery development, and to the effects of biofuel activities on communities. Another layer of N_{lc} local communities is added to the model, in which each local community lc is associated with one or more of the potential refinery locations. Positive effects of the refineries, Pos_{lc} (effects that may increase the willingness of the community to accept the refinery) such as the increase in employment, increase in household spending, and reliability of the refinery as a crop consumer are identified and quantified as a function of the outputs of other subsystems. Negative effects Neg_{lc} that may trigger community resistance, such as congestion, water limitation, water pollution and noise, are also identified and quantified. The final willingness to accept the refinery by the local communities is given in Eq. (39) as

$$AC_{lc} = AC_{lc}^{0} + (\alpha_{lc}^{pos})^{T} \cdot Pos_{lc} - (\alpha_{lc}^{neg})^{T} \cdot Neg_{lc} \quad \forall\ lc \tag{39}$$

where AC_{lc}^{0} = initial willingness to accept the refinery by each local community without considering the possible positive and negative aspects of the biofuels development; and α_{lc}^{pos} and α_{lc}^{neg} = contribution and penalization, respectively, to the willingness to accept refinery activities neighboring the local communities. For a development plan to be feasible, the willingness of the local community to accept the activities of the refineries should be higher than a prespecified threshold, as in Eq. (40)

$$AC_{lc} \geq AC_{lc}^{\min} \quad \forall\ lc \tag{40}$$

Overall Objective

The objective of the multiyear model is to maximize the present value of the entire system profit over the planning horizon, as in Eq. (41). The revenue is subject to the gains achieved from selling corn and soybeans and refineries' products in the demand zones. The costs are associated with the investment and the operation of the different subsystems

$$\max \sum_{t} \frac{\mathrm{Re}^{t} - K_{\text{Land}}^{t} - K_{\text{Capital}}^{t} - K_{\text{Operation}}^{t} - K_{\text{Shipment}}^{t} - K_{\text{Expansion}}^{t} - K_{\text{Miantenance}}^{t}}{(1+r)^{t-1}} \tag{41}$$

where r = discount rate.

Model Solution

The resultant model is a very large-scale, mixed integer linear programming (MILP) problem. A special structure of the model is considered to reduce its size. One dependent decision variable is extracted from each equality constraint. The dependent variables are then substituted in the objective function and the constraints to obtain a smaller model with fewer decision variables. The linear equality constraints are associated with a balance requirement in the different subsystems, such as the products and raw material shipment balance in the transportation network and the water flow and nitrate load in the subwatersheds network.

The dependent variables can be extracted efficiently from the problem using some of the concepts of the graph theory, which relates the extraction of dependent variables to the spanning tree (ST) that could be generated by the breadth-first-search (BFS) algorithm (Boulos et al. 2006) or any other ST-generating algorithm.

Following this procedure, each ST link in the transportation network could be eliminated from the optimization. The number of ST links is dramatically larger than the number of non-ST links, which is equal to the network loops. Therefore, the number of decision variables is significantly reduced. The reduction procedure is even more beneficial in the watershed because the watershed network is

ST by definition; thus, the water flow and nitrate load variables could be eliminated from the optimization problem.

Although the size-reduction technique is very useful, it only reduces the number of continuous variables in the model. Attempts to solve the model using a commercial mixed integer linear programming solver (CPLEX) resulted in inefficient performance because of the size of the model and the large number of binary decision variables. To cope with this computational burden, a heuristic approach called the successive smoothing algorithm (SSA) was developed to solve the mixed integer linear problem formulated in this study. The SSA solves optimization problems with a large number of fixed-cost variables using successive linear programming (LP) approximations for the problems. The SSA uses the smoothing techniques (Tishler and Zang 1982) to convert the problem to a sequence of nonlinear minimization problems. A specially designed linearization technique is then used to solve the nonlinear optimization problem and to identify the zero/nonzero pattern of the decision variables.

The performance of the SSA was tested on several scenarios of the integrated model and on a series of randomly generated problems. The SSA generated high-quality solutions in less computational time than the MILP approach. The results indicate that the maximum optimality gap obtained by the SSA algorithm is as low as 0.5%, and that the computation time is reduced by a factor of 15 when compared with CPLEX MILP solver.

Demonstration of Model Application and Corresponding Results

The model was applied to the Sangamon River basin, a 15,000-km^2 typical agricultural catchment in Central Illinois, to demonstrate its capability and potential outcomes. The model required several input data sets to represent the various subsystems, including crop areas and yields, crop production costs and crop prices, transportation capacities and costs, water quantity and quality, and refinery capacity and capital and operations costs. Available historical data were used and other required data were estimated using simulation models such as the hydrological model, economical model, and crop yield simulation model. Candidate locations for cellulosic and corn refineries, and demand nodes, were specified exogenously for the model. Decisions on the land-use subsystems were taken at the land parcel level (each land parcels is 10 × 10 km). Therefore, data on costs and yields for producing conventional crops and cellulosic biofuel feedstock include great spatial heterogeneity. A detailed description of the historical data sets used in the modeling work may be found in the Supplemental Data. The transportation mode within the Sangamon River Basin is limited to trucking for both feedstock and ethanol. However, the model makes no assumptions about the final transportation distance and the mode outside of the study area, i.e., if the ethanol production is to be exported out of the study area by other transportation modes (e.g., rail), then this exporting activity will be originated from the demand/export nodes (identified in the transportation subsystem), which are supplied by trucking.

For feedstock supply, use of alternative transportation modes (e.g., barge, rail) relies heavily on the geographical feature (e.g., rivers) of the study region and the infrastructure's availability. Moreover, past studies show that rail or barge is cost-effective only in cases of long-distance shipment (e.g., more than 200 km); however, the typical collection radius of a study area is 80 km (Mahmudi and Flynn 2006; Searcy et al. 2007).

The results presented in this section serve as a demonstration of (1) the integrated model outputs, and (2) the effect of subsystem interdependency on the overall system performance. The model was used for a 10-year planning horizon (2013–2022). The 2022 results are presented for demonstration purposes unless otherwise indicated. Because of a lack of data for the social subsystem, it was excluded from all runs presented in the paper.

Model Outputs

The model was used under a business-as-usual (BAU) scenario. No mandate of cellulosic biofuel was imposed and a scenario with hypothetical mandate was estimated according to the U.S. Renewable Fuel Standard (RFS) mandate. RFS requires 60 billion liters of cellulosic ethanol production in the United States by 2022. The share of the State of Illinois is estimated to be 20% of the total U.S. mandate (Kang et al. 2010); 2% of the U.S. mandate is estimated for the study site, the Sangamon River watershed in Central Illinois, which contains approximately 10% of the Illinois cropland. Specifically, a minimum annual production is established for different types of feedstock in each year in the planning horizon. Moreover, the model restricts the land allocated to *Miscanthus* in each land parcel to a maximum fraction of 30% as explained previously.

Land Allocation: BAU versus RFS

The land allocation under the BAU scenario is shown in Fig. 2(a). The economically optimal solution does not involve any *Miscanthus*, and the traditional corn-soybean pattern remains. When the ethanol mandate is imposed under the RFS scenario, the optimal solution in 2022 includes *Miscanthus* because corn stover alone cannot meet the cellulosic ethanol mandate specified for the watershed. Figs. 2(b–f) show the agriculture land converted to *Miscanthus* along the planning horizon. In the first year, *Miscanthus* is planted in two land parcels [Fig. 2(b)]. *Miscanthus* needs three years to produce its maximum yield; thus, the first two years depend primarily on corn stover to meet the cellulosic ethanol mandates. In the second year, *Miscanthus* conversion starts to grow gradually in the northern and southern regions of the watershed [Figs. 2(c–f)]. The land parcels chosen for *Miscanthus* have high yield potential as estimated by the crop yield simulation model used in this study. Between years 2 and 7, additional land parcels are converted to *Miscanthus* every other year. For example, some new land parcels are chosen for *Miscanthus* in year 2 but not in year 3.

Refineries and Transportation System: BAU versus RFS

Under the BAU scenario, it is profitable to produce 1.43 billion liters of corn ethanol at the assumed price of \$0.64/L. Seven corn refineries with different sizes are determined in the watershed [Fig. 3(a)], which are all put into use in the first year with the capacities shown in Fig. 3(a). The location of the refineries is primarily affected by the relative cost of shipping the feedstock from land parcels to refineries and shipping ethanol and DDGS to demand zones. The RFS scenario includes both corn-based and cellulosic-based refineries in year 10 [Fig. 3(b)]. Under the RFS scenario, the locations of the corn ethanol refineries are the same as those under BAU, but some of the corn refineries have smaller capacities (under the RFS scenario, 1.3 billion liters of corn ethanol is produced). By the end of the planning horizon, four cellulosic refineries are built. However, unlike the corn ethanol refineries, cellulosic-based refineries are built in different years throughout the planning horizon. Fig. 3(c) shows the capacity expansion of the refineries in Site A [Fig. 3(b)]. This cellulosic refinery is built

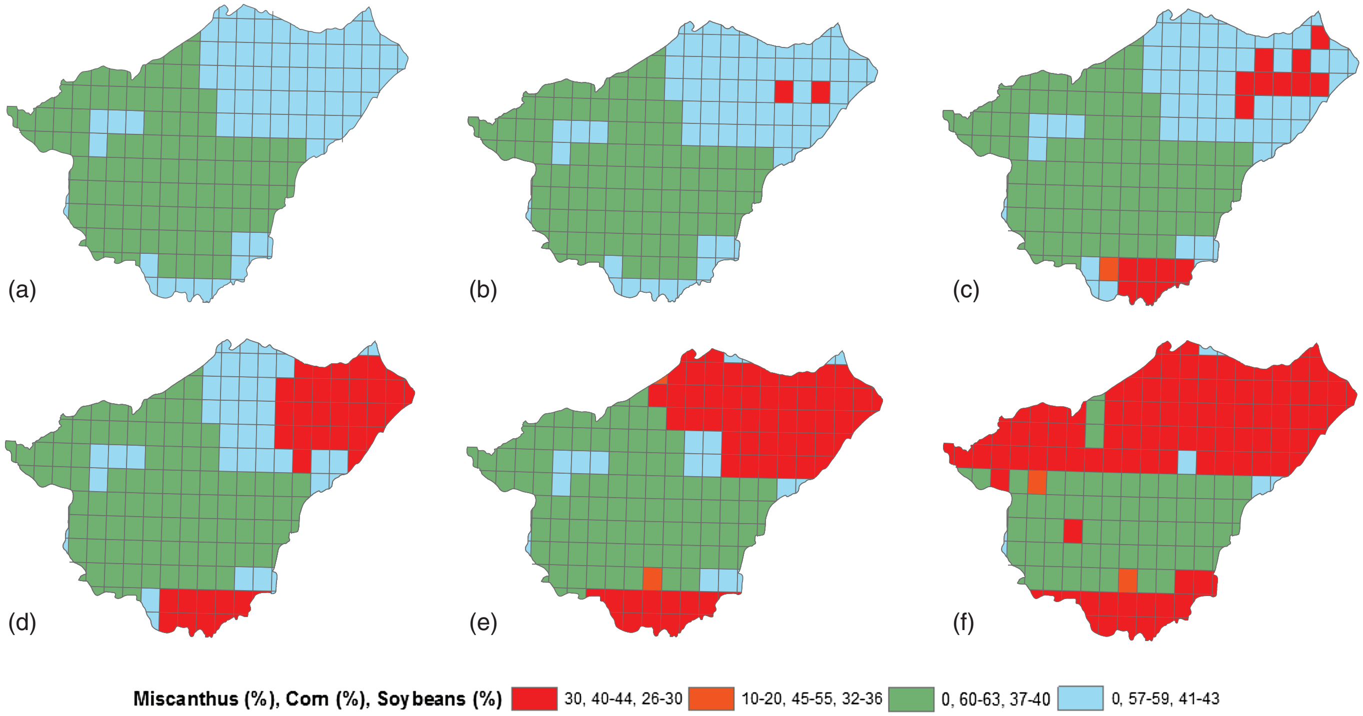

Fig. 2. (Color) Land allocation: (a) BAU scenario and RFS scenario in (b) year 1; (c) years 2–3; (d) years 4–5[a]; (e) years 6–7[a]; (f) years 8–10; [a]land allocation in successive years is approximately the same

in the first year and is then expanded in the fifth and the sixth year to reach its maximum capacity. Similar expansions are observed in other refineries in Sites B, C, and D. Thus, the model chooses to build cellulosic refineries in the vicinity of *Miscanthus* supplies because of the cost of shipping *Miscanthus* and corn stover to refineries. In the earlier years, *Miscanthus* is introduced in the south and the northeast of the watershed. Thus, the cellulosic refineries in Sites A and D are built earlier in the planning horizon to make use of *Miscanthus*.

Figs. 3(a and b) display the traffic situation induced by the activities related to biofuel development under the BAU and RFS scenarios, respectively. The RFS scenario induces heavier traffic in the system, especially in regions having both corn and cellulosic refineries. This system traffic increase becomes more critical when investigating regions with heavy background traffic.

Under the RFS scenario, one cellulosic refinery is built in the first year of the planning horizon and *Miscanthus* feedstock supply is available only after the first year. Fig. 3(c) shows the feedstock supply along the planning horizon, indicating that the main feedstock used in earlier years is corn stover. The use of corn stover decreases along the planning horizon; therefore, some of the land parcels that are economically viable for harvesting corn stover in the early years will not be exploited for corn stover after *Miscanthus* becomes available [Fig. 3(c)]. Starting from the fourth year, *Miscanthus* supply in the cellulosic refinery in Site A will exceed the corn stover supply. A similar behavior is noted in the cellulosic refinery plant in Site D, as shown in Fig. 3(c). In the first year, corn stover is used in the whole watershed because there is no *Miscanthus* supply in the watershed at that stage. In this scenario, it is assumed that farmers plant *Miscanthus* only after the mandate was in effect. However, if farmers are provided with sufficient information and incentives before the implementation of the mandate, it is likely that some will plant *Miscanthus* before year 1, and the corn stover–use pattern will change accordingly. Therefore, the model can be used to evaluate the effects of alternative start-up assumptions.

Costs and Revenues: BAU versus RFS

The multiyear costs and revenues associated with the system under the BAU and RFS scenarios are summarized in Table 1. Under BAU, the highest cost is associated with crop production, followed by the operation of the refineries. The results also show that the revenue from selling crops (soybeans and corn) for other demands (food and feed) exceeds that for biofuel. Although the revenue under the RFS scenario is higher than that under BAU, the profit from the system is lower because of the higher costs of meeting the cellulosic biofuel target under the RFS. The main cost increments are associated with the following: (1) Capital and operational costs of the cellulosic refineries under RFS; (2) production costs because *Miscanthus* production costs are higher than those of conventional crops; and (3) transportation costs because cellulosic feedstock materials are bulkier, which entails higher transportation costs.

As given in Table 1, despite the increase in the shipments under RFS, only a modest increase in the transportation capacity is suggested by the model because the cellulosic refineries are located far from congested links and only a limited transportation infrastructure expansion or renewal is required, compared with the BAU scenario.

The total net profit under the RFS scenario is expected to be lower than the BAU because the RFS represents a more constrained application of the model. The total net profit under RFS is 8% lower than that under BAU (Table 1). This reduction in the total profit implies that the price of cellulosic ethanol will need to increase from \$0.64 per liter (the ethanol price assumed in the model) to \$0.80 per liter to provide incentives for refineries to meet the RFS requirement in the watershed. Thus, cellulosic ethanol will need to be sold at a premium of \$0.16/L relative to corn ethanol, to induce its production. This premium could be interpreted as the minimum difference in the value of a renewable identification number (RIN) between cellulosic and corn-based ethanol required for cellulosic ethanol to be profitable.

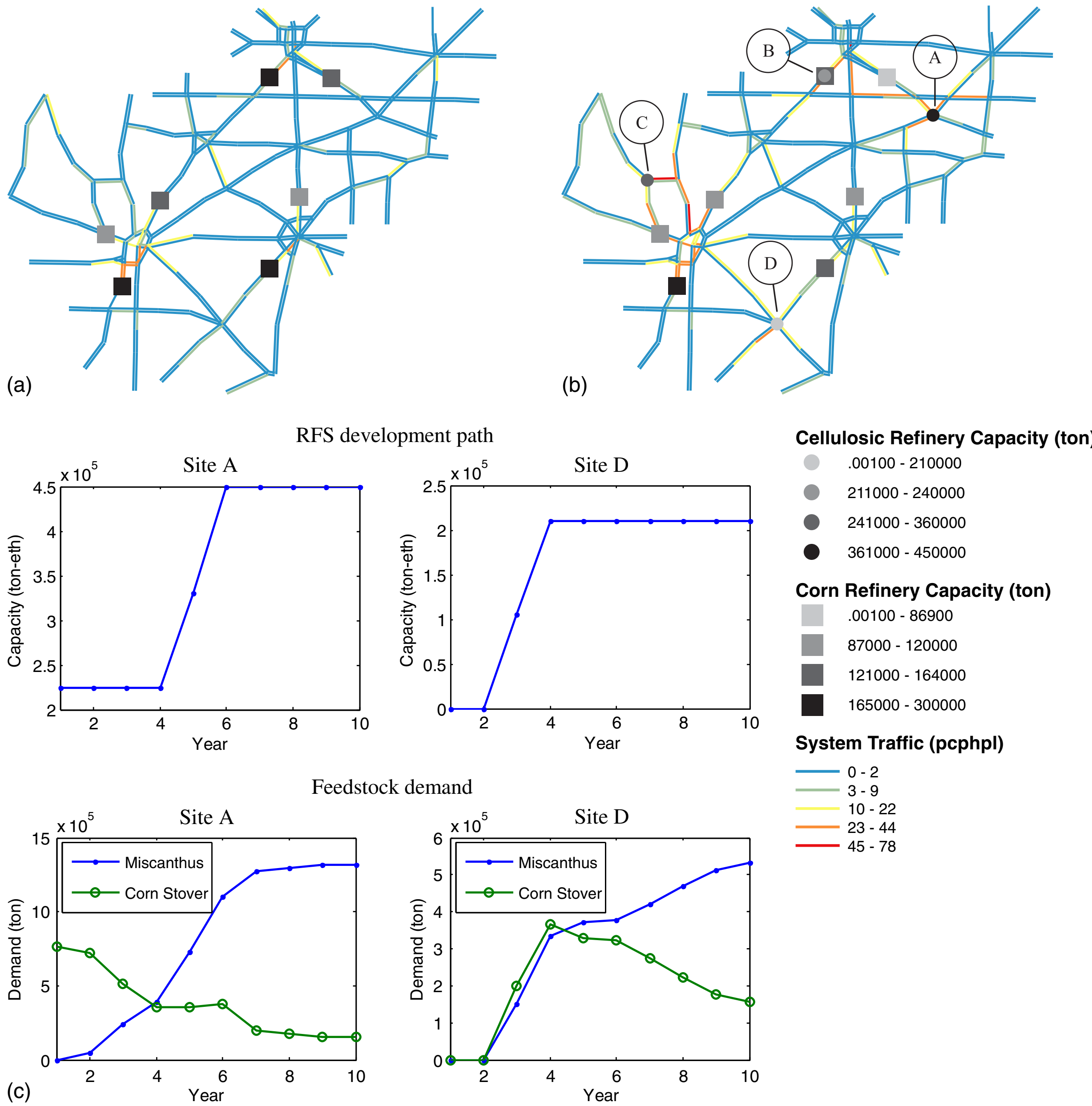

Fig. 3. System development in year 10: (a) BAU; (b) RFS; (c) RFS development path for sites A and D and the feedstock demand; pcphpl = passenger car per hour per lane

Streamflow and Nitrate-N Load: BAU versus RFS

The development of the biorefineries and the increase in the land fraction converted to *Miscanthus* has a direct effect on the quality and quantity of water. With the increase of the cellulosic ethanol mandate, the change in the land use will alter the water flow and the nitrate load. The demand for water for both refineries and for growing *Miscanthus* will cause a decline in the streamflow in the watershed. However, *Miscanthus* requires less nitrogen fertilization and therefore causes less nitrate pollution. Fig. 4 shows the yearly reduction in the flow and the nitrate load from each subwatershed under RFS, compared with the BAU scenario. Although the developed model has monthly temporal resolution, only the annual results are presented in this paper.

Fig. 4(a) shows the reduction in the water flow in each subwatershed. The results show that in the areas that contain 30% *Miscanthus*, the annual water flow reduction compared with the BAU scenario ranges from 3 to 9%, whereas the nitrate-N reduction ranges from 24 to 41%. The reduction in flow and nitrate-N load in the upstream subwatersheds also affects flow and water quality in the downstream subwatershed, as shown in Fig. 4.

Subsystems' Interdependency

To demonstrate the effect of the subsystems' interdependency on the overall system performance, the model was applied in four additional scenarios, each imposing a direct change on one of the subsystems and indirectly affecting the other subsystems and the entire system. The first scenario is run with the land-use subsystem by allowing 60% of the land to be converted to *Miscanthus* (instead of 30% under the RFS scenario). This scenario is denoted as the RFS 60% scenario. This new setting allows the model to allocate more land to *Miscanthus*, and the *Miscanthus* supply is

Table 1. Summary of Costs and Revenues

Cost and revenue items	BAU	RFS
Revenue: Total (US$ billion)	29.02	31.30
Corn	9.49	8.50
Soybeans	7.82	7.01
Ethanol	9.06	13.38
DDGS	2.65	2.41
Cost: Total (US$ billion)	11.31	14.95
Refinery capital—cellulosic[a]	0.00	0.86
Refinery capital—corn[a]	0.39	0.36
Refinery operation—cellulosic	0.00	1.23
Refinery operation—corn	2.12	2.62
Crop production	8.26	9.01
Crop transportation	0.33	0.60
Biofuel transportation	0.07	0.13
Transportation infrastructure[a]	0.14	0.15
Profit: Total (US$ billion)	17.71	16.34

[a]These cost items account for the time of the investment and the lifetime of the facility.

therefore larger in some areas of the watershed. Unlike the RFS scenario, in which all land parcels in the northern part of the watershed contain 30% *Miscanthus* at the end of the planning horizon, the RFS 60% scenario concentrates on *Miscanthus* cultivation in the northeastern part of the watershed.

This direct change in the land-use subsystem changes the solutions of other subsystems such as the refineries' locations and sizes, traffic congestion in the transportation network, transportation infrastructure expansion/renewal, and the flow and nitrate load in the watershed. Fig. 5(a) shows the changes in the system development in year 10 as a result of the change in the land-use subsystem. The cellulosic refineries move to the vicinity of the area with more intensive production of *Miscanthus*. Additionally, the cellulosic refinery chosen under RFS 30% in the southwest of the watershed is no longer constructed because the northern area can meet the requirement for *Miscanthus* after allowing for 60% land conversion. Corn refinery locations and sizes in the north of the watershed are also changed compared with the RFS scenario.

In the RFS 60% scenario, the model aggregates the two corn refineries from the RFS scenario into one large refinery located outside the *Miscanthus* supply area. The cellulosic refinery activity in the *Miscanthus* area increases traffic volumes compared with the RFS scenario as shown in Fig. 5(a). This traffic volume increment explains the relocation of the corn refinery and aggregates the two corn refineries in the north into one large refinery as suggested by the model. The model also places the cellulosic refinery in the southwest of the watershed chosen under the RFS scenario, so that the traffic congestion added to that area is relieved. However, comparing the traffic volumes surrounding the southeast cellulosic refinery shows that the traffic volume increases as a result of the increase in the refinery capacity.

The model results also demonstrate the importance of including the transportation infrastructure expansion/renewal in the transportation subsystem. A no-expansion scenario in which the model runs without the ability to increase the transportation infrastructure capacity is also considered. However, this scenario needs to meet the LOS by avoiding biofuel activities in the links experiencing congestion by public traffic. The lack of infrastructure development will force the model to route raw materials and product shipments to links that do not exceed the prespecified LOS by public traffic. This changes the optimal refinery locations and sizes in highly congested areas. Once the refinery locations change, the land allocation will also change and raw material supply becomes closer to the new refinery sites.

In addition, changes in land allocation will result in changes in the flow and nitrate load regime in the watershed. Fig. 5(b) shows the development of the system at the end of the planning horizon in the no-expansion scenario. As shown in Fig. 5(b), most of the congested links is located in the south of the watershed. Comparing the system development under the RFS scenario, two of the corn refineries, which are built in the vicinity of congested links, are omitted in the no-expansion scenario, and the capacity of these refineries is distributed among other corn refineries. The development of the cellulosic refinery does not change, compared with the RFS scenario; the routing for the selected location does not use the congested links.

Considering the factor of congestion affects the optimal solution of the integrated system. A no-congestion scenario, without LOS constraints, was also tested. The model still takes the transportation cost into account as the refinery location is being decided. Therefore, locations and routing decisions are made regardless of potential congestion; transportation costs increase only as a function of travel distance. The optimal development at the end of the planning horizon is shown in Fig. 5(c). Without congestion considerations, the model solution suggests building two refinery plants in highly congested links, which in turn result in higher congestion in these links. These results demonstrate the importance of including congestion and travel costs as a function of the distance between the supply and demand zones.

Finally, the effect of the environmental subsystem on the overall system is demonstrated by imposing a hypothetical maximum monthly nitrate load from the watershed. Examining the nitrate load

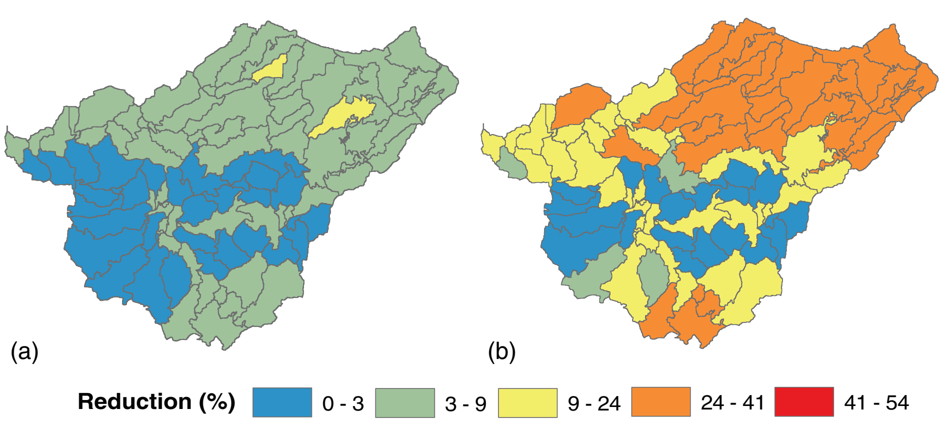

Fig. 4. (Color) (a) Flow; (b) nitrate yearly reduction in year 10 compared with BAU

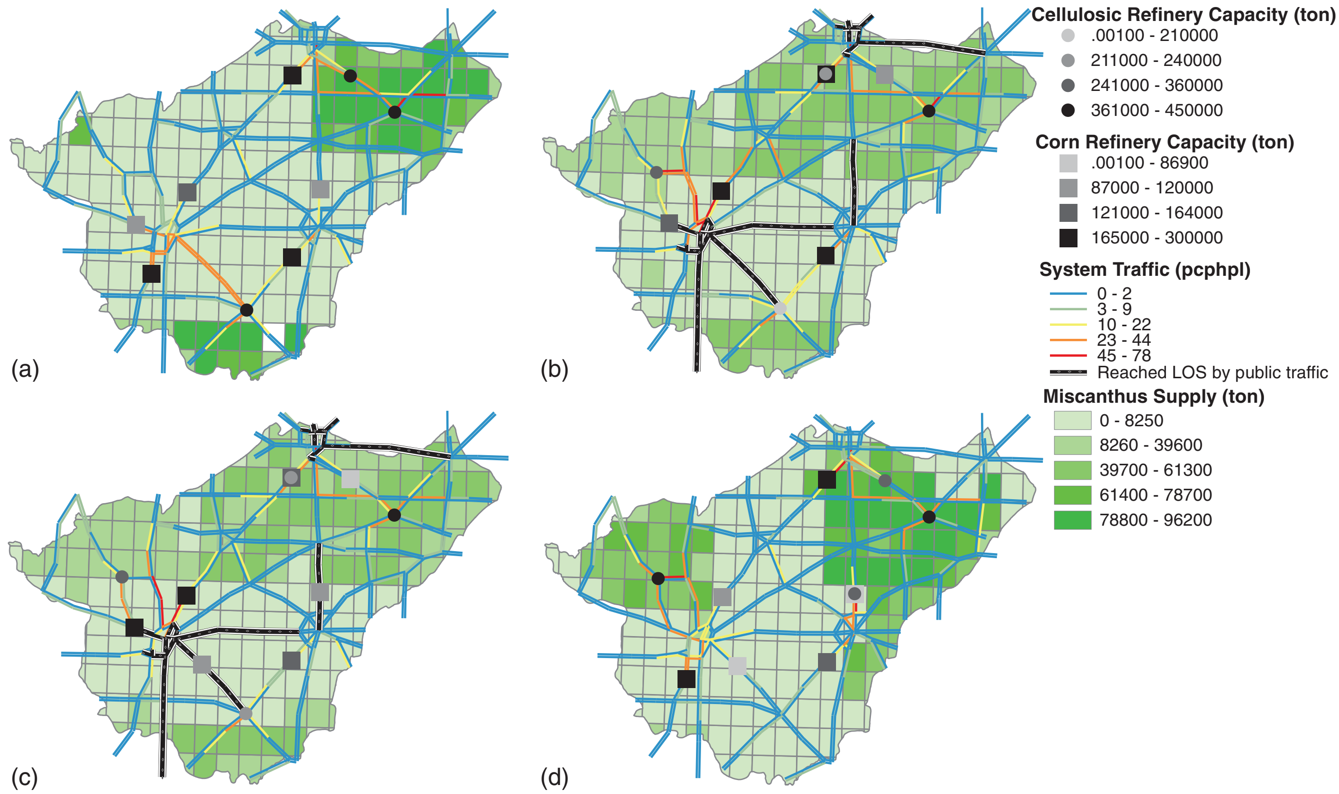

Fig. 5. (Color) Refinery development and *Miscanthus* supply in year 10 for (a) RFS mandate with 60% maximum *Miscanthus* fraction; (b) scenario without infrastructure investment allowed but with LOS constraints; (c) scenario without infrastructure investment and without LOS constraints; (d) scenario with maximum monthly nitrate load from the watershed

in the watershed outlet shows that the maximum load is obtained in May, which coincides with corn fertilization application and the maximum monthly flow in the watershed. For this scenario, the monthly maximum nitrate load constraint is set to reduce the critical monthly load (in May) by 20% of the nitrate load under the RFS 60% scenario.

More *Miscanthus* results in less nitrate pollution (Smith et al. 2013); therefore, the optimal biofuel development plan under the environmental scenario relies more on *Miscanthus* for ethanol production. Fig. 5(d) shows the *Miscanthus* supply and system development under the environmental scenario. The results show that the *Miscanthus* planting pattern is different from the pattern presented in the RFS 60% scenario—not only with regard to the *Miscanthus* planting area in the watershed, but also with regard to the locations chosen for the *Miscanthus* supply.

Under the RFS 60% scenario, the north and south of the watershed are the favorable locations for *Miscanthus*. However, under the scenario with the water-quality constraint, *Miscanthus* is planted in the west, closer to the outlet. This area has high potential nitrate load if it is planted with corn and soybeans (according to the estimates obtained by a hydrologic simulation model). In summary, imposing environmental constraints will change the biofuel development plan dramatically. Refineries' location and size, transportation loads, and land allocation are changed as a result of environmental constraints.

Summary and Conclusions

A model was developed to determine the optimal development of interdependent systems related to the biofuel development over a multiyear planning horizon. Unlike a subsystem-level analysis, in which the effects of one subsystem on the others may be overlooked, the developed integrated modeling approach results in a SoS model that is essentially based on the formulation of interdependence among the subsystems.

The supply side of the model is represented by the land-use subsystem, in which the model determines the optimal allocation between conventional crops and dedicated energy crops. The land-use subsystem is linked to food demand zones and biofuel refinery subsystems by transportation network (transportation subsystem). In the refinery subsystem, the model determines the optimal locations and sizes in addition to operation and capacity expansion of the refineries along the planning horizon. For the transportation subsystem, the model determines the optimal routing schemes and infrastructure-capacity expansion along the planning horizon. Land-use and refinery subsystems are connected to the watershed subsystem; therefore, decisions on land use, refinery, and transportation subsystems are taken under environmental constraints on water flow and nitrate load in the watershed.

Flexibility of model building and the ease of use are achieved by a user front-end processor that receives system data and automatically builds the model without the need for further programming—even for a different case study. This enables easy evaluation of the various scenarios by an interactive mode, which is important for building such a SoS model with an interdisciplinary research team through the conceptual and theoretical, numerical, and implementation stages of the model development.

The model demonstrates the interdependence of the subsystems through its application at the Sangamon watershed. The results quantify the importance of the integrated modeling approach in analyzing questions involving multiple interdependent infrastructures. A change in one subsystem (e.g., the land-use subsystem) propagates into the entire subsystems. Thus, the subsystem interdependencies play a key role in determining optimal system development. For example, changing the *Miscanthus* maximum conversion fraction from 30 to 60% has a tremendous effect not only on land allocation but also on cellulosic and corn-refinery capacities, locations, long-term development, and transportation and environmental subsystems. The transportation infrastructure expansion/renewal and the congestion considerations are both key factors in shaping the final solution. Omitting these elements from the modeling framework by exclusively including transportation travel cost as a function of distance will result in an inadequate development plan.

The effects on environment and infrastructure systems are externalities to societies as a result of biofuel development. Thus, policy mandates in some emerging biofuel production countries may drive biofuel production in an unsustainable fashion without being buffered by policies to minimize, e.g., potentially disastrous congestion and water pollution, in certain areas. In the context of interdependent infrastructure systems that interact with socioeconomic and environmental systems, every policy decision may introduce risks that impose new burdens on other sectors of the economy or external costs on society; the case with biofuel development as presented in this paper is just one example.

Acknowledgments

This work is funded by the U.S. National Science Foundation (NSF) (U.S. NSF grant EFRI-083598 led by XM Cai). The authors thank the anonymous reviewers for their helpful comments to the early version of this paper.

Supplemental Data

Variables and parameters in the SoS-Biofuel model, input data for the SoS-Biofuel model, and abbreviations are available online in the ASCE Library (http://www.ascelibrary.org).

References

ArcGIS version 10.1 [Computer software], Redlands, CA, Esri.

Bai, Y., Hwang, T., Kang, S., and Ouyang, Y. (2011). "Biofuel refinery location and supply chain planning under traffic congestion." *Transp. Res. Part B*, 45(1), 162–175.

Boulos, P. F., Lansey, K. E., and Karney, B. W. (2006). *Comprehensive water distribution systems analysis handbook for engineers and planners*, MWH Soft.

Chen, X., Huang, H., Khanna, M., and Önal, H. (2011). "Meeting the mandate for biofuels: Implications for land use and food and fuel prices." *NBER Working Paper No. 16697*, National Bureau of Economic Research (NBER), Cambridge, MA.

Chen, X., and Önal, H. (2012a). "An economic analysis of the future U.S. biofuel industry, facility location, and supply chain network." ⟨http://ssrn.com/abstract=2084313⟩ (Sep. 20, 2014).

Chen, X., and Önal, H. (2012b). "Modeling agricultural supply response using mathematical programming and crop mixes." *Am. J. Agric. Econ.*, 94(3), 674–686.

David, M. B., Wall, L. G., Royer, T. V., and Tank, J. L. (2006). "Denitrification and the nitrogen budget of a reservoir in an agricultural landscape." *Ecol. Appl.*, 16(6), 2177–2190.

Davis, G. A., and Sanderson, K. (2002). "Building our way out of congestion? Highway capacity in the twin cities." *Rep. No. 2002-01*, Minnesota Dept. of Transportation, MN.

Eksioglu, S., Li, S., Zhang, S., Sokhansanj, S., and Petrolia, D. (2010). "Analyzing impact of intermodal facilities on design and management of biofuel supply chain." *Transportation Research Record 2191*, Transportation Research Board, Washington, DC, 144–151.

Energy Independence and Security Act (EISA). (2007). *Energy Independence and Security Act of 2007*, ⟨http://www.govtrack.us/congress/bills/110/hr6⟩ (Sep. 20, 2014).

Energy Information Administration (EIA). (2010). *Annual energy outlook 2010 with projections to 2030*, Washington, DC.

Foulds, L. R. (1976). *Critical link identification in a network*, Massey Univ., Palmerston North, New Zealand.

Hajibabai, L., and Ouyang, Y. (2013). "Integrated planning of supply chain networks and multimodal transportation infrastructure expansion: Model development and application to the biofuel industry." *Comput. Aided Civ. Infrastruct. Eng.*, 28(4), 247–259.

Heller, M. (2001). "Interdependencies in civil infrastructure systems." *Bridge*, 31(4), 9–15.

Huang, Y., Chen, C. W., and Fan, Y. (2010). "Multistage optimization of the supply chains of biofuels." *Transp. Res. Part E*, 46(6), 820–830.

Kang, S., Önal, H., Ouyang, Y., Scheffran, J., and Tursun, D. (2010). "Optimizing the biofuels infrastructure: Transportation networks and biorefinery locations in Illinois." *Handbook of bioenergy economics and policy, series: Natural resource management and policy*, Vol. 33, M. Khanna, et al., eds., Springer, New York.

LeBlanc, L. J. (1979). "Global solutions for a nonconvex, nonconcave rail network model." *Manage. Sci.*, 23(2), 131–139.

Magis, K. (2008). "Community resilience measurement protocol: A system to measure the resilience of forest-based communities." *National Rep. on Sustainable Forests*, U.S. Dept. of Agriculture, Forest Service, Washington, DC.

Mahmudi, H., and Flynn, P. C. (2006). "Rail vs truck transport of biomass." *Appl. Biochem. Biotechnol.*, 129(1–3), 88–103.

McIsaac, G. F., David, M. B., and Mitchell, C. A. (2010). "Miscanthus and switchgrass production in central Illinois: Impacts on hydrology and inorganic nitrogen leaching." *J. Environ. Qual.*, 39(5), 1790–1799.

Ng, T. L., Cai, X., and Ouyang, Y. (2011). "Some implications of biofuel development for engineering infrastructures in the United States." *Biofuels, Bioprod. Biorefin.*, 5(5), 581–592.

Ng, T. L., Eheart, J. W., Cai, X., and Miguez, F. (2010). "Modeling Miscanthus in the soil and water assessment tool (SWAT) to simulate its water quality effects as a bioenergy crop." *Environ. Sci. Technol.*, 44(18), 7138–7144.

Richard, T. L. (2010). "Challenges in scaling up biofuels infrastructure." *Science*, 329(5993), 793–796.

Roess, R. G., McShane, W. R., and Prassas, E. S. (1998). *Traffic engineering*, Prentice Hall, Upper Saddle River, NJ.

Searcy, E., Flynn, P., Ghafoori, E., and Kumar, A. (2007). "The relative cost of biomass energy transport." *Appl. Biochem. Biotechnol.*, 137–140 (1–12), 639–652.

Smith, C. M., et al. (2013). "Reduced nitrogen losses following conversion of row crop agriculture to perennial biofuel crops." *J. Environ. Qual.*, 42(1), 219–228.

Tishler, A., and Zang, I. (1982). "An absolute deviations curve fitting algorithm for nonlinear models." *Optimization in statistics: TIMS studies in management science*, Vol. 19, S. H. Zanakis and J. S. Rustagi, eds., North Holland, Oxford, NY.

Tittmann, P. W., Parker, N. C., Hart, Q. J., and Jenkins, B. M. (2010). "A spatially explicit techno-economic model of bioenergy and biofuels production in California." *J. Transp. Geog.*, 18(6), 715–728.

Unnikrishnan, A., Valsaraj, V., and Damnjanovic, I. (2009). "Design and management strategies for mixed public private transportation networks: A meta-heuristic approach." *Comput. Aided Civ. Infrastruct. Eng.*, 24(4), 266–279.

Wright, M., and Brown, R. (2007). "Establishing the optimal sizes of different kinds of biorefineries." *Biofuel, Bioprod. Bioprocess.*, 1(3), 191–200.